

Préambule

Ce travail pratique explore l'Analyse par composantes factorielles, avec tous les codes contenus dans le script *Tp_2_ACF_Rendu.R*. Le rapport est structuré en trois sections principales : la première se concentre sur une première exploration de la base de données. La deuxième section revisite un exemple de cours pour valider le code élaboré dans la première partie, et enfin, la troisième réalise une analyse approfondie de la base de données de travail.

Étude d'une base de données

Cette première partie consiste à étudier la table des données liées au travail d'une série d'individus, et d'en faire une première ACF sur un jeu de données choisi. La table est composée de 5 catégories correspondant à l'identifiant de l'employé, l'âge de ce dernier, la fonction occupée dans l'entreprise, le temps de travail et la qualité de vie. Les trois dernières catégories sont détaillées ci-dessous :

Modalité par *Fonction* : Administratif : modalité 1 Technicien (OS) : modalité 2 ingénieur : 3 technicien supérieur :4
direction : modalité 5 contractuel_S1 : modalité 6 contractuel_S2 : modalité 7 non_repondu : 0
Modalité par *Sexe* : F : 2 H : 1 non_repondu : 0

A la question 'temps nécessaire pour effectuer mon travail' : les modalités sont

- modalité 8 : je suis en sous charge de travail
- modalité 2 : la charge de travail correspond à mes horaires de travail
- modalité 4 : je ne dépasse pas mes horaires de travail
- modalité 16 : je suis surchargé et je dépasse mes horaires de travail
- non_repondu : 99

A la question : la qualité de vie au travail a été évaluée selon 3 modalités

- modalité 2 : améliorer
- modalité 4 : maintenue
- modalité 16 : dégradée

La première étape implique la création des matrices essentielles pour l'Analyse en Composantes Factorielle (ACF). Toutes les étapes sont explicitées dans le code. Dans notre exemple, nous avons choisi deux variables spécifiques - la fonction de travail et le temps de travail - qui offrent une analyse intéressante.

Pour réduire l'espace, on détermine les axes de projection en diagonalisant la matrice S . Cette matrice est obtenue en effectuant des opérations avec différentes matrices telles que F (matrice des fréquences relatives), D_n (matrice diagonale des marges en ligne) de taille $n*n$, et D_p (matrice diagonale des marges en colonne) de taille $p*p$. En identifiant les vecteurs propres les plus impactant dans cette matrice, nous choisissons les deux qui influencent le plus la représentation des valeurs. En recalculant les coordonnées des points dans cet espace, nous obtenons un graphique qui représente le profil des lignes.

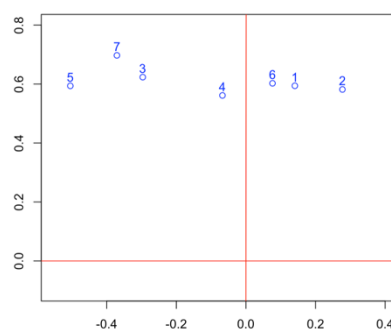


Figure 1 : Projection profil ligne

De manière similaire, nous répétons ces étapes pour les profils des colonnes en utilisant la matrice TT. Cela nous permet d'obtenir un graphique représentant le profil des colonnes.

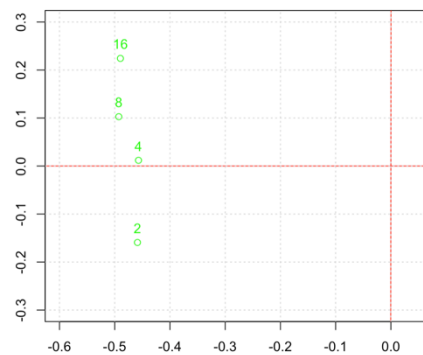


Figure 2 : Projection profil colonne

Pour évaluer la qualité de l'ACP réalisée sur le profil des lignes, nous utilisons la fonction `quality_ACP()` élaborée dans des travaux pratiques ACP antérieurs. Cela nous donne des informations sur la qualité de la représentation.

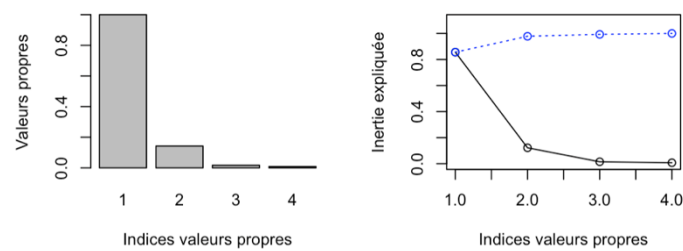


Figure 3 : Qualité de la projection du profil ligne

En utilisant la fonction de FactoMineR, nous obtenons un graphique qui offre une perspective supplémentaire sur l'analyse effectuée.

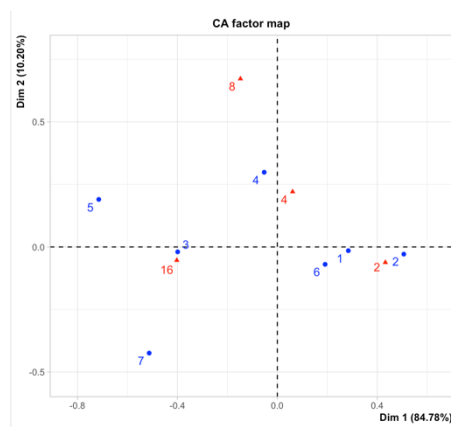


Figure 4 : Résultat de la fonction FactoMineR

Nous pouvons finalement effectuer un test du Chi2 de la façon suivante comme indiqué dans l'énoncé du TP. Cela permet de nous renseigner sur la dépendance possible de

certaines variables entre elles sans nous en amené la certitude. On obtient un tableau ci-contre, à partir de ce dernier, on pourrait supposer par exemple la dépendance de la variable 1 de la fonction de travail avec la variable 2 et 16 du temps de travail. Mais ce n'est pas suffisant pour conclure.

	2	4	8	16
1	7.70	0.09	0.08	6.09
2	11.54	0.46	1.27	11.71
3	14.63	0.29	0.00	15.42
4	0.64	1.08	4.45	0.07
5	7.18	1.44	0.25	3.44
6	0.91	0.01	0.62	0.67
7	0.99	3.67	0.29	5.00

Exemple de cours

Dans cette partie nous souhaitons revenir à l'étude données (issues de Pagès) déjà sous forme de table de contingence suivante :

	<i>Perçu sucré</i>	<i>Perçu acide</i>	<i>Perçu amer</i>
<i>Sucré</i>	10	0	0
<i>Acide</i>	0	8	2
<i>Amer</i>	0	4	6

Comme précédemment, on débute par le calcul du tableau des fréquence relative du tableau de contingence puis de la détermination du profil ligne et du profil colonne. On veut réaliser l'ACP sur ces 2 profils.

On commence par celui sur le profil ligne. Pour cela on doit déterminer les valeurs et vecteurs propres de la matrice $S = F' * D_n^{-1} * F * D_p^{-1}$ avec F la matrice des fréquences relatives, D_n la matrice diagonale des marges en ligne $f_{i.}$ de dimension $n*n$ et D_p la matrice diagonale des marges en colonne $f_{.j}$ de dimension $p*p$. Ensuite, les composantes principales sont déterminées en projetant les points dans le plan formé par les deux vecteurs propres associés aux deux valeurs propres dominantes. La figure correspondante est ensuite obtenue comme illustré ci-contre :

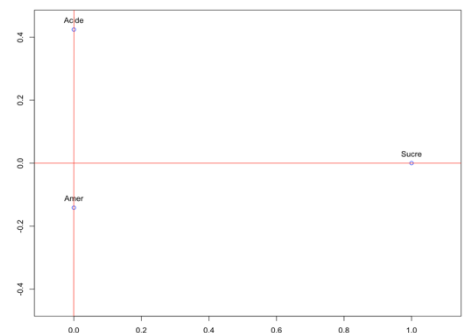


Figure 5 : Projection profil ligne

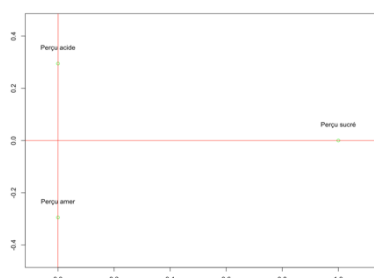


Figure 6 : Projection profil colonne

On effectue le même travail pour le profil colonne, mais cette fois-ci avec la matrice $TT = F * D_p^{-1} * F' * D_n^{-1}$. On obtient les composantes principales en projetant les points dans le plan formé par les deux vecteurs propres associés aux deux valeurs propres dominantes. La figure correspondante est ensuite obtenue comme illustré ci-contre :

Ces deux graphiques combinés avec la table de contingence offrent une clé d'interprétation de la base de données. En observant ces résultats, on peut conclure que la présence d'un produit sucré est généralement associée à une perception sucrée. De même, la présence d'un aliment amer tend à induire une perception amère, tandis qu'un aliment acide est fortement lié à une perception acide. Cependant, il est à noter qu'une petite proportion des produits acides et amers peut occasionnellement entraîner une perception opposée.

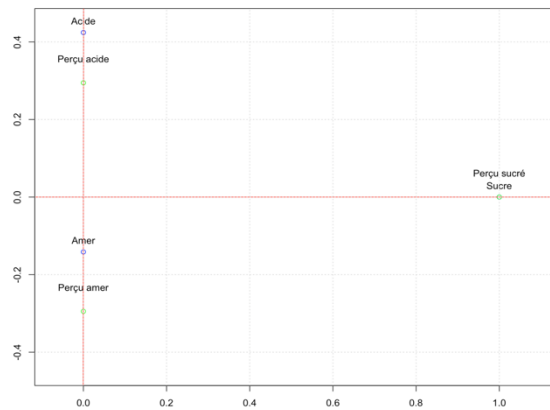


Figure 7 : Superposition des projections

Avec la fonction CA de FactoMineR, il est possible de réaliser l'ACF directement, cette fonction nous renvoie ce graphique :

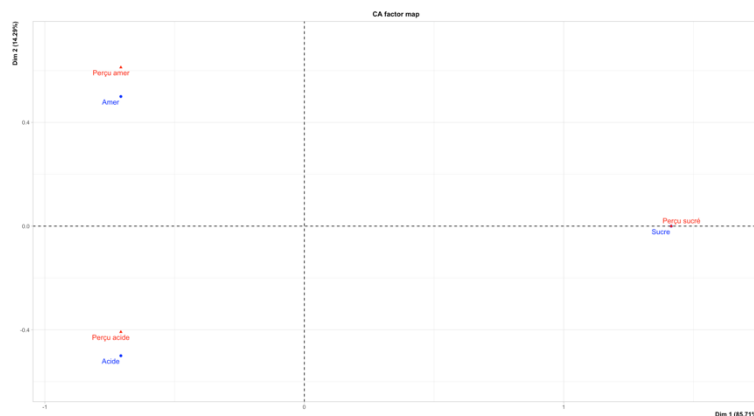


Figure 8 : ACF avec la fonction CA()

Dans le graphique, les coordonnées des points diffèrent de nos précédentes coordonnées en raison d'un ajustement du barycentre des profils selon les colonnes et les lignes dans la fonction CA. En effet pour l'affichage avec la fonction CA(), il procède à une opération de transition barycentrique qui permet de s'assurer que les deux soit projeté dans les mêmes dimensions. Cependant, cette variation n'impacte pas notre analyse, car le résultat global demeure le même. Néanmoins pour l'analyse de la dernière partie nous avons décidé de nous baser sur le graphique renvoyé par la fonction de FactoMineR.

Analyse de la base de travail

Pour effectuer une analyse plus poussée de la base de données de travail, on va utiliser le graphique obtenu grâce à la fonction CA() de FactomineR et on décide d'analyser l'adéquation entre temps de travail et travail à réaliser en choisissant les variables Fonctions et temps de travail.

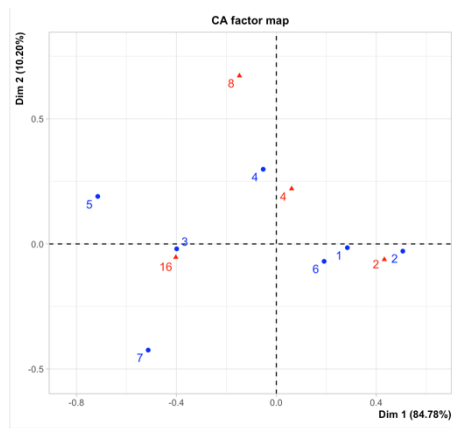


Figure 9 : ACF avec la fonction CA()
de la base de travail.

Sur le plan graphique, plusieurs observations se dégagent. Les ingénieurs et les cadres semblent avoir les charges de travail les plus élevées, tandis que les techniciens et les travailleurs administratifs semblent maintenir un équilibre entre leur charge de travail et leurs horaires. Aucun lien spécifique n'est observable entre une profession donnée et une sous-charge de travail déclarée. Les contractuels ne montrent pas de tendance claire en termes de charge de travail, cela varie considérablement. Enfin, les techniciens supérieurs semblent ne pas dépasser leurs horaires de travail.

En ce qui concerne le traitement de plus de deux mobilités simultanément, cette approche n'est pas envisageable. Les projections se déroulant dans des dimensions distinctes, toute comparaison perdrait sa signification. Il est donc nécessaire de traiter les mobilités deux par deux pour une analyse pertinente.