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

Sophisticated Theory and Practice in Quality Improvement

G. A. Barnard

Phil. Trans. R. Soc. Lond. A 1989 **327**, 581-589
doi: 10.1098/rsta.1989.0013

Phil. Trans. R. Soc. Lond. A **327**, 581–589 (1989)
Printed in Great Britain

Sophisticated theory and practice in quality improvement

By G. A. Barnard
Department of Mathematics, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, U.K.

Quality improvement methods must be understood by those who operate them. This has often been taken to imply that only crude statistical methods are suitable for use on the shop floor. But with the important proviso that proper explanation and motivation are needed, practical understanding of subtle procedures may well run ahead of theoretical work. Many of the major advances in statistical method made in this century have arisen in industrial contexts; and there is no reason to suppose matters will change in this respect in the future. The implications of these facts for industrial management and for the training of statisticians are explored.

1. Shop-floor statistics can be sophisticated

When, during World War II, I was concerned to introduce sequential statistical methods into the inspection of ordnance factory output, I was told by many of the factory managers that the young women who were doing the work would find the new methods hard to operate, and that for this reason the older, less efficient methods should be retained. Imagine my satisfaction when I learned from a colleague that after arguing into the night he had finally persuaded one manager to allow him to leave my instructions with the shift forewoman. When my colleague arrived late for the 6.00 a.m. shift next morning he found the inspectors enjoying the change from the old routine and in fact welcoming the new methods as being much more commonsensical than the old ones. The precise mathematical theory of the new test procedures involved the then unfamiliar theory of random walks and it was this that led the managers to take the view they did. They failed to see that, while the mathematics was necessary to determine just what numerical combinations should lead to acceptance or rejection, the general ideas involved were no different from those applied in scoring a tennis match which is not played over a fixed length of time, but rather stops when the evidence of superiority of one player or the other is sufficiently strong.

The example of sequential testing is not alone in providing an instance in which advances in 'shop-floor' practice have been based on recent advances in statistical theory, or even where such practice has promoted such advances. My theme in this paper is to help counter the misapprehension, easily derived from much of the literature on quality control (though not, of course, from Dr Deming's books), that the average engineer who has taken a course or two in the basics of statistics and probability is capable of solving on his own all the statistical problems likely to present themselves. That he should be capable of understanding solutions to such problems has already been indicated; but recognition of the existence of such problems, their correct formulation, and the working out of solutions typically call for statistical expertise of a high order. Such expertise is in short supply in this country. Much of it is currently being taken up by the expanding areas of medical statistics, econometric forecasting, and so forth, with the result that industry is failing to acquire its proper share. A dangerous gap has opened up between university training in statistics and industrial applications with the result that such courses as are given in our universities and polytechnics often leave the impression that statistical quality control is a matter of following simple rule-of-thumb procedures with very little in the way of intellectual challenge. I want to suggest that not only does industry stand in need of high-level statistics and statisticians; but high-level statistics needs to have regard to its origins, much of them in the industrial field, as a foundation for healthy development.

2. Two examples of sophisticated statistics

Two instances may be given to show how relatively sophisticated probabilistic considerations may enter into the solution of practical problems in industry. The first may be regarded as an early example of the type of design principle with which Dr Taguchi's name is now associated. One of the devices with which the writer was once concerned was a resistance–capacitance (RC) circuit whose time constant $k = RC$ was required to be held within very close limits. Neither the capacitors nor the resistors could be manufactured within the limits required, so that selective assembly, matching the resistance R to the measured capacitance C, was necessary. Selective assembly in such cases produces tremendous difficulties in balancing the numbers of components to be used. These could be overcome by forming the required resistance from two resistors, $R_1+R_2 = R$, connected in series. The problem then arose of writing a specification for the resistor manufacturers in such a way that the known distribution $\Phi(R)$ of the required values of R could be obtained without waste of materials. Figure 1 shows the assembly bench on which the resistors were put together with the capacitors. The resistors R_i were sorted by size into the boxes with coded values 1, 2, 3, ..., 12 and 12, 11, 10, ..., 1 as indicated. The pointer attached to the lower set of boxes indicates that the total resistance of any pair of resistors taken from boxes in line with each other will be 6. If a capacitor is found to require a total resistance of 18, then, moving the lower set of boxes to the left (thus causing the upper set to move right), twelve steps will arrange the boxes so as to present the assembler with sets of pairs adding to 18. There will always be a number of options from which to choose, corresponding to various values of the difference $D = R_2-R_1$.

We therefore have $R_1 = \frac{1}{2}(R-D)$ and $R_2 = \frac{1}{2}(R+D)$, and if D is statistically independent of R and its distribution is symmetric, the distribution of R_i is formed by the convolution of Φ with

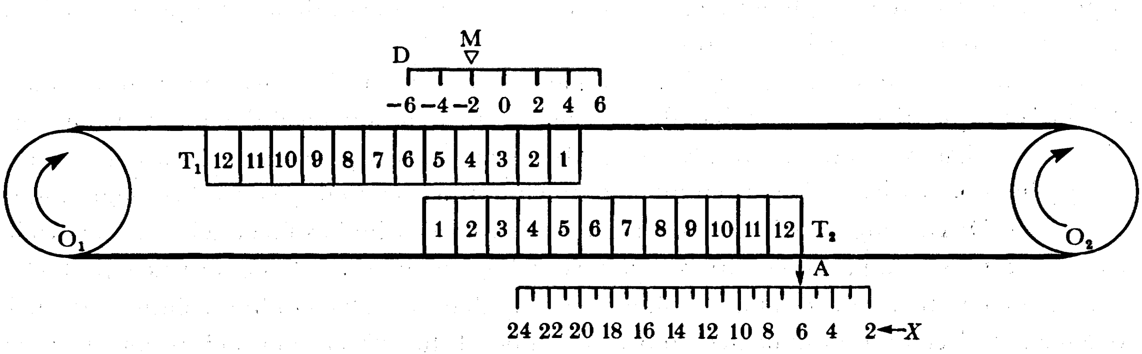

Figure 1. The circles O_1 and O_2 represent pulleys carrying the rope to which the sets of trays T_1 and T_2 are attached. When a pair of resistors R_1, R_2 is required with $R_1+R_2 = X$, the lower arrow A, fixed to the rope and the trays, is moved to the mark X on the scale (fixed to the bench). In the figure shown, $X = 6$. Then the resistors R_1 and R_2 from any pair of opposite trays will total X. The difference R_2-R_1 is then set on the scale D, fixed to the bench. The marker M specifying which pair is to be chosen is moved from day to day to generate the required distribution of R_2-R_1.

the distribution of D and rescaling by $\frac{1}{2}$. To avoid component imbalance it is therefore enough that the distribution of the resistors should be capable of representation as such a convolution. If, as was the case in the problem cited, the distribution of required resistors and manufactured resistors is very nearly normal, it is enough that the mean of the required distribution and the mean of the manufactured distribution should agree, and that the variance of the manufactured distribution should be larger than that of Φ. By putting a mark on the bench, moved from day to day to follow a normal distribution with zero mean and variance the difference of the two variances, it was possible to ensure efficient assembly without component imbalance. Such manual operations as described here would, of course, be out of place in a modern factory, but mechanization of the processes involved is easy to envisage. It may come as a surprise to some, however, that the theory of divisibility of distributions, developed by the late Paul Lêvy and implicit in this analysis, should prove relevant to such issues.

A second example is provided by the need to design a quality control system to cover the application of the British Standard Kite Mark to condoms. Special features of this product are that production is essentially continuous, with no rational basis for subdivision into lots and that as the use of the product is singular there is no possibility, should an item fail, of replacing it with another good item. A less uncommon feature is the extremely low permissible rate of defectives, which is secured in this, as in other cases, by over-stressing; for example, items tested are required to withstand inflation to a length of about four feet. The control scheme proposed for this problem, to be found in British Standard 3604, calls for testing not less than 1% of production, and keeping a record of the cumulative number N tested, and the cumulative number R of items failing any one of the tests. It is required that, at all times, the inequality

$$R < 0.01N + 3\surd(0.01N) \tag{2.1}$$

should remain satisfied. If (2.1) is rewritten in terms of the standardized variable

$$Z = (R - Np)/\surd(Npq),$$

where $p = 1-q$ is the probability that an item will fail the test, it becomes

$$Z < (0.01-p)\surd N/\surd(pq) + 0.3/\surd(pq). \tag{2.2}$$

Now Khintchine's law of the iterated logarithm tells us that as N increases, the sequence of successive maxima of the Z values increases like $\surd(2\ln\ln N)$, with probability 1. But if, the coefficient of $\surd N$ in (2.2) is negative, the right-hand side decrease like $-\surd N$, and the inequality is bound (with probability 1) to be violated. If, on the other hand, the coefficient of $\surd N$ is positive, the probability that the condition (2.2) will be violated will be very small (except early on, when special provisions may be made). This scheme, published in 1964 (but see Barnard 1969), provided the first example of a 'test of power 1', in so far as it can be regarded as a test of the hypothesis $p < 0.01$ against alternative $p > 0.01$. Further refinements of such tests were made by Robbins & Siegmund (1969). One may remark that similar ideas could be applied in the supervision of weights and measures, where much unnecessary expenditure on the accuracy of weighing and measuring could be avoided.

It may well be argued that in this second example it was not necessary that those concerned should permanently employ a statistician. It was in fact done by way of consultancy. But for the first example it might well not have been clear to the production engineers concerned with the assembly problem that statisticians could prove useful in solving the problem of component balance. Another example with a similar message is provided by the use by the Honda Motor Cycle Company of power spectrum analysis for the adaptive design of front suspensions to suit the differing road characteristics in various parts of the world. Such a design problem is primarily one for a mechanical engineer who could not be expected to be aware of the techniques available. The possibility of doing this was in fact pointed out by a statistician charged by Honda with general oversight for quality in design as well as in production. As Dr Deming has repeatedly pointed out, the management of any major concern that thinks it is enough to issue pamphlets on control charts to its foremen, without employing in a senior position a really first-class statistician, will be making a serious mistake.

3. Modern statistics had its origin in industrial problems

The examples we have given might be said to represent extremals in applied statistics, in so far as they are not representative of the wide range of more common statistical methods: regression analysis, forecasting and control theory, experimental design, not to mention the more recent developments coming from the Center for Quality and Productivity Improvements at Wisconsin, which are, or one hopes soon will be, in daily use in many industries. But references to these more central statistical techniques, with their widespread use in econometrics, commercial, medical and agricultural applications should prompt a reminder that many of these techniques, in their twentieth-century form, were developed by statisticians working in industry. Indeed, the man who more than any other can be regarded as the founder of twentieth-century statistics – the statistics of small samples – was the industrial brewer W. S. Gosset. Sent by the Guinness Company to learn statistical method from Karl Pearson, Gosset soon saw the inapplicability of the large sample pearsonian ideas to the testing of malting samples of barley, and in 1908 he published, under the pseudonym of Student, the paper on the probable error of a mean which, after further development by him and by the two men who can both be regarded as his followers (Egon Pearson and Ronald Fisher), laid the foundations of the modern theory of statistical inference. While Fisher found a fruitful field of application in a biometric direction, the younger Pearson applied ideas to some extent suggested to him by Student within the industrial field.

It is a matter for speculation how much of the doctrinal difference which grew up between Egon Pearson and Fisher owed its origin to the differences between their areas of primary interest. Personal relations were never easy, largely because of Egon's very strong sense of filial duty to Karl Pearson. The essence of their doctrinal difference seems to me to lie in the fact that Fisher adopted a conditional attitude to statistical inference, whereas Pearson retained a marginal view. I should perhaps point out that neither Pearson nor Fisher ever shared Neyman's view that the term 'statistical inference' is a misnomer, and that the proper study of mathematical statisticians is 'inductive behaviour'. That a conditional approach should seem natural to Fisher can be understood because the interval between the planning of an agricultural experiment and the collection of the final results is typically at least a year, during which there is plenty of time for ancillary information, bearing on the interpretation of the final results, to become available. In industry, on the other hand, the interval between the planning and the results stages is typically short, and should, for example, one of the planned observations be lost, it is usually feasible to make up the lack by a further observation. This fact, together with the semi-routine character of much of the testing that is done in industry, means

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

that in an industrial context a marginal approach will often appear more natural than a conditional one.

It is sometimes thought that the distinction between a conditional and a marginal approach is important only if we are concerned to focus on a particular case rather than on long-run behaviour. This is not so. In the problem considered in the next section, we may want our confidence intervals for μ to have shortest average length, subject to a given overall coverage frequency. To do this we need to keep constant, from case to case, the value of h in (4.12), not the value of α. This will mean, except in the normal case, that the conditional confidence will need to vary from sample to sample. Alternatively, we may want our conditional confidence to stay constant, at the expense of average length. In practice some compromise is likely.

The relative rapidity with which results are typically obtained in industry as compared with biometrics is not the only reason why a marginal approach to industrial problems is often appropriate. If we are laying down a set of rules for routine inspection, it clearly must be the long-run behaviour of the rules which is of primary concern. This fact was understood by Fisher (1955), though sadly his expression of understanding suffers from gross overemphasis of the distinction between pure scientific research and the solution of industrial problems. A full understanding requires both the marginal and the conditional properties of statistical procedures, and indeed it is one reason why industrial applications are of importance in the foundations of statistical inference that such applications serve to illustrate this need particularly well. One may in this connection refer to the work of the late Jack Kiefer (1982) on what might be termed a marginal approach to conditional inference.

4. A small statistical novelty: Fechner distributions

While very many, if not the majority of distributions met with in biometric contexts, may be taken without serious error as normal or transformable to normality, this is by no means the case with the distributions met with in industrial contexts. The breaking strengths of cotton fibres or of steel bars provide examples of distributions which are skew in ways which are not readily corrected by simple transformations such as logarithmic or power transformations. It is therefore not surprising that one of the earliest questions which Gosset put to Egon Pearson was to enquire into the effect of departures from normality on the distribution of Student's t. Pearson, 'assisted by N. K. Adyanthaya', responded with two papers (1928; 1929). By laborious random sampling procedures he showed that, provided the distributions involves were symmetrical, the t test was what would now, after the work of Box, be called robust. But this comforting conclusion would not hold if there were a serious degree of skewness. Both Fisher and Pearson then made attempts to derive further results in this direction, but they were defeated by lack of computational facilities. Only very recently has it become possible in practice to look at this problem again and to obtain useful results using the vastly increased computing power now available from desk-top computers. I hope it will not be thought out of place to give a summary account of some results in this area; the more so because use of the Taguchi and related methods of engineering design, which aim to eliminate first-order terms in the input–output relationships, may be expected to produce even more products in which departures from specification values exhibit skewness in their distribution.

We need first of all to specify a family of distributions which may be expected to represent most of the types of skewness liable to arise in practice. An approach to such a family can be obtained by replacing the function u^2 involved in the standard normal density by the easily programmed family of functions,

$$M^a(u) = \begin{cases} u^a & \text{for} \quad u \geqslant 0, \\ (-Mu)^a & \text{for} \quad u \leqslant 0. \end{cases} \tag{4.1}$$

So we define our family of distribution shapes by the family of densities,

$$\phi(u) = K \exp\left(-\tfrac{1}{2}M^a(u)\right), \tag{4.2}$$

where K is given by

$$1/K = (1+1/M)\, 2^{1/a}\, \Gamma(1/a)/a. \tag{4.3}$$

This might be called the Fechner family, because by allowing M to differ from 1 it embodies the idea, suggested in the posthumous *Kollectivmasslehre* of G. T. Fechner (1897), of having different scales for deviations in a positive direction from those in a negative direction. It also allows for 'lepto-' or 'platy-kurtosis' by allowing a to differ from 2. The key property of the family which we use subsequently is, that the logarithm of the density is homogeneous of degree a. For any positive λ,

$$M^a(\lambda u) = \lambda^a M^a(u).$$

If we have a sample x_i, $i = 1, 2, \ldots, n$, of n items following a Fechner distribution with location parameter μ and scale parameter σ, we put

$$p_i = (x_i - \mu)/\sigma, \tag{4.4}$$

and the joint density of the p_i will then be

$$K^n \exp\left(-\tfrac{1}{2}\sum_i M^a(p_i)\right). \tag{4.5}$$

We suppose that M and a are known, for the time being, and μ and σ are unknown. Note that μ is the mode of the distribution and not necessarily the mean, while σ is a scale parameter equal to the standard deviation if $M = 1$ and $a = 2$ but not, in general, otherwise. The mean and the standard deviation, if needed, can be calculated as $\mu + \kappa\sigma$ and $\nu\sigma$, where κ and ν are given by

$$\kappa = [1-(1/M)]\, 2^{1/a}\left(\frac{\Gamma(2/a)}{\Gamma(1/a)}\right),$$

$$\nu^2 + \kappa^2 = \left[\frac{(M^3+1)}{(M^3+M)}\right] 2^{2/a}\left[\frac{\Gamma(3/a)}{\Gamma(1/a)}\right].$$

In particular, we shall use below the values $\kappa = 0.3989$ and $\nu = 0.7687$ for $(M, a) = (2, 2)$.

To obtain the distribution of Student's t we set

$$p_i = s(t + c_i), \tag{4.6}$$

where s is required to be non-negative and the c_i are required to satisfy

$$\sum_i c_i = 0, \quad \sum_i c_i^2 = n(n-1). \tag{4.7}$$

A little algebra then shows that our t is indeed Student's.

$$t = (\bar{x} - \mu)\sqrt{n}/s_x,$$

while

$$s = s_x/\sigma\sqrt{n} \quad \text{and} \quad c_i = (x_i - \bar{x})\sqrt{n}/s_x. \tag{4.8}$$

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

Here $\bar{x}$ and s_x are the usual sample mean and standard deviation. The jacobian of the transformation (4.6) is $s^{n-1}\Delta(c)$, where $\Delta(c)$ is a determinant depending only on the c_i, so we find the joint density of (s, t, c_i) as

$$K^n \Delta(c)\, s^{n-1} \exp\left[-\tfrac{1}{2}\sum_i M^a(s(t+c_i))\right],$$

which, using the homogeneity of M^a can be written

$$K^n \Delta(c)\, s^{n-1} \exp\left[-\tfrac{1}{2}s^a \sum_i M^a(t+c_i)\right]. \tag{4.9}$$

Now if we adopt the fisherian approach, we can imagine that we learn the values of the c_i before we learn the values of $\bar{x}$ and s_x. The c_i tell us nothing about μ or σ, but they do tell us something about the relevant distribution of t and s which, when we later learn the values of $\bar{x}$ and s_x, shall in turn tell us about μ and σ. The relevant joint density of t and s is obtained from (4.9) by treating the c_i as known, and it is therefore

$$\psi(s, t \mid c) = K' s^{n-1} \exp\left[-\tfrac{1}{2}s^a \sum_i M^a(t+c_i)\right], \tag{4.10}$$

where K' is a normalizing constant determined by the condition that ψ should integrate to 1. If we want only the distribution of t we integrate out s and obtain

$$\xi(t \mid c) = K'' / \left[\sum_i M^a(t+c_i)\right]^{n/a}, \tag{4.11}$$

where K'' is another normalizing constant, determined by the condition that ξ should integrate to 1.

Just by way of confirmation, we may note that for the case of normality $(M, a) = (1, 2)$,

$$\left[\sum_i M^a(t+c_i)\right]^{n/2} = [n(n-1)]^{n/2}[1+t^2/(n-1)]^{n/2},$$

and so the density (4.11) can be seen to be equivalent to the traditional form of Student's distribution on $n-1$ degrees of freedom. Because of the conditions (4.7) imposed on the c_i, the observed values of these do not appear in the expression for ξ in the case of normality. This is why the normal distribution is exceptional in making the conditional and the marginal approaches equivalent.

In practice we use t to find a 'confidence distribution' $\Xi(\mu)$ for μ defined by the property that if h is chosen so that

$$\int_{\Xi > h} \Xi(\mu)\, d\mu = 1-\alpha, \tag{4.12}$$

then the set $\{\mu : \Xi(\mu) > h)$ is the shortest confidence set for μ having confidence coefficient $1-\alpha$. With a little algebra it can be seen from (4.11) that for any Fechner distribution the confidence distribution for μ has density

$$\Xi(\mu) = K^* / \left[\sum_i M^a(x_i-\mu)\right]^{n/a}, \tag{13}$$

where, again, K^* is a normalizing constant.

For actual computation we replace K^* in (4.13) by $[\sum_i M^a(x_i-\bar{x})]^{n/a}$ and denote the result by $\bar{\Xi}(\mu)$. We then use an integration routine to evaluate the integral of $\bar{\Xi}(\mu)$ over a wide range, bearing in mind that, if necessary, we can estimate tail areas of the integral of $\bar{\Xi}(\mu)$ by using the fact that the denominator is larger than $n^{n/a}(\max x_i - \mu)^n$ in the upper tail, and similarly

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

in the lower tail. The constant K^* is thus evaluated, whereupon further use of integration routines will give the confidence coefficients. Although the description of these procedures may sound complicated, the operations involved are all practicable on a hand-held calculator.

The confidence coefficients have the standard property that if used repeatedly with the same value of α, the true value μ will be covered with frequency $1-\alpha$. But the procedure suggested has the advantage that it possible to adjust the confidence coefficient to the precision of the sample to hand in such a way as to maximize the long-run precision. When α is varied from case to case, of course, the coverage frequency is the average of the $1-\alpha$ used. As remarked above, if we really wish to minimize the average length of our confidence intervals, subject to a given long-run coverage frequency, we should determined h as a function of the c_i so as to achieve this. In practice an empirical approach could be used.

Thus the question put by Student to Pearson a little more than fifty years ago can now be answered in a reasonably satisfactory manner. The logic of the approach given here can, of course, be used with a completely arbitrary distribution; the advantage of the Fechner family is that, while a broad spectrum of unimodal distributions can be satisfactorily approximated within this family, the integral involved in passing from (4.11) to (4.12) can be expressed in closed form, leaving only one numerical quadrature to be performed. Had Pearson lived to see such procedures in practical use, it is hard to believe he would have failed to see the theoretical advantages of the conditional approach to inference.

If we are concerned to estimate the mean of the distribution rather than the mode, the theory becomes more complicated, but the practical application remains within range of modern computers. For $(M, a) = (2, 2)$, for instance, the mean is $\mu+0.3989\sigma = \theta$, say. Then we need to find a function of (s, t) which involves θ and the observed $\bar{x}$ and s_x alone. The only such function is

$$u = t-0.3989/s = (\bar{x}-\theta)\sqrt{n}/s_x.$$

Transforming (s, t) to (u, r), with $r = 1/s$, the jacobian is $1/r^2$ and the conditional joint density of (u, r) is

$$K^* r^{-(n+1)} \exp\left[-\tfrac{1}{2}\sum_i M^a (u+0.3989r+c_i)/r^a\right], \tag{4.13}$$

from which we can integrate r to obtain the marginal density of u. This then can be used in the same way as was t to obtain the confidence distribution of θ.

5. Industrial application and the foundations of statistics

Box (1980) has repeatedly emphasized the iterative nature of scientific inference, involving as it does experimentation, model building, estimation, model criticism, further experimentation and so on. The fact that this is the procedure by which both science and technology are advanced is more apparent in industrial contexts than in areas such as biometry where a single experimentation stage may take up most of the total time, and hence become the sole focus of attention. This has resulted in far too much of the existing statistical literature and textbooks being concentrated on the interpretation of single experiments considered alone. When we are considering a single experiment, the marginal and the conditional aspects present themselves as exclusive alternatives; but if the single experiment is viewed in the context of the iterative sequence of which it is a member, marginal aspects can be seen a supplementing conditional aspects, both being necessary to the proper exercise of judgement. One way of

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

integrating the two aspects is to imagine that we know something of the long-run behaviour of the system with which we are dealing. Such supposed information may be expressible as a bayesian prior distribution for the unknown parameters involved; but a realistic assessment of what is actually known and what remains unknown will usually require that any such prior distribution should be regarded as no more than a potentially useful guess, to be abandoned should model criticism show it to be implausible. In this way we may arrive at a view of the foundations of statistics which does not differ essentially from the 'ecumenical' view propounded in Box's paper.

6. Conclusion

The main part of my remarks has been addressed to statisticians, in the hope that more of them will see the possibilities in the industrial field, in spite of the attractions of other fields in which the value of mathematical statistics is better recognized. But the death last month of Frank Nixon, who – as chief quality engineer for Rolls Royce – was so active in the promotion years ago of the National Council for Quality and Reliability, will remind those of us who remember him what an uphill struggle it has been to persuade management in this country of the need for proper attention to quality. The message has at last sunk in; but long-term difficulties in solving the problem have been allowed to develop. Before World War I, and for a time after it, this country had a reputation second to none for the quality of its manufactures, largely based on the existence here of a large body of highly skilled and intelligent mechanics and artisans. As Dr S. J. Prais has repeatedly pointed out (1988), our educational system has for a long time been failing to produce a body of people possessing the kinds of skill which nowadays correspond to those of our former artisans. It will not be easy in the immediate future to remedy this lack. And this is to say nothing of the pressing problems in higher education of which most of my audience will be all too well aware. We need to re-create the sense of urgency which some of us experienced in the 1950s and 1960s. Fortunately this meeting, and other recent signs of progress, suggest that this time our efforts will be better received.

References

Barnard, G. A. 1969 Practical applications of tests with power one. *Bull. Inst. int. statist.* **43**, 389–393.

Box, G. E. P. 1980 Sampling and Bayes' inference in scientific modelling and robustness. *Jl R. statist. Soc.* A **143**, 383–404.

Fechner, G. T. 1897 *Kollectivemasslehre*. Leipzig: Wilhelm Englemann.

Fisher, R. A. 1955 Statistical methods and scientific induction. *Jl R. statist. Soc.* B **17**, 69–78.

Kiefer, J. C. 1982 Conditional inference. In *Encyclopedia of Statistical Sciences*, vol. 2 (ed. N. L. Johnson, S. Kotz & C. Read). New York: Wiley.

Pearson, E. S. & Adyanthaya, N. K. 1928 The distribution of frequency constants in small samples from symmetrical populations. *Biometrika* **20**, 356–360.

Pearson, E. S. & Adyanthaya, N. K. 1929 The distribution of frequency constants in small samples from non-normal symmetrical and skew populations. *Biometrika* **21**, 259–286.

Prais, S. J. 1988 *Rev. natn. Inst. econ soc. Res.* Feb. 1988.

Robbins, H. & Siegmund, D. 1969 Confidence sequences and interminable tests. *Bull. Inst. int. statist.* **43**, 379–387.