Case Study: Analysing and Plotting from Large Datasets

We have frequently seen examples where data is being carefully collected from multiple sources, carefully recorded in an Excel workbook and well, not much more than that. Various people might try building their own charts to look for trends and disturbances.

All too often the result is a bit of a mess, there are about 15 different charts, everything gets plotted on the primary access so some plots look like flat lines bouncing along the bottom of the chart. There is no statistical analysis of the data, not even any indication of which data points are within acceptable limits, and which aren't.

At the end of all of that, even if someone does come up with some nice looking charts that tell the story that you are interested in, they never update as new information is loaded into the workbook. If they ever do update the plots they have to painfully go though every plot line revising its start and end point.

So what do you do? In our case we built a generic trend workbook which has the following characteristics:
- Point and click selection of the data to be plotted;
- Point and click selection of start time and end time;
- For each data point selected, these are plotted individually plotting on a Run Chart have upper and lower control limits calculated from the dataset itself;
- Hard wired upper and lower specification limits with immediate feedback on any data that falls outside those limits;
- Calculation of various statistical information from the data, providing direct feedback on process capability measures, variation and step shifts in the process.

Initially this was done by looking up the data requested, which added a significant overhead to the workbook in terms of size and the time taken to update. This also had the effect of limiting the maximum number of records that could be examined at one time. It worked well enough, but had its limitations.

Since that time we have developed a trending and analysis package which does not use any lookups for the data, it simply looks at exactly the right place in the main data table. This enhancement means that the workbooks are only ever very slightly bigger than the size of the data table. It also means that any period of time can be examined up to and including as many records as one has in the main data table.

With a very basic introduction into the trending workbook users can very quickly start looking at the data that is of interest to them, knowing only how to point a click to see what they want. In addition, the use of upper and lower specification limits, upper and lower control limits gives users immediate feedback on whether they have a problem or not, in a very clear and unambiguous manner.