# Regression analysis to determine significant predictors in estimating students' final grade

Ricky Doucette

## Attributes for both student-mat.csv (Math course) and student-por.csv (Portuguese language course) datasets:

1 school - student's school (binary: "GP" - Gabriel Pereira or "MS" - Mousinho da Silveira). 2 sex - student's sex (binary: "F" - female or "M" - male)
3 age - student's age (numeric: from 15 to 22)
4 address - student's home address type. (binary: "U" - urban or "R" - rural)
5 famsize - family size (binary: "LE3" - less or equal to 3 or "GT3" - greater than 3)
6 Pstatus - parent's cohabitation status (binary: "T" - living together or "A" - apart)
7 Medu - mother's education (numeric: 0 - none, 1 - primary education (4th grade), 2 – 5th to 9th grade, 3 – secondary education or 4 – higher education)
8 Fedu - father's education (numeric: 0 - none, 1 - primary education (4th grade), 2 – 5th to 9th grade, 3 – secondary education or 4 – higher education)
9 Mjob - mother's job (nominal: "teacher", "health" care related, civil "services" (e.g. administrative or police), "at_home" or "other")
10 Fjob - father's job (nominal: "teacher", "health" care related, civil "services" (e.g. administrative or police), "at_home" or "other")
11 reason - reason to choose this school (nominal: close to "home", school "reputation", "course" preference or "other")
12 guardian - student's guardian (nominal: "mother", "father" or "other")
13 traveltime - home to school travel time (numeric: 1 - <15 min., 2 - 15 to 30 min., 3 - 30 min. to 1 hour, or 4 - >1 hour)
14 studytime - weekly study time (numeric: 1 - <2 hours, 2 - 2 to 5 hours, 3 - 5 to 10 hours, or 4 - >10 hours)
15 failures - number of past class failures (numeric: n if 1<=n<3, else 4)
16 schoolsup - extra educational support (binary: yes or no). 17 famsup - family educational support (binary: yes or no)
18 paid - extra paid classes within the course subject (Math or Portuguese) (binary: yes or no)
19 activities - extra-curricular activities (binary: yes or no)
20 nursery - attended nursery school (binary: yes or no)
21 higher - wants to take higher education (binary: yes or no). 22 internet - Internet access at home (binary: yes or no). 23 romantic - with a romantic relationship (binary: yes or no). 24 famrel - quality of family relationships (numeric: from 1 - very bad to 5 - excellent)
25 freetime - free time after school (numeric: from 1 - very low to 5 - very high)
26 goout - going out with friends (numeric: from 1 - very low to 5 - very high)
27 Dalc - workday alcohol consumption (numeric: from 1 - very low to 5 - very high)
28 Walc - weekend alcohol consumption (numeric: from 1 - very low to 5 - very high)
29 health - current health status (numeric: from 1 - very bad to 5 - very good)
30 absences - number of school absences (numeric: from 0 to 93)

**these grades are related with the course subject, Math or Portuguese:**

31 G1 - first period grade (numeric: from 0 to 20)
31 G2 - second period grade (numeric: from 0 to 20)
32 G3 - final grade (numeric: from 0 to 20, output target)

Additional note: there are several (382) students that belong to both datasets . These students can be identified by searching for identical attributes that characterize each student, as shown in the annexed R file.

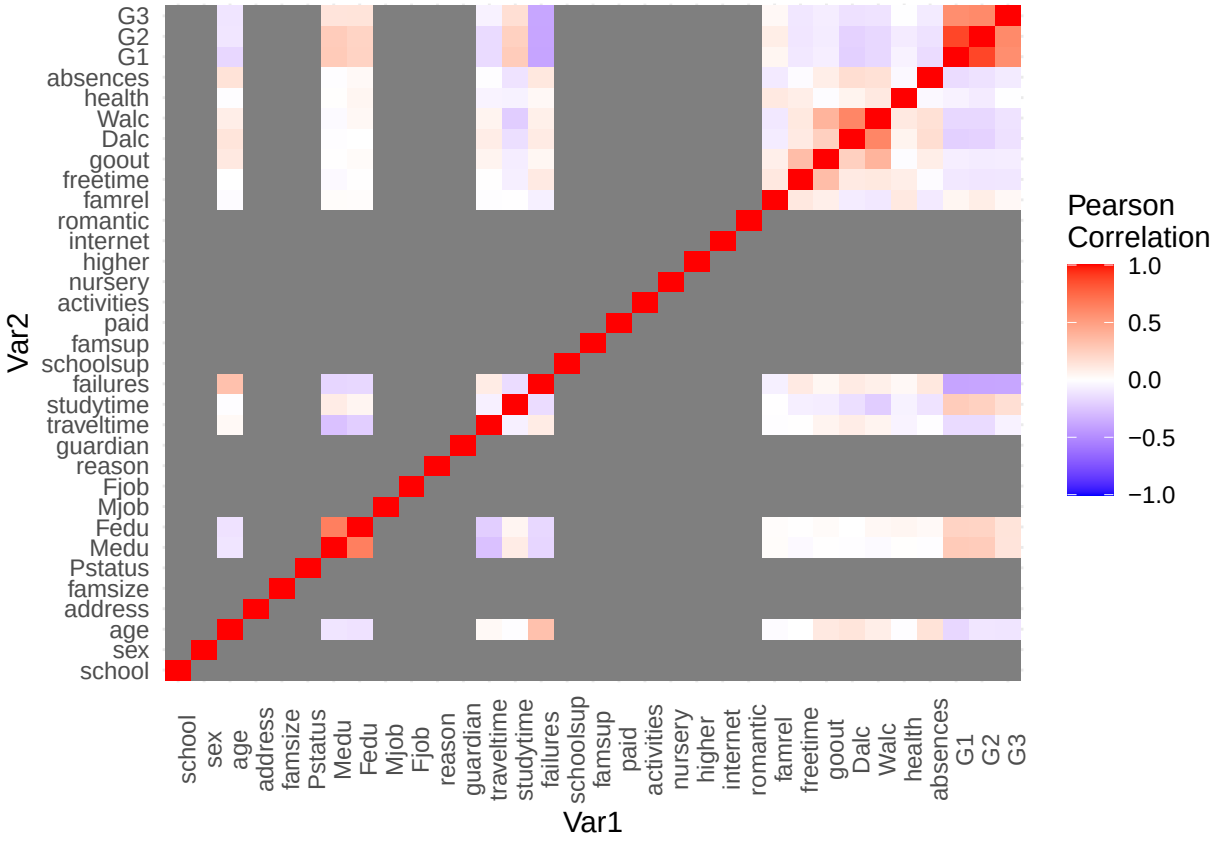
```
#load dataset
mydata <- read.csv("student-por.csv", sep=";", header = TRUE)
```

```
# G3 is being set to a categorical variable. 0 if <10 (fail), 1 if >=10 (pass)
mydata['G3'][mydata['G3']<10] <- 0
mydata['G3'][mydata['G3']>=10] <- 1

# Removing points that are problematic in the graphs
mydata <- mydata[-647,]
mydata <- mydata[-1,]
```

```
attach(mydata)

#Remove outlier rows
# mydata <- mydata[!mydata$G3 == 0, ]
# mydata <- mydata[!mydata$G2 == 0, ]
# mydata <- mydata[!mydata$G1 == 0, ]
# mydata <- mydata[!mydata$G3 == 1,]

# Compute the correlation matrix
# Testing multicolinearity
#change data to factors

cols_to_factor <- c("school", "sex", "address", "famsize", "Pstatus",
                    "Mjob", "Fjob", "reason", "guardian", "schoolsup",
                    "famsup","paid","activities",
                    "nursery","higher","internet", "romantic")

# Use lapply to apply as.factor() to each column
mydata[, cols_to_factor] <- lapply(mydata[, cols_to_factor], as.factor)
```

```
# Converting factor variables to numeric
mydata_numeric <- data.frame(lapply(mydata, function(x) {
  if(is.factor(x)) {
    as.numeric(as.character(x))
  } else {
    x
  }
}))
corr_matrix <- cor(mydata_numeric)
# Visualize the correlation matrix as a heatmap
library(ggplot2)
library(reshape2)
```

```
ggplot(melt(corr_matrix), aes(x=Var1, y=Var2, fill=value)) +
  geom_tile() +
  scale_fill_gradient2(low="blue", mid="white", high="red",
                       midpoint=0, limit=c(-1,1), space="Lab",
                       name="Pearson\nCorrelation") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 90))
```




```
# Using VIF to test for multicolinearity

library(car)
vif_model <- glm(G3 ~ ., data = mydata, family = "binomial")
vif(vif_model)
```

```
##                GVIF Df GVIF^(1/(2*Df))
## school     2.410067  1        1.552439
## sex        3.099991  1        1.760679
## age        2.850028  1        1.688203
## address    1.763693  1        1.328041
## famsize    1.676672  1        1.294864
## Pstatus    2.240670  1        1.496887
## Medu       2.523979  1        1.588704
## Fedu       2.259116  1        1.503035
## Mjob      10.434353  4        1.340628
## Fjob       9.356908  4        1.322487
```

```
## reason      4.020349  3        1.260987
## guardian    2.983972  2        1.314313
## traveltime  1.664615  1        1.290200
## studytime   1.491383  1        1.221222
## failures    1.773751  1        1.331822
## schoolsup   1.650389  1        1.284675
## famsup      1.474189  1        1.214162
## paid        1.316542  1        1.147407
## activities  1.632195  1        1.277574
## nursery     1.583747  1        1.258470
## higher      1.549498  1        1.244788
## internet    2.187811  1        1.479125
## romantic    2.159165  1        1.469410
## famrel      1.912996  1        1.383111
## freetime    1.860761  1        1.364097
## goout       2.694035  1        1.641351
## Dalc        2.974753  1        1.724747
## Walc        4.576055  1        2.139172
## health      1.978325  1        1.406529
## absences    1.686848  1        1.298787
## G1          2.881252  1        1.697425
## G2          2.578408  1        1.605742
```

```
model <- glm(G3 ~ ., data = mydata, family = "binomial")
model
```

```
##
## Call:  glm(formula = G3 ~ ., family = "binomial", data = mydata)
##
## Coefficients:
##      (Intercept)          schoolMS              sexM               age
##       -43.130129         -1.561525         -1.316500          0.958236
##         addressU        famsizeLE3          PstatusT              Medu
##        -0.135195          0.362254          1.009166         -0.107259
##             Fedu        Mjobhealth         Mjobother      Mjobservices
##        -0.126239          2.105626          1.454211          0.764383
##      Mjobteacher        Fjobhealth         Fjobother      Fjobservices
##         4.120146         -1.853057         -0.507461         -0.317733
##      Fjobteacher        reasonhome       reasonother  reasonreputation
##        -4.743629         -0.090618          0.190070          0.556349
##   guardianmother     guardianother        traveltime         studytime
##        -0.697524         -0.021122          0.337291          0.026488
##         failures      schoolsupyes         famsupyes           paidyes
##        -0.424025         -0.714431         -0.300532         -1.470155
##    activitiesyes        nurseryyes         higheryes       internetyes
##         0.670398         -0.842482          0.818902         -0.743096
##      romanticyes            famrel          freetime             goout
##        -1.566373         -0.168652         -0.008593         -0.376421
##             Dalc              Walc            health          absences
##         0.011560         -0.139496         -0.236034         -0.003088
##               G1                G2
##         1.351662          2.259987
##
## Degrees of Freedom: 646 Total (i.e. Null);  605 Residual
```

```
## Null Deviance:       553.7
## Residual Deviance: 122.5     AIC: 206.5
```

```r
summary(model)
```

```
##
## Call:
## glm(formula = G3 ~ ., family = "binomial", data = mydata)
##
## Coefficients:
##                    Estimate Std. Error z value Pr(>|z|)
## (Intercept)      -43.130129   9.644444  -4.472 7.75e-06 ***
## schoolMS          -1.561525   0.719013  -2.172  0.02987 *
## sexM              -1.316500   0.809941  -1.625  0.10407
## age                0.958236   0.306838   3.123  0.00179 **
## addressU          -0.135195   0.621576  -0.218  0.82782
## famsizeLE3         0.362254   0.650803   0.557  0.57778
## PstatusT           1.009166   1.107169   0.911  0.36204
## Medu              -0.107259   0.351612  -0.305  0.76033
## Fedu              -0.126239   0.328066  -0.385  0.70039
## Mjobhealth         2.105626   1.462680   1.440  0.14999
## Mjobother          1.454211   0.782281   1.859  0.06304 .
## Mjobservices       0.764383   1.005019   0.761  0.44692
## Mjobteacher        4.120146   1.682276   2.449  0.01432 *
## Fjobhealth        -1.853057   2.025209  -0.915  0.36019
## Fjobother         -0.507461   1.309577  -0.387  0.69839
## Fjobservices      -0.317733   1.355460  -0.234  0.81467
## Fjobteacher       -4.743629   2.037570  -2.328  0.01991 *
## reasonhome        -0.090618   0.771653  -0.117  0.90652
## reasonother        0.190070   0.799595   0.238  0.81211
## reasonreputation   0.556349   0.921155   0.604  0.54586
## guardianmother    -0.697524   0.784536  -0.889  0.37395
## guardianother     -0.021122   1.380883  -0.015  0.98780
## traveltime         0.337291   0.356898   0.945  0.34463
## studytime          0.026488   0.372974   0.071  0.94338
## failures          -0.424025   0.339520  -1.249  0.21170
## schoolsupyes      -0.714431   0.906437  -0.788  0.43059
## famsupyes         -0.300532   0.568749  -0.528  0.59722
## paidyes           -1.470155   1.186401  -1.239  0.21528
## activitiesyes      0.670398   0.609126   1.101  0.27108
## nurseryyes        -0.842482   0.697944  -1.207  0.22740
## higheryes          0.818902   0.682350   1.200  0.23009
## internetyes       -0.743096   0.711329  -1.045  0.29618
## romanticyes       -1.566373   0.681527  -2.298  0.02154 *
## famrel            -0.168652   0.291493  -0.579  0.56287
## freetime          -0.008593   0.273516  -0.031  0.97494
## goout             -0.376421   0.283338  -1.329  0.18401
## Dalc               0.011560   0.357985   0.032  0.97424
## Walc              -0.139496   0.360689  -0.387  0.69894
## health            -0.236034   0.241389  -0.978  0.32816
## absences          -0.003088   0.063867  -0.048  0.96143
## G1                 1.351662   0.339139   3.986 6.73e-05 ***
## G2                 2.259987   0.442163   5.111 3.20e-07 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 553.7  on 646  degrees of freedom
## Residual deviance: 122.5  on 605  degrees of freedom
## AIC: 206.5
##
## Number of Fisher Scoring iterations: 10
```

```
colnames(mydata)
```

```
##  [1] "school"     "sex"        "age"        "address"    "famsize"
##  [6] "Pstatus"    "Medu"       "Fedu"       "Mjob"       "Fjob"
## [11] "reason"     "guardian"   "traveltime" "studytime"  "failures"
## [16] "schoolsup"  "famsup"     "paid"       "activities" "nursery"
## [21] "higher"     "internet"   "romantic"   "famrel"     "freetime"
## [26] "goout"      "Dalc"       "Walc"       "health"     "absences"
## [31] "G1"         "G2"         "G3"
```

Remove column G2 due to multicolinearity

```
mydata <- mydata[, -32]
#fit full model again
model <- glm(G3 ~ ., data = mydata, family = "binomial")
summary(model)
```

```
##
## Call:
## glm(formula = G3 ~ ., family = "binomial", data = mydata)
##
## Coefficients:
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)      -22.939683   4.771657  -4.807 1.53e-06 ***
## schoolMS          -0.851035   0.518312  -1.642 0.100603
## sexM              -0.364493   0.509678  -0.715 0.474520
## age                0.720298   0.212425   3.391 0.000697 ***
## addressU           0.081205   0.474616   0.171 0.864149
## famsizeLE3         0.005795   0.479188   0.012 0.990352
## PstatusT           0.587881   0.688232   0.854 0.392999
## Medu              -0.147823   0.262574  -0.563 0.573452
## Fedu               0.194665   0.239481   0.813 0.416297
## Mjobhealth         0.708419   0.976207   0.726 0.468032
## Mjobother          0.164312   0.524409   0.313 0.754031
## Mjobservices      -0.116002   0.692660  -0.167 0.866998
## Mjobteacher        1.779290   1.184845   1.502 0.133173
## Fjobhealth        -1.570506   1.414314  -1.110 0.266811
## Fjobother         -0.561683   0.817060  -0.687 0.491803
## Fjobservices      -1.471553   0.869012  -1.693 0.090386 .
## Fjobteacher       -3.929084   1.566197  -2.509 0.012118 *
## reasonhome        -0.096357   0.549096  -0.175 0.860701
## reasonother       -0.206161   0.599180  -0.344 0.730792
```

```
## reasonreputation   0.264348   0.647650   0.408 0.683153
## guardianmother    -0.536507   0.545913  -0.983 0.325721
## guardianother     -0.354621   0.919603  -0.386 0.699775
## traveltime         0.219780   0.279665   0.786 0.431944
## studytime         -0.014198   0.283614  -0.050 0.960073
## failures          -0.789291   0.270206  -2.921 0.003488 **
## schoolsupyes       0.040885   0.684573   0.060 0.952376
## famsupyes          0.060525   0.424290   0.143 0.886567
## paidyes           -0.814722   0.729983  -1.116 0.264386
## activitiesyes      0.693174   0.436471   1.588 0.112256
## nurseryyes        -0.577126   0.506545  -1.139 0.254563
## higheryes          0.701244   0.498253   1.407 0.159307
## internetyes       -0.990099   0.519837  -1.905 0.056828 .
## romanticyes       -0.642399   0.430734  -1.491 0.135855
## famrel             0.159489   0.203568   0.783 0.433352
## freetime          -0.001409   0.205460  -0.007 0.994527
## goout             -0.240293   0.196847  -1.221 0.222195
## Dalc               0.188624   0.247362   0.763 0.445737
## Walc              -0.206794   0.222339  -0.930 0.352327
## health            -0.239744   0.157005  -1.527 0.126764
## absences          -0.043431   0.040927  -1.061 0.288615
## G1                 1.671495   0.215163   7.769 7.94e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 553.7  on 646  degrees of freedom
## Residual deviance: 198.6  on 606  degrees of freedom
## AIC: 280.6
##
## Number of Fisher Scoring iterations: 8
```

**model selection using stepwise forward selection with BIC, k =log(n)**

```
model.empty <- glm(G3~1,data=mydata, family = "binomial")
model.step.bic <- step(model.empty,
direction = "forward",scope = list(lower = model.empty, upper = model), k=2.8, trace =1)
```

```
## Start:  AIC=556.5
## G3 ~ 1
##
##              Df Deviance    AIC
## + G1          1   247.07 252.67
## + failures    1   484.70 490.30
## + school      1   500.62 506.22
## + higher      1   506.11 511.71
## + studytime   1   533.97 539.57
## + Fedu        1   540.20 545.80
## + reason      3   535.12 546.32
## + Medu        1   540.86 546.46
## + address     1   543.46 549.06
```

```
## + Dalc        1   544.73 550.33
## + Walc        1   544.74 550.34
## + age         1   546.28 551.88
## + freetime    1   546.92 552.52
## + romantic    1   549.26 554.86
## + absences    1   549.34 554.94
## + internet    1   549.44 555.04
## + sex         1   549.50 555.10
## + goout       1   550.05 555.65
## <none>            553.70 556.50
## + Mjob        4   542.65 556.65
## + traveltime  1   551.73 557.33
## + guardian    2   549.02 557.42
## + activities  1   551.95 557.55
## + paid        1   551.97 557.57
## + famsize     1   552.00 557.60
## + famsup      1   552.95 558.55
## + famrel      1   552.96 558.56
## + schoolsup   1   553.01 558.61
## + nursery     1   553.68 559.28
## + health      1   553.68 559.28
## + Pstatus     1   553.70 559.30
## + Fjob        4   550.80 564.80
##
## Step:  AIC=252.67
## G3 ~ G1
##
##              Df Deviance    AIC
## + age         1   242.34 250.74
## + failures    1   243.70 252.10
## + Walc        1   244.15 252.54
## <none>            247.07 252.67
## + sex         1   245.02 253.42
## + goout       1   245.27 253.67
## + higher      1   245.87 254.27
## + health      1   246.09 254.50
## + paid        1   246.14 254.54
## + romantic    1   246.21 254.61
## + traveltime  1   246.21 254.61
## + freetime    1   246.53 254.94
## + activities  1   246.56 254.96
## + school      1   246.63 255.03
## + Dalc        1   246.70 255.10
## + absences    1   246.80 255.20
## + schoolsup   1   246.81 255.21
## + nursery     1   246.82 255.22
## + internet    1   246.83 255.23
## + studytime   1   246.90 255.30
## + famsize     1   246.95 255.35
## + Fedu        1   247.02 255.42
## + Medu        1   247.03 255.43
## + famrel      1   247.03 255.43
## + address     1   247.04 255.44
## + famsup      1   247.06 255.46
```

```
## + Pstatus     1   247.06 255.46
## + Fjob        4   239.38 256.18
## + reason      3   243.15 257.15
## + guardian    2   246.36 257.56
## + Mjob        4   243.39 260.19
##
## Step:  AIC=250.74
## G3 ~ G1 + age
##
##              Df Deviance    AIC
## + failures    1   234.21 245.41
## + higher      1   239.10 250.30
## + Walc        1   239.15 250.35
## <none>            242.34 250.74
## + romantic    1   240.22 251.42
## + goout       1   240.36 251.56
## + sex         1   240.56 251.76
## + health      1   241.11 252.31
## + paid        1   241.36 252.56
## + Dalc        1   241.41 252.60
## + schoolsup   1   241.50 252.70
## + traveltime  1   241.51 252.72
## + absences    1   241.66 252.86
## + activities  1   241.73 252.93
## + internet    1   241.78 252.98
## + freetime    1   241.79 252.99
## + school      1   241.95 253.15
## + Medu        1   242.11 253.31
## + studytime   1   242.14 253.34
## + nursery     1   242.16 253.36
## + famsize     1   242.23 253.43
## + famrel      1   242.29 253.49
## + address     1   242.31 253.51
## + Fedu        1   242.33 253.53
## + Pstatus     1   242.34 253.54
## + famsup      1   242.34 253.54
## + Fjob        4   235.65 255.25
## + reason      3   239.14 255.94
## + guardian    2   242.05 256.05
## + Mjob        4   238.84 258.44
##
## Step:  AIC=245.41
## G3 ~ G1 + age + failures
##
##              Df Deviance    AIC
## + Walc        1   231.16 245.16
## <none>            234.21 245.41
## + goout       1   231.73 245.73
## + higher      1   231.75 245.75
## + romantic    1   232.23 246.23
## + activities  1   233.00 247.00
## + schoolsup   1   233.04 247.04
## + sex         1   233.09 247.09
## + health      1   233.11 247.11
```

```
## + traveltime  1   233.16 247.16
## + school      1   233.22 247.22
## + Dalc        1   233.23 247.23
## + absences    1   233.52 247.52
## + paid        1   233.56 247.56
## + internet    1   233.56 247.56
## + nursery     1   233.75 247.75
## + freetime    1   234.06 248.06
## + address     1   234.10 248.10
## + famsize     1   234.12 248.12
## + studytime   1   234.15 248.15
## + Medu        1   234.16 248.16
## + famsup      1   234.21 248.21
## + famrel      1   234.21 248.21
## + Fedu        1   234.21 248.21
## + Pstatus     1   234.21 248.21
## + Fjob        4   227.13 249.53
## + guardian    2   233.80 250.60
## + reason      3   231.43 251.03
## + Mjob        4   231.98 254.38
##
## Step:  AIC=245.16
## G3 ~ G1 + age + failures + Walc
##
##              Df Deviance    AIC
## <none>           231.16 245.16
## + romantic    1   228.54 245.34
## + higher      1   228.87 245.66
## + activities  1   229.48 246.28
## + traveltime  1   230.10 246.90
## + school      1   230.18 246.98
## + internet    1   230.36 247.16
## + schoolsup   1   230.43 247.23
## + health      1   230.44 247.25
## + goout       1   230.48 247.28
## + nursery     1   230.58 247.38
## + absences    1   230.67 247.47
## + paid        1   230.68 247.48
## + famsize     1   230.95 247.75
## + address     1   230.97 247.76
## + sex         1   231.09 247.89
## + Pstatus     1   231.11 247.91
## + Medu        1   231.12 247.91
## + Fedu        1   231.14 247.94
## + Dalc        1   231.15 247.95
## + famrel      1   231.16 247.96
## + freetime    1   231.16 247.96
## + famsup      1   231.16 247.96
## + studytime   1   231.16 247.96
## + Fjob        4   223.94 249.14
## + guardian    2   230.72 250.32
## + reason      3   228.63 251.03
## + Mjob        4   229.02 254.22
```

**model selection using stepwise forward selection with AIC, k = 2**

```
model.empty <- glm(G3~1,data=mydata, family = "binomial")
model.step.aic <- step(model.empty,
direction = "forward",scope = list(lower = model.empty, upper = model), k=2, trace =1)
```

```
## Start:  AIC=555.7
## G3 ~ 1
##
##              Df Deviance    AIC
## + G1          1   247.07 251.07
## + failures    1   484.70 488.70
## + school      1   500.62 504.62
## + higher      1   506.11 510.11
## + studytime   1   533.97 537.97
## + reason      3   535.12 543.12
## + Fedu        1   540.20 544.20
## + Medu        1   540.86 544.86
## + address     1   543.46 547.46
## + Dalc        1   544.73 548.73
## + Walc        1   544.74 548.74
## + age         1   546.28 550.28
## + freetime    1   546.92 550.92
## + Mjob        4   542.65 552.65
## + romantic    1   549.26 553.26
## + absences    1   549.34 553.34
## + internet    1   549.44 553.44
## + sex         1   549.50 553.50
## + goout       1   550.05 554.05
## + guardian    2   549.02 555.02
## <none>            553.70 555.70
## + traveltime  1   551.73 555.73
## + activities  1   551.95 555.95
## + paid        1   551.97 555.97
## + famsize     1   552.00 556.00
## + famsup      1   552.95 556.95
## + famrel      1   552.96 556.96
## + schoolsup   1   553.01 557.01
## + nursery     1   553.68 557.68
## + health      1   553.68 557.68
## + Pstatus     1   553.70 557.70
## + Fjob        4   550.80 560.80
##
## Step:  AIC=251.07
## G3 ~ G1
##
##              Df Deviance    AIC
## + age         1   242.34 248.34
## + failures    1   243.70 249.70
## + Walc        1   244.15 250.15
## + sex         1   245.02 251.02
## <none>            247.07 251.07
## + goout       1   245.27 251.27
```

```
## + Fjob        4   239.38 251.38
## + higher      1   245.87 251.87
## + health      1   246.09 252.09
## + paid        1   246.14 252.14
## + romantic    1   246.21 252.21
## + traveltime  1   246.21 252.21
## + freetime    1   246.53 252.53
## + activities  1   246.56 252.56
## + school      1   246.63 252.63
## + Dalc        1   246.70 252.70
## + absences    1   246.80 252.80
## + schoolsup   1   246.81 252.81
## + nursery     1   246.82 252.82
## + internet    1   246.83 252.83
## + studytime   1   246.90 252.90
## + famsize     1   246.95 252.95
## + Fedu        1   247.02 253.02
## + Medu        1   247.03 253.03
## + famrel      1   247.03 253.03
## + address     1   247.04 253.04
## + famsup      1   247.06 253.06
## + Pstatus     1   247.06 253.06
## + reason      3   243.15 253.15
## + guardian    2   246.36 254.36
## + Mjob        4   243.39 255.39
##
## Step:  AIC=248.34
## G3 ~ G1 + age
##
##              Df Deviance    AIC
## + failures    1   234.21 242.21
## + higher      1   239.10 247.10
## + Walc        1   239.15 247.15
## + romantic    1   240.22 248.22
## <none>            242.34 248.34
## + goout       1   240.36 248.36
## + sex         1   240.56 248.56
## + health      1   241.11 249.11
## + paid        1   241.36 249.36
## + Dalc        1   241.41 249.41
## + schoolsup   1   241.50 249.50
## + traveltime  1   241.51 249.51
## + Fjob        4   235.65 249.65
## + absences    1   241.66 249.66
## + activities  1   241.73 249.73
## + internet    1   241.78 249.78
## + freetime    1   241.79 249.79
## + school      1   241.95 249.95
## + Medu        1   242.11 250.11
## + studytime   1   242.14 250.14
## + nursery     1   242.16 250.16
## + famsize     1   242.23 250.23
## + famrel      1   242.29 250.29
## + address     1   242.31 250.31
```

```
## + Fedu        1   242.33 250.33
## + Pstatus     1   242.34 250.34
## + famsup      1   242.34 250.34
## + reason      3   239.14 251.14
## + guardian    2   242.05 252.05
## + Mjob        4   238.84 252.84
##
## Step:  AIC=242.21
## G3 ~ G1 + age + failures
##
##              Df Deviance    AIC
## + Walc        1   231.16 241.16
## + goout       1   231.73 241.73
## + higher      1   231.75 241.75
## <none>            234.21 242.21
## + romantic    1   232.23 242.23
## + activities  1   233.00 243.00
## + schoolsup   1   233.04 243.04
## + sex         1   233.09 243.09
## + health      1   233.11 243.11
## + Fjob        4   227.13 243.13
## + traveltime  1   233.16 243.16
## + school      1   233.22 243.22
## + Dalc        1   233.23 243.23
## + absences    1   233.52 243.52
## + paid        1   233.56 243.56
## + internet    1   233.56 243.56
## + nursery     1   233.75 243.75
## + freetime    1   234.06 244.06
## + address     1   234.10 244.10
## + famsize     1   234.12 244.12
## + studytime   1   234.15 244.15
## + Medu        1   234.16 244.16
## + famsup      1   234.21 244.21
## + famrel      1   234.21 244.21
## + Fedu        1   234.21 244.21
## + Pstatus     1   234.21 244.21
## + reason      3   231.43 245.43
## + guardian    2   233.80 245.80
## + Mjob        4   231.98 247.98
##
## Step:  AIC=241.16
## G3 ~ G1 + age + failures + Walc
##
##              Df Deviance    AIC
## + romantic    1   228.54 240.54
## + higher      1   228.87 240.87
## <none>            231.16 241.16
## + activities  1   229.48 241.48
## + Fjob        4   223.94 241.94
## + traveltime  1   230.10 242.10
## + school      1   230.18 242.18
## + internet    1   230.36 242.36
## + schoolsup   1   230.43 242.43
```

```
## + health      1   230.44 242.44
## + goout       1   230.48 242.48
## + nursery     1   230.58 242.58
## + absences    1   230.67 242.67
## + paid        1   230.68 242.68
## + famsize     1   230.95 242.95
## + address     1   230.97 242.97
## + sex         1   231.09 243.09
## + Pstatus     1   231.11 243.11
## + Medu        1   231.12 243.12
## + Fedu        1   231.14 243.14
## + Dalc        1   231.15 243.15
## + famrel      1   231.16 243.16
## + freetime    1   231.16 243.16
## + famsup      1   231.16 243.16
## + studytime   1   231.16 243.16
## + reason      3   228.63 244.63
## + guardian    2   230.72 244.72
## + Mjob        4   229.02 247.02
##
## Step:  AIC=240.54
## G3 ~ G1 + age + failures + Walc + romantic
##
##              Df Deviance    AIC
## + activities  1   226.47 240.47
## + higher      1   226.49 240.49
## <none>            228.54 240.54
## + traveltime  1   227.34 241.34
## + internet    1   227.64 241.64
## + school      1   227.65 241.65
## + goout       1   227.68 241.68
## + nursery     1   227.78 241.78
## + paid        1   227.79 241.79
## + Fjob        4   221.84 241.84
## + health      1   227.97 241.97
## + absences    1   228.10 242.10
## + schoolsup   1   228.15 242.15
## + sex         1   228.24 242.24
## + famsize     1   228.38 242.38
## + address     1   228.42 242.42
## + Medu        1   228.47 242.47
## + famrel      1   228.49 242.49
## + Dalc        1   228.51 242.51
## + Pstatus     1   228.53 242.53
## + famsup      1   228.53 242.53
## + freetime    1   228.53 242.53
## + studytime   1   228.54 242.54
## + Fedu        1   228.54 242.54
## + guardian    2   228.24 244.24
## + reason      3   226.50 244.50
## + Mjob        4   226.15 246.15
##
## Step:  AIC=240.47
## G3 ~ G1 + age + failures + Walc + romantic + activities
```

```
##
##               Df Deviance    AIC
## + higher       1   224.40 240.40
## <none>             226.47 240.47
## + internet     1   225.30 241.30
## + goout        1   225.31 241.31
## + school       1   225.53 241.53
## + traveltime   1   225.54 241.54
## + nursery      1   225.63 241.63
## + paid         1   225.66 241.66
## + health       1   225.88 241.88
## + sex          1   226.11 242.11
## + absences     1   226.12 242.12
## + address      1   226.19 242.19
## + schoolsup    1   226.20 242.20
## + freetime     1   226.32 242.32
## + famsize      1   226.32 242.32
## + Fjob         4   220.32 242.32
## + Medu         1   226.42 242.42
## + Fedu         1   226.44 242.44
## + Dalc         1   226.44 242.44
## + famsup       1   226.45 242.45
## + famrel       1   226.45 242.45
## + Pstatus      1   226.47 242.47
## + studytime    1   226.47 242.47
## + guardian     2   225.89 243.89
## + reason       3   224.71 244.71
## + Mjob         4   223.75 245.75
##
## Step:  AIC=240.4
## G3 ~ G1 + age + failures + Walc + romantic + activities + higher
##
##               Df Deviance    AIC
## <none>             224.40 240.40
## + internet     1   223.03 241.03
## + goout        1   223.32 241.32
## + traveltime   1   223.36 241.36
## + paid         1   223.37 241.37
## + school       1   223.39 241.39
## + health       1   223.49 241.49
## + nursery      1   223.79 241.79
## + absences     1   224.11 242.11
## + sex          1   224.17 242.17
## + Fjob         4   218.21 242.21
## + schoolsup    1   224.22 242.22
## + address      1   224.23 242.23
## + famsize      1   224.23 242.23
## + famsup       1   224.30 242.30
## + famrel       1   224.33 242.33
## + studytime    1   224.34 242.34
## + freetime     1   224.35 242.35
## + Dalc         1   224.40 242.40
## + Pstatus      1   224.40 242.40
## + Fedu         1   224.40 242.40
```

```
## + Medu         1   224.40 242.40
## + guardian     2   223.59 243.59
## + reason       3   222.87 244.87
## + Mjob         4   221.80 245.80
```

```
#library(bestglm)
#model.bestglm.bic <- bestglm(mydata,IC = 'BIC') #subset regression BIC
#model.bestglm.bic$BestModels
```

**model selection using stepwise backward selection**

```
model_full <- glm(G3 ~ ., data = mydata, family = "binomial")
model_backward <- step(model_full, direction = "backward", trace =1)
```

```
## Start:  AIC=280.6
## G3 ~ school + sex + age + address + famsize + Pstatus + Medu +
##     Fedu + Mjob + Fjob + reason + guardian + traveltime + studytime +
##     failures + schoolsup + famsup + paid + activities + nursery +
##     higher + internet + romantic + famrel + freetime + goout +
##     Dalc + Walc + health + absences + G1
##
##              Df Deviance    AIC
## - reason      3   199.01 275.01
## - Mjob        4   202.17 276.17
## - guardian    2   199.60 277.60
## - freetime    1   198.60 278.60
## - famsize     1   198.60 278.60
## - studytime   1   198.60 278.60
## - schoolsup   1   198.60 278.60
## - famsup      1   198.62 278.62
## - address     1   198.63 278.63
## - Medu        1   198.91 278.91
## - sex         1   199.11 279.11
## - Dalc        1   199.18 279.18
## - famrel      1   199.21 279.21
## - traveltime  1   199.23 279.23
## - Fedu        1   199.26 279.26
## - Pstatus     1   199.32 279.32
## - Walc        1   199.47 279.47
## - absences    1   199.73 279.73
## - paid        1   199.84 279.84
## - nursery     1   199.94 279.94
## - goout       1   200.11 280.11
## - higher      1   200.58 280.58
## <none>            198.60 280.60
## - romantic    1   200.85 280.85
## - health      1   200.99 280.99
## - activities  1   201.20 281.20
## - school      1   201.35 281.35
## - Fjob        4   208.02 282.02
## - internet    1   202.41 282.41
```

```
## - failures    1   207.62 287.62
## - age         1   211.60 291.60
## - G1          1   376.97 456.97
##
## Step:  AIC=275.01
## G3 ~ school + sex + age + address + famsize + Pstatus + Medu +
##     Fedu + Mjob + Fjob + guardian + traveltime + studytime +
##     failures + schoolsup + famsup + paid + activities + nursery +
##     higher + internet + romantic + famrel + freetime + goout +
##     Dalc + Walc + health + absences + G1
##
##              Df Deviance    AIC
## - Mjob        4   202.76 270.76
## - guardian    2   200.24 272.24
## - address     1   199.01 273.01
## - freetime    1   199.01 273.01
## - studytime   1   199.01 273.01
## - schoolsup   1   199.02 273.02
## - famsize     1   199.02 273.02
## - famsup      1   199.04 273.04
## - Medu        1   199.32 273.32
## - Dalc        1   199.48 273.48
## - traveltime  1   199.53 273.53
## - Fedu        1   199.64 273.64
## - sex         1   199.64 273.64
## - famrel      1   199.72 273.72
## - Pstatus     1   199.74 273.74
## - Walc        1   199.74 273.74
## - absences    1   200.13 274.13
## - nursery     1   200.33 274.33
## - paid        1   200.49 274.49
## - goout       1   200.68 274.68
## <none>            199.01 275.01
## - higher      1   201.23 275.23
## - romantic    1   201.51 275.51
## - health      1   201.64 275.64
## - activities  1   201.94 275.94
## - Fjob        4   208.59 276.59
## - school      1   202.62 276.62
## - internet    1   202.76 276.76
## - failures    1   208.06 282.06
## - age         1   213.86 287.86
## - G1          1   378.01 452.01
##
## Step:  AIC=270.76
## G3 ~ school + sex + age + address + famsize + Pstatus + Medu +
##     Fedu + Fjob + guardian + traveltime + studytime + failures +
##     schoolsup + famsup + paid + activities + nursery + higher +
##     internet + romantic + famrel + freetime + goout + Dalc +
##     Walc + health + absences + G1
##
##              Df Deviance    AIC
## - guardian    2   203.56 267.56
## - schoolsup   1   202.76 268.76
```

```
## - famsize     1   202.76 268.76
## - studytime   1   202.77 268.77
## - Medu        1   202.78 268.78
## - freetime    1   202.78 268.78
## - famsup      1   202.80 268.80
## - address     1   202.82 268.82
## - Dalc        1   202.98 268.98
## - Walc        1   203.10 269.10
## - sex         1   203.27 269.27
## - Pstatus     1   203.38 269.38
## - Fedu        1   203.39 269.39
## - famrel      1   203.43 269.43
## - nursery     1   203.53 269.53
## - traveltime  1   203.55 269.55
## - absences    1   204.26 270.26
## - health      1   204.63 270.63
## - paid        1   204.68 270.68
## <none>            202.76 270.76
## - goout       1   204.87 270.87
## - romantic    1   204.87 270.87
## - activities  1   205.24 271.24
## - higher      1   205.46 271.46
## - internet    1   205.68 271.68
## - Fjob        4   212.09 272.09
## - school      1   206.11 272.11
## - failures    1   212.31 278.31
## - age         1   216.97 282.97
## - G1          1   379.16 445.16
##
## Step:  AIC=267.56
## G3 ~ school + sex + age + address + famsize + Pstatus + Medu +
##     Fedu + Fjob + traveltime + studytime + failures + schoolsup +
##     famsup + paid + activities + nursery + higher + internet +
##     romantic + famrel + freetime + goout + Dalc + Walc + health +
##     absences + G1
##
##              Df Deviance    AIC
## - famsize     1   203.57 265.57
## - schoolsup   1   203.57 265.57
## - Medu        1   203.57 265.57
## - studytime   1   203.57 265.57
## - freetime    1   203.62 265.62
## - famsup      1   203.62 265.62
## - address     1   203.64 265.64
## - Dalc        1   203.86 265.86
## - Walc        1   203.91 265.91
## - sex         1   204.08 266.08
## - Pstatus     1   204.11 266.11
## - nursery     1   204.17 266.17
## - famrel      1   204.24 266.24
## - traveltime  1   204.50 266.50
## - Fedu        1   204.53 266.53
## - absences    1   205.26 267.26
## - paid        1   205.48 267.48
```

```
## <none>            203.56 267.56
## - health      1   205.60 267.60
## - activities  1   205.62 267.62
## - romantic    1   205.75 267.75
## - goout       1   205.75 267.75
## - higher      1   205.94 267.94
## - internet    1   205.99 267.99
## - school      1   207.05 269.05
## - Fjob        4   213.12 269.12
## - failures    1   212.86 274.86
## - age         1   218.66 280.66
## - G1          1   384.15 446.15
##
## Step:  AIC=265.57
## G3 ~ school + sex + age + address + Pstatus + Medu + Fedu + Fjob +
##     traveltime + studytime + failures + schoolsup + famsup +
##     paid + activities + nursery + higher + internet + romantic +
##     famrel + freetime + goout + Dalc + Walc + health + absences +
##     G1
##
##              Df Deviance    AIC
## - schoolsup   1   203.57 263.57
## - Medu        1   203.57 263.57
## - studytime   1   203.57 263.57
## - freetime    1   203.62 263.62
## - famsup      1   203.62 263.62
## - address     1   203.64 263.64
## - Dalc        1   203.86 263.86
## - Walc        1   203.91 263.91
## - sex         1   204.08 264.08
## - Pstatus     1   204.14 264.14
## - nursery     1   204.17 264.17
## - famrel      1   204.24 264.24
## - traveltime  1   204.51 264.51
## - Fedu        1   204.53 264.53
## - absences    1   205.30 265.30
## - paid        1   205.49 265.49
## <none>            203.57 265.57
## - health      1   205.61 265.61
## - activities  1   205.64 265.64
## - goout       1   205.76 265.76
## - romantic    1   205.77 265.77
## - higher      1   205.94 265.94
## - internet    1   206.01 266.01
## - school      1   207.13 267.13
## - Fjob        4   213.18 267.18
## - failures    1   212.93 272.93
## - age         1   218.67 278.67
## - G1          1   385.42 445.42
##
## Step:  AIC=263.57
## G3 ~ school + sex + age + address + Pstatus + Medu + Fedu + Fjob +
##     traveltime + studytime + failures + famsup + paid + activities +
##     nursery + higher + internet + romantic + famrel + freetime +
```

```
##     goout + Dalc + Walc + health + absences + G1
##
##               Df Deviance    AIC
## - Medu         1   203.57 261.57
## - studytime    1   203.57 261.57
## - freetime     1   203.62 261.62
## - famsup       1   203.62 261.62
## - address      1   203.64 261.64
## - Dalc         1   203.86 261.86
## - Walc         1   203.92 261.92
## - sex          1   204.08 262.08
## - Pstatus      1   204.14 262.14
## - nursery      1   204.17 262.17
## - famrel       1   204.24 262.24
## - traveltime   1   204.51 262.51
## - Fedu         1   204.53 262.53
## - absences     1   205.31 263.31
## - paid         1   205.49 263.49
## <none>             203.57 263.57
## - health       1   205.61 263.61
## - activities   1   205.64 263.64
## - goout        1   205.76 263.76
## - romantic     1   205.82 263.82
## - higher       1   205.94 263.94
## - internet     1   206.02 264.02
## - Fjob         4   213.19 265.19
## - school       1   207.24 265.24
## - failures     1   212.94 270.94
## - age          1   218.80 276.80
## - G1           1   386.52 444.52
##
## Step:  AIC=261.57
## G3 ~ school + sex + age + address + Pstatus + Fedu + Fjob + traveltime +
##     studytime + failures + famsup + paid + activities + nursery +
##     higher + internet + romantic + famrel + freetime + goout +
##     Dalc + Walc + health + absences + G1
##
##               Df Deviance    AIC
## - studytime    1   203.57 259.57
## - freetime     1   203.62 259.62
## - famsup       1   203.62 259.62
## - address      1   203.64 259.64
## - Dalc         1   203.88 259.88
## - Walc         1   203.93 259.93
## - sex          1   204.08 260.08
## - Pstatus      1   204.14 260.14
## - nursery      1   204.18 260.18
## - famrel       1   204.24 260.24
## - traveltime   1   204.51 260.51
## - Fedu         1   204.81 260.81
## - absences     1   205.32 261.32
## - paid         1   205.50 261.50
## <none>             203.57 261.57
## - health       1   205.64 261.64
```

```
## - activities  1   205.70 261.70
## - goout       1   205.77 261.77
## - romantic    1   205.82 261.82
## - higher      1   205.97 261.97
## - internet    1   206.04 262.04
## - Fjob        4   213.25 263.25
## - school      1   207.29 263.29
## - failures    1   213.01 269.01
## - age         1   218.83 274.83
## - G1          1   386.52 442.52
##
## Step:  AIC=259.57
## G3 ~ school + sex + age + address + Pstatus + Fedu + Fjob + traveltime +
##     failures + famsup + paid + activities + nursery + higher +
##     internet + romantic + famrel + freetime + goout + Dalc +
##     Walc + health + absences + G1
##
##              Df Deviance    AIC
## - freetime    1   203.62 257.62
## - famsup      1   203.63 257.63
## - address     1   203.65 257.65
## - Dalc        1   203.88 257.88
## - Walc        1   203.94 257.94
## - sex         1   204.10 258.10
## - Pstatus     1   204.14 258.14
## - nursery     1   204.18 258.18
## - famrel      1   204.25 258.25
## - traveltime  1   204.52 258.52
## - Fedu        1   204.82 258.82
## - absences    1   205.35 259.35
## - paid        1   205.50 259.50
## <none>            203.57 259.57
## - health      1   205.64 259.64
## - activities  1   205.70 259.70
## - goout       1   205.77 259.77
## - romantic    1   205.82 259.82
## - higher      1   205.99 259.99
## - internet    1   206.10 260.10
## - school      1   207.34 261.34
## - Fjob        4   213.53 261.53
## - failures    1   213.05 267.05
## - age         1   218.94 272.94
## - G1          1   387.64 441.64
##
## Step:  AIC=257.62
## G3 ~ school + sex + age + address + Pstatus + Fedu + Fjob + traveltime +
##     failures + famsup + paid + activities + nursery + higher +
##     internet + romantic + famrel + goout + Dalc + Walc + health +
##     absences + G1
##
##              Df Deviance    AIC
## - famsup      1   203.68 255.68
## - address     1   203.70 255.70
## - Dalc        1   203.97 255.97
```

```
## - Walc        1   204.04 256.04
## - sex         1   204.12 256.12
## - Pstatus     1   204.19 256.19
## - nursery     1   204.32 256.32
## - famrel      1   204.40 256.40
## - traveltime  1   204.55 256.55
## - Fedu        1   204.87 256.87
## - absences    1   205.40 257.40
## - paid        1   205.61 257.61
## <none>            203.62 257.62
## - health      1   205.66 257.66
## - romantic    1   205.93 257.93
## - activities  1   205.94 257.94
## - goout       1   205.96 257.96
## - higher      1   206.00 258.00
## - internet    1   206.13 258.13
## - school      1   207.34 259.34
## - Fjob        4   213.63 259.63
## - failures    1   213.10 265.10
## - age         1   218.95 270.95
## - G1          1   388.60 440.60
##
## Step:  AIC=255.68
## G3 ~ school + sex + age + address + Pstatus + Fedu + Fjob + traveltime +
##     failures + paid + activities + nursery + higher + internet +
##     romantic + famrel + goout + Dalc + Walc + health + absences +
##     G1
##
##              Df Deviance    AIC
## - address     1   203.76 253.76
## - Dalc        1   204.04 254.04
## - Walc        1   204.11 254.11
## - Pstatus     1   204.25 254.25
## - sex         1   204.27 254.27
## - nursery     1   204.39 254.39
## - famrel      1   204.46 254.46
## - traveltime  1   204.67 254.67
## - Fedu        1   205.02 255.02
## - absences    1   205.41 255.41
## - paid        1   205.62 255.62
## <none>            203.68 255.68
## - health      1   205.68 255.68
## - activities  1   206.03 256.03
## - goout       1   206.03 256.03
## - romantic    1   206.06 256.06
## - higher      1   206.11 256.11
## - internet    1   206.13 256.13
## - school      1   207.34 257.34
## - Fjob        4   213.63 257.63
## - failures    1   213.10 263.10
## - age         1   218.99 268.99
## - G1          1   388.87 438.87
##
## Step:  AIC=253.76
```

```
## G3 ~ school + sex + age + Pstatus + Fedu + Fjob + traveltime +
##     failures + paid + activities + nursery + higher + internet +
##     romantic + famrel + goout + Dalc + Walc + health + absences +
##     G1
##
##              Df Deviance    AIC
## - Dalc        1   204.10 252.10
## - Walc        1   204.15 252.15
## - Pstatus     1   204.31 252.31
## - sex         1   204.36 252.36
## - famrel      1   204.49 252.49
## - nursery     1   204.50 252.50
## - traveltime  1   204.68 252.68
## - Fedu        1   205.11 253.11
## - absences    1   205.49 253.49
## - health      1   205.74 253.74
## <none>            203.76 253.76
## - paid        1   205.84 253.84
## - activities  1   206.05 254.05
## - internet    1   206.15 254.15
## - romantic    1   206.21 254.21
## - goout       1   206.22 254.22
## - higher      1   206.23 254.23
## - Fjob        4   213.71 255.71
## - school      1   207.97 255.97
## - failures    1   213.11 261.11
## - age         1   219.10 267.10
## - G1          1   389.60 437.60
##
## Step:  AIC=252.1
## G3 ~ school + sex + age + Pstatus + Fedu + Fjob + traveltime +
##     failures + paid + activities + nursery + higher + internet +
##     romantic + famrel + goout + Walc + health + absences + G1
##
##              Df Deviance    AIC
## - Walc        1   204.26 250.26
## - sex         1   204.59 250.59
## - Pstatus     1   204.67 250.67
## - famrel      1   204.79 250.79
## - nursery     1   204.79 250.79
## - traveltime  1   205.08 251.08
## - Fedu        1   205.55 251.55
## - absences    1   205.65 251.65
## - paid        1   206.06 252.06
## <none>            204.10 252.10
## - health      1   206.17 252.17
## - activities  1   206.39 252.39
## - internet    1   206.44 252.44
## - goout       1   206.48 252.48
## - romantic    1   206.54 252.54
## - higher      1   206.81 252.81
## - Fjob        4   213.81 253.81
## - school      1   208.42 254.42
## - failures    1   213.48 259.48
```

```
## - age         1   220.74 266.74
## - G1          1   389.61 435.61
##
## Step:  AIC=250.26
## G3 ~ school + sex + age + Pstatus + Fedu + Fjob + traveltime +
##     failures + paid + activities + nursery + higher + internet +
##     romantic + famrel + goout + health + absences + G1
##
##              Df Deviance    AIC
## - Pstatus     1   204.83 248.83
## - nursery     1   204.92 248.92
## - famrel      1   205.02 249.02
## - sex         1   205.18 249.18
## - traveltime  1   205.38 249.38
## - Fedu        1   205.65 249.65
## - absences    1   205.86 249.86
## <none>            204.26 250.26
## - paid        1   206.32 250.32
## - health      1   206.42 250.42
## - internet    1   206.53 250.53
## - activities  1   206.54 250.54
## - romantic    1   206.73 250.73
## - higher      1   207.09 251.09
## - goout       1   207.97 251.97
## - Fjob        4   214.08 252.08
## - school      1   208.80 252.80
## - failures    1   213.62 257.62
## - age         1   220.83 264.83
## - G1          1   390.63 434.63
##
## Step:  AIC=248.83
## G3 ~ school + sex + age + Fedu + Fjob + traveltime + failures +
##     paid + activities + nursery + higher + internet + romantic +
##     famrel + goout + health + absences + G1
##
##              Df Deviance    AIC
## - nursery     1   205.43 247.43
## - sex         1   205.53 247.53
## - famrel      1   205.57 247.57
## - traveltime  1   205.96 247.96
## - Fedu        1   206.09 248.09
## - absences    1   206.56 248.56
## - internet    1   206.71 248.71
## - health      1   206.73 248.73
## <none>            204.83 248.83
## - paid        1   206.98 248.98
## - activities  1   207.26 249.26
## - romantic    1   207.44 249.44
## - higher      1   207.59 249.59
## - Fjob        4   214.36 250.36
## - goout       1   208.45 250.45
## - school      1   209.13 251.13
## - failures    1   214.24 256.24
## - age         1   221.20 263.20
```

```
## - G1          1   390.63 432.63
##
## Step:  AIC=247.43
## G3 ~ school + sex + age + Fedu + Fjob + traveltime + failures +
##     paid + activities + higher + internet + romantic + famrel +
##     goout + health + absences + G1
##
##              Df Deviance    AIC
## - sex         1   206.03 246.03
## - famrel      1   206.06 246.06
## - traveltime  1   206.70 246.70
## - Fedu        1   206.77 246.77
## - absences    1   206.96 246.96
## - internet    1   207.25 247.25
## - paid        1   207.41 247.41
## <none>            205.43 247.43
## - health      1   207.56 247.56
## - activities  1   207.83 247.83
## - romantic    1   207.84 247.84
## - higher      1   208.52 248.52
## - Fjob        4   214.87 248.87
## - goout       1   209.39 249.39
## - school      1   209.74 249.74
## - failures    1   214.72 254.72
## - age         1   221.83 261.83
## - G1          1   391.88 431.88
##
## Step:  AIC=246.03
## G3 ~ school + age + Fedu + Fjob + traveltime + failures + paid +
##     activities + higher + internet + romantic + famrel + goout +
##     health + absences + G1
##
##              Df Deviance    AIC
## - famrel      1   206.63 244.63
## - traveltime  1   207.13 245.13
## - Fedu        1   207.41 245.41
## - absences    1   207.64 245.64
## - internet    1   207.81 245.81
## <none>            206.03 246.03
## - romantic    1   208.07 246.07
## - activities  1   208.19 246.19
## - paid        1   208.21 246.21
## - health      1   208.76 246.76
## - higher      1   209.46 247.46
## - school      1   209.99 247.99
## - Fjob        4   216.45 248.45
## - goout       1   210.60 248.60
## - failures    1   215.72 253.72
## - age         1   222.78 260.78
## - G1          1   394.96 432.96
##
## Step:  AIC=244.63
## G3 ~ school + age + Fedu + Fjob + traveltime + failures + paid +
##     activities + higher + internet + romantic + goout + health +
```

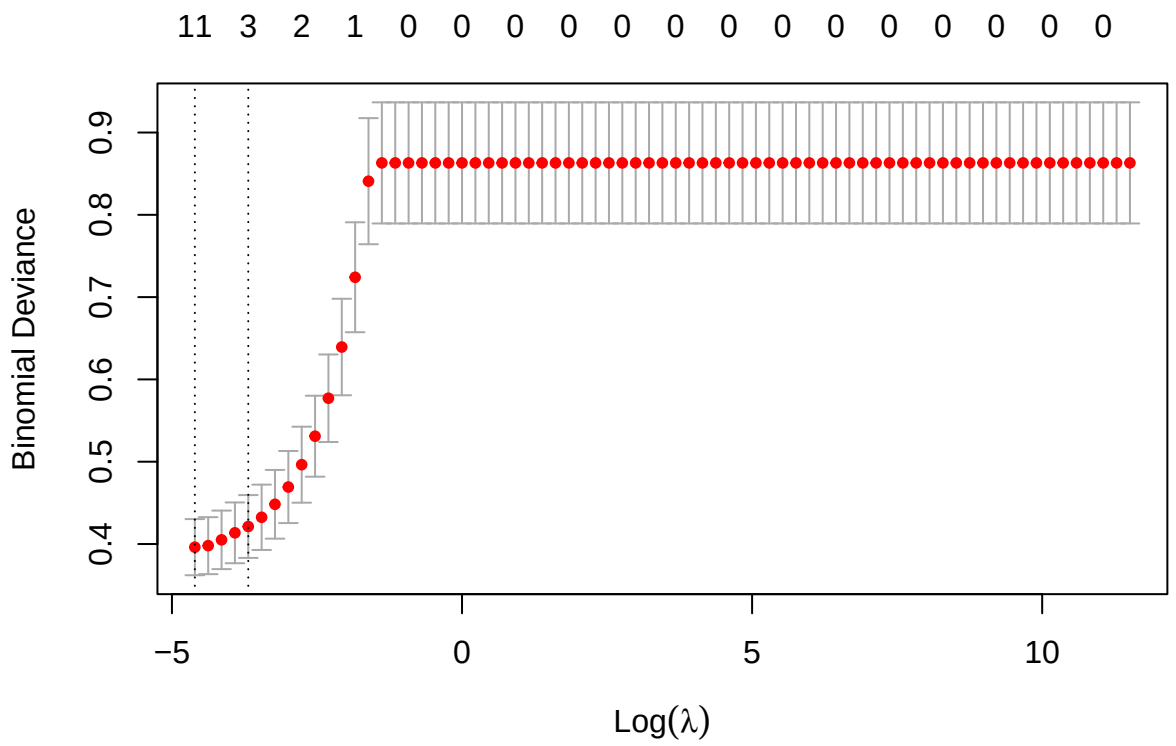
```
##     absences + G1
##
##              Df Deviance    AIC
## - Fedu        1   207.84 243.84
## - traveltime  1   207.90 243.90
## - internet    1   208.12 244.12
## - romantic    1   208.42 244.42
## - absences    1   208.53 244.53
## - paid        1   208.62 244.62
## <none>            206.63 244.63
## - activities  1   208.79 244.79
## - health      1   208.97 244.97
## - higher      1   209.71 245.71
## - Fjob        4   216.76 246.76
## - goout       1   210.89 246.89
## - school      1   211.11 247.11
## - failures    1   217.00 253.00
## - age         1   223.07 259.07
## - G1          1   394.99 430.99
##
## Step:  AIC=243.84
## G3 ~ school + age + Fjob + traveltime + failures + paid + activities +
##     higher + internet + romantic + goout + health + absences +
##     G1
##
##              Df Deviance    AIC
## - internet    1   208.96 242.96
## - traveltime  1   209.15 243.15
## - paid        1   209.56 243.56
## - absences    1   209.57 243.57
## - activities  1   209.72 243.72
## - romantic    1   209.81 243.81
## <none>            207.84 243.84
## - health      1   210.06 244.06
## - Fjob        4   216.79 244.79
## - higher      1   211.49 245.49
## - goout       1   211.71 245.71
## - school      1   212.66 246.66
## - failures    1   218.33 252.33
## - age         1   223.67 257.67
## - G1          1   397.21 431.21
##
## Step:  AIC=242.96
## G3 ~ school + age + Fjob + traveltime + failures + paid + activities +
##     higher + romantic + goout + health + absences + G1
##
##              Df Deviance    AIC
## - traveltime  1   210.48 242.48
## - activities  1   210.61 242.61
## - paid        1   210.66 242.66
## - absences    1   210.71 242.71
## - romantic    1   210.79 242.79
## <none>            208.96 242.96
## - health      1   211.01 243.01
```

```
## - higher      1   212.36 244.36
## - Fjob        4   218.47 244.47
## - goout       1   213.01 245.01
## - school      1   213.18 245.18
## - failures    1   218.83 250.83
## - age         1   223.77 255.77
## - G1          1   397.23 429.23
##
## Step:  AIC=242.48
## G3 ~ school + age + Fjob + failures + paid + activities + higher +
##     romantic + goout + health + absences + G1
##
##              Df Deviance    AIC
## - romantic    1   212.27 242.27
## - activities  1   212.27 242.27
## - paid        1   212.45 242.45
## <none>            210.48 242.48
## - absences    1   212.55 242.55
## - health      1   212.67 242.67
## - higher      1   213.62 243.62
## - goout       1   213.97 243.97
## - Fjob        4   220.02 244.02
## - school      1   214.10 244.10
## - failures    1   219.90 249.90
## - age         1   224.95 254.95
## - G1          1   398.28 428.28
##
## Step:  AIC=242.27
## G3 ~ school + age + Fjob + failures + paid + activities + higher +
##     goout + health + absences + G1
##
##              Df Deviance    AIC
## - activities  1   213.77 241.77
## - paid        1   213.82 241.82
## <none>            212.27 242.27
## - absences    1   214.47 242.47
## - health      1   214.80 242.80
## - goout       1   215.40 243.40
## - higher      1   215.79 243.79
## - Fjob        4   222.39 244.39
## - school      1   216.46 244.46
## - failures    1   222.19 250.19
## - age         1   225.54 253.54
## - G1          1   399.60 427.60
##
## Step:  AIC=241.77
## G3 ~ school + age + Fjob + failures + paid + higher + goout +
##     health + absences + G1
##
##            Df Deviance    AIC
## - paid      1   215.27 241.27
## <none>          213.77 241.77
## - absences  1   215.97 241.97
## - health    1   216.11 242.11
```

```
## - goout     1   216.39 242.39
## - higher    1   217.14 243.14
## - school    1   217.93 243.93
## - Fjob      4   224.55 244.55
## - failures  1   223.00 249.00
## - age       1   226.36 252.36
## - G1        1   401.37 427.37
##
## Step:  AIC=241.27
## G3 ~ school + age + Fjob + failures + higher + goout + health +
##     absences + G1
##
##            Df Deviance    AIC
## <none>          215.27 241.27
## - absences  1   217.34 241.34
## - goout     1   217.78 241.78
## - health    1   217.80 241.80
## - higher    1   218.36 242.36
## - school    1   219.39 243.39
## - Fjob      4   225.50 243.50
## - failures  1   224.65 248.65
## - age       1   227.79 251.79
## - G1        1   403.81 427.81
```

cannot use best subset as p $>$ 15

```
library(glmnet)
lambda.seq <- 10^seq(-2, 5, by = .1)
X <- data.matrix(mydata[,-length(mydata)])
y <- mydata[,length(mydata)]
cv.lasso <- cv.glmnet(X, y,
lambda = lambda.seq,
alpha = 1, #Lasso penalty
family = 'binomial')
plot(cv.lasso)
```




```
coef(cv.lasso, s = "lambda.min")
```

```
## 32 x 1 sparse Matrix of class "dgCMatrix"
##                       s1
## (Intercept) -11.01015748
## school       -0.21094354
## sex           .
## age           0.18857645
## address       .
## famsize       .
## Pstatus       .
## Medu          .
## Fedu          .
## Mjob          .
## Fjob         -0.21464698
## reason        .
## guardian      .
## traveltime    0.01131966
## studytime     .
## failures     -0.37503604
## schoolsup     .
## famsup        .
## paid         -0.06885427
## activities    .
## nursery       .
```

```
## higher        0.31040661
## internet      .
## romantic     -0.04981902
## famrel        .
## freetime      .
## goout        -0.02862320
## Dalc          .
## Walc         -0.06009627
## health        .
## absences      .
## G1            1.08155835
```

```
#reload dataset
mydata <- read.csv("student-por.csv", sep=";", header = TRUE)
```

```
# G3 is being set to a categorical variable. 0 if <10 (fail), 1 if >=10 (pass)
mydata['G3'][mydata['G3']<10] <- 0
mydata['G3'][mydata['G3']>=10] <- 1

# Removing points that are problematic in the graphs
mydata <- mydata[-647,]
mydata <- mydata[-1,]
```

```
attach(mydata)

#Remove outlier rows
# mydata <- mydata[!mydata$G3 == 0, ]
# mydata <- mydata[!mydata$G2 == 0, ]
# mydata <- mydata[!mydata$G1 == 0, ]
# mydata <- mydata[!mydata$G3 == 1,]

# Compute the correlation matrix
# Testing multicolinearity
#change data to factors

cols_to_factor <- c("school", "sex", "address", "famsize", "Pstatus",
                    "Mjob", "Fjob", "reason", "guardian", "schoolsup",
                    "famsup","paid","activities",
                    "nursery","higher","internet", "romantic")

# Use lapply to apply as.factor() to each column
mydata[, cols_to_factor] <- lapply(mydata[, cols_to_factor], as.factor)
```

```
lassofit <- glm(G3 ~ G2 + G1 + Walc + famrel + romantic + activities + failures, data=mydata, family =

forwardfit <- glm (G3 ~ G2 + famrel + absences + G1 + age + activities + Walc +
    romantic + school, data= mydata, family = "binomial")

summary(lassofit)
```

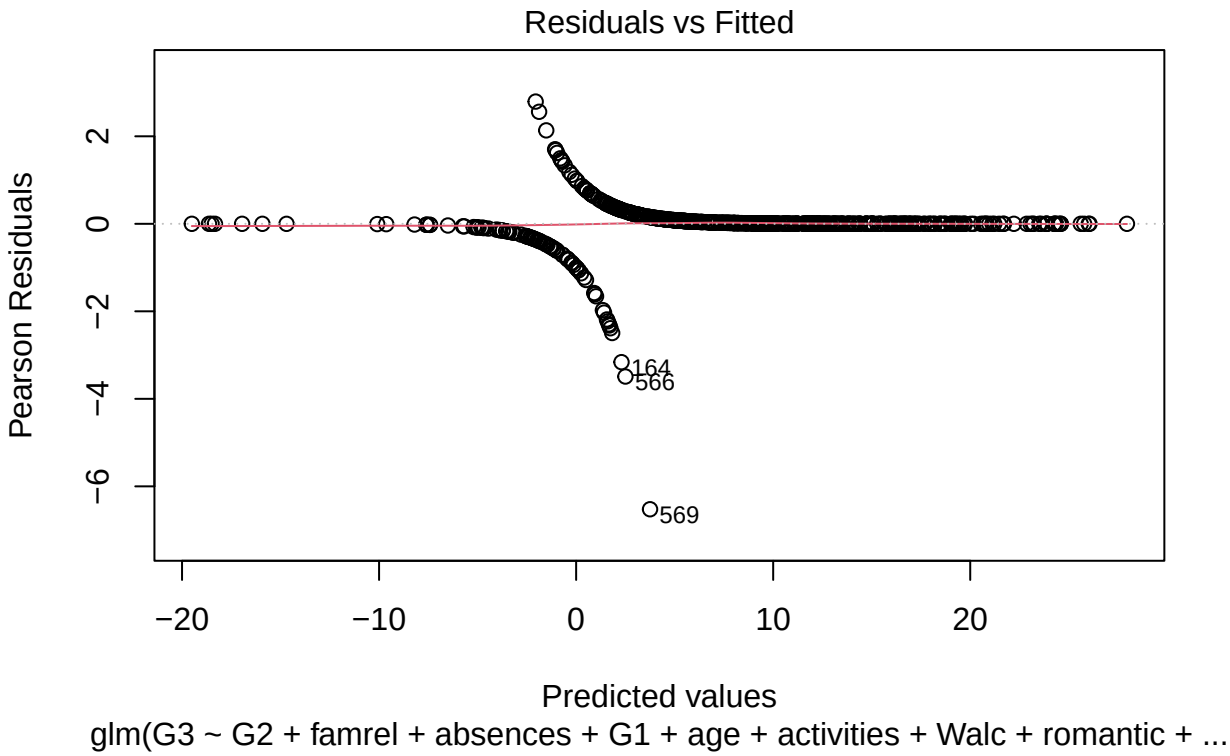
```
##
## Call:
```

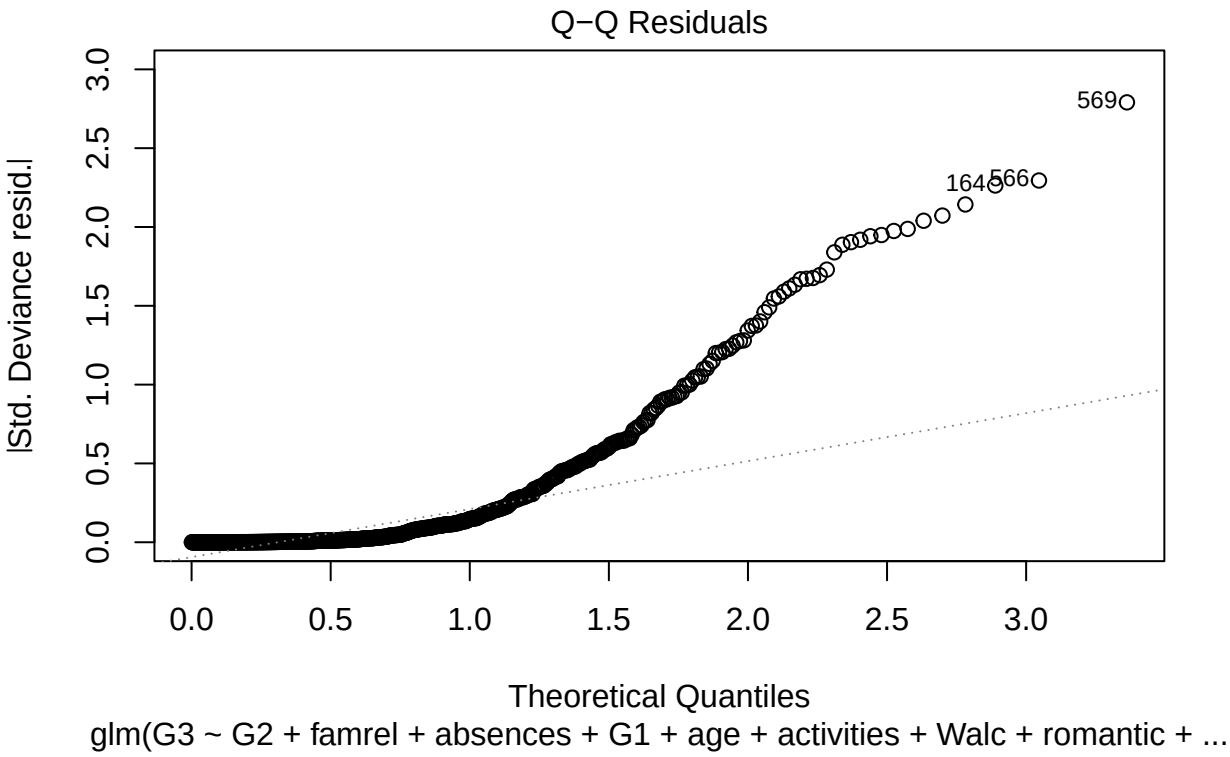
```
## glm(formula = G3 ~ G2 + G1 + Walc + famrel + romantic + activities +
##     failures, family = "binomial", data = mydata)
##
## Coefficients:
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -19.41365    2.56816  -7.559 4.05e-14 ***
## G2              1.59999    0.25905   6.176 6.56e-10 ***
## G1              0.83387    0.17544   4.753 2.00e-06 ***
## Walc           -0.29954    0.15360  -1.950   0.0512 .
## famrel         -0.26090    0.19629  -1.329   0.1838
## romanticyes    -0.49868    0.42691  -1.168   0.2428
## activitiesyes   0.26530    0.42347   0.626   0.5310
## failures       -0.06839    0.23484  -0.291   0.7709
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 553.70  on 646  degrees of freedom
## Residual deviance: 164.09  on 639  degrees of freedom
## AIC: 180.09
##
## Number of Fisher Scoring iterations: 9
```

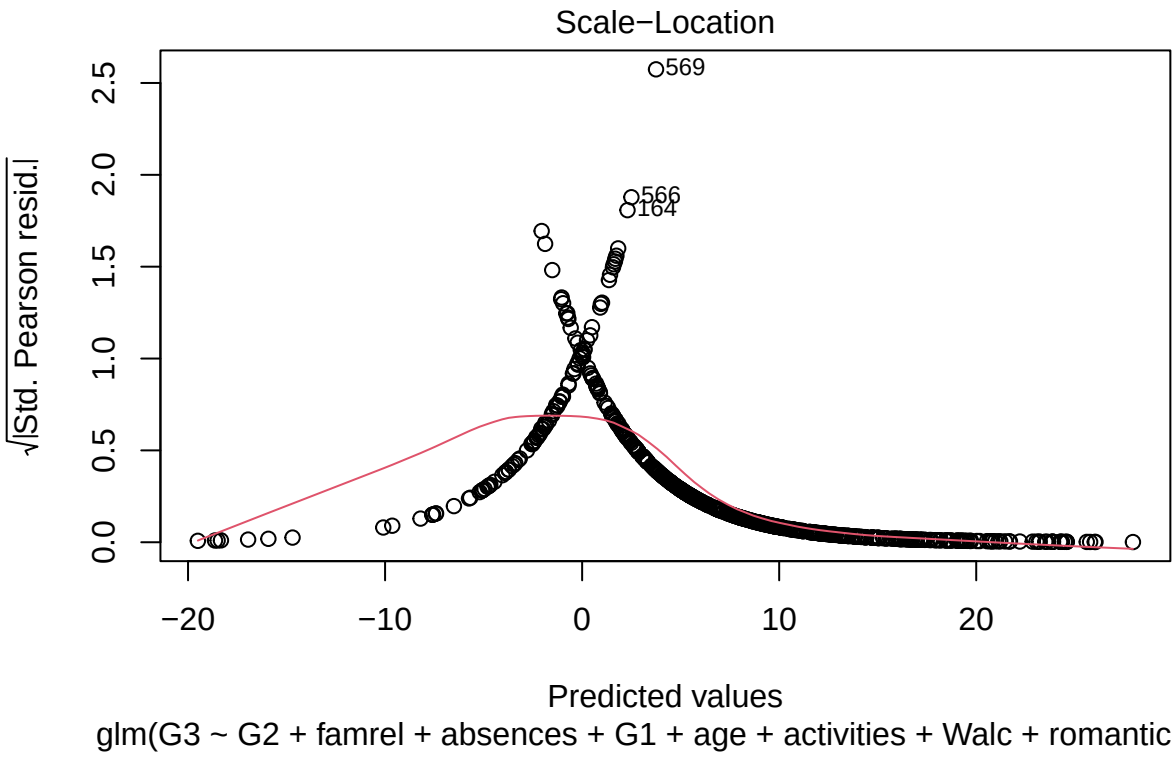
```
summary(forwardfit)
```

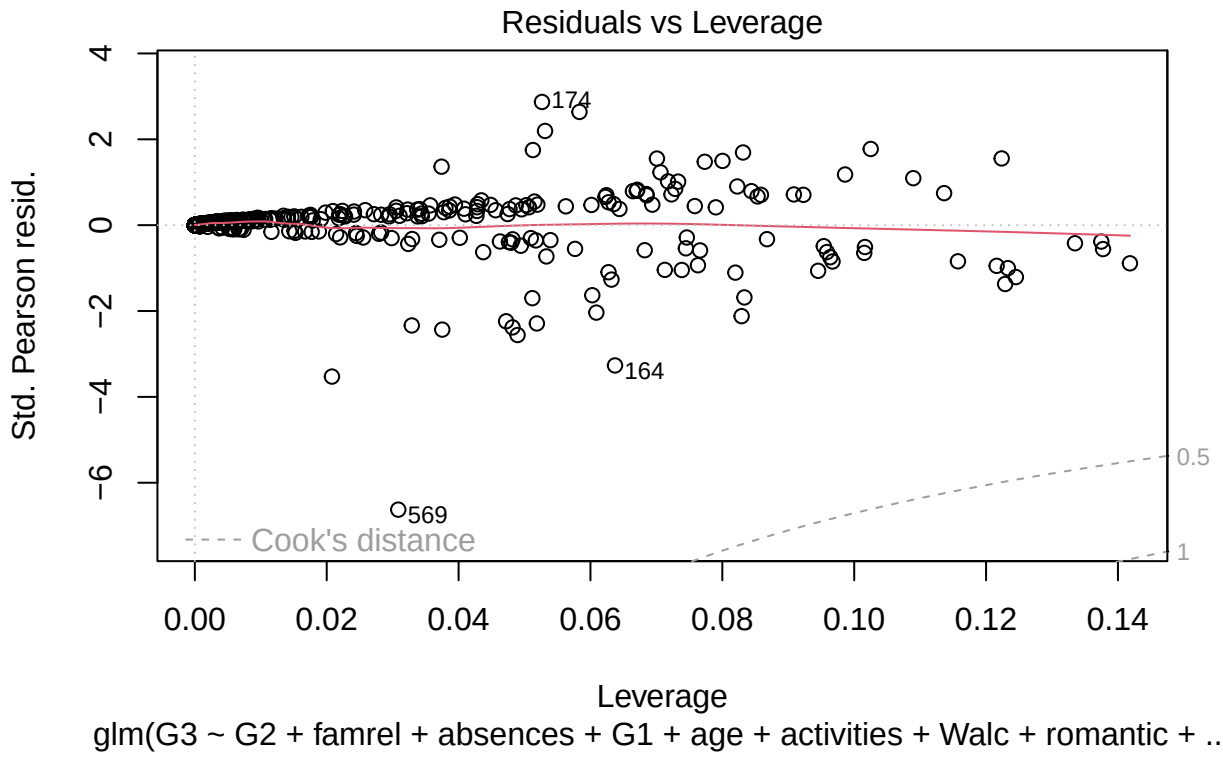
```
##
## Call:
## glm(formula = G3 ~ G2 + famrel + absences + G1 + age + activities +
##     Walc + romantic + school, family = "binomial", data = mydata)
##
## Coefficients:
##                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -30.105407   5.088372  -5.917 3.29e-09 ***
## G2              1.763783   0.294560   5.988 2.13e-09 ***
## famrel         -0.302084   0.205993  -1.466  0.14252
## absences       -0.008551   0.048935  -0.175  0.86129
## G1              0.908398   0.205217   4.427 9.58e-06 ***
## age             0.558481   0.194354   2.874  0.00406 **
## activitiesyes   0.416171   0.445234   0.935  0.34993
## Walc           -0.392442   0.166164  -2.362  0.01819 *
## romanticyes    -0.945009   0.478297  -1.976  0.04818 *
## schoolMS       -0.690233   0.478017  -1.444  0.14875
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 553.70  on 646  degrees of freedom
## Residual deviance: 153.49  on 637  degrees of freedom
## AIC: 173.49
##
## Number of Fisher Scoring iterations: 9
```

```
plot(forwardfit)
```

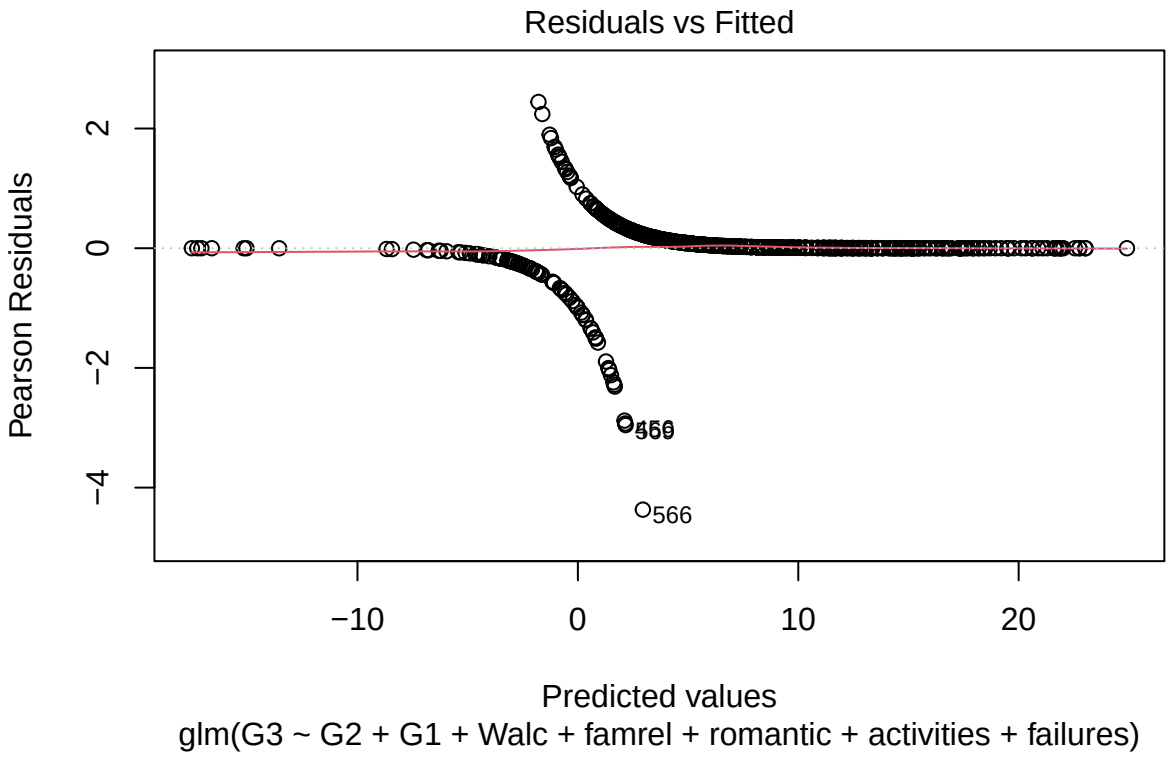


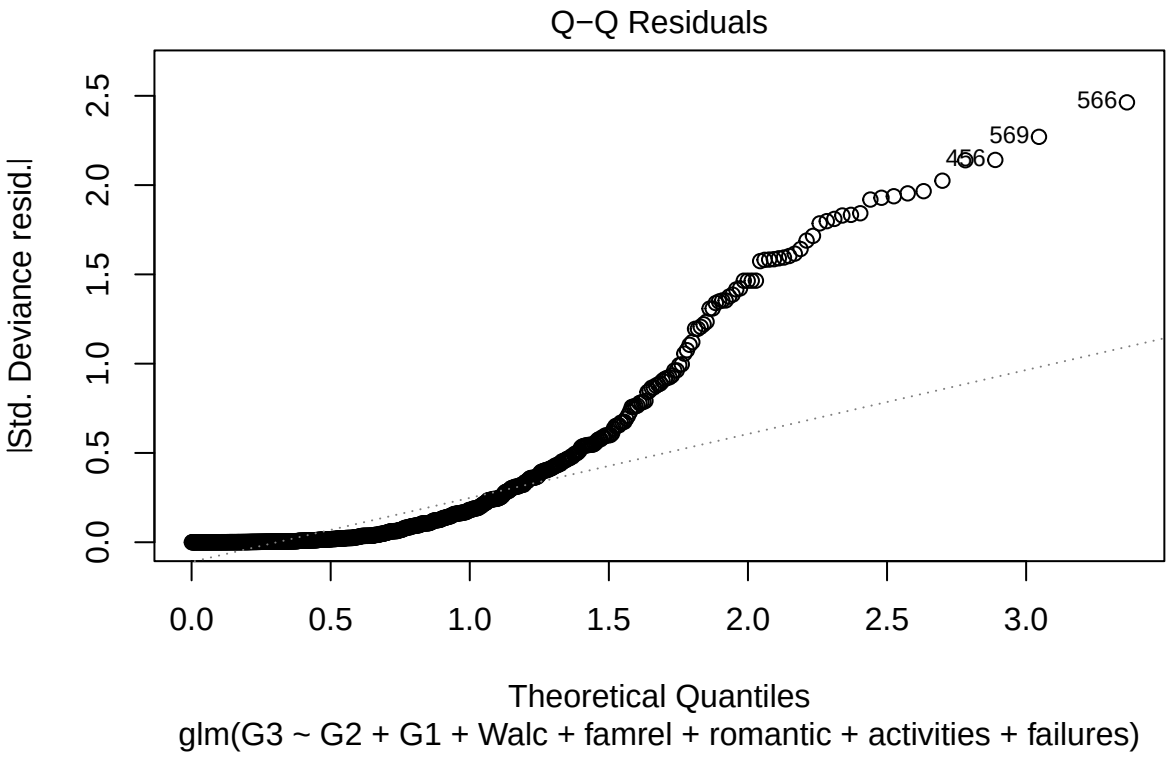

Q−Q Residuals
|Std. Deviance resid.|
0.0
0.5
1.0
1.5
2.0
2.5
3.0
569
164
566
Theoretical Quantiles
glm(G3 ~ G2 + famrel + absences + G1 + age + activities + Walc + romantic + ...

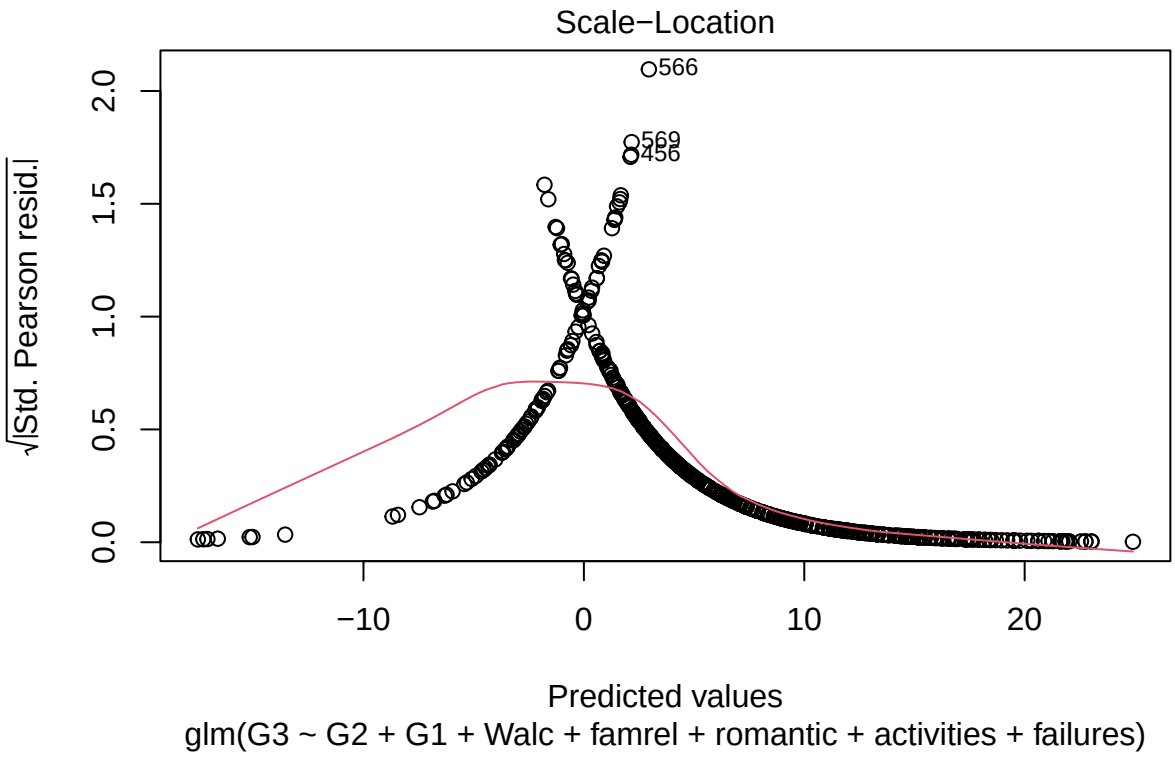

Scale−Location
√|Std. Pearson resid.|
0.0
0.5
1.0
1.5
2.0
2.5
569
566
164
−20
−10
0
10
20
Predicted values
glm(G3 ~ G2 + famrel + absences + G1 + age + activities + Walc + romantic + ...

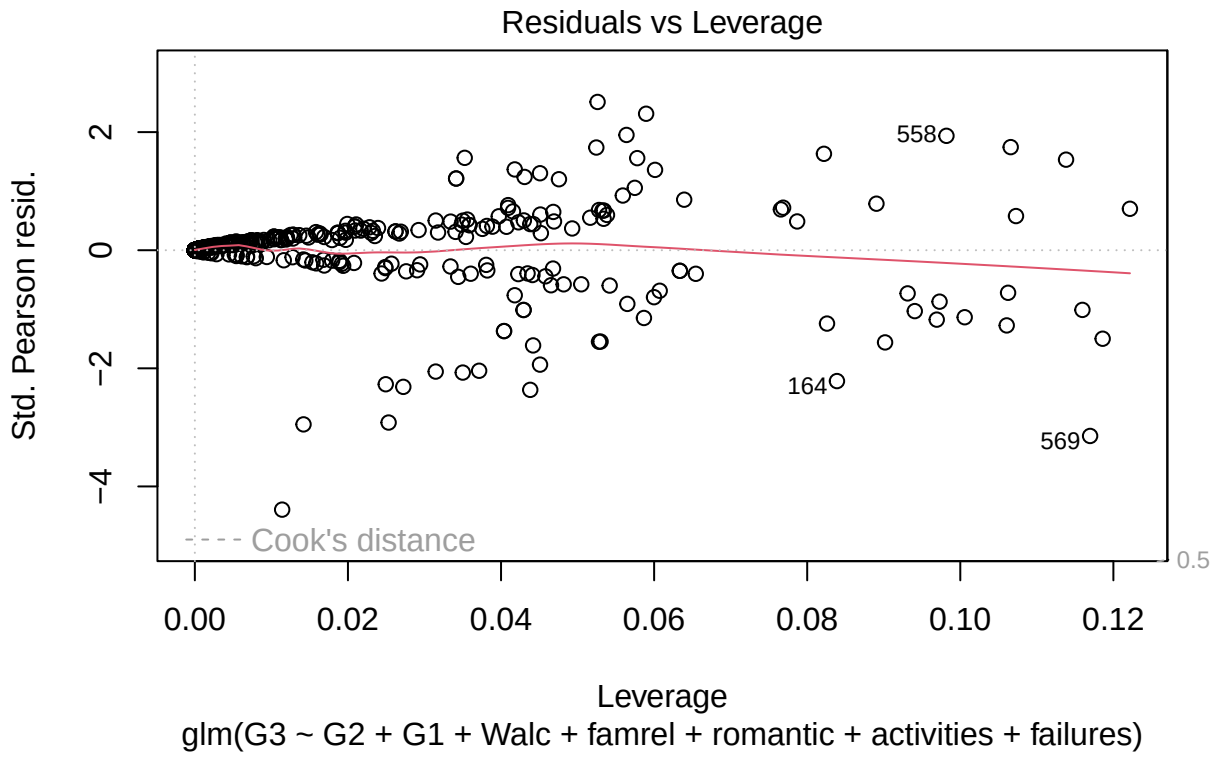



```
plot(lassofit)
```

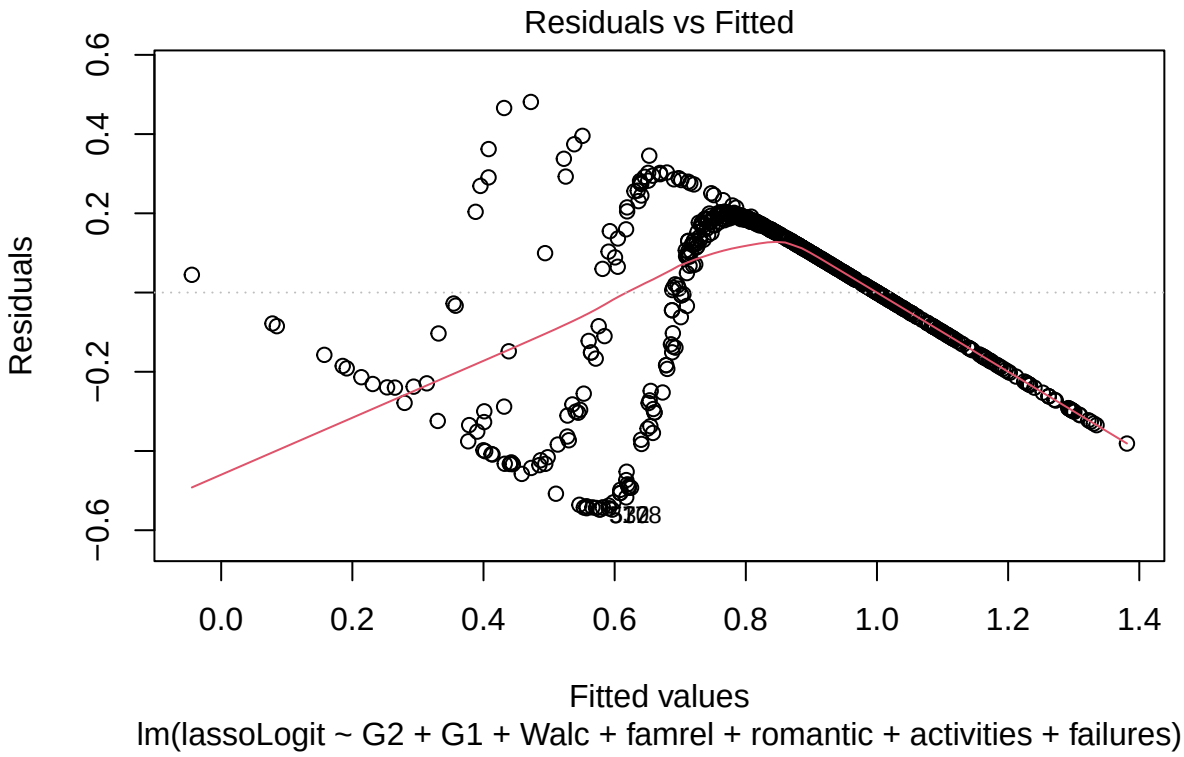

Residuals vs Fitted
Pearson Residuals
2
0
−2
−4
−10
0
10
20
566
Predicted values
glm(G3 ~ G2 + G1 + Walc + famrel + romantic + activities + failures)

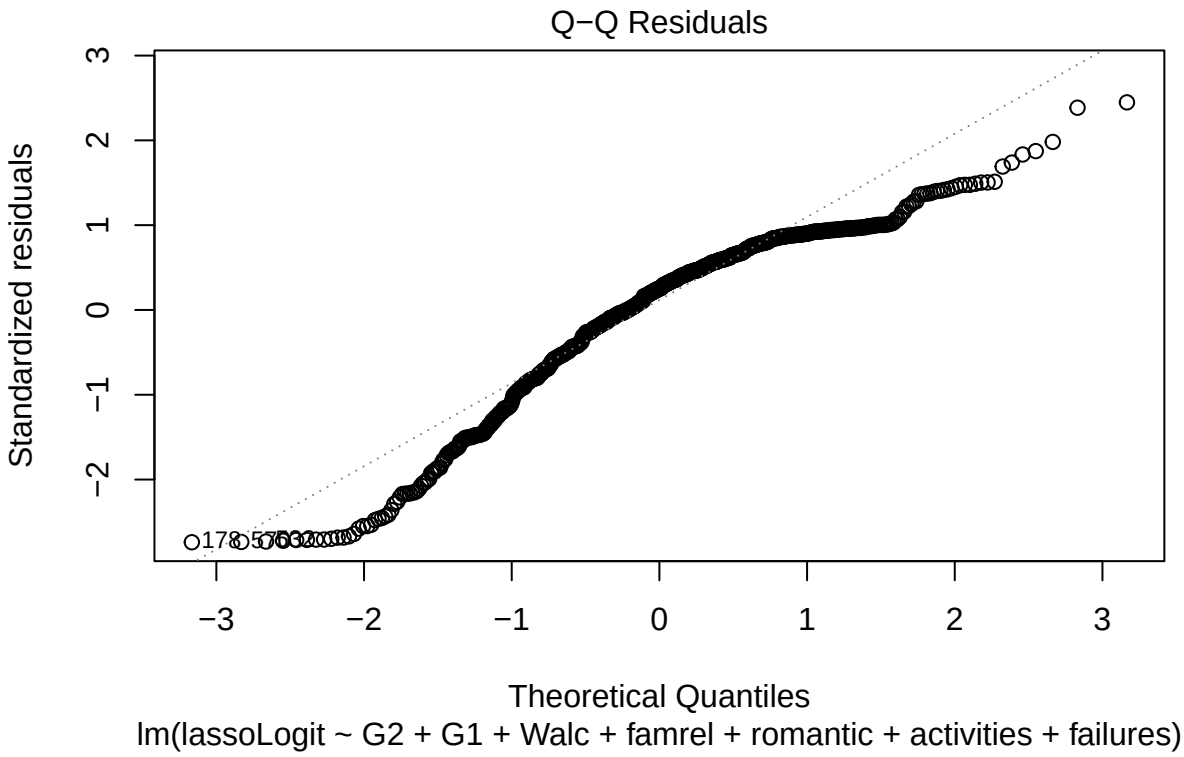

Q−Q Residuals
|Std. Deviance resid.|
0.0
0.5
1.0
1.5
2.0
2.5
3.0
566
569
456
Theoretical Quantiles
glm(G3 ~ G2 + G1 + Walc + famrel + romantic + activities + failures)

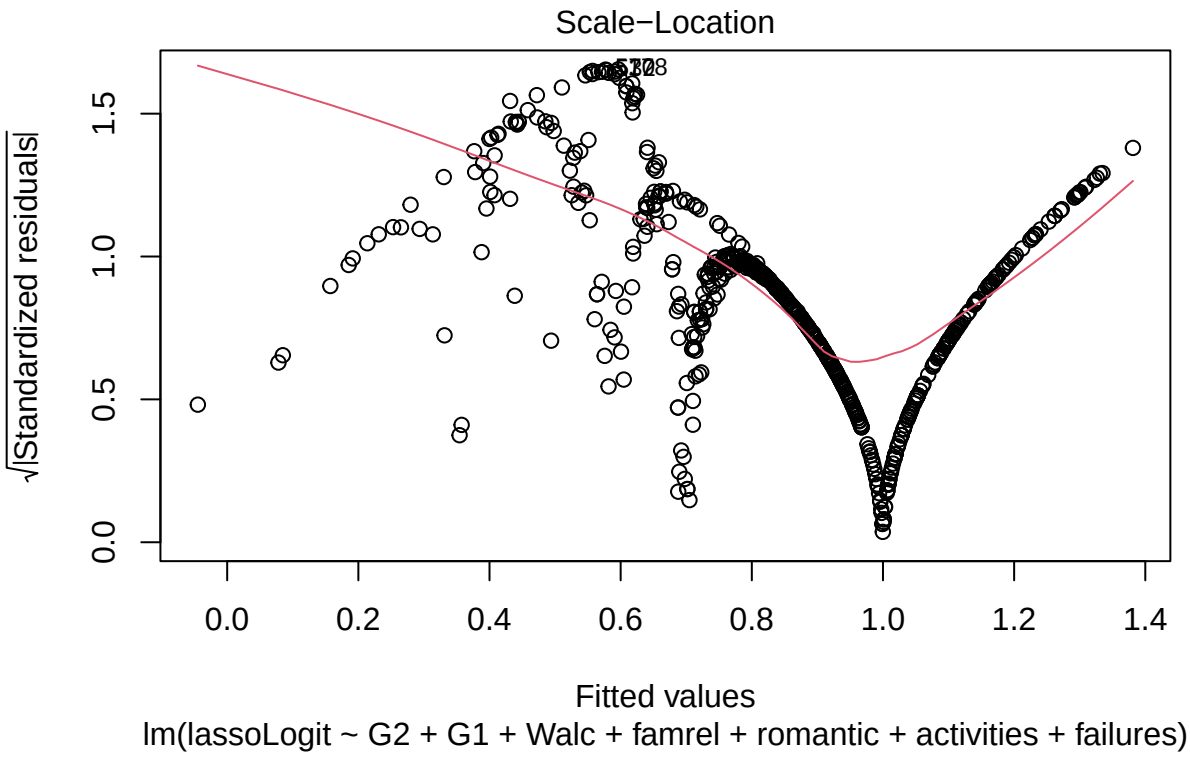

Scale−Location
√|Std. Pearson resid.|
0.0
0.5
1.0
1.5
2.0
566
569
456
−10
0
10
20
Predicted values
glm(G3 ~ G2 + G1 + Walc + famrel + romantic + activities + failures)

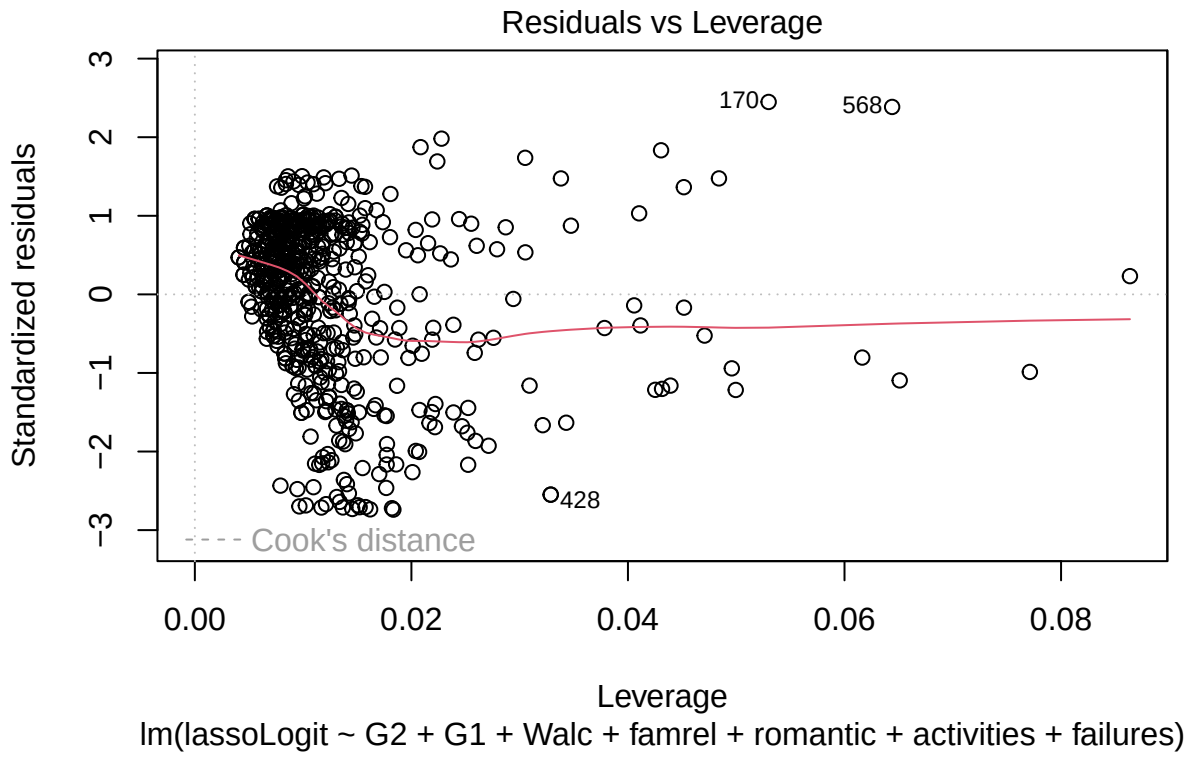



```
rom <- ifelse(romantic=="yes",1,0)
act<- ifelse(activities=="yes",1,0)
scoo <- ifelse(school =="GP",1,0)

lassoLogit <- exp(coef(lassofit)[1] + G2*coef(lassofit)[2] + G1*coef(lassofit)[3] + Walc*coef(lassofit)

forwardLogit <- exp(coef(forwardfit)[1] + G2*coef(forwardfit)[2] + famrel*coef(forwardfit)[3] + absence


#plotting inverse logit
plot(lm(lassoLogit~ G2 + G1 + Walc + famrel + romantic + activities + failures))
```

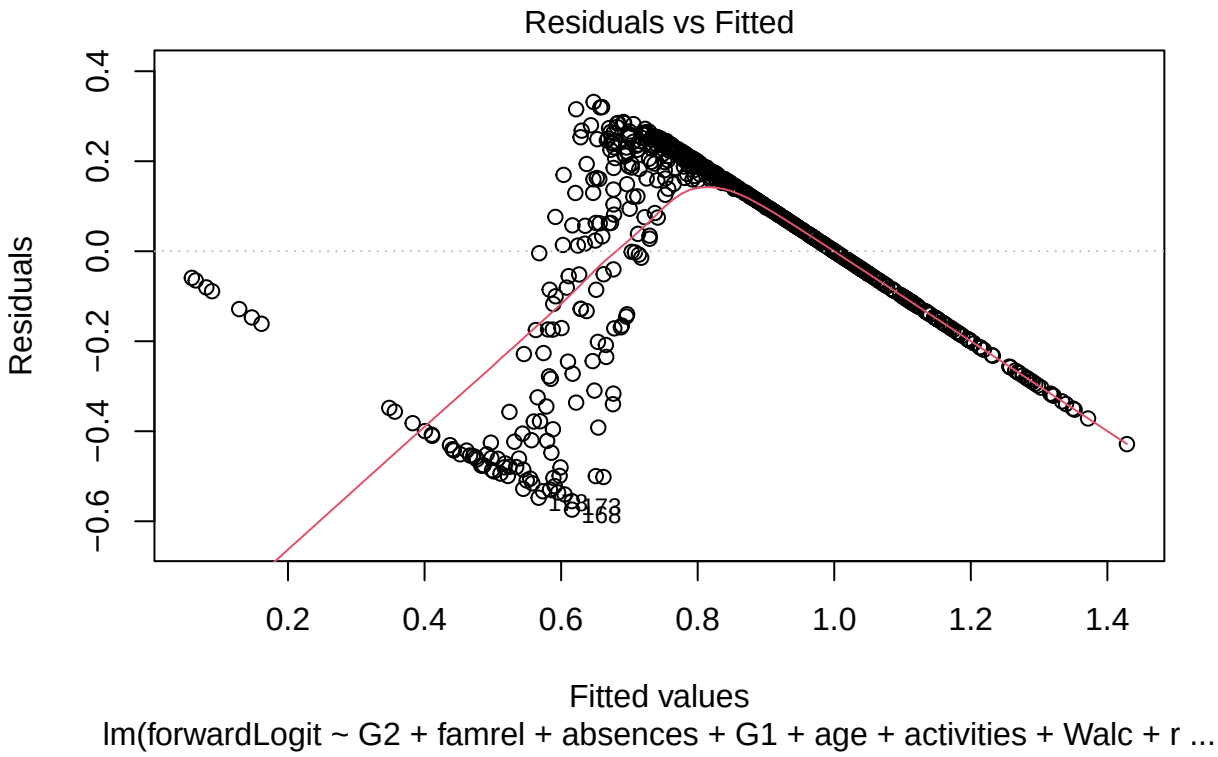

Residuals vs Fitted
Residuals
0.6
0.4
0.2
−0.2
−0.6
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
Fitted values
lm(lassoLogit ~ G2 + G1 + Walc + famrel + romantic + activities + failures)

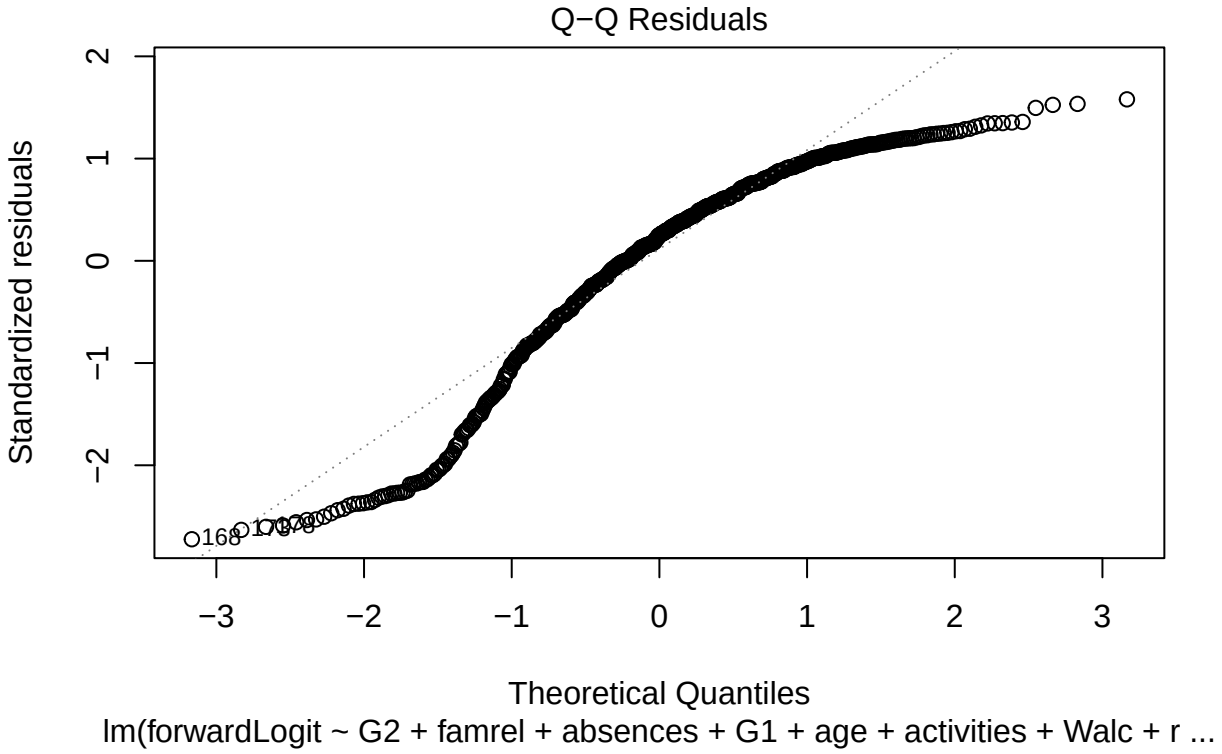

Q−Q Residuals
Standardized residuals
3
2
1
0
−1
−2
−3
−2
−1
0
1
2
3
Theoretical Quantiles
lm(lassoLogit ~ G2 + G1 + Walc + famrel + romantic + activities + failures)

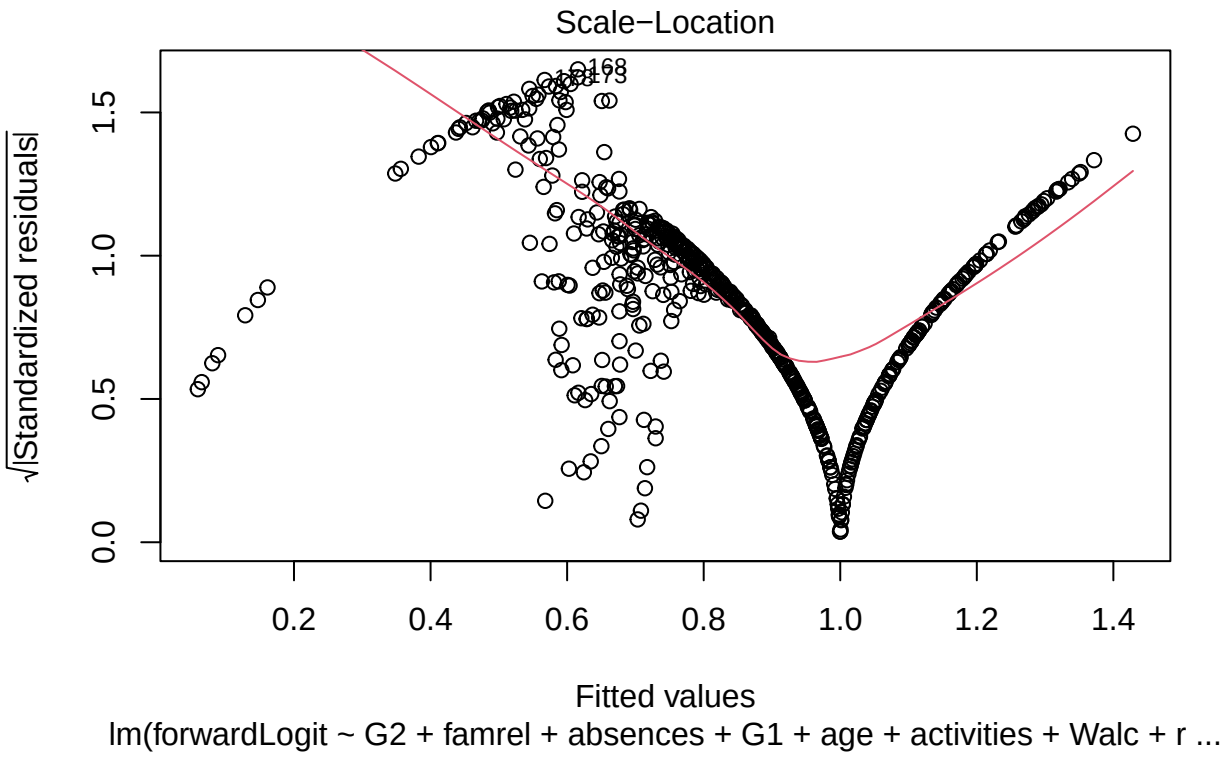

Scale−Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
Fitted values
lm(lassoLogit ~ G2 + G1 + Walc + famrel + romantic + activities + failures)

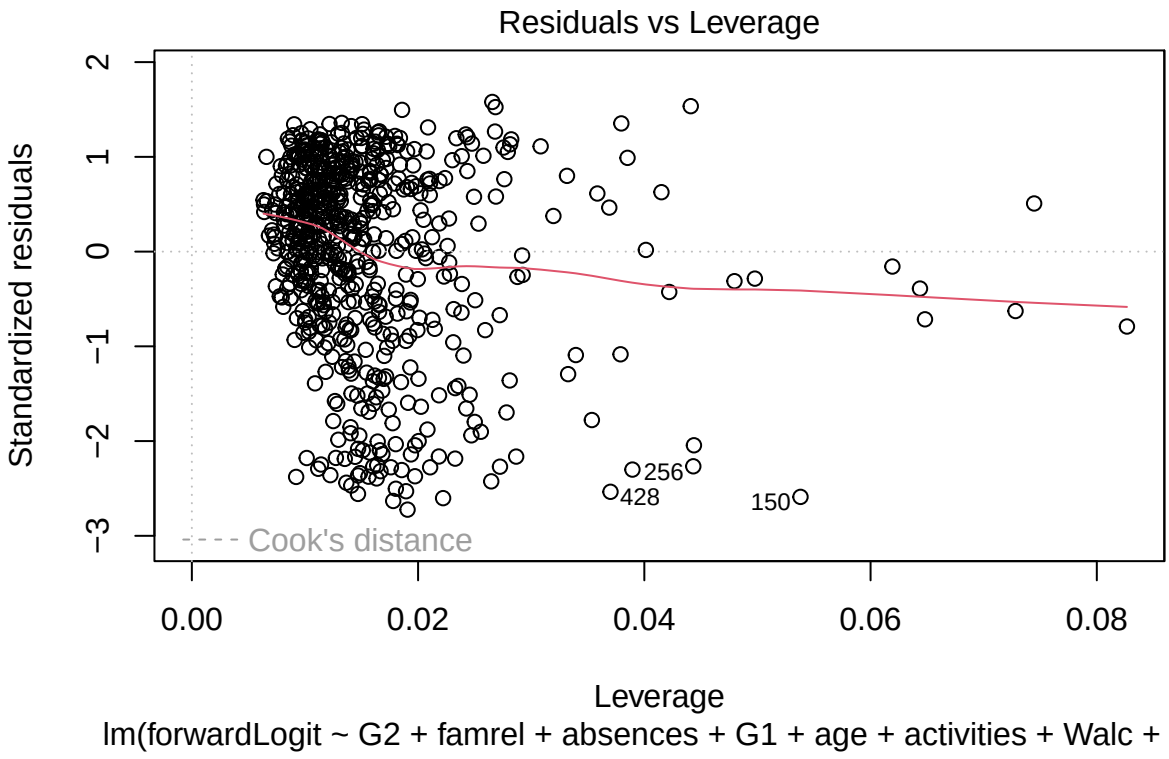



```
plot(lm (forwardLogit ~ G2 + famrel + absences + G1 + age + activities + Walc + romantic + school))
```


Residuals vs Fitted
Residuals
0.4
0.2
0.0
−0.2
−0.4
−0.6
0.2
0.4
0.6
0.8
1.0
1.2
1.4
Fitted values
lm(forwardLogit ~ G2 + famrel + absences + G1 + age + activities + Walc + r ...


Q−Q Residuals
Standardized residuals
2
1
0
−1
−2
−3
−2
−1
0
1
2
3
Theoretical Quantiles
lm(forwardLogit ~ G2 + famrel + absences + G1 + age + activities + Walc + r ...


Scale−Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
0.2
0.4
0.6
0.8
1.0
1.2
1.4
Fitted values
lm(forwardLogit ~ G2 + famrel + absences + G1 + age + activities + Walc + r ...




```
#performing log transformations as the plots for the forwardfit and lassofit violates linear assumption

# mydata$G3_log <- log(mydata$G3)
# fit1 <- glm(G3_log ~ G1 + G2+ Walc + famrel + romantic + activities + failures, data = mydata)
# fit2 <- glm (G3_log ~ G2 + famrel + absences + G1 + age + activities + Walc + romantic + school, data
#
#
# fit_log <- glm(fit2)
#
# summary(fit_log)
#
# plot(fit_log)
```

removing G2 due to high multicolinearity

```
lassofit <- glm(G3 ~  G1 + Walc + famrel + romantic + activities + failures, family = "binomial")
forwardfit <- glm (G3 ~  famrel + absences + G1 + age + activities + Walc +
    romantic + school, family = "binomial")
summary(lassofit)
```

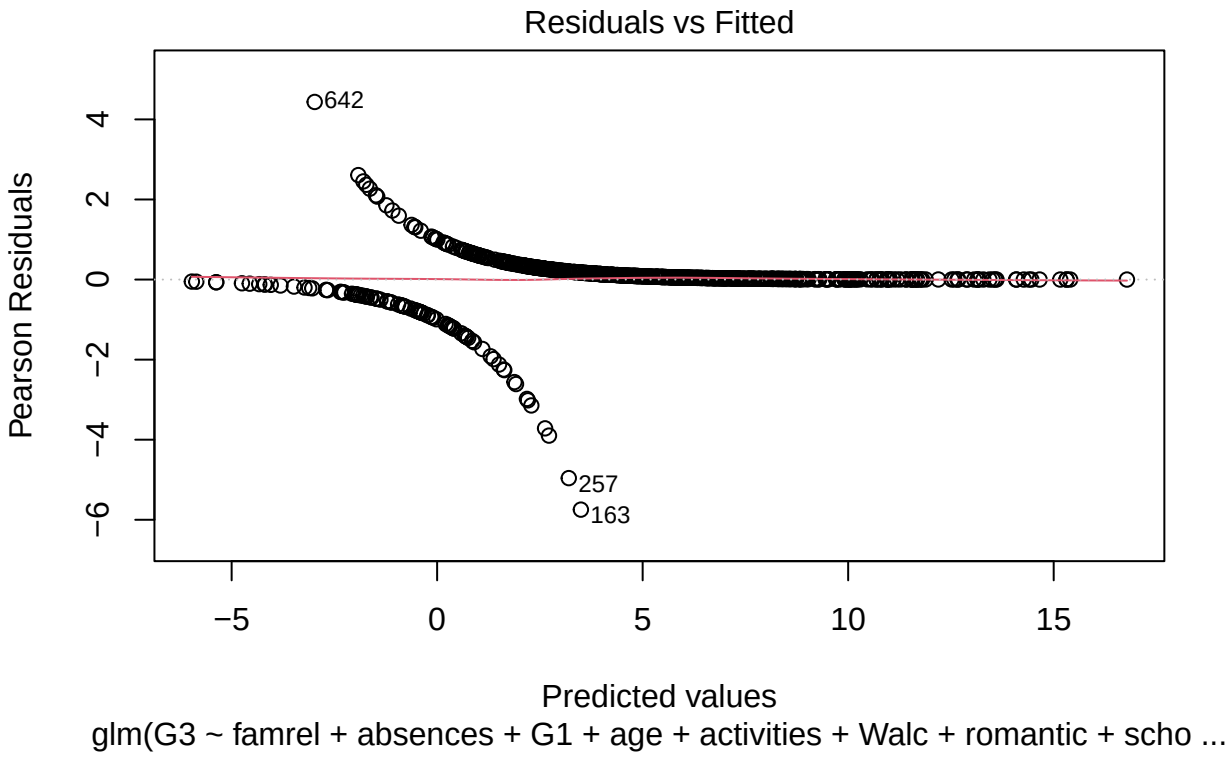
```
##
## Call:
## glm(formula = G3 ~ G1 + Walc + famrel + romantic + activities +
##     failures, family = "binomial")
##
```

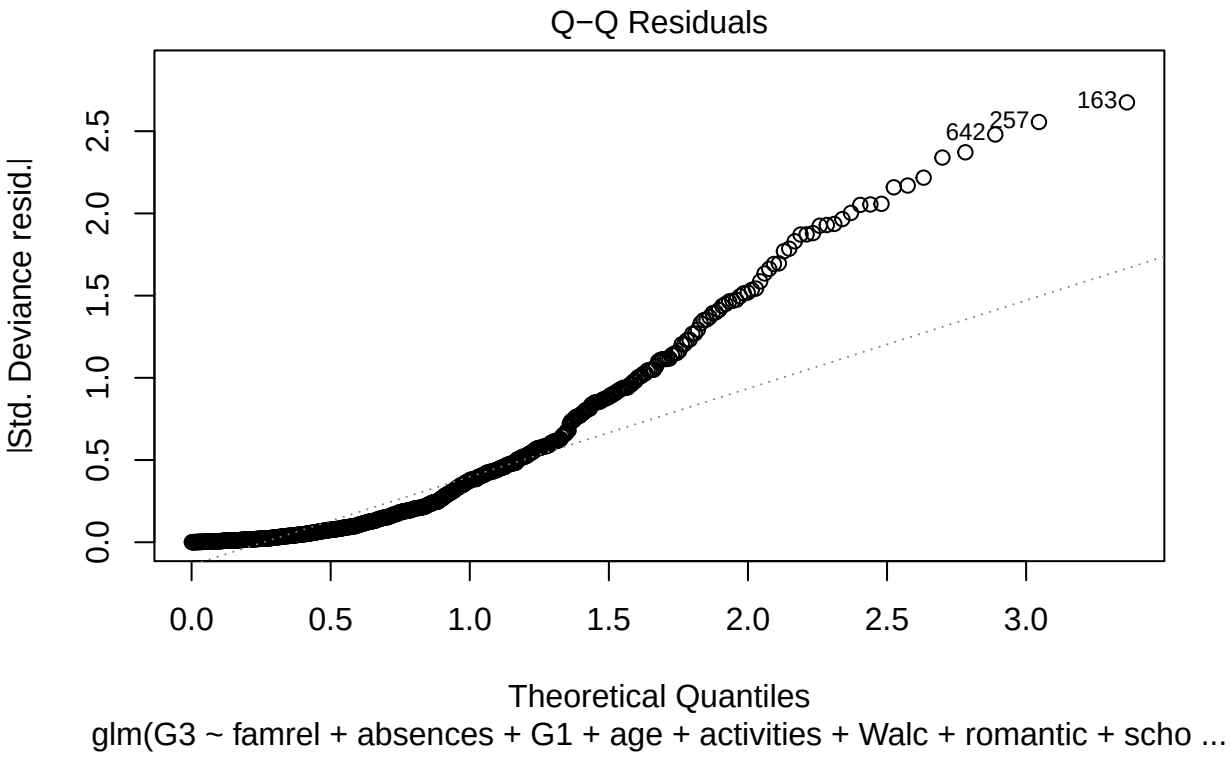
```
## Coefficients:
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -10.63693    1.47775  -7.198 6.11e-13 ***
## G1              1.37010    0.14507   9.444  < 2e-16 ***
## Walc           -0.23047    0.12233  -1.884   0.0596 .
## famrel          0.00852    0.16020   0.053   0.9576
## romanticyes    -0.36258    0.34054  -1.065   0.2870
## activitiesyes   0.38291    0.34379   1.114   0.2654
## failures       -0.37154    0.20194  -1.840   0.0658 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 553.70  on 646  degrees of freedom
## Residual deviance: 238.55  on 640  degrees of freedom
## AIC: 252.55
##
## Number of Fisher Scoring iterations: 7
```

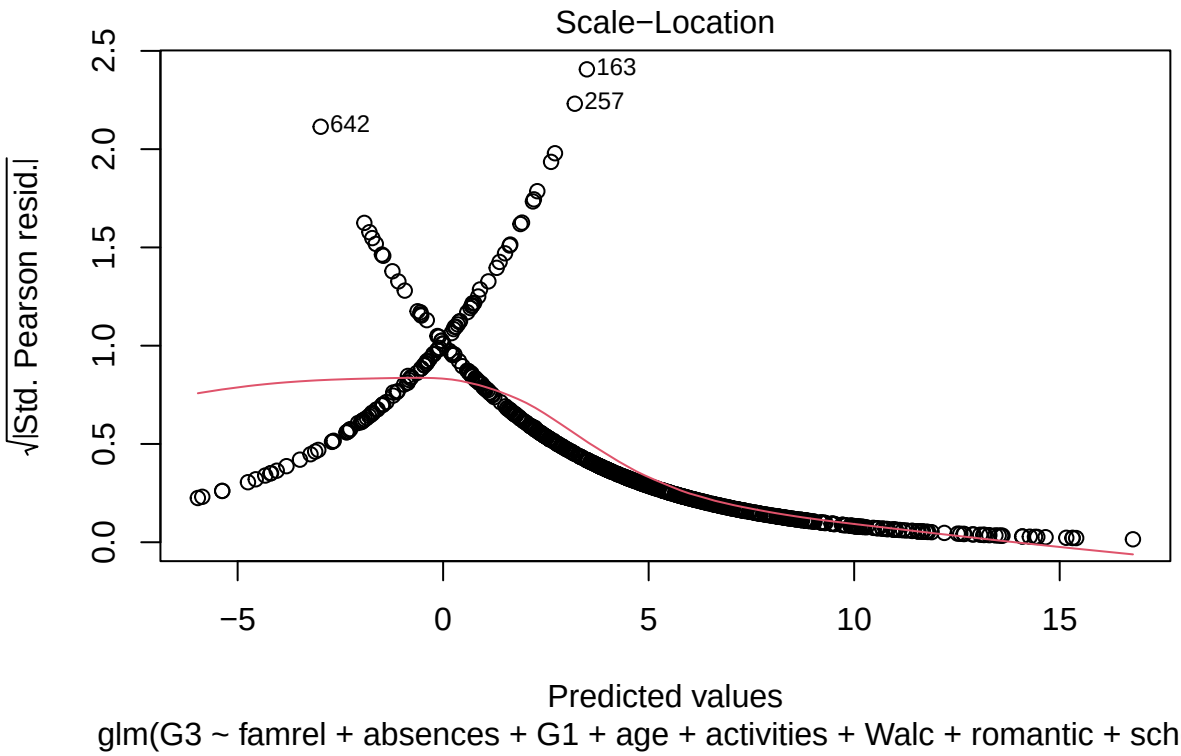
```
summary(forwardfit)
```

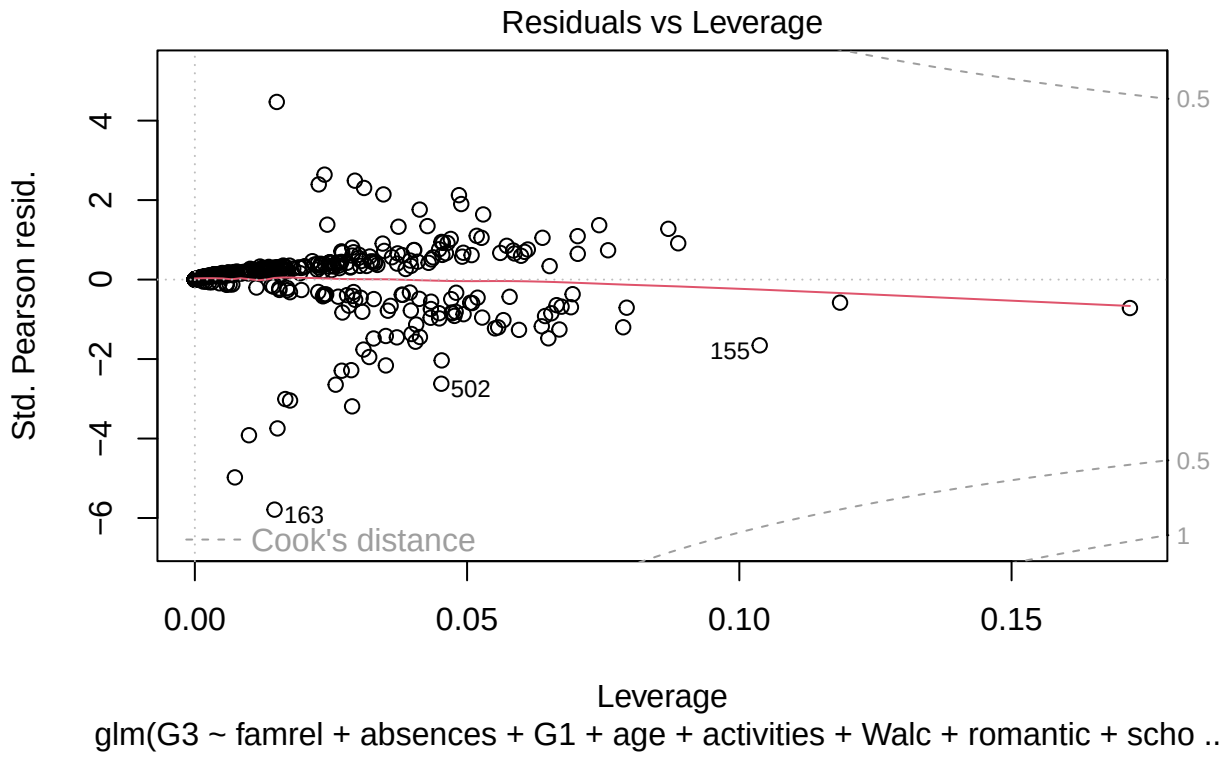
```
##
## Call:
## glm(formula = G3 ~ famrel + absences + G1 + age + activities +
##     Walc + romantic + school, family = "binomial")
##
## Coefficients:
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -18.05893    3.33307  -5.418 6.02e-08 ***
## famrel          0.01381    0.16253   0.085  0.93231
## absences       -0.02870    0.03632  -0.790  0.42940
## G1              1.49026    0.16200   9.199  < 2e-16 ***
## age             0.38824    0.14740   2.634  0.00844 **
## activitiesyes   0.37006    0.34699   1.066  0.28621
## Walc           -0.24713    0.12418  -1.990  0.04659 *
## romanticyes    -0.61707    0.35589  -1.734  0.08294 .
## schoolMS       -0.28397    0.36821  -0.771  0.44057
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 553.70  on 646  degrees of freedom
## Residual deviance: 234.05  on 638  degrees of freedom
## AIC: 252.05
##
## Number of Fisher Scoring iterations: 8
```

```
plot(forwardfit)
```

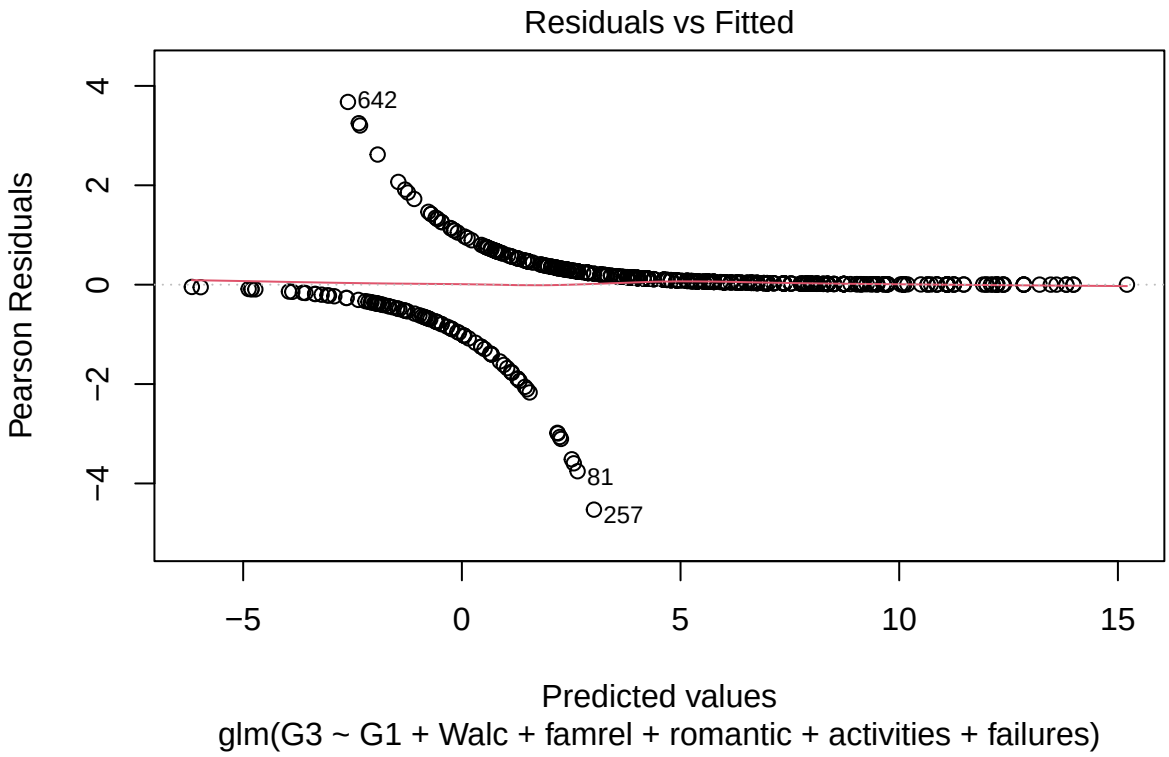

Residuals vs Fitted
642
257
163
Pearson Residuals
4
2
0
−2
−4
−6
−5
0
5
10
15
Predicted values
glm(G3 ~ famrel + absences + G1 + age + activities + Walc + romantic + scho ...

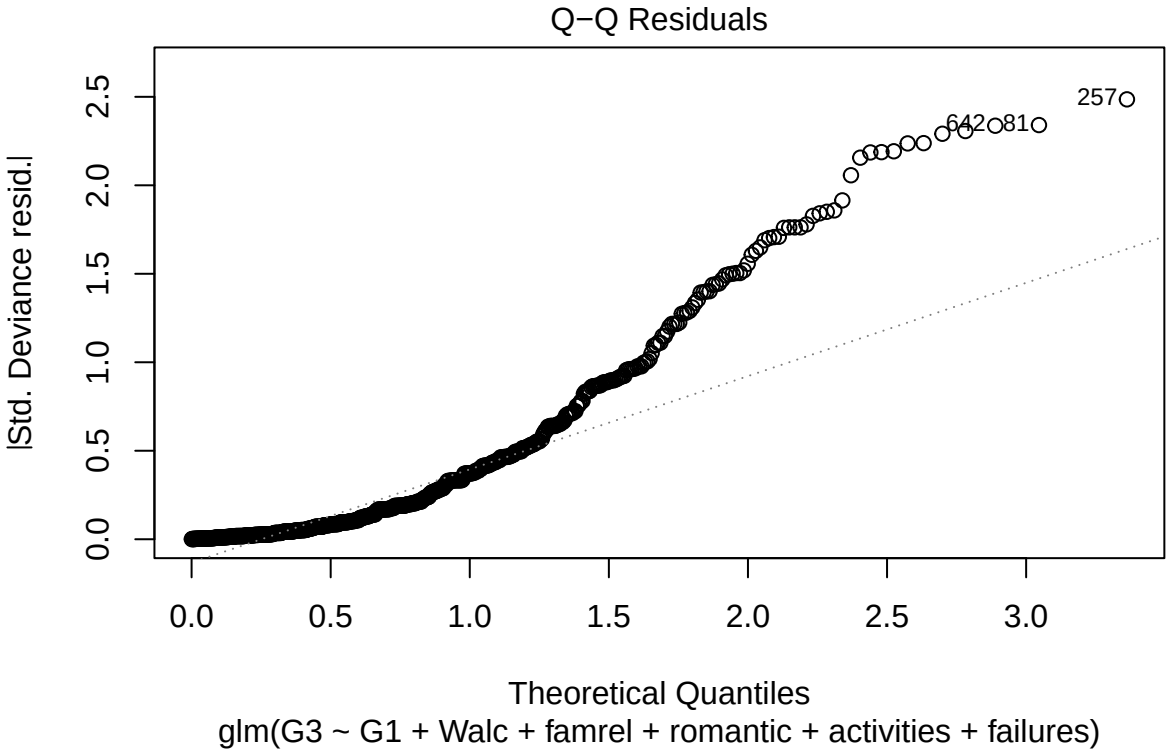

Q−Q Residuals
|Std. Deviance resid.|
0.0
0.5
1.0
1.5
2.0
2.5
0.0
0.5
1.0
1.5
2.0
2.5
3.0
163
257
642
Theoretical Quantiles
glm(G3 ~ famrel + absences + G1 + age + activities + Walc + romantic + scho ...

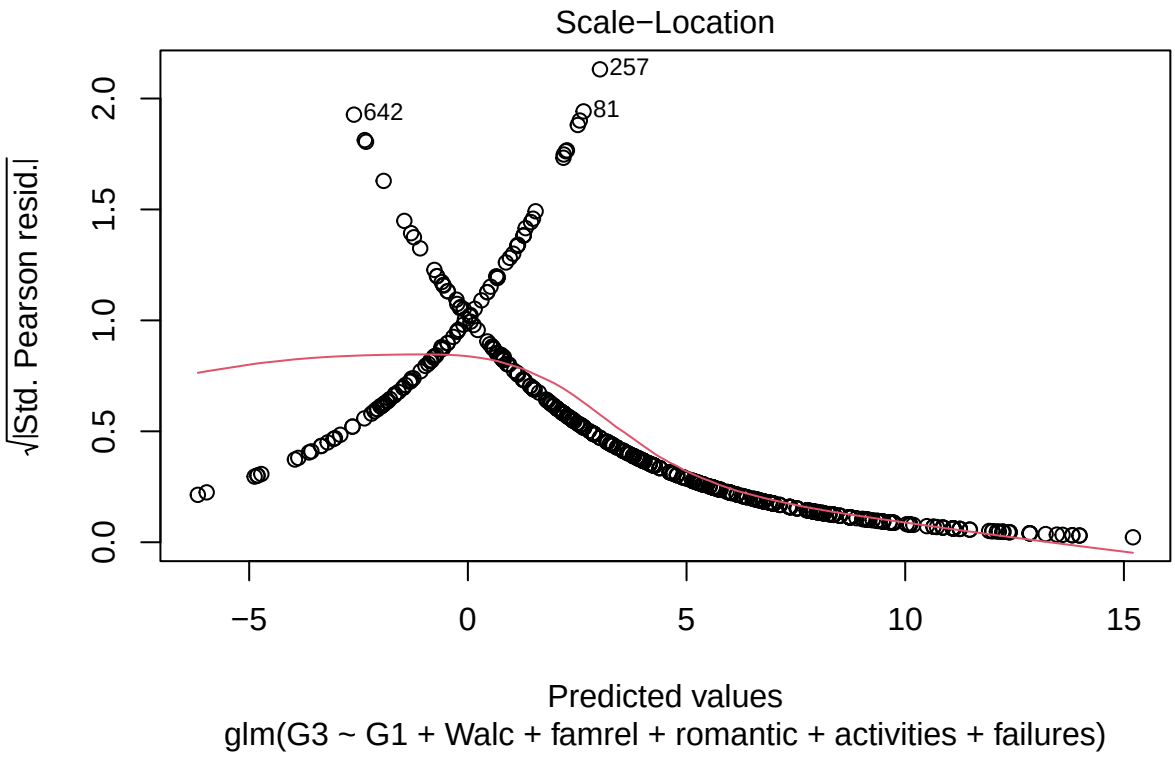

Scale−Location
√|Std. Pearson resid.|
0.0
0.5
1.0
1.5
2.0
2.5
163
257
642
−5
0
5
10
15
Predicted values
glm(G3 ~ famrel + absences + G1 + age + activities + Walc + romantic + scho ...

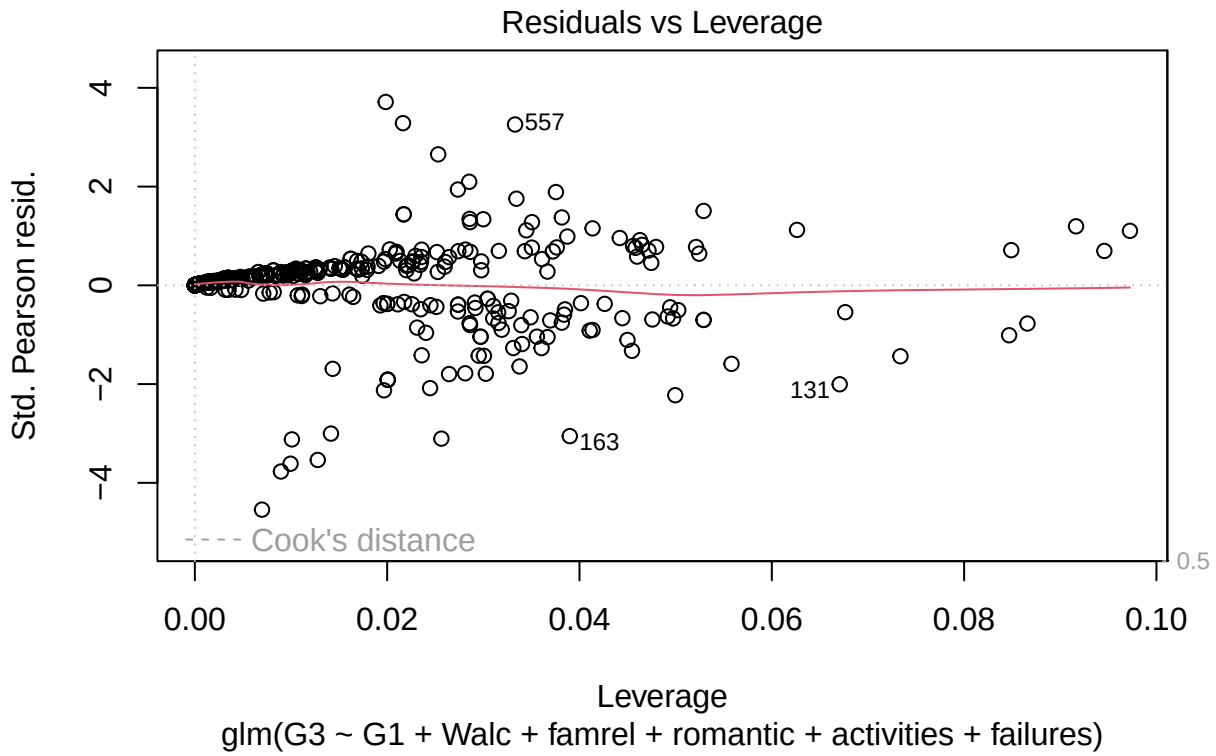



```
plot(lassofit)
```

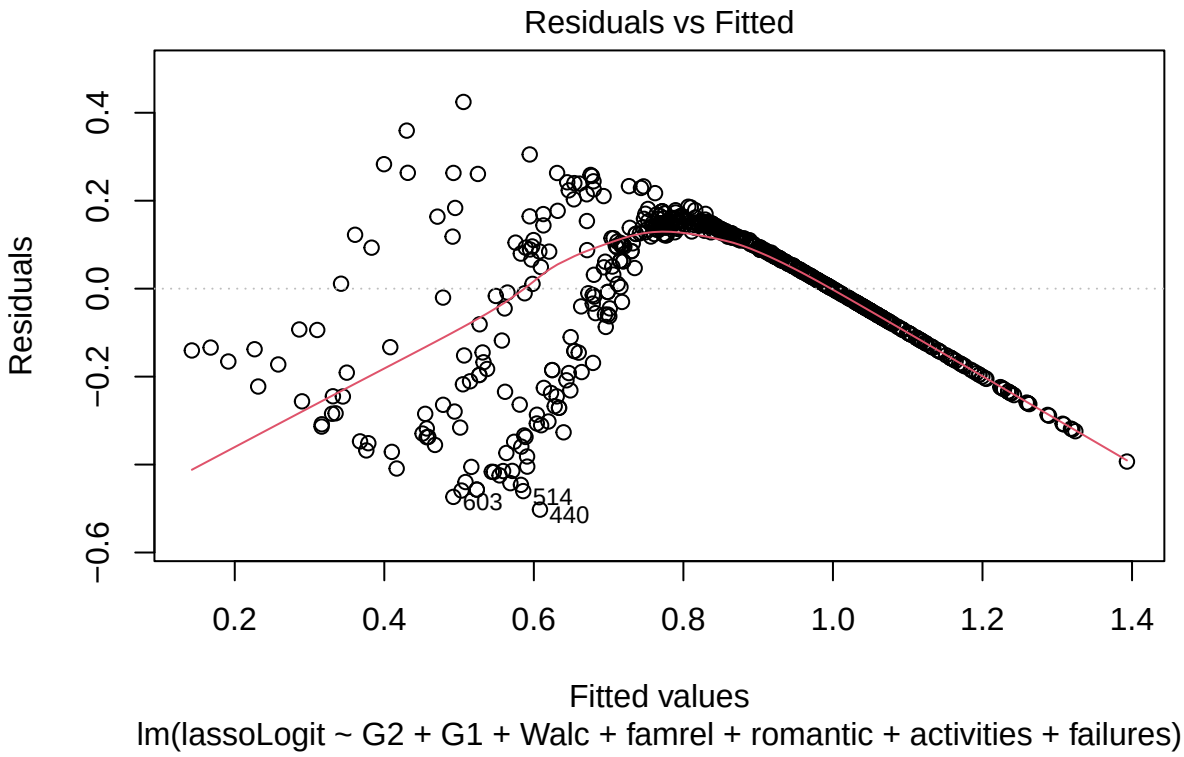

Residuals vs Fitted
Pearson Residuals
4
2
0
−2
−4
642
81
257
−5
0
5
10
15
Predicted values
glm(G3 ~ G1 + Walc + famrel + romantic + activities + failures)

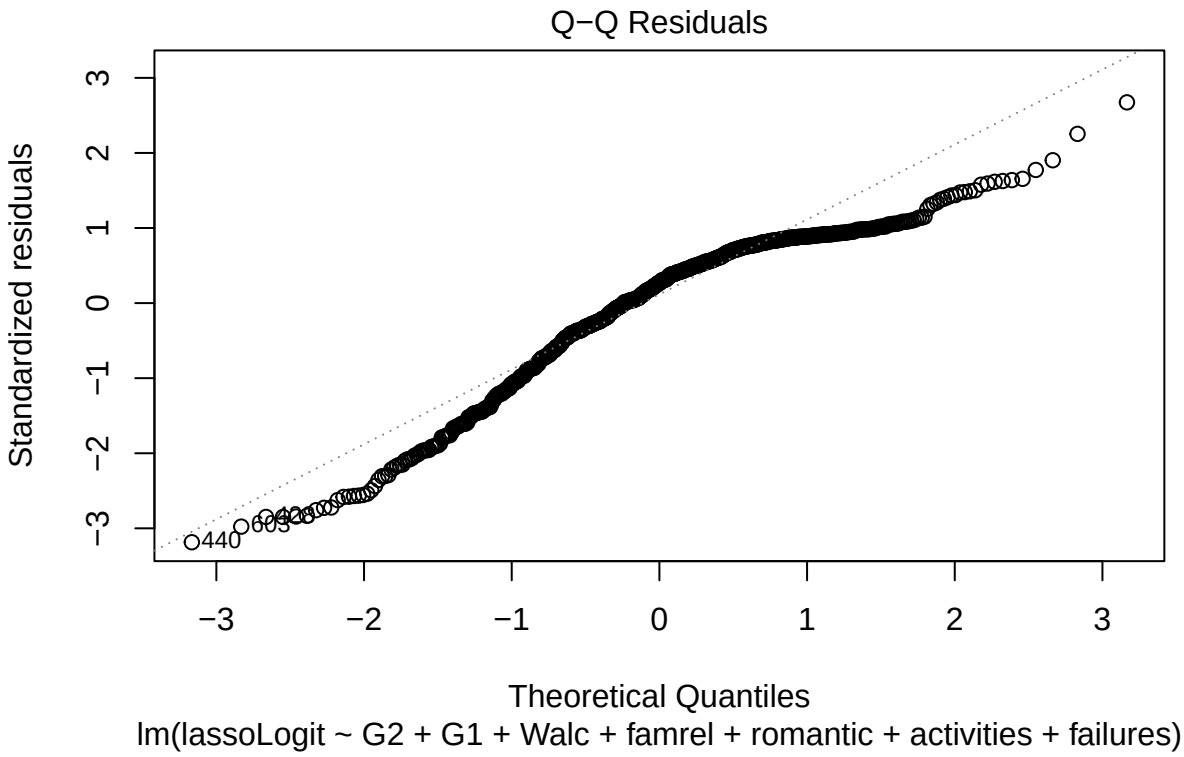

Q−Q Residuals
|Std. Deviance resid.|
0.0
0.5
1.0
1.5
2.0
2.5
3.0
257
642
81
Theoretical Quantiles
glm(G3 ~ G1 + Walc + famrel + romantic + activities + failures)

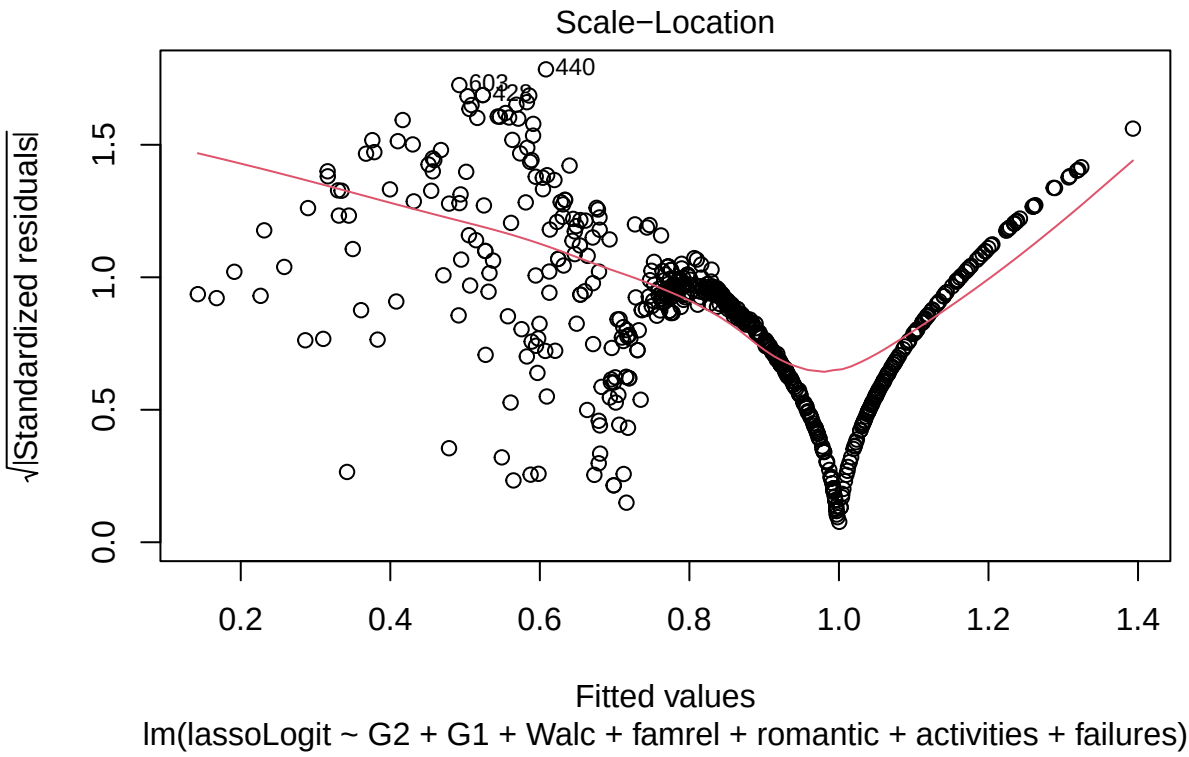

Scale−Location
257
642
81
√|Std. Pearson resid.|
0.0
0.5
1.0
1.5
2.0
−5
0
5
10
15
Predicted values
glm(G3 ~ G1 + Walc + famrel + romantic + activities + failures)

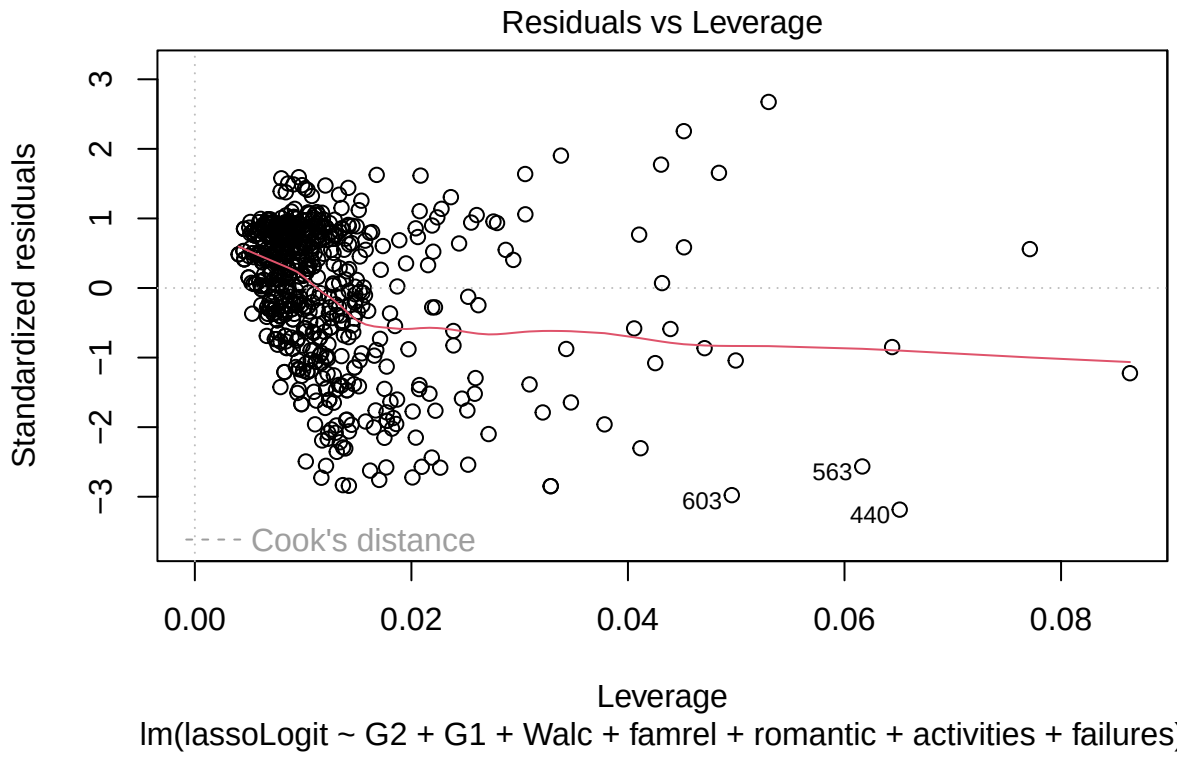



```
lassoLogit <- exp(coef(lassofit)[1] + G1*coef(lassofit)[2] + Walc*coef(lassofit)[3] + famrel*coef(lasso

forwardLogit <- exp(coef(forwardfit)[1] + famrel*coef(forwardfit)[2] + absences*coef(forwardfit)[3] + G

#plotting inverse logit
plot(lm(lassoLogit~ G2 + G1 + Walc + famrel + romantic + activities + failures))
```

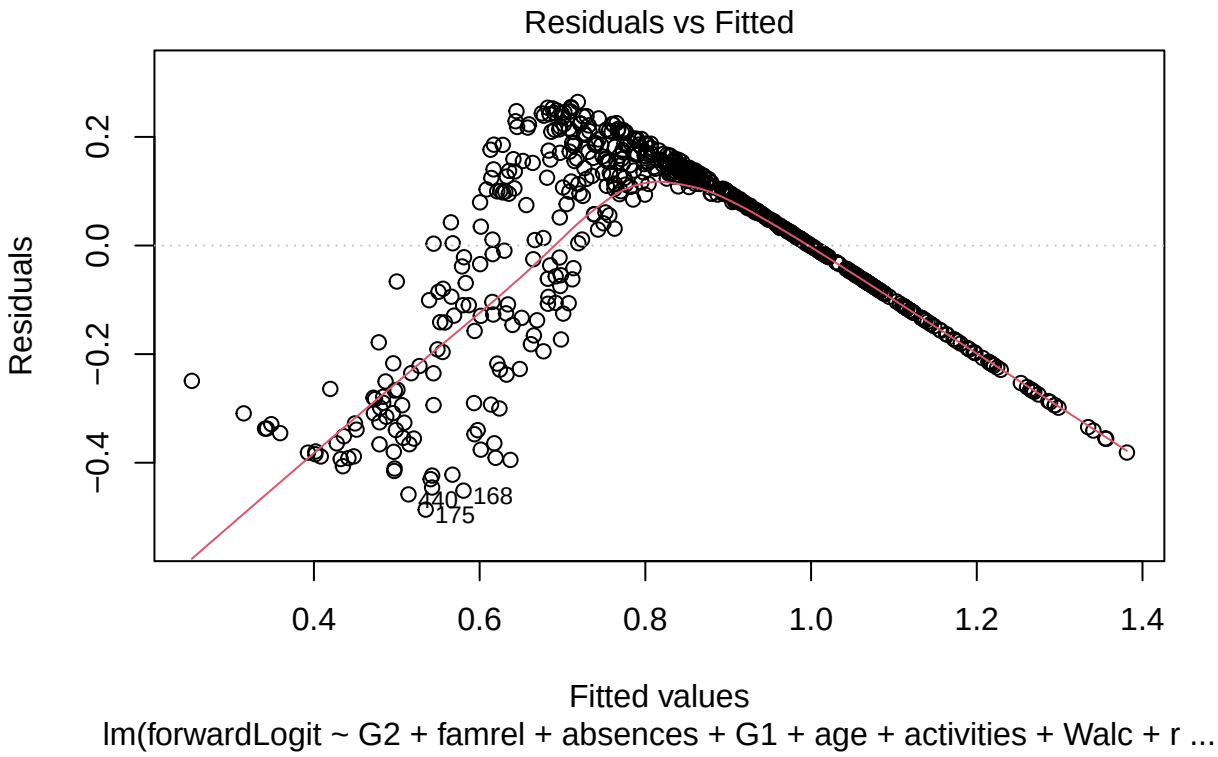

Residuals vs Fitted
Residuals
0.4
0.2
0.0
−0.2
−0.6
603
514
440
0.2
0.4
0.6
0.8
1.0
1.2
1.4
Fitted values
lm(lassoLogit ~ G2 + G1 + Walc + famrel + romantic + activities + failures)

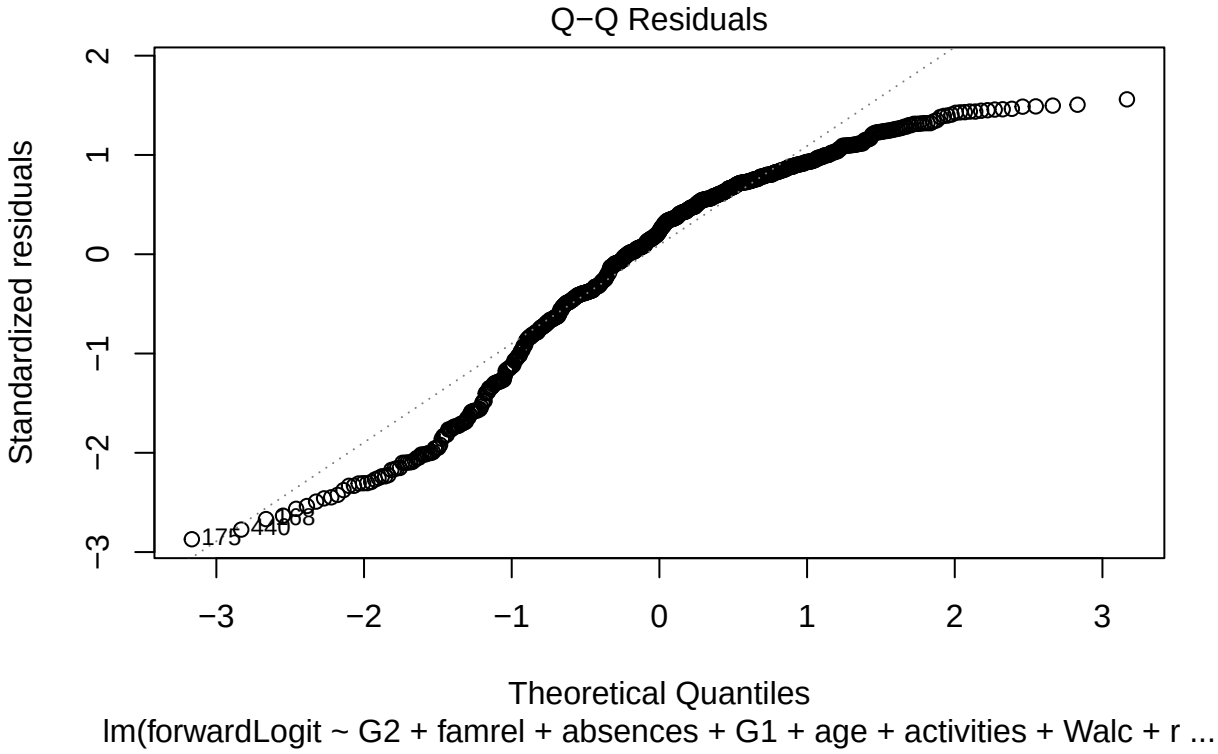

Q−Q Residuals
Standardized residuals
Theoretical Quantiles
lm(lassoLogit ~ G2 + G1 + Walc + famrel + romantic + activities + failures)
440
603

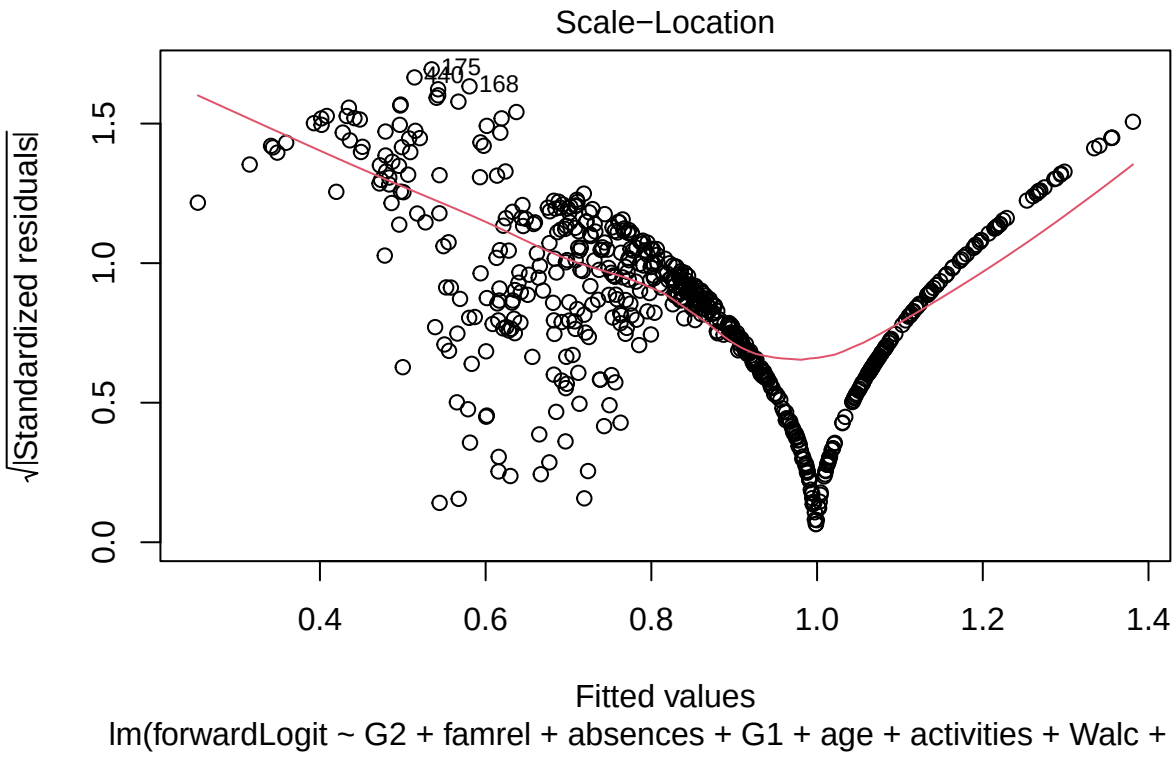

Scale−Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
0.2
0.4
0.6
0.8
1.0
1.2
1.4
603
428
440
Fitted values
lm(lassoLogit ~ G2 + G1 + Walc + famrel + romantic + activities + failures)

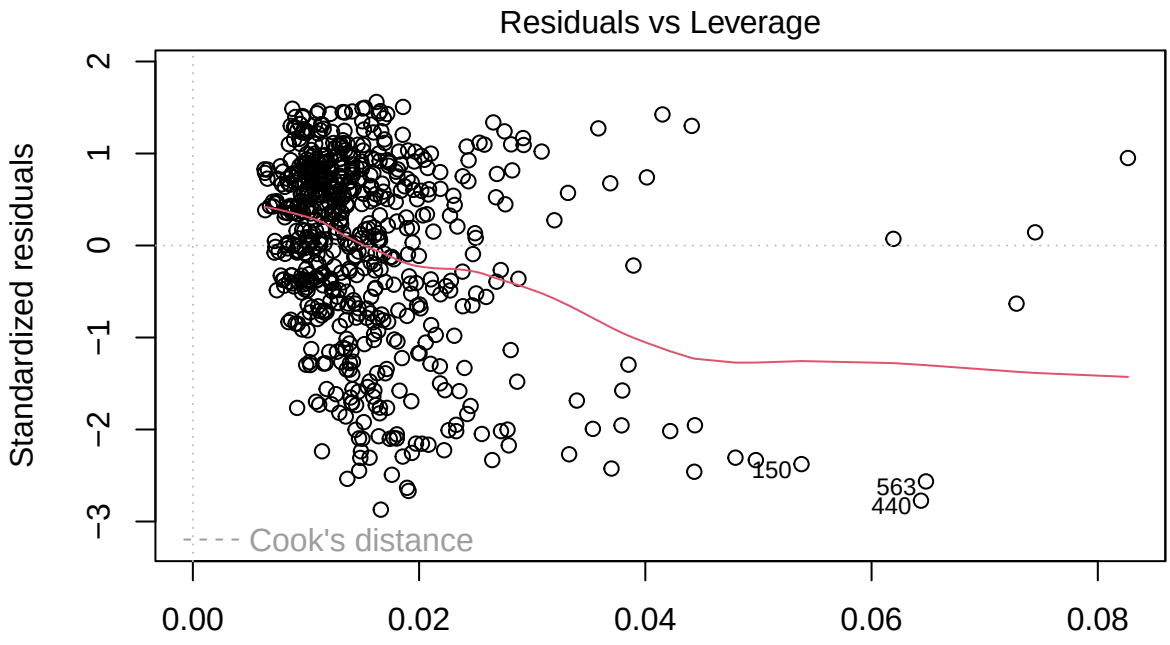



```
plot(lm (forwardLogit ~ G2 + famrel + absences + G1 + age + activities + Walc + romantic + school))
```


Residuals vs Fitted
Residuals
0.2
0.0
−0.2
−0.4
0.4
0.6
0.8
1.0
1.2
1.4
440
168
175
Fitted values
lm(forwardLogit ~ G2 + famrel + absences + G1 + age + activities + Walc + r ...


Q−Q Residuals
Standardized residuals
Theoretical Quantiles
lm(forwardLogit ~ G2 + famrel + absences + G1 + age + activities + Walc + r ...
175
440
668
−3
−2
−1
0
1
2
3


Scale−Location
175
240
168
√|Standardized residuals|
0.0
0.5
1.0
1.5
0.4
0.6
0.8
1.0
1.2
1.4
Fitted values
lm(forwardLogit ~ G2 + famrel + absences + G1 + age + activities + Walc + r ...




```
#performing log transformations as the plots for the forwardfit and lassofit violates linear assumption

# mydata$G3_log <- log(mydata$G3)
# fit1 <- glm(G3_log ~  Walc + famrel + romantic + activities + failures, data = mydata)
# fit2 <- glm (G3_log ~ famrel + absences + G1 + age + activities + Walc + romantic + school, data = my
#
#
# fit_log <- glm(fit2)
#
# summary(fit_log)
#
# plot(fit_log)


vif(forwardfit)
```

```
##     famrel   absences         G1        age activities       Walc   romantic
##   1.063805   1.185406   1.255852   1.282080   1.031272   1.056713   1.112700
##     school
##   1.210438
```