

Des vecteurs à la précision des coefficients d'un modèle

Comment utiliser ce fascicule

Ce fascicule reprend depuis le début les outils mathématiques dont vous aurez besoin pendant le module : vecteurs, matrices, rang, inverse, variance et covariance, méthode des moindres carrés. Aucune connaissance préalable en calcul matriciel n'est supposée.

Chaque section suit la même progression. Elle part d'un calcul concret, tiré de l'étude d'extraction utilisée pendant la séance, puis elle généralise. Les encadrés « Méthode » donnent les gestes à refaire. Les encadrés « Vérifiez votre compréhension » permettent de contrôler que vous avez compris avant de continuer ; les réponses figurent juste en dessous. Les démonstrations sont regroupées à la fin de chaque section, dans une partie « Pour aller plus loin » que vous pouvez lire dans un second temps.

Les sections 1 à 6 sont à lire avant la séance. Les sections 7 à 9 seront utilisées pendant la seconde moitié de la séance. La section 10 propose des exercices corrigés.

Table des matières

1	Un tableau d'essais et ce que l'on veut en tirer	3
2	Les vecteurs	4
2.1	Une colonne de nombres	4
2.2	Additionner des vecteurs et les multiplier par un nombre	4
2.3	Le produit scalaire	5
2.4	Deux colonnes orthogonales	5
3	Les matrices et le produit matriciel	7
3.1	Un tableau rectangulaire de nombres	7
3.2	Le produit d'une matrice par un vecteur	7
3.3	Le produit de deux matrices	8
4	La transposée et la matrice $\mathbf{X}^\top \mathbf{X}$	10
4.1	Échanger lignes et colonnes	10
4.2	La matrice $\mathbf{X}^\top \mathbf{X}$	10
5	Le rang d'une matrice	12
5.1	Six essais qui ne permettent pas de conclure	12
5.2	Colonnes dépendantes et rang	12
5.3	Calculer un rang	13
5.4	Le critère d'identifiabilité	13
6	L'inverse d'une matrice	15
6.1	L'analogue de la division	15
6.2	Deux cas simples	15
6.3	La méthode de Gauss–Jordan	16
6.4	Quand une matrice n'est-elle pas inversible ?	16
7	Variance et covariance	18
7.1	Une mesure est une variable aléatoire	18
7.2	Espérance et variance	18
7.3	La variance d'une combinaison de mesures	18
7.4	La covariance de deux coefficients	19
7.5	Estimer σ^2 à partir de répétitions	20
8	La méthode des moindres carrés	21
8.1	Le principe	21
8.2	Ajuster une droite	21
8.3	La formule générale	21

9	La précision des coefficients	23
9.1	Un résultat déjà connu, écrit autrement	23
9.2	La formule générale	23
9.3	Le rôle de l'orthogonalité	24
9.4	Le rôle de la position des points	24
9.5	Cinq notions à ne pas confondre	24
10	Exercices corrigés	26

1 Un tableau d'essais et ce que l'on veut en tirer

Un laboratoire étudie le rendement d'une extraction. Il réalise quatre essais en faisant varier deux réglages : la température T , égale à 40 °C ou à 60 °C, et la durée t , égale à 1 h ou à 2 h. Les résultats sont les suivants.

Essai	T	t	Rendement (%)
1	40 °C	1 h	23
2	60 °C	1 h	37
3	40 °C	2 h	41
4	60 °C	2 h	59

Pour faciliter les calculs, on remplace chaque réglage par un nombre sans unité, appelé **variable codée**, qui vaut -1 au niveau bas et $+1$ au niveau haut. On note x_1 la température codée et x_2 la durée codée :

$$x_1 = \frac{T - 50^\circ\text{C}}{10^\circ\text{C}}, \quad x_2 = \frac{t - 1,5\text{ h}}{0,5\text{ h}}.$$

Par exemple, 40 °C donne $x_1 = (40 - 50)/10 = -1$ et 2 h donne $x_2 = (2 - 1,5)/0,5 = +1$. Le tableau devient :

Essai	x_1	x_2	y (%)
1	-1	-1	23
2	+1	-1	37
3	-1	+1	41
4	+1	+1	59

Pour mesurer l'influence de la température, on compare la moyenne des deux essais réalisés à 60 °C (essais 2 et 4) à celle des deux essais réalisés à 40 °C (essais 1 et 3). On divise ensuite par deux, pour obtenir la variation correspondant à un pas de $+1$ sur l'échelle codée :

$$b_1 = \frac{1}{2} \left(\frac{37 + 59}{2} - \frac{23 + 41}{2} \right) = \frac{-23 + 37 - 41 + 59}{4} = \frac{32}{4} = 8.$$

Observons les signes placés devant les quatre rendements dans le numérateur : $-$, $+$, $-$, $+$. Ce sont exactement les valeurs de la colonne x_1 du tableau. Le calcul de b_1 consiste donc à multiplier chaque valeur de la colonne x_1 par le rendement de la même ligne, à additionner ces quatre produits, puis à diviser par 4.

Ce type de calcul revient sans cesse en plans d'expériences. Avec deux facteurs et quatre essais, on peut tout écrire à la main. Avec onze facteurs et douze essais, comme dans la séance 2, cela devient impossible sans une écriture plus compacte. Le calcul matriciel fournit cette écriture. Il permet aussi de répondre à des questions que le simple calcul ne pose pas : peut-on séparer l'influence de la température de celle de la durée ? avec quelle précision ? un autre choix de quatre essais aurait-il été meilleur ?

L'objectif de ce fascicule est de vous conduire, pas à pas, jusqu'aux deux formules qui résument la séance :

$$\hat{\beta} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y} \quad \text{et} \quad \text{Cov}(\hat{\beta}) = \sigma^2 (\mathbf{X}^\top \mathbf{X})^{-1}.$$

La première donne les coefficients d'un modèle, la seconde leur précision. Vous comprendrez chaque symbole et vous saurez démontrer ces deux résultats.

2 Les vecteurs

Dans cette section

Écrire une colonne de tableau sous forme de vecteur ; calculer un produit scalaire ; reconnaître deux colonnes orthogonales.

2.1 Une colonne de nombres

Dans le tableau de la section précédente, les quatre rendements forment une liste ordonnée : 23, 37, 41, 59. On l'écrit en colonne et on la désigne par une lettre grasse :

$$\mathbf{y} = \begin{pmatrix} 23 \\ 37 \\ 41 \\ 59 \end{pmatrix}.$$

De même, la colonne des températures codées s'écrit

$$\mathbf{x}_1 = \begin{pmatrix} -1 \\ +1 \\ -1 \\ +1 \end{pmatrix}.$$

Définition 2.1 : Vecteur

Un **vecteur** de dimension n est une liste ordonnée de n nombres, appelés ses composantes, écrits en colonne. On note u_i la composante de rang i du vecteur $\mathbf{u}$.

L'ordre des composantes compte : la i -ième composante de $\mathbf{y}$ et celle de $\mathbf{x}_1$ se rapportent toutes deux à l'essai i . Pour gagner de la place, on écrira parfois un vecteur en ligne en le faisant suivre du symbole $^\top$: ainsi $\mathbf{y} = (23, 37, 41, 59)^\top$. Ce symbole est expliqué à la section 4.

2.2 Additionner des vecteurs et les multiplier par un nombre

On additionne deux vecteurs de même dimension composante par composante, et l'on multiplie un vecteur par un nombre en multipliant chacune de ses composantes. Par exemple :

$$\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} + \begin{pmatrix} 4 \\ 0 \\ -1 \end{pmatrix} = \begin{pmatrix} 5 \\ 2 \\ 2 \end{pmatrix}, \quad 3 \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} = \begin{pmatrix} 3 \\ -3 \\ 0 \end{pmatrix}.$$

On ne peut pas additionner deux vecteurs de dimensions différentes.

Ces deux opérations suffisent pour écrire un modèle. Notons $\mathbf{1} = (1, 1, 1, 1)^\top$ le vecteur dont toutes les composantes valent 1 et $\mathbf{x}_2 = (-1, -1, +1, +1)^\top$ la colonne des durées codées. Le vecteur

$$40 \mathbf{1} + 8 \mathbf{x}_1 + 10 \mathbf{x}_2 = \begin{pmatrix} 40 - 8 - 10 \\ 40 + 8 - 10 \\ 40 - 8 + 10 \\ 40 + 8 + 10 \end{pmatrix} = \begin{pmatrix} 22 \\ 38 \\ 42 \\ 58 \end{pmatrix}$$

contient les rendements que prévoit le modèle $y = 40 + 8x_1 + 10x_2$ pour les quatre essais. Un tel vecteur, obtenu en additionnant des vecteurs multipliés par des nombres, s'appelle une **combinaison linéaire** de ces vecteurs.

2.3 Le produit scalaire

Le calcul de b_1 de la section 1 consistait à multiplier deux colonnes terme à terme, puis à additionner. Cette opération porte un nom.

Définition 2.2 : Produit scalaire

Le **produit scalaire** de deux vecteurs $\mathbf{u}$ et $\mathbf{v}$ de même dimension n est le nombre

$$\mathbf{u} \cdot \mathbf{v} = u_1v_1 + u_2v_2 + \cdots + u_nv_n.$$

Méthode : calculer un produit scalaire

1. Écrire les deux vecteurs côte à côte et vérifier qu'ils ont la même dimension.
2. Multiplier les composantes qui se trouvent sur la même ligne.
3. Additionner tous ces produits. Le résultat est un nombre, pas un vecteur.

Exemple résolu 2.3 : Le numérateur de b_1

Pour $\mathbf{x}_1 = (-1, 1, -1, 1)^\top$ et $\mathbf{y} = (23, 37, 41, 59)^\top$, les produits ligne par ligne valent -23 , 37 , -41 et 59 . Leur somme donne $\mathbf{x}_1 \cdot \mathbf{y} = 32$. On retrouve le numérateur du calcul de b_1 , et donc $b_1 = \mathbf{x}_1 \cdot \mathbf{y} / 4 = 8$.

Le produit scalaire d'un vecteur avec lui-même est la somme des carrés de ses composantes. Sa racine carrée est la **norme** du vecteur, notée $\|\mathbf{u}\| = \sqrt{\mathbf{u} \cdot \mathbf{u}}$. Une colonne codée qui ne contient que des -1 et des $+1$ sur N essais vérifie toujours $\|\mathbf{x}\|^2 = N$, puisque chaque carré vaut 1.

2.4 Deux colonnes orthogonales

Définition 2.4 : Orthogonalité

Deux vecteurs sont **orthogonaux** lorsque leur produit scalaire est nul.

Le mot vient de la géométrie. Dans le plan, le produit scalaire de deux flèches est nul exactement lorsqu'elles forment un angle droit. En dimension supérieure, on ne peut plus dessiner les vecteurs, mais on conserve le mot.

Pour des colonnes de -1 et de $+1$, l'orthogonalité a une signification concrète. Le produit $u_i v_i$ vaut $+1$ quand les deux colonnes ont le même signe à l'essai i et -1 quand elles ont des signes opposés. Le produit scalaire est donc nul lorsqu'il y a autant d'essais où les signes sont identiques que d'essais où ils sont opposés. Autrement dit, les deux facteurs sont « équilibrés » l'un par rapport à l'autre.

Exemple résolu 2.5 : Deux plans de quatre essais

Dans le plan du tableau précédent, $\mathbf{x}_1 = (-1, 1, -1, 1)^\top$ et $\mathbf{x}_2 = (-1, -1, 1, 1)^\top$. Les produits valent 1 , -1 , -1 , 1 , et leur somme est nulle : les deux colonnes sont orthogonales.

Considérons maintenant les trois essais $(-1, -1)$, $(+1, -1)$ et $(-1, +1)$, où l'on fait varier un seul facteur à la fois à partir du premier essai. Les colonnes sont $\mathbf{x}_1 = (-1, 1, -1)^\top$ et $\mathbf{x}_2 = (-1, -1, 1)^\top$, et leur produit scalaire vaut $1 - 1 - 1 = -1$. Ces colonnes ne sont pas orthogonales.

Nous verrons dans la section 9 que cette différence a des conséquences directes sur la précision des coefficients estimés.

Vérifiez votre compréhension

1. Calculez $\mathbf{u} \cdot \mathbf{v}$ pour $\mathbf{u} = (1, 2, -1)^\top$ et $\mathbf{v} = (3, 0, 3)^\top$.
2. On définit la colonne $\mathbf{x}_{12}$ en multipliant terme à terme $\mathbf{x}_1 = (-1, 1, -1, 1)^\top$ et $\mathbf{x}_2 = (-1, -1, 1, 1)^\top$. Écrivez-la, puis vérifiez qu'elle est orthogonale à $\mathbf{x}_1$.
3. Quelle est la norme au carré d'une colonne de huit valeurs égales à ± 1 ?

Réponses. 1. $3 + 0 - 3 = 0$: les deux vecteurs sont orthogonaux. 2. $\mathbf{x}_{12} = (1, -1, -1, 1)^\top$; $\mathbf{x}_1 \cdot \mathbf{x}_{12} = -1 - 1 + 1 + 1 = 0$. 3. 8.

Pour aller plus loin : les démonstrations

L'inégalité de Cauchy–Schwarz. Pour tous vecteurs $\mathbf{u}$ et $\mathbf{v}$ de même dimension, on a $(\mathbf{u} \cdot \mathbf{v})^2 \leq \|\mathbf{u}\|^2 \|\mathbf{v}\|^2$, avec égalité si et seulement si l'un des vecteurs est un multiple de l'autre. Ce résultat servira à la section 9.

Démonstration. Si $\mathbf{v}$ est nul, les deux membres sont nuls. Sinon, pour tout nombre λ , la quantité $\|\mathbf{u} - \lambda \mathbf{v}\|^2$ est une somme de carrés, donc positive ou nulle. En la développant, on obtient

$$0 \leq \|\mathbf{u}\|^2 - 2\lambda \mathbf{u} \cdot \mathbf{v} + \lambda^2 \|\mathbf{v}\|^2.$$

Choisissons $\lambda = \mathbf{u} \cdot \mathbf{v} / \|\mathbf{v}\|^2$. L'inégalité devient $0 \leq \|\mathbf{u}\|^2 - (\mathbf{u} \cdot \mathbf{v})^2 / \|\mathbf{v}\|^2$, et il suffit de multiplier par $\|\mathbf{v}\|^2$. L'égalité impose $\|\mathbf{u} - \lambda \mathbf{v}\| = 0$, c'est-à-dire $\mathbf{u} = \lambda \mathbf{v}$. ■

3 Les matrices et le produit matriciel

Dans cette section

Construire la matrice d'un modèle ; calculer le produit d'une matrice par un vecteur, puis le produit de deux matrices ; vérifier que des dimensions sont compatibles.

3.1 Un tableau rectangulaire de nombres

Définition 3.1 : Matrice

Une **matrice** de dimensions $n \times m$ est un tableau de nombres comportant n lignes et m colonnes. L'élément situé à la ligne i et à la colonne j est noté a_{ij} : le premier indice désigne toujours la ligne.

Un vecteur de dimension n est une matrice à une seule colonne, de dimensions $n \times 1$. Une matrice qui a autant de lignes que de colonnes est dite carrée. La matrice carrée qui porte des 1 sur sa diagonale et des 0 partout ailleurs s'appelle la **matrice identité**, notée **I**.

En plans d'expériences, on range dans une matrice, appelée **matrice du modèle** et notée **X**, tout ce qui multiplie les coefficients du modèle. Pour le modèle $y = b_0 + b_1x_1 + b_2x_2$, chaque ligne de **X** correspond à un essai et contient trois nombres : 1 (qui multiplie b_0), puis la valeur de x_1 , puis celle de x_2 . Pour les quatre essais de l'étude d'extraction :

$$\mathbf{X} = \begin{pmatrix} 1 & -1 & -1 \\ 1 & +1 & -1 \\ 1 & -1 & +1 \\ 1 & +1 & +1 \end{pmatrix}.$$

C'est une matrice 4×3 : quatre essais, trois coefficients. Si l'on ajoute au modèle le terme $b_{12}x_1x_2$, on ajoute une quatrième colonne qui contient, pour chaque essai, le produit x_1x_2 , soit $(1, -1, -1, 1)^\top$.

3.2 Le produit d'une matrice par un vecteur

Reprenons le modèle $y = 40 + 8x_1 + 10x_2$. Pour le premier essai, réalisé à 40°C pendant 1 h, on a $x_1 = -1$ et $x_2 = -1$, donc

$$y = 40 \times 1 + 8 \times (-1) + 10 \times (-1) = 22.$$

On refait le même calcul pour les trois autres essais, avec chaque fois une autre ligne du tableau. Plutôt que d'écrire quatre fois ce calcul, on range les coefficients dans le vecteur $\boldsymbol{\beta} = (40, 8, 10)^\top$, et l'on dit que le vecteur des quatre prévisions est le produit de **X** par **β** :

$$\mathbf{X}\boldsymbol{\beta} = \begin{pmatrix} 1 & -1 & -1 \\ 1 & +1 & -1 \\ 1 & -1 & +1 \\ 1 & +1 & +1 \end{pmatrix} \begin{pmatrix} 40 \\ 8 \\ 10 \end{pmatrix} = \begin{pmatrix} 40 - 8 - 10 \\ 40 + 8 - 10 \\ 40 - 8 + 10 \\ 40 + 8 + 10 \end{pmatrix} = \begin{pmatrix} 22 \\ 38 \\ 42 \\ 58 \end{pmatrix}.$$

Méthode : calculer le produit $\mathbf{X}\mathbf{v}$

1. Vérifier que **X** a autant de colonnes que **v** a de composantes. Sinon, le produit n'existe pas.
2. Pour chaque ligne de **X**, multiplier ses éléments un à un par les composantes de **v**, puis additionner : c'est le produit scalaire de cette ligne et de **v**.
3. Le résultat est un vecteur qui a autant de composantes que **X** a de lignes.

On peut aussi lire ce produit colonne par colonne. Le résultat est égal à 40 fois la première colonne de $\mathbf{X}$, plus 8 fois la deuxième, plus 10 fois la troisième : c'est la combinaison linéaire des colonnes de $\mathbf{X}$ calculée à la section précédente. Cette seconde lecture sera utile pour comprendre le rang (section 5).

Vérifiez votre compréhension

1. Avec la matrice $\mathbf{X}$ ci-dessus et $\beta = (39, 7, 9)^\top$, calculez la deuxième composante de $\mathbf{X}\beta$.
2. Peut-on calculer le produit d'une matrice 4×3 par un vecteur de dimension 4 ?

Réponses. 1. $39 + 7 - 9 = 37$. 2. Non : la matrice a 3 colonnes et le vecteur 4 composantes.

3.3 Le produit de deux matrices

Le produit de deux matrices généralise le calcul précédent : on fait le produit de la première matrice par chacune des colonnes de la seconde.

Définition 3.2 : Produit matriciel

Soit $\mathbf{A}$ une matrice $n \times k$ et $\mathbf{B}$ une matrice $k \times m$. Le produit $\mathbf{C} = \mathbf{AB}$ est la matrice $n \times m$ dont l'élément c_{ij} est le produit scalaire de la ligne i de $\mathbf{A}$ et de la colonne j de $\mathbf{B}$:

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{ik}b_{kj}.$$

Le produit n'existe que si le nombre de colonnes de $\mathbf{A}$ est égal au nombre de lignes de $\mathbf{B}$. Une façon simple de le vérifier consiste à écrire les dimensions côte à côte : dans $(n \times k)(k \times m)$, les deux nombres du milieu doivent être égaux, et ceux des extrémités donnent la taille du résultat, $n \times m$.

Méthode : poser un produit de matrices à la main

1. Écrire les dimensions et vérifier qu'elles sont compatibles.
2. Dessiner le tableau vide du résultat.
3. Pour remplir la case située à la ligne i et à la colonne j , parcourir la ligne i de la première matrice et la colonne j de la seconde, multiplier les nombres deux à deux et additionner.
4. Contrôler une case au hasard en refaisant le calcul.

Exemple résolu 3.3 : Un produit de deux matrices 2×2

Calculons $\mathbf{AB}$ avec $\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$ et $\mathbf{B} = \begin{pmatrix} 5 & 6 \\ 7 & 8 \end{pmatrix}$. Les dimensions $(2 \times 2)(2 \times 2)$ sont compatibles et le résultat est une matrice 2×2 .

- Case (1, 1) : ligne 1 de $\mathbf{A}$ et colonne 1 de $\mathbf{B}$, soit $1 \times 5 + 2 \times 7 = 19$.
- Case (1, 2) : ligne 1 et colonne 2, soit $1 \times 6 + 2 \times 8 = 22$.
- Case (2, 1) : ligne 2 et colonne 1, soit $3 \times 5 + 4 \times 7 = 43$.
- Case (2, 2) : ligne 2 et colonne 2, soit $3 \times 6 + 4 \times 8 = 50$.

On obtient $\mathbf{AB} = \begin{pmatrix} 19 & 22 \\ 43 & 50 \end{pmatrix}$. Si l'on calcule $\mathbf{BA}$ de la même façon, on trouve $\begin{pmatrix} 23 & 34 \\ 31 & 46 \end{pmatrix}$, qui est différent.

Cet exemple montre que l'ordre des facteurs compte : en général $\mathbf{AB}$ et $\mathbf{BA}$ sont différents, et souvent l'un des deux produits n'existe même pas. En revanche, les autres règles de calcul habituelles restent vraies : on peut regrouper les produits comme on veut, $(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$, on peut développer, $\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$, et la matrice identité ne change rien, $\mathbf{IA} = \mathbf{AI} = \mathbf{A}$.

Vérifiez votre compréhension

1. Quelles sont les dimensions de $\mathbf{AB}$ si $\mathbf{A}$ est 3×2 et $\mathbf{B}$ est 2×4 ?

2. Calculez $\begin{pmatrix} 2 & -1 & 0 \\ 1 & 3 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 2 & 1 \\ -1 & 4 \end{pmatrix}$.

Réponses. 1. 3×4 . 2. $\begin{pmatrix} 0 & -1 \\ 6 & 7 \end{pmatrix}$; par exemple, la case $(2, 1)$ vaut $1 \times 1 + 3 \times 2 + 1 \times (-1) = 6$.

Pour aller plus loin : les démonstrations

Associativité. L'élément (i, j) de $(\mathbf{AB})\mathbf{C}$ vaut $\sum_l (\sum_k a_{ik}b_{kl})c_{lj}$, et celui de $\mathbf{A}(\mathbf{BC})$ vaut $\sum_k a_{ik}(\sum_l b_{kl}c_{lj})$. Ces deux sommes finies contiennent exactement les mêmes termes $a_{ik}b_{kl}c_{lj}$, dans un ordre différent; elles sont donc égales. La distributivité se vérifie de la même manière, élément par élément. ■

4 La transposée et la matrice $\mathbf{X}^\top \mathbf{X}$

Dans cette section

Transposer une matrice ; calculer $\mathbf{X}^\top \mathbf{X}$ comme un tableau de produits scalaires de colonnes ; lire sur $\mathbf{X}^\top \mathbf{X}$ si les colonnes de $\mathbf{X}$ sont orthogonales.

4.1 Échanger lignes et colonnes

Définition 4.1 : Transposée

La **transposée** d'une matrice $\mathbf{A}$ de dimensions $n \times m$ est la matrice $\mathbf{A}^\top$ de dimensions $m \times n$ obtenue en écrivant les lignes de $\mathbf{A}$ en colonnes.

Par exemple,

$$\mathbf{A} = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \quad \text{a pour transposée} \quad \mathbf{A}^\top = \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix}.$$

La transposée d'un vecteur colonne est un vecteur ligne. C'est ce qui justifie la notation $\mathbf{y} = (23, 37, 41, 59)^\top$: on écrit la liste en ligne, puis on la transpose pour obtenir une colonne. Elle permet aussi d'écrire le produit scalaire comme un produit de matrices, $\mathbf{u} \cdot \mathbf{v} = \mathbf{u}^\top \mathbf{v}$: une matrice $1 \times n$ multipliée par une matrice $n \times 1$ donne une matrice 1×1 , c'est-à-dire un nombre.

On retiendra une règle qui sert dans les démonstrations : la transposée d'un produit est le produit des transposées *dans l'ordre inverse*, $(\mathbf{AB})^\top = \mathbf{B}^\top \mathbf{A}^\top$.

4.2 La matrice $\mathbf{X}^\top \mathbf{X}$

En plans d'expériences, on calcule très souvent le produit de la transposée de $\mathbf{X}$ par $\mathbf{X}$ elle-même. Si $\mathbf{X}$ a N lignes et p colonnes, $\mathbf{X}^\top$ est une matrice $p \times N$, et le produit $\mathbf{X}^\top \mathbf{X}$ est une matrice carrée $p \times p$.

Ce produit se calcule plus simplement qu'il n'y paraît. La ligne j de $\mathbf{X}^\top$ n'est autre que la colonne j de $\mathbf{X}$. L'élément (j, k) de $\mathbf{X}^\top \mathbf{X}$ est donc le produit scalaire de la colonne j et de la colonne k de $\mathbf{X}$.

Méthode : calculer $\mathbf{X}^\top \mathbf{X}$

1. Nommer les colonnes de $\mathbf{X}$: pour le modèle $b_0 + b_1 x_1 + b_2 x_2$, ce sont $\mathbf{1}$, $\mathbf{x}_1$ et $\mathbf{x}_2$.
2. Remplir un tableau carré dont la case (j, k) contient le produit scalaire de la colonne j et de la colonne k .
3. Il suffit de calculer la diagonale et les cases situées au-dessus : la case (k, j) est égale à la case (j, k) , car le produit scalaire ne dépend pas de l'ordre des vecteurs. La matrice $\mathbf{X}^\top \mathbf{X}$ est donc **symétrique**.

Exemple résolu 4.2 : Le plan à quatre essais

Les colonnes sont $\mathbf{1} = (1, 1, 1, 1)^\top$, $\mathbf{x}_1 = (-1, 1, -1, 1)^\top$ et $\mathbf{x}_2 = (-1, -1, 1, 1)^\top$. Sur la diagonale, on trouve $\mathbf{1} \cdot \mathbf{1} = 4$, $\mathbf{x}_1 \cdot \mathbf{x}_1 = 4$ et $\mathbf{x}_2 \cdot \mathbf{x}_2 = 4$. Hors de la diagonale, $\mathbf{1} \cdot \mathbf{x}_1 = -1 + 1 - 1 + 1 = 0$, $\mathbf{1} \cdot \mathbf{x}_2 = 0$ et $\mathbf{x}_1 \cdot \mathbf{x}_2 = 1 - 1 - 1 + 1 = 0$. Ainsi

$$\mathbf{X}^\top \mathbf{X} = \begin{pmatrix} 4 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 4 \end{pmatrix}.$$

Exemple résolu 4.3 : Le plan « un facteur à la fois » à trois essais

Les essais sont $(-1, -1)$, $(+1, -1)$ et $(-1, +1)$. Les colonnes sont $\mathbf{1} = (1, 1, 1)^\top$, $\mathbf{x}_1 = (-1, 1, -1)^\top$ et $\mathbf{x}_2 = (-1, -1, 1)^\top$. La diagonale contient trois fois le nombre 3. Hors de la diagonale, $\mathbf{1} \cdot \mathbf{x}_1 = -1$, $\mathbf{1} \cdot \mathbf{x}_2 = -1$ et $\mathbf{x}_1 \cdot \mathbf{x}_2 = 1 - 1 - 1 = -1$. Ainsi

$$\mathbf{X}^\top \mathbf{X} = \begin{pmatrix} 3 & -1 & -1 \\ -1 & 3 & -1 \\ -1 & -1 & 3 \end{pmatrix}.$$

La comparaison de ces deux exemples est instructive. Dans le premier cas, tous les termes situés hors de la diagonale sont nuls : les colonnes de $\mathbf{X}$ sont deux à deux orthogonales. Dans le second, ce n'est pas le cas. De façon générale, *les colonnes de $\mathbf{X}$ sont deux à deux orthogonales si et seulement si $\mathbf{X}^\top \mathbf{X}$ est diagonale*, puisque les termes hors de la diagonale sont précisément les produits scalaires de deux colonnes différentes.

On utilisera aussi le vecteur $\mathbf{X}^\top \mathbf{y}$. Sa composante j est le produit scalaire de la colonne j de $\mathbf{X}$ et du vecteur des réponses. Pour le plan à quatre essais et $\mathbf{y} = (23, 37, 41, 59)^\top$, on obtient $\mathbf{X}^\top \mathbf{y} = (160, 32, 40)^\top$.

Vérifiez votre compréhension

1. Quelles sont les dimensions de $\mathbf{X}^\top \mathbf{X}$ si $\mathbf{X}$ est une matrice 12×5 ?
2. On réalise deux fois chacun des trois essais « un facteur à la fois ». Que devient $\mathbf{X}^\top \mathbf{X}$?
3. Pour le plan à quatre essais et le modèle avec le terme $x_1 x_2$, que vaut $\mathbf{X}^\top \mathbf{X}$?

Réponses. 1. 5×5 . 2. Chaque produit scalaire est une somme sur les essais ; en doublant tous les essais, toutes les cases doublent : la diagonale vaut 6 et les autres cases -2 . 3. La matrice $4\mathbf{I}$, car les quatre colonnes sont deux à deux orthogonales.

Pour aller plus loin : les démonstrations

Transposée d'un produit. Soit $\mathbf{A}$ de dimensions $n \times k$ et $\mathbf{B}$ de dimensions $k \times m$. L'élément (i, j) de $(\mathbf{AB})^\top$ est l'élément (j, i) de $\mathbf{AB}$, c'est-à-dire $\sum_l a_{jl} b_{li}$. L'élément (i, j) de $\mathbf{B}^\top \mathbf{A}^\top$ vaut $\sum_l (\mathbf{B}^\top)_{il} (\mathbf{A}^\top)_{lj} = \sum_l b_{li} a_{jl}$. Les deux sommes sont égales. ■

Symétrie de $\mathbf{X}^\top \mathbf{X}$. D'après la règle précédente, $(\mathbf{X}^\top \mathbf{X})^\top = \mathbf{X}^\top (\mathbf{X}^\top)^\top = \mathbf{X}^\top \mathbf{X}$. ■

5 Le rang d'une matrice

Dans cette section

Reconnaître des colonnes dépendantes ; calculer un rang par élimination ; savoir à quelle condition les coefficients d'un modèle sont déterminés de façon unique.

5.1 Six essais qui ne permettent pas de conclure

Un journal de production contient six extractions. Chaque fois que les opérateurs ont augmenté la température, ils ont aussi allongé la durée, si bien que $x_1 = x_2$ pour chaque lot. Pour le modèle $y = b_0 + b_1x_1 + b_2x_2$, la matrice du modèle est

$$\mathbf{X} = \begin{pmatrix} 1 & -1 & -1 \\ 1 & -0,5 & -0,5 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0,5 & 0,5 \\ 1 & 1 & 1 \end{pmatrix}.$$

Ses deux dernières colonnes sont identiques. Considérons trois jeux de coefficients : $(40, 18, 0)$, $(40, 0, 18)$ et $(40, 9, 9)$. Pour chaque lot, ils donnent exactement la même prévision, puisque seule la somme $b_1 + b_2$ intervient : $b_0 + b_1x_1 + b_2x_2 = b_0 + (b_1 + b_2)x_1$. Aucune mesure, si précise soit-elle, ne permet donc de savoir si l'effet observé vient de la température, de la durée, ou des deux. On dit que les coefficients b_1 et b_2 ne sont pas **identifiables** avec ce plan.

Pourtant, le journal compte six essais pour trois coefficients. Le nombre d'essais ne suffit donc pas : il faut regarder comment les colonnes de $\mathbf{X}$ dépendent les unes des autres.

5.2 Colonnes dépendantes et rang

Définition 5.1 : Colonnes linéairement dépendantes

Des colonnes sont **linéairement dépendantes** si l'une d'elles peut s'écrire comme une combinaison linéaire des autres. Dans le cas contraire, elles sont dites indépendantes.

Dans le journal, la troisième colonne est égale à la deuxième : les colonnes sont dépendantes. Une dépendance n'est pas toujours aussi visible. Par exemple, une colonne égale à la somme de deux autres crée aussi une dépendance.

Définition 5.2 : Rang

Le **rang** d'une matrice, noté $\text{rang}(\mathbf{X})$, est le nombre maximal de colonnes indépendantes que l'on peut choisir parmi ses colonnes.

Le rang de la matrice du journal vaut 2 : les deux premières colonnes sont indépendantes, mais la troisième n'apporte rien de nouveau. Deux propriétés sont utiles en pratique. D'une part, le rang ne peut dépasser ni le nombre de lignes ni le nombre de colonnes. D'autre part, **répéter un essai ne change pas le rang** : une ligne recopiée n'apporte aucune direction nouvelle. Une répétition améliore la précision, comme on le verra plus loin, mais elle n'aide pas à séparer des effets confondus.

5.3 Calculer un rang

Pour des matrices plus grandes, on ne voit pas toujours les dépendances à l'œil nu. On utilise alors la méthode d'élimination de Gauss, qui transforme la matrice sans changer son rang.

Méthode : calculer un rang par élimination

1. On travaille colonne par colonne, de gauche à droite.
2. Dans la colonne traitée, on choisit une ligne dont l'élément n'est pas nul : c'est le **pivot**.
3. On retranche aux lignes situées en dessous un multiple de la ligne du pivot, choisi pour annuler leur élément dans cette colonne.
4. On passe à la colonne suivante en ignorant les lignes qui ont déjà fourni un pivot. Si une colonne ne contient plus que des zéros, on la saute.
5. Le rang est égal au nombre de pivots trouvés.

Exemple résolu 5.3 : Le rang de la matrice du journal

On note $L_1, \dots, L_6$ les lignes.

- Colonne 1 : le pivot est le 1 de L_1 . On retranche L_1 à chacune des autres lignes. La deuxième ligne devient $(0; 0,5; 0,5)$, la troisième $(0; 1; 1)$, et ainsi de suite jusqu'à la sixième, $(0; 2; 2)$.
- Colonne 2 : le pivot est le 0,5 de L_2 . On retranche à L_3 deux fois L_2 , à L_4 deux fois L_2 , à L_5 trois fois L_2 et à L_6 quatre fois L_2 . Toutes ces lignes deviennent nulles.
- Colonne 3 : sous L_2 , il ne reste que des zéros. Il n'y a pas de pivot.

On a trouvé deux pivots : $\text{rang}(\mathbf{X}) = 2$, alors que le modèle compte trois coefficients.

Exemple résolu 5.4 : Le plan « un facteur à la fois »

Pour $\mathbf{X} = \begin{pmatrix} 1 & -1 & -1 \\ 1 & 1 & -1 \\ 1 & -1 & 1 \end{pmatrix}$, on retranche L_1 à L_2 et à L_3 . On obtient $(0, 2, 0)$ et $(0, 0, 2)$. Les pivots sont 1, 2 et 2 : le rang vaut 3. Avec seulement trois essais, le modèle à trois coefficients est donc identifiable.

5.4 Le critère d'identifiabilité

Revenons à la question de départ. Deux jeux de coefficients β et γ donnent les mêmes prévisions lorsque $\mathbf{X}\beta = \mathbf{X}\gamma$. On peut montrer que cela n'arrive que pour $\beta = \gamma$ exactement lorsque les colonnes de $\mathbf{X}$ sont indépendantes, c'est-à-dire lorsque le rang de $\mathbf{X}$ est égal au nombre de colonnes.

Théorème 5.5 : Identifiabilité

Soit $\mathbf{X}$ la matrice d'un modèle à p coefficients. Les coefficients sont identifiables, c'est-à-dire déterminés de façon unique par les prévisions du modèle, si et seulement si $\text{rang}(\mathbf{X}) = p$. Il faut pour cela au moins p essais distincts, mais cette condition ne suffit pas.

Il faut bien noter que l'identifiabilité concerne un *modèle* associé à un *plan*, jamais un plan seul. Les trois essais « un facteur à la fois » identifient le modèle à trois coefficients, mais pas le modèle qui ajoute l'interaction $b_{12}x_1x_2$: quatre coefficients ne peuvent pas être déterminés par trois conditions expérimentales distinctes, et répéter ces trois essais n'y changerait rien.

Vérifiez votre compréhension

1. Quel est le rang de $\begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 1 & 0 & 1 \end{pmatrix}$? Quelle relation lie ses colonnes ?
2. On ajoute au journal un septième essai au point $(1, -1)$, c'est-à-dire à 60°C pendant 1 h. Le modèle devient-il identifiable ?

Réponses. 1. Le rang vaut 2 : la troisième colonne est la somme des deux premières. 2. Oui : les lignes $(1, -1, -1)$, $(1, 1, 1)$ et $(1, 1, -1)$ donnent trois pivots, donc le rang vaut 3. Un seul essai bien placé suffit là où six essais mal placés échouaient.

Pour aller plus loin : les démonstrations

Démonstration du critère. Par distributivité, $\mathbf{X}\boldsymbol{\beta} = \mathbf{X}\boldsymbol{\gamma}$ équivaut à $\mathbf{X}(\boldsymbol{\beta} - \boldsymbol{\gamma}) = \mathbf{0}$. Or, pour tout vecteur $\mathbf{v} = (v_0, \dots, v_{p-1})^\top$, le produit $\mathbf{X}\mathbf{v}$ est la combinaison linéaire $v_0\mathbf{x}_0 + \dots + v_{p-1}\mathbf{x}_{p-1}$ des colonnes. Si les colonnes sont indépendantes, seule la combinaison à coefficients tous nuls donne le vecteur nul, donc $\boldsymbol{\beta} = \boldsymbol{\gamma}$. Si elles sont dépendantes, il existe un $\mathbf{v}$ non nul tel que $\mathbf{X}\mathbf{v} = \mathbf{0}$, et $\boldsymbol{\beta}$ et $\boldsymbol{\beta} + \mathbf{v}$ donnent les mêmes prévisions. Dans le journal, $\mathbf{v} = (0, -1, 1)^\top$ convient. ■

Rang de $\mathbf{X}^\top\mathbf{X}$. La matrice $\mathbf{X}^\top\mathbf{X}$ a le même rang que $\mathbf{X}$. En effet, si $\mathbf{X}^\top\mathbf{X}\mathbf{v} = \mathbf{0}$, alors $0 = \mathbf{v}^\top\mathbf{X}^\top\mathbf{X}\mathbf{v} = (\mathbf{X}\mathbf{v})^\top(\mathbf{X}\mathbf{v}) = \|\mathbf{X}\mathbf{v}\|^2$, donc $\mathbf{X}\mathbf{v} = \mathbf{0}$; la réciproque est immédiate. Les deux matrices ont donc les mêmes relations de dépendance entre colonnes. ■

6 L'inverse d'une matrice

Dans cette section

Savoir ce qu'est l'inverse d'une matrice carrée; inverser une matrice 2×2 et une matrice diagonale; inverser une matrice 3×3 par la méthode de Gauss–Jordan; vérifier un inverse.

6.1 L'analogie de la division

Pour un nombre a non nul, diviser par a revient à multiplier par $1/a$, le nombre qui vérifie $a \times (1/a) = 1$. Il n'existe pas de division entre matrices, mais on peut définir l'équivalent de $1/a$.

Définition 6.1 : Matrice inverse

Une matrice carrée $\mathbf{A}$ est **inversible** s'il existe une matrice $\mathbf{A}^{-1}$, de même taille, telle que $\mathbf{A}^{-1}\mathbf{A} = \mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$. Cette matrice, lorsqu'elle existe, est unique : c'est l'**inverse** de $\mathbf{A}$.

Seule une matrice carrée peut avoir un inverse. La matrice du modèle $\mathbf{X}$ est en général rectangulaire, puisqu'elle compte plus d'essais que de coefficients; c'est donc la matrice carrée $\mathbf{X}^\top \mathbf{X}$ que l'on inversera.

Pour vérifier qu'une matrice $\mathbf{B}$ est bien l'inverse de $\mathbf{A}$, il suffit de calculer le produit $\mathbf{AB}$ et de constater que l'on obtient la matrice identité.

6.2 Deux cas simples

Les matrices diagonales. Une matrice diagonale dont tous les termes diagonaux sont non nuls a pour inverse la matrice diagonale formée des inverses de ces termes. Par exemple, l'inverse de $\mathbf{X}^\top \mathbf{X} = 4\mathbf{I}$, obtenue pour le plan à quatre essais, est $\frac{1}{4}\mathbf{I}$. C'est la situation la plus simple que l'on puisse rencontrer, et elle correspond à des colonnes de $\mathbf{X}$ orthogonales.

Les matrices 2×2 . Pour $\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, on calcule d'abord le nombre $ad - bc$, appelé **déterminant**. Si ce nombre n'est pas nul, la matrice est inversible et

$$\mathbf{A}^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

On échange donc les deux termes de la diagonale, on change le signe des deux autres, et l'on divise par le déterminant. Si le déterminant est nul, la matrice n'est pas inversible.

Exemple résolu 6.2 : Une matrice 2×2

Pour $\mathbf{A} = \begin{pmatrix} 4 & 2 \\ 2 & 2 \end{pmatrix}$, le déterminant vaut $4 \times 2 - 2 \times 2 = 4$. Donc

$$\mathbf{A}^{-1} = \frac{1}{4} \begin{pmatrix} 2 & -2 \\ -2 & 4 \end{pmatrix} = \begin{pmatrix} 0,5 & -0,5 \\ -0,5 & 1 \end{pmatrix}.$$

Vérification : la première ligne de $\mathbf{AA}^{-1}$ vaut $(4 \times 0,5 - 2 \times 0,5; -4 \times 0,5 + 2 \times 1) = (1; 0)$, et la seconde vaut $(0; 1)$.

6.3 La méthode de Gauss–Jordan

Pour une matrice plus grande, on écrit côte à côte la matrice à inverser et la matrice identité. On effectue ensuite des opérations sur les lignes, toujours sur les deux blocs à la fois, jusqu'à ce que le bloc de gauche devienne l'identité. Le bloc de droite est alors l'inverse cherché.

Méthode : inverser par Gauss–Jordan

1. Écrire le tableau $[\mathbf{A} \mid \mathbf{I}]$.
2. Utiliser trois opérations sur les lignes : échanger deux lignes ; multiplier une ligne par un nombre non nul ; ajouter à une ligne un multiple d'une autre ligne.
3. Faire apparaître des zéros sous la diagonale, puis au-dessus, puis des 1 sur la diagonale.
4. Lire l'inverse dans le bloc de droite, puis le vérifier par un produit.

Exemple résolu 6.3 : Inverser $\mathbf{X}^\top \mathbf{X}$ du plan « un facteur à la fois »

On part de

$$\left(\begin{array}{ccc|ccc} 3 & -1 & -1 & 1 & 0 & 0 \\ -1 & 3 & -1 & 0 & 1 & 0 \\ -1 & -1 & 3 & 0 & 0 & 1 \end{array} \right).$$

- On remplace L_1 par $L_1 + L_2 + L_3$. Comme chaque colonne de gauche a pour somme 1, on obtient $L_1 = (1, 1, 1 \mid 1, 1, 1)$.
- On ajoute la nouvelle L_1 à L_2 et à L_3 : $L_2 = (0, 4, 0 \mid 1, 2, 1)$ et $L_3 = (0, 0, 4 \mid 1, 1, 2)$.
- On divise L_2 et L_3 par 4 : $L_2 = (0, 1, 0 \mid \frac{1}{4}, \frac{1}{2}, \frac{1}{4})$ et $L_3 = (0, 0, 1 \mid \frac{1}{4}, \frac{1}{4}, \frac{1}{2})$.
- On remplace L_1 par $L_1 - L_2 - L_3$: $L_1 = (1, 0, 0 \mid \frac{1}{2}, \frac{1}{4}, \frac{1}{4})$.

Le bloc de gauche est devenu l'identité. L'inverse est

$$(\mathbf{X}^\top \mathbf{X})^{-1} = \begin{pmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{pmatrix}.$$

Vérification de la première ligne du produit : $3 \times \frac{1}{2} - \frac{1}{4} - \frac{1}{4} = 1$, puis $3 \times \frac{1}{4} - \frac{1}{2} - \frac{1}{4} = 0$, puis $3 \times \frac{1}{4} - \frac{1}{4} - \frac{1}{2} = 0$.

6.4 Quand une matrice n'est-elle pas inversible ?

La méthode de Gauss–Jordan échoue lorsqu'on ne trouve pas un pivot dans chaque colonne, c'est-à-dire lorsque le rang de la matrice est inférieur à sa taille. Une matrice carrée est donc inversible si et seulement si son rang est égal à son nombre de colonnes. Comme $\mathbf{X}^\top \mathbf{X}$ a le même rang que $\mathbf{X}$ (section 5), on obtient un résultat très utile : $\mathbf{X}^\top \mathbf{X}$ est inversible exactement lorsque le modèle est identifiable. Pour le journal de production, $\mathbf{X}^\top \mathbf{X}$ a deux colonnes égales et ne peut pas être inversée.

Vérifiez votre compréhension

1. Inversez $\begin{pmatrix} 3 & 7 \\ 7 & 21 \end{pmatrix}$.
2. Quel est l'inverse de $\begin{pmatrix} 6 & -2 & -2 \\ -2 & 6 & -2 \\ -2 & -2 & 6 \end{pmatrix}$? (Remarquez que c'est le double de la matrice de l'exemple.)

Réponses. 1. Le déterminant vaut $63 - 49 = 14$, et l'inverse est $\frac{1}{14} \begin{pmatrix} 21 & -7 \\ -7 & 3 \end{pmatrix}$. 2. Si l'on double une matrice, son inverse est divisé par deux : on obtient la matrice de l'exemple divisée par 2, c'est-à-dire $\frac{1}{4}$ sur la diagonale et $\frac{1}{8}$ ailleurs.

Pour aller plus loin : les démonstrations

Unicité de l'inverse. Si $\mathbf{B}$ et $\mathbf{C}$ vérifient toutes deux la définition, alors $\mathbf{B} = \mathbf{B}(\mathbf{A}\mathbf{C}) = (\mathbf{B}\mathbf{A})\mathbf{C} = \mathbf{C}$. ■

Formule 2×2 . On vérifie directement que $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} = (ad - bc)\mathbf{I}$. Si $ad - bc \neq 0$, il suffit de diviser par ce nombre. Si $ad - bc = 0$, les deux colonnes sont proportionnelles, il existe $\mathbf{v} \neq \mathbf{0}$ tel que $\mathbf{A}\mathbf{v} = \mathbf{0}$, et l'on ne peut pas avoir $\mathbf{v} = \mathbf{A}^{-1}\mathbf{A}\mathbf{v} = \mathbf{0}$: la matrice n'est pas inversible. ■

Inverse d'une matrice symétrique. En transposant $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$, on obtient $(\mathbf{A}^{-1})^\top \mathbf{A}^\top = \mathbf{I}$, donc $(\mathbf{A}^{-1})^\top$ est l'inverse de $\mathbf{A}^\top$. Si $\mathbf{A}$ est symétrique, cela donne $(\mathbf{A}^{-1})^\top = \mathbf{A}^{-1}$: l'inverse est symétrique lui aussi. C'est le cas de $(\mathbf{X}^\top \mathbf{X})^{-1}$. ■

7 Variance et covariance

Dans cette section

Décrire la dispersion d'une mesure par sa variance ; calculer la variance d'une combinaison de mesures ; comprendre d'où viennent les facteurs $\frac{1}{2}$, $\frac{1}{4}$ et $\frac{1}{N}$; savoir ce que signifie la covariance de deux coefficients.

7.1 Une mesure est une variable aléatoire

Si l'on refait deux fois le même essai d'extraction, on n'obtient jamais exactement le même rendement : la pesée, le prélèvement, l'intégration du pic chromatographique introduisent de petites erreurs. On décrit donc la mesure de l'essai i comme la somme d'une valeur vraie, inconnue, et d'une erreur aléatoire :

$$Y_i = \mu_i + \varepsilon_i.$$

Avant d'être réalisée, la mesure Y_i est une **variable aléatoire** : on ne connaît pas sa valeur, mais on peut décrire comment elle se répartit. Après la mesure, on dispose d'un nombre, noté y_i .

Les coefficients d'un modèle sont calculés à partir des mesures. Ils sont donc eux aussi aléatoires : une autre campagne d'essais aurait donné d'autres valeurs. Comparer la précision de deux plans revient à comparer la dispersion de leurs coefficients.

7.2 Espérance et variance

L'**espérance** $E(Y)$ d'une variable aléatoire est sa valeur moyenne, celle que l'on obtiendrait en répétant la mesure un très grand nombre de fois. La **variance** $\text{Var}(Y)$ mesure la dispersion autour de cette moyenne ; c'est la moyenne des carrés des écarts :

$$\text{Var}(Y) = E[(Y - EY)^2].$$

Sa racine carrée est l'**écart-type**, qui s'exprime dans la même unité que la mesure. Dans ce module, on note σ^2 la variance d'une mesure ; σ est donc l'écart-type de répétabilité.

Deux règles de calcul suffisent pour la suite.

- L'espérance se calcule « comme une moyenne » : $E(aY + bZ) = aE(Y) + bE(Z)$ pour tous nombres a et b .
- Multiplier une mesure par un nombre a multiplie sa variance par a^2 : $\text{Var}(aY) = a^2 \text{Var}(Y)$. Ajouter une constante ne change pas la variance.

Par exemple, si un rendement a un écart-type de 1,5 point, sa moitié a un écart-type de 0,75 point, et sa variance passe de 2,25 à $2,25/4 \approx 0,56$.

7.3 La variance d'une combinaison de mesures

Tous les coefficients calculés dans ce cours sont des combinaisons de la forme $L = a_1Y_1 + a_2Y_2 + \dots + a_NY_N$. Le résultat suivant est le plus utilisé du module.

Théorème 7.1 : Variance d'une combinaison

On suppose que les mesures ont toutes la même variance σ^2 et que les erreurs de deux essais différents n'ont aucun lien entre elles. Alors

$$\text{Var}(a_1Y_1 + \dots + a_NY_N) = \sigma^2 (a_1^2 + a_2^2 + \dots + a_N^2).$$

Autrement dit, on additionne les carrés des poids et l'on multiplie par σ^2 . Les signes des poids n'interviennent pas, puisqu'ils sont élevés au carré.

Exemple résolu 7.2 : D'où viennent $\sigma^2/2$, $\sigma^2/4$ et σ^2/N

- *Stratégie « un facteur à la fois »*. L'influence de la température est estimée par $b_1 = \frac{y_2 - y_1}{2} = -\frac{1}{2}y_1 + \frac{1}{2}y_2$. Les poids sont $-\frac{1}{2}$ et $\frac{1}{2}$, leurs carrés valent $\frac{1}{4}$, et donc $\text{Var}(b_1) = \sigma^2\left(\frac{1}{4} + \frac{1}{4}\right) = \frac{\sigma^2}{2}$.
- *Plan à quatre essais*. On a $b_1 = \frac{-y_1 + y_2 - y_3 + y_4}{4}$. Les quatre poids valent $\pm\frac{1}{4}$, leurs carrés $\frac{1}{16}$, et donc $\text{Var}(b_1) = \sigma^2 \times 4 \times \frac{1}{16} = \frac{\sigma^2}{4}$.
- *Plan à N essais dont la colonne ne contient que des ± 1 , équilibrée*. Les N poids valent $\pm\frac{1}{N}$, d'où $\text{Var}(b_1) = \sigma^2 \times N \times \frac{1}{N^2} = \frac{\sigma^2}{N}$.

Le facteur $\frac{1}{2}$ de la stratégie « un facteur à la fois » ne vient pas d'un défaut particulier : il vient de ce que le coefficient y est calculé avec deux mesures seulement, alors que dans le plan à quatre essais chaque mesure sert au calcul de chaque coefficient.

Il faut distinguer le coefficient b_1 de ce que l'on appelle souvent l'**effet total** du facteur, $\Delta_1 = \bar{y}_+ - \bar{y}_-$, c'est-à-dire la différence entre la moyenne des essais au niveau haut et celle des essais au niveau bas. On a $\Delta_1 = 2b_1$, et donc $\text{Var}(\Delta_1) = 4\text{Var}(b_1)$. Dans le plan à quatre essais, $\text{Var}(\Delta_1) = \sigma^2$. Lorsqu'on compare des valeurs ou des précisions, il faut toujours savoir laquelle de ces deux grandeurs est utilisée.

7.4 La covariance de deux coefficients

Deux coefficients calculés à partir des mêmes mesures ne varient pas indépendamment l'un de l'autre. La **covariance** mesure ce lien :

$$\text{Cov}(Y, Z) = E[(Y - EY)(Z - EZ)].$$

Elle est positive si les deux grandeurs ont tendance à s'écarter de leur moyenne dans le même sens, négative dans le cas contraire. On la ramène souvent à un nombre compris entre -1 et 1 , le **coefficient de corrélation** $\text{Corr}(Y, Z) = \text{Cov}(Y, Z) / \sqrt{\text{Var}(Y) \text{Var}(Z)}$.

Sous les mêmes hypothèses que le théorème précédent, la covariance de deux combinaisons $L = \sum a_i Y_i$ et $M = \sum c_i Y_i$ vaut

$$\text{Cov}(L, M) = \sigma^2 (a_1 c_1 + a_2 c_2 + \dots + a_N c_N).$$

Elle est donc nulle lorsque les vecteurs des poids sont orthogonaux.

Exemple résolu 7.3 : Deux coefficients qui partagent une mesure

Dans la stratégie « un facteur à la fois », $b_1 = \frac{y_2 - y_1}{2}$ et $b_2 = \frac{y_3 - y_1}{2}$. Les poids sont $(-\frac{1}{2}, \frac{1}{2}, 0)$ et $(-\frac{1}{2}, 0, \frac{1}{2})$. Le seul produit non nul est celui de la première mesure, $\frac{1}{4}$. Donc $\text{Cov}(b_1, b_2) = \frac{\sigma^2}{4}$ et

$$\text{Corr}(b_1, b_2) = \frac{\sigma^2/4}{\sqrt{(\sigma^2/2) \times (\sigma^2/2)}} = \frac{1}{2}.$$

Les deux coefficients utilisent la mesure du point de départ : une erreur sur cette mesure se reporte sur les deux. Dans le plan à quatre essais, les poids de b_1 et de b_2 sont proportionnels à $\mathbf{x}_1$ et à $\mathbf{x}_2$, qui sont orthogonaux ; la covariance est nulle.

7.5 Estimer σ^2 à partir de répétitions

En pratique, σ^2 est inconnu. Si l'on réalise r fois le même essai et que l'on obtient des valeurs de moyenne $\bar{y}$, on l'estime par

$$s^2 = \frac{1}{r-1} \sum_{i=1}^r (y_i - \bar{y})^2.$$

On divise par $r-1$ et non par r ; ce nombre $r-1$ s'appelle le nombre de degrés de liberté. Avec deux répétitions, la formule se simplifie en $s^2 = (y_1 - y_2)^2/2$. Lorsque plusieurs essais sont répétés, on fait la moyenne des estimations obtenues pour chacun d'eux, ce qui augmente le nombre de degrés de liberté. Aucune de ces formules ne suppose que les erreurs suivent une loi normale ; cette hypothèse ne servira qu'à construire des tests, en séance 2.

Vérifiez votre compréhension

1. Les mesures $Y_1, \dots, Y_4$ ont la même variance σ^2 et des erreurs indépendantes. Calculez la variance de $\frac{Y_1+Y_2}{2} - \frac{Y_3+Y_4}{2}$.
2. Dans le plan à quatre essais, que vaut la variance de $b_{12} = \frac{y_1-y_2-y_3+y_4}{4}$?
3. On répète deux fois chacun des trois essais « un facteur à la fois » et l'on calcule b_1 à partir des moyennes. Que vaut $\text{Var}(b_1)$?

Réponses. 1. Quatre poids égaux à $\pm \frac{1}{2}$: $\sigma^2 \times 4 \times \frac{1}{4} = \sigma^2$. 2. $\sigma^2/4$. 3. Chaque moyenne a pour variance $\sigma^2/2$, et b_1 en est la demi-différence : $\frac{1}{4}(\frac{\sigma^2}{2} + \frac{\sigma^2}{2}) = \frac{\sigma^2}{4}$. C'est la précision du plan à quatre essais, mais il a fallu six essais.

Pour aller plus loin : les démonstrations

Hypothèses. On suppose $\text{Var}(Y_i) = \sigma^2$ pour tout i et $\text{Cov}(Y_i, Y_j) = 0$ pour $i \neq j$ (les erreurs de deux essais différents sont non corrélées).

Bilinéarité de la covariance. En notant $\tilde{Y} = Y - \mathbb{E}Y$, la linéarité de l'espérance donne $\text{Cov}(aY + bZ, W) = \mathbb{E}[(a\tilde{Y} + b\tilde{Z})\tilde{W}] = a \text{Cov}(Y, W) + b \text{Cov}(Z, W)$, et de même pour le second argument.

Démonstration du théorème. Par bilinéarité, $\text{Cov}(L, M) = \sum_i \sum_j a_i c_j \text{Cov}(Y_i, Y_j)$. Les termes avec $i \neq j$ sont nuls et ceux avec $i = j$ valent $a_i c_i \sigma^2$. On obtient $\text{Cov}(L, M) = \sigma^2 \sum_i a_i c_i$, et la variance s'en déduit avec $M = L$. ■

Pourquoi diviser par $r-1$. Pour deux répétitions, $\mathbb{E}(s^2) = \frac{1}{2} \mathbb{E}[(Y_1 - Y_2)^2] = \frac{1}{2} \text{Var}(Y_1 - Y_2) = \frac{1}{2} \times 2\sigma^2 = \sigma^2$, car $\mathbb{E}(Y_1 - Y_2) = 0$. L'estimation est donc juste en moyenne. Avec un diviseur égal à 2, elle serait trop faible de moitié. ■

8 La méthode des moindres carrés

Dans cette section

Comprendre le principe des moindres carrés ; ajuster une droite à la main ; utiliser la formule $\hat{\beta} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}$.

8.1 Le principe

Lorsqu'on dispose de plus d'essais que de coefficients, aucun jeu de coefficients ne reproduit exactement toutes les mesures, à cause des erreurs aléatoires. On cherche alors les coefficients qui s'en approchent le mieux. Pour chaque essai, on calcule l'écart entre la mesure et la prévision du modèle ; on élève ces écarts au carré et on les additionne. La **méthode des moindres carrés** consiste à choisir les coefficients qui rendent cette somme la plus petite possible.

Les écarts entre mesures et prévisions obtenus avec ces coefficients s'appellent les **résidus**.

Le cas le plus simple est celui d'un modèle réduit à une constante, $y = b_0$. On cherche alors le nombre b_0 le plus proche de toutes les mesures au sens des moindres carrés. La réponse est la moyenne $\bar{y}$ des mesures : la méthode des moindres carrés généralise donc le calcul d'une moyenne.

8.2 Ajuster une droite

Exemple résolu 8.1 : Une droite par trois points

On cherche la droite $y = b_0 + b_1x$ la plus proche des trois points $(-1; 2)$, $(0; 3)$ et $(1; 7)$. La somme des carrés des écarts vaut

$$S(b_0, b_1) = (2 - b_0 + b_1)^2 + (3 - b_0)^2 + (7 - b_0 - b_1)^2.$$

Au minimum, les dérivées par rapport à b_0 et à b_1 sont nulles. En dérivant et en simplifiant par -2 , on obtient deux équations :

$$\begin{cases} 3b_0 + 0b_1 = 12 \\ 0b_0 + 2b_1 = 5 \end{cases}$$

La première donne $b_0 = 4$, la seconde $b_1 = 2,5$. La droite des moindres carrés est $y = 4 + 2,5x$. Ses prévisions valent 1,5 ; 4 et 6,5 ; les résidus valent donc 0,5 ; -1 et 0,5.

Ces deux équations s'appellent les **équations normales**. Il est remarquable qu'on puisse les écrire directement avec les matrices de la section 4. Avec

$$\mathbf{X} = \begin{pmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 1 \end{pmatrix} \quad \text{et} \quad \mathbf{y} = \begin{pmatrix} 2 \\ 3 \\ 7 \end{pmatrix},$$

on trouve $\mathbf{X}^\top \mathbf{X} = \begin{pmatrix} 3 & 0 \\ 0 & 2 \end{pmatrix}$ et $\mathbf{X}^\top \mathbf{y} = \begin{pmatrix} 12 \\ 5 \end{pmatrix}$: ce sont exactement les nombres qui apparaissent dans les équations. Les équations normales s'écrivent donc $\mathbf{X}^\top \mathbf{X} \beta = \mathbf{X}^\top \mathbf{y}$.

8.3 La formule générale

Ce résultat est général : quel que soit le modèle linéaire, les coefficients des moindres carrés vérifient

$$\mathbf{X}^\top \mathbf{X} \hat{\beta} = \mathbf{X}^\top \mathbf{y}.$$

Le chapeau sur $\hat{\beta}$ rappelle qu'il s'agit d'une estimation, calculée à partir des mesures. Lorsque le modèle est identifiable, $\mathbf{X}^\top \mathbf{X}$ est inversible (section 6) et l'on peut multiplier les deux membres par son inverse :

$$\hat{\beta} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}.$$

Méthode : estimer les coefficients d'un modèle

1. Écrire la matrice $\mathbf{X}$ (une colonne par coefficient) et le vecteur $\mathbf{y}$ des mesures.
2. Calculer $\mathbf{X}^\top \mathbf{X}$ et $\mathbf{X}^\top \mathbf{y}$ à l'aide des produits scalaires de colonnes.
3. Inverser $\mathbf{X}^\top \mathbf{X}$, puis multiplier cet inverse par $\mathbf{X}^\top \mathbf{y}$.
4. Contrôler : calculer les résidus et vérifier que leur somme est nulle lorsque le modèle contient une constante.

Exemple résolu 8.2 : Le plan à quatre essais

Avec les rendements $\mathbf{y} = (23, 37, 41, 59)^\top$ et le modèle $b_0 + b_1x_1 + b_2x_2$, on a vu que $\mathbf{X}^\top \mathbf{X} = 4\mathbf{I}$ et $\mathbf{X}^\top \mathbf{y} = (160, 32, 40)^\top$. L'inverse de $\mathbf{X}^\top \mathbf{X}$ vaut $\frac{1}{4}\mathbf{I}$, donc

$$\hat{\beta} = \frac{1}{4} (160, 32, 40)^\top = (40, 8, 10)^\top.$$

Chaque coefficient est simplement le produit scalaire de sa colonne avec $\mathbf{y}$, divisé par 4 : on retrouve le calcul de la section 1. Les résidus valent $(23 - 22, 37 - 38, 41 - 42, 59 - 58) = (1, -1, -1, 1)$.

Quand le nombre d'essais est égal au nombre de coefficients et que le modèle est identifiable, le modèle passe exactement par tous les points et les résidus sont nuls. On dit que le plan est **saturé**. Un tel plan ne laisse aucune possibilité d'estimer σ^2 , ni de vérifier que le modèle convient.

Vérifiez votre compréhension

1. Ajustez le modèle $b_0 + b_1x_1 + b_2x_2$ aux trois essais « un facteur à la fois » de rendements 23, 37 et 41, à l'aide de l'inverse calculé à la section 6. Que valent les résidus ?
2. Pourquoi la somme des résidus est-elle nulle lorsque $\mathbf{X}$ contient une colonne de 1 ?

Réponses. 1. $\mathbf{X}^\top \mathbf{y} = (101, -27, -19)^\top$, puis $\hat{\beta} = (39, 7, 9)^\top$. Les résidus sont nuls : trois essais pour trois coefficients, le plan est saturé. 2. Les équations normales signifient que le vecteur des résidus est orthogonal à chaque colonne de $\mathbf{X}$; pour la colonne de 1, ce produit scalaire est la somme des résidus.

Pour aller plus loin : les démonstrations

Démonstration des équations normales. Soit $\hat{\beta}$ tel que $\mathbf{X}^\top \mathbf{X} \hat{\beta} = \mathbf{X}^\top \mathbf{y}$, et soit $\mathbf{e} = \mathbf{y} - \mathbf{X} \hat{\beta}$ le vecteur des résidus. On a $\mathbf{X}^\top \mathbf{e} = \mathbf{X}^\top \mathbf{y} - \mathbf{X}^\top \mathbf{X} \hat{\beta} = \mathbf{0}$: le résidu est orthogonal à chaque colonne de $\mathbf{X}$. Pour tout autre vecteur β , on écrit $\mathbf{y} - \mathbf{X} \beta = \mathbf{e} + \mathbf{X}(\hat{\beta} - \beta)$ et l'on développe :

$$\|\mathbf{y} - \mathbf{X} \beta\|^2 = \|\mathbf{e}\|^2 + 2\mathbf{e}^\top \mathbf{X}(\hat{\beta} - \beta) + \|\mathbf{X}(\hat{\beta} - \beta)\|^2 = \|\mathbf{e}\|^2 + \|\mathbf{X}(\hat{\beta} - \beta)\|^2,$$

car le terme du milieu est nul. La somme des carrés est donc minimale pour $\beta = \hat{\beta}$. Si le modèle est identifiable, $\mathbf{X}(\hat{\beta} - \beta) = \mathbf{0}$ impose $\beta = \hat{\beta}$: le minimum est unique. ■

Interprétation géométrique. Lorsque β varie, les prévisions $\mathbf{X} \beta$ parcourent l'ensemble des combinaisons linéaires des colonnes de $\mathbf{X}$. Les moindres carrés choisissent la combinaison la plus proche de $\mathbf{y}$: c'est la projection orthogonale de $\mathbf{y}$ sur cet ensemble, et le résidu lui est perpendiculaire.

Plan saturé. Si $\mathbf{X}$ est carrée et inversible, $(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top = \mathbf{X}^{-1}(\mathbf{X}^\top)^{-1} \mathbf{X}^\top = \mathbf{X}^{-1}$, donc $\hat{\beta} = \mathbf{X}^{-1} \mathbf{y}$ et $\mathbf{X} \hat{\beta} = \mathbf{y}$. ■

9 La précision des coefficients

Dans cette section

Comprendre et utiliser la formule $\text{Cov}(\hat{\beta}) = \sigma^2(\mathbf{X}^\top \mathbf{X})^{-1}$; lire les variances et les corrélations des coefficients dans $(\mathbf{X}^\top \mathbf{X})^{-1}$; comprendre pourquoi l'orthogonalité et la position des points comptent.

9.1 Un résultat déjà connu, écrit autrement

Pour la stratégie « un facteur à la fois », nous avons trouvé à la section 7 que $\text{Var}(b_1) = \text{Var}(b_2) = \sigma^2/2$ et $\text{Cov}(b_1, b_2) = \sigma^2/4$. Rangeons ces valeurs dans un tableau, en ajoutant la constante b_0 :

	b_0	b_1	b_2
b_0	$\sigma^2/2$	$\sigma^2/4$	$\sigma^2/4$
b_1	$\sigma^2/4$	$\sigma^2/2$	$\sigma^2/4$
b_2	$\sigma^2/4$	$\sigma^2/4$	$\sigma^2/2$

La diagonale contient les variances des trois coefficients, et les autres cases leurs covariances deux à deux. Un tel tableau s'appelle la **matrice de covariance** des coefficients, notée $\text{Cov}(\hat{\beta})$.

Comparons-le à l'inverse de $\mathbf{X}^\top \mathbf{X}$ calculé à la section 6 pour ce même plan : $\frac{1}{2}$ sur la diagonale, $\frac{1}{4}$ ailleurs. Les deux tableaux sont identiques, au facteur σ^2 près. Ce n'est pas une coïncidence.

9.2 La formule générale

Théorème 9.1 : Covariance des coefficients

On suppose que les mesures ont toutes la même variance σ^2 , que les erreurs de deux essais différents ne sont pas liées, que les niveaux des facteurs sont réglés exactement, et que le modèle est identifiable. Alors

$$\text{Cov}(\hat{\beta}) = \sigma^2(\mathbf{X}^\top \mathbf{X})^{-1}.$$

La variance du coefficient b_j se lit sur la diagonale de $(\mathbf{X}^\top \mathbf{X})^{-1}$, multipliée par σ^2 ; la covariance de b_j et b_k se lit dans la case (j, k) .

Cette formule a une conséquence importante pour l'expérimentateur. La matrice $(\mathbf{X}^\top \mathbf{X})^{-1}$ ne dépend que de $\mathbf{X}$, c'est-à-dire des essais choisis et du modèle, et pas des résultats. On peut donc calculer la précision d'un plan *avant* de réaliser la moindre expérience, et comparer plusieurs plans sur le papier.

Exemple résolu 9.2 : Comparer deux plans de quatre essais

Le plan factoriel à quatre essais donne $\mathbf{X}^\top \mathbf{X} = 4\mathbf{I}$, donc $\text{Cov}(\hat{\beta}) = \frac{\sigma^2}{4}\mathbf{I}$: chaque coefficient a une variance $\sigma^2/4$, et les coefficients ne sont pas corrélés.

Le plan « un facteur à la fois » avec une répétition du point de départ, formé des essais $(-1, -1)$, $(-1, +1)$, $(+1, -1)$ et $(+1, +1)$, donne

$$\mathbf{X}^\top \mathbf{X} = \begin{pmatrix} 4 & -2 & -2 \\ -2 & 4 & 0 \\ -2 & 0 & 4 \end{pmatrix}, \quad (\mathbf{X}^\top \mathbf{X})^{-1} = \frac{1}{8} \begin{pmatrix} 4 & 2 & 2 \\ 2 & 3 & 1 \\ 2 & 1 & 3 \end{pmatrix}.$$

On lit $\text{Var}(b_1) = \frac{3}{8}\sigma^2$, soit 50 % de plus qu'avec le plan factoriel, et $\text{Corr}(b_1, b_2) = \frac{1/8}{3/8} = \frac{1}{3}$. Les deux plans ont le même nombre d'essais et permettent tous deux d'estimer le modèle, mais ils n'ont pas la même précision.

9.3 Le rôle de l'orthogonalité

Lorsque les colonnes de $\mathbf{X}$ sont deux à deux orthogonales, $\mathbf{X}^\top \mathbf{X}$ est diagonale, et son inverse aussi. Le calcul devient alors très simple : chaque coefficient vaut $\hat{b}_j = \mathbf{x}_j \cdot \mathbf{y} / \|\mathbf{x}_j\|^2$, sa variance vaut $\sigma^2 / \|\mathbf{x}_j\|^2$, et les coefficients ne sont pas corrélés. Pour des colonnes de ± 1 , on a $\|\mathbf{x}_j\|^2 = N$, et donc $\text{Var}(\hat{b}_j) = \sigma^2 / N$.

On peut démontrer que, si les niveaux codés restent compris entre -1 et $+1$, aucun plan de N essais ne fait mieux que σ^2 / N pour un coefficient, et que cette valeur n'est atteinte que par une colonne de ± 1 orthogonale à toutes les autres (voir « Pour aller plus loin »). C'est ce qui fait l'intérêt des plans orthogonaux, que l'on retrouvera en séance 2.

9.4 Le rôle de la position des points

Considérons le plan factoriel resserré, dont les quatre essais sont placés aux niveaux $\pm 0,5$ au lieu de ± 1 . Ses colonnes restent orthogonales, mais leur norme au carré ne vaut plus que $4 \times 0,25 = 1$. La variance des coefficients b_1 et b_2 vaut alors σ^2 , quatre fois plus que pour le plan aux niveaux ± 1 . Le nombre d'essais n'a pas changé, ni l'orthogonalité : seule la position des points a changé. Pour mesurer une pente avec des mesures bruitées, il vaut mieux placer les essais loin les uns des autres, dans la limite d'un domaine réaliste et où le modèle reste valable.

9.5 Cinq notions à ne pas confondre

Les sections précédentes ont fait apparaître cinq propriétés différentes. Un plan est *identifiable* pour un modèle lorsque le rang de $\mathbf{X}$ est égal au nombre de coefficients ; il est *orthogonal* lorsque $\mathbf{X}^\top \mathbf{X}$ est diagonale ; il est *précis* lorsque les termes diagonaux de $\sigma^2(\mathbf{X}^\top \mathbf{X})^{-1}$ sont petits. Ces trois propriétés dépendent du plan et du modèle. Les deux autres dépendent du phénomène étudié : le modèle est *valide* s'il ne néglige aucun terme important, et les facteurs sont *sans interaction* si l'effet de l'un ne dépend pas du niveau de l'autre. Ces propriétés ne doivent pas être confondues. Des colonnes non nulles et deux à deux orthogonales sont indépendantes : elles garantissent donc l'identifiabilité du modèle. Cela ne garantit ni sa validité physique, ni une précision suffisante, ni l'absence d'interaction. Le plan « un facteur à la fois » est identifiable mais pas orthogonal ; le plan resserré est orthogonal mais peu précis ; un plan orthogonal peut donner des coefficients faux si le modèle oublie un terme.

Vérifiez votre compréhension

1. Pour le plan factoriel dupliqué (huit essais), que vaut $\text{Var}(\hat{b}_1)$?
2. Vérifiez, en calculant une ligne du produit, que la matrice $\frac{1}{8} \begin{pmatrix} 4 & 2 & 2 \\ 2 & 3 & 1 \\ 2 & 1 & 3 \end{pmatrix}$ est bien l'inverse de $\begin{pmatrix} 4 & -2 & -2 \\ -2 & 4 & 0 \\ -2 & 0 & 4 \end{pmatrix}$.
3. Un plan est orthogonal. Peut-on en conclure que ses coefficients sont toujours justes ?

Réponses. 1. $\mathbf{X}^\top \mathbf{X} = 8\mathbf{I}$, donc $\text{Var}(\hat{b}_1) = \sigma^2 / 8$. 2. Deuxième ligne de la matrice multipliée par les colonnes de l'inverse : $\frac{1}{8}(-8 + 8 + 0) = 0$, $\frac{1}{8}(-4 + 12 + 0) = 1$, $\frac{1}{8}(-4 + 4 + 0) = 0$. 3. Non : si le modèle néglige un terme réellement présent, les coefficients peuvent être biaisés (voir « Pour aller plus loin »).

Pour aller plus loin : les démonstrations

Matrice de covariance d'un vecteur. Pour un vecteur aléatoire $\mathbf{Y}$, on note $\text{Cov}(\mathbf{Y})$ la matrice dont la case (i, j) est $\text{Cov}(Y_i, Y_j)$. Les hypothèses du théorème s'écrivent $\text{Cov}(\mathbf{Y}) = \sigma^2 \mathbf{I}$. Pour une matrice $\mathbf{A}$ non aléatoire, la bilinéarité de la covariance donne

$$\text{Cov}(\mathbf{AY}) = \mathbf{A} \text{Cov}(\mathbf{Y}) \mathbf{A}^\top,$$

car la case (k, l) du membre de gauche vaut $\sum_i \sum_j a_{ki} \text{Cov}(Y_i, Y_j) a_{lj}$.

Démonstration du théorème. Posons $\mathbf{M} = \mathbf{X}^\top \mathbf{X}$ et $\mathbf{A} = \mathbf{M}^{-1} \mathbf{X}^\top$, de sorte que $\hat{\boldsymbol{\beta}} = \mathbf{AY}$. Comme $\mathbf{M}$ est symétrique, $\mathbf{M}^{-1}$ l'est aussi, et $\mathbf{A}^\top = \mathbf{XM}^{-1}$. Donc

$$\text{Cov}(\hat{\boldsymbol{\beta}}) = \mathbf{A}(\sigma^2 \mathbf{I}) \mathbf{A}^\top = \sigma^2 \mathbf{M}^{-1} \mathbf{X}^\top \mathbf{X} \mathbf{M}^{-1} = \sigma^2 \mathbf{M}^{-1} \mathbf{M} \mathbf{M}^{-1} = \sigma^2 \mathbf{M}^{-1}. \quad \square$$

Absence de biais. Si le modèle est juste, $E(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta}$ et $E(\hat{\boldsymbol{\beta}}) = \mathbf{AX}\boldsymbol{\beta} = \mathbf{M}^{-1} \mathbf{M}\boldsymbol{\beta} = \boldsymbol{\beta}$: en moyenne, l'estimation est égale à la vraie valeur.

La borne σ^2/N . Notons $\mathbf{c}^\top$ la ligne j de $\mathbf{A}$, de sorte que $\hat{b}_j = \mathbf{c} \cdot \mathbf{y}$ et $\text{Var}(\hat{b}_j) = \sigma^2 \|\mathbf{c}\|^2$. Comme $\mathbf{AX} = \mathbf{I}$, on a $\mathbf{c} \cdot \mathbf{x}_j = 1$ et $\mathbf{c} \cdot \mathbf{x}_k = 0$ pour $k \neq j$. L'inégalité de Cauchy-Schwarz (section 2) donne $1 \leq \|\mathbf{c}\|^2 \|\mathbf{x}_j\|^2$, donc $\text{Var}(\hat{b}_j) \geq \sigma^2 / \|\mathbf{x}_j\|^2$. Si tous les niveaux sont compris entre -1 et 1 , $\|\mathbf{x}_j\|^2 \leq N$, d'où $\text{Var}(\hat{b}_j) \geq \sigma^2 / N$. L'égalité exige que $\mathbf{c}$ soit proportionnel à $\mathbf{x}_j$, ce qui, avec $\mathbf{c} \cdot \mathbf{x}_k = 0$, impose que $\mathbf{x}_j$ soit orthogonale aux autres colonnes, et que tous ses éléments valent ± 1 . ■

Quand le modèle oublie un terme. Si la vraie réponse moyenne est $\mathbf{X}_1 \boldsymbol{\beta}_1 + \mathbf{X}_2 \boldsymbol{\beta}_2$ mais que l'on n'ajuste que $\mathbf{X}_1 \boldsymbol{\beta}_1$, on obtient $E(\hat{\boldsymbol{\beta}}_1) = \boldsymbol{\beta}_1 + (\mathbf{X}_1^\top \mathbf{X}_1)^{-1} \mathbf{X}_1^\top \mathbf{X}_2 \boldsymbol{\beta}_2$. Le second terme est le biais. Deux exemples. Avec la stratégie « un facteur à la fois », oublier l'interaction donne $E(\hat{b}_1) = b_1 - b_{12}$: on mesure l'effet de la température au point de départ, et non son effet moyen. Avec le plan factoriel, oublier un terme de courbure $b_{11} x_1^2$ biaise la constante : comme $x_1^2 = 1$ pour les quatre essais, la colonne oubliée est identique à la colonne de 1, et $E(\hat{b}_0) = b_0 + b_{11}$. Le plan est orthogonal, et pourtant rien dans les résultats ne signale l'erreur.

10 Exercices corrigés

Les exercices suivent l'ordre du fascicule. Essayez de les faire sans regarder la solution ; si vous bloquez, relisez la méthode indiquée.

Exercice 10.1 (produit matriciel, section 3) Calculer $\begin{pmatrix} 1 & 0 & 2 \\ -1 & 3 & 1 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 1 & 0 \\ 0 & 4 \end{pmatrix}$. Le produit dans l'autre ordre existe-t-il ? Quelles seraient ses dimensions ?

Exercice 10.2 (matrice $\mathbf{X}^\top \mathbf{X}$, section 4) Un plan comporte les cinq essais codés $(-1, -1)$, $(1, -1)$, $(-1, 1)$, $(1, 1)$ et $(0, 0)$. Écrire la matrice $\mathbf{X}$ du modèle $y = b_0 + b_1x_1 + b_2x_2$, puis calculer $\mathbf{X}^\top \mathbf{X}$. Le plan est-il orthogonal ?

Exercice 10.3 (rang, section 5) Un collègue propose quatre essais aux points $(-1, -1)$, $(-1, 1)$, $(1, 1)$ et $(1, 1)$ « pour avoir des répétitions ». Calculer le rang de $\mathbf{X}$ pour le modèle $y = b_0 + b_1x_1 + b_2x_2$. Que peut-on estimer avec ce plan ?

Exercice 10.4 (inverse, section 6) Inverser la matrice $\mathbf{X}^\top \mathbf{X}$ obtenue à l'exercice 10.2, puis en déduire les variances des trois coefficients.

Exercice 10.5 (variance, section 7) On estime l'influence d'un facteur par $\frac{Y_1+Y_2+Y_3}{3} - Y_4$, où les mesures ont la même variance σ^2 et des erreurs indépendantes. Calculer la variance de cette quantité, puis comparer avec $\frac{Y_1+Y_2}{2} - \frac{Y_3+Y_4}{2}$.

Exercice 10.6 (moindres carrés, section 8) Ajuster la droite $y = b_0 + b_1x$ aux points $(1; 3)$, $(2; 4)$ et $(4; 8)$ par la formule $\hat{\beta} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}$.

Exercice 10.7 (précision, section 9) Pour la droite de l'exercice précédent, calculer $\text{Var}(\hat{b}_0)$, $\text{Var}(\hat{b}_1)$ et la corrélation des deux coefficients. Recommencer en remplaçant x par $x' = x - \frac{7}{3}$ (abscisses centrées). Que constate-t-on ?

Solutions

Solution de l'exercice 10.1. Le produit d'une matrice 2×3 par une matrice 3×2 donne une matrice 2×2 . Première ligne : $1 \times 2 + 0 \times 1 + 2 \times 0 = 2$ et $1 \times 1 + 0 \times 0 + 2 \times 4 = 9$. Seconde ligne : $-2 + 3 + 0 = 1$ et $-1 + 0 + 4 = 3$. Le résultat est $\begin{pmatrix} 2 & 9 \\ 1 & 3 \end{pmatrix}$. Le produit dans l'autre ordre existe aussi, puisque $(3 \times 2)(2 \times 3)$ est compatible, et il donne une matrice 3×3 : les deux produits n'ont même pas la même taille.

Solution de l'exercice 10.2. $\mathbf{X}$ a cinq lignes $(1, x_1, x_2)$; la dernière est $(1, 0, 0)$. On trouve $\mathbf{1} \cdot \mathbf{1} = 5$, $\mathbf{x}_1 \cdot \mathbf{x}_1 = \mathbf{x}_2 \cdot \mathbf{x}_2 = 4$ (le point central n'ajoute rien), et tous les produits hors de la diagonale sont nuls. Donc $\mathbf{X}^\top \mathbf{X} = \begin{pmatrix} 5 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 4 \end{pmatrix}$: le plan est orthogonal.

Solution de l'exercice 10.3. Les colonnes $\mathbf{x}_1$ et $\mathbf{x}_2$ sont toutes deux égales à $(-1, -1, 1, 1)^\top$; le rang vaut 2. On ne peut estimer que b_0 et la somme $b_1 + b_2$. Les répétitions permettent d'estimer le bruit, avec 2 degrés de liberté, mais elles ne séparent pas les deux facteurs.

Solution de l'exercice 10.4. La matrice est diagonale ; son inverse s'obtient en inversant chaque terme : $\frac{1}{5}$, $\frac{1}{4}$, $\frac{1}{4}$. Donc $\text{Var}(\hat{b}_0) = \sigma^2/5$ et $\text{Var}(\hat{b}_1) = \text{Var}(\hat{b}_2) = \sigma^2/4$. Le point central améliore la précision de la constante, pas celle des pentes.

Solution de l'exercice 10.5. Les poids sont $\frac{1}{3}$, $\frac{1}{3}$, $\frac{1}{3}$ et -1 , d'où une variance $\sigma^2(\frac{1}{9} + \frac{1}{9} + \frac{1}{9} + 1) = \frac{4}{3}\sigma^2$. Pour la comparaison équilibrée, les quatre poids valent $\pm\frac{1}{2}$ et la variance vaut σ^2 . La comparaison équilibrée est plus précise, avec le même nombre de mesures.

Solution de l'exercice 10.6. $\mathbf{X}^\top \mathbf{X} = \begin{pmatrix} 3 & 7 \\ 7 & 21 \end{pmatrix}$ et $\mathbf{X}^\top \mathbf{y} = \begin{pmatrix} 15 \\ 43 \end{pmatrix}$. Le déterminant vaut 14, d'où $(\mathbf{X}^\top \mathbf{X})^{-1} = \frac{1}{14} \begin{pmatrix} 21 & -7 \\ -7 & 3 \end{pmatrix}$. On obtient $\hat{b}_0 = \frac{21 \times 15 - 7 \times 43}{14} = \frac{14}{14} = 1$ et $\hat{b}_1 = \frac{-7 \times 15 + 3 \times 43}{14} = \frac{24}{14} = \frac{12}{7}$.

Solution de l'exercice 10.7. On lit $\text{Var}(\hat{b}_0) = \frac{21}{14}\sigma^2 = \frac{3}{2}\sigma^2$ et $\text{Var}(\hat{b}_1) = \frac{3}{14}\sigma^2$. La corrélation vaut $\frac{-7}{\sqrt{21 \times 3}} \approx -0,88$: les deux coefficients sont fortement liés. Avec les abscisses centrées x' , la somme des x' est nulle, $\mathbf{X}^\top \mathbf{X}$ devient diagonale (3 et $\frac{14}{3}$), la pente garde la même variance $\frac{3}{14}\sigma^2$, la constante a une variance $\frac{\sigma^2}{3}$ et la corrélation disparaît. Centrer les facteurs, comme on le fait avec les variables codées, rend la colonne de chaque facteur orthogonale à celle de la constante.