

ISFA 2° année 2002-2003 Septembre

Les questions sont en grande partie indépendantes.

Merci d'indiquer clairement, sur votre copie, les numéros des questions traitées.

Soit un tableau à 3 lignes (individus) et 3 colonnes (variables) :

$$A = \begin{bmatrix} \sqrt{2}/2 & -1/\sqrt{6} & 0 \\ -\sqrt{2}/2 & 2/\sqrt{6} & -\sqrt{2}/2 \\ 0 & -1/\sqrt{6} & \sqrt{2}/2 \end{bmatrix}$$

1. Quelle est la matrice de corrélation associée à A ?

2. Montrer que pour toute matrice X le rang de X est celui de $X'X$.

3. Montrer que le rang d'une matrice de corrélation calculée entre p variables mesurées sur n individus est inférieur ou égal à $n-1$ si $p > n-1$.

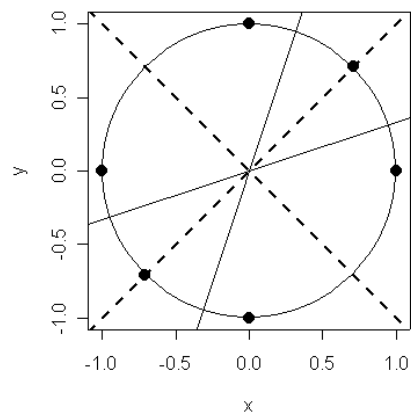
Soit 6 points de \mathbb{R}^2 dont les coordonnées sont :

A	B	C	D	E	F
1	$\sqrt{2}/2$	0	-1	$-\sqrt{2}/2$	0
0	$\sqrt{2}/2$	1	0	$-\sqrt{2}/2$	-1

4. Sur une figure, placer les 6 points, leur centre de gravité, les deux droites de régression et les deux axes principaux. Expliciter les calculs.

Notes pour la solution

```
w=sqrt(2)/2
x=c(1,w,0,-1,-w,0)
y=c(0,w,1,0,-w,-1)
plot(x,y,asp=1,pch=20,cex=2)
symbols(0,0,circle=1,inch=F,add=T)
abline(c(0,1),lty=2,lwd=2)
abline(c(0,-1),lty=2,lwd=2)
abline(c(0,1/3))
abline(c(0,3))
```



Le tableau de données est proposé par Aurélie Despeyroux dans son travail personnel.



Le sujet de cette analyse porte sur la qualité de vie à Lyon. Les données sont issues d'un article paru dans L'Express intitulé «Où vit-on le mieux à Lyon?». Les auteurs de cet article se sont basés sur 80 critères répartis en douze catégories pour fournir un classement général des arrondissements de Lyon. Dans le cadre de cette étude, je me suis intéressée particulièrement à onze critères, qui sont les suivants:

- Rich* Richesse des habitants : revenu moyen imposable (F).
- Dyna* Dynamisme démographique : part des moins de 20 ans dans la population (%)
- Prix* Prix des logements : prix moyen d'achat du m².
- Secu* Sécurité : taux de vols avec violence pour 1000 habitants.
- Trans* Transports en communs : nombre de stations de métro au km².
- Envi* Environnement : part de superficie d'espaces verts (%).
- Cult* Culture : nombre de théâtres pour 10 000 habitants
- Spor* Sport : nombre de salles de sports pour 10 000 habitants.
- Sant* Santé : nombre de médecins généralistes pour 10 000 habitants.
- Comm* Commerce : nombre de restaurants pour 1 000 habitants
- Inve* Investissements : dépenses en éclairage public, espaces verts et équipements publics en F par habitant.

Les données sont tapées dans Excel :

	A	B	C	D	E	F	G	H	I	J	K	L
1		Rich	Dyna	Prix	Secu	Trans	Envi	Cult	Spor	Sant	Comm	Inve
2	Arr1	144545	19.09	16500	9.3	1.99	3	6.33	2.23	16.7	8.23	19059
3	Arr2	180983	19.13	11600	13.8	1.49	0.3	2.14	0.71	26.8	11.15	6838
4	Arr3	146543	20.51	12500	3.5	0.94	1.7	0.24	2.06	13	3.08	4354
5	Arr4	157337	20.25	15000	3.1	0.68	1.8	0.3	1.78	9.8	2.07	3338
6	Arr5	152447	21.12	13700	2.5	0.65	1	1.28	1.49	10.4	3.32	3194
7	Arr6	200977	19.63	20000	4.6	0.53	28	0.21	2.08	22.2	4.34	7661
8	Arr7	126540	19.11	11500	3.4	0.42	0.3	0.16	1.94	12.5	3.47	11225
9	Arr8	119853	21.6	11800	3.3	0.3	0.5	0.14	2.56	10.5	0.98	4585
10	Arr9	121627	25.36	11700	3	0.4	2.3	0.21	2.98	17.6	1.87	5421

puis lues et éditées dans R :

```
> lyon
      Rich Dyna Prix Secu Trans Envi Cult Spor Sant Comm Inve
Arr1 144545 19.09 16500 9.3 1.99 3.0 6.33 2.23 16.7 8.23 19059
Arr2 180983 19.13 11600 13.8 1.49 0.3 2.14 0.71 26.8 11.15 6838
Arr3 146543 20.51 12500 3.5 0.94 1.7 0.24 2.06 13.0 3.08 4354
Arr4 157337 20.25 15000 3.1 0.68 1.8 0.30 1.78 9.8 2.07 3338
Arr5 152447 21.12 13700 2.5 0.65 1.0 1.28 1.49 10.4 3.32 3194
Arr6 200977 19.63 20000 4.6 0.53 28.0 0.21 2.08 22.2 4.34 7661
Arr7 126540 19.11 11500 3.4 0.42 0.3 0.16 1.94 12.5 3.47 11225
```

```

Arr8 119853 21.60 11800 3.3 0.30 0.5 0.14 2.56 10.5 0.98 4585
Arr9 121627 25.36 11700 3.0 0.40 2.3 0.21 2.98 17.6 1.87 5421

```

5. Que faut-il faire pour passer d'une forme à l'autre de la même information ?

```

> round(cor(lyon),dig=2)
      Rich Dyna Prix Secu Trans Envi Cult Spor Sant Comm Inve
Rich  1.00 -0.50 0.66 0.43 0.28 0.69 0.07 -0.60 0.63 0.51 -0.04
Dyna -0.50  1.00 -0.34 -0.47 -0.49 -0.16 -0.37 0.64 -0.17 -0.56 -0.44
Prix  0.66 -0.34  1.00 0.01 0.18 0.84 0.26 0.04 0.21 0.11 0.30
Secu  0.43 -0.47 0.01  1.00 0.80 -0.06 0.63 -0.59 0.77 0.96 0.45
Trans 0.28 -0.49 0.18 0.80  1.00 -0.15 0.90 -0.40 0.42 0.84 0.65
Envi  0.69 -0.16 0.84 -0.06 -0.15  1.00 -0.13 0.11 0.42 0.01 0.07
Cult  0.07 -0.37 0.26 0.63 0.90 -0.13  1.00 -0.14 0.25 0.68 0.80
Spor -0.60 0.64 0.04 -0.59 -0.40 0.11 -0.14  1.00 -0.35 -0.66 0.10
Sant  0.63 -0.17 0.21 0.77 0.42 0.42 0.25 -0.35  1.00 0.75 0.25
Comm  0.51 -0.56 0.11 0.96 0.84 0.01 0.68 -0.66 0.75  1.00 0.52
Inve -0.04 -0.44 0.30 0.45 0.65 0.07 0.80 0.10 0.25 0.52  1.00

```

6. Que signifie cette matrice et comment est-elle calculée ?

7. Quelle est la signification des deux valeurs mises en gras ?

8. Cette matrice est-elle diagonalisable ?

9. Donner sa trace et un majorant de son rang.

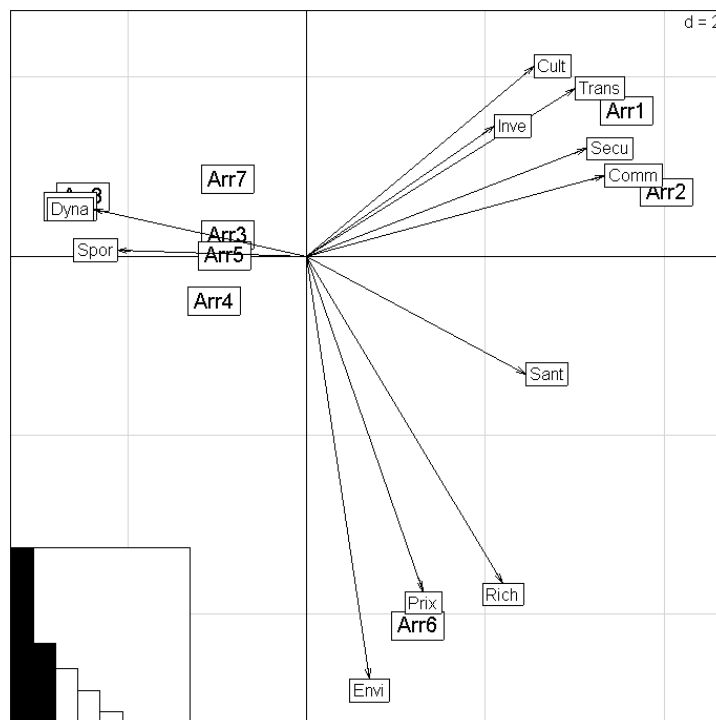
10. Ses valeurs propres sont-elles toutes du même signe ?

Le biplot de l'ACP normée du tableau donne (librairie ade4) :

```

> scatter(dudi.pca(lyon,scan=F),clab.r=1.25,posieig="bottom")

```



11. Donner une légende technique de cette figure.

12. Donner un commentaire de cette figure.

```
> princomp(lyon, cor=T)
Error in princomp.default(lyon, cor = T) :
  princomp can only be used with more units than variables
```

13. Que signifie ce message ?

```
> prcomp(lyon, scal=T)

Standard deviations:
[1] 2.291e+00 1.563e+00 1.304e+00 1.020e+00 6.448e-01 2.965e-01 2.122e-01 1.284e-01 7.800e-16
```

14. Quelle est la plus grande valeur propre de la matrice de la question 3 ?

```
Rotation:
      PC1      PC2      PC3      PC4      PC5      PC6      PC7      PC8      PC9
Rich  0.27438 -0.45643  0.19974 -0.04727  0.23717  3.078e-02  0.171092  0.19262  0.50277
Dyna  -0.29717  0.06507 -0.05233  0.63382  0.46687 -3.729e-01  0.071608  0.16048 -0.15317
Prix  0.16287 -0.46804 -0.38379 -0.16710  0.30970 -6.719e-02 -0.224844  0.45022 -0.11563
Secu  0.39161  0.14985  0.16549  0.24369 -0.06865  3.164e-01 -0.668215  0.19870 -0.20299
Trans 0.37545  0.23415 -0.12891 -0.02647  0.39128  4.799e-01  0.535781 -0.02170 -0.33490
Envi  0.08753 -0.59025 -0.20948  0.12590 -0.16694  9.418e-05  0.003502 -0.58632 -0.34385
Cult  0.31785  0.26483 -0.36463 -0.04002  0.36615 -2.876e-01 -0.302150 -0.44584  0.17910
Spor  -0.26410  0.00793 -0.53618  0.33368 -0.09445  5.360e-01 -0.049912 -0.02510  0.47753
Sant  0.30674 -0.16478  0.15210  0.59958 -0.24579 -1.720e-02  0.163653  0.05261 -0.04297
Comm  0.41712  0.11209  0.13394  0.13697 -0.04745 -2.067e-01  0.109687 -0.16609  0.41387
Inve  0.26275  0.18108 -0.51055 -0.03111 -0.48840 -3.347e-01  0.221044  0.34780 -0.07449
```

15. Quelles sont les propriétés de cette matrice ? Que contient-elle ? à quoi sert-elle ?

16. Une matrice X a n lignes et 2 colonnes. On suppose que les moyennes par colonnes sont nulles (nuage centré). A la ligne i , on trouve les coordonnées (x_i, y_i) du point M_i . On suppose que le vecteur $\mathbf{u}' = (a, b)$ unitaire est le premier axe principal du nuage des n points M_i . Soit le nouveau nuage de n points P_i de coordonnées (z_i, t_i) définies par :

$$\begin{aligned}z_i &= x_i \cos \alpha + y_i \sin \alpha \\t_i &= x_i \sin \alpha - y_i \cos \alpha\end{aligned}$$

Donner le premier axe principal du nuage des n points P_i .
