

Algèbre matricielle

Badia Jacques¹

avec la collaboration de

Bastien C., Durier C., Gasqui P., Goyeau H., Jayet P-A.,
Lefort-Buson M., Malet P., Menbre J-M., Oury F-X.,
Pellerin S., Pinochet X., Solari A., Volaire F.

Table des matières

Avant-Propos	ix
1 Espace vectoriel - Géométrie	1
1.1 Représentations géométriques	4
1.2 Espace vectoriel	6
1.3 Projections	7
1.4 Géométrie analytique	8
2 Géométrie - Statistique - Probabilité	13
2.1 Statistique et Géométrie	13
2.1.1 Résumer des observations	13
2.1.2 Mesurer des liaisons	17
2.1.3 Expliquer ou prédire des variables	20
2.1.4 Comparer des populations	23
2.2 Probabilité et Géométrie	27
2.2.1 Espérance mathématique	27
2.2.2 Variance	28
2.2.3 Corrélation	28
2.2.4 Espérance conditionnelle	28
2.2.5 Variance et variance conditionnelle	29
2.3 Tableau résumé et compléments bibliographiques	33
3 Formalisation des concepts	35
3.1 Espace vectoriel	35
3.1.1 Sous espace vectoriel	36

3.1.2	Combinaison linéaire	36
3.1.3	Indépendance linéaire	36
3.1.4	Base d'un espace vectoriel	36
3.2	Le langage géométrique	37
3.2.1	Sous espace affine	38
3.2.2	Décomposition en somme directe	38
3.3	Exemples	39
3.3.1	Représentations dans l'espace des variables	39
3.3.2	Base d'un espace vectoriel	40
3.3.3	Indépendance linéaire	41
3.3.4	Application linéaire et représentation matricielle	42
3.3.5	Représentation des sous espaces engendrés par les colonnes d'une matrice	44
4	Calcul matriciel	49
4.1	Définitions	49
4.2	Opérations élémentaires	51
4.2.1	Matrice des variables centrées réduites	55
4.2.2	Indépendance linéaire	56
4.3	Représentation géométrique	56
4.3.1	Exemples	58
4.4	Valeurs et vecteurs propres	62
4.4.1	Définition	63
4.4.2	Résultats généraux	63
4.4.3	Recherche de valeurs et de vecteurs propres	65
4.4.4	Vecteurs propres et droites de régression	65
4.5	Projecteurs associés à une décomposition en somme directe	68
4.5.1	Projecteurs associés à des décompositions en somme directe	68
5	Partitionnement - Différenciation - Extremum	73
5.1	Partitionnement	73
5.1.1	Opérations élémentaires	73

5.1.2	Inverse et déterminant d'une matrice partitionnée	75
5.2	Différenciation	76
5.3	Extremum	78
5.3.1	Maximisation sous contrainte	78
5.4	Interprétation géométrique des valeurs et des vecteurs propres .	78
5.4.1	Relations entre deux ellipsoïdes	80
5.4.2	A propos de la maximisation	81
5.5	Dérivée d'un déterminant	82
5.6	Quelques formules utiles	82
5.7	Description d'un ensemble d'observations	83
5.7.1	Illustration géométrique	89
5.8	Références	92
	Index	94
	Index	95

Table des figures

1.1	Représentation des individus puis des variables.	3
1.2	Addition de deux vecteurs.	4
1.3	Différence de deux vecteurs.	4
1.4	Multiplication par un scalaire.	5
1.5	La règle des cosinus.	5
1.6	Projections orthogonales.	8
1.7	Coordonnées et axes.	10
2.1	Ecart à la moyenne.	15
2.2	Représentation des vecteurs observations, moyenne et erreur. . .	16
2.3	Coefficient de corrélation simple.	17
2.4	Corrélations simple, multiple et partielle.	18
2.5	Représentation de T_c , T_f et P	19
2.6	Représentation de T'_c , T'_f et P'	20
2.7	Régression et coefficient de corrélation.	20
2.8	Projection sur le plan XOJ.	21
2.9	Projection sur OJ.	22
2.10	Projections sur le sous espace des variables centrées.	23
2.11	Comparaison de deux populations P_1 et P_2	25
2.12	Espérance conditionnelle.	29
2.13	Rotation orthogonale dans le plan.	31
3.1	Construction du vecteur $\alpha_1 B_1 + \alpha_2 B_2$	37
3.2	Sous espace vectoriel (E_1) et sous espace affine (E_2) de l'espace vectoriel E de dimension 2.	38

3.3	Représentation dans \mathbb{R}^3 des sous espaces complémentaires D et P	39
3.4	Sous espaces P_1 et P_2 engendrés par les colonnes de la matrice X	46
3.5	Sous espaces P'_1 et P'_2 engendrés par les colonnes de la matrice X'	46
4.1	Ellipses d'inertie centrées au point moyen μ pour différentes valeurs du coefficient de corrélation (ρ).	59
4.2	Ellipses d'inertie centrées au point moyen μ pour des valeurs différentes du coefficient de corrélation (ρ) et des écarts types (S_1 et S_2).	61
4.3	Ellipses dans les bases orthonormales OB_1, OB_2 et OB_1^*, OB_2^*	62
4.4	Régression de x en y (1), droite des moindres rectangles (2), régression orthogonale (3), régression de y en x (4).	67
4.5	Décomposition de \mathbb{R}^3 en une somme directe: 1) droite D_1 et plan P , 2) droite D_2 et plan P	69
5.1	Ellipses E_1 et E_2 centrées en μ . D_1 et D_2 définissent un couple d'axes conjugués. Les valeurs propres de E_1 relatives à E_2 sont $\lambda_i = [\mu A_{i2}/\mu A_{i1}]^2$ ($i = 1, 2$).	80
5.2	Même chose que pour la Figure 5.1 avec μA_{21} et μA_{11} perpendiculaires, même unité de longueur sur ces axes de telle sorte que E_1 soit un cercle unité.	81
5.3	Représentation des données: iris Versicolor (\bullet), iris Setosa (\circ).	84
5.4	Premier vecteur propre de W_1, W_2 et W , points moyens (\star).	86
5.5	Premier vecteur propre de $W^{-1}B$	88
5.6	Ellipses d'inertie centrées aux points moyens μ_1 (iris Versicolor) et μ_2 (iris Setosa), direction de projection (D_1) et axe conjugué (D_2) de $\mu_1\mu_2$ par rapport aux ellipses.	89
5.7	Ellipses d'inertie des populations P_1, P_2 et P_3 centrées respectivement aux points moyens μ_1, μ_2 et μ_3	90
5.8	Représentation de l'ellipse d'inertie E_1 commune aux trois populations et de l'ellipse d'inertie E_2 des points moyens.	90
5.9	Direction qui maximise le rapport des longueurs de deux segments $\mu M_2/\mu M_1$	91

Liste des tableaux

2.1	Correspondance entre Espace vectoriel, Géométrie, Statistique et Probabilité.	33
3.1	Sociétés d'investissements dont les portefeuilles sont composés des titres de trois grandes compagnies (C.E.A, I.N.R.A, C.N.R.S).	41
4.1	Longueur du corps (x) et de la profondeur de la poitrine (y) de 22 vaches laitières.	65
5.1	longueur (X), largeur (Y) des sépales mesurées sur iris Setosa et iris Versicolor.	83
5.2	Matrices d'inertie, de variance-covariance, degrés de liberté (ddl).	85
5.3	Estimations des moyennes.	85
5.4	Valeurs propres et vecteurs propres des matrices d'inertie W_1 , W_2 et W	86

Avant-Propos

L'algèbre matricielle présente l'intérêt d'un formalisme simple, elle permet de nombreuses représentations géométriques dans la mesure où on ne la considère pas comme une simple technique de calcul. Les qualités qui viennent d'être soulignées sont particulièrement intéressantes en statistique puisqu'elles permettent d'une part, d'unifier la présentation des théories statistiques et d'automatiser les calculs et d'autre part, de rendre plus concrets les résumés statistiques par le truchement des représentations graphiques. Ce dernier point est d'importance puisqu'il touche au domaine essentiel de l'interprétation. Dans ces conditions, on ne peut pas dissocier l'algèbre matricielle de la théorie géométrique qui la sous-tend.

Ce document a donc pour but de présenter les outils matriciels indispensables à une bonne compréhension des méthodes statistiques du niveau 2 du plan de formation à la statistique mis en place par la Formation Permanente et le Département de Biométrie de l'INRA. Il s'agit d'un travail collectif, l'exposé des concepts a été rédigé par l'animateur, les formateurs ont conçu les transparents et les guides d'utilisation qui leurs sont associés. Son contenu est classique, on le trouve dans de nombreux ouvrages ou revues tels que ceux qui ont été utilisés et qui sont cités en références. Son originalité réside uniquement dans sa présentation délibérément géométrique, il comprend cinq chapitres, numérotés de 1 à 5.

1. Le chapitre 1 qui est un rappel des définitions élémentaires sur les espaces vectoriels et l'explicitation du concept de géométrie qui est la conséquence de la définition d'une distance qui dérive elle-même de la définition d'un produit scalaire.
2. Le chapitre 2 qui applique les principes présentés dans le chapitre 1 aux représentations géométriques des résumés statistiques et probabilistes couramment utilisés.

Le contenu de ces deux chapitres est, pour l'essentiel, emprunté à Bryant (1984). Des compléments ont été apportés, ils concernent les représentations des divers coefficients de corrélation, les notions de dualité (Bachacou *et al.*, 1981) et de test.

3. Dans le chapitre 3 les principes géométriques présentés dans les chapitres 1 et 2 sont formalisés et illustrés par des exemples puisés dans Dempster (1969), Cailliez et Pages (1976).
4. Le calcul matriciel fait l'objet des chapitres 4 et 5. Leur contenu a été en partie tiré de Cyffers (1965), Kendall et Stuart (1966), Dagnélie (1969), Cailliez et Pages (1976), Morrisson (1976).

Cette rédaction a bénéficié des conseils et des critiques de collègues biométriciens que nous tenons à remercier. En particulier R. Faivre qui a fait de nombreuses remarques sur la partie relative à l'exposé des concepts, S. Junca-Holmes, E. de Turckheim et G. Philippeau (ITCF¹) qui nous ont fait bénéficier de leurs expériences de formateurs en critiquant le contenu et la présentation des transparents.

1. Institut Technique des Céréales et des Fourrages

Chapitre 1

Espace vectoriel - Géométrie

L'algèbre matricielle est d'un usage courant en statistique, c'est un moyen d'expression à la fois simple et imagé. Simple parce que le formalisme est réducteur et imagé parce qu'il est possible de représenter les objets manipulés.

Qu'est-ce qu'une matrice?

Quand un échantillon de n individus et p variables est observé, on collecte des données sous la forme de np nombres réels x_{ij} . Chaque x_{ij} est la valeur observée de la $j^{\text{ème}}$ variable ($j = 1, 2, \dots, p$) sur le $i^{\text{ème}}$ individu ($i = 1, 2, \dots, n$).

- Ces valeurs peuvent être rangées dans un tableau rectangulaire à n lignes et p colonnes dans lequel x_{ij} est l'élément à l'intersection de la $i^{\text{ème}}$ ligne et de la $j^{\text{ème}}$ colonne.

Ce tableau rectangulaire est la *matrice* X des données, on la représente par :

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}$$

- Chaque ligne i de X :

$$X_{i.} = [x_{i1} \quad x_{i2} \quad \cdots \quad x_{ip}]$$

peut être considérée comme les coordonnées d'un point ou comme un *vecteur* dans un espace de dimension p .

Les n points $i = 1, 2, \dots, n$ ainsi définis véhiculent la même information que celle qui est contenue dans la matrice des données. Chaque point de

l'espace de dimension p représente un *échantillon individuel* et l'espace qui les contient est appelé *espace des individus*.

- Chaque colonne j de X :

$$X_{.j} = \begin{bmatrix} x_{1j} \\ x_{2j} \\ \vdots \\ x_{nj} \end{bmatrix}$$

détermine les coordonnées d'un point ou un vecteur dans un espace de dimension n .

Les p points ainsi définis correspondent aux variables, l'espace qui les contient est appelé *espace des variables*. Ces p points véhiculent la même information que celle qui est contenue dans le tableau des données.

- L'espace des individus et l'espace des variables sont dits *espaces duaux*. Ainsi, chaque concept ou raisonnement dans l'un de ces espaces a une image duale dans l'autre. L'une de ces images peut paraître plus simple que l'autre, c'est ce qui justifie les représentations ou les raisonnements faits dans l'un ou l'autre de ces espaces.

Illustration numérique

Soit la matrice :

$$X = \begin{bmatrix} 2 & 4 \\ 4 & 3 \\ 6 & 5 \end{bmatrix}$$

Les lignes de cette matrice représentent des individus (il y en a trois, on les désigne par *ind.1*, *ind.2*, *ind.3*), les colonnes représentent des variables (il y en a deux, on les désigne par *var.1*, *var.2*).

Pour chaque individu (chaque ligne), les valeurs prises par *var.1* et *var.2* représentent des coordonnées dans un système de deux axes. On peut donc représenter *ind.1*, *ind.2*, *ind.3* dans le système d'axes défini par *var.1* et *var.2*, c'est l'espace des individus. De même, pour chaque variable (chaque colonne), les valeurs prises par *ind.1*, *ind.2*, *ind.3* représentent les coordonnées de *var.1* et *var.2* dans le système d'axes défini par les trois individus, c'est l'espace des variables. Ces données sont représentées sur FIG. 1.1, successivement dans l'espace des individus et dans l'espace des variables.

Les lignes (resp. les colonnes) d'une matrice sont des vecteurs lignes (resp. colonnes), nous allons nous intéresser à eux. Notre objectif est de montrer qu'un vecteur peut être interprété géométriquement. Habituellement, dans le plan de la feuille, on représente un vecteur par une flèche. Deux flèches de même longueur, de même direction et de même sens représentent des vecteurs égaux

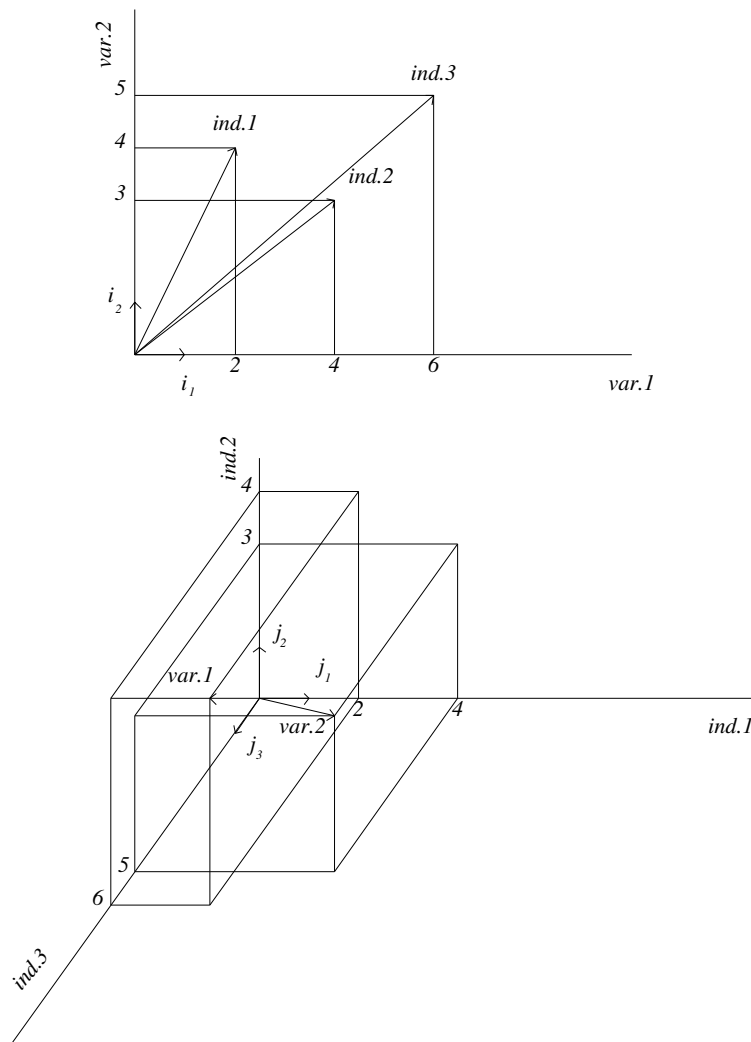


FIG. 1.1 - Représentation des individus puis des variables.

que l'on désigne par des lettres surmontées de flèches telles que : $\vec{U} = \vec{V}$. Cette notation est d'autant plus lourde qu'on est parfois amené à préciser l'origine et l'extrémité d'un vecteur en écrivant par exemple le vecteur \vec{V} sous la forme \vec{OA} où O est l'origine et A l'extrémité du vecteur \vec{V} . Pour éviter cet écueil et pour simplifier l'écriture :

- Les vecteurs seront représentés soit par des lettres majuscules telles que U, V, W, X, \dots , soit par OU, OV, UV, \dots , si l'on veut préciser l'origine et l'extrémité de ces vecteurs.
- Les composantes d'un vecteur seront désignées par des lettres minuscules indicées. Les k composantes du vecteur U seront : (u_1, u_2, \dots, u_k) .

Par ailleurs, les scalaires et les angles seront désignés par des lettres grecques minuscules telles que $\alpha, \beta, \gamma, \dots, \theta$.

1.1 Représentations géométriques

Dans le plan P de la feuille (FIG. 1.2), nous avons représenté :

- un point particulier O appelé *origine* du plan.
- des segments de droites orientés U , V , $U + V$ d'origine commune O et d'extrémités respectives U , V et $U + V$.

Ces segments de droites ont une *longueur*, une *direction* et un *sens*, ce sont les représentations géométriques des vecteurs U , V , $U + V$.

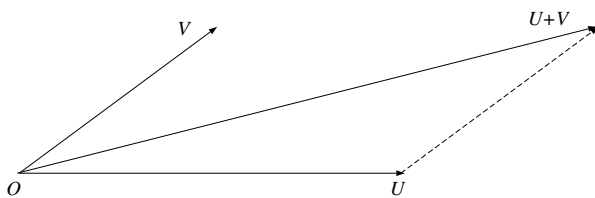


FIG. 1.2 - Addition de deux vecteurs.

Egalité de deux vecteurs

Deux vecteurs sont égaux s'ils ont même longueur, même direction et même sens.

Addition de deux vecteurs

Géométriquement, l'addition de deux vecteurs U et V est réalisée en traçant à l'extrémité de U (resp. de V) un vecteur égal au vecteur V (resp. de U) et en reliant l'origine O de U (resp. V) à l'extrémité du vecteur placé après U . Sur FIG. 1.2 le vecteur somme est noté $U + V$.

Différence de deux vecteurs

Sur FIG. 1.3, sont représentés les vecteurs U , V , $-V$ et $V - U$.

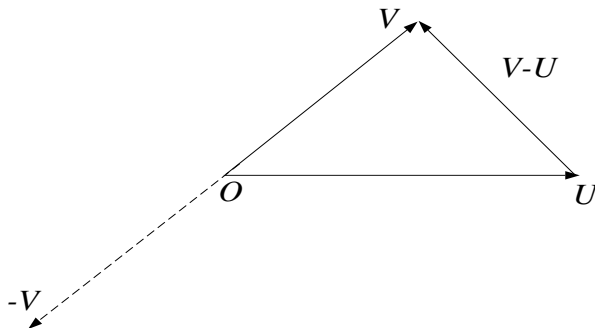


FIG. 1.3 - Différence de deux vecteurs.

Multiplication par un scalaire

Sur FIG. 1.4 sont représentés différents cas de multiplications de vecteurs (U , V , W , ...) et de scalaires (α , β , ...).

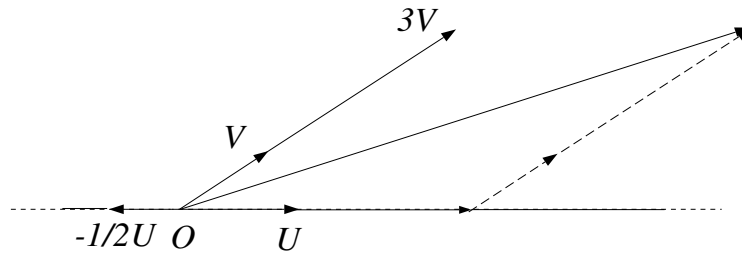


FIG. 1.4 - *Multiplication par un scalaire.*

Soient U et V deux vecteurs non nuls ayant des directions différentes alors, tout point de P est de la forme $\alpha U + \beta V$.

Longueur, distance, angle, produit scalaire

C'est à partir des éléments de FIG. 1.5 que nous rappelons les notions de longueur, de distance, d'angle et de produit scalaire.

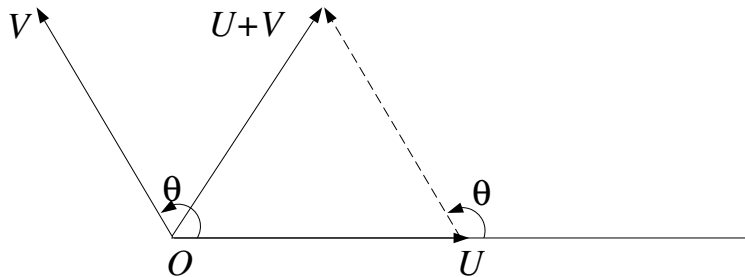


FIG. 1.5 - *La règle des cosinus.*

On note :

$|U|$ la longueur du vecteur U

$|V - U|$ la distance qui sépare les extrémités U et V des vecteurs OU et OV (cf. FIG. 1.3).

On considère sur FIG. 1.5 le triangle de sommets O , U et $U + V$. La règle des cosinus dit que :

$$|U + V|^2 = |U|^2 + |V|^2 + 2|U||V|\cos(\theta) \quad (1.1)$$

où θ est l'angle formé par les vecteurs U et V . Quand l'angle θ est droit, on dit que les vecteurs U et V sont *orthogonaux*. Dans ce cas la formule (1.1) se réduit à :

$$|U + V|^2 = |U|^2 + |V|^2 \quad (1.2)$$

on reconnaît le théorème de Pythagore.

C'est $2|U||V|\cos(\theta)$ qui fait la différence entre les formules (1.1) et (1.2). C'est cette quantité qui décrit fondamentalement la relation entre les vecteurs U et V . Cette quantité est (au coefficient 2 près) appelée *produit scalaire* des vecteurs U et V . On la note $\langle U, V \rangle$:

$$\langle U, V \rangle = |U||V|\cos(\theta) \quad (1.3)$$

En tenant compte de (1.3), (1.1) s'écrit :

$$|U + V|^2 = |U|^2 + |V|^2 + 2\langle U, V \rangle \quad (1.4)$$

qui est l'analogie de l'identité remarquable bien connue : $(a+b)^2 = a^2 + b^2 + 2ab$.

L'intérêt fondamental du produit scalaire apparaît clairement quand on sait que sa connaissance implique celles des longueurs, des distances et des angles entre deux vecteurs.

- Longueur : $|U| = \langle U, U \rangle^{1/2}$
- Distance : $|V - U| = [\langle U, U \rangle + \langle V, V \rangle - 2\langle U, V \rangle]^{1/2}$
- Angle : $\cos(\theta) = \frac{\langle U, V \rangle}{(\langle U, U \rangle \langle V, V \rangle)^{1/2}}$

Le produit scalaire satisfait aux propriétés suivantes (notées P_1 à P_4) :

- P_1 : $\langle U, V \rangle = \langle V, U \rangle$
- P_2 : $\langle U, U \rangle \geq 0$, $\langle U, U \rangle = 0$ si et seulement si $U = 0$
- P_3 : $\langle \alpha U + \beta V, W \rangle = \alpha \langle U, W \rangle + \beta \langle V, W \rangle$
- P_4 : U et V sont orthogonaux si et seulement si : $\langle U, V \rangle = 0$

1.2 Espace vectoriel

La géométrie plane peut s'exprimer en termes de vecteurs, d'addition vectorielle, de multiplication d'un vecteur par un scalaire et de produit scalaire ; deux vecteurs non colinéaires déterminent un plan ; \dots . Toutes ces notions sont généralisables à trois dimensions et au-delà de trois dimensions. Par exemple,

trois vecteurs non coplanaires engendrent un espace à trois dimensions. La *dimension* d'un espace est en général égale au nombre minimum de vecteurs requis pour déterminer l'espace.

Un espace vectoriel est une abstraction de cette approche géométrique. On dit qu'on a un espace vectoriel quand on a une collection d'éléments, appelés vecteurs, qui peuvent être additionnés et multipliés par des scalaires. Ici, on ne décrit pas ce que sont les vecteurs, on décrit comment ils se comportent. Toute chose qui se comporte de cette façon est un vecteur.

Deux concepts complémentaires émergent :

- Bien que d'inspiration géométrique, les éléments d'un espace vectoriel n'ont pas toujours besoin d'apparaître dans une représentation géométrique. On peut représenter un plan sans représenter les vecteurs qui déterminent ce plan.
- Quand un produit scalaire est donné, on définit une *géométrie*. Dans ce cas, on peut interpréter géométriquement les éléments d'un espace vectoriel quelle que soit leur vraie nature.

Par exemple, si l'on considère deux variables aléatoires X et Y , leur somme et leur différence sont aussi des variables aléatoires, il en est de même pour $\alpha X + \beta Y$ où α et β sont des scalaires. Des variables aléatoires peuvent donc être perçues comme des vecteurs et si un produit scalaire approprié est défini, on peut les interpréter géométriquement (des combinaisons linéaires de deux variables aléatoires sont représentées par un plan, ...).

1.3 Projections

Dans ce paragraphe, nous notons $\hat{U}, \hat{V}, \hat{W}, \dots$ les projections de U, V, W, \dots sur un plan P (représenté par un parallélogramme) contenant une origine O (FIG. 1.6). Ces projections faites parallèlement à une direction D perpendiculaire au plan P sont dites *projections orthogonales*.

Deux propriétés (notées P_1, P_2) définissent les projections orthogonales sur le plan P :

- P_1 : la projection orthogonale \hat{U} de U sur le plan P est dans le plan P .
- P_2 : le vecteur $U - \hat{U}$ est orthogonal à tout vecteur de P ($\langle V, U - \hat{U} \rangle = 0$ pour tout vecteur V dans le plan P).

opérateurs, les calculs, il faut traduire ces concepts en termes de coordonnées. Nous allons passer brièvement en revue la notion de coordonnées dans un espace à trois dimensions que nous généraliserons ensuite à des espaces à n dimensions.

Coordonnées

Considérons l'espace vectoriel E de dimension 3 défini de la façon suivante :

1. Les vecteurs de E sont définis par les triplets de la forme $U = (u_1, u_2, u_3)$, $V = (v_1, v_2, v_3), \dots$. Soit :

$$U = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} \quad V = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} \quad \dots$$

2. L'addition de deux vecteurs est définie composante par composante :

$$(U + V) = \begin{bmatrix} u_1 + v_1 \\ u_2 + v_2 \\ u_3 + v_3 \end{bmatrix}$$

3. La multiplication par un scalaire α est définie par :

$$\alpha U = \begin{bmatrix} \alpha u_1 \\ \alpha u_2 \\ \alpha u_3 \end{bmatrix}$$

4. Le produit scalaire de deux vecteurs U et V est défini par :

$$\langle U, V \rangle = u_1 v_1 + u_2 v_2 + u_3 v_3 \quad (1.5)$$

On peut vérifier que les définitions d'un espace vectoriel de dimension finie sont satisfaites pour E . La longueur d'un vecteur U est alors donnée par :

$$|U|^2 = u_1^2 + u_2^2 + u_3^2 \quad (1.6)$$

Considérons les vecteurs :

$$I_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \quad I_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \quad I_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

D'après (1.5) et (1.6), nous voyons que :

$$|I_1|^2 = |I_2|^2 = |I_3|^2 = 1$$

$$\langle I_1, I_2 \rangle = \langle I_1, I_3 \rangle = \langle I_2, I_3 \rangle = 0 \quad (1.7)$$

Les vecteurs I_1 , I_2 et I_3 sont des vecteurs *unitaires* mutuellement orthogonaux. Ainsi, tout vecteur U peut s'écrire comme une combinaison de I_1 , I_2 et I_3 :

$$U = u_1 I_1 + u_2 I_2 + u_3 I_3 \quad (1.8)$$

On dit que I_1 , I_2 et I_3 forment une *base* de l'espace vectoriel. La figure FIG. 1.7 montre géométriquement la relation (8). Les axes I_1 , I_2 et I_3 sont mutuellement orthogonaux. Les coefficients de (1.8) sont les coordonnées de U pour ce système d'axes.

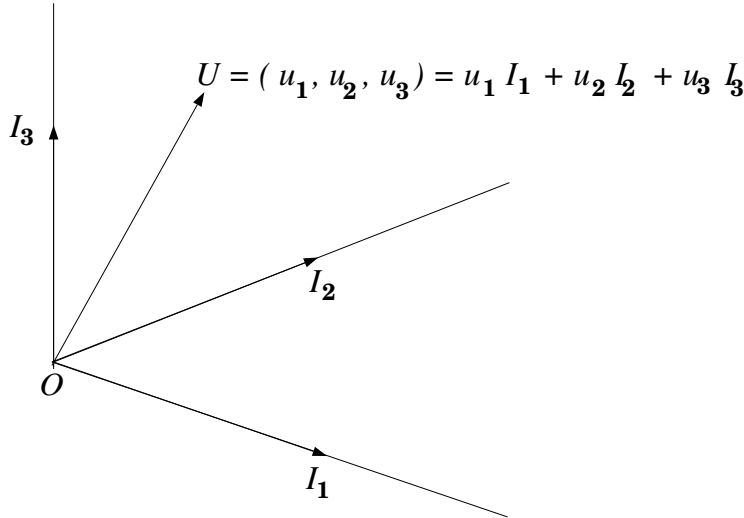


FIG. 1.7 - *Coordonnées et axes.*

Tout triplet de nombres (u_1, u_2, u_3) peut être interprété géométriquement, on peut le considérer comme un vecteur dans un espace déterminé par (1.8). On peut vérifier que toutes les propriétés qui dérivent du produit scalaire sont en accord avec les concepts de la géométrie ordinaire (géométrie *Euclidienne*). En particulier, les vecteurs U et V sont orthogonaux quand la relation (1.5) est égale à zéro; le résultat classique de géométrie analytique:

$$\cos^2(\theta) = \frac{(u_1 v_1 + u_2 v_2 + u_3 v_3)^2}{(u_1^2 + u_2^2 + u_3^2)(v_1^2 + v_2^2 + v_3^2)} \quad (1.9)$$

est en accord avec celui que suggère (1.5).

L'équation (1.6) est la généralisation à trois dimensions du théorème de Pythagore.

Nous avons vu que des vecteurs et un produit scalaire déterminent une géométrie. De ce point de vue, il n'y a pas de raisons de se limiter à trois dimensions. Toute collection de n nombres u_1, u_2, \dots, u_n peut être considérée comme un vecteur

$$U = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$$

- L'addition et la multiplication par un scalaire sont définies sur les coordonnées.
- Les axes ont pour vecteurs unitaires :

$$I_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad I_2 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} \quad \dots \quad I_n = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$$

- L'espace vectoriel à n dimensions est constitué par l'ensemble de toutes les combinaisons linéaires de I_1, I_2, \dots, I_n .
- Le produit scalaire des vecteurs U et V est défini par :

$$\langle U, V \rangle = u_1 v_1 + u_2 v_2 + \dots + u_n v_n \quad (1.10)$$

On peut calculer des produits scalaires au moyen de la formule (1.10) si l'on connaît les coordonnées. Cette formule correspond à un ensemble particulier d'axes. Si nous étions dans un autre système d'axes orthogonaux, chaque point U aurait un nouvel ensemble de coordonnées qu'il faudrait utiliser pour calculer le produit scalaire et donc, les longueurs, les angles.

Cette formule (1.10) n'est valable que si l'on est dans un système d'axes orthogonaux, si tel n'était pas le cas, la formule serait plus compliquée.

Les principes géométriques fondamentaux qui découlent du produit scalaire tels que les longueurs, les angles sont indépendants du système d'axes choisi. Par exemple, l'angle que font deux vecteurs entre eux ne dépend pas du système d'axes de référence.

Le produit scalaire (1.10) est dit produit scalaire Euclidien car, pour $n = 3$, il correspond à la géométrie Euclidienne classique. C'est de loin le produit

scalaire le plus utilisé pour des n – *uplets* de nombres, mais beaucoup d'autres produits scalaires peuvent être définis.

Toute fonction des U, V, \dots , qui satisfait à $\langle \alpha U + \beta V, W \rangle = \alpha \langle U, W \rangle + \beta \langle V, W \rangle$ est un produit scalaire convenable.

Chapitre 2

Géométrie - Statistique - Probabilité

Objectif: Montrer que les concepts statistiques, probabilistes s'expriment géométriquement.

2.1 Statistique et Géométrie

Soient (x_1, x_2, \dots, x_n) , (y_1, y_2, \dots, y_n) , \dots respectivement n observations indépendantes des variables X, Y, \dots faites sur n unités statistiques différentes appelées *individus*.

La statistique a pour but de répondre à des questions telles que :

1. quels nombres résument au mieux l'information véhiculée par les observations? que valent ces résumés?
2. quelle est l'intensité de la relation qui lie les variables? comment peut-on la mesurer à partir des observations?
3. quelle est la fonction linéaire qui permet d'expliquer (resp. prédire) au mieux une variable en fonction d'autres variables? quelle est la qualité de l'explication (resp. prédiction)?

Pour répondre à ces questions on utilise le critère des *moindres carrés* exprimé d'un point de vue géométrique.

2.1.1 Résumer des observations

D'après le critère des moindres carrés, le meilleur résumé des n observations x_i , $i = 1, \dots, n$ est le nombre α qui minimise :

$$\sum_{i=1}^n (x_i - \alpha)^2 \quad (2.1)$$

Soit X le vecteur de coordonnées (x_1, x_2, \dots, x_n) , J le vecteur de coordonnées $(1, 1, \dots, 1)$. La formule (2.1) s'écrit comme un produit scalaire :

$$\langle X - \alpha J, X - \alpha J \rangle = |X - \alpha J|^2$$

Le critère des moindres carrés signifie que parmi les résumés possibles, de la forme αJ , on choisit celui qui est le plus proche de X (*i.e.*, celui situé à la plus petite distance quand on utilise le produit scalaire Euclidien).

Le point de J le plus proche de X est le projeté (orthogonalement) de X sur J . On peut donc déterminer la valeur de α en explicitant la projection orthogonale au moyen du produit scalaire. En effet, soit \hat{X}_J une projection quelconque de X sur J , alors :

- $\hat{X}_J = \alpha J$ pour tout α
- Parmi les projections de X sur J , il en existe une (la projection orthogonale notée \hat{X}_J) qui satisfait aux propriétés P_1 et P_2 du § 1.3 :

$$\hat{X}_J = \alpha J \quad (2.2)$$

$$\langle X - \hat{X}_J, J \rangle = 0 \quad (2.3)$$

Si l'on substitue (2.2) dans (2.3), alors :

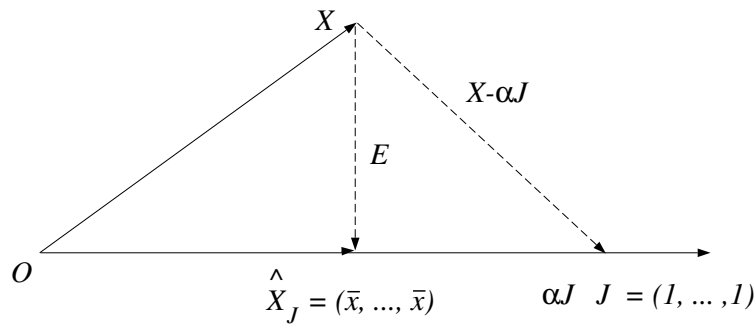
$$\langle X - \hat{X}_J, J \rangle = \langle X - \alpha J, J \rangle = 0 \quad \iff \quad \langle X, J \rangle = \alpha \langle J, J \rangle$$

En utilisant le produit scalaire Euclidien, on déduit que :

$$\sum_{i=1}^n x_i = \alpha n \quad \iff \quad \alpha = \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

On peut donc conclure que la moyenne d'échantillonnage (notée \bar{x}) est le meilleur nombre qui résume les données. Sur FIG. 2.1 est représentée la solution suggérée par la géométrie ainsi que la décomposition de X en deux composantes orthogonales : $\hat{X}_J = \bar{x}J = (\bar{x}, \bar{x}, \dots, \bar{x})$ et l'erreur qu'on commet quand on résume les données par \bar{x} .

$$E = X - \bar{x}J = (x_1 - \bar{x}, \quad x_2 - \bar{x}, \quad \dots, \quad x_n - \bar{x})$$

FIG. 2.1 - *Ecart à la moyenne.*

Si toutes les observations étaient identiques, X et J seraient colinéaires, on aurait $|E|^2 = 0$ et X serait une constante. On interprète donc E comme la variation des composantes de X ($X = \bar{x}J + (X - \bar{x}J)$).

La décomposition correspondante des carrés des longueurs est obtenue en appliquant le théorème de Pythagore :

$$|X|^2 = |\bar{x}J|^2 + |X - \bar{x}J|^2 \iff \sum_{i=1}^n x_i^2 = n\bar{x}^2 + \sum_{i=1}^n (x_i - \bar{x})^2$$

elle correspond à la décomposition classique de la somme des carrés.

$|\bar{x}J|^2$ est la *somme des carrés* expliquée par la moyenne d'échantillonnage, elle dépend du nombre n d'observations.

$|E|^2 = |X - \bar{x}J|^2$ est la *somme des carrés des erreurs (SCE)*, elle dépend du nombre n d'observations.

Le vecteur $X = \bar{x}J + (X - \bar{x}J)$ est dans un espace vectoriel de dimension n . On dit que le vecteur $\bar{x}J$ est dans un *sous espace vectoriel* de dimension 1 (engendré par J) de l'espace vectoriel de dimension n car :

- J a toutes les propriétés d'un espace vectoriel (addition et multiplication par un scalaire).
- J est contenu dans l'espace de dimension n .

Le vecteur E quant à lui est dans un sous espace de dimension $n - 1$ de l'espace de dimension n car :

- E est dans le sous espace orthogonal à J (*complémentaire orthogonal*).
- La somme des dimensions de deux sous espaces complémentaires orthogonaux est égale à la dimension de l'espace vectoriel qui contient ces sous espaces.

La moyenne d'échantillonnage sera un bon résumé de l'information véhiculée par les observations si chaque composante du vecteur E est petite (proche de zéro). Pour apprécier cela, on estime la moyenne par dimension de la somme des carrés des erreurs en divisant SCE par $n - 1$ (dimension du sous espace vectoriel qui contient E). On obtient ainsi le *carré moyen d'erreur (CME)* :

$$CME = \frac{|E|^2}{n-1} = \frac{1}{(n-1)} \sum_{i=1}^n (x_i - \bar{x})^2$$

on reconnaît l'expression de la variance d'échantillonnage.

Les dimensions des espaces sont appelées les *degrés de liberté*.

Illustration numérique

Soit $X = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$ le vecteur des observations faites sur deux individus (*ind.1* et *ind.2*). Sur FIG. 2.2 on a représenté les vecteurs X , J et $\hat{X}_J = \bar{x}J$.

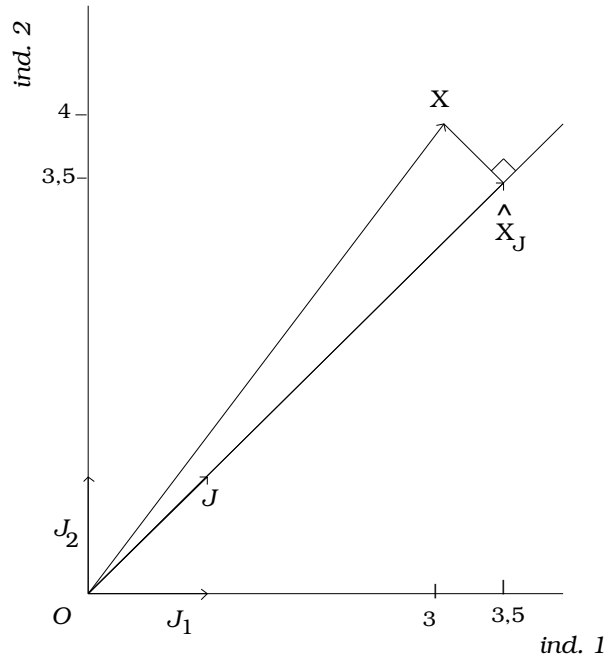


FIG. 2.2 - Représentation des vecteurs observations, moyenne et erreur.

On veut vérifier que $|X|^2 = |\hat{X}_J|^2 + |X - \hat{X}_J|^2$ en calculant chaque élément de l'égalité.

$$|X|^2 = \langle X, X \rangle = 3^2 + 4^2 = 25$$

$$\hat{X}_J = \begin{bmatrix} 3,5 \\ 3,5 \end{bmatrix}$$

$$|\hat{X}_J|^2 = \langle \hat{X}_J, \hat{X}_J \rangle = (3,5)^2 + (3,5)^2 = 24,5$$

$$|X - \hat{X}_J|^2 = \langle X - \hat{X}_J, X - \hat{X}_J \rangle = (3 - 3,5)^2 + (4 - 3,5)^2 = 0,5$$

$$|\hat{X}_J|^2 + |X - \hat{X}_J|^2 = 24,5 + 0,5 = 25 = |X|^2$$

2.1.2 Mesurer des liaisons

Apprécier l'intensité de la relation qui lie deux variables X et Y consiste à mesurer l'influence des variations de X sur Y .

Ces variations liées de X et Y sont appréhendées à partir des composantes orthogonales à J ($X - \hat{X}_J = X - \bar{x}J$ et $Y - \hat{Y}_J = Y - \bar{y}J$), la mesure utilisée est le cosinus de l'angle θ entre ces composantes (FIG. 2.3). Cet angle θ est insensible à tout changement d'échelle ou translation sur X ou Y , son cosinus n'est autre que le coefficient de *corrélation simple* de X et Y .

Le produit scalaire de $X - \bar{x}J$ et $Y - \bar{y}J$ divisé par les degrés de liberté appropriés ($n - 1$) n'est autre que la covariance d'échantillonnage de X et Y .

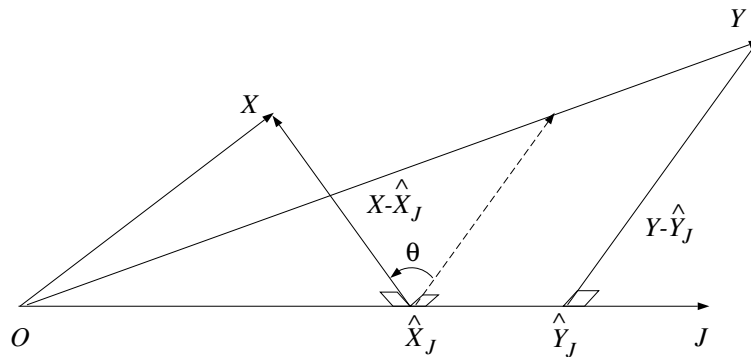


FIG. 2.3 - Coefficient de corrélation simple.

D'un point de vue très général :

- les différentes mesures d'intensité des relations entre des variables X, Y, Z, \dots sont appréhendées dans le *sous espace vectoriel orthogonal* à J .
- le sous espace vectoriel orthogonal à J est le sous espace vectoriel des *variables centrées*. Il permet de visualiser simplement les différents coefficients de corrélation (corrélations simple, multiple et partielle).

On note :

$X' = X - \bar{x}J$, $Y' = Y - \bar{y}J$ et $Z' = Z - \bar{z}J$ trois variables centrées.

$\hat{X}'_{Z'}$, $\hat{Y}'_{Z'}$ respectivement les projections de X' et Y' sur Z' .

$\hat{X}'_{Y',Z'}$ la projection de X' sur le plan déterminé par Y' et Z' .

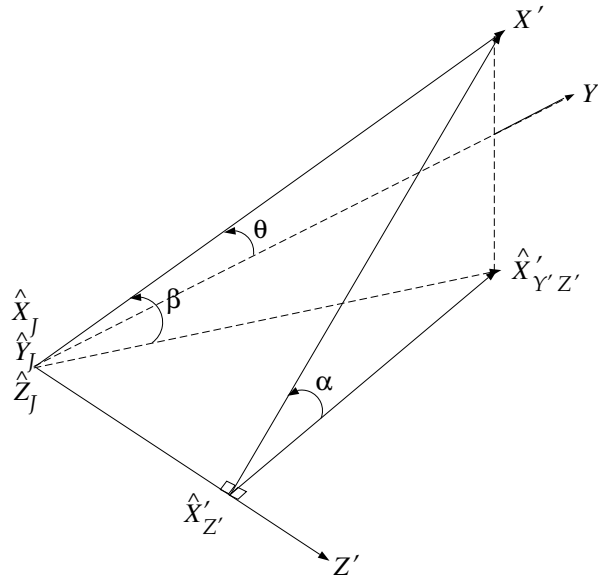


FIG. 2.4 - *Corrélations simple, multiple et partielle.*

La figure FIG. 2.4 visualise dans l'espace des variables centrées.

1. Le coefficient de corrélation simple de X et Y . C'est le cosinus de l'angle que font les vecteurs X' et Y' ($\cos(\theta)$).
2. Le coefficient de corrélation multiple de X et Y, Z . C'est le coefficient de corrélation de X avec Y et Z considérés conjointement. C'est donc le cosinus de l'angle des vecteurs X' et $\hat{X}'_{Y'Z'}$ ($\cos(\beta)$).
3. Le coefficient de corrélation partielle de X et Y . C'est le coefficient de corrélation de X et Y quand l'effet de Z est éliminé. Le vecteur Z' étant orthogonal à $X' - \hat{X}'_{Z'}$ et à $\hat{X}'_{Y'Z'} - \hat{X}'_{Z'}$, le coefficient de corrélation partielle est le cosinus de l'angle que font ces deux derniers vecteurs ($\cos(\alpha)$).

Illustration numérique

Soit la matrice :

$$X = \begin{bmatrix} 10 & 5 & 41 \\ 20 & 15 & 59 \\ 35 & 20 & 69 \end{bmatrix}$$

Les lignes représentent des individus (notés *ind.1*, *ind.2*, *ind.3*), les colonnes représentent des variables. De la première à la troisième colonne on a les va-

riables pression en atmosphères (P), température en degrés centigrades (T_c) et température en degrés Fahrenheit (T_f). On veut vérifier graphiquement que :

1. l'angle que fait T_c avec P est différent de l'angle que fait T_f avec P
2. les angles que font les variables centrées T'_c et T'_f avec P' sont identiques.

Sur FIG. 2.5 on a représenté les vecteurs T_c , T_f et P . On observe que T_c et T_f ne sont pas colinéaires.

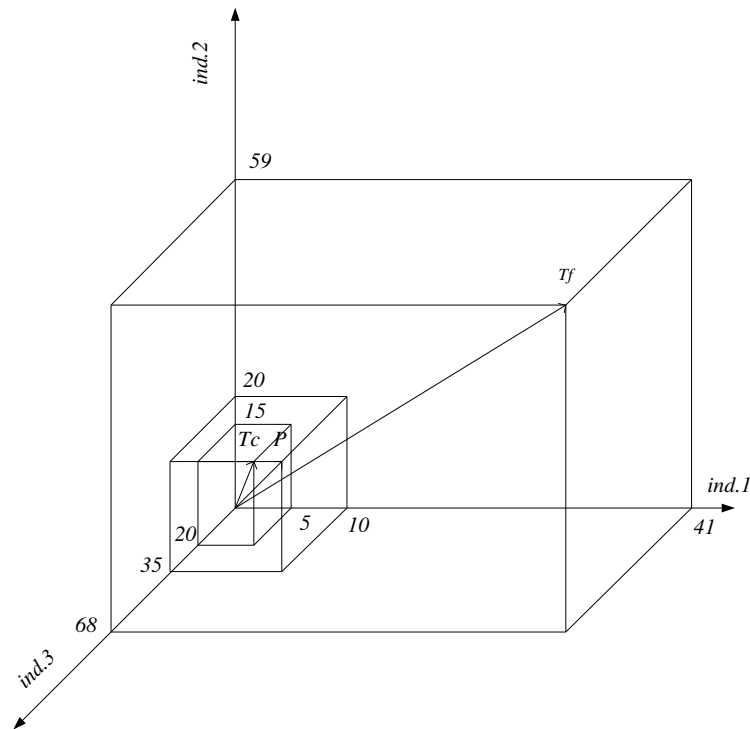
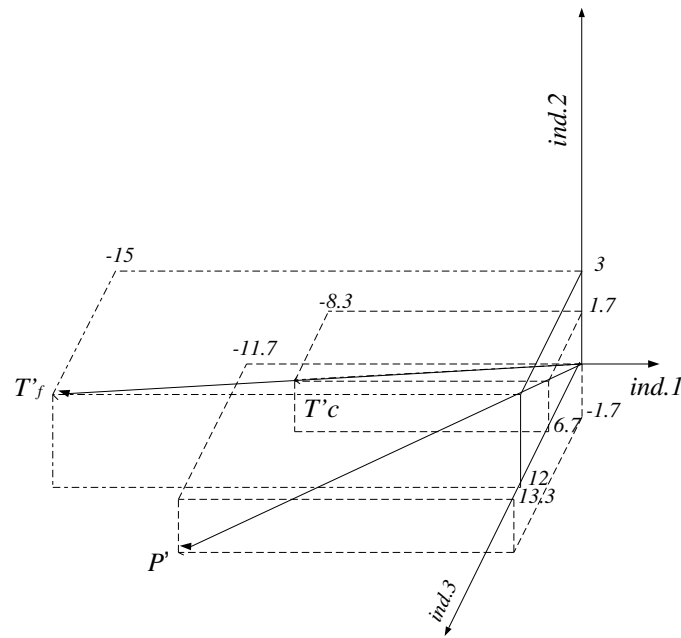


FIG. 2.5 - Représentation de T_c , T_f et P .

Soient P' , T'_c et T'_f les variables centrées. La matrice des observations centrées est :

$$X' = \begin{bmatrix} -11,7 & -8,4 & -15 \\ -1,7 & 1,7 & 3 \\ 13,4 & 6,7 & 12 \end{bmatrix}$$

Sur la figure FIG. 2.6 on a représenté les vecteurs des variables centrées, on observe que T'_c et T'_f sont colinéaires.

FIG. 2.6 - Représentation de T'_c , T'_f et P' .

2.1.3 Expliquer ou prédire des variables

Considérons le modèle de régression $Y = \alpha + \beta X + \dots$ dans lequel X, Y, \dots sont interprétés comme des vecteurs. Le critère des moindres carrés consiste à choisir α, β, \dots tels que $|Y - \alpha J - \beta X - \dots|^2$ soit minimum. Si P est l'hyperplan déterminé par J, X, \dots alors la meilleure estimation de $\alpha J + \beta X + \dots$ est la projection orthogonale \hat{Y}_P de Y sur P . Pour mesurer la qualité de l'estimation on utilise soit $SCE = |Y - \hat{Y}_P|^2$ soit l'angle que font $Y - \hat{Y}_P$ et $\hat{Y}_P - \hat{Y}_J$.

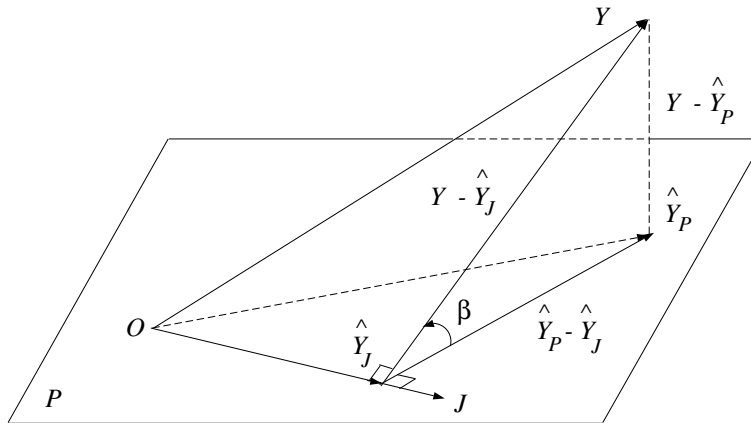


FIG. 2.7 - Régression et coefficient de corrélation.

Sur la figure FIG. 2.7, on a représenté l'angle β que font les vecteurs $Y - \hat{Y}_P$ et $\hat{Y}_P - \hat{Y}_J$. Ces deux vecteurs sont dans le sous espace des variables centrées

car ils sont tous les deux orthogonaux à J . Une fonction de cet angle peut donc servir pour mesurer la qualité du modèle utilisé. En particulier, $\cos^2(\beta)$ qui n'est autre que le rapport des carrés des longueurs des vecteurs $\hat{Y}_P - \hat{Y}_J$ et $Y - \hat{Y}_J$. $\cos^2(\beta)$ est le carré du coefficient de corrélation multiple c'est, le *coefficient de détermination* que l'on note R^2 .

Le coefficient de détermination est donc le rapport du carré de la longueur d'un côté de l'angle droit et du carré de la longueur de l'hypoténuse du triangle limité par les extrémités Y , \hat{Y}_P et \hat{Y}_J des vecteurs OY , $O\hat{Y}_P$ et $O\hat{Y}_J$. Dans ces conditions, il est clair, que le coefficient de détermination est, au même titre que $|Y - \hat{Y}_P|^2$, une mesure de la qualité du modèle utilisé. En effet, le théorème de pythagore permet lui aussi d'exprimer l'erreur qu'on commet quand on utilise un modèle car $|Y - \hat{Y}_P|^2$ est lié à $|Y - \hat{Y}_J|^2$ et $|\hat{Y}_P - \hat{Y}_J|^2$ par la relation $|Y - \hat{Y}_P|^2 = |Y - \hat{Y}_J|^2 + |\hat{Y}_P - \hat{Y}_J|^2$.

Notion de dualité

On illustre la notion de dualité sur l'exemple de la droite de régression (Bachacou *et al.*, 1981). Classiquement, on estime les paramètres α et β de la droite de régression de Y en X sur un échantillon de n individus ($y_i = \alpha + \beta x_i + e_i$; $i = 1, \dots, n$) par la méthode des moindres carrés, en se plaçant dans l'espace des individus (\mathbb{R}^2). Nous allons montrer que ces estimations peuvent aussi être obtenues en raisonnant dans l'espace des variables (\mathbb{R}^n). Soient OY , OX et OJ les vecteurs de composantes (y_1, \dots, y_n) , (x_1, \dots, x_n) et $(1, \dots, 1)$, nous cherchons le vecteur $O\hat{Y} = \alpha OJ + \beta OX$ tel que $|OY - O\hat{Y}|^2$ soit minimum ($O\hat{Y}$ projection orthogonale de OY sur le plan défini par OJ et OX).

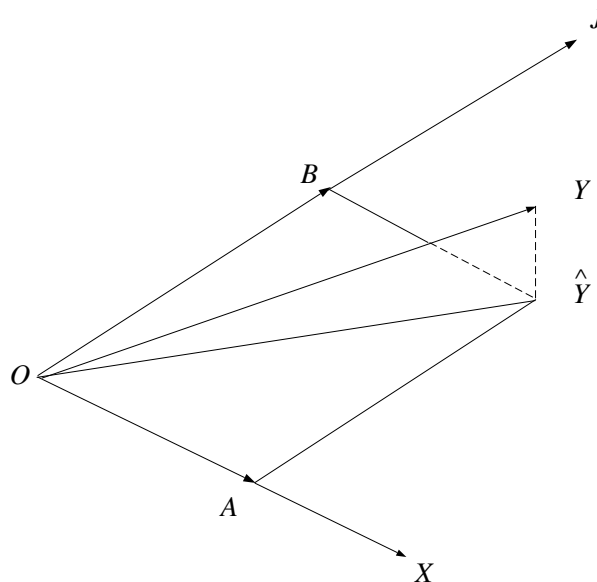


FIG. 2.8 - Projection sur le plan XOJ .

On cherche donc α et β tels que $OA = \beta OX$ et $OB = \alpha OJ$

1 - Estimation de α

On projette orthogonalement OY , $O\hat{Y}$, OX et OA sur OJ (FIG. 2.9. D'après le théorème des trois perpendiculaires, Y et \hat{Y} se projettent au même point \bar{Y} ; X et A se projettent respectivement en \bar{X} et \bar{A} . De $OB = A\hat{Y}$ (FIG. 2.8) et $A\hat{Y} = \bar{A}\bar{Y}$ (FIG. 2.9), on déduit que :

$$OB = O\bar{Y} + \bar{Y}B = O\bar{Y} - O\bar{A} \text{ car } \bar{Y}B = -O\bar{A}$$

$$OB = O\bar{Y} - \beta O\bar{X} \implies \alpha = \bar{y} - \beta \bar{x}$$

$$\text{avec } \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i, \quad \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

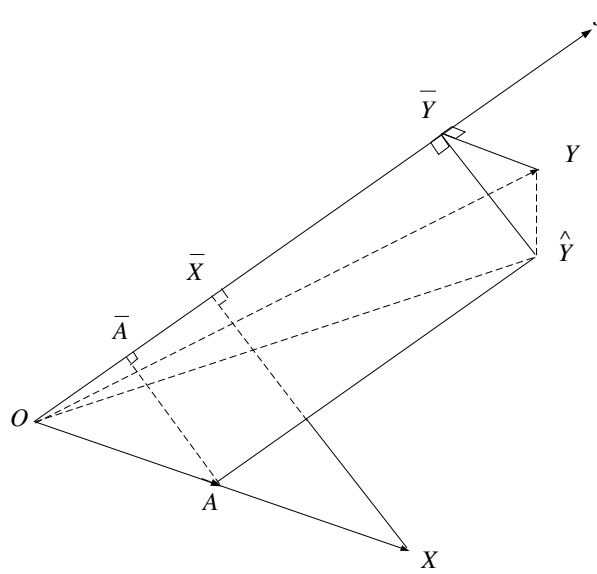


FIG. 2.9 - Projection sur OJ .

2 - Estimation de β

On projette orthogonalement Y , \hat{Y} , X et A sur le sous espace des variables centrées orthogonal à OJ . On obtient les vecteurs OX' et OY' des variables centrées ainsi que la projection orthogonale de A et de \hat{Y} en A' (FIG. 2.10).

La régression sur les variables centrées se résout en projetant orthogonalement OY' sur OX' . Comme OA' est le projeté de OY' sur OX' , on écrit que le produit scalaire des vecteurs OA' et $OY' - OA'$ est nul :

$$\langle OA', OY' - OA' \rangle = \langle \beta OX', OY' - \beta OX' \rangle$$

$$\langle \beta OX', OY' - \beta OX' \rangle = \beta \langle OX', OY' \rangle - \beta^2 \langle OX', OX' \rangle = 0$$

$$\beta = \frac{\langle OX', OY' \rangle}{\langle OX', OX' \rangle} = \frac{\langle OX', OY' \rangle}{\langle OX', OX' \rangle^{1/2} \langle OY', OY' \rangle^{1/2}} \times \frac{\langle OY', OY' \rangle^{1/2}}{\langle OX', OX' \rangle^{1/2}}$$

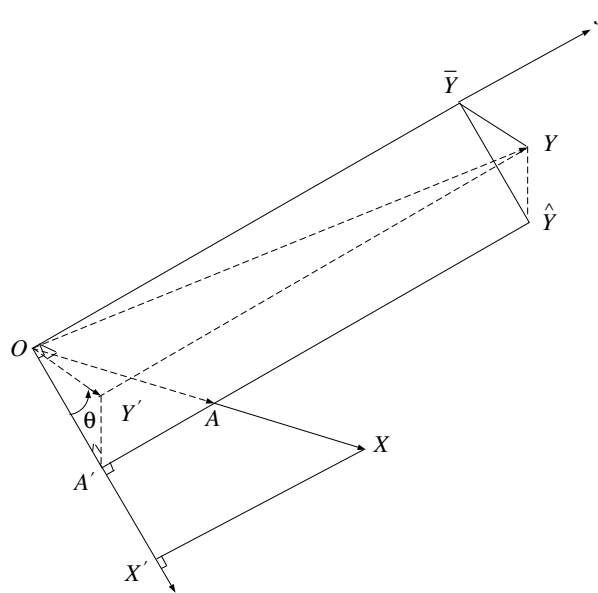


FIG. 2.10 - Projections sur le sous espace des variables centrées.

$$\beta = \rho \frac{S_y}{S_x}$$

ρ est l'estimation du coefficient de corrélation, S_y et S_x les estimations des écarts types des y et des x . En effet :

$$\langle OX, OY' \rangle = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

$$\langle OX, OX' \rangle = \sum_{i=1}^n (x_i - \bar{x})^2$$

$$\langle OY, OY' \rangle = \sum_{i=1}^n (y_i - \bar{y})^2$$

Le sous espace des variables centrées est de dimension $n - 1$ (*degrés de liberté*).

Les résultats ainsi obtenus sont identiques à ceux qu'on obtient quand on raisonne sur l'espace des individus.

2.1.4 Comparer des populations

On veut comparer deux populations P_1 et P_2 . Soit Y une variable aléatoire observée sur P_1 et P_2 : $Y = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix}$. Supposons Y_1 et Y_2 normalement distribuées ($Y_1 \sim N(\mu_1, \sigma_1^2)$ et $Y_2 \sim N(\mu_2, \sigma_2^2)$) et notons y_{11}, \dots, y_{1n_1} et y_{21}, \dots, y_{2n_2} les n_1 et n_2 observations de Y_1 et Y_2 respectivement.

Pour comparer les moyennes des populations P_1 et P_2 :

1. On estime les moyennes (μ_1 et μ_2) et la variance (σ^2). Soient $y_1.$ et $y_2.$ les estimateurs de μ_1 et μ_2 ; S_1^2 , S_2^2 les estimateurs des variances des populations P_1 , P_2 et S^2 l'estimateur de la variance commune aux deux populations :

$$\text{Moyennes : } y_1. = \sum_{j=1}^{n_1} \frac{y_{1j}}{n_1} \quad y_2. = \sum_{j=1}^{n_2} \frac{y_{2j}}{n_2}$$

$$\text{Variances : } S_1^2 = \sum_{j=1}^{n_1} \frac{(y_{1j} - y_1.)^2}{(n_1 - 1)} \quad S_2^2 = \sum_{j=1}^{n_2} \frac{(y_{2j} - y_2.)^2}{(n_2 - 1)}$$

$$\begin{aligned} \text{Variance commune : } S^2 &= \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} \\ S^2 &= \frac{\sum_{j=1}^{n_1} (y_{1j} - y_1.)^2 + \sum_{j=1}^{n_2} (y_{2j} - y_2.)^2}{n_1 + n_2 - 2} \end{aligned}$$

2. On utilise la statistique de test :

$$T = \frac{y_1. - y_2.}{S\left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{\frac{1}{2}}} \implies T^2 = \frac{n_1 \times n_2}{n_1 + n_2} (y_1. - y_2.) S^{-2} (y_1. - y_2.)$$

T est la statistique de Student à $n_1 + n_2 - 2$ degrés de liberté.

3. On interprète géométriquement la statistique de Student.

Considérons le terme $(y_1. - y_2.) S^{-2} (y_1. - y_2.)$ de T^2 , il contient :

- (a) le terme $(y_1. - y_2.) (y_1. - y_2.)$ qui n'est autre que le carré de la distance entre les points moyens des deux populations P_1 et P_2 .
- (b) le terme S^{-2} .

Donc dans T^2 , le carré de la distance entre les points moyens des deux populations est pondéré par l'inverse de la variance commune aux deux populations. Faisons le changement de variable $y'_1. = y_1./S$ et $y'_2. = y_2./S$ alors :

$$(y_1. - y_2.) S^{-2} (y_1. - y_2.) = (y'_1. - y'_2.) (y'_1. - y'_2.)$$

Le deuxième membre de cette expression est le carré d'une distance au sens où nous l'avons défini jusqu'à présent. On en déduit que le premier membre est aussi le carré d'une distance. Comme une distance dérive d'un produit scalaire, on peut définir le produit scalaire de deux vecteurs U et V par $\langle U, V \rangle = U' M V$ avec U' le vecteur U transposé et M une matrice *symétrique définie positive* désignée sous le nom de *métrique* (ces notions seront définies ultérieurement).

Dans le cas qui nous intéresse, $U = U' = V = y_{1.} - y_{2.}$ et $M = S^{-2}$. Ainsi, le carré de la statistique de Student n'est autre (au coefficient $n_1 n_2 / (n_1 + n_2)$ près) qu'une distance au carré qui prend en compte la variabilité des observations par le truchement d'une métrique dite *intra-population*. Cette notion de métrique est essentielle en analyse multidimensionnelle (analyses factorielles, classifications). En analyse de variance multidimensionnelle, on retrouve un équivalent de la distance mise en évidence dans la statistique de Student, c'est la distance dite de Mahalanobis. D'un point de vue géométrique :

(a) on représente les variables par des vecteurs. Soit :

$$Y = \begin{bmatrix} y_{11} \\ \vdots \\ y_{1n_1} \\ y_{21} \\ \vdots \\ y_{2n_2} \end{bmatrix} \quad \hat{Y}_P = \begin{bmatrix} y_{1.} \\ \vdots \\ y_{1.} \\ y_{2.} \\ \vdots \\ y_{2.} \end{bmatrix} \quad \hat{Y}_J = \begin{bmatrix} y_{..} \\ \vdots \\ y_{..} \\ y_{..} \\ \vdots \\ y_{..} \end{bmatrix} \quad J = \begin{bmatrix} 1 \\ \vdots \\ 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \quad X_1 = \begin{bmatrix} 1 \\ \vdots \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad X_2 = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

Ces vecteurs sont dans des espaces vectoriels :

- $Y \in \mathbb{R}^n$ (espace de dimension $n = n_1 + n_2$).
- $\hat{Y}_P \in P$ (sous espace de \mathbb{R}^n de dimension 2 engendré par X_1 et X_2).
- $\hat{Y}_J \in$ sous espace de P , de dimension 1, engendré par J .
- $Y - \hat{Y}_P \in P^\perp$ (sous espace orthogonal à P , de dimension $n - 2$).
- $\hat{Y}_P - \hat{Y}_J \in$ sous espace de P , orthogonal à J , de dimension 1.

On les représente par (FIG. 2.11) :

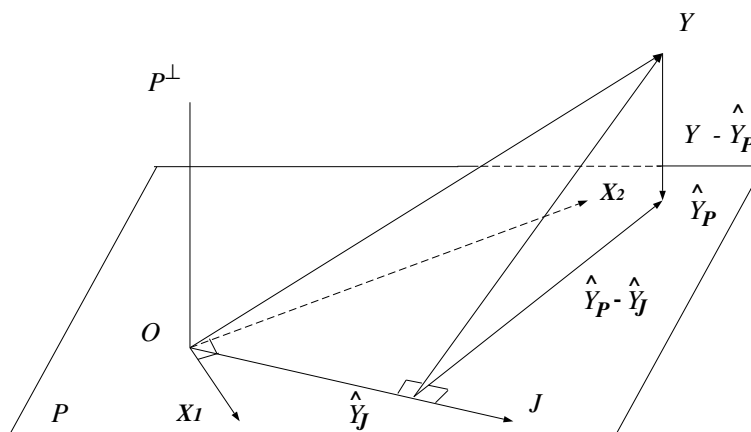


FIG. 2.11 - Comparaison de deux populations P_1 et P_2 .

- (b) pour chaque observation, on écrit que la valeur observée est égale à la moyenne générale (m) plus un écart à la moyenne générale spécifique à chaque population (m_1 pour les observations de P_1 et m_2 pour les observations de P_2) plus un terme d'erreur. Soit :

$$\begin{aligned} y_{11} &= m + m_1 + e_{11} \\ \dots &\quad \dots \quad \dots \quad \dots \\ y_{1n_1} &= m + m_1 + e_{1n_1} \\ y_{21} &= m + m_2 + e_{21} \\ \dots &\quad \dots \quad \dots \quad \dots \\ y_{2n_2} &= m + m_2 + e_{2n_2} \end{aligned}$$

L'écriture matricielle de ce système est :

$$\begin{bmatrix} y_{11} \\ \vdots \\ y_{1n_1} \\ y_{21} \\ \vdots \\ y_{2n_2} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ \vdots & \vdots & \vdots \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ \vdots & \vdots & \vdots \\ 1 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} m \\ m_1 \\ m_2 \end{bmatrix} + \begin{bmatrix} e_{11} \\ \vdots \\ e_{1n_1} \\ e_{21} \\ \vdots \\ e_{2n_2} \end{bmatrix}$$

On l'écrit symboliquement : $Y = X \times \Theta + E$

Pour estimer les paramètres (m , m_1 , m_2) et la variance (S^2), on minimise :

$$|E|^2 = |Y - \hat{Y}_P|^2 = \langle Y - X\Theta, Y - X\Theta \rangle$$

- (c) pour éprouver l'hypothèse que les deux populations P_1 et P_2 sont identiques ($\mu_1 = \mu_2$) contre l'hypothèse alternative, elles sont différentes ($\mu_1 \neq \mu_2$), on évalue $|\hat{Y}_P - \hat{Y}_J|^2$ et $|Y - \hat{Y}_P|^2$ (si P_1 et P_2 sont identiques, les écarts (m_1 et m_2) à la moyenne m sont nuls et Y se projette sur J). Soit :

$$|\hat{Y}_P - \hat{Y}_J|^2 = \langle \hat{Y}_P - \hat{Y}_J, \hat{Y}_P - \hat{Y}_J \rangle = \sum_{i=1}^2 (y_{i.} - y_{..})^2 = \frac{n_1 \times n_2}{n_1 + n_2} (y_{1.} - y_{2.})^2$$

$$\text{avec } y_{..} = \frac{1}{n_1 + n_2} (n_1 y_{1.} + n_2 y_{2.})$$

$$|Y - \hat{Y}_P|^2 = \langle Y - \hat{Y}_P, Y - \hat{Y}_P \rangle = \sum_{j=1}^{n_1} (y_{1j} - y_{1.})^2 + \sum_{j=1}^{n_2} (y_{2j} - y_{2.})^2$$

On utilise la statistique F de Fisher à 1 et $n_1 + n_2 - 2$ degrés de liberté :

$$F = \frac{\frac{|\hat{Y}_P - \hat{Y}_J|^2}{1}}{\frac{|Y - \hat{Y}_P|^2}{n_1 + n_2 - 2}} = \frac{(y_{1.} - y_{2.})^2}{S^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}$$

Conclusion

1. Les sommes des carrés sont des longueurs de vecteurs au carré.
2. Les degrés de liberté sont les dimensions des espaces qui contiennent les vecteurs.
3. L'hypothèse testée est :

$$H_0 : \mu_1 = \mu_2 = \mu$$

$$\bar{H}_0 : \mu_1 \neq \mu_2$$
 Sous H_0 , F suit une loi de Fisher $F_{(1, n_1+n_2-2)}$ degrés de liberté. Si F est supérieure à une certaine valeur (lue dans une table) \iff Si $|\hat{Y}_P - \hat{Y}_J|$ est grand, on rejette H_0 .
4. La statistique de Fisher à 1 et $n_1 + n_2 - 2$ degrés de liberté est identique au carré de la statistique de Student à $n_1 + n_2 - 2$ degrés de liberté.

2.2 Probabilité et Géométrie

A condition de bien identifier les notions de vecteur et de produit scalaire, on peut raisonner géométriquement avec des variables aléatoires. Supposons que nous avons des variables aléatoires X, Y, Z, \dots conjointement distribuées $F(x, y, z, \dots)$.

2.2.1 Espérance mathématique

L'espérance d'une variable aléatoire est calculée par :

$$E(X) = \int x dF(x)$$

avec $F()$ la fonction de répartition des variables impliquées.

Supposons que les variables aléatoires X, Y, \dots ont pour moyenne $E(X) = \mu_X, E(Y) = \mu_Y, \dots$. Ces variables aléatoires peuvent être représentées par des vecteurs qu'on peut additionner et multiplier par des scalaires. La quantité :

$$\langle X, Y \rangle = E(XY) = \int xy dF(x, y)$$

définit le produit scalaire de X et Y .

Ici, l'équivalent du vecteur J (vecteur dont toutes les composantes sont égales à 1) est la constante $\mathbf{1}$. Par exemple, la projection d'une variable aléatoire X

sur $\mathbf{1}$ doit être telle que $\hat{X}_1 = \alpha \mathbf{1}$ quelque soit α (cf. propriété P_1 du § 1.3) et telle que $\langle X - \hat{X}_1, \mathbf{1} \rangle = 0$ (cf. propriété P_2 du § 1.3). On en déduit que $E(X) = \alpha$ de telle sorte que $\mu_X \mathbf{1} = \mu_X$ est la projection de X sur $\mathbf{1}$.

Le vecteur $X - \mu_X$ est une variable aléatoire centrée.

2.2.2 Variance

Pour toute variable aléatoire X , $\langle X, X \rangle = E(X^2)$ est la longueur au carré du vecteur X . Le carré de la longueur de la variable aléatoire $X - \mu_X$ est :

$$|X - \mu_X|^2 = \langle X - \mu_X, X - \mu_X \rangle = E(X - \mu_X)^2 = \sigma_X^2$$

c'est la variance de X .

2.2.3 Corrélation

L'angle θ que font les variables centrées $X - \mu_X$ et $Y - \mu_Y$ joue un rôle important car :

$$\cos(\theta) = \frac{\langle X - \mu_X, Y - \mu_Y \rangle}{|X - \mu_X| |Y - \mu_Y|}$$

représente la corrélation entre X et Y .

Quand en théorie des probabilités on écrit :

$$\text{var}(X + Y) = \text{var}(X) + \text{var}(Y) + 2\text{cov}(X, Y)$$

on ne fait qu'écrire la loi du cosinus.

Dire que deux variables aléatoires ne sont pas corrélées est équivalent à dire que les vecteurs correspondants sont orthogonaux. Les interprétations géométriques sont analogues à celles qui ont été faites précédemment, la seule différence est qu'ici, on interprète un modèle théorique alors qu'en statistique, on interprète les données.

2.2.4 Espérance conditionnelle

D'un point de vue plus général, l'espérance conditionnelle de X sachant Y, Z, \dots , notée $E(X|Y, Z, \dots)$, est identique à la projection de X sur un plan P déterminé par Y, Z, \dots . En fait, P est l'espace des fonctions de Y, Z, \dots de carré intégrable, il est fermé pour les opérations linéaires. C'est pour cette raison qu'il est interprété comme un sous espace linéaire.

Considérons les trois propriétés des espérances conditionnelles :

$$E(X|Y, Z, \dots) \text{ est une fonction de } Y, Z, \dots \quad (2.4)$$

$$E[E(X|Y, Z, \dots)] = E(X) \quad (2.5)$$

Si g est une fonction de Y, Z, \dots , alors :

$$E(Xg|Y, Z, \dots) = gE(X|Y, Z, \dots) \quad (2.6)$$

De (2.5) et (2.6), on déduit que si g est une fonction de Y, Z, \dots alors :

$$\langle g, X - E(X|Y, Z, \dots) \rangle = 0 \quad (2.7)$$

Donc, (2.4) dit que $E(X|Y, Z, \dots)$ est dans le plan P , tandis que (2.7) précise que $X - E(X|Y, Z, \dots)$ est orthogonal à tout vecteur g de P donc en particulier à $E(X|Y, Z, \dots)$.

L'espérance conditionnelle est donc l'analogie d'une projection (FIG. 2.12).

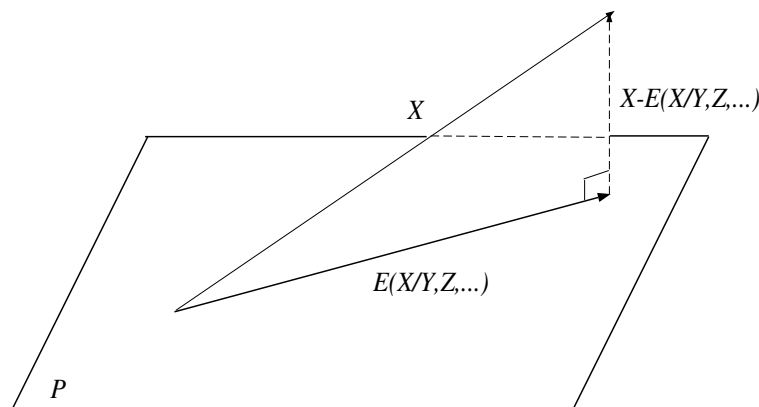


FIG. 2.12 - *Espérance conditionnelle.*

Les concepts géométriques permettent d'interpréter la projection \hat{X}_P de X sur P comme le point de P le plus proche de X . Ainsi, si Y, Z, \dots représentent la connaissance et X une valeur à prédire, alors la projection est la fonction de la connaissance qui prédit le mieux X car elle donne le plus petit carré moyen d'erreur.

2.2.5 Variance et variance conditionnelle

Un autre théorème de la théorie des probabilités dit que :

$$\text{var}(X) = \text{var}[E(X|Y, Z, \dots)] + E[\text{var}(X|Y, Z, \dots)] \quad (2.8)$$

D'après (2.5), $E[X - E(X|Y, Z, \dots)] = 0$ alors :

$$\text{var}[X - E(X|Y, Z, \dots)] = E[X - E(X|Y, Z, \dots)]^2$$

Puisque la variance conditionnelle est la variance relative à la distribution conditionnelle, (2.5) implique que c'est aussi égal à $E[\text{var}(X|Y, Z, \dots)]$. Donc (2.8) s'écrit :

$$\text{var}(X) = \text{var}[E(X|Y, Z, \dots)] + \text{var}[X - E(X|Y, Z, \dots)]$$

qui n'est autre qu'une extension du théorème de Pythagore.

Exemple

Il y a différentes façons de démontrer qu'une combinaison linéaire de deux variables aléatoires normales, indépendantes suit une loi normale (Anderson, 1958 ; Mood *et al.*, 1974). Cette démonstration proposée par Hombas (1989) a pour objectif d'appliquer les notions de géométrie analytique qui ont été présentées, elle a le mérite d'être plus simple et plus imagée que les démonstrations faites classiquement.

- Soient X_1 et X_2 deux variables aléatoires distribuées indépendamment suivant une loi normale centrée réduite $N(0, 1)$. Soit $X = (X_1, X_2)$.

La densité conjointe de X_1 et X_2 est :

$$f_X(x_1, x_2) = (2\pi)^{-1} \exp(-(x_1^2 + x_2^2)/2) \quad (-\infty < x_1 < \infty, -\infty < x_2 < \infty)$$

- Soient X'_1 et X'_2 deux autres variables aléatoires, indépendantes et distribuées normalement $N(0, 1)$. Soit $X' = (X'_1, X'_2)$.

La densité conjointe de X'_1 et X'_2 est :

$$f_{X'}(x'_1, x'_2) = (2\pi)^{-1} \exp(-(x'^2_1 + x'^2_2)/2) \quad (2.9)$$

$$(-\infty < x'_1 < \infty, -\infty < x'_2 < \infty)$$

- Dans les deux cas, les densités conjointes ont pour contours des cercles centrés sur 0. Pour les espaces échantillons (X_1, X_2) et (X'_1, X'_2) qui sont coplanaires, les surfaces définies par les densités de probabilité coïncident si $x'^2_1 + x'^2_2 = x^2_1 + x^2_2$. Une transformation $(X_1, X_2) \longrightarrow (X'_1, X'_2)$ qui satisfait cette condition est la rotation orthogonale montrée sur FIG. 2.13 :

$$\begin{aligned} X'_1 &= X_1 \cos\theta + X_2 \sin\theta \\ X'_2 &= -X_1 \sin\theta + X_2 \cos\theta \end{aligned} \quad (2.10)$$

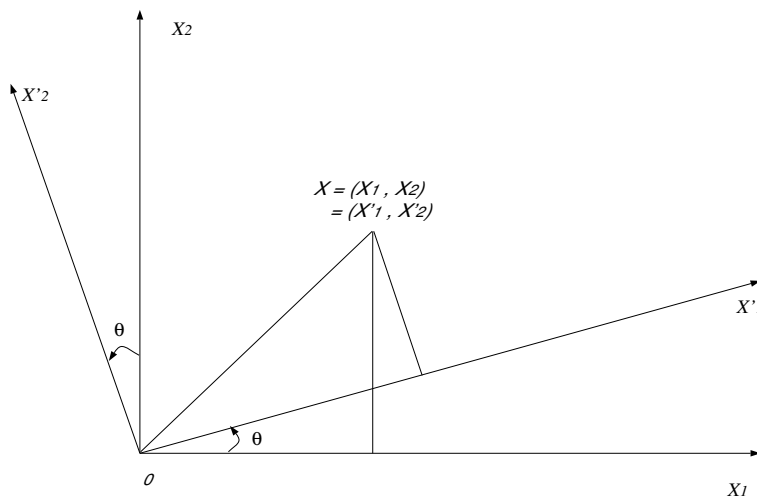


FIG. 2.13 - Rotation orthogonale dans le plan.

- La factorisation de (2.9) donne :

$$f_{X'}(x'_1, x'_2) = (2\pi)^{-1/2} \exp(-x'^2_1/2) \times (2\pi)^{-1/2} \exp(-x'^2_2/2)$$

Puisque $X'_1 \sim N(0, 1)$ il en est de même pour $X_1 \cos \theta + X_2 \sin \theta \sim N(0, 1)$. De plus, X'_1 et X'_2 sont indépendantes.

- Donc les deux variables indépendantes, normalement distribuées $N(0, 1)$, obéissent à la propriété d'invariance rotationnelle suivante : la rotation orthogonale d'un système de coordonnées déterminé par deux variables aléatoires indépendantes $N(0, 1)$ préserve l'indépendance et la normalité.

Cette discussion prouve aussi que la variable aléatoire $X' = \alpha_1 X_1 + \alpha_2 X_2$ avec $\alpha_1^2 + \alpha_2^2 = 1$, qui représente l'une des composantes de (2.10), est elle-même normalement distribuée $N(0, 1)$.

Considérons maintenant les scalaires β_1 et β_2 et soient α_1 et α_2 tels que :

$$\alpha_1 = \frac{\beta_1}{(\beta_1^2 + \beta_2^2)^{1/2}} \quad \text{et} \quad \alpha_2 = \frac{\beta_2}{(\beta_1^2 + \beta_2^2)^{1/2}}$$

Ils satisfont à $\alpha_1^2 + \alpha_2^2 = 1$. Ainsi, si X_1 et X_2 sont $N(0, 1)$, on déduit que :

$$\frac{(\beta_1 X_1 + \beta_2 X_2)}{(\beta_1^2 + \beta_2^2)^{1/2}} \sim N(0, 1)$$

Donc :

$$\beta_1 X_1 + \beta_2 X_2 \sim N(0, \alpha_1^2 + \alpha_2^2)$$

Soient X et Y deux variables aléatoires normales indépendantes $N(\mu_i, \sigma_i^2)$, $i = 1, 2$. Alors, la variable $Z = \beta_1 X + \beta_2 Y$ peut être écrite :

$$Z = \beta_1 \sigma_1 (X - \mu_1) / \sigma_1 + \beta_2 \sigma_2 (Y - \mu_2) / \sigma_2 + \beta_1 \mu_1 + \beta_2 \mu_2 \quad (2.11)$$

Les variables aléatoires $(X - \mu_1) / \sigma_1$ et $(Y - \mu_2) / \sigma_2$ sont $N(0, 1)$. En appliquant à (2.11) les résultats précédents, on déduit que :

$$Z \sim N(\beta_1 \mu_1 + \beta_2 \mu_2, \beta_1^2 \sigma_1^2 + \beta_2^2 \sigma_2^2)$$

La généralisation à plus de deux variables aléatoires est évidente.

2.3 Tableau résumé et compléments bibliographiques

Espace vectoriel	Expression géométrique		Expression	
	Représentation	Forme analytique	Statistique	Probabiliste
Vecteur U	Segment orienté de 0 vers U	$U = (u_1, u_2, \dots, u_n)$	Données u_1, u_2, \dots, u_n	V. aléatoire U
Vecteur $U - V$	Segment orienté de V vers U	$U - V = (u_1 - v_1, \dots, u_n - v_n)$	Différences $u_1 - v_1, \dots$	V. aléatoire $U - V$
$\langle U, V \rangle$	Mesure de non orthogonalité	$u_1 v_1 + \dots + u_n v_n$	Somme des produits	$E(UV)$
$\langle U, U \rangle$	Longueur au carré	$u_1^2 + \dots + u_n^2$	Somme des carrés	$E(U^2)$
$ U - V ^2$	Distance au carré	$(u_1 - v_1)^2 + \dots$	Somme carrés différences	$E(U - V)^2$
$\langle U, V \rangle = 0$	Orthogonalité: U, V perpendiculaires			
$\alpha U + \beta V + \dots$	Hyperplans contenant O , engendrés par U, V, \dots		Modèles $\alpha U + \beta V + \dots$	Fonctions de carré intégrable
$\langle U - \hat{U}, V \rangle = 0$ $\forall V$ dans P	Projection de U sur P		Meilleur modèle $\alpha U + \beta V + \dots$	$E(U V, W, \dots)$
Vecteur J	Segment orienté incliné à 45°	$J = (1, 1, \dots, 1)$	Constante $(1, 1, \dots, 1)$	Constante 1
$ U - \hat{U}_J ^2$		$\sum (u_i - \bar{u})^2$	$(n - 1) \times$ variance d'échantillonnage de U	Variance de U
$\frac{\langle U - \hat{U}_J, V - \hat{V}_J \rangle}{ U - \hat{U}_J \cdot V - \hat{V}_J }$	Cosinus de l'angle entre $U - \hat{U}_J$ et $V - \hat{V}_J$	Corrélation d'échantillonnage de U et V		Corrélation de U et V
$\langle U - \hat{U}_J, V - \hat{V}_J \rangle = 0$		$U - \hat{U}_J$ et $V - \hat{V}_J$ perpendiculaires	U et V non corrélées	U et V non corrélées

TAB. 2.1 - Correspondance entre Espace vectoriel, Géométrie, Statistique et Probabilité.

La présentation géométrique des concepts utilisés en statistique a fait l'objet de quelques articles. Elle a été initiée par Fisher (1915) avec son article sur la distribution du coefficient de corrélation, elle s'est poursuivie jusqu'à la parution de l'article de Kruskal (1975) sur la géométrie des inverses généralisées. L'impopularité relative de cette approche a pour origine l'idée fautive que géométrie est synonyme d'abstraction alors que, au contraire, elle permet

de mieux comprendre la signification des objets manipulés par la statistique. De ce dernier point de vue, les articles de Herr (1980) et de Margolis (1979) présentent de l'intérêt.

Chapitre 3

Formalisation des concepts

3.1 Espace vectoriel

Un espace vectoriel de dimension p sur le corps des réels \mathbb{R} est une construction mathématique dont les éléments sont des vecteurs. Il est défini par deux opérations :

- *L'addition.* Soient U et V deux vecteurs de l'espace, $W = U + V$ est unique et appartient à l'espace.
- *La multiplication par un scalaire.* Soit U un vecteur et λ un nombre réel (scalaire), $Z = \lambda U$ est unique et appartient à l'espace.

L'addition et la multiplication par un scalaire obéissent aux règles suivantes :

- *L'addition*
 - 1- Il existe un vecteur nul (noté ϕ) tel que :
 $\phi + V = V + \phi = V$ pour tout V .
 - 2- Pour chaque V , il existe un vecteur opposé $-V$ tel que :
 $V + (-V) = \phi$
 - 3- Le vecteur addition est commutatif :
 $U + V = V + U$ pour tout U et tout V .
 - 4- L'addition est associative :
 $(V + U) + W = V + (U + W)$ pour tout U, V et W .
- *La multiplication par un scalaire*

Pour tout scalaire α, β et tout vecteur U, V , les multiplications de vecteurs par des scalaires obéissent à :

$$\alpha(U + V) = \alpha U + \alpha V.$$

$$(\alpha + \beta)V = \alpha V + \beta V.$$

$$(\alpha\beta)V = \alpha(\beta V).$$

$$\mathbf{1}V = V \text{ (où } \mathbf{1} \text{ est la constante } \mathbf{1}\text{)}.$$

– De ces axiomes de base on peut déduire des règles de travail telles que :

$$0V = \phi \text{ pour tout } V.$$

$$\alpha\phi = \phi \text{ pour tout } \alpha.$$

$$(-1)V = -V \text{ pour tout } V.$$

3.1.1 Sous espace vectoriel

Un sous ensemble de vecteurs d'un espace vectoriel est un *sous espace vectoriel* si pour les opérations définies sur tout l'espace il forme un espace vectoriel (exemple : le sous espace dont le seul élément est ϕ).

3.1.2 Combinaison linéaire

Le vecteur V est une *combinaison linéaire* des vecteurs U_1, U_2, \dots, U_r s'il existe des scalaires $\alpha_1, \alpha_2, \dots, \alpha_r$ tels que :

$$V = \alpha_1 U_1 + \alpha_2 U_2 + \dots + \alpha_r U_r$$

3.1.3 Indépendance linéaire

Les vecteurs U_1, U_2, \dots, U_r sont dits *linéairement indépendants* si la relation :

$$\alpha_1 U_1 + \alpha_2 U_2 + \dots + \alpha_r U_r = \phi$$

implique que $\alpha_1 = \alpha_2 = \dots = \alpha_r = 0$. Autrement, ils sont linéairement dépendants et au moins l'un d'entre eux peut être exprimé comme une combinaison des autres.

3.1.4 Base d'un espace vectoriel

1- Une *base* d'un espace vectoriel est un ensemble de vecteurs linéairement indépendants qui engendrent l'espace dans sa totalité.

2- Tout vecteur est défini de façon *unique* comme une combinaison linéaire des vecteurs d'une base donnée.

- 3- Un espace vectoriel a toujours une base (elle n'est pas unique).
- 4- Chaque base a le même nombre d'éléments. Si ce nombre est fini, c'est la dimension de l'espace.

3.2 Le langage géométrique

Souvent il est naturel d'abandonner le langage des espaces vectoriels et de lui préférer celui de la géométrie dans lequel on associe bijectivement un *point* à un vecteur ; un *hyperplan* de dimension r passant par l'origine à un sous espace P (cf. § 2.1.3) de dimension r (la droite et le plan sont des hyperplans de dimensions 1 et 2 respectivement). Dans ces conditions, une base B_1, B_2, \dots, B_p détermine un système de coordonnées au sens de la géométrie analytique. Les segments de droites OB_1, OB_2, \dots, OB_p sont les axes de coordonnées et le point $U = \alpha_1 B_1 + \alpha_2 B_2 + \dots + \alpha_p B_p$ a pour coordonnées $\alpha_1, \alpha_2, \dots, \alpha_p$.

Considérons un plan de la géométrie ordinaire (la feuille de papier). On veut représenter géométriquement un espace vectoriel de dimension 2 dans la base B_1, B_2 . Pour cela, on choisit un point du plan qu'on appelle *origine*. Deux droites du plan sont alors choisies pour représenter les axes déterminés par OB_1 et OB_2 . Avec ces choix, la correspondance entre les vecteurs de l'espace et les points du plan est définie. Pour construire le point correspondant au vecteur $\alpha_1 B_1 + \alpha_2 B_2$ on a seulement besoin des notions de multiplication d'un segment par un scalaire et de translation d'un segment parallèlement à lui-même (FIG. 3.1). Multiplication d'un segment par un scalaire et translation sont les expressions géométriques des opérations fondamentales sur les vecteurs (multiplication par un scalaire et addition).

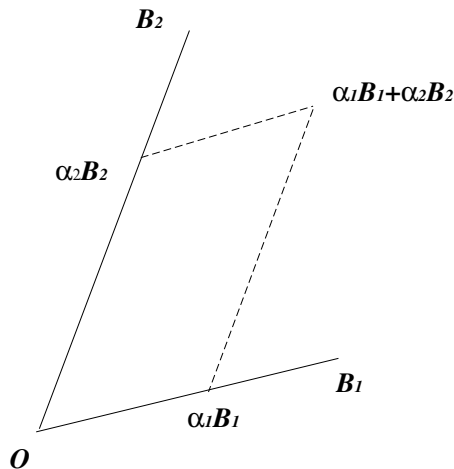


FIG. 3.1 - Construction du vecteur $\alpha_1 B_1 + \alpha_2 B_2$

Ce qui vient d'être dit peut être généralisé à plus de deux dimensions.

Dans ce contexte de représentation, l'origine O est un sous espace vectoriel de dimension 0, toute droite passant par l'origine est un sous espace vectoriel de dimension 1. Une droite qui ne passe pas par l'origine n'est pas un sous espace vectoriel, c'est un sous espace *affine*.

3.2.1 Sous espace affine

Soit E un espace vectoriel, E_1 un sous espace vectoriel de E et M un élément quelconque de E . L'ensemble des vecteurs MV_2 colinéaires aux vecteurs OV_1 de E_1 appartiennent au sous espace affine E_2 . En d'autres termes, le sous espace affine E_2 se déduit du sous espace vectoriel E_1 par une translation. Sur FIG. 3.2 sont représentés, dans le cas d'un espace vectoriel E de dimension 2, le sous espace vectoriel E_1 et le sous espace affine E_2 .

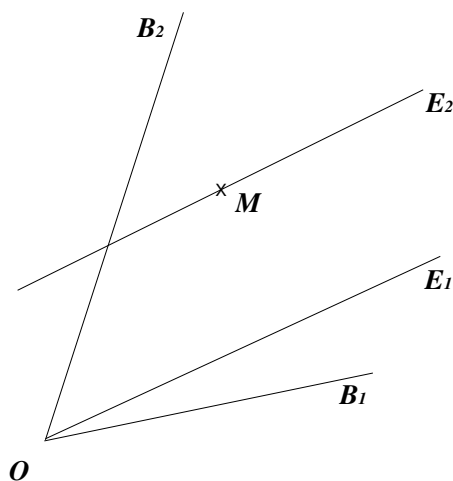


FIG. 3.2 - Sous espace vectoriel (E_1) et sous espace affine (E_2) de l'espace vectoriel E de dimension 2.

Les concepts d'espace vectoriel et d'espace affine sont différents. Cependant, on peut identifier un espace affine, dans lequel on a choisi une origine, à un espace vectoriel.

3.2.2 Décomposition en somme directe

Une base $B = [B_1, B_2, \dots, B_p]$ d'un espace vectoriel E est telle que tout vecteur V de E s'écrit de façon unique $\sum_{i=1}^p \alpha_i B_i$. Le vecteur $\alpha_i B_i$ appartient au sous espace vectoriel E_i engendré par le vecteur de base B_i .

On dit que E est *somme directe* des sous espaces vectoriels E_1, E_2, \dots, E_p et l'on écrit que :

$$E = E_1 \oplus E_2 \oplus \dots \oplus E_p$$

En particulier, quand E est somme directe des sous espaces vectoriels E_1 de dimension p_1 et E_2 de dimension p_2 tels que $E = E_1 \oplus E_2$ ($p = p_1 + p_2$), on dit que E_1 et E_2 sont *complémentaires*.

Considérons dans \mathbb{R}^3 le plan P et la droite D non contenue dans P . La droite D et le plan P sont des sous espaces vectoriels de \mathbb{R}^3 respectivement de dimensions 1 et 2 (FIG. 3.3).

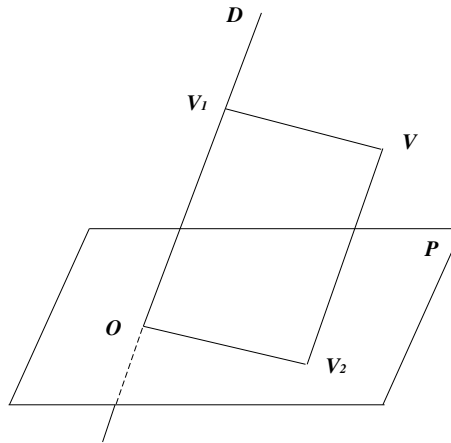


FIG. 3.3 - Représentation dans \mathbb{R}^3 des sous espaces complémentaires D et P .

$$\mathbb{R}^3 = D \oplus P$$

De façon unique, on a la décomposition $V = V_1 + V_2$ avec $V_1 \in D$ et $V_2 \in P$. Nous avons déjà rencontré ce type de décomposition au paragraphe 2.1.1.

3.3 Exemples

3.3.1 Représentations dans l'espace des variables

Soit T la variable température exprimée en degrés centigrades.

1- Les températures exprimées en degrés centigrades, en degrés Fahrenheit ($32 + 1.8T$) et en degrés Kelvin ($T - 273$) sont-elles dans le même espace des variables ?

2- Si tel n'est pas le cas, comment peut-on définir l'espace des variables pour qu'il contienne les trois variables températures ?

- 3- On suppose que chaque vecteur température est à n composantes.
- 3.1- Comment peut-on représenter les moyennes, les sommes des carrés d'écart à la moyenne et les corrélations de ces expressions différentes de la température?
 - 3.2- Quelle est la dimension du sous espace des variables centrées?
 - 3.3- Qu'est-ce qui caractérise les variables centrées dans ce sous espace?

⇒

1- Non car on ne peut pas passer d'une expression de la température à une autre par le jeu des définitions données.

2- Il faut ajouter à l'espace des variables la variable T_0 dont la valeur observée est toujours égale à 1. Ainsi, les températures en degrés centigrades, Fahrenheit et Kelvin s'écrivent :

$$0T_0 + 1T, 32T_0 + 1.8T \text{ et } -273T_0 + T$$

3.1- Moyennes : Projections des vecteurs température sur J . Somme des carrés d'écart à la moyenne et corrélations : Projections sur le sous espace des variables centrées orthogonal à J .

3.2- Dimension du sous espace des variables centrées : $n - 1$.

3.3- Caractéristique des variables centrées : la colinéarité.

3.3.2 Base d'un espace vectoriel

En première approximation, la valeur d'un engrais est caractérisée par sa concentration dans les trois principaux éléments fertilisants N, P, K (Cailliez et Pages, 1976). On désigne les engrais en indiquant les quantités d'éléments fertilisants dans l'ordre azote, acide phosphorique et potasse contenus dans 100 kilogrammes d'engrais. Les diverses compositions possibles d'engrais forment un espace vectoriel de dimension 3. Parmi les listes d'engrais ci-après, quelles sont celles qui forment une base de cet espace vectoriel?

1- (10,0,0), (0,12,0) et (0,0,20)

2- (20,20,0) et (0,12,12)

3- (0,12,12), (0,12,18) et (0,20,10)

4- (3,6,9), (4,8,8) et (4,15,0)

5- (10,10,10), (15,15,15) et (3,6,9)

⇒

En appliquant la définition : 1 et 4 forment une base de l'espace vectoriel.

3.3.3 Indépendance linéaire

Supposons (Cailliez et Pages, 1976) qu'il n'existe que trois sociétés dont les actions sont cotées en Bourse :

- 1- La Compagnie Egyptienne des Automobiles (C.E.A)
- 2- Les Industries Navales de la Region Aquitaine (I.N.R.A)
- 3- La Compagnie de Navigation Romande Suisse (C.N.R.S)

Un portefeuille formé des actions de ces diverses sociétés est un point d'un espace vectoriel si l'on admet :

- 1- La possibilité de posséder n'importe quelle fraction de titre.
- 2- La possibilité de devoir, aussi bien que de posséder, une action.

La question qu'on peut se poser est de savoir s'il est plus avantageux de "boursicoter" soi-même (en achetant ou en vendant, suivant ses pronostics sur les cours, des actions de chacune des trois sociétés) ou d'acheter des titres de sociétés d'investissements. Supposons (TAB. 3.1) qu'on trouve sur le marché six sociétés d'investissements dont les portefeuilles sont composés des titres des trois grandes compagnies dans les proportions ci-après (en %).

	C.E.A	I.N.R.A	C.N.R.S
Compagnie pour la Sécurité des Investissements	40	40	20
Centrale d'Investisseurs	70	10	20
Société Interprofessionnelle des Commerçants	30	50	20
Société Centrale d'Investissements	60	30	10
Intergroupe des Sociétés de Crédit	50	30	20
Institut Coopératif de Souscription	50	20	30

TAB. 3.1 - *Sociétés d'investissements dont les portefeuilles sont composés des titres de trois grandes compagnies (C.E.A, I.N.R.A, C.N.R.S).*

1- Qu'est-ce qui peut bien pousser le *boursicoteur* à rechercher trois sociétés d'investissements dont les portefeuilles forment une base de l'espace vectoriel des portefeuilles de titres des trois compagnies C.E.A, I.N.R.A et C.N.R.S?

2- Quels sont les ensembles de trois sociétés d'investissements qui sont les plus intéressants pour le *boursicoteur*?

⇒

1- Maximiser ses gains (ou minimiser ses pertes) en jouant (en fonction des fluctuations du marché) sur le nombre de titres en sa possession de chacune des trois sociétés d'investissement.

Si les portefeuilles d'une société d'investissement forment une combinaison linéaire des portefeuilles des deux autres sociétés d'investissement, il est bien évident qu'elle n'apporte rien au boursicoteur. Par contre, si les portefeuilles des trois sociétés d'investissement sont linéairement indépendants le boursicoteur a plus de possibilités en jouant sur les ventes et les achats des titres de ces trois sociétés (il a plus de degrés de liberté).

2- Toutes les combinaisons de trois sociétés d'investissement à l'exception des combinaisons (CSI, CIS, SIC); (CSI, CIS, ISC); (CSI, SIC, ISC); (CIS, SIC, ISC) sont intéressantes pour le boursicoteur.

Par exemple les trois sociétés d'investissements (CSI, CIS, SIC) sont des vecteurs de l'espace des trois sociétés (C.E.A, I.N.R.A, C.N.R.S) dont les actions sont cotées en bourse. Pour que ces vecteurs forment une base de cet espace vectoriel, il faut qu'ils soient linéairement indépendants. On peut facilement vérifier que tel n'est pas le cas puisque les trois vecteurs sont liés par la relation : $4 \text{ CSI} - 3 \text{ SIC} = \text{CIS}$.

On utilise une démarche identique pour montrer que (CSI, CIS, ISC); (CSI, SIC, ISC); (CIS, SIC, ISC) ne forment pas des bases de l'espace vectoriel des trois sociétés C.E.A, I.N.R.A et C.N.R.S.

3.3.4 Application linéaire et représentation matricielle

Cet exemple a pour objectifs de récapituler les concepts présentés, introduire l'addition et la multiplication matricielle (Cailliez et Pages, 1976).

1- Soit E l'ensemble des sacs d'engrais contenant N, P et K. Soit X un sac :

$$X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \in E$$

x_1 : quantité de nitrate

x_2 : quantité de phosphate

x_3 : quantité de potasse

On peut définir :

Le sac *somme* de deux sacs

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} x'_1 \\ x'_2 \\ x'_3 \end{bmatrix} = \begin{bmatrix} x_1 + x'_1 \\ x_2 + x'_2 \\ x_3 + x'_3 \end{bmatrix}$$

Le sac *homothétique* de deux sacs

$$\alpha \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} \alpha x_1 \\ \alpha x_2 \\ \alpha x_3 \end{bmatrix}$$

Si l'on admet la possibilité de quantités d'engrais négatives ou nulles, E est un espace vectoriel de dimension 3.

2- E est muni de la *base* :

$$B_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \quad B_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \quad B_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

B_1 est le sac qui contient la quantité unité de N, ... Si un sac X contient x_1 unités de N, x_2 unités de P et x_3 unités de K, on écrit que $X = x_1 B_1 + x_2 B_2 + x_3 B_3$. On dit aussi que :

$$X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \in \mathbb{R}^3$$

3- A chaque X , associons son prix y_1 et son rendement y_2 , c'est-à-dire un élément Y de l'espace F des couples (prix, rendement). Soit :

$$Y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \in F$$

On peut définir sur F une addition et une multiplication par un scalaire. Si prix et rendement peuvent être négatifs ou nuls, F est un espace vectoriel de dimension 2, il a pour base :

$$B'_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} : \text{sac de prix 1 et de rendements 0}$$

$$B'_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix} : \text{sac de prix 0 et de rendement 1}$$

A un sac d'engrais de composition donnée, on peut associer son prix et son rendement. Si l'on suppose que prix et rendements sont additifs :

$$\begin{aligned} y_1 &= a_{11}x_1 + a_{12}x_2 + a_{13}x_3 \\ y_2 &= a_{21}x_1 + a_{22}x_2 + a_{23}x_3 \end{aligned}$$

a_{11} , a_{12} , a_{13} sont respectivement les prix d'une unité de N, de P, de K.

a_{21} , a_{22} , a_{23} sont respectivement les rendements d'une unité de N, de P, de K.

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix}$$

est une matrice à 2 lignes et 3 colonnes ($A(2 \times 3)$), on écrit matriciellement que $Y = AX$. Chaque composante de $Y = AX$ est obtenue en multipliant la ligne correspondante de A par la colonne X .

Multiplier *ligne par colonne*, c'est faire la somme des produits des termes de même rang dans la ligne et la colonne considérée.

Dans cet exemple, on a supposé qu'à un sac ne correspondent qu'un prix et qu'un rendement :

$$E \xrightarrow{f} F$$

$$X \longrightarrow Y = f(X)$$

A chaque vecteur de E , f fait correspondre un et un seul vecteur de F . f est une application de E dans F , elle est linéaire (on suppose que prix et rendements sont additifs).

Des bases ayant été choisies dans E et F , on a associé une matrice A à 2 lignes et 3 colonnes. Ce résultat est général :

Pour écrire des *modèles linéaires*, pour faire des *calculs*, on associe une matrice à chaque application linéaire (une fois précisées les bases des espaces de départ et d'arrivée).

3.3.5 Représentation des sous espaces engendrés par les colonnes d'une matrice

On considère la matrice :

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

Cette matrice est telle que la somme des deux premières colonnes est égale à la somme des deux dernières colonnes. Le résultat de ces sommations est égal au vecteur J .

Par ailleurs, les colonnes de X sont des vecteurs colonnes non colinéaires. Ainsi, les deux premières colonnes définissent un plan P_1 et les deux dernières colonnes un plan P_2 . Ces deux plans sont des sous espaces vectoriels de \mathbb{R}^4 , ils ont le vecteur J en commun.

Si l'on veut représenter ces deux sous espaces, il faut déterminer la position de P_1 par rapport à P_2 et pour cela chercher la valeur de l'angle de ces deux

plans.

Le vecteur J commun à P_1 et à P_2 , est à l'intersection de ces deux plans. Soit U un vecteur quelconque de P_1 et V un vecteur quelconque de P_2 , l'angle des plans P_1 et P_2 est égal à l'angle des vecteurs $U - U_J$ et $V - V_J$ avec U_J et V_J les projections de U et V sur J .

Les deux premières colonnes de X sont des vecteurs de base pour P_1 , les deux dernières colonnes sont des vecteurs de base pour P_2 . Les vecteurs U et V sont donc de la forme :

$$U = a_1 \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} + b_1 \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} a_1 \\ a_1 \\ b_1 \\ b_1 \end{bmatrix}$$

$$V = a_2 \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix} + b_2 \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} a_2 \\ b_2 \\ a_2 \\ b_2 \end{bmatrix}$$

avec a_1, b_1, a_2 et b_2 des scalaires. On déduit que :

$$U_J = \begin{bmatrix} \frac{a_1 + b_1}{2} \\ \frac{a_1 + b_1}{2} \\ \frac{a_1 + b_1}{2} \\ \frac{a_1 + b_1}{2} \end{bmatrix} \quad V_J = \begin{bmatrix} \frac{a_2 + b_2}{2} \\ \frac{a_2 + b_2}{2} \\ \frac{a_2 + b_2}{2} \\ \frac{a_2 + b_2}{2} \end{bmatrix}$$

$$U - U_J = \begin{bmatrix} \frac{a_1 - b_1}{2} \\ \frac{a_1 - b_1}{2} \\ -\frac{a_1 - b_1}{2} \\ -\frac{a_1 - b_1}{2} \end{bmatrix} \quad V - V_J = \begin{bmatrix} \frac{a_2 - b_2}{2} \\ -\frac{a_2 - b_2}{2} \\ \frac{a_2 - b_2}{2} \\ -\frac{a_2 - b_2}{2} \end{bmatrix}$$

$$\langle U - U_J, V - V_J \rangle = \left(\frac{a_1 - b_1}{2} \right) \left(\frac{a_2 - b_2}{2} \right) [1 \quad 1 \quad -1 \quad -1] \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \end{bmatrix} = 0$$

Le produit scalaire $\langle U - U_J, V - V_J \rangle$ étant égal à zéro, on déduit que P_1 et P_2 sont orthogonaux, on a la représentation de la figure FIG. 3.4.

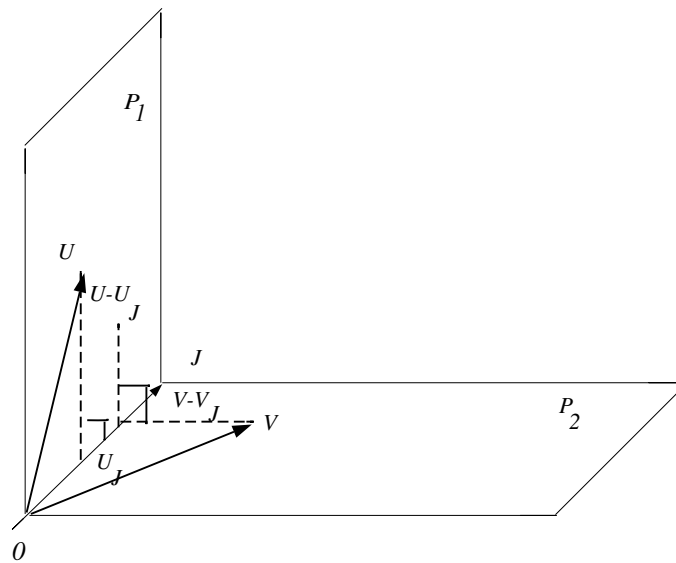


FIG. 3.4 - Sous espaces P_1 et P_2 engendrés par les colonnes de la matrice X .

Supposons maintenant qu'on s'intéresse à la matrice suivante :

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

et que, comme nous venons de le faire avec X , nous cherchions à représenter les sous espaces P'_1 et P'_2 engendrés respectivement par les deux premières colonnes et les deux dernières colonnes de X' . En désignant par U' un vecteur quelconque de P'_1 et par V' un vecteur quelconque de P'_2 on vérifie facilement que le produit scalaire $\langle U' - U'_J, V' - V'_J \rangle$ est différent de zéro (U'_J et V'_J sont les projections de U' et de V' sur J). On déduit que les plans P'_1 et P'_2 ne sont plus orthogonaux, on a la représentation de la figure FIG. 3.5.

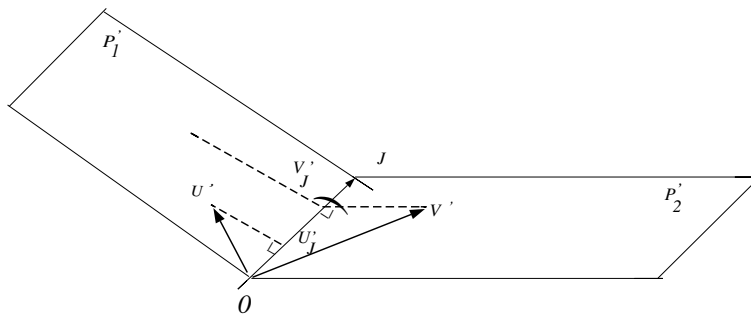


FIG. 3.5 - Sous espaces P'_1 et P'_2 engendrés par les colonnes de la matrice X' .

La matrice X' a été obtenue en supprimant la dernière ligne de X . Cette suppression a rompu l'équilibre de X qui était caractérisé par le nombre de

0 et de 1 et par la disposition des zéros et des uns dans les colonnes. Cette rupture est traduite graphiquement par la non orthogonalité des deux sous espaces vectoriels P'_1 et P'_2 . Cette représentation est intéressante pour visualiser et comprendre ce que sont des sommes des carrés ajustées et non ajustées en analyse de la variance par exemple.

Chapitre 4

Calcul matriciel

Le tableau rectangulaire A à n lignes et p colonnes où a_{ij} est l'élément à l'intersection de la i ème ligne et de la j ème colonne est une matrice qu'on représente par :

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{np} \end{bmatrix}$$

Chaque ligne i de A est un vecteur ligne : $[a_{i1} \quad a_{i2} \quad \cdots \quad a_{ip}]$ et chaque colonne

j est un vecteur colonne : $\begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{nj} \end{bmatrix}$

4.1 Définitions

1. *Dimension*: Le nombre de lignes n et le nombre de colonnes p de la matrice (notation : $(n \times p)$).
2. *Matrice transposée*: Matrice obtenue en interchangeant les rôles des lignes et des colonnes (notation : A'). La dimension de A' est $(p \times n)$.

$$A' = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{np} \end{bmatrix}' = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{12} & a_{22} & \cdots & a_{n2} \\ \vdots & \vdots & \vdots & \vdots \\ a_{1p} & a_{2p} & \cdots & a_{np} \end{bmatrix}$$

3. *Matrice carrée*: Matrice qui a autant de lignes que de colonnes ($n = p$).

4. *Matrices triangulaires* :

- *Triangulaire haute* : Matrice carrée dont les éléments au-dessous de la diagonale principale sont égaux à zéro.

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & a_{nn} \end{bmatrix}$$

- *Triangulaire basse* : Matrice carrée dont les éléments au-dessus de la diagonale principale sont égaux à zéro.

$$A = \begin{bmatrix} a_{11} & 0 & \cdots & 0 \\ a_{21} & a_{22} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

- 5. *Matrice symétrique* : Matrice carrée égale à sa transposée ($a_{ij} = a_{ji}$ pour tout couple i, j). En général, on ne représente que la partie triangulaire haute (ou basse), c'est-à-dire les éléments situés sur la diagonale principale et au-dessus (au-dessous) de la diagonale principale.

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ & a_{22} & \cdots & a_{2n} \\ & & \vdots & \vdots \\ & & & a_{nn} \end{bmatrix}$$

- 6. *Matrice diagonale* : Matrice carrée dont les seuls éléments éventuellement non nuls sont sur la diagonale principale (notation : $\text{diag}(A)$).

$$D = \text{diag}(A) = \begin{bmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & a_{22} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & a_{nn} \end{bmatrix}$$

- 7. *Matrice identité* : Matrice diagonale dont les éléments diagonaux sont tous égaux à 1 (notation : I).

- 8. *Matrice nulle* : Matrice dont tous les éléments sont égaux à zéro.

- 9. *Matrice J* : Matrice dont tous les éléments sont égaux à 1.

4.2 Opérations élémentaires

1. Addition:

$$C = A + B \iff c_{ij} = a_{ij} + b_{ij}$$

C , A et B sont de mêmes dimensions.

2. Multiplication par un scalaire:

$$C = \lambda A \iff c_{ij} = \lambda a_{ij}$$

C et A sont de mêmes dimensions

3. Produit de deux matrices:

Pour que le produit $C = AB$ de deux matrices A et B soit défini, il faut que le nombre de colonnes de A soit égal au nombre de lignes de B . Si A est de dimensions $(n \times p)$ et B de dimensions $(p \times m)$, le résultat C du produit de A par B est de dimensions $(n \times m)$.

$$C = AB \iff c_{ij} = \sum_{k=1}^p a_{ik} b_{kj}$$

En général, la multiplication n'est pas commutative, $AB \neq BA$. Pour cette raison, on dit parfois du produit AB que c'est la prémultiplication de B par A ou la postmultiplication de A par B .

4. Matrice idempotente:

Une matrice $A(n \times n)$ est idempotente si: $AA = A$.

5. Matrice inverse:

On appelle matrice inverse de A la matrice A^{-1} telle que:

$$AA^{-1} = A^{-1}A = I$$

Si A^{-1} existe, elle est *unique*, on dit alors que A est *non singulière*.

Si A n'a pas d'inverse elle est dite *singulière*.

- L'inverse d'une matrice symétrique est symétrique.
- L'inverse de la transposée de A est la transposée de A^{-1} .
- $(ABC)^{-1} = C^{-1}B^{-1}A^{-1}$
- $(\alpha A)^{-1} = \frac{1}{\alpha}A^{-1}$

$$- D = \begin{bmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & a_{22} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & a_{nn} \end{bmatrix} \implies D^{-1} = \begin{bmatrix} a_{11}^{-1} & 0 & \cdots & 0 \\ 0 & a_{22}^{-1} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & a_{nn}^{-1} \end{bmatrix}$$

6. *Matrice orthogonale :*

$A(n \times n)$ est orthogonale si : $AA' = I \iff A' = A^{-1}$

7. *Rang :*

Le nombre de vecteurs lignes linéairement indépendants si $n \leq p$, le nombre de vecteurs colonnes linéairement indépendants si $n > p$ (notation : $rg(A)$).

- $rg(A') = rg(A)$
- $rg(A'A) = rg(AA') = rg(A)$
- $rg(AB) \leq \min(rg(A), rg(B))$

8. *Inverse généralisée :*

Soit A une matrice ($n \times p$) de rang r . $G(p \times n)$ est une *inverse généralisée* de A si G est de rang r et si elle satisfait à $AGA = A$. En statistique, on utilise souvent l'inverse généralisée de Penrose (1955) qui satisfait aussi aux conditions :

$$GAG = G \quad (GA)' = GA \quad (AG)' = AG$$

9. *Transposée d'un produit de matrices :*

$$(AB)' = B'A'$$

10. *Trace :*

Soit A une matrice carrée, la somme des éléments diagonaux de A s'appelle la trace de A qu'on note : $tr(A)$.

$$- tr(A) = tr(A')$$

Soient $A(n \times p)$ et $B(p \times n)$ deux matrices, alors :

$$- tr(AB) = tr(BA)$$

$$- tr(AA') = tr(A'A) = \sum_{i=1}^n \sum_{j=1}^p a_{ij}^2$$

11. Déterminant d'une matrice :

Considérons le système d'équations :

$$\begin{cases} 3x_1 + 5x_2 + 1x_3 = 7 \\ 2x_1 + 4x_2 + 5x_3 = 4 \\ 1x_1 + 2x_2 + 2x_3 = 6 \end{cases}$$

On l'écrit matriciellement :

$$\begin{bmatrix} 3 & 5 & 1 \\ 2 & 4 & 5 \\ 1 & 2 & 2 \end{bmatrix} \times \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 7 \\ 4 \\ 6 \end{bmatrix}$$

On l'écrit symboliquement : $AX = B$.

Si la matrice A est inversible (A^{-1} existe) alors le système d'équations admet la solution : $X = A^{-1}B$ obtenue en prémultipliant $AX = B$ par A^{-1} . Pour savoir si A^{-1} existe, on calcule un scalaire appelé le *déterminant* de A qu'on note $\det(A)$. On sait en effet que si $\det(A) \neq 0$ alors A^{-1} existe.

Calcul du déterminant de A

Pour préciser le mode de calcul notons :

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

avec $a_{11} = 3$, $a_{12} = 5$, \dots , etc

Le déterminant de cette matrice est :

$$\det(A) = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33}$$

En remplaçant les a_{ij} par leurs valeurs, on a :

$$\det(A) = 3 \times 4 \times 2 + 5 \times 5 \times 1 + 1 \times 2 \times 2 - 1 \times 4 \times 1 - 3 \times 5 \times 2 - 5 \times 2 \times 2 = -1$$

D'un point de vue général, si A une matrice carrée ($n \times n$), le déterminant de A est le scalaire $\det(A) = \sum \pm a_{1j(1)}a_{2j(2)} \cdots a_{nj(n)}$, la sommation portant sur les $n!$ permutations $j(1), j(2), \dots, j(n)$ des entiers $1, 2, \dots, n$. Les signes $+$ et $-$ sont associés aux nombres pair et impair d'interchanges de paires nécessaires pour modifier la permutation $1, 2, \dots, n$ en $j(1), j(2), \dots, j(n)$. On peut aussi définir le déterminant d'une matrice carrée A à partir des *mineurs* et des *cofacteurs* de A .

- Le mineur de l'élément a_{ij} de A est le déterminant obtenu en supprimant la i ème ligne et la j ème colonne de A .

- Le cofacteur de a_{ij} est le mineur multiplié par $(-1)^{i+j}$, on le note A_{ij} .

En terme de cofacteurs, le déterminant de A s'écrit de la façon suivante :

$$\det(A) = a_{i1}A_{i1} + \cdots + a_{in}A_{in} \quad i = 1, \dots, n$$

$$\det(A) = a_{1j}A_{1j} + \cdots + a_{nj}A_{nj} \quad j = 1, \dots, n$$

- Si $A(n \times n)$ est une matrice triangulaire :

$$\det(A) = \prod_{i=1}^n a_{ii}$$

- Si les éléments d'une ligne ou d'une colonne d'une matrice $A(n \times n)$ sont multipliés par un scalaire λ , le déterminant de la nouvelle matrice est égal à $\lambda \det(A)$. Si chaque élément est multiplié par λ , alors $\det(\lambda A) = \lambda^n \det(A)$.
- Si deux lignes (ou deux colonnes) d'une matrice sont interchangées, le signe du déterminant est inversé.
- Si A et B sont deux matrices de dimension $(n \times n)$, alors :

$$\det(AB) = \det(A) \times \det(B)$$

- $A(n \times n)$ est non singulière si et seulement si $\det(A) \neq 0$.
- Si $A(n \times n)$ est non singulière, alors :

$$\det(A^{-1}) = \frac{1}{\det(A)}$$

12. Matrice adjointe :

C'est la matrice des cofacteurs, on la note $A^*(n \times n)$.

$$A^* = \begin{bmatrix} A_{11} & \cdots & A_{1n} \\ \cdots & \cdots & \cdots \\ A_{n1} & \cdots & A_{nn} \end{bmatrix}$$

On montre que si $\det(A) \neq 0$:

$$A^{-1} = \frac{1}{\det(A)} A^{*'} = \frac{1}{\det(A)} \begin{bmatrix} A_{11} & \cdots & A_{n1} \\ \cdots & \cdots & \cdots \\ A_{1n} & \cdots & A_{nn} \end{bmatrix}$$

4.2.1 Matrice des variables centrées réduites

Quand les grandeurs mesurées et les unités sont hétérogènes, on est conduit à transformer les données de telle sorte qu'elles aient une origine commune et s'expriment dans une même unité. Pour cela, on centre et on réduit ces grandeurs. Si $X(n \times p)$ est la matrice des grandeurs mesurées, écrire la matrice $Z(n \times p)$ des variables centrées réduites.

\implies

Soient :

- x_{ij} la $j^{\text{ème}}$ variable observée sur le $i^{\text{ème}}$ individu
- \bar{x}_j la moyenne de la $j^{\text{ème}}$ variable
- s_j l'écart type de la $j^{\text{ème}}$ variable
- $z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j}$ la variable centrée réduite

La matrice X des données s'écrit :

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}$$

On définit les matrices D et J de la façon suivante :

$$D = \begin{bmatrix} \frac{1}{s_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{s_2} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \frac{1}{s_p} \end{bmatrix}$$

$$J = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix}$$

L'écriture matricielle de la transformation "centrage, réduction" est alors :

$$Z = \left(I - \frac{1}{n}J\right)XD$$

avec I la matrice identité.

4.2.2 Indépendance linéaire

Dans l'exemple des trois sociétés dont les actions sont cotées en bourse, on a vu que les sociétés d'investissements CSI, CIS, SIC ne forment pas une base de l'espace vectoriel des trois sociétés C.E.A, I.N.R.A et C.N.R.S. Calculons, par la méthode des cofacteurs, le déterminant de la matrice carrée A de dimension 3 formée par CSI, CIS, SIC :

$$A = \begin{bmatrix} 40 & 40 & 20 \\ 70 & 10 & 20 \\ 30 & 50 & 20 \end{bmatrix}$$

$$\det(A) = 40(10 \times 20 - 20 \times 50) - 40(70 \times 20 - 20 \times 30) + 20(70 \times 50 - 10 \times 30) = 0$$

La matrice A est telle que :

- Une de ses lignes est combinaison linéaire des deux autres.
- Son déterminant est nul.

A partir de cette remarque, on admettra que quand une matrice a une ligne (ou une colonne) qui est combinaison linéaire des autres lignes (ou colonnes) alors, son déterminant est égal à zéro. Ce résultat est intéressant pour vérifier l'indépendance linéaire de vecteurs.

4.3 Représentation géométrique

Dans un espace vectoriel E de dimension p , on définit généralement le produit scalaire sur des vecteurs de base B_i , ($i = 1, \dots, p$). La base d'un espace vectoriel n'est pas unique. Supposons que les vecteurs de la base choisie sont des vecteurs colonnes rangés dans une matrice. Soit :

$$B = [B_1 \quad B_2 \quad \dots \quad B_p]$$

la matrice des vecteurs de base de dimension $(p \times p)$.

Dans E muni de la base B , deux vecteurs quelconques U et V sont de la forme :

$$U = \sum_{i=1}^p x_i B_i \text{ et } V = \sum_{i=1}^p y_i B_i$$

avec x_i et y_i des nombres réels, i variant de 1 à p .

Les x_i et les y_i , ($i = 1, \dots, p$) peuvent être considérés comme les composantes de deux vecteurs X et Y de dimension $(p \times 1)$, donc $U = BX$ et $V = BY$. On en déduit que :

$$\langle U, V \rangle = \langle BX, BY \rangle$$

Soit Q la matrice $(p \times p)$ dont l'élément (i, j) est $\langle B_i, B_j \rangle$ alors :

$$\langle U, V \rangle = \langle BX, BY \rangle = X'B'BY = X'QY$$

On dit que Q est la *métrique* relative à la base B .

Puisqu'une métrique est liée à une base particulière, on déduit que tout changement de base définit un produit scalaire différent.

- Dans l'espace Euclidien E muni de la base B , il est clair que $X'QX > 0$ pour tout vecteur X de dimension $(p \times 1)$ différent de zéro. Dans ce cas, on dit que la matrice Q est *définie positive*. On peut vérifier par ailleurs que Q est *symétrique* : $Q = BB' = (BB')'$.
- $X'QX$ est désignée sous le nom de *forme quadratique*.
- D'un point de vue géométrique, il existe une correspondance étroite entre la classe des matrices symétriques définies positives et la classe des *ellipsoïdes* centrés à l'origine. Nous allons montrer cette correspondance dans le cas d'un espace de dimension $p = 2$.

Soit un nuage de points constitué par n observations de la variable X_1 et de la variable X_2 . La forme du nuage de points, plus ou moins allongée dans une direction donnée, est une représentation de la relation qui lie les observations de X_1 aux observations de X_2 ; le coefficient de corrélation, la covariance sont des mesures de cette liaison. Le problème de la représentation géométrique des observations revient à chercher la figure géométrique qui enveloppe au mieux le nuage de points. Pour cela, on écrit que la distance d'un point au centre du nuage est inférieure ou égale à une constante positive. Comme la distance dérive du produit scalaire et que $X'QX$ est le produit scalaire associé à la métrique Q , on déduit que, les points à une distance du centre du nuage inférieure à C , sont tels que $X'QX \leq C^2$, soit :

$$X'QX = [X_1 - \bar{X}_1 \quad X_2 - \bar{X}_2] \begin{bmatrix} 1 & b \\ b & 1 \end{bmatrix} \begin{bmatrix} X_1 - \bar{X}_1 \\ X_2 - \bar{X}_2 \end{bmatrix} \leq C^2$$

avec X_1 et X_2 les coordonnées d'un point, \bar{X}_1 et \bar{X}_2 les coordonnées du point moyen et $Q = \begin{bmatrix} 1 & b \\ b & 1 \end{bmatrix}$ la métrique (matrice symétrique, définie positive). Si

l'on développe cette expression, on trouve : $(X_1 - \bar{X}_1)^2 + 2b(X_1 - \bar{X}_1)(X_2 - \bar{X}_2) + (X_2 - \bar{X}_2)^2 \leq C^2$ qui n'est autre que l'équation d'une surface limitée par une conique. Cette conique sera une ellipse si $b^2 \leq 1$.

On observe que, si le coefficient b est de signe opposé à celui du produit $(X_1 - \bar{X}_1)(X_2 - \bar{X}_2)$, l'ellipse est orientée dans le sens des variations conjointes des variables X_1 et X_2 (par exemple, si X_1 et X_2 augmentent simultanément le grand axe de l'ellipse est une droite de pente positive). Intuitivement, on dira que le terme $2b(X_1 - \bar{X}_1)(X_2 - \bar{X}_2)$ doit être négatif pour que l'ellipse et le nuage de points aient des orientations identiques. Dans ces conditions, quand X_1 et X_2 sont liées positivement (resp. négativement), le produit $(X_1 - \bar{X}_1)(X_2 - \bar{X}_2)$ est positif (resp. négatif) et donc b doit être négatif (resp. positif). En d'autres termes, si la mesure de liaison est le coefficient de corrélation, il interviendra avec le signe “-” (resp. “+”) si X_1 et X_2 sont liées positivement (resp. négativement). Ainsi, la métrique Q associée à l'enveloppe elliptique du nuage de points ne peut pas être la matrice de corrélation $\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$ car ρ et $(X_1 - \bar{X}_1)(X_2 - \bar{X}_2)$ sont de même signe, mais doit être l'inverse de la matrice de corrélation $\frac{1}{1 - \rho^2} \begin{bmatrix} 1 & -\rho \\ -\rho & 1 \end{bmatrix}$ car dans ce cas $-\rho$ et $(X_1 - \bar{X}_1)(X_2 - \bar{X}_2)$ sont de signes opposés.

D'un point de vue général, on admettra que $X'Q^{-1}X \leq 1$ est l'ellipsoïde d'inertie associé à la forme quadratique $X'QX$. La justification de cette affirmation est donnée au paragraphe § 4.4 (Q et Q^{-1} ont les mêmes vecteurs propres).

4.3.1 Exemples

Pour illustrer ce qui vient d'être dit considérons les variables X_1 et X_2 . Soient S_1^2 et S_2^2 les variances estimées de X_1 et X_2 , $X_1^* = \frac{X_1 - \bar{X}_1}{S_1}$ et $X_2^* = \frac{X_2 - \bar{X}_2}{S_2}$ les variables X_1 et X_2 centrées réduites.

1. Plaçons nous d'abord dans l'espace des individus repérés par leurs coordonnées centrées réduites X_1^* et X_2^* . Considérons les deux cas suivants :

- (a) Les variables X_1^* et X_2^* ne sont pas liées par une relation linéaire (*i.e.* le coefficient de corrélation $\rho = 0$). Le lieu des points M de coordonnées X_1^* et X_2^* , à une distance unité de l'origine est le cercle d'équation : $X_1^{*2} + X_2^{*2} = 1$

Sous forme matricielle l'équation du cercle s'écrit $X^{*'} Q^{-1} X^* = 1$ avec $X^{*'}$ le vecteur transposé de X^* et

$$X^* = \begin{bmatrix} X_1^* \\ X_2^* \end{bmatrix}, \quad Q^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

- (b) Les variables X_1^* et X_2^* sont liées par une relation linéaire (*i.e.* le coefficient de corrélation $\rho \neq 0$). L'inverse de la matrice de corrélation est :

$$Q^{-1} = \frac{1}{(1 - \rho^2)} \begin{bmatrix} 1 & -\rho \\ -\rho & 1 \end{bmatrix}$$

Après développement, l'équation de l'ellipse $X^{*'} Q^{-1} X^* = 1$ est :

$$X_1^{*2} - 2\rho X_1^* X_2^* + X_2^{*2} = 1 - \rho^2 \quad (4.1)$$

Si $\rho = 0$ on retrouve l'équation du cercle, si $\rho = \pm 1$, on obtient des équations de droites qui passent par le point moyen μ de coordonnées (X_1, X_2) . Pour chacune d'elle, le signe de la pente est celui du coefficient de corrélation ($\rho = \pm 1 \implies X_1^* = \pm X_2^*$). Suivant les valeurs du coefficient de corrélation, (4.1) permet de passer de l'équation d'un cercle ($\rho = 0$) à celle d'une ellipse ($|\rho| < 1$) puis, à celle d'une droite ($\rho = \pm 1$). Cette représentation géométrique (FIG. 4.1) rend compte du degré de liaison entre X_1^* et X_2^* car la valeur du coefficient de corrélation ρ caractérise l'aplatissement de l'ellipse c'est-à-dire *l'excentricité* de l'ellipse.

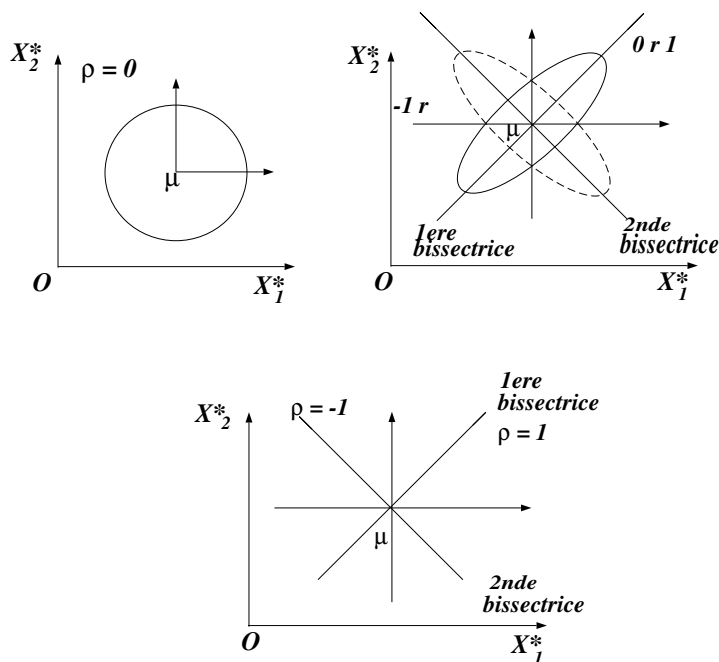


FIG. 4.1 - Ellipses d'inertie centrées au point moyen μ pour différentes valeurs du coefficient de corrélation (ρ).

2. Plaçons nous maintenant dans l'espace des individus repérés par leurs coordonnées X_1 et X_2 . L'équation (4.1) s'écrit :

$$\left(\frac{X_1 - \bar{X}_1}{S_1}\right)^2 - 2\rho \left(\frac{X_1 - \bar{X}_1}{S_1}\right) \left(\frac{X_2 - \bar{X}_2}{S_2}\right) + \left(\frac{X_2 - \bar{X}_2}{S_2}\right)^2 = 1 - \rho^2$$

Soit :

$$[S_2(X_1 - \bar{X}_1)]^2 - 2\rho[S_2(X_1 - \bar{X}_1)][S_1(X_2 - \bar{X}_2)] + [S_1(X_2 - \bar{X}_2)]^2 = S_1^2 S_2^2 (1 - \rho^2) \quad (4.2)$$

C'est l'équation d'une ellipse, d'excentricité ρ , centrée au point moyen μ de coordonnées (\bar{X}_1, \bar{X}_2) . On considère les deux cas suivants :

- (a) Le coefficient de corrélation $\rho = 0$:
- Si $S_1 = S_2 = S$, l'équation (4.2) se réduit à l'équation d'un cercle.
 - Si $S_1 \neq S_2$, l'équation (4.2) est soit l'équation d'une ellipse de grand axe parallèle à X_1 si $S_1 > S_2$ soit de grand axe parallèle à X_2 si $S_2 > S_1$.
- (b) Le coefficient de corrélation $\rho \neq 0$:
- Si $\rho = \pm 1$, l'équation (4.2) se réduit à l'équation générale d'une droite passant par le point moyen μ de coordonnées (\bar{X}_1, \bar{X}_2) , de pente $\pm \frac{S_1}{S_2}$ selon que $\rho = \pm 1$.
 - Si $|\rho| < 1$, l'équation (4.2) est l'équation d'une ellipse pour laquelle la pente du grand axe dépend du signe du coefficient de corrélation et du rapport des valeurs de S_1 et S_2 .

Ces ellipses d'inertie, plus ou moins dégénérées en fonction des valeurs du coefficient de corrélation, dont les orientations dans le plan dépendent du signe du coefficient de corrélation et du rapport des variances, représentent bien l'allure générale des nuages de points (FIG. 4.2).

Quelques remarques s'imposent :

- Dans les deux cas qui viennent d'être présentés, les métriques utilisées sont différentes si l'on considère que sur l'espace des individus, les points individus sont repérés par :
 1. des coordonnées centrées réduites X_1^* et X_2^* . La métrique est alors la matrice de corrélation.

2. des coordonnées centrées $X_1 - \bar{X}_1$ et $X_2 - \bar{X}_2$. La métrique est alors la matrice de variance-covariance.
- Il existe d'autres métriques qui permettent d'interpréter des représentations géométriques en fonction de la nature des données et de l'objectif fixé par l'analyse. On peut citer par exemple la métrique du χ^2 utilisée en analyse factorielle des correspondances (Cailliez et Pages, 1976). En fait, il est presque toujours possible de travailler avec la métrique du *géomètre* à condition de faire les transformations idoines sur les données.

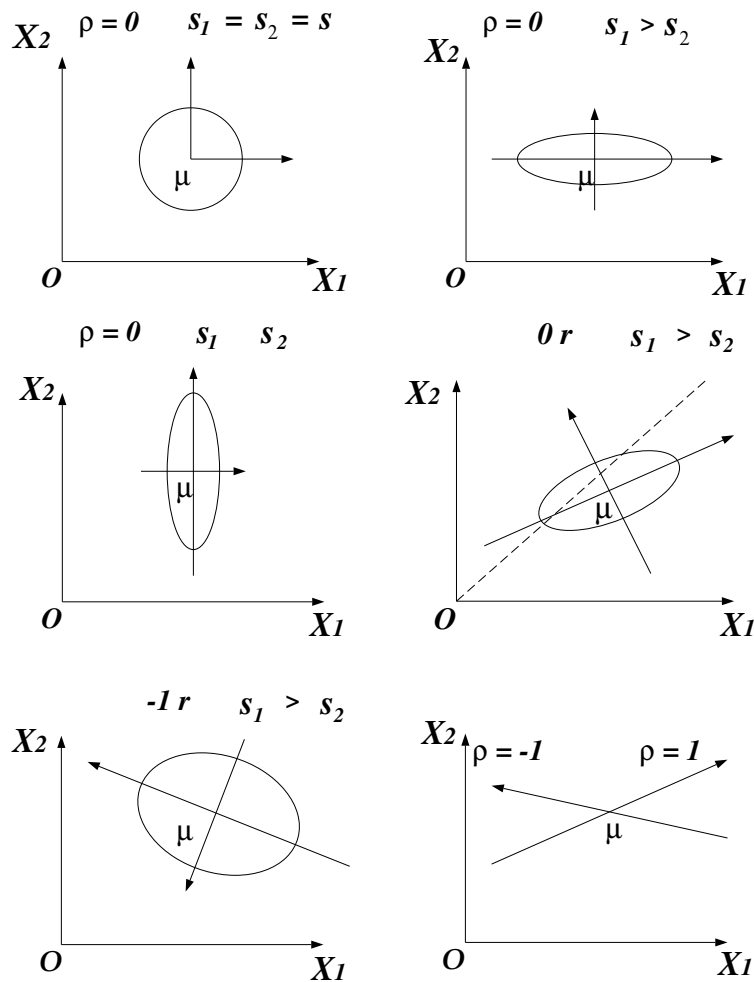


FIG. 4.2 - Ellipses d'inertie centrées au point moyen μ pour des valeurs différentes du coefficient de corrélation (ρ) et des écarts types (S_1 et S_2).

Quand un produit scalaire est donné dans un espace, il est possible de considérer une gamme plus étendue de produits scalaires et la gamme correspondante d'ellipses associées dans laquelle le cercle ne représente qu'un cas particulier. Sur FIG. 4.3 on a représenté un plan dans lequel les vecteurs OB_1 et OB_2 sont

perpendiculaires. Si l'on prend comme produit scalaire dans ce plan celui qui est défini en choisissant pour base OB_1 et OB_2 alors, l'ensemble des points à une distance unité de l'origine O est le cercle de rayon unité centré en O . Supposons maintenant qu'un couple de vecteurs arbitraires OB_1^* , OB_2^* est choisi comme base orthonormale pour une autre définition du produit scalaire sur le même espace. Alors, on peut vérifier que l'ensemble des points à une distance unité de l'origine O est une ellipse. Cette ellipse est telle que la tangente à l'ellipse, au point d'intersection de l'ellipse et de OB_1^* , est parallèle à OB_2^* et que la tangente à l'ellipse, au point d'intersection de l'ellipse et de OB_2^* , est parallèle à OB_1^* , aucun autre axe de l'ellipse n'a cette propriété. On dit que OB_1 et OB_2 d'une part et OB_1^* et OB_2^* d'autre part sont les *demi-axes conjugués* du cercle et de l'ellipse respectivement.

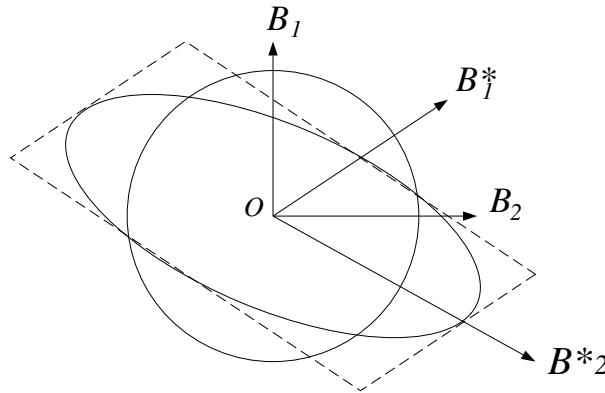


FIG. 4.3 - Ellipses dans les bases orthonormales OB_1, OB_2 et OB_1^*, OB_2^* .

Les résultats précédents se généralisent à des espaces de dimensions $p > 2$.

Conséquence

Un objectif de la statistique étant de décrire et de résumer des nuages de points, on utilise le concept d'ellipsoïde centré à l'origine pour rechercher des directions et des points particuliers qui caractérisent ces nuages de points. Ces directions et ces points particuliers sont : les axes principaux, les axes conjugués, les points moyens des ellipsoïdes. Pour déterminer ces axes on raisonne en termes de changement de base en utilisant les notions de *valeurs propres* et de *vecteurs propres*.

4.4 Valeurs et vecteurs propres

Soient D et D^\perp deux droites orthogonales du plan passant par l'origine O . On note S la symétrie par rapport à D . S est une application linéaire de \mathbb{R}^2 dans \mathbb{R}^2 . Cherchons les points de \mathbb{R}^2 tels que $S(X) = \lambda X$.

$$V_1 \in D \implies S(V_1) = \lambda_1 V_1 = 1 \times V_1$$

$$V_2 \in D^\perp \implies S(V_2) = \lambda_2 V_2 = -1 \times V_2$$

Le point 0 joue un rôle particulier : $S(0) = \lambda 0$ pour tout λ .

V_1 et V_2 sont appelés *vecteurs propres* de S associés aux *valeurs propres* λ_1 et λ_2 . Ainsi, tout vecteur de D est un vecteur propre associé à la valeur propre $\lambda_1 = 1$ et tout vecteur de D^\perp est un vecteur propre associé à la valeur propre $\lambda_2 = -1$. Il y a donc *deux infinités* de vecteurs propres associés à l'application linéaire S , ils se répartissent sur les deux sous espaces vectoriels de dimension 1 : D et D^\perp . D est le *sous espace propre* associé à la valeur propre $\lambda_1 = 1$ et D^\perp est le *sous espace propre* associé à la valeur propre $\lambda_2 = -1$. En généralisant cet exemple de la symétrie par rapport à D , on définit les vecteurs propres et les valeurs propres de la façon suivante :

4.4.1 Définition

Soit E un espace vectoriel de dimension p , f une application linéaire de E dans E . $V \in E$ est vecteur propre de f associé à la valeur propre $\lambda \in \mathbb{R}$ si $f(V) = \lambda V$ ($V \neq O$). Si E est muni de la base $B = [B_1, B_2, \dots, B_p]$, $V = \sum_{i=1}^p v_i B_i$.

Soit A la matrice associée à l'application linéaire f , la base B étant choisie dans E , $f(V) = \lambda V \iff AV = \lambda V \iff (A - \lambda I)V = 0$. V est vecteur propre de la matrice A de valeur propre associée λ . Le système homogène $(A - \lambda I)V = 0$ admettra une solution non triviale si et seulement si $\det(A - \lambda I) = 0$.

4.4.2 Résultats généraux

1. A toute matrice carrée A d'ordre p sont associées au plus p valeurs propres réelles. Ces valeurs propres sont les racines d'un polynôme de degré p : $\det(A - \lambda I) = 0$ appelé *polynôme caractéristique*.

Tout vecteur propre V_i associé à la valeur propre λ_i de la matrice carrée A satisfait au système d'équations : $(A - \lambda_i I)V_i = 0$.

2. Si A est symétrique :
 - (a) Toutes les racines du polynôme caractéristique sont réelles.
 - (b) A la racine simple λ_i correspond un sous espace propre de dimension 1.

- (c) L'ordre de multiplicité d'une valeur propre est la dimension du sous espace propre associé.
- (d) Si λ_i et λ_j sont deux valeurs propres distinctes, les vecteurs propres associés sont orthogonaux. En effet, si λ_i et λ_j sont deux valeurs propres distinctes et V_i, V_j les vecteurs propres associés :

$$AV_i = \lambda_i V_i \text{ et } AV_j = \lambda_j V_j$$

En prémultipliant la première de ces équations par V_j' (vecteur propre transposé de V_j) et la seconde par V_i' (vecteur propre transposé de V_i), on obtient :

$$V_j'AV_i = \lambda_i V_j'V_i \text{ et } V_i'AV_j = \lambda_j V_i'V_j$$

En faisant la différence de ces deux expressions, on déduit :

$$\lambda_i V_j'V_i = \lambda_j V_i'V_j \iff (\lambda_i - \lambda_j)V_j'V_i = 0$$

car $V_j'V_i = V_i'V_j \in \mathbb{R}$ donc :

$$\lambda_i \neq \lambda_j \implies V_j'V_i = 0$$

Le produit scalaire des vecteurs V_i et V_j est égal à zéro, ces deux vecteurs sont orthogonaux.

- (e) Il existe une matrice orthogonale P telle que : $A = P'DP$. D est une matrice diagonale. Les éléments situés sur la diagonale sont les valeurs propres de A . Les colonnes de la matrice P sont les vecteurs propres normalisés de A . Même si les valeurs propres ne sont pas distinctes, il est encore possible de choisir les éléments de leurs vecteurs propres permettant d'obtenir un ensemble de vecteurs propres mutuellement orthogonaux. Ces propriétés sont importantes car si l'on applique la transformation orthogonale : $X = PY$ aux p variables d'une forme quadratique $X'AX$, la forme quadratique devient :

$$X'AX = Y'P'APY = Y'DY = \lambda_1 y_1^2 + \dots + \lambda_r y_r^2$$

les λ_i sont les valeurs propres de A et r est le rang de la forme quadratique. Toute forme quadratique peut être réduite à une somme de carrés pondérés en calculant les valeurs et les vecteurs propres de la matrice A .

3. Si une matrice A d'ordre p admet p valeurs propres différentes λ_i , le système de vecteurs propres est une base de E .

4. Si A est une matrice carrée inversible d'ordre p qui admet p valeurs propres non nulles λ_i et les vecteurs propres V_i . Alors, la matrice inverse A^{-1} admet pour valeurs propres les inverses des valeurs propres de A et a les mêmes vecteurs propres que A . En effet, les vecteurs propres V_i de A , associés aux valeurs propres λ_i , satisfont au système d'équations $(A - \lambda_i I)V_i = 0$. En prémultipliant par A^{-1} on obtient :

$$(I - \lambda_i A^{-1})V_i \iff (A^{-1} - \lambda_i^{-1} I)V_i = 0$$

4.4.3 Recherche de valeurs et de vecteurs propres

Chercher les valeurs propres et les vecteurs propres des matrices :

$$A = \begin{bmatrix} 1 & 1 \\ 8 & 3 \end{bmatrix} \quad B = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \quad C = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}$$

\implies

La matrice A a deux valeurs propres *distinctes* : $\lambda_1 = 5$ et $\lambda_2 = -1$ (solutions de $(1 - \lambda)(3 - \lambda) - 8 = 0$).

La matrice B a deux valeurs propres *confondues* : $\lambda_1 = \lambda_2 = 1$ (solutions de $(1 - \lambda)^2 = 0$). 1 est valeur propre d'ordre 2.

La matrice C n'a qu'une valeur propre *réelle* : $\lambda = 1$ (solution de $(1 - \lambda)(\lambda^2 + 1) = 0$).

4.4.4 Vecteurs propres et droites de régression

On veut étudier la relation entre la longueur du corps x et la profondeur y de poitrine de 22 vaches laitières (Dagnélie, 1969). Les données sont présentées dans TAB. 4.1 :

x	y	x	y	x	y	x	y	x	y	x	y
168	71	169	68	159	70	159	73	148	68	161	69
150	65	148	67	151	69	155	71	151	70	176	74
154	67	145	66	169	74	158	70	146	71	150	65
165	69	163	69	157	71	161	73				

TAB. 4.1 - Longueur du corps (x) et de la profondeur de la poitrine (y) de 22 vaches laitières.

Comme on veut étudier les variations simultanées des deux caractéristiques x et y il n'y a pas plus de raisons d'utiliser la droite de régression de y en x que la droite de régression de x en y . Dagnelie (1969) propose d'utiliser le critère des moindres rectangles (minimum de la somme des produits des écarts $|x_i - x(y_i)|$ et $|y_i - y(x_i)|$) pour obtenir une droite de régression dite droite des *moindres rectangles*. Cette dernière droite de régression est comprise entre les droites de régression de y en x et x en y . Par ailleurs, en minimisant perpendiculairement à la droite elle-même on obtiendrait une nouvelle droite de régression dite droite de *régression orthogonale*. Ces quatre droites de régression passent toutes par le point moyen (\bar{x}, \bar{y}) , elles ont pour équations :

$$\text{-- régression de } x \text{ en } y : y = 0,519x - 12,2 \quad (1)$$

$$\text{-- droite des moindres rectangles : } y = 0,310x + 20,8 \quad (2)$$

$$\text{-- droite de régression orthogonale : } y = 0,196x + 38,7 \quad (3)$$

$$\text{-- régression de } y \text{ en } x : y = 0,185x + 40,5 \quad (4)$$

Sachant $\bar{x} = 157,41$, $\bar{y} = 69,55$, $var(x) = 68,95$, $var(y) = 6,614$, $cov(x, y) = 12,73$, retrouver l'équation :

1- de la droite de régression orthogonale en calculant les valeurs et les vecteurs propres de la matrice de variance-covariance V .

2- de la droite des moindres rectangles en calculant les valeurs propres et les vecteurs propres de la matrice de corrélation.

⇒

1- La matrice de variance-covariance V représente (à une constante multiplicative près) la matrice d'inertie du nuage de points (x, y) . Chercher à déterminer la droite de régression orthogonale est équivalent à chercher la direction de plus grande variabilité de ce nuage de points, c'est-à-dire le vecteur propre associé à la plus grande valeur propre de V . Pour cela, on résout : $det(V - \lambda I) = 0$ avec

$$V = \begin{bmatrix} 68,950 & 12,730 \\ 12,730 & 6,614 \end{bmatrix}$$

On déduit l'équation caractéristique $\lambda^2 - 75,564\lambda + 293,982 = 0$ qui admet deux racines : $\lambda_1 = 71,45$ et $\lambda_2 = 4,11$. Le vecteur propre de composantes v_1 et v_2 associé à la première valeur propre satisfait au système d'équations :

$$\begin{cases} -2,50v_1 + 12,73v_2 = 0 \\ 12,73v_1 - 64,84v_2 = 0 \end{cases}$$

Ces deux équations sont identiques. Si l'on choisit $v_1 = 1$, on déduit que $v_2 = 0,196$. La pente de ce vecteur propre est égale à $0,196$ (rapport v_2 sur

v_1). La droite de régression orthogonale $y = ax + b$ passe par le point moyen, la valeur du terme constant est donc : $b = 69,55 - 0,196 \times 157,41 = 38,70$.

2- La matrice de corrélation est :

$$C = \begin{bmatrix} 1 & 0,596 \\ 0,596 & 1 \end{bmatrix}$$

L'équation caractéristique est : $(1 - \lambda)^2 - 0,355 = 0$, elle admet deux racines $\lambda_1 = 1,596$ et $\lambda_2 = 0,404$. Le vecteur propre de composantes v_1 et v_2 associé à la première valeur propre satisfait au système d'équations :

$$\begin{cases} -0,596v_1 + 0,596v_2 = 0 \\ 0,596v_1 - 0,596v_2 = 0 \end{cases}$$

La pente du premier vecteur propre est égale à 1, c'est le rapport $\frac{v_2}{v_1} = \frac{0,596}{0,596}$. La matrice de corrélation n'est autre que la matrice de variance-covariance des variables centrées réduites. On déduit donc que $[(y - 69,55)/2,572]/[(x - 157,41)/8,304] = 1$ et $y = 0,310x + 20,8$ est l'équation de la droite des moindres rectangles. Ainsi, la droite des moindres rectangles n'est autre que la droite de régression orthogonale calculée sur les variables centrées réduites. Sur FIG. 4.4 on a représenté, les données et les différentes droites de régression.

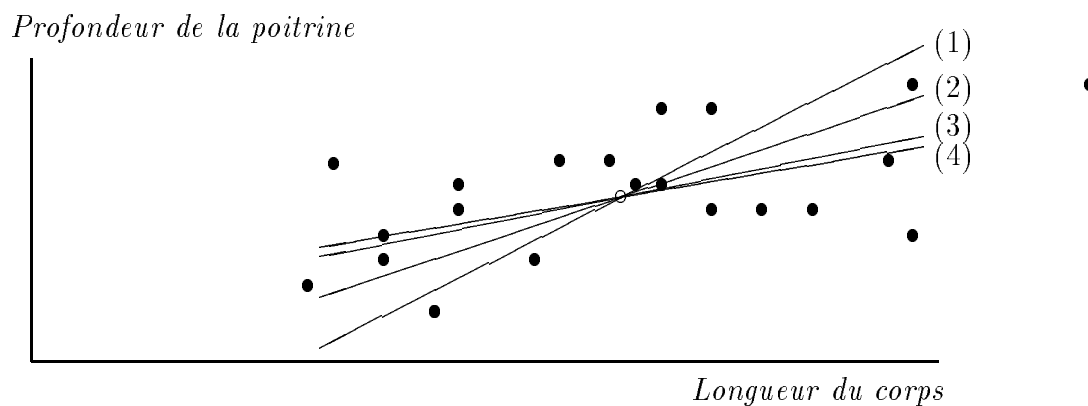


FIG. 4.4 - Régression de x en y (1), droite des moindres rectangles (2), régression orthogonale (3), régression de y en x (4).

4.5 Projecteurs associés à une décomposition en somme directe

Au paragraphe § 3.2.2 nous avons vu la décomposition en somme directe d'un espace vectoriel E de dimension p . Supposons qu'on ait une décomposition en somme directe de E en k sous espaces E_1, \dots, E_k de dimensions p_1, \dots, p_k ($E = E_1 \oplus \dots \oplus E_k$ et $p = p_1 + \dots + p_k$). Tout vecteur V de E s'écrit de façon

unique $V = \sum_{i=1}^k V_i$ avec $V_i \in E_i, i = 1, \dots, k$. Les k applications linéaires f_i

de E dans E qui font correspondre $f_i(V) = V_i \in E_i \subset E$ à V sont appelées *projecteurs* associés à la somme directe $E = E_1 \oplus \dots \oplus E_k$. Ces applications linéaires sont telles que $f_i(f_i(V)) = f_i(V_i) = V_i$, elles sont dites *idempotentes*.

Les projecteurs sont des applications linéaires idempotentes et réciproquement toute application linéaire idempotente est un projecteur. Les matrices A_i associées aux projecteurs f_i sont idempotentes ($A_i A_i = A_i$), leurs valeurs propres sont toutes égales à 1 ou à 0.

4.5.1 Projecteurs associés à des décompositions en somme directe

On considère dans \mathbb{R}^3 la base orthonormale

$$B = [B_1 \quad B_2 \quad B_3] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

et les décompositions en somme directe $\mathbb{R}^3 = D_1 \oplus P$ et $\mathbb{R}^3 = D_2 \oplus P$ avec P le sous espace vectoriel de \mathbb{R}^3 engendré par les vecteurs de base B_1 et B_2 , D_1 le sous espace de \mathbb{R}^3 engendré par le vecteur B_3 orthogonal à P et D_2 le sous espace de \mathbb{R}^3 engendré par la bissectrice de l'angle (B_1, B_3) .

La base B étant choisie, écrire les matrices associées :

1. aux projecteurs sur D_1 et P d'une part et aux projecteurs D_2 et P d'autre part. Vérifier que ces matrices sont idempotentes.
2. Ecrire les matrices associées aux différents projecteurs dans la base :

$$B' = [B_1 \quad B_2 \quad B'_3] = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

⇒

1. Sur FIG. 4.5 on a représenté les sous espaces impliqués dans les deux décompositions en somme directe. Soit f_D le projecteur sur le sous espace vectoriel de dimension 1 (la droite D_1 ou la droite D_2 suivant le cas étudié) et f_P le projecteur sur le sous espace vectoriel de dimension 2 (le plan P). f_D et f_P sont des applications linéaires de \mathbb{R}^3 dans \mathbb{R}^3 .

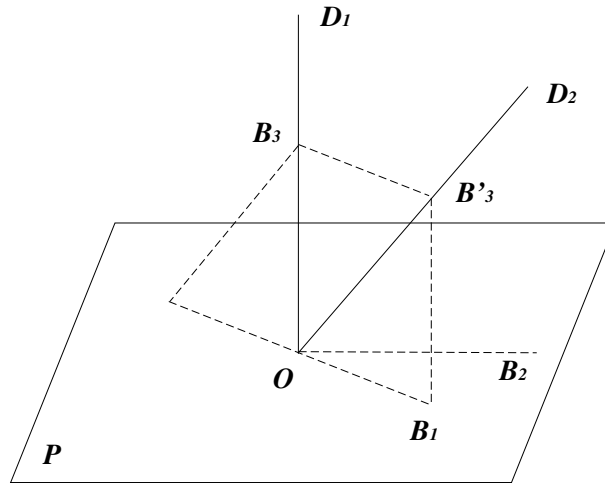


FIG. 4.5 - Décomposition de \mathbb{R}^3 en une somme directe : 1) droite D_1 et plan P , 2) droite D_2 et plan P .

Dans la décomposition en somme directe $\mathbb{R}^3 = D_1 \oplus P$ les sous espaces vectoriels D_1 et P sont complémentaires orthogonaux. Les projections sur ces sous espaces sont telles que :

- A tout point V de coordonnées v_1, v_2 et v_3 dans \mathbb{R}^3 , l'application linéaire f_{D_1} fait correspondre sa projection sur OB_3 c'est-à-dire le vecteur de composantes $(0, 0, v_3)$. La matrice associée à cette application linéaire est :

$$A_{D_1} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

car

$$A_{D_1}V = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ v_3 \end{bmatrix}$$

ce qui est le résultat attendu.

- A tout point V de coordonnées v_1, v_2 et v_3 dans \mathbb{R}^3 , l'application linéaire f_P fait correspondre sa projection sur le plan P , c'est-à-dire le vecteur de composantes $(v_1, v_2, 0)$. La matrice associée à cette application linéaire est :

$$A_P = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

car

$$A_P V = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} v_1 \\ v_2 \\ 0 \end{bmatrix}$$

ce qui est le résultat attendu.

- On vérifie, en effectuant les produits matriciels, que :

$$A_{D_1} A_{D_1} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} = A_{D_1}$$

$$A_P A_P = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} = A_P$$

Les matrices A_{D_1} et A_P sont idempotentes.

- Soit I la matrice unité, on remarque que $A_{D_1} = I - A_P$ et $A_P = I - A_{D_1}$. Ce résultat est général.

Dans la décomposition en somme directe $\mathbb{R}^3 = D_2 \oplus P$ les sous espaces vectoriels D_2 et P sont complémentaires, ils ne sont pas orthogonaux. Les projections sur ces sous espaces sont telles que :

- A tout point V de coordonnées v_1, v_2 et v_3 dans \mathbb{R}^3 , l'application linéaire f_{D_2} fait correspondre sa projection sur la bissectrice de l'angle (B_1, B_3) . Dans la base B , cette bissectrice est de composantes $(1, 0, 1)$. La matrice associée à cette application linéaire est :

$$A_{D_2} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

- En appliquant le résultat que $A_P = I - A_{D_2}$, on déduit que :

$$A_P = I - A_{D_2} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

- On vérifie, en effectuant les produits matriciels, que :

$$A_{D_2}A_{D_2} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} = A_{D_2}$$

$$A_P A_P = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} = A_P$$

Les matrices A_{D_2} et A_P sont idempotentes.

Dans les deux décompositions de \mathbb{R}^3 qui viennent d'être faites, on remarque que les matrices A_p associées aux projecteurs sur P , sont différentes.

2. Soit M la matrice des coordonnées des vecteurs B_1 , B_2 et B_3 dans la base B' (M est dite matrice de changement de base) :

$$M = \begin{bmatrix} 0 & 0 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Les matrices associées aux applications linéaires f_{D_1} et f_P d'une part, f_{D_2} et f_P d'autre part dans la nouvelle base sont :

$$A'_{D_1} = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad A'_P = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$A'_{D_2} = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad A'_P = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Ces matrices sont idempotentes, ce sont des matrices associées à des projecteurs.

Chapitre 5

Partitionnement - Différenciation - Extremum

5.1 Partitionnement

Il est souvent utile de considérer (quand des lignes ou des colonnes ont des caractéristiques communes) qu'une matrice est constituée d'éléments qui sont eux-mêmes des matrices. De telles matrices s'écrivent :

$$A = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1p} \\ A_{21} & A_{22} & \cdots & A_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ A_{n1} & A_{n2} & \cdots & A_{np} \end{bmatrix}$$

A est une matrice *partitionnée*, les matrices A_{ij} à n_i lignes et p_j colonnes sont appelées *sous-matrices*. Toutes les sous-matrices d'une ligne donnée de A ont le même nombre de lignes. De même, toutes les sous-matrices d'une colonne donnée de A ont le même nombre de colonnes.

Les opérations sur des matrices partitionnées ressemblent à celles déjà présentées, la seule différence réside dans la nature non scalaire des éléments de la matrice.

5.1.1 Opérations élémentaires

1. *Transposition*: On note A' la matrice transposée de la matrice A .

$$A' = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1p} \\ A_{21} & A_{22} & \cdots & A_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ A_{n1} & A_{n2} & \cdots & A_{np} \end{bmatrix}' = \begin{bmatrix} A'_{11} & A'_{21} & \cdots & A'_{n1} \\ A'_{12} & A'_{22} & \cdots & A'_{n2} \\ \vdots & \vdots & \vdots & \vdots \\ A'_{1p} & A'_{2p} & \cdots & A'_{np} \end{bmatrix}$$

2. *Addition*: Soient A et B deux matrices partitionnées formées de sous-matrices de *dimensions similaires*. L'addition $A + B$ est :

$$A + B = \begin{bmatrix} A_{11} + B_{11} & A_{12} + B_{12} & \cdots & A_{1p} + B_{1p} \\ A_{21} + B_{21} & A_{22} + B_{22} & \cdots & A_{2p} + B_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ A_{n1} + B_{n1} & A_{n2} + B_{n2} & \cdots & A_{np} + B_{np} \end{bmatrix}$$

3. *Multiplication*: Quand les dimensions des sous-matrices de A et B sont conformes à la règle du produit matriciel, le produit de A par B s'écrit :

$$AB = \begin{bmatrix} \sum_{i=1}^p A_{1i}B_{i1} & \cdots & \sum_{i=1}^p A_{1i}B_{im} \\ \dots & \dots & \dots \\ \sum_{i=1}^p A_{ni}B_{i1} & \cdots & \sum_{i=1}^p A_{ni}B_{im} \end{bmatrix}$$

4. *Inversion*: Il est possible d'exprimer l'inverse d'une matrice partitionnée non singulière en termes de sous-matrices. Un cas important en statistique est le suivant :

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

où A_{11} et A_{22} sont toutes deux des sous-matrices carrées non singulières. Si l'on note :

$$B = A^{-1} = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}$$

on peut vérifier que :

$$B_{11} = (A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1}$$

$$B_{22} = A_{22}^{-1} + A_{22}^{-1}A_{21}B_{11}A_{12}A_{22}^{-1}$$

$$B_{12} = -A_{11}^{-1}A_{12}B_{22}$$

$$B_{21} = -A_{22}^{-1}A_{21}B_{11}$$

Une autre expression de A^{-1} peut être obtenue en permutant les rôles de A_{11} et A_{22} . On déduit alors de ces deux expressions de A^{-1} une identité matricielle très utile :

$$(A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1} = A_{11}^{-1} + A_{11}^{-1}A_{12}(A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1}A_{21}A_{11}^{-1}$$

En particulier, si A est une matrice non singulière ($n \times n$), B un vecteur ($n \times 1$), α un scalaire, on a l'identité de Bartlett (1951) :

$$(A + \alpha BB')^{-1} = A^{-1} - \frac{\alpha}{1 + \alpha B' A^{-1} B} A^{-1} B B' A^{-1}$$

5. *Déterminant*: Il est parfois nécessaire de calculer le déterminant de la matrice partitionnée (1). Si A_{11} est non singulière,

$$\det(A) = \det(A_{11}) \times \det(A_{22} - A_{21}A_{11}^{-1}A_{12})$$

Si A_{22} est non singulière,

$$\det(A) = \det(A_{22}) \times \det(A_{11} - A_{12}A_{22}^{-1}A_{21})$$

5.1.2 Inverse et déterminant d'une matrice partitionnée

Calculer l'inverse et le déterminant de :

$$A = \begin{bmatrix} 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 2 & 2 & 2 \\ 3 & -1 & 4 & 0 & 0 \\ 3 & -1 & 0 & 4 & 0 \\ 3 & -1 & 0 & 0 & 4 \end{bmatrix}$$

\Rightarrow

$$A_{11}^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad A_{22}^{-1} = \begin{bmatrix} \frac{1}{4} & 0 & 0 \\ 0 & \frac{1}{4} & 0 \\ 0 & 0 & \frac{1}{4} \end{bmatrix}$$

$$A_{11} - A_{12}A_{22}^{-1}A_{21} = \begin{bmatrix} -\frac{5}{4} & \frac{3}{4} \\ -\frac{9}{2} & \frac{5}{2} \end{bmatrix}$$

$$(A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1} = \begin{bmatrix} 10 & -3 \\ 18 & -5 \end{bmatrix}$$

$$A_{12}A_{22}^{-1} = \begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{bmatrix} \quad A_{22}^{-1}A_{21} = \begin{bmatrix} \frac{3}{4} & -\frac{1}{4} \\ \frac{3}{4} & -\frac{1}{4} \\ \frac{3}{4} & -\frac{1}{4} \end{bmatrix}$$

Ainsi,

$$A^{-1} = \begin{bmatrix} 10 & -3 & -1 & -1 & -1 \\ 18 & -5 & -2 & -2 & -2 \\ -3 & 1 & \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ -3 & 1 & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ -3 & 1 & \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{bmatrix}$$

$$\det(A) = 6$$

5.2 Différenciation

On a parfois besoin de dériver des fonctions de vecteurs ou de matrices. Si l'on désigne par $f(X)$ une fonction continue des éléments d'un vecteur $X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$,

$\frac{\partial f(X)}{\partial x_i}$ et $\frac{\partial^2 f(X)}{\partial x_i \partial x_j}$ les dérivées partielles première et seconde. Le vecteur des dérivées partielles est :

$$\frac{\partial f(X)}{\partial X} = \begin{bmatrix} \frac{\partial f(X)}{\partial x_1} \\ \frac{\partial f(X)}{\partial x_2} \\ \vdots \\ \frac{\partial f(X)}{\partial x_n} \end{bmatrix}$$

1. Si $f(X)$ est constante pour tout X :

$$\frac{\partial f(X)}{\partial X} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

2. Soit C un vecteur $(n \times 1)$ constant et soit

$$f(X) = C'X = [c_1 \quad c_2 \quad \cdots \quad c_n] \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$$\frac{\partial f(X)}{\partial X} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}$$

La forme colonne du vecteur dérivée est inchangée si l'on écrit $f(X) = X'C$.

3. Vecteur des dérivées d'une forme quadratique $X'AX$:

$$\frac{\partial X'AX}{\partial X} = 2AX$$

Plus généralement, si $f(X) = (C - BX)'A(C - BX)$ avec A une matrice symétrique de dimension $(p \times p)$, $B = [B_1 \quad \cdots \quad B_n]$ une matrice de constantes de dimension $(p \times n)$ et C un vecteur de dimension $(p \times 1)$ alors :

$$\frac{\partial f(X)}{\partial X} = -2B'A(C - BX)$$

4. La matrice des dérivées partielles du second ordre, d'une fonction de n variables, est appelée matrice *hessienne* :

$$H = \frac{\partial^2 f(X)}{\partial X' \partial X} = \begin{bmatrix} \frac{\partial^2 f(X)}{\partial x_1^2} & \cdots & \frac{\partial^2 f(X)}{\partial x_1 \partial x_n} \\ \cdots & \cdots & \cdots \\ \frac{\partial^2 f(X)}{\partial x_1 \partial x_n} & \cdots & \frac{\partial^2 f(X)}{\partial x_n^2} \end{bmatrix}$$

Par exemple, la matrice hessienne de $f(X) = X'AX$ est $2A$. La matrice hessienne est symétrique si les conditions de continuité et d'existence de toutes les dérivées première et seconde sont satisfaites par $f(X)$.

5.3 Extremum

La condition nécessaire pour obtenir un maximum ou un minimum de $f(X)$ au point $X = X_0$ est $\frac{\partial f(X)}{\partial X} = 0$ en ce point. Une condition suffisante pour avoir en X_0 :

1. un *Maximum*, est que la matrice hessienne évaluée en X_0 soit *définie négative*.
2. un *Minimum*, est que la matrice hessienne évaluée en X_0 soit *définie positive*.

5.3.1 Maximisation sous contrainte

Pour maximiser (minimiser) une fonction $f(X)$ sous la contrainte $g(X) = c$ on utilise la méthode des *multiplicateurs de Lagrange* qui consiste à écrire la nouvelle fonction :

$$h(X, \lambda) = f(X) - \lambda[g(X) - c]$$

On cherche l'extremum en résolvant :

$$\frac{\partial h(X, \lambda)}{\partial X} = \frac{\partial f(X)}{\partial X} - \lambda \frac{\partial g(X)}{\partial X} = 0$$

$$\frac{\partial h(X, \lambda)}{\partial \lambda} = -g(X) + c = 0$$

Cette seconde équation n'étant autre que la condition initiale.

5.4 Interprétation géométrique des valeurs et des vecteurs propres

Considérons dans l'espace des individus de dimension $p = 2$ l'ensemble des points M situés sur l'ellipse centrée en μ d'équation $(X - \mu)' \Sigma^{-1} (X - \mu) = 1$. Avec :

$$X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \quad \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \quad \Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$$

On veut déterminer les axes principaux de cette ellipse.

Le premier axe principal est la droite qui passe par le centre μ de l'ellipse et qui la traverse dans sa plus grande dimension. Le second axe principal de l'ellipse passe aussi par μ , c'est une droite perpendiculaire au premier axe principal. Donc, le demi-premier axe principal de l'ellipse est celui qui maximise la distance d'un point de l'ellipse au centre μ de cette ellipse. Soit :

$$(X - \mu)'(X - \mu) \text{ maximum sachant } (X - \mu)'\Sigma^{-1}(X - \mu) = 1$$

Pour maximiser la forme quadratique $(X - \mu)'(X - \mu)$ sous la contrainte $(X - \mu)'\Sigma^{-1}(X - \mu) = 1$ on utilise la méthode des multiplicateurs de Lagrange. La fonction à maximiser est :

$$h(X, \lambda) = (X - \mu)'(X - \mu) - \lambda[(X - \mu)'\Sigma^{-1}(X - \mu) - 1]$$

on annule sa dérivée par rapport à X . soit :

$$\frac{\partial h(X, \lambda)}{\partial X} = 2(X - \mu) - 2\lambda\Sigma^{-1}(X - \mu) = 0 \quad (5.1)$$

En prémultipliant par Σ on obtient le système d'équations homogène $(\Sigma - \lambda I)(X - \mu) = 0$ qui admet des solutions non triviales si et seulement si $\det(\Sigma - \lambda I) = 0$. Cette dernière équation n'est autre que le polynôme caractéristique. Ainsi,

- le premier axe principal de l'ellipse est un vecteur propre de Σ .

En prémultipliant (5.1) par $(X - \mu)'$ on obtient :

$$\lambda = \frac{(X - \mu)'(X - \mu)}{(X - \mu)'\Sigma^{-1}(X - \mu)}$$

La valeur propre obtenue apparaît donc comme un rapport de carrés de longueurs associées à deux métriques définies sur E (la métrique unité I et la métrique Σ^{-1}). D'autre part, en utilisant la contrainte, on obtient $(X - \mu)'(X - \mu) = \lambda$. On peut donc conclure que le premier axe principal n'est autre que le vecteur propre associé à la plus grande valeur propre de Σ .

- Le second axe principal est le vecteur propre associé à la plus petite valeur propre car il n'y a que deux valeurs propres distinctes et on a démontré que des vecteurs propres associés à des valeurs propres distinctes étaient orthogonaux. Pour les mêmes raisons que celles qui viennent d'être évoquées, la seconde valeur propre est aussi un rapport de carrés de longueurs définies par les métriques I et Σ^{-1} .

5.4.1 Relations entre deux ellipsoïdes

Les résultats précédents se généralisent au cas de deux ellipsoïdes dans un espace vectoriel de dimension quelconque muni de métriques Σ_1 et Σ_2 .

Soit E un espace vectoriel de dimension p , deux produits scalaires π_1 et π_2 peuvent être définis sur E en spécifiant leurs ellipsoïdes dans E . Soient E_1 et E_2 les ellipsoïdes centrés au point moyen μ :

$$(X - \mu)' \Sigma_1^{-1} (X - \mu) = 1 \text{ et } (X - \mu)' \Sigma_2^{-1} (X - \mu) = 1$$

Pour ces deux ellipsoïdes il existe un ensemble de p droites passant par μ qui définissent des ensembles d'axes conjugués communs à E_1 et E_2 . Les directions de ces p droites sont déterminées par les segments de droite $\mu B_1, \dots, \mu B_p$ où B_1, \dots, B_p est une base de vecteurs propres de π_1 relative à π_2 .

La valeur propre λ_i ($i = 1, \dots, p$) est le rapport du carré de la longueur du demi-axe de E_2 dans la direction B_i au carré de la longueur du demi-axe de E_1 dans la même direction.

Pour illustrer ce qui vient d'être dit, nous considérons l'espace de dimension $p = 2$, la situation qui vient d'être décrite est représentée sur la figure 5.1.

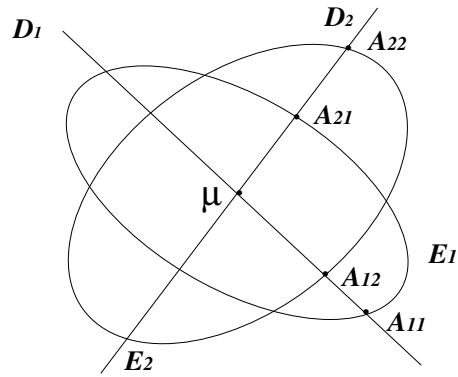


FIG. 5.1 - Ellipses E_1 et E_2 centrées en μ . D_1 et D_2 définissent un couple d'axes conjugués. Les valeurs propres de E_1 relatives à E_2 sont $\lambda_i = [\mu A_{i2} / \mu A_{i1}]^2$ ($i = 1, 2$).

Soient maintenant les ellipses obtenues en considérant que E est un espace Euclidien pour E_1 (figure 5.2). Dans cet espace Euclidien E_1 devient un cercle et les axes conjugués de E_2 correspondent aux axes principaux. Les valeurs propres λ_i , ($i = 1, 2$) sont les inverses des carrés des longueurs des demi-axes principaux.

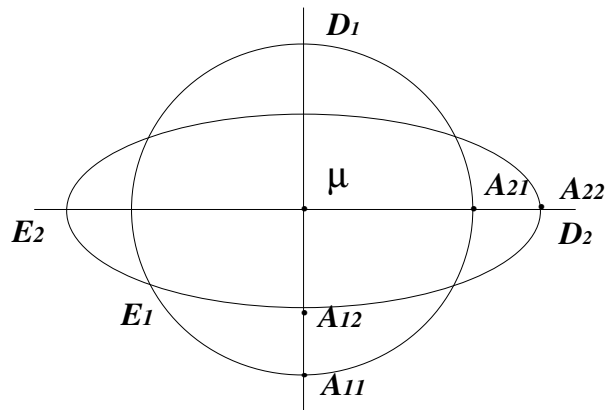


FIG. 5.2 - Même chose que pour la Figure 5.1 avec μA_{21} et μA_{11} perpendiculaires, même unité de longueur sur ces axes de telle sorte que E_1 soit un cercle unité.

5.4.2 A propos de la maximisation

D'un point de vue plus général il est facile de montrer que les valeurs propres rangées dans l'ordre décroissant assurent les maxima du rapport de deux formes quadratiques $X'\Sigma_1 X$ et $X'\Sigma_2 X$ dans des directions orthogonales. Pour cela, il suffit de différencier, par rapport à X :

$$\frac{X'\Sigma_1 X}{X'\Sigma_2 X}$$

puis, d'annuler cette dérivée et enfin d'utiliser la propriété d'orthogonalité des vecteurs propres.

$$\begin{aligned} \frac{\partial}{\partial X} \left(\frac{X'\Sigma_1 X}{X'\Sigma_2 X} \right) &= \frac{(X'\Sigma_2 X)(2\Sigma_1 X) - (X'\Sigma_1 X)(2\Sigma_2 X)}{(X'\Sigma_2 X)^2} = 0 \\ &\Downarrow \\ \Sigma_1 X &= \left(\frac{X'\Sigma_1 X}{X'\Sigma_2 X} \right) \Sigma_2 X \\ &\Downarrow \\ [\Sigma_2^{-1} \Sigma_1 - \left(\frac{X'\Sigma_1 X}{X'\Sigma_2 X} \right) I] X &= 0 \end{aligned}$$

On a obtenu le système homogène d'équations relatif aux vecteurs propres X de la matrice $\Sigma_2^{-1} \Sigma_1$ et aux valeurs propres associées :

$$\lambda = \frac{X'\Sigma_1 X}{X'\Sigma_2 X}$$

5.5 Dérivée d'un déterminant

La dérivée du déterminant d'une matrice $A(n \times n)$ par rapport à ses éléments a_{ij} peut être obtenue à partir du développement de A en cofacteurs de la i ème ligne ou de la j ème colonne.

$$\frac{\partial \det(A)}{\partial a_{ij}} = \frac{\partial}{\partial a_{ij}}(a_{i1}A_{i1} + \cdots + a_{ij}A_{ij} + \cdots + a_{in}A_{in})$$

Si A est symétrique :

$$\frac{\partial \det(A)}{\partial a_{ii}} = A_{ii}, \quad \frac{\partial \det(A)}{\partial a_{ij}} = 2A_{ij}$$

5.6 Quelques formules utiles

1. Soit $X(n \times p)$ une matrice d'éléments x_{ij} , sa dérivée par rapport à x_{ij} est $\frac{\partial X}{\partial x_{ij}} = J_{ij}$ avec $J_{ij}(n \times p)$ est la matrice ayant un 1 en ij ème position et des zéros partout ailleurs. Si X est symétrique :

$$\frac{\partial X}{\partial x_{ij}} = J_{ij} + J_{ji} \quad i \neq j$$

2. La règle pour différencier des produits de matrices est identique à celle utilisée pour les scalaires. Supposons que les éléments $x_{ij}(z)$ et $y_{ij}(z)$ des matrices $X(n \times p)$ et $Y(p \times q)$ sont fonctions d'une variable z . Alors,

$$\frac{\partial XY}{\partial z} = \frac{\partial X}{\partial z}Y + X\frac{\partial Y}{\partial z}$$

3. Cette dernière formule permet de différencier l'inverse d'une matrice carrée non singulière. Soit $I = XX^{-1}$, alors :

$$\frac{\partial I}{\partial x_{ij}} = J_{ij}X^{-1} + X\frac{\partial X^{-1}}{\partial x_{ij}} = 0$$

On en déduit que : $\frac{\partial X^{-1}}{\partial x_{ij}} = -X^{-1}J_{ij}X^{-1}$

Si X est symétrique : $\frac{\partial X^{-1}}{\partial x_{ij}} = -X^{-1}J_{ii}X^{-1} \quad i = j$

$$\frac{\partial X^{-1}}{\partial x_{ij}} = -X^{-1}(J_{ij} + J_{ji})X^{-1} \quad i \neq j$$

5.7 Description d'un ensemble d'observations

Soit le tableau 5.1 des données iris constitué par des longueurs (X) et des largeurs (Y) des sépales mesurées sur $n_1 = 50$ iris de la variété *Versicolor* et $n_2 = 50$ iris de la variété *Setosa* (R. A. Fisher (1936); Kendall et Stuart, vol 3).

iris <i>Versicolor</i>				iris <i>Setosa</i>			
X	Y	X	Y	X	Y	X	Y
7,0	3,2	6,6	3,0	5,1	3,5	5,0	3,0
6,4	3,2	6,8	2,8	4,9	3,0	5,0	3,4
6,9	3,1	6,7	3,0	4,7	3,2	5,2	3,5
5,5	2,3	6,0	2,9	4,6	3,1	5,2	3,4
6,5	2,8	5,7	2,6	5,0	3,6	4,7	3,2
5,7	2,8	5,5	2,4	5,4	3,9	4,8	3,1
6,3	3,3	5,5	2,4	4,6	3,4	5,4	3,4
4,9	2,4	5,8	2,7	5,0	3,4	5,2	4,1
6,6	2,9	6,0	2,7	4,4	2,9	5,5	4,2
5,2	2,7	5,4	3,0	4,9	3,1	4,9	3,1
5,0	2,0	6,0	3,4	5,4	3,7	5,0	3,2
5,9	3,0	6,7	3,1	4,8	3,4	5,5	3,5
6,0	2,2	6,3	2,3	4,8	3,0	4,9	3,6
6,1	2,9	5,6	3,0	4,3	3,0	4,4	3,0
5,6	2,9	5,5	2,5	5,8	4,0	5,1	3,4
6,7	3,1	5,5	2,6	5,7	4,4	5,0	3,5
5,6	3,0	6,1	3,0	5,4	3,9	4,5	2,3
5,8	2,7	5,8	2,6	5,1	3,5	4,4	3,2
6,2	2,2	5,0	2,3	5,7	3,8	5,0	3,5
5,6	2,5	5,6	2,7	5,1	3,8	5,1	3,8
5,9	3,2	5,7	3,0	5,4	3,4	4,8	3,0
6,1	2,8	5,7	2,9	5,1	3,7	5,1	3,8
6,3	2,5	6,2	2,9	4,6	3,6	4,6	3,2
6,1	2,8	5,1	2,5	5,1	3,3	5,3	3,7
6,4	2,9	5,7	2,8	4,8	3,4	5,0	3,3

TAB. 5.1 - longueur (X), largeur (Y) des sépales mesurées sur iris *Setosa* et iris *Versicolor*.

Sur la figure 5.3 on a représenté le nuage des points observés. Pour décrire ces données, on estimera des paramètres de position (moyenne générale, moyenne de chacune des populations d'iris), des paramètres de dispersion (des matrices d'inertie).

Largeur des sépales

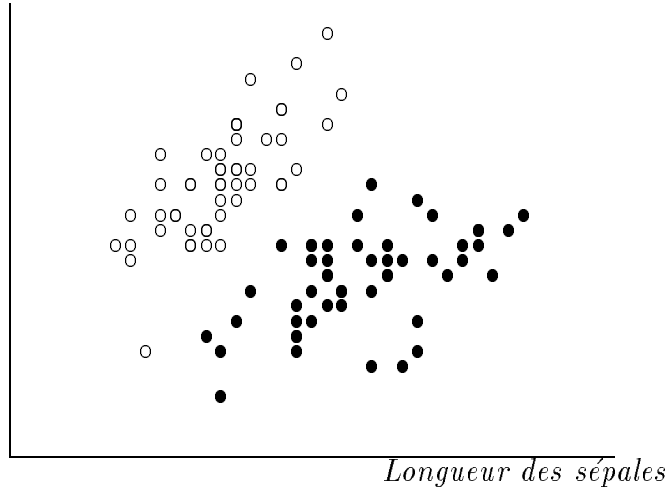


FIG. 5.3 - Représentation des données : iris Versicolor (●), iris Setosa (○).

1. Les matrices d'inertie des populations :

$$W_i, (i = 1 : \text{iris Versicolor}, i = 2 : \text{iris Setosa})$$

$$W_i = \begin{bmatrix} \sum_{j=1}^{n_i} (x_{ij} - x_{i.})^2 & \sum_{j=1}^{n_i} (x_{ij} - x_{i.})(y_{ij} - y_{i.}) \\ \sum_{j=1}^{n_i} (x_{ij} - x_{i.})(y_{ij} - y_{i.}) & \sum_{j=1}^{n_i} (y_{ij} - y_{i.})^2 \end{bmatrix}$$

2. La matrice somme des inerties des populations :

$$W = W_1 + W_2$$

$$W = \begin{bmatrix} \sum_{i=1}^2 \sum_{j=1}^{n_i} (x_{ij} - x_{i.})^2 & \sum_{i=1}^2 \sum_{j=1}^{n_i} (x_{ij} - x_{i.})(y_{ij} - y_{i.}) \\ \sum_{i=1}^2 \sum_{j=1}^{n_i} (x_{ij} - x_{i.})(y_{ij} - y_{i.}) & \sum_{i=1}^2 \sum_{j=1}^{n_i} (y_{ij} - y_{i.})^2 \end{bmatrix}$$

3. La matrice d'inertie des points moyens des populations : B

$$B = \begin{bmatrix} \sum_{i=1}^2 n_i (x_{i.} - x_{..})^2 & \sum_{i=1}^2 n_i (x_{i.} - x_{..})(y_{i.} - y_{..}) \\ \sum_{i=1}^2 n_i (x_{i.} - x_{..})(y_{i.} - y_{..}) & \sum_{i=1}^2 n_i (y_{i.} - y_{..})^2 \end{bmatrix}$$

$$B = \frac{n_1 \times n_2}{n_1 + n_2} \begin{bmatrix} (x_{1.} - x_{2.})^2 & (x_{1.} - x_{2.})(y_{1.} - y_{2.}) \\ (x_{1.} - x_{2.})(y_{1.} - y_{2.}) & (y_{1.} - y_{2.})^2 \end{bmatrix}$$

4. La matrice d'inertie totale: $T = B + W$

A chacune de ces matrices d'inertie est associé un nombre de *degrés de liberté* (*ddl*) et on estime les matrices de variance-covariance en divisant ces matrices d'inertie par les degrés de liberté correspondants (TAB. 5.2).

	<i>ddl</i>	<i>Matrices</i>	
		<i>Inertie</i>	<i>Var-cov</i>
Populations ($i = 1, 2$)	$n_i - 1$	W_i	$\frac{W_i}{n_i - 1}$
Somme populations	$n_1 + n_2 - 2$	$W = W_1 + W_2$	$\frac{W}{n_1 + n_2 - 2}$
Points moyens populations	1	B	B
Totale	$n - 1$	T	$\frac{T}{n - 1}$

TAB. 5.2 - *Matrices d'inertie, de variance-covariance, degrés de liberté (ddl).*

Estimations des paramètres

1. Paramètres de position

Les estimations des moyennes de chaque variété d'iris et de la moyenne générale sont données dans le tableau 5.3.

	<i>Moyennes</i>		
	<i>Versicolor</i>	<i>Setosa</i>	<i>Générale</i>
Longueur sépales	5,936	5,006	5,471
Largeur sépales	2,770	3,428	3,099

TAB. 5.3 - *Estimations des moyennes.*

2. Paramètres de dispersion

(a) *Inertie population*: $W_1 = \begin{bmatrix} 13,055 & 4,175 \\ 4,175 & 4,825 \end{bmatrix}$ $W_2 = \begin{bmatrix} 6,088 & 4,865 \\ 4,865 & 7,046 \end{bmatrix}$

$$W = \begin{bmatrix} 19,143 & 9,036 \\ 9,036 & 11,866 \end{bmatrix}$$

$$(b) \text{ Inertie des points moyens des populations: } B = \begin{bmatrix} 21,623 & -15,299 \\ -15,299 & 10,824 \end{bmatrix}$$

$$(c) \text{ Inertie totale: } T = \begin{bmatrix} 40,766 & -6,263 \\ -6,263 & 22,690 \end{bmatrix}$$

Pour décrire les populations d'iris, on utilise l'information véhiculée par les paramètres de position et de dispersion (résumés des données). Pour cela, on cherche les directions de plus grande variabilité des nuages de points c'est-à-dire les vecteurs propres des matrices d'inertie W_1 , W_2 et W . Les résultats sont donnés dans le tableau 5.4 :

	W_1		W_2		W	
Valeurs propres	14,802	3,078	11,455	1,679	25,246	5,764
Vecteurs propres	1	1	1	1	1	1
	0,418	-2,390	1,103	-0,906	0,675	-1,481

TAB. 5.4 - Valeurs propres et vecteurs propres des matrices d'inertie W_1 , W_2 et W .

On peut vérifier que les deux vecteurs propres associés à chaque matrice W_1 , W_2 et W sont orthogonaux entre eux (le produit scalaire de ces deux vecteurs propres est égal à zéro).

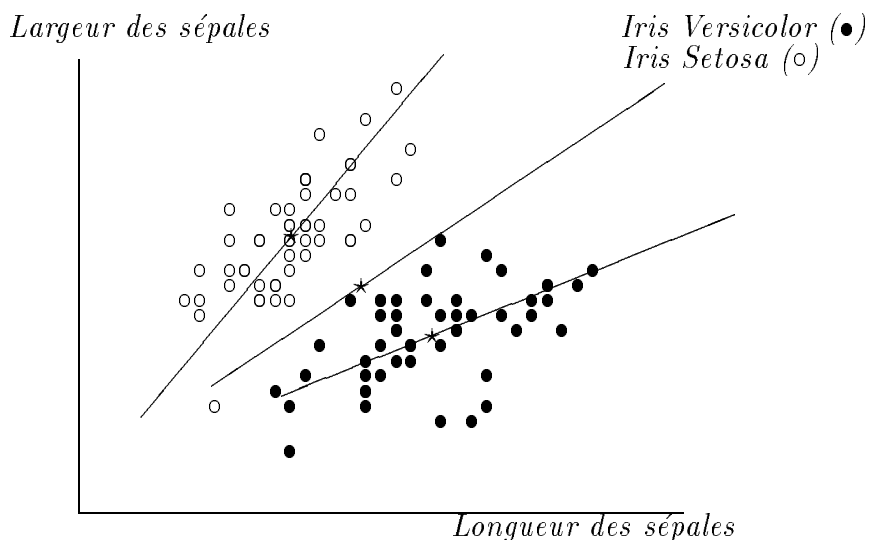


FIG. 5.4 - Premier vecteur propre de W_1, W_2 et W , points moyens (*).

On a représenté la première direction propre de W_1 , W_2 et W sur la figure 5.4. De même, la direction d'inertie maximum des points moyens des populations est obtenue en calculant les valeurs et vecteurs propres de la matrice B . Comme par deux points, il ne passe qu'une droite et qu'une seule, il faut s'attendre à ne trouver qu'une valeur propre non nulle et donc une seule direction propre. On obtient :

$$\lambda = 32,447 \quad \begin{bmatrix} 1 \\ -0,707 \end{bmatrix}$$

Les résultats obtenus montrent que les deux populations d'iris sont représentées par des nuages de points qui ont des orientations et des dispersions comparables. On peut donc admettre que $W = W_1 + W_2$ représente la matrice d'inertie commune aux deux populations. Par ailleurs, si l'on trace la direction propre de B et si l'on projette orthogonalement sur cette direction les points individuels de chacune des populations d'iris, on constate que les iris *Versicolor* sont positionnés à droite et les iris *Setosa* à gauche du point moyen. Le vecteur propre de B est donc une direction de l'espace sur laquelle les projections orthogonales des observations relatives à chacune des deux populations d'iris sont séparées mais on peut mieux faire. En effet, cette direction ne prend pas en compte l'inertie W commune aux deux populations, ce n'est donc pas la direction de l'espace qui permet de séparer au mieux les deux populations d'iris. La question qui se pose est donc de savoir s'il existe, dans le plan défini par les variables X et Y , une direction optimale pour séparer les populations d'iris.

Pour cela, on cherche une combinaison linéaire des X et Y ($Z = v_1X + v_2Y$) telle que, la distance entre les points moyens des populations d'iris soit maximum compte tenu des caractéristiques de chacune de ces populations. Autrement dit, on cherche à maximiser l'inertie des points moyens des populations (B) par rapport à l'inertie des populations (W). Soient :

$$W_Z = \sum_{j=1}^{n_1} (Z_{1j} - Z_{1.})^2 + \sum_{j=1}^{n_2} (Z_{2j} - Z_{2.})^2 = V'WV$$

$$B_Z = \frac{n_1 n_2}{n_1 + n_2} (Z_{1.} - Z_{2.})^2 = V'BV$$

$$V = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$$

On résout ce problème soit en utilisant les résultats présentés au § 5.4.2 soit en maximisant $\phi = \frac{V'BV}{V'WV}$ sous la contrainte $V'WV = 1$. Dans ce cas, on

utilise la méthode des multiplicateurs de Lagrange présentée au § 5.3.1. Nous avons choisi d'utiliser la méthode des multiplicateurs de Lagrange :

1. On écrit : $h(V, \lambda) = V'BV - \lambda(V'WV - 1) = V'[B - \lambda W]V - \lambda$
2. On dérive $h(V, \lambda)$ par rapport à V : $\frac{\partial h(V, \lambda)}{\partial V} = 2[B - \lambda W]V$
3. On annule la dérivée : $[B - \lambda W]V = 0 \iff [W^{-1}B - \lambda I]V = 0$

On retrouve un problème de valeur et de vecteur propre.

Pour l'exemple des iris, on a le résultat suivant :

$$W^{-1} = \begin{bmatrix} 0,082 & -0,062 \\ -0,062 & 0,132 \end{bmatrix} \quad W^{-1}B = \begin{bmatrix} 2,713 & -1,920 \\ -3,356 & 2,374 \end{bmatrix}$$

Equation caractéristique : $\lambda^2 - 5,088\lambda = 0$

Une seule valeur propre non nulle : $\lambda = 5,088$

Un seul vecteur propre : $V = \begin{bmatrix} 1 \\ -1,237 \end{bmatrix}$

Cette direction est représentée sur FIG. 5.5.

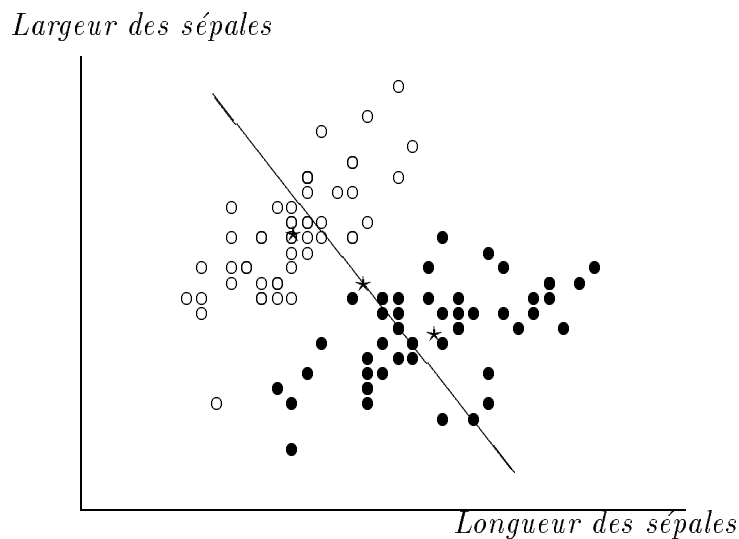


FIG. 5.5 - Premier vecteur propre de $W^{-1}B$.

5.7.1 Illustration géométrique

On veut décrire par la géométrie les différentes étapes qui ont été suivies dans l'exemple des iris. Nous avons considéré deux familles d'iris, elles peuvent être représentées par deux ellipses centrées en μ_1 et μ_2 . On passe de la première à la seconde par une translation $\mu_1\mu_2$. Si l'on projette ces ellipses sur un axe parallèlement à une direction donnée D_1 , les centres des ellipses projetées sont les projections de μ_1 et μ_2 parallèlement à cette direction. Les écarts types s'obtiennent à partir des tangentes aux ellipses parallèles à cette direction. Dans l'ensemble des directions du plan, il en existe une telle que le rapport de la distance $\mu_1\mu_2$ à l'écart type commun soit maximum, c'est la direction D_2 conjuguée de $\mu_1\mu_2$ par rapport aux ellipses (Cyffers, 1965). La situation qui vient d'être décrite est représentée sur FIG. 5.6.

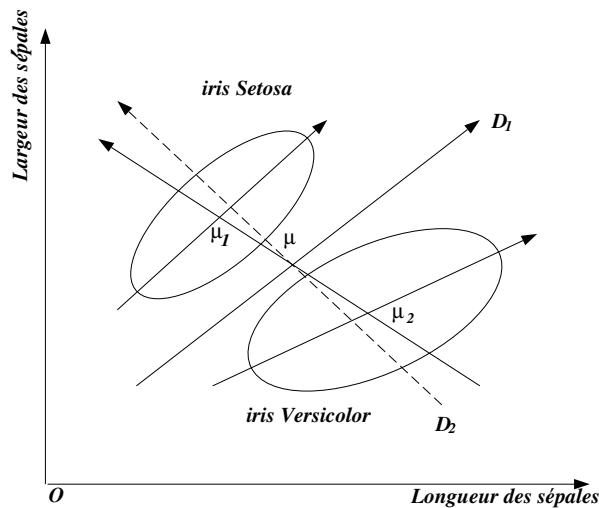


FIG. 5.6 - Ellipses d'inertie centrées aux points moyens μ_1 (*iris Versicolor*) et μ_2 (*iris Setosa*), direction de projection (D_1) et axe conjugué (D_2) de $\mu_1\mu_2$ par rapport aux ellipses.

Considérons maintenant un cas moins pathologique en nous plaçant dans un cadre un peu plus général tel que : l'espace des individus est toujours défini par deux variables X_1 et X_2 , mais dans lequel nous avons trois populations P_1, P_2, P_3 au lieu de deux. On fait cette supposition pour pouvoir représenter l'ellipse d'inertie des points moyens μ_1, μ_2, μ_3 des populations et on suppose que ces points moyens ne sont pas alignés.

1. Représentation des données

On représente les données par les contours des ellipses d'inertie associées à chacune des populations P_1, P_2 et P_3 . Ces ellipses sont centrées aux points moyens des populations (FIG. 5.7).

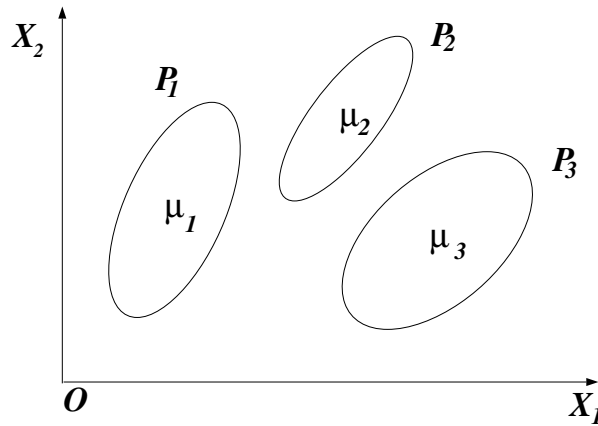


FIG. 5.7 - Ellipses d'inertie des populations P_1 , P_2 et P_3 centrées respectivement aux points moyens μ_1 , μ_2 et μ_3 .

2. Représentation des ellipses : 1) inertie commune aux trois populations, 2) inertie des points moyens μ_1 , μ_2 et μ_3

Les trois populations étant représentées par des ellipses d'inertie, approximativement orientées dans la même direction et de tailles comparables, on représente l'inertie commune aux trois populations par une ellipse E_1 centrée au point moyen μ et l'inertie des points moyens par une ellipse E_2 elle aussi centrée au point moyen μ (FIG. 5.8).

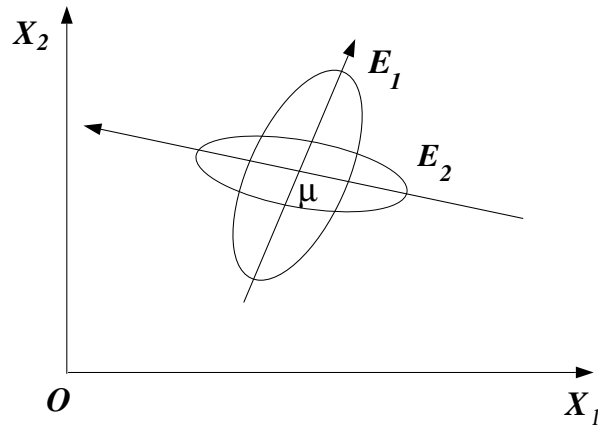


FIG. 5.8 - Représentation de l'ellipse d'inertie E_1 commune aux trois populations et de l'ellipse d'inertie E_2 des points moyens.

3. Que cherche-t-on?

On veut trouver, en tenant compte de l'inertie intra-population, la direction de l'espace telle que les projections de μ_1 , μ_2 et μ_3 sur cette direction soient les plus éloignées possible. Autrement dit, M_1 étant un point de

l'ellipse E_1 et M_2 le point de l'ellipse E_2 à l'intersection du demi-axe μM_1 et de E_2 , on cherche à maximiser la longueur du segment μM_2 par rapport à la longueur du segment μM_1 (FIG. 5.9).

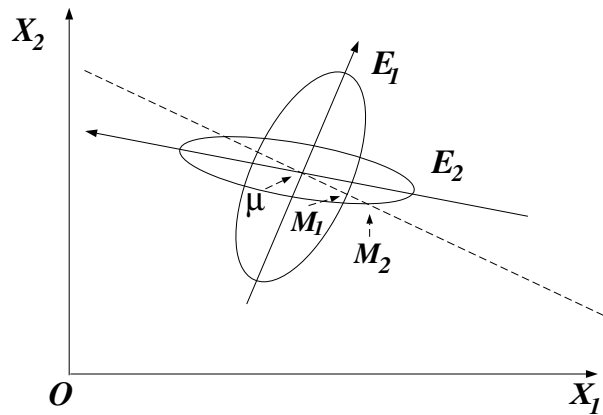


FIG. 5.9 - Direction qui maximise le rapport des longueurs de deux segments $\mu M_2 / \mu M_1$.

Cette dernière illustration, qui généralise l'exemple des iris, suggère que la géométrie est un outil intéressant pour explorer le champ sémantique de la statistique.

5.8 Références

- ANDERSON, T. W. (1958). An introduction to Multivariate Statistical Analysis. New York: John Wiley and Sons, Inc.
- BACHACOU, J., MASSON, J-P., MILLIER C. (1981). Manuel de la programmation mathématique AMANCE 81. I.N.R.A., Département de Biométrie.
- BARTLETT, M. S. (1951). An inverse matrix adjustment arising in discriminant analysis. *Annals of Mathematical Statistics*, 22, 107-111.
- BRYANT, P. (1984). Geometry, Statistics, Probability: Variations on a Common Theme. *The American Statistician*, 38, 1, 38-48.
- CAILLIEZ, F. ET PAGES, J-P. (1976). Introduction à l'Analyse des Données. SMASH, Paris.
- CYFFERS, B. (1965). Analyse discriminatoire. *Revue de statistique appliquée*, 13.2, 29-46 et 13.3, 39-65.
- DAGNÉLIE, P. (1969). Théorie et Méthodes Statistiques. Editions J. Duculot, S. A. - Gembloux, Vol. 1.
- DEMPSTER, A. P. (1969). Elements of Continuous Multivariate Analysis. Addison-Wesley, 388.
- FISHER, R. A. (1915). Frequency Distribution of the Values of the Correlation Coefficient in Samples From an Indefinitely Large Population. *Biometrika*, 10, 507-521.
- HERR, D. G. (1980). On the History of the Use of Geometry in the General Linear Model. *The American Statistician*, 34, 1, 43-47.
- HOMBAS, V. C. (1989). Combinations of Independent Normal Random Variables. *Teaching Statistics*, Autumn 1989, 11, 3, 74-75.
- KENDALL, M.G. AND STUART, A. (1966). The Advanced Theory of Statistics. Vol. 3, 2nd Edition, Griffin: London.
- KRUSKAL, W. H. (1975). The Geometry of Generalized Inverses. *Journal of the Royal Statistical Society, Ser. B*, 37, 272-283.
- MARGOLIS, M. S. (1979). Perpendicular Projections and Elementary Statistics. *The American Statistician*, 33, 3, 131-135.
- MOOD, A. M., GRAYBILL, F. A. AND BOES, D. C. (1974). Introduction to the theory of Statistics. 3rd Edition. New York: McGraw-Hill.

MORRISSON, D. F. (1976). *Multivariate Statistical Methods*. Second Edition. New York : McGraw-Hill.

PENROSE, R. A. (1955). A generalized inverse for matrices. *Proceedings of the Cambridge Philosophical Society*, 51, 406-413.

Index des auteurs

A

Anderson, T. W. 30, 92

B

Bachacou, J. 21, 92

Bartlett, M. S. 75, 92

Boes, D. C. 92

Bryant, P. 92

C

Cailliez, F. 40–42, 61, 92

Cyffers, B. 89, 92

D

Dagnélie, P. 65, 92

Dempster, A. P. 92

F

Fisher, R. A. 26, 27, 33, 83, 92

G

Graybill, F. A. 92

H

Herr, D. G. 34, 92

Hombas, V. C. 30, 92

K

Kendall, M. G. 83, 92

Kruskal, W. H. 33, 92

M

Margolis, M. S. 34, 92

Masson, J-P. 92

Millier, C. 92

Mood, A. M. 30, 92

Morrisson, D. F. 93

P

Pages, J-P. 40–42, 61, 92

Penrose, R. A. 52, 93

S

Stuart, A. 83, 92

Index des sujets

A

Application linéaire 44, 62, 63
 idempotente 68
Axes conjugués 62

C

Coefficient
 de corrélation multiple . 17, 18, 21
 de corrélation partielle 17, 18
 de corrélation simple 17, 18
 de détermination 21
Cofacteur 53

E

Ellipsoïde 57
Espérance 27
 conditionnelle 28, 29
Espace
 (dimension) 7
 des individus 2
 des variables 2
 dual 2
 vectoriel 1, 7
 base 10, 36

F

Forme quadratique 57
 rang 64

I

Indépendance linéaire 36

M

Métrique 24, 57
 intra-p0pulation 25

Matrice

J 50
 (addition) 51, 74
 (dimension) 49
 (déterminant) 53, 75
 (multiplication) 51, 74
 (présentation) 1, 49
 (rang) 52
 (trace) 52
adjointe 54
carrée 49
définie négative 78
définie positive 57, 78
d'inertie 66, 84, 85, 87
digonale 50
hessienne 77
idempotente 51
identité 50
inverse 51, 74
inverse généralisée 52
nulle 50
orthogonale 52
partitionnée 73
singulière 51
symétrique 50
transposée 49, 52, 73
triangulaire 50
Mineur 53
Moindres carrés 13, 20
Moindres rectangles 66
Multiplicateurs de Lagrange . . . 78, 88

P

Polynôme caractéristique 63
Produit scalaire . . 6, 9, 11, 12, 14, 17,
 22, 45, 46
Projection orthogonale . 7, 8, 14, 20–22,
 87

R

Régression	20
orthogonale	66

S

Somme directe	39
Sous espace	
affine	38
complémentaires	39
propre	63
vectoriel	15, 36, 39, 78
base	43
orthogonal	15, 17
Sous-matrices	73

V

Valeur propre	62
Vecteur	
(addition)	4
(différence)	4
(direction)	4
(extrémité)	4
(longueur)	4
(multiplication)	5
(origine)	4
(ortogonalité)	5
(sens)	4
(égalité)	4
propre	62