

Objectif

Dans ce didacticiel, nous montrons l'utilisation du composant ONE WAY MANOVA (Multivariate Analysis of Variance). Nous verrons notamment que ce test multivarié peut amener des conclusions différentes de la combinaison des tests univariés menés sur chaque descripteur de l'espace de représentation.

Fichier

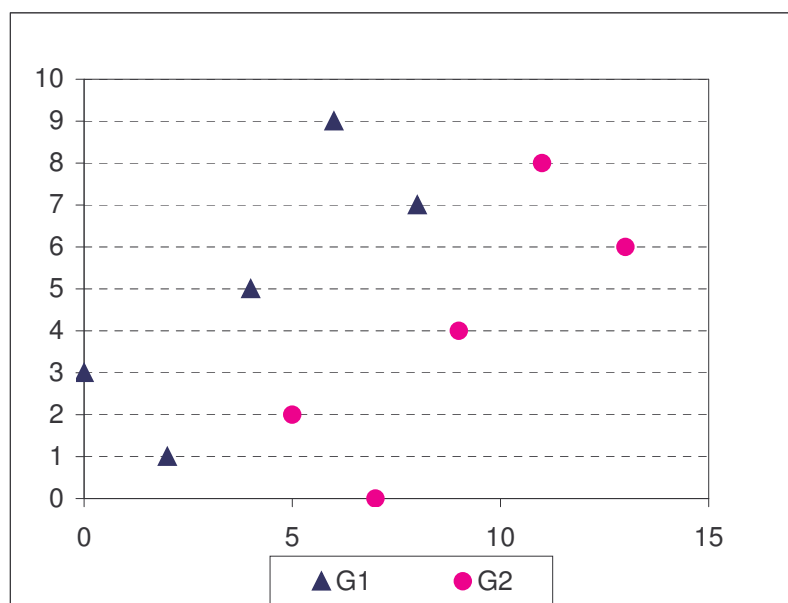
Le fichier TOMASSONE_P_29.XLS¹ recense 10 individus décrits par 2 variables (X1, X2) et répartis en deux groupes (G1, G2). Nous voulons mettre en place le test suivant

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

$\mu_{gk} = \begin{pmatrix} \mu_{gk,x1} \\ \mu_{gk,x2} \end{pmatrix}$ est le centre de gravité du groupe Gk.

Il s'agit bien d'un test de comparaison de moyennes de plusieurs populations (2 populations dans notre fichier) telle que nous le pratiquons en analyse de variance, sauf que nous sommes en présence de plusieurs variables dépendantes (X1 et X2).

Pour mieux appréhender les résultats, nous proposons une représentation des données dans le plan.

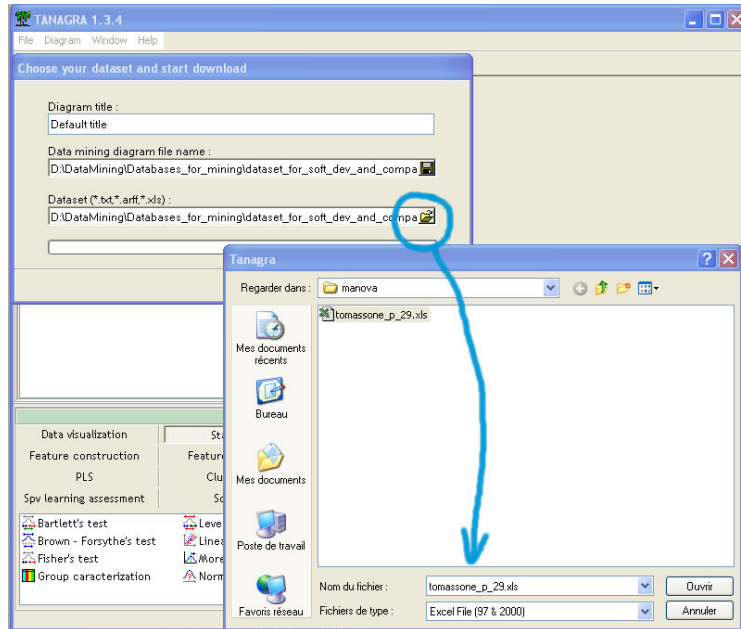


¹ R. Tomassone, M. Danzart, J. Daudin, J. Masson, « Discrimination et Classement », Masson, 1988, page 29.

MANOVA

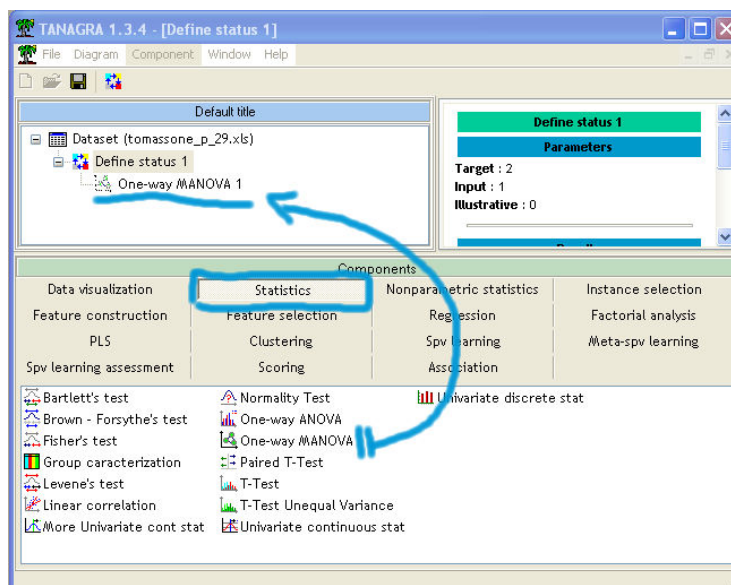
Importer les données

Nous créons un nouveau diagramme et importons le fichier de données.



MANOVA

Pour mettre en place le test MANOVA, il faut dans un premier temps insérer le composant DEFINE STATUS dans le diagramme, placer les variables X1 et X2 en TARGET, ce sont les variables dépendantes ; et la variable GROUP en INPUT, c'est la variable indépendante. Puis, dans un deuxième temps, nous insérons le composant ONE WAY MANOVA qui se trouve dans la palette STATISTICS.



Le Lambda de WILKS est la principale statistique de test produite par cette procédure, elle a été tabulée par son auteur, il n'existe pas en revanche de fonction approchant la fonction de répartition. Pour cette raison, dans la pratique, pour juger de la significativité du test, nous utilisons les transformations de BARTLETT qui suit une loi du CHI-2, et de RAO, plus précise, qui suit une loi de FISHER. C'est approximations sont relativement performantes pour des tailles d'échantillon modérées.

Dans notre cas, nous constatons qu'au niveau de signification de 1%, la différence du paramètre de localisation entre les deux populations constaté sur l'échantillon n'est pas dû au hasard. Le résultat est souligné en rouge dans ce cas².

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

Ce test reposant sur une hypothèse d'égalité des matrices de variance co-variance, en toute rigueur nous aurions dû tester ce postulat au préalable. Dans la pratique, nous passons souvent outre pour deux raisons : le test MANOVA est relativement robuste ; et le test BOX-M³ qui permet de tester l'égalité des matrices de variance co-variance des groupes est lui-même très peu robuste, il dépend exagérément de la conformité à la loi normale de la fonction de répartition.

ANOVA sur chaque variable

Il est tentant de retraduire le test ci-dessus en une série de tests individuels

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

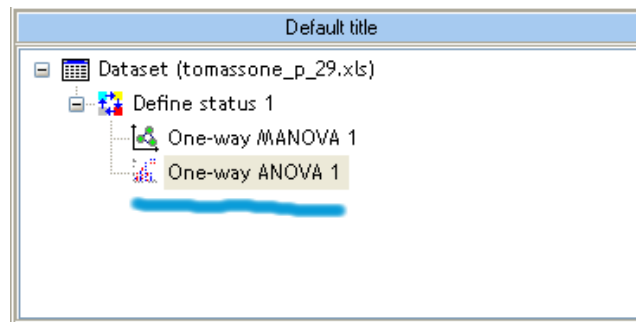
Nous concluons alors au rejet d'égalité globale des moyennes si, dans l'un des tests au moins, nous décidons de rejeter de l'hypothèse nulle.

² L'égalité des 2 p-values est une simple coïncidence ici.

³ Le test BOX-M est une généralisation au cas multidimensionnel du test de BARTLETT d'égalité des variances

Bien que séduisante au premier abord, elle permet de simplifier grandement les calculs, cette stratégie est erronée car elle s'appuie sur une mauvaise estimation de la variance : elle ne tient pas compte de la co-variance entre les variables dépendantes.

Pour nous en persuader, ajoutons le composant ONE WAY ANOVA dans notre diagramme. TANAGRA procédera à une analyse de variance univariée pour chaque variable dépendante.



Les résultats sont édifiants. Si l'on s'en tient à cette procédure, les deux tests individuels considèrent que l'hypothèse d'égalité des moyennes est compatible avec les données : nous concluons donc à tort que les centres de gravité des populations sont confondus au risque de 1%.

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				