USER-FRIENDLY STATISTICAL CONCEPTS FOR PROCESS MONITORING

Draft of article published in the *Journal of the Operational Research Society, 49*(9), 976-985, 1998.

Michael Wood, Nick Capon and Michael Kaye
Management and Decision Support Unit
Department of Accounting and Management Science
University of Portsmouth
13 January 1998

Abstract
Many standard statistical process control techniques involve sophisticated mathematical concepts, which are frequently misunderstood and misused by their users. This means, the paper argues, that the techniques, and the terminology and concepts underlying them, are inappropriate for their intended uses and users. The paper considers the areas of the statistical background which cause difficulties. It goes on to describe some alternative concepts, techniques and terminology - which are likely to be conceptually simpler and more "user-friendly" (and in some cases more accurate and robust). None of the ideas require users to be familiar with the standard deviation. We also suggest that the approach of reengineering the conceptual background to suit the context of users and uses may be appropriate to other areas of practical statistics and OR. The paper identifies some general principles for achieving this.

Key words: OR education, Quality, Statistical process control, Statistics.

Introduction
The difficulties non-mathematicians often experience with mathematical areas of statistics and OR have been noted frequently. The consequences are obvious and important: for example, the techniques in question are often not used when they would yield substantial benefits, and when they are used they are sometimes applied in ways that are - to the experienced practitioner - silly or counter-productive.

A related issue is the widely acknowledged importance of emphasising the *process* of solving problems, and the *spirit* or *philosophy* of statistics and operational research, as opposed to the view of statistics and OR as simply collections of mathematical techniques[1-4]. Without this broader view, collections of techniques are generally of little value.

However, mastery of the appropriate mathematical concepts and techniques is clearly necessary to implement the philosophy and solve real problems successfully. The danger is that novices will fail to master the relevant techniques, and so not be in a position to consider the broader issues and use their understanding fruitfully. Even if they do manage to master the techniques, there is still a danger that the effort expended may mislead them into thinking that the main problems in implementing OR techniques are mathematical, and that any other issues are relatively trivial.

This paper considers this issue in relation to statistical process control (SPC) - an area in which sophisticated mathematical concepts are widely used by non-specialists. It proposes a radical solution: reengineering the statistical terminology, concepts and techniques used to make them more appropriate for the context of use - ie conceptually simpler and more "user-friendly". The hope is then that this will enable novices to develop a fuller and more flexible understanding of the basic statistical principles, which in turn should lead to a better appreciation of the philosophy of SPC, and more effective process monitoring.

From the perspective of most beginners, the statistical theory necessary for SPC divides into two categories: ideas which are easy to understand; and ideas which are difficult to understand and may be almost incomprehensible to many novices. Line graphs, histograms, averages, scatter plots, and so on, fall in the "easy" category. (This is not to say that they will necessarily occur to novices without guidance.) In the "difficult" category we may put:

1 the standard deviation and the normal distribution;
2 capability indices (eg c_{pk});
3 the interpretation of control charts.
4 the standard mathematical models and their associated algorithms for calculating limits for control charts (mean charts, range charts, proportion defective charts and so on);

This may seem unduly pessimistic - especially the first item on this list. However, there is evidence to suggest that, in the UK at least, even the simpler items on this list cause very considerable difficulties for novices. In a previous paper[5] we have described some of the problems experienced with concepts such as the standard deviation - described as "horrendous" by one manager - and the use of the standard algorithms for calculating control chart limits - where mistakes in calculations or interpretation may render the results useless.

The usual (and laudable) rhetoric behind SPC generally emphasises the fact that the procedures are owned, used and preferably initiated and designed by process operators themselves. This is clearly difficult if the underlying statistical theory is too complex for the process operators to understand. (In practice SPC, and specific statistical procedures, may be forced on organisations by customers or regulating authorities[6].) This paper suggests some alternative approaches to the difficult areas in the list above. These alternatives are based on more "user-friendly" concepts than the usual ones. Some of these alternatives have been proposed in an earlier paper[7]; the present paper extends these suggestions and provides further examples of their use. (The suggested alternatives for 1 and 2 are not covered in the earlier paper, and the approach to 3 presented here is rather different.) One of the proposals is an approach to 4 using the technique of resampling[8-10]; space limitations mean that this is described only very briefly in the present paper. For further details the reader is referred to the earlier paper[7]; we also hope to produce another paper exploring this issue in more depth soon.

This paper is not concerned with the philosophy of SPC or with what is worth monitoring and the likely benefits. Our concern is simply with suggesting more appropriate alternatives to the statistical concepts and techniques in the list above. However, these more user-friendly alternatives should enable novice users to make more realistic judgments on these broader issues.

The paper explains the proposed methods by giving a brief example of the methods in action. However, this is just one example: different approaches will be appropriate to different situations. The potential benefit of the approaches suggested here is that the improved understanding of the statistical aspects of the procedures means that users can be more flexible than they can be with procedures which are only vaguely understood.

These methods are intended for beginners, but the reader should bear in mind that the explanations given here are much briefer and more abstract than would be appropriate in an explanation for novices. The reader should also note that we have delayed a discussion of some more technical issues, and potential objections to the proposals here, to the section on *queries and objections answered*.

The standard deviation and the normal distribution

The first difficult concept on the list is the standard deviation. This, as opposed to the variation which it measures, is seen by many novices as the core concept of statistics, but it is also the point at which many start to get lost. There are alternative, more transparent, ways of assessing variability. The most useful are based round order statistics like quartiles and percentiles, the interquartile range, or box-and-whisker plots. These have the advantage that their meaning can be easily described in simple language (eg 75% of the sample are above the lower quartile) whereas this is not true of the standard deviation. The definition of the standard deviation inevitably prompts questions about why the deviations from the mean are squared - answers to which must be based on considerations of mathematical convenience, which is an unsatisfactory basis for a user-friendly concept.

There is another reason why measures based on order statistics are preferable to concepts like the standard deviation. The definition of action lines on control charts in terms of 0.1% of normal variation lying above the upper line, and another 0.1% lying below the lower line, is, in effect, a statement about percentiles. The same is true of other concepts such as confidence intervals. *In this sense percentiles, quartiles, and quantiles in general, are the approach to assessing variability which provides a direct link to the intuitive idea of variability or spread.* The choice is not between learning about percentiles or the standard deviation, it is between learning about *percentiles only*, or about *the standard deviation and percentiles*. (There is little point in understanding what the standard deviation is without an appreciation of the fact that the "three sigma" limits extend from the 0.1 percentile to the 99.9 percentile.) The novice's task is clearly easier if (s)he can make do with percentiles and quartiles, and forget the standard deviation..

We will take the difference between the median and one of the quartiles - the *quartile deviation* or *QD* - as the basic measure of spread in place of the standard deviation. For example, for a distribution with

Upper quartile	= 24 units
Median	= 20 units
Lower quartile	= 17 units

the QD is 4 units for the upper part of the distribution and 3 units for the lower part. The reason for the discrepancy may be that the underlying distribution is asymmetrical, or it may be due to the particular sample we are using. In the first case we can distinguish between the upper and lower values:

Upper quartile deviation (UQD) = 4 units
Lower quartile deviation (LQD) = 3 units

In the second case we can assume the values are the same and use their mean as an estimate:

Quartile deviation (QD) = 3.5 units

The QD is obviously equal to half the difference between the quartiles - which is sometimes referred to as the *semi-interquartile range*. It does not depend on the median.

One of the main purposes of using the standard deviation is to estimate probabilities from the normal distribution. It is a simple matter to rescale the normal distribution in terms of QDs instead of SDs - as in the Appendix.

Explaining the rationale behind the table in the Appendix in mathematical terms is not easy. However, experiments with real or simulated data which show the normal, bell-shaped pattern should convince even sceptical beginners that the relationship between the number of quartile deviations from the median and the percentiles of the distribution follows the Appendix very closely. (It is not always necessary to use tables such as these even when working with distributions which are, in effect, normal: the resampling approach[7-10], discussed briefly below, involves simulating distributions, which means that abstract models, such as the normal distribution, are not necessary.)

None of the methods advocated below makes use of the standard deviation. Where a quantitative assessment of spread is required they refer to the QD, UQD or LQD instead of the standard deviation. The advantage is that these statistics are likely to be more transparent, easier to calculate, and also avoid the problem that the standard deviation is not a *resistant* measure: it is likely to be unduly affected by extreme values[11].

Measures of capability

The remaining examples in this paper are based on data which refer to samples from a colour printing process within a packaging company. The values recorded are densitometer readings (multiplied by 1000 to reduce the number of decimal places) showing the variation in a particular colour over a representative sampling period. The agreed specification is 1600 - 1800. Figure 1 is a histogram of 34 samples of six densitometer readings (204 readings in all) with the tolerance limits superimposed.

FIGURE 1 HERE

Figure 1 is a very simple way of assessing capability. Diagrams such as these are simple to understand, easily drawn, and give a clear indication if the process is capable of meeting the required specification: in the present case it is clearly not capable.

Capability is often measured by means of a capability index such as c_{pk}. This is only meaningful if the measurements are normally distributed and the process is in statistical control[12], neither of which is true of the data shown in Figure 1. (It should be obvious that the data is not normal from the histogram, and the control chart for the mean set up to see if the process was in statistical control found that 9 of the 34 sample means were outside the action limits.) Table 1 shows part of a simulated data set which is normally distributed and in statistical control: we will use this in our initial discussion of capability measures and control charts because we can be sure that it satisfies the necessary assumptions. We can reasonably assume that this data is typical of the printing process once it has been thoroughly stabilised by eliminating all "special causes".

TABLE 1 HERE

The value of c_{pk} derived from the data in Table 1 is

$c_{pk} = 0.45$

using the standard method involving estimating the standard deviation from the average range of the sample[13]. This low value indicates that the process is not capable of meeting the specification.

There are two difficulties with this method of measuring the capability. First, it requires an understanding of the meaning of the index. This is not trivial: a value of 1 would indicate a defect rate of around one or two parts per thousand, with higher values of the index corresponding to more capable processes with lower defect rates. Second, the index is only useful for normally distributed measurements, and provides an assessment of "potential capability" *if* the process is "in control". It cannot be used for the data shown in Figure 1. The use of a diagram such as Figure 1 is far more transparent and avoids both of these difficulties.

However, a numerical measure may sometimes be useful to provide a simple summary of the situation. The obvious alternative to capability indices, such as c_{pk}, is to use the much more straightforward measure "parts per million out of tolerance"[14] (or percent out of tolerance). It is quite difficult to see what advantages c_{pk} has over a simple *parts per million* measurement except that the former is more esoteric and serves to mystify what is essentially a simple concept.

For these reasons we suggest estimating the parts per million defective (ppmd) as an *incapability index*. There are several ways in which this could be done, of which we will now review two.

1. Directly from the empirical data

49 of the 600 measurements recorded are outside the specification limits, so the estimated *incapability* from this method is

ppmd = 81,667, or *8.2% defective*.

This is obviously not likely to be accurate over the long term, and does not enable extrapolation to rare events not represented in the sample (eg the chance of a measurement being less than 1500), but it is a very simple method. When the only data available is attribute data, or when there is no *a priori* distribution which can be assumed, this is the only possible method.

2. Using the data to fit a normal distribution

Obviously the first thing to do is to check that the data is roughly normally distributed - in the case of the data in Table 1, this is clearly true. To use the normal table in the Appendix, we need to work out the quartiles (1718 and 1641) and the median (1680) of the data (simply by arranging the data in order of size and reading of the appropriate quantiles). The UQD and LQD can then be worked out:

UQD = 38

LQD = 39

We can combine these to calculate a combined estimate of the quartile deviation (QD) of 38.5. This means that the upper specification limit is $(1800-1680)/38.5 = 3.1$ QDs above the median, and the lower specification limit is $(1680-1600)/38.5 = 2.1$ QDs below the median. The Appendix then indicates that 1.82% of the output from the process will be above the upper specification limit, and 7.84% will be below the lower limit. This gives a predicted defect rate, or *estimated incapability*, of 9.7% - which is close to the empirical rate of 8.2%. If the lower specification limit had been 1500 and we had wanted to estimate the percentage of the output less than 1500, this method can be used to produce such an estimate (0.1%).

It is important to remember that the data is used to derive the median and quartiles only; estimates of the tails of the distribution are worked out from the central 50%. If there are outliers which would not be expected from the pattern of this central 50%, the incapability index, in effect, assumes that these outliers do not exist.

Interpretation of control charts

Shewhart control charts (mean charts, range charts, "p" charts, "c" charts, etc) are a widely used way of monitoring a process. They comprise a line graph showing the values of the appropriate statistic (mean, range, etc) estimated from samples taken at regular points in time. *Control* lines are superimposed to indicate statistically significant departures from the normal performance - corresponding to *special causes* of variation which should be investigated. Special causes detected by charts for the colour printing process, for example, included set-up problems and problems with the ink.

On a conventional control chart the statistical control lines are marked as values on the vertical axis, so the uninformed viewer might expect them to refer to critical values of, in the example here, the densitometer reading. This leads on to the natural, and very common, misinterpretation of the statistical control lines as some kind of tolerance interval. A process which is "out of control" is assumed not to be delivering what customers want.

This interpretation is not, of course, accurate at all: the statistical control lines mark points on a very different scale measuring the strength of the evidence for a change from normal performance. At first sight this may appear to be a minor quibble. However, anecdotal evidence suggests that this misinterpretation is very common, and as the charting procedures are designed to be used and "owned" by the operators of the process and not by expert statisticians, this problem deserves to be taken very seriously.

To try and avoid these problems, we suggest that statistical control limits should be clearly labelled with the (obviously approximate) percentage of points which are likely to lie within each band if the process is behaving according to hypothesis. Figure 2 below shows how this could be done. The term "control" itself is a misleading term which may be responsible for some of the misconceptions[7, 15]. We suggest using the words *expected zone* and *unexpected zone* instead. This seems more natural, and less potentially confusing, that the conventional format. The statistical limits could be called *surprise limits* to clarify their meaning. (One possible confusion here is that the natural acronym, *usl*, may be confused with *upper specification limit*: this is the reason for referring to the specification limits as *tolerance limits* in Figure 1.)

FIGURE 2 HERE

It is important to be clear about the basis of the expectations on which these zones are based. We will say that the expected zone is based on the assumption of *ordinary conditions*: by definition, special causes only occur in extraordinary conditions, so points in the unexpected zone indicate the possibility that conditions are not ordinary and special causes may be acting. It is obviously important that users should be in a position to see what "ordinary" incorporates in practice - although this may involve more statistical insight than many users can achieve, and is a source of confusion for beginners[7] and disagreements among experts[16]. *For the models used in the conventional charts ordinary conditions comprise the situation when there are no factors differing between samples which do not differ to a similar extent within each sample.*

In the rest of this paper we will avoid the word control - calling the charts *monitoring charts* and the "control" limits statistical or *surprise* limits.

Calculation of surprise limits

Plotting a monitoring chart (as a line graph) is easy. Calculating the statistical surprise limits is not. Most charts use a mathematical model of the distribution of the data used to estimate the variability of the statistic plotted and so the statistical limits. For example, the binomial distribution is used for the *p* chart, the Poisson distribution for the *c* chart, and the central limit theorem for the mean chart. There are a number of difficulties with this: the situation has to fit one of these standard models, and the user needs to understand the model used. The number and statistical sophistication of these models makes the latter problem particularly serious.

We propose two, alternative, approaches. In both cases the conceptual basis is more transparent and more generally applicable than the standard models. The latter point is important in reducing the amount of detail learners need to master - each of the methods will produce estimates of surprise limits for mean, range and proportion defective charts[7], and other possibilities as well.

The quartiles approach

Our first suggested method of simplification is based on the idea of using the points plotted (means and ranges in the present example) directly to estimate the appropriate statistical limits.

There are two main difficulties with this. First, if there are special causes operating in the samples used to estimate the statistical limits, the limits will reflect the variability produced by these special causes as well as the common (within sample) causes of variation. This means that the chart will be an insensitive method of diagnosing special causes because the control lines are set wider than they should be if based only on "ordinary" variability. (Normally the chart should be set up only when the process is stable in the sense that "surprising" points have been eliminated. However, when the "trial limits" are set up, it may be necessary to estimate the limits from an unstable process. It is also likely that there may be special causes which are not strong enough to drive the monitoring chart beyond the surprise limits.) Second, a very large sample of points (each of which may depend on a sample of data taken at a particular time) would be required to estimate the 0.1 and 99.9 percentiles as the extremes (by definition) only occur occasionally.

For these reasons, the alternative proposed here is to base the estimates of statistical limits on the median and quartiles only. These can be estimated from a smaller sample of points, and furthermore, the fact that outliers do not directly affect the values found means that the influence of special causes will be eliminated - providing that these special causes do not occur in the middle 50% of samples. *Ordinary conditions - and so expected performance - are implicitly defined as the state of the process when the sample statistics lie in the middle half of the distribution.* This has the advantage of being a much simpler definition than the corresponding definition for a conventional chart (above).

The procedure is similar to that used for the capability calculation, except that the values used are the sample means or ranges, not the individual measurements. In the case of the range chart, in particular, the asymmetry of the distribution - the difference between LQD and UQD - points to the importance of keeping these two parameters separate. Appendix 1 indicates that the 99.8% statistical limits are 4.7 QDs above and below the median. The resulting surprise limits for mean and range charts based on the data in Table 1 are shown in Table 2.

TABLE 2 HERE

This is very simple and transparent procedure and requires very little arithmetic. The principle is simply that of extrapolating the pattern found in the central part of the distribution. On the other hand, the estimates produced are very rough and it does assume a "normal shape" in each half of the distribution - an assumption which is theoretically justified in the case of the mean chart, but only on the very superficial grounds of the rough pattern of the distribution in the case of the range chart. A more detailed analysis of the accuracy of this method will be given below in the section on *Some questions and objections answered*.

The resampling approach

Resampling is an approach to estimating sampling error by using a computer to draw random "resamples" from a sample of data[8-10]. In an earlier paper[7] one of the present authors suggested how this approach could be used for calculating statistical limits for "control" charts. The advantages are that the rationale behind the approach is transparent; it does not depend on mathematical probability theory; it is not necessary to make questionable assumptions about distributions and so the results are often more mathematically rigorous than the conventional methods; and a single method will cover a chart based on any statistic calculated from a random sample of data (including mean, range and proportion defective). There is not sufficient space in the present paper to discuss this further; for further details the reader is referred to the earlier paper[7]. We also hope to make it the subject of a future paper.

Some queries and objections answered

The standard deviation, and its associated concept the variance, is an important concept for further statistical theory. Therefore novices should learn about it and use it.

In the short term this objection is valid if we consider that the standard deviation is part of the language of statistics with which everyone needs to be familiar. However, in the longer term, this argument does not hold because, for a normal distribution, the standard deviation is equal to about 1.48 QD, and this relationship could easily be used to rewrite any equations involving the standard deviation or variance. The resulting equations may lack elegance, but this is unlikely to be of concern to non-mathematicians. The relevant parts of statistical theory could easily be translated into a more user friendly framework. The standard deviation is also the basis of a widely used slogan, "six sigma"[17], but this slogan would - again - be more comprehensible if it referred to the variability itself, rather than an unintuitive measure of variability.

As the QD depends on only two values, it is likely to provide a less reliable estimate of variability than is the standard deviation which depends on all values in a sample.

In some circumstances (if the sample is small or the variable is discrete-valued) the sample QD may be a relatively unreliable means of assessing variation, but on the other hand it is more resistant[11] (to the influence of outliers) than the standard deviation. It would, in principle, be possible to devise a more reliable estimator of the population quartiles than the sample quartiles - but this would have the disadvantage of lacking transparency, although it would still be preferable to the standard deviation in that it would refer to an intuitively accessible quantity.

The unreliability of estimates of the QDs, particularly when based on a small number of samples, means that the quartiles method of estimating surprise limits will only provide very rough answers. Further, treating each half of the distribution as a separate half normal distribution is at best a very rough approximation.

It is true that this method will not produce reliable estimates from a short sequence of data points, or from data which take only a few discrete values (such as some *p* chart and *c* chart data), since the median and quartiles are likely to be unreliable or insensitive estimators of population parameters under these circumstances. However, rough approximations properly understood are likely to be better than misused or misinterpreted sophistication.

The data in Table 1, analysed in Table 2, was simulated by a spreadsheet (Microsoft Excel) from a normal distribution with a mean of 1677 and a standard deviation of 53. These are, in effect, the process mean and standard deviation which can be used to calculate the "true" limits: ie the limits which would result from an infinite amount of data. This can be achieved by using the constants based on the population standard deviation (σ)[13]. Table 3 gives the errors in the limits produced by the quartiles method (by comparing these with the "true" limits). This table indicates that, for example, the upper limit for the mean chart produced by the quartiles method (1736) is 6% of the width of the "true" interval (134) less than the "true" limit (1744).

TABLE 3 HERE

Table 4 shows the distribution of these errors for 1000 simulated sets of data similar to Table 1, and also the corresponding figures for the standard method. As would be expected, the median errors for the mean chart limits are close to zero, but the difference between the 2.5 and 97.5 percentiles (ie a the central 95% probability interval) indicates that the quartiles method is not very reliable - even when the data consists of 100 samples. With smaller numbers of samples, the errors in the quartiles method are likely to be greater. (A similar simulation based on 20 samples shows that - to take one example - the 95% interval for the lower limit for the mean chart is roughly twice the width: -37% to +28%.) The range chart shows a similar pattern except that the medians are *not* zero: the quartiles method has a systematic bias when used to estimate statistical limits for range charts.

TABLE 4 HERE

These simulation results suggest that the quartiles method for estimating statistical limits for mean and range charts should only be viewed as an approximate method. The same is obviously likely to be true for other charts - *p* charts, *c* charts and so on. In practice this difficulty may be less important than it may seem in principle. If, as is often the case, points are plotted on the chart every few hours, the sequence of sample results available may be very long - hundreds rather than the conventional twenty of textbooks. The probability level of 99.8% is just an arbitrary (but useful) convention - which raises the question of the logic of worrying unduly about the precision with which this arbitrary convention is applied. Furthermore, the inaccuracies of the quartiles method may be negligible compared with errors due to the misuse of conventional methods - eg the (unnoticed) use of an incorrect formula in one of the case studies described in [5] produced errors of several hundred percent.

Clearly, when a greater degree of accuracy is wanted, or when there are only a few samples of data available, the quartiles method is not appropriate. The alternative we would suggest is the resampling approach, which arguably makes more rigorous and efficient use of the data than do the standard methods[7].

The surprise limits derived by the two methods may be very different for unstable processes. Why is this?

As we noted above, the data on which Figure 1 is based indicates an unstable process. Clearly this data is not a suitable basis for setting up a monitoring chart, but it may still be necessary to derive statistical limits from such data to check if a process is stable. As this data is not simulated, we have no way of knowing (or defining) the "true" limits; all that is possible is to compare the quartiles method with the standard one.

TABLE 5 HERE

Table 5 shows that there are very substantial differences between the two estimates of surprise limits for the mean chart in particular. One explanation is the potential unreliability of the quartiles method with a small number (34) of samples. However, the size of the discrepancies suggests that there may be a more fundamental cause. The assumptions about ordinary conditions are summarised in the sections (above) on each of the methods. As these are different, differences in the answers should be anticipated. It seems likely that there are substantial sources of variation between the samples in the middle half of the distribution of samples: this variation is reflected in the surprise limits derived from the quartiles method but not from the standard method. *Ordinary conditions*, as defined by the quartiles method, include circumstances which are not treated as ordinary by the conventional method. The conventional method - in this situation - is thus more sensitive than the quartiles method. It would be difficult to decide which approach is preferable without a much more detailed analysis of the printing process; in general there are arguments for different definitions of "ordinary conditions"[16]. However, there is a very strong argument for a transparent procedure - like the quartiles method - so that the exact sense in which the surprise limits should cause surprise is as clear as possible.

Conclusions

The terminology, concepts and methods suggested in this paper are designed for conceptual simplicity. They are also computationally simple - with the exception of the resampling approach. It should be obvious to users how and why the methods work, so the frequency and seriousness of mistakes and misunderstandings should be reduced. In addition, concepts and methods which are designed to be transparent should enable users to adapt methods to their own purposes, and to concentrate on designing methods of analysis of their own for their own purposes. In short, the spirit or philosophy of OR is more likely to flourish if the conceptual and mathematical basis is more user-friendly.

The aim of this paper is not simply to propose an alternative set of recipes for SPC - although we have suggested alternative approaches to *all* the conceptually difficult areas identified in the introduction. We believe the philosophy behind our approach is of far wider applicability than SPC. There are five general themes underlying the approach.

First, concepts, such as the standard deviation, which are appropriate and useful for mathematicians, may be quite inappropriate for novices. There is little reason, beyond the inertia in any cognitive framework, not to redesign some of these concepts.

Second, the underlying concepts and terminology are very important and sometimes inappropriate. We have suggested measuring incapability in terms of parts per million defective instead of capability in terms of C_{pk}, and the concept of "ordinary conditions" as an alternative for the notions of "within samples variation" and "common causes". Similarly the terminology of "surprise" limits and the "expected zone" seems more natural than the very widely misinterpreted notion of "statistical control" (which is not the same as "real" control). A clear and simple description, or image, of technical concepts in terms familiar to users is obviously crucial.

Third, there are large gains to be made by using one general principle or method for a range of different situations. The quartiles and the resampling methods are both of much more general applicability than the conventional mathematical models. Similarly, the table in the Appendix and the quartile deviation are used for both the quartiles method for estimating surprise limits, and for the incapability index.

Fourth, there is the possibility of using informal methods. For example, Figure 1 provides a simple, graphical capability assessment of the process. Flexible informal procedures seem more appropriate to the task of encouraging process owners to diagnose and improve their own process. On the other hand, formal automatic procedures are quicker to perform once the rationale has been grasped.

The fifth point (which space limitations have prevented us from discussing in detail in the present paper) is that computer intensive methods such as resampling, and simulation in general, often provide a powerful but transparent way of avoiding the necessity of understanding and using subtle mathematical models. Iterative methods, perhaps implemented on a spreadsheet, for solving equations or optimising functions are another example of this strategy. These approaches all have the advantage that users with very little mathematical background can literally see what is going on.

Acknowledgment - We would like acknowledge the help of Roger Reeves with an earlier version of this paper. Sadly, Roger died in February, 1995.

References

1 Mar Molinero, C. (1996). Rethinking statistics teaching for operational research. Journal of the Operational Research Society, 47(6), 723-730.

2 Stuart, M. (1995). Changing the teaching of statistics. The Statistician, 44(1), 45-54.

3 Bailey, T., & Weal, S. (1993). Introducing undergraduates to the spirit of OR while imparting substantive skills. J. Opl Res. Soc., 44(9), 897-908.

4 Eilon, S. (1987). OR is not mathematics. Omega, 15, 87-92.

5 Wood, M., & Preece, D. (1992). "Using quality measures: practice, problems and possibilities." International Journal of Quality and Reliability Management, 9(7), 42-53.

6 Preece, D, & Wood, M. (1995). "Quality measurements: who is using the sums and for what purpose?" Human Resource Management Journal, 5(3), 41-55.

7 Wood, M. (1995). "Three suggestions for improving control charting procedures." International Journal of Quality and Reliability Management.

8 Gunter, B. (1992, December). "Bootstrapping: how to make something from almost nothing and get statistically valid answers - Part 1: Brave New World." Quality Progress, 97-103.

9 Kennedy, K., & Schumacher, P. (1993). "The introduction of bootstrapping techniques for finding confidence intervals using simulation." Journal of Computers in Mathematics and Science Teaching, 12(1), 83-92.

10 Noreen, Eric W. (1989). Computer-intensive methods for testing hypotheses: an introduction. Chichester: Wiley.

11 Erickson, B. H., & Nosanchuk, T. A. (1979). Understanding Data. Milton Keynes: Open University Press, 43-50.

12 Oakland, J. S. (1986). Statistical process control. London: Heinemann, 70-103.

13 Oakland, J. S. (1986). Statistical process control. London: Heinemann, 232-233.

[14] Gilchrist, W. (1993). "Modelling capability." J. Opl Res. Soc., 44(9), 909-923.

[15] Wood, M. (1994). "Statistical methods for monitoring service processes." International Journal of Service Industry Management, 5(4), 53-68.

[16] Correspondence on Limits for Shewhart's control charts by Wheeler, D. and others RSS News, April, June and October, 1996.

[17] Wiggans, T., & Turner, G. (1991, June). Breaking down the walls. Total Quality Management, 183-186.

Appendix: Normal table based on quartile deviations

Table 1: Simulated normal data

Table 2: Median and quartiles of the sample means and ranges, and surprise limits for simulated data in Table 1

Table 3: Comparison of "true" limits and limits estimated by the quartiles method for data in Table 1

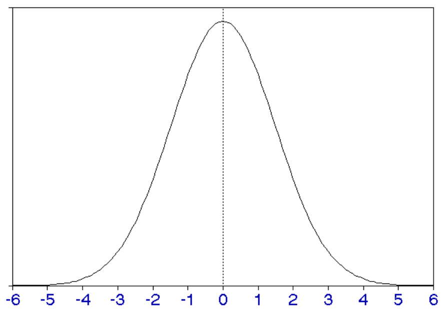
Table 4: Errors* in statistical limits calculated by standard and quartiles method for 1000 simulated data sets like Table 1

Table 5: Comparison of different methods of estimating statistical limits for the unstable process on which Figure 1 is based

Figure 1: Histogram of printing process data

Figure 2: Range monitoring chart

Appendix: Normal table based on quartile deviations



The table gives the proportion in each tail of the distribution. For example, it indicates that 36.79% are more than 0.5 quartile deviations (QDs) above the median, and a similar number are less than 0.5 QDs below the median.

QDs from median	Proportion in tail	QDs from median	Proportion in tail	QDs from median	Proportion in tail
0.0	50.00%	2.2	6.90%	4.4	0.17%
0.1	47.33%	2.3	6.04%	4.5	0.14%
0.2	44.66%	2.4	5.27%	4.6	0.12%
0.3	42.00%	2.5	4.58%	4.7	0.10%
0.4	39.37%	2.6	3.96%	4.8	0.080%
0.5	36.79%	2.7	3.41%	4.9	0.066%
0.6	34.27%	2.8	2.93%	5.0	0.054%
0.7	31.83%	2.9	2.51%	5.1	0.045%
0.8	29.46%	3.0	2.14%	5.2	0.037%
0.9	27.19%	3.1	1.82%	5.3	0.030%
1.0	25.00%	3.2	1.54%	5.4	0.025%
1.1	22.92%	3.3	1.30%	5.5	0.021%
1.2	20.93%	3.4	1.10%	5.6	0.017%
1.3	19.06%	3.5	0.92%	5.7	0.014%
1.4	17.28%	3.6	0.77%	5.8	0.012%
1.5	15.62%	3.7	0.64%	5.9	0.010%
1.6	14.06%	3.8	0.54%	6.0	0.008%
1.7	12.61%	3.9	0.45%	6.1	0.007%
1.8	11.27%	4.0	0.37%	6.2	0.005%
1.9	10.03%	4.1	0.31%	6.3	0.004%
2.0	8.89%	4.2	0.25%	6.4	0.004%
2.1	7.84%	4.3	0.21%	6.5	0.003%

Table 1: Simulated normal data

Sample							Mean	Range
1	1623	1549	1695	1708	1605	1722	1650.3	173
2	1744	1638	1621	1828	1631	1676	1689.7	207
3	1673	1729	1693	1658	1719	1661	1688.8	71
...	...	...	...	...	...	...	...	...
100	1653	1626	1710	1685	1645	1662	1663.5	84

(This table shows only 4 of the 100 samples.)

Table 2: Median and quartiles of the sample means and ranges, and surprise limits for simulated data in Table 1

	Mean	Range
Lower quartile	1666.7	112
Median	1681.3	141
Upper quartile	1692.9	180
LQD	14.6	29
UQD	11.6	39

Limits of 99.8% zone (*median – 4.7 LQD* to *median + 4.7 UQD*)

	Mean	Range
Lower	1613	5
Upper	1736	324

Table 3: Comparison of "true" limits and limits estimated by the quartiles method for data in Table 1

	Mean Chart		Range Chart	
	Percentile		Percentile	
	0.1	99.9	0.1	99.9
"True" limits	1610	1744	29	297
Quartiles method	1613	1736	5	324
Error*	+2%	-6%	-9%	+10%

* The error is the discrepancy from the "true" limits expressed as a percentage of the width of the "true" interval.

Table 4: Errors* in statistical limits calculated by quartiles and standard methods for 1000 simulated data sets like Table 1

	Mean Chart		Range Chart	
	percentile		percentile	
	0.1	99.9	0.1	99.9
Quartiles method				
2.5 percentile	-19%	-16%	-28%	-24%
median	0%	-1%	-11%	-7%
97.5 percentile	15%	18%	2%	13%
Standard method**				
2.5 percentile	-5%	-5%	-1%	-7%
median	-1%	0%	0%	0%
97.5 percentile	4%	4%	1%	7%

* Errors are the discrepancies from the "true" limits expressed as a percentage of the width of the "true" interval (as in Table 3).
** Standard refers to the methods using mean sample mean and mean sample range.

Table 5: Comparison of different methods of estimating statistical limits for the unstable process on which Figure 1 is based

	Mean Chart		Range Chart	
	percentile		percentile	
	0.1	99.9	0.1	99.9
Standard method*	1713	1796	17	183
Quartiles method	1651	1909	13	225
Difference**	-75%	+136%	-2%	+25%

* Standard refers to the methods using mean sample mean and mean sample range.
** The difference is expressed as a percentage of the interval calculated by the standard method.

Figure 1: Histogram of printing process data

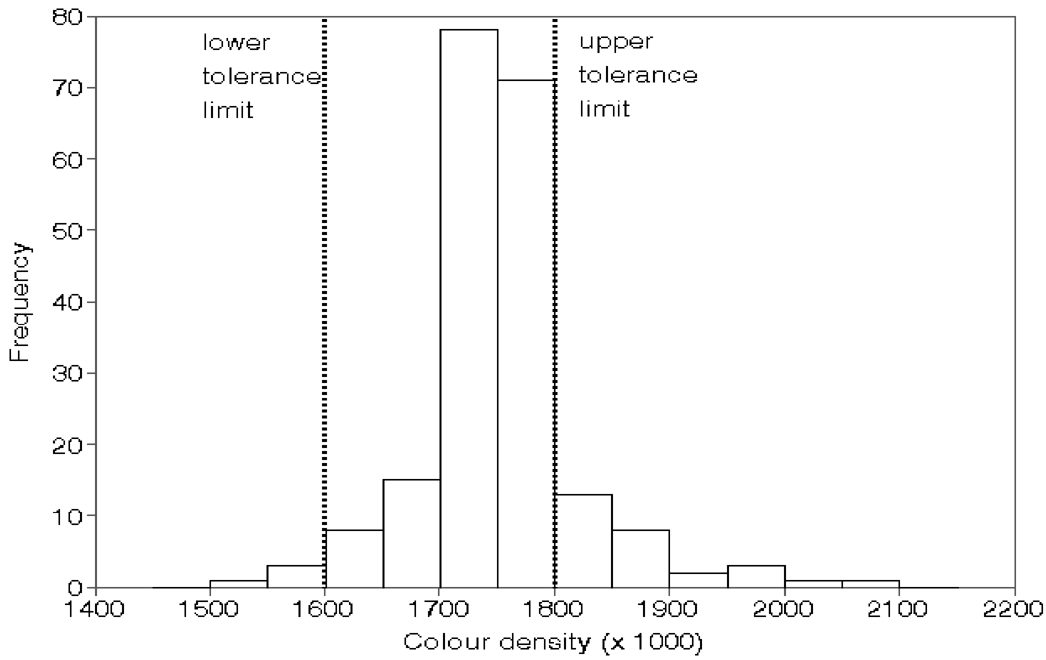


Figure 2: Range monitoring chart