

REGRESSION LINEAIRE

©Année Universitaire 2021-2022 ©

PLAN

01

Introduction

02

Regression
linéaire simple

03

Regression
linéaire multiple



Classification

Prédire des valeurs
discrets

VS

Régression

Prédire des valeurs
continues(numériques)


INTRODUCTION

01

La régression linéaire est un modèle de régression qui cherche à établir une relation linéaire entre une variable, dite cible, et une ou plusieurs variables, dites explicatives.

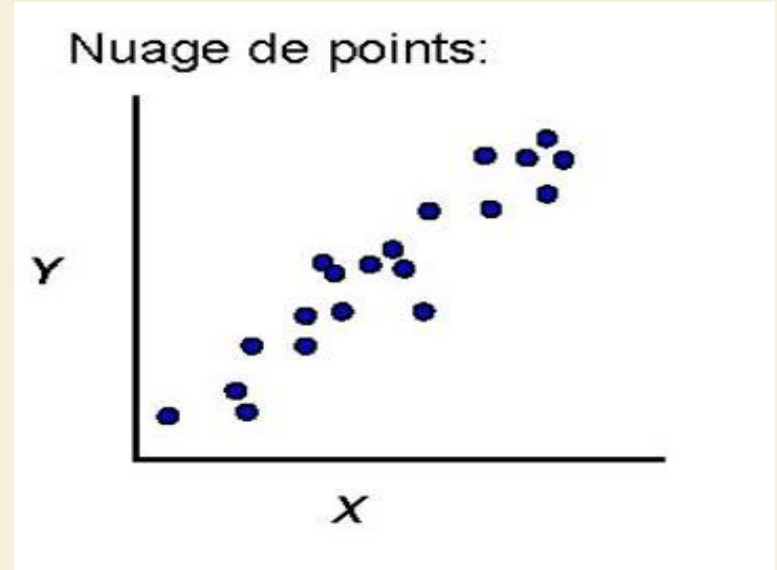
Regression linéaire Simple

02

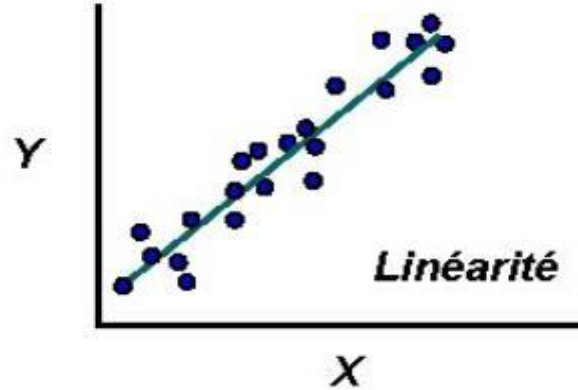


➤ La régression linéaire consiste à déterminer une droite qui se rapproche le plus possible d'un ensemble de points.

- Les points (x_i, y_i) sont les données d'entrée
- Les valeurs x_i sont les variables prédictives.
- La valeur y_i est la valeur observée



La relation est linéaire



- On cherche à trouver une droite : $y = ax + b$ qui soit le plus proche possible de tous les points de nos données d'apprentissage

- Pour trouver a et b on utilise le **critère des moindres carrés**:

$$\begin{cases} \hat{a} = \frac{\sum_i (y_i - \bar{y})(x_i - \bar{x})}{\sum_i (x_i - \bar{x})^2} \\ \hat{b} = \bar{y} - \hat{a}\bar{x} \end{cases}$$

Exemple de régression simple :

Expliquer le rendement de maïs Y (en quintal) à partir de la quantité d'engrais utilisé (en kilo) sur des parcelles de terrain similaires.

Modèle de régression simple :

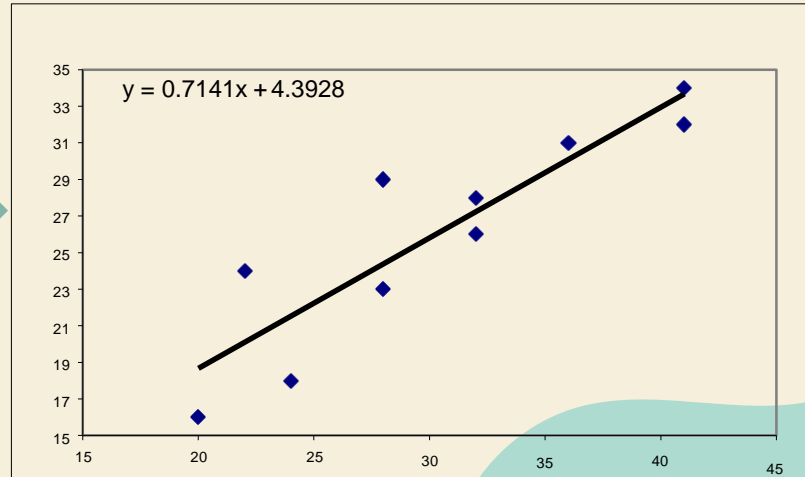
$$Y = ax + b$$

N° de parcelle	Y	X
1	16	20
2	18	24
3	23	28
4	24	22
5	28	32
6	29	28
7	26	32
8	31	36
9	32	41
10	34	41

	Y	X	(Y-YB)	(X-XB)	(Y-YB)(X-XB)	(X-XB) ²
1	16	20	-10.1	-10.4	105.04	108.160
2	18	24	-8.1	-6.4	51.84	40.960
3	23	28	-3.1	-2.4	7.44	5.760
4	24	22	-2.1	-8.4	17.64	70.560
5	28	32	1.9	1.6	3.04	2.560
6	29	28	2.9	-2.4	-6.96	5.760
7	26	32	-0.1	1.6	-0.16	2.560
8	31	36	4.9	5.6	27.44	31.360
9	32	41	5.9	10.6	62.54	112.360
10	34	41	7.9	10.6	83.74	112.360
Moyenne	26.1	30.4	Somme		351.6	492.4

$$a = 351,6 / 492,4 = 0,714$$

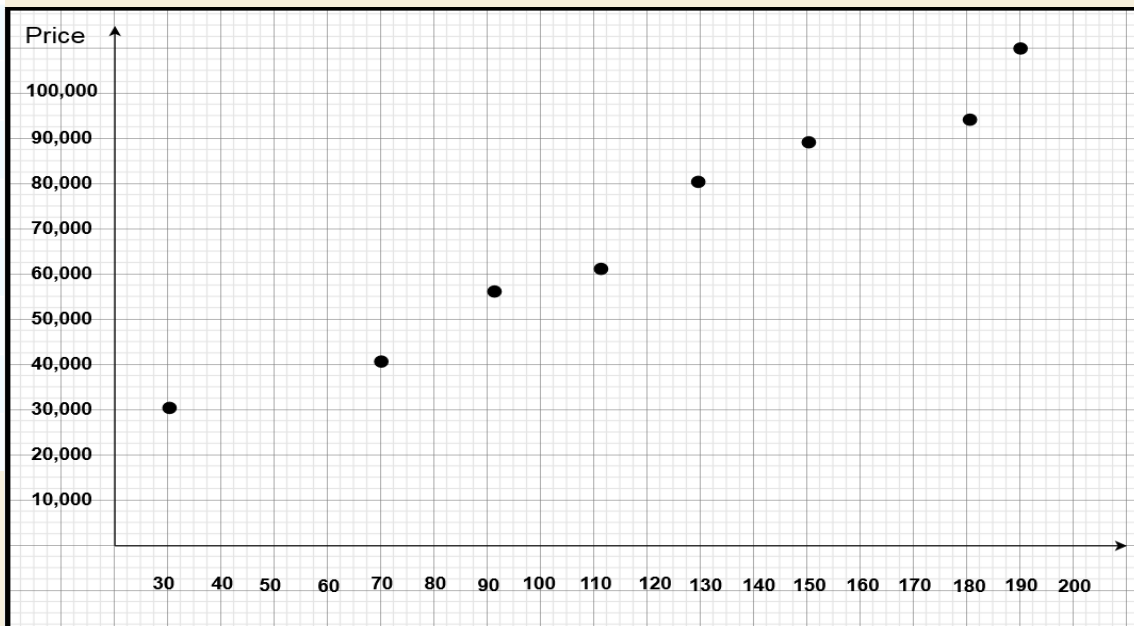
$$b = 26,1 - 0,714 * 30,4 = 4,39$$





ACTIVITE

Size	Price
30	30,000
70	40,000
90	55,000
110	60,000
130	80,000
150	90,000
180	95,000
190	110,000



Pouvez-vous deviner quelle est la meilleure estimation pour le prix d'une maison de 140 mètres carrés?

A.\$60,000

B.\$95,000

C.\$85,000

X	Y	X-XB	Y-YB	(X-XB)(Y-YB)	(X-XB) ²
30	30,000	-88,75	-40	3550	7876,5625
70	40,000	-48,75	-30	1462,5	2376,5625
90	55,000	-28,75	-15	431,25	826,5625
110	60,000	-8,75	-10	87,5	76,5625
130	80,000	11,25	10	112,5	126,5625
150	90,000	31,25	20	625	975,5625
180	95,000	61,25	25	1531,25	3751,5625
190	110,000	71,25	40	2850	5076,5625
$XB=118,75$	$YB=70$	Moyenne	Somme	10650	21086,5

Regression linéaire Multiple

03

- Dans la régression linéaire simple on a une variable dépendante (y) et une variable indépendante (x)
- Dans la régression linéaire multiple on a plusieurs variables indépendantes (plusieurs x)

- Imaginons que l'on veuille déterminer le prix d'une voiture en fonction de nombre de cylindres et du nombre de portes
- Avec une régression linéaire simple c'est impossible!

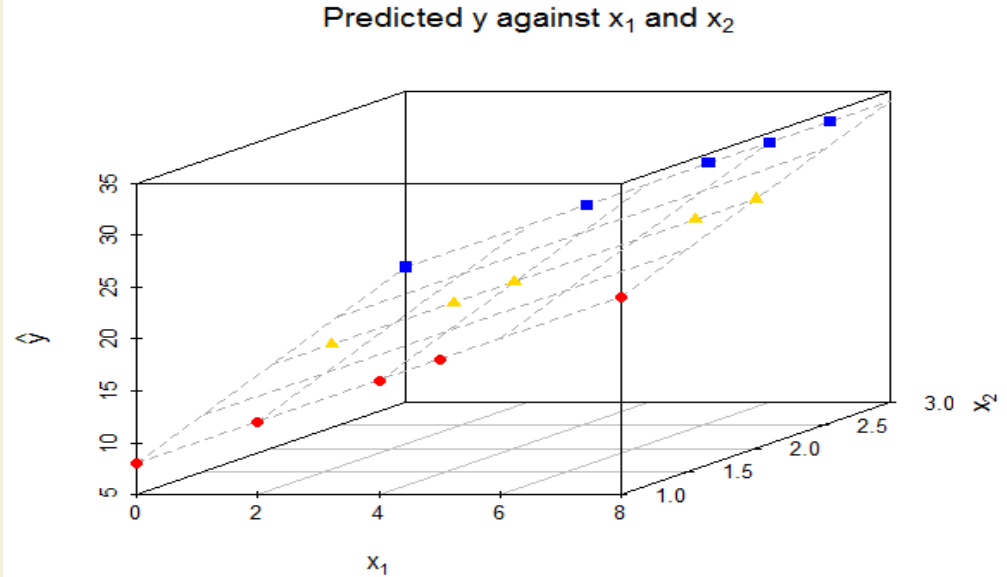
- Dans le cas d'une régression linéaire multivariée, la fonction prédictive s'écrit sous la forme :

$$Y = \alpha x_1 + \beta x_2 + \gamma x_3 + \dots + \omega x + b$$

- On rajoute des variables indépendantes càd on rajoute des dimensions

Et pour que ça soit visuel!

On va juste passer sur deux variables indépendantes (x_1 et x_2)





Simple

On va prédire une droite



Multiple

On va prédire un plan
(avec deux x)

MERCI POUR VOTRE ATTENTION!

Elaboré par:

ABID Ameni

ZOUARI Asma

SADDOUD Oumaima