Statistical Methods for Quality Control
II. Statistical Process Control [Control Charts]

In general:

Methods for the control of a process (usually a single, continuing manufacturing process) to determine when the process needs adjusment or correction (when it is *out of control*).
The problem is that there will always be random variation (ideally quite small, but always present) in the output, even when the process is in control, and we cannot test all the output, so we must rely on data from samples. When the process is in control, the only variation is the variation due to *common causes* ; when it is out of control, we look for the *assignable causes* so that we can correct them. This can be seen seen as a hypothesis test situation with the null hypohtesis H_0 "The process is in control" (Not directly a statement about a mean, or proportion) and H_a "The process is out of control". There are the usual four possible combinations of reality (H_0 is true or H_a is true) and our conclusion (We believe H_0 or we believe H_a). Type I error (believing the process is out of control when it isn't) leads to lost production because we stop production to make adjustments that aren't needed (though we think they are). Type II error (believing the process is in control when it isn't) leads to production of defective goods because we let the process run although it is producing defective goods.
We will look at control charts based on four different statistics: $\bar{x}$ (sample mean), R (sample range) p (really $\bar{p}$ – the sample proportion) and np (really $n\bar{p}$ – number of defectives in sample) . The $\bar{x}$ and R charts are used when we are considering measurements (length, width, diameter, weight) on the items produced, the p and np charts are used when we are looking at the proportion (or number) of items produced that do not meet specifications (which may be more complex).

The Scheme:

We will draw samples, all of the same size, from the production line at different times and calculate the sample mean (or range, or proportion of defective items, or number of defective items). If the value is more than three standard errors above or below the expected value, we will conclude the process is out of control and take appropriate action. [If there is a suspicious pattern, over time, to the sample values, we believe the process is moving out of control]

To implement the scheme(described in terms of $\bar{x}$) :

We set up a chart with time (or sample number - that is first, second, third, etc.) on the horizontal axis, a scale for $\bar{x}$ on the vertical axis. We plot a horizontal line at $E(\bar{x})$ (the process mean) , a horizontal line at $E(\bar{x}) + 3\sigma_{\bar{x}}$ (the *upper control limit* or UCL) and one at $E(\bar{x}) - 3\sigma_{\bar{x}}$ (the *lower control limit* or LCL). [Recall that $\sigma_{\bar{x}} = \sigma_X/\sqrt{n}$]. This is our x-bar chart. As we take each sample, we plot the value of $\bar{x}$ for that sample on the chart (moving from left to right). A sample that gives a point (a value of $\bar{x}$) above the upper control limit or below the lower control limit indicates the process is out of control [If the process were in control, the probability of such a sample is less than 1% - we show at the 1% level that the process is not in control] - we stop the process and make corrections.
In addition, a pattern of 1. many sample values on the same side of the process mean line or 2. a trend line from six or seven points showing a movement in the same direction are taken to indicate the process is headed out of control and corrective action should be taken.

Extensions, details:(other statistics, estimating the process mean and control limits from sample data)

Setting up process mean and control limits from sample data (x-bar chart):

Since the process mean and standard deviation are theoretical values, we usually (always, in practice) have to estimate them from sample data. We do this by taking several samples while the process is in control, and using the sample means and ranges to estimate $E(\bar{x})$ and $\sigma_{\bar{x}}$. If we have k samples, all of the same size n , we approximate $E(\bar{x})$ with the overall sample mean $\bar{\bar{x}} = \dfrac{\bar{x}_1 + \bar{x}_2 + \cdots + \bar{x}_k}{k}$. This gives the value we use for the process mean line.

The control limits are a bit more complicated. We use sample ranges (easier to calculate than sample standard deviations) to approximate the process standard deviation σ and use this to estimate both $\sigma_{\bar{x}} = \dfrac{\sigma_X}{\sqrt{n}}$ and $3\sigma_{\bar{x}} = 3\dfrac{\sigma_X}{\sqrt{n}}$. First we need the mean of the sample ranges $\bar{R} = \dfrac{R_1 + R_2 + \cdots + R_k}{k}$ There is a rule of thumb for estimating σ from $\bar{R}$; the major calculations are hidden behind a table (table 20.3 p. 858) - we use $\dfrac{\bar{R}}{d_2}$ to estimate σ, where d_2 depends on the sample size and is obtained from the first column in table 20.3. Since we want our control limits to be $E(\bar{x}) \pm 3\frac{\sigma}{\sqrt{n}}$ we get $E(\bar{x}) \pm 3\frac{\sigma}{\sqrt{n}} = E(\bar{x}) \pm 3\frac{\bar{R}}{d_2\sqrt{n}}$ Fortunately for us, table 20.3 includes values for $\frac{3}{d_2\sqrt{n}}$ in the column labeled A2 , so we finally obtain (for calculating control limits from sample data - obtained while the process is in control) Process mean $\bar{\bar{x}}$, UCL $= \bar{\bar{x}} + A_2\bar{R}$, LCL $= \bar{\bar{x}} - A_2\bar{R}$

R–chart:

To monitor the *variability* of the value being measured , we use an R (range) chart. In practice, this is usually used with an x-bar chart, and is usually read first - if the R-chart shows the process is out of control (values outside the control limits), the x-bar chart will not even be completed (the process will be adjusted). For an R-chart, we set up a chart to plot the sample ranges on a chart with the same design as an x-bar chart. For our "process mean" we want the mean range for samples of size n— we get this from $\bar{R}$ (defined above) and set our control limits $3\sigma_{\bar{R}}$ above and below $\bar{R}$ – our estimate for $\sigma_{\bar{R}}$ is $d_3\frac{\bar{R}}{d_2}$with d_3 coming from table 20.3. Thus our control limits are $\bar{R} \pm 3\left(d_3\frac{\bar{R}}{d_2}\right) = \bar{R}\left(1 \pm 3\frac{d_3}{d_2}\right)$. To make life yet better, table 20.3 includes columns for $D_3 = 1 - 3\frac{d_3}{d_2}$ and $D_4 = 1 + 3\frac{d_3}{d_2}$
Thus we have, for the R-chart: Centerline at $\bar{R}$, LCL $= \bar{R}D_3$, UCL $= \bar{R}D_4$

p-chart:

The p-chart (and its sibling the np -chart) is used when quality is mesured in terms of items being either satisfactory or defective. The chart is made to plot values of $\bar{p}$ — the proportion of defective items in the a sample . Our process mean (of proportions) is p , the standard deviation is $\sigma_p = \sqrt{\frac{p(1-p)}{n}}$ so, if our sample size is large enough for the binomial distribution to be approximated by the normal, the control limits are : LCL $= p - 3\sqrt{\frac{p(1-p)}{n}}$, UCL $= p + \sqrt{\frac{p(1-p)}{n}}$ (Otherwise tables - which we will not look at - are used) If we do not have a historical (or pre-set) value for p, we use several samples when the process is known to be in control and approximate p and $\sigma_{\bar{p}}$ by pooling these.

np-chart:

The np-chart works like the p-chart but is set up for using the *count*count of the number of defective items in each sample, rather than the *proportion*. The center line is at np and the control limits are: LCL $= np - 3\sqrt{np(1-p)}$ and UCL $= np + 3\sqrt{np(1-p)}$