

# TD : Dimension

## 1 Exercice 1 : Compréhension de l'ACP

Soit le vecteur aléatoire  $X = (X_1, X_2, X_3, X_4)^\top$  dont l'espérance et la variance sont, respectivement, données par

$$\mu = \begin{pmatrix} 0 \\ 1 \\ -1 \\ 0 \end{pmatrix} \text{ et } \Sigma = \begin{pmatrix} 9 & 1 & -1 & 2 \\ 1 & 4 & -1 & 1 \\ -1 & -1 & 16 & 0 \\ 2 & 1 & 0 & 9 \end{pmatrix}.$$

Les valeurs et vecteurs propres (normés) correspondants de  $\Sigma$  sont

$$\lambda_1 = 16.27, \quad \alpha_1 = (0.165, 0.098, -0.980, 0.059)^\top, \quad (1)$$

$$\lambda_2 = 11.12, \quad \alpha_1 = (0.665, 0.169, 0.171, 0.707)^\top, \quad (2)$$

$$\lambda_3 = 6.95, \quad \alpha_1 = (0.718, -0.017, 0.077, -0.691)^\top, \quad (3)$$

$$\lambda_4 = 3.67, \quad \alpha_1 = (0.118, -0.981, -0.070, 0.139)^\top. \quad (4)$$

$$(5)$$

1. Trouvez une combinaison linéaire de  $X_1, X_2, X_3$  et  $X_4$  dont la variance est 6.95.
2. Soit  $Z = a_0 + a_1 X_1 + a_2 X_2 + a_3 X_3 + a_4 X_4$ . Trouvez les valeurs de  $a_0, a_1, a_2, a_3, a_4$  telles que  $Z$  est de moyenne 0 et de variance maximale. Quelle est la valeur de cette variance ? Quelle est la covariance entre cette combinaison linéaire et celle trouvée dans la question 1.
3. Donnez une matrice diagonale  $M$  telle que  $M\Sigma M = R$ , où  $R$  est la matrice des corrélations de  $X$ .

## 2 Exercice 2 : Matrice de covariance et de corrélation

Démontrer que la matrice de variance-covariance échantillonnale des données centrées et réduites est égale à la matrice de corrélation échantillonnale des données initiales.

## 3 Exercice 3 : Pratique de l'ACP

Soit un vecteur de trois variables aléatoires  $X = (X_1, X_2, X_3)^\top$  dont la matrice de variance est donnée par  $\text{Var}(X) = \Sigma$ . On vous dit que les valeurs propres de  $\Sigma$  sont 3, 2 et 1 et que les vecteurs propres normés correspondants sont, respectivement,  $v_1 = (0, -1/\sqrt{2}, 1/\sqrt{2})^\top$ ,  $v_2 = (1, 0, 0)^\top$  et  $v_3 = (0, 1/\sqrt{2}, 1/\sqrt{2})^\top$ .

1. En fait,  $X_1, X_2$  et  $X_3$  représentent, respectivement, la circonférence du poignet droit, un score de capacité pulmonaire et l'indice de masse corporelle des individus d'une certaine population. Les médecins aimeraient résumer ces trois mesures en un seul score, noté  $C$ , qui est une combinaison linéaire des trois variables originales, i.e.  $C = c_1 X_1 + c_2 X_2 + c_3 X_3$  pour des nombres réels  $c_1, c_2$  et  $c_3$ ) et qui capture le plus de variabilité possible. Trouvez les valeurs de  $c_1, c_2$  et  $c_3$  telles que  $\sum_{j=1}^3 c_j^2 = 1$  et que la variance de  $C$  sera maximale. Quelle est la valeur de cette variance maximale?
2. Quelle proportion de la variabilité totale de  $X_1, X_2$  et  $X_3$  est capturée par  $C$ ?
3. Retrouvez la matrice  $\Sigma$  à partir des informations données dans l'énoncé de la question.

## 4 Exercice 4 : Taux de survie de patients

La tableau suivant donne la répartition de patients selon l'issue observée en ligne (survie ou décès codés respectivement 0 et 1) en fonction de l'observation de troubles cognitifs en phase initiale en colonne (absence codée 0, présence de troubles codée 1 et troubles sévères codée 2).

	Absence	Troubles	Troubles sévères	Total
Survie	108	61	13	182
Décès	8	4	2	14
Total	116	65	15	196

1. Quelle méthode vous semble la mieux adaptée à l'analyse de ce tableau?
2. Donner le tableau des fréquences relatives.
3. Donner le profil moyen des lignes. À quoi correspond-il?
4. Donner le profil moyen des colonnes. À quoi correspond-il?

5. Combien d'axes obtient-on en effectuant l'analyse donnée en question 1 ? Justifier.
6. Quel test vous permet de détecter une éventuelle indépendance entre l'issue observée et la présence ou l'absence de troubles cognitifs en phase initiale ? Sachant que l'inertie totale du nuage est égale à 0.0050, donner la valeur de la statistique du test correspondant.

## 5 Exercice 5 : Théorie de l'ACM

1. Parmi les tableaux suivants, quels sont ceux qui peuvent être étudiés par une ACM ? Justifier.
  - Un tableau issu d'une enquête où toutes les questions engendrent des réponses qualitatives.
  - Un tableau où les individus répondent pour chaque question par "trop peu", "peu", "assez", "trop", "beaucoup trop".
  - Un tableau dans lequel on dénombre le nombre de personnes ayant répondu à deux questions particulières.
  - Un tableau contenant diverses caractéristiques d'individus (taille, poids, couleur des yeux, couleur des cheveux, etc.).
2. Quels sont les énoncés vrais dans le cadre de l'ACM ? Justifier.
  - Le tableau de données brut contient autant de colonnes que de variables.
  - Le tableau disjonctif complet contient autant de colonnes que de variables.
  - Le tableau de Burt contient autant de colonnes que de variables.
  - Le tableau disjonctif complet contient autant de colonnes que de modalités (toutes variables confondues).
3. Quels sont les énoncés vrais dans le cadre de l'ACM ? Justifier. La distance entre deux points représentant les individus  $i$  et  $j$  est
  - d'autant plus petite que ces deux individus prennent à peu près les mêmes modalités.
  - d'autant plus grande que les individus prennent des modalités différentes.
  - d'autant plus grande que les individus prennent des modalités différentes et que ces modalités sont fréquentes.
  - d'autant plus grande que les individus prennent des modalités différentes et que ces modalités sont rares.
4. Quels sont les énoncés vrais dans le cadre de l'ACM ? Justifier.
  - La distance entre deux points représentant les modalités  $k$  et  $l$  est d'autant plus petite que ces deux modalités ont en commun un grand nombre d'individus.
  - La distance entre deux points représentant les modalités  $k$  et  $l$  est d'autant plus grande que ces deux modalités sont rares.

- La distance entre deux modalités ayant de gros effectifs est nécessairement grande.
  - La distance entre deux modalités rares est nécessairement grande.
5. Quels sont les énoncés vrais dans le cadre de l'ACM ? Justifier.
- Le graphe biplot fournit une représentation optimale du nuage des individus.
  - Le graphe biplot fournit une représentation optimale du nuage des modalités.
  - Si un individu est proche d'une modalité alors il prend cette modalité.
  - Un individu est du côté des modalités qu'il a choisi.
  - Une modalité est du côté des individus qui ont choisi cette modalité.