

5

Control charts

The fact that the criterion which we happen to use has a fine ancestry in highbrow statistical theorems does not justify its use. Such justification must come from empirical evidence that it works. (Shewhart, 1931, p. 18)

Overview

Control charts or process behaviour charts have been used for nearly 90 years to monitor process performance. Although originally developed for use in manufacturing industry they are now widely applied to processes involving the provision of services in fields such as finance and healthcare.

This chapter deals with a wide variety of control charts and with their creation, interpretation and maintenance via Minitab. Variables charts enable the monitoring of continuous random variables (measurements), while attribute charts monitor discrete random variables (counts). The consequences of tampering with processes are illustrated. Reference is made to auto-correlated data and feedback adjustment. Time-weighted control charts will be introduced, as will multivariate charts for the simultaneous monitoring of two or more variables.

The term ‘control charts’ suggests that these tools have a role only in the control phase of Six Sigma projects. However, as Figure 1.4 indicates with reference to the transient ischaemic attack and stroke clinic project, they may also be employed during the measure, analyse and improve phases. Indeed, the team involved with the project resolved that the control charts should continue to be maintained as a control measure once the project had formally ended.

5.1 Shewhart charts for measurement data

5.1.1 I and MR charts for individual measurements

‘The general idea of a control chart was sketched out in a memorandum that Walter Shewhart of Bell Labs wrote on May 16, 1924’ (Ryan, 2000, p. 22) Reference has already been made in

Table 5.1 Starch temperature data.

Observation	1	2	3	4	5	6	7	8	9	10
Time	08:00	08:15	08:30	08:45	09:00	09:15	09:30	09:45	10:00	10:15
Temperature	27.2	27.6	26.8	27.2	27.1	26.6	27.6	27.7	27.5	26.6
Observation	11	12	13	14	15	16	17	18	19	20
Time	10:30	10:45	11:00	11:15	11:30	11:45	12:00	12:15	12:30	12:45
Temperature	27.2	26.7	25.9	27.1	27.6	27.5	28.3	26.5	29.0	27.2

Chapter 2 to variation due to common causes and due to special causes. Caulcutt (2004, p. 37) referred to Shewhart’s thinking as follows:

He suggested that a process is acted upon by countless factors, many of which have little effect on the measured performance. Nonetheless, these minor factors, or ‘common causes’, are important because they are jointly responsible for the random variation in performance. If, in addition a process is acted upon by a major factor, or a ‘special cause’, the process will change and this change may be revealed by a violation of the control chart rules.

Liquid starch is used in the packaging industry in the manufacture of corrugated paper. Starch temperature is monitored at a manufacturing plant, which operates continuously, by recording temperature (°C) at intervals of 15 minutes. A set of 20 consecutive observations of temperature on 2 August 2010 while the Blue shift team (one of three) was running the process is given in Table 5.1, along with the time of observation.

One of the new features in Release 16 of Minitab is the **Assistant** menu. Selection of **Assistant > Control Charts...** yields the flow chart displayed in Figure 5.1. The first question to consider is: what is the data type? Temperature is a continuous random variable – one may think of measuring temperature using a mercury thermometer and the endpoint of the column

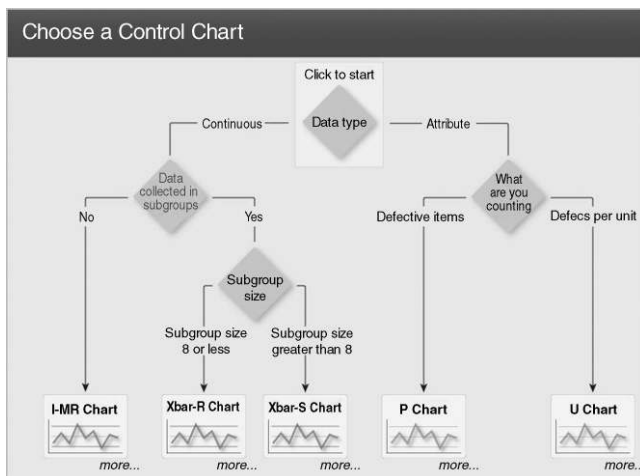


Figure 5.1 Flow chart for chart selection.

being located at any position on the continuous scale marked on the glass body of the instrument. The next question to address is whether or not the data are collected in subgroups. The answer is negative as a single temperature measurement is made every 15 minutes. This leads to the choice **I-MR Chart** – *individual* values and the *moving range* of these values are plotted in what is actually a pair of charts.

Clicking on **more...** underneath the **I-MR Chart** icon yields guidelines on collecting the data and using the chart. Clicking on the icon itself yields a simplified menu for creation of the chart with the assumption that the individual values to be charted have already been set up in a column of a worksheet. The author has opted to introduce the reader immediately to the full menus for the creation of individual value and moving range charts.

Columns called Date, Shift, Time and Temperature were set up in a Minitab worksheet. (In order to enter the times at which the temperatures were recorded, i.e. 8:00 to 12:45 in intervals of 15 minutes, one may use **Calc > Make Patterned Data > Simple Set of Date/Time Values...** with **Patterned Sequence** specified as **From first date/time: 08:00 To last date/time: 12:45 In steps of: 15** with **Step unit: Minute**.) The temperature data must first be entered into a column, along with any other relevant data in other columns. Use of **Stat > Control Charts > Variables Charts for Individuals > Individuals...** yields the dialog box in Figure 5.2.

Temperature is entered in **Variables:** to be charted. Under **I Chart Options...**, clicking on the **Estimate** tab, clicking on the down arrow to select **Use the following subgroups when estimating parameters** and inserting 1:20 in the window ensures that all 20 measurements will be used in the calculation of the chart limits. Defaults were accepted otherwise. In addition, **Stamp** was checked under **Scale...** and Time selected under **Stamp columns:** This yields the basic individuals control chart for the starch temperature displayed in Figure 5.3.

Prior to discussion of the chart the reader is invited to enter Date, Shift, Time and Temperature into four columns of a worksheet and to recreate the chart. Note that up to three Stamp columns may be selected, so as more data come to hand one could select Shift and Date, in addition to Time, in order to aid chart interpretation. On creating the chart the reader may find that the \bar{X} , LCL and UCL reference lines are labelled outwith the chart area. In order to have the chart appear as in Figure 5.3 select **Tools > Options...**, double-click on **Control**

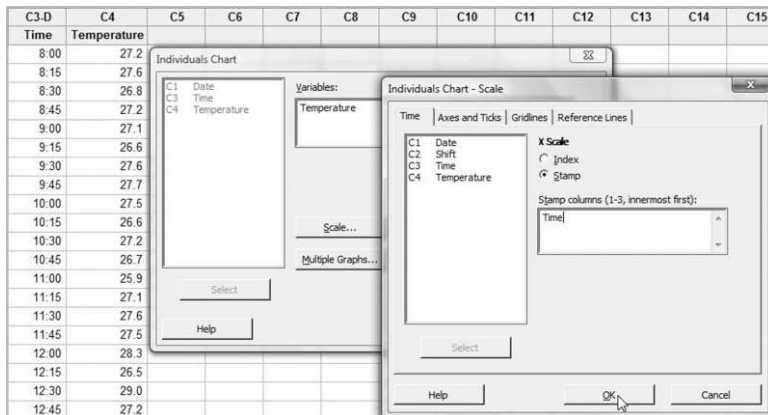


Figure 5.2 Creation of an individuals chart for temperature.

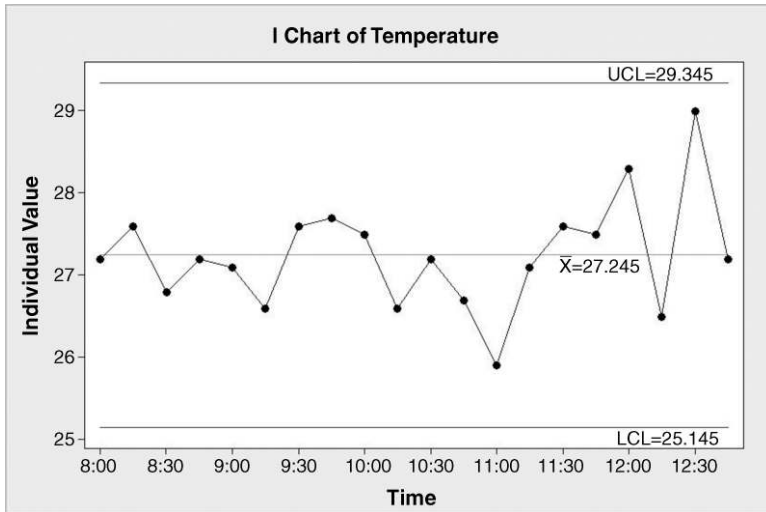


Figure 5.3 Individuals chart for temperature.

Charts and Quality Tools and then click on **Other** and check **Display control limit/center line labels for all stages**. The author recommends use of this setting as, when stages are used as in creating the chart in Figure 1.4, it is useful to have an indication of process performance levels given on the chart for all the stages.

The ‘naked run chart’ has been ‘clothed’ through the addition of a centre line (CL) corresponding to the mean, $\bar{X} = 27.245$, of the 20 observations of temperature together with the LCL and UCL reference lines. These are the upper and lower control limits (or upper and lower chart limits), respectively. They are ‘three sigma limits’ placed at ‘three sigma’ below and above the centre line respectively. Sigma in this context refers to an estimate of the process standard deviation obtained from the data. Since all 20 points lie between the ‘tramlines’ formed by the LCL and UCL it is conventional to conclude that the process is exhibiting only common cause variation. (Signals of evidence of special cause variation other than the occurrence of a point beyond the chart limits will be considered later in the chapter.) The process can be deemed to be in a state of statistical control and to be behaving in a stable and predictable manner within the natural limits of variation determined by the upper and lower chart limits. (None of the *P*-values on the Minitab run chart of the data is less than 0.05, which supports the conclusion from the control chart that there is no evidence of anything other than common cause variation affecting the process for maintenance of the starch temperature.)

The estimate of the process standard deviation used in the computation of the chart limits is *not* the sample standard deviation of the 20 Temperature observations. The estimate is obtained by calculating the 19 moving range (*MR*) values as indicated in Table 5.2. The reason for use of this method of estimating standard deviation is that the process data used to compute chart limits are often ‘contaminated’ by some special cause variation of which the creator of the chart is unaware. The moving range method of estimation of process standard deviation is influenced less by such contamination than the sample standard deviation method.

Each successive pair of temperatures is regarded as a sample of $n = 2$ values. The first pair has range $27.6 - 27.2 = 0.4$, the second pair has range $27.6 - 26.8 = 0.8$ and so on. The mean of the 19 moving ranges is 0.7895. Values of the factor, d_2 , which may be used to convert a mean

Table 5.2 Calculation of moving ranges.

Observation	1	2	3	4	5	6	7	8	9	10
Temperature	27.2	27.6	26.8	27.2	27.1	26.6	27.6	27.7	27.5	26.6
<i>MR</i>	*	0.4	0.8	0.4	0.1	0.5	1	0.1	0.2	0.9
Observation	11	12	13	14	15	16	17	18	19	20
Temperature	27.2	26.7	25.9	27.1	27.6	27.5	28.3	26.5	29.0	27.2
<i>MR</i>	0.6	0.5	0.8	1.2	0.5	0.1	0.8	1.8	2.5	1.8

range for a set of samples from a normal distribution into a standard deviation estimate, can be found in Appendix 2 or obtained via Help. Use **Help > Methods and Formulas > Quality and process improvement > Control charts > Variable Charts for Individuals > Methods for estimating standard deviation**. Clicking on the table link at the end of the Average moving range heading reveals that, for sample size $n = 2$, the value of d_2 is 1.128. (Reference to **Estimate** under **I Chart Options . . .**, in the dialog involved in the creation of Figure 5.2, reveals that the default **Method for estimating standard deviation** with **Subgroup size = 1** is to use **Average moving range** with **Length of moving range: 2**. The user may, if desired, specify that the standard deviation be estimated using moving ranges of length greater than 2 and may specify use of median moving range rather than average moving range.) The calculation of the chart limits is displayed in Box 5.1.

The values obtained in Box 5.1 agree with those displayed on the Minitab control chart in Figure 5.3. Since $3/1.128 = 2.66$ the calculations may be streamlined by use of the formulae:

$$\text{LCL} = \bar{X} - 2.66 \times \overline{MR},$$

$$\text{UCL} = \bar{X} + 2.66 \times \overline{MR}.$$

An estimate of process standard deviation (process sigma) is given by

$$\frac{\text{Mean moving range}}{d_2} = \frac{\overline{MR}}{d_2} = \frac{0.7895}{1.128} = 0.6999.$$

The lower chart limit is

$$\begin{aligned} \text{LCL} &= \bar{X} - 3 \times \text{Estimated sigma} \\ &= 27.245 - 3 \times 0.6999 = 27.245 - 2.0997 = 25.145, \end{aligned}$$

and the upper chart limit is

$$\begin{aligned} \text{UCL} &= \bar{X} + 3 \times \text{Estimated sigma} \\ &= 27.245 + 3 \times 0.6999 = 27.245 + 2.0997 = 29.345. \end{aligned}$$

Box 5.1 Calculation of chart limits.

These formulae would be required should the reader wish to create an individual chart using pencil and paper or via a spreadsheet package. The formulae, together with those for other control charts covered later in this chapter, are given in Appendix 3. The above formulae apply only when the moving range used is of length 2.

Having found no evidence of any special cause variation on the control chart, it could be adopted for process monitoring – with the limits calculated from the first 20 observations being employed. (It is not desirable, in general, to update the chart limits as new data become available.) Thus when the next observation of temperature becomes available all that is required is for the value to be plotted on the chart. In order to do this via Minitab, right-click on the active chart and from the menu click on **Update Graph Automatically**. On typing the next temperature value of 26.2 into the Temperature column in the worksheet the control chart will be automatically updated. Employment earlier of the option **Use the following subgroups when estimating parameters: 1 : 20** under **Estimate** ensures that the chart limits remain those calculated from the initial 20 observations.

The reader will have observed that, under **Estimate**, one may select the default option **Omit the following subgroups when estimating parameters:**. The author prefers generally to specify the data to be used in the calculations rather than the data to be omitted. Had the chart initial set of observations yielded a chart with, say, the 17th point outside the chart limits and there was a known special cause associated with that observation, then revised chart limits could be obtained with the 17th observation omitted from the calculations. This could be achieved with **Use the following subgroups when estimating parameters: 1 : 16 18 : 20** under **Estimate**.

The additional point lies within the chart limits so there is no signal of a possible special cause. On plotting the next four values 26.5, 25.6, 26.3 and 24.1 the chart will be as shown in Figure 5.4. The reader is invited to enter the data as described above and to create the chart in Figure 5.4 for her/himself. The reader should note that if a centre line label or chart limit label is obscured it may readily be moved to a better location by left-clicking, keeping the mouse button depressed and dragging.

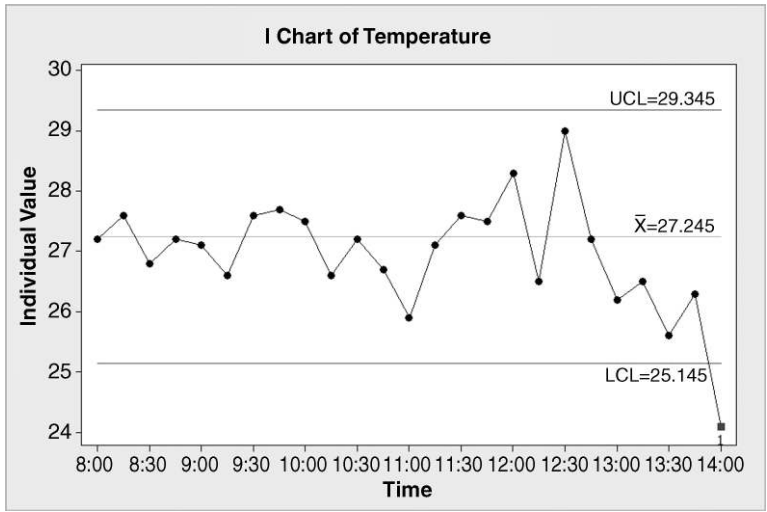


Figure 5.4 Individuals chart for temperature with additional data points.

The 25th point on the chart lies below the lower chart limit, so this provides evidence that a special cause may be affecting the process. Note that the plotting symbol for this ‘out of control’ point is a (red) square annotated with a 1. The reason for this is that there are a number of tests for special causes of which the first on the list provided by Minitab is the occurrence of a point outside the chart limits, i.e. of a point lying at a greater distance than three standard deviations from the centre line.

The maintenance engineer was subsequently called in and found a defective heating element, which he replaced. Thus the special cause of variation in the process was removed. One could then proceed to continue to monitor temperature using the chart with the limits established using the first 20 observations. In the case of major changes to the process it might be advisable to start the whole charting process again, i.e. to gather a series of initial temperature readings and to plot an initial chart. If there are no points outside the limits on this new chart then it can be adopted for further routine monitoring. If there are points outside the limits then a search should be made for potential special causes.

The moving ranges may also be plotted in a second control chart. For the 25 temperatures the chart of the 24 moving ranges is shown in Figure 5.5. It was created using **Stat > Control Charts > Variables Charts for Individuals > Moving Range...**, clicking on **MR Options...** and then on the **Estimate** tab, entering **Variables:** Temperature and specifying **Use the following subgroups when estimating parameters: 1 : 20**. The upper limit was calculated from the initial set of 19 moving ranges as follows:

$$UCL = 3.267 \times \overline{MR} = 3.267 \times 0.7895 = 2.579.$$

No reference will be made in this book to underlying theory concerning the distribution of ranges of samples from a normal distribution and to the derivation of the above upper limit formula. The ‘centre’ line on the chart is plotted at the level of the mean moving range. The lower chart limit is effectively 0 when ranges of pairs of measurements are used to estimate

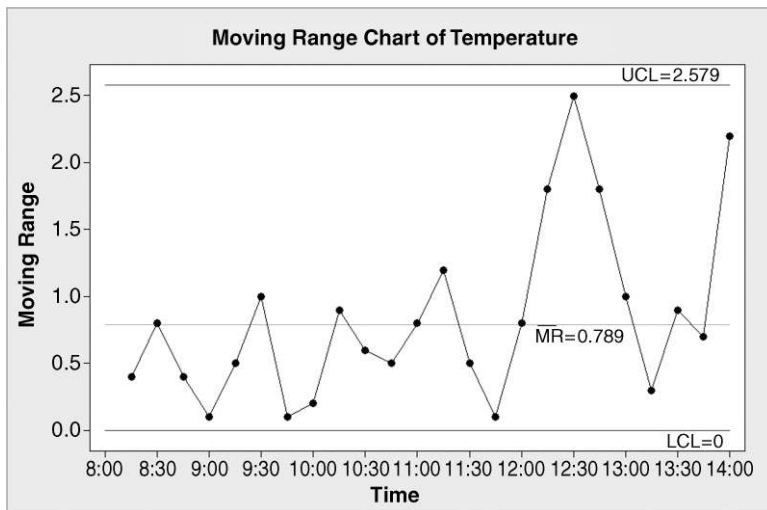


Figure 5.5 Moving range chart with limits based on first 20 temperatures.

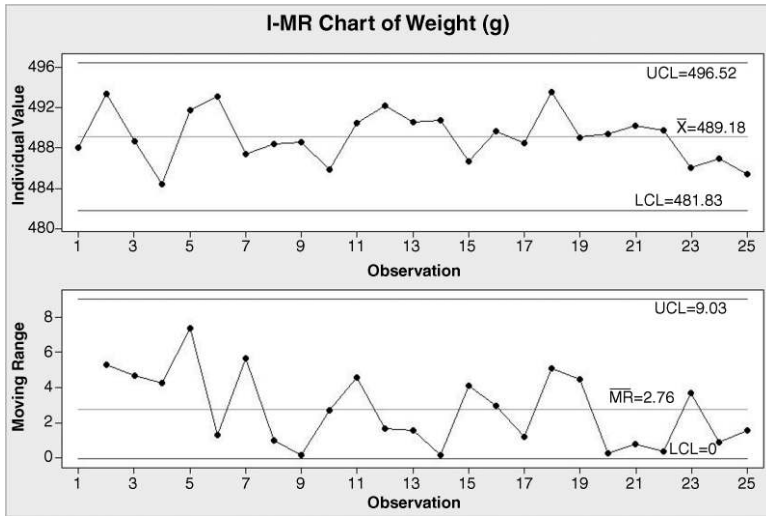


Figure 5.6 Individual and moving range charts for weight.

process standard deviation. All the points plotted lie within the chart limits. Note that, since the moving range can never take a negative value, a point below the (effective) lower chart limit is impossible.

The individuals chart or X chart can signal changes in the process location. Ideally the moving range chart would only signal changes in the process variability. However, changes in the process location, which are not accompanied by any change in process variability, can also yield points above the upper limit on the moving range chart. Montgomery (2009, p. 264) urges caution in the use of Shewhart control charts for individual measurements.

Note that Minitab offers the facility to create the charts separately or as a pair. The I-MR pair of charts – individuals and moving range – for the 25 bottle weights given in Table 2.1 is shown in Figure 5.6. The chart limits were calculated using all 25 data values. The charts were created using **Stat > Control Charts > Variables Charts for Individuals > I-MR...**, selecting **I-MR Options...**, clicking on the **Estimate** tab and specifying **Use the following subgroups when estimating parameters: 1 : 25**. The data are available in Weight1A.MTW.

No points fall outside the three-sigma control chart limits on either chart, so extended charts with those limits could subsequently be used to monitor the process. The updated chart with the addition of the data for the further Weights recorded in Table 2.2 is shown in Figure 5.7. The extended data set is available in Weight1B.MTW and the reader should note that, in creating the chart in Figure 5.7, it is necessary to specify **Use the following subgroups when estimating parameters: 1 : 25** under **Estimate via I-MR Options...**

5.1.2 Tests for evidence of special cause variation on Shewhart charts

The run chart in Figure 2.14 provided evidence of special cause variation, yet no points fall outside the limits on either control chart in Figure 5.7. In addition to a point outside chart limits providing a signal of evidence of the presence of special cause variation, there are a number of other tests used to provide evidence of the presence of special causes of variation. The tests available in Minitab for the Shewhart charts provided are accessed via the **Tests** tab under

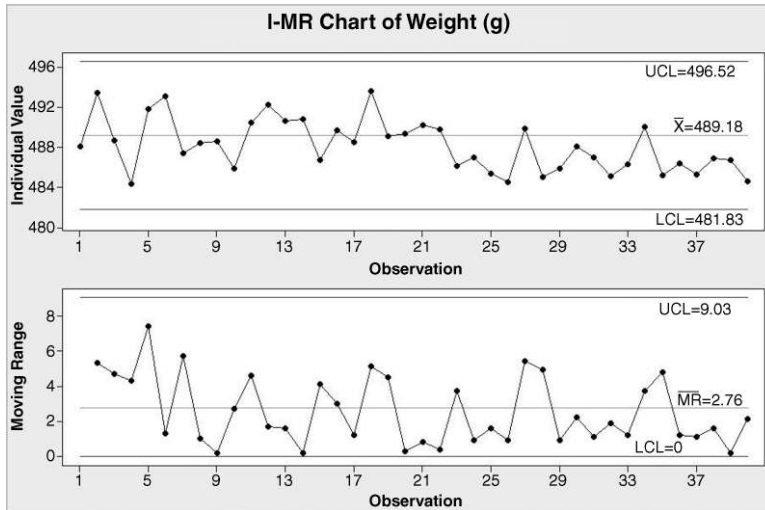


Figure 5.7 Updated individual and moving range charts for weight.

<Chart Type> **Options...** (<Chart Type> represents **I Chart** or **MR** or **I-MR** etc., depending on the particular chart or charts being used.) The tests available for the individuals chart are displayed in Figure 5.8.

The default test checks for the occurrence of a point more than three standard deviations from the centre line, i.e. for a point outside the chart limits. This test is referred to as Test 1 in Minitab. Note that it is listed first in the dialog box displayed in Figure 5.8. The default versions of all eight tests used in Minitab are listed in Box 5.2. The user can select which of the tests

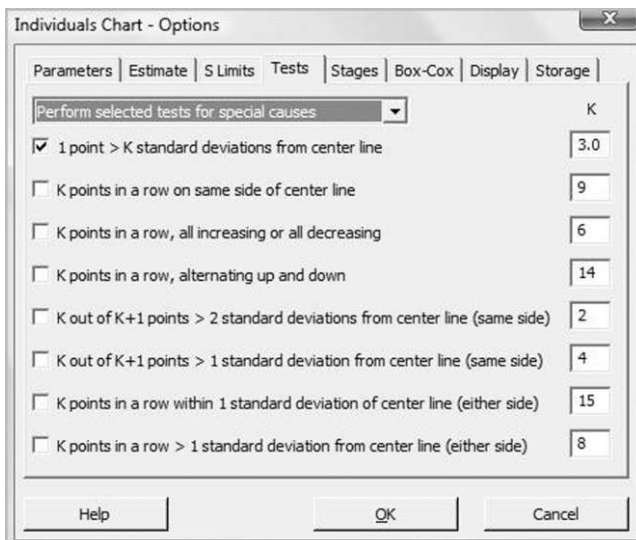


Figure 5.8 Tests for evidence of special causes.

<u>Test 1</u>	1 point more than 3 standard deviations from center line
<u>Test 2</u>	9 points in a row on same side of center line
<u>Test 3</u>	6 points in a row, all increasing or all decreasing
<u>Test 4</u>	14 points in a row, alternating up and down
<u>Test 5</u>	2 out of 3 points > 2 standard deviations from center line (same side)
<u>Test 6</u>	4 out of 5 points > 1 standard deviation from center line (same side)
<u>Test 7</u>	15 points in a row within 1 standard deviation of center line (either side)
<u>Test 8</u>	8 points in a row > 1 standard deviation from center line (either side)

Box 5.2 Tests for evidence of special causes available in Minitab.

he/she wishes to apply or use the drop-down menu to select the application of either all tests or no tests. All eight tests are available for the individuals chart, but only the first four are available for the moving range chart. The reader is strongly recommended to refer to Appendix 4, where an example of evidence of special cause variation being provided by each one of the eight tests is displayed. (If desired, the reader may define the tests differently from the defaults listed in Box 5.2. For example, many practitioners use 8 rather than 9 points in a row in Test 2. The change may be made locally by changing the value of K from 9 to 8 in the second box in the column on the right of the dialog box in Figure 5.8, but it will revert to the default value of 9 in any new Minitab project. If desired the change may be made global by using **Tools > Options . . .**, double-clicking on **Control Charts and Quality Tools**, clicking on **Tests** and changing 9 to 8 in the window for Test 2.)

Another option provided for control charts is the positioning of horizontal reference lines/control limits at any number of standard deviations from the centre line the user desires. For the weight data in Weight1B.MTW select **Stat > Control Charts > Variables Charts for Individuals > Individuals . . .** and click on **I Chart Options . . .** On the **Estimate** tab, select **Use the following subgroups when estimating parameters** and enter 1 : 25 beneath it. On the **S Limits** tab, select **Display control limits at** and enter **These multiples of the standard deviation: 1 2 3**. Finally, click on the **Tests** tab and select **Perform all tests for special causes**. This yields the chart in Figure 5.9.

The 29th point plotted is labelled with the digit 6, indicating that there is a signal of a possible special cause affecting the process from Test 6 on the list in Box 5.2. Test 6 involves checking for the occurrence of four out of five consecutive points that are more than one standard deviation away from the centre line. The reader should verify from scrutiny of Figure 5.9 that, of the ringed points 25–29, four are more than one standard deviation away from the centre line, lying below the one-sigma lower limit of 486.73. The figure also shows the message displayed on moving the mouse pointer to the label 6 associated with the 29th data point. The Session window displays the report, shown in Panel 5.1, on the chart just created.

Thus when 29 observations have been plotted there is evidence that a special cause is affecting the process. The warning means that one has to take care that the limits have been calculated using the desired observations in the creation of any subsequent chart. (A run chart of the first 29 observations created using Minitab provides no evidence of any special cause of variation affecting the process. Thus, with the use of additional tests, the individuals control chart provides evidence of special cause variation and it does so with fewer observations than the run chart in this case.)

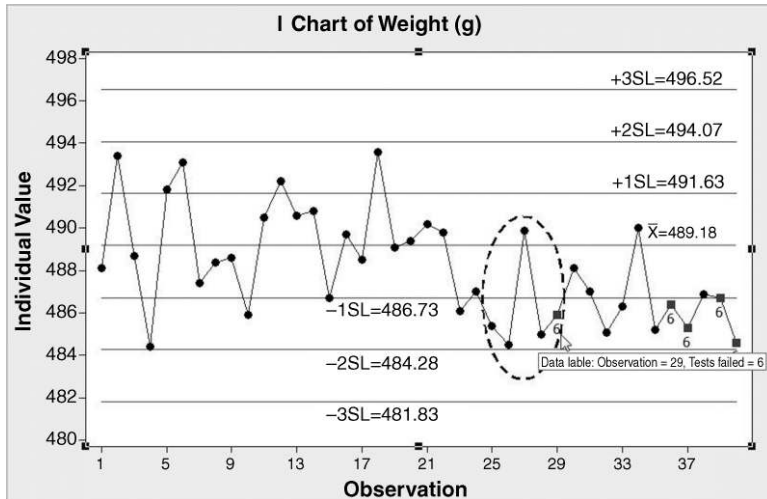


Figure 5.9 Alternative individuals chart of weight data.

Test Results for I Chart of Weight (g)

TEST 6. 4 out of 5 points more than 1 standard deviation from center line (on one side of CL).

Test Failed at points: 29, 36, 37, 39, 40

* WARNING * If graph is updated with new data, the results above may no longer be correct.

Panel 5.1 Session window report on chart in Figure 5.9.

In the next section charts for samples or subgroups of measurements are introduced, together with the facility in Minitab to chart data from different stages in the history of a process on the same diagram. This facility may be used with all of the Shewhart charts considered in this chapter.

5.1.3 Xbar and R charts for samples (subgroups) of measurements

In many situations processes are monitored using samples or subgroups of product. The third column in the Minitab worksheet Camshaft.MTW (available in the Minitab Sample Data folder supplied with the software) gives the length (mm) of a series of 20 samples (or subgroups) of size $n = 5$ camshafts taken from supplier 2. Reference to the flow chart from the **Assistant** menu displayed in Figure 5.1 leads to the widely used procedure of computing the sample means (\bar{X}) and the sample ranges (R) and plotting both series of values in sequence with appropriate centre lines and control limits added. The tests available in Minitab for the \bar{X} chart of means and the R chart of ranges match those for the individuals and moving range charts, respectively.

To create the charts using Minitab use **Stat > Control Charts > Variables Charts for Subgroups > Xbar-R...** In this case, where all the measurements are in a single column, the default **All observations for a chart are in one column:** option is used. Clicking on the

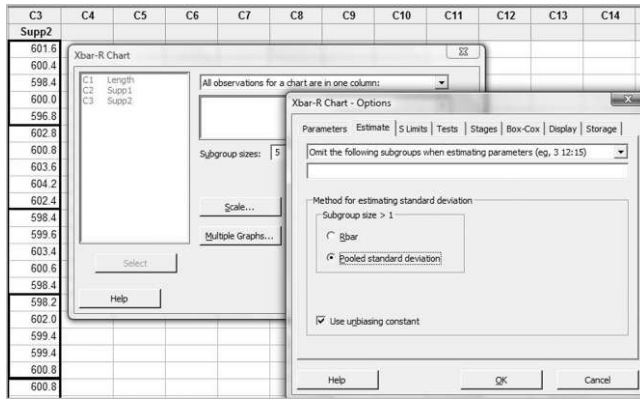


Figure 5.10 Creation of mean and range charts for subgroups of size 5.

Estimate tab under **Xbar-R Options...** reveals that the default **Method for estimating standard deviation** with **Subgroup size > 1** is to use **Pooled standard deviation**. Note that **Use unbiasing constant** is also checked by default.

One could opt to use **Rbar**, the mean of all the sample ranges, in order to estimate the standard deviation of camshaft length. Historically this was a widely used method, but statistical theory shows that the use of pooled standard deviation yields better estimates than use of mean sample ranges. Readers with an interest in the technical details may find it helpful to consult the paper by Mahmoud *et al.* (2010). Throughout this chapter the Xbar-R charts presented have all been created using the default option to estimate standard deviation using pooled standard deviation and the unbiasing constant.

The completed dialog is displayed in Figure 5.10, the option to perform all tests for special causes having been selected. Bold rectangles have been added to the image to indicate the first four samples/subgroups of camshaft length. The charts in Figure 5.11 were obtained. There are three signals on the Xbar chart indicating potential special cause variation affecting supplier 2's process – there is evidence that the process is not in a state of statistical control, i.e. that it is not behaving in a stable, predictable manner.

Let us suppose that discussion with those responsible for the process led to identification of assignable causes of variation for subgroup 2 (machine fault) and subgroup 14 (operator error) but not for subgroup 9. When signals of potential special cause variation lead to identification of actual special causes it is normal to recalculate the chart limits with the corresponding subgroups, 2 and 14 in this case, omitted from the calculations and to scrutinize the revised charts. This may be achieved using **Xbar-R Options...**, clicking on the **Estimate** tab and specifying 2 14 under **Omit the following subgroups when estimating parameters**, but the author prefers to employ **Estimate** specifying 1 3:13 15:20 under **Use the following subgroups when estimating parameters**. The **Labels** facility was used to add footnotes indicating the actions taken. The resulting chart pair is shown in Figure 5.12. (If the special causes identified have not been eliminated then one should be cautious about recalculating limits. The Xbar chart with the existing limits has successfully detected these causes and so has the potential to do so again should they recur. Otherwise there is a risk of 'tightening' the Xbar chart limits to such an extent that the chart starts to yield too many false alarm signals of special cause variation.)

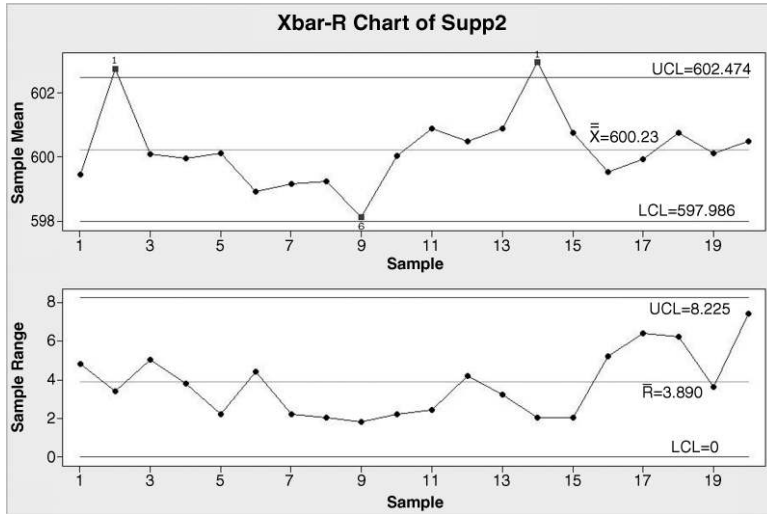


Figure 5.11 Mean and range charts for supplier 2.

The means for subgroups 2 and 14 fall outside the new control limits on the Xbar chart, but these subgroups can now effectively be ignored. Subgroup 9 no longer gives a signal on the Xbar chart, but subgroup 15 now does. However, this is a spurious signal since the point corresponding to subgroup 14 has been counted as one of the four from five consecutive points more than one standard deviation from the centre line (same side). There are no signals on the R chart, so the decision might well now be taken to begin monitoring the process by taking further subgroups of five camshafts at regular intervals and plotting the means and ranges on charts using the limits displayed on the charts in Figure 5.12.

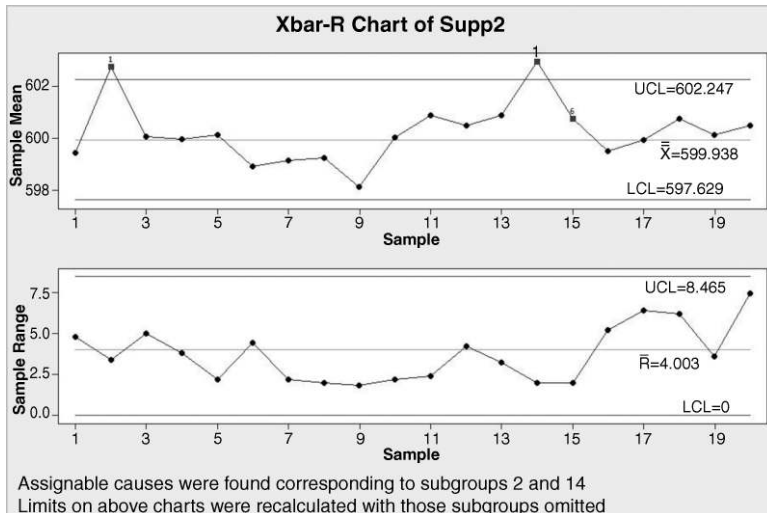


Figure 5.12 Xbar and R charts with revised limits.

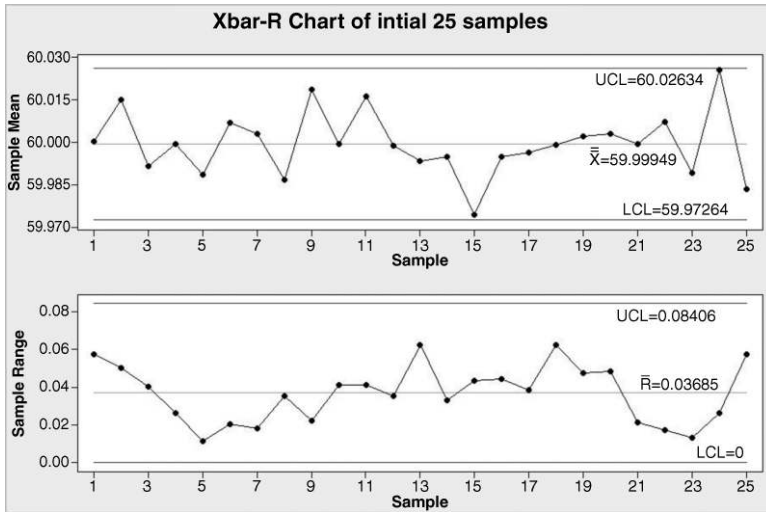



Figure 5.13 Initial Xbar and R charts for rods process.

Some simulated data will now be used to demonstrate how different types of process changes are indicated by signals on Xbar and R charts. Imagine a process which cuts extruded plastic into rods and that, under ideal conditions, the rods have length (mm) which is normally distributed with mean 60.00 and standard deviation 0.02, i.e. $N(60.00, 0.02^2)$. **Calc > Random Data > Normal...** was used to generate four columns of 25 values from the $N(60.00, 0.02^2)$ distribution. The values were rounded to three decimal places and stored in the supplied worksheet Rods.MTW. Each row of four values from the columns may be regarded as a sample/subgroup of size $n = 4$ from the normal distribution specified. Xbar and R charts of the initial data are shown in Figure 5.13.

In creating these charts **Observations are in one row of columns: x1-x4** was specified and, from the **Estimate** tab under **Xbar-R Options...**, one may select **Use the following subgroups when estimating parameters:** and specify 1 : 25. With all available tests applied there are no signals of potential special cause variation on the charts, so the decision could be taken to use Xbar and R charts with the limits shown to monitor the process. (Suppose that you forgot to select Perform all tests for special causes on the **Tests** tab. The Edit Last Dialog icon  may be used, or alternatively Ctrl + E, to access the most recently used dialog box and to make any desired changes.)

Before proceeding to look at further simulated data for the process, details of the calculation of the limits for the Xbar and R charts are presented in Boxes 5.3 and 5.4. Readers may skip the details in these boxes as the charts may be employed effectively without familiarity with technical details. If required, the formulae and constants involved are available from the **Help** menu via **Help > Methods and Formulas > Quality and process improvement > Control charts > Variable Charts for Subgroups > Methods for estimating standard deviation.**

A further 15 subgroups of four length measurements were generated using the same distribution as for the initial 25 subgroups that were charted in Figure 5.12. Then a further 20 subgroups were generated using different distributions to illustrate four different scenarios

The means, \bar{x} , of the 25 samples of $n = 4$ lengths have mean $\bar{\bar{x}} = 59.999\,49$, the double bar notation indicating that the value is the mean of a set of means. (Since the normal distribution used to generate the simulated data had mean $\mu = 60.00$ it is not surprising that the mean of the 25 means is close to 60.00 and thus provides a good estimate of the process mean μ .)

The pooled standard deviation is the square root of the mean of the 25 sample variances in this case, where the samples all have the same size, and is 0.017 837 1. This has to be divided by the unbiasing constant $c_4(76) = 0.996\,672$, available in the linked table in **Methods and Formulas**, to yield the estimate 0.017 896 7, of the process standard deviation (process sigma). (This estimate is close to the standard deviation $\sigma = 0.02$ of the distribution used to generate the simulated data.)

The theory of the distribution of the sample mean from Section 3.4 yields three-sigma Xbar chart limits of

$$\mu_{\bar{x}} \pm 3\sigma_{\bar{x}} = \mu \pm 3 \frac{\sigma}{\sqrt{n}}.$$

The chart limits are

$$\begin{aligned} \bar{\bar{x}} \pm 3 \times \frac{\text{Estimated sigma}}{\sqrt{n}} &= 59.999\,49 \pm 3 \frac{0.017\,896\,7}{\sqrt{4}} \\ &= 59.999\,49 \pm 0.026\,845\,0 = (59.972\,64, 60.026\,34). \end{aligned}$$

These limits are in agreement with those on the Xbar chart in Figure 5.13.

Box 5.3 Calculation of Xbar chart limits.

The three-sigma R chart limits are

$$\mu_R \pm 3\sigma_R = d_2\sigma \pm 3d_3\sigma,$$

where the constants d_2 and d_3 may be read from the linked table in **Methods and Formulas**. The centre line is placed at the estimated μ_R given by chart limits are

$$d_2 \times \text{Estimated sigma} = 2.059 \times 0.017\,896\,7 = 0.036\,85).$$

The chart limits are

$$\begin{aligned} d_2 \times \text{Estimated sigma} \pm 3d_3 \times \text{Estimated sigma} \\ &= 2.059 \times 0.017\,896\,7 \pm 3 \times 0.8794 \times 0.017\,896\,7 \\ &= (-0.010\,36, 0.084\,06). \end{aligned}$$

Since range cannot be negative, the lower chart limit is effectively 0. These limits and the centre line are in agreement with those on the R chart in Figure 5.13. Sample size has to be at least 7 for nonzero lower limits to occur on an R chart.

Box 5.4 Calculation of R chart limits.

in terms of process changes. Thus one can think of the process change, i.e. that a special cause of variation took effect, occurring at some time between the taking of the 40th and 41st samples. The data sets are supplied as RodsScenario1.MTW, RodsScenario2.MTW, RodsScenario3.MTW and RodsScenario4.MTW. In all four cases:

- limits on the charts (see Figures 5.14–5.16) are those calculated from data for the first 25 samples;
- all tests for evidence of special cause variation were performed.

In scenario 1 (process mean increased, process standard deviation unchanged), subgroup 42 gives rise to the first signal of the process change on the Xbar chart in Figure 5.14. In practice, on plotting the data for this sample, action would be taken to seek a special cause of variation affecting the process. Thus it took just two subgroups to signal the process change. However, if the samples were taken at 15-minute intervals this could correspond to up to half an hour of production of less satisfactory rods from the point of view of the customer. The number of samples required to signal a process change is referred to as the *run length*. (It must be emphasized that chart limits, or control limits as they are referred to by some, are *not* specification limits.)

In scenario 2 (process mean decreased, process standard deviation unchanged), it may be seen from Figure 5.15 that the first signal of the process change was from the 44th sample. Thus it took four subgroups to flag the change in this scenario. With a smaller change in the process mean in this scenario than in the first, it is not surprising that the run length is greater. (Note that, as in scenario 1, there is a spurious signal on the R chart. Recall that, in simulating the data for both scenarios 1 and 2, no change was made to the process standard deviation, i.e. to process variability.)

In scenario 3 (process standard deviation increased, process mean unchanged) it took just one sample for the R chart to signal the likely occurrence of a special cause affecting process variability. With increased variability the limits on the Xbar chart in Figure 5.16 are too close together from sample 41 onwards. Thus an increase in process variability typically yields signals on the Xbar chart as well as on the R chart. Thus it is advisable to examine the R chart first when employing Xbar and R charts for process monitoring purposes and to interpret the charts as a pair.

A major increase in variability, such as the one illustrated in this scenario, could have a major impact on process capability. Thus the process owners would most likely wish to take action quickly after the data for subgroup 41 were plotted to eliminate any special cause found to be affecting the process. The changes in location in the first two scenarios could also impact on process capability in a detrimental manner. The final scenario illustrated refers to a situation where process variability is reduced. Reduction of variability is fundamental to achieving quality improvement.

In scenario 4 (process standard deviation decreased, process mean unchanged) Figure 5.17 indicates that it took 13 subgroups for a signal to appear on the R chart. Note that the reduction in process variability does not give rise to signals on the Xbar chart. The reason for this is that from sample 40 onwards the Xbar chart limits are too far apart. Were we dealing with real as opposed to simulated data then perhaps a deliberate change was made to the process after subgroup 40 was taken, possibly as the result of a Six Sigma project undertaken to identify ways to reduce process variability. In any case it would clearly be desirable to maintain the reduced variability.

For the purposes of illustration let us suppose that, between the times at which samples 40 and 41 were taken, a new feed control system was fitted to the machine which cut the

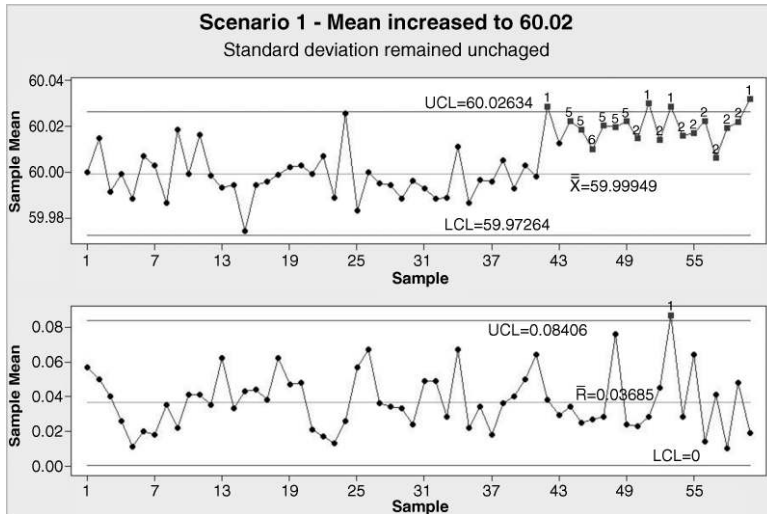


Figure 5.14 Xbar and R charts for an increase in process mean (scenario 1).

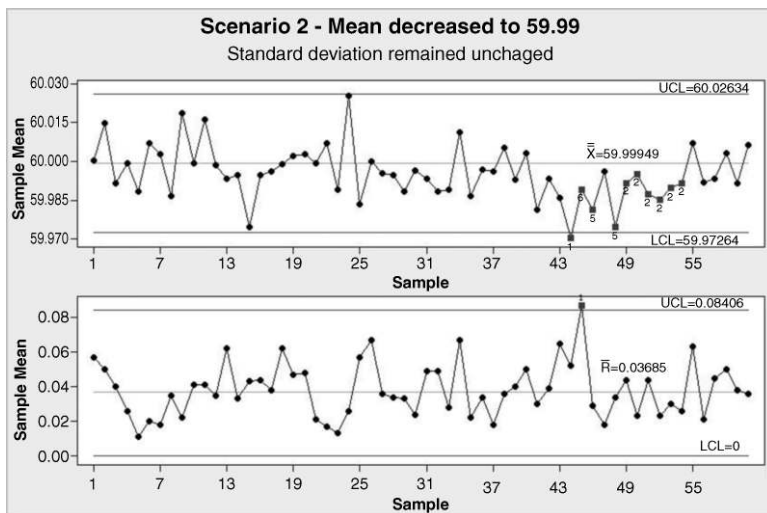


Figure 5.15 Xbar and R charts for a decrease in process mean (scenario 2).

extruded plastic into rods. In addition to the four columns containing the subgroups of four length measurements across their rows, a fifth column consisting of 40 values of 1 followed by 20 values of 2, indicating the two phases of operation monitored, has been added to the worksheet. Some of the dialog involved may be viewed in Figure 5.18. Points to note are:

- **Observations are in one row of columns:** x1-x4 has been specified
- Under **Options . . .** and **Estimate** one has to select **Use the following subgroups when estimating parameters:** and specify **both 1 : 25 and 41 : 60.**

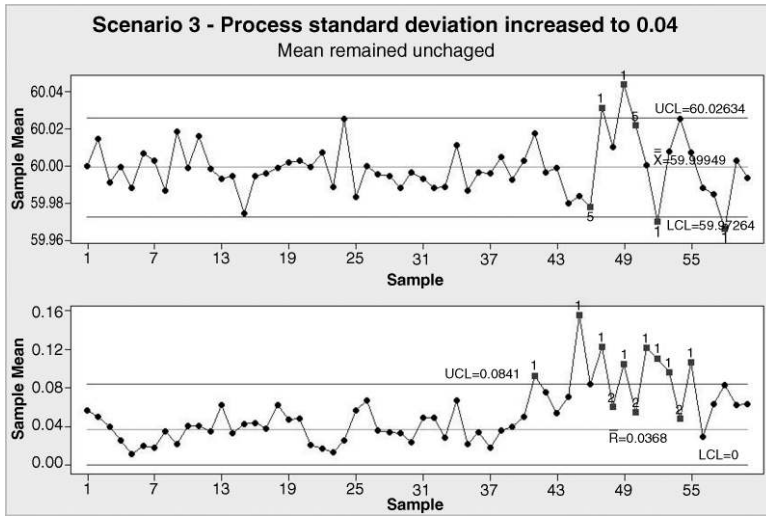


Figure 5.16 Xbar and R charts for an increase in process standard deviation (scenario 3).

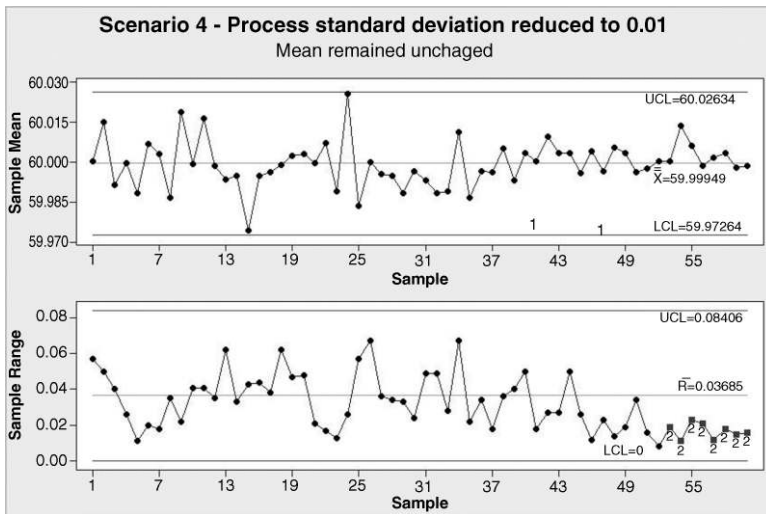


Figure 5.17 Xbar and R charts for a decrease in process standard deviation (scenario 4).

- Under **Options...** and **Stages** one has to specify selection of **Phase** in order to **Define stages (historical groups) with this variable:**. The default option to use both **When to start a new stage** and **With each new value** may be used. (Alternatively, **When to start a new stage** and **With the first occurrence of these values:** 1 2 could be employed.)

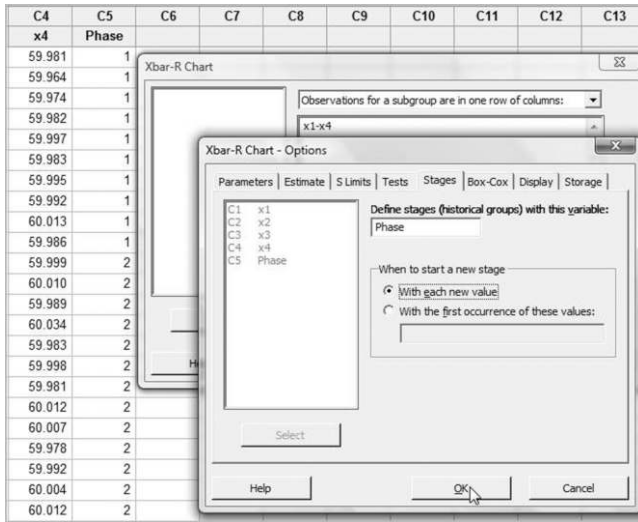


Figure 5.18 Defining stages in the creation of Xbar and R charts.

The charts are displayed in Figure 5.19. In order to have the limits and centre lines labelled for all stages use <Chart> **Options** > **Display** > **Other** and check **Display control limits/center line labels for all stages**.

The 20 subgroups plotted for the second stage may be regarded as initial data for a new Xbar-R chart pair. There is a signal from the fourth of the second-phase subgroups on the R chart, so many experts would recommend reviewing the situation after another few subgroups have been obtained before adopting the new charts for further routine monitoring. Montgomery (2009, p. 297) states that when used in this manner ‘the control chart becomes a

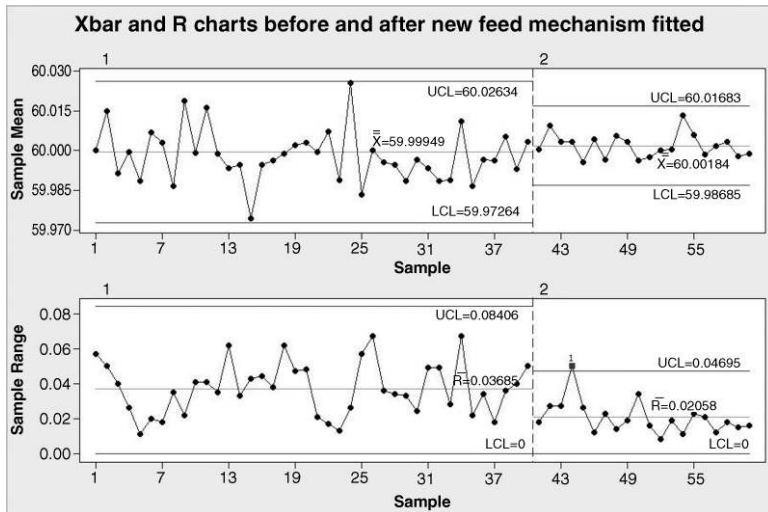


Figure 5.19 Xbar and R charts with two stages.

logbook in which the timing of process interventions and their subsequent effect on process performance is easily seen’.

As an alternative to mean and range (Xbar and R) charts, mean and standard deviation (Xbar and S) charts may be used when data for process monitoring are collected in subgroups. The flow chart in Figure 5.1 recommends the use of Xbar and S charts whenever sample size is 8 or more. An example of the use of an Xbar and S chart pair is provided as an exercise. Other topics on Shewhart control charts for measurement data, such as the use of individuals charts to check that a process is operating ‘on target’ and triple charts (I-MR-R/S charts) are referred to in follow-up exercises. No reference will be made to zone charts in this book.

Tests 1, 2 (with 8 points in a row), 5 and 6 are referred to as the Western Electric Company (WECO) rules, and some practitioners prefer to use these four tests rather than all eight available in Minitab. Ultimately the decision on which tests to use lies with the process team. The following comments are made in NIST/SEMATECH (2005, Section 6.3.2):

While the WECO rules increase a Shewhart chart’s sensitivity to trends or drifts in the mean, there is a severe downside to adding the WECO rules to an ordinary Shewhart control chart that the user should understand. When following the standard Shewhart ‘out of control’ rule (i.e., signal if and only if you see a point beyond the plus or minus 3 sigma control limits) you will have ‘false alarms’ every 371 points on the average . . . Adding the WECO rules increases the frequency of false alarms to about once in every 91.75 points, on the average . . . The user has to decide whether this price is worth paying (some users add the WECO rules, but take them ‘less seriously’ in terms of the effort put into troubleshooting activities when out of control signals occur).

Readers wishing to construct control charts for measurement data without using Minitab may find the factors in Appendix 2 and the formulae in Appendix 3 of value.

5.2 Shewhart charts for attribute data

5.2.1 P chart for proportion nonconforming

Consider a large e-commerce company at which there is concern over complaints from customers concerning inaccurate invoices being e-mailed to them. During the measure phase of a Six Sigma project aimed at improving the situation, random samples of 200 invoices were checked for inaccuracies, each week, for 20 weeks. The data, together with the calculated proportions, are shown in Table 5.3 and are available in the worksheet Inaccurate1.MTW.

Table 5.3 Invoice data.

Week no.	1	2	3	4	5	6	7	8	9	10
No. inaccurate	23	23	20	21	17	22	24	20	18	17
Proportion	0.115	0.115	0.100	0.105	0.085	0.110	0.120	0.100	0.090	0.085
Week no.	11	12	13	14	15	16	17	18	19	20
No. inaccurate	24	17	15	19	19	22	27	23	23	18
Proportion	0.120	0.085	0.075	0.095	0.095	0.110	0.135	0.115	0.115	0.090

The upper chart limit is given by

$$\begin{aligned} \text{UCL} &= \bar{p} + 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}} \\ &= 0.103 + 3\sqrt{\frac{0.103 \times 0.897}{200}} = 0.103 + 0.0645 = 0.1675. \end{aligned}$$

The lower chart limit is

$$\begin{aligned} \text{LCL} &= \bar{p} - 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}} \\ &= 0.103 - 3\sqrt{\frac{0.103 \times 0.897}{200}} = 0.103 - 0.0645 = 0.0385. \end{aligned}$$

Box 5.5 Calculation of P chart limits.

In this scenario counts are being made of the number of items (invoices) that are defective (contain one or more inaccuracies), so according to the flow chart in Figure 5.1 the appropriate control chart in this situation is a P chart or chart for proportion defective. Some refer to items as being nonconforming rather than defective. The mean of the 20 proportions is $\bar{p} = 0.103$ and the chart centre line is plotted at this value. This is taken as an estimate of the population proportion, p , and the calculation of the three-sigma limits for the chart are made using the formulae, given in Appendix 3, incorporating the standard deviation of a proportion stated in Chapter 4. The calculations are displayed in Box 5.5.

To create the chart with Minitab, the number of inaccurate invoices is entered into a column labelled No. Inaccurate and use made of **Stat > Control Charts > Attributes Charts > P . . .** No. Inaccurate is entered under **Variables:** as the variable to be charted and the sample/subgroup size is specified using **Subgroup sizes:** 200. The default versions of all four available tests were implemented under **P Chart Options . . . > Tests**. This may be achieved by checking each of the four tests or by selecting **Perform all tests for special causes** from the menu under **Tests**. The four available tests are the same as the first four of the eight available with both charts for individuals (X) and for means (Xbar). Examples of patterns yielding signals of possible special cause variation from the tests are given in Appendix 4. The chart is displayed in Figure 5.20. The reader should note that Minitab did not require a column of proportions in order to create the chart.

There are no signals of any potential special causes affecting the process so it can be deemed to be in a state of statistical control, operating in a stable and predictable manner with approximately 10% of invoices having inaccuracies. It can therefore be agreed to use the chart with the centre line and limits shown for future process monitoring.

Before proceeding to look at further data from the process the reader is invited to re-create the P chart using **Assistant > Control Charts** and clicking on **more . . .** under **P chart**.

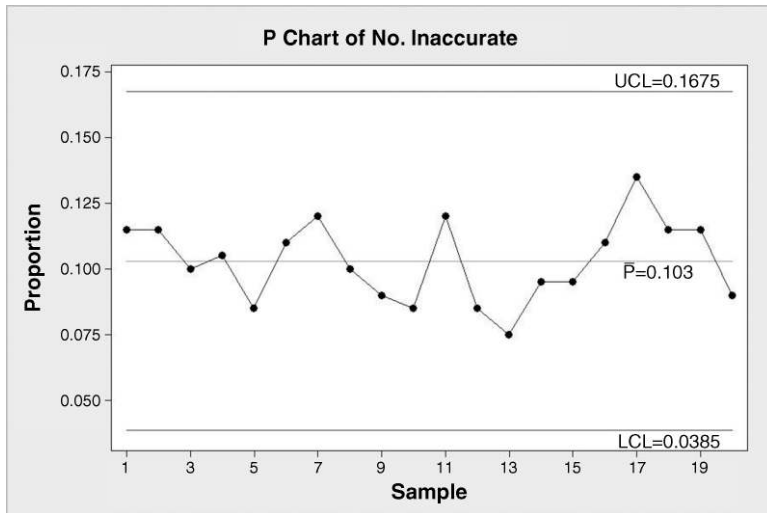


Figure 5.20 P chart of invoice data.

Click on **Attribute data** and read the two descriptions, then under **Attribute data** click on **Next**. Having read the two descriptions, under **Defective items** click on **Next** to obtain the screen shown in Figure 5.21. Clicking on the + icons yields further details. Finally the reader is invited to click on the create chart icon and reproduce the chart in Figure 5.20. (Hint: You will need to select **Estimate from the data** under **Control limits and center line**.) This leads to the creation of three items in the Graphs folder – a Stability Report that includes the P chart, a Summary Report and a Report Card. The clean bill of health from the Report Card supports the decision to ‘roll out’ the chart in Figure 5.20 for further monitoring.

P Chart

Home > Attribute data > Defective items

Click to create chart

The P chart monitors the proportion (or percentage) of defective items.

Guidelines

Collecting the data

- Collect data in subgroups (samples, lots). +
- Collect subgroups at appropriate time intervals. +
- Collect enough subgroups. +
- Subgroups must be large enough. +
- Subgroup sizes can be unequal. +
- Count the number of defective items in each subgroup. +

Using the chart

- Minitab automatically tests for special causes. +
- Estimate new control limits only when the process changes. +

Figure 5.21 P chart guidelines from the Assistant menu.

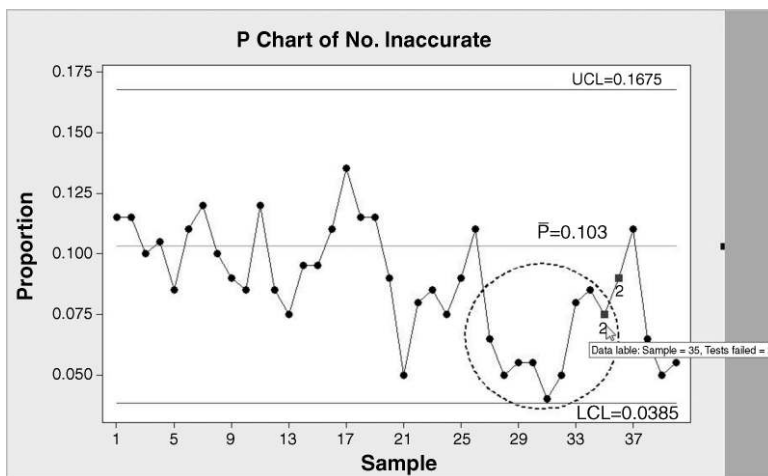
Table 5.4 Further invoice data.

Week No.	21	22	23	24	25	26	27	28	29	30
No. Inaccurate	10	16	17	15	18	22	13	10	11	11
Proportion	0.05	0.08	0.085	0.075	0.09	0.11	0.065	0.05	0.055	0.055
Week No.	31	32	33	34	35	36	37	38	39	40
No. Inaccurate	8	10	16	17	15	18	22	13	10	11
Proportion	0.04	0.05	0.08	0.085	0.075	0.09	0.11	0.065	0.05	0.055

At the end of the 20-week period during which the above data were gathered, changes planned by the project team were introduced. Data for the next 20 weeks are given in Table 5.4 and the data for all 40 weeks are available in the worksheet Inaccurate2.MTW.

The extended control chart for proportions was created using **Stat > Control Charts > Attributes Charts > P...** and is shown in Figure 5.22. Note that the limits are those based on the data for the first 20 weeks so this was indicated using **Estimate** under **P Chart Options...** (The P chart dialog available via the **Assistant** menu does not permit specification of the samples to be used in the calculation of the limits.) The samples to be used in the calculation of the limits may be specified in two ways: either **Estimate > Omit the following subgroups...** and enter 21 : 40 or **Estimate > Use the following subgroups...** and enter 1 : 20. The author prefers always to use the latter method.

The first signal appears from the sample taken during week 35. Reference to the list of tests in Appendix 4 indicates that a signal arising from Test 2 results from the occurrence of nine consecutive points on the same side of the centre line. These points have been ringed in the plot, and on moving the mouse pointer to the label 2 beside the ninth point the sample number and the test failed are displayed. Thus the chart provides evidence that the process changes have led to improvement, in the form of a reduction in the proportion of inaccurate invoices. Thus it is appropriate to introduce an additional column named Phase in the worksheet to indicate the two phases of operation of the process, before and after changes.

**Figure 5.22** Extended P chart of invoice data.

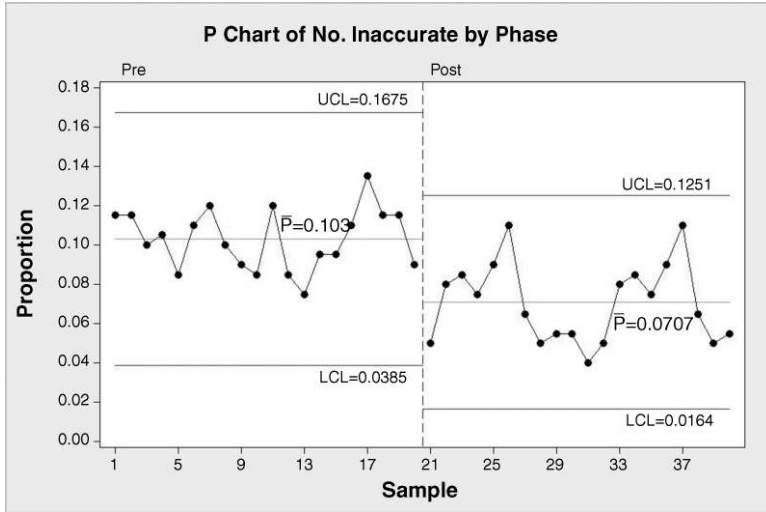


Figure 5.23 P charts for the two phases.

The first 20 cells in the column named Phase could contain the text Pre and the remaining 20 the text Post. This may be achieved using **Calc > Make Patterned Data > Text Values...** with **Store patterned data in:** Phase, **Text values:** Pre Post, **Number of times to list each value:** 20, and **Number of times to list the sequence:** 1. In order to create the revised chart displayed in Figure 5.23 all the data values were used in the computation of limits. Again this may be specified in two ways: either **Estimate > Omit the following subgroups...** and leave the window blank, or **Estimate > Use the following subgroups...** and enter 1:40. In addition, under **P Chart Options...** > **Stages** one has to enter Phase under **Define stages (historical groups) with this variable:**.

Figure 5.23 indicates that the changes have reduced the proportion of inaccurate invoices to around 7%. It also indicates that the process is behaving in a stable, predictable manner following the changes and that the second chart could be adopted for further monitoring.

Clearly there is room for further improvement. Let us suppose that at a later date the proportion has dropped to around 2%, with mean proportion for a series of 25 samples being 0.018. The calculation of the Lower Chart Limit is shown in Box 5.6.

A negative proportion is impossible so there is, strictly speaking, no lower control limit on the P chart with subgroup size 200. However, Minitab inserts a horizontal line at zero on the

$$\begin{aligned}
 LCL &= \bar{p} - 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}} \\
 &= 0.018 - 3\sqrt{\frac{0.018 \times 0.982}{200}} = 0.018 - 0.028 = -0.010
 \end{aligned}$$

Box 5.6 Calculation yielding a negative lower chart limit.

$$n > \frac{9(1 - \bar{p})}{\bar{p}}$$

Box 5.7 Criterion to ensure a nonzero lower limit on a P chart.

P chart in such cases, labelled $LCL = 0$. With no lower limit, the possibility of evidence of a further drop in the proportion of nonconforming invoices being signalled by a point falling below the lower limit is not available. To reinstate this option the sample size can be increased. Some mathematical manipulation shows that, to ensure a lower limit exists on a P chart with three-sigma limits, the inequality in Box 5.7 must be satisfied. For $\bar{p} = 0.018$, the formula gives $n > 491$. Thus, once the monthly proportion of nonconforming invoices had dropped to around 2%, monthly samples of, say, 500 invoices would provide the opportunity to detect further improvement through a signal from a point on the chart falling below the lower limit.

In some situations it is not possible to have constant sample size but it is still possible to create a P chart. The chart limits are no longer horizontal parallel lines but have a stepped appearance, the limits being closer together for larger samples and wider apart for smaller samples. As an example, consider the data in Table 5.5 giving monthly admissions of stroke patients to a major hospital together with the numbers of those patients treated in the acute stroke unit.

The data were set up in three columns in the supplied Minitab worksheet ASU.MTW. The first contains the month in which the data were collected in the date format Jan-02, Feb-02 etc. and was set up using **Calc > Make Patterned Data > Simple Set of Date/Time Values...** with **Patterned Sequence** specified as **From first date/time:** Jan-02 **To last date/time:** Dec-03 **In steps of:** 1 with **Step unit:** Month and defaults otherwise. The second column contained the monthly counts of patients admitted with a diagnosis of stroke, and the third the number of those patients who receive treatment in the acute stroke unit.

Part of the dialog involved in creating a P chart for the proportion of patients receiving treatment in the acute stroke unit is shown in Figure 5.24. **Subgroup size:** is specified by selecting the column named Strokes. Use of the **Scale...** facility enables the horizontal axis of the chart to be 'stamped' with the months when the data were collected. Limits were based on the first 15 observations as changes were made to the process of managing stroke patients at the hospital at the end of March 2003, corresponding to observation 15. The chart is displayed in Figure 5.25. The centre line is placed at 0.6352 and labelled $\bar{P} = 0.6352$. However, this value is not the mean of the first 15 proportions; it is the total number of stroke patients receiving ASU care for the first 15 months (477) divided by the total number of stroke patients admitted in the

Table 5.5 Monthly stroke admissions.

2002	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Strokes	47	58	35	49	58	56	50	45	51	53	61	47
ASU	31	34	29	28	30	35	31	37	26	33	37	32
2003	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Strokes	51	38	52	43	43	44	49	42	38	58	39	36
ASU	32	32	30	34	31	28	32	37	32	41	26	31

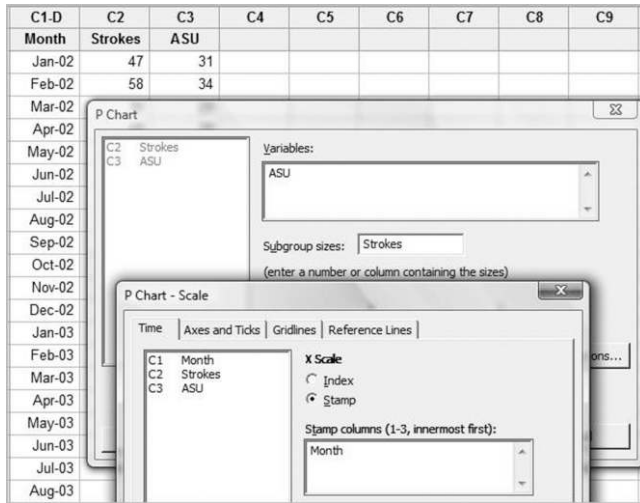


Figure 5.24 Creation of a P chart with variable sample size.

first 15 months (751). In the case of constant sample size the result of this calculation is the same as the result of taking the mean of the corresponding proportions. The UCL and LCL values displayed apply to the final sample that had size 36. The reader is invited to use the formulae in Box 5.5 to confirm the LCL of 0.3945 and UCL of 0.8758 displayed.

From the signals on the chart it would appear that the process changes have led to a greater proportion of stroke patients receiving acute stroke unit care. (Of course the theory underlying the P chart is based on the binomial distribution for which the probability of care in the acute stroke unit would remain constant from patient to patient. In reality this is unlikely to be the case. Wheeler and Poling (1998, pp. 182–184) and Henderson *et al.* (2008) refer to this issue.

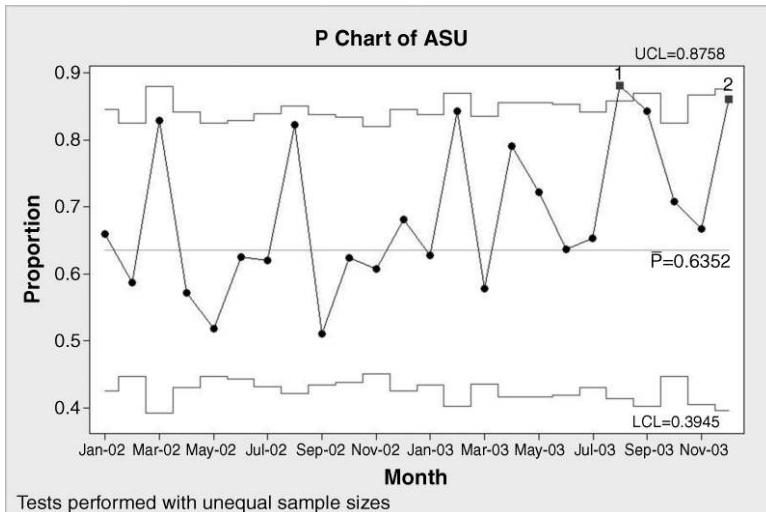


Figure 5.25 P chart with variable subgroup size.

However, the chart does display the data in an informative way and methods presented in Chapter 7 can be used to test formally whether or not the proportion of patients receiving acute stroke unit care has increased.)

The assumptions underlying valid use of a P chart are as follows:

1. Samples of n (not necessarily a constant) items provide the areas of opportunity.
2. Each item may be classified to either possess, or not possess, an attribute. Usually the attribute is nonconformance with specifications.
3. The probability, p , that an item possesses the attribute of interest is constant.
4. The status of any item with regard to possession of the attribute is independent of that of any other item.

5.2.2 NP chart for number nonconforming

The number defective or NP chart is exactly equivalent to the P chart, the only difference being that the number defective is plotted instead of the proportion defective. The NP charts for the number of inaccurate invoices data in Tables 5.3 and 5.4 in the previous section are displayed in Figure 5.26. Note that this is simply a scaled version of the chart in Figure 5.23 – e.g. the upper chart limit in the post-change phase in Figure 5.26 is 200 times the upper chart limit in the post-change phase on the P chart in Figure 5.23.

The author believes that, since the NP chart plots the number of defective items rather than the proportion of defective items, it is less directly informative than the P chart. One advantage of the NP chart over a P chart is that it is much simpler to update a pencil and paper version of an NP chart as no calculation is required – the count of defective items in the sample is plotted directly on to the chart. The underlying assumptions for valid use of an NP chart are the same as for the P chart.

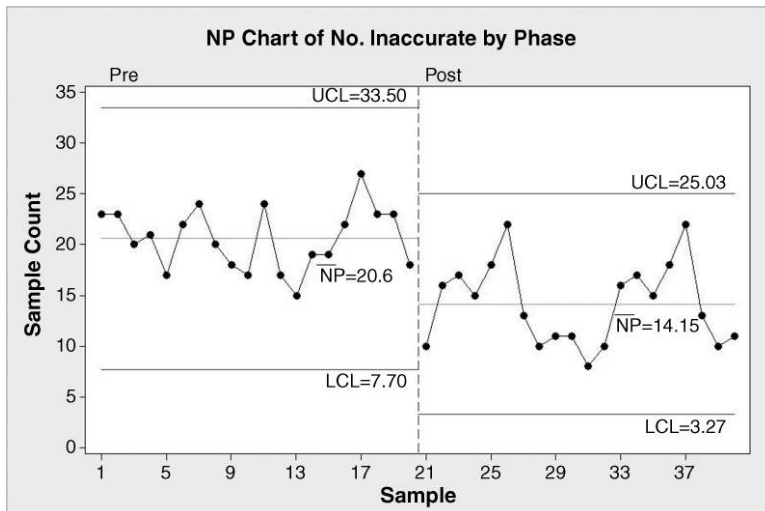


Figure 5.26 NP charts for the two phases.

Table 5.6 Counts of nonconformities in ATMs.

ATM	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
No. of nonconformities	5	4	7	9	4	6	5	8	9	11	5	10	6	6	5
ATM	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
No. of nonconformities	4	7	10	6	9	8	8	4	8	8	4	4	7	3	12

5.2.3 C chart for count of nonconformities

The C chart is used to plot the count of defects/nonconformities in equal ‘areas of opportunity’ for these to occur. These ‘areas of opportunity’ may be in time, space or segments of product. The number of yarn breakages per hour on a monofilament spinning frame, the number of nonconformities (imperfect solder joints, missing or damaged components etc.) on a printed circuit board taken at regular intervals from a process are respectively time and space examples.

Table 5.6 gives counts of the number of nonconformities detected during verification of automatic telling machines of a particular type. The data are in time sequence and are available in ATM.MTW. A C chart of the data is shown in Figure 5.27 created using **Stat > Control Charts > Attributes Charts > C...** with the counts from the above table previously entered into a column named No. Nonconformities. The chart limits calculations are shown in Box 5.8. Tests 1 to 4 inclusive (see Appendix 4) are available for the C chart in Minitab. These were all applied in the creation of the chart in Figure 5.27. There is no evidence from the chart of any special cause affecting the process. Thus the chart could be employed for further monitoring of the process.

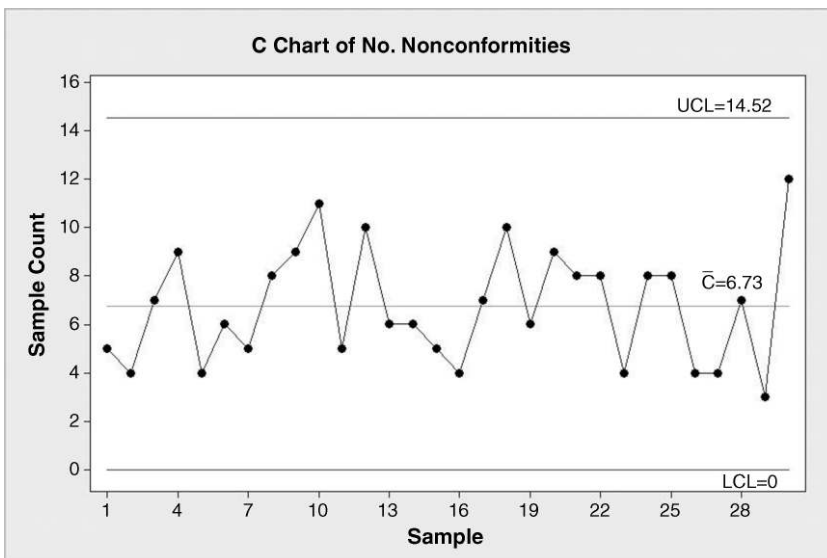


Figure 5.27 C chart of ATM nonconformity counts.

The upper chart limit is given by

$$\begin{aligned} \text{UCL} &= \bar{p} + 3\sqrt{\bar{p}} \\ &= 6.733 + 3 \times 2.595 = 14.52. \end{aligned}$$

The lower chart limit is

$$\begin{aligned} \text{LCL} &= \bar{c} - 3\sqrt{\bar{c}} \\ &= 6.733 - 3 \times 2.595 = -1.51. \end{aligned}$$

As a count of nonconformities can never be negative, Minitab sets the LCL to 0.

Box 5.8 Calculation of limits for a C chart.

The assumptions underlying valid use of a C chart are as follows:

1. The counts are of nonconformities or events.
2. The nonconformities or events occur in a defined region of space or period of time or segment of product referred to as the area of opportunity.
3. The nonconformities or events occur independently of each other, and the probability of occurrence of a nonconformity or event is proportional to the size of the area of opportunity.

5.2.4 U chart for nonconformities per unit

The U chart may be employed when counts of nonconformities are made over a number of units of product. The worksheet *Faults.MTW* contains the number of faults detected in each of 30 consecutive hourly samples of 40 retractable single-use syringes from a pilot manufacturing process. A U chart of the data is shown in Figure 5.28 created using **Stat > Control**

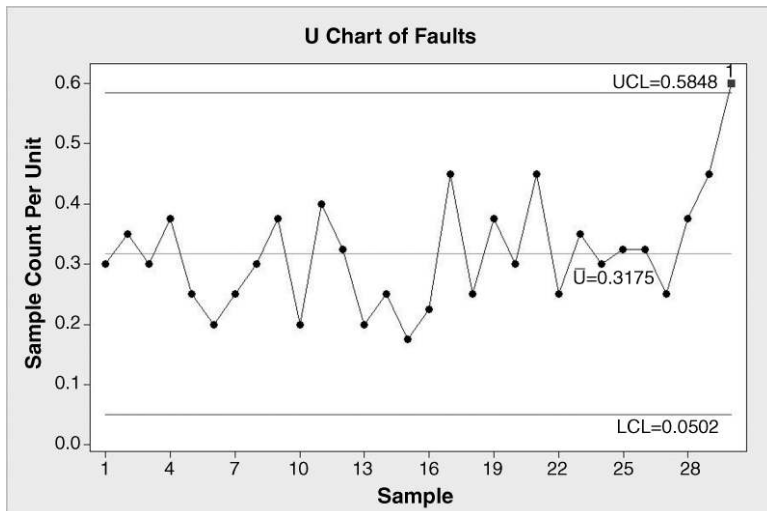


Figure 5.28 U chart of faults per syringe.

Charts > Attributes Charts > U... Tests 1 to 4 inclusive (see Appendix 4) are available for the U chart in Minitab. These were all applied in the creation of the chart in Figure 5.28. There is evidence from the chart of a special cause affecting the process since the last point is above the upper chart limit. The sample size for a U chart can be variable and Minitab enables U charts with variable sample size to be created in the same way as in the case of the P chart. A column that indicates the sample sizes has to be specified in the **Subgroup sizes:** window in the dialog, as was done in the dialog displayed in Figure 5.24 for the chart in Figure 5.25.

Reference to the Assistant flow chart for Shewhart control chart selection that is displayed in Figure 5.1 reveals no reference to the C chart. However, a C chart is the special case of the U chart with constant size 1 for all subgroups. The reader is invited to re-create the C chart in Figure 5.27 using the U chart facility.

Readers wishing to construct control charts for attribute data without using Minitab may find the formulae in Appendix 3 of value. This completes the material on Shewhart control charts in this book. Montgomery (2009, pp. 330–344) provides guidelines for the implementation of these charts. After a brief discussion of funnel plots, we turn our attention to time-weighted control charts.

5.2.5 Funnel plots

Although not strictly control charts, funnel plots will be included here because the underlying statistical modelling is identical to that for the P chart. In order to introduce the funnel plot, consider the situation where a customer has records of counts of nonconforming units for a number of suppliers as shown in Table 5.7. (Having the data in time sequence is crucial for the correct use of control charts, but is not so for funnel plots; indeed the data for a funnel plot typically applies to the same time period.)

The funnel plot, with three sigma limits, is shown in Figure 5.29. The plot gets its name from the funnel-like shape of the curves defining the limits. The proportions for each supplier are plotted against the number of units tested. The centre line corresponds to the mean proportion nonconforming across all ten suppliers of $\bar{p} = 228/2000 = 0.114$, approximately 11%. The three-sigma limits for a supplier are calculated using the formula $\bar{p} \pm 3\sqrt{\bar{p}(1-\bar{p})/n}$, where n is the number of units tested for that supplier. Should limit values greater than 1 or negative values be obtained then they should be set to 1 or 0, respectively.

Table 5.7 Records of nonconforming units for ten suppliers.

Supplier	Units tested	No. nonconforming
A	200	19
B	150	41
C	60	8
D	400	48
E	80	11
F	250	13
G	160	19
H	200	18
I	360	38
J	140	13

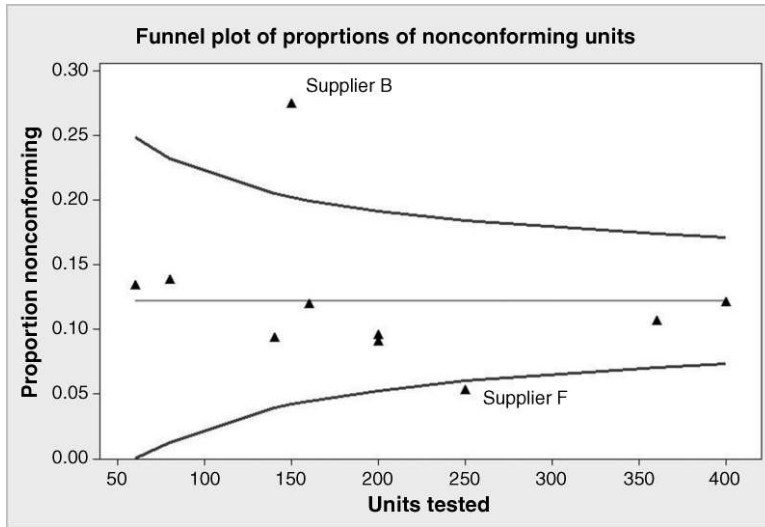


Figure 5.29 Funnel plot of proportions of nonconforming units.

Suppliers B and F plot outside the limits. As supplier B falls above the upper limit the chart provides evidence that supplier B produces nonconforming units at a significantly higher rate (27.3%) than the overall rate of 11.4%. Similarly, there is evidence that supplier F performs significantly better (3.6%) than the overall rate of 11.4%.

Spiegelhalter (2002) discusses the use of funnel plots for institutional comparisons in healthcare and also calculation of the limits using the binomial probability distribution rather than the normal approximation method used above. League tables are often produced when comparing performance across institutions, but some argue that identifying the institutions that stand out from the crowd via a funnel plot analysis and then investigating the performance of these can lead to insights that lead to quality improvement. Minitab does not have a facility for the direct creation of funnel plots. A follow-up exercise is provided and details of how a funnel plot may be created using Minitab are provided in the notes on the exercise on the book's website.

5.3 Time-weighted control charts

5.3.1 Moving averages and their applications

Time-weighted control charts plot information derived not only from the most recent sample obtained but also from the most recent and earlier samples. Two types of time-weighted control charts will be discussed: the exponentially weighted moving average (EWMA) chart and the cumulative sum (CUSUM) chart.

Before considering the EWMA chart, moving averages will be introduced.

Consider daily sales of pizzas at Halcro Snacks, which operates a fast-food kiosk, on Mondays to Fridays inclusive each week, in a business park. The data (Table 5.8) are available in Pizzas.xls.

Having set up the data in two columns of a worksheet, **Graph > Time Series Plot... > Simple** may be used to display the data. The **Time/Scale...** button may be used to **Stamp**

Table 5.8 Daily sales data.

Day	Sales	Day	Sales
5-Jul-10	54	19-Jul-10	63
6-Jul-10	93	20-Jul-10	110
7-Jul-10	55	21-Jul-10	70
8-Jul-10	59	22-Jul-10	83
9-Jul-10	143	23-Jul-10	177
12-Jul-10	86	26-Jul-10	65
13-Jul-10	95	27-Jul-10	112
14-Jul-10	58	28-Jul-10	66
15-Jul-10	75	29-Jul-10	56
16-Jul-10	146	30-Jul-10	163

the horizontal time axis with Day, selected in the **Stamp columns:** window, as shown in Figure 5.30.

A cyclical pattern is evident in the level of daily sales, with Friday having the highest level each week. A moving average of length 5 is a natural way to summarize the data in view of the five-day operation. (In this context ‘average’ implies the mean.) Each consecutive set of five daily sales data includes data for a Monday, Tuesday, Wednesday, Thursday and Friday. By the end of the first five days of trading, total daily sales were

$$54 + 93 + 55 + 59 + 143 = 404,$$

so the first moving average of length 5 is

$$54/5 + 93/5 + 55/5 + 59/5 + 143/5 = 404/5 = 80.8.$$

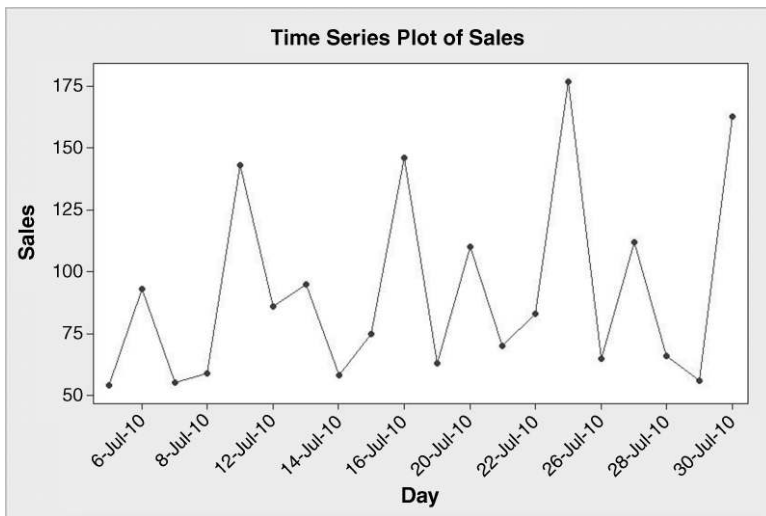


Figure 5.30 Time series plot of sales data.

Day	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Weight	0.2	0.2	0.2	0.2	0.2															
Sales	54	93	55	59	143	86	95	58	75	146	63	110	70	83	177	65	112	66	56	163
MA (Length 5)	*	*	*	*	80.8															
etc.																				
Day	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Weight		0.2	0.2	0.2	0.2	0.2										0.2	0.2	0.2	0.2	0.2
Sales	54	93	55	59	143	86	95	58	75	146	63	110	70	83	177	65	112	66	56	163
MA (Length 5)	*	*	*	*	80.8	87.2	87.6	88.2	91.4	92	87.4	90.4	92.8	94.4	100.6	101	101.4	100.6	95.2	92.4

Figure 5.31 Schematic for calculation of moving averages.

An equivalent way to obtain this is to multiply each one of five consecutive daily sales counts by 0.2 and sum the five products. The set of five factors 0.2, 0.2, 0.2, 0.2 and 0.2 are known as **weights**. The set of weights for calculation of a moving average must sum to 1. (With quarterly data, for example, one could employ a moving average of length 4 with weights 0.25, 0.25, 0.25 and 0.25.) The first two and the final moving average calculations are displayed in schematic form in Figure 5.31.

The first moving average becomes available on day 5 and the final one on day 20. They can be readily calculated and plotted in Minitab. Use **Stat > Time Series > Moving Average . . .**, select Sales as the **Variable**: and specify the **MA length**: as 5. Under **Graphs . . .** check **Plot smoothed vs. actual**. Under **Storage . . .** check **Moving averages**, accept defaults otherwise and click **OK, OK**. The plot in Figure 5.32 is obtained.

In Figure 5.31 the first moving average value of 80.8 appears below 5 in the row giving the day number. This moving average is plotted (default in Minitab) against the 5th day. The moving average provides a smoothing technique for data in the form of time series. Scrutiny of the moving average plot enables any underlying trend or other major pattern in the series to be

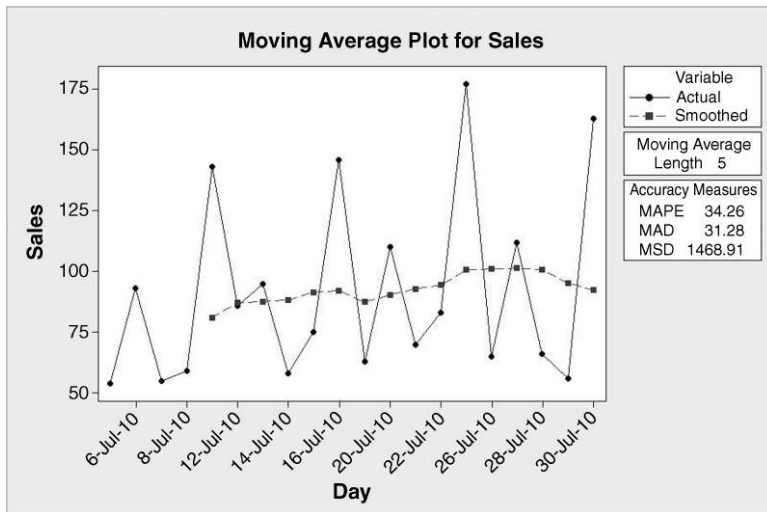


Figure 5.32 Time series plot of sales with moving average of length 5.

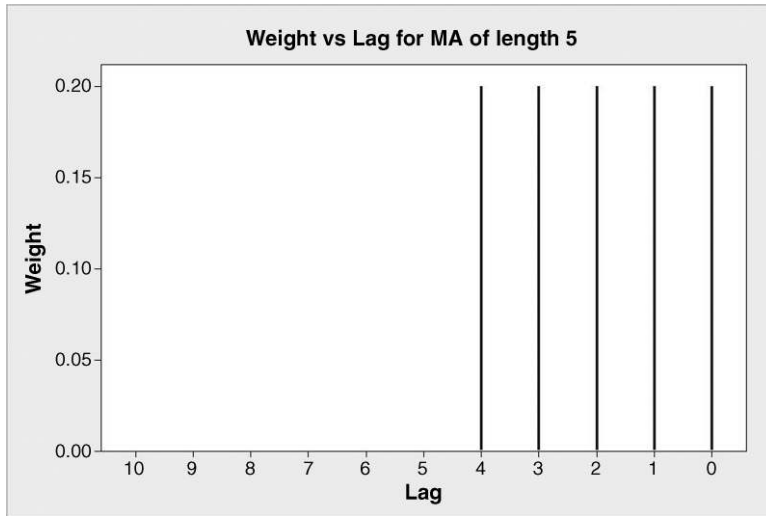


Figure 5.33 Weights for moving average of length 5.

identified. In this case the plot of the moving average plot indicate that sales appear to follow an initial upward trend, with a suggestion of a downward trend latterly. The accuracy measures in the panel to the right of the plot will be explained later in the section.

As far as the calculation of the moving average values is concerned, the reader is invited to visualize the shaded template in Figure 5.31 being moved from one group of five sales figures to the next. With each group, the sales figures are multiplied by the adjacent weights and the products summed to give the moving average value. The current observation may be considered to have lag 0, the immediately prior observation lag 1 and so on. A plot of the weights versus lag for the moving average of length 5 is given in Figure 5.33. The scale has been reversed as a lag value of 1 means one step back in time, i.e. in the negative direction on a conventional horizontal axis.

Consider now the situation where, in calculating a moving average, the most recently observed value is assigned weight α , the observation prior to that weight $\alpha(1 - \alpha)$, the observation prior to that weight $\alpha(1 - \alpha)^2$ and so on, where the number α is selected such that $0 < \alpha \leq 1$. The sum of an infinite sequence of weights of this type is 1. In the case where $\alpha = 0.4$, for example, the sequence of weights would be 0.400, 0.240, 0.144, 0.086, . . . A plot of weight versus lag is shown in Figure 5.34. Minitab refers to α as the *weight*, but the term *smoothing constant* is also widely used.

The weights form a geometric series and decrease exponentially. The more recent the observation then the greater the influence, or weight, it has in the calculation of the moving average. A moving average of this type is referred to as an *exponentially weighted moving average* (EWMA). The direct calculation of some EWMA is displayed in Figure 5.35.

Each weight, as we move further back in time, is 0.6 times the previous one in this case. This factor of 0.6 is known as the *discount factor* $\theta = 1 - \alpha$. Minitab can be used to calculate and plot the EWMA, which may also be referred to as the smoothed value. Use **Stat > Time Series > Single Exp Smoothing . . .**, select **Sales** as the **Variable:** and specify the **Weight to Use in Smoothing** by Use: 0.4 (the smoothing parameter, denoted by Alpha in

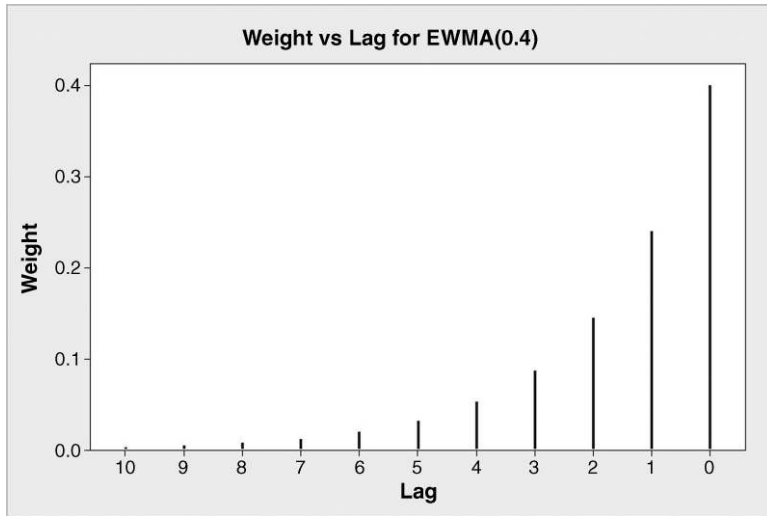


Figure 5.34 Weights for an exponentially weighted moving average.

Minitab). Values for the smoothing parameter usually lie between 0 and 1. Under **Graphs...** select **Plot smoothed vs. actual**. Under **Storage...** check **Smoothed data**, under **Time...** enter **Stamp: Day** and click **OK, OK**. The plot in Figure 5.36 is obtained. Again the plot of the moving average indicates that Sales appear to follow an initial upward trend with a suggestion of a downward trend latterly.

The calculation, in practice, of EWMA's (the smoothed values) in Minitab is outlined in Box 5.9. These first two smoothed values may be confirmed from the column of smoothed data created in the worksheet and named SMOO1 by the software. Note, too, that the final three smoothed values agree with those obtained in the schematic in Figure 5.37 (Under **Options...** the user may change the number of observations used to calculate a smoothed value for time 0 from the default number of six.)

Day	...	11	12	13	14	15	16	17	18	19	20
Weight	...	0.004	0.007	0.011	0.019	0.031	0.052	0.086	0.144	0.240	0.400
Sales	...	63	110	70	83	177	65	112	66	56	163
EWMA(0.4)											110.7

Day	...	11	12	13	14	15	16	17	18	19	20
Weight	...	0.007	0.011	0.019	0.031	0.052	0.086	0.144	0.240	0.400	
Sales	...	63	110	70	83	177	65	112	66	56	163
EWMA(0.4)										75.8	110.7

Day	...	11	12	13	14	15	16	17	18	19	20
Weight	...	0.011	0.019	0.031	0.052	0.086	0.144	0.240	0.400		
Sales	...	63	110	70	83	177	65	112	66	56	163
EWMA(0.4)									88.9	75.8	110.7

etc.

Figure 5.35 Schematic for calculation of an EWMA.

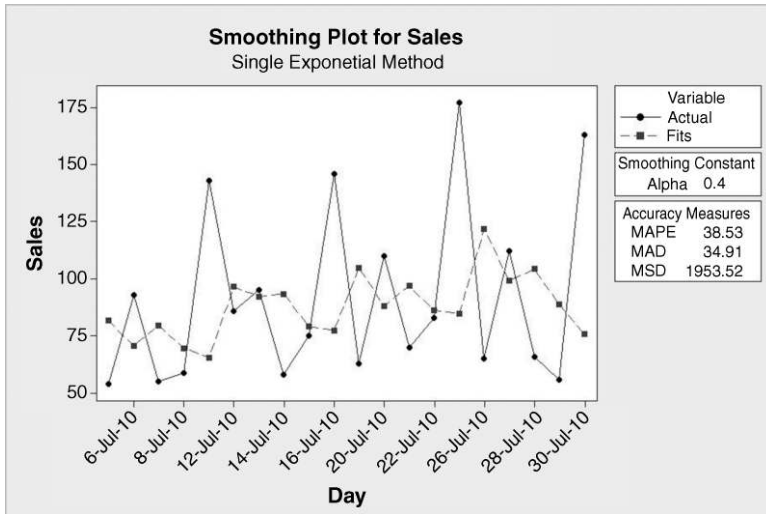


Figure 5.36 Plot of EWMA and original data.

The cyclical nature of the sales data made the choice of length 5 for the simple moving average a natural one. An arbitrary choice of 0.4 was made for the smoothing constant in order to introduce the concept of the exponentially weighted moving average. So far we have considered the moving average as a means of smoothing time series data. The moving average evaluated at the current point in time may be used as a forecast of the value at the next point in

For single exponential smoothing it may be shown that the smoothed value at time t is given by

$$\alpha \times (\text{Data value at time } t) + (1 - \alpha)(\text{Smoothed value at time } t - 1).$$

The observed data values are considered to be at times 1, 2, 3, . . . , so to ‘kick-start’ the calculations a smoothed value for time $t = 0$ is required. The default in Minitab is to take the mean of the first six data values to be the smoothed value at time $t = 0$. From Table 4.8 it may be verified that this value is 81.667. Now the smoothed value at time 1 is

$$\begin{aligned} &\alpha \times (\text{Data value at time 1}) + (1 - \alpha)(\text{Smoothed value at time 0}) \\ &= 0.4 \times 54 + 0.6 \times 81.667 \\ &= 70.600, \end{aligned}$$

and the smoothed value at time 2 is

$$\begin{aligned} &\alpha \times (\text{Data value at time 2}) + (1 - \alpha)(\text{Smoothed value at time 1}) \\ &= 0.4 \times 93 + 0.6 \times 70.600 \\ &= 79.560, \end{aligned}$$

etc.

Box 5.9 Calculation of exponentially weighted moving averages.

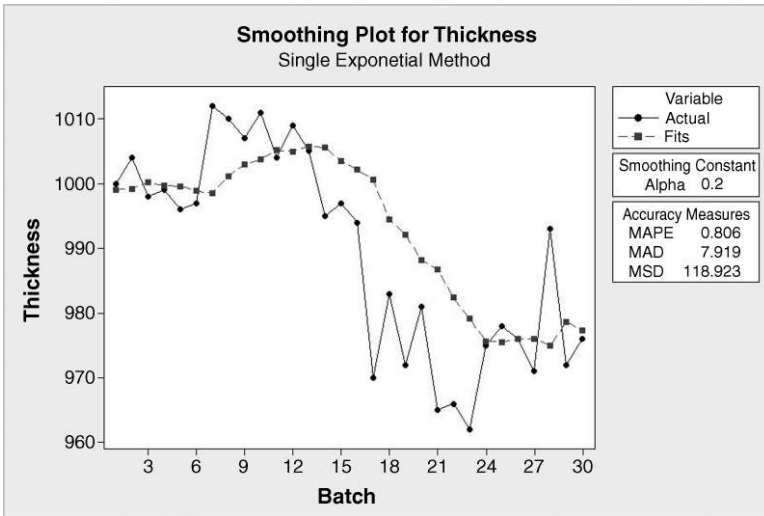


Figure 5.37 Forecasts of Thickness.

time, i.e. as a ‘one-period-ahead forecast’. By way of illustration, consider the data in Table 5.9 and worksheet Thickness.MTW, giving thickness (in angstroms, Å) of a nitride layer measured at a fixed location on the first wafer in each one of 30 successive batches of wafers from a microelectronics fabrication process. In terms of quality, the ability to forecast is of value – if the forecast thickness for the next batch does not lie within specifications for thickness, then there is the possibility of taking some action to adjust the process in some way to ensure that thickness for the next batch is satisfactory.

With the data set up in two columns, Batch and Thickness, use **Stat > Time Series > Single Exp Smoothing . . .**, select Thickness in **Variable:** and specify the **Weight to Use in Smoothing** as **Use: 0.2** (the default value for the smoothing parameter in Minitab). A title can be created under **Options . . .** Under **Time . . .** select **Stamp** and specify Batch. Under **Graphs**, check **Plot predicted vs. actual**. Under **Storage . . .** check **Fits (one-period-ahead forecasts)** and **Residuals**. The plot in Figure 5.37 is obtained.

The lowest panel to the right of the plot gives three measures of accuracy for the forecasts. The first few rows of the worksheet, which includes the stored fits (one-period-ahead forecasts) and residuals and four additional columns, calculated by the author in order to indicate how the measures of accuracy are computed, are displayed in Figure 5.38.

The error (or deviation or residual – yes, all three names appear in this context in Minitab!) is the actual value observed minus the forecast (fitted value) made of that value, that is to say, $RESIDUAL = DATA - FIT$; this is a simple rearrangement of the formula $DATA = FIT + RESIDUAL$. Thus, for example, for batch 11 the error is $1004 - 1005.2114 = -1.2114$. This

Table 5.9 Thickness data for nitride layer.

Batch	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Thickness	1000	1004	998	999	996	997	1012	1010	1007	1011	1004	1009	1005	995	997
Batch	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Thickness	994	970	983	972	981	965	966	962	975	978	976	971	993	972	976

THICKNESS.MTW ***								
↓	C1	C2	C3	C4	C5	C6	C7	C8
	Batch	Thickness	FITS1	RES11	% Error	APE	AD	SD
1	1	1000	999.0000	1.0000	0.10000	0.10000	1.0000	1.000
2	2	1004	999.2000	4.8000	0.47809	0.47809	4.8000	23.040
3	3	998	1000.1600	-2.1600	-0.21643	0.21643	2.1600	4.666
4	4	999	999.7280	-0.7280	-0.07287	0.07287	0.7280	0.530
5	5	996	999.5824	-3.5824	-0.35968	0.35968	3.5824	12.834
6	6	997	998.8659	-1.8659	-0.18715	0.18715	1.8659	3.482
7	7	1012	998.4927	13.5073	1.33471	1.33471	13.5073	182.446
8	8	1010	1001.1942	8.8058	0.87186	0.87186	8.8058	77.542
9	9	1007	1002.9554	4.0446	0.40165	0.40165	4.0446	16.359
10	10	1011	1003.7643	7.2357	0.71570	0.71570	7.2357	52.356
11	11	1004	1005.2114	-1.2114	-0.12066	0.12066	1.2114	1.468
12	12	1009	1004.9691	4.0309	0.39949	0.39949	4.0309	16.248

Figure 5.38 Columns C5–C8 indicate computation of accuracy measures.

error as a percentage of the observed thickness for batch 11 of 1004 gives the percentage error as -0.12066% . APE in Figure 5.38 is the absolute percentage error, AD the absolute deviation (error) and SD the squared deviation (error). Mean APE (MAPE), mean AD (MAD) and mean SD (MSD) may all be used as measures of forecast accuracy and are displayed to the right of the plot in Figure 5.37. Interested readers are invited to calculate the entries displayed in columns C5–C8 and to verify the values for MAPE, MAD and MSD displayed in Figure 5.37.

Forecasts were also generated using values 0.4, 0.6 and 0.8 for the smoothing constant alpha. The accuracy measures obtained with the four values for the smoothing constant are displayed in Table 5.10. Of the four smoothing parameters tested, an alpha of 0.6 performs best in that it gives the lowest values for MAPE, MAD and MSD. Minitab also provides a procedure for the selection of an optimal value for the smoothing constant based on the fitting of an autoregressive integrated moving average (ARIMA) time series model to the data. (These models are not considered in this book.) This procedure was implemented by checking **Optimal ARIMA as Weight to Use in Smoothing**. For the thickness data this yields an alpha of 0.5855. In addition, **Generate forecasts** was checked with **Number of forecasts:** set to 1 and **Starting from origin:** 30 specified. The corresponding plots are shown in Figure 5.39.

Comparison of this plot with the one in Figure 5.37 reveals the superior performance of the smoothing parameter 0.5855 over that of the smoothing parameter 0.2. The reader should note

Table 5.10 Accuracy measures for a series of values of the smoothing constant.

Smoothing constant α	MAPE	MAD	MSD
0.2	0.806	7.92	118.9
0.4	0.684	6.73	82.8
0.6	0.650	6.40	76.5
0.8	0.662	6.51	82.0

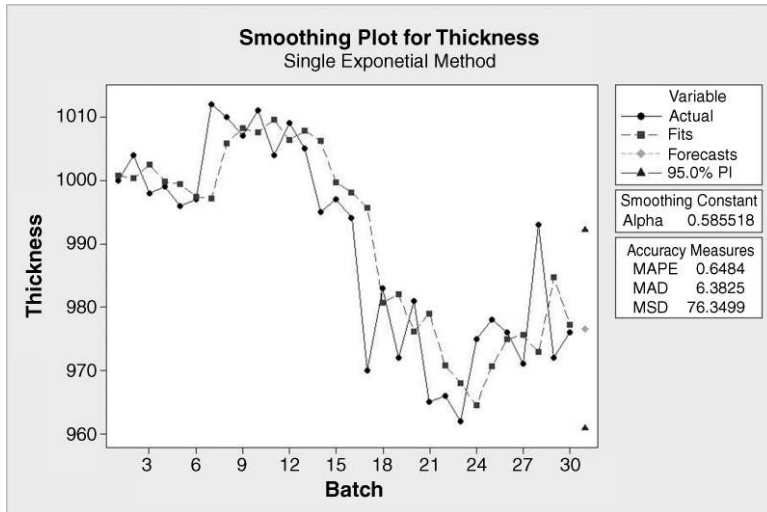


Figure 5.39 Forecasts with optimal alpha.

that the accuracy measures displayed in the panel to the right of the plot are marginally lower than those listed in Table 5.10 for smoothing constant 0.6. The diamond shaped symbol on the right of the plot represents the forecast thickness of 976.5 for the 31st batch. The triangular symbols define lower and upper 95% prediction limits of 960.9 and 992.1 for the forecast. These limits are such that 95 times out of 100 in the long term the actual observation will lie in the interval. The forecast and the prediction limits are displayed in the Session window.

In addition to their application for smoothing and forecasting, moving averages may be plotted in control charts. As moving averages are linear combinations of random variables the results on linear combinations of random variables presented in Chapter 4 may be used to obtain formulae for standard deviations of moving averages. The mathematics will not be presented. Details may be found in Montgomery (2009, pp. 419–430). Only the EWMA control chart will be considered in this book.

5.3.2 Exponentially weighted moving average control charts

The EWMA control chart performs well at detecting small changes in a process mean. For example, for a normally distributed random variable, the average run length (ARL) for detection of a one standard deviation shift in the process mean using an individuals chart is approximately 44. For an EWMA chart with smoothing constant 0.4 the ARL for detection of such a shift is approximately 11. However, the EWMA chart does not detect relatively large shifts in a process mean as quickly as a Shewhart chart. Hunter (1989, pp. 13–19) demonstrated that the EWMA chart with smoothing constant 0.4 performs similarly to the Shewhart chart employing Minitab tests 1, 2 (with eight points in a row employed rather than the default nine), 5 and 6, i.e. the Western Electric Rules. A discussion can be found in Montgomery (2009, pp. 422–424).

It can be shown that the standard deviation of the EWMA is given by the expression in Box 5.10. For the temperature data of Table 5.1 the estimates of process mean and standard deviation, based on the first 20 data values, were 27.245 and 0.699 9, respectively.

The standard deviation of the EWMA is

$$\sigma \sqrt{\frac{\alpha}{2 - \alpha} [1 - (1 - \alpha)^{2i}]},$$

where i is the sample number and α is the smoothing parameter. In the long term the value of $(1 - \alpha)^{2i}$ becomes negligible since $0 < \alpha < 1$ and the standard deviation is effectively

$$\sigma \sqrt{\frac{\alpha}{2 - \alpha}}.$$

Box 5.10 Standard deviation of EWMA.

$$27.245 \pm 3 \times 0.6999 \sqrt{\frac{0.4}{2 - 0.4}} = 27.245 \pm 1.04985$$

$$\text{UCL} = 28.295$$

$$\text{LCL} = 26.195$$

Box 5.11 Calculation of long-term EWMA chart limits.

The long-term three-sigma limits for an EWMA control chart with smoothing parameter 0.4, for example, could therefore be calculated as shown in Box 5.11.

When all 25 data values (available in Temperature.MTW) have been entered into a column, to create the chart in Minitab use is made of **Stat > Control Charts > Time-Weighted Charts > EWMA...** Under **Subgroup sizes:** 1 was specified. The **Weight of EWMA:** is the smoothing constant and 0.4 was selected. Under **EWMA Options...** and **Estimate,** use of samples 1 : 20 and the **Rbar** method were indicated so that the estimates of process mean and standard deviation would be the same as those used in the setting up of the individuals chart of the data earlier. The chart in Figure 5.40 was obtained. As with the individuals chart in Figure 5.3 the 25th point is below the lower chart limit, thus providing evidence of a potential special cause affecting the process. (The reader should note the narrower limits for the first few samples.)

Consider again the sequence of weights $\alpha, \alpha(1 - \alpha), \alpha(1 - \alpha)^2, \alpha(1 - \alpha)^3, \dots$ used in exponential smoothing. With $\alpha = 1$ these weights become 1, 0, 0, 0 ... so that the EWMA would simply be the most recent observation. Just for fun, the reader is invited to create the EWMA chart of temperature as in Figure 5.40 but with smoothing constant 1 and to check that it is identical to the Shewhart chart in Figure 5.4.

Further simulated data for the rod cutting process described in Section 5.1 will be used to illustrate the use of the EWMA for data collected in subgroups. In the scenario illustrated in Figure 5.41 the process mean was changed from 60.000 to 60.008 after sample 40. The procedure computes the sample means and applies the EWMA methodology to the sample means. The subgroups comprised the rows of columns x1, x2, x3 and x4. Observe that the EWMA chart, with limits smoothing constant 0.2, signalled this process change at sample number 54. (Limits were based on the first 25 samples and the default pooled standard deviation

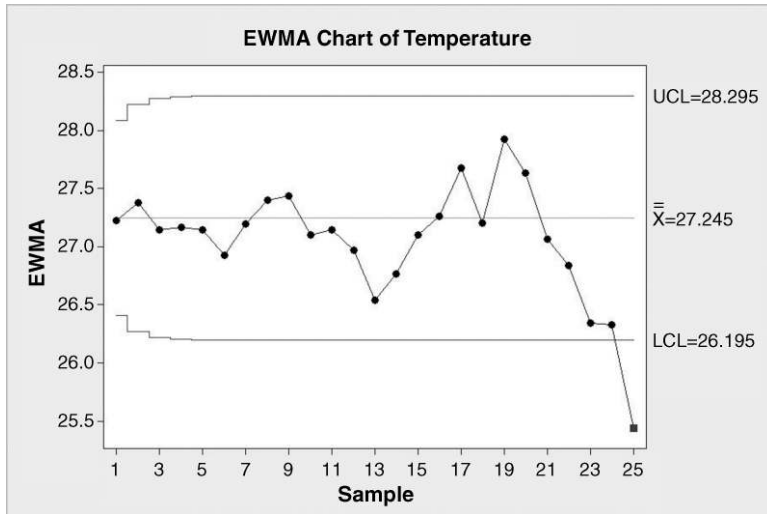


Figure 5.40 EWMA control chart of temperature.

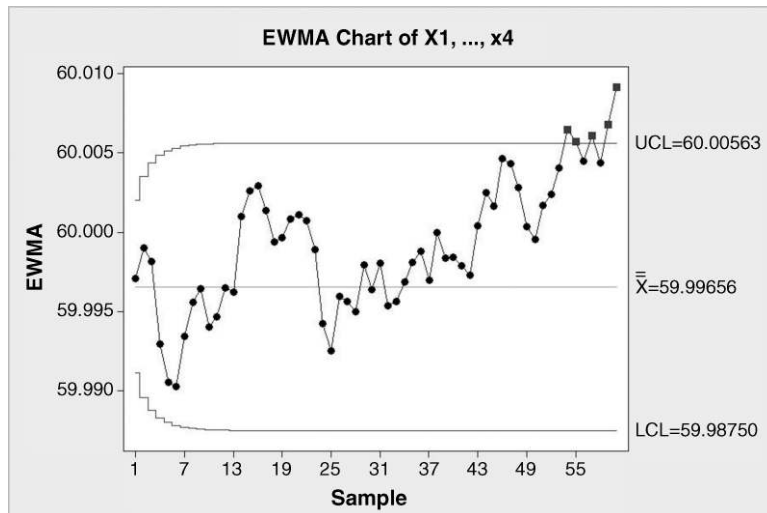


Figure 5.41 EWMA control chart for rod length.

with unbiasing constant method of estimation of process standard deviation was used.) The data are available in RodsScenario5.MTW. The reader is invited to verify that an Xbar chart also signals the process change at sample number 59 via Test 2. This example illustrates the ability of the EWMA chart to detect a small shift in the process mean earlier than a Shewhart chart.

5.3.3 Cumulative sum control charts

The most recently plotted point in a Shewhart control chart contains only information from the most recent sample. The latest plotted point in a control chart of moving averages of length 5

Table 5.11 CUSUM for yield.

Run No. i	Yield	Target	Deviation	CUSUM S_i
0	*	*	*	0
1	115	100	15	15
2	95	100	-5	10
3	110	100	10	20
4	105	100	5	25
5	95	100	-5	20
6	100	100	0	20
7	110	100	10	30
8	100	100	0	30
9	90	100	-10	20
10	85	100	-15	5

contains equally weighted information from the five most recent samples. The latest plotted point in an EWMA control chart contains weighted information from all the samples, but the weights decrease exponentially with the age of the sample. The cumulative sum (CUSUM) chart is such that in the latest plotted point information from all the samples is included with equal weight. The concept of the CUSUM chart is radically different from all charts encountered so far in this chapter. It was proposed by Ewan Page (1954) in the UK thirty years after Walter Shewhart proposed the charts that bear his name in the USA. A simple set of data will be used to introduce the basic principle of the chart.

Consider a batch process with target for yield of 100 units. The yield values for runs 1 to 10 are tabulated in Table 5.11. The first step in computing the CUSUM values is the calculation of the deviation of yield from target for each run. It is necessary to define a CUSUM value of zero corresponding to run no. 0. The CUSUM value corresponding to, say, run no. 3 is denoted by S_3 and is the sum of the deviation for that run and the deviations for all previous runs, i.e. in this case $S_3 = 15 + (-5) + 10 = 20$. Calculation can be speeded up on observing that once data for a run is available then:

$$\text{New CUSUM} = \text{Previous CUSUM} + \text{New deviation.}$$

The basic CUSUM chart is a plot of CUSUM (S_i) versus run number (i) as displayed in Figure 5.42.

Table 5.12 Mean yields for some sets of consecutive runs.

Set	Runs	Mean yield	Segment joining	Slope of segment
1	4 to 9 inclusive	100	P3 to P9	0
2	1 to 5 inclusive	104	P0 to P5	4
3	5 to 9 inclusive	99	P4 to P9	-1

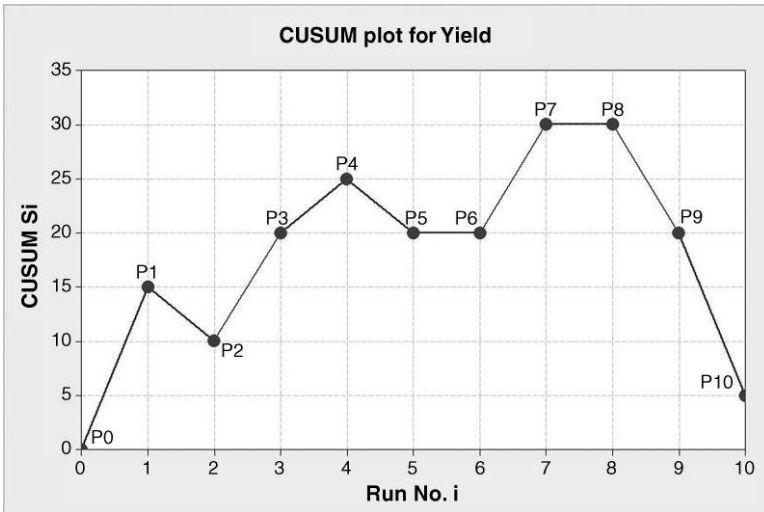


Figure 5.42 CUSUM plot for yield.

The reader is invited to confirm the mean yield values given in the third column of Table 5.12. The set of runs 4 to 9 inclusive has mean 100. The slope of the line segment joining the points P3 and P9 in the plot may be calculated as

$$\frac{S_9 - S_3}{9 - 3} = \frac{20 - 20}{6} = 0.$$

The reader is invited to check the other two slopes given in Table 5.12. Observe that the slope indicates by how much the mean yield for a consecutive set of runs differs from the target yield of 100. Thus the fundamental property of the CUSUM plot is that the slope indicates the process mean performance over the corresponding time period.

In order to illustrate this, a set of 90 yields was simulated using Minitab. The first 40 were from the $N(100, 10^2)$ distribution, the next 20 were from the $N(95, 10^2)$ distribution and the final 30 from the $N(103, 10^2)$ distribution. The target was again taken to be 100. The CUSUM plot is shown in Figure 5.43. The data are available in the worksheet Yields.MTW.

Reference lines have been added to indicate the three phases in terms of the distribution of yield. In the first phase the ‘horizontal’ appearance of the CUSUM plot corresponds to the process operating ‘on target’. In the second phase the downward trend in the plot corresponds to a process operating ‘below target’. In the final phase the upward trend corresponds to a process operating ‘above target’. Formal detection of signals of possible special cause variation may be carried out using a V-mask. To illustrate, let us suppose that we decide to set up a CUSUM chart with a V-mask when the first 25 yields are available from the simulation referred to above. The data are available in Yields25.MTW. In Minitab we would need to use **Stat > Control Charts > Time-Weighted Charts > CUSUM...** With **All observations for a chart are in one column:** selected, Yield is specified as the variable to be charted. **Subgroup sizes:** 1 and **Target:** 100 are entered. Using **CUSUM Options...**, under **Estimate**, 1 : 25 may be entered via **Use the following subgroups when estimating parameters**. Under **Plan/Type**, for **Type of CUSUM**, choose **Two-sided (V-mask)** and for **Center on subgroup:** enter 25. Accept the default **CUSUM plan** and all other defaults. The

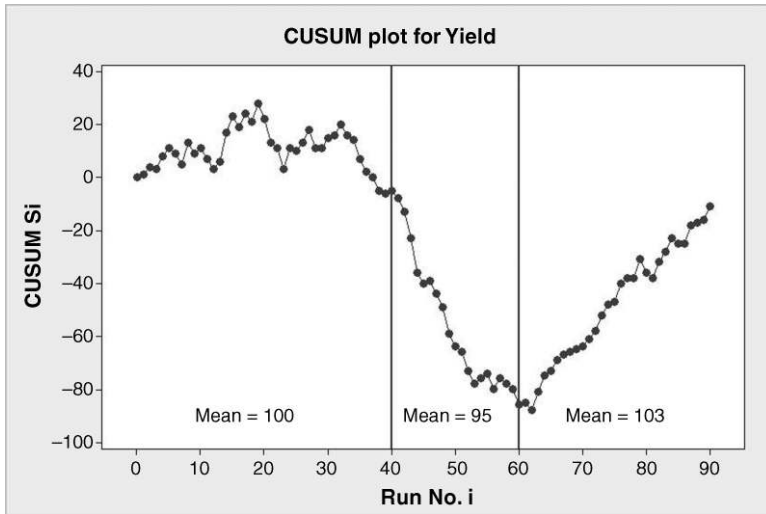


Figure 5.43 CUSUM for simulated Yield data.

resulting chart is displayed in Figure 5.44. (The author considers it potentially misleading that, having specified **Target:** as 100 during the creation of the chart, Target = 0 appears as a legend to the right of the plot, so he clicks on it and deletes it.)

The midpoint of the line segment, which forms the blunt end of the V-mask, is placed on the point specified using **Center on subgroup:**. If all the previously plotted points are ‘embraced’ by the arms of the mask then the process may be deemed to be in a state of statistical control, exhibiting no signal of any potential special cause variation. The mask can then be adopted for further monitoring of the process and can be thought of as being moved to each new point as it is plotted.

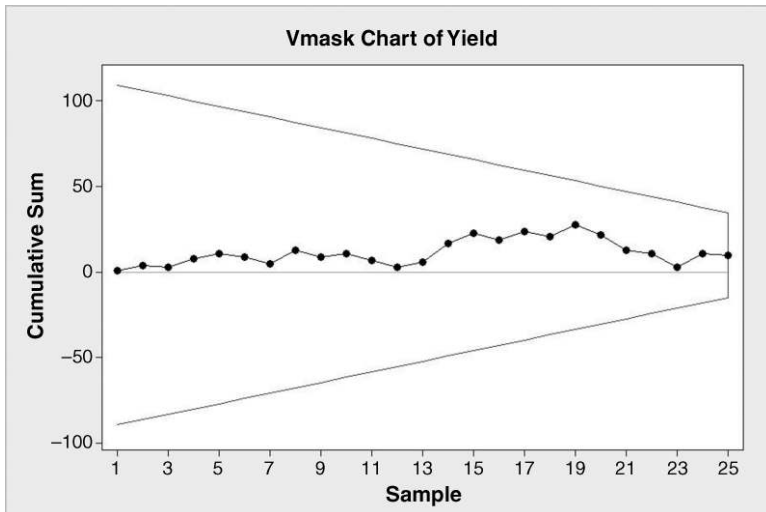


Figure 5.44 CUSUM chart with V-mask for first 25 simulated Yields.

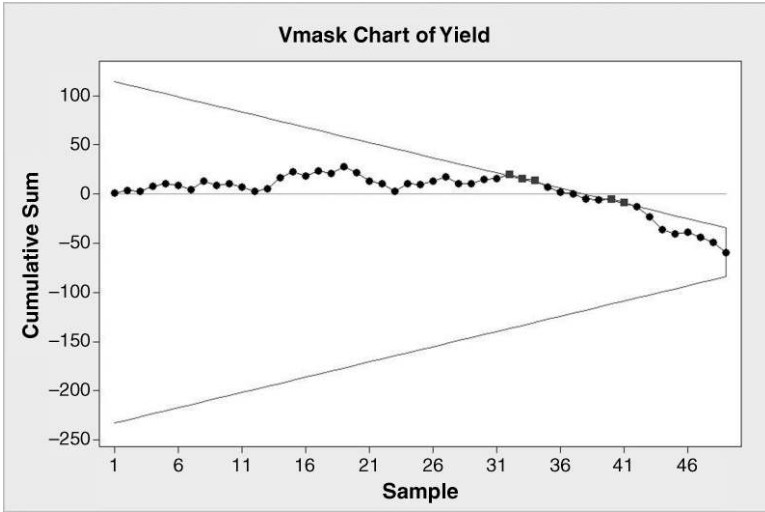


Figure 5.45 CUSUM chart with signal of potential special cause variation.

On plotting the point for the 49th run the chart appears as shown in Figure 5.45. The fact that at least one point is not embraced by the arms of the V-mask is the signal of potential special cause variation from this type of CUSUM chart. As with the EWMA chart, the CUSUM chart is very sensitive to small changes in the process mean.

If one selects **One-sided (LCL,UCL)** as the **Type of CUSUM**, with all other choices as before then the chart in Figure 5.46 is the result. This version of the CUSUM chart consists of two one-sided CUSUM plots. Here the point below the LCL for run 49 signals potential special

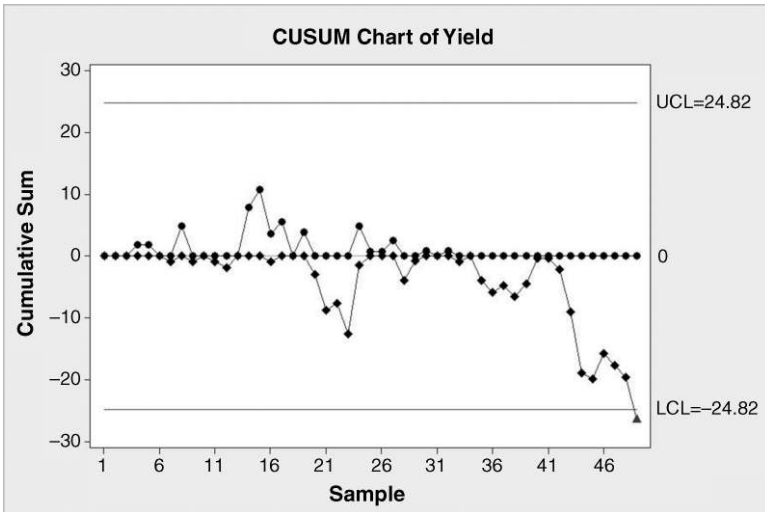


Figure 5.46 CUSUM chart with signal of potential special cause variation.

cause variation affecting the process. CUSUM charts may also be used with measurements recorded in subgroups. Montgomery (2009, pp. 400–419) provides comprehensive details. Caulcutt (1995, pp. 108–109) refers to the use of CUSUM charts for ‘post mortem’ analysis of process data. This approach may even be used with data for which there is no specific target value. By creating a CUSUM chart of the first type considered, with target set equal to the overall mean for the data series, one can often gain useful insights into process performance – marked changes in slope indicate the likelihood of changes in the process mean. A paper by Henderson *et al.* (2010) provides examples and data sets.

5.4 Process adjustment

5.4.1 Process tampering

The late Dr W. Edwards Deming often carried out funnel experiments during his presentations. These experiments were developed in order to illustrate the assertion that ‘if anyone adjusts a stable process to try to compensate for a result that is undesirable, or for a result that is extra good, the output will be worse than if he had left the process alone’ (Deming, 1986, pp. 327–331). In the experiments a target point is set up on a sheet of foam placed on a table and marbles are dropped, one by one, through a funnel onto the foam. Initially the funnel is aimed directly at the target. Following each drop the point of impact of the marble is recorded and one of a series of four rules is applied to determine the next point of aim.

1. Leave aim unchanged.
2. Adjust aim from previous aim position to ‘compensate’ for the deviation from target of the last bead dropped.
3. Adjust aim to opposite side of target from point of rest of last bead.
4. Adjust aim to point of rest of last bead.

Results from the experiments may be simulated, and output from simulations of the four scenarios, generated using a Minitab macro written by Terry Zeimer in 1991 (<http://www.minitab.com/en-GB/support/macros/default.aspx?q=deming+funnel&collection=LTD>), is displayed in Figure 5.47. Further information on the macro will be provided in Chapter 11. (With one of the four plots active, **Editor > Layout Tool...** was used to display all four plots in a single Minitab graph.) The points represent 100 impact points under each scenario.

The scaling of both axes is the same in each plot and the target is at the central point of the grid in both. The top right-hand plot (Rule 2) exhibits greater variation about the target than does the top left-hand plot (Rule 1). Rule 1 is optimum, Rule 2 yields stability but increased variability, Rule 3 leads to instability with oscillation and Rule 4 to what is known as random walk behaviour. The fundamental point of the experiment is to demonstrate that tampering with a stable process leads to increased variability in performance.

In order to demonstrate the effect of Rule 2 type ‘tampering’ on a process, consider again the rod-cutting process operating ‘on target’ and producing rods with lengths which are normally distributed with mean 60.00 mm and standard deviation 0.02 mm. For the purposes of illustration consider an individuals chart, with three-sigma limits, for the length of a single rod selected at random from the process output at regular intervals. The chart limits are at 59.94

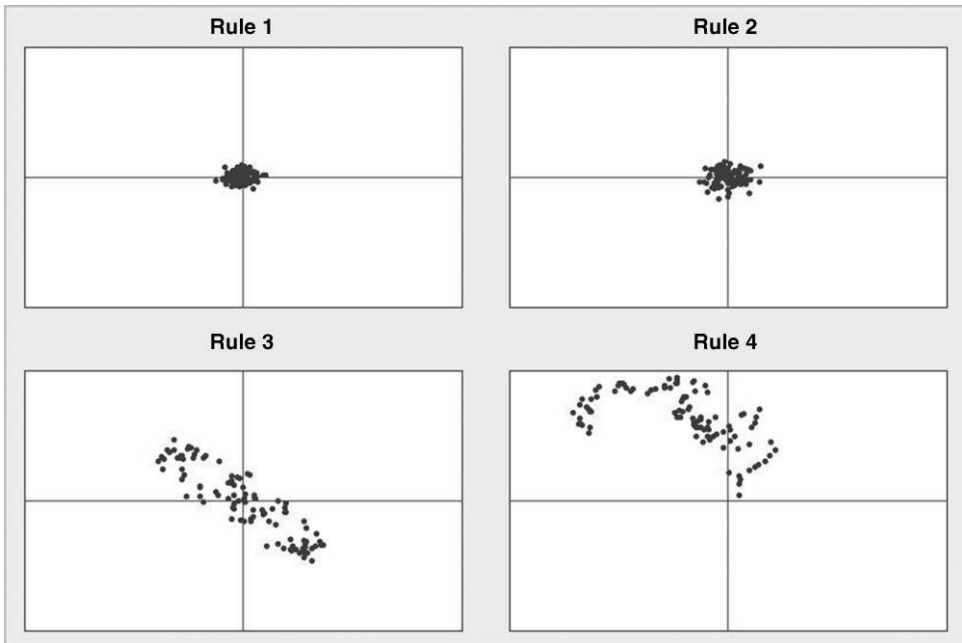


Figure 5.47 Simulations of the Deming funnel experiments.

and 60.06. If an operator of the process applied Rule 2 type tampering then he would respond to an observed length of 60.03 mm by *reducing* the process aim by 0.03 mm *from the previous aim* and to an observed length of 59.98 by *increasing* the process aim by 0.02 mm *from the previous aim*. Were an operator of the process to apply Rule 3 type tampering then he would respond to an observed length of 60.03 mm by *changing* the process aim to 0.03 mm *below target*, i.e. to 59.97, and to an observed length of 59.98 by *changing* the process aim to 0.02 mm *above target*, i.e. to 60.02. Under Rule 4 type tampering an operator would respond to an observed length of 60.03 mm by *changing* the process aim to 60.03 mm and to an observed length of 59.98 mm by *changing* the process aim to 59.98 mm. Individuals control charts for 100 simulated values under all four rules are shown in Figure 5.48. In each case the scaling on the vertical axis is the same to facilitate visual comparison of performance.

The increased variability under Rule 2 is again apparent, and some of the plotted points lie outside the historical chart limits in the case of Rule 2. The standard deviation of the set of 100 lengths obtained under Rule 1 is 0.0187, which is close to the specified standard deviation of 0.02. The standard deviation of the 100 lengths obtained under Rule 2 is 0.0263. Theoretically it can be shown that, under Rule 2, the variability, as measured by standard deviation, is increased by a factor of $\sqrt{2}$, i.e. by approximately 40%. Thus the tampering leads to increased variability, with a consequent reduction in the capability of the process to meet customer specifications.

Under Rule 1 the process displays what Shewhart referred to as controlled variation and what is widely referred to as common cause variation. A point outside the control limits on the Shewhart individuals chart would be taken as a signal of the possible occurrence of a special cause of variation. Such a signal would lead those involved in running the process to search for

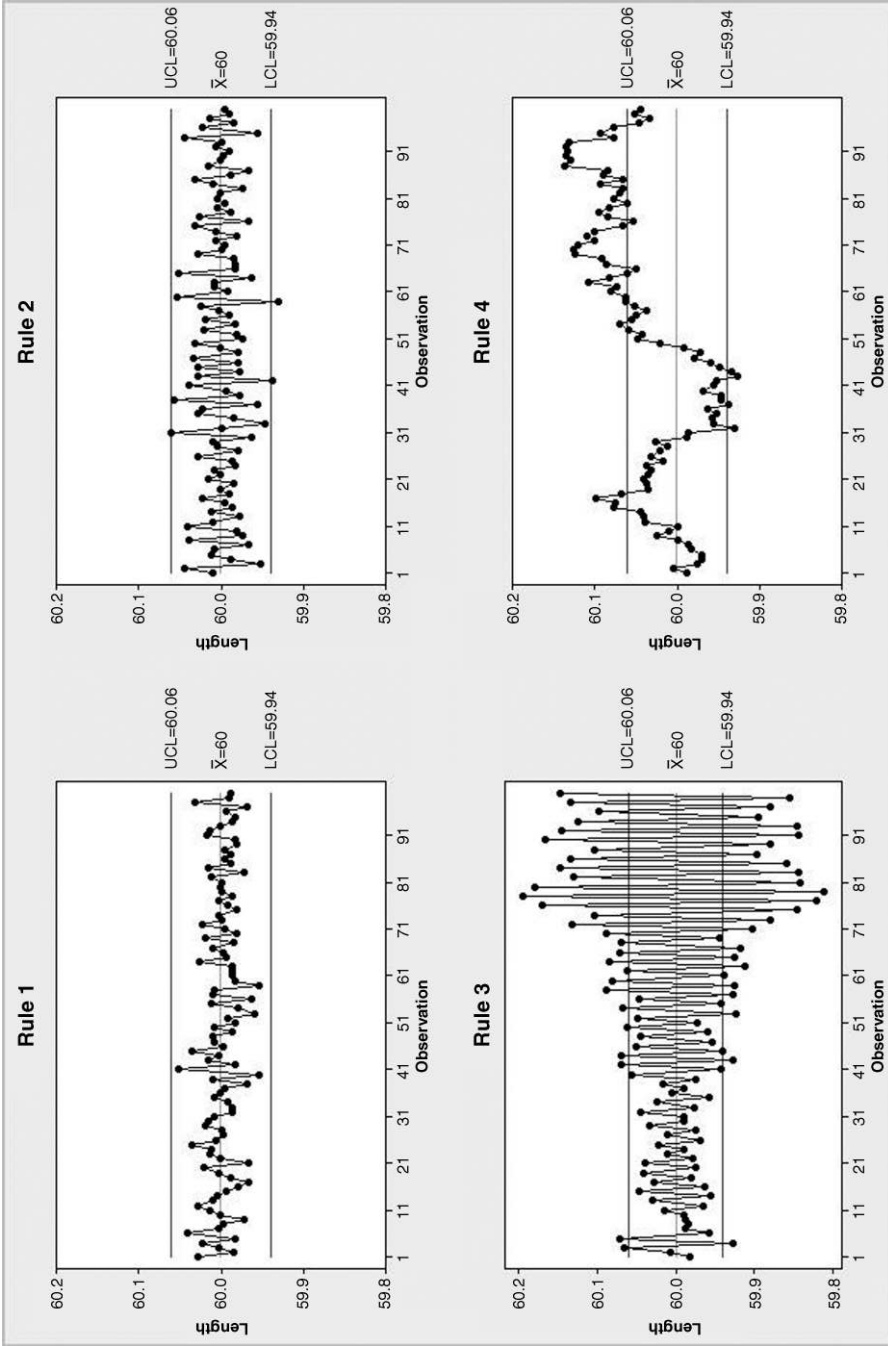


Figure 5.48 Individuals charts for the four funnel rules.

any special causes, e.g. a damaged cutting tool or a new and poorly trained operator. Having identified such special causes, effort would typically be made to eliminate them. Of course, some special causes may correspond to evidence of improved process performance, in which case it would be desirable to retain rather than eliminate.

5.4.2 Autocorrelated data and process feedback adjustment

George Box and Alberto Luceño applaud the Six Sigma strategy for quality improvement for the recognition that ‘even when best efforts are made using standard quality control methods, the process mean can be expected to drift’ (Box and Luceño, 2000, pp. 297–298). When successive process observations are not independent the data is said to be *autocorrelated*. The autocorrelation structure in the data enables a forecast of the next observation to be made from the available data. With the availability of a compensating factor whose effect on the process is known, appropriate adjustment to the level of the factor can be made in order to correct the predicted deviation from target. The procedure may be referred to as *feedback adjustment*.

Independence may be checked informally by examining a scatterplot of X_i versus X_{i-1} , i.e. of each observation plotted against the previous observation. The scatterplot in Figure 5.49 is for the rod cutting process operating without tampering under Rule 1. Clearly there is no lag 1 autocorrelation, i.e. no correlation between X_i and X_{i-1} .

The type of scatterplot in Figure 5.49 may be thought of as the ‘fingerprint’ of a typical process for which successive observations are independent. A formal analysis may be carried out by constructing a correlogram or autocorrelation function, consisting of a line graph of the autocorrelations at lag k plotted against k . The lag 1 correlation is the correlation between X_i and X_{i-1} , -0.134 in this case, and the lag 2 autocorrelation is the correlation between X_i and X_{i-2} , 0.168 in this case, etc. Correlograms generated using Minitab include significance limits indicating any autocorrelations which differ significantly from 0. The plot in Figure 5.50 was created using **Stat > Time Series > Autocorrelation**. None of the autocorrelation line

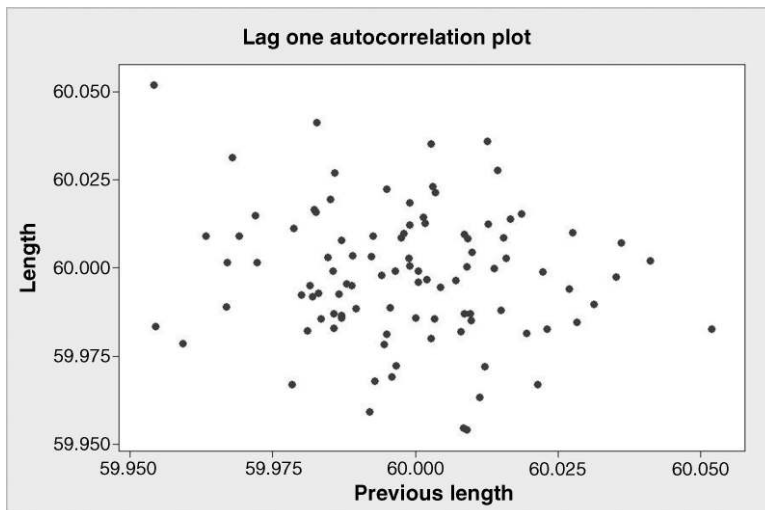


Figure 5.49 Lag 1 autocorrelation plot for data under Rule 1.

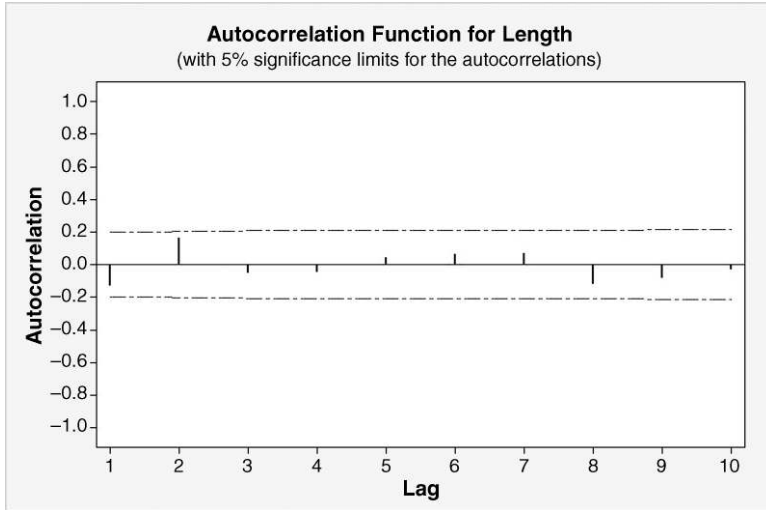


Figure 5.50 Correlogram for lengths obtained under Rule 1.

segments protrude beyond the limits, so there is no evidence of dependence in the time series of rod lengths in this case.

Consider a low-pressure chemical vapour deposition (LPCVD) process used in the fabrication of microelectronic circuits. A nitride layer is to be built up to a target thickness of 1000 Å on successive batches of silicon wafers. Let X_i represent a measurement of the thickness of the layer on a test wafer selected from the i th batch. Data from a simulated realization of such a process are plotted in Figure 5.51.

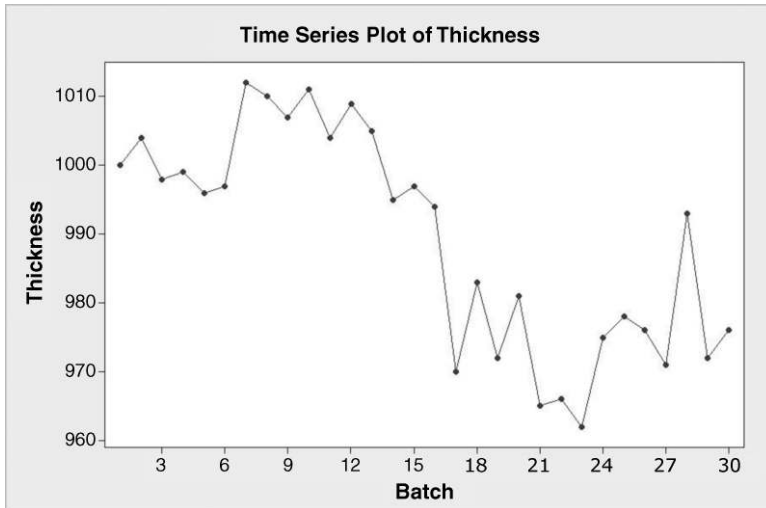


Figure 5.51 Time series plot of thickness data.

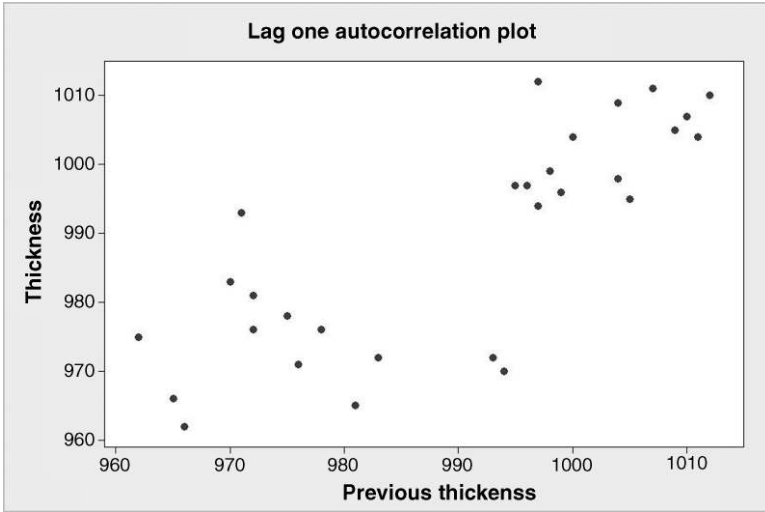


Figure 5.52 Lag 1 autocorrelation plot for thickness.

Here the thickness appears to shift and drift with time. A scatterplot of X_i versus X_{i-1} is shown in Figure 5.52. Unlike the scatterplot in Figure 5.49, the one in Figure 5.52 exhibits positive autocorrelation at lag 1. This scatterplot may be thought of as a typical fingerprint of a process for which successive observations are not independent.

The autocorrelation function (correlogram) is shown in Figure 5.53. Montgomery (2009, p. 446) comments that for such variables, even with moderately low levels of autocorrelation, conventional control charts will ‘give misleading results in the form of too many false alarms’.

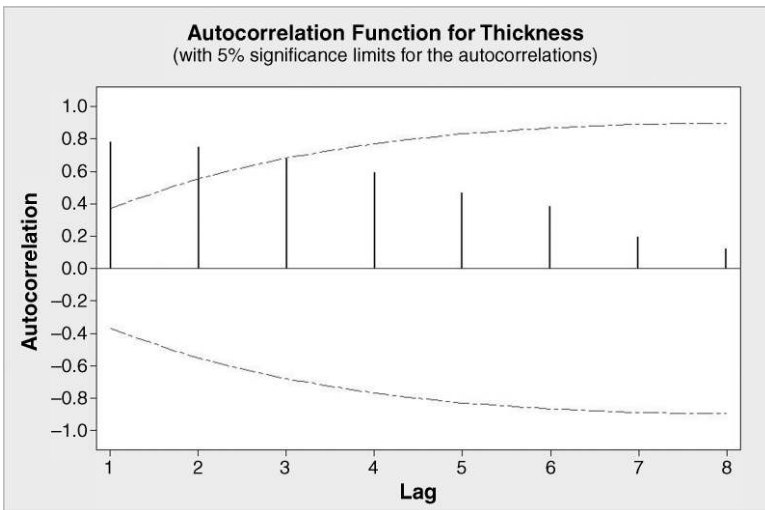


Figure 5.53 Correlogram for thickness.

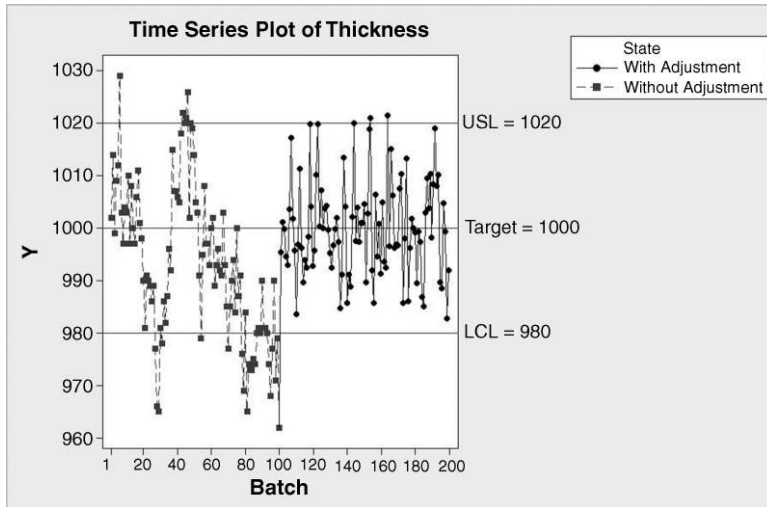


Figure 5.54 Run chart for batches both with and without adjustment.

Here the line segments representing the autocorrelation at both lag 1 and lag 2 protrude through the upper limit. Thus there is evidence that successive observations are not independent for this process. The autocorrelation structure in the data enables a forecast of the next observation to be made from the data. With the availability of both a forecasting procedure and a compensating factor whose effect on the process is known, appropriate adjustment to the level of the factor can be made in order to correct the predicted deviation from target. The exponentially weighted moving average is a forecasting tool used in industry in this context.

The assumption is made that the effect of any change in the level of the compensating factor will be complete by the time the next observation is made, i.e. that the process may be considered to be a *responsive system*. For the LPCVD process referred to earlier, processing time is a potential compensating factor with a gain of 30, i.e. for every extra minute the wafers remain in the LPCVD reactor another 30 Å can be expected to be added to the thickness of the nitride layer. If the most recent batch had spent 34 minutes in the reactor and the forecast thickness for the next batch was 940 Å, then, provided there were no random errors involved, a control action or adjustment of +2 minutes to the processing time would yield the required target thickness of 1000 Å. However, if the forecast was obtained from an EWMA with smoothing constant 0.4 then the actual adjustment made would be $0.4 \times (+2) = +0.8$ minutes. A simulated realization of the process without the adjustment procedure in operation is shown in Figure 5.54, together with a simulated realization of the process with the adjustment procedure in operation.

Comparison of the plot of thickness for the second 100 batches (adjustment in operation) with that for the first 100 batches (no adjustment in operation) reveals the benefit. The earlier ‘wandering mean’ behaviour has been replaced with much more stable behaviour and reduced process variability, which in turn leads to increased process capability. The procedure may be referred to as *feedback adjustment*, and further details, including discussion of choice of a suitable value for the smoothing constant and of applications, may be found in Henderson (2001), Montgomery (2009, p. 529) and Box and Luceño (1997).

5.5 Multivariate control charts

When monitoring the location of a single measured quality characteristic that remains stable and predictable, using a Shewhart chart with three-sigma limits and no other tests for evidence of special cause variation, the frequency of false alarm signals is 1 in 371 in the long term. In other words, one sample in 371 would give rise to a point plotting outside the chart limits although no special cause was affecting the process. If the locations of six independent measured quality characteristics which all remained stable and predictable were monitored using Shewhart charts the false alarm rate would be 1 in 62 in the long term, i.e. there would be six times as many false alarms to deal with. Typically a multivariate set of quality characteristics will not be independent.

For dependent bivariate random variables having a bivariate normal distribution, the equivalent of a point lying between the three-sigma limits in the univariate case is a point lying within a control ellipse in the scatterplot. Evidence of special cause variation could be overlooked through monitoring of the quality characteristics separately. Although the creation of control ellipses for bivariate data is quite feasible, the time sequence of the observations cannot be readily indicated. With three or more variables, representation of control ellipsoids and hyper-ellipsoids is impractical.

Hotelling's T^2 statistic may be plotted in a control chart in order to monitor a group of measured dependent quality characteristics. The theory underlying the chart assumes that the variables have a multivariate normal distribution. In order to construct the chart, estimates have to be made from the data of the means of the variables and of their covariance matrix. Data on short-circuit current (x) and fill factor (y) for photovoltaic cells, where a single cell was sampled at regular intervals from production, are in PV.MTW. The T^2 chart may be thought of as the multivariate equivalent of the Xbar chart, and Minitab also provides a generalized variance chart, which may be thought of as the multivariate equivalent of the R chart or S chart.

In order to create the charts, use **Stat > Control Charts > Multivariate Charts > Tsquared-Generalized Variance...** The two variables to be charted are selected and **Subgroup sizes: 1** specified. Estimation was based on the first 30 samples. The charts are shown in Figure 5.55.

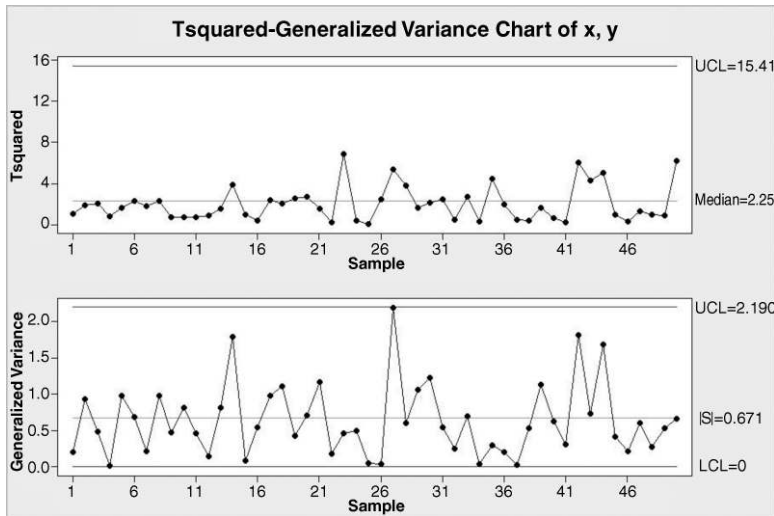


Figure 5.55 T^2 and generalized variance charts.

There are no signals of potential special cause variation affecting the process. Montgomery (2009, p. 499) gives a detailed account of multivariate control charts.

5.6 Exercises and follow-up activities

1. For the data you collected for Exercise 1 in Chapter 2, use Minitab to create an appropriate control chart.
2. Weekly checks of water quality are made at a chemical plant that manufactures products for use in the manufacture of microelectronic circuits. Values of the water quality index (WQI) are provided in Water.MTW. Create an individuals and moving range pair of control charts and verify that the WQI appears to be stable and predictable. Verify also that it is reasonable to consider WQI to be normally distributed by creating a normal probability plot of the data. Obtain estimates of the mean and standard deviation of the distribution from information on the control charts and compare with the estimates given on the normal probability plot.
3. In a continuous process for the manufacture of glass the soda level in the molten glass in the furnace is monitored daily. A series of 20 consecutive daily values are given in the worksheet Soda.MTW.
 - (i) Create an individuals chart of the data, specifying that estimation is to be carried out using subgroups 1 to 20 and with all the available tests implemented. Observe that the process is in a state of statistical control and that therefore the decision may be taken to monitor soda level using the individuals chart created.
 - (ii) Right-click on the chart and select **Update Graph Automatically**. Use **Window > Tile** to ensure that the control chart and the worksheet with the data may be viewed simultaneously.
 - (iii) Add the next five data values 12.96, 12.88, 12.89, 13.09 and 12.79 to the worksheet and observe the data points being added to the plot as they are entered into the worksheet. You should observe that Test 6 signals possible special cause variation affecting soda level on plotting the final value.
 - (iv) Recreate the chart using **S Limits** under **I Chart Options...** with 1 2 3 inserted. Observe from the revised chart how Test 6 has given rise to the signal.
4. Control charts may be used to ascertain whether or not there is evidence that a process is not operating on target through use of **Parameters** under **Options...** Suppose that the target level for soda in the previous exercise was 13.00. With the 20 daily soda levels in Soda.MTW create an individuals chart using **Parameters** to set the mean at 13.00. (Do not enter a value in the **Standard deviation:** window.) Note how the chart yields a signal from the eleventh sample that the soda level is off target. The use of a control chart in setting or checking process aim is described in detail by Wheeler and Chambers (1992, pp. 194–204).
5. The worksheet PilotOD.MTW gives data on samples of four output shafts taken at regular intervals from a production process. (Data reproduced by permission of the

Statistics and Actuarial Science Department, University of Waterloo, Canada, from Steiner *et al.*, 1997, p. 6). The values represent deviations from nominal (micrometers) for the diameter of the pilot.

- (i) Create Xbar and R charts of the data and verify that, when only Test 1 is used, the 15th sample signals a possible special cause affecting the process.
 - (ii) Verify that when all tests are used the 4th and 15th samples both provide signals.
 - (iii) Given that there was a problem identified with the process during the period when the 15th sample was taken, create the charts with the 15th sample omitted from the calculations of the limits and comment on process behaviour.
6. Refer again to the camshaft length data discussed in Section 5.1.3 and create Xbar and R charts for supplier 2 as in Figure 5.12, but before doing so create a column named Subgroup containing the numbers 1 to 20 each repeated 5 times using **Calc > Make Patterned Data > Simple Set of Numbers . . .** The following entries are required:

Store patterned data in: Subgroup

From first value: 1

To last value: 20

In steps of: 1

Number of times to list each value: 5

Number of times to list the sequence: 1

In addition, use the **Data Options . . .** facility to exclude the points corresponding to subgroups 2 and 14 from the plots. The dialog required is shown in Figure 5.56.

A text box was added to the charts indicating the nature of the special cause identified for subgroup 2. In Figure 5.57 the mouse pointer is shown positioned over the text tool icon in the process of creating text indicating the nature of the special cause – operator error – corresponding to the omitted subgroup 14. Once text has been entered double-clicking on it yields a menu that may be used, for example, to change font size.

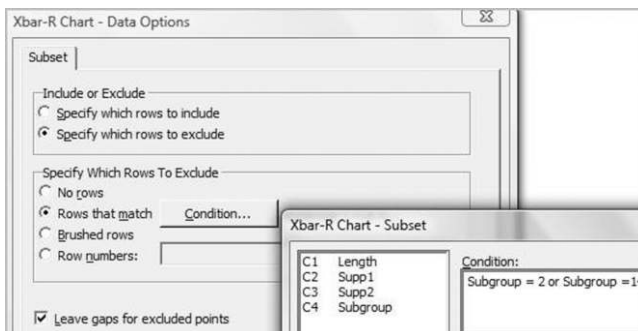


Figure 5.56 Specify subgroups for exclusion from plot.

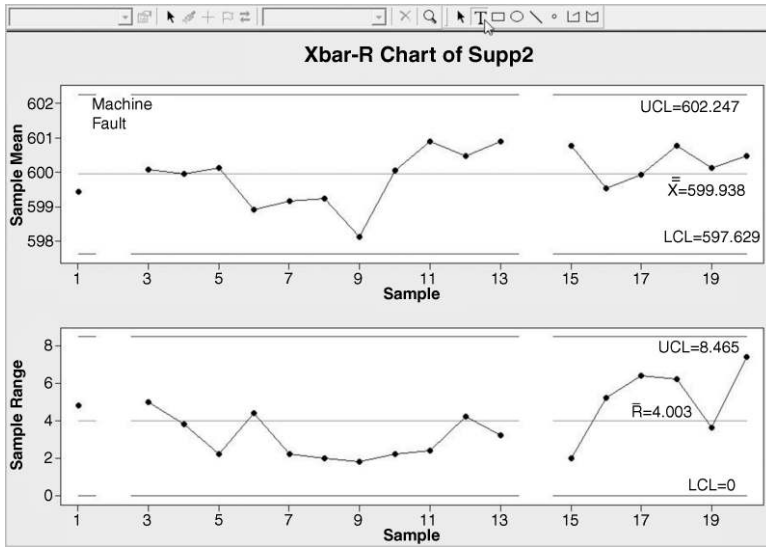


Figure 5.57 Xbar and R charts with gaps for excluded subgroups.

Although Minitab’s Assistant flow chart suggests that the subgroup size should be at least 9 (see Figure 5.1) for Xbar and S charts to be used, analyse the data for supplier 2 using them and demonstrate that the same conclusions would be reached as via the Xbar and R charts.

- The file Etch.MTW contains data on a dry etch process which etches silicon dioxide off silicon wafers during a batch microelectronic fabrication process (Lynch and Markle, 1997, pp. 81–83). The data are ©Society for Industrial and Applied Mathematics and ©American Statistical Association and are reproduced by permission of both organizations. During each batch run 18 wafers were processed and etch rate (angstroms per minute) was measured at nine positions on each wafer in a selection of six wafers from the batch. We will assume that the set of 54 measurements from each run constitutes a rational subgroup for the creation of Xbar and S (mean and standard deviation) charts of the 27 subgroups. In order to set up the data for charting use **Data > Stack > Rows...** as indicated in Figure 5.58. There were three phases involved. The first nine runs were carried out when the multi-wafer etch tool was only use intermittently, and the second nine runs were carried out when the tool was in regular use. Before the final nine runs were made the mass flow controller for CHF₃ was recalibrated. Note how it is necessary to expand the Phase column during the stacking operation into a new column named Stage. Note too that in specifying the name Etch Rate for the column in which the stacked data is to be stored it is necessary to enclose the name in single quotes.

Each consecutive group of 54 values in the Etch Rate column constitutes a subgroup for charting purposes. Create the Xbar and S charts by Stage and comment on the changes in process performance. Montgomery (2009, p. 251) advises that Xbar and S charts should be used in preference to Xbar and R charts when either the subgroup size is greater than 10 or the sample size is variable. Compare Xbar and R charts of the data with Xbar and S charts.

Phase	RUN	Wafer	Site 1	Site 2	Site 3	Site 4	Site 5	Site 6	Site 7	Site 8	Site 9
1	1	1	581.14	568.41	571.89	577.14	581.63	545.45	543.04	575.72	540.14
1	1	2	560.81	550.83	554.86	553.39	555.16	537.36	533.59	537.71	526.40
1	1	3	577.13	568.28	569.96	588.71	577.45	556.18	553.25	572.21	541.76
1	1	1							567.79	583.45	540.25
1	1	1							563.81	577.58	552.99
1	1	1							534.69	573.29	532.84
1	1	2							567.05	562.78	548.15
1	1	2							567.74	567.09	561.59
1	1	2							548.91	550.90	547.61
1	1	2							546.28	552.99	549.34
1	1	2							568.09	566.96	547.36
1	1	2							555.05	557.69	550.54
1	1	3							534.11	544.95	534.46
1	1	3							523.70	540.61	530.49
1	1	3							529.33	544.65	530.11
1	1	3							522.44	543.35	522.38
1	1	3							512.58	532.25	517.39
1	1	3							516.04	541.09	524.59
1	1	4							516.09	548.30	521.63
1	1	4							544.72	564.22	552.53

Figure 5.58 Stacking rows.

8. Bisgaard and Kulachi (2000) refer to a problem with off-centre bottle labels that had ‘bothered management for some time’. The excessive variation in the position of the labels detracted from the appearance of an expensive product and there was concern that this was affecting the company’s share of the decorating market.

The line foreman believed the labels were off centre because there was a lot of variation in bottle diameter. He said that the Quality Control Department had attempted a capability study but had ‘got nowhere’. The Maintenance Department claimed that the specifications were too tight and that the labels varied as well as the bottles. In an attempt to gain some insight into the problem, the deviations of label heights from target for 60 consecutive bottles were measured. The data for this example, available in the file Labels.xls, are from ‘Finding assignable causes’ by Bisgaard and Kulachi and are reproduced with permission from *Quality Engineering* (© 2000 American Society for Quality).

- Treat the data as 12 consecutive subgroups of size 5 and create Xbar and R charts and comment.
- Treat the data as 60 consecutive individual measurements and create an individuals chart. Note the repetitive pattern.
- The schematic diagram in Figure 5.59 indicates how labels were applied to bottles by a rotating drum with six label applicators spaced around its surface. Given that the first bottle in the data set had its label applied by the first applicator, set up a column named Applicator containing the sequence of numbers 1 to 6 repeated 10 times using **Calc > Make Patterned Data > Simple Set of Numbers...** The following entries are required:

Store patterned data in: Applicator
From first value: 1
To last value: 6
In steps of: 1
Number of times to list each value: 1
Number of times to list the sequence: 10

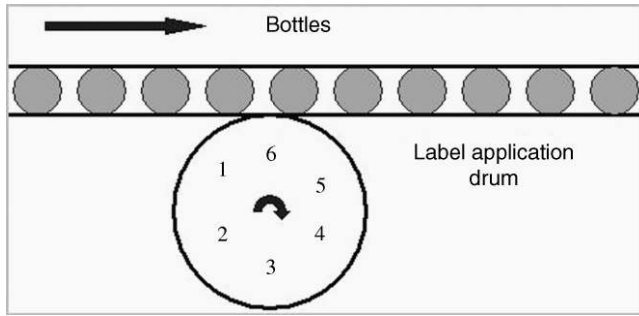


Figure 5.59 Label applicator.

- (iv) Unstack the data to obtain columns of Deviations for each applicator and verify, by creating individuals charts for each, that the individual applicators appear to be performing in a stable, predictable manner.
- (v) Create a boxplot of Deviation by Applicator and comment.

According to the authors, the moral of this story is that for process investigations, data should be plotted not only in time order but also in any other way that makes sense and preferably as individuals and not just as averages.

9. Montgomery (2009, pp. 292–298) gives an example on the manufacture of cans for frozen orange juice. The cans were spun from cardboard stock and a metal base panel attached. Every 30 minutes during production a sample of 50 cartons was inspected and the number of nonconforming cans recorded. Data for the first 30 samples are given in Cans1.xls and are reproduced by permission of John Wiley & Sons Inc., New York.

- (i) Create a P chart of the data and verify that samples 15 and 23 signal the occurrence of possible special cause variation.

Process records indicated that there was a problem with a new batch of raw material at the time that sample 15 was taken and that an inexperienced operator had been involved in running the process at the time sample 23 was taken. As assignable causes could be found for these two ‘out of control’ points it was decided to recalculate the limits with those samples excluded.

- (ii) Create the revised P chart. The author suggests that under **P Chart Options...** > **Estimate** you specify the sample to be used by employing **Use the following subgroups ...** and entering 1 : 14 16 : 22 24 : 30.

You should find that sample 21 now signals possible special cause variation. No assignable cause could be determined so it was decided to use the current chart with centre line at 0.215 and lower and upper limits of 0.041 and 0.389 for further process monitoring. At the same time, as the proportion of non-onforming cans was running at over 20%, it was decided to have adjustments made to the machine which produced the cans. A further 24 samples were taken and the extended data set is provided in Cans2.xls.

- (iii) Update the P chart to show the additional data and note how it indicates that the adjustments were beneficial. (Note that under **P Chart Options... > Estimate** you will need to specify the sample to be used by employing **Use the following subgroups ...** and entering 1 : 14 16 : 22 24 : 30).
- (iv) Create a Phase column with value 1 in the first 30 rows and value 2 in the next 24 rows and use it to create P charts for the two phases. The author suggests that under **P Chart Options... > Estimate** you specify the sample to be used by employing **Use the following subgroups ...** and entering 1 : 14 16 : 22 24 : 30 31 : 54. You should find that the chart for Phase 2 has centre line at 0.111 and lower and upper chart limits of 0 and 0.244, respectively.

Further data from the second phase of operation of the process are provided in Cans3.xls. Chart all the available data and comment on process performance.

Repeat the exercise using individuals charts of the actual proportions of nonconforming cans in the samples.

10. A department within a major organization prepares a large number of documents each week, with the numbers being similar from week to week. Table 5.13 gives the number of errors detected each week during final checks for a series of 15 weeks.
- (i) Create a C chart of the data.
- (ii) Given that a senior member of staff responsible for document preparation was on sick leave during week 4, explain why the chart with revised upper limit of 14.83, obtained on omitting the data for Week 4, could be 'rolled out' for routine monitoring.
Additional data are given in Table 5.14.
- (iii) Plot the additional data, with the revised limits used in (ii), explain how the chart provides evidence of process improvement and state what action you would recommend.
11. The worksheet PCB1.MTW gives counts of nonconformities on samples of 10 printed circuit boards taken from successive batches of a particular type of board built in work cell A at a factory.
- (i) Create a U chart of the data and use the formulae in Appendix 3 to check the centre line and chart limits.

Table 5.13 Data for weeks 1–15.

Week	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
No. of errors	7	8	3	22	1	3	10	3	13	9	13	10	4	2	7

Table 5.14 Data for weeks 16–30.

Week	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
No. of errors	3	4	6	7	6	4	3	6	3	4	4	2	5	9	2

- (ii) Since the sample size is constant here a C chart may be used. Create a C chart of the data and note that it is a scaled version of the U chart.

The advantage of the U chart, in terms of assessing process performance, is that it displays nonconformities *per unit*.

In work cell B a different type of board is manufactured and the sample size used for the monitoring of nonconformities varies. The worksheet PCB2.MTW gives counts of nonconformities on a series of samples of boards.

- (iii) Create a U chart of these data and comment on process performance.

12. In Section 5.2.1 an example on the use of a P chart to monitor the proportion of stroke patients receiving acute stroke unit care was given. It was also noted that the assumption of a binomial distribution is unlikely to be valid. An alternative approach to the use of a P chart in this case is to compute the proportion of patients receiving acute stroke unit care for each month and to create an individuals chart of these proportions.

- (i) Retrieve the data from the worksheet ASU.MTW, calculate the proportions and create an individuals chart of the proportions with limits based on the first 15 samples and all available tests implemented.

There are various points to note. First, the centre line on the individuals chart is at 0.646 4 as opposed to 0.635 2 in Figure 5.25. This is because the P chart procedure calculates the centre line as the total number receiving acute stroke unit care in the first 15 months (477) divided by the total number of stroke patients in the first 15 months (751). Second, the UCL is 1.024 7, an impossible value for a proportion! Third, unlike the P chart in Figure 5.25 there are no signals indicating an improved proportion of patients receiving acute stroke unit care. However, note that the last nine points are very close to being on the upper side of the centre line.

- (ii) Re-create the chart using **S Limits** under **I Chart Options**. . . to **Place bounds on control limits**, check the two boxes and enter 0 and 1 respectively since the variable to be charted is a proportion.

13. In the manufacture of aerosol cans height is a critical dimension and is measured at three locations equally spaced round the can. During a production run a can was selected every 10 minutes and three height measurements obtained for a sequence of 40 cans. The data are available in Aerosols.MTW.

- (i) Treat each row of the three columns of heights as a subgroup/sample of three heights and create Xbar and R charts.

Note that there are many signals on the Xbar chart. However, this is an incorrect approach. The problem is that the underlying assumption of independence is violated. The three heights in each subgroup/sample are from the same can. Had the samples/subgroups comprised a single height measurement from each of three different cans then use of Xbar and R charts would have been valid.

The correct approach is to use:

- an individuals chart of the means of the sets of three height measurements;
- a moving range chart for these means;
- a range (or standard deviation) chart for the sets of three height measurements.

(ii) Use **Calc > Row Statistics...** to create a column of means for the sets of three measurements and display the means in individuals and moving range charts.

Scrutiny of these two charts and the earlier R chart reveals no signals of possible special cause variation. The use of the ranges of the sets of three heights in (i) gave an estimate of standard deviation that is too small because it only measured variation *within* cans. This gave rise to limits on the Xbar chart that were too close together, hence the signals noted earlier.

(iii) Use **Stat > Control Charts > Variables Charts for Subgroups > I-MR-R/S (Between/Within)...** with subgroups specified across the three height columns to create the triple chart display of the data and verify that the charts obtained are the three discussed earlier.

Wheeler and Chambers (1992, pp. 221–226) discuss these under the heading three-way control charts.

14. Set up the funnel plot data in Table 5.7 in Minitab and create the funnel plot in Figure 5.29.