Potentially Capable but not Actual

Neil S Barnett - Melbourne, Australia

1. Introduction

Many service courses in statistics, in my experience, have degenerated into the presentation of a set of recipes. The prevailing situation extant in many courses was summed up recently by my son, himself a trainee engineer, when he said, "I learn the recipes to pass the exam but I don't really understand the material". In fact the rapidity with which the material in his course has been presented would defy understanding by the most statistically astute. Ironically, if past form is anything to go by, he'll do exceptionally well!

With many manufacturing organisations pursuing quality improvement programmes, I've had numerous opportunities over recent years to teach practicing engineers the intricacies of statistical process control. This has often been done in their workplace using material that is directly relevant to them and their work environment. Generally, I have been surprised by how little they understood about the true nature of statistics and the potential benefits that a statistical approach could offer them. I appreciate that practicing engineers who may not have graduated in recent years or who have reached their current positions by working up from the shop-floor, will have had little or no formal statistical training. All recent engineering graduates, however, will almost certainly have endured some compulsory statistics course as part of their training. The disarming fact, however, is that neither by virtue of previous training nor through interaction with those who have been so trained have they come to see the benefits or relevance of statistical techniques. One must, therefore, reasonably ask what has gone wrong? Is the fault with them, with the courses they have pursued, a combination of both, or something else altogether?

With the current surge of interest in statistical techniques engendered by the "quality" push, many design, maintenance and process engineers are being forced to embrace ideas that many have met before but haven't mastered.

2. Statistical process control

I was recently discussing the problems of teaching statistics to undergraduate engineers with a bright young graduate chemical engineer who had a far greater appreciation of statistical methods than (in my experience) most with his background. He lamented the fact that he wasn't taught the techniques of statistical process control during his tertiary studies. I suggest that statistical process control should feature much more prominently in engineering courses than it does currently. It is even feasible to build statistical ideas around the fundamentals of statistical process control since this material embraces the ideas of point and interval estimation, hypothesis testing and a detailed study of normality and its properties. It would be possible to bring in the elements of independence and correlation, the study of non-normal distributions and even introduce some non-parametric statistical tests, as well. Transformation techniques for handling non-normal distributions could feature, with emphasis on the importance of sound data integrated into quality management principles. Of course, these techniques have a much more general relevance than merely for process control, but at least to aspiring process engineers such an approach has much to commend it. There is, in addition, the whole area of reliability, a "smattering" of which also could also be covered in such a course.

"Statistics" is peculiar in that most people using statistics in practice are not statisticians! It seems that in order to be deemed respectable any piece of research or investigative work must be accompanied by a statistical analysis - right or wrong does not always seem to matter too much. Respectability is gained by the statistical analysis, thus many non-statisticians have been drawn into the statistical arena. So it is in the manufacturing industry, many of those overseeing the use of statistical techniques aren't statisticians, many are engineers. Given that there are engineers with the background and perceptions alluded to previously, there is obvious potential for misunderstandings to "creep in".

Introductory plant training in statistical techniques, when given, is often appended to training in quality management and appears under the heading of "Tools of T.Q.M.". Often this material is presented by trainers who themselves are inadequately trained or are so severely limited by time constraints that it is not possible to really do justice to the techniques they are expounding.

When one today examines certain commonly used procedures in statistical process control there is an undeniable focus on point estimation. This pre-occupation is to the point (pardon the pun!) that undue credence is given to estimates as reliable determinants of process parameters. Although fundamental to statistical reasoning, the notion that processes are perceived through the "key-hole" of a sample or samples is often not fully grasped. As a consequence of viewing the process through samples, judgements made about the process are susceptible to error. The degree of this uncertainty is vital knowledge if we are to make far reaching and expensive changes based on these judgements. It is one thing to obtain a point estimate of process variability and yet another to gauge its reliability. The time honoured manner in which this is done is by obtaining appropriate confidence intervals. Failure to do this or to run some appropriate simulations or repetitions, can give us a false sense of knowing which can further lead to all manner of unreasonable practices and claims. If it is to be engineers and non-statisticians who are to be the main users of statistical process control and allied techniques, then our training courses must become more effective in conveying the essentials of statistical sampling and estimation.

Provided we are reasonably confident that we are sampling from a normal process, a sample of 10 process values that yields $s = 1.2$ as a point estimate of the process standard deviation has a probability symmetric 95% confidence interval of $0.83 < \sigma < 2.19$. The corresponding interval for a sample of size 41 is $0.99 < \sigma < 1.54$. It is common in process control to estimate process variability (standard deviation) but rare indeed in my experience to examine confidence intervals to gain a "feel" of the reliability of these estimates.

3. Capability indices

This backdrop leads onto the main example of this presentation, the obsession that industry has with capability indices. As I write this paper I have three fairly popular textbooks on statistical process control sitting on my desk and none of them makes anything but passing reference to capability indices. One would in fact be excused, if being introduced to statistical process control for the first time, for believing that no great importance is to be placed on such indices. However, reality is that many industries are placing undue emphasis on them and it has been my experience that in certain industries the whole statistical process control effort has become focussed on them.

It is very important, therefore, that we who teach courses to engineers and technical people should appraise very carefully what we teach and how we teach it. We need in this area, I believe, to counter some wrong and misleading practices. A wrong practice in the quality area is unlikely to stay "in-house" especially if it is within a large company, as one important concept in quality improvement is to pass the "quality buck" onto suppliers; to stipulate the measures and performance they need to achieve to meet recipient requirements. In this way it is possible for bogus ideas to be passed from source to a whole chain of supplier companies. Management of supplier companies are often only too willing to comply in order to keep their customers happy. However, wrong practices passed onto the shop-floor for use and as a measure of evaluating quality performance can rapidly become sources of discouragement, discontent and to lead to a further deterioration in management/shop-floor relations. Further, they can lead to a contempt for statistical methods. It is frequently those who have little understanding of statistical methods who see the obvious practical fallacies that often escape those working in a management capacity - who have embraced quality ideas as commendable but who look for simple measures of quality status and improvement.

The manufacture of any product, however simple or complex, is done to a number of manufacturing specifications; let's assume that such specifications are a direct measure of customer satisfaction. For simplicity, suppose further that we are examining just one product characteristic that can be numerically measured, be it diameter, length, the result of a chemical analysis or whatever. Suppose upper and lower specifications of this measure are denoted by U and L respectively. If the manufacturing process standard deviation is σ then the capability index, C_p, is defined as:

$$C_p = \frac{U-L}{6\sigma}.$$

For a meaningful interpretation of this index the process has to be producing products with a characteristic that follows closely a Gaussian distribution. One property of such a distribution is that the vast majority of items will have a characteristic value in a 6σ band symmetric round the process mean, μ. Loosely then, 6σ can be described as the expected oscillation band of the process. Since we want the variability of the process to be small compared with the requirements placed on the product, C_p is reasonably required to be > 1, the bigger the better. If we regard U and L as fixed, the only way to increase C_p is to decrease σ which may represent excessive cost. Hence C_p should be greater than 1 to the extent that it can be achieved cost effectively. Now conceptually this is fine, it's plausible, and it's not difficult to understand. However, the best behaved process in the world will have its problems from time to time and non-normality, shifts in μ and changes in σ will occur. Provided, however, the process has been observed to behave stably for reasonable length periods, C_p measures the real potential of the process - the potential not actual performance in relation to making products within specifications. Of course σ is unknown and therefore so is C_p. The best we can do is to estimate C_p by $\hat{C}_p$ where

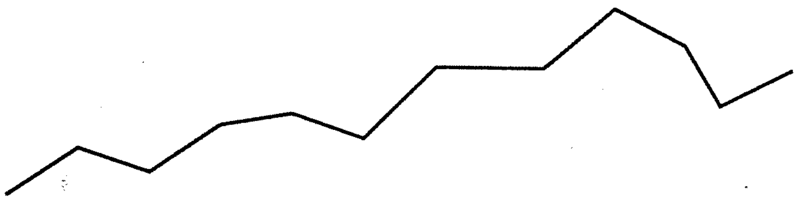
$$\hat{C}_p = \frac{U-L}{6\hat{\sigma}}.$$

Therefore, assessment of process potential involves obtaining a point estimate of C_p, $\hat{C}_p$ from a sample or samples. A confidence band for C_p can then be obtained by utilising the chi-squared distribution. If using s, from a single sample to estimate σ then 95% confidence bands for C_p can be obtained as exemplified in the following: If $n = 10$, $s = 1.2$, $U-L = 10$, $\hat{C}_p = 1.39$ then $0.76 < C_p < 2.02$. If $n = 41$, $s = 1.2$, $U-L = 10$, $1.09 < C_p < 1.69$.

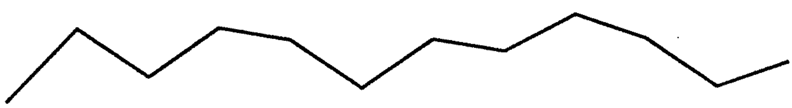
Since $C_p = 1$ is a threshold value there is some merit in obtaining the necessary value of $\hat{C}_p$ to ensure that $C_p > 1$ with a high degree of probability. When $C_p > 1$ we say the process is capable (of producing product consistently within specifications), for example, for $Pr(C_p > 1) \geq 0.99$ and $n = 10$, $\hat{C}_p$ must be > 2.08 and for $n = 41$, $\hat{C}_p$ must be > 1.34. It is somewhat unusual, in my experience, to find confidence bands for C_p calculated.

The issue of potential not actual becomes even more critical when we examine how σ is usually estimated. The most effective method to do this is to run one or several capability trials. The value of σ refers to the standard deviation of the process when it is running stably with as little outside interference as possible. It is therefore desirable to contrive this situation and sample intensively for the duration, using the obtained sample values to estimate σ. If the process is being adjusted during the data collection the observed standard deviation will be a combination of natural variation and that induced by interference, making a reliable estimate of σ tenuous. A capability trial can also provide the opportunity to examine just how long a process will remain stable and help establish a reasonable sampling period for subsequent monitoring and adjustment. It is not uncommon for the process to undergo changes during the capability trials and these changes have to be acknowledged and built in to estimation of σ.

Besides this, we wish to use the observed data to determine a mechanism for future process control that gets in harmony with the natural variation of the process. These realities mean that invariably s, the sample standard deviation, isn't the best means of estimating σ. When studying the process capability of continuous processes (common in the chemical industry) it is likely to have to examine the process over a period of several days, taking single samples every two or three hours. For illustration purposes suppose that these single sample test results are plotted on a chart and that the following situation has occurred.



A straight calculation of s for this data involves finding an average and obtaining the deviations of each data point from this average. This, however, ignores the obvious fact that the sample points are trending as a consequence of a moving process mean, so that the s calculation measures the combined effect overall, due to the trending mean of the process and the process variability. Considering, as we are, the potential of the process, we would reason that, could we arrest the mean trend, we would have observed the following pattern.



The calculation of the standard deviation from the data giving this plot would be considerably less than that calculated using the original. The estimation of σ, $\hat{\sigma}$ should, therefore, be done in a manner that serves the producer in assessing the capability of his process knowing that if he can control it as required, he can satisfy customer requirements. The standard deviation s, as calculated for the former situation, accurately "describes" the sample standard deviation of the products actually produced during the capability trial - the actual which, if a customer was buying he would get! It should be apparent from this that the estimating of standard deviation for the producer and consumer are different things. This subtlety is seldomly acknowledged. The issue is pertinent to the establishing of control lines for process control also, where we aim to control to the ideal situation.

If production monitoring is performed (as in most discrete item manufacturing processes) by use of mean and range charts, the issue of standard deviation estimation is again present. The standard estimation method of using

$$\frac{\bar{R}}{d_2} \text{ or } c_4\bar{s}$$

where $\overline{R}$ and $\overline{s}$ are respectively the mean range and mean standard deviation of the individual sample ranges and standard deviations $R_1, R_2, \ldots, R_n$, and $s_1, \ldots, s_n$. This method of calculation is what is meaningful to the producer in order to ascertain his capability and for subsequent process monitoring. As to what the consumer receives then

$$s = \sqrt{\frac{1}{(nk-1)} \sum_{ij} (x_{ij} - \overline{x}_{..})^2}$$

is the more meaningful estimate. This translates, in ANOVA parlance, to using the within group variation as being appropriate to the producer and total variation as that appropriate to the customer.

For single sample values, as are common in the chemical industry, $\overline{R}/d_2$ calculation where $\overline{R}$ is the moving average range of size 2 is virtually equivalent to calculating the square root of the average variance of successive overlapping data pairs. Let the data points be $y_1, y_2, \ldots, y_n$. The mean of the first two points is $(y_1 + y_2)/2$, giving a variance of $(y_1 - y_2)^2/2$. Now if this is repeated for all successive overlapping pairs we have:

$$(y_1 - y_2)^2/2,\ (y_2 - y_3)^2/2,\ \ldots,\ (y_{n-1} - y_n)^2/2.$$

Averaging, these give the standard deviation as

$$\sqrt{\frac{1}{2(n-1)} \sum_{i=1}^{n-1} (y_i - y_{i+1})^2}\ .$$

The point that is usually missed is that whilst $\overline{R}$ can be converted to an estimate of σ via d_2 for a Gaussian distribution, the estimate so obtained is an estimate of the within sample standard deviation. This estimate is usually appropriate for control chart use, i.e. control lines are located at $\overline{\overline{x}} \pm 3\overline{R}/(d_2\sqrt{n})$, usually denoted by $\overline{\overline{x}} \pm A_2\overline{R}$ where A_2 values are obtained from standard tables. Producer assessment of process capability likewise utilises the estimate, $\overline{R}/d_2$. The consumer, however, is more concerned with the actual rather than producer potential.

The following table of simulations from a Gaussian distribution illustrates the effects on estimation of σ when it is estimated first from the producer's perspective and secondly from the customer's. When the distribution is Gaussian and stable there is no appreciable difference but when the mean moves linearly the discrepancies are marked. Conceptually, excessive movements in process mean render meaningful interpretation of C_p tenuous anyway but this is rarely considered in practice.

μ	Sample Size	Producer (σ)	Consumer (σ)
100	40	.8161	.8863
100	40	.9550	.9812
100	40	.7715	.7187
100	40	.6896	.7654
98 - 100	20	.8264	.9464
98 - 100	40	.7367	.9337
98 - 100	60	.8553	.9668
98 - 100	80	.8892	1.0215
96 - 100	20	.7996	1.4247
96 - 100	40	.7247	1.4810
96 - 100	60	.8026	1.4717
96 - 100	80	.8881	1.4188

Note: $\sigma = 0.8$

4. Enter C_{pk}

Engineers at Ford Motor Company were early to embrace Taguchi ideas; these included extending or rather modifying the definition of quality. With this modification, priority of conformance to target was given a high profile. Taguchi maintains that there is a societal loss when products vary to any degree from the target dimension. Processes with small variability and mean close to target value were deemed highly desirable. Crosby's perception of "zero defects", where a defect is measured purely as conformance to manufacturing specifications, became outmoded. Rather, small variability and mean close to target became the order of the day.

With this concept, and with the confusion of whether C_p estimation is for producer use, for consumer use, or for both, came the desire to improve on the definition of C_p since this index makes no reference to process mean. Ford engineers thus conceived the C_{pk} index as

$$C_{pk} = \min\left\{\frac{U-\mu}{3\sigma}, \frac{\mu-L}{3\sigma}\right\}$$

where (U+L)/2 is assumed to be the desired target value. If $\mu = (U+L)/2$, $C_{pk} = C_p$, otherwise $C_{pk} < C_p$. As with the C_p index the C_{pk} depends for its interpretation on the process values following a Gaussian distribution. Again, as with the C_p index, C_{pk} can only be estimated and this is usually done using

$$\hat{C}_{pk} = \min\left\{\frac{U-\bar{x}_{..}}{3\hat{\sigma}}, \frac{\bar{x}_{..}-L}{3\hat{\sigma}}\right\}$$

for repeated samples of fixed size or by

$$\hat{C}_{pk} = \min\left\{\frac{U-\bar{x}}{3\hat{\sigma}}, \frac{\bar{x}-L}{3\hat{\sigma}}\right\}$$

for single values, as are commonly collected in the chemical industry. In both instances two parameters need to be estimated. Should these be estimated from the perspective of the producer or the consumer? This question is rarely addressed. As with the C_p index it would seem desirable to assess the reliability of point estimates of C_{pk} (whether from the producer's or consumer's perspective) by use of confidence intervals. This, however, is by no means trivial (Barnett, 1990). In the absence of any gauge of reliability it would seem unwise to give too much credence to point estimates of C_{pk}. However, it is common practice in many industries for production quality to be assessed on the basis of point estimates of C_{pk} with no regard to the most appropriate estimate of σ. In addition, point estimates of C_{pk} are often given absolute significance as if they were unwavering measurements of process quality performance. Production goals in particular industries are being based on obtaining specific $\hat{C}_{pk}$ values. Certain corporations are selecting suppliers on the basis of numerical $\hat{C}_{pk}$ values. The most prevalent abuse of this index that I have observed is in the chemical industry where samples are often, by necessity, small, distributions can often be non-normal and even sample values correlated, all of which make meaningful use of C_{pk} impossible.

5. Concluding remarks

I have addressed a very fundamental statistical issue but one that is nonetheless currently causing considerable concern. It has come about as a consequence of those using the techniques not fully grasping their nature and through their desire for simplicity. Effective education and training is the only lasting cure. Hard though it is, we must constantly strive in our courses to impart the true essentials of statistical reasoning which, hopefully, will offset the tendency for practitioners to slip into the erroneous use of techniques.

Reference

Barnett, N S (1990) Process control and product quality - the C_p and C_{pk} re-visited. *International Journal of Quality and Reliability Management* 7(4).