Examen d'Analyse des données

L'examen se compose de deux parties: la première de 55 mn est personnelle, la seconde de 1h15mn s'effectue en binôme et en salle des machines. Le cours, les feuilles de T.P. ainsi que tout autres documentations sur Splus sont autorisés.

Première partie

On considère des données présentant deux caractères quantitatifs X et Y de n individus de même poids. On notera $X = {}^t(x_1, .., x_n)$ et $Y = {}^t(y_1, .., y_n)$, où les x_i et y_i sont des réels. On suppose que X et Y ne sont pas centrées. On note aussi I_n la matrice identit d'ordre n.

1. Montrer que l'application linaire de matrice $\dfrac{X.{}^tX}{{}^tX.X}$ est un projecteur dont vous prciserez les lments. Même question pour la matrice $\left(I_n - \dfrac{X.{}^tX}{{}^tX.X}\right)$.

2. Dterminer a tel que $d_Y(a) = \displaystyle\sum_{i=1}^{n}(y_i - ax_i)^2$ soit minimale (en fonction des x_i et des y_i). Que vaut alors $d_Y(a)$?

3. Dterminer de même b tel que $d_X(b) = \displaystyle\sum_{i=1}^{n}(x_i - by_i)^2$ soit minimale. Que vaut alors $d_X(b)$ (comparer avec $d_Y(a)$)?

4. Dterminer α tel que $e_Y(\alpha, \beta, \gamma) = \displaystyle\sum_{i=1}^{n}(y_i - \alpha x_i^2 - \beta x_i - \gamma)^2$ soit minimale (en fonction des x_i et des y_i). Comparer alors $e_Y(\alpha, \beta, \gamma)$ et $d_Y(a)$.

Seconde partie

1. On considre la matrice **note**, compose des notes obtenues en mathmatiques (m), sciences (s), franais (f), langue (l) et lors d'un devoir la maison (discipline non prcise), par 9 lves. Indiquer les variables vous semblant fortement corrélées entre elles (justifier le numériquement). Faire une ACP de ce jeu de donnes (expliquer vos choix). Rajouter ces donnes l'individu PH dont les notes sont: 11 (m), 8 (s), 16 (f), 13 (l) et 11 (dm). Recommencer alors l'ACP et conclure.

2. On travaille sur le data.frame **consoprote**, dont une explication est donnée par **?consoprote**. Effectuer une ACP sur ce jeu de données (on précisera en justifiant les options et l'ordre de l'ACP, les contributions aux nouvelles variables, le nombre pertinent et la composition de groupes d'individus pouvant être mis en évidence).