

Université de Kairouan
Institut Supérieur des Mathématiques Appliquées & Informatique
Examen Juin 2013

Module	Analyse des données
Auditoire	1ière Année Mastère Ingénierie financière
Enseignant	Mohamed Essaied Hamrita
Durée	Deux heures

La présentation, la lisibilité, l'orthographe, la qualité de la rédaction et la clarté entreront pour une part importante dans l'appréciation des copies. Aucun document n'est autorisé.

Exercice 1 (8 points) :

Les rendements de 103 semaines de cinq actifs financiers listés dans New York Stock Exchange sont analysés. Dans l'annexe 1, vous trouverez les sorties du logiciel suite à l'application d'une analyse en composantes principales (ACP) à la matrice des corrélations.

- 1) Décrire le jeu de données (nombre d'individus, nombre de variables, nature des variables). (0.5 pts)
- 2) Rappeler le principe de la méthode ACP. Est-elle applicable dans cette analyse? (1 pt)
- 3) En examinant le tableau 1 de l'annexe 1 :
 - a) Compléter le tableau. (0.5 pts)
 - b) Combien d'axes factoriels doit-on retenir? Justifier votre réponse. (0.5 pts)
- 4) En examinant le diagramme des variables :
 - a) Identifier les principales liaisons entre les variables. (1 pt)
 - b) Interpréter la position des variables par rapport au cercle des corrélations. (1 pt)
- 5) Donner une interprétation au premier axe principal suivant la position des variables. (1 pt)
- 6) Préciser les variables les mieux représentées sur le premier plan factoriel. (1.5 pts)
- 7) Interpréter le graphique des projections des individus. (1 pt)

Exercice 2 (7 points) :

On propose d'étudier un tableau de données qui croise 10 catégories socio-professionnelles (CSP) avec 8 modes d'hébergements lors des vacances. Dans l'annexe 2, vous trouverez les sorties du logiciel R.

- 1) En examinant le tableau 1 et la figure d'éboullis des valeurs propres, déterminer le nombre d'axes principaux à retenir. Justifier. (0.5 pts)
- 2) Quels sont les modes d'hébergements les plus fréquemment cités? Interpréter. (0.5 pts)
- 3) Quelles sont les modes d'hébergements les mieux représentées sur le premier plan factoriel? (2 pts)

4) Quels sont les catégories socio-professionnelles contribuant beaucoup au premier axe factoriel ? Au second axe factoriel ? Interpréter chaque axe en termes du mode d'hébergement et de classe CSP. (1.5 pts + 1.5 pts + 1 pt = 4 pts)

Exercice 3 (5 points) :

Une enquête réalisée sur un échantillon comprenant 10 individus qui ont répondu à trois questions :

- Q1 : Lieu de résidence (V si Ville et B si Banlieue).
- Q2 : CSP (C si Cadre, O si Ouvrier et P si Profession libérale).
- Q3 : Type d'habitat (Pr si Propriétaire, L si Locataire et S si Sans logement).

Le résultat de cette enquête est donnée dans le tableau suivant :

	1	2	3	4	5	6	7	8	9	10
Lieu de résidence	B	V	V	V	B	B	V	B	V	B
CSP	P	P	O	C	O	C	P	O	C	O
Type d'habitat	Pr	Pr	L	L	Pr	Pr	Pr	L	L	S

- 1) Transformer ce tableau en tableau codé. Pour le codage on procède comme suit :
Q1 : 1 si ville et 2 si banlieue ; Q2 : 1 si ouvrier, 2 si cadre et 3 si profession libérale ; Q3 : 1 si sans logement, 2 si locataire et 3 si propriétaire. (1 pt)
- 2) Transformer le tableau codé en un tableau logique. (1pt)
- 3) En déduire le tableau de Burt. (1.5 pts)
- 4) Déterminer la distribution marginale des lignes du dernier tableau. (1.5 pts)

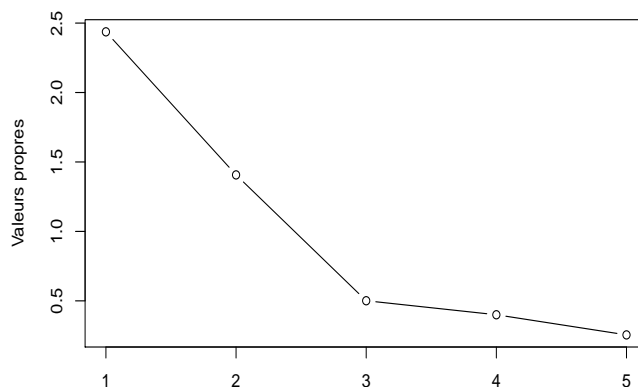
Bon Travail

Annexe 1 : listening exercice 1

Tableau 1: Valeurs propres
 Variances % Variances % Var-Cum

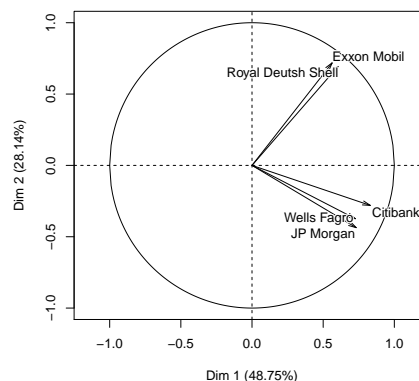
Axe	Variances	% Variances	% Var-Cum
Axe 1	2.437	48.745	48.745
Axe 2	1.407	28.140	76.886
Axe 3	0.501	10.010	86.896
Axe 4	0.400	8.001	94.897
Axe 5	?	?	100.000

Eboulis des valeurs propres

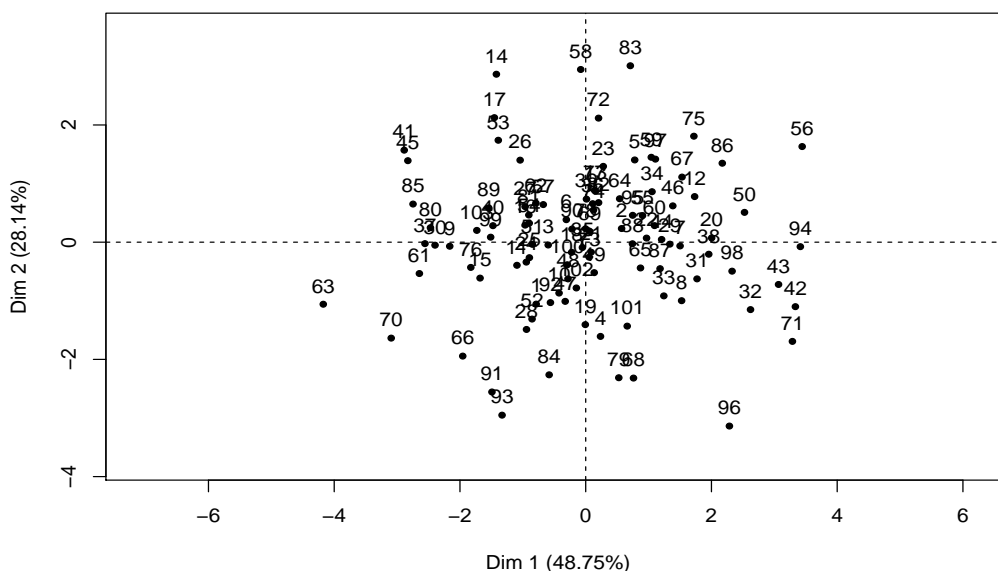


	Coordonnées		Contribution		Cos2	
	Dim.1	Dim.2	Dim.1	Dim.2	Dim.1	Dim.2
JP Morgan	0.732	-0.437	22.004	13.543	0.536	0.191
Citibank	0.831	-0.280	28.346	5.591	0.691	0.079
WellsFagro	0.726	-0.374	21.638	9.934	0.527	0.140
RoyalDeutsh Shell	0.605	0.694	15.004	34.227	0.366	0.482
ExxonMobil	0.563	0.719	13.009	36.705	0.317	0.516

Variables factor map (PCA)



Individuals factor map (PCA)

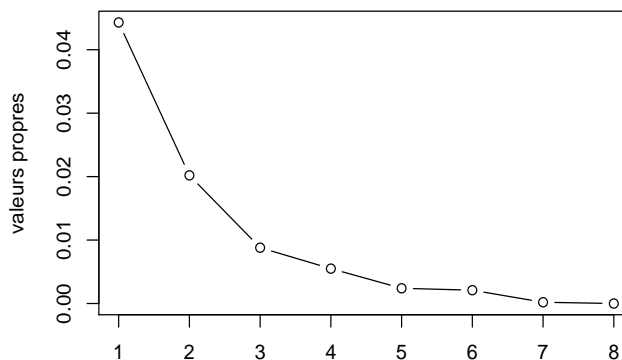


Annexe 2 : listening exercice 2

Tableau 1

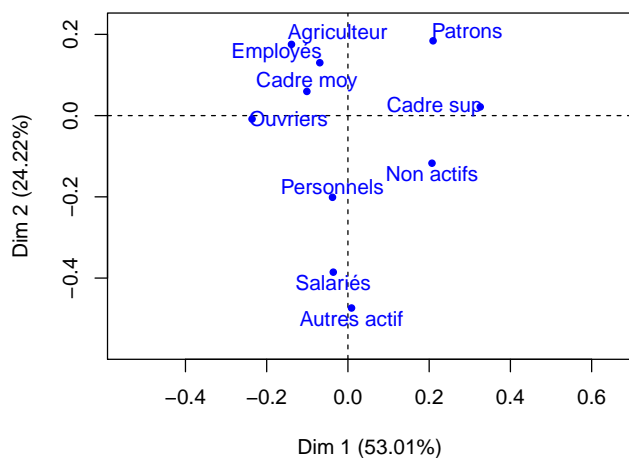
	val. propres	% variance	% cum. var
dim 1	0.0443	53.0077	53.0077
dim 2	0.0202	24.2209	77.2285
dim 3	0.0088	10.5319	87.7605
dim 4	0.0055	6.5373	94.2977
dim 5	0.0024	2.8486	97.1463
dim 6	0.0021	2.5659	99.7122
dim 7	0.0002	0.2878	100.000
dim 8	0.0000	0.0000	100.000

Eboulis des valeurs propres



	Poids	Coordonnées		Contributions		Cos2	
		Dim 1	Dim 2	Dim 1	Dim 2	Dim 1	Dim 2
hôtel	0.146	0.329	0.175	35.724	21.975	0.729	0.205
locat	0.109	-0.078	0.138	1.478	10.246	0.116	0.368
propri	0.072	0.408	-0.131	26.934	6.105	0.655	0.068
parent	0.411	-0.064	-0.143	3.753	41.741	0.153	0.778
amis	0.060	0.113	-0.009	1.718	0.022	0.202	0.001
tente	0.124	-0.301	0.141	25.436	12.121	0.724	0.158
villag	0.043	-0.213	0.163	4.410	5.615	0.471	0.274
divers	0.036	-0.083	0.111	0.547	2.176	0.066	0.120

Projections du classe CSP



Projections du mode d'hébergement

