



An Excel Add-In for Statistical Process Control Charts

Samuel E. Buttrey
Naval Postgraduate School

Abstract

Statistical process control (SPC) describes a widely-used set of approaches used to detect shifts in processes in, for example, manufacturing. Among these are “control charts”. Control charts and other SPC techniques have been in use since at least the 1950s, and, because they are comparatively unsophisticated, are often used by management or operations personnel without formal statistical training. These personnel will often have experience with the popular spreadsheet program **Excel**, but may have less training on a mainstream statistical package. Base **Excel** does not provide the ability to draw control charts directly, although add-ins for that purpose are available for purchase.

We present a free add-in for **Excel** that draws the most common sorts of control charts. It follows the development of the textbook of [Montgomery \(2005\)](#), so it may be well-suited for instructional purposes.

Keywords: statistical process control, SPC, control charts, **Excel**.

1. The Excel spreadsheet program and control charts

1.1. Introduction

This paper describes a free, open-source add-in for the popular **Excel** spreadsheet program which draws control charts. The add-in should allow non-technical users to produce charts of interest; since it follows many of the directives in [Montgomery’s \(2005\)](#) textbook, it may also be useful for instruction.

1.2. Control charts

A control chart is a graph with time on the X -axis and whose Y -axis shows the output of a particular process. The chart, together with control limits established from probability

theory, provide a visual clue as to the stability of the process. When a point on the chart lands outside the control limits, or when certain anomalous behaviours are observed, the operators can be alerted to investigate or perform other actions. Although control charts can be used in many contexts, it will be convenient to imagine the output as representing measurements of quality in manufacturing.

1.3. Types of control charts

Our add-in draws eight popular control charts. Charts are described as being relevant to *attribute* (categorical) data or *variable* (continuous) data. For attribute data we draw the:

- p charts, for the proportion of defective items in binomial data;
- np charts, for the number of defective items in binomial data;
- c charts, for the number of defective items in a fixed-size sample from Poisson data; and
- u charts, for the rate of production of defective items in Poisson data.

For variable data, we draw the:

- \bar{X} charts, for the average of a set of continuous measurements;
- R charts, for the range of a set of continuous measurements;
- S charts, for the standard deviation of a set of continuous measurements; and
- X charts, for individual measurements.

In each case we provide appropriate control limits. We also allow the user to identify observations that fall outside control limits, and to re-draw the graph excluding those points. For attribute data we implement a set of so-called “trend rules,” which call for observations to be identified as worthy of further examination if they form patterns obeying certain criteria. For example, one rule says that a pattern is suspicious if two of three consecutive points fall in the region between $\bar{x} + s$ and $\bar{x} + 2s$. Our add-in allows the user to identify when some or all of these patterns are present.

1.4. Implementation

The add-in is implemented in Visual Basic Applications (see, for example, [Walkenbach 1999](#)), so no compiled code is necessary. We believe that will make the add-in easy to install and use, even for users whose machines are constrained by, for example, security considerations. Implementation in **Excel** makes the add-in accessible to that program’s huge user base and means that the familiar spreadsheet interface, rather than specialized software, can be used in courses on quality assurance. Our use of **Excel**’s well-developed drawing functions produces graphs with a look that should hopefully be familiar to many users.

1.5. Installation

Excel versions prior to 2007

A user with a version of **Excel** prior to the 2007 version starts the installation process by downloading the add-in file to a location on his or her computer. Then he or she selects **Tools | Add-Ins**, clicks on **Browse** to locate the addin, and then presses **Ok** to complete the installation. Subsequently, a **Control Charts** item will appear under the **Tools** menu, and selecting that entry produces the start-up screen.

Excel 2007

Installation of add-ins in **Excel 2007** also starts with the user downloading the add-in file to some known location on his or her computer. He or she should then click on the **Office** button in the top-left corner of the screen. The choice of **Excel Options** at the bottom of that window produces the **Excel Options** page, from which the choice of **Add-Ins** can be made in the left panel. With **Excel Add-Ins** selected in the **Manage** drop-down list at the bottom of the page, the user clicks **Go** to activate the **Add-In** window, then **Browse** to locate the add-in file, and then **Ok** to complete the installation. In this version of **Excel** the add-in is made available under the **Add-In** tab, where an entry named **Control Charts** will be visible under **Menu Commands**. Clicking that entry will produce the start-up screen.

1.6. Organization

The next section describes the user interface and the way that the user should organize his or her data in order to use the add-in. Sections 3 and 4 give brief descriptions of the charts, for users unfamiliar with the details of their construction. Finally, Section 5 gives conclusions and directions for further development.

2. The user interface

2.1. Example: Attribute data

Figure 1 shows the window that appears when the add-in is invoked. This window is intended for attribute data; a user who needs a chart for variable data will click the **Variables** button in the top center of the window (see Section 4). The data for an attribute control chart will consist of a set of counts of defective items and a set of sample sizes. The data behind the form is taken from Table 6-7 of [Montgomery \(2005\)](#) and counts the number of defects found in constant-size sets of printed circuit boards. Here the user has requested a c chart.

Data and sample size

The user enters a reference to a column of count data, giving the number of defective units (or defects) produced, into the **Enter Count Data** field, most usually by clicking in that box and then highlighting the relevant area on the spreadsheet with the mouse. If the sample size is constant for all observations, that number can be typed directly into the **Enter sample size** box; alternatively the user can use the mouse to input a column of sample sizes the same

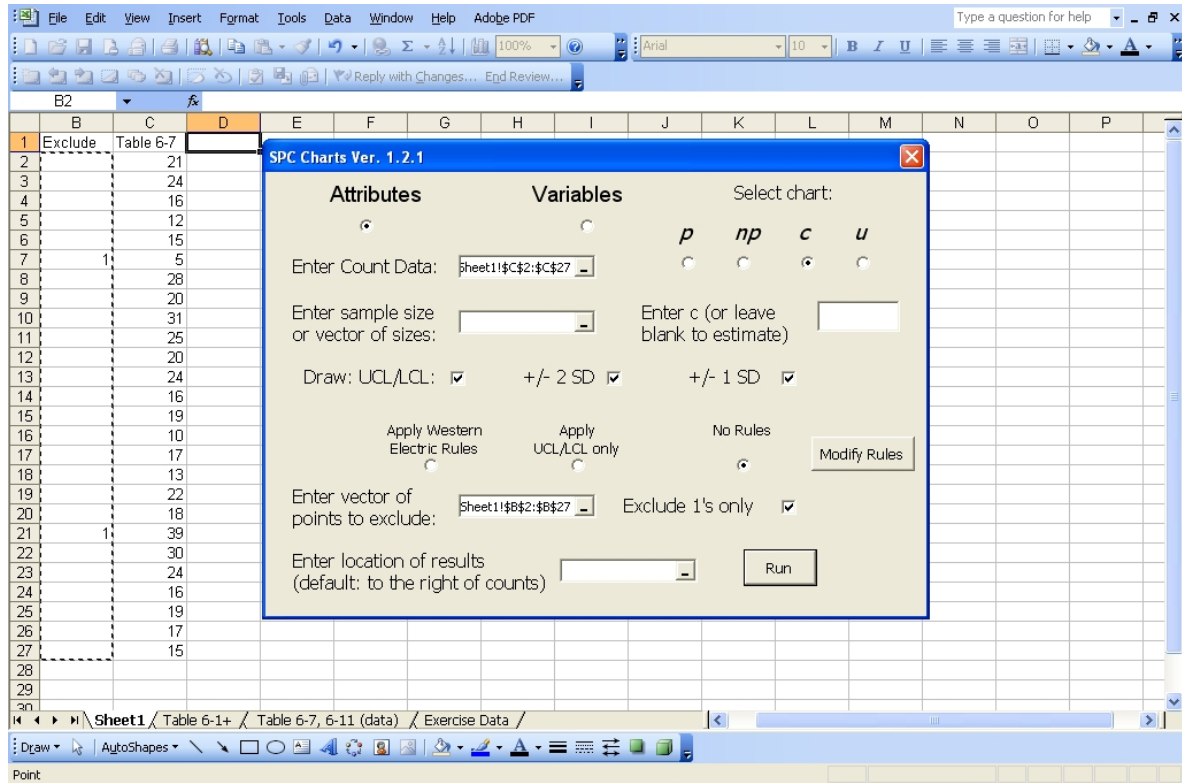


Figure 1: Main window, showing data input for attribute chart.

size as the first column. The c chart measures counts (rather than rates) and therefore acts as if every observation is the count in a sample of size 1 unit. So for that chart (as shown in Figure 1) no sample size need be entered.

The user should generally have the count data be the rightmost column of interest on the sheet, since by default results are placed in a set of columns starting immediately to the right of the counts (see also below).

Choosing the chart

In the top right portion of the window the user selects the sort of chart desired. For p and np charts, the user can also enter a value of the p parameter in the box underneath. For c and u charts the box's label changes (as in Figure 1) to remind the user that in this case it is the c parameter that can be entered. In either case the relevant parameter is estimated from the data if the box is left blank.

Control limits

Across the middle of the window is a set of three checkboxes that allow the user to draw the control limits (UCL/LCL for “upper control limit” and “lower control limit”), lines at ± 2 SD, and lines at ± 1 SD. For attribute data all three of these boxes are checked by default.

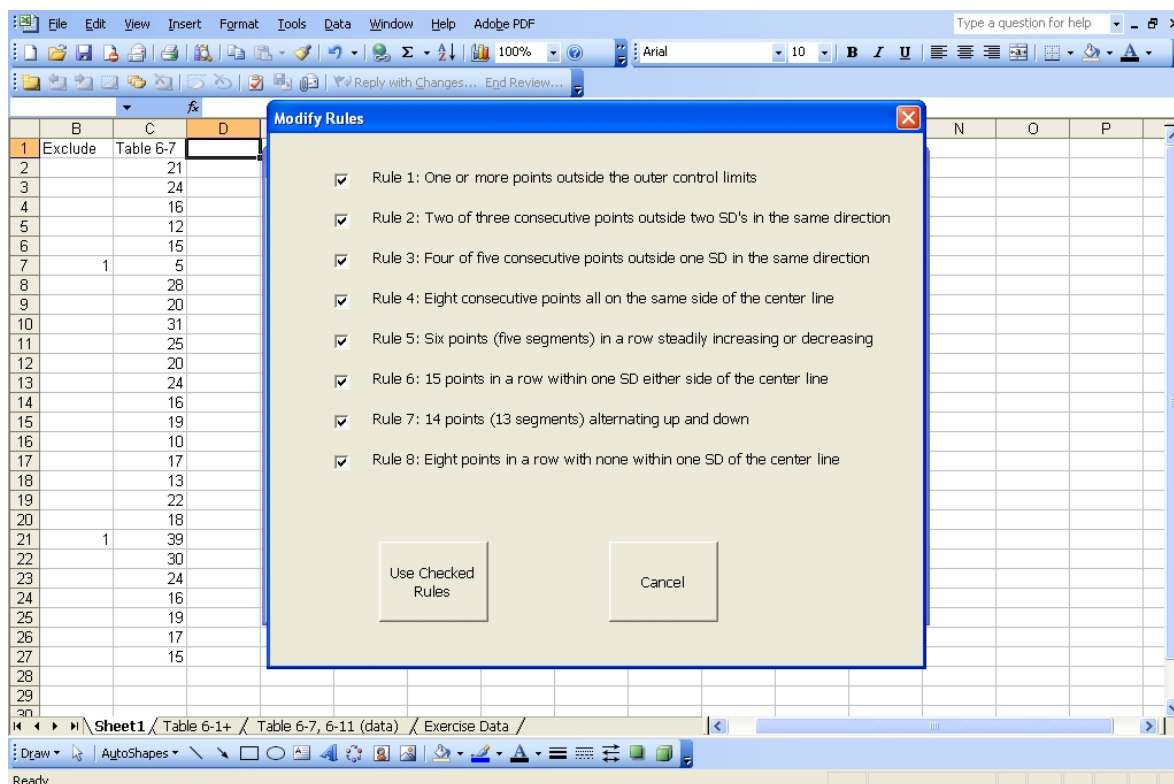


Figure 2: Selecting trend rules.

Trend rules

Beneath the control limit checkboxes is a set of three radio buttons that allow the user to apply different subsets of control rules. The **Apply Western Electric Rules** applies each of a set of eight rules listed in [Montgomery \(2005, p. 167\)](#). (Technically the term “Western Electric rules” more commonly refers just to the first four of these, but we have kept the phrase to evoke the entire set.) When this option is selected the program will “flag” points that produce violations of any of the either rules. When the **UCL/LCL only** choice is selected, only points that fall outside the lower or upper control limits will be flagged, and, of course, **No Rules** means that no points should be flagged. When two plots are produced simultaneously (for example, \bar{X} and R charts), the rules are applied separately.

The **Modify Rules** button brings up a separate window that allows the user to choose a subset of the rules to be applied. [Figure 2](#) shows the **Modify Rules** window. In the current version of the add-in, this allows users to select the set of rules that are to be applied when the **Apply Western Electric rules** button is pressed. In future versions we envision giving the user slightly more specific control over individual rules, allowing him or her to, for example, change the “six” in rule 5 (“Six points...in a row steadily increasing or decreasing”) to “seven”.

Excluding points

When one or more points fall outside a control limit (or violate one of the other rules) it is common to investigate them, to see if the violation can be ascribed to some known

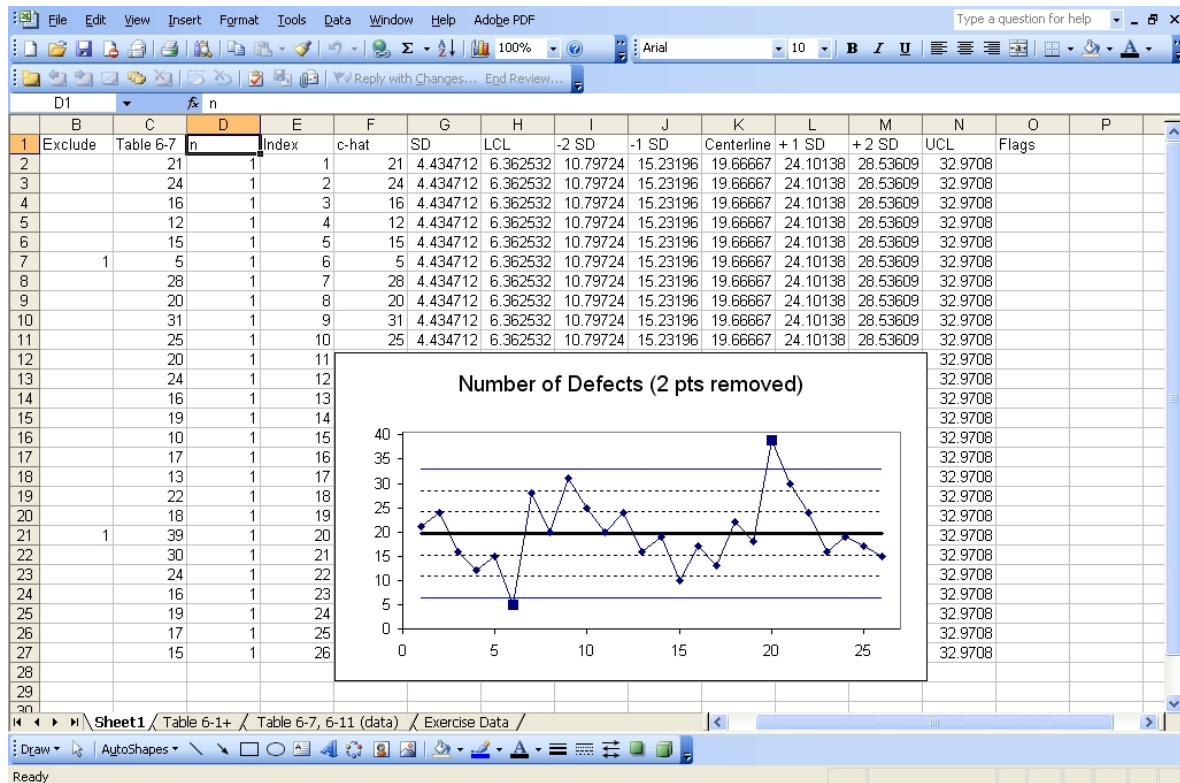


Figure 3: Example of attribute chart with excluded points.

cause. Measurements produced when the process is known to be out of control are often excluded and the chart re-drawn. By entering an additional column of data into the **points to exclude** box, the user can direct that points with entries in that column be excluded from any calculations. Two such points have been excluded in Figure 1. Figure 3 shows the results of pressing Run in Figure 1. Notice that the excluded points still appear on the graph, marked with a special symbol; they are still flagged if they contribute to a violation of a relevant trend rule.

The “1” in **Exclude 1’s only** refers to the control limit violation rule. When a point lands outside the control limits it is flagged with a “1,” since rule number 1 has been violated. When the **Exclude 1’s only** box is checked, the user is asking to exclude only points flagged in that way. Points that violate other rules, and which are therefore flagged with a different number, continue to be included in the computations when the **Exclude 1’s only** box is checked.

In the current version of the add-in, results are placed immediately to the right of the data. Therefore the **points to exclude** column should not be placed immediately to the right of the data; doing that will over-write that column and confuse the algorithm. We recommend placing the column of points to exclude to the left of the data.

Result location

By default, results are presented in a set of columns immediately to the right of the data on the workbook. If a location is entered in the **location of results** box, results will someday

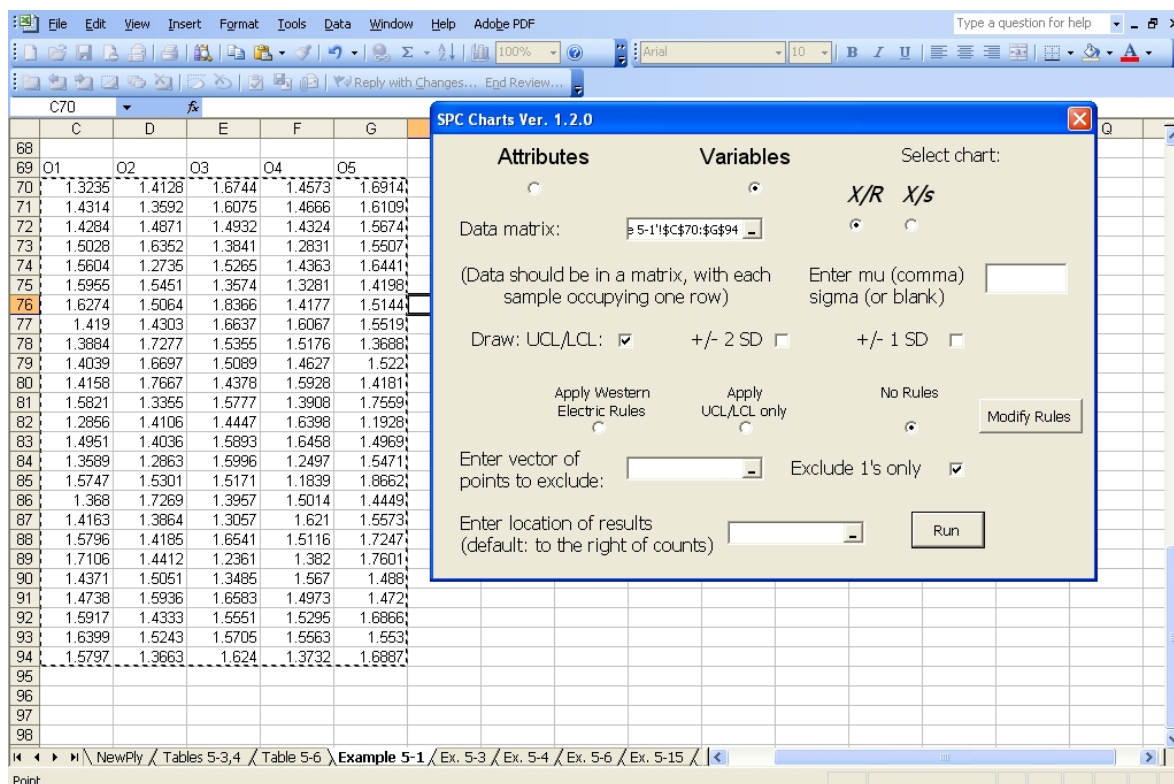


Figure 4: Example of data input for variable data.

be placed in that location instead. In fact, in the current version of the add-in this entry is ignored.

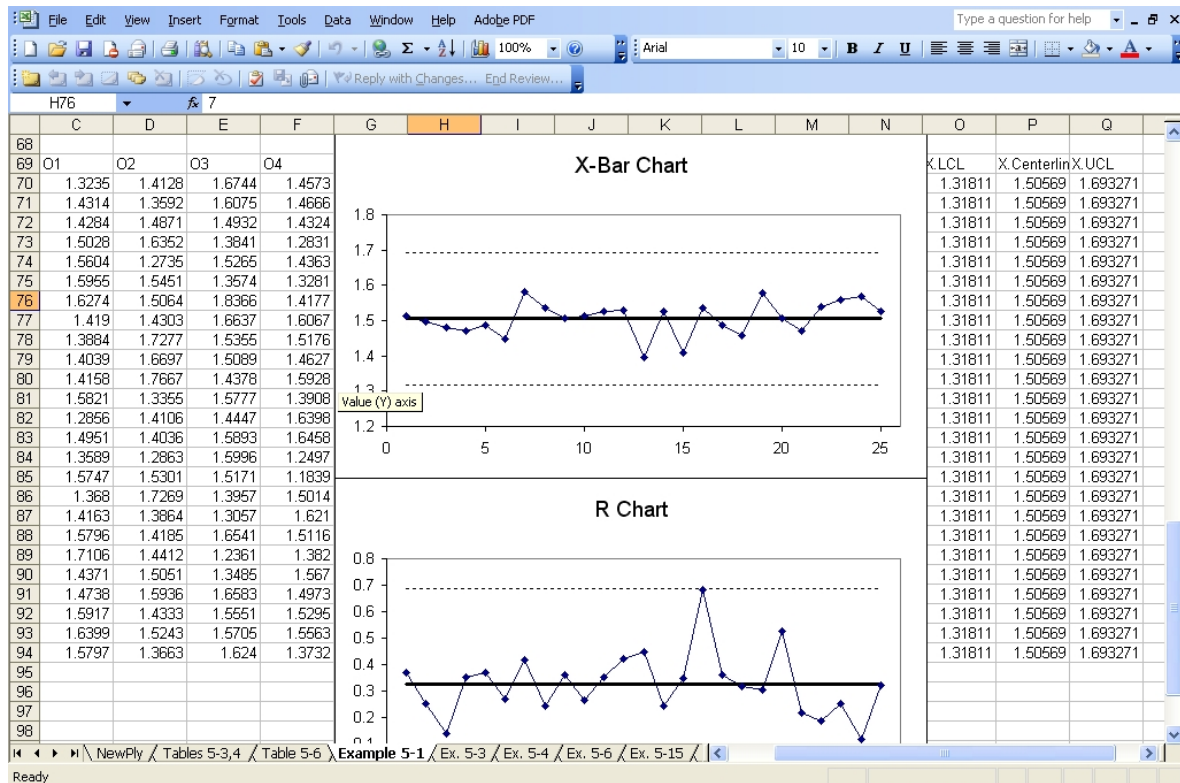
2.2. Variable data

Figure 4 shows the window that allows users to draw control charts for variables. The changes from the attribute case are largely minor. For variable data, we expect a matrix of data, each of whose rows describe a set of observations from a particular run (for example, from a manufactured lot). The number of columns in the data matrix then gives the size of the largest sample in the data; in some rows, some entries may be blank if sample sizes differ.

Since there are two parameters in these charts (μ and σ) the user may enter one or both in the box labeled **Enter mu (comma) sigma (or blank)**. That instruction tells the user to enter both μ and σ separated by a comma. A single number is interpreted as a value of μ , with σ to be estimated; a comma followed by a number, conversely, is taken to be a value of σ , with μ to be estimated.

For variables data, only upper and lower control limits can be drawn, and the Western Electric trend rules cannot be imposed.

Figure 4 shows, on the spreadsheet behind the add-in window, the data in Example 5-1 of Montgomery (2005). For this example, the user has made the default choice of X/R (that is, she has asked to produce the \bar{X} and R charts) and, by leaving the **Enter mu** box blank, has asked that μ and σ be estimated. Figure 5 shows the \bar{X} and R charts produced from the

Figure 5: Example of \bar{X} and R charts for variable data.

set-up shown in Figure 4.

2.3. Preparing the data

For attribute data, the data representing the counts of defective items needs to be arrayed in a column. The computations are performed on the sheet, so they are visible after the picture has been drawn. This has the advantage of making the computations more transparent to the user, though it also means that if the values in the new cells change (perhaps because the user re-draws the chart with updated information) the graph will change as well. Figure 2 shows the way the data from Montgomery (2005, Table 6-4) would be laid out. Note that the sample size data is located to the left of the column giving the number of nonconforming units.

For variable data, the user will pass in a matrix of observations, each row giving the observations for a particular sample. If the matrix of observations has only one column we produce the X chart for individual observations (below). No sample size needs to be supplied for variable data, since that information is carried in the numbers of non-empty entries in the rows.

In both cases, different types of results appear in the spreadsheet in one of two different ways. In some cases numbers are computed inside Visual Basic for Applications (VBA), and then displayed as numbers on the spreadsheet. For example, the lower control limit for the \bar{X} chart is computed internally and displayed as a number. In other cases VBA creates formulas and

relies on the underlying spreadsheet to evaluate them. For example, the “X-bar” column on the same chart is computed through a call to **Excel**’s built-in `=AVERAGE` function. This choice has been made for programming convenience, but it has the effect that if the user changes the underlying data, some columns will change and others will not. Of particular note is that the “Flags” columns are computed in VBA and do not update when the data changes. Consequently users cannot change the data and expect the results to be properly updated automatically; instead, they will need to re-run the add-in to produce an updated chart.

3. Statistical overview: Charts for attribute data

This section gives a brief overview of the control charts produced by the add-in, for practitioners who might be unfamiliar with their use. In each case we envision drawing a random sample from some distribution and counting the number of defective items in the sample (for attribute data) or measuring each item in the sample with a continuous measurement (for variable data). The following paragraphs describe the different charts.

The p chart

The p chart helps practitioners look at variations in the rate at which defective items are produced. We envision that sample j consists of n_j observations, each of which is either defective or not defective. The sample proportion of defectives within each lot, \hat{p}_j , is computed in the usual way. The chart plots these \hat{p}_j ’s against sample number. The centerline of the graph is given by \bar{p} , which is the defect rate from all samples combined (or the value of p provided by the user, if there is one). Control lines for sample j are plotted at $\bar{p} \pm \hat{\sigma}_j$, $\bar{p} \pm 2\hat{\sigma}_j$, and $\bar{p} \pm 3\hat{\sigma}_j$, where $\hat{\sigma}_j$ is the usual estimate of standard deviation given by $\sqrt{\bar{p}(1-\bar{p})/n_j}$. Very often all the sample sizes will be the same, in which case all the control lines will be horizontal.

The np chart

The np chart is simply a p chart (for the constant sample size case) in which everything has been multiplied by n . This means that the chart’s vertical axis is expressed in terms of numbers of defectives per lot, rather than in defective rate, which might be easier for non-technical personnel to interpret. The np chart is not appropriate for variable sample-size data, however, since defect counts will naturally be expected to increase with sample size. This gives rise to a variable centerline, which can be difficult to interpret. The add-in will not permit such a chart to be drawn.

The c chart

The c chart counts the number of defects (not defective items) in a particular sample. This would be appropriate for data arising from a Poisson process, for example. We require a constant sample size (so as to ensure a constant centerline) and we compute the defective rate c_j within each sample. The plot’s centerline is set to a value \bar{c} , which is either the given value c or, if no value is supplied, the average of the c_j . Following the Poisson assumption, we compute the standard deviation of \bar{c} as $\sqrt{\bar{c}}$, and draw control lines at $\bar{c} \pm \sqrt{\bar{c}}$, $\bar{c} \pm 2\sqrt{\bar{c}}$, and $\bar{c} \pm 3\sqrt{\bar{c}}$.

The u chart

The u chart measures the defect rate, rather than the count, in Poisson-type data. For a constant sample size n , then, it is simply a c chart in which everything has been divided by n . When sample sizes vary and no centerline is provided, the centerline \bar{u} is set at the total number of defects divided by the total sample size. Control lines for sample j are then drawn at $\bar{u} \pm 1, 2$, or $3\sqrt{\bar{u}/n_j}$.

4. Statistical overview: Charts for variable data

 \bar{X} and R chart, σ supplied

The \bar{X} chart is used when the data consists of a set of samples of continuous measurements, each measurement supposed to be iid from some distribution. The centerline of the chart is set to a user-supplied μ , if there is one, or to the overall average of all the observations in the sample, denoted \bar{x} . If a σ is supplied, control limits are conventionally drawn at μ (or \bar{x}) $\pm 3\sigma/\sqrt{n}$.

The R chart plots the ranges of the individual samples. The random variable $W = R/\sigma$, based on an iid sample from a normal distribution, itself has a known distribution whose mean and standard deviation are denoted by d_2 and d_3 , respectively. These values, which depend only on sample size, have been tabulated in a number of locations. We have taken the values of those constants (and others described below) from a package in the R statistical environment (R Development Core Team 2009) named **qcc** and due to Scrucca (2004).

When σ is given, the centerline of the chart is set at $d_2\sigma$, and upper and lower control limits set at $d_2\sigma \pm 3d_3\sigma$. Often the values $d_2 - 3d_3$ and $d_2 + 3d_3$ are written D_1 and D_2 and tabulated separately.

 \bar{X} and R chart, σ not supplied

When no σ is supplied, it can be estimated. One way to do this is to note that, since d_2 is the mean of the distribution of R/σ , \bar{R}/d_2 serves as an estimate of σ . This estimate (call it $\hat{\sigma}$) is in fact unbiased (Montgomery 2005). Then the \bar{X} chart can be drawn with centerline at \bar{x} and with control limits at $\bar{x} \pm 3(\bar{R}/d_2)/\sqrt{n}$. The value $3(\bar{R}/d_2)/\sqrt{n}$ is denoted by A_2 and again we have taken values of this expression from R.

Meanwhile, the standard deviation of the range R is $d_3\sigma$, which can be estimated by $d_3(\bar{R}/d_2)$. The R chart is therefore drawn with the centerline at \bar{R} and the control limits at $\bar{R} \pm 3d_3(\bar{R}/d_2)$. The symbols D_3 and D_4 are often used to denote $1 - 3(d_3/d_2)$ and $1 + 3(d_3/d_2)$; in that case the control limits of the \bar{R} chart can be written $D_3\bar{R}$ and $D_4\bar{R}$.

 \bar{X} and S chart, σ supplied

The R chart is not well suited for variable sample sizes, since in that case it produces a changing centerline. When sample sizes vary, or when they are large (say, bigger than around 10 or 12), Montgomery (2005) recommends using an \bar{X} chart based directly on the standard deviation, together with an S chart, which shows the changing values of sample standard deviations across time. When a standard value of σ is supplied, the \bar{X} chart is drawn as

before.

Our estimate of the standard deviation for a single sample is the usual one based on the unbiased estimator of variance, $s^2 = 1/(n-1) \sum (x_i - \bar{x})^2$. Of course s is not unbiased for σ ; we write $\mathbf{E}(s) = c_4\sigma$, where c_4 is another value that depends on sample size. (This one we compute directly in our VBA code, using the expression in [Montgomery \(2005, p. 95\)](#)). Furthermore we have $\text{SD}(s) = c_4\sqrt{1 - c_4^2}$. Therefore the S chart is drawn with the centerline at $c_4\sigma$ and control limits at $c_4\sigma - 3\sigma\sqrt{1 - c_4^2} = B_5\sigma$ and $c_4\sigma + 3\sigma\sqrt{1 - c_4^2} = B_6\sigma$.

\bar{X} and S chart, σ not supplied

When no value of σ is given, we compute \bar{s} , the average of the sample standard deviations (weighted by sample size if necessary), and estimate σ by \bar{s}/c_4 . The \bar{X} chart is drawn with the centerline at \bar{x} and control limits at $\bar{x} \pm 3(\bar{s}/c_4)(1/\sqrt{n}) = \bar{x} \pm A_3\bar{s}$, where A_3 is another value that can be looked up in a table or computed from c_4 .

In this case the S chart is drawn with centerline at \bar{s} and control limits at $\bar{s} \pm 3(\bar{s}/c_4)\sqrt{1 - c_4^2} = (B_3\bar{s}, B_4\bar{s})$ for suitably chosen B_3, B_4 .

For all four of these charts, users should be aware that the algorithm estimates variance by pooling variability within samples. If the samples are taken within batches produced within lots, there can be an extra (batch-to-batch) source of variability that the algorithm will not capture, and, as a result, the control limits will be too narrow. See [Montgomery \(2005\)](#), exercise 5-70 for an example.

X chart for individual observations

When it is necessary to draw a control chart for individual measurements, it is common to use, as an estimate of the range, the moving range and its average. That is, we define $\text{MR}_i = |x_i - x_{i-1}|$ and compute the average of these, $\overline{\text{MR}} = 1/(n-1) \sum \text{MR}_i$. Then the control chart has its centerline at \bar{x} and control limits at $\bar{x} \pm 3\overline{\text{MR}}/d_2$. For this chart extra care must be taken to track exclusions; if, for example, observation 5 is excluded, we must ensure that neither $|x_5 - x_4|$ nor $|x_6 - x_5|$ is used in the calculation of $\overline{\text{MR}}$.

5. Conclusion and future development

5.1. Conclusion

We have demonstrated an add-in to the widely-used spreadsheet program **Excel** that provides control chart functionality, previously available only in commercial products, in a free open-source form. The add-in draws many of the major sorts of process control charts for attribute and variable data with a straightforward interface. It follows the development in the textbook of [Montgomery \(2005\)](#) and is suitable for use in a course on quality control methods.

5.2. Future development

Future releases of the add-in will concentrate on improvement in two areas. First, the display of the computed values of confidence levels and so on, which currently appears on the spread-

sheet behind the chart, should be able to be suppressed. If results are desired, the user should be able to select the location in which they are displayed. Second, the ability to modify the trend rules (not just select a subset of them) might be useful.

Beyond user interface issues, the addition of the s^2 , the cumulative sum (CUSUM), and exponentially weighted moving average (EWMA) charts would complete the set of univariate charts detailed in Montgomery (2005). The add-in might be even more useful in an instructional setting if it incorporated some of the process control tools not related to control charts, like, for example, a calculator to implement the ANSI sampling standards.

References

- Montgomery DC (2005). *Introduction to Statistical Quality Control*. 5th edition. John Wiley & Sons, New York, NY, USA.
- R Development Core Team (2009). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.
- Scrucca L (2004). “**qcc**: An R Package for Quality Control Charting and Statistical Process Control.” *R News*, 4(1), 11–17. URL <http://CRAN.R-project.org/doc/Rnews/>.
- Walkenbach J (1999). *Microsoft Excel 2000 Power Programming with VBA*. IDG Books, Foster City, CA, USA.

Affiliation:

Samuel E. Buttrey
Code OR/Sb
Department of Operations Research
Naval Postgraduate School
Monterey, CA 93943, United States of America
E-mail: buttrey@nps.edu
Telephone: +1/831/656-3035