# Objectif

Mettre en œuvre l'analyse des correspondances multiples (ACM) avec TANAGRA.

L'analyse des correspondances multiples (ou analyse factorielle des correspondances multiples – AFCM) est une technique de visualisation très populaire en analyse de données. Dans ce tutoriel, nous montrons comment la mettre en œuvre avec TANAGRA.

## Fichier de données

Nous utilisons le fichier RACES\_CANINES\_ACM.XLS tiré de l'ouvrage de TENENHAUS¹ (Tableau 1, page 200). L'intérêt de ce fichier est que nous pouvons comparer directement nos résultats avec ceux du livre (pages 212 à 222). Nous nous contentons de montrer l'enchaînement des opérations et la lecture des tableaux de résultats dans ce tutoriel. Pour ce qui de l'interprétation, le mieux est de se référer à l'ouvrage.

Le tableau de données est le suivant :

Chien	Taille	Poids	Velocite	Intelligence	Affection	Agressivite	Fonction
Beauceron	Taille++	Poids+	Veloc++	Intell+	Affec+	Agress+	utilite
Basset	Taille-	Poids-	Veloc-	Intell-	Affec-	Agress+	chasse
Berger All	Taille++	Poids+	Veloc++	Intell++	Affec+	Agress+	utilite
Boxer	Taille+	Poids+	Veloc+	Intell+	Affec+	Agress+	compagnie
Bull-Dog	Taille-	Poids-	Veloc-	Intell+	Affec+	Agress-	compagnie
Bull-Mastif	Taille++	Poids++	Veloc-	Intell++	Affec-	Agress+	utilite
Caniche	Taille-	Poids-	Veloc+	Intell++	Affec+	Agress-	compagnie
Chihuahua	Taille-	Poids-	Veloc-	Intell-	Affec+	Agress-	compagnie
Cocker	Taille+	Poids-	Veloc-	Intell+	Affec+	Agress+	compagnie
Colley	Taille++	Poids+	Veloc++	Intell+	Affec+	Agress-	compagnie
Dalmatien	Taille+	Poids+	Veloc+	Intell+	Affec+	Agress-	compagnie
Doberman	Taille++	Poids+	Veloc++	Intell++	Affec-	Agress+	utilite
Dogue All	Taille++	Poids++	Veloc++	Intell-	Affec-	Agress+	utilite
Epag. Breton	Taille+	Poids+	Veloc+	Intell++	Affec+	Agress-	chasse
Epag. Français	Taille++	Poids+	Veloc+	Intell+	Affec-	Agress-	chasse
Fox-Hound	Taille++	Poids+	Veloc++	Intell-	Affec-	Agress+	chasse
Fox-Terrier	Taille-	Poids-	Veloc+	Intell+	Affec+	Agress+	compagnie
Gd Bleu Gasc	Taille++	Poids+	Veloc+	Intell-	Affec-	Agress+	chasse
Labrador	Taille+	Poids+	Veloc+	Intell+	Affec+	Agress-	chasse
Levrier	Taille++	Poids+	Veloc++	Intell-	Affec-	Agress-	chasse
Mastiff	Taille++	Poids++	Veloc-	Intell-	Affec-	Agress+	utilite
Pekinois	Taille-	Poids-	Veloc-	Intell-	Affec+	Agress-	compagnie
Pointer	Taille++	Poids+	Veloc++	Intell++	Affec-	Agress-	chasse
St-Bernard	Taille++	Poids++	Veloc-	Intell+	Affec-	Agress+	utilite
Setter	Taille++	Poids+	Veloc++	Intell+	Affec-	Agress-	chasse
Teckel	Taille-	Poids-	Veloc-	Intell+	Affec+	Agress-	compagnie
Terre-Neuve	Taille++	Poids++	Veloc-	Intell+	Affec-	Agress-	utilite

La première colonne correspond à l'identifiant des observations, les variables actives sont en vert, la variable illustrative en bleu. Notons que dans l'ouvrage, le tableau est déjà codé en disjonctif complet. Dans notre tutoriel, nous avons mis en forme les données telles qu'elles seront présentées à TANAGRA qui, en interne, effectuera le codage 0/1 approprié.

21/12/06 Page 1 sur 10

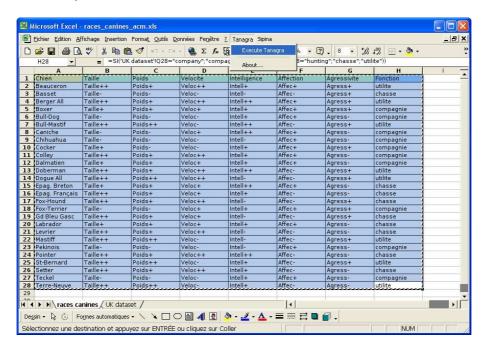
\_

<sup>&</sup>lt;sup>1</sup> M. TENENHAUS, « Méthodes Statistiques en gestion », DUNOD, 1996. C'est un ouvrage qui, par sa clarté et son accessibilité, fait référence en France s'agissant de la pratique de l'analyse de données.

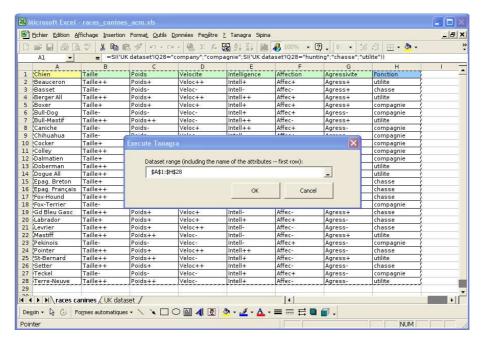
# Analyse des Correspondances Multiples avec TANAGRA

#### Créer un diagramme

Depuis la version 1.4.11, il est possible de démarrer TANAGRA à partir du tableur EXCEL<sup>2</sup>. C'est la procédure que nous choisissons ici : le diagramme est automatiquement créé, et les données importées. Pour ce faire, nous sélectionnons la plage de données dans EXCEL, puis nous cliquons sur le menu TANAGRA/ EXECUTE TANAGRA.



Une boîte de dialogue vient confirmer la sélection en affichant les références de la plage de cellules. Nous validons en cliquant sur OK.



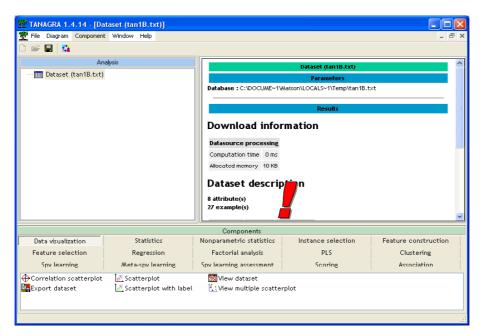
<sup>&</sup>lt;sup>2</sup> Il faut bien entendu avoir référencé la macro-complémentaire (Add-In) TANAGRA dans EXCEL, voir le didacticiel adéquat sur le site web. La démarche est également valable avec le tableur CALC de OPEN OFFICE.

21/12/06 Page 2 sur 10

\_

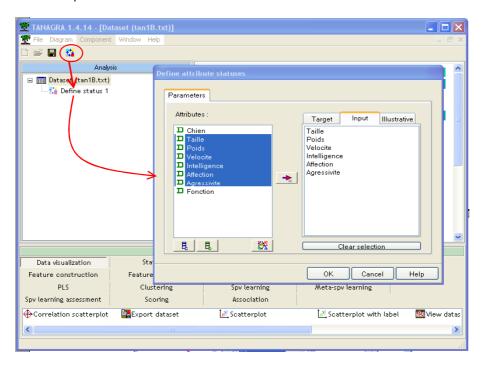
TANAGRA est alors démarré, nous vérifions que l'ensemble de données comprend bien 27 observations et 8 variables<sup>3</sup>.

R.R.



## ACM (AFCM)

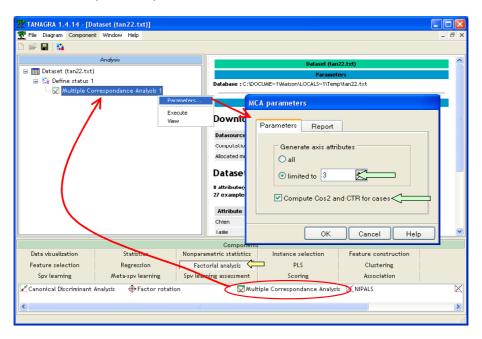
Pour initier une analyse, nous devons tout d'abord définir le rôle des variables. C'est le rôle du composant DEFINE STATUS accessible dans la barre d'outil. Nous mettons en INPUT les variables actives. Nous verrons plus tard comment utiliser les variables illustratives.



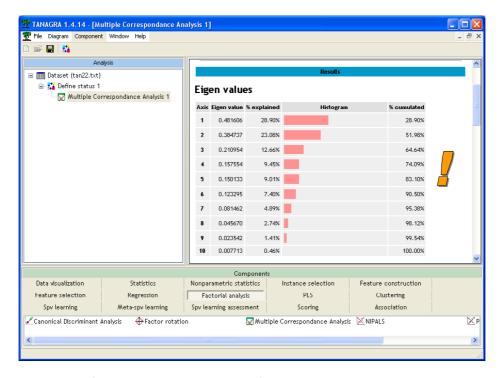
<sup>&</sup>lt;sup>3</sup> Le type identifiant n'est pas reconnu par TANAGRA, la première colonne est donc codée comme une variable discrète, avec autant de modalités que d'observations. Cela ne pose pas de problème dans la pratique... sauf si nous traitons plus de 255 observations. Mais l'identifiant n'est vraiment utile que si l'on veut produire les plans factoriels en étiquetant les points. Si les observations sont nombreuses, ce graphique ne sera pas très lisible de toute manière, les étiquettes se superposant les uns sur les autres.

21/12/06 Page 3 sur 10

Puis nous plaçons le composant MULTIPLE CORRESPONDANCE ANALYSIS (onglet FACTORIAL ANALYSIS) dans le diagramme de traitements. Nous cliquons sur le menu PARAMETERS : nous spécifions alors le nombre d'axes à produire (3) ; nous activons l'option qui permet de calculer les COS2 et les contributions pour chaque individu.



**Valeurs propres.** Nous activons le menu contextuel VIEW pour accéder aux résultats. La première partie des résultats correspond au tableau des valeurs propres (Tableau 3, p.214).



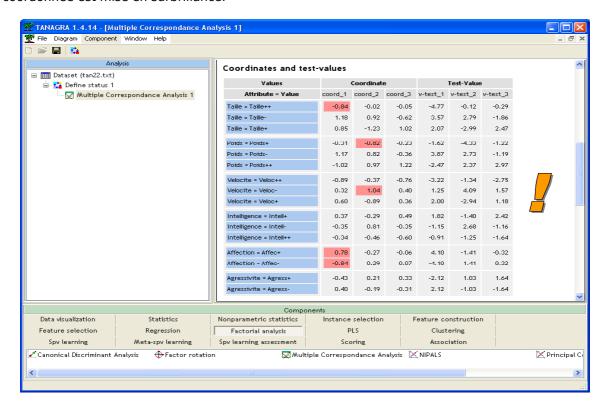
**Analyse des modalités.** La seconde partie des résultats indique le positionnement des modalités sur les axes factoriels, leurs contributions et leurs COS<sup>2</sup> (Tableau 5<sup>4</sup>, p.216).

21/12/06 Page 4 sur 10

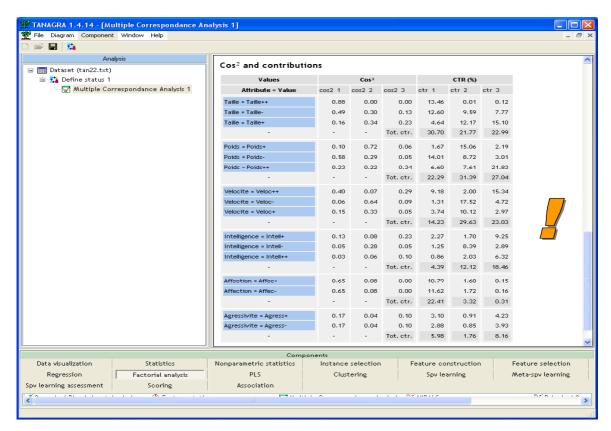
\_

<sup>&</sup>lt;sup>4</sup> Dans son ouvrage, M. Tenenhaus reprend les corrélations. Ils correspondent à la racine du COS², avec le signe de la coordonnée factorielle. La lecture des résultats n'en est pas affectée.

Le premier tableau indique les coordonnées factorielles des modalités et les valeurs tests associées. Lorsque la valeur test est supérieure au seuil défini par l'utilisateur (par défaut 4), la coordonnée est mise en surbrillance.



Le second tableau fournit les cosinus au carré et les contributions.



Coordonnées factorielles : projection, COS<sup>2</sup> et contribution. TANAGRA ne fournit pas les vecteurs propres pour l'ACM. Il calcule directement les coordonnées factorielles, les contributions et

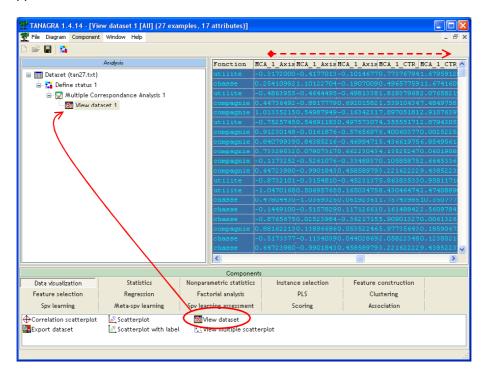
21/12/06 Page 5 sur 10

Didacticiel - Études de cas

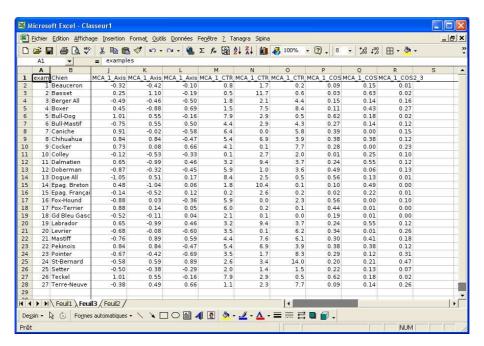
les COS² pour chaque axe demandé. Ces nouvelles variables sont ajoutées aux données et sont disponibles en aval du composant.

R.R.

Pour visualiser le tableau de données associé, nous plaçons dans le diagramme le composant VIEW DATASET (onglet DATA VISUALIZATION). Nous cliquons sur le menu VIEW. Dans la dernière partie de la grille apparaissent les nouvelles colonnes.



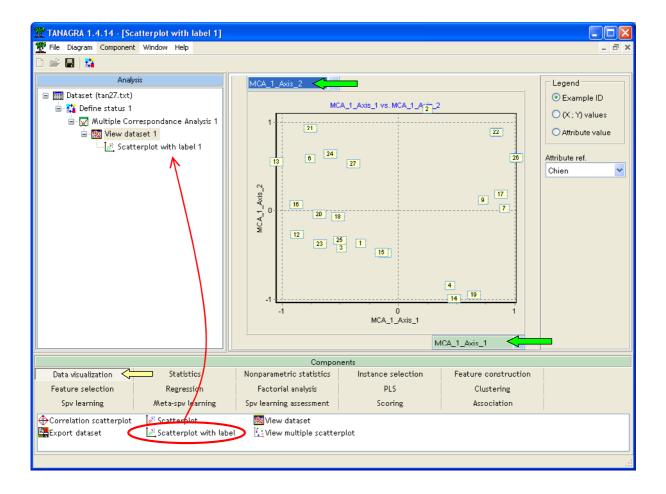
TANAGRA adopte le format scientifique. Cela peut gêner la lecture. Une manière simple de s'en sortir est de copier (menu COMPONENT / COPY RESULTS) et de coller les données de la grille dans le tableur de votre choix. En adaptant la précision à notre convenance, nous obtenons le tableau 4, p.215.



L'intérêt du tableur est manifeste ici. En effet, nous disposons de multiples possibilités de tri qui permettent de mettre en évidence les informations pertinentes.

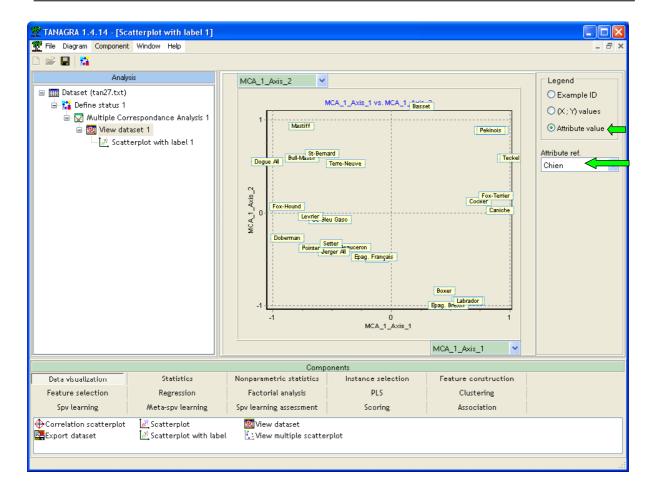
21/12/06 Page 6 sur 10

**Plans factoriels.** La popularité des méthodes factorielles repose en grande partie sur les représentations graphiques qu'elle propose. Elles nous permettent d'apprécier visuellement les proximités entre les observations. Dans notre cas, nous projetons les observations dans le premier plan factoriel. Nous voulons associer les identifiants aux points. Nous utilisons pour cela le composant SCATTERPLOT WITH LABEL (onglet DATA VISUALIZATION) que nous plaçons en dessous de l'ACM. Nous le paramétrons de manière à avoir en abscisse le premier facteur, en ordonnée le second facteur. Notons qu'il est très aisé de passer d'un plan factoriel à un autre.



Au départ, les points sont identifiés par leur numéro. En activant l'option LEGEND / ATTRIBUTE VALUE, et en choisissant la référence MODELE, nous obtenons la carte des points étiquetés par leurs identifiants (Figure 1, p.218). Bien entendu, cette option est pratique tant que le nombre de points reste raisonnable. Au-delà d'un certain nombre d'observations, le graphique serait illisible. Notons également que sur cet exemple précis, certains points se superposent parce qu'ils ont des coordonnées identiques. Il est très difficile de les discerner dans le graphique, il est plus indiqué d'inspecter le tableau des coordonnées factorielles dans ce cas (toujours dans un tableur par exemple, en exploitant les fonctionnalités de tri).

21/12/06 Page 7 sur 10



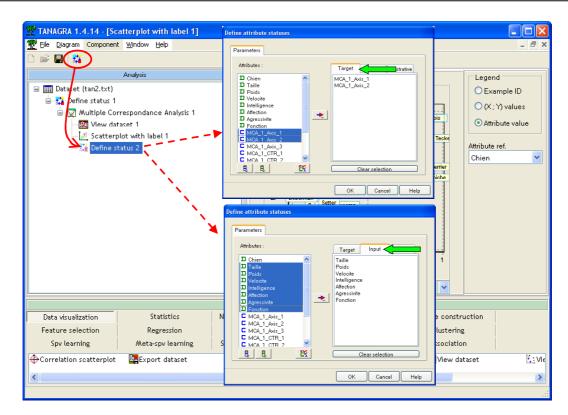
R.R.

Il est possible de modifier la taille des étiquettes avec les raccourcis CTRL+Q et CTRL+W.

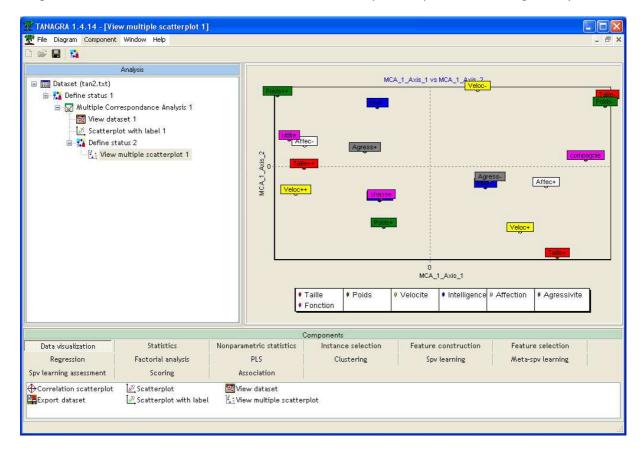
Carte des modalités. Nous pouvons projeter les individus, nous pouvons aussi projeter les variables dans les plans factoriels. Ca peut être utile lorsque l'on veut interpréter les axes. Les informations sont disponibles dans le tableau des sorties, mais une représentation graphique est toujours plus avenante. De plus, nous nous donnons la possibilité de projeter les variables illustratives dans le nouveau repère.

Nous procédons en deux temps. Tout d'abord nous ajoutons le composant DEFINE STATUS en dessous de l'ACP. Nous définissons les deux premiers axes comme TARGET, les variables actives et les variables illustratives sont placées en INPUT.

21/12/06 Page 8 sur 10



Dans un deuxième temps, nous ajoutons le composant VIEW MULTIPLE SCATTERPLOT dans le diagramme. Nous obtenons la carte des variables dans le premier plan factoriel (Figure 2, p.220).



Nous retrouvons bien ici les commentaires page 217-219. Sur le premier axe, les chiens de grande taille, véloces et peu affectueux, utilitaires, sont opposés aux chiens de compagnie, de petite taille et de faible poids (cf. l'ouvrage pour les commentaires détaillés).

21/12/06 Page 9 sur 10

Didacticiel - Études de cas

Individus illustratifs. Nous ne l'avons pas mis en œuvre dans ce tutoriel, mais il est tout à fait possible de travailler avec les données illustratives en analyse factorielle. Le plus simple serait alors de définir une variable indicatrice dans une nouvelle colonne, avec deux modalités (« actif » et « illustratif ») ; puis d'utiliser les composants INSTANCE SELECTION pour sélectionner les individus actifs (DISCRETE SELECT EXAMPLES ou RULE BASED SELECTION) lors de la construction des axes ; et enfin, de projeter la totalité des observations, ou uniquement les individus illustratifs, dans les différents plans factoriels en définissant le sous-ensemble adéquat avec le composant RECOVER EXAMPLES.

#### Conclusion

TANAGRA ne prétend pas fournir des outils de reporting et de déploiement à la hauteur des logiciels commerciaux. En se contentant de proposer des résultats standards, repris dans des ouvrages qui font référence, nous essayons de donner aux utilisateurs les principaux codes de lecture d'une analyse factorielle.

Pouvoir reprendre les résultats dans un tableur est certainement une des fonctionnalités les plus intéressantes du logiciel. En effet, il nous donne accès à des outils (tri, mise en forme, etc.) dans un environnement bien connu des praticiens du traitement des données. Par exemple, la possibilité de trier les différents tableaux selon les contributions et les COS2 s'avère réellement pratique lorsque l'on souhaite interpréter les axes.

21/12/06 Page 10 sur 10