Machine Learning^{4DS}

Notions de base sur l'apprentissage supervisé
REGRESSION

Mohamed Heny SELMI

medheny.selmi@esprit.tn

Enseignant et Responsable option Data Science à ESPRIT

Régression
Linéaire Multiple

PROBLÉMATIQUE – ÉCHANTILLON D'APPRENTISSAGE

Modèle Véhicule	X1 (Frs) Prix	X2 (cm3) Cylindrée	X3 (kW) Puissance	X4 (kg) Poids	Y (l/100km) Consommation
Daihatsu Cuore	11600	846	32	650	5,7
Suzuki Swift 1.0 GLS	12490	993	39	790	5,8
Fiat Panda Mambo L	10450	899	29	730	6,1
VW Polo 1.4 60	17140	1390	44	955	6,5
Opel Corsa 1.2i Eco	14825	1195	33	895	6,8
Subaru Vivio 4WD	13730	658	32	740	6,8
Toyota Corolla	19490	1331	55	1010	7,1
Ferrari 456 GT	285000	5474	325	1690	21,3
Mercedes S [illegible]	183900	5987	300	2250	18,7
Maserati G[illegible]li GT	92500	2789	209	1485	[illegible]
Opel Ast[illegible].6i 16[illegible]	[illegible]	1597	74	[illegible]	7,4
Peuge[illegible]6 X[illegible]108	[illegible]	1761	74	[illegible]	9,0
Renau[illegible] 2.[illegible]V	[illegible]	[illegible]	[illegible]	[illegible]	11,7
Seat I[illegible]o GTI	22500	1983	85	1075	9,5
VW G[illegible].0 GTI	31580	1984	85	1155	9,5
Citroen ZX Volcane	28750	1998	89	1140	8,8
Fiat Tempra 1.6 Liberty	22600	1580	65	1080	9,3
Fort Escort 1.4i PT	20300	1390	54	1110	8,6
Honda Civic Joker 1.4	19900	1396	66	1140	7,7
Volvo 850 2.5	39800	2435	106	1370	10,8
Ford Fiesta 1.2 Zetec	19740	1242	55	940	6,6
Hyundai Son[illegible]a 3000	38990	2972	107	1400	11,7
Lancia K 3.[illegible] LS	50800	2958	[illegible]	[illegible]	11,9
Mazda Ha[illegible]back V	[illegible]	2497	122	[illegible]	10,8
Mitsubishi Galant	31990	1998	66	1300	7,6
Opel Omega 2.5i V6	47700	2496	125	1670	11,3
Peugeot 806 2.0	36950	1998	89	1560	10,8
Nissan Primera 2.0	26950	1997	92	1240	9,2
Seat Alhambra 2.0	36400	1984	85	1635	11,6
Toyota Previa salon	50900	2438	97	1800	12,8
Volvo 960 Kombi aut	49300	2473	125	1570	12,7

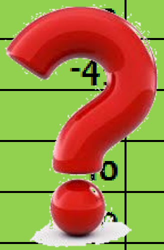
$$f(X_1, X_2, X_3, \dots \dots \dots \dots \dots, X_n) = Y$$

f(Prix, Cylindrée, Puissance, Poids) = **Consommation**

PROBLÉMATIQUE - ÉCHANTILLON D'APPRENTISSAGE

$$f(x_1, x_2, x_3, \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.., x_n) = Y$$

1	MT	RG	PRIX	BR	INV	PUB	FV	TPUB	VENTES
2	369	118	59	9	17	89	177	225	5439
3	476	138	71	18	4	63	279	206	5149
4	432	152	73	16	-50	16	245	309	4704
5	418	135	79	35	[illegible]	74	270	83	5036
6	383	104	60	21	[illegible]	32	201	298	4110
7	554	138	81	20	[illegible]	93	324	161	6180
8	320	147	66	15	[illegible]	48	154	305	4888
9	268	129	57	29	[illegible]	51	166	263	4290
10	359	106	69	27	71	74	196	414	5397
11	461	132	82	27	-18	91	267	170	5272
12	420	136	70	10	8	91	213	429	4989
13	536	111	73	27	128	74	296	273	5927
14	311	143	67	22	-25	27	[illegible]	[illegible]	4033
[illegible]	[illegible]	[illegible]	74	[illegible]	27	75	[illegible]	345	6124
[illegible]	332	140	60	11	61	21	180	247	4708
17	336	136	60	25	-30	40	213	328	4627
18	394	146	59	13	143	52	209	407	4872
19	415	148	69	8	47	29	207	80	5151



$$f(MT, RG, PRIX, BR, INV, PIB, FV, TPUB) = VENTES$$

Objectifs

- Trouver le meilleur modèle (linéaire) liant Y et X
- Qualifier la liaison par rapport à chaque X_i
- Comparer les modèles de Prédiction : globale ou réduit
- Détecter les individus atypiques

$$\mathcal{R}(X_1, X_2, \ldots\ldots\ldots\ldots\ldots, X_n) = Y$$

Objectifs de la régression linéaire

- ✓ Le modèle de prédiction *LINEAIRE* consiste à prédire la valeur d'une variable cible continue, en fonction des valeurs d'un certain nombre d'autres variables prédictives

- ✓ Cette variable « cible » peut être par exemple :
  - le poids : en fonction de la taille
  - le prix d'un appartement : en fonction de sa superficie, de l'étage et du quartier
  - la consommation d'électricité : en fonction de la température extérieure et de l'épaisseur de l'isolation

Exemple de régression linéaire

$$Y = f(X_1, X_2, X_3, \ldots, X_n)$$

Coût du loyer

Nombre de pièces

Services offerts
(piscine, stationnement intérieur, etc.)

L'étage dans l'immeuble

...

Estimer le coût du loyer en fonction :

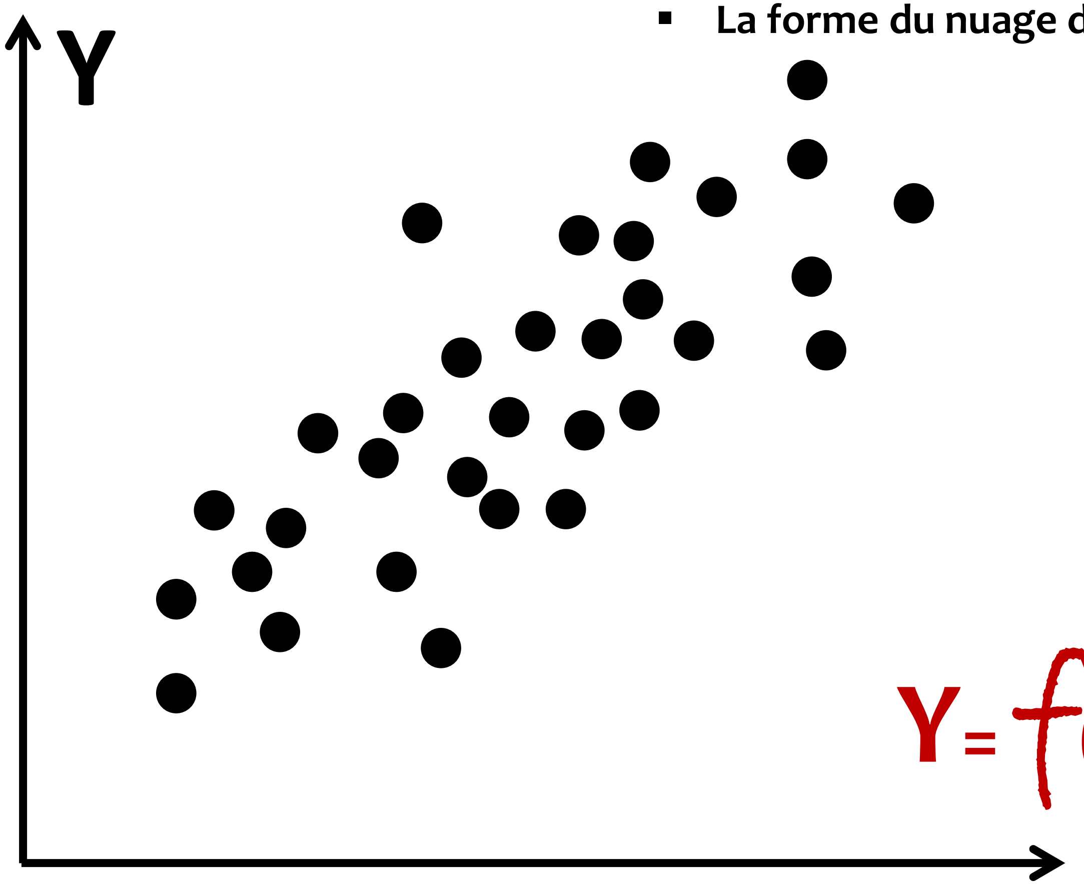
du nombre de pièces,

du niveau d'étage dans l'immeuble,

des services offerts ...

Représentation par un nuage de points

- N couples de points (x_i, y_i)
- X_i : quantitatives
- Y : quantitative
- La forme du nuage de points suit la nature des variables



X_i

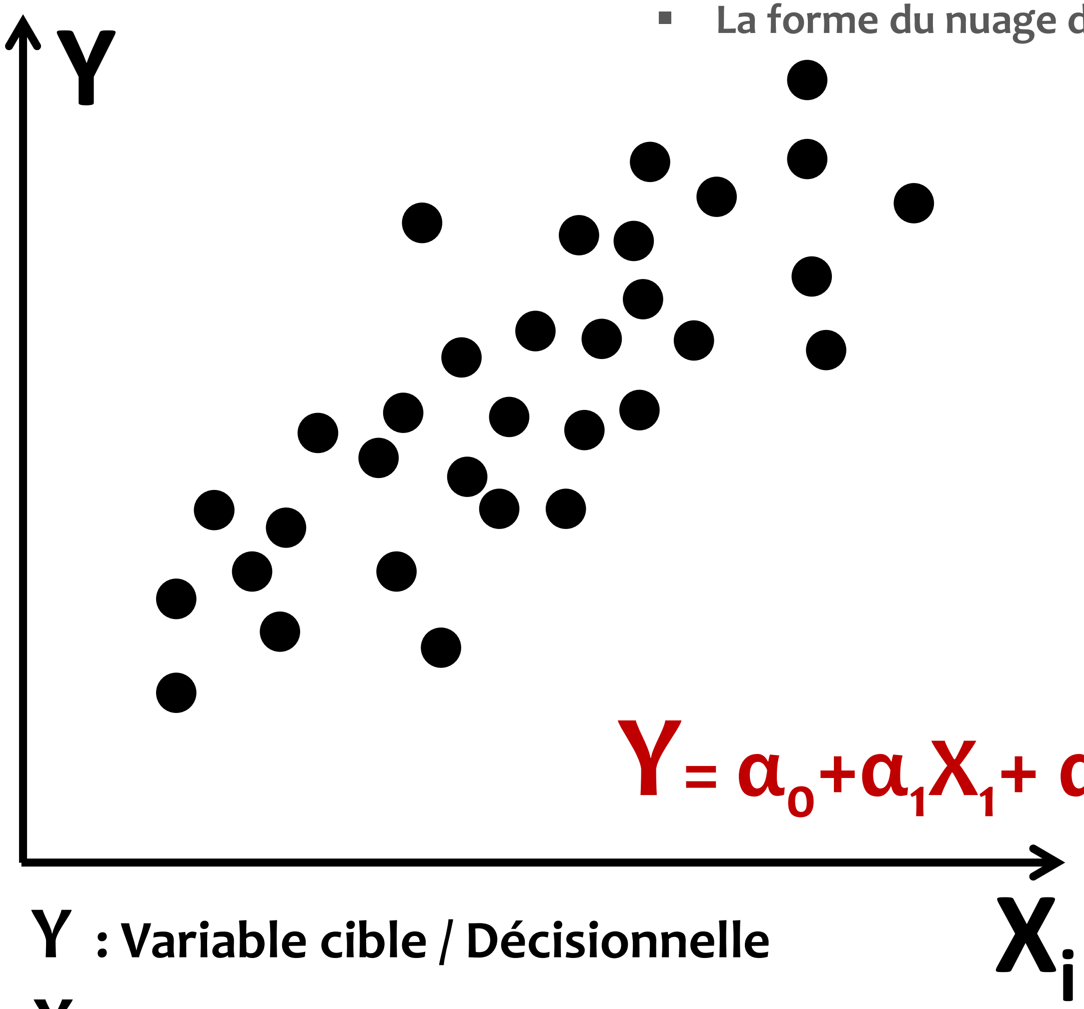
$$Y = f(X_1, X_2, \ldots\ldots\ldots, X_n)$$

Y : Variable cible / Décisionnelle

X_i : Variables prédictives

Représentation par un nuage de points

- N couples de points (x_i, y_i)
- Xi : quantitatives
- Y : quantitative
- La forme du nuage de points suit la nature des variables



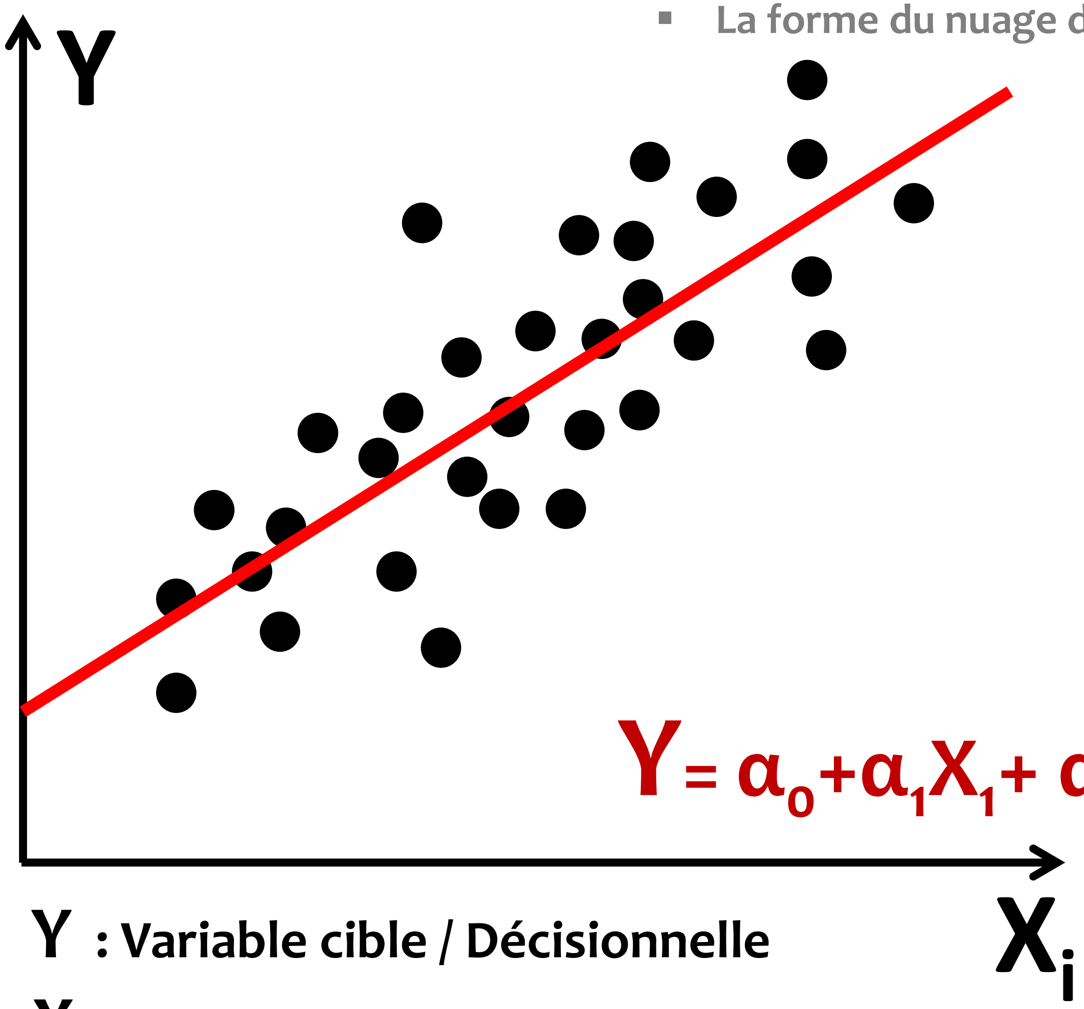
$$Y = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \dots\dots\dots + \alpha_n X_n$$

Y : Variable cible / Décisionnelle

X_i : Variables prédictives

Représentation par un nuage de points

- N couples de points (x_i, y_i)
- Xi : quantitatives
- Y : quantitative
- La forme du nuage de points suit la nature des variables



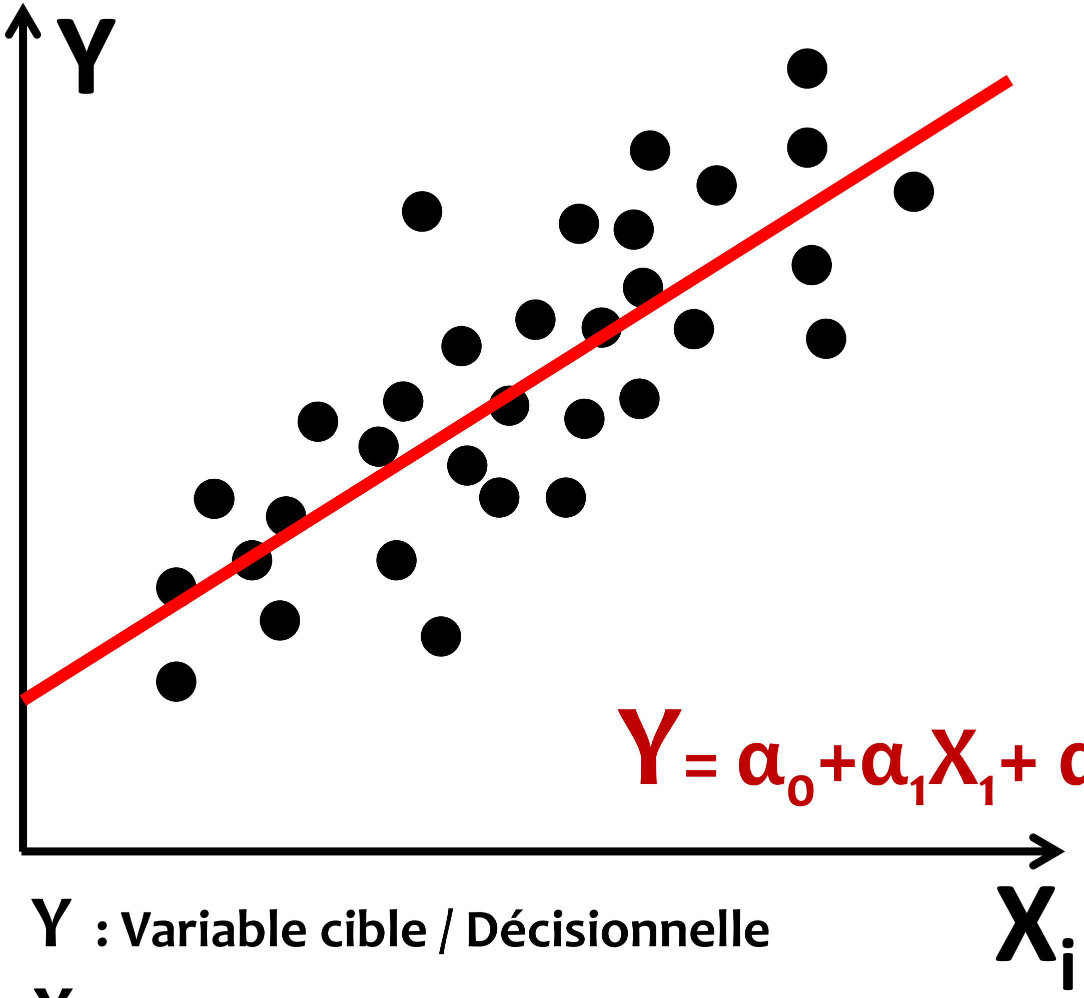
$$Y = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \ldots\ldots\ldots + \alpha_n X_n$$

Y : Variable cible / Décisionnelle

X_i : Variables prédictives

Représentation par un nuage de points

- La droite qui représente mieux les données
- La droite qui résume le mieux le nuage des points
- La droite qui explique mieux les Y en fonctions des X_i



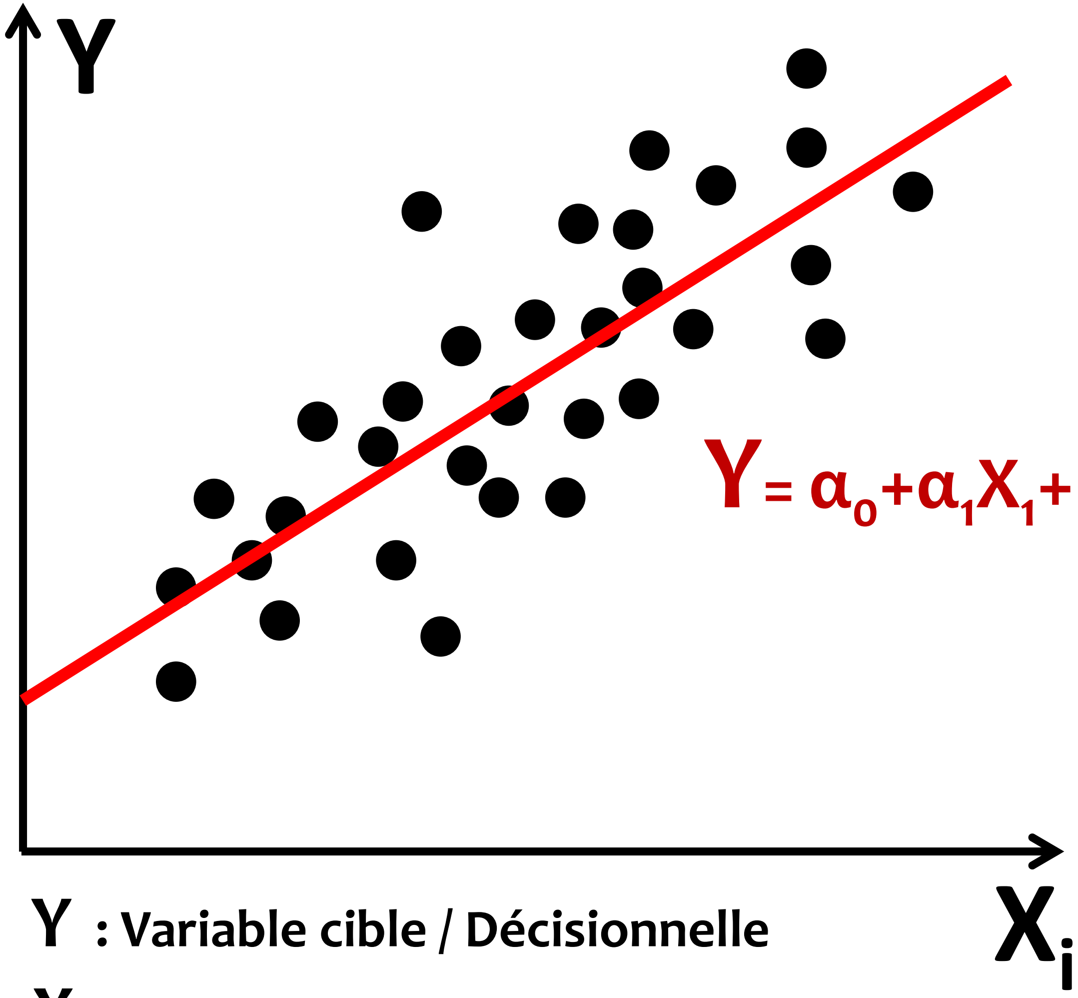
$$Y = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \ldots\ldots\ldots + \alpha_n X_n$$

Y : Variable cible / Décisionnelle

X_i : Variables prédictives

Méthodes des moindres carrés

- La droite qui représente mieux les données
- La droite qui résume le mieux le nuage des points
- La droite qui explique mieux les Y en fonctions des X_i



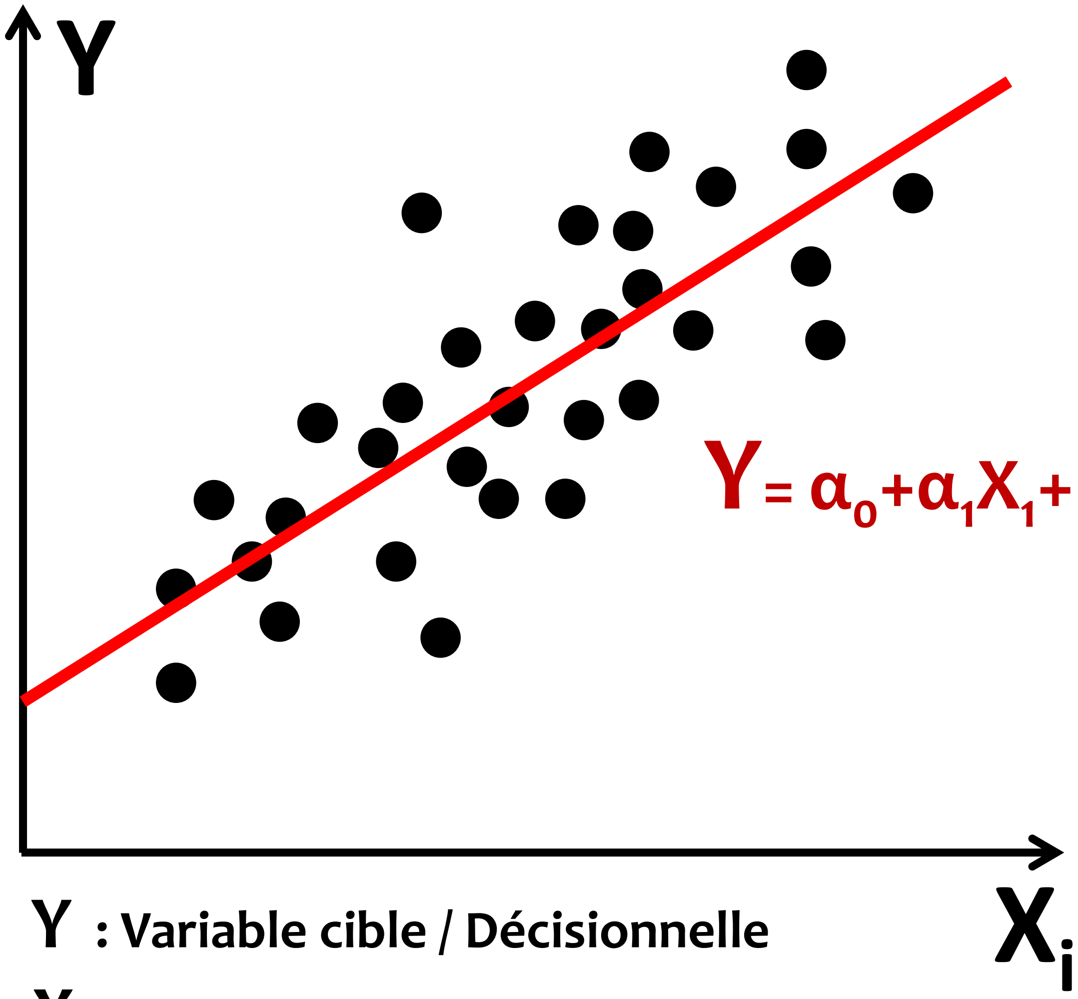
$$Y = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \ldots\ldots\ldots + \alpha_n X_n + \varepsilon$$

Y : Variable cible / Décisionnelle

X_i : Variables prédictives

Méthodes des moindres carrés

- la droite dont les points du nuage sont en moyenne les plus proches
- la droite qui passe à la plus faible distance de chaque point du nuage



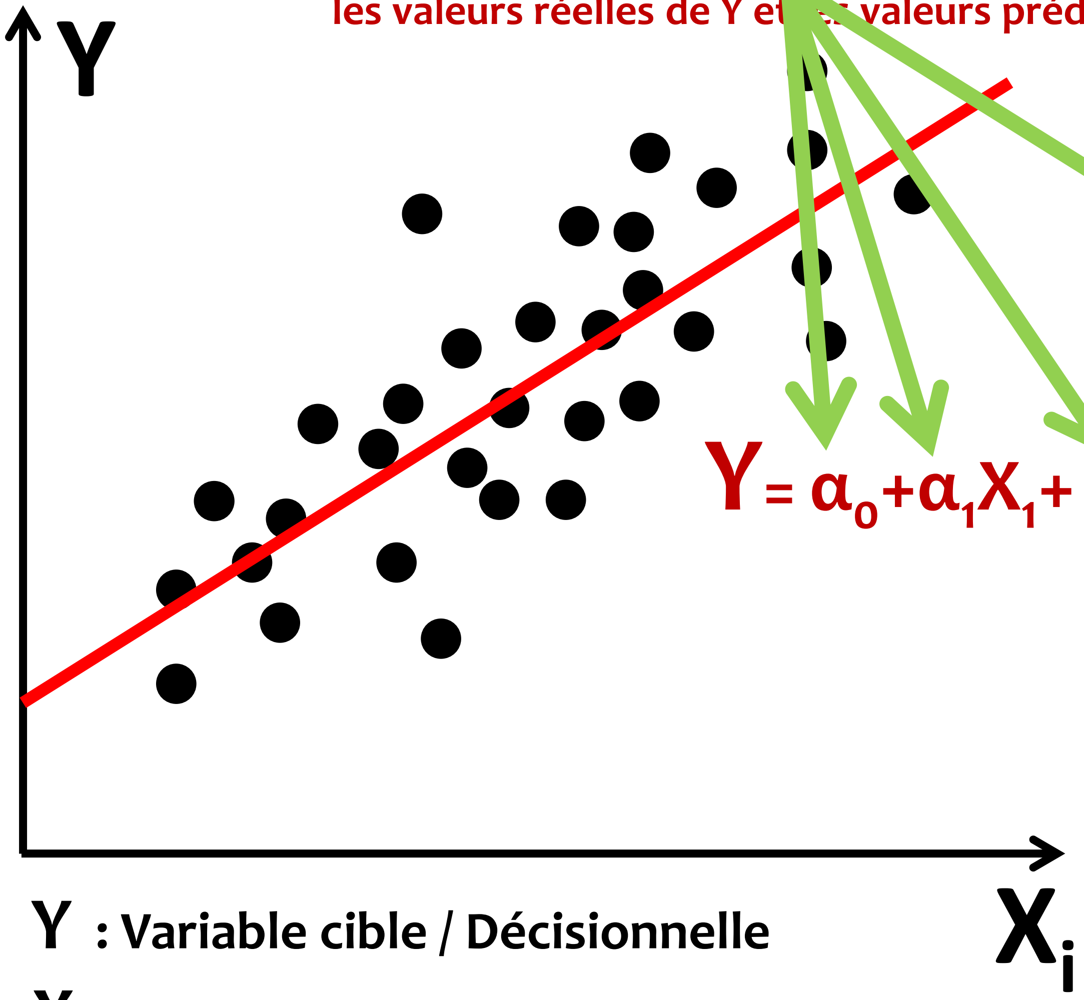
$$Y = \alpha_0+\alpha_1X_1+ \alpha_2X_2+\ldots\ldots\ldots+ \alpha_nX_n+\varepsilon$$

Y : Variable cible / Décisionnelle

X_i : Variables prédictives

Méthodes des moindres carrés

- la droite dont les points du nuage sont en moyenne les plus proches
- la droite qui passe à la plus faible distance de chaque point du nuage
- **Trouver les valeurs des α_i qui minimise la somme des carrés des écarts entre les valeurs réelles de Y et les valeurs prédites avec le modèle de prédiction**



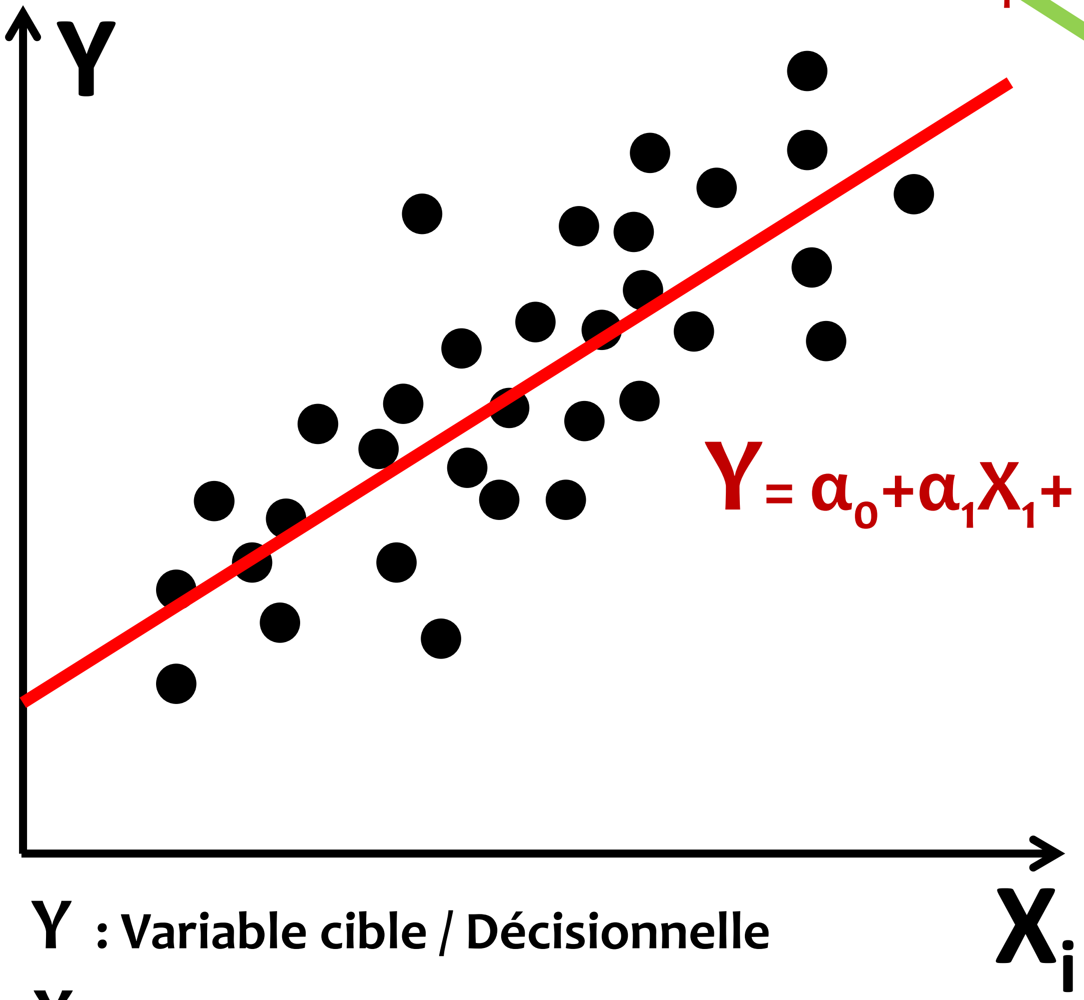
$$Y = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \dots\dots\dots + \alpha_n X_n + \varepsilon$$

Y : Variable cible / Décisionnelle

X_i : Variables prédictives

Méthodes des moindres carrés

- la droite dont les points du nuage sont en moyenne les plus proches
- la droite qui passe à la plus faible distance de chaque point du nuage
- **Trouver les valeurs des α_i qui <u>minimise</u> la somme des carrés des écarts entre les valeurs réelles de Y et les valeurs prédites avec le modèle de prédiction**



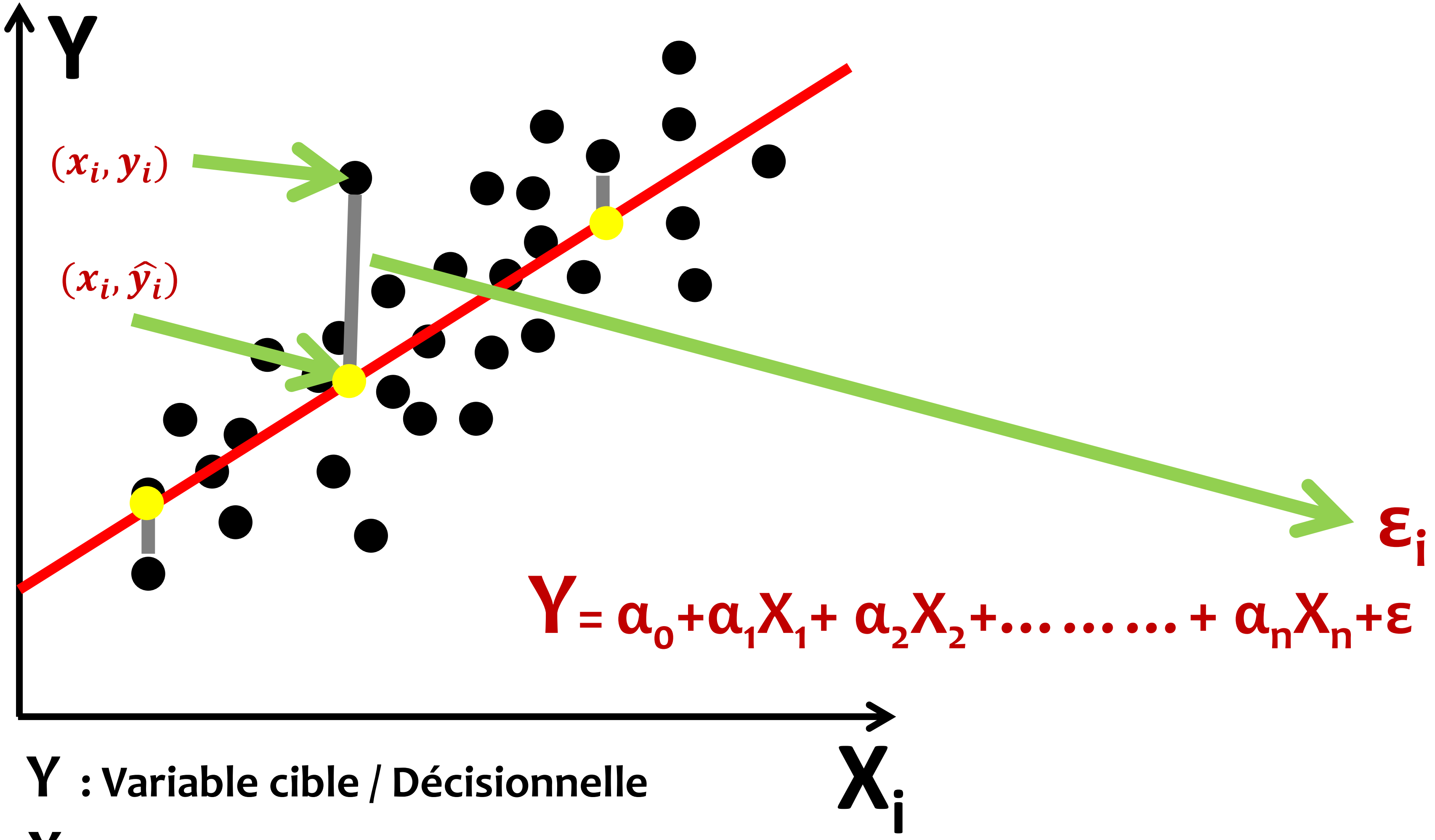
$$Y = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \ldots\ldots\ldots + \alpha_n X_n + \varepsilon$$

Y : Variable cible / Décisionnelle

X_i : Variables prédictives

Estimation par la méthode des moindres carrés

- **La distance d'un point à la droite est la distance verticale entre l'ordonnée du point observé (x_i, y_i) et l'ordonnée du point correspondant sur la droite $(x_i, \hat{y}_i)$**
- Trouver les valeurs des α_i qui minimise la somme des carrés des écarts entre les valeurs réelles de Y et les valeurs prédites avec le modèle de prédiction



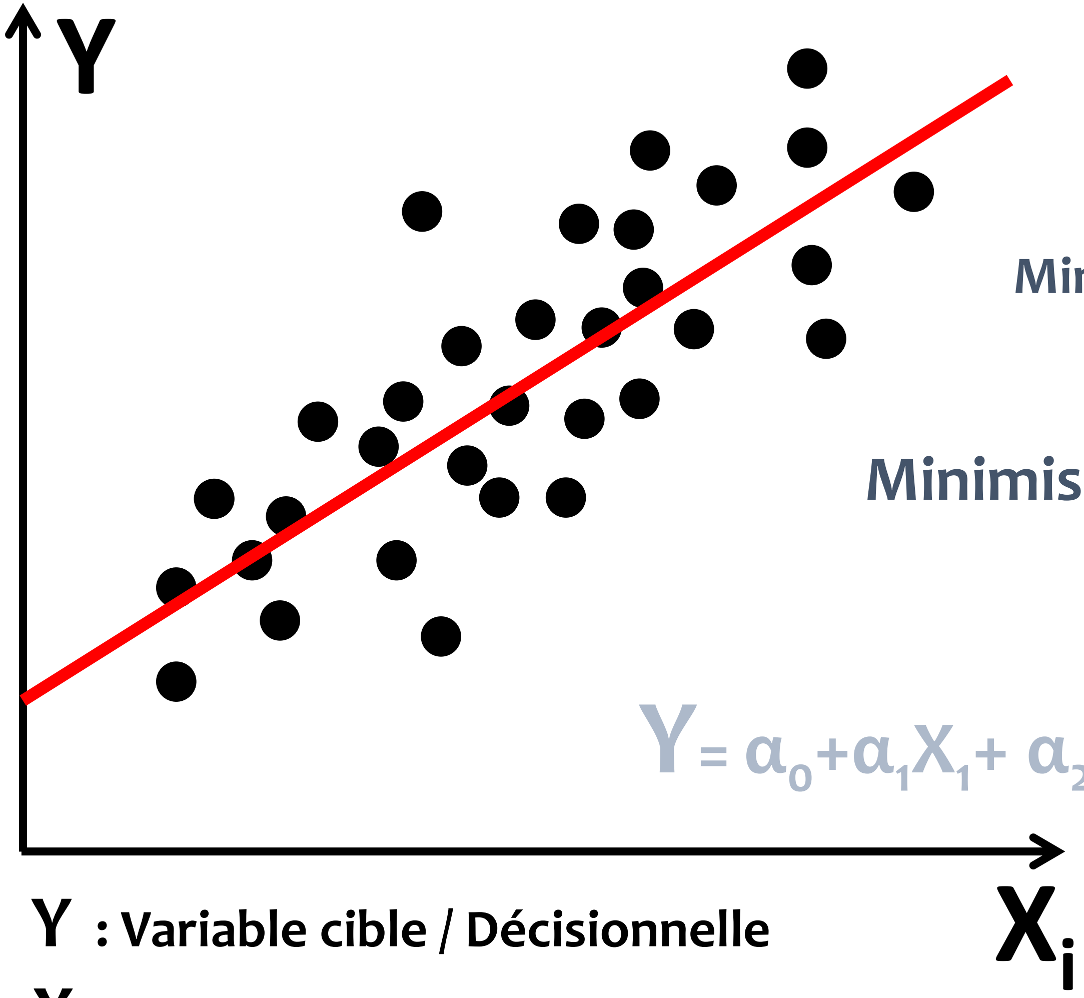
$$Y = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \ldots\ldots\ldots + \alpha_n X_n + \varepsilon$$

Y : Variable cible / Décisionnelle

X_i : Variables prédictives

Objectifs de la méthode des moindres carrés

- La distance d'un point à la droite est la distance verticale entre l'ordonnée du point observé (x_i, y_i) et l'ordonnée du point correspondant sur la droite $(x_i, \hat{y}_i)$
- **Minimiser toutes les erreurs => minimiser les ε_i**



Minimiser $\varepsilon_i = y_i - \widehat{y_i}$

Minimiser $\sum \varepsilon_i{}^2 = \sum (y_i - \hat{y}_i)^2$

$Y = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \ldots\ldots\ldots + \alpha_n X_n + \varepsilon$

Y : Variable cible / Décisionnelle

X_i : Variables prédictives

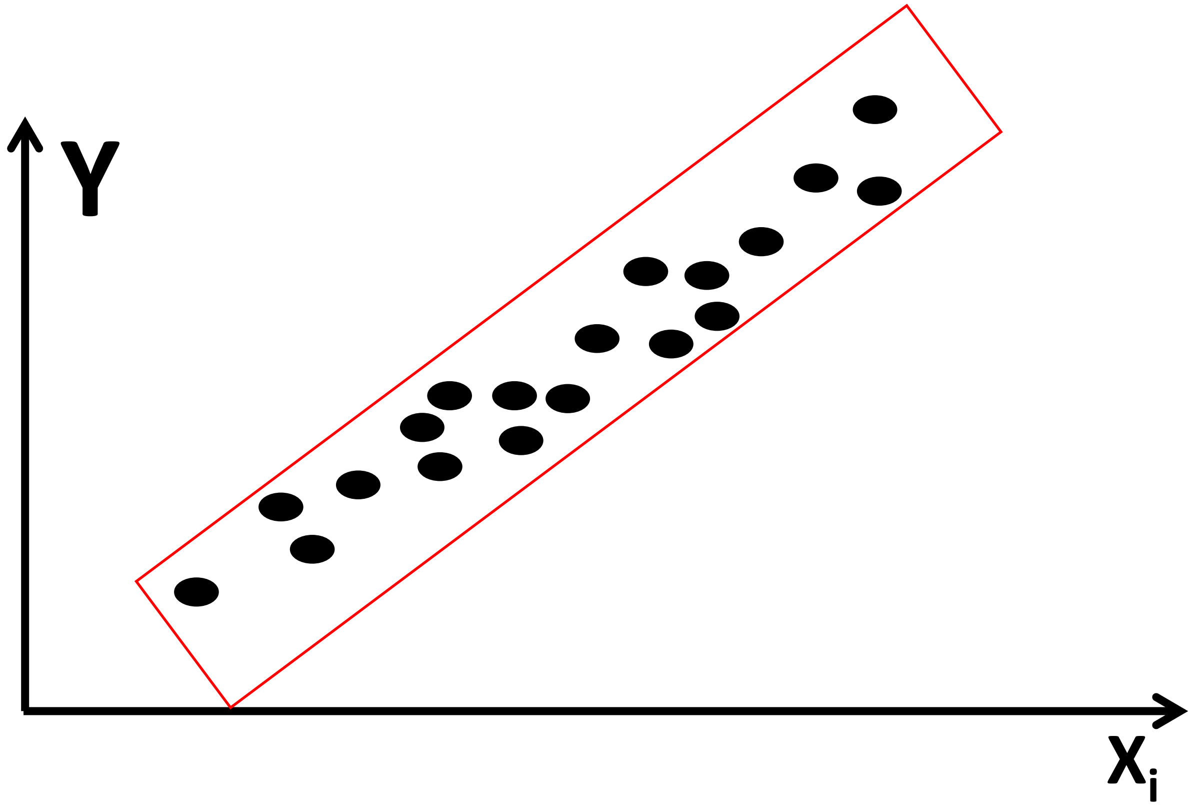
Diagramme Quantile-Quantile
QQ-plot

le diagramme Quantile-Quantile ou diagramme Q-Q ou Q-Q plot est un outil graphique permettant d'évaluer la pertinence de l'ajustement d'une distribution donnée à un modèle théorique.

Le terme de quantile-quantile provient du fait que l'on compare la position de certains quantiles dans la population observée avec leurs positions dans la population théorique.

Le diagramme quantile-quantile permet également de comparer deux distributions que l'on estime semblables.

Diagramme Quantile-Quantile
QQ-plot



Recours à l' écriture matricielle

$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{1,1} & \cdots & x_{1,p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1,n} & \cdots & x_{n,p} \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Entrepôt d'apprentissage

VENTES
5439
5149
4704
5036
4110
6180
4888
4290
5397
5272
4989
5927
4033
6124
4708
4627
4872
5151

1	MT	RG	PRIX	BR	INV	PUB	FV	TPUB
2	369	118	59	9	17	89	177	225
3	476	138	71	18	4	63	279	206
4	432	152	73	16	-50	16	245	309
5	418	135	79	35	142	74	270	83
6	383	104	60	21	-45	32	201	298
7	554	138	81	20	42	93	324	161
8	320	147	66	15	10	48	154	305
9	268	129	57	29	89	51	166	263
10	359	106	69	27	71	74	196	414
11	461	132	82	27	-18	91	267	170
12	420	136	70	10	8	91	213	429
13	536	111	73	27	128	74	296	273
14	311	143	67	22	-25	27	181	60
15	517	142	74	27	27	75	307	345
16	332	140	60	11	61	21	180	247
17	336	136	60	25	-30	40	213	328
18	394	146	59	13	143	52	209	407
19	415	148	69	8	47	29	207	80

Les termes de l'équation

$$y_i = \alpha_0 + \alpha_1 x_{1i} + \alpha_2 x_{2i} + \ldots + \alpha_p x_{pi} + \varepsilon_i$$

$i^{\text{ème}}$ observation de Y

Terme constant

Influence de la variable X_1

Influence de la variable X_p

Résidu de la $i^{\text{ème}}$ observation

Recours à l’ écriture matricielle

$$\boxed{\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}} = \begin{bmatrix} 1 & \boxed{\begin{matrix} x_{1,1} & \cdots & x_{1,p} \\ \vdots & \ddots & \vdots \\ x_{1,n} & \cdots & x_{n,p} \end{matrix}} \\ \vdots & \\ 1 & \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Entrepôt d’apprentissage

SI $\begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix} \rightarrow 0$ **ALORS**

$$\left\{ \begin{array}{l} s = \sum \varepsilon_i{}^2 = \sum (y_i - \hat{y}_i)^2 \\ = \sum (y_i - \alpha x_i \; - \; \alpha_0 \,)^2 \\ \\ \dfrac{\partial s}{\partial \alpha} = 0 \; et \; \dfrac{\partial s}{\partial \alpha_0} = 0 \end{array} \right.$$

Recours à l’ écriture matricielle

$$\begin{bmatrix} y_1 \\ \mathrm{M} \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{1,1} & \Lambda & x_{1,p} \\ \mathrm{M} & \mathrm{M} & \mathrm{O} & \mathrm{M} \\ 1 & x_{1,n} & \Lambda & x_{n,p} \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \mathrm{M} \\ \alpha_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \mathrm{M} \\ \varepsilon_n \end{bmatrix}$$

Entrepôt d’apprentissage

SI $\begin{bmatrix} \varepsilon_1 \\ \mathrm{M} \\ \varepsilon_n \end{bmatrix} \rightarrow 0$ **ALORS**

$$\frac{\partial s}{\partial \alpha} = 0 \; et \; \frac{\partial s}{\partial \alpha_0} = 0$$

$$\sum x_i y_i - \alpha \sum x_i{}^2 - \alpha_0 \, \bar{x} = 0$$

$$et \; \bar{y} - \alpha \bar{x} - \alpha_0 = 0$$

Mohamed Heny SELMI ©

Recours à l' écriture matricielle

$$\begin{bmatrix} y_1 \\ \mathrm{M} \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{1,1} & \Lambda & x_{1,p} \\ \mathrm{M} & \mathrm{M} & \mathrm{O} & \mathrm{M} \\ 1 & x_{1,n} & \Lambda & x_{n,p} \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \mathrm{M} \\ \alpha_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \mathrm{M} \\ \varepsilon_n \end{bmatrix}$$

Entrepôt d'apprentissage

SI $\begin{bmatrix} \varepsilon_1 \\ \mathrm{M} \\ \varepsilon_n \end{bmatrix} \rightarrow 0$ **ALORS**

$$\sum x_i y_i - \alpha \sum x_i{}^2 - \alpha_0\, \bar{x} = 0$$

$$et\ \bar{y} - \alpha\bar{x} - \alpha_0 = 0$$

$$\hat{\alpha} = \frac{\sum(y_i - \bar{y})(x_i - \bar{x})}{\sum(x_i - \bar{x})^2}$$

$$\widehat{\alpha_0} = \bar{y} - \hat{\alpha}\bar{x}$$

Mohamed Heny SELMI ©

Recours à l' écriture matricielle

$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{1,1} & \cdots & x_{1,p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1,n} & \cdots & x_{n,p} \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Entrepôt d'apprentissage

SI $\begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix} \rightarrow 0$ **ALORS**

$$\left\{ \begin{array}{c} \boldsymbol{Y = X\,\alpha} \\ X^tY = X^tX\,\alpha \\ X^tY = [X^tX]\,\alpha \\ [X^tX]^{-1}X^tY = [X^tX]^{-1}\,[X^tX]\,\alpha \\ [X^tX]^{-1}X^tY = \alpha \\ [X^tX]^{-1}X^tY = \hat{\alpha} \end{array} \right.$$

Recours à l’ écriture matricielle

$$\boxed{\begin{bmatrix} y_1 \\ M \\ y_n \end{bmatrix}} = \begin{bmatrix} 1 & \boxed{\begin{matrix} x_{1,1} & \Lambda & x_{1,p} \\ M & O & M \\ x_{1,n} & \Lambda & x_{n,p} \end{matrix}} \\ M & \\ 1 & \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ M \\ \alpha_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ M \\ \varepsilon_n \end{bmatrix}$$

Entrepôt d’apprentissage

SI $\begin{bmatrix} \varepsilon_1 \\ M \\ \varepsilon_n \end{bmatrix} \to 0$ **ALORS**

2

$$\left\{ \begin{aligned} Y &= X\,\alpha \\ X^tY &= X^tX\,\alpha \\ X^tY &= [X^tX]\,\alpha \\ [X^tX]^{-1}X^tY &= [X^tX]^{-1}\,[X^tX]\,\alpha \\ [X^tX]^{-1}X^tY &= \alpha \\ [X^tX]^{-1}X^tY &= \hat{\alpha} \end{aligned} \right.$$

Recours à l’ écriture matricielle

$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{1,1} & \cdots & x_{1,p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1,n} & \cdots & x_{n,p} \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Entrepôt d’apprentissage

SI $\begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix} \rightarrow 0$ **ALORS**

2

$$\left\{ \begin{array}{c} \boldsymbol{Y = X\,\alpha} \\ \boldsymbol{X^tY = X^tX\,\alpha} \\ \boldsymbol{X^tY = [X^tX]\,\alpha} \\ [X^tX]^{-1}X^tY = [X^tX]^{-1}\,[X^tX]\,\alpha \\ [X^tX]^{-1}X^tY = \alpha \\ [X^tX]^{-1}X^tY = \hat{\alpha} \end{array} \right.$$

Recours à l' écriture matricielle

$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{1,1} & \cdots & x_{1,p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1,n} & \cdots & x_{n,p} \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Entrepôt d'apprentissage

SI $\begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix} \rightarrow 0$ **ALORS**

2

$$Y = X\,\alpha$$
$$X^tY = [illegible]\,\alpha$$
$$X^tY = [illegible]\,\alpha$$
$$[X^tX]^{-1}X^tY = [X^tX]^{-1}\,[X^tX]\,\alpha$$
$$[X^tX]^{-1}X^tY = \alpha$$
$$[X^tX]^{-1}X^tY = \hat{\alpha}$$

Recours à l' écriture matricielle

$$\boxed{\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}} = \begin{bmatrix} 1 & \boxed{\begin{matrix} x_{1,1} & \cdots & x_{1,p} \\ \vdots & \ddots & \vdots \\ x_{1,n} & \cdots & x_{n,p} \end{matrix}} \\ \vdots & \\ 1 & \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Entrepôt d'apprentissage

SI $\begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix} \to 0$ **ALORS**

$$\left\{ \begin{aligned} Y &= X\,\alpha \\ X^tY &= X^tX\,\alpha \\ X^tY &= [X^tX]\,\alpha \\ [X^tX]^{-1}X^tY &= [X^tX]^{-1}\,[X^tX]\,\alpha \\ [X^tX]^{-1}X^tY &= \alpha \\ [X^tX]^{-1}X^tY &= \hat{\alpha} \end{aligned} \right.$$

Mohamed Heny SELMI ©

Recours à l' écriture matricielle

$$\boxed{\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}} = \begin{bmatrix} 1 & \boxed{\begin{matrix} x_{1,1} & \cdots & x_{1,p} \\ \vdots & \ddots & \vdots \\ x_{1,n} & \cdots & x_{n,p} \end{matrix}} \\ \vdots & \\ 1 & \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Entrepôt d'apprentissage

SI $\begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix} \rightarrow 0$ **ALORS**

2

$$\left\{ \begin{aligned} Y &= X\,\alpha \\ X^tY &= X^tX\,\alpha \\ X^tY &= [X^tX]\,\alpha \\ [X^tX]^{-1}X^tY &= [X^tX]^{-1}\,[X^tX]\,\alpha \\ [X^tX]^{-1}X^tY &= \alpha \\ [X^tX]^{-1}X^tY &= \hat{\alpha} \end{aligned} \right.$$

Recours à l’ écriture matricielle

$$\boxed{\begin{bmatrix} y_1 \\ M \\ y_n \end{bmatrix}} = \left[\begin{matrix} 1 \\ M \\ 1 \end{matrix}\;\boxed{\begin{matrix} x_{1,1} & \Lambda & x_{1,p} \\ M & O & M \\ x_{1,n} & \Lambda & x_{n,p} \end{matrix}}\right] \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ M \\ \alpha_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ M \\ \varepsilon_n \end{bmatrix}$$

Entrepôt d’apprentissage

<u>SI</u> $\begin{bmatrix} \varepsilon_1 \\ M \\ \varepsilon_n \end{bmatrix} \rightarrow 0$ **<u>ALORS</u>**

$$\begin{cases} \boldsymbol{Y = X\,\alpha} \\ \boldsymbol{X^tY = X^tX\,\alpha} \\ \boldsymbol{X^tY = [X^tX]\,\alpha} \\ \boldsymbol{[X^tX]^{-1}X^tY = [X^tX]^{-1}\,[X^tX]\,\alpha} \\ \boldsymbol{[X^tX]^{-1}X^tY = \alpha} \\ \boldsymbol{[X^tX]^{-1}X^tY = \hat{\alpha}} \end{cases}$$

Mohamed Heny SELMI ©

Recours à l' écriture matricielle

$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{1,1} & \cdots & x_{1,p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1,n} & \cdots & x_{n,p} \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Entrepôt d'apprentissage

SI $\begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix} \rightarrow 0$ **ALORS**

Si on exige que ε ne contient plus de l'information X

$Y - X\alpha$ a pu absorber l'information des X contenue dans Y

Inter – Indépendance entre X et $Y - X\alpha$

X et $Y - X\alpha$ sont géométriquement orthogonaux

$$\langle X | Y - X\alpha \rangle = 0$$

Recours à l' écriture matricielle

$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{1,1} & \cdots & x_{1,p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1,n} & \cdots & x_{n,p} \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Entrepôt d'apprentissage

SI $\begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix} \to 0$ **ALORS**

Si on exige que ε ne contient plus de l'information X

$Y - X\alpha$ a pu absorber l'information des X contenue dans Y

Inter – Indépendance entre X et $Y - X\alpha$

X et $Y - X\alpha$ sont géométriquement orthogonaux

$$\langle X | Y - X\alpha \rangle = 0$$

Recours à l' écriture matricielle

$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{1,1} & \cdots & x_{1,p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1,n} & \cdots & x_{n,p} \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Entrepôt d'apprentissage

SI $\begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix} \to 0$ **ALORS**

3

Si on exige que ε ne contient plus de l'information X

$Y - X\alpha$ **a pu absorber l'information des X contenue dans Y**

Inter – Indépendance entre X et $Y - X\alpha$

$X \; et \; Y - X\alpha$ sont géométriquement orthogonaux

$\langle X | Y - X\alpha \rangle = 0$

Mohamed Heny SELMI ©

Recours à l’ écriture matricielle

$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{1,1} & \cdots & x_{1,p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1,n} & \cdots & x_{n,p} \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Entrepôt d’apprentissage

SI $\begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix} \to 0$ **ALORS**

3

Si on exige que ε ne contient plus de l’information X

$Y - X\alpha$ a pu absorber l’information des X contenue dans Y

Inter – Indépendance entre X et $Y - X\alpha$

$X \text{ et } Y - X\alpha$ sont géométriquement orthogonaux

$$\langle X | Y - X\alpha \rangle = 0$$

Recours à l' écriture matricielle

$$\begin{bmatrix} y_1 \\ M \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{1,1} & \Lambda & x_{1,p} \\ M & M & O & M \\ 1 & x_{1,n} & \Lambda & x_{n,p} \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ M \\ \alpha_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ M \\ \varepsilon_n \end{bmatrix}$$

Entrepôt d'apprentissage

SI $\begin{bmatrix} \varepsilon_1 \\ M \\ \varepsilon_n \end{bmatrix} \rightarrow 0$ **ALORS**

Si on exige que ε ne contient plus de l'information X

$Y - X\alpha$ a pu absorber l'information des X contenue dans Y

Inter – Indépendance entre X et $Y - X\alpha$

$X\ et\ Y - X\alpha$ sont géométriquement orthogonaux

$$\langle \boldsymbol{X} | \boldsymbol{Y} - \boldsymbol{X\alpha} \rangle = \mathbf{0}$$

Recours à l’ écriture matricielle

$$\boxed{\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}} = \begin{bmatrix} 1 & x_{1,1} & \cdots & x_{1,p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1,n} & \cdots & x_{n,p} \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Entrepôt d’apprentissage

<u>**SI**</u> $\begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix} \to 0$ <u>**ALORS**</u> $\left\{ \langle X | Y - X\alpha \rangle = 0 \right.$

Mohamed Heny SELMI ©

Recours à l' écriture matricielle

$$\begin{bmatrix} y_1 \\ M \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{1,1} & \Lambda & x_{1,p} \\ M & M & O & M \\ 1 & x_{1,n} & \Lambda & x_{n,p} \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ M \\ \alpha_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ M \\ \varepsilon_n \end{bmatrix}$$

Entrepôt d'apprentissage

SI $\begin{bmatrix} \varepsilon_1 \\ M \\ \varepsilon_n \end{bmatrix} \rightarrow 0$ **ALORS**

$$\langle \boldsymbol{X} | \boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\alpha} \rangle = \boldsymbol{0}$$

$$X^t(Y - X\alpha) = 0$$

$$X^tY - X^tX\alpha = 0$$

$$X^tY = X^tX\alpha$$

$$X^tY = [X^tX]\alpha$$

$$[X^tX]^{-1}X^tY = [X^tX]^{-1}\,[X^tX]\,\alpha$$

$$[X^tX]^{-1}X^tY = \alpha$$

$$[X^tX]^{-1}X^tY = \hat{\alpha}$$

Recours à l’ écriture matricielle

$$\boxed{\begin{bmatrix} y_1 \\ M \\ y_n \end{bmatrix}} = \begin{bmatrix} 1 & \boxed{\begin{matrix} x_{1,1} & \Lambda & x_{1,p} \\ M & O & M \\ x_{1,n} & \Lambda & x_{n,p} \end{matrix}} \\ M & \\ 1 & \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ M \\ \alpha_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ M \\ \varepsilon_n \end{bmatrix}$$

Entrepôt d’apprentissage

SI $\begin{bmatrix} \varepsilon_1 \\ M \\ \varepsilon_n \end{bmatrix} \to 0$ **ALORS**

$$\langle \boldsymbol{X} | \boldsymbol{Y} - \boldsymbol{X\alpha} \rangle = \boldsymbol{0}$$
$$\boldsymbol{X^t}(\boldsymbol{Y} - \boldsymbol{X\alpha}) = \boldsymbol{0}$$
$$X^tY - X^tX\alpha = 0$$
$$X^tY = X^tX\alpha$$
$$X^tY = [X^tX]\alpha$$
$$[X^tX]^{-1}X^tY = [X^tX]^{-1}\,[X^tX]\,\alpha$$
$$[X^tX]^{-1}X^tY = \alpha$$
$$[X^tX]^{-1}X^tY = \hat{\alpha}$$

Mohamed Heny SELMI ©

Recours à l’ écriture matricielle

$$\boxed{\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}} = \left[\begin{matrix} 1 \\ \vdots \\ 1 \end{matrix}\ \boxed{\begin{matrix} x_{1,1} & \cdots & x_{1,p} \\ \vdots & \ddots & \vdots \\ x_{1,n} & \cdots & x_{n,p} \end{matrix}}\right] \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Entrepôt d’apprentissage

SI $\begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix} \to 0$ **ALORS**

$$\langle X | Y - X\alpha \rangle = 0$$
$$X^t(Y - X\alpha) = 0$$
$$X^tY - X^tX\alpha = 0$$
$$X^tY = X^tX\alpha$$
$$X^tY = [X^tX]\alpha$$
$$[X^tX]^{-1}X^tY = [X^tX]^{-1}\,[X^tX]\,\alpha$$
$$[X^tX]^{-1}X^tY = \alpha$$
$$[X^tX]^{-1}X^tY = \hat{\alpha}$$

Recours à l' écriture matricielle

$$\boxed{\begin{bmatrix} y_1 \\ \mathrm{M} \\ y_n \end{bmatrix}} = \begin{bmatrix} 1 & \boxed{\begin{matrix} x_{1,1} & \Lambda & x_{1,p} \\ \mathrm{M} & \mathrm{O} & \mathrm{M} \\ x_{1,n} & \Lambda & x_{n,p} \end{matrix}} \\ \mathrm{M} & \\ 1 & \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \mathrm{M} \\ \alpha_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \mathrm{M} \\ \varepsilon_n \end{bmatrix}$$

Entrepôt d'apprentissage

SI $\begin{bmatrix} \varepsilon_1 \\ \mathrm{M} \\ \varepsilon_n \end{bmatrix} \rightarrow 0$ **ALORS**

3

$$\left\{ \begin{array}{c} \langle X | Y - X\alpha \rangle = 0 \\ X^t(Y - X\alpha) = 0 \\ X^tY - X^tX\alpha = 0 \\ X^tY = X^tX\alpha \\ X^tY = [X^tX]\alpha \\ [X^tX]^{-1}X^tY = [X^tX]^{-1}\,[X^tX]\,\alpha \\ [X^tX]^{-1}X^tY = \alpha \\ [X^tX]^{-1}X^tY = \hat{\alpha} \end{array} \right.$$

Recours à l' écriture matricielle

$$\boxed{\begin{bmatrix} y_1 \\ \mathrm{M} \\ y_n \end{bmatrix}} = \begin{bmatrix} 1 & \boxed{\begin{matrix} x_{1,1} & \Lambda & x_{1,p} \\ \mathrm{M} & \mathrm{O} & \mathrm{M} \\ x_{1,n} & \Lambda & x_{n,p} \end{matrix}} \\ \mathrm{M} & \\ 1 & \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \mathrm{M} \\ \alpha_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \mathrm{M} \\ \varepsilon_n \end{bmatrix}$$

Entrepôt d'apprentissage

SI $\begin{bmatrix} \varepsilon_1 \\ \mathrm{M} \\ \varepsilon_n \end{bmatrix} \rightarrow 0$ **ALORS**

$$\left\{ \begin{aligned} &\langle X | Y - X\alpha \rangle = 0 \\ &X^t(Y - X\alpha) = 0 \\ &X^tY - X^tX\alpha = 0 \\ &X^tY = X^tX\alpha \\ &X^tY = [X^tX]\alpha \\ &[X^tX]^{-1}X^tY = [X^tX]^{-1}\,[X^tX]\,\alpha \\ &[X^tX]^{-1}X^tY = \alpha \\ &[X^tX]^{-1}X^tY = \hat{\alpha} \end{aligned} \right.$$

Recours à l’ écriture matricielle

$$\begin{bmatrix} y_1 \\ M \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{1,1} & \Lambda & x_{1,p} \\ M & M & O & M \\ 1 & x_{1,n} & \Lambda & x_{n,p} \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ M \\ \alpha_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ M \\ \varepsilon_n \end{bmatrix}$$

Entrepôt d’apprentissage

SI $\begin{bmatrix} \varepsilon_1 \\ M \\ \varepsilon_n \end{bmatrix} \rightarrow 0$ **ALORS**

$$\left\{ \begin{array}{c} \langle X | Y - X\alpha \rangle = 0 \\ X^t(Y - X\alpha) = 0 \\ X^tY - X^tX\alpha = 0 \\ X^tY = X^tX\alpha \\ X^tY = [X^tX]\alpha \\ [X^tX]^{-1}X^tY = [X^tX]^{-1}\,[X^tX]\,\alpha \\ [X^tX]^{-1}X^tY = \alpha \\ [X^tX]^{-1}X^tY = \hat{\alpha} \end{array} \right.$$

Recours à l' écriture matricielle

$$\boxed{\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}} = \begin{bmatrix} 1 & x_{1,1} & \cdots & x_{1,p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1,n} & \cdots & x_{n,p} \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Entrepôt d'apprentissage

SI $\begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix} \rightarrow 0$ **ALORS**

3

$$\left\{ \begin{array}{c} \langle X | Y - X\alpha \rangle = 0 \\ X^t(Y - X\alpha) = 0 \\ X^tY - X^tX\alpha = 0 \\ X^tY = X^tX\alpha \\ X^tY = [X^tX]\alpha \\ [X^tX]^{-1}X^tY = [X^tX]^{-1}\,[X^tX]\,\alpha \\ [X^tX]^{-1}X^tY = \alpha \\ [X^tX]^{-1}X^tY = \hat{\alpha} \end{array} \right.$$

Mohamed Heny SELMI ©

Recours à l' écriture matricielle

$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{1,1} & \cdots & x_{1,p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1,n} & \cdots & x_{n,p} \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Entrepôt d'apprentissage

SI $\begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix} \to 0$ **ALORS**

$$\begin{cases} \langle X | Y - X\alpha \rangle = 0 \\ X^t (Y - X\alpha) = 0 \\ X^t Y - X^t X\alpha = 0 \\ X^t Y = X^t X\alpha \\ X^t Y = [X^t X]\alpha \\ [X^t X]^{-1} X^t Y = [X^t X]^{-1} [X^t X]\,\alpha \\ [X^t X]^{-1} X^t Y = \alpha \\ [X^t X]^{-1} X^t Y = \hat{\alpha} \end{cases}$$

Recours à l’ écriture matricielle

$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{1,1} & \cdots & x_{1,p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1,n} & \cdots & x_{n,p} \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Entrepôt d’apprentissage

SI $\begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix} \to 0$ **ALORS**

3

$$\begin{cases} \langle X \mid Y - X\alpha \rangle = 0 \\ X^t(Y - X\alpha) = 0 \\ X^tY - X^tX\alpha = 0 \\ X^tY = X^tX\alpha \\ X^tY = [X^tX]\alpha \\ [X^tX]^{-1}X^tY = [X^tX]^{-1}\,[X^tX]\,\alpha \\ [X^tX]^{-1}X^tY = \alpha \\ [X^tX]^{-1}X^tY = \hat{\alpha} \end{cases}$$

Mohamed Heny SELMI ©

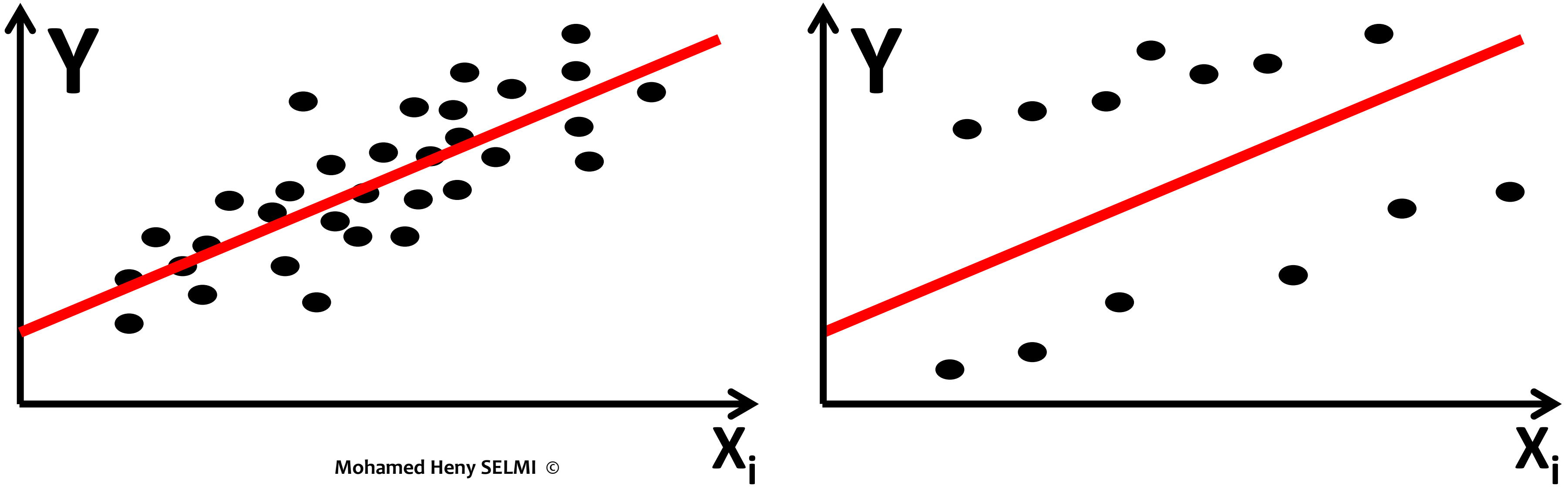
$$[X^t X]^{-1} X^t Y = \hat{\alpha}$$

Les coefficients estimateurs sont d'autant plus précis que :

i. **La variance de l'erreur est faible :**
la droite de régression passe bien au milieu des points

ii. **La dispersion des X est forte :**
les X couvrent bien l'espace de représentation

$\sum \varepsilon_i{}^2 = \sum (y_i - \hat{y}_i)^2$ **est élevé !**



caractéristiques des coefficients estimateurs

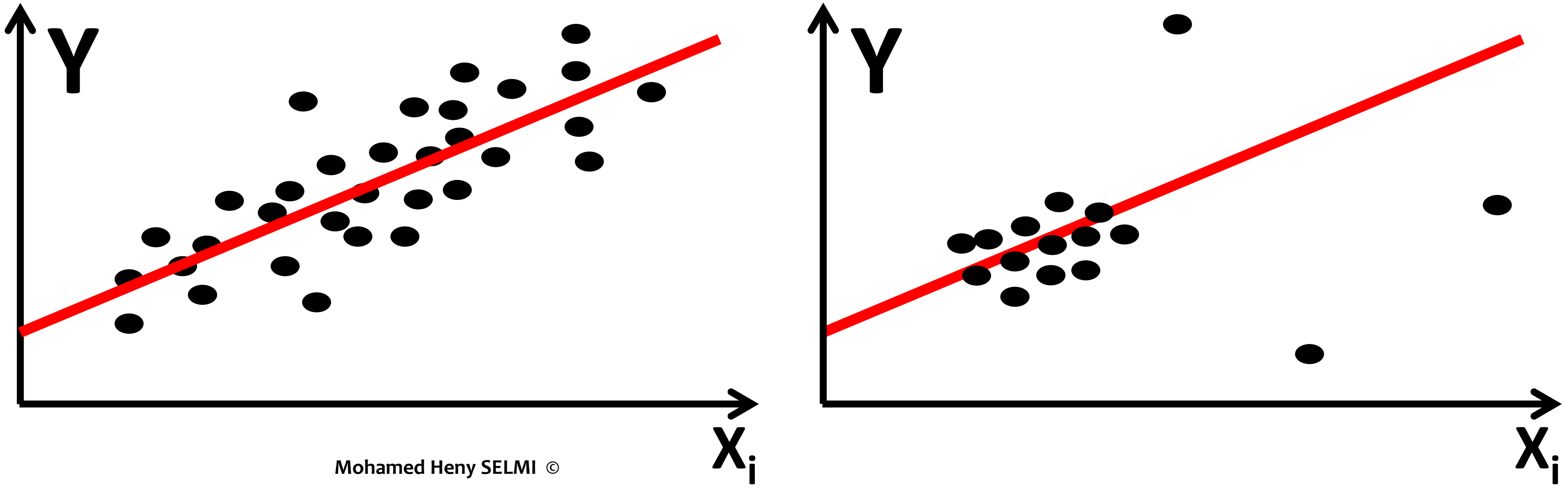
$$[X^t X]^{-1} X^t Y = \hat{\alpha}$$

Les coefficients estimateurs sont d'autant plus précis que :

i. La variance de l'erreur est faible :
la droite de régression passe bien au milieu des points

ii. La dispersion des X est forte :
les X couvrent bien l'espace de représentation

$\sum(x_i - \overline{x})^2$ est faible !



Modèle Global à base de coefficients estimateurs

$$[X^tX]^{-1}X^tY = \hat{\alpha}$$

Y (l/100km) Consommation
5,7
5,8
6,1
6,5
6,8
6,8
7,1
21,3
18,7
14,5
7,4
9,0
11,7
9,5
9,5
8,8
9,3
8,6
7,7
10,8
6,6
11,7
11,9
10,8
7,6
11,3
10,8
9,2
11,6
12,8
12,7

=

X1 (Frs) Prix	X2 (cm3) Cylindrée	X3 (kW) Puissance	X4 (kg) Poids
11600	846	32	650
12490	993	39	790
10450	899	29	730
17140	1390	44	955
14825	1195	33	895
13730	658	32	740
19490	1331	55	1010
285000	5474	325	1690
183900	5987	300	2250
92500	2789	209	1485
25000	1597	74	1080
22350	1761	74	1100
36600	2165	101	1500
22500	1983	85	1075
31580	1984	85	1155
28750	1998	89	1140
22600	1580	65	1080
20300	1390	54	1110
19900	1396	66	1140
39800	2435	106	1370
19740	1242	55	940
38990	2972	107	1400
50800	2958	150	1550
36200	2497	122	1330
31990	1998	66	1300
47700	2496	125	1670
36950	1998	89	1560
26950	1997	92	1240
36400	1984	85	1635
50900	2438	97	1800
49300	2473	125	1570

X

$$\begin{bmatrix} \widehat{\alpha_1} \\ \widehat{\alpha_2} \\ \widehat{\alpha_3} \\ \widehat{\alpha_4} \end{bmatrix}$$

Modèle Global à base de coefficients estimateurs

$$[X^tX]^{-1}X^tY = \hat{\alpha}$$

VENTES
5439
5149
4704
5036
4110
6180
4888
4290
5397
5272
4989
5927
4033
6124
4708
4627
4872
5151

=

1	MT	RG	PRIX	BR	INV	PUB	FV	TPUB
2	369	118	59	9	17	89	177	225
3	476	138	71	18	4	63	279	206
4	432	152	73	16	-50	16	245	309
5	418	135	79	35	142	74	270	83
6	383	104	60	21	-45	32	201	298
7	554	138	81	20	42	93	324	161
8	320	147	66	15	10	48	154	305
9	268	129	57	29	89	51	166	263
10	359	106	69	27	71	74	196	414
11	461	132	82	27	-18	91	267	170
12	420	136	70	10	8	91	213	429
13	536	111	73	27	128	74	296	273
14	311	143	67	22	-25	27	181	60
15	517	142	74	27	27	75	307	345
16	332	140	60	11	61	21	180	247
17	336	136	60	25	-30	40	213	328
18	394	146	59	13	143	52	209	407
19	415	148	69	8	47	29	207	80

X

$$\begin{matrix} \widehat{\alpha_1} \\ \widehat{\alpha_2} \\ \widehat{\alpha_3} \\ . \\ . \\ . \\ \widehat{\alpha_8} \end{matrix}$$

Modèle Global à base de coefficients estimateurs

$$[X^tX]^{-1}X^tY = \hat{\alpha}$$

VENTES
5439
5149
4704
5036
4110
6180
4888
4290
5397
5272
4989
5927
4033
6124
4708
4627
4872
5151

=

1	MT	RG	PRIX	BR	INV	PUB	FV	TPUB
2	369	118	59	9	17	89	177	225
3	476	138	71	18	4	63	279	206
4	432	152	73	16	-50	16	245	309
5	418	135	79	35	142	74	270	83
6	383	104	60	21	-45	32	201	298
7	554	138	81	20	42	93	324	161
8	320	147	66	15	10	48	154	305
9	268	129	57	29	89	51	166	263
10	359	106	69	27	71	74	196	414
11	461	132	82	27	-18	91	267	170
12	420	136	70	10	8	91	213	429
13	536	111	73	27	128	74	296	273
14	311	143	67	22	-25	27	181	60
15	517	142	74	27	27	75	307	345
16	332	140	60	11	61	21	180	247
17	336	136	60	25	-30	40	213	328
18	394	146	59	13	143	52	209	407
19	415	148	69	8	47	29	207	80

X

$\widehat{\alpha_1}$
$\widehat{\alpha_2}$
$\widehat{\alpha_3}$

$\widehat{\alpha_8}$

Critères de sélection de variables pertinentes

AIC

- **Critère d'information d'Akaike**
- la différence entre 2 fois le nombre de paramètres (k) deux fois la log-vraisemblance du modèle estimé.
- $$\boldsymbol{AIC = 2k - 2\ \ln L}$$

BIC

- **critère d'information Bayésien**
- pénalité dépend de la taille de l'échantillon et pas seulement du nombre de paramètres
- $$\boldsymbol{AIB = -2\ln L + k\ln N}$$

Les méthodes pas à pas consistent à considérer d'abord un modèle faisant intervenir toutes les variables explicatives: puis on procède par élimination ou ajout successif de variables.

- la méthode descendante ou élimination en arrière lorsque on élimine des variables

- la méthode ascendante ou sélection en avant lorsque on ajoute des variables

- La méthode stepwise est une combinaison de ces deux méthodes

Exemple : cas de Ventes semestrielles

- Variable à prédire : VENTES = Ventes semestrielles
- Variables prédictives :

MT = Marché total
RG = Remises aux grossistes

PRIX = Prix

BR = Budget de Recherche

INV = Investissement

PUB = Publicité

FV = Frais de ventes

TPUB = Total budget publicité de la branche

	A	B	C	D	E	F	G	H	I
1	MT	RG	PRIX	BR	INV	PUB	FV	TPUB	VENTES
2	369	118	59	9	17	89	177	225	5439
3	476	138	71	18	4	63	279	206	5149
4	432	152	73	16	-50	16	245	309	4704
5	418	135	79	35	142	74	270	83	5036
6	383	104	60	21	-45	32	201	298	4110
7	554	138	81	20	42	93	324	161	6180
8	320	147	66	15	10	48	154	305	4888
9	268	129	57	29	89	51	166	263	4290
10	359	106	69	27	71	74	196	414	5397
11	461	132	82	27	-18	91	267	170	5272
12	420	136	70	10	8	91	213	429	4989
13	536	111	73	27	128	74	296	273	5927
14	311	143	67	22	-25	27	181	60	4033
15	517	142	74	27	27	75	307	345	6124
16	332	140	60	11	61	21	180	247	4708
17	336	136	60	25	-30	40	213	328	4627
18	394	146	59	13	143	52	209	407	4872
19	415	148	69	8	47	29	207	80	5151

Première étape

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.898[a]	.806	.752	256.29

a. Predictors: (Constant), Total publicité de la branche, Marché total, Remises aux grossistes, Budget de recherche, Investissements, Publicité, Prix, Frais de ventes

TPUB = Total budget publicité de la branche

Coefficients[a]

Model		Unstandardized Coefficients		t	Sig.
		B	Std. Error		
1	(Constant)	3129.231	641.355	4.879	.000
	MT	4.423	1.588	2.785	.009
	RG	1.676	3.291	.509	.614
	PRIX	-13.526	8.305	-1.629	.114
	BR	-3.410	6.569	-.519	.608
	INV	1.924	.778	2.474	.019
	PUB	8.547	1.826	4.679	.000
	FV	1.497	2.771	.540	.593
	TPUB	-2.15E-02	.401	-.054	.958

a. Dependent Variable: VENTES

Modèle complet (sans restriction de variables)

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	t	Sig.
1	(Constant)	3129.231	641.355	4.879	.000
	MT	4.423	1.588	2.785	.009
	RG	1.676	3.291	.509	.614
	PRIX	-13.526	8.305	-1.629	.114
	BR	-3.410	6.569	-.519	.608
	INV	1.924	.778	2.474	.019
	PUB	8.547	1.826	4.679	.000
	FV	1.497	2.771	.540	.593
	TPUB	-2.15E-02	.401	-.054	.958

a. Dependent Variable: VENTES

VENTE$_1$ = 3129,231 + 4,423 X MT + 1,676 X RG - 13,526 X PRIX – 3,410 X BR +1,924 X INV + 8,328 X PUB + 1,497 X FV – 0,00215 X TPUB

deuxième étape

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.898[a]	.806	.760	251.99

a. Predictors: (Constant), Frais de ventes, Remises aux grossistes, Publicité, Investissements, Budget de recherche, Prix, Marché total

b. Dependent Variable: Ventes

BR = Budget de Recherche

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	t	Sig.
1	(Constant)	3115.648	579.517	5.376	.000
	MT	4.426	1.561	2.836	.008
	RG	1.706	3.191	.535	.597
	PRIX	-13.445	8.029	-1.675	.104
	BR	-3.392	6.451	-.526	.603
	INV	1.931	.756	2.554	.016
	PUB	8.558	1.784	4.798	.000
	FV	1.482	2.710	.547	.588

a. Dependent Variable: VENTES

Troisième étape

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.897[a]	.804	.766	249.04

a. Predictors: (Constant), Frais de ventes, Remises aux grossistes, Publicité, Investissements, Prix, Marché total

b. Dependent Variable: Ventes

FV = Frais de ventes

Coefficients[a]

Model		Unstandardized Coefficients		t	Sig.
		B	Std. Error		
1	(Constant)	3137.547	571.233	5.493	.000
	MT	4.756	1.412	3.368	.002
	RG	1.705	3.153	.541	.593
	PRIX	-14.790	7.521	-1.966	.058
	INV	1.885	.742	2.539	.016
	PUB	8.519	1.761	4.837	.000
	FV	.950	2.484	.382	.705

a. Dependent Variable: VENTES

Quatrième étape

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.896[a]	.803	.772	245.69

a. Predictors: (Constant), Publicité, Remises aux grossistes, Marché total, Investissements, Prix

b. Dependent Variable: Ventes

RG = Remises aux grossistes

Coefficients[a]

Model		Unstandardized Coefficients		t	Sig.
		B	Std. Error		
1	(Constant)	3084.009	546.374	5.645	.000
	MT	5.222	.704	7.415	.000
	RG	1.700	3.111	.546	.589
	PRIX	-13.467	6.589	-2.044	.049
	INV	1.984	.686	2.893	.007
	PUB	8.328	1.666	4.998	.000

a. Dependent Variable: VENTES

Modèle avec sélection de variables

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	t	Sig.
1	(Constant)	3084.009	546.374	5.645	.000
	MT	5.222	.704	7.415	.000
	PRIX	-13.467	6.589	-2.044	.049
	INV	1.984	.686	2.893	.007
	PUB	8.328	1.666	4.998	.000

a. Dependent Variable: VENTES

$$VENTE_2 = 3084{,}009 + 5{,}222 \text{ X } MT - 13{,}467 \text{ X } PRIX + +1{,}984 \text{ X } INV + 8{,}328 \text{ X } PUB$$

PRÉDICTION ET ÉVALUATION SUR UN ENTREPÔT DE TEST

Modèle global

1	MT	RG	PRIX	BR	INV	PUB	FV	TPUB	VENTES	pred.full
2	328	123	77	20	59	88	211	141	4787	4722,93
3	285	105	63	8	-28	12	176	218	4123	3646,53
4	441	120	80	16	-22	50	267	405	4801	4847,31
5	462	112	73	15	68	93	283	212	5712	5406,63
6	417	120	81	35	148	83	257	111	5512	5330,98
7	408	131	66	13	120	62	235	141	5313	5138,27
8	362	145	67	23	117	73	220	239	4942	5092,38
9	436	123	73	32	100	43	276	280	5366	5004,36
10	456	128	65	22	144	52	253	93	5741	5401,29
11	364	120	64	14	128	96	195	107	5383	5269,58
12	433	124	68	8	122	25	258	291	5140	4839,73
13	277	135	62	11	76	68	175	410	4842	4476,9
14	455	126	78	22	18	95	233	118	5316	5626,29
15	398	138	56	12	50	77	229	98	5540	5175,03
16	412	149	78	36	30	26	258	124	4647	4752,77
17	415	119	75	20	-40	41	211	315	4630	4844,78
18	484	111	58	13	107	40	258	321	5502	5358,06
19	515	120	77	23	126	21	328	398	5288	5139,33
20	429	125	74	11	88	83	218	118	5095	5452,81
21	355	131	65	24	113	77	208	307	5094	5082,23

PRÉDICTION ET ÉVALUATION SUR UN ENTREPÔT DE TEST

Modèle global

Modèle réduit

1	MT	RG	PRIX	BR	INV	PUB	FV	TPUB	VENTES	pred.full	pred.sel
2	328	123	77	20	59	88	211	141	4787	4722,93	4976,03
3	285	105	63	8	-28	12	176	218	4123	3646,53	4070,95
4	441	120	80	16	-22	50	267	405	4801	4847,31	5139,34
5	462	112	73	15	68	93	283	212	5712	5406,63	5636,19
6	417	120	81	35	148	83	257	111	5512	5330,98	5336,96
7	408	131	66	13	120	62	235	141	5313	5138,27	5100,06
8	362	145	67	23	117	73	220	239	4942	5092,38	4991,94
9	436	123	73	32	100	43	276	280	5366	5004,36	5051,22
10	456	128	65	22	144	52	253	93	5741	5401,29	5226,66
11	364	120	64	14	128	96	195	107	5383	5269,58	5215,42
12	433	124	68	8	122	25	258	291	5140	4839,73	4869,75
13	277	135	62	11	76	68	175	410	4842	4476,9	4556,15
14	455	126	78	22	18	95	233	118	5316	5626,29	5622,77
15	398	138	56	12	50	77	229	98	5540	5175,03	5194,05
16	412	149	78	36	30	26	258	124	4647	4752,77	4782,91
17	415	119	75	20	-40	41	211	315	4630	4844,78	4936,43
18	484	111	58	13	107	40	258	321	5502	5358,06	5243,04
19	515	120	77	23	126	21	328	398	5288	5139,33	5207,94
20	429	125	74	11	88	83	218	118	5095	5452,81	5391,9
21	355	131	65	24	113	77	208	307	5094	5082,23	4997,16

PRÉDICTION ET ÉVALUATION SUR UN ENTREPÔT DE TEST

Modèle global

Modèle réduit

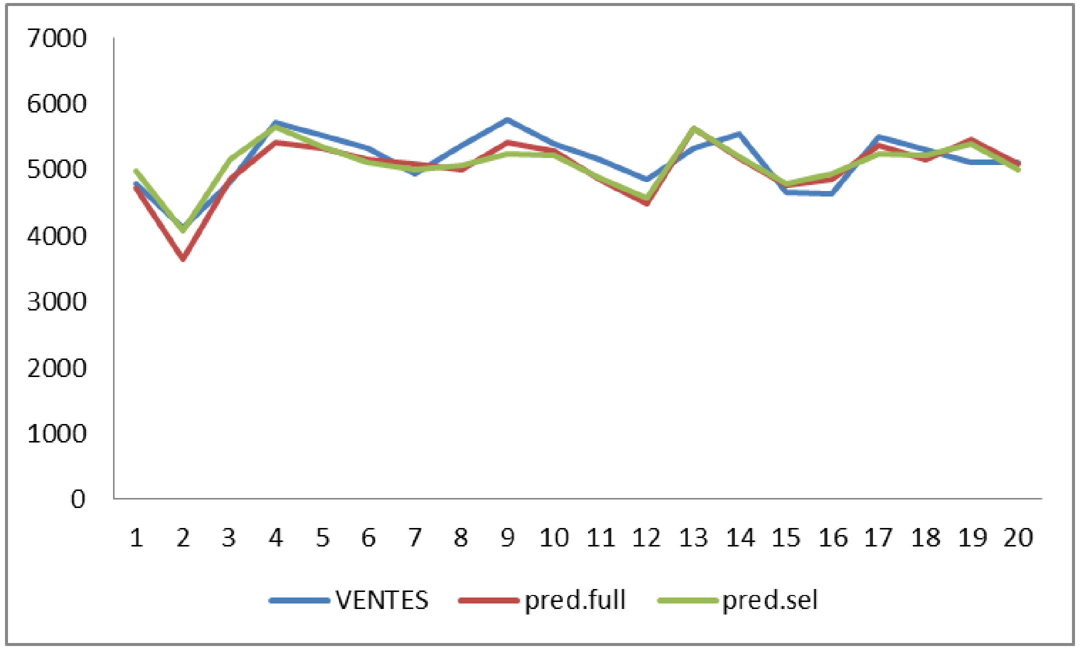
1	MT	RG	PRIX	BR	INV	PUB	FV	TPUB	VENTES	pred.full	pred.sel
2	328	123	77	20	59	88	211	141	4787	4722,93	4976,03
3	285	105	63	8	-28	12	176	218	4123	3646,53	4070,95
4	441	120	80	16	-22	50	267	405	4801	4847,31	5139,34
5	462	112	73	15	68	93	283	212	5712	5406,63	5636,19
6	417	120	81	35	148	83	257	111	5512	5330,98	5336,96
7	408	131	66	13	120	62	235	141	5313	5138,27	5100,06
8	362	145	67	23	117	73	220	239	4942	5092,38	4991,94
9	436	123	73	32	100	43	276	280	5366	5004,36	5051,22
10	456	128	65	22	144	52	253	93	5741	5401,29	5226,66
11	364	120	64	14	128	96	195	107	5383	5269,58	5215,42
12	433	124	68	8	122	25	258	291	5140	4839,73	4869,75
13	277	135	62	11	76	68	175	410	4842	4476,9	4556,15
14	455	126	78	22	18	95	233	118	5316	5626,29	5622,77
15	398	138	56	12	50	77	229	98	5540	5175,03	5194,05
16	412	149	78	36	30	26	258	124	4647	4752,77	4782,91
17	415	119	75	20	-40	41	211	315	4630	4844,78	4936,43
18	484	111	58	13	107	40	258	321	5502	5358,06	5243,04
19	515	120	77	23	126	21	328	398	5288	5139,33	5207,94
20	429	125	74	11	88	83	218	118	5095	5452,81	5391,9
21	355	131	65	24	113	77	208	307	5094	5082,23	4997,16

COMPARAISON DES MODÈLES DE PRÉDICTION

Modèle global

Modèle réduit

VENTES	pred.full	pred.sel
4787	4722,93	4976,03
4123	3646,53	4070,95
4801	4847,31	5139,34
5712	5406,63	5636,19
5512	5330,98	5336,96
5313	5138,27	5100,06
4942	5092,38	4991,94
5366	5004,36	5051,22
5741	5401,29	5226,66
5383	5269,58	5215,42
5140	4839,73	4869,75
4842	4476,9	4556,15
5316	5626,29	5622,77
5540	5175,03	5194,05
4647	4752,77	4782,91
4630	4844,78	4936,43
5502	5358,06	5243,04
5288	5139,33	5207,94
5095	5452,81	5391,9
5094	5082,23	4997,16

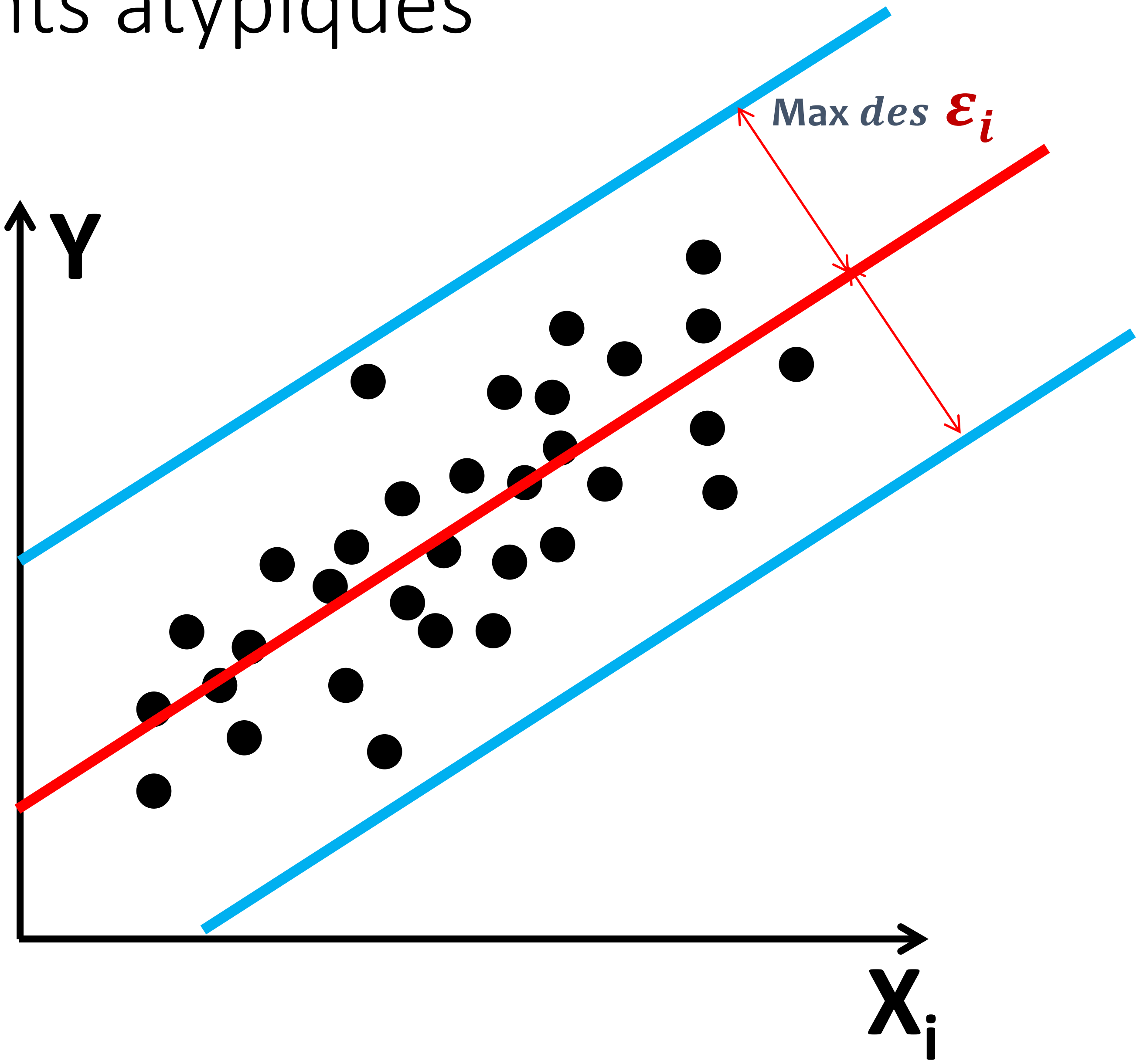


Les points atypiques

- Repérer les observations qui jouent un rôle anormal dans la régression

Atypique (aberrant)	• Elle prend une valeur inhabituelle sur une variable • Elle prend une combinaison de valeurs inhabituelles sur plusieurs variables
Influent	• Elle pèse de manière exagérée dans la régression • les résultats sont très différents selon que le point est pris en compte ou pas dans la régression
Atypique (régression)	• Elle est très mal reconstituée (expliquée) par la régression • le résidu observé est très élevé, le point n'obéit pas à la relation qui a été établie par la régression

Les points atypiques



Régression *Logistique Binaire*

Variable cible à K (K = 2) modalités

PROBLÉMATIQUE

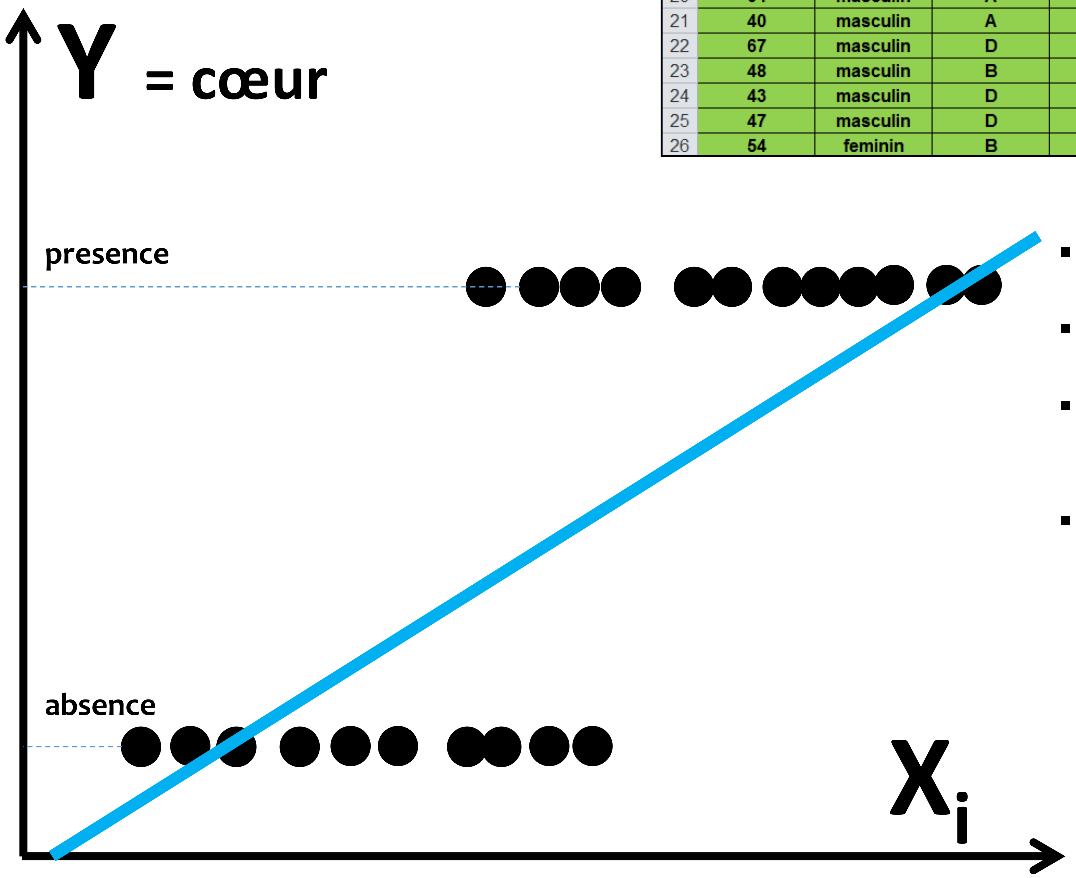
MT	RG	PRIX	BR	INV	PUB	FV	TPUB	VENTES
369	118	59	9	17	89	177	225	5439
476	138	71	18	4	63	279	206	5149
432	152	73	16	-50	16	245	309	4704
418	135	79	35	142	74	270	83	5036
383	104	60	21	-45	32	201	298	4110
554	138	81	20	42	93	324	161	6180
320	147	66	15	10	48	154	305	4888
268	129	57	29	89	51	166	263	4290
359	106	69	27	71	74	196	414	5397
461	132	82	27	-18	91	267	170	5272
420	136	70	10	8	91	213	429	4989
536	111	73	27	128	74	296	273	5927
311	143	67	22	-25	27	181	60	4033
517	142	74	27	27	75	307	345	6124
332	140	60	11	61	21	180	247	4708
336	136	60	25	-30	40	213	328	4627
394	146	59	13	143	52	209	407	4872
415	148	69	8	47	29	207	80	5151

1	age	sexe	typedouleur	sucre	tauxmax	angine	depression	coeur
2	70	masculin	D	A	109	non	24	presence
3	67	feminin	C	A	160	non	16	absence
4	57	masculin	B	A	141	non	3	presence
5	64	masculin	D	A	105	oui	2	absence
6	74	feminin	B	A	121	oui	2	absence
7	65	masculin	D	A	140	non	4	absence
8	56	masculin	C	B	142	oui	6	presence
9	59	masculin	D	A	142	oui	12	presence
10	60	masculin	D	A	170	non	12	presence
11	63	feminin	D	A	154	non	40	presence
12	59	masculin	D	A	161	non	5	absence
13	53	masculin	D	A	111	oui	0	absence
14	44	masculin	C	A	180	non	0	absence
15	61	masculin	A	A	145	non	26	presence
16	57	feminin	D	A	159	non	0	absence
17	71	feminin	D	A	125	non	16	absence
18	46	masculin	D	A	120	oui	18	presence
19	53	masculin	D	B	155	oui	31	presence
20	64	masculin	A	A	144	oui	18	absence
21	40	masculin	A	A	178	oui	14	absence
22	67	masculin	D	A	129	oui	26	presence
23	48	masculin	B	A	180	non	2	absence
24	43	masculin	D	A	181	non	12	absence
25	47	masculin	D	A	143	non	1	absence
26	54	feminin	B	B	159	oui	0	absence

Si la variable à prédire est une variable Binaire ? Peut-on faire une régression linéaire ?

PROBLÉMATIQUE

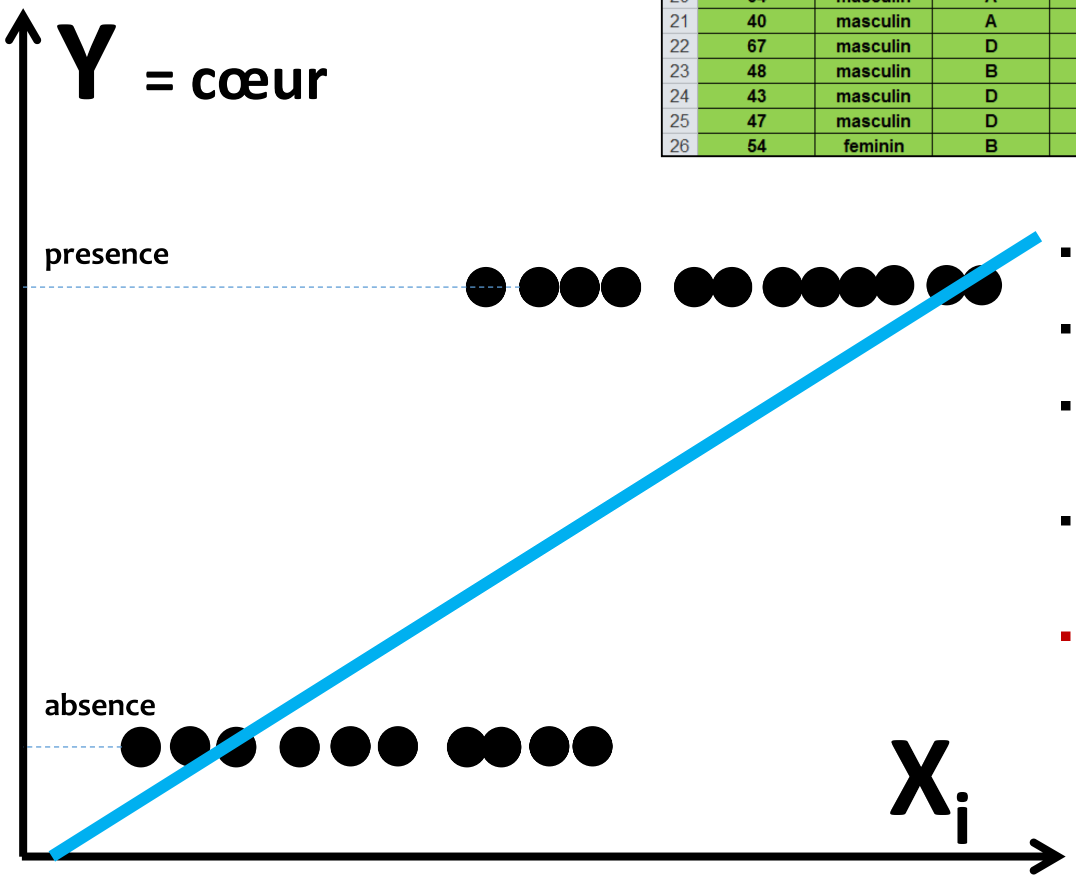
1	age	sexe	typedouleur	sucre	tauxmax	angine	depression	coeur
2	70	masculin	D	A	109	non	24	presence
3	67	feminin	C	A	160	non	16	absence
4	57	masculin	B	A	141	non	3	presence
5	64	masculin	D	A	105	oui	2	absence
6	74	feminin	B	A	121	oui	2	absence
7	65	masculin	D	A	140	non	4	absence
8	56	masculin	C	B	142	oui	6	presence
9	59	masculin	D	A	142	oui	12	presence
10	60	masculin	D	A	170	non	12	presence
11	63	feminin	D	A	154	non	40	presence
12	59	masculin	D	A	161	non	5	absence
13	53	masculin	D	A	111	oui	0	absence
14	44	masculin	C	A	180	non	0	absence
15	61	masculin	A	A	145	non	26	presence
16	57	feminin	D	A	159	non	0	absence
17	71	feminin	D	A	125	non	16	absence
18	46	masculin	D	A	120	oui	18	presence
19	53	masculin	D	B	155	oui	31	presence
20	64	masculin	A	A	144	oui	18	absence
21	40	masculin	A	A	178	oui	14	absence
22	67	masculin	D	A	129	oui	26	presence
23	48	masculin	B	A	180	non	2	absence
24	43	masculin	D	A	181	non	12	absence
25	47	masculin	D	A	143	non	1	absence
26	54	feminin	B	B	159	oui	0	absence



- Visiblement la régression linéaire ne convient pas
- La droite linéaire ne représente pas bien les données
- La droite linéaire n'est pas la meilleure courbe qui résume mieux le nuage de points
- La droite linéaire n'explique pas bien les Y en fonction des X_i

PROBLÉMATIQUE

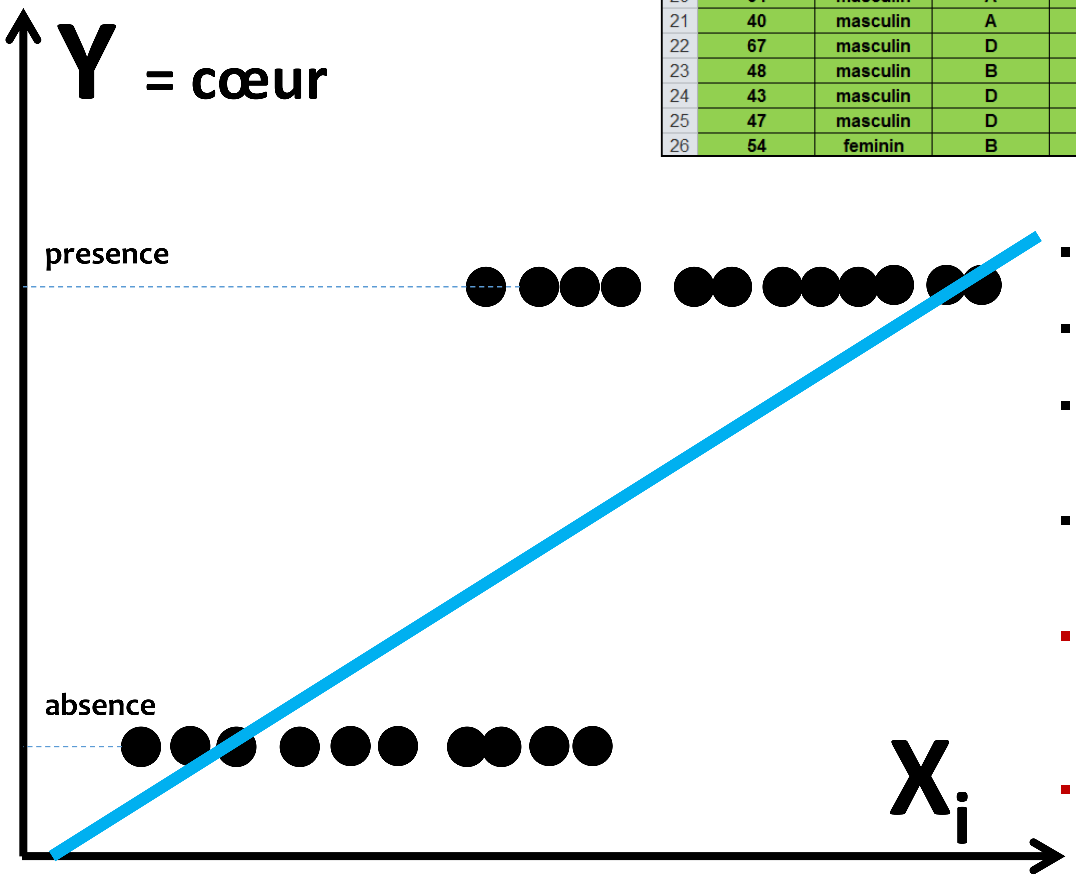
1	age	sexe	typedouleur	sucre	tauxmax	angine	depression	coeur
2	70	masculin	D	A	109	non	24	presence
3	67	feminin	C	A	160	non	16	absence
4	57	masculin	B	A	141	non	3	presence
5	64	masculin	D	A	105	oui	2	absence
6	74	feminin	B	A	121	oui	2	absence
7	65	masculin	D	A	140	non	4	absence
8	56	masculin	C	B	142	oui	6	presence
9	59	masculin	D	A	142	oui	12	presence
10	60	masculin	D	A	170	non	12	presence
11	63	feminin	D	A	154	non	40	presence
12	59	masculin	D	A	161	non	5	absence
13	53	masculin	D	A	111	oui	0	absence
14	44	masculin	C	A	180	non	0	absence
15	61	masculin	A	A	145	non	26	presence
16	57	feminin	D	A	159	non	0	absence
17	71	feminin	D	A	125	non	16	absence
18	46	masculin	D	A	120	oui	18	presence
19	53	masculin	D	B	155	oui	31	presence
20	64	masculin	A	A	144	oui	18	absence
21	40	masculin	A	A	178	oui	14	absence
22	67	masculin	D	A	129	oui	26	presence
23	48	masculin	B	A	180	non	2	absence
24	43	masculin	D	A	181	non	12	absence
25	47	masculin	D	A	143	non	1	absence
26	54	feminin	B	B	159	oui	0	absence



- Visiblement la régression linéaire ne convient pas
- La droite linéaire ne représente pas bien les données
- La droite linéaire n'est pas la meilleure courbe qui résume mieux le nuage de points
- La droite linéaire n'explique pas bien les Y en fonction des X_i
- La résolution : trouver une autre régression dont la forme de sa représentation est plus proche de la nature du nuage de points

PROBLÉMATIQUE

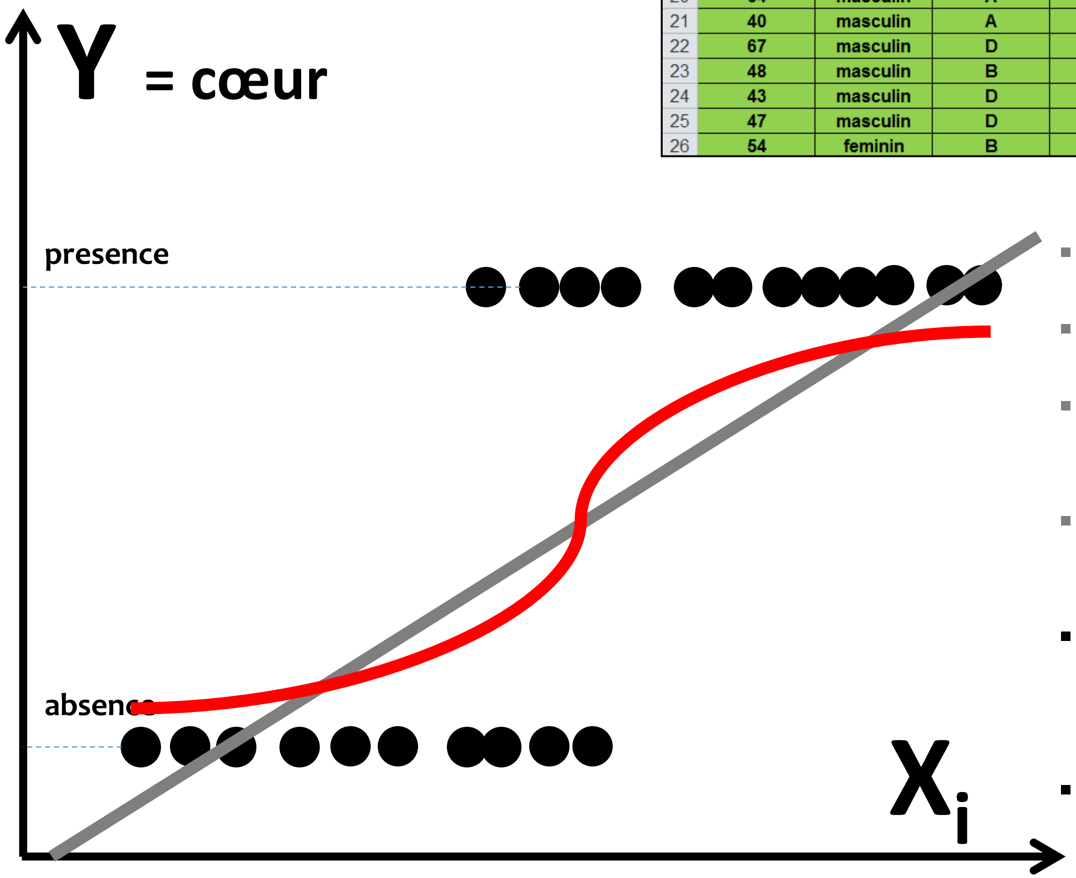
1	age	sexe	typedouleur	sucre	tauxmax	angine	depression	coeur
2	70	masculin	D	A	109	non	24	presence
3	67	feminin	C	A	160	non	16	absence
4	57	masculin	B	A	141	non	3	presence
5	64	masculin	D	A	105	oui	2	absence
6	74	feminin	B	A	121	oui	2	absence
7	65	masculin	D	A	140	non	4	absence
8	56	masculin	C	B	142	oui	6	presence
9	59	masculin	D	A	142	oui	12	presence
10	60	masculin	D	A	170	non	12	presence
11	63	feminin	D	A	154	non	40	presence
12	59	masculin	D	A	161	non	5	absence
13	53	masculin	D	A	111	oui	0	absence
14	44	masculin	C	A	180	non	0	absence
15	61	masculin	A	A	145	non	26	presence
16	57	feminin	D	A	159	non	0	absence
17	71	feminin	D	A	125	non	16	absence
18	46	masculin	D	A	120	oui	18	presence
19	53	masculin	D	B	155	oui	31	presence
20	64	masculin	A	A	144	oui	18	absence
21	40	masculin	A	A	178	oui	14	absence
22	67	masculin	D	A	129	oui	26	presence
23	48	masculin	B	A	180	non	2	absence
24	43	masculin	D	A	181	non	12	absence
25	47	masculin	D	A	143	non	1	absence
26	54	feminin	B	B	159	oui	0	absence



- Visiblement la régression linéaire ne convient pas
- La droite linéaire ne représente pas bien les données
- La droite linéaire n'est pas la meilleure courbe qui résume mieux le nuage de points
- La droite linéaire n'explique pas bien les Y en fonction des X_i
- La résolution : trouver une autre régression dont la forme de sa représentation est plus proche de la nature du nuage de points
- L'estimation des proportions par **régression logistique**

PROBLÉMATIQUE

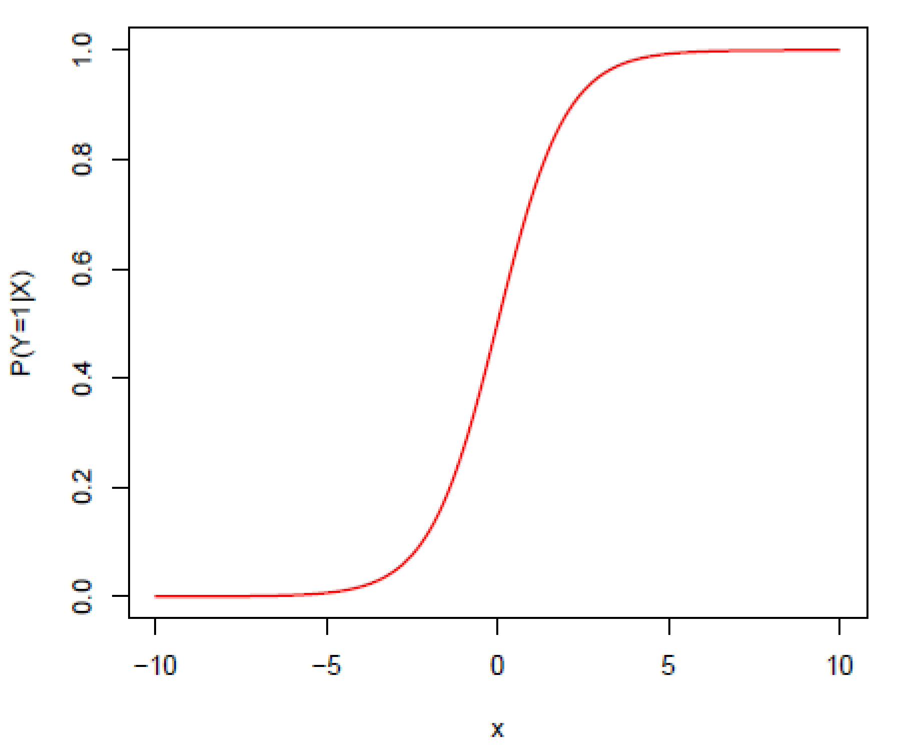
1	age	sexe	typedouleur	sucre	tauxmax	angine	depression	coeur
2	70	masculin	D	A	109	non	24	presence
3	67	feminin	C	A	160	non	16	absence
4	57	masculin	B	A	141	non	3	presence
5	64	masculin	D	A	105	oui	2	absence
6	74	feminin	B	A	121	oui	2	absence
7	65	masculin	D	A	140	non	4	absence
8	56	masculin	C	B	142	oui	6	presence
9	59	masculin	D	A	142	oui	12	presence
10	60	masculin	D	A	170	non	12	presence
11	63	feminin	D	A	154	non	40	presence
12	59	masculin	D	A	161	non	5	absence
13	53	masculin	D	A	111	oui	0	absence
14	44	masculin	C	A	180	non	0	absence
15	61	masculin	A	A	145	non	26	presence
16	57	feminin	D	A	159	non	0	absence
17	71	feminin	D	A	125	non	16	absence
18	46	masculin	D	A	120	oui	18	presence
19	53	masculin	D	B	155	oui	31	presence
20	64	masculin	A	A	144	oui	18	absence
21	40	masculin	A	A	178	oui	14	absence
22	67	masculin	D	A	129	oui	26	presence
23	48	masculin	B	A	180	non	2	absence
24	43	masculin	D	A	181	non	12	absence
25	47	masculin	D	A	143	non	1	absence
26	54	feminin	B	B	159	oui	0	absence



- Visiblement la régression linéaire ne convient pas
- La droite linéaire ne représente pas bien les données
- La droite linéaire n'est pas la meilleure courbe qui résume mieux le nuage de points
- La droite linéaire n'explique pas bien les Y en fonction des X_i
- **La résolution : trouver une autre régression dont la forme de sa représentation est plus proche de la nature du nuage de points**
- **L'estimation des proportions par**

  régression logistique

FONCTION LOGISTIQUE



- ✓ **Un nuage de points dont la variable décisionnelle est une variable qualitative binaire {0,1} ne peut pas être résumer par une droite**

- ✓ **Un tel nuage ne peut être représenté que par une fonction mathématique qui donne une courbe en S :**

- ✓ **Solution :**

Fonction Logistique π

$$\pi(x) = P(Y = 1/\ X) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}$$

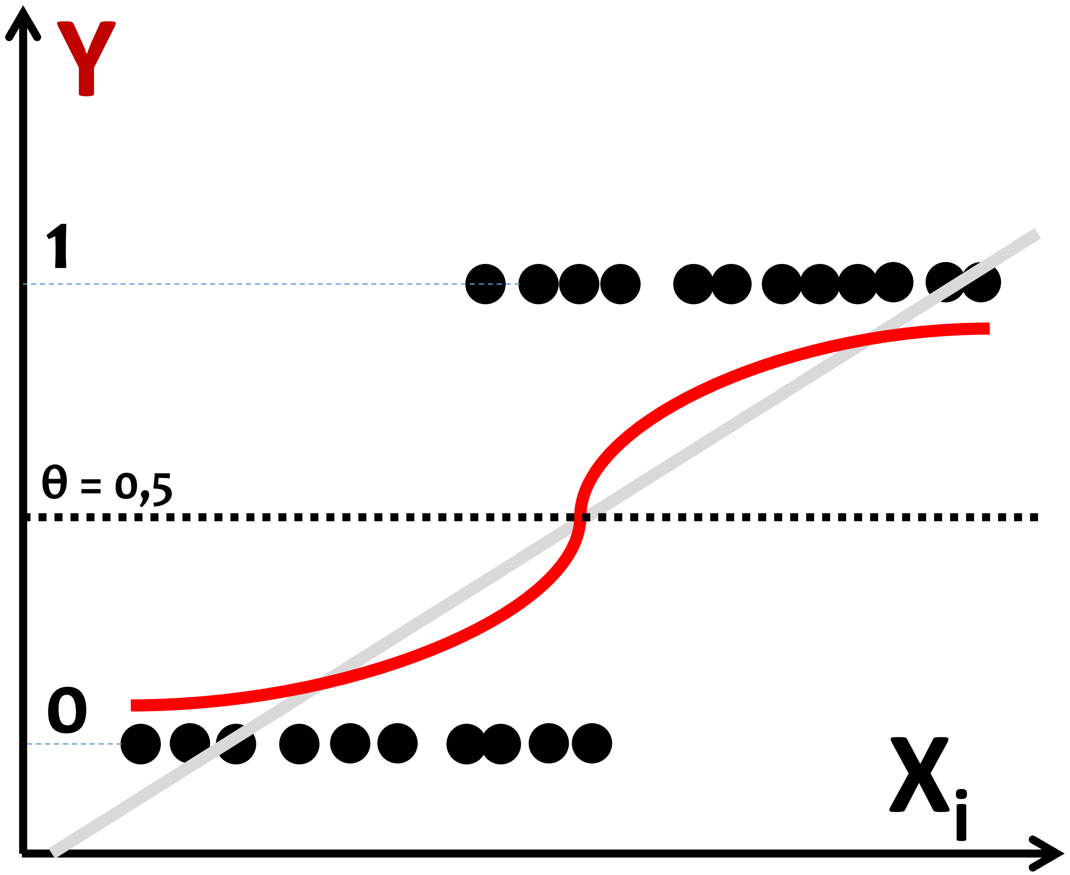
: régression logistique binaire simple

$$\pi(x) = P(Y = 1/\ X = x) = \frac{e^{\beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k}}{1 + e^{\beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k}}$$

: régression logistique binaire multiple

INTERPRÉTATION DE Y – PROBABILITÉ DE SUCCÈS

- **Prédire une variable décisionnelle ayant deux modalités Y = {0 (absence), 1 (présence)}**
- **L'une désigne un succès (Y = 1) et l'autre un échec (Y = 0)**
- **Le principe de la régression dans ce cas est de *chercher la probabilité* d'obtenir le succès P(Y = 1)**
- **Obtenir la probabilité du cas succès → Obtenir la probabilité de l'échec**
- **P(Y = 0) = 1 - P(Y = 1)**



Pour décider :

- ✓ **Se munir d'une règle de décision**
- ✓ **Pour un seuil θ :**

$$Y = \begin{cases} 1 \text{ si } P(Y = 1) > \theta \\ 0 \text{ si } P(Y = 1) \leq \theta \end{cases}$$

- ✓ **En approximation : θ = 0,5**

Objectifs : trouver les meilleurs β_i

$$\pi(X) = P(Y = 1|X = x_i) = \frac{e^{\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k}}{1 + e^{\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k}}$$

$$\pi(X) + \pi(X).\, e^{\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k} = e^{\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k}$$

$$\pi(X) = e^{\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k} . (1 - \pi(X))$$

$$e^{\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k} = \frac{\pi(X)}{1 - \pi(X)}$$

$$\ln(e^{\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k}) = \ln(\frac{\pi(X)}{1 - \pi(X)})$$

$$\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k = \ln(\frac{\pi(X)}{1 - \pi(X)})$$

Objectifs : trouver les meilleurs β_i

$$\pi(X) = P(Y = 1|X = x_i) = \frac{e^{\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k}}{1 + e^{\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k}}$$

$$\pi(X) + \pi(X).\, e^{\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k} = e^{\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k}$$

$$\pi(X) = e^{\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k}.\left(1 - \pi(X)\right)$$

$$e^{\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k} = \frac{\pi(X)}{1 - \pi(X)}$$

$$\ln(e^{\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k}) = \ln(\frac{\pi(X)}{1 - \pi(X)})$$

$$\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k = \ln(\frac{\pi(X)}{1 - \pi(X)})$$

OBTENTION DES COEFFICIENTS CLASSIFIEURS $\boldsymbol{\beta_i}$

Objectifs : trouver les meilleurs $\boldsymbol{\beta_i}$

$$\pi(X) = P(Y = 1|X = x_i) = \frac{e^{\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k}}{1 + e^{\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k}}$$

$$\pi(X) + \pi(X).e^{\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k} = e^{\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k}$$

$$\pi(X) = e^{\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k}.\left(1 - \pi(X)\right)$$

$$e^{\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k} = \frac{\pi(X)}{1 - \pi(X)}$$

$$\ln(e^{\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k}) = \ln(\frac{\pi(X)}{1 - \pi(X)})$$

$$\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k = \ln(\frac{\pi(X)}{1 - \pi(X)})$$

Objectifs : trouver les meilleurs $\boldsymbol{\beta_i}$

$$\pi(X) = P(Y = 1|X = x_i) = \frac{e^{\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k}}{1 + e^{\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k}}$$

$$\pi(X) + \pi(X).e^{\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k} = e^{\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k}$$

$$\pi(X) = e^{\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k}.\left(1 - \pi(X)\right)$$

$$e^{\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k} = \frac{\pi(X)}{1 - \pi(X)}$$

$$\ln(e^{\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k}) = \ln(\frac{\pi(X)}{1 - \pi(X)})$$

$$\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k = \ln(\frac{\pi(X)}{1 - \pi(X)})$$

OBTENTION DES COEFFICIENTS CLASSIFIEURS β_i

Objectifs : trouver les meilleurs β_i

$$\pi(X) = P(Y = 1|X = x_i) = \frac{e^{\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k}}{1 + e^{\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k}}$$

$$\pi(X) + \pi(X).\, e^{\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k} = e^{\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k}$$

$$\pi(X) = e^{\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k}.\left(1 - \pi(X)\right)$$

$$e^{\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k} = \frac{\pi(X)}{1 - \pi(X)}$$

$$\ln(e^{\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k}) = \ln(\frac{\pi(X)}{1 - \pi(X)})$$

$$\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k = \ln(\frac{\pi(X)}{1 - \pi(X)})$$

Objectifs : trouver les meilleurs $\boldsymbol{\beta_i}$

$$\pi(X) = P(Y = 1|X = x_i) = \frac{e^{\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k}}{1 + e^{\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k}}$$

$$\pi(X) + \pi(X).e^{\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k} = e^{\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k}$$

$$\pi(X) = e^{\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k}.\left(1 - \pi(X)\right)$$

$$e^{\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k} = \frac{\pi(X)}{1 - \pi(X)}$$

$$\ln(e^{\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k}) = \ln(\frac{\pi(X)}{1 - \pi(X)})$$

$$\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k = \ln(\frac{\pi(X)}{1 - \pi(X)})$$

Objectifs : trouver les meilleurs β_i

$$\pi(X) = P(Y = 1|X = x_i) = \frac{e^{\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k}}{1 + e^{\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k}}$$

$$\pi(X) + \pi(X).e^{\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k} = e^{\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k}$$

$$\pi(X) = e^{\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k}.\left(1 - \pi(X)\right)$$

$$e^{\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k} = \frac{\pi(X)}{1 - \pi(X)}$$

$$\ln(e^{\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k}) = \ln(\frac{\pi(X)}{1 - \pi(X)})$$

$$\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k = \ln(\frac{\pi(X)}{1 - \pi(X)})$$

OBTENTION DES COEFFICIENTS CLASSIFIEURS $\boldsymbol{\beta_i}$

$$\boldsymbol{\beta_0} + \boldsymbol{\beta_1}\,\mathbf{x_1} + \cdots + \boldsymbol{\beta_k}\,\mathbf{x_k} = \ln\left(\frac{\boldsymbol{\pi(X)}}{\mathbf{1} - \boldsymbol{\pi(X)}}\right)$$

PROBABILITÉ DU CAS 'SUCCÈS'

X_1	X_2	Y
1	2	OUI
1	1	OUI
2	2	OUI
2	2	OUI
		NON
		NON
		NON

$$P(Y = OUI \mid X_1 = 1 \ et \ X_2 = 2) = \frac{1}{4} = 0,25$$

$$P(Y = OUI \mid X_1 = 1 \ et \ X_2 = 1) = \frac{1}{4} = 0,25$$

$$P(Y = OUI \mid X_1 = 2 \ et \ X_2 = 2) = \frac{2}{4} = 0,5$$

$$P(Y = OUI \mid X_1 = 2 \ et \ X_2 = 1) = \frac{0}{4} = 0$$

UTILITÉ DE LA FONCTION $\pi(X)$

X_1	X_2	Y	P(Y=OUI | X_i)
1	2	OUI	0,25
1	1	OUI	0,25
2	2	OUI	0,5
2	2	OUI	0,5
		NON	
		NON	
		NON	

$$P(Y = OUI \mid X_1 = 1 \; et \; X_2 = 2\,) = \frac{1}{4} = 0,25$$

$$P(Y = OUI \mid X_1 = 1 \; et \; X_2 = 1\,) = \frac{1}{4} = 0,25$$

$$P(Y = OUI \mid X_1 = 2 \; et \; X_2 = 2\,) = \frac{2}{4} = 0,5$$

$$P(Y = OUI \mid X_1 = 2 \; et \; X_2 = 1\,) = \frac{0}{4} = 0$$

Supposons qu'on va considérer les valeurs des probabilités de Y : [0,1]

Alors on va construire un modèle linéaire qui explique les probabilités P(Y|X_i) par les X_i

Lors de la prédiction : on risque d'avoir des valeurs de probabilités

P(Y|X_i) < 0 ou P(Y|X_i) > 1

Absurde : Modèle non significatif / erroné ?

UTILITÉ DE LA FONCTION $\pi(X)$

X_1	X_2	Y	P(Y=OUI | X_i)
1	2	OUI	0,25
1	1	OUI	0,25
2	2	OUI	0,5
2	2	OUI	0,5
		NON	
		NON	
		NON	

$$P(Y = OUI \mid X_1 = 1 \; et \; X_2 = 2) = \frac{1}{4} = 0,25$$

$$P(Y = OUI \mid X_1 = 1 \; et \; X_2 = 1) = \frac{1}{4} = 0,25$$

$$P(Y = OUI \mid X_1 = 2 \; et \; X_2 = 2) = \frac{2}{4} = 0,5$$

$$P(Y = OUI \mid X_1 = 2 \; et \; X_2 = 1) = \frac{0}{4} = 0$$

- ✓ **Passer à la transformation Logistique**
- ✓ **Transformer l'intervalle des probabilités en des valeurs réelles moyennant une fonction inversible**

UTILITÉ DE LA FONCTION $\pi(X)$

X_1	X_2	Y	P(Y=OUI | X_i)
1	2	OUI	0,25
1	1	OUI	0,25
2	2	OUI	0,5
2	2	OUI	0,5
		NON	
		NON	
		NON	

$$P(Y = OUI \mid X_1 = 1 \ et \ X_2 = 2) = \frac{1}{4} = 0,25$$

$$P(Y = OUI \mid X_1 = 1 \ et \ X_2 = 1) = \frac{1}{4} = 0,25$$

$$P(Y = OUI \mid X_1 = 2 \ et \ X_2 = 2) = \frac{2}{4} = 0,5$$

$$P(Y = OUI \mid X_1 = 2 \ et \ X_2 = 1) = \frac{0}{4} = 0$$

- ✓ **Passer à la transformation Logistique**
- ✓ **Transformer l'intervalle des probabilités en des valeurs réelles moyennant une fonction inversible**
- ✓ **La Fonction LOGIT est une fonction bijective**
- ✓ **Elle permet de récupérer les probabilités dans un sens inverse**

APPLICATION DE LA FONCTION $ln(\frac{\pi(X)}{1-\pi(X)})$

X_1	X_2	Y	P(Y=OUI | X_i)
1	2	OUI	0,25
1	1	OUI	0,25
2	2	OUI	0,5
2	2	OUI	0,5
		NON	
		NON	
		NON	

$$\pi_1(Y = OUI \mid X_1 = 1\ et\ X_2 = 2) = \frac{1}{4} = 0,25$$

$$\pi_2(Y = OUI \mid X_1 = 1\ et\ X_2 = 1) = \frac{1}{4} = 0,25$$

$$\pi_3(Y = OUI \mid X_1 = 2\ et\ X_2 = 2) = \frac{2}{4} = 0,5$$

$$\pi_4(Y = OUI \mid X_1 = 2\ et\ X_2 = 1) = \frac{0}{4} = 0$$

$$ln(\frac{\pi_1(X)}{1-\pi_1(X)}) = ln(\frac{0,25}{1-0,25}) = -1,098612287$$

$$ln(\frac{\pi_2(X)}{1-\pi_2(X)}) = ln(\frac{0,25}{1-0,25}) = -1,098612287$$

$$ln(\frac{\pi_3(X)}{1-\pi_3(X)}) = ln(\frac{0,5}{1-0,5}) = 0$$

$$ln(\frac{\pi_4(X)}{1-\pi_4(X)}) = ln(\frac{0}{1-0}) = -\infty$$

APPLICATION DE LA FONCTION $ln(\frac{\pi(X)}{1-\pi(X)})$

X_1	X_2	Y	$P(Y=OUI \mid X_i)$	$ln(\frac{\pi(X)}{1-\pi(X)})$
1	2	OUI	0,25	$-1,098612287$
1	1	OUI	0,25	$-1,098612287$
2	2	OUI	0,5	0
2	2	OUI	0,5	0
		NON		
		NON		
		NON		

$]-\infty, +\infty[$

Information purement quantitative

Régression Logistique Polytomique

Variable cible à K (K > 2) modalités

Famille de méthodes

Régression logistique multinomiale

- La variable cible est nominale à K >2 modalités
- Les modalités ne sont pas ordonnées
- On ne veut pas tenir compte d’un éventuel ordonnancement
- LOGITS par rapport à une modalité de référence

Régression logistique ordinale

- La variable cible est nominale à K >2 modalités
- Plus que la prédiction, c’est l’interprétation des coefficients qui importe
- LOGITS adjacents : on applique des régressions logit entre chaque couple de deux modalités voisines.
- ODDS-RATIO cumulatifs : on choisit un seuil et on applique la régression logit entre les modalités avant ce seuil, et le restant.

Régression logistique multinomiale

Principe :

- Modéliser la probabilité d'appartenance d'un individu à une catégorie « k »

$$\pi_k(\omega) = P[Y(\omega) = y_k / X(\omega)]$$

Avec toujours la contrainte $\sum_k \pi_k(\omega) = 1$

- Prendre une modalité de référence, la dernière par exemple, et estimer (K-1) LOGITS

$$LOGIT_k(\omega) = \ln\left[\frac{\pi_k(\omega)}{\pi_K(\omega)}\right] = a_{0,k} + a_{1,k}X_1(\omega) + \cdots + a_{J,k}X_J(\omega)$$

Stratégie de modélisation :

- **modéliser (K-1) rapports de probabilités**
- **Prendre une modalité comme référence** (exp. la dernière)
- **exprimer (K-1) LOGIT par rapport à cette référence**

$$\ln\frac{P(Y = k / X)}{P(Y = K / X)} = a_{0,k} + a_{1,k}X_1 + \cdots + a_{J,k}X_J, k = 1, \ldots, K-1$$

Régression logistique Ordinale
LOGITS adjacents

Principe :

- Calculer le LOGIT du passage d'une catégorie à l'autre
- (K-1) équations LOGIT sont calculés → (K-1) x (J+1) paramètres à estimer

$$\begin{cases} LOGIT_1(\omega) = \ln\left[\dfrac{\pi_1(\omega)}{\pi_2(\omega)}\right] = a_{0,1} + a_{1,1}X_1(\omega) + \cdots + a_{J,1}X_J(\omega) \\ \cdots \\ LOGIT_{K-1}(\omega) = \ln\left[\dfrac{\pi_{K-1}(\omega)}{\pi_K(\omega)}\right] = a_{0,K-1} + a_{1,K-1}X_1(\omega) + \cdots + a_{J,K-1}X_J(\omega) \end{cases}$$

$$\begin{cases} \ln\left[\dfrac{\pi_2(\omega)}{\pi_1(\omega)}\right] = -LOGIT_1(\omega) \\ \ln\left[\dfrac{\pi_3(\omega)}{\pi_1(\omega)}\right] = -LOGIT_2(\omega) - LOGIT_1(\omega) \\ \cdots \\ \ln\left[\dfrac{\pi_K(\omega)}{\pi_1(\omega)}\right] = -LOGIT_{K-1}(\omega) - \cdots - LOGIT_2(\omega) - LOGIT_1(\omega) \end{cases}$$

Même idée que le modèle multinomial, sauf que la catégorie de référence change à chaque étape : on procède par des couples de références en alternance.

On évalue le passage de la modalité (k) à (k-1)

Régression logistique Ordinale
LOGITS cumulatifs

Principe : Calculer le LOGIT d'être au delà ou en deçà du niveau y_k de la variable Y

LOGITS cumulatifs $$LOGIT_k = \ln\left(\frac{P(Y \le k / X)}{P(Y > k / X)}\right) = \ln\left(\frac{P(Y \le k / X)}{1 - P(Y \le k / X)}\right) = \ln\left(\frac{\pi_1 + \cdots + \pi_k}{\pi_{k+1} + \cdots + \pi_K}\right)$$

$$\ln \frac{P(Y \le k / X)}{P(Y > k / X)} = a_{0,k} + a_{1,k} X_1 + \cdots + a_{J,k} X_J , k = 1, \ldots, K - 1$$

Evaluation

Mesure de la qualité de la modélisation

- R^2 de Cox & Snell

$$R^2 = 1 - \left[\frac{l(cte)}{l(cte, X)}\right]^{\frac{2}{n}}$$

$$Max \ \ R^2 = 1 - [l(cte)]^{\frac{2}{n}}$$

- Pseudo R^2 (McFadden)

$$Pseudo - R^2 = 1 - \left[\frac{-2L(cte, X)}{-2L(cte)}\right]$$

- R^2 ajusté de Nagelkerke

$$R^2_{adj} = \frac{R^2}{R^2_{max}}$$