

# Introduction à la régression

## cours n°1

ENSM.SE – axe MSA

# Présentation

- **Objectifs**
  - Comprendre la problématique générale de la régression.
  - Savoir **mettre en œuvre** la régression linéaire dans un cas simple, mais réaliste.
- **Evaluation**
  - TP le 21 novembre : 20%
  - Etude critique le 23 novembre : 20%
  - TP le 12 décembre : 20%
  - Examen avec documents : 40%

# Calendrier

- mardi 8 novembre (8h15-11h30) :  
introduction, régression, moindres carrés
- mercredi 9 novembre (8h15-11h30) :  
analyse de variance (ANOVA), modèle probabiliste
- lundi 21 novembre (8h15-11h30) :  
TP sur les données de taille
- mardi 22 novembre (8h15-11h30) :  
ANOVA complète, résidus

# Calendrier (suite)

- mercredi 23 novembre (8h15-11h30)

étude critique

- mardi 29 novembre (8h15-11h30)

prévision + régression non paramétrique

- lundi 12 décembre (8h15-9h45)

TP régression non paramétrique

- vendredi 16 décembre (8h15-9h45)

examen

# Problématique de la régression

Lien entre une **réponse**  $y$

et des **prédicteurs**  $x_1, \dots, x_p$

... dans un contexte incertain

... à partir d'expériences

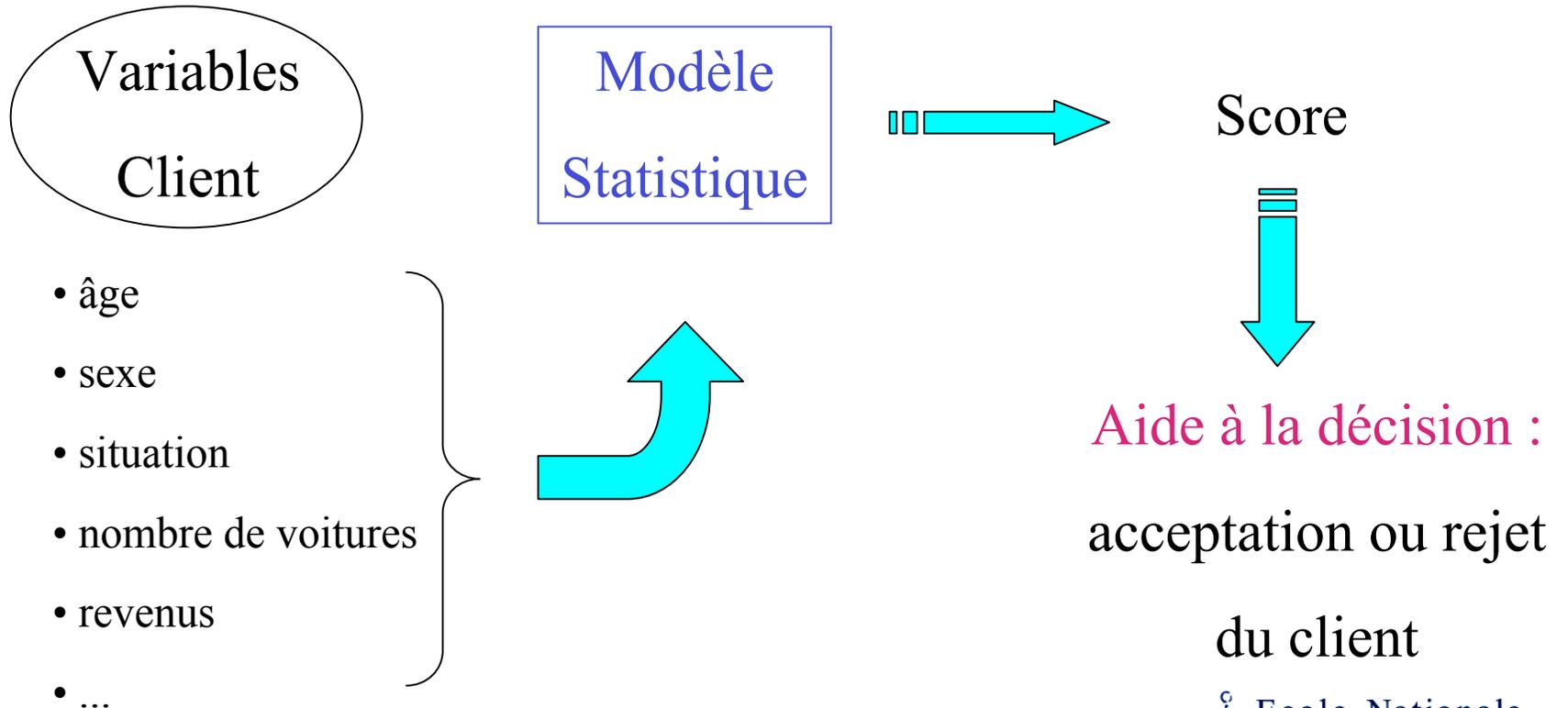
... dans un but **prédictif**

# Classification

- réponse quantitative/qualitative
- régression paramétrique/non paramétrique
- régression linéaire/non linéaire
- prédicteurs contrôlés/non contrôlés

## Exemple 1 : Société de crédit

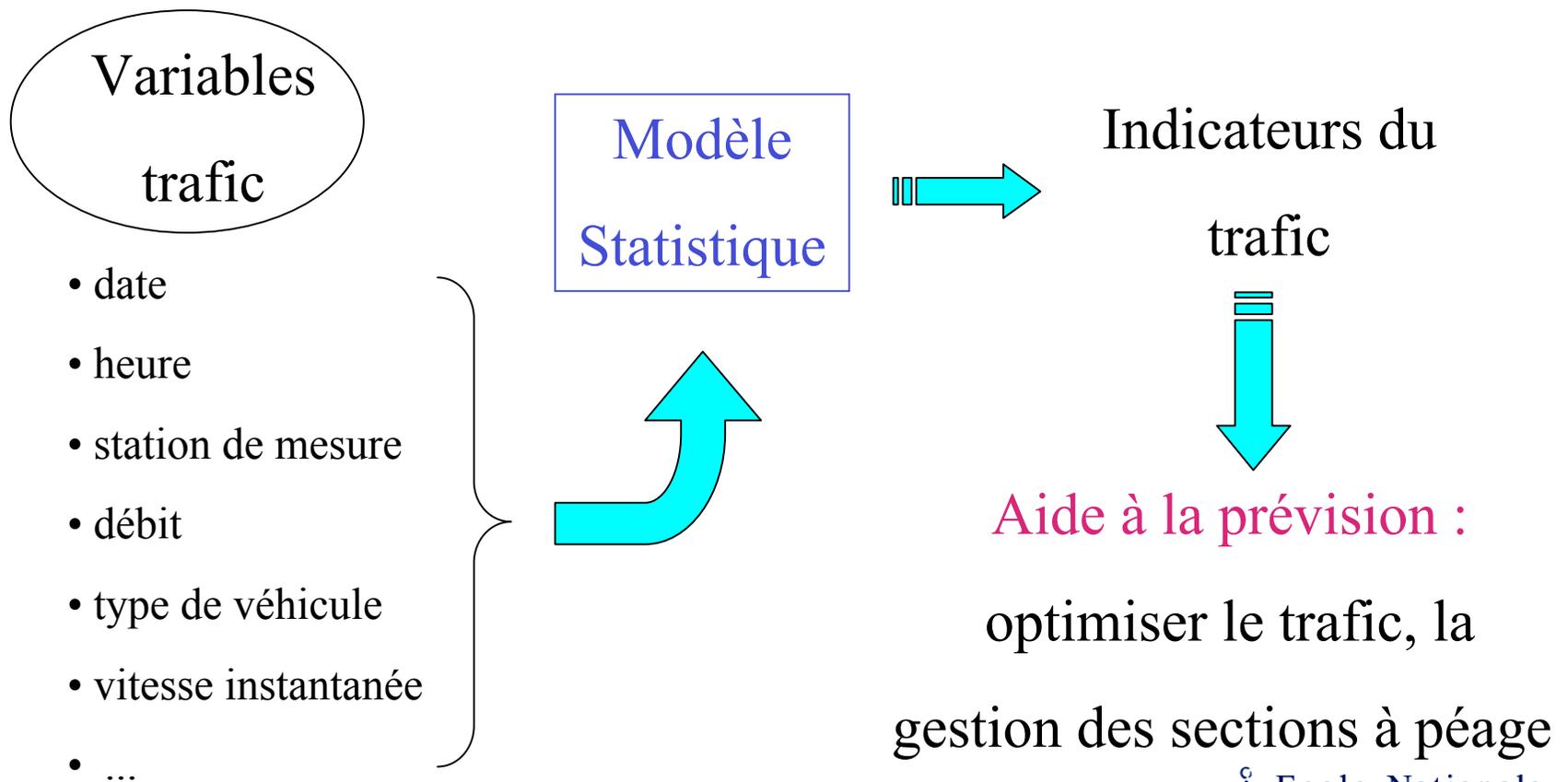
But : cibler au mieux un client potentiel



## Exemple 2 : Société d'autoroute



But : prévoir le trafic sur une section d'autoroute de son réseau



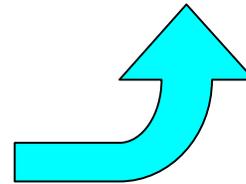
## Exemple 3 : Calcination de la chaux

But : optimiser la réaction

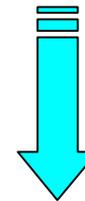
Variables  
four

- température
- pression
- ...

Modèle  
Statistique



quantité produite



aide à la production :  
maximiser la plus-value



# Aspects formels

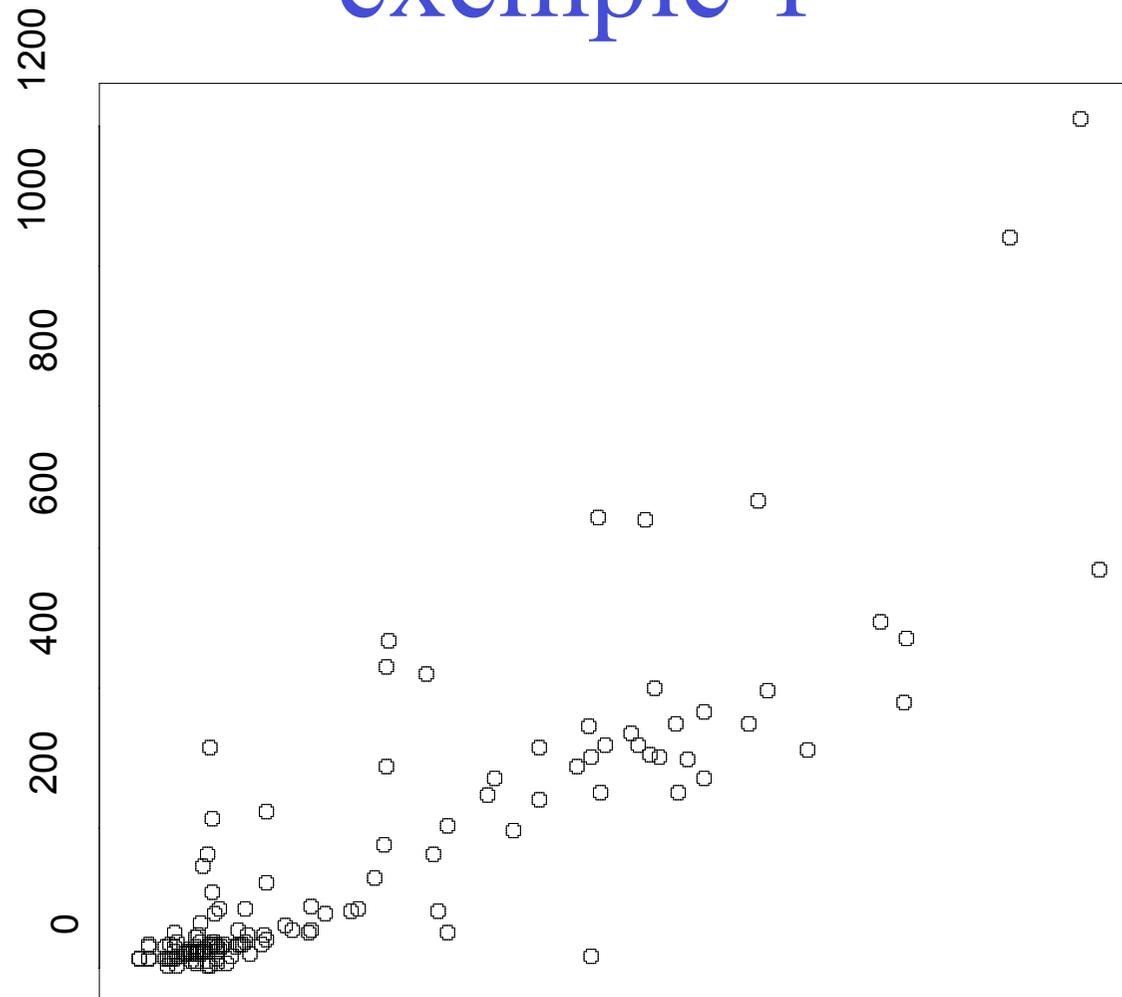
- L'espérance conditionnelle comme espérance
- L'espérance conditionnelle comme projection
- L'espérance conditionnelle linéaire
- Cas gaussien
- Prédicteurs qualitatifs

# Exemple 1

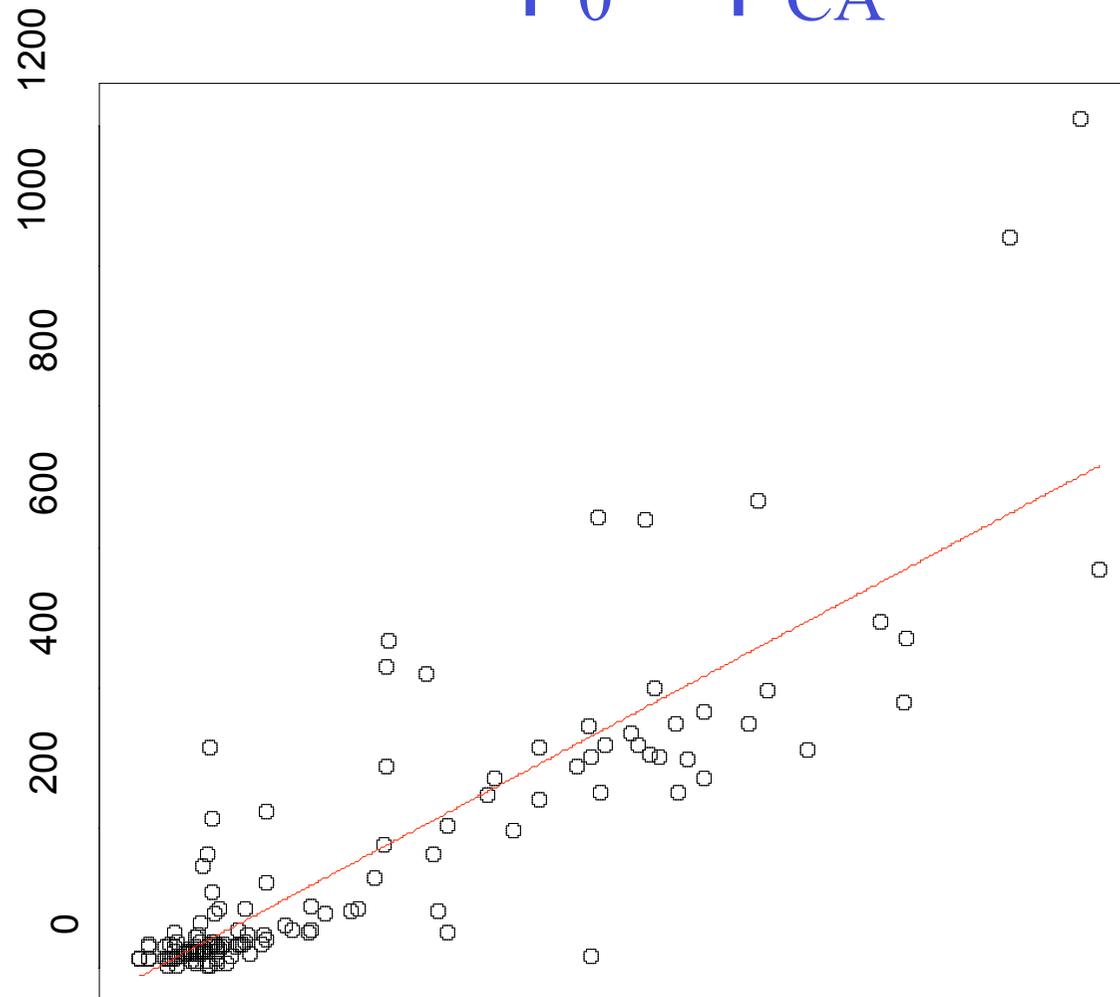
## Récupération de cartons dans la grande distribution

- Réponse = tonnage de cartons annuel
- Prédicteur = CA HT
  - + surface commerciale + prestataire + catégorie
- 102 observations

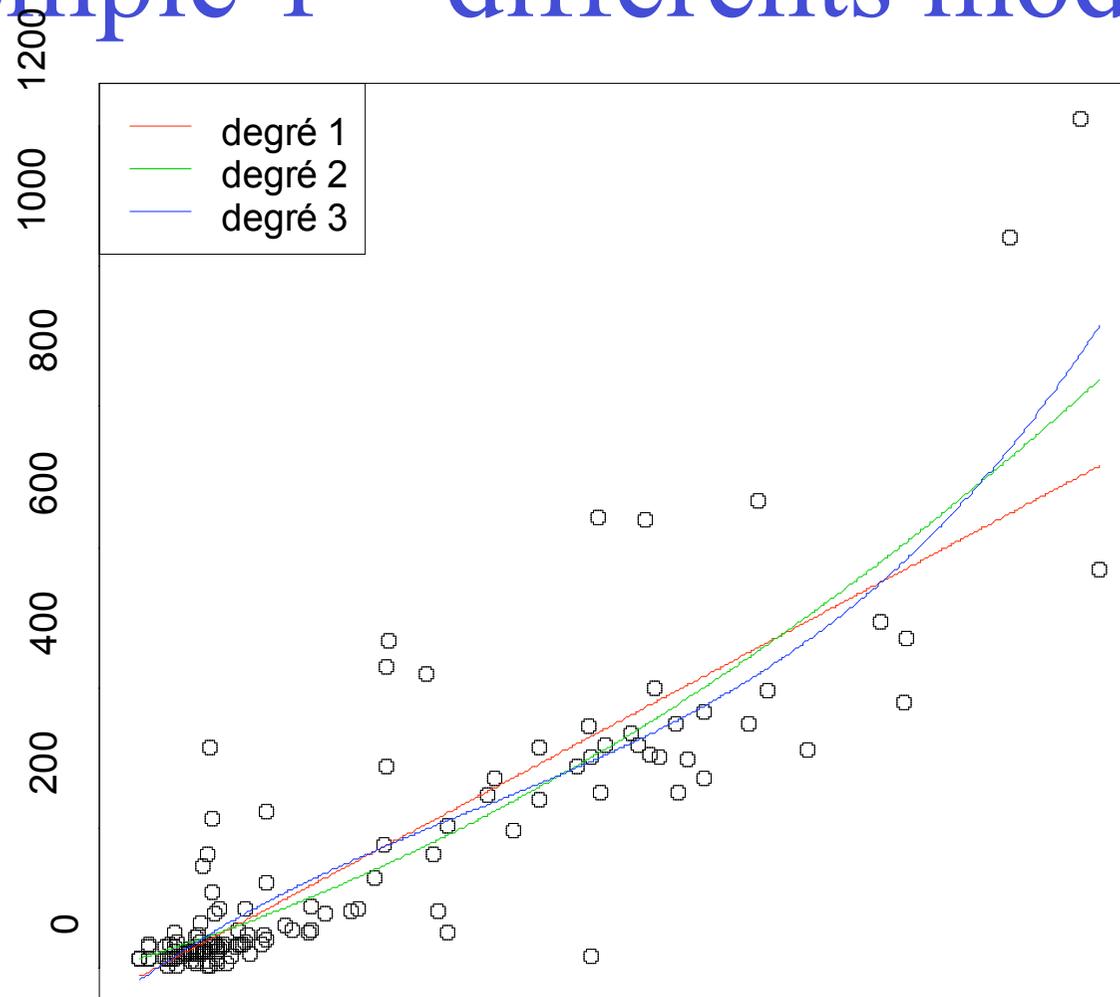
# exemple 1

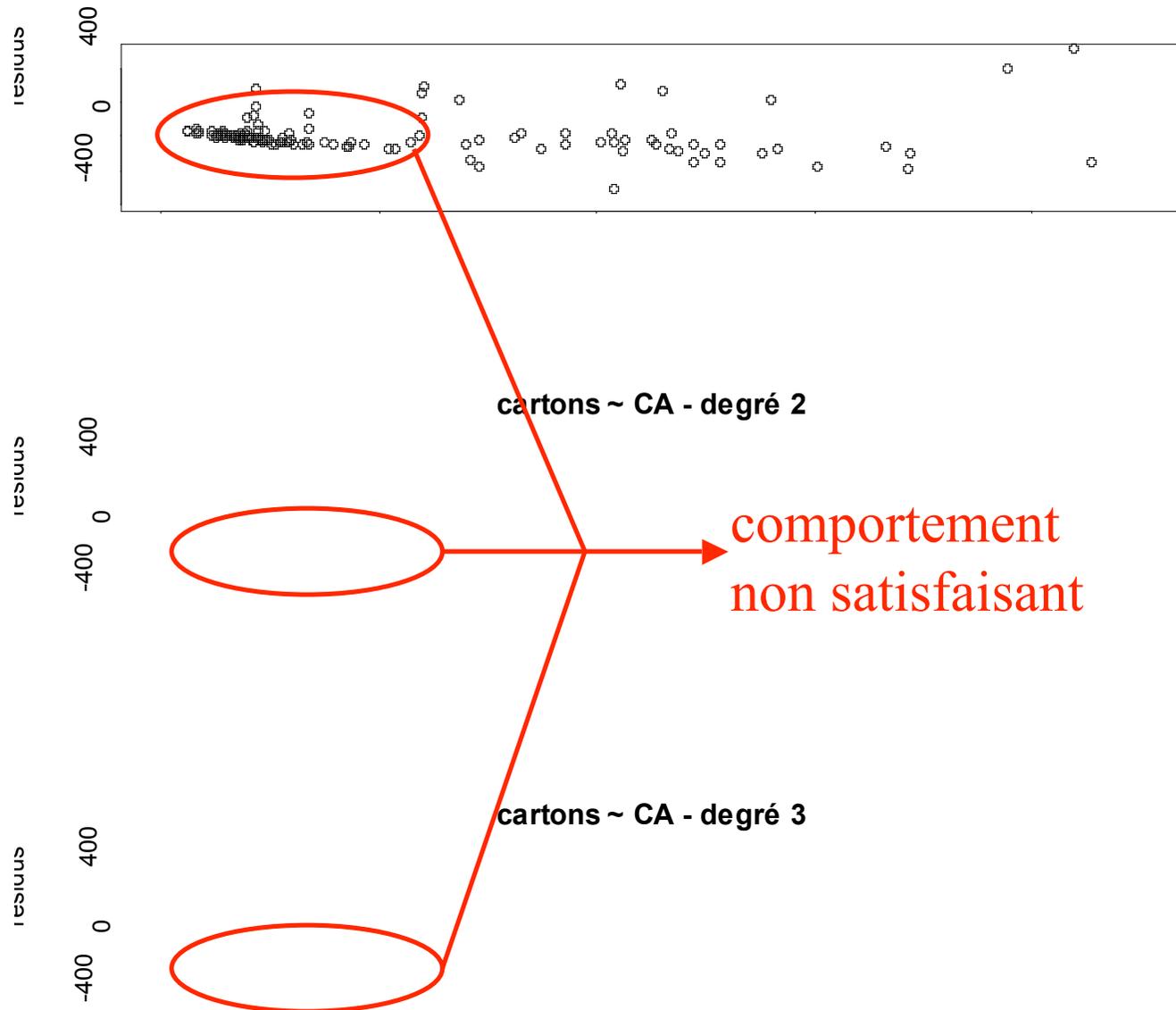


$$\text{cartons} = \beta_0 + \beta_{CA} CA$$



# exemple 1 – différents modèles



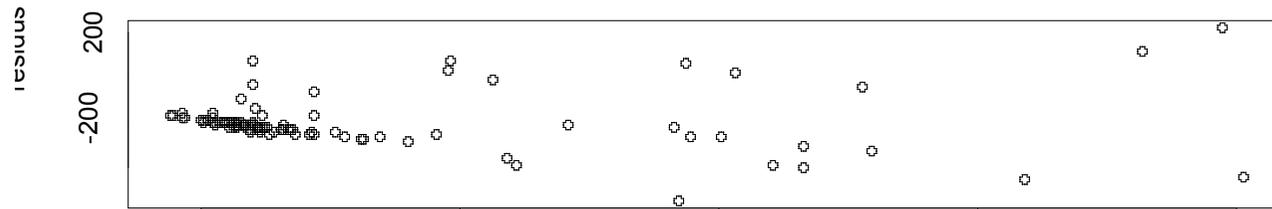


# Questions/objectifs

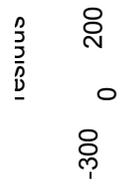
- Estimer "au mieux" l'équation de la droite
- Utilité de la variable CA ?
- Effet des autres variables ?
- Validation
- Précision des résultats
- Prévisions et contrôles
- ...

# Les moindres carrés

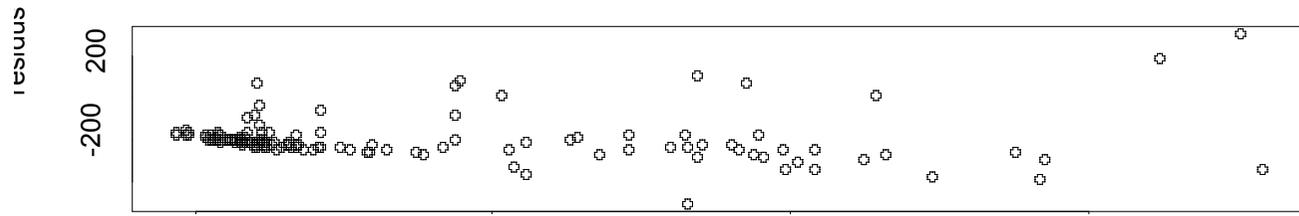
- Notations
- Equation normale
- Interprétation géométrique
- Un premier indicateur de validation



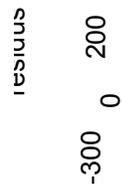
**cartons ~ CA - degré 2**



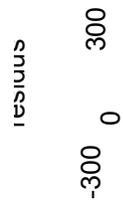
**cartons ~ CA - degré 3**



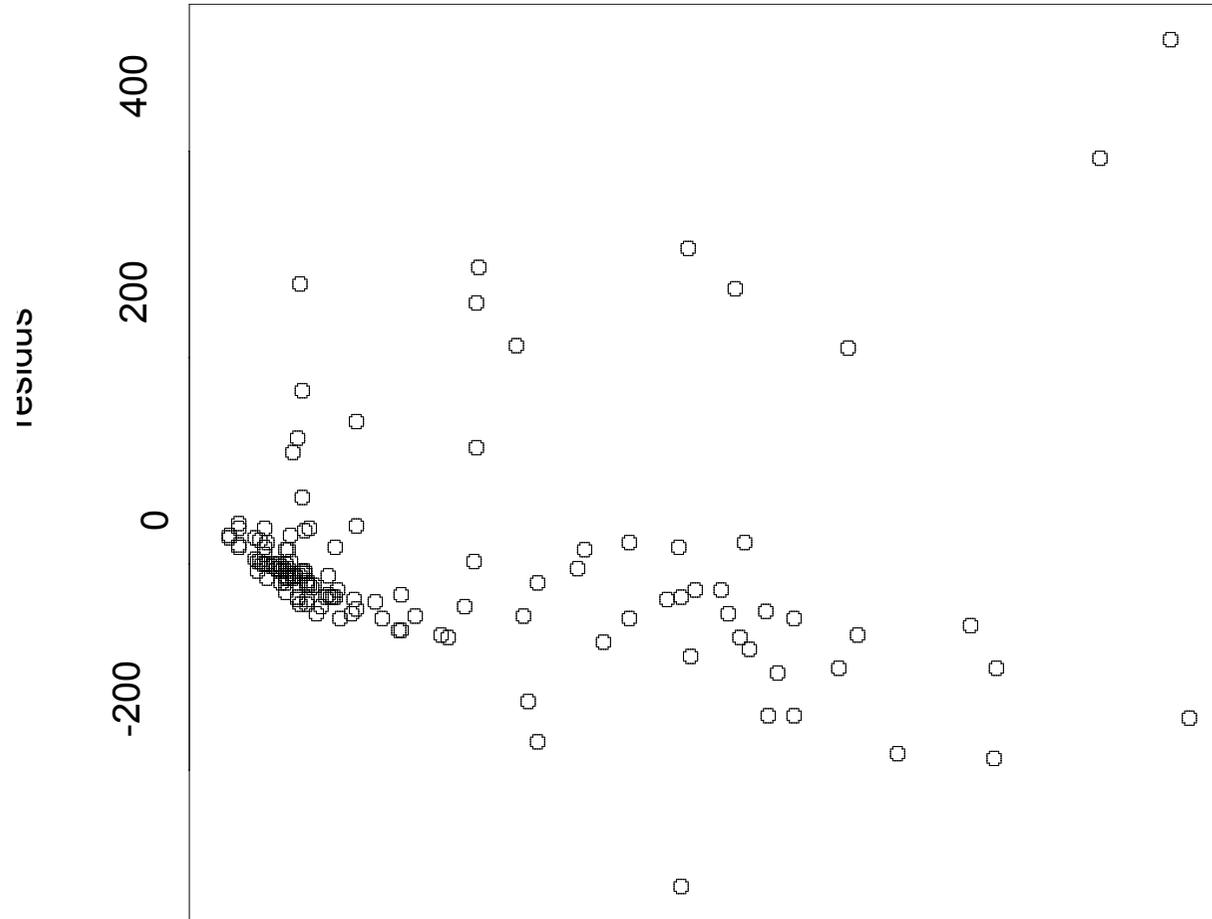
**cartons ~ CA - degré 2**



**cartons ~ CA - degré 3**

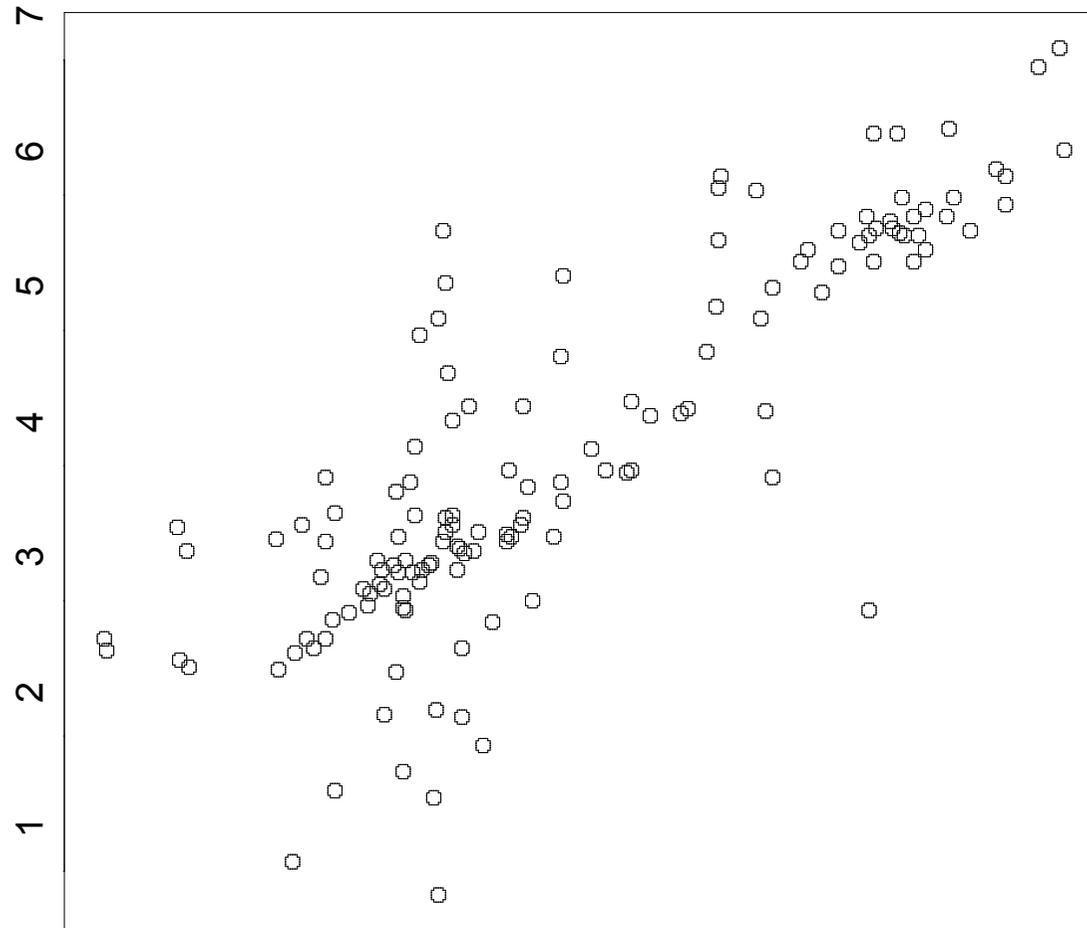


# zoom sur le modèle de degré 1

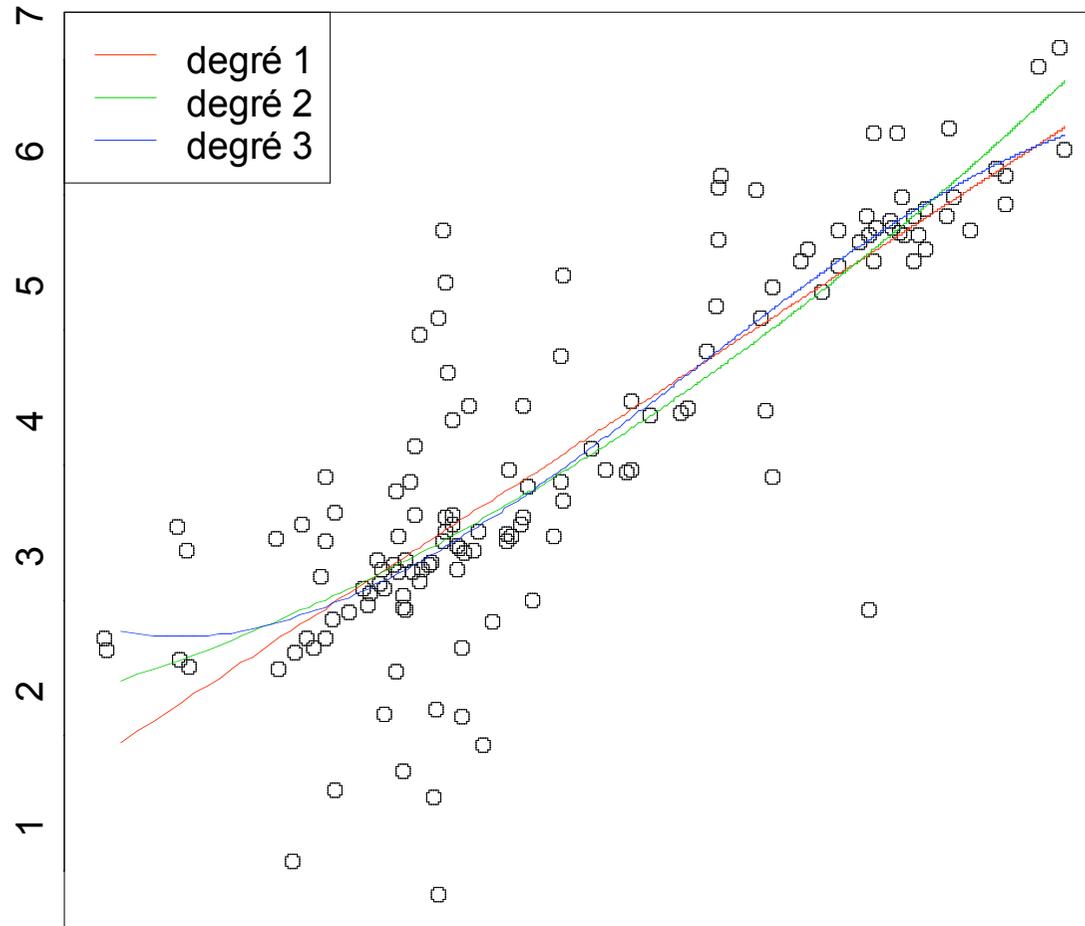


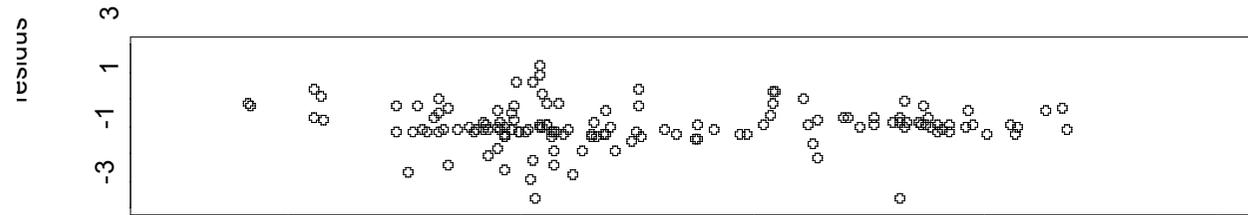
# exemple 1

## après transformation logarithmique

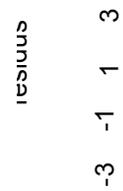


# exemple 1 transformé – modèles

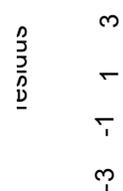


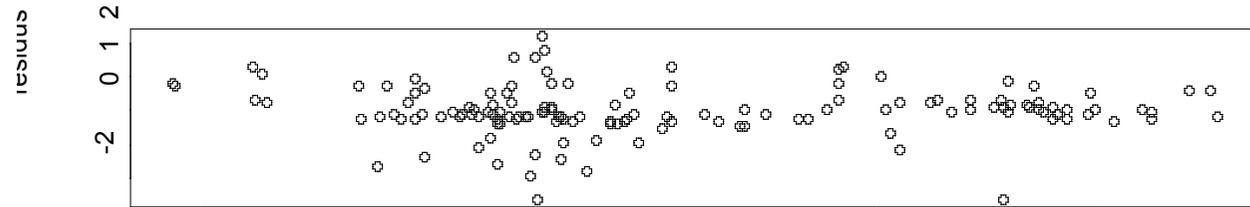


**log(cartons) ~ log(CA) - degré 2**

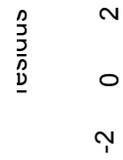


**log(cartons) ~ log(CA) - degré 3**

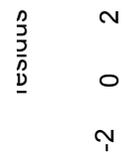




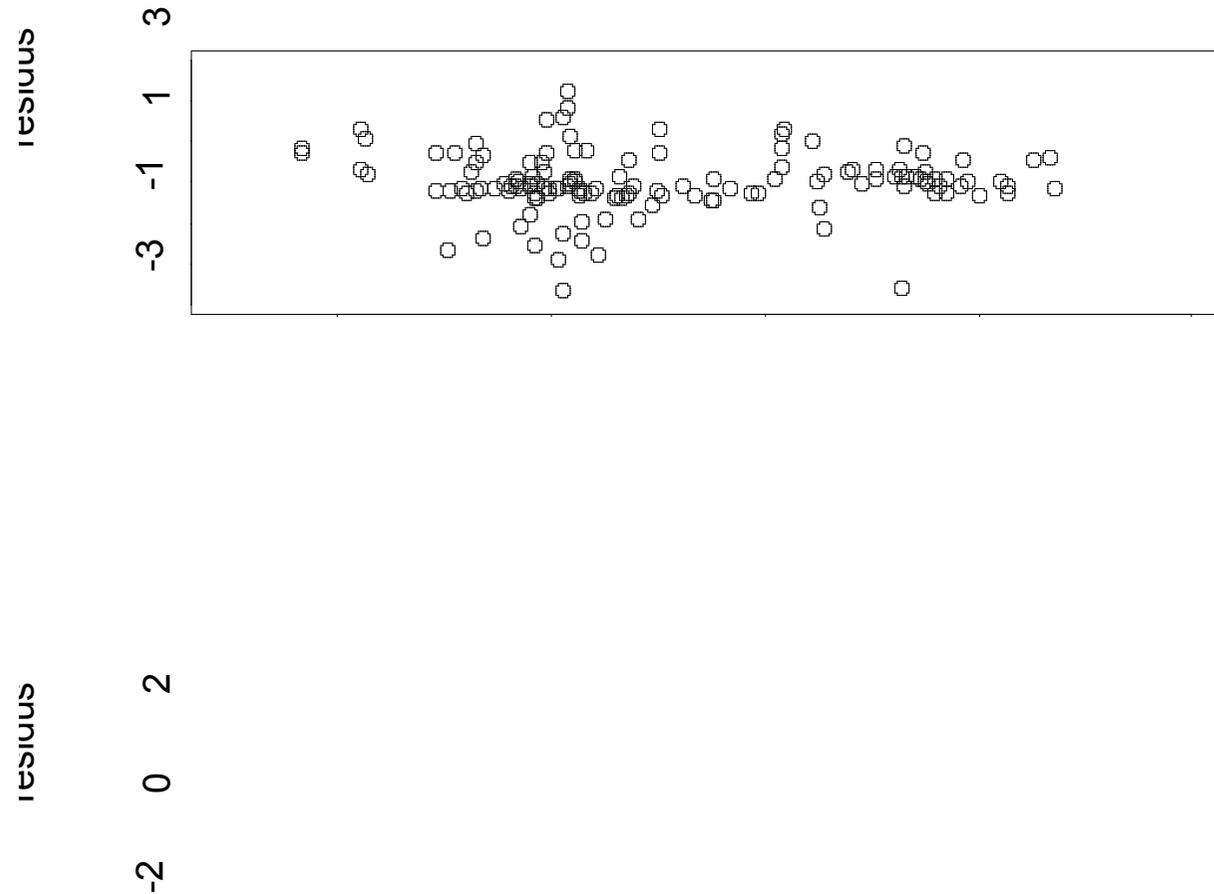
$\log(\text{cartons}) \sim \log(\text{CA})$  - degré 2



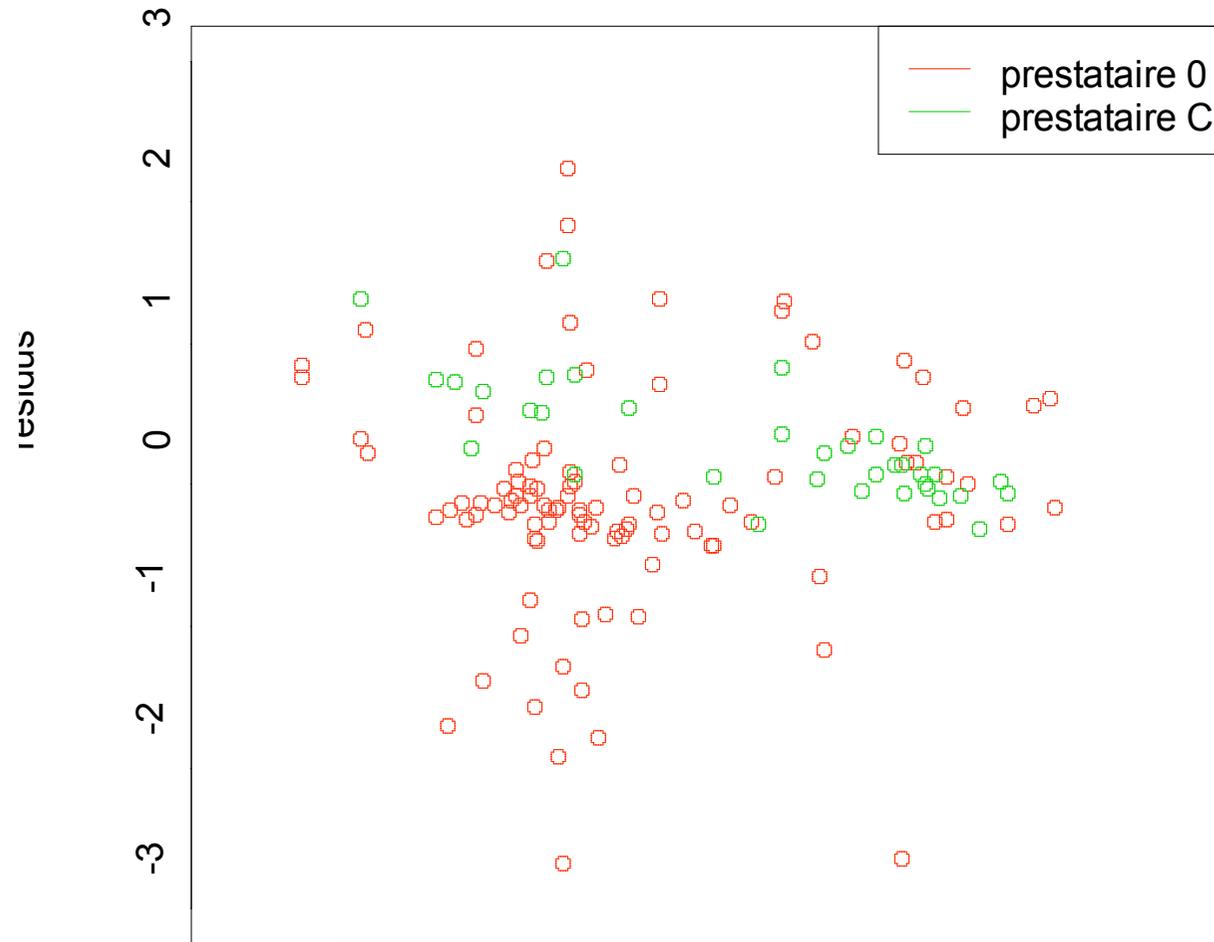
$\log(\text{cartons}) \sim \log(\text{CA})$  - degré 3



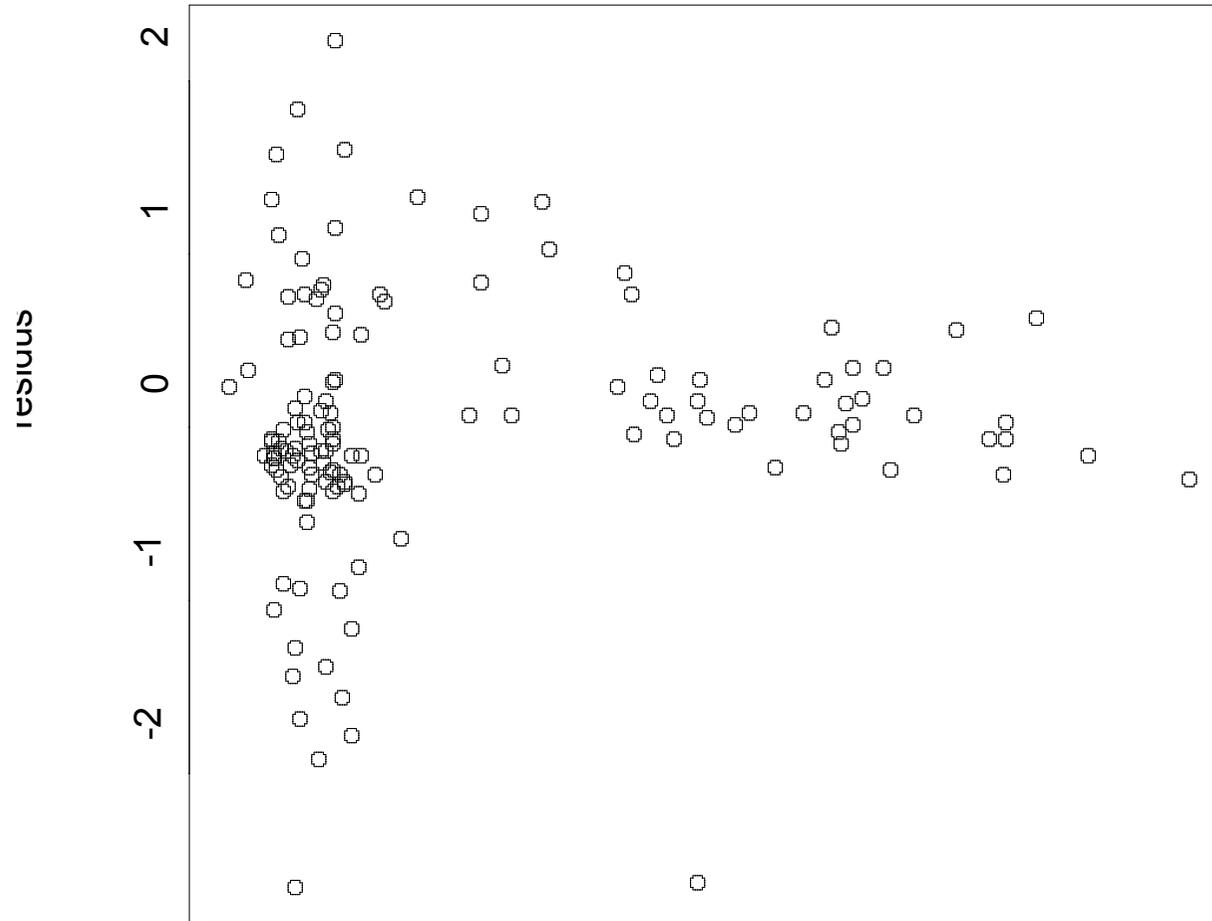
# Zoom sur le modèle de degré 1



# Et les autres variables ?

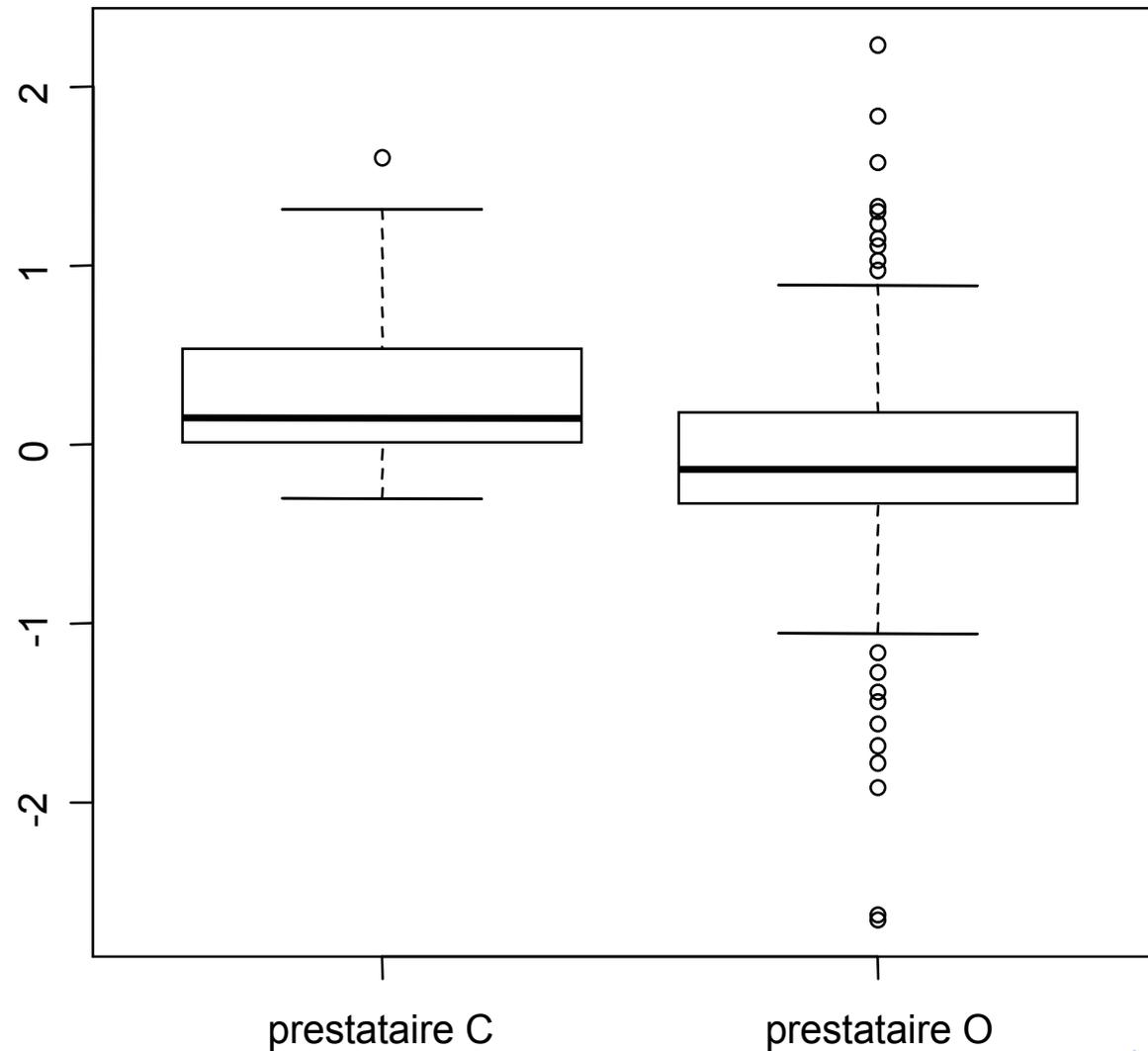


# Variable surface



# Variable prestataire

$\log(\text{cartons}) \sim \log(\text{CA}) - \text{degré } 1$   
résidus



# Variable catégorie

