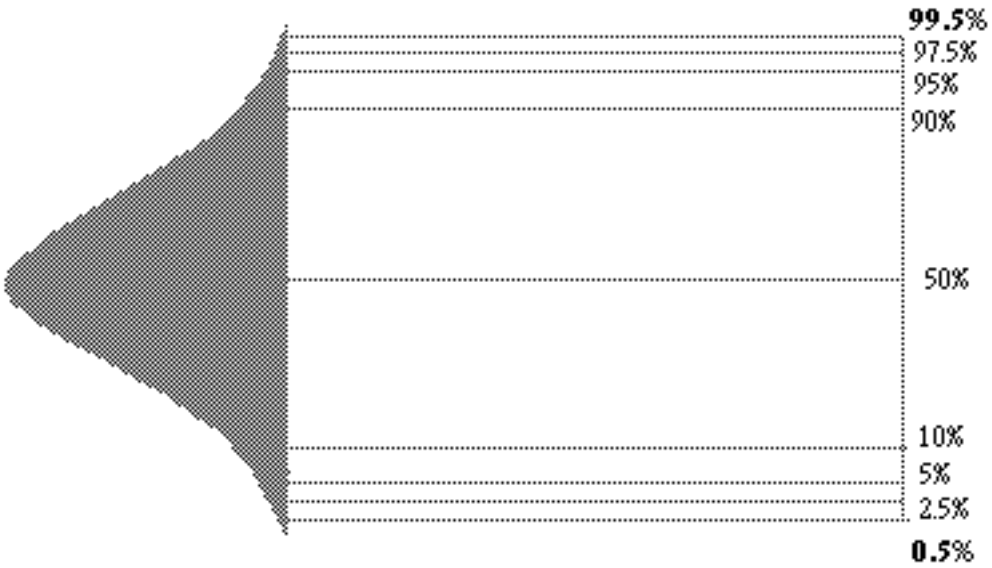


Monitoring Data using Statistical Process Control (SPC)

Introduction

Statistical inference is primarily concerned with drawing samples from some population or process of interest, and concluding from the sample data what might be expected if we sampled all of the population. For example, market research firms sample about 1,000 electors before each election, and evaluate from their replies the political preferences for all the electorate in the country. What we sample from is generally referred to as the **population**; i.e., all people over 18 and eligible to vote is the population of interest in this example. Sampling, if done correctly, can provide very good information about summary characteristics of the population under study (sometimes referred to as the **population parameter**) allowing statistics to be estimated in a cost efficient manner without having to go to the effort and cost of sampling every single item. Statistical Process Control (SPC) uses the same principles as those of sampling. Sample measurements are take from a process and measured for some characteristic of interest, the data are plotted and a decision is taken whether the parameter has changed or not.

SPC has its origins in manufacturing industry where it was first introduced by **Walter Shewhart** in the 1920s when working for the Bell Telephone company. He outlined the procedure in his classic text *The Economic Control of Manufactured Product* published in 1926 and the most popular SPC charts - **Shewhart charts** - are named after him. In recent decades SPC has spread from manufacturing industry to many non-manufacturing sectors as we can adapt the production line analogy to a wide range of industries and organisations concerned with monitoring data over time. Healthcare related applications are one such sector where the use of SPC has become widespread in the last decade where processes like waiting times for a procedure, user satisfaction levels, emergency service response times or number of user falls can be subject to SPC methods. See for example Chapter 13 (The use of control charts in healthcare) in the text Statistical Methods in Healthcare by Flatten et. al. (2011) and the paper by Woodall (2004) for more information.

In this section of the Data Visualisation module the procedure for generating one of the many types of SPC charts will be illustrated. The theoretical principles underpinning SPC will be illustrated using simulation software while users will be able to create charts and input sample data using customised software and Tableau.

1. The Concept of Variation in SPC

To summarise a data set we can calculate measures of average like the mean and the median and also measures of variability such as the range and the standard deviation. The mean or median indicates the typical or average value while the standard deviation gives us an idea of how spread out from the mean or how variable the data set is. Nearly all data sets have variation or spread unless all the data have the same value, in which case there is hardly any need for statistical analysis!

Variation can be divided into two types known as **Common** and **Special**. Common variation also known as **random** or **noise** variation is generally small in magnitude and is usually explained by a large number of causes, most of which remain unknown. It is regarded as a natural feature of processes. Examples of common variation might be small fluctuations in the response times of an ambulance to an emergency or weekly differences in the number of post operative infections. Examples from a manufacturing perspective might be small fluctuations in the length of a memory card or small differences in the number of units that fail a functional test in batches of 100 newly manufactured computers.

Special variation - also known as **signal** or **assignable** variation on the other hand is generally large in size relative to common variation, and suggests **real** significant change in the process. It is normally due to some assignable cause that can be tracked down and identified - unlike common variation which is explained by many causes. Examples of special variation might be a large decrease in the percentage of school children being immunised due to a media scare story. From the manufacturing sector examples of special variation might be a reported increase in the number of failures of logic boards which can be traced to the introduction of a new component placement machine in the manufacturing process or say, increased Hard Drive failures at a functional test station which are explained by a change of supplier of this component.

It is important to note that it is possible to have common variation at unacceptable levels in a process. To reduce this requires a fundamental system level process review. For example, to reduce the waiting time for a procedure may require investment in new facilities, recruitment of additional staff or the purchase of additional hardware/software. **System level causes tend to require system level solutions**.

Identification and control of variation is crucial to adopting a continuous improvement programme. The purpose of SPC is to distinguish between common and special variation, and in so doing, assist in the gradual reduction of the variation of the process.

In the next section we will examine techniques which allow us to do just this based on charts known as **Control Charts**. But first we will examine common variation in more detail through the use of a sampling experiment.

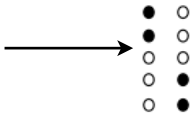
Sampling Experiment

In this section we will illustrate the difference between random and common variation. We will simulate an example based on an auditor sampling 10 users of a software application from a population of 100 users and counting the number of users who express dissatisfaction with the user interface.

The population of 100 users will be represented by a basket containing 100 beads. Fifty of the beads are black in colour representing the number of dissatisfied users. The remaining fifty are white representing satisfied users.

The sample of 10 users taken by the auditor will be represented by each member of the class selecting 10 beads from the basket and counting the number which are black. The 10 beads are then returned to the basket and the next class member repeats the process by selecting another 10 beads and so on. The results of 20 selections from the basket reported by 20 members of a different class is shown in the table overleaf.

Sample Number	Dark Beads Selected (i.e. dissatisfied users)	Sample Size
1	5	10
2	7	10
3	6	10
4	3	10
5	3	10
6	8	10
7	3	10
8	4	10
9	6	10
10	2	10
11	4	10
12	6	10
13	6	10
14	3	10
15	5	10
16	5	10
17	7	10
18	6	10
19	4	10
20	5	10



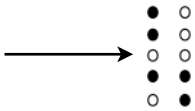


Table 1.1: Number of black beads reported in samples of 10

From the above table we can see that the number of black beads selected ranges from 2 to 8. This variability in the number of black beads sampled is called **random** or **sampling variation**.

Plotting the number of black beads on the chart below we can examine the pattern of random variation visually. The key point to note is that the increase, say, from 3 black beads (or 30%) in sample 5 to eight black beads in sample 6 (or 80%) is **not** explained by an increase in the overall level of dissatisfaction in the population which remained at 50 (or 50%).

Similarly, the decrease from 6 black beads in sample 9 to 2 black beads in sample 10 is **not** due to a decrease in overall user satisfaction. Again it is explained by random or sampling variation- that is the differing number of black beads in each sample is explained by the sampling process.

The sample results ranging from 2 out of 10 (or 20%) to 8 out of 10 (or 80%) is not explained by real changes in user satisfaction but by sampling or random variation.

An interesting observation is that if we add up the total number of black beads selected in Table 1.1 and divide by the total number of black beads obtained we get 98/200 which is 0.49. This result is very close to the true proportion of black beads in the basket (0.5). Therefore, accumulating the results of samples taken over and over again can provide a good approximation to the overall or true level of dissatisfaction. A rough rule of thumb is to select a minimum of 25 samples for a good estimate of the true rate. In reality the accuracy of the estimate will depend on the sample size and the true (but unknown) underlying failure rate.

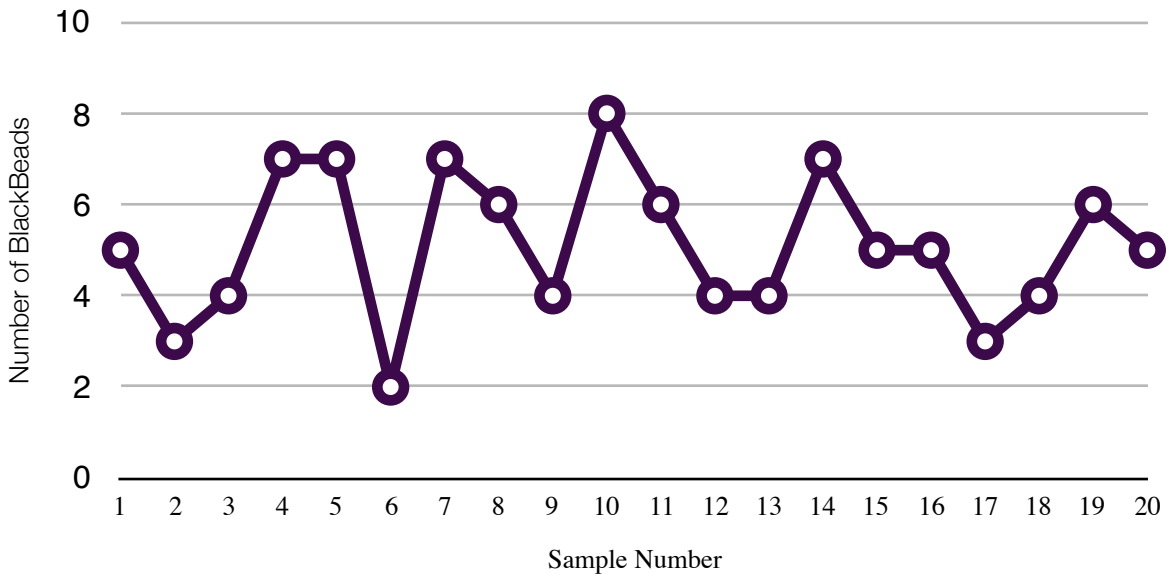


Figure 1.1: Run plot of the number of black beads obtained in samples of 10 beads

Random Variation and the Normal Curve

An important question that arises from the previous section is how can we determine if the true satisfaction rate has **really** changed i.e. if there are real increases or decreases in user satisfaction. For example, in the previous sampling experiment we saw that the large fluctuations in our sampling results - ranging from 20% of samples to 80% - is explained by random variation rather than any real change in the % of dissatisfied users. However, if users satisfaction really deteriorate and starts to generate rates in excess of 50% how can we detect this change - when we are sampling only a small fraction of users - and conclude that the increased number of negative responses is explained by real process change rather than random variation? We can answer this question by repeating the sampling experiment that we have just performed many many times and examining the pattern of the number of black beads obtained. If the number of black beads in our sample deviates from the pattern we can then safely attribute this result to real change in the number of black beads in the basket (or population). The plot below represents a simulation of the number of black beads found in 10,000 selections of 10 beads from the basket.

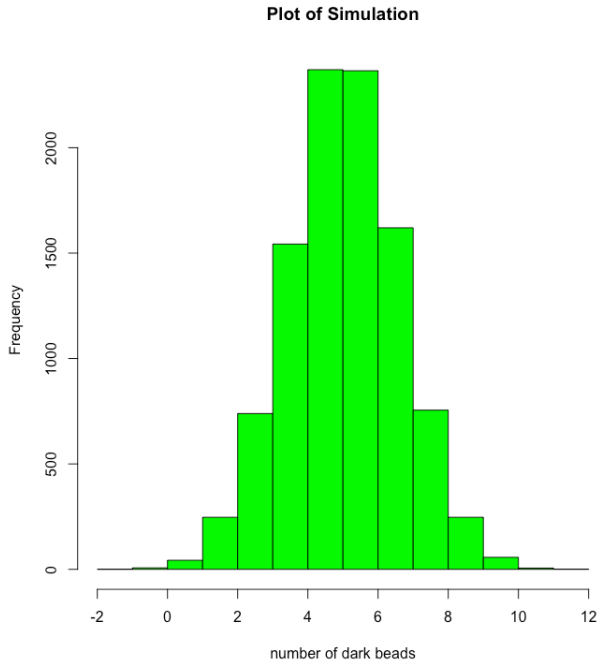


Figure 1.2: Number of black beads found in 10,000 simulated samples of size 10

The shape of this curve is what is known as a **Bell** or **Normal Curve**. The more data that are collected the smoother the curve becomes. This bell shape is the signature of random variation. We can see from the curve that it is very unusual to obtain more than nine (less than 0.1%) or less than one black bead (less than 0.1%) in a sample of 10 from a basket containing 50% black beads.

Therefore, if 10 black beads are obtained in a sample of 10 this suggests a **real** increase in the number of black beads in the basket rather than a result explained by random variation. We will be wrong in our decision less than 0.1% of the time as it is possible to obtain a sample which contains 10 black beads when the number of black beads has not changed - but it is a very small risk. In SPC we call this boundary of no more than nine beads an **Upper Control Limit (UCL)**. Similarly, it is very rare to obtain less than 1 black bead in a sample of 10 if the true number of black beads in the basket has remained stable at 50%. This should also arise in about 0.1% of samples. Therefore, if 0 black beads are obtained in a sample of 10 this suggests a real decrease in the number of black beads in the basket has occurred rather than the results of random variation. In SPC we call this boundary of no less than one black bead a **Lower Control Limit (LCL)**.

Repeating the sampling experiment but in this case **increasing** the number of black beads in the basket from 50% to 70% (i.e 50 to 70) we obtain the numbers shown in samples labelled 21 to 24 in Table 1.2. Sample 23 is recorded with 10 black beads out of 10 beads sampled. This is outside the limits of random variation as might be expected because the proportion of black beads ***did*** increase. Note also that the other 3 samples recorded high numbers - all above the mean of 5 beads - which is another potential indicator of special variation which will be discussed later.

Decreasing the number of black beads in the basket to 20% and taking four more samples we see from Table 1.2 that sample number 27 reported 0 black beads. This falls outside the lower limit of random variation. This signals a **real** decrease in the proportion as again the proportion of black beads ***did*** decrease.

Sample	Number Black Selected in samples of 10	True Number of Black Beads	
21	7	**70**	
22	8	**70**	**70 black beads in basket**
23	**10**	**70**	
24	8	**70**	
25	2	**20**	
26	4	**20**	**20 black beads in basket**
27	**0**	**20**	
28	1	**20**	

Table 1.2: Number of black beads selected from a samples of 10

The earlier sampling results together with the results in Table1.2 are illustrated in Figure 1.3. We can see the impact of real rather than random variation in the chart with sample 23 above the upper limit of variation and sample 27 below the lower limit where real changes did occur. It should be noted that when the proportion does increase/ decrease it may take several samples to be taken before a sample breaches the limits. However, other clues like the samples starting to fall consistently above/below the mean are also be an indicator of real process change.

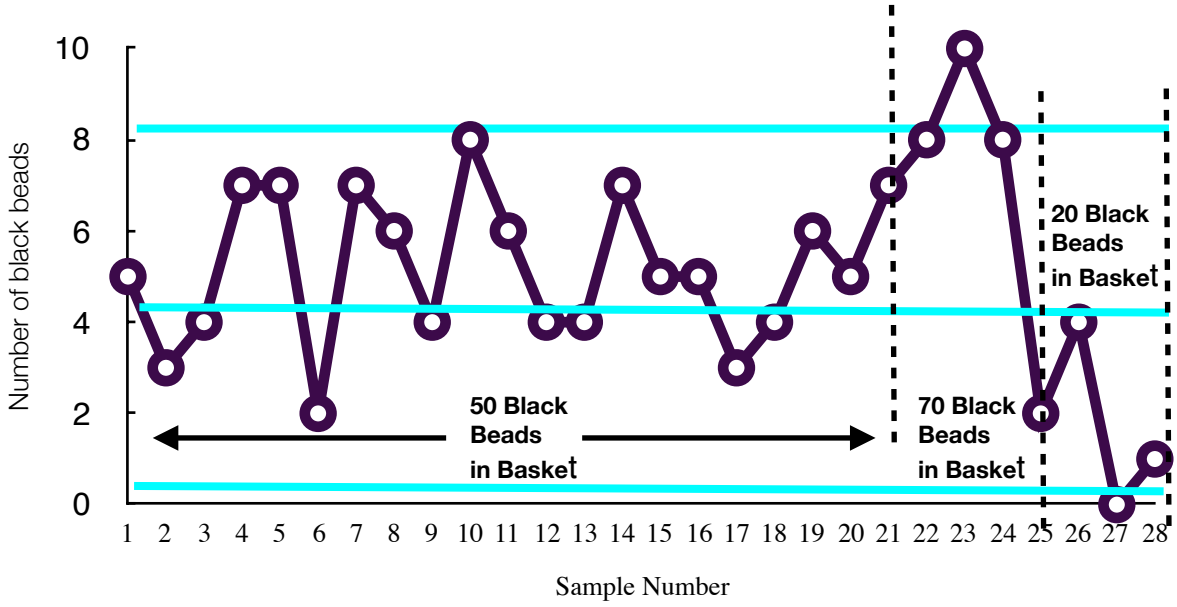


Figure 1.3: Run plot of the number of black beads in samples of 10 from a basket containing 50%,70% and 20% black beads

As about 99.9 per cent of samples from a basket with 50 per cent black beads should contain between 1 and 9 black beads we call the limits 99.97% control limits. These control limits can be approximated by calculating the mean ± 3 times the standard deviation of the data which we will examine in the next section. Please refer to **Computer Activity 1** at the end of this Chapter to explore the sampling experiment dynamically using the software *Sampling Experiment*.

Special Properties of the Normal Curve

The Normal curve is the basis behind most of the traditional SPC charts. It can be shown that approximately 68% of observations in a normally distributed data set fall within plus or minus one standard deviation of the mean of the data set. For example, in Figure 1.4 the mean number of black beads is 5 and the standard deviation is 1.6. So 68 per cent of the values in this data set should fall between 5 + 1.6 and 5 -1.6, i.e. between 3.4 and 6.6 black beads. Ninety-five per cent will fall within ± 2 standard deviations, and finally approximately 99.97% of the values will fall within the mean ± 3 standard deviations or between 5 + 3(1.6) and 5 - 3(1.6)1.6 or between 0.2 and 9.8 black beads as shown in the diagram below.

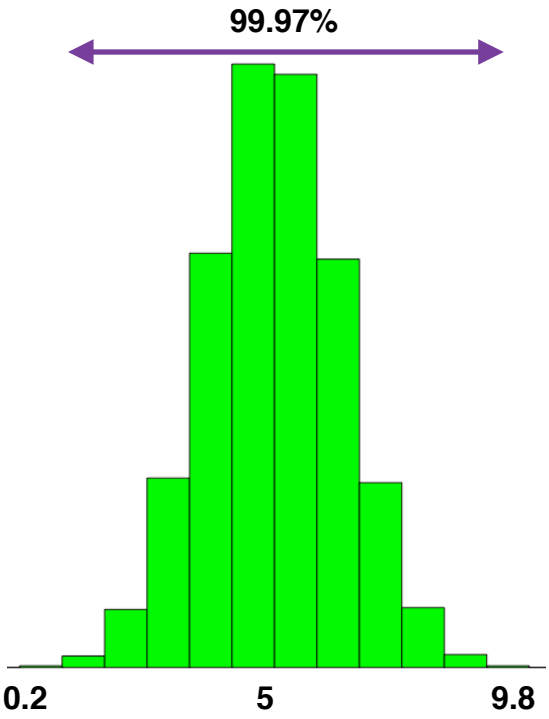


Figure 1.4: 99.97 per cent limits on the number of black beads

These important properties are the key to understanding SPC.

As 99.97% of the data values should be contained within the limits ± 3 standard deviations of the mean (which in this example would be 0.2 to 9.8 dark beads) this suggests that it is extremely unlikely that we should obtain a sample with 0 or 10 dark beads fall unless the mean has really increased or decreased.

So if we obtain 10 black beads or 0 black beads we can be almost certain that the process mean has increased or decreased, respectively explained by the development of a special cause of variation. This, in a nutshell, is the main purpose of SPC:

Finding out at the earliest possible time if the process has changed by determining if any points fall outside the critical value of ± 3 standard deviations of the mean. This critical value in SPC is generally referred to as a Control Limit.

Control Limits for SPC

SPC systems generally have two control limits, an upper control limit (UCL) and a lower control limit (LCL). Points falling outside these limits denote **special variation**.

When the upper and lower control limits are plotted we call the chart a **Control Chart**. Points falling beyond the upper control limit signal a process for which the mean has increased as shown in the plot below and in the discussion in the previous section. Points falling beyond the lower control limit may signal a process for which the mean has significantly decreased. While this is generally a positive development in the context of defect rates, it can also imply the development of more serious issues. For example, the process may appear to have improved, but this is only because it is not being inspected/ tested properly and so we are not detecting the true level of defects.

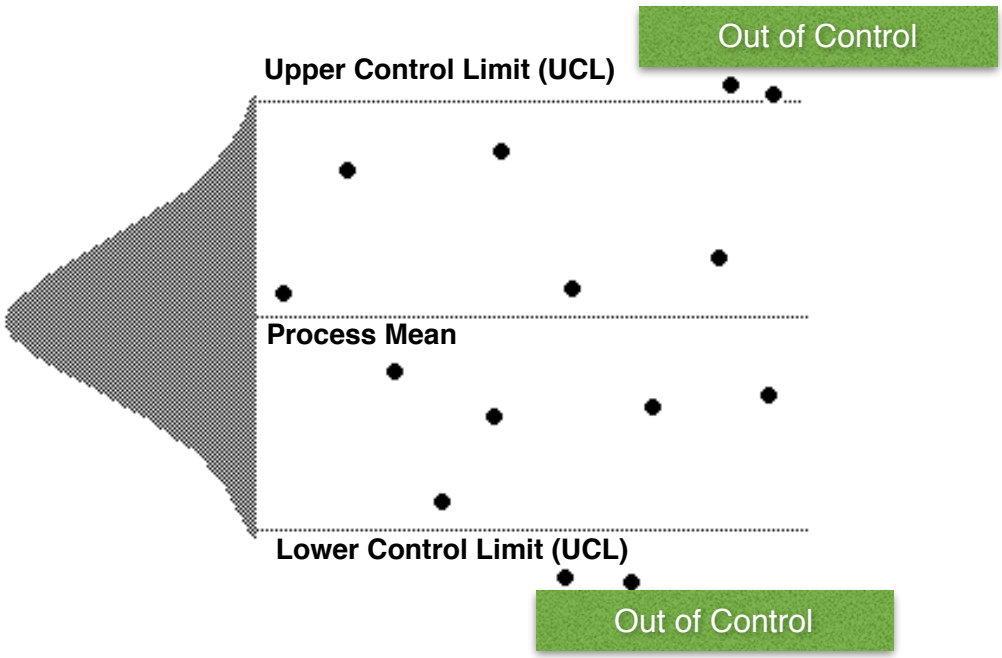


Figure 1.5: Illustration of a Statistical Process Control Chart

Supplemental Rule for Diagnosing Special Variation

As previously discussed, if a point falls outside the UCL and/or LCL we conclude that a special cause of variation is present in the process, which has led to an upward/ downward shift in the process mean.

However, a point does not have to fall outside the control limits for a special cause of variation to be detected. For example, seven consecutive points all falling on one side of the mean on the chart, signals the presence of special variation as shown below. Eight consecutive points either increasing or decreasing in magnitude is another rule for detecting the presence of special variation.

There are many other rules for detecting process drift before a point crosses the control limits which are discussed in most textbooks on SPC. However, while the use of such rules has the advantage of increasing the sensitivity of the chart to detecting process drift, they have the drawback of complicating somewhat the simplicity of operating and interpreting the control chart.

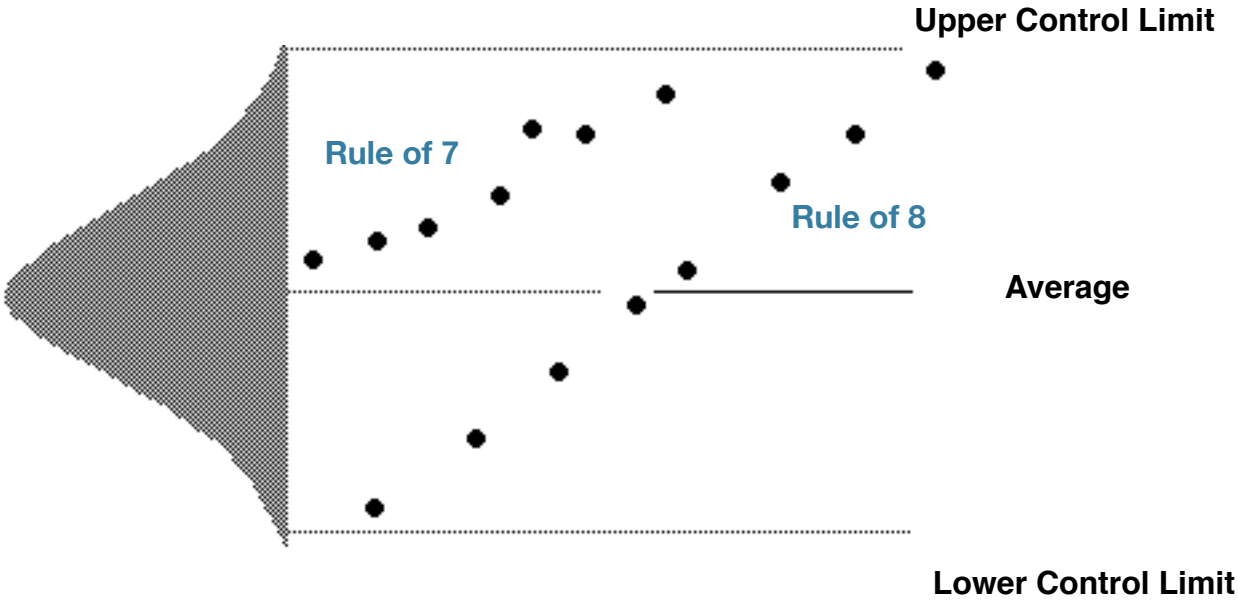


Figure 1.6: Supplemental rules for operating SPC Charts

1.2 Classification of SPC Charts

SPC charts can be divided into those that are more appropriate for **discrete data** and those that are suitable for **continuous data**. Discrete data has a limited number of distinct possible values or categories i.e. reject/accept, yes/no/maybe, sick/well, on/off etc. Continuous data has a large number of possible values i.e. the weights of people, heights of people, times taken to do a particular task in minutes, emergency response times etc. Control charts based on continuous data are called *variable* control charts, whereas charts using discrete data are called *attribute* charts. The most common variable control charts are Shewhart Mean and Range charts sometimes called X&R charts. The most common attribute charts are number defective (nP) charts, proportion defective or P charts and nonconformity charts or C charts. In this introduction we will examine SPC charts for discrete data only.

SPC Charts for Discrete Data

Monitoring discrete data e.g. data based on a small number of categories for example, a user is either satisfied with their level of care or not we can use an nP to monitor user quality. nP charts monitor the number of samples reporting a characteristic of interest (e.g. satisfaction) by taking samples of a **fixed size** over time and plotting the number with the characteristic of interest on the **y-axis** of the chart.

The procedure for calculating a nP Chart is outlined in the next section.

Procedure for Calculating nP Chart

i) Calculate, based on at least 25 samples an estimate of the average proportion of observations with the characteristic of interest, **P**. **P** is calculated as:

$$\mathbf{P} = \frac{\text{Total number of observations } \mathbf{with} \text{ the characteristic of interest}}{\text{Total number of observations taken}}$$

ii) Calculate **n**P which gives the average number of observations with the characteristic of interest per sample size. P is as above while **n** is the fixed number of units inspected per sample.

iii) Calculate the Upper Control Limit (UCL) and the Lower Control Limit (LCL) using the following formulae:

$$UCL = nP + 3\sqrt{nP(1-P)}$$
$$LCL = nP - 3\sqrt{nP(1-P)}$$

where **n** is the fixed number of units inspected per sample

This formula is equivalent to calculating the Mean ± 3 times the standard deviation as discussed earlier in our sampling experiment. The control limits are the boundaries of the number of defectives we should expect for random variation only. Points falling outside these boundaries represent special variation.

iv) Plot the number of observations with the characteristic of interest on the chart.

Worked Example

The following table shows the number of units rejected from samples of size 100 taken from 30 batches from a computer manufacturing process. For example, sample 1 reported 8 rejects from the batch of 100 inspected.

Sample Number	Inspected (n)	Rejected	Sample Number	Inspected (n)	Rejected
1	100	8	16	100	7
2	100	7	17	100	3
3	100	8	18	100	6
4	100	6	19	100	8
5	100	4	20	100	12
6	100	1	21	100	1
7	100	9	22	100	7
8	100	0	23	100	6
9	100	5	24	100	3
10	100	8	25	100	7
11	100	9	26	100	6
12	100	4	27	100	6
13	100	4	28	100	8
14	100	3	29	100	7
15	100	7	30	100	6

i) **P** is calculated as:

$$\frac{\text{Total number rejected}}{\text{Total Number inspected}} = \frac{176}{3{,}000} = 0.058666 = 0.059$$

ii) **nP** is calculated as: 100(0.059) = 5.9

iii) The control limits are calculated as:

$$UCL = nP + 3\sqrt{nP(1-P)} = 100(0.059) + 3\sqrt{100(0.059)(1-0.059)} = 5.9 + 7.1 = 13$$

$$LCL = nP - 3\sqrt{nP(1-P)} = 5.9 - 3\sqrt{100(0.059)(1-0.059)} = 5.9 - 7.1 = -1.2$$

As the LCL is less than 0 we set it equal to 0 as we can't have a negative number of defectives! This arises as the normal approximation, which the control limits are based on, is **not strictly valid** in this circumstance (technically the shape is described by the binomial distribution). The number of defective units is plotted on the nP chart as shown below.

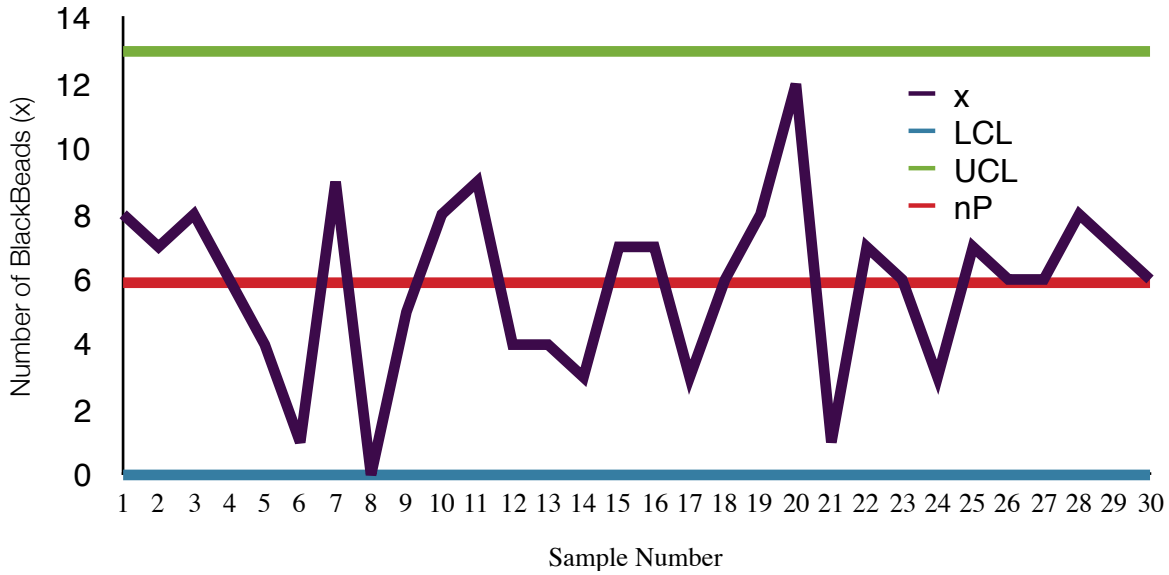


Figure 1.7: SPC nP Chart created from data in worked example

The nP chart shown above suggests a stable process with all the data points in control. The average number of defectives per 100 inspected (nP), at 5.9 appears to be a realistic estimate of the quality of the manufacturing process.

Recalculation of Control Limits

Control limits should only be recalculated and when it is clear that the process has stabilised (usually if about 25 points have seen to be free of special variation). If the process has shifted upwards then the limits should not be loosened as this defeats the purpose of SPC, instead, we should find the assignable causes of variation that led to this decrease in quality, and apply corrective action to bring the process under control again. Points falling out of control arising from the presence of special variation should not be used in the re-computation of control limits, as these points will ultimately lead to widening of the control limits.

1.3 Software for Creating and Visualising SPC Charts

The DataDesk programmes accompanying this Chapter are designed to allow users a greater understanding of some of the important features of nP SPC charts through the use of visualisation and simulation. In this section two applications will be introduced *nP Chart - Simulation* and *nP Chart - Data Entry*.

i) nP Chart - simulation

On opening this application the interface in Figure 1.8 is observed:

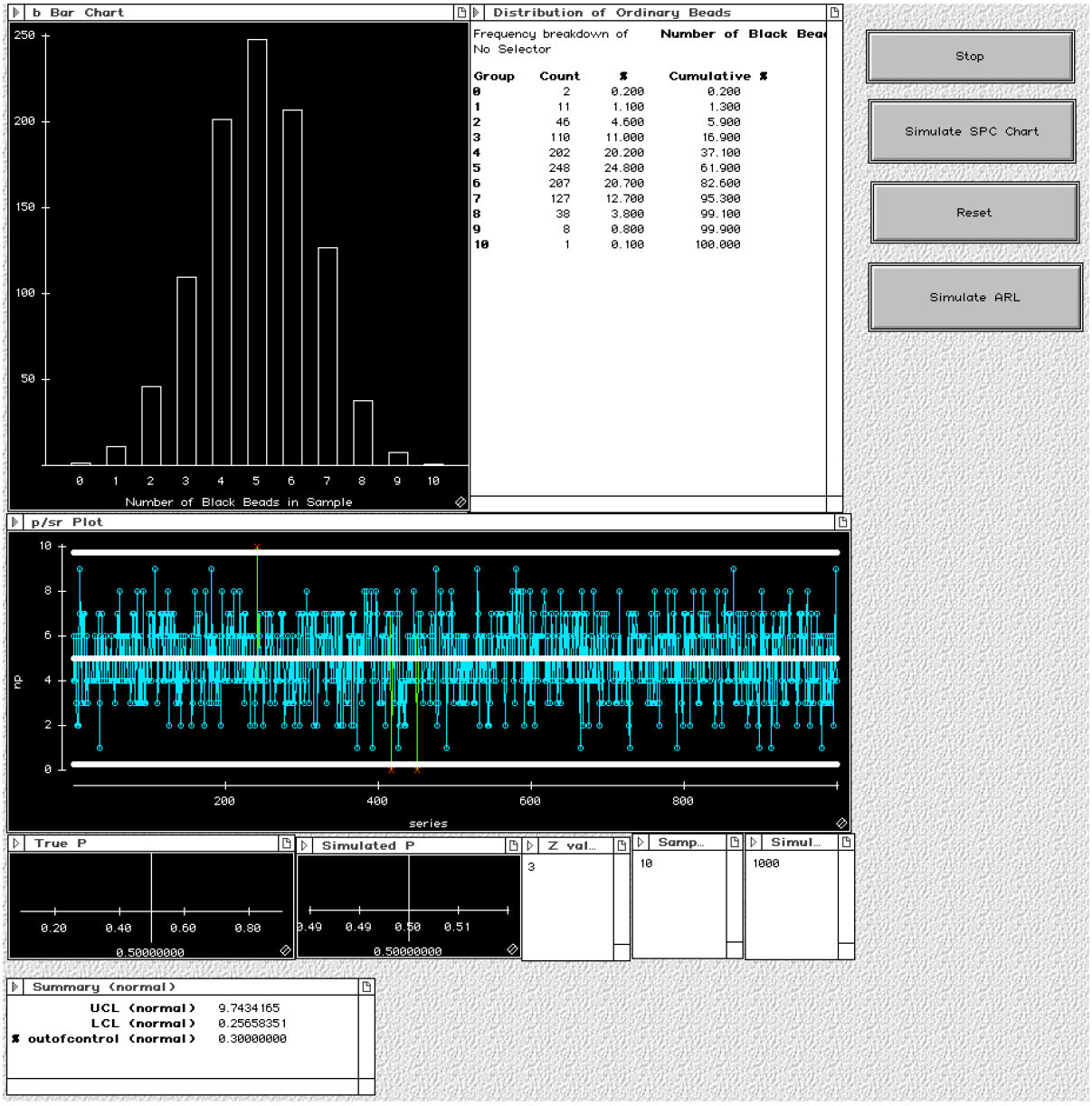


Figure 1.8 Screenshot of application *nP Chart - simulation*

The bottom panel of the interface plot is where the user inputs the required information to generate the simulated chart. The first slider sets the estimate of the proportion of samples with the characteristic of interest (*True P*) for the chart. This estimate can be based on historical data and is used to compute the upper and lower control limits.

The second slider allows us to input a value for **P** that may differ from *True P*. It is called *Simulated P* and allows us to see how the control chart reacts to actual changes in **P**. The remaining three panels are for inputting the desired value of Z (options are 1.64, 2.57, 1.96 and 3), sample size (n) and the number of samples we wish to simulate.

In this example we will simulation a process with P = 0.5, n = 10 and Z = 3.0 (i.e. using 99.97% limits). We will assume the process is in control at P = 0.5 and set the *Simulated P* to the same value as *True P* i.e. 0.5. A total of 1,000 points will be generated.

Selecting *Simulate SPC Chart* an SPC P chart is generated as shown in Figure 1.8.
The upper and lower control limits together with the percentage of points that exceed the control limits are computed. The upper and lower limits are 9.74 and 0.25 using Z = 3 (or 99.97 per cent limits) and n = 10. The results of the simulation in Figure 1.8 are that 3 samples (shown in red) out of 1,000 exceeded the limits. This is equivalent to 0.3 per cent of sample results which is slightly higher than what the theory predicts (i.e. 100% - 99.97% = 0.03%). This arises because the normal distribution is only an approximation (albeit a good one) for this process.

ii) nP Chart - data entry

The application *nP chart - data entry* allows users to input data and the application will construct the chart. The programme interface in shown in Figure 1.9. Data can be copied and pasted into the right hand panel and the chart created. The data in Figure 1.9 uses data from the worked example (page 16) to plot the nP chart automatically.

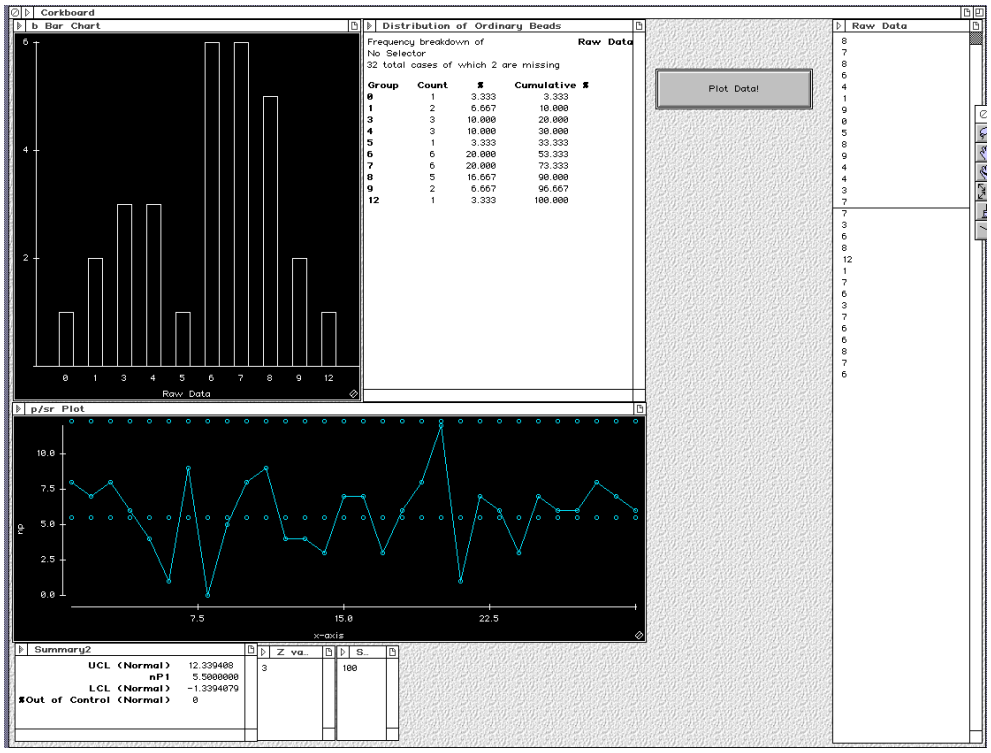


Figure 1.9 Screenshot of application *nP Chart - data entry*

1.4 Summary

Statistical Process Control (SPC) involves, broadly speaking, investigating the state of a process, deriving limits for acceptable variation known as control limits, and ensuring that if the process variation increases beyond the set controls, action is taken and the source of the increased variation is removed. Thus SPC's function is to maintain a stable process within acceptable limits of variation. The concept of SPC is now readily accepted as a beneficial tool for controlling variation, and as an effective tool for continually increasing the quality of a process.

Control charts have the following main advantages:

- The graphical representation of process performance is a considerable aid to evaluating the **stability** of the process.
- SPC charts information on product trends, often allowing for corrective action before an out of control situation arises.
- By determining the type of variation in our process (common or special), SPC charts allow us to determine the appropriate strategies for quality improvement programmes.
- SPC is an objective scientific approach on the stability or otherwise of the process in contrast to 'gut feel' opinion.

1.5 Exercises

1. The data in Table 1.3 is the number of user deaths recorded in a GP's practice by quarter from 1974 to 1998. The practice comprised on average 1,500 users over this time period.

i) Using Excel/DataDesk or Tableau draw an nP chart for this data set

ii) From the plot in i) do you think the monthly deaths are explained by random variation only?

iii) In reaching your conclusion in ii) list briefly what assumptions are you making?

iv) State, giving a reason(s) if you think the assumptions in iii) are realistic?

v) What additional information do you think would be useful in drawing inferences from this SPC plot.

vi) Using the application *nP Chart - data entry* compute an SPC chart for this data set.

Quarter	Year	Number of User Deaths	Quarter	Year	Number of User Deaths
Q2	1974	2	Q2	1987	7
Q3	1974	2	Q3	1987	3
Q4	1974	5	Q4	1987	3
Q1	1975	9	Q1	1988	8
Q2	1975	6	Q3	1988	1
Q3	1975	6	Q4	1988	8
Q4	1977	3	Q1	1989	3
Q1	1978	4	Q2	1989	3
Q2	1978	1	Q3	1989	4
Q3	1978	10	Q4	1989	5
Q4	1978	10	Q2	1990	1
Q1	1979	5	Q3	1990	3
Q2	1979	7	Q4	1990	3
Q3	1979	13	Q1	1991	3
Q4	1979	5	Q2	1991	2
Q1	1980	7	Q3	1991	2
Q2	1980	5	Q4	1991	5
Q3	1980	2	Q1	1992	1
Q4	1980	4	Q2	1992	1
Q1	1981	4	Q3	1992	2
Q2	1981	11	Q4	1992	4
Q3	1981	8	Q1	1993	6
Q4	1981	3	Q2	1993	11
Q1	1982	10	Q3	1993	4
Q2	1982	3	Q4	1993	7
Q3	1982	4	Q1	1994	3
Q4	1982	1	Q2	1994	4
Q1	1983	3	Q3	1994	2
Q2	1983	4	Q4	1994	8
Q3	1983	3	Q1	1995	14
Q4	1983	2	Q2	1995	12
Q1	1984	9	Q3	1995	8
Q2	1984	7	Q4	1995	8
Q3	1984	6	Q1	1996	9
Q4	1984	8	Q2	1996	16
Q1	1985	13	Q3	1996	10
Q2	1985	3	Q4	1996	9
Q3	1985	2	Q1	1997	15
Q4	1985	6	Q2	1997	11
Q1	1986	3	Q3	1997	11
Q2	1986	6	Q4	1997	10
Q3	1986	1	Q1	1998	19
Q4	1986	5	Q2	1998	4
Q1	1987	4			

Table 1.3 User deaths by month

2. The number of users that experienced *Catheter Associated Neoecomial Infection* from monthly samples of 50 users are shown in Table 1.6.

i) Using Excel/DataDesk or Tableau draw an nP chart for this data set

ii) From the plot in i) do you think the monthly variation in infections are explained by random variation only?

iii) In reaching your conclusion in ii) list briefly what assumptions are you making?

iv) State, giving a reason(s) if you think the assumptions in iii) are realistic?

v) What additional information do you think would be useful in drawing inferences from the SPC plot

vi) Using the application *nP Chart - data entry* compute an SPC chart for this data set.

Month	Number of Users Sampled	Total that resulted in Catheter Associated Neoecomial Infections	Month	Number of Users Sampled	Total that resulted in Catheter Associated Neoecomial Infections
1	50	8	19	50	5
2	50	4	20	50	4
3	50	5	21	50	0
4	50	3	22	50	2
5	50	3	23	50	7
6	50	1	24	50	2
7	50	6	25	50	1
8	50	4	26	50	5
9	50	5	27	50	0
10	50	5	28	50	11
11	50	2	29	50	12
12	50	3	30	50	3
13	50	6	31	50	4
14	50	1	32	50	6
15	50	2	33	50	3
16	50	6	34	50	7
17	50	2	35	50	4
18	50	8	36	50	2

Table 1.6: Number of catheter associated neoecomial infections

3. a) The Supplier Quality Department of a high-volume computer manufacturer inspects incoming hard discs at its materials audit area. Samples of 100 discs are inspected each day and the number which do not conform to the manufacturing specification are recorded. The table below is based on the results of seven days' data.

Day	Failed Units	Inspected Units
1	36	100
2	40	100
3	36	100
4	76	100
5	80	100
6	18	100
7	16	100
TOTAL	302	700

i) Draw by hand using graph paper to the nearest whole number a nP Statistical Process Control (SPC) Chart for this data set, clearly specifying the control limits and the centre line.

ii) Plot the sample results on the chart.

iii) State whether the process appears stable or unstable giving a reason for your answer.

b) The operations manager, who is not familiar with SPC charts, is shown the chart constructed in i). The manager is keen to understand the theory of the charts and in particular the basis behind the control limits. Write a short paragraph explaining to the manager the theoretical underpinning of the control limits.

4. The data set below is the number of units inspected and the number rejected per day in a customer audit station in a high volume manufacturing process.

Day	Failed Units	Inspected Units
1	5	50
2	10	50
3	12	50
4	0	50
5	3	50
6	0	50
7	1	50
8	5	50
9	2	50
10	2	50
11	4	50
12	3	50
13	0	50
14	3	50
15	0	50
16	3	50
17	2	50
18	2	50
19	1	50
20	2	50
21	4	50
22	7	50
23	3	50
24	1	50
25	2	50
26	6	50
27	2	50
28	1	50
29	1	50
30	0	50
31	1	50

i) Calculate a nP Chart for this data

a) Using pen and graph paper

b) Using Data Desk, EXCEL or Tableau

ii) Write a short paragraph on the stability or otherwise of the customer audit process.

5. The following data set represents the number of units tested for 48 hours and the number that failed the test over 18 weeks at a product reliability test station.

Week	Failed	Tested
1	605	1000
2	311	1000
3	837	1000
4	354	1000
5	269	1000
6	461	1000
7	137	1000
8	200	1000
9	67	1000
10	660	1000
11	391	1000
12	335	1000
13	220	1000
14	282	1000
15	317	1000
16	237	1000
17	156	1000
18	108	1000

i) Calculate a nP Chart for this data

a) Using pen and graph paper

b) Using Data Desk, EXCEL or Tableau

ii) Write a short paragraph on the stability or otherwise of the Pre-Burn in process.

Appendix 1 - Exploring the Sampling Experiment

Open the application *Sampling Experiment* which is a simulation programme that replicates the sampling experiment and requires the user to enter three quantities as follows:

1. The number of beads selected in each sample. In our example it is set to 10 in the field *SampleSize* but can be set to any number.
2. Proportion of dark beads in the basket. In this example it is set to 0.5 but can be changed to any value between 0 and 1 using using the slider *Proportion Dark Beads in Basket.*
3. Number of simulations i.e. the number of times we want to replicate the experiment. In our example it is set to 10,000 in the field *Simulations* - but it can be set to any number bearing in mind the greater the number of simulations the more accurate the results.

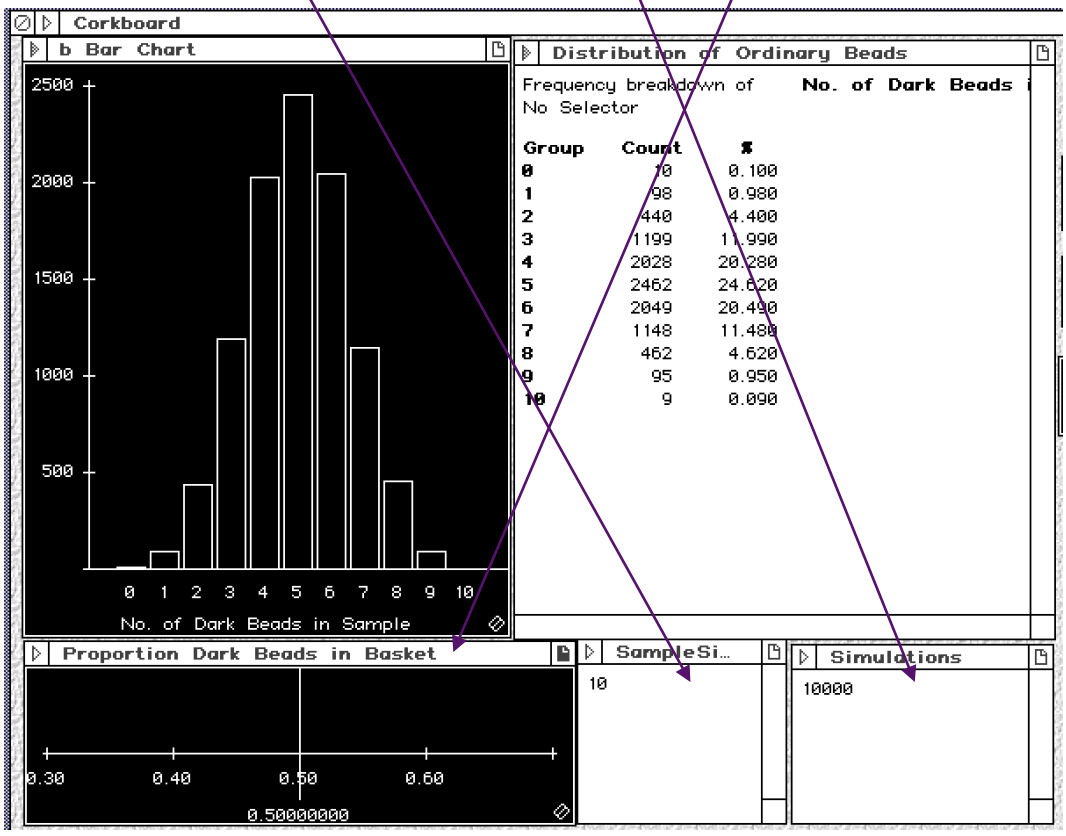


Figure 1.10: Simulation of 10,000 black beads in selections of 10

The table and bar chart located at the top of Figure 1.10 shows the number of dark beads selected in 10,000 simulations - it's as if we asked 10,000 participants to select a sample of 10 beads from a basket, to count the number which are dark and then constructed a table (top left) or a plot (top right) of the results!

From the table we can see for example that just 10 of 10,000 simulations reported 0 dark beads. This accounts for just 0.1 % of all samples and is clearly a rare outcome.

Experiment with this software by changing the of proportion of dark beads (P) and sample size (n). What happens the shape of the distribution as the proportion changes from 0.5? Can you explain why this change of shape occurs?

Appendix 2: Other Common SPC Charts

There are a large number of Shewhart charts named after Walter Shewhart who was instrumental in introducing the charts to Bell telephones in the 1920s. The charts include **P** and **nP** charts for the monitoring the proportion defective for constant and variable sample sizes. **C charts** for the distribution of the number of defects per unit and **X & R charts** for the distribution of the mean and range of sample measurements for continuous data.

Cusum methods developed during the mid 1950s by ES Page are optimal for detecting small shifts or drift in process quality. Popular Cusum charts include charts for monitoring the proportion defective (**Cusum P**), defects per unit (**Cusum C**), sample means (**Cusum X**) and for monitoring data that have been transformed to the standard normal distribution (**Cusum Z**).

The **Sequential Probability Ratio Test** (SPRT) was first proposed by Abraham Wald during the late 1940s. The SPRT is very responsive to upward movements in the proportion defective. A chart similar known as a weighted binomial monitor can detect both upward and downward movements in the proportion defective can also be usefully deployed where the proportion parameter is very low. The weighted binomial and SPRT are very much related to Cusum charts but involve the use of the exact probability distribution rather than using the simpler normal approximation

All the charts outlined in above are available from the writer complete with worked examples using DataDesk applications. The applications incorporate dynamic, interactive and visualisation features which should allow users to obtain a greater understanding of SPC and the role of statistical reasoning in the decision making process.

Appendix 3: Create an SPC Chart Using Tableau

To create an SPC chart of the worked example using Tableau import the data set from Excel. The variables *Sample Number*, *White Beads* and *Sample Size* can then be seen on the measures shelf. To create an nP SPC chart we need to create a number of calculated fields which will compute P, nP, UCL and LCL. The latter three will be displayed on the graph as the centre line (nP) and the control limits UCL and LCL, respectively.

1. Calculate P

The overall proportion P is calculated as:

$$P = \frac{\text{Total number of black beads}}{\text{Total number of samples}}$$

Create a calculated field called P and enter the following code:

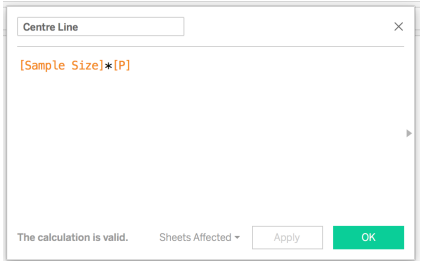
{(SUM([White Beads])/(SUM([*Sample Size*])}

Note the **curly brackets** which is required as it tells Tableau that this is a fixed number rather than an aggregation.

2. Calculate nP

The centre line is nP.

Create a calculated field called nP and enter the following code:



3. Calculate UCL

The upper control limit (UCL) is:

UCL = nP + 3√nP(1-P)

Create a calculated field called UCL and enter the following code:

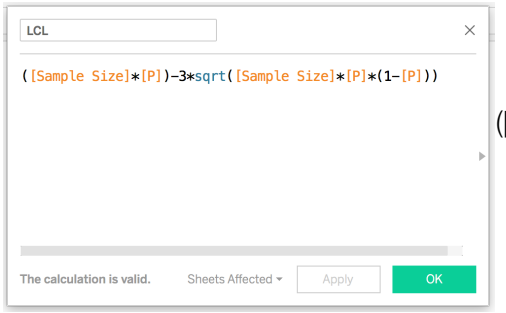
([Sample Size]*[P])+ 3*sqrt([Sample Size]*[P]*(1-[P]))

4. Calculate LCL

The upper control limit (UCL) is:

UCL = nP - 3√nP(1-P)

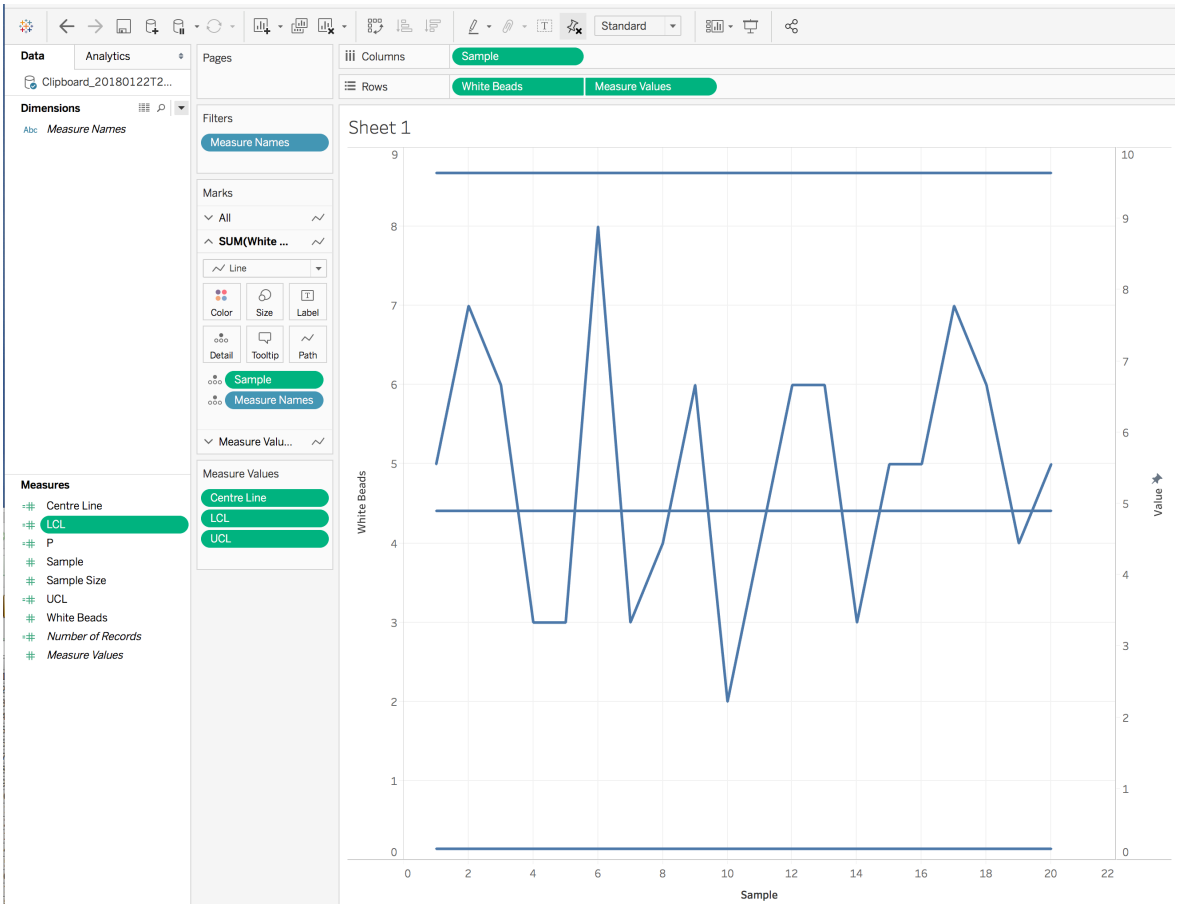
Create a calculated field called UCL and enter the following code:



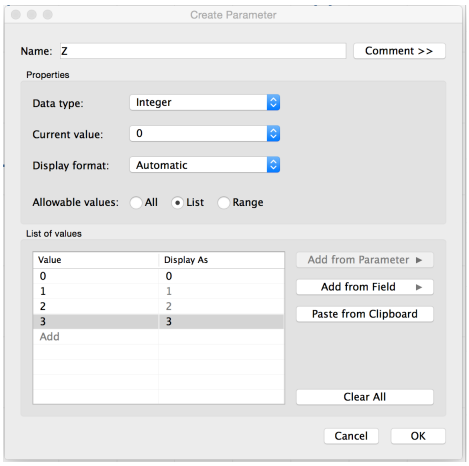
([Sample Size]*[P])- 3*sqrt([Sample Size]*[P]*(1-[P]))

Once these three calculated fields are set up place *Sample Number* to row and *Number of Black Beads* to column. A line plot is generated. Now drag the pill called *measure values* to the row shelf. A graph appears below the time plot. Now remove the measure value pills that are not required until the only pils left which are nP, UCL and LCL.

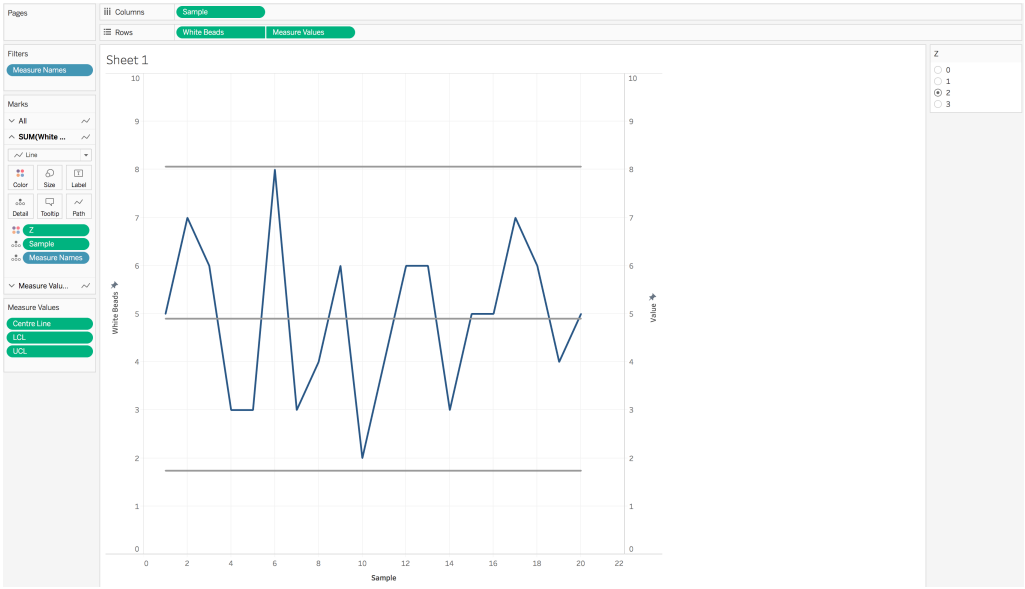
Now select *measure values* pill and the option *use dual axis* from its dropdown menu. Both graphs are now merged. Click on each of the two y-axis of this plot select the option **use independent axis**. Set the same upper and lower bands for both axis i.e 0 and 10. The SPC chart is then generated as shown below.



The standard deviation in the formula is set to 3 but sometimes it can be useful to set it to a different value which is usually 2. The upper and lower limits are then called **warning limits**. To allow the standard deviation to be changed we can create a **parameter** for the standard deviation from the *Create Parameter* menu item attached to the dimensions drop down menu. We will call this parameter Z. The following dialog box appears:



We can list the values we would like by entering 0,1,2 and 3 under the list of values. Z now appears at the bottom of the screen and by selecting **show parameter control** a panel appears on the canvas with values 0, 1, 2 and 3. Changing the values of Z allows updates the chart as shown below where Z = 2 is selected.



Appendix 4: Note on Process Correlation

Most SPC applications depend to a large extent on developing appropriate models that account for the random variation in a process. The distributional assumptions such as **normality**, the existence of **stable process parameters** for generating control limits, and the selection of **random samples** from the process have been stated. However, one important assumption behind the use of SPC methods is that sample outcomes are independently distributed. This means that the sample results over time are independent and not related to previous samples. For example, if 10 dissatisfied users are recorded on Monday this should not influence the number that will be recorded on a Tuesday. If this is not the case the process measurement are **correlated**. Process correlation can lead to control limits that are too narrow or too wide, and can reduce the benefits of a process monitoring programme as too many or too few signals of assignable variation will occur.

Some practitioners interpret process correlation as an assignable form of variation which should be removed. However, in some processes it cannot be removed easily. If this is the case it is possible to construct a correlation free control chart.

The detection of correlation can be difficult and may require the use of more advanced statistical methods known as time series analysis. Nevertheless, it is important to go to the effort of testing for process correlation, as not only will the control limits based on the presence of correlation be more realistic, the pattern of correlation detected may lead to a greater understanding of the underlying process.

1.6 References

Faltin, F. et.al. (2012). Statistical Methods in Healthcare, Wiley.

D. Spiegelhalter, C. Sherlaw-Johnson, M. Bardsley, I. Blunt, C. Wood, and O. Grigg (2012), Statistical methods for healthcare regulation: Rating, screening and surveillance, *Journal of the Royal Statistical Society* (A)*,* **175**(1):1-47

Woodall, W.H. (2006). "The Use of Control Charts in Health-Care and Public-Health Surveillance", *Journal of Quality Technology*, **38**, No 2.