Introduction

In a hypothesis test, a random sample should be used for making a statement about the property of the population.

Purpose and benefit

Making a decision on the basis of a defined "allowed" error probability.

Basics

A so-called null hypothesis and an alternative hypothesis are defined:

- The **null hypothesis** H_o maintains, for example:
  There is equality or the average values of two random samples are the same, or the average value of a batch corresponds to the specification, or the consumption of two vehicles is the same, etc.

- The **alternative hypothesis** H_1 maintains, for example:
  There is a difference, or the average values of two random samples are unequal, or the delivered parts are faulty, etc.

Two kinds of errors can be made when carrying out a statistical test:

1. The null hypothesis H_o is correct, but is not accepted!

   ⇨ This error is referred to as a 1st order error, or the α error, or the producer's risk.

2. The null hypothesis H_o is accepted although it is incorrect!

   ⇨ This error is referred to as a 2nd order error, or the β error, or the consumer's risk.

Overall, the following four situations exist:

		Reality	
		H_o	H_1
Decision	H_o	correct	β error (2nd order)
	H_1	α error (1st order)	Correct

H_o : Null hypothesis; H_1 or H_1 : Alternative hypothesis

The α error describes the risk of assuming an "effect" that does not actually exist. - The β error describes the risk of overseeing an "effect".

Statistical hypothesis tests

> **Note:**
>
> The actually calculated error probability for the 1st order error is referred to as the p_{value}. The defined error probability $\square$ as the basis for decision is generally $0.05 = 5\%$.
>
> ⇨ The null hypothesis is <u>accepted</u> if $\quad p_{value} \geq \alpha$
>
> ⇨ The null hypothesis is <u>not accepted</u> if $\quad p_{value} < \alpha$

Example 1: Test for normal distribution

Many test processes and statistical calculations are only to be used if the data has a normal distribution (distributions are symmetrical and randomly distributed around the mean – see section ***The normal distribution*** in the appendix).

The null hypothesis H_0 states: The data has a normal distribution.

There is a large number of test processes. The basic principle of all test processes is the evaluation of deviations between the observed frequencies and the expected ones.
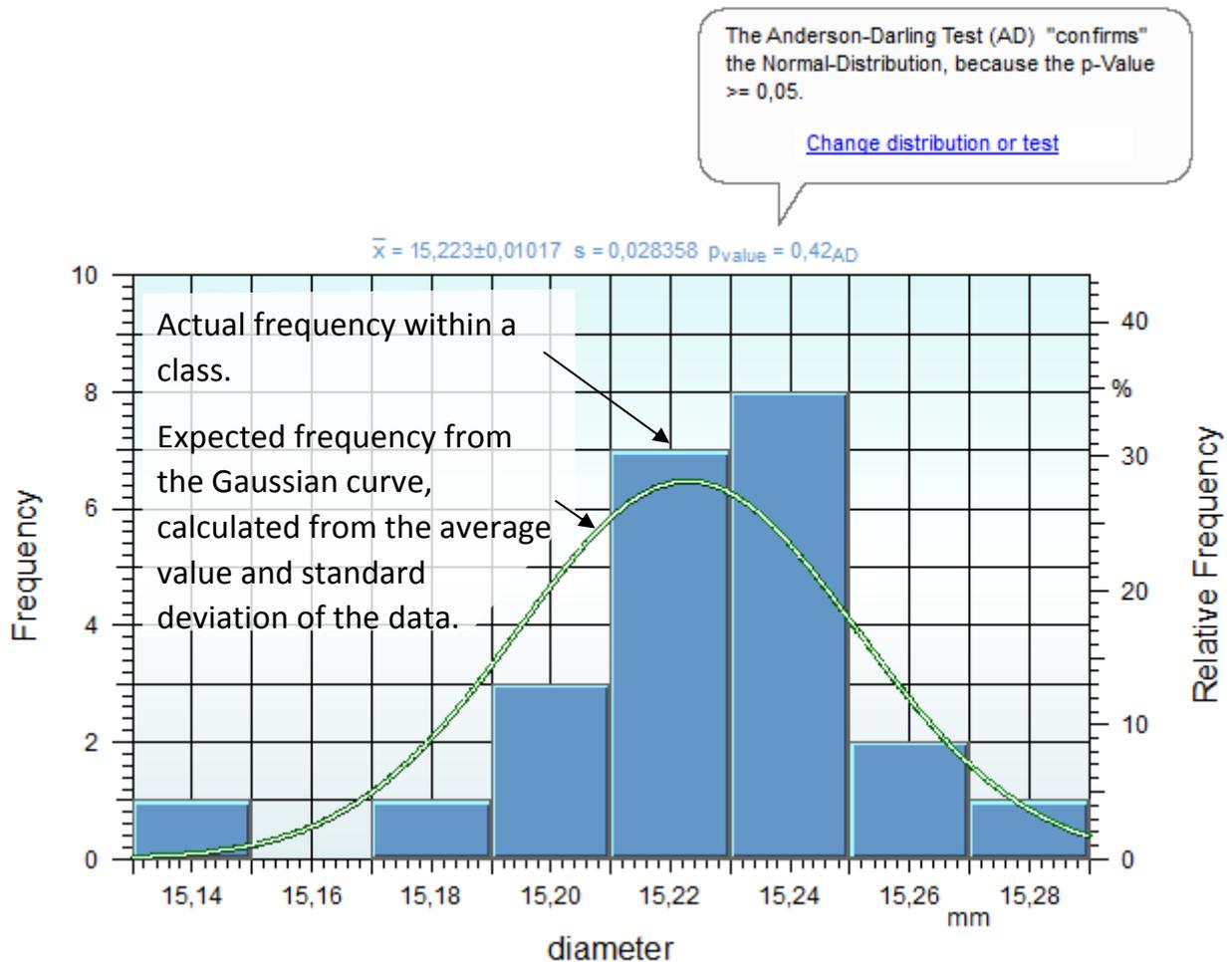
The most important tests are stated below. Each has its specific advantages and disadvantages (for more information, refer to the statistical literature).

- Shapiro-Wilk (SW) $n < 50$, standard in German-speaking countries.

- χ^2-test (CH)   from $n \geq 50$, classification of data required.

- Anderson-Darling (AD)   Standard in Minitab & Six Sigma.

- Kolmogorov-Smirnov (KS)   Classic, possible with and without classification.

- Cramér von Mises (CM)   Relatively highly dependent on the random sample size.

- Epps-Pulley (EP)   Less widespread, tolerant against deviations.

- Jarque-Bera test (JB)   Less widespread, very tolerant against deviations.

The Anderson-Darling and the KS test are recommended. It is a good idea to have random samples of up to 100 at most. Beyond that, even slight deviations almost always result in a rejection of the normal distribution.

Example:

23 shafts were measured during turning. In spite of the skewed distribution, the Anderson-Darling test results in a error probability of $p_{value} = 0.42$. This is greater than $\alpha = 0.05$, as the defined "significance limit". The normal distribution is accepted.

The Anderson-Darling Test (AD) "confirms" the Normal-Distribution, because the p-Value >= 0,05.

Change distribution or test

$\bar{x} = 15{,}223 \pm 0{,}01017 \quad s = 0{,}028358 \quad p_{value} = 0{,}42_{AD}$

Actual frequency within a class.

Expected frequency from the Gaussian curve, calculated from the average value and standard deviation of the data.

Note:

It should be considered that all tests are only sensible if the data is also to be accepted as having a normal distribution from a "technical" perspective. For example, no normal distribution is to be expected from running times in service life tests and from the field, but rather a Weibull distribution (see appendix).

It is important to know about the technical context and the expected distribution! Analyzing data of unknown origin and circumstances is highly inadvisable!
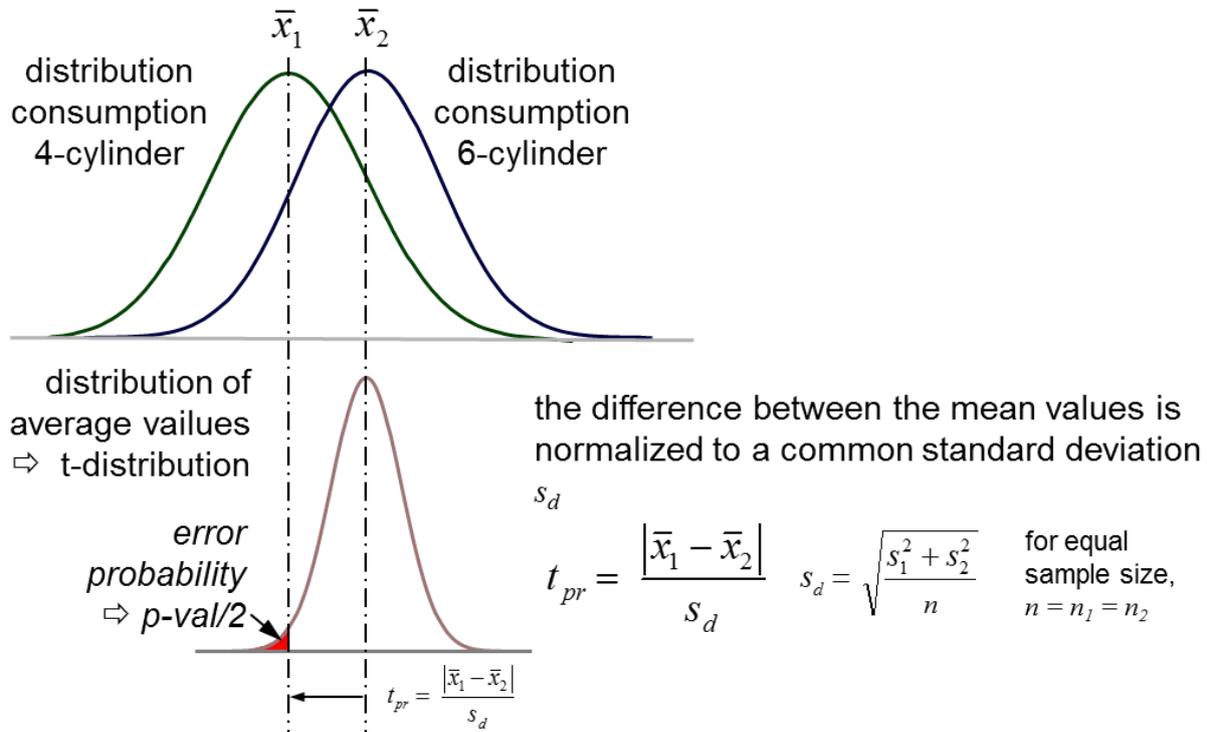
Example 2:
Test for the same average values (t-test)

In the t-test, measured values are compared between two groups. The null hypothesis is established that the average values are the same. This is defined even if you are actually looking for a difference. The null hypothesis is rejected if the effect is not sufficiently pronounced due to the distribution.

Statistical hypothesis tests

The precondition for using the t-test is that the data must have a normal distribution (see previous example). Differences in the distributions (uneven standard deviations) are possible, however.

In a comparison between vehicle consumptions, the null hypothesis should be: There is no difference in consumption between 4 and 6-cylinder engines. The distributions show significant differences in the average values, although some engines in the middle do have the same or similar consumptions.

$$\bar{x}_1 \qquad \bar{x}_2$$

distribution
consumption
4-cylinder

distribution
consumption
6-cylinder

distribution of
average vailues
⇨ t-distribution

the difference between the mean values is
normalized to a common standard deviation
s_d

error
probability
⇨ p-val/2

$$t_{pr} = \frac{\left|\bar{x}_1 - \bar{x}_2\right|}{s_d} \qquad s_d = \sqrt{\frac{s_1^2 + s_2^2}{n}} \qquad \begin{array}{l}\text{for equal}\\\text{sample size,}\\ n = n_1 = n_2\end{array}$$

$$t_{pr} = \frac{\left|\bar{x}_1 - \bar{x}_2\right|}{s_d}$$

The deviation from average values follows what is referred to as the Student distribution (or t-distribution). Now, the probability of the calculated test value is compared with a defined error probability of, for example, 5%. The probability of t_{pr} is the p_{value}.

As described initially, the null hypothesis is rejected if the following applies: $p_{value} < \alpha$. In this case:

$\bar{x}_1$ = 5.7 l/100km $\bar{x}_2$ = 6.7 l/100km s_d = 0.315 l/100km

t_{pr} = 3.34

As a result of which the t-distribution produces:

p_{value} = 0.003

The difference between 4 and 6-cylinder engines is thus significant, because 0.003 < 0.05.

Statistical hypothesis tests

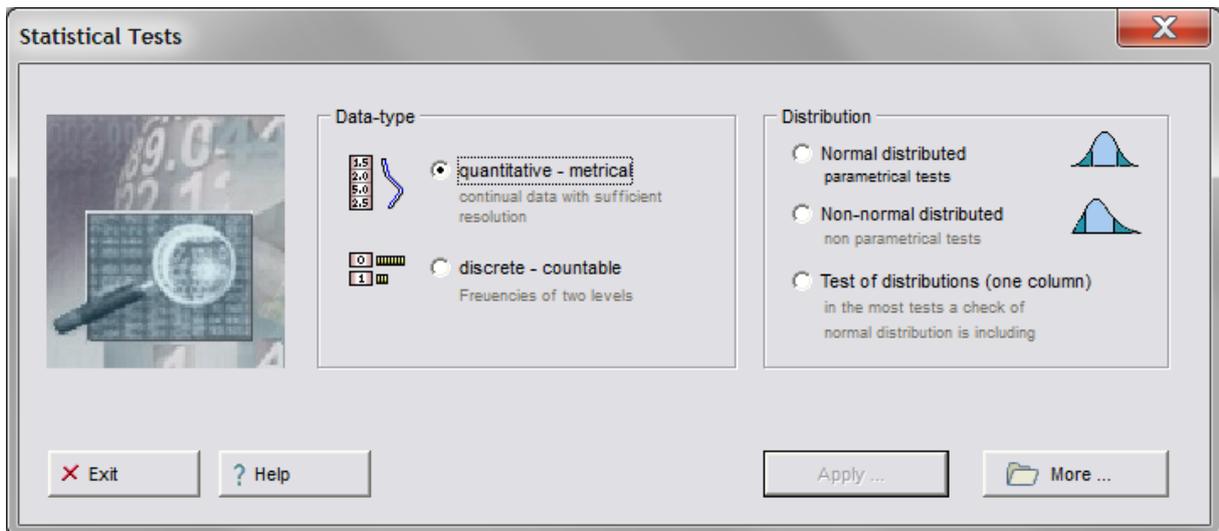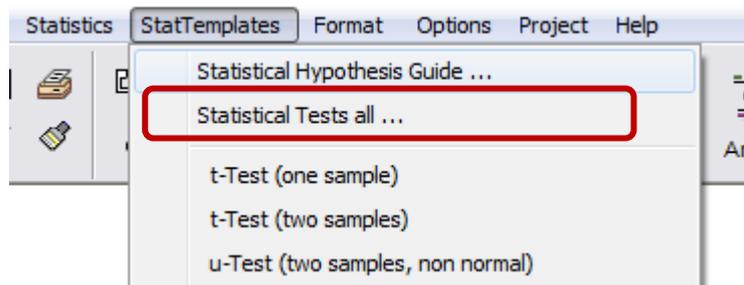| Test	Purpose	Requirement	Template/Menu Visual-XSel
Binomialtest	Comparison of faults to a specification	sample size $n > 9/(p*(1-p))$	Main Guide Statistical Tests *StatTest_Defects_1Sample_to_Specification.vxg*
Chi²-Contingency	Comparison of two groups concerning its frequencies (e.g. failure ratio)	sample size $n > 9/(p*(1-p))$	Main Guide Statistical Tests *StatTest_Defects_2Sample_Contingency.vxg*
Chi²-Homogeneity-test	Comparison of more groups concerning its frequencies	sample size $n > 9/(p*(1-p))$	Main Guide Statistical Tests *StatTest_Chi2Homogen.vxg*
t-Test	Comparison of mean from a measurement with a specification or between to measurements	normally distributed $n \geq 30$	Main Guide Statistical Tests *StatTest_t_1(2)Samples.vxg*
Multi-t-test	Comparison of mean values for each measurement series with those of the other.	normally distributed, $n \geq 30$; independent samples	Main Guide Statistical Tests *StatTest_Multi_t.vxg*
u-test	Comparison of mean values for equality ranking	Non normally distributed $n_1 \geq 4$; $n_2 \geq 4$; $n_1+n_2 \geq 20$	Main Guide Statistical Tests - *StatTest_u_2Samples.vxg*
F-test	Comparison of the variances of two samples for equality.	normally distributed, $n \geq 30$; independent samples	Main Guide Statistical Tests - *StatTest_F.vxg*
Correlation	Detection of a linear relationship between two or more continuous variable.	Continuing and ascending order	Menu Statistics - Correlation …
Linear Regression	Detection of a linear relationship between x and y (independent and dependent var.)	Random deviation	Menu Statistics - Regression…

Statistical hypothesis tests

Using Visual-XSel

Start Visual-XSel and use the icon below..
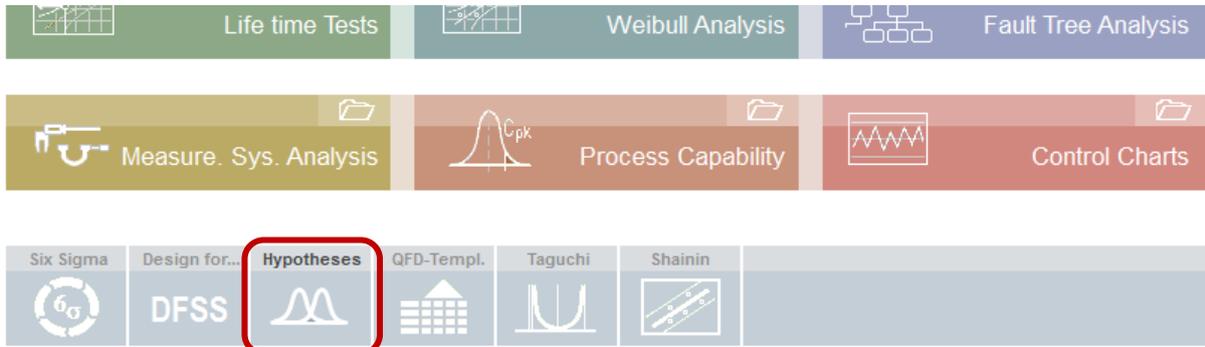
Or use the menu
StatTemplates

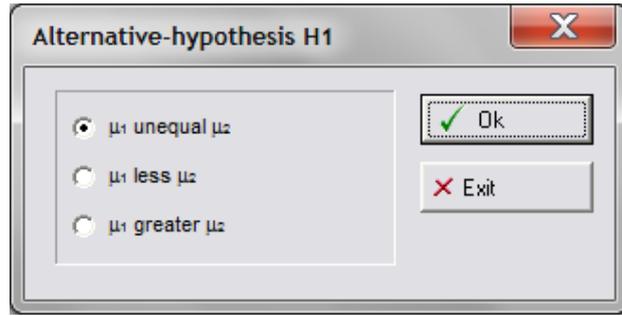Select the wanted data type
and use finally the button
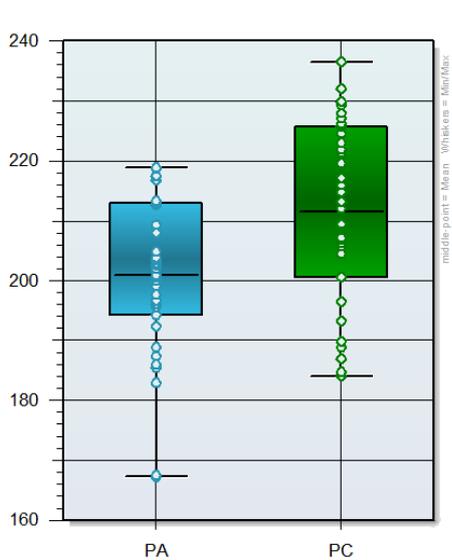*Apply*

Use the link *Paste* for the data from the clipboard and …

…follow the next speech
bubble to start the macro

Statistical hypothesis tests

Select the wanted alternative hypothesis.

Alternative-hypothesis H1

- ● μ₁ unequal μ₂
- ○ μ₁ less μ₂
- ○ μ₁ greater μ₂

✓ Ok

✗ Exit

The results are shown in the main window…

middle point = Mean   Whiskers = Min/Max

PA	PC
167,2	184,1
167,5	184,7
182,9	184,7
185,5	187
186	188,9
187,4	189,9
188,9	193,3
192,4	196,5
194,3	200,6
195,6	204,5
196,1	205,7
196,7	206,2
197,7	206,8
199,1	207,1
200	209,2
200,5	209,5
200,9	212
202	213,2
202,6	214,8
203	215,9
203,5	217,2
203,8	219,6
204,8	221,3
208	221,9

t test	3,11
t crit	1,99
mean m1	200,860
mean m2	211,509
significance (alpha)	0,05
p-value	0,003
beta-risk	0,136
power	0,864

Ho : M1=M2 versus H1 : M1<>M2
Ho is rejected in favor of H1.
The variances are equal

PA	normal
PC	normal

Tip: If p-Value < Significance the null hypothesis that the amples are equal must be rejected.

If the data are not normal distributed, the StatTest_u.vxg has to be used.