

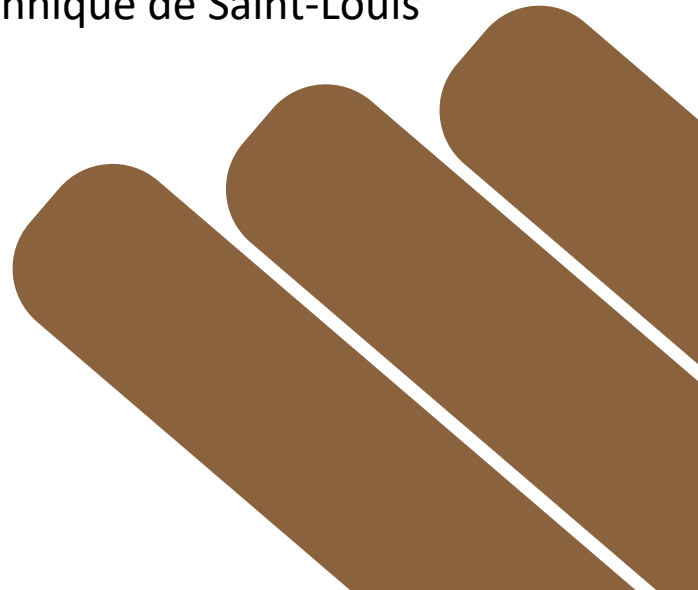


# **Rapport d'étude sur l'efficacité des techniques utilisées en agriculture : Analyse en composantes principales**

Présenté par :

**Michel Nassalang**

Elève-Ingénieur à l'Institut polytechnique de Saint-Louis



## Données des cultures à Podor

Notre sujet porte sur des données provenant d'une enquête sur l'agriculture dans la région de Podor.

Cette enquête avait pour but d'évaluer l'efficacité des techniques et méthodes utilisées dans l'agriculture ou acquises au cours d'une formation. Avec le soutien de l'ONG Enda Pronal, les données ont été récoltées sur un territoire de 43 villages dans la commune de Guédé Village. Elles ont été collectées par un étudiant de l'université Gaston Berger du nom de Félix Charlemagne Sagna. L'étude se base sur les cultures et les terres de chaque famille logeant dans ces villages.

Ce jeu de données contient **93 individus et 9 variables**, 1 variable qualitative est illustrative.

**NB** : les numéros sur le document caractérisent les identifiants des familles ; Chaque famille détenant des terres et pratiquant l'agriculture est caractérisée par un numéro bien défini.

---

### 1. Observation d'individus extrêmes

L'analyse des graphes ne révèle aucun individu singulier.

---

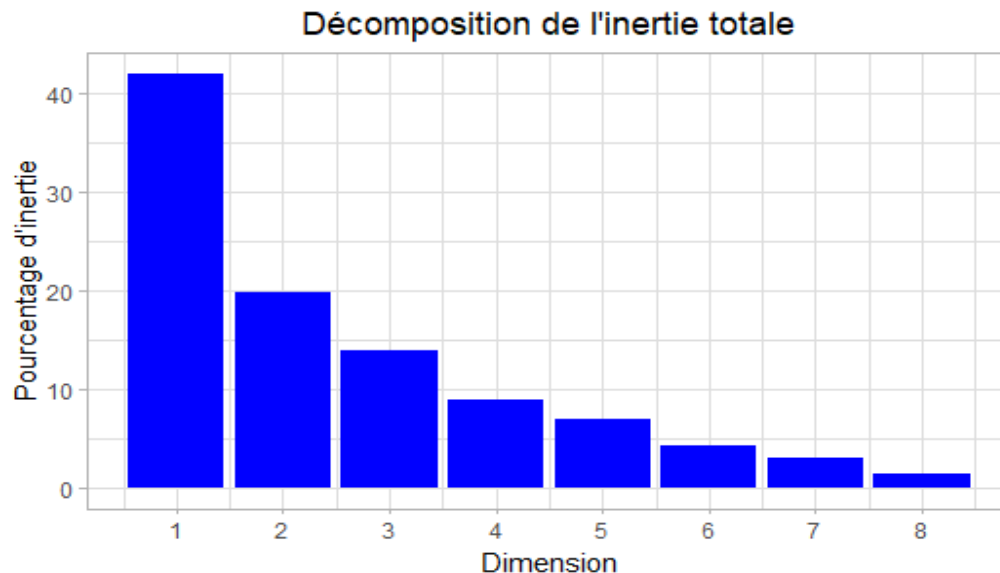
### 2. Distribution de l'inertie

L'inertie des axes factoriels indique d'une part si les variables sont structurées et suggère d'autre part le nombre judicieux de composantes principales à étudier.

Les 2 premiers axes de l'analyse expriment **61.66%** de l'inertie totale du jeu de données ; cela signifie que 61.66% de la variabilité totale du nuage des individus (ou des variables) est représentée dans ce plan. C'est un pourcentage assez important, et le premier

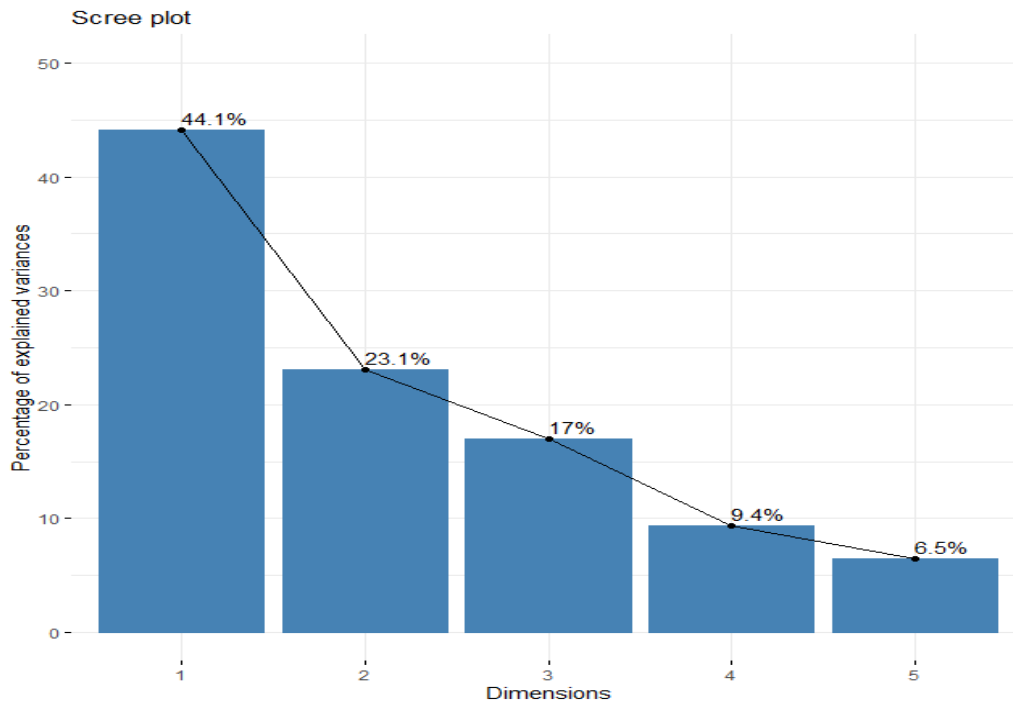
plan représente donc convenablement la variabilité contenue dans une grande part du jeu de données actif. Cette valeur est nettement supérieure à la valeur référence de **36.75%**, la variabilité expliquée par ce plan est donc hautement significative (cette inertie de référence est le quantile 0.95-quantile de la distribution des pourcentages d'inertie obtenue en simulant 2668 jeux de données aléatoires de dimensions comparables sur la base d'une distribution normale).

Du fait de ces observations, il serait tout de même probablement préférable de considérer également dans l'analyse les dimensions supérieures ou égales à la troisième.



**Figure 2 - Décomposition de l'inertie totale**

Une estimation du nombre pertinent d'axes à interpréter suggère de restreindre l'analyse à la description des 2 premiers axes. Ces composantes révèlent un taux d'inertie supérieur à celle du quantile 0.95-quantile de distributions aléatoires (61.66% contre 36.75%). Cette observation suggère que seuls ces axes sont porteurs d'une véritable information. En conséquence, la description de l'analyse sera restreinte à ces seuls axes.



### 3. Description du plan 1:2

## Warning: ggrepel: 79 unlabeled data points (too many overlaps). Consider  
## increasing max.overlaps

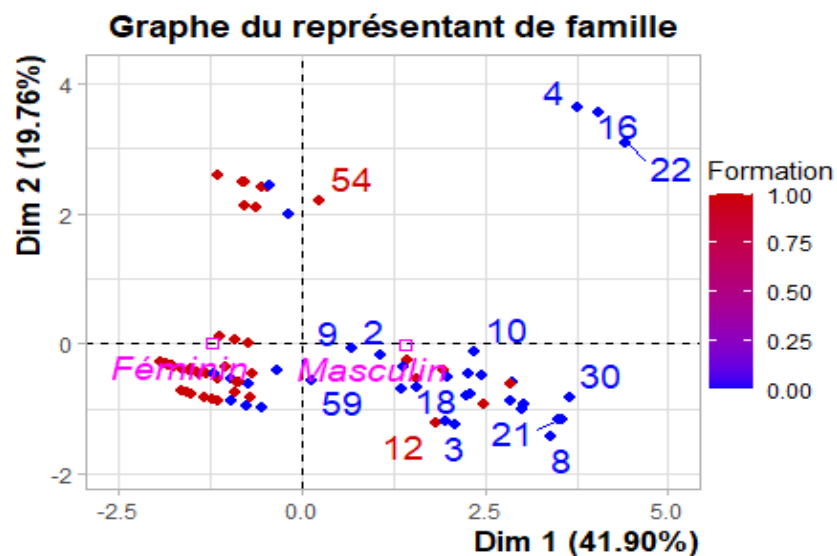


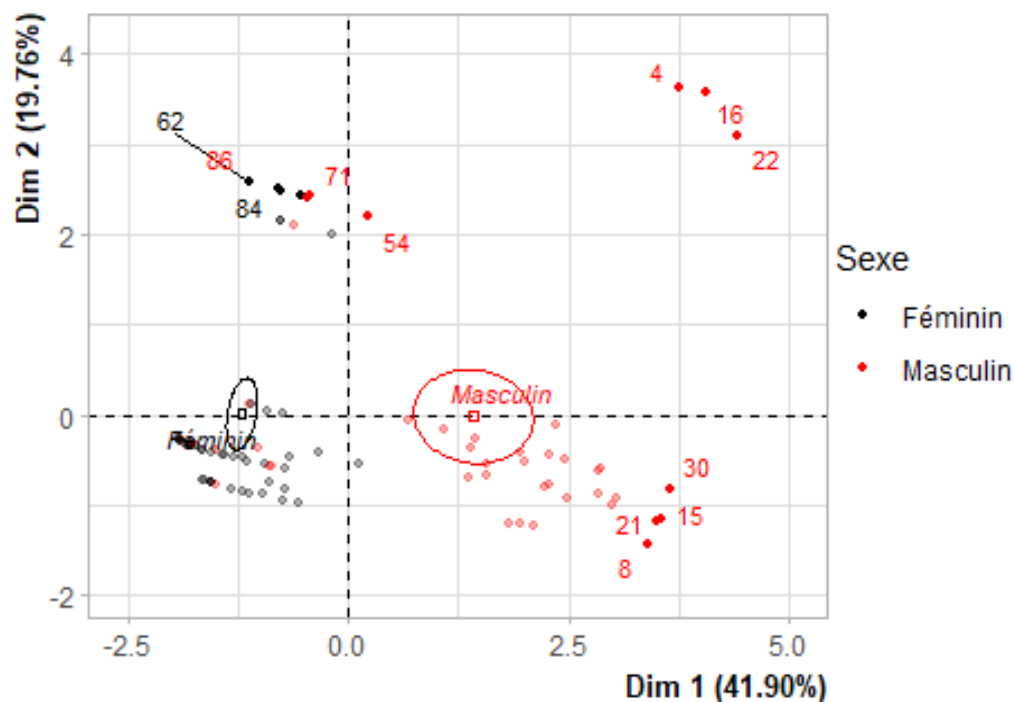
Figure 3.1 - Graphe des individus (ACP)

La probabilité critique du test de Wilks indique la variable dont les modalités séparent au mieux les individus sur le plan (i.e. qui explique au mieux les distances entre individus).

```
##      Sexe  
## 7.535629e-15
```

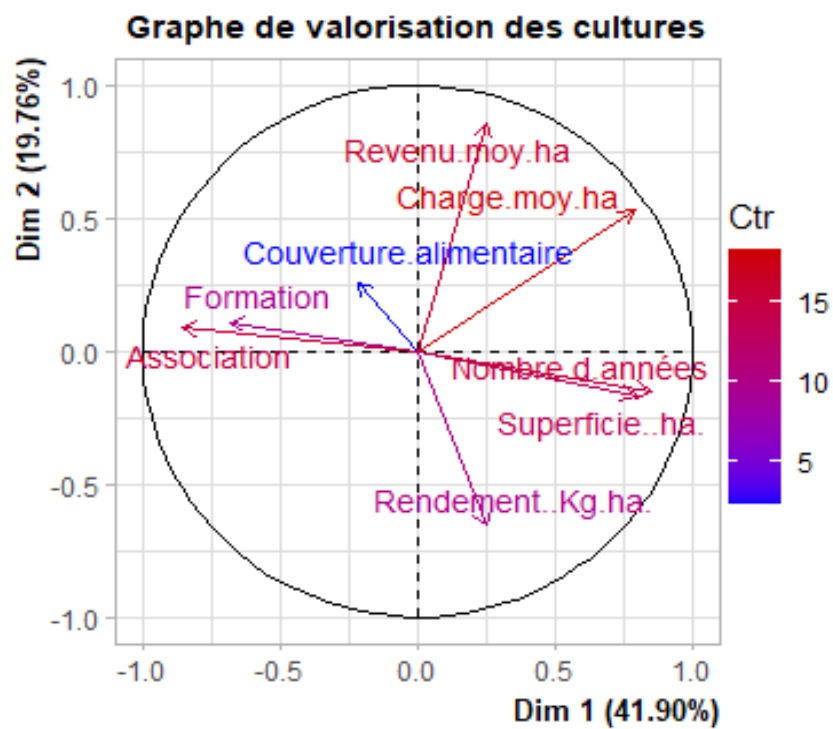
Il n'y a qu'une unique variable qualitative possible pour illustrer les distances entre individus : *Sexe*.

```
## Warning: ggrepel: 5 unlabeled data points (too many overlaps). Consider  
## increasing max.overlaps
```



**Figure 3.2 - Graphe des individus (ACP)**

*Les individus libellés sont ceux ayant la plus grande contribution à la construction du plan. Les individus sont colorés selon leur appartenance aux modalités de la variable Sexe.*



**Figure 3.3 - Graphe des variables (ACP)**



**Figure 3.3 - Graphe de contribution**

Dans le graphe ci-dessus, on observe la mise en évidence des variables les plus contributives pour chaque dimension.

Les variables libellées sont celles les mieux représentées sur le plan.

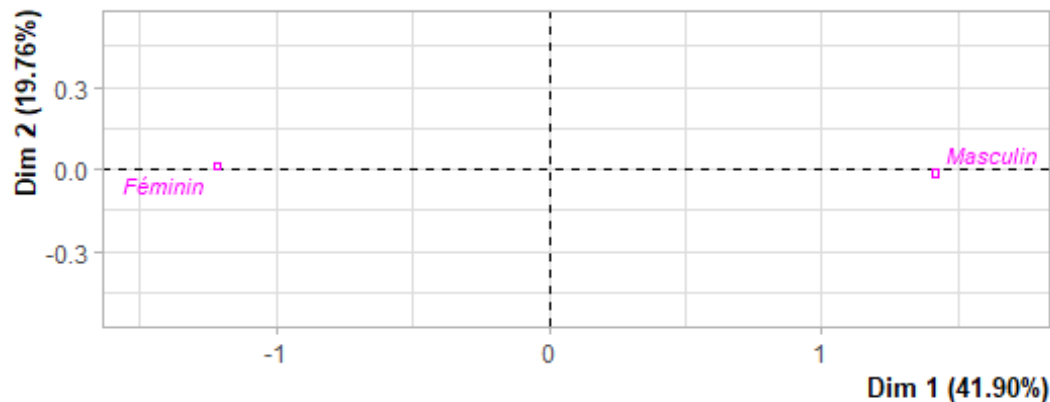


Figure 3.4 - Graphe des modalités (ACP)

Les facteurs libellés sont ceux les mieux représentés sur le plan.

---

La **dimension 1** oppose les familles des numéros 16, 4, 22, 15, 21, 30 et 8 (à droite du graphe, caractérisés par une coordonnée fortement positive sur l'axe) à des familles caractérisées par une coordonnée fortement négative sur l'axe (à gauche du graphe).

Le groupe des familles avec les numéros 15, 21, 30 et 8 (caractérisés par une coordonnée positive sur l'axe) partage :

- De fortes valeurs pour les variables *Nombre.d.années*, *Superficie..ha.*, *Rendement..Kg.ha.* et *Charge.moy.ha* (de la plus extrême à la moins extrême).
- De faibles valeurs pour les variables *Association*, *Formation* et *Couverture.alimentaire* (de la plus extrême à la moins extrême).

Le groupe des familles avec les numéros 16, 4 et 22 (caractérisés par une coordonnée positive sur l'axe) partage :

- De fortes valeurs pour les variables *Charge.moy.ha*, *Revenu.moy.ha* et *Superficie..ha*. (de la plus extrême à la moins extrême).
- De faibles valeurs pour les variables *Association*, *Formation* et *Rendement..Kg.ha*. (de la plus extrême à la moins extrême).

Le groupe 3 (caractérisé par une coordonnée négative sur l'axe) partage :

- De fortes valeurs pour les variables *Formation* et *Association* (de la plus extrême à la moins extrême).
- De faibles valeurs pour les variables *Charge.moy.ha*, *Revenu.moy.ha*, *Superficie..ha*. et *Nombre.d.années* (de la plus extrême à la moins extrême).

Notons que les variables *Féminin* et *Masculin* sont extrêmement corrélées à cette dimension (corrélations respectives de 0.98, 0.98). Ces variables pourraient donc résumer à elles seules la dimension 1.

---

La **dimension 2** oppose les familles avec les numéros tels que 62, 84, 86, 38, 46, 16, 4, 32, 22 et 56 (en haut du graphe, caractérisés par une coordonnée fortement positive sur l'axe) aux familles avec les numéros comme 15, 21, 30, 8 (en bas du graphe, caractérisés par une coordonnée fortement négative sur l'axe).

Le groupe des familles avec les numéros 62, 84, 86, 38, 46, 32, 56, 48, 71 et 54 (caractérisés par une coordonnée positive sur l'axe) partage :

- De fortes valeurs pour les variables *Revenu.moy.ha*, *Couverture.alimentaire* et *Association* (de la plus extrême à la moins extrême).
- De faibles valeurs pour la variable *Rendement..Kg.ha..*



Le groupe des familles avec les numéros 16, 4 et 22 (caractérisés par une coordonnée positive sur l'axe) partage :

- De fortes valeurs pour les variables *Charge.moy.ha*, *Revenu.moy.ha* et *Superficie..ha*. (de la plus extrême à la moins extrême).
- De faibles valeurs pour les variables *Association*, *Formation* et *Rendement..Kg.ha*. (de la plus extrême à la moins extrême).

Le groupe 3 (caractérisé par une coordonnée négative sur l'axe) partage :

- De fortes valeurs pour les variables *Formation* et *Association* (de la plus extrême à la moins extrême).
- De faibles valeurs pour les variables *Charge.moy.ha*, *Revenu.moy.ha*, *Superficie..ha*. et *Nombre.d.années* (de la plus extrême à la moins extrême).

Le groupe des familles avec les numéros 15, 21, 30 et 8 (caractérisé par une coordonnée négative sur l'axe) partage :

- De fortes valeurs pour les variables *Nombre.d.années*, *Superficie..ha.*, *Rendement..Kg.ha*. et *Charge.moy.ha* (de la plus extrême à la moins extrême).
  - De faibles valeurs pour les variables *Association*, *Formation* et *Couverture.alimentaire* (de la plus extrême à la moins extrême).
-

## 4. Classification

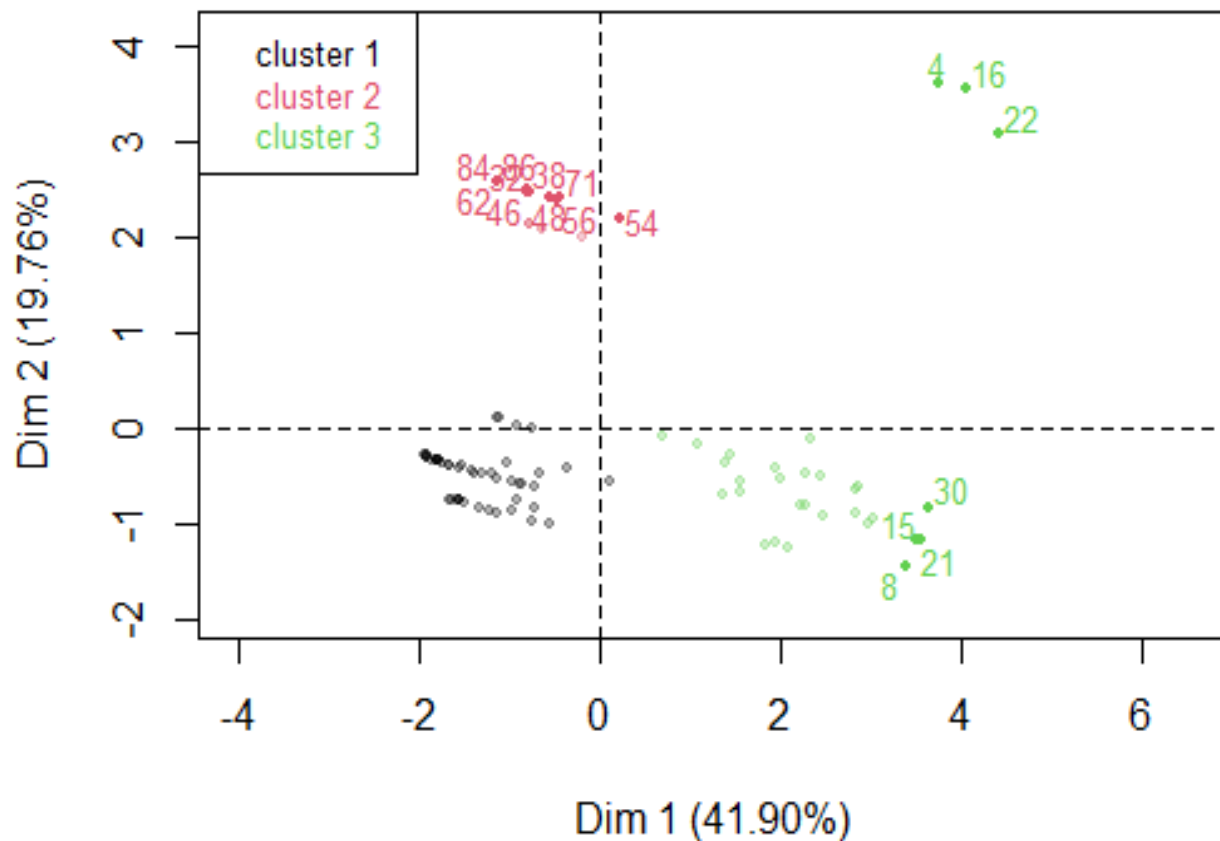


Figure 4 - Classification Ascendante Hiérarchique des individus.

*La classification réalisée sur les familles fait apparaître 3 classes.*

La **classe 1** est composé d'individus partageant :

- de fortes valeurs pour les variables *Association* et *Formation* (de la plus extrême à la moins extrême).
- de faibles valeurs pour les variables *Charge.moy.ha*, *Nombre.d.années*, *Superficie..ha.* et *Revenu.moy.ha* (de la plus extrême à la moins extrême).

La **classe 2** est composée des familles avec les numéros tels que 32, 38, 46, 48, 54, 56, 62, 71, 84 et 86. Ce groupe est caractérisé par :

- de fortes valeurs pour les variables *Revenu.moy.ha* et *Association* (de la plus extrême à la moins extrême).

- de faibles valeurs pour les variables *Rendement..Kg.ha.*, *Nombre.d.années* et *Superficie..ha.* (de la plus extrême à la moins extrême).

La **classe 3** est composée des familles avec les numéros tels que 4, 8, 15, 16, 21, 22 et 30. Ce groupe est caractérisé par :

- de fortes valeurs pour les variables *Nombre.d.années*, *Superficie..ha.*, *Charge.moy.ha* et *Rendement..Kg.ha.* (de la plus extrême à la moins extrême).
- de faibles valeurs pour les variables *Association* et *Formation* (de la plus extrême à la moins extrême).

## Conclusion

On retiendra de cette analyse que les données nous ont permis de recueillir une somme considérable d'informations pour en dégager l'essentiel, à savoir les formations et les associations offertes aux villageois de Guédé-village n'ont pas un apport considérable sur les cultures des familles. Par contre les fonds déployés pour les cultures continuent d'être un poids important sur les bénéfices obtenus après les cultures.

## Annexes

### Liste des variables caractéristiques des dimensions de l'analyse

\$Dim.1

\$Dim.1\$quanti

	<b>correlation</b>	<b>p.value</b>
Nombre.d.années	8.494E-01	5.434E-27
Superficie..ha.	7.980E-01	1.012E-21
Charge.moy.ha	7.909E-01	4.110E-21
Rendement..Kg.ha.	2.541E-01	1.398E-02
Revenu.moy.ha	2.519E-01	1.487E-02
Couverture.alimentaire	-2.153E-01	3.818E-02
Formation	-6.813E-01	5.699E-14
Association	-8.543E-01	1.355E-27

\$Dim.2

\$Dim.2\$quanti

	<b>correlation</b>	<b>p.value</b>
Revenu.moy.ha	8.563E-01	7.528E-28
Charge.moy.ha	5.291E-01	4.971E-08
Couverture.alimentaire	2.596E-01	1.197E-02
Rendement..Kg.ha.	-6.552E-01	1.042E-12

\$Dim.3  
\$Dim.3\$quanti

	<b>correlation</b>	<b>p.value</b>
Couverture.alimentaire	7.904E-01	4.508E-21
Formation	3.510E-01	5.617E-04
Superficie..ha.	2.876E-01	5.183E-03
Nombre.d.années	2.178E-01	3.598E-02
Rendement..Kg.ha.	-2.631E-01	1.082E-02
Association	-2.659E-01	9.997E-03
Revenu.moy.ha	-2.918E-01	4.540E-03

## Liste des variables caractéristiques des clusters de la classification

res.hcpc\$desc.var

Eta2 P-value

Revenu.moy.ha 0.9517416 5.764855e-60

Rendement..Kg.ha. 0.8931296 1.988754e-44

Charge.moy.ha 0.2978271 1.230219e-07

\$`1`

v.test Mean in category Overall mean sd in category Overall sd

Charge.moy.ha -5.008191 212954.5 260483.9 104388.5 142319.6

Revenu.moy.ha -8.820833 972556.8 1509698.4 177817.3 913194.3

p.value

Charge.moy.ha 5.494399e-07

Revenu.moy.ha 1.136116e-18

\$`2`

v.test Mean in category Overall mean sd in category

Rendement..Kg.ha. 7.651600 27048 10754.12 5272.826

Revenu.moy.ha 2.163855 2103220 1509698.39 291116.179

Overall sd p.value

Rendement..Kg.ha. 7089.689 1.984944e-14

Revenu.moy.ha 913194.264 3.047545e-02

\$`3`

v.test Mean in category Overall mean sd in category

Revenu.moy.ha 8.624995 3245941.176 1509698.39 218057.86350

Charge.moy.ha 4.728290 408823.529 260483.87 181079.51951

Rendement..Kg.ha. -6.050320 1298.412 10754.12 87.29939

Overall sd p.value

Revenu.moy.ha 913194.264 6.409538e-18

Charge.moy.ha 142319.617 2.264189e-06

Rendement..Kg.ha. 7089.689 1.445580e-09