



Technique de prédiction : La régression linéaire

Cadre du modèle linéaire

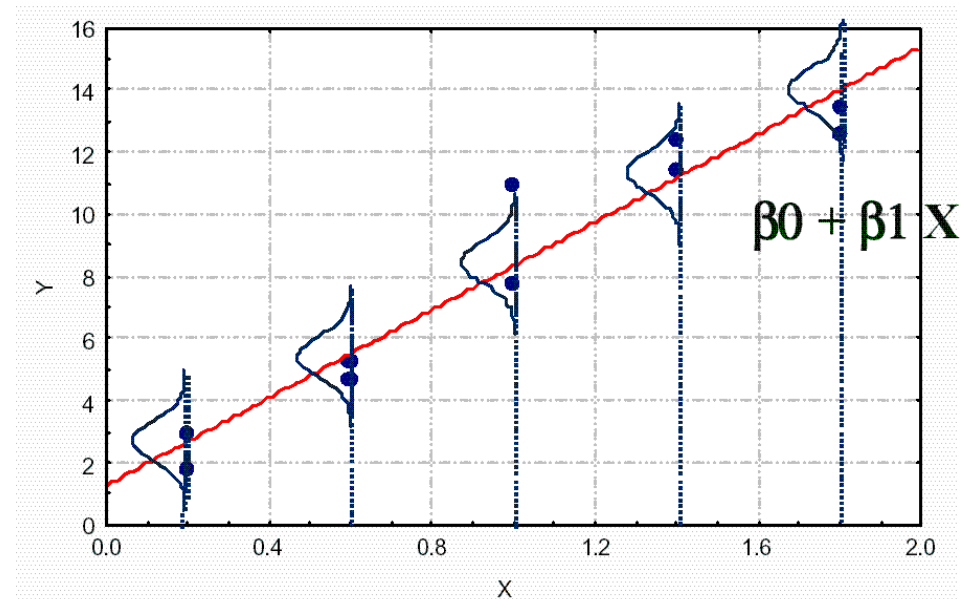
- Dans le modèle simple : X et Y deux variables continues
- Les valeurs x_i de X sont contrôlées et sans erreur de mesure
- On observe les valeurs correspondantes y_1, \dots, y_n de Y
- Exemples :
 - X peut être le temps et Y une grandeur mesurée à différentes dates
 - Y peut être la différence de potentiel mesurée aux bornes d'une résistance pour différentes valeurs de l'intensité X du courant

Hypothèse fondamentale du modèle linéaire

- X et Y ne sont pas indépendantes et la connaissance de X permet d'améliorer la connaissance de Y
- Savoir que $X = x$ permet rarement de connaître exactement la valeur de Y, mais on suppose que cela de connaître la valeur moyenne $E(Y|X=x)$, l'espérance conditionnelle de Y sachant que $X = x$
- On suppose plus précisément que $E(Y|X=x)$ est une fonction linéaire de x, ce qui permet d'écrire
 - $E(y_i) = \alpha + \beta x_i$ pour tout $i = 1, \dots, n$
 - $\Leftrightarrow y_i = \alpha + \beta x_i + \varepsilon_i$, avec $E(\varepsilon_i) = 0$ pour tout $i = 1, \dots, n$
 - $n = \text{nb d'observations}$ et $\varepsilon_i = \ll \text{résidu} \gg$ de l'observation i
- Régression linéaire multiple :
 - $Y = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k + \varepsilon$
 - important : on suppose l'indépendance linéaire des X_i

Autres hypothèses du modèle linéaire

- La variance des résidus est la même pour toutes les valeurs de X (homoscédasticité)
 - $V(\varepsilon_i) = s^2$
- Les résidus sont linéairement indépendants
 - $\text{cov}(\varepsilon_i, \varepsilon_j) = 0 \quad \forall i \neq j$
- Les résidus sont normalement distribués
 - $\varepsilon_i \sim N(0, s^2)$



La composante stochastique

- L'existence de la composante stochastique (ε_i) correspond au fait que :
 - des individus avec même valeur x_i peuvent avoir des réponses Y différentes (variation synchronique)
 - OU un même individu mesuré à plusieurs reprises avec la même valeur x_i peut avoir des réponses Y différentes (variation diachronique)
- On a équivalence de $\varepsilon_i \sim N(0, s^2)$ et $Y/X=x_i \sim N(\alpha + \beta x_i, s^2)$
- Cette hypothèse de normalité classe la régression linéaire dans la famille des modèles linéaires généraux (GLM)
- Dans les modèles linéaires généralisés, la loi de $Y/X=x_i$ n'est plus nécessairement normale

Que signifie la variance des estimateurs ?

- Après avoir postulé l'existence d'une relation $E(Y) = \alpha + \beta X$, on recherche des estimateurs a et b de α et β
- On n'atteint jamais les véritables coefficients α et β car :
 - le modèle linéaire n'est le plus souvent qu'une approximation de la réalité
 - on ne travaille que sur des échantillons et non la population entière
 - on commet des erreurs de mesure
- Des modèles sur des échantillons différents donneront des estimateurs a' et b' différents
- D'où une variance des estimateurs a et b

Méthode des moindres carrés ordinaires (MCO)

- On recherche des estimateurs a et b de α et β qui minimisent les résidus $\varepsilon_i^2 = (Y_i - \hat{Y}_i)^2$, où \hat{Y}_i est prédit par la droite $\hat{Y} = a + bX$

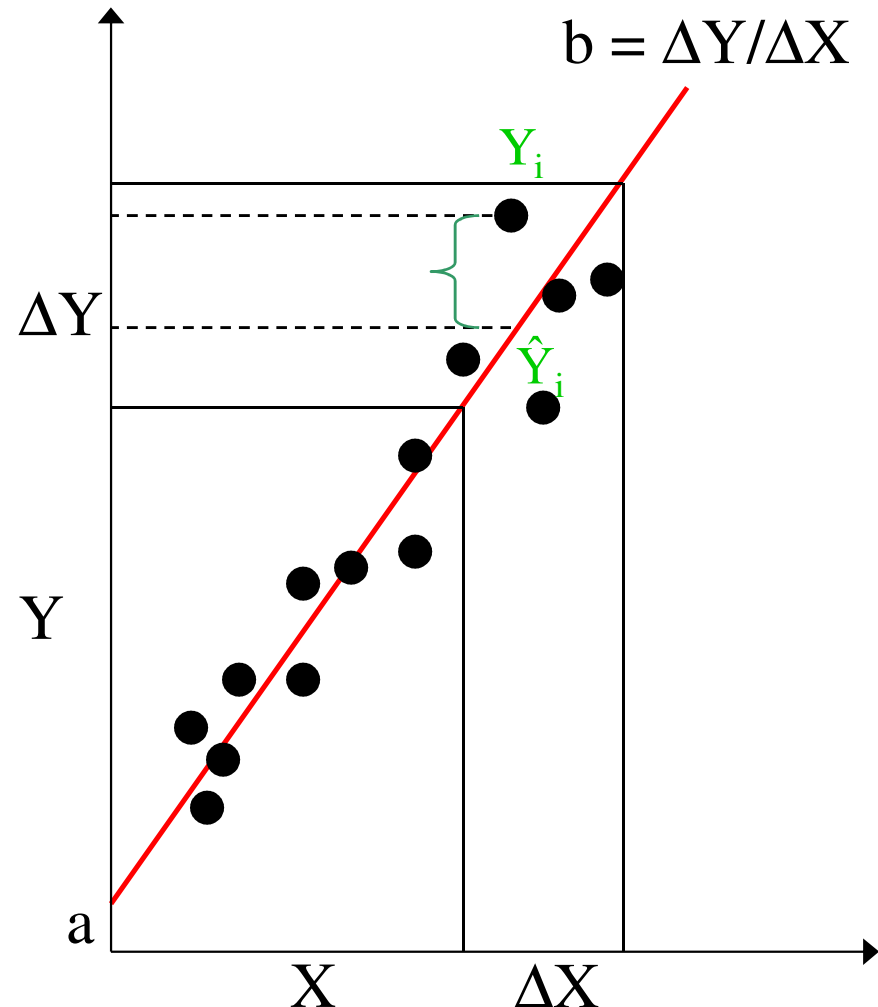
- L'estimateur b de la pente est :

$$b = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_i (x_i - \bar{x})^2} = \frac{\text{cov}(X, Y)}{\sigma^2_X}$$

- L'estimateur a de la constante vaut :

$$a = \bar{y} - b \cdot \bar{x}$$

- La droite $\hat{Y} = a + b.X$ ajuste le nuage de points



Propriétés des estimateurs MCO

- Les estimateurs MCO des coefficients ont :
 - une moyenne : $E(a)$ et $E(b)$
 - une variance :
 - constante : $\sigma_a^2 = s^2 [1/n + \bar{X}^2 / \sum (x_i - \bar{X})^2]$
 - avec : $s^2 =$ variance des résidus
 - > IC au niveau $100(1-\alpha)\% = a \pm t_{\alpha/2, n-p-1} \cdot \sigma_a$
 - pente : $\sigma_b^2 = s^2 [1 / \sum (x_i - \bar{X})^2]$
 - > IC au niveau $100(1-\alpha)\% = b \pm t_{\alpha/2, n-p-1} \cdot \sigma_b$
- La méthode MCO est optimale car :
 - les estimateurs sont sans biais : $E(a) = \alpha$ et $E(b) = \beta$
 - de variance minimale parmi tous les estimateurs linéaires
 - on dit qu'ils sont « BLUE » : best linear unbiased estimators
- Hypothèse de normalité $\varepsilon_i \sim N(0, s^2) \Rightarrow$ les estimateurs sont de variance minimale parmi tous les estimateurs

Conséquence des formules de variance

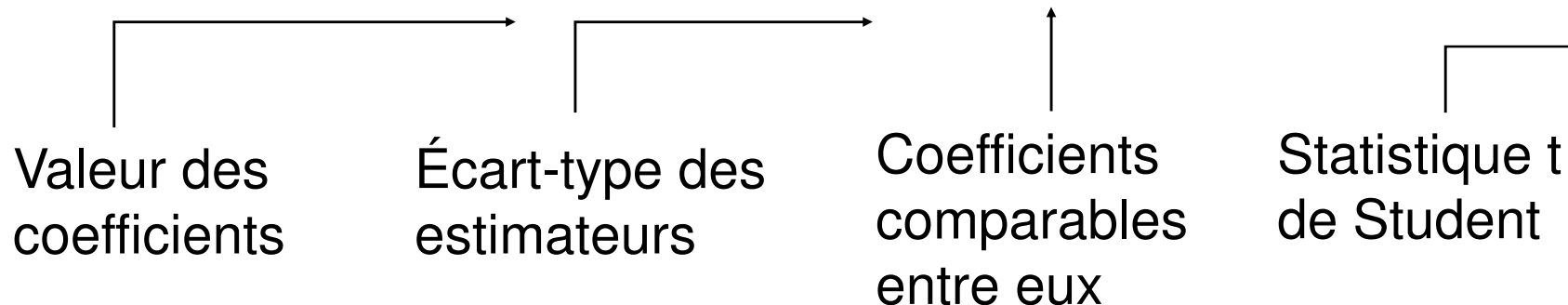
- Pour diminuer les variances :
 - diminuer la variance résiduelle s^2 de l'échantillon
 - augmenter la taille n de l'échantillon
 - augmenter l'étendue des valeurs observées de X
- Mais : on accepte parfois (régression ridge) des estimateurs légèrement biaisés pour diminuer leur variance

Coefficients de régression et tests

Coefficients^a

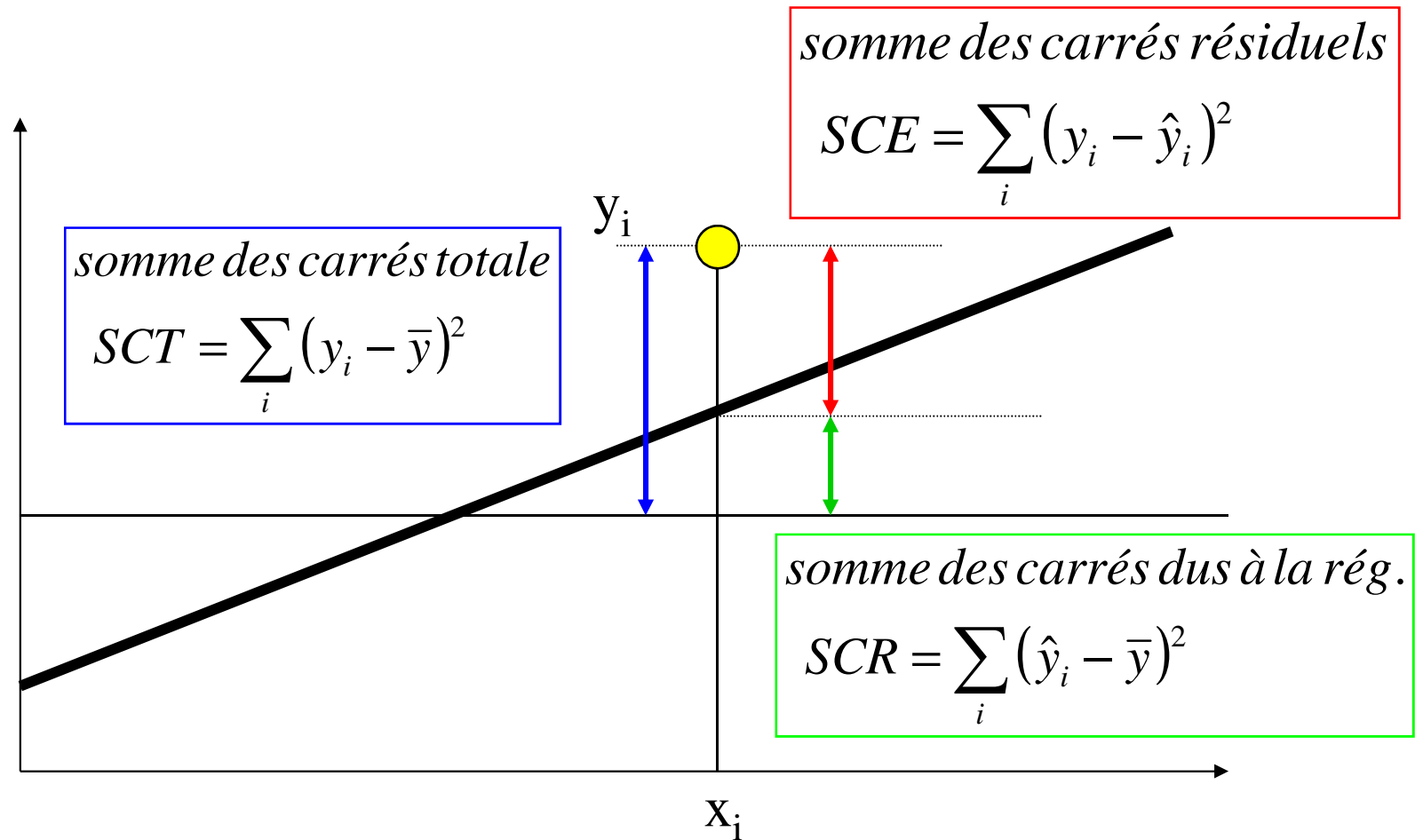
Modèle	Coefficients non standardisés		Coefficients standardisés	t	Signification
	B	Erreur standard	Bêta		
1 (constante)	1467,643	62,422		23,512	,000
TEMPERAT	-37,060	2,295	-,866	-16,147	,000
ISOLATIO	-29,774	3,492	-,457	-8,526	,000

a. Variable dépendante : CONSOMMA



Une valeur $t > 2$ ou $t < -2$ est significative à 95 % d'un coeff $\neq 0$

Sommes des carrés



Test global du modèle

ANOVA^b

Modèle		Somme des carrés	ddl	Carré moyen	F	Signification
1	Régression	3267046,7	2	1633523,333	167,933	,000 ^a
	Résidu	116727,068	12	9727,256		
	Total	3383773,7	14			

a. Valeurs prédites : (constantes), ISOLATIO, TEMPERAT

b. Variable dépendante : CONSOMMA

2 prédicteurs \Rightarrow régression linéaire multiple

$$F = \frac{\frac{SCR}{p}}{\frac{SCE}{n-p-1}}$$

SCR = somme des carrés « Régression »

SCE = somme des carrés « Erreurs »

p = nombre de variables

n = nombre d'observations

suit une loi F de ddl ($p, n-p-1$)
sous l'hypothèse nulle (H_0) :
($b_1 = b_2 = 0$)

$$R^2 = SCR / SCT = 1 - (SCE / SCT)$$

variance s^2 du terme d'erreur = 98,627²

Coefficient de détermination

- $R^2 = SCR / SCT$
- R^2 = proportion de variation de la variable cible expliquée par tous les prédicteurs (syn : régresseurs)
- Bon ajustement si R^2 proche de 1
- R^2 est biaisé (optimiste car croissant avec le nb de variables) et on lui substitue le R^2 ajusté :

$$R^2_{ajusté} = 1 - \frac{(1 - R^2)(n - 1)}{n - p - 1}$$

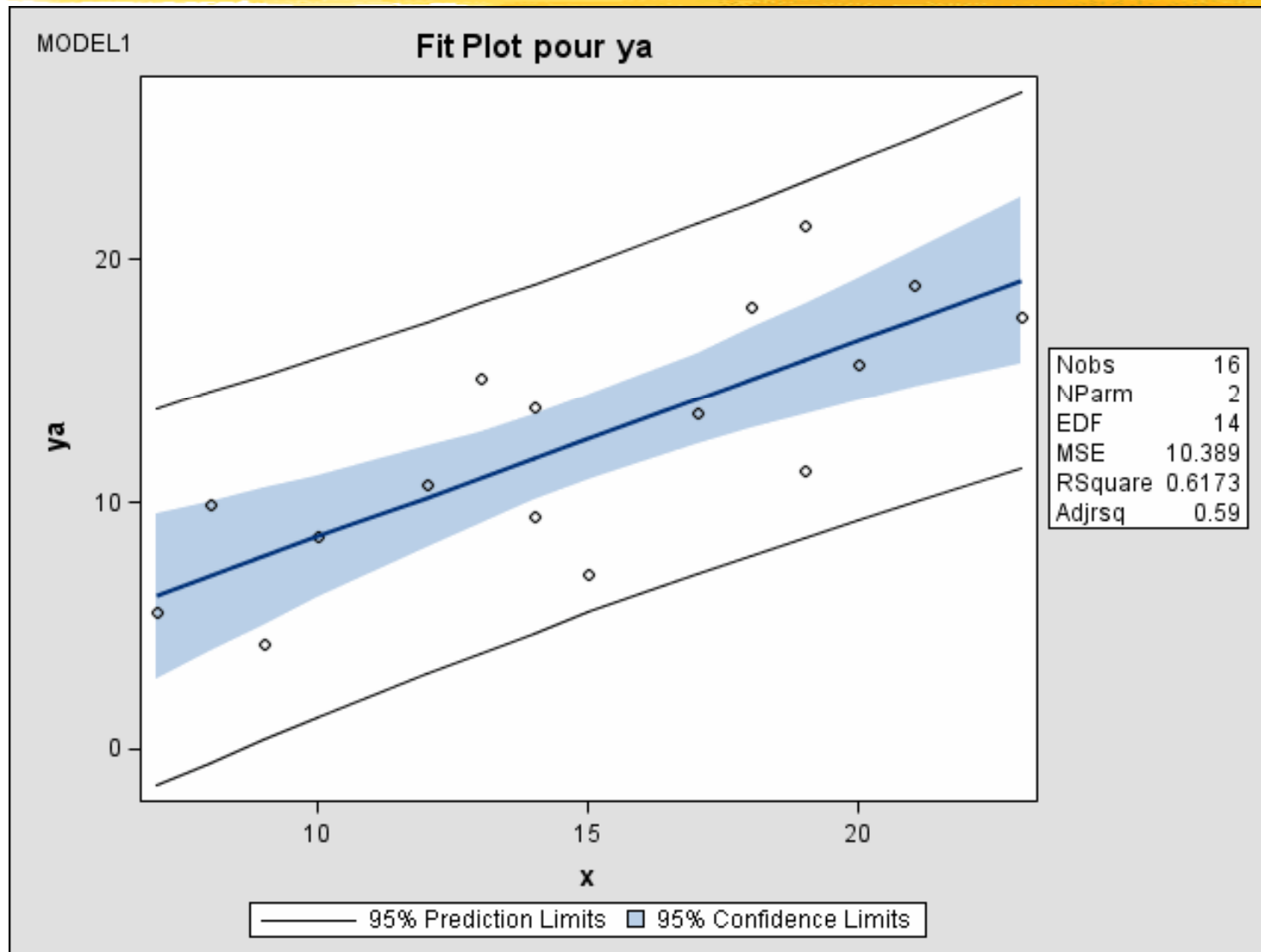
- R^2 ajusté est toujours $< R^2$ et peut être < 0

Modèle	R	R-deux	R-deux ajusté	Erreur standard de l'estimation
1	,983 ^a	,966	,960	98,627

Intervalles de confiance

- $\hat{y}_0 = a + bx_0$ est une prévision de Y et de la moyenne $E(Y)$ en tout point x_0 de l'intervalle de mesure (car $E(\varepsilon_i) = 0$)
- D'après les formules sur les variances des estimateurs, les IC à $(100-\alpha) \%$ de $E(Y)$ et Y au point X_0 sont :
 - $\hat{y}_0 \pm t_{\alpha/2, n-p-1} \cdot s [1/n + (x_0 - \bar{X})^2 / \sum (x_i - \bar{X})^2]^{1/2}$ pour $E(Y)$
 - $\hat{y}_0 \pm t_{\alpha/2, n-p-1} \cdot s [1 + 1/n + (x_0 - \bar{X})^2 / \sum (x_i - \bar{X})^2]^{1/2}$ pour Y (on a ajouté la variance du terme d'erreur)
- Autrement dit, la variance de la valeur prédite pour une observation est :
 - $s^2 [1 + 1/n + (x_0 - \bar{X})^2 / \sum (x_i - \bar{X})^2]$
- > Plus difficile d'estimer une valeur possible de Y sachant $X=x_0$ que la moyenne des valeurs possibles sachant $X=x_0$
- > L'IC augmente quand x_0 s'éloigne de \bar{X}

IC de la moyenne et des observations



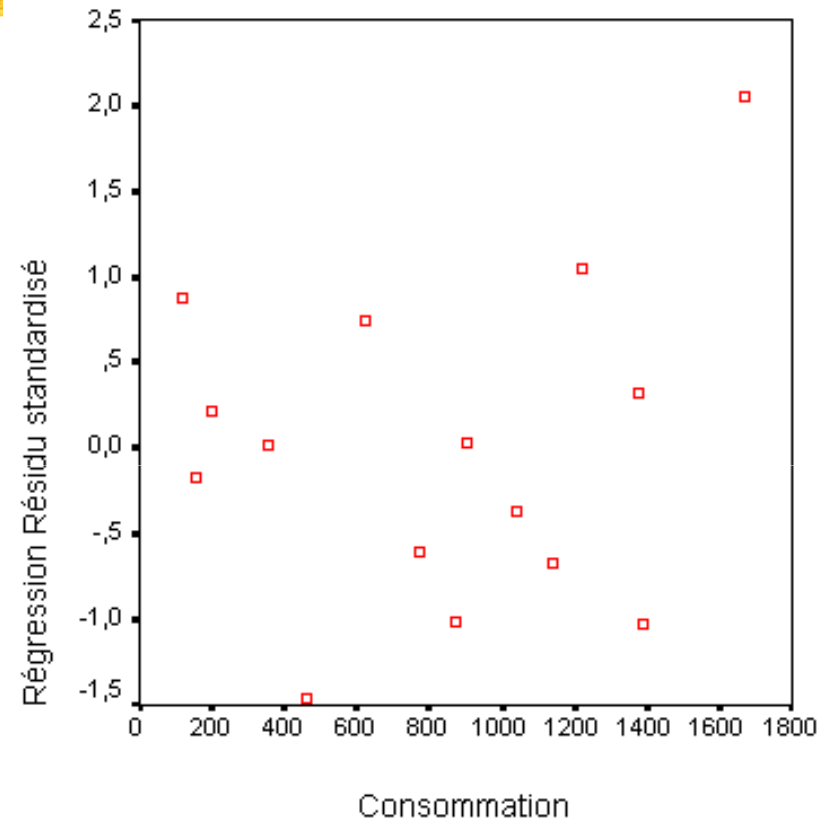
Précautions d'utilisation

- Le modèle n'est valide que sur l'étendue des observations, et surtout près de la moyenne de X
- Un petit échantillon (< 20) ne détecte que les relations fortes ; un grand détecte toutes les relations même faibles (rejet de H_0 malgré petit R^2)
- Minimum de 5 observations (mieux vaut en avoir > 15)
- Attention aux résidus standardisés (résidu / s) > 3
- Pour savoir si les extrêmes ont une influence : les enlever et voir les coeff. restent dans les IC des coeff. initiaux
- Attention aux distances de Cook > 1
 - la distance de Cook d'une observation i mesure l'écart des coefficients avec et sans cette observation
- Régression multiple : vérifier l'absence de multicolinéarité

Analyse des résidus

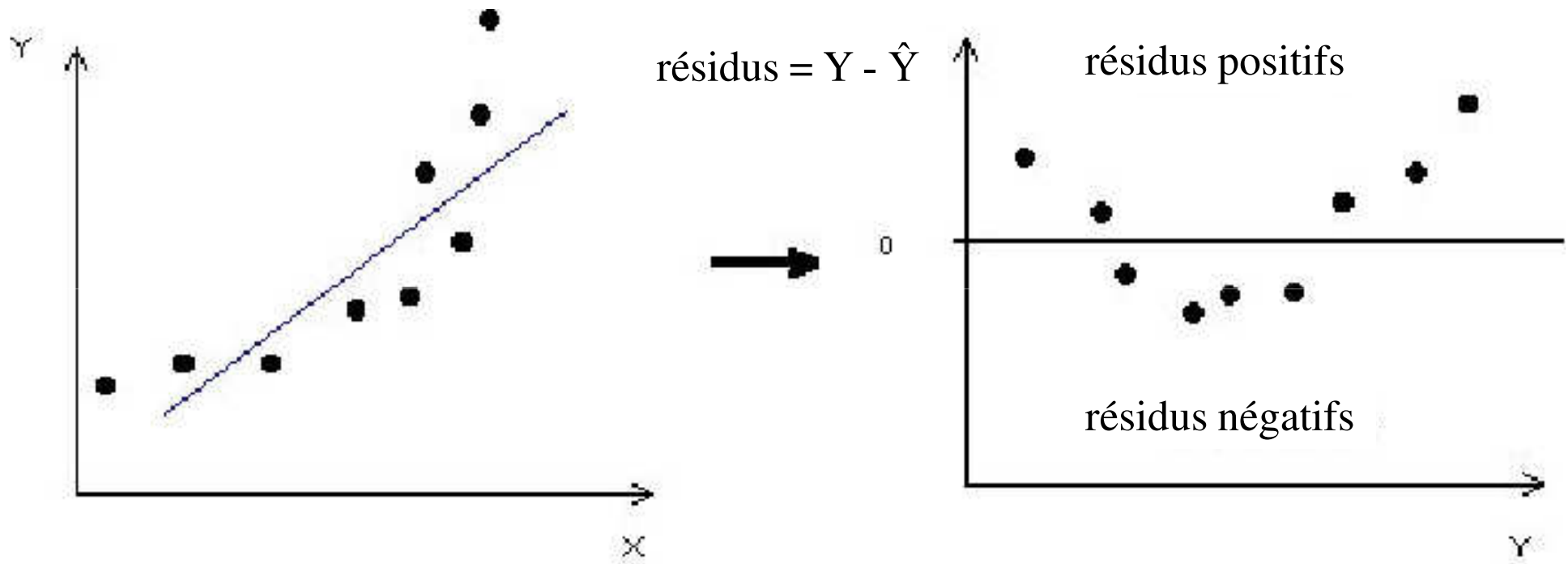
Vérification du respect des hypothèses de base

- Test d'**autocorrélation**
(statistique de Durbin-Watson comprise entre 1,5 et 2,5)
- Test d'**homoscédasticité**
(égalité de la variance en fonction de y)
- Test de **normalité** (test de Kolmogorov)
- Vérification d'absence de points extrêmes
- Un diagramme des résidus est souvent très parlant



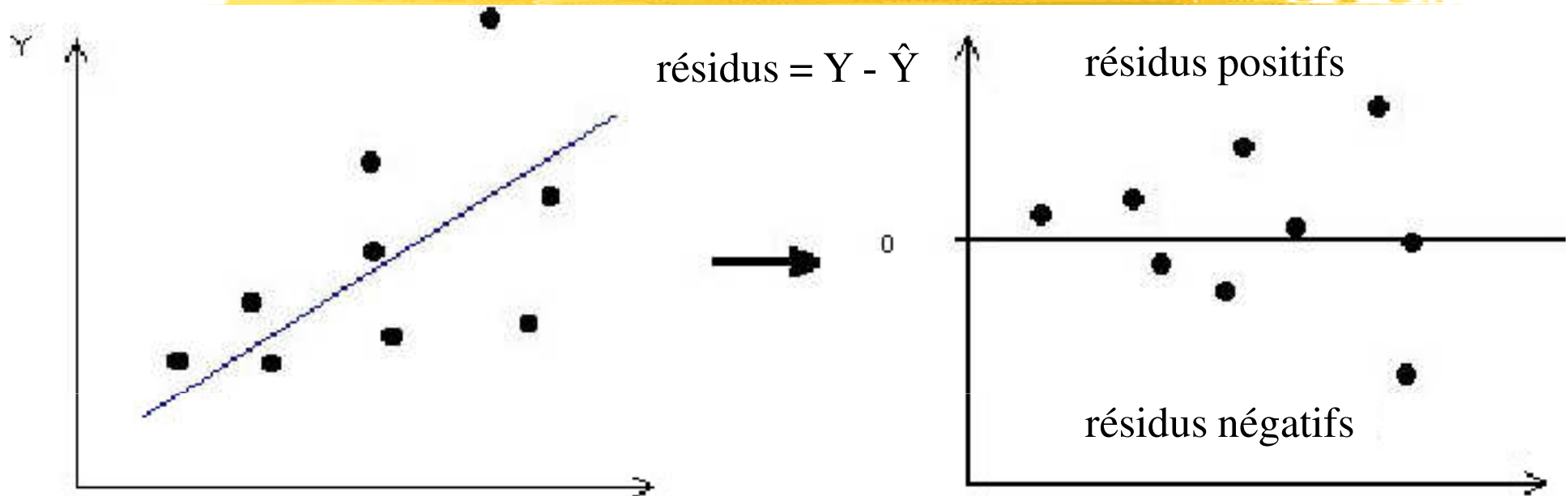
Les résidus standardisés doivent être répartis aléatoirement autour de 0 et rester dans les bornes $[-3 ; +3]$

Problème 1 : Autocorrélation des résidus



Corrélation entre ε_i et $\varepsilon_{i+1} \Rightarrow$ les valeurs moyennes de Y sont sur-estimées ; les autres sont sous-estimées

Problème 2 : Hétéroscédasticité des résidus



Appliquer le test de Levene en regroupant en classes les valeurs de Y

Estimation précise de Y en fonction de X lorsque Y est petit ; grande incertitude quand Y est grand

⇒ remplacer Y par son log, son inverse ou sa racine carrée

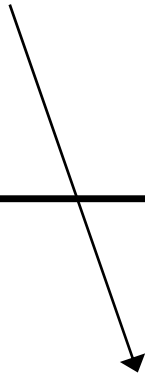
(ou par le carré ou l'exponentielle quand la variance diminue)

⇒ ou utiliser la méthode des moindres carrés pondérés

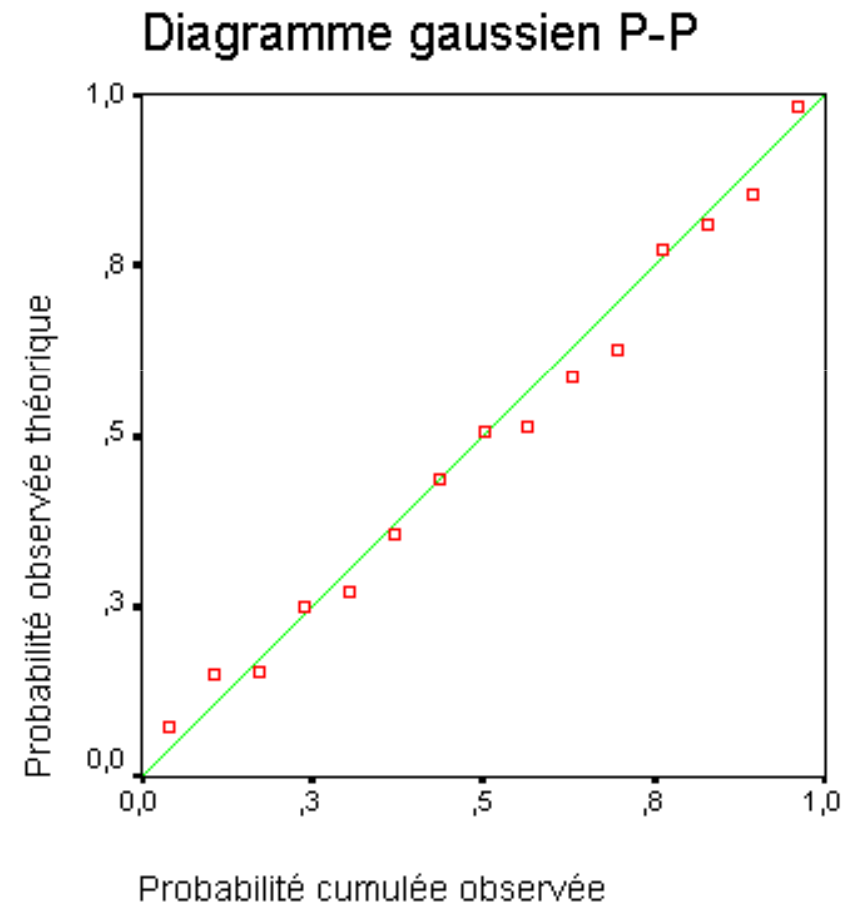
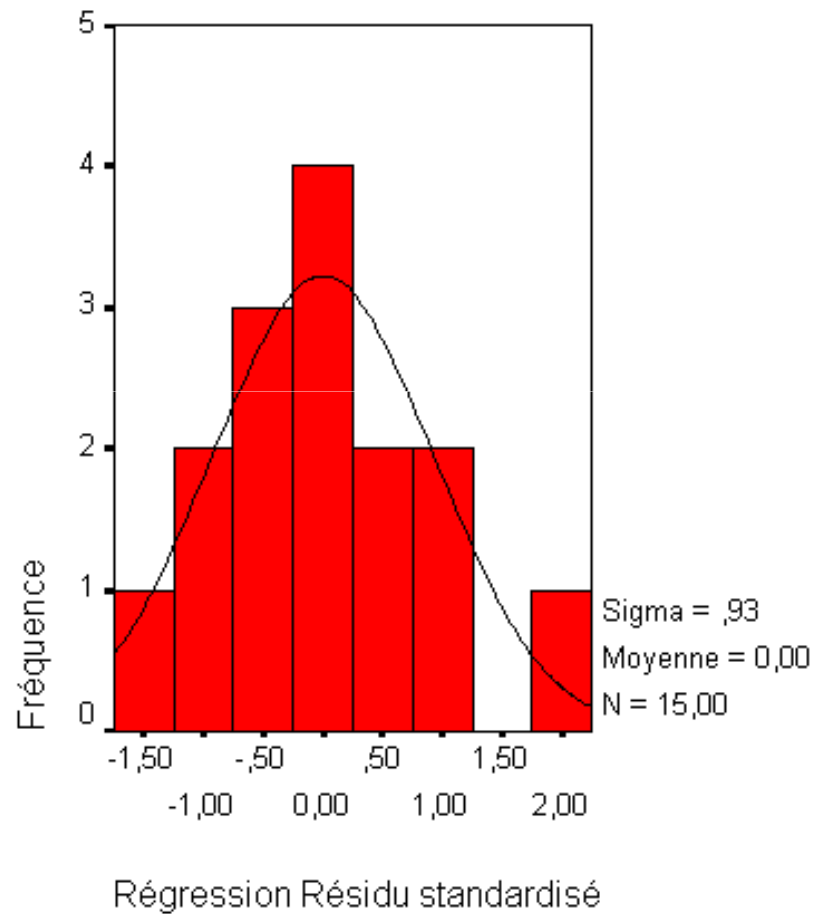
Homoscédasticité et autocorrélation des résidus

- Utiliser un diagramme des résidus pour vérifier l'homoscédasticité et l'absence d'autocorrélation
- Statistique de Durbin-Watson pour l'autocorrélation :
 - $= \sum (\varepsilon_i - \varepsilon_{i-1})^2 / \sum \varepsilon_i^2$
 - vaut entre 0 et 4
 - proche de 2 si pas d'autocorrélation (OK entre 1,5 et 2,5)
 - < 2 pour des corrélations positives
 - > 2 pour des corrélations négatives

	R	R-deux	R-deux ajusté	Durbin-Watson
	,983 ^a	,966	,960	1,819



Normalité des résidus



Utilité des tests sur les résidus 1/3

- Exemple tiré de :
 - Tomassone, Lesquoy, Millier : *La Régression nouveaux regards sur une ancienne méthode statistique*, 1986
 - Anscombe F.J. : *Graphs in Statistical Analysis*, 1973

	x	ya	yb	yc	yd	xe	ye
1	7	5.535	0.113	7.399	3.864	13.715	5.654
2	8	9.942	3.77	8.546	4.942	13.715	7.072
3	9	4.249	7.426	8.468	7.504	13.715	8.491
4	10	8.656	8.792	9.616	8.581	13.715	9.909
5	12	10.737	12.688	10.685	12.221	13.715	9.909
6	13	15.144	12.889	10.607	8.842	13.715	9.909
7	14	13.939	14.253	10.529	9.919	13.715	11.327
8	14	9.45	16.545	11.754	15.86	13.715	11.327
9	15	7.124	15.62	11.676	13.967	13.715	12.746
10	17	13.693	17.206	12.745	19.092	13.715	12.746
11	18	18.1	16.281	13.893	17.198	13.715	12.746
12	19	11.285	17.647	12.59	12.334	13.715	14.164
13	19	21.365	14.211	15.04	19.761	13.715	15.582
14	20	15.692	15.577	13.737	16.382	13.715	15.582
15	21	18.977	14.652	14.884	18.945	13.715	17.001
16	23	17.69	13.947	29.431	12.187	33.281	27.435

Utilité des tests sur les résidus 2/3

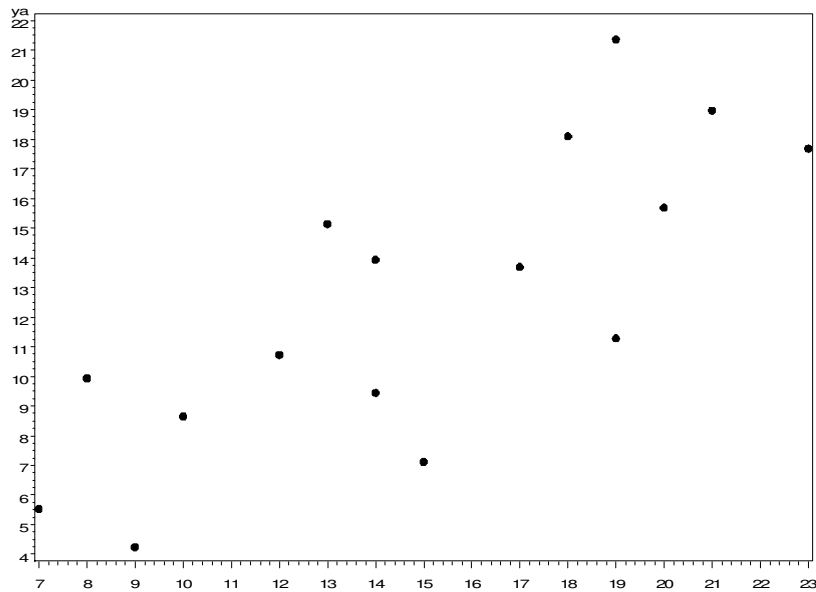
- Dans les 5 régressions : mêmes sommes de carrés, même variance résiduelle, même F-ratio, mêmes R^2 , même droite de régression, mêmes écarts-types des coefficients...

Analyse de variance					
Source	DF	Somme des carrés	Carré moyen	Valeur F	Pr > F
Model	1	234.6	234.6	22.6	0.0003
Error	14	145.4	10.4		
Corrected Total	15	380.1			

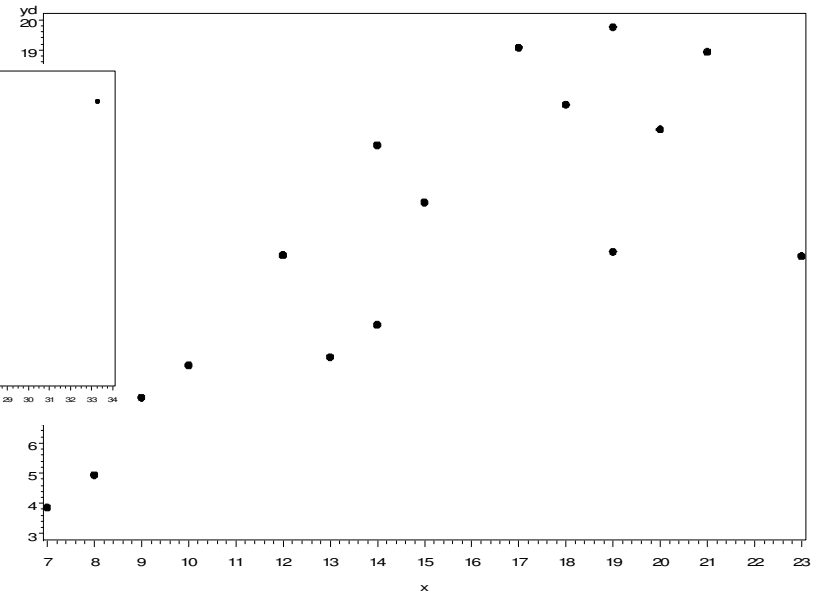
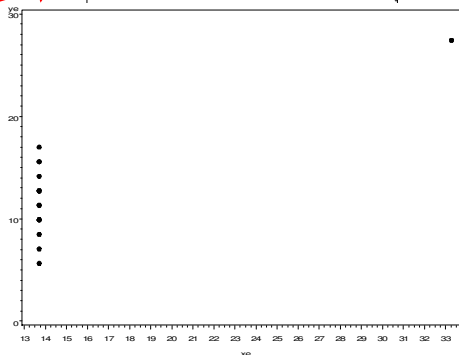
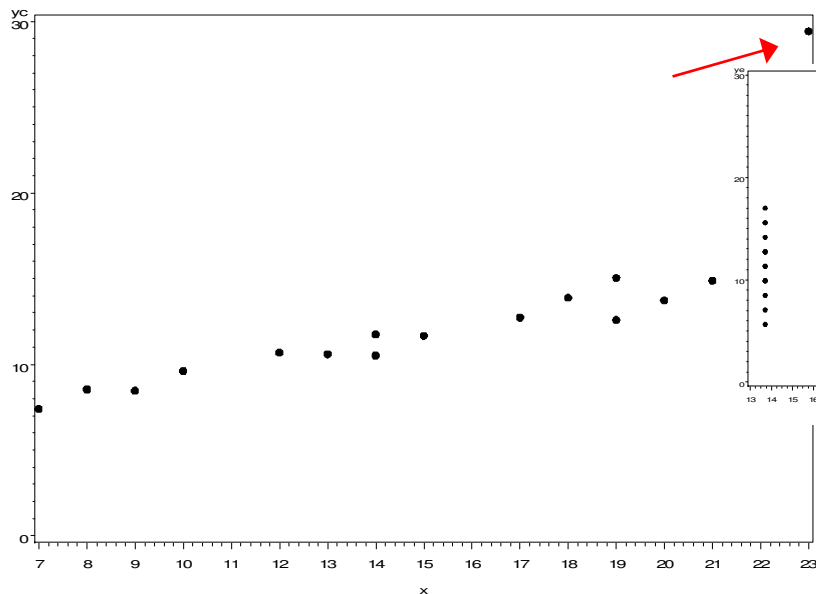
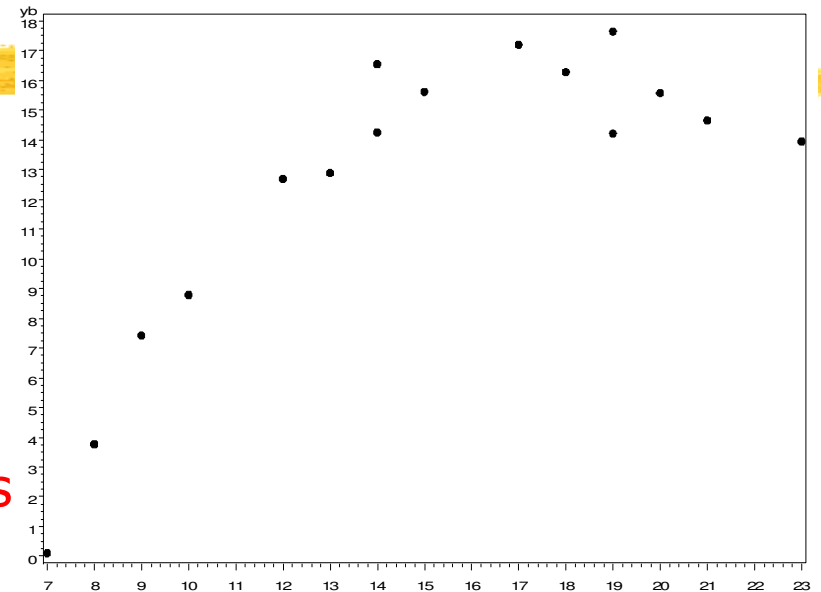
Root MSE	3.22	R-Square	0.62
Dependent Mean	12.60	Adj R-Sq	0.59
Coeff Var	25.60		

Résultats estimés des paramètres							
Variable	D F	Résultat estimé des paramètres	Erreur std	Valeur du test t	Pr > t	Tolérance	Inflation de variance
Intercept	1	0.52	2.67	0.20	0.8476	.	0
x	1	0.81	0.17	4.75	0.0003	1.00	1.00

Utilité des tests sur les résidus 3/3

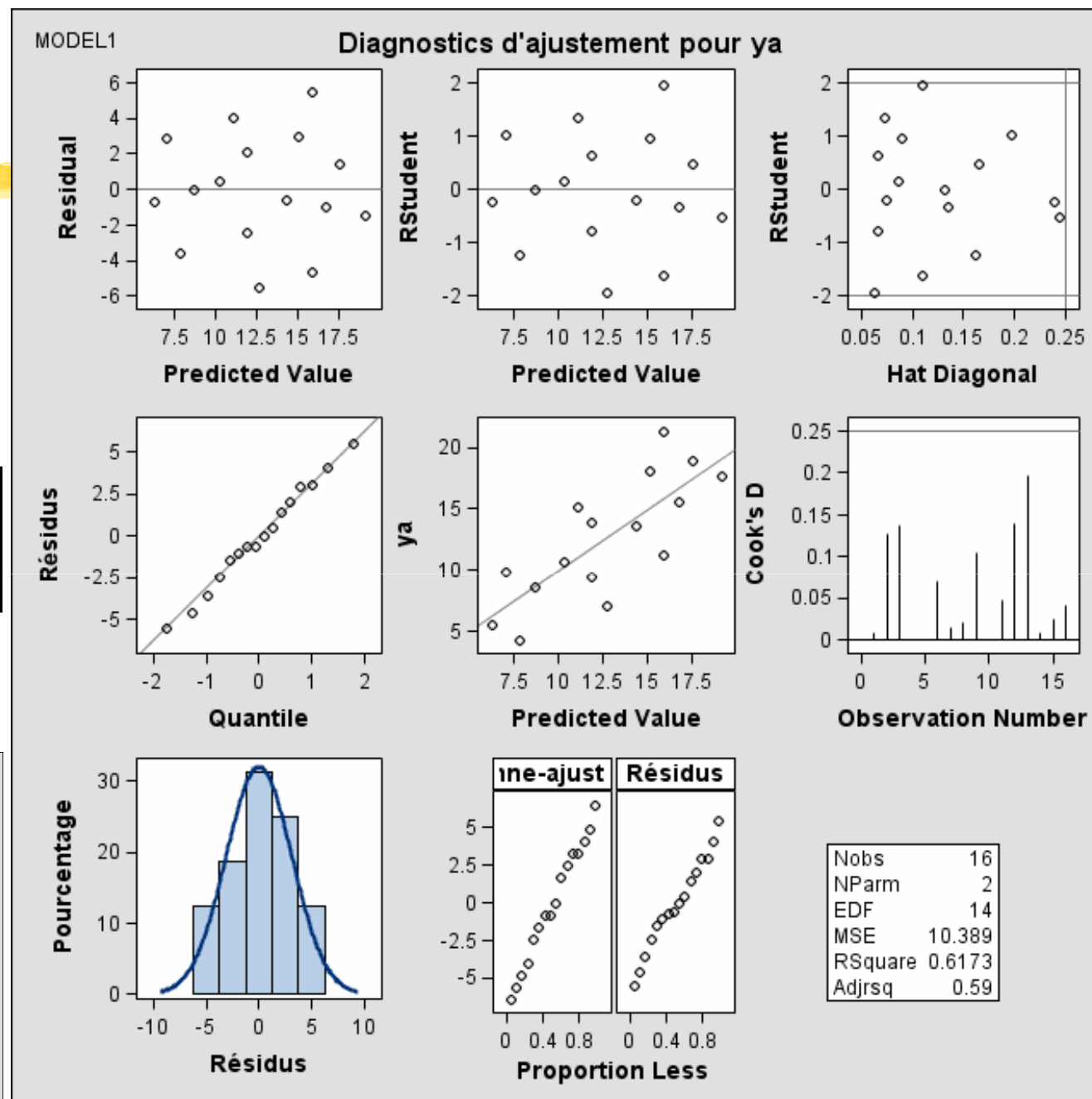
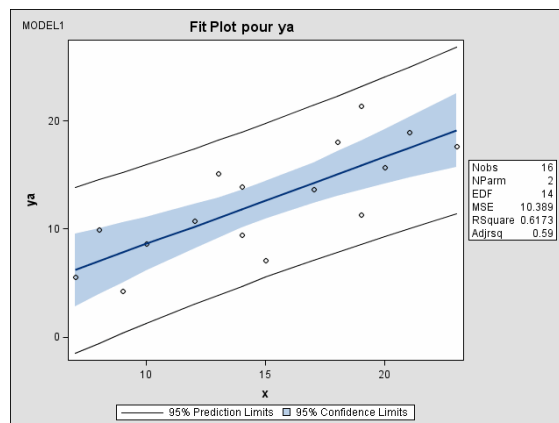


Et
pourtant !
Les
situations
sont bien
différentes



Régression 1 :

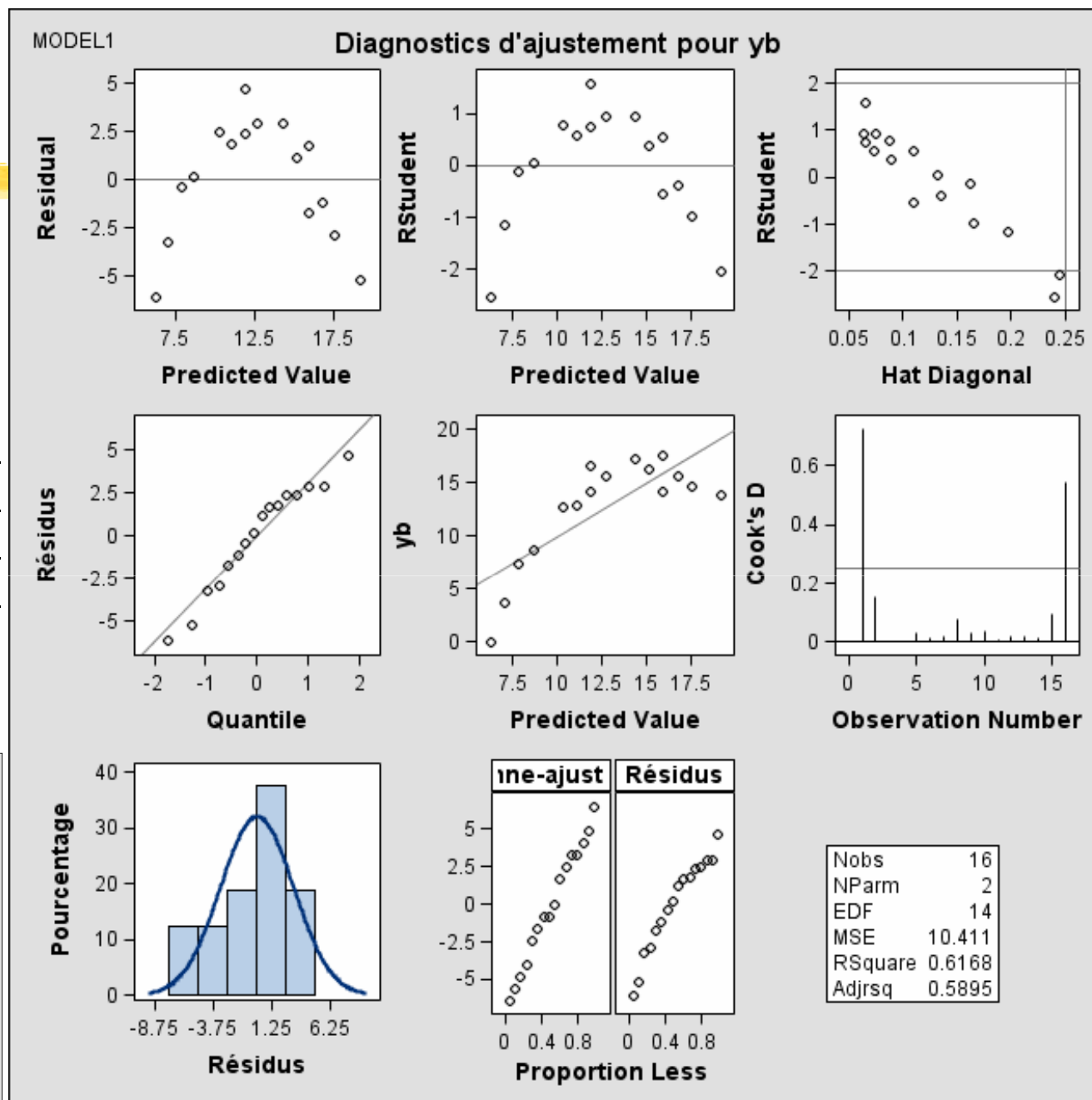
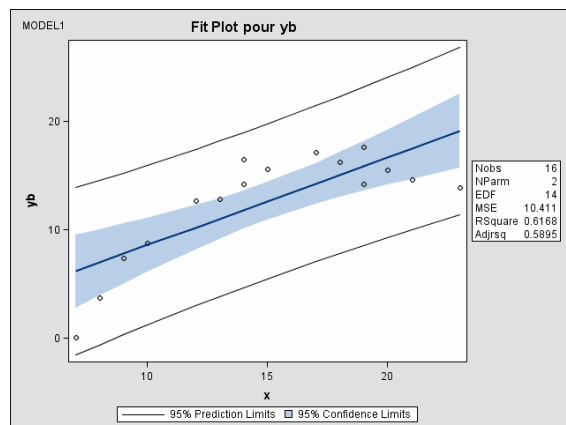
Durbin-Watson D	2.538
Number of Observations	16
1st Order Autocorrelation	-0.277



Régression 2 :

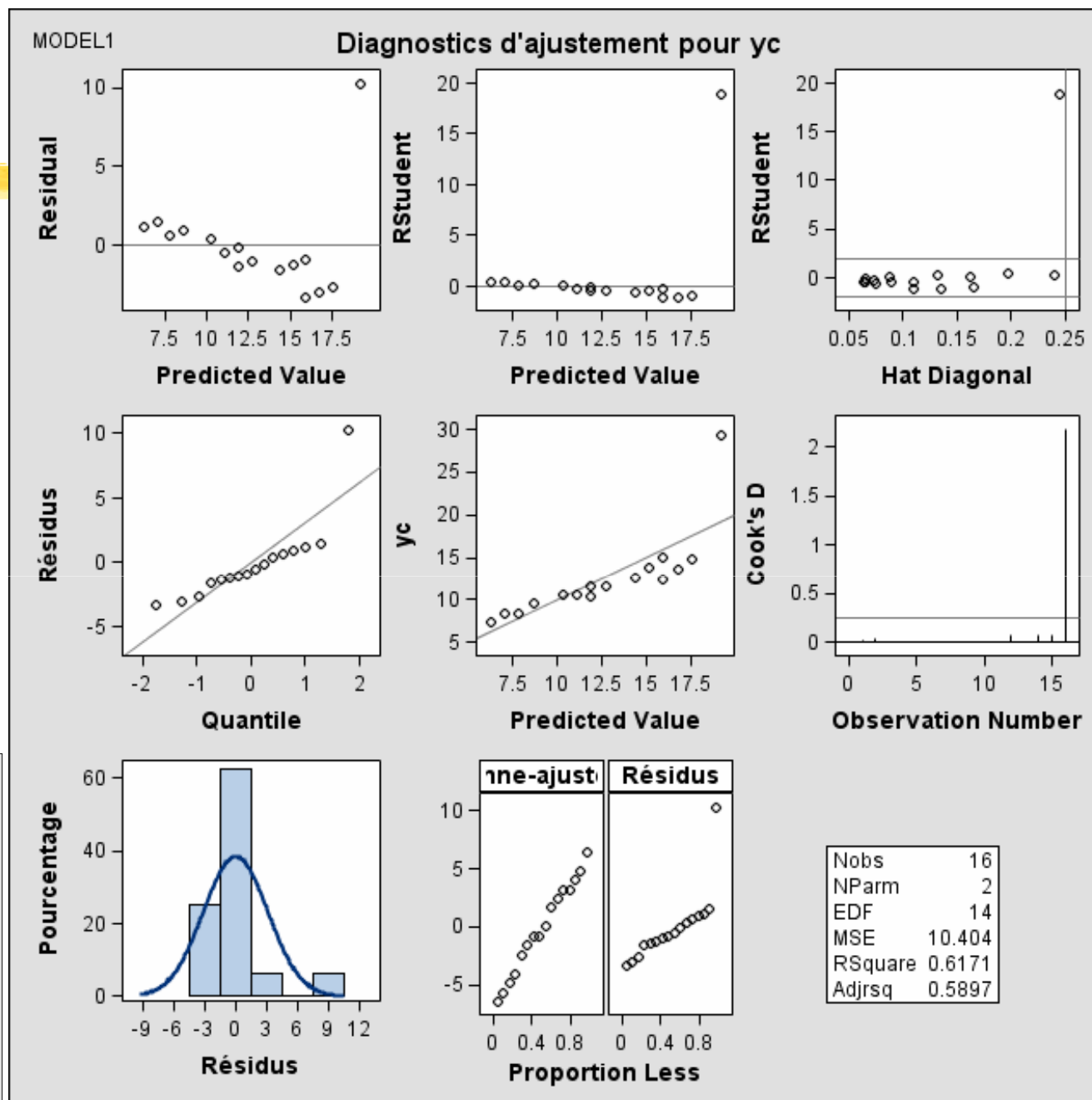
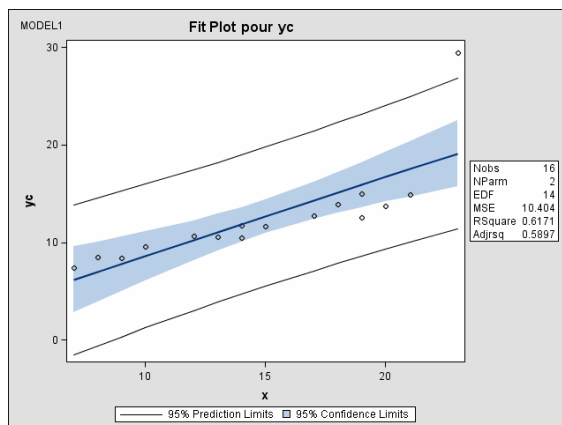
Forte auto-corrélation positive !

Durbin-Watson D	0.374
Number of Observations	16
1st Order Autocorrelation	0.595



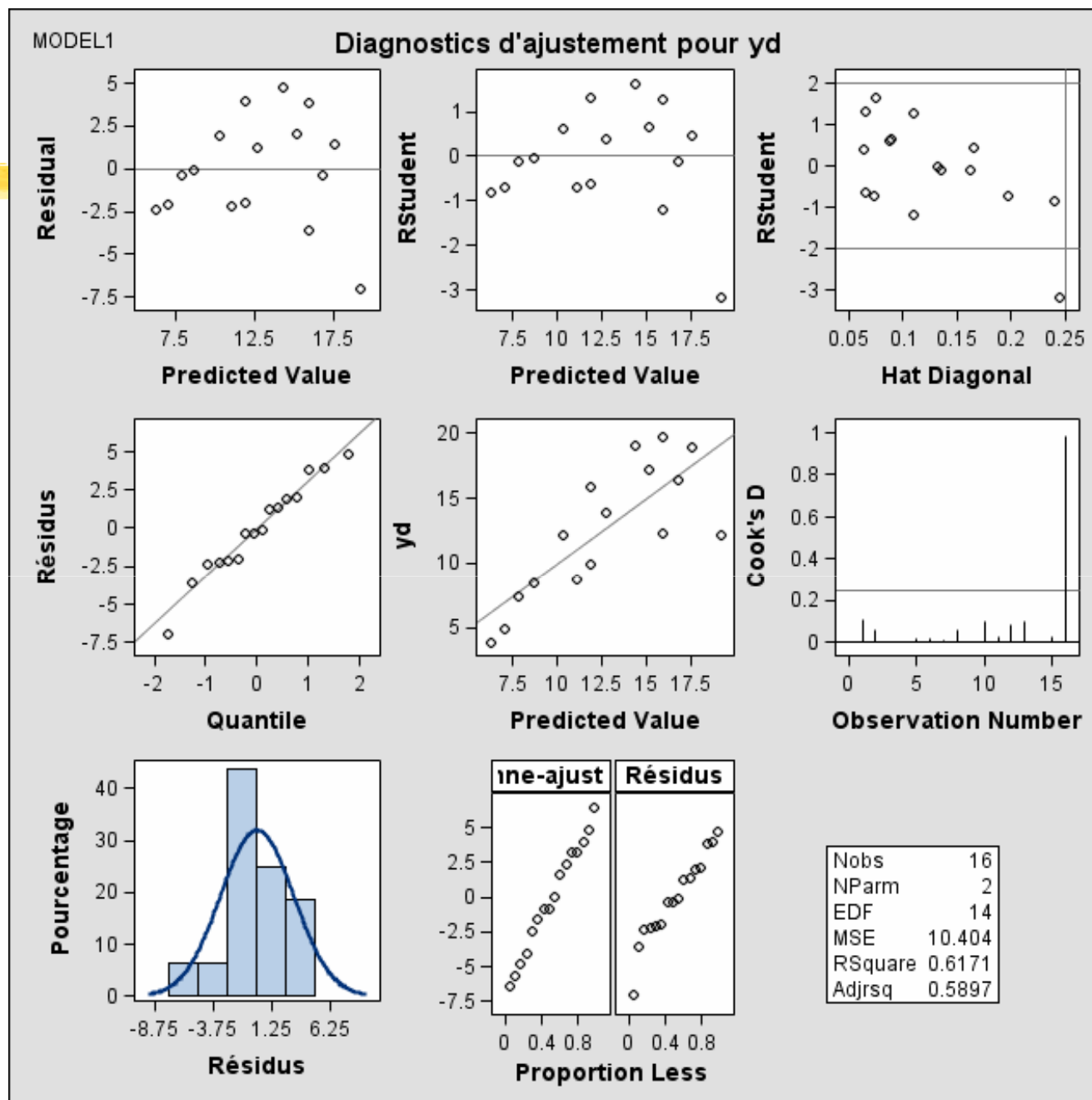
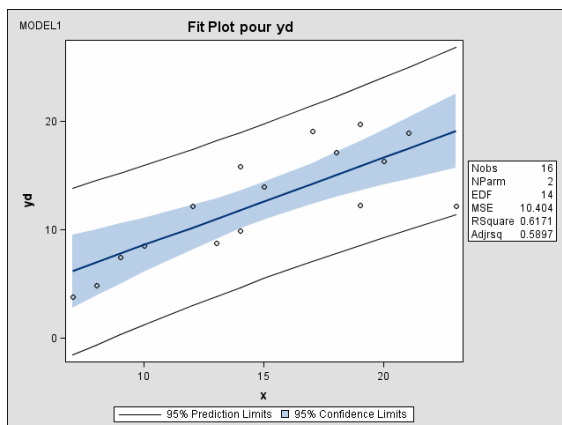
Régression 3 :

Durbin-Watson D	1.289
Number of Observations	16
1st Order Autocorrelation	-0.015



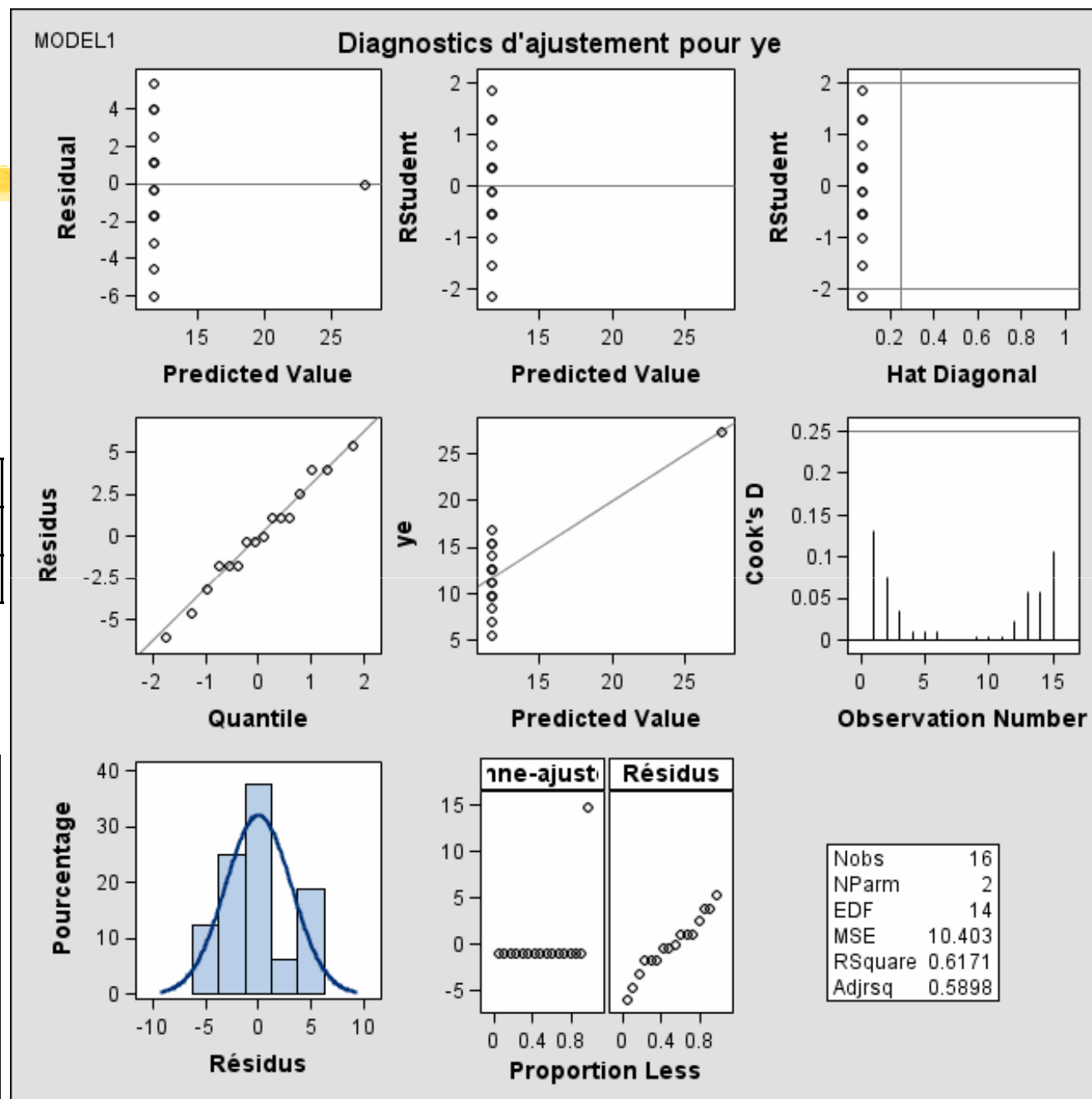
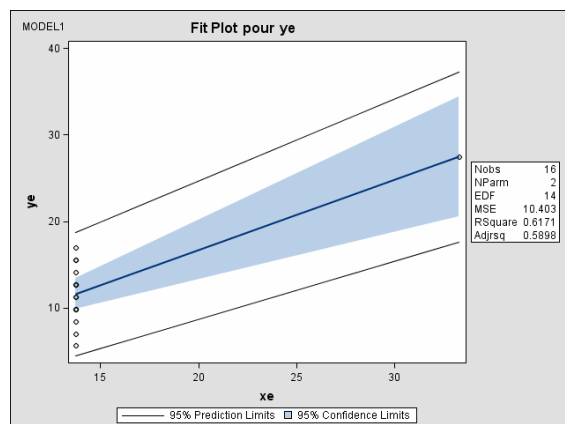
Régression 4 :

Durbin-Watson D	1.821
Number of Observations	16
1st Order Autocorrelation	-0.094



Régression 5 :

Durbin-Watson D	0.310
Number of Observations	16
1st Order Autocorrelation	0.723



Attention à la multicolinéarité

- Multicolinéarité = plusieurs variables explicatives (fortement) corrélées entre elles.
- Cela entraîne :
 - des coefficients de régression très sensibles aux fluctuations même faibles des données
 - des écarts-types élevés pour les coefficients de régression
 - une dégradation de la précision des prévisions
- Mesurée par :
 - tolérance $X_i = 1 - (\text{coefficient de détermination de la régression de } X_i \text{ sur les autres variables})$
 - doit être $> 0,2$
 - $VIF = 1 / \text{tolérance}$
 - doit être < 5

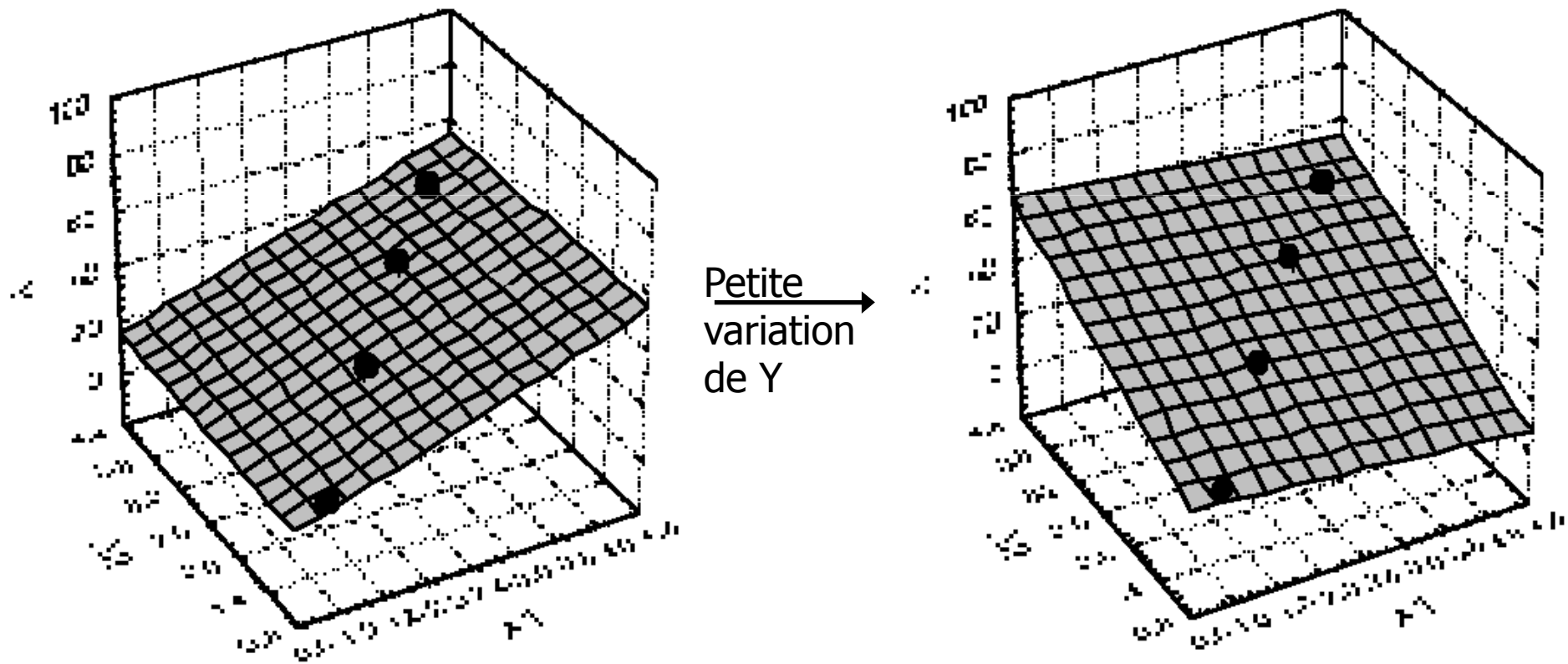
Attention à la multicolinéarité

- Autre mesure possible : les indices de conditionnement de la matrice des corrélations
 - on a multicolinéarité modérée (resp. forte) si présence d'indices $\eta_k > 10$ (resp. 30)
 - on regarde si on peut relier la valeur propre correspondante à une forte contribution ($> 50\%$) de la composante à la variance de 2 ou plusieurs variables

Modèle	Dimension	Valeur propre	Indice de conditionnement	Proportions de la variance		
				(constante)	TEMPERAT	ISOLATIO
1	1	2,145	1,000	,03	,07	,03
	2	,766	1,673	,02	,92	,02
	3	,089	4,915	,95	,01	,95

Effets de la multicollinéarité

- X_1 et X_2 presque colinéaires \Rightarrow coefficients de la régression très sensibles à de petites variations de Y



Solutions à la multicolinéarité

- Suppression des variables concernées
 - accepter de baisser un peu R^2 pour baisser la multicolinéarité
- Transformation (logarithme...) des variables concernées
- Régression biaisée (ridge)
 - l'erreur quadratique de l'estimation de la pente β de la régression = $\text{variance_estimateur} + (\text{biais_estimateur})^2$, d'où une « erreur quadratique avec biais » < « erreur sans biais » si le biais est compensé par une faible variance
- Régression sur composantes principales
 - passer ensuite des coefficients de régression des composantes principales à ceux des variables initiales
- Régression PLS (Partial Least Squares)
 - utilisable même si : nb observations \ll nb variables
 - on démontre (De Jong, 1993) que la régression PLS sur k composantes est toujours plus prédictive que la régression sur les k premières composantes principales



Technique de prédiction : La régression PLS

La méthode Partial Least Squares

- C'est une méthode qui se juxtapose à d'autres méthodes de régression (linéaire, logistique, analyse discriminante)
- Utile en présence d'un grand nombre de variables présentant de la colinéarité ou des valeurs manquantes
- Algorithme simple (suite de régressions simples, sans inversion ni diagonalisation de matrices) \Rightarrow efficace sur de grands volumes de données
- Utilisation en chimie, industrie pétrolière, cosmétique, biologie, médecine, agroalimentaire
 - en cosmétique : conserver tous les ingrédients d'un produit \Rightarrow très nombreuses variables explicatives
 - en agroalimentaire (analyse sensorielle) : expliquer le classement d'un produit par plusieurs dégustateurs (variable Y), en fonction de ses propriétés (jusqu'à plusieurs centaines) physico-chimiques et de saveur

Principe de la régression PLS

- Régression PLS inventée par Herman et Svante Wold (1983)
- On a Y variable à expliquer et X_i variables explicatives
- Le choix des variables transformées résulte d'un compromis entre :
 - maximisation de la variance des X_i (ACP)
 - maximisation de la corrélation entre X_i et Y (régression)
 - donc : on cherche les combinaisons linéaires T_j des X_i maximisant $\text{cov}^2(T_j, Y) = r^2(T_j, Y) \cdot \text{var}(T_j) \cdot \text{var}(Y)$

Étape 1 de la régression PLS

- On cherche une combinaison $T_1 = \sum_i \lambda_{1i} X_i$ des X_i qui maximise la variance de T_1 et la corrélation entre T_1 et Y
 - \Leftrightarrow maximiser $\text{cov}^2(T_1, Y) = r^2(T_1, Y) \cdot \text{var}(T_1) \cdot \text{var}(Y)$
- La solution est $\lambda_{1i} = \text{cov}(Y, X_i)$
 - en normant $\|(\lambda_{11}, \dots, \lambda_{1p})\| = 1$
 - on a donc $T_1 = \sum_i \text{cov}(Y, X_i) \cdot X_i$
- La régression de Y sur T_1 donne un résidu Y_1 :
 - $Y = c_1 T_1 + Y_1$
- La régression de X_i sur T_1 donne aussi des résidus X_{1i} :
 - $X_i = c_{1i} T_1 + X_{1i}$
- On réitère en remplaçant Y par Y_1 et les X_i par les $X_{1i} \Rightarrow$ étape 2

Etape 2 de la régression PLS

- On répète la même opération en remplaçant Y par son résidu Y_1 et les X_i par leurs résidus X_{1i}
- On obtient une combinaison $T_2 = \sum_i \lambda_{2i} X_i$ en normant $||(\lambda_{21}, \dots, \lambda_{2p})|| = 1$
- Puis on régresse Y_1 sur T_2 et les X_{1i} sur T_2 : on obtient des résidus Y_2 et X_{2i}
 - $Y_1 = c_2 T_2 + Y_2$
 - $X_{1i} = c_{2i} T_2 + X_{2i}$
- On réitère jusqu'à ce que le nb de composantes T_k donne un résultat satisfaisant (vérifié par validation croisée)
- A la fin, on a :
 - $Y = c_1 T_1 + Y_1 = c_1 T_1 + c_2 T_2 + Y_2 = \sum_j c_j T_j + \text{résidu}$
- Et on remplace cette expression par une expression de la régression de Y en fonction des X_i

Choix du nombre de composantes 1/2

- On procède généralement par validation croisée
- On se place à l'étape h et on veut décider de conserver ou non la composante h
- On calcule la somme des carrés résiduels (REsidual Sum of Squares), comme en régression linéaire :

$$\text{RESS}_h = \sum_k (y_{(h-1),k} - \hat{y}_{(h-1),k})^2$$

où $\hat{y}_{(h-1),k} = c_h t_{h,k}$ = prévision de $y_{(h-1),k}$ calculée pour chaque observation k

- Ensuite, les observations sont partagées en G groupes, et on réalise G fois l'étape courante de l'algorithme PLS sur Y_{h-1} et les $X_{h-1,i}$ en ôtant chaque fois un groupe

Choix du nombre de composantes 2/2

- Puis on calcule la somme *prédite* des carrés résiduels (Predicted REsidual Sum of Squares) $PRESS_h$
- Analogue à la précédente mais qui évite le sur-apprentissage en remplaçant la prévision $\hat{y}_{(h-1),k}$ par la prévision $\hat{y}_{(h-1),-k}$ déduite de l'analyse réalisée sans le groupe contenant l'observation k
- $PRESS_h = \sum_k (y_{(h-1),k} - \hat{y}_{(h-1),-k})^2$
- On retient la composante h si : $PRESS_h \leq \gamma \cdot RESS_{h-1}$
en posant $RESS_0 = \sum (y_i - \bar{y})^2$
Souvent : on fixe $\gamma = 0,95$ si $n < 100$, et $\gamma = 1$ si $n \geq 100$

Nombre de composantes PLS

- Cette sélection par validation croisée permet de retenir un nombre de composantes :
 - suffisamment grand pour expliquer l'essentiel de la variance des X_i et de Y
 - suffisamment petit pour éviter le sur-apprentissage
- En pratique le nombre de composantes dépasse rarement 3 ou 4
- Notons également que la régression PLS sur k composantes est toujours plus prédictive que la régression sur les k premières composantes principales

Généralisations de la régression PLS

- Régression PLS2 développée pour prédire plusieurs Y_j simultanément
 - on peut avoir $\text{nb}(Y_j) \gg \text{nb observations}$
- Régression logistique PLS développée par Michel Tenenhaus (2000)
 - algorithme analogue au précédent
- Et régression logistique sur composantes PLS, équivalente à la régression logistique PLS mais plus simple :
 - on commence par une régression PLS de l'indicatrice de Y sur les X_i (ou des indicatrices de Y , si Y a plus de 2 modalités)
 - on obtient k composantes PLS (éventuellement : $k = 1$)
 - puis on effectue une régression logistique de Y sur les composantes PLS



Technique de prédiction : La régression robuste

Régression robuste



- Méthodes valides quand les résidus des observations ne suivent pas une loi normale
- Peu sensibles aux « outliers »
- De plus en plus répandues dans les logiciels statistiques
 - SAS, R, S-PLUS, STATA...

Algorithmes de régression robuste

- Moindres médianes de carrés
- Moindres carrés winsorisés (least winsored squares)
 - remplacement des x centiles extrêmes par Q_x
- Moindres carrés écrêtés (least trimmed squares)
 - suppression des x centiles extrêmes
- Moindres carrés pondérés
 - par l'inverse de la variance de la variable à expliquer, pour compenser l'hétéroscédasticité, en posant par ex. $p_i = s^2/s_i^2$ au voisinage d'un point x_i
- Moindres carrés localement pondérés sur les voisins (LOESS)
- Doubles moindres carrés

- Régression spline
- Méthode du noyau

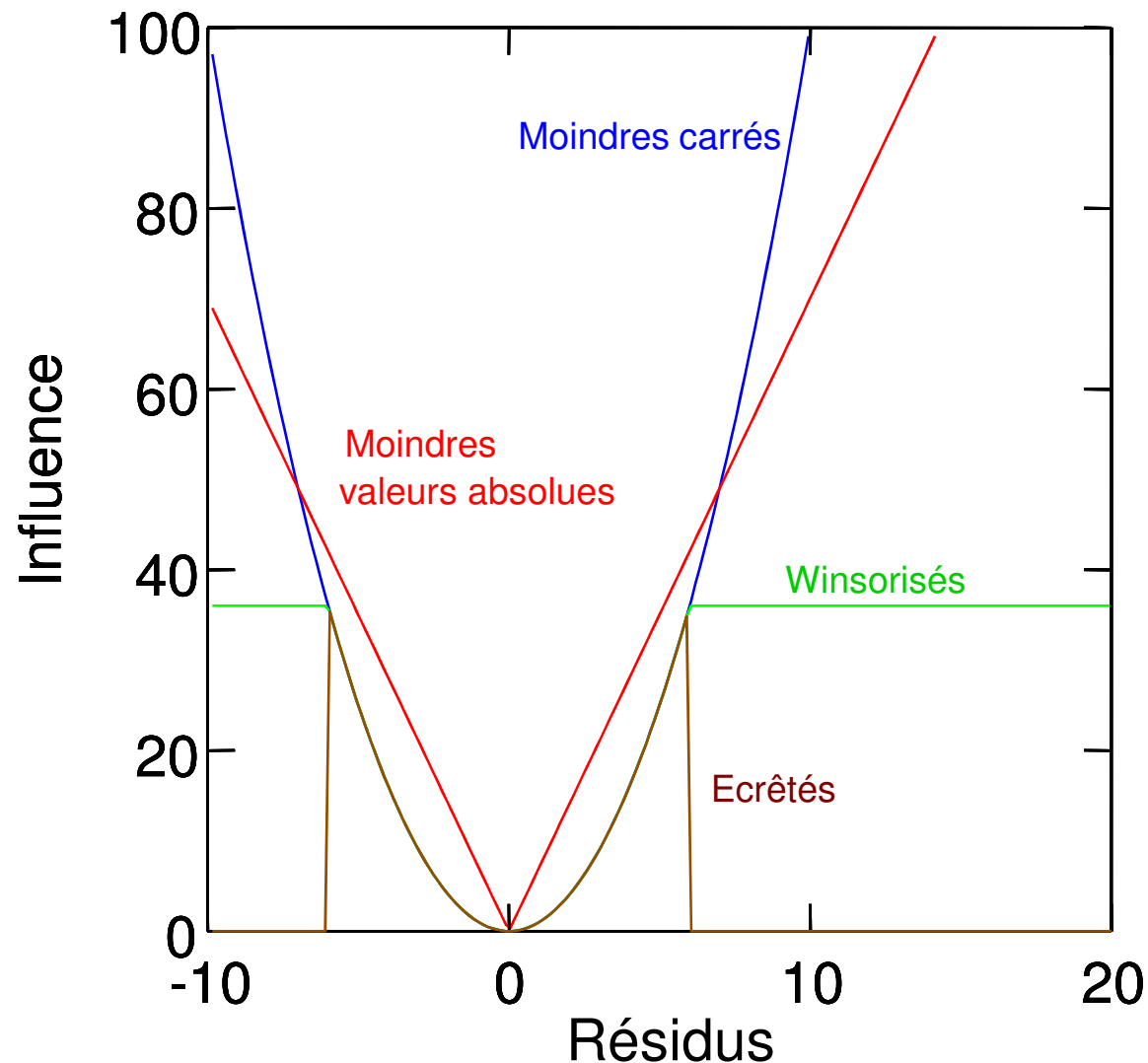
Autres algorithmes de régression

- Moindres valeurs absolues

$$\sum_i |x_i - \bar{x}|$$

- Régression polynomiale
- Régression sur variables qualitatives par codage optimal (moindres carrés alternés)

Influence des résidus



Principe de la régression LOESS

- Pour chaque point x : on prend les n voisins
 - le nombre n est choisi pour représenter un certain % de l'ensemble des points
 - ce % est appelé « paramètre de lissage » (« smoothing parameter »)
 - il existe des critères pour le choix de ce paramètre
- On pondère chacun de ces n points selon une fonction décroissante de leur distance à x
- On calcule la régression pondérée sur les n voisins pour prédire x
- LOESS utilisable avec plusieurs régresseurs
- Initiateur : Cleveland (1979)

Exemples de régressions LOESS

