

# Régression robuste

# Régression non-paramétrique

Ndèye Niang- Gilbert Saporta  
Conservatoire National des Arts et Métiers  
Mars 2014

# Plan

- Introduction: rappels de régression linéaire
  - Les Moindres carrés ordinaires
  - Valeurs aberrantes - Observations influentes
- Régression robuste
  - Notions de robustesse-Estimateurs robustes
  - Régression robuste
- Régression non paramétrique
  - MA / PPV / méthode fenêtre mobile
  - Noyau kernel
  - Régression locale /polynomiale

# Introduction

- Etude de la liaison entre variables quantitatives observées sur les mêmes individus  $Y = f(X) + e$

Y variable dépendante  
variable à expliquer  
variable endogène

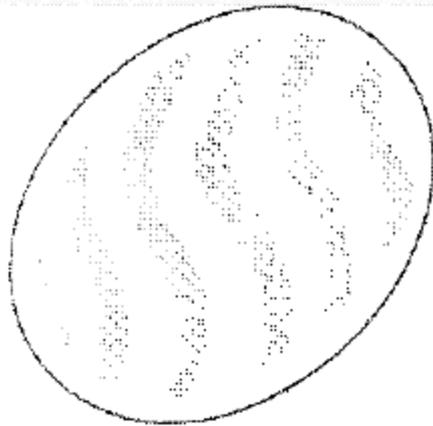
X variables indépendantes  
variable explicatives  
variable exogènes

- Hypothèses:  
f connue , linéaire, estimation des paramètres

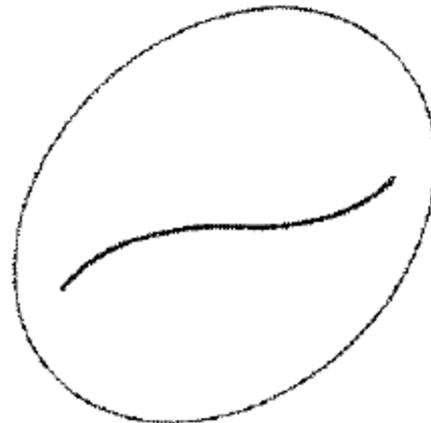
Erreurs iid et gaussiennes, homoscedasticité

# Introduction

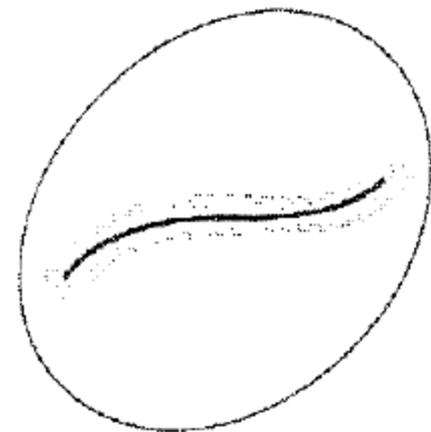
Non paramétrique    paramétrique    robuste



Nonparametric statistics



Parametric statistics



Robust statistics

# Introduction

- Régression linéaire:
- Aspects descriptifs
  - Ajustement linéaire : droite de régression
  - Coefficient  $R^2$
- Aspects inférentiels
  - Estimation des paramètres (MCO- MV)
  - Vérification des hypothèses
  - Tests sur les coefficients
  - Etude de la stabilité des résultats
  - Prévision

# 1. Les moindres carrés

- Hypothèse de régression linéaire:
  - simple  $E(Y/x) = f(x) = \beta_0 + \beta_1 x$
  - multiple:  $E(Y/\mathbf{X}) = f(\mathbf{X}) = \mathbf{X}'\beta$
- Estimation de  $\beta$  à partir d'un échantillon
- Méthode des moindres carrés
- Fonction de perte:  $L(y; f(x)) = (y - f(x))^2$

- Estimateur des moindres carrés

$$\hat{\mathbf{y}} = \mathbf{X}\mathbf{b} = \mathbf{A}\mathbf{y}$$

$$\mathbf{y} - \mathbf{X}\mathbf{b} \perp W \quad (\mathbf{y} - \mathbf{X}\mathbf{b})' \mathbf{X}\mathbf{u} = 0 \quad \forall \mathbf{u}$$

$$\mathbf{X}'\mathbf{y} = \mathbf{X}'\mathbf{X}\mathbf{b} \quad \text{Equations normales}$$

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}$$

$$\text{projecteur } \mathbf{A} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'$$

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - p - 1}$$

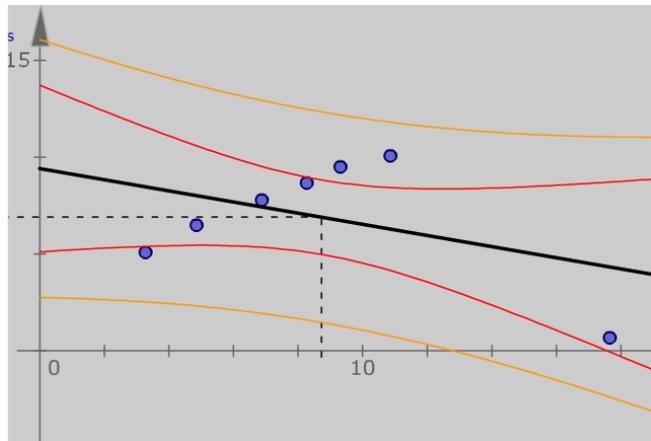
- $\mathbf{b}$  estimateur de variance minimale de  $\beta$  parmi les estimateurs linéaires sans biais (BLUE best linear unbiased estimators)
- estimateur du maximum de vraisemblance si résidus gaussiens iid

$$V(\mathbf{b}) = \sigma^2 (\mathbf{X}'\mathbf{X})^{-1}$$

Estimations imprécises si multicolinéarité

# Prévisions d'une valeur ultérieure

- Intervalle de prévision



- Problèmes d'imprécision , grande variance et surtout sensible aux valeurs aberrantes

- Démo

[cours de Marc Bourdeau \(Montréal\) sur le modèle linéaire](#)

[http://www.agro-montpellier.fr/cnam-lr/statnet/cours\\_autre.htm](http://www.agro-montpellier.fr/cnam-lr/statnet/cours_autre.htm)

## 2. Analyse des résultats

On s'intéresse ici à la stabilité des résultats

Distinction entre questions relatives à l'influence d'observations particulières et questions relatives à l'influence des variables sur les estimations (multicolinéarité)

Analyse des résidus pour vérifier les hypothèses de base du modèle : variance constante et pas autocorrélation

## 2. Analyse des résultats

- Etude des résidus est fondamentale:
  - permet la **détection des valeurs aberrantes et des observations influentes**
  - Vérification empirique des hypothèses:  
graphique des résidus en fonction des  $X$  ne doivent pas laisser apparaître de tendance ou de dépendance dans l'étude de 2 résidus successifs (**Test de Durbin Watson**)  
( sinon estimateurs toujours sans biais mais on perd la variance minimale)

## 2. Analyse des résultats

- Résidu : vecteur  $\mathbf{y} - \hat{\mathbf{y}}$
- Matrice de variance des résidus:
- La « hat matrix » ou projecteur

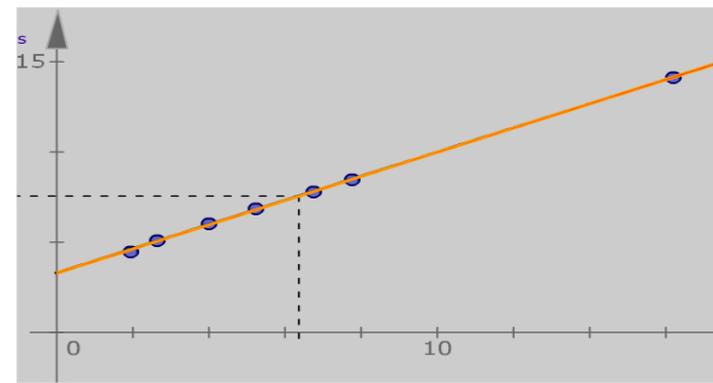
$$\mathbf{A} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \quad \hat{\mathbf{y}} = \mathbf{A}\mathbf{y}$$

- Les termes diagonaux  $h_i$

$$\frac{1}{n} \leq h_i \leq 1 \quad \sum_{i=1}^n h_i = p + 1$$

- Résidu:
  - espérance nulle,
- Résidu studentisé 
$$\frac{y_i - \hat{y}_i}{\hat{\sigma} \sqrt{1 - h_i}}$$
- Si n grand les résidu studentisés doivent être compris entre -2 et 2
- Résidu fort peut indiquer une valeur aberrante mais une valeur peut être aberrante sans que son résidu soit important.

Ne pas confondre observations aberrantes et observations influentes (au sens de l'écart au modèle)



- Etude influence de chaque observation sur les résultats:
  - **Sur sa propre prédiction**: Résidu prédit (en enlevant i)

$$y_i - \hat{y}_{(-i)} = \frac{y_i - \hat{y}_i}{1 - h_i}$$

- Prudence avec les observations à  $h_i$  grand
- **Press**: somme des carrés des résidus prédits est une mesure du pouvoir prédictif du modèle
- **Influence d'une observation sur les estimations des coefficients**: la distance de Cook

$$D_i = \frac{(\mathbf{b} - \mathbf{b}_{(-i)})' \mathbf{X}'\mathbf{X}(\mathbf{b} - \mathbf{b}_{(-i)})}{(p+1)\hat{\sigma}^2} = \frac{1}{p+1} (\hat{y}_i - \hat{y}_{-i})^2 \frac{h_i}{1-h_i}$$

- Devrait rester  $< 1$

- Les méthode de diagnostics classiques sont insuffisantes pour détecter toutes les valeurs aberrantes en particulier multidimensionnelles:
  - effet de masque
  - Effet de débordement (swamping)
  - estimations non robustes
- **Solution en deux étapes:**
  - Eliminer les points aberrants après les avoir détectés
  - Refaire analyse avec estimateurs classiques
- **Solution directe: autres critères que MCO**

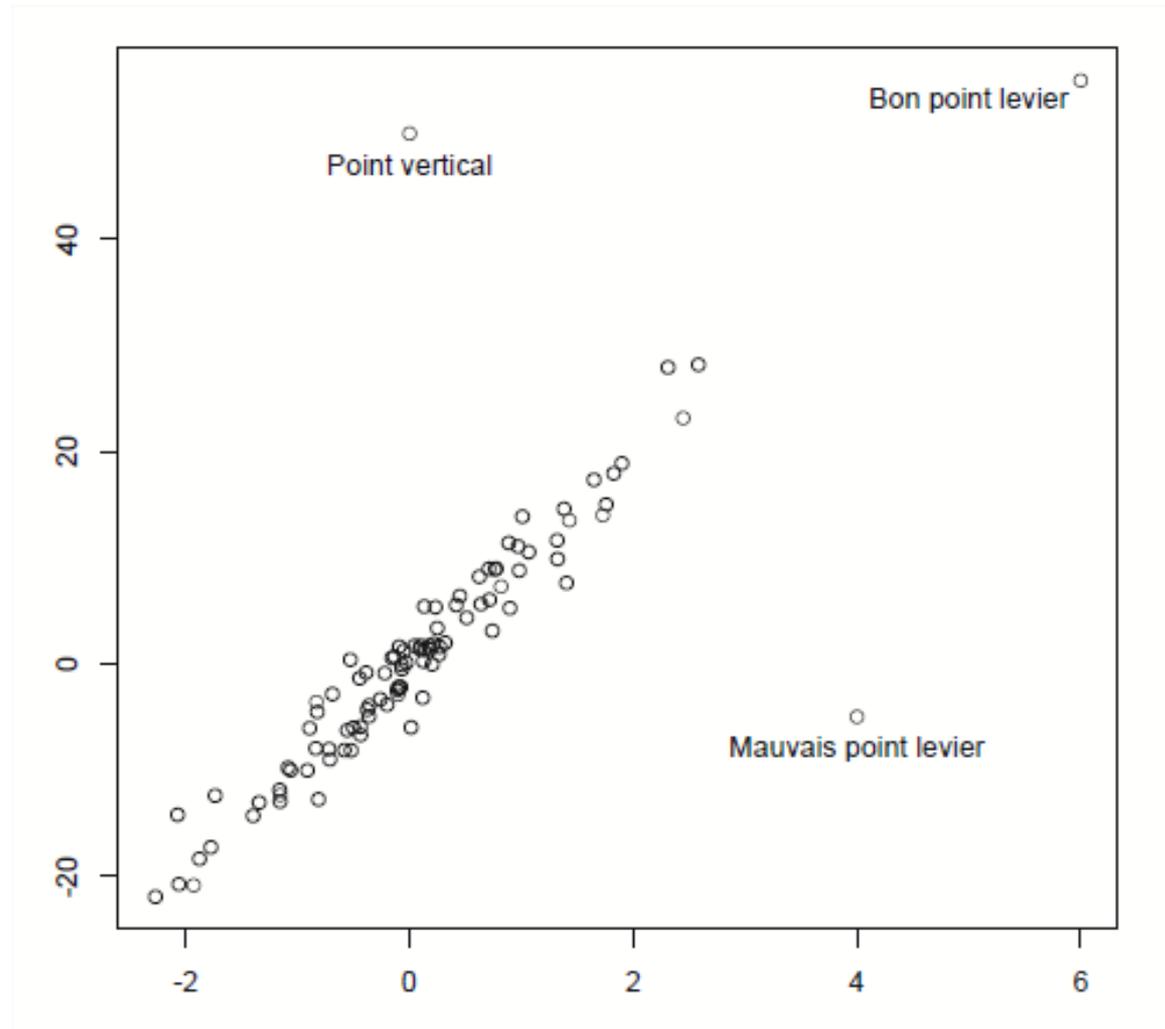
### 3 - Solution en deux étapes

- Détection de valeurs aberrantes
  - Nécessité d'utiliser des indicateurs robustes de tendance centrale et de dispersion pour calculer une variante robuste de la distance de Mahalanobis

### Notions de robustesse

- Différents types de points aberrants
- Point de rupture
- Estimateurs robustes

- Différents types de points aberrants:



# Notion de « point de rupture » d'un estimateur

- Fraction des données qui peuvent être arbitrairement changées sans changer arbitrairement la valeur de l'estimateur.
- Deux cas : n fini, n infini (pt de rupture asymptotique)
  - Ne peut être  $> 0.5$
  - Asymptotiquement:
    - Nul pour la moyenne (si une valeur devient infinie, la moyenne aussi)
    - 0.5 pour la médiane
  - n fini
    - $1/n$  pour la moyenne,  $(n-1)/2n$  pour la médiane,  $(n-h)/h$  pour la régression LTS

# Estimateurs robustes

- **Moyenne tronqué d'ordre  $\alpha$ :**
  - Moyenne arith des observations dont on éliminé les  $\alpha \times n$  plus grandes et petites ( $\alpha = 15\%$ )  
(ex la médiane est un cas extrême  $\alpha = 50\%$  )
- « **Winsorisation** »
  - Au lieu de les éliminer on les prend toutes égales au dernières valeurs prises
- Autre approche: M-**estimateur**

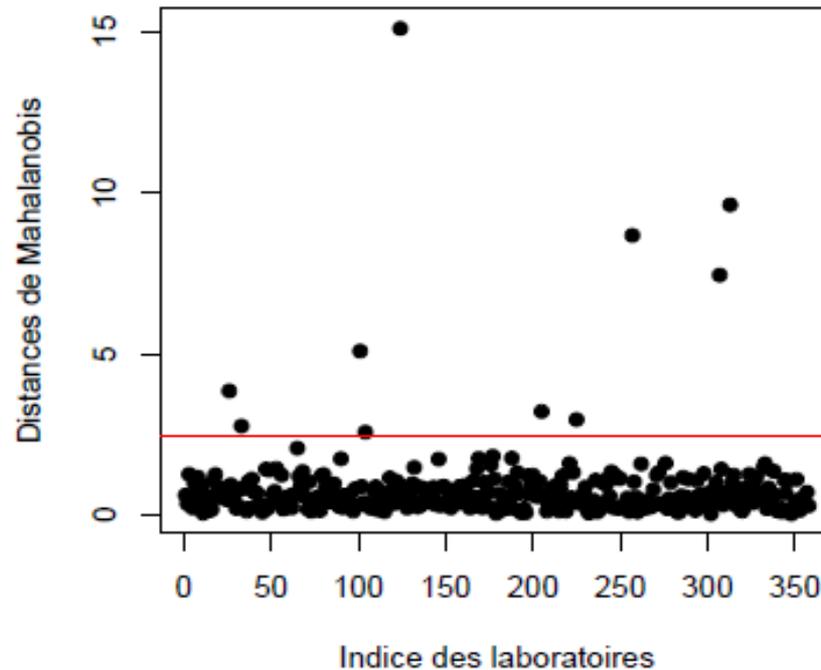
# Solution en deux étapes:

- Détection des points aberrants avec les **estimateurs classiques**:
  - Soit  $x_1, x_2, \dots, x_n$ , un échantillon de taille  $n$ 
    - Calculer  $\bar{x}$  et  $S$
    - Pour chaque observation  $x_i$  calculer la distance de Mahalanobis

$$d_i^2 = (x_i - \bar{x})' S^{-1} (x_i - \bar{x})$$

- Détecter celles dont les distances sont « trop » grandes c'est-à-dire supérieures au quantile 97,5% de la loi du chi carré à  $p$  degrés de liberté

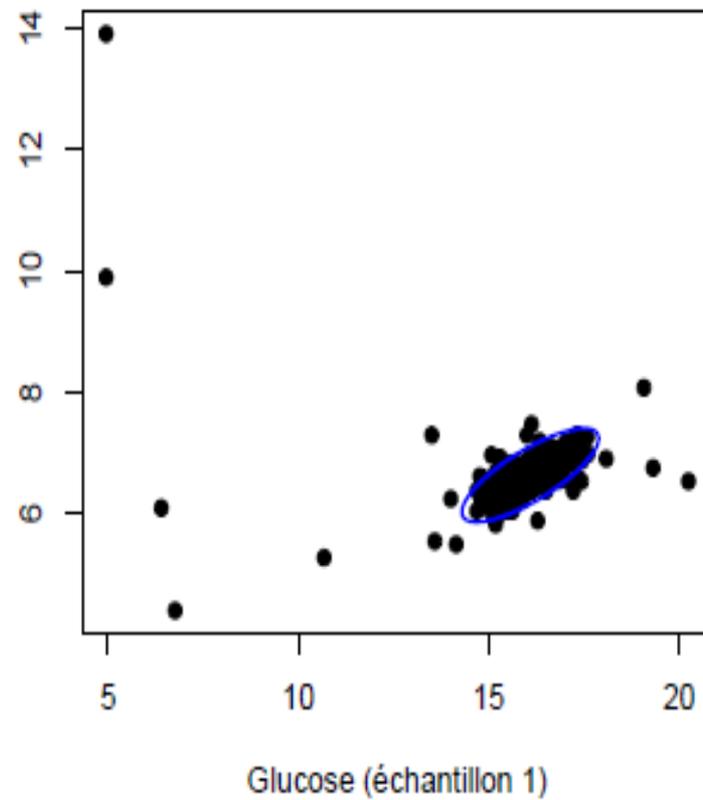
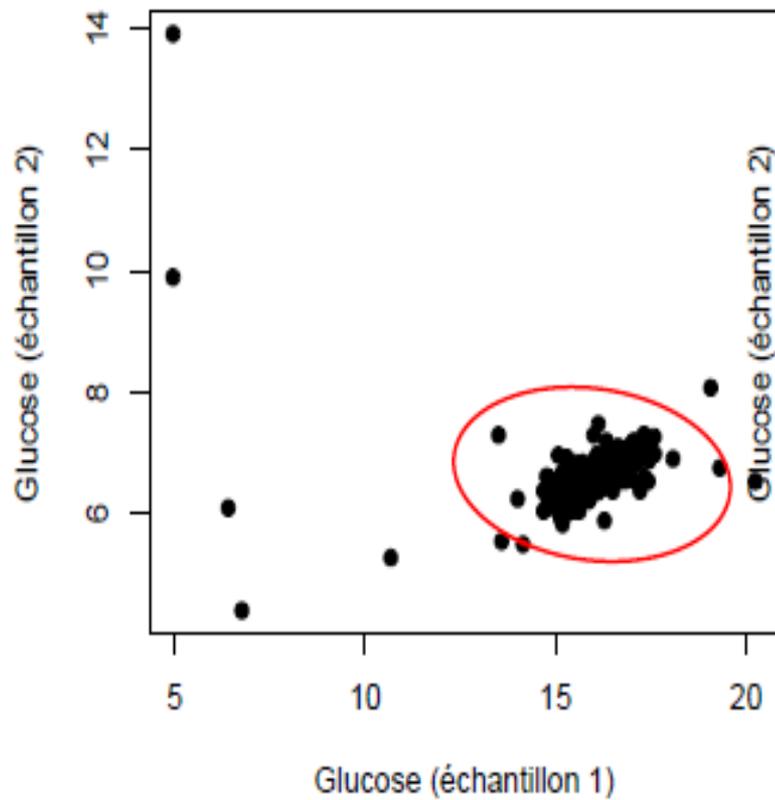
- Solutions en deux étapes :



D'après Haesbroeck

- problème:  $\bar{x}$  et  $S$  estimations **non robustes** influencées par les points aberrants

# Masking effect



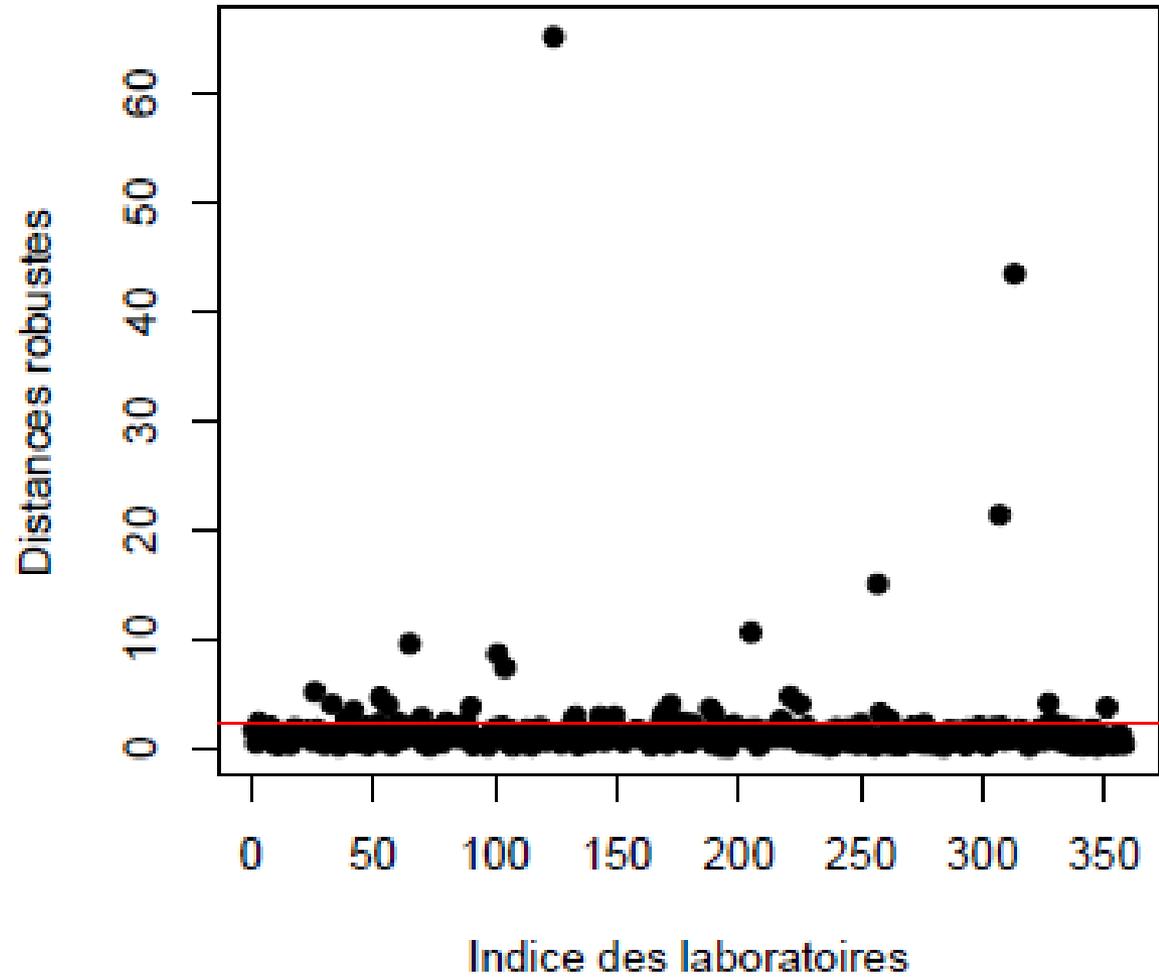
D'après Haesbroeck

- Solution en deux étapes:
- Détection des points aberrants avec les **Estimateurs robustes**:
  - Calculer des estimations robustes  $\hat{\mu}$  et  $\hat{\Sigma}$  de la moyenne et de la matrice de covariance  $\mu$  et  $\Sigma$
  - Calculer pour chaque observation  $x_i$  la distance robuste

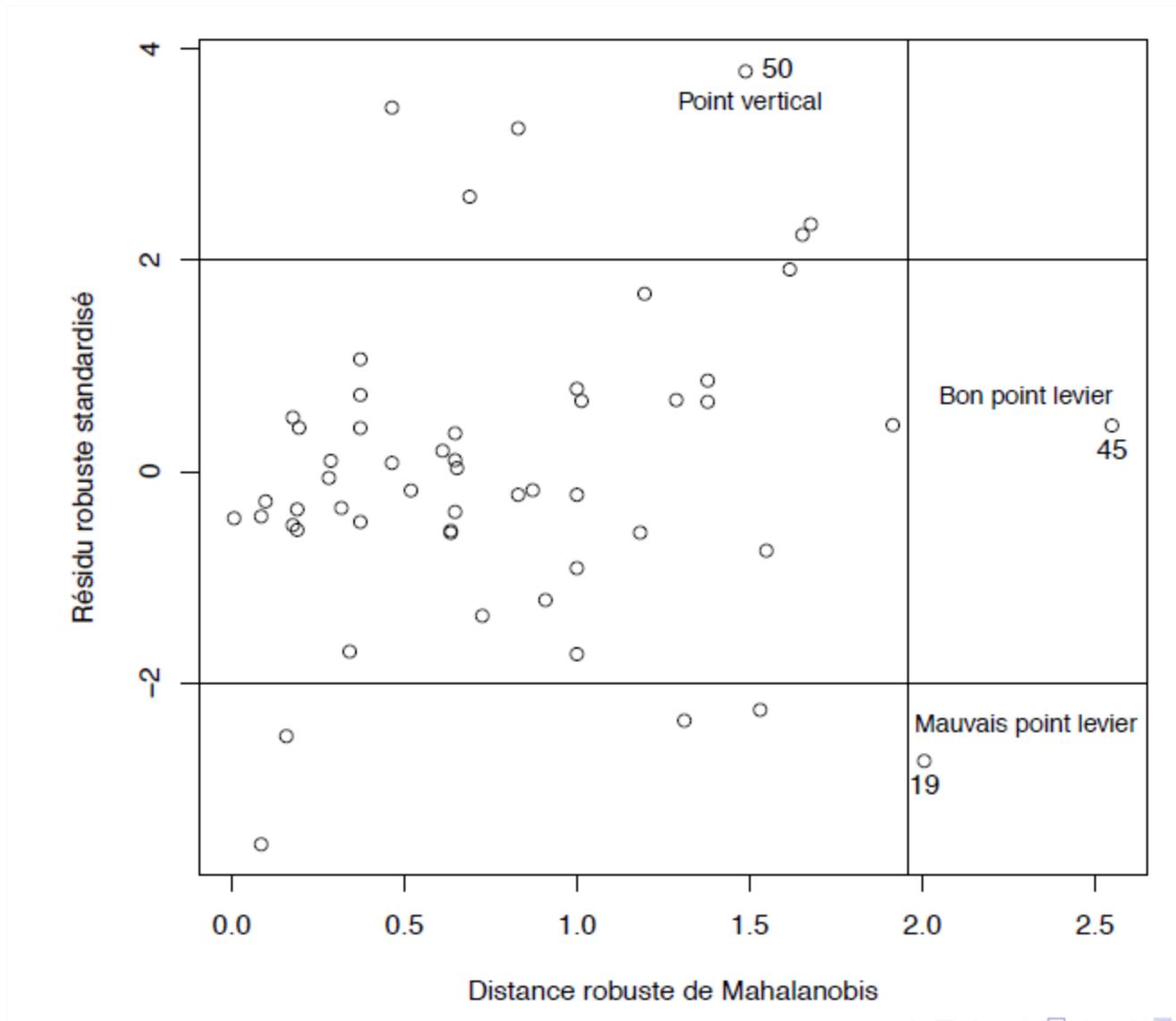
$$d_i^2 = (x_i - \hat{\mu})' \hat{\Sigma}^{-1} (x_i - \hat{\mu})$$

- Détection des observations aberrantes, celles dont les distances sont « trop » grandes c'est-à-dire supérieures au quantile 97,5% de la loi du chi carré à  $p$  degrés de liberté

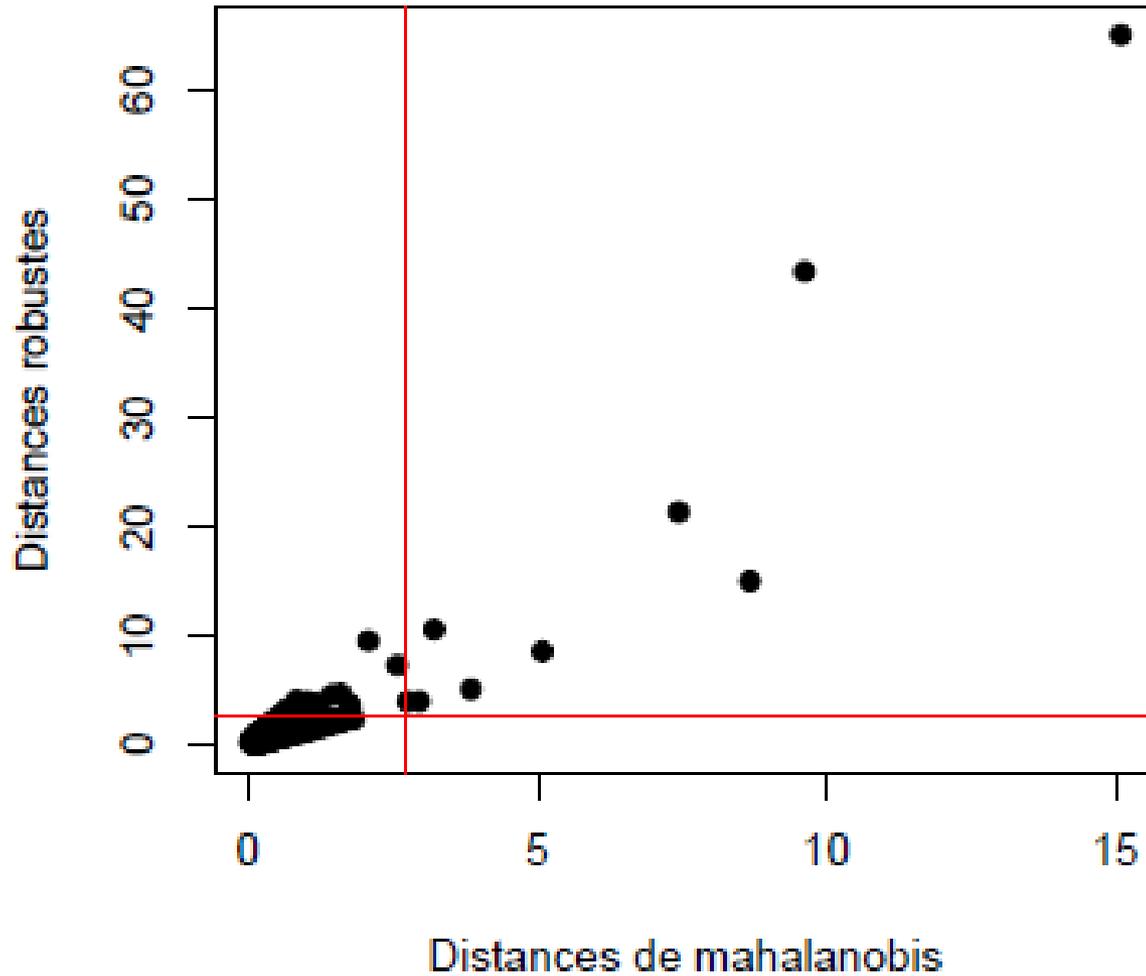
- Solutions:



D'après Haesbroeck



- Outil graphique DD plot



D'après Haesbroeck

- **Solution directe :**
  - Utiliser d'autres critères que les moindres carrés :
    - Régression L1 avec les valeurs absolues
    - M régression avec des M-estimateurs de Huber
    - Régression LST (Least trimmed squares) de Rousseuw

# 4. Régression en norme $L_1$ (LAD)

- Utiliser les valeurs absolues

$$L(y; f(x)) = |y - f(x)|$$

- Régression linéaire simple sur échantillon

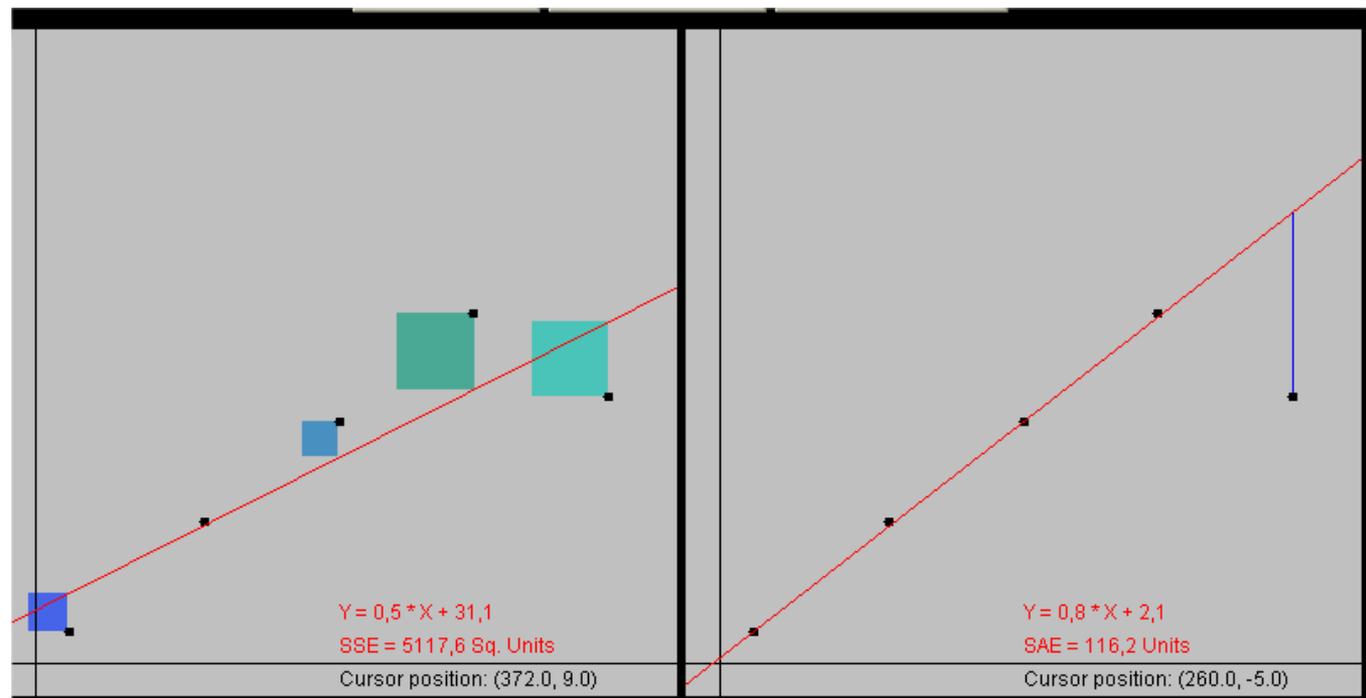
$$\min \sum_{i=1}^n |y_i - a - bx_i|$$

- Propriété:
  - La droite de régression  $L_1$  passe par deux des points de l'échantillon mais on ne peut pas savoir à l'avance lesquels
- 50 ans plus ancien que les moindres carrés (Boscovitch 1757)

## 4 - Régression en norme $L_1$ (LAD)

- Résolution plus difficile (surtout en régression multiple)
  - Pas de solution analytique
  - Nécessité d'un algorithme spécifique
- Algorithme LAD simple (Birkes & Dodge, 1993)
- Possibilité de non-unicité ou de dégénérescence
- Régression LAD multiple
  - Programmation linéaire
  - $p+1$  résidus sont nuls

- [http://www.math.wpi.edu/Course\\_Materials/SAS/tablets/7.3/7.3c/index.html](http://www.math.wpi.edu/Course_Materials/SAS/tablets/7.3/7.3c/index.html) voir si ça marche sur carri java



- Erreurs standard asymptotiques

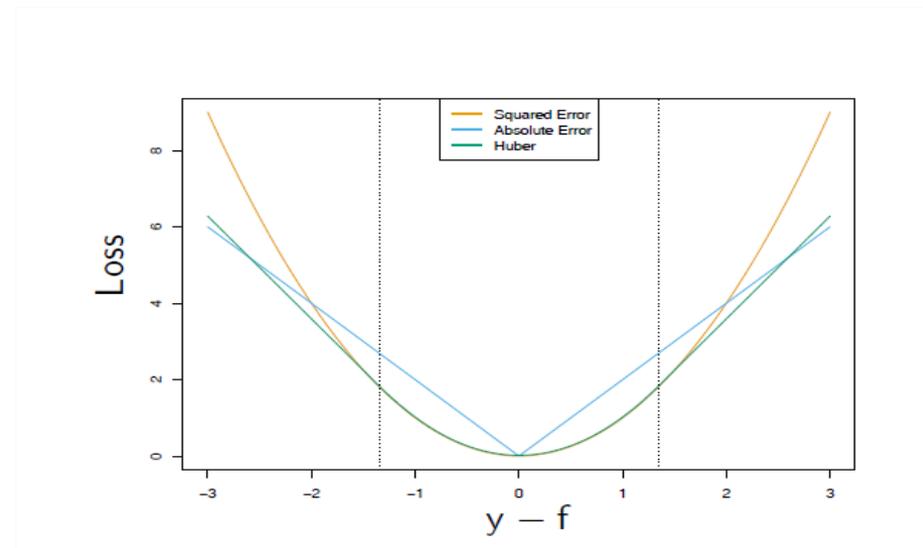
$$V(\hat{\boldsymbol{\beta}}) ; \frac{1}{4(f'(0))^2} (\mathbf{X}'\mathbf{X})^{-1}$$

avec  $f$  densité de  $\varepsilon$

- Maximum de vraisemblance avec densité des erreurs loi de Laplace

# 5 - M-régression

- Issue des M-estimateurs de **Huber**
- Exemple: fonction de perte quadratique jusqu'à  $c$ , linéaire au-delà
- Pour  $c$  grand, on retrouve les mco, pour  $c=0$  la régression  $L_1$



- $\min \sum_{i=1}^n \rho(y_i - \mathbf{x}_i \boldsymbol{\beta})$      $\rho$  fonction convexe paire
  - M-estimateur: maximum de vraisemblance avec erreurs de densité proportionnelle à  $\exp(-\rho(u))$
- En dérivant:  $\sum_{i=1}^n \rho'(y_i - \mathbf{x}_i \hat{\boldsymbol{\beta}}) \mathbf{x}'_i = 0$
- Notation usuelle  $\psi = \rho'$
- Moindres carrés pondérés

$$\sum_{i=1}^n w_i (y_i - \mathbf{x}_i \hat{\boldsymbol{\beta}}) \mathbf{x}'_i = 0 \quad \text{avec } w_i = \frac{\rho'(y_i - \mathbf{x}_i \hat{\boldsymbol{\beta}})}{y_i - \mathbf{x}_i \hat{\boldsymbol{\beta}}}$$

# Propriétés

- Si  $\sigma$  est inconnu on minimise  $\sum_{i=1}^n \left[ \rho \left( \frac{y_i - \mathbf{x}_i \boldsymbol{\beta}}{\sigma} \right) + a \right] \sigma$   
avec  $a > 0$
- La matrice de covariance des estimateurs est asymptotiquement proportionnelle à celle des moindres carrés

## Asymptotic Covariance and Confidence Intervals

The following three estimators of the asymptotic covariance of the robust estimator are available in PROC ROBUSTREG:

$$\text{H1: } K^2 \frac{[1/(n-p)] \sum (\psi(r_i))^2}{[(1/n) \sum (\psi'(r_i))]^2} (X^T X)^{-1}$$

$$\text{H2: } K \frac{[1/(n-p)] \sum (\psi(r_i))^2}{[(1/n) \sum (\psi'(r_i))]} W^{-1}$$

$$\text{H3: } K^{-1} \frac{1}{(n-p)} \sum (\psi(r_i))^2 W^{-1} (X^T X) W^{-1}$$

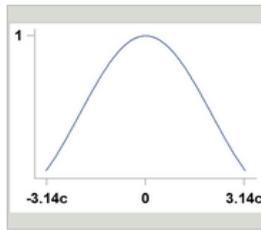
where  $K = 1 + \frac{\rho}{n} \frac{\text{Var}(\psi')}{(E\psi')^2}$  is a correction factor and  $W_{jk} = \sum \psi'(r_i) x_{ij} x_{ik}$ . Refer to Huber (1981, p. 173) for more details.

# Fonction de poids

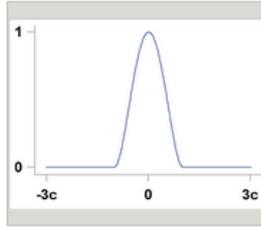
## Proc ROBUSTREG de SAS

Weight Function	Option	Default a, b, c
andrews	WF=ANDREWS<(C=c)>	1.339
bisquare	WF=BISQUARE<(C=c)>	4.685
cauchy	WF=CAUCHY<(C=c)>	2.385
fair	WF=FAIR<(C=c)>	1.4
hampel	WF=HAMPEL<( <A=a> <B=b> <C=c>)>	2, 4, 8
huber	WF=HUBER<(C=c)>	1.345
logistic	WF=LOGISTIC<(C=c)>	1.205
median	WF=MEDIAN<(C=c)>	0.01
talworth	WF=TALWORTH<(C=c)>	2.795
welsch	WF=WELSCH<(C=c)>	2.985

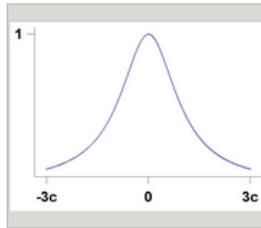
$$\text{andrews } W(x,c) = \begin{cases} \frac{\sin(\frac{x}{c})}{\frac{x}{c}} & \text{if } |x| \leq \pi c \\ 0 & \text{otherwise} \end{cases}$$



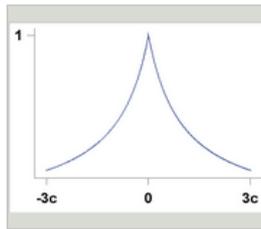
$$\text{bisquare } W(x,c) = \begin{cases} (1 - (\frac{x}{c})^2)^2 & \text{if } |x| < c \\ 0 & \text{otherwise} \end{cases}$$



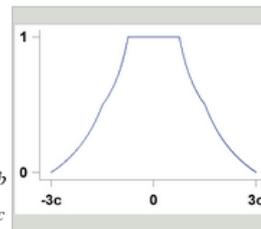
$$\text{cauchy } W(x,c) = \frac{1}{1 + (\frac{x}{c})^2}$$



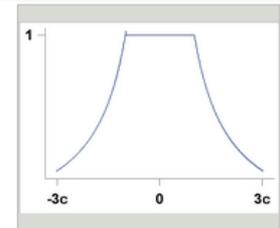
$$\text{fair } W(x,c) = \frac{1}{(1 + \frac{|x|}{c})^2}$$



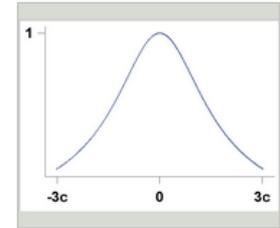
$$\text{hampel } W(x,a,b,c) = \begin{cases} 1 & |x| < a \\ \frac{a}{|x|} & a < |x| \leq b \\ \frac{a}{|x|} \frac{c-|x|}{c-b} & b < |x| \leq c \\ 0 & \text{otherwise} \end{cases}$$



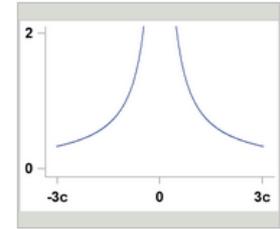
$$\text{huber } W(x,c) = \begin{cases} \frac{1}{c} & \text{if } |x| < c \\ \frac{1}{|x|} & \text{otherwise} \end{cases}$$



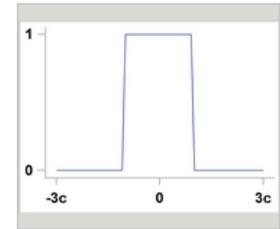
$$\text{logistic } W(x,c) = \frac{\tanh(\frac{x}{c})}{c}$$



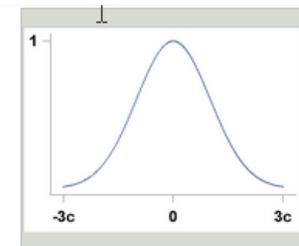
$$\text{median } W(x,c) = \begin{cases} \frac{1}{c} & \text{if } x = 0 \\ \frac{1}{|x|} & \text{otherwise} \end{cases}$$



$$\text{talworth } W(x,c) = \begin{cases} 1 & \text{if } |x| < c \\ 0 & \text{otherwise} \end{cases}$$



$$\text{welsch } W(x,c) = \exp(-\frac{1}{2}(\frac{x}{c})^2)$$



# 6. Régression LTS (Least Trimmed Squares) de Rousseuw

- Basée sur le sous ensemble de  $h$  individus (parmi  $n$ ) où les mco donnent la plus petite somme des carrés des résidus.  $h$  est choisi entre  $n/2$  et  $n$ . La valeur  $h=(3n+p+1)/4$  est recommandée.
- Illustration sur un exemple avec la proc `robustreg` de SAS
- Sous R package `robustbase`
- <http://cran.r-project.org/web/packages/robustbase/robustbase.pdf>

# Bibliographie

- Birkes, D., Dodge, Y. (1993) *Alternative methods of regression*, Wiley
- Rousseeuw, P.J. and Leroy, A.M. (1987), *Robust Regression and Outlier Detection*, New York: John Wiley & Sons, Inc.
- Marona R.A., Martin R.D. and Yohai V.J. (2006) *Robust Statistics*, John Wiley & Sons Ltd.,

# Régression non-paramétrique (univariée)

Gilbert Saporta

Conservatoire National des Arts et Métiers

<http://cedric.cnam.fr/~saporta>

Décembre 2012

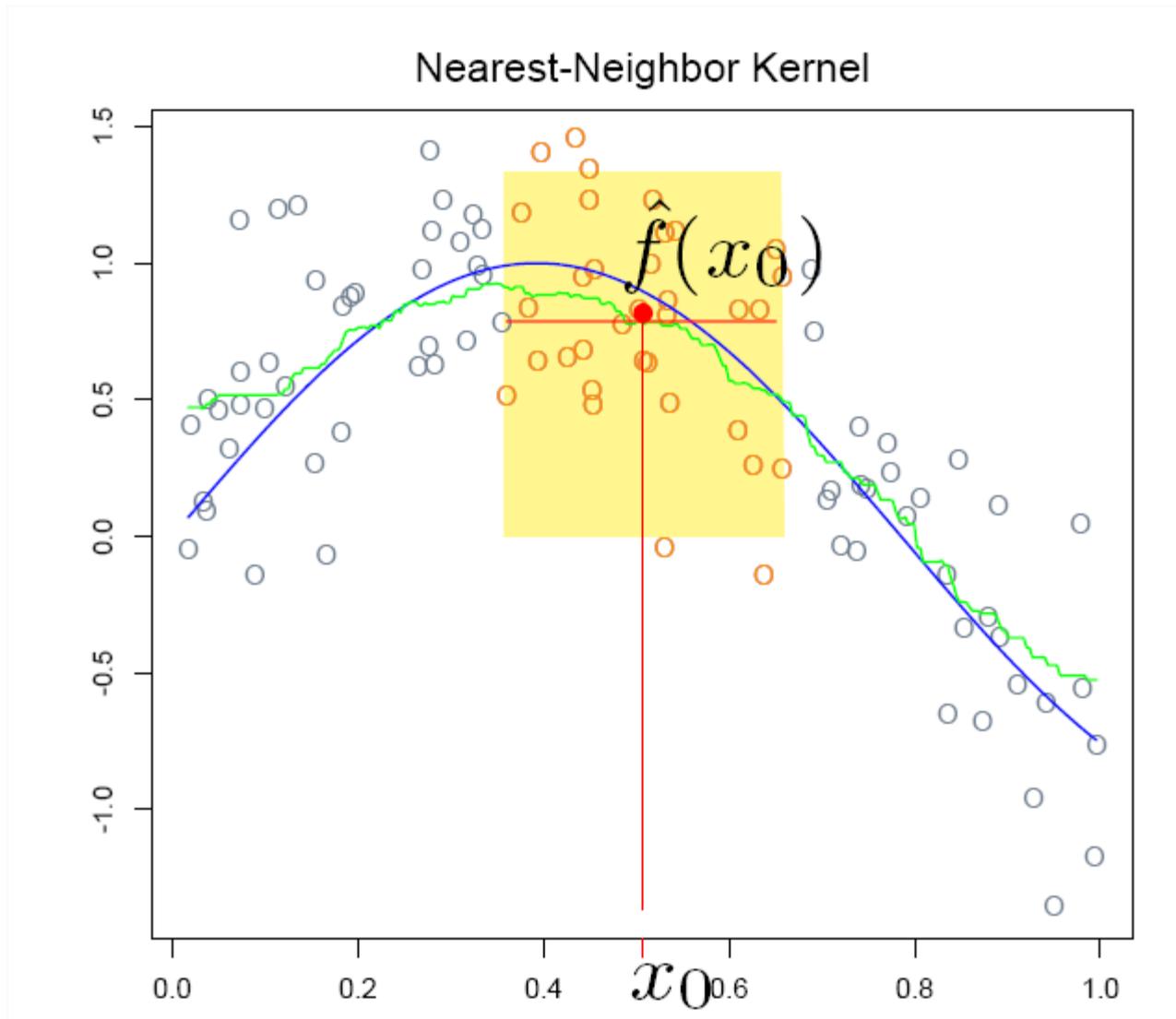
# Introduction

- Forme de la courbe de régression totalement inconnue
- Aucune hypothèse linéarité, loi des erreurs, ...
- Principes anciens, premiers travaux modernes datent des années 50
- Premières applications estimation de fonctions de densité par des méthodes d'opérateurs à noyau Rosenblatt (1956) et de Parzen (1962)
- Ces premiers travaux ont étendus à la notion de régression kernel (lissage par opérateur à noyau en français)

# Introduction

- Résultats: représentation graphique de la relation entre  $X$  et  $Y$
- Pas de forme analytique de la fonction de lien  $f(x)$
- Le problème: estimer  $f(x)$  inconnue et pas seulement ses paramètres
- Deux familles de méthodes
  - Kernel (Nadaraya Watson 1964)
  - Régression locale polynomiale (Cleveland 1979 , Cleveland et Delvin 1988)

- Estimation de l'espérance conditionnelle  $E(Y/x_0) = f(x_0)$  , ou fonction de régression.
  - Approche similaire à l'estimation de densité par la méthode du noyau
  - Premières tentatives inspirées des moyennes mobiles:
    - les k plus proches voisins: moyenne de y pour les k-ppv de  $x_0$
    - moyenne des y sur une fenêtre de largeur fixe centrée sur  $x_0$
- Inconvénient majeur: discontinuité**



©Hastie, Tibshirani & Friedman 2009 Chap 6

# Méthode de la fenêtre mobile

- Moyenne des  $y_i$  dans un voisinage autour de  $x_0$ :  $[x_0-h/2; x_0+h/2]$

$$\hat{f}(x_0) = \frac{\sum_{i=1}^n y_i \mathbf{1}_{[x_0-h/2; x_0+h/2]}(x_i)}{\sum_{i=1}^n \mathbf{1}_{[x_0-h/2; x_0+h/2]}(x_i)}$$

- On construit autour de chaque  $x$  une classe (comme dans un histogramme) de longueur  $h$  centrée sur  $x$ , on compte le nombre de points appartenant à l'intervalle et on calcule la moyenne des  $y$  correspondants

on peut l'écrire

$$\hat{E}(Y / X = x_0) = \frac{\sum_{i=1}^n K\left(\frac{x_0 - x_i}{h}\right) y_i}{\sum_{i=1}^n K\left(\frac{x_0 - x_i}{h}\right)}$$

$K(u) = 1$  si  $|u| < 1/2$  noyau uniforme

En utilisation un noyau continu on obtient l'estimateur de **Nadaraya-Watson** donc une estimation continue de  $f(x)$

Noyaux classiques:

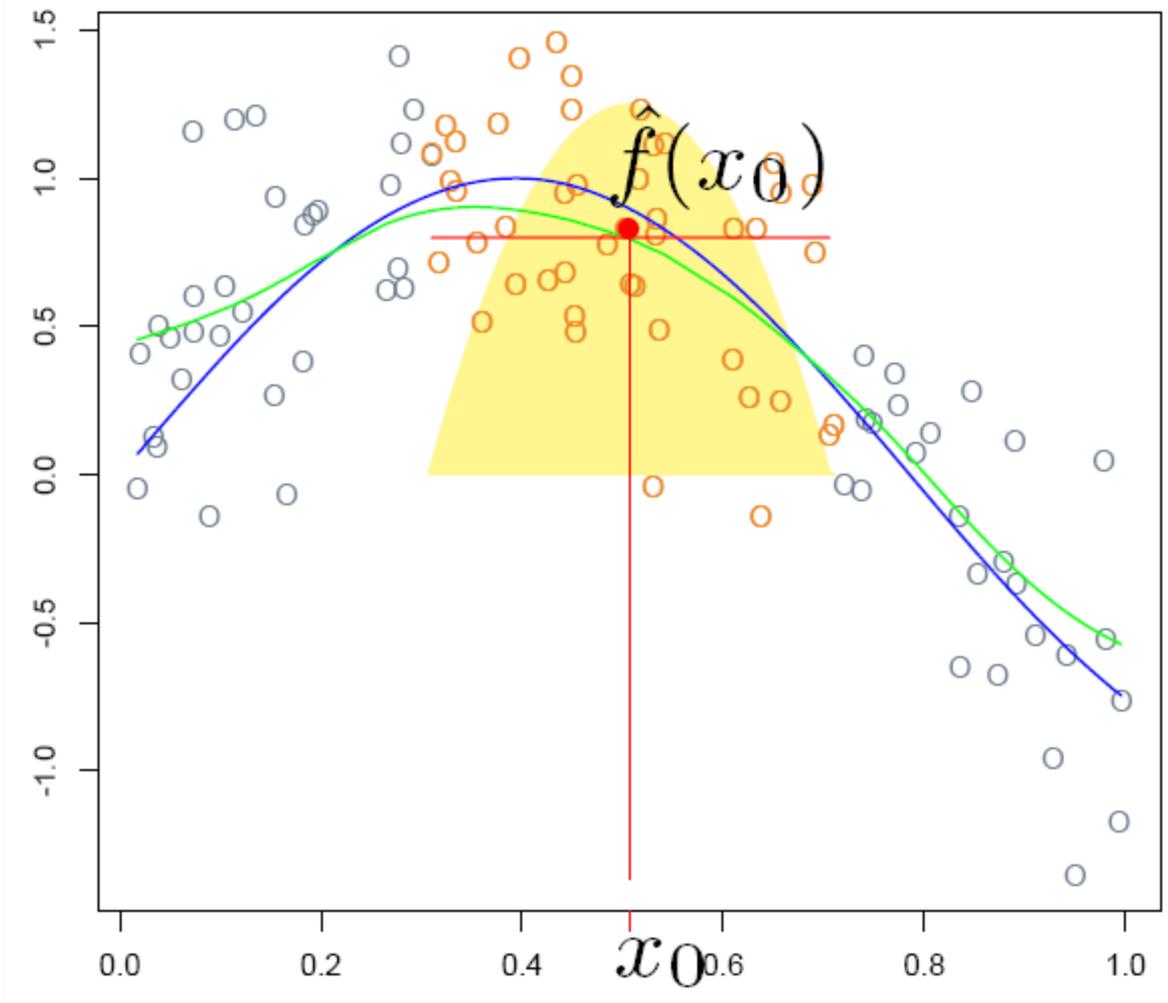
– Epanechnikov

$$K(u) = \frac{3}{4}(1 - u^2) \text{ si } |u| \leq 1, 0 \text{ sinon}$$

– Tricube

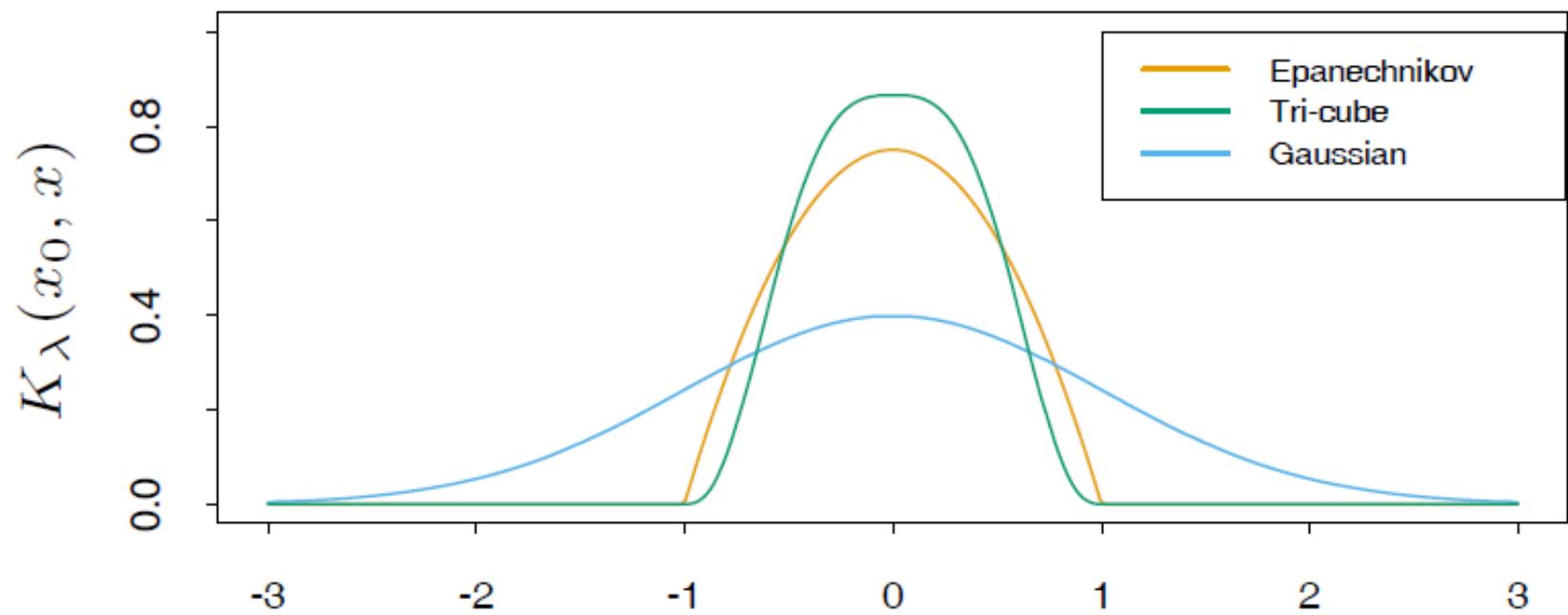
$$K(u) = \left(1 - |u|^3\right)^3 \text{ si } |u| \leq 1, 0 \text{ sinon}$$

# Epanechnikov Kernel



©Hastie, Tibshirani & Friedman 2009 Chap 6

- <http://www.cs.uic.edu/~wilkinson/Applets/smoothers.html>



©Hastie, Tibshirani & Friedman 2009 Chap 6

- Biais et variance pour des  $x_i$  fixés

$$w_i = K\left(\frac{x_0 - x_i}{h}\right)$$

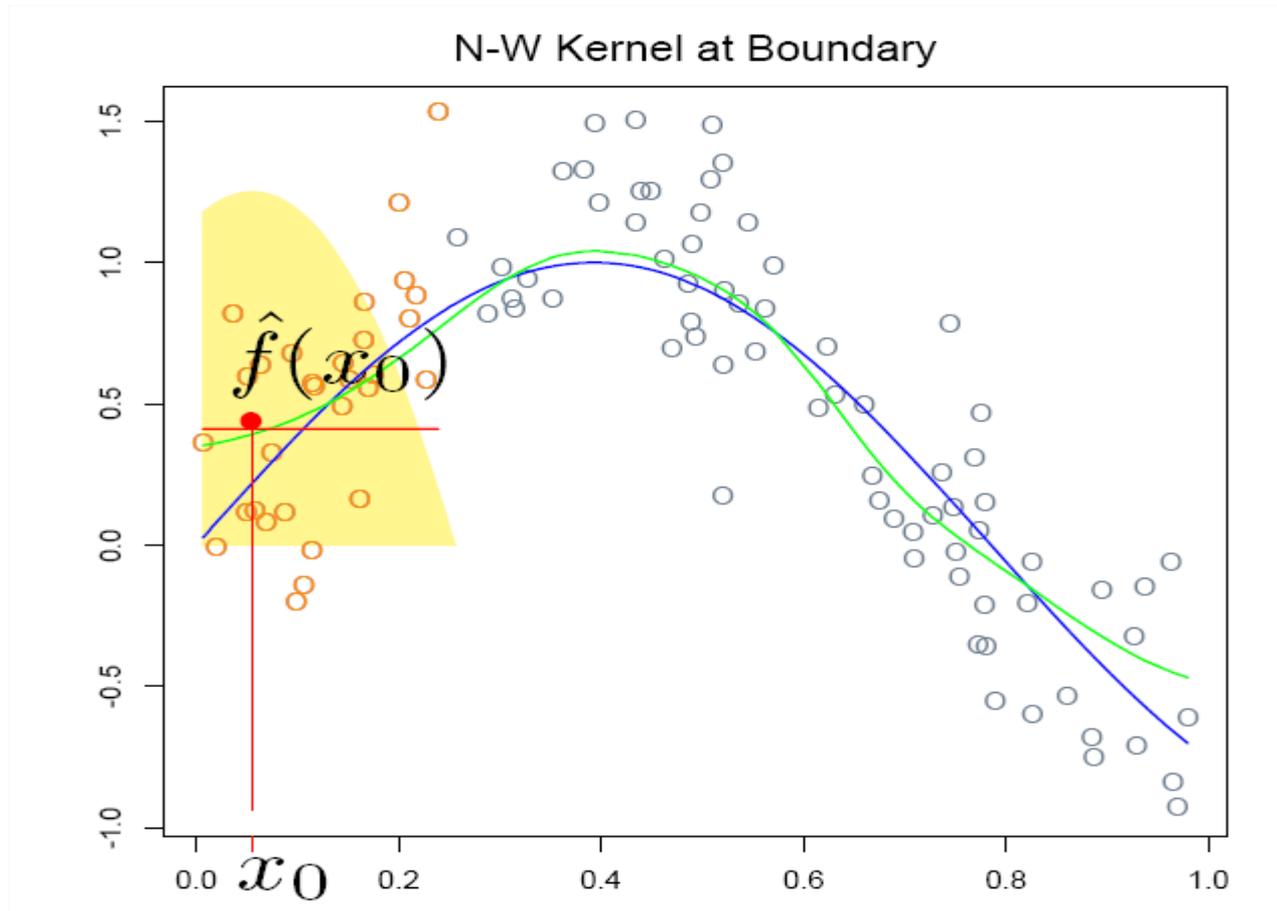
$$E\left(\hat{f}(x_0) - f(x_0)\right) = \frac{\sum_{i=1}^n w_i (f(x_i) - f(x_0))}{\sum_{i=1}^n w_i}$$

$$V\left(\hat{f}(x_0)\right) = \sigma^2 \frac{\sum_{i=1}^n w_i^2}{\left(\sum_{i=1}^n w_i^2\right)^2}$$

# Choix de $h$

- Arbitrage biais variance:
  - Plus  $h$  grand et plus la courbe sera lisse. La variance est limitée mais l'estimateur peut être fortement biaisé
  - Plus  $h$  faible plus la courbe est irrégulière. Les biais d'estimation sont faibles mais la variance est très forte
- Critère MISE et GVC

- Problèmes d'estimation aux bornes valeurs extrêmes de  $x$

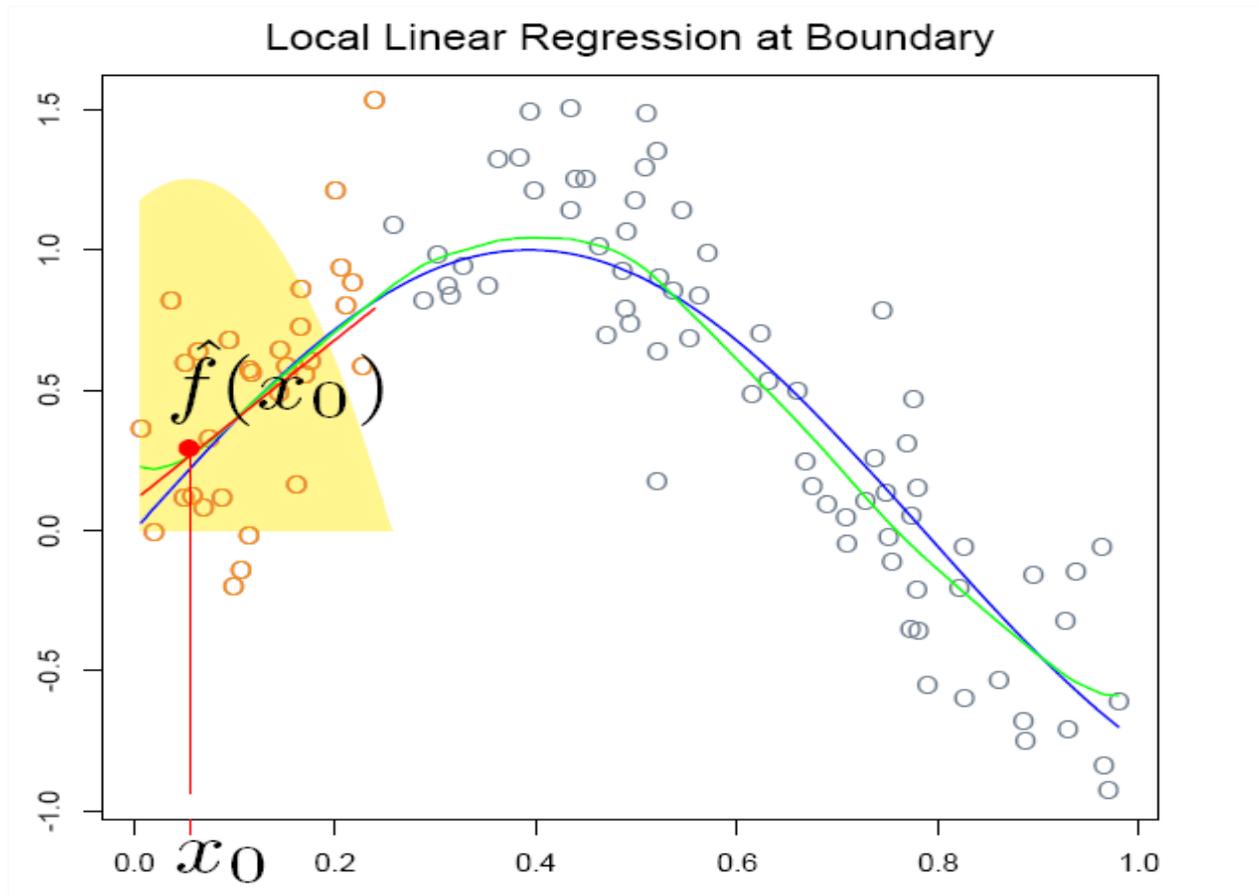


- Alternative: régression linéaire locale
  - Idée générale: utiliser un modèle de régression linéaire défini uniquement dans un voisinage du point d'intérêt  $x_0$
  - On résout en chaque point  $x_0$  le problème de moindres carrés pondérés:

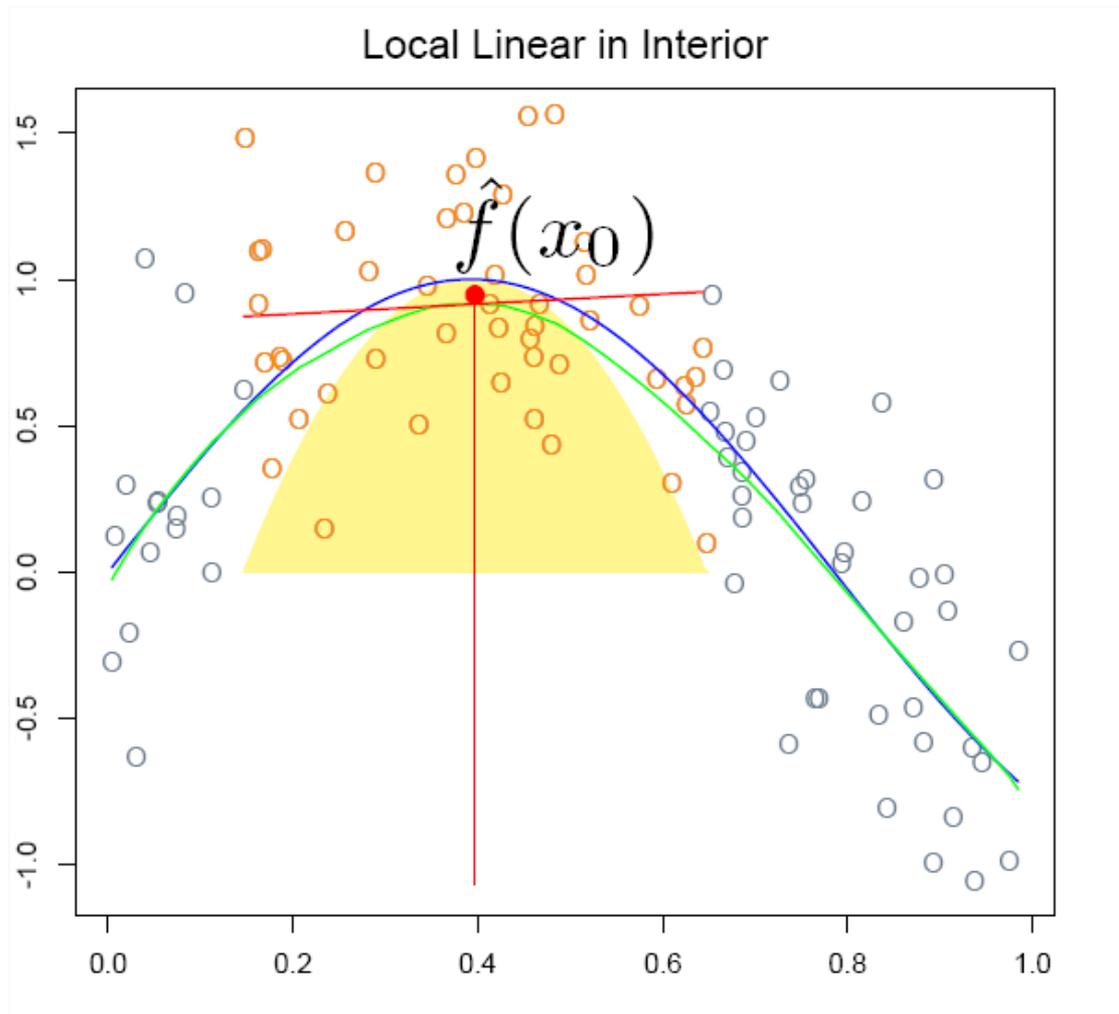
$$\min_{\alpha(x_0), \beta(x_0)} \sum_{i=1}^n K \left( \frac{x_0 - x_i}{h} \right) [y_i - \alpha(x_0) - \beta(x_0)x_i]^2$$

- Nota: formules globales (pour tout  $x$ ), mais chacune utilisée seulement en  $x_0$

- Régression linéaire locale
  - Résout le problème de l'asymétrie du noyau tronqué aux bornes.

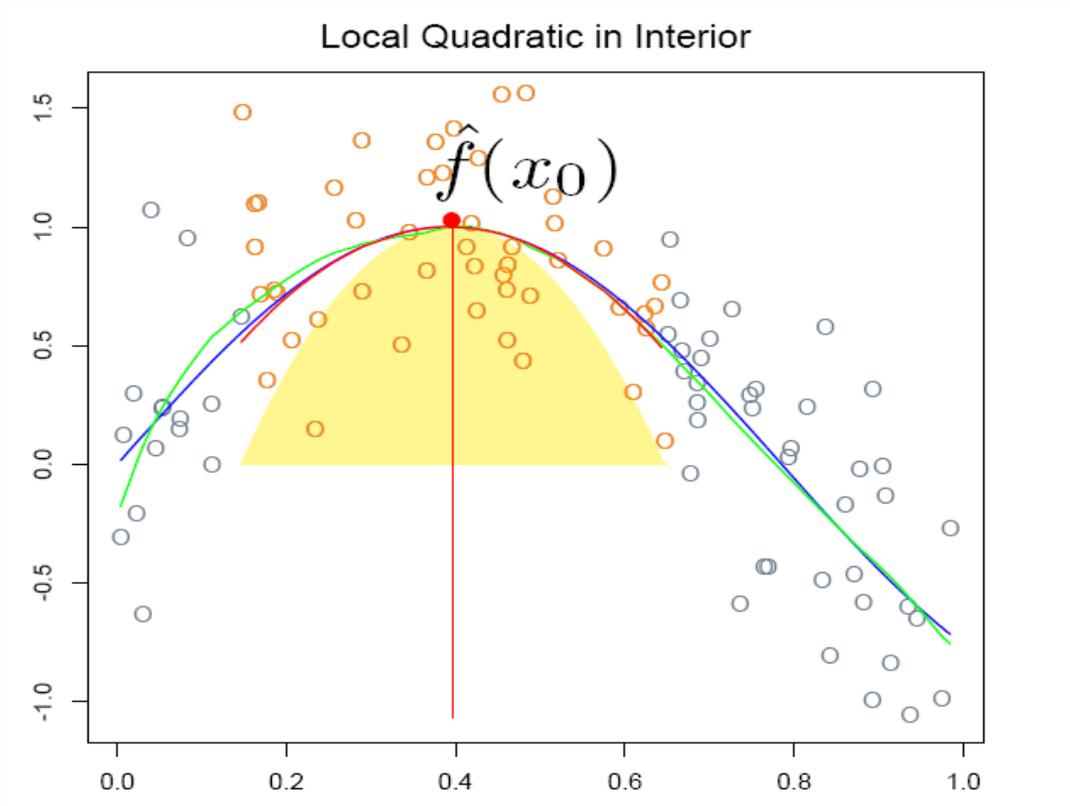


- Les creux et les bosses
  - La régression linéaire locale est biaisée dans les zones de courbure forte



- Régression polynomiale locale

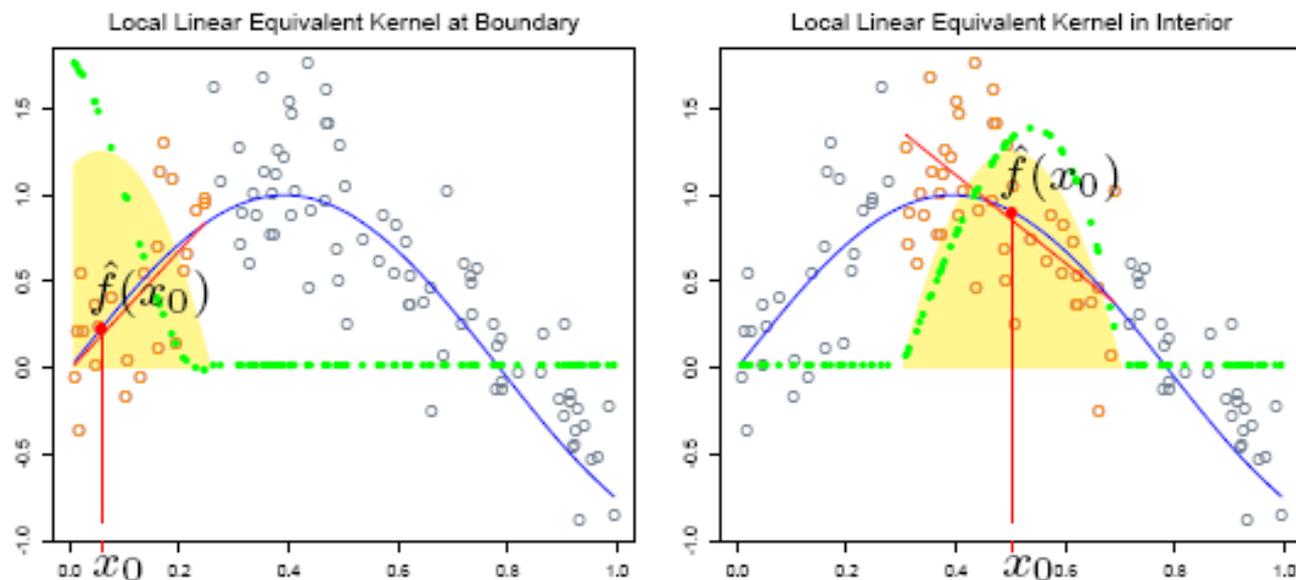
$$\min_{\alpha(x_0), \beta(x_0)} \sum_{i=1}^n K\left(\frac{x_0 - x_i}{h}\right) \left[ y_i - \alpha(x_0) - \sum_{j=1}^d \beta_j(x_0) x_i^j \right]^2$$



- Estimation linéaire et noyau équivalent
  - moindres carrés pondérés
  - $\mathbf{X}$  matrice à  $n$  lignes et  $d+1$  colonnes des  $x^j$

$\mathbf{W}(x_0)$  matrice diagonale  $n, n$  de terme  $K\left(\frac{x_0 - x_i}{h}\right)$

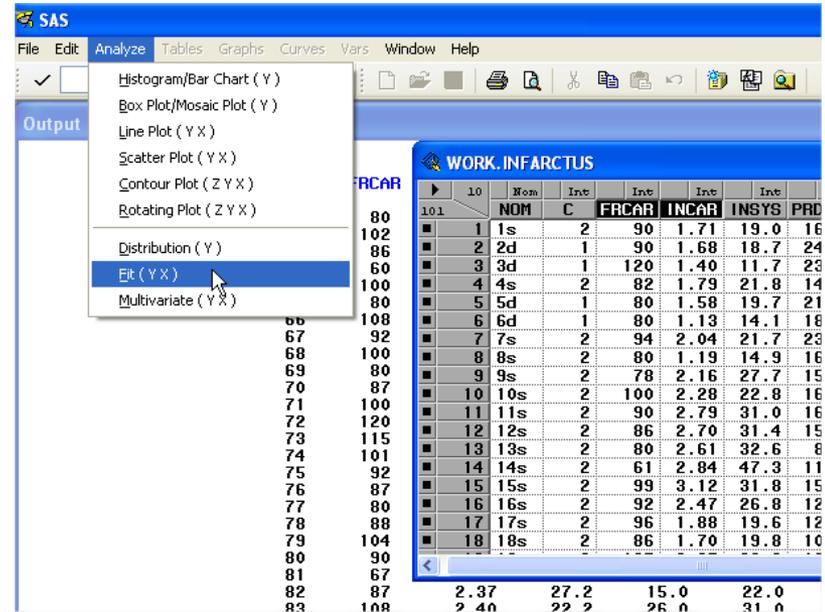
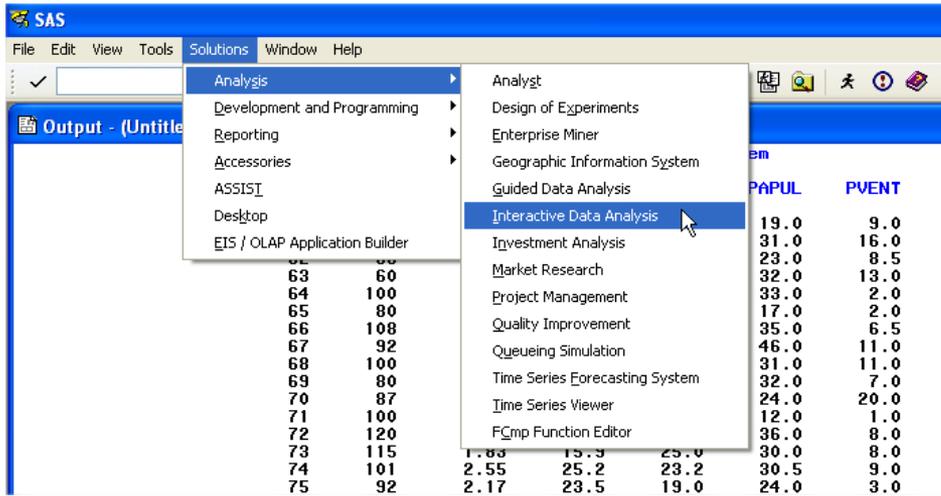
$$\begin{aligned}\hat{f}(x_0) &= \mathbf{x}'_0 (\mathbf{X}'\mathbf{W}(x_0)\mathbf{X})^{-1} \mathbf{X}'\mathbf{W}(x_0)\mathbf{y} \\ &= \sum_{i=1}^n l_i(x_0) y_i\end{aligned}$$



**FIGURE 6.4.** The green points show the equivalent kernel  $l_i(x_0)$  for local regression. These are the weights in  $\hat{f}(x_0) = \sum_{i=1}^N l_i(x_0)y_i$ , plotted against their corresponding  $x_i$ . For display purposes, these have been rescaled, since in fact they sum to 1. Since the yellow shaded region is the (rescaled) equivalent kernel for the Nadaraya–Watson local average, we see how local regression automatically modifies the weighting kernel to correct for biases due to asymmetry in the smoothing window.

- Choix de  $h$ 
  - validation croisée
- Méthode proche: LOESS ou LOWESS
- Extension possible à la régression logistique

# SAS INSIGHT

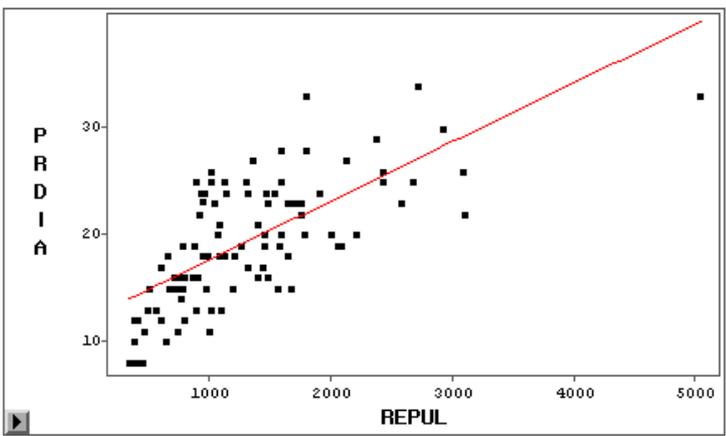


WORK.INFARCTUS

	12	Nom	Int	Int	Int	Int	Int	Int
101	NOM	C	FRCAR	INCAR	INSYS	PRDIA	PAPUL	P
30	1	1s	2	90	1.71	19.0	16.0	19.5
32	2	2d	1	90	1.68	18.7	24.0	31.0
36	3	3d	1	120	1.40	11.7	23.0	29.0
50	4	4s	2	82	1.79	21.8	14.0	17.5
30	5	5d	1	80	1.58	19.7	21.0	28.0
38	6	6d	1	80	1.13	14.1	18.0	23.5
32	7	7s	2	94	2.04	21.7	23.0	27.0
30	8	8s	2	80	1.19	14.9	16.0	21.0
30	9	9s	2	78	2.16	27.7	15.0	20.5
37	10	10s	2	100	2.28	22.8	16.0	23.0
30	11	11s	2	90	2.79	31.0	16.0	25.0
15	12	12s	2	86	2.70	31.4	15.0	23.0
31	13	13s	2	80	2.61	32.6	8.0	15.0
32	14	14s	2	61	2.84	47.3	11.0	17.0
37	15	15s	2	99	3.12	31.8	15.0	20.0
30	16	16s	2	92	2.47	26.8	12.0	19.0
38	17	17s	2	96	1.88	19.6	12.0	19.0
34	18	18s	2	86	1.70	19.8	10.0	14.0
30								
37			2.37	27.2	15.0	22.0	10.0	
38			2.40	22.2	26.0	31.0	4.0	
20			1.91	15.9	18.0	27.0	15.0	
38			1.50	13.9	28.0	43.0	16.0	
36			2.36	27.4	24.0	34.0	8.0	
12			1.56	13.9	24.0	29.0	4.0	
30			1.34	17.0	16.0	25.0	16.0	
35			1.65	17.4	20.0	33.0	7.0	
30			2.04	22.7	28.0	41.0	10.0	
30			3.03	33.6	17.0	23.5	7.0	
34			1.21	12.9	17.0	22.0	3.0	
51			1.34	26.3	11.0	17.0	6.0	

PRDIA = REPUL  
 Response Distribution: Normal  
 Link Function: Identity

Model Equation  
 PRDIA = 11.9816 + 0.0055 REPUL



Parametric Regression Fit				
Curve	Degree(Polynomial)	Model		Error
		DF	Mean Square	DF Mean
—	1	1	1660.4355	99

SAS

File Edit Analyze Tables Graphs **Curves** Vars Window Help

Output - (Untitled)

Confidence Ellipse

Confidence Curves

Polynomial...

Spline...

Kernel...

**Loess...**

Local Polynomial, Fixed Bandwidth...

			Int	
			FRCAR	I
64	100	4	4s	2
65	80	5	5d	1
66	108	6	6d	1
67	92	7	7s	2

Loess Fit

Type:

Mean

Linear

Quadratic

Method:

GCV

DF

Alpha

Weight:

Normal

Triangular

Quadratic

Tri-Cube

Number of Intervals: 128

Convergence Criterion (DF):

Relative Difference

Absolute Difference

DF: 3

Alpha: 0.5

Relative Difference: 0.005

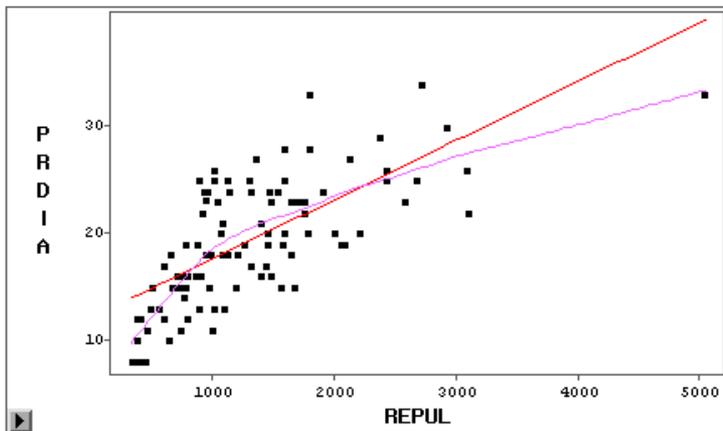
Absolute Difference: 0.05

OK Cancel



► PRDIA = REPUL  
 Response Distribution: Normal  
 Link Function: Identity

► Model Equation  
 PRDIA = 11.9816 + 0.0055 REPUL



► Parametric Regression Fit

Curve	Degree(Polynomial)	Model		Error			Pr > F	
		DF	Mean Square	DF	Mean Square	R-Square		
<span style="color: red;">—</span>	1	1	1660.4355	99	17.3173	0.4920	95.88	<.0001

► Loess Fit

Curve	Type	Weight	N_Intervals	Method	Alpha		K	DF	R-Square	MSE	MSE(GCV)
<span style="color: purple;">—</span>	Linear	Tri-Cube	128	GCV	0.7583		76	4.339	0.5841	14.5196	15.1714

# Avantages et inconvénients

- Utile si la forme de la régression est totalement inconnue
- Méthode adaptative qui s'ajuste automatiquement
- Pas de formule explicite, prévision délicate en dehors du domaine (extrapolation)

# Bibliographie

- Cleveland, W.S.; Devlin, S.J. (1988). Locally-Weighted Regression: An Approach to Regression Analysis by Local Fitting . *Journal of the American Statistical Association* 83 (403): 596–610.
- Dreesbeke, J.J., Saporta G. (éditeurs) (2011) *Approches non paramétriques en régression*, Editions Technip
- Hastie, T., Tibshirani, R., Friedman, J. (2009): *The Elements of Statistical Learning* , 2nd edition, chapitre 6, Springer, <http://www-stat.stanford.edu/~hastie/Papers/ESLII.pdf>
- Lejeune, M. (1985), Estimation non paramétrique par noyaux : régression polynomiale mobile, *Revue de Statistique Appliquée*, 33, 43-68.