

## Objectif

Montrer l'utilisation du composant CANONICAL DISCRIMINANT ANALYSIS, connue en français sous l'appellation « Analyse Factorielle Discriminante » ou encore « Analyse Discriminante Descriptive ».

L'objectif de cette méthode est de produire, à partir d'un ensemble d'individus répartis dans des groupes décrits par plusieurs descripteurs, de nouvelles variables, les axes factoriels, on parle aussi de variables discriminantes, qui séparent au mieux les groupes.

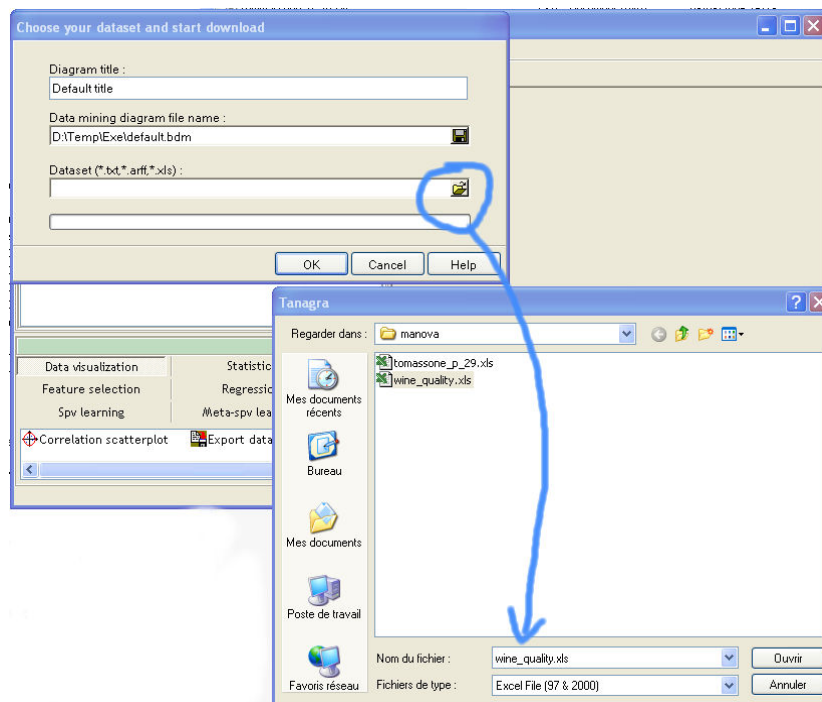
## Fichier

Nous travaillerons sur le fichier WINE\_QUALITY.XLS<sup>1</sup>. Il recense 34 crus du bordelais répartis en 3 groupes « bon », « moyen », « médiocre » ; les descripteurs correspondent à des variables météorologiques (somme des températures journalières, jours d'ensoleillement, jours de chaleur, pluie).

## Analyse Factorielle Discriminante

### Importer les données

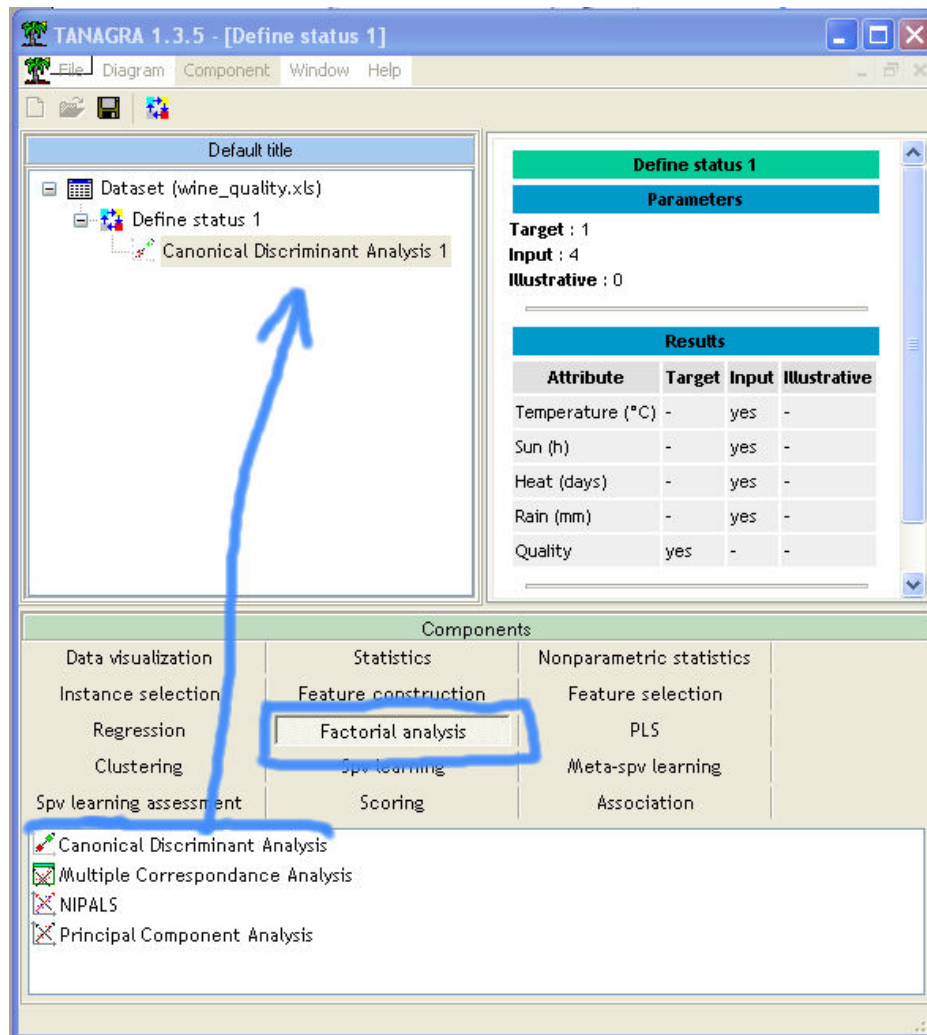
Première étape toujours, créez un nouveau diagramme et importez les données à l'aide du menu FILE / NEW.



<sup>1</sup> Extrait de M. Tenenhaus, « Méthodes Statistiques en Gestion », Edition Dunod, 1996, p. 244 (Tableau 1) – Il s'agit de données annuelles sur la période 1924 – 1957.

## Définir l'analyse

A l'aide du composant DEFINE STATUS, placez la variable QUALITY en TARGET, et les autres variables en INPUT ; puis insérez le composant CANONICAL DISCRIMINANT ANALYSIS.



## Lecture des résultats

Les résultats sont répartis dans 3 tableaux qui indiquent respectivement le pouvoir discriminant des axes factoriels ; les coefficients des équations de projections sur les axes factoriels ; les corrélations entre les variables et les axes factoriels.

Dans notre cas, nous constatons que deux axes factoriels ont été produits, le premier résume déjà 95,9% de la variance expliquée, le second axe n'est pas significatif, la p-value du test de Bartlett est de 0,28.

Results							
<b>Roots and Wilks' Lambda</b>							
Root	Eigenvalue	Proportion	Canonical R	Wilks Lambda	CHI-2	d.f.	p-value
1	3.27886	0.95945	0.875382	0.205263	46.7122	8	0.000000
2	0.13857	1.00000	0.348867	0.878292	3.8284	3	0.280599
<b>Canonical Discriminant Function</b>							
Coefficients	Unstandardized		Standardized				
	Attribute	Root n°1	Root n°2	Root n°1	Root n°2		
Temperature (°C)	-0.0086	0.0000	-0.7509	0.0041			
Sun (h)	-0.0068	0.0053	-0.5476	0.4309			
Heat (days)	0.0271	-0.1278	0.1984	-0.9362			
Rain (mm)	0.0059	-0.0062	0.4456	-0.4690			
constant	32.91135	-2.16759	-				
<b>Factor Structure Matrix - Correlations</b>							
Root	Root n°1			Root n°2			
	Descriptors	Total	Within	Between	Total	Within	Between
Temperature (°C)	-0.901	-0.724	-0.987	-0.375	-0.584	-0.164	
Sun (h)	-0.897	-0.701	-0.999	0.116	0.176	0.052	
Heat (days)	-0.771	-0.525	-0.956	-0.590	-0.780	-0.292	
Rain (mm)	0.663	0.398	0.977	-0.361	-0.421	-0.212	

Les coefficients non-standardisés (bruts) nous permettent d'effectuer une projection pour une nouvelle observation. Par exemple, sur le premier axe Z1, avec TEMPERATURE = 3000, SUN = 1100, HEAT = 20 et RAIN = 300, nous obtiendrons sa coordonnée sur le premier axe avec :  $-0.0086 \times 3000 + -0.0068 \times 1100 + 0.0271 \times 20 + 0.0059 \times 300 + 32.91135 = 1.9435$

Les coefficients standardisés correspondent au produit des coefficients bruts avec l'écart type intra-classes des variables. Ils permettent de comparer l'importance des variables dans la construction des axes factoriels en supprimant les disparités dues à l'utilisation d'échelles différentes pour mesurer les données.

Enfin, dernier tableau, la matrice de structure indique les corrélations totales, intra-classes et inter-classes, entre les variables et les axes factoriels. Tout comme les coefficients standardisés, ils fournissent de bonnes indications pour l'interprétation des axes factoriels.

## Représentation graphique

Il est possible de représenter les données dans le plan formé par les 2 axes factoriels. Pour ce faire, ajoutez un composant SCATTERPLOT dans le diagramme.

Sélectionnez les variables adéquates dans les listes déroulantes (CDA\_AXIS), le graphique est d'autant plus intéressant que nous pouvons illustrer les points selon leur groupe d'appartenance (QUALITY). Nous constatons que le premier axe factoriel permet de discerner la qualité des vins : plus la valeur est forte, moins bon sera le cru. Pour la nouvelle observation ci-dessus, nous pouvons prédire que ce cru ( $Z_1 = 1.9435$ ) sera vraisemblablement de qualité médiocre.

