What the Hell Is an Evaluation Strategy?

Stephan Conrad, TEQ® Training & Consulting GmbH

Jean-Claude Riber, a Franco-German stage director of opera said that statistics was like a Spanish guesthouse – everybody found in it what they were searching for. What a wise saying! After all, statistics is not a land of milk and honey where we just have to open our mouth and a roast chicken flies to us. However, it is even a dangerous saying since everybody might find what they seek, whether that is good or bad…

Therefore, Carl Hahn, former chairman of the Volkswagen group, thought that statistics had to be taken with a pinch of salt and should be applied with reason. This is a very good approach. There is even a quote of Paul Schnitker, former president of the German Confederation of Skilled Crafts, which supports it. He once said that statistics was like a building made of figures. Like a good house, statistics needed a solid foundation, clear outlines and the proof that they would keep their value in the course of time if they should be good. Bad statistics, by contrast, would collapse like a house of cards.

However, how can we ensure that statistics are built on a solid foundation? The aim is to describe available data through statistics gained in statistical data analyses, such as capability analyses, and to evaluate whether they meet the specified requirements. In order to do so, you generally have to take the following steps. You have to check the systematic and random variations of the data, describe them by means of suitable distribution (time) models and calculate capability indices from the statistics of the distribution.

Since there are numerous common distribution models, test procedures and ways to calculate statistics, this is where we enter the Spanish guesthouse. You just have to know what you want in order to be able –sometimes a little bit forcefully – to deduce the desired results. That's a piece of cake – of course, we will succeed somehow in retrieving a green smiley…

But that is certainly not our request! We would like to gain reliable results based on a firm and solidly built foundation and we still want to take them with a pinch of salt and apply them with reason. That is why we have to arrange and define which methods and procedures to apply and how to apply them. This is the only chance that each colleague reaches the same correct and reliable results that can be used in later evaluations.

Of course, there is the question of "right or wrong" in the background. In most cases, this question is quite easy to answer if the application contradicts the mathematical requirements for the respective methods and procedures. However, this does not mean that there is only one possible solution for each application. Sometimes several roads lead "to Rome" or at least in the close vicinity of Rome, but maybe they even lead to Stockholm since the results achieved on the way might differ. As long as a company or organization does not determine a valid structure, all these options are correct. Only the definition of an evaluation structure creates a binding framework in the company or organization and specifies what is considered to be "right" or "wrong".

This fixed structure for the application of statistical methods is called evaluation strategy. Companies and organizations may define it in work and process instructions or document it in guidelines. It is also included in the Q-DAS® software and forms the "heart" of the Q-DAS® software products qs-STAT®, solara.MP and destra®.

The Evaluation Strategy in Q-DAS® Software Products

The evaluation strategy is responsible for all decisions the software has to make in an automated capability or data analysis. Even though the respective strategy is discreetly hidden in the "Configuration" menu, the quality of the evaluation stands and falls with the settings adjusted in the evaluation strategy; the settings actually decide on the quality of the evaluation of measurement systems, machines and processes. Many customers include their quality guidelines in their own evaluation strategies and provide them to their suppliers. All the evaluation strategies that have already been included in the standard software versions on customer request are protected against modification and thus guarantee that a report indicating the "Smith and Jones Inc." strategy has really been evaluated according to the specifications of the Smith and Jones Inc. company. This is the reason why administrators shall protect customized strategies against modification by adjusting user rights.

The Evaluation Strategy Decides...

In order to achieve correct and consistent results in a company, you first have to adapt the configuration of evaluation to the general concerns of the company. If there is a customer with a specified evaluation strategy having high priority in your company, you may use this evaluation strategy as a template. Unfortunately, in most cases you have to consider several customers, i.e. you have to create several standards or steer a middle course based on all available specifications. Since these strategies sometimes differ considerably, the templates the
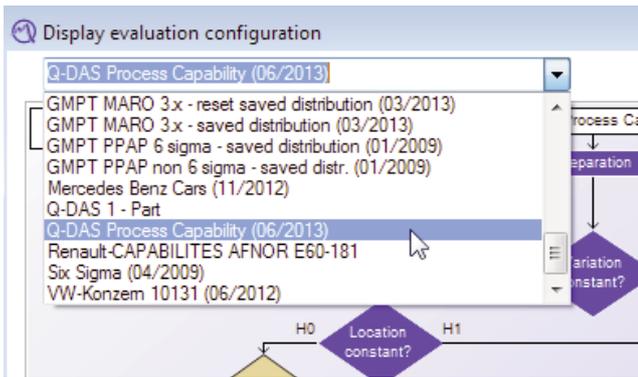
Figure 1: Evaluation strategies

software includes even though they have not been created by a certain customer only provide an example of a simplistic approach to a solution.

Even though a specific customer dominates the orders in your company and thus has the leading strategy, it is often necessary to change designations, minimum quantity or limits due to internal specifications and process descriptions. This is the only way to still apply with internal process descriptions. TEQ® offers services in order to help you with this topic.

Basics of the Evaluation Strategy in qs-STAT®

qs-STAT® shows the evaluation strategy in the form of a flow chart. There are a few options regarding the preparation of the evaluation but then the evaluation strategy is all about defining a reasonable distribution model that describes the data as correctly as possible. The available methods range from "best fit" (i.e. the main point is that the model adapts optimally to the data) to the calculation of distribution time models in accordance with ISO 22514-2.

Due to the distribution time models, the program classifies the processes automatically into the main groups A to D and sometimes even into the subgroups A1/2 and

Figure 2: Distribution time models
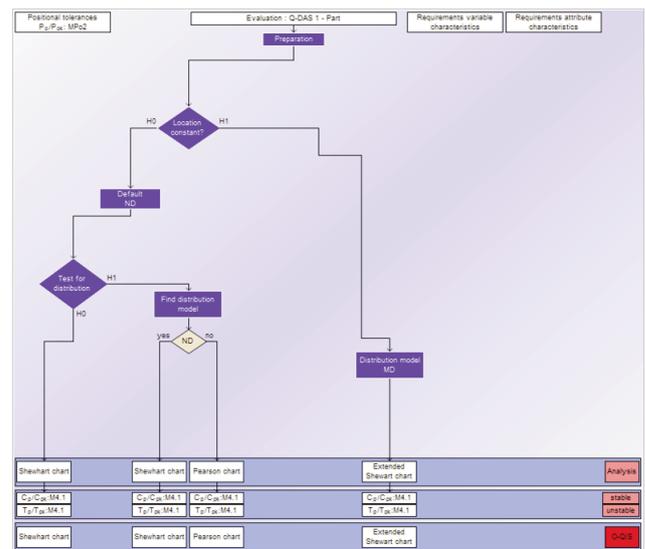
C1 to C4. This classification helps you detecting instabilities and risks immediately. According to ISO, process models B, C and D are considered to be "unstable" processes. However, do not confuse this stability assessment with the stability analysis based on a control chart as described below.

One of the disadvantages of the "best fit" method is that the process is assigned quite randomly – i.e. based on statistical procedures only - to a distribution that currently fits. You can obtain a better and more stable selection by trying to use "reasonable" and "expected" distributions first. The next step is to search for the "best fit" in case the expected distribution does not apply. In the easiest case, you start with a test for normality. The desired distribution always depends on the size of the characteristics in complex configurations. It is assumed that e.g. in case of linear measures the normal distribution shall be first choice, whereas it is the folded normal distribution in case of roundness and the Rayleigh distribution for concentricity. When the expected distribution is challenged for statistical reasons, you search for a suitable distribution by using the "best fit" method.

Figure 3: "Best fit"
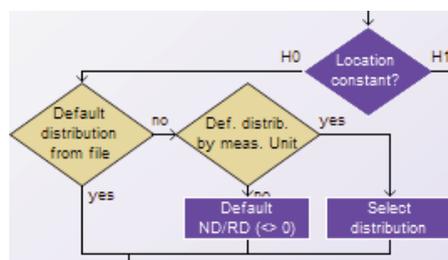
Figure 4: Desired distribution

The decision which distribution fits best depends on many factors. It starts with the question as to which distributions are permissible and reasonable at all from a corporate perspective. The program determines the ideal parameters for these distributions according to

Figure 5: Searching for a distribution

Figure 7: Requirements

strategies that need to be defined. Afterwards you have to select the criteria on which the test for best fit is based. On rare occasions, you might have to deal with a tough nut to crack demanding for an alternative strategy. Anyway, the program always decides in favor of a specific distribution model in the end.

Most companies adopt these decisions and document the results in the report. Others create a kind of "emergency break" they put on when the distribution model does not meet the expectations. This "emergency break" stops qs-STAT® without any further decisions so that the user has to interact.

As soon as the program finds the suitable distribution, you have to evaluate the stability in a quality control chart. The program calculates a retrospective analysis control chart corresponding to the found distribution (time) model. At the same time, qs-STAT® also proposes a SPC control chart on site that you may adopt as necessary.
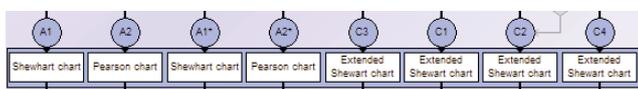
Figure 6: Distribution time models and control charts

Based on this information about distribution and stability, we can start the process capability analysis. There are numerous methods available in order to calculate capabilities. Select the method in qs-STAT® that seems to be "correct" according to your regulations. Broadly speaking, we may distinguish between methods in accordance with DIN/ISO VDA that are used as decision criteria in almost every corporate guideline and so-called "overall" or "within" capabilities according to AIAG Core Tools. You also define the designation of capability
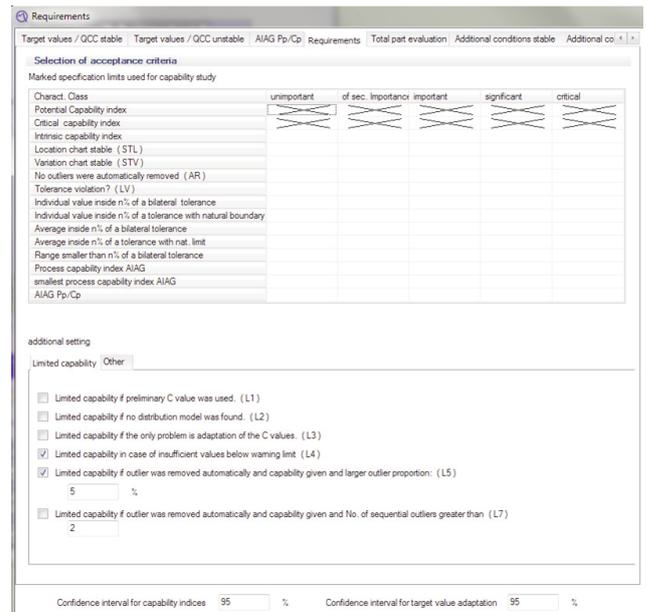
indices in the requirements of the evaluation strategy. It is up to you whether you select C_m / C_{mk}, P_m / P_{mk}, P_p / P_{pk}, C_p / C_{pk}, T_p / T_{pk} or choose any other exotic name. The corresponding criteria might be number of samples, number of values or stability criteria. You can even consider some boundary conditions. How do you evaluate e.g. natural limits, insufficient values or automatically removed outliers? Do you adopt the limits to the number of values, and if so, what is the respective level of confidence? Would you like to make total part evaluations in addition to the characteristic evaluation? If so, what are the grading scale and the characteristic class weights you want to apply ?

There are many questions but luckily you have to answer them just once in order to tap the magic keys <F9> and <F10> to achieve the correct result in a split second. As explained at the beginning, business before pleasure! And as mentioned before, TEQ® is pleased to help you with your business so that you do not have to wait for the pleasure of your own evaluations too long.
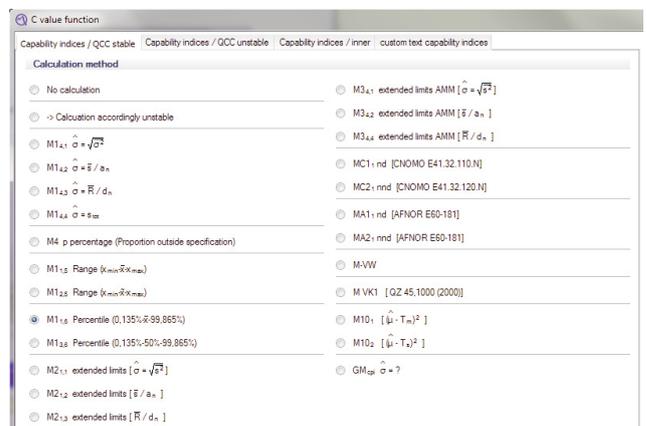
Figure 8: Calculation methods for capability indices
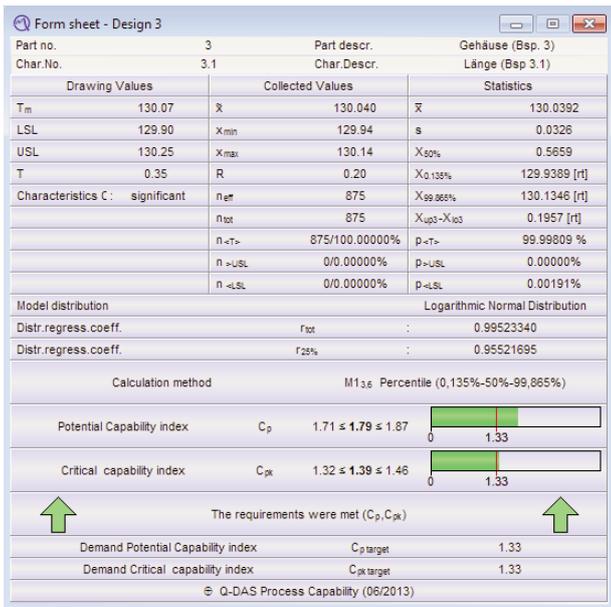
Figure 9: Form sheet Design 3 (<F10> key)

But stop – there is still something to explain. When you use the SPC control chart on site in procella® or O-QIS, you also have to define warning and alarm criteria. Based on these criteria, the program informs operators about difficult cases. However, for users who apply nothing but qs-STAT® this is just a voluntary exercise but not an obligation.

And What about solara.MP …?

After opening the evaluation strategy in solara.MP, you will also see a flow chart first. However, if you have a closer look, you will find out that it is only a list of single procedures some of which are formally arranged in the form of a flow chart. Technically speaking, you only activate the procedures of the measurement system analysis that shall be available to the user. On a deeper level, you have to define the rules applying to the recording of data, the type of evaluation and the requirements the results have to meet.
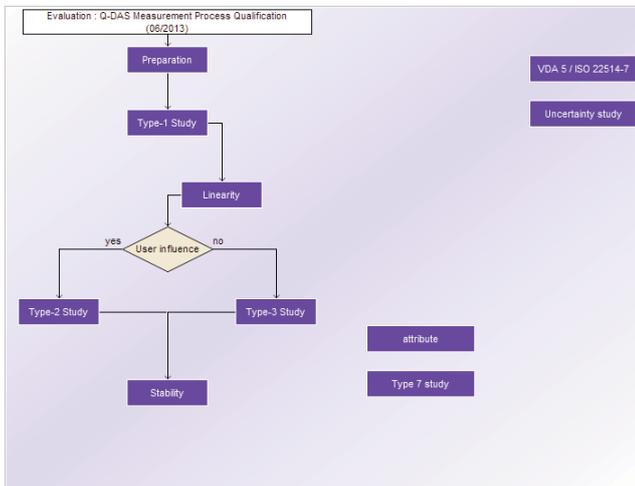
Figure 10: Evaluation strategy in solara.MP

I have to admit that this leads to the infinite depth of the statistical outer space again, of course. Thus you have to consider carefully how to parameterize all these procedures. Even in this case, many users do the "same" thing which is far from doing the "identical" thing. Please select whether you want to calculate type-1 study with "4s or 6s", whether you conduct type-2 study based on ANOVA (which actually should be the case) and whether you want to perform tests of significance in these studies. Especially type-2 study leads to spirited discussions about whether the ndc should be a requirement (see PIQ® International Issue 2013 "The Crux of the ndc"). And this is how it continues…
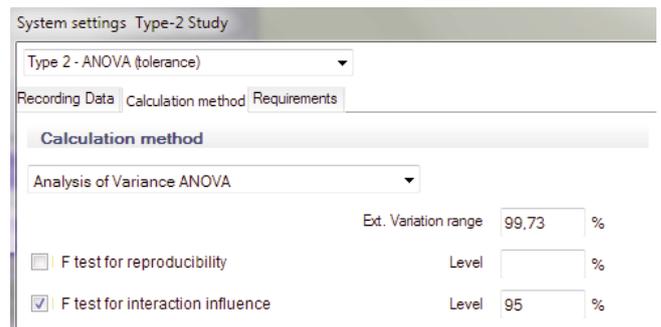
Figure 11: Recording data / Calculation method / requirements

The three options at the far right (figure 12) are of particular importance. Here you may allow for the calculation of the measurement uncertainty in accordance with VDA Volume 5 (2nd edition or ISO 22514-7). Right below is the "Uncertainty study" option sounding like a very general approach but actually representing the all-in-one version of this study as given in Daimler's "Leitfaden 05". The last option is "measurement uncertainty" and refers to the previous and 1st edition of VDA 5. Only activate this option in case you inevitably have to evaluate historical VDA volume 5 studies.

Do you have any questions? The author will be pleased to answer them. Send an e-mail to

stephan.conrad@teq.de .
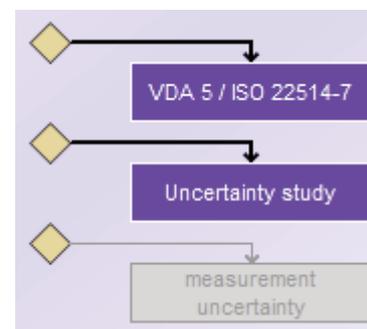
Figure 12: VDA Volume 5 and others