

Mai 2018

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

### Exercice 1

Considérons les variables suivantes :  $A(1, 0, 4)$ ,  $B(3, 1, 0)$ ,  $C(2, 1, 2)$  et  $D(1, 0, 1)$ .

1. Peuvent-elles être linéairement indépendantes ? Justifier votre réponse.
2. Expliquer linéairement la variable  $A$  par les variables  $B$ ,  $C$  et  $D$ .
3. La variable  $A$  appartient-elle au plan engendré par  $B$  et  $C$  ?
4. En considérant le produit scalaire euclidien usuel, calculer la projection  $\hat{A}$  de la variable  $A$  sur le plan engendré par  $B$  et  $C$ . Calculer  $\langle A - \hat{A}, B \rangle$ . Le résultat était-il prévisible ?
5. Calculer le cosinus de l'angle  $(A, \hat{A})$  et estimer la qualité de la représentation de  $A$  par  $\hat{A}$ .

### Exercice 2

On considère les réponses de 6 individus à un sondage dont les deux questions sont « Combien d'années d'étude avez-vous effectué après le bac ? » et « Quelle est votre ancienneté dans l'entreprise ? ». Les résultats sont regroupés dans le tableau suivant :

	I <sub>1</sub>	I <sub>2</sub>	I <sub>3</sub>	I <sub>4</sub>	I <sub>5</sub>	I <sub>6</sub>
Années d'étude	5	3	4	2	3	2
Ancienneté	1	3	0	1	2	4

1. Représenter les individus dans le plan  $(O, x, y)$ , en prenant une échelle de 1 cm pour 1 an d'étude sur l'axe des abscisses et 1 cm pour 1 an d'ancienneté sur l'axe des ordonnées.
2. Tous les individus sont munis du même poids. Calculer l'inertie du nuage par rapport à l'origine  $O$ .
3. Calculer les inerties du nuage projeté sur l'axe des abscisses puis sur l'axe des ordonnées.
4. Laquelle des inerties calculées représente l'inertie par rapport à l'axe  $(Ox)$  ?
5. Quel lien unit ces trois inerties ? En vertu de quel résultat ?

### Exercice 3

On considère le nombre d'appareils connectés possédés par des familles. Ces données sont rassemblées dans le tableau suivant. Pour simplifier, on suppose que chaque famille représente le même poids.

Nombre d'appareils connectés	Domicile	Travail
A	5	3
B	2	2
C	3	1
D	4	3
E	1	1
F	3	2

1. Déterminer le centre de gravité du nuage et la matrice des données du nuage centré.

- Calculer la matrice d'inertie ainsi que l'inertie totale du nuage (sans calcul supplémentaire).  
On considérera  $Q = I_2$ .
- Donner la part d'inertie expliquée par chacune des composantes principales.

#### Exercice 4

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	1	0
B	2	1
C	3	3
D	2	3
E	-1	1
F	4	2

- Représenter ces observations sur un dessin.
- Classifier ces observations en deux groupes en utilisant la méthode des centres mobiles à partir des centres initiaux B et F.
- En utilisant la distance euclidienne usuelle, établir le tableau des distances (au carré) entre ces observations.
- 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 ?
- En utilisant le critère du diamètre, effectuer une classification hiérarchique ascendante de ces 6 observations. Donner l'arbre hiérarchique et suggérer une coupure possible.
- 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 5

Un institut de sondage a recueilli les données suivantes pour étudier la relation entre la catégorie socio-professionnelle (CSP) d'une personne (agriculteur : AGRI, cadre supérieur : CSUP, cadre moyen : CMOY, employé : EMPL, ouvrier : OUVR, retraité : RETR et chômeur : CHOM) et sa principale source d'information concernant les problèmes environnementaux (télévision : TEL, journaux : JOU, radio : RAD, livres : LIV, associations : ASS et mairie : MAI). L'objectif de cet exercice est de justifier et commenter l'analyse ci-jointe produite sur ces données.

CSP	TEL	JOU	RAD	LIV	ASS	MAI	Total
AGRI	26	18	9	5	4	6	68
CSUP	19	49	4	16	5	3	96
CMOY	44	87	4	39	14	3	191
EMPL	83	87	13	24	5	1	213
OUVR	181	107	16	31	7	7	349
RETR	167	95	29	15	7	7	320
CHOM	27	9	4	2	2	2	46
Total	547	452	79	132	44	29	1283

- Combien d'axes principaux proposez-vous de choisir ? Pour quelles raisons ?
- En notant  $n$  le nombre d'individus et  $p$  et le nombre de variables, on a vu que l'analyse par composantes principales déterminait tout d'abord les axes principaux (dirigés par les vecteurs propres notés  $U_i$ , dans  $\mathbb{R}^p$ ), puis les facteurs principaux (notés  $C_i$ , dans  $\mathbb{R}^p$ ), puis les composantes principales (notés  $D_i$ , dans  $\mathbb{R}^n$ ).
  - Que représentent concrètement ces différents vecteurs ?

- (b) L'analyse ci-jointe effectuée par le logiciel SAS procure 4 tableaux : la matrice de corrélation (*correlation matrix*), les valeurs propres (*eigenvalues*), les vecteurs propres (*eigenvectors*) et enfin un dernier. Deux d'entre eux correspondent aux familles de vecteurs rappelés ci-dessus, lesquels ?
- (c) Expliquer comment obtenir la troisième famille de vecteurs en utilisant les données jointes et calculer explicitement les deux premiers vecteurs de cette famille.
- (d) Comment calculer la qualité de la représentation des individus sur les axes principaux ?  
La calculer pour le premier individu.
- (e) Comment calculer la qualité de la représentation des variables sur les axes principaux ?  
La calculer pour la première variable.
- (f) Conclure.

## The SAS System

### The PRINCOMP Procedure

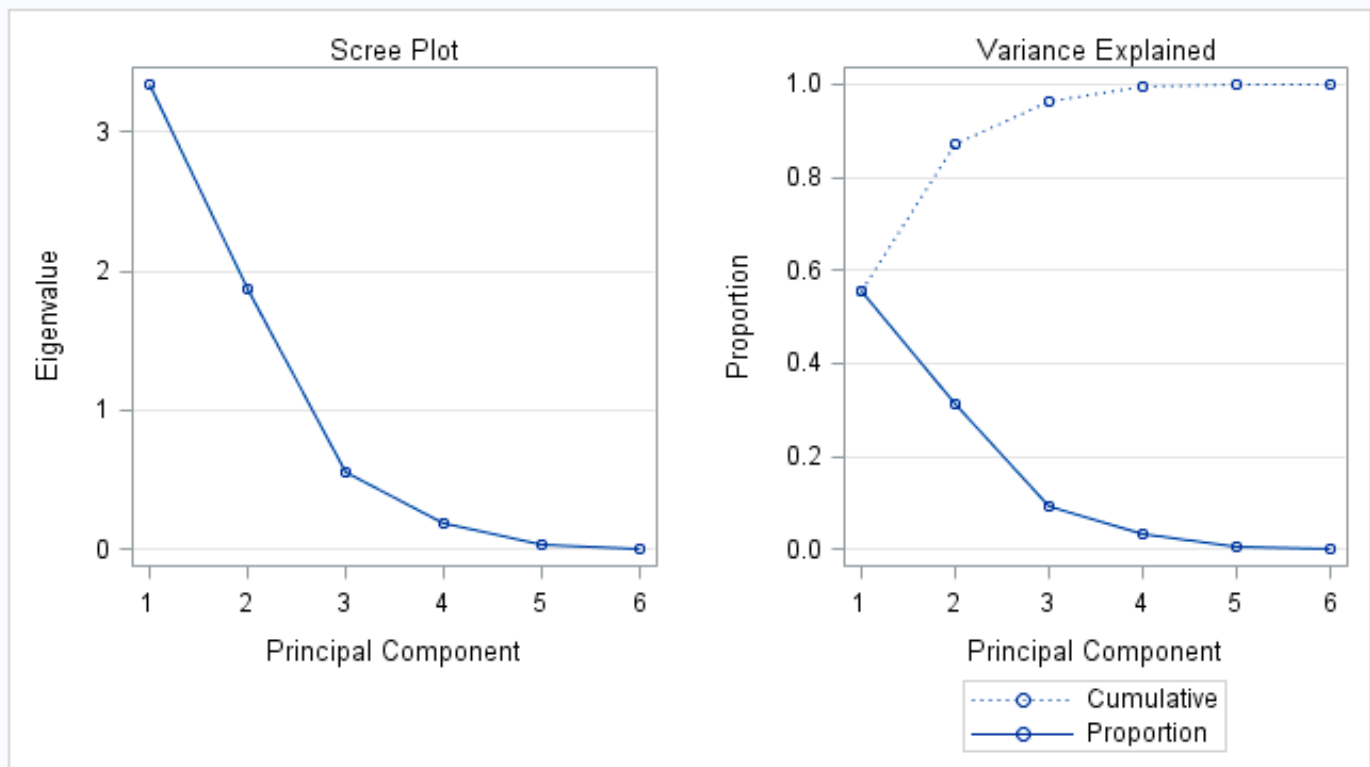
<b>Observations</b>	7
<b>Variables</b>	6

Simple Statistics						
	TEL	JOU	RAD	LIV	ASS	MAI
<b>Mean</b>	78.14285714	64.57142857	11.28571429	18.85714286	6.285714286	4.142857143
<b>StD</b>	68.91644354	39.22523483	9.15995425	13.40930736	3.817254062	2.478478796

Correlation Matrix						
	TEL	JOU	RAD	LIV	ASS	MAI
<b>TEL</b>	1.0000	0.7770	0.8551	0.3602	0.1785	0.6361
<b>JOU</b>	0.7770	1.0000	0.5747	0.8212	0.6243	0.2785
<b>RAD</b>	0.8551	0.5747	1.0000	0.0357	0.0163	0.6366
<b>LIV</b>	0.3602	0.8212	0.0357	1.0000	0.8540	-.0093
<b>ASS</b>	0.1785	0.6243	0.0163	0.8540	1.0000	0.1007
<b>MAI</b>	0.6361	0.2785	0.6366	-.0093	0.1007	1.0000

Eigenvalues of the Correlation Matrix				
	Eigenvalue	Difference	Proportion	Cumulative
<b>1</b>	3.34296696	1.46627345	0.5572	0.5572
<b>2</b>	1.87669351	1.32525447	0.3128	0.8699
<b>3</b>	0.55143904	0.36283621	0.0919	0.9618
<b>4</b>	0.18860283	0.15160794	0.0314	0.9933
<b>5</b>	0.03699489	0.03369212	0.0062	0.9994
<b>6</b>	0.00330277		0.0006	1.0000

Eigenvectors						
	Prin1	Prin2	Prin3	Prin4	Prin5	Prin6
<b>TEL</b>	0.482196	-.281956	-.240531	-.348144	-.711974	-.044962
<b>JOU</b>	0.513715	0.163761	-.321066	-.079799	0.469458	-.615985
<b>RAD</b>	0.394737	-.444602	-.229573	0.640353	0.183380	0.387463
<b>LIV</b>	0.378392	0.513874	-.035476	-.343804	0.191132	0.660880
<b>ASS</b>	0.327685	0.502925	0.446734	0.520516	-.371542	-.176457
<b>MAI</b>	0.311784	-.423098	0.765198	-.270624	0.253967	-.022688



### The SAS System

Obs	CSP	TEL	JOU	RAD	LIV	ASS	MAI	Prin1	Prin2	Prin3	Prin4	Prin5	Prin6
1	CHOM	27	9	4	2	2	2	-1.37430	-0.37519	-0.40747	-0.12925	-1.69132	-0.26473
2	AGRI	26	18	9	5	4	6	-0.78041	-0.74411	1.29682	0.09025	0.78406	1.29441
3	CSUP	19	49	4	16	5	3	-0.69268	0.32615	0.02791	-0.35957	1.27946	-1.67021
4	EMPL	83	87	13	24	5	1	0.02242	0.40496	-1.85552	0.19726	0.67054	1.02317
5	CMOY	44	87	4	39	14	3	0.45271	1.87626	0.82799	0.63961	-0.54874	0.14761
6	RETR	167	95	29	15	7	7	1.14609	-1.19551	-0.03670	1.40901	-0.16066	-0.62260
7	OUVR	181	107	16	31	7	7	1.22616	-0.29256	0.14698	-1.84731	-0.33334	0.09234