Majeure *Science des données* 2017-18

Compte-rendu du TP n°4 Analyse factorielle discriminante

**Andi WANG**

1. Représenter chaque groupe (modalité de la variable qualitative type :‘chien’ ou ‘loup’) :

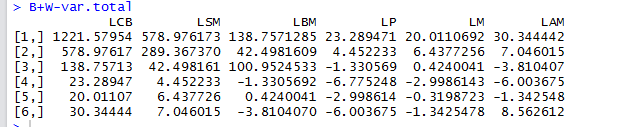
Valeur de discrimination d’une variable Xj avant projection

*Etudiez les matrices V-1 et B : dimension, caractéristiques*

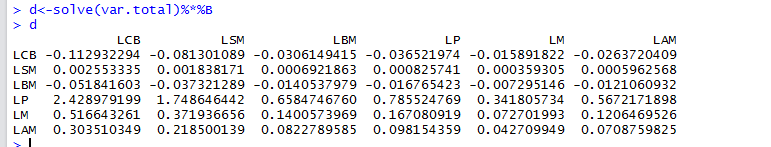
*Pouvez toujours diagonaliser directement V^(-1) B ?*

On calcule la matrice V, B et W pour les 2 groupes de données et obtenir des résultats de cette modèle initiale.

Quand on utilise des données originales et sans les centrer, on calcule (W+B)-V, et on obtient:



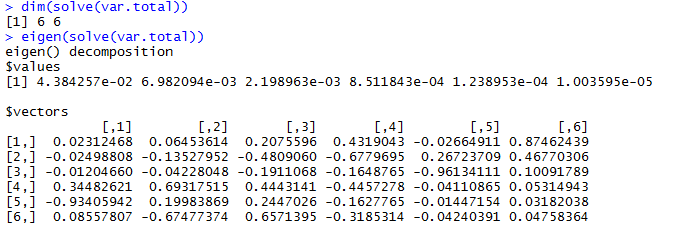
Et la matrice d:



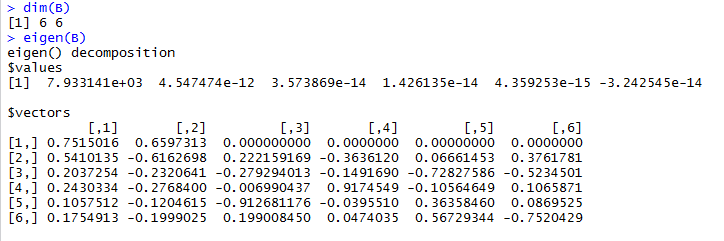
Les résultats ne sont pas zéros ou dans vers zéro, donc, cette modèle ne peut pas bien séparer des données initiales de ces 2 groupes.

Et puis on étude la matrice V-1 et B:

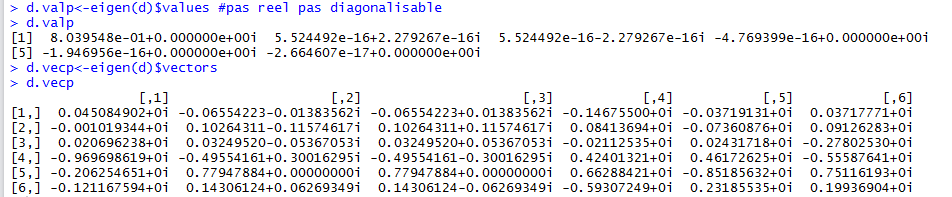
Pour la matrice V-1 (on a utilisé solve(V) pour calculer V-1):



Pour la matrice B:



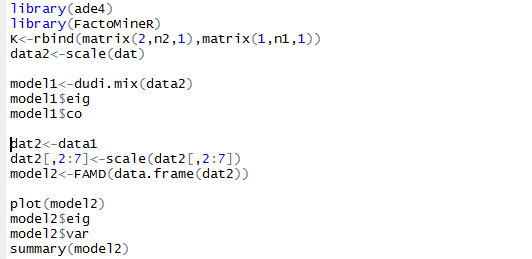
Quand on calcule directement eigen de V ^(-1)\*B, on va obtenir:



Dans cette condition, les valeurs propres de la matrice ne sont pas toutes réel. Donc, on ne peut pas directement diagonaliser cette matrice.

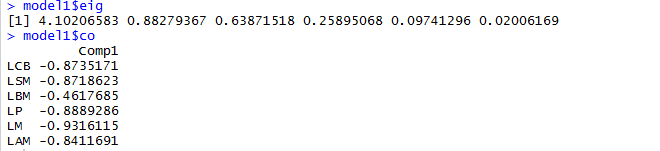
2) Implémenter l’AFD centrée et tester sur le fichier de données fournies avec les deux packages.

*code ou pseudo code*

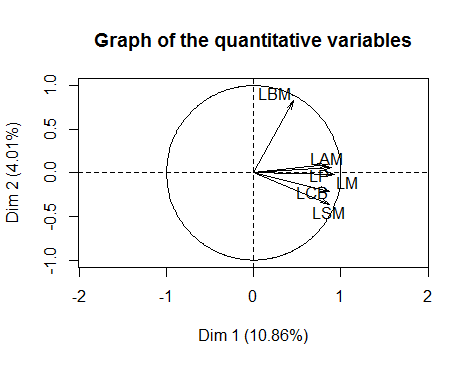
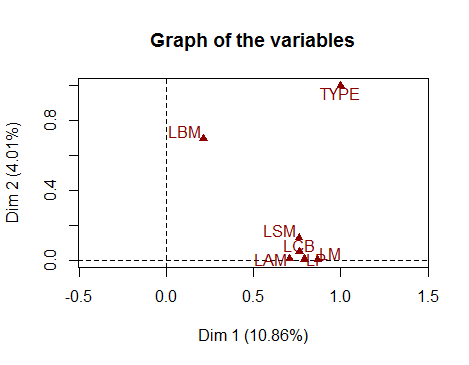
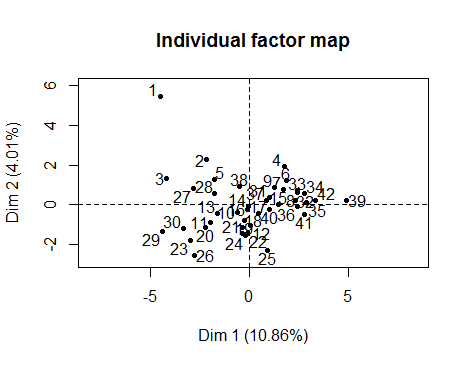


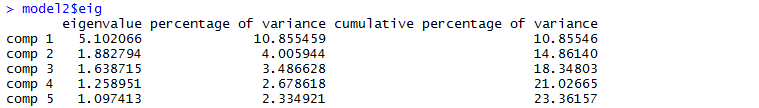
*Principaux résultats de l’AFD descriptive:*

Pour modèle de dudi.mix, on a des résultats:



Pour modèle de FAMD, on a des figures et des résultats:

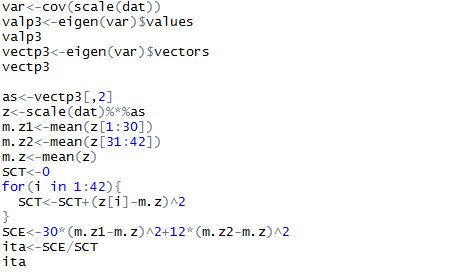




1. calcul du pouvoir discriminant (rapport de corrélation = η = SCE/SCT) de chaque axe factoriel *as* est évalué en calculant les nouvelles coordonnées zij des individus xij à partir de l’équation de l’axe *as*puis les barycentres conditionnels des groupes projetés sur cet axe soit :

*code de calcul du SCE SCT*

*code de calcul de* η = SCE/SCT

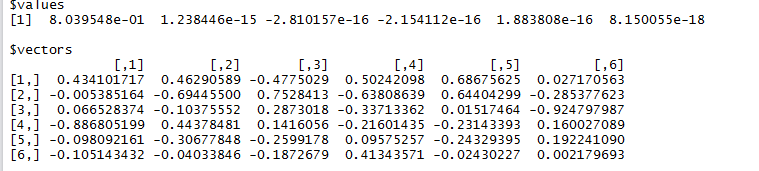


La matrice vectp3 stocke les vecteur propres, ce sont aussi des axes correspondues

On fait changement de la colonnes de vectp3 pour changer des axes différentes

*Principaux résultats de l’AFD descriptive:*

Des valeurs propres et des vecteurs propres:



Les 6 valeurs de rapport corrélation des 6 axes:

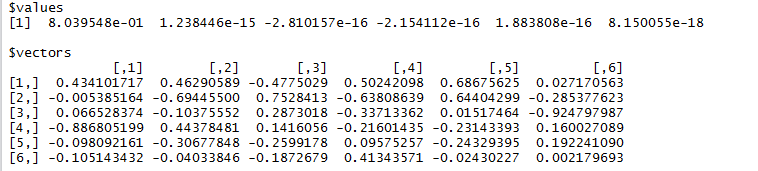
0.8235635 3.113587e-31 2.928718e-32 9.563876e-32 1.875833e-33 1.675519e-32

*A partir du calcul de ηs  de chacun des axes factoriels : choisir le nombre d’axes à retenir*

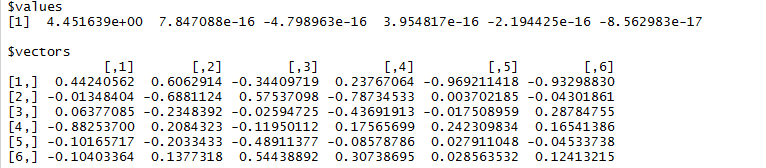
Apres des calculs des valeurs de rapport correlation des 6 axes, on a trouve que il y a seulement une axe dont le rapport de correlation est proche de 1, les autres sont toutes trop petites, donc le nombre d’axes a retenir est 1.

*Comparer les valeurs propres et vecteurs propres des matrices V-1B , et de W-1B*

Pour la matrice V-1B:



Pour la matrice W-1B:



Pour chaque valeur propres de V-1B on note λ, on peut trouver que  est la valeur propre de W-1B.

Et en plus, les vecteurs propres de ces 2 matrice sont proches.

*Que remarquez-vous entre la valeur propre de chaque axe et du rapport de corrélation*

la valeur propre de V^(-1)\*B sont proche du rapport de correlation

*Fournir les équations des K-1 axes factoriels (K nombre max de groupes)*

Pour cette question K=2, donc on a seulement 1 axe factoriel:



4) Qualité de la projection de l’AFD.

*Qualité de la projection d’un nuage par axe as*



La qualite de la projection d’un nuage par axe 1 sont presque 100%.

*Contribution absolue du centre de gravité gq à l’axe as*



*Contribution relative du centre de gravité gq à l’axe as*



*Calculer ces 3 indicateurs pour chaque axe et chaque groupe*

1. AFD sur données centrée-réduites *contributions partielles de variables aux nouvelles variables cad leurs corrélations.*