

Projet Modèles de Régression

Université de Paris Nanterre
M1 ISÉFAR

Radwan SARMINI DET SATOUF - Sirine GABRIEL- Anissa GOULIF

3 Modèle de Poisson

Question 1:

Vérifions que $E(Y_i) = \exp(x_i; \beta)$:

$$E(Y_i) = \exp\left(\beta_0 + \sum_{j=1}^p \beta_j x_{ij}\right)$$

$$E(Y_i) = \exp(\beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \dots + \beta_p x_{i,p})$$

$$E(Y_i) = \exp((1, x_{i,1}, \dots, x_{i,p}) \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_p \end{pmatrix})$$

nous avons bien: $E(Y_i) = \exp(x_i; \beta)$

Car en effet, $x_i = (1, x_{i,1}, \dots, x_{i,p})$ et $\beta = \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_p \end{pmatrix}$

ainsi $x_i \beta = \beta_0 + x_{i,1} \beta_1 + x_{i,2} \beta_2 + \dots + x_{i,p} \beta_p$

$$x_i \beta = \beta_0 + \sum_{j=1}^p \beta_j x_{ij}$$

Question 2:

Il s'agit bien d'un modèle linéaire généralisé car:

① Les lois de Y_i appartiennent à une même famille exponentielle:
→ Les Y_i sont indépendants et suivent une loi de Poisson de paramètre θ_i (supposé); de fonction de masse $\frac{e^{-\theta_i} \theta_i^y}{y!} : \frac{\exp(y \log(\theta_i) - \theta_i)}{y!}$

Ainsi, les Y_i appartiennent à la famille exponentielle caractérisée par

$$f(y_i; m_i) = c(y_i, \phi) \exp\left(\frac{y_i a(m_i) - b(m_i)}{\phi}\right)$$

avec $c(y_i, \phi) = \frac{1}{y_i!}$; $a(m_i) = \log(\theta_i)$; $b(m_i) = \theta_i$; $m_i = \theta_i = E(Y_i)$

et $\phi = 1$ où ϕ est la variance.

avec x_i, y_i suivant la

Poisson

$i \in \{1, \dots, n\}$

$1 \leq p \leq n$

$j \in \{1, \dots, p\}$

x_{ij} variable explicative

$$\beta = \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_p \end{pmatrix}; x_i = (1, x_{i,1}, \dots, x_{i,p})$$

② Il existe $\beta \in \mathbb{R}^p$ inconnu tel que $g(m_i) = x_i \cdot \beta$

où g est dérivable et strictement monotone.

en effet nous avons $E(Y_i) = \exp(x_i \cdot \beta)$

$\Leftrightarrow g(E(Y_i)) = x_i \cdot \beta$ où $g(x) = \log(x)$

La fonction \log étant dérivable sur $[0; +\infty[$ et strictement monotone.

La fonction lien pour le modèle linéaire généralisé est donc $g(x) = \log(x)$.

Question 3:

Pour déclarer un modèle linéaire généralisé il est nécessaire de spécifier les 2 éléments vus à la question 2 mais il est également nécessaire de considérer la matrice $X_{(n,p)}$

Les lignes de X sont les x_i ; il s'agit de la matrice du plan d'expérience contenant les variables explicatives et fictives.

$$X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} 1 & x_{12} & \dots & x_{1p} \\ \vdots & & & \\ 1 & x_{n2} & \dots & x_{np} \end{pmatrix} \quad \text{Rg}(X) = p.$$

Question 4 :

Montrons que le vecteur score admet comme première composante :

$$\sum_{i=1}^n (Y_i - \exp(x_i \beta))$$

$U_n(\beta)$ est le vecteur score et la j^{th} composante est

$$u_j(\beta) = \sum_{i=1}^n \frac{\partial}{\partial \beta_j} \log f(Y_i, g^{-1}(x_i \beta)). \quad \begin{aligned} g(x) &= \log(x) \\ g^{-1}(x) &= \exp(x) \end{aligned}$$

Ainsi nous avons :

$$\begin{aligned} u_1(\beta) &= \sum_{i=1}^n \frac{\partial}{\partial \beta_1} \log f(Y_i, g^{-1}(x_i \beta)) \\ &= \sum_{i=1}^n \frac{\partial}{\partial \beta_1} \log \left(\frac{1}{Y_i!} \exp(Y_i x_i \beta) - \exp(x_i \beta) \right) \\ &= \sum_{i=1}^n \frac{\partial}{\partial \beta_1} \left[\log \left(\frac{1}{Y_i!} \right) + \log \left(\exp(Y_i x_i \beta) - \exp(x_i \beta) \right) \right] \\ u_1(\beta) &= \sum_{i=1}^n \left(\frac{\partial}{\partial \beta_1} \left(Y_i x_i \beta \right) - \frac{\partial}{\partial \beta_1} \exp(x_i \beta) + \frac{\partial}{\partial \beta_1} \log \left(\frac{1}{Y_i!} \right) \right) \end{aligned}$$

ne dépend pas de β_1

$$\frac{\partial Y_i x_i \beta}{\partial \beta_1} = \frac{\partial}{\partial \beta_1} Y_i (\beta_1 + x_{i2} \beta_2 + \dots + x_{ip} \beta_p) = Y_i$$

$$\begin{aligned} \frac{\partial \exp(x_i \beta)}{\partial \beta_1} &= \frac{\partial}{\partial \beta_1} \exp(\beta_1 + x_{i2} \beta_2 + \dots + x_{ip} \beta_p) = 1 \times \exp(\beta_1 + x_{i2} \beta_2 + \dots + x_{ip} \beta_p) \\ &= 1 \times \exp(x_i \beta) \\ &= \exp(x_i \beta) \end{aligned}$$

D'où $u_1(\beta) = \sum_{i=1}^n (Y_i - \exp(x_i \beta))$

Question 5:

On suppose $\hat{\beta}$ l'estimateur du maximum de vraisemblance

Comme $\hat{\beta}$ est solution des équations normales $U_n(\hat{\beta})=0$

alors cela implique $u_j(\hat{\beta})=0 \quad \forall j \in \{1, \dots, p\}$.

$$\Leftrightarrow U_1(\hat{\beta}) = 0$$

$$\Leftrightarrow \sum_{i=1}^n (Y_i - \exp(x_i \hat{\beta})) = 0$$

$$\Leftrightarrow \sum_{i=1}^n Y_i - \sum_{i=1}^n \exp(x_i \hat{\beta}) = 0$$

$$\Leftrightarrow \sum_{i=1}^n Y_i = \sum_{i=1}^n \exp(x_i \hat{\beta}) \quad \square$$

Question 6:

Calcul de la déviance du modèle

$$D(Y) = 2(\hat{L}_S - \hat{L}) \quad \text{avec} \quad \hat{L}_S : \text{la log-vraisemblance maximale du modèle suivri}$$

\hat{L} : la log-vraisemblance maximale dans le modèle \mathcal{M} .

$$\hat{L} = \log \left(\prod_{i=1}^n \frac{\exp(-\exp(x_i \hat{\beta})) \exp(x_i \hat{\beta})^{Y_i}}{Y_i!} \right)$$

$$= \log \left(\prod_{i=1}^n \exp(-\exp(x_i \hat{\beta})) \prod_{i=1}^n \exp(x_i \hat{\beta})^{Y_i} \right) + \log \left(\prod_{i=1}^n \frac{1}{Y_i!} \right) \quad \Delta Y_i \neq 0$$

$$= \sum_{i=1}^n \log(\exp(-\exp(x_i \hat{\beta}))) + \sum_{i=1}^n \log(\exp(x_i \hat{\beta})^{Y_i}) + \sum_{i=1}^n -\log(Y_i!)$$

$$= \sum_{i=1}^n -\exp(x_i \hat{\beta}) + \sum_{i=1}^n Y_i \exp(x_i \hat{\beta}) + \sum_{i=1}^n -\log(Y_i!)$$

$$= \sum_{i=1}^n [Y_i x_i \hat{\beta} - \exp(x_i \hat{\beta}) - \log(Y_i!)] \quad \Delta Y_i \neq 0$$

On définit la modalité saturée comme le modèle ayant le plus grand nombre de paramètres possible, ainsi, ici il sera le cas de celui pour lequel la moyenne de la variable sera définie par l'observation elle-même : (c'est à dire $E(Y_i) = Y_i$).

$$\begin{aligned}\hat{\mathcal{L}}_s &= \log \left(\prod_{i=1}^n \frac{\exp(-Y_i)}{Y_i!} Y_i^{Y_i} \right) \\ &= \sum_{i=1}^n \left[\log \exp(-Y_i) + \log(Y_i^{Y_i}) - \log(Y_i!) \right] \quad \Delta Y_i \neq 0 \\ &= \sum_{i=1}^n \left[-Y_i + Y_i \log(Y_i) - \log(Y_i!) \right]\end{aligned}$$

Calcul de la déviance :

$$\begin{aligned}\mathcal{D}(Y) &= 2(\hat{\mathcal{L}}_s - \hat{\mathcal{L}}) \\ &= 2 \left(\sum_{i=1}^n (-Y_i + Y_i \log(Y_i) - \log(Y_i!)) \right. \\ &\quad \left. - \sum_{i=1}^n (x_i \hat{\beta} Y_i - \exp(x_i \hat{\beta}) - \log(Y_i!)) \right) \\ &= 2 \left(\sum_{i=1}^n (-Y_i + Y_i \log(Y_i) - \log(Y_i!)) - x_i \hat{\beta} Y_i + \exp(x_i \hat{\beta}) + \log(Y_i!) \right) \\ &= 2 \left(\sum_{i=1}^n Y_i (\log(Y_i) - x_i \hat{\beta}) + \sum_{i=1}^n \exp(x_i \hat{\beta}) - Y_i \right) \quad \text{montre à la question 5} \\ &= 2 \left(\sum_{i=1}^n Y_i (\log(Y_i) - x_i \hat{\beta}) \right) \quad \text{avec } Y_i \neq 0\end{aligned}$$

Nous avons bien $\mathcal{D}(Y) = 2 \left(\sum_{i=1}^n Y_i (\log(Y_i) - x_i \hat{\beta}) \right)$ avec $Y_i \neq 0$.

Question 7

On teste $H_0: \beta_2 = \beta_3$ contre $H_1: \beta_2 \neq \beta_3$. On note $\hat{\beta}_0$ l'estimateur du maximum de vraisemblance de β calculé sous l'hypothèse H_0 (on suppose qu'il est bien défini).

a) Montrons que la statistique du test du rapport des vraisemblances est $-2 \sum_{i=1}^n Y_i (\ln(\hat{Y}_i) - \ln(Y_i))$

Le rapport de vraisemblance est l'écart entre les déviations du modèle réduit sous H_0 et du modèle complet :

$$\begin{aligned} T_q &= D(\text{modèle sous } H_0) - D(\text{modèle complet}) \\ &= -2 \left(\sum_{i=1}^n Y_i (\ln(Y_i) - \ln(\hat{Y}_i)) \right) - 2 \left(\sum_{i=1}^n Y_i (\ln(Y_i) - \ln(\hat{Y}_i)) \right) \\ &= 2 \left(\sum_{i=1}^n Y_i (\ln(Y_i) - \ln(\hat{Y}_i) - \ln(Y_i) + \ln(\hat{Y}_i)) \right) \\ T_q &= -2 \left(\sum_{i=1}^n Y_i (\ln(\hat{Y}_i) - \ln(Y_i)) \right) \end{aligned}$$

Nous obtenons bien le résultat souhaité avec $g=2$ car $h(\beta) = \begin{pmatrix} \beta_2 \\ \beta_3 \end{pmatrix}$ alors $h: \mathbb{R}^2 \rightarrow \mathbb{R}^2 : \operatorname{rg}(h) = 2$.

b) p-value du test

$$p\text{-value} = 1 - P(X^2(2) \leq -2 \sum_{i=1}^n Y_i (\ln(\hat{Y}_i) - \ln(Y_i)))$$

Il s'agit d'un test asymptotique car le test est basé sur l'estimateur du maximum de vraisemblance de β .

Si la statistique de test (T_q) converge en loi vers une loi de $X^2(2)$ sous H_0 ; la p-value est donc également asymptotique. La taille de l'échantillon est à prendre en compte (il faut qu'elle soit finie).

Question 8

- a) Ces commandes nous servent à simuler une loi de Poisson. De plus, la commande « set.seed » assure la reproductibilité de la séquence de nombres aléatoires qui vont être générer à la suite de cette commande. En effet, on simule ici un dataset de taille n=30 et on génère dans deux variables x1 et x2 des variables aléatoires normales multivariées. On « assemble » ensuite les x1 et x2 dans une variable grand X (X est donc composé de 3 colonnes : une colonne de 1, la colonne des x1 et la colonne des x2 créés précédemment). Ensuite nous créons la « réponse » Y de tel sorte à ce que cela suive une loi de Poisson en utilisant « rpois » qui prend en paramètre n (taille de l'échantillon de variable aléatoire à retourner que nous avons créé précédemment) ainsi que m qui est finalement l'espérance des variables indépendante. « m » est généré grâce à la formule que nous avons vu précédemment c'est-à-dire l'exponentielle des xi (donc ici X) fois bêta (paramètres inconnus de la loi). Bêta est généré par une concaténation de 1, 2 et 2.1, paramètres inconnus de notre loi de Poisson.
- b) Ici nous retrouvons un modèle linéaire avec comme variable à expliquer Y et comme variables explicatives x1 et x2. On utilise ici un modèle linéaire généralisé tout comme dans les questions précédentes. Le type de loi appliqué est alors la loi de Poisson. Le modèle établi par régression est donc $Y = 2.02078x_1 + 2.10530x_2 + 0.95990$ (c'est la commande summary qui nous informe de ces valeurs). De plus, ce que l'on remarque également c'est que la « Null deviance » est très forte (4235) donc on peut en conclure que le modèle « nul » (avec comme variable à expliquer Y mais sans variable explicative, donc avec uniquement l'intercept) ne semble pas être pertinent.
- c) A la question 5 nous avions trouvé que $\sum_{i=1}^n Y_i = \sum_{i=1}^n \exp(x_i \hat{\beta})$. C'est exactement ce que nous affirme encore une fois ces lignes de code. En effet, on a p qui vaut les prédictions « réponses » des valeurs du modèle fit créé précédemment et Y est la valeur créée à la question 8 a) et on trouve ici que la somme de l'ensemble des valeurs de p vaut la somme de l'ensemble des valeurs de Y pour i allant de 1 à 30 (car n=30 ici) ce qui appuie encore une fois ce que nous avions trouvé à la question 5 (on traduit ici $\exp(x_i \hat{\beta})$ comme étant la prédition de la variable réponse).
- d) Ici, le « l » créé nous informe des prédictions produites à l'échelle des prédicteurs additifs. C'est donc « l'inverse » des prédictions de p (p rend compte des prédictions de la variable réponse, donc variable à expliquer Y) puisqu'ici l rend compte des prédictions des variables explicatives additives (x_1+x_2 dans le cas de notre modèle fit). De plus, « dev » rend compte de la déviance du modèle, calculé tout comme à la question 6. On retrouve ici pour cette valeur, exactement la même valeur trouvée dans le summary du modèle fit (avec la commande summary(fit) faite à la question 1, on s'intéresse ici à la « residual deviance »).
- e) Intéressons-nous d'abord à l'anova qui, grâce à une analyse des variances, compare 2 modèles : notre modèle fit créé précédemment et un modèle fit0 où, au lieu d'avoir x_1+x_2 comme variables explicatives nous avons seulement un x qui rend lui-même compte de la somme de x1 et x2 mais calculé avant la création du modèle glm. On voit

sur l'anova que le modèle fit0 obtient une residual deviance plus importante que pour le modèle fit ce qui témoigne d'un manque d'ajustement par rapport au modèle fit. Le modèle fit semble donc être un peu plus pertinent que le modèle fit0. De plus, avec l'anova on trouve une p-value du test réalisé (test chisq, en effet on met ce test en paramètre car la statistique de test converge en loi vers une loi de chi deux) de 0.05166 et on retrouve exactement la même valeur en réutilisant la méthode de la question 7, c'est-à-dire en calculant la statistique de test grâce à la formule de la question 7 a) puis en exprimant la p-valeur à partir de cette statistique de test. En effet, l'objet « stat » créé ici est la statistique de test du rapport de vraisemblances (et est calculé, comme expliqué précédemment, avec la formule énoncée à la question 7 a). Et tout comme à la question 7 b), nous retrouvons la p-valeur grâce à l'objet stat en réutilisant ce que nous avons trouvé à la question 7 b) (cf cette question) en s'aidant de la commande pchisq (pour avoir la fonction de distribution et donc la probabilité de la formule que nous avons écrit à la question 7 b)). Enfin, ici la p-value est supérieure à 5%, nous pouvons donc en conclure que le meilleur modèle entre le modèle réduit (fit0) et le modèle non réduit (fit), le meilleur modèle entre les deux est le modèle réduit (la p-value étant supérieure à 5%, on conserve l'hypothèse H0 qui est que le meilleur modèle est le modèle réduit).

4 Données GermanCredit : régression logistique

Question 1

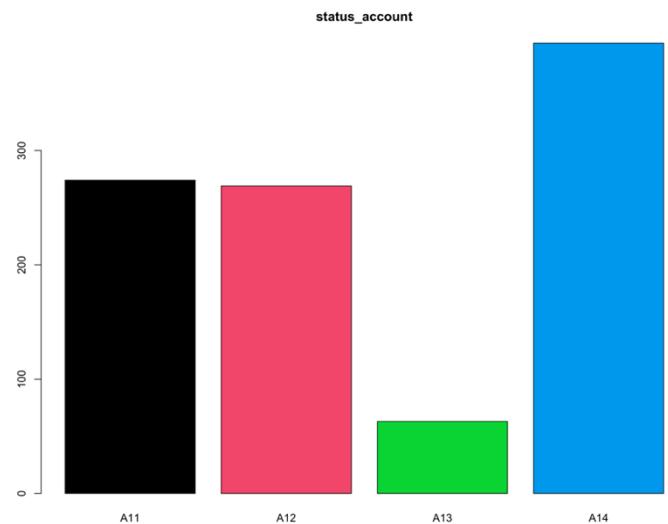
Class est une variable qualitative binaire. C'est une variable à expliquer qui représente le défaut de crédit soit un bon (Good) risque crédit ou un mauvais risque de crédit (Bad). Et 0 signifie Good et 1 signifie Bad.

Question 2

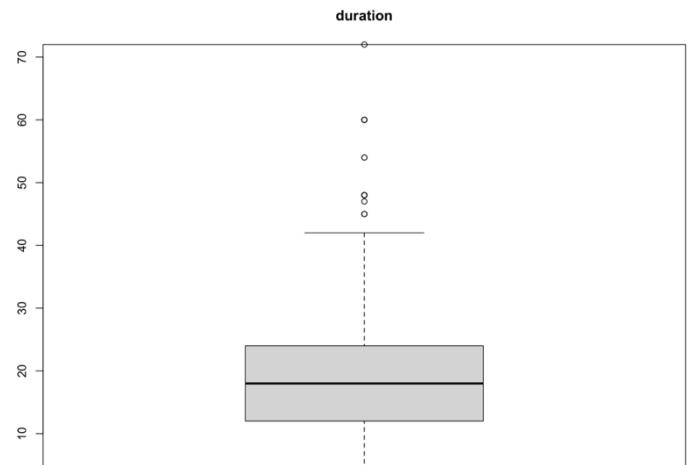
Ce jeu de données contient 1000 observations et 20 variables comme la suite :

- 1- « status_account » est une variable qualitative composée de 4 valeurs uniques :
 - a. A11 avec une fréquence de 274
 - b. A12 avec une fréquence de 269
 - c. A13 avec une fréquence de 63
 - d. A14 avec une fréquence de 394

Et on remarque que A14 est le mode



- 2- « duration » est une variable quantitative discrète avec les caractéristiques suivantes :
 - a. Min. : 4.0
 - b. 1er Quantile : 12.0
 - c. Médian : 18.0
 - d. Moyenne : 20.9
 - e. 3em Quantile. : 24.0
 - f. Max. : 72.0
 - g. Variance : 145.415



- 3- « history » est une variable qualitative composée de 5 valeurs uniques :
 - a. A30 avec une fréquence de 40
 - b. A31 avec une fréquence de 49
 - c. A32 avec une fréquence de 530
 - d. A33 avec une fréquence de 88
 - e. A34 avec une fréquence de 293

Et on remarque que A32 est le mode

- 4- « purpose» est une variable qualitative composée de 6 valeurs uniques :
 - a. A43 avec une fréquence de 280
 - b. A40 avec une fréquence de 234
 - c. A42 avec une fréquence de 181
 - d. A41 avec une fréquence de 103
 - e. A49 avec une fréquence de 97

- f. A46 avec une fréquence de 50
- Et on remarque que A43 est le mode

5- « amount» est une variable quantitative continue avec les caractéristiques suivantes :

- a. Min. : 250
- b. 1er Quantile : 1366
- c. Médian : 2320
- d. Moyenne : 3271
- e. 3em Quantile. : 3972
- f. Max. : 18424
- g. Variance : 7967843.47

6- « savings» est une variable qualitative composée de 5 valeurs uniques :

- a. A61 avec une fréquence de 603
- b. A62 avec une fréquence de 103
- c. A63 avec une fréquence de 63
- d. A64 avec une fréquence de 48
- e. A65 avec une fréquence de 183

Et on remarque que A61 est le mode

7- « employment» est une variable qualitative composée de 5 valeurs uniques :

- a. A71 avec une fréquence de 62
- b. A72 avec une fréquence de 172
- c. A73 avec une fréquence de 339
- d. A74 avec une fréquence de 174
- e. A75 avec une fréquence de 253

Et on remarque que A73 est le mode

8- « amount» est une variable quantitative discrète avec les caractéristiques suivantes :

- a. Min. : 1
- b. 1er Quantile : 2
- c. Médian : 3
- d. Moyenne : 2.973
- e. 3em Quantile. : 4
- f. Max. : 4
- g. Variance : 1.25

9- « status_personnal» est une variable qualitative composée de 4 valeurs uniques :

- a. A91 avec une fréquence de 50
- b. A92 avec une fréquence de 310
- c. A93 avec une fréquence de 548
- d. A94 avec une fréquence de 92

Et on remarque que A93 est le mode

10- « other_debtors» est une variable qualitative composée de 3 valeurs uniques :

- a. A101 avec une fréquence de 907
- b. A102 avec une fréquence de 41

- c. A103 avec une fréquence de 52
- Et on remarque que A101 est le mode

11- « residence_since » est une variable quantitative discrète avec les caractéristiques suivantes :

- a. Min. : 1
- b. 1er Quantile : 2
- c. Médian : 3
- d. Moyenne : 2.873
- e. 3em Quantile. : 4
- f. Max. : 4
- g. Variance : 1.21

12- « property » est une variable qualitative composée de 4 valeurs uniques :

- a. A121 avec une fréquence de 282
- b. A122 avec une fréquence de 232
- c. A123 avec une fréquence de 332
- d. A124 avec une fréquence de 154

Et on remarque que A123 est le mode

13- « age » est une variable quantitative discrète avec les caractéristiques suivantes :

- a. Min. : 19
- b. 1er Quantile : 27
- c. Médian : 33
- d. Moyenne : 35.55
- e. 3em Quantile. : 42
- f. Max. : 75
- g. Variance : 129.40

14- « other_installment » est une variable qualitative composée de 3 valeurs uniques :

- a. A141 avec une fréquence de 139
- b. A142 avec une fréquence de 47
- c. A143 avec une fréquence de 814

Et on remarque que A143 est le mode

15- « housing » est une variable qualitative composée de 3 valeurs uniques :

- a. A151 avec une fréquence de 179
- b. A152 avec une fréquence de 713
- c. A153 avec une fréquence de 108

Et on remarque que A152 est le mode

16- « nb_credits » est une variable quantitative discrète avec les caractéristiques suivantes :

- a. Min. : 1
- b. 1er Quantile : 1
- c. Médian : 1
- d. Moyenne : 1.4

- e. 3em Quantile. : 2
- f. Max. : 4
- g. Variance : 0.33

17- « job » est une variable qualitative composée de 4 valeurs uniques :

- a. A171 avec une fréquence de 22
- b. A172 avec une fréquence de 200
- c. A173 avec une fréquence de 630
- d. A174 avec une fréquence de 148

Et on remarque que A173 est le mode

18- « nb_credits » est une variable quantitative discrète avec les caractéristiques suivantes :

- a. Min. : 1
- b. 1er Quantile : 1
- c. Médian : 1
- d. Moyenne : 1.5
- e. 3em Quantile. : 2
- f. Max. : 2
- g. Variance : 0.13

19- « telephone » est une variable qualitative composée de 2 valeurs uniques :

- a. A191 avec une fréquence de 596
- b. A192 avec une fréquence de 404

Et on remarque que A173 est le mode

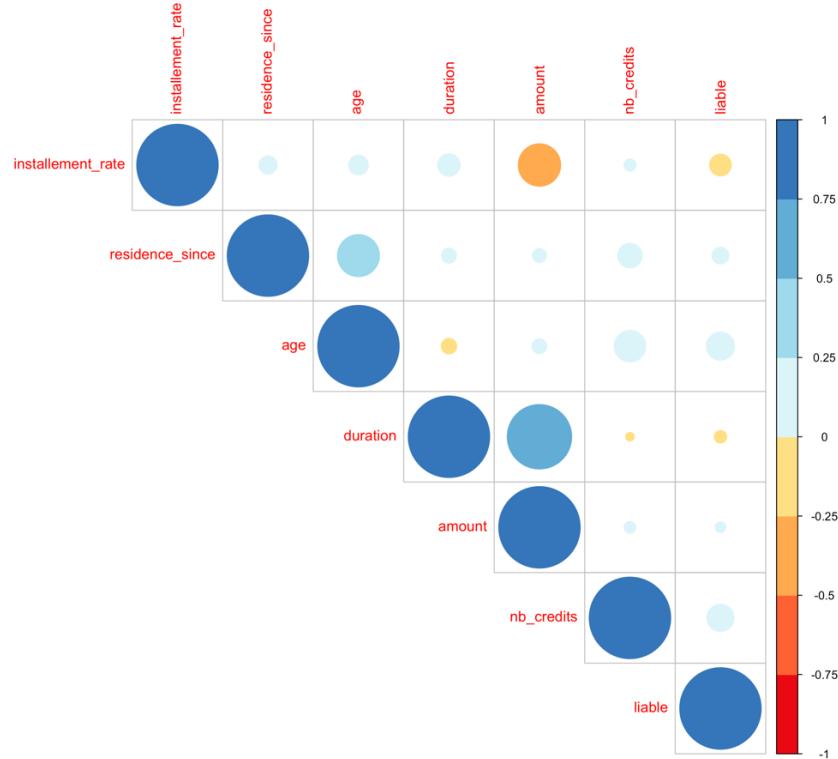
20- « class » est une variable qualitative composée de 2 valeurs uniques :

- a. A171 avec une fréquence de 700
- b. A172 avec une fréquence de 300

Et on remarque que A173 est le mode

Note : Une fonction s'appelle « descr » dans le code R où on peut trouver tous les graphiques de ce jeu de données.

La Corrélation entre les variables quantitatives :



On remarque que la variance est très grande pour duration, amount et age il sera préférable de réduire ces écarts en prenant la somme de la différence entre les valeurs d'une variable et la moyenne de cette variable en divisant par la variance de cette variable (équivalent à l'ACP normé) (Comme dans la méthode de LASSO)

	duration	amount	installement_rate	residence_since	age	nb_credits	liable
duration	145.4150060	21273.74978	1.00838939	0.45341842	-4.9569950	-0.07859960	-0.10406907
amount	21273.7497758	7967843.47091	-856.77080480	90.12011011	1050.5226547	33.90690090	17.52053053
installement_rate	1.0083894	-856.77080	1.25152252	0.06087588	0.7414835	0.01400300	-0.02884384
residence_since	0.4534184	90.12011	0.06087588	1.21819319	3.3449750	0.05714214	0.01704204
age	-4.9569950	1050.52265	0.74148348	3.34497497	129.4012853	0.98075876	0.48685686
nb_credits	-0.0785996	33.90690	0.01400300	0.05714214	0.9807588	0.33368468	0.02293794
liable	-0.1040691	17.52053	-0.02884384	0.01704204	0.4868569	0.02293794	0.13110611

Question 3

```
modReg <- glm( class ~ ., data = credit1, family = binomial )
```

Question 4

Si on fait `summary(modReg)` on trouve la suite :

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	1.688e-01	1.076e+00	0.157	0.875351
status_accountA12	-3.238e-01	2.157e-01	-1.501	0.133320
status_accountA13	-9.252e-01	3.677e-01	-2.516	0.011854 *
status_accountA14	-1.685e+00	2.311e-01	-7.291	3.07e-13 ***
duration	2.965e-02	9.272e-03	3.198	0.001384 **
historyA31	2.212e-01	5.443e-01	0.406	0.684375
historyA32	-5.121e-01	4.276e-01	-1.198	0.231008
historyA33	-7.838e-01	4.696e-01	-1.669	0.095050 .
historyA34	-1.410e+00	4.378e-01	-3.220	0.001281 **
purposeA41	-1.635e+00	3.749e-01	-4.361	1.30e-05 ***
purposeA40	-1.534e+00	8.005e-01	-1.917	0.055280 .
purposeA42	-7.355e-01	2.587e-01	-2.843	0.004467 **
purposeA43	-8.410e-01	2.457e-01	-3.424	0.000618 ***
purposeA44	-4.477e-01	7.618e-01	-0.588	0.556760
purposeA45	-1.692e-01	5.453e-01	-0.310	0.756392
purposeA46	9.472e-02	3.957e-01	0.239	0.810804
purposeA48	-1.951e+00	1.207e+00	-1.616	0.106089
purposeA49	-6.829e-01	3.321e-01	-2.056	0.039764 *
amount	1.232e-04	4.446e-05	2.771	0.005581 **
savingsA62	-3.684e-01	2.847e-01	-1.294	0.195679
savingsA63	-3.851e-01	4.008e-01	-0.961	0.336560
savingsA64	-1.330e+00	5.237e-01	-2.540	0.011100 *
savingsA65	-9.662e-01	2.601e-01	-3.715	0.000203 ***
employmentA72	-1.246e-01	4.253e-01	-0.293	0.769543
employmentA73	-2.290e-01	4.096e-01	-0.559	0.576113
employmentA74	-8.664e-01	4.444e-01	-1.950	0.051226 .
employmentA75	-2.975e-01	4.130e-01	-0.720	0.471307
installement_rate	3.347e-01	8.784e-02	3.810	0.000139 ***
status_personnalA92	-2.553e-01	3.856e-01	-0.662	0.507904
status_personnalA93	-8.150e-01	3.791e-01	-2.150	0.031563 *
status_personnalA94	-3.857e-01	4.530e-01	-0.852	0.394473
other_debtorsA102	3.766e-01	4.032e-01	0.934	0.350283
other_debtorsA103	-1.057e+00	4.251e-01	-2.487	0.012874 *
residence_since	1.207e-02	8.592e-02	0.140	0.888302
propertyA122	2.713e-01	2.523e-01	1.075	0.282192
propertyA123	2.195e-01	2.350e-01	0.934	0.350367
propertyA124	7.121e-01	4.205e-01	1.694	0.090348 .
age	-1.488e-02	9.216e-03	-1.614	0.106506
other_installmentA142	-9.305e-02	4.120e-01	-0.226	0.821318
other_installmentA143	-6.242e-01	2.389e-01	-2.613	0.008966 **
housingA152	-4.143e-01	2.319e-01	-1.786	0.074051 .
housingA153	-6.097e-01	4.714e-01	-1.294	0.195819
nb_credits	2.925e-01	1.894e-01	1.544	0.122558
jobA172	5.076e-01	6.721e-01	0.755	0.450119
jobA173	5.550e-01	6.482e-01	0.856	0.391896
jobA174	4.634e-01	6.570e-01	0.705	0.480549
liable	2.550e-01	2.481e-01	1.028	0.304128
telephoneA192	-2.730e-01	2.007e-01	-1.360	0.173840

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1221.73 on 999 degrees of freedom
Residual deviance: 901.88 on 952 degrees of freedom
AIC: 997.88

Number of Fisher Scoring iterations: 5

On remarque que les variables suivantes particulièrement, ont des P-value très élevées : « residence_since », « property », « job », « age », « liable », « telephone» elles ne sont pas statistiquement significatives et on peut douter de la fiabilité comme elles ont des P-value plus grandes que 0.05. Et c'est pourquoi le modèle ne semble pas parcimonieux. (Et comme leurs Z-value est entre -2 et 2)

Note : d'après de la table :

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.3324	-0.7020	-0.3783	0.7198	2.6073

On voit que le modèle est bien centré et symétrique.

Deviance

La différence entre « deviance nul » et « deviance résiduel » montre comment notre modèle se comporte par rapport au modèle nul (un modèle avec seulement l'intercepte). Plus cet écart est large, mieux c'est. Dans notre cas l'écart est de 319.85 (26.18%)

McFadden's Pseudo R2

Pseudo R2 prend une valeur supérieure ou égale à 0 et strictement inférieure à 1, avec une valeur plus proche de zéro indiquant que le modèle n'a pas de pouvoir prédictif et inversement si la valeur est plus proche de 1.

```
ll.null <- modReg$null.deviance/-2  
ll.propse<- modReg$deviance/-2  
Pseudo_R2 <- (ll.null-ll.propse )/ll.null  
Pseudo_R2 # 0.26
```

Avec le modèle général on trouve que Pseudo R2 est de 0.26 donc un pouvoir prédictif moins que la moyenne.

P-value pour chi2 de R2

```
Chi2_R2 <- 1- pchisq(2*(ll.propse-ll.null),df=length(modReg$coefficients)-1)  
Chi2_R2 # 0
```

Comme Chi2_R2 est 0 donc la relation entre « class » et les autres variables n'est pas due au hasard, et la valeur R2, 0.26 nous indique la taille de l'effet de cette relation.

Note : Il faut noter que certaines valeurs dans les variables : « purpose, history, other_debtors job et other_installment » ne sont pas bien représentées par rapport aux autres variables. Par exemple A410, A44 et A48 dans la variable purpose, ne sont pas de tout bien représentées :

```
[1] "purpose"  
    credit1[, g]  
class A40 A41 A410 A42 A43 A44 A45 A46 A48 A49  
  0 145  86     7 123 218    8 14  28    8 63  
  1  89  17     5  58  62    4   8  22    1 34
```

Wald Test

Un test de Wald est utilisé pour évaluer la signification statistique de chaque coefficient dans le modèle et est calculé en prenant le rapport du carré du coefficient de régression au carré de l'erreur standard du coefficient. L'idée est de tester l'hypothèse que le coefficient d'une variable indépendante dans le modèle est significativement différent de zéro. Si le test ne rejette pas l'hypothèse nulle, cela suggère que la suppression de la variable du modèle ne nuira pas substantiellement à l'ajustement de ce modèle.

Par exemple:

```
Wald test for residence_since
in glm(formula = class ~ ., family = binomial(link = "logit"), data = credit1)
F = 0.01972717 on 1 and 952 df: p= 0.88833
```

Donc `residence_since` a une P-value très élevé et son coefficient risque d'être zéro. Et il faut l'éliminer.

Question 5

Et pour trouver un modèle parcimonieux on doit sélectionner les variables significatives et qui contribuent le plus au modèle. D'après le Test de Wald, il est prudent d'éliminer certaines variables comme :

`employment - age residence_since - property - housing - nb_credits- job - liable - telephone`

Anova

		Df	Deviance	Resid.	Df	Resid.	Dev	Pr(>Chi)
Si on exécute la fonction anova(modReg, test="Chisq")	NULL			999		1221.73		
	status_account	3	131.336	996	1090.39	< 2.2e-16	***	
	duration	1	38.497	995	1051.90	5.485e-10	***	
	history	4	29.311	991	1022.58	6.759e-06	***	
	purpose	9	33.509	982	989.08	0.0001089	***	
	amount	1	1.504	981	987.57	0.2200237		
	savings	4	19.068	977	968.50	0.0007623	***	
	employment	4	12.496	973	956.01	0.0140190	*	
	installement_rate	1	11.907	972	944.10	0.0005594	***	
	status_personnal	3	9.459	969	934.64	0.0237759	*	
	other_debtors	2	8.137	967	926.51	0.0171062	*	
	residence_since	1	0.155	966	926.35	0.6934956		
	property	3	2.520	963	923.83	0.4717369		
	age	1	3.725	962	920.11	0.0536071	.	
	other_installment	2	8.357	960	911.75	0.0153184	*	
	housing	2	3.517	958	908.23	0.1723184		
	nb_credits	1	2.328	957	905.90	0.1270420		
	job	3	1.110	954	904.79	0.7745757		
	liable	1	1.049	953	903.74	0.3057414		
	telephone	1	1.863	952	901.88	0.1722319		

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

En analysant le tableau, nous pouvons voir la baisse de l'écart lors de l'ajout de chaque variable une à la fois. Et en ajoutant :

- `status_account, duration, history, purpose, savings, employment, installement_rate, status_personnal, other_debtors, other_installment`

Réduit considérablement la déviance résiduelle et les autres variables semblent moins importantes.

Pour trouver un bon modèle prédictif on doit diviser le jeu de donnée sous Train et Test :

- Train 80% de données
- Test 20% de données

Car il ne faut pas avoir un modèle parfait que pour ces données (Overfitting) on a besoin d'être un peu plus sûr que ce modèle va marcher pour des nouvelles données et c'est pourquoi on teste et on peut utiliser l'accuracy (la qualité ou l'état d'être correct) comme un critère d'un bon modèle.

Nous avons créé une fonction qui s'appelle « `Trying_all` » et qui prend en entrée les données, la « link function » et le nombre d'itération et qui return data.frame de 16 colonnes et ces colonnes sont regroupées en 4 techniques dans lesquelles on va trouver le bon modèle final.

Et chaque technique possède 4 critères ($\frac{16 \text{ colonnes}}{4 \text{ critères}} = 4 \text{ techniques}$).

De plus il y a 100 lignes, car nous avons initialisé d'une manière aléatoire 100 fois Train data et Test Data pour avoir une vision globale et modéliser sur Test Data et tester toutes les variables. De plus on a implémenté 3 méthodes de « link function : logit, probit , cloglog »

En bref :

Nous avons utilisé trois méthodes de « link function » :

- link function de logit
- link function de probit
- link function de cloglog

Et dans chaque méthode il y a quatre techniques pour modéliser et les techniques sont :

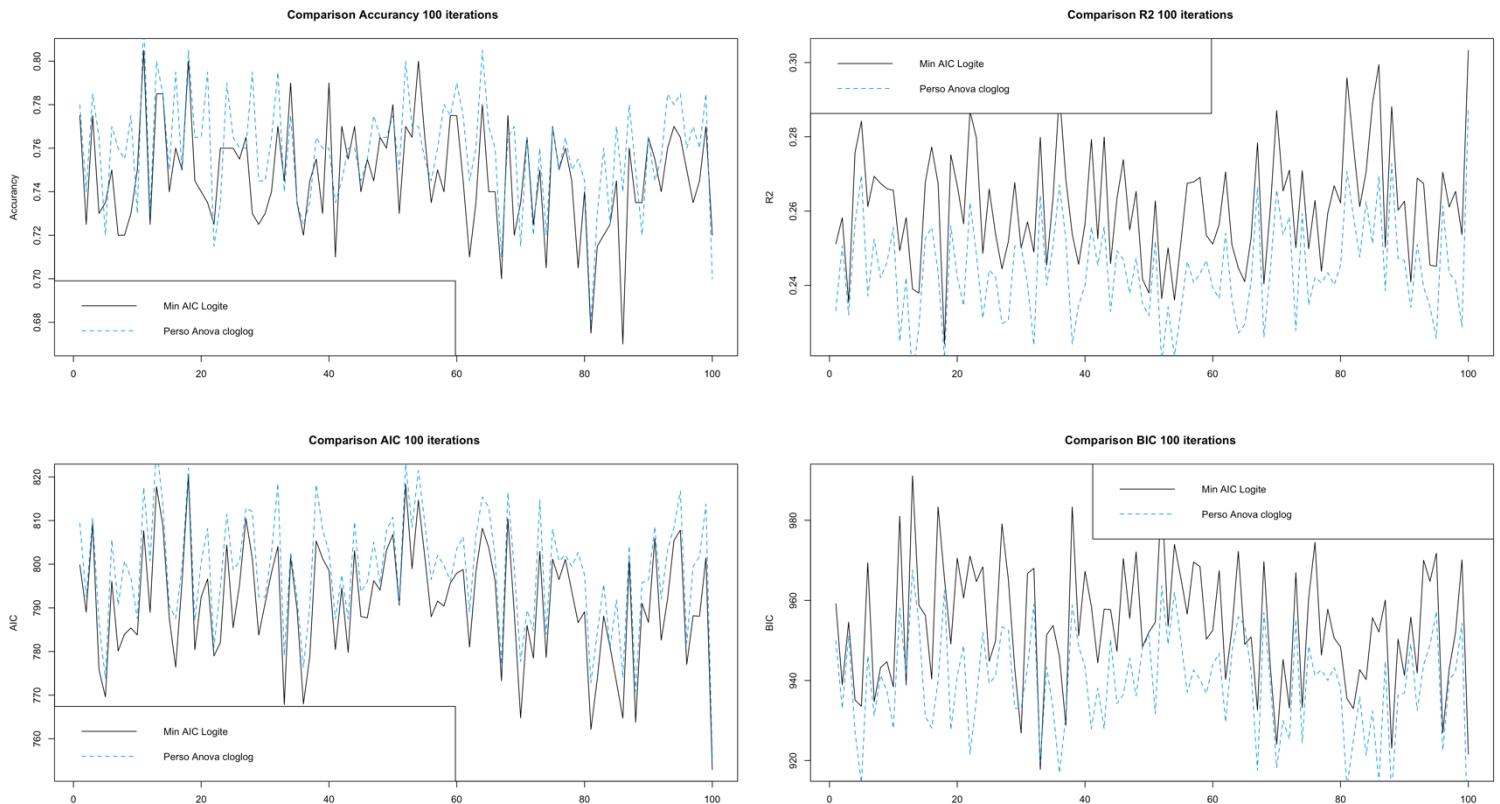
- 1- Toutes les variables d'une manière générale
- 2- Stepwise pour minimiser l'AIC avec une direction forward and backward sélection
- 3- Stepwise pour minimiser la BIC avec une direction forward and backward sélection
- 4- Méthode personnelle élaborée grâce à Anova pour sélectionner toutes les variables qui ont une Déviance la plus élevée et qui sont significatives d'après Wald test.

Pour comparer on prend en compte 4 critères :

- AIC
- Pseudo R2
- Accuracy
- BIC

Mais avant de comparer nous nous sommes inspirés de la simulation de monte Carlo. Donc nous avons décidé de simuler 100 fois d'une manière aléatoire les 80% de données qui permettent d'établir le modèle et utiliser 20% restantes pour prédire et tester. Pour modéliser 100 fois les 4 techniques (chaque fois avec des données différentes d'une manière aléatoire) la sortie de la fonction est les 4 critères * 4 techniques = 16 colonnes.

Ensuite on prend la Moyenne, la Quantile "97.5%" et la Quantile "2.5%" de ces critères : d'AIC, de Pseudo R2, d'Accuracy et de BIC pour chaque technique de modélisation et pour les 3 méthodes de link functions de 100 simulations pour comparer et choisir la méthode et la technique idéale pour sélectionner les variables .



Après avoir analysé et observé par rapport à la Moyenne, la Quantile "97.5%" et la Quantile "2.5%" on trouve deux techniques intéressantes : celle de stepwise-min-AIC avec link function de type « logit » et celle d'anova-perso avec un link function de type « cloglog »

Step-min-AIC logit	AIC	Pseudo R2	Accuracy	BIC
Moyenne	791.3831	0.2612	0.7472	953.6112
Quantile "97.5%"	816.3110	0.2941	0.7952	983.3337
Quantile "2.5%"	764.2374	0.2362	0.7024	923.5422

Anova-perso- cloglog	AIC	Pseudo R2	Accuracy	BIC
Moyenne	798.7325	0.2442	0.7590	939.2708
Quantile "97.5%"	821.7801	0.2702	0.8026	962.3185
Quantile "2.5%"	773.3210	0.2206	0.7124	913.8594

On remarque l'AIC et Pseudo R2 sont un peu meilleurs pour le modèle Step-min-AIC logit que le modèle Anova-perso- cloglog et inversement pour l' Accuracy et BIC. Mais comme on ne cherche pas forcement que le modèle soit parfait uniquement pour ces données on va donner l'Accuracy la priorité maximale puis Pseudo R2 puis AIC et BIC. Donc on va choisir le modèle Anova-perso qui a une link fonction de type cloglog et ses variables sont les suivantes :

- status_account
- duration
- history
- purpose
- savings
- installement_rate

- `tatus_personnal`
- `other_debtors`
- `other_installment`

Et toutes ces variables sont significatives et contribuent le plus au modèle.

Question 6

Y appartient à Class

On sait que chaque $Y \sim Ber, 0 \leq p \leq 1$ si :

$$Y = \begin{cases} 1, & \text{avec probability } p, \\ 0, & \text{avec probability } 1 - p, \end{cases}$$

Donc si $P[Y = 1] = 1$ et $P[Y = 0] = 1 - p$,

Et $P[Y = y] = p^y(1 - p)^{1-y}, y = 0, 1$

Mais class est de taille $n = 1000$ donc on a $class \sim Bi(n, p)$ (en ajoutant n Indépendants $Ber(p)$ donc $Ber(p)$ est la même que $Bi(1, p)$)

On sait que Y est complètement déterminé par p donc on a :

- $E[Y] = p \cdot 1 + (1 - p) \cdot 0 = p$
- $V[Y] = p(1 - p)$

Et en régression logistique on a :

$$p(x) = P[Y = 1 | X = x] = E[Y | X = x]$$

$$E[p(\text{tout } x)] = E[\text{E}[Class = 0 | X = x]] = \frac{700}{1000} = 0.7$$

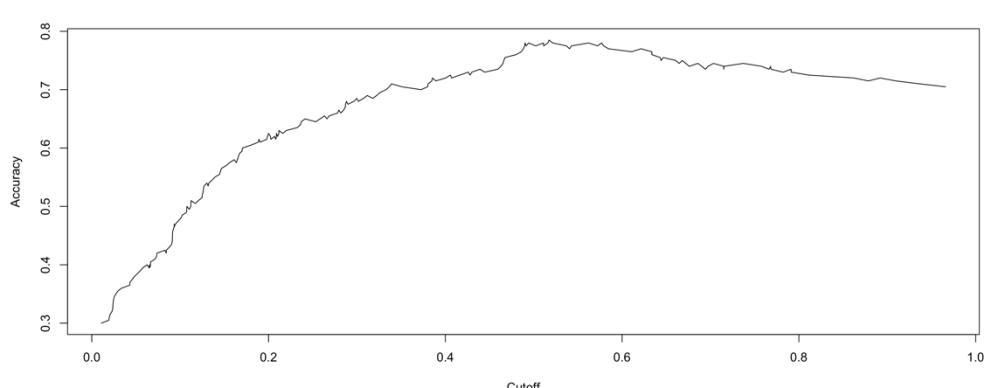
Trouver la meilleure Cutoff pour améliorer la prédiction :

Lors du développement de modèles de prédiction, la mesure la plus critique concerne l'efficacité du modèle dans la prédiction de la variable cible sur les observations hors échantillon. Le processus implique l'utilisation des estimations du modèle pour prédire les valeurs de l'ensemble d'apprentissage normalisé.

```
> print(c(Accuracy = acc, Cutoff = cut))
```

Accuracy Cutoff.476

0.7850000 0.5173151



Annexe : sorties R

Partie 2

Question 1 :

```
> #Question 1
> #-----
> #on donne un nom pour chaque variable
> noms_variables <- c("status_account","duration","history","purpose",
+ "amount","savings","employment","installement_rate",
+ "status_personnal","other_debtors","residence_since",
+ "property","age","other_installment","housing",
+ "nb_credits","job","liable","telephone","foreign","class")
> credit0 <- read.table("GermanCredit.txt", col.names = noms_variables)
> credit0$class <- as.factor(credit0$class-1) # rendre les valures de class à 0, 1
> credit <- subset(credit0, select=-c(foreign))#supprimer la variable foreign
> any(!is.na(credit))# pas de manque de données
[1] TRUE
> #changer de char à factor pour les variables qualitatives
> credit1<- lapply(credit, function(x) if(class(x) == "character") as.factor(x) else x)
> credit1 <- data.frame(credit1) # rendre en data.frame
> head(credit1,4) #pour avoir une idée
  status_account duration history purpose amount savings employment installement_rate status_personnal
1          A11       6     A34     A43    1169     A65       A75                 4             A93
2          A12      48     A32     A43    5951     A61       A73                 2             A92
3          A14      12     A34     A46   2096     A61       A74                 2             A93
4          A11      42     A32     A42   7882     A61       A74                 2             A93
  other_debtors residence_since property age other_installment housing nb_credits job liable
1          A101            4     A121   67           A143     A152      2 A173     1
2          A101            2     A121   22           A143     A152      1 A173     1
3          A101            3     A121   49           A143     A152      1 A172     2
4          A103            4     A122   45           A143     A153      1 A173     2
  telephone class
1        A192      0
2        A191      1
3        A191      0
4        A191      0
> str(credit1$class)
Factor w/ 2 levels "0","1": 1 2 1 1 2 1 1 1 1 2 ...
```

Question 2 :

```

> #Question 2
> # -----
> #Cette fonction va décrire les données et va faire une représentation graphiques
> descr <- function(data){
+   str(data)# bref des variables
+   print("Les fréquences marginales : ")
+   for (i in 1:dim(data)[2]){
+     if (class(data[,i]) == "factor"){ #pour les variables qualitatives
+       tab =table(data[,i])#fréquences
+       barplot(tab,col=1:length(tab),main=names(data)[i])
+
+       print(names(data)[i]) # nom de variable
+       print(prop.table(tab)) #fréquences marginales
+       print("")
+     }
+     else { #pour les variables quantitatives
+       boxplot(data[,i],main=names(data)[i])
+     }
+   }
+   print("")
+   summary(data) #bref statistique
+
+ }
> descr(credit1) # fonction pour décrire le jeu de données
'data.frame': 1000 obs. of 20 variables:
 $ status_account : Factor w/ 4 levels "A11","A12","A13",..: 1 2 4 1 1 4 4 2 4 2 ...
 $ duration       : int 6 48 12 42 24 36 24 36 12 30 ...
 $ history        : Factor w/ 5 levels "A30","A31","A32",..: 5 3 5 3 4 3 3 3 3 5 ...
 $ purpose        : Factor w/ 10 levels "A40","A41","A410",..: 5 5 8 4 1 8 4 2 5 1 ...
 $ amount         : int 1169 5951 2096 7882 4870 9055 2835 6948 3059 5234 ...
 $ savings        : Factor w/ 5 levels "A61","A62","A63",..: 5 1 1 1 5 3 1 4 1 ...
 $ employment     : Factor w/ 5 levels "A71","A72","A73",..: 5 3 4 4 3 3 5 3 4 1 ...
 $ installement_rate: int 4 2 2 2 3 2 3 2 2 4 ...
 $ status_personnal : Factor w/ 4 levels "A91","A92","A93",..: 3 2 3 3 3 3 3 3 1 4 ...
 $ other_debtors   : Factor w/ 3 levels "A101","A102",..: 1 1 1 3 1 1 1 1 1 ...
 $ residence_since : int 4 2 3 4 4 4 4 2 4 2 ...
 $ property       : Factor w/ 4 levels "A121","A122",..: 1 1 1 2 4 4 2 3 1 3 ...
 $ age            : int 67 22 49 45 53 35 53 35 61 28 ...
 $ other_installment: Factor w/ 3 levels "A141","A142",..: 3 3 3 3 3 3 3 3 3 3 ...
 $ housing         : Factor w/ 3 levels "A151","A152",..: 2 2 2 3 3 3 2 1 2 2 ...
 $ nb_credits     : int 2 1 1 1 2 1 1 1 1 2 ...
 $ job             : Factor w/ 4 levels "A171","A172",..: 3 3 2 3 3 2 3 4 2 4 ...
 $ liable          : int 1 1 2 2 2 2 1 1 1 1 ...

```

```
$ telephone      : Factor w/ 2 levels "A191","A192": 2 1 1 1 1 2 1 2 1 1 ...
$ class         : Factor w/ 2 levels "0","1": 1 2 1 1 2 1 1 1 2 ...
[1] "Les fréquences marginales : "
[1] "status_account"

  A11   A12   A13   A14
0.274 0.269 0.063 0.394
[1] ""
[1] "history"

  A30   A31   A32   A33   A34
0.040 0.049 0.530 0.088 0.293
[1] ""
[1] "purpose"

  A40   A41   A410  A42   A43   A44   A45   A46   A48   A49
0.234 0.103 0.012 0.181 0.280 0.012 0.022 0.050 0.009 0.097
[1] ""
[1] "savings"

  A61   A62   A63   A64   A65
0.603 0.103 0.063 0.048 0.183
[1] ""
[1] "employment"

  A71   A72   A73   A74   A75
0.062 0.172 0.339 0.174 0.253
[1] ""
[1] "status_personnal"

  A91   A92   A93   A94
0.050 0.310 0.548 0.092
[1] ""
[1] "other_debtors"

  A101  A102  A103
0.907 0.041 0.052
[1] ""
[1] "property"

  A121  A122  A123  A124
0.282 0.232 0.332 0.154
[1] ""
[1] "other_installment"

  A141  A142  A143
```

```

[1] ""
[1] "housing"

A151 A152 A153
0.179 0.713 0.108
[1] ""
[1] "job"

A171 A172 A173 A174
0.022 0.200 0.630 0.148
[1] ""
[1] "telephone"

A191 A192
0.596 0.404
[1] ""
[1] "class"

0 1
0.7 0.3
[1] ""
[1] ""

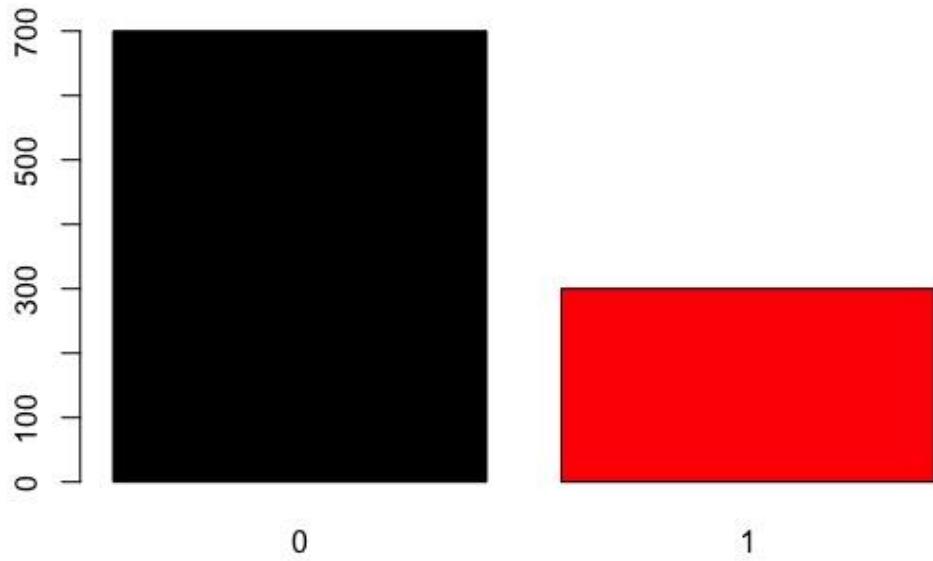
status_account duration history purpose amount savings employment
A11:274 Min. : 4.0 A30: 40 A43 :280 Min. : 250 A61:603 A71: 62
A12:269 1st Qu.:12.0 A31: 49 A40 :234 1st Qu.: 1366 A62:103 A72:172
A13: 63 Median :18.0 A32:530 A42 :181 Median : 2320 A63: 63 A73:339
A14:394 Mean :20.9 A33: 88 A41 :103 Mean : 3271 A64: 48 A74:174
3rd Qu.:24.0 A34:293 A49 : 97 3rd Qu.: 3972 A65:183 A75:253
Max. :72.0 A46 : 50 Max. :18424
(Other): 55

installement_rate status_personnal other_debtors residence_since property age
Min. :1.000 A91: 50 A101:907 Min. :1.000 A121:282 Min. :19.00
1st Qu.:2.000 A92:310 A102: 41 1st Qu.:2.000 A122:232 1st Qu.:27.00
Median :3.000 A93:548 A103: 52 Median :3.000 A123:332 Median :33.00
Mean :2.973 A94: 92 Mean :2.845 A124:154 Mean :35.55
3rd Qu.:4.000 Mean :3.845 3rd Qu.:4.000 3rd Qu.:42.00
Max. :4.000 Max. :4.000 Max. :75.00

other_installment housing nb_credits job liable telephone class
A141:139 A151:179 Min. :1.000 A171: 22 Min. :1.000 A191:596 0:700
A142: 47 A152:713 1st Qu.:1.000 A172:200 1st Qu.:1.000 A192:404 1:300
A143:814 A153:108 Median :1.000 A173:630 Median :1.000
Mean :1.407 A174:148 Mean :1.155
3rd Qu.:2.000 3rd Qu.:1.000
Max. :4.000 Max. :2.000

```

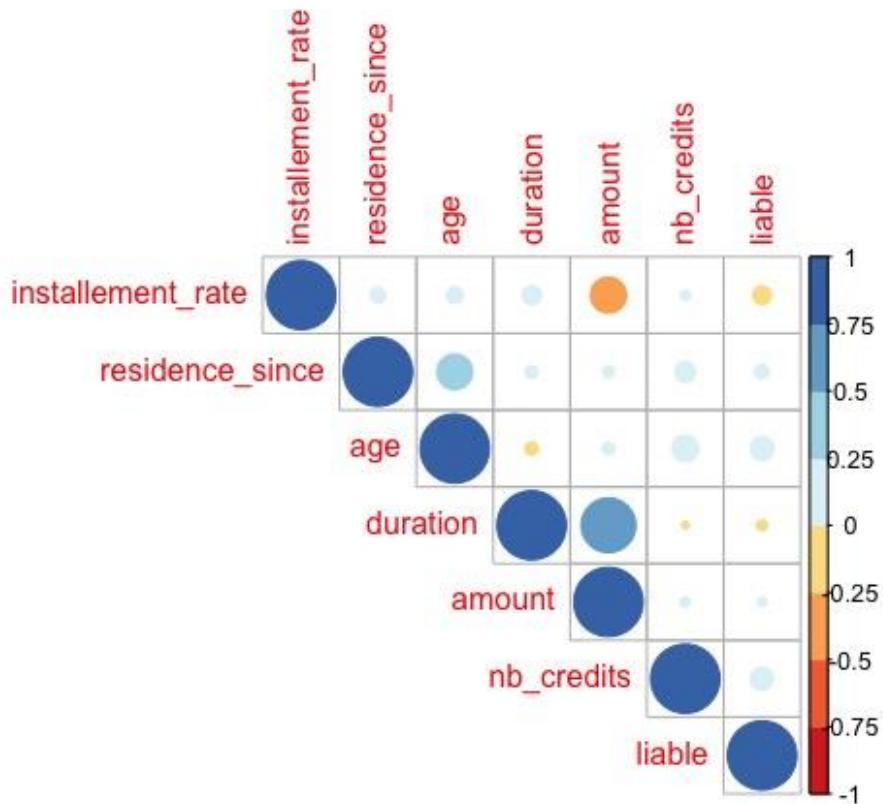
class



```
> cov(qunt)#covaraince et varaince
```

	duration	amount	installement_rate	residence_since	age	nb_credits	liable
duration	145.4150060	21273.74978	1.00838939	0.45341842	-4.9569950	-0.07859960	-0.10406907
amount	21273.7497758	7967843.47091	-856.77080480	90.12011011	1050.5226547	33.906900090	17.52053053
installement_rate	1.0083894	-856.77080	1.25152252	0.06087588	0.7414835	0.01400300	-0.02884384
residence_since	0.4534184	90.12011	0.06087588	1.21819319	3.3449750	0.05714214	0.01704204
age	-4.9569950	1050.52265	0.74148348	3.34497497	129.4012853	0.98075876	0.48685686
nb_credits	-0.0785996	33.90690	0.01400300	0.05714214	0.9807588	0.33368468	0.02293794
liable	-0.1040691	17.52053	-0.02884384	0.01704204	0.4868569	0.02293794	0.13110611

```
corrplot(cov(qunt), type="upper", order="hclust",
         col=brewer.pal(n=8, name="RdYlBu"))
```



Question 3 et question 4 :

```

> #question 3
> #-----
> modReg <- (glm(class ~ ., data = credit1, family = binomial(link='logit')))
> #question 4
> #-----
> summary(modReg)

Call:
glm(formula = class ~ ., family = binomial(link = "logit"), data = credit1)

Deviance Residuals:
    Min      1Q   Median      3Q      Max 
-2.3324 -0.7020 -0.3783  0.7198  2.6073 

Coefficients:
            Estimate Std. Error z value Pr(>|z|)    
(Intercept) 1.688e-01 1.076e+00  0.157 0.875351  
status_accountA12 -3.238e-01 2.157e-01 -1.501 0.133320  
status_accountA13 -9.252e-01 3.677e-01 -2.516 0.011854 *  
status_accountA14 -1.685e+00 2.311e-01 -7.291 3.07e-13 *** 
duration          2.965e-02 9.272e-03  3.198 0.001384 **  
historyA31        2.212e-01 5.443e-01  0.406 0.684375  
historyA32        -5.121e-01 4.276e-01 -1.198 0.231008  
historyA33        -7.838e-01 4.696e-01 -1.669 0.095050 .  
historyA34        -1.410e+00 4.378e-01 -3.220 0.001281 **  
purposeA41         -1.635e+00 3.749e-01 -4.361 1.30e-05 *** 
purposeA410        -1.534e+00 8.005e-01 -1.917 0.055280 .  
purposeA42         -7.355e-01 2.587e-01 -2.843 0.004467 **  
purposeA43         -8.410e-01 2.457e-01 -3.424 0.000618 *** 
purposeA44         -4.477e-01 7.618e-01 -0.588 0.556760  
purposeA45         -1.692e-01 5.453e-01 -0.310 0.756392  
purposeA46         9.472e-02 3.957e-01  0.239 0.810804  
purposeA48         -1.951e+00 1.207e+00 -1.616 0.106089  
purposeA49         -6.829e-01 3.321e-01 -2.056 0.039764 *  
amount            1.232e-04 4.446e-05  2.771 0.005581 **  
savingsA62        -3.684e-01 2.847e-01 -1.294 0.195679  
savingsA63        -3.851e-01 4.008e-01 -0.961 0.336560  
savingsA64        -1.330e+00 5.237e-01 -2.540 0.011100 *  
savingsA65        -9.662e-01 2.601e-01 -3.715 0.000203 *** 
employmentA72     -1.246e-01 4.253e-01 -0.293 0.769543  
employmentA73     -2.290e-01 4.096e-01 -0.559 0.576113  
employmentA74     -8.664e-01 4.444e-01 -1.950 0.051226 .  
employmentA75     -2.975e-01 4.130e-01 -0.720 0.471307  
installement_rate 3.347e-01 8.784e-02  3.810 0.000139 *** 
status_personnalA92 -2.553e-01 3.856e-01 -0.662 0.507904  
status_personnalA92 -2.553e-01 3.856e-01 -0.662 0.507904  
status_personnalA93 -8.150e-01 3.791e-01 -2.150 0.031563 *  
status_personnalA94 -3.857e-01 4.530e-01 -0.852 0.394473  
other_debtorsA102  3.766e-01 4.032e-01  0.934 0.350283  
other_debtorsA103 -1.057e+00 4.251e-01 -2.487 0.012874 *  
residence_since   1.207e-02 8.592e-02  0.140 0.888302  
propertyA122      2.713e-01 2.523e-01  1.075 0.282192  
propertyA123      2.195e-01 2.350e-01  0.934 0.350367  
propertyA124      7.121e-01 4.205e-01  1.694 0.090348 .  
age                -1.488e-02 9.216e-03 -1.614 0.106506  
other_installmentA142 -9.305e-02 4.120e-01 -0.226 0.821318  
other_installmentA143 -6.242e-01 2.389e-01 -2.613 0.008966 ** 
housingA152        -4.143e-01 2.319e-01 -1.786 0.074051 .  
housingA153        -6.097e-01 4.714e-01 -1.294 0.195819  
nb_credits         2.925e-01 1.894e-01  1.544 0.122558  
jobA172            5.076e-01 6.721e-01  0.755 0.450119  
jobA173            5.550e-01 6.482e-01  0.856 0.391896  
jobA174            4.634e-01 6.570e-01  0.705 0.480549  
liable             2.550e-01 2.481e-01  1.028 0.304128  
telephoneA192     -2.730e-01 2.007e-01 -1.360 0.173840  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1221.73 on 999 degrees of freedom
 Residual deviance: 901.88 on 952 degrees of freedom
 AIC: 997.88

Number of Fisher Scoring iterations: 5

```

> #Ensuite on veut savoir si les variables sont représentatives
> for(g in 1:dim(credit1)[2]){
+   if (class(credit1[,g])=="factor"){
+     print(names(credit1[g]))
+     print(xtabs(~class+ credit1[,g],data=credit1)) #représentative ?
+     print(''))
+   }
[1] "status_account"
  credit1[, g]
class A11 A12 A13 A14
  0 139 164 49 348
  1 135 105 14 46
[1] ""
[1] "history"
  credit1[, g]
class A30 A31 A32 A33 A34
  0 15 21 361 60 243
  1 25 28 169 28 50
[1] ""
[1] "purpose"
  credit1[, g]
class A40 A41 A410 A42 A43 A44 A45 A46 A48 A49
  0 145 86 7 123 218 8 14 28 8 63
  1 89 17 5 58 62 4 8 22 1 34
[1] ""
[1] "savings"
  credit1[, g]
class A61 A62 A63 A64 A65
  0 386 69 52 42 151
  1 217 34 11 6 32
[1] ""
[1] "employment"
  credit1[, g]
class A71 A72 A73 A74 A75
  0 39 102 235 135 189
  1 23 70 104 39 64
[1] ""
[1] "status_personnal"
  credit1[, g]
class A91 A92 A93 A94
  0 30 201 402 67
  1 20 109 146 25
[1] ""
[1] "other_debtors"
  credit1[, g]
class A101 A102 A103

```

```

1 20 109 140 65
[1] ""
[1] "other_debtors"
credit1[, g]
class A101 A102 A103
0 635 23 42
1 272 18 10
[1] ""
[1] "property"
credit1[, g]
class A121 A122 A123 A124
0 222 161 230 87
1 60 71 102 67
[1] ""
[1] "other_installment"
credit1[, g]
class A141 A142 A143
0 82 28 590
1 57 19 224
[1] ""
[1] "housing"
credit1[, g]
class A151 A152 A153
0 109 527 64
1 70 186 44
[1] ""
[1] "job"
credit1[, g]
class A171 A172 A173 A174
0 15 144 444 97
1 7 56 186 51
[1] ""
[1] "telephone"
credit1[, g]
class A191 A192
0 409 291
1 187 113
[1] ""
[1] "class"
credit1[, g]
class 0 1
0 700 0
1 0 300
[1] ""

```

```
> # wald test
> t=1
> while(t < 20){
+ print(regTermTest(modReg,names(credit1)[t])); t=t+1 ; print('')}
Wald test for status_account
in glm(formula = class ~ ., family = binomial(link = "logit"), data = credit1)
F = 19.61969 on 3 and 952 df: p= 2.3977e-12
[1] ""
Wald test for duration
in glm(formula = class ~ ., family = binomial(link = "logit"), data = credit1)
F = 10.22727 on 1 and 952 df: p= 0.0014295
[1] ""
Wald test for history
in glm(formula = class ~ ., family = binomial(link = "logit"), data = credit1)
F = 5.428025 on 4 and 952 df: p= 0.00025306
[1] ""
Wald test for purpose
in glm(formula = class ~ ., family = binomial(link = "logit"), data = credit1)
F = 3.491133 on 9 and 952 df: p= 0.000293
[1] ""
Wald test for amount
in glm(formula = class ~ ., family = binomial(link = "logit"), data = credit1)
F = 7.680757 on 1 and 952 df: p= 0.0056901
[1] ""
Wald test for savings
in glm(formula = class ~ ., family = binomial(link = "logit"), data = credit1)
F = 4.737312 on 4 and 952 df: p= 0.00086705
[1] ""
Wald test for employment
in glm(formula = class ~ ., family = binomial(link = "logit"), data = credit1)
F = 1.946066 on 4 and 952 df: p= 0.10075
[1] ""
Wald test for installement_rate
in glm(formula = class ~ ., family = binomial(link = "logit"), data = credit1)
F = 14.51573 on 1 and 952 df: p= 0.00014791
[1] ""
Wald test for status_personnal
in glm(formula = class ~ ., family = binomial(link = "logit"), data = credit1)
F = 3.201179 on 3 and 952 df: p= 0.022697
[1] ""
Wald test for other_debtors
in glm(formula = class ~ ., family = binomial(link = "logit"), data = credit1)
F = 3.706437 on 2 and 952 df: p= 0.02492
[1] ""
Wald test for residence_since
in glm(formula = class ~ ., family = binomial(link = "logit"), data = credit1)
```

```

Wald test for residence_since
in glm(formula = class ~ ., family = binomial(link = "logit"), data = credit1)
F = 0.01972717 on 1 and 952 df: p= 0.88833
[1] ""
Wald test for property
in glm(formula = class ~ ., family = binomial(link = "logit"), data = credit1)
F = 1.052529 on 3 and 952 df: p= 0.36848
[1] ""
Wald test for age
in glm(formula = class ~ ., family = binomial(link = "logit"), data = credit1)
F = 2.605312 on 1 and 952 df: p= 0.10684
[1] ""
Wald test for other_installment
in glm(formula = class ~ ., family = binomial(link = "logit"), data = credit1)
F = 3.93012 on 2 and 952 df: p= 0.019961
[1] ""
Wald test for housing
in glm(formula = class ~ ., family = binomial(link = "logit"), data = credit1)
F = 1.857899 on 2 and 952 df: p= 0.15657
[1] ""
Wald test for nb_credits
in glm(formula = class ~ ., family = binomial(link = "logit"), data = credit1)
F = 2.384318 on 1 and 952 df: p= 0.12289
[1] ""
Wald test for job
in glm(formula = class ~ ., family = binomial(link = "logit"), data = credit1)
F = 0.2592461 on 3 and 952 df: p= 0.85476
[1] ""
Wald test for liable
in glm(formula = class ~ ., family = binomial(link = "logit"), data = credit1)
F = 1.056007 on 1 and 952 df: p= 0.30439
[1] ""
Wald test for telephone
in glm(formula = class ~ ., family = binomial(link = "logit"), data = credit1)
F = 1.849517 on 1 and 952 df: p= 0.17416
[1] ""

> # McFadden's Pseudo R^2
> ll.null <- modReg$null.deviance/-2
> ll.propse<- modReg$deviance/-2
> Pseudo_R2 <- (ll.null-ll.propse )/ll.null
> Pseudo_R2 # et donc tres faible juste 26% represent le modele
[1] 0.2617991

```

Question5 :

```
> anova(modReg, test="Chisq")
Analysis of Deviance Table
```

Model: binomial, link: logit

Response: class

Terms added sequentially (first to last)

	Df	Deviance	Resid. Df	Resid. Dev	Pr(>Chi)
NULL			999	1221.73	
status_account	3	131.336	996	1090.39	< 2.2e-16 ***
duration	1	38.497	995	1051.90	5.485e-10 ***
history	4	29.311	991	1022.58	6.759e-06 ***
purpose	9	33.509	982	989.08	0.0001089 ***
amount	1	1.504	981	987.57	0.2200237
savings	4	19.068	977	968.50	0.0007623 ***
employment	4	12.496	973	956.01	0.0140190 *
installement_rate	1	11.907	972	944.10	0.0005594 ***
status_personnal	3	9.459	969	934.64	0.0237759 *
other_debtors	2	8.137	967	926.51	0.0171062 *
residence_since	1	0.155	966	926.35	0.6934956
property	3	2.520	963	923.83	0.4717369
age	1	3.725	962	920.11	0.0536071 .
other_installment	2	8.357	960	911.75	0.0153184 *
housing	2	3.517	958	908.23	0.1723184
nb_credits	1	2.328	957	905.90	0.1270420
job	3	1.110	954	904.79	0.7745757
liable	1	1.049	953	903.74	0.3057414
telephone	1	1.863	952	901.88	0.1722319

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

```

> # Monto Carlo de logit
> data_sel.logit<-Trying_all(credit1,"logit",n.iter=100)#Attention ca dure 8 min
> data_sel.logit
   gen.AIC    gen.R2 gen.Acc  BIC_gen setp.AIC   setp.R2 setp.Acc BIC_setp AIC_BIC   R2_BIC Acc_BIC  BIC_BIC anova.AIC
1  810.2129 0.2692599  0.720 1035.074 796.3514 0.2547942  0.715 955.6282 829.4638 0.1697585  0.725 871.6253 805.9570
2  807.1395 0.2724044  0.760 1032.001 795.7646 0.2553946  0.775 955.0414 860.9229 0.1293864  0.760 884.3459 801.6895
3  817.9959 0.2612967  0.770 1042.857 801.8838 0.2450412  0.765 951.7914 841.9182 0.1529234  0.705 874.7105 814.3929
4  808.5933 0.2709169  0.730 1033.455 797.1279 0.2621848  0.745 975.1432 855.3422 0.1350962  0.735 878.7653 803.1813
5  831.8604 0.2471115  0.800 1056.722 818.6704 0.2319587  0.770 977.9472 868.6658 0.1214642  0.740 892.0889 829.7883
6  822.3000 0.2568930  0.775 1047.161 806.6055 0.2422565  0.770 961.1977 853.1509 0.1373382  0.730 876.5740 811.3635
7  800.9652 0.2787215  0.745 1025.827 788.5073 0.2669123  0.730 957.1534 824.6369 0.1767434  0.695 871.4830 802.2304
8  804.5058 0.2750990  0.745 1029.367 793.6828 0.2575245  0.745 952.9596 836.2046 0.1710469  0.750 897.1045 802.4473
9  797.9598 0.2817965  0.740 1022.821 784.3375 0.2773175  0.725 967.0373 845.1945 0.1454787  0.700 868.6175 797.3736
10 820.5533 0.2586802  0.760 1045.415 809.4625 0.2475186  0.775 982.7931 860.7144 0.1295997  0.765 884.1374 815.3206
11 822.4817 0.2567072  0.780 1047.343 807.9295 0.2429482  0.770 967.2063 853.5430 0.1410295  0.755 886.3353 820.7556
12 810.0689 0.2694072  0.740 1034.930 795.1927 0.2518871  0.750 945.1003 848.2505 0.1423520  0.745 871.6736 798.1210
13 800.6217 0.2790730  0.725 1025.483 784.4901 0.2669300  0.710 943.7669 821.8347 0.1796105  0.730 868.6808 787.5315
14 798.5963 0.2811452  0.750 1023.458 780.5363 0.2689290  0.758 935.1285 837.7269 0.1551654  0.705 865.8346 788.6956
15 809.8473 0.2696339  0.765 1034.709 797.4827 0.2515905  0.755 952.0749 835.3561 0.1637299  0.705 877.5176 802.7793
16 804.8163 0.2747813  0.770 1029.678 790.1347 0.2672936  0.765 963.4653 845.4014 0.1452670  0.745 868.8244 796.1544
17 806.0859 0.2734824  0.755 1030.947 792.8249 0.2543098  0.760 942.7324 829.6449 0.1695733  0.720 871.8064 795.9172
18 810.5235 0.2689421  0.765 1035.385 795.8956 0.2552606  0.760 955.1724 849.1584 0.1434693  0.715 877.2661 800.8954
19 807.5362 0.2719985  0.725 1032.398 795.4485 0.2598106  0.725 964.0945 856.7020 0.1337049  0.730 880.1251 800.3293
20 798.4337 0.2813117  0.745 1023.295 779.9044 0.2675293  0.720 929.8119 825.2138 0.1741069  0.675 867.3753 786.5778
21 803.7880 0.2758334  0.730 1028.649 786.7741 0.2605006  0.740 936.6816 809.0140 0.2009130  0.725 874.5986 793.2139
22 795.4726 0.2843413  0.740 1020.334 782.4651 0.2710481  0.740 946.4265 828.6778 0.1644238  0.710 856.7855 786.6119
23 812.1348 0.2672935  0.755 1036.996 796.2799 0.2528211  0.760 950.8721 849.1282 0.1435003  0.695 877.2359 804.7089
24 805.4475 0.2741355  0.730 1030.309 794.2354 0.2671905  0.730 976.9353 852.8813 0.1396603  0.755 880.9889 808.8886
25 830.0968 0.2489158  0.830 1054.958 818.8157 0.2359027  0.825 987.4617 859.5965 0.1307434  0.730 883.0196 826.2151
26 785.3971 0.2946499  0.725 1010.258 770.9383 0.2787491  0.695 925.5305 805.3635 0.1964628  0.715 852.2097 779.8160
27 797.6471 0.28221164 0.740 1022.508 783.9755 0.2674565  0.745 943.2523 819.5877 0.1839557  0.685 871.1184 805.7217
28 792.8451 0.2870295  0.735 1017.707 777.7951 0.2737799  0.715 937.0719 822.6400 0.1726477  0.710 855.4322 800.1151
29 811.4125 0.2680326  0.740 1036.274 795.3676 0.2496619  0.765 940.5906 850.3902 0.1401628  0.730 873.8133 797.8850
30 801.0328 0.2786524  0.730 1025.894 787.7015 0.2656905  0.725 951.6629 832.2079 0.1689972  0.730 879.0540 803.3277
31 803.0218 0.2766174  0.735 1027.883 789.3210 0.2640335  0.720 953.2824 848.4849 0.1462047  0.710 881.2772 802.4916
32 802.6574 0.2769902  0.755 1027.519 791.5415 0.2474377  0.765 922.7107 820.2159 0.1792205  0.740 862.3774 794.0821
33 815.2925 0.2640627  0.790 1040.154 799.9412 0.2531676  0.785 963.9026 858.1353 0.1322384  0.740 881.5584 808.2623
34 802.7146 0.2769317  0.755 1027.576 786.8001 0.2584277  0.760 932.0230 827.3571 0.1719139  0.695 869.5186 794.0441
35 777.9207 0.3022993  0.730 1002.782 768.6736 0.2892513  0.715 942.0042 837.8220 0.1530218  0.710 861.2450 775.3551
36 782.0174 0.2981078  0.710 1006.879 771.7304 0.2861238  0.700 945.0610 824.4013 0.1749382  0.700 866.5628 788.9886
37 805.7627 0.2738130  0.735 1030.624 791.3655 0.2558029  0.760 941.2731 839.6666 0.1511345  0.695 863.0897 796.8366
38 812.7339 0.2666805  0.745 1037.595 796.6413 0.2524513  0.750 951.2335 830.7210 0.1684723  0.730 872.8825 800.4726
39 810.3921 0.2690765  0.750 1035.253 795.2231 0.2559486  0.760 954.4999 851.1242 0.1394118  0.755 874.5472 799.9947
40 802.1080 0.2775523  0.740 1026.969 792.3120 0.2671122  0.730 970.3272 845.2874 0.1453837  0.715 868.7104 802.1227
41 811.1918 0.2682583  0.735 1036.053 798.2390 0.2487704  0.750 948.1465 857.7230 0.1326603  0.720 881.1460 803.3693
42 816.2741 0.2630584  0.745 1041.135 801.6051 0.2453263  0.750 951.5127 857.7181 0.1326652  0.760 881.1412 808.3045

```

```
-- 
20 0.2566089  0.750 927.1161
21 0.2498191  0.745 933.7523
22 0.2565740  0.710 927.1502
23 0.2380581  0.795 945.2473
24 0.2337818  0.745 949.4270
25 0.2160543  0.835 966.7534
26 0.2635271  0.710 920.3544
27 0.2370220  0.760 946.2600
28 0.2427583  0.765 940.6534
29 0.2450400  0.760 938.4233
30 0.2394713  0.750 943.8661
31 0.2403268  0.770 943.0300
32 0.2489309  0.750 934.6204
33 0.2344225  0.765 948.8007
34 0.2489697  0.745 934.5825
35 0.2680913  0.720 915.8934
36 0.2541422  0.745 929.5270
37 0.2461127  0.770 937.3749
38 0.2423925  0.750 941.0109
39 0.2428815  0.740 940.5330
40 0.2407043  0.750 942.6610
41 0.2394288  0.760 943.9076
42 0.2343794  0.770 948.8429
43 0.2279234  0.790 955.1528
44 0.2440635  0.740 939.3778
45 0.2720581  0.735 912.0163
46 0.2291050  0.760 953.9979
47 0.2429191  0.760 940.4963
48 0.2496626  0.735 933.9053
49 0.2473601  0.760 936.1557
50 0.2382133  0.775 945.0956
51 0.2526481  0.765 930.9873
52 0.2396880  0.770 943.6543
53 0.2437006  0.780 939.7325
54 0.2459879  0.785 937.4969
55 0.2375951  0.750 945.6998
56 0.2452463  0.775 938.2217
57 0.2514689  0.730 932.1399
58 0.2429584  0.745 940.4578
59 0.2490622  0.740 934.4921
60 0.2378890  0.780 945.4126
61 0.2412884  0.770 942.0900
62 0.2356270  0.770 947.6234
[ reached 'max' / getOption("max.print") -- omitted 38 rows ]
```

```

> # Monto Carlo de probit
> data_sel.probit<-Trying_all(credit1,"probit",n.iter=100)#Attention ca dure 8 min
There were 50 or more warnings (use warnings() to see the first 50)
> data_sel.probit
   gen.AIC  gen.R2 gen.Acc BIC_gen setp.AIC  setp.R2 setp.Acc BIC_setp AIC_BIC   R2_BIC Acc_BIC BIC_BIC anova.AIC
1 808.8538 0.2706505 0.725 1033.715 795.6517 0.2596026 0.715 964.2977 829.5031 0.1697183 0.725 871.6646 806.0328
2 805.5462 0.2740345 0.760 1030.408 794.2987 0.2568943 0.775 953.5755 861.3131 0.1289871 0.760 884.7362 801.4204
3 817.9301 0.2613640 0.775 1042.792 802.8009 0.2461492 0.780 957.3930 842.0479 0.1527906 0.705 874.8402 816.3677
4 808.8250 0.2706799 0.750 1033.686 798.5653 0.2607142 0.750 976.5806 855.7969 0.1346310 0.735 879.2199 805.1161
5 831.1188 0.2478702 0.805 1055.980 818.3453 0.2343377 0.790 982.3067 868.8158 0.1213108 0.740 892.2389 830.8863
6 823.4963 0.2556691 0.770 1048.358 807.9248 0.2388604 0.770 957.8324 853.4806 0.1370008 0.730 876.9037 812.8761
7 800.4330 0.2792661 0.750 1025.294 788.4572 0.2669636 0.735 957.1032 839.4056 0.1554941 0.695 872.1978 803.4763
8 805.1179 0.2744728 0.755 1029.979 794.2232 0.2569717 0.750 953.5000 845.2267 0.1577235 0.760 896.7574 804.1682
9 798.2410 0.2815088 0.740 1023.102 784.5357 0.2730222 0.720 957.8663 845.8198 0.1448389 0.700 869.2429 797.8348
10 819.7754 0.2594761 0.760 1044.637 808.9486 0.2480443 0.775 982.2792 860.6646 0.1296506 0.765 884.0877 815.6583
11 822.2265 0.2569682 0.770 1047.088 808.7582 0.2421003 0.770 968.0350 853.6174 0.1409534 0.730 886.4097 821.3384
12 809.0177 0.2704827 0.745 1033.879 794.8180 0.2522705 0.755 944.7256 847.9920 0.1426165 0.745 871.4150 799.1485
13 801.1166 0.2785667 0.720 1025.978 785.6102 0.2657839 0.720 944.8870 826.3674 0.1729266 0.730 868.5289 788.8097
14 798.8634 0.2808720 0.755 1023.725 781.1477 0.2683034 0.745 935.7399 833.9265 0.1611000 0.720 866.7188 789.4457
15 808.6694 0.2708390 0.775 1033.531 798.0522 0.2510078 0.760 952.6444 835.1931 0.1638966 0.705 877.3546 803.9618
16 804.9153 0.2746800 0.780 1029.777 791.0058 0.2704949 0.770 973.7056 845.7046 0.1449568 0.745 869.1277 797.9759
17 806.0259 0.2735437 0.750 1030.887 793.6558 0.2575494 0.770 952.9353 829.8659 0.1693471 0.720 872.0274 797.2204
18 811.0845 0.2683681 0.765 1035.946 796.3675 0.2547777 0.760 955.6443 849.3006 0.1433239 0.715 877.4982 801.9508
19 806.6074 0.2729488 0.725 1031.469 795.3198 0.2640347 0.735 973.3351 856.9387 0.1334627 0.730 880.3618 802.3773
20 797.4200 0.2823488 0.745 1022.281 779.3787 0.2701134 0.720 933.9709 833.6113 0.1614225 0.670 866.4036 787.8517
21 805.0194 0.2745735 0.725 1029.881 789.0344 0.2581880 0.740 938.9420 823.0967 0.1803655 0.745 874.6275 795.4762
22 797.1740 0.2826005 0.745 1022.035 784.3200 0.2691503 0.745 948.2814 829.8046 0.1632710 0.710 857.9123 790.1640
23 812.3748 0.2670479 0.765 1037.236 797.5943 0.2576151 0.780 966.2403 849.3402 0.1432834 0.695 877.4478 806.4224
24 805.6679 0.2739100 0.720 1030.529 794.4581 0.2669628 0.720 977.1579 850.5782 0.1440630 0.750 883.3705 811.1666
25 829.3351 0.2496952 0.835 1054.196 818.7725 0.2359468 0.830 987.4186 860.0696 0.1302594 0.730 883.4927 827.4855
26 784.8969 0.2951617 0.720 1009.758 771.3055 0.2845123 0.730 939.9515 805.2596 0.1965691 0.710 852.1057 780.1695
27 797.7807 0.2819797 0.735 1022.642 783.6093 0.2678312 0.730 942.8861 811.3581 0.1964684 0.700 872.2580 807.0382
28 792.3517 0.2875343 0.730 1017.213 776.1841 0.2754281 0.715 935.4609 822.2574 0.1730392 0.715 855.0496 801.3577
29 812.2352 0.2671908 0.760 1037.097 797.2645 0.2518137 0.740 951.8567 850.8197 0.1397233 0.730 874.2427 800.0883
30 801.7888 0.2777890 0.730 1026.650 788.2915 0.2671332 0.710 956.9375 832.2993 0.1689037 0.730 879.1454 805.7115
31 801.8331 0.2778336 0.725 1026.694 788.9660 0.2664431 0.720 957.6120 848.7609 0.1459223 0.720 881.5532 803.3511
32 802.4549 0.2771974 0.755 1027.316 791.9208 0.2470497 0.750 923.0899 819.8502 0.1795946 0.740 862.0117 795.2837
33 815.2008 0.2641566 0.790 1040.062 800.1239 0.2529806 0.780 964.0854 858.3360 0.1320331 0.740 881.7590 810.1371
34 801.1899 0.2784916 0.755 1026.051 785.9079 0.2593405 0.730 931.1308 827.4695 0.1717990 0.695 869.6310 795.1172
35 778.6272 0.3015765 0.725 1003.489 770.2362 0.2876526 0.710 943.5668 838.0353 0.1528035 0.710 861.4584 778.0911
36 783.1646 0.2969340 0.710 1008.026 772.7661 0.2850641 0.705 946.0967 823.8408 0.1755116 0.705 866.0023 790.7521
37 805.1068 0.2744842 0.735 1029.968 790.3013 0.2589381 0.745 944.8934 840.2772 0.1505098 0.705 863.7002 797.5826
38 813.1157 0.2662899 0.750 1037.977 797.7445 0.2513226 0.755 952.3367 830.8563 0.1683338 0.730 873.0179 802.6556
39 810.5876 0.2688765 0.755 1035.449 796.1750 0.2549746 0.755 955.4518 850.8798 0.1396619 0.755 874.3028 801.3416
40 802.4552 0.2771971 0.730 1027.317 793.4037 0.2659952 0.735 971.4190 845.6130 0.1450505 0.715 869.0361 803.2732

```

40	802.4552	0.2771971	0.730 1027.317 793.403 0.2659952	0.735 971.4190 845.6130 0.1450505	0.715 869.0361 803.2732
41	811.4551	0.2679890	0.730 1036.316 799.6153 0.2473622	0.755 949.5229 858.5477 0.1318165	0.720 881.9708 804.8381
42	815.5813	0.2637672	0.745 1040.443 802.2254 0.2487843	0.760 961.5022 857.6524 0.1327326	0.760 881.0754 809.5966
43	821.8123	0.2573921	0.775 1046.674 810.0487 0.2428262	0.780 974.0101 849.9827 0.1405797	0.760 873.4057 815.6028
44	802.3295	0.2773257	0.740 1027.191 791.1005 0.2683517	0.720 969.1158 827.3779 0.1718927	0.735 869.5394 801.6941
45	780.5540	0.2996050	0.710 1005.415 764.7713 0.2911976	0.710 933.4173 837.7973 0.1530471	0.735 861.2203 772.8804
46	818.3529	0.2609315	0.735 1043.214 806.7314 0.2482665	0.745 975.3775 849.1900 0.1413907	0.735 872.6131 815.5786
47	800.5351	0.2791616	0.730 1025.396 788.9607 0.2644574	0.745 952.8681 836.6312 0.1624253	0.765 878.7927 801.1412
48	799.9114	0.2797997	0.750 1024.773 780.1224 0.2713987	0.725 939.3992 820.4418 0.1789893	0.685 862.6033 794.5147
49	804.7939	0.2748043	0.750 1029.655 789.1103 0.2703880	0.740 967.1255 830.5901 0.1686062	0.775 872.7516 797.0779
50	821.7341	0.2574721	0.815 1046.595 802.6333 0.2463207	0.795 957.2255 861.3106 0.1289896	0.745 884.7337 805.5110
51	794.8645	0.2849634	0.760 1019.726 781.6807 0.2738969	0.740 950.3268 842.9404 0.1477850	0.720 866.3634 792.4976
52	807.6439	0.2718883	0.750 1032.505 791.3628 0.2598983	0.720 950.6396 831.2871 0.1678930	0.740 873.4486 804.6707
53	802.3444	0.2773105	0.775 1027.206 787.0340 0.2622809	0.745 941.6262 844.1774 0.1485656	0.750 872.2850 799.3606
54	803.3104	0.2763221	0.755 1028.172 790.3962 0.2608872	0.770 949.6730 838.8543 0.1519657	0.710 862.2773 797.5884
55	811.8704	0.2675641	0.745 1036.732 795.2059 0.2621050	0.735 968.5366 857.1659 0.1332303	0.745 880.5889 807.3905
56	792.8424	0.2870323	0.765 1017.704 778.7836 0.2686760	0.725 928.6912 832.8357 0.1622161	0.715 865.6280 799.3532
57	795.2105	0.2846094	0.735 1020.072 780.7968 0.2727551	0.735 944.7582 834.9455 0.1621037	0.720 872.4224 793.0879
58	800.0444	0.2796636	0.740 1024.906 785.8351 0.2655538	0.745 945.1119 836.2190 0.1587544	0.665 869.0113 800.1767
59	793.2755	0.2865892	0.750 1018.137 783.0101 0.2704905	0.725 946.9715 841.4716 0.1492878	0.715 864.8946 793.5747
60	802.5638	0.2770860	0.730 1027.425 791.5466 0.2719878	0.735 978.9311 855.4351 0.1350011	0.740 878.8582 805.2325
61	803.8091	0.2758119	0.750 1028.670 795.6559 0.2616446	0.745 968.9866 861.5332 0.1287619	0.755 884.9562 803.7870
62	814.2972	0.2650810	0.770 1039.159 798.8788 0.2522083	0.775 958.1556 861.7356 0.1285548	0.775 885.1587 807.8512
	anova.R2	anova.Acc	BIC_anova		
1	0.2367037	0.760	946.5711		
2	0.2414227	0.780	941.9588		
3	0.2261296	0.810	956.9061		
4	0.2376415	0.750	945.6545		
5	0.2112750	0.800	971.4246		
6	0.2297020	0.780	953.4144		
7	0.2393193	0.765	944.0146		
8	0.2386114	0.765	944.7065		
9	0.2450913	0.725	938.3732		
10	0.2268554	0.780	956.1966		
11	0.2210438	0.795	961.8768		
12	0.2437472	0.745	939.6869		
13	0.2543253	0.725	929.3480		
14	0.2536745	0.760	929.9841		
15	0.2388225	0.780	944.5002		
16	0.2449470	0.750	938.5142		
17	0.2457200	0.765	937.7588		
18	0.2408801	0.760	942.4892		
19	0.2404437	0.740	942.9157		
20	0.2553055	0.750	928.3900		
21	0.2475045	0.740	936.0145		
22	0.2529396	0.720	930.7024		

20	0.2553055	0.750	928.3900
21	0.2475045	0.740	936.0145
22	0.2529396	0.720	930.7024
23	0.2363050	0.805	946.9608
24	0.2314511	0.745	951.7049
25	0.2147545	0.835	968.0239
26	0.2631654	0.710	920.7079
27	0.2356749	0.760	947.5766
28	0.2414869	0.765	941.8961
29	0.2427857	0.755	940.6267
30	0.2370324	0.750	946.2498
31	0.2394473	0.775	943.8895
32	0.2477015	0.745	935.8220
33	0.2325044	0.760	950.6754
34	0.2478718	0.740	935.6556
35	0.2652920	0.725	918.6294
36	0.2523380	0.745	931.2904
37	0.2453494	0.760	938.1209
38	0.2401590	0.760	943.1939
39	0.2415034	0.745	941.8800
40	0.2395271	0.745	943.8115
41	0.2379260	0.770	945.3765
42	0.2330574	0.785	950.1350
43	0.2269122	0.785	956.1411
44	0.2411427	0.735	942.2325
45	0.2706232	0.730	913.4187
46	0.2269370	0.755	956.1169
47	0.2417084	0.765	941.6795
48	0.2484882	0.730	935.0531
49	0.2458658	0.755	937.6162
50	0.2372376	0.780	946.0493
51	0.2505520	0.765	933.0360
52	0.2380972	0.765	945.2091
53	0.2435302	0.775	939.8990
54	0.2453434	0.785	938.1267
55	0.2353145	0.750	947.9289
56	0.2435378	0.785	939.8915
57	0.2499481	0.730	933.6262
58	0.2426953	0.750	940.7150
59	0.2494500	0.755	934.1131
60	0.2375225	0.780	945.7708
61	0.2390014	0.770	944.3254
62	0.2348431	0.775	948.3896

[reached 'max' / getOption("max.print") -- omitted 38 rows]

```

> # Monto Carlo de cloglog
> data_sel.cloglog<-Trying_all(credit1,"cloglog",n.iter=100)#Attention ca dure 8 min
There were 50 or more warnings (use warnings() to see the first 50)
> data_sel.cloglog
   gen.AIC    gen.R2 gen.Acc  BIC_gen setup.AIC  setup.R2 setup.Acc BIC_setup AIC_BIC    R2_BIC Acc_BIC  BIC_BIC anova.AIC
1 810.8012 0.2686580  0.720 1035.663 797.2401 0.2559312  0.710 961.2015 830.6161 0.1685795  0.725 872.7776 805.2843
2 808.9488 0.2705533  0.775 1033.810 795.7711 0.2594805  0.780 964.4171 860.1699 0.1301567  0.760 883.5930 800.5981
3 816.8254 0.2624944  0.765 1041.687 802.1283 0.2468373  0.755 956.7205 841.6169 0.1532316  0.700 874.4092 813.5656
4 807.8724 0.2716545  0.755 1032.734 796.2048 0.2651756  0.750 978.9047 854.4580 0.1360009  0.735 877.8810 802.8854
5 830.8698 0.2481249  0.800 1055.731 819.9471 0.2326988  0.790 983.9085 868.3054 0.1218330  0.750 891.7285 828.6104
6 820.4918 0.2587431  0.790 1045.353 806.7155 0.2421440  0.775 961.3077 852.5138 0.1379900  0.720 875.9368 810.3054
7 796.3856 0.2834071  0.735 1021.247 785.6673 0.2698181  0.730 954.3133 822.8790 0.1785420  0.680 869.7251 800.3282
8 801.1663 0.2785158  0.730 1026.028 792.3911 0.2711238  0.750 979.7755 848.9236 0.1518947  0.750 895.7697 803.1244
9 799.5141 0.2802063  0.745 1024.375 786.9578 0.2746366  0.730 969.6576 843.2895 0.1474278  0.700 866.7125 798.0581
10 820.7801 0.2584481  0.760 1045.641 810.4840 0.2382882  0.770 965.0762 860.4521 0.1298681  0.765 883.8751 813.8636
11 821.0489 0.2581731  0.765 1045.910 809.0731 0.2479169  0.770 982.4038 857.4129 0.1350239  0.740 885.5205 820.5444
12 810.4652 0.2690017  0.750 1035.327 797.5879 0.2514828  0.765 952.1801 848.0159 0.1425920  0.745 871.4390 799.1281
13 797.1202 0.2826556  0.715 1021.982 783.2896 0.2681582  0.735 942.5664 838.6540 0.1542168  0.715 866.7616 786.3797
14 797.2725 0.2824998  0.750 1022.134 782.9392 0.2685168  0.760 942.2160 835.7329 0.1572055  0.700 863.8406 788.5243
15 811.6035 0.2678371  0.780 1036.465 799.7873 0.2492325  0.770 954.3795 852.9725 0.1375208  0.700 876.3955 803.2868
16 806.5552 0.2730022  0.760 1031.417 793.1244 0.2642347  0.750 966.4551 844.9244 0.1457551  0.745 868.3474 798.8789
17 800.9877 0.2786985  0.755 1025.849 790.0422 0.2592031  0.750 944.6344 830.0813 0.1691267  0.715 872.2428 792.5938
18 810.0991 0.2693763  0.770 1034.960 796.0893 0.2550623  0.765 955.3661 847.4439 0.1452235  0.715 875.5516 801.7732
19 805.2237 0.2743645  0.730 1030.085 792.3792 0.2609046  0.755 956.3406 854.8393 0.1356107  0.730 878.2624 797.0378
20 794.3336 0.2855066  0.745 1019.195 780.0947 0.2796123  0.735 958.1099 819.1343 0.1823735  0.705 865.9802 783.8332
21 800.3893 0.2793107  0.735 1025.251 782.8536 0.2645118  0.740 932.7611 799.6887 0.2125004  0.720 869.9578 789.0931
22 791.8882 0.2880086  0.735 1016.750 779.5300 0.2740511  0.700 943.4914 826.9223 0.1662200  0.700 855.0299 783.0633
23 810.5893 0.2688747  0.765 1035.451 794.7084 0.2544289  0.765 949.3006 843.2965 0.1515132  0.725 876.0888 802.6788
24 803.9315 0.2756866  0.735 1028.793 794.4275 0.2690403  0.725 981.8119 851.7344 0.1408337  0.755 879.8421 808.4078
25 829.0435 0.2499935  0.820 1053.905 819.3314 0.2374213  0.825 992.6620 858.4649 0.1319012  0.730 881.8879 826.5931
26 789.3972 0.2905573  0.715 1014.259 775.0356 0.2725106  0.700 924.9432 805.4060 0.1964193  0.715 852.2522 783.9920
27 796.1643 0.2836336  0.735 1021.026 784.3623 0.2752458  0.750 962.3776 819.8159 0.1837223  0.685 871.3466 805.7100
28 796.2580 0.2835377  0.755 1021.119 781.0035 0.2643584  0.715 926.2265 824.4564 0.1707893  0.715 857.2487 798.5883
29 811.5632 0.2678783  0.770 1036.425 795.1243 0.2560497  0.755 954.4011 849.7645 0.1408029  0.730 873.1876 797.1208
30 798.9585 0.2807747  0.720 1023.820 788.1068 0.2734610  0.730 970.8066 848.1080 0.1445440  0.730 876.2157 800.7881
31 801.7490 0.2779197  0.735 1026.610 790.7366 0.2707703  0.745 973.4365 846.8180 0.1479181  0.715 879.6103 800.7099
32 804.2085 0.2754032  0.765 1029.070 791.1307 0.2580894  0.765 945.7229 817.7435 0.1837963  0.730 864.5896 792.7965
33 813.0646 0.2663421  0.795 1037.926 796.6899 0.2483090  0.760 941.9129 856.8712 0.1335318  0.735 880.2943 802.6786
34 800.7049 0.2789879  0.755 1025.566 786.1500 0.2693242  0.750 954.7961 846.3092 0.1443382  0.690 869.7322 793.2276
35 777.6135 0.3026136  0.725 1002.475 769.7799 0.2840268  0.735 933.7413 838.5334 0.1522939  0.710 861.9565 774.8107
36 779.4870 0.3006968  0.715 1004.348 770.2508 0.2896839  0.720 948.2661 824.6032 0.1747316  0.710 866.7647 785.7114
37 804.1026 0.2755116  0.735 1028.964 793.7386 0.2636063  0.740 967.0692 838.6063 0.1522194  0.705 862.0293 798.4222
38 811.8480 0.2675869  0.755 1036.709 796.9259 0.2521601  0.745 951.5181 830.3633 0.1688382  0.740 872.5248 799.0941
39 811.3959 0.2680495  0.755 1036.257 798.0715 0.2509880  0.745 952.6637 851.1095 0.1394268  0.755 874.5326 801.2857
40 800.4551 0.2792435  0.745 1025.316 790.7632 0.2686969  0.730 968.7784 844.2374 0.1464579  0.715 867.6604 799.4305
41 806.2990 0.2732643  0.750 1031.160 795.4241 0.2536967  0.750 950.0163 856.1486 0.1342711  0.720 879.5716 801.6971

```

40	800.4551	0.2792435	0.745	1025.316	790.7632	0.2686969	0.730	968.7784	844.2374	0.1464579	0.715	867.6604	799.4305
41	806.2990	0.2732643	0.750	1031.160	795.4241	0.2536967	0.750	950.0163	856.1486	0.1342711	0.720	879.5716	801.6971
42	816.6400	0.2626841	0.745	1041.501	804.2395	0.2528624	0.750	977.5701	857.2435	0.1331509	0.750	880.6665	808.2843
43	822.3203	0.2568723	0.790	1047.182	809.6238	0.2412146	0.780	968.9006	849.7386	0.1408294	0.760	873.1617	812.8726
44	798.6144	0.2811267	0.745	1023.476	789.7396	0.2738367	0.755	977.1240	828.4033	0.1708436	0.745	870.5648	796.3032
45	775.1568	0.3051272	0.700	1000.018	760.8218	0.2993311	0.700	938.8370	836.8131	0.1540540	0.735	860.2362	769.9012
46	816.2382	0.2630951	0.730	1041.100	805.8345	0.2491842	0.750	974.4805	847.2036	0.1434231	0.735	870.6266	811.4404
47	800.4835	0.2792145	0.730	1025.345	789.2699	0.2681784	0.745	962.6005	855.6052	0.1348270	0.740	879.0283	796.7068
48	793.6299	0.2862266	0.735	1018.491	775.7438	0.2758787	0.740	935.0206	815.7981	0.1857868	0.695	862.6442	786.8187
49	808.1902	0.2713294	0.755	1033.052	792.5067	0.2628204	0.720	961.1527	826.7869	0.1745436	0.755	873.6330	796.0902
50	819.1925	0.2600724	0.785	1044.054	801.8966	0.2511669	0.770	965.8580	860.1299	0.1301977	0.745	883.5530	803.5311
51	791.8750	0.2880221	0.755	1016.736	779.8703	0.2757492	0.755	948.5163	839.5179	0.1512867	0.720	862.9409	787.2740
52	806.0600	0.2735089	0.735	1030.921	791.1270	0.2662784	0.735	964.4576	851.8973	0.1386208	0.755	875.3203	803.5224
53	802.4059	0.2772476	0.765	1027.267	788.9345	0.2623828	0.745	948.2113	840.1753	0.1547066	0.755	872.9676	798.2255
54	807.4706	0.2720656	0.790	1032.332	793.3918	0.2578222	0.785	952.6686	833.1057	0.1598935	0.720	861.2134	798.8202
55	810.2136	0.2692591	0.750	1035.075	793.6990	0.2575080	0.760	952.9758	848.7008	0.1439375	0.725	876.8085	799.6112
56	790.0420	0.2898975	0.755	1014.903	780.3580	0.2752503	0.740	949.0040	829.1581	0.1680251	0.735	866.6350	795.9887
57	795.0846	0.2847382	0.750	1019.946	781.2354	0.2723062	0.745	945.1968	842.4202	0.1483171	0.700	865.8433	792.6187
58	805.2188	0.2743695	0.735	1030.080	789.7357	0.2615630	0.750	949.0125	834.0534	0.1609701	0.675	866.8457	800.6659
59	795.2504	0.2845686	0.755	1020.112	785.8066	0.2717219	0.725	959.1372	840.4803	0.1503019	0.715	863.9034	796.2905
60	803.4799	0.2761487	0.730	1028.341	793.3053	0.2701885	0.740	980.6897	854.8014	0.1356495	0.740	878.2244	803.4895
61	800.2783	0.2794244	0.755	1025.140	792.3952	0.2670271	0.745	970.4105	860.2938	0.1300300	0.755	883.7169	799.9023
62	815.9999	0.2633390	0.785	1040.861	797.4357	0.2536848	0.775	956.7125	859.9644	0.1303670	0.775	883.3875	803.8145
	anova.R2		anova.ACC		BIC_anova								
1	0.2374695	0.780	945.8226										
2	0.2422641	0.790	941.1365										
3	0.2289965	0.790	954.1040										
4	0.2399239	0.735	943.4238										
5	0.2136036	0.795	969.1488										
6	0.2323322	0.785	950.8437										
7	0.2425402	0.755	940.8666										
8	0.2396793	0.770	943.6628										
9	0.2448628	0.745	938.5965										
10	0.2286916	0.790	954.4019										
11	0.2218562	0.800	961.0828										
12	0.2437682	0.750	939.6664										
13	0.2568115	0.720	926.9181										
14	0.2546173	0.750	929.0626										
15	0.2395132	0.770	943.8251										
16	0.2440231	0.760	939.4172										
17	0.2504536	0.770	933.1322										
18	0.2410618	0.770	942.3115										
19	0.2459067	0.735	937.5762										
20	0.2594169	0.740	924.3716										
21	0.2540353	0.745	929.6314										
22	0.2602046	0.710	923.6017										

```
20 0.2594169  0.740  924.3716
21 0.2540353  0.745  929.6314
22 0.2602046  0.710  923.6017
23 0.2401352  0.790  943.2172
24 0.2342736  0.725  948.9462
25 0.2156676  0.835  967.1314
26 0.2592544  0.710  924.5304
27 0.2370339  0.750  946.2484
28 0.2443204  0.765  939.1266
29 0.2458219  0.760  937.6592
30 0.2420697  0.750  941.3265
31 0.2421497  0.775  941.2483
32 0.2502462  0.760  933.3349
33 0.2401355  0.765  943.2169
34 0.2498052  0.745  933.7660
35 0.2686482  0.715  915.3490
36 0.2574953  0.735  926.2498
37 0.2444903  0.775  938.9606
38 0.2438029  0.735  939.6324
39 0.2415606  0.750  941.8241
40 0.2434587  0.755  939.9689
41 0.2411396  0.760  942.2355
42 0.23444000 0.775  948.8226
43 0.2297055  0.780  953.4110
44 0.2466584  0.760  936.8416
45 0.2736713  0.730  910.4396
46 0.2311709  0.770  951.9788
47 0.2462455  0.760  937.2451
48 0.2563623  0.745  927.3571
49 0.2468763  0.760  936.6286
50 0.2392632  0.750  944.0695
51 0.2558965  0.775  927.8123
52 0.2392721  0.770  944.0608
53 0.2446916  0.780  938.7639
54 0.2440832  0.795  939.3585
55 0.2432738  0.765  940.1496
56 0.2469802  0.775  936.5270
57 0.2504281  0.750  933.1571
58 0.2421947  0.770  941.2042
59 0.2466714  0.740  936.8289
60 0.2393058  0.765  944.0279
61 0.2429760  0.770  940.4407
62 0.2389732  0.775  944.3529
[ reached 'max' / getOption("max.print") -- omitted 38 rows ]
```

Comparaison :

```

> # mean de tous les link fonction ensemble
> data.mean=rbind(mean_data_sel.logit,mean_data_sel.probit,mean_data_sel.cloglog)
> row.names(data.mean)<-c("logit","probit","cloglog")
> #premiere 8 valeurs (pour mieux voir)
> knitr::kable(data.mean[1:8],digits=4,
+   caption = "Mean pour les Modeles : general x 4, stepwise min AIC x 4")

```

Table: Mean pour les Modeles : general x 4, stepwise min AIC x 4

	gen.AIC	gen.R2	gen.Accl	BIC_gen	setp.AIC	setp.R2	setp.Accl	BIC_setp
llogit	804.72541	0.27491	0.74941	1029.5871	790.71251	0.26171	0.74661	952.65951
lprobit	804.51191	0.27511	0.74961	1029.3731	790.94571	0.26181	0.74681	953.59551
lcloglog	803.81551	0.27581	0.75101	1028.6771	791.08651	0.26311	0.74831	957.06231

```

> #deuxieme 8 valeurs (pour mieux voir)
> knitr::kable(data.mean[9:16],digits=4,
+   caption = "Mean pour les Modeles :stepwise min BIC x 4 , perso Anova x 4")

```

Table: Mean pour les Modeles :stepwise min BIC x 4 , perso Anova x 4

	AIC_BIC	R2_BIC	Acc_BIC	BIC_BIC	anova.AIC	anova.R2	anova.Accl	BIC_anova
llogit	840.38531	0.15411	0.72471	872.28751	799.27611	0.24361	0.75891	939.81451
lprobit	840.19931	0.15451	0.72601	872.57001	800.52441	0.24231	0.75981	941.06281
lcloglog	840.89861	0.15301	0.72661	871.58281	798.38231	0.24451	0.75971	938.92061

```

> #Quantile "97.5%"
> data.97.5=rbind(Up_data_sel.logit,Up_data_sel.probit ,Up_data_sel.cloglog)
> row.names(data.97.5)<-c("logit","probit","cloglog")
> knitr::kable(data.97.5[1:8],digits=4,
+   caption = "Quantile 97.5% pour les Modeles : general x 4, stepwise min AIC x 4")

```

Table: Quantile 97.5% pour les Modeles : general x 4, stepwise min AIC x 4

	gen.AIC	gen.R2	gen.Accl	BIC_gen	setp.AIC	setp.R2	setp.Accl	BIC_setp
llogit	827.74641	0.30031	0.80261	1052.6081	814.41561	0.28761	0.79261	983.20061
lprobit	826.56171	0.30041	0.80501	1051.4231	814.40441	0.28751	0.79261	983.02421
lcloglog	826.65201	0.30201	0.79761	1051.5131	815.12891	0.29021	0.79001	983.39431

```

> knitr::kable(data.97.5[9:16],digits=4,
+   caption = "Quantile 97.5% pour les Modeles :stepwise min BIC x 4 , perso Anova x 4")

```

Table: Quantile 97.5% pour les Modeles :stepwise min BIC x 4 , perso Anova x 4

	AIC_BIC	R2_BIC	Acc_BIC	BIC_BIC	anova.AIC	anova.R2	anova.Accl	BIC_anova
llogit	863.10311	0.18891	0.76761	887.26551	823.62181	0.26781	0.80761	964.16021
lprobit	861.63941	0.19471	0.76761	887.23981	824.45411	0.26651	0.80501	964.99241
lcloglog	862.23671	0.20491	0.76501	886.48641	822.62811	0.26791	0.80261	963.16641

```

> #Quantile "2.5%"
> data.2.5=rbind(Low_data_sel.logit,Low_data_sel.probit ,Low_data_sel.cloglog)
> row.names(data.2.5)<-c("logit","probit","cloglog")
> knitr::kable(data.2.5[1:8],digits=4,
+ caption = "Quantile 2.5% pour les Modeles : general x 4, stepwise min AIC x 4")

```

Table: Quantile 2.5% pour les Modeles : general x 4, stepwise min AIC x 4

	gen.AIC	gen.R2	gen.Acc	BIC_gen	step.AIC	step.R2	step.Acc	BIC_step
llogit	779.9083	0.2513	0.7021	1004.770	766.8467	0.2377	0.6974	923.6133
lprobit	779.8203	0.2525	0.6997	1004.682	766.4372	0.2373	0.7050	923.3988
lcloglog	778.2260	0.2524	0.7024	1003.087	767.3324	0.2378	0.7000	927.0247

```

> knitr::kable(data.2.5[9:16],digits=4,
+ caption = "Quantile 2.5% pour les Modeles :stepwise min BIC x 4 , perso Anova x 4")

```

Table: Quantile 2.5% pour les Modeles :stepwise min BIC x 4 , perso Anova x 4

	AIC_BIC	R2_BIC	Acc_BIC	BIC_BIC	anova.AIC	anova.R2	anova.Acc	BIC_anova
llogit	810.8637	0.1272	0.6724	852.9531	775.6042	0.2187	0.7174	916.1425
lprobit	810.1227	0.1287	0.6724	853.7958	776.9113	0.2179	0.7200	917.4496
lcloglog	802.4044	0.1280	0.6824	853.4893	775.5277	0.2197	0.7150	916.0661

#Plot

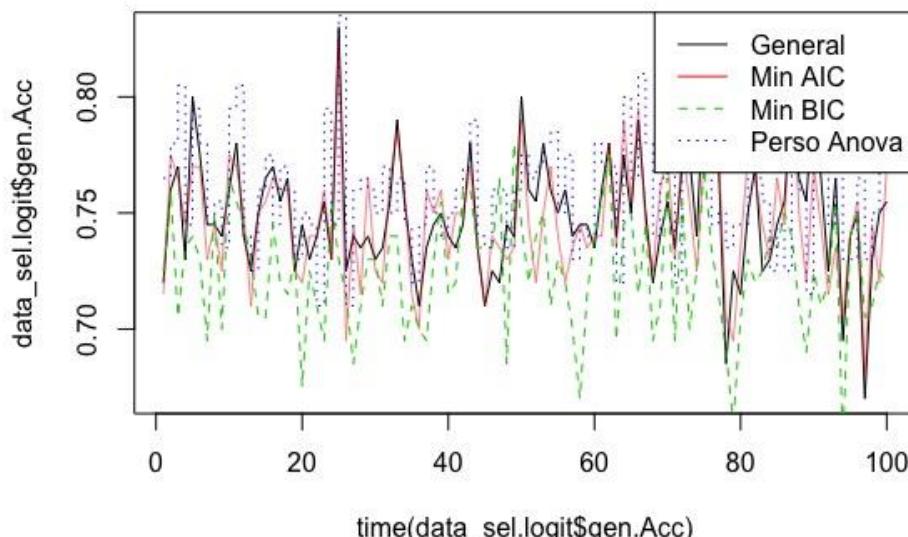
#Accuracy

```

plot(time(data_sel.logit$gen.Acc),data_sel.logit$gen.Acc,type="l",
      main="Comparison Accuracy modelisation Logite")
lines(time(data_sel.logit$gen.Acc),data_sel.logit$step.Acc,type="l",col=2,lwd=.5)
lines(time(data_sel.logit$gen.Acc),data_sel.logit$Acc_BIC,type="l",col=3,lty=2)
lines(time(data_sel.logit$gen.Acc),data_sel.logit$anova.Acc,type="s",col=4,lty=3)
legend("topright",col = c(1,2,3,4),
      legend = c("General","Min AIC","Min BIC","Perso Anova"),lty =c(1,1,2,3))

```

Comparison Accuracy modelisation Logite



```

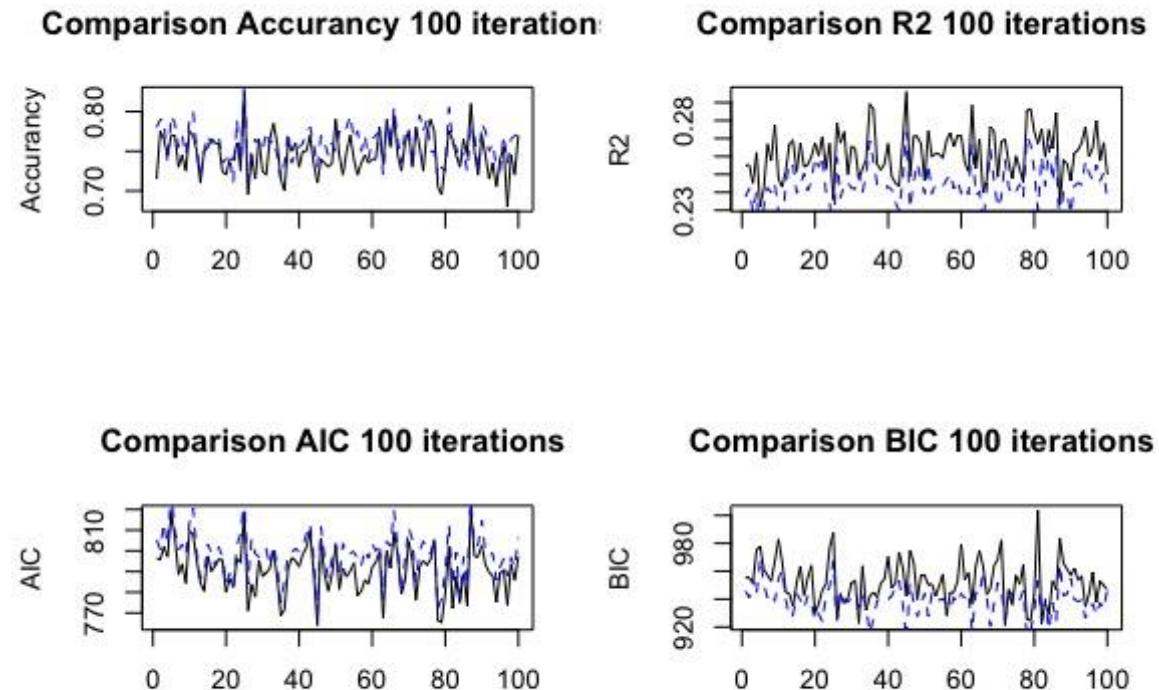
#Comparison entre le modèle Step-min-AIC logit et le modèle Anova-perso- cloglog
par(mfrow=c(2,2))
plot(time(data_sel.logit$gen.Acc),data_sel.logit$setp.Acc,type="l",
     main="Comparison Accuracy 100 iterations ",ylab = "Accuracy",xlab = "")
lines(time(data_sel.logit$gen.Acc),data_sel.cloglog$anova.Acc,type="l",col=4,lty=2)
legend("bottomleft",col = c(1,4),
       legend = c("Min AIC Logite","Perso Anova cloglog"),lty =c(1,2))

plot(time(data_sel.logit$gen.Acc),data_sel.logit$setp.R2,type="l",
     main="Comparison R2 100 iterations ",ylab = "R2",xlab = "")
lines(time(data_sel.logit$gen.Acc),data_sel.cloglog$anova.R2,type="l",col=4,lty=2)
legend("topleft",col = c(1,4),
       legend = c("Min AIC Logite","Perso Anova cloglog"),lty =c(1,2))

plot(time(data_sel.logit$gen.Acc),data_sel.logit$setp.AIC,type="l",
     main="Comparison AIC 100 iterations ",ylab = "AIC",xlab = "")
lines(time(data_sel.logit$gen.Acc),data_sel.cloglog$anova.AIC,type="l",col=4,lty=2)
legend("bottomleft",col = c(1,4),
       legend = c("Min AIC Logite","Perso Anova cloglog"),lty =c(1,2))

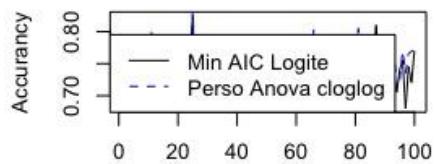
plot(time(data_sel.logit$gen.Acc),data_sel.logit$BIC_setp,type="l",
     main="Comparison BIC 100 iterations ",ylab = "BIC",xlab = "")
lines(time(data_sel.logit$gen.Acc),data_sel.cloglog$BIC_anova,type="l",col=4,lty=2)
legend("topright",col = c(1,4),
       legend = c("Min AIC Logite","Perso Anova cloglog"),lty =c(1,2))

```

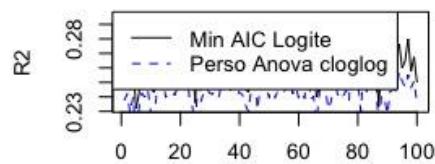


Légendes :

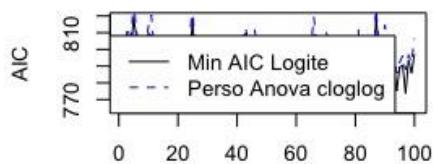
Comparison Accuracy 100 iteration:



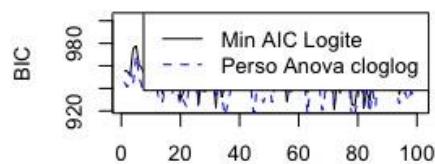
Comparison R2 100 iterations



Comparison AIC 100 iterations



Comparison BIC 100 iterations



Modèle final:

```

> summary(model_Step)

Call:
glm(formula = class ~ status_account + duration + history + purpose +
    amount + savings + employment + installement_rate + status_personnal +
    other_debtors + age + other_installment + housing, family = binomial(link = "logit"),
    data = credit1)

Deviance Residuals:
    Min      1Q  Median      3Q     Max 
-2.3609 -0.7069 -0.3904  0.7401  2.7304 

Coefficients:
            Estimate Std. Error z value Pr(>|z|)    
(Intercept) 1.601e+00  8.279e-01  1.934 0.053113 .  
status_accountA12 -3.784e-01  2.119e-01 -1.786 0.074171 .  
status_accountA13 -1.020e+00  3.625e-01 -2.814 0.004892 ** 
status_accountA14 -1.716e+00  2.287e-01 -7.503 6.25e-14 *** 
duration        3.088e-02  8.963e-03  3.445 0.000570 *** 
historyA31      2.587e-02  5.228e-01  0.049 0.960538    
historyA32      -7.222e-01  4.071e-01 -1.774 0.076075 .  
historyA33      -8.169e-01  4.636e-01 -1.762 0.078082 .  
historyA34      -1.407e+00  4.307e-01 -3.267 0.001088 ** 
purposeA41      -1.618e+00  3.663e-01 -4.417 1.00e-05 *** 
purposeA410     -1.686e+00  7.789e-01 -2.165 0.030407 *  
purposeA42      -7.248e-01  2.529e-01 -2.865 0.004167 ** 
purposeA43      -8.626e-01  2.434e-01 -3.543 0.000395 *** 
purposeA44      -5.040e-01  7.488e-01 -0.673 0.500099    
purposeA45      -9.066e-02  5.377e-01 -0.169 0.866114    
purposeA46      1.546e-01  3.930e-01  0.393 0.694027    
purposeA48      -1.902e+00  1.208e+00 -1.575 0.115329    
purposeA49      -7.335e-01  3.266e-01 -2.246 0.024714 *  
amount          1.138e-04  4.117e-05  2.764 0.005710 ** 
savingsA62      -2.791e-01  2.766e-01 -1.009 0.312814    
savingsA63      -4.142e-01  3.985e-01 -1.039 0.298652    
savingsA64      -1.271e+00  5.101e-01 -2.492 0.012692 *  
savingsA65      -9.699e-01  2.566e-01 -3.780 0.000157 *** 
employmentA72   3.342e-02  3.783e-01  0.088 0.929606    
employmentA73   -8.461e-02  3.530e-01 -0.240 0.810609    
employmentA74   -6.988e-01  3.949e-01 -1.769 0.076817 .  
employmentA75   -8.750e-02  3.653e-01 -0.240 0.810699    
installement_rate 3.250e-01  8.517e-02  3.816 0.000135 *** 
status_personnalA92 -2.586e-01  3.773e-01 -0.685 0.493163    
status_personnalA93 -7.423e-01  3.678e-01 -2.018 0.043578 *  
status_personnalA94 -4.424e-01  4.450e-01 -0.994 0.320108    
other_debtorsA102 4.322e-01  3.992e-01  1.083 0.278967    
status_personnalA94 -4.424e-01  4.450e-01 -0.994 0.320108 
other_debtorsA102 4.322e-01  3.992e-01  1.083 0.278967 
other_debtorsA103 -1.046e+00  4.151e-01 -2.520 0.011751 *  
age              -1.549e-02  8.937e-03 -1.733 0.083064 .  
other_installmentA142 -1.322e-01  4.094e-01 -0.323 0.746823 
other_installmentA143 -6.805e-01  2.353e-01 -2.891 0.003835 ** 
housingA152      -4.312e-01  2.213e-01 -1.949 0.051350 .  
housingA153      -1.746e-01  3.410e-01 -0.512 0.608690 

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

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1221.73 on 999 degrees of freedom
 Residual deviance: 910.64 on 962 degrees of freedom
 AIC: 986.64

Number of Fisher Scoring iterations: 5

```

> summary(model_Anova_perso)

Call:
glm(formula = class ~ status_account + duration + history + purpose +
    savings + installement_rate + status_personnal + other_debtors +
    other_installment, family = binomial(link = "cloglog"), data = credit1)

Deviance Residuals:
    Min      1Q   Median      3Q      Max 
-2.3574 -0.7019 -0.4267  0.7358  2.6055 

Coefficients:
            Estimate Std. Error z value Pr(>|z|)    
(Intercept) 0.324730  0.420845  0.772 0.440343  
status_accountA12 -0.246793  0.148336 -1.664 0.096164 .  
status_accountA13 -0.894845  0.293497 -3.049 0.002297 ** 
status_accountA14 -1.390608  0.181507 -7.661 1.84e-14 *** 
duration        0.031668  0.005034  6.291 3.15e-10 *** 
historyA31     -0.177901  0.320895 -0.554 0.579312  
historyA32     -0.637166  0.250480 -2.544 0.010966 *  
historyA33     -0.680047  0.303707 -2.239 0.025146 *  
historyA34     -1.216372  0.278018 -4.375 1.21e-05 *** 
purposeA41      -1.014103  0.275923 -3.675 0.000238 *** 
purposeA40      -0.955127  0.524373 -1.821 0.068536 .  
purposeA42      -0.459800  0.180500 -2.547 0.010854 *  
purposeA43      -0.674263  0.178092 -3.786 0.000153 *** 
purposeA44      -0.485717  0.535644 -0.907 0.364518  
purposeA45      -0.191039  0.391749 -0.488 0.625794  
purposeA46      0.070567  0.268105  0.263 0.792390  
purposeA48      -1.702464  1.048875 -1.623 0.104561  
purposeA49      -0.611151  0.232533 -2.628 0.008583 ** 
savingsA62     -0.215849  0.202279 -1.067 0.285933  
savingsA63     -0.567657  0.343672 -1.652 0.098588 .  
savingsA64     -1.112270  0.424101 -2.623 0.008725 ** 
savingsA65     -0.820670  0.202937 -4.044 5.26e-05 *** 
installement_rate 0.207802  0.058618  3.545 0.000393 *** 
status_personnalA92 -0.132080  0.256979 -0.514 0.607270  
status_personnalA93 -0.629951  0.254153 -2.479 0.013189 *  
status_personnalA94 -0.325453  0.318767 -1.021 0.307266  
other_debtorsA102  0.274945  0.274268  1.002 0.316118  
other_debtorsA103  -0.914037  0.330505 -2.766 0.005682 ** 
other_installmentA142 -0.045985  0.286935 -0.160 0.872674  
other_installmentA143 -0.457493  0.169583 -2.698 0.006981 ** 
---
Signif. codes:  0 '****' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1221.73 on 999 degrees of freedom
 Residual deviance: 932.31 on 970 degrees of freedom
 AIC: 992.31

Number of Fisher Scoring iterations: 8

```

> model_Anova_perso

Call: glm(formula = class ~ status_account + duration + history + purpose +
  savings + installement_rate + status_personnal + other_debtors +
  other_installment, family = binomial(link = "cloglog"), data = credit1)

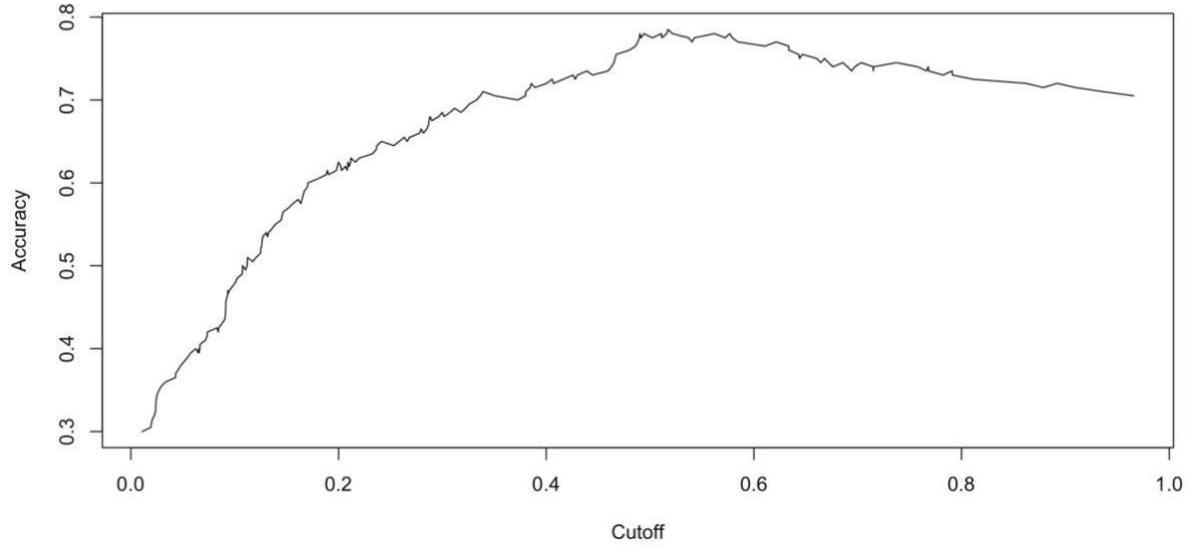
Coefficients:
(Intercept)      status_accountA12      status_accountA13      status_accountA14      duration
               0.32473                 -0.24679                 -0.89484                -1.39061                0.03167
historyA31          historyA32          historyA33          historyA34      purposeA41
              -0.17790                 -0.63717                 -0.68005                -1.21637                -1.01410
purposeA410         purposeA42         purposeA43         purposeA44      purposeA45
             -0.95513                 -0.45980                 -0.67426                -0.48572                -0.19104
purposeA46         purposeA48         purposeA49         savingsA62      savingsA63
              0.07057                 -1.70246                 -0.61115                -0.21585                -0.56766
savingsA64         savingsA65     installement_rate  status_personnalA92  status_personnalA93
             -1.11227                 -0.82067                 0.20780                -0.13208                -0.62995
status_personnalA94 other_debtorsA102 other_debtorsA103 other_installmentA142 other_installmentA143
            -0.32545                 0.27495                -0.91404                -0.04599                -0.45749

Degrees of Freedom: 999 Total (i.e. Null);  970 Residual
Null Deviance:    1222
Residual Deviance: 932.3      AIC: 992.3

```

Question 6:

```
> print(c(Accuracy = acc, Cutoff = cut))
Accuracy Cutoff.476
0.7850000 0.5173151
>
```



```
> model_Anova_perso
```

```
Call: glm(formula = class ~ status_account + duration + history + purpose +
savings + installement_rate + status_personnal + other_debtors +
other_installment, family = binomial(link = "cloglog"), data = train1)
```

Coefficients:

	(Intercept)	status_accountA12	status_accountA13	status_accountA14	duration
	0.154088	-0.191839	-0.779266	-1.375131	0.031511
historyA31		historyA32	historyA33	historyA34	purposeA41
	-0.002182	-0.432546	-0.188618	-1.027810	-1.175766
purposeA410		purposeA42	purposeA43	purposeA44	purposeA45
	-1.295211	-0.620100	-0.782514	-0.582304	-0.186421
purposeA46		purposeA48	purposeA49	savingsA62	savingsA63
	-0.190211	-1.388482	-0.553215	-0.429025	-0.690169
savingsA64		savingsA65	installement_rate	status_personnalA92	status_personnalA93
	-1.358705	-0.792295	0.256229	-0.210919	-0.677744
status_personnalA94		other_debtorsA102	other_debtorsA103	other_installmentA142	other_installmentA143
	-0.565015	0.274215	-0.840744	-0.246971	-0.463415

Degrees of Freedom: 799 Total (i.e. Null); 770 Residual

Null Deviance: 977.4

Residual Deviance: 736.8 AIC: 796.8