PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A
MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

Standards and Statistics [and Discussion]

D. J. G. Farlie, M. Gerson, C. Chatfield, F. A. Wood and K. Sachs

Phil. Trans. R. Soc. Lond. A 1989 **327**, 573-579
doi: 10.1098/rsta.1989.0012

Phil. Trans. R. Soc. Lond. A **327**, 573–579 (1989) 573
Printed in Great Britain

Standards and statistics

By D. J. G. Farlie

Department of Mathematics, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, U.K.

Currently, about 2000 British standards have some statistical or quality control content. The paper outlines the development of the interaction between statistics and standardization from the beginnings in 1932 to the recent large-scale international cooperation in developing international standards for the application of statistics.

In May 1932, the British Standards Institution arranged a meeting between Dr W. A. Shewhart and representatives of manufacturing industries with a view to a closer cooperation between those developing methods from theoretical considerations and those interested in the practical application of these methods to the problems in standardization. The meeting gave rise to a committee charged with, among other things, the responsibility to produce a report on the application of statistical methods in standardization and the specification of quality. With commendable speed the committee produced the famous BS600: 1935. The text of the report was primarily the work of Dr E. S. Pearson, later to be Professor Egon Pearson, but produced after much discussion with the committee. The foreword to this standard indicates how little the style of such committees changes with time. Those who have served with me on successor committees will recognize the equal relevance of the following quotation to present day practice as well as to that of the original committee.

> The treatment of the subjects and the methods advocated have been debated at length by the committee, who are grateful to Dr Pearson for his patient consideration and willing cooperation in the shaping of the publication in its present form.

The foreword continued by looking forward to a time when the proposals in the publication would interest workers outside Great Britain, but disclaimed any intention of creating a comprehensive introduction to statistics. The emphasis was, and still remains, on the demonstration of the benefits of using statistical methods in standardization and in the specification of quality.

The exposition given in BS600: 1935 is firmly rooted in the practical world of standards; three examples serve to illustrate the contribution of statistical methods to the control of quality. The production of malleable iron castings is used to motivate a search for rational subgroups of manufacture, and also to encourage the recognition of the contribution of both average strength and variability of strength to the quality of the product. The breaking strengths of cotton strips are used to point out the care that is needed to give meaning to a specification such as 'the strength must be greater than 164 lb' (1 lb = 0.45 kg). Is this a statement about every sample or about the average of a predetermined number of test pieces? How will compliance with the standard be decided? The analysis of ash content in samples of coal is used to direct attention to the effects of time trends on the estimation of overall average quality. If an average ash content is specified, should this permit the inclusion of a long period of ash content above this average because of an earlier period of below average values?

Although the normal distribution occurs as a mathematical equation in an appendix, the major use of the distribution is as a simple scaling device to interpret the mean and standard deviation. Within one standard deviation of the mean there will be 68.27 % of the distribution, within two standard deviations, 95.45 %, and within three standard deviations, 99.73 %. These theoretical percentages are compared with the actual percentages for empirical distributions of ash content in coal, strength of malleable iron castings and depth of sapwood in telegraph poles, and they are used to comment on skewness and the smoothing effect of large sample sizes. Consignment sampling, control charts and the use of substitute quality indices for direct performance measurement are all covered. There is an uncomfortably modern sound to the concluding sentence of the introduction; it describes the application of statistical methods as 'a field which covers the development on scientific lines of methods which seek to eliminate waste of energy in production, and to supply goods soundly made and such as are best fitted to satisfy human wants'. It is uncomfortable because that vision of the benefits to manufacturing industries has fired fewer people than one would have liked during the intervening half century. An excellent review of the intervening years in a variety of countries, written from the perspective of quality management rather than quality statistics, is provided by S. J. Morrison (1987).

Browsing through the background papers of the QMS/1 committee of the British Standards Institution (BSI), it is clear that there have been repeated efforts to integrate statistical thinking into the committee framework of the Institution. As an example, in October 1955, BSI staff met with Dr Dudding (General Electric Company), Mr Tippett (British Cotton Industry Research Association) and Mr Wainwright (Bakelite Ltd) in the presence of Professors Barnard and Pearson to discuss the problems of sampling for industrial specifications. During the discussion, the merits of process control over the sampling of finished products were advocated, particularly in the context of a certification or Mark scheme. Nevertheless, a note of caution was sounded at the meeting on the grounds that a standard might well deal with a greater variety of industries than those represented at the meeting and the standard 'might have to adopt relatively rough and ready methods in some cases'. That element of caution still applies with individuals, while acknowledging the attraction of statistical methods for others, claiming that their industry or their product was special and needed a different treatment. During the immediate postwar period, British Standards BS1313: 1947, BS2564: 1955, BS2635: 1955 and BS2846: 1957 were all produced with a view to making statistical methods more readily available to industry. It is interesting to note that, in contrast with the current fashion for anonymity, the last three standards all had named authors. In a talk given to the South Wales local group of the Royal Statistical Society in March 1972, Professor Pearson commented on the difficulty of maintaining a steady review of standards without the necessary resources of a statistician with engineering experience or an engineer with statistical experience to act as interpreter, and an adequate supporting staff within BSI. The position regarding the generation and review of standards was made both better and worse by the decision of the BSI committee in 1966 to participate in the activities of a rejuvenated Technical Committee of the International Standards Organization (ISO), TC/69, under a French secretariat. This participation introduced a new international dimension to the discussions, providing a genuine stimulus for new work as well as generating frustrations that mirror, on the small scale, those prevalent in the working of the EEC. Different countries have different priorities for the items of work and a different attitude towards the role of explicit theoretical reasoning within a standard.

A wide variety of standards in the general area of quality management and the use of statistics in quality control were produced after the mid-fifties. To provide a ready reference to seven standards dealing with terminology and the management of quality assurance, BSI compiled and issued *Handbook* 22: *quality assurance* in 1981. This handbook was followed in 1985 by three others: *Handbook* 23: *general management*, *Handbook* 24: *quality control* and *Handbook* 25: *statistical interpretation of data*. Together, they represent all facets of the commitment to quality. Individual standards of all kinds are listed in the annual catalogue, but as this contains thousands of entries, shorter sectional lists of particular interests are available. For example, *Sectional list* 44 gives better access to those titles particularly relevant to general management, quality assurance, and quality and statistics. The different standards available can be broadly classified into five categories.

(*a*) Guides to quality management advocating the use of statistics.
(*b*) Statistically based quality procedures.
(*c*) Individual statistical tests.
(*d*) Product related tests based on statistics.
(*e*) Standards devoid of any statistical contribution.

Predominant among the standards in category (*a*) is BS5750 with its several parts. This standard sets out general guidance on the required components of a quality management system in either the full context of design/development, production, installation and servicing or the reduced contexts of production and installation or final inspection and test. The standard as a whole forms the basis of assessment for registration as a 'quality assessed British company', and parts 0, 1, 2, 3 correspond to a short series ISO9000–9004. The last, ISO9004, makes explicit reference to the use of statistical methods in market analysis, product design, reliability studies, process control, acceptance sampling/inspection plans and performance assessment. It recognizes a greater role for statistics in quality management than is usually allowed, namely statistical process control and sampling inspection, and this is a challenge to statisticians to produce a code of good practice to meet this extended role. Also in category (*a*) are the various parts of BS5760 covering the reliability of products, systems and components. The statistical elements of this standard are concentrated mainly in *Part* 2: *guide to the assessment of reliability*.

Category (*b*) is, however, the one that is most often under consideration when standards and statistics are mentioned together because this category includes the control chart procedures and acceptance sampling plans that outsiders think to be the sole contribution of statistics to quality management. The two uses are intended to be served by two short series of standards, BS5700–5703 for control charts and BS6000–6002 for acceptance sampling plans. ISO equivalents for these standards either exist or are in the process of development by international working groups drawn from Europe, North America and Japan. A general introduction to control charts is given by BS5700 and this refers to control charts for attributes described in BS5701, to control charts for normally distributed variables in BS2564 curently under revision to appear as BS5702, and finally to the parts of BS5703 dealing with the application of cumulative sum (CUSUM) charts for control of mean, variability and attribute data. Those working with multinational companies may well be aware of the different attitudes which exist with regard to control chart limits. One is based on neat and tidy probabilities of falling in the warning and action zones with an in-control process and the other based on neat and tidy multiples (2 and 3) of the standard error of the plotted statistic to create these zones. For the mean there is little problem, but for other statistics the discrepancy in treatment can and often

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

does create confusion. Largely on the initiative of Great Britain, CUSUM chart methods have been introduced to the international scene. A draft recommendation, concerning an International Technical Report on CUSUM charts, has been prepared and is based on parts 1 and 2 of BS5703. It is likely to appear soon, soon that is on the international time scale where time is measured in years.

The corresponding series for acceptance sampling begins with BS6000, an introduction to the ideas contained in BS6001 such as acceptable quality level and operating characteristics. The basic BS6001, part 1, is a very close relative of the military standard 105D, but since the adoption of BS6001, part 1, *Supplementary standards* parts 2 and 3 have been produced to provide plans indexed by limiting quality and to provide procedures for the controlled skipping of sampling inspection. The international equivalents will be with the parts of ISO2859. Sampling inspection by variables is contained in BS6002 and ISO3951. The revision of this standard is leading to interesting new methods covering multivariate acceptance sampling plans described by Baillie (1987). These standards form the bulk of *Handbook* 24, the remaining standard BS5309 refers to methods of sampling from bulk material. Other standards in category (*b*) are available in *Handbook* 25. These standards relate to the use of statistics in paper testing (BS2987), in fatigue testing (BS3518), in rubber testing (BS5324) and in the assessment of precision of test methods (BS5497).

Statisticians often regard category (*c*) as a restraint upon trade, a standardization of statistics to be resisted at all costs. The motives for producing standards that appear to duplicate the textbooks are twofold. The first is to provide a reference that will not disappear from bookshelves; the second is to standardize the method of testing so that the tests can reliably be invoked by other standards. The aim is not to put standards into statistics but to put statistics into standards. The standards under this heading form the seven parts of the revised BS2846 and also the standards ISO2602, 3207, 2854, 3494 and 3301.

Category (*d*) is wide ranging. An audit carried out in 1976 indicated that at that time 35% of British Standards had some statistical or quality control content, representing 2219 standards. Of these standards, the largest proportion (75%) were specifications, with 19% devoted to methods of sampling or testing and the rest being guides, glossaries and codes of practice. The size of the problem is obvious if every revision of these standards were to require the services of a statistician. The aim is instead to provide, within a general framework, the following:

a collection of standards that can be cited for sampling inspection;

test procedures to determine significant discrepancies when comparing samples with a control or each other;

reliability assessment for equipment;

repeatability assessment for test methods.

As Stephens (1977) pointed out, standards are a new frontier for quality, being themselves the subject of standards relating to clarity of expression, appropriateness of test criteria and acceptance criteria that are unambiguous. This advice repeats that given in BS600 40 years ago, but with fresh examples of the perils of poorly specified criteria in product standards.

Looking back, I am sure that the members of that original committee in 1932 would be pleased by the level of international cooperation in the statistical community and the recent conference activity of the Institute of Statisticians and the Royal Statistical Society in the area of quality and productivity, cooperation which has culminated in the current meeting. They might be surprised by the extent of disclosure of information implicit in modern quality management systems, although they were themselves in favour of certification schemes based on quality records from production. Looking forward, there is much that remains to be done. Some computer programs appear in individual standards, but these programs form simple aids to procedures which owe nothing to computers. With expert systems being developed for statistical analysis, there is a temptation to believe that limited-objective expert systems could be produced to guide standards' writers. One thing is certain, the quality of future standards is in the hands of the users. The membership of committees that produce standards is drawn from the users and all draft standards are circulated for public comment before issue. When there is an international dimension to the standard, as is increasingly the case, this review is worldwide. The best way to influence the production of standards is to be in the committee during the gestation period; the next best way is to participate in the public comment on the draft; the least effective is to complain at the finished product. While describing the structure of standard generation, I should take this opportunity to state that all opinions expressed in this talk are personal views and do not represent the views or opinions of the British Standards Institution.

References

Morrison, S. J. 1987 *The Statistician* **36**, 439–464.
Baillie, D. H. 1987 *The Statistician* **36**, 405–478.
Stephens, K. 1977 *J. Quality Technol.* **9**, 70–81.

Discussion

M. Gerson (*Industrial Statistics Research Unit, University of Newcastle upon Tyne, U.K.*). In preparing advice concerning statistical process control (SPC), we have read quite thoroughly the British Standard *Handbook* 24 – not an easy read even for professional statisticians – and so we are some of your customers. Much of the material and advice included there is of considerable value.

However, there are areas in which the advice is either incorrect or misleading. One major concern is that the standards ignore the possible existence of different sources of variability when setting up various control charts. Some of the standards' recommendations imply important assumptions about the generating process, notably concerning stationarity of the mean. This means essentially that only variation across the process is allowed for, while short-term or long-term variation along the process is assumed to be non-existent or to consist only of 'special causes'. In some cases, notably in component manufacture, this may be justified. In many others, especially in the process industry, it is far from true and we have been teaching people to recognize and deal with such cases.

Charts which are set up to control the average level frequently indicate action with no discoverable cause. The danger is that many users will conclude either that their process is 'too uncontrolled for the method to work' or, even worse, that SPC is of no use at all.

To be fair, many commercial SPC training manuals contain the same fault. It is to be hoped that a policy of continuous improvement will rapidly be applied to these products.

D. J. G. Farlie. I am sure that the members of the sub-committee dealing with control charts will be very interested in Ms Gerson's comments, both the favourable and unfavourable aspects. The now rather elderly BS2564 does concentrate attention on the special case with no external random variability, in common with many industrial manuals. The newer standard BS5700, which provides guidance to all forms of control charts, anticipates the publication of BS5702 as a replacement for the earlier BS2564 and §4.6 of this guidance document does discuss the identification of additional sources of variation and the modification of control charts to take account of this variation. The recognition of additional sources of variation has always been part of the BS5703 standard dealing with cumulative sum procedures.

C. Chatfield (*School of Mathematical Sciences, University of Bath, U.K.*). I have served on three BSI committees over the last four years as a nominee of the Royal Statistical Society. When I first joined, I did not realize that such a wide variety of standards were not only available in quality control, but covered statistical methods more generally. For example, many statisticians are probably not aware that there is a standard, 'Tests for departures from normality', and may indeed be surprised that this is deemed a suitable topic for a standard. This raises the general question referred to by the speaker as to which statistical topics are suitable for standardization. As a member of one committee I argued strongly against a proposal to write a standard on outliers.

Initially, I was also unaware of the complex committee structure at both the national (BSI) and international (ISO) level. While I have been impressed by the dedicated efforts of many individuals in writing and revising standards, I have been disturbed by a variety of administrative and organizational problems. It appears that BSI is seriously underfunded and, for example, the statistical committees are serviced by one permanent secretary when there should really be two. At the international level there are also severe problems. For example, some of the statistical committees do not have a permanent international secretariat. As a result, development is very slow and draft documents can take years to be processed. Would Dr Farlie like to comment on these important practical issues?

D. J. G. Farlie. It is true that the support available at national and international level is less than one would like. This must force a consideration of priorities in developing or revising standards. It would help in the settling of these priorities if users were more actively concerned in the working of the various sub-committees. As I mentioned, the most effective place to influence a new standard is within the committee itself; the next most effective way is with comments when the draft is circulated for public comment and the least effective is to grumble afterwards.

F. A. Wood (*Mathematical Association, Leicester, U.K.*). Dr Farlie commented in the discussion that it is possible to have a no-inspection system, provided it is of a random type. Could he clarify this point?

D. J. G. Farlie. There is in the context of BS6001, a newly published part 3 referring to skip–lot sampling. In this, lots that form a continuing sequence of lots from the same source can be subject to a reduced form of inspection in which some lots may not be sampled at all. This extends the principle applied to sampling items from within a lot to sampling lots within a sequence of lots. Needless to say the system is subject to a number of checks including minimum rates of sampling and control on the quality management system.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

K. Sachs (*GKN Technology Limited, Wolverhampton, U.K.*). Taguchi has emphasized that quality improvement is concerned with an ever closer approach to specified target values rather than keeping within specified tolerances. The use of statistics in quality management has a long and honourable history. The postwar wave was initiated by the transfer to industry of scientists experienced in operational research during the war. It lost its impetus when industry, more concerned with output than quality, misused statistics to estimate the smallest number of samples required to convince customers that products were within tolerance limits.

The economic climate has changed and the perception that industry's salvation lies in quality improvement has generated a new wave of interest in the application of statistics. However, the danger of misuse and misunderstanding has not gone away. It applies with special force to standards.

Standards must make it clear that *products* must meet specification; it is not sufficient that the recommended test procedures are followed. Sampling patterns for acceptance, indeed all statistical techniques, are guides to help the supplier meet his obligation to produce quality products and not a substitute for it.

The revised British Standard for engineering steels (BS970: 1983) is one example of a standard that makes this point clearly, largely as a result of the efforts of the Society of Motor Manufacturers and Traders. It is due to be replaced by a new European Standard based on an unsatisfactory ISO document and there is a risk that this statement may disappear.

I hope that the committees concerned with statistical methods, in BSI, in the EEC standardization committees and in ISO, will ensure that all product specifications offer statistical sampling methods explicitly as a guide to quality improvement and not as a substitute for quality performance.

D. J. G. Farlie. I am sure that Mr Sachs will find BS5750 conforms to his wishes in the emphasis it places on the attainment of quality. Acceptance sampling standards are not designed to encourage the supply of substandard products but to emphasize the risks involved in making judgements on small samples. The surest way not to find defectives is not to look for them, or to look for them in very small samples. The standards relating to control charts help to put supervision where it is more effective, namely during the process of manufacture.