



L'analyse des données multivariées à l'aide du logiciel

L'analyse factorielle des correspondances simples (A.F.C.s.)

INRA - Formation Permanente
Juillet 2005

André Bouchier

Copyright 2005 André Bouchier

Permission is granted to make and distribute verbatim copies of this manual provided the copyright notice and this permission notice are preserved on all copies.

Permission is granted to copy and distribute modified versions of this manual under the conditions for verbatim copying, provided that the entire resulting derived work is distributed under the terms of a permission notice identical to this one.

1. Avant propos :

- Pré-recquis

Ce support de cours concerne la mise en oeuvre d'une **analyse factorielle des correspondances simples** l'aide du logiciel de statistiques R. Ce n'est ni un cours de statistique ni une initiation à R. On suppose donc connues les bases de ce logiciel ainsi que quelques notions de statistiques concernant les analyses factorielles.

- Les fonctions utilisées :

Les fonctions utilisées pour cette analyse sont disponibles dans les bibliothèques standard de R et dans la bibliothèque **ade4**. Pour aider à la compréhension, l'écriture des «programmes» sera détaillée. Par la suite, vous pourrez condenser cette écriture. Mais n'oubliez pas de les commenter abondamment.

2. Rappels :

- L'analyse en composantes principales nous permet de projeter sur un plan un résumé pertinent d'un tableau de données quantitatives.
- Nous avons 2 variables qualitatives dont le croisement des modalités donne le tableau de fréquence. Nous travaillons ici sur des effectifs.
- L'analyse factorielle des correspondances simples nous permettra de projeter sur un plan les structures d'un tableau de contingence.

Attention : l'analyse factorielle des correspondances étudie la structure des écarts à l'indépendance, pas leur intensité.

3. Les données d'exemple, un tableau de contingence :

- Le tableau des données d'exemple :

Le tableau de données *housetasks* contient 13 « tâches ménagères » et leur répartition dans le couple.

Ce tableau contient 4 colonnes numériques :

wife: travaux effectués par la femme

alternating: travaux effectués alternativement par l'homme et la femme

husband: travaux effectués par l'homme

jointly: travaux effectués ensemble

- Chaque valeur numérique du tableau de données est un effectif

4. Lecture des données :

- Le tableau de données est fourni avec la librairie **ade4**

```
library(ade4)
data(housetasks)
housetasks
```

	<i>Wife</i>	<i>Alternating</i>	<i>Husband</i>	<i>Jointly</i>
<i>Laundry</i>	156	14	2	4
<i>Main_meal</i>	124	20	5	4
<i>Dinner</i>	77	11	7	13
<i>Breakfeast</i>	82	36	15	7
<i>Tidying</i>	53	11	1	57
<i>Dishes</i>	32	24	4	53
<i>Shopping</i>	33	23	9	55
<i>Official</i>	12	46	23	15
<i>Driving</i>	10	51	75	3
<i>Finances</i>	13	13	21	66
<i>Insurance</i>	8	1	53	77
<i>Repairs</i>	0	3	160	2
<i>Holidays</i>	0	1	6	153

5.L'A.F.C.s. - fonction `dudi.coa()` de la bibliothèque `ade4` :

- Les résultats de l'AFC sont stockés dans la variable `z`

```
z<-dudi.coa(df = housetasks, scannf = F, nf = 3)
```

- L'éboulis des valeurs propres

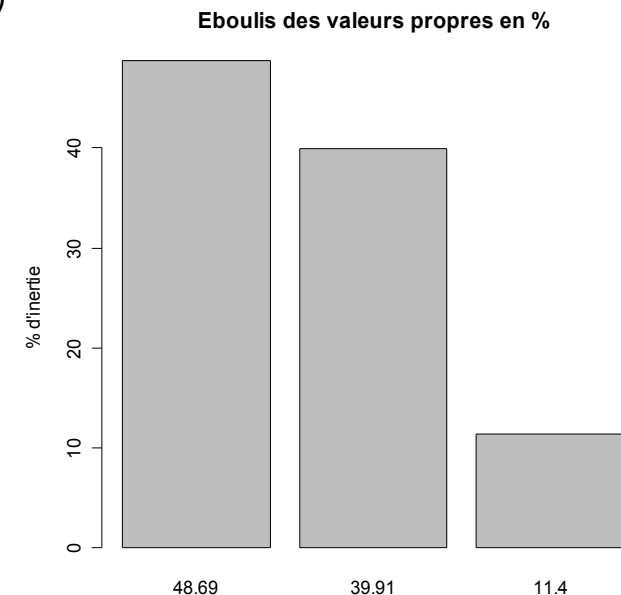
```
inertie<-z$eig/sum(z$eig)*100  
barplot(inertie,ylab="% d'inertie",names.arg=round(inertie,2))  
title("Eboulis des valeurs propres en %")
```

- Les valeurs propres

```
round(z$eig,2)  
  
[1] 0.54 0.45 0.13
```

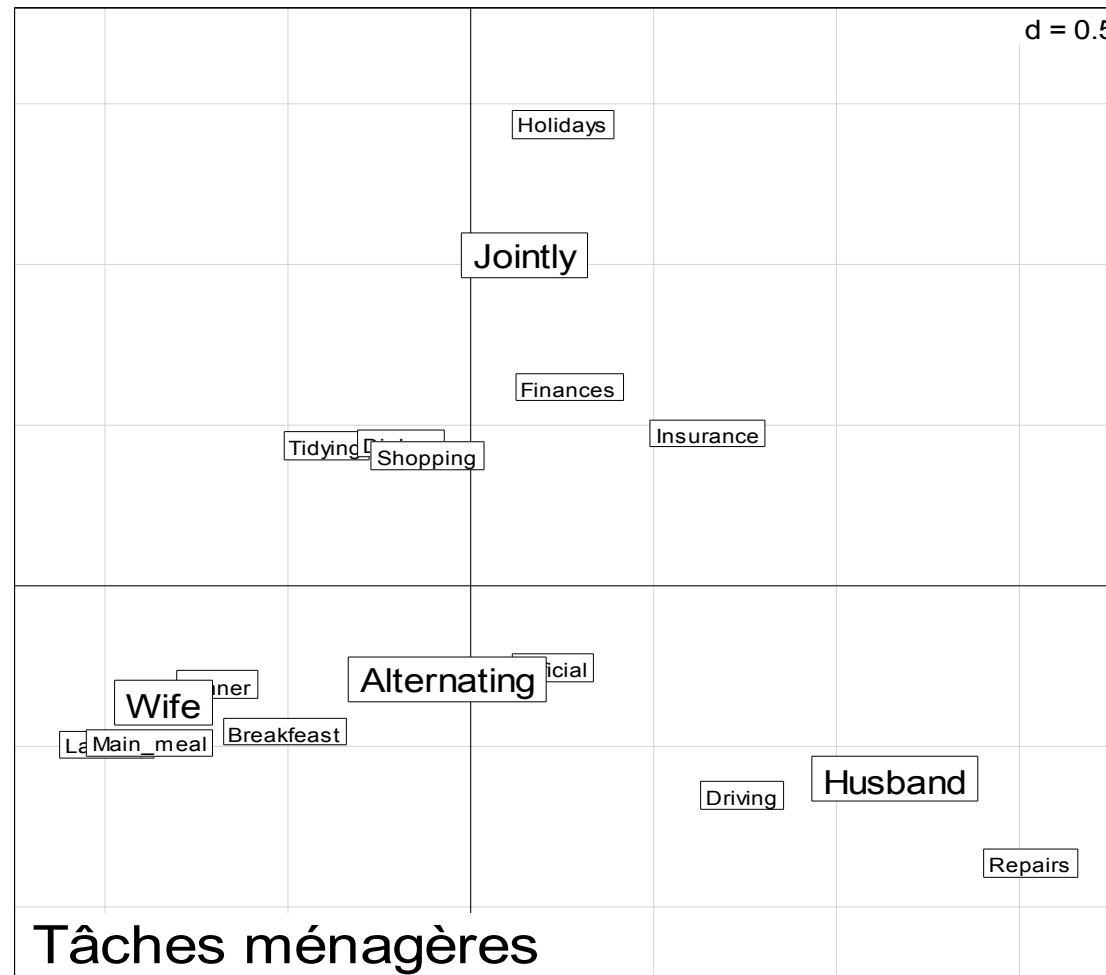
- Les valeurs propres en %

```
round(z$eig/sum(z$eig)*100,2)  
  
[1] 48.69 39.91 11.40
```



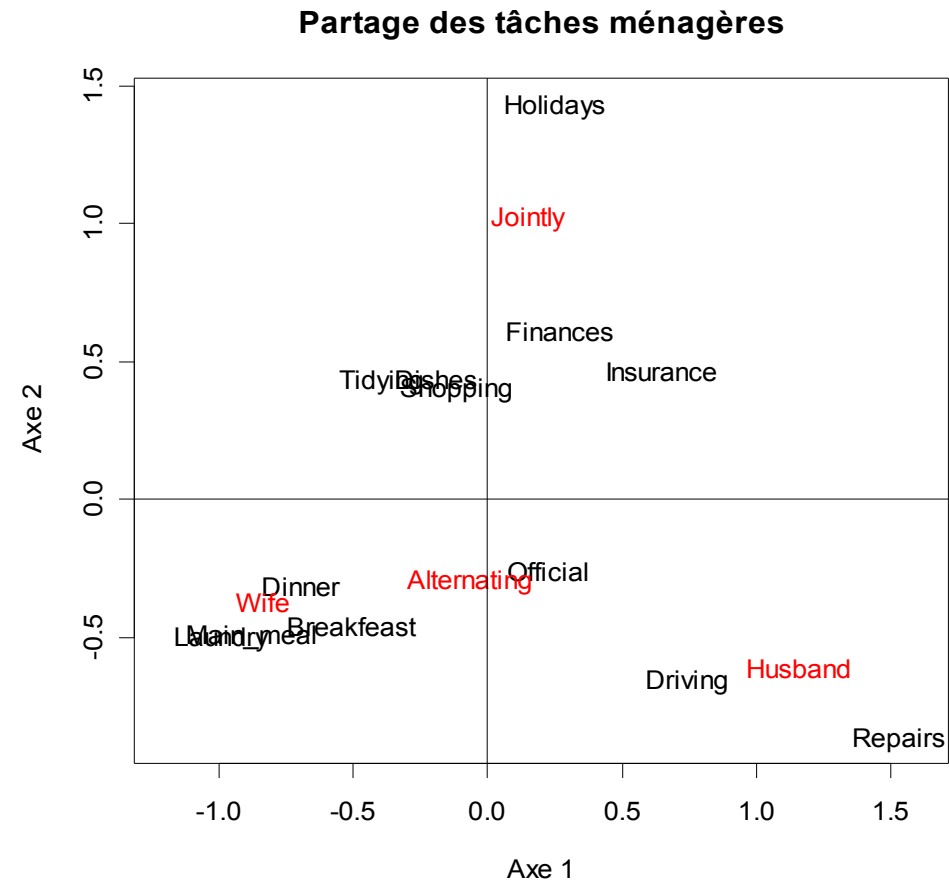
6. Une représentation du plan factoriel :

```
scatter.coa(z, method = 1, sub = "Tâches ménagères", posieig = "none")
```



7. Une autre représentation du plan factoriel :

```
plot(z$li[,1],z$li[,2],type="n",xlab="Axe 1",ylab="Axe 2",xlim=c(-1.2,1.6))
text(z$li[,1], z$li[,2], label=row.names(housetasks))
text(z$co[,1], z$co[,2], label= colnames(housetasks),col="red")
title("Partage des tâches ménagères")
abline(h=0,v=0)
```



8. Interprétation des facteurs : contributions "relatives" des colonnes :

- Contributions des colonnes à la construction des axes :

```
inertia.dudi(z, col.inertia = T)$col.abs
```

	<i>Comp1</i>	<i>Comp2</i>	<i>Comp3</i>
<i>Wife</i>	4446	1031	1082
<i>Alternating</i>	10	278	8255
<i>Husband</i>	5423	1779	613
<i>Jointly</i>	120	6912	50
<i>Somme</i>	10000	10000	10000

9. Interprétation des facteurs : contributions "relatives" des lignes :

- Contributions des lignes à la construction des axes :

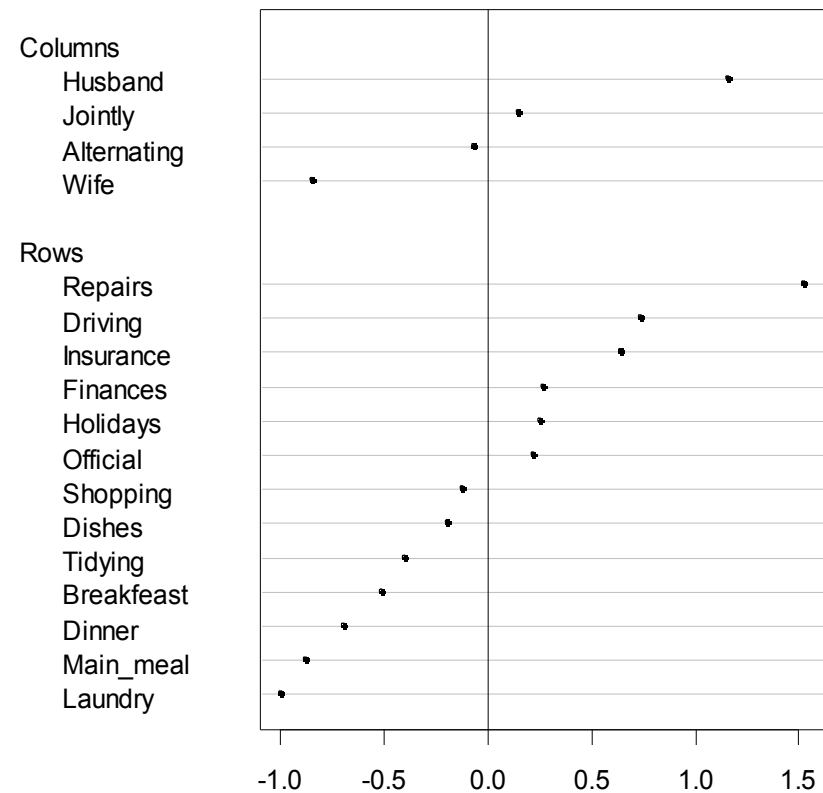
```
inertia.dudi(z, row.inertia = T)$row.abs
```

	<i>Axis1</i>	<i>Axis2</i>	<i>Axis3</i>
<i>Laundry</i>	1829	556	797
<i>Main_meal</i>	1239	474	186
<i>Dinner</i>	547	132	210
<i>Breakfeast</i>	382	370	307
<i>Tidying</i>	200	297	49
<i>Dishes</i>	43	284	363
<i>Shopping</i>	18	252	222
<i>Official</i>	52	80	3694
<i>Driving</i>	808	765	1860
<i>Finances</i>	88	556	6
<i>Insurance</i>	615	402	525
<i>Repairs</i>	4073	1588	1660
<i>Holidays</i>	108	4245	121
Somme	10000	10000	10000

10.A.F.C. Simples, une aides à l'interprétation - axe 1 :

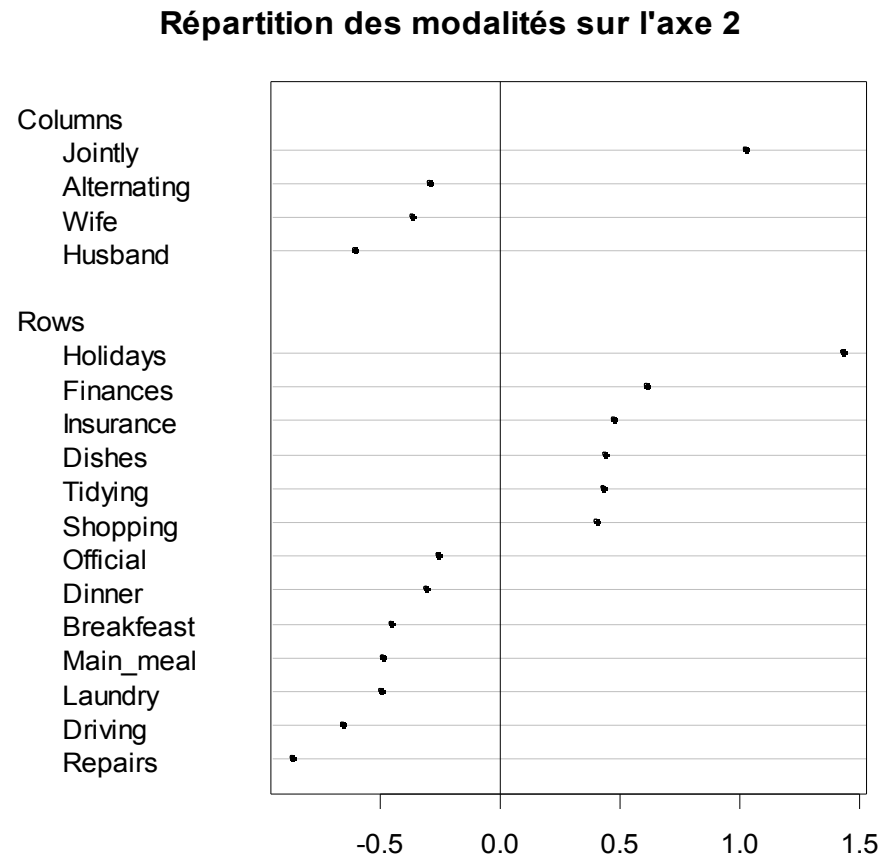
```
score.coa (corr,xax = 1,dotchart = TRUE)  
title("Répartition des modalités sur l'axe 1")  
abline(v=0)
```

Répartition des modalités sur l'axe 1



11.A.F.C. Simples, une aides à l'interprétation - axe 2 :

```
score.coa (corr,xax = 2,dotchart = TRUE)  
title("Répartition des modalités sur l'axe 2")  
abline(v=0)
```



12.L'A.F.C.s, le programme complet :

```
#les données
library(ade4)
data(housetasks)
#l'AFC simple
z<-dudi.coa(df = housetasks, scannf = F, nf = 3)
#L'éboulis des valeurs propres
inertie<-z$eig/sum(z$eig)*100
barplot(inertie,ylab="% d'inertie",names.arg=round(inertie,2))
title("Eboulis des valeurs propres en %")
#Les valeurs propres en %
round(z$eig/sum(z$eig)*100,2)
#une représentation du plan factoriel
scatter.coa(z, method = 1, sub = "Tâches ménagères", posieig = "none")
#les contributions relatives
inertia.dudi(z,col.inertia = T)$col.abs
inertia.dudi(z,row.inertia = T)$row.abs
#aide graphique à l'interprétation des axes
score.coa (corr,xax = 1,dotchart = TRUE)
title("Répartition des modalités sur l'axe 1")
abline(v=0)
score.coa (corr,xax = 2,dotchart = TRUE)
title("Répartition des modalités sur l'axe 2")
abline(v=0)
```

13. Quelques références :

- **Logiciel R :**

R Development Core Team (2004). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
ISBN 3-900051-07-0, URL

<http://www.R-project.org>.

- **Bibliothèque ade4 :**

Jean Thioulouse, Anne-Beatrice Dufour and Daniel Chessel (2004). ade4: Analysis of Environmental Data : Exploratory and Euclidean methods in Environmental sciences. R package version 1.3-3.

<http://pbil.univ-lyon1.fr/ADE-4>

Mailing list: <http://pbil.univ-lyon1.fr/ADE-4/adelist.html>

14. Table des matières

1. Avant propos :.....	3
2. Rappels :.....	4
3. Les données d'exemple, un tableau de contingence :.....	5
4. Lecture des données :.....	6
5. L'A.F.C.s. - fonction dudi.coa () de la bibliothèque ade4 :.....	7
6. Une représentation du plan factoriel :.....	8
7. Une autre représentation du plan factoriel :.....	9
8. Interprétation des facteurs : contributions "relatives" des colonnes :.....	10
9. Interprétation des facteurs : contributions "relatives" des lignes :.....	11
10. A.F.C. Simples, une aide à l'interprétation - axe 1 :.....	12
11. A.F.C. Simples, une aide à l'interprétation - axe 2 :.....	13
12. L'A.F.C.s, le programme complet :.....	14
13. Quelques références :.....	15