

Septembre 2019

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

### Exercice 1

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

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 E 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 diamètre 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 2

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

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 3

On considère le taux de rejet en pourcentages des différentes chaînes d'une usine de production de composants électroniques sur deux années consécutives. Les données sont rassemblées dans le tableau ci-dessous. Pour simplifier, on suppose que chaque chaîne représente le même poids :

Chaîne	Taux 2017	Taux 2018
A	4	3
B	1	2
C	5	7
D	2	1
E	3	2

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 4

Les données ci-jointes sont relatives à certaines catégories de crimes aux États-Unis.

1. Justifier et commenter l'analyse ci-jointe produite sur ces données. Vous préciserez en particulier le nombre d'axes principaux choisis et les raisons qui vont ont conduit à faire ce choix.
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  $U_i$ ), puis les facteurs principaux (notés  $C_i$ ), puis les composantes principales (notées  $D_i$ ).
  - (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 (sans titre, après les graphiques). 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.

### The PRINCOMP Procedure

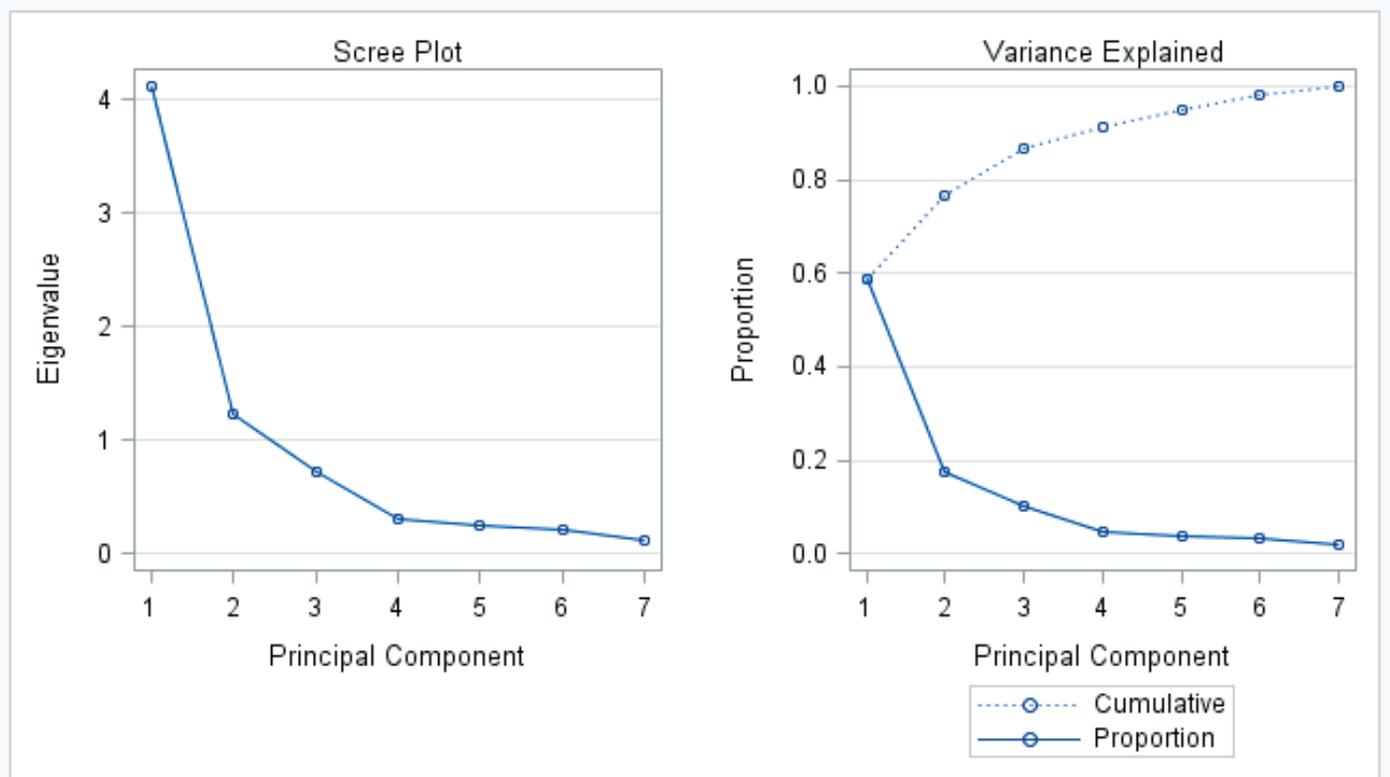
<b>Observations</b>	50
<b>Variables</b>	7

Simple Statistics							
	<b>Murder</b>	<b>Rape</b>	<b>Robbery</b>	<b>Assault</b>	<b>Burglary</b>	<b>Larceny</b>	<b>Auto_Theft</b>
<b>Mean</b>	7.444000000	25.73400000	124.0920000	211.3000000	1291.904000	2671.288000	377.5260000
<b>StD</b>	3.866768941	10.75962995	88.3485672	100.2530492	432.455711	725.908707	193.3944175

Correlation Matrix							
	<b>Murder</b>	<b>Rape</b>	<b>Robbery</b>	<b>Assault</b>	<b>Burglary</b>	<b>Larceny</b>	<b>Auto_Theft</b>
<b>Murder</b>	1.0000	0.6012	0.4837	0.6486	0.3858	0.1019	0.0688
<b>Rape</b>	0.6012	1.0000	0.5919	0.7403	0.7121	0.6140	0.3489
<b>Robbery</b>	0.4837	0.5919	1.0000	0.5571	0.6372	0.4467	0.5907
<b>Assault</b>	0.6486	0.7403	0.5571	1.0000	0.6229	0.4044	0.2758
<b>Burglary</b>	0.3858	0.7121	0.6372	0.6229	1.0000	0.7921	0.5580
<b>Larceny</b>	0.1019	0.6140	0.4467	0.4044	0.7921	1.0000	0.4442
<b>Auto_Theft</b>	0.0688	0.3489	0.5907	0.2758	0.5580	0.4442	1.0000

Eigenvalues of the Correlation Matrix				
	<b>Eigenvalue</b>	<b>Difference</b>	<b>Proportion</b>	<b>Cumulative</b>
<b>1</b>	4.11495951	2.87623768	0.5879	0.5879
<b>2</b>	1.23872183	0.51290521	0.1770	0.7648
<b>3</b>	0.72581663	0.40938458	0.1037	0.8685
<b>4</b>	0.31643205	0.05845759	0.0452	0.9137
<b>5</b>	0.25797446	0.03593499	0.0369	0.9506
<b>6</b>	0.22203947	0.09798342	0.0317	0.9823
<b>7</b>	0.12405606		0.0177	1.0000

Eigenvectors							
	Prin1	Prin2	Prin3	Prin4	Prin5	Prin6	Prin7
<b>Murder</b>	0.300279	-.629174	0.178245	-.232114	0.538123	0.259117	0.267593
<b>Rape</b>	0.431759	-.169435	-.244198	0.062216	0.188471	-.773271	-.296485
<b>Robbery</b>	0.396875	0.042247	0.495861	-.557989	-.519977	-.114385	-.003903
<b>Assault</b>	0.396652	-.343528	-.069510	0.629804	-.506651	0.172363	0.191745
<b>Burglary</b>	0.440157	0.203341	-.209895	-.057555	0.101033	0.535987	-.648117
<b>Larceny</b>	0.357360	0.402319	-.539231	-.234890	0.030099	0.039406	0.601690
<b>Auto_Theft</b>	0.295177	0.502421	0.568384	0.419238	0.369753	-.057298	0.147046



State	Murder	Rape	Robbery	Assault	Burglary	Larceny	Auto_Theft	Prin1	Prin2	Prin3	Prin4	Prin5	Prin6	Prin7
Alabama	14.2	25.2	96.8	278.3	1135.5	1881.9	280.7	-771.15	234.25	-95.08	46.75	-38.21	0.69	4.16
Alaska	10.8	51.6	96.8	284.0	1331.7	3369.8	753.3	676.81	-172.59	358.54	69.76	-125.11	22.00	1.68
Arizona	9.5	34.2	138.2	312.3	2346.1	4467.4	439.5	2072.71	-8.31	-213.88	-87.28	-45.44	-7.31	2.31
Arkansas	8.8	27.6	83.2	203.4	972.6	1862.1	183.4	-881.24	70.04	-119.48	18.60	2.92	6.56	-0.92
California	11.5	49.4	287.0	358.0	2139.4	3499.8	663.5	1170.45	404.94	14.69	22.24	20.10	7.69	-1.76
Colorado	6.3	42.0	170.7	292.9	1935.2	3903.2	477.1	1394.36	-40.15	-64.64	-12.58	-15.34	3.91	-2.96
Connecticut	4.2	16.8	129.5	131.8	1346.0	2620.7	593.2	2.72	124.74	188.88	-84.70	2.72	-4.12	-0.11
Delaware	6.0	24.9	157.0	194.2	1682.6	3678.4	467.0	1072.59	-150.49	13.47	-50.38	21.03	-5.59	0.86
Florida	10.2	39.6	187.9	449.1	1859.9	3840.5	351.4	1299.54	-79.89	-179.33	145.59	-50.83	-6.62	-0.90
Georgia	11.7	31.1	140.5	256.5	1351.1	2170.2	297.9	-415.84	271.53	-128.52	7.72	6.72	4.42	0.74
Hawaii	7.2	25.5	128.0	64.1	1911.5	3920.4	489.4	1383.72	-109.85	-14.25	-220.51	52.90	1.19	3.71
Idaho	5.5	19.4	39.6	172.5	1050.8	2599.6	237.6	-198.92	-219.47	-52.78	-12.42	-29.25	-2.15	0.22
Illinois	9.9	21.8	211.3	209.0	1085.0	2828.5	528.6	64.08	-171.82	229.49	95.90	49.57	-4.19	3.77
Indiana	7.4	26.5	123.2	153.5	1086.2	2498.7	377.4	-249.85	-86.41	69.25	-5.75	26.37	5.29	0.32
Iowa	2.3	10.6	41.2	89.8	812.5	2685.1	219.9	-242.22	-465.35	31.16	-17.57	14.55	-6.45	0.10
Kansas	6.6	22.0	100.7	180.5	1270.4	2739.3	244.3	30.00	-102.92	-108.66	-26.45	18.51	-2.51	0.02
Kentucky	10.1	19.1	81.1	123.3	872.2	1662.1	245.4	-1099.94	94.82	-27.93	-35.99	27.33	4.24	2.65
Louisiana	15.5	30.9	142.9	335.5	1165.5	2469.9	337.7	-230.96	10.80	-21.63	137.20	-34.81	0.49	5.14
Maine	2.4	13.5	38.7	170.0	1253.1	2350.7	246.9	-321.17	59.53	-128.64	-78.98	-31.09	-8.20	-2.76
Maryland	8.0	34.8	292.1	358.9	1400.0	3177.7	428.5	516.91	-89.65	35.88	199.38	67.29	-4.63	-2.96
Massachusetts	3.1	20.8	169.1	231.6	1532.2	2311.3	1140.1	-106.12	623.43	597.25	-51.52	-109.66	-3.36	-1.97
Michigan	9.3	38.9	261.9	274.6	1522.7	3159.0	545.5	565.21	33.55	109.04	80.61	59.27	5.02	-1.10
Minnesota	2.7	19.5	85.9	85.8	1134.7	2559.3	343.1	-184.75	-107.52	30.81	-89.61	32.11	1.42	-2.06
Mississippi	14.3	19.6	65.7	189.1	915.6	1239.9	144.4	-1457.50	307.27	-172.46	-17.41	0.92	3.15	4.55
Missouri	9.6	28.3	189.0	233.5	1318.3	2424.2	378.4	-198.25	154.46	-19.29	29.87	45.20	1.08	0.00
Montana	5.4	16.7	39.2	156.8	804.9	2773.2	309.2	-154.11	-472.62	110.60	43.83	-34.68	-3.26	2.09
Nebraska	3.9	18.1	64.7	112.7	760.0	2316.1	249.1	-582.43	-310.56	53.11	7.84	18.39	1.14	-0.95
Nevada	15.8	49.1	323.1	355.0	2453.1	4212.6	559.2	1927.82	275.36	-142.45	-6.85	72.17	3.21	2.94
New Hampshire	3.2	10.7	23.2	76.0	1041.7	2343.9	293.4	-426.30	-104.47	-0.87	-113.36	-8.59	-4.32	0.22
New Jersey	5.6	21.0	180.4	185.1	1435.8	2774.5	511.5	174.93	110.87	89.60	-30.03	36.63	-5.06	-0.71
New Mexico	8.8	39.1	109.6	343.4	1418.7	3008.6	259.5	346.22	-82.69	-153.61	92.27	-52.83	4.27	-2.10
New York	10.7	29.4	472.6	319.1	1728.0	2782.0	745.8	370.55	483.24	216.04	140.97	183.57	-9.50	-1.21
North Carolina	10.6	17.0	61.3	318.3	1154.1	2037.8	192.1	-636.90	145.11	-183.33	68.12	-71.37	-10.47	2.15

North Dakota	0.9	9.0	13.3	43.8	446.1	1843.0	144.7	-1161.22	-381.02	42.34	-16.91	25.01	-1.12	-2.11
Ohio	7.8	27.3	190.5	181.1	1216.0	2696.8	400.4	-8.27	-57.71	60.59	25.00	67.54	2.07	-0.06
Oklahoma	8.6	29.2	73.8	205.0	1288.2	2228.1	326.8	-397.93	184.43	-76.37	-48.27	-31.90	6.42	-0.33
Oregon	4.9	39.9	124.1	286.9	1636.4	3506.1	388.9	894.96	-118.31	-70.25	18.35	-36.95	5.08	-4.23
Pennsylvania	5.6	19.0	130.3	128.0	877.5	1624.1	333.2	-1116.91	155.31	53.29	-14.01	51.66	2.54	-2.04
Rhode Island	3.6	10.5	86.5	201.0	1489.5	2844.1	791.4	290.14	198.84	319.43	-74.06	-102.69	-13.49	1.54
South Carolina	11.9	33.0	105.9	485.3	1613.6	2342.4	245.1	-138.20	408.49	-295.66	128.87	-120.97	-7.09	-1.47
South Dakota	2.0	13.5	17.9	155.7	570.5	1704.4	147.5	-1217.04	-196.65	-22.74	42.07	-24.32	-2.57	-3.57
Tennessee	10.1	29.7	145.8	203.9	1259.7	1776.5	314.0	-802.89	387.45	-97.14	-31.00	32.64	7.32	-0.95
Texas	13.3	33.8	152.4	208.2	1603.1	2988.7	397.6	425.81	99.03	-66.69	-52.74	22.28	6.28	4.27
Utah	3.5	20.3	68.8	147.3	1171.6	3004.6	334.5	222.74	-288.44	25.60	-32.53	-9.42	-2.36	-0.71
Vermont	1.4	15.9	30.8	101.2	1348.2	2201.0	265.2	-409.86	198.09	-144.18	-172.62	-9.32	-1.90	-3.99
Virginia	9.0	23.3	92.1	165.7	986.2	2521.2	226.7	-296.11	-227.75	-38.40	18.77	21.28	2.00	2.29
Washington	4.3	39.6	106.2	224.8	1605.6	3386.9	360.3	768.42	-108.02	-85.59	-40.24	-18.25	8.80	-4.39
West Virginia	6.0	13.2	42.2	90.9	597.4	1341.7	163.3	-1521.64	-3.07	-26.78	-26.61	23.72	2.09	-0.22
Wisconsin	2.8	12.9	52.2	63.7	846.9	2614.2	220.7	-288.90	-406.98	20.89	-46.78	36.00	-2.82	0.02
Wyoming	5.4	21.9	39.7	173.9	811.6	2772.2	282.0	-154.08	-472.85	80.69	57.47	-37.37	0.75	0.82