

TP4 - Analyse des correspondances multiples

Guillaume Bernard-Reymond et Lorenzo Gaggini

Janvier 2024

Dans ce TP, nous avons à notre disposition un jeu données concernant 300 consommateurs de thé ayant répondu à un questionnaire de 36 questions dont les réponses sont qualitatives, hormis une concernant l'âge.

Nous avons décidé d'utiliser la bibliothèque FactomineR pour son confort d'utilisation. On trouvera le code R utilisé en appendice.

1. L'ACM

(a) Description des commandes

Les 18 premières variables correspondent aux modes de consommation du thé, la 19-ème est l'âge des personnes. Enfin les 17 dernières variables nous informent sur la personne et sa perception du thé. On met donc en relation le mode consommation avec l'individu et sa perception du thé. C'est ce qui justifie la mise en supplémentaire de ces 17 variables.

(b) Les deux premières dimensions captent 17,988% de la part de variance totale comme l'illustre le tableau ci-dessous :

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5	Dim.6	Dim.7
Variance	0.148	0.122	0.090	0.078	0.074	0.071	0.068
% of var.	9.885	8.103	6.001	5.204	4.917	4.759	4.522
Cumulative % of var.	9.885	17.988	23.989	29.192	34.109	38.868	43.390

De plus sur le graphique des valeurs propres, on peut observer un décrochement après la deuxième :

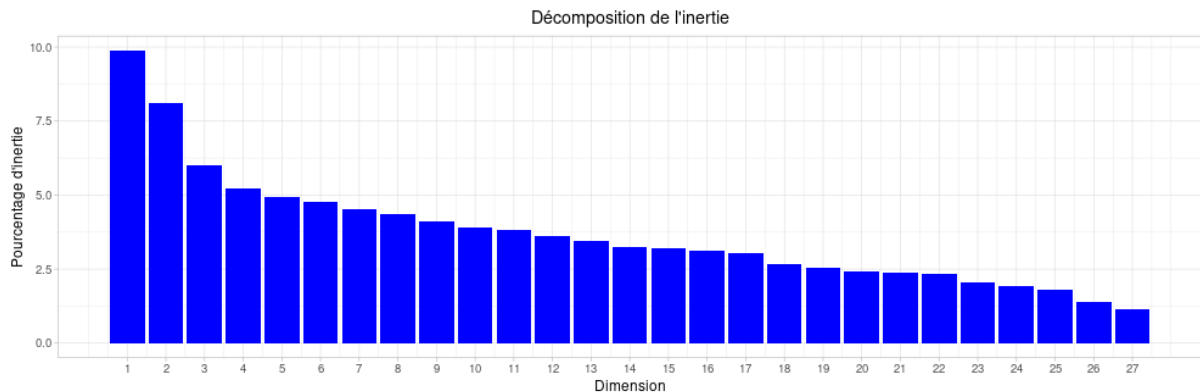
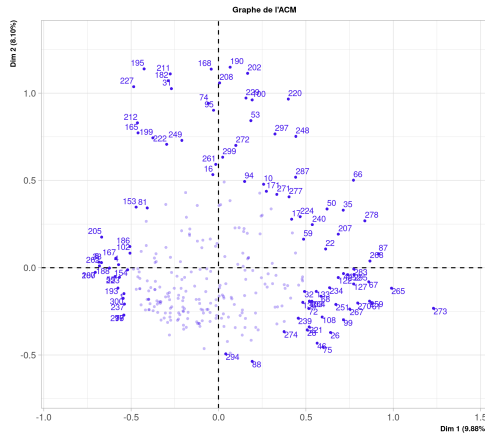
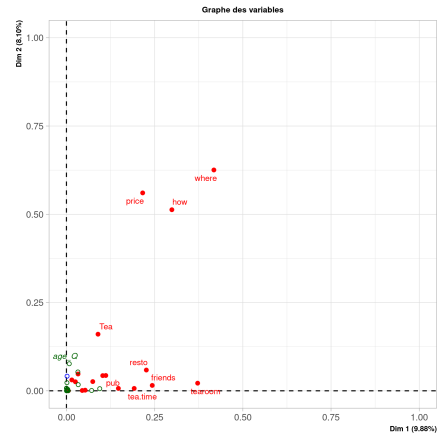


Tableau des individus les plus contributeurs pour l'axe 1 :

Côté -				Côté +			
Individus	Coordonnées	CTR	COS2	COS2	CTR	Coordonnées	Individus
262	-0.71	5.55	0.51	0.75	8.19	0.86	265
205				0.71	7.85	0.85	273



Graphique des 100 individus les plus contributeurs



Graphique des variables

Pour ce qui est des variables, les plus liées à la première dimension sont : "where" ($R^2 = 0.418$); "tearoom" ($R^2 = 0.372$) et "how" ($R^2 = 0.299$)

Dimension 2

Sur le second, on peut voir s'opposer des individus (168; 190) en haut et (88; 294) en bas. Toutefois, on ne distingue pas d'individus particulier qui aurait contribué de façon majeure au positionnement de cette axe.

Du côté des variables, on retrouve deux des trois variables précédentes : "where" ($R^2 = 0.626$), "price" ($R^2 = 0.561$) et "how" ($R^2 = 0.513$).

2. Etude des individus

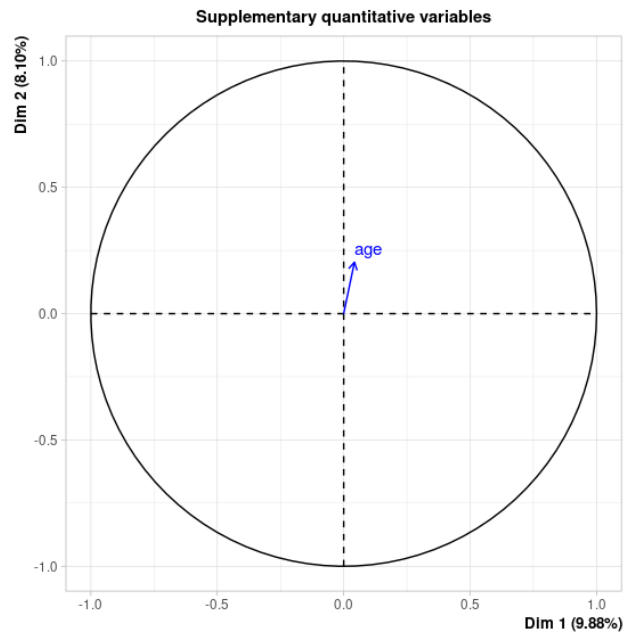
- Aucun groupe ne se dégage de manière suffisamment significative pour être interprété. Le nuage est étalé de manière homogène.
- Les individus atypiques sont ceux remarqués à la question 1)b) à savoir les individus 273 et 265. Ce sont des consommateurs buvant du thé de manière régulière sans contexte particulier. L'individu 205 illustre en opposition sur cet axe le mode de consommation.

On voit sur le tableau suivant l'opposition des individus 273 et 205.

Individu 273	Individu 205
breakfast	Not.breakfast
tea time	Not.tea time
evening	Not.evening
Not.lunch	Not.lunch
Not.dinner	dinner
always	Not.always
home	home
work	Not.work
tearoom	Not.tearoom
friends	friends
resto	Not.resto
pub	Not.pub
Earl Grey	green
other	alone
sugar	No.sugar
tea bag+unpackaged	tea bag
chain store+tea shop	chain store
p_variable	p_branded

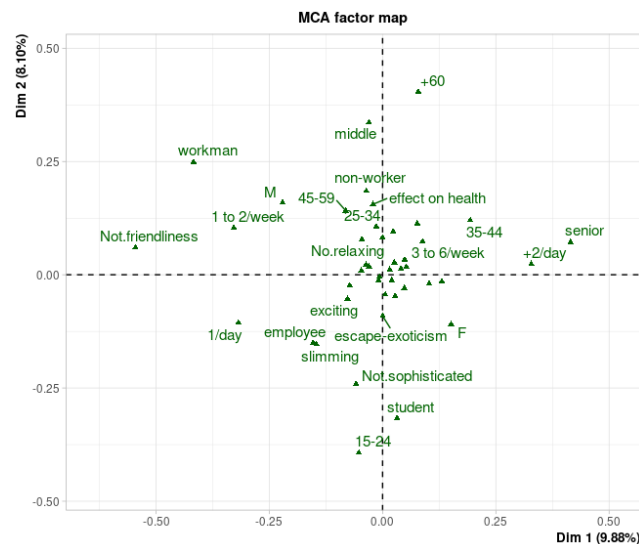
3. Etude des variables et des modalités :

- (a) Les modalités opposées le long du premier axes sont : "tea room, chain store tea-shop" à "not friends, not tearoom, not tea-time". Cela appuie les oppositions relevées par les individus atypiques relevés précédemment entre les buveurs de thé réguliers et ceux ne l'étant pas.
- (b) La seconde dimension oppose "teashop, unpackaged, p-upscale" à "p-unknow". C'est donc un mode de consommation propre à des amateurs avertis de thé.
- (c) Le graphique ci-dessous montre que l'âge n'est pas bien représenté le long de la dimension 2. Sa corrélation est seulement de 0,2.



On comprend alors l'opposition entre les jeunes consommateurs et les plus âgés. Les premiers achètent du thé sans nécessairement se soucier de la qualité du thé consommé. A la différence des personnes plus âgées qui ont tendance à s'orienter vers des magasins spécialisés vendant du thé haut de gamme.

- (d) Voici le graphique des modalités illustratives :



On peut observer que les modalités de la variable age-Q sont ordonnées de bas en haut, des plus jeunes aux plus âgés. Ceci corrobore l'orientation de la variable "âge" le long de l'axe 2. "Age-Q" et "age" étant deux représentations différentes d'une même quantité intrinsèque aux individus, elles sont en toute cohérence semblables.

4. **Ellipse de confiance**

qu'est ce que l'on doit faire ???????