AICOS Technologies AG
Christian Keller & Dr. Yves-L. Grize
Efringerstrasse 32
CH-4057 Basel, Switzerland
Tel.: +41 61 686 98 77
Fax: +41 61 686 98 88
Web: http://www.aicos.com

Visualization Methods in Statistical Quality Control[1]

Quality technology today

Quality control has strongly changed during the last 10 years. Today's way of thinking in quality control is no longer product-oriented but process-oriented. Quality cannot any more be achieved by inspection alone, but most of all by planning and by process monitoring. Further, striving after permanent improvements has become a basic requirement for quality.

Because of these changes, collaborators are forced to work more efficiently and to evaluate by themselves often complex relationships. They are asked to think pro-actively and to deliver results quickly. In this context, methods for visualizing the information hiding in the data play in particular an important role. They support a competent evaluation, facilitate the communication necessary and help to detect problems and their solutions more quickly. Finally, with the development of computers and the more and more user-friendly interfaces, informative graphs can easily be produced.

In this article four visualization methods which are important in quality control are presented: parallel boxplots, scatter plot matrix, control charts and cusum charts. All of these graphs can be generated easily and quickly by the validated Excel-macros EasyStat of the company AICOS Technologies AG.

Evaluation of an interlaboratory experiment: parallel boxplots

Problem: If a test is analyzed in several different laboratories, one will never obtain the same result. This indicates that there are random or systematic influence factors leading to differences between the laboratories. There are several reasons for such differences like e.g., different measuring instruments or environmental influences. At least a part of this variation is inevitable.

+ Questions: The purpose of interlaboratory experiments is to recognize and quantify the differences between laboratories. The following questions are in the forefront:

1. How large are the differences within a laboratory (repeatability) and between the laboratories (reproducibility)?
2. Are there systematic differences between the laboratories?

These questions and other practical considerations determine the organization of an interlaboratory experiment. In the simplest case this leads to conduct repeated measurements of the same test in some laboratories. The repetitions within a laboratory should be done under the same conditions.

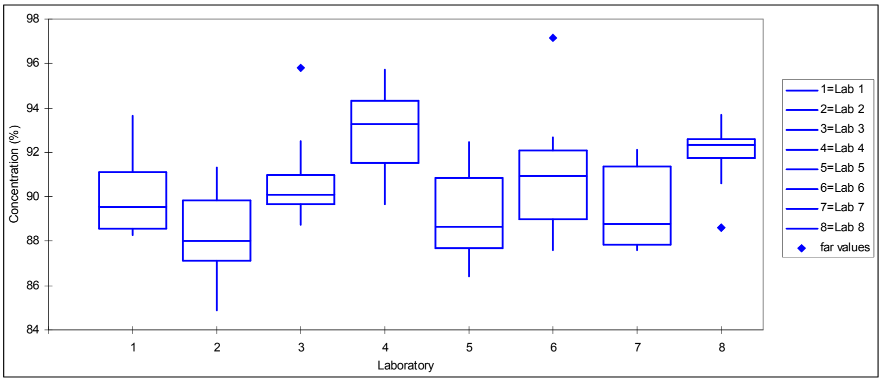

Figure 1: Parallel boxplot (with EasyStat 4.1) of the measurements from 8 laboratories.

[1] This is an English summary of a paper published in the German magazine *LaborPraxis* in April 1999.

+ Methodology: The statistical evaluation of an interlaboratory experiment consist mainly of splitting the total variability in a part within laboratories and in one between laboratories. An analysis of variance (ANOVA), as it can be carried out in Excel, does this job.

Before every statistical evaluation, it is advisable to represent the data graphically by proper techniques. For the analysis of interlaboratory experiments, parallel boxplots, as presented in Figure 1, prove themselves particularly valuable. Unusual measurements can be recognized more easily than in tables [2]. Moreover, the variability between and within laboratories can be estimated visually.

+ Example: In the example of Figure 1, the results of an analysis of a test by HPLC in eight different laboratories are represented, whereby in each laboratory ten measurements were carried out.

A boxplot delivers a compact graphical representation of a data set. Half of the data lies in the box whose lower and upper border are defined by the two quartiles. The median is drawn as a line within the box. The largest part of the data lies between the end points of the vertical lines (whiskers). If extreme observations, so-called outliers exist, they are represented separately as individual points (see Figure 2).

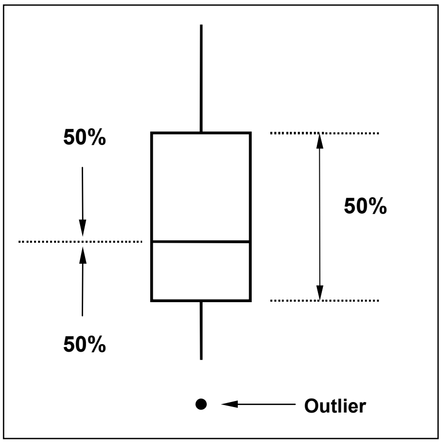

Figure 2: Schematic representation of a boxplot with one extreme observation.

+ Evaluation: Figure 1 shows that especially Laboratories 4 and 8 deliver systematically higher measurements. In addition, the spread in Laboratory 8 is significantly smaller than in other laboratories. Finally, in Laboratories 3, 6 and 8 three extreme values deviating strongly from the remaining measurements of the corresponding laboratories attract notice. The F-test of an ANOVA would only show that the differences between laboratories are statistically significant. Hence, the question of which laboratories are significantly different is not yet answered. On the contrary, parallel boxplots can quickly give this information.

Parallel boxplots deliver much more valuable information than an ANOVA does. They allow statements about the distribution of the data and comparisons between groups (laboratories).

Discovery of hidden relationships: scatter plot matrix

+ Problem: Often the quality of a product is evaluated by several different characteristics. In such multivariate situations, it is interesting to find out relationships between the features which can lead to new insights.

+ Questions:

1. Can expensive measurements be eliminated without losing essential information, if there is a strong correlation between single quality features?
2. Can a product be optimized and its quality be improved through the knowledge of the relationships between quality features and process parameters?

+ Technique: A measure of the (linear) relationship of two quantities is the correlation. For multivariate data sets the correlation coefficients computed pairwise can be written in a matrix form. An isolated judgement of these numbers however can often lead to totally wrong interpretations. By the aim of a scatter plot matrix relationships between several quantities can be visualized in a very useful way. It consists of a systematic arrangement of all the scatter plots between every two variables (see Figure 3). The graph instructs not only about the strength and direction of the relationship, but also informs about its form (linear/nonlinear).

+ Example: As an example let us consider the quality analysis of a dye-stuff (see [3]). During the production of the dye-stuff, several analytic properties have been determined by liquid chromatography. Different quality features like color intensity, tone or brightness are determined at the end of the production process by technical or visual measurements. Since the corrections of a faulty batch are cost and time intensive, one tries to predict the quality of the dye-stuff from the analytical properties during production.

+ Evaluation: Figure 3 shows the scatter plot matrix between three analytic measurements (sum of the red and of the green dye-stuff as well as the characteristic wave-length) and two quality features (hue and brightness). There are more or less strong (linear) relationships between the variables. The hue can for instance be predicted right well by the characteristic wave-length. The greater the wave-length, the deeper the value of the hue. However, for the other quality feature brightness, there is only a weak (positive) correlation with the characteristic wave-length.

In the scatter plot matrix, special points attract attention at once, like e.g., that observation having a great value for the wave-length as well as for the hue (see the third scatter plot from left in the second line of Figure 3). It obviously is an outlier. Such information will be lost by the computation of a correlation coefficient.

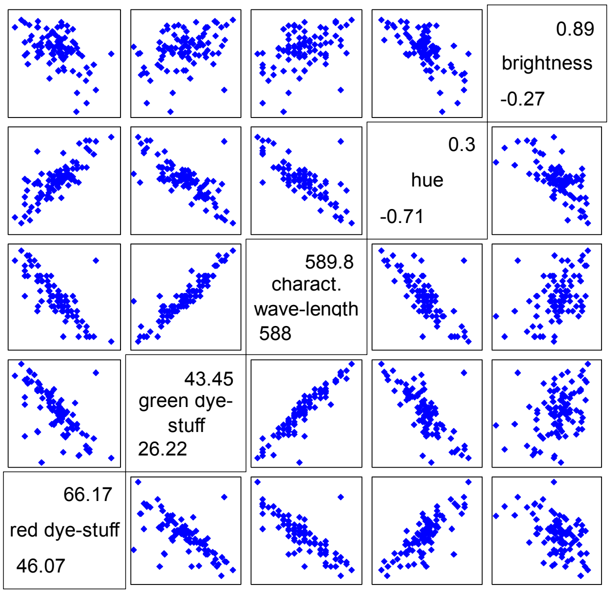

Figure 3: Scatter plot matrix with EasyStat 4.1

Identification of small shifts in the process: control and cusum charts

+ Problem: A process in operation is influenced by random and systematic effects that affect its quality. Random influences are the sum of several smaller influences (error in measurement, external disturbances) which are permanently there. They form the natural process variability and cannot be reduced. On the other hand, systematic, non-random influences (internal disturbances, wear and tear) are not predictable and appear unregularly.

+ Questions: The task of statistical quality and process control is to detect such disturbances and to help to find their causes. Typical questions in this context are:

1. Is the process under control?
2. Has a shift in the process mean value taken place? If yes, one would like to identify the time of the shift and to determine its causes.

+ Methods: In order to avoid that a process produces results not consistent with the quality requirements, it should be permanently supervised by statistical process control. In addition to statistical models such as regression or time series analysis, two important graphical tools also help to give answers to the questions above: the control chart and the cusum chart (see [4]). The continuous use of these techniques make it possible to discover shifts in the process in an early stadium. Scrap can then be avoided before it is produced at all.

+ Example: Consider as an example the production of a drug in capsule form. During production, attention should be paid that the active substance concentration reached in the capsules is great enough, which is a difficult task because of the small weight of the substance. The active substance is therefore stamped in the capsule matrix mechanically. The stamp volume of the filled powder has to be controlled (in ml/100 mg).

+ Evaluation: The data are represented in Figure 4 as a control chart. The control limits define the area within which the values should lie, if the process is under statistical control. If there is a point outside the control limits, then it suggests a process disturbance.

Often, two types of control limits are computed which are based either on the global variability (standard deviation of all observations) or on the local variability (mean difference of consecutive values). If these limits are different, an indication of process perturbations like, e.g., trends or cycles is present. In our example the two control borders are almost identical and overlap each other. The process seems to be under control; all the points lie clearly inside the control limits.

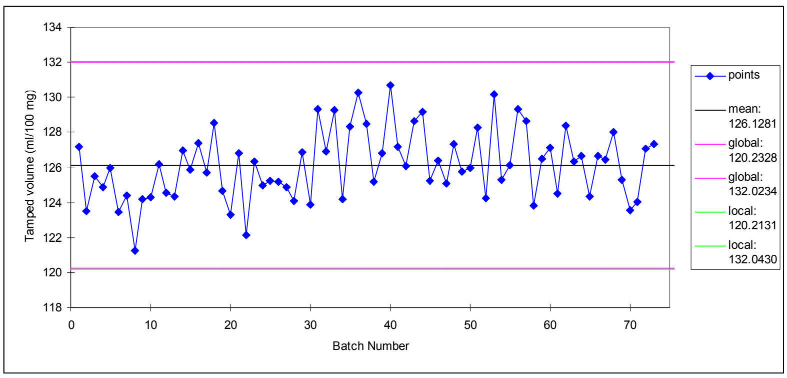

Figure 4: Control chart for the stamp volumes of 73 Batches with EasyStat 4.1.

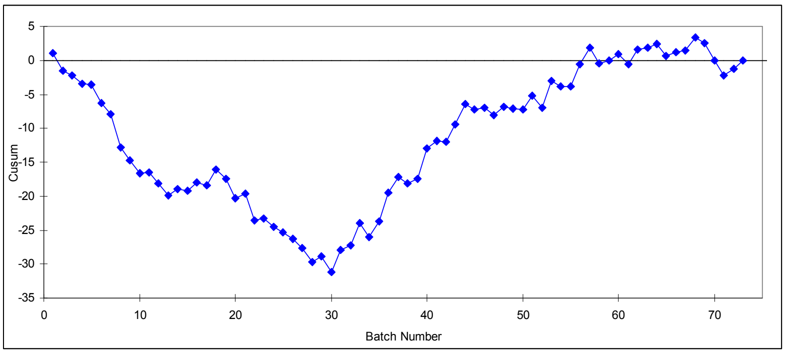

Figure 5: Cusum chart for the data of Figure 4 with EasyStat 4.1.

In process control, it is also important to be able to detect small shifts over a long time period. For this task control charts with memory are suitable, like for instance the cusum-chart of Figure 5. It uses the history of the process such that the deviations from the process mean or nominal value are cumulatively summed up. In this way, small but persistent deviations from the nominal value can quickly be detected. A turning-point such as in Figure 5 shows a strong indication that at Batch 30 there was a shift in the process mean value. However, with the usual control chart in Figure 4 such small shifts can hardly be detected.

In the example considered, a reason could in fact be found to explain the process shift. Inquiries showed that exactly after Batch 30 the cooling water was changed from tap-water to river-water. The temperature of tap-water varies less than the one of river-water. Using that information, the production was from then on always cooled by tap-water, and this eliminated the difficulties with the stamp volume.

Further methods

In addition to the methods explained in this article, several other visualization techniques relevant for quality control are available. Worth mentioning are especially

+ steam and leaf representation,
+ histogram,
+ quantile-plot,
+ stratification.

Because of their efficiency, these tools together with the three methods of this article build the so-called Magnificent Seven tools of quality control (see [5]).

All these techniques as well as further statistical routine analyses are implemented in the very user-friendly and validated Excel macros EasyStat (see [6]).

Summary

Decisions must be based on objective information and verifiable knowledge. Statistical and other quantitative methods help to extract the information contained in the data. Through visualization, this information can be accessed easily and is directly interpretable. Simple statistical techniques help

+ to understand the data,
+ to recognize the information in the data,
+ to make clear documentation available,
+ to judge quality,
+ to control processes,
+ to formulate well-founded statements and
+ to work rationally.

References

[1] Schmidli, H. und Grize, Y. L. (1996), Messen wir alle gleich gut? Ringversuche!, Qualität 7+8, S. 48–53.

[2] Keller, C. (1999), Ausreisser – und dann?, Qualität und Zuverlässigkeit 1, S. 91–96.

[3] Weihs, C. und Jessenberger, J. (1999), Statistische Methoden zur Qualitätssicherung und -optimierung, Wiley-VCH, Weinheim.

[4] Grize, Y. L. (1994), Kontrollkarten und was sie aussagen, Bulletin der Schweizerischen Arbeitsgemeinschaft für Qualitätsförderung (SAQ), Vol. 1, S. 18–23 und Vol. 2, S. 24–29.

[5] Flühler, H. (1992), Die sieben Glorreichen, Chimia 46, S. 251–254.

[6] EasyStat Version 4.1 Benutzeranleitung (1999), AICOS Technologies AG, Efringerstrasse 32, CH-4057 Basel.