Subject

In this tutorial, we show how to use TANAGRA in an analysis of variance problem. We test also homogeneity of variances assumption on the same dataset.

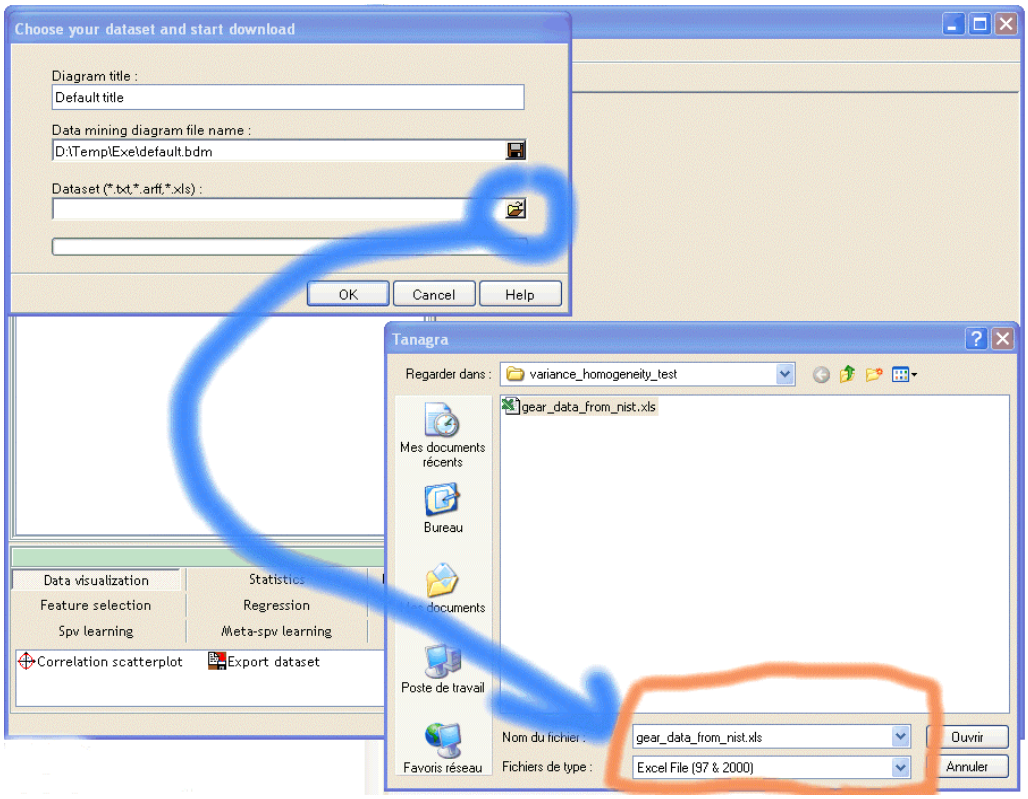
Dataset

We use the GEAR dataset (NIST/SEMATECH e-Handbook of Statistical Methods, http://www.itl.nist.gov/div898/handbook/). We consider that we have 10 machine tools, which produce gears[1]. We have 10 batches of 10 observations. We want to test various assumptions: (1) the average diameter of the gears is the same one for the whole of the machines? (2) The variability of the gear diameter is the same for the whole of the machines?

ANOVA

Download the dataset

The first step is to download the dataset (gear_data_from_nist.xls). In order to create a new diagram, we select the FILE / NEW menu.



[1] http://www.itl.nist.gov/div898/handbook/eda/section3/eda3581.htm

Analysis of variance

We want to know if the diameter of the gears is significantly different from one batch to another with a significance level of 1%, in other words, the machines produce gears with identical characteristics?

We add a DEFINE STATUS component in the diagram; we set GEAR as TARGET (the dependant variable) and BATCH ID as INPUT (the independent / group variable).

Define status 1

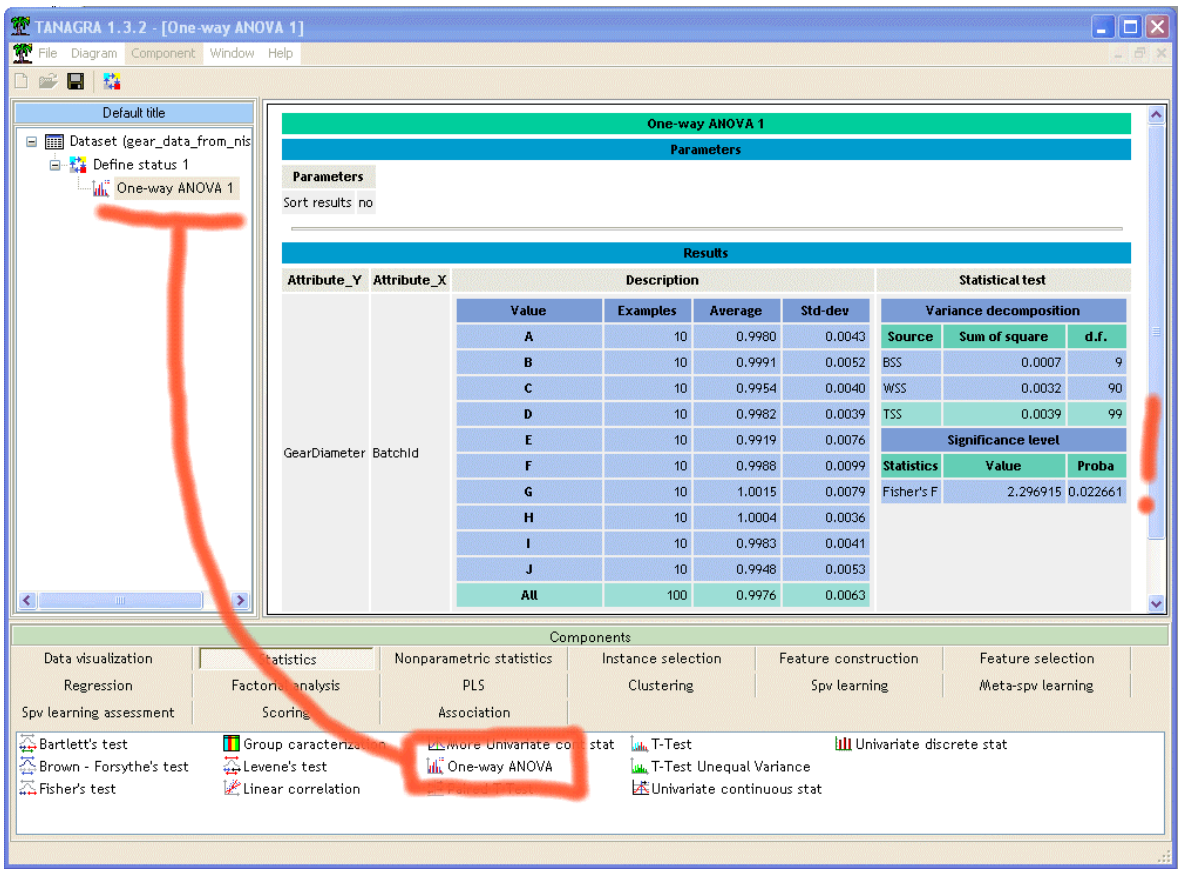
Parameters

Target : 1
Input : 1
Illustrative : 0

Results

Attribute	Target	Input	Illustrative
GearDiameter	yes	-	-
BatchId	-	yes	-

The ONE WAY ANOVA component is available in the STATISTICS components.



For a significance level of 1%, we see that the assumption "the gear diameter is the same for all the machines" is not rejected. The p-value is 2.26%.

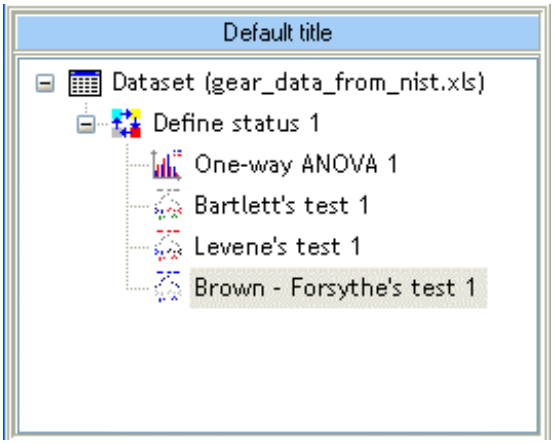
We give below the results of DATAPLOT from the NIST website.

```
                          *****************
                          *  ANOVA TABLE  *
                          *****************

SOURCE              DF SUM OF SQUARES    MEAN SQUARE   F STATISTIC    F CDF SIG
-------------------------------------------------------------------------------
TOTAL (CORRECTED)   99       0.003903       0.000039
-------------------------------------------------------------------------------
FACTOR  1            9       0.000729       0.000081        2.2969  97.734%   *
-------------------------------------------------------------------------------
RESIDUAL            90       0.003174       0.000035

      RESIDUAL    STANDARD DEVIATION =        0.00593857840
      RESIDUAL    DEGREES OF FREEDOM =            90
      REPLICATION STANDARD DEVIATION =        0.00593857747
      REPLICATION DEGREES OF FREEDOM =            90
```

Tests for equality of variances

We want to check that the precision of the manufacturing process is the same. The null hypothesis is "the variance of gear diameter is the same for all the machines".

There are various tests in TANAGRA: BARTLETT, LEVENE and BROWN & FORSYTHE.



Bartlett's test is very sensitive to departures from normality assumption; it must be used with caution. In our dataset, we accept that the variances are equal across machines (practical significance level is 1.36%).

Bartlett's test 1

Parameters

Parameters	
Sort results	no

Results

Attribute_Y	Attribute_X	Description				Statistical test	
		Value	Examples	Average	Std-dev	Test	
GearDiameter	BatchId	A	10	0.9980	0.0043	Pooled var.	0.0000
		B	10	0.9991	0.0052	Bartlett's T	20.7858
		C	10	0.9954	0.0040	df	9
		D	10	0.9982	0.0039	p-value	0.0136
		E	10	0.9919	0.0076		
		F	10	0.9988	0.0099		
		G	10	1.0015	0.0079		
		H	10	1.0004	0.0036		
		I	10	0.9983	0.0041		
		J	10	0.9948	0.0053		
		All	100	0.9976	0.0063		

Levene's test, especially the Brown & Forsythe's alternative, are more robust. These tests are available in TANAGRA.

Levene's test gives the following results.

Levene's test 1

Parameters

Parameters	
Sort results	no

Results

Attribute_Y	Attribute_X	Description				Statistical test	
		Value	Examples	Average	Std-dev	Test	
GearDiameter	BatchId	A	10	0.9980	0.0043	Levene's W	2.159444
		B	10	0.9991	0.0052	df	9/90
		C	10	0.9954	0.0040	p-value	0.032238
		D	10	0.9982	0.0039		
		E	10	0.9919	0.0076		
		F	10	0.9988	0.0099		
		G	10	1.0015	0.0079		
		H	10	1.0004	0.0036		
		I	10	0.9983	0.0041		
		J	10	0.9948	0.0053		
		All	100	0.9976	0.0063		

Brown & Forsythe's test gives the following results.

Brown - Forsythe's test 1

Parameters

Parameters	
Sort results	no

Results

Attribute_Y	Attribute_X	Description				Statistical test	
		Value	Examples	Average	Std-dev	Test	
GearDiameter	BatchId	A	10	0.9980	0.0043	Brown & Forsythe's W	1.705920
		B	10	0.9991	0.0052	df	9/90
		C	10	0.9954	0.0040	p-value	0.099082
		D	10	0.9982	0.0039		
		E	10	0.9919	0.0076		
		F	10	0.9988	0.0099		
		G	10	1.0015	0.0079		
		H	10	1.0004	0.0036		
		I	10	0.9983	0.0041		
		J	10	0.9948	0.0053		
		All	100	0.9976	0.0063		

For this problem, **with a significance level of 1%**, these tests are coherent: the variance of gear diameter is the same for all the machines.