TD4 : Variables couplées

Exercice 1

Nous observons p variables réelles $x^1, \ldots, x^p$ sur n individus pondérés par $p_1, \ldots, p_n > 0$ normalisés. Notons Σ la matrice de covariance et $\mathfrak{C}$ la matrice de corrélation,

$$\forall i, j \in \{1, \ldots, p\},\ \Sigma_{ij} = \mathrm{Cov}(x^i, x^j) \qquad \text{et} \qquad \mathfrak{C}_{ij} = \mathfrak{C}(x^i, x^j) = \frac{\mathrm{Cov}(x^i, x^j)}{\sqrt{\mathrm{Var}(x^i)\mathrm{Var}(x^j)}}\ .$$

Monter que Σ est symétrique et définie positive. En déduire que $\mathfrak{C}$ est aussi symétrique et définie positive.

Exercice 2

Nous étudions le tableau des notes de 6 élèves uniformément pondérés :

Français	Anglais	Maths	Sport
5	5	6	15
8	7	4	11
10	9	5	8
15	14	10	7
11	12	13	9
5	7	10	16

1. Calculer le centre de gravité et établir le tableau des données centrées.
2. Calculer la matrice de covariance des quatre variables.
3. Calculer l'inertie standard I.
4. Calculer les contributions des individus et les contributions des variables à I.

Exercice 3

Nous disposons des observations de 2 variables couplées x et y chez 4 individus pondérés par les poids normalisés $p_1 = p_2 = 1/6$ et $p_3 = p_4 = 1/3$. Voici les valeurs observées :

	x	y
Individu 1	-3	2
Individu 2	1	4
Individu 3	6	-1
Individu 4	4	-5

Les individus 1 et 2 forment le groupe A et les individus 3 et 4 forment le groupe B.

Pour toute variable z, nous noterons $\overline{z}$ la moyenne de z, $\overline{z}^A$ la moyenne de z dans le groupe A et $\overline{z}^B$ la moyenne de z dans le groupe B et $\sigma^2(z)$ la variance de z, $\sigma_A^2(z)$ la variance de z dans le groupe A et $\sigma_B^2(z)$ la variance de z dans le groupe B.

1. Calculer les moyennes $\overline{x}$, $\overline{x}^A$ et $\overline{x}^B$ ainsi que les moyennes $\overline{y}$, $\overline{y}^A$ et $\overline{y}^B$.
2. Calculer les variances $\sigma^2(x)$, $\sigma_A^2(x)$ et $\sigma_B^2(x)$ ainsi que les variances $\sigma^2(y)$, $\sigma_A^2(y)$ et $\sigma_B^2(y)$.

3. En déduire les valeurs des variances intra-groupes $\sigma^2_{intra}(x)$ et $\sigma^2_{intra}(y)$ et celles des variances inter-groupes $\sigma^2_{inter}(x)$ et $\sigma^2_{inter}(y)$ associées aux variables x et y respectivement.
4. Donner la valeur de l'inertie standard I associée aux variables x et y.
5. Vérifier sur cet exemple la formule de décomposition par groupes de l'inertie standard.

Exercice 4

Voici les observations de 6 individus uniformément pondérés pour 3 variables :

x^1	x^2	x^3
21	1	1
27	1	2
15	7	1
21	1	5
3	1	1
3	1	2

Calculer l'inertie et les contributions des individus à l'inertie avec

1. la distance euclidienne,
2. la distance des variables réduites.

Exercice 5 [Formule de décomposition de l'inertie]

Nous observons p variables quantitatives $x^1, \ldots, x^p$ sur n individus pondérés par les poids $p_1, \ldots, p_n$ normalisés. Les individus sont regroupés en N groupes $G_1, \ldots, G_N$ formant une partition de $\{1, \ldots, n\}$. Au sein du groupe G_k, on définit la moyenne et la variance de la variable x^j par

$$\bar{x}^j_k = \frac{1}{q_k} \sum_{i \in G_k} p_i x^j_i \qquad \text{et} \qquad \sigma^2_k(x^j) = \frac{1}{q_k} \sum_{i \in G_k} p_i \left(x^j_i - \bar{x}^j_k \right)^2$$

où $q_k = \sum_{i \in G_k} p_i$.

Préliminaire : rappeler comment nous avons obtenu la formule de décomposition de la variance d'une variable x^j

$$\begin{aligned} \mathrm{Var}(x^j) &= \sum_{k=1}^{N} q_k \left(\bar{x}^j_k - \bar{x}^j \right)^2 + \sum_{k=1}^{N} q_k \sigma^2_k(x^j) \\ &= \mathrm{Var}_{inter}(x^j) + \mathrm{Var}_{intra}(x^j) \ . \end{aligned}$$

Dans le cas de l'inertie standard, en déduire que nous avons

$$I = I^{inter} + I^{intra} = \sum_{j=1}^{p} \mathrm{Var}_{inter}(x^j) + \sum_{j=1}^{p} \mathrm{Var}_{intra}(x^j) \ .$$

Nous proposons ici de généraliser cette décomposition à l'inertie définie à partir d'une distance donnée par une matrice symétrique et définie positive. Nous considérons donc une telle matrice M de taille $p \times p$ et nous notons

$$\langle x, y \rangle_M = {}^t x M y, \ \forall x, y \in \mathbb{R}^p \ .$$

1. Rappeler comment on construit une distance d_M à partir de $\langle \cdot, \cdot \rangle_M$.

2. Que vaut l'inertie I_M définie à partir de cette distance d_M ?
3. Montrer la formule de décomposition de l'inertie

$$\begin{aligned} I_M &= I_M^{inter} + I_M^{intra} \\ &= \sum_{k=1}^{N} q_k d_M^2(g^{(k)}, g) + \sum_{k=1}^{N} q_k I_M^{(k)} \end{aligned}$$

   où $g^{(k)}$ est le centre de gravité des observations du groupe G_k et $I_M^{(k)}$ est l'inertie définie à partir de d_M calculée sur les individus de G_k.