Matrices et corrélations

Exercice 1
On note $\boldsymbol{X}$ un tableau de données de taille $(n \times p)$, les n individus étant en ligne, et $\boldsymbol{D}$ une matrice de poids de taille $(n \times n)$, et $\boldsymbol{M}$ une métrique de taille $(p \times p)$. Parmi les formes matricielles suivantes, indiquer celles qui ont un sens :

1. $\boldsymbol{X}\boldsymbol{X}^\top$
2. $\boldsymbol{X}\boldsymbol{M}\boldsymbol{X}^\top$
3. $(\boldsymbol{X}+\boldsymbol{X}^\top)\boldsymbol{D}$
4. $\boldsymbol{X}^\top\boldsymbol{D}$
5. $\boldsymbol{X}\boldsymbol{M}^{\frac{1}{2}}$
6. $(\boldsymbol{D}\boldsymbol{X})\boldsymbol{M}(\boldsymbol{D}\boldsymbol{X})^\top$
7. $\boldsymbol{D}(\boldsymbol{X}+\boldsymbol{X}^\top)$

Exercice 2
Calculer avec les relations présentées en cours le point moyen $\boldsymbol{m}_y$ du tableau centré suivant

$$\boldsymbol{Y} = \boldsymbol{X} - \boldsymbol{1}\,\boldsymbol{m}_x^\top$$

où $\boldsymbol{m}_x$ désigne le point moyen du tableau $\boldsymbol{X}$.
Démontrer la propriété suivante de la matrice de variance-covariance

$$\boldsymbol{\Sigma} = \boldsymbol{X}^\top\boldsymbol{D}\boldsymbol{X} - \boldsymbol{m}\boldsymbol{m}^\top = \boldsymbol{Y}^\top\boldsymbol{D}\boldsymbol{Y}$$

Exercice 3
On note $\boldsymbol{Z} = \boldsymbol{Y}\boldsymbol{D}_{1/\sigma}$, où $\boldsymbol{D}_{1/\sigma}$ est la matrice diagonale $(p \times p)$ de termes diagonaux $\frac{1}{\sigma_i}$.
Calculer la matrice de variance-covariance de $\boldsymbol{Z}$.

Exercice 4
En 1973, F. J. Anscombe a publié dans le numéro 27 de American Statistician un jeu de données très intéressantes pour montrer les pièges du calcul aveugle du coefficient de corrélation.

	x1	x2	x3	x4	y1	y2	y3	y4
1	10	10	10	8	8.04	9.14	7.46	6.58
2	8	8	8	8	6.95	8.14	6.77	5.76
3	13	13	13	8	7.58	8.74	12.74	7.71
4	9	9	9	8	8.81	8.77	7.11	8.84
5	11	11	11	8	8.33	9.26	7.81	8.47
6	14	14	14	8	9.96	8.10	8.84	7.04
7	6	6	6	8	7.24	6.13	6.08	5.25
8	4	4	4	19	4.26	3.10	5.39	12.50
9	12	12	12	8	10.84	9.13	8.15	5.56
10	7	7	7	8	4.82	7.26	6.42	7.91
11	5	5	5	8	5.68	4.74	5.73	6.89

1. Calculer la moyenne et la variance des 8 variables $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_4$ et $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_4$.

2. Calculer les coefficients de corrélation des couples $(\boldsymbol{x}_i, \boldsymbol{y}_i)$. Que note-t-on ?

3. Tracer la représentation des couples $(\boldsymbol{x}_i, \boldsymbol{y}_i)$. Commenter.

Exercice 5
Un magasin utilise pour ses produits une base de données indiquant en particulier pour chaque produit son prix de revient r et son prix de vente v. Plutôt que comparer les produits sur ces deux données, on s'intéresse au bénéfice sur chaque vente $v - r$. On cherche donc à construire une métrique qui mette en valeur ce bénéfice, c'est-à-dire telle que deux produits sur lesquels les bénéfices sont similaires soient considérés comme proches.

1. Montrer que l'on peut définir une matrice symétrique $\boldsymbol{M}$ de taille (2×2) telle que, pour chaque produit $\boldsymbol{x}^\top = (r \quad v)$,

$$\boldsymbol{x}^\top \boldsymbol{M} \boldsymbol{x} = (v - r)^2$$

   Est-ce que la matrice $\boldsymbol{M}$ définit une métrique ? Pourquoi ? En quoi est-ce un problème en pratique ?

2. Pour un nombre $a > 0$ donné, on considère la matrice

$$\boldsymbol{M}' = \boldsymbol{M} + \begin{pmatrix} a & 0 \\ 0 & 0 \end{pmatrix}$$

   Montrer que $\boldsymbol{M}'$ définit une métrique.