

TD2 - Analyse factorielle des correspondances et analyse des correspondances multiples

MAF 2025-2026

Ce TD a pour but de reprendre les notions de base de l'analyse factorielle des correspondances (AFC) et de l'analyse des correspondantes multiples vues en cours mais à partir d'exemples simples. Ce TD permet aussi de voir les commandes principales utiles pour mettre en oeuvre ce type de méthode.

1 Dégustation de fromages

On s'intéresse à 8 fromages AOP, évalués par un panel de dégustateurs à l'aide de 7 descripteurs sensoriels. Pour chaque fromage, les dégustateurs devaient choisir le descripteur jugé le plus caractéristique.

Les descripteurs sont doux, salé, fruité, piquant, crèmeux, sec, odorant

Les fromages sont Comté, Beaufort, Camembert, Roquefort, Munster, Chèvre frais, Reblochon, Emmental

Le tableau suivant donne, pour chaque fromage, le nombre de fois où chaque descripteur a été choisi.

1.1 Tableau 1 : Fromages × descripteurs

| | doux | salé | fruité | piquant | crèmeux | sec | odorant |
|--------------|------|------|--------|---------|---------|-----|---------|
| Emmental | 8 | 3 | 10 | 0 | 9 | 1 | 2 |
| Beaufort | 7 | 4 | 8 | 0 | 8 | 2 | 1 |
| Camembert | 5 | 6 | 1 | 1 | 7 | 0 | 6 |
| Roquefort | 0 | 9 | 0 | 10 | 1 | 0 | 9 |
| Munster | 1 | 7 | 0 | 8 | 2 | 0 | 10 |
| Chèvre frais | 9 | 1 | 4 | 0 | 11 | 0 | 1 |
| Reblochon | 4 | 6 | 3 | 1 | 5 | 2 | 5 |
| Comté | 6 | 5 | 2 | 0 | 3 | 6 | 1 |

On note ce tableau de contingence ($X = (x_{ij})$) avec

- ($i = 1, \dots, I$) les fromages,
- ($j = 1, \dots, J$) les descripteurs.

- En reprenant les notations du cours combien, donner les valeurs de k_{1+} , k_{+2} , n , k_{12} , f_{1+} , f_{+2} , f_{ij}
- Sous l'hypothèse où le type de fromage et le descripteur sont deux variables indépendantes, quelle relation devrait vérifier les fréquences ?

1.2 Profils lignes

- Donner pour chaque descripteur le nombre de fois où il a été choisi.
- Rappeler la forme de la matrice de données utilisées pour étudier les profils lignes.
- Rappeler le poids attribué à chaque individu (chaque ligne).
- Déterminer le profil ligne moyen.
- Rappeler la métrique utilisée et calculer l'inertie.
- Quel est le nombre d'axes principaux maximum ?

1.3 Profils colonnes

Pour vérifier votre compréhension, essayer de répondre aux mêmes questions sur les profils colonnes.

1.4 Mise en oeuvre de l'AFC

- A quelle valeur correspond la somme des valeurs propres ?
- Dans quel cas, une valeur propre vaut 1 ?
- Voici les valeurs propres obtenues sur les données de dégustation fromage, Combien d'axes avez-vous envie de considérer ?

| | eigenvalue | percentage of variance | cumulative percentage of variance |
|-------|--------------|------------------------|-----------------------------------|
| dim 1 | 0.4925864639 | 71.29381599 | 71.29382 |
| dim 2 | 0.1342574414 | 19.43156385 | 90.72538 |
| dim 3 | 0.0430350236 | 6.22861422 | 96.95399 |
| dim 4 | 0.0178732104 | 2.58685423 | 99.54085 |
| dim 5 | 0.0029255339 | 0.42342308 | 99.96427 |
| dim 6 | 0.0002468579 | 0.03572864 | 100.00000 |

1.5 Analyse des profils colonnes

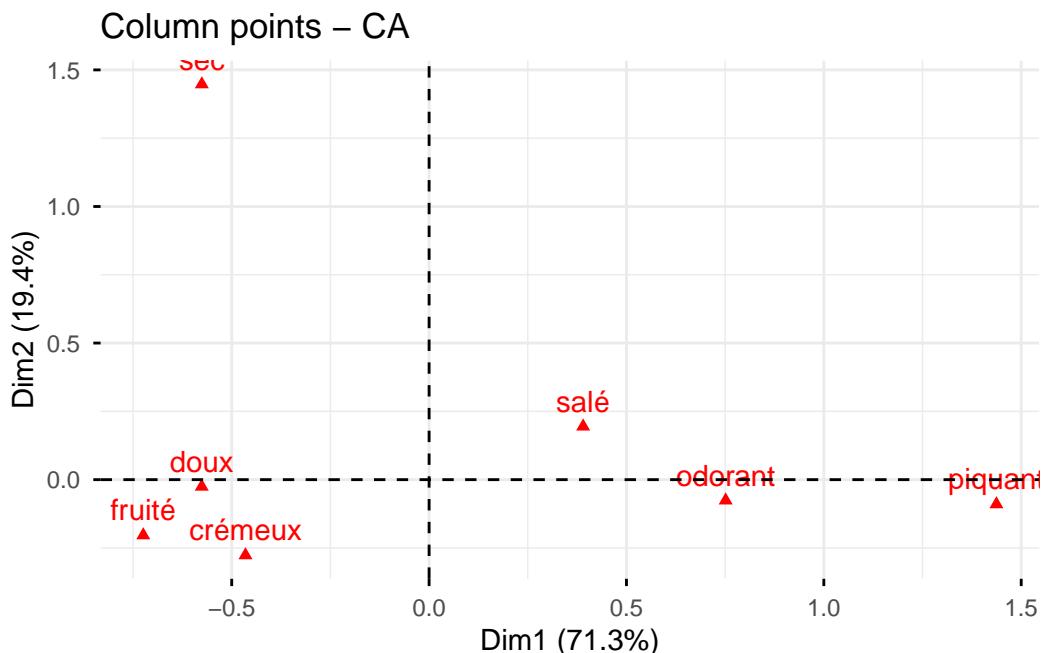
- Les contributions à la construction des axes de chacun des descripteurs sont données ci-dessous. Quels sont les descripteurs les plus contributifs pour le premier axe, pour le second ?

| | Dim 1 | Dim 2 | Dim 3 | Dim 4 | Dim 5 | Dim 6 |
|---------|-----------|------------|------------|-----------|------------|------------|
| doux | 12.187581 | 0.0877877 | 7.0322716 | 12.879109 | 1.8845841 | 47.8291191 |
| salé | 5.736433 | 5.2345594 | 0.6389908 | 16.237836 | 50.1845312 | 3.4156131 |
| fruité | 13.485453 | 3.9072547 | 62.0039410 | 7.062007 | 0.8643095 | 0.0073517 |
| piquant | 37.949380 | 0.5451211 | 12.3265268 | 39.561058 | 0.2088129 | 0.3593277 |
| crémeux | 9.143291 | 11.8691832 | 12.4980939 | 2.651583 | 2.4058488 | 40.6175206 |
| sec | 3.349047 | 77.6759893 | 0.3411195 | 1.734490 | 4.1735057 | 7.7484730 |
| odorant | 18.148815 | 0.6801046 | 5.1590565 | 19.873917 | 40.2784077 | 0.0225948 |

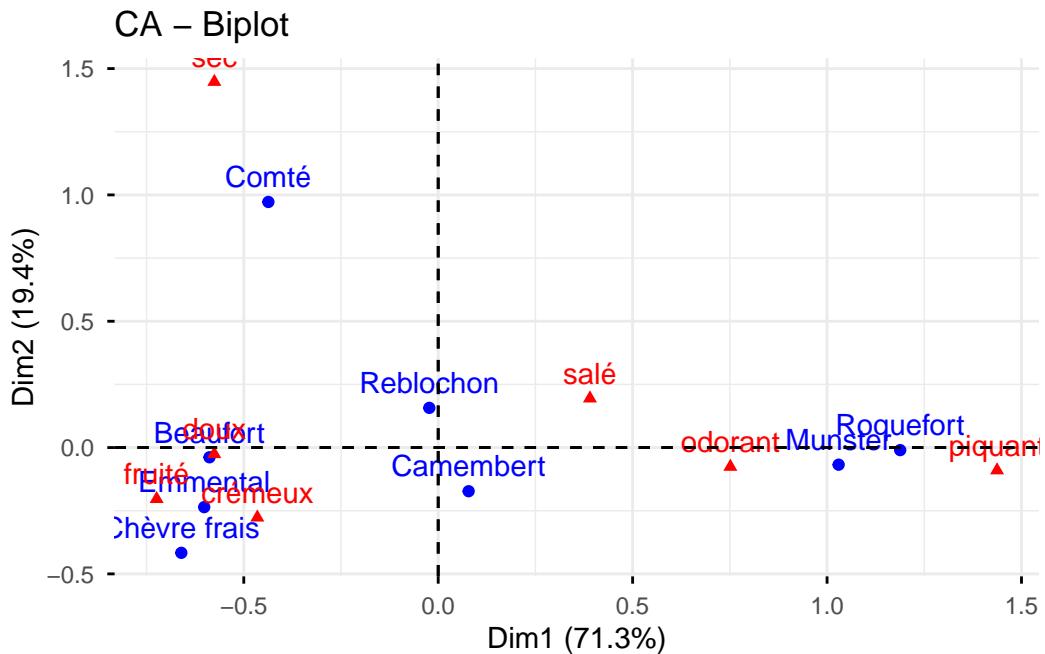
13. Le tableau ci dessous donne la qualité de la représentation. Identifier un descripteur peu contributif à la création de l'axe 1 et pourtant bien représenté.

| | Dim 1 | Dim 2 | Dim 3 | Dim 4 | Dim 5 | Dim 6 |
|---------|-----------|-----------|-----------|-----------|-----------|-----------|
| doux | 0.9144098 | 0.0017952 | 0.0460955 | 0.0350614 | 0.0008398 | 0.0017984 |
| salé | 0.7075103 | 0.1759652 | 0.0068853 | 0.0726673 | 0.0367607 | 0.0002111 |
| fruité | 0.6666470 | 0.0526451 | 0.2677868 | 0.0126671 | 0.0002538 | 0.0000002 |
| piquant | 0.9344436 | 0.0036585 | 0.0265173 | 0.0353457 | 0.0000305 | 0.0000044 |
| crémeux | 0.6722481 | 0.2378503 | 0.0802806 | 0.0070738 | 0.0010506 | 0.0014966 |
| sec | 0.1359107 | 0.8591623 | 0.0012094 | 0.0025540 | 0.0010059 | 0.0001576 |
| odorant | 0.9191485 | 0.0093879 | 0.0228269 | 0.0365209 | 0.0121152 | 0.0000006 |

14. On obtient la projection suivante pour les profils colonnes dans le premier plan principal. Que pouvez-vous dire des liens entre les descripteurs ?



15. A votre avis où devrait-on trouver le munster ?
16. Voici la représentation conjointe des profils lignes et colonnes, rédiger 5 lignes d'analyse obtenue à partir de cette représentation.



1.6 Variables supplémentaires

17. Selon vous, à quel endroit du plan s'attend-on à trouver un crottin de chèvre bien sec, le livarot, le caprice des Dieux ?

2 Analyse des activités de Loisir

La base loisirs est un jeu de données individuel destiné à l'étude des comportements de pratique des loisirs et de leurs déterminants socio-démographiques. Chaque observation correspond à un individu, et les variables renseignent :

les pratiques de loisirs Un ensemble de variables binaires ou qualitatives décrivant la participation à différentes activités culturelles, récréatives ou domestiques : Lecture, Écouter.musique, Cinéma, Spectacle, Exposition, Ordinateur, Sport, Marche, Voyage, Jouer.musique, Collection, Activité.bénévole, Bricolage, Jardinage, Tricot, Cuisine, Pêche, TV (niveau d'utilisation de 0 à 4).

Des indications socio-démographiques Sexe, Âge, Situation.matrimoniale et Profession fournissent des informations permettant de caractériser les individus et d'expliquer les variations de comportements de loisirs. Il y a quelques valeurs manquantes.

Une Variable synthétique Nb.activités est une variable dérivée indiquant, pour chaque individu, le nombre total d'activités pratiquées parmi celles listées.

Ces données sont disponibles sur le site de [Francois Husson](#)

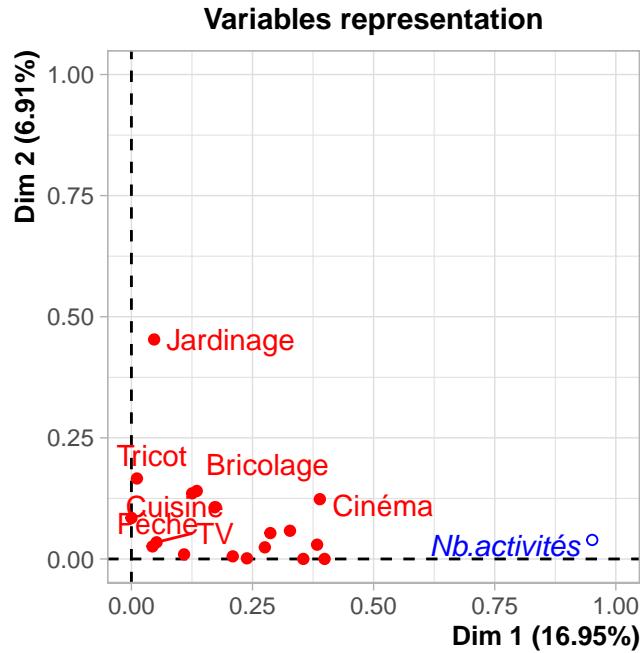
1. On cherche à construire une visualisation de cette base, permettant de mettre en évidence des liens potentiels entre différentes activités. Quelle est la méthode adaptée ?

2. Rappeler les éléments clés pour construire une ACM (tableau de données, métrique et poids).
3. L'analyse des correspondances multiples est mise en oeuvre à l'aide de la fonction MCA ci-dessous. Seules les activités de loisirs sont utilisées pour construire ACM. Indiquez toutes les variables supplémentaires.
4. A partir du graphique d'inertie ci-dessous, discuter la répartition de l'inertie sur les 2 premières dimensions.

| | eigenvalue | percentage of variance | cumulative percentage of variance |
|--------|------------|------------------------|-----------------------------------|
| dim 1 | 0.1977116 | 16.946704 | 16.94670 |
| dim 2 | 0.0806491 | 6.912781 | 23.85948 |
| dim 3 | 0.0720218 | 6.173298 | 30.03278 |
| dim 4 | 0.0628724 | 5.389066 | 35.42185 |
| dim 5 | 0.0584600 | 5.010860 | 40.43271 |
| dim 6 | 0.0558124 | 4.783924 | 45.21663 |
| dim 7 | 0.0555234 | 4.759147 | 49.97578 |
| dim 8 | 0.0533082 | 4.569278 | 54.54506 |
| dim 9 | 0.0530444 | 4.546664 | 59.09172 |
| dim 10 | 0.0491301 | 4.211153 | 63.30288 |
| dim 11 | 0.0464933 | 3.985139 | 67.28802 |
| dim 12 | 0.0450743 | 3.863513 | 71.15153 |
| dim 13 | 0.0435131 | 3.729690 | 74.88122 |
| dim 14 | 0.0433607 | 3.716628 | 78.59785 |
| dim 15 | 0.0408000 | 3.497142 | 82.09499 |
| dim 16 | 0.0379816 | 3.255564 | 85.35055 |
| dim 17 | 0.0373349 | 3.200134 | 88.55069 |
| dim 18 | 0.0362192 | 3.104506 | 91.65519 |
| dim 19 | 0.0349682 | 2.997272 | 94.65246 |
| dim 20 | 0.0323421 | 2.772180 | 97.42464 |
| dim 21 | 0.0300458 | 2.575355 | 100.00000 |

2.1 Représentation des variables

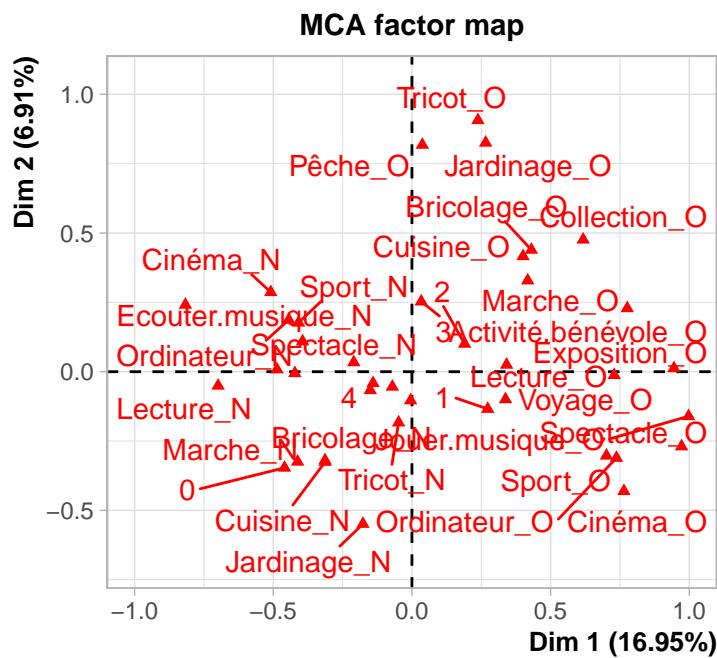
5. Expliquer ce qui est représenté sur le graphique suivant et discuter du rôle particulier de la variable quantitative supplémentaire.



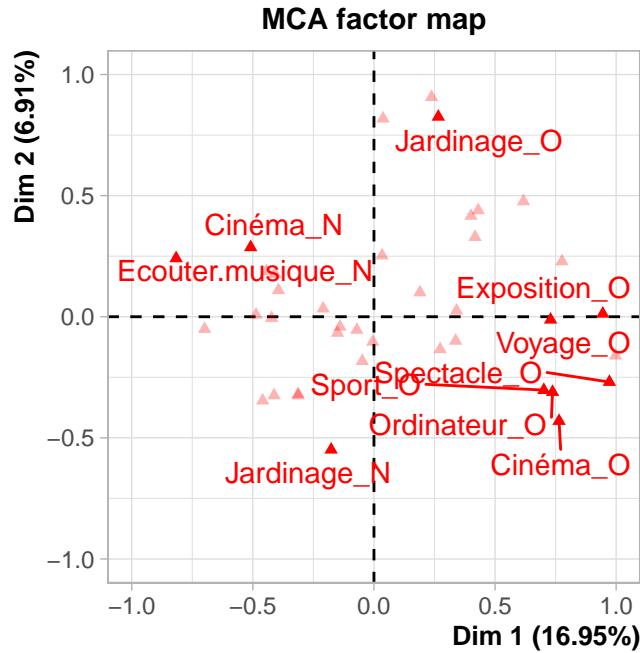
6. Quelles sont les variables les mieux représentées sur l'axe 1, l'axe 2 ?

2.2 Représentation des différentes modalités.

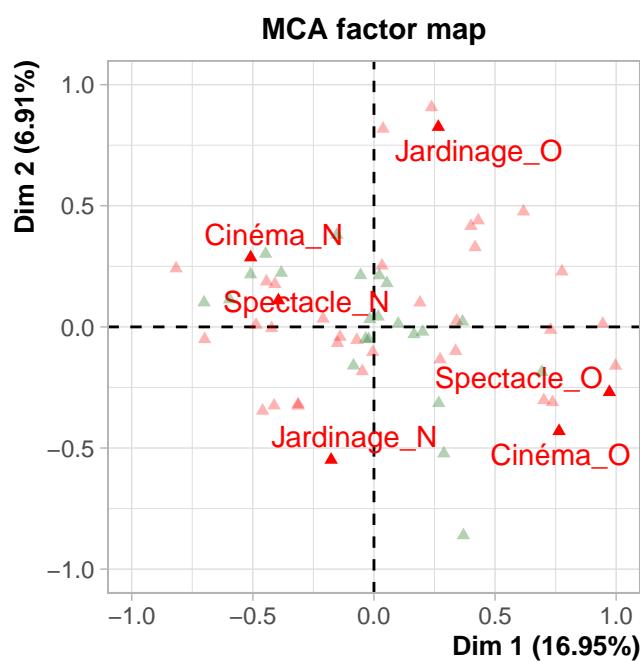
Le graphique ci-dessous représente l'ensemble des modalités



7. Pour y voir plus clair, on s'intéresse au 10 variables les plus contributives, quel enseignement pouvez-vous tirer de ce graphique ?



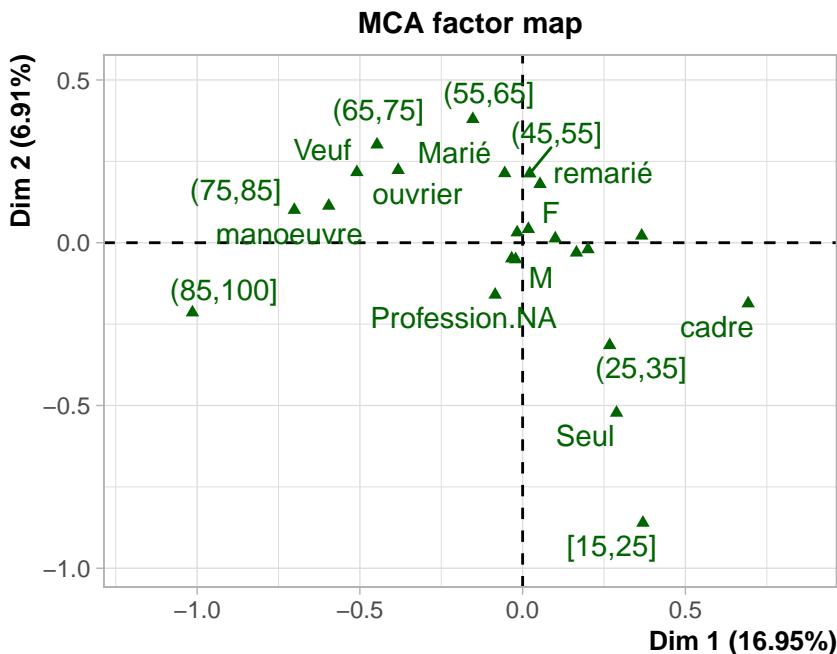
8. Peut-on affirmer que les personnes qui vont à des expositions ont également tendance à aller au spectacle ?
9. Les personnes adeptes du jardinage passent-elles peu de temps sur leur ordinateur ?
10. A votre avis quelles sont les caractéristiques démographiques des personnes adeptes du Jardinage ?
11. Dans le graphique ci-dessous, on ne représente que les modalités dont le \cos^2 est supérieur à 0.4. Expliquer de quel \cos^2 il est question. Comparer avec le graphique précédent. Est ce surprenant ?



2.3 Prise en compte de variables supplémentaires.

On s'intéresse à la variable supplémentaire Age.

12. Selon votre intuition à la question 10, où devrait-on retrouver les personnes de 65 à 75 ans ?



13. Que constatez-vous pour la variable Age ?
14. Le tableau ci-dessous donne la qualité de représentation pour chacune des modalités de la variable Age. Quelle information peut-on en tirer ?
- | | | | | | |
|-------------|-------------|-------------|-------------|-------------|-------------|
| (25,35] | (35,45] | (45,55] | (55,65] | (65,75] | (75,85] |
| 0.031285680 | 0.009937587 | 0.012782877 | 0.029487348 | 0.036519518 | 0.030560953 |
15. On refait l'analyse en gardant plus de dimensions principales (argument ncp de MCA) et on illustre la qualité de la représentation des (25–35] pour chaque axe principal. Comment expliquer cette qualité de représentation ?
16. Quel est le plan le plus pertinent pour bien visualiser les sondés de la catégories (25–35] ?
17. Dans le graphique ci-dessous, seules les modalités dont le cos2 sont supérieures à 0.2 sont représentées. Discuter ce que vous voyez en 5 lignes max.

