*Statistical Science*
1993, Vol. 8, No. 4, 356–409

# Is Industrial Statistics Out of Control?

**David Banks**

*Abstract.* This paper is a critique of the current direction of industrial statistics in the U.S. The sweep includes not just the toolkit of statistical methods most often employed in industry, but also international competitiveness, the corporate climate in which statistical solutions are sought and the educational process which trains applied statisticians. Much is found that is good, but bland endorsement does little to advance a field. Therefore most of the paper is deliberately iconoclastic. Almost all of the extreme viewpoints are rooted in the author's direct experience, working with statisticians and managers across a range of industries. Some of the perspective must be attributed to knowledgable gossip, from friends and students who are now employed by various companies. Incidentally, the paper reviews four new statistical textbooks for engineers and an edited volume of papers on experimental design for industry.

*Key words and phrases:* Total Quality Management, control charts, statistical education, consulting.

## 1. INTRODUCTION

Industrial statistics is a growth industry. Everyone reading this paper has surely received a small mountain of mail on the subject. There are advertisements for short courses on control charts, Total Quality Management (TQM), Taguchi methods and statistical process control (SPC); there are flyers for books, periodicals and newsletters and mass mailings from concerned members of the statistical profession. Recently, our journals have begun to devote more serious attention to this area; we have developed new professional jargon and an entire subgenre of after-dinner speech has flowered on the subject of statistical contributions to international industrial competitiveness. Some of us have parlayed the boom into consultancies, funded research and corporate/university affiliations. But an examination of this trajectory of rapid development, to decide how well the interests of industry and academics are served, is overdue.

This paper speaks to three audiences. One group is traditional academic statisticians, whose research has barely flirted with industrial problems. Whatever interest they have in this paper is driven by the desire to see what all the noise is about and glean a quick understanding of recent themes in this area. The second group is those university statisticians who have bent their efforts to embrace the new enthusiasm; these people have written papers and proposals on industrial statistics and have attempted to develop industrial partnerships. The third group consists of industrial (or, slightly more generally, corporate) statisticians. In conversations with representatives of this group, I have sensed a delicate dilemma—they realize that much of the quality control climate that rains down on them from upper echelons is wishful hoopla, but they also recognize a kernel of value, which, if properly cultivated, could lead to measurable (but usually unmiraculous) product improvement. In some degree, this paper tries to address all three groups.

My own perspective is that of an academic statistician who has enjoyed broad consulting experiences and a relatively large number of interactions with various companies. Although industrial statisticians will surely have a better sense of practice and practicalities than I, it is my belief that we share a common concern that the tools of our profession be used appropriately, that the value of statistical methodology be accurately recognized and represented and that rising students receive the kind of education that will enable them to solve real problems in honestly useful ways. Most of this paper relates these issues to the manufacturing sector, but a large part is pertinent to service industries; very little aims at statisticians working in such areas as high finance.

The trigger for this paper was an invitation to review several textbooks for courses in industrial statistics. A number of idiosyncratic personal opinions were spawned, and in the course of conversations with various colleagues, both from academics and industry, it

*David Banks is Associate Professor, Department of Statistics, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213.*

emerged that these opinions were not as orthogonal to the rest of the world's as is the usual case. The perspective expanded when a fourth book, an edited volume on experimental design in industry, also arrived for review. Although this paper grew from a critique of some books into a critique of an entire field, specific attention is given to these five impetuses:

1. *Statistical Methods for Quality Improvement*, by Thomas P. Ryan (1989). Wiley, New York.
2. *Introduction to Statistical Quality Control*, 2nd ed., by Douglas C. Montgomery (1991). Wiley, New York.
3. *Engineering Statistics*, 2nd ed., by Robert Hogg and Johannes Ledolter (1991). Macmillan, New York.
4. *Statistical Quality Design and Control*, by Richard DeVor, Tsong-how Chang and John W. Sutherland (1992). Macmillan, New York.
5. *Statistical Design and Analysis of Industrial Experiments*, edited by Subir Ghosh (1990). Dekker, New York.

Much in these books is found to be good, but there are aspects which reflect current practices, in both industry and academics, that some may find troubling.

Section 2 describes the development and current state of industrial statistics in Japan and the U.S. Although some details are well known within the statistical process control community, one cumulative effect of the many quality salesmen has been to imbue the history with a mythological glow. This survey attempts a more realistic appraisal, which will inform subsequent comments on the potential contributions of academic statisticians.

After establishing the historical context and current tension, Section 3 offers short commentary on several topical areas in the industrial statistics purview: experimental design, control charts, Taguchi methods and process capability indexes. Section 4 points out new directions in which statistical researchers may tackle important industrial problems, and Section 5 examines the appropriate education for industrial statisticians and quality control engineers, which segues into an appraisal of the books under review. Section 6 offers some closing reflections on the different cultures of industrial and academic statistics.

## 2. STATISTICAL MYTHOLOGY

There has always been a warm symbiosis between research statistics and industrial applications. The births of nonparametrics, sequential analysis and experimental design testify to the generative influence of practical problems on the body of theoretical statistics. However, the recent spurt of activity has been driven by new influences, such as the emerging industrial primacy of Japan (with Europe close behind), the shift in the search for research funding toward corporate institutions and the expanding scientific perspective of industry executives.

Regarding Japan, several misunderstandings seem current, especially at the interface of the statistical and management communities. These include:

1. the belief that the Japanese explosion of industrial capability is attributable to their emphasis on statistical methodology;
2. the belief that they use sophisticated quality control methodology;
3. all the usual accusations and countercharges about unfair trade practices (we do it too), racist management (we do it too), and governmental support of key industries (they do it rather better).

This article speaks only to the first two points, since politicians, CEOs and Op-Ed columnists have cast so much light on the last.

The myth of the importance of statistical methodology has been fostered by the 1980 NBC television special "If Japan Can, Why Can't We?" and in various articles in the popular press. However, economic historians (in particular, Japanese economic historians) emphasize entirely different factors. Uno (1987), in a definitive study of Japanese industrial change between 1955 and 1985, never mentions trends in product quality or the contributions of industrial statistics. Instead, he identifies the following defining influences on the postwar Japanese economy:

1. The Korean War, pumping foreign dollars into Japan, created an industrial boom and the capital base for subsequent expansion.
2. From 1955 to 1964, government planners led and abetted a deliberate drive away from agriculture and light industry, towards heavy and chemical industries. There was intensive capital investment; economic growth was strong until 1961, then stagnated.
3. Between 1965 and 1970, the Japanese government liberalized laws governing trade and capital flow. The economy became substantially more open and international; economic growth resumed.
4. In the 1970s, the key domestic influence on Japanese corporations was the need to reduce environmental damage. Government and industry diverted substantial resources to this end. (At the beginning of the period, the ratio of pollution prevention investment to total private investment was about 3%; it peaked at 17% in 1975 and has recently declined to about 5%.)
5. In the 1980s, partly in response to the previous

> decade's environmental difficulties, government and industry committed themselves to a new plan for growth, relying upon high-tech and service industries. One measure of their success is that the majority of new international patents are granted to Japanese corporations.

Today, Japan's industries are expanding their investment abroad and systematically defining a global role. Their companies are becoming more multinational; transnational vertical growth, by owning foreign suppliers, assemblers or retailers, is now a standard strategy.

Uno's broad summary agrees substantially with other experts. Kosai and Ogino (1984) put some emphasis on the importance of Japan's well-educated and disciplined workforce, and on their relatively high rate of personal savings. Okita (1980) mentions that in the first stage, Japan got a boost from the returns available from their "peace dividend," occasioned by the demilitarization after World War II; he also details several fortuitous aspects of the economic climate (cheap resources, wage structures that favored personal savings) which fostered growth through 1961. Lincoln (1988) points out that the disruption of oil supplies in 1973–74 reinforced the concern with pollution that prompted Japan's eventual movement away from heavy industry. Saso and Kirby (1990) give considerable credit to the intelligent planning of the Japanese government; they are also the only authors who mention quality control. They spend three of their 203 pages on the subject, and their tone is lukewarm. They point out the high costs (workers are paid overtime wages for their participation in such TQM enterprises as quality circles, and there is considerable expense for on-site education) and suggest that the primary dividend is nothing more than some temporary morale building among the workers.

Several features of this historical sketch merit emphasis. First, the conventional sense of process control applies to the chemical and manufacturing industries that took root immediately after the Korean War. The comparable American industries are still struggling to regain a competitive standing in those domains, while Japanese planners have moved their attention to new arenas [recall the law of the multiplication of advantage which appears in the AI chess literature (cf. Good, 1977): *the further one is ahead, the more easily one's lead increases*]. Second, the Japanese have learned a lot about environmentally sensitive manufacture and waste management – this is crucial to the world's industrial future, and we will all benefit if they can market their expertise abroad. From a TQM perspective, pollution production is a process like any other, and hence susceptible to planned control. Third, the sustained growth in Japan over four decades reflects the intelligence and self-discipline of management more than the use of SPC, Taguchi methods or industrial statistics.

This is not to suggest that statistics has played no role, only that its role has been minor compared to the sculpting influences of widespread urban pollution, rising oil costs and the conjunction of capital with farsighted management. But even in the narrow domain of product quality, statistics has played a small role. From the Japanese perspective, industrial statistics is an arrow in the quiver of TQM techniques that is part of their arsenal of success. Other arrows include quality circles, consensus-style management, a corporate commitment to continual improvement (as opposed to U.S. "home run" corporate strategies) and clear divisions of process responsibility. Other weapons include well-educated workers, a population minded toward domestic investment and relatively consistent government policies for directing industrial growth.

It is noteworthy that Japanese success has rested almost entirely upon rudimentary statistical methods, all current in the 1930s. Instead of depending upon a few strategically placed statistical experts, as is usual in the U.S., their industries rely upon the ubiquitous use of simple procedures. Specifically, Japan's production engineers (Ishikawa, 1985) point to seven key tools that have advanced the quality of their output. These tools are sometimes called "The Magnificent Seven" among U.S. industries trying to transplant Japanese methods – the name is a joking reference to Hollywood's remake of Akira Kurosawa's film *Seven Samurai*. Specifically, the seven tools are the scatterplot, the checksheet, the Pareto diagram, the control chart, the histogram, the Ishikawa fishbone (or cause and effect chart) and primitive graphics, such as the bar chart, the pie graph and box plots.

The scatterplot, control chart, histogram and graphics are familiar to all statisticians, and need no explanation. The checksheet is a generic procedure for capturing data in a systematic way – it indicates that someone has put enough thought into the process that a form has been designed to minimize reporting errors and human reporting time. As far as I know, there is no formal definition of the checksheet.

In the manufacturing context, the Pareto diagram is a bar graph that indicates the percentage of occurrences of each of several kinds of product defect. The categories to which the bars correspond are ordered by descending frequency (except for the last category "Other," which consists of all defects not previously listed; this often breaks the monotonicity). The Pareto diagram easily extends to other situations; for example, one could use it to track different kinds of customer complaints. To illustrate the generality, consider the Pareto diagram in Figure 1. It might have been constructed by a whimsical journal editor who is examin-

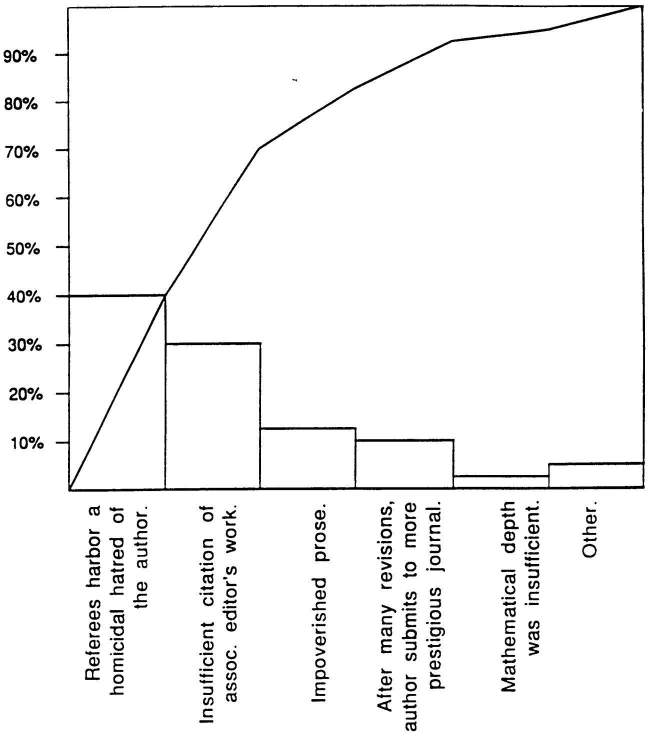


FIG. 1. *An example of the Pareto diagram. (This diagram is entirely fictitious; any resemblence to the editorial policy of a real journal is entirely coincidental.)*

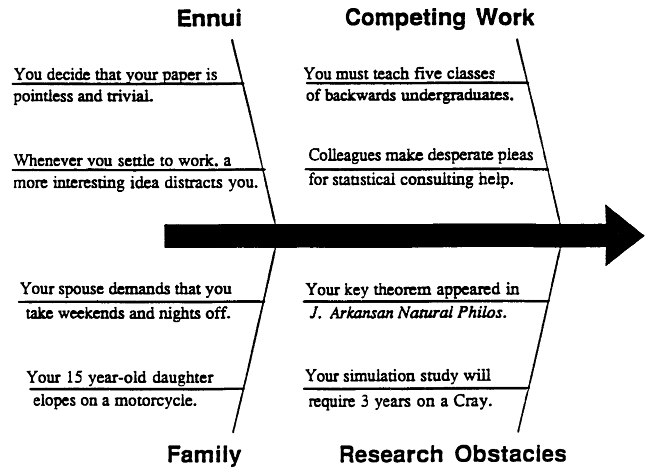


FIG. 2. *An example of an Ishikawa fishbone diagram, also known as the cause and effect diagram.*

ing the leading reasons that papers are rejected. Note that the cumulative percentage line has virtually no informational value [Tufte (1983) might call it chart-junk]; also, note that the one might want to study the possibility that a single item displays more than one kind of defect.

Obviously, the example is trivial, but my point is that the method is also trivial. Any business that does not use some comparably sensible way of identifying key problems will ultimately fall victim to natural selection. My chief concern regarding reliance upon this simple tool is that the problem becomes trivialized in the mind of the unsophisticated user. For example, I have heard an engineer insist that the Pareto diagram determines which defect problem must be solved first, but that is a gross oversimplification that ignores the cost of correction. In Figure 1, it is very difficult for the editor to enforce a spirit of professional fellowship among referees, but it is relatively easy to instruct the associate editors not to insist upon citational flattery. Similarly, an industry should tackle defects in the order determined by a cost-benefit judgment on obtaining a solution. Often it is better to solve the easy (or inexpensive, or nonpolitical) defects first, even if they are not the most numerous, and then re-examine the problem.

The Ishikawa fishbone, or cause and effect diagram, is a method for listing the possible causes for a particular problem, in a hierarchy that ranges from broad areas to very specific triggers. Sometimes it is used as an organizational aid in running a brainstorming session, to ensure that key features are not overlooked. Usually, it serves as a starting point for a systematic search for the cause of some problem; this problem need not be industrial – the folklore mentions one quality expert who used it to debug a failing marriage. Its name derives from its appearance, as shown in Figure 2. This particular example might have been drawn by an exasperated author, trying to discover why papers take so much longer to write than originally anticipated. It shows only two levels of hierarchical causes, represented as large and small tines branching from the central spine. In industry practice, there might be several more.

As a tool, the Ishikawa fishbone is susceptible to misuse, and the misuse is most likely to occur among those who trust it too blindly. In general, it only offers an avenue for organizing one's thoughts about a problem. It does not guarantee that a problem source has not been overlooked, and it does not highlight interactions that may occur between two problem sources. For example, in Figure 2, the author has somehow overlooked the possibility that the author might be lazy. Also, it is reasonable to guess that combinations of circumstances in the competing work tine and the ennui tine might reinforce each other's influence. In the U.S., it may be a fundamental difficulty that this tool needs team spirit to thrive; when used by a group, it is easy for team members to elaborate the tines in other members' areas of responsibility, but one imagines that their own may not receive the same careful scrutiny.

Regrettably, these seven tools have been embued with an aura of infallibility, and American industries that are self-consciously mimicking Japanese methods have largely adopted this attitude. At best, these seven tools are nothing more than a codification of common sense; however, their use has become so ritualized that many practitioners apply them without thought, which necessarily undermines the value. For example, I know of industries in which much discussion has been squandered on determining the appropriate location of upper and lower warning lines on the control chart and what kind of stopping rules should be mandated (e.g., seven points in a row above the centerline, four points in a row above a warning line, etc.). However, this level of distinction is enormously sensitive to underlying model assumptions, which typically receive only cursory examination. Essentially, all the value in the control chart accrues from the fact that a process expert periodically plots the recent history and thinks about it—any additional contribution from elaborate stopping rules is nearly negligible.

Recently, a second set of tools is climbing into the spotlight. These are the seven management tools for quality control: the affinity diagram, the relation diagram, the tree diagram, the matrix diagram, the process decision program chart, matrix-data analysis method and the arrow diagram [cf. Futami (1986) for detailed descriptions and a brief account of their history in Japan]. The first three have no statistical component and are not discussed here; they simply consist of rules for writing down sentence fragments and then drawing circles around some, and lines or arrows between others. Such sophisticated doodling can be valuable in problem solving, although it is unclear that a manager actually benefits from the level of formalism that these tools enforce.

Regarding the other members of the second seven, some short comments may be useful.

> **The matrix diagram**: This is simply an organized way to compel the manager to consider each possible pairwise combination of problem elements (in product manufacture, problem elements might include performance requirements, cost constraints and material capabilities, but the strategy generalizes readily to other domains). Its chief limitation is the ability of the user to anticipate all of the salient problem elements which must be captured in the rows and columns of the matrix. To an extent, it repairs a defect in the use of the Ishikawa fishbone, in that the user seeks interactions among the problem elements.
>
> **The arrow diagram**: In the West, this graphic is known as the PERT (program evaluation and review technique) chart. This is minimally quantitative, but its value in juggling complex schedules for subtasks in larger projects has long been recognized in MBA curricula.
>
> **The process decision program chart**: One can regard this as a prospective flowchart. The manager uses it to anticipate future problems and appropriate responses. It begins at the start of the process and moves sequentially through each important decision point; at each decision point, the diagram branches according to the decisions that might possibly be made. The result is a graphical game plan that covers the major contingencies.
>
> **Matrix-data analysis method**: This is just principal components analysis. As the practice in Japan is described, it is unsophisticated by statistical standards. Much attention is spent on the correlation matrix, but none is given to the validity of the underlying assumptions. There is considerable emphasis on graphics, such as the examination of 2-dimensional plots of the data on the space spanned by the first two eigenvectors. It is remarkable that Japanese management is making primitive use of a reasonably sophisticated statistical perspective; however, I doubt that their training enables good inference. It may be the statistical equivalent of letting children play with a loaded gun.

On balance, there is value in these tools, as long as no one takes them too dogmatically. Japanese engineers who write papers for Japanese journals claim that the methods work well, and perhaps they do in Japan; however, there is the potential for considerable selection bias in their reports. It is very uncertain that U.S. management can achieve significant benefit from these methods. At best, much of their value results from the Nipponification of statistical ideas, making them palatable to managements eager to learn corporate Kung Fu.

It would be easy to suggest that American industry should leapfrog the antiquated Japanese methods, adopting visualization tools that employ dynamic color graphics to represent complex processes and using new wave multivariate procedures (such as MARS, AVAS and PPR as discussed in subsection 4.1). Often, this is what researchers suggest when seeking corporate support. However, these methods require substantial investment, both in new technology and in people trained in its use, and the marginal improvement in process understanding seems generally unlikely to justify the cost. I would guess that intelligent use of simple tools will achieve about 95% of the knowledge that could be obtained through more sophisticated techniques, at much smaller cost. Also, the simple tools can be applied more quickly to all problems, whereas the complex tools are unlikely to be ubiquitously used.

In addition to the 14 quality tools, U.S. firms are sprinting to adopt TQM. This is a kind of corporate

EST (EST is a consciousness-raising, self-help discipline that became briefly prominent in the 1970s); it has become as popular among upper management today as planned obsolescence was in the early sixties. There are dozens of definitions of TQM, but none with any mathematical precision. Essentially, the gist is that TQM refers to any company-wide plan to systematically boost the quality of their product. Connotations of TQM practice include greater emphasis on assessment of and adaptability to customer requirements, a broadened sense of the economic value of high-quality output, performance benchmarking against competitors, team problem solving and a tendency to empower the employee. In practice, companies tailor the details to their own needs, often combining weird jargon with a born-again belief in the universal applicability of the core philosophy.

There is nothing deep about TQM – any 14-year-old can immediately apprehend the key ideas (of course, the same can be said of the theory of natural selection, so simplicity is not equivalent to triviality). But despite and because of this simplicity, TQM has worked very successfully in diverse industries, and the nation's economic engine would doubtless run more smoothly if it were more widely employed. Some of the valuable aspects of TQM include:

- TQM correctly shifts corporate attention to the customer, at every level of the business. From the TQM perspective, each employee is a two-legged profit machine, interacting with a private set of customers. Often the customers are internal to the company – the stock clerk's customer is the foreman who needs a bin of bolts, and the foreman's customer is the sales agent who must market the product. Often there are multiple customers; a secretary's customers include the boss, people who read typewritten letters, and anyone who calls with an inquiry. Of course, there is no direct revenue from internal customers, but one presumes that closer attention is paid to Brownie points when making promotions, to reward employees who go the extra mile in satisfying colleagues' requirements.
- TQM self-consciously commits a company to the unending quest for process improvement. In theory, each person constantly scans their work environment for opportunities to better meet the customers' requirements. To further this goal, employees are trained in simple pseudo-statistical procedures (such as Pareto charts, Ishikawa fishbones and control charts) so that everyone can better recognize and embrace these opportunities.
- TQM emphasizes the need for discovering the customers' actual requirements. Historically, American industries have been notorious for putting their own convenience ahead of the end-user, and in a competitive environment, this is a short path to disaster. Therefore TQM typically develops a range of methods for soliciting customer input and measuring customer satisfaction, and these affect (impact) even the earliest phases of research and development.

These ideas should be completely obvious, and it is unclear whether TQM is intrinsically more effective than alternative theories of management. It could well be that the general success of this strategy simply reflects the Cooley or Hawthorne or Heisenberg principle (the name depends upon whether one is a sociologist, psychologist or physicist, respectively), and that all the productivity benefits that accrue from TQM could have been equally well realized by hyping any new theory of management, because the response is not due to the kind of manipulation, but simply the fact that a manipulation has occurred.

However, there are dangers in TQM. As implemented, it tends to be enshrined, and this stifles creative solutions. One fears that corporate executives walk their corridors clutching copies of Deming's *Out of the Crisis*, just as a generation of Chinese bureaucrats sought universal applicability from Mao's *Little Red Book*.

As an example of the ideology that overreliance on seemingly insightful soundbites from quality gurus can breed, consider the collision between the zero defect philosophy and acceptance sampling plans. The zero defect philosophy (cf. Crosby, 1991) maintains that since all defective product is a drain on corporate assets, the proper goal of a quality management program is to produce defect-free output, by building quality directly into the production process. In contrast, the acceptance sampling strategy attempts to "inspect in" product quality. A particular example is the Dodge–Romig plan, which guarantees an average outgoing quality level for the product by sampling outgoing lots, rejecting those with too many defectives, performing rectifying inspection of rejected lots and shipping the corrected product. In discussing these ideas, DeVor, Chang and Sutherland (1992) write:

> One of the by-products of the Dodge–Romig approach to quality control was the acceptance of the concept of an acceptable quality level (AQL), a target level of defective material, supposedly based on economic grounds. The most devastating effect of the AQL concept is that it promotes quality improvement up to an acceptable plateau beyond which further improvements appear to be unjustified economically. Scrap became an accepted part of the business, unfortunate but inevitable. Certainly, scrap reduction wars were waged from time to time, but by and large, scrap became an integral part of the production and business

> planning activity. Deming has clearly pointed out the fallacies in this way of thinking, but for many years (the 1930s through the 1970s) this basic approach to quality control prevailed. (p. 11)

This is an unreasonably optimistic view of the value of zero defect manufacturing and an unfair bashing of the average outgoing quality approach.

Regarding zero defect goals, there is always the problem of diminishing return in efforts at quality improvement. No company in the world has achieved zero defects, and none will spend infinite sums to attain it. There is always a point at which the cost of additional defect reduction outweighs the return, but strict advocates of zero defect production overlook this in their zeal for a slogan. Regarding acceptance sampling, these strategies do precisely what they claim. When the decision rules honestly reflect economic trade-offs, then acceptance sampling has broadly desirable features. In many instances, it may offer the only immediate way to maintain contractually obligated quality levels; moreover, it gives a regular check on outgoing quality which can guide subsequent efforts at quality improvement. There is no necessary contradiction between the zero defect strategy and the acceptance sampling tactic, but attempts to boil down the ideas into oversimplified principles have produced a spurious conflict.

There is a pervasive mistake (especially among people who teach short courses or write research proposals) that Japan is working at the leading edge of industrial statistics and that American firms can only compete by funding statistical research to produce the next generation of quality control. This is a useful error to propagate when seeking money, but in fact Japanese industry rarely employs sophisticated statistics. Typically, the Japanese use simple methods from the 1930s, such as control charts and histograms; the effectiveness of these antiques stems from their almost universal application, at every stage of the manufacturing cycle.

From the perspective of a research statistician, it is notable that no Japanese university has a department of statistics (although the Institute of Statistical Mathematics is specially empowered to grant Ph.D.'s). Itô (1979) indicates that the suggestion to establish one has been made on several occasions, but aside from a short-lived experiment at the University of Tokyo in the 1960s, it has never come to pass [Kitagawa (1983) points out that several Departments of Information Science were established at various Japanese universities during the 1960s, and these typically contain a strong statistical component]. The great Japanese mathematical statisticians have not emphasized industrial applications, but rather differential geometry, econometrics, efficient estimation and time series analysis.

The intellectual push behind the quality movement came from JUSE (the Japanese Union of Scientists and Engineers), which quickly seized upon Deming's ideas as a method whereby their group could contribute to rebuilding Japan's postwar economy. This organization founded the Deming Prizes, which became a catalyst for change in Japanese industries. Ken-ichi Koyanagi, the president of JUSE, invited Deming to visit Japan in 1949 to teach statistical quality control. When Deming accepted that invitation in 1950, he seems to have taken a two-pronged approach in putting his ideas across. On the one hand, he taught statistics to large numbers of the members of JUSE; on the other hand, he met with Ishikawa and other leaders of the Kei-danren (the Japanese association of senior management). Mann (1988) reports that this combination of contacts provided both the will at the top and the capability in the middle that was necessary to establish the quality revolution.

U.S. corporate leaders should learn from the Japanese example that their quality budget can best be spent by distributing simple statistics throughout their corporation (this includes internal paper flow, accounting, hiring, advertisement and personnel benefits, not just the processing stream). Chucking money at university statisticians is a laudable charity, but may be a breach of fiduciary responsibility unless it specifically targets industry problems that cannot be addressed by in-house talent. However, this is one of the few aspects of the American manufacturing debacle in which it is difficult to legitimately lay the entire blame at the feet of myopic management. So many academic statisticians have written such brilliant research proposals arguing the case for corporate sponsorship on grounds of quality competitiveness that our profession must surely shoulder some responsibility for misleading our prospective clients.

## 3. RESEARCH IN INDUSTRIAL STATISTICS: CONVENTIONAL TOPICS

Industrial statistics covers such a range of applications and methods that enumerative treatment is beyond this paper. However, there are several classical areas of research that deserve some special comment, for reasons indicated below.

### 3.1 Experimental Design

For decades, experimental design has been the hallmark training of the industrial statistician. Although the original theory was propounded for agricultural applications, industry rapidly became the major consumer and its problems led directly to the invention of the related areas of evolutionary operation and response surface methodology.

My sense, from a research standpoint, is that the bloom is off the experimental design rose. It is hard to see profoundly new ideas on this horizon – most of the attention is given to the working out of old ideas for relatively more complex cases. Mixture experiments, optimal design, sequential design and trend-free design have risen in their turn, and none of these has entirely fallen, but the publications in these areas have grown steadily more particular. Currently, the Latin hypercube design research (cf. Stein, 1987; Owen, 1992) is a wave that responds to recent concerns with high-dimensional computer models for complex industrial processes; perhaps it will temporarily reinvigorate the field.

As previously mentioned, one of the triggers for this survey was an invitation to review a collection of papers on the interface between modern experimental design research and industrial applications. That collection is *Statistical Design and Analysis of Industrial Experiments*, edited by Subir Ghosh (1990). Its preface says that "This book will be indispensable for everybody concerned with experimentation in industries and serve as a guide for students, instructors and researchers at colleges and universities." Although such an introduction is permissible hyperbole, the book fell short of markedly less grand expectations, largely because of the disjunction in interests between industrial statisticians and university researchers. To show how this occurs, we begin with a list of the papers and a short account of their content.

*Experimental Design for Product Design*, by G. Taguchi. The paper is diffuse and offers a blend of elusive philosophy (Section 1), low-level theory (Sections 2 and 3.1–3.4) and practice (subsection 3.5). It ties together many of Taguchi's themes, including signal-to-noise ratios, orthogonal arrays and parameter design, and climaxes with a practical problem in air gauge design.

*Designing Experiments in Research and Development: Four Case Studies*, by K. Kafadar. Kafadar writes to persuade people of the value of experimental design methods in the R&D phase of manufacture. This reviewer (and most readers of *Statistical Science*) already agree with her thesis, but she gives four solidly practical examples, of which the power-meter problem in Section 2 seemed most interesting.

*Biotechnology Experimental Design*, by P. D. Haaland. The first two sections are written at a super-low level, and reflect excessive saturation in the corporate TQM credo. The third section gives three examples, but these are less exciting than the examples contained in other chapters of the book. Statisticians are advised to skip this portion of the book.

*Expert Systems for the Design of Experiments*, by C. Nachtsheim, P. Johnson, K. Kotnour, R. Meyer and I. Zualkernan. Sections 1 and 2 give a cursory survey of the history of REX, STATPATH and BUMP and strategies that underlie the development of an expert system. Section 3 develops a process model for consultation by a human expert and describes a somewhat fuzzy attempt to validate that model. Section 4 describes the expert system MELAMPUS (named from Greek mythology's equivalent of Dr. Dolittle, on the grounds that Melampus was the first seer). This reviewer was not impressed by the paper and is skeptical of the value of MELAMPUS.

*The Effect of Ozone on Asthmatics and Normals: An Unbalanced ANOVA Example*, T. Lorenzen. Lorenzen describes an unbalanced, mixed effects design, focusing on the development of expected mean square terms. The analysis appears technically correct, but the application is not obviously industrial nor especially general. It is unclear why this book was an appropriate vehicle for the paper.

*Sensitivity of an Air Pollution and Health Study to the Choice of a Mortality Index*, D. Gibbons and G. McDonald. Previous studies have used regression techniques to examine the link between air pollution and mortality. Previous critiques of these studies have focused on the validity of the assumptions, the sensitivity of the results, the accuracy of the data and the complexities of the model-fitting task. This chapter extends those critiques, by examining the effect of changes in the definition of the mortality index (it turns out to be small). The technical features of the work are routine, and the paper's concerns are distant from the topic of the book.

*Mixture Experiments*, J. Cornell. This paper is an excellent review of issues and methods in the design of mixture experiments. It includes practical theory and good industrial examples.

*Response Surface Designs and the Prediction Variance Function*, R. Myers. This paper is an excellent review of the issues and methods in response surface analysis. It includes practical theory and good industrial examples.

*The Analysis of Multiresponse Experiments: A Review*, A. Khuri. This paper is a good review of the issues and methods in multiresponse experiments. The coverage of industrial applications is nearly nonexistent.

*The Role of Experimentation in Quality Engineering: A Review of Taguchi's Contributions*, V. Nair and A. Shoemaker. Aspects of Taguchi's methods have generated considerable controversy in the industrial statistics community. This paper outlines his ideas (more clearly than Taguchi himself in Chapter 1), presents an application from AT&T VLSI manufacture and then surveys the areas that have generated discussion. For statisticians, this is the best short account of Taguchi's work that I have encountered.

*SEL: A Search Method Based on Orthogonal Arrays*, C. F. J. Wu, S. S. Mao and F. S. Ma. The proposed SEL is an intriguing strategy for finding good response regions quickly. The paper describes the idea and illustrates its use in three applications. Many questions are unresolved, but this paper repays the reader's attention.

*Modern Factorial Design Theory for Experimenters and Statisticians*, J. N. Srivastava. Although the first section is painfully elementary, this paper turns into a very good review of several important results in modern experimental design. There is virtually no discussion of industrial applications.

*New Properties of Orthogonal Arrays and Their Statistical Applications*, A. S. Hedayat. For orthogonal array fans, this chapter is a tease. It is too brief to do the subject justice, but it gives an intriguing bird's-eye survey and lightly develops some new ideas, such as designs of fractional strength [readers are referred to Hedayat, Sloane and Stufken (1994) for a fuller treatment]. No industrial examples are given.

*Construction of Run Orders of Factorial Designs*, C.-S. Cheng. The theory behind this area is sufficiently abstract that it seems unlikely that the method will see much use in practice. This is unfortunate, since, in rare cases, it is plausible that industry could benefit from the technique. However, the effort needed to sell the strategy, devise the design and execute the analysis is so great that it is probably not cost effective for most practical problems.

*Methods for Constructing Trend-Resistant Run Orders of 2-Level Factorial Experiments*, M. Jacroux. The topic in this paper is similar to the previous paper, and the previous reservations apply.

*Measuring Dispersion Effects of Factors in Factorial Experiments*, S. Ghosh and E. S. Lagergren. This chapter considers three methods for measuring dispersion effects in factorial models, motivated largely by Taguchi's concern with variance minimization. A brief industrial example is given, taken from previous literature. It would be good if the comparisons among the methods were more explicit; in particular, this reviewer is concerned about model robustness. However, it appears that at least the first method offers a usefully simple approach and is unlikely to be unduly sensitive to nonnormality.

*Designing Factorial Experiments: A Survey of the Use of Generalized Cyclic Designs*, A. Dean. Slightly more attention to industrial datasets would have been welcome, but overall, this is a very good review of cyclic designs.

*Crossover Designs in Industry*, D. Raghavarao. Raghavarao states that cross-over designs have not been used in industry and proceeds to develop plausible applications and corresponding designs. The exegesis is moderately clear, and the section on nonparametric analysis was a welcome change from the model-centric theory in the preceding chapters.

To summarize the results of these synopses, this edited volume shows the usual vital few and trivial many (to lapse into a TQM buzzphrase). The vital few are very few and emphasize area reviews and applications over theoretical derivation. Essentially, what is good is not new, and what is new is not good.

Of course, part of this evaluation depends upon the viewpoint of the audience. I think the book is, regrettably, more successful in speaking to academic than industrial statisticians. Most industrial statisticians lack the time, background and management support needed to undertake the very complex experimental designs that are outlined in several of the book's chapters. Moreover, their practical sense may persuade them to put their resources into planning analyses that are robust to the violation of model assumptions, rather than clever minimalist designs that are sensitive to minor model failures.

In fairness, there are some industrial statisticians who will benefit from more than one or two of the chapters. Several of these people are, in fact, contributors to this volume. However, industrial statisticians would probably do their companies better service by teaching large numbers of low-level employees how to use simple statistics, rather than by spending the same amount of time developing a trend-resistant run order for a partially balanced incomplete block design with four associate classes.

On the other hand, an academic researcher would find several of the papers in this volume to be honest work, but the topics are not hot. There is no consensus on a central problem whose solution would affect the future research agenda. From that perspective, which is no fault of the book's but rather a natural result of the maturity of this field, the appeal of the collection is limited.

### 3.2 Control Charts

A different area of classical research concerns control charts. These were developed almost simultaneously by Shewhart (1926) in the U.S. and by Dudding and Jennett (1937) in England. (Both groups were using them as early as 1924, although formal descriptions did not arise in the statistical literature until later.) Control charts are a prime example of a good idea that has been corrupted by formal theorists and then resold to industrial practitioners.

The earliest control chart was the $\overline{X}$-chart; other workers quickly developed the $p$-chart, $R$-chart and similar variations. The next intellectual milestone was the CUSUM chart, invented by Page (1954); this looks at the cumulative sum of the deviations from the centerline, enabling more rapid detection of small shifts in the process. This was followed by the EWMA chart, an exponentially weighted moving average technique developed by Roberts (1959); it examines the residuals from a forecast of the process based upon the discounted past and therefore confronts the problem of trends in control chart data.

In the wake of these key ideas, there has been a subgenre of statistical literature that specializes in variations on control chart themes. People write papers about control charts for data that have double exponential, $t$, Weibull and other distributions [recent examples include Kaminsky et al. (1992) and Bannerjee and Rahim (1988)], or about average run length calculations (cf. Champ and Woodall, 1987; Davis and Woodall, 1988; Hawkins, 1992) or about control charts that track multivariate product features (cf. Crosier, 1988; Pignatiello and Runger, 1990). This theoretical particularization has steadily removed the procedure from its initial application as a broadly practical tool, sequestering it into the fold of arcane statistical techniques.

The result of this tendency to play off of other research rather than practical problems has been that many process engineers believe that $3\sigma$ limits on control charts have reliable physical significance. They are recklessly unmindful of the distributional and independence assumptions that underlie their use of the control chart and blindly apply dozens of rules (runs criteria, warning lines, etc.) for declaring out-of-controlness to the same set of data. Under the canonical assumptions, this increases the false alarm rate over the nominal level; more importantly, it obscures the real value of the control chart exercise, which consists of having a knowledgable process engineer regularly examine the trajectory of the product quality. If the expert sees a hiccup in the process the day after the foreman was shoved into the vat of raw chemicals, the expert knows not to stop the process and recalibrate the heaters—the source is clear and one just lets the impurities filter out. However, it drift emerges when the inputs are putatively homogeneous, a canny engineer may intuit a problem before the control chart formally signals. In such cases, meek subservience to the rule of the chart can hurt the process.

More recently, and more encouragingly, there has been some movement back toward realism. Reynolds et al. (1988) have pointed out that it is desirable to let the sampling rate in a control chart depend upon the recent history; that is, if the process is looking dodgy, inspect it more often. This type of result combines good theory with a laudable attention to practicality. Although the theoretical properties are, as usual, regretably sensitive to a busload of assumptions, the strategy is transparently sensible and should inform almost any application of control chart methodology.

From a research perspective, control charts are dinosaurs. I do not see issues there that are going to provide scope for heroic improvement. It is hoped that statisticians' slow disengagement will have a salutary effect upon industry practice, as engineers and quality control managers devote more attention to the performance of their process and less attention to the arcana of control chart theory.

### 3.3 Taguchi Methods

The Taguchi method (cf. Taguchi, 1986) emphasizes the use of experimental designs in identifying ways to minimize the mean squared error of product features that affect quality. His fundamental revolution was to dethrone the conventional acceptance of tolerance specifications, arguing that any deviation of a part's value from the specified target was a "cost to society." The ripples from that stone have spread internationally and spawned a growing body of work in robust parameter design (cf. Freeny and Nair, 1992), signal-to-noise ratios (cf. León and Wu, 1992), orthogonal arrays for experimental design (cf. Wang and Wu, 1992), accumulation analysis for categorical data (cf. Hamada and Wu, 1990) and other areas. Since Taguchi's writings are sometimes difficult, readers who want a brief introduction to the ideas might prefer the review given by Kackar (1985).

Although it is difficult to formally characterize the procedures developed by Taguchi and his school, key components include:

- emphasis upon designing the product to have robust quality in the teeth of inescapable variation in manufacture and usage;
- the use of a signal-to-noise ratio to measure deviations between current and desired process performance;
- the reliance upon inner and outer orthogonal arrays (experimental designs), corresponding to factors that are susceptible to manufacturing control and factors that reflect variation to which product quality should be insensitive, respectively.

The signal-to-noise strategy can be flexibly tailored to the situation—Vining and Myers (1990) report that

Taguchi has devised more than 60 such measures. These measures are not derived from information theoretic considerations, but rather respond to judgments about costs incurred as a function of the production process mean and variance with respect to a specific quality characteristic.

The Taguchi method has been controversial, partly because it has often been formulated very vaguely and partly because of criticisms levelled at it by various statisticians. In particular, Box (1985) points out that Taguchi's predilection for saturated or nearly saturated orthogonal arrays overlooks simpler experimental designs that often enable better handling of interaction effects. Also, Taguchi's strategy pays meager attention to sequential experimentation, which can be tremendously important when experimental runs are expensive or large numbers of factors must be considered.

Nonetheless, most industrial statisticians agree that Taguchi's perspective is valuable (although some grind their teeth in frustration over its adoption as a management shibboleth). The concentration upon designing the product to make quality easier to achieve, and the recognition that minimizing financial loss requires simultaneous control of the mean and the variance of the production process, are key ideas. Vining and Myers (1990) have recently recast the latter insight into a more standard statistical formulation, by viewing the problem as a dual-response surface analysis. Here one seeks to optimize with respect to one criterion (the mean) while fixing another criterion (the variance) at an acceptable level. It would probably be slightly closer to the spirit of Taguchi's intention if the variance were not held fixed, but rather allowed to vary so as to obtain the least cost compromise between the value of the mean and the value of the variance. This might be obtained by plotting a curve of the total cost as a function of both mean and variance.

The real danger in the ubiquitous Taguchi cult is that industry users may copy the inefficiencies as well as the insights. There is a large statistical toolkit of pertinent designs and alternative methods, and it would be unfortunate to overlook these on grounds of ideological commitment.

### 3.4 Process Capability Indices

A process capability index attempts to measure whether the error in manufacture is so great that an economically excessive amount of product falls outside the tolerance specifications. Several indices are available; the two used most commonly in the U.S. are $C_p$ and $C_{pk}$:

$$C_p = (\mathrm{USL} - \mathrm{LSL}) / 6\hat{\sigma}$$

$$C_{pk} = \min\{\mathrm{USL} - \hat{\mu}, \hat{\mu} - \mathrm{LSL}\} / 3\hat{\sigma},$$

where USL is the upper specification limit, LSL is the lower specification limit, $\hat{\mu}$ is the sample mean of the process and $\hat{\sigma}$ is the sample standard deviation of the process.

Clearly, the $C_p$ index does not involve the sample mean and thus implicitly assumes that process performance is centered at the midpoint of the upper and lower specification limits. Otherwise, simply by shifting the location, a process with an arbitrarily good (large) index could be generating arbitrarily bad amounts of noncompliant product. Since $C_{pk}$ avoids this presumption, it should be strongly preferred; note that it automatically handles the case of asymmetry in the upper and lower specification limits with respect to the process mean. However, in most ways, the $C_p$ and $C_{pk}$ indices are discouragingly similar—both are aimed at processes that have Gaussian variation, and both inherit the magic $3\sigma$ concept of acceptable behavior.

In Japan, a process is said to be capable if its $C_{pk}$ value is at least 1.33 (cf. Sullivan, 1984). Under the Gaussian assumption, this suggests that not more than 6 items per 100,000 are nonconforming. However, this Gaussian approximation is entirely unreasonable in practice, since it depends sensitively upon the tail behavior of the actual unknown distribution. In control charts, the fact that one averages several values when plotting a single point makes the Gaussian assumption marginally agreeable—in the process capability context, it is completely unsatisfactory. So management should be skeptical of estimated numbers of nonconforming items obtained from Japanese manufacturing processes having particular capability indices, and they should be skeptical of studies that compare different processes in terms of their capability indices. The value of a capability index lies in tracking improvement in a particular process and is untrustworthy when comparing the performances of different processes.

It can happen that a process is capable, but not in control, or that a process is in control, but not capable. For the $C_{pk}$ index, this occurs when the process standard deviation is small relative to the distance between the mean and the nearest specification limit, but the mean of the process is far from the target value. Conversely, a process can be strictly in control, but have large variance with respect to the separation between the upper and lower specification limits. I have been assured by industry statisticians that confusion of these concepts is rife within U.S. management.

The process capability index emphasizes that in most situations, significant improvement requires reduction in process variance. This perspective is one of the chief motivations for Taguchi's work in parameter design, where the goal is to design a process that is insensitive to perturbations in input streams that are difficult to control. Concomitant to this perspective is the hope that it is less difficult to shift the mean of a low variance process than to reduce the variance in a process that already has the proper mean.

As used, the process capability index is trying to serve three separate purposes, and this causes some interpretational difficulty. One purpose is to track improvement in the control of a particular process over time. The second is to compare the performance of different processes, and the third is to roughly estimate the number of defective parts produced. All of these are sensible questions to ask, but the index is not particularly useful regarding the second two.

The first purpose is approximately well answered by something like $C_{pk}$. Large numbers here are good rough indicators of the comfort zone of the recent operating history of the process with respect to the prescribed specifications. Conveniently, it is easy to calculate the index from quantities used in constructing the control chart.

It is much less clear that a process capability index is a reasonable guide for comparing different processes. When examining very similar processes, perhaps between economic competitors, it may have value. But if a manager wants to evaluate problem areas in a complex process with many subprocesses, the indices of the subprocesses may grossly mislead attention. A direct comparison implicitly assumes that the quality of the final product is equally sensitive to variation in each of the subprocesses, which seems very unlikely.

Respecting the third purpose, both $C_p$ and $C_{pk}$ are essentially mute. The conventional interpretation leans too heavily upon assumptions about independence and tail behavior to be useful. Nonetheless, it is important for management to monitor the numbers of out-of-specification parts closely, since changes in these extremes are harbingers of process degradation. In most cases, managers eventually receive information from the customer or assembler regarding noncompliant parts, and this seems the only appropriate basis for estimating change in the numbers of bad parts produced.

## 4. RESEARCH IN INDUSTRIAL STATISTICS: NEW TOPICS

The methodologies discussed in Section 3 are either intellectually stale or quickly becoming so. One does not expect new research in these areas to make astonishing discoveries that can provide enormous benefit to industry; rather, gains will be small and become smaller as researchers tackle increasingly narrow problems. Nonetheless, competitive industrial performance requires a strong understanding of those methodologies, both their potential and their limitations. To achieve this, companies rarely need academic researchers; rather they must have access to Ph.D. level statisticians internally, who may not be developing new theory, but can comfortably command the old. Usually the bottlenecks in effective corporate use of statistical techniques are managerial, operational and structural, rather than due to any lack of available technical expertise.

So what role is left for the researchers? Fortunately, there are dozens of statistical problems that have intrinsic mathematical interest and whose solution would provide enormous profit potential to industry; moreover, many of these problems (at some abstract level of formulation) arise across many industries. The examples listed below are not intended to be exhaustive or representative—they only reflect a sampling of the more provocative cases that have arisen in recent forays into industrial consulting.

### 4.1 High-Dimensional Response Surface Analysis

For many industries, automated systems regularly measure very large numbers of independent variables. In continuous manufacturing (e.g., the continuous pouring of sheet glass or crude oil fractionation), one typically records several hundred control variables, such as temperature, quantities of raw materials and pressure, and tens of quality variables, such as optical clarity, purity and boiling point. The same is true for batch-continuous processes (e.g., sheet aluminium or VLSI chip manufacture), in which blocks of raw material are subjected to a fixed sequence of steps, each of which has a continuous processing character.

The immediate purpose of this data collection effort is to maintain control of the production system. Processes that are bones of aggessive international contention tend to be unstable; otherwise, control is easy and profit margins are small (and there is little contribution a research statistician can make). Therefore research attention and funding focuses on processes in which small drifts lead rapidly to catastrophic quality degradation. Sensitive feedback loops are needed to regulate the process, and the tightness of control is often the key factor in attaining high quality product at competitive manufacturing cost. The control loops may depend upon the values of the output quality variables, but more often are based upon inline measurments, which are compared to target values. In very mature processes, or ones in which a sophisticated understanding of the physics or chemistry of each stage exists, it is sometimes possible to use feedforward control to guide drifting product back onto the target trajectory.

The secondary purpose of the data collected during manufacture is to inform future efforts at process optimization. The data are archived into superlarge datasets (cf. Banks and Parmigiani, 1992); e.g., one dataset from Alcoa captured 212 control variables and 16 quality variables and held 240 records for each day of operation. The usual hope is that quality engineers can apply the principles of response surface analysis to obtain better setpoints for the control variables and feedback loops used in subsequent production.

Classical response surface analysis works well when the dimensionality of the data is low. However, the "curse of dimensionality" implies that in all practical situations, the value of statistical techniques, no matter how clever, rapidly decreases as the problem becomes more multivariate and as less structure is known. The following are three nearly equivalent heuristic descriptions of the curse:

- *The complexity of the possible signal increases rapidly with dimension.* The simplest way to see this is to note that if one looks just at the class of polynomial regression models, then the number of possible models of degree $r$ developed from $p$ variables increases combinatorially fast in $p$.
- *In high dimensions, all data are sparse.* If one distributes 20 points uniformly on [0, 1], the gaps between the data tend to be very small; however, if one distributes 20 points uniformly on $[0, 1]^{10}$, nearly all points are very distant. Thus for a fixed sample size $n$, the amount of information about local functional relationships between the control and quality variables goes to zero as the dimensionality, $p$, increases. Figure 3 indicates how swiftly the expected distance grows in $p$; it shows the side of a cube that is expected to contain $q\%$ of data that are uniformly distributed in the unit cube in $\boldsymbol{R}^p$, for $p = 1,2,5,10,15$.
- *In high dimensions, nearly all datasets show generalized multicollinearity.* Suppose $p = 3$. Traditionally, a dataset is multicollinear if the independent variables concentrate on a plane or line in $\boldsymbol{R}^3$; one result is that parameter estimates in the fitted model are wildly unstable. For general $p$, the data are multicollinear if the independent variables concentrate on a lower dimensional affine subspace. In the context of nonparametric modeling, multicollinearity generalizes to concurvity, in which the independent variables concentrate on a smooth submanifold of lower dimensionality.

Hastie and Tibshirani (1990) and Scott and Wand (1991) provide additional details on these different facets of the curse.

The gist of the curse is that the hope of using archived industrial data to build a predictive model of process response is almost surely doomed to failure. Conventional linear modeling is hopeless. At best, one can try using new nonparametric regression methods such as MARS [for multivariate additive regression splines; cf. Friedman (1991)], GAM [for generalized additive models; cf. Hastie and Tibshirani (1990)], PPR

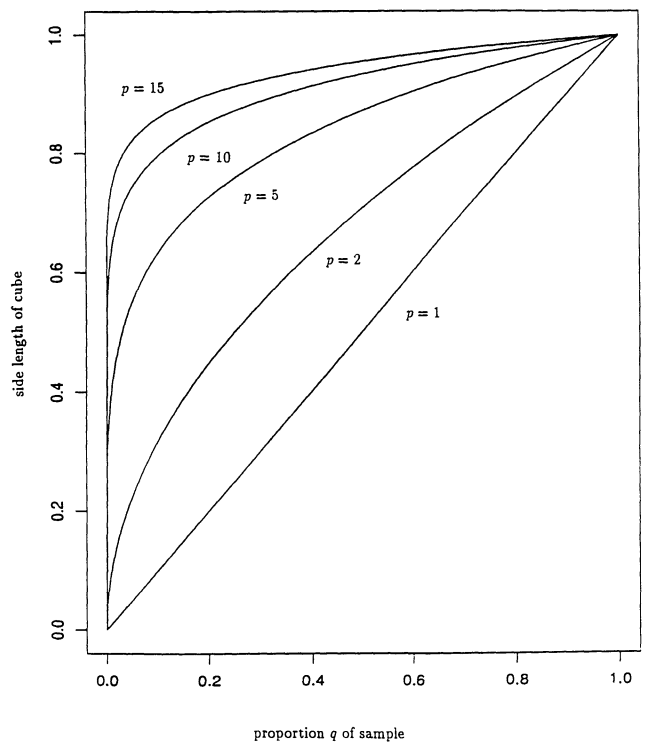


FIG. 3. *A graph illustrating the curse of dimensionality. The plotted functions show the side-length y of subcube in the p-dimensional unit cube that is expected to contain proportion q of a uniform random sample, for* $p = 1,2,5,10,15$.

[for projection pursuit regression; cf. Friedman and Stuetzle (1981)] and neural nets [Barron and Barron (1988) and Ripley (1993) give reviews of this area from interestingly different statistical perspectives], but the process structure must be serendipitously simple before these methods have any chance of succeeding, even with the superlarge datasets available. Also, it is worth emphasizing that most industrial data of this kind show strong serial correlation, partly because of feedback control, partly for all the usual reasons; this complication makes the new wave nonparametric methods even less likely to succeed.

However, the prospects are not unvaryingly dismal. One aspect of the response control heuristic is that it is unwise to waste explanatory simplicity in building good descriptions of process response in regions of poor quality. From a practical standpoint, the process engineer wants a model that better enables good quality output. It is reasonable to hope that in that relatively small region of control space, the model may be locally low dimensional and hence susceptible to an analysis that sidesteps the curse of dimensionality.

Researchers have already begun pushing this strategy, although not in an explicitly industrial context. Duan and Li (1991) and Li (1989) have pioneered a technique called SIR (sliced inverse regression) that attempts to reduce the effective dimensionality in nonparametric modeling by restricting attention to one subset (slice) of the data at a time, where the subsets correspond to quantiles of the response variable. Thus one might focus attention upon modeling the process structure using only the control data that gave the 10% best quality output; after all, why spend effort understanding exactly how to produce output with quality in the bottom 90%? So far, work in this area has focused upon performing principal component analysis on the slices (cf. Li, 1991). The door is open for more general approaches, perhaps by extending the new-wave techniques so that adaptive partitioning strategies can depend upon the response variable, by the more obvious device of applying the new wave methods directly to the slices, or by humbler practices, such as doing partial least squares or ridge regression on each slice.

In industrial applications, one expects that for a relatively mature process, the model fitted by a SIR method will find that most control variables are (locally) unimportant. These variables are the ones which are so well understood that they are tightly controlled about their target values. Few variables would show interesting main effects – if these effects were important, then the combination of constant international competition and decades of manufacturing experience would have already revealed their influence. The best hope for process improvement lies in identifying the few variables that show simple interactions, and industries that apply SIR analysis should be alert for these. Of course, if the industrial process is not mature, then there is greater scope for heroic change. This potential probably exists in such industries as VLSI wafer fabrication.

### 4.2 Software Reliability

Most competitive companies place great reliance upon millions of lines of software code. However, as every software manager is keenly aware, such code almost surely contains hundreds of hidden bugs; moreover, the code is difficult to maintain, and it is fraught with the potential for dangerous/expensive errors [cf. Humphrey (1989) or, for an earlier but wonderfully readable account, Brooks (1982)]. Writing a typical large program may cost decades of human-years. The code is so interactive that an error in one portion may affect many other portions, in entirely unpredictable ways. It is so complex that the number of logical paths can easily exceed the number of lawful chess games (approximately $10^{120}$), rendering exhaustive testing impossible. Software is unsmooth, in that small changes in code can cause enormous changes in the output – thus program maintenance is like blindman's buff in a minefield. It is as unique as artwork, since each program is written for a specific task, and so EVOP strategies that succeed when making ball bearings over and over again are nearly pointless for code; in some sense, there is only one chance to get it right. Finally, software is invisible. Most manufacturing products can be broadly inspected with the eye, a task for which humans are evolutionarily adapted; but we are not hardwired for the kinds of logic checks needed in software verification. These aspects distinguish the software quality problem from conventional quality problems.

There are many examples of serious software errors. For example, the American Society for Quality Control mailed out thousands of dues bills in 1992 which were unpayable because the membership ID was omitted. Also, the judicial software in Connecticut drew no jurors from the town of Hartford for three years, because the terminal "d" in the town's name extended into the status field, and thus all residents appeared to be dead. And in two separate instances in 1991, telephone service to 10 million customers in Washington, Pennsylvania, West Virginia, and Los Angeles was disrupted (I could not place calls for an entire afternoon) by what the *New York Times* (July 10, 1991) describes as a few lines of erroneous code in the Switching System 7 traffic management program written by DSC Communications.

Besides simple software errors, there is the broader problem of system performance. The nature and use of industrial software presents unique problems in operational reliability. For example, the difficulty in inspecting software opens the door to subtle sabotage or unprofessional work – a disgruntled employee of the *Encyclopedia Britannica* substituted the names of company executives for historical figures, and there is a persistent myth that penny round-off errors in the financial transactions of a large bank are secretly credited to the architect of their computer system. Also, software is so powerful that unless it is written to include elaborate sanity checks on input, minor input errors can have devastating consequences. For example, Salomon Brothers recently sold several million dollars worth of stock by mistake, apparently provoking a small Stock Market tumble in the process – a tired clerk had transposed the value of the stock with the number of shares to be sold, and the supervisor did not notice. Similarly, TRW's credit reports accidentally listed virtually every resident of Norwich, Vermont, as being delinquent in their property taxes, due to a one-keystroke clerical error.

Unlike most manufactured objects, software bugs are hard to apprehend through direct measurement. Catastrophes may be caused typographically, by logical errors on the part of the programmer, by sabotage, by simple human error during key inputs or by interactive failures between separate modules (e.g., two modules may each be logically correct, but the first passes information to the second more rapidly than the second can handle). The first type of error can be very hard to detect during inspection; the other kinds are nearly impossible to discover or prevent. Much of the difficulty arises from the tremendous complexity of a realistically large software program.

Any software module can be conceptualized as a directed decision graph. Nodes correspond to if–then branches, and edges correspond to direct manipulations on the inputs. Starting from a base state, the module takes a vector of inputs, then negotiates a series of if–then statements (nodes) to produce a vector of outputs. Since the number of possible paths in the graph increases combinatorially, enumerative verification of software accuracy is almost always impossible. Thus anyone who purchases or commissions software is buying a lottery ticket, and, at a minimum, needs to understand the odds.

Previous work in this area has been fragmented and diverse. Some of the key threads are as follows:

- Dalal and Mallows (1988) offer the most statistical treatment – it is a Bayesian analysis of the problem of deciding how long one should test software before release. They descibe a practical implementation, but the method is sufficiently mathematical that it requires repackaging before conventional industries will be able to avail themselves of its virtues. (By the way, as a nonindustrial digression, the decision problem they address can be mapped directly into the problem of deciding when to stop the simulated annealing algorithm in optimization.)
- Humphrey (1989), Cho (1987) and Musa, Iannino and Okumoto (1990) attempt to transport TQM ideas into the software arena. The details differ, but the first author emphasizes the process of software production, with inline measurements to check accuracy, control charts to track error rates for individual programmers, metrics for software quality and built-in protocols for software validation. The other authors put more of a statistical reliability spin on their treatment by including failure rate models and the associated parameter estimation. My own sense is that although these devices will, on balance, reduce errors in the code, they do not offer the sweeping solution that one hopes to find.
- Transparent proofs, a new mathematical technique described by Arora and Safra (1992), re-expresses calculations so that any error appears nearly everywhere. This has enormous potential for facilitating program checking, although the technique is still in its infancy and remains infeasible for routine application. A particular problem is that identifying the source of errors is hard, especially when the software program may involve multiple bugs. In a similar vein, there are methods for producing provably correct code (cf. Dijkstra, 1982), but these are curiosities rather than practicalities.
- Some people (Jewell, 1985; Goudie, 1990) have used capture–recapture methods to estimate the number of undiscovered bugs and an appropriate stopping rule for terminating the inspection. This has solid statistical roots, but does not directly help the producer in achieving process improvement.
- There is a smattering of statistical literature that looks at human factors in code quality, such as Humphrey and Singpurwalla (1990). In a similar vein, there have been some empirical studies of where errors are most likely to appear in code (cf. Shen et al., 1985); it appears that the dominant component in the number of faults is proportional to code length, but another component reflects the complexity of the function or the interface with other modules.

Humphrey (1989) indicates that the Japanese rely chiefly upon "software factories," in which highly modular code is produced. Programmers specialize in certain modules and tailor them to particular requirements. Over time, workers whose modules interact build up a good understanding of each other's requirements. This structure and the attention to the software production process is why Humphrey views the Japanese as the world leaders in software quality, despite the distance between this domain and their more conventional manufacturing arenas. If this modular production turns out to be the best avenue for swift improvement, then there are pertinent OC-curve methods, balancing Type I and II error, that have been developed for modular manufacture (Easterling et al., 1991). These might be exported to the software domain.

### 4.3 Partial Least Squares

Recently, especially in the chemical industries, there has been an organized challenge to traditional procedures for fitting multiple linear regression models. Instead of ordinary least squares or variable selection techniques, or the more sophisticated principal components regression and ridge regression, practitioners have been experimenting with partial least squares.

The partial least squares approach was developed by Wold (1966, 1975). It has many different flavors, but the most common implementation for multivariate regression is described by Helland (1988). The method is designed to work well in data-rich, theory-poor situations, such as those that commonly arise in industrial datasets. The framework is very similar to principal components regression; comparative studies of these and other methods appear in Stone and Brooks (1990) and Frank and Friedman (1993).

Suppose one has $n$ $p$-vectors of explanatory variables, $\mathbf{X}_1, \ldots, \mathbf{X}_n$, corresponding to the $n$-dependent values $Y_1, \ldots, Y_n$. Denote the centered (mean zero) explanatory vectors by $\tilde{\mathbf{X}}_i$, $i = 1, \ldots, n$, the centered vector of dependent values by $\tilde{\mathbf{Y}}$, and the $n \times p$ matrix whose $i$th row is $\tilde{\mathbf{X}}_i$ by $\tilde{\mathbf{X}}$.

To perform principal components regression, one uses $p^*$ pseudo-regressors; these are constructed according to

$$X^*_{ij} = \mathbf{c}_j{}^T\mathbf{X}_i \qquad \text{for } i = 1, \ldots, n \text{ and } j = 1, \ldots, p^*,$$

where the $\mathbf{c}_j$ is the normalized eigenvector corresponding to the $j$th largest eigenvalue of the sample covariance matrix $\mathbf{S} = \tilde{\mathbf{X}}^T\tilde{\mathbf{X}}$. Equivalently, $\mathbf{c}_j$ maximizes $\mathbf{c}^T\mathbf{S}\mathbf{c}$ subject to the constraints that $\mathbf{c}_j{}^T\mathbf{c}_i = 0, \ldots, \mathbf{c}_j{}^T\mathbf{c}_{j-1} = 0$. Using these $p^*$ pseudo-regressors, one then applies the ordinary least-squares method to find coefficients that minimize the squared error in fitting the data; this gives the principal components regression function. In applications, one wants $p^*$ to be smaller than $p$. Several methods are available for determining $p^*$; cross-validation is currently popular.

In contrast, for partial least-squares regression, one finds $p^{**}$ pseudo-regressors of the form

$$X^{**}_{ij} = \mathbf{d}_j{}^T\mathbf{X}_i \qquad \text{for } i = 1, \ldots, n \text{ and } j = 1, \ldots, p^{**}.$$

Here $\mathbf{d}_1$ is the normalized vector that maximizes $(\mathbf{d}^T\tilde{\mathbf{X}}^T\tilde{\mathbf{Y}})^2$, the squared sample covariance, and $\mathbf{d}_j$ maximizes the same quantity subject to the constraints that $\mathbf{d}_j{}^T\mathbf{S}\mathbf{d}_1 = 0, \ldots, \mathbf{d}_j{}^T\mathbf{S}\mathbf{d}_{j-1} = 0$. Note the close parallel to the constraints imposed in principal components regression. As before, one uses ordinary least squares on the pseudo-regressors to find the partial least-squares regression function. Again, one wants $p^{**}$ to be relatively smaller than $p$, and cross-validation offers a sensible basis for choosing $p^{**}$.

Partial least squares is a hot area in industrial statistics. There have been several studies in which it has performed very well compared to conventional techniques, and it is now widely used in chemometrics, especially for calibration problems (cf. Martens and Naes, 1989). Currently, considerable excitement has been generated by a combination of Bayesian analysis with extensive simulation reported in Frank and Friedman (1993). Their results strongly imply that variable selection methods perform very badly (almost as badly as ordinary least squares) and that ridge regression is marginally superior to partial least squares and principal components regression across the range of conditions examined. No doubt more will be heard on this topic, but it appears that the murmurs of dissatisfaction with variable selection methods commonly available to industry analysts via standard software packages have been justified, and industrial practitioners may be expected to continue their inclination to explore alternatives.

### 4.4 Geometric Conformance

Manufacturing industries are encountering competitive challenges in miniaturization. When making and assembling very small components, it is crucial that each part's geometry conform to tolerance specifications. (Tolerance specifications are usually defined in terms of Taylor's principle: a part geometry is out of conformance if its surface cannot be inscribed within a similar shape that is larger by a specified amount, or if it does not circumscribe a similar shape that is smaller by a specified amount.) Assessing this conformance is a key issue in modern inspection—one must do it rapidly and very accurately, at scales that are increasingly difficult to measure.

Until the last decade, part conformance would be crudely assessed by checking some key dimensions with a ruler; if precision were wanted, the inspector might use a dial gauge or calipers. Nowadays, leading-edge applications employ coordinate measuring machines (CMMs). These consist of a platform upon which the part in question is fixtured (this is a technical term, meaning that the part is placed in a standard position for measurement), and a computer-controlled arm, equipped with a measuring stylus, that moves over the surface of the part, touching it at various points. The touch-off points are recorded as $(x, y, z)$ coordinates and one uses these vectors to decide whether the part geometry is satisfactory. In the future, it may happen that this verification process can be done with an electronic eye, which is logically very similar to the CMM, except that the number of touch-off points is larger and one must combine different 2-D perspectives to reconstruct the 3-D shape.

The statistical issue in the geometric conformance problem is to infer the shape of the object from the measurements. The solution should incorporate the measurement error in the CMM device (since these are carefully calibrated, this problem offers an unusual instance in which the error covariance matrix $\Sigma$ is actually known), the errors in the fixturing of the part, and models for errors in the part manufacture. Chapman and Kim (1992) and Chen and Chen (1992) describe the use of spherical regression methods in addressing aspects of this problem, and Hulting (1992) proposes graphical techniques for analyzing fixturing errors. A complete solution, for realistically complicated part geometries, has yet to be found.

## 5. WHAT INDUSTRIAL STATISTICIANS SHOULD KNOW

Hogg (1985) made suggestions as to what engineers should be expected to know about statistics. (The momentum of that articulation carried Hogg and Ledolter through the vicissitudes of authorship, culminating in *Engineering Statistics*.) I think those recommendations serve well as a platform for what industrial statisticians should know.

It goes without saying that statisticians with doctorates should know everything. But for those at the Master's level, Hogg's committee would emphasize design of experiments, graphical methods, linear regression and response surface methodology; also, they should not appear embarrassingly ignorant when the coffee-urn conversation turns to acceptance sampling, control charts or W. E. Deming. Although this may not have been as current a topic when Hogg's committee met, a smattering of Taguchi's ideas are sufficiently fashionable to assist a green career. I think a grasp of the bootstrap and cross-validation is invaluable, and a good understanding of applied Bayesian analysis could be a useful ticket to future developments. At the micro level, every practicing statistician should know such tricks as the arcsine-square root transformation for proportions or Fisher's rule for combining the results of independent experiments.

It is fundamentally important for any new industrial statistician to be serenely competent with at least one major statistical package, and it is highly desirable that this package enjoy good graphics capabilities. No amount of theoretical proficiency will (at the Master's level) compensate for inadequate computational skills. Graphics are useful because senior statisticians in industry have assured me that upper management cannot deal with statistical issues that are not susceptible to simple visual renderings.

Industrial statisticians should be proficient problem solvers. In terms of their university training, this points up the desirability of building some exposure to statistical consulting into the Master's program. Obviously, the Master's programs are already tightly packed; nonetheless, graduates who can claim some experience in realistic consulting enjoy such enormous advantages in obtaining and maintaining employment that it seems an essential part of a competitive program. The immediate advantage appears on the resumé and in the kinds of letters of recommendation that the faculty can write. The more important advantage is that, in a well-constructed program, they have a chance to learn nontextbook skills that can give them a flying start in their career.

I digress briefly, on a personal soapbox: Boen and Zahn (1982) have spent considerable attention on the teaching of statistical consulting, stressing interpersonal transactions in establishing meaningful dialogues with the client. I find their perspective too Caliifornia-ish to recommend, but agree with their emphasis on good oral and written communications skills, over and above mere technical competence. Contrary to most students' beliefs, defining the problem can be the greatest service the statistical consultant provides; this task is often fluid, difficult and without unique solution. Unless students see this process for themselves, they usually think in terms of end-of-the-chapter exercises and struggle to find an exact match between a recently taught tool and the client's naive statement of the problem.

Sadly, amusingly, technical proficiency can be a liability in statistical practice. I am often surprised that mathematically brilliant statisticians can be so bad at consulting, but I have listened as a time series expert tried to cast a simple repeated measures design into a Box–Jenkins model, I have heard probabilists blithely make unrealistic distributional assumptions on no better grounds than personal convenience and I have seen Bayesians undertake analyses that their clients do not understand and cannot hope to persuade their colleagues to endorse. A senior statistician at a major petrochemical firm has mentioned world class statisticians whom he would not trust with his data and whose students he would never hire. The only real corrective is breadth and consulting experience and (sometimes) a complete change in one's *Weltanschuaang*, but I urge all practitioners to read Breiman (1985).

Finally, new industrial statisticians should be prepared to quickly build a very deep understanding of the industry that employs them. This emphasis upon the partnership between domain knowledge and good analysis is fundamental in statistics, but too easy to underemphasize in a crowded Master's curriculum.

Although it is rhetorically awkward to attempt to list topics that Master's level industrial statisticians need not know (it opens the door for endless quibbling clarifications), I strongly feel that there is no need to explicitly teach TQM. Of more conventional subjects, it seems that survey sampling is not ubiquitously used, nor are any of the more mathematically athletic aspects of time series analysis. Only a weak knowledge is required of continuous multivariate statistics, and there is even less demand for discrete multivariate (contingency table) methods. Most of formal probability may be quickly forgotten upon graduation, and much of inference and decision theory. Typically, one hopes that students retain the flavors of the ideas long after the details have faded and that their education will serve as a platform rather than a straitjacket. From this perspective, superficial exposure to a great range of methods (cluster analysis, ridge regression, CART, the bootstrap etc.) may be better education than the current drill in the mathematical consequences of sufficiency, Fisher information and the strong law of large numbers.

Regarding people with bachelor's degrees, their backgrounds are so diverse and their capabilities so varied that detailed specifications are difficult. In the modern world, for any career that requires a combination of competencies in statistics and some aspect of engineering, it is probably unreasonable to view a bachelor's degree as a terminal degree. Nonetheless, certain things are obviously important for good undergraduate preparation, and these clearly include an understanding of basic statistics (probability, $p$-values, confidence intervals and multiple regression). If there is additional strength in experimental design and some exposure to a computer package, that constitutes a strong starting point at this level. If the student were a statistics major with an interest in industry, then the topics list would include much more regression, applied statistics, computer skills, Bayesian methods and so forth. If the student were an engineer with an interest in statistical tools, then this core list is as much as one can reasonably expect.

Slightly tangentially, it is my impression that students at all levels should anticipate a credential deflation upon graduation. A person who is hired with a bachelor's degree in statistics typically finds an employer who expects only trainability and a skill set equal to what a good high school student ought to achieve, but rarely does. Similarly, the employer of a new Master's student often anticipates no greater knowledge than what a bachelor's degree in statistics should certify, and a Ph.D. student who takes a nonacademic position may not be called upon to use any statistical tools that had not been taught in the Master's program. Of course, employers may have a very different view of this, and surely they always hope to hire people whose capabilities exceed expectations.

Regarding undergraduates, several new texts have recently appeared that aim at the undergraduate engineer, and these were one of the prods that produced this review. The texts might also be useful in a low-level course for graduate engineers, or for a short course taught to working engineers, but all of them are too elementary to be used in a Master's program in statistics. The four texts considered are:

1. Thomas P. Ryan's *Statistical Methods for Quality Improvement* (1989). Ryan's book emphasizes control charts and experimental design and could be used for a first course in statistics for engineering students.
2. Douglas Montgomery's *Introduction to Statistical Quality Control* (1991). Montgomery assumes slightly more mathematical background than does Ryan, strongly emphasizes acceptance sampling methodology and aims at undergraduates pursuing engineering and management studies. Perhaps the second half could be used in a Master's course for engineers.
3. Hogg and Ledolter's *Engineering Statistics* (1991). This book offers a first course in statistics for engineers. Although it deliberately straddles the domains of statistics and quality management, compared to the other introductory texts in this sample it leans more heavily upon its statistical foot. It could be used as a text in a low-level graduate course for engineers that have had no prior exposure to statistics.
4. DeVor, Chang and Sutherland's *Statistical Quality Design and Control* (1992). These authors address seniors and Master's students in engineering. They assume no prior background in statistics and take a more strongly engineering perspective than Ryan, Montgomery or Hogg and Ledolter. However, the differences in statistical coverage between this and the other three basic texts are slight, although the tone sometimes varies.

None of these books requires prior exposure to statistics, but all assume a level of mathematical sophistication comparable to that provided by a two-semester calculus sequence. Hogg and Ledolter are somewhat more advanced than the rest (for example, they discuss expected mean squares in ANOVA tables), but it is nearly pointless to differentiate the others.

Each of the four books delivers the usual first-course coverage of topics in probability models, estimation and hypothesis testing, although DeVor, Chang and Sutherland tend to be relatively more cursory. The major surprise is that Montgomery and DeVor, Chang and Sutherland do not include linear regression, which seems a very unfortunate omission. Ryan and Hogg and Ledolter cover multiple linear regression and prediction intervals. Only Hogg and Ledolter treat Bayes' theorem (and this was absent in their first edition). Tolerance intervals as estimated from a sample are not usually treated in an introductory statistics course, but they are covered by Hogg and Ledolter; the other authors treat tolerance from an engineering perspective, in the context of design specifications. DeVor, Chang and Sutherland do not mention some of the less familiar probability models, such as the hypergeometric or exponential.

Regarding the specialty topics, each book offers extensive treatment of various flavors of control charts (cf. this reviewer's comments in subsection 3.2). Also, each has an initial chapter or two propounding the philosophy of quality control, and each devotes serious attention to experimental design up through fractional factorials. All but Ryan give fairly detailed treatment of response surface analysis, and all but Hogg and Ledolter offer some explicit treatment of Taguchi's ideas and process capability indices. Other divergences in coverage are that Ryan devotes considerable space to multivariate control charts, CUSUM charts, the analysis of means and evolutionary operation, whereas Montgomery emphasizes acceptance sampling and a managerial perspective. DeVor, Chang and Sutherland put emphasis on case studies with an engineering twist, while Hogg and Ledolter tend to provide usefully greater mathematical depth.

Regarding computational issues, Hogg and Ledolter and DeVor, Chang and Sutherland both include exercises that push the students to use a statistical package. Hogg and Ledolter do not specify a particular package, but their problems are so data intensive that any instructor who uses them should plan to provide software support. Upon request, DeVor, Chang and Sutherland provide free software tailored for their book to instructors who order their text. Montgomery does not emphasize software as a teaching tool, but points the reader to pertinent software reviews in the *Journal of Quality Technology* for each of the major topics he considers. Ryan's text can be covered without resort to any tool more sophisticated than a typical engineering student's scientific calculator.

In terms of efficient prose, Ryan covers all of the topics in the fewest number of pages, with Hogg and Ledolter a very concise second. The level of coverage is comparable for all four books, and entirely satisfactory for the target undergraduate (or junior graduate) audience.

All of the books have good exercise sets. Montgomery's problems tend to be more numerous and more deliberately couched in terms of real or hypothetical industrial applications than are Ryan's, which often have a philosophical feel. Hogg and Ledolter have both group project problems and exercises for individuals. The former are quite good, but the latter seem to have a somewhat narrow band of difficulty and often failed to fire my imagination. DeVor, Chang and Sutherland provide the largest number and greatest diversity of exercises, ranging from routine drill to the use of computer programs to open-ended questions about case studies. Depending on the application, different styles of exercises will be preferrable.

Ryan makes minimal use of real data, which is a refreshing departure from conventional thinking. Colleagues in engineering have assured me that their students have sufficient imagination that one need not provide real data to illustrate every variation of the factorial design. Moreover, one must search through heaps of industrial examples to find something uncluttered enough to use in a textbook – that kind of selection bias is likely to undermine the intended exposure to reality by making the world seem simpler than it is. However, at the other extreme of pedagogic practice, Hogg and Ledolter and DeVor, Chang and Sutherland make regular use of real examples, and it appears to work well. A particular strength of both is the inclusion of case studies. Montgomery takes the middle path. This reviewer is inclined to favor Ryan's minimalism, but imagines that most readers/teachers/students subscribe to the fashionable preference for case studies and real data.

Regarding the presentational styles of the authors, they are all admirably effective. My annoyance with Ryan's misuse of the adverb "hopefully" was offset by the strong sense of narrative trajectory that underlay his writing – this feature is particularly valuable for classroom use. Montgomery's style was direct and very clear. His prose was terse and declarative, giving the feel of a well-written handbook. Hogg and Ledolter's voice was somewhat closer to a conventional text in introductory statistics – it moved with purposeful pedagogy through each chapter's agenda, avoiding both dullness and digression. In contrast, DeVor, Chang and Sutherland were somewhat more vague – their sentences tended to be longer and hedged about with softening subordinate clauses and parenthetical commentary [this is an almost inescapable result of multiple authorship (except with unfortunately discursive writers)].

In the spirit of *Consumer Reports*, one of the mainsprings of quality consciousness in this country and a pioneer in the presentation of complex multivariate information, I summarize the performances of the four books in Figure 4. A fully darkened circle indicates excellence; an empty circle shows serious inadequacy.

Broadly, my sense is that Hogg and Ledolter's book is the most desirable text for a one-semester first course in quality control statistics for engineers. Ryan's book is a close second, largely because I am undaunted by his parsimonious use of real problems. Montgomery has written a valuable reference book for practitioners, and it could be used successfully in the classroom or industrial short course, especially if the emphasis were on acceptance sampling. DeVor, Chang and Sutherland's text probably requires a two-semester sequence and would probably be preferred by teachers who are themselves engineers rather than statisticians.

## 6. TWO CULTURES

Industrial statistics has an entirely different agenda than academic statistics. It concentrates upon economic results and ad hoc applications and often (but not as often as it ought) seeks simple solutions that are expedient if suboptimal. In contrast, academic statisticians need to prove new theorems, of almost any kind, and it often seems that the least practical and most impenetrable receive the best peer response. To academics, the economic progress of an industry or a nation is a distant consideration, except possibly when one is rationalizing a proposal for funding.

The rift goes deeper. Academic statisticians are drawn from all nations, spend sabbaticals in all nations and obtain students from all nations. The occasionally jingoistic tone of corporate rhetoric grates on their cosmopolitan vision. It sounds narrow and selfish, especially as it comes from managers whose own stupidity has squandered the vast capital and technological advantage they are now scrambling to recover. (Of course, this representation is extreme – most major

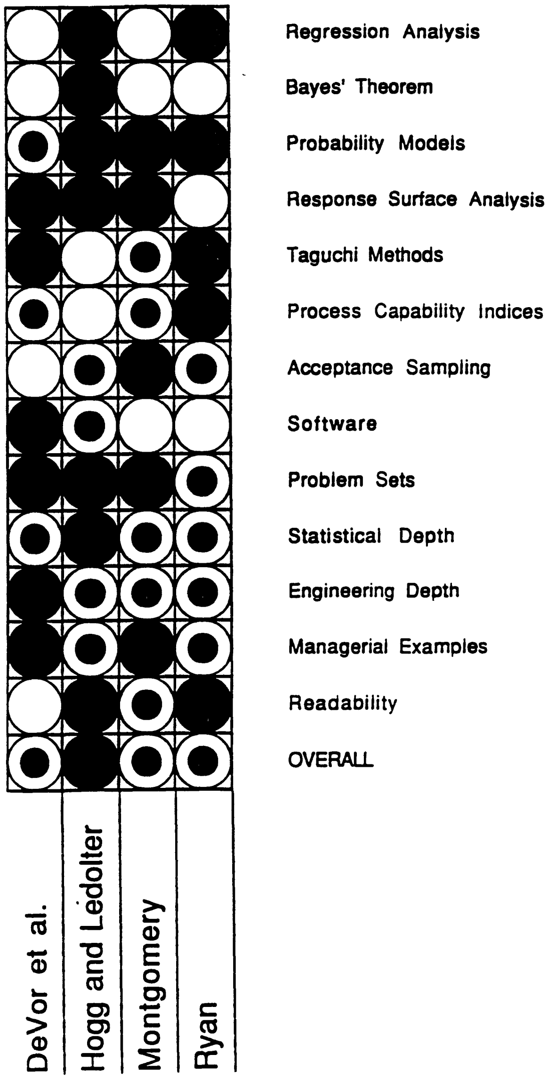


FIG. 4. *A comparative summary of the four introductory books under review, along axes that may be relevant to potential users.*

industries are multinational, and chief executives are hiring coaches to teach them the rudiments of courtesy abroad.)

The greatest gulf of all is one of respect. TQM strikes many of us as something between a foolish fad and much ado about the trivial. Worse, some academic statisticians regard most industrial statisticians as the ones who were not good enough to succeed at a university. Very often, in the biannual Black Friday faculty meetings at which all the graduate students are reviewed, comments surface to the effect that a barely adequate thesis might prepare someone for a career in industry, whereas the advisors of strong students should be sure to pace their students' completions to the academic calender so that candidates can interview at universities in January and February.

The same snobbish point gets made in other ways. The number of industrial statisticians who give invited talks at American Statistical Association meetings is dramatically underproportionate to their membership. Few internationally famous statisticians seek sabbaticals at industry, and many statisticians feel that applied research is generally not as prestigious as work in pure theory. And it does not soothe feelings that industrial statisticians typically command better wages than the professors who labored to teach them their trade.

Inescapably, this academic arrogance has sparked a reaction. Applied statisticians are often openly skeptical of academics' practicality, and have been assured that some practitioners view us as pretentious parasites, spawning uselessly complex theory while they deliver the goods. Industry drives them up a different ladder of advancement and forces their attention in other directions, and they discover that the statistics they learned has only a small value in rarified management echelons. Statistical skill carries one a few steps on that ladder, but then personality, experience, drive and political skills become significantly more important. (This is also true in universities, but to a much lesser degree.)

In large part, the industrial statisticians are correct. We have not shaped our mutual field according to their interests, and our greatest failure is that we have not provided (certainly at the Master's level!) the kind of education they need most. Our academic programs are chiefly designed to replicate university professors; even the relatively applied programs take their major themes from research institutions. It would be exciting to try to develop a program that aims at applied statisticians – this would entail endless heresies, such as purging all measure theory, admissibility, Hilbert spaces, sufficiency, differential geometry and so forth. In their place would doubtless appear the bootstrap, extensive computing, a deep study of exploratory data analysis and much more emphasis upon consulting experiences. I realize that such a complete restructuring of the curriculum is unlikely to ever occur, but I urge departments to consider the possibilities.

Also, we have set a research agenda that often ignores messy real problems or abstracts such problems to the point of sterility. I cannot believe that, aside from a few exceptions noted before, the contribution of academic researchers to control chart theory has served the interests of the industries that generated the questions. Similarly, the time series models that are most studied do not model the kinds of control loops that commonly operate to create autocorrelation in industrial processes. I would like to hope that some of this is being remedied by the emerging attention upon robust parameter design, software reliability and partial least squares, but it is noteworthy that in all of these cases, much of the impetus and effort arose outside conventional academic circles.

There are too many university researchers who have worn a rut in the theoretical landscape by repeated publications in a narrow domain. Even by the broader standards of intrinsic mathematical interest, this relentless elaboration of old ideas is hardly honest work. The best researchers know this well; David Aldous (1989) ends his book with the following postscript:

> In this book I have tried to explain the heuristics and direct the reader to what has been proved, in various special cases. I shall be well-satisfied if applied researchers are convinced to add the heuristic as one more little tool in their large toolkit. For theoreticians, I have already made some remarks on the relationship between heuristics and theory: let me end with one more. A mathematical idea develops best when it faces hard concrete problems that are not in the "domain of attraction" of existing proof techniques. An area develops worst along the "lines of least resistance" in which existing results are slightly generalized or abstracted. I hope this book will discourage theoreticians from the pursuit of minor variations of the known and the formalization of the heuristically obvious, and encourage instead the pursuit of the unknown and the unobvious.

Our ivory tower windows survey a broader demesne than is allowed to most corporate statisticians, but we need to open the windows and let in the wind.

These divisions in our profession cannot be repaired soon, if at all. It will help if universities design curriculums that take better account of the needs of the majority of our Ph.D. and Master's students. It is desirable for academic researchers to pursue research avenues that link to industrial problems, but there is enormous inertia which will slow movement in that direction. It is probably past time for university researchers to drop stale pseudo-applied activities (such as control charts and oddly balanced designs) that only win us a reputation for the recondite. And it would be very kind if industry statisticians were to understand that professors have different purposes and that statistical practice is an unsafe road for academic advancement.

## REFERENCES


Aldous, D. J. (1989). *Probability Approximations via the Poisson Clumping Heuristic.* Springer, New York.

Arora, S. and Safra, S. (1992). Probabilistic checking of proofs. In *33rd Annual IEEE Conferences on the Foundations of Computer Science.* To appear.

Banks, D. L. and Parmigiani, G. (1992). Pre-analysis of superlarge industrial datasets. *Journal of Quality Technology* **24** 115–129.

Bannerjee, P. K. and Rahim, M. A. (1988). Economic design of $\overline{X}$-control charts under Weibull shock models. *Technometrics* **30** 407–414.

Barron, A. R. and Barron, R. L. (1988). Statistical learning networks: A unifying view. In *Computing Science and Statistics: Proceedings of the 20th Symposium on the Interface* (E. J. Wegman, D. T. Gantz and J. J. Miller, eds.) 192–203. Amer. Statist. Assoc., Alexandria, VA.

Boen, J. and Zahn, D. (1982). *The Human Side of Consulting.* Lifetime Learning Press, Belmont, CA.

Box, G. E. (1985). Discussion of "Off-line Quality Control, Parameter Design and the Taguchi Method" by R. N. Kackar. *Journal of Quality Technology* **17** 189–190.

Breiman, L. (1985). Nail finders, edifices and Oz. In *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer* (L. M. LeCam and R. A. Olshen, eds.) **1** 201–204. Wadsworth, Belmont, CA.

Brieman, L. and Friedman, J. (1985). Estimating optimal transformations for multiple regression and correlation (with discussion). *J. Amer. Statist. Assoc.* **80** 580–619.

Brooks, F. P., Jr. (1982). *The Mythical Man-Month.* Addison-Wesley, New York.

Champ, C. W. and Woodall, W. H. (1987). Exact results for Shewhart control charts with supplementary runs rules. *Technometrics* **29** 393–400.

Chapman, G. R. and Kim, P. T. (1992). Spherical regression in quality assurance. Institute for Improvement in Quality and Productivity Research Report RR-92-01, Univ. Waterloo.

Chen, G. and Chen, J. (1992). Geometric quality assurance. Institute for Improvement in Quality and Productivity Research Report RR-92-06, Univ. Waterloo.

Cho, C.-K. (1987). *Quality Programming, Developing and Testing Software with Statistical Quality Control.* Wiley, New York.

Crosby, P. B. (1991). The pragmatic philosophy of Phil Crosby. In *The Quest for Competitiveness* (Y. K. Shetty and V. M. Buehler, eds.) 107–114. Quorum Books, New York.

Crosier, R.B. (1988). Multivariate generalizations of cumulative sum quality-control schemes. *Technometrics* **30** 291–304.

Dalal, S. and Mallows, C. (1988). When should one stop testing software? *J. Amer. Statist. Assoc.* **83** 872–879.

Davis, R. B. and Woodall, W. H. (1988). Performance of the control chart trend rule under linear shift. *Journal of Quality Technology* **20** 260–262.

DeVor, R. Chang, T.-h. and Sutherland, J. W. (1992). *Statistical Quality Design and Control.* Macmillan, New York.

Dijkstra, E. (1982). Formal techniques and sizable programs. In *Selected Writings on Computing: A Personal Perspective* 205–210. Springer, New York.

Duan, N. and Li, K. C. (1991). Slicing regression: A link-free regression method. *Ann. Statist.* **19** 505–530.

Dudding, B. P. and Jennett, W. J. (1937). Statistical techniques as a tool in industry. *General Electric Corporation Journal* **8** 1–23.

Easterling, R., Spencer, F., Mazumdar, M. and Diegert, K. (1991). System-based component-test plans and operating characteristics: Binomial data. *Technometrics* **33** 287–298.

Frank, I. E. and Friedman, J. H. (1993). A statistical view of some chemometrics regression tools (with discussion). *Technometrics* **35** 109–148.

Freeny, A. E. and Nair, V. N. (1992). Robust parameter design with uncontrolled noise parameters. *Statist. Sinica* **2** 313–334.

Friedman, J. H. (1991). Multivariate additive regression splines (with discussion). *Ann. Statist.* **19** 1–141.

Friedman, J. H. and Stuetzle, W. (1981). Projection pursuit regression. *J. Amer. Statist. Assoc.* **76** 817–823.

Futami, R. (1986). The outline of seven management tools for QC. *Rep. Statist. Appl. Res. Un. Japan. Sci. Engrs.* **33** 7–26.

Ghosh, S. (1990). *Statistical Design and Analysis of Industrial Experiments.* Dekker, New York.

Good, I. J. (1977). Dynamic probability, computer chess, and the measurement of knowledge. In *Machine Intelligence 8* (E. W. Elcock and D. Michie, ed.) 139–150. Wiley, New York.

GOUDIE, I. B. J. (1990). A likelihood-based stopping rule for recapture debugging. *Biometrika* **77** 203–206.

HAMADA, M. and WU, C. F. J. (1990). A critical look at accumulation analysis and related methods (with discussion). *Technometrics* **32** 119–162.

HASTIE, T. J. and TIBSHIRANI, R. J. (1990). *Generalized Additive Models.* Chapman & Hall, New York.

HAWKINS, D. M. (1992). A fast accurate approximation for average run length of CUSUM control charts. *Journal of Quality Technology* **24** 37–43.

HEDAYAT, A. S., SLOANE, N. and STUFKEN, J. (1994). *Orthogonal Arrays: Theory and Applications.* To appear.

HELLAND, I. S. (1988). On the structure of partial least squares regression. *Comm. Statist. B* **17** 581–607.

HOGG, R. V. (1985). Statistical education for engineers: An initial task force report. *Amer. Statist.* **39** 168–175.

HOGG, R. and LEDOLTER, J. (1992). *Engineering Statistics*, 2nd ed. Macmillan, New York.

HULTING, F. L. (1992). Methods for the analysis of coordinate measurement data. In *Computing Science and Statistics: Proceeds of the 24th Symposium on the Interface* (H. J. Newton, ed.) 160–169. Interface Foundation, Fairfax Station, VA.

HUMPHREY, W. S. (1989). *Managing the Software Process.* Addison-Wesley, Reading, MA.

HUMPHREY, W. S. and SINGPURWALLA, N. (1990). Predicting (individual) software productivity. Statistical Center for Quality Improvement Technical Report 7, Carnegie Mellon Univ.

ISHIKAWA, K. (1985). *What is Total Quality Control?* Prentice-Hall, Englewood Cliffs, NJ. [Translated by D. J. Lu.]

ITO, P. K. (1979). Statistics in Japanese universities. *Environmental Health Perspectives* **32** 21–23.

JEWELL, W. S. (1985). Bayesian estimation of undetected errors. *Bayesian Statistics 2* (J. Bernardo, M. DeGroot, D. Lindley and A. Smith, eds.) 663–671. North-Holland, Amsterdam.

KACKAR, R. N. (1985). Off-line quality control, parameter design and the Taguchi method. *Journal of Quality Technology* **17** 176–209.

KAMINSKY, F. C., BENNEYAN, J. C., DAVIS, R. D. and BURKE, R. J. (1992). Statistical control charts based on a geometric distribution. *Journal of Quality Technology* **24** 63–64.

KITAGAWA, T. (1983). The forty years of statistical sciences in Japan—My retrospect and prospect. *Bull. Inform. Cybernet.* **20** 1–21.

KOSAI, Y. and OGINO, Y. (1984). *The Contemporary Japanese Economy.* M. E. Sharpe, Armonk, NY.

LEÓN, R. V. and WU, C. F. J. (1992). A theory of performance measures in parameter design. *Statist. Sinica* **2** 335–358.

LI, K. C. (1989). Sight-seeing with SIR: A transformation based projection pursuit method. Technical Report, Dept. Mathematics, UCLA.

LI, K. C. (1991). Sliced inverse regression for dimension reduction (with discussion). *J. Amer. Statist. Assoc.* **86** 316–327.

LINCOLN, E. J. (1988). *Japan: Facing Economic Maturity.* The Brookings Institution, Washington, DC.

MANN, N. R. (1988). W. Edwards Deming: Why it happened in Japan and not in the U.S. *Chance* **1** 8–15.

MARTENS, H. and NAES, T. (1989). *Multivariate Calibration.* Wiley, New York.

MONTGOMERY, D. C. (1991). *Introduction to Statistical Quality Control*, 2nd ed. Wiley, New York.

MUSA, J. D., IANNINO, A. and OKUMOTO, K. (1990). *Software Reliability.* McGraw-Hill, New York.

OKITA, S. (1980). *The Developing Economies and Japan.* Univ. Tokyo Press.

OWEN, A. B. (1992). Orthogonal arrays for computer experiments, integration and visualization. *Statist. Sinica* **2** 439–452.

PAGE, E. S. (1954). Continuous inspection schemes. *Biometrika* **41** 100–115.

PIGNATIELLO, JR., J. J. and RUNGER, G. C. (1990). Comparison of multivariate CUSUM charts. *Journal of Quality Technology* **22** 173–186.

REYNOLDS, M. R., JR., AMIN, R. W., ARNOLD, J. C. and NACHLAS, J. A. (1988). $\overline{X}$ charts with variable sampling intervals. *Technometrics* **30** 181–192.

RIPLEY, B. D. (1993). Statistical aspects of neural networks. *Proceedings of the Séminaire Européen de Statistique.* Chapman & Hall, London. To appear.

ROBERTS, S. W. (1959). Control chart tests based on geometric moving averages. *Technometrics* **1** 239–251.

RYAN, T. P. (1989). *Statistical Methods for Quality Improvement.* Wiley, New York.

SASO, M. and KIRBY, S. (1990). *Japanese Industrial Competition to 1990.* Abt Books, Cambridge, MA.

SCOTT, D. W. and WAND, M. P. (1991). Feasibility of multivariate density estimates. *Biometrika* **78** 197–205.

SHEN, V., YU, T., THEBAUT, S. and PAULSEN, L. (1985). Identifying error-prone software—An empirical study. *IEEE Trans. Software Engrg.* **SE-11** 317–324.

SHEWHART, W. A. (1926). Quality control charts. *Bell System Technical Journal* **5** 593–602.

STEIN, M. (1987). Large sample properties of simulations using Latin hypercube sampling. *Technometrics* **29** 143–151.

STONE, M. and BROOKS, R. J. (1990). 'Continuum regression': Cross-validated sequentially constructed prediction embracing ordinary least squares, partial least squares and principal components regression (with discussion). *J. Roy. Statist. Soc. Ser. B* **52** 237–270.

SULLIVAN, L. P. (1984). Reducing variability: A new approach to quality. *Quality Progress* **17** 15–21.

TAGUCHI, G. (1986). *Introduction to Quality Engineering: Designing Quality into Products and Processes.* UnipubKraus, New York.

TUFTE, E. R. (1983). *The Visual Display of Quantitative Information.* Graphics Press, Cheshire, CT.

UNO, K. (1987). *Japanese Industrial Performance.* North-Holland, Amsterdam.

VINING, G. G. and MYERS, R. H. (1990). Combining Taguchi and response surface philosophies: A dual response approach. *Journal of Quality Technology* **22** 38–45.

WANG, J. C. and WU, C. F. J. (1992). Nearly orthogonal arrays with mixed levels and small runs. *Technometrics* **34** 409–422.

WOLD, H. (1966). Estimation of principal components and related models by iterative least squares. In *Multivariate Analysis* (P. R. Kirshnaiah, ed.) 391–420. Academic, New York.

WOLD, H. (1975). Soft modelling by latent variables: The nonlinear iterative partial least squares approach. In *Perspectives in Probability and Statistics: Papers in Honor of M. S. Bartlett* (J. Gani, ed.) 117–142. Academic, London.