Fiche de Biostatistique

Exercices

D. Chessel & A.B. Dufour

Résumé

La fiche donne des énoncés d'exercices d'algèbre et d'analyse des données. Quand une phrase commence par **?** décider si elle est vraie ou fausse et justifier.

Plan

Indépendance, générateur, dimension, bases

? Dans un espace vectoriel E, si **a**, **b**, **c** et **d** sont des vecteurs indépendants, alors **a** - **b**, **b** - **c**, **c** - **d** et **d** - **a** sont des vecteurs indépendants.

? Dans $\mathbb{R}^4$, les vecteurs (2, 14, -34, 7), (1, 4, -5, 2) et (1, 2, 3, 1) engendrent un sous-espace vectoriel de dimension 2.

? Dans un espace vectoriel E, si **a** et **b** sont des vecteurs indépendants et si **a** et **c** sont des vecteurs indépendants, alors **a**, **b** et **c** sont des vecteurs indépendants.

? Un espace vectoriel ne peut pas être constitué d'un nombre fini d'éléments.

? Les fonctions de [0,1] dans [0,1] forment un sous-espace vectoriel de l'ensemble des fonctions de $\mathbb{R}$ dans $\mathbb{R}$.

? On note E l'ensemble des polynômes à coefficients réels de degré inférieur ou égal à 3. E est un espace vectoriel sur $\mathbb{R}$. Dans E, un polynôme et sa dérivée forment toujours un système libre.

? Les vecteurs colonnes de la matrice $\begin{bmatrix} 2 & 1 & 3 & 2 & 1 & 13 \\ 0 & 2 & 1 & 4 & -1 & 15 \\ 4 & -2 & 3 & 9 & 7 & 17 \\ 2 & -3 & 4 & 5 & 22 & 23 \end{bmatrix}$ sont indépendants.

Méthode du pivot

Donner la dimension et une base des sous-espaces engendrés par les vecteurs colonnes des matrices :

$$\begin{bmatrix} 1 & 5 & -1 & 0 \\ 2 & 6 & -3 & 4 \\ 2 & 6 & 4 & 3 \\ 1 & 5 & 0 & -1 \end{bmatrix} \quad \begin{bmatrix} 2 & 1 & 3 \\ -5 & -1 & 3 \\ 3 & 1 & 1 \\ 10 & 3 & 1 \end{bmatrix} \quad \begin{bmatrix} 1 & 3 & 2 \\ 2 & 6 & 4 \\ 5 & 5 & 0 \\ -1 & -6 & -2 \end{bmatrix} \quad \begin{bmatrix} 2 & 1 & 3 & 2 \\ 0 & 2 & 1 & 4 \\ 4 & -2 & 3 & 9 \\ 2 & -3 & 4 & 5 \end{bmatrix}$$

? La matrice $\mathbf{A} = \begin{bmatrix} -2 & 0 & 1 & 2 & -1 \\ 0 & -2 & 1 & -1 & 2 \\ -2 & 1 & 2 & -1 & 0 \\ 1 & 0 & 2 & -2 & -1 \\ -1 & 2 & -2 & 0 & 1 \end{bmatrix}$ est inversible.

Produits scalaires

? Dans un espace euclidien $\|v+w\|^2+\|v-w\|^2=2\|v\|^2+2\|w\|^2$ n'est vrai que si v et w sont orthogonaux.

? Si x et y sont deux vecteurs de $\mathbb{R}^n$ et si $\|x+y\|=\|x\|+\|y\|$, alors l'un des deux vecteurs est nul.

On considère dans $\mathbb{R}^3$ la fonction h qui, aux vecteurs $\mathbf{v}=(x_1,x_2,x_3)$ et $\mathbf{w}=(y_1,y_2,y_3)$, associe le nombre réel :

$$h(\mathbf{v},\mathbf{w})=[x_1,x_2,x_3]\begin{bmatrix}1&0&2\\0&1&0\\2&0&1\end{bmatrix}\begin{bmatrix}y_1\\y_2\\y_3\end{bmatrix}$$

? La fonction h est un produit scalaire.

? La fonction $w(\mathbf{x},\mathbf{y})=[x_1,x_2,x_3]\begin{bmatrix}2&1&0\\1&5&3\\0&3&2\end{bmatrix}\begin{bmatrix}y_1\\y_2\\y_3\end{bmatrix}$ est un produit scalaire ?

? La fonction $g(\mathbf{x},\mathbf{y})=[x_1,x_2,x_3]\begin{bmatrix}1&-1&0\\-1&2&-1\\0&-1&1\end{bmatrix}\begin{bmatrix}y_1\\y_2\\y_3\end{bmatrix}$ est un produit scalaire ?

? Si l'inverse d'une matrice carrée est égale à sa transposée, ses colonnes forment une base orthonormée pour le produit scalaire canonique.

? Une base orthonormale pour un produit scalaire donné est orthogonale pour tous les autres produits scalaires.

? Si $s(\mathbf{x},\mathbf{y})=\mathbf{x}^t\mathbf{A}\mathbf{y}$ est un produit scalaire de $\mathbb{R}^3$, alors la matrice $\mathbf{A}$ est inversible et la fonction $t(\mathbf{x},\mathbf{y})=\mathbf{x}^t\mathbf{A}^{-1}\mathbf{y}$ est un produit scalaire.

Orthonormalisation

En partant de la base canonique, donner une base orthonormée pour le produit scalaire défini par :

$$\Psi:\mathbb{R}^3\times\mathbb{R}^3\to\mathbb{R}$$

$$(\mathbf{x},\mathbf{y})\quad\mapsto\Psi(\mathbf{x},\mathbf{y})=(x_1-2x_2)(y_1-2y_2)+x_2y_2+(x_2+x_3)(y_2+y_3)$$

En partant de la base :

$$\mathbf{f}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix} \quad \mathbf{f}_2 = \begin{bmatrix} 1 \\ 3 \\ 1 \\ 0 \end{bmatrix} \quad \mathbf{f}_3 = \begin{bmatrix} 3 \\ 2 \\ 1 \\ 0 \end{bmatrix} \quad \mathbf{f}_4 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}$$

donner une base orthonormée pour le produit scalaire canonique.

? $\mathbf{e}_1$, $\mathbf{e}_2$ et $\mathbf{e}_3$ désignent les vecteurs de la base canonique de R^3, muni du produit scalaire canonique. L'angle entre $\mathbf{e}_1+\mathbf{e}_2+\mathbf{e}_3$ et son projeté orthogonal sur la plan ($\mathbf{e}_1$,$\mathbf{e}_2$) vaut $p\ /4$.

? Si f est un projecteur orthogonal de E = R^n, il existe une base pour laquelle sa matrice ne contient que des valeurs égales à 0 ou à 1.

? Si f est un endomorphisme de E = R^n et si p est son rang, il existe une base pour laquelle sa matrice comporte p colonnes de 0.

? Si ϕ est un produit scalaire de E = R^n, il existe une base pour laquelle sa matrice est la matrice identité.

Applications linéaires, matrices, inverse, projecteurs

E est l'espace vectoriel $\mathbb{R}^3$, {e} est la base canonique de E. E est muni du produit scalaire canonique. On note **a**, **b** et **c** les trois vecteurs définis dans {e} par les colonnes de **H**. Une matrice **A** et la matrice **H** sont définies par :

$$\mathbf{H} = \begin{bmatrix} 1/\sqrt{3} & -1/\sqrt{2} & -1/\sqrt{6} \\ 1/\sqrt{3} & 1/\sqrt{2} & -1/\sqrt{6} \\ 1/\sqrt{3} & 0 & 2/\sqrt{6} \end{bmatrix} \quad \mathbf{A} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Soit f l'application linéaire de E dans E définie par :

$$\mathbf{u} = (x, y, z) \mapsto f(\mathbf{u}) = \left(x/\sqrt{3} + y/\sqrt{3} + z/\sqrt{3}, -x/\sqrt{2} + y/\sqrt{2}, -x/\sqrt{6} - y/\sqrt{6} + 2z/\sqrt{6} \right)$$

? La matrice de f dans la base canonique est **H**.

? La matrice **H** est la matrice d'un produit scalaire.

? La matrice **A** est la matrice d'un produit scalaire.

? L'opérateur associé à **H** est un projecteur.

? Les vecteurs **a**, **b** et **c** forment une base orthonormée de E.

? L'opérateur associé à **H** est bijectif.

? Le projecteur orthogonal sur le sous-espace vectoriel engendré par **a** et **b** a pour matrice **A** par rapport à une base convenablement choisie.

? L'opérateur associé à **A** est le projecteur orthogonal sur le sous-espace vectoriel engendré par **c**.

? La matrice $\mathbf{HAH}^t$, où $\mathbf{H}^t$ est la transposée de **H**, est de rang 3.

? L'image et le noyau de l'opérateur associé à $\mathbf{HAH}^t$ forment une somme directe orthogonale.

? La matrice **HA** peut être considérée comme la matrice d'un projecteur par rapport à deux bases convenablement choisies.

Donner une propriété de l'application dont la matrice dans la base {**a**, **b**, **c**} est $\mathbf{H}^t\mathbf{AH}$.

Diagonalisation

$$l = 2+\sqrt{2} \quad \mathbf{u} = \begin{bmatrix} 1 \\ -1 \\ -1 \\ 1 \end{bmatrix} \quad \mathbf{D} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 1 \end{bmatrix} \quad \mathbf{E} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \quad \mathbf{F} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 4 \end{bmatrix}$$

$$\mathbf{A} = \begin{bmatrix} a & b & c \\ b & a & d \\ c & d & a \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} a \\ b \\ c \end{bmatrix} \begin{bmatrix} a & b & c \end{bmatrix} \quad \mathbf{C} = \begin{bmatrix} 0 & 0 & 0 \\ 2 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

? Le vecteur **u** est un vecteur propre de la matrice **D**.

? Le nombre l est une valeur propre de la matrice **D**.

? Les matrices **E** et **F** sont semblables.

? Il existe des nombres réels a, b et c pour lesquels **A** n'est pas diagonalisable.

? La matrice **B** a une valeur propre triple.

? La matrice **C** est diagonalisable.

? La matrice d'un produit scalaire de $\mathbb{R}^n$ est diagonalisable.

? La somme d'une matrice carrée réelle et de sa transposée est diagonalisable.

? Si le produit de deux matrices carrées réelles est diagonalisable, chacune de ces deux matrices est diagonalisable.

? La somme de deux matrices diagonalisables est diagonalisable.

$$\mathbf{D} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 1 \end{bmatrix} \quad \mathbf{a} = \begin{bmatrix} 1 \\ -1-\sqrt{2} \\ 1+\sqrt{2} \\ -1 \end{bmatrix} \quad \mathbf{b} = \begin{bmatrix} 1 \\ -1 \\ -1 \\ 1 \end{bmatrix}$$

? **a** et **b** sont des vecteurs propres de **D**.

? La matrice **D** a 4 valeurs propres distinctes.

? La matrice **D** n'est pas inversible.

? La matrice **D** admet une base unique de vecteurs propres orthonormés.

? Si la matrice **D** est considérée comme un tableau de données associé à 4 individus (lignes) portant chacun le poids 1/4 et à 4 variables (colonnes) portant chacune le poids 1, ce tableau est centré.

? La matrice $\frac{1}{4}\mathbf{D}^2$ est une matrice de variances-covariances.

? La matrice **D** est une matrice de variances-covariances.

? La matrice **D** n'est pas une matrice de corrélation.

? Si la matrice **D** est considérée comme un tableau de données associé à 4 individus (lignes) portant chacun le poids 1/4 et à 4 variables (colonnes) portant chacune le poids 1, l'analyse en composantes principales de ce tableau présente une valeur propre nulle.

? En général, une matrice de variances-covariances inversible est la matrice d'un produit scalaire.

$$\mathbf{A} = \begin{bmatrix} a & b & c \\ b & a & d \\ c & d & a \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} a \\ b \\ c \end{bmatrix} \begin{bmatrix} a & b & c \end{bmatrix} \quad \mathbf{C} = \begin{bmatrix} 0 & 0 & 0 \\ 2 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

? Il existe des nombres réels a, b et c pour lesquels **A** n'est pas diagonalisable.

? La matrice **B** a une valeur propre triple.

? La matrice **C** est diagonalisable.

? Une matrice de corrélation est toujours inversible.

? Une matrice de corrélation inversible est une matrice de produit scalaire.

E est un espace vectoriel de dimension finie sur $\mathbb{R}$. Soit f un endomorphisme sur E. Id est l'application identité.

? Si $f^2 = f$, alors f est inversible.

? Si f est inversible, alors f^{-1} est diagonalisable.

? Si f est diagonalisable, alors f est inversible.

? Si f est inversible, alors f^2 aussi.

? Si f est diagonalisable, alors f^2 aussi.

? Si $f^2 = f$, alors f est diagonalisable.

? Si $f^2 = \mathrm{Id}$, alors f est inversible.

? Si f est inversible et diagonalisable, alors f^{-1} est diagonalisable.

? Si la matrice de f pour une base {u} est symétrique, alors f est diagonalisable.

? Le noyau de f est un sous-espace propre de f.

A est une matrice carrée réelle symétrique à p lignes et p colonnes ($p > 1$).

? **A** est toujours diagonalisable.

? **A** est toujours inversible.

? Le rang de **A** et celui de $\mathbf{A}^2$ sont les mêmes.

? **A** et $\mathbf{A}^2$ ont les mêmes valeurs propres.

? **A** et $\mathbf{A}^2$ ont les mêmes vecteurs propres.

? Si $\mathbf{A}^2$ est diagonale **A** est la matrice d'un produit scalaire.

$$\mathbf{A} = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} 3 & 0 & 8 \\ 3 & 1 & -6 \\ -2 & 0 & -5 \end{bmatrix}$$

Diagonaliser **A** si c'est possible.

Diagonaliser **B** si c'est possible .

? $\mathbf{BB}^t$ et $\mathbf{B}^t\mathbf{B}$ sont toutes les deux diagonalisables.

? $\mathbf{B}+\mathbf{B}^t$ est diagonalisable.

? $\mathbf{B}^t\mathbf{AB}$ est diagonalisable.

? $\mathbf{BA}$ est inversible.

? $\mathbf{AB}$ est inversible.

? Une des matrices **A** ou **B** est une matrice de produit scalaire.

? Les deux matrices **A** et **B** sont inversibles.

? Ni la matrice **A** ni la matrice **B** ne sont des matrices de projecteurs.

On note E l'ensemble des polynômes à coefficients réels de degré inférieur ou égal à 3. E est un $\mathbb{R}$-espace vectoriel. On note {e} la base formée par 1, x, x^2 et x^3 et {g} la base formée par les polynômes 1, (1-x), $(1-x)^2$ et $(1-x)^3$.

? La matrice de l'application identité de {g} dans {e} est :

$$\mathbf{B} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & -1 & -2 & -3 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & -1 \end{bmatrix}$$

? L'application f de E dans E qui à un polynôme associe sa dérivée est un endomorphisme de E.

? L'application f n'a pas de vecteurs propres.

? La matrice $\mathbf{A}=\begin{bmatrix}0&1&0&0\\0&0&2&0\\0&0&0&3\\0&0&0&0\end{bmatrix}$ est diagonalisable.

? La matrice **A** est une matrice de f .

? La matrice **B** est diagonalisable.

? La matrice $\mathbf{B}^{-1}\mathbf{AB}$ n'est pas diagonalisable.

? Les matrices $\mathbf{A}\,\mathbf{A}^t$ et $\mathbf{A}+\mathbf{A}^t$ sont diagonalisables.

? Si une matrice **X** est semblable à une matrice **Y** et si la matrice **Y** est semblable à une matrice **Z**, alors les matrices **X** et **Z** sont semblables.

? Si **X** et **Y** sont des matrices carrées, symétriques, réelles et si **XY** = **YX**, alors **XY** est diagonalisable.

E est l'espace vectoriel $\mathbb{R}^4$,{e} est sa base canonique et $\langle \mathbf{x} | \mathbf{y} \rangle$ est son produit scalaire canonique. On considère la matrice A et *f* le R-opérateur associé à **A** :

$$\mathbf{A}=\begin{bmatrix}-1&1&-1&1\\1&-1&1&-1\\-1&1&1&-1\\1&-1&-1&1\end{bmatrix}$$

? La matrice **A** est symétrique et inversible.

? La dimension du noyau de *f* vaut 3.

? Il existe un produit scalaire de E dont la matrice par rapport à {e} est **A**.

? L'application f est un projecteur.

? Les colonnes de A sont des vecteurs orthogonaux.

? L'application définie par $g(\mathbf{x},\mathbf{y})=\langle \mathbf{x} \,|\, f(\mathbf{y})\rangle$ est bilinéaire mais non symétrique

? L'application définie par $h(\mathbf{x},\mathbf{y})=\langle f(\mathbf{x}) \,|\, f(\mathbf{y})\rangle$ est un produit scalaire

A est maintenant considérée comme un tableau de mesures sur 4 individus (lignes) et 4 variables (colonnes). Chaque ligne a le poids 1/4 et chaque colonne a le poids 1.

? Ce tableau est centré pour la pondération uniforme.

? Les variances et les covariances sont toutes égales.

? La matrice de corrélation est de rang 2.

? L'angle entre deux variables vaut au maximum $p/2$.

? L'inertie totale des nuages associés vaut 16.

Statistique

? L’écart-type de la somme de deux variables est égale à la somme des écarts-types de chacune d’entre elles si et seulement si leur corrélation est nulle.

? L’écart-type de la somme de deux variables est supérieur ou égal à la somme des écarts-types de ces deux variables.

On considère les variables $\mathbf{x} = (x_1, x_2, x_3, x_4) = (0,1,2,3)$ et $\mathbf{y} = (y_1, y_2, y_3, y_4) = (1,3,3,9)$

Donner l’équation de la droite $f(x) = ax$ qui minimise la quantité $E_f = \sum_{i=1}^{4} (y_i - f(x_i))^2$

Donner l’équation de la droite $g(x) = bx$ qui minimise la quantité $E_g = \sum_{i=1}^{4} \frac{(y_i - g(x_i))^2}{x_i + 1}$

La pondération implicite est la pondération uniforme. Tous les résultats seront donnés sans approximation numérique.

La variable **x** prend n = 4 valeurs : $\mathbf{x} = (-3, 0, 1, 2)$.

La variable **y** prend n = 4 valeurs : $\mathbf{y} = (0, 2, 2, 4)$.

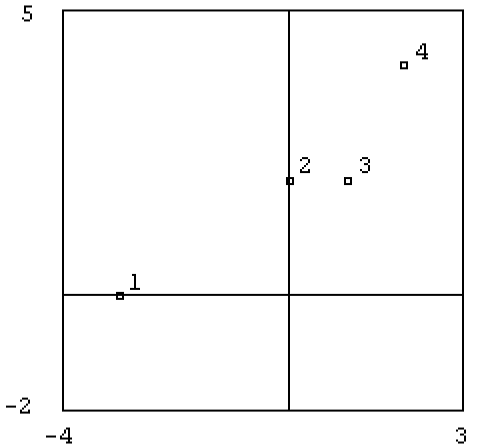
Donner la moyenne et la variance de **x**, la moyenne et la variance de **y**, la covariance et la corrélation des deux variables.

Donner la fonction $y = bx$ qui minimise $E(b) = \frac{1}{n}\sum_{i=1}^{n}(y_i - bx_i)^2$.

Donner la fonction $y = bx + c$ qui minimise $E(b,c) = \frac{1}{n}\sum_{i=1}^{n}(y_i - bx_i - c)^2$.

Donnerla fonction $x = by + c$ qui minimise $E(b,c) = \frac{1}{n}\sum_{i=1}^{n}(x_i - by_i - c)^2$.

Donner la fonction $y = ax^2 + bx + c$ qui minimise $E(a,b,c) = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - ax_i^2 - bx_i - c\right)^2$.



Tracer les solutions trouvées sur la figure. Quel est le modèle qui donne l'erreur minimum ?

? Les valeurs propres d'une matrice de corrélation sont toujours positives ou nulles.

? Le rang de la matrice de corrélation d'une ACP normée sur *p* variables vaut *p*.

? Dans une matrice de corrélation théorique sur p variables dans laquelle tous les coefficients non diagonaux sont égaux à a une valeur propre vaut $1+(p-1)a$.

? L'inertie projetée sur le premier axe principal d'une ACP normée sur *p* variables vaut au moins 1.

? Le pourcentage de variance expliquée par la prédiction linéaire de **x** par **y** est toujours égal au pourcentage de variance expliquée par la prédiction linéaire de **y** par **x**.

? Pour deux variables statistiques centrées et de variances non nulles, le nuage des points est sur une droite si et seulement si le coefficient de corrélation égale à 1.

Droite curieuse

On a réalisé sur 22 individus des observations sur deux caractères X et Y et obtenu les résultats suivants :

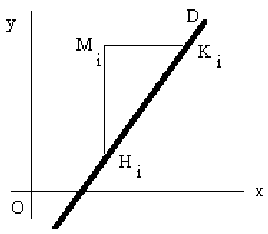
X	168	169	150	148	154	145	165	163	148	161	151
Y	71	68	65	67	67	66	69	69	68	69	70

X	176	159	159	151	155	169	158	157	161	146	150
Y	74	70	73	69	71	74	70	71	73	71	65

a) Calculer les moyennes, les écarts-types et le coefficient de corrélation linéaire.

b) Donner les deux droites de régression et leur représentation graphique associée au nuage des donnée brutes.

c) M_i est un point du nuage (données brutes) et D une droite du plan d'équation $y = ax + b$.



On note $d\left(\mathrm{M},\mathrm{P}\right)$ la distance euclidienne (métrique canonique) entre deux points M et P et on souhaite déterminer D pour rendre minimum la quantité :

$$S = \sum_{i=1}^{n} d\left(\mathrm{M}_i, \mathrm{H}_i\right) d\left(\mathrm{M}_i, \mathrm{K}_i\right)$$

d) Etablir que D doit passer par le centre d'inertie du nuage des points pondérés uniformément et exprimer S en fonction de a seul.

e) Déterminer en fonction de la corrélation et des écarts-types la ou les valeurs de a rendant S le plus faible possible et calculer la valeur correspondante de S dans les cas où :

– un des écart-type est nuls ;

– la covariance est nulle ;

– les écarts-types et la covariance sont non nuls.

f) Comparer le coefficient angulaire de D avec les coefficients similaires des droites de régression.

g) Appliquer ces résultats aux données numériques proposées et tracer D.

Régression par l'origine

x est le nombre de pièces et y est la surface en m^2 d'un appartement construit entre 1958 et 1962.

X Y	[10, 20]	[20, 30]	[30, 40]	[40, 50]	[50, 60]	[60, 70]	[70, 80]	[80, 90]	[90, 100]	[100, 160
1	1	4	2	1	0	0	0	0	0	0
2	1	12	22	8	3	1	1	0	0	0
3	0	0	1	15	64	23	10	5	0	0
4	0	0	0	1	10	55	24	10	7	4
5	0	0	0	0	0	4	13	12	6	6
6	0	0	0	0	0	0	1	2	2	7
7	0	0	0	0	0	0	0	0	2	1
8	0	0	0	0	0	0	0	0	0	2

La liaison la plus simple qu'on peut supposer entre les deux variables est du type $y = ax$ (régression par l'origine). Tracer le nuage des données, la droite de régression ordinaire, la courbe de régression (prédiction par la moyenne conditionnelle). Calculer et comparer les variances résduelles.

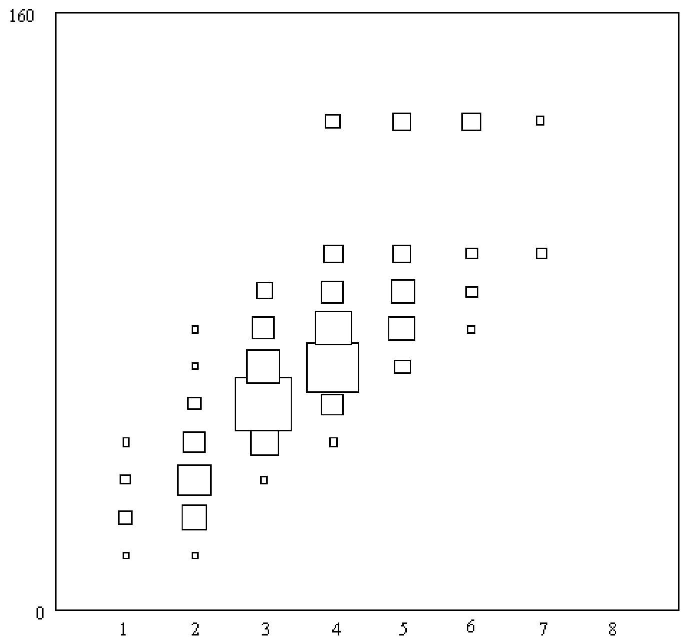
Donner la solution des moindres carrés des modèles :

$$y = a_1 x \quad \text{qui minimise } S_1 = \sum_{i=1}^{n} \left(y_i - a_1 x_i \right)^2$$

$$y = a_2 x \quad \text{qui minimise } S_2 = \sum_{i=1}^{n} \frac{\left(y_i - a_2 x_i \right)^2}{x_i}$$

$$y = a_3 x \quad \text{qui minimise } S_3 = \sum_{i=1}^{n} \frac{\left(y_i - a_2 x_i \right)^2}{x_i^2}$$

Tracer les droites correspondantes en utilisant le graphe ci-dessous. Commenter la notion de "modèle utile" et de "modèle vrai".



Jury de sélection

On considère un jury formé de L juges sélectionnant chacun R produits parmi C. Le résultat de la sélection est consigné dans un tableau $\mathbf{X}$ à L lignes et C colonnes où x_{ij} vaut 1 si le juge i sélectionne le produit j et 0 sinon.

a) Soit A_{ik} le nombre des sélections communes aux juges i et k. Donner le coefficient de corrélation linéaire r_{ik} entre les juges i et k, en fonction de A_{ik} , R et C.

b) On peut définir sur l'ensemble des C produits une statistique S en posant :

$$S_j = \sum_{i=1}^{L} x_{ij}$$

Quelle signification donner à S_j ? Calculer la moyenne des valeurs de S_j en fonction de L, C et R. Exprimer la variance V des valeurs S_j en fonction de L, C, R et la somme pour $i \neq k$ des A_{ik} .

c) Exprimer en fonction de L, C, R et V la moyenne Z des coefficients de corrélation linéaire r_{ik} sur l'ensemble des $L(L-1)/2$ paires de juges.

d) Vérifier le résultat précédent dans le cas où les L jugements concordent.

Vrai ou Faux ? Dans un jury de sélection comportant 10 juges sélectionnant exactement 4 produits sur 8 proposés, la variance du nombre de sélection par produits ne peut pas dépasser 20.

Vrai ou Faux ? Dans un jury de sélection 24 juges choisissent 4 produits parmi 6. La variance du nombre de sélection par produit ne peut pas dépasser 128.

Jury de classement

On considère un jury de dégustation de L juges qui ont à classer C produits pour un critère donné. Chacun des juges classe les produits par ordre de préférence (sans ex æquo) et attribue à chacun un entier compris entre 1 et C : B_{ij} est le rang proposé par le juge i pour le produit j. Le résultat de la dégustation est consigné dans un tableau de L lignes et C colonnes où chaque ligne est une permutation des entiers 1, ..., C. On note m_i et s_i^2 la moyenne et la variance des rangs attribués par le juge i. De plus Sj est la somme des rangs associés au produit j et on pose :

$$T_{ik} = \sum_{j=1}^{c} B_{ij} B_{kj}$$

a) Calculer m_i et s_i^2.

b) Calculer le coefficient de corrélation linéaire r_{ik} entre les juges i et k (coefficient de Spearman) en fonction de C et T_{ik} .

c) Calculer la moyenne M et la variance V de la statistique $(S_j)_{j=1,...,C}$ en fonction de L, C et la somme pour $i \neq k$ des T_{ik} .

d) Exprimer en fonction de L, C, V la moyenne Z des coefficients de Spearman pour tous les couples (i,k) de juges.

e) Vérifier le résultat obtenu lorsque tous les jugements concordent.

f) Application : 3 juges ont classé 7 vins de la manière suivante :

/	1	2	3	4	5	6	7
1 /	5	2	3	4	1	6	7
2 /	7	1	3	6	2	5	4
3 /	2	1	4	7	3	6	5

Calculer le coefficient de corrélation moyen entre les juges. Sachant que seuls 3 vins sont retenus pour le tour suivant on peut réduire les données aux choix effectué par chacun des juges soit :

```
  / 1 2 3 4 5 6 7
_ _ _ _ _ _ _ _ _
1 / 0 1 1 0 1 0 0
2 / 0 1 1 0 1 0 0
3 / 1 1 0 0 1 0 0
```

Calculer le nouveau coefficient de corrélation moyen. Commenter alors les trois méthodes de sélection pour le tour suivant :

A- Seront sélectionnés les produits ayant obtenu la meilleure somme des rangs;
B- Seront sélectionnés les produits ayant obtenu le plus grand nombre de sélections par un juge;
C- Sera d'abord sélectionné le produit ayant obtenu le plus grand nombre de places de premier, puis en cas d'ex æquo le plus grand nombre de places de second, puis en cas d'ex æquo le plus grand nombre de places de troisième, puis...

Coefficient de concordance

Les notations sont celles de l'exercice précédent.

a) Pour quelle situation la variance $\underline{V}$ de S sera-t-elle maximale? Quelle est alors la variance $Vmax$?

b) On pose $K = V/Vmax$. K est appelé coefficient de concordance de Kendall. Exprimer Z en fonction de K et K en fonction de Z. Justifier le nom donné à K.

c) On pose $U_{ik} = \sum_{j=1}^{c}\left(B_{ij} - B_{kj}\right)^2$. Quelle relation existe t'il entre r_{ik} et U_{ik} ?

d) Application : 9 juges classent 6 vins de consommation courante pour la finesse de l'arôme (tableau 1) et l'harmonie générale (tableau 2)

	Tableau 1						Tableau 2					
	A	B	C	D	E	F	A	B	C	D	E	F
1	6	5	4	2	1	3	6	3	2	1	4	5
2	6	5	4	3	1	2	5	6	4	2	1	3
3	1	4	5	6	2	3	3	6	1	2	5	4
4	4	3	2	1	5	6	5	3	1	2	4	6
5	4	5	3	2	1	6	6	5	4	2	1	3
6	1	5	6	3	4	2	4	2	1	5	3	6
7	6	5	2	3	1	4	6	5	2	3	1	4
8	5	4	3	2	1	6	5	4	3	2	1	6
9	5	3	1	2	4	6	5	2	1	3	4	6

- Pour chacun des deux tableaux, calculer V, Z, K. Pour quel critère le jury est-il le plus homogène? Commentaires.
- D'après le tableau 1, déterminer un jugement moyen et le comparer à chacun des 9 jugements observés. Montrer que l'on peut classer les juges en deux catégories pertinentes. Quels sont le produits qui différencient les catégories précédentes?
- Etudier la cohérence des sous-groupes. Commentaires.

Pour en savoir plus : R. Tomassone et C. Flanzy (1977) Présentation synthétique de diverses méthodes d'analyse de données fournies par un jury de dégustateurs. *Ann. Technol. Agric.*, 26, 373-418.

SMIG

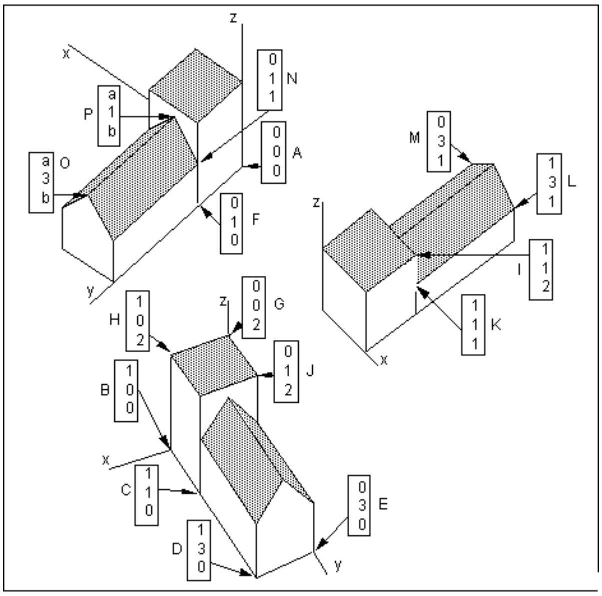
En 1970 le SMIG a été remplacé par le SMIC. De nombreuses rentes viagères étant indexées sur le SMIG, on a conservé l'indice ancien à cet usage sous le nom de minimum garanti. Depuis 1968, le minimum garanti a pris au mois de juillet les valeurs suivantes :

1968	3.00	1973	3.98	1978	6.82
1969	3.20	1974	4.63	1979	?
1970	3.42	1975	5.20		
1971	3.61	1976	5.69		
1972	3.80	1977	6.25		

Commenter l'évolution de cet indice sur la période considérée.

Maison

Représenter cet objet dans la directions de votre choix ($a = \frac{1}{2}$, b=$\sqrt{3}$)



Parabole

On considère les variables :

$$\mathbf{x} = (x_1, \ldots, x_5) = (-2, -1,0,1,2)$$
$$\mathbf{y} = (y_1, \ldots, y_5) = (-3,1,4,1,\text{-}3)$$

Donner l'équation de la parabole $f(x) = ax^2 + bx + c$ qui minimise la quantité :

$$E = (1/5)\sum_{i=1}^{5}(y_i - f(x_i))^2$$

Donner la valeur du minimum atteint et tracer cette parabole. N.B. Aucune approximation numérique n'est autorisée.

Même exercice pour $\mathbf{x} = (x_1, x_2, x_3, x_4) = (0,1,2,3)$ et $\mathbf{y} = (y_1, y_2, y_3, y_4) = (1,3,3,9)$

Tables de contingence

On a relevé les notes de 50 étudiants. **x** est la note de mathématiques et **y** celle d'informatique. On obtient le tableau de contingence :

	y	[5, 7[	[7, 9[	[9, 11[	[11, 13]	[13, 15[
x		6	8	10	12	14
[4, 6]	5	0	5	0	0	1
[6, 8[	7	2	2	2	1	2
[8, 10[	9	1	4	6	0	2
10, 12[	11	3	1	3	5	3
[12, 14[	13	0	0	0	4	4

a) Donner les moyennes et les variances marginales.
b) Calculer les effectifs, les moyennes et les variances conditionnels.
c) Dessiner le nuage et les courbes de régression. Calculer les rapports de corrélation.
d) Déterminer les droites de régression. Calculer le coefficient de corrélation.
e) Commenter les décompositions de la variance.

---

x et **y** sont deux variables statistiques qui prennent les valeurs 0 ou 1. On peut les considérer soit comme variables quantitatives, soit comme variables qualitatives (non/oui).

a) Pour n individus on recueille a observations du type (0,0), b observations du type (0,1), c observations du type (1,0) et d observations du type (1,1). La statistique est alors donnée par la table de contingence :

$$\begin{array}{cccc} & y=0 & y=1 & total \\ x=0 & a & b & a+b \\ x=1 & c & d & c+d \\ total & a+c & b+d & n \end{array}$$

Donner en fonction de a, b, c, d et n les moyennes et variances marginales, les moyennes et variances conditionnelles, la covariance, le carré du coefficient de corrélation linéaire et les rapports de corrélation.

En vrac

Indiquer clairement quelle est votre opinion sur les assertions suivantes. Elles peuvent être toujours vraies, toujours fausses, parfois vraies et parfois fausses, sans signification ou non définies, et vous pouvez ne pas avoir d'opinion sur la question (ce qui est parfois prudent). Dans chaque cas donner un argument qui justifie votre position.

B1 - La moyenne d'une série statistique de nombres positifs est positive.

B2 - La variance d'une série statistique de deux nombres est nulle.

B3 - Le coefficient de corrélation linéaire est une quantité positive ou nulle

B4 - Le coefficient de corrélation ne dépend pas des unités de mesure.

B5 - Une matrice carrée est symétrique.

B6 - Une matrice carrée inversible est diagonalisable.

B7 - La somme des valeurs propres d'une matrice est nulle.

B8 - Un système linéaire de deux équations à trois inconnues a des solutions.

B9 - Il y a deux droites de régression pour un nuage de points bivarié.

Analyse en composantes principales

On veut faire l'analyse en composantes principales centrée (ACP) de la matrice :

$$\mathbf{X} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

On a 4 lignes-individus et 4 colonnes-variables. La pondération des lignes est uniforme, la pondération des colonnes est unitaire, la transformation préalable est le centrage par colonne.

Donner les moyennes des 4 variables. Donner les variances des 4 variables. Donner la matrice des covariances du tableau **X**.

Donner les valeurs propres de la matrice $\mathbf{A} = \begin{bmatrix} 3 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 3 \end{bmatrix}$.

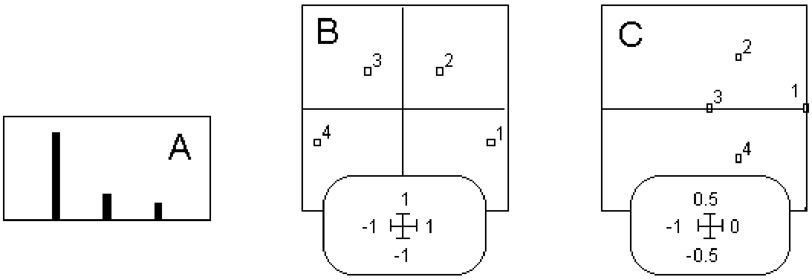
En déduire les valeurs propres de l'ACP de **X**. Donner l'inertie totale de l'ACP de **X**.

Donner le **second** axe principal de l'ACP de **X**.

Donner les coordonnées des lignes sur le **second** axe principal de l'ACP de **X**.

Donner les coordonnées des colonne sur la **seconde** composante principale de l'ACP de **X**.

Donner une légende pour la figure :


A
B
3
2
4
1
1
-1
1
-1
C
2
3
1
4
0.5
-1
0
-0.5