THE USE OF RESAMPLING FOR ESTIMATING CONTROL CHART LIMITS

Michael Wood, Michael Kaye and Nick Capon
Management and Decision Support Unit
Department of Accounting and Management Science
University of Portsmouth
Email: michael.wood@port.ac.uk
14 January 1999

This paper proposes a resampling procedure for estimating the control limits of any statistical control chart which is based on a statistic calculated from a random sample of data. This includes charts for the mean, range, standard deviation, median, proportion defective, number defective and so on. This procedure has advantages over conventional methods in its conceptual simplicity and transparency, flexibility, generality and robustness. The paper describes the operation of a simple program (available from the internet) for carrying out the resampling procedure.

Keywords: *Quality, Statistics, Statistical process control, Simulation, OR education.*

THE USE OF RESAMPLING FOR ESTIMATING CONTROL CHART LIMITS

This paper proposes a resampling procedure for estimating the control limits of any statistical control chart which is based on a statistic calculated from a random sample of data. This includes charts for the mean, range, standard deviation, median, proportion defective, number defective and so on. This procedure has advantages over conventional methods in its conceptual simplicity and transparency, flexibility, generality and robustness. The paper describes the operation of a simple program (available from the internet) for carrying out the resampling procedure.

Introduction

Statistical process control (Shewhart) charts are an important tool for quality management. The control limits on these charts depend, however, on a sophisticated mathematical and conceptual basis: users without an adequate grasp of this basis are likely to have difficulties using control charts to their full effect[1,2,3]. Even among experts, there are disagreements over which methods are correct or useful, and which are "silly"[4]. In two earlier papers[5,6] we argued that the mathematical and statistical basis of conventional statistical process control techniques is too sophisticated for the intended users, and that there is a very strong case for devising a more user friendly framework. Our previous papers made some specific suggestions for such a framework. One of these suggestions was the use of a resampling (bootstrap) procedure instead of the conventional algorithms for estimating statistical control limits.

Resampling (or bootstrap) procedures are fairly widely discussed in the literature but relatively little used in practice. In relation to control charts, Bajgier[7] has suggested a bootstrap procedure for estimating control limits for X-bar (mean) charts because of its greater accuracy for modelling non-normal distributions, and Seppala et al[8] have advocated a more complex bootstrap procedure for mean and variance charts for non-stationary or strongly skewed processes. In both papers, the motivation for the bootstrap method is to improve the accuracy with which probability levels are estimated, not to provide a more user-friendly approach, and the analysis is limited to mean and range charts.

The purpose of the present paper is to put forward a *general* resampling procedure for estimating control limits for *any* statistic calculated from a random sample of data. This

includes the mean, range, standard deviation, median, proportion defective, number defective and so on. The fact that a single method can replace several different formulae (for mean charts, range charts, *p* charts and so on) has obvious advantages for training and ease of use.

In addition to its greater accuracy in modelling non-normally distributed statistics, and its generality, resampling also has the potential for far greater levels of transparency and user-friendliness than are possible with conventional techniques[9,10]. The conventional methods for calculating limits for Shewhart charts for the mean, range, proportion defective and so on, are reasonably easy to perform. Their conceptual basis is, however, relatively complex, which means that *it is not easy for non-specialists to understand why the methods work and how to interpret the statistical limits*. This is important because this lack of understanding (sometimes not recognised as such by the people involved) can lead to errors and an inability to adapt approaches to new situations (see Reference[2] for some examples of this). Comments from many delegates at conferences at which we have presented papers on this theme confirm these observations.

This paper argues that resampling can provide a solution to this problem. The procedure is transparent in that users can see what is happening and the rationale behind it is fairly obvious (Figure 1 below) - so misuse is much less likely. The resampling procedure makes no reference to the standard deviation, the normal distribution, the central limit theorem, the binomial distribution formulae, or to tables whose origin is likely to be - to the user - a mystery. (These are all areas which novices are likely to find difficult[6].) All that is assumed is an understanding of random sampling and percentiles. (Control limits are *defined* in terms of percentiles, and random or chance variation, so they cannot be properly interpreted - however they are derived - without an appreciation of these concepts.) This has clear and important implications for education and training: users of the resampling procedure need much less in the way of technical background than do users of conventional methods.

Resampling does, however, require computational assistance, and the software clearly has to be easy to use if it is not to negate the potential advantages of the transparency of the method. We describe the operation of a simple menu-driven program (available on the web from http://www.pbs.port.ac.uk/~woodm/resample.htm) which leads users through the resampling procedure. Trials with this program suggest it is easy to use (with appropriate instruction) despite the fact that it could clearly be improved substantially.

For all these reasons, we believe that resampling is worthy of consideration for estimating control chart limits. (In an earlier paper[6] we made some suggestions for more user-

friendly approaches to other aspects of SPC such as process capability analysis - these do not involve the resampling principle.)

We start by giving a formal description of the proposed procedure and its interpretation. This is then illustrated by an example which also shows how the software works. We then discuss the generality and flexibility of the procedure in terms of the statistics which can be modelled (mean, range, median, proportion defective, etc) and the possibilities for obtaining graphical output and experimenting with different probability levels and sample sizes, and the issues raised by "out of control" processes.

The resampling procedure: formal description

The construction of a Shewhart statistical control chart typically starts with j samples of n observations of a process:

$x_{11}, \ldots x_{1n}$

...

$x_{i1}, \ldots x_{in}$

...

$x_{j1}, \ldots x_{jn}$

A chart for a statistic s for this process comprises a line graph of $s_i(x_{i1}, \ldots x_{in})$ and two *control limits* designed to encompass 99.8% of fluctuations due to *chance* or *common causes* of variation.

In a standard statistical control chart these limits are estimated using probability theory; there are a number of different models depending on the nature of the observations (usually a numerical measurement, a count of defects, or a categorisation as "defective" or "not defective") and the statistic plotted, s. Examples include *mean* charts, *range* charts, *p* charts, *c* charts, *u* charts, *median* charts, and so on[11]: each of these involves a different mathematical model and a different set of formulae.

If s is normally distributed, the natural limits to use to achieve this probability level are 3.09 standard deviations above and below the mean: conventional practice in some contexts (particularly in the USA[11]) is to round this off to 3 standard deviations and to *assume* that the distribution of s is normal - even if s is the proportion defective or the range, in which case the assumption is not necessarily accurate. For the resampling procedure (which avoids any reference to the standard deviation of s) we avoid these approximations and return to first principles by assuming that the control limits are the 0.1 and 99.9 percentiles of the appropriate distribution - leaving 99.8% between the two limits.

If some of the j points plotted fall outside the control limits - indicating a *special cause* - the reasons for this are sought, and, if it is considered appropriate, the points in question may be excluded from the data set and the limits recalculated to obtain a more realistic assessment of the state of the process under *ordinary conditions*[6]. This may require judgment by those familiar with the process.

The purpose of the proposed resampling procedure is to estimate the control limits for the chart. The principle is to use the power of the computer to simulate a large number of samples as a way of estimating sampling error, instead of the conventional reliance on probability theory.

The basic resampling procedure can be used with any statistic, s, which can be regarded as a function of a random sample of data, $x_{i1}, ...x_{in}$. The procedure is:

1. Decide which samples of observations are to be used to estimate the control limits. These may be all of the samples, or some samples may have been removed on the grounds that they are not typical of the process - as described in the section on unstable processes below. (Sometimes this decision may depend on an earlier, trial run of the resampling procedure.) Suppose there are k samples remaining after any exclusions.

2. Combine all observations from the k samples as a single *combined sample*. If the samples are all of size n there will be kn observations in the combined sample.

3. Use an appropriate computer simulation package to take *r random samples of size m* (with replacement) from the combined sample, and calculate the value of s for each of these "resamples". The "resamples" are so called because they are samples from a sample. Normally m and n will be equal but this is not necessary - data from samples of one size can be used to set up control limits for samples of another size (although obviously the values of s based on the original data samples of size n should not be plotted against the control limits for samples of size m). Each observation in each resample is chosen at random from the whole of the combined sample, so there is a possibility that the same observation may occur more than once in the same resample. The value of r must be sufficient for the simulation to provide reliable results bearing in mind that we are interested in simulating the frequency of fairly rare (0.1%) events: something between 10,000 and 100,000 is suggested - the limiting factor being the speed of the computer and software.

4. Plot the frequency distribution of the r values of s. Strictly this stage is not necessary - and it has no equivalent in the standard procedures (although it is possible to plot a

histogram of the actual sample values, which is not quite the same) - but, as we will see, it is useful.

5 Estimate the 0.1 and 99.9 percentiles of this distribution. These correspond to the limits for the control chart. Points should only be considered to indicate a special cause if they are greater than the upper limit or less than the lower limit: points on the line should not be regarded as indicating a special cause.

The justification for this procedure rests on the assumptions that the combined sample provides an adequate picture of the distribution of observations from the process, and that the process is *stable*. (If the process is unstable it cannot realistically be represented by the combined sample: we will consider this below.) Given these assumptions, random samples from this combined sample with replacement (since an observation in the combined sample represents a possibility which may occur again) will provide a picture of the distribution of the statistic s - provided that there are no special causes of differences between samples. The procedure is in effect a direct simulation to show the sampling error resulting from a stable process. (Note that it is not necessary to invoke the concept of population parameters: all we need is a statistic defined on a finite sample.)

The stipulation that points *on* the control limits should *not* be treated as indicating special causes (step 5 above) follows from the requirement that the probability of the stable process yielding a value of s in the upper "out of control" zone should be no more than 0.1%. The difficulty is that if, for example (Figure 3 below), 8% of the resampled values of s take the maximum value of 5, the 0.1 percentile is 5, but the probability of s being equal to 5 is about 8%. This means that the "out of control" zone needs to be defined as *greater than 5*. A similar argument applies to the lower limit.

Bootstrap methods have been criticised for their unreliability[12]. The resampling method proposed here has two features which suggest it may be more straightforward than other bootstrap methods. First, the combined sample from which the resamples are taken is typically fairly large (84 and 198 in the two examples used in this article); for large samples bootstrapping is likely to be reliable[12]. Second, when bootstrapping is used for estimating confidence intervals or significance levels, the sample is used as a surrogate for the population in order to see the likely discrepancy between the sample and the population - the dual role of the sample is obviously a potential problem. In the present proposal the combined sample is used as the surrogate for the population, which is then compared with statistics calculated from individual samples. For these reasons, the resampling procedure proposed here is likely to be more accurate and straightforward than the bootstrap procedures used for

significance tests and confidence intervals.

The interpretation of control limits

In conventional control charts the control limits are defined by a probability model of the "in control" process. In hypothesis testing terms, this is the hypothesis under test[13]. In an earlier paper[6] we have proposed the phrase *ordinary conditions* for this state of the process because it seems an accurate and easily comprehended phrase to describe the underlying concept. If the control limits are calculated by the resampling procedure "ordinary conditions" means that the process is stable in the sense that every sample can be regarded as a random sample from the same pool of potential observations: there are no samples which special causes make different from the norm. As will become clear in the section on unstable processes, the meaning of "out of control" and "ordinary conditions" for such a process depends on the model used. In a few situations the resampled limits are not equivalent in magnitude or meaning to the conventional limits.

Software

As resampling is not (yet) a widely used approach in statistics there are few purpose-built packages available, and none that we found was really suitable. Implementing the procedure on a spreadsheet would be possible but cumbersome. This meant it was necessary to write a program, RESAMPLE, to implement the procedure. This program is menu-driven and can be downloaded from the world wide web (from http://www.pbs.port.ac.uk/~woodm/resample.htm).

Our experience using it with groups of postgraduate students suggests that it is reasonably easy to use; in a recent class we found that 90 minutes was sufficient time to explain the resampling principle from scratch, to allow the students time to experiment using the program to derive control limits for several different types of charts, and to discuss the implications and utility of the procedure. With a more sophisticated program (with better editing facilities and on-line help, for example), this learning process would doubtless be even easier.

An example

We will illustrate the resampling procedure by showing how it can be applied to Process 1 (see Appendix). There are seven samples of twelve observations, all 84 of which are used.

The statistic chosen for the first chart was the mean. The analysis was carried out by the program RESAMPLE: Figure 1 shows an edited version of the output from this program. This has the dual advantages that the reader can see both how the procedure works, and how the program works.

FIGURE 1 HERE

The first screen in Figure 1 explains the procedure to the user. The second shows the first resample, the value of the mean calculated from this resample, and this value plotted on a crude histogram or tally chart. This gives the user the opportunity to see exactly what the program is doing: if necessary more resamples can be viewed in the same way.

The third screen shows the distribution of 10,000 resamples - this is step 4 of the resampling procedure. This enables the user to see the pattern of the distribution, and incidentally demonstrates that many sampling distributions are approximately normal.

The fourth screen gives the 0.1 and 99.9 percentiles of this distribution as well as a few hints on interpretation.

The number of resamples in this example is 10,000, although the control limits estimated from the first 1000 of these resamples were very similar (30.96 and 32.71). In practice users can continue to increase the number of resamples until the pattern has stabilised. As we would expect, the results from the resampling procedure are very similar to those from the conventional procedure: 31.1 and 32.8.

In this example the means of all seven samples of data were within the control limits: no special causes were flagged. If special causes had been flagged, it may have been appropriate to repeat the estimation of the control limits with the relevant sample excluded.

Having estimated the control limits, a chart to monitor the process can be constructed in the usual manner. Figure 2 shows the seven samples used to estimate and control limits, and some later samples. Sample 14 is flagged as worthy of investigation.

FIGURE 2 HERE

As well as the mean, RESAMPLE can also work with the median, sum, standard deviation, range, interquartile range or any specified percentile of the resamples.

Resampling methods tend to be more general than conventional ones both in the range of statistics which they model, and in the range of questions they can answer about these

statistics. The next two section discuss how general the procedure is in these two senses.

Statistics which can be modelled

The resampling procedure can be carried out with any statistic which can be calculated from a random sample of observations. The procedure will model the sampling distribution which would be expected from a stable process. This means the model of "ordinary conditions" reflects the actual process rather than a mathematical model which may not be appropriate. The paragraphs below list some of the possibilities.

Mean, median, range, standard deviation of a numerical variable

For a sample of numerical observations such as those in the Appendix, the statistic, *s*, could also be the *median*, or the *range* (Figure 3) or the *standard deviation*.

FIGURE 3 HERE

The formal procedure above indicates that a value of 5 for the range should not be taken to indicate a special cause; in step 5 of the resampling procedure, only values greater than 5 indicate such causes. The reason for this is obvious from the graph in Figure 3: there is a substantial chance - around 8% - of a range of 5 arising randomly. (The control limits from the conventional procedure are 1.3 and 5.9; assuming integer values, the only difference in practice is that a range of 1 would be flagged as a special cause by the conventional procedure but not by the resampling procedure.)

Proportion or number defective

For observations which comprise the categorisation of an item as "defective" or not, there are two obvious statistics - the *proportion defective (p)* in a sample, and the *number defective (np)*. The resampling procedure can be applied with both statistics. As an example, we will consider ten samples of 50 items from a process: the combined sample of 500 items contains 11 defectives. The convention used here is that 1 indicates a defective item, and 0 an item which is not defective: this means that the sum of a sample, or resample, of 50 observations gives the number defective, and the mean gives the proportion defective. Simulating 10,000 resamples yielded estimated control limits for the number defective of 0 and 5: in accordance with step 5 of the resampling procedure this means that a special cause is flagged by *fewer than* 0 defectives (obviously impossible) or by *more than* 5. This is

effectively a simulation of the binomial distribution, so it will yield probabilities which can be made as close to the exact answers as required simply by taking more resamples. As would be expected these results are consistent with the exact binomial probabilities:

P(0) = 32.88%, P(more than 4) = 0.48%, P(more than 5) = 0.08%

The conventional approach to these charts uses three sigma (standard deviation) limits, which are often inaccurate (the upper limit in the example above is *4.2*, for example). Ryan and Schwertman[14] suggest a more accurate approach, but at the cost of considerable complexity. Resampling provides accurate estimates by a very simple procedure.

Count of defects or defects per unit

The *c* (count) chart is based on a Poisson model: typical observations on which these charts are based are the following numbers of customer complaints (defects) per week received by an organisation in successive weeks:

7, 10, 5, 9, 14, 6, 6, 12, 8, 5, 9, 4

The resampling procedure cannot be used directly here because the statistics are not based on a random sample of observations but are a count of defects in the entire process each week. It is possible to transform the data to *number defective* form by splitting the week into, say, 72 working hours and regarding an hour as defective if a complaint is received during the hour - so the first datum would be replaced by 7 1's and 65 0's. One might object that more than one complaint could be received in an hour, in which case the week could be split into minutes or even seconds. Defects per unit could be treated in a very similar manner. Once the data are transformed, we can apply the same procedure as described in the previous section.

This procedure is, in effect, modelling the Poisson distribution from first principles. The advantage of doing this is that users are forced to appreciate how and why the model works; the disadvantage is obviously the extra complexity of the method.

The above statistics include most of the standard models in texts on statistical control charts. The only other commonly used ones are the single value chart, and moving average and moving range charts[11] - which cannot usefully be simulated by the resampling procedure since the statistics in question are not based on a random sample of data. (We have suggested another "user-friendly" approach appropriate to cases like these in another paper[6].) There are also further possibilities - e.g. the interquartile range or geometric mean - which could be simulated by the resampling procedure.

Questions which can be answered

In the section above we looked at the different statistics that can be modelled by the resampling procedure. The procedure is also flexible in terms of the questions which can be asked about these statistics.

First, and most obviously, different probability levels could be used. The conventional levels are generally regarded as appropriate, but there may be contexts in which it is sensible to disregard the conventions.

Second, the entire sampling distribution of the statistic is obtained, not simply a few probability levels. This means that users can see, for example, that a mean sample hole size of 30.0 is well outside the pattern in Figure 1 without recourse to formal control limits. This has the potential advantage that the use of control limits could operate on a more intuitive level.

Finally, it is easy to experiment with different sample sizes. The resampling program enables the user to run off control limits for samples of any size. Ten thousand resamples of size 3, for example, give control limits of 30.0 and 33.3; the user may decide that these control limits indicate that samples of three are sufficient to monitor the mean of this process. (Obviously, statistics based on samples of 12 should not be plotted on the same chart as control limits for samples of 3.) A similar experiment could be carried out with the limits for the range chart.

Conventional procedures could be extended in all three of these ways, but at the cost of considerable extra complexity: different probabilities would require extended tables, graphing the sampling distributions would require software incorporating the relevant mathematical models, and even experimenting with different sample sizes is not easy with some of the formulae that are traditionally used for mean and range charts (those which use the mean range whose value depends strongly on the sample size). The advantage of resampling is that it provides an entirely obvious approach to these issues, which works regardless of the statistic being monitored.

Unstable "out of control" processes

If the original observations on which the control limits are based come from a process which is not stable - some samples differ substantially from the norm and there may be "out of control" points on the graph - this causes difficulties for estimating control limits. The conventional advice is that these samples should be eliminated so that the remaining samples provide an accurate picture of "ordinary conditions". (In practice this may not solve the

problem because a chart may be drawn before the diagnosis of special causes is made, and also there may be some special causes which do disturb the process but not sufficiently to trigger the "out of control" signal.)

In all standard charts the centre line is based on the mean of the statistics calculated from all the original samples: this, of course implies that any unusual samples will have some influence on the baseline of the chart. The variability of the statistic - as reflected in the width of the control lines - is calculated from probability theory in the case of range charts (the usual assumption being that individual data points are normally distributed) and p, c and u charts, and empirically from the *within* sample variability (using the central limit theorem) in the case of the mean chart.

The resampling procedure for the statistics p, np, c and u is, in effect, a computer simulation of the model described by the appropriate probability distribution. This means there should be no difference between the results of the resampling procedure (using a suitably large number of resamples) and the use of the appropriate probability distribution, even if the process is out of control. (In the case of the p chart, for example, the input to the resampling procedure is the number of defectives and the number of non-defectives; for the probability model the overall value of p is calculated from just the same data.) However, as noted above, the conventional approximations used in these charts are fairly rough[14]: as the resampling procedure does not employ any approximations, it will provide *more* accurate estimates.

When the resampling procedure is applied to the mean, the centre line is based on the mean or median of the resample means, which will be similar to the overall mean of the original samples. This is also identical to the conventional procedures. For measures of variability, on the other hand, resampling will give an answer based on the overall variability in the *combined* sample rather than the *within* sample variability of conventional procedures. If the process is seriously "out of control" *from the perspective of the mean*, the conventional procedures for mean and range charts in effect eliminate this source of variability by considering the variation within each sample. The resampling procedure will not do this: ordinary conditions are defined in terms of random samples from the same pool of potential observations; resamples taken from this pool will reflect overall variability. This means that estimates of variability from resampling are likely to be greater if the process is out of control. This, in turn, implies that the mean chart control limits will be more widely spaced, and both control limits on the range chart will be higher.

The resampling procedure, then, is only sensible for stable processes. If there are

substantial instabilities these must be filtered out. There are a number of ways in which this could be done. Knowledge of the process may suggest that some samples are likely to be unreliable indicators of ordinary conditions. A cusum chart[11] could be used to detect long term changes so that the resampling procedure can be applied to relatively homogeneous subgroups. Seppala et al[8] suggest a bootstrap procedure which avoids the difficulty by resampling the deviations from the sample means instead of the raw data. On the other hand, it is a fairly complex procedure.

A much cruder approach - which could be applied to *p* and *c* chart data as well - is a *restricted resampling procedure*. This involves ranking the samples of data in order of their mean (or median if this is the statistic used), the top 25% (or thereabouts) and the bottom 25% discarded, and the remaining middle half used for resampling. The rationale is that special causes are likely to be eliminated; *ordinary conditions* now have a simple interpretation in terms of "middle samples". Obviously this will not be effective if there are special causes acting most of the time, and it also means that the size of the combined sample is halved: but these are unlikely to be serious difficulties in practice. (There is, of course, no reason for taking the middle *half* - and not, for example, the middle 60% - except that a half is a "round" number.)

From the point of view of the central tendency of the resampled control limits, the restricted resampling procedure is similar to the idea of a trimmed mean or an interquartile mean[15] - an acknowledged way of reducing the influence of outliers. However, the estimate of variability may be slightly reduced, even if the process is in fact stable, because samples with high or low means - which are likely to have high or low individual values - have been eliminated.

Process 2 in the Appendix is out of control on both mean and range according to standard methods (ie using the UK control chart factors[11]). Table 1 shows the control limits for mean and range charts using these standard methods, the resampling procedure and the restricted resampling procedure. As expected, the limits for the mean from the resampling procedure are wider, and both limits for the range higher, than those from the conventional procedure. The limits from restricted resampling, on the other hand, are very close to the conventional ones for the mean, but noticeably lower for the range. The reason for the latter point is that, as one might expect, the ranges of samples whose means are high or low (and so were eliminated from the resampling procedure) were higher (mean range = 102) than for samples in the central band (mean range = 61). The same factor might be expected to reduce the estimated variability for the mean chart: the reason this is not observed is probably that

this effect is counterbalanced by a substantial within-sample variation even within the central half of the samples.

The standard methods for mean and range charts estimate variability using *all* the samples (even the "out of control" ones for the initial limits before these samples have been removed), whereas the restricted resampling procedure bases the estimate on the central samples only. The latter perhaps seems a more useful and natural interpretation of ordinary conditions, and possibly one less prone to misinterpretation.

TABLE 1 HERE

Table 2 gives the same information for Process 1, which is in control: this shows little difference between the three procedures.

TABLE 2 HERE

Table 3 gives a comparison of the three methods of analysis applied to normal, simulated data - to represent a stable process. This shows that restricted resampling gives similar results to resampling for samples of 20, but for small samples the two procedures do diverge. This suggests that the restricted resampling procedure is likely to underestimate slightly the variability in mean and range charts for small samples. If a process is being monitored by mean and range charts, and these charts are based on small samples - as is often the case, it would clearly be a good idea to re-estimate the control limits using a complete set of data (not restricted to the middle samples) once the process has been stabilised. This would be normal practice anyway.

TABLE 3 HERE

The reader should, of course, bear in mind that statistical limits are only meaningful with stable processes. The transparent interpretation of ordinary conditions (compared with the conventional probability models) means that this should be obvious to users. The suggestions discussed in this section for inferring the stable core of an unstable process are very much a last resort.

Conclusions: comparisons between different ways of estimating control

limits

We have proposed a general resampling procedure for estimating control limits for any statistic calculated from a random sample of observations. The results of this procedure can always be interpreted in terms of the behaviour of random samples under *ordinary conditions*: successive samples are simulated as random resamples from the same population of potential observations. Special causes are then indicated by any sample which falls outside the pattern of the simulation.

This procedure has a number of very clear advantages over the standard methods:

Generality

One procedure, and one interpretation, replaces a multitude of different algorithms, statistical tables and interpretations for particular cases (mean charts, range charts, p charts and so on). This makes the learner's task, and the user's, much simpler.

Conceptual simplicity and transparency

Understanding the rationale behind the conventional algorithms requires a fairly advanced knowledge of probability theory. The *p* and *c* charts, for example, require an understanding of binomial and Poisson processes, and the rationale behind the conventional method for the mean chart is notorious for producing disagreements among experts and confusions among beginners[4]. On the other hand, the resampling procedure requires only an understanding of the statistic used (mean, etc), histograms, percentiles and random samples, both to see what to do and how to interpret the answers in terms of the null hypothesis of ordinary conditions. Resampling has the advantage that users can literally see it working, resample by resample, whereas conventional probability theory is rarely fully understood and often misapplied.

This means that resampling is likely to have the twin advantages of reducing training requirements, while improving novices' ability to see (literally) what is going on.

Less distortion by unrealistic assumptions

As well as being easier to understand, resampling methods are likely to be less distorted by unrealistic assumptions. There are many such assumptions in conventional methods. Range and small sample mean charts depend on the assumption that individual measurements are normally distributed, and *p* and *c* charts use often inaccurate "three sigma" limits. Resampling uses the combined sample as an empirical distribution and so makes no such assumptions.

Greater flexibility

Experiments to see the effects of different sample sizes and different probability levels are easy to perform. It is also straightforward to produce charts for statistics not usually employed - eg median or inter-quartile range.

Intuitively accessible graphical results

The graphical display of the results of resampling opens up the possibility of dispensing with formal control limits, and simply comparing successive sample results directly with the resample distribution.

There are two possible disadvantages with resampling. First, it does require a computer: it is convenient to be able to do calculations in one's head or on the back of an envelope. But the computer power required is not great - an ordinary PC is more than adequate - and software is available (and can easily be written) which is relatively easy to use. Second, estimates of control limits for some statistics (including the mean and range, but not p and np) derived from unstable processes may be slightly less useful than those derived by the usual methods (although there are counter-arguments). We suggested several ways of resolving this difficulty, and also pointed out that the greater transparency of resampling means that users are more likely to appreciate the issues and respond in an appropriate manner than are users of conventional methods.

Appendix: Processes and data

Process 1: Flow rates through aerosol nozzles
```
Sample 1: 32 32 32 33 32 33 33 33 34 32 33 32
Sample 2: 32 32 31 31 32 32 29 32 32 32 33 31
Sample 3: 32 29 32 32 32 31 32 31 33 34 33 32
Sample 4: 32 31 31 32 32 31 31 31 33 32 32 31
Sample 5: 32 32 32 32 33 30 32 33 32 31 32 32
Sample 6: 32 31 31 29 31 32 33 33 32 32 33 32
Sample 7: 32 33 31 32 32 32 33 33 33 32 31 32
```

```
Process 2: Densitometer readings (x 1000) for a colour
printing process (5 of 34 samples shown)
Sample 1   1731 1696 1699 1790 1739 1798
Sample 2   1726 1753 1741 1736 1748 1740
Sample 3   1770 1762 1761 1716 1743 1720
Sample 4   1747 1747 1737 1737 1779 1716
Sample 5   1747 1763 1737 1722 1743 1751
........   ............................
```

References

1. Dale, B. G., & Shaw, P. (1991). Statistical process control: an examination of some common queries. International Journal of Production Economics, 22(1(?)), 33-41.

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

3. Hoerl, R. W., & Palm, A. C. (1992). Discussion: integrating SPC and APC. Technometrics, 34(3), 268-272.

4. Wheeler, D. (April, June and October 1996) and others. Correspondence on Limits for Shewhart's control charts. RSS News.

5. Wood, M. (1995). Three suggestions for improving control charting procedures. International Journal of Quality and Reliability Management, 12(5), 61-74.

6. Wood, M., Capon, N., & Kaye, M. (1998). User-friendly statistical concepts for process monitoring. Journal of the Operational Research Society, 49, 976-985.

7. Bajgier, S. M. (1992). The use of bootstrapping to construct limits on control charts. Proceedings of the Decision Science Institute, 1611-1613.

8. Seppala, T., Moskowitz, H., Plante, R., & Tang, J. (1995, April). Statistical process control via the subgroup bootstrap. Journal of Quality Technology, 27(2), 139-153.

9. Ricketts, C., & Berry, J. (1994). Teaching statistics through resampling. Teaching Statistics, 16(2), 41-44.

10. Simon, J. L. (1992). Resampling: the new statistics. Arlington, VA: Resampling Stats, Inc.

11. Bissell, D. (1994). Statistical methods for SPC and TQM. London: Chapman and Hall.

12.	Noreen, E. W. (1989). Computer intensive methods for testing hypotheses. Chichester: Wiley.

13.	Box, G., & Kramer, T. (1992). Statistical process monitoring and feedback adjustment - a discussion. Technometrics, 34(3), 251- 267, especially 252-253.

14.	Ryan, T. P., & Schwertman, N. C. (1997). Optimal limits for attributes control charts. Journal of Quality Technology, 29(1), 86-98.

15.	Erickson, B. H., & Nosanchuk, T. A. (1979). Understanding Data. Milton Keynes: Open University Press, p. 40.

Table 1: Comparison of different methods of estimating statistical limits for Process 2

	Mean Chart		Range Chart	
Method	Control limits	Special	Control limits	Special
Standard	1713 – 1796	9	17 – 183	3
Resampling	1669 – 1863	5	23 – 492	0
Restricted resampling	1719 – 1787	11	13 – 131	4

Standard refers to the UK methods using mean sample mean and mean sample range (11). Both resampling procedures involved 10,000 resamples; in the restricted case the 9 samples with the highest means and the 9 with the lowest means were eliminated. The figures in the Special columns correspond to the number of the 34 samples which were outside the control limits.

Table 2: Comparison of different methods of estimating statistical limits for Process 1

	Mean Chart	Range Chart
Method	Control limits	Control limits
Standard	31.1 – 32.8	1.3 – 5.9
Resampling	31.0 – 32.7	1 – 5
Restricted resampling	30.9 – 32.7	1 – 5

Standard refers to the UK methods using mean sample mean and mean sample range (11). Both resampling procedures involved 10,000 resamples; in the restricted case the 2 samples with the highest means and the 2 with the lowest means were eliminated.

Table 3: Comparison of resampling and restricted resampling from 100 samples of normal simulated data (n= 20, 6, 2)

	Mean Chart	*Range Chart*
	Control limits	*Control limits*
Samples of 20 values		
Resampling	28.7 – 31.2	3.4 – 11.5
Restricted resampling	28.7 – 31.2	3.3 – 10.9
Samples of 6 values		
Resampling	27.8 – 32.5	1.0 – 10.3
Restricted resampling	28.1 – 32.4	1.0 – 8.7
Samples of 2 values		
Resampling	25.9 – 34.2	0.0 – 8.9
Restricted resampling	27.5 – 32.6	0.0 – 5.5

Both resampling procedures involved 10,000 resamples; in the restricted case the 25 samples with the highest means and the 25 with the lowest means were eliminated.

Figure 1: Output from the resampling program for mean chart for process 1

(The dashed lines separate four separate screen displays. Some intermediate screens – for data input, etc – are not shown. The version of the program available from http://www.pbs.port.ac.uk/~woodm/resample.htm is not customized for control charts and has different text in the first and fourth screens above.)

The program will take random samples from the data you have entered. These are called 'resamples' because the samples are being sampled again. It will calculate the mean of each of these resamples.

The purpose of this is to see how much these means will vary if the resamples are drawn at random. You can then compare the means of actual samples with this pattern. If the mean of a sample is right outside the pattern of the random resamples, it is a reasonable conclusion that there is a "special cause" operating and the process should be checked.

```
----------------------------------------------------------------
Average (mean) of resamples of size  12.
 1st resample:   29  31  32  32  32  29  33  32  32  32  32  32
Average (mean) of 1st resample is  31.50
 30.89 to 31.15
 31.16 to 31.42
 31.43 to 31.69  X
 31.70 to 31.96
 31.97 to 32.23
(X represents 1  resamples.)
----------------------------------------------------------------
Average (mean) of resamples of size 12.
Number of resamples=  10000
 30.35 to 30.61
 30.62 to 30.88  –
 30.89 to 31.15  –
 31.16 to 31.42  XXXX–
 31.43 to 31.69  XXXXXXXXXXXXXXXXXXX–
 31.70 to 31.96  XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX–
 31.97 to 32.23  XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX–
 32.24 to 32.50  XXXXXXXXXXXXXXXX–
 32.51 to 32.77  –
 32.78 to 33.04
(X represents 83  resamples.  – represents fewer than this.)
----------------------------------------------------------------
Control limits are set at the 0.1 and 99.9 percentiles:
     Estimate of lower limit is 31.00
     Estimate of upper limit is 32.67
```

This means, for example, that about 99.8% of random resamples will fall between 31.00 and 32.67. Similarly about 99.8% of actual samples from the process should fall within this range – if the process has not changed. Any points outside this range mean that the process mean has (almost certainly) changed in some way. If this happens you should look for the cause.

Figure 2: Statistical control chart for the mean of process 1
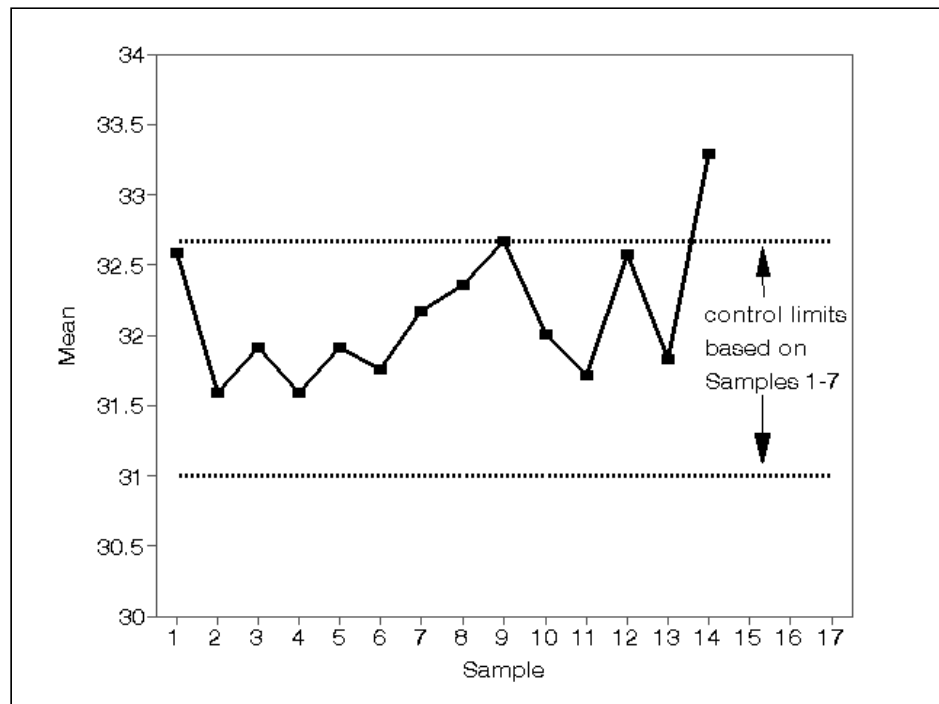
Figure 3: Resampling results for range chart for process 1

```
Range of resamples of size  12.
Number of resamples=  10000
 0
 1      XXX-
 2      XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX-
 3      XXXXXXXXXXXXXXXXXXXXXXX-
 4.     XXXXXXXXXXXXXXXXXXXXXXXXXXXX-
 5      XXXXXXXX-
 6
(X represents 98  resamples.  - represents fewer than this.)

Central 99.8% interval goes from 0.1 to 99.9 percentiles:
 1.00 to 5.00
```