

Mai 2019

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

Exercice 1

On considère le nombre d'ouvertures de magasins pour une franchise pendant deux années consécutives, 2017 et 2018. Ces données sont rassemblées dans le tableau suivant. Pour simplifier, on suppose que chaque département représente le même poids.

Département	2017	2018
A Ain	0	2
B Calvados	1	3
C Charente	3	2
D Haute-Savoie	2	1
E Lozère	1	0
F Var	3	4

1. Représenter les magasins dans le plan (O, x, y) , en prenant une échelle de 1 cm pour 1 magasin ouvert sur les deux axes.
2. Déterminer le centre de gravité du nuage et la matrice des données du nuage centré.
3. Calculer la matrice d'inertie.
4. Calculer l'inertie totale du nuage par rapport à l'origine O de deux façons différentes :
 - (a) en exploitant les questions précédentes sans aucun calcul supplémentaire ;
 - (b) par un calcul direct.
5. Calculer les inerties :
 - (a) du nuage projeté sur l'axe des abscisses ;
 - (b) du nuage projeté sur l'axe des ordonnées ;
 - (c) du nuage par rapport à l'axe (Ox) .
6. Lesquelles de ces inerties sont égales et pourquoi ?
7. Montrer le lien unissant ces différentes inerties et le justifier.
8. On considérera $Q = I_2$. Calculer les deux composantes principales et donner la part d'inertie expliquée par chacune d'elles. Les faire figurer sur le dessin.
9. Classifier ces observations en deux groupes en utilisant la méthode des centres mobiles à partir des centres initiaux E et F.
10. En utilisant la distance euclidienne usuelle, établir le tableau des distances (au carré) entre ces observations.
11. Rappeler la différence entre les méthodes de classification hiérarchique ascendante 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 ?
12. En utilisant le critère du diamètre, effectuer une classification hiérarchique ascendante de ces 6 observations. Donner l'arbre hiérarchique et suggérer une coupure possible.
13. Effectuer une nouvelle classification avec le critère de Ward. Donner l'arbre hiérarchique et suggérer une coupure possible. Quel pourcentage d'inertie la partition proposée explique-t-elle ?

Exercice 2

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

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

Exercice 3

On s'intéresse à un ensemble de données représentant les relevés de températures mensuelles ($p = 12$) dans 15 villes de France ($n = 15$). Ces données sont regroupées dans le dernier tableau de l'annexe jointe à cet énoncé.

L'objectif de cet exercice est de justifier et commenter l'analyse par composantes principales produite sur ces données.

1. Combien d'axes principaux proposez-vous de choisir ? Pour quelles raisons ?
2. 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 composantes principales (notés C_i , dans \mathbb{R}^n), puis les facteurs principaux (notés D_i , dans \mathbb{R}^p).
 - (a) 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 sur le premier axe.
 - (e) Comment calculer la qualité de la représentation des variables sur les axes principaux ? La calculer pour les deux premiers axes principaux.
 - (f) Conclure.

The SAS System

The PRINCOMP Procedure

Observations	15
Variables	12

Simple Statistics

	JANV	FEVR	MARS	AVRI	MAI	JUIN	JUIL	AOUT	SEPT	OCTO	NOVE	DECE
Mean	3.973333333	4.833333333	8.233333333	10.98000000	14.43333333	17.83333333	19.83333333	19.56666667	16.98666667	12.32000000	7.926666667	4.846666667
STD	2.007296215	1.868409051	1.529083136	1.41481952	1.50459613	1.79350681	2.12893893	2.01198788	1.84811358	1.82998829	1.803752702	1.958449338

Correlation Matrix

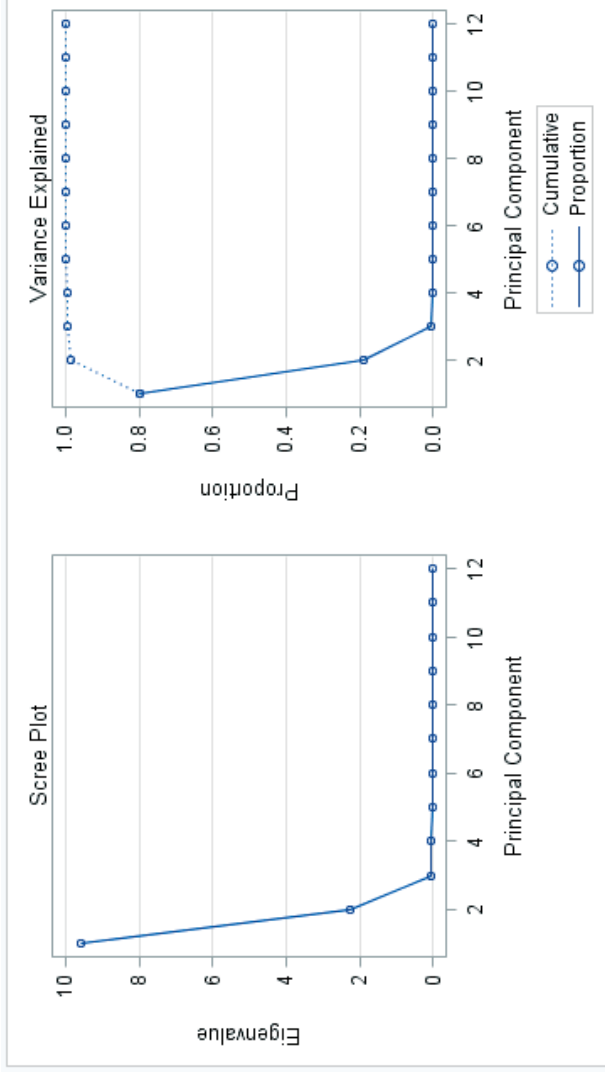
	JANV	FEVR	MARS	AVRI	MAI	JUIN	JUIL	AOUT	SEPT	OCTO	NOVE	DECE
JANV	1.0000	0.9735	0.8353	0.6115	0.3626	0.3389	0.2969	0.4051	0.6037	0.8505	0.9509	0.9939
FEVR	0.9735	1.0000	0.9311	0.7612	0.5466	0.5244	0.4903	0.5880	0.7597	0.9403	0.9861	0.9701
MARS	0.8353	0.9311	1.0000	0.9195	0.7667	0.7567	0.7215	0.7981	0.9106	0.9685	0.9273	0.8336
AVRI	0.6115	0.7612	0.9195	1.0000	0.9526	0.9422	0.9123	0.9490	0.9784	0.9053	0.7820	0.6193
MAI	0.3626	0.5466	0.7667	0.9526	1.0000	0.9938	0.9812	0.9803	0.9421	0.7702	0.5910	0.3807
JUIN	0.3389	0.5244	0.7567	0.9422	0.9938	1.0000	0.9915	0.9887	0.9404	0.7567	0.5725	0.3613
JUIL	0.2969	0.4903	0.7215	0.9123	0.9812	0.9915	1.0000	0.9906	0.9329	0.7387	0.5472	0.3239
AOUT	0.4051	0.5880	0.7981	0.9490	0.9803	0.9887	0.9906	1.0000	0.9701	0.8121	0.6366	0.4302
SEPT	0.6037	0.7597	0.9106	0.9784	0.9421	0.9404	0.9329	0.9701	1.0000	0.9256	0.7961	0.6224
OCTO	0.8505	0.9403	0.9685	0.9053	0.7702	0.7567	0.7387	0.8121	0.9256	1.0000	0.9643	0.8657
NOVE	0.9509	0.9861	0.9273	0.7820	0.5910	0.5725	0.5472	0.6366	0.7961	0.9643	1.0000	0.9609
DECE	0.9939	0.9701	0.8336	0.6193	0.3807	0.3613	0.3239	0.4302	0.6224	0.8657	0.9609	1.0000

Eigenvalues of the Correlation Matrix

	Eigenvalue	Difference	Proportion	Cumulative
1	9.58177958	7.30536118	0.7985	0.7985
2	2.27641840	2.20640399	0.1897	0.9882
3	0.07001440	0.03033967	0.0058	0.9940
4	0.03967473	0.02562944	0.0033	0.9973
5	0.01404529	0.00606375	0.0012	0.9985
6	0.00798154	0.00193228	0.0007	0.9992
7	0.00604926	0.00430236	0.0005	0.9997
8	0.00174689	0.00025471	0.0001	0.9998
9	0.00149218	0.00100004	0.0001	0.9999

Eigenvectors

	Prin1	Prin2	Prin3	Prin4	Prin5	Prin6	Prin7	Prin8	Prin9	Prin10	Prin11	Prin12
JANV	0.245922	0.427059	-.080941	0.255970	0.265538	0.284783	0.093813	-.168578	0.551036	0.429087	0.091650	-.009160
FEVR	0.284437	0.310899	-.126829	-.058499	-.045030	0.535807	0.383891	0.141322	-.356465	-.307576	-.216898	0.278843
MARS	0.312966	0.103400	-.583333	-.548320	0.104990	-.369850	0.079001	0.229694	0.022120	0.124097	0.165477	-.018552
AVRI	0.313149	-.134992	-.460103	0.271469	-.126236	0.038977	-.194587	-.582654	-.248592	-.095774	-.095764	-.351253
MAI	0.281951	-.314631	-.145718	0.464837	-.406869	0.085619	-.015533	0.502604	0.037961	0.117558	0.366536	0.094177
JUIN	0.278982	-.330960	-.046941	0.245318	0.298566	-.333248	0.070354	0.005261	0.246339	-.160201	-.511605	0.443835
JUIL	0.271862	-.352218	0.310353	-.059148	0.220178	0.016914	0.540370	0.074100	-.148783	0.250109	-.084066	-.514008
AOUT	0.290299	-.284958	0.232574	-.201691	0.328882	0.173554	-.181967	-.306133	-.188164	0.019902	0.526853	0.401580
SEPT	0.314665	-.137927	0.155875	-.321604	-.067321	0.314792	-.376496	0.167378	0.454942	-.405646	-.122455	-.307130
OCTO	0.316646	0.112976	0.280357	-.201207	-.404743	-.052382	-.302835	-.032703	-.188967	0.545738	-.383445	0.167271
NOVE	0.291962	0.274345	0.321566	0.015053	-.407183	-.416414	0.333424	-.284457	0.180913	-.330041	0.249433	0.026495
DECE	0.250153	0.413775	0.212706	0.303537	0.393740	-.263574	-.353255	0.312276	-.334556	-.151554	0.055052	-.210001



VILLES	JANV	FEBV	MARS	AVRI	MAI	JUIN	JUIL	AOUT	SEPT	OCTO	NOVE	DECE	Prin1	Prin2	Prin3	Prin4	Prin5	Prin6	Prin7	Prin8	Prin9	Prin10	Prin11	Prin12
Bordeaux	5.6	6.6	10.3	12.8	15.8	19.3	20.9	21.0	18.6	13.8	9.1	6.2	5.21	0.27	1.15	-0.08	0.00	0.07	-0.22	0.03	-0.04	0.03	0.00	0.01
Brest	6.1	5.8	7.8	9.2	11.6	14.4	15.6	16.0	14.7	12.0	9.0	7.0	-3.70	7.52	-0.23	-0.08	0.19	0.25	0.05	-0.03	-0.03	0.05	0.01	0.00
Clermont	2.6	3.7	7.5	10.3	13.8	17.3	19.4	19.1	16.2	11.2	6.6	3.6	-3.01	-1.12	0.01	-0.22	-0.29	-0.03	0.01	0.11	-0.05	0.00	-0.01	0.01
Grenoble	1.5	3.2	7.7	10.6	14.5	17.8	20.1	19.5	16.7	11.4	6.5	2.3	-2.85	-3.14	0.18	-0.73	0.45	0.04	0.11	0.02	0.01	-0.02	-0.02	0.00
Lille	2.4	2.9	6.0	8.9	12.4	15.3	17.1	17.1	14.7	10.4	6.1	3.5	-7.21	1.13	-0.58	-0.04	0.03	-0.09	-0.33	0.04	-0.03	-0.06	0.01	0.00
Lyon	2.1	3.3	7.7	10.9	14.9	18.5	20.7	20.1	16.9	11.4	6.7	3.1	-1.57	-3.28	0.03	0.01	-0.08	0.15	0.09	-0.03	0.04	0.00	0.09	0.01
Marseille	5.5	6.6	10.0	13.0	16.8	20.8	23.3	22.8	19.9	15.0	10.2	6.9	8.46	-1.61	-0.56	0.31	-0.02	0.27	-0.06	0.06	0.07	0.02	-0.04	0.00
Montpellier	5.6	6.7	9.9	12.8	16.2	20.1	22.7	22.3	19.3	14.6	10.0	6.5	7.26	-0.91	-0.30	0.07	-0.06	-0.01	0.17	-0.03	-0.19	-0.02	0.00	0.00
Nantes	5.0	5.3	8.4	10.8	13.9	17.2	18.8	18.6	16.4	12.2	8.2	5.5	-0.48	2.10	0.38	0.21	-0.05	-0.01	0.11	-0.14	0.05	-0.08	-0.03	0.01
Nice	7.5	8.5	10.8	13.3	16.7	20.1	22.7	22.5	20.3	16.0	11.5	8.2	10.55	1.36	-0.29	0.02	0.29	-0.32	-0.03	0.01	0.05	0.00	0.03	0.00
Paris	3.4	4.1	7.6	10.7	14.3	17.5	19.1	18.7	16.0	11.4	7.1	4.3	-2.26	-0.16	0.37	0.54	0.05	0.07	0.03	0.09	0.02	-0.05	0.01	-0.02
Rennes	4.8	5.3	7.9	10.1	13.1	16.2	17.9	17.8	15.7	11.6	7.8	5.4	-2.41	3.10	0.26	0.05	-0.20	-0.26	0.16	0.07	0.03	0.04	-0.01	0.00
Strasbourg	0.4	1.5	5.6	9.8	14.0	17.2	19.0	18.3	15.1	9.5	4.9	1.3	-7.30	-3.82	-0.01	0.55	0.20	-0.12	-0.06	-0.09	-0.04	0.07	-0.02	0.00
Toulouse	4.7	5.6	9.2	11.6	14.9	18.7	20.9	20.9	18.3	13.3	8.6	5.5	3.07	-0.35	-0.08	-0.43	-0.31	-0.01	-0.17	-0.13	0.03	0.02	0.00	-0.02
Vichy	2.4	3.4	7.1	9.9	13.6	17.1	19.3	18.8	16.0	11.0	6.6	3.4	-3.77	-1.10	-0.34	-0.17	-0.20	-0.02	0.14	0.01	0.06	0.02	-0.02	0.01