

Séries d'exercices ANAD,2CS SIT ACP

Mme L;Hamdad

Exercice1: Six élèves ont obtenus les notes suivantes en trois matières:

	Français	Maths	Histoire
1	9	12	10
2	15	9	10
3	5	10	8
4	11	13	14
5	11	13	8
6	3	15	10

- Effectuer une ACP normée sur ce tableau, avec représentation simultanée des élèves et des matières sur le plan d'inertie maximum.

- Donner une interprétation à partir du graphe.

Soit les notes des six élèves en Anglais:

Ang	10	11	5	14	3	16
-----	----	----	---	----	---	----

- Représenter Cette matière supplémentaire dans le graphe précédent.

• On vérifiera que sont vecteurs propres $(1, -1, 0)$, $(1, 1, 1 + \sqrt{3})$

Exercice2:

Soit le tableau suivant qui représente les notes obtenues par 10 candidats à l'examen du permis de conduire: 2(Bon), 1(Moyen), Mauvais(0).

	1	2	3	4	5	6	7	8	9	10
R	1	1	1	1	1	1	1	1	2	0
C	2	2	2	2	0	0	0	0	1	1
V	1	1	1	1	1	1	1	1	0	2

- Représenter au mieux les candidats par l'ACP. Quelle est la qualité de représentation avec 1 ou 2 facteurs.

- Représenter les 3 variables

- Représenter les sujets supplémentaires suivant:

11	0	0	1
12	2	1	2

Exercice3:

Une étude réalisée par un médecin dans deux services hospitaliers spécialisés dans le traitement et l'accompagnement d'enfants handicapés a été menée sur une dizaine d'années auprès de 96 patient en tout qui ont été traités au moins pendant 2 ans. Pour chacun de ces enfants, on a relevé les variables suivantes

- educ : Educatif . Nombre de séances de psychologie.
- reed : Rééducatif .Nombre de séances kiné, .
- hand : Acceptation du handicap.
- barr : Barrières. Mesure l'importance des barrières (trottoirs,escaliers. . .) .
- soc : Sociabilité. Variable qui traduit l'attitude de l'enfant vis-à-vis des autres
- inte : Intellectuel. Mesure du handicap intellectuel de l'enfant.
- supe : Membres supérieurs. Mesure de l'aisance avec laquelle l'enfant peut se servir de ses membres sup :
- auto : Autonomie. Mesure, l'autonomie de l'enfant.

On traite ces données par une analyse en composantes principales normée. Les principaux résultats de cette ACP sont indiqués ci-dessous.

Moyennes 78,23 162,68 107,75 10,08 10,66 5,69 7,43 9,18

Ecart type 36,41 64,31 111,86 3,46 3,70 5,53 3,04 5,62

Matrice de Corrélations

	<i>educ</i>	<i>reed</i>	<i>hand</i>	<i>barr</i>	<i>soc</i>	<i>inte</i>	<i>sup e</i>	<i>auto</i>
<i>educ</i>	1	-0,11	-0,21	0,01	0,23	0,13	0,02	0,26
<i>reed</i>		1	0,22	0,03	0,36	0,22	0,44	0,37
<i>hand</i>			1	0,37	0,36	0,06	0,16	0,17
<i>barr</i>				1	0,26	0,12	0,21	0,10
<i>soc</i>					1	0,47	0,46	0,68
<i>inte</i>						1	0,30	0,43
<i>sup e</i>							1	0,56
<i>auto</i>								1

Val. Propres & stat. associées.

	Val prop	%d'inertie	%d'inertie cum
1	3	37,56	37,56
2	1,39	-	54,91
3	1,09	13,62	68,53
4	0,74	9,21	77,74
5	0,67	8,41	86,15
6	0,55	6,86	93,01
7	0,33	-	97,17
8	0,23	2,83	100

Projections, contributions et qualités de représentation des variables

	W ₁	W ₂	W ₃	C _{ab1}	C _{ab2}	C _{ab3}	cos ² ₁	cos ² ₁₂	cos ² ₁₂₃
<i>educ</i>	-0,18	-0,71	-0,49	0,01	-	0,22	0,03	0,54	0,77
<i>reed</i>	-0,58	0,17	0,56	0,11	0,02	0,29	-	0,36	0,68
<i>hand</i>	-0,41	0,71	-0,21	0,06	0,36	0,04	0,17	0,67	-
<i>barr</i>	-0,35	0,47	-0,63	0,04	0,16	0,37	0,12	0,34	0,74
<i>soc</i>	-0,84	-0,06	-0,16	0,24	0	0,02	0,71	-	0,74
<i>inte</i>	-0,60	-0,25	-0,04	0,12	0,05	0	0,36	0,42	0,42
<i>supe</i>	-0,76	-0	0,24	0,19	0	0,05	0,57	0,57	0,63
<i>auto</i>	-0,84	-0,26	0,06	0,24	0,05	0	0,71	0,78	0,78

Questions:

- 1) Représenter les variables sur le premier et le troisième plan factoriel en représentant le cercle de corrélation.
- 2) Dans l'espace des individus, calculer le vecteur propre associé à la plus grande valeur propre
- 3) Calculer les valeurs manquantes
- 4) Etude des valeurs propres
 - a) A quoi correspond la somme des valeurs propres ?
 - b) On choisit de n'étudier que les trois premières composantes principales. Justifier ce choix en analysant le tableau des valeurs propres.

- 5) Etude du tableau des corrélations
- Quelles sont les variables le plus fortement corrélées entre elles ?
 - Y a-t-il des variables pratiquement non corrélées ?
- 6) Etude des qualités de représentation des variables
- Quelles sont les variables les mieux représentées sur le premier axe factoriel ? Quelle est celle qui est la plus mal représentée sur cet axe.
 - Deux des variables sont pratiquement indépendantes de la deuxième composante principale. Lesquelles ?
 - Globalement, comment peut-on qualifier la représentation de l'ensemble des variables lorsqu'on se limite aux trois premières composantes principales ?
- 7) Etude du premier axe principal, du point de vue des variables.
- Quelles sont les variables qui ont le plus contribué à la formation du premier axe principal ?
 - Quelle remarque peut-on faire concernant le signe des saturations des variables sur le premier axe ? Comment appelle-t-on l'effet ainsi mis en évidence ?

Quelle interprétation peut-on faire de ce premier axe ?

- 8) Etude du deuxième axe principal, du point de vue des variables.

Quelles sont les variables dont la contribution à la formation de la deuxième composante principale est supérieure à la moyenne ? Pour chacune d'elles, préciser le signe de la projection correspondante. Caractériser cet axe en termes d'opposition entre variables.

- 9) Etude du troisième axe principal, du point de vue des variables.

Quelles sont les variables dont la contribution à la formation de la troisième composante principale est supérieure à la moyenne (des contributions) ? Pour chacune d'elles, préciser le signe de la projection correspondante. Caractériser cet axe en termes d'opposition entre variables.

- 10) Etude du nuage des individus:

On donne ci-dessous le tableau (I) des scores des individus, (groupe1) en se limitant aux scores les plus extrêmes sur le premier axe et le tableau (II) des scores des individus, (groupe2) en se limitant aux scores les plus extrêmes sur le deuxième axe.

(I)

	Axe1	Axe2	Axe3
Victor	3,27	0,26	1,61
benoit	3,1	-0,5	0,35
nico	-3,21	0,47	1,65
john	-3,39	0,87	0,53
floran	-3,87	0,73	0,98

(II)

	Axe1	Axe2	Axe3
sébastien	0,36	2,05	-0,22
loriane	-1,99	-2,03	0,19
serge	2,04	-2,55	-1,29
daniel	-1,33	-2,95	-0,41
laurent	-0,85	-3,33	-0,26
grégoire	-1,99	2,53	17,2

Indiquer quelles sont les caractéristiques des sujets du groupe1 et celle des sujets du groupe2.

- 11) On donne l'individu suivant zinedine (68 126 105 11 5 1 5 3 2), calculer sa projection sur l'axe1, et caractériser-le.

Exercice 4: Une étude sur les téléphones mobiles aboutissant à un tableau comportant un nombre important de caractéristiques touchant à l'utilisation de

ces appareils a été effectuée. Une part de ces caractéristiques sont mesurées selon une echelle comportant cinq degrés. Dans le but de fournir une cartographie de ces appareils pouvant servir à les différencier selon le maximum de ces critères il a été construit un tableau de données ci dessous, retenant les variables jugées les plus pertinentes :

	commodité	acoustique	tps charge	ato veille	autoconv	puisGSM	sensiGSM	sensides
philips	5	3	5	5	5	5	3	3
nokia3310	4	4	2	5	5	4	4	4
nokia6210	4	3	3	5	5	4	4	4
sony	5	4	4	4	2	4	4	4
siemensSL45	5	3	5	5	1	4	4	4
eircssonT20	3	4	3	5	5	3	3	3
siemensS35I	4	4	3	5	2	4	4	4
alcatel	3	4	4	5	2	5	3	3
samsung3	4	3	4	3	2	5	3	3
trium	3	3	4	5	2	5	4	4
motorola	3	4	3	5	2	4	4	4
panasonic	4	4	5	4	3	3	1	2
alcatel501	4	2	4	5	2	5	2	3
ericsson2628	3	3	1	3	3	4	3	4
sagem	3	3	3	5	3	3	1	2

L'analyse sera faite sur les propriétés numériques des variables, pas leur interprétation. Une ACP normée a été effectuée sur ce tableau, la sortie avec le logiciel R donne les résultats suivant (Matrice de corrélation, valeurs propres, contribtions absolues des individus (en 10000 ième) et projection des variables, sur les cinq axes):

$$\mathbf{R} = \begin{pmatrix} 1 & -0.12 & 0.54 & -0.03 & -0.01 & 0.17 & 0.21 & 0.15 \\ & 1 & -0.12 & 0.06 & 0.11 & -0.43 & 0.23 & 0.09 \\ & & 1 & 0.19 & -0.30 & 0.25 & -0.18 & -0.36 \\ & & & 1 & 0.18 & -0.03 & 0.16 & 0.05 \\ & & & & 1 & -0.27 & -0.04 & -0.11 \\ & & & & & 1 & 0.34 & 0.29 \\ & & & & & & 1 & 0.93 \\ & & & & & & & 1 \end{pmatrix}$$

Valeurs propres.

2.23	1.94	1.30	1.01	0.89	0.47	0.13	0.02
------	------	------	------	------	------	------	------

	Axe1	Axe2	Axe3	Axe4	Axe5
Philips	0	1096	1115	905	1342
Nokia3310	345	1340	389	178	519
Nokia6210	263	208	188	1118	793
Sony	536	65	161	2330	346
SiemensSL45	583	1168	381	327	65
EricssonT20	479	479	1637	66	0
SiemensS35i	478	163	118	302	311
Alcatel	3	20	4	1	209
Samsung 3	3	800	1687	513	543
Trium	549	96	243	461	1234
Motorola	303	526	2	90	1429
Panasonic	3127	80	356	1735	49
Alcatel501	102	1720	436	1634	49
Ericsson2628	12	987	4341	29	965
Sagem	3215	95	29	295	266

	V ₁	V ₂	V ₃	V ₄	V ₅
Commodite	-0.29	0.57	-0.46	0.23	0.52
Acoustiq	-0.05	-0.56	-0.44	0.58	-0.19
TpsCharge	0.13	0.82	-0.45	0.17	-0.14
AutoVeille	-0.11	-0.04	-0.66	-0.53	-0.49
Autoconv	0.18	-0.47	-0.38	-0.47	0.56
PuisGSM	-0.52	0.57	0.30	-0.31	-0.07
SensiGSM	-0.96	-0.19	-0.14	0.04	-0.01
SensiDCS	-0.95	-0.24	0.07	0.02	0.04

- 1- Quel est le couple de variables qui sont les plus corrélées entre elles ? Les moins corrélées entre elles ? Les plus opposées ?
- 2- Faire une représentation en histogramme des valeurs propres. Combien de composantes principales faut-il retenir ? Quel est le pourcentage d'inertie expliquée par le sous-espace principal correspondant ? Si on ne retient que le premier plan principal, quel est le pourcentage d'inertie expliquée ?
- 3- Représenter les variables sur le premier plan principal avec cercle de corrélation.
- 4- Quelles sont les variables qui déterminent le premier axe principal ? La proximité de ces variables sur le cercle est-elle cohérente avec ce que l'on sait des données d'origine ?
- 5- D'après la représentation (figure1), quels sont les individus qui déterminent le plus la première composante principale ? Quelle est leur contribution ?
- 6- Quel est le lien entre les variables de la question 4 et les individus de la question 5 ?

Une nouvelle ACP a été effectuée, en supprimant les individus perturbateurs et en les considérant comme éléments supplémentaires, les résultats de l'analyse (valeurs propres, corrélation entre variables et composantes et contributions absolues des individus) sont les suivants

Valeurs propres

2.73	1.72	1.32	0.93	0.68	0.43	0.17	0.02
------	------	------	------	------	------	------	------

	V_1	V_2	V_3	V_4	V_5
Commodite	-0.24	-0.68	0.12	-0.63	0.19
Acoustiq	0.71	-0.04	0.21	0.32	0.55
TpsCharge	-0.63	-0.67	0.22	0.13	0.18
AutoVeille	0.10	-0.40	0.70	0.41	-0.36
Autoconv	0.24	0.38	0.70	-0.45	-0.16
PuisGSM	-0.78	-0.14	-0.22	0.20	-0.19
SensiGSM	0.75	-0.59	-0.11	0.02	-0.06
SensiDCS	0.74	-0.38	-0.42	-0.12	-0.34

	Axe1	Axe2	Axe3	Axe4	Axe5
Philips	1382	195	1920	922	3
Nokia3310	1273	29	431	318	166
Nokia6210	334	12	324	865	1712
Sony	199	1064	320	563	2008
SiemensSL45	13	2968	118	93	1
EricssonT20	344	1676	2626	40	837
SiemensS35i	608	295	29	181	60
Alcatel	341	116	96	3310	557
Samsung 3	1295	226	1160	217	1473
Trium	6	244	385	1306	1922
Motorola	776	9	73	1425	6
Alcatel501	3200	182	6	4	1104
Ericsson2628	229	2981	2539	758	150

- 7.- Quels sont les individus qui ont été supprimés,.Expliquer pourquoi.
 8- Combien de valeurs propres faut-il retenir, et quelles proportion cela représente-t-il ? Comparer à la question 2.
 9- Pour chacune des trois premières composantes principales, donner les variables qui les caractérisent le plus (en fixant le seuil à 0.60).
 10.- Dans la projection des individus sur le premier plan principal, où se trouvent ceux qui ont des notes supérieures à la moyenne pour **PuisGSM** ? et pour **TpsCharge** ?
 11- Comment mesure-t-on la qualité de la représentation d'un individu par un axe principal ? Quels sont les deux individus les plus mal représentés sur le premier axe principal ?
 12- Comment mesure-t-on la qualité de la représentation d'un individu par un sous-espace principal ?
 13- Que peut-on dire de la position des deux éléments supplémentaires sur la projection des individus ? Est-ce qu'ils sont bien représentés par le sous-espace engendré par les 3 premiers axes ?

Leurs coordonnées sur les axes principaux sont :

panasonic	2.43	-2.34	1.99	0;12	3.52	2.43	0.03	0.92
sagem	2.13	-3;59	2.04	-0.52	1.15	3.08	-0.36	-0.20

Exercice5: Le tableau suivant résume le nombre de spécimens de 3 espèces de chenilles sur 7 arbres:

	1	2	3	4	5	6	7
Espèce 1	276	132	123	108	91	63	55
Espèce 2	29	49	49	69	37	328	20
espèce 3	14	12	32	42	12	77	16

- Faire une ACP à l'aide du logiciel R.

Exercice6: On effectue une ACP sur le tableau représentant la consommation annuelles pour 8 catégories socioprofessionnelles, les résultats sont comme suit::

Matrice de Corrélation

$$R = \begin{pmatrix} 1 & & & & & & & \\ -0.75 & 1 & & & & & & \\ 0.83 & -0.57 & 1 & & & & & \\ -0.89 & -0.90 & -0.73 & 1 & & & & \\ -0.66 & -0.30 & 0.52 & 0.40 & 1 & & & \\ -0.90 & -0.66 & 0.80 & 0.84 & 0.61 & 1 & & \\ -0.82 & -0.96 & -0.65 & 0.91 & -0.74 & -0.82 & 1 & \\ -0.85 & -0.78 & -0.82 & 0.72 & -0.55 & -0.73 & 0.85 & 1 \end{pmatrix}$$

Moyennes: 127.30 4.94 96.09 58.81 38.58 6.53 10.30 24.17

Ecart type: 27.90 3.03 38.66 31.23 7.87 2.81 4.45 12.33

Valeurs propres: 6.2059 0.8966 0.4204 0.3191 0.1437 0.0092 0.0048 0.0000

Taux d'inertie cumulés:

77.57 88.78 94.03 98.02 99.82 99.93 100 100

Composantes principales, projections des variables et contributions relatives:

	Y₁	Y₂		V¹	V²	C_{re}¹²
AGRI	3.366	-0.209	PAO	0.975	0.135	-
SAAG	3.529	-0.76	PAA	-0.86	0.440	0.9333
PRIN	-1.497	-0.407	BO	0.866	0.214	0.7957
CSUP	-4.285	0.189	AB	-0.927	0.250	0.9218
CMOY	-1.775	-0.801	POT	0.616	0.689	0.8541
EMPL	-0.886	-0.800	LEC	0.914	0.117	0.8489
OUVR	0.993	-0.387	RAI	-0.930	0.293	0.9501
INAC	0.565	2.343	PLP	-0.905	-0.038	-

1- L'ACP est-elle normée Analyser la matrice de corrélation

2- Dans l'espace des variables, donner le vecteur propre associé à la plus grande valeur propre

3- Quelles sont les individus et les variables qui contribuent le plus à la construction des deux premiers axes. Donner une signification à ces axes.

4- Compléter la colonne des contributions relatives. Que constatez-vous.

5- Représenter graphiquement les individus et les variables sur le premier plan factoriel, en déterminant sa qualité globale de représentation. La variable "POT" est-elle bien représentée par le premier axe?

6- Reconstituer la première ligne et la première colonne du tableau initial à partir de la première valeur propre.

Exercice7: Il s'agit de visualiser la teneur en **Al, Fe, Mg, Ca, Na** de 26 poteries par une ACP. Les résultats sont comme suit

Matrice de corrélation

$$\begin{pmatrix} 1 & & & & \\ -0.79 & 1 & & & \\ 0.80 & 0.90 & 1 & & \\ 0.76 & 0.77 & 0.84 & 1 & \\ 0.47 & 0.66 & 0.64 & 0.48 & 1 \end{pmatrix}$$

Valeurs propres:

3,88	•	0.24	0.18	0.08
------	---	------	------	------

Les vecteurs propres et projections des variables:

u₁	u₂	u₃	u₄	u₅		V₁	V₂	V₃	V₄	V₅
0.45	•	0.69	-0.44	-0.04	Al	•	0.28	0.34	-0.19	-0.01
-0.48	-0.04	-0.16	-0.66	-0.56	Fe	-0.94	-0.03	-0.08	-0.28	-0.16
-0.49	0.05	0.16	-0.35	0.78	Mg	-0.96	0.04	0.08	•	0.23
-0.45	0.34	0.68	0.37	-0.27	Ca	-0.88	0.27	0.34	0.16	-0.08
-0.37	-0.87	0	0.34	-0.02	Na	-0.72	-0.68	0	0.15	-0.01

Composantes principales

y₁	-1.78	-0.61	-0.72	-1.42	-1.49	-1.26	-1.18	-1.49	-2.08	-2.31	-2.62	-2.16	2.42
y₂	-2.29	-0.12	-0.37	0.56	0	0.09	0.61	0.5	0	0.38	-1.95	0.22	0.11
	-1.43	-1.41	-1.15	-1.15	2.67	2.14	2.59	2.56	2.62	2.42	2.51	2.57	1.79
	-0.01	0.01	1.63	1.49	-0.03	0.12	-0.26	-0.14	-0.65	-0.05	•	-0.03	0.24

Questions

- 1- Compléter les vides.
- 2- Calculer les inerties cumulées. Quel est le nombre d'axe à retenir?
- 3- Déterminer les individus et variables qui contribuent à 1er et au 2ème axe factoriel. Donner une signification à ces axes.
- 4- Donner les valeurs de $r(Y_1, Al)$, $r(Y_2, Mg)$; $r(Y_1, Mg)$; qu'indiquent ces coefficients de corrélation.
- 6- Représenter les minéraux et les individus (dans la question3) graphiquement sur le premier plan factoriel. et interpréter.
- 7- La poterie 24 a-t-elle contribué à la construction de l'axe1? Est ce qu'elle est bien représentée par cet axe? Elle est caractérisée par quels minéraux?
- 8- Reconstituer la première ligne du tableau à partir de la 1 ère valeur propre.
- 9- Enumérer les étapes d'une ACP avec R.