Australian Centre of Excellence for Risk Analysis

Report Cover Page

ACERA Project
2006 Round 1, Project 05
Title
Statistical methods for biosecurity monitoring
Author(s) / Address (es)
David Fox, School of Civil and Environmental Engineering, University of Melbourne
Material Type and Status (Internal draft, Final Technical or Project report, Manuscript, Manual, Software)
Project first final report
Summary
The aim of the first report from this project was to review and evaluate methods for monitoring systems in circumstances such as those that might be encountered in routine biosecurity inspections. The report intended to produce recommendations for potential development and testing of alternative monitoring approaches based on 'process control charts'. This report provides a review of basic statistical concepts that underlie these methods. In absence of relevant data and contexts for Australian quarantine, the report uses data from New Zealand on inspections and detections. It introduces some common control-charting techniques from manufacturing and veterinary epidemiology. It concludes that control charts are particularly well suited to visualising moderate to large volumes of time-based data such as would be expected in container inspection systems, and to detecting 'unusual' or 'aberrant' trends. Their utility for detecting the occurrence (in space) of an invasive species is more limited. For very low probability events, (eg. exotic disease outbreak) monitoring 'time between outbreaks' is a potentially more useful quantity, although the report shows that the statistical power (ability to correctly identify real 'shifts' in the mean time between events) of current charting techniques is relatively low. The report provides examples and makes suggestions regarding how these tools might be tested more effectively.

ACERA Use only	Received By:	Date:
ACERA Use only	ACERA / AMSI SAC Approval:	Date:
ACERA Use only	ACERA / AMSI SAC Approval:	Date:

STATISTICAL METHODS FOR BIOSECURITY MONITORING & SURVEILLANCE

I: CONTROL CHARTING

David R. Fox
Australian Centre of Excellence for Risk Analysis
University of Melbourne

March 2007

Acknowledgement

The author is grateful to Mark Burgman and two anonymous referees for their thoughtful comments on an earlier draft of this report.

Table of Contents

List of Figures

EXECUTIVE SUMMARY

This paper provides a review of basic statistical concepts as well as introducing some common control-charting techniques that have been successfully applied in areas as diverse as the manufacturing industries and veterinary epidemiology. The present review focus on the applicability of control charts for monitoring temporal trends and aberrations in bio-security related applications. Control charts are particularly well suited to the visualisation and assessing of moderate to large volumes of time-based data and as such would be expected to have greater utility for container inspection regimes say, than for detecting the occurrence (in space) of an invasive species. Control charts need to be viewed as just one method in a tool-kit of available techniques which can potentially assist field officers and quarantine risk assessors in identifying 'unusual' or 'aberrant' trends. For events having very low probabilities of occurrence (eg. exotic disease outbreak) the monitoring of 'time between outbreaks' is a potentially more useful quantity to be charting although as shown in this report, the statistical power (ability to correctly identify real 'shifts' in the mean time between events) of current charting techniques is relatively low.

An attempt has been made to illustrate the various forms of charting in the context of a quarantine inspection service. Due to a lack of availability of locally relevant data, the examples in this report have made use of publicly available data from new Zealand's container inspection program.

1. INTRODUCTION

This report is intended to provide managers, field officers, and researchers who have some responsibility for biosecurity monitoring and surveillance with an introduction into the principles and procedures of statistical process control (SPC) and in particular, control chart techniques.

Statistical Process Control can be broadly defined as the (statistical) methodologies and tools by which 'quality' is monitored and managed. With this definition, the original context of SPC is clear – to 'control' the quality of manufactured items in an industrial process or setting, although these days the word *control* is de-emphasised and is usually either dropped[1] or replaced by *improvement*. The development of SPC techniques can be traced back to the First World War and shortly thereafter with the introduction of the Shewhart control chart in the 1920s. General acceptance and uptake of SPC tools in the West was relatively slow until it was realised that a major contributing factor to the high productivity *and* quality of Japanese manufactured goods was due to that country's enthusiastic embrace of a total quality philosophy as espoused by leading (American) statistician and quality advocate– Edwards Deming.

The 1980s saw a resurgence of interest in SPC under the banner of 'Total Quality Management' or TQM. The 'six-sigma' philosophy was conceived during this time in response to Motorola's desire to achieve a tenfold reduction in product-failure levels within five years. The Six Sigma methodology (based on the steps Define - Measure - Analyse - Improve – Control) underpins the objectives of process improvement, reduced costs, and increased profits.

While much good work was done in the ensuing years with numerous examples of demonstrable success attributed to the SPC/TQM paradigm, the trend attracted some poorly credentialed 'experts' offering radical transformations to new (and often times unrealisable) levels of profitability. Not surprisingly, there were some failures and residual ill-feeling towards TQM. For example, one large company found that two thirds of the TQM programs it examined had been halted due to lack of results

[1] The American Society for Quality Control (ASQC) changed its name to the American Society for Quality (ASQ) on July 1, 1997.

while a 1994 American Electronics Association survey showed that TQM implementation had dropped from 86 percent to 63 percent, and that reductions in defect rates were not being realised (Dooley and Flor, 1998).

Despite some negative experiences in the manufacturing sector, TQM made a substantial contribution to quality improvement. The issue these days is one of how much quality is enough? For example, would anyone be interested in driving a 50-year old vehicle even if it was still under warranty? Or is there any value in ensuring that a computer will run reliably for 5 years when generational change in the computer industry is measured in months?

Environmental applications of TQM and SPC techniques have only more recently been identified despite the need for robust and reliable monitoring and surveillance systems. Fox (2001) attributes this to a lack of cross-talk between the 'brown' (industrial) statisticians and the 'green' (environmental) statisticians. Whatever the reasons, the slow uptake of SPC tools for environmental monitoring meant that critical assessments about environmental condition and important decisions about responses were being made on the basis of often-times flawed statistical advice. The ecologists' statistical toolkit was generally standard issue – t-tests, ANOVA, ANOSIM, and MDS were invariably represented and much used while the relatively simple techniques of Xbar/S charts, EWMA charts, and capability analysis were virtually unheard of.

The Australian Guidelines for Fresh and Marine Water Quality (ANZECC/ARMCANZ 2000) advocated a more 'risk-based' approach to water quality monitoring and assessment and in particular, a reduced emphasis on binary decision-making (t-tests and ANOVA) and increased emphasis on early-warning systems underpinned by dynamic visualizations (control charting). Shortly after the release of the Guidelines the events of 9/11 and the subsequent discovery of high-grade anthrax sent via the U.S. Postal service elevated the importance of early warning systems. In response, U.S. State and Federal governments have invested hundreds of millions of dollars in developing advanced surveillance systems to detect, among other things, another anthrax attack. Recognising that existing monitoring systems which relied on the gathering and processing of hospital records had unacceptable latencies, a new area of research emerged which aimed to provide close to real-time monitoring and warning of 'aberrant' events.

Syndromic surveillance is underpinned by a belief that signals of an emerging 'syndrome' such as a flu outbreak can be identified by an analysis of multiple time-series of ancillary variables such as absenteeism records and sales of non-prescription cold and flu medications together with an analysis of spatial clustering of outbreaks. Kulldorf (1997) developed a spatial scan statistic to help with the latter, while control charts were an obvious first candidate for the former. A major barrier at present is the difficulty in 'proving' that any of these new systems has made a difference or even do what they're meant to. As noted by Mostashari and Hartman (2003), no syndromic surveillance system has provided early warning of bioterrorism, and no large-scale bioterrorist attack has occurred since existing systems were instituted.

While the use of syndromic surveillance for counter-terrorism (see for example, http://www.bt.cdc.gov/surveillance/ears/) is a recent development, similar systems have been used for some time now to detect outbreaks, patterns, and trends in diseases and epidemics (see for example, http://www.satscan.org/). These techniques do not appear to have had any appreciable uptake in Australia or elsewhere around the world in quarantine inspection and bio-security. While adoption and uptake of SPC techniques in the Australian Quarantine and Inspection Service has been low, a search of the Department of Agriculture, Fisheries and Forestry (DAFF) website (www.daff.gov.au) reveals two (publicly available) documents that discuss the use of simple control charts (Korth 1997, Commonwealth of Australia 2002). One of these documents (Commonwealth of Australia, 2002) devotes a chapter to the use of control charting as an effective means of detecting trends for meat hygiene assessments. Control charting has also been recommended for detecting spatial and temporal clusters in veterinary monitoring (Carpenter, 2001) as well as testing waste streams from wastewater treatment plants (Hall and Golding, 1998). Stark et al. (2006) discuss risk-based veterinary surveillance approaches to protecting livestock and consumer health, although control-charting was not mentioned.

This report provides a review of basic SPC techniques (focusing on control charting methods) with a view to introducing our Client Agencies to 'new' or alternative approaches to monitoring that may offer enhanced anomaly detection capabilities. With a shared understanding of the control charting principles and an understanding of the strength and weaknesses of these methods, it is hoped that

opportunities for implementation and evaluation in an actual quarantine inspection / bio-surveillance environment will emerge.

2. BASIC STATISTICAL CONCEPTS

Statistics is concerned with random variation. Moreover, statistics is concerned with *random* variation. This does not mean that life for a statistician is totally unpredictable. The quantities exhibiting the random variation (the **random variables**) can be 'predicted' or described to the extent that in repeated 'trials' or observations, the values assumed by the random variable can be described by a *frequency distribution*. Figure 1 shows the **histogram** for the proportion of containers inspected each week during the period 15-Jan-2002 to 24-Dec-2006 that were found to have a quarantine risk[2].

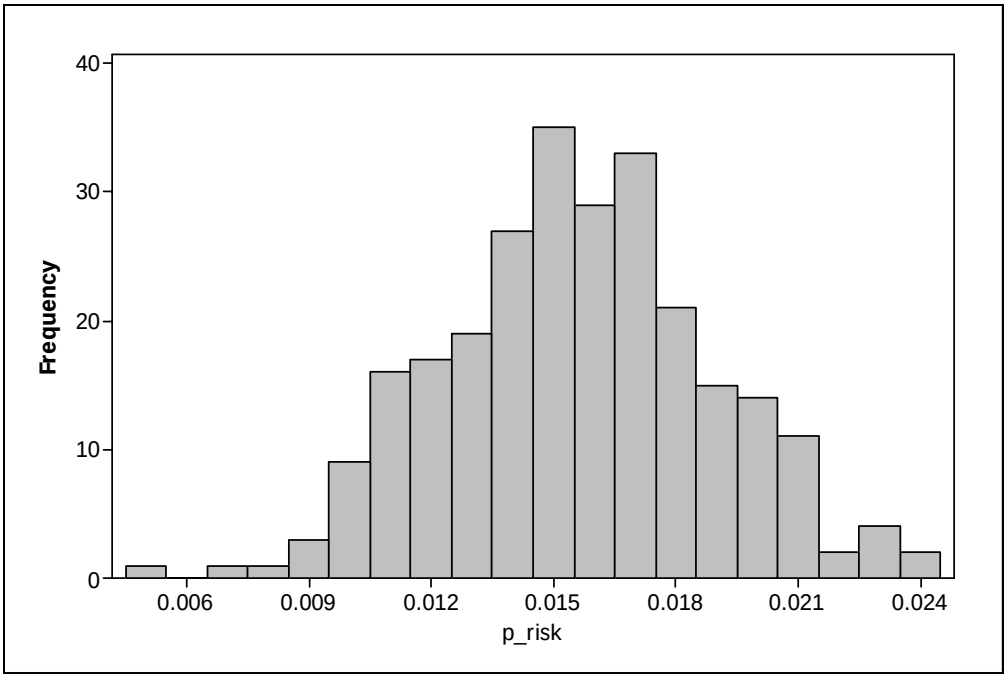

Figure 1. Histogram of weekly proportions of containers found to have quarantine risk.

A number of features are immediately apparent from Figure 1: (i) the *range* of observed proportions is from close to zero to approximately 2.5%; (ii) most proportions are around 1.5% - this is called the *modal* value; (iii) the distribution of

[2] Simulated data based on parameters derived from New Zealand's *Container Watch* publication – available at www.biosecurity.govt.nz .

proportions exhibits a high degree of symmetry (i.e. a lack of *skewness*); (iv) an eyeball estimate of an *average* or *mean* value is about 1.5%.

As a companion to the graphical summary provided by the histogram, we generally compute relevant *statistics* for our data. Some software tools (such as MINITAB) combine both the graphical and numerical summaries. Such a combined summary for the proportion data is shown in Figure 2.

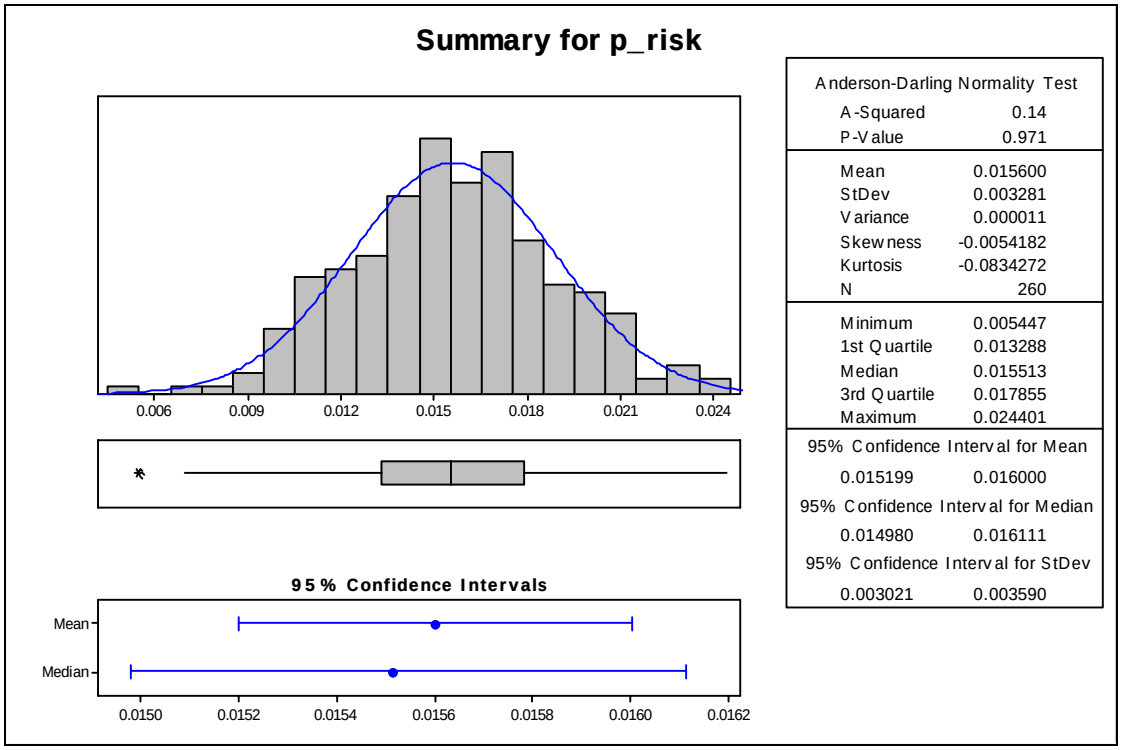

Figure 2. Statistical summary of proportion data. Curve overlaying the histogram is the best-fitting normal distribution.

The numerical summaries presented on the right-hand side of Figure 2 are explained below:

Anderson-Darling Normality Test

Because many statistical procedures (including control charting) rely on an implicit assumption of underlying normality in the distribution of the quantity of interest, the best-fitting normal distribution for the data at hand has been identified and overlaid on the sample histogram. While this allows for a visual inspection of the plausibility of the normality assumption, a more formal statistical test can be performed. There are many such tests – some good, others not so good (see for example Stephens, 1974). One such test that has been shown to perform well under a wide range of conditions is the Anderson-Darling test. The implicit hypothesis being tested is that the sample data has been drawn (at random) from a much larger population of

values which is normally distributed. In this case, the *test-statistic* is a value of A^2 (0.14 in Figure 2). The numerical value is rather meaningless by itself. In order to gauge the *significance* of the result we look at the companion *p-value*. The rule-of-thumb is that small p-values lead to a rejection of the implicit hypothesis (otherwise known as the *null hypothesis*). So, how small is small? Convention dictates that p-values of less than 0.05 are 'significant'. However you are cautioned against the unthinking adoption of this 0.05 norm. Since the p-value of 0.971 is well above the nominal 0.05, we do not reject the hypothesis of normality.

Mean

This is the usual *arithmetic mean* or *average* of the data. Other means are available (geometric and harmonic) although they are not widely used.

StDev and Variance

This is the *standard deviation* and is the most common measure of spread (or *dispersion*) of a data set. It is defined as the positive square root of the *variance*. Since the variance is computed by summing the square of differences formed by subtracting the (sample) mean from each data value, it (the variance) can only ever take on non-negative values. The reason for preferring the standard deviation as a measure of spread is that it has the same units as the original data values.

Skewness

As the name suggests, this is a measure of *skewness* or *asymmetry*. Skewness can be *negative* (long-tail to the left) or *positive* (long-tail to the right). Symmetrical distributions have *zero* skewness.

Kurtosis

This is not a quantity that is used often in its own right. It is a numerical measure of '*peakedness'* of a distribution. A flat distribution is said to have low kurtosis while a highly peaked distribution has high kurtosis. Different software packages will compute skewness slightly differently. The normal distribution has a skewness of 3. MINITAB and other software tools subtract 3 from the computed skewness so as to make the measure relative to a normal distribution. The skewness of -0.0834 in Figure 2 is very close to zero, implying that this sample data is about as peaked as a normal distribution.

Minimum and Maximum

These are self-evident.

First, second, and third quartiles

The *quartiles* are numerical values that divide the distribution into four equal parts. The first quartile (Q_1) is such that 25% of all values are less than or equal to Q_1; the second quartile (Q_2) is such that 50% of all values are less than or equal to Q_2; while the third quartile (Q_3) is such that 75% of all values are less than or equal to Q_3. Q_2 is also known as the *median* value.

The difference Q_3 - Q_1 is referred to as the *inter-quartile range* (*IQR*) and is a measure of spread or variation.

Confidence Intervals: mean, median, and standard deviation

An explanation of a confidence interval would require considerably more than a few lines. However, a practical interpretation is that we may assert that the *true parameter value* lies within the stated interval with stated degree of confidence.

Another way of presenting the data is the *box-plot*. The box-plot for the proportion data is shown in Figure 3. There are a number of variations on how the box-plot is constructed so you should check your computer software to be clear on the specifics. The information given by MINITAB (the statistical software used to produce Figure 3) is shown in Box 1 and general information on graphical summaries from SYSTAT is shown in Box 2.

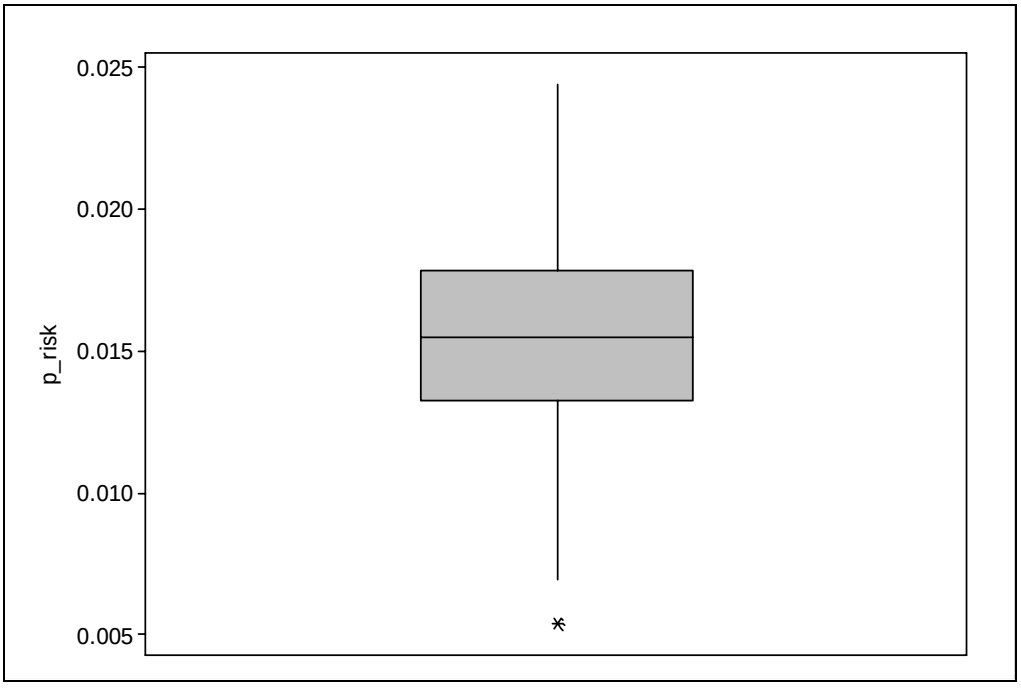

Figure 3. Boxplot for weekly proportion of containers found to have quarantine risk.

A number of features are immediately apparent from Figure 3: (i) the distribution of the proportion of containers having a quarantine risk is generally symmetrical about a mean/median of 0.015; (ii) the range is about 0.005 to 0.025; (iii)

the first and third quartiles are approximately 0.013 and 0.018 respectively and thus; (iv) the IQR is about 0.005; and (v) there is no evidence of extremes in the data (the single 'outlier' indicated by an asterisk in Figure 3 is of no real concern).

Box 1. MINITAB's Help on Box-plots

Graphical Summary
Boxplot (1 of 1)

Boxplots summarize information about the shape, dispersion, and center of your data. They can also help you spot outliers.

- The left edge of the box represents the first quartile (Q1), while the right edge represents the third quartile (Q3). Thus the box portion of the plot represents the interquartile range (IQR), or the middle 50% of the observations.
- The line drawn through the box represents the median of the data.
- The lines extending from the box are called whiskers. The whiskers extend outward to indicate the lowest and highest values in the data set (excluding outliers).
- Extreme values, or outliers, are represented by dots. A value is considered an outlier if it is outside of the box (greater than Q3 or less than Q1) by more than 1.5 times the IQR.

Use the boxplot to assess the symmetry of the data:

- If the data are fairly symmetric, the median line will be roughly in the middle of the IQR box and the whiskers will be similar in length.
- If the data are skewed, the median may not fall in the middle of the IQR box, and one whisker will likely be noticeably longer than the other.

In the boxplot of the precipitation data the median is centered in the IQR box, and the whiskers are the same length. This indicates that except for the outlier (asterisk), the data are symmetric. This is a good indication that the outlier may not be from the same population as the rest of the sample data.

Box 2. SYSTAT's help on graphical summaries.

Histogram, Box plot, Dot density, Density function

The density of a sample is the relative concentration of data points in intervals across the range of the distribution. A histogram is one way to display the density of a quantitative variable; box plots, dot or symmetric dot density, frequency polygons, fuzzygrams, jitter plots, density stripes, and histograms with data-driven bar widths are others.

A histogram is the most familiar one among these displays. The word comes from a Greek word (*histos*) for a straight standing beam, like a mast or loom frame, and a word (*gram*) for a drawn picture. Thus, a histogram is a pictorial display of vertically standing bars. It is a crude density estimator because the shape of a histogram depends upon the choice of the number of bars. Most other graphical density estimation methods depend on subjective choices of parameters (or settings) as well, which is one reason the general field of density estimation is rather controversial (Wegman, 1982).

The software can use the sample mean and standard deviation to construct a normal curve (or cumulative normal curve) for comparison against the actual anomalies of the sample distribution. A kernel curve is also available for density and distribution curves.

Rather than comparing sample values to the normal distribution (mean, standard deviation, and so on), box plots show robust statistics (median, quartiles, and so on). Some complain that box plots or the choice of intervals for bars in a histogram can mask gaps or separations in the distribution. Dot histograms (dit) and symmetric dot displays (dot) answer this problem because they display every value in the sample. It is often useful to examine both a box plot and a dit display. A gap histogram is another alternative. Its bar widths vary across the range of the distribution--when there are gaps, the neighboring bar is made wider to include the gap.

Fuzzygrams superimpose a probability distribution on each bar of a histogram. Bars for histograms based on small samples are fuzzier than bars for large sample histograms. Jittered dot density displays points by calculating the exact locations of the data values and then, to keep points from colliding, jittering them randomly on a short vertical axis. These displays work better for large samples than small samples. Density stripes are vertical lines placed at the location of data values along a horizontal data scale and look like supermarket bar codes. For large samples, the stripes tend to collide, so you should consider a jitter dot density instead.

Bivariate densities can be displayed as 3-D histograms and as 2-D surfaces or contours constructed using normal theory sample statistics or a nonparametric kernel estimator.

These displays can be stratified across the levels of a grouping variable; for 2-D displays, if the grouping variable has only two values, a dual (or back-to-back) version is available.

The Dynamic Explorer is available for rotating 3-D displays and also for fine tuning all displays.

Graphics

In addition to the graphical summaries already presented, it will be useful to display the time-series plot (ie. some numerical value plotted over time) for the monitored data. Figure 4 shows a plot of the number of containers inspected each week and the number of containers found to be a quarantine concern. Given that the former series is about three orders of magnitude greater than the latter, plotting these two variables on the same axes is of limited value. Given that both the number of

containers inspected and the number of 'detects' vary with time, a more useful quantity to plot is the *proportion*, $p = \frac{N_{\text{detect}}}{N_{\text{inspect}}}$ (Figure 5).

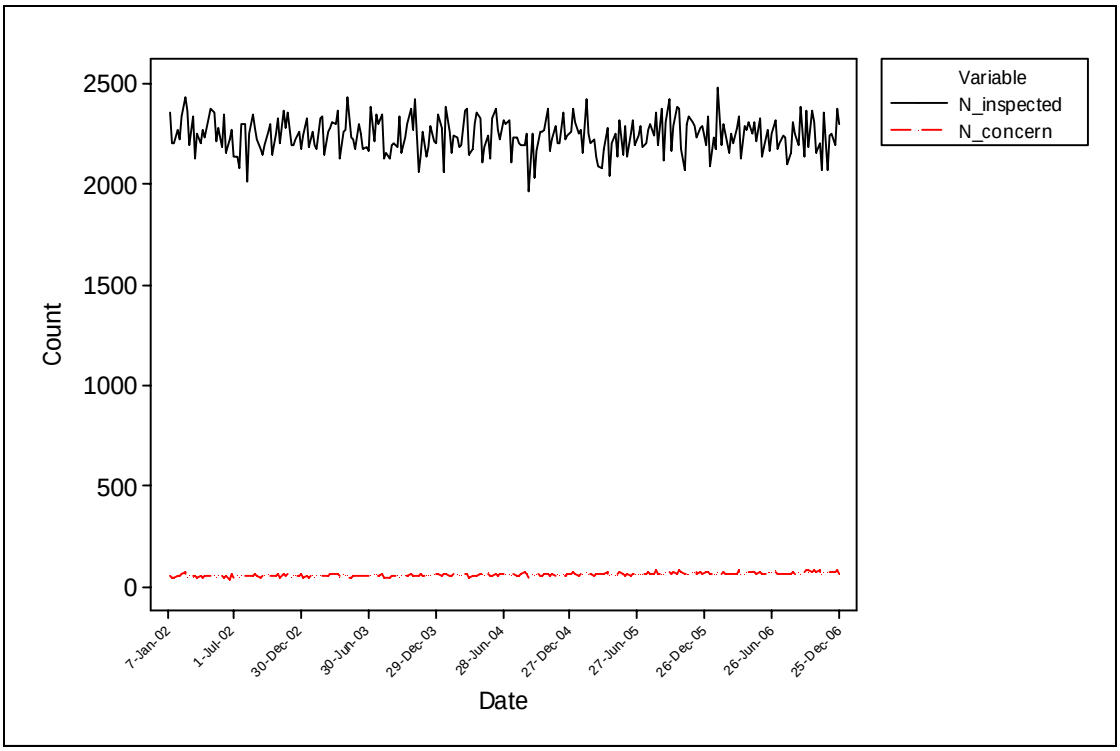

Figure 4. Time series plot of number of containers inspected per week (black line) and number of containers found to be of quarantine concern (red line).

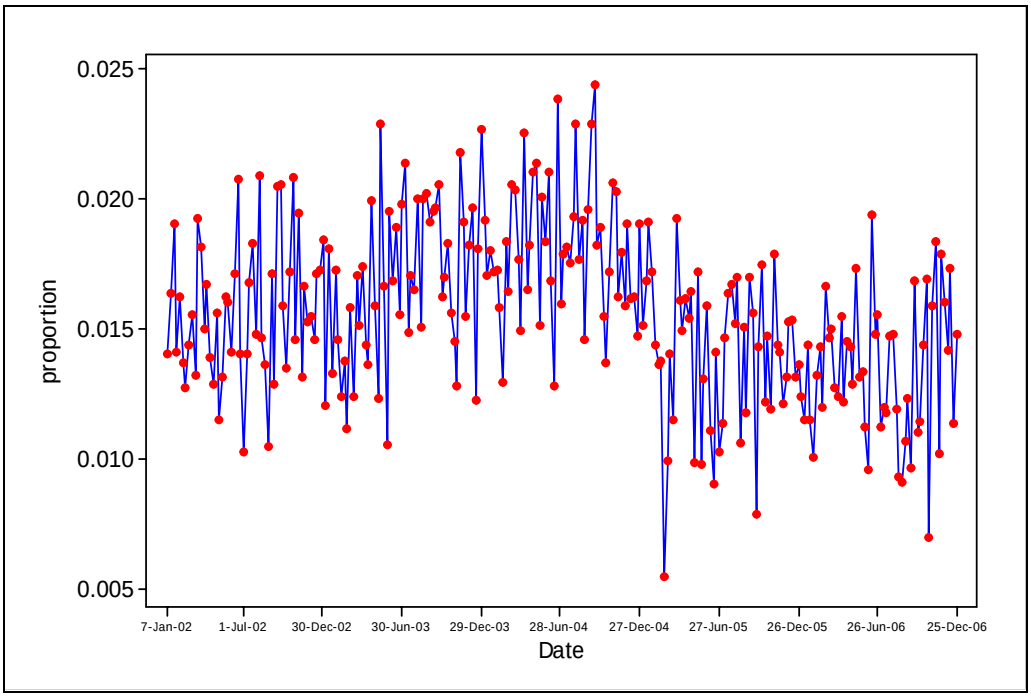

Figure 5. Times series plot of proportion of inspected containers (per week) found to be a quarantine concern.

Figure 5 is potentially far more revealing. The weekly oscillations in the proportion of detects is quite apparent. It also appears that on or around February 2005 there was a 'step-change' (a drop) in the proportion. A useful technique for exploring trends in graphs like Figure 5 is to superimpose a smoothed version of the data. There are many statistical 'smoothers' although one that is commonly used is the *loess* smooth which stands for locally weighted regression. The main idea behind loess and other smoothers is to subset the data and replace the individual data points by the value of a fitted function or some statistic such as the median of the data. Figure 6 shows a smoothed version of the proportion data of Figure 5. The step-change discussed above is now clearly evident and in fact one might speculate that there are three distinct periods evident in the series: prior to about March 2003; March 2003 to around February 2005; and post-February 2005.

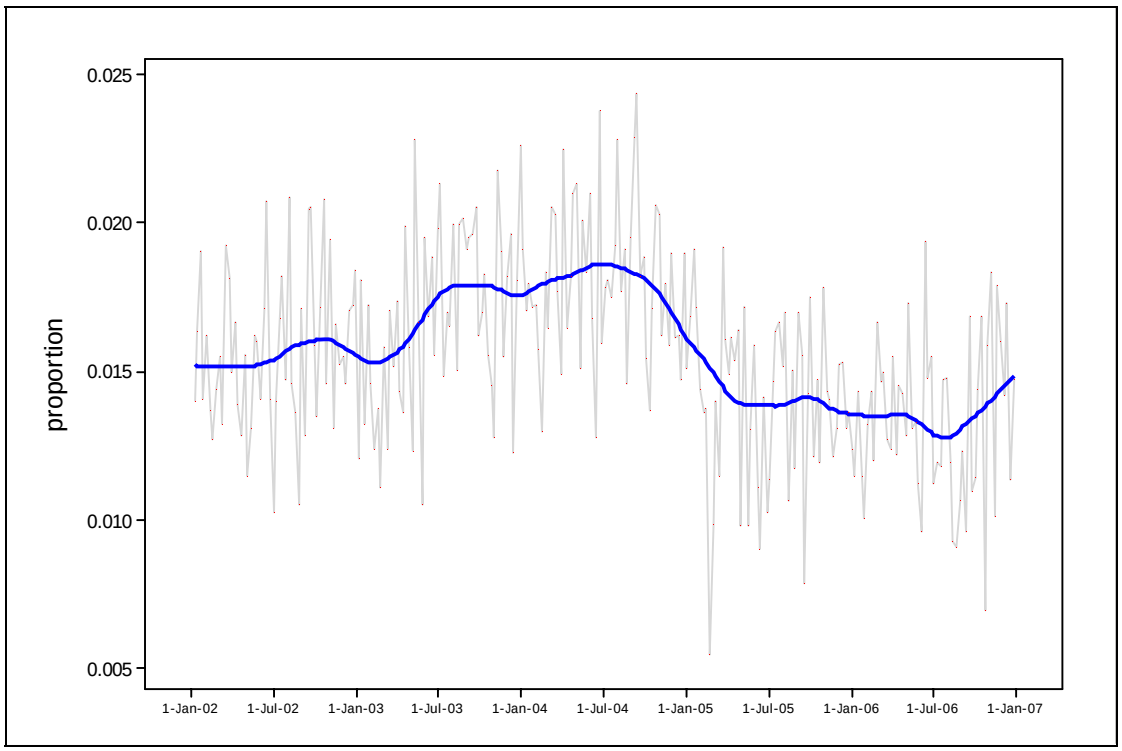

Figure 6. Proportion data of Figure 5 with loess smooth superimposed.

Whether or not these epochs are real or an artefact of random fluctuations is something that cannot be immediately answered. One way of helping address the 'significance' of these variations is through the use of *control charts*.

3. Statistical significance

While a comprehensive treatment of statistical inference is outside the scope of this introductory note, a brief discussion of the main ideas underpinning 'statistical significance' is warranted.

To place the discussion in context, suppose that it is known from lengthy monitoring that the proportion of containers classified as a 'quarantine risk' is normally distributed with a mean of 0.0156 and a standard deviation of 0.0033. Statisticians write this in an abbreviated way as $\theta \sim N(0.0156, 0.0033^2)$ where θ denotes the true proportion. Figure 7 shows a plot of the *probability density function* (or *pdf*) for this particular normal distribution.

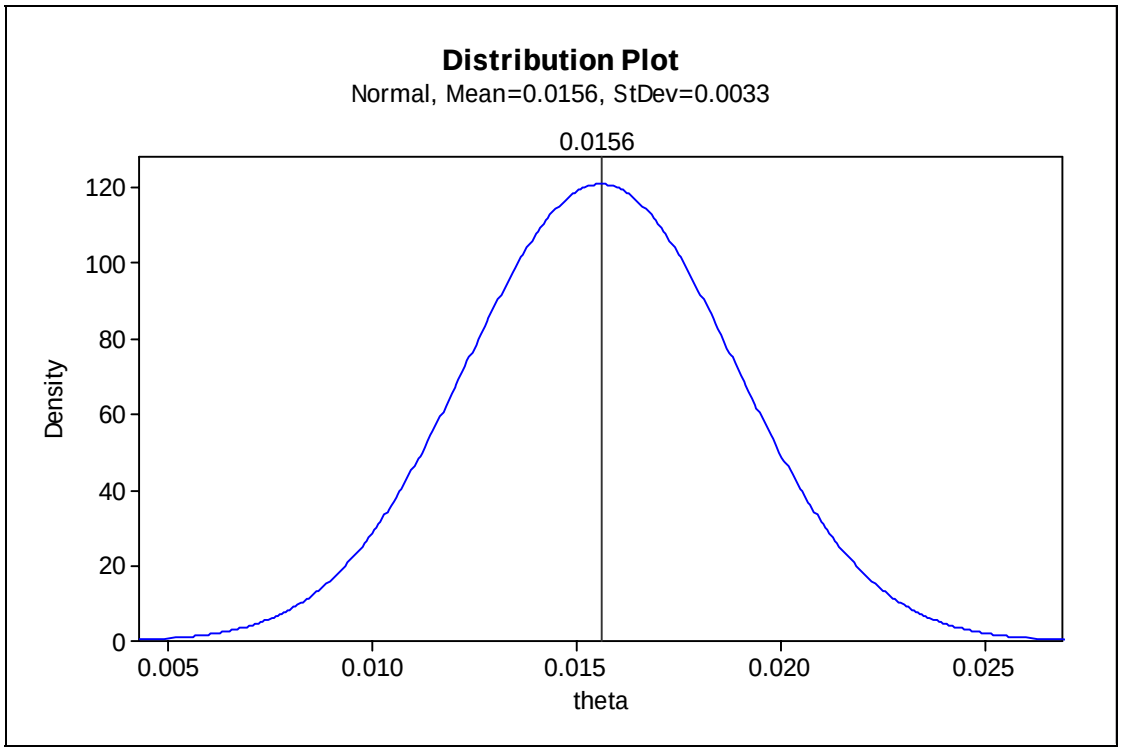

Figure 7. Theoretical distribution for the true proportion of quarantine risk containers.

Now, suppose we collected *n=52* weekly proportions and looked at the average (arithmetic mean) proportion of containers classified as a quarantine risk. Statistical theory tells us that if $\theta \sim N(0.0156, 0.0033^2)$,then the average of *n* sample proportions is distributed as $\bar{\theta} \sim N(0.0156, 0.0033^2/n)$. For *n=52*, this distribution is shown in Figure 8 (together with the original distribution of Figure 7).

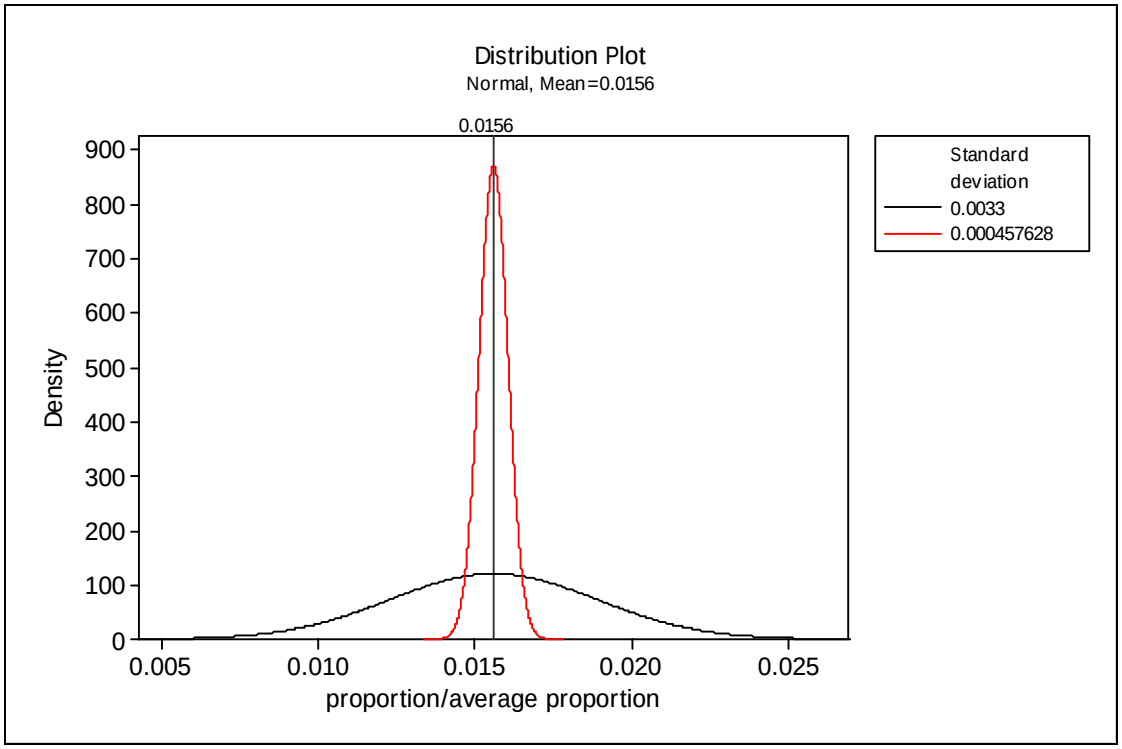

Figure 8. Theoretical distributions for an individual proportion (black) and the average of 52 proportions (red).

It is readily seen from Figure 8 that while both distributions are centred about the same mean value, the distribution of the average is more compactly so. Thus, while an *individual* proportion of say, 0.018 or more is expected to occur quite frequently it is highly unlikely that the average of 52 readings would be this high if the individual values had been drawn from a population described by the red curve in Figure 8. In statistical parlance, we would say that an average (of *n=52*) proportion of 0.018 or greater is a highly *significant* result. This immediately raises the question of 'how significant is significant' or where do we draw the line between statistical significance and non-significance? Convention dictates (and this is a potentially dangerous approach without understanding the ramifications) that we choose a cut-off value (call it θ^*) such that the probability $P\left[\bar{\theta} > \theta^*\right]$ is 'small' – and small is taken to mean 0.05. For our example, we find that a θ^* equal to 0.0164 has an associated 'tail probability' of 0.05 (Figure 9). So, if we are only interested in large proportions, then we will declare a sample average (of *n=52*) to be *statistically significant* if it is numerically greater than 0.0164. If we are interested in both increases and decreases in the assumed proportion, we split the 0.05 area into two 'tail' areas each of 0.025. This generates a *two-sided* test of significance (Figure 10).

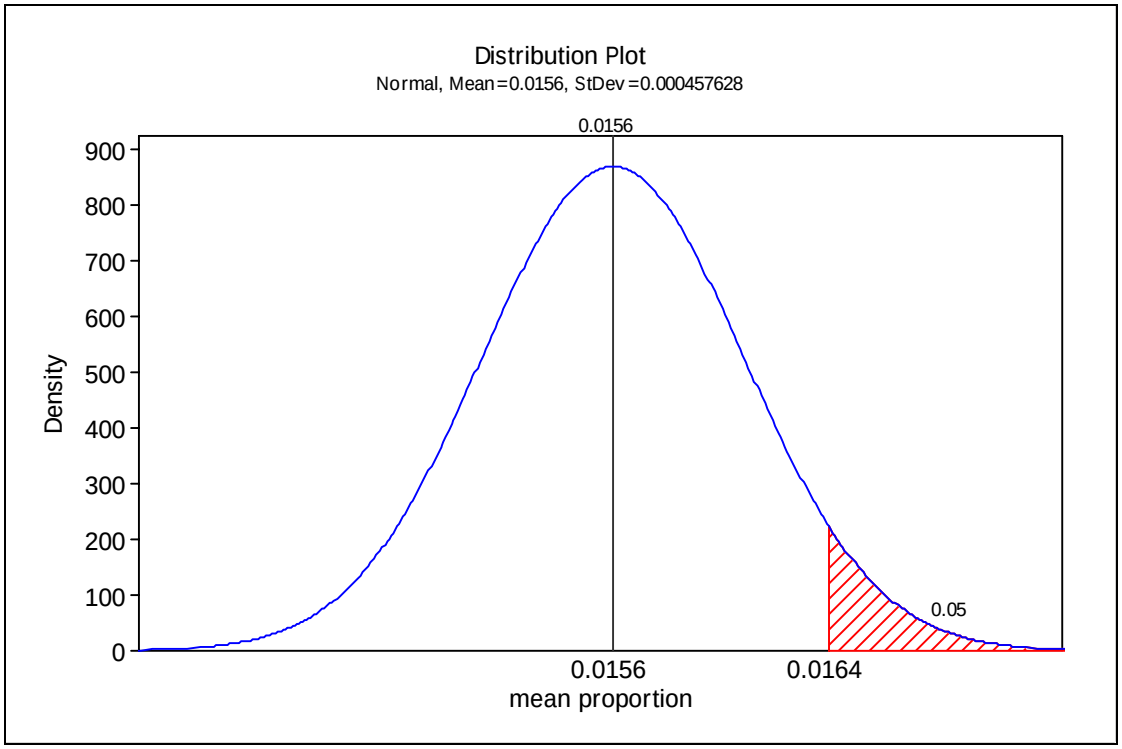

Figure 9. Assumed distribution for mean of *n=52* sample proportions. One-tail, 5% 'critical region' identified by red shading.

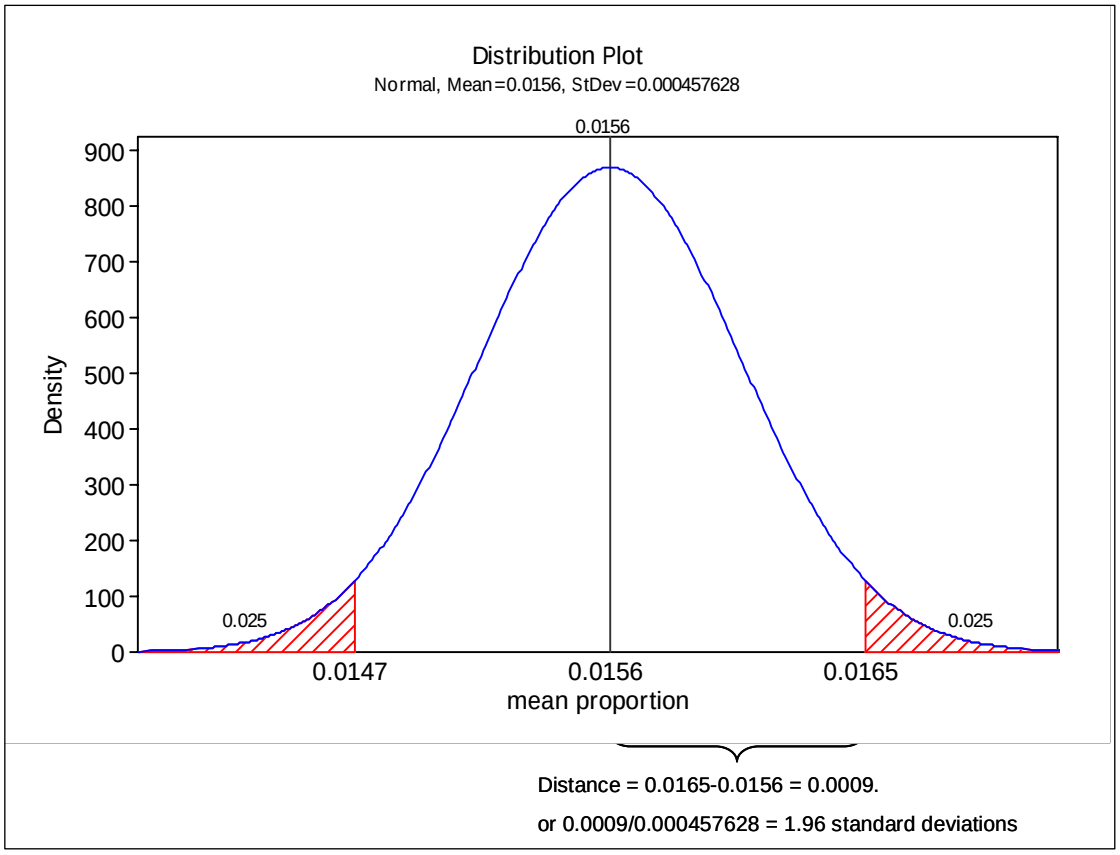

Figure 10. Assumed distribution for mean of n=52 sample proportions. Two-tail, 5% 'critical regions' identified by red shading.

It is also seen from Figure 10 that the upper limit of 0.0165 is equivalent to 1.96 standard deviations from the mean (it is easily verified that the lower limit is also 1.96 standard deviations from the mean, but obviously in the opposite direction). These ± 1.96 'sigma limits' or 'control limits' form the basis of 'early warning' or detection limits on control charts. The multiplier (1.96) determines the width of the limits which in turn determines important statistical properties associated with 'false triggering'. Thus, there is a trade-off to be struck: very narrow limits will give high sensitivity to shifts in the 'process' but at the expense of increased rates of false-triggering. By default, limits on control charts are placed at either 2 or 3 standard deviations from the mean (centre-line). Limits determined through non-statistical considerations (eg. biological or ecological significance) can also be used, although the implications for 'statistical significance' would need to be determined if the chart was to be used in an inferential mode.

4. Control Charts

The Basic Shewhart Chart

The simplest control chart is essentially Figure 5 with the addition of upper and/or lower control limits. This could be done manually, although it is easier to have computer software do it. The output from the MINITAB statistical software package is shown in Figure 11.

Note that the upper and lower control limits in Figure 11 are not constant. This is because the denominator in the expression $p = \frac{N_{\text{detect}}}{N_{\text{inspect}}}$ is not constant (as noted by the warning message in the lower right corned of Figure 11). The red plotting symbols in Figure 11 indicate 'violations' or 'excursions' outside the control limits. By itself, this chart raises no particular concerns, other than there were two occasions when the proportion was 'significantly' high and three occasions when it was 'significantly' low. Various modifications and options are available in the presentation of control charts such as Figure 11. For example, if we think there are two 'epochs' as discussed earlier, we can provide separate limits in each epoch (Figure 12).

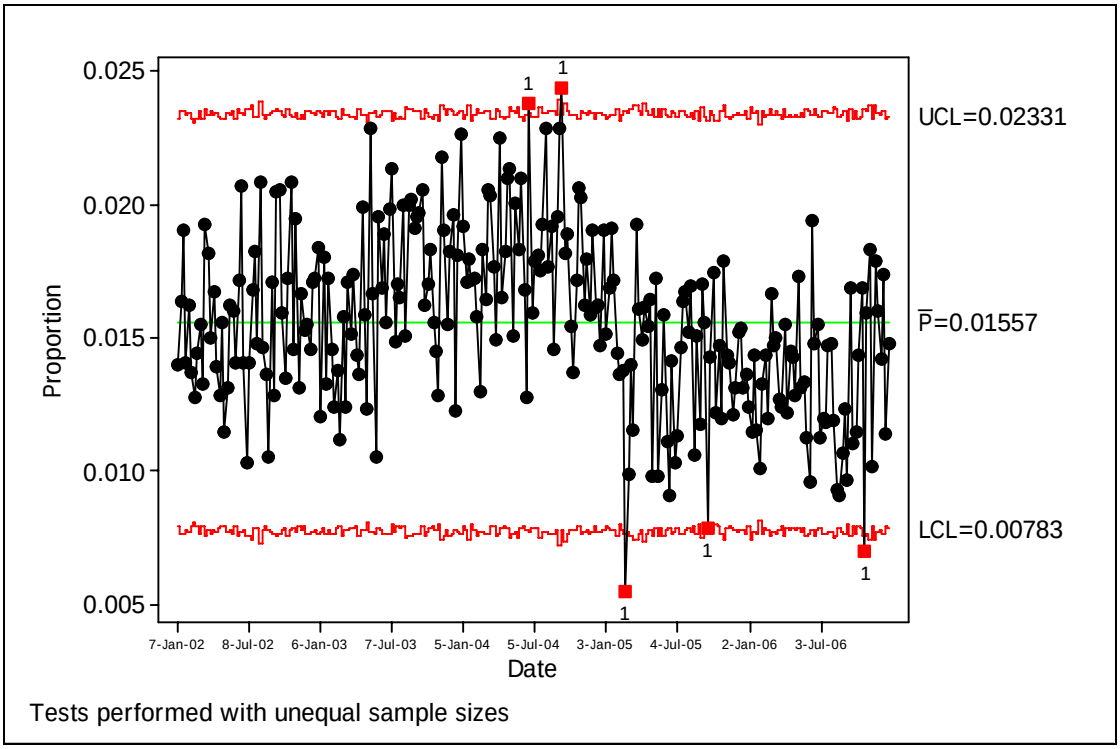

Figure 11. P-chart for proportion data.

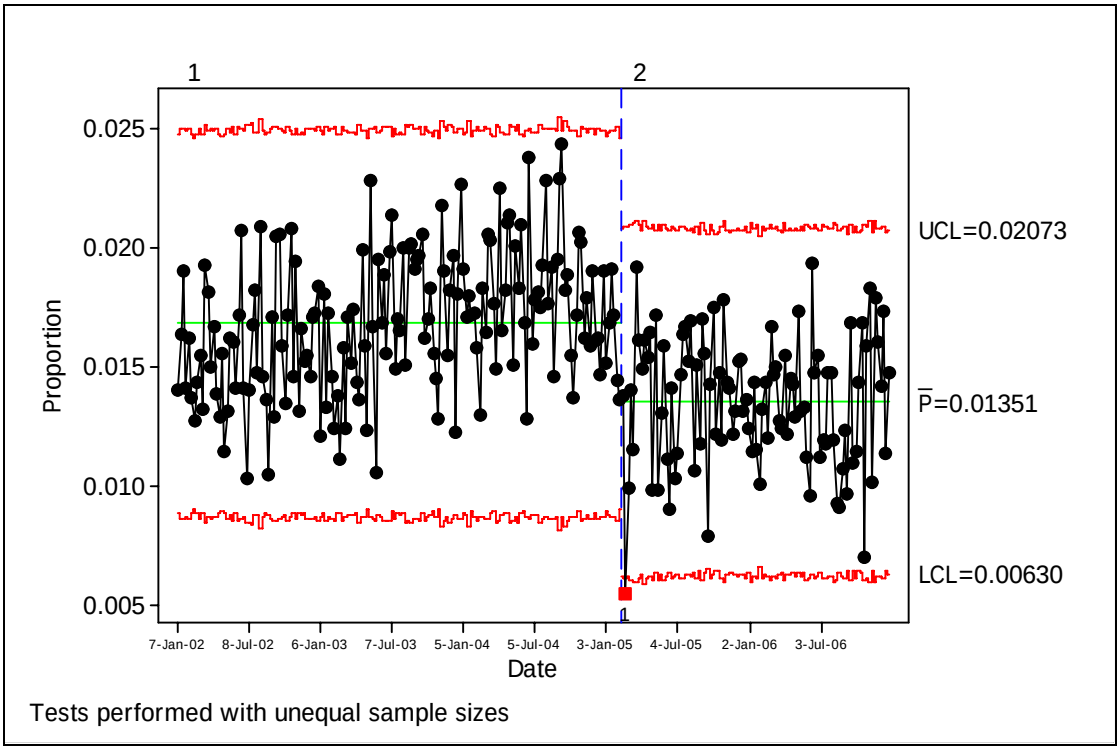

Figure 12. As for Figure 11 but with two periods identified: before (period 1) and after (period2) February 14, 2005.

For data that is collected over *time* a number of time-based control charts are available. Some of the more common/useful are described below.

Time-based charts

The simplest way of smoothing over time is by 'block-averaging'. Figure 13 shows the series of Figure 5 divided into a number of non-overlapping 'blocks' of constant width. The average of the data in each block is computed and plotted at the centre of the block and these points can be connected by straight line segments to reveal a smoother version of the series.

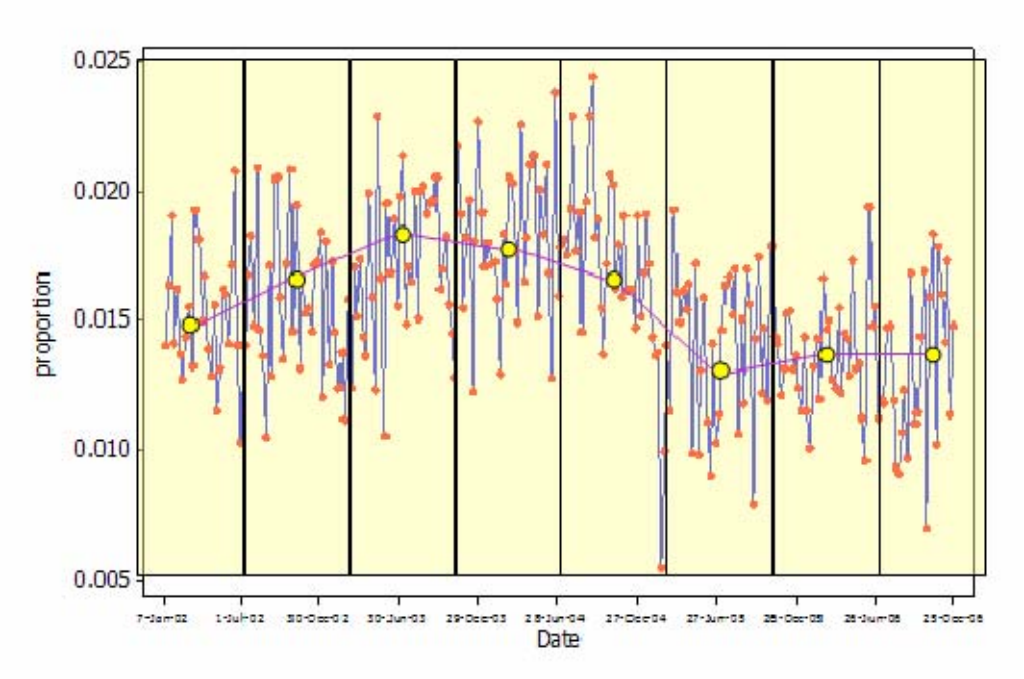

Figure 13. Smoothing using block averaging.

Block averaging is a relatively unsophisticated way of smoothing and has some potential difficulties – not least of which is that the mean of each block is computed without reference to the rest of the series. In other words there is no 'history' built in to the mean of an individual block and so the 'smoothed' series can still exhibit some erratic jumps. To overcome this, we can take the basic 'block' or 'window' and step it across the series so that there is overlap. This is achieved by replacing the 'oldest' k observations with the most recent k observations. The block averages in this case result in a *moving average* (MA) of the original series (Figure 14). A MA plot for the proportion data is shown in Figure 15. The MINITAB help for this procedure is given in Box 3.

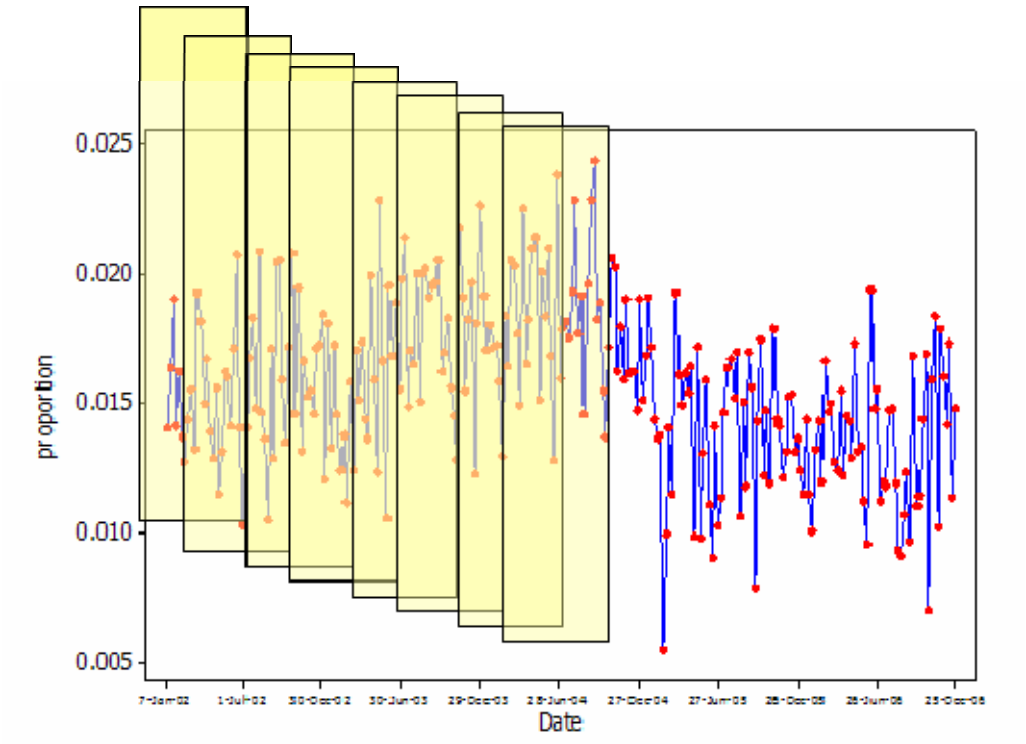

Figure 14. Moving average scheme. A block or 'window' is stepped incrementally over the series and the block mean computed and plotted.

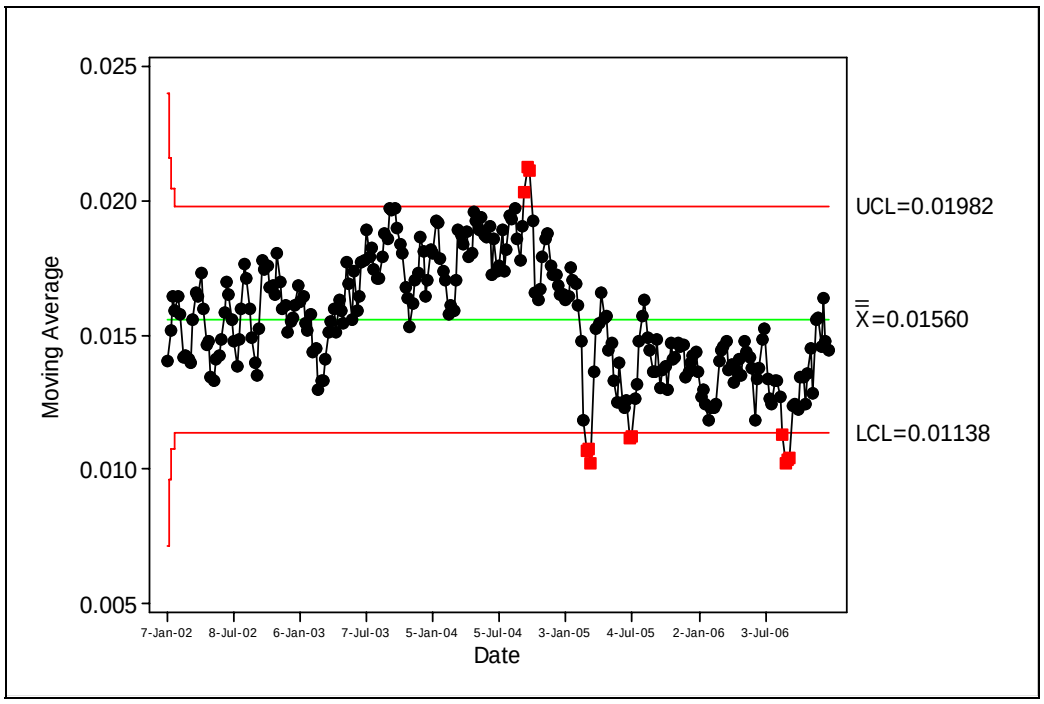

Figure 15. Moving average chart of proportion. Sub-group size =1; MA length=4.

Box 3. MINITAB's Help on moving average chart

Moving Average Chart

overview how to example data see also

Stat > Control Charts > Time-weighted Charts > Moving Average

A moving average chart is a chart of moving averages – averages calculated from artificial subgroups that are created from consecutive observations. The observations can be either individual measurements or subgroup means. This chart is generally not preferred over an EWMA chart because it does not weight the observations as the EWMA does.

When data are in subgroups, the mean of all the observations in each subgroup is calculated. Moving averages are then formed from these means. By default, the process standard deviation, σ, is estimated using a pooled standard deviation. You can also base the estimate on the average of subgroup ranges or subgroup standard deviations, or enter a historical value for σ.

When you have individual observations, moving averages are formed from the individual observations. By default, σ is estimated σ, with $\overline{MR}$ / d2, the average of the moving range divided by an unbiasing constant. Moving ranges are artificial subgroups created from consecutive measurements. The moving range is of length 2, since consecutive values have the greatest chance of being alike. You can also estimate σ using the median of the moving range, change the length of the moving range, or enter a historical value for σ.

For more information, see Control Charts Overview and Time-Weighted Control Charts Overview.

Dialog box items

All observations for a chart are in one column: Choose if data are in one or more columns, then enter the columns.

Subgroup sizes: Enter a number or a column of subscripts. If the subgroups are not equal, each control limit is not a single straight line but varies with the subgroup size. If the subgroup sizes do not vary much, you may want to force the control limits to be constant by specifying a fixed subgroup size using MA Option > Estimate.

Observations for a subgroup are in one row of columns: Choose if subgroups are arranged in rows across several columns, then enter the columns.

Length of MA: Enter the length of the moving averages. The value you enter is the number of subgroup means to be included in each average. If you have individual observations (that is, you specified a subgroup size of 1), Minitab uses them in place of the subgroup means in all calculations.

By adjusting the parameters of the moving average plot, different levels of smoothing can be achieved. For example, Figure 16 shows a MA plot for the proportion data using the averages of 3 weekly sub-groups. With this degree of smoothing, the three periods are clearer.

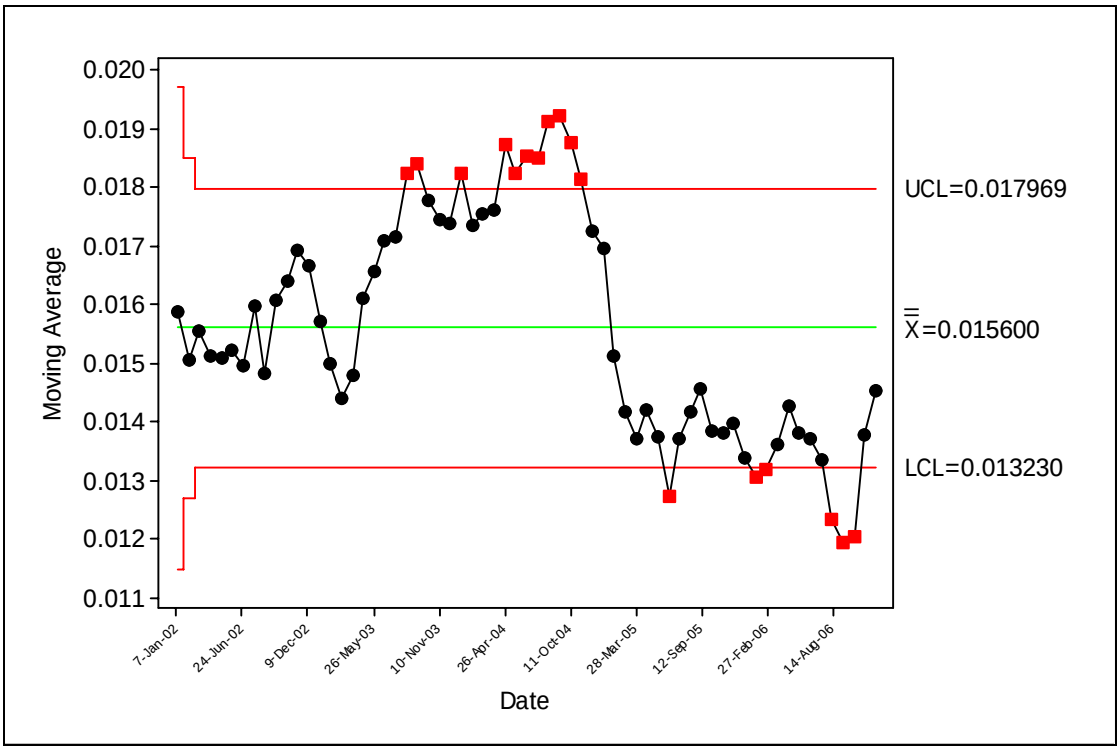

Figure 16. Moving average for proportion. Subgroup size defined by weeks in each month (usually 4); MA length=3.

Time-weighted charts

The block or moving average charts just discussed give equal importance or weighting to all data in the current 'window'. While this may be appropriate in some situations, it doesn't accord with the usual notion that the greater the time separation, the less influential the data becomes. Time-weighted control charts such as the *Exponentially Weighted Moving Average* (EWMA) are more flexible in that the relative weightings given to recent and historical data can be specified.

By way of example, suppose we wish to form a weighted average of the current observation and the (*k-1*) most recent values. That is, we're interested in forming the *weighted mean* $\bar{X}_1 = \alpha X_1 + \alpha^2 X_2 + \ldots + \alpha^k X_k$ where the weighting factor is α. The requirement that $\bar{X}_1$ is an unbiased estimator of the true mean imposes the constraint $\sum_{i=1}^{k} \alpha^i = 1$ and for a given *k* the solution to this is the root of the equation $\alpha^k - 2\alpha + 1 = 0$. For example, if *k=10*, we find $\alpha = 0.5002$. A plot of these weights compared to the simple arithmetic mean is shown in Figure 17.

The recursive formula for computing values of the EWMA chart is

$$EWMA_t = \alpha X_t + (1-\alpha) EWMA_{t-1} \qquad 0<\alpha<1$$

In other words, the current EWMA is a weighted average of the current data value and the EWMA in the preceding period. Figure 18 shows the EWMA chart for the weekly proportion data with $\alpha = 0.2$ while a more smoothed version based on aggregated (monthly) data is shown in Figure 19.

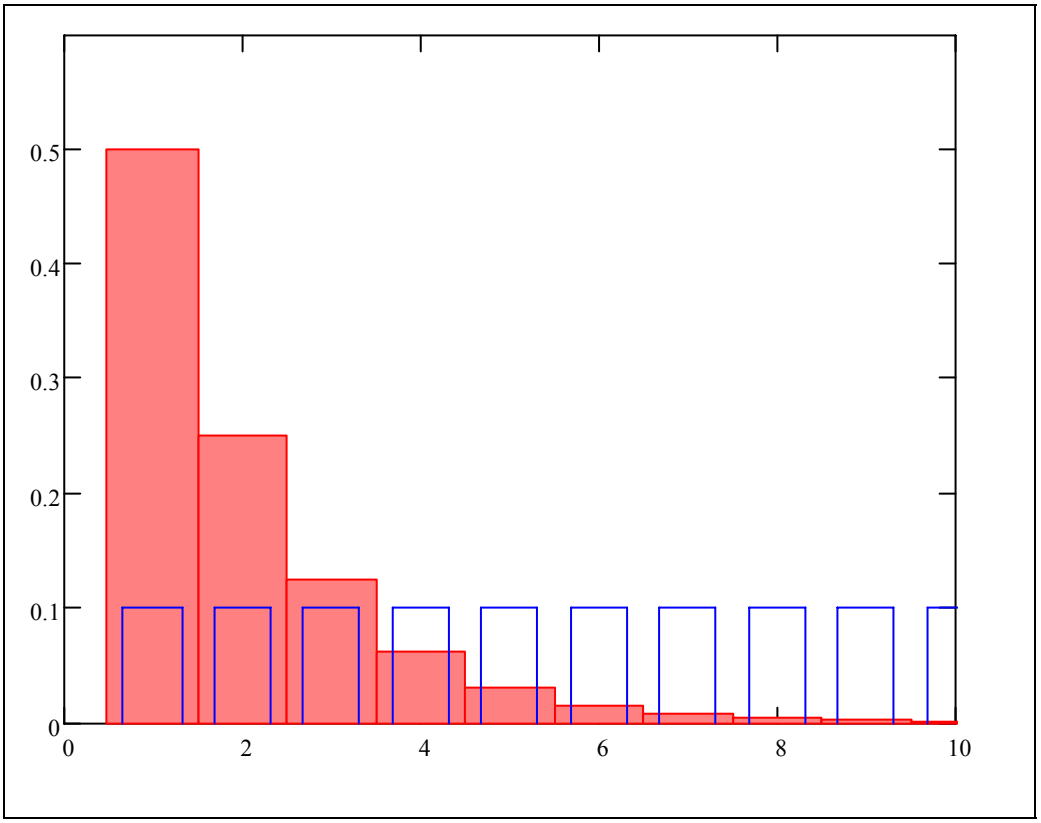

Figure 17. Comparison of exponentially declining weights (red bars) compared with equal-weighting scheme (blue rectangles) for *k=10*.

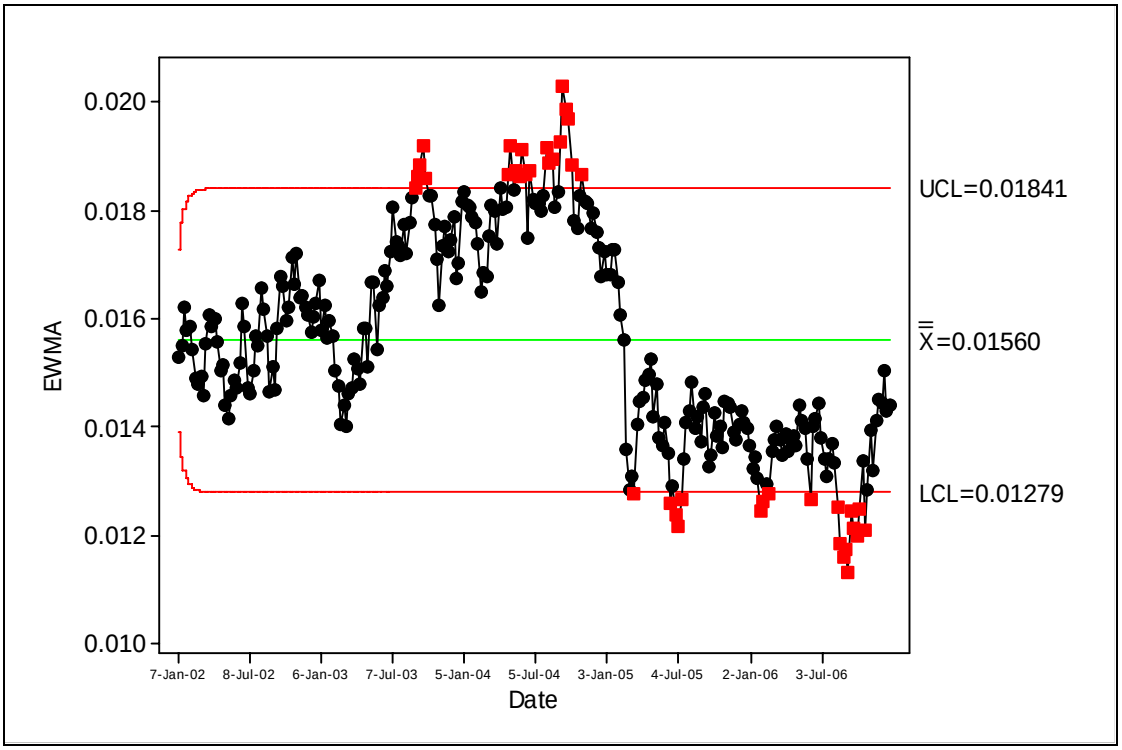

Figure 18. EWMA chart for proportion. Subgroup size=1; EWMA weight=0.2. Red plotting symbols denote 'out of control' points.

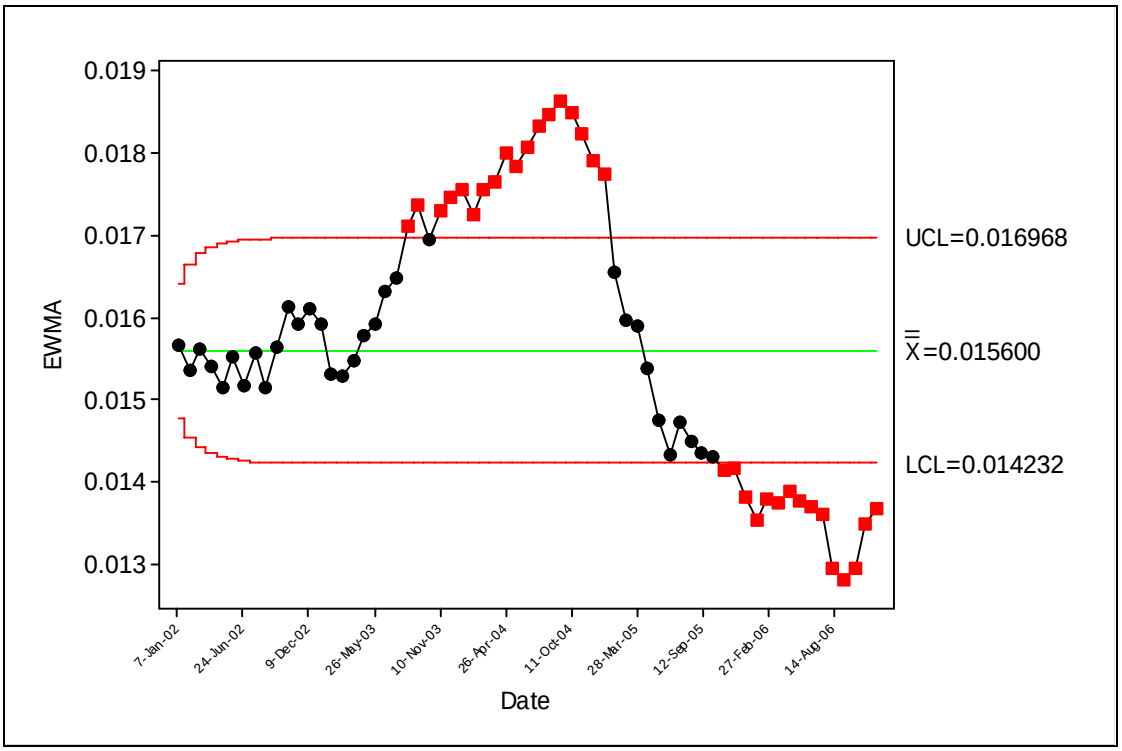

Figure 19. EWMA chart for proportion. Subgroup size= number of weeks in month (generally 4); EWMA weight=0.2.

5. Time between events

We have seen how control charts can be used to monitor *variables* (such as the *number* of quarantine threats detected) or *attributes* (eg. the *proportion* of containers having a quarantine threat). When 'events' (eg. the 'arrival' of a quarantine risk at a port of entry) occur randomly in time, an alternative approach is to monitor the *inter-arrival time*. One advantage of this approach is that the inter-arrival time is available at the time of the arrival whereas if we are tracking the number of arrivals then these need to be aggregated over some time period before a meaningful analysis can be performed. However, some modifications to the standard charts are required to accommodate the fact that the distribution of *inter-arrival* times is usually (highly) non-normal. Details of the theoretical development can be found in Radaelli (1998). More recently, control charts for the number of 'cases' between events (so-called *'g'* and *'h'* charts) have been developed and applied to monitoring hospital-acquired infections and other relatively rare adverse health-related events (Benneyan 2001a, 2001b).

By way of example, consider Figure 20 which depicts the 'arrival' of a quarantine threat over time. Figure 21 shows a more detailed 'slice' through this pattern of arrivals. By measuring the 'white-spaces' in Figure 21 we obtain data on the inter-arrival times. The complete listing of data for this example is given in the Appendix.

Our analysis of the inter-arrival data commences with an inspection of basic distributional properties (Figure 22).

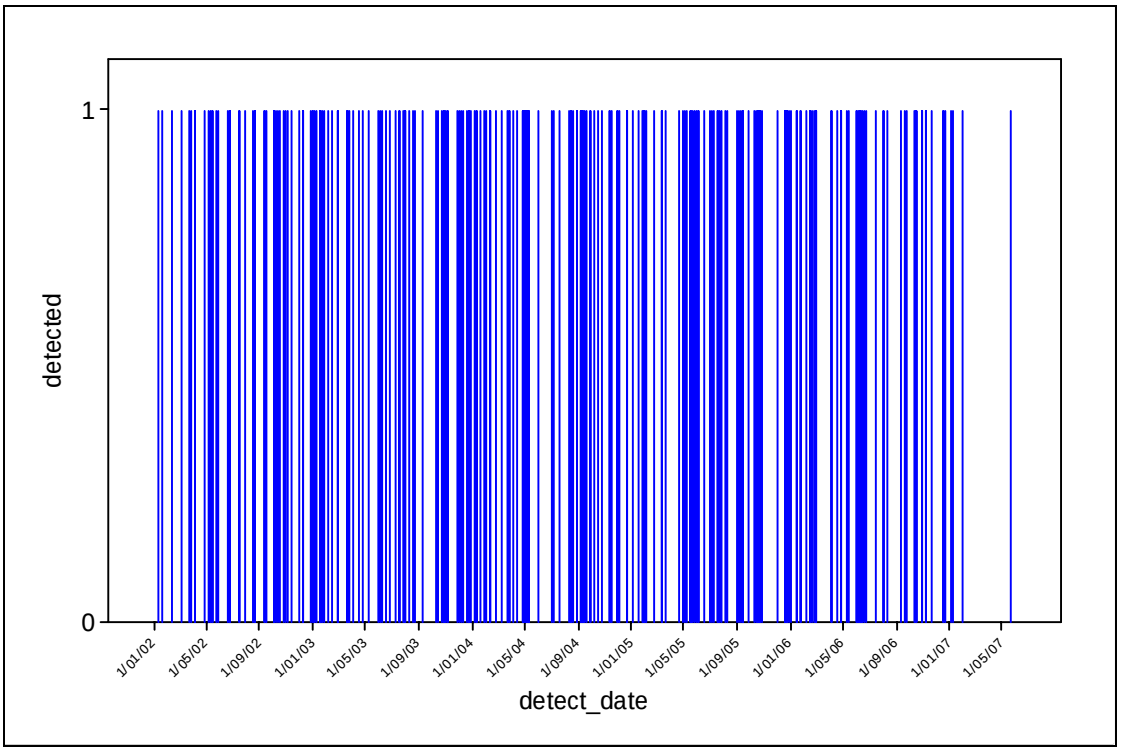

Figure 20. Time sequence of detection of quarantine threats.

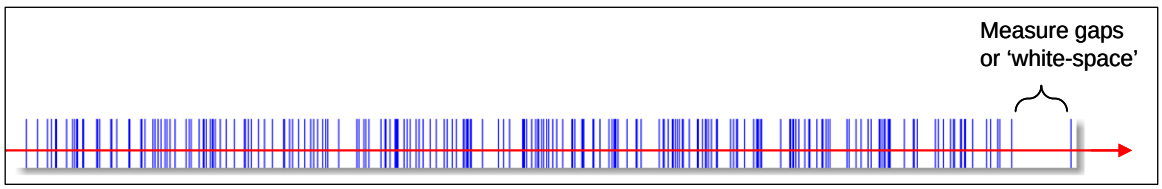

Figure 21. Pattern of inter-arrival times as measured by the 'white space' between blue lines.

It is clear from Figure 22 that these data are highly skewed and that the normal distribution is not an appropriate probability model.

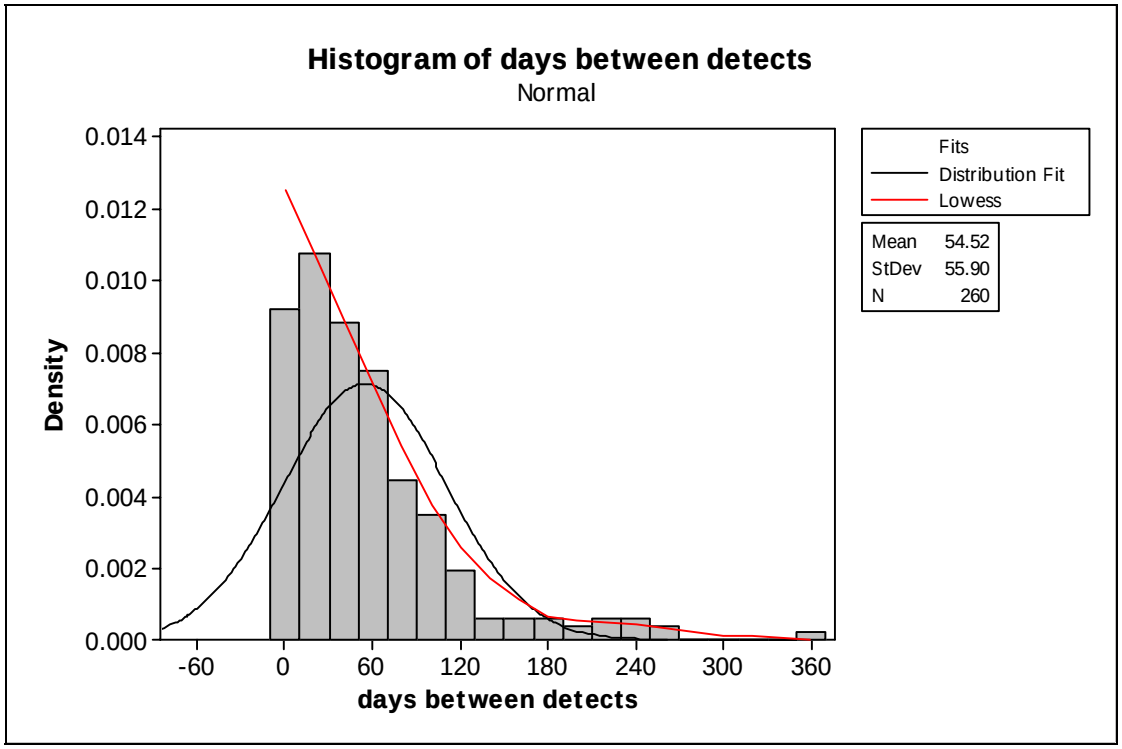

Figure 22. Histogram of inter-arrival times with smoothed version (red line) and theoretical normal distribution (black line) overlaid. The normal distribution provides a poor description of this data (evidenced by the both the shape and probability mass associated with negative values of days between detects).

The smoothed histogram in Figure 22 suggests a 'J-shaped' probability model is more appropriate. One such model is the *negative exponential* probability distribution. This choice is also supported by statistical theory which says that if events arrive randomly in time according to a *Poisson* probability model with an average rate of arrival of λ per unit time, then the distribution of the inter-arrival time is negative exponential with parameter λ. The probability density function (*pdf*) for the negative exponential is given by Equation 1 and the corresponding cumulative distribution function *(cdf)* is given by Equation 2.

$$f_X(x) = \lambda e^{-\lambda x}, \quad \lambda, x > 0 \tag{1}$$

$$F_X(x) = 1 - e^{-\lambda x}, \quad \lambda, x > 0 \tag{2}$$

For this distribution, the mean is $1/\lambda$ and the variance is $1/\lambda^2$. Notice, that this immediately implies that the variance increases/decreases with an increasing

/decreasing mean – in contradiction to many 'conventional' statistical techniques which assume constant variance.

One simple way of estimating the parameter λ is to equate the theoretical and sample means. In this case, we have $1/\lambda = 54.42$ and hence our estimate is $\hat{\lambda} = 1/_{54.42} = 0.1838$. A plot of the histogram of the data with a negative exponential distribution having $\hat{\lambda} = 0.1838$ overlaid is shown in Figure 23. The adequacy of this fit is more readily seen by comparing the empirical and theoretical *cumulative distribution functions* (Figure 24).

The 'false-triggering' due to the non-normality of the data is evident in the I-Chart[3] of Figure 25. There are two ways of over-coming this. The first is to modify the control chart itself to account for the non-normality. The second approach is to *transform* the data so that the transformed data are normally distributed (or approximately so) and then apply standard control charting techniques to the transformed data. We consider each of these approaches in turn.

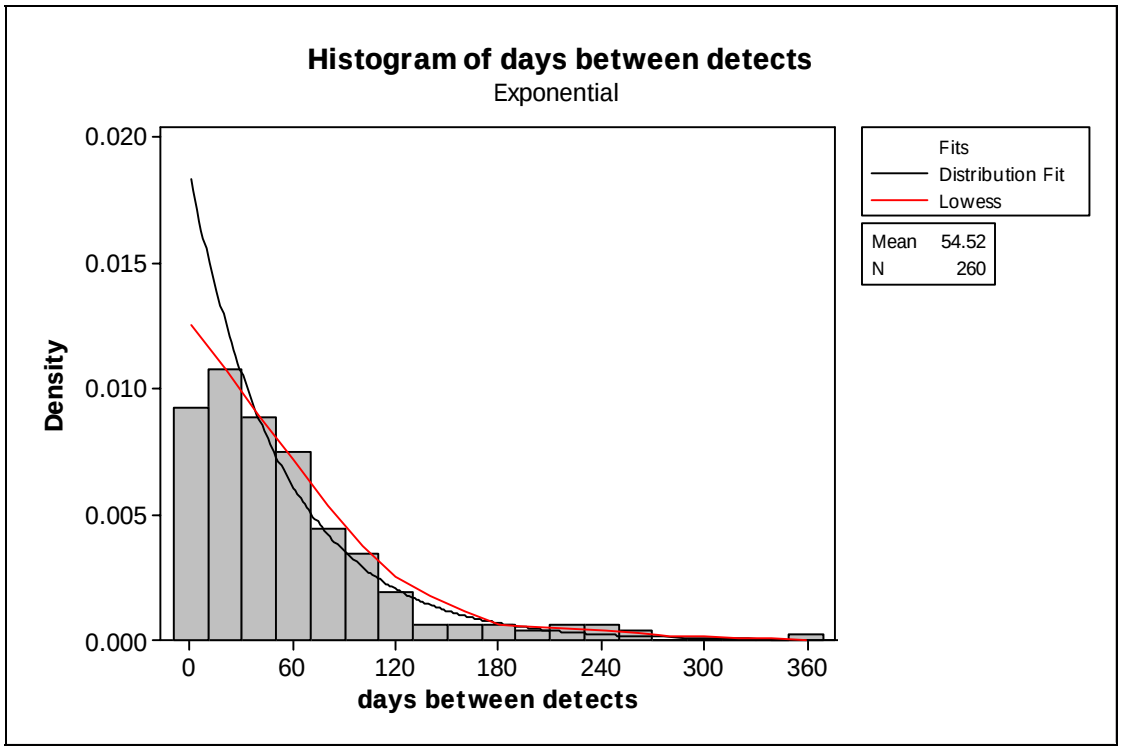

Figure 23. Histogram of days between detects. Smoothed histogram indicated by red curve, theoretical exponential distribution depicted by black curve.

[3] An "I-Chart" is simply a control chart for *individual* observations ie. ungrouped data.

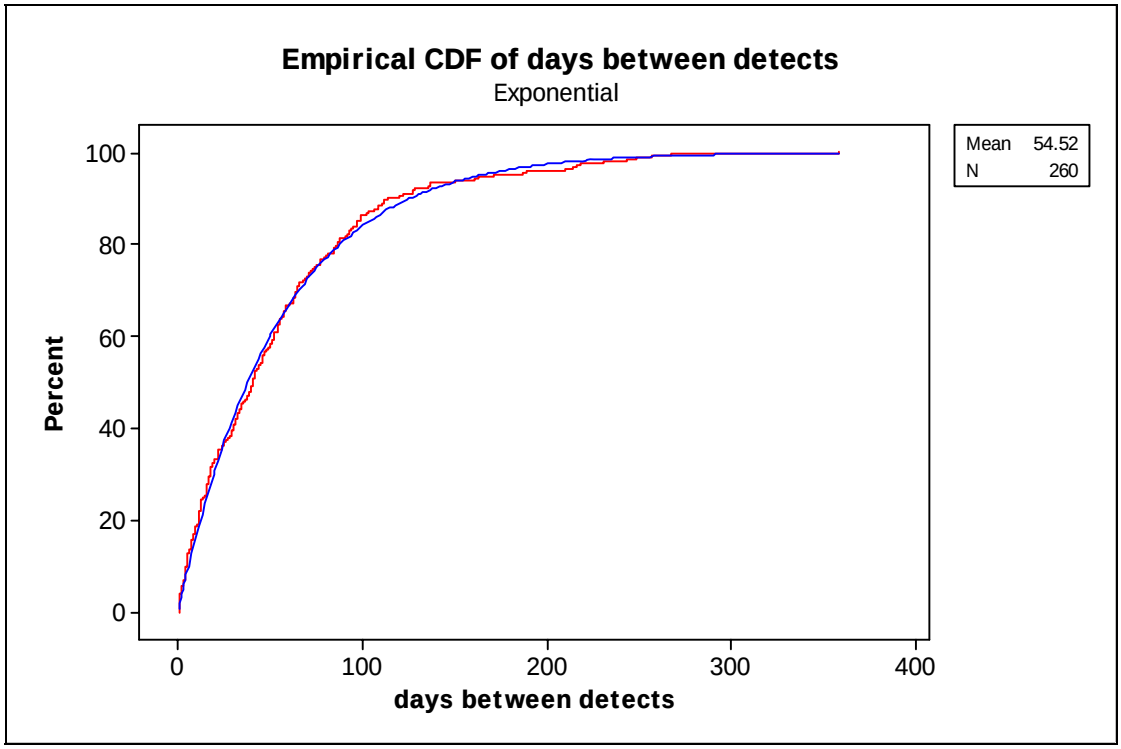

Figure 24. Empirical cdf for days between detects (red curve) and theoretical exponentail cdf (blue curve).

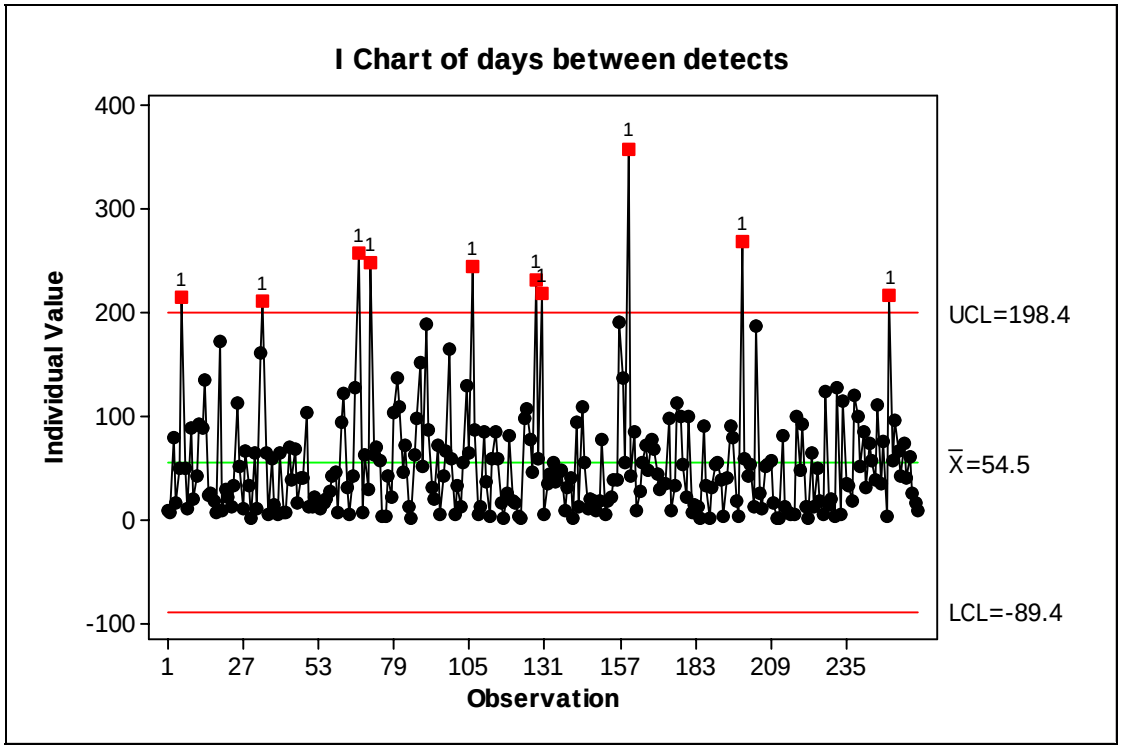

Figure 25. Chart of individual values of days between detects (I-Chart).

Transformations to Normality

In section 3 we looked briefly at the issue of statistical significance. The discussion focussed on a normally distributed random variable. Violations of the normality assumption will tend to invalidate the results of any statistical procedure which invokes this assumption[4]. One way to overcome this problem is to identify a mathematical transformation of the data so that the transformed data are normally distributed or approximately so. There is a tendency among practitioners to spend an inordinate amount of time on the identification of the 'best' transformation (best in the sense that the resulting data are most nearly normal). This is often wasted effort since many statistical procedures (including control charting) are relatively robust to mild to moderate departures from normality. The over-riding objective should be to identify a *simple* mathematical transformation that at least results in data that is approximately

[4] The severity of the violation cannot be anticipated in advance since it is a function of the degree to which the assumption is violated, the manner in which it is violated, and the robustness of the statistical procedure to such violations.

symmetrical. The identification process can be by trial and error or by some 'automated' procedure. An example of the latter is the so-called *Box-Cox* family of transformations. The idea is simple enough: we wish to find the value of the transformation parameter λ (not to be confused with the λ in equations 1 and 2) so that data transformed according to

$$Y = \begin{cases} \dfrac{X^{\lambda}}{\lambda} & \lambda \neq 0 \\ \ln(X) & \lambda = 0 \end{cases}$$

exhibit a greater degree of normality than the untransformed data (the *X*s). Having found this λ we proceed to work with the transformed values, Y. MINITAB and other software packages simplify the task of determining λ for a given data set. The 'optimal' λ is identified as the abscissa value at the minimum on a Box-Cox 'profile plot' (Figure 26) – in this case we find $\lambda = 0.24$. With this value of λ we then transform the data and then use control charting methods on the *transformed* data.

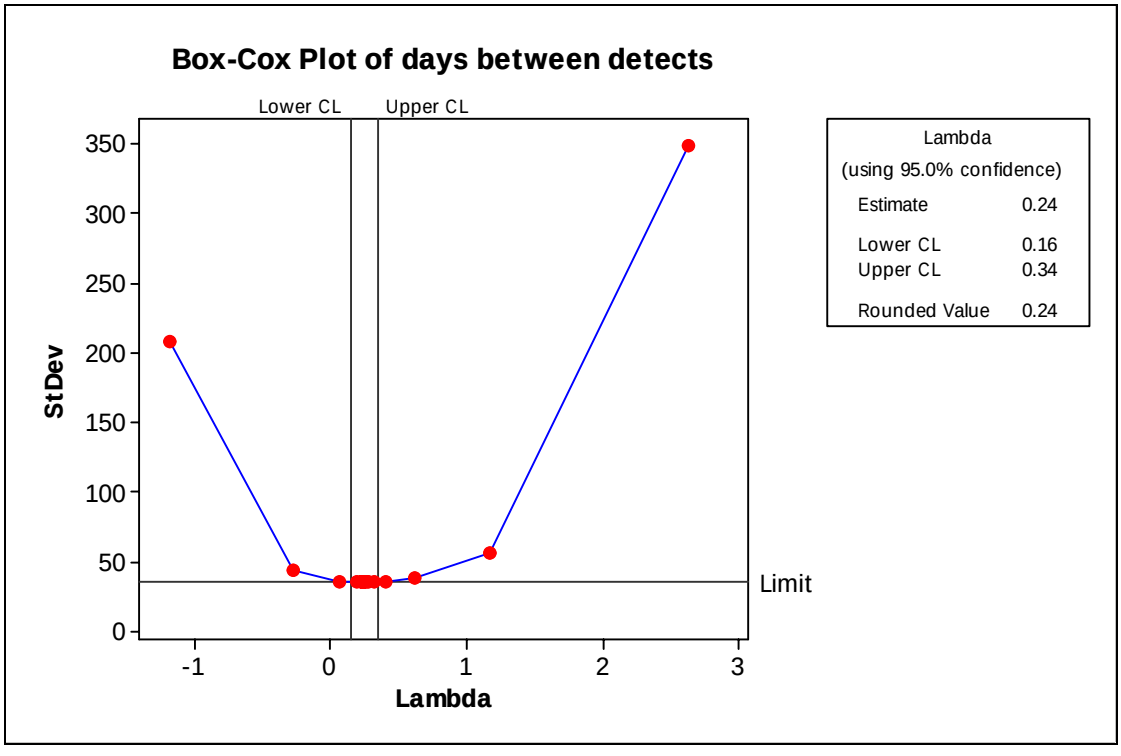

Figure 26. Box-Cox profile plot for the days between detects. Optimal lambda is 0.24.

The effectiveness of the transformation is evident from the histogram of the transformed data (Figure 27).

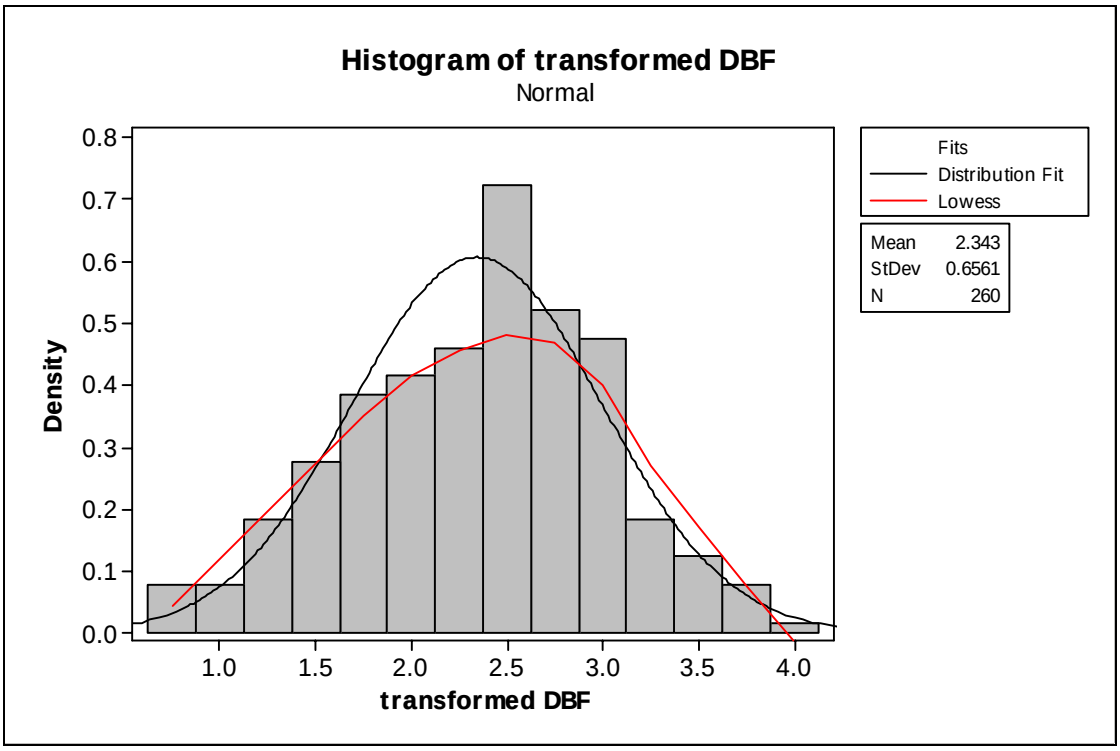

Figure 27. Histogram of transformed days between detection with smoothed version (red curve) and theoretical normal (black curve) overlaid.

To see how the transformation process works, consider setting an upper control limit on the days between detects such that this limit would only be exceeded 10% of the time when there has been no change in the underlying response-generating mechanism. In other words, on the *transformed scale*, we wish to find that value y^* which satisfies the following $P\left[Y > y^*\right] = 0.10$. From Figure 27, we see that the transformed data (Y) are well described by a normal distribution having mean 2.343 and standard deviation 0.6561. Either using tables of the normal distribution or computer software (as in Figure 28) we determine that $y^* = 3.18$. We can 'back-transform' this y^* to determine an equivalent x^* on the *untransformed* scale by noting that

$$P\left[Y > y^*\right] = P\left[X^{\lambda} > y^*\right] = P\left[X > y^{*1/\lambda}\right]$$

That is, $x^* = y^{*1/\lambda}$. With $y^* = 3.18$ and $\lambda = 0.24$ we obtain $x^* = 124$ which compares favourably to the theoretical result obtained directly from the negative exponential distribution (Figure 29).

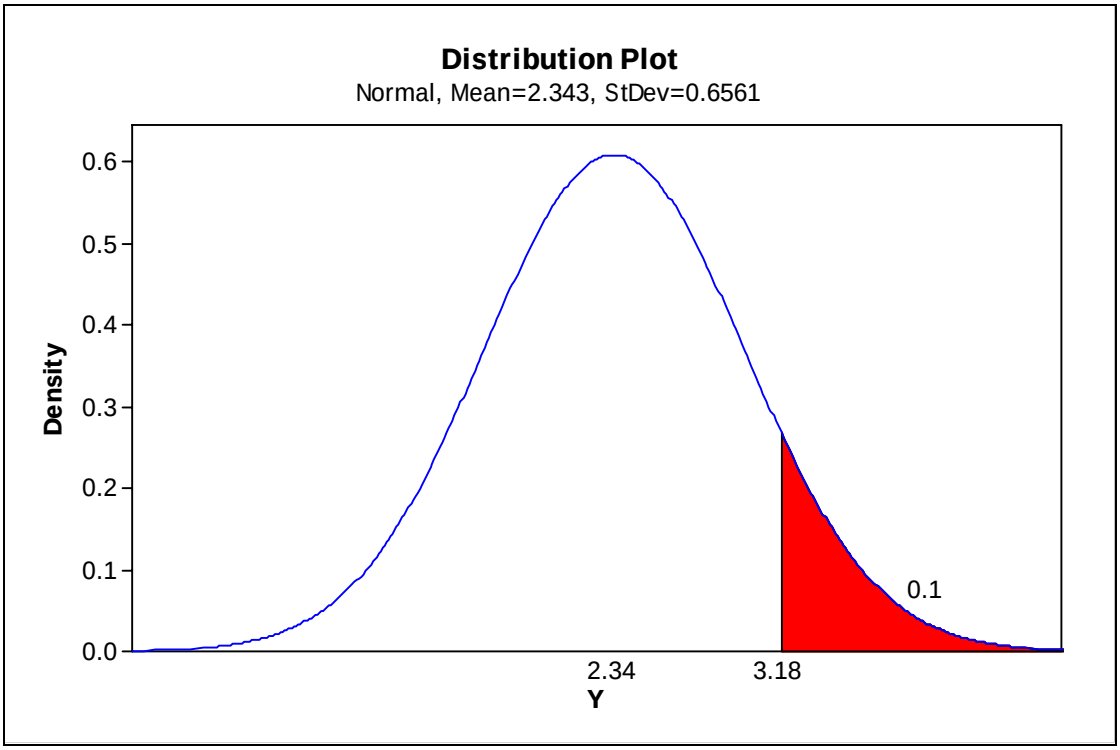

Figure 28. Fitted normal distribution to transformed days between detects with upper 10% point indicated.

Finally, we plot the I-Chart for the transformed days between detects and note that there is now no 'out-of-control' situation indicated (Figure 30).

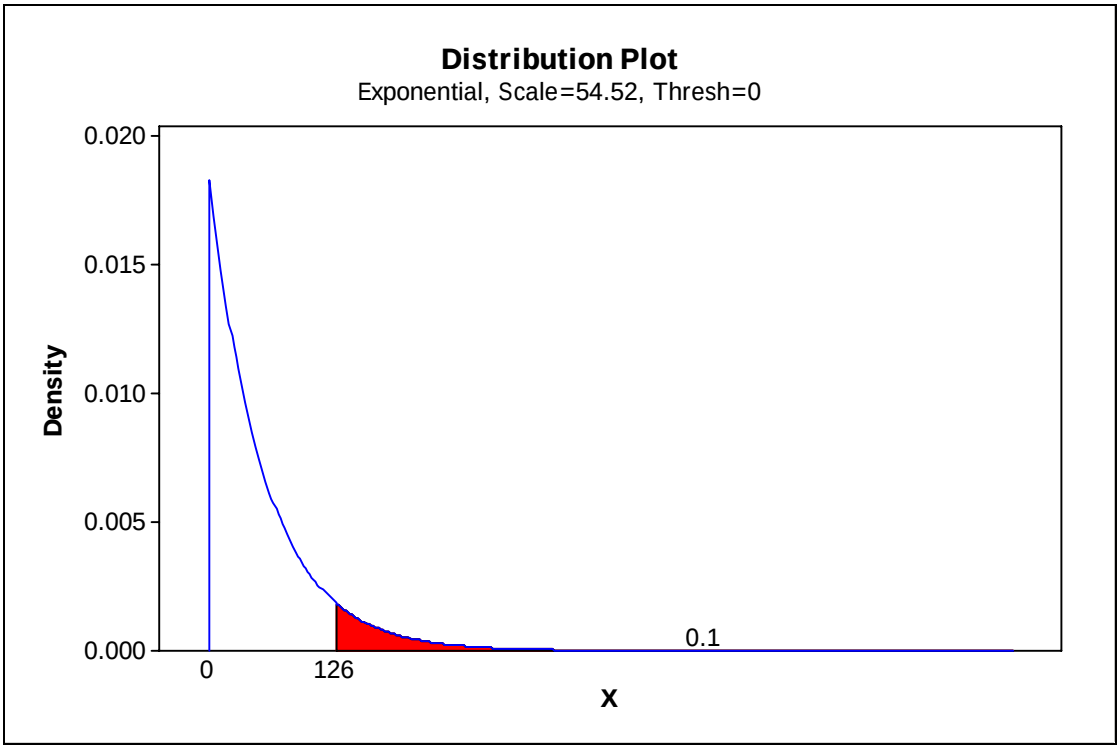

Figure 29. Theoretical negative exponential distribution for untransformed days between detects and upper 10% point indicated.

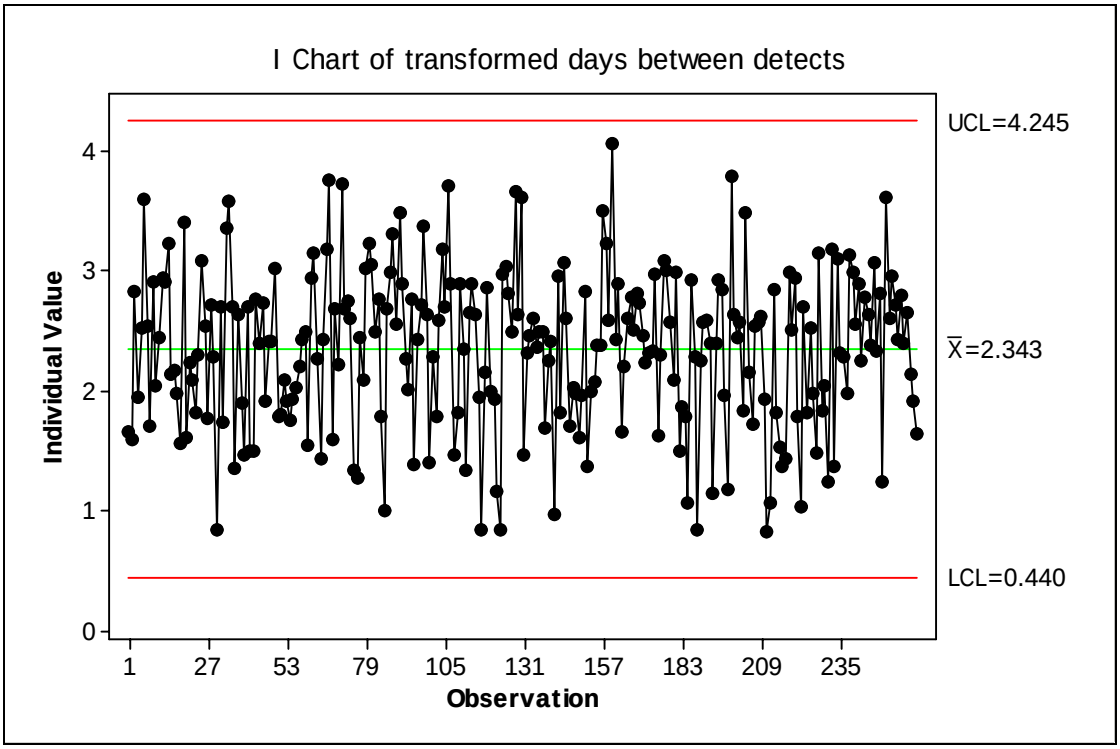

Figure 30. I-Chart for transformed days between detects.

Control chart for time-between-events

Rather than transform the data as described in the preceeding section, alternative methods have been developed which modify existing control charts for use with untransformed (and non-normal) data. Radaelli (1998) describes procedures for setting control limits for both one and two-sided control charts for inter-arrival times. Only the one-sided case is considered here since we are generally only interested in tracking significant deviations in one direction (eg. where the inter-arrival time between quarantine risk detects is decreasing).

Let X_i be the i^{th} inter-arrival time. An 'out-of-control' situation is declared if $X_i < T_L$ in the case of *decreasing* inter-arrival times (ie. increasing counts) or $X_i > T_U$ in the case of *increasing* inter-arrival times (ie. decreasing counts) where T_L and T_U are suitably chosen positive constants. Suppose that an 'in-control' situation corresponds to a mean inter-arrival time of λ_0^{-1} (where λ is the parameter in equation 1) then using Equation 2, it can be determined that

$$P\left[X_i < T_L \,|\, \lambda = \lambda_0 \right] = 1 - e^{-\lambda_0 T_L} \tag{3}$$

$$P\left[X_i > T_U \,|\, \lambda = \lambda_0 \right] = e^{-\lambda_0 T_U} \tag{4}$$

Equations 3 and 4 are analogous to the Type I error in a hypothesis test: it's the probability of a false-positive. As in statistical hypothesis testing, the Type I error-rate (α) is set to be some arbitrarily small value (eg. α =0.05). Thus, the upper and lower control limits can be determined by setting Equations 3 and 4 equal to α and solving for either T_L or T_U . Thus we have:

$$T_L = -\lambda_0^{-1} \ln\left(1-\alpha\right) \tag{5}$$

$$T_U = -\lambda_0^{-1} \ln\left(\alpha\right) \tag{6}$$

In addition to having a low α, we also require our control chart to correctly signal an important deviation from 'in-control' conditions. Suppose we wish to detect a change from λ_0 to λ_1 with some high probability, $(1-\beta)$ where $\lambda_1 = k\lambda_0$ ($k>1$ for a one-sided lower chart; $k<1$ for a one-sided upper chart). That is:

$$P\left[X_i < T_L \,\middle|\, \lambda = \lambda_1\right] = 1 - e^{-k\lambda_0 T_L} = (1-\beta) \qquad \text{(lower chart)} \qquad (7)$$

$$P\left[X_i > T_U \,\middle|\, \lambda = \lambda_1\right] = e^{-k\lambda_0 T_U} = (1-\beta) \quad \text{(upper chart)} \qquad (8)$$

Substituting T_L and T_U in Equations 7 and 8 respectively, we obtain:

$$(1-\beta) = 1 - e^{k\ln(1-\alpha)} \qquad \text{(lower chart)} \qquad (9)$$

$$(1-\beta) = e^{k\ln(\alpha)} \qquad \text{(upper chart)} \qquad (10)$$

The performance characteristics for both lower and upper one-sided charts are shown in Figures 31 and 32. Both of these figures show that the ability to detect even relatively large shifts (eg. a doubling or halving) in the mean inter-arrival time is low (typically less than 0.2) for values of α less than 0.1. For example, using a 10% level of significance (ie. alpha =0.10), the one-sided chart of Figure 31 suggests that there is a less than 20% chance of detecting a doubling (ie. k=2) of the mean arrival rate (or a halving of the inter-arrival time).

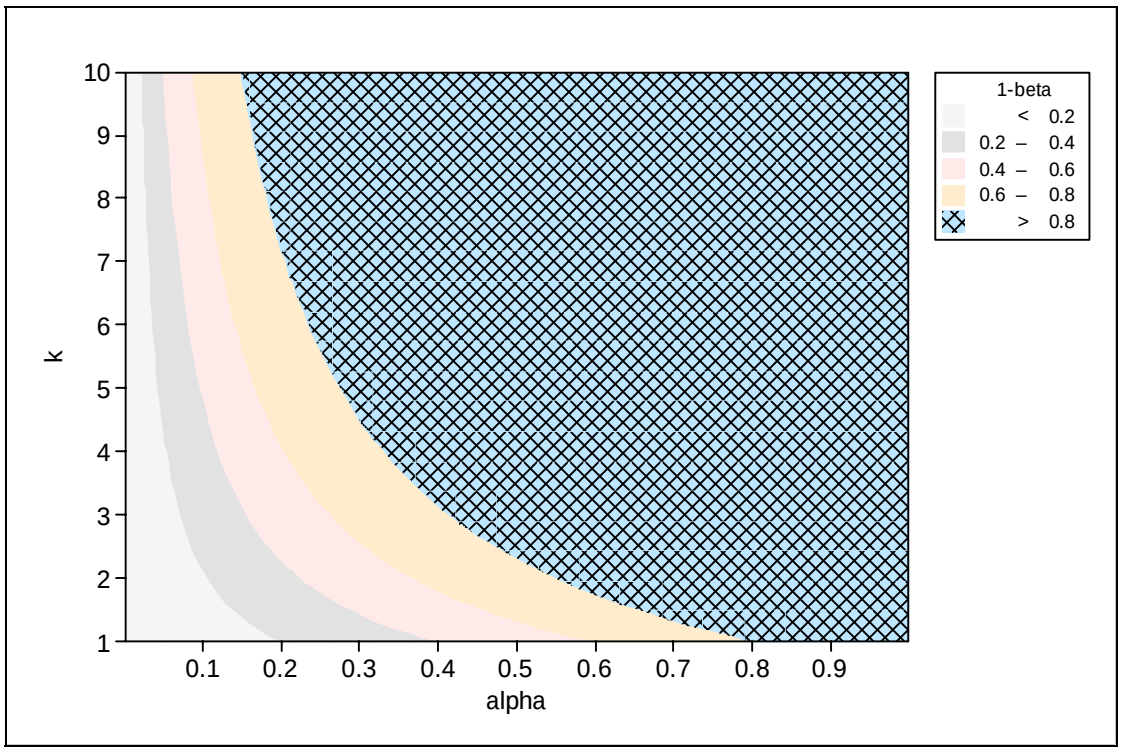

Figure 31. Performance characteristics (as measured by equation 9) for a one-sided, lower control chart.

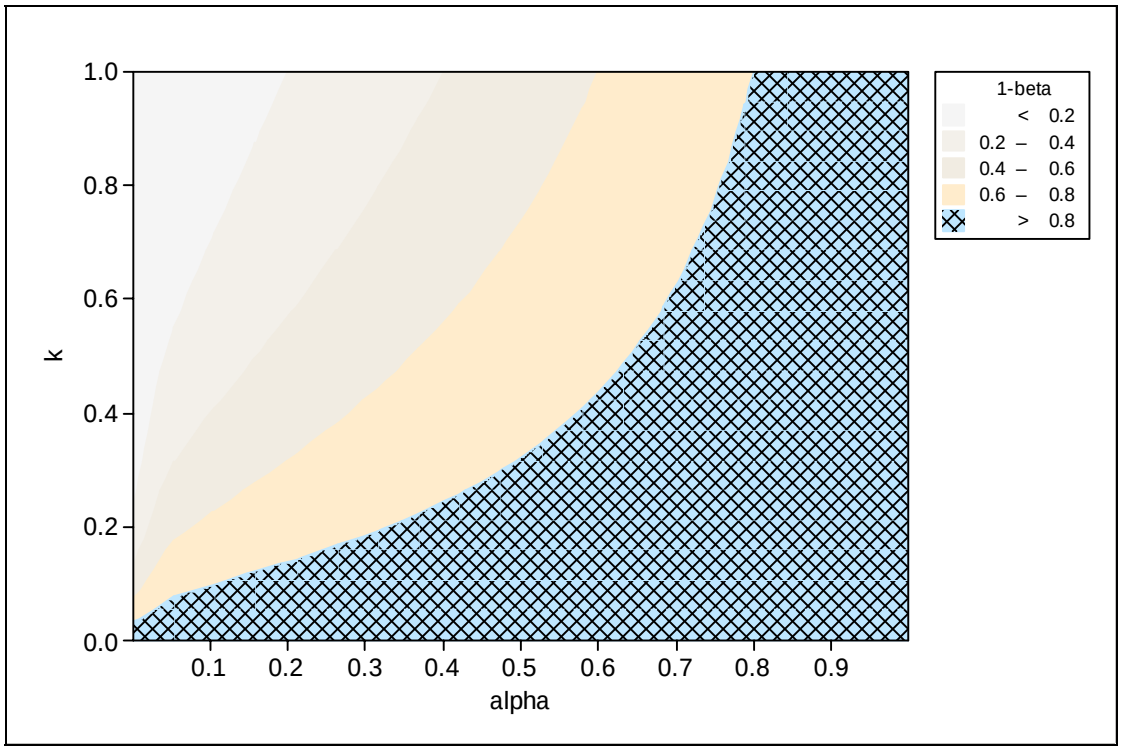

Figure 32. Performance characteristics (as measured by equation 10) for a one-sided, upper control chart.

6. DISCUSSION

This paper provides a review of basic statistical concepts as well as introducing some common control-charting techniques that have been advocated elsewhere (Carpenter 2001, Commonwealth of Australia 2002) as being particularly suited to monitoring for temporal trends and aberrations in bio-security related applications. Control charts are particularly well suited to the visualisation and assessing of moderate to large volumes of time-based data and as such would be expected to have greater utility for container inspection regimes say, than for detecting the occurrence (in space) of an invasive species. Control charts need to be viewed as just one method in a tool-kit of available techniques which can potentially assist field officers and quarantine risk assessors in identifying 'unusual' or 'aberrant' trends. For events having very low probabilities of occurrence (eg. exotic disease outbreak) the monitoring of 'time between outbreaks' is a potentially more useful quantity to be charting although as shown in this report, the statistical power (ability to correctly

identify real ‘shifts’ in the mean time between events) of current charting techniques is relatively low.

Since the events of September 2001, there has been a substantial research push in the area of ‘syndromic surveillence’ with the accompanying development of new approaches and methods to detecting unusual patterns in a space-time continuum. Some of these techniques would appear to have direct applicability to the activities of Biosecurity Australia and AQIS and will be the subject of a separate report.

7. REFERENCES

Benneyan, J.C. (2001a) Number-Between g-type statistical quality control charts for monitoring adverse events. *Health Care Management Science*, **4**, 305-318.

Benneyan, J.C. (2001b) Performance of Number-Between g-type statistical quality control charts for monitoring adverse events. *Health Care Management Science*, **4**, 319-336.

Carpenter, T.E. (2001) Methods to investigate spatial and temporal clustering in veterinary epidemiology. *Preventative Veterinary Medicine*, **48**, 303-320.

Commonwealth of Australia (2002) Meat hygiene assessment: Objective methods for the monitoring of processes and products. 2nd. Edition. Canberra, Australia. http://www.daff.gov.au/corporate_docs/publications/pdf/quarantine/mid/msqa2nedition.pdf

Dooley, K., and R. Flor (1998), Perceptions of Success and Failure in Total Quality Management Initiatives, *J. of Quality Management*, **3(2)**: 157-174.

Hall, J.A. and Golding, L. (1998) Standard methods for whole effluent toxicity testing: Development and Application. NIWA Client report MfE80205, National Institute of Water and Atmospheric Research, Hamilton, New Zealand.

Korth, W. (1997) Determination and application of recovery factors in chemical residue testing of food products by Australian laboratories associated with the national residue survey. http://www.daff.gov.au/corporate_docs/publications/pdf/animalplanthealth/nrs/recovery_factors_korth_1997.pdf

Kulldorff, M. (1997) A spatial scan statistic. *Communications in Statistics: Theory and Methods*, **26**:1481-1496.

Kulldorff, M. (2001) Prospective time-periodic geographical disease surveillance using a scan statistic. *Journal of the Royal Statistical Society* (Series A), **164**, 61-72.

Mostashari, F. and Hartman, J. (2003) Syndromic Surveillance: a Local Perspective. *Journal of Urban Health: Bulletin of the New York Academy of Medicine* **80(2)**, Supplement 1.

Radaelli, G. (1998) Planning time-between-events Shewhart control charts. *Total Quality Management*, **9(1)**, 133-140.

Stark, K.D.C., Regula, G., Hernandez, J., Knopf, L., Fuchs, K., Morris, R.S. and Davies, P. (2006) Concepts for risk-based surveillance in the field of veterinary medicine and veterinary public health: Review of current approaches BMC Health

Services Research **6(20).** Available at http://www.biomedcentral.com/1472-6963/6/20

Stephens. M.A. (1974) EDF Statistics for Goodness of Fit and Some Comparisons, *Journal of the American Statistical Association*, **69(347)**, 730-737

8. APPENDIX : TIME BETWEEN DETECTS DATA

8.2	10.9	15.1	103.5	63.4	4.9	190.0	13.8	56.5	33.4
7.1	65.4	10.3	136.8	243.5	33.3	136.9	11.3	15.9	31.9
78.4	31.4	15.8	108.1	86.4	43.4	53.8	1.2	0.4	17.5
16.4	0.5	19.3	45.4	4.9	54.5	357.9	90.3	1.3	119.9
48.7	64.7	27.3	70.6	12.4	36.3	41.5	31.4	80.4	98.7
214.0	10.2	40.9	11.4	84.8	45.3	85.1	0.5	12.3	51.5
49.2	160.7	45.9	1.0	35.6	46.2	8.4	30.0	5.9	84.9
9.4	209.8	6.2	62.7	3.3	8.9	27.2	52.3	3.6	29.5
87.2	63.7	92.8	97.5	58.7	29.6	54.4	53.9	4.5	72.6
19.9	3.5	122.0	150.8	84.8	40.3	72.0	38.7	99.1	57.3
41.8	58.4	30.7	51.1	57.9	0.8	47.1	1.7	47.4	37.5
91.8	14.6	4.5	188.7	16.2	93.6	76.7	39.2	92.4	110.1
87.8	4.9	41.2	86.1	0.5	12.0	67.6	89.2	11.2	34.6
135.2	63.7	126.8	30.5	24.8	108.7	43.2	78.9	1.1	75.5
23.8	5.4	257.2	18.7	81.0	55.2	28.7	16.5	64.4	2.4
25.6	5.5	7.1	71.1	18.0	9.2	34.1	2.0	12.1	216.5
17.1	70.2	62.4	4.0	15.8	18.9	34.6	267.7	48.4	55.5
6.3	38.5	28.0	40.8	1.8	17.3	96.9	57.4	17.1	94.5
170.6	67.7	248.0	65.3	0.5	7.2	7.7	42.2	5.1	65.1
7.4	15.3	62.6	163.1	97.0	16.5	32.4	51.8	122.4	41.6
28.9	40.2	68.5	57.7	106.2	77.0	111.6	12.5	12.8	73.9
21.8	39.8	55.5	4.1	76.9	3.7	99.6	186.6	19.8	39.1
12.0	102.0	3.3	32.1	45.4	18.0	52.2	24.9	2.4	60.1
32.6	11.5	2.8	11.3	230.3	21.3	22.0	9.7	127.0	24.1
111.7	11.8	41.9	53.7	58.4	37.4	99.0	50.0	3.8	15.1
50.1	21.6	21.9	128.6	218.1	38.1	5.5	52.1	113.9	8.1

Table entries are days