Statistica Neerlandica (2006) Vol. 60, nr. 3, pp. 270–282

Industrial statistics: a discipline with opportunities and challenges

Jeroen de Mast* and Ronald J. M. M. Does†

Institute for Business and Industrial Statistics of the University of Amsterdam (IBIS UvA), Plantage Muidergracht 24, 1018 TV Amsterdam, The Netherlands

The development of industrial statistics as a scientific discipline, as well as its contributions to improving quality and efficiency in modern industry are discussed. Furthermore, the nature, purpose, and research paradigm of scientific research in industrial statistics are profiled. These accounts form the basis for a discussion of the discipline's future and key elements therein.

Keywords and Phrases: history, rational reconstruction, grounding research.

1 Introduction

This paper is an opinion paper discussing the current state and future of industrial statistics as a scientific discipline. The emphasis is not on the discipline itself, but rather on the context in which it operates. The context of application of industrial statistics is described against the background of the emergence of management practice and science, and the rise of modern industry in the 20th century. Besides the historical development, we discuss current practices of industrial statistics applications.

Industrial statistics as a scientific discipline is characterized by outlining the nature and methodological framework of its research. The paradigms of reconstruction research and grounding research are discussed, and in addition, the nature of the relationship to mathematics is specified. The characterizations of industrial statistics as an applied and a scientific discipline are the canvas for a discussion of the future of the field. Key elements of this future are highlighted in the form of opportunities and challenges.

2 The emergence of industrial statistics and its contributions to quality and efficiency improvement

The 20th century witnessed incredible increases in the quality of products, while in the same period prices dropped dramatically. These important improvements in

*jdemast@science.uva.nl
†rjmmdoes@science.uva.nl

quality and efficiency in industry were the result of innovations in management, engineering, and economics. The industry is also indebted to industrial statistics for these advances (and vice versa, the statistical sciences owe a great deal to the opportunities that industry created for their development). We describe below the emergence of industrial statistics against the background of the rise of modern industry. More elaborate descriptions and references can be found in WREN (2005) and DE MAST (2006).

The invention of the mass fabrication paradigm in the US by the end of the 19th century stimulated the development of the discipline of management. Scholars like Henri Fayol and Frederick Taylor elevated management from an art to a more professional and scientific discipline. The economic focal points of mass fabrication are volume and productivity. Companies improved efficiency by mechanization, standardization, productivity improvement (Taylor's *scientific management*), and especially the exploitation of economies of scale. Parallel to and interwoven with the development of mass fabrication, the disciplines of industrial statistics and quality engineering emerged. One of the key events was stimulated by the Guinness Breweries, the then largest brewery in the world. Guinness hired some of the brightest young men they could find to make a rigorous and scientific study of the brewing process and the associated processes of growing barley and hop. For want of an adequate methodology for making reliable inferences from the experiments they ran, William Gosset (Student) developed the basics of the t-test, the first modern data analysis procedure. His 1908 paper *The Probable Error of the Mean* inspired statisticians like Sir Ronald Fisher, Egon Pearson and Jerzy Neyman to develop much of the now standard machinery for the design and analysis of the results of comparative experiments (hypothesis testing, the analysis of variance, regression analysis, design of experiments).

Other early contributions of industrial statistics were in the field of process control. In the 1920s, Walter Shewhart – who worked for the Bell Laboratories (now Lucent Technologies), the research arm of AT&T – introduced the notions of assignable cause of variation and chance cause. 'Assignable cause' was Shewhart's term for patterns in process measurements that are to be interpreted as a signal that something in the process has changed. All other patterns in data are to be interpreted as random noise, which operators should ignore. Shewhart developed an easy-to-use tool that helps operators discern between the two: the well-known control chart. Where Gosset's t-test helped researchers draw reliable conclusions from experiments in the presence of uncontrollable influences and noise, Shewhart's control chart helped operators interpret patterns in process data reliably. The foundations that Shewhart laid for statistical process control – or SPC as it is perhaps even better known – still represent a standard practice in quality control. Shewhart and his followers (such as W. Edwards Deming) took up the scientific study of quality control quite seriously and their importance goes beyond the control chart. An important element of Shewhart's philosophy on process control is that measurements and the use of simple statistical tools for data analysis enable a team of operators to learn from their process.

Despite Shewhart's emphasis on process control, quality inspection continued to be the primary way to ensure quality. Because 100% inspection became very costly, time consuming, and was often simply impossible (tests would destroy products), quality inspection based on samples became a necessity. Accepting or rejecting batches of products based on inspection of a sample creates risks: the sample could give too optimistic or too pessimistic an impression of the batch. Probability theory was called upon to find a balance between sample size, the consumer's risk, and the producer's risk. In the 1930s, Harold Dodge, who worked – like Shewhart – for the Bell Laboratories, developed standard schemes for sampling inspection and published tables that related sample sizes to risks. Furthermore, Dodge was one of the initiators of the movement that came to be known as Statistical Quality Control, and that resulted, among others, in the present American Society for Quality (ASQ), a society with more than 100,000 members.

After the Second World War, Western companies continued to work in the mass fabrication paradigm, focussing on productivity and volume, and not on quality, and with millions of consumers eager to buy they got away with it. Economical conditions, as well as knowledge of management and production, had changed, however. As Europe's centuries-old manufacturing paradigm based on craftmanship had been made obsolete by the American innovation of mass fabrication, so the Americans were overtaken by the Japanese in the decades after the Second World War. Forced by the situation in which they found themselves, Toyota and other Japanese companies created a production system whose economic focal points were low inventory levels, speed, and flexibility. In order for the system to work, Toyota needed manufacturing processes that run like clockwork: optimized changeovers to enable low inventory levels, aggressive defect reduction to eliminate inefficiencies and enable short cycle times, and partnerships with suppliers. Having made all processes more reliable, Toyota did not need excessive buffers of inventory: suppliers delivered the exact number of components needed just in time. The prevailing Western organizational structures – which were basically derived from Ford's command-and-control structure – did not suit Toyota's needs. Running one's operation like clockwork implied delegating authority to the operators to intervene when problems arise. Furthermore, problem-solving was put in the hands of shop-floor workers in the form of quality circles. Instead of seeing management as the source of process improvement, Toyota mobilized shop-floor workers to participate in continuous improvement.

The Japanese government hired Western experts like Deming and Joseph Juran to teach them the theory of statistics and quality management, which had been largely ignored in the West. Deming promoted sound problem-solving techniques and statistical methods, such as his plan-do-check-act (PDCA) cycle of continuous improvement; see his book *Out of the Crisis* (Deming, 1986). Juran introduced his principles of quality management to help integrate quality activities in all layers of an organization (Juran, 1989). The Japanese implemented Juran's tenet that top management should play an active role in quality activities. Furthermore, Juran introduced the quality trilogy of quality planning, quality control, and quality improvement as three complementary (but integrated) aspects of quality management. Among the important new innovations were Juran's project-by-project approach to quality improvement, and his ideas on operator controllability. A flood of new practices arose in Japanese companies, which became known in the West only decades later: kaizen, lean manufacturing, JIT, quality circles, and many more.

In the 1970s, the Japanese assaulted the world markets with their clockwork manufacturing machines, and appeared to have significant competitive advantages: they sold similar products as Western companies, but at lower cost, with less defects, and superior reliability and durability. The Japanese had learned that quality and manufacturing virtuosity are strategic weapons to be reckoned with. The Japanese quality revolution brought the Western industries to a state of recession. In 1980, NBC broadcasted a documentary with the title "If Japan can … Why can't we?" The first reactions to the Japanese assault were confused and often besides the point. For years in a row, the success of the Japanese competition was attributed to a superior cost structure due to cheap labour, low quality and imitation. When the West finally realized that it was facing a completely different, and clearly superior manufacturing paradigm, the first reactions were unfocused, rash, and confused. Within a few years, a flood of quality gurus (Crosby, Feigenbaum, Ishikawa, Taguchi, Shainin, Shingo) came over to the Western business world, and each month there was a new magical trick: quality circles, JIT, kanban, pull-systems, kaizen, … Some of the 'flavours of the month' appeared to be just fads; many more had their valuable points, but were less generic than claimed and failed to endure. Most of the valuable ideas have been integrated into more generic theories, which is probably their right place. Six Sigma is perhaps the most comprehensive one of these generic theories.

These developments set the stage for developments in industrial statistics. After World War II, George Box made important modifications to the theory of the design of experiments, which made it better applicable in industry. The theory of experimental design was developed in the 1930s mostly for use in agricultural research, where the focus is on large, long-lasting one-shot experiments. Industry, on the other hand, has a need for sequences of smaller scale experiments, where each test elaborates on the results of the previous tests. In their seminal paper of 1951, Box and Wilson introduced response surface methodology, which rapidly became popular in the chemical and other industries for process optimization and product development (Box and Wilson, 1951). The proliferation of computers made statistical modeling approaches – such as regression analysis – easy to apply, and their use in industry expanded.

Post-war Japan saw statistics as the secret weapon that had helped the Allies win the war, and they were eager to learn from experts like Deming. In line with their approach of continuous improvement led by teams of shop-floor workers, the Japanese invested heavily in training their work-force in simple statistical and problem-solving tools. Kaoru Ishikawa developed seven basic tools of quality: the histogram, Pareto chart, cause-and-effect diagram, run chart, scatter diagram, flow chart, and control chart (Ishikawa, 1982). Shewhart's and Deming's ideas on statistical process control were commonly applied in Japan, including control charts and other statistical techniques. Genichi Taguchi introduced in engineering and product design the notion of robustness, that is, the extent to which the quality of products is sensitive to sources of variation during manufacturing and usage. Although the methodology for experimental design that Taguchi proposed to study robustness is generally considered less effective than the alternatives that Western industrial statistics had developed, he deserves at least the credit for getting systematic experimentation and the study of sources of variation widely accepted in industry, in Japan as well as in the West.

With the renewed interest in the West in quality, industrial statistics flourished and the arsenal of techniques for quality engineering expanded rapidly. Standard methods to study the precision of measurement systems were developed, as well as methods for analysing non-normal data, statistical models and techniques for reliability engineering, multivariate methods, capability analysis, variance component estimation, and graphical methods. The older theories about design and analysis of experiments were still being extended to adapt them to new applications. Research into methodologies for process monitoring, control, and adjustment thrived. The development of powerful computers and user-friendly statistical software brought statistical methods within reach of a large public.

3 The present: Six Sigma

The last decades of the 20th century witnessed economic developments in many industries which have come to be labeled 'hypercompetition'. Industries such as consumer electronics, the automotive industry, and the foods industry increasingly competed with each other on higher and higher levels of efficiency and quality. The main winner of this race has been the consumer; for the participating companies, the race resulted in profit margins under pressure. Falling behind in the race for quality and efficiency means being out of business. Against this background, the Six Sigma programme emerged (De Mast, Does and De Koning, 2006). Rather than being a completely new approach, it is simply the next step in the development of more and more professional and scientific approaches to the organization and management of processes. It incorporates many of the innovations in quality and efficiency improvement of the 20th century (Bisgaard and De Mast, 2006).

The basic principle of Six Sigma is that organizations should invest substantial efforts in the systematic improvement of their routine functions, not only manufacturing and service delivery, but also sales, accounting, marketing, and product and process development. Six Sigma offers a managerial framework to organize such an endeavor, *de facto* turning the systematic improvement of routine operations in companies itself into a routine operation. Besides a managerial framework, it offers a detailed and comprehensive methodology for improvement projects, consisting of a framework for inquiry and many statistical and non-statistical techniques (De Koning and De Mast, 2006). Better than any quality improvement initiative before, it incorporates a sound business economic attitude and effectively bridges the gap between engineering and management frames of reference. All these at a scale that is enormous: a substantial number of multinational enterprises have started the implementation of the programme, and in more recent years, the programme is obtaining the status of standard approach in health care and the service industries as well. Six Sigma, more than any initiative before, has brought statistical methods of considerable level to a general public. Thousands of black belts and green belts (as Six Sigma project leaders are called) worldwide are trained in the use of design of experiments, non-normal capability analysis, and analysis of variance, to mention just a few (Hoerl, 2001, gives an overview).

4 Industrial statistics as a scientific discipline

Industrial statistics could be defined as 'the discipline which develops quantitative methods and paradigms for inquiry and routine decision making in industry.' Empirical inquiry is the context in which statistical techniques are applied in improvement projects (such as Six Sigma projects); routine decision-making is the context for industrial statistics in such applications as sampling inspection and control charting. In essence, industrial statistics is methodological research. Methodology is not an empirical science, and researchers in industrial statistics cannot turn to research paradigms and methods that empirical scientists use. In fact, the statistical literature is not quite elaborate or explicit when it comes to defining its research paradigm, but we shall make an attempt to profile the objectives and research methods of the discipline.

In the literature, methodological research is sometimes characterized as rational reconstruction, and sometimes as grounding research. A rational reconstruction presents problematically formulated knowledge – in a form that is not explicit, precise or consistent enough – in a similar, but more precise and more consistent formulation. The given problematic complex is typically intuitive, tacit knowledge. The simplest form of rational reconstruction is explication: the formulation of exact definitions for loosely defined concepts. Linguistic research is often reconstruction research (where one attempts to make explicit the grammatical rules that native speakers of a language know intuitively), as well as research in law (trying to reconstruct intuitive notions of right and wrong) and aspects of mathematics (e.g., the axiomatic and measure-theoretic set-up of probability as an attempt to formalize intuitive notions of probability). To illustrate the reconstruction nature of research in industrial statistics: acceptance sampling existed before industrial statistics, and users will have had an intuitive understanding of consumer's and producer's risks. But concepts such as acceptable quality level (AQL) and limiting quality level (LQL), and tools such as the operating characteristic (OC-) curve have brought a precise and consistent framework for reasoning about the suitability of sampling schemes. Likewise, control engineers had developed PID (Proportional Integral Derivative) controllers based on trial and error and gut feeling, but time series analysis and ARIMA (Autoregressive Integrated Moving Average) models offered a precise and consistent framework to understand and design process controls. Shewhart's assignable and chance causes, Fisher's principles of design of experiments, Neyman and Pearson's procedure for hypothesis testing: all these innovations offer a precise and consistent framework to supersede implicit, vague, and imprecise understanding, and they are rational reconstructions.

Rational reconstructions could have a purely descriptive impetus. The emphasis is on reconstruction as 'again' construction, i.e. making the object 'more equal to itself' by extracting essential elements and reformulating and restructuring them. The main criteria for adequacy in this case are clarity, exactness, and similarity to the original. One step further is a rational reconstruction with a prescriptive impetus. The emphasis is on 'new' construction. The original material is taken as a starting point, but based on critical examination (on the basis of external formal criteria such as logic), it is corrected. Besides clarity and exactness, we have in this case the criterion of consistency, which replaces the criterion of similarity.

Where reconstruction research focuses on explication and consistency, grounding research is an investigation into the rationality of actions. In the paradigm of Habermas (1981, pp. 25ff.), actions are called rational if they imply a validity claim which can be justified. Grounding research seeks to make explicit the validity claims that an action makes in order to investigate whether they can be justified. Seeing methods as the products of research in industrial statistics, we note that the validity claim that these usually make is 'usefulness'. This claim is composed of two claims: that the method's intended purpose is a legitimate objective, and second, that the method is effective in attaining the intended purpose. Grounding of a method thus amounts to providing an argument that shows that the method is effective (either by providing empirical evidence or theoretical argumentation that demonstrates the method's effectiveness) and its purpose legitimate. De Koning and De Mast (2005) discuss how a methodology such as Six Sigma's DMAIC method can be grounded. They provide additional references about rational reconstruction as well as grounding research.

Mathematics and, in particular, mathematical statistics play dominant roles in industrial statistics. The paradigms and methods that industrial statistics has developed rest significantly on mathematical machinery. For instance, the developments of the concepts of AQL, LQL and OC-curve is rational reconstruction, but their elaboration is mathematical research. It would be a mistake, though, to regard industrial statistics as a branch of mathematics. The relationship is more complex. First, the mindset and reasoning in which statistics is applied are quite far from mathematical reasoning. With the objective of arriving at inferences about a real system, statistical reasoning typically proceeds by these steps:

(1) *Experimental model:* Advance a hypothetical system – a 'model' – that mimics the real system under study (that is, on essential aspects, its behavior is similar to the behavior of the real system). Questions of interest for the real system are translated to questions in terms of parameters of the model. Whereas the real system under study (defined in terms of a population, experimental units, variates, and population parameters) is governed by the laws of nature, the experimental model (defined in terms of a probability triple $(\Omega, \mathcal{F}, \mathbf{P})$ and including a canonical model for sampling error and a test statistic) is governed by the axioms and laws of probability.

(2) *Deduction:* A sample is taken from the population of interest, and based on the findings a realization of the test statistic is computed. Using the standard machinery of reasoning that comes with the experimental model (such as hypothesis testing and confidence interval estimation) the inquirer deduces conclusions that hold for the hypothetical system.

(3) *Inference:* Finally, assuming that the model's behaviour is indicative for the real system's behaviour, the conclusions for the model are carried over to the real system (MAYO, 1996, discusses from the perspective of modern philosophy of science how statistical inference works, elaborating, among others, Fisher's example of a lady tasting tea). Step 2 is mathematical deductive reasoning. But this step is part of a larger whole, namely an *inductive* argument (steps 1, 2, and 3 together), which involves a number of extra-mathematical inductive steps: the definition of the population and experimental unit, assumptions made to go from raw observations to a realization of the test statistic (such as assumptions concerning the validity of the measurement procedure and data-cleaning steps), *ceteris paribus* assumptions, and the translation from a statistical conclusion (e.g. about correlation) to an inference about the real system (e.g. about causation). The mindset and rationale of these more inductive aspects of empirical inquiry are quite distinct from and in many aspects even contrary to mathematical deductive thinking.

What holds for the application of statistics can be said as well for research in industrial statistics: mathematical reasoning is part of but not the whole of the story. Whether one takes a reconstruction view or a grounding view on the development of statistical methods, the starting point for the development of statistical methods are not mathematical axioms, but assumptions about what would constitute a useful and effective method. Where mathematical problems are valued for the logical challenge they pose and the elegance of their solution, these qualities are almost irrelevant from an industrial statistics point of view. For industrial statistics what defines 'good' is how useful it is. For example, Shewhart's control chart is mathematically trivial, but the underlying concepts and the graphical technique are effective and useful, and therefore justify their prominent place in industrial statistics.

Rather than saying that industrial statistics is a branch of mathematics, one should say that industrial statistics *uses* mathematics, just as for instance physics does (and even probability: 'Probability is no more a branch of mathematics than is physics, although it owes a great debt to mathematics for its formulation and development'; FINE, 1988).

5 Industrial statistics: opportunities and challenges

5.1 Observation 1

Efficiency has been the focus of Western industry up until the 1980s. Many initiatives in the last decades of the 20th century sought to optimize quality. Economically speaking, the driver for growth and profitability in the West in the 21st century will be innovation (this is claimed, for example, in the 2004 report of the US Council on Competitiveness entitled *Innovate America: Thriving in a World of Challenge and Change*). The attribution of economic growth to innovation is due to such economists as Schumpeter and Solow. Rosenberg (1983) describes the significant cumulative impact of small-scale innovations such as small product modifications and process adjustments. Innovations may pertain to quality or efficiency, but they need not be. They can improve processes and products, as well as sales strategies, accounting policies, or business models. In an economy that is determined more and more by dynamics than by static advantages, it is company-wide innovative capabilities which drive a company's competitiveness.

5.2 Observation 2

Continuous improvement and incremental innovation should be a decentralized activity, which means that line rather than staff personnel execute this function. The idea is based on the work of management scholars and economists such as Hayek, Mintzberg (1994), and Jensen (1998), who concluded that these activities cannot take place in the absence of detailed, intimate knowledge of the system under study (what economists call *specific knowledge*).

5.3 Observation 3

A critical part of innovation is inquiry (building understanding of the system one works with). Statistics provides vital tools for inquiry (De Mast, 2003).

The first two observations imply that continual improvement and innovation of one's own work environment will become more and more everyone's task. Combined with point 3, we conclude that mastery of statistical tools and thinking will be important for a large number of professionals, whether line managers, engineers, marketers or, for example, nurses in a hospital (Van den Heuvel, Does and Vermaat, 2004). In addition, statistical competencies as part of a company's organization-wide innovative capabilities gain strategic importance. Industrial statistics may find itself at the heart of economic developments in the West (Bisgaard and De Mast, 2006).

To seize this opportunity, however, industrial statistics should face some challenges. Decentralized organization-wide innovation efforts need a managerial framework, because unbridled decentralization can easily lead to pointlessness and pet projects (as experiences with *Total Quality Management* and other programmes have demonstrated). Six Sigma's organizational structure for programme and project management is a big step forward and has proved to be effective. They are in line with commendations from the scientific literature on the subject (cf. Jensen, 1998; OGC, 1999). Next, the benefits of using statistical methods should be framed in a terminology that managers can relate to (in other words: industrial statistics needs a good *value proposition*). Our main concern, however, relates to the more and more exclusive dominance of mathematics in industrial statistics, at the expense of extra-mathematical aspects, and that this stakes the usefulness of some of the methods that it develops. We wish to illustrate this point from an example.

6 An example

Consider a project that seeks to reduce the lead time of a certain administrative process (say, processing of address changes). Following an approach like Six Sigma, project leaders are taught to tackle such an issue by developing an explanatory model, which describes the causal influence factors that determine lead time. Before the possible effects of conjectured influence factors can be experimentally verified and quantified, the project leader has to identify potential influence factors for further study – what methodologists call *hypothesis generation* (see De Mast and Bergman, 2006). One of the approaches to generate hypotheses is to collect data from the process in operation and to look for salient patterns.

Suppose the process in question is executed at four sites, and that the project leader recorded the lead times of a number of cases at each site – the plotted data might look like Figure 1. Given these data, what sort of analysis would you do? In our experience, statisticians tend to attack such a dataset with the analysis of

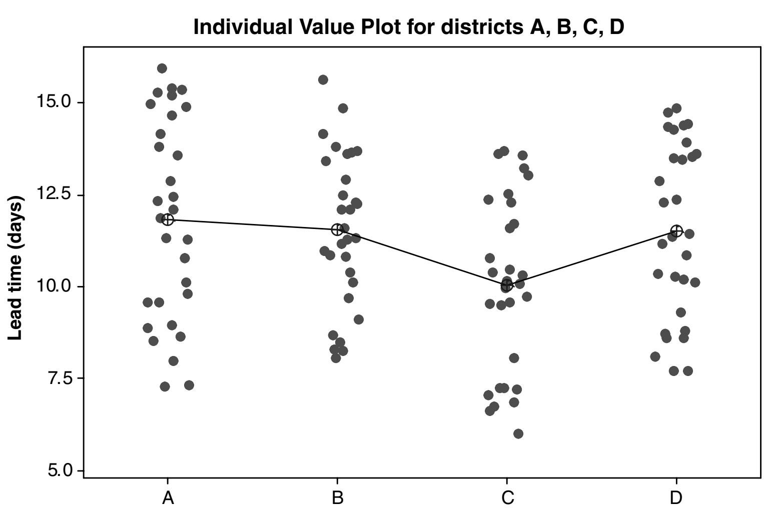

Fig. 1. Hypothetical dataset.

variance (ANOVA) procedure. After all, we have data in four groups, and ANOVA is the standard procedure to study group differences. For most statisticians, ANOVA is a procedure to compute a *p*-value, thus testing a certain hypothesis (and this is the way ANOVA is usually explained in textbooks). For a project leader who tries to identify the dominant influence factors for lead time, however, the research question is a different one: are there *relevant* differences among the four sites regarding lead time? For such a project leader, the group means are the results of interest, not the *p*-value. In case the sample size is large it is hard to see that the *p*-value has any value at all: even tiny differences will be reported as significant, although for the project leader they are too small to be of any relevance. For small datasets, *p*-values do have value: they indicate to the project leader whether he is looking at a real and repeatable pattern or only at an artifact of noise. But even then, statistical significance is just one of the validity checks, and should not be promoted to being the primary research question. Moreover, the big majority of nullhypotheses tested in practice are incredible on *a priori* grounds: the four sites are different in a very real sense (different persons, different layout of the building, different interpretations of procedures) and although the resulting differences in lead time may be very small, it is incredible they are 0.0, as the nullhypothesis states. In such cases, where the null-hypothesis is incredible beforehand, the only information the *p*-value conveys is whether the dataset is large enough to prove it.

So far for ANOVA; let us turn back to the question how would you analyse these data? The point is that if we let the data guide our choice of analysis procedure, we would not use ANOVA. The salient feature in this dataset are not the differences among groups, but the differences *within* groups. The data show that lead times well below 10 days are realistic, but yet, lead times around 15 days occur frequently. For a project leader who aims to understand the factors that determine lead time, it is essential to understand these within-group differences, whereas factors that could explain the between-group differences are simply irrelevant because their effect is dominated by factors that operate within groups. The project leader is likely to be better off using procedures like pairwise comparison or autopsy (see DE MAST and BERGMAN, 2006). Autopsy means that the project leader isolates some of the cases that had a lead time around 15 days and closely studies them. Pairwise comparison is similar, but has the project leader compare these cases to ones that have a lead time below 10 days. Procedures as these are informal and are not based on mathematics, but they should be part of the industrial statistician's standard tool kit just like ANOVA is.

If usefulness, and not just mathematical correctness, is taken as the criterion to evaluate methods, it follows that research (as well as teaching) in industrial statistics must focus on extra-mathematical aspects just as well as on mathematics:

– Research should be based on correct assumptions about the application context. This includes assumptions about practical restrictions to data collection and experimentation, assumptions about the type of research questions that are relevant, about the type of mathematical models that are likely to work, and the sort of solutions that are expected. The consequence is that researchers and teachers of industrial statistics should have a basic insight into the engineering and business economic context in which their methods are applied.

– Methods should be made operational in a form that non-statisticians can apply them. Modern statistical software provides an excellent platform for the implementation of statistical methods. But industrial statisticians must make a clear distinction between what is important for them as the developers of these methods and what is important for the user. Hypothesis testing may be the most interesting aspect of ANOVA for statisticians, for the user estimation of the group means and a simple plot are usually more relevant.
– Methods should be integrated into a methodology. Users apply them as part of inquiry, and that means that they should be placed in this framework. Six Sigma, with its tools integrated in its DMAIC (Define, Measure, Analyse, Improve, Control) method, provides a good example, but see DE MAST (2003) for a more general description. And more in general: mathematical aspects of methods should be integrated with extra-mathematical aspects. For example, correlation is a mathematical concept, whereas the definition of causation requires extra-mathematical concepts (counterfactual logic). But causation is important for inquirers, and consequently, statisticians should go beyond vague remarks along the lines that correlation does not imply causation, and provide methods to the inquirer that help him make inferences about causality. The Problem, Plan, Data, Analysis, Conclusion (PPDAC) structure proposed by MACKAY and OLDFORD (2000) is a good example of a useful integration of statistical methods in a framework for inquiry.

7 Conclusion

The emergence of industrial statistics as a scientific and applied discipline has been interwoven with the evolution of industry, and industrial statistics has made important contributions to quality and efficiency. On the one hand, prospects are bright, with economic developments in the West that emphasize the need for general mastery of research skills, and Six Sigma as a widely accepted program which brings statistical methods of considerable level to a wide public. On the other hand is the danger that industrial statistics as a scientific discipline isolates itself from its application context. This concern reflects similar worries by BENNIS and O'TOOLE (2005) about the trend that business schools sacrifice relevance of their research to the favour of rigor and scientific prestige. In our view, industrial statistics should see itself more as a methodological than as a mathematical discipline. This stance does not imply that mathematics is any less important than the extra-mathematical aspects of the discipline, but rather that they are both important. The Six Sigma program provides a useful vehicle and statisticians should not disqualify it out of scepticism about the slick and hyped form in which it is often presented. Furthermore, disciplines like econometrics and psychometrics could serve as an example (had it not been such an ugly word, we would even have suggested the term *industrometrics*).

References

BENNIS, W. G. and J. O'TOOLE (2005), How business schools lost their way, *Harvard Business Review* **83**, 96–104.

BISGAARD, S. and J. DE MAST (2006), After Six Sigma – What's next?, *Quality Progress* **39**, 30–36.

BOX, G. E. P. and K. B. WILSON (1951), On the experimental attainment of optimum conditions, *Journal of the Royal Statistical Society Series* B **13**, 1–45.

DE KONING, H. and J. DE MAST (2005), Grounding of Six Sigma's Breakthrough Cookbook: How to research a methodology?, *International Journal of Six Sigma and Competitive Advantage* **1**, 263–275.

DE KONING, H. and J. DE MAST (2006), A rational reconstruction of Six Sigma's Breakthrough Cookbook. *International Journal of Quality and Reliability Management* **23**, to appear.

DE MAST, J. (2003), Quality improvement from the viewpoint of statistical method, *Quality and Reliability Engineering International* **19**, 255–264.

DE MAST, J. (2006), A history of industrial statistics and quality and efficiency improvement, in: S. COLEMAN, T. GREENFIELD, D. MONTGOMERY, D. STEWARDSON (eds.), *Statistical practice in business and industry*, Wiley, New York, in press.

DE MAST, J. and M. BERGMAN (2006), Hypothesis generation in improvement projects: approaches for exploratory studies, *Quality and Reliability Engineering International* **22**, to appear.

DE MAST, J., R. J. M. M. DOES and H. DE KONING (2006), *Lean Six Sigma for service and healthcare*, Beaumont Quality Publications, Alphen aan den Rijn.

DEMING, W. E. (1986), *Out of the crisis*, MIT, Cambridge, MA.

FINE, T. (1988), Foundations of probability, in: S. KOTZ and N. JOHNSON (eds.), *Encyclopedia of statistical sciences*, 8th edn, Vol. 3, pp. 175–184. Wiley, New York.

HABERMAS, J. (1981), *Theorie des kommunikativen Handelns, Band 1*, Suhrkamp, Frankfurt.

HOERL, R. (2001), Six Sigma black belts: what do they need to know, *Journal of Quality Technology* **33**, 391–406.

ISHIKAWA, K. (1982) *Guide to quality control*, Asian Productivity Organization, Tokyo.

JENSEN, M. C. (1998), *Foundations of organizational strategy*, Harvard University Press, Cambridge, MA.

JURAN, J. M. (1989), *Juran on leadership for quality*, The Free Press, New York.

MACKAY, R. J. and R. W. OLDFORD (2000), Scientific method, statistical method and the speed of light, *Statistical Science* **15**, 254–278.

MAYO, D. G. (1996), *Error and the growth of experimental knowledge*, The University of Chicago Press, Chicago, IL.

MINTZBERG, H. (1994), *The rise and fall of strategic planning*, The Free Press, New York.

OFFICE OF GOVERNMENT COMMERCE (OGC) (1999), *Managing successful programmes*, The Stationery Office, London.

ROSENBERG, N. (1983), *Inside the black box: technology and economics*, Cambridge University Press, Cambridge.

VAN DEN HEUVEL, J., R. J. M. M. DOES and M. B. VERMAAT (2004), Six Sigma in a Dutch hospital: does it work in the nursing department?, *Quality and Reliability Engineering International* **20**, 419–426.

WREN, D. A. (2005), *The History of management thought*, 5th ed., Wiley, New York.

Received: September 2005. Revised: October 2005.