

Septembre 2017

Aucun document n'est admis. Les calculatrices sont autorisées.  
Tous les calculs doivent être justifiés.

Exercice 1

On considère des informations rassemblées sur les candidats pour un poste dans une entreprise. Les données sont rassemblées dans le tableau ci-dessous:

Candidat	Nombre d'années d'expérience	Nombre d'années d'étude
A	2	3
B	0	3
C	0	5
D	2	4
E	1	5

1. Déterminer le centre de gravité du nuage et la matrice des données du nuage centré.
2. Calculer la matrice d'inertie ainsi que l'inertie totale du nuage (sans calcul supplémentaire).  
On considérera  $Q = I_2$ .
3. Donner la part d'inertie expliquée par chacune des composantes principales.
4. Quel est le lien entre l'inertie du nuage et les valeurs propres de la matrice d'inertie?

Exercice 2

On considère 6 observations  $A, B, C, D, E$  et  $F$  décrites par 2 variables quantitatives appelées  $x$  et  $y$ . Les résultats sont donnés dans le tableau suivant:

	$x$	$y$
A	2	3
B	-2	4
C	1	5
D	3	4
E	0	2
F	3	0

1. Représenter ces observations sur un dessin.
2. Classifier ces observations en deux groupes en utilisant la méthode des centres mobiles à partir des centres initiaux B et F.
3. Rappeler la différence entre les méthodes utilisant le critère du diamètre, du saut minimum et de la moyenne. Quel sont les avantages de la méthode de Ward par rapport à ces dernières?
4. En utilisant la distance de Manhattan, établir le tableau des distances entre ces observations.
5. En utilisant le critère du saut minimum et le tableau des distances construit précédemment, effectuer une classification hiérarchique ascendante de ces 6 observations. Donner l'arbre hiérarchique et suggérer une coupure possible.

6. En utilisant la distance euclidienne usuelle, établir le tableau des distances (au carré) entre ces observations.
7. Effectuer une nouvelle classification en utilisant le critère de Ward. Donner l'arbre hiérarchique et suggérer une coupure possible. Quel pourcentage d'inertie la partition proposée explique-t-elle?

### Exercice 3

Une banque en ligne demande à ses nouveaux clients de répondre aux trois questions suivantes:

- Comment avez-vous connu notre banque?  
(Amis/connaissances, Publicité, Autres)
- À combien s'élèvent vos revenus annuels bruts?  
(Moins de 25 000 euros, Entre 25 000 et 50 000 euros, Plus de 50 000 euros)
- Disposez-vous d'un compte dans une banque classique?  
(Oui, Non, Plus maintenant)

Une seule réponse est acceptée par question. Voici les résultats de ce questionnaire sur 5 personnes:

A	Amis/connaissances	Plus de 50 000 euros	Oui
B	Publicité	Entre 25 000 et 50 000 euros	Plus maintenant
C	Publicité	Moins de 25 000 euros	Non
D	Amis/connaissances	Entre 25 000 et 50 000 euros	Plus maintenant
E	Autres	Moins de 25 000 euros	Oui

1. En modélisant ces résultats par trois variables qualitatives numériques (représentant les résultats aux trois questions), donner ces résultats sous la forme d'un tableau de données chiffrées à 5 lignes et 3 colonnes.
2. Donner le tableau disjonctif complet correspondant.
3. Écrire le tableau de Burt associé à ces données.
4. Expliquer soigneusement comment déterminer à partir du tableau de Burt:
  - le nombre de personnes qui disposent de revenus annuels bruts supérieurs à 50 000 euros ;
  - le nombre de personnes qui ont connu la banque via la publicité et qui disposent toujours d'un compte dans une banque classique;
  - le nombre de personnes qui n'ont plus de compte dans une banque classique.

Vérifier ces valeurs sur les données initiales.

### Exercice 4

On considère les classements, effectués par 10 magazines sportifs, de 20 équipes de Basket. Justifier et commenter l'analyse produite sur ces données.

## 1. Les données et le programme

Preseason 1985 College Basketball Rankings(rankings of 20 teams by 10 news services)

Note:

- (a) news services rank varying numbers of teams;
- (b) not all teams are ranked by all news services;
- (c) each team is ranked by at least one service;
- (d) rank 20 is missing for UPI.

```
Option nodate nonumber nocenter ;
data Basketball;
input School $13. CSN DurSun DurHer WashPost USAToday
Sport InSports UPI AP SI;
label CSN = 'Community Sports News '
DurSun = 'Durham Sun'
DurHer = 'Durham Morning Herald'
WashPost = 'Washington Post'
USAToday = 'USA Today'
Sport = 'Sport Magazine'
InSports = 'Inside Sports'
UPI = 'United Press International'
AP = 'Associated Press'
SI = 'Sports Illustrated'
;
format CSN--SI 5.1;
datalines;

Louisville      1 8 1 9 8 9 6 10 9 9
Georgia Tech    2 2 4 3 1 1 1 2 1 1
Kansas          3 4 5 1 5 11 8 4 5 7
Michigan        4 5 9 4 2 5 3 1 3 2
Duke            5 6 7 5 4 10 4 5 6 5
UNC             6 1 2 2 3 4 2 3 2 3
Syracuse        7 10 6 11 6 6 5 6 4 10
Notre Dame      8 14 15 13 11 20 18 13 12 11
Kentucky        9 15 16 14 14 19 11 12 11 13
LSU             10 9 13 13 13 15 16 9 14 8
DePaul          11 10 21 15 20 15 19 12 16 19
Georgetown      12 7 8 6 9 2 9 8 8 4
Navy            13 20 19 10 18 13 15 17 20 14
Illinois        14 3 3 7 7 3 10 7 7 6
Iowa            15 16 15 18 19 15 16 14 15 20
Arkansas        16 17 16 20 20 11 16 12 18 16
Memphis         17 12 11 16 16 8 20 11 15 12
Washington      18 16 12 18 19 17 20 17 15 18
UAB             19 13 10 11 12 17 15 16 16 15
UNLV            20 18 18 19 20 17 14 18 18 19
;
run;
proc princomp data=Basketball n=2 out=PCBasketball standard;
var CSN--SI;
run;
proc sort data=PCBasketball; by prin1;run;
proc print data=PCBasketball;run;
```

## 2. Les résultats

### The PRINCOMP Procedure

Observations 20  
Variables 10

#### Simple Statistics

	CSN	DurSun	DurHer	WashPost	USAToday
Mean	10.50000000	10.30000000	10.55000000	10.75000000	11.35000000
StD	5.91607978	5.72253164	5.98660786	5.96370602	6.65127609

#### Simple Statistics

	Sport	InSports	UPI	AP	SI
Mean	10.90000000	11.40000000	9.85000000	10.75000000	10.60000000
StD	5.95509512	6.31122642	5.23424957	5.93717104	6.04718290

#### Correlation Matrix

		CSN	DurSun	DurHer	WashPost
CSN	Community Sports News	1.0000	0.6700	0.5610	0.7317
DurSun	Durham Sun	0.6700	1.0000	0.8046	0.8397
DurHer	Durham Morning Herald	0.5610	0.8046	1.0000	0.7293
WashPost	Washington Post	0.7317	0.8397	0.7293	1.0000
USAToday	USA Today	0.7764	0.8655	0.8356	0.9059
Sport	Sport Magazine	0.4153	0.7361	0.7117	0.6498
InSports	Inside Sports	0.7513	0.7339	0.7405	0.8096
UPI	United Press International	0.7725	0.9012	0.7065	0.8030
AP	Associated Press	0.7754	0.8776	0.8111	0.8246
SI	Sports Illustrated	0.7120	0.8508	0.7537	0.8800

#### Correlation Matrix

	USAToday	Sport	In Sports	UPI	AP	SI
CSN	0.7764	0.4153	0.7513	0.7725	0.7754	0.7120
DurSun	0.8655	0.7361	0.7339	0.9012	0.8776	0.8508
DurHer	0.8356	0.7117	0.7405	0.7065	0.8111	0.7537
WashPost	0.9059	0.6498	0.8096	0.8030	0.8246	0.8800
USAToday	1.0000	0.6720	0.8829	0.8845	0.9446	0.9236
Sport	0.6720	1.0000	0.6957	0.7593	0.7034	0.7544
InSports	0.8829	0.6957	1.0000	0.8129	0.8821	0.7987
UPI	0.8845	0.7593	0.8129	1.0000	0.9184	0.8826
AP	0.9446	0.7034	0.8821	0.9184	1.0000	0.8576
SI	0.9236	0.7544	0.7987	0.8826	0.8576	1.0000

#### Eigenvalues of the Correlation Matrix

	Eigenvalue	Difference	Proportion	Cumulative
1	8.09876884	7.44509887	0.8099	0.8099
2	0.65366997		0.0654	0.8752

Eigenvectors

		Prin1	Prin2
CSN	Community Sports News	0.280076	-.653165
DurSun	Durham Sun	0.324304	0.136337
DurHer	Durham Morning Herald	0.299065	0.303827
WashPost	Washington Post	0.320282	-.115147
USAToday	USA Today	0.340732	-.109475
Sport	Sport Magazine	0.276402	0.645043
InSports	Inside Sports	0.317144	-.114554
UPI	United Press International	0.330525	-.026392
AP	Associated Press	0.336771	-.074533
SI	Sports Illustrated	0.329586	0.042176

		W	U	I								
		a	S	n								
S	D	D	s	A	S					P	P	
c	u	u	h	T	S	p				r	r	
h	r	r	P	o	p	o				i	i	
o	C	S	H	o	d	o	r	U				
b	S	u	e	s	a	r	t	P	A	S	n	n
s	N	n	r	t	y	t	s	I	P	I	1	2
1 Georgia Tech	2.0	2.0	4.0	3.0	1.0	1.0	1.0	2.0	1.0	1.0	-1.65171	-0.07457
2 UNC	6.0	1.0	2.0	2.0	3.0	4.0	2.0	3.0	2.0	3.0	-1.47605	-0.41767
3 Michigan	4.0	5.0	9.0	4.0	2.0	5.0	3.0	1.0	3.0	2.0	-1.29667	0.48515
4 Kansas	3.0	4.0	5.0	1.0	5.0	11.0	8.0	4.0	5.0	7.0	-1.01772	1.03669
5 Duke	5.0	6.0	7.0	5.0	4.0	10.0	4.0	5.0	6.0	5.0	-0.93515	0.78970
6 Illinois	14.0	3.0	3.0	7.0	7.0	3.0	10.0	7.0	7.0	6.0	-0.74840	-1.97952
7 Syracuse	7.0	10.0	6.0	11.0	6.0	6.0	5.0	6.0	4.0	10.0	-0.66004	-0.10272
8 Louisville	1.0	8.0	1.0	9.0	8.0	9.0	6.0	10.0	9.0	9.0	-0.65334	0.61913
9 Georgetown	12.0	7.0	8.0	6.0	9.0	2.0	9.0	8.0	8.0	4.0	-0.62726	-1.44206
10 LSU	10.0	9.0	13.0	13.0	13.0	15.0	16.0	9.0	14.0	8.0	0.22518	0.47487
11 Kentucky	9.0	15.0	16.0	14.0	14.0	19.0	11.0	12.0	11.0	13.0	0.49704	1.65146
12 Notre Dame	8.0	14.0	15.0	13.0	11.0	20.0	18.0	13.0	12.0	11.0	0.51380	1.71854
13 Memphis	17.0	12.0	11.0	16.0	16.0	8.0	20.0	11.0	15.0	12.0	0.57426	-1.67199
14 UAB	19.0	13.0	10.0	11.0	12.0	17.0	15.0	16.0	16.0	15.0	0.69036	-0.48038
15 DePaul	11.0	10.0	21.0	15.0	20.0	15.0	19.0	12.0	16.0	19.0	0.93598	0.65747
16 Navy	13.0	20.0	19.0	10.0	18.0	13.0	15.0	17.0	20.0	14.0	0.99459	0.39904
17 Arkansas	16.0	17.0	16.0	20.0	20.0	11.0	16.0	12.0	18.0	16.0	1.02934	-0.77778
18 Iowa	15.0	16.0	15.0	18.0	19.0	15.0	16.0	14.0	15.0	20.0	1.04592	-0.06083
19 Washington	18.0	16.0	12.0	18.0	19.0	17.0	20.0	17.0	15.0	18.0	1.17468	-0.51663
20 UNLV	20.0	18.0	18.0	19.0	20.0	17.0	14.0	18.0	18.0	19.0	1.38517	-0.30790