Today: Quality Control and various related topics:
i.e. management by exception
profiting from reduced variability
graphical methods of monitoring variability

Start with Tanur article on colour tolerances (pp 170-176):

Automobile body parts are produced to match a certain colour standard, and the study described was to try to relate objective measurements of colour hues to what the subjects thought was a good or a bad match of colour. Instruments measure colour but a slight difference can often be detected by a human eye. The question was, how much of a difference in colour from the nominal standard, as measured by the instruments, would be judged as a negligible difference by a human observer? The instruments measured three aspects of each sample of colour (a piece of coloured plastic) and subtracted these from the measurements on the standard (the target colour represented by a particular design specification). So only the differences of the three indexes were looked at.

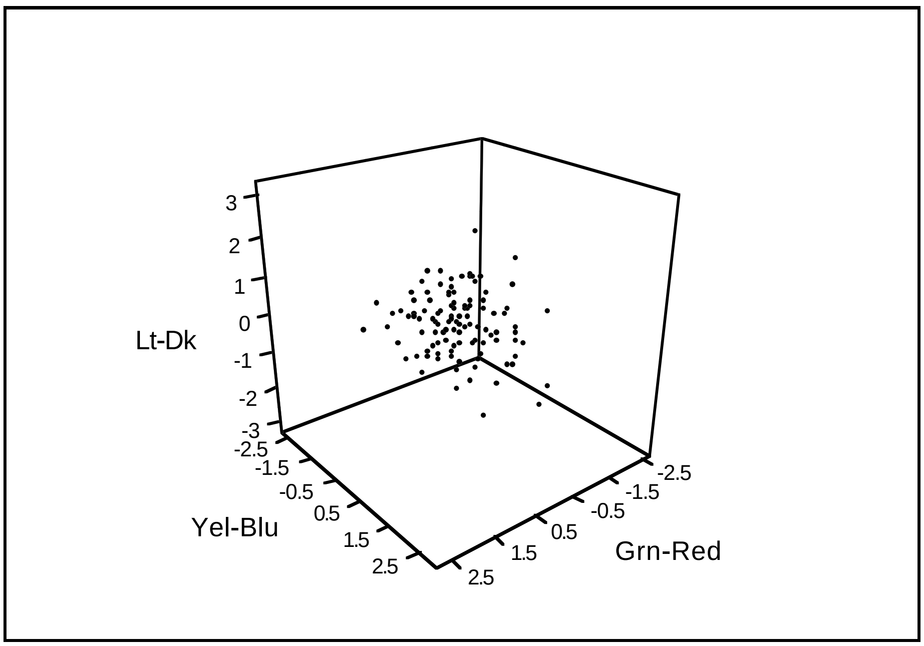
Here is some data of the type described in the article (but fictitious):

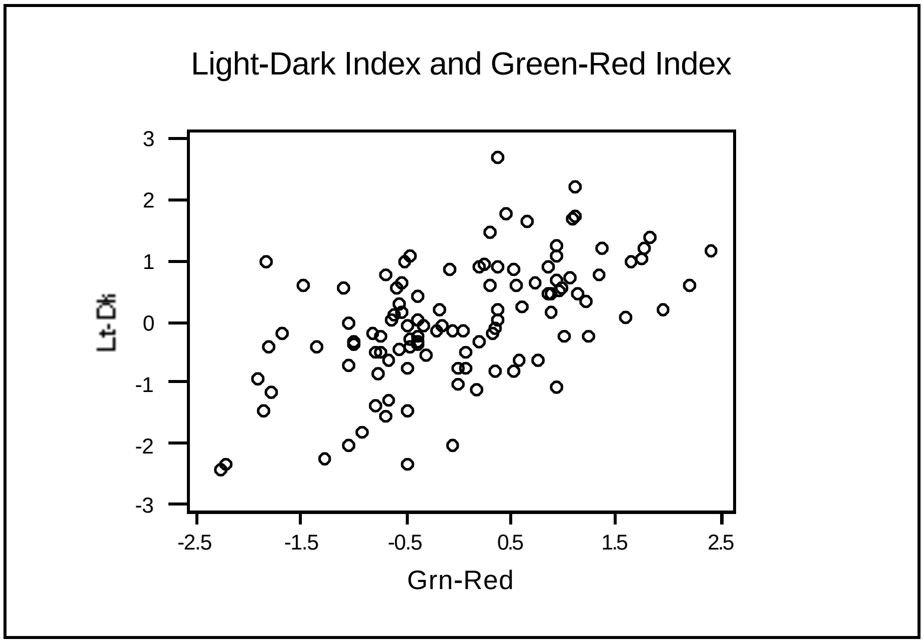
| Row | Lt-Dk | Grn-Red | Yel-Blu |
|---|---|---|---|
| 1 | 0.56 | 0.99 | 0.89 |
| 2 | 1.21 | 1.37 | 1.93 |
| 3 | 0.03 | 0.38 | -0.32 |
| 4 | 0.63 | -0.54 | -0.94 |
| 5 | -1.83 | -0.93 | 1.49 |
| . | . | . | . |
| . | . | . | . |
| . | . | . | . |
| 107 | 0.91 | 0.37 | -0.77 |
| 108 | -0.05 | -0.15 | 0.99 |
| 109 | -0.14 | -0.05 | -0.44 |
| 110 | -0.06 | -0.34 | -1.43 |
| 111 | -0.00 | -1.05 | 0.19 |

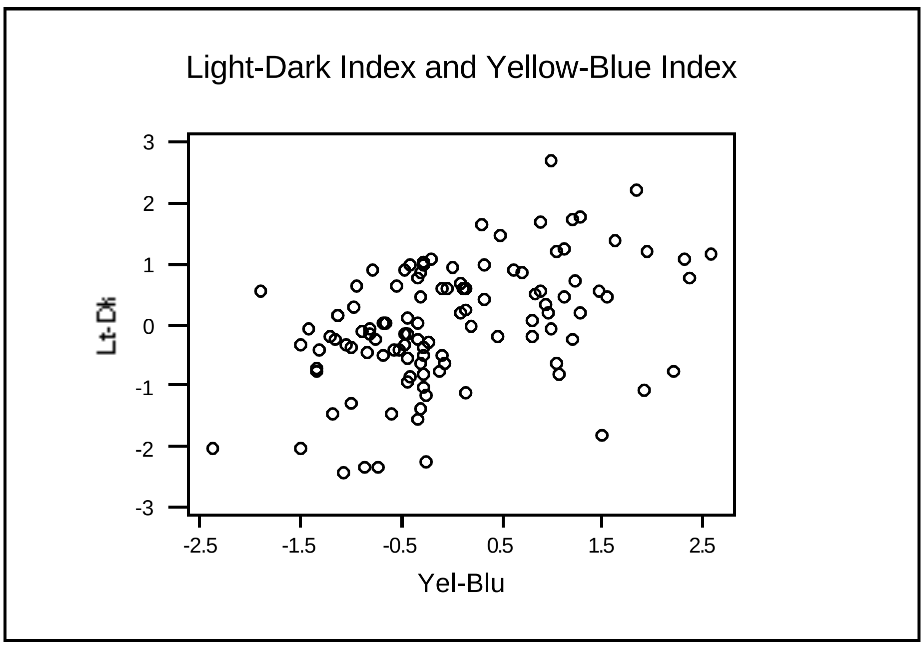
The schematic diagram in the article might have included data that looks like this (see diagram next page):

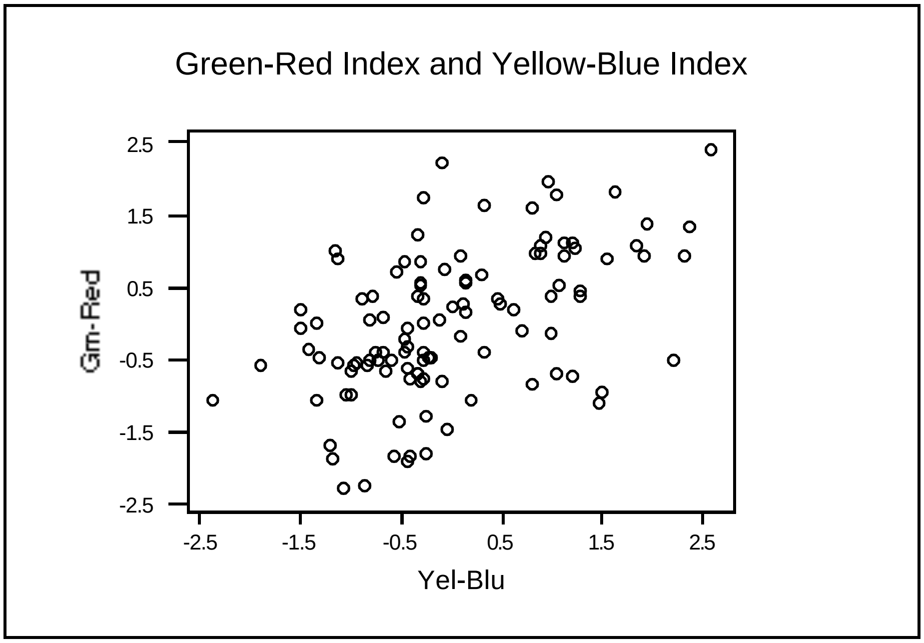
These points represent the actual colours relative to a standard. A value of (0,0,0) would be a perfect match to the target colour, and a value like (-1, 2, 3) would be low on the Lt-Dk index, high on the Grn-Red and on the Yel-Blu index. In fact one way to summarize the extent to which a colour (a,b,c) departs from the desired target colour is to report the distance of the point (a,b,c) from (0,0,0). But the formula for this distance is just $\sqrt{a^2+b^2+c^2}$ which in the case of a colour sample (-1,2,3) would be a distance of $\sqrt{14}=3.8$

We can try to draw a diagram of the data using a three dimensional scatter plot (see next page), but it is hard to really see the shape of the scatter in this 2-dimensional diagram. However we can look at the variables two at a time. (see below)




Light-Dark Index and Green-Red Index
Lt-Dk
3
2
1
0
-1
-2
-3
-2.5
-1.5
-0.5
0.5
1.5
2.5
Grn-Red


Light-Dark Index and Yellow-Blue Index
Lt-Dk
3
2
1
0
-1
-2
-3
-2.5
-1.5
-0.5
0.5
1.5
2.5
Yel-Blu




Note that the correlations are positive – this is not logically necessary – but whatever the correlations are in the real data would be of interest to the people responsible for controlling the colours. With my simulated data shown here, I have arranged that all three variables have SD = 1 since the actual scale is arbitrary. Note that the average colour indices should be about 0, since these are average differences from the target colour, and while a perfect match would usually bepossible, getting it right on average might be achievable. (called "unbiasedness").

Let us now return to the great simplification of the problem that the "distance" idea achieves. Instead of struggling with the three plots, and the hidden links between them, let's use the distance as a measure of how good the colour match is. The first few were:

| Row | Lt-Dk | Grn-Red | Yel-Blu | Distance |
|---|---|---|---|---|
| 1 | 0.56 | 0.99 | 0.89 | 2.08 |
| 2 | 1.21 | 1.37 | 1.93 | 7.07 |
| 3 | 0.03 | 0.38 | -0.32 | 0.25 |
| 4 | 0.63 | -0.54 | -0.94 | 1.59 |
| 5 | -1.83 | -0.93 | 1.49 | 6.45 |

The distance summarizes how bad the colour match is for that particular sample of plastic. Obviously, sample 3 is a very goog one while sample 2 is not so good. But how do we decide what is good enough?

We have to relate what the data is telling us with what our subjects say they can detect as a real colour mismatch. The article explains that it would be an advantage to have the index as a good predictor of the human response. Just like the Jury article, we try to use a training sample to link the measurements to the human responses. Then we would use this link without human input for future manufactured parts, to decide on their acceptability (colour-wise).

The distances for the 11 samples have a distribution like this:

```
           :
          .:
        :.::
       :::::. : :
       ::::::::::. : : . :   :
      .:::::::::::::.: :.:.. : :..:.  : .    .     ..   .     .
      +---------+---------+---------+---------+---------+------Distance
    0.0       2.5       5.0       7.5      10.0      12.5
```

A wonderful result for the auto part company would be that for some threshold (e.g. > 5.0 say) the human subjects say that there is a mismatch while for distances less than the threshold, the colour is declared a good match. But in reality, this did not happen. Nevertheless, the use of the automatic colour match test would be helpful even if it were not perfect, since having a number of people representing potential customers to judge each part was not practical.

But suppose the data looked like this:

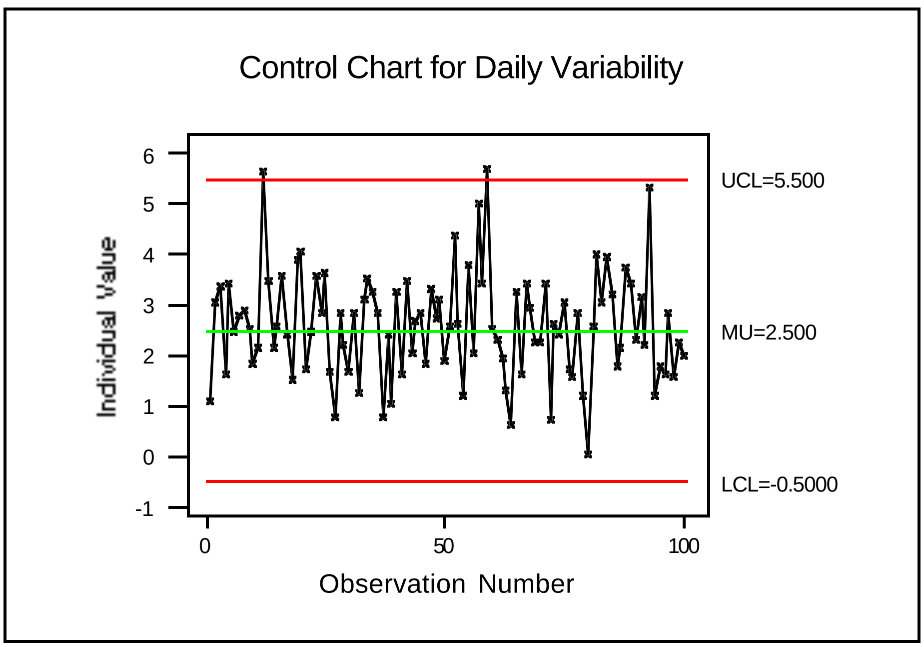
```
           :
           :
        :.::
       :::::  . .
       :::::: :::  : .   .   .
      .::::::.::::::.: :.:   : . .:
      +---------+---------+---------+---------+---------+-------OK
             .
          . .:. ..   . . ... . ..  .  : .    .     ..   .     .
      +---------+---------+---------+---------+---------+-------notOK
       0.0       2.5       5.0       7.5      10.0      12.5
```

If we used 5.0 as our threshold, we would be correctly classify about 13 samples as not-OK but would reject incorrectly 12 samples. Our OK decision based on the distance measurement would be correct in 79 samples but incorrect in 7 samples. Our overall percentage of correct decisions would be (13+79)/111 or 83%. So our robot that is automatically measuring colours is successfully replacing the human judge 83% of the time. (We assume that the samples met in production are roughly the same quality as those in the study with the 111 colour samples).

The term "**quality control**" usually refers to the regular monitoring of a production process to ensure that its output is acceptable. Often a quality measurement will reflect the degree to which a product matches its specifications. One of the dogmas of quality control is that variability is a bad thing. This is closely related to the principle of **management by exception –** management concentrates on the unusual happenings in the company. For example, an oil company examines those gas stations that have been **most** or **least** profitable in the latest month, rather than try to look at the performance of all stations.

Control charts are used to display quality measurements on a periodic basis (hourly, daily, weekly …) When variability thresholds are crossed, employees discuss the conditions which pertained to the timing of the increased variability, to try to detect the cause of the unusual variability, and to eliminate the cause.

Here is an example of a control chart:




Note that both high (bad) and low (good) values of the SD are considered worth detecting. Day 12 and Day 59 had exceptionally large SD and would have been followed up.

In the case of out colour matching, the distance index is already measure a “deviation” so this measure, rather than its SD, would likely be used.

Why is reduction of variability so important in manufacturing? Consider the sale of lumber, perhaps the familiar “2 x 4”. The milling produces boards of various qualities – not only do different species of wood have different characteristics, but some pieces are slightly bent, some have unsightly knots, some are nicked or eaten by insects … If we sell unsorted wood, we can only guarantee the lowest quality and so must accept the lowest price. But if we sort by quality, we can charge higher prices for higher quality. In other words, reducing variability would have a direct impact on improved profits. Similar stories can be told of the profitability of reducing waste, reducing the cost of warranty claims, and improving customer satisfaction. Quality control has been a big part of industrial culture over the past 50 years. Lately the philosophy of quality control has expanded to include more general management practices. A name you will hear in this context is “Six-Sigma” management. This name has its roots in the ± 3 SDs that we have already mentioned.