Régression linéaire dans $\mathbb{R}^2$

L3$\{$*MA*, *Info_SI*, *Info_Reseaux*$\}$ - UFR S.A.T

Pr. Ousmane THIARE

othiare@ugb.edu.sn
[www.ousmanethiare.com]

16 avril 2020

Régression linéaire dans $\mathbb{R}^2$

Soit n un entier strictement positif et $\mathbb{R}$ le corps des nombres réels.
L'ensemble $\mathbb{R}^n$ des n-uples $(x_1, \cdots, x_n)$ de nombres réels est muni de sa structure usuelle d'**espace vectoriel réel**, définie par les opérations :

- $(x_1, \cdots, x_n) + (x'_1, \cdots, x'_n) = (x_1 + x'_1, \cdots, x_n + x'_n)$

Soit n un entier strictement positif et $\mathbb{R}$ le corps des nombres réels.
L'ensemble $\mathbb{R}^n$ des n-uples $(x_1, \cdots, x_n)$ de nombres réels est muni de sa structure usuelle d'**espace vectoriel réel**, définie par les opérations :

- $(x_1, \cdots, x_n) + (x'_1, \cdots, x'_n) = (x_1 + x'_1, \cdots, x_n + x'_n)$
- $l(x_1, \cdots, x_n) = (lx_1, \cdots, lx_n), l \in \mathbb{R}$.

Soit n un entier strictement positif et $\mathbb{R}$ le corps des nombres réels.
L'ensemble $\mathbb{R}^n$ des n-uples $(x_1, \cdots, x_n)$ de nombres réels est muni de sa structure usuelle d'**espace vectoriel réel**, définie par les opérations :

- $(x_1, \cdots, x_n) + (x'_1, \cdots, x'_n) = (x_1 + x'_1, \cdots, x_n + x'_n)$
- $l(x_1, \cdots, x_n) = (lx_1, \cdots, lx_n), l \in \mathbb{R}$.
- On identifiera un élément $X = (x_1, \cdots, x_n)$ de $\mathbb{R}^n$ avec la matrice $X = \begin{pmatrix} x_1 \\ \vdots \\ x_i \\ \vdots \\ x_n \end{pmatrix}$ à n lignes et 1 colonne.

Les opérations dans $\mathbb{R}^n$ sont alors définies par des opérations sur les matrices :

- **Addition :**

$$\begin{pmatrix} x_1 \\ \vdots \\ x_i \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} x'_1 \\ \vdots \\ x'_i \\ \vdots \\ x'_n \end{pmatrix} = \begin{pmatrix} x_1 + x'_1 \\ \vdots \\ x_i + x'_i \\ \vdots \\ x_n + x'_n \end{pmatrix}$$

$$(x_1, \cdots, x_n) + (x'_1, \cdots, x'_n) = (x_1 + x'_1, \cdots, x_n + x'_n)$$

- **Multiplication par un scalaire :**

$$l \begin{pmatrix} x_1 \\ \vdots \\ x_i \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} lx_1 \\ \vdots \\ lx_i \\ \vdots \\ lx_n \end{pmatrix}$$

$$l(x_1, \cdots, x_n) = (lx_1, \cdots, lx_n)$$

- **Multiplication par un scalaire :**

$$l \begin{pmatrix} x_1 \\ \vdots \\ x_i \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} lx_1 \\ \vdots \\ lx_i \\ \vdots \\ lx_n \end{pmatrix}$$

$$l(x_1, \cdots, x_n) = (lx_1, \cdots, lx_n)$$

- Dans $\mathbb{R}^n$, les n éléments $e_i, i \in \{1, \cdots, n\}$, dont toutes les coordonnées sont nulles, sauf la i^{me} qui vaut 1, forment une base, appelée la **base canonique** de $\mathbb{R}^n$.

  Tout élément $X = (x_1, \cdots, x_n)$ de $\mathbb{R}^n$ s'écrit : $X = \sum_{i=1}^{i=n} x_i e_i$

Produit scalaire dans $\mathbb{R}^n$

Définition

Soit F une application de $\mathbb{R}^n \times \mathbb{R}$ dans $\mathbb{R}$.
On notera aussi $\langle X|F|Y\rangle$ ou $\langle X|Y\rangle_F$, le nombre réel F(X,Y).
On appelle produit scalaire dans $\mathbb{R}^n$, toute application F de $\mathbb{R}^n \times \mathbb{R}^n$ dans $\mathbb{R}$ qui possède les propriétés suivantes :
a) Bilinéarité

- Linéarité par rapport à la première variable : $F(X+X', Y) = F(X, Y) + F(X', Y)$ et $F(lX, Y) = lF(X, Y)$, quels que soient l dans $\mathbb{R}$, $X, X', Y \in \mathbb{R}^n$; cette propriété s'écrit aussi

$$\langle X + X'|F|Y\rangle = \langle X|F|Y\rangle + \langle X'|F|Y\rangle$$

a) Bilinéarité

- Linéarité par rapport à la deuxième variable : $F(X, Y + Y') = F(X, Y) + F(X, Y')$ et $F(X, lY) = lF(X, Y)$, quels que soient l dans $\mathbb{R}$, $X, X', Y \in \mathbb{R}$; cette propriété s'écrit aussi

$$\langle X|F|Y + Y'\rangle = \langle X|F|Y\rangle + \langle X|F|Y'\rangle$$

b) Symétrie

c) Positivité

Produit scalaire dans $\mathbb{R}^n$

Définition

a) Bilinéarité

- Linéarité par rapport à la deuxième variable :
  $F(X, Y + Y') = F(X, Y) + F(X, Y')$ et
  $F(X, lY) = lF(X, Y)$, quels que soient l dans $\mathbb{R}$,
  $X, X', Y \in \mathbb{R}$; cette propriété s'écrit aussi

$$\langle X|F|Y + Y'\rangle = \langle X|F|Y\rangle + \langle X|F|Y'\rangle$$

b) Symétrie

- F(X,Y)=F(Y,X), quels que soient X et Y dans $\mathbb{R}^n$:
  $\langle X|F|Y\rangle = \langle Y|F|X\rangle$

c) Positivité

a) Bilinéarité

- Linéarité par rapport à la deuxième variable : $F(X, Y + Y') = F(X, Y) + F(X, Y')$ et $F(X, lY) = lF(X, Y)$, quels que soient l dans $\mathbb{R}$, $X, X', Y \in \mathbb{R}$; cette propriété s'écrit aussi

$$\langle X|F|Y + Y'\rangle = \langle X|F|Y\rangle + \langle X|F|Y'\rangle$$

b) Symétrie

- F(X,Y)=F(Y,X), quels que soient X et Y dans $\mathbb{R}^n$: $\langle X|F|Y\rangle = \langle Y|F|X\rangle$

c) Positivité

- F(X,X) est un nombre réel supérieur ou égal à 0, quel que soit X dans $\mathbb{R}^n$: $\langle X|F|X\rangle = 0$

d) Non dégénérescence

- F(X,X)=0 entraine X=$0_{\mathbb{R}^n}$:

$$\langle X|F|X\rangle = 0 \Rightarrow X = 0$$

Autrement dit, le vecteur $0 = (0, \cdots, 0, \cdots, 0)$ de $\mathbb{R}^n$ est l'unique solution de l'équation F(X,X)=0.
On dit aussi qu'un produit scalaire sur $\mathbb{R}^n$ est une **forme bilinéaire symétrique positive non dégénérée**. Le mot "forme" fait simplement référence au fait que les valeurs sont des scalaires. Lorsqu'il est muni d'un produit scalaire, $\mathbb{R}^n$ est appelé un **espace vectoriel euclidien**.

a) Produit scalaire canonique

- L'application de $\mathbb{R}^n \times \mathbb{R}^n$ dans $\mathbb{R}$ définie par : $((x_1, \cdots, x_n), (y_1, \cdots, y_n)) \mapsto \langle X|Y \rangle = X^t Y =$
$$\begin{pmatrix} x_1 \cdots, x_j, \cdots, x_n \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_i \\ \vdots \\ y_n \end{pmatrix} = \sum_{i=1}^{i=n} x_i y_i$$
est un produit scalaire sur $\mathbb{R}^n$ qu'on appelle le **produit scalaire canonique** de $\mathbb{R}^n$.

b) Produit scalaire défini par une matrice diagonale à éléments positifs

Soit une matrice réelle M à n lignes et n colonnes dont tous les éléments en dehors de la diagonale principale sont nuls et dont tous les éléments de la diagonale principale sont des **nombres réels strictement positifs.**

Alors l'application : $(X, Y) \mapsto \langle X|M|Y\rangle = X^t MY =$

$$(x_1 \cdots, x_j, \cdots, x_n)M \begin{pmatrix} y_1 \\ \vdots \\ y_i \\ \vdots \\ y_n \end{pmatrix} =$$

$m_{ij}x_jy_i = m_{ii}x_iy_i$ est un produit scalaire sur $\mathbb{R}^n$. M **mat des poids.**

b) Produit scalaire défini par une matrice diagonale à éléments positifs

- Le produit scalaire canonique correspond au cas où la matrice M est la matrice unité I_n (tous les éléments de la diagonale sont égaux à 1 et les élements en dehors de la diagonale sont 0) : tous les poids sont égaux à 1.

b) Produit scalaire défini par une matrice diagonale à éléments positifs

- Le produit scalaire canonique correspond au cas où la matrice M est la matrice unité I_n (tous les éléments de la diagonale sont égaux à 1 et les élements en dehors de la diagonale sont 0) : tous les poids sont égaux à 1.
- Autre exemple : $M = D_{\frac{1}{n}} = \frac{1}{n} I_n$. Tous les poids sont égauxà$\frac{1}{n}$ et la somme des poids vaut 1.

Pour tout produit scalaire F de $\mathbb{R}^n$, on peut écrire :

$$F(X,Y) = F(\sum_{i=1}^{i=n} x_i e_i, \sum_{j=1}^{j=n} y_j e_j) = \sum_{i,j=1}^{n} F(e_i,e_j) x_i y_j =$$

$$(x_1, \cdots, x_i, \cdots, x_n) M_F \begin{pmatrix} y_1 \\ \vdots \\ y_j \\ \vdots \\ y_n \end{pmatrix}$$ La matrice $M_F = [F(e_i,e_j)]$

s'appelle la **matrice du produit scalaire** F dans la base canonique. Cette matrice est symétrique :
$F(e_i,e_j) = F(e_j,e_i)$.
Les éléments de sa diagonale sont : $F(e_i,e_j) > 0$.

Remarquons que ces propriétés ne sont pas suffisantes : une matrice symétrique dont les éléments de la diagonale sont des nombres réels strictement positifs bilinéaire $((x_1, x_2), (y_1, y_2)) \mapsto (x_1, x_2) \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$ qu'elle définit n'est pas un produit scalaire car le "produit scalaire" du vecteur propre (1,-1) pour la valeur propre négative, par lui même, est un nombre réel strictement négatif $(1, -1) \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = -2$

La matrice $\begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$ n'est donc pas la matrice d'un produit scalaire sur $\mathbb{R}^2$, bien qu'elle soit symétrique et que les éléments de sa diagonale soient strictement positifs.

Théorème

En réalité, pour qu'une matrice carrée symétrique réelle soit la matrice d'un produit scalaire, il faut et il suffit que toutes ses valeurs propres, qui sont toujours des nombres réels, soient strictement positives.

Si F est un produit scalaire sur $\mathbb{R}^n$, le nombre réel positif $\|X\|_F = \sqrt{F(X,X)}$ s'appelle la **F-norme** de X, ou **F-longueur** de X. Quand il n'y a pas de confusion à craindre, on parlera simplement de norme ou de longueur, qu'on notera $\|X\|$ au lieu de $\|X\|_F$.

On dit qu'un vecteur est **normé pour F** si sa F-longueur est 1.

Par exemple, dans $\mathbb{R}^2$ muni du produit scalaire canonique, la longueur de $X = (x_1, x_2)$ est $\|X\|_F = \sqrt{x_1^2 + x_2^2}$ et le vecteur (1,0) est normé.

Etant donnés deux vecteurs X et Y de $\mathbb{R}^n$ et un produit scalaire F sur $\mathbb{R}^n$, pour tout nombre réel l, on a :

- $F(X + lY, X + lY) = \|X + lY\|_F^2 \geq 0$

Comme cette relation est vraie pour tout nombre réel l, c'est que le déterminant de ce trinôme du second degré est négatif :
$(\langle X|Y\rangle_F)^2 - \|X\|_F^2\|Y\|_F^2 \leq 0 \Longrightarrow |\langle X|Y\rangle_F| \leq \|X\|_F\|Y\|_F$
(**inégalité de Schwarz**).

Etant donnés deux vecteurs X et Y de $\mathbb{R}^n$ et un produit scalaire F sur $\mathbb{R}^n$, pour tout nombre réel l, on a :

- $F(X + lY, X + lY) = \|X + lY\|_F^2 \geq 0$
- $l^2 F(Y, Y) + l(F(Y, X) + F(X, Y)) + F(X, X) \geq 0$

Comme cette relation est vraie pour tout nombre réel l, c'est que le déterminant de ce trinôme du second degré est négatif :
$(\langle X|Y\rangle_F)^2 - \|X\|_F^2\|Y\|_F^2 \leq 0 \Longrightarrow |\langle X|Y\rangle_F| \leq \|X\|_F\|Y\|_F$
(**inégalité de Schwarz**).

Etant donnés deux vecteurs X et Y de $\mathbb{R}^n$ et un produit scalaire F sur $\mathbb{R}^n$, pour tout nombre réel l, on a :

- $F(X+lY, X+lY) = \|X+lY\|_F^2 \geq 0$
- $l^2 F(Y,Y) + l(F(Y,X) + F(X,Y)) + F(X,X) \geq 0$
- $l^2 F(Y,Y) + 2lF(X,Y) + F(X,X) \geq 0$

Comme cette relation est vraie pour tout nombre réel l, c'est que le déterminant de ce trinôme du second degré est négatif :
$(\langle X|Y\rangle_F)^2 - \|X\|_F^2\|Y\|_F^2 \leq 0 \Longrightarrow |\langle X|Y\rangle_F| \leq \|X\|_F\|Y\|_F$ (**inégalité de Schwarz**).

Etant donnés deux vecteurs X et Y de $\mathbb{R}^n$ et un produit scalaire F sur $\mathbb{R}^n$, pour tout nombre réel l, on a :

- $F(X + lY, X + lY) = \|X + lY\|_F^2 \geq 0$
- $l^2F(Y, Y) + l(F(Y, X) + F(X, Y)) + F(X, X) \geq 0$
- $l^2F(Y, Y) + 2lF(X, Y) + F(X, X) \geq 0$
- $\|Y\|_F^2 l^2 + 2\langle X|Y\rangle_F l + \|X\|_F^2 \geq 0$

Comme cette relation est vraie pour tout nombre réel l, c'est que le déterminant de ce trinôme du second degré est négatif :
$(\langle X|Y\rangle_F)^2 - \|X\|_F^2\|Y\|_F^2 \leq 0 \Longrightarrow |\langle X|Y\rangle_F| \leq \|X\|_F\|Y\|_F$
(**inégalité de Schwarz**).

Si les deux vecteurs X et Y sont différents de 0, leur longueur n'est pas nulle, le produit de leurs longueurs n'est pas nul, le rapport $\frac{\langle X|Y\rangle_\phi}{\|X\|_\phi\|Y\|_\phi}$ est compris entre -1 et 1, et il existe donc un angle compris entre 0 et p radians dont le cosinus est égal au rapport $\frac{\langle X|Y\rangle_\phi}{\|X\|_\phi\|Y\|_\phi}$.
Par définition, cet angle unique a compris entre 0 et p, vérifiant :

$$cos\ a = \frac{\langle X|Y\rangle_\phi}{\|X\|_\phi\|Y\|_\phi} = \frac{\langle X|\phi|Y\rangle}{\|X\|_\phi\|Y\|_\phi}$$

est appelé l'**angle des deux vecteurs non nuls** X et Y.

Etant donnés deux vecteurs X et Y de $\mathbb{R}^n$ et un produit scalaire F sur $\mathbb{R}^n$, on dit que X et Y sont **F-orthogonaux** (ou simplement "orthogonaux" s'il n'y a pas de confusion à craindre) si, et seulement si, leur produit scalaire est nul :

$$F(X, Y) = \langle X|Y\rangle_F = 0$$

Exemples :

- 0 est F-orthogonal à tout vecteur de $\mathbb{R}^n$.

Etant donnés deux vecteurs X et Y de $\mathbb{R}^n$ et un produit scalaire F sur $\mathbb{R}^n$, on dit que X et Y sont **F-orthogonaux** (ou simplement "orthogonaux" s'il n'y a pas de confusion à craindre) si, et seulement si, leur produit scalaire est nul :

$$F(X, Y) = \langle X|Y\rangle_F = 0$$

Exemples :

- 0 est F-orthogonal à tout vecteur de $\mathbb{R}^n$.
- L'angle de deux vecteurs non nuls F-orthogonaux est $\frac{\pi}{2}$

Notions d'espace vectoriel euclidien
Produit scalaire dans $\mathbb{R}^n$
Propriétés
Approche euclidienne de la régression

Etant donnés deux vecteurs X et Y de $\mathbb{R}^n$ et un produit scalaire F sur $\mathbb{R}^n$, on dit que X et Y sont **F-orthogonaux** (ou simplement "orthogonaux" s'il n'y a pas de confusion à craindre) si, et seulement si, leur produit scalaire est nul :

$$F(X, Y) = \langle X|Y\rangle_F = 0$$

Exemples :

- 0 est F-orthogonal à tout vecteur de $\mathbb{R}^n$.
- L'angle de deux vecteurs non nuls F-orthogonaux est $\frac{\pi}{2}$
- La base canonique de $\mathbb{R}^n$ muni du produit scalaire canonique est formée de vecteurs normés orthogonaux deux à deux : on parle alors de **base orthonormée.**

Notions d'espace vectoriel euclidien

Produit scalaire dans $\mathbb{R}^n$

Propriétés

Approche euclidienne de la régression

Soient X et Y deux vecteurs non nuls de $\mathbb{R}^n$ et F un produit scalaire sur $\mathbb{R}^n$.
Il existe un unique vecteur Z de $\mathbb{R}^n$, proportionnelàY tel que X-Z soit orthogonalàY.

Preuve

Pour tout vecteur Z, on peut écrire :

- $\langle X - Z | Y \rangle_F = \langle X | Y \rangle_F - \langle Z | Y \rangle_F$
  Si l'on prend un Z proportionnel à Y, on a Z=aY, donc :

Pour que X-Z soit orthogonal à Y, soit $\langle X - Z | Y \rangle_F = 0$, il faut et il suffit que l'on prenne $a = \dfrac{\langle X | Y \rangle_\phi}{\| Y \|_\phi^2}$

Soient X et Y deux vecteurs non nuls de $\mathbb{R}^n$ et F un produit scalaire sur $\mathbb{R}^n$.
Il existe un unique vecteur Z de $\mathbb{R}^n$, proportionnelàY tel que X-Z soit orthogonalàY.

Preuve

Pour tout vecteur Z, on peut écrire :

- $\langle X - Z|Y\rangle_F = \langle X|Y\rangle_F - \langle Z|Y\rangle_F$
  Si l'on prend un Z proportionnel à Y, on a Z=aY, donc :
- $\langle X - Z|Y\rangle_F = \langle X|Y\rangle_F - a\langle Y|Y\rangle_F = \langle X|Y\rangle_F - a\|Y\|_F^2$

Pour que X-Z soit orthogonal à Y, soit $\langle X - Z|Y\rangle_F = 0$, il faut et il suffit que l'on prenne $a = \dfrac{\langle X|Y\rangle_\phi}{\|Y\|_\phi^2}$

Le projeté orthogonal Z_0 de X sur Y est le vecteur Z de $\mathbb{R}^n$ proportionnel à Y, qui minimise $\|X - Z\|_F^2$.

Preuve

Soit Z un vecteur proportionnel à Y

Soit $Z_0 = \dfrac{\langle X|Y\rangle_\phi}{\|Y\|_\phi^2} Y$ le projeté orthogonal de X sur Y.

$\|X - Z\|_F^2 = \|X - Z_0 + Z_0 - Z\|_F^2$

Comme Z est proportionnel à Y et que Z_0 est proportionnel à Y, la différence $Z_0 - Z$ est proportionnelleàY. Or $X - Z_0$ est orthogonal à Y, donc $X - Z_0$ est orthogonal à $Z_0 - Z$ qui est proportionnel àY. Il en résulte que l'on a :

$\|X - Z\|_F^2 = \|X - Z_0 + Z_0 - Z\|_F^2 =$
$\|X - Z_0\|_F^2 + \|Z_0 - Z\|_F^2 \geq \|X - Z_0\|_F^2$. $\|X - Z\|_F^2$ atteint son minimum lorsque $Z = Z_0$.

Régression linéaire dans $\mathbb{R}^2$

Considérons une variable statistique quantitative bidimensionnelle (X,Y)àvaleurs dans $\mathbb{R}^2$, définie dans une population Ω de taille n. Elle est définie par l'ensemble des couples $\{(X(w), Y(w))\}_{w\in\Omega}$. $\mathbb{R}^2$ est l'**espace des variables**.

La variable statistique est représentée par un nuage de points dans $\mathbb{R}^2$ et chaque point du nuage statistique représente un individu dans la population Ω.

Les n valeurs X(w) de X pour les n individus de la population peuvent être considérées comme les coordonnées d'un vecteur de $\mathbb{R}^n$. Ce vecteur est noté encore $X = \begin{pmatrix} x_1 \\ \vdots \\ x_i \\ \vdots \\ x_n \end{pmatrix}$

Même chose pour les n valeurs Y(w) de Y pour les n individus de la population.

L'espace $E = \mathbb{R}^n$ apparait alors comme l'**espace des variables.**

Dans cet espace des variables, la matrice $D_{\frac{1}{n}} = \frac{1}{n} I_n$, o˘ I_n est la matrice unité à n lignes et n colonnes, définit un **produit scalaire** :

$$\langle X|Y\rangle_{D_{\frac{1}{n}}} = \langle X|D_{\frac{1}{n}}|Y\rangle = \sum_{i=1}^{i=n} \frac{1}{n} x_i y_i = \frac{1}{n} \sum_{i=1}^{i=n} x_i y_i = \frac{1}{n} \langle X|Y\rangle$$

en notant $\langle X|Y\rangle$ le produit scalaire canonique de $\mathbb{R}^n$. On note $1_n = \begin{pmatrix} 1 \\ \vdots \\ 1 \\ \vdots \\ 1 \end{pmatrix}$ le **vecteur unité** de $\mathbb{R}^n$. Le vecteur unité est normé.

la moyenne $\overline{X}$ de la variable statistique X est donné par :

$$\overline{X} = \frac{1}{n}\sum_{w} X(w) = \frac{1}{n}\sum_{i=1}^{i=n} x_i = \frac{1}{n}\sum_{i=1}^{i=n} x_i \times 1 = \langle X|D_{\frac{1}{n}}|1_n\rangle = \langle X|1$$

La moyenne de X est le produit scalaire de X par le vecteur unité 1_n.
Notons X_0 la variable centrée correspondant à X : pour chaque individu w de la population, sa valeur est $X(w) - \overline{X}$:

$$X_0 = \begin{pmatrix} x_1 - \overline{X} \\ \vdots \\ x_i - \overline{X} \\ \vdots \\ x_n - \overline{X} \end{pmatrix} = \begin{pmatrix} x_1 \\ \vdots \\ x_i \\ \vdots \\ x_n \end{pmatrix} - \overline{X} \begin{pmatrix} 1 \\ \vdots \\ 1 \\ \vdots \\ 1 \end{pmatrix} = X - \overline{X} 1_n$$

$$X = X_0 + \overline{X} 1_n = X_0 + \langle X | 1_n \rangle D_{\frac{1}{n}} 1_n$$

$$s^2(X) = \overline{X_0^2} = \frac{1}{n}\sum_{i=1}^{i=n}(x_i - \overline{X})^2 = \langle X_0 | D_{\frac{1}{n}} | X_0 \rangle = \|X_0\|^2$$

$$s^2(X) = \|X_0\|^2$$

La variance de X est le carré de la norme de la variable centrée.

La covariance de deux variables quantitatives réelles X et Y définies sur Ω est la moyenne du produit des variables centrées :

$$Cov(X, Y) = \frac{1}{n}\sum_{i=1}^{i=n}(x_i - \overline{X})(y_i - \overline{Y}) = \langle X_0 | D_{\frac{1}{n}} | Y_0 \rangle = \langle X_0 | Y_0 \rangle D_{\frac{1}{n}}$$

$$Cov(X, Y) = \langle X_0 | D_{\frac{1}{n}} | Y_0 \rangle \langle X_0 | Y_0 \rangle D_{\frac{1}{n}}$$

La covariance est le produit scalare des variables centrées.

$$r_{XY} = \frac{Cov(X, Y)}{s(X)s(Y)} = \frac{\langle X_0 | D_{\frac{1}{n}} | Y_0 \rangle}{\|X_0\|_\phi \|Y_0\|_\phi} = cos(X_0, Y_0)$$
$$r_{XY} = cos(X_0, Y_0)$$

Le coefficient de corrélation linéaire est le cosinus de l'angle des variables centrées.

Soient Y la variable à expliquer, X la variable explicative, $X0$ et Y_0 les variables centrées.
Le prédicteur linéaire $D_{Y|X}$ est $y^* = a + bx$ ou $y^* - \overline{Y} = b(x - \overline{X})$, soit $y_0^* = bx_0$.
Il est représenté par la **droite de régression** de Y en X dans l'espace des individus.
Le coefficient b s'obtient par

$$b = \frac{Cov(X, Y)}{s^2(X)} = \frac{Cov(X, Y)}{s^2(X_0)} = \frac{\langle X_0 | Y_0 \rangle_{D_{\frac{1}{n}}}}{\|X_0\|^2_{D_{\frac{1}{n}}}}.$$

D'après ce qui précède (I.1.2.3.e), $bX_0 = \frac{\langle X_0 | Y_0 \rangle_{D_{\frac{1}{n}}}}{\|X_0\|^2_{D_{\frac{1}{n}}}} X_0$

est le projeté orthogonal de Y_0 sur X_0, $Y_0 - bX_0$ est orthogonal à X_0 et b est la valeur qui minimise l'expression

$$S^2 = \frac{1}{n} \sum_{i=1}^{i=n} (Y_{0i} - bX_{0i})^2 = \|Y_0 - bX_0\|^2_{D_{\frac{1}{n}}} = s^2(Y - bX) =$$
$$s^2(Y - a - bX) = s^2(Y - Y^*) = s^2(Y_0 - Y_0^*)$$

Régression linéaire dans $\mathbb{R}^2$

Prédicteur linéaire (suite)

Le prédicteur linéaire de la variable centrée Y_0 est le projeté orthogonal de Y_0 sur X_0 dans $\mathbb{R}^n$. C'est la variable Y_0^* qui minimise la variance de $Y_0 - Y_0^*$.
Nous avons alors :

$$s^2(Y) = \|Y_0\|^2_{D_{\frac{1}{n}}} = \|Y_0 - bX_0 + bX_0\|^2_{D_{\frac{1}{n}}} = \|Y_0 - bX_0\|^2_{D_{\frac{1}{n}}} + \|bX_0|$$

$$s^2(Y) = S^2_{min} + b^2\|X_0\|^2_{D_{\frac{1}{n}}} = S^2_{min} + (\frac{Cov(X,Y)}{s^2(X)})s^2(X) = S^2_{min} + (\frac{Cov(X,Y)}{s(X)s(Y)})s^2(Y)$$

$$s^2(Y) = S^2_{min} + r^2_{XY}s^2(X).$$

Nous retrouvons la variance résiduelle S^2_{min} et la variance expliquée par la régression $r^2_{XY}s^2(X)$.
De facon symétrique, si X est la variable explicative et Y la variable à expliquer, nous aurons une expression :

$$s^2(X) = S^2_{min} + r^2_{XY}s^2(X).$$

avec la variance résiduelle S'^2_{min} et la variance expliquée par la régression $r^2_{XY}s^2(X)$.