

# Projet semestriel

## Consignes :

- **Résoudre** les différents exercices à l'aide du **logiciel R sous R-studio**
- Ecrire les **scripts** pour chaque traitement et préciser les fonctions utilisées
- Tous les résultats des fonctions à implémenter seront arrondis à  $10^{-2}$  près
- Présenter les résultats dans un **pdf** avec tous les tracés et tableaux de calcul
- **Commenter** et interpréter les résultats lors de la soutenance

## Exercice 1 : matrice diagonalisable

Soit la matrice

$$A = \begin{pmatrix} 0 & 1/2 & 1/2 \\ -2/3 & 1 & 2/3 \\ -1/3 & 1/2 & 5/6 \end{pmatrix}$$

1. Donner les valeurs propres de  $A$  ainsi que leur multiplicités. Pourquoi  $A$  est diagonalisable? Justifier.
2. Ecrire une fonction qui demande à l'utilisateur de saisir un entier naturel  $p$  non nul. Retourner une matrice random d'entiers compris entre 1 et 10, puis afficher toutes les valeurs propres ainsi que leurs multiplicités.
3. Donner les vecteurs propres de  $A$  pour chaque valeur propre.
4. Donner  $D$  la matrice des valeurs propres de  $A$  ainsi que  $P$  une matrice des vecteurs propres permettant d'effectuer le changement de bases entre  $A$  et  $D$ . Calculer  $P^{-1}$  et vérifier le changement de base  $A = PDP^{-1}$
5. On pose

$$\begin{cases} a_{n+1} = 0, 5b_n + 0, 5c_n \\ b_{n+1} = -(2/3)a_n + b_n + (2/3)c_n \\ c_{n+1} = -(1/3)a_n + 0, 5b_n + (5/6)c_n \end{cases}, \text{ avec } a_0 = 1, b_0 = 2, c_0 = 3$$

Ecrire une fonction qui demande de à l'utilisatuer de saisir un entier naturel non nul  $n$ . Puis, à l'aide du changement de base de la question 4, retourner les valeurs  $a_n, b_n, c_n$ . Que constate-t-on pour  $n \rightarrow \infty$ ?

## Eléments de cours : ensembles orthogonaux et projections orthogonales

Soit  $W$  un sous-espace de  $\mathbb{R}^n$ . Tout vecteur  $y$  de  $\mathbb{R}^n$  peut s'écrire de façon unique sous la forme

$$y = \hat{y} + z$$

avec  $\hat{y}$  appartient à  $W$  et  $z$  à  $W^\perp$ , le sous-espace de  $\mathbb{R}^n$ , orthogonal à  $W$ . Si  $\{u_1, \dots, u_p\}$  est une base orthogonale de  $W$ , on a

$$\hat{y} = \frac{y \bullet u_1}{u_1 \bullet u_1} u_1 + \dots + \frac{y \bullet u_p}{u_p \bullet u_p} u_p$$

Le vecteur  $\hat{y}$  est appelé projection orthogonale de  $y$  sur  $W$  et  $z = y - \hat{y}$ . Le vecteur  $\hat{y}$  définit le point de  $W$  le plus proche de  $y$ . Remarque : Si  $\{u_1, \dots, u_p\}$  est une base orthonormée de  $W$ , alors

$$\hat{y} = (y \bullet u_1)u_1 + \dots + (y \bullet u_p)u_p$$

Avec  $U = [u_1 \ u_2 \ \dots \ u_p]$ , on a :

$$\hat{y} = U(U^T)y$$

## Exercice 2 : étude de cas

Soit la matrice

$$A = \begin{pmatrix} -6 & -3 & 6 & 1 \\ -1 & 2 & 1 & -6 \\ 3 & 6 & 3 & -2 \\ 6 & -3 & 6 & -1 \\ 2 & -1 & 2 & 3 \\ -3 & 6 & 3 & 2 \\ -2 & -1 & 2 & -3 \\ 1 & 2 & 1 & 6 \end{pmatrix}$$

1. Montrer que les colonnes de  $A$  sont orthogonales. Préciser les calculs choisis
2. Construire la matrice  $U$  en normant chaque colonne de  $A$
3. Calculer  $U^T U$  et  $U U^T$ . En quoi diffèrent-elles ?
4. Choisir au hasard un vecteur  $y$  de  $\mathbb{R}^8$  et calculer  $p = U U^T y$  et  $z = y - p$ . Pourquoi  $p$  appartient à  $Col(A)$ , l'espace vectoriel engendré par les vecteurs colonnes de  $A$ . Vérifiez que  $z$  est orthogonal à  $p$ .
5. Vérifiez que  $z$  est orthogonal à chaque colonne de  $U$
6. Expliquez pourquoi  $z$  appartient à  $(Col(A))^\perp$
7. Déterminer le point de  $Col(U)$  le plus proche de  $y = (1, 1, 1, 1, 1, 1, 1, 1)$
8. Calculez la distance entre  $b = (1, 1, 1, 1, -1, -1, -1, -1)$  et  $Col(U)$

## Exercice 3 : Interpolation de données

Soit un nuage de points 2D :  $\Delta = \{(x_i, y_i), i = 0, \dots, n\}$ . On suppose que  $x_0 < x_1 < \dots < x_n$  sont des entiers relatifs triés et  $y_0, y_1, \dots, y_n$  des entiers relatifs quelconques

On souhaite interpoler  $\Delta$  par un polynôme de degré inférieure ou égal à  $n$ .

1. Créer une fonction qui génère graphiquement  $\Delta$  aléatoirement tel que l'utilisateur saisisse le nombre de points et  $a, b, c, d$  pour que  $\forall i = 0, \dots, n, a \leq x_i \leq b, c \leq y_i \leq d$ . Proposer un exemple.
2. Donner la matrice de vandermonde permettant une première interpolation polynomiale de  $\Delta$ , avec  $n = 9$ , puis  $n = 19$  et enfin  $n = 29$ . Que constate-t-on ? Retourner et tracer si possible le polynôme d'interpolation.
3. Donner la matrice de Newton permettant une seconde interpolation polynomiale de  $\Delta$ , avec  $n = 9$ , puis  $n = 19$  et enfin  $n = 29$ . Comparer avec la méthode précédente. Retourner et tracer si possible le polynôme d'interpolation.
4. Toujours, pour l'interpolation de Newton, implémenter le tableau des différences divisées afin de calculer et tracer le polynôme d'interpolation.
5. Appliquer les fonctions des questions 2, 3 et 4 à

$$\Delta = \{(0, 2), (1, 6), (2, 12), (3, 20), (4, 30), (5, 42), (6, 56), (7, 72), (8, 90), (9, 110)\}$$

Comparer les résultats et donner le polynôme d'interpolation  $P$  dans la base canonique des polynômes. Prévion : calculer  $P(10), P(15)$  et  $P(20)$

## Exercice 4 : Palindromes

1. Créer une fonction qui identifie les mots ou phrases palindromiques (mots ou phrases qui se lisent dans les deux sens, par exemple RADAR). Cette fonction prendra comme argument le mot "mot" et retournera les phrases : "mot est un palindrome" si le mot = palindrome et "mot n'est pas un palindrome", sinon.
2. Appliquer votre fonction sur les mots "radar", "bonne année", "sept", "kayak", "la mariée ira mal", "statistiques", "engage le jeu que je le gagne", "esope reste ici et se repose".
3. Créer un dictionnaire de 8000 mots aléatoires ayant un nombre de lettres compris entre 2 et 9. Mettre 1000 mots de 2 lettres, 1000 mots de 3 lettres, ... ainsi de suite
4. Retourner tous les mots palindromiques de ce dictionnaire.

## Exercice 5 : ACP

Voici les données des performances d'athlètes ayant participé à des épreuves du décathlon aux Jeux Olympiques et au Décastar (compétition annuelle d'athlétisme se tenant au Stade de Thouars de Talence, en Gironde)

	c100	long	poids	haut	c400	c110	disq	perche	javel	c1500	RANG	POINTS	COMPET
SEBRLE	11.0	7.58	14.8	2.07	49.8	14.7	43.8	5.02	63.2	292	1	8217	DS
CLAY	10.8	7.40	14.3	1.86	49.4	14.0	50.7	4.92	60.2	302	2	8122	DS
KARPOV	11.0	7.30	14.8	2.04	48.4	14.1	49.0	4.92	50.3	300	3	8099	DS
BERNARD	11.0	7.23	14.2	1.92	48.9	15.0	40.9	5.32	62.8	280	4	8067	DS
YURKOV	11.3	7.09	15.2	2.10	50.4	15.3	46.3	4.72	63.4	276	5	8036	DS
WARNERS	11.1	7.60	14.3	1.98	48.7	14.2	41.1	4.92	51.8	278	6	8030	DS
ZSIVOCZKY	11.1	7.30	13.5	2.01	48.6	14.2	45.7	4.42	55.4	268	7	8004	DS
McMULLEN	10.8	7.31	13.8	2.13	49.9	14.4	44.4	4.42	56.4	285	8	7995	DS
MARTINEAU	11.6	6.81	14.6	1.95	50.1	14.9	47.6	4.92	52.3	262	9	7802	DS
HERNU	11.4	7.56	14.4	1.86	51.1	15.1	45.0	4.82	57.2	285	10	7733	DS
BARRAS	11.3	6.97	14.1	1.95	49.5	14.5	42.1	4.72	55.4	282	11	7708	DS
NOOL	11.3	7.27	12.7	1.98	49.2	15.3	37.9	4.62	57.4	267	12	7651	DS
BOURGUIGNON	11.4	6.80	13.5	1.86	51.2	15.7	40.5	5.02	54.7	292	13	7313	DS
Sebrle	10.8	7.84	16.4	2.12	48.4	14.0	48.7	5.00	70.5	280	1	8893	JO
Clay	10.4	7.96	15.2	2.06	49.2	14.1	50.1	4.90	69.7	282	2	8820	JO
Karpov	10.5	7.81	15.9	2.09	46.8	14.0	51.6	4.60	55.5	278	3	8725	JO
Macey	10.9	7.47	15.7	2.15	49.0	14.6	48.3	4.40	58.5	265	4	8414	JO
Warners	10.6	7.74	14.5	1.97	48.0	14.0	43.7	4.90	55.4	278	5	8343	JO
Zsivoczky	10.9	7.14	15.3	2.12	49.4	15.0	45.6	4.70	63.4	270	6	8287	JO
Hernu	11.0	7.19	14.6	2.03	48.7	14.2	44.7	4.80	57.8	264	7	8237	JO
Nool	10.8	7.53	14.3	1.88	48.8	14.8	42.0	5.40	61.3	276	8	8235	JO
Bernard	10.7	7.48	14.8	2.12	49.1	14.2	44.8	4.40	55.3	276	9	8225	JO
Schwarzl	11.0	7.49	14.0	1.94	49.8	14.2	42.4	5.10	56.3	274	10	8102	JO
Pogorelov	11.0	7.31	15.1	2.06	50.8	14.2	44.6	5.00	53.4	288	11	8084	JO
Schoenbeck	10.9	7.30	14.8	1.88	50.3	14.3	44.4	5.00	60.9	279	12	8077	JO
Barras	11.1	6.99	14.9	1.94	49.4	14.4	44.8	4.60	64.6	267	13	8067	JO
Smith	10.8	6.81	15.2	1.91	49.3	14.0	49.0	4.20	61.5	273	14	8023	JO
Averyanov	10.6	7.34	14.4	1.94	49.7	14.4	39.9	4.80	54.5	271	15	8021	JO
Ojaniemi	10.7	7.50	15.0	1.94	49.1	15.0	40.4	4.60	59.3	276	16	8006	JO
Smirnov	10.9	7.07	13.9	1.94	49.1	14.8	42.5	4.70	60.9	263	17	7993	JO
Qi	11.1	7.34	13.6	1.97	49.6	14.8	45.1	4.50	60.8	273	18	7934	JO
Drews	10.9	7.38	13.1	1.88	48.5	14.0	40.1	5.00	51.5	274	19	7926	JO
Parkhomenko	11.1	6.61	15.7	2.03	51.0	14.9	41.9	4.80	65.8	278	20	7918	JO
Terek	10.9	6.94	15.2	1.94	49.6	15.1	45.6	5.30	50.6	290	21	7893	JO
Gomez	11.1	7.26	14.6	1.85	48.6	14.4	41.0	4.40	60.7	270	22	7865	JO
Turi	11.1	6.91	13.6	2.03	51.7	14.3	39.8	4.80	59.3	290	23	7708	JO
Lorenzo	11.1	7.03	13.2	1.85	49.3	15.4	40.2	4.50	58.4	263	24	7592	JO
Karlivans	11.3	7.26	13.3	1.97	50.5	15.0	43.3	4.50	52.9	279	25	7583	JO
Korkizoglou	10.9	7.07	14.8	1.94	51.2	15.0	46.1	4.70	53.0	317	26	7573	JO
Uldal	11.2	6.99	13.5	1.85	51.0	15.1	43.0	4.50	60.0	282	27	7495	JO
Casarsa	11.4	6.68	14.9	1.94	53.2	15.4	48.7	4.40	58.6	296	28	7404	JO

Les dix épreuves du décathlon :

- course sur 100 m (c100),
- saut en longueur (long),
- lancer de poids (poids),
- saut en hauteur (haut),
- course sur 400 m (c400),
- course de haies sur 110 m (c110),
- lancer de disque (disq),
- saut à la perche (perch),
- lancer de javelot (javel),
- course sur 1500 m (c1500)

Autres variables

- rang de classement (RANG),
- nombre de points (POINTS),
- compétition (COMPET),
  - Jeux Olympiques (JO),
  - Décastar (DS)

**Attention !** Les noms des participants sont en majuscule pour le Décastar, afin de permettre de différencier les participations d'un même athlète aux deux épreuves.  
(exemple : SERBLE/Serble).

### A. Analyse rapide

- Récupérer les données du fichier "decathlon" et donner la matrice corrélation des variables quantitatives (ne pas prendre COMPET)
- Quelles sont les couples de variables les plus corrélées, les moins corrélées, les plus opposées ? Justifier.
- Comment se groupent les variables du point de vue des signes de corrélation ? Expliquez pourquoi.

### B. ACP : dans cette partie, vous allez procéder à une analyse en composantes principales des performances centrées-réduites, en excluant les variables RANG, POINTS et COMPET.

- Donner les valeurs propres de la matrice de corrélation. Trier ces valeurs propres et donner le nombre de vecteurs propres qui expliquent le plus l'inertie du nuage des individus. Quelle

règle peut-on utiliser ? Donner le pourcentage d'inertie totale en conservant les trois premiers vecteurs propres.

5. Déterminer les trois composantes principales (projection des individus sur les trois vecteurs propres), que l'on note  $C1, C2, C3$  dans l'ordre décroissant d'inertie
6. Déterminer le tableau des corrélations des variables par rapport à  $C1, C2, C3$  et donner les deux cercles de corrélation des variables par rapport à  $(C1, C2)$  et  $(C2, C3)$
7. Quelles sont les variables qui déterminent les 3 composantes principales ? Proposez un seuil.
8. Expliquez comment les données peuvent être modifiées pour faire apparaître un effet de taille. Comment peut-on alors interpréter les axes principaux de la question 5 ?