Analyse des données
Filière: IAGI

Année universitaire: 2021-2022

AZMI Mohamed

Mohamed.azmi@ensam-casa.ma

Chapitre III : ACM

Analyse des Correspondances Multiples

Introduction:
L'Analyse des Correspondances Multiples

- L'Analyse des Correspondances Multiples (ACM) permet d'étudier une population de **I** individus décrits par **J** variables qualitatives.
- L'application la plus courante de l'ACM est **le traitement de l'ensemble des réponses à une enquête**. Chaque question constitue une variable dont les modalités sont les réponses proposées (parmi lesquelles chaque enquêté doit faire un choix unique).

	Question_1				Question_2				…				Question_J				
	Rép_{11}	Rep_{12}	…	Rép_{1k_1}	Rép_{21}	Rep_{22}	…	Rép_{1k_2}	…	…	…	…	$\text{Rép}_{\text{J}1}$	$\text{Rep}_{\text{J}2}$	…	Rép_{1k_J}	
Ind1																	J
Ind2																	J
Ind3																	J
…																	J
IndI	I_1	I_2	…	…	…	…	…	…	…	…	…	…	…	…	…	I_K	I$\times J$

Introduction:
L'Analyse des Correspondances Multiples

	Var$_1$ =Question$_1$				Var$_2$ =Question$_2$				…				Var$_J$ =Question$_J$				
	Rép$_{11}$	Rep$_{12}$	…	Rép$_{1k_1}$	Rép$_{21}$	Rep$_{22}$	…	Rép$_{1k_2}$	…	…	…	…	Rép$_{J1}$	Rep$_{J2}$	…	Rép$_{1k_J}$	
Ind1	x_{11}	x_{12}	…	…	…	…	…	…	…	…	…	…	…	…	…	x_{1K}	J
Ind2	x_{21}	x_{22}	…	…	…	…	…	…	…	…	…	…	…	…	…	x_{2K}	J
Ind3	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	J
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	J
IndI	I_1	I_2	…	…	…	…	…	…	…	…	…	…	…	…	…	I_K	I$\times J$

- $\begin{cases} x_{ik} = 1 \text{ ;si l'individu i choisit la catégorie k} \\ x_{ik} = 0 \text{ ; } sinon \end{cases}$ / i varie de 1 à I et k de 1 à K.
- Nous considérons que la variable j (Question$_j$) possède k_j catégories. J varie de 1 à J
- $\sum_{j=1}^{J} k_j = K$ Le nombre total des catégories égale à K
- $\sum_{k=1}^{\boldsymbol{K}} x_{ik} = J$ Le nombre total des catégories choisies par chaque individu égale au nombre de questions J

Introduction:

Objectifs

La problématique de l'ACM peut être considérée comme :

- Apparentée à celle de l'ACP (étude d'un tableau où les lignes représentent des Individus et les colonnes représentent des Variables)
- Une généralisation de celle de l'AFC (étude de la liaison entre plusieurs variables qualitatives).

Ces deux aspects sont toujours plus ou moins explicitement présents dans les objectifs de l'ACM, présentés ici à partir des trois familles d'objets qui interviennent en ACM : les individus, les variables et les modalités des variables.

L'étude des modalités permet à la fois d'étudier **les liaisons entre les variables** et d'aborder **la typologie des individus** en examinant le comportement moyen de classes d'individus (chaque modalité représente une classe regroupant les individus ayant cette modalité).

ACM

Etude des individus

- Étudier les individus signifie comprendre **les similitudes entre les individus** en termes de toutes les variables. En d'autres termes, **fournir une typologie des individus** : quels sont les individus les plus semblables (et les plus dissemblables) ?

- Les individus sont comparés sur la base de la présence-absence des catégories qu'ils ont sélectionnées. De ce seul point de vue, la distance entre deux individus dépend entièrement de leurs caractéristiques et non de celles des autres individus. **Cependant, il est important de tenir compte des caractéristiques des autres individus lors du calcul de cette distance.**

ACM

Etude des individus

- Exemples pour comprendre comment on peut calculer la distance entre deux individus:

  1. Si deux personnes choisissent les **mêmes catégories**, la distance qui les sépare devrait être **nulle**.
  2. Si deux personnes choisissent toutes les deux **un grand nombre de catégories**, elles devraient être **proches** l'une de l'autre.
  3. Si deux personnes choisissent toutes les mêmes catégories, **sauf une qui est choisie par l'une des personnes et rarement par toutes les autres personnes,** elles devraient être **distancées** pour tenir compte du **caractère unique** de l'une des deux.
  4. Si deux individus **partagent une catégorie rare**, ils devraient être **proches** l'un de l'autre malgré leurs différences, afin de rendre compte de leur caractère distinctif commun.

=> les individus doivent être comparés catégorie par catégorie tout en tenant compte de la rareté ou du caractère universel de cette catégorie.

ACM

Etude des variables et des catégories

- Comme dans ACP et l’AFC, l’objectif est de résumer les relations entre les variables.
- Dans l’ACM, nous nous concentrons principalement sur l’étude des catégories, car les **catégories** représentent à la fois les **variables** et des **groupes d’individus** (groupe = tous les individus qui choisissent une catégorie).

- Une façon de comparer deux catégories est de compter les individus qui choisissent les deux catégories.
  - **Moins deux catégories ont d’individus en commun, plus elles sont éloignées (en termes de distance).**
  - On désigne par $\boldsymbol{I_{K_1 \neq K_2}}$ le nombre d’individus qui ont soit la catégorie K_1, soit la catégorie K_2.

ACM

Etude des variables et des catégories

- Il est important de tenir compte de la taille de chaque groupe d'individus lors du calcul de la distance $\boldsymbol{I_{K_1 \neq K_2}}$.
- Prenons un exemple avec trois catégories k_1, k_2 et k_3, chacune composée de 10, 100 et 100 individus, respectivement. (I_{K_1} = 10, I_{K_2}=100, I_{K_3}=100)
  - Si les catégories k_1 et k_2 ne partagent aucun individu, $I_{K_1 \neq K_2}$ = 110.
  - Si les catégories k_2 et k_3 partagent 45 individus $I_{K_2 \neq K_3}$ = 110.
  - Cependant, dans le 1er cas k_1 et k_2 se partagent 0% des individus alors que dans le 2ème k_2 et k_3 se partagent 45% des individus. Les catégories k_1 et k_2 doivent être plus éloignées que les catégories k_2 et k_3 .

ACM

Etude des variables et des catégories

L'étude des variables consiste à résumer l'ensemble des variables (qualitatives) par un petit nombre **de variables numériques (Facteurs synthétiques)**.

Par exemple, on peut chercher à résumer un ensemble de variables socio-professionnelles par un indicateur de « statut social ».
L'intérêt de ces variables synthétiques provient de ce qu'elles sont liées à l'ensemble des variables étudiées. Ainsi, une variable synthétique ne pourra être considérée comme un indicateur de « statut social » que si elle est liée à la fois à la catégorie socio-professionnelle, au type de diplôme, etc.

Remarque. Par rapport à l'ACP, on cherche, selon ce point de vue, une variable quantitative pour synthétiser un ensemble de variables qualitatives (et non quantitatives) ce qui implique, d'une façon ou d'une autre, d'affecter un coefficient à chaque modalité de chaque variable ; pour un individu, la valeur de la variable synthétique est alors la somme des coefficients des modalités qu'il possède.

ACM

Nuage des individus

	Var_1 =$Question_1$				Var_2 =$Question_2$				…				Var_J =$Question_J$				
	$Rép_{11}$	Rep_{12}	…	$Rép_{1k_1}$	$Rép_{21}$	Rep_{22}	…	$Rép_{1k_2}$	…	…	…	…	$Rép_{J1}$	Rep_{J2}	…	$Rép_{1k_J}$	
Ind1	$\frac{x_{11}}{J}$	$\frac{x_{12}}{J}$	…	…	…	…	…	…	…	…	…	…	…	…	…	$\frac{x_{1k}}{J}$	1
Ind2	$\frac{x_{21}}{J}$	$\frac{x_{22}}{J}$	…	…	…	…	…	…	…	…	…	…	…	…	…	$\frac{x_{2k}}{J}$	1
Ind3	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	1
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	1
IndI	$\frac{I_1}{\text{I} \times J}$	$\frac{I_2}{\text{I} \times J}$	…	…	…	…	…	…	…	…	…	…	…	…	…	$\frac{I_k}{\text{I} \times J}$	1

$$\boldsymbol{d^2(i,l) = \sum_{k=1}^{K} \frac{\left(\frac{x_{ik}}{J} - \frac{x_{lk}}{J}\right)^2}{\frac{I_k}{\text{I} \times J}} = \frac{1}{J} \sum_{k=1}^{K} \frac{I}{I_k} \left(x_{ik} - x_{lk}\right)^2}$$

Le centre de gravité de ce nuage, noté G_I, a pour coordonnée $\frac{\text{I}_\text{k}}{\text{I} \times \text{J}}$ pour la modalité k. proportion, au coefficient J près, des individus ayant choisi la modalité k .

ACM

Nuage des individus

$$d^2(i,l) = \sum_{k=1}^{K} \frac{\left(\frac{x_{ik}}{J} - \frac{x_{lk}}{J}\right)^2}{\frac{I_k}{I \times J}} = \frac{1}{J} \sum_{k=1}^{K} \frac{I}{I_k} \left(x_{ik} - x_{lk}\right)^2$$

- L'expression $(x_{ik} - x_{lk})^2$ vaut 0 ou 1 et ne diffère de 0 que pour les modalités *k* possédées par un seul des deux individus *i* et *l*.
- La distance d(i , l) **croît avec le nombre de modalités qui diffèrent** pour les individus *i* et *l* (ce qui est logique !).
- Une modalité *k* intervient dans cette distance avec le poids I / I_k, i.e la **présence d'une modalité rare éloigne son ou ses possesseurs de tous les autres individus** (y compris le centre de gravité G_I).

ACM

Nuage des modalités

	Var$_1$ =Question$_1$				Var$_2$ =Question$_2$				…				Var$_J$ =Question$_J$				
	Rép$_{11}$	Rep$_{12}$	…	Rép$_{1k_1}$	Rép$_{21}$	Rep$_{22}$	…	Rép$_{1k_2}$	…	…	…	…	Rép$_{J1}$	Rep$_{J2}$	…	Rép$_{1k_J}$	
Ind1	$\frac{x_{11}}{I_1}$	$\frac{x_{12}}{I_2}$	…	…	…	…	…	…	…	…	…	…	…	…	…	$\frac{x_{1K}}{I_K}$	$\frac{J}{I \times J}=1/I$
Ind2	$\frac{x_{21}}{I_1}$	$\frac{x_{22}}{I_2}$	…	…	…	…	…	…	…	…	…	…	…	…	…	$\frac{x_{2K}}{I_K}$	$\frac{J}{I \times J}=1/I$
Ind3	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	$\frac{J}{I \times J}=1/I$
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	$\frac{J}{I \times J}=1/I$
Ind$_I$	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

$$d^2(k,h) = \sum_{i=1}^{I} \frac{\left(\frac{x_{ik}}{I_k} - \frac{x_{ih}}{I_h}\right)^2}{\frac{J}{I \times J}} = I \sum_{i=1}^{I} \left(\frac{x_{ik}}{I_k} - \frac{x_{ih}}{I_h}\right)^2$$

ACM

Nuage des modalités

$$d^2(k,h) = I\ \sum_{i=1}^{I}\left(\frac{x_{ik}}{I_k} - \frac{x_{ih}}{I_h}\right)^2$$

- Le centre de gravité du nuage des modalités, noté G_K, est caractérisé par un profil constant égal à $1/I$ (équivalent à une modalité que tous les individus auraient choisie).
- $d^2(k, G_k) = \frac{I}{I_k} - 1 \Rightarrow$ une modalité rare sera toujours loin du centre de gravité G_K
- Deux modalités d'une même variable sont obligatoirement assez éloignées l'une de l'autre.
- Deux modalités possédées par les mêmes individus sont confondues.
- Les modalités rares sont éloignées de toutes les autres.

ACM

Nuage des modalités

$d^2(k, G_k) = \frac{I}{I_k} - 1 \;\Rightarrow\;$ **Inertie de k par rapport à G_k = $\frac{I_k}{IJ} d^2(k, G_k) = \frac{1}{J}\left(1 - \frac{I_k}{I}\right)$**

- Le poids de la modalité k vaut $\frac{I_k}{IJ}$; il est proportionnel à l'effectif I_k.
- **Une modalité influence la construction des axes par l'intermédiaire de son inertie par rapport au centre de gravité.**
- Concrètement, il est courant de voir les premiers facteurs d'une ACM déterminés presque exclusivement par quelques modalités très rares partagées par les mêmes individus.
  - Par exemple, une modalité présente dans 1 % seulement de la population possède une inertie (c'est-à-dire une influence) presque deux fois plus grande qu'une modalité présente dans 50 % de la population.

ACM

Nuage des modalités

$$d^2(k, G_k) = \frac{I}{I_k} - 1 \qquad \text{Inertie de } k = \frac{1}{J}\left(1 - \frac{I_k}{I}\right)$$

- **L'inertie d'une variable j** est liée directement au nombre de ses modalités:

$$\text{Inertie}(\text{var}_j) = \sum_{k=1}^{K_j} Inertie(modalit\text{é}\ k) = \sum_{k=1}^{K_j} \frac{1}{J}\left(1 - \frac{I_k}{I}\right) = \frac{K_j - 1}{J}$$

- **L'inertie totale du nuage** des modalités est:

$$\text{Inertie}\ (N_k) = \sum_{j=1}^{J} \text{Inertie(varj)} = \sum_{j=1}^{J} \frac{K_j - 1}{J} = \frac{K}{J} - 1$$

- **Ainsi, l'inertie globale ne dépend que de la structure du questionnaire, et plus précisément du nombre moyen de catégories par variable**

ACM

Relations de transition et représentation simultanée

$$F_s(i) = \frac{1}{\sqrt{\lambda_s}} \sum_{k \in K} \frac{x_{ik}}{J} G_s(k)$$

L'individu i est placé, au coefficient $1/\sqrt{\lambda_s}$ près, au barycentre des modalités qu'il possède.

$$G_s(k) = \frac{1}{\sqrt{\lambda_s}} \sum_{i \in I} \frac{x_{ik}}{I_k} F_s(i)$$

La modalité k est placée, au coefficient $1/\sqrt{\lambda_s}$ près, au barycentre des individus qui la possèdent.

Dans l'étude de sa projection, on peut considérer une modalité aussi bien comme barycentre d'une classe d'individus (**Etude de ressemblance**) que comme indicatrice d'une variable (**Etude d'association**)

ACM

Relations de transition et représentation simultanée

En pratique, les deux notions de proximité (association ou ressemblance) s'utilisent conjointement :

- on interprète souvent la proximité entre modalités **de variables différentes** en tant **qu'association** de modalités
- et on interprète la proximité entre modalités **d'une même variable** en tant que **ressemblance** entre deux classes d'individus.

Par exemple, en décrivant un plan factoriel sur lequel apparaissent différents repères sociaux, on interprète

- la proximité entre les modalités *retraités* et *plus de 65 ans* en terme d'association (ce sont presque les mêmes individus qui possèdent ces deux modalités)
- la proximité entre *60 à 65 ans* et *plus de 65 ans* en terme de ressemblance (ces deux classes d'individus possèdent des caractéristiques identiques quant aux autres variables).

ACM

Synthèse des variables qualitatives

Ici, nous utilisons le rapport de corrélation pour mesurer la liaison entre une variable numérique (ici le facteur) et une variable qualitative.

$$Inertie_{\ Totale} = \ Inertie_{\ Inter} + Inertie_{\ Intra}$$

- $Inertie_{\ Totale} = \sum_{k=1}^{k_j} \sum_{i \, \epsilon k_j} (F_{ik} \ - \ \bar{F})^2$
- $Inertie_{\ Intra} = \sum_{k=1}^{k_j} \sum_{i \, \epsilon k_j} (F_{ik} \ - \bar{F}_{.k})^2$
- $Inertie_{\ Inter} = \sum_{k=1}^{k_j} \sum_{i \, \epsilon k_j} I_k (\bar{F}_{.k} \ - \ \bar{F})^2$

Le carré du rapport de corrélation est donné par : $\eta^2\left(var_j \, , F\right) = \frac{Inertie_{\ Inter}}{Inertie_{\ Totale}}$

ACM

Synthèse des variables qualitatives

$$\eta^2(var_j\,,F) = \frac{Inertie_{\ Inter}}{Inertie_{\ Totale}} = \frac{\sum_{k=1}^{k_j} \sum_{i\,\epsilon k_j} I_k(\bar{F}_{.k} - \bar{F})^2}{\sum_{k=1}^{k_j} \sum_{i\,\epsilon k_j} (F_{ik} - \bar{F})^2}$$

- Lorsqu'il est proche de 1, les individus d'une même classe sont très regroupés et les classes sont séparées les unes des autres : c'est une situation de liaison très forte entre la variable qualitative et la variable numérique.

- Lorsqu'il est proche de 0, les moyennes des classes sont très proches de la moyenne générale et les individus d'une même classe sont très dispersés : la variable qualitative et la variable numérique ne sont pas liées.

ACM

Synthèse des variables qualitatives

Exemple

Calculer les rapports de corrélation entre chacune des variables qualitative et le facteurs F1

Interpréter les résultats

$$\eta^2(var_j\,, F) = \frac{Inertie_{\ Inter}}{Inertie_{\ Totale}} = \frac{\sum_{k=1}^{k_j}\sum_{i\,\epsilon k_j} I_k(\bar{F}_{.k} \ - \ \bar{F})^2}{\sum_{k=1}^{k_j}\sum_{i\,\epsilon k_j}(F_{ik} \ - \ \bar{F})^2}$$

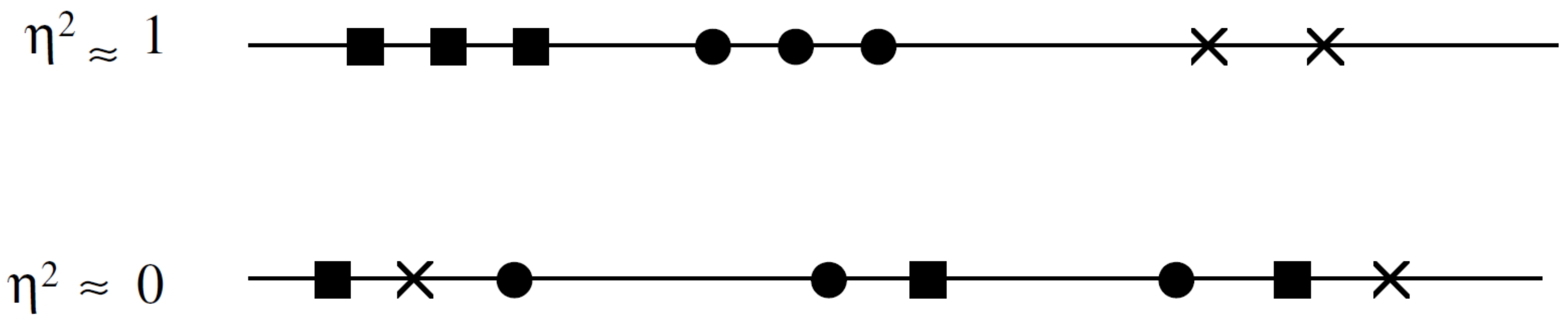
Individus	Genre	C S P	F1
1	F	Ouvrier non qualifié	1,2
2	F	Ouvrier non qualifié	1,1
3	F	Ouvrier qualifié	0,6
4	F	Ouvrier qualifié	0,5
5	F	Cadre	-0,8
6	H	Ouvrier non qualifié	0,9
7	H	Ouvrier qualifié	0,5
8	H	Ouvrier qualifié	0,4
9	H	Cadre	-0,96
10	H	Cadre	-1,6

ACM

Synthèse des variables qualitatives

Illustration des deux valeurs extrêmes du rapport de corrélation.

Huit individus, représentés par un symbole différent selon leur modalité pour une variable qualitative, figurent sur un axe représentant une variable numérique.

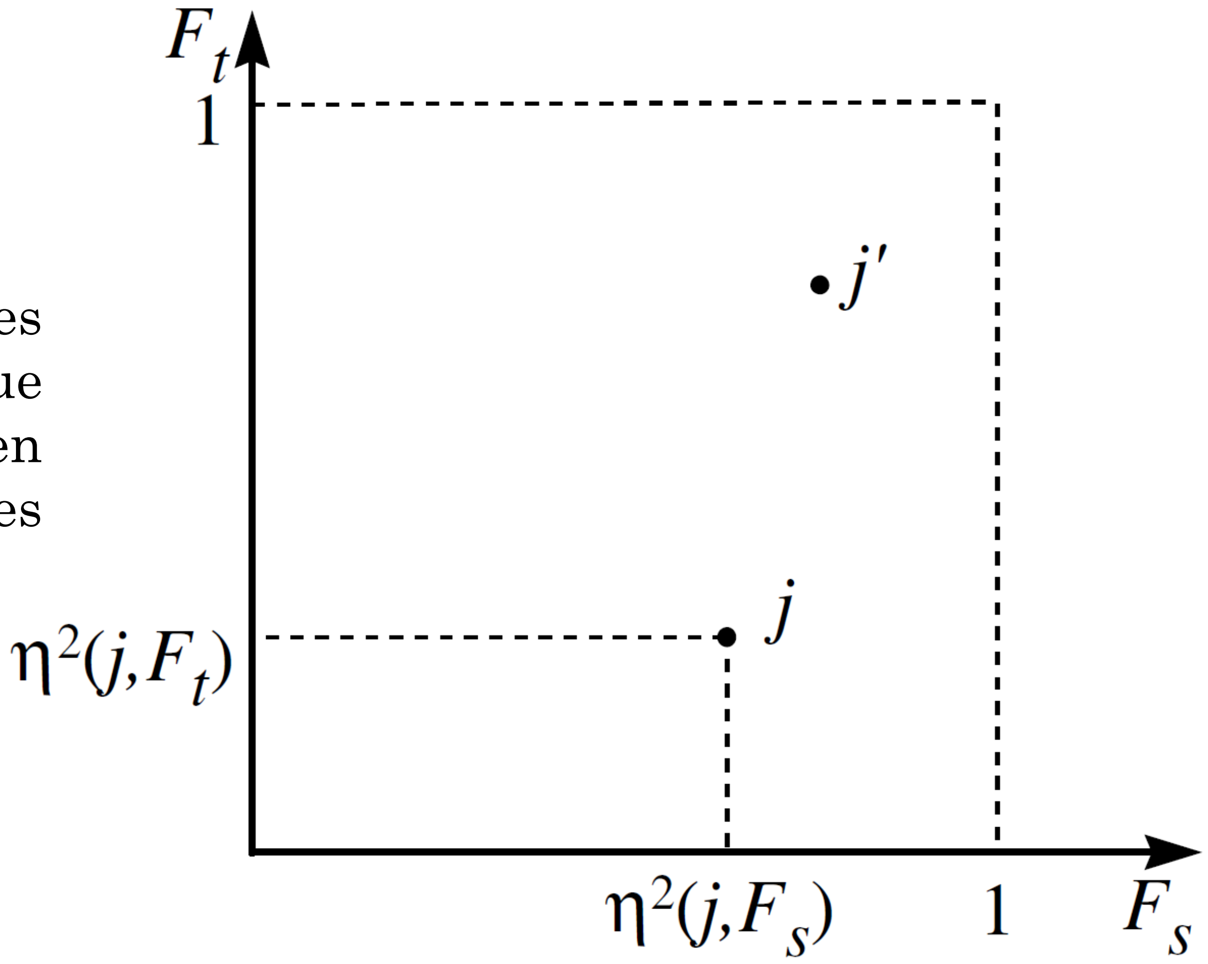


Représentation des variables en ACM

Graphique des carrés des liaisons

Il est intéressant de commencer l'analyse des résultats d'une ACM par la consultation systématique du graphique des carrés des liaisons qui met en évidence les variables les plus liées à chacun des facteurs.

La proximité entre deux points-variables traduisant la ressemblance entre les partitions engendrées par les deux variables.



ACM

Interprétation : L'inertie

- Dans ACM, les pourcentages d'inertie associés aux premiers composants sont généralement beaucoup plus faibles que dans l'ACP.
- Dans l'ACP, seules les relations linéaires sont étudiées : une seule composante devrait être suffisante pour représenter toutes les variables si elles sont fortement corrélées.
- Dans l'ACM, nous étudions des relations beaucoup plus générales et au moins $\min(K_j, K_L) - 1$ dimensions sont nécessaires pour représenter la relation entre deux variables ayant, respectivement, K_j et K_L catégories.
- Il est donc courant que beaucoup plus de dimensions soient étudiées dans l'ACM que dans l'ACP.

ACM

Interprétation : L'inertie

Barycentre des modalités d'une variable : $\sum_{i=1}^{k_j} \frac{I_k}{I} \frac{x_{ik}}{I_k} = \frac{1}{I}$

- L'ensemble des modalités d'une même variable est donc centré sur l'origine; les facteurs opposent entre elles à la fois l'ensemble de toutes les modalités et l'ensemble des modalités de chaque variable.
- Pour représenter parfaitement les r modalités d'une même variable, au moins $(r-1)$ facteurs sont nécessaires.
- **L'inertie d'une variable à r modalités (égale à $(r-1)/J$) est donc répartie dans un sous-espace à $r-1$ dimensions.**
  - **le pourcentage d'inertie associé à chaque facteur, est nécessairement faible lorsque les variables présentent beaucoup de modalités ;**
  - **même si un facteur est très lié à une variable, il est impossible que toutes ses modalités soient bien représentées par ce seul facteur ;**
  - **il n'est pas utile de multiplier de façon importante les modalités d'une même variable : le gain de finesse obtenu risque de ne pas être valorisé dans l'analyse.**

ACM

Interprétation : Contribution et qualité de représentation

- Les calculs et les interprétations pour la qualité de la contribution et de la représentation des individus et des catégories sont **les mêmes que dans l'AFC**.
- Cependant, en raison de la dimensionalité des données, **la qualité de représentation sur un plan donné est souvent beaucoup plus faible** par rapport aux qualités de représentation obtenues en AFC (ou ACP).
- **La dimensionalité de l'ensemble de données n'affecte pas la contribution**, car cet aspect est calculé pour chaque composante. Il est à noter que la contribution d'une variable catégorielle à une composante (facteur) donnée peut être calculée en additionnant les contributions de ces catégories.

ACM

Interprétation : Les éléments supplémentaires

- Les éléments supplémentaires présentent un grand intérêt en analyse des données pour
  - **Ne pas perdre d'information** sur un cas dont on ne désire pas qu'il participe activement aux analyses (cas nouveau, observation aberrante ou douteuse, élément de nature différente de ceux analysés, etc.),
  - **Interpréter plus facilement l'analyse** (à l'aide de variables illustratives par exemple, comme l'appartenance à une C.S.P. dans une analyse portant sur des questions d'opinion),
  - **Construire des modèle** de prévision, etc.
- Ces éléments sont représentés sur les axes issus de l'analyse factorielle, sans qu'ils aient contribué à la formation de ces axes.

ACM

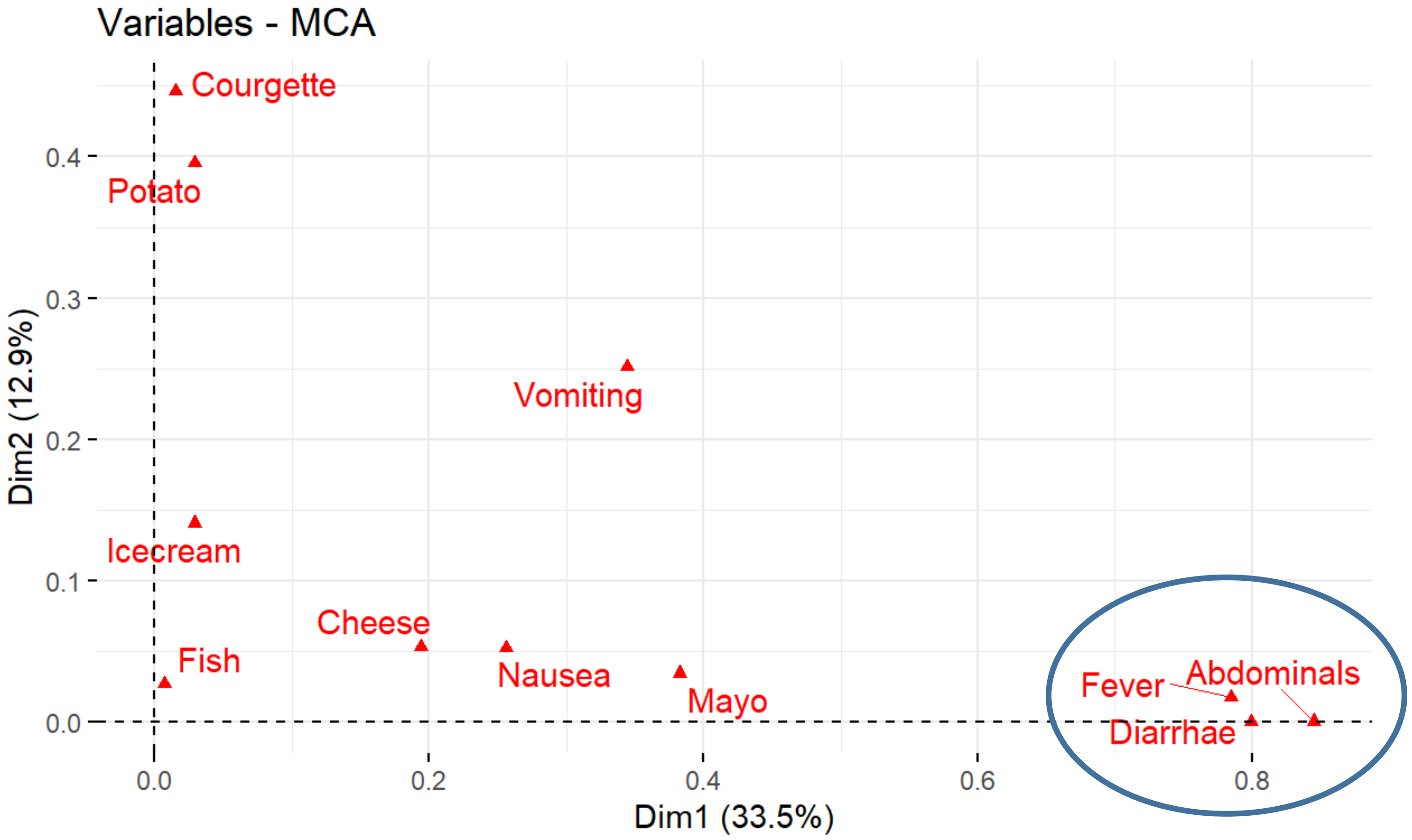
Interprétation : Description des facteurs (axes)

Description par les variables:

```
dimdesc(res.mca, axes =1)
```

```
$`Dim 1`
$quali
                   R2      p.value
Abdominals 0.8451157 4.055640e-23
Diarrhae   0.7994680 3.910776e-20
Fever      0.7846788 2.600566e-19
Mayo       0.3829749 4.756234e-07
Vomiting   0.3442016 2.510738e-06
Nausea     0.2562007 8.062777e-05
Cheese     0.1944181 7.534834e-04
```

$$R^2 = \eta^2(var_j, F) = \frac{Inertie_{\,Inter}}{Inertie_{\,Totale}}$$



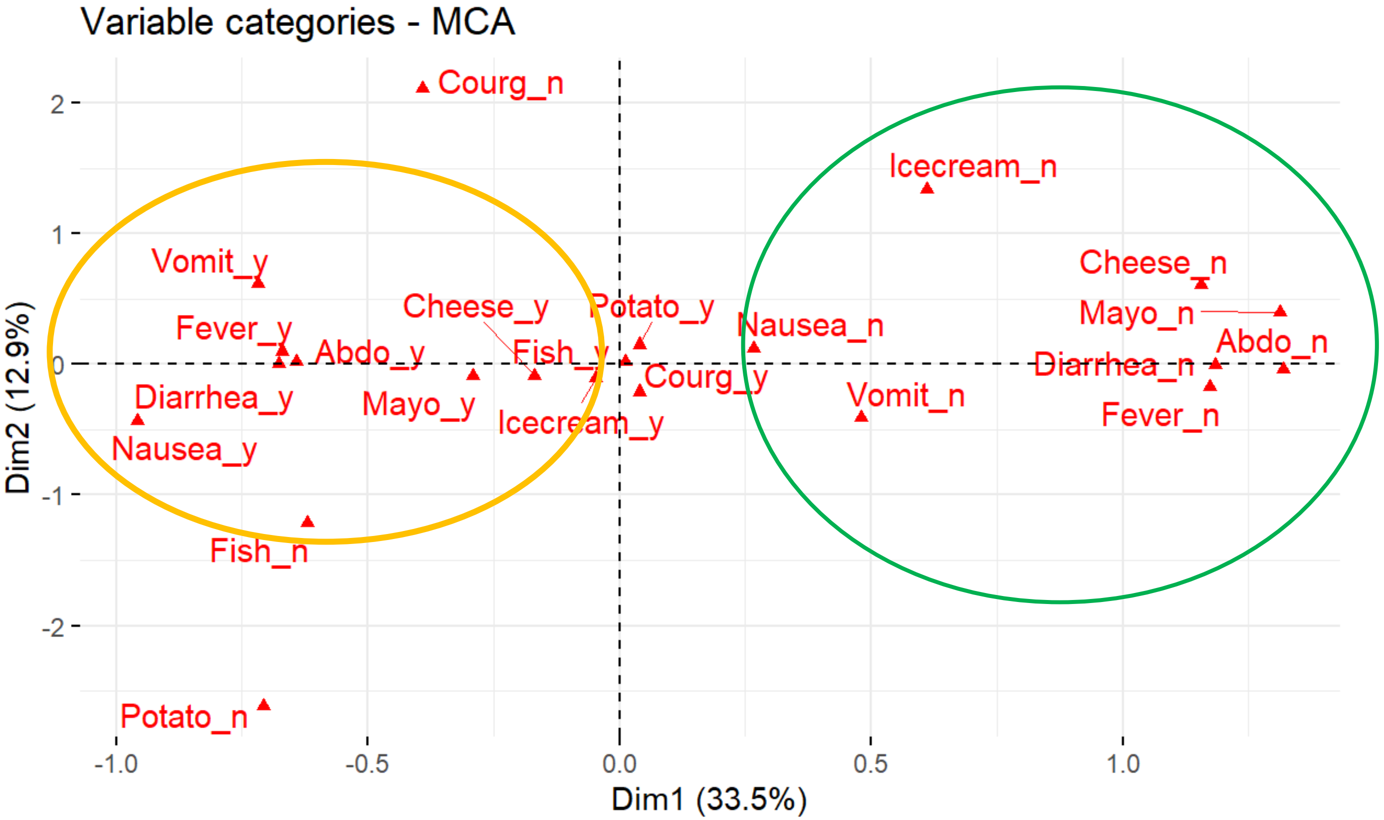
ACM

Interprétation : Description des facteurs (axes)

Description par les catégories:

```
dimdesc(res.mca, axes =1)
$category
                        Estimate      p.value
Abdominals=Abdo_n      0.5671866 4.055640e-23
Diarrhae=Diarrhea_n    0.5380920 3.910776e-20
Fever=Fever_n          0.5330918 2.600566e-19
Mayo=Mayo_n            0.4644981 4.756234e-07
Vomiting=Vomit_n       0.3466915 2.510738e-06
Nausea=Nausea_n        0.3547892 8.062777e-05
Cheese=Cheese_n        0.3830043 7.534834e-04
Cheese=Cheese_y       -0.3830043 7.534834e-04
Nausea=Nausea_y       -0.3547892 8.062777e-05
Vomiting=Vomit_y      -0.3466915 2.510738e-06
Mayo=Mayo_y           -0.4644981 4.756234e-07
Fever=Fever_y         -0.5330918 2.600566e-19
Diarrhae=Diarrhea_y   -0.5380920 3.910776e-20
Abdominals=Abdo_y     -0.5671866 4.055640e-23
```

Comme la plupart des variables ont deux catégories, la caractérisation par catégorie est similaire à celle calculée à partir des variables, mais elle spécifie la tendance de la position sur l’axe.



ACM

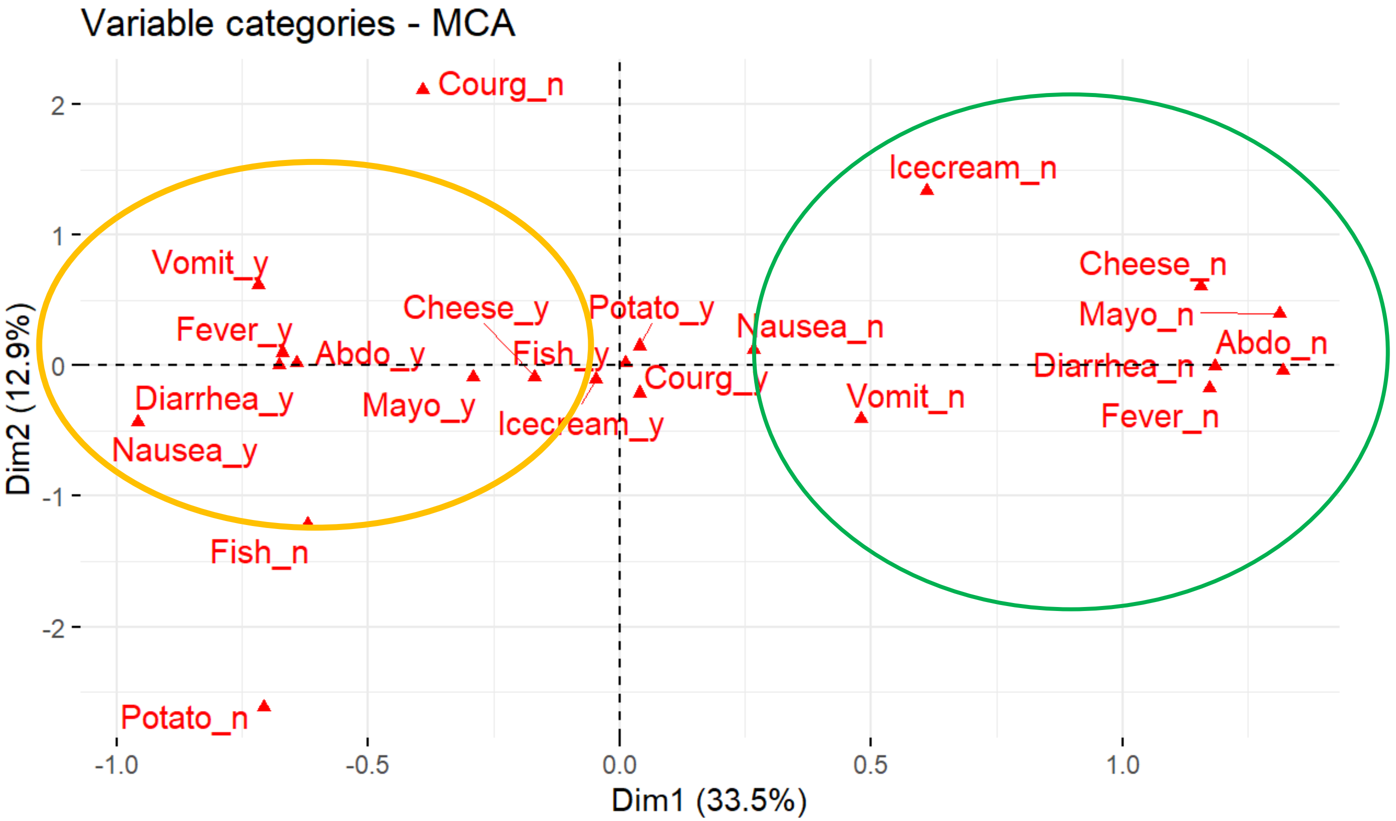
Interprétation : Description des facteurs (axes)

Description par les catégories:

```
dimdesc(res.mca, axes =1)
```

```
$category
                          Estimate      p.value
Abdominals=Abdo_n        0.5671866 4.055640e-23
Diarrhae=Diarrhea_n      0.5380920 3.910776e-20
Fever=Fever_n            0.5330918 2.600566e-19
Mayo=Mayo_n              0.4644981 4.756234e-07
Vomiting=Vomit_n         0.3466915 2.510738e-06
Nausea=Nausea_n          0.3547892 8.062777e-05
Cheese=Cheese_n          0.3830043 7.534834e-04
Cheese=Cheese_y         -0.3830043 7.534834e-04
Nausea=Nausea_y         -0.3547892 8.062777e-05
Vomiting=Vomit_y        -0.3466915 2.510738e-06
Mayo=Mayo_y             -0.4644981 4.756234e-07
Fever=Fever_y           -0.5330918 2.600566e-19
Diarrhae=Diarrhea_y     -0.5380920 3.910776e-20
Abdominals=Abdo_y       -0.5671866 4.055640e-23
```

Les individus ayant des coordonnées plus positives seraient donc plus susceptibles d'avoir les catégories dont les coordonnées sont positives.



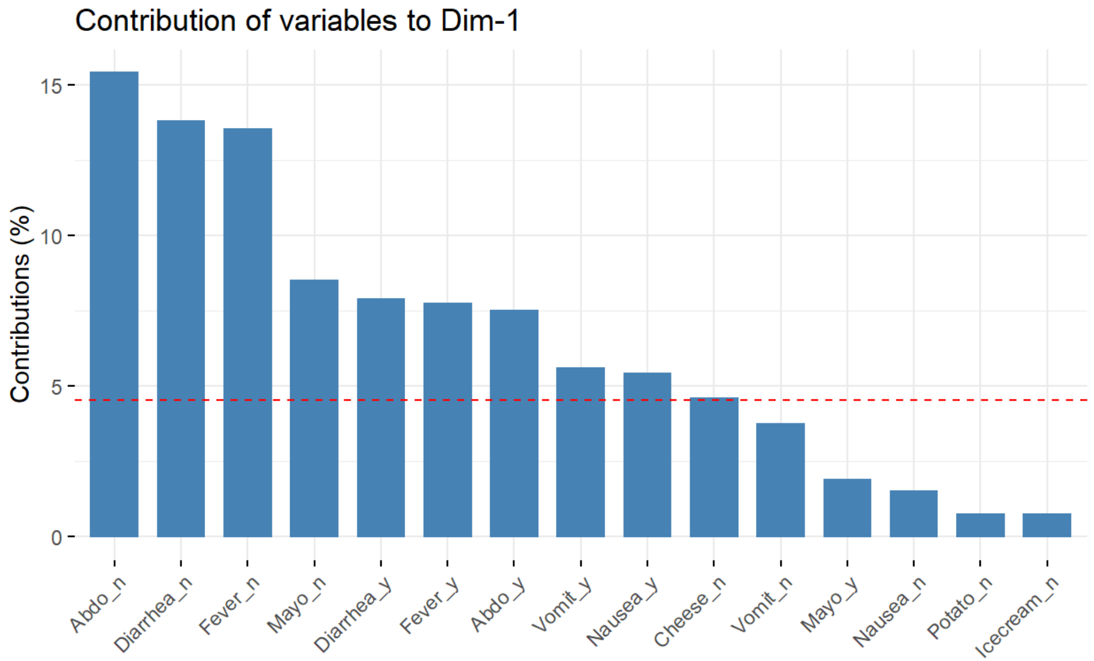
ACM

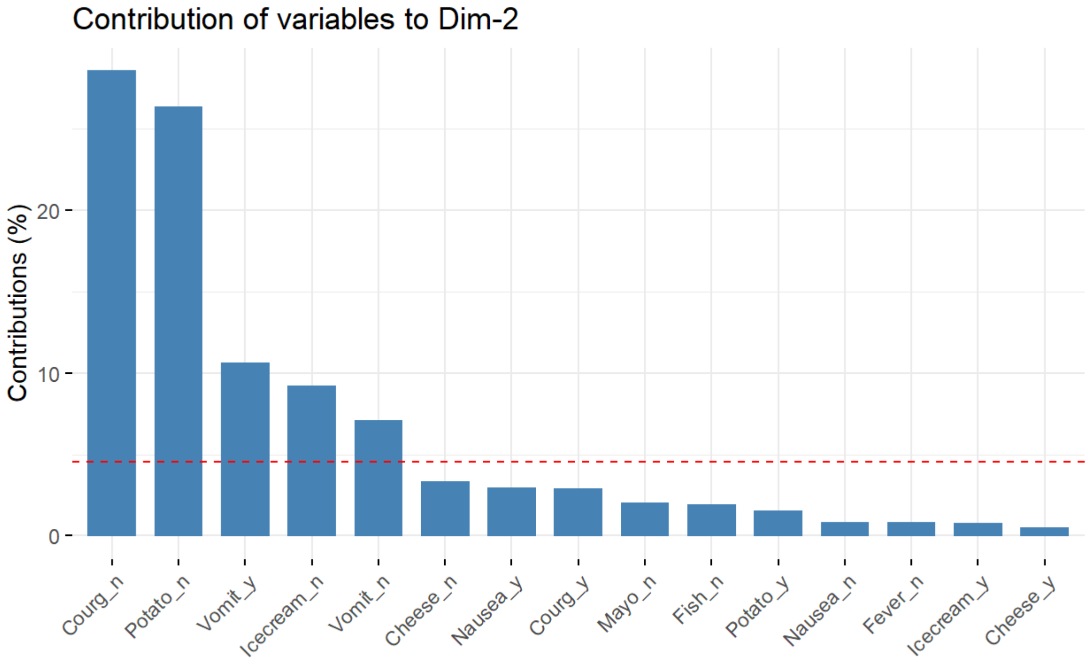
Interprétation : Description des facteurs (axes)

Contributions des variables

Les variables qui contribuent le plus à la définition des dimensions 1 et 2, sont les plus importantes pour expliquer la variabilité dans le jeu de données.

```
fviz_contrib (res.mca, choice = "var", axes = c(1,2), top = 15)
```





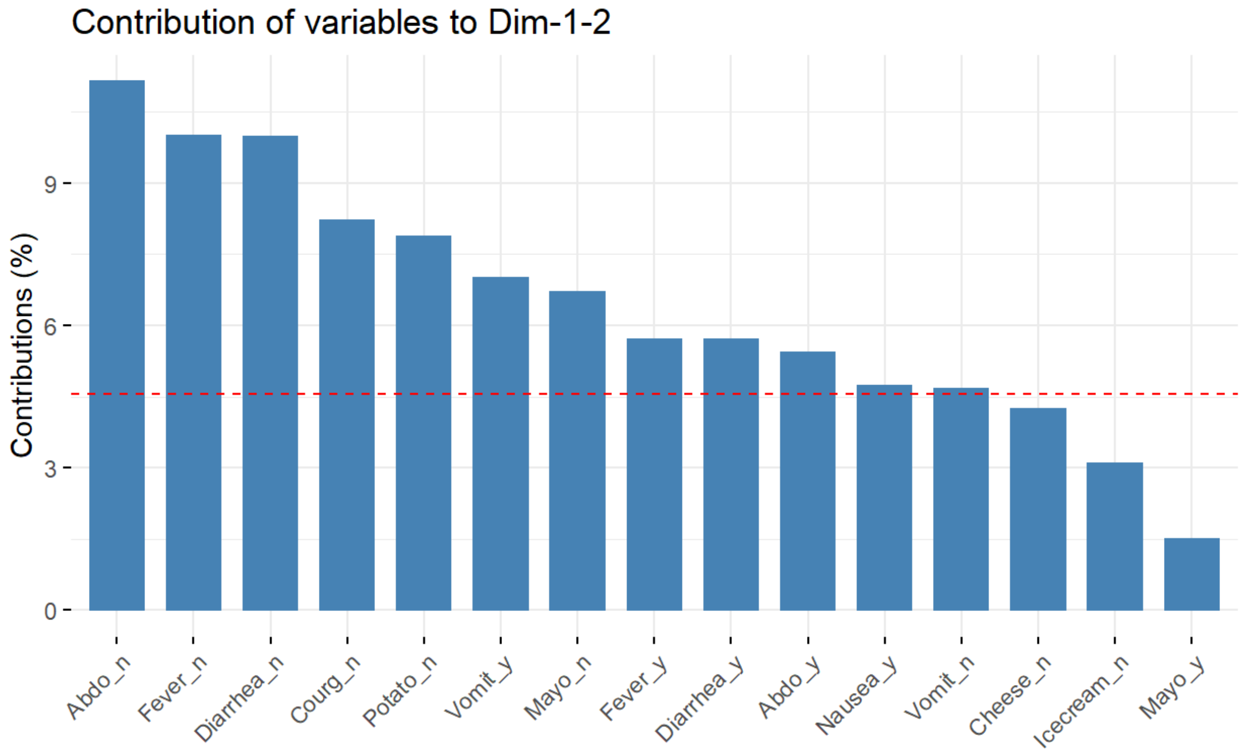
ACM

Interprétation : Description des facteurs (axes)

Contributions des variables

Les contributions totales aux dimensions 1 et 2 sont obtenues comme suit:

```
fviz_contrib (res.mca, choice = "var", axes = 1:2, top = 15)
```

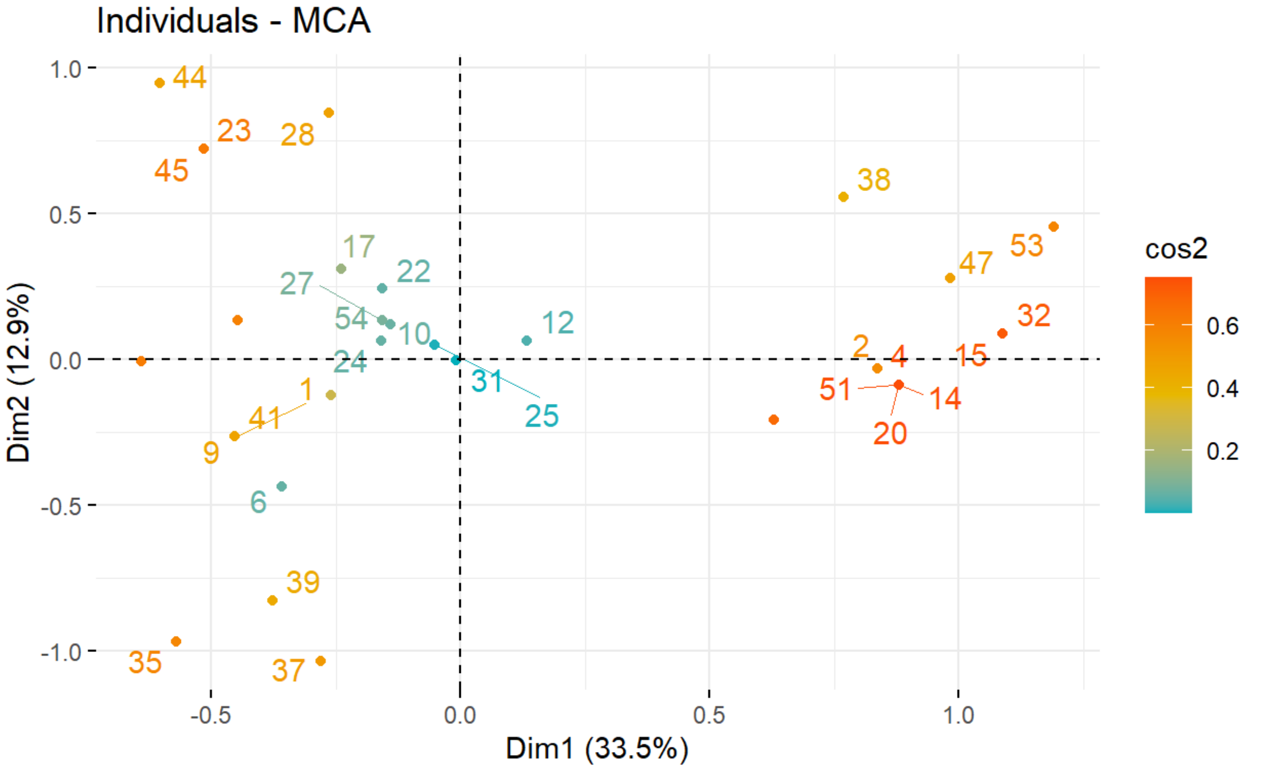


ACM

Interprétation : Description des facteurs (axes)

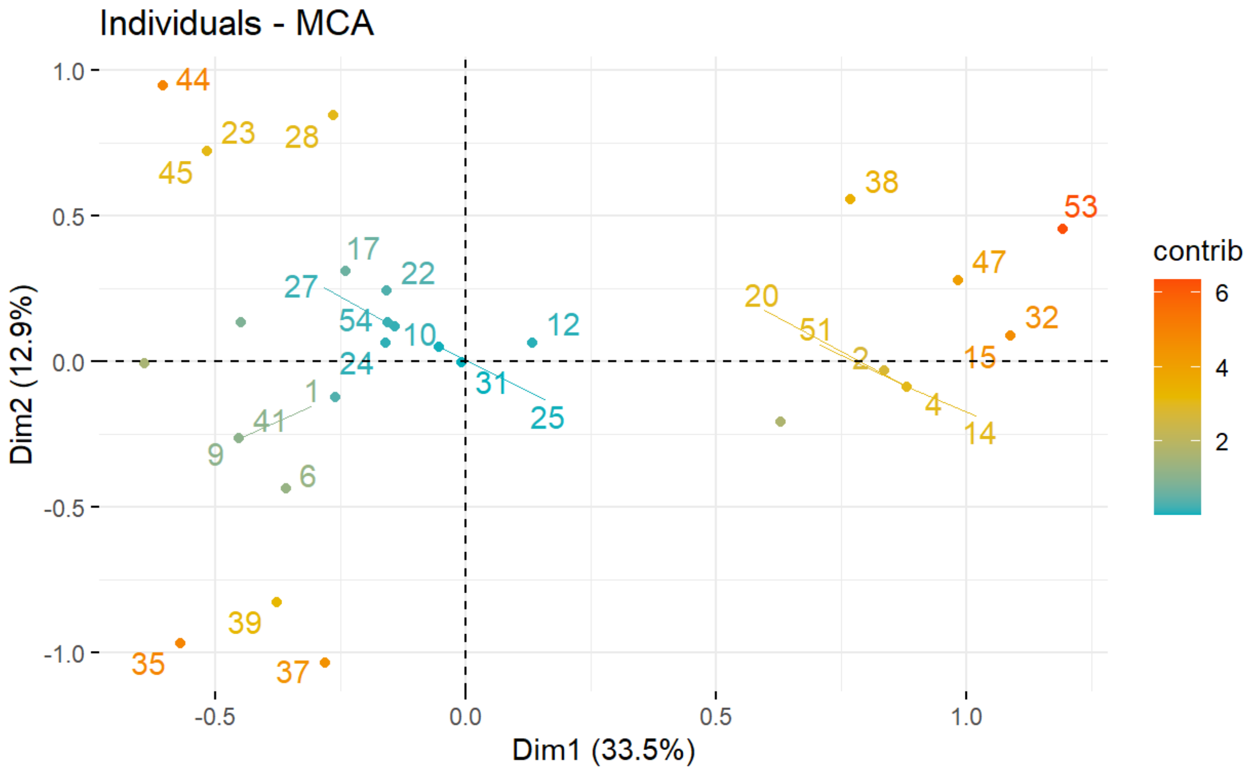
Qualité des individus

```
fviz_mca_ind(res.mca, col.ind = "cos2", gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"), repel = TRUE)
```



Contributions des individus

```
fviz_mca_ind(res.mca, col.ind = "contrib", gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"), repel = TRUE)
```



ACM

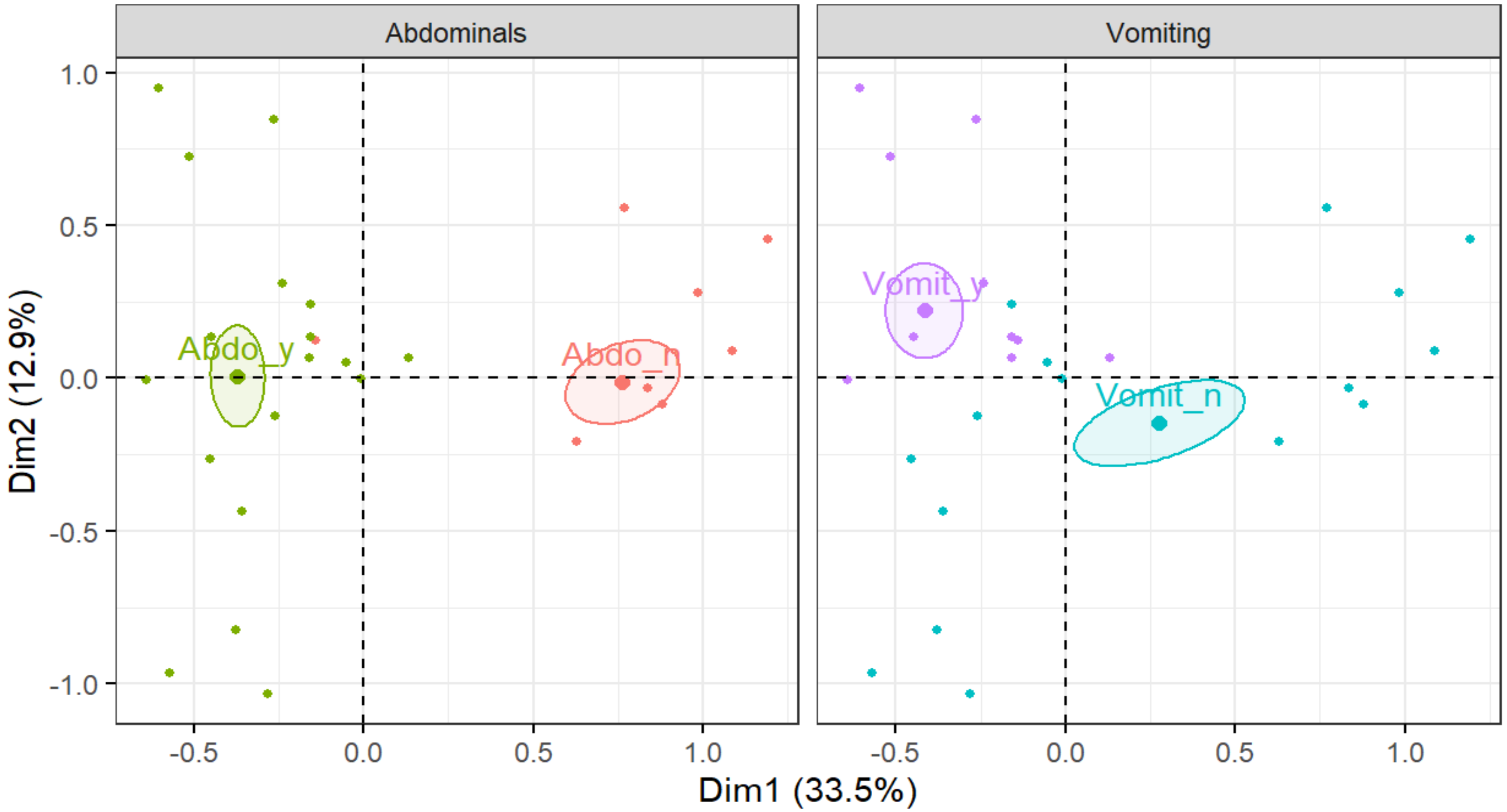
Interprétation : Description des facteurs (axes)

<u>Colorer les individus par groupes</u>

```
fviz_ellipses(res.mca, c("Vomiting", "Abdominals"), geom = "point")
```

Les ellipses de confiance peuvent être dessinées **du barycentre des individus portant une catégorie**.

Les ellipses sont utilisées pour **visualiser si deux catégories sont significativement différentes** ou non.



ACM

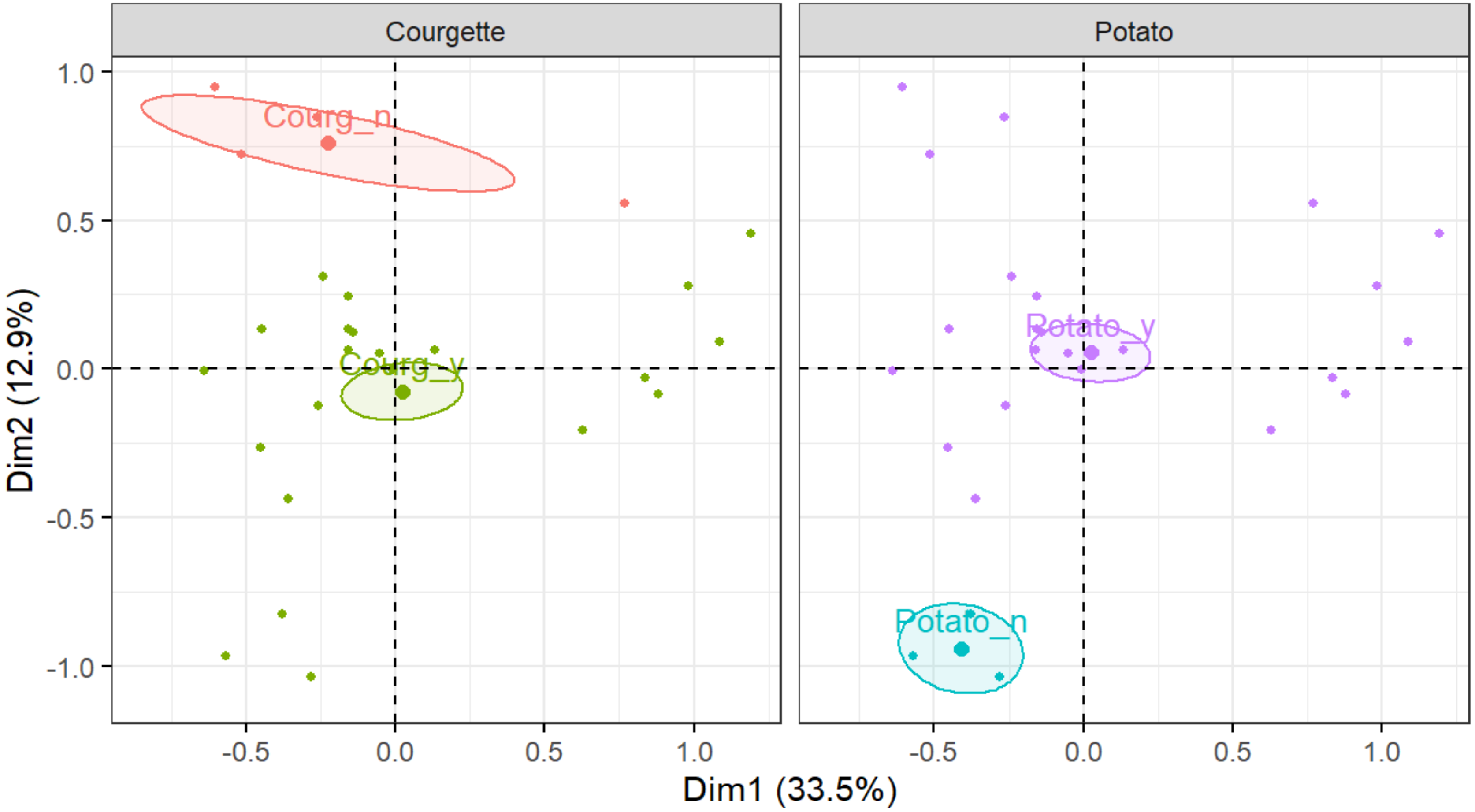
Interprétation : Description des facteurs (axes)

Colorer les individus par groupes

```
fviz_ellipses(res.mca, c("Courgette", "Potato"), geom = "point")
```

Les ellipses de confiance peuvent être dessinées du barycentre des individus portant une catégorie.

Les ellipses sont utilisées pour visualiser si deux catégories sont significativement différentes ou non.

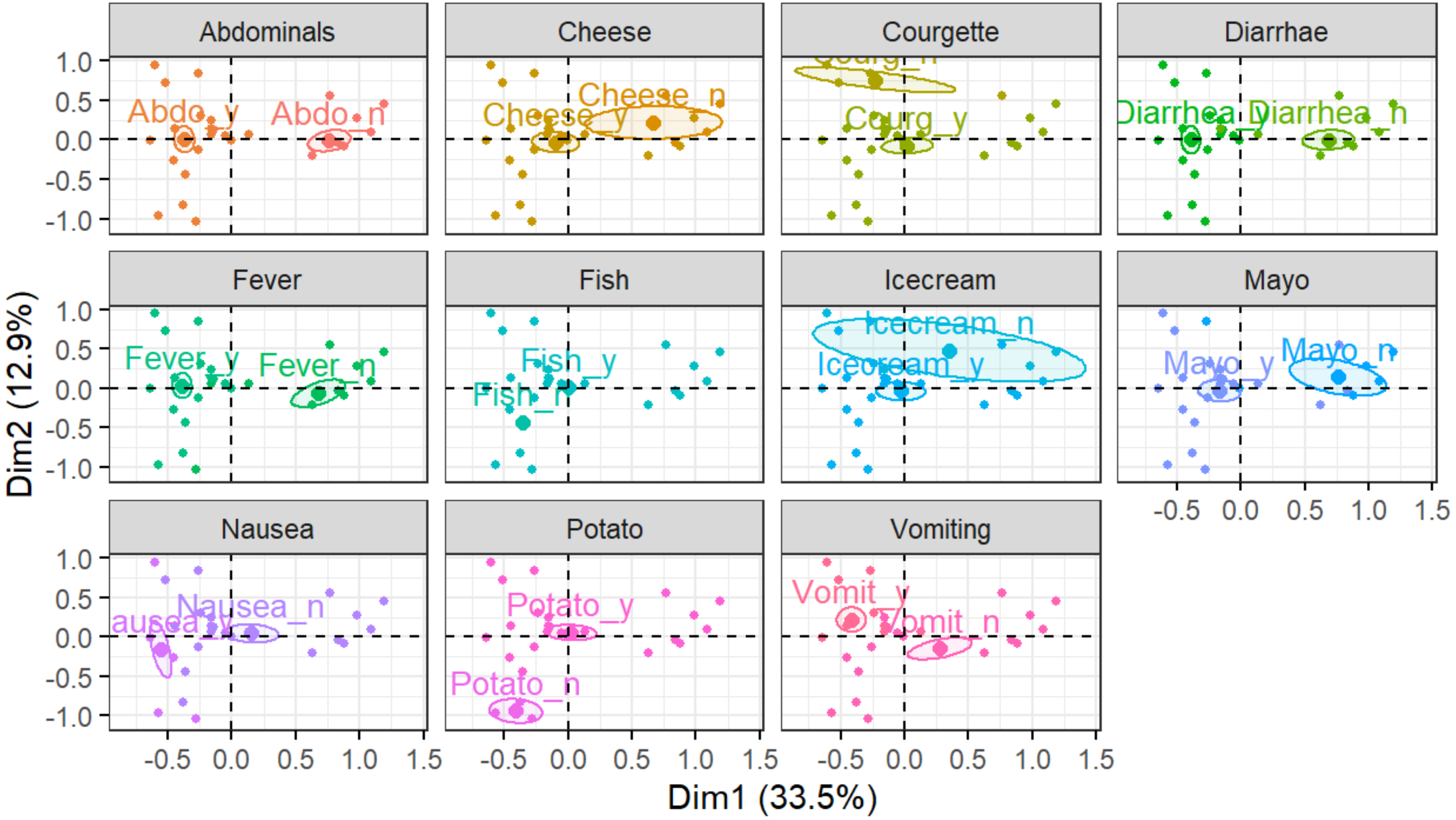


ACM

Interprétation : Description des facteurs (axes)

Colorer les individus par groupes

```
fviz_ellipses(res.mca, colnames(poison.active), geom = "point")
```



ACM

Interprétation : Description des facteurs (axes)

<u>Eléments supplémentaires : Rappel</u>

- Les éléments supplémentaires présentent un grand intérêt en analyse des données pour
  - **Epargner ces élément pour l'interprétation**
  - **Construire des modèle** de prévision, etc.
- Ces éléments sont représentés sur les axes issus de l'analyse factorielle, sans qu'ils aient contribué à la formation de ces axes.
- Ci-après les paramètres spécifiant les éléments supplémentaires dans la fonction MCA

```
MCA(X, ind.sup = …, quanti.sup = …, quali.sup = …, graph = TRUE)
```

ind.sup	a vector indicating the indexes of the supplementary individuals
quanti.sup	a vector indicating the indexes of the quantitative supplementary variables
quali.sup	a vector indicating the indexes of the categorical supplementary variables