Partie 1 : TD

Question 1

1) La droite de projection orthogonale a pour but de minimiser la somme des carrés des distances $M_i m_i$, $\frac{1}{m}\sum_{i=1}^{m} M_i m_i^2$.
Cherchons à retrouver pourquoi cette droite D doit passer par le point moyen du nuage.

En débutant avec le calcul de l'inertie d'une variable statistique par rapport à un point A.
L'inertie est définie comme la moyenne du carré des distances de la variable X au point A.

On a : $I_T = \sum_{i=1}^{m} \frac{1}{m} d^2(A, X_i) = \sum_{i=1}^{m} \frac{1}{m}\left[\sum_{j=1}^{m}(a_j - x_{ij})^2\right] = \frac{1}{m}\sum_{i=1}^{m}\sum_{j=1}^{m}(a_j - x_{ij})^2$

Or l'inertie I_T est dérivable puisqu'elle s'écrit comme une somme d'éléments eux-mêmes dérivables.

Ainsi : $I_T' = \frac{1}{m}\sum_{i=1}^{m} 2\sum_{j=1}^{m}(a_j - x_{ij}) = \frac{1}{m}\sum_{i=1}^{m} 2\sum_{j=1}^{m}(a_j - X_{ij}) = \frac{1}{m}\sum_{i=1}^{m} 2ma_j \sum_{j=1}^{m}(a_j - X_{ij})$

On cherche à résoudre $I_T' = 0 \iff a_j = \frac{1}{m}\sum_{j=1}^{m} X_{ij} = g$ avec g, le centre de gravité du nuage de point.

Ainsi I_T' s'annule en g, qui est donc un minimum globale de I_T

On peut donc en conclure que D doit passer par le point moyen du nuage

Question 2 : Etude de la projection associée à l'ACP en 2D

2.1 Posons : $\varphi : \mathbb{R}^2 \longmapsto \mathbb{R}^{2c}$

$(u_1, u_2) \longmapsto [(A - \lambda I_2)u_1 ; (A - \lambda I_2)u_2]$ ainsi φ définit le noyau de l'endomorphisme
On a à retrouver les conditions pour lesquelles l'application linéaire du noyau de l'endomorphisme, soit φ ne se réduise pas à 0.

Comme $\varphi : \mathbb{R}^2 \longmapsto \mathbb{R}^{2e}$, nous sommes en dimension finie. Alors il suffit de montrer que si $\varphi(u_1, u_2) = (0;0)$ n'implique pas que (u_1, u_2) soit égal au vecteur nul, c'est à dire que $(u_1, u_2) = (0,0)$

Ainsi : $\varphi(u_1, u_2) = (0;0) \Rightarrow \begin{cases} (A - \lambda I_2) u_1 = 0 \\ (A - \lambda I_2) u_2 = 0 \end{cases} \Rightarrow \begin{cases} A u_1 - \lambda u_1 = 0 \\ A u_2 - \lambda u_2 = 0 \end{cases} \Rightarrow \begin{cases} \lambda_1 = \lambda_1 u_1 \\ \lambda_2 = \lambda_2 u_2 \end{cases}$

D'après l'écriture précédente on a $(u_1, u_2) = (1, 1)$ ainsi nous avons un contre exemple.

L'application n'est donc pas injective, donc bijective dans ces termes conditions.

la condition : $\text{Ker}(A - \lambda I_2) \neq 0 \iff \det(A - \lambda I_2) = 0$

2.2 L'équation du 2nd degré dont les valeurs λ_1 et λ_2 sont solutions est : $\lambda^2 - \text{tr}(A)\lambda + \det(A) = 0$
où A est la matrice de variance / covariance. $A = \begin{pmatrix} \text{var}(X) & \text{cov}(X,Y) \\ \text{Cov}(Y,X) & \text{var}(Y) \end{pmatrix}$

donc $\text{tr}(A) = V(X) + V(Y) = S^2(X) + S^2(Y)$

$$\det(A) = V(X).V(Y) - \text{Cov}^2(X,Y) = S^2(X)S^2(Y) - \text{cov}^2(X,Y)$$

2.3 Calculons le discriminant du polynôme.

ainsi : $\Delta = \left(V(X)+V(Y)\right)^2 - 4\left[V(X)V(Y) - \text{cov}^2(X,Y)\right]$
$= V^2(X) + 2V(X)V(Y) + V^2(Y) - 4V(X)V(Y) + 4\text{cov}^2(X,Y)$
$= V^2(X) + V^2(Y) - 2V(X)V(Y) + 4\text{cov}^2(X,Y)$
$= \left[V(X) - V(Y)\right]^2 + 4\text{cov}^2(X,Y)$

ainsi $\left[V(X) - V(Y)\right]^2 > 0$ et $4\text{cov}^2(X,Y) > 0$
donc $\Delta > 0$, ainsi il y a deux racines $(\lambda_1, \lambda_2) \in \mathbb{R}^2$

ainsi $\lambda_1 = \frac{1}{2}\left[V(X) + V(Y) + \sqrt{(V(X)-V(Y))^2 + 4\,Cov^2(X,Y)} \right]$

$\lambda_2 = \frac{1}{2}\left[V(X) + V(Y) - \sqrt{(V(X)-V(Y))^2 + 4\,Cov^2(X,Y)} \right]$

2.4 $(u_1 ; u_2)$ une base orthonormée ; $Au_1 = \lambda_1$ et $Au_2 = \lambda_2$.

et le vecteur propre U associé à la valeur propre λ, avec $U = \begin{pmatrix} S^2(Y) - \lambda \\ -Cov(X,Y) \end{pmatrix}$

et U vecteur propre associé à λ, de la matrice A ssi $AU = \lambda U$

Ainsi : $AU = \begin{pmatrix} V(X) & cov(Y,X) \\ Cov(X,Y) & V(Y) \end{pmatrix} \begin{pmatrix} S^2(Y) - \lambda \\ -Cov(X,Y) \end{pmatrix}$

$$\Rightarrow \begin{pmatrix} V(X)(S^2(Y)-\lambda) - cov^2(X,Y) \\ cov(Y,X)(S^2(Y)-\lambda) - V(Y)cov(X,Y) \end{pmatrix}$$

$$\Rightarrow \begin{pmatrix} S^2(X)S^2(Y) - \lambda S^2(X) - cov^2(X,Y) \\ cov(X,Y)(S^2(Y)-\lambda) - S^2(Y)Cov(X,Y) \end{pmatrix}$$

d'après q. 2.2

$$\Rightarrow \begin{pmatrix} \det(A) - \lambda S^2(X) \\ -\lambda Cov(X,Y) \end{pmatrix} \quad \text{et } \lambda^2 - (S^2(X)+S^2(Y))\lambda + \det(A) = 0$$

$$\Rightarrow \begin{pmatrix} -\lambda^2 + (S^2(X)+S^2(Y))\lambda - \lambda S^2(X) \\ -\lambda Cov(X,Y) \end{pmatrix}$$

$$\Rightarrow \begin{pmatrix} -\lambda^2 + \lambda S^2(X) \\ -\lambda\, Cov(X,Y) \end{pmatrix}$$

$$\Rightarrow \lambda \begin{pmatrix} -\lambda + S^2(X) \\ -Cov(X,Y) \end{pmatrix}$$

$$\Rightarrow \lambda U$$

Ainsi, nous avons bien $AU = \lambda U$, avec U un vecteur propre relatif à la valeur propre λ

2.5 $(u_1 ; u_2)$ étant une BON, $\|u_1\|^2 = 1$, $\|u_2\|^2 = 1$

et par ailleurs $\|U\| = \sqrt{\|U\|^2} = \sqrt{{}^tUU}$

avec ${}^tUU = \begin{pmatrix} S^2(Y) - \lambda & -Cov(X,Y) \end{pmatrix} \begin{pmatrix} S^2(Y) - \lambda \\ -cov(X,Y) \end{pmatrix}$

$$= \begin{pmatrix} (S^2(Y) - \lambda)^2 + Cov^2(X,Y) \end{pmatrix}$$

ainsi $\|U\| = \sqrt{(S^2(Y) - \lambda)^2 + Cov^2(XY)}$

2.6 Le vecteur propre relatif à la valeur propre λ :

$$U_{normé} = \begin{pmatrix} \frac{S^2(Y) - \lambda}{\sqrt{(S^2(Y) - \lambda)^2 + Cov^2(X,Y)}} \\ \frac{-Cov(X,Y)}{\sqrt{(S^2(Y) - \lambda)^2 + Cov^2(X,Y)}} \end{pmatrix}$$

2.7 Ainsi, $u_1 = \frac{1}{(S^2(Y) - \lambda_1)^2 + Cov^2(X,Y)} \begin{pmatrix} S^2(Y) - \lambda_1 \\ -Cov(X,Y) \end{pmatrix}$

$$u_2 = \frac{1}{(S^2(Y) - \lambda_2)^2 + Cov^2(X,Y)} \begin{pmatrix} S^2(Y) - \lambda_2 \\ -Cov(X,Y) \end{pmatrix}$$

TP n°1 _ ACP

Question n°3.1.

On définit la matrice par $X = \begin{pmatrix} 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \end{pmatrix}^T$ ainsi que la matrice des poids $D = \frac{1}{6} I_n$, puisqu'il y a 6 individus dans notre jeu de données.

Une rapide boucle sur les colonnes de X permet de nous donner la moyenne de chacune des variables, soit le vecteur $M = (0.5 \quad 0.5)$.

La matrice Z centrée est alors définie par $Z = \begin{pmatrix} -0.5 & 0.5 & -0.5 & 0.5 & 0.5 & -0.5 \\ -0.5 & 0.5 & 0.5 & -0.5 & 0.5 & -0.5 \end{pmatrix}^T$.

Finalement, en "re-biaisant" la fonction cov() sur R, on obtient la matrice de covariance $A = \frac{n-1}{n} cov(Z) = \begin{pmatrix} 0.25 & 0.8333 \\ 0.8333 & 0.25 \end{pmatrix}$ il vient donc directement $S^2(X) = S^2(Y) = 0.25$ et $cov(X, Y) = 0.83$.

Question n°3.2.

Pour la suite du TP, nous avons développé une fonction générique qui réalise les calculs nécessaires à l'ACP pour une matrice de taille n x p.

Pseudo-code :

```
ACP(X, centrer = FALSE) :
        Récupérer les dimensions n et p de X
        Pour chaque colonne de X : Z <= colonne – moyenne(colonne)
        Si (centrer == TRUE) :
                Z <= Z / var(colonne)
        Afficher Z
        Calculer la matrice de covariance A
        Afficher A
        Diagonaliser A
        U <= vecteurs propres
        Lambda <= valeurs propres
        Composantes <= X.u
Retourner Z, A, U, Lambda, Composantes
```

Après exécution du code avec notre matrice X, on obtient $\lambda_1 = 0.33334$ et $\lambda_2 = 0.16667$.
On a donc bien l'égalité $tr(A) = \lambda_1 + \lambda_2 = 0.5$ ainsi que $det(A) = \lambda_1 \lambda_2$.

Question n°3.3.

D'après le TD : $u_i = (S^2(Y) - \lambda_i \,;\, -cov(X, Y))$, d'où les expressions des vecteurs propres $u_1 = (0.7071 \,;\, 0.7071)$ et $u_2 = (-0.7071 \,;\, 0.7071)$.

Notre fonction R nous renvoie les mêmes valeurs de vecteurs propres.

Question n°3.4.

On a $I_s(u_1)$ que l'on calcule directement avec une boucle sur R, d'où $I_s(u_1) = 0.3333 = \lambda_1$. De même avec $I_s(u_2) = 0.16667 = \lambda_2$.

On obtient aisément par calcul matriciel sur R les égalités $u^T Au = \lambda u^T u = \lambda \|u\|^2 = \lambda$.

Question n°3.5.

Après calcul matriciel sur R, on obtient une inertie totale de 0.5, ce qui correspond bien à la somme des deux inerties statistiques portées par chacun des axes principaux.

Question n°3.6.

Le graphique ci-dessous représente les 6 points ainsi que le tracé des deux vecteurs portant les axes principaux.

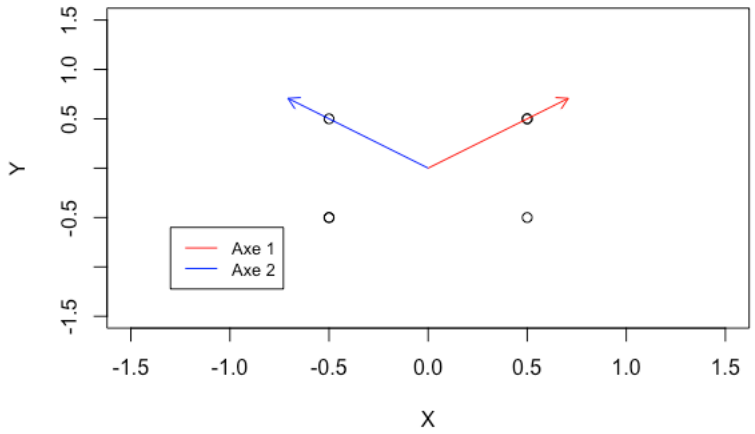

Figure 1 : Tracé des deux axes principaux dans le plan

Question n°3.7.

Le taux d'inertie expliqué par la 1ère composante est d'environ 66%, alors que la 2nd composante explique quant à elle 33%. On remarque ici que la deuxième composante ne compte que pour 1/3 de l'inertie totale, on ne considérant que la proportion d'inertie expliquée par chacun des axes, il pourrait alors être judicieux de ne conserver que la première composante.

Question n°4.

On obtient les coordonnées des individus dans notre repère (O, C1, C2) par un rapide calcul vectoriel. Ces coordonnées correspondent aux lignes de la matrice Composantes retournées par notre fonction ACP.

On possède alors la matrice $C = \begin{pmatrix} -0.707 & 0.707 & 0 & 0 & 0.707 & -0.707 \\ 0 & 0 & 0.707 & -0.707 & 0 & 0 \end{pmatrix}^T$. La figure ci-dessous représente les points projetés dans notre nouveau repère. On constate qu'en 2 dimensions, en conservant les deux composantes, notre ACP revient simplement à faire une rotation de notre nuage de points.

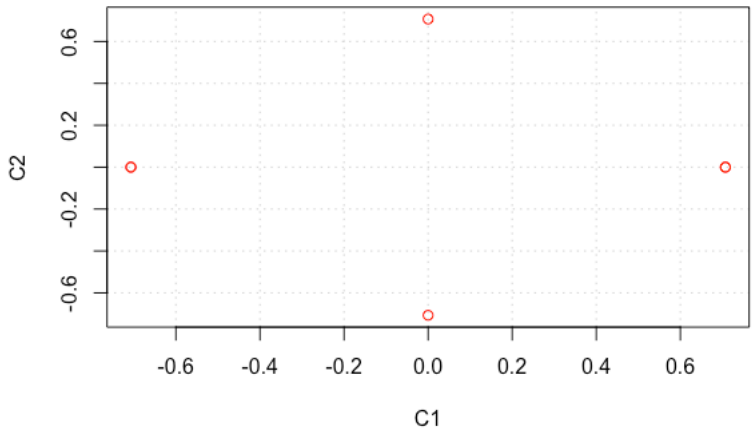

Figure 2 : Points projetés dans le repère (O, C1, C2)

Question n°5.

Sur l'espace E à 2 dimensions, la qualité de projection des individus vaut forcément 1 puisque l'on ne réduit pas la dimension. En revanche, en calculant pour chaque individu la qualité de sa projection sur l'espace E à une dimension engendré par le 1er vecteur propre, on obtient un vecteur de la forme $Q = (1\ 1\ 0\ 0\ 1\ 1)$. Pour la qualité des individus si l'on considère le sous espace engendré par le 2nd vecteur propre, on obtient $Q = (0\ 0\ 1\ 1\ 0\ 0)$.

On remarque que les individus sont parfaitement projetés sur un axe ou bien sur l'autre : effectivement les axes C1 et C2 correspondent en fait aux deux diagonales reliant les sommets du rectangle formés par les 6 individus. Les individus 3 et 4 appartiennent donc à l'axe C2, alors que les autres appartiennent à l'axe C1. Le repère (0, C1, C2) étant orthonormé, chaque axe correspond au donc noyau de la projection orthogonale sur l'autre axe.

Question n°6 et n°7.

De la même manière que précédemment, on obtient une matrice de contribution des individus à une composante notée $Ct = \begin{pmatrix} 0.25 & 0 \\ 0.25 & 0 \\ 0 & 0.5 \\ 0 & 0.5 \\ 0.25 & 0 \\ 0.25 & 0 \end{pmatrix}$. Comme 4 des individus se trouvent sur l'axe C1, ils n'y contribuent chacun qu'à 25%, alors que les deux individus de l'axe C2 y contribuent à 50% chacun (effectivement chaque individu est affecté d'un même poids).