Corrigé (Série 2)

1) Le nuage des individus sera représenté dans $\mathbb{R}^5$ (nombre de variable)

2) Le nuage des variables sera représenté dans $\mathbb{R}^8$ (nombre d'individus).

3) Compléter le tableau:

Variable	Moyenne	Ecart-type.
S	1.85	0.2179
QS	6.05	0.3808
QT	6.65	1.1292
A	6.70	1.8009
Fu	6.81.	0.5644

Par définition: $\bar{x}^j = \sum_{i=1}^{8} p_i x_{ij}$ ou $p_i = \frac{1}{8}$ $\forall i = 1,8$.

$$s^j = \sum_{i=1}^{8} p_i (x_{ij} - \bar{x}^j)^2$$

4) Les coordonnées du centre de gravité sont donc!

$$\bar{g} = (1.85, 6,05, 6,65, 6.70, 6.81)^t$$

5) $M = D_{1/s^2} = \begin{pmatrix} 0.0475 & 0 & 0 & 0 & 0 \\ 0 & 0.1450 & 0 & 0 & 0 \\ 0 & 0 & 1.2750 & 0 & \\ 0 & 0 & 0 & 0.3186 & \end{pmatrix}$

6) 03 valeurs significative du coefficients de corrélation, entre: S et QS et QS et QT par conséquent entre S et QT

6) La somme des valeurs propres :
$2.5837 + 1.648 + 0.5060 + 0.2668 + 0.0187 = 5$.
Comme il s'agit d'une ACP normée donc il est question de diagonaliser la matrice de ~~[illegible]~~ corrélation, R avec :
$trace(R) = 5 = \sum_{k=1}^{5} \lambda_k$, donc les résultats sont en accord avec ceux du cours.

7) le pourcentage d'inertie expliqué par un axe k vaut :

$$I_k = \frac{\lambda_k}{\sum_{k=1}^{5} \lambda_k} \times 100.$$

le pourcentage cumulé est donc :

$$I_k(\text{cumulé}) = I_1 + I_2 + \dots + I_k.$$

Ce qui donne :

	I_k	I_k (cumulé)
c_1	51,65	51,65
c_2	32,58	84.23
c_3	10,08	94.31
c_4	5,31	99.62
c_5	0,38	100.00

8)
En calcul les première différence :

$\varepsilon_1 = \lambda_2 - \lambda_1 = 0.95$
$\varepsilon_2 = \lambda_3 - \lambda_2 = 1.12$
$\varepsilon_3 = \lambda_4 - \lambda_3 = 0.23$
$\varepsilon_4 = \lambda_5 - \lambda_4 = 0.24$

Ensuite les différences seconde :

$\delta_1 = \varepsilon_2 - \varepsilon_1 = -0,17$.
$\delta_2 = \varepsilon_3 - \varepsilon_2 = 0.88$.
$\delta_3 = \varepsilon_4 - \varepsilon_3 = -0.0089$

La première seconde différence négative est ε_2 donc on retient 2 axes (2 valeurs propres) 84,23 %

9) Une composante principale est définie par :

$$C_k = X u_k$$ où u_k est le vecteur propre de R associé à la $k^{ième}$ plus grande valeur propre :

$$C_k = \sum_{j=1}^{p} u_{kj} X^j$$

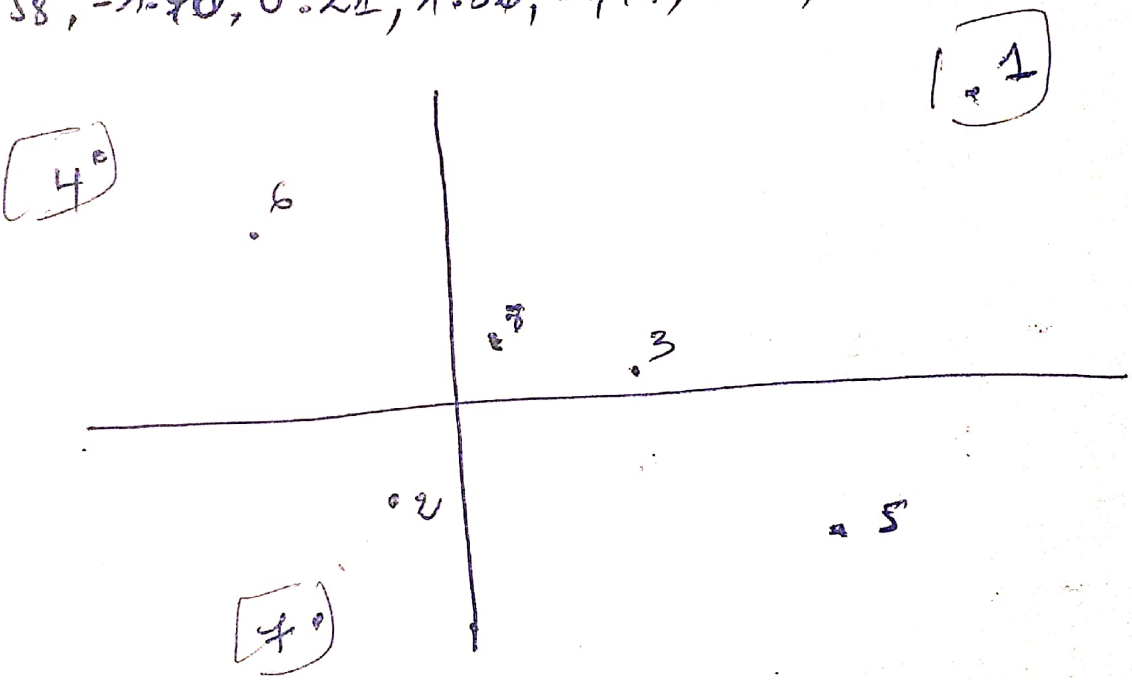
X^j : La $j^{ème}$ colonne du tableau centré et réduit

u_{kj} : La $j^{ème}$ coordonnée de u_k.

10) $C_1 = (2.68, -0.42, 1.21, -2.30, 1.44, -1.70, -1.2, 0.2814)^t$.

$C_2 = (1.58, -1.40, 0.21, 1.50, -1.47, 0.98, -1.73, 0.31)^t$.

11)



- Les individus 1 et 4 possèdent les plus forte coordonnées ~~pos~~ (valeur absolue) donc ils contribuent plus que les autres à l'inertie de l'axe 1.
- Sur l'axe 2 1, 4 et 7 sont les plus importants en terme de contribution.

Scanné avec CamScanner

définition, la contribution d'un individu à l'inertie d'un axe (D_k) porté par le vecteur propre a_k de R, associé à la plus grande valeur propre λ_k est :

$$Cont(e_i, D_k) = \frac{p_i C_{ik}^2}{\lambda_k} = \frac{1}{8} \cdot \frac{C_{ik}^2}{\lambda_k}, \text{ où}$$

C_{ik} est la coordonnée de (e_i) sur l'axe (D_k).

Pour l'axe 1 et 2 on a obtenu le tableau suivant :

individu	axe 1	axe 2
1	$\boxed{39{,}97\%}$	$\boxed{19{,}32\%}$
2	0,87 %	15,01 %
3	7,11 %	0,33 %
4	$\boxed{25{,}46\%}$	$\boxed{17{,}46\%}$
5	10,08 %	16,65 %
6	14,12 %	7,56 %
7	6,97 %	$\boxed{22{,}96\%}$
8	0,38 %	0,73 %

~~13) La qualité de~~

Donc on retrouve bien les résultats du plan 1, 2.

c'est-à-dire : 1 et 4 pour l'axe 1
1, 4 et 7 pour l'axe 2.

13) La qualité de représentation d'un individu e_i sur un axe (D_k) est définie par :

$$qual(e_i, D_k) = \cos^2(\vec{e_i}, \vec{a_k}), \text{ où } \vec{a_k} \text{ est le}$$

vecteur propre de R.

Scanné avec CamScanner

$$Qual(e_i, axe_k) = \cos^2(\vec{e_i}, \vec{a_k}) = \frac{\langle e_i, a_k \rangle_M}{\|\vec{a_k}\|_M \cdot \|\vec{e_i}\|_M}$$

et Comme $M = I_{R^P}$ (données centrées et réduite) donc:

$$Qual(e_i, axe_k) = \frac{C_{ik}^2}{\sum_{k=1}^{5} C_{ik}^2}$$

La qualité de représentation sur le plan (k,m) est:

$$Qualité(e_i, plan(k,m)) = \frac{C_{ik}^2 + C_{im}^2}{\sum_{k=1}^{5} C_{ik}^2}.$$

Les résultats pour les 2 premiers axes sont donnés dans le tableau suivant:

individus	axe1	axe2	plan(1,2)
1	$\boxed{0,71}$	0,25	$\boxed{0,96}$
2	0,058	$\boxed{0,63}$	0,69
3	$\boxed{0,87}$	0,03	$\boxed{0,90}$
4	0,64	0,28	$\boxed{0,92}$
5	0,32	0,34	0,66
6	0,60	0,20	$\boxed{0,80}$
7	0,30	$\boxed{0,62}$	$\boxed{0,92}$
8	0,094	0,11	0,20.

sur l'axe 1 : individu 1 et 3

sur l'axe 2 : ~~0,63~~ individu 2 et 7

sur le plan (1,2) : 1, 3, 4, 7, 6.

le cos est proche de 1 donc un angle entre ei et l'axe proche de 0.

14] Les relations de transition sont:

$$u_k = \cancel{M} a_k \text{ et } c_k = X u_k.$$

Comme les données sont centrées et réduite et $M = I_{\mathbb{R}^5}$ donc le facteur est le même que l'axe principale, donc $\boxed{u_k = a_k}$

$$u_1 = a_1 = (-0{,}56, -0{,}56, -0{,}59, +0{,}06, -0{,}02)^t$$

$$u_2 = a_2 = (-0{,}13, 0{,}033, 0.195, 0{,}68, -0{,}69)^t.$$

Dans le cas d'une ACP Normée, la correlation d'une variable X^j avec un axe k vaut:

$$cor(X^j, c_k) = \sqrt{\lambda_k}\, u_{kj}$$

où u_{kj} est la $j^{ème}$ coordonnées du facteur u_k.

~~les axes~~ le tableau suivant donne les correlations des 5 variables avec les 2 premiers axes: (composantes).

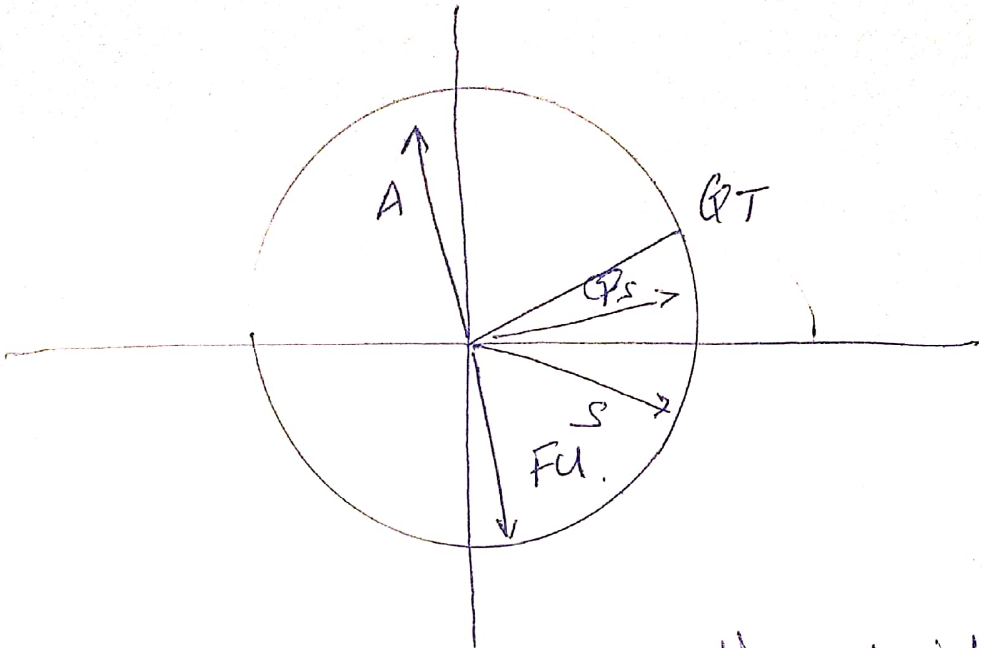
variables	c_1	c_2
S	0,91	−0,17
QS	0.90	−0,04
QT	0.95	0.24
A	−0.10	0,86
FU	0,038	−0,88

Scanné avec CamScanner

les variables : S, QS et QT sont très corrélés avec l'axe 1.

Les variables A et FU avec l'axe 2 avec une corrélation négative entre FU et C_2.

15) La représentation des variables se fait par le moyen du cercle des corrélations (Voir Cours).



16) La qualité de représentation d'une variable sur un plan est définie par la longueur de la flèche. Donc les variables les mieux représentés sont QT et S avec des normes proche de 1.

18). Interprétation: On interprete conjointement:

1- Le Cercle des correlations. ($C_1 - C_2$).

2- La position des individus sur le plan (1-2)

Scanné avec CamScanner

- Pour le cercle des corrélations : la première composante principale $C_1 = Xu_1$ est fortement corrélée avec les variables QT, QS, S (qu'on peut classer comme variable de bon usage), et pas corrélée avec A et FU.
- La 2ème composante $C_2 = Xu_2$ est fortement corrélée positivement avec A et négativement FU, donc l'axe 2 oppose 2 types de téléphones :
  - Un téléphone facile à utiliser mais qui consomme plus ~~de batterie~~ de batterie.

Sur le plan des individus, on constate une opposition sur l'axe 2 de 2 groupes de téléphones :

- 1, 4, 6 bonne batterie et mauvaise utilisation
- 7, 2, 6 mauvaise " " bonne " .

- Sur l'axe 1, on constate une opposition entre de téléphones bon pour les transmissions (appel ?) : 1, 8, 3, 5 et des téléphones mauvais pour les appels 4, 6, 7, 2.

Il y a lieu de noter que l'information contenue dans l'axe 1 est indépendante de l'axe 2, ce qui se traduit par :

Il y pas pas de relation entre la qualité des appels et l'autonomie de batterie d'une part et la ~~qualité de~~ facilité d'utilisation d'autre part.

— fin —

(08)