SPC Training

Datalyzer

https://datalyzer.com

Contents

Introduction

Since 1979 Datalyzer is involved with implementation of Statistical Process Control in many companies by providing training, implementation support and advanced SPC software.

During many implementations one of the key success factors is training people in SPC knowledge. Extensive training is not always possible so we have developed a training manual so people can strengthen their knowledge about SPC.

This manual should ideally be used in combination with the training software SPC Wizard but we have changed the manual in such a way it can also be used without Wizard by providing examples.

For further knowledge about SPC please check the website datalyer.com.

What is SPC all about?

To do anything properly we need to understand why we are doing it.

Statistical Process Control is a tool to help you to improve processes and maintain the improvements. Do not confuse this with other activities which involve inspecting or measuring - people test or count things for a variety of reasons. Here are three of these reasons - We have called them selection, numerical and analysis.

1. Selection

This is where someone looks at what has been created in order to separate the good from the bad.

Examples of testing for selection are:

- products are tested and the rejects put on the rework line
- despatch documents are checked and the late delivery is given a special fast track
- pupils sit examinations and the ones with the best marks or the worst marks get additional education

The big problem with this is that any damage has already been done. It is very expensive to rework manufactured products and because screening is never 100% effective, some substandard product will get through to customers.

Another problem is that selection tends to make us think that everything which passes the test is perfect - even the things which pass by a tiny degree. Similarly we are likely to see all "fails" as worthless. SPC helps us to continually improve towards the ideal without having to impose arbitrary pass / fail criteria.

> "Cease dependence on inspection to achieve quality. Eliminate the need for inspection on a mass basis by building quality into the product in the first place." Deming

2. Numerical
Lots of people now want the outcome of processes described in exact numbers.

Examples of the results of processes being turned into simple numbers are:

- a quality meeting discusses how many products failed to conform to specifications last month.
- a government department publishes how well it has performed.
- league tables which claim to show the best worker, the best school or the worst hospital

There is an assumption that things will get better after discussing the numbers but there many pitfalls here. Here are just some of the pitfalls:

1. We may be assuming that the performance of the worker, school or company is the only thing which affects the outcome. In fact these are only the final elements in complex systems - and every part of a system can affect the results.

2. We may be fooled into thinking there is something significant about an unusually good result or an unusually bad result. To make sense of data from processes we need to understand special cause variation and common cause variation. This understanding comes with using SPC properly.

3. We may make important decisions based on the assumption that results from the past are an indicator to future results, but this assumption is only reasonable for stable processes. Only SPC can tell us whether our processes are stable.

4. We may be assuming that bad results will shame people into trying harder but when people try harder in one area, they often make things worse somewhere else. Efforts must be focused on improving process rather than just pushing the existing system harder.

> "Eliminate management by objective. Eliminate management by numbers, numerical goals. Substitute leadership." Deming

The biggest deficiency with testing for selection or for numerical reasons is that it does not actually improve anything. Just knowing how big our problems are, or fixing faulty products does nothing to make them right in the first place.

3 Analysis
SPC is a technique for analyzing processes, with a view to improving them. SPC helps us to answer questions like:

- is the process stable and therefore predictable?

- where should we look to get ideas for improving the process (should we look at special causes of variation or the common causes of variation)?
- has our latest idea really improved the process?
- is the process deteriorating?

Getting answers to questions like these actually increases our understanding of the processes we work with. This leads to real improvements in quality and productivity.

We need to inspect or measure things for SPC but the way we take the samples and measure is likely to be different from the way we would do it for other purposes. For example:

1. We usually do not to need to test and measure a very large number of products for SPC. A small number of samples is enough in most cases.
2. The samples must be tested and measured in EXACTLY the same way each time - standardization of the tests is vital.
3. For selection purposes or for making numerical summaries it may be satisfactory to use go / no-go gauges or turn product or service measurements into a simple OK / not OK. For SPC, measurements from a continuous scale should always be used in preference to pass / fail counts.
4. The data must be plotted on the charts quickly and near to where it has been collected. When special variation is detected we need to work fast to find the cause before evidence disappears.

For these reasons SPC should be seen as an exercise in its own right and not just an extension of existing selection and reporting activities. You might be able to draw SPC charts from data originally gathered for other purposes, but keep in mind that the data not be ideal.

Lesson 1 - Variation and control charts

Before starting any of the lessons below, go through the simulations of Deming's red beads experiment and funnel experiment (click on the buttons on the Training & simulations panel).

In this lesson we are going to reinforce some of the points made in Deming's famous red beads experiment.

Imagine that you are interested in a process and you want to get better results from that process. It may be a manufacturing process or a service process, it may be in the public sector or you may work for a private company. We are going to count a particular type of thing which you would prefer did not happen. We will scoop beads from a box to simulate the process and any red beads scooped are going to represent the thing you would prefer to avoid.

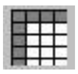
Click the Red Beads button on the Training & simulations panel (If the panel is not open, click the Training & simulations button  on the Main Panel toolbar).

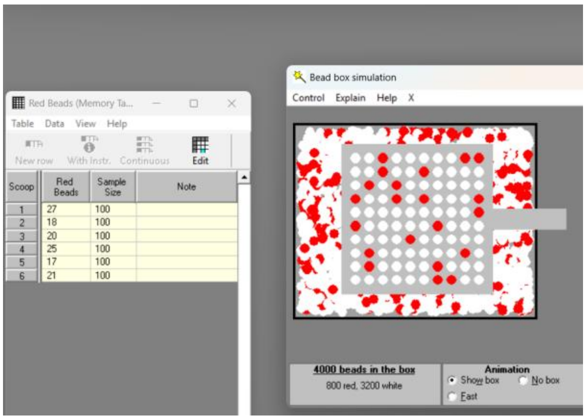
In the Paddle Size frame, click the 100 button

Click the Start button on the simulation window six or seven times and look at the number of red beads drawn with every scoop.

The point to understand here is that it is random variation which is producing the different number of red beads in each scoop. Every process contains some random variation and the people who operate the process have no control over it.

Let's now look at a table of the results.

Click the data table button  on the Main Panel toolbar.



Now, suppose that we did not know that these results were due to random variation. Say, for example, that the figures come from a school and represent the number of pupils expelled for bad behavior each year.

Now click the start button a few more times.

After each scoop look at the bottom 3 or 4 rows of results in the data table.

Scoop	Red Beads	Sample Size	
1	27	100	
2	18	100	
3	20	100	
4	25	100	
5	17	100	
6	21	100	
7	16	100	
8	20	100	
9	17	100	
10	16	100	
11	20	100	
12	23	100	
13	25	100	
14	16	100	
15	22	100	

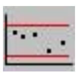
Translate the numbers you find into headlines in your local newspaper:

"Big jump in expellings from school - officials want to know why"

"The number of expelled pupils has fallen since the new headmaster took over."

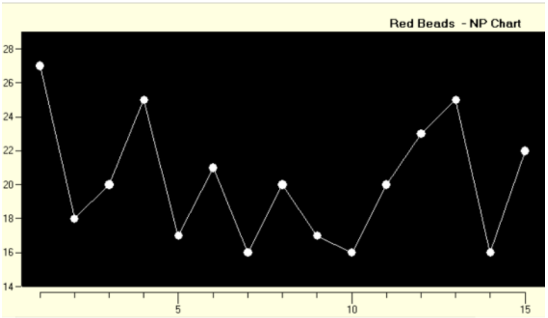
"Number of expelled pupils has increased for 3 years in a row. Experts blame video games."

It is very easy to fall into the trap of assuming that there is always a reason for figures going up or down. There is no reason - other than random variation - to explain why the number of red beads drawn by the paddle changes with every scoop. Every year, random variation alone will cause a different number of pupils to enter the school system with behaviour problems.

Click the control chart button  on the Main Panel toolbar.

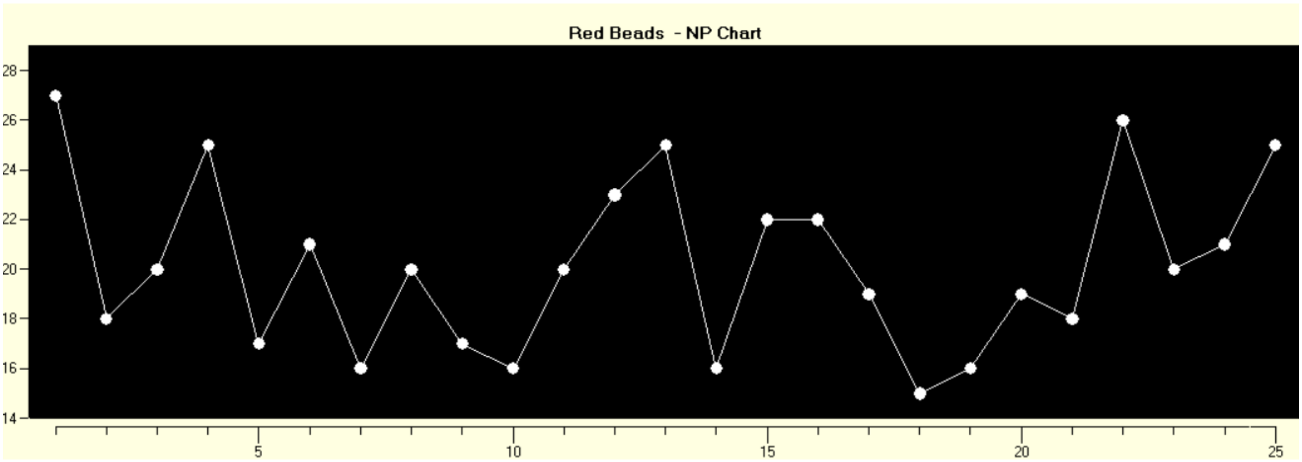
Confirm that Binomial is selected as the chart type

Click the OK button



We now have the results - in time order - on a chart.

Press the start button a few more times on the simulation window. Look at the chart as you press the button.



This is what random variation looks like. Putting past results in time sequence on a chart like this should make us less likely to jump to conclusions about an individual result. We might be less likely to assume that an upward or downward trend of just 3 or 4 results means that a long-term change is taking place.

However, we do need to know if a new policy, or a change in procedures, or a change to a process really affects the results. So we need something which points out the significant changes and encourages us to ignore the random variation. This is where the control limits come in.

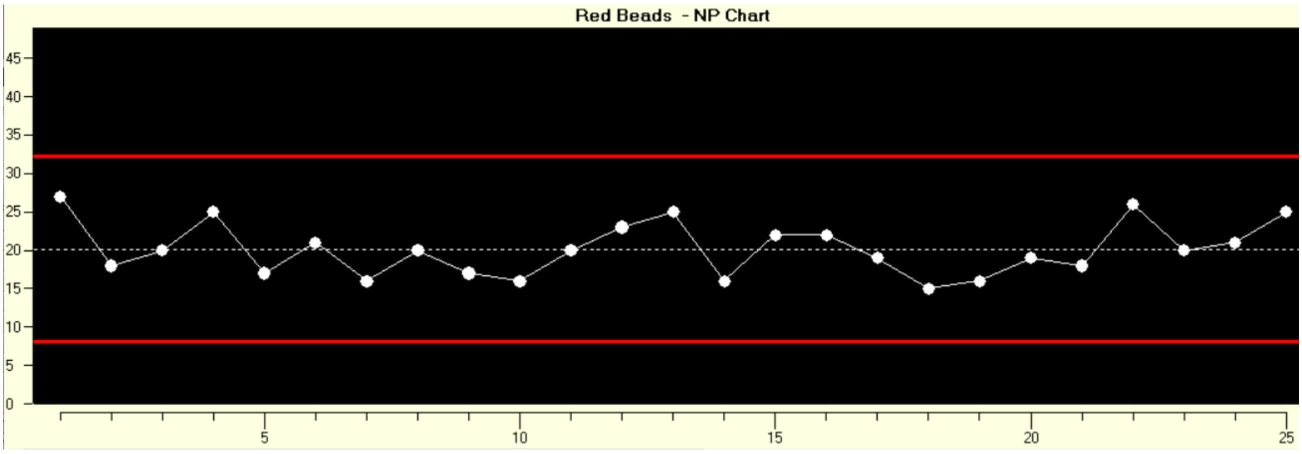
Press the control limits button  on the chart

Put "1" in the "From" box and ensure that the last scoop number is in the "To" box.

Click OK

Press the + key on the keyboard once or twice (if you have a numberpad use that + key). This will allow you to see more points on the chart at once. If the points are too squashed up, press the - key.

Now click the start button again a few more times and keep your eye on the chart.



The new points on the chart should all be between the control limit lines (although you might be unlucky and get a false signal). For now, we will not worry about how the position of the lines was determined.

Continue scooping until you have about 50 points on the chart.

Let's now look at all these points

Make the chart the active window by clicking on it with the mouse.

Although there are lots of ups and downs on the chart, you should be able to get an overall impression that this process is stable. There is no obvious change over the long term.

In this process, whatever causes the low counts also causes the high counts. The things which cause the variation are common to all the results. That is why this type of variation is called common cause variation.

We also say that the results are "in statistical control". We say this because as long as nothing changes, we can predict that this process will continue with approximately the same average, and the control limits will continue to show the maximum and minimum results that we can expect.

Now let's find the average using the Information box.

Ensure that the Information box button  on the control chart toolbar is in the depressed position. Look at the information box. The Avg figure shows the average for the points which were included in the control limits calculations.

Remember the average figure (approximately).

If you want, you could recalculate the control limits using all the points to get a more accurate average.

OPTIONAL:

Click the control limits button  again

Put the last scoop in the "To" box (a quick way to do this is to mouse click on the small button under the box).

Click OK

Remember this average figure (approximately).

Now we will look at a set of results from a very different type of process.

Click on the bead box simulation window with the mouse to make it the active window

Choose 'Existing results' and 'Clear all' from the Control menu. This will remove all the existing results.

Choose "Variation" from the 'Control menu.

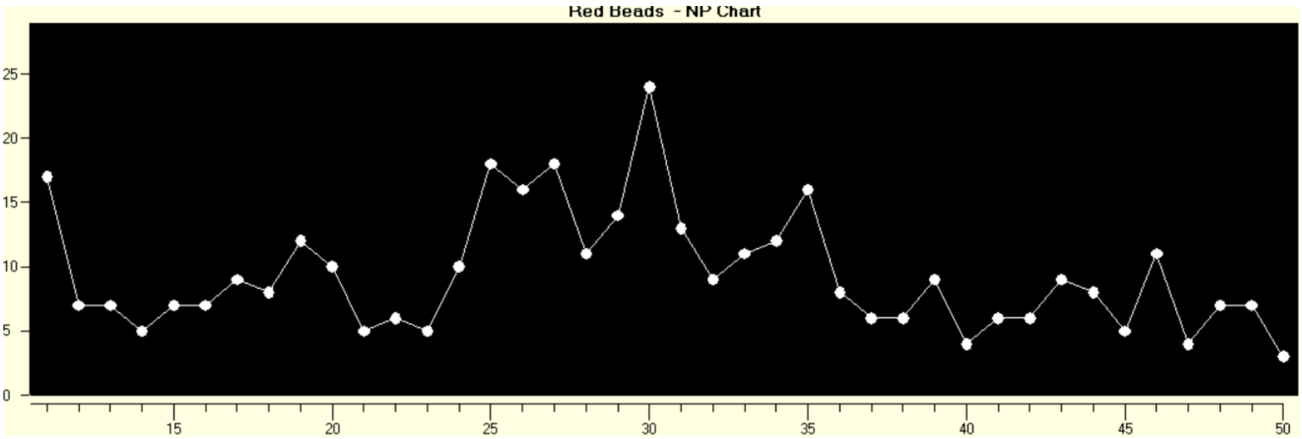
Click the 'Open file' button.

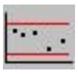
Load the file BOX.PVF

Click OK on the special variation panel to return to the simulation window.

Put the number 50 in the Batch size box Click
the No box button in the Animation frame.
Click Start.

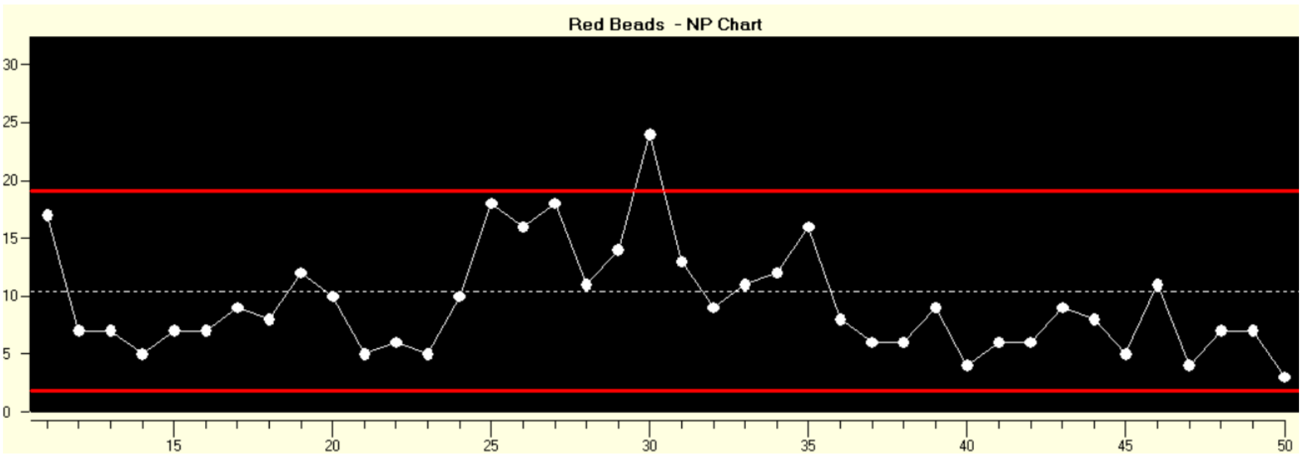
When the 50 scoops finish, look at the results on a control chart.



Click the control chart button  on the Main Panel toolbar.
Confirm that Binomial is selected as the chart type Click
the OK button.

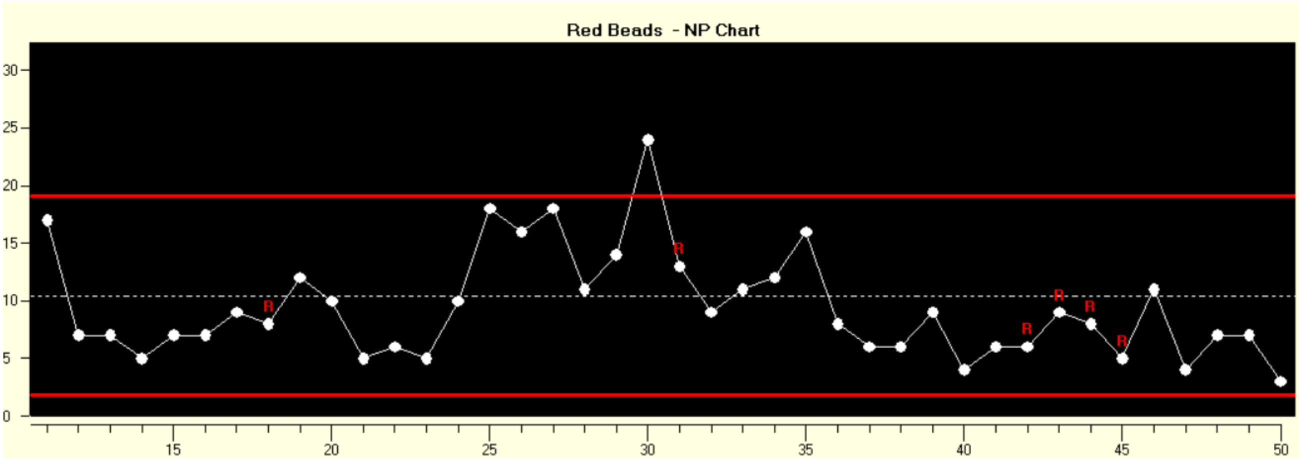
Press the control limits button  on the chart
Enter "1" in the "From" box and ensure that the last scoop number is in the "To" box.
Click OK.
Click the + key to make all the points visible (if you have a numberpad use that + key,).

Ensure that the  button on the chart toolbar is in the depressed position to show the information box.
Look at the average (Avg) figure.



You will notice that it is about the same as the average figure for the previous results but the chart looks very different.

Click the small down arrow beside the 'Runs' button  on the control chart toolbar. Select the number 8 from the list which drops down.



This chart is showing that the process is unstable. There are two things on the chart which indicate instability:

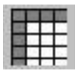
- some points are outside the control limit lines
- some points have an "R" above them which indicates a "run". The previous 6 points are on the same side of the center line - it is unusual for random variation to produce 7 consecutive points above average or 7 consecutive points below average.

With this process, there is something else affecting it alongside the common causes of variation. Look at the results between 25 and 37 on the horizontal scale and compare them with the results between 37 and 50. Something must be causing them to be different. We call this a "special" cause of variation.

When special causes of variation are present in a process we say that the process is "unstable". With an unstable process you cannot predict future results because we do not know when the special causes of variation will occur. If we have an unstable process, we should, whenever possible:

- investigate what is causing the special variation,
- learn whatever we can from the investigation
- improve the process by making the best conditions permanent
- put controls in place to prevent the special variation from returning

We have arranged that no more special variation will occur, so let's scoop some more beads. But first we will write a note on the control chart.

Click the data table button  on the Main Panel toolbar.

Click on the cell in the Note column and sample number 50 row.

Click the Edit button  at the top of the table.

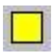
Type "Special variation removed" and hit the ENTER key.

Return to the simulation by clicking on it with the mouse Put

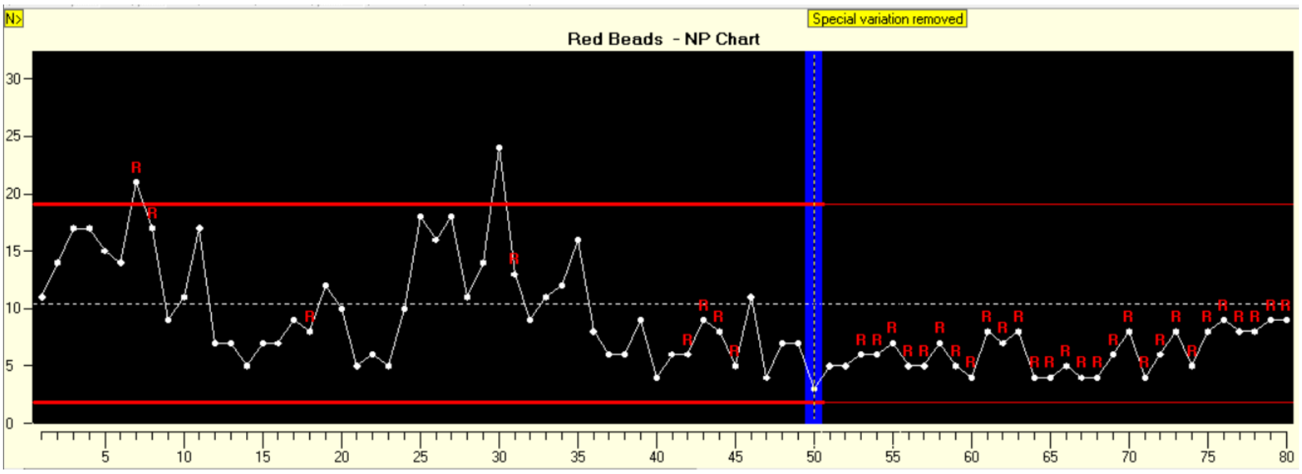
the number 30 in the Batch size box.

Click the Start button.

Look again at the control chart.

Make sure that the Show Notes button  on the control chart toolbar is in the depressed position.

Mouse click on the yellow square which contains the letter "S" above the dotted line.



This was the point that the special variation was removed.

There seems to be consistently fewer red beads now and the red beads represent something that we do not want. So there is reason to believe that we have improved this process. We should now recalculate the control limits using results which come after the improvement.

Click the control limits button  on the chart

Put the number 50 in the "From" box or click the button under the box

Click the button under the 'To' box, this puts the final scoop on the "To" box.

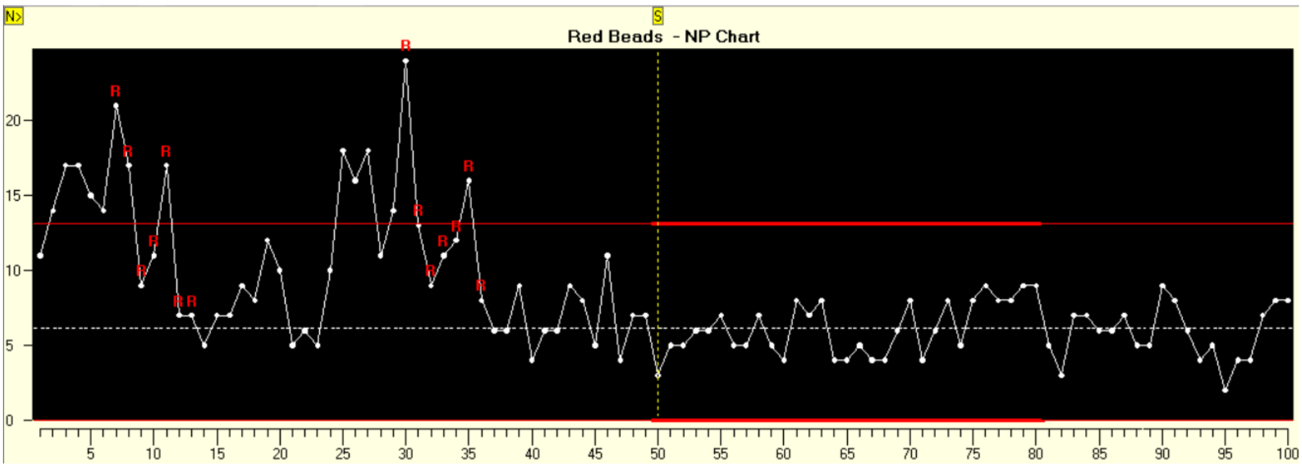
Click OK

We now have new control limits which indicate the limits of common cause variation for the new improved process.

Return to the simulation window by clicking on it with the mouse

Put the number 1 in the Batch size box

Scoop some more beads. Keep your eye on the control chart.



The results should all be between the new limits.

These new limits will show if any special causes of variation appear. If one does appear, we should investigate it and remove it.

If we want to improve a process which contains only common cause variation, we will need to investigate the factors which are constantly affecting the process and influence every result.

Lesson 1 summary:

1. All processes contain variation.
2. We must distinguish between special cause variation and common cause variation. We need to know this difference because the things we will have to do to remove or reduce the two types of variation are very different. We need to reduce variation to improve processes.
3. The way to distinguish between common cause variation and special cause variation is to use a control chart.
4. Before we can consider a process to be "under control", efforts must be made to remove special causes of variation. We must also learn from each incident of special variation and take action to make sure that these types of changes do not happen again.

5 If we want to improve a process which contains only common cause variation, we need to investigate the factors which affect every result.

End of Lesson 1

Lesson 2 - Xbar & Range control chart

In lesson 1 we discovered why we need control charts. Now we are going to learn how to draw the control charts.

We use different types of control charts for different types of data. Data can be divided into two major categories, variables and attributes

Variable data is any measurement which has a continuous scale. For example:

- length
- weight
- temperature
- pressure

Attribute data is based on discrete counts. For example:

- the number of blemishes on a surface,
- the number of faulty products
- the number of unpaid invoices

With variable data we can measure to any accuracy that we want, for example 12.5, 3.075 etc. Attribute data, on the other hand, can only have whole number values like 1, 3, 12 etc.

In this lesson we are going to look at a process where we have been given a preferred target value for a variable measurement. Our job is to try to get results as close as possible to that target.

Click the Tennis ball launcher button on the Training & simulations panel.

Choose "Variation" from the "Control" menu

Click the "Open file" button

Load the file "UNSTABLE.PVF"

Click the "OK" button on the Set Variation panel.

Imagine that you are the operator of a machine - the launcher. Your job is to fire balls at the target and get them to land as close as possible to the ideal value of 500. First we need to 'center' the process.

Fire one ball from the launcher.

Ensure that '1' is selected in the 'Shots' frame in the launcher simulation.

Press 'Start'.

Look at the figure for Landing position.

The customer of this process wants the landing position to be 500, so what do we do now? Should we compensate for the error by moving the launcher?

The answer is NO. There is always a certain amount of variation in every process and Deming's funnel experiment showed us that trying to adjust for this variation can actually cause more variation on the output.

Fire four more balls.

Press 'Start' on the launcher simulation four times.

Look at the display of Landing positions at the right if the simulation (if some parts of the simulation have been erased, press F5 to refresh the display).'

Each landing position is different and the variation may due to common causes which are always present in this process.

We still do not have enough information to make adjustments to accurately land as many balls as possible as close as possible to 500.

Fire off 50 more shots.

Click on '50' in the 'Shots' frame in the launcher simulation.

Press 'Start'

When all the balls have been fired, calculate the mean or average landing position of all 55 shots.

Choose Average from the Show menu.

Use all the results in the calculations (i.e. accept the default values in the From and To boxes).

Click on Calculate.

Note the Average value - whole number only, ignore what comes after the decimal point.

Press Cancel or Close.

Now move the launcher so that the spread of future shots will centre on the 500 mark.

Subtract the Average value from 500. You have to do this manually

Move the launcher by this amount (e.g. if you need to move the launcher 93 to the right, change the text box in the Launcher Position frame to 1093 or +93, then click 'Move').

Assuming that nothing changes in this process, future output should now be centered on 500 (do not worry if you got the calculations wrong, it is not important for the rest of the exercise).

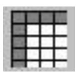
"Assuming that nothing changes" is a big assumption. At this stage we do not know very much about our process and we do not know if things are likely to change over time. So, we need to find out if the process is "statistically stable". Fire off another 100 shots then we will create a control chart.

With 50 selected in the Shots frame, press Start twice

The landing position is a continuous scale. We are actually using complete numbers for the results (423, 657 etc.) but there is nothing to stop us using more accurate measurements if we wanted to (423.45, 657.09 etc.).

This type of data is called variable data. We will use an Xbar and Range chart as the control chart for this process. In an Xbar and Range chart, the data is arranged into subgroups.

Let's take a quick look at the data table.

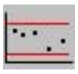
Click the data table button  on the Main Panel toolbar.

Notice that there are 5 columns with "Landing position ( )" at the top. The number in brackets means that the column is part of a subgroup. For each row of the table, the data in these five columns represents one subgroup. Look at subgroup 25.

Look at the row with the number 25 in the grey column at the left. Click on the scroll bar at the right hand side of the table if necessary to bring the row into view.

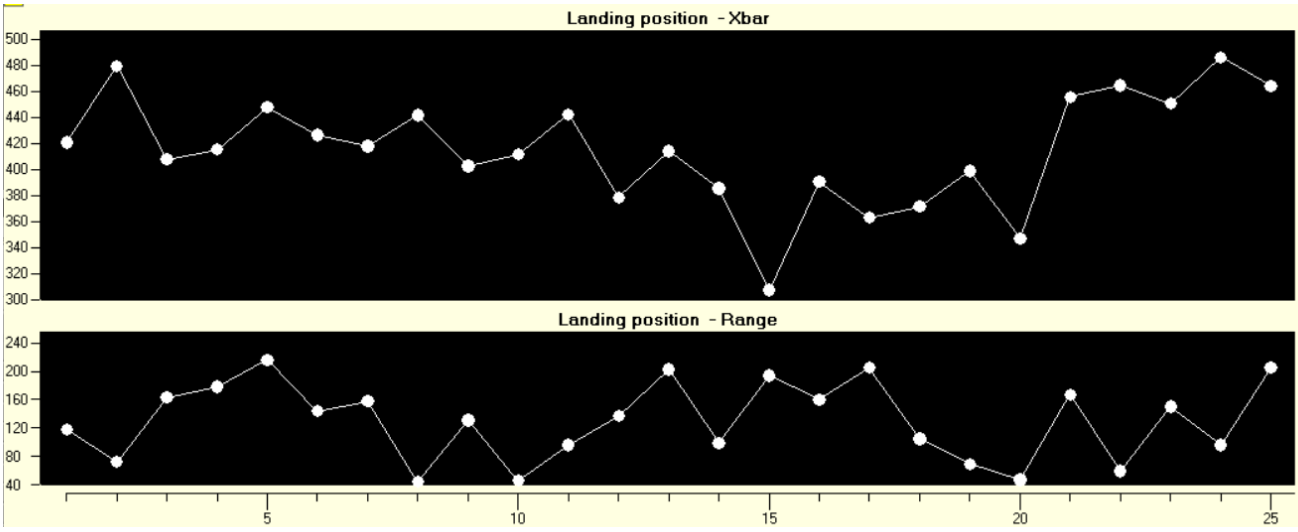
This subgroup starts at shot number 121. The subgroup columns contain the results of shots 121, 122, 123, 124 and 125.

Now let's look at the control chart of this data:

Click the control chart button  on the Main Panel toolbar.

Confirm that Xbar and range is selected as the chart type

Click the OK button



An Xbar and range chart contains two graphs. For each row in the data table, the subgroup average is plotted (Xbar) along with the largest value in the subgroup minus the smallest value (Range). Let's confirm this:

Ensure that the information button  button on the chart toolbar is in the depressed position. This will show the information box at the right of the charts.

Click on any point on the chart.

Confirm that the value on the Xbar chart (shown beside "Xbar" in the information box) is the average of the five "Landing position" columns in the highlighted row in the data table (add up the five values then divide by 5).

Confirm that the value on the Range chart (shown beside "Rng" in the information box) is the subgroup range (the highest value in the subgroup minus the lowest value).

Next we will calculate the control limits and draw control lines on the chart. The purpose of these lines is to show when we should suspect that something has changed which affects the process (in other words, a special cause of variation has occurred).

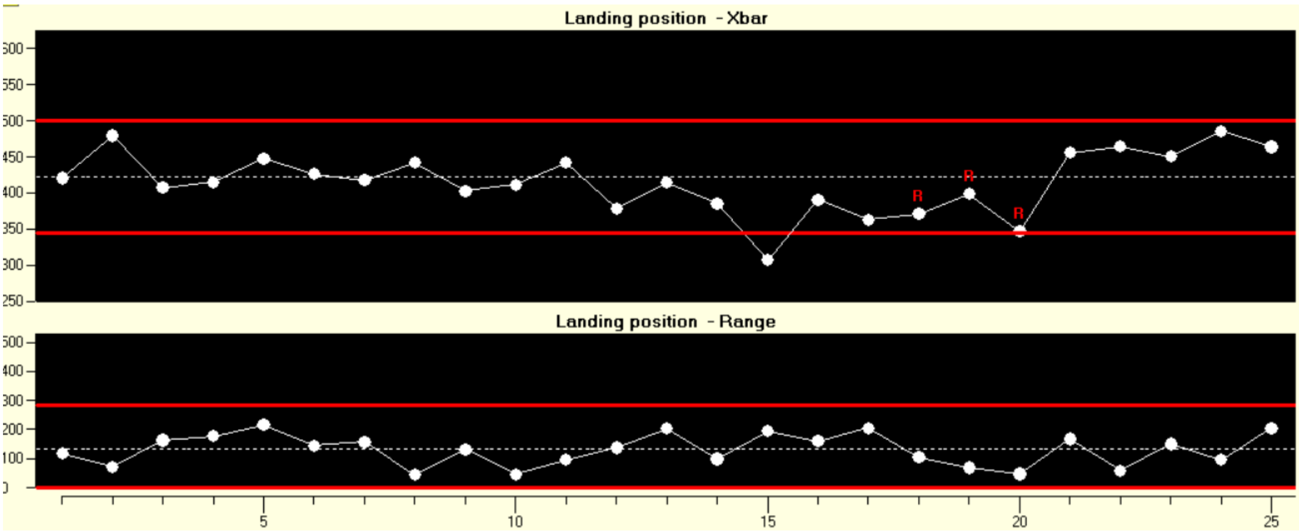
We calculate the control limits from a section of the data. Of course we know that the launcher has been moved and moving the launcher is a special cause of variation, so we should calculate the limits with results which come after the launcher move. For now, we will not worry about how the calculations are made.

Press the control limits button  on the chart

In the From box, enter 12, this subgroup number was just after the launcher was moved (if you clicked on the note line on the control chart before opening this dialogue box, you can enter the number simply by clicking on the button under the From box)..

Make sure that the last subgroup is in the To box (you can do this quickly by clicking on the button under the box).

Click OK



If all the results after moving the launcher are within the control limits, then the process is probably stable or "in statistical control". This means that all the variation comes from common causes. Common cause variation is just the normal random variation which is inherent in the process.

If some of the results after moving the launcher are outside the lines, then this would probably be due to a special cause of variation. The process would be described as "unstable" and because there is every reason to believe that this type of change will happen again, it will be very difficult to keep the average centered on 500.

We know from this process the launcher was moved at subgroup 14 and at subgroup 20.

If we want to be in full control of a process we must use the charts to identify when special cause variation occurs, determine if things were better before or after the change, then make one of these situations permanent.

Let's carry on producing. Fire off another 100 shots

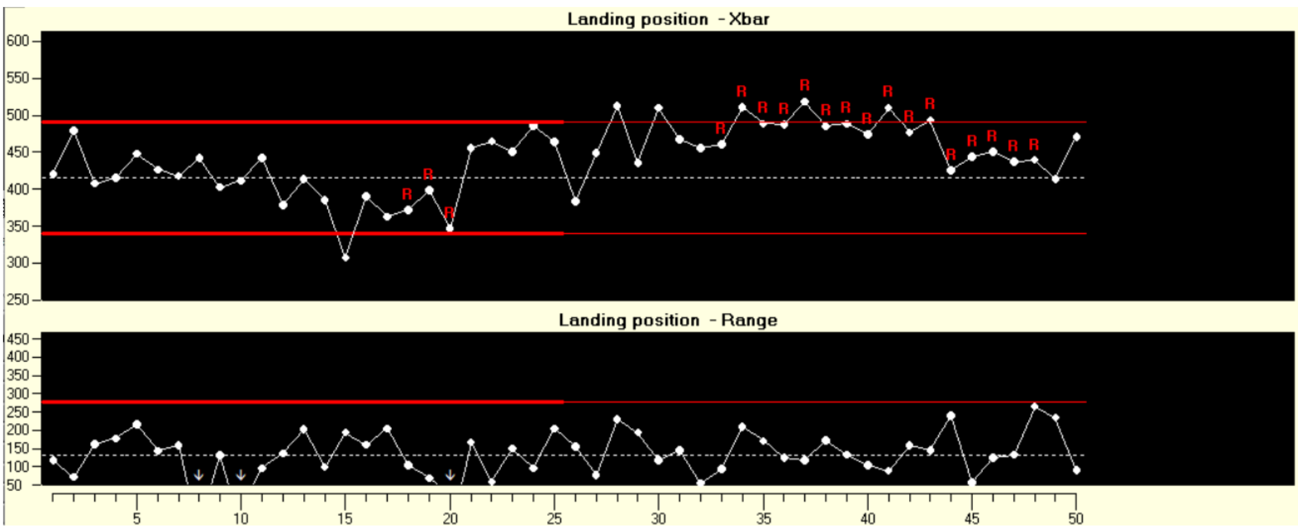
Click the simulation button  on the main panel.

With 50 selected in the Shots frame, press Start twice

Look at the control chart and do not recalculate the control limits. All we need to know now is whether there has been any change in the process since the lines were calculated. Is the output stable?

Click the control chart to bring it to the foreground

Look to see if any of the points are outside the control limits (ignore the points before the control limits were calculated)



It looks as if something unusual happened around subgroup 44 (launcher shot 220). Subgroup average drops below the control line so a special cause of variation has occurred. As an operator, your job is to produce results as close as possible to 500 but the average landing position has suddenly changed.

You could, of course re-center the process (move the launcher). This might help in the short term, but you have no idea whether things might suddenly change back to normal. The only really satisfactory solution is to carry out an investigation, find the source of the special cause of variation, learn from what happened, then make sure that this kind of change does not occur again.

A word now about specification or tolerance limits. In most industrial processes, the operator is given specification or tolerance limits as well as the target value. However, these specification limits should always be looked on as representing the MINIMUM acceptable quality from the process. World class quality does not come from treating everything within the specification limits as equally acceptable. We must try to produce as close as we can to the target value. This is what the customer really wants.

In our bouncing ball process, an unknown special cause of variation made the subgroup average fall at around subgroup 40 (shot 200). It might be that the individual results are still within the specified tolerance limits, but our customer would prefer the results to be 500. So we must make efforts to produce with the average output at 500 and the minimum variation that our process is capable of. So we must investigate and remove special causes of variation even if we are still producing within the specification or tolerance limits.

At this point you should be at subgroup 51 and shot 255. I can tell you that investigations show that that one batch of balls has slightly less bounce than normal. We discard this batch and demand from our supplier that they supply us with statistically stable product (they can only be sure of doing this by using control charts).

I have removed this special cause of variation so things should return to normal from shot 256. Fire off another 50 shots to see this. So this time, there is no need to change the position of the launcher.

Click the simulation button  on the main panel Fire off another 50 shots.

The control chart should show clearly that a change occurred around subgroup 44 (shot 220) and things returned to normal around subgroup 51 (shot 256).

Bring the control chart to the foreground

Look at the plots on the Xbar chart

Now we will look at how the control limits were calculated.

Click the View menu on the control chart

Select "Show control limit calculations"

Look at how the limits are calculated

The mathematics are not difficult. You need to find the average Xbar (average subgroup average) and the average range for the section of the data that you use to calculate the limits. You also need to look up a table to find constants called "A2" "D3" and "D4" for the subgroup size that you are using.

Click the Close button on the panel showing the calculations.

For the purposes of teaching, the computer can generate large quantities of data easily and then plot the data onto charts. In real processes, operators should plot the results without delay onto control charts. If the process operator keeps an eye on the charts, then he or she should know quickly when a change occurs. This can make it easier to find the special cause of the variation. Because the mathematics of control charts is not difficult, you do not need to have a computer at every point that a control chart would be useful. Simple charts on graph paper are just as effective.

By distinguishing between special cause variation and common cause variation, control charts can help operators and managers to run processes which produce on-target with minimum variation.

If special cause variation is present, we must find the root cause and stop this from occurring again in the future. We ask the questions:

"What happened at approximately that point to change the results?". and

"How can we prevent this from happening again?"

If no special causes are present and we want to get better results, we ask different questions:

"Looking at **all** the results, is the average off-target?" and

"Looking at **all** the results, why is there so much variation?"

To reduce common cause variation we might need better machinery, more frequent maintenance or less common cause variation within raw materials.

Lesson 2 summary:

1. When we are given a target or ideal value for something, we should always try to get results with the Average on target and with the minimum amount of variation. We should not just try to get results within the specification limits.
2. Control charts can help to distinguish between common cause variation and special cause variation.
3. An good type of control chart for variables is the Xbar and Range chart. Xbar and Range charts use data arranged into small subgroups.
4. It is not enough just to react to special causes of variation by adjusting the process to compensate. World class quality comes from removing the special causes of variation and preventing them from returning. Because this often requires management action, SPC will only work properly when managers understand the role they have to play in creating stability and reducing variation.

End of Lesson 2

Lesson 3 Histograms and distributions

A histogram is a way of showing a set of measurements as a picture.

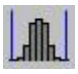
Let's fire some balls from the tennis ball simulation and then look at a histogram of the landing positions.

Click the Tennis ball launcher button on the Training & simulations panel.

Click on '10' in the 'Shots' frame in the launcher simulation.

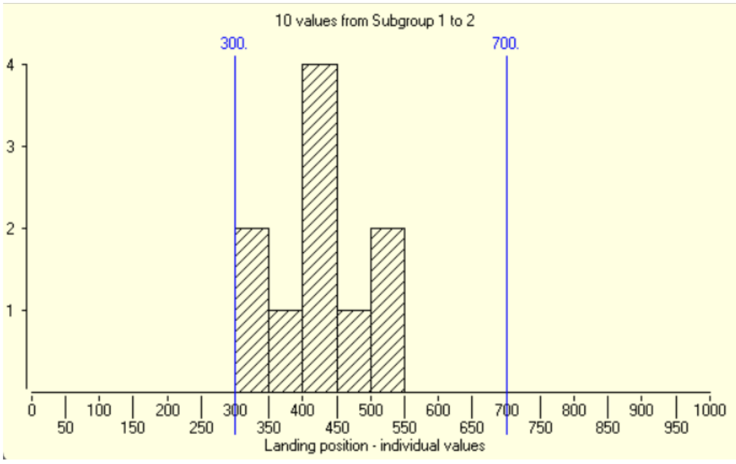
Press 'Start'

Look at the histogram

Click the histogram button  on the Main Panel toolbar.

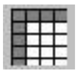
Confirm that "Individual values" is selected under "Make the histogram from"

Click the OK button



In this histogram, the measurements are the landing positions of the balls from the launcher simulation. The possible landing positions are set out on the horizontal scale and this is divided into a number of sections. For each section, a column is drawn and the height of the column represents the number of balls which have fallen within that section of landing positions. For example if the column between "400" and "450" on the horizontal scale is 2 units high, this means that 2 balls have landed between 400 and 450.

Let's check that the histogram has been drawn correctly for our data.

Click the data table button  on the Main Panel toolbar.

Choose one of the columns on the histogram which is 1 unit high

Look at the numbers at either side of the column on the horizontal scale.

Confirm that there is only one result between these two values in the 5 "Landing position" columns of the data table (check the 10 cells in the two rows under the "Landing position" columns).

If there is a column which is 2 or 3 units high, confirm that there is 2 or 3 results between the two values for that column in the "Landing position" columns of the data table.

You have fired 10 balls in the launcher simulation, confirm that the histogram columns add up to 10.

Now we will make the histogram contain more results.

Click the 50 option in the shots frame of the simulation.

Click "Start"

Click on the histogram to see the results

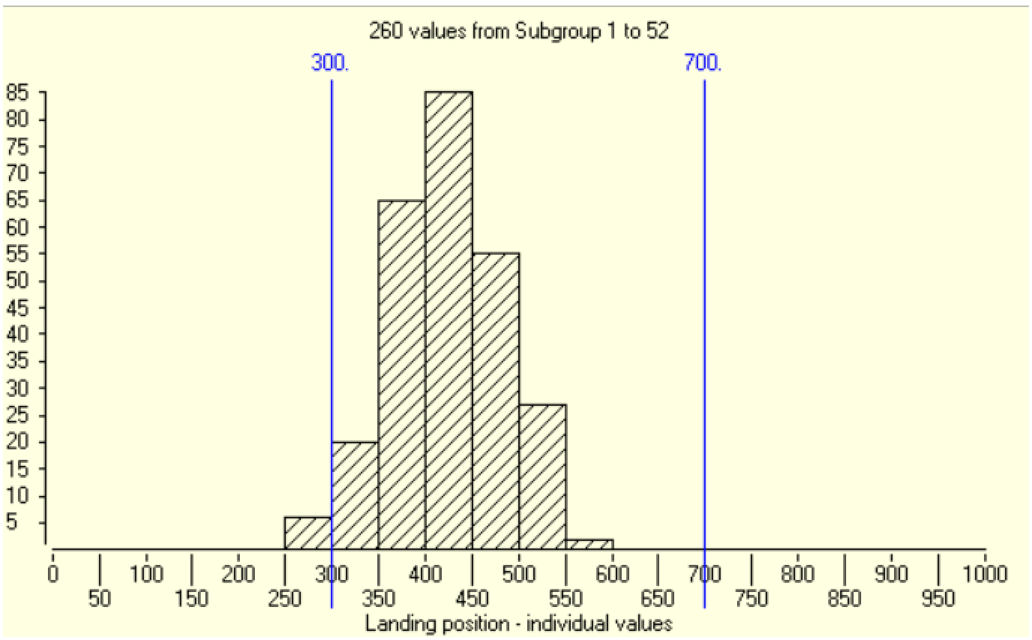
You may need to adjust the horizontal scale to see all the columns,

Hold down the "Control" key and click on the histogram below the horizontal scale Lets put a still more results in the histogram.

Click "Start" a few more times,

Keep looking back at the histogram after each batch of shots

Control click the horizontal scale to make sure all the columns are visible



It should become clear after you have fired a hundred or so balls that the highest columns are near the middle of the histogram. This means that most of the balls land near the middle of the range of possible results.

The shape of the histogram is call the "distribution". It shows how the measurements are "distributed" among the range of possible measurements.

A particular bell-shaped histogram pattern is known as the 'normal' distribution. It occurs frequently in nature and is common in industrial processes. We know a great deal about normal distributions and this helps us to make some general statements about the outcome of processes.

We will now produce a normal distribution.

Choose "Existing Results" and "Clear all" from the simulation "Control" menu.

Move the launcher to position 1078 (click the help button in the Launcher Position frame if you do not know how to do this)

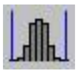
Hold down the "Control" key and click in the "Shots" frame

Put the number 1000 in the rightmost box in the "Options for Number of Shots" frame then click OK

Select the 1000 option in the shots frame

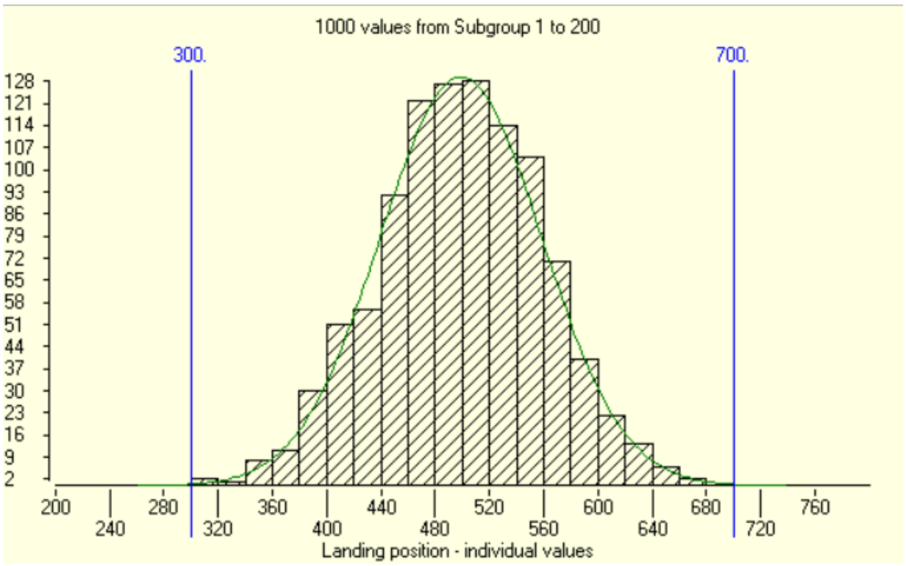
Click "Start" to fire the 1000 shots

Now we will look at a histogram of the 1000 shots

Click on the histogram button  on the main panel.

Confirm that "Individual values" is selected to make the histogram from Click

the OK button.



Notice that the histogram shape looks like a bell with a high middle and tails at each end (screw up your eyes and use you imagination a bit).

Click the indices button  on the histogram toolbar.

This box shows some figures for the data which is used to make the histogram. We are going to describe the first two figures under the "Statistics" section.

Average:
The Average is calculated in the normal way from the individual results:

The individual results are added up then divided by the number of results.

Standard Deviation:
'St dev' means Standard Deviation.

Standard Deviation gives us a figure for how much the individual values in a set of measurements are spread around the Average. A set of measurements where most of the values are near the Average has a low Standard

Deviation, a set of measurements where most of the values are far away from the Average has a high Standard Deviation.

You can create and use control charts without knowing how to calculate Standard Deviation. For those who want to know, here is how to calculate a Standard Deviation:

First, find the Average of all the values

For every individual value, find the distance from the Average.

Square this number (multiply it by itself).

Add all the squares together

Divide by the number of measurements minus one.*

Find the square root of this number.

* You may come across calculators which divide the sum of the squares by the number of measurements rather than the number of measurements minus 1. Throughout SPC Wizard, the "n minus 1" method is used.

I repeat that you do not have to remember how to calculate Standard Deviation to draw control charts or to use the charts to improve processes. All you need to know is that Standard Deviation is a measurement of spread.

For Normal distributions, we can use Standard Deviation to make some useful statements about a set of measurements. We can also make some predictions about future measurements from the same process if it is reasonable to assume that the process will not change (it will only be reasonable to make this assumption if the process is stable).

If past measurements show a normal distribution, and the process is stable, then we can say :

As long as the process stays stable:

about 68% of results will lie between one Standard Deviation below Average and one Standard Deviation above Average,

about 27% will lie between one Standard Deviation and two Standard Deviations from the Average,

about 4.5% will lie between two Standard Deviations and three Standard Deviations from the Average,

only a very small proportion (about 0.3%) will be more than three Standard Deviations from the Average.

Let's see if this is true

Click the indices button on the histogram toolbar to remove the information box.

Click the colours button on the histogram toolbar

The green zone is up to 1 Standard Deviation either side of the Average.

The yellow zone is more than 1 but less than 2 Standard Deviations from Average The

purple zone is more than 2 but less than 3 Standard Deviations from Average

The red zone is more than 3 Standard Deviations from Average

Ensure that the specification limits button is in the depressed position.

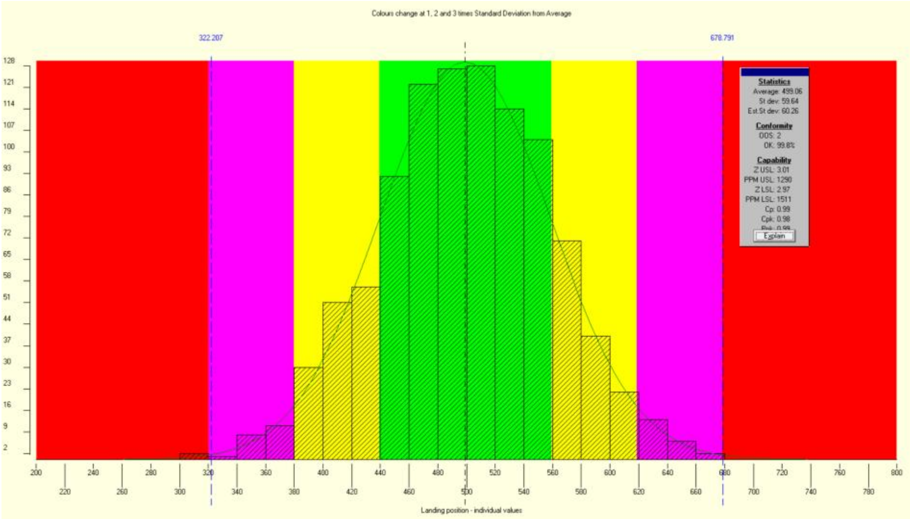
Click the indices button on the histogram toolbar to show the information box again.

The blue lines show the specification limits for the launcher simulation. In the "Conformity" section of the information box you will see figures for the number of results which are out of specification (OOS) and the percentage of the results which are within specification.

SPC Wizard has a special feature to allow you to drag the specification limits to different positions on the histogram. The figures in the Information box recalculate as you drag the lines This does not change the real specification limits, you can return to them at any time - just click anywhere on the histogram..

Use the mouse to drag specification limit lines, place the lines at 3 times Standard Deviation from Average (this is at the join between the red zones and the purple zones).

If the Information box is obscuring one of the lines, switch it off by clicking the button on the histogram toolbar, move the specification line then click the Information button again. Alternatively, you could move the Information box to another location on the histogram- drag the blue bar at the top using the mouse.



The percentage OK figure now shows how many of the results are within 3 Standard Deviations from Average. If you subtract this from 100% you will get the percentage outside 3 Standard Deviations. The exact figure will be different each time you generate results from the launcher but it should be very small (in the region of 0.3%). Now drag the specification lines to the join between the yellow and the purple zones.

The lines are now at 2 times Standard Deviations from Average. The percentage OK figure should still be large
(around 95%). So for a normal distribution, only around 4% or 5% of the results are between 2 Standard Deviations and 3 Standard Deviations from Average.

Drag the lines to the join between the yellow and the green zones.

The lines are now at 1 Standard Deviation from average. The percentage OK figure will be somewhere around 68%.

So, we now know that for a normal distribution, the majority of results will be less than one Standard Deviation from Average. However, we also know that there will be a small number of results more than 3 times Standard Deviation from Average. We cannot tell WHEN these extreme results will happen, but we know that they will happen sometime.

Click the histogram to restore the correct specification limits.

Using statistics to predict the future:
These percentages tell us approximately what has happened in the past. but often we are asked what sort of results we will get in the future from a process.

We can predict that the process will continue to produce roughly the same proportion of results in the 1, 2 and 3 standard deviation zones IF THE PROCESS DOES NOT CHANGE IN ANY WAY.

Also, the percentages given above for the normal distribution are only true over the very long term. It would be wrong to suggest that we can tell with confidence what the measurements will be in any one particular batch of goods.

Lesson 3 summary:

1. A histogram is a way of showing a set of measurements as a picture.
2. The shape of the histogram is know as the distribution.
3. Standard Deviation gives a figure for how much spread there is in a set of measurements.

4. If past measurements show a "normal" distribution, and the process is stable, then as long as the process remains stable, we can predict the approximate number of results which will be at different distances from the Average.

End of Lesson 3

Lesson 4 - Early instability and the X chart

In lesson 2, we used a section of data to calculate control limits for a process. In that example, the process was stable in the early stages and we used data from that period to calculate the control limits.

In this exercise we are going to investigate what happens if the process is unstable while producing the data which is used to calculate control limits.

Click the Tennis ball launcher button on the Training & simulations panel.

Choose "Variation" from the "Control" menu

Click the "Open file" button

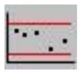
Load the file "XBAR2.PVF"

Click "OK" on the Set variation panel.

Fire off 100 shots and then create a control chart. We will not worry about centring the process:

Select 50 in the shots frame.

Click "Start" twice.

After all the balls have fired, click the control chart button  on the Main Panel toolbar.

Confirm that Xbar and range is selected as the chart type

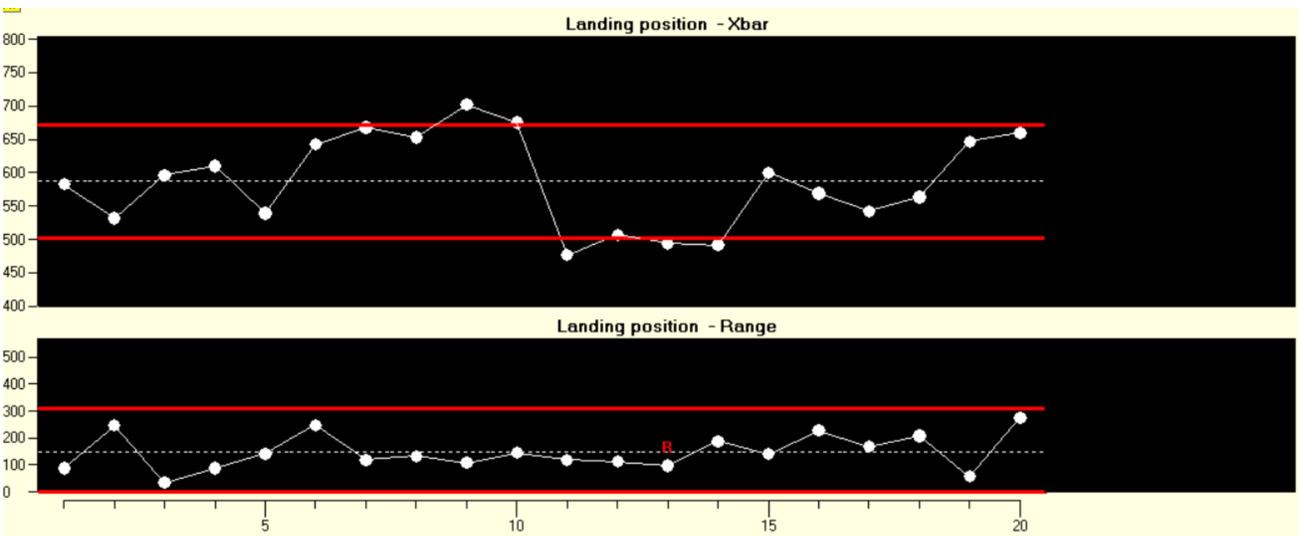
Click the OK button

Now calculate the limits. We have not moved the launcher so we can use all results in the calculations.

Press the control limits button  on the chart

Make sure that 1 is in the "From" box and the last subgroup is in the "To" box.

Click OK



You should now see that the process is indicating instability even though the data used to calculate the lines contains instability.

This ability of Shewhart control charts to detect special causes of variation, even when these special causes are present in the data used to calculate the control limits, is very important. Most commercial processes are not naturally in a state of statistical control.

The control limits are set at 3 times sigma from the average. Sigma is similar to standard deviation but uses calculations based on the average range of the subgroups (range is the maximum value minus the minimum value).

It is a mistake to calculate the control limits from individual results. See for yourself:

In the launcher simulation choose Average from the Show menu

Click the Calculate button

Click on Show 3 x S. Dev. lines on Run Chart

Click the Close & Apply button

Note the warning and click OK

Lines at 3 times standard deviation from the average will now be drawn on the Run Chart on the launcher simulation window. Notice that this chart does NOT detect the special variation whereas the Xbar chart does detect it (by 'detecting' we mean that there are points plotting outside the control limits). You can spool back the Run Chart to see previous shots by clicking the scroll bar to the right of the chart.

The reason that the Xbar chart detects special variation is because the control limits are calculated using an estimate of standard deviation based on the average subgroup range. Since the subgroups are small, this means that all the long-term changes are filtered out.

When using charts with real processes it is important to ensure that subgroups contain mostly common cause variation. Normally this can be done by measuring a small number of consecutive products for each subgroup, and having a time gap between the subgroups.

X or Individual Value chart:
Sometimes it is not possible to ensure that the subgroups will contain mostly common cause variation. For example, it may not be possible to ensure that all samples in the subgroup will be from the same batch. In this case, we will have to use an X & mr (individual value and moving range) chart.

In this type of chart we plot the individual measurements on one graph and the differences between the consecutive measurements on the other graph, this is called the Moving Range (sometimes only the individual points are shown, the moving range chart is omitted).

Here is an example to show how moving ranges are calculated.

Measurement	2.5	3.1	3.3	2.4	2.9	2.3	2.4
Moving range		0.6	0.2	0.9	0.5	0.6	0.1

The average moving range for the above data is 0.4833. The distance of the control limits from the average is calculated from the average moving range.

We will now run the launcher simulation with the same pattern of special variation (XBAR2.PVF):

Choose Existing results and Clear all from the Control menu on the launcher simulation.

Choose Sampling from the Control menu

Click the Send individual results option in the Subgrouping frame

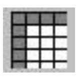
Click OK

Fire off 100 shots. Again, we will not worry about centring the process:

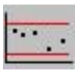
Select 50 in the shots frame.

Click "Start" twice.

Let's have a quick look at how the data is arranged in the data table

Click the data table button  on the main panel

Notice that the data is not arranged in subgroups, each individual result is in a row on its own. Now we will look at the data on an X & mr control chart.

Click the control chart button  on the Main Panel toolbar.

Choose X & mr as the chart type

Click the OK button

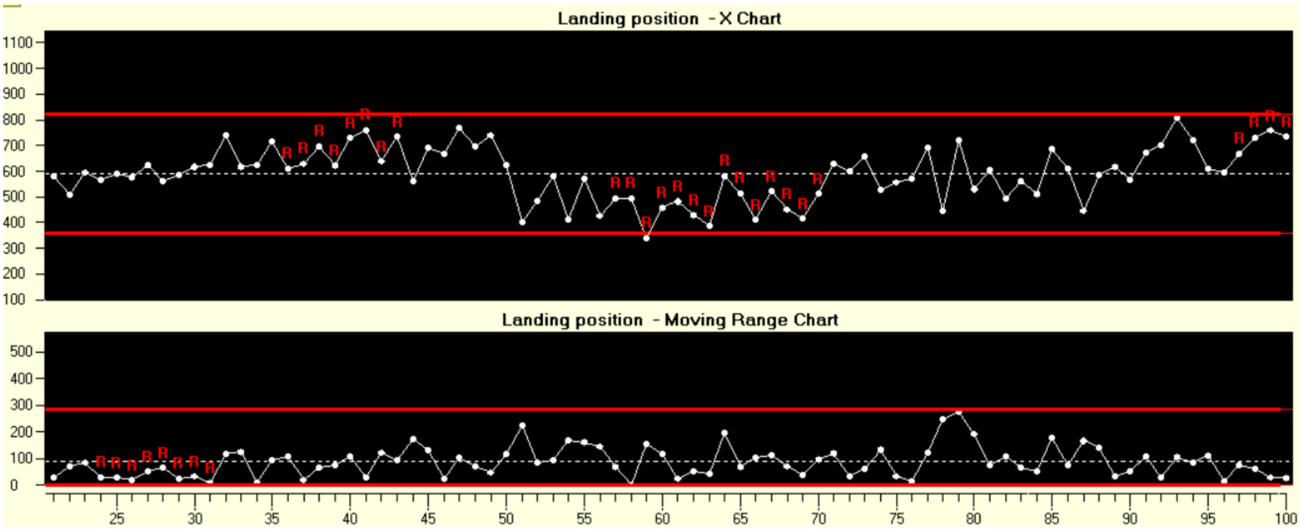
Press the + key two or three times to get all the values on the chart at once. Use the + key on the numberpad if you have one.

Now calculate the control limits.

Click the control limits button  on the chart

Make sure that 1 is in the "From" box and the last shot is in the "To" box.

Click OK



The chart should show some instability, either by having some points outside the control limits or because there are long runs in the data. A run is where a number of consecutive results are all above average or all below average.

Lets look at how control limits for individual value chart are calculated:

Click the View menu on the control chart

Select "Show control limit calculations"

Look at how the limits are calculated

If you would like to check the calculations, do so now.

Click the "Close" button

Lesson 4 summary:

1. Shewhart control charts will indicate instability even if instability is present in the data used to calculate the control limits.

2. We must use our knowledge of the process when deciding how to sample results and arrange them into subgroups. We should do this in a way which we know will reduce the chances of special cause variation occurring within subgroups.

3. If we do not know much about the process or we cannot be confident that little special cause variation will be present within subgroups, then we should use an X (individual values) chart or and X & mr (moving range) chart. With these charts, the control limits are based on the average difference between each individual result.

4. It is a mistake to calculate the control limits from the deviation of the individual results from the Average. The distance of the control limits from Average is always calculated from a short-term dispersion statistic (subgroup range or moving range).

End of Lesson 4

Lesson 5 - Binomial control charts

In lesson 2 we looked at Xbar and Range control charts. In lesson 4 the X (individual value) chart was introduced. In both these cases, we used variable or measurement data. This is data which comes from a continuous scale.

There is a different type of data called "attribute" data. Attribute data comes from discrete counts. For example:

- the number of blemishes on a surface,
- the number of faulty products
- the number of unpaid invoices

With attribute type data, in order to choose the correct type of control chart, we have to look at the way the data was generated. If we know in advance that the set of data will exhibit the characteristics of Binomial data or Poisson data then these types of charts should be used.

Binomial data:
Binomial data is where individual items are inspected and each item either possesses the attribute in question or it does not. Binomial means "two names" so if each item can be put down as either a pass or a fail then we can consider the data gathered to be Binomial data.

For example, consider the attribute 'blue' in samples of beads scooped from a box which contains beads of many colours. Each bead scooped is either blue or it is not blue - so if we create a stream of samples taken from the box and we count the number of blue beads in the samples, then we can assume that the resulting data will be Binomial type data.

Other examples of counts which would generate binomial data are:

- Late deliveries
- Non-conforming goods
- Out of specification components.

The random variation of Binomial data acts in a particular way, because of this we can calculate where to put the control limits. All we need to know is the average of the data set and the sample size.

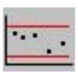
SPC Wizard can generate Binomial data using a simulation of scooping beads from a box - so let's do this now:

Click the Red Beads button on the Training & simulations panel.

Put the number 30 in the batch size box

Click the Start button

Now let's look at this data on a control chart.

Click the control chart button  on the Main Panel toolbar.

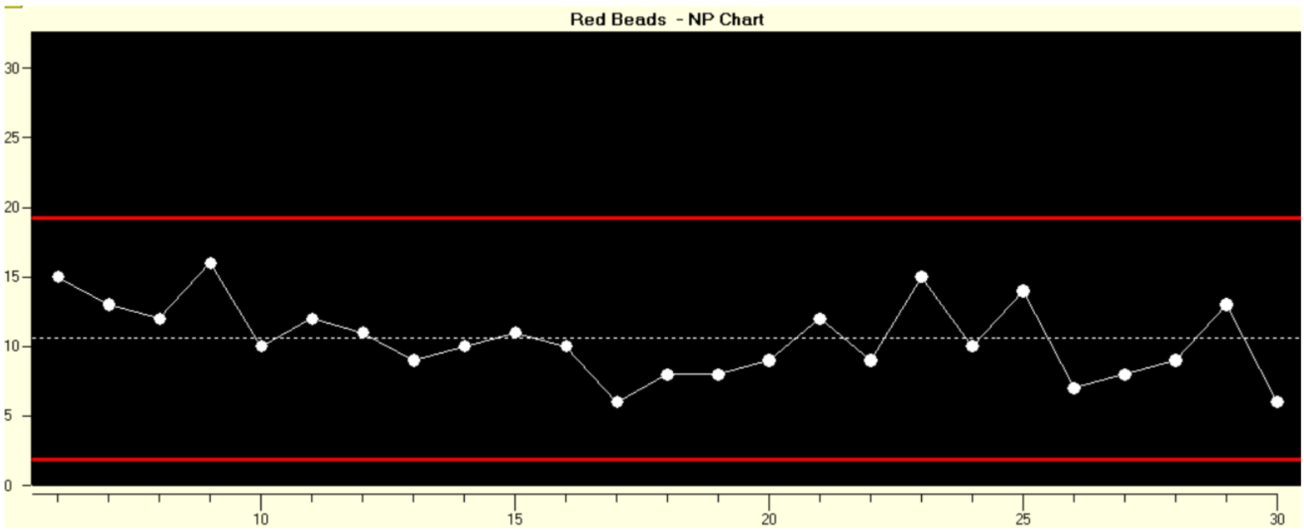
Confirm that Binomial is selected as the chart type

Click the OK button

Press the control limits button  on the chart

Put "1" in the "From" box and ensure that the last scoop number is in the "To" box.

Click OK

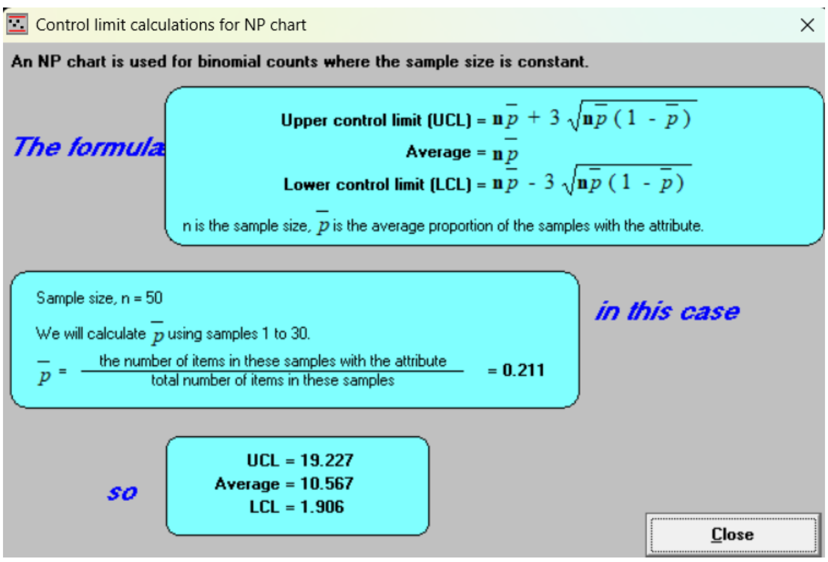


This type of chart is called an "np" chart. It is used when we know we have Binomial data and the sample size does not change. The points on the np chart are simply the number of items in the sample which have the attribute being counted (in this case we are counting beads with the attribute "red") Let's look briefly at how the control limits were calculated:

Click the View menu on the chart.

Select "Show control limit calculations" Look

at how the limits are calculated

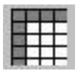


In these formulae "n" is the sample size (in this case 50, the size of the paddle), and "p bar" is the average proportion of the samples which have the attribute being counted.

Click the Close button.

You could try the calculations yourself:

OPTIONAL:

Click the data table button  on the Main Panel toolbar.

Total up the number of red beads in all the samples and divide this by the total number beads scooped in all the samples. This is "p bar".

Select "Show control limit calculations" again from the Menu button on the control chart Use

the formulae to calculate the upper control limit (UCL) and lower control limit (LCL).

Compare these with the results SPC Wizard gets.

When you are finished, click the Close button on the show calculations panel.

Tip: If you like to use a spreadsheet for calculations you can copy the contents of the data table to the Windows clipboard and paste this into your spreadsheet. The copy command is under the Data menu on the Data Table.

Binomial data with different sample sizes:
If we have binomial data but the sample size is not constant, then we cannot use an np chart. We will now use the simulation to add new samples to the data we have already started, but we will change the sample size:

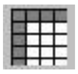
Click 100 in the Paddle size frame

Check that 30 is still in the Batch size box

Click the start button

Read the message which shows after the last scoop.

When the sample size is not constant for every scoop we have to convert counts to a rate or proportion. The resulting chart is called a "p" chart. We convert to a rate by dividing the attribute count by the sample size. Let's confirm that this is the case:

Click the data table button  on the Main Panel toolbar.

Ensure that Information box button  on the control chart toolbar is in the depressed position.

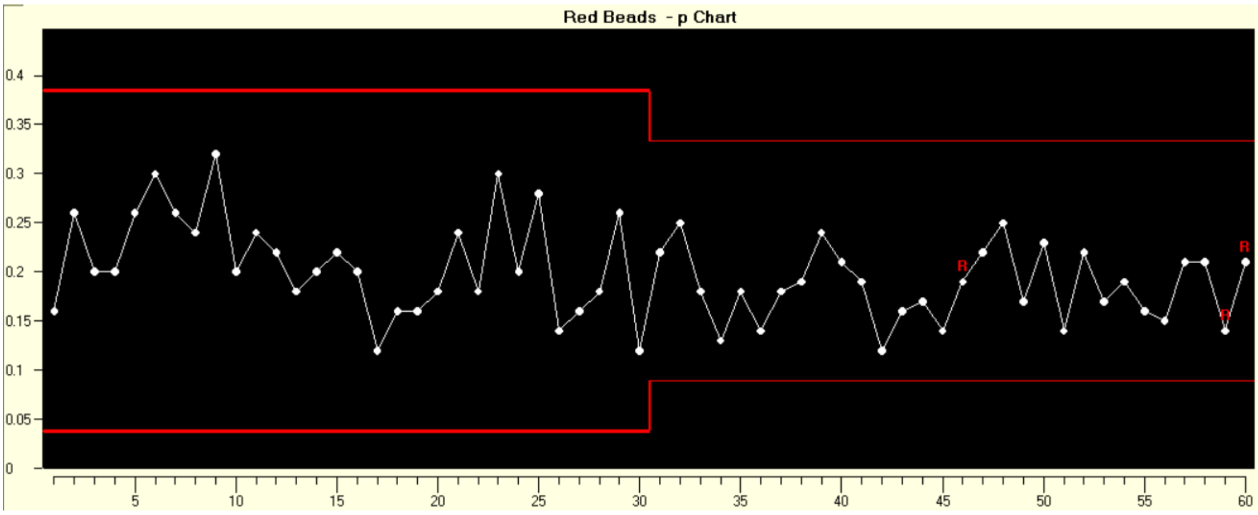
Click on any point on the control chart after sample no 30.

Look at the number next to "p" in the Information box

Look at the highlighted row on the Data Table. Get the number of red beads and the sample size for that point.

Confirm that the point plotted on the control chart and the proportion in the information box is the number of red beads divided by the sample size.

Click on a point before sample number 30 and repeat the above check.



You will notice that there is a step in the control limit lines at the point where the sample size changed. Before we look at the mathematics of the control limits, let's try to understand why there is a step.

The purpose of the control limits is to show the maximum and minimum values that we can put down to random common cause variation. Any points outside the limits indicate that something else has probably occurred to cause the result to be further from the average.

As we have said before, the random common cause variation of Binomial data acts in a particular way. The variation with large sample sizes is smaller than the variation with small sample sizes. We can use the simulation to demonstrate this.

Click the ". . ." button in the Paddle Size frame and select 5

Make sure that 30 is in the Batch size frame and click the Start button

Look at the results in the Data Table and keep in mind that the proportion of red beads in the box has not changed. In this exercise there are always 800 red beads in a box of 4000 (20%).

When the sample size is 5, a lot of times the number of red beads scooped is 1 (20% of the sample size), but it is not unusual to get 0 or occasionally 3 (60% of the sample size).

Now let's use a very big sample size:

Click the ". . ." button in the Paddle Size frame and select "150"

Make sure that 30 is in the Batch size frame and click the Start button

Look again at the results in the Data Table. Remember that 20% of the beads in the box are red and 20% of the of the sample size is now 30.

As you would expect most of the results are near 30, but even the most extreme results are nowhere near 0% or 60% of the sample size (60% of the sample size would be 90).

Now let's see how the control chart handles these extreme sample sizes.

Make the chart the active window by clicking on it with the mouse.

Press the + key a couple of times to see more samples on the chart

Look at the way the points which correspond to the small sample size (samples 60 - 90) vary up and down, then compare this with the variation with the large sample size (after 90). Keep in mind that we are not looking at absolute numbers here, we are looking the proportion of the sample which is red.

If necessary, use the scrollbar to look at different sections of the chart.

Look at the position of the control limits for the small sample size and the large sample size.

This illustrates one of the basic points about using control charts for attributes. Small sample sizes produce control charts which are not sensitive because there is so much random common cause variation in small sample sizes. Large sample sizes produce more sensitive control charts.

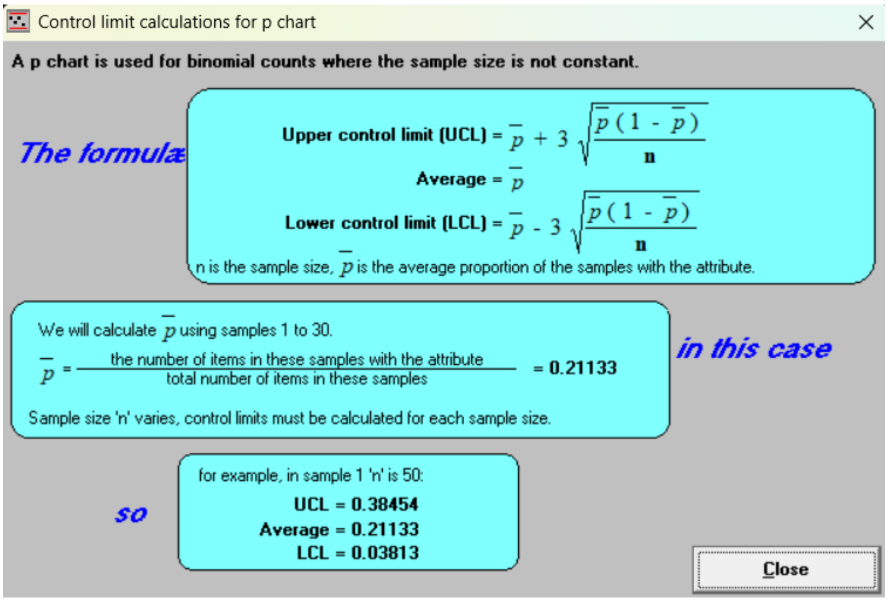
What this means is that if a process has a special cause of variation acting on it from time to time, it may not produce any points outside the control limits if the sample size is small. The same special cause of variation is more likely to produce points outside the control limits if we use a large sample size.

Let's have a quick look at the mathematics for the control limits:

Click the View menu on the control chart

Select "Show control limit calculations"

Look at how the limits are calculated



Notice that the limits have to be separately calculated for each sample size. If you want to try the calculations for yourself do so now. The example given is for sample number 1 but you can check your calculations for other samples using the Information box, just click on any point on the control chart and the control limits will be shown on the Information box.

Click the Close button on the Show Calculations panel

Criteria for binomial data:
We can only use an np chart or a p chart if we know in advance that the data produced will be binomial data. The full conditions which have to be satisfied before we can consider a set of data to be Binomial are:

1. The count must arise from a known number of discrete products (goods or services).
2. Each product inspected must either have, or not have, the attribute which we are counting.
3. The products inspected must not influence each another. If one item has the attribute, this fact must not change the likelihood of its neighbors having the attribute.

Lesson 5 summary:

1. Data from process can be divided into two major categories, variables and attributes.
2. Binomial data is attribute data where individual items are inspected and each item either possesses the attribute in question or it does not.
3. An "np" chart is used for Binomial data if the sample size is constant.
4. A "p" chart is used for Binomial data if the sample size is not constant.
5. Before using an "np" chart or a "p" chart we have to make sure that all the conditions for Binomial data are met.
6. With applying SPC to attribute counts, small sample sizes make it difficult to distinguish between common cause variation and special cause variation.

End of Lesson 5

Lesson 6 - Poisson and "X" charts using attribute data

In lesson 5 we created control charts for Binomial data.

Binomial data is where we look at products or services, and for each, decide it as a "pass" or a "fail". It is not always best to classify a whole product in this "all or nothing" way. For example, we might want to count the number of blemishes on a surface. We will never know the number of 'non-blemishes' on the surface so the data gathered is not Binomial data.

Criteria for Poisson data:
We can consider data to be Poisson type data if:

1. Discrete counts of an attribute can be made. e.g. tears in material, cracks on surfaces etc.
2. The counts arise from a known area of opportunity.
3. As with Binomial data, the attributes must arise independently of one another. In other words, there must be no mechanism which makes the attribute normally occur in clusters.
4. There are relatively few incidents of the attribute appearing compared with what might happen in the worst possible circumstances.

Another way of looking at this is that Binomial counts represent DEFECTIVES whereas Poisson counts represents DEFECTS.

With Poisson data, we use a "c" chart if the sample size is constant, and a "u" chart if it is not. With Poisson type data, the sample size is sometimes called the "area of opportunity".

Let's look at a file with Poisson type data:

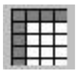
Click the Database Explorer button  on the Main Panel.

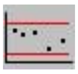
Click the Open Database .

button Select the file 'TutData.mdb'.

Select English in the Language list and Poisson 1 in the Data Table list.

Click the Open button.

If the data table is not open click the table button  on the Main Panel toolbar.

Click the control chart button  on the Main Panel toolbar.

Confirm that Poisson is selected as the chart type

Click the OK button

Press the control limits button  on the chart

Confirm that "1" is in the "From" box and the last sample number is in the "To" box. Alter them if necessary.

Click the OK button

A "c" chart is very similar to an "np" chart, the points plotted are simply the numbers in the data column. The only difference is the way the control limits are calculated:

Click the View menu on the chart

Select "Show control limit calculations"

Look at how the limits are calculated

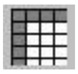
The control limits for a "c" chart are calculated from the average attribute count for all the samples. Notice that the sample size is not used anywhere in these calculations. If you would like to check the calculations for the control limits on the "c" chart, do so now.

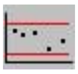
Click the "Close" button

Poisson data with different "areas of opportunity":
Now let's look at a chart for Poisson type data where the sample size or area of opportunity is not constant.

Click the Database Explorer button  on the Main Panel Select Poisson 2 in the Data Table list.

Click the Open button.

If the data table is not already open click the table button  on the Main Panel toolbar.

Click the control chart button  on the Main Panel toolbar.

Confirm that Poisson is selected as the chart type

Click the OK button

Press the control limits button  on the chart

Confirm that "1" is in the "From" box and the last sample number is in the "To" box. Alter them if necessary.

Click the OK button

Because the "area of opportunity" is not the same for all samples, we need to convert each attribute count into a rate before plotting the points on the chart. The resulting chart is called a "u" chart. The rate is simply the attribute count divided by the sample size or area of opportunity for the sample.

Click the View menu on the chart

Select "Show control limit calculations"

Look at how the limits are calculated

Notice that the control limits are tighter for larger areas of opportunity. This is for the same reasons that the control limits vary for different sample sizes in "p" charts.

If you would like to check the calculations for the control limits on the "u" chart, do so now. Use the Information box to see the control limits are for any point on the chart.

Click the "Close" button

The X (individual value) control chart with attribute data:
If a Binomial or Poisson chart is not appropriate, then we can use an X or 'individual' chart.

Let's compare a binomial chart with an X chart using the same data. First we will generate some data:

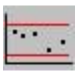
Click the training / simulations button  on the Main Panel toolbar.

Click the Red Beads button.

In the Paddle Size frame, click the 100 button

Put 30 in the Batch size box Click the Start button.

Now we will create a Binomial chart and an X (individual values) chart from same data.

Click the control chart button  on the Main Panel toolbar.

Confirm that Binomial is selected as the chart type.

Click the OK button

Choose Add Chart from the Chart menu on the control chart window.

Select 'X (individual) as the chart type

Click OK

Press the control limits button on the chart

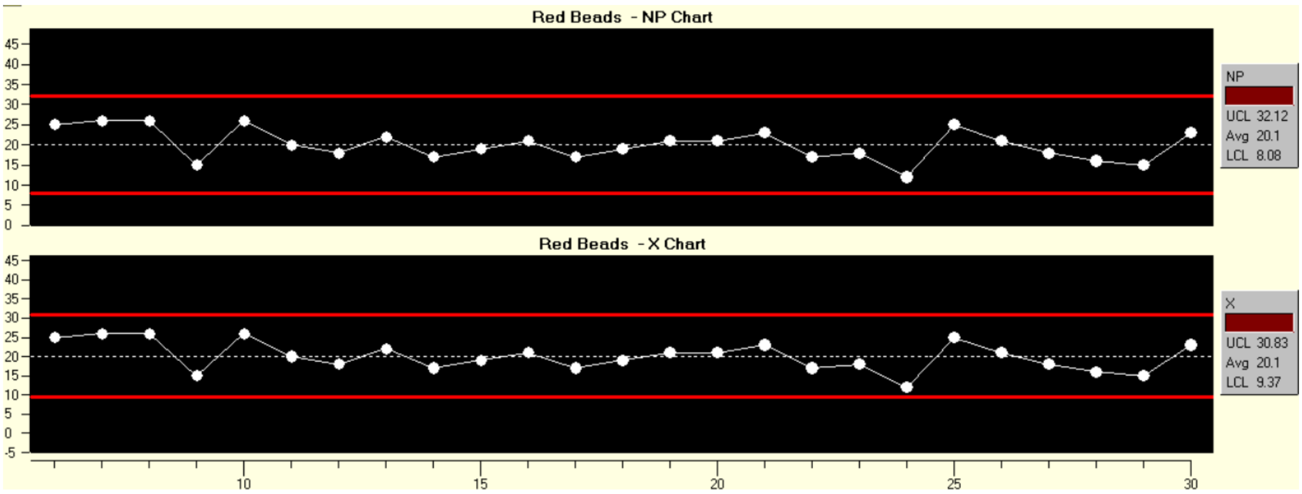
Enter 1 in the "From" box and ensure that the last scoop number is in the "To" box.

Click OK

Usually when you do this exercise, the control limits for the two charts are similar. Let's find the exact values:

Make sure that the Information box button on the control chart toolbar is in the depressed position.

Look at the upper control limit (UCL) and lower control limits (LCL) for each chart.



If we cannot be confident that the data we have fulfils the conditions to be binomial or Poisson data, then we can usually rely on an X chart to do a pretty good job. However there are limitations: Click 20 in the Paddle size frame on the simulation window Click the start button.

When the warning message appears click the OK button.

We now have a non constant sample size. The warning message informs you that. SPC Wizard automatically changes Binomial and Poisson charts to rate type charts when the sample size is no longer constant..

Sometimes X charts should be rate charts when the sample size is not constant and sometimes they should not - it depends on what the measurement represents. In our case the number of red beads scooped is definitely dependent on the sample size so we should look at an X chart based on rates.

Choose Add Chart from the Chart menu.

Select X (individual) as the chart type.

Click the 'Rate' box so that it has a tick in it.

Click OK

Press the control limits button on the chart.

Select 'Red beads - X Chart (rate)' in the Chart drop down list Enter

1 in the "From" box and 30 in the "To" box.

Click OK

The p chart and the X rate chart are both showing proportions and the control limits have been calculated using scoops 1 - 30 (ignore the other X chart in the middle which is not based on rates)

Compare the two charts. You can press the '+' key to see more data points on the charts.

Look at the data and the control limits before and after the change of sample size (the change was at scoop number 30).

Because we have not changed the number of beads in the box, we are looking at the results of a stable process so in theory control charts should not show any points outside the control limits.

There is always more random common cause variation with small sample sizes and you can see that the points on both charts jump up and down more after we change to a smaller sample size.

Because the control limits on a binomial chart are based on a theoretical knowledge of the way binomial data behave, the control limits change to accommodate the different sample sizes.

On X charts, the control limits are based on the variation between successive points in the data stream. When this variation changes due to altering the sample size, this can be misinterpreted as a process change.

Check Advanced Topics in Statistical Process Control by Donald Wheeler for a more background information.

X charts with low average:
When the average count is very small, another problem prevents us from using X charts. With attribute counts, the data can only take integer values such as 6, 12, 8 etc. Values such as 1.45 cannot occur. The discreteness of the values is not a problem when the average is large, but when the average is small (less than 1) then the only values which are likely to appear are 0, 1, 2 and occasionally 3.

The whole idea of control charts is that we want to gain insight into the physical variations which are happening in a process by looking at the variation of some measurement at the output of the process. When the measurements are constrained to a few discrete values then the results are not likely to reflect subtle physical changes within the process. For this reason X charts should not be used for attribute counts when the average count is low.

Lesson 9 gives more information about using attribute control charts when the average count is low.

Lesson 6 summary:

1. Poisson data is where we are counting defects (whereas binomial data is where we count "defectives")
2. With Poisson data, we use a "c" chart if the sample size is constant.
3. With Poisson data, we use a "u" chart if the sample size is not constant.
4. With Poisson type data, the sample size is sometimes called the "area of opportunity".
5. Before using an "c" chart or a "u" chart we have to make sure that all the conditions for Poisson data are met.
6. If we cannot be sure that the data will meet all the conditions to be Binomial or Poisson data, then we may be able to use an X chart, but the average count must be greater than 1.

End of Lesson 6

Lesson 7 - Pareto chart

A Pareto chart helps us to identify priorities for tackling problems. The Pareto principle (named after a 19th century Italian economist) states that 80% of defects or problems usually arise from about 20% of the causes.

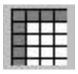
Let's look at data from an imaginary process.

Click the Database Explorer button  on the Main Panel

Click the Open Database . button Select the file 'TutData.mdb'.

Select English in the Language list and Pareto in the Data Table list.

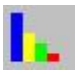
Click the Open button.

If the data table is not already open click the table button  on the Main Panel toolbar.

Sample	Surface tear	Rollock warp	Impure lid	Scrim pitted	Shaft smear	Shaft cracked	Gear leak	Smidgers	Scrim missing	Bar loose
3	1	3	2	10	1	2	1	18	2	0
4	2	0	0	6	0	0	1	21	1	0
5	1	1	2	3	0	1	0	25	1	0
6	6	1	0	8	0	3	2	35	1	0
7	2	2	1	5	1	1	0	27	0	1
8	1	3	2	10	1	2	1	18	2	0
9	2	0	0	6	0	0	1	21	1	0
10	1	1	2	3	0	1	0	25	1	0
11	6	1	0	8	0	3	2	35	1	0
12	2	2	1	5	1	1	0	27	0	1
13	1	3	2	10	1	2	1	18	2	0
14	2	0	0	6	0	0	1	21	1	0
15	1	1	2	3	0	1	0	25	1	0
16	6	1	0	8	0	3	2	35	1	0
17	2	2	1	5	1	1	0	27	0	1
18	1	3	2	10	1	2	1	18	2	0
19	2	0	0	6	0	0	1	21	1	0
20	1	1	2	3	0	1	0	25	1	0
21	6	1	0	8	0	3	2	35	1	0
22	2	2	1	5	1	1	0	27	0	1
23	1	3	2	10	1	2	1	18	2	0
24	2	0	0	6	0	0	1	21	1	0
25	1	1	2	3	0	1	0	25	1	0

The columns in the table represent 10 types of non-conformity or imperfection which can occur in Assembly M412. Each of the 25 rows contains the results of one inspection.

Let's look at this data using a Pareto chart:

Click the Pareto chart button  on the Main Panel toolbar.

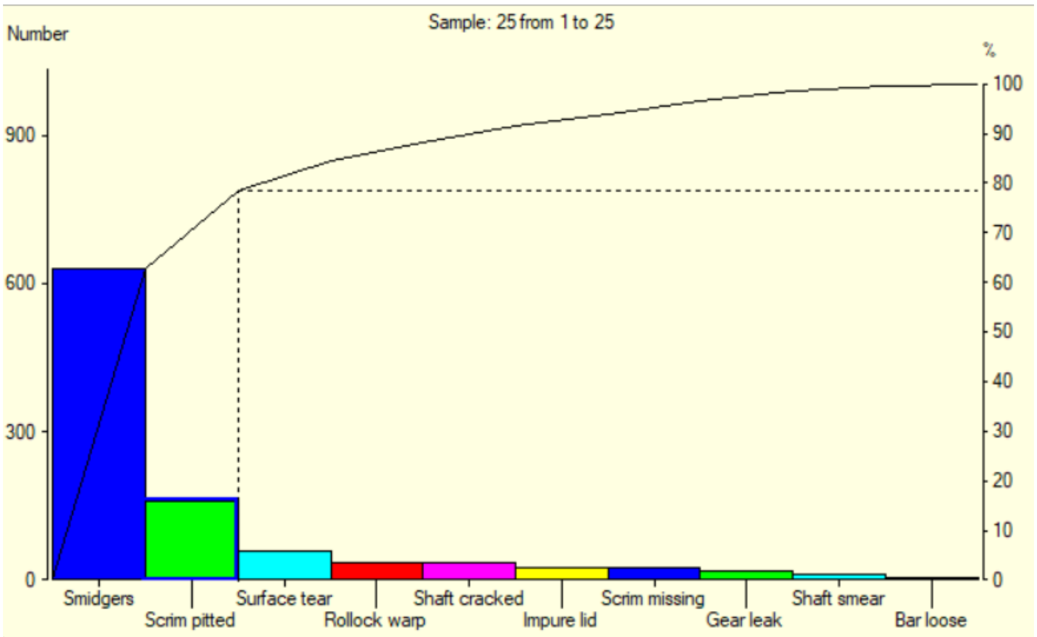
In a Pareto chart, the categories of data are shown as columns and the height of each column represents the total from all the samples.

The order of the columns is arranged so that the largest is shown on the left, the second largest next and so on. Since these counts usually represent defects or non-conformities, the biggest problems are therefore the categories on the left of the chart.

Our Pareto chart makes it immediately obvious that the most frequent problem is Smidgers appearing on the assembly.

It is common practice on Pareto charts to superimpose a cumulative percentage curve:

Click the "%" button



At each point on this curve you can see the percentage of the overall number of non-conformities or imperfections which are caused by the categories to the left of the point. This is best illustrated by example:

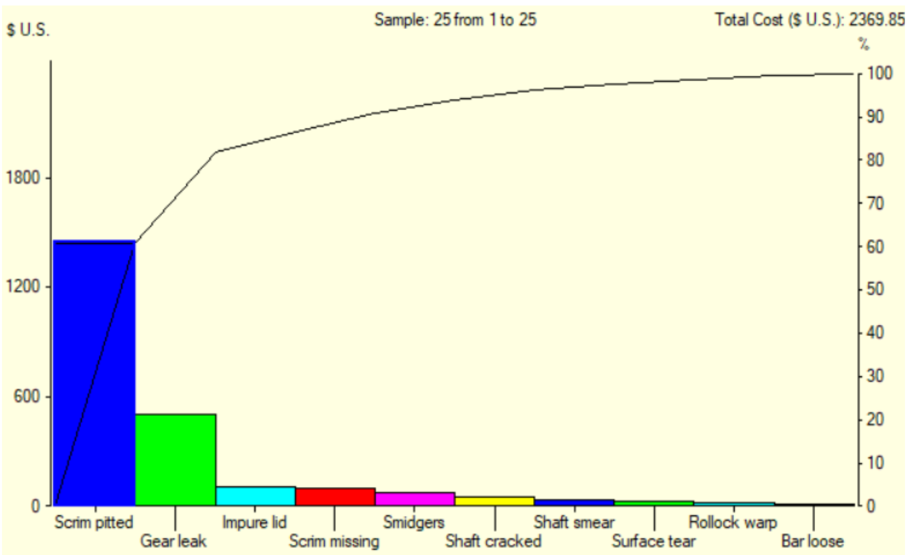
Mouse click on the second column from the left on the Pareto chart

By eye, follow the dotted lines from the column to the % scale on the right.

The dotted line points to just under 80%. This means the first two columns account for nearly 80% of all the non-conformities or imperfections that occur in Assembly M412. If we concentrate our efforts on reducing the number of smidgers and pitted scrims, then even if we are only partially successful, we are likely to make a substantial difference to the number of assembles which we have to send for rework.

Of course, all types of problems do not have an equal impact in terms of cost or importance. So if we know the cost of putting right each type of problem, then it is better to draw the Pareto chart with the column heights representing the total cost.

Click the Costs button $.



A "unit cost" value has been entered for each column. We are now looking at the total cost associated with each column (total number multiplied by unit cost). You will see that the left vertical scale is now labelled " US".

Click the Costs button again so that the button is not depressed.

The chart returns to showing numbers again.

Click the Cost button a few more times and look at the names under the first few columns as you click the button.

When the chart is showing costs, we get a different picture from the picture we get when it is showing numbers Although smidgers are the most common imperfection in assembly M412, they are easy to remove - a quick wipe with a cloth is all that is needed. A pitted scrim, on the other hand, needs the assembly to be dismantled. A leaking gear housing is the big nightmare - but fortunately they are not very common.

Make sure the chart is showing costs (the Costs button should appear depressed). If it is showing numbers, click the Costs button again.

Mouse click on the highest column.

Now we see that pitted scrims account for about 60% of all rework costs, so this is the problem which is creating the highest cost to the company.

Mouse click on the "Gear leak" column

Although we do not get many gear leaks, they are actually the second biggest problem in terms of costs. Smidgers do not cost the company a lot of money despite the fact that they occur in large numbers.

So looking at Pareto chart for M412 costs, we should concentrate our efforts in eliminating pitted scrims and gear leaks. But we should also take into account the effort required to solve an issue. We could go for low hanging fruit if we see an item which would be very easy to solve.

A table must have columns for at least two different attributes before you can use the Pareto chart. You can use the Pareto chart with data generated by the Colored beads simulation.

Lesson 7 summary:

1. A Pareto chart helps us to determine priorities.
2. If the cost associated with one unit of each column is known then we can choose between displaying total numbers or total costs for each category.
3. The columns on the left have the highest totals.

End of Lesson 7

Lesson 8 - Scatter chart

When we want to reduce or eliminate a problem, we will need to come up with ideas or theories about what is causing the problem. One way to check if a theory should be taken seriously is to use a scatter chart.

To use a scatter chart, we first have to take a series of measurements of two things over a period of time. The two things that we would measure are the problem itself, and the thing that we think may be causing the problem. We then plot the measurements on a scatter chart. The scatter chart will help us to see whether there is a mathematical relationship between two sets of measurements.

We will look at how to use a scatter chart using an example:

Flaking plugs:
A company makes large cylindrical casings known as "plugs" for a chemical process.

Analysis using a Pareto chart showed that the problem of surface flaking of the plugs was costing the company a lot of money.

A process improvement team was set up to try to reduce the number of "flakers". The team quickly found that everyone had a different opinion of what was OK and what was a flaker. The first job, therefore, was to come up with a good definition of a flaker which everyone could use.

The process operators were shown how to use control charts and they started keeping a chart of the number of flakers produced in each batch. This chart showed that the process was unstable. so they knew that they had to look for special causes of variation. Mary, one of the process operators on the team, said she always feels cold on days that they have a lot of flakers.

The process operators started keeping records of the air temperature at the time the plugs were made. At one of the team meetings Jack pointed out that on at least two occasions when the number of flakers was outside the control limits, it was raining.

The team asked the lab for help to test the theory that rain was a factor. One of the engineers pointed out that it was actually raining that very day but there were very few flakers. Nevertheless he still suggested that it might be a good idea to measure the moisture content of the main ingredient. The resulting records are in file SCATTER.SPC

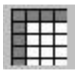
Click the Database Explorer button  on the Main Panel

Click the Open Database .

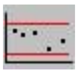
button Select the file 'TutData.mdb'.

Select English in the Language list and Scatter in the Data Table list.

Click the Open button.

If the data table is not already open click the table button  on the Main Panel toolbar.

First, let's have a look at the control chart. Because each plug is either a flaker or it is not a flaker, the chart we should use is a binomial chart.

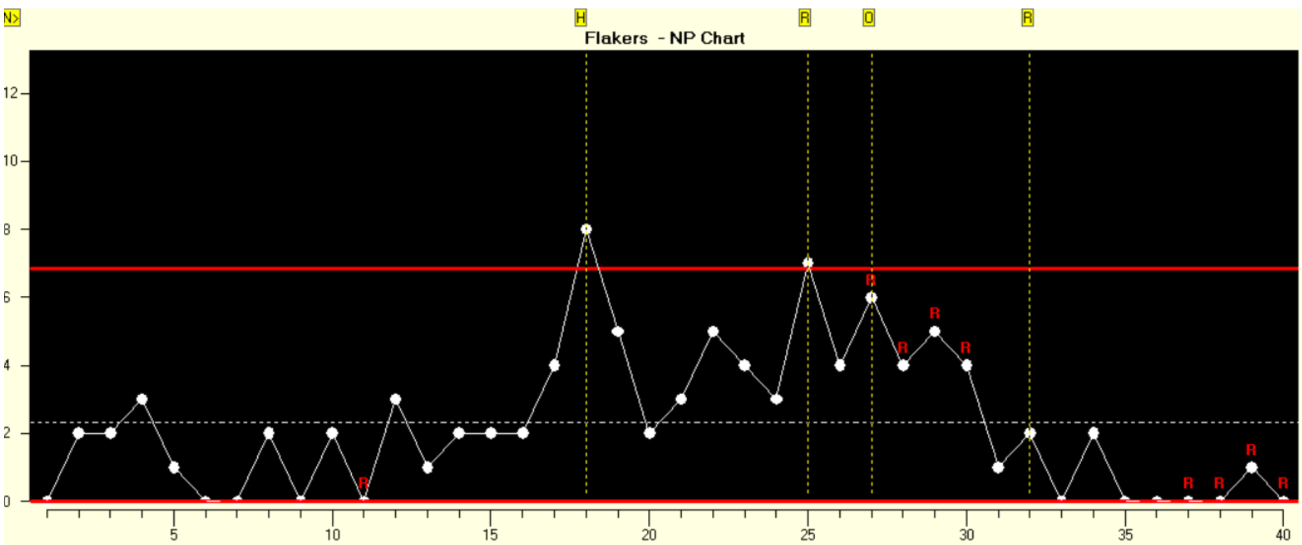
Click the control chart button  on the Main Panel toolbar.

Confirm that Flakers is selected as data column and binomial is the chart type.

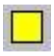
Click the OK button

Press the control limits button  on the chart

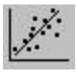
Enter "1" in the "From" box and ensure that 40 is in the "To" box.

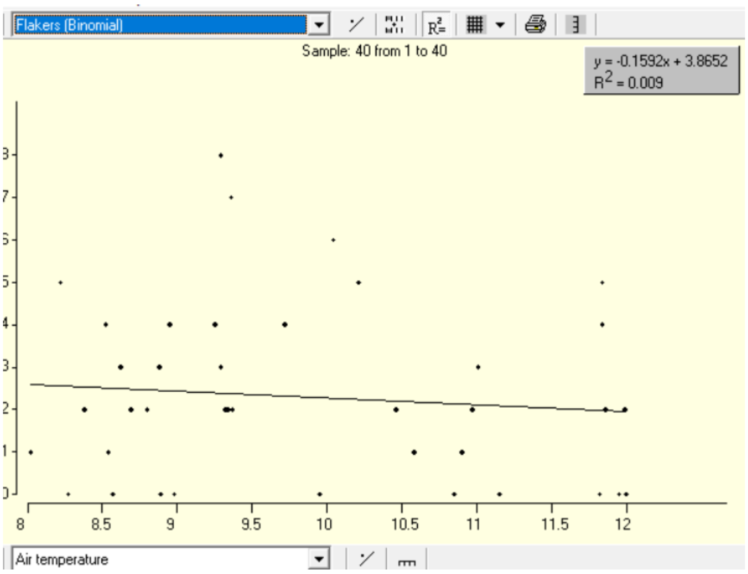


If you are wondering why there is no sample size column in the data table this is because a column is not necessary if sample size is constant throughout all the samples. SPC Wizard still needs to know the sample size for binomial data but this time it was entered once only for the whole file (in the Table Properties panel).

The data is out of control because some points are outside the control limits. There are also runs of 10 consecutive points above and below the average line - these also indicates instability. You can mouse click on the square at the top of the vertical dotted lines to see the notes that the operators have recorded (you may need to click the "Show notes" button  on the toolbar to see the notes).

Now let's look at a scatter chart.

Click the scatter chart button  on the Main Panel toolbar



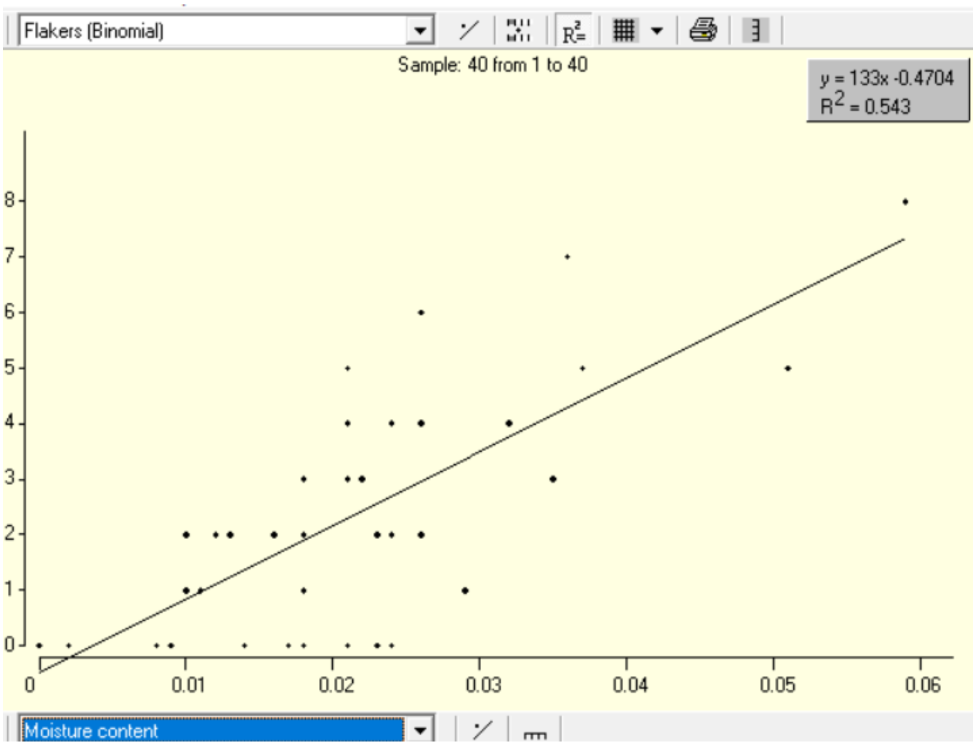
On this chart, the number of flakers is on the vertical axis and the air temperature is on the horizontal axis. For each row in the data table, a dot is put where the two values meet.

Mouse click on any sample number in the data table (click on a cell in the grey column on the left).

On the scatter chart a point is highlighted with a small blue circle. Confirm that the point is at the same height as the number of flakers for that sample number. Also, the distance along the x axis should show the air temperature.

In a scatter chart, if the measurements on the horizontal axis are not related in any way to the measurements on the vertical axis, then the dots will appear at random, with no pattern visible. If there is a mathematical relationship between them then the dots will tend to group into a fuzzy line or curve.

In this case there does not seem to be any pattern to the points on the scatter chart. We can conclude, therefore, that there is no correlation between air temperature and the number of flakers produced. This means that we can say that the air temperature is not a factor in producing flakers. Mouse click on the small triangle at the end of the box under the x axis Select "Moisture content" from the list by clicking on it with the mouse.



We now have a scatter chart with flakers on the vertical axis and moisture content on horizontal axis. There appears to be a correlation between the two sets of numbers because we can see the dots have formed into a fuzzy line. This chart is showing that flakers increase when the moisture content increases. This still does not prove that one causes the other. There could be a third factor which causes BOTH to change at the same time. Still, we seem to have a clue here.

Click the Equation button

What you see now is the "best fit" line through the points. The equation for this line is shown at the top right of the chart. The R-squared figure is a measure of how well the data fits the line. If R squared = 1 then all the points lie on the line. If R-squared is 0 or near 0 then there is no correlation between the data on the two axes so the line and the equation has no relevance.

Mouse click on the small down arrow at the end of the box under the x axis Select

"Air temperature" from the list by clicking on it with the mouse.

You can now see the best fit line through these points. The R-squared value is low showing that there is no correlation between the two sets of data.

Flaking plugs - a solution:
In our example it was found that the moisture content of the ingredient was a very large factor in causing the plugs to flake. The solution was simple and cheap - a better area was found to store the ingredient and a machine was purchased to reseal partly-used packages. The savings from these new procedures run into tens of thousands of pounds every year.

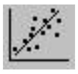
Scatter charts with changing sample size:
In the above example, the sample size was constant. When the sample size is variable we should convert attribute counts into proportions or rates before plotting the data on to a scatter chart. We do this by dividing counts by the sample size. Of course we would not divide measurements like temperature or moisture content by the sample size. Let's look at how the scatter chart handles rates.

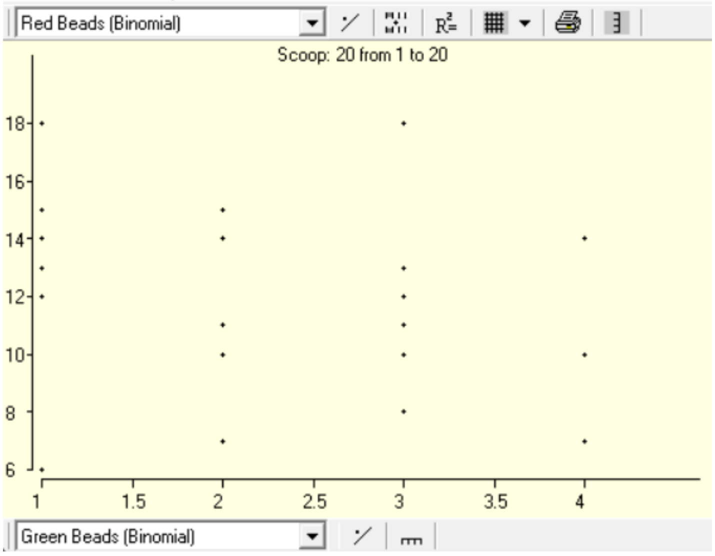
Click the Training & simulations button  on the Main Panel toolbar.

Click the Coloured Beads button

Put 20 in the Batch size box

Click the Start button

Click the scatter chart button  on the Main Panel toolbar



The chart is showing the number of Red beads on the vertical axis and the number of Green beads on the horizontal axis.

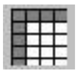
Click the 100 button in the Paddle size frame on the simulation

Click the Start button

Click the 20 button in the Paddle size frame

Click the Start button

We now have data with changing sample size. A "Rate" button  has appeared for each axis on the scatter chart. When this is depressed each data point on that axis is converted to a rate by dividing it by the sample size.

Click the data table button  on the Main Panel.

Return to the scatter chart and make sure that the both Rate buttons  are depressed.

Click on one of the points on the scatter chart with the mouse

Look at the highlighted row on the Data Table

Confirm that the red and green bead counts have been converted into rates (count / sample size).

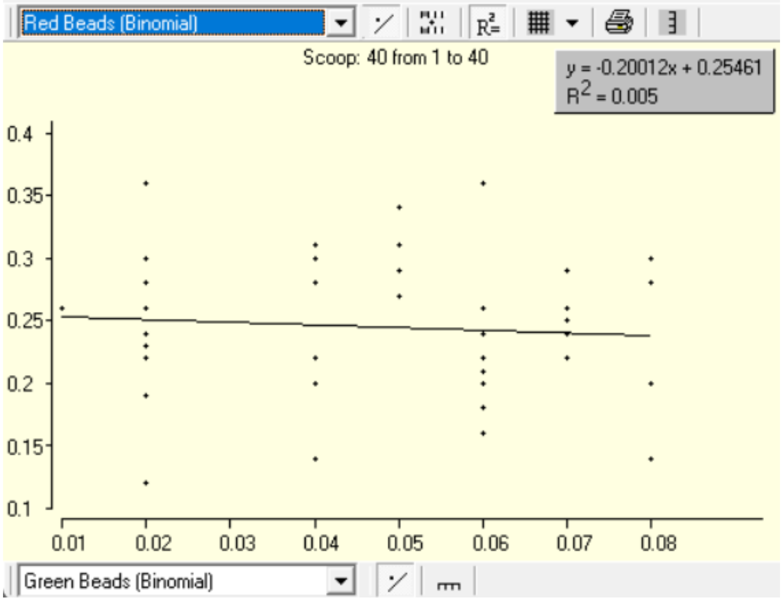
Click on both rate buttons to make them not depressed.

Look at the scatter chart. The dots on the chart do not seem to be at random. They seem to form into a fuzzy line. This means that there is a mathematical correlation between red beads and green beads. These beads are chosen at random so are you surprised by this correlation?

Check the Rate button for each axis Now

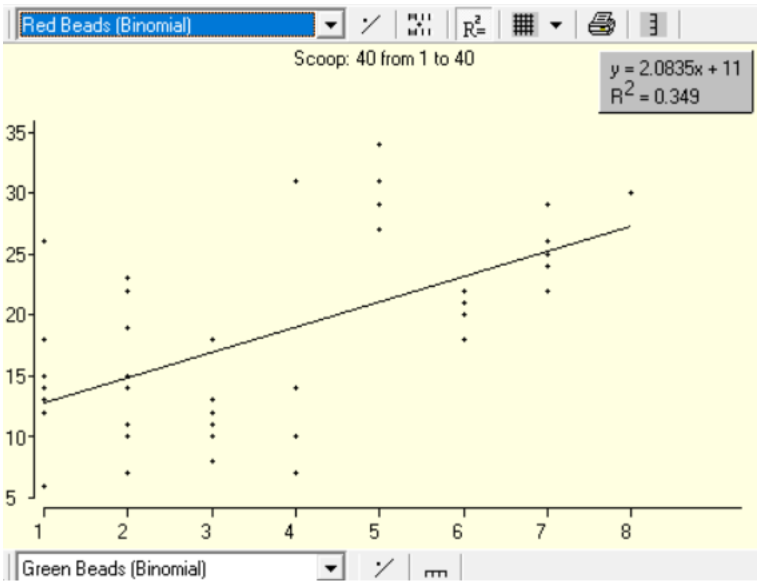
the dots seem to be at random..

Click the Equation button.



Notice that the R-squared value is near 0. This is what we would expect.

Uncheck the Rate checkbox again for each axis



The R-squared value is now much higher suggesting that there IS a correlation between the number of red beads and the number of green beads. Why is there a correlation? Does this mean that more red beads in a sample causes more green beads in the same sample?

If we have different subgroupsizes it is logical both nr of defects will increase and it seems to be a correlation but there actually isn't.

Lesson 8 summary:

1. A scatter chart helps us to see whether there is a mathematical relationship (correlation) between two things which we have measured. This may help us to find the causes of problems.

2. Even if we find a mathematical relationship this does not necessarily mean that one of them causes the other.

End of Lesson 8

Lesson 9 - Attribute control charts with low average

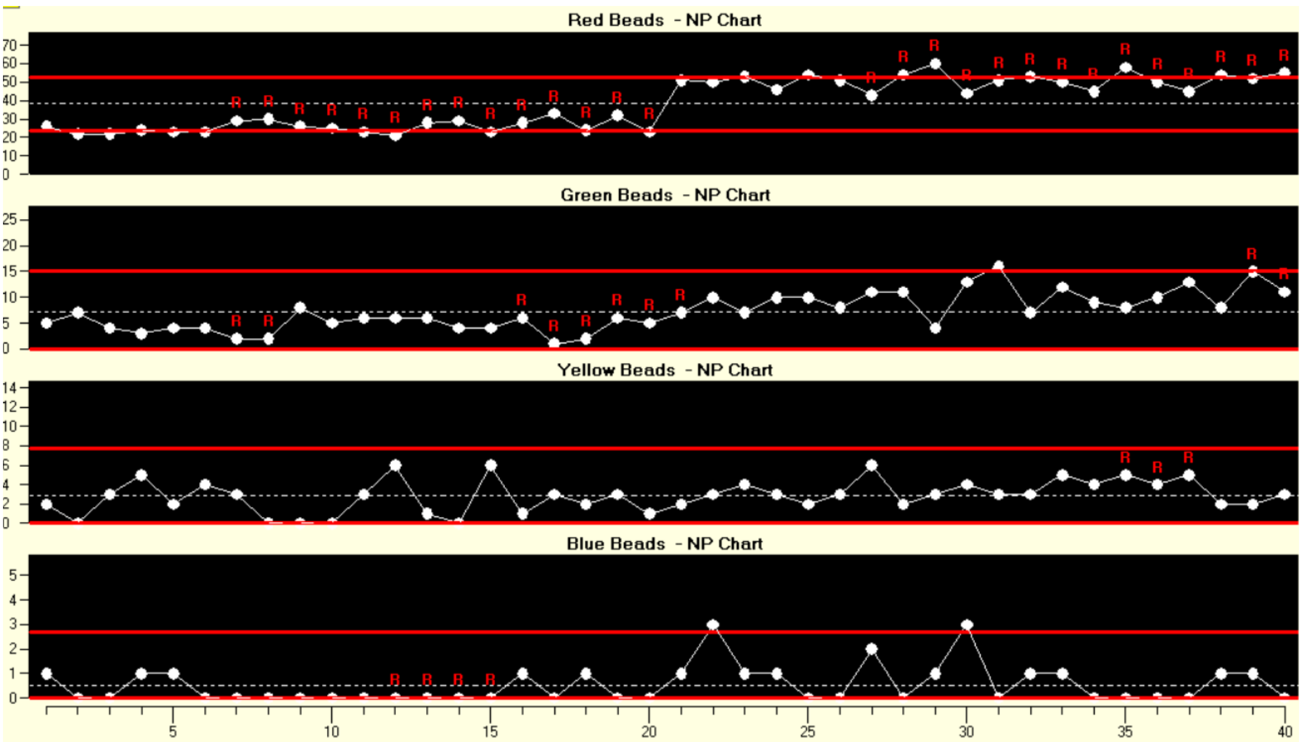
We are now going to look at a particular problem you can encounter with attribute control charts.

We will generate four streams of data from a process and create a special cause of variation in each data stream.

> Click the Coloured Beads button on the Training & simulations panel.
>
> In the Paddle Size frame, click the 100 button
>
> Put 20 in the Batch size box Click
>
> the Start button.
>
> When the 20 scoops have finished, Choose "Variation" from the 'Control' menu.
>
> In the "Beads in the box" frame, put double the current numbers in the box for each colour. i.e. Red 2000, Green 400, Yellow 160, Blue 20.
>
> Click OK to return to the simulation.
>
> Click the Start button again.
>
> Click "X" on the menu bar to hide the simulation (you can bring the simulation back at any time by clicking simulation button on the main panel

Now we will look at control charts for all four colours of beads.

> Click Multiple Control Charts from the Show menu on the Main Panel
>
> Click the control limits button  on the control chart window.
>
> Put 1 in the "From" box and ensure that 40 is in the "To" box then click OK.



For all four colors, the number of beads in the box doubles after shot 20, therefore we would expect to see a clear signal that the special type of variation has occurred.

The control limits are calculated using all the data so scoops 1-20 should give results below average and scoops 21 - 40 should give results above average.

Look at the chart for red beads.

The chart is showing points outside the control limits and there are long runs below and above average - so the special variation is clearly visible on this chart.

Now look at the charts for Green, Yellow and Blue beads.

The special cause of variation is not so obvious on these charts, especially on the chart for blue beads.

Now we will look at the position of the upper control limit for each chart:

Ensure that the Information box button  on the control chart window is in the depressed position.

Look at the average (Avg.) figure for the red beads chart.

Look at the Upper control limit (UCL) figure for red beads chart.

The upper control limit is probably about 1.4 times the average.

Look at the Average and Upper control limit values for the other three charts.

Work out approximately how many times greater the UCL is than the Average.

As the average figure gets lower, the upper control limit gets further from the average. In the Blue bead chart, the Upper control limit is many times greater than the average.

The only thing that is different between the four charts is the average number of beads scooped. This demonstration shows one of the inherent problems with control charts for attributes. If the average of the samples is low, then attribute control chart are not sensitive at detecting special variation.

Because we usually count problems or failures, this means that as we get more successful at removing problems, the charts become less good at separating special cause variation from common cause variation.

The best way to overcome this problem is - whenever possible - use a measurement from a continuous scale and plot this on a chart for variables (Xbar & range or X) rather than use count data with an attribute chart.

For example, you are trying to produce a product or service within a given specification of time, weight or length, then make control charts from the time, weight or length measurement. This will indicate special variation much better than an attribute chart showing the number of out-of-specification products.

If it is not possible to use a variable, then there are other possible solutions.

1. Use large sample sizes to make the average count as high as possible.
2. If you have already collected the data, you could combine a number of the original samples into a smaller number of large samples. This would increase the average count per sample, but there is a danger here. If many of the new large samples contained products from before and after process changes, then the special variation could be hidden. It would only be sensible to combine samples if the new larger samples still contain products made at approximately the same time.
3. We could measure the interval between occurrences of the attribute and plot this on an X chart. The interval could be the number of products which do not have the attribute, or the total volume of good product between each occurrence of bad product. We could also measure time intervals (months, days or minutes) between occurrences of the thing we are interested in.

SPC Wizard Rare Events Mode makes it easy to calculate intervals or combine samples.

Click the Database Explorer button  on the Main Panel

Click the Open Database .

button Select the file 'TutData.mdb'.

Select English in the Language list and 'Attributes - low counts' in the Data Table list.

Click the Open button.

This file contains special variation between scoop 200 and scoop 201. The change is identical to the change you produced in the previous example (i.e. the colored beads were doubled from the standard default numbers) the only difference is the number of scoops before and after the change. The file has been provided in case you have a slow computer. If you prefer you could produce a set of results yourself.

OPTIONAL:

Click the Coloured Beads button on the 'Training / simulations' panel (If the panel is not open, click the  button on the Main Panel toolbar).

In the Paddle Size frame, click the 100 button

Put 200 in the Batch size box Click

the Start button.

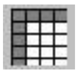
When the 200 scoops have finished, Click "Variation" on the menu bar.

In the "Beads in the box" frame, put double the current numbers in the box for each colour. i.e. Red 2000, Green 400, Yellow 160, Blue 20.

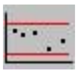
Click OK to return to the simulation.

Click the Start button again.

Whichever way you have produced the data, now open the data table:

If the data table is not already open click the table button  on the Main Panel toolbar.

Let's look at control charts for Yellow and Blue beads. This time we will create the control charts by drag and drop.

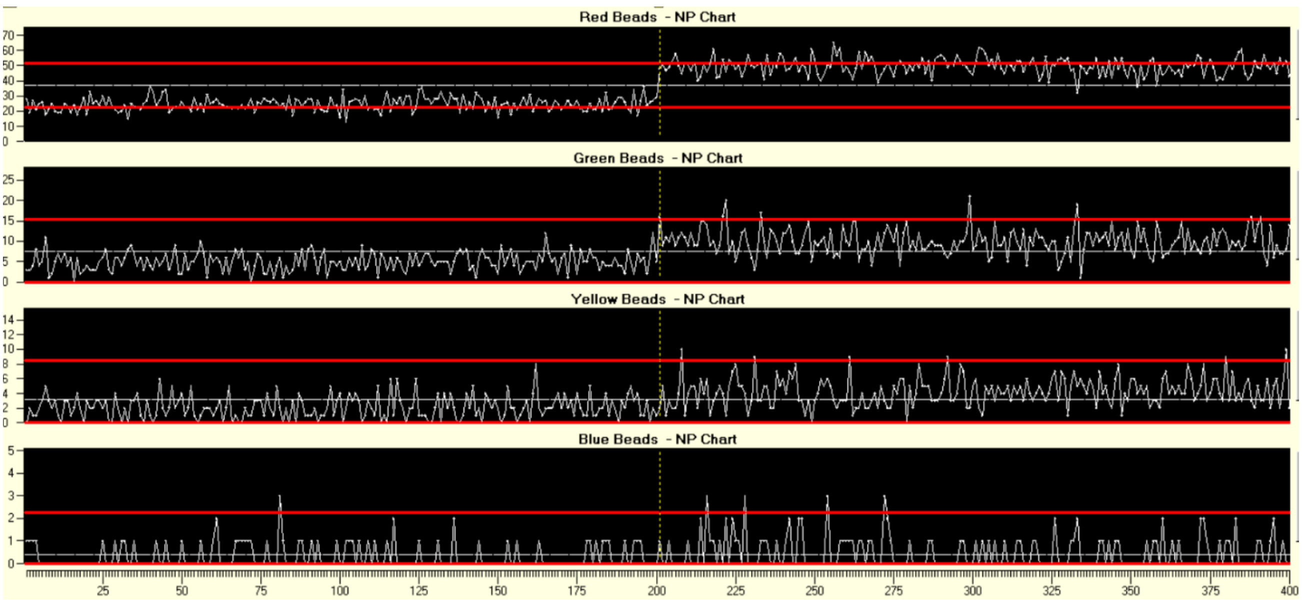
Use the mouse to drag the button on the table marked "Yellow Beads" and drop it on the Control chart button  on the main panel. (stop the mouse over the Yellow Beads button, hold down the mouse button and move the mouse with the button held down until it is over the control chart button then release the mouse button).

Use the mouse to drag the Blue Beads button. Drop it over the existing control chart.

Click the control limits button  button on the control chart window

Put 1 in the "From" box and ensure that 400 is in the "To" box then click OK

Use the scroll bars on both charts to look at some data points before and after scoop 201 (you can press the + key to see more points at a time).



These charts will look very like the charts for Yellow and Blue beads we produced earlier in the lesson except there are more data points. We know that the special variation occurred at scoop 201 but this is not very obvious, especially on the chart for blue beads.

Now we will look at the same data using combined samples.

Choose Special Modes then Rare Events Mode from the File menu on the Main Panel.

Click the "Combine Samples" option

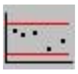
Confirm that the number 10 is in the box

Click "OK"

Click 'No to all' when prompted to save the control limits.

The data table now has only 40 rows, and each row, now contains the results from 10 of the original scoops.

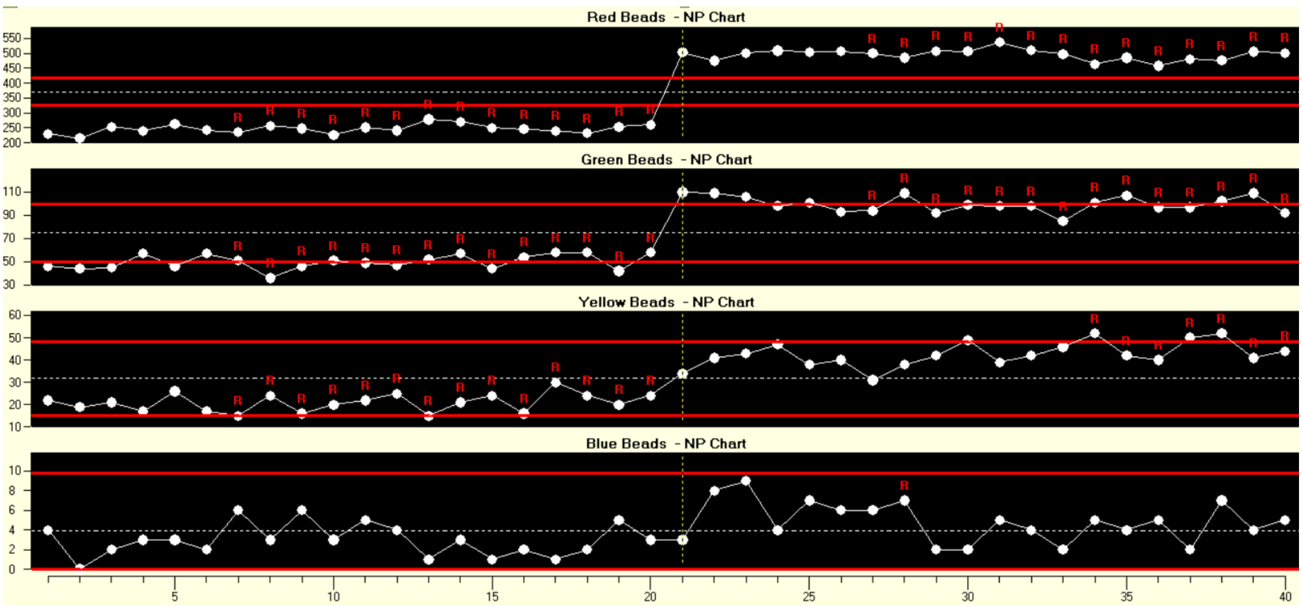
Let's look at control charts using this data.

Use the mouse to drag the button on the table marked "Yellow Beads" and drop it on the Control chart button  on the main panel

Use the mouse to drag the Blue Beads button. Drop it over the existing control chart.

Click the control limits button  button on the chart

Put 1 in the "From" box and ensure that 40 is in the "To" box then click OK



It is more obvious now that the number of Yellow beads is running below average up to sample 20 and above average from sample 21. So combining samples has worked for the Yellow beads.

But what about the Blue beads, we know that the number of Blue beads also doubled at what is now sample 21. Is this obvious from the chart?

Let's try another approach. Reload the original table. The easiest way to do this is to choose item 1 in the recent tables list.

Choose file 1. at the bottom of the 'File' menu on the Main Panel.

The original table file has been loaded.

Choose Special Modes then Rare Events Mode from the File menu.

Click the "Calculate Intervals" option

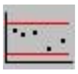
Choose Blue Beads in the "Calculate intervals between" list

Make sure that the "Other items" option is selected

Click OK

The data table now has a row for every blue bead which appeared in the original scoops. The 'Interval between . . .' column shows the number of non-blue beads which were scooped between each occurrence of a blue bead.

We cannot use a Binomial chart for this type of data, so we will look at it with an X chart.

Click the control chart button  on the Main Panel toolbar.

Confirm that "Intervals between Blue Beads" is selected and "X (individual)" is chosen as the chart type

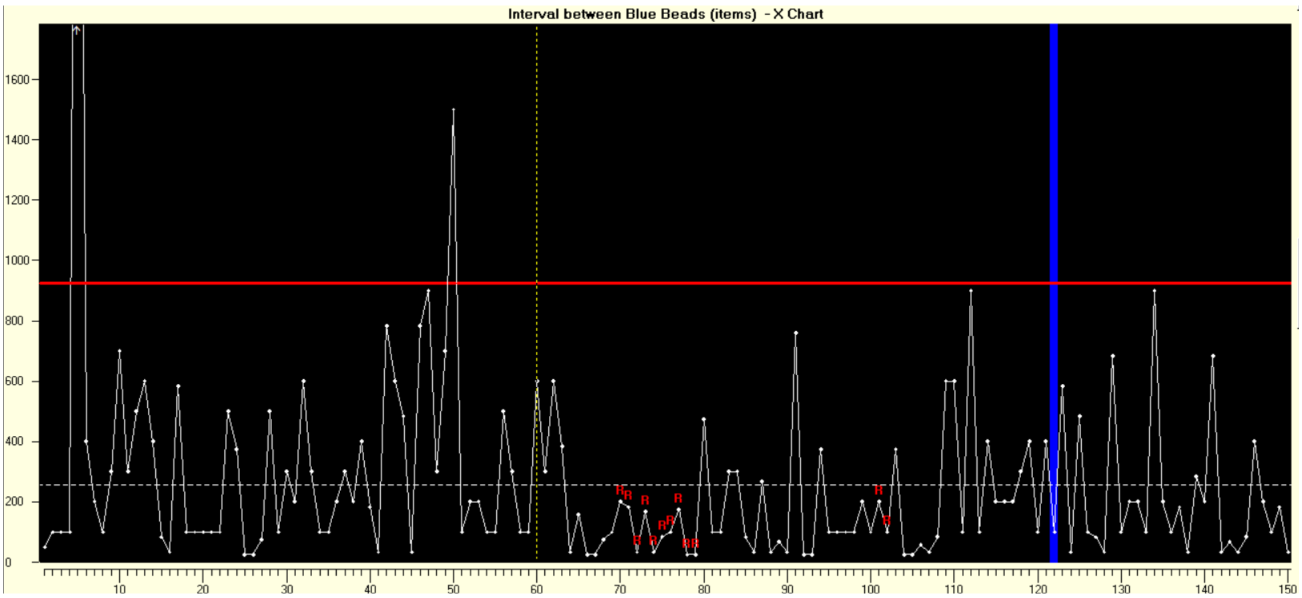
Click the OK button

Press the control limits button  on the chart

Put "1" the "From" box and ensure that the last interval number is in the "To" box.

Click OK

Press the + key on the keyboard once or twice (if you have a numberpad use that + key). This will allow you to see more points on the chart at once. If the points are too squashed up, press the - key.



We know that the number of Blue beads in the box increased at original scoop number 201. This means that the interval between blue beads should decrease at this point. You may or may not consider this to be apparent in the chart. In the example it is from subgroup 60.

One final thing to keep in mind. In these exercises we have been looking at control charts to illustrate how special variation becomes visible to us from the data, and we know in advance when a special cause of variation occurred. In a real life situation, we will not know in advance when or even **if** special variation is present in the data.

Lesson 9 summary:

1. Beware of any attribute control chart which has a very small average. Special causes of variation may be present in the process but not showing on the chart.
2. Always try to use control charts with variable measurements rather than simple pass / fail counts.

3. If it is not possible to use variable measurements, then, if you have enough data, you may try combining samples or calculating intervals to see if there is any indication of special variation.

End of Lesson 9

Lesson 10 - Process capability

When Specification or Tolerance Limits have been set for a product, it is important to know if our processes are capable of producing products which will be consistently within these limits.

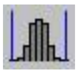
One of the simplest ways to get a picture of the capability of a process is to superimpose the Specification Limits on a histogram. Let's do this now.

Click the Database Explorer button  on the Main Panel

Click the Open Database . button Select the file 'TutData.mdb'.

Select English in the Language list and Stable in the Data Table list.

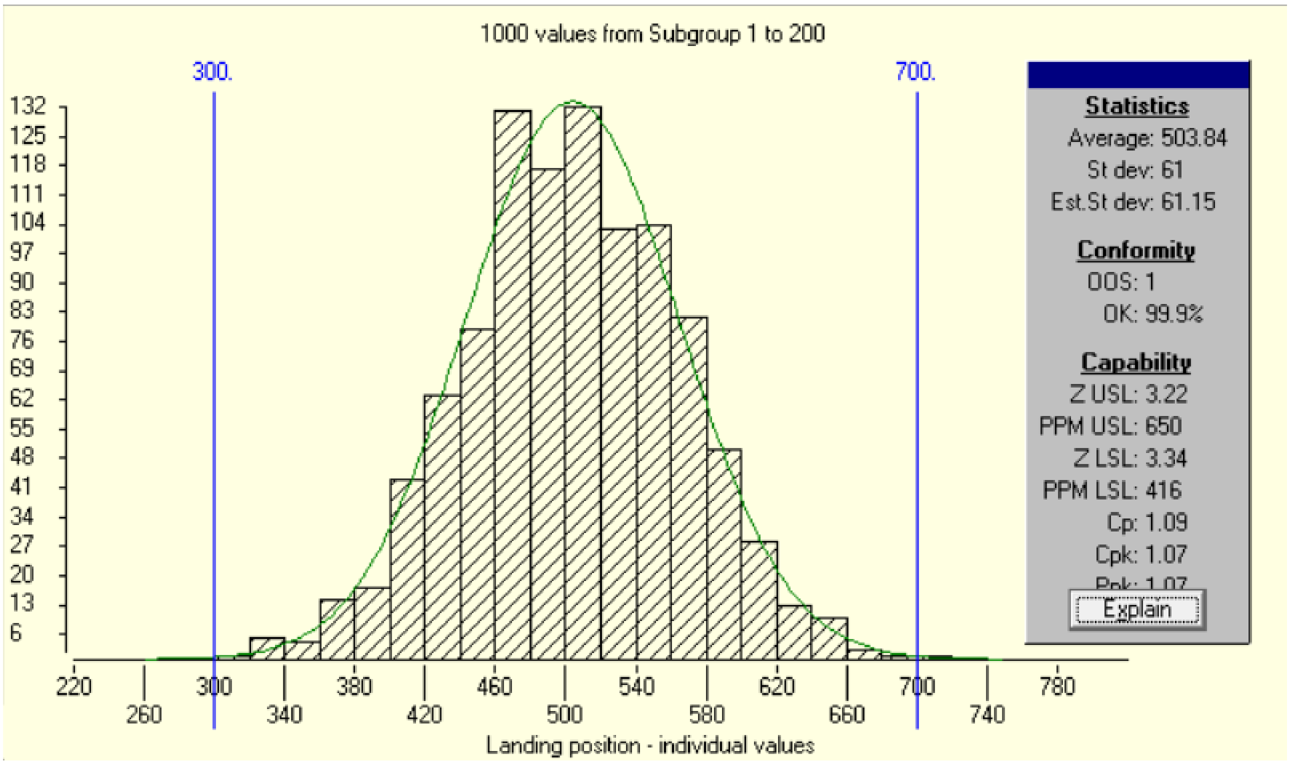
Click the Open button.

Click the histogram button  on the main panel.

Make sure that "Individual values" is selected under "Make histogram from"

Click OK

Ensure that the specification limits button  on the histogram toolbar is in the depressed position.



We can now get an instant picture of the spread of the measurements and compare this with the specification limits.

In many situations this will be all we really need but if we want to be more precise, we can use special indices which have been designed to put process capability into exact numbers.

Before we look at the capability indices, I want to explain about Estimated Standard Deviation.

In a previous exercise we looked at Standard Deviation. This is a figure for how much a set of measurements are spread around the average. The calculations are based on how far each result is from the average.

Now we will read about Estimated Standard Deviation. The text for this is available in the histogram information box.

Click the indices button  on the histogram toolbar

Click on the words "Est St dev" Read

the explanation carefully.

Let's see if the estimate is a good one.

Click the colours button  on the histogram toolbar.

Click on the small down arrow to the right of the button and select "Estimated std dev"

Again click on the small down arrow to the right of the button and this time select "Real std dev"

Although it may be difficult to see, the coloured bands are changing between true Standard Deviation and Estimated Standard Deviation each time you click. I think you will agree that in this case, Estimated Standard Deviation is very close to the real thing.

Make sure that we are displaying colored bands based on Estimated Standard Deviation.

Click on the small down arrow to the right of the colors button and select "Estimated std dev" Now

we will look at three of the indices.

Capability indices:
The purpose of all capability indices is to give a figure for how well the spread of the results fits within the specification limits. If the specification limits are well beyond the two tails of the histogram, the process will be "capable" because it should be easy to produce consistently within specification . If the specification limits are sitting on top of parts of the histogram, then we will produce a lot of products which are out of specification, so we would say that the process is not capable.

Cp:
Cp is one of the indices in common use. Read now about Cp from the histogram information box.

Click on "Cp" in the histogram information box Read the explanation carefully.

Let's look at how this index changes with different specification limits.

Use the mouse to slowly drag both specification limit lines outwards until they are well inside the red zones.

Look at the value for Cp.

It should be showing a figure greater than 1 to indicate that this process can easily produce goods within the new specification limits.

Now drag the specification lines until they are within the yellow zones.

Look at the value of Cp now.

With these specification limits, our process is clearly going to produce a lot of goods out of specification. This is reflected in the value of Cp - it is less than 1.

Drag the specification limits until they are at the boundary between the red and purple zones.

The limits are now at 3 times Estimated Standard Deviation from Average. The value of Cp should be about 1.
This marks the boundary between a process which is capable and one which is not capable of producing within specification.

There is, however, an important point to remember about Cp.

Hold down the "Control" key and move one of the limit lines in any direction. Because this is a big file, you will need to go slowly if you have a slow computer. The capability indices will recalculate frequently, watch them as you drag.

All of the capability indices change except Cp.

Stop dragging when you have moved the lines as far away from centre of the histogram as you can.

It is possible to produce 100% of goods out of specification and still have a Cp of one or less. This is because Cp only compares the distance between the specification limits with the spread of the results. This index does not take into account whether the process is properly "centred" between the specification limits.

Now return to the specification limits to their correct value.

Click on the histogram anywhere.

Cpk:

Click on "Cpk" in the information box.

Read page 1 of the explanation carefully.

Drag both specification limits to 3 times Estimated Standard Deviation from Average (the boundary between the red and the purple zones).

The figure for Cpk is very similar to Cp - approximately 1.

Drag one of the limits well into the red zone (do NOT hold down the Control key) Drag

the limit back towards the original position and carry on well into the purple zone.

Repeat this a few times and look at the figure for Cpk while you move the line.

The value of Cpk remains static when the limit you are dragging is further form the Average than the other limit. The value of Cpk changes when the limit you are dragging is closer to the average than the other limit. This is because the Cpk is calculated using only the specification limit which is closest to the average.

Because of this, Cpk is an index which takes into account how well the process is "centred".

Hold down the "Control" key and move one of the limit lines. You will need to go slowly if you have a slow computer. Compare Cp and Cpk as you drag.

Cp does not change when you do this but Cpk does. It is only possible to get a high value of Cpk if the process is reasonably "centred".

It is possible to have a negative value for Cpk

Drag the limits so that they are both on the same side of the average.

When the limits are both on the same side of the average, Cpk is negative.

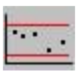
Ppk:

Drag the limit lines individually and compare Ppk with Cpk

Hold down the "Control" key and drag one of the lines. Again compare Ppk with Cpk as both lines move.

You will notice that the two values are very similar all of the time. We will read more about Ppk later.

Data from stable and unstable processes:
Histograms tell us a lot about a set of data but they does not tell us everything. Histograms do not show the time sequence of the measurements. The time sequence is shown in a control chart.

Click the control chart button  on the main panel toolbar.

Confirm that Xbar & range is selected as the data type and click OK

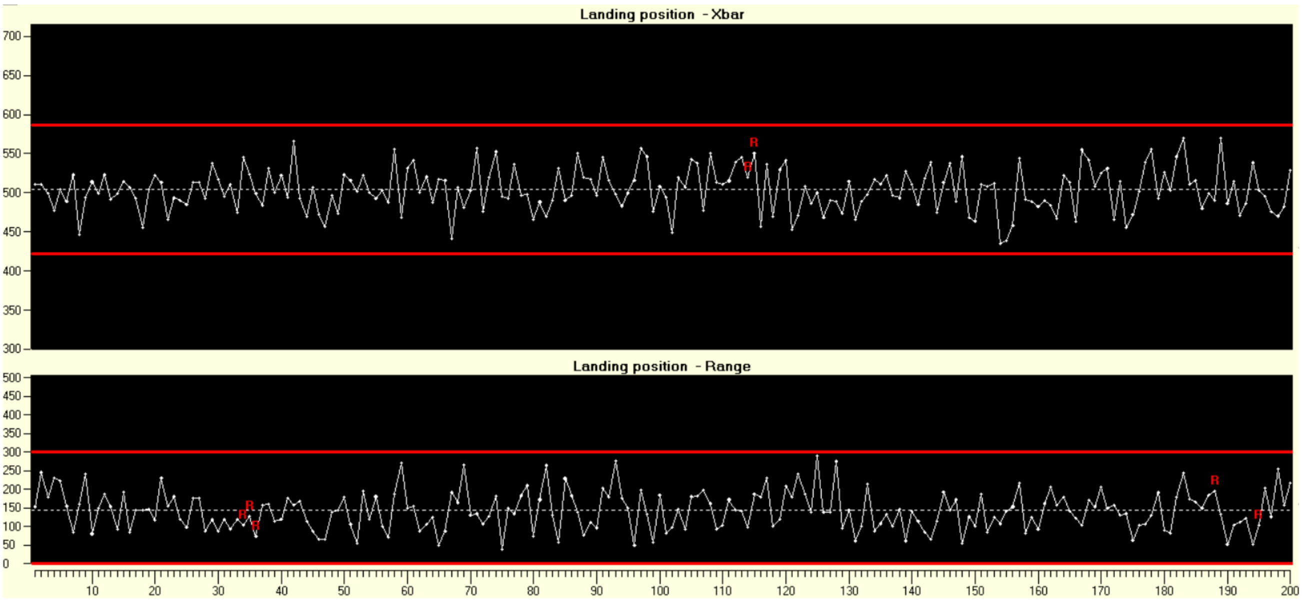
Press the control limits button  on the chart toolbar

Put "1" in the "From" box and ensure that the last subgroup number is in the "To" box.

Click OK

Press the + key a few time so that you can see about 100 data points on the chart (if you have a numberpad use that + key,). Drag the right edge of the chart if you want to make the display wider.

Use the scroll bar at the bottom of the chart or press 4 and 6 on the numberpad (Num lock must be on) to look all 200 data points.



You can see that this set of data is basically stable (keep in mind that there can be occasional false signals in control charts). So, up till now in this lesson, we were looking at data from a stable process.

We will return to the histogram and note some figures.

Click on the control chart then press the Escape key or click on the X button to remove the control chart.

We now need to record some of the values from the information box on the histogram.

Click anywhere on the histogram to return to the real specification limits of 300 and 700.

Write down on paper the values in the information box for Average, St dev, Est St dev, OOS, Cp, Cpk and Ppk.

Before we are finished with this set of data. we will take a last look at the position of the real specification limits relative to the spread of the histogram.

Look at the histogram.

Notice that the specification limits are near the ends of the tails of the histogram. In fact there is only 1 point out of specification in the 1000 measurements included in the histogram.

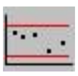
Now we will look at some different data.

Click the Database Explorer button  on the Main Panel

Click the Open Database . button Select the file 'TutData.mdb'.

Select English in the Language list and Unstable in the Data Table list.

Click the Open button.

Click the Control Chart button  on the Main Panel

Check that Xbar and Range is selected as the chart type and click OK

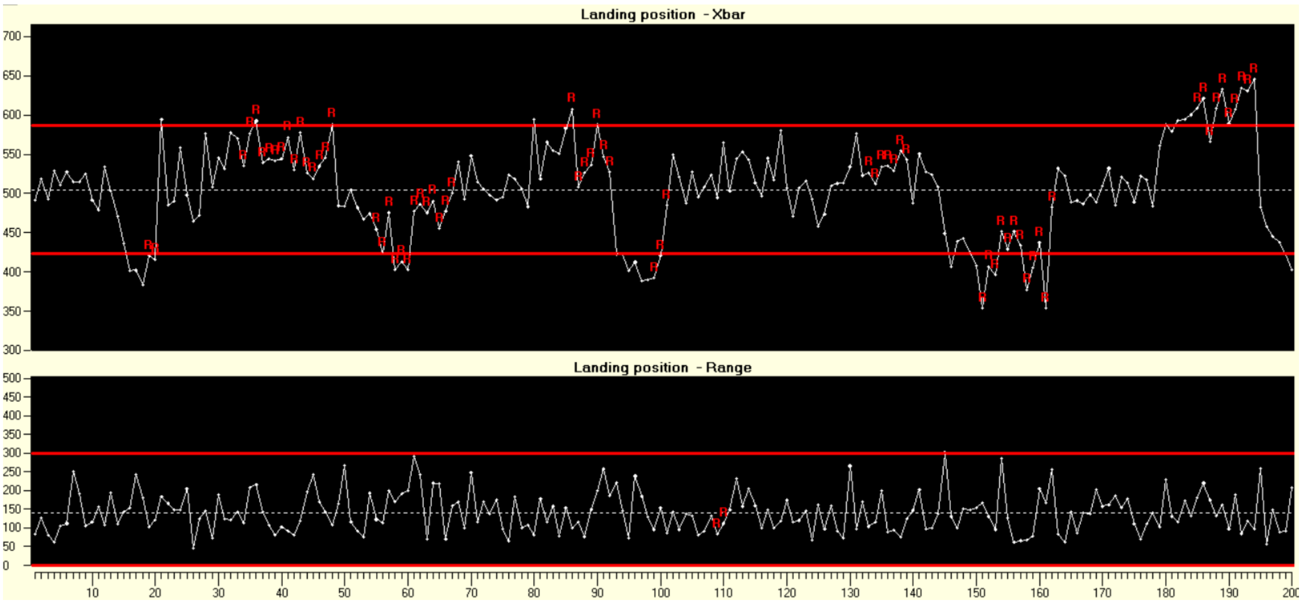
Press the control limits button  on the chart toolbar.

Put "1" in the "From" box and ensure that the last subgroup number is in the "To" box.

Click OK

Click the + key a few times so that you can see 100 or more data points (if you have a numberpad use that + key,). Drag the right edge of the chart if you want to make the display wider.

Use the scroll bar at the bottom of the chart or press 4 and 6 on the numberpad (Num lock must be on) to look all 200 data points.

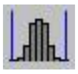


This data is showing signs of instability. There are many points outside the control limits and clusters of runs where 8 or more points are on the same side of the average line. So, keep in mind for the rest of this exercise that we are looking at data from an unstable process.

Close down the control chart.

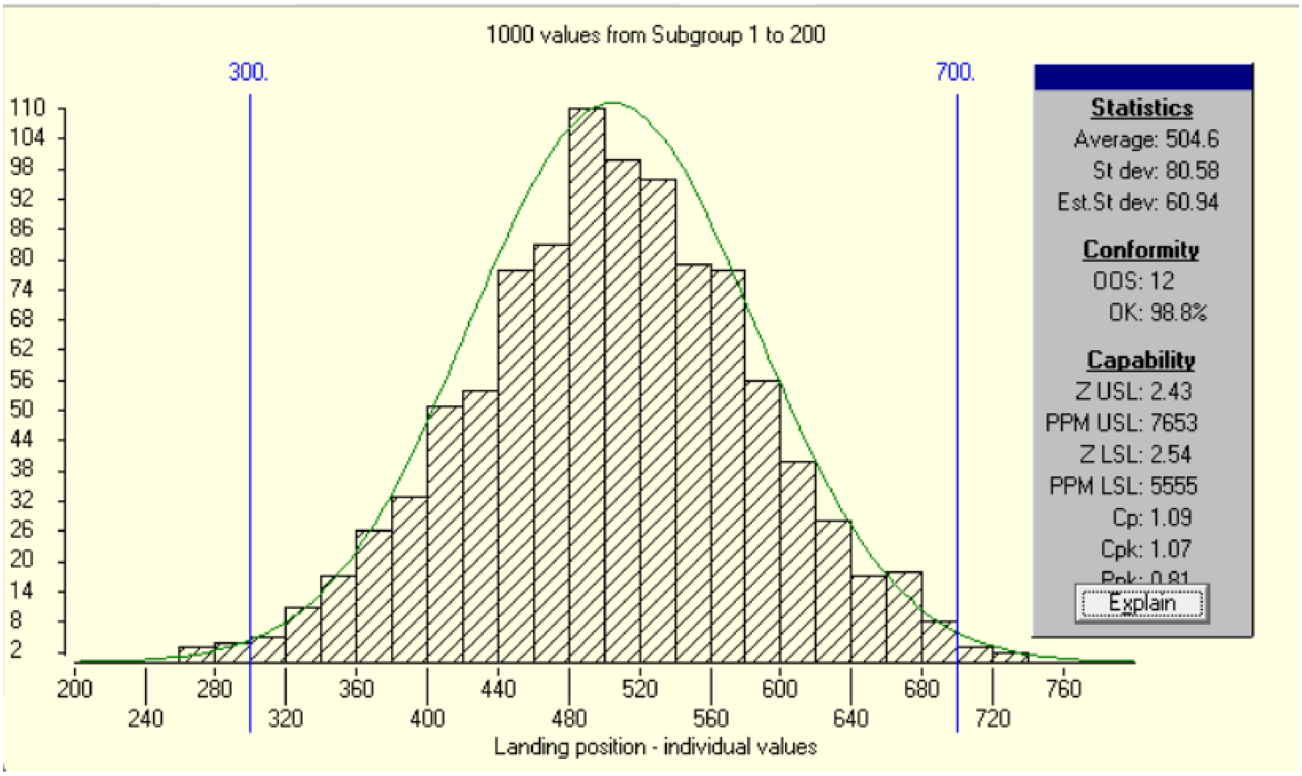
Press the Escape key or click the X button on the chart.

Now let's look at this data on a histogram.

Click the histogram button  on the Main panel toolbar

Ensure that "Individual values" is selected under "Make the histogram from" & click OK

Ensure that the Specification limits button  on the histogram toolbar is in the depressed position.



Notice that the tails of the histogram spread over the specification limits more than they did with the previous data.

Click the indices button on the histogram toolbar

The out of specification figure is 12, with the previously data it was 1. So it is clear that this process is performing less well than the process which produced the previous set of data.

Look at the values in the Statistics section of the information box and compare them with the values you recorded for the stable process.

You will notice that the values for Average and Estimated Standard Deviation are very similar to the values for the stable process. The value for real Standard Deviation, however, is higher.

Click the colours button on the histogram toolbar.

Click on the small down arrow to the right of the button and select "Estimated std dev".

Again click on the small down arrow to the right of the button and this time select "Real std dev".

Do this a few times and look at the coloured bars.

The coloured bands are changing between true Standard Deviation and Estimated Standard Deviation each time you click. This time, the real standard deviation is wider than the estimated standard deviation.

Leave the histogram showing Estimated Standard Deviation zones.

Click on the small down arrow to the right of the button and select "Estimated std dev".

Now we will look at the capability indices with this unstable process.

Compare the figures for Cp, Cpk and Ppk in the Information box with those values you recorded from the previous data.

Cp and Cpk are very near the figures for the previous process, however, the value for Ppk is lower now. Let's look at this in more detail.

Click on "Cpk" in the information box.

Read the explanation again.

Click on "Ppk" in the information box.

Read the explanation carefully.

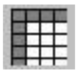
You might expect that ALL the indices would be lower for this process than the previous process. After all, the spread of results is wider now and there are more out of specification points. To understand these indices fully, you have to understand what they were designed to tell us about processes.

Cp and Cpk are CAPABILITY indices. They tell us what the process will be capable of producing after we have made it fully stable. The process which produced the present data has special causes of variation which are making the histogram wider than it could be. Because this process has a Cp and Cpk of about 1, we can expect to get very few out of specification points **after we have removed the special causes of variation**.

Ppk is a PERFORMANCE index. It tells us what the process has produced in the past. For this index it does not matter if the spread of the results was caused by special type variation or common type variation.

As we saw with the previous data, for a fully stable process, Ppk is approximately equal to Cpk.

Indexes when there is only one specification limit:
Some products or services have only one specification limit. For example, a "24-hour" parcel delivery service has only one specification limit - nobody minds how early the goods are delivered.

If the data table is not already opened click the table button  on the Main Panel toolbar.

Click on any of the 5 "Launcher position" column headers (the grey cell at the top of the column) with the RIGHT mouse key.

Select Column Properties

Switch off the one of the specification limits.

Click the checkbox beside "USL" in the Specification limits frame.

Click OK

When there is only one specification limit, we still calculate Cpk and Ppk in the same way. Cpk and Ppk only take into account the nearest specification limit. When there is only one limit, that is the nearest one.

Cp is a little different. When there are two limits, the calculation is:

(Upper specification limit - Lower specification limit) divided by 6 times Estimated Standard Deviation

When there is only an upper specification limit, the calculation is:

(Specification limit - Average) divided by 3 times Estimated Standard Deviation When

there is only a lower specification limit, the calculation is:

(Average - Specification limit) divided by 3 times Estimated Standard Deviation

These are exactly the same equations as for Cpk. Confirm this by dragging the specification limit line.

Drag the specification limit line

Compare Cp with Cpk as you drag

Other indices:
Other indices which you may come across are Z (distance to specification) and PPM (parts per million). The new AIAG VDA SPC manual has introduced new capability indices based on the geometric distribution. For further reading check our blog https://datalyzer.com/resources/blog-what-the-new-vda-aiag-spc-manuals-capability-indices-mean-in-practice-cp-cpk-pp-ppk-cw-cwk-pm-g-pmk-g-cp-g-cpk-g-pp-g-ppk-g/

Lesson 10 summary:

1. The easiest way to determine the capability of a process is to superimpose the Specification Limits on a histogram.
2. Cp is a capability index which indicates how well the process will perform after the process has been made fully stable and after it has been correctly "centered" between the specification limits.

3. Cpk is a capability index which indicates how well the process will perform after the process has been made fully stable, assuming it will be "centered" as it is now.
4. Ppk is a performance index which indicates how well the process has performed up till now. It takes into account the spread of the results and how well the process has been centred between the specification limits.
5. If there is only one specification limit, Cp = Cpk
6. For Cp, Cpk and Ppk, a value of 1 is the minimum value to consider the process "capable". The higher the value, the better.

End of Lesson 10

Appendix: SPC in non manufacturing

Distinguishing between common cause variation and special cause variation is just as important in non-manufacturing processes. For example, if the latest quarter's sales figures in an area are up compared with the previous period, we need to know if this is simply part of the natural ups and downs inherent in the selling process (common cause variation), or due to something unusual which has significantly altered the success rate - such as an improved advertising campaign or some extra-special effort from a salesperson (special cause variation).

Imagine what happens if we think that the advertising campaign or the extra efforts were responsible for the improved figures but, in fact, they were simply due to the normal common cause variation. We will probably declare the campaign a success and adopt it across the country, and we might give the salesperson a prize.

Next quarter, we are astonished to find that some areas get reduced sales with the new advertising campaign and the 'top' salesperson is well down the league of achievers. We do not understand why we did so badly and disillusion sets in. In fact we did not do badly, this is simply common cause variation at work again.

Summary

Let us go over the main points that you should now understand about processes, and the use of this type of control chart:

1. All processes contain variation.
2. To achieve the above, we must distinguish between special cause variation and common cause variation. We need to know this difference because the things we will have to do to remove or reduce the two types of variation are very different.
3. The way to distinguish between common cause variation and special cause variation is to use a control chart.
4. The calculations for the distance of the control limits from the average are based on an average short-term dispersion measurement.
5. We must use our knowledge of the process when deciding how to sample results and arrange them into subgroups. We should do this in a way which we know will reduce the chances of special cause variation occurring within subgroups. The range of each subgroup is a convenient dispersion measurement.
6. If we do not know much about the process and we cannot be confident that little special cause variation will be present within subgroups, then we should use a Moving Range chart where the dispersion measurement is based on the difference between subsequent individual results.
7. Operators should understand the charts. They should be involved in producing them and plotting the results.
8. Before we can consider a process to be "under control", efforts must be made to remove special causes of variation. We must also learn from each incident of special variation and take action to make sure that these types of changes do not happen again - this may require close co-operation with suppliers. Some

companies are not capable of this type of management - they will always exist in a state of chaos even if they have control charts.

9. If the process is statistically stable, any adjustments that we make to the process average will create more variation on the output.