date ..... / .....

la vraie réponse moyenne au point $x = x_{n+1}$;
Soit une paramètre de la population et non une nouvelle observation
- L'IP en x_{n+1} est toujours plus grand que Ic en x_{n+1} car il dépend de l'erreur associée aux futures observations
- L'IP n'est valide que pour une nouvelle observation à la fois.

I/2) 2/ Régression linéaire multiple et modèle matriciels :

2.1/ Rappels d'Algèbre linéaire

Vecteur : un vecteur est un élément d'un espace vectoriel représente souvent sous forme de colonne ou de ligne : $\vec{V} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$

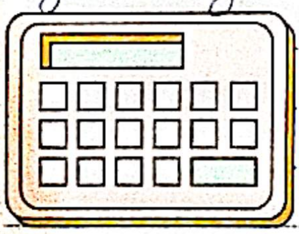
- $\forall \lambda \in \mathbb{R}$, $\lambda \vec{V} = \lambda \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} \lambda x \\ \lambda y \\ \lambda z \end{pmatrix}$

- Soit $\vec{u} = \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix}$ une vecteur dans l'espace vectoriel, alors :

$$\vec{V} + \vec{u} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} + \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} x + x' \\ y + y' \\ z + z' \end{pmatrix}.$$

- Combinaison linéaire : un Combinaison linéaire des vecteurs $V_1, \dots, V_n$ donner par : $a_1 V_1 + \dots + a_n V_n$; Tq: a_i sont des scalaires.
- Matrice : Tableau de nombres organisé en lignes et colonnes sert à représenter des applications linéaires ou des systèmes d'équations.
On note $(M_{n,m})$ une matrice de n lignes et m colonnes, ou on peut note $(a_{ij})_{\substack{1 \le i \le n \\ 1 \le j \le m}}$, tq: a_{ij} sont des coefficient réels.

- matrice carrée : est une matrice pour laquelle le nombre de lignes est égal au nombre de colonnes, on note : M_n, matrice carrée de taille $n \times n$



Produit matriciel : soient deux matrices $A_{m,n}$ et $B_{n,p}$.
Produit $C = A \times B$ est définie seulement si le nombre

de colonne de A égale à le nombre de lignes de B, Alors C une matrice de taille m × p : C_{mp}.

Exemples: Calculer le produit des matrices:

$$A = \begin{pmatrix} 1 & 6 & 3 \\ 4 & 5 & 2 \end{pmatrix} ; B = \begin{pmatrix} 5 & 1 \\ 3 & 0 \\ -1 & 2 \end{pmatrix}$$

$$C = \begin{pmatrix} 20 & 7 \\ 33 & 8 \end{pmatrix}$$

Remarque: $AB \neq BA$

le produit de deux matrice n'est pas commutatif.

. Somme des matrices:

Soient deux matrices A et B de même taille n × m, leur somme $C = A + B$ est définie élément par élément $c_{ij} = a_{ij} + b_{ij}$ (C est de taille n × m)

Ex: $A = \begin{pmatrix} 1 & 0 \\ -1 & 2 \end{pmatrix}$; $B = \begin{pmatrix} 1 & 1 \\ 2 & -3 \end{pmatrix}$; leurs somme est : $C = \begin{pmatrix} 2 & 1 \\ 1 & -1 \end{pmatrix}$

. la somme de deux matrice est commutatif c.à.d $A+B = B+A$

soit $K \in \mathbb{R}$, $K(A+B) = KA + KB$.

. Système linéaire:

est un ensemble d'équations où les variables sont toutes exprimées au premier degrée, Cela signifie que l'on peut combiner les équation pour trouver une ou plusieurs solutions:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n \end{cases}$$

ou ($AX = b$ écriture matricielle)

Système linéaire de n équations à p inconnues si $b_i = 0$, $i = \overline{1,n}$ on appelle système homogène

Déterminant: est un nombre (scalaire) associé à une matrice carrée (n × n), il se note det A ou $|A|$.

date .... / ....

- Si $\det A \neq 0$ on dit A est inversible
- Si $A = a \Rightarrow \det A = a$
- Si A de taille 2×2 ; $A = \begin{pmatrix} a & c \\ b & d \end{pmatrix}$

$\Rightarrow \det A = ad - cb$.

Exemple : Calculer le det B tq :

$$B = \begin{pmatrix} 1 & 1 & 3 \\ 2 & 2 & 4 \\ 3 & 2 & 1 \end{pmatrix}$$

$$\det B = 1 \times \begin{vmatrix} 2 & 4 \\ 2 & 1 \end{vmatrix} - 1 \times \begin{vmatrix} 2 & 4 \\ 3 & 1 \end{vmatrix} + 3 \begin{vmatrix} 2 & 2 \\ 3 & 2 \end{vmatrix}$$

$\det B = -6 + 10 - 6 = -2$

- La matrice transposée :

La transposée d'une matrice A notée A^T (ou A'), est la matrice obtenue en échangeant les lignes et les colonnes de A, ie

$A = a_{ij} \Rightarrow A^T = a_{ji}$

Remarque : $\det A = \det A^T$

propriétés : $(A^T)^T = A$

$(AB)^T = B^T A^T$

$(A+B)^T = A^T + B^T$

Ex : la transposée de la matrice : $A = \begin{pmatrix} -1 & 0 & 1 \\ 2 & 3 & 4 \\ 6 & 3 & 1 \end{pmatrix}$ est $A' = \begin{pmatrix} -1 & 2 & 6 \\ 0 & 3 & 3 \\ 1 & 4 & 1 \end{pmatrix}$.

- La matrice inversible : une matrice carrée A est dite inversible ($\det A \neq 0$) ou régulière s'il existe une matrice A^{-1} carrée (appelée matrice inverse) ; Tq : $A^{-1}A = AA^{-1} = I$

où I est la matrice identité. $A^{-1} = \frac{1}{\det A} \, \mathrm{com}(A)^T$.

Ex : Calculer la matrice inversible de A.

$$A = \begin{pmatrix} +1 & -1 & +3 \\ -2 & +2 & 4 \\ 3 & -2 & 1 \end{pmatrix}. \quad \det A = -2.$$

$$\mathrm{com}(A) = \begin{vmatrix} +\begin{vmatrix} 2 & 4 \\ 2 & 1 \end{vmatrix} & -\begin{vmatrix} 2 & 4 \\ 3 & 1 \end{vmatrix} & +\begin{vmatrix} 2 & 2 \\ 3 & 2 \end{vmatrix} \\ -\begin{vmatrix} 1 & 3 \\ 2 & 1 \end{vmatrix} & +\begin{vmatrix} 1 & 3 \\ 3 & 1 \end{vmatrix} & -\begin{vmatrix} 1 & 1 \\ 3 & 2 \end{vmatrix} \\ +\begin{vmatrix} 1 & 3 \\ 2 & 4 \end{vmatrix} & -\begin{vmatrix} 1 & 3 \\ 2 & 4 \end{vmatrix} & +\begin{vmatrix} 1 & 1 \\ 3 & 2 \end{vmatrix} \end{vmatrix}$$

date .... / ....

$$Com(A) = \begin{vmatrix} -6 & 10 & 4 \\ 5 & -9 & 1 \\ -2 & 2 & -1 \end{vmatrix}$$

$$Com(A)^T = \begin{pmatrix} -6 & 5 & -2 \\ 10 & -9 & 2 \\ 4 & 2 & -1 \end{pmatrix}$$

$$A^{-1} = -\frac{1}{2} \begin{pmatrix} -6 & 5 & -2 \\ 10 & -9 & 2 \\ 4 & 2 & -1 \end{pmatrix} = \begin{pmatrix} 3 & -\frac{5}{2} & +1 \\ -5 & +\frac{9}{2} & -1 \\ -2 & -1 & +\frac{1}{2} \end{pmatrix}$$

2-2) Régression linéaire multiple: la régression linéaire multiple est identique à la régression linéaire simple sauf que dans cette cas, on va expliquer la variable à expliquer par plusieurs variable explicatives. (16/11/2025)

2.2.1) forme généralisé de le modèle de R.L.M: On souhait prédire et expliquer les valeurs d'une variable ~~quali~~ quantitative, ~~(on dit alors)~~ y à partir des valeur de p variables $X_1, X_2, \dots, X_p$ quantitative. on dit alors que on explique y par $X_1, \dots, X_p$. pour ce faire on dispose des données qui sont n observations de $(y, X_1, \dots, X_p)$ de réalisation $(y_1, x_{11}, \dots, x_{p1}); \dots; (y_n, x_{1n}, \dots, x_{pn})$; Elle se présentent généralement sous forme d'une tableau

Y	X_1	---	X_p
y_1	x_{11}	- -	x_{p1}
y_2	x_{12}	- -	x_{p2}
⋮	⋮	⋮	⋮
y_n	x_{1n}	- -	x_{pn}

s'il y a une liaison linéaire entre y et $X_1, \dots, X_p$, on peut considérer le modèle de régression linéaire multiple (R.L.M) de forme générale:

$$y = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p + \varepsilon$$

où $\beta_0, \beta_1, \dots, \beta_p$ sont des coefficient réels inconnus à déterminer et ε est un variable quantitatives indépendante de $X_1 \dots X_p$ avec $E(\varepsilon_i) = 0$ pour $i = \overline{1,n}$, pour une réalisation $(x_{1i}, \dots, x_{pi})$, y_i de $(X_1 \dots X_p)$ et y on peut c.à.d: $y_i = \beta_0 + \beta_1 x_{1i} + \dots + \beta_p x_{pi} + \varepsilon_i$ pour $i = \overline{1,n}$

Exemple: la dépendance de le goût d'un fromage (y) et la concentraction de plusieur composés chimiques dont: la concentraction de l'acide acétique (variable X_1).

date 16 / 11 / 2025.

- la concentration d'hydrogène sulfuré (X_2).
- " " d'acide lactique (X_3)

* écriture matricielle du modèle de RLM

on peut transformer la forme générale de la modèle de RLM à la forme matricielle suivante :

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{p1} \\ 1 & x_{12} & \cdots & x_{p2} \\ \vdots & \vdots & & \vdots \\ 1 & x_{1n} & \cdots & x_{pn} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_p \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

23/11/2025

2.2.2) Estimation des moindres carrées ordinaires

Notre principale objectif est d'estimer convenablement $\beta_0, \dots, \beta_p$ à l'aide des données

Soient $\| \, \|$ la norme euclidienne pour tout vecteurs colonne X ;
$\|X\|^2 = X^t X$

Partant du modèle de RLM écrit sous la forme matricielle $Y = X\beta + \varepsilon$, un estimateur des moindres carrées ordinaires (EMCO), $\hat{\beta}$ de β vérifier $\hat{\beta} \in \arg\min_{\beta \in \mathbb{R}^{p+1}} \|\underbrace{Y - X\beta}_{\varepsilon}\|^2$

on suppose que X est de rang colonnes plein : il n'existe pas de vecteur colonne x à $p+1$ composantes non null tel que $Xx =$ le vecteur null cela entraine l'existence de $(X^tX)^{-1}$. Alors $\hat{\beta}$ est unique, il est donnée par la formule $\hat{\beta} = (X^TX)^{-1}X^TY$

preuve : voir (Sur l'estimateur des moindres carrées ordinaires (EMCO) Christophe Chesneau page 8.

$\hat{\beta} = \arg\min \|\underbrace{Y - X\beta}_{\varepsilon_i}\|^2$

$S = \sum_{i=1}^{n} \varepsilon_i^2 = \|\varepsilon\|^2 = \varepsilon^T\varepsilon$

$= (Y - X\beta)^T (Y - X\beta)$

$= Y^TY - 2\beta^TX^TY + \beta^TX^TX\beta$

$\frac{\partial S}{\partial \beta} = -2X^TY + 2X^TX\beta = 0$

$\Rightarrow \hat{\beta} = (X^TX)^{-1}X^TY$

date .... / ....

Propriétés des estimateurs :

- L'estimateur $\hat{\beta}$ est le meilleur estimateur non biaisé de β avec : $E(\hat{\beta}) = \beta$
  - $Var(\hat{\beta}) = \hat{\sigma}^2_{\varepsilon} [X^T X]^{-1}$

- **Coefficient de détermination R^2 :**

R^2 ~~est~~ donnée par : $R^2 = 1 - \dfrac{\|\hat{Y} - Y\|^2}{\|\bar{Y}1_n - Y\|}$

où : $\hat{Y} = X\hat{\beta}$ et $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} y_i$

plus R^2 est proche de 1 plus la liaison linéaire entre Y et $X_1, X_2, \dots, X_p$ est forte.

- **Coefficient de détermination ajusté :**

une version améliorée du R^2 défini par : $\bar{R}^2 = 1 - \dfrac{n-1}{n-(p+1)}(1 - R^2)$.

Exp : on suppose qu'on a le ~~ta~~ tableau suivant :

Année	Y	X_1	X_2
2000	4	1	6
2001	4	3	6
2002	8	4	7
2003	12	6	7
2004	12	6	7

Question : estimer les coefficients de ce modèle.

$Y = \beta X + \varepsilon$

?

2.3) Estimateur par maximum de vraisemblance :

2.3.1) la fct vraisemblance :

Déf : $x_1, \dots, x_n$ un échantillon de vraisemblance aléatoires de loi de probabilité $P_{X_i}(x_i)$, la fct de vraisemblance est la fct de θ définie par : $L(\theta, x_1, \dots, x_n) = P(x_1, \dots, x_n / \theta)$, cette fct ne calcule pas la probabilité des données mais elle mesure à quel point un paramètre θ rend les données observées plausibles.

- C'est l'outil de base pour l'estimation du EMV

2.3.2) Estimation par MV en RLMS

Dans le cas général de la régression linéaire multiple, on considère le modèle $Y = X\beta + \varepsilon$

où $\varepsilon_i \sim N(0, \sigma^2 I_n)$.

Eq: Y est le vecteur des observation $(n \times 1)$.

$\hat{Y}$ le vecteur des prédictions $(n \times 1)$

X est la matrice des prédicteurs $(n \times p)$.

β est le vecteur des paramètres $(p \times 1)$.

ε est un vecteur de bruit gaussien.

Chaque y_i suit distribution normale $y_i \sim N(X_i\beta, \sigma^2)$.

$\Rightarrow Y \sim N(X\beta \,;\, \Sigma\, \sigma^2 I_n)$. 30/11

Sous l'hypothèse d'indépendance des erreurs on écrire la vraisemblance sous la forme d'un produit,

$$L(\beta, \sigma^2) = P(Y \mid \beta, \sigma^2) =$$

$$= P(y_1, \dots, y_n \mid \beta, \sigma^2) = \prod_{i=1}^{n} P(y_i \mid \beta, \sigma^2) \rightarrow \varepsilon_i$$

$$= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - x_i\beta)^2}{2\sigma^2}\right)$$

Comme $\prod_{i=1}^{n} \exp(a_i) = \exp\left(\sum_{i=1}^{n} a_i\right)$,

on obtient:

$$L(\beta, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - x_i\beta)^2\right)$$

La somme des carrées des erreurs est la produit scalaire:

$$\sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - x_i\beta)^2 = (Y - X\beta)^T (Y - X\beta) = \|Y - X\beta\|^2$$

Ainsi, en réécrivant sous forme matricielle, on obtient:

$$L(\beta, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(\frac{1}{2\sigma^2} (Y - X\beta)^T (Y - X\beta)\right)$$

maximiser la log-vraisemblance est équivalente à maximiser la vraisemblance, donc on obtient $\log(L(\beta, \sigma^2)) = \ell(\beta, \sigma^2)$

$$= -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} (Y - X\beta)^T (Y - X\beta)$$

maximiser la log-vraisemblance revient à trouver l'argument de β qui minimiser la somme des carrées des erreurs.

date 30 / 11

$\hat{\beta}_{MV} = \arg\max_{\beta} l(\beta, \delta^2)$

$= \arg\max_{\beta} \left(-\frac{n}{2} \log(2\pi\delta^2) - \frac{1}{2\delta^2} (y - X\beta)^T (y - X\beta) \right)$

$= \arg\min_{\beta} (y - X\beta)^T (y - X\beta)$

$= \hat{\beta}_{MCO}$

on retrouve ici le critère des moindres carrés ordinaires qui donne les estimateurs

$\hat{\beta}_{MV} = (X^T X)^{-1} X^T y = \hat{\beta}_{MCO}$.

Ce qui explique pourquoi les estimateurs MV et MCO coïncident sous l'hypothèse de normalité et d'indépendance des erreurs

- Pourquoi ne pas chercher la dérivée seconde?

Dans les deux méthode MCO et MV, la fonction de coût à minimiser une fonction quadratique en β

$C(\beta) = (y - X\beta)^T (y - X\beta)$

- Pour s'assurer qu'un minimum est bien atteint, on pourrait vérifier la matrice Hessiene : $\nabla C(\beta) = \frac{\partial C}{\partial \beta}$

$= -2X^T y + 2X^T X\beta \Rightarrow \nabla^2 C(\beta) = 2X^T X$.

la matrice de Gram $X^T X$ est semi-définie positive si les colonnes de X sont linéairement indépendantes, alors $X^T X$ est définie positive cela garantie que la fct est convexe et atteint minimum unique.

2.3.3) Multicolinéarité et VIF : (VIF : facteur d'inflation de la variance).

Déf : "Multicolinéarité

On parle de multicolinéarité lorsque deux variables explicatives ou plus d'un modèle de regression sont fortement corrélées entre elles (c.à.d) : l'une des variables explicatives est une combinaison linéaire d'une plusieurs autres variables explicative introduites dans le même modèle

- Une multicolinéarité prononcé s'avère problématique, car quand on a ce cas, les conséquences peuvent être les suivantes :

- Les coefficients deviennent instables (petite

date .... / ....

changement dans les données → grands changement dans les coefficients).

- Les erreurs standards augement, donc les tests deviement non significatifs alors que la variable est réellement utile.
- Il devient difficile d'interpréter l'effet individuel des variables corrélées.
- le modèle peut avoir une faible capacité de généralisation.

- Facteur d'inflation de la variance V.I.F.

V.I.F est la méthode la plus courant pour détecter la multicolinéairité.

Déf: "VIF" : La facteur V.I.F est une mesure statistique qui quantifie l'étendu de multicolinéarité, en indéquant dans quelle mesur la variance d'une coefficient de régression est gonflée par la corrélations.

- Si l'on considère une variable explicative X_i, $V.I.F = \frac{1}{1-R_i^2}$

où: R_i^2 est le coefficient de détermination obtenu en regressant X_i sur toutes les autres variables.

Interprétation:

VIF: proche de 1: (VIF ≤ 1) Aucune corrélation avec d'autres prédicteurs.

$1 < VIF < 5$: corrélation modérée généralement aceptable.

$VIF > 5$: multicolinéarité importante, à surveiller.

$VIF > 10$: Multicolinéarité sévère

⇒ Problème sérieux pour la Fiabilité du modèle.

date 07 / 12

Les solutions pour le VIF et trop élevé :

Supprimer une des variables corrélées (garder seulement l'une des deux variables très proches)

- Combiner les variables par : Moyenne ACP
- Centrer ou normaliser les donnés (cela n'enléve pas la multicolinéarite mais peut réduire les problémes numériques
- utiliser une régression pénalisée :
  - Ridge
  - Lasso

Exp : On souhaite analyser l'effet de plusieurs variables explicatives sur le prix d'une maison, les données suivantes ont été observées

prix (\$)	Surface (m^2)	Nombre de chambre	age
200	100	3	10
260	150	4	6
320	200	5	3
210	120	3	12

Question : il y a une multicolinéarité entre les variables explicatives ?

II) Diagonistic et Validations des modéles :

1) Contrôle de l'aptitude du modéle :

1-1) Analyse des résidus :

Dans une modèle de R L le résidus ε_i est la déffirance entre la valeur réel et la valeur ajustée $\varepsilon_i = y_i - \hat{y}_i$

l'analyse des résidus permet de vérifier la valiation des hypothéses au modéle de R L :

- Les Hypothéses du modéle linéaire :