M5_AI2_CANOJORGE

JORGE CANO

2024-10-12

```
## package 'gamlss' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\txell\AppData\Local\Temp\RtmpYdfdaX\downloaded_packages

## package 'ggplot2' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\txell\AppData\Local\Temp\RtmpYdfdaX\downloaded_packages

## 'data.frame':    1000 obs. of  21 variables:
##  $ chk_acct        : chr  "A11" "A12" "A14" "A11" ...
##  $ duration        : int  6 48 12 42 24 36 24 36 12 30 ...
##  $ credit_his      : chr  "A34" "A32" "A34" "A32" ...
##  $ purpose         : chr  "A43" "A43" "A46" "A42" ...
##  $ amount          : int  1169 5951 2096 7882 4870 9055 2835 6948 3059 5234 ...
##  $ saving_acct     : chr  "A65" "A61" "A61" "A61" ...
##  $ present_emp     : chr  "A75" "A73" "A74" "A74" ...
##  $ installment_rate: int  4 2 2 2 3 2 3 2 2 4 ...
##  $ sex             : chr  "A93" "A92" "A93" "A93" ...
##  $ other_debtor    : chr  "A101" "A101" "A101" "A103" ...
##  $ present_resid   : int  4 2 3 4 4 4 4 2 4 2 ...
##  $ property        : chr  "A121" "A121" "A121" "A122" ...
##  $ age             : int  67 22 49 45 53 35 53 35 61 28 ...
##  $ other_install   : chr  "A143" "A143" "A143" "A143" ...
##  $ housing         : chr  "A152" "A152" "A152" "A153" ...
##  $ n_credits       : int  2 1 1 1 2 1 1 1 1 2 ...
##  $ job             : chr  "A173" "A173" "A172" "A173" ...
##  $ n_people        : int  1 1 2 2 2 2 1 1 1 1 ...
##  $ telephone       : chr  "A192" "A191" "A191" "A191" ...
##  $ foreign         : chr  "A201" "A201" "A201" "A201" ...
##  $ response        : Factor w/ 2 levels "0","1": 1 2 1 1 2 1 1 1 1 2 ...

## [1] 0
```

1. Propón un modelo lineal logit en el que la variable respuesta (crédito bueno=0, crédito malo=1), lo expliquen el resto de variables.

```
##
## Call:
```

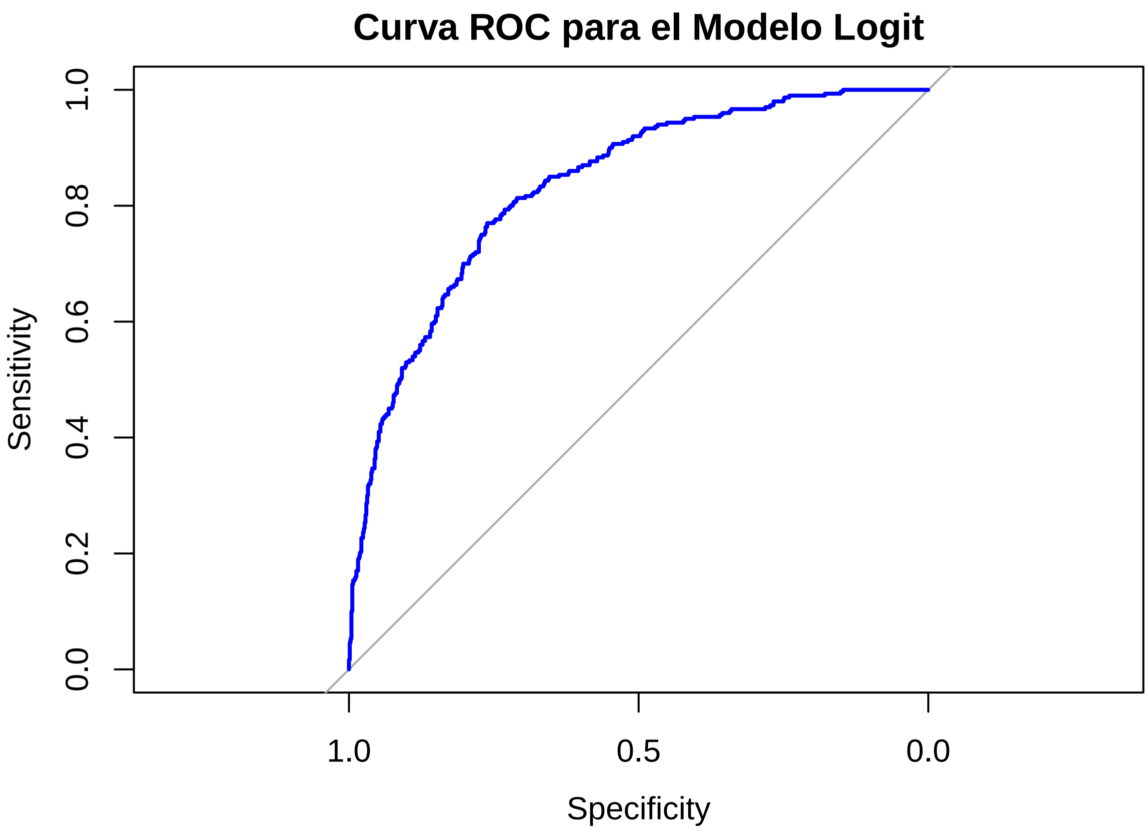
```
## glm(formula = response ~ ., family = binomial(link = "logit"),
##     data = german_credit)
##
## Coefficients:
##                      Estimate Std. Error z value Pr(>|z|)
## (Intercept)         4.005e-01  1.084e+00   0.369 0.711869
## chk_acctA12        -3.749e-01  2.179e-01  -1.720 0.085400 .
## chk_acctA13        -9.657e-01  3.692e-01  -2.616 0.008905 **
## chk_acctA14        -1.712e+00  2.322e-01  -7.373 1.66e-13 ***
## duration            2.786e-02  9.296e-03   2.997 0.002724 **
## credit_hisA31       1.434e-01  5.489e-01   0.261 0.793921
## credit_hisA32      -5.861e-01  4.305e-01  -1.362 0.173348
## credit_hisA33      -8.532e-01  4.717e-01  -1.809 0.070470 .
## credit_hisA34      -1.436e+00  4.399e-01  -3.264 0.001099 **
## purposeA41         -1.666e+00  3.743e-01  -4.452 8.51e-06 ***
## purposeA410        -1.489e+00  7.764e-01  -1.918 0.055163 .
## purposeA42         -7.916e-01  2.610e-01  -3.033 0.002421 **
## purposeA43         -8.916e-01  2.471e-01  -3.609 0.000308 ***
## purposeA44         -5.228e-01  7.623e-01  -0.686 0.492831
## purposeA45         -2.164e-01  5.500e-01  -0.393 0.694000
## purposeA46          3.628e-02  3.965e-01   0.092 0.927082
## purposeA48         -2.059e+00  1.212e+00  -1.699 0.089297 .
## purposeA49         -7.401e-01  3.339e-01  -2.216 0.026668 *
## amount              1.283e-04  4.444e-05   2.887 0.003894 **
## saving_acctA62     -3.577e-01  2.861e-01  -1.250 0.211130
## saving_acctA63     -3.761e-01  4.011e-01  -0.938 0.348476
## saving_acctA64     -1.339e+00  5.249e-01  -2.551 0.010729 *
## saving_acctA65     -9.467e-01  2.625e-01  -3.607 0.000310 ***
## present_empA72     -6.691e-02  4.270e-01  -0.157 0.875475
## present_empA73     -1.828e-01  4.105e-01  -0.445 0.656049
## present_empA74     -8.310e-01  4.455e-01  -1.866 0.062110 .
## present_empA75     -2.766e-01  4.134e-01  -0.669 0.503410
## installment_rate    3.301e-01  8.828e-02   3.739 0.000185 ***
## sexA92             -2.755e-01  3.865e-01  -0.713 0.476040
## sexA93             -8.161e-01  3.799e-01  -2.148 0.031718 *
## sexA94             -3.671e-01  4.537e-01  -0.809 0.418448
## other_debtorA102    4.360e-01  4.101e-01   1.063 0.287700
## other_debtorA103   -9.786e-01  4.243e-01  -2.307 0.021072 *
## present_resid       4.776e-03  8.641e-02   0.055 0.955920
## propertyA122        2.814e-01  2.534e-01   1.111 0.266630
## propertyA123        1.945e-01  2.360e-01   0.824 0.409743
## propertyA124        7.304e-01  4.245e-01   1.721 0.085308 .
## age                -1.454e-02  9.222e-03  -1.576 0.114982
## other_installA142 -1.232e-01  4.119e-01  -0.299 0.764878
## other_installA143 -6.463e-01  2.391e-01  -2.703 0.006871 **
## housingA152        -4.436e-01  2.347e-01  -1.890 0.058715 .
## housingA153        -6.839e-01  4.770e-01  -1.434 0.151657
## n_credits           2.721e-01  1.895e-01   1.436 0.151109
## jobA172             5.361e-01  6.796e-01   0.789 0.430160
## jobA173             5.547e-01  6.549e-01   0.847 0.397015
## jobA174             4.795e-01  6.623e-01   0.724 0.469086
## n_people            2.647e-01  2.492e-01   1.062 0.288249
## telephoneA192      -3.000e-01  2.013e-01  -1.491 0.136060
## foreignA202        -1.392e+00  6.258e-01  -2.225 0.026095 *
```

```
## ---
## 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:  895.82  on 951  degrees of freedom
## AIC: 993.82
##
## Number of Fisher Scoring iterations: 5
```

Para ver el grado de ajuste del modelo, calculamos la curva ROC y el AUC, obsevando que el ajuste es bastante bueno, concretamente con un AUC de (0.834), lo que nos sugiere que al ser valor cercano a 1 tenemos una mejor predicción que un modelo aleatorio. Una curva que se acerca más a la esquina superior izquierda (donde sensibilidad y especificidad son ambas altas) indica un mejor rendimiento de clasificación del modelo.

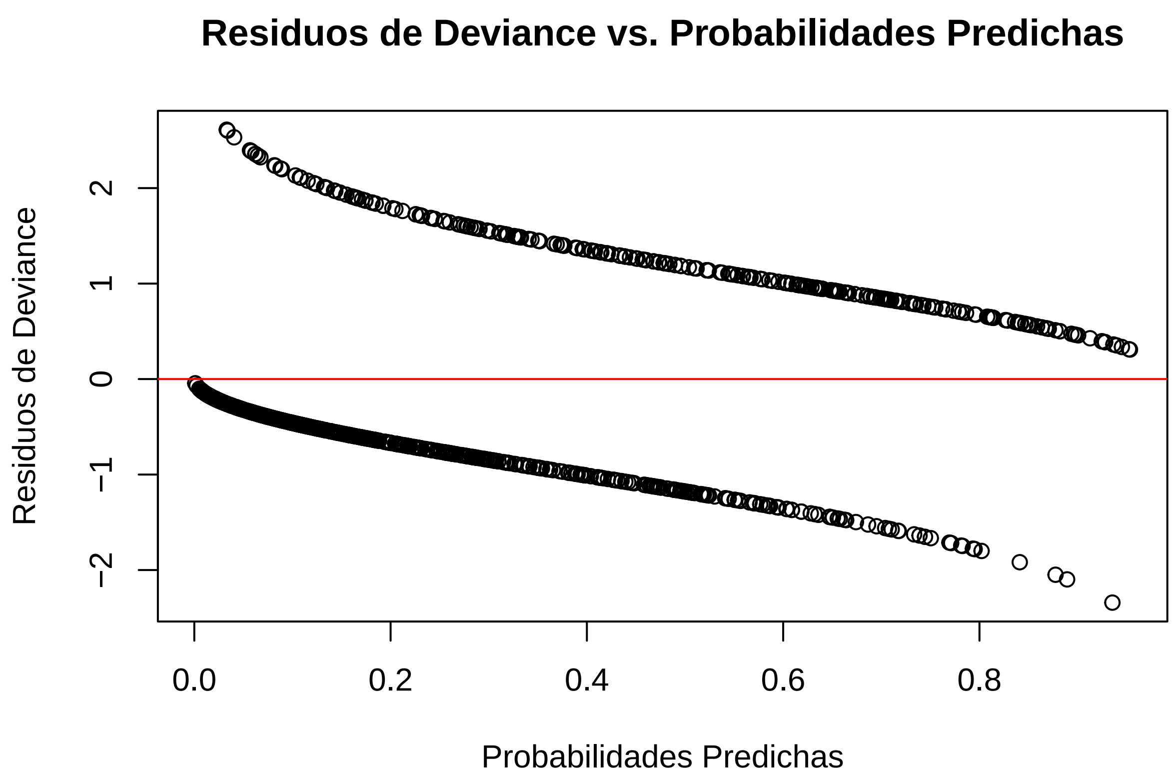
```
## package 'pROC' successfully unpacked and MD5 sums checked
```

```
##
## The downloaded binary packages are in
##  C:\Users\txell\AppData\Local\Temp\RtmpYdfdaX\downloaded_packages
```




```
## Area under the curve: 0.8338
```

Analisis de los residuos del modelo vs las probabilidades predichas



2. Interpreta la variable duration. ¿Es significativa? ¿A partir de qué nivel de significación deja de ser significativa?

Dado que la variable mantiene una elevada correlación con otras variables (0.65 con amount) y no está mostrando los resultado, forzamos que salga informados informando en la GLM las variables a considerar. Para confirmar esta teoría aplicamos la matriz de correlación:

```
##                    duration      amount installment_rate present_resid
## duration         1.00000000  0.62498420       0.07474882    0.03406720
## amount           0.62498420  1.00000000      -0.27131570    0.02892632
## installment_rate  0.07474882 -0.27131570       1.00000000    0.04930237
## present_resid     0.03406720  0.02892632       0.04930237    1.00000000
## age              -0.03613637  0.03271642       0.05826568    0.26641918
## n_credits        -0.01128360  0.02079455       0.02166874    0.08962523
## n_people         -0.02383448  0.01714215      -0.07120694    0.04264343
##                         age   n_credits    n_people
## duration         -0.03613637 -0.01128360 -0.02383448
## amount            0.03271642  0.02079455  0.01714215
```

```
## installment_rate  0.05826568  0.02166874 -0.07120694
## present_resid     0.26641918  0.08962523  0.04264343
## age               1.00000000  0.14925358  0.11820083
## n_credits         0.14925358  1.00000000  0.10966670
## n_people          0.11820083  0.10966670  1.00000000
```

```
##    Estimate  Std. Error     z value    Pr(>|z|)
## 0.027863324 0.009296278 2.997256078 0.002724218
```

Para la variable duration, observamos un P valor de 0.0027, siendo inferior a P<0.05 podemos considerarlo significativa la variable. En concreto con este valor lo podemos considerar muy significativo, lo que representa que valor sea producto del azar es muy bajo. Para contestar el nivel de consideración de significación debemos contextualizar el estudio que estemos realizando, siendo más críticos en ensayos médicos que en otro tipo de estudios no tan determinantes, donde podemos aceptar P valores < a 0.1.

Otro contraste que podemos realizar es analizando el z value, donde con un valor cercano a 3 (2.99) nos indica una alta significancia con el coeficiente. Por otro lado el signo positivo, nos indica que la observación se encuentra por encima del valor de la media.

3. Si eliminamos la variable amount del modelo, ¿crees que alguna otra variable incrementaría el sesgo provocado por la falta de amount en el modelo? Es decir, identifica el sesgo en otra variable producido por eliminar la variable amount.

Si omitimos la variable amount del modelo, observamos un cambio significativo en el intercepto, lo que implica un potencial sesgo al incluir esta variable.

A continuación muestro en formato tabla y gráficamente los impactos de omitir la variable.

```
##
## Call:
## glm(formula = response ~ ., family = binomial(link = "logit"),
##     data = german_credit)
##
## Coefficients:
##                      Estimate Std. Error z value Pr(>|z|)
## (Intercept)         4.005e-01  1.084e+00   0.369 0.711869
## chk_acctA12        -3.749e-01  2.179e-01  -1.720 0.085400 .
```

```
## chk_acctA13       -9.657e-01  3.692e-01  -2.616 0.008905 **
## chk_acctA14       -1.712e+00  2.322e-01  -7.373 1.66e-13 ***
## duration           2.786e-02  9.296e-03   2.997 0.002724 **
## credit_hisA31      1.434e-01  5.489e-01   0.261 0.793921
## credit_hisA32     -5.861e-01  4.305e-01  -1.362 0.173348
## credit_hisA33     -8.532e-01  4.717e-01  -1.809 0.070470 .
## credit_hisA34     -1.436e+00  4.399e-01  -3.264 0.001099 **
## purposeA41        -1.666e+00  3.743e-01  -4.452 8.51e-06 ***
## purposeA410       -1.489e+00  7.764e-01  -1.918 0.055163 .
## purposeA42        -7.916e-01  2.610e-01  -3.033 0.002421 **
## purposeA43        -8.916e-01  2.471e-01  -3.609 0.000308 ***
## purposeA44        -5.228e-01  7.623e-01  -0.686 0.492831
## purposeA45        -2.164e-01  5.500e-01  -0.393 0.694000
## purposeA46         3.628e-02  3.965e-01   0.092 0.927082
## purposeA48        -2.059e+00  1.212e+00  -1.699 0.089297 .
## purposeA49        -7.401e-01  3.339e-01  -2.216 0.026668 *
## amount             1.283e-04  4.444e-05   2.887 0.003894 **
## saving_acctA62    -3.577e-01  2.861e-01  -1.250 0.211130
## saving_acctA63    -3.761e-01  4.011e-01  -0.938 0.348476
## saving_acctA64    -1.339e+00  5.249e-01  -2.551 0.010729 *
## saving_acctA65    -9.467e-01  2.625e-01  -3.607 0.000310 ***
## present_empA72    -6.691e-02  4.270e-01  -0.157 0.875475
## present_empA73    -1.828e-01  4.105e-01  -0.445 0.656049
## present_empA74    -8.310e-01  4.455e-01  -1.866 0.062110 .
## present_empA75    -2.766e-01  4.134e-01  -0.669 0.503410
## installment_rate   3.301e-01  8.828e-02   3.739 0.000185 ***
## sexA92            -2.755e-01  3.865e-01  -0.713 0.476040
## sexA93            -8.161e-01  3.799e-01  -2.148 0.031718 *
## sexA94            -3.671e-01  4.537e-01  -0.809 0.418448
## other_debtorA102   4.360e-01  4.101e-01   1.063 0.287700
## other_debtorA103  -9.786e-01  4.243e-01  -2.307 0.021072 *
## present_resid      4.776e-03  8.641e-02   0.055 0.955920
## propertyA122       2.814e-01  2.534e-01   1.111 0.266630
## propertyA123       1.945e-01  2.360e-01   0.824 0.409743
## propertyA124       7.304e-01  4.245e-01   1.721 0.085308 .
## age               -1.454e-02  9.222e-03  -1.576 0.114982
## other_installA142 -1.232e-01  4.119e-01  -0.299 0.764878
## other_installA143 -6.463e-01  2.391e-01  -2.703 0.006871 **
## housingA152       -4.436e-01  2.347e-01  -1.890 0.058715 .
## housingA153       -6.839e-01  4.770e-01  -1.434 0.151657
## n_credits          2.721e-01  1.895e-01   1.436 0.151109
## jobA172            5.361e-01  6.796e-01   0.789 0.430160
## jobA173            5.547e-01  6.549e-01   0.847 0.397015
## jobA174            4.795e-01  6.623e-01   0.724 0.469086
## n_people           2.647e-01  2.492e-01   1.062 0.288249
## telephoneA192     -3.000e-01  2.013e-01  -1.491 0.136060
## foreignA202       -1.392e+00  6.258e-01  -2.225 0.026095 *
## ---
## 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:  895.82  on 951  degrees of freedom
```

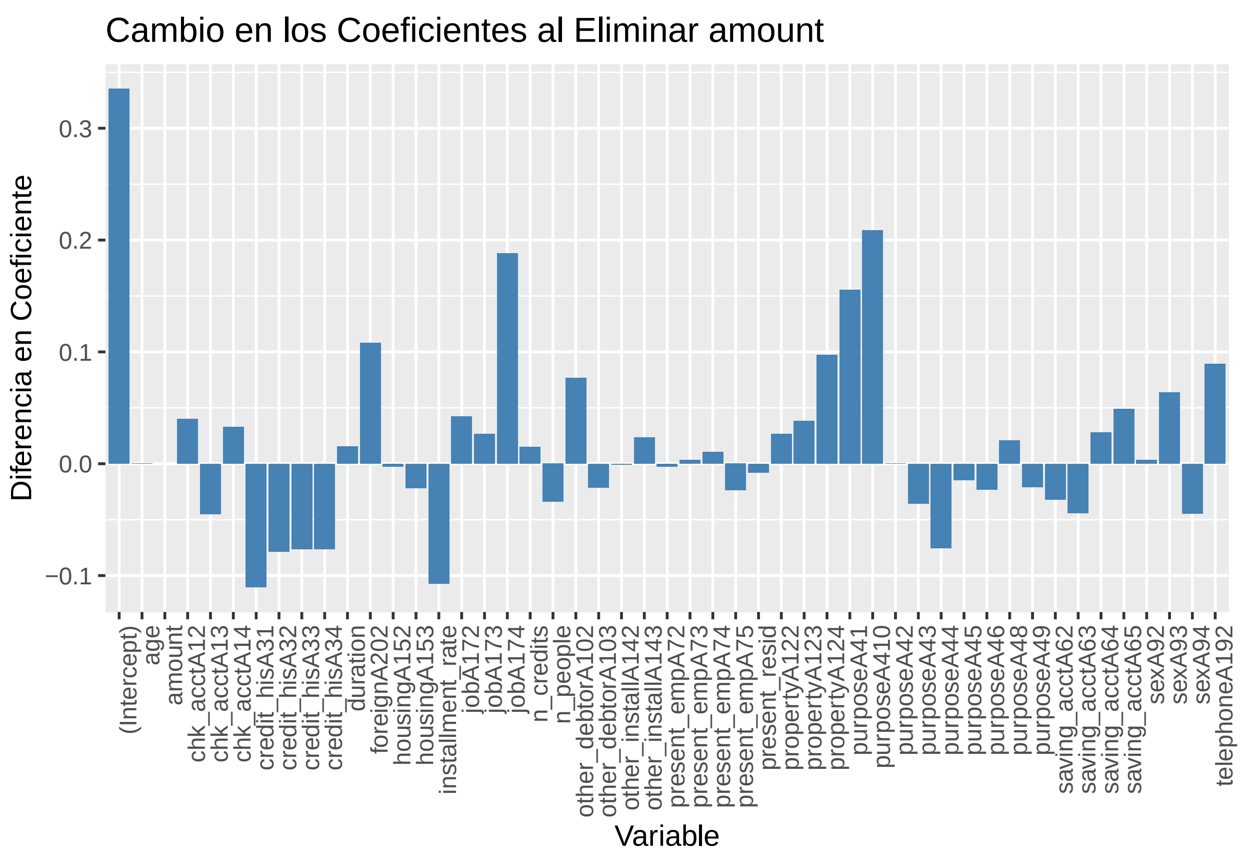
```
## AIC: 993.82
##
## Number of Fisher Scoring iterations: 5
```

```
##
## Call:
## glm(formula = response ~ . - amount, family = binomial(link = "logit"),
##     data = german_credit)
##
## Coefficients:
##                      Estimate Std. Error z value Pr(>|z|)
## (Intercept)          0.735577   1.084058   0.679 0.497429
## chk_acctA12         -0.334793   0.215930  -1.550 0.121030
## chk_acctA13         -1.010444   0.367713  -2.748 0.005998 **
## chk_acctA14         -1.679050   0.230131  -7.296 2.96e-13 ***
## duration             0.043200   0.007700   5.610 2.02e-08 ***
## credit_hisA31        0.033121   0.543002   0.061 0.951362
## credit_hisA32       -0.664639   0.425722  -1.561 0.118475
## credit_hisA33       -0.929478   0.467501  -1.988 0.046792 *
## credit_hisA34       -1.511875   0.435588  -3.471 0.000519 ***
## purposeA41          -1.510888   0.363500  -4.157 3.23e-05 ***
## purposeA410         -1.280260   0.719681  -1.779 0.075252 .
## purposeA42          -0.791414   0.260204  -3.042 0.002354 **
## purposeA43          -0.927151   0.246018  -3.769 0.000164 ***
## purposeA44          -0.598289   0.762031  -0.785 0.432381
## purposeA45          -0.231046   0.553188  -0.418 0.676194
## purposeA46           0.013150   0.392597   0.033 0.973281
## purposeA48          -2.038610   1.200515  -1.698 0.089486 .
## purposeA49          -0.761048   0.331502  -2.296 0.021690 *
## saving_acctA62      -0.389630   0.285663  -1.364 0.172584
## saving_acctA63      -0.420244   0.398863  -1.054 0.292063
## saving_acctA64      -1.311038   0.517790  -2.532 0.011342 *
## saving_acctA65      -0.897587   0.260475  -3.446 0.000569 ***
## present_empA72      -0.069255   0.424366  -0.163 0.870364
## present_empA73      -0.179541   0.405985  -0.442 0.658318
## present_empA74      -0.820499   0.442087  -1.856 0.063458 .
## present_empA75      -0.300354   0.408312  -0.736 0.461974
## installment_rate     0.222983   0.079244   2.814 0.004895 **
## sexA92              -0.271900   0.382754  -0.710 0.477470
## sexA93              -0.752239   0.375267  -2.005 0.045012 *
## sexA94              -0.411578   0.450180  -0.914 0.360584
## other_debtorA102     0.512717   0.405800   1.263 0.206420
## other_debtorA103    -1.000046   0.423580  -2.361 0.018229 *
## present_resid       -0.002927   0.085961  -0.034 0.972835
## propertyA122         0.307907   0.251806   1.223 0.221408
## propertyA123         0.232759   0.234427   0.993 0.320766
## propertyA124         0.827707   0.417375   1.983 0.047353 *
## age                 -0.014147   0.009149  -1.546 0.122061
## other_installA142   -0.123686   0.411643  -0.300 0.763819
## other_installA143   -0.622631   0.237472  -2.622 0.008744 **
## housingA152         -0.446099   0.233362  -1.912 0.055926 .
## housingA153         -0.705536   0.471264  -1.497 0.134364
## n_credits            0.287224   0.188962   1.520 0.128509
## jobA172              0.578398   0.686712   0.842 0.399637
```

```
## jobA173            0.581191   0.662886   0.877 0.380618
## jobA174            0.667596   0.666836   1.001 0.316759
## n_people           0.230638   0.248814   0.927 0.353954
## telephoneA192     -0.210575   0.196510  -1.072 0.283911
## foreignA202       -1.284315   0.607564  -2.114 0.034526 *
## ---
## 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:  904.28  on 952  degrees of freedom
## AIC: 1000.3
##
## Number of Fisher Scoring iterations: 5
```

```
##                             Variable    Con_Amount   Sin_Amount    Diferencia
## (Intercept)              (Intercept)  0.4005027032  0.735576808  0.3350741049
## chk_acctA12              chk_acctA12 -0.3748533845 -0.334792817  0.0400605671
## chk_acctA13              chk_acctA13 -0.9656768269 -1.010443588 -0.0447667615
## chk_acctA14              chk_acctA14 -1.7118879549 -1.679050243  0.0328377117
## duration                    duration  0.0278633245  0.043199600  0.0153362752
## credit_hisA31          credit_hisA31  0.1433777014  0.033121015 -0.1102566865
## credit_hisA32          credit_hisA32 -0.5861135632 -0.664638913 -0.0785253501
## credit_hisA33          credit_hisA33 -0.8531614098 -0.929477993 -0.0763165831
## credit_hisA34          credit_hisA34 -1.4357715801 -1.511875466 -0.0761038860
## purposeA41               purposeA41 -1.6664669545 -1.510888106  0.1555788483
## purposeA410             purposeA410 -1.4887859369 -1.280260167  0.2085257702
## purposeA42               purposeA42 -0.7916103762 -0.791413629  0.0001967473
## purposeA43               purposeA43 -0.8915834370 -0.927151139 -0.0355677019
## purposeA44               purposeA44 -0.5227827424 -0.598289376 -0.0755066339
## purposeA45               purposeA45 -0.2163959040 -0.231045730 -0.0146498256
## purposeA46               purposeA46  0.0362838335  0.013149514 -0.0231343195
## purposeA48               purposeA48 -2.0594327737 -2.038610249  0.0208225248
## purposeA49               purposeA49 -0.7400868495 -0.761047613 -0.0209607631
## amount                       amount  0.0001282747           NA            NA
## saving_acctA62        saving_acctA62 -0.3577405779 -0.389629633 -0.0318890546
## saving_acctA63        saving_acctA63 -0.3760728784 -0.420244289 -0.0441714105
## saving_acctA64        saving_acctA64 -1.3391988399 -1.311038154  0.0281606863
## saving_acctA65        saving_acctA65 -0.9466891929 -0.897587288  0.0491019051
## present_empA72        present_empA72 -0.0669104342 -0.069254850 -0.0023444157
## present_empA73        present_empA73 -0.1828309822 -0.179541334  0.0032896486
## present_empA74        present_empA74 -0.8310018182 -0.820498623  0.0105031949
## present_empA75        present_empA75 -0.2766245208 -0.300354492 -0.0237299714
## installment_rate    installment_rate  0.3300898152  0.222983422 -0.1071063931
## sexA92                       sexA92 -0.2754548085 -0.271900010  0.0035547985
## sexA93                       sexA93 -0.8160779448 -0.752238828  0.0638391167
## sexA94                       sexA94 -0.3670718835 -0.411578236 -0.0445063520
## other_debtorA102  other_debtorA102  0.4360476126  0.512717088  0.0766694755
## other_debtorA103  other_debtorA103 -0.9786160157 -1.000046249 -0.0214302337
## present_resid          present_resid  0.0047760501 -0.002927226 -0.0077032760
## propertyA122            propertyA122  0.2814382403  0.307906551  0.0264683109
## propertyA123            propertyA123  0.1945346780  0.232758944  0.0382242661
## propertyA124            propertyA124  0.7304477374  0.827707467  0.0972597296
```

```
## age                                  age -0.0145354910 -0.014146515  0.0003889759
## other_installA142 other_installA142 -0.1232005664 -0.123686251 -0.0004856844
## other_installA143 other_installA143 -0.6463286585 -0.622631235  0.0236974238
## housingA152             housingA152 -0.4436209848 -0.446099212 -0.0024782272
## housingA153             housingA153 -0.6838601772 -0.705535928 -0.0216757506
## n_credits                 n_credits  0.2720759275  0.287223775  0.0151478478
## jobA172                     jobA172  0.5361303832  0.578397688  0.0422673048
## jobA173                     jobA173  0.5547174978  0.581190704  0.0264732061
## jobA174                     jobA174  0.4794752439  0.667596230  0.1881209858
## n_people                   n_people  0.2646713870  0.230638044 -0.0340333429
## telephoneA192         telephoneA192 -0.3000079729 -0.210575325  0.0894326482
## foreignA202             foreignA202 -1.3922159416 -1.284315079  0.1079008626
```

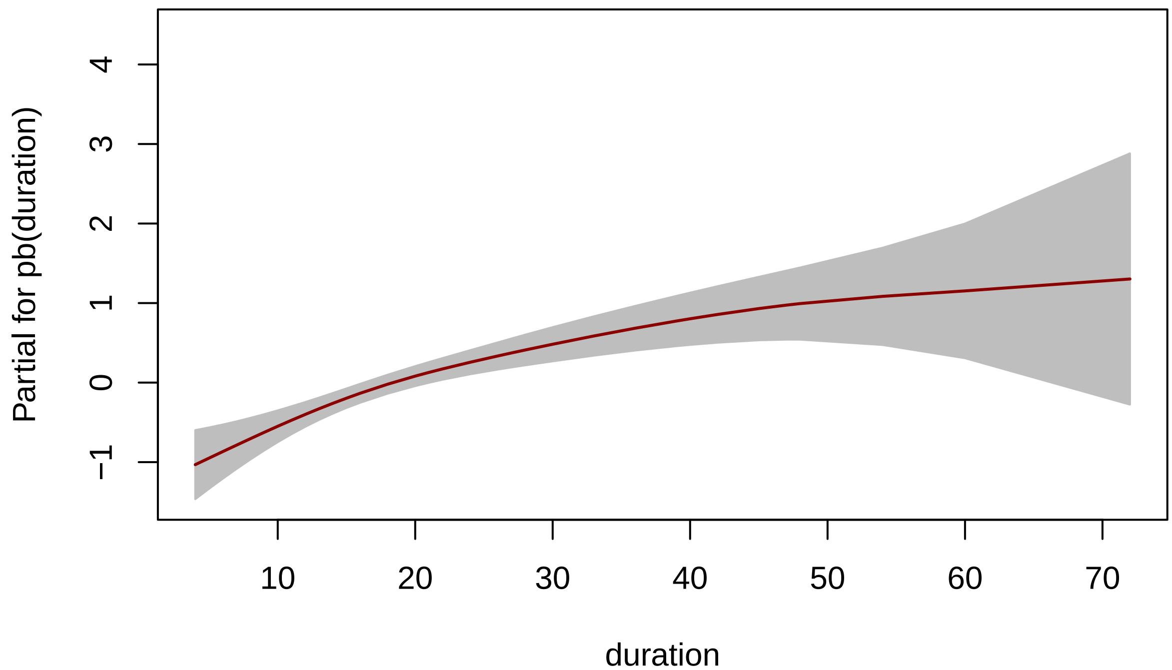


4. Identifica efectos no lineales en la variable duration y amount. Interpreta los nuevos resultados después de meter, en el modelo, estas no linealidades.

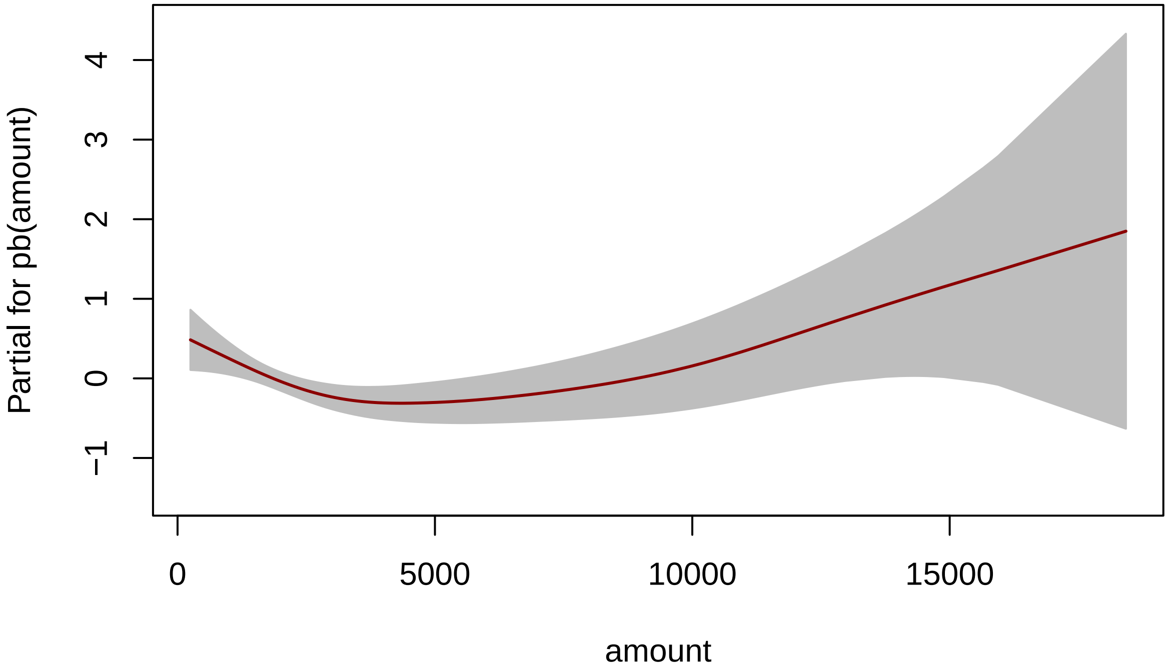
```
## GAMLSS-RS iteration 1: Global Deviance = 1156.908
## GAMLSS-RS iteration 2: Global Deviance = 1156.908

## ******************************************************************
## Family:  c("BI", "Binomial")
##
## Call:  gamlss(formula = response ~ pb(duration) + pb(amount),
##     family = BI, data = german_credit)
```

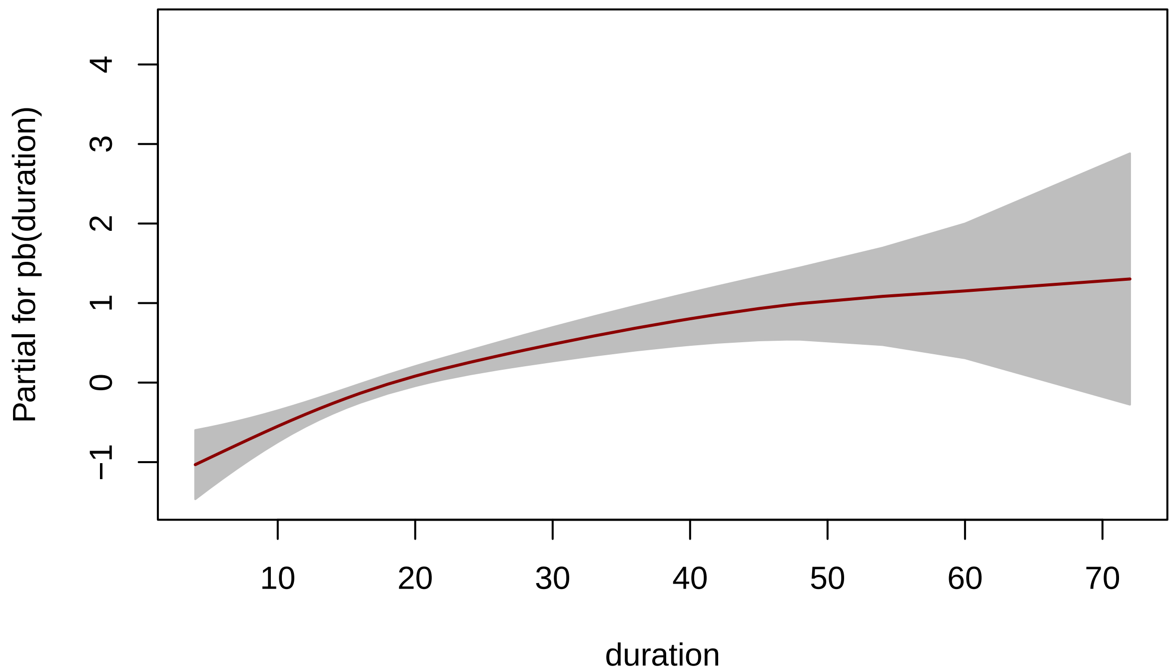
```
##
## Fitting method: RS()
##
## -------------------------------------------------------------------
## Mu link function:  logit
## Mu Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -1.765e+00  1.596e-01 -11.056  < 2e-16 ***
## pb(duration)  4.182e-02  5.827e-03   7.178 1.39e-12 ***
## pb(amount)    8.946e-07  3.988e-05   0.022    0.982
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## -------------------------------------------------------------------
## NOTE: Additive smoothing terms exist in the formulas:
##  i) Std. Error for smoothers are for the linear effect only.
## ii) Std. Error for the linear terms maybe are not accurate.
## -------------------------------------------------------------------
## No. of observations in the fit:  1000
## Degrees of Freedom for the fit:  6.737082
##       Residual Deg. of Freedom:  993.2629
##                      at cycle:  2
##
## Global Deviance:     1156.908
##             AIC:     1170.382
##             SBC:     1203.446
## *******************************************************************
```

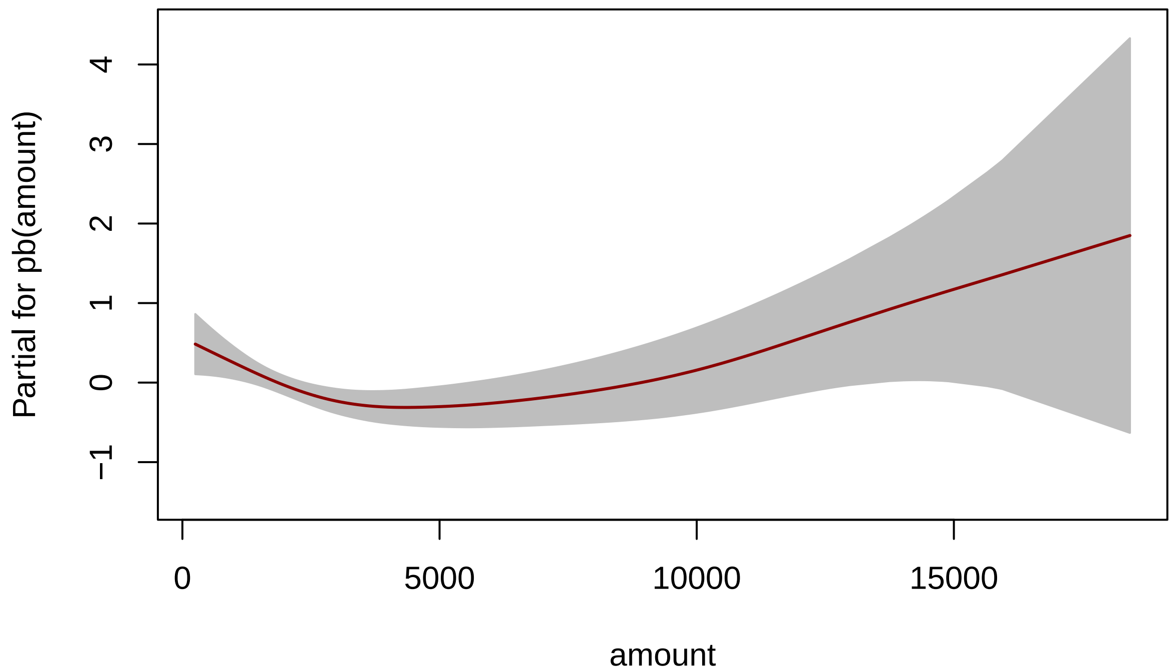


Partial for pb(duration)
−1
0
1
2
3
4
10
20
30
40
50
60
70
duration



Partial for pb(amount)
−1
0
1
2
3
4
0
5000
10000
15000
amount


Partial for pb(duration)
−1
0
1
2
3
4
10
20
30
40
50
60
70
duration



Analizando los datos observamos que tanto el intercepto como duration son variables altamente significativas, con p valores inferiores al 0.01. La variable amount, nos indica lo contrario, con un p valor muy elevado (0.982) nos indica que es muy poco significativo.

Dado que duration se introduce en el modelo suavizado a través de pb() (penalized B-splines), el efecto de duration en la probabilidad de que el crédito sea malo no es lineal, sino que sigue una forma más flexible.

```
##
## Call:
## glm(formula = response ~ duration + amount, family = binomial,
##     data = german_credit)
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.670e+00  1.466e-01 -11.390  < 2e-16 ***
## duration     3.412e-02  7.282e-03   4.685  2.8e-06 ***
## amount       2.300e-05  3.059e-05   0.752    0.452
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1221.7  on 999  degrees of freedom
## Residual deviance: 1176.6  on 997  degrees of freedom
## AIC: 1182.6
##
## Number of Fisher Scoring iterations: 4
```

```
## [1] 1182.552
```

```
## [1] 1197.276
```

Por último comparamos los resultados de ambos modelos (con y sin suavizar) contrastando el AIC y BIC no podemos concluir un mejor ajuste, pues en el modelo ajustado el AIC es mas bajo (lo que generalmente se considera como mejor modelo), pero en el caso de BIC presenta un mejor ajuste en el modelo sin suavizar.

Como conclusión, el modelo sugiere que duration es un predictor significativo para la probabilidad de que el crédito sea malo, mientras que amount no muestra un efecto significativo.

5. ¿Cuál es la probabilidad estimada media de que el crédito sea malo para mayores de 50 años?

```
## 'data.frame':    1000 obs. of  21 variables:
##  $ chk_acct        : chr  "A11" "A12" "A14" "A11" ...
##  $ duration        : int  6 48 12 42 24 36 24 36 12 30 ...
##  $ credit_his      : chr  "A34" "A32" "A34" "A32" ...
##  $ purpose         : chr  "A43" "A43" "A46" "A42" ...
##  $ amount          : int  1169 5951 2096 7882 4870 9055 2835 6948 3059 5234 ...
##  $ saving_acct     : chr  "A65" "A61" "A61" "A61" ...
##  $ present_emp     : chr  "A75" "A73" "A74" "A74" ...
##  $ installment_rate: int  4 2 2 2 3 2 3 2 2 4 ...
##  $ sex             : chr  "A93" "A92" "A93" "A93" ...
##  $ other_debtor    : chr  "A101" "A101" "A101" "A103" ...
##  $ present_resid   : int  4 2 3 4 4 4 4 2 4 2 ...
##  $ property        : chr  "A121" "A121" "A121" "A122" ...
##  $ age             : int  67 22 49 45 53 35 53 35 61 28 ...
##  $ other_install   : chr  "A143" "A143" "A143" "A143" ...
##  $ housing         : chr  "A152" "A152" "A152" "A153" ...
##  $ n_credits       : int  2 1 1 1 2 1 1 1 1 2 ...
##  $ job             : chr  "A173" "A173" "A172" "A173" ...
##  $ n_people        : int  1 1 2 2 2 2 1 1 1 1 ...
##  $ telephone       : chr  "A192" "A191" "A191" "A191" ...
```

```
##  $ foreign         : chr  "A201" "A201" "A201" "A201" ...
##  $ response        : Factor w/ 2 levels "0","1": 1 2 1 1 2 1 1 1 1 2 ...

## [1] 0

##
## Call:
## glm(formula = formula_new, family = gaussian, data = german_credit)
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)        0.542886   0.169039   3.212 0.001364 **
## duration           0.007026   0.001193   5.888 5.42e-09 ***
## chk_acctA12       -0.072293   0.036829  -1.963 0.049948 *
## chk_acctA13       -0.194947   0.057835  -3.371 0.000780 ***
## chk_acctA14       -0.266407   0.034675  -7.683 3.86e-14 ***
## credit_hisA31     -0.009205   0.090403  -0.102 0.918921
## credit_hisA32     -0.161295   0.070969  -2.273 0.023264 *
## credit_hisA33     -0.202374   0.077981  -2.595 0.009600 **
## credit_hisA34     -0.275153   0.071764  -3.834 0.000134 ***
## purposeA41        -0.221328   0.050087  -4.419 1.11e-05 ***
## purposeA410       -0.217504   0.125190  -1.737 0.082642 .
## purposeA42        -0.129365   0.041533  -3.115 0.001896 **
## purposeA43        -0.140669   0.037389  -3.762 0.000179 ***
## purposeA44        -0.090382   0.120199  -0.752 0.452277
## purposeA45        -0.042838   0.090971  -0.471 0.637827
## purposeA46         0.019685   0.064013   0.308 0.758520
## purposeA48        -0.261063   0.139476  -1.872 0.061549 .
## purposeA49        -0.125204   0.051846  -2.415 0.015926 *
## saving_acctA62    -0.055321   0.044518  -1.243 0.214299
## saving_acctA63    -0.085910   0.054985  -1.562 0.118518
## saving_acctA64    -0.148225   0.061834  -2.397 0.016715 *
## saving_acctA65    -0.118110   0.035743  -3.304 0.000987 ***
## present_empA72    -0.004015   0.069648  -0.058 0.954047
## present_empA73    -0.029742   0.066580  -0.447 0.655185
## present_empA74    -0.116852   0.069632  -1.678 0.093651 .
## present_empA75    -0.051492   0.066658  -0.772 0.440017
## installment_rate   0.031637   0.012083   2.618 0.008980 **
## sexA92            -0.064286   0.063503  -1.012 0.311632
## sexA93            -0.128827   0.061801  -2.085 0.037378 *
## sexA94            -0.097648   0.073555  -1.328 0.184643
## other_debtorA102   0.089713   0.066465   1.350 0.177411
## other_debtorA103  -0.166238   0.060778  -2.735 0.006351 **
## present_resid     -0.001846   0.013283  -0.139 0.889478
## propertyA122       0.050725   0.037574   1.350 0.177338
## propertyA123       0.037980   0.035617   1.066 0.286528
## propertyA124       0.136087   0.063411   2.146 0.032117 *
## other_installA142  0.001935   0.069328   0.028 0.977733
## other_installA143 -0.079054   0.039121  -2.021 0.043584 *
## housingA152       -0.071151   0.037203  -1.913 0.056109 .
## housingA153       -0.119929   0.072490  -1.654 0.098370 .
## n_credits          0.038850   0.028102   1.382 0.167150
## jobA172            0.091455   0.104305   0.877 0.380817
## jobA173            0.100242   0.101227   0.990 0.322294
```

```
## jobA174            0.105878   0.102051   1.037 0.299769
## n_people           0.035190   0.038721   0.909 0.363682
## telephoneA192     -0.033131   0.029677  -1.116 0.264532
## foreignA202       -0.140328   0.070717  -1.984 0.047503 *
## EDAD_hasta_50      0.002699   0.001904   1.417 0.156676
## EDAD_despues_50    0.001602   0.004288   0.374 0.708771
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.1596286)
##
##     Null deviance: 210.00  on 999  degrees of freedom
## Residual deviance: 151.81  on 951  degrees of freedom
## AIC: 1052.7
##
## Number of Fisher Scoring iterations: 2
```

```
## [1] 0.2482003
```

Dado que ambos valores p son superiores a 0.05, no hay evidencia estadísticamente significativa de que la edad (ya sea inferior o superior a 50 años) tenga un ajuste o un efecto significativo en el modelo. Esto significa que, en el contexto del modelo, la edad no parece influir de manera importante en la probabilidad de la respuesta para personas menores o mayores de 50 años.

la probabilidad estimada media de que el crédito sea malo para mayores de 50 años es de 24,8%

6. ¿Crees que hay discriminación de género en este último modelo creado?

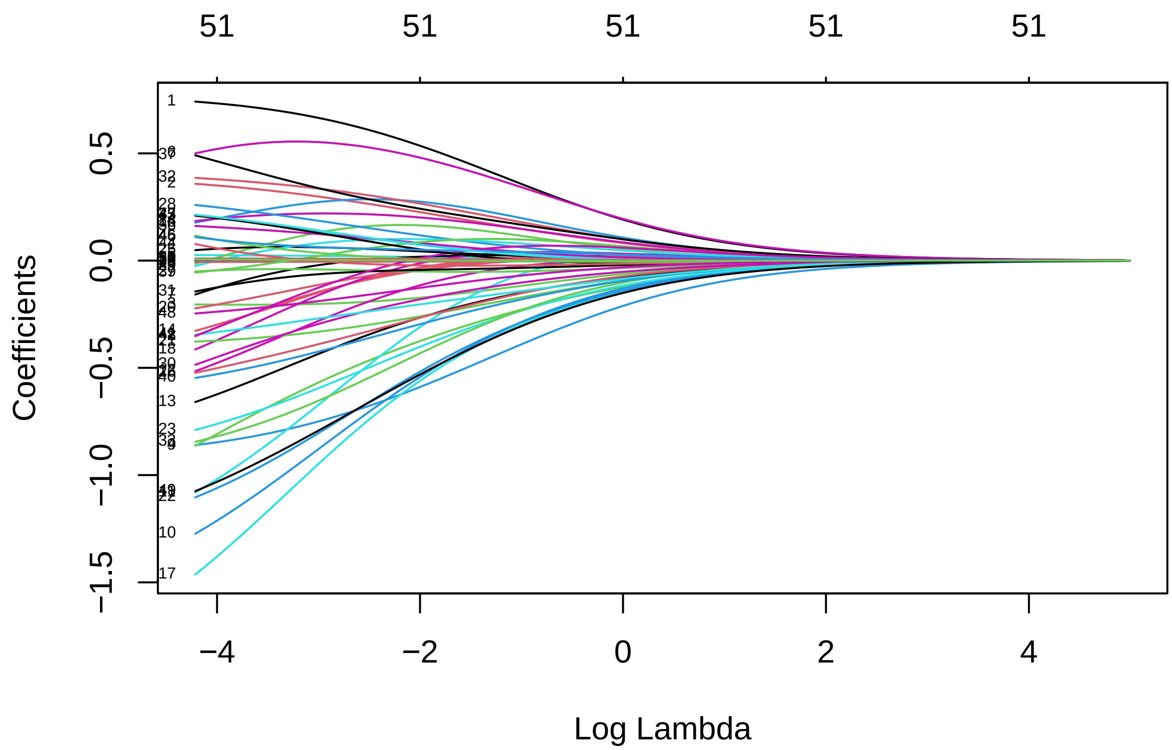
Si analizamos las variable que recogen la información de genero (SexA92,93 y 94) vemos que que únicamente podemos tomar como representativo analizando su p valor el caso de SexA93, con un p.value < 0.05. Esta categoría corresponde a 'male: single'. Analizando únicamente esta variable como significativa para el modelo, no podemos aterrizar una respuesta directa, pues hay otra categoría correspondiente al mismo sexo que no son significativas. Como consecuencia de esto, podemos indicar que no existe una discriminación directa por género.

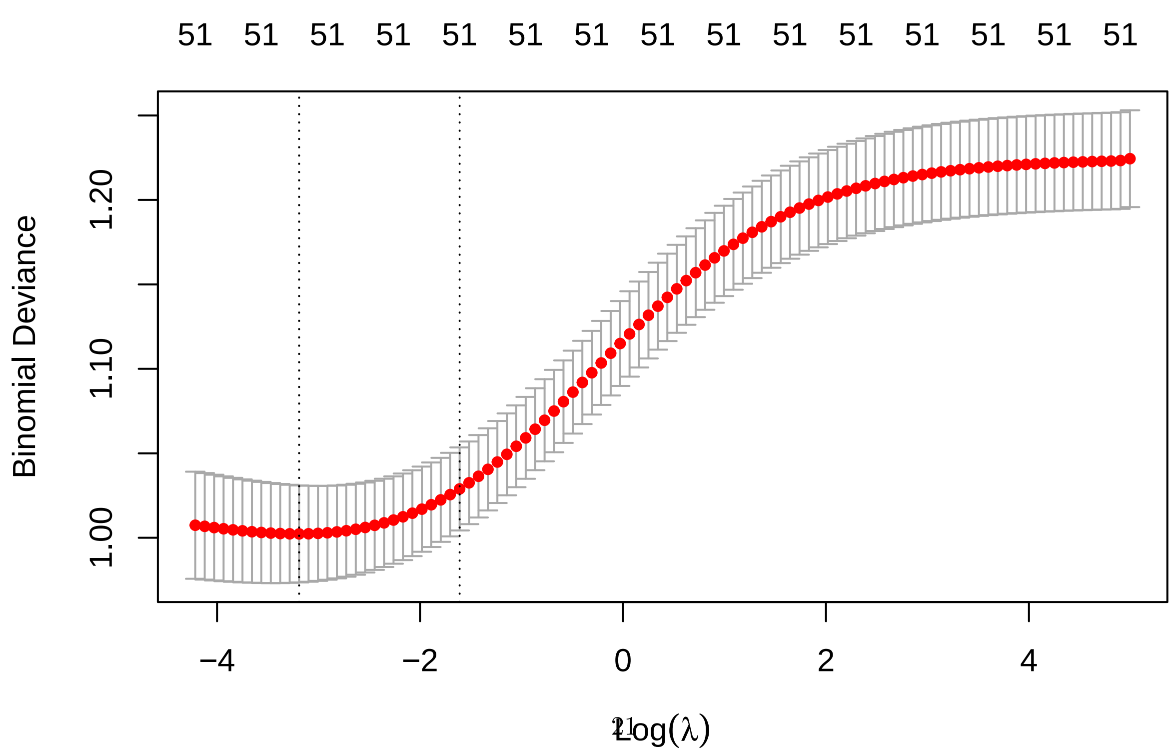
```
##
## Call:
```

```
## glm(formula = formula_logit, family = binomial, data = german_credit)
##
## Coefficients:
##                      Estimate Std. Error z value Pr(>|z|)
## (Intercept)          0.735577   1.084058   0.679 0.497429
## duration             0.043200   0.007700   5.610 2.02e-08 ***
## chk_acctA12         -0.334793   0.215930  -1.550 0.121030
## chk_acctA13         -1.010444   0.367713  -2.748 0.005998 **
## chk_acctA14         -1.679050   0.230131  -7.296 2.96e-13 ***
## credit_hisA31        0.033121   0.543002   0.061 0.951362
## credit_hisA32       -0.664639   0.425722  -1.561 0.118475
## credit_hisA33       -0.929478   0.467501  -1.988 0.046792 *
## credit_hisA34       -1.511875   0.435588  -3.471 0.000519 ***
## purposeA41          -1.510888   0.363500  -4.157 3.23e-05 ***
## purposeA410         -1.280260   0.719681  -1.779 0.075252 .
## purposeA42          -0.791414   0.260204  -3.042 0.002354 **
## purposeA43          -0.927151   0.246018  -3.769 0.000164 ***
## purposeA44          -0.598289   0.762031  -0.785 0.432381
## purposeA45          -0.231046   0.553188  -0.418 0.676194
## purposeA46           0.013150   0.392597   0.033 0.973281
## purposeA48          -2.038610   1.200515  -1.698 0.089486 .
## purposeA49          -0.761048   0.331502  -2.296 0.021690 *
## saving_acctA62      -0.389630   0.285663  -1.364 0.172584
## saving_acctA63      -0.420244   0.398863  -1.054 0.292063
## saving_acctA64      -1.311038   0.517790  -2.532 0.011342 *
## saving_acctA65      -0.897587   0.260475  -3.446 0.000569 ***
## present_empA72      -0.069255   0.424366  -0.163 0.870364
## present_empA73      -0.179541   0.405985  -0.442 0.658318
## present_empA74      -0.820499   0.442087  -1.856 0.063458 .
## present_empA75      -0.300354   0.408312  -0.736 0.461974
## installment_rate     0.222983   0.079244   2.814 0.004895 **
## sexA92              -0.271900   0.382754  -0.710 0.477470
## sexA93              -0.752239   0.375267  -2.005 0.045012 *
## sexA94              -0.411578   0.450180  -0.914 0.360584
## other_debtorA102     0.512717   0.405800   1.263 0.206420
## other_debtorA103    -1.000046   0.423580  -2.361 0.018229 *
## present_resid       -0.002927   0.085961  -0.034 0.972835
## propertyA122         0.307907   0.251806   1.223 0.221408
## propertyA123         0.232759   0.234427   0.993 0.320766
## propertyA124         0.827707   0.417375   1.983 0.047353 *
## age                 -0.014147   0.009149  -1.546 0.122061
## other_installA142 -0.123686   0.411643  -0.300 0.763819
## other_installA143 -0.622631   0.237472  -2.622 0.008744 **
## housingA152         -0.446099   0.233362  -1.912 0.055926 .
## housingA153         -0.705536   0.471264  -1.497 0.134364
## n_credits            0.287224   0.188962   1.520 0.128509
## jobA172              0.578398   0.686712   0.842 0.399637
## jobA173              0.581191   0.662886   0.877 0.380618
## jobA174              0.667596   0.666836   1.001 0.316759
## n_people             0.230638   0.248814   0.927 0.353954
## telephoneA192       -0.210575   0.196510  -1.072 0.283911
## foreignA202         -1.284315   0.607564  -2.114 0.034526 *
## ---
## 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:  904.28  on 952  degrees of freedom
## AIC: 1000.3
##
## Number of Fisher Scoring iterations: 5
```

7. Propón un modelo Ridge para modelizar el fenómeno crediticio. ¿Cuál es el lambda que minimiza el error? Compara este modelo con el logit que teníamos, anteriormente, con la curva ROC.





```
##                                   s1
## (Intercept)       -1.302771e+00
## chk_acctA11        6.827608e-01
## chk_acctA12        3.104376e-01
## chk_acctA13       -2.038631e-01
## chk_acctA14       -7.732971e-01
## duration           2.270137e-02
## credit_hisA31      5.551010e-01
## credit_hisA32     -3.512957e-02
## credit_hisA33     -1.721300e-01
## credit_hisA34     -6.111135e-01
## purposeA41        -9.459211e-01
## purposeA410       -7.442498e-01
## purposeA42        -3.185209e-01
## purposeA43        -4.716351e-01
## purposeA44        -1.776475e-01
## purposeA45         1.130481e-01
## purposeA46         2.623685e-01
## purposeA48        -1.031568e+00
## purposeA49        -1.995620e-01
## amount             8.241508e-05
## saving_acctA62    -1.310065e-01
## saving_acctA63    -3.418234e-01
## saving_acctA64    -8.585203e-01
## saving_acctA65    -6.301729e-01
## present_empA72     2.184779e-01
## present_empA73     6.320594e-02
## present_empA74    -4.121612e-01
## present_empA75    -3.987872e-02
## installment_rate   1.987849e-01
## sexA92             7.024738e-02
## sexA93            -3.219851e-01
## sexA94            -7.105071e-02
## other_debtorA102   3.459813e-01
## other_debtorA103  -6.909910e-01
## present_resid      3.295812e-03
## propertyA122       1.483207e-01
## propertyA123       1.307869e-01
## propertyA124       3.598759e-01
## age               -5.598306e-03
## other_installA142  7.376757e-03
## other_installA143 -4.470756e-01
## housingA152       -2.824001e-01
## housingA153       -1.597832e-01
## n_credits          1.436214e-01
## jobA172            5.299777e-03
## jobA173            4.460143e-02
## jobA174            6.737067e-02
## n_people           1.537346e-01
## telephoneA192     -1.997274e-01
## foreignA202       -8.415977e-01
## EDAD_hasta_50      8.982922e-03
## EDAD_despues_50    3.071181e-03
```

Conclusión modelo Ridge:

Los coeficientes de las variables menos relevantes tienden a acercarse a cero debido a la penalización, mientras que los coeficientes importantes se mantienen relativamente altos. Esto es evidente en variables como amount, que tiene un coeficiente cercano a cero, sugiriendo que su impacto es mínimo. Por otro lado, variables como chk_acctA11 y purposeA41 tienen coeficientes altos (positivos y negativos, respectivamente), lo cual indica una influencia importante en la probabilidad de que el crédito sea malo.

Coeficientes Positivos: Indican un aumento en la probabilidad de que el crédito sea malo. **Coeficientes Negativos:** Indican una disminución en la probabilidad de que el crédito sea malo. **Coeficientes Cercanos a Cero:** Variables con poco o ningún efecto sobre la respuesta en este modelo regularizado.

Este análisis nos ayuda a identificar las variables clave que influyen en la probabilidad de un crédito malo y, eliminando el ruido de las variables menos relevantes mediante la regularización Ridge.