

Mardi 15 avril 2008

page 112

Analyse de données

Be et ACP

Revoir le calcul
de detExercice 1

$$A = \begin{pmatrix} 14 & -2 & -2 \\ -3 & 15 & 1 \\ -1 & 1 & 19 \end{pmatrix}$$

Les valeurs propres $\lambda_i, i \in \llbracket 1, 3 \rrbracket$ de A sont les solutions de $|A - \lambda I_3| = 0$.

$$|A - \lambda I_3| = \begin{vmatrix} 14-\lambda & -2 & -2 \\ -3 & 15-\lambda & 1 \\ -1 & 1 & 19-\lambda \end{vmatrix} \begin{vmatrix} 14-\lambda & -2 \\ -3 & 15-\lambda \\ -1 & 1 \end{vmatrix}$$

D'après la règle de Sarrus:

$$|A - \lambda I_3| = (14-\lambda)(15-\lambda)(19-\lambda) + (-2) \cdot 1 \cdot (-1) + (-2)(-3) \cdot 1 \\ - (-1)(15-\lambda)(-2) - 1 \cdot 1(14-\lambda) - (19-\lambda)(-3)(-2)$$

$$|A - \lambda I_3| = (210 + \lambda^2 - 15\lambda - 14\lambda)(19-\lambda) + 2 + 6 \\ - (30 - 2\lambda) - (14 - \lambda) - (114 - 6\lambda)$$

$$|A - \lambda I_3| = (\lambda^2 - 29\lambda + 210)(19-\lambda) + 8 \\ - 30 + 2\lambda - 14 + \lambda - 114 + 6\lambda$$

$$|A - \lambda I_3| = (\lambda^2 - 29\lambda + 210)(19-\lambda) + 9\lambda - 150$$

$$|A - \lambda I_3| = (19\lambda^2 - 29 \times 19\lambda + 19 \times 210 - \lambda^3 + 29\lambda^2 + 210\lambda) \\ + 9\lambda - 150$$

$$|A - \lambda I_3| = -\lambda^3 + (29 + 19)\lambda^2 + (210 + 9 - 29 \times 19)\lambda \\ + 19 \times 210 - 150$$

$$|A - \lambda I_3| = -\lambda^3 + 48\lambda^2 - 332\lambda + 3840$$

Exercice 2

10) On réalise une ACP centrée et réduite de manière à travailler sur des nombres centrés autour de la moyenne et dont les écarts ont été réduits. Ainsi, les écarts entre les ordres de grandeur des variables n'ont pas d'implication sur l'étude et aucune variable n'a de poids plus important que les autres.

20) Les matrices 2 et 3 diffèrent par la valeur encadrée ($i=7, j=2$).

Recalculons $A_{7,2}^s$.

$$A_{7,2}^s = \frac{y_{7,2} - \bar{y}}{\sigma_y}$$

$$\bar{y} = \frac{354 + 388 + 562 + 341 + 396 + 689 + 367 + 484 + 621 + 423 + 548 + 887}{12}$$

$$\bar{y} = \frac{6060}{12} = 505$$

$$\text{On } y_{7,2} = 367$$

$$\text{d'où } y_{7,2} - \bar{y} < 0$$

$$\text{donc } A_{7,2}^s < 0$$

La matrice correcte est la deuxième

$$3. R = \frac{1}{n} A^S A^S$$

~~Recalculons $R_{1,5}$.~~

3°) La matrice des corrélations doit être symétrique. Seule la troisième matrice proposée l'est.

La matrice des corrélations est donc la troisième.

4°) La matrice 2 comporte une valeur de moins. Il y a 7 variables, il doit donc y avoir 7 valeurs propres. Donc la matrice 2 est fautive.

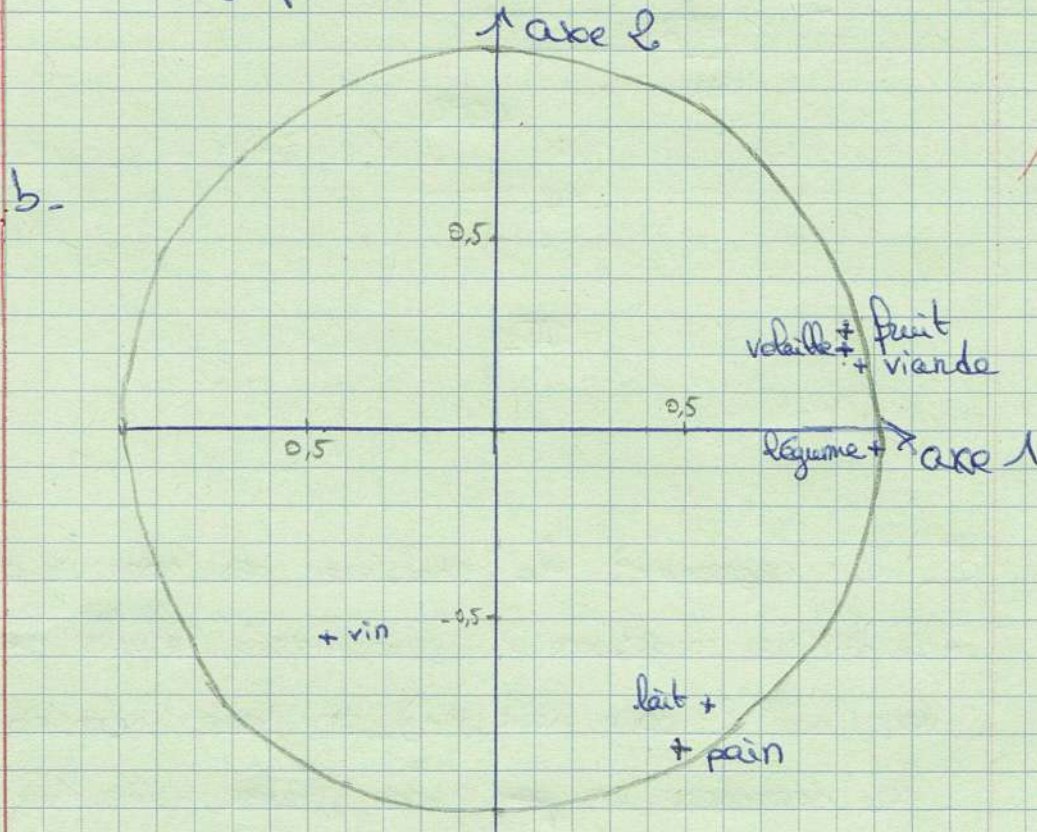
5°) Part d'information portée par chaque axe choisi:

- axe 1: $\frac{\lambda_1}{n} \times 100 = \frac{4,333}{7} \times 100 \approx 61\%$

- axe 2: $\frac{1,830}{7} \times 100 \approx 25\%$

Avec deux axes seulement, on obtient 86% d'information, ce qui permet une bonne analyse (le minimum acceptable étant 60%). En ne choisissant que deux axes, on perd donc 14% d'information.

7) a. Le "qlt" représente l'angle entre la ~~l'~~ individu et l'axe. Une bonne interprétation ne peut se faire que sur les valeurs les plus élevées, celles-ci représentant les données les plus significatives.



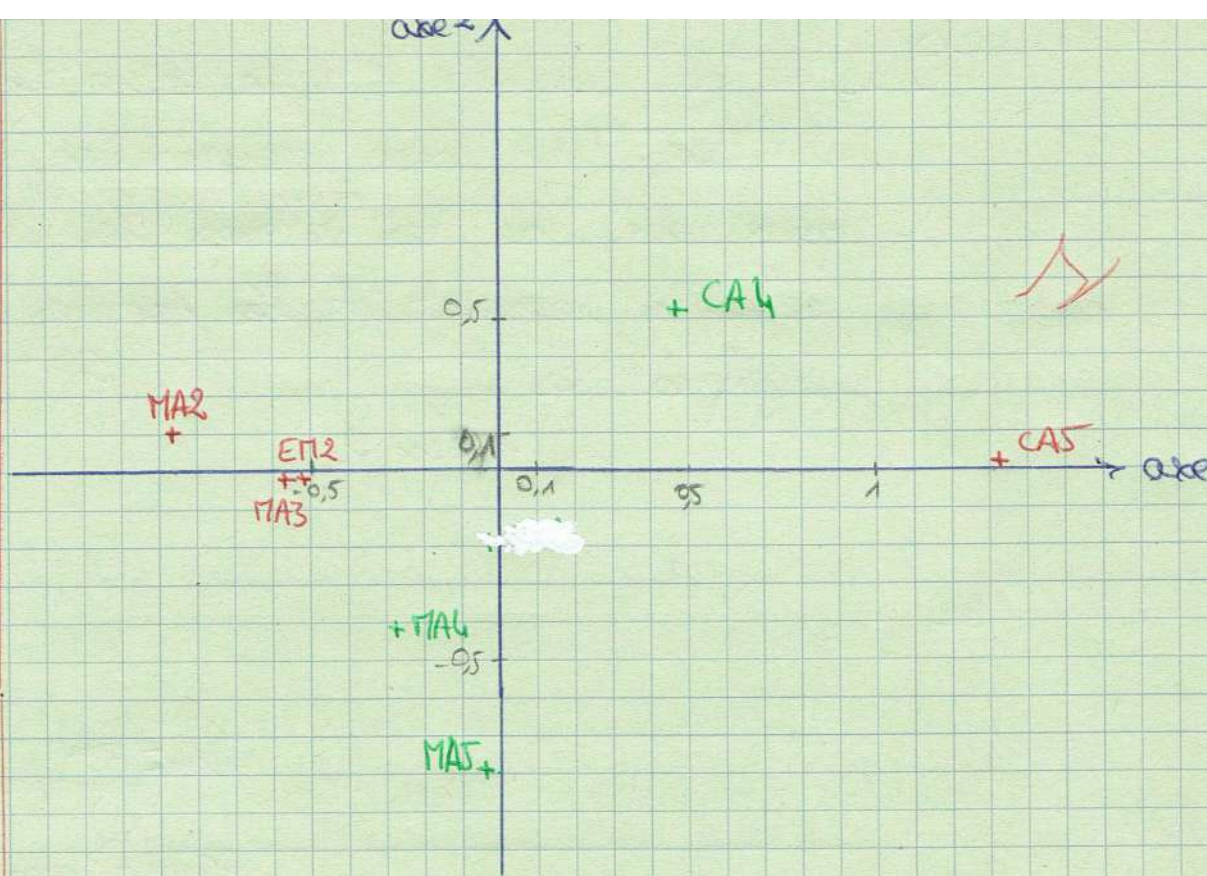
Des valeurs très élevées sur l'axe 1 pour la viande, la volaille, les fruits et les légumes montrent que cet axe sera très représentatif de ces variables. De même, l'analyse du pain, du lait et du vin ne seront faites que sur l'axe 2, leurs valeurs absolues étant moyennes sur l'axe 1 mais élevées sur l'axe 2.

c. Pour tracer la représentation des individus, on privilégie ceux dont la qualité est élevée et

la coordonnée extrême:

- axe 1: MA2, MA3, ET3, CA5(+)

- axe 2: MA4(+), ~~MA4~~, CA4, MA5



- d. En reprenant la matrice des données pour chaque individu (en rouge: axe 1, fruit, viande, volaille, légumes; en vert: axe 2: pain, lait, vin), on peut remarquer que:
- les dépenses en ~~pa~~ viande, volaille, fruits, légumes augmentent très nettement avec le nombre d'enfants
 - les ménages de travailleurs manuels consomment bien plus de vin, pain, lait que les autres.

Ainsi, l'axe 1 représente le nombre d'enfants et l'axe 2 leur catégorie socio-professionnelle.

NOM : CRAVERO
Prénom Emilie

Groupe : B

Annexe à rendre avec la copie

Question 2 :

Voici trois propositions pour la matrice centrée réduite associée aux données étudiées. Choisissez celle qui vous semble correcte, en justifiant votre choix.

1.

-0,323	-0,276	-0,343	-0,335	-0,286	-0,485
-0,432	-0,214	-0,274	-0,285	-0,307	-0,276
-0,210	0,104	0,047	0,150	-0,317	0,056
-0,114	-0,300	-0,289	-0,313	-0,088	-0,269
-0,171	-0,199	-0,294	-0,296	-0,101	-0,198
-0,024	0,336	0,349	0,417	-0,297	0,177
0,246	-0,252	-0,203	-0,200	0,144	-0,115
0,038	-0,038	-0,023	-0,050	0,107	-0,053
-0,174	0,212	0,365	0,418	-0,140	0,091
0,586	-0,150	-0,030	-0,053	0,352	0,070
0,386	0,079	0,129	0,109	0,411	0,419

2.

$j=2$

-0,323	-0,276	-0,343	-0,335	-0,286	-0,485
-0,432	-0,214	-0,274	-0,285	-0,307	-0,276
-0,210	0,104	0,047	0,150	-0,317	0,056
-0,114	-0,300	-0,289	-0,313	-0,088	-0,269
-0,171	-0,199	-0,294	-0,296	-0,101	-0,198
-0,024	0,336	0,349	0,417	-0,297	0,177
0,246	-0,252	-0,203	-0,200	0,144	-0,115
0,038	-0,038	-0,023	-0,050	0,107	-0,053
-0,174	0,212	0,365	0,418	-0,140	0,091
0,586	-0,150	-0,030	-0,053	0,352	0,070
0,386	0,079	0,129	0,109	0,411	0,419
0,192	0,698	0,566	0,440	0,522	0,582

$i=7$

3.

$j=2$

-0,323	0,276	-0,343	-0,335	-0,286	-0,485
-0,432	0,214	-0,274	-0,285	-0,307	-0,276
-0,210	0,104	0,047	0,150	-0,317	0,056
-0,114	-0,300	-0,289	-0,313	-0,088	-0,269
-0,171	-0,199	-0,294	-0,296	-0,101	-0,198
-0,024	0,336	0,349	0,417	-0,297	0,177
0,246	0,252	-0,203	-0,200	0,144	-0,115
0,038	-0,038	-0,023	-0,050	0,107	-0,053
-0,174	0,212	0,365	0,418	-0,140	0,091
0,586	-0,150	-0,030	-0,053	0,352	0,070
0,386	0,079	0,129	0,109	0,411	0,419
0,192	0,698	0,566	0,440	0,522	0,582

$i=7$

Réponse : La matrice des données comportant 12 individus, la matrice centrée réduite associée doit également comporter 12 lignes. La matrice n'en comporte que 11, elle est donc fautive.

Voir suite sur la copie 2. (réponse 2)

Question 3 :

Voici trois propositions pour la matrice de corrélation associée aux données étudiées. Choisissez celle qui vous semble correcte, en justifiant votre choix.

1.

1,000	0,196	0,321	0,248	0,856	0,593
0,196	1,000	0,959	0,926	0,332	0,856
0,321	0,959	1,000	0,982	0,375	0,881
0,248	0,926	0,982	1,000	0,233	0,827
0,856	0,332	0,375	0,233	1,000	0,663
0,593	0,856	0,881	0,827	0,663	1,000

2.

1,000	0,196	0,321	0,248	0,856	0,593	0,304
0,196	1,000	0,959	0,926	0,332	0,856	-0,486
0,321	0,959	1,000	0,982	0,375	0,881	-0,437
0,248	0,926	0,982	1,000	0,233	0,827	-0,400
0,856	0,332	0,375	0,233	1,000	0,663	0,007
0,593	0,856	0,881	0,827	0,663	1,000	-0,356
0,304	-0,486	-0,437	-0,400	1,345	-0,356	1,000

3.

1,000	0,196	0,321	0,248	0,856	0,593	0,304
0,196	1,000	0,959	0,926	0,332	0,856	-0,486
0,321	0,959	1,000	0,982	0,375	0,881	-0,437
0,248	0,926	0,982	1,000	0,233	0,827	-0,400
0,856	0,332	0,375	0,233	1,000	0,663	0,007
0,593	0,856	0,881	0,827	0,663	1,000	-0,356
0,304	-0,486	-0,437	-0,400	0,007	-0,356	1,000

Réponse : La matrice 1 est incomplète, donc fautive. Les matrices 2 et 3 diffèrent par une unique valeur encadrée. Voir suite sur la copie 2 (réponse 3)

Question 4 :

Voici trois propositions pour les valeurs propres associées aux données étudiées. Choisissez celle qui vous semble correcte, en justifiant votre choix.

1.

λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7
4,333	1,830	0,631	0,128	0,058	0,019	0,001

2.

λ_1	λ_2	λ_3	λ_4	λ_5	λ_6
4,333	1,830	0,631	0,128	0,058	0,019

3.

λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7
4,333	1,830	1,631	0,128	0,058	0,019	0,001

Réponse : Voir copie 2.

Question 6 :

Voici trois propositions pour les vecteurs propres associée aux données étudiées. Choisissez celle qui vous semble correcte, en justifiant votre choix.

1.

λ_1	λ_2
u1	u2
0,240	0,622
0,446	0,205
0,462	0,141
0,438	0,197
0,281	0,523
0,466	0,098
0,206	0,479

2.

λ_1	λ_2
u1	u2
0,240	-0,622
0,446	0,205
0,462	0,141
0,438	0,197
0,281	-0,523
0,466	-0,098
-0,206	-0,479

3.

λ_1	λ_2
u1	u2
0,240	-0,622
0,446	0,205
0,462	0,141
0,438	0,197
0,281	-0,523
0,466	-0,098

Réponse : La matrice 3 ne comporte que 6 coordonnées par vecteur (au lieu de 7), elle est donc fautive.