

Université de Carthage
Ecole Supérieure de la Statistique et de l'Analyse de l'Information

Examen d'Analyse des Données

1 ère année du cycle de formation d'ingénieurs

Durée de l'épreuve : 1 heure 30 - Documents non autorisés

Nombre de pages : 4 - Date de l'épreuve : 17 mai 2018

Exercice 1 : On a effectué une étude sur le temps de travail personnel hebdomadaire consacré par 210 étudiants d'une promotion à l'approche d'une session d'examen. Les questions posées étaient les suivantes :

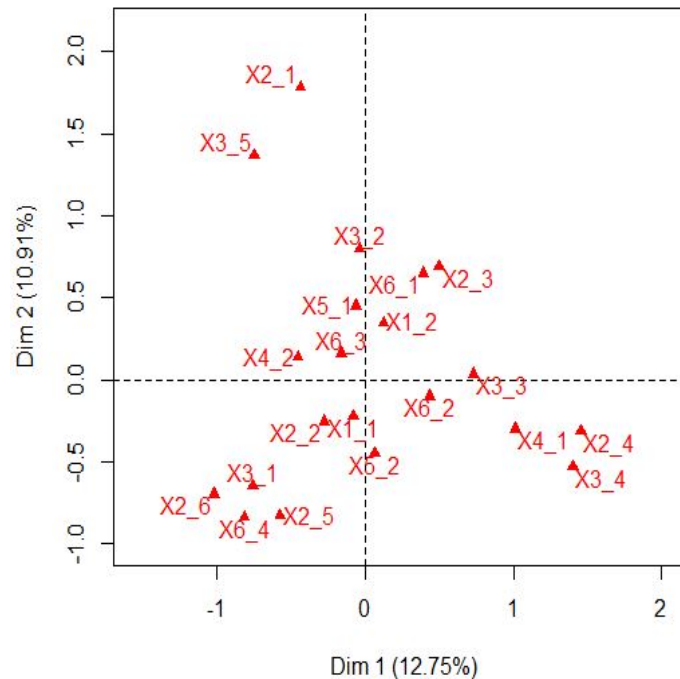
- Sexe (X_1) : 1-Masculin, 2-Féminin ;
- Catégorie socio-professionnelle du père (X_2) : 1- Sans profession ou chômeur, 2-Salarié, 3-Cadre salarié, 4-Profession libérale, 5-Commerçant ou artisan, 6-Agriculteur ;
- Catégorie socio-professionnelle de la mère (X_3) : 1- Sans profession ou chômeur, 2-Salarié, 3-Cadre salarié, 4-Profession libérale, 5-Commerçant ou artisan ;
- Etes vous membre d'une association sportive, musicale ou autre ? (X_4) : 1-Oui, 2-Non ;
- Pratiquez vous souvent des activités de bricolage ou jardinage ou lecture non scolaire ou autre ? (X_5) : 1-Oui, 2-Non ;
- Combien d'heures de travail personnel avez vous consacré à vos études ? (X_6) : 1- Moins de 10 heures, 2-Entre 10 et 20 heures, 3-Entre 20 et 30 heures, 4-Plus de 30 heures.

Soit `temps` la base de données sur le logiciel R obtenue à l'issue de cette étude. On a effectué une Analyse des Correspondances Multiples (ACM) sur cette base de données sur R :

```
library(FactoMineR)
temps.acm <- MCA(temps,ncp=2,graph=T)
temps.acm$eig[,1]
[1] 0.31863241 0.27270015 0.25624849 0.21782230 0.19793709 0.18718962
[7] 0.18432574 0.16792033 0.15610855 0.14479893 0.09612180 0.09050967
[13] 0.08030605 0.07196292 0.05741593
```

1. Rappeler la formule donnant le nombre de valeurs propres non nulles et non triviales d'une ACM.
2. Calculer de 2 manières différentes l'inertie totale de cette ACM.

3. Combien d'axes devrait-on retenir ? Justifier votre réponse.
4. Interpréter la carte donnée ci-dessous.



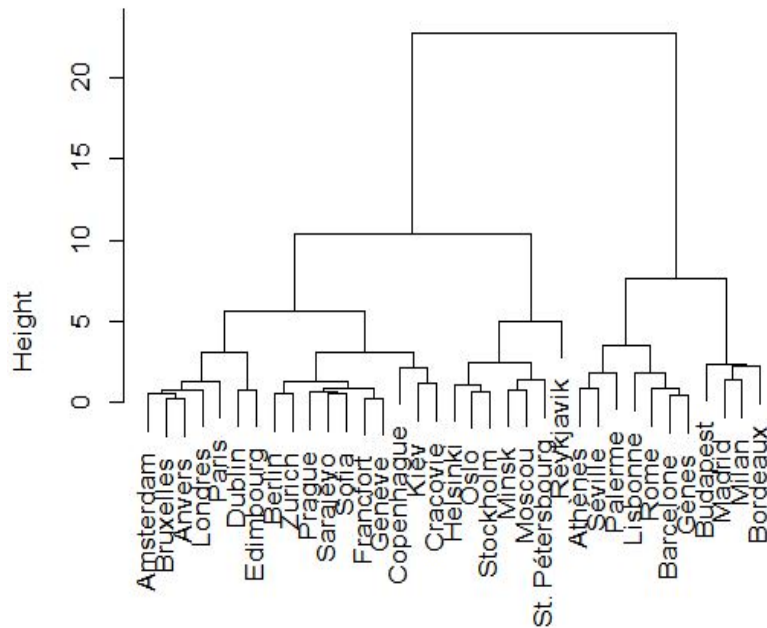
5. On voudrait effectuer une classification automatique avec le critère de Ward sur les 210 étudiants à partir de leurs coordonnées sur les deux premiers axes factoriels issus de cette ACM. Donner la commande à exécuter pour obtenir une telle classification.

Exercice 2 : On s'intéresse au climat des différents pays d'Europe. Pour cela, on a recueilli les températures moyennes des 12 mois de l'année (en degré Celcius) pour 23 grandes villes européennes. En plus des températures mensuelles, on donne pour chaque ville, sa région (Nord, Sud, Est, Ouest) .

On a effectué une classification automatique des 23 villes en exécutant le script suivant :

```
library(FactoMineR)
library(cluster)
temperature <- read.table("temperat.csv",header=TRUE, sep=";", dec=".", row.names=1)
classif<-agnes(scale(temperature[,1:12]), method="ward")
plot(classif,xlab="individuals",main="")
title("Dendrogram")
```

1. La hiérarchie obtenue suite à l'exécution de ce script est donnée ci-dessous. Déterminer le meilleur nombre de classes à retenir.



Dans la suite on considère la partition en 3 classes.

```
P3<-as.factor(cutree(classif,k=3))
```

```
P3
```

```
[1] 1 2 1 1 2 1 1 3 1 1 2 1 2 3 3 3 1 1 3 2 1 1 3 1 2 2 1 1 1 2 2 2 2 3 1
```

```
Levels: 1 2 3
```

```
summary(P3)
```

```
1 2 3
```

```
17 11 7
```

2. On voudrait décrire les classes obtenues à l'aide de la fonction `catdes` du package `FactoMineR`. Indiquer la démarche à suivre.

3. En utilisant la fonction `catdes` pour décrire les 3 classes, nous avons obtenu les résultats suivants :

```
$category
```

```
$category$'1'
```

	Cla/Mod	Mod/Cla	Global	p.value	v.test
Région=Ouest	88.88889	47.058824	25.71429	0.006885958	2.702310
Région=Sud	10.00000	5.882353	28.57143	0.004979077	-2.808384

```
$category$'2'
```

	Cla/Mod	Mod/Cla	Global	p.value	v.test
Région=Sud	90	81.81818	28.57143	7.310189e-06	4.484451

```
Région=Nord      0  0.00000 22.85714 3.124901e-02 -2.153887
```

```
$category$'3'
```

```
      Cla/Mod  Mod/Cla  Global    p.value  v.test
Région=Nord    50 57.14286 22.85714 0.03652157 2.091072
```

```
$quanti
```

```
$quanti$'1'
```

```
      v.test Mean in category Overall mean sd in category Overall sd    p.value
Août    -2.219019      17.54118      18.98000      1.372014      3.674297 0.026485453
Juin     -2.326752      16.07059      17.41429      1.413283      3.272495 0.019978451
Juillet  -2.591233      18.01176      19.62286      1.443803      3.523236 0.009563284
```

```
$quanti$'2'
```

```
      v.test Mean in category Overall mean sd in category Overall sd    p.value
Septembre 5.001680      20.763636      15.631429      2.246724      4.050592 5.683295e-07
Août      4.934312      23.572727      18.980000      2.089407      3.674297 8.043381e-07
Juillet   4.914510      24.009091      19.622857      1.983777      3.523236 8.900480e-07
Avril     4.848655      13.890909      9.282857      1.662667      3.751684 1.243015e-06
Juin      4.829854      21.418182      17.414286      1.785710      3.272495 1.366335e-06
Octobre   4.797991      16.181818      11.002857      2.629489      4.261018 1.602651e-06
Mai       4.713150      17.763636      13.911429      1.703618      3.226477 2.439162e-06
Novembre  4.558695      11.263636      6.065714      3.253606      4.501107 5.147236e-06
Mars      4.491300      10.681818      5.228571      2.419173      4.793065 7.078985e-06
Février   4.291443      8.109091      2.217143      3.316749      5.419831 1.775155e-05
Décembre  4.157556      8.036364      2.880000      3.639998      4.895934 3.216699e-05
Janvier   4.069622      6.936364      1.345714      3.768837      5.422985 4.708957e-05
```

```
$quanti$'3'
```

```
      v.test Mean in category Overall mean sd in category Overall sd    p.value
Juillet   -2.465948      16.6428571      19.622857      2.3243169      3.523236 1.366510e-02
Juin      -2.698164      14.3857143      17.414286      2.2216008      3.272495 6.972317e-03
Août      -2.954008      15.2571429      18.980000      2.0091627      3.674297 3.136764e-03
Mai       -3.327865      10.2285714      13.911429      1.9789711      3.226477 8.751418e-04
Septembre -3.570036      10.6714286      15.631429      1.2925217      4.050592 3.569320e-04
Décembre  -3.595053      -3.1571429      2.880000      2.0091627      4.895934 3.243250e-04
Janvier   -3.664996      -5.4714286      1.345714      2.8348415      5.422985 2.473429e-04
Novembre  -3.688326      0.3714286      6.065714      0.9602296      4.501107 2.257340e-04
Octobre   -3.814040      5.4285714      11.002857      0.5897076      4.261018 1.367131e-04
Février   -3.889984      -5.0142857      2.217143      2.6117474      5.419831 1.002507e-04
Avril     -4.049860      4.0714286      9.282857      1.1335334      3.751684 5.124836e-05
Mars      -4.170992      -1.6285714      5.228571      1.3477116      4.793065 3.032764e-05
```

En vous basant sur ces résultats, donner une description des 3 classes.

4. Afin d'étudier de manière plus précise le lien entre la partition en 3 classes et la région, quelle méthode préconiserez-vous ?