

## Perceived Truths, Risks and Statistical Tests

Roman Wenig, Frank Stockhaus,  
TEQ® Training & Consulting GmbH

Our modern world is no longer imaginable without electronic data processing. You generate information from data. These pieces of information help you make decisions. We virtually face the decision-making process everywhere and permanently. Decisions control our processes and create our services and material products. However, this is not just about making decisions; it is also about how good these decisions are. You reach a decision on the basis of your experience or your personal assessment of the respective situation, but you may also come to a decision by applying a well-structured method based on hard figures, data and facts.

For economic reasons, companies normally control processes based on sample inspections, especially in industrial production. This fact compels us to accept risks. A sample inspection always incurs the risk to make a wrong decision. In order to calculate this risk, you necessarily have to master statistical methods. The basic methods include statistical tests.

There is currently no simple method dealing with the variation of our processes in an elegant and easier way but statistical tests. These tests offer a binary result provided with a calculated risk. Statistical tests separate chance from changes affecting a process systematically, they distinguish between signal and noise, they sort the wheat from the chaff. It is not possible to imagine a single step of the product life cycle without these tests. So they should be a standard tool for anyone making decisions based on figures, data and facts.

In order that the statistical test provides the desired result reliably, users have to master the methods on the one hand and must even meet certain requirements for the preparation, implementation and finishing process on the other hand. The following article describes this process step by step for parameter tests, i.e. tests determining whether the parameters of the population are significant. Each step includes specific questions that you have to answer and tasks you have to perform before taking the next step. Finally, an example illustrates this approach.

### 1. Which probability distribution best describes the current situation?

- Find out about the following aspects:
  - o Do you have a sample? Does your sample describe the situation “reliably”?
  - o Do you work with discrete or continuous data?
- In case of discrete data, what is the suitable distribution model? (The probability distribution and its parameters depend on this model.)
  - o Number of defective units - binomial distribution
  - o Number of defects per unit – Poisson distribution
- In case of continuous data, test whether the normal distribution model describes the data.

### 2. Which parameter of the distribution do you want to test?

- What do you want to know? What do you want to know about your sample? How much information do you have about the population – maybe you even know its variation?
- Find out about the following aspects:
  - o Do you want to test a shift parameter or
  - o a parameter of variation or
  - o do you want to test both parameters or
  - o just an error rate?
- Make the data analysis. Calculate the statistics for the shift parameter and parameter of variation or the error rate based on your sample.

### 3. Which kind of comparison do you want to make?

- Define whether you want to
  - o make a nominal-actual value comparison (sample vs. customer requirement),
  - o compare two actual values (two samples) or
  - o draw a comparison between several samples.

### 4. Which hypotheses do you have to formulate?

- Formulate a null and an alternative hypothesis.
  - o The null hypothesis  $H_0$  describes that nothing has changed. The expected signal disappears in the noise (i.e. in the variation or confidence interval of the parameter to be tested).
  - o The alternative hypothesis  $H_1$  is always opposed to the null hypothesis  $H_0$ . ( $H_1$  always reflects what we are attempting to demonstrate!)

- o The parameter  $x$  either shows a significant deviation from  $y$  (i.e. it is unequal to  $y$ ), is significantly less than  $y$  or is significantly greater than  $y$ .

**5. Which test should you choose?**

- Select the right statistical test. Statistical software will help you find a suitable test since they provide decision trees supporting you in gaining the appropriate test.
- Define the error probability  $\alpha$ .  $\alpha$  usually amounts to 5%, 1% or 0.1%.
- Conduct the statistical test.

**6. What is the test result?**

- The test result also provides the P-value. Compare this value to the error probability  $\alpha$ .
- Alternative: The test result also provides the test statistic. Compare this statistic to the critical quantile.
- Based on this comparison, draw a conclusion about the “statistically formal” decision regarding the null hypothesis.
  - o  $P \geq \alpha$ : The null hypothesis is not rejected.
  - o  $P < \alpha$ : The null hypothesis is rejected in favor of the alternative hypothesis. You have to accept the alternative hypothesis.

**7. How does the test result refer to the real situation?**

- Formulate an answer describing the test result as follows: “The test result means that... (the test parameter is significant / is not significant or the result of the data analysis is random / is not random).”
- Transfer the test result to your real situation.

**8. What is the power of the test? How “good” is it?**

- Find out whether the test result is plausible depending on
  - o the sample size,
  - o error probability  $\alpha$ ,
  - o the probability of a type II error or power of the test
  - o a “technologically reasonable” shift in location that can be proved, a “technologically reasonable” change in variation or a “technologically reasonable” change in the error rate.

These eight steps provide a sort of guidance for performing statistical tests and shall help you get along with it. The following example illustrates these steps. The software qs-STAT® shows the required statistical evaluations.

For the width of a keyway, all calculated  $C_p$ -values and  $C_{pk}$ -values only amounted to about 1. These values did not meet the requirement of at least 1.33. In order to reach better values, i.e. higher values, a reduction of the variation was required. The responsible process engineer immediately had an idea of how to improve the process and implemented it. The next 20 parts were produced under modified process conditions and measured afterwards.

Graphics, such as value chart, histogram or box plot frequently visualized the distribution of continuous characteristics. Process engineer, machine operator and production manager quickly reached an agreement in this example: The box plot in Figure 1 clearly showed that the corrective action reduced the variation of the characteristic considerably!

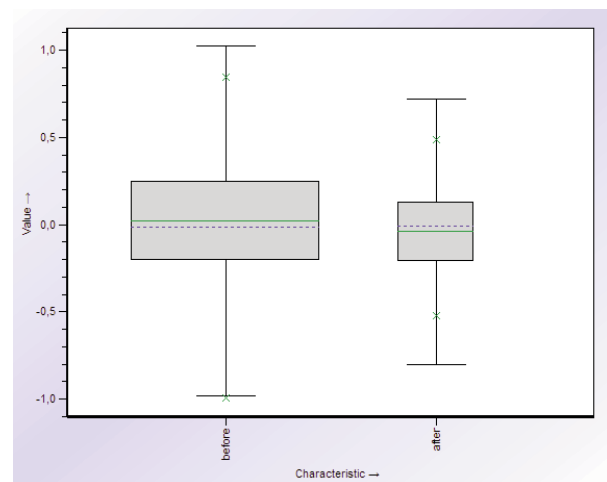


Figure 1: Box plot showing the characteristic before and after taking corrective action

Even the numerical values of the standard deviations from both samples proved that the process variation was less now, see Figure 2. So the standard deviation was reduced and only amounted to 76% of the original standard deviation.

The process engineer had a good feeling and asked for a new process capability analysis based on a total of 125 parts. After several shifts had done their work, the new result was available – but suddenly all people involved were a bit confused when hearing about the recalculated values.  $C_p = 0.99$  and  $C_{pk} = 0.95$ .

Part no.	Part descr.	swivel	
Char.No.	Char.Descr.	$\bar{x}$	s
1	before	10.002222	0.0334
2	after	9.995945	0.0254

Figure 2: Average  $\bar{x}$  and standard deviation  $s$  before and after taking corrective action

### What happened?

When dealing with the “perceived truths” mentioned above, the engineer, operator and manager ignored two important aspects:

- The standard deviations  $s_{\text{before}}$  and  $s_{\text{after}}$  calculated from the two samples are just estimates of the true but unknown standard deviations  $\sigma$  of the two populations. Based on the fact alone that the standard deviation  $s_{\text{after}}$  is less than  $s_{\text{before}}$  you cannot suggest that this circumstance also complies with the population. The calculated deviation might even be pure chance!
- The **sample size** after taking corrective action is relatively **small**, it just includes 20 parts. In general, the following rule applies. The smaller the sample size, the higher the uncertainty in the range in which the true standard deviation of the population is expected.

Only a statistical test can evaluate whether the corrective action led to a significantly reduced standard deviation  $\sigma$  of the population; only such a test provides an objective answer to this question.

In order to conduct a statistical test, we take the eight steps mentioned above.

#### 1. Which probability distribution best describes the current situation?

The normal distribution describes the distribution of the continuous characteristic “width of the keyway”. The qs-STAT® software defines normal distribution as suitable model for both samples. This histogram shown in Figure 3 confirms this result.

#### 2. Which parameter of the distribution do you want to test?

We want to test whether the parameter of the standard deviation was reduced due to the measure taken by the process engineer.

#### 3. Which kind of comparison do you want to make?

We have to compare two populations with respect to their standard deviation. So we need a test comparing two actual values. It is also referred to as 2-sample test.

#### 4. Which hypothesis do you have to formulate?

Mathematically speaking, the null hypothesis  $H_0$  and alternative hypothesis  $H_1$  are as follows:

$$H_0 : \sigma_{\text{before}}^2 \leq \sigma_{\text{after}}^2$$

$$H_1 : \sigma_{\text{before}}^2 > \sigma_{\text{after}}^2$$

With regard to the contents of these hypotheses, the test shall confirm that the variation after taking corrective action  $\sigma_{\text{after}}^2$  is less than the variation  $\sigma_{\text{before}}^2$  before taking these measures. This variation is often also referred to as variance.

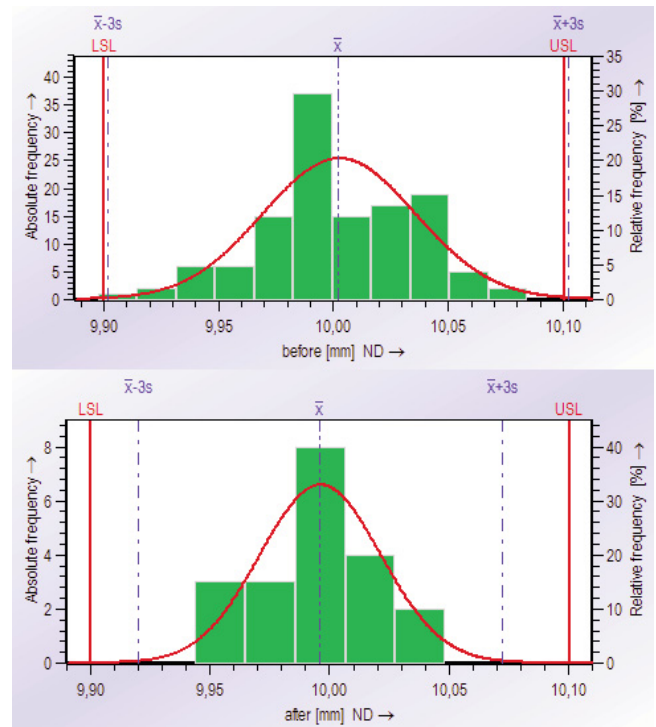


Figure 3: Histogram before and after taking corrective action

### 5. Which test should you choose?

Experts in statistics already know which test to choose after taking steps 1 to 4. It is the “two-sample F-test”.

But even if you are not an expert in statistics, the assistant for test procedures in the Sample Analysis module of qs-STAT® will help you find the suitable test. A tree diagram facilitates the selection of an adequate test. Click on the respective box to reach the next subordinate level. This is how you will find the right test step by step. Figure 4 shows the result for this example.

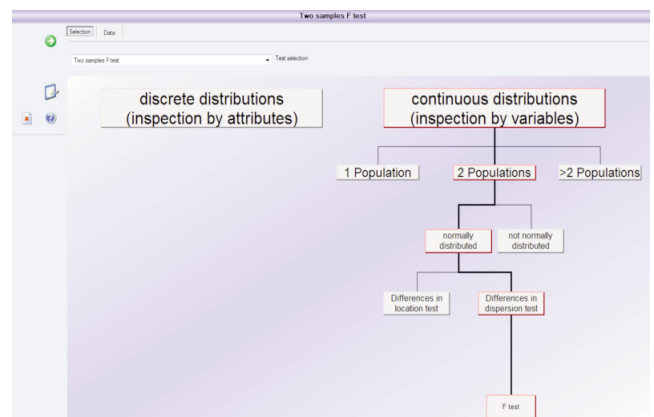


Figure 4: Selection of the two-sample F-test

### 6. What is the test result?

The test procedure assistant in qs-STAT® shows the test result in the “Test” tab (Figure 5).

Note: You select the desired hypotheses (see step 4) in the lower left-hand corner of the window!

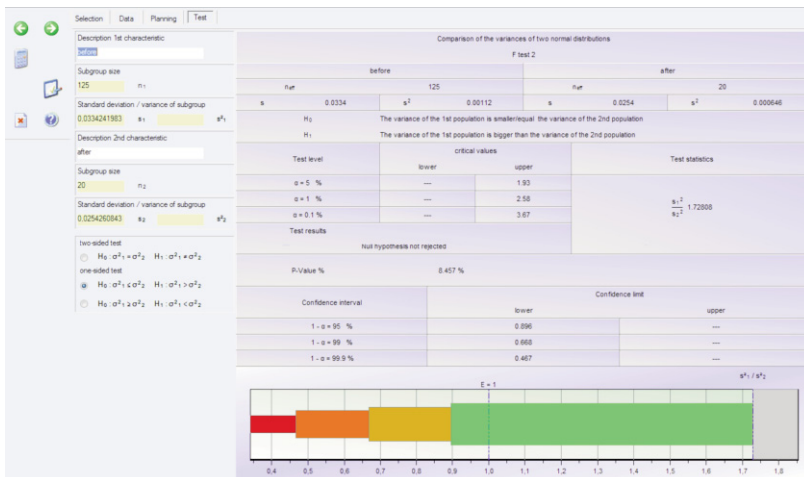


Figure 5: Selection of hypotheses and test results

An elegant way to evaluate the test result is the comparison of the P-value to the specified error probability  $\alpha$ . This example applies the error probability  $\alpha = 5\%$ . Figure 6 shows that the P-value amounts to 8.5%, i.e. this value is greater than the error probability  $\alpha = 5\%$ . So the null hypothesis is not rejected. The “test result” field also puts this result in words.

Comparison of the variances of two normal distributions				
F test 2				
before			after	
n	s	s <sup>2</sup>	n	s <sup>2</sup>
125	0.0334	0.00112	20	0.000646
H <sub>0</sub> : The variance of the 1st population is smaller/equal the variance of the 2nd population H <sub>1</sub> : The variance of the 1st population is bigger than the variance of the 2nd population				
Test level	critical values		Test statistics	
	lower	upper		
$\alpha = 5\%$	---	1.93	$\frac{s_1^2}{s_2^2}$ 1.72808	
$\alpha = 1\%$	---	2.58		
$\alpha = 0.1\%$	---	3.67		
Test results				
Null hypothesis not rejected				
P-Value %	8.457 %			

Figure 6: Test result (extract from Figure 5)

Please consider that this step just provides a “statistically formal” decision regarding the null hypothesis. There are only two possible outcomes.

- The null hypothesis is rejected or
- the null hypothesis is not rejected.

## 7. How does the test result refer to the real situation?

The result achieved in step 6 means that the program did not identify any significant reduction of the variation.

In other words, it is pure coincidence that the value of the standard deviation in the sample reached after taking corrective action ( $s_{\text{after}} = 0.0254$  mm) is less than the standard deviation of the sample before taking these measures ( $s_{\text{before}} = 0.0334$  mm).

Eventually, the “perceived truth” of an assumed improvement, as considered by the persons involved based on the sample data, was not confirmed by the objective statistical test!



## ... accurate, robust and reliable

The new MG50 multichannel interface for incremental gauges with EtherCAT for fast data transfer.

### MG50

- Up to 30 gauges connectable
- EtherCAT interface
- Compatible with gauges of DF and DT series

### DF Series

- 60 mil. strokes achieved
- Measuring range: 5 mm / 12 mm
- Max. resolution: 0.1  $\mu\text{m}$
- Accuracy:  $\pm 0.5 \mu\text{m}$
- Vacuum retraction

The optimal combination for fast data transfer in production and assembling lines.

## Contact us for more information!

Magnescale Europe GmbH  
 Tel. +49-(0)7153-934-291  
 info-eu@magnescale.com  
 www.magnescale.com



**8. What is the power of the test? How “good” is it?**

The “power” of a statistical test depends on the probability of possibly reaching a wrong decision. The correlations are complicated and we will not go into too much detail. We just want to mention the fact that, from a statistical perspective, you can make two different types of wrong decisions:

- You reject the null hypothesis even though it is correct (type I error or  $\alpha$ -error).
- You do not reject the null hypothesis even though it is not correct (type II error or  $\beta$ -error).

The “Planning” tab in the qs-STAT® software shows the operating characteristic curve. It refers to  $\beta$ , the probability of making the type II error. In this example, this probability depends on the variations in the populations

$$\frac{\sigma_{\text{before}}}{\sigma_{\text{after}}}$$

and is applied as a function of this ratio.

Now you have to specify a value for this ratio, determine the probability  $\beta$  or the power of the statistical test  $1 - \beta$  and calculate the minimum sample size.

In this example, the company specified

$$\frac{\sigma_{\text{before}}}{\sigma_{\text{after}}} = 1,5 \quad \text{and} \quad \beta = 10 \% \quad \text{or} \quad 1 - \beta = 90 \%$$

These specifications define that the test shall identify the reduction to  $2/3 \approx 67\%$  of the variation with a probability of 90%. Figure 7 shows the respective operating characteristic curve in the form of a red line. The required sample size amounts to at least 54 parts.

However, the project engineer only drew a sample of 20 parts in this example. What are the consequences? A green line illustrates the operating characteristic curve in Figure 7.

The x-coordinate  $\frac{\sigma_{\text{before}}}{\sigma_{\text{after}}} = 1,5$

leads to a y-coordinate representing a probability of about 47%, i.e. a reduced standard deviation of about 67% leads to the acceptance of the null hypothesis with a probability of 47% ( $\beta$ ). Based on this situation, the reduction of the standard deviation was not identified! In other words, the probability of identifying a standard deviation reduced to 67% amounts to 53% ( $1 - \beta$ ), so the test is not likely to detect this reduction!

Now, we reach the conclusion that the sample size of  $n = 20$  parts selected by the operator was too small to identify the respective reduction of the standard deviation in a statistical test with certainty.

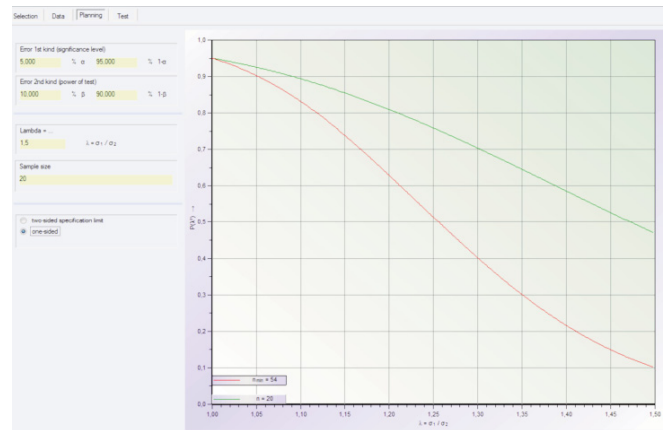


Figure 7: Operating characteristic curves for the two-sample F-test

**Conclusion**

Statistical tests are an essential tool in inferential statistics. They are part and parcel of many applied statistical procedures, e.g. measurement system analysis, statistical process control SPC or design of experiments DoE. You cannot even do without statistical tests in Six Sigma projects and when deciding about measures for process improvement.

This article shall provide an insight into the application of statistical tests and the corresponding opportunities. We did not intend to discuss all aspects regarding the application of statistical tests in detail.

We hope this article piqued your interest in this topic. In this case, we recommend you to attend one of our seminars on statistical procedures. These seminars provide you with a basic understanding of statistical procedures and corresponding mathematical relationships, but they also focus on a certain variety of graphics and diagrams available to illustrate these tests. The Microsoft® Excel standard software helps you apply methods requiring lots of computational effort. Please contact the German training and consulting company TEQ® or the respective Q-DAS® subsidiary for information about available trainings.

In case of on-site trainings you may decide whether you want to use the commercial statistical packages of qs-STAT® or MINITAB®.

Do you have any questions? Please feel free to contact the authors. Send an e-mail to [frank.stockhaus@teq.de](mailto:frank.stockhaus@teq.de) or [roman.wenig@teq.de](mailto:roman.wenig@teq.de).