

# project

Yue Zhang et G  r  mi Bridonneau

## Introduction

On va   tudier une transformations non lin  aires et en particulier la transformation de Box-Cox r  guli  rement utilis   pour stabiliser la variance et corriger les asym  tries des donn  es. On va dans un premier temps   tudier th  oriquement cette transformation et comment obtenir les param  tres optimaux (dans ce travail on met en oruvre l'estimation maximum vraisemblance) de cette transformation, et puis   tudier l'intervalle de confiance et tests statistiques sur l'estimateur. Dans un second temps on testera notre transformation sur des donn  es simul  es puis nous terminerons par un cas pratique sur l'  tude du nombre de cycles    rupture d'un fil peign   en fonction de certains param  tres, en   tudiant la regression mod  le lin  aire d'ordre 1, d'ordre 2 et les choix des variables.

## 1 La transformation de Box-Cox

On s'int  resse au mod  le d'observation  $(x_i, Y_i)$ :

$$h_\lambda(Y_i) = Z_i = x_i\theta + \varepsilon_i, \varepsilon_i \stackrel{iid}{\sim} \mathcal{N}(0, \sigma^2) \quad (1)$$

o    $(h_\lambda)$  est une famille de transformations param  tr  es par  $\lambda$ .

### Etude de la tranformation de Box-Cox.

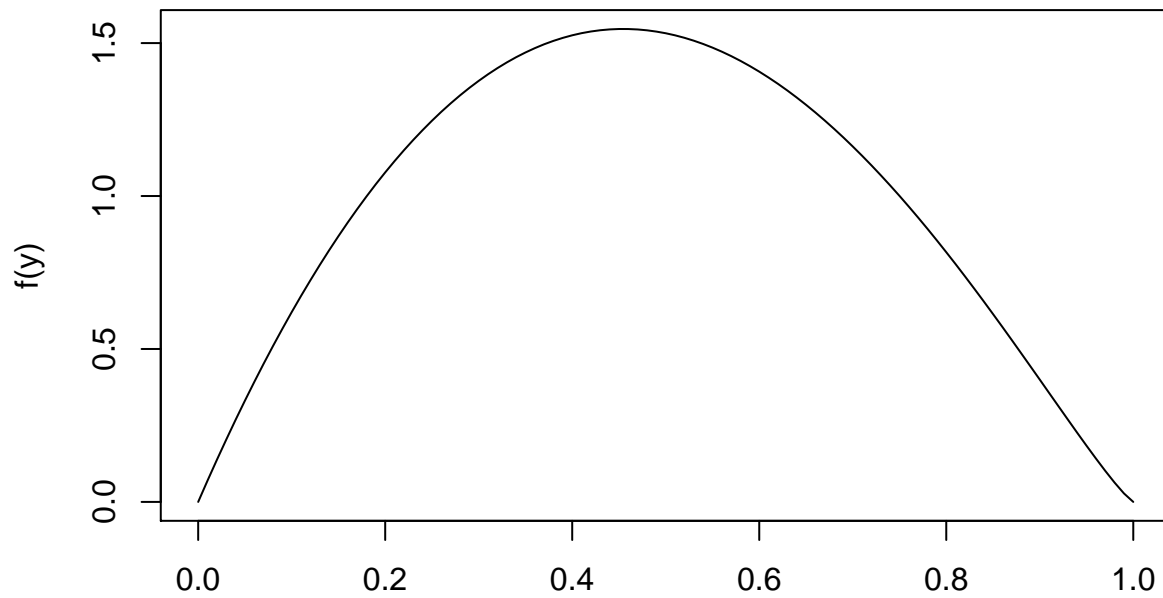
Ici on s'int  resse    la transformation de Box-Cox d  finie par

$$\forall \lambda \in \mathbb{R}, \forall y > 0, \tilde{h}_\lambda(y) = \begin{cases} \frac{y^\lambda - 1}{\lambda} & \lambda \neq 0 \\ \log y & \lambda = 0 \end{cases}$$

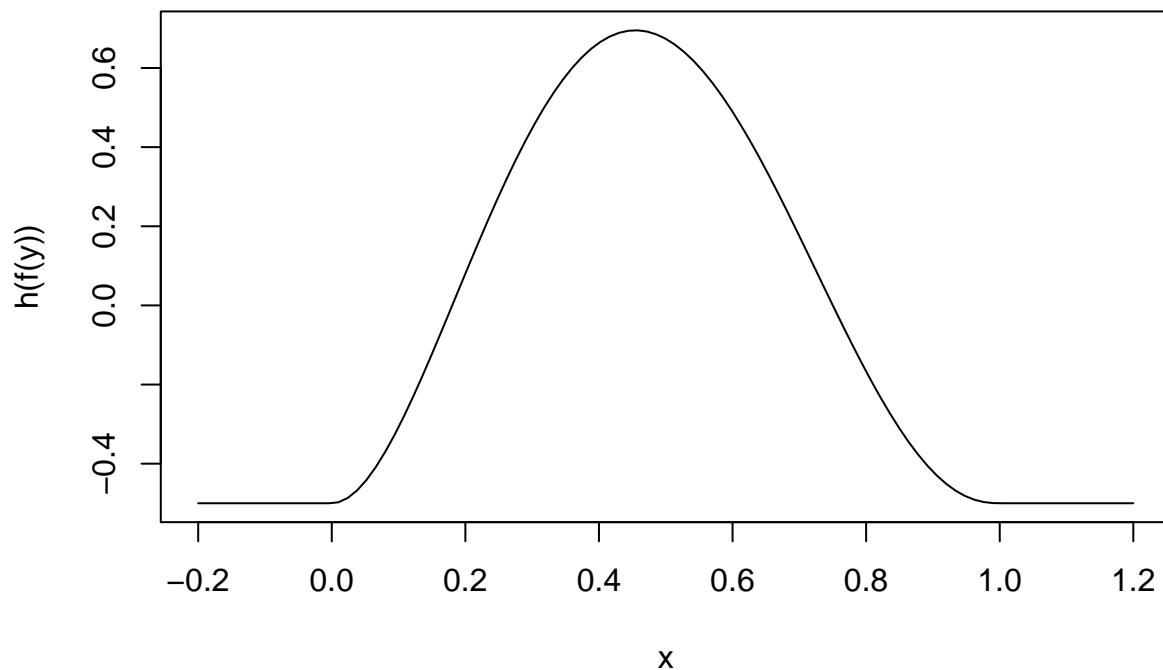
On remarque que th  oriquement cette transformation est incompatible avec le mod  le 1. En effet la transformation  $\tilde{h}_\lambda$  est valable seulement pour  $y > 0$ . De plus pour tout  $\lambda \neq 0$ , la transformation  $\tilde{h}_\lambda$  est born   et donc la transformation ne peut pas   tre gaussienne. Pour  $\lambda = 0$  on n'a pas ce probl  me gr  ce    la surjectivit   du logarithme.

Si toute les observations sont positives on peut quand m  me utiliser cette transformation car on perdra qu'une faible partie des donn  es normalement dans la queue    gauche de la r  partition. Par exemple si les donn  es ne suivent pas une loi normale mais une loi beta de param  tre  $\alpha = 2, \beta = 2.2$  et qu'on utilise la transformation de Box et Cox avec  $\lambda = 2$  on obtient:

### Loi beta de paramètres $\alpha=2$ , $\beta=2.2$



La même loi beta après une transformation de Box et Cox avec  $\lambda$



On voit qu'on a aucune valeur négative et que donc la gaussianisation n'est pas parfaite mais cette transformation reste raisonnable.

## 2. Déterminer la fonction de vraisemblance

Supposons que pour  $\beta = (\theta, \lambda, \sigma^2)'$  a  $p \times 1$  vecteur de paramètres, on ait  $h_\lambda(Y_i) = Z_i = x_i\theta + \varepsilon_i$ ,  $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$ ,  $\varepsilon_i$  suivent une loi gaussien i.i.d. Donc par la définition de vraisemblance:

$$\begin{aligned}
 L(\lambda, \theta, \sigma^2; Y) &= \prod_{i=1}^n \frac{\partial F}{\partial Y_i} \\
 &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(h_\lambda(Y_i) - x_i\theta)^2}{2\sigma^2}\right) \left| \frac{\partial h_\lambda(Y_i)}{\partial Y_i} \right| \\
 &= \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} \exp\left(-\frac{\sum_{i=1}^n (h_\lambda(Y_i) - x_i\theta)^2}{2\sigma^2}\right) \prod_{i=1}^n \left| \frac{\partial h_\lambda(Y_i)}{\partial Y_i} \right| \\
 &= \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} \exp\left(-\frac{(h_\lambda(Y) - X\theta)'(h_\lambda(Y) - X\theta)}{2\sigma^2}\right) \prod_{i=1}^n |Y_i^{\lambda-1}|
 \end{aligned} \tag{2}$$

Donc le terme  $J(\lambda; Y) = \prod_{i=1}^n \left| \frac{\partial h_\lambda(Y_i)}{\partial Y_i} \right| = \prod_{i=1}^n |Y_i^{\lambda-1}|$ , est la transformation de Jacobian du terme  $(h_\lambda(Y) - X\theta)$  à  $Y$ .

## 3. Estimation du maximum de vraisemblance

A  $\lambda$  fixé, on souhaite déterminer l'estimateur du maximum de vraisemblance  $\hat{\theta}(\lambda)$  et  $\hat{\sigma}^2(\lambda)$ . Donc tout d'abord, depuis l'équation 2 on calcule la log-vraisemblance.

$$\begin{aligned}
 \ell &= \log L(\lambda, \theta, \sigma^2; Y) \\
 &= -\frac{n}{2} \log(2\pi\sigma^2) - \frac{(h_\lambda(Y) - X\theta)'(h_\lambda(Y) - X\theta)}{2\sigma^2} + \sum_{i=1}^n \log |Y_i^{\lambda-1}| \\
 &= -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\sigma^2) - \frac{\|h_\lambda(Y) - X\theta\|^2}{2\sigma^2} + (\lambda - 1) \sum_{i=1}^n \log |Y_i|
 \end{aligned} \tag{3}$$

Ensuite, étant donné que la log-vraisemblance  $\ell$  l'équation (3) est une transformation monotone de la vraisemblance  $L$  dans l'équation (2), on maximise la log-vraisemblance  $\ell$  respectivement pour  $\theta$ ,  $\sigma^2$  et  $\lambda$ , donc on obtient le premier ordre dérivation ci-dessous:

$$\frac{\partial \ell}{\partial \sigma^2} = -\frac{n}{2} \frac{1}{\sigma^2} + \frac{(h_\lambda(Y) - X\theta)'(h_\lambda(Y) - X\theta)}{2\sigma^4} = 0 \tag{4}$$

Donc on a  $\hat{\sigma}^2 = \frac{(h_\lambda(Y) - X\theta)'(h_\lambda(Y) - X\theta)}{n} = \frac{h_\lambda(Y)'(I_n - H)h_\lambda(Y)}{n}$ , avec  $H = X(X'X)^{-1}X'$  et  $I_n$  matrice identité.

$$\begin{aligned}
 \frac{\partial \ell}{\partial \theta} &= -\frac{2(-X)'(h_\lambda(Y) - X\theta)}{2\sigma^2} \\
 &= \frac{X'(h_\lambda(Y) - X\theta)}{\sigma^2} = 0
 \end{aligned} \tag{5}$$

Donc,  $\hat{\theta} = (X'X)^{-1}X'h_\lambda(Y)$ , par  $X'h_\lambda(Y) = X'X\theta$ .

Pour vérifier la formule avec  $L_{max}(\lambda)$ , on remplace nos env  $\hat{\sigma}^2$  et  $\hat{\theta}$  calculés dans les équations (4) et (5) dans la log-vraisemblance  $\ell$ :

$$\begin{aligned} L_{max}(\lambda) &:= \ell = \log L(\lambda, \hat{\theta}(\lambda), \hat{\sigma}^2(\lambda)) \\ &= -\frac{n}{2} \log\left(\frac{\|h_\lambda(Y) - X\theta\|^2}{n}\right) - \frac{\|h_\lambda(Y) - X\theta\|^2 n}{2\|h_\lambda(Y) - X\theta\|^2} + (\lambda - 1) \sum_{i=1}^n \log |Y_i| - \frac{n}{2} \log(2\pi) \\ &= -\frac{n}{2} \log(\hat{\sigma}^2(\lambda)) + (\lambda - 1) \sum_{i=1}^n \log |Y_i| - \frac{n}{2} - \frac{n}{2} \log(2\pi) \end{aligned} \quad (6)$$

Donc  $a(n) = -\frac{n}{2} - \frac{n}{2} \log(2\pi)$  qui est bien une constante ne dépendant que de  $n$ . Maintenant on calcule l'env  $\hat{\lambda}$ :

$$\frac{\partial \ell}{\partial \lambda} = -\frac{2(h_\lambda(Y) - X\theta) \left| \frac{\partial h_\lambda(Y)}{\partial \lambda} \right|}{2\sigma^2} + \sum_{i=1}^n \log |Y_i| = 0 \quad (7)$$

Et

$$\begin{aligned} \frac{\partial L_{max}}{\partial \lambda} &= -\frac{n}{2} \frac{1}{\hat{\sigma}^2(\lambda)} \left| \frac{\partial \hat{\sigma}^2(\lambda)}{\partial \lambda} \right| + \sum_{i=1}^n \log |Y_i| \\ &= -\frac{n}{2} \frac{1}{\hat{\sigma}^2(\lambda)} \frac{2(h_\lambda(Y) - X\theta) \left| \frac{\partial h_\lambda(Y)}{\partial \lambda} \right|}{n} + \sum_{i=1}^n \log |Y_i| \\ &= -\frac{(h_\lambda(Y) - X\theta) \left| \frac{\partial h_\lambda(Y)}{\partial \lambda} \right|}{\hat{\sigma}^2} + \sum_{i=1}^n \log |Y_i| = 0 \end{aligned} \quad (8)$$

On peut bien vérifier que  $\frac{\partial \ell}{\partial \lambda}$  et  $\frac{\partial L_{max}}{\partial \lambda}$  sont égaux par calcul de l'équation du maximum de vraisemblance.

Par l'équation (4), on sait que  $\hat{\sigma}^2(\lambda) = \frac{h_\lambda(Y)'(I_n - H)h_\lambda(Y)}{n} = \frac{SCR(\lambda)}{n}$  avec  $H = X(X'X)^{-1}X'$ , est la somme de carrés résiduels de variance  $h_\lambda(Y)$  divisée par  $n$ . Depuis l'équation (8), on peut continuer ce calcul en remplaçant  $\hat{\sigma}^2$  avec pour rappel  $h_\lambda(Y) = \frac{Y^\lambda - 1}{\lambda}$ :

$$\begin{aligned} \frac{\partial L_{max}}{\partial \lambda} &= -\frac{n}{2} \frac{n}{h_\lambda(Y)'(I_n - H)h_\lambda(Y)} \frac{2h_\lambda(Y)'(I_n - H)}{n} \left( \frac{Y^\lambda \log Y}{\lambda} - \frac{Y^\lambda - 1}{\lambda^2} \right) + \sum_{i=1}^n \log |Y_i| \\ &= -n \frac{h_\lambda(Y)'(I_n - H)}{h_\lambda(Y)'(I_n - H)h_\lambda(Y)} \left( \frac{Y^\lambda \log Y}{\lambda} - \frac{h_\lambda(Y)}{\lambda} \right) + \sum_{i=1}^n \log |Y_i| \\ &= -n \frac{h_\lambda(Y)'(I_n - H)\lambda^{-1}Y^\lambda \log Y}{h_\lambda(Y)'(I_n - H)h_\lambda(Y)} + n \frac{h_\lambda(Y)'(I_n - H)h_\lambda(Y)}{h_\lambda(Y)'(I_n - H)h_\lambda(Y)\lambda} + \sum_{i=1}^n \log |Y_i| \\ &= -n \frac{h_\lambda(Y)'(I_n - H)u_\lambda(Y)}{h_\lambda(Y)'(I_n - H)h_\lambda(Y)} + \frac{n}{\lambda} + \sum_{i=1}^n \log |Y_i| \end{aligned} \quad (9)$$

avec  $u_\lambda(Y) = \lambda^{-1} Y^\lambda \log Y$ . Le numérateur dans l'équation (9) est la somme résiduelle des produits dans l'analyse de la covariance de  $h_\lambda(Y)$  et  $u_\lambda(Y)$ . Maintenant on utilise la transformation normalisée afin de simplifier le résultat, on définit  $z_\lambda(Y)$  ci-dessous:

$$\begin{aligned} z_\lambda(Y) &= \frac{h_\lambda(Y)}{J(\lambda; Y)^{1/n}} \\ &= \frac{h_\lambda(Y)}{(\prod_{i=1}^n |Y_i|)^{\lambda-1/n}} \end{aligned} \quad (10)$$

Donc  $\hat{\sigma}^2$  devient  $\hat{\sigma}^2(\lambda; z) = \frac{z_\lambda(Y)'(I_n - H)z_\lambda(Y)}{n} = \frac{SCR(\lambda; z)}{n}$ ,  $SCR(\lambda; z)$  est la somme des carrées résiduelle de  $z_\lambda(Y)$ . De plus,  $L_{max} = -\frac{n}{2} \log(\hat{\sigma}^2(\lambda; z)) + a(n)$ , donc on propose de trouver  $\hat{\lambda}$  qui maximise  $L_{max}(\lambda)$ , c'est à dire minimize  $\hat{\sigma}^2(\lambda; z)$ . Donc on cherche l'emc (estimateur des moindres carrées)

$$\hat{\lambda} = \arg \min_{\lambda} SCR(\lambda; z) \quad (11)$$

Par le théorème du cours, l'emv est asymptotiquement normal, donc la distribution de  $\sqrt{n}(\hat{\beta} - \beta)$ , quand  $n \rightarrow \infty$ , elle converge en une loi normale.

$$\hat{V}^{-1/2} \sqrt{n}(\hat{\beta} - \beta) \rightarrow \mathcal{N}(0, I_{dp}) \quad (12)$$

$I_1(\beta)^{-1}$  est la matrice de l'information de Fisher, noté que  $\hat{V} = I_1(\beta)^{-1}$  et  $I_{dp}$  est la matrice identité de taille  $p$ . Quand  $n \geq 30$ , par le théorème *TCL*,  $\hat{\beta}$  tends vers un vecteur gaussien, donc à distance finie la distribution de  $\hat{\beta}$  approche la loi gaussienne.

#### 4. Distribution asymptotique de l'emv

##### Estimer la variance de $\hat{\lambda}$

Par la propriété de l'emv, quand  $\hat{\beta}$  tend à devenir gaussien, on peut prendre pour loi approchée à distance finie la loi asymptotique

$$\begin{aligned} \hat{V}^{-1/2}(\hat{\beta} - \beta) &\overset{appr}{\sim} \mathcal{N}(0, I_{dp}) \\ \hat{\beta} &\overset{appr}{\sim} \mathcal{N}(\beta, I_1(\beta)^{-1}) \end{aligned} \quad (13)$$

Par la définition, la matrice de l'information de Fisher est écrite ci-dessous:

$$\begin{aligned} I_1(\beta) &= \mathbb{E}_\beta[\dot{\ell}\dot{\ell}'] \\ &= -\mathbb{E}_\beta[\ddot{\ell}] \end{aligned} \quad (14)$$

où  $\ddot{\ell}$  est la matrice Hessienne  $\ddot{\ell} = \frac{\partial^2 \ell}{\partial \beta \partial \beta'}$  (pour rappelle on a définit  $\beta = (\theta, \lambda, \sigma^2)'$  a  $p \times 1$  vecteur de paramètres). En particulier, on n'a pas forcément besoin d'estimer  $\sigma^2$  simultanément avec  $\theta$  et  $\lambda$ , donc pour

simplifier les calculs, on décide de calculer la matrice Hessienne de  $L_{max}(\lambda)$ . Dans l'équation (8), on a calculé

$$\frac{\partial L_{max}}{\partial \lambda} = -\frac{(h_\lambda(Y) - X\theta) \left| \frac{\partial h_\lambda(Y)}{\partial \lambda} \right|}{\hat{\sigma}^2} + \sum_{i=1}^n \log |Y_i|, \text{ et on obtient sans souci } \frac{\partial L_{max}}{\partial \theta} = \frac{X'(h_\lambda(Y) - X\theta)}{\hat{\sigma}^2}.$$

$$\begin{aligned} \ddot{\ell} := H(\beta) &= \frac{\partial^2 L_{max}}{\partial \beta \partial \beta'} \\ &= -\hat{\sigma}^{-2} \begin{bmatrix} X'X & -X' \left| \frac{\partial h_\lambda(Y)}{\partial \lambda} \right| \\ -\left| \frac{\partial h_\lambda(Y)}{\partial \lambda} \right| X & \left| \frac{\partial h_\lambda(Y)}{\partial \lambda} \right|' \left| \frac{\partial h_\lambda(Y)}{\partial \lambda} \right| + \left| \frac{\partial^2 h_\lambda(Y)}{\partial^2 \lambda} \right| (h_\lambda(Y) - X\theta) \end{bmatrix} \end{aligned} \quad (15)$$

$H(\beta)$  est bien une matrice définie négative. Etant donnée la distribution asymptotique normale de l'emv, on peut conclure que  $\widehat{Var}(\hat{\beta}) = -[H(\hat{\beta})]^{-1}$ , maintenant on calcul  $\widehat{Var}(\hat{\lambda})$ :

$$\widehat{Var}(\hat{\lambda}) = -H(\hat{\lambda})^{-1} \quad (16)$$

$$\text{où } H(\lambda) = \frac{\partial^2 L_{max}}{\partial^2 \lambda} = \frac{\left| \frac{\partial h_\lambda(Y)}{\partial \lambda} \right|' \left| \frac{\partial h_\lambda(Y)}{\partial \lambda} \right| + \left| \frac{\partial^2 h_\lambda(Y)}{\partial^2 \lambda} \right| h_\lambda(Y)' (I_n - H)}{-\hat{\sigma}^2}.$$

### Intervalle de confiance

L'emv est asymptotiquement normalement distribué, par la propriété dans l'équation (13), on peut construire le test  $T = \frac{\hat{\beta} - \beta}{\sqrt{\widehat{Var}(\hat{\beta})}} \sim \mathcal{N}(0, I_{dp})$ . Par définition,  $P(q_{\alpha/2} < \frac{\hat{\beta} - \beta}{\sqrt{\widehat{Var}(\hat{\beta})}} < q_{1-\alpha/2}) = 1 - \alpha$ ,

donc on peut obtenir l'intervalle de confiance  $[\hat{\beta} - q_{1-\alpha/2} \sqrt{\widehat{Var}(\hat{\beta})}, \hat{\beta} + q_{1-\alpha/2} \sqrt{\widehat{Var}(\hat{\beta})}]$ , où  $q_{\alpha/2}$  et  $q_{1-\alpha/2}$  sont quantiles d'ordre  $\alpha/2$  et  $1 - \alpha/2$  sous la loi normale  $\mathcal{N}(0, 1)$ . La distribution est symétrique par rapport à 0, donc l'IC estimateur de  $\beta$  est également  $[\hat{\beta} - q_{1-\alpha/2} \sqrt{\widehat{Var}(\hat{\beta})}, \hat{\beta} + q_{1-\alpha/2} \sqrt{\widehat{Var}(\hat{\beta})}]$ .

L'intervalle de confiance de  $\lambda$  est donc  $[\hat{\lambda} - q_{1-\alpha/2} \sqrt{\widehat{Var}(\hat{\lambda})}, \hat{\lambda} + q_{1-\alpha/2} \sqrt{\widehat{Var}(\hat{\lambda})}]$ , où  $\widehat{Var}(\hat{\lambda})$  est calculé dans l'équation (16).

### Test de Wald

On définit  $A = [0, 1, 0]$ ,  $\beta = (\theta, \lambda, \sigma^2)'$ ,  $\beta_0 = (\theta_0, \lambda_0, \sigma_0^2)'$

$H_0 : A(\beta - \beta_0) = 0$  (i.e.  $\lambda - \lambda_0 = 0$ ), contre  $H_1 : A(\beta - \beta_0) \neq 0$  (i.e.  $\lambda - \lambda_0 \neq 0$ )

Sous  $H_0$ :

$$T = [AVA']^{-1/2} A(\hat{\beta} - \beta_0) \rightarrow \mathcal{N}(0, I_{dp}) \quad (17)$$

En utilisant la propriété de la statistique de Wald,  $W$  est la carré de la norme de  $T$  et sa loi asymptotique sous  $H_0$  est:

$$\begin{aligned}
W &= (A\hat{\beta} - A\beta_0)(A\hat{V}A')^{-1}(A\hat{\beta} - A\beta_0)' \\
&= \frac{(\hat{\lambda} - \lambda_0)^2}{\widehat{Var(\hat{\lambda})}} \rightarrow \chi^2(1)
\end{aligned} \tag{18}$$

où  $\hat{V} = I_1(\hat{\beta})^{-1}$ , et  $W \geq 0$ , la région de rejet est unilatère à droite de niveau asymptotique  $\alpha$  pour une hypothèse bilatère est  $\mathcal{R} = \left\{ W > q_{1-\alpha}^{\chi^2(1)} \right\}$  avec  $P_{(H_0)}(\mathcal{R}) \rightarrow \alpha$ .

## 5. Test du rapport vraisemblance

Par le théorème asymptotique du RV, sous  $H_0$ :

$$TRV = -2\log(RV) \rightarrow \chi^2(1) \tag{19}$$

$TRV \geq 0$ , la région de rejet  $\mathcal{R} = \left\{ TRV > q_{1-\alpha}^{\chi^2(1)} \right\}$  du test de rapport de vraisemblances maximales est asymptotiquement de niveau  $\alpha$ ,  $P_{(H_0)}(\mathcal{R}) \rightarrow \alpha$ .

Par la définition de rapport de vraisemblance:

$$RV = \frac{L(\lambda_0; Y)}{L(\hat{\lambda}; Y)} \tag{20}$$

où  $L$  est la fonction de vraisemblance.

$$\begin{aligned}
TRV &= -2\log\left(\frac{L(\lambda_0; Y)}{L(\hat{\lambda}; Y)}\right) \\
&= -2(\log L(\lambda_0; Y) - \log L(\hat{\lambda}; Y)) \\
&= 2(L_{max}(\hat{\lambda}; Y) - L_{max}(\lambda_0; Y)) \\
&= 2\left(-\frac{n}{2}\log(\hat{\sigma}^2(\hat{\lambda})) + \frac{n}{2}\log(\hat{\sigma}^2(\lambda_0))\right) \\
&= n\log\left(\frac{\hat{\sigma}^2(\lambda_0)}{\hat{\sigma}^2(\hat{\lambda})}\right)
\end{aligned} \tag{21}$$

## 2 Test de la méthode sur des données simulées

### 1. Modélisation la regression linéaire simple

#### Condition convergence

La condition de convergence indiquée dans la section 1 est bien vérifiée.

En effet on a que si  $x_1, x_2, \dots \stackrel{i.i.d}{\sim} \mathcal{N}(0, 1)$ ,

$$\frac{X'X}{n} = \frac{1}{n} \begin{bmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_n \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & \frac{1}{n} \sum_{i=1}^n x_i \\ \frac{1}{n} \sum_{i=1}^n x_i & \frac{1}{n} \sum_{i=1}^n x_i^2 \end{bmatrix} \quad (22)$$

On a de plus  $\mathbb{E}[x_i^2] = \text{Var}(x_i) + \mathbb{E}[x_i]^2 = 1$ . Ainsi d'après la loi des grands nombres on a  $(X'X)/n$  qui converge vers la matrice identité de taille  $2 \times 2$  qui est bien définie positive.

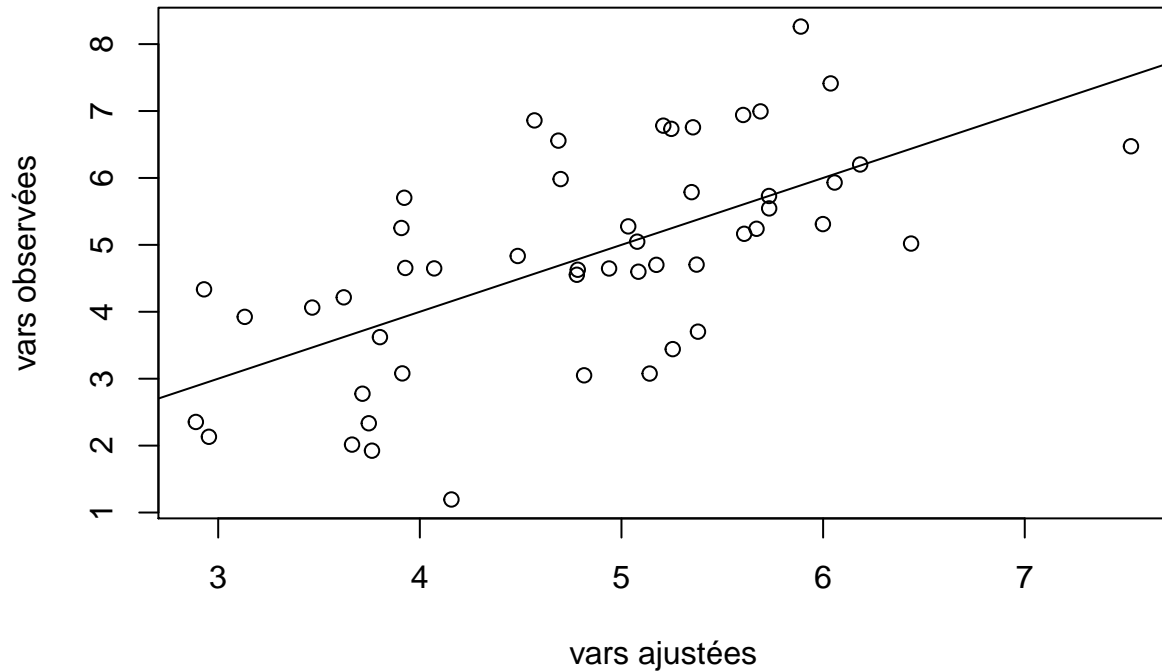
### Estimation de la regression linéaire simple

Le modèle linéaire simple donc est  $z_i = \mu + \theta x_i + \sigma \varepsilon_i$ ,  $\varepsilon_i \sim \mathcal{N}(0, 1)$  i.i.d. Le plan d'expérience  $X$  est en taille  $[n, p]$  i.e.  $[50 \times 2]$ , et  $\hat{\theta} = (X'X)^{-1}X'Z$ ,  $\hat{\sigma}^2 = \frac{1}{n-p} \|Z - X\hat{\theta}\|^2$ .

Par la définition,  $\frac{y_i^\lambda - 1}{\lambda} = h_\lambda(y_i) = z_i$ ,  $y_i = (\lambda z_i + 1)^{1/\lambda}$ .

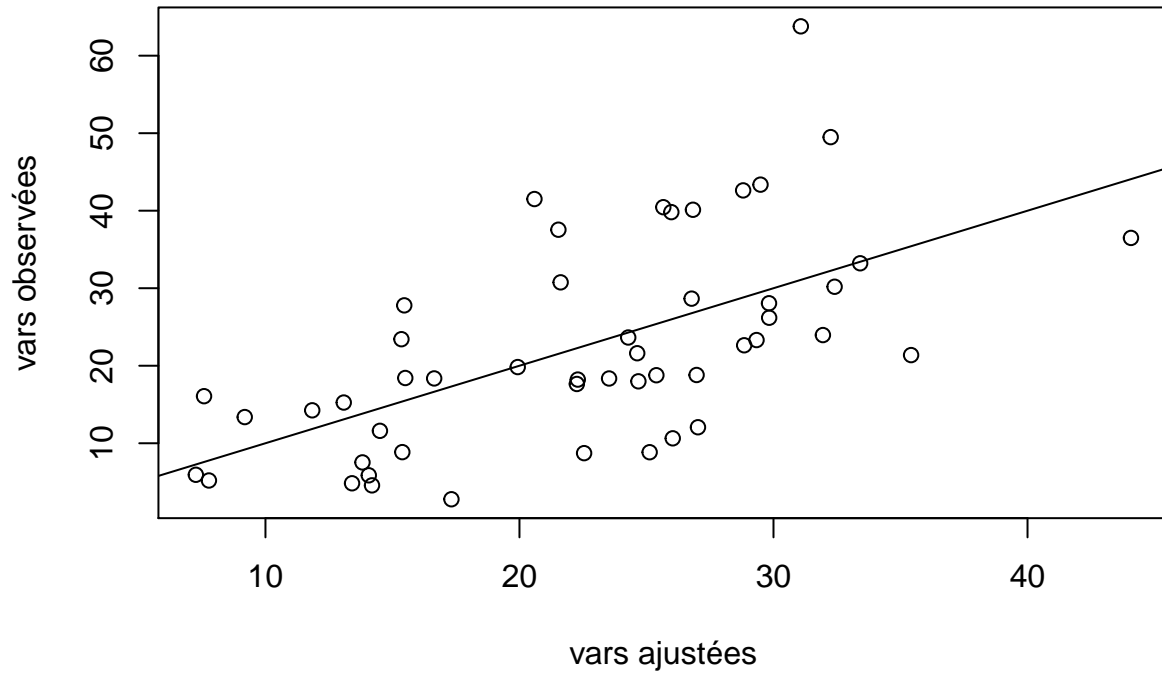
En Traçant la profondeur ajustée en fonction de la profondeur observée, on peut observer que le modèle n'est pas bien ajustée, les points s'allongent autour de la première bissectrice, mais les ajustements parfois assez s'éloignent valeurs observées,  $Y$  plus grave que  $Z$  visuellement, donc ce qui implique aussi ses bruts résidus fortes et potentiellement celle de  $Y$  est plus forte que celle de  $Z$ .

### Pour Z





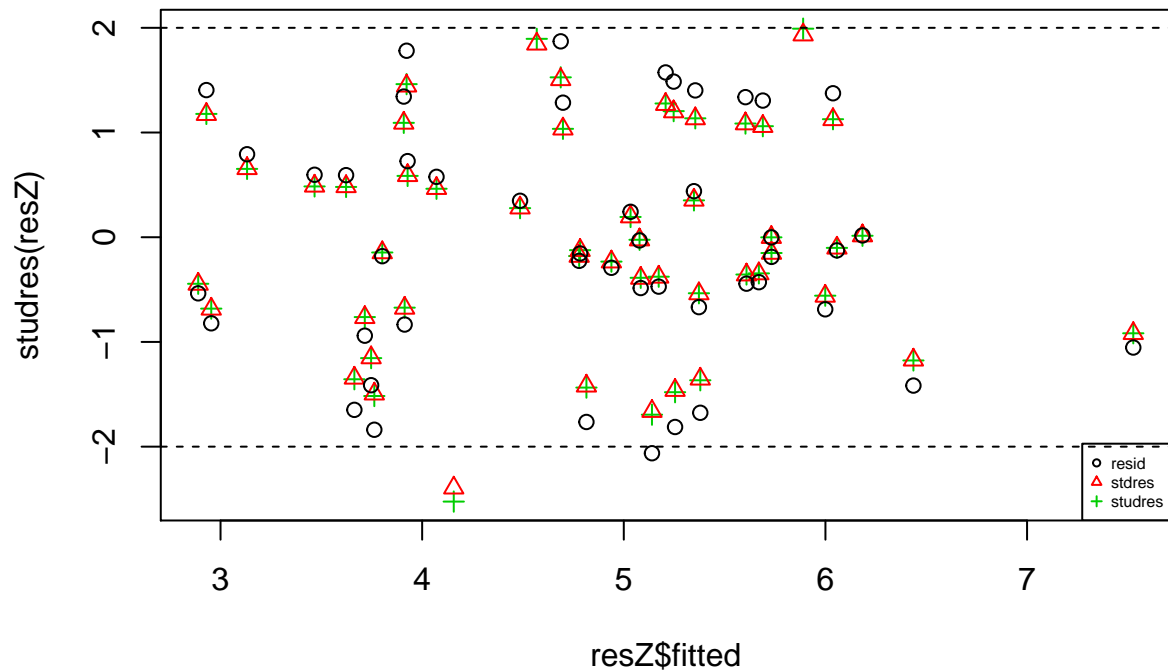
## Pour Y



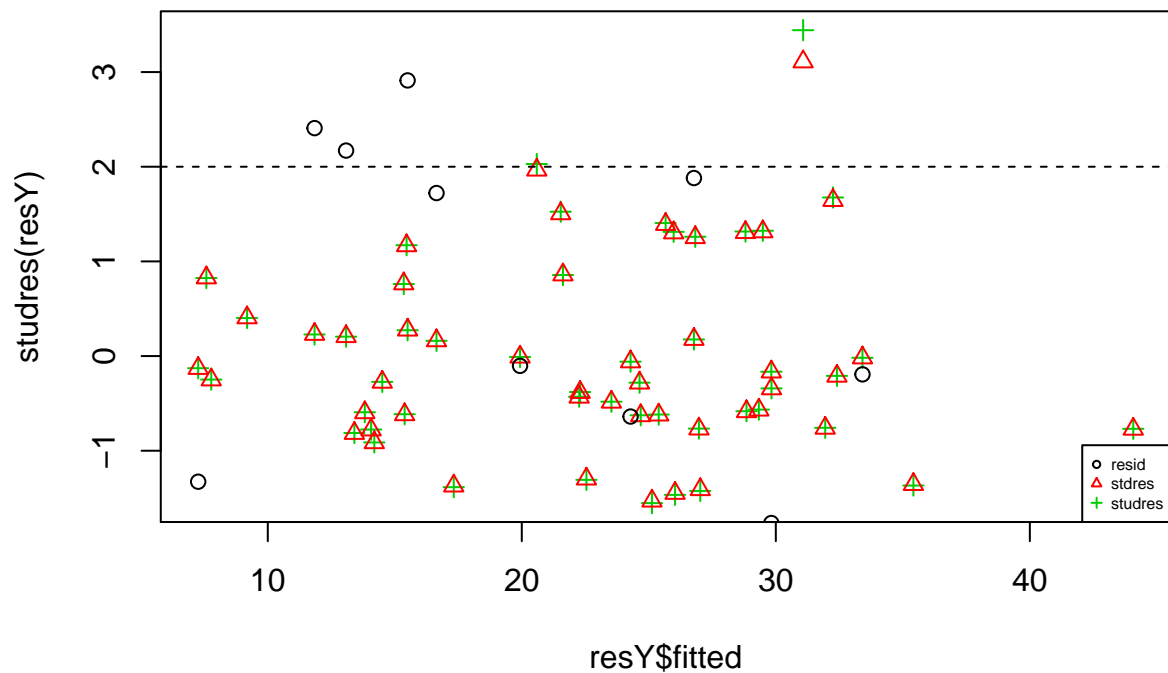
### Etude les résidus

Par la définition, les résidus  $\hat{\varepsilon} = Z - X\hat{\theta}$ . Et les résidus studentisés donc est  $t_i = \frac{\hat{\varepsilon}_i}{\hat{\sigma}\sqrt{1-h_{i,i}}}$  où  $h_{i,i}$  sont les éléments diagonaux de  $H = X(X'X)^{-1}X'$ .

Pour  $Z \sim X$ , on voit que les résidus bruts sont fortes parce que les valeurs observées sont elles-mêmes assez éloignés de 0 et estimées avec une faible précision. La plupart d'eux sont raisonnablement compris entre  $-2$  et  $2$  sauf qu'un seul point, en respectant la règle empirique d'appartenance de 95% des résidus à l'intervalle  $-2$  et  $2$ .

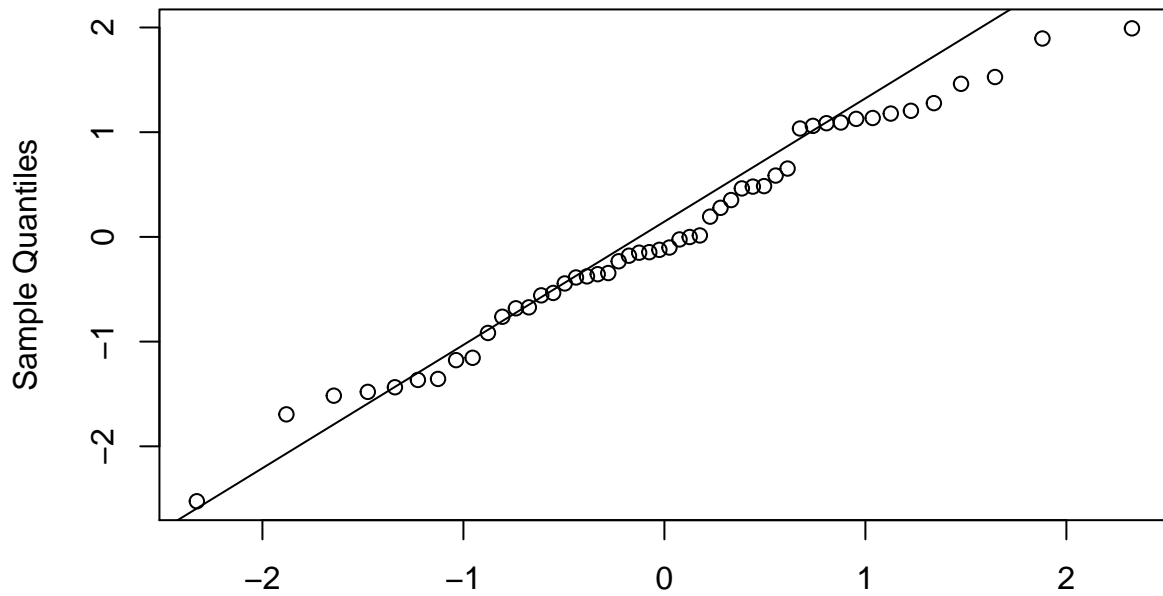


Pour  $Y \sim X$ , on voit que les résidus bruts sont plus fortes que celles de  $Z$ , parce que les valeurs observées sont elles-mêmes beaucoup plus éloignées de 0 et estimées avec une très faible précision. La plupart d'eux sont raisonnablement compris entre  $-2$  et  $2$  mais sauf trois points, en respectant aussi la règle empirique d'appartenance de 95% des résidus à l'intervalle  $-2$  et  $2$ .

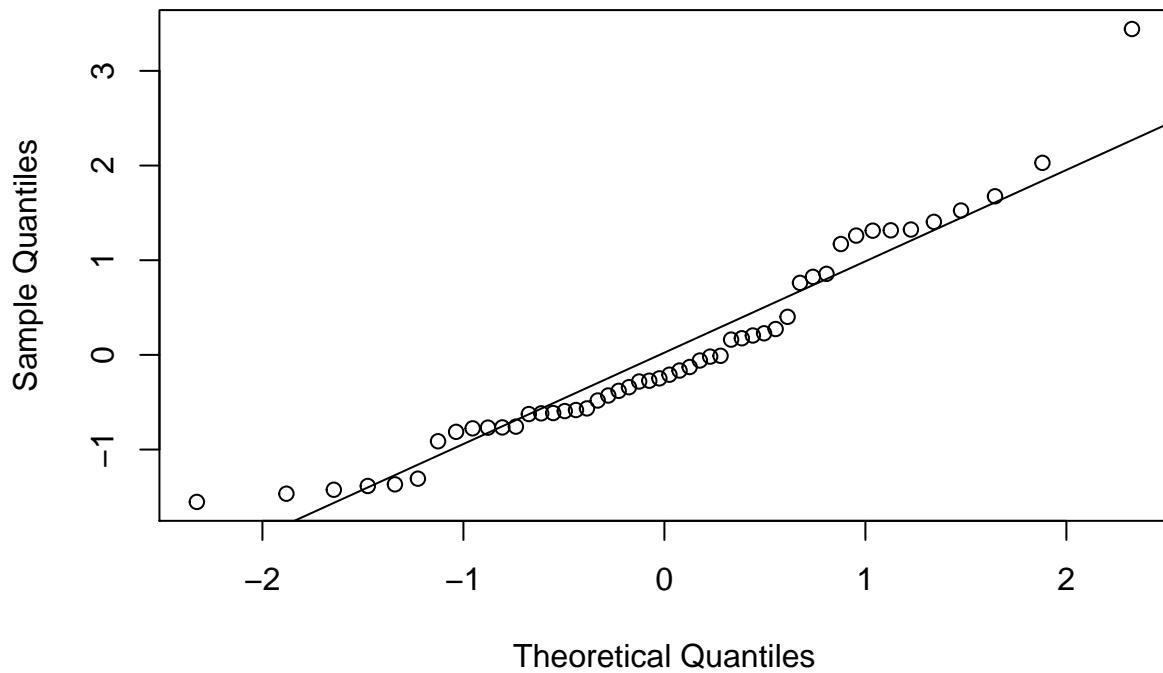


Pour  $Z \sim X$  les points s'alignent, mais pas pour les points qui se trouvent près des deux extrémités, donc on s'en doute la gaussianité des résidus. De même observation pour  $Y \sim X$ .

**graphe quantile-quantile Z~X**



Theoretical Quantiles  
**graphe quantile-quantile Y~X**



Donc on s'intéresse à faire le test Shapiro pour ses normalités. Pour  $Z \sim X$ , p-valeur égale à  $0.6019 > \alpha = 0.05$ , donc on garde l'hypothèse de gaussianité avec une risque de seconde espèce inconnue. Pour  $Y \sim X$ , p-valeur égale à  $0.007836 < \alpha = 0.05$ , donc on rejette l'hypothèse de gaussianité avec une risque de première espèce  $\alpha = 0.05$ .

##

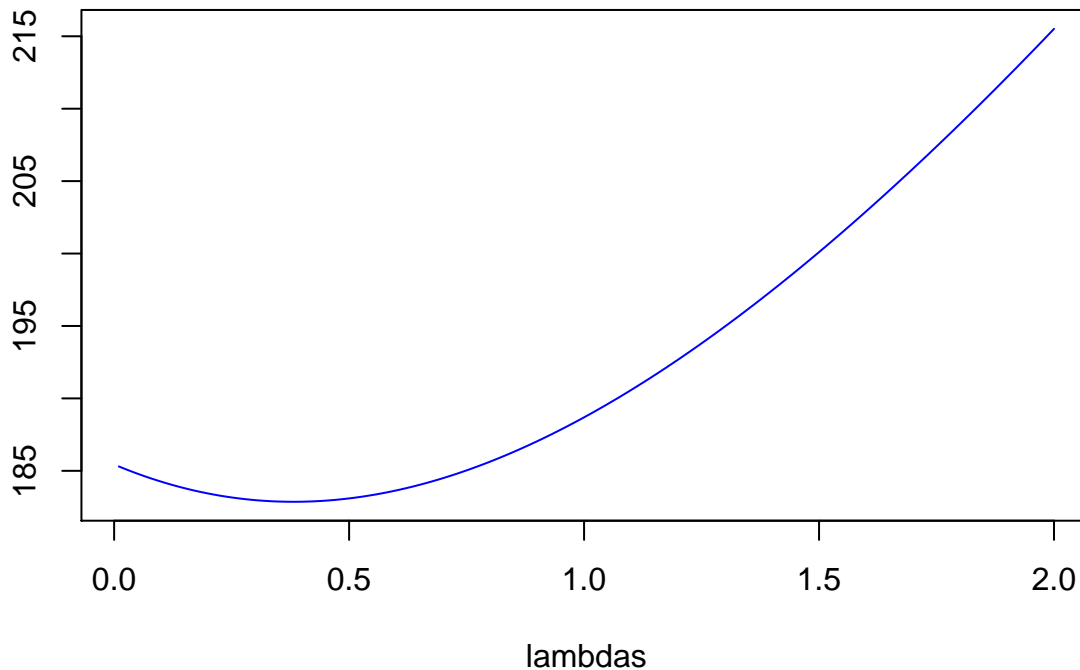
```
## Shapiro-Wilk normality test
##
## data: studres(resZ)
## W = 0.98116, p-value = 0.6019

##
## Shapiro-Wilk normality test
##
## data: studres(resY)
## W = 0.93397, p-value = 0.007836
```

## 2. Mise en oeuvre le calcul $\hat{\lambda}$

La variable  $Q$  est  $(I_n - H)$ , où  $I_n$  est la matrice identité de taille  $50 \times 50$ ,  $H = X(X'X)^{-1}X'$ . Le variable  $sig2$  égale à  $\frac{h_\lambda(Y)'(I_n - H)h_\lambda(Y)}{n}$  qui est exactement  $\hat{\sigma}^2$  où on a démontré dans l'équation (4). La fonction  $Lmle$  retourne le terme  $-\frac{n}{2} \log(\hat{\sigma}^2)$ .

### -Lmax en fonction de lambda



On voit bien que la fonction  $-L_{max}$  est une fonction quadratique convexe, au près de  $\lambda$  à 0.4,  $-L_{max}$  atteint au minimum.

## 3. Calcul $\hat{\lambda}$ et $\widehat{Var}(\hat{\lambda})$

Etant donné que  $Z = h_\lambda(Y)$  est la transformation de  $Y$ , par la définition donc  $\lambda \neq 0$ , pour la Méthodes Newton, on commence la itération à partir de 2. En utilisant la fonction optimisation *nlm*, on obtient la valeur estimée  $\hat{\lambda}$  est *resopt\$estimate* = 0.3817253. Comme démontré dans l'équation (16), la variance de  $\hat{\lambda}$  est l'inverse de hessian donc est 0.02948455. (Pour la minimisation  $-L_{max}$ , la matrice hessian pour  $\lambda$  est définie positive 33.91607)

## 4. Tests hypothèses

### Intervalle de confiance

Comme on a montré dans le premier section, l'intervalle de confiance pour  $\lambda$  est  $[\hat{\lambda} - q_{1-\alpha/2} \sqrt{\widehat{Var}(\hat{\lambda})}, \hat{\lambda} + q_{1-\alpha/2} \sqrt{\widehat{Var}(\hat{\lambda})}]$  au niveau asymptotiquement  $1 - \alpha$ . Ici on fixe  $\alpha$  à 0.05, donc  $q_{1-\alpha/2}$  est 1.959964 la quantile d'ordre  $1 - \alpha/2$  sous loi normale, donc on obtien l'IC  $[0.04517863, 0.718272]$ .

### Test de Wald

Das l'équation (18), on a montré  $W = \frac{(\hat{\lambda} - \lambda_0)^2}{\widehat{Var}(\hat{\lambda})} \rightarrow \chi^2(1)$ . Donc on veut tester  $H_0 : \lambda = \lambda_0$ , contre  $H_1 : \lambda \neq \lambda_0$  par la statistique de Wald pour les quatres cas ci-dessous:

- 1) Par la définition de  $h_\lambda(Y)$ , quand  $\lambda = 1$ , les données  $Y$  ne nécessitent pas de transformation, donc dans le test  $\lambda_0 = 1$ . On a  $W_{obs} > q_{1-\alpha}^{\chi^2}$  qui se trouve dans la région de rejet (unilatère à droite), et p-valeur  $0.0003173887 < \alpha$ , donc rejette  $H_0$  et accepte  $H_1$  avec risque d'erreur première espèce  $\alpha$ .
- 2) De même raison, quand  $\lambda = 0.5$ , la transformation à appliquer aux observations est en racine carrée. Donc dans le test on veut  $\lambda_0 = 0.5$ . On a  $W_{obs} < q_{1-\alpha}^{\chi^2}$  qui ne se trouve pas dans la région de rejet (unilatère à droite), et p-valeur  $0.4909476657 > \alpha$ , donc rejette  $H_1$  et accepte  $H_0$  avec risque d'erreur seconde espèce inconnue.
- 3) Quand  $\lambda_0 = 0.3$ , on a  $W_{obs} < q_{1-\alpha}^{\chi^2}$  qui ne se trouve pas dans la région de rejet, et p-valeur  $0.6341115306 > \alpha$ , donc rejette  $H_1$  et accepte  $H_0$  avec risque de seconde espèce inconnue.
- 4) Quand  $\lambda_0 = 0$ , on a  $W_{obs} > q_{1-\alpha}^{\chi^2}$  qui se trouve dans la région de rejet (unilatère à droite), et p-valeur  $0.0262112618 < \alpha$ , donc rejette  $H_0$  et accepte  $H_1$  avec risque d'erreur premier espèce  $\alpha = 5\%$ . (Note que dans  $h_\lambda(Y)$ ,  $\lambda \neq 0$ , donc on test 0.000001 ici)

## 5. Test de rapport de vraisemblance

```
## [1] "lambda= 1"
## [1] "TRV: 11.6612671670046"
## [1] "p_value: 0.000638148590272717"
## [1] "conserve H_0? FALSE"
## [1] "lambda= 0.5"
## [1] "TRV: 0.466380664196947"
## [1] "p_value: 0.494656961439759"
## [1] "conserve H_0? TRUE"
## [1] "lambda= 0.3"
## [1] "TRV: 0.228985473733189"
## [1] "p_value: 0.632277106839235"
## [1] "conserve H_0? TRUE"
## [1] "lambda= 1e-06"
## [1] "TRV: 5.14845895351846"
## [1] "p_value: 0.0232670132748466"
## [1] "conserve H_0? FALSE"
```

Par l'équation (21),  $TRV$  suit une loi  $\chi^2(1)$ , donc de même façon, on calcule les  $TRV$  observés et les p-valeurs. En comparant p-valeurs et  $\alpha$ , les conclusions obtenues sont mêmes que ce qui données par le test Wald dans la question précédente.

## 6. Vérification par fonction “powerTransform”

Etant donné que l’on a toujours des problèmes sur l’installation du package “car”, on l’exécute sur compilateur en ligne et voici dessous les résultats obtenus.

```
res <- powerTransform(Y~X, family="bcPower")  
summary(res)
```



Run (Ctrl-Enter)

Any scripts or data that you put into this service are public.

```
Loading required package: carData  
bcPower Transformation to Normality  
Est Power Rounded Pwr Wald Lwr Bnd Wald Upwr Bnd  
Y1 0.3817 0.5 0.0452 0.7183  
  
Likelihood ratio test that transformation parameter is equal to 0  
(log transformation)  
LRT df pval  
LR test, lambda = (0) 5.148486 1 0.023267  
  
Likelihood ratio test that no transformation is needed  
LRT df pval  
LR test, lambda = (1) 11.66127 1 0.00063815
```

FIGURE 1 – powerTransform

On peut voir que l’estimateur de la puissance transformation appliquée sur  $Y$  est la même valeur que l’on a obtenu par maximisation le vraisemblance  $L_{max}$  calculé par *nlm* dans la question 3. Puis l’intervalle de confiance  $[0.0452, 0.7183]$  est identique que l’on calcule dans question 4 au niveau asymptotique  $\alpha = 0.05$ . Comme on a montré dans question 5,  $TRV \sim \chi^2(1)$ , pour  $\lambda = 1$  et  $\lambda = 0$ , notre  $TRV$  observés et les p-valeurs sont les mêmes que les résultats donnés par fonction *powerTransform*.

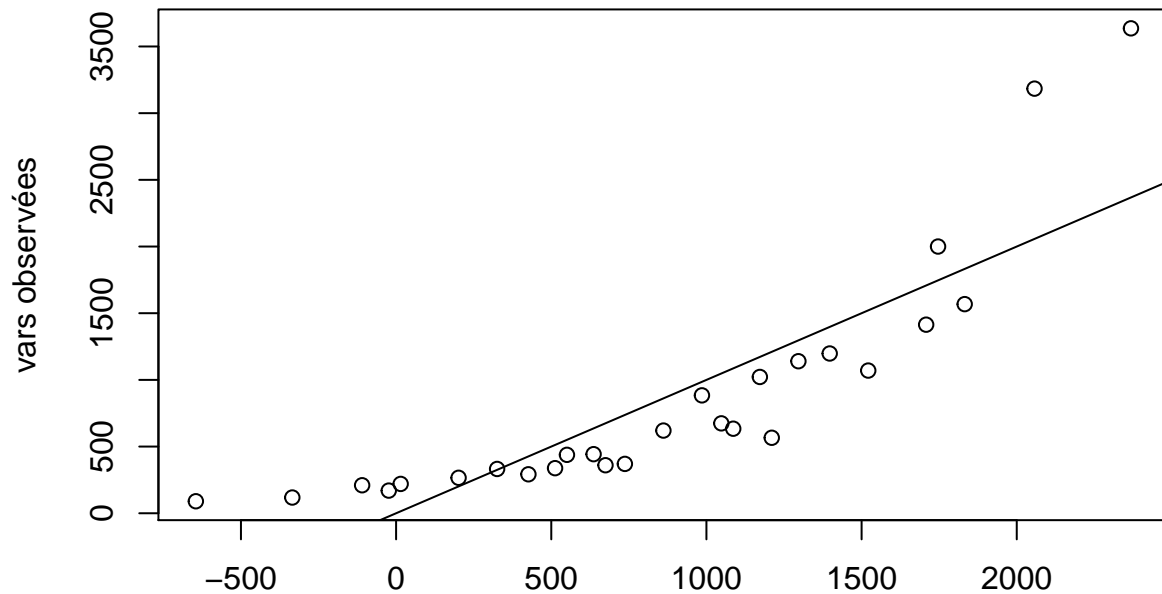
## 3 Cas pratique

### 1. Modèle regression linéaire multiple

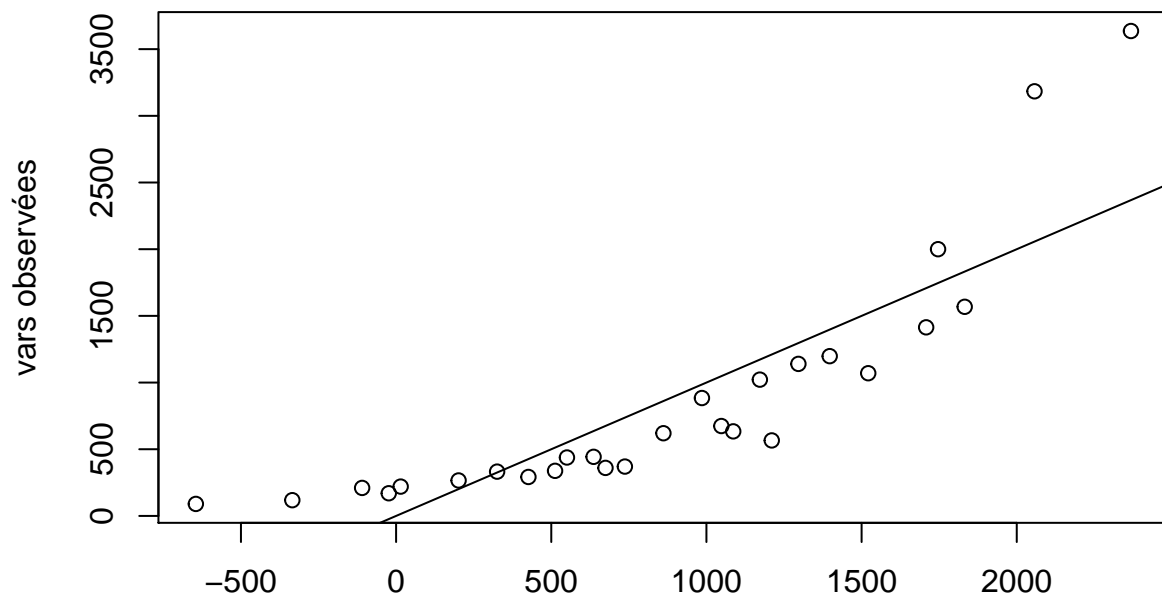
#### Analyse regression LM multiple

On définit le modèle linéaire multiple ( $M1$ )  $y = \mu + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \sigma \varepsilon$ ,  $\varepsilon \sim \mathcal{N}(0, I_n)$  i.i.d, on note que  $\theta = [\mu, \beta_1, \beta_2, \beta_3]$ .

**M1:  $y \sim x_1 + x_2 + x_3$**



**y ~ longueur + amplitude + chargement**



D'abord, on voit que le modèle pour les variables transformées  $M1 : y = \mu + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \sigma \varepsilon$  n'est pas bien ajusté, comme le montre visuellement la figure ci-dessus: certains les points s'allongent presque autour de la première bissectrice, mais les premiers données et les derniers données s'éloignent assez loin que la droite  $y = x$ , pour la plupart données, les ajustements un peu plus proches des valeurs observées mais avec certain des bruits assez évidents.

Et puis par la définition dans l'énoncé, on obtient les valeurs pour les variables non transformées (longueur, amplitude et chargement). De même façon, on applique aussi modèle linéaire multiple, mais les deux ajustements sont identiques. Bien sûr que les estimateurs de coefficient ( $\hat{\theta}$ ) changent, parce que les variables explicatives (i.e. la nouvelle matrice expérience  $X = [1, \text{longueur}, \text{amplitude}, \text{chargement}]$ ) changent. Mais les valeurs ajustées ne changent pas, donc la performance de modèle ne change pas non plus. Comme on peut vérifier par les données ci-dessous, les estimateurs changent mais le  $R^2$  et F statistique ne changent pas.

```
##
## Call:
## lm(formula = y ~ ., data = df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -644.5 -279.1 -150.2  199.5 1268.0
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   861.37      93.94   9.169 3.83e-09 ***
## x1             660.00     115.06   5.736 7.66e-06 ***
## x2            -535.83     115.06  -4.657 0.000109 ***
## x3            -310.83     115.06  -2.702 0.012734 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 488.1 on 23 degrees of freedom
## Multiple R-squared:  0.7291, Adjusted R-squared:  0.6937
## F-statistic: 20.63 on 3 and 23 DF,  p-value: 1.028e-06
##
## Call:
## lm(formula = df$y ~ longueur + amplitude + chargement)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -644.5 -279.1 -150.2  199.5 1268.0
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 4521.370   1621.721   2.788 0.010454 *
## longueur     13.200     2.301   5.736 7.66e-06 ***
## amplitude   -535.833    115.057  -4.657 0.000109 ***
## chargement   -62.167     23.011  -2.702 0.012734 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 488.1 on 23 degrees of freedom
## Multiple R-squared:  0.7291, Adjusted R-squared:  0.6937
## F-statistic: 20.63 on 3 and 23 DF,  p-value: 1.028e-06
```

A partir ici, on n'analyse que les valeurs pour  $y \sim x_1 + x_2 + x_3$ . Parce que pour  $y \sim \text{longueur} + \text{amplitude} + \text{chargement}$  c'est la même façon et les conclusions sont les mêmes que pour  $y \sim x_1 + x_2 + x_3$ .



### Analyse la significativité des variables

Ici on veut tester chaque composante  $\theta_i$  dans  $\theta$ .  $H_0 : \theta_i = 0$  contre  $H_1 : \theta_i \neq 0$ , on suppose que  $\alpha = 0.05$ .

- 1)  $\mu$ : p-valeur  $3.83e - 09 < \alpha$ , donc on rejette  $H_0$  et accepte  $H_1$ , c'est-à-dire on décide que le coefficient  $\mu$  est non nul avec risque de première espèce  $\alpha = 0.05$ , et l'intercept est utile dans  $M1$  donc elle est significative.
- 2)  $\beta_1$ : p-valeur  $7.66e - 06 < \alpha$ , donc on rejette  $H_0$  et accepte  $H_1$ , c'est-à-dire on décide que le coefficient  $\beta_1$  est non nul avec risque de première espèce  $\alpha = 0.05$ , et  $x_1$  est utile dans  $M1$  donc elle est significative.
- 3)  $\beta_2$ : p-valeur  $0.000109 < \alpha$ , donc on rejette  $H_0$  et accepte  $H_1$ , c'est-à-dire on décide que le coefficient  $\beta_2$  est non nul avec risque de première espèce  $\alpha = 0.05$ , et  $x_2$  est utile dans  $M1$  donc elle est significative.
- 4)  $\beta_3$ : p-valeur  $0.012734 < \alpha$ , donc on rejette  $H_0$  et accepte  $H_1$ , c'est-à-dire on décide que le coefficient  $\beta_3$  est non nul avec risque de première espèce  $\alpha = 0.05$ , et  $x_3$  est utile dans  $M1$  donc elle est significative.

En conclusion, on considère de garder tous les composantes variables dans  $M1$ .

### Analyse la significativité globale de la regression

Il s'agit de test Fisher sur un sous modèle linéaire du  $M1$ . On veut tester  $H_0 : \mu = \beta_1 = \beta_2 = \beta_3 = 0$ , contre  $H_1$  : l'un des paramètres n'est pas nul.

On observe que le F-statistique observé est 20.63, et son p-valeur est  $1.028e - 06 < \alpha = 0.05$ , donc on rejette  $H_0$  et accepte  $H_1$ , avec risque d'erreur de première espèce  $\alpha$ . On peut conclure que au moins l'un des quatres coefficient (composante paramètre) n'est pas nul, au moins l'un des quatres variables (intercept et  $x_1, x_2, x_3$ ) est significative.

### Analyse R squared

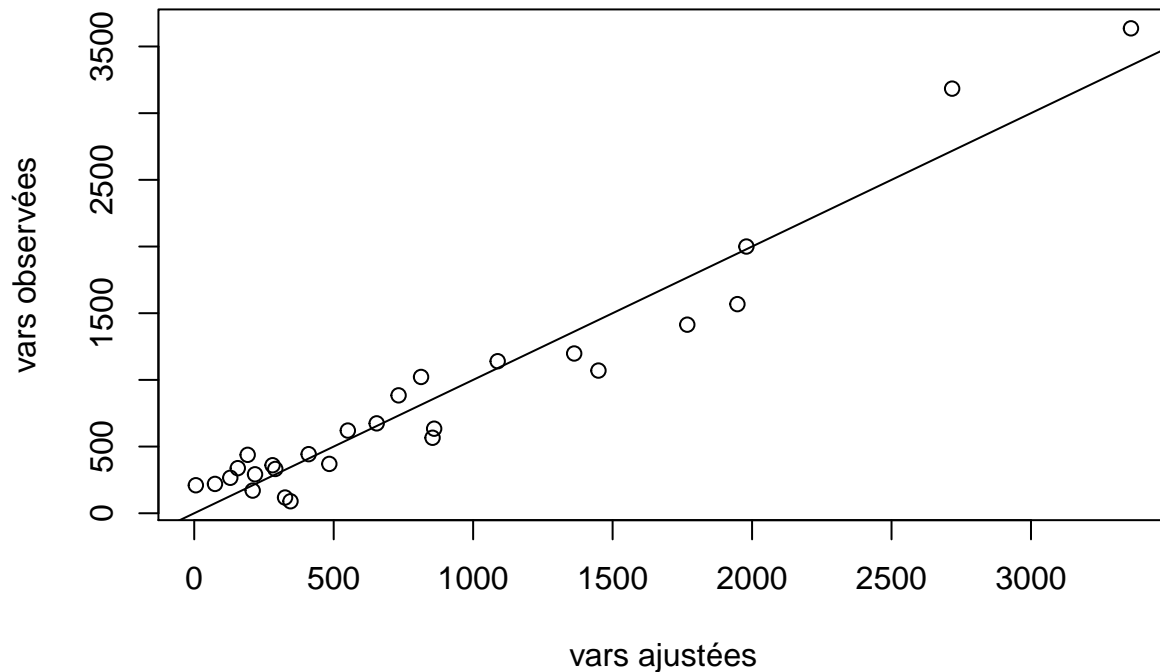
$R^2$  s'interprète comme la part de variance expliquée par les régresseurs supplémentaires. Donc 0.7291 signifie que la population de 72.91% de données peuvent être expliquées par notre modèle  $M1$ , la mauvaise qualité de l'ajustement et aussi énormément des brutes. En conclusion,  $M1$  le LM multiple n'est pas un modèle idéal.

## 2. Modèle regression linéair d'ordre 2

### Etude redression LM d'ordre 2

Maintenant, on voudrais modéliser un modèle regression linéair d'ordre 2  $M2 : y = \mu + \sum_i \beta_i x_i + \sum_i \beta_{ii} x_i^2 + \sum_{i < j} \beta_{ij} x_i x_j + \varepsilon$ ,  $\varepsilon \sim \mathcal{N}(0, I_n)$  i.i.d. Voici dessous le graphe de l'ajustement:

## M2



On voit bien visuellement que l'ajustement du modèle *M2* est beaucoup plus amélioré que *M1*! Les points s'allongent autour de la première bissectrice, les ajustements beaucoup proches des valeurs observées. En plus, il n'existe plus les valeurs négatives.

Maintenant, on analyse rapidement les valeurs données par R de même façon que l'on a fait dans la question précédente.

```
##
## Call:
## lm(formula = y ~ df$x1 + df$x2 + df$x3 + I(df$x1^2) + I(df$x2^2) +
##      I(df$x3^2) + I(df$x1 * df$x2) + I(df$x1 * df$x3) + I(df$x2 *
##      df$x3), data = df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -379.48 -185.95   41.41  148.48  466.69
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      550.70     138.44   3.978 0.000973 ***
## df$x1           660.00      64.09  10.299 1.00e-08 ***
## df$x2          -535.83      64.09  -8.361 1.99e-07 ***
## df$x3          -310.83      64.09  -4.850 0.000150 ***
## I(df$x1^2)       238.56     111.00   2.149 0.046317 *
## I(df$x2^2)       275.72     111.00   2.484 0.023712 *
## I(df$x3^2)      -48.28     111.00  -0.435 0.669081
## I(df$x1 * df$x2) -456.50      78.49  -5.816 2.06e-05 ***
## I(df$x1 * df$x3) -235.67      78.49  -3.003 0.008011 **
## I(df$x2 * df$x3)  142.92      78.49   1.821 0.086278 .
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 271.9 on 17 degrees of freedom
## Multiple R-squared:  0.9379, Adjusted R-squared:  0.905
## F-statistic: 28.51 on 9 and 17 DF,  p-value: 1.564e-08
```

### Analyse la significativité des variables

Ici on veut tester chaque composante  $\theta_i$  dans  $\theta$ .  $H_0 : \theta_i = 0$  contre  $H_1 : \theta_i \neq 0$ , on suppose que  $\alpha = 0.05$ . Etant donné que y'a dix composantes paramètres dans l'estimateur  $\theta$ , on n'analyse pas un par un variables ici.

En vérifiant chaque p-valeur correspondantes les variables, on peut conclure que seulement la variable  $x_3^2$  et la variable  $x_2x_3$  ne sont pas significatives. Parce que ses p-valeurs sont supérieures à  $\alpha$  donc on regarde  $H_0$  (avec risque de seconde espèce inconnue) et décide que ses coefficients sont nuls, donc les deux variables ne sont pas utiles dans  $M2$  (i.e. ne sont pas significatives). Pour la reste, ils sont tous significatives. En conclusion, on considère de ne pas garder  $x_3^2$  et  $x_2x_3$  dans  $M2$ .

### Analyse la significativité globale de la regression

On veut tester  $H_0$  : tous les composantes de l'estimateur sont égaux à 0, contre  $H_1$  : l'un des paramètres n'est pas nul.

On observe que le F-statistique observé est 28.51, et son p-valeur est  $1.564e - 08 < \alpha = 0.05$ , donc on rejette  $H_0$  et accepte  $H_1$ , avec risque d'erreur de première espèce  $\alpha$ . On peut conclure que au moins l'un des dix coefficients (composantes paramètre) n'est pas nul, au moins l'un des dix variables est significative.

### Analyse R squared

$R^2$  s'interprète comme la part de variance expliquée par les régresseurs supplémentaires. Donc 0.9379 signifie que la population de 93.79% de données sont expliquées par notre modèle  $M2$ , ce qui implique un très bonne qualité de l'ajustement. En conclusion,  $M2$  le LM multiple d'ordre 2 est un modèle assez performante.

### Test modèle

On veut construire un test  $M1$  contre  $M2$  en utilisant "anova" fonction. En théorie, le F statistique est le même que le test de significativité globale de la regression. On voit bien le  $F_{obs} = 9.5227$ , p-valeur égale à  $0.000115 < \alpha = 0.05$ , donc on conclut que les variables d'interaction ajoutées sont significatives.

## 3. Test de variance

### Conclusion