Subject

In this tutorial, we show how to use the ONE WAY MANOVA component (Multivariate Analysis of Variance). We will see that a multivariate test and a combination of univariate tests give a different conclusion.

Dataset

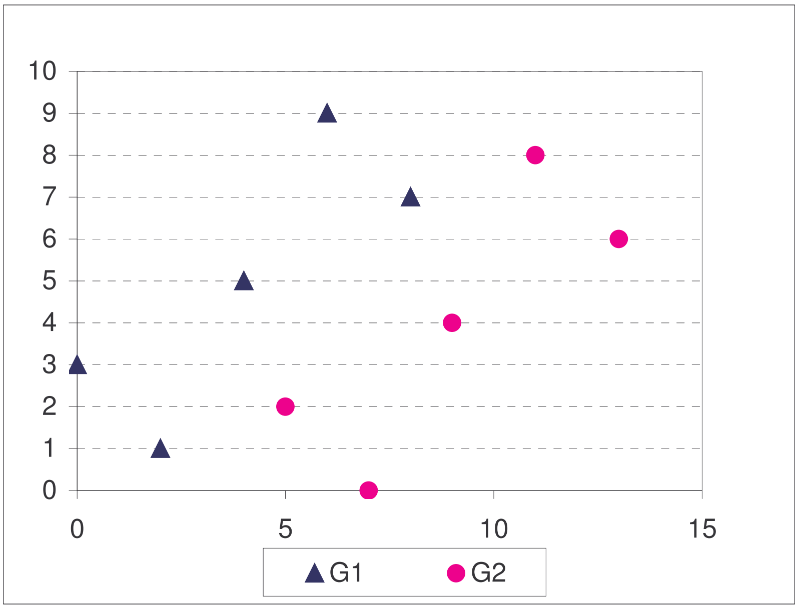
The dataset TOMASSONE_P_29.XLS[1] contains 10 examples with 2 dependent variables (X1, X2) divided into two groups (G1, G2). We want to test

$$\begin{cases} H0: \mu_{g1} = \mu_{g2} \\ H1: \mu_{g1} \neq \mu_{g2} \end{cases}$$

$\mu_{gk} = \begin{pmatrix} \mu_{gk,x1} \\ \mu_{gk,x2} \end{pmatrix}$ is the vector of means of the group Gk.

This is a multivariate generalization of the one-way analysis of variance; we treat two or more dependent variables.

We see below a graphical representation of the dataset.

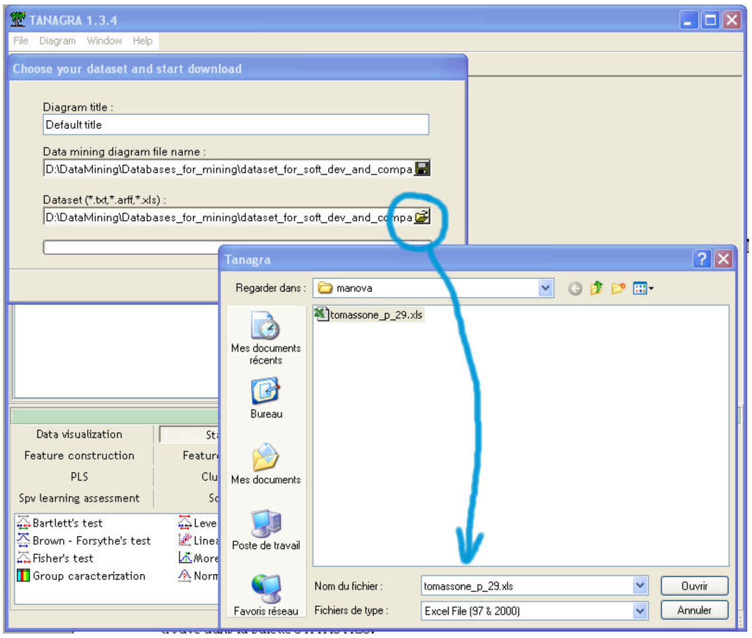


[1] R. Tomassone, M. Danzart, J. Daudin, J. Masson, « Discrimination et Classement », Masson, 1988, page 29.

MANOVA

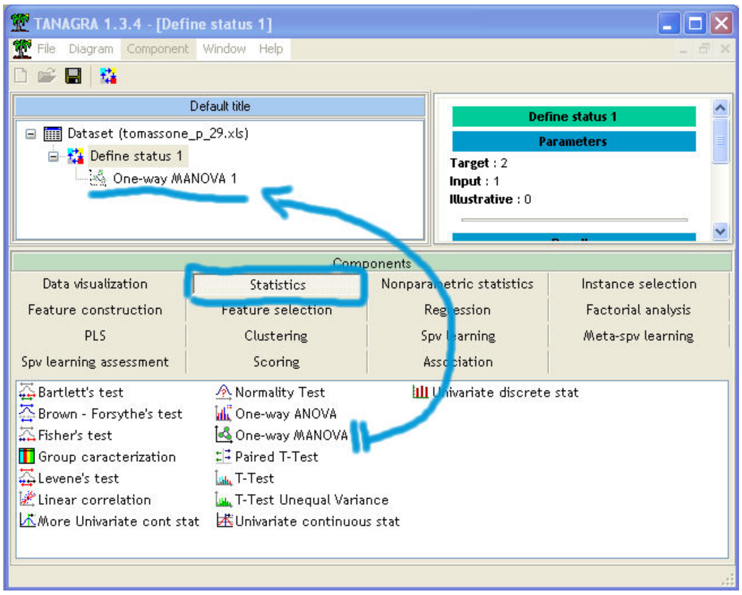
Download the dataset

We build a new diagram and import the dataset.



MANOVA

First of all, we insert a DEFINE STATUS component in the diagram and set (X1, X2) as TARGET, GROUP as INPUT. Next, we insert the ONE WAY MANOVA component.



WILKS' Lambda is the traditional test, the smaller the lambda, the greater the difference between vectors of means. We use two transformation of lambda in order to check the significance of the difference: BARTLETT transformation uses a CHI-2 distribution; RAO transformation follows a FISHER distribution, it is more accurate.

In our dataset, with a significance level of 0.01, the measured difference on the sample does not seem due to chance[2].

One-way MANOVA 1

Parameters

Results

Descriptive stat. (Mean)				Tests results		
Group	**G1**	**G2**	**ALL**	**Stat**	**Value**	**p-value**
Group Size	5	5	10	Wilks' Lambda	0.25308	-
X1	4.0000	9.0000	6.5000	Bartlett -- C(2)	9.61847	0.00815
X2	5.0000	4.0000	4.5000	Rao -- F(2, 7)	10.32986	0.00815

This test assumes that the variances in the different groups are identical. But, it is often not necessary to test this hypothesis because (1) the MANOVA is quite robust and (2) the usual BOX-M test used for testing the homogeneity of variances is not at all robust when the dataset is not normally distributed[3].

Combination of univariate ANOVA

The transformation of the MANOVA test in a combination of univariate test seems attractive, especially because it enables to simplify the computation.

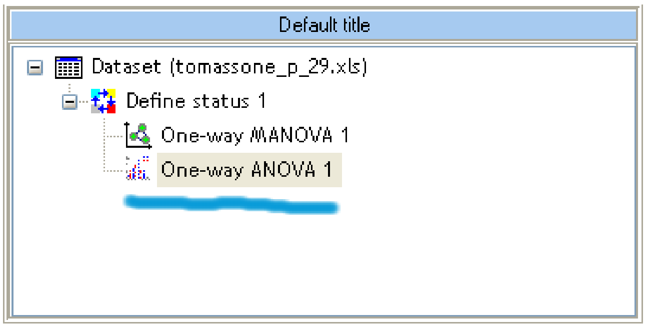
$$\begin{cases} H0: \mu_{g1,xj} = \mu_{g2,xj} \\ H1: \mu_{g1,xj} \neq \mu_{g2,xj} \end{cases}, j = 1, \ldots, J$$

We reject the whole null hypothesis if one the test, at least, rejects their null hypothesis. We see below that this approach can lead to a misleading result because we do not take into account the covariance between the dependant variables.

We insert the ONE-WAY ANOVA in the diagram. TANAGRA performs a univariate ANOVA on each dependant variables.

[2] The two tests have the same p-value; it is an artifact here.

[3] The BOX-M test is a multivariate generalization of the Bartlett's test for equality of variance.



For a significance level of 0.01, we see that the difference between means is not significance on each dependant variable, with this approach we will conclude wrongly that the vectors of means are not significantly different.

Results

Attribute_Y	Attribute_X	Description				Statistical test		
X1	Group	**Value**	**Examples**	**Average**	**Std-dev**	**Variance decomposition**		
		G1	5	4.0000	3.1623	**Source**	**Sum of square**	**d.f.**
		G2	5	9.0000	3.1623	BSS	62.5000	1
		All	10	6.5000	3.9791	WSS	80.0000	8
						TSS	142.5000	9
						Significance level		
						Statistics	**Value**	**Proba**
						Fisher's F	6.250000	0.036942
X2	Group	**Value**	**Examples**	**Average**	**Std-dev**	**Variance decomposition**		
		G1	5	5.0000	3.1623	**Source**	**Sum of square**	**d.f.**
		G2	5	4.0000	3.1623	BSS	2.5000	1
		All	10	4.5000	3.0277	WSS	80.0000	8
						TSS	82.5000	9
						Significance level		
						Statistics	**Value**	**Proba**
						Fisher's F	0.250000	0.630536