

Quelques éléments d'algèbre pour l'analyse de données

A.B. Dufour, D. Chessel et M. Royer

L'algèbre matricielle est fondamentale dans la compréhension de la théorie liée à l'analyse exploratoire des données.

Table des matières

1	Moyennes et variances	2
1.1	Définitions élémentaires	2
1.2	Pondérations	2
2	Produits scalaires	4
3	Longueur, angle et distance	6
3.1	Norme d'un vecteur	6
3.2	Angle entre 2 vecteurs	7
3.3	Distance entre 2 vecteurs	7
4	Exercices d'application	7
4.1	Dépenses journalières	7
4.2	Consommation d'alcool	9
5	Réduction des endomorphismes	9
5.1	Quelques généralités	9
5.2	Exemple	10

1 Moyennes et variances

L'analyse des données manipule des variables quantitatives ou qualitatives mesurées ou observées sur n individus. On commence ici par les variables quantitatives, celles dont les valeurs appartiennent à \mathbb{R} .

Une variable $\mathbf{x} = (x_1, x_2, \dots, x_n)$ est donc un vecteur de \mathbb{R}^n .

Exercice

On désire implanter une station-service entre Lyon et Lausanne. On récupère les prix de 10 stations avoisinantes. Les données se trouvent dans le fichier `essence.txt`.

(cf site pédagogique <http://pbil.univ-lyon1.fr/R/enseignement.html>)

1. Importer le fichier dans votre répertoire de travail.
2. Afficher les prix du gazole.

1.1 Définitions élémentaires

La moyenne descriptive de \mathbf{x} est $m(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n x_i$. On note souvent $m(\mathbf{x}) = \bar{\mathbf{x}}$. La variance descriptive de \mathbf{x} est $v(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n (x_i - m(\mathbf{x}))^2$. On peut utiliser la notation $v(\mathbf{x}) = s_{\mathbf{x}}^2$.

L'écart-type descriptif de \mathbf{x} est $s_{\mathbf{x}} = \sqrt{v(\mathbf{x})}$.

Ces paramètres caractérisent la position (moyenne) et la dispersion (variance) des mesures. La médiane, les quartiles sont d'autres paramètres de position; l'intervalle interquartile est un autre paramètre de dispersion. Ces paramètres décrivent la variable vue comme n points de \mathbb{R} .

Exercice 1

1. Ecrire la fonction permettant de calculer la moyenne. Calculer la moyenne de la variable `Gazole`. Comparer le résultat à `mean`.
2. Ecrire la fonction permettant de calculer la variance. Calculer la variance de la variable `Gazole`. Comparer le résultat à `var`.
3. Ecrire la fonction permettant de calculer l'écart-type. Calculer l'écart-type de la variable `Gazole`. Comparer le résultat à `sd`.

Exercice 2

1. Calculer la moyenne et la variance descriptives de $\mathbf{1}_n = (1, 1, \dots, 1)$ (a) en posant $n = 10$, à l'aide de `R`, (b) en généralisant.
2. Calculer la moyenne et la variance descriptives de $\mathbf{x} = (1, 2, \dots, n)$ (a) en posant $n = 10$ à l'aide de `R`, (b) en généralisant.

1.2 Pondérations

Une pondération des n individus est un vecteur de \mathbb{R}^n dont toutes les composantes sont positives et dont la somme vaut 1. Une pondération est notée :

$$\mathbf{p} = (p_1, p_2, \dots, p_n) \text{ avec } \sum_{i=1}^n p_i = 1 \text{ et } 1 \leq i \leq n \Rightarrow p_i > 0$$

La moyenne pondérée de \mathbf{x} est : $m_{\mathbf{p}} = m_{\mathbf{p}}(\mathbf{x}) = \sum_{i=1}^n p_i x_i$.

La variance pondérée de \mathbf{x} est : $v_{\mathbf{p}} = v_{\mathbf{p}}(\mathbf{x}) = \sum_{i=1}^n p_i (x_i - m_{\mathbf{p}}(x))^2$.

L'écart-type pondéré de \mathbf{x} est $\sqrt{v_{\mathbf{p}}(\mathbf{x})}$.

Exemple

Huit étudiants (3 hommes et 5 femmes) se sont retrouvés autour d'un apéritif chez Alain, un autre étudiant et lui ont attribué une note (sur 10). On désire une pondération sur les étudiants invités tels que l'on ait un poids global de $\frac{1}{2}$ pour les hommes et de $\frac{1}{2}$ pour les femmes. Le vecteur de pondération est donc : $3p_H = \frac{1}{2}$ et $5p_F = \frac{1}{2}$

```
pH <- 1/(3*2)
pF <- 1/(5*2)
round(pH,4)
[1] 0.1667
round(pF,4)
[1] 0.1
poids <- rep(c(pH,pF),c(3,5))
sum(poids)
[1] 1
```

En considérant le vecteur `Notes` de ces étudiants invités, on calcule la moyenne, la variance et l'écart-type pondérés.

```
Notes <- c(8,10,4,8,6,7,7,9)
moyN <- sum(poids*Notes)
varN <- sum(poids*(Notes-moyN)^2)
sdN <- sqrt(varN)
moyN
[1] 7.366667
varN
[1] 3.632222
sdN
[1] 1.905839
```

Exercice

On souhaite implanter la station au bord de la nationale. Cependant les concurrents ne sont pas tous implantés dans le même contexte géographique. On accorde plus d'importance aux prix des stations (1, 3 et 7) qui se situent également sur une nationale soit un poids 2 fois supérieur.

Calculer la moyenne, la variance et l'écart-type pondérés de la variable `Gazole`.

La moyenne (resp. variance) descriptive est le cas particulier de la moyenne (resp. variance) pondérée pour la pondération uniforme définie par :

$$p_i = \frac{1}{n}.$$

Quand aucune ambiguïté n'est possible, on note simplement $m_{\mathbf{p}}(\mathbf{x}) = m(\mathbf{x})$ et $v_{\mathbf{p}}(\mathbf{x}) = v(\mathbf{x})$.

2 Produits scalaires

Un vecteur de l'espace vectoriel \mathbb{R}^n est un n -uplet de nombres réels, soit $\mathbf{x} = (x_1, x_2, \dots, x_n)$.

Etant donnés $\omega_1, \omega_2, \dots, \omega_n$, n nombres strictement positifs, on appelle ω -produit scalaire diagonal associé à $\omega = (\omega_i)_{1 \leq i \leq n}$ l'application qui, à un couple (\mathbf{x}, \mathbf{y}) de points de \mathbb{R}^n , associe le nombre réel :

$$\langle \mathbf{x} | \mathbf{y} \rangle_\omega = \sum_{i=1}^n \omega_i x_i y_i.$$

L'application ω -produit scalaire vérifie les propriétés :

- ★ PS1 $\langle \mathbf{x} | \mathbf{y} \rangle_\omega = \langle \mathbf{y} | \mathbf{x} \rangle_\omega$ pour tout \mathbf{x} et \mathbf{y} de \mathbb{R}^n ,
- ★ PS2a $\langle \mathbf{x} | \mathbf{y} + \mathbf{z} \rangle_\omega = \langle \mathbf{x} | \mathbf{y} \rangle_\omega + \langle \mathbf{x} | \mathbf{z} \rangle_\omega$ pour tout $\mathbf{x}, \mathbf{y}, \mathbf{z}$ de \mathbb{R}^n ,
- ★ PS2b $\langle \mathbf{x} | \alpha \mathbf{y} \rangle_\omega = \alpha \langle \mathbf{x} | \mathbf{y} \rangle_\omega$ pour tout \mathbf{x} et \mathbf{y} de \mathbb{R}^n et pour tout α de \mathbb{R} ,
- ★ PS3 $\langle \mathbf{x} | \mathbf{x} \rangle_\omega \geq 0$ pour tout \mathbf{x} de \mathbb{R}^n ,
- ★ PS4 $\langle \mathbf{x} | \mathbf{x} \rangle_\omega = 0 \Rightarrow \mathbf{x} = \mathbf{0}_n = (0, 0, \dots, 0)$ pour tout \mathbf{x} de \mathbb{R}^n .

Plus généralement, on considère une fonction Φ de $\mathbb{R}^n \times \mathbb{R}^n$ dans \mathbb{R} qui, à un couple de points (\mathbf{x}, \mathbf{y}) , associe un nombre réel noté indifféremment

$$\langle \mathbf{x} | \mathbf{y} \rangle_\Phi \text{ ou } \Phi(\mathbf{x}, \mathbf{y}).$$

On dit que c'est un produit scalaire si elle vérifie les propriétés PS1, ..., PS4.

Exemple

Alain a invité à nouveau les étudiants la semaine d'après. Soient $\mathbf{x}^T = (8 \ 10 \ 4 \ 8 \ 6 \ 7 \ 7 \ 9)$ et $\mathbf{y}^T = (9 \ 8 \ 7 \ 6 \ 4 \ 7 \ 9 \ 10)$, les deux vecteurs **Notes** obtenus. \mathbf{x} et \mathbf{y} sont deux vecteurs de \mathbb{R}^8 .

Chacun des vecteurs est muni de la pondération uniforme $\omega_i = \frac{1}{8}$.

Le produit scalaire sous \mathbb{R} est :

```
x <- c(8,10,4,8,6,7,7,9)
y <- c(9,8,7,6,4,7,9,10)
omega <- rep(1/8,8)
prodscal <- sum(omega*x*y)
prodscal
[1] 56.75
```

L'écriture matricielle du produit scalaire est donnée par $\mathbf{x}^T \mathbf{D} \mathbf{y}$ où \mathbf{D} est la matrice contenant les pondérations ω_i sur la première diagonale et des zéros partout ailleurs.

```
D <- diag(omega)
prodscal <- t(x) %*% D %*% y
prodscal
      [,1]
[1,] 56.75
```

On vérifie ainsi les propriétés du produit scalaire. Soient $\mathbf{z}^T = (8 \ 4 \ 8 \ 10 \ 7 \ 6 \ 8 \ 10)$ un autre vecteur de \mathbb{R}^8 et $\alpha = 2$ un scalaire quelconque.

```
z <- c(8,4,8,10,7,6,8,10)
alpha <- 2
```

Propriété 1 :

```
t(x) %*% D %*% y
      [,1]
[1,] 56.75
t(y) %*% D %*% x
      [,1]
[1,] 56.75
t(x) %*% D %*% y == t(y) %*% D %*% x
      [,1]
[1,] TRUE
```

Propriété 2a :

```
t(x) %*% D %*% (y+z)
      [,1]
[1,] 112.5
t(x) %*% D %*% y + t(x) %*% D %*% z
      [,1]
[1,] 112.5
```

Propriété 2b :

```
t(x) %*% D %*% (alpha * y)
      [,1]
[1,] 113.5
alpha * (t(x) %*% D %*% y)
      [,1]
[1,] 113.5
```

Propriété 3 :

```
t(x) %*% D %*% x >= 0
      [,1]
[1,] TRUE
```

Exercice

Pour établir les prix, on prend en compte le prix du Sans Plomb 95 (SP95) et du Sans Plomb 98 (SP98).

La matrice diagonale des poids (cf résultats de l'exercice page 3) **D** va permettre de définir une métrique.

1. Calculer le produit scalaire du **Gazole** avec le carburant sans plomb SP95 puis avec le carburant sans plomb SP98.
2. Trouver une solution pour réaliser les calculs en une seule fois.

Les produits scalaires sont des fonctions *BSPND* (*B*ilinaires, *S*ymétriques, *P*ositives et *N*on *D*égénérés).

Quand on manipule un seul produit scalaire, si aucune confusion n'est possible, on note simplement $\langle \mathbf{x} | \mathbf{y} \rangle_{\Phi} = \langle \mathbf{x} | \mathbf{y} \rangle$.

3 Longueur, angle et distance

3.1 Norme d'un vecteur

Un produit scalaire permet de mesurer la longueur d'un vecteur et l'angle de deux vecteurs. La Φ -norme associée à un produit scalaire est définie par :

$$\|\mathbf{x}\|_{\Phi} = \sqrt{\langle \mathbf{x} | \mathbf{x} \rangle_{\Phi}}.$$

On a $\|\alpha \mathbf{x}\|_{\Phi} = |\alpha| \|\mathbf{x}\|_{\Phi}$. Quand aucune confusion est possible, on note simplement $\|\mathbf{x}\|_{\Phi} = \|\mathbf{x}\|$.

Exemple

On reprend les vecteurs \mathbf{x} , \mathbf{y} et \mathbf{z} munis cette fois, des pondérations uniformes. On calcule la norme de chaque vecteur.

```
omega <- rep(1/8,8)
norme <- function(x,pds) sqrt(sum(pds*x*x))
norme(x,omega)
[1] 7.574629
norme(y,omega)
[1] 7.713624
norme(z,omega)
[1] 7.850159
```

Exercice

Calculer la norme de chacun des carburants.

Projection sur un vecteur

Si \mathbf{x} et \mathbf{y} sont deux vecteurs de \mathbb{R}^n et si \mathbf{x} est non nul, il existe un unique vecteur \mathbf{z} de \mathbb{R}^n proportionnel à \mathbf{x} tel que $\mathbf{y} - \mathbf{z}$ soit orthogonal à \mathbf{x} .

On dit que \mathbf{z} est le projeté Φ -orthogonal de \mathbf{y} sur \mathbf{x} . Il vaut :

$$\mathbf{z} = \frac{\langle \mathbf{x} | \mathbf{y} \rangle_{\Phi}}{\langle \mathbf{x} | \mathbf{x} \rangle_{\Phi}} \mathbf{x}.$$

Exemple

Donner le coefficient de la projection orthogonale de \mathbf{y} sur \mathbf{x} .

```
sum(omega*x*y)/sum(omega*x*x)
[1] 0.9891068
```

Exercice

1. Calculer la projection du prix du gazole sur le vecteur $\mathbf{1}_n$.
2. On note \mathbf{g} le prix du gazole et \mathbf{z} le vecteur de cette projection. Ecrire le vecteur $\mathbf{g} - \mathbf{z}$. Calculer sa norme. Que reconnait-on ?

3.2 Angle entre 2 vecteurs

Deux vecteurs de \mathbb{R}^n sont Φ -orthogonaux si et seulement si $\langle \mathbf{x} | \mathbf{y} \rangle_\Phi = 0$.

Si \mathbf{x} et \mathbf{y} sont 2 points de \mathbb{R}^n , la Φ -mesure de l'angle de \mathbf{x} et \mathbf{y} est notée

$$A_\Phi(\mathbf{x}, \mathbf{y}) = a$$

Elle est définie par :

$$0 \leq a \leq \pi \text{ et } \cos a = \frac{\langle \mathbf{x} | \mathbf{y} \rangle_\Phi}{\|\mathbf{x}\|_\Phi \|\mathbf{y}\|_\Phi}.$$

Exemple

L'angle entre les deux vecteurs \mathbf{x} et \mathbf{y} est donné par la relation :

```
sum(omega*x*y)/(norme(x,omega)*norme(y,omega))
[1] 0.9712836
cosangle <- sum(omega*x*y)/(norme(x,omega)*norme(y,omega))
acos(cosangle)
[1] 0.2402289
```

Exercice

Calculer le cosinus de l'angle entre le Gazole et les carburants sans plomb. Que constate-t-on ?

3.3 Distance entre 2 vecteurs

La Φ -distance de deux vecteurs est définie par :

$$d_\Phi(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|_\Phi = \|\mathbf{y} - \mathbf{x}\|_\Phi.$$

On sait donc mesurer les angles et les distances entre points de \mathbb{R}^n au sens d'un produit scalaire donné.

Exemple

```
norme(x-y, omega)
[1] 1.837117
```


Exercice

Calculer les écarts entre les prix des carburants. Qu'observe-t-on ?

4 Exercices d'application

4.1 Dépenses journalières

On donne les dépenses journalières moyennes de 6 individus (en euros) pour divers produits : le transport, la boisson, la nourriture, les journaux et les cigarettes. Les données se trouvent dans le fichier `conso.txt`.

- 1) Charger le fichier dans votre dossier de travail. Afficher les données sous  et donner la nature de l'objet.

- 2) On suppose que toutes les variables ont la même importance ; on munit donc l'espace des variables de la métrique uniforme \mathbf{U} .
 - a) Calculer la norme de la dépense de chaque individu ; mettre les résultats dans la variable \mathbf{nU} . Qui consomme le plus ?
 - b) Calculer la distance entre les individus.
- 3) On estime que le Transport et la Nourriture sont des dépenses plus importantes (vitales). Les autres dépenses seront considérées comme annexes, facultatives. On estime donc qu'elles pèsent plus sur le portefeuille et on augmente le poids de ces variables.

Du coup, seront très différents deux individus qui diffèrent par rapport à ces variables.

Construire la métrique, notée \mathbf{D} qui donne aux dépenses de transport et nourriture un poids 5 fois moins important que pour les autres dépenses.

 - a) Calculer la norme de la dépense de chaque individu ; mettre les résultats dans la variable \mathbf{nD} . Qui consomme le plus ? Comparer avec les valeurs de \mathbf{nU} .
 - b) Calculer la distance entre les individus.
- 4) On souhaite comparer deux individus au sens de la répartition de leurs dépenses. Pour cela, on va calculer l'angle entre les individus pour la métrique uniforme. La valeur de la dépense totale journalière ne joue pas de rôle (on pourrait aussi diviser chaque ligne par la somme de la ligne en question).
 - a) Faire le calcul des cosinus des angles.

2 vecteurs de $\cos = 1$ sont proportionnels. Ils représentent des individus qui ont même comportement au sens des répartitions dans les variables.

2 vecteurs orthogonaux ont un $\cos = 0$. Comportement de logique différente.
 - b) Commenter le cas des individus 5 et 6.
- 5) On introduit une métrique uniforme \mathbf{V} sur les individus.
 - a) Calculer la norme de chaque variable, qu'on mettra dans le vecteur \mathbf{nV} . Quel est le produit le plus consommé ?
 - b) Calculer la variance des variables, qu'on mettra dans le vecteur \mathbf{varV} . Pour quel produit le comportement des consommateurs est le plus homogène ? le moins homogène ?
 - c) Recalculer la norme des dépenses des individus après centrage des variables, sous la métrique uniforme \mathbf{U} (vecteur \mathbf{nUc}). Recalculer aussi la distance entre les individus dans ce contexte (matrice \mathbf{dUc}). Commentaires.
 - d) On va maintenant réduire les variables. Calculer la norme des produits réduits (vecteur \mathbf{nVr}).
 - e) Faut-il d'abord réduire puis centrer les variables ou le contraire ?

4.2 Consommation d'alcool

Utilisons le jeu de données sur la consommation d'alcool dans les différents pays d'Europe entre 1961 et 1999. Ce tableau de données croise 20 pays et 39 années pour la consommation individuelle de bière, de vin et de spiritueux.

Chaque pays, chaque année présente un profil moyen de consommation : chaque valeur est une consommation annuelle moyenne en litres d'alcool, par habitant pour les 3 types d'alcool.

```
load("pps066.rda")
```

On a calculé la variance de chaque produit consommé, ce qui donne le résultat suivant :

```
apply(pps066$beer,2,vardes) -> beer
apply(pps066$spir,2,vardes) -> spir
apply(pps066$wine,2,vardes) -> wine
cbind(beer,spir,wine) -> varconso
t(varconso)
```

	1961	1962	1963	1964	1965	1966	1967	1968
beer	5.078029	5.362275	5.820570	6.528349	6.360501	6.50979	6.170965	6.187269
spir	2.887225	1.307075	2.419223	2.223434	2.515511	2.52144	1.889813	1.901719
wine	30.616925	30.449283	34.465155	27.759134	28.020495	31.02768	31.100080	32.841375
	1969	1970	1971	1972	1973	1974	1975	1976
beer	6.520599	6.810506	7.360501	7.299389	7.291390	6.867456	6.965456	6.905059
spir	2.284223	1.377185	2.418575	1.711759	1.858406	2.382285	2.451375	2.434479
wine	33.279039	28.121133	33.467994	28.027136	27.542353	27.889393	26.375435	27.218159
	1977	1978	1979	1980	1981	1982	1983	1984
beer	6.533169	5.711869	5.739933	5.944695	5.612145	5.930395	5.775386	5.411035
spir	2.617863	2.383000	2.115301	2.435573	2.086311	2.107253	2.332979	2.378059
wine	24.106935	19.860206	18.274235	19.316115	18.514769	16.473569	14.959823	14.056925
	1985	1986	1987	1988	1989	1990	1991	1992
beer	4.739285	4.670879	4.248801	4.060270	3.992859	4.491750	4.131633	4.386329
spir	2.658263	2.420453	2.115564	1.925086	2.200699	1.646711	1.676949	1.453709
wine	13.153635	11.281599	10.782761	9.918441	10.152984	10.026599	9.378003	9.250765
	1993	1994	1995	1996	1997	1998	1999	
beer	4.021089	4.100759	4.114204	4.219726	4.380503	3.912845	3.844851	
spir	1.463835	1.655774	1.622696	1.569093	1.591645	1.626713	1.678379	
wine	9.573604	9.475615	9.942104	9.711694	10.065525	11.433483	11.121715	

- 1) Quelle information obtient-on ?
- 2) Donner une représentation graphique de l'évolution de la variance pour chaque type d'alcool.
- 3) Que se passera-t-il si on réduit les variables ?

5 Réduction des endomorphismes

5.1 Quelques généralités

L'analyse exploratoire des données peut être considérée comme la recherche de bases dans lesquelles la forme de la matrice est la plus simple possible c'est-à-dire diagonales voire triangulaires.

La clef de la diagonalisation est la notion de **vecteur propre**. On dit que \mathbf{v} est vecteur propre si et seulement si :

1. $\mathbf{v} \neq 0$
2. $\exists \lambda \in \mathbb{R}, \quad \mathbf{X}\mathbf{v} = \lambda\mathbf{v}$

Le scalaire λ est dit **valeur propre** associée à \mathbf{v} .

5.2 Exemple

Soit la matrice $\mathbf{X} = \begin{pmatrix} 2 & 0 & 4 \\ 3 & -4 & 12 \\ 1 & -2 & 5 \end{pmatrix}$

Son polynôme caractéristique est donné par :

$$|\mathbf{X} - \lambda \mathbf{I}_3| = 0 \text{ soit } \lambda(\lambda - 1)(2 - \lambda) = 0$$

On obtient trois valeurs propres que l'on classe par ordre décroissant : $\lambda_1 = 2$, $\lambda_2 = 1$ et $\lambda_3 = 0$.

La recherche du vecteur propre \mathbf{v}_1 associée à λ_1 conduit au vecteur $\begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix}$, à un facteur de proportionalité près.

La recherche du vecteur propre \mathbf{v}_2 associée à λ_2 conduit au vecteur $\begin{pmatrix} -4 \\ 0 \\ 1 \end{pmatrix}$, à un facteur de proportionalité près.

La recherche du vecteur propre \mathbf{v}_3 associée à λ_3 conduit au vecteur $\begin{pmatrix} -4 \\ 3 \\ 2 \end{pmatrix}$, à un facteur de proportionalité près.

Sous \mathbb{R} , il existe une procédure automatique de recherche des valeurs et vecteurs propres : `eigen`.

```
matX <- matrix(c(2,3,1,0,-4,-2,4,12,5),ncol=3)
matX
      [,1] [,2] [,3]
[1,]    2    0    4
[2,]    3   -4   12
[3,]    1   -2    5

eigen(matX)

$values
[1] 2.000000e+00 1.000000e+00 5.093148e-15
$vectors
      [,1] [,2] [,3]
[1,] 8.944272e-01 -9.701425e-01 0.7427814
[2,] 4.472136e-01 3.523360e-15 -0.5570860
[3,] -1.119535e-16 2.425356e-01 -0.3713907

res <- eigen(matX)
```

Exercice 1

1. Comparer les vecteurs propres obtenus dans les deux procédures.
2. Calculer les normes de \mathbf{v}_1 , \mathbf{v}_2 , \mathbf{v}_3
3. Calculer les normes des trois vecteurs obtenus avec la procédure `eigen`.
4. Conclure.

Exercice 2

On note \mathbf{X} la matrice associée aux données `essence`. Elle est de dimension 10×3 .

1. Calculer la matrice $\frac{1}{n-1} \mathbf{X}_*^T \mathbf{X}_*$ où \mathbf{X}_* représente la matrice associées aux données centrées réduites et donner ses dimensions.
2. Comparer le résultat à `cor(essence)`. Conclure.
3. Rechercher les valeurs propres et les vecteurs propres de $\frac{1}{n-1} \mathbf{X}_*^T \mathbf{X}_*$.
4. Calculer la matrice $\frac{1}{n-1} \mathbf{X}_* \mathbf{X}_*^T$ et donner ses dimensions.
5. Rechercher les valeurs propres et les vecteurs propres de $\frac{1}{n-1} \mathbf{X}_* \mathbf{X}_*^T$. Comparer avec les résultats précédents. Conclure.