

Etude et mise en oeuvre de la transformation de Box-Cox sur des données simulées et expérimentales

BLANC Hugo
MAMFOUMBI Jean-Michel

November 2020

Introduction

L'objectif du projet est de mettre en oeuvre la transformation de Box-Cox dans l'étude de jeu de données simulées ou expérimentales pour mettre en évidence ses propriétés. Dans un premier temps nous allons étudier théoriquement la transformation de Box-Cox qui possède des propriétés intéressantes (stabilisation de la variance, capacité à gaussianiser les données lorsqu'elles sont fortement dissymétrique) : pour cela nous tenterons de justifier son utilisation avant de calculer sa log-vraisemblance et nous déterminerons l'estimateur du maximum de vraisemblance de son paramètre $\hat{\lambda}$ pour construire des tests bilatéraux. Dans un second temps nous testerons cette transformation sur un jeu de données simulées et nous procéderons aux tests proposés dans la première parties. Nous tenterons aussi de simuler le niveau du test de rapport de vraisemblance et de calculer sa puissance. Enfin, nous passerons à l'étude du cas pratique pour lequel nous comparerons plusieurs modèles différents avant d'utiliser à nouveau en oeuvre la transformation de Box-Cox et de voir si ses propriétés (variance stable, données gaussiennes) fonctionnent là aussi.

1 La transformation de Box-Cox

1) On a définit la transformation de Box-Cox suivante, pour $\lambda \in \mathbb{R}$:

$$\tilde{h}_\lambda : y \mapsto \begin{cases} \frac{y^\lambda - 1}{\lambda} & \lambda \neq 0 \\ \log(\lambda) & \lambda = 0 \end{cases} \text{ définie pour } y > 0$$

A $\lambda > 0$ fixé, cette fonction est à valeurs dans $] -\frac{1}{\lambda}; +\infty[$, tandis que pour $\lambda = 0$ la fonction est à valeurs dans \mathbb{R} et pour $\lambda < 0$ cette fonction est à valeurs dans $] -\infty; -\frac{1}{\lambda}[$.

Aussi, si on suppose que pour tout i :

$$\tilde{h}_\lambda(Y_i) = Z_i = x_i\theta + \epsilon_i, \epsilon_i \sim_{iid} \mathcal{N}(0, \sigma^2)$$

Alors dans ce cas, $\tilde{h}_\lambda(Y_i) \sim \mathcal{N}(x_i\theta, \sigma^2)$, aussi théoriquement les réalisations de la variable aléatoire $\tilde{h}_\lambda(Y_i)$ est à valeurs dans \mathbb{R} ce qui n'est pas compatible avec le domaine d'arrivée de \tilde{h}_λ lorsque $\lambda \neq 0$.

Aussi, la transformation semble compatible seulement lorsque $\lambda = 0$.

2) Afin de déterminer la vraisemblance des paramètres $(\lambda, \theta, \sigma^2)$ à l'observation de $Y = (Y_1, \dots, Y_n)$, on peut déjà remarquer que d'après le modèle d'observation, on a :

$$h_\lambda(Y_i) = Z_i = x_i\theta + \epsilon_i, \epsilon_i \sim_{iid} \mathcal{N}(0, \sigma^2)$$

où ϵ_i est le bruit iid gaussien et x_i est le vecteur ligne des conditions d'expérience.

Aussi, on en déduit qu'en notant $Z = (Z_1, \dots, Z_n)$, on a :

$$Z = X\theta + \epsilon, \epsilon \sim \mathcal{N}(0, \sigma^2 I_n)$$

où ϵ est le bruit gaussien et X est la matrice du plan d'expérience (correspondant à la concaténation des vecteurs lignes x_i).

Ainsi :

$$Z \sim \mathcal{N}(X\theta, \sigma^2 I_n)$$

Et Z admet pour densité f_Z par rapport à la mesure de Lebesgue sur \mathbb{R}^n , la fonction :

$$\forall z \in \mathbb{R}^n, f_Z(z) = \left(\frac{1}{2\pi\sigma^2} \right)^{\frac{n}{2}} \exp - \frac{(z - X\theta)'(z - X\theta)}{2\sigma^2}$$

Or, on a aussi que $h_\lambda(Y) = Z$, où $h_\lambda(Y)$ est le vecteur des composantes $h_\lambda(Y_i)$, de plus par formule de changement de variables dans les intégrales multiples en prenant h_λ , C^1 -difféomorphisme, généralisée pour n variables (cf. $h_\lambda(Y) = Z$, où $h_\lambda(Y)$ est le vecteur des composantes $h_\lambda(Y_i)$), on obtient que:

$$\forall y = (y_1, \dots, y_n) \in \mathbb{R}^n, f_Y(y) = |\det J_{h_\lambda}(y)| (f_Z \circ h_\lambda)(y) \quad (1)$$

où f_Y est la densité de Y par rapport à la mesure de Lebesgue sur \mathbb{R}^n et J_{h_λ} est la jacobienne de h_λ . Aussi, par la généralisation précédente de h_λ , on a que la i -ème fonction coordonnée vaut:

$$h_{\lambda,i} := (y_1, \dots, y_n) \mapsto h_\lambda(y_i) \quad (*)$$

De plus,

$$\forall i, j \in \{1, \dots, n\}^2, \text{ on a } (J_{h_\lambda})_{i,j} = \frac{\partial h_{\lambda,i}}{\partial y_j}$$

Aussi, pour $i \neq j$:

$$\frac{\partial h_{\lambda,i}}{\partial y_j} = 0 \text{ (d'après } (*), h_{\lambda,i} \text{ est indépendant de } y_j)$$

Et pour $i = j$:

$$\frac{\partial h_{\lambda,i}}{\partial y_i}(y) = |y_i|^{\lambda-1}$$

Ainsi, on en déduit que:

$$\det J_{h_\lambda}(y) = \prod_{i=1}^n |y_i|^{\lambda-1}$$

Et la vraisemblance des paramètres $(\lambda, \theta, \sigma^2)$ à l'observation de $Y = (Y_1, \dots, Y_n)$ s'écrit donc :

$$\begin{aligned} L(\lambda, \theta, \sigma^2; y) &= f_Y(y) \\ &= |\det J_{h_\lambda}(y)| (f_Z \circ h_\lambda)(y) \\ &= \prod_{i=1}^n |y_i|^{\lambda-1} \left(\frac{1}{2\pi\sigma^2} \right)^{\frac{n}{2}} \exp - \frac{(h_\lambda(y) - X\theta)'(h_\lambda(y) - X\theta)}{2\sigma^2} \end{aligned}$$

Aussi, en posant: $J(\lambda; Y) = \prod_{i=1}^n |Y_i|^{\lambda-1}$, on a finalement:

$$L(\lambda, \theta, \sigma^2; Y) = J(\lambda; Y) \left(\frac{1}{2\pi\sigma^2} \right)^{\frac{n}{2}} \exp - \frac{(h_\lambda(Y) - X\theta)'(h_\lambda(Y) - X\theta)}{2\sigma^2}$$

item Déterminons l'estimateur du maximum de vraisemblance (EMV) à λ fixé, $\hat{\theta}(\lambda)$ et $\hat{\sigma}(\lambda)$ en fonction de X et λ : On cherche les valeurs de θ et σ pour lesquels la log-vraisemblance est maximale (car elle a les mêmes variations que la vraisemblance). On calcule la log-vraisemblance puis on la dérive par rapport à θ

$$\log L = \log(J(\lambda; Y)) - \frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} (h_\lambda(y) - X\theta)'(h_\lambda(y) - X\theta)$$

D'où

$$\frac{\partial \log L}{\partial \theta} = 0$$

$$\implies \frac{1}{2\sigma^2} (2X'h_\lambda(Y) - 2X'X\theta) = 0$$

$$\implies X'h_\lambda(Y) - X'X\theta = 0$$

$$\implies X'X\theta = X'h_\lambda(Y)$$

Dans un modèle supposé identifiable on en déduit :

$$\theta = (X'X)^{-1}X'h_\lambda(Y)$$

Puis on calcule la dérivée seconde :

$$\frac{\partial^2 \log L(\theta, \sigma^2)}{\partial \theta^2} = \frac{-X'X}{2\sigma^2} \leq 0$$

Donc on a bien que la log-vraisemblance est maximale en $\theta = (X'X)^{-1}X'h_\lambda(Y)$ donc

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

On détermine ensuite $\hat{\sigma}^2$:

$$\frac{\partial \log L}{\partial \sigma^2} = 0$$

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

$$\implies \sigma^2 = \frac{\|h_\lambda(Y) - X\theta\|^2}{n}$$

$$\frac{\partial^2 \log L(\theta, \sigma^2)}{\partial \sigma^2} = \frac{n}{2\sigma^4} - \frac{\|h_\lambda(Y) - X\theta\|^2}{\sigma^6}$$

$$\frac{\partial^2 \log L(\theta, \sigma^2)}{\partial \sigma^2} = \frac{n}{2\sigma^6} \left(\sigma^2 - 2 \frac{\|h_\lambda(Y) - X\theta\|^2}{n} \right) \leq 0 \text{ en } \sigma^2 = \frac{\|h_\lambda(Y) - X\theta\|^2}{n}$$

Donc $\hat{\sigma}^2 = \frac{\|h_\lambda(Y) - X\hat{\theta}\|^2}{n}$

On en conclut que notre EMV est donné par $(\hat{\theta}, \hat{\sigma}^2) = \left((X'X)^{-1}X'h_\lambda(Y), \frac{\|h_\lambda(Y) - X\hat{\theta}\|^2}{n} \right)$

Calculons $L_{max}(\lambda)$:

$$L_{max}(\lambda) = \log(J(\lambda; Y)) - \frac{n}{2} \log(2\pi\hat{\sigma}^2) - \frac{\|h_\lambda(Y) - X\hat{\theta}\|^2}{2\hat{\sigma}^2}$$

$$\implies L_{max}(\lambda) = (\lambda - 1) \sum_{i=1}^n \log |y_i| - \frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\hat{\sigma}^2) - \frac{n}{2}$$

$$\implies L_{max}(\lambda) = (\lambda - 1) \sum_{i=1}^n \log |y_i| - \frac{n}{2} \log(\hat{\sigma}^2) + a(n) \text{ avec } a(n) = -\frac{n}{2} - \frac{n}{2} \log(2\pi)$$

Pour déterminer $\hat{\lambda}$, on cherche le maximum de L_{max} , pour cela on dérive L_{max} par rapport à λ et on cherche un point d'annulation:

$$\frac{\partial L_{max}}{\partial \lambda}(\lambda) = \sum_{i=1}^n \log |y_i| - \frac{n}{2} \frac{d\hat{\sigma}^2}{d\lambda} \frac{1}{\hat{\sigma}^2}$$

Aussi, $\hat{\lambda}$ est solution de:

$$\sum_{i=1}^n \log |y_i| - \frac{n}{2} \frac{d\hat{\sigma}^2}{d\lambda} \frac{1}{\hat{\sigma}^2} = 0$$

ou plus simplement de:

$$\frac{\partial L_{max}}{\partial \lambda}(\hat{\lambda}) = 0$$

et une dérivée seconde négative au voisinage de $\hat{\lambda}$.

En théorie, pour rechercher $\hat{\lambda}$, on peut effectuer une recherche de maximum sur la fonction L_{max} .

Déterminons désormais si l'estimateur du maximum de vraisemblance est gaussien à distance finie:

Par hypothèse $h_\lambda(Y)$ est de loi gaussienne $h_\lambda(Y) \sim \mathcal{N}(X\theta, \sigma^2 I_n)$. Comme $\hat{\theta} = (X'X)^{-1}X'h_\lambda(Y)$ est une fonction linéaire de $h_\lambda(Y)$, la loi de $\hat{\theta}$ est gaussienne d'espérance $\mathbb{E}(\hat{\theta}) = \theta$ et de variance $\text{var}(\hat{\theta}) = \sigma^2(X'X)^{-1}$

On a que $\hat{\sigma}^2$ suit une loi du χ_2 (par somme de carré de loi gaussienne) donc ne suit pas une loi gaussienne.

Enfin, pour en revenir à $\hat{\lambda}$, tout d'abord par théorème sous les conditions de régularité qui si elles sont vérifiées ici, on a que cet estimateur du maximum de vraisemblance est asymptotiquement gaussien. De plus, nous avons vu que pour de grands échantillons (par exemple pour $n \geq 30$, où n est la taille de l'échantillon) la loi de l'estimateur converge rapidement et l'on peut approximer que $\hat{\lambda}$ est gaussien à distance finie.

- 3) Nous sommes sous les conditions de convergence spécifiées précédemment, ainsi l'estimateur du maximum de vraisemblance $\hat{\lambda}$ est asymptotiquement gaussien. Ainsi, $\hat{\lambda}$ tend à devenir gaussien et d'après le cours on a à distance finie (pour $n \leq 30$ par exemple, ce qui sera le cas par la suite):

$$\hat{\lambda} \sim_{\text{appr}} \mathcal{N}(\lambda, I_n^{-1}(\lambda)) \text{ et } \text{var}(\hat{\lambda}) \approx I_n^{-1}(\lambda) = \frac{1}{n} I_1^{-1}(\lambda)$$

De plus, par la loi forte des grands nombres:

$$\begin{aligned} \frac{1}{n} \frac{\partial^2 L_{\max}}{\partial \lambda^2}(\hat{\lambda}) &= \frac{1}{n} \frac{\partial^2 L(\lambda, \hat{\theta}, \hat{\sigma}^2; y)}{\partial \lambda^2}(\hat{\lambda}) \\ &= \frac{1}{n} \sum_{i=1}^n \frac{\partial^2 \log f(y_i)}{\partial \lambda^2}(\hat{\lambda}) \\ &\xrightarrow{n \rightarrow +\infty} \mathbb{E}\left(\frac{\partial^2 \log f(Y_1)}{\partial \lambda^2}(\hat{\lambda})\right) = -I_1(\hat{\lambda}) \text{ (par la loi forte des grands nombres)} \end{aligned}$$

Aussi:

$$-n \left(\frac{\partial^2 L_{\max}}{\partial \lambda^2}(\hat{\lambda}) \right)^{-1} \xrightarrow{n \rightarrow +\infty} I_1^{-1}(\hat{\lambda})$$

Ainsi, pour ce qui est de l'estimateur de la variance de $(\hat{\lambda})$, on peut prendre:

$$\hat{V}_n = - \left(\frac{\partial^2 L_{\max}}{\partial \lambda^2}(\hat{\lambda}) \right)^{-1}$$

En effet, pour n grand (pour un échantillon au moins supérieur à 30 observations d'après le cours 2 diaporama 19):

$$\begin{aligned} \hat{V}_n &= - \frac{n}{n} \left(\frac{\partial^2 L_{\max}}{\partial \lambda^2}(\hat{\lambda}) \right)^{-1} \\ &\approx \frac{1}{n} I_1^{-1}(\lambda) \\ &= I_n^{-1}(\lambda) \\ &\approx \text{var}(\hat{\lambda}) \end{aligned}$$

d'après le cours et l'étude des estimateurs de maximums de vraisemblance, de plus il y a la convergence asymptotiquement vers $\text{var}(\hat{\lambda})$.

De plus, pour la fonction, $\frac{\partial^2 L_{\max}}{\partial \lambda^2}$, on note H_n son hessien (qui est donc un scalaire égale à la dérivée seconde car la fonction n'a qu'une seule variable λ).

Ainsi, on se dote de l'estimateur de $\text{var}(\hat{\lambda})$ suivant: $\hat{V}_n = -(H_n(\hat{\lambda}))^{-1}$

D'après le théorème de Slutsky et la normalité asymptotique, on a:

$$\hat{V}_n^{-1/2}(\hat{\lambda} - \lambda) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$$

Aussi on peut construire un intervalle de confiance de λ , en effet en notant $q_{1-\alpha/2}^*$ le quantile d'ordre $1 - \frac{\alpha}{2}$ d'une loi gaussienne centrée réduite, on obtient l'intervalle de confiance de niveau $1 - \alpha$ suivant:

$$IC_n(\lambda) = [\hat{\lambda} - q_{1-\alpha/2}^* \hat{V}_n^{1/2}, \hat{\lambda} + q_{1-\alpha/2}^* \hat{V}_n^{1/2}]$$

On souhaite tester:

$$(H_0) : \lambda = \lambda_0 \text{ contre } (H_1) : \lambda \neq \lambda_0$$

Sous l'hypothèse de test (H_0) , on peut construire la statistique de notre test de Wald, on a tout d'abord:

$$\hat{V}_n^{-1/2}(\hat{\lambda} - \lambda_0) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$$

car $\hat{\lambda}$ est l'estimateur du maximum de vraisemblance de λ et \hat{V}_n est un estimateur de la variance de $\hat{\lambda}$. Ainsi, on a donc (comme carré de lois normales):

$$W = \hat{V}_n^{-1}(\hat{\lambda} - \lambda_0)^2 \xrightarrow{\mathcal{L}} \chi(1) \text{ (car on a pris } A = 1, \text{ matrice scalaire de rang } r = 1)$$

La région de rejet du test s'en déduit: $\mathcal{R} = \{W > q_{1-\alpha}^{\chi^2(1)}\}$ de niveau approximativement égale à α .

4) En notant $LRV = 2(L_{max}(\hat{\lambda}) - L_{max}(\lambda))$, on a finalement:

$$\begin{aligned} LRV &= 2(L_{max}(\hat{\lambda}) - L_{max}(\lambda)) \\ &= -2(\log L(\lambda, \hat{\theta}, \hat{\sigma}^2; Y) - \log L(\hat{\lambda}, \hat{\theta}, \hat{\sigma}^2; Y)) \\ &= -2\log\left(\frac{L(\lambda, \hat{\theta}, \hat{\sigma}^2; Y)}{L(\hat{\lambda}, \hat{\theta}, \hat{\sigma}^2; Y)}\right) \\ &= -2\log(RV) \xrightarrow{\mathcal{L}} \chi^2(1) \end{aligned}$$

,en notant RV le rapport de vraisemblance généralisé dans ce modèle et en utilisant le théorème asymptotique du rapport de vraisemblance.

On se place sous les hypothèses de test:

$$(H_0) : \lambda = \lambda_0 \text{ contre } (H_1) : \lambda \neq \lambda_0$$

On choisit la statistique de test: $LRV = 2(L_{max}(\hat{\lambda}) - L_{max}(\lambda_0)) \xrightarrow{\mathcal{L}} \chi^2(1)$ sous l'hypothèse (H0).

La région de rejet du test s'en déduit: $\mathcal{R} = \{LRV > q_{1-\alpha}^{\chi^2(1)}\}$ de niveau asymptotiquement égale à α , où $q_{1-\alpha}^{\chi^2(1)}$ est le quantile d'ordre $1 - \alpha$ d'une loi du $\chi^2(1)$.

2 Test de la méthode sur des données simulées

1) On génère le germe du générateur aléatoire et on génère les x_i ainsi que les conditions initiales du problème et les termes d'erreurs:

```
set.seed(999)
help("set.seed")

#initialisation des données
n=50
lambda=0.3
a=5
b=1
var=2
x_i=rnorm(n)
eps_i=rnorm(n,0,sqrt(var))
```

Vérifions que la condition de convergence est vérifié:

En notant X la matrice du plan d'expérience, on a:

$$\frac{X'X}{n} = \begin{pmatrix} 1 & \frac{\sum_{i=1}^n x_i}{n} \\ \frac{\sum_{i=1}^n x_i}{n} & \frac{\sum_{i=1}^n x_i^2}{n} \end{pmatrix}$$

où en notant que (x_1, \dots, x_n) sont issue d'un n-échantillon gaussien $\mathcal{N}(0, 1)$: (X_1, \dots, X_n) . Aussi par la loi forte des grands nombres:

$$\mathbb{P}\left(\lim_{n \rightarrow +\infty} \bar{X}_n = \mathbb{E}(X)\right) = 1$$

En appliquant cette propriété à notre n-échantillon gaussien $\mathcal{N}(0, 1)$, on déduit que:

$$\lim_{n \rightarrow +\infty} \frac{\sum_{i=1}^n x_i}{n} = 0 \text{ (espérance du n-échantillon gaussien)}$$

$$\lim_{n \rightarrow +\infty} \frac{\sum_{i=1}^n x_i^2}{n} = 1 \text{ (variance du n-échantillon gaussien)}$$

Ainsi, on a:

$$\lim_{n \rightarrow +\infty} \frac{X'X}{n} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Aussi, la matrice identité est définie positive donc la **condition de convergence est vérifiée dans cette exemple**.

On peut d'ailleurs vérifier numériquement cette convergence pour des grandes valeurs de n , pour $n = 10^5$ on trouve:

```

                                x_i
      1.0000000000 -0.0005703712
x_i -0.0005703712  1.0041733074

```

On calcule ensuite l'échantillon Z grâce à la formule: $Z_i = a + bx_i$, puis l'échantillon Y en inversant la relation de

Bickel et Docksum tel que: $Y_i = \text{sgn}(0.3Z_i + 1)|0.3Z_i + 1|^{1/0.3}$:

```

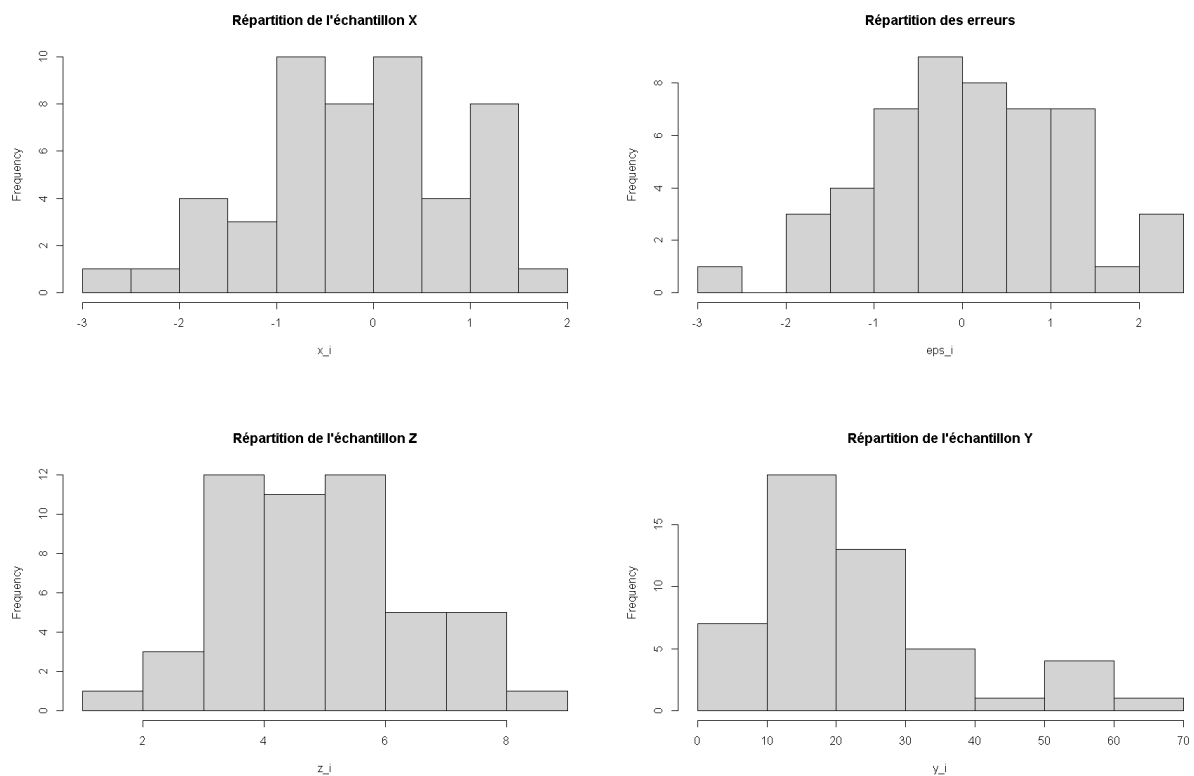
#On calcule l'échantillon z, d'après les informations de l'énoncé
z_i=a+b*x_i+eps_i

#Puis l'échantillon y, grâce à la relation h_lambda, formule de transformation de
#Bickel et Docksum:
y_i=sign(0.3*z_i+1)*abs(0.3*z_i+1)^(1/0.3)

#vérification:
z_i=sign(y_i)*((abs(y_i)^lambda)-1)/lambda
# [1] 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00
# [9] 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 1.665335e-16 0.000000e+00
# [17] -8.881784e-16 -8.881784e-16 0.000000e+00 0.000000e+00 -8.881784e-16 0.000000e+00 -8.881784e-16 0.000000e+00
# [25] 0.000000e+00 0.000000e+00 -4.440892e-16 -8.881784e-16 4.440892e-16 0.000000e+00 0.000000e+00 0.000000e+00
# [33] -8.881784e-16 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 4.440892e-16
# [41] 0.000000e+00 8.881784e-16 0.000000e+00 0.000000e+00 0.000000e+00 -4.440892e-16 0.000000e+00 4.440892e-16
# [49] 0.000000e+00 8.881784e-16

```

Ainsi, nous obtenons une répartition des différents échantillons suivantes:



Une simple analyse graphique semble permettre d'estimer que l'échantillon Y ne suit pas une répartition gaussienne. Estimons désormais les régressions linéaires:

La régression linéaire de z en fonction de x , donne:

```
Call:
lm(formula = z_i ~ x_i)

Residuals:
    Min       1Q   Median       3Q      Max
-2.43920 -0.66126  0.02716  0.72144  2.19484

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   5.0717     0.1571  32.283 < 2e-16 ***
x_i           1.1764     0.1567   7.507 1.23e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.095 on 48 degrees of freedom
Multiple R-squared:  0.54,    Adjusted R-squared:  0.5305
F-statistic: 56.36 on 1 and 48 DF, p-value: 1.233e-09
```

Aussi, on obtient une estimation du paramètre $a = 5.0717$ avec une erreur type de 0.1571 et du paramètre $b = 1.1764$ avec une erreur type de 0.1567, ce qui est proche des valeurs réelles ($a = 5$ et $b = 1$). Le test de Fisher de significativité globale de la régression donne une p-value de $1.233e^{-9} < 0.05$, donc l'on rejette l'hypothèse de non significativité globale des coefficients de la régression.

De plus, les statistiques du test de Student de significativité des coefficients et leurs p-value respectives ($< 2e^{-16}$ et $1.23e^{-9}$ toutes les deux inférieurs à 0.05) permettent de déduire que l'on rejette l'hypothèse de non-significativité pour les deux coefficients respectivement, aussi il faut les conserver tous les deux dans notre modèle.

On peut tracer cette régression linéaire:



La régression linéaire de y en fonction de x, donne:

```
Call:
lm(formula = y_i ~ x_i)

Residuals:
    Min       1Q   Median       3Q      Max
-20.060  -7.388  -1.948   6.782  28.002

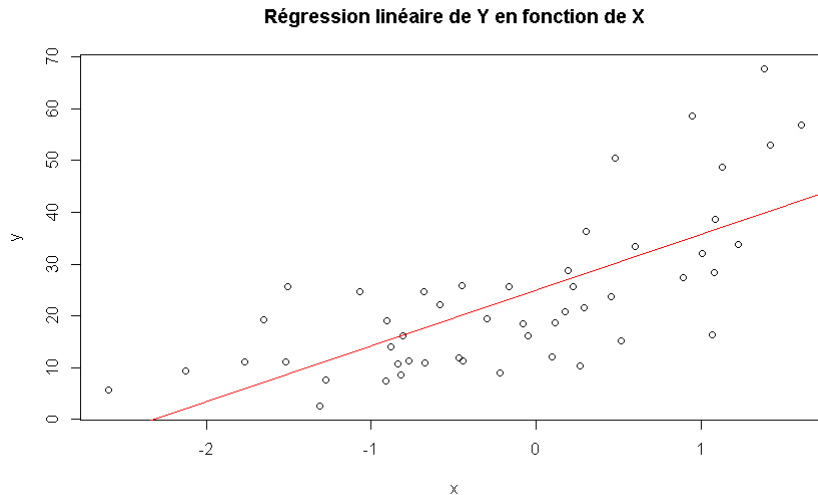
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  24.943     1.517  16.445 < 2e-16 ***
x_i          10.737     1.513   7.096 5.23e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10.58 on 48 degrees of freedom
Multiple R-squared:  0.512,    Adjusted R-squared:  0.5018
F-statistic: 50.36 on 1 and 48 DF, p-value: 5.233e-09
```

Aussi, on obtient une estimation du paramètre associé à x qui vaut 24.943 avec une erreur type de 1.517 et du paramètre de l'intercept qui vaut 10.737 avec une erreur type de 1.513. Le test de Fisher de significativité globale de la régression donne une p-value de $5.233e^{-9} < 0.05$, donc l'on rejette l'hypothèse de non significativité globale des coefficients de la régression.

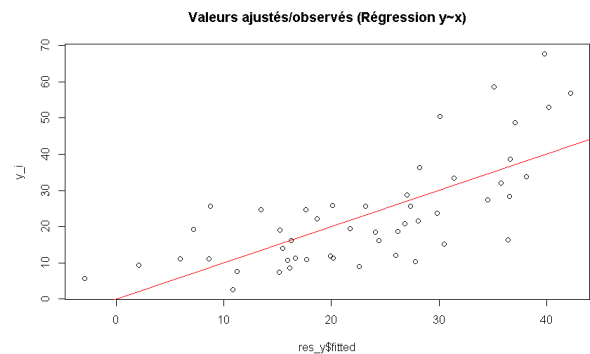
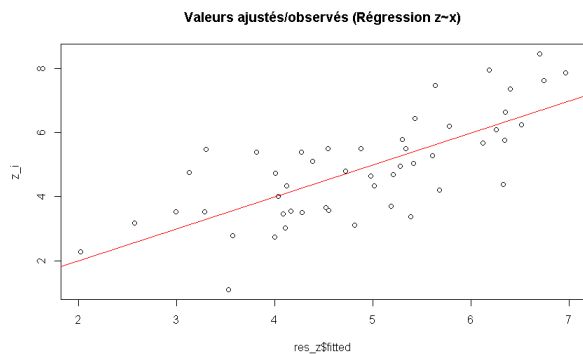
De plus, les statistiques du test de Student de significativité des coefficients et leurs p-value respectives ($< 2e^{-16}$ et $5.23e^{-9}$ toutes les deux inférieurs à 0.05) permettent de déduire que l'on rejette l'hypothèse de non-significativité

pour les deux coefficients respectivement, aussi il faut les conserver tous les deux dans notre modèle.
On peut tracer cette régression linéaire:



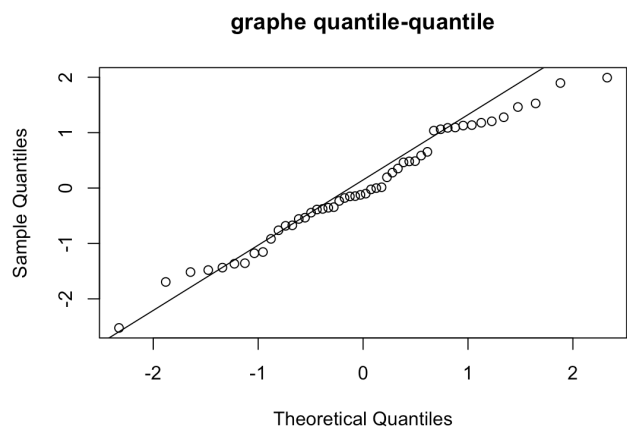
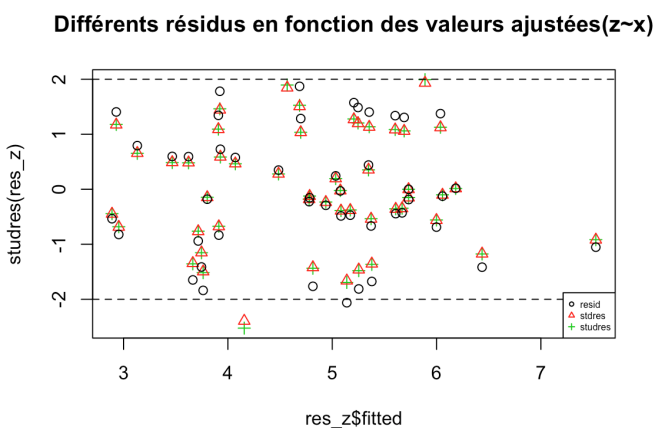
Passons à l'étude des résidus:

Pour le graphe des valeurs exactes par rapport aux valeurs estimées, on obtient:



On peut voir que les points s'alignent assez bien suivant la première bissectrice, dans la régression simple de z fonction de x , ce qui est un gage visuel que l'ajustement est assez bon.

Tandis que dans le cas cas la régression simple de y fonction de x (image de droite), les points sont légèrement plus dispersés par rapport à la première bissectrice donc l'ajustement est moins bon.

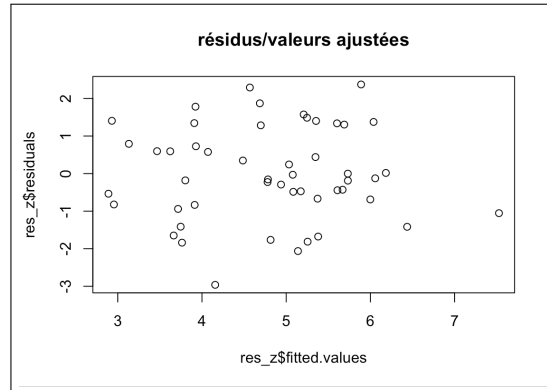


Passons à l'étude des résidus. Dans le cas de la régression linéaire avec z , on remarque que les résidus studentisés (triangle rouge) sont globalement compris entre $[-2, 2]$ pour notre échantillon de 50 observation : la **règle empirique d'appartenance** est bien respectée et on n'observe aucune tendance ou dépendance significative des résidus studentisés. Pour la normalité de l'erreur le graphe quantile quantile, les points s'alignent plutôt bien donc a priori il n'y a pas de raison de remettre en question la gaussianité des résidus. Pour être plus confiant quant à cet hypothèse on peut aussi réaliser un test de Shapiro-Wilk.

Shapiro-Wilk normality test

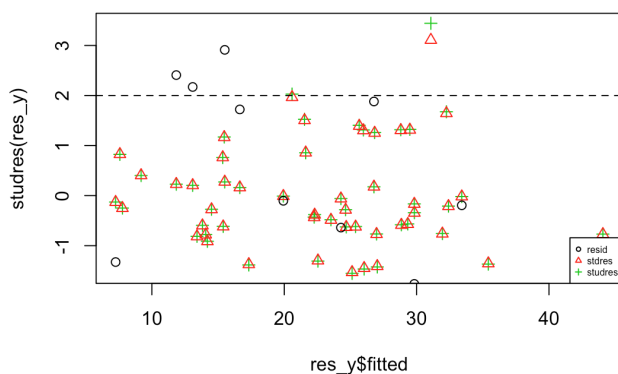
```
data: studres(res_z)
W = 0.98116, p-value = 0.6019
```

La p-value est supérieure à 0.05 on conserve l'hypothèse de gaussienne avec un **risque de seconde espèce inconnu**. On peut donc estimer que l'hypothèse gaussienne est correcte.

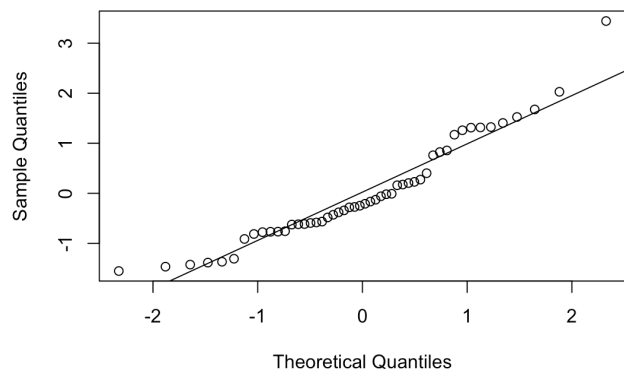


Les résidus sont dispersés de manière uniforme tout le long de l'axe des valeurs ajustées, on peut donc supposer que l'hypothèse d'**homoscédasticité** est correcte.

Différents résidus en fonction des valeurs ajustées(y~x)



graphe quantile-quantile

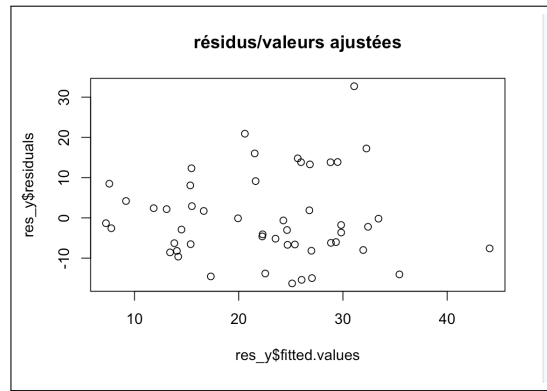


On remarque pour la régression linéaire avec y : la **règle empirique d'appartenance** est respectée pour la très grande majorité des résidus studentisés, il y a une seule valeur extrême qui semble en dehors (vers la valeur ajustée qui vaut environ 33). Toutefois, les points sont moins alignées avec notre droite de référence que dans le graphe quantile quantile précédent et il y a aussi une possible tendance dans la répartition des résidus studentisés. On peut supposer que l'hypothèse gaussienne est mauvaise, ce que l'on va essayer de confirmer avec un test de Shapiro-Wilk:

Shapiro-Wilk normality test

```
data: studres(res_y)
W = 0.93397, p-value = 0.007836
```

On obtient une p-value inférieure à 0.05 donc on rejette l'hypothèse gaussienne avec un **risque de première espèce 0.05**.



On constate une augmentation de la dispersion des résidus le long de l'axe x = valeurs ajustées. On retrouve une forme de cône caractéristique de **l'hétéroscédasticité**, la variance ne semble pas constante.

2) On calcule la matrice X du plan d'expérience grâce au logiciel R:

```
> X = as.matrix(cbind(1,x_1))
> X
      x_1
[1,] 1 -0.281740159
[2,] 1 -1.312559630
[3,] 1  0.795183982
[4,] 1  0.270070494
[5,] 1 -0.277306422
[6,] 1 -0.566023739
[7,] 1 -1.878658258
[8,] 1 -1.266791144
[9,] 1 -0.967749684
[10,] 1 -1.121009362
[11,] 1  1.325463712
[12,] 1  0.133977391
[13,] 1  0.938749446
[14,] 1  0.172538101
[15,] 1  0.957650447
[16,] 1 -1.362686245
[17,] 1  0.068335133
[18,] 1  0.100657648
[19,] 1  0.901344754
[20,] 1 -2.074337106
[21,] 1 -1.228563304
[22,] 1  0.643044319
[23,] 1 -0.359762914
[24,] 1  0.294035586
[25,] 1 -1.125268488
[26,] 1  0.642265650
[27,] 1 -1.106737580
[28,] 1 -0.884840392
[29,] 1 -1.554095144
[30,] 1 -0.126678989
[31,] 1  2.382664156
[32,] 1  0.601276101
[33,] 1  0.179361267
[34,] 1  1.080531484
[35,] 1 -0.246812106
[36,] 1 -2.113736990
[37,] 1 -0.370527471
[38,] 1  0.522867793
[39,] 1  0.517805536
[40,] 1 -1.402510873
[41,] 1 -0.485636726
[42,] 1  0.008498139
[43,] 1 -1.282113287
[44,] 1 -1.11578841
[45,] 1  0.300665411
[46,] 1  0.276478845
[47,] 1 -2.050877659
[48,] 1  0.014190211
[49,] 1  0.582266478
[50,] 1 -0.034726387
```

Par la suite, nous allons interpréter le code suivant:

```
Q = diag(1,n) - X%*%solve(t(X)%*%X)%*%t(X)
Lmle = function(Z){
  n = length(Z)
  sig2 = ( t(Z)%*%Q%*%Z )/n
  -n/2*log(sig2)
}
```

On peut remarquer que la matrice Q du code correspond en fait à: $I_n - H$ où H correspond à ce que nous avons appelé "Hat Matrix".

Aussi, pour un vecteur Z donné, on a:

$$\begin{aligned}
sig_2 &= \frac{{}^tZ(I_n - H)Z}{n} \\
&= \frac{{}^tZ(I_n - H)Z}{n} \\
&= \frac{{}^tZ(I_n - H)^2Z}{n} && \text{(car H est une matrice de projection donc I-H aussi et ainsi } (I - H)^2 = I - H\text{)} \\
&= \frac{({}^tZ - {}^tZH)(Z - HZ)}{n} \\
&= \frac{{}^t(Z - HZ)(Z - HZ)}{n} && \text{(car H est une matrice symétrique donc } {}^tH = H\text{)} \\
&= \frac{\|Z - \hat{Z}\|^2}{n} && \text{(par définition du produit scalaire et de la matrice H)} \\
&= \hat{\sigma}^2 && \text{(pour une observation de variable réponse Z et par définition de } \hat{\sigma}^2\text{)}
\end{aligned}$$

Aussi, on en déduit que $Lmle$ permet de calculer $-\frac{n}{2} \log(\hat{\sigma}^2)$, pour une observation de variable réponse Z .

On programme ensuite la fonction h_lambda et on en déduit la fonction $Lmin(lambda, Y)$:

```

h_lambda=function(lambda,y){
  (sign(y)*abs(y)^(lambda-1))/lambda
}

lmin= function(lambda,Y){
  n=length(Y)
  lmax=(lambda-1)*sum(log(abs(Y)))+Lmle(h_lambda(lambda,y))- (n/2)*(1+log(2*pi))
  -lmax
}

```

Par la suite, on peut donc tracer la fonction $-L_{max}(\lambda)$ pour $\lambda \in [0; 2]$ grâce à la commande:

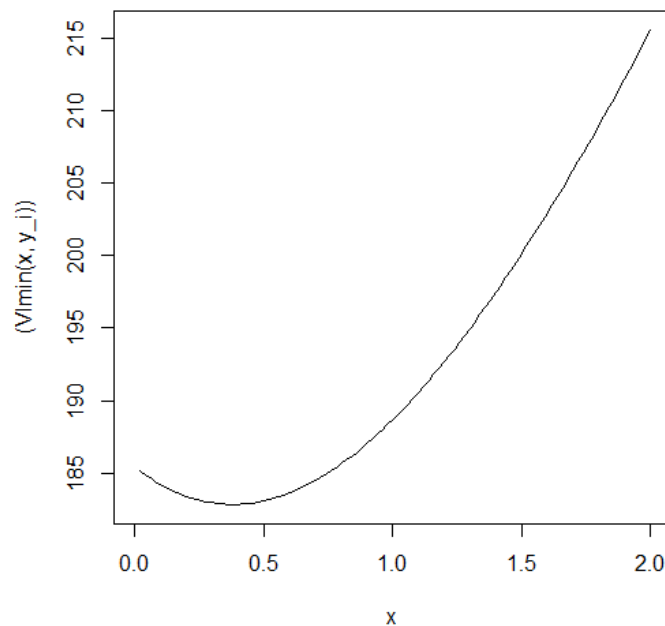
```

vlmin = vectorize(lmin,"lambda")
curve(expr=(vlmin(x,y_i)),from=0, to= 2)

```

et on obtient finalement:

Tracé de $-L_{max}(\lambda)$ (avec $\lambda = x$)



Aussi, sachant que $\hat{\lambda}$ maximise $L_{max}(\lambda)$, on peut en déduire que $\hat{\lambda}$ minimise $L_{min}(\lambda)$ et donc on a:

$$\hat{\lambda} \approx 0.4$$

- 3) Nous allons désormais utiliser la fonction `nlm` afin d'optimiser $-L_{max}(\lambda)$, après un appel à la notice R on obtient que cette fonction permet de minimiser la fonction prise en argument:

Non-Linear Minimization

Description

This function carries out a minimization of the function f using a Newton-type algorithm. See the references for details.

Usage

```
nlm(f, p, ..., hessian = FALSE, typsize = rep(1, length(p)),
    fscale = 1, print.level = 0, ndigit = 12, gradtol = 1e-6,
    stepmax = max(1000 * sqrt(sum((p/typsize)^2)), 1000),
    steptol = 1e-6, iterlim = 100, check.analyticals = TRUE)
```

Ceci correspond exactement à ce que nous cherchons ici pour obtenir $\hat{\lambda}$, afin que l'algorithme converge le plus rapidement possible, nous l'initialiserons avec une valeur proche de la valeur minimisant $-L_{max}(\lambda)$, ainsi nous prendrons $p = 0.4$.

Nous obtenons ainsi: $\hat{\lambda} = 0.38$, de plus le hessien de $-L_{max}(\lambda)$ correspond à l'opposée du hessien de $L_{max}(\lambda)$, aussi le hessien obtenu est directement l'opposé de l'estimation de la variance de $\hat{\lambda}$ d'après la question 1 de l'exercice 1 et nous obtenons donc:

$$\hat{V}_n(\hat{\lambda}) = 0.03$$

- 4) Maintenant que nous avons obtenu ces différentes données nous allons pouvoir déterminer un intervalle de confiance de niveau $1-\alpha$ pour λ . En effet, nous avons obtenu à l'exercice 1 (question 4) que sous les conditions de convergence qui sont ici vérifiées (cf question 1), nous avons:

$$IC_n(\lambda) = [\hat{\lambda} - q_{1-\alpha/2}^* \hat{V}_n^{1/2}, \hat{\lambda} + q_{1-\alpha/2}^* \hat{V}_n^{1/2}]$$

Aussi l'application ici, nous permet d'obtenir:

$$IC_n(\lambda) = [0.38 - 0.17q_{1-\alpha/2}^*, 0.38 + 0.17q_{1-\alpha/2}^*]$$

Si l'on applique cela pour $\alpha = 0.05$, on obtient:

$$IC_n(\lambda) = [0.045, 0.72]$$

Nous allons désormais tester plusieurs hypothèses grâce à une statistique de Wald:

- Sous (H_0) : les données de Y ne nécessitent pas de transformation (soit $\lambda = 1$)
On va donc tester $A\lambda = 1$ contre $A\lambda \neq 1$ avec $A = (1)$ (matrice de rang $r = 1$)
Or on sait que les conditions de régularité sont vérifiées, on a donc l'estimateur du maximum de vraisemblance asymptotiquement normal et on peut donc introduire la statistique de Wald :

$$W = [(\hat{V}_n)^{-1/2}(\hat{\lambda} - 1)]^2 \text{ qui suit une loi du } \chi^2(1) \text{ par Slutsky et passage au carré de loi normale}$$

On rejette l'hypothèse si $W_{obs} > q_{1-\alpha}^{\chi^2(1)} = 3.84$

$W_{obs} = 12.96$ donc on **rejette l'hypothèse avec un risque de première espèce 0.05**

On peut vérifier cela en calculant la p-value, on trouve p-value=0.00032 < 0.05 donc cela confirme qu'on peut rejeter l'hypothèse (H_0) avec un risque de première espèce 0.05.

- Sous (H_0) les données de Y nécessitent une transformation en racine carré soit $\lambda = 0.5$
On va donc tester $A\lambda = 0.5$ contre $A\lambda \neq 0.5$ avec $A = (1)$
On effectue le même raisonnement que précédemment en adaptant notre statistique de test

$W = [(\hat{V}_n)^{-1/2}(\hat{\lambda} - 0.5)]^2$ qui suit une loi du $\chi^2(1)$ par Slutsky et passage au carré de loi normale

On rejette l'hypothèse si $W_{obs} > q_{1-\alpha}^{\chi^2(1)} = 3.84$

$W_{obs} = 0.47$ donc on **conserve avec un risque de seconde espèce inconnu**

On peut vérifier cela en calculant la p-value, on trouve p-value=0.49 > 0.05 donc cela confirme qu'on peut conserver l'hypothèse (H_0) avec un risque de seconde espèce inconnu.

- Sous (H_0) : $\lambda = 0.3$
On va donc tester $A\lambda = 0.3$ contre $A\lambda \neq 0.3$ avec $A = (1)$
On effectue le même raisonnement que précédemment en adaptant notre statistique de test

$W = [(\hat{V}_n)^{-1/2}(\hat{\lambda} - 0.3)]^2$ qui suit une loi du $\chi^2(1)$ par Slutsky et passage au carré de loi normale

On rejette l'hypothèse si $W_{obs} > q_{1-\alpha}^{\chi^2(1)} = 3.84$

$W_{obs} = 0.23$ donc on **conserve l'hypothèse avec un risque de seconde espèce inconnu**

On peut vérifier cela en calculant la p-value, on trouve p-value=0.63 > 0.05 donc cela confirme qu'on peut conserver l'hypothèse (H_0) avec un risque de seconde espèce inconnu.

- Sous (H_0) : la transformation à appliquer aux observations est \tilde{h}_0 (on supposera qu'il est possible d'utiliser les résultats de la section 1 dans ce cas, sans chercher de preuve théorique).
On suppose qu'il est possible d'appliquer les résultats de la question 1
On va donc tester $A\theta = 0$ contre $A\theta \neq 0$ avec $A = (1)$
On effectue le même raisonnement que précédemment en adaptant notre statistique de test

$W = [(\hat{V}_n)^{-1/2}(\hat{\lambda})]^2$ qui suit une loi du $\chi^2(1)$ par Slutsky et passage au carré de loi normale

On rejette l'hypothèse si $W_{obs} > q_{1-\alpha}^{\chi^2(1)} = 3.84$

$W_{obs} = 4.94$ donc on **rejette l'hypothèse avec un risque de première espèce 0.05**

On peut vérifier cela en calculant la p-value, on trouve p-value=0.026 < 0.05 donc cela confirme qu'on peut rejeter l'hypothèse (H_0) avec un risque de première espèce 0.05.

```
> lambda_0
[1] 1.0 0.5 0.3 0.0
> w_vect
[1] 12.9647946 0.4744339 0.2265361 4.9420964
> w_vect > qchisq(0.95,1)
[1] TRUE FALSE FALSE TRUE
```

5) Testons les mêmes hypothèses avec les rapports de vraisemblance. On introduit la statistique

$$T_{RV} = -2 \log \frac{L(\lambda, \theta \sigma^2; Y)}{L(\hat{\lambda}, \theta \sigma^2; Y)} \sim \chi^2(1)$$

D'où:

On rejette l'hypothèse si $T_{obs} > q_{1-\alpha}^{\chi^2(1)} = 3.84$

- Sous (H_0) : les données de Y ne nécessitent pas de transformation (soit $\lambda = 1$)
Nous obtenons $T_{obs} = 11.66$ donc on rejette l'hypothèse (H_0) **avec un risque de première espèce 0.05**
On peut vérifier cela en calculant la p-value, on trouve p-value=0.00063815 < 0.05 donc cela confirme qu'on peut rejeter l'hypothèse (H_0) avec un risque de première espèce de 0.05.
- Sous (H_0) les données de Y nécessitent une transformation en racine carré (soit $\lambda = 0.5$)
Nous obtenons $T_{obs} = 0.47$ donc on conserve l'hypothèse (H_0) **avec un risque de seconde espèce inconnu**
On peut vérifier cela en calculant la p-value, on trouve p-value=0.49466 > 0.05 donc cela confirme qu'on peut conserver l'hypothèse (H_0) avec un risque de seconde espèce inconnu.
- Sous (H_0) : $\lambda = 0.3$
Nous obtenons $T_{obs} = 0.23$ donc on conserve l'hypothèse (H_0) **avec un risque de seconde espèce inconnu**
On peut vérifier cela en calculant la p-value, on trouve p-value=0.63 > 0.05 donc cela confirme qu'on peut conserver l'hypothèse (H_0) avec un risque de seconde espèce inconnu.

```
> lambda_0
[1] 1.0 0.5 0.3
> LRV_vect
[1] 11.6612672 0.4663807 0.2289855
> LRV_vect>qchisq(0.95,1)
[1] TRUE FALSE FALSE
```

- (H_0) : la transformation à appliquer aux observations est \tilde{h}_0 (on supposera qu'il est possible d'utiliser les résultats de la section 1 dans ce cas, sans chercher de preuve théorique).

Après avoir modifié les fonctions afin de les adapter à ce test particulier nécessitant \tilde{h}_0 et en prenant (H_0) : $\lambda = 0$, nous obtenons: $T_{obs} = 5.15$ donc on rejette l'hypothèse (H_0) **avec un risque de première espèce 0.05**

On peut vérifier cela en calculant la p-value, on trouve $p\text{-value}=0.023 < 0.05$ donc cela confirme qu'on peut rejeter l'hypothèse (H_0) avec un risque de première espèce de 0.05.

```
> LRV_0
      [,1]
[1,] 5.148486
> LRV_0>qchisq(0.95,1)
      [,1]
[1,] TRUE
```

- 6) Après une recherche sur le fonctionnement de la fonction **powerTransform** du package **car**, on obtient:

powerTransform (car)
R Documentation

Finding Univariate or Multivariate Power Transformations

Description

`powerTransform` uses the maximum likelihood-like approach of Box and Cox (1964) to select a transformation of a univariate or multivariate response for normality, linearity and/or constant variance. Available families of transformations are the default Box-Cox power family and two additional families that are modifications of the Box-Cox family that allow for (a few) negative responses. The summary method automatically computes two or three likelihood ratio type tests concerning the transformation powers.

Aussi, cette fonction permet de trouver le coefficient λ permettant d'obtenir la transformation \tilde{h}_λ la plus adapté pour un modèle linéaire donné. Dans notre cas, comme les observations Y sont positives ont a les transformations de Box-Cox et de Bickel et Docksum qui correspondent pour $\lambda > 0$, aussi nous devrions retrouver le même coefficient que précédemment mais aussi le même intervalle de confiance et des résultats des tests de rapport de vraisemblances similaires:

```
> rep=powerTransform(y_i~x_i)
> summary(rep)
bcPower Transformation to Normality
  Est Power Rounded Pwr wald Lwr Bnd wald upr 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
```

Effectivement, nous retrouvons les mêmes résultats, l'estimation de λ vaut 0.38, tandis que l'intervalle de confiance est similaire à celui que nous avons obtenu à la question 4 ($IC_n(\lambda) = [0.045, 0.72]$). Tandis que les statistiques prennent les mêmes valeurs que précédemment et nous confirme que l'on rejette l'hypothèse nulle (car $0.23 < 0.05$ et $0.0006 < 0.05$) dans chacun des deux cas illustrés séparément.

D'ailleurs, on peut aussi vérifier la cohérence avec les autres hypothèses que nous avons testé (non affiché par défaut) en utilisant la fonction `testTransform`:

```

> testTransform(rep,0.5)
              LRT df    pval
LR test, lambda = (0.5) 0.4663807 1 0.49466
> testTransform(rep,0.3)
              LRT df    pval
LR test, lambda = (0.3) 0.2289855 1 0.63228

```

On retrouve les mêmes statistiques de test que précédemment et les p-value nous confirment que l'on peut conserver les hypothèses (H_0) dans les deux cas $\lambda = 0.3$ et $\lambda = 0.5$ avec des risque de seconde espèce inconnu.

- 7) Afin de simuler le niveau du test du rapport de vraisemblance de (H_0): $\lambda = 0.3$ contre (H_1): $\lambda \neq 0.3$, on réalise le code suivant:

```

n=50
lambda=0.3
a=5
b=1
var=2
compteur=0
for (i in 1:100){
  x_test=rnorm(n)
  eps_test=rnorm(n,0,sqrt(var))
  z_test=a+b*x_test+eps_test
  y_test=sign(lambda*z_test+1)*abs(lambda*z_test+1)^(1/lambda)
  reptest=powerTransform(y_test~x_test)
  if (testTransform(reptest,0.3)$pval<0.05)
  {
    compteur=compteur+1
  }
}
compteur
Nb_moyen_rejets=compteur/100

```

Et on observe le nombre moyen de rejet du test suivant:

```

> compteur
[1] 4
> Nb_moyen_rejets
[1] 0.04

```

On observe un niveau simulé égal à 0.04, le résultat semble cohérent car les tests ont été effectué avec un risque $\alpha = 0.05$ de première espèce. Pour simuler la puissance du même test avec en alternative $\lambda = \lambda_1$, il suffit de générer $B=100$ n-échantillons sous (H_1) donc en les créant en prenant $\lambda = \lambda_1$ et on observe le nombre moyens de rejets du test en testant toujours pour test $\lambda = 0.3$, la puissance correspondra à la fréquence de rejets du test. Ainsi, on obtient pour $\lambda_1 = 0.1$:

```

#Pour la puissance
n=50
lambda=0.1
a=5
b=1
var=2
compteur=0
for (i in 1:100){
  x_test=rnorm(n)
  eps_test=rnorm(n,0,sqrt(var))
  z_test=a+b*x_test+eps_test
  y_test=sign(lambda*z_test+1)*abs(lambda*z_test+1)^(1/lambda)
  reptest=powerTransform(y_test~x_test)
  if (testTransform(reptest,0.3)$pval<0.05)
  {
    compteur=compteur+1
  }
}
compteur
Nb_moyen_rejets=compteur/100

```

Et on observe le nombre moyen de rejet du test suivant:

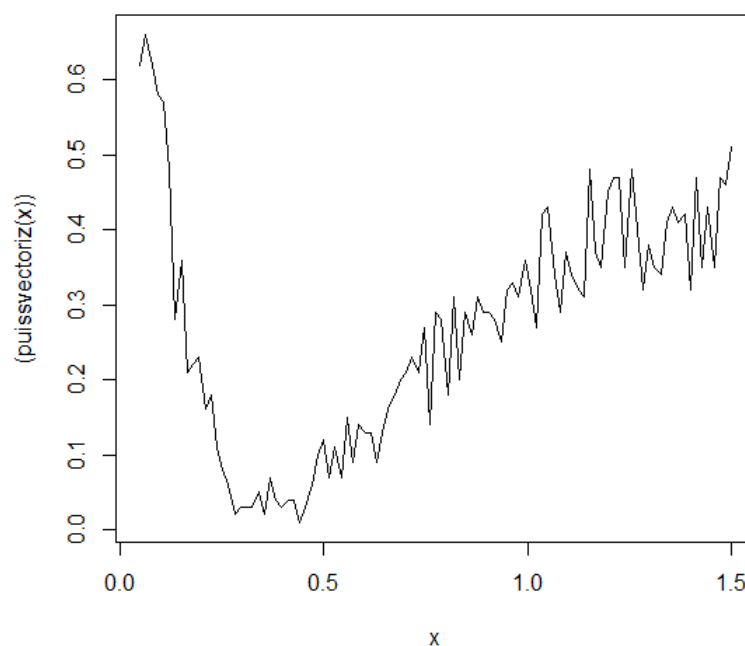
```
> compteur  
[1] 51  
> Nb_moyen_rejets  
[1] 0.51
```

Afin de tracer l'évolution de la puissance pour λ_1 compris entre 0.05 et 1.5, on transforme la démarche précédente en fonction et on la vectorise afin de pouvoir enfin tracer la courbe:

```
#On transforme cela en fonction:  
puissance=function(lambda)  
{  
  n=50  
  a=5  
  b=1  
  var=2  
  compteur=0  
  for (i in 1:100){  
    x_test=rnorm(n)  
    eps_test=rnorm(n,0,sqrt(var))  
    z_test=a+b*x_test+eps_test  
    y_test=sign(lambda*z_test+1)*abs(lambda*z_test+1)^(1/lambda)  
    reptest=powerTransform(y_test~x_test)  
    if (testTransform(reptest,0.3)$pval<0.05)  
    {  
      compteur=compteur+1  
    }  
  }  
  compteur  
  Nb_moyen_rejets=compteur/100  
  Nb_moyen_rejets  
}  
puissvectoriz = vectorize(puissance,"lambda")  
  
curve(expr=(puissvectoriz(x)),from=0.05, to= 1.5)
```

et on obtient la représentation de la puissance suivante:

Evolution de la puissance du test étudié ((H0) : $\lambda = 0.3$ contre (H1) : $\lambda = x$)



On peut remarquer que plus on s'éloigne de $\lambda = 0.3$ (hypothèse (H_0)), plus la puissance augmente et qu'elle est minimale au voisinage de $\lambda = 0.3$. De plus, le test semble sans biais (car la puissance est supérieur au niveau α).

- 8) **Bonus:** On reprend la question 7 pour le test de (H_0) : $\lambda = 0$, on modifie donc la construction des nos échantillons, on les génère donc en prenant $\lambda = 0$ et en modifiant la construction de y , en prenant cette fois l'inverse de la transformation de Cox-Box (pour $\lambda = 0$, donc l'exponentielle) .

```
#Bonus
n=50
lambda=0
a=5
b=1
var=2
compteur=0
for (i in 1:100){
  x_test=rnorm(n)
  eps_test=rnorm(n,0,sqrt(var))
  z_test=a+b*x_test+eps_test
  y_test=exp(z_test)
  reptest=powerTransform(y_test~x_test)
  if (testTransform(reptest,0)$pval<0.05)
  {
    compteur=compteur+1
  }
}
compteur
Nb_moyen_rejets=compteur/100
```

Aussi, après exécution on obtient:

```
> compteur
[1] 5
> Nb_moyen_rejets
[1] 0.05
```

On observe un niveau simulé égal à 0.05, le résultat semble cohérent car les tests ont été effectué avec un risque $\alpha = 0.05$ de première espèce.

Pour simuler la puissance du même test avec en alternative $\lambda = \lambda_1$, il suffit de générer B=100 n-échantillons sous (H_1) donc en les créant en prenant $\lambda = \lambda_1$ et on observe le nombre moyens de rejets du test en testant pour toujours pour le test $\lambda = 0$, la puissance correspondra à la fréquence de rejets du test.

Ainsi, on obtient pour $\lambda_1 = 0.1$:

```
#Pour la puissance
n=50
lambda=0.1
a=5
b=1
var=2
compteur=0
for (i in 1:100){
  x_test=rnorm(n)
  eps_test=rnorm(n,0,sqrt(var))
  z_test=a+b*x_test+eps_test
  y_test=sign(lambda*z_test+1)*abs(lambda*z_test+1)^(1/lambda)
  reptest=powerTransform(y_test~x_test)
  if (testTransform(reptest,0)$pval<0.05)
  {
    compteur=compteur+1
  }
}
compteur
Nb_moyen_rejets=compteur/100
```

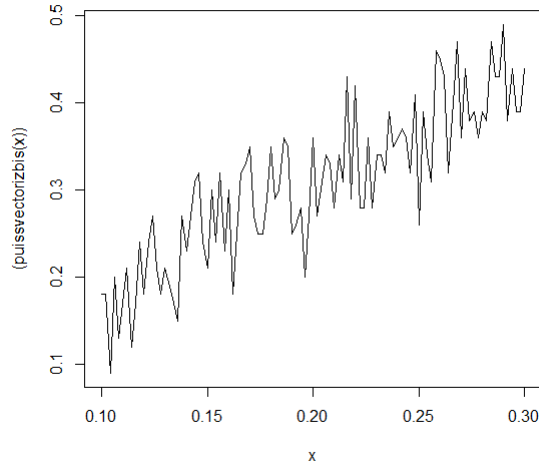
Et on observe le nombre moyen de rejet du test suivant:

```
> compteur
[1] 16
> Nb_moyen_rejets
[1] 0.16
```

Aussi, la puissance du test est basse donc le risque de seconde espèce $\beta = 0.84$ est élevé ce qui est mauvais signe. Le test semble assez mauvais.

Afin de tracer l'évolution de la puissance pour λ_1 compris entre 0.1 et 0.3, on transforme la démarche précédente en fonction comme à la question précédente et on la vectorise afin de pouvoir enfin tracer la courbe, on obtient ainsi:

Evolution de la puissance du test étudié ((H0) : $\lambda = 0$ contre (H1) : $\lambda = x$)



On observe que l'évolution de la puissance augmente globalement lorsque λ_1 s'éloigne de 0, ce qui semble cohérent. De plus, le test semble sans biais (car la puissance est supérieur au niveau α), il faudrait vérifier en observant les résultats pour λ_1 compris entre 0 et 0.1.

3 Cas pratique

Pour cet exercice, on dispose de 27 couples (x_i, Y_i) pour lesquels Y_i représente la i -ème observation de notre caractéristique d'intérêt dans l'exercice, c'est-à-dire le nombre de cycles de rupture d'un fil peigné en fonction de trois variables explicatives $x_i = (x_{i1}, x_{i2}, x_{i3})$

- 1) On définit le modèle de régression avec effets additifs (M1) suivant :

$$Y_i = \alpha + x_{i1}\theta_1 + x_{i2}\theta_2 + x_{i3}\theta_3 + \epsilon_i, \text{ où } \epsilon_i \sim \mathcal{N}(0, \sigma^2)$$

Comme indiqué ci-dessus, on se place dans le cadre d'un bruit gaussien pour connaître les lois des estimateurs à distance finie. On pourra vérifier cette hypothèse sur R à partir de différents tests dont notamment le test de Shapiro-Wilk.

Après avoir défini le modèle, on l'ajuste sur R à partir de la fonction **lm** à partir de l'instruction suivante :

```
reg_multi<- lm(cycle$y~a1 + a2 + a3, data = df)
```

Ensuite, on ajuste un autre modèle (bruit gaussien) dans lequel on applique la transformation aux variables explicatives :

$$\begin{cases} f_1(x_1) = \frac{x_1 - 300}{50} \\ f_2(x_2) = x_2 - 9 \\ f_3(x_3) = \frac{x_3 - 45}{5} \end{cases} \quad (1)$$

On crée dans R un **dataframe** qu'on a nommé df qui contient les données initiales obtenues à partir de la transformation inverse appliquée aux variables (on nomme les vecteurs a_1, a_2, a_3):

```

a1 = 50*cycle$x1 + 300
a2 = cycle$x2 + 9
a3 = 5*cycle$x3 + 45
df = data.frame(cycle$y,a1,a2,a3)

```

On obtient alors les résultats suivants :

<pre> Call: lm(formula = cycle\$y ~ a1 + a2 + a3, 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) 4521.370 1621.721 2.788 0.010454 * a1 13.200 2.301 5.736 7.66e-06 *** a2 -535.833 115.057 -4.657 0.000109 *** a3 -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 </pre>	<pre> Call: lm(formula = y ~ x1 + x2 + x3, data = cycle) 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 </pre>
---	--

(a) Données de la régression linéaire de M1 avant transformation (b) Données de la régression linéaire de M1 après transformation

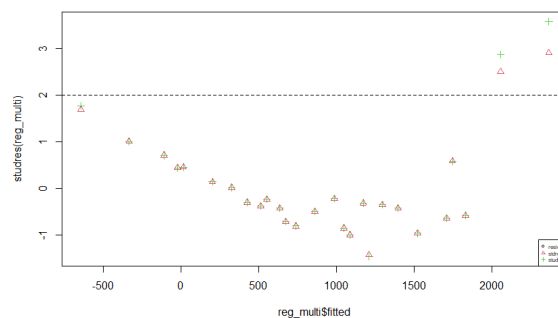
Commentons les résultats obtenus. On expliquera dans un premier temps quelles valeurs changent dans le modèle M1bis et quelles valeurs ne sont pas modifiées (on parlera des estimateurs et des valeurs de statistique de test) puis on conclura sur la validité de nos modèles à travers les indicateurs à notre disposition (R et les p-values pour chaque test).

Entre le modèle M1 et le modèle M1bis sont conservées t_1, t_2, t_3 , les statistiques des tests de Student de significativité du coefficient devant les trois premières variables explicatives, $\hat{\theta}_2$ l'estimateur du coefficient devant la seconde variable explicative et $\hat{\sigma}_2$. L'erreur résiduelle, le coefficient de détermination et la valeur de la statistique utilisées pour le test de validité globale sont aussi conservés. On aboutira donc aux mêmes conclusions en ce qui concerne la significativité globale et la significativité de chaque paramètre pour les deux modèles.

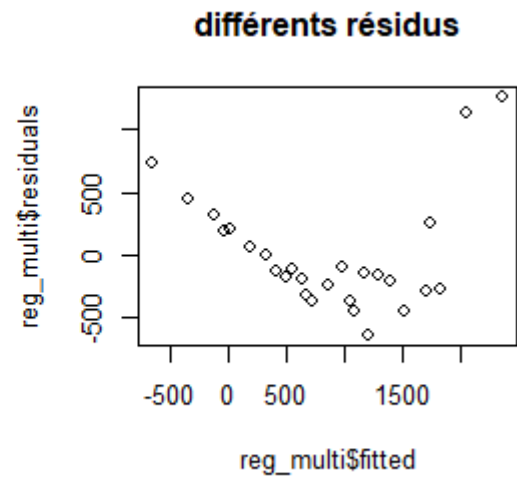
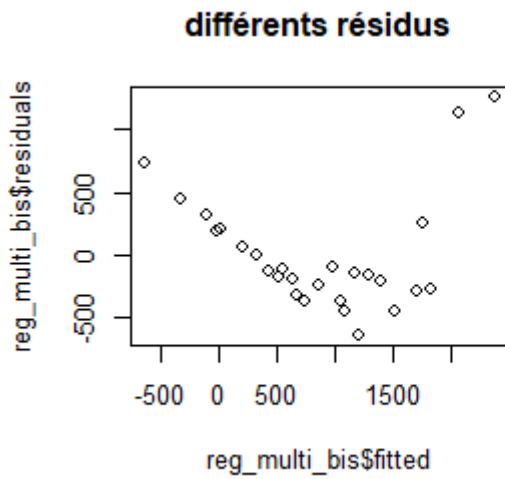
Comme le **coefficient de détermination** R^2 est le même dans chaque modèle : la part de variance expliquée par les variables explicatives est égale dans les deux modèles. Il en va de même pour le coefficient de corrélation **ajusté**. De plus, dans les deux modèles on rejette l'**hypothèse de nullité globale** avec un **risque de première espèce de 0.05** car, dans les deux modèles, pour ce test la **p-value est très faible** (p-value $\simeq 10^{-6}$). On considère aussi que chacune de nos variables est significative. En effet, on rejette l'hypothèse de nullité pour chaque variable avec un **risque de première espèce de 0.05** car les p-values associées sont toutes inférieures à 0.05.

Vérifions si les hypothèses sur le modèle sont vérifiées:

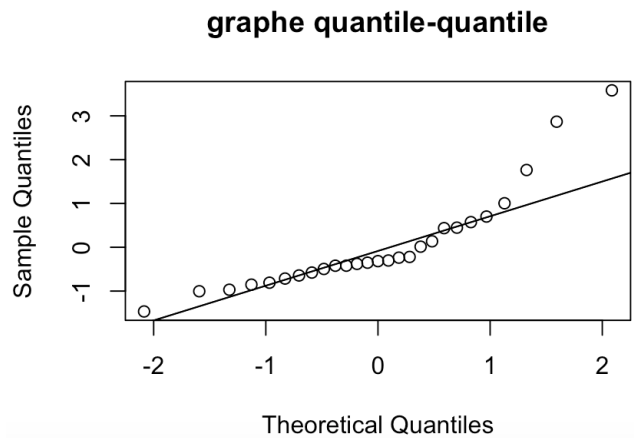
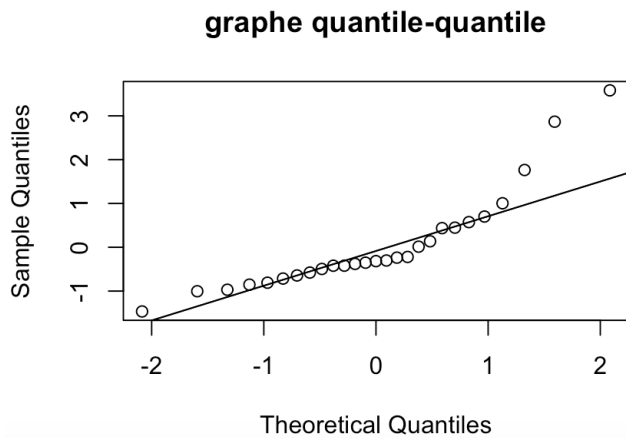
Pour cela, nous allons notamment étudier les résidus, pour ce qui est du modèle (M1) avant transformation, on obtient la représentation suivante des résidus en fonction des valeurs ajustées:



On peut remarquer que les résidus bruts sont assez élevés (supérieures à 2 en valeur absolue) ce qui est dû au fait que les valeurs observées sont déjà élevées (de l'ordre de 10^3) et que l'ajustement n'est pas des plus élevés.



On peut remarquer que les résidus du modèle avant transformation et après transformation sont égaux donc la figure précédente permet une interprétation des deux modèles. Aussi, on observe que la dispersion n'est pas du tout uniforme le long de l'axe des valeurs ajustées. On peut en conclure que l'hypothèse d'**homoscédasticité** n'est pas vérifiée. On peut dégager une **tendance dans les résidus** (forme parabolique).



Le graphe quantile quantile semble nous indiquer que l'hypothèse gaussienne est fautive, les résidus ne s'alignent pas bien le long de notre droite. On fait ensuite un test de Shapiro-Wilk et on obtient ce résultat pour les deux modèles (on montre ici la commande sur M1bis):

```
> shapiro.test(reg_multi_bis$residuals)
```

Shapiro-Wilk normality test

data: reg_multi_bis\$residuals
W = 0.86383, p-value = 0.002191

On rejette l'hypothèse des résidus gaussiens avec un **risque de première espèce de 0.05**

2) On construit un nouveau modèle où on prend en compte les interactions entre les différentes variables :

$$Y_i = x_{i1}\theta_1 + x_{i2}\theta_2 + x_{i3}\theta_3 + x_1^2\theta_4 + x_2^2\theta_5 + x_3^2\theta_6 + x_{i1}x_{i2}\theta_7 + x_{i1}x_{i3}\theta_8 + x_{i2}x_{i3}\theta_9\epsilon_i, \text{ où } \epsilon_i \sim \mathcal{N}(0, \sigma^2)$$

```
reg_multi2<-lm(y~x1 + x2 + x3 + I(x1^2) + I(x2^2) + I(x3^2) + I(x1*x2) + I(x1*x3) + I(x2*x3),data=cycle)
```

On obtient les résultats suivants :

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	550.70	138.44	3.978	0.000973 ***
x1	660.00	64.09	10.299	1.00e-08 ***
x2	-535.83	64.09	-8.361	1.99e-07 ***
x3	-310.83	64.09	-4.850	0.000150 ***
I(x1^2)	238.56	111.00	2.149	0.046317 *
I(x2^2)	275.72	111.00	2.484	0.023712 *
I(x3^2)	-48.28	111.00	-0.435	0.669081
I(x1 * x2)	-456.50	78.49	-5.816	2.06e-05 ***
I(x1 * x3)	-235.67	78.49	-3.003	0.008011 **
I(x2 * x3)	142.92	78.49	1.821	0.086278 .

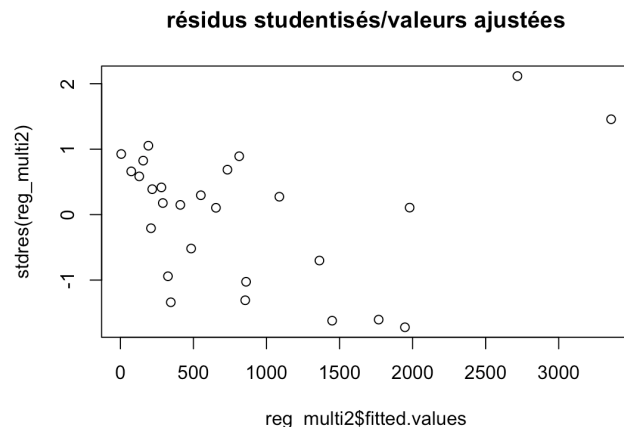
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(a) Données de la régression linéaire de M2

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

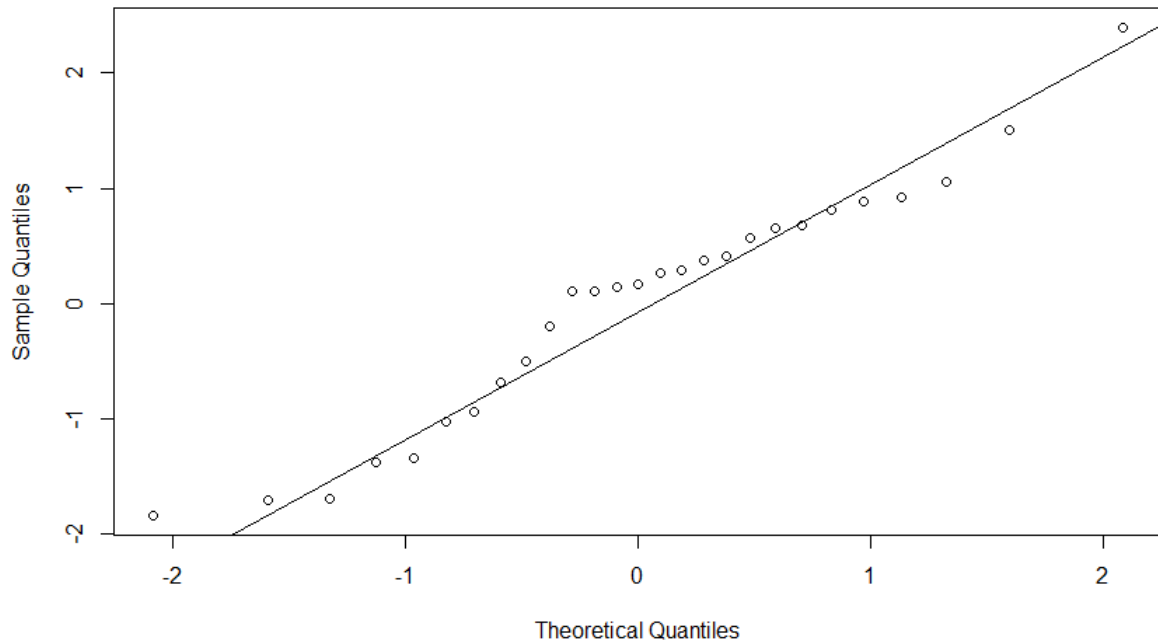
(b) Données de la régression linéaire de M2

On remarque que l'erreur résiduelle est plus faible que dans les modèles M1 avant et après transformation, on obtient donc un R^2 beaucoup plus proche de 1 : $R^2 = 0.9379$ ce qui traduit un très bon ajustement. La part de variance expliquée par le modèle est plus importante. Ensuite, on remarque que la variable x_3^2 n'est pas significative (p-value $\simeq 0.66$), de même pour la variable x_2x_3 on ne peut pas rejeter l'hypothèse de nullité car la p-value vaut environ 0.08. On la conserve donc avec un **risque de seconde espèce β inconnu**.



On remarque que la dispersion des résidus n'est pas uniforme car plus les valeurs ajustées augmentent, plus les résidus augmentent en norme. On **rejette donc l'hypothèse d'homoscédasticité**

graphe quantile-quantile



Le graphe quantile quantile semble nous indiquer que l'hypothèse gaussienne est plausible, on fait ensuite un test de Shapiro-Wilk pour confirmer cela et on obtient ce résultat:

```
> shapiro.test(stdres(reg_multi2))

      shapiro-wilk normality test

data:  stdres(reg_multi2)
W = 0.9555, p-value = 0.2904
```

On conserve donc l'hypothèse des résidus gaussiens avec un **risque de seconde espèce inconnu**.

Construisons le test de (M1) contre (M2): Il s'agit en fait de tester l'hypothèse nulle H_0 (modèle iid, tous les coefficients d'ordre 2 sont nuls) contre l'hypothèse alternative H_1 . On traduit les hypothèses comme suit :

$$\begin{cases} H_0 : \forall i \in \{4, \dots, 9\} : \theta_i = 0 \\ \text{vs } H_1 : \exists i \in \{4, \dots, 9\} : \theta_i \neq 0 \end{cases} \quad (2)$$

On peut pour cela utiliser la **statistique de Wald** qui s'utilise pour l'estimateur du maximum de vraisemblance or on sait que l'estimateur du maximum de vraisemblance est égal à l'estimateur des moindres carrés dans le cadre du modèle de régression linéaire gaussien multiple.

En choisissant $A = (0_{6,3} \quad \mathbb{I}_6)$ et $\theta = (\theta_1, \dots, \theta_9)^T$, cela revient à tester :

$$(H_0) : A\theta = 0 \text{ vs } (H_1) : A\theta \neq 0$$

où $r = \text{rang}(A) = 6$, de plus $n = 27$ observations et $p = 10$

On en déduit la statistique

$$W_{\text{exact}} = \frac{W}{r} = \frac{(A\hat{\theta})'(A\hat{V}A')^{-1}(A\hat{\theta})}{r} \sim \mathcal{F}(r, n-p)$$

On va donc rejeter l'hypothèse du modèle (M1) si $|W_{\text{exact}}| > f_{(r, n-p), 1-\alpha}$

Après le calcul grâce au logiciel R, on trouve:

```
> w=(t(A%%theta)%%solve(A%%V%%t(A))%% A%%theta)/r
> w
      [,1]
[1,] 9.522668
```

Calcul de la statistique W pour le test des modèles

```
> qf(0.95,r,n-p)
[1] 2.69866
```

Quantile de fisher d'ordre 0.95 pour r=6, n-p=17

```
> abs(w)>qf(0.95,r,n-p)
      [,1]
[1,] TRUE
```

Résultat du test de (M1) contre (M2)

Par le calcul on obtient $W_{exact} = 9.52 > 2.70$ donc on rejette l'hypothèse nulle avec un risque de première espèce de $\alpha = 0.05$ et on considère que l'ajout des variables d'interactions est à préconiser. Cela va dans le sens du résultat obtenu sur le coefficient de détermination de M2 qui est beaucoup plus proche de 1.

- 3) D'après l'énoncé, la transformation de Box-Cox peut être intéressante pour stabiliser la variable. On va donc utiliser la fonction **powerTransform** pour déterminer le λ qui permet le meilleur ajustement pour le modèle linéaire défini dans la partie 1 ($h_\lambda = x_i\theta + \epsilon_i$, $\epsilon_i \text{ iid } \sim \mathcal{N}(0, \sigma^2)$).

```
> repex3_1bis=powerTransform(reg_multi_bis)
> summary(repex3_1bis)
bcPower Transformation to Normality
  Est Power Rounded Pwr wald Lwr Bnd wald Up Bnd
y1   -0.0592      0   -0.1789   0.0606

Likelihood ratio test that transformation parameter is equal to 0
(log transformation)
              LRT df      pval
LR test, lambda = (0) 0.9213384 1 0.33712

Likelihood ratio test that no transformation is needed
              LRT df      pval
LR test, lambda = (1) 84.07566 1 < 2.22e-16
```

On obtient $\lambda = 0$ (en prenant la valeur arrondi, *Rounded Pwr*) ce qui nous donne le modèle suivant en appliquant la transformation de Box-Cox pour $\lambda = 0$:

$$\log(y_i) = x_i\theta + \epsilon_i \text{ (M1bis)}$$

En estimant le nouveau modèle on obtient la régression suivante:

```
> trans_y=log(cycle$y)
> trans_reg_1bis=lm(trans_y~cycle$x1+cycle$x2+cycle$x3)
> summary(trans_reg_1bis)

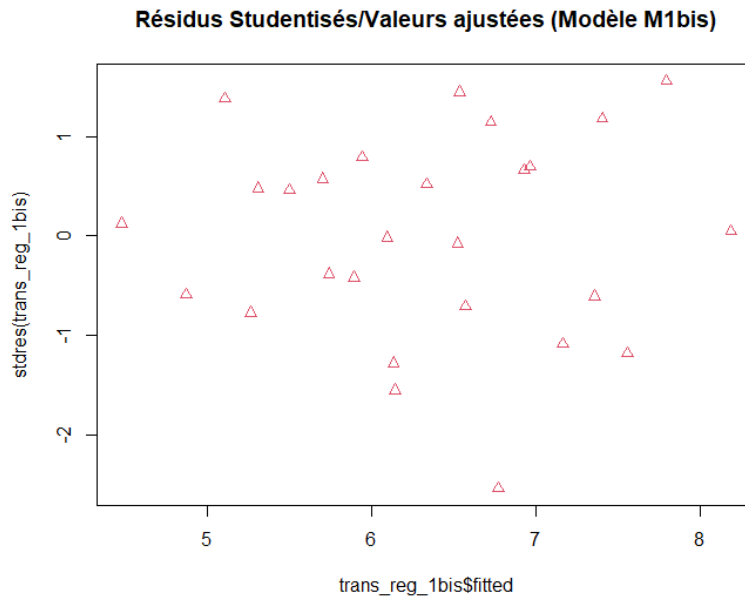
Call:
lm(formula = trans_y ~ cycle$x1 + cycle$x2 + cycle$x3)

Residuals:
    Min       1Q   Median       3Q      Max
-0.43592 -0.11250  0.00802  0.11635  0.26790

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.33475    0.03572  177.327 < 2e-16 ***
cycle$x1     0.83238    0.04375   19.025 1.43e-15 ***
cycle$x2    -0.63087    0.04375  -14.419 5.22e-13 ***
cycle$x3    -0.39262    0.04375   -8.974 5.66e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1856 on 23 degrees of freedom
Multiple R-squared:  0.9658,    Adjusted R-squared:  0.9614
F-statistic: 216.8 on 3 and 23 DF,  p-value: < 2.2e-16
```

Aussi, on peut observer que le test de Fischer de significativité globale de la régression donne une p-value $2.2e^{-16} < 0.05$ donc on rejette l'hypothèse de non significativité globale de la régression. De plus, les statistiques du test de Student de significativité des coefficients et leurs p-value respectives ($< 2e^{-16}$, $1.43e^{-15}$, $5.22e^{-13}$ et $5.66e^{-9}$ sont tous inférieurs à 0.05) permettent de déduire que l'on rejette l'hypothèse de non-significativité pour chacun des coefficients respectivement, aussi il faut les conserver tous dans notre modèle. et on obtient la figure suivante pour les résidus studentisés en fonction des valeurs ajustées :



On constate donc que pour (M1bis) la transformation stabilise bien la variance, en effet la dispersion des résidus est uniforme le long de l'axe des valeurs ajustées. On peut donc conserver **l'hypothèse d'homoscédasticité**. De plus, désormais les résidus peuvent être considérés comme gaussien (en prenant un risque de seconde espèce inconnu) car l'on obtient le résultat suivant en effectuant un test de normalité de Shapiro:

```
> shapiro.test(stdres(trans_reg_1bis))

Shapiro-wilk normality test

data:  stdres(trans_reg_1bis)
W = 0.96287, p-value = 0.4287
```

Etudions désormais le modèle (M2bis):

Nous allons lui appliquer une transformation similaire à celle utilisé pour obtenir le modèle (M1) précédemment, toutefois par curiosité nous pouvons vérifier si cette transformation est optimale pour le modèle (M2):

```
> #Pour le modèle (M2bis)
> repex3_2bis=powerTransform(reg_multi2)
> summary(repex3_2bis)
bcPower Transformation to Normality
  Est Power Rounded Pwr Wald Lwr Bnd Wald Up Bnd
Y1  -0.2188      0    -0.5179    0.0803

Likelihood ratio test that transformation parameter is equal to 0
(log transformation)
              LRT df    pval
LR test, lambda = (0) 1.743789  1 0.18666

Likelihood ratio test that no transformation is needed
              LRT df    pval
LR test, lambda = (1) 50.97355  1 9.3614e-13
```

Aussi, on retrouve $\lambda = 0$ (en prenant la valeur arrondi, *Rounded Pwr*) et désormais en appliquant le même transformation de Box-Cox que précédemment on obtient le nouveau modèle et la régression suivante:


```

> trans_reg_2bis=lm(log(y)~x1+x2+x3+I(x1^2) + I(x2^2) + I(x3^2) + I(x1*x2) + I(x1*x3) + I(x2*x3),data = cycle)
> summary(trans_reg_2bis)

Call:
lm(formula = log(y) ~ x1 + x2 + x3 + I(x1^2) + I(x2^2) + I(x3^2) +
  I(x1 * x2) + I(x1 * x3) + I(x2 * x3), data = cycle)

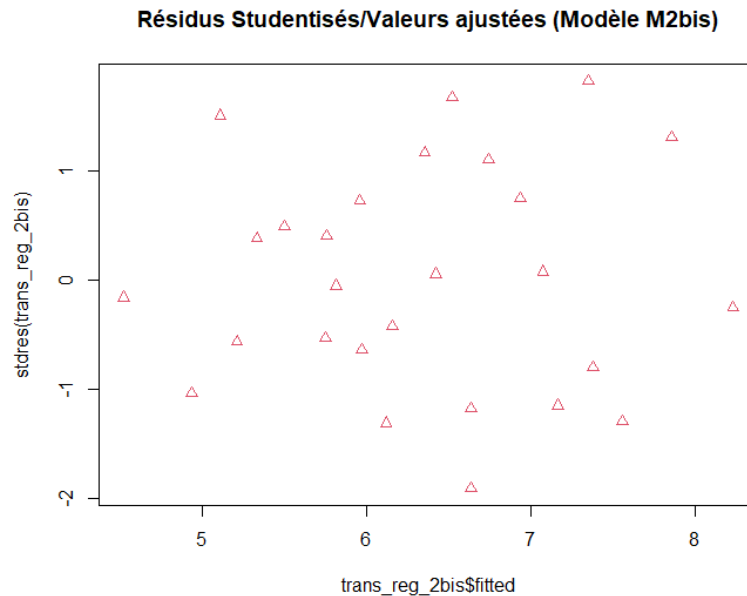
Residuals:
    Min       1Q   Median       3Q      Max
-0.300182 -0.106445 -0.009047  0.111572  0.262081

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.42070    0.09867   65.072 < 2e-16 ***
x1           0.83238    0.04568   18.224 1.36e-12 ***
x2          -0.63087    0.04568  -13.812 1.14e-10 ***
x3          -0.39262    0.04568   -8.596 1.35e-07 ***
I(x1^2)      -0.08595    0.07911   -1.086  0.292
I(x2^2)       0.02434    0.07911    0.308  0.762
I(x3^2)      -0.06733    0.07911   -0.851  0.407
I(x1 * x2)   -0.03824    0.05594   -0.684  0.503
I(x1 * x3)   -0.06841    0.05594   -1.223  0.238
I(x2 * x3)   -0.02102    0.05594   -0.376  0.712
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1938 on 17 degrees of freedom
Multiple R-squared:  0.9725,    Adjusted R-squared:  0.9579
F-statistic: 66.76 on 9 and 17 DF,  p-value: 1.724e-11

```

Aussi, on peut observer que le test de Fischer de significativité globale de la régression donne une p-value $1.724e^{-11} < 0.05$ donc on rejette l'hypothèse de non significativité globale de la régression. De plus, les statistiques du test de Student de significativité des coefficients et leurs p-value respectives permettent de déduire que l'on rejette l'hypothèse de non-significativité pour chacun des coefficients d'ordre 1, toutefois on accepte l'hypothèse de non significativité car les p-values associées sont strictement supérieures à 0.05 pour les coefficients d'ordre 2 (en prenant un risque de seconde espèce inconnu) aussi ces coefficients ne sont pas très pertinent dans notre modèle. De plus, nous obtenons la figure suivante pour les résidus studentisés en fonction des valeurs ajustées :



On constate aussi dans (M2bis) la transformation stabilise bien la variance, en effet la dispersion des résidus est uniforme le long de l'axe des valeurs ajustées. On peut donc conserver **l'hypothèse d'homoscédasticité**.

- 4) Pour déterminer le modèle le plus on va construire un test de (M1bis) contre (M2bis). Il s'agit en fait de tester l'hypothèse nulle H_0 (modèle iid, tous les coefficients d'ordre 2 sont nuls) contre l'hypothèse alternative H_1 . On traduit les hypothèses comme suit :

$$\begin{cases} H_0 : \forall i \in \{4, \dots, 9\} : \theta_i = 0 \\ \text{vs } H_1 : \exists i \in \{4, \dots, 9\} : \theta_i \neq 0 \end{cases} \quad (3)$$

On peut pour cela utiliser la **statistique de Wald** qui s'utilise avec l'estimateur du maximum de vraisemblance or on sait que l'estimateur du maximum de vraisemblance est égal à l'estimateur des moindres carrés dans le cadre du modèle de régression linéaire gaussien multiple.

En choisissant $A = (0_{6,3} \quad \mathbb{I}_6)$ et $\theta = (\theta_1, \dots, \theta_9)^T$, cela revient à tester :

$$(H_0) : A\theta = 0 \text{ vs } (H_1) : A\theta \neq 0$$

où $r = \text{rang}(A) = 6$, de plus $n = 27$ observations et $p = 10$

On en déduit la statistique

$$W_{\text{exact}} = \frac{W}{r} = \frac{(A\hat{\theta})'(A\hat{V}A')^{-1}(A\hat{\theta})}{r} \sim \mathcal{F}(r, n - p)$$

On va donc rejeter l'hypothèse du modèle (M1) si $|W_{exact}| > f_{(r,n-p),1-\alpha}$
 L'implémentation sur R correspond au code suivant:

```
#Question 4

p=10
n=27
r=6
A=matrix(c(0,0,0,0,1,0,0,0,0,0,
           0,0,0,0,0,1,0,0,0,0,
           0,0,0,0,0,0,1,0,0,0,
           0,0,0,0,0,0,0,1,0,0,
           0,0,0,0,0,0,0,0,1,0,
           0,0,0,0,0,0,0,0,0,1,
           0,0,0,0,0,0,0,0,0,1
           ),nrow=6,ncol=10,byrow=TRUE)
theta=trans_reg_2bis$coefficients
V=vcov(trans_reg_2bis)

sigma=sqrt(sum((log(cycle$y)-trans_reg_2bis$fitted.values)^2)/(n-p))

w=(t(A%%theta)%%solve(A%%V%%t(A))%% A%%theta)/r

qf(0.95,r,n-p)

abs(w)>qf(0.95,r,n-p)
```

Après le calcul grâce au logiciel R, on trouve:

```
> w
      [,1]
[1,] 0.6839074
> qf(0.95,r,n-p)
[1] 2.69866
> abs(w)>qf(0.95,r,n-p)
      [,1]
[1,] FALSE
```

On conserve donc l'hypothèse H_0 avec un **risque de seconde espèce inconnu** et donc le modèle que l'on proposera est (M1bis).

Remarque: On pouvait se douter de ce résultat car nous avons obtenu à la question précédente lors de l'analyse de la régression linéaire pour le modèle (*M2bis*) que l'usage des termes d'ordre 2 dans ce modèle n'était pas pertinent.

Conclusion

Tout au long de cette étude nous avons pu étudier un domaine spécifique de l'étude des régressions linéaires: les transformations en puissance de la variable à expliquer. Aussi, nous avons pu observer deux types de transformations de ce type: la famille de transformations de Box-Cox et celle de Bickel et Docksum (qui sont en fait équivalentes pour des observations y positives et pour $\lambda > 0$). Dans un premier temps, nous avons fait l'étude théorique de ce type de transformations en étudiant notamment la cohérence de ce type de transformation ainsi que plusieurs propriétés vérifiées par les estimateurs du maximum de vraisemblance dans ce genre de modèle, notamment l'estimateur du maximum de vraisemblance $\hat{\lambda}$ de λ , ce qui nous a permis de construire des statistiques de test et des intervalles de confiance associés à λ . Par la suite, nous avons pu tester nos découvertes théoriques sur un échantillon de données simulées grâce au langage R. Aussi, nous avons notamment étudié la transformation de Bickel et Docksum sur ce jeu de données simulées afin de calculer notamment $L_{max}(\lambda)$, de le minimiser afin de trouver $\hat{\lambda}$ et un estimateur de sa variance, ce qui nous a finalement permis d'effectuer différents tests statistiques. Enfin, dans la dernière partie nous avons étudié un cas pratique permettant d'utiliser les transformations de Box-Cox afin de notamment stabiliser la variance d'un échantillon statistique. Ainsi, ce projet nous a permis une première approche des transformations en la variable à expliquer, permettant d'améliorer les qualités de nos échantillons de données afin de vérifier au mieux les hypothèses du modèle de régression linéaire.