

Trabalho MLG - Titanic

Carlos Manchini

26 de novembro de 2017

Contents

Importando dados	1
Definindo as variáveis	1
Retirando tripulantes com dados faltantes (NA)	2
1. Ajustando modelo 1	2
1.1. Análise de diagnóstico e influência fit1	3
1.2. Verificação da função de ligação - teste RESET	8
1.3. Qualidade do ajuste	9
2. Ajustando modelo 2	9
2.1. Análise de diagnóstico e influência fit2	10
2.2. Verificação da função de ligação - teste RESET	15
2.3. Qualidade do ajuste	16
3. Ajustando modelo 3 - full	16
3.1. Seleção do modelo	17
3.2. Análise de diagnóstico e influência fit3_step	18
3.3. Verificação da função de ligação - teste RESET	23
3.4. Qualidade do ajuste	24
4. Modelo fit3 sem outliers	24
5. Probabilidade de sobrevivência	27

Importando dados

```
data<-read.csv('titanic.csv',header=T,sep = ",")
head(data)
class(data)
```

Definindo as variáveis

```
surv<-data$Survived #891
sex<-data$Sex
class<-data$Pclass
age<-data$Age

surv<-as.numeric(surv)

library(varhandle)
sex<-unfactor(sex)
sex[sex=="male"]=0
```

```
sex[sex=="female"]=1
sex<-as.numeric(sex)

names<-c("Classe")
class<-to.dummy(class, names)
C1<-class[,1]
C2<-class[,2]
C3<-class[,3]
#ou
#C1<-(Pclass==1)+(Pclass==0)
#C2<-(Pclass==2)+(Pclass==0)
```

Retirando tripulantes com dados faltantes (NA)

```
frame<-data.frame(surv,sex,class,age)
frame<-na.exclude(frame)

surv<-frame$surv
sex<-frame$sex
C1<-frame$Classe.1
C2<-frame$Classe.2
C3<-frame$Classe.3
age<-frame$age
```

1. Ajustando modelo 1

```
fit1<-glm(surv~sex+C1+C2+age, family=binomial(link="logit"))
summary(fit1)
```

```
##
## Call:
## glm(formula = surv ~ sex + C1 + C2 + age, family = binomial(link = "logit"))
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.7303  -0.6780  -0.3953   0.6485   2.4657
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.326394   0.247892  -5.351 8.76e-08 ***
## sex          2.522781   0.207391  12.164 < 2e-16 ***
## C1           2.580625   0.281442   9.169 < 2e-16 ***
## C2           1.270826   0.244048   5.207 1.92e-07 ***
## age        -0.036985   0.007656  -4.831 1.36e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 964.52  on 713  degrees of freedom
```

```
## Residual deviance: 647.28 on 709 degrees of freedom
## AIC: 657.28
##
## Number of Fisher Scoring iterations: 5
```

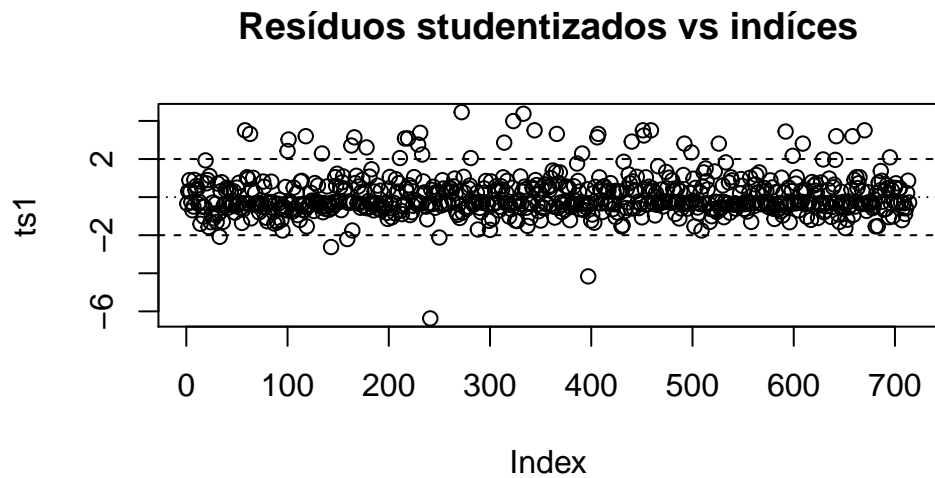
Modelo obtido (coerente com o site):

$$g(X) = -1.33 + 2.55 \times I_{feminino} + 1.27 \times I_{2classe} + 2.58 \times I_{1classe} - 0.04 \times I_{idade}$$

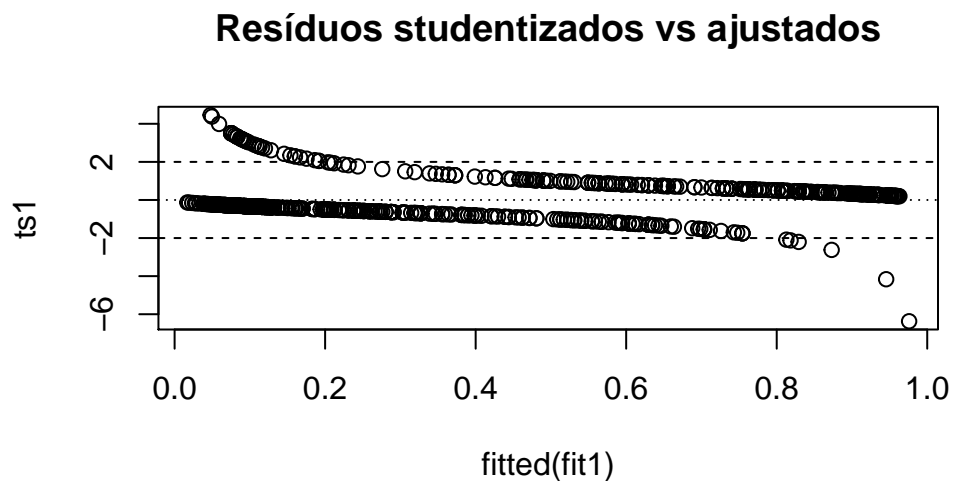
1.1. Análise de diagnóstico e influência fit1

1.1.1. Resíduos

```
ts1<-resid(fit1,type="pearson") #studentizados
abline(plot(ts1, main='Resíduos studentizados vs índices'), lty=c(2,2,3), h=c(-2,2,0))
```

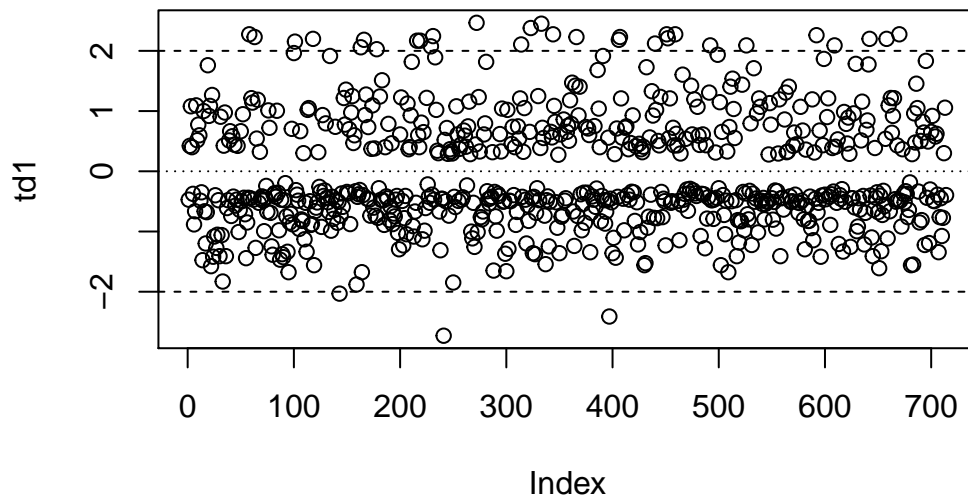


```
abline(plot(fitted(fit1),ts1,main='Resíduos studentizados vs ajustados') , lty=c(2,2,3), h=c(-2,2,0))
```



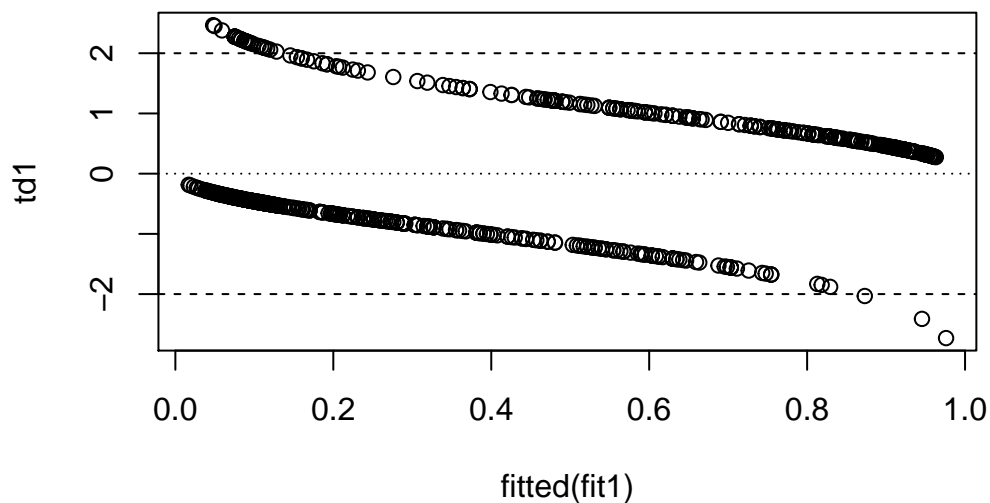
```
td1<-resid(fit1,type="deviance") #componente do desvio
abline(plot(td1,main='Resíduos componente do desvio vs índices'), lty=c(2,2,3), h=c(-2,2,0))
```

Resíduos componente do desvio vs índices



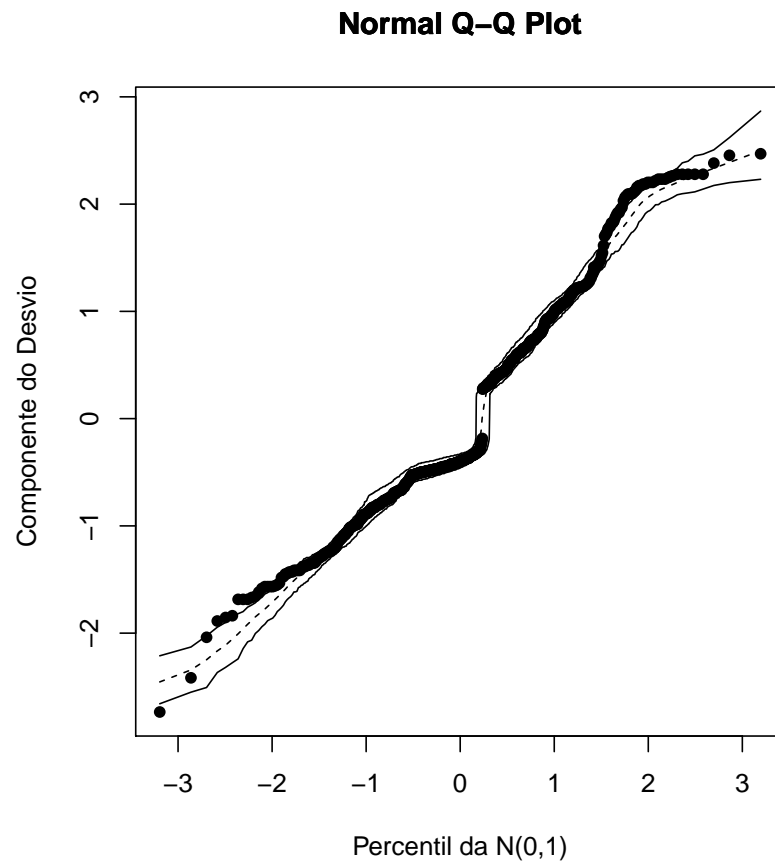
```
abline(plot(fitted(fit1),td1,main='Resíduos componente do desvio vs ajustados') ,
      lty=c(2,2,3), h=c(-2,2,0))
```

Resíduos componente do desvio vs ajustados



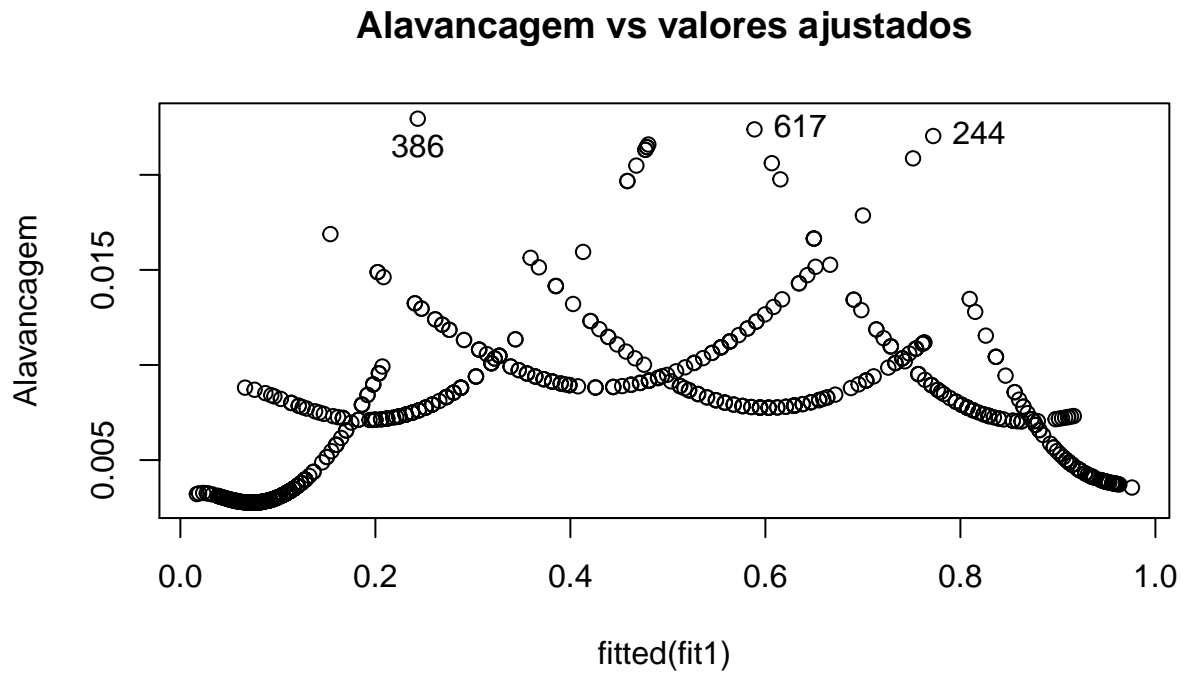
1.1.2. Envelope Simulado

```
source("envel_bino.r")  
envelope.bin(fit1) #default: residuo deviance  
##envelope.bin(fit1,liga=binomial(link= * )) *alterar funcao de ligacao
```



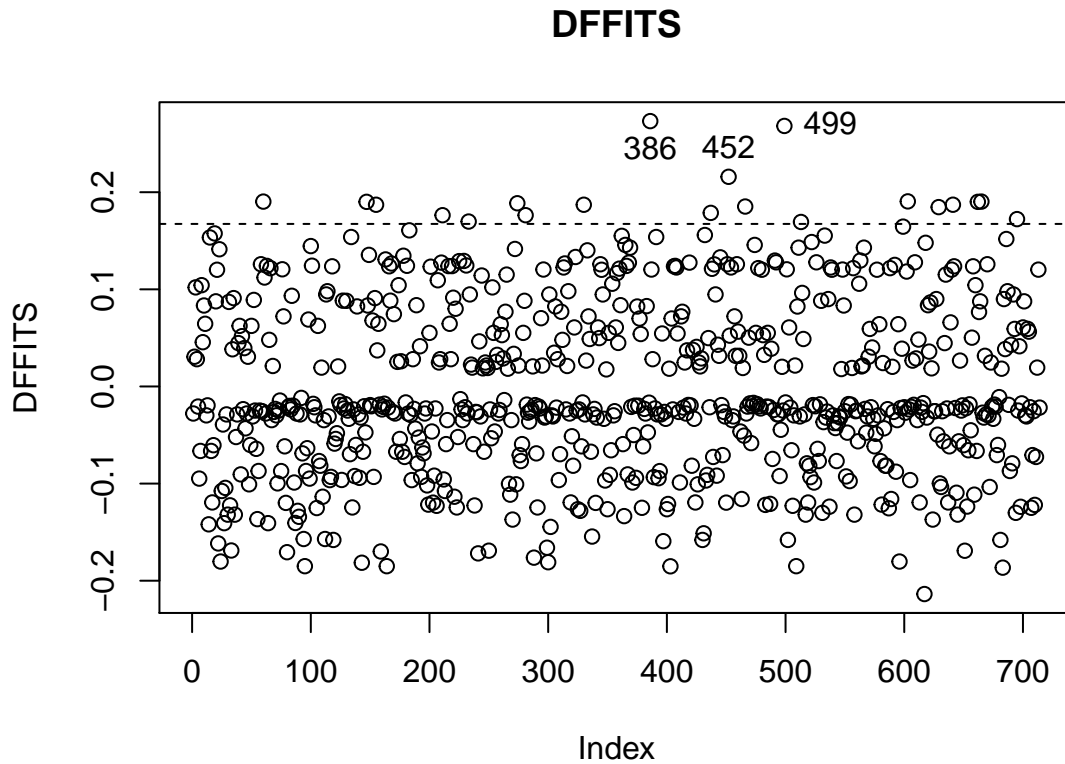
1.1.3. Alavancagem

```
hatvalues(fit1)
plot(fitted(fit1),hatvalues(fit1),ylab="Alavancagem",
     main='Alavancagem vs valores ajustados')
identify(fitted(fit1),hatvalues(fit1), n=3)
```



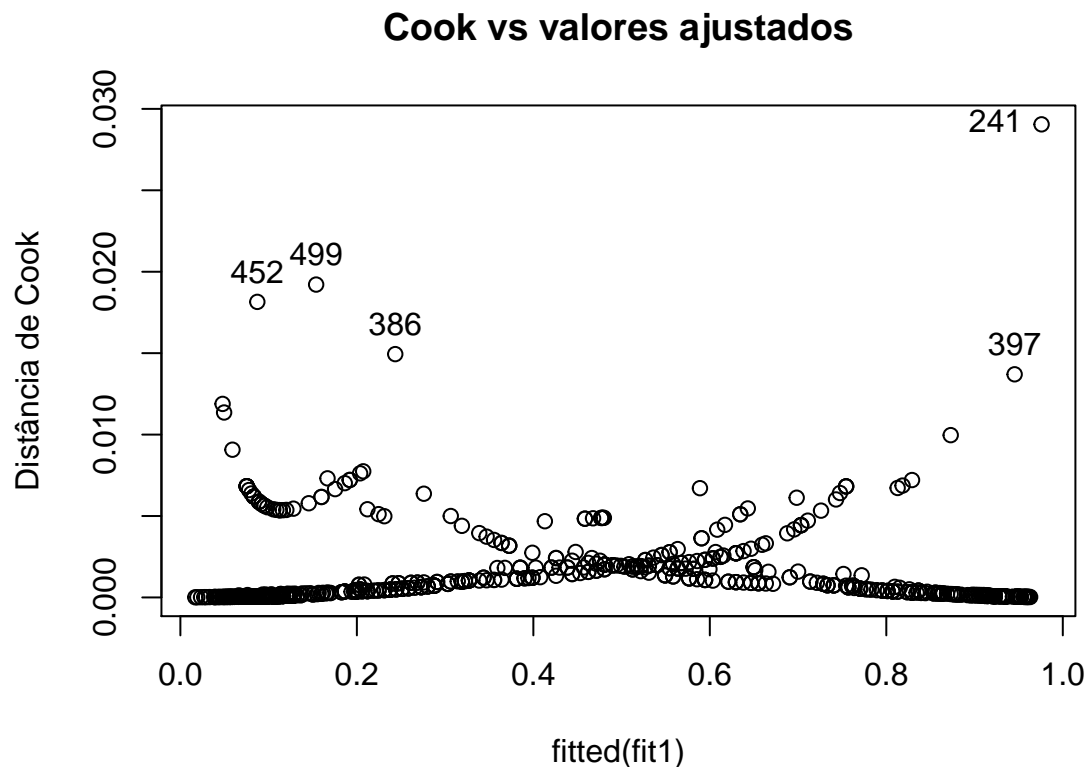
1.1.4. DFFIT

```
dffits(fit1)
limite<-2*sqrt(fit1$rank / n)
limite
abline(plot(dffits(fit1),ylab="DFFITS",main="DFFITS"), h=limite,lty=2)
identify(dffits(fit1), n=3)
```



1.1.5. Distância de Cook

```
cooks.distance(fit1)
plot(fitted(fit1),cooks.distance(fit1),ylab="Distância de Cook",
     main="Cook vs valores ajustados")
identify(fitted(fit1),cooks.distance(fit1), n=5)
```



1.2. Verificação da função de ligação - teste RESET

```
eta.hat<-(fit1$linear.predictors)^2
verify.link<-glm(surv~sex+C1+C2+age+eta.hat, family=binomial(link="logit"))
summary(verify.link)
```

```
##
## Call:
## glm(formula = surv ~ sex + C1 + C2 + age + eta.hat, family = binomial(link = "logit"))
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.6078  -0.6383  -0.4806   0.5698   2.1541
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.776798   0.284367  -6.248 4.15e-10 ***
## sex          2.829822   0.253908  11.145 < 2e-16 ***
## C1           3.115581   0.365759   8.518 < 2e-16 ***
## C2           1.567466   0.271072   5.782 7.36e-09 ***
## age        -0.044969   0.008405  -5.350 8.79e-08 ***
## eta.hat      0.177069   0.054255   3.264  0.0011 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```



```
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 964.52  on 713  degrees of freedom
## Residual deviance: 635.11  on 708  degrees of freedom
## AIC: 647.11
##
## Number of Fisher Scoring iterations: 5
```

$\hat{\eta}^2$ é significativo quando adicionado no modelo como covariável extra. Portanto, o modelo está incorretamente especificado.

1.3. Qualidade do ajuste

```
source("pseudo_R2_disc.r")
pseudo_R2(fit1)
```

```
## [1] pseudo R2 = 0.4841
```

O modelo foi capaz de explicar 48,4% da variabilidade dos dados.

2. Ajustando modelo 2

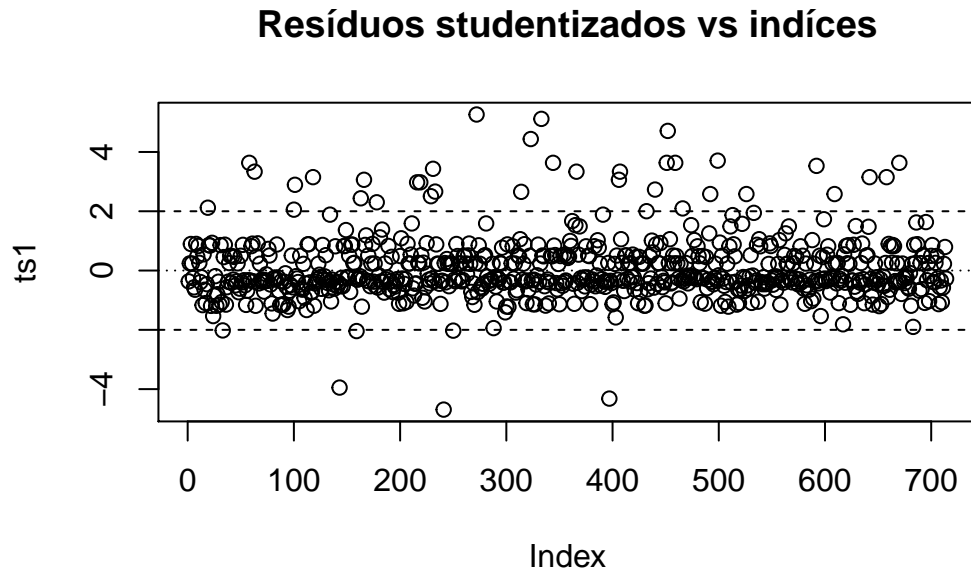
```
SeC2<-sex*age #Interação
fit2<-glm(surv~sex+age+C1+C2+SeC2, family = binomial(link="logit"))
summary(fit2)
```

```
##
## Call:
## glm(formula = surv ~ sex + age + C1 + C2 + SeC2, family = binomial(link = "logit"))
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.5043  -0.6389  -0.3718   0.6579   2.5913
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.75521    0.29501  -2.560 0.010469 *
## sex          1.16344    0.43762   2.659 0.007848 **
## age         -0.05704    0.01010  -5.647 1.63e-08 ***
## C1           2.69756    0.29582   9.119 < 2e-16 ***
## C2           1.19186    0.24323   4.900 9.58e-07 ***
## SeC2         0.04985    0.01478   3.373 0.000745 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 964.52  on 713  degrees of freedom
## Residual deviance: 635.39  on 708  degrees of freedom
## AIC: 647.39
##
## Number of Fisher Scoring iterations: 5
```

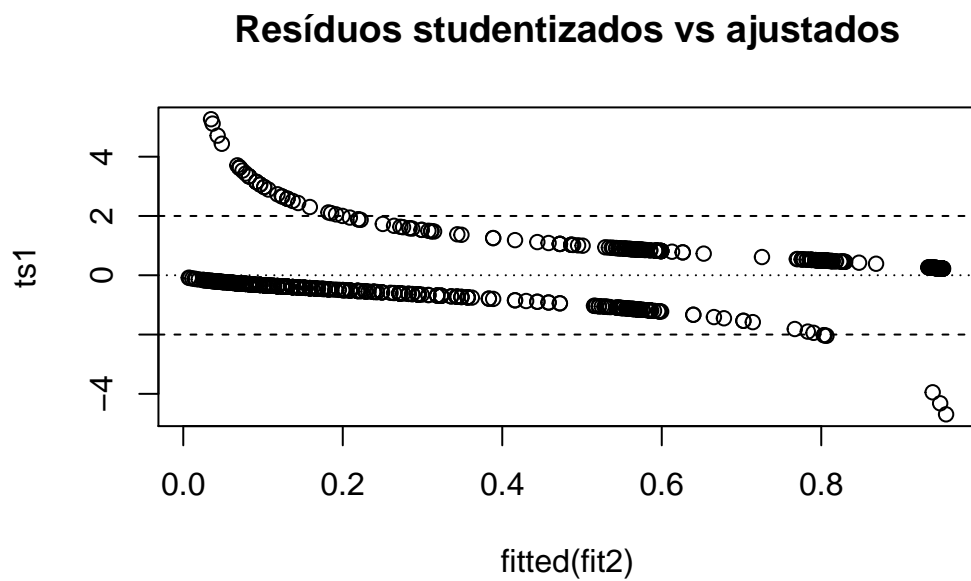
2.1. Análise de diagnóstico e influência fit2

2.1.1. Resíduos

```
ts1<-resid(fit2,type="pearson") #studentizados  
abline(plot(ts1, main='Resíduos studentizados vs índices'), lty=c(2,2,3), h=c(-2,2,0))
```

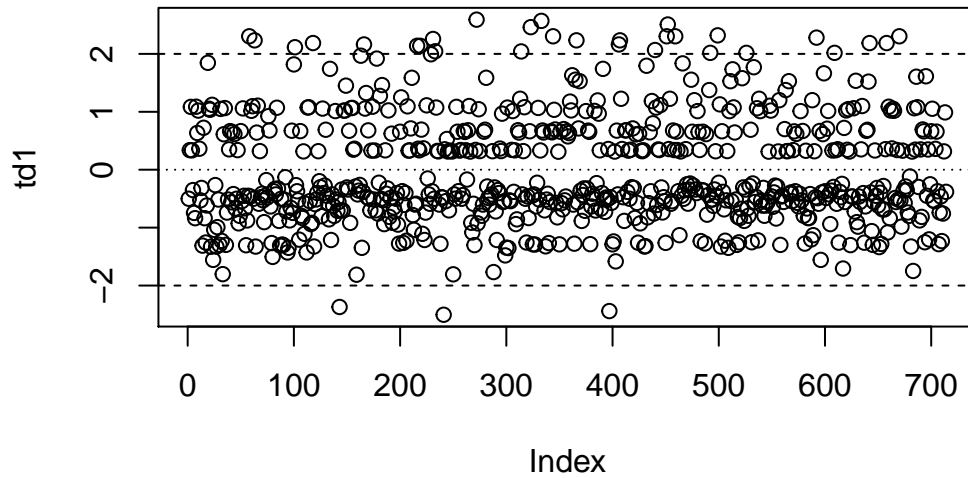


```
abline(plot(fitted(fit2),ts1,main='Resíduos studentizados vs ajustados') ,  
        lty=c(2,2,3), h=c(-2,2,0))
```



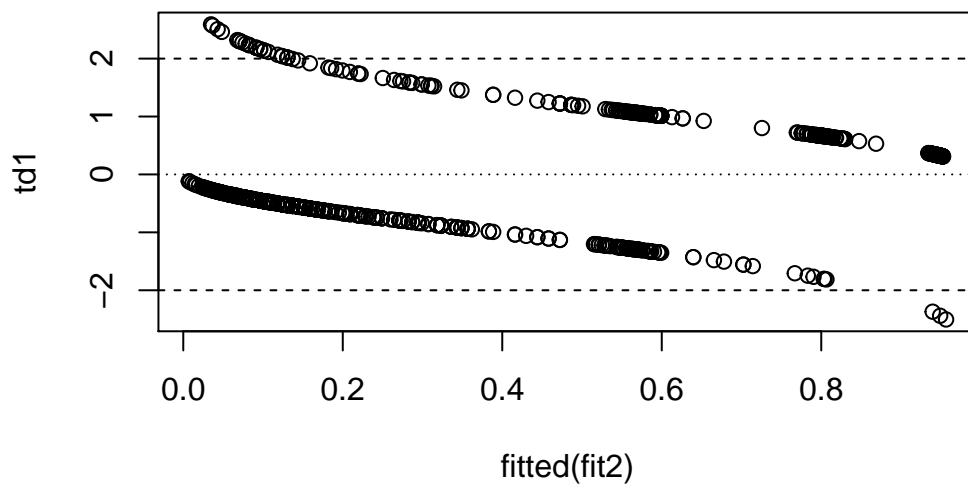
```
td1<-resid(fit2,type="deviance") #componente do desvio
abline(plot(td1,main='Resíduos componente do desvio vs índices'),lty=c(2,2,3),h=c(-2,2,0))
```

Resíduos componente do desvio vs índices



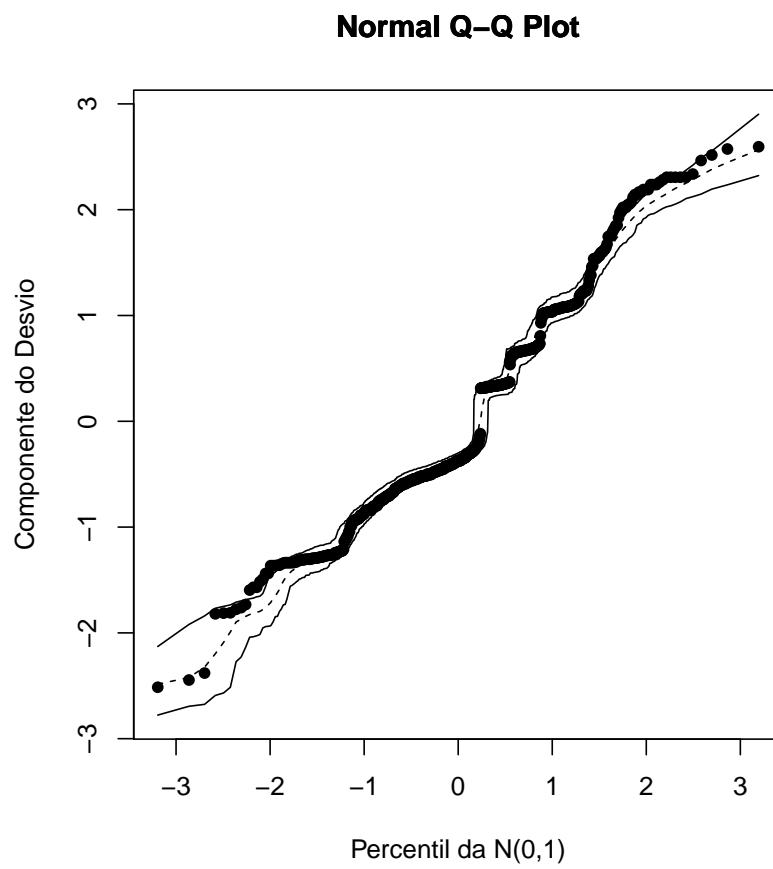
```
abline(plot(fitted(fit2),td1,main='Resíduos componente do desvio vs ajustados'),
      lty=c(2,2,3), h=c(-2,2,0))
```

Resíduos componente do desvio vs ajustados



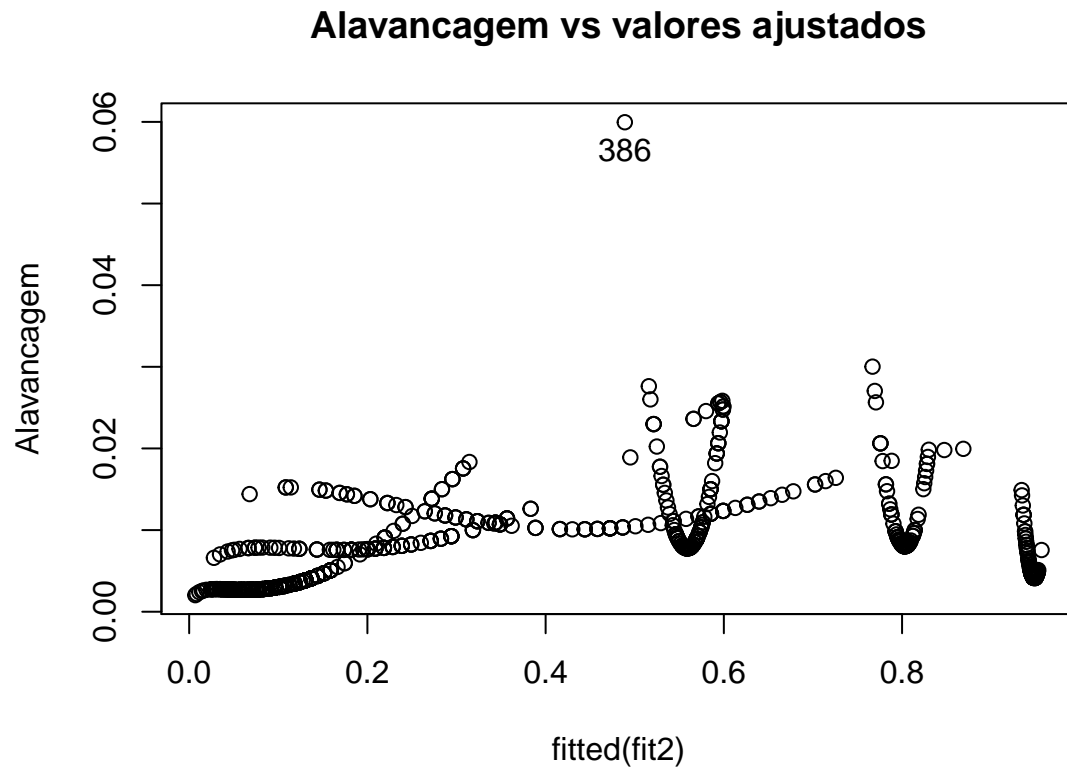
#####2.1.2. Envelope Simulado

```
source("envel_bino.r")  
envelope.bin(fit2) #default: residuo deviance  
##envelope.bin(fit2,liga=binomial(link= * )) *alterar funcao de ligacao  
  
n<-length(surv)
```



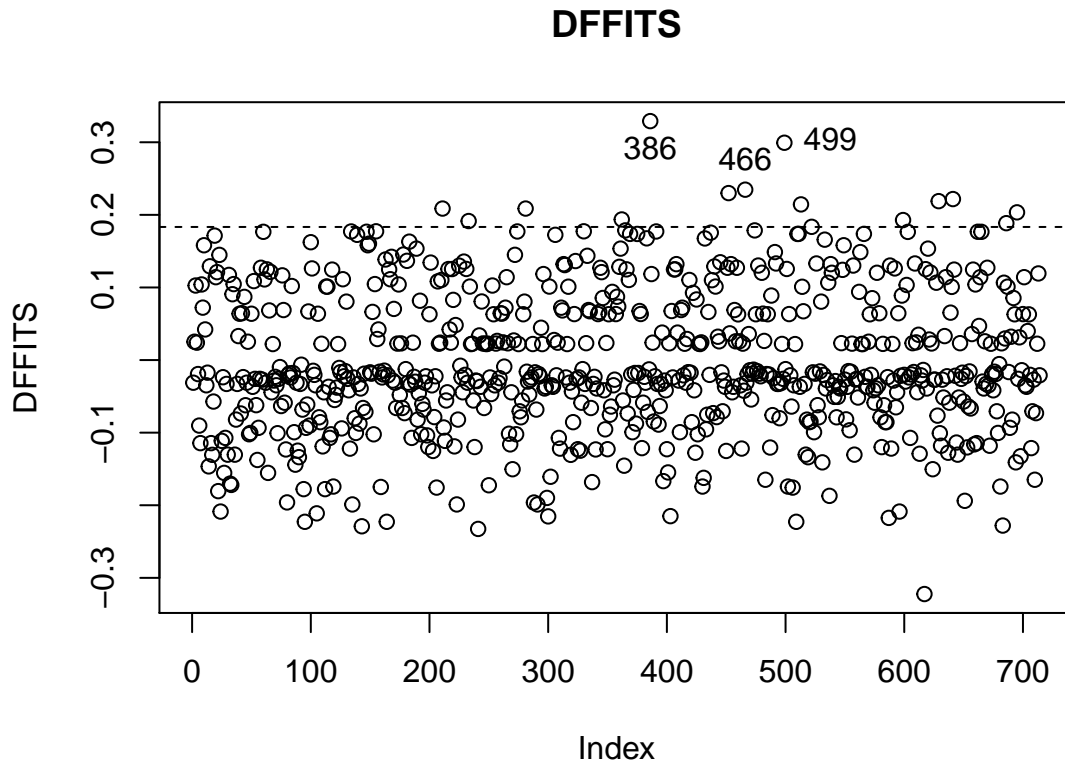
2.1.3. Alavancagem

```
hatvalues(fit2)
plot(fitted(fit2),hatvalues(fit2),ylab="Alavancagem",
     main='Alavancagem vs valores ajustados')
identify(fitted(fit2),hatvalues(fit2), n=1)
```



2.1.4. DFFIT

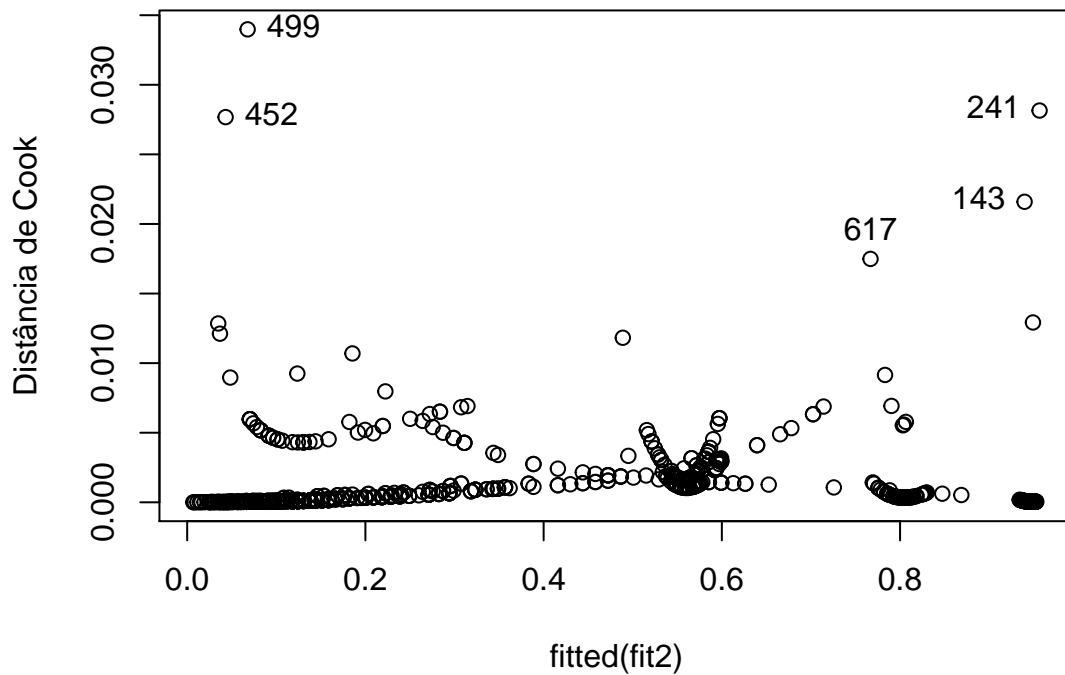
```
dffits(fit2)
limite<-2*sqrt(fit2$rank / n)
limite
abline(plot(dffits(fit2),ylab="DFFITS",main="DFFITS"), h=limite,lty=2)
identify(dffits(fit2), n=3)
```



2.1.5. Distância de Cook

```
cooks.distance(fit2)
plot(fitted(fit2),cooks.distance(fit2), ylab="Distância de Cook",
     main="Cook vs valores ajustados")
identify(fitted(fit2),cooks.distance(fit2), n=5)
```

Cook vs valores ajustados



2.2. Verificação da função de ligação - teste RESET

```
eta.hat<-(fit2$linear.predictors)^2
verify.link<-glm(surv~sex+C1+C2+age+eta.hat, family=binomial(link="logit"))
summary(verify.link)
```

```
##
## Call:
## glm(formula = surv ~ sex + C1 + C2 + age + eta.hat, family = binomial(link = "logit"))
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.3313  -0.6515  -0.4739   0.5693   2.1396
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.487178   0.254188  -5.851 4.89e-09 ***
## sex          2.751961   0.239746  11.479 < 2e-16 ***
## C1           2.994020   0.341608   8.765 < 2e-16 ***
## C2           1.474874   0.255140   5.781 7.44e-09 ***
## age         -0.050873   0.009129  -5.572 2.51e-08 ***
## eta.hat      0.145204   0.042462   3.420 0.000627 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 964.52 on 713 degrees of freedom
## Residual deviance: 634.80 on 708 degrees of freedom
## AIC: 646.8
##
## Number of Fisher Scoring iterations: 5
```

$\hat{\eta}^2$ é significativo quando adicionado no modelo como covariável extra. Portanto, o modelo está incorretamente especificado.

2.3. Qualidade do ajuste

```
source("pseudo_R2_disc.r")
pseudo_R2(fit2)
```

```
## [1] pseudo R2 = 0.4984
```

O modelo foi capaz de explicar aproximadamente 50% da variabilidade dos dados.

Observações potencialmente influentes

```
age[c(368,452,499)]
surv[c(368,452,499)]
C1[c(368,452,499)]
C2[c(368,452,499)]
```

Neste contexto não será necessário retirarmos as devidas observações pois há alta quantidade de dados e a influência é mínima.

3. Ajustando modelo 3 - full

```
SeA<-sex*age
SeC1<-sex*C1
SeC2<-sex*C2
AeC1<-C1*age
AeC2<-C2*age

fit3<-glm(surv~sex+age+C1+C2+SeA+SeC1+SeC2+AeC1+AeC2, family = binomial(link="logit"))
summary(fit3)
```

```
##
## Call:
## glm(formula = surv ~ sex + age + C1 + C2 + SeA + SeC1 + SeC2 +
##      AeC1 + AeC2, family = binomial(link = "logit"))
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.6219  -0.6405  -0.3971   0.3376   3.2713
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
```



```
## (Intercept) -0.55620    0.38204   -1.456    0.14543
## sex         0.79386    0.49194    1.614    0.10658
## age        -0.04861    0.01521   -3.197    0.00139 **
## C1          1.48300    0.68348    2.170    0.03003 *
## C2          1.26911    0.64685    1.962    0.04977 *
## SeA         0.03034    0.01887    1.608    0.10777
## SeC1        1.68959    0.73022    2.314    0.02068 *
## SeC2        3.14883    0.71867    4.381 1.18e-05 ***
## AeC1         0.01534    0.01972    0.778    0.43684
## AeC2        -0.04911    0.02492   -1.971    0.04872 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##    Null deviance: 964.52  on 713  degrees of freedom
## Residual deviance: 604.34  on 704  degrees of freedom
## AIC: 624.34
##
## Number of Fisher Scoring iterations: 6
```

3.1. Seleção do modelo

```
fit3_step <- step(fit3)
```

```
summary(fit3_step)
```

```
##
## Call:
## glm(formula = surv ~ sex + age + C1 + C2 + SeA + SeC1 + SeC2 +
##      AeC2, family = binomial(link = "logit"))
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.7807  -0.6323  -0.4211   0.3313   3.2475
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.73944    0.30442  -2.429 0.015141 *
## sex          0.92429    0.46462   1.989 0.046662 *
## age         -0.04046    0.01083  -3.737 0.000186 ***
## C1           1.95168    0.33135   5.890 3.86e-09 ***
## C2           1.41172    0.62148   2.272 0.023115 *
## SeA          0.02465    0.01728   1.426 0.153782
## SeC1         1.74017    0.73538   2.366 0.017964 *
## SeC2         3.19946    0.71883   4.451 8.55e-06 ***
## AeC2        -0.05535    0.02350  -2.356 0.018485 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##    Null deviance: 964.52  on 713  degrees of freedom
```

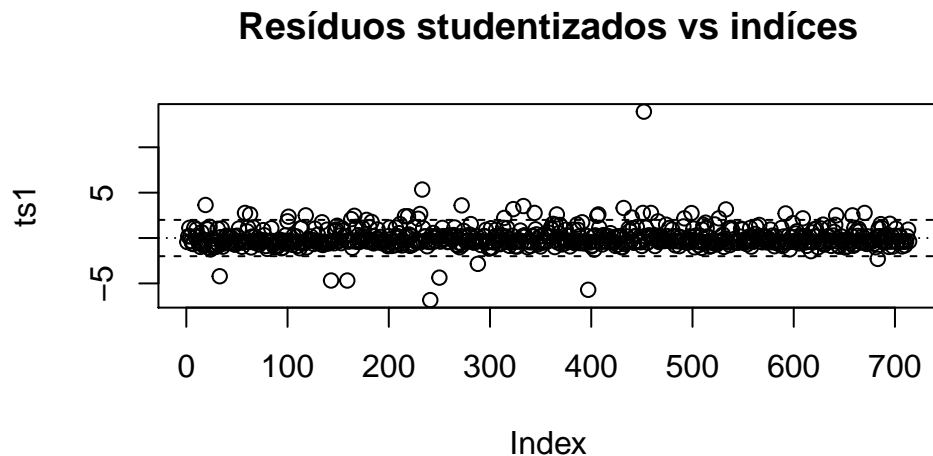
```
## Residual deviance: 604.94 on 705 degrees of freedom
## AIC: 622.94
##
## Number of Fisher Scoring iterations: 6
```

Aplicamos o método *stepwise* com intenção de redução de dimensionalidade no modelo e covariáveis significativas. O método considera como critério de seleção do melhor modelo o AIC. Obtemos que apenas as interações *sexo * idade* e *idade * classe2* não foram significativas a 5%.

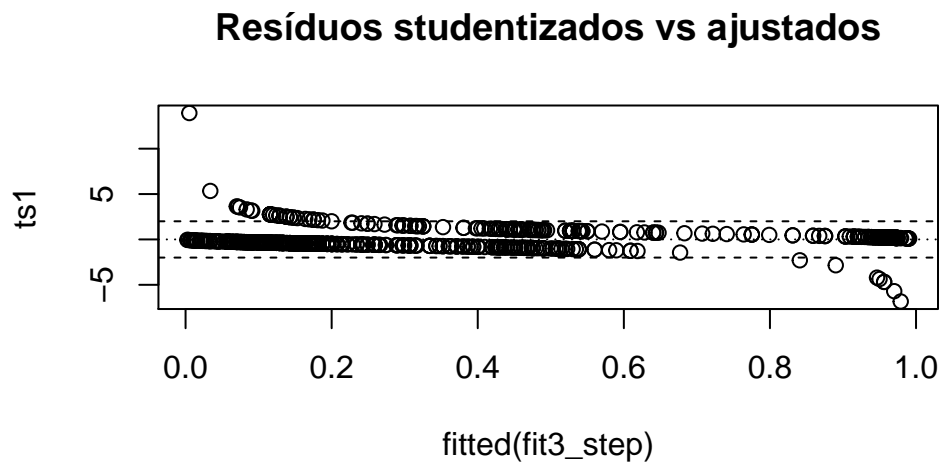
3.2. Análise de diagnóstico e influência fit3_step

3.2.1. Resíduos

```
ts1<-resid(fit3_step,type="pearson") #studentizados
abline(plot(ts1, main='Resíduos studentizados vs índices'), lty=c(2,2,3), h=c(-2,2,0))
```

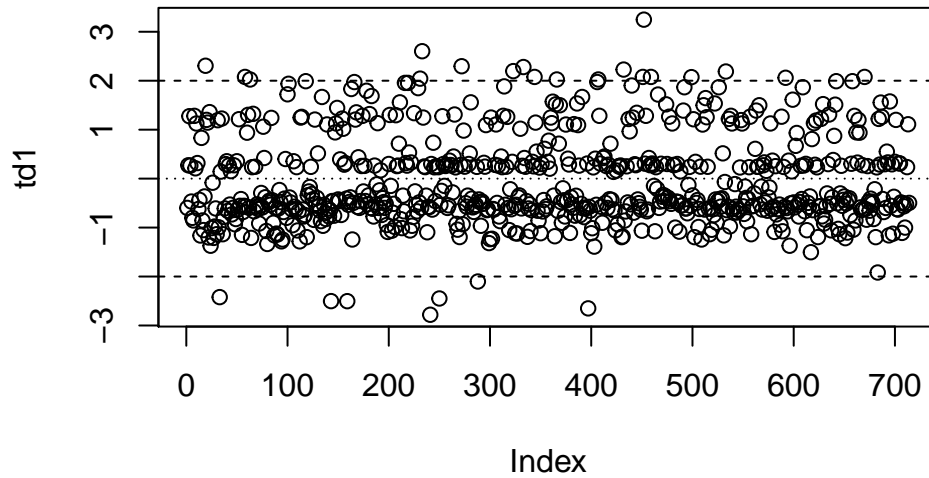


```
abline(plot(fitted(fit3_step),ts1,main='Resíduos studentizados vs ajustados') ,
       lty=c(2,2,3), h=c(-2,2,0))
```



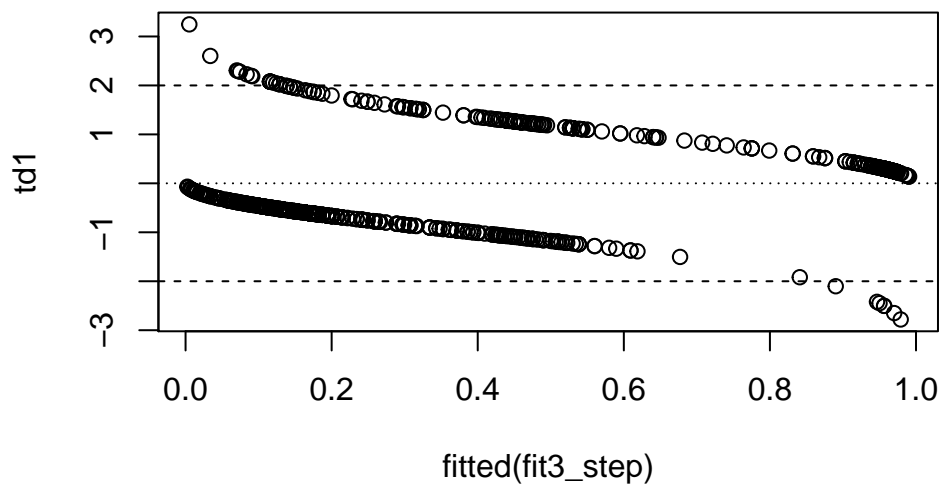
```
td1<-resid(fit3_step,type="deviance") #componente do desvio
abline(plot(td1,main='Resíduos componente do desvio vs índices'),lty=c(2,2,3),h=c(-2,2,0))
```

Resíduos componente do desvio vs índices



```
abline(plot(fitted(fit3_step),td1,main='Resíduos componente do desvio vs ajustados'),
      lty=c(2,2,3), h=c(-2,2,0))
```

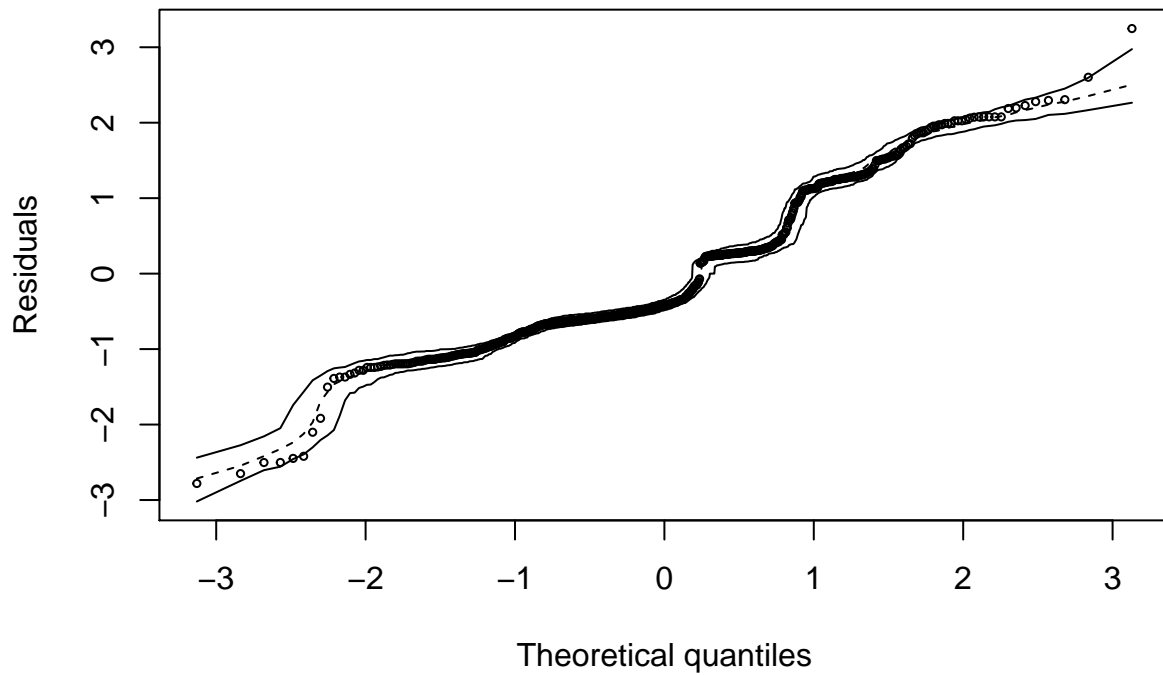
Resíduos componente do desvio vs ajustados



3.2.2. Envelope Simulado

```
hnp(fit3_step, halfnormal = F, conf=0.9)
```

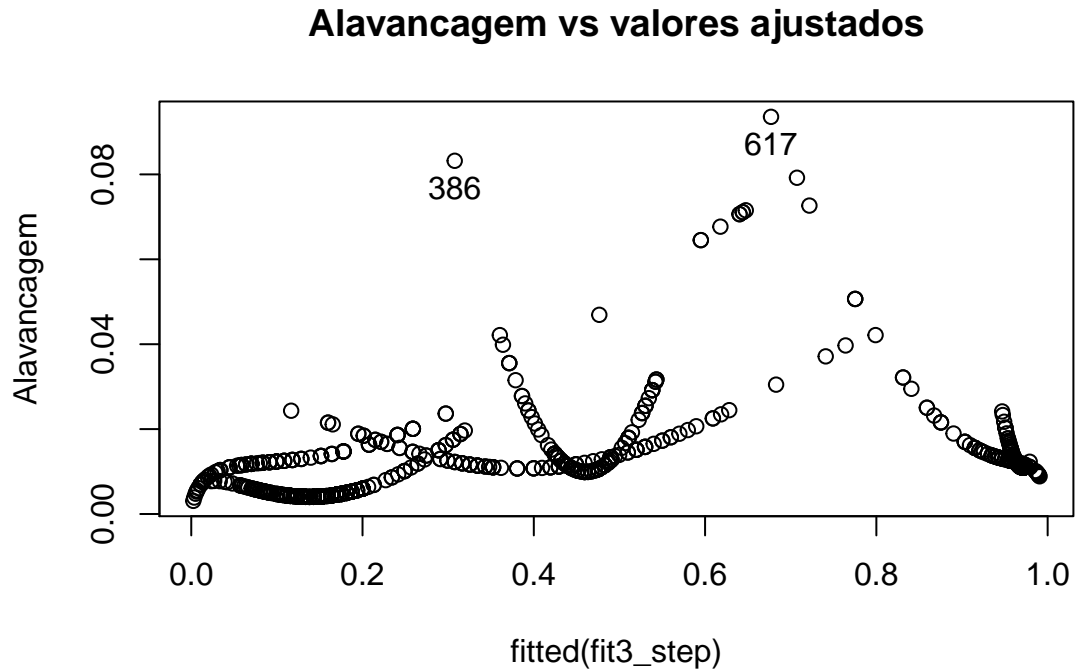
```
## Binomial model
```



Foi detectado pela função `hnp` o modelo binomial. O envelope simulado usado para diagnóstico de ajustes de modelos demonstrou bom ajuste para o modelo `fit3_step`. Os pontos, em grande maioria, se encontram inseridos entre as bandas de 90% confiança.

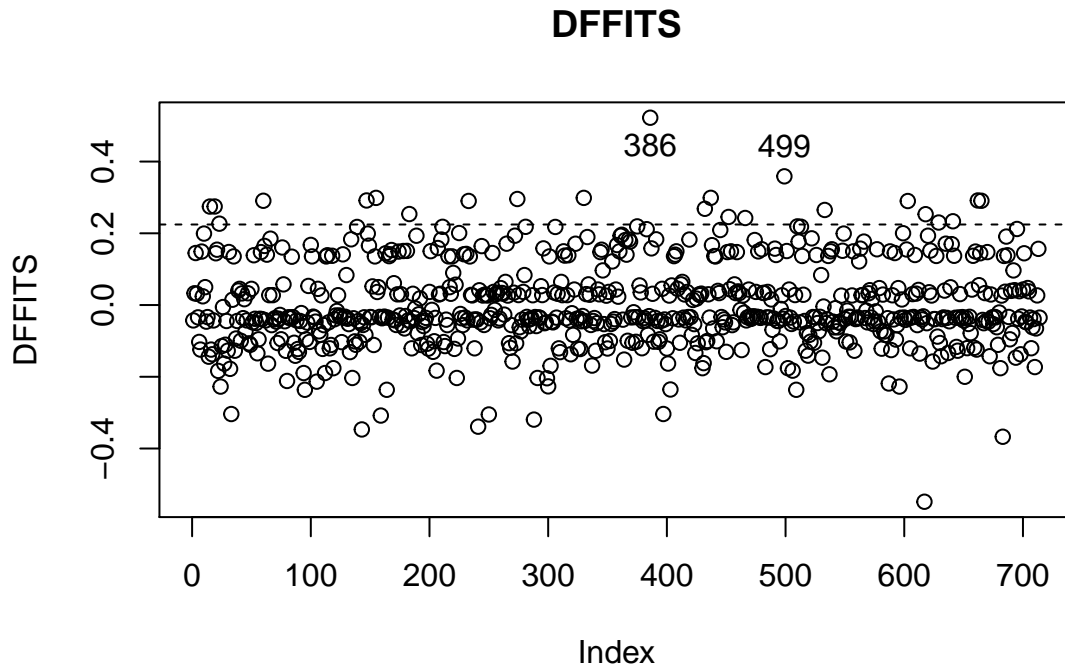
3.2.3. Alavancagem

```
hatvalues(fit3_step)
plot(fitted(fit3_step), hatvalues(fit3_step), ylab="Alavancagem",
     main='Alavancagem vs valores ajustados')
identify(fitted(fit3_step), hatvalues(fit3_step), n=1)
```



3.2.4. DFFIT

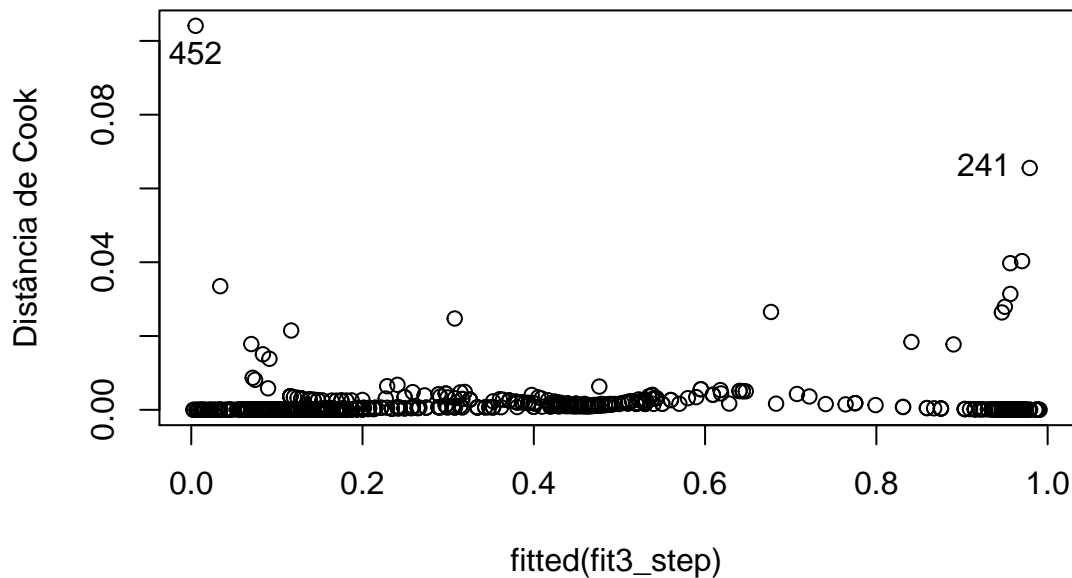
```
dffits(fit3_step)
limite<-2*sqrt(fit3_step$rank / n)
limite
abline(plot(dffits(fit3_step),ylab="DFFITS",main="DFFITS"), h=limite,lty=2)
identify(dffits(fit3_step), n=2)
```



3.2.5. Distância de Cook

```
cooks.distance(fit3_step)
plot(fitted(fit3_step),cooks.distance(fit3_step), ylab="Distância de Cook",
     main="Cook vs valores ajustados")
identify(fitted(fit3_step),cooks.distance(fit3_step), n=2)
```

Cook vs valores ajustados



3.3. Verificação da função de ligação - teste RESET

```
eta.hat<-(fit3_step$linear.predictors)^2
verify.link<-glm(surv~sex+C1+C2+age+SeC1+SeC2+AeC2+eta.hat, family=binomial(link=logit))
summary(verify.link)
```

```
##
## Call:
## glm(formula = surv ~ sex + C1 + C2 + age + SeC1 + SeC2 + AeC2 +
##      eta.hat, family = binomial(link = logit))
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.0293  -0.6222  -0.4576   0.3422   3.0123
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.970687   0.296700  -3.272 0.001069 **
## sex          1.526830   0.327323   4.665 3.09e-06 ***
## C1           1.886190   0.381108   4.949 7.45e-07 ***
## C2           1.454186   0.625105   2.326 0.020002 *
## age         -0.033256   0.009503  -3.500 0.000466 ***
## SeC1         1.888114   1.068022   1.768 0.077084 .
## SeC2         3.419473   0.726954   4.704 2.55e-06 ***
## AeC2        -0.057052   0.023363  -2.442 0.014604 *
## eta.hat       0.021243   0.063926   0.332 0.739659
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 964.52   on 713   degrees of freedom
## Residual deviance: 606.85   on 705   degrees of freedom
## AIC: 624.85
##
## Number of Fisher Scoring iterations: 6
```

Inserimos a covariável η^2 e testamos se o parâmetro associado à essa covariável é igual a zero. Ela não foi significativa para o modelo, ou seja, não rejeitamos a hipótese nula e concluímos que

$\hat{\eta}^2 = 0$. Portanto, o modelo está corretamente especificado.

3.4. Qualidade do ajuste

```
source("pseudo_R2_disc.r")
pseudo_R2(fit3_step)
```

```
## [1] pseudo R2 = 0.534
```

O modelo foi capaz de explicar 53,4% da variabilidade dos dados.

4. Modelo fit3 sem outliers

Irei testar como ficaria esse ultimo modelo ajustado sem as possíveis observações influentes.

Retirando observações potencialmente influentes

```
new_age<-age[-c(368,452,499)]
new_sex<-sex[-c(368,452,499)]
new_C1<-C1[-c(368,452,499)]
new_C2<-C2[-c(368,452,499)]
new_surv<-surv[-c(368,452,499)]
```

Novas interações

```
new_SeC1<-new_sex*new_C1
new_SeC2<-new_sex*new_C2
new_AeC2<-new_age*new_C2
```

Modelo full sem outliers

```
final_fit<-glm(new_surv~new_sex+new_C1+new_C2+new_age+new_SeC1+new_SeC2+new_AeC2, family=binomial(link=
#source("diag_binomial.r")
#diag_binomial(final_fit)
summary(final_fit)
```

```
##
## Call:
## glm(formula = new_surv ~ new_sex + new_C1 + new_C2 + new_age +
##      new_SeC1 + new_SeC2 + new_AeC2, family = binomial(link = logit))
```



```

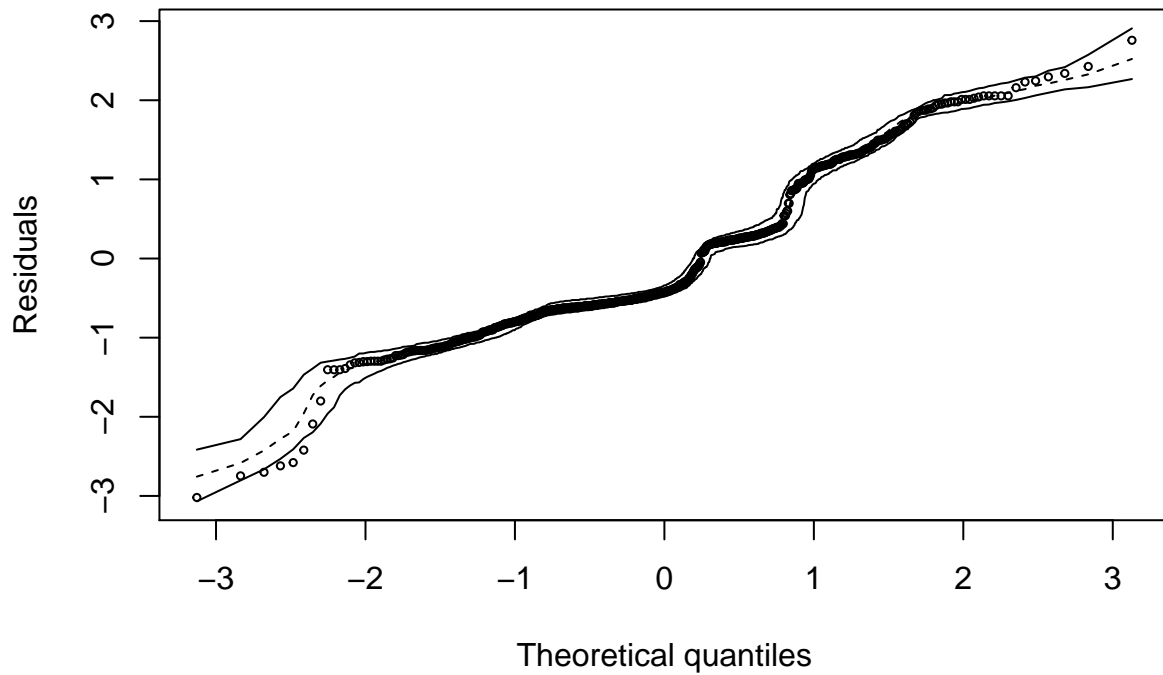
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -3.0196  -0.6259  -0.4295   0.3284   2.7580
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.868571   0.274749  -3.161 0.001571 **
## new_sex      1.459674   0.270578   5.395 6.87e-08 ***
## new_C1       1.819929   0.312334   5.827 5.65e-09 ***
## new_C2       1.752247   0.655029   2.675 0.007471 **
## new_age     -0.034847   0.009017  -3.865 0.000111 ***
## new_SeC1     2.207269   0.683059   3.231 0.001232 **
## new_SeC2     3.943332   0.791951   4.979 6.38e-07 ***
## new_AeC2    -0.076212   0.025869  -2.946 0.003219 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 959.09  on 710  degrees of freedom
## Residual deviance: 591.51  on 703  degrees of freedom
## AIC: 607.51
##
## Number of Fisher Scoring iterations: 6

```

Envelope Simulado

```
hnp(final_fit, halfnormal=F, conf=0.9)
```

```
## Binomial model
```



Houve uma pequena melhora do modelo `fit3_step` para o `final_fit` onde apenas 3 observações influentes foram retiradas e o ajuste refeito. Mais pontos se mantiveram inseridos entre as bandas de 90% de confiança.

Pseudo R^2

```
source("pseudo_R2_disc.r")  
pseudo_R2(final_fit)
```

```
## [1] pseudo R2 = 0.5452
```

Em termos de qualidade de ajuste, com a remoção das observações influentes obteve-se pequeno aumento da explicação da variabilidade dos dados pelo modelo final ajustado (54,5%).

Não seria viável avaliar resíduos, influência dado que o processo já foi considerado OK no modelo anterior.

5. Probabilidade de sobrevivência

Segundo site: Para calcular a probabilidade de sobrevivência de um indivíduo, $E(Y)$, é necessário aplicar uma função inversa a função de ligação, sendo representada por:

$$E(Y) = \frac{e^{g(x)}}{1 + e^{g(x)}}$$

O modelo final é dado por:

$$g(X) = -0.8686 + Sexo \times 1.4597 + C1 \times 1.8199 + C2 \times 1.7522 - Idade \times 0.0349 \\ + (Sexo \times C1) \times 2.2073 + (Sexo \times C2) \times 3.9433 - (Idade \times C2) \times 0.076,$$

em que $sexofem = 1$ e $sexomasc = 0$.

Minha probabilidade de sobrevivência estando na 1 classe seria:

```
g <- -0.868571 + 0*1.459674 + 1*1.819929 + 0*1.752247 - 22*0.0349 + 0*1*2.207 + 0*0*3.9433 - 22*0*0.0762*  
exp(g)/(1+exp(g))  
## [1] 0.5457611
```

O valor retornado pelo site foi de 59,1% nessas mesmas condições.

Minha probabilidade de sobrevivência estando na 2 classe seria:

```
g <- -0.868571 + 0*1.459674 + 0*1.819929 + 1*1.752247 - 22*0.0349 + 0*0*2.207 + 0*1*3.9433 - 22*1*0.0762*  
exp(g)/(1+exp(g))  
## [1] 0.1735715
```