Ridge and LASSO

Lab 9 Handout

Statistics 139

Topics

- Ridge and LASSO
- Tuning λ
- Comparing Prediction Models

Question 1: Data wrangling/cleaning and Initial Model

This lecture handout uses data from the Prevention of REnal and Vascular END-stage Disease (PREVEND) study, which was featured in previous labs and lectures. The documentation for the variables is stored in the `oibiostat` package and can be accessed through Git.

a) Load the data set and convert the categorical variables (with more than 2 classes) into factors. The variable `Ethnicity` is the self-reported ethnicity of the respondent: (0) Western European, (1) African, (2) Asian, and (3) Other. `Gender` is 0 for males and 1 for females.

```
prevend=read.csv("data/prevend.csv")
summary(prevend)
```

```
##       X                Age             Gender          Ethnicity
##  Min.   :   1    Min.   :35.00    Min.   :0.0000    Min.   :0.00000
##  1st Qu.:1024    1st Qu.:45.00    1st Qu.:0.0000    1st Qu.:0.00000
##  Median :2048    Median :54.00    Median :0.0000    Median :0.00000
##  Mean   :2048    Mean   :54.65    Mean   :0.4774    Mean   :0.06789
##  3rd Qu.:3072    3rd Qu.:63.00    3rd Qu.:1.0000    3rd Qu.:0.00000
##  Max.   :4095    Max.   :82.00    Max.   :1.0000    Max.   :3.00000
##    Education           RFFT              VAT              CVD
##  Length:4095      Min.   : 11.00    Min.   :-1.000    Length:4095
##  Class :character  1st Qu.: 48.00    1st Qu.: 9.000    Class :character
##  Mode  :character  Median : 67.00    Median :10.000    Mode  :character
##                    Mean   : 68.62    Mean   : 9.364
##                    3rd Qu.: 87.00    3rd Qu.:11.000
##                    Max.   :157.00    Max.   :12.000
##       DM              Smoking         Hypertension          BMI
##  Length:4095      Min.   :-1.000    Min.   :0.0000    Min.   :16.75
##  Class :character  1st Qu.: 0.000    1st Qu.:0.0000    1st Qu.:23.74
##  Mode  :character  Median : 0.000    Median :0.0000    Median :26.12
##                    Mean   : 0.232    Mean   :0.3392    Mean   :26.80
```

1

```
##                       3rd Qu.: 0.000    3rd Qu.:1.0000    3rd Qu.:29.20
##                       Max.    : 1.000    Max.    :1.0000    Max.     :60.95
##       SBP                 DBP               MAP                 eGFR
##   Min.    : 77.5   Min.    : 43.50   Min.    : 58.00   Min.     : 20.85
##   1st Qu.:112.5   1st Qu.: 67.00   1st Qu.: 84.00   1st Qu.: 69.52
##   Median :123.0   Median : 73.00   Median : 91.00   Median : 77.91
##   Mean    :125.9   Mean    : 73.09   Mean    : 92.14   Mean     : 78.57
##   3rd Qu.:136.5   3rd Qu.: 78.50   3rd Qu.: 99.00   3rd Qu.: 87.71
##   Max.    :218.5   Max.    :108.00   Max.    :151.50   Max.     :149.52
##       Chol                HDL              Statin               Days
##   Min.    : 1.580   Min.    :-1.000   Length:4095        Min.     :  -1.0
##   1st Qu.: 4.640   1st Qu.: 1.140   Class :character   1st Qu.:  -1.0
##   Median : 5.310   Median : 1.370   Mode  :character   Median :  -1.0
##   Mean    : 5.363   Mean    : 1.402                      Mean     : 278.5
##   3rd Qu.: 6.050   3rd Qu.: 1.630                      3rd Qu.:  -1.0
##   Max.    :10.020   Max.    : 3.440                      Max.     :6223.0
##       FRS
##   Min.    :-3.00
##   1st Qu.: 5.00
##   Median :10.00
##   Mean    : 9.99
##   3rd Qu.:15.00
##   Max.    :31.00
```

```r
prevend$Ethnicity = as.factor(prevend$Ethnicity)
summary(prevend$Ethnicity)
```

```
##     0     1     2     3
## 3958    30    73    34
```

What is the best predictive model of RFFT score based on the data in `prevend`? To do this, we'll have to do a little more data processing first.

Run the code below to count the number of missing values in each variable (denoted by the value `-1` in this data set).

```r
apply(prevend==(-1),2,sum)
#apply(prevend=="-1",2,sum)
```

b) Drop any variables that have more missing values than observed values (use `data$variable = NULL`). Convert any categorical variables with several levels to factors (if you have not done that already), remove any remaining rows with missing observations, and finally split into train-test with 2000 observations in the training set (use `set.seed(139)`). Use the training set for the rest of the analyses in this handout.

```r
prevend$Days = NULL
prevend$X = NULL

prevend = prevend[apply(prevend==(-1),1,sum)==0,]
```

```
set.seed(139)
ids = sample(nrow(prevend),2000)
prevend.train = prevend[ids,]
prevend.test = prevend[-ids,]
```

c) Explore the data graphically and assess whether the response variable or any predictor variable(s) need to be transformed.

```
quants = prevend.train[,c(1,5:6,11:17,19)]
prevend.scaled = apply(quants,2,scale)
boxplot(prevend.scaled,cex.axis=0.5,las=2)
```
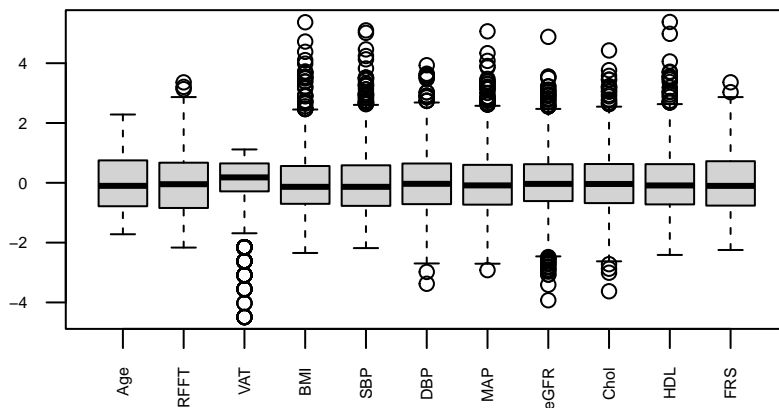

d) Fit an overly-expressed regression model (call it **complexModel**) that contains all main effects and 2-way interaction terms between all predictors. How many predictors in total are in this model?

```
complexModel = lm(RFFT~.^2,data=prevend.train)
length(complexModel$coef)
```

```
## [1] 248
```

```
table(is.na(complexModel$coef))
```

```
##
## FALSE  TRUE
##   242     6
```

```
which(is.na(complexModel$coef))
```

```
## Ethnicity1:EducationUniv        Ethnicity1:CVDYes        Ethnicity3:CVDYes
##                      71                       77                       79
##        Ethnicity1:DMYes         Ethnicity3:DMYes          Ethnicity1:FRS
##                      80                       82                      113
```

```
summary(complexModel)
```

```
##
## Call:
## lm(formula = RFFT ~ .^2, data = prevend.train)
```

```
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -60.171 -13.237  -0.811  12.574  71.626
##
## Coefficients: (6 not defined because of singularities)
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                2.720e+02  1.309e+02   2.079   0.0378 *
## Age                       -3.205e+00  1.704e+00  -1.881   0.0602 .
## Gender                     1.808e+01  2.690e+01   0.672   0.5015
## Ethnicity1                 4.744e+04  7.333e+04   0.647   0.5177
## Ethnicity2                -1.866e+02  1.392e+02  -1.340   0.1804
## Ethnicity3                -1.589e+02  2.677e+02  -0.594   0.5529
## EducationLowerSecond       9.692e+00  2.830e+01   0.342   0.7321
## EducationPrimary          -2.034e+01  4.275e+01  -0.476   0.6343
## EducationUniv              5.956e+01  2.906e+01   2.049   0.0406 *
## VAT                       -8.043e+00  5.409e+00  -1.487   0.1372
## CVDYes                    -8.451e+01  4.335e+01  -1.949   0.0514 .
## DMYes                     -2.930e+01  5.292e+01  -0.554   0.5800
## Smoking                   -6.387e+01  3.139e+01  -2.035   0.0420 *
## Hypertension               1.776e+01  3.094e+01   0.574   0.5659
## BMI                        1.972e-01  2.638e+00   0.075   0.9404
## SBP                       -1.974e+00  1.586e+00  -1.245   0.2133
## DBP                       -3.434e+00  2.898e+00  -1.185   0.2361
## MAP                        3.327e+00  3.681e+00   0.904   0.3663
## eGFR                       2.247e-01  7.329e-01   0.307   0.7592
## Chol                      -1.604e+01  1.153e+01  -1.391   0.1644
## HDL                        2.312e+01  3.596e+01   0.643   0.5204
## StatinUser                 1.872e+01  2.942e+01   0.636   0.5247
## FRS                        7.075e+00  5.041e+00   1.404   0.1606
## Age:Gender                 1.034e-01  2.873e-01   0.360   0.7190
## Age:Ethnicity1            -1.210e+02  1.887e+02  -0.642   0.5212
## Age:Ethnicity2             1.382e+00  1.299e+00   1.064   0.2876
## Age:Ethnicity3             2.347e+00  3.606e+00   0.651   0.5152
## Age:EducationLowerSecond   2.868e-02  3.010e-01   0.095   0.9241
## Age:EducationPrimary       3.343e-01  4.427e-01   0.755   0.4502
## Age:EducationUniv         -4.301e-01  3.066e-01  -1.403   0.1609
## Age:VAT                    4.106e-02  5.597e-02   0.734   0.4633
## Age:CVDYes                 4.689e-01  5.588e-01   0.839   0.4015
## Age:DMYes                  3.370e-01  5.512e-01   0.611   0.5410
## Age:Smoking                1.319e-01  2.870e-01   0.460   0.6459
## Age:Hypertension          -4.147e-01  3.052e-01  -1.359   0.1745
## Age:BMI                   -2.618e-03  2.703e-02  -0.097   0.9228
## Age:SBP                    1.195e-02  1.831e-02   0.653   0.5138
## Age:DBP                    5.420e-02  3.378e-02   1.605   0.1087
## Age:MAP                   -3.396e-02  4.408e-02  -0.770   0.4412
## Age:eGFR                  -2.269e-03  8.077e-03  -0.281   0.7788
## Age:Chol                   1.086e-01  1.050e-01   1.034   0.3011
```

4

```
## Age:HDL                           -1.600e-01  3.054e-01  -0.524   0.6004
## Age:StatinUser                     5.769e-02  3.097e-01   0.186   0.8522
## Age:FRS                           -1.666e-02  1.674e-02  -0.995   0.3198
## Gender:Ethnicity1                 -2.549e+03  4.012e+03  -0.635   0.5253
## Gender:Ethnicity2                 -1.163e+01  1.325e+01  -0.878   0.3801
## Gender:Ethnicity3                  9.041e+00  2.933e+01   0.308   0.7579
## Gender:EducationLowerSecond       -2.137e+00  3.706e+00  -0.577   0.5643
## Gender:EducationPrimary           -9.080e+00  5.561e+00  -1.633   0.1027
## Gender:EducationUniv               1.345e+00  3.373e+00   0.399   0.6900
## Gender:VAT                        -8.388e-02  6.426e-01  -0.131   0.8962
## Gender:CVDYes                      9.030e+00  6.320e+00   1.429   0.1533
## Gender:DMYes                       3.538e+00  7.434e+00   0.476   0.6342
## Gender:Smoking                     2.729e+00  4.639e+00   0.588   0.5564
## Gender:Hypertension                5.582e+00  4.187e+00   1.333   0.1827
## Gender:BMI                        -3.420e-01  3.393e-01  -1.008   0.3136
## Gender:SBP                         1.106e-01  2.309e-01   0.479   0.6319
## Gender:DBP                        -1.602e-02  4.004e-01  -0.040   0.9681
## Gender:MAP                        -9.964e-02  5.242e-01  -0.190   0.8493
## Gender:eGFR                       -1.801e-01  1.009e-01  -1.784   0.0746 .
## Gender:Chol                        1.383e+00  1.699e+00   0.814   0.4160
## Gender:HDL                        -5.137e+00  4.792e+00  -1.072   0.2838
## Gender:StatinUser                 -1.003e+00  3.790e+00  -0.265   0.7914
## Gender:FRS                        -7.150e-01  8.365e-01  -0.855   0.3928
## Ethnicity1:EducationLowerSecond   -8.806e+02  1.310e+03  -0.672   0.5016
## Ethnicity2:EducationLowerSecond    1.576e+01  1.352e+01   1.165   0.2441
## Ethnicity3:EducationLowerSecond    3.072e+01  2.341e+01   1.312   0.1896
## Ethnicity1:EducationPrimary       -4.625e+03  7.184e+03  -0.644   0.5198
## Ethnicity2:EducationPrimary        1.621e+01  1.650e+01   0.982   0.3260
## Ethnicity3:EducationPrimary        1.781e+01  5.906e+01   0.302   0.7630
## Ethnicity1:EducationUniv                  NA         NA      NA       NA
## Ethnicity2:EducationUniv           7.008e+00  1.754e+01   0.399   0.6896
## Ethnicity3:EducationUniv           3.617e+01  2.609e+01   1.386   0.1658
## Ethnicity1:VAT                    -3.192e+02  4.791e+02  -0.666   0.5054
## Ethnicity2:VAT                     3.737e+00  2.701e+00   1.383   0.1667
## Ethnicity3:VAT                     1.848e-01  4.503e+00   0.041   0.9673
## Ethnicity1:CVDYes                         NA         NA      NA       NA
## Ethnicity2:CVDYes                  8.161e+00  2.491e+01   0.328   0.7433
## Ethnicity3:CVDYes                         NA         NA      NA       NA
## Ethnicity1:DMYes                          NA         NA      NA       NA
## Ethnicity2:DMYes                   3.109e+01  2.257e+01   1.377   0.1686
## Ethnicity3:DMYes                          NA         NA      NA       NA
## Ethnicity1:Smoking                -1.426e+03  2.233e+03  -0.639   0.5232
## Ethnicity2:Smoking                 3.196e+01  2.112e+01   1.513   0.1304
## Ethnicity3:Smoking                 2.464e+01  3.877e+01   0.635   0.5252
## Ethnicity1:Hypertension           -4.881e+02  7.373e+02  -0.662   0.5080
## Ethnicity2:Hypertension            7.890e+00  1.804e+01   0.437   0.6620
## Ethnicity3:Hypertension            1.740e+01  1.696e+01   1.026   0.3050
## Ethnicity1:BMI                    -2.343e+02  3.621e+02  -0.647   0.5176
```

```
## Ethnicity2:BMI                       2.256e+00  1.944e+00   1.160   0.2461
## Ethnicity3:BMI                      -6.488e-01  2.104e+00  -0.308   0.7578
## Ethnicity1:SBP                       1.875e+02  2.884e+02   0.650   0.5156
## Ethnicity2:SBP                      -3.520e-01  1.183e+00  -0.298   0.7661
## Ethnicity3:SBP                       2.808e+00  1.998e+00   1.405   0.1601
## Ethnicity1:DBP                      -6.206e+02  9.767e+02  -0.635   0.5252
## Ethnicity2:DBP                       6.043e-01  1.835e+00   0.329   0.7419
## Ethnicity3:DBP                      -8.833e-01  2.393e+00  -0.369   0.7121
## Ethnicity1:MAP                       4.168e+01  7.948e+01   0.524   0.6001
## Ethnicity2:MAP                      -2.400e-01  2.378e+00  -0.101   0.9196
## Ethnicity3:MAP                      -1.692e+00  3.415e+00  -0.495   0.6204
## Ethnicity1:eGFR                     -7.277e+01  1.120e+02  -0.650   0.5159
## Ethnicity2:eGFR                      4.529e-01  3.469e-01   1.306   0.1918
## Ethnicity3:eGFR                     -1.103e-01  9.508e-01  -0.116   0.9077
## Ethnicity1:Chol                      3.292e+02  5.362e+02   0.614   0.5393
## Ethnicity2:Chol                      3.938e+00  5.826e+00   0.676   0.4991
## Ethnicity3:Chol                     -5.000e+00  9.034e+00  -0.554   0.5800
## Ethnicity1:HDL                      -3.775e+03  5.864e+03  -0.644   0.5199
## Ethnicity2:HDL                      -5.238e+00  2.388e+01  -0.219   0.8264
## Ethnicity3:HDL                      -2.842e+01  2.956e+01  -0.961   0.3365
## Ethnicity1:StatinUser                2.938e+03  4.583e+03   0.641   0.5216
## Ethnicity2:StatinUser               -1.153e+01  1.242e+01  -0.929   0.3531
## Ethnicity3:StatinUser                8.328e+00  2.724e+01   0.306   0.7599
## Ethnicity1:FRS                              NA         NA      NA       NA
## Ethnicity2:FRS                      -3.739e+00  3.991e+00  -0.937   0.3490
## Ethnicity3:FRS                      -4.804e+00  8.430e+00  -0.570   0.5689
## EducationLowerSecond:VAT             2.591e-01  6.995e-01   0.370   0.7111
## EducationPrimary:VAT                 9.371e-01  9.388e-01   0.998   0.3183
## EducationUniv:VAT                   -2.835e-01  7.105e-01  -0.399   0.6899
## EducationLowerSecond:CVDYes         -1.799e+00  5.698e+00  -0.316   0.7522
## EducationPrimary:CVDYes             -9.820e-01  6.805e+00  -0.144   0.8853
## EducationUniv:CVDYes                 1.784e-01  6.385e+00   0.028   0.9777
## EducationLowerSecond:DMYes          -1.104e+01  7.867e+00  -1.403   0.1608
## EducationPrimary:DMYes              -8.131e-01  9.461e+00  -0.086   0.9315
## EducationUniv:DMYes                 -8.536e+00  8.755e+00  -0.975   0.3297
## EducationLowerSecond:Smoking        -1.299e+00  4.479e+00  -0.290   0.7718
## EducationPrimary:Smoking             3.232e+00  6.545e+00   0.494   0.6215
## EducationUniv:Smoking                2.089e+00  4.501e+00   0.464   0.6427
## EducationLowerSecond:Hypertension  -4.684e-01  4.193e+00  -0.112   0.9111
## EducationPrimary:Hypertension        5.120e+00  5.987e+00   0.855   0.3925
## EducationUniv:Hypertension           4.119e+00  4.392e+00   0.938   0.3485
## EducationLowerSecond:BMI             7.722e-01  3.663e-01   2.108   0.0351 *
## EducationPrimary:BMI                 4.176e-01  4.744e-01   0.880   0.3787
## EducationUniv:BMI                    3.399e-01  3.804e-01   0.893   0.3718
## EducationLowerSecond:SBP            -6.777e-02  2.659e-01  -0.255   0.7988
## EducationPrimary:SBP                -6.087e-02  3.365e-01  -0.181   0.8565
## EducationUniv:SBP                   -3.673e-02  2.832e-01  -0.130   0.8968
## EducationLowerSecond:DBP            -3.813e-01  4.735e-01  -0.805   0.4208
```

```
## EducationPrimary:DBP              -7.388e-01  6.318e-01  -1.169    0.2424
## EducationUniv:DBP                  9.970e-02  4.787e-01   0.208    0.8350
## EducationLowerSecond:MAP           4.154e-02  6.361e-01   0.065    0.9479
## EducationPrimary:MAP               6.572e-01  7.765e-01   0.846    0.3974
## EducationUniv:MAP                 -3.964e-01  6.589e-01  -0.602    0.5475
## EducationLowerSecond:eGFR         -1.634e-01  1.035e-01  -1.578    0.1148
## EducationPrimary:eGFR             -3.385e-01  1.530e-01  -2.213    0.0271 *
## EducationUniv:eGFR                -2.257e-01  1.058e-01  -2.133    0.0331 *
## EducationLowerSecond:Chol         -2.187e-01  1.766e+00  -0.124    0.9014
## EducationPrimary:Chol              1.263e+00  2.502e+00   0.505    0.6138
## EducationUniv:Chol                -2.298e+00  1.748e+00  -1.315    0.1888
## EducationLowerSecond:HDL           1.593e+00  5.316e+00   0.300    0.7645
## EducationPrimary:HDL               2.761e+00  7.258e+00   0.380    0.7037
## EducationUniv:HDL                  1.241e+01  4.994e+00   2.484    0.0131 *
## EducationLowerSecond:StatinUser   -6.214e-01  3.706e+00  -0.168    0.8669
## EducationPrimary:StatinUser        5.774e+00  5.098e+00   1.132    0.2576
## EducationUniv:StatinUser          -8.426e+00  3.966e+00  -2.124    0.0338 *
## EducationLowerSecond:FRS           3.979e-01  8.785e-01   0.453    0.6506
## EducationPrimary:FRS              -1.405e+00  1.327e+00  -1.059    0.2897
## EducationUniv:FRS                  1.274e+00  8.778e-01   1.451    0.1469
## VAT:CVDYes                         7.171e-01  9.493e-01   0.755    0.4501
## VAT:DMYes                          1.377e+00  1.205e+00   1.143    0.2534
## VAT:Smoking                        1.260e+00  8.512e-01   1.480    0.1391
## VAT:Hypertension                   3.954e-01  7.817e-01   0.506    0.6131
## VAT:BMI                           -2.789e-03  7.042e-02  -0.040    0.9684
## VAT:SBP                            2.841e-02  4.563e-02   0.623    0.5336
## VAT:DBP                            5.409e-03  8.743e-02   0.062    0.9507
## VAT:MAP                            3.415e-02  1.117e-01   0.306    0.7599
## VAT:eGFR                          -8.612e-04  1.962e-02  -0.044    0.9650
## VAT:Chol                           2.964e-01  3.226e-01   0.919    0.3584
## VAT:HDL                            2.199e-01  9.478e-01   0.232    0.8165
## VAT:StatinUser                     1.246e-01  6.733e-01   0.185    0.8532
## VAT:FRS                           -2.460e-01  1.660e-01  -1.482    0.1386
## CVDYes:DMYes                      -1.390e+00  8.471e+00  -0.164    0.8697
## CVDYes:Smoking                     4.737e+00  7.987e+00   0.593    0.5532
## CVDYes:Hypertension                3.913e+00  6.544e+00   0.598    0.5499
## CVDYes:BMI                         1.773e-01  5.271e-01   0.336    0.7366
## CVDYes:SBP                         4.305e-01  3.587e-01   1.200    0.2302
## CVDYes:DBP                         1.252e-01  6.035e-01   0.207    0.8357
## CVDYes:MAP                        -5.363e-01  7.935e-01  -0.676    0.4992
## CVDYes:eGFR                        2.013e-01  1.641e-01   1.227    0.2201
## CVDYes:Chol                        2.078e-01  2.818e+00   0.074    0.9412
## CVDYes:HDL                         8.499e+00  1.053e+01   0.807    0.4195
## CVDYes:StatinUser                 -2.155e+00  4.998e+00  -0.431    0.6664
## CVDYes:FRS                        -2.408e-01  1.602e+00  -0.150    0.8806
## DMYes:Smoking                     -2.884e-01  9.059e+00  -0.032    0.9746
## DMYes:Hypertension                -1.412e+01  7.922e+00  -1.783    0.0748 .
## DMYes:BMI                         -4.044e-01  6.245e-01  -0.648    0.5174
```

```
## DMYes:SBP               5.645e-01  4.213e-01   1.340   0.1804
## DMYes:DBP               6.781e-01  7.607e-01   0.891   0.3728
## DMYes:MAP              -1.051e+00  9.676e-01  -1.086   0.2776
## DMYes:eGFR             -2.587e-01  2.005e-01  -1.290   0.1971
## DMYes:Chol             -2.870e+00  2.774e+00  -1.035   0.3010
## DMYes:HDL               1.422e+00  1.022e+01   0.139   0.8894
## DMYes:StatinUser        5.133e+00  7.120e+00   0.721   0.4711
## DMYes:FRS               1.313e+00  1.535e+00   0.855   0.3925
## Smoking:Hypertension   -7.367e+00  5.023e+00  -1.467   0.1426
## Smoking:BMI             4.451e-02  4.264e-01   0.104   0.9169
## Smoking:SBP             2.215e-01  3.039e-01   0.729   0.4662
## Smoking:DBP             5.347e-01  5.341e-01   1.001   0.3169
## Smoking:MAP            -2.615e-01  7.093e-01  -0.369   0.7125
## Smoking:eGFR            2.482e-02  1.232e-01   0.201   0.8404
## Smoking:Chol            3.543e-01  1.859e+00   0.191   0.8489
## Smoking:HDL            -5.173e-01  5.395e+00  -0.096   0.9236
## Smoking:StatinUser      9.299e-02  4.929e+00   0.019   0.9850
## Smoking:FRS            -2.287e-01  8.167e-01  -0.280   0.7795
## Hypertension:BMI        6.191e-01  4.002e-01   1.547   0.1221
## Hypertension:SBP       -1.520e-01  3.031e-01  -0.501   0.6161
## Hypertension:DBP       -1.190e-01  5.355e-01  -0.222   0.8242
## Hypertension:MAP        2.035e-01  7.210e-01   0.282   0.7778
## Hypertension:eGFR      -2.191e-01  1.208e-01  -1.814   0.0699 .
## Hypertension:Chol      -1.350e+00  1.854e+00  -0.728   0.4665
## Hypertension:HDL        4.720e-02  5.652e+00   0.008   0.9933
## Hypertension:StatinUser 8.710e+00  3.993e+00   2.181   0.0293 *
## Hypertension:FRS        1.378e+00  8.983e-01   1.535   0.1251
## BMI:SBP                 4.650e-03  2.375e-02   0.196   0.8448
## BMI:DBP                -2.426e-02  4.316e-02  -0.562   0.5741
## BMI:MAP                 5.722e-03  5.511e-02   0.104   0.9173
## BMI:eGFR                1.532e-02  1.061e-02   1.444   0.1490
## BMI:Chol               -4.243e-02  1.616e-01  -0.263   0.7930
## BMI:HDL                -1.672e-01  4.660e-01  -0.359   0.7198
## BMI:StatinUser         -5.672e-01  3.577e-01  -1.586   0.1129
## BMI:FRS                -4.626e-02  7.841e-02  -0.590   0.5553
## SBP:DBP                 1.222e-02  1.515e-02   0.807   0.4198
## SBP:MAP                -8.938e-03  7.670e-03  -1.165   0.2440
## SBP:eGFR                3.658e-03  7.339e-03   0.498   0.6183
## SBP:Chol                1.771e-01  1.156e-01   1.532   0.1258
## SBP:HDL                -1.394e-01  3.398e-01  -0.410   0.6818
## SBP:StatinUser         -1.747e-01  2.513e-01  -0.695   0.4872
## SBP:FRS                -8.601e-03  4.918e-02  -0.175   0.8612
## DBP:MAP                 4.659e-03  1.360e-02   0.343   0.7320
## DBP:eGFR               -9.978e-03  1.275e-02  -0.783   0.4340
## DBP:Chol                2.647e-01  2.134e-01   1.241   0.2149
## DBP:HDL                -3.830e-01  5.998e-01  -0.638   0.5233
## DBP:StatinUser          3.724e-01  4.614e-01   0.807   0.4196
## DBP:FRS                -1.157e-01  9.847e-02  -1.175   0.2403
```

```
## MAP:eGFR                    1.059e-03  1.712e-02   0.062   0.9507
## MAP:Chol                    -3.247e-01  2.790e-01  -1.164   0.2446
## MAP:HDL                      1.577e-01  8.018e-01   0.197   0.8441
## MAP:StatinUser               9.309e-02  5.988e-01   0.155   0.8765
## MAP:FRS                      4.517e-02  1.227e-01   0.368   0.7129
## eGFR:Chol                   -2.701e-02  4.728e-02  -0.571   0.5679
## eGFR:HDL                     6.728e-02  1.385e-01   0.486   0.6271
## eGFR:StatinUser             -5.911e-03  1.068e-01  -0.055   0.9559
## eGFR:FRS                     7.943e-03  2.379e-02   0.334   0.7385
## Chol:HDL                     2.781e-01  1.601e+00   0.174   0.8621
## Chol:StatinUser             -9.133e-01  1.730e+00  -0.528   0.5976
## Chol:FRS                    -3.935e-02  2.582e-01  -0.152   0.8789
## HDL:StatinUser              -6.492e+00  5.201e+00  -1.248   0.2121
## HDL:FRS                      1.001e+00  8.227e-01   1.217   0.2238
## StatinUser:FRS              -8.399e-01  8.938e-01  -0.940   0.3475
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 20.8 on 1758 degrees of freedom
## Multiple R-squared:  0.4552, Adjusted R-squared:  0.3805
## F-statistic: 6.095 on 241 and 1758 DF,  p-value: < 2.2e-16
```

```
summary(complexModel)$coef
```

```
##                             Estimate    Std. Error      t value
## (Intercept)             2.720372e+02 1.308639e+02   2.078778936
## Age                    -3.205472e+00 1.704247e+00  -1.880873121
## Gender                  1.808460e+01 2.690118e+01   0.672260506
## Ethnicity1              4.744300e+04 7.333079e+04   0.646972462
## Ethnicity2             -1.865793e+02 1.392187e+02  -1.340188325
## Ethnicity3             -1.588543e+02 2.676526e+02  -0.593509424
## EducationLowerSecond    9.692140e+00 2.830485e+01   0.342419759
## EducationPrimary       -2.033842e+01 4.275224e+01  -0.475727609
## EducationUniv           5.956151e+01 2.906253e+01   2.049426517
## VAT                    -8.042547e+00 5.408618e+00  -1.486987446
## CVDYes                 -8.451294e+01 4.335462e+01  -1.949340832
## DMYes                  -2.929524e+01 5.292297e+01  -0.553544834
## Smoking                -6.387281e+01 3.139349e+01  -2.034587680
## Hypertension            1.776358e+01 3.093717e+01   0.574182534
## BMI                     1.971523e-01 2.638317e+00   0.074726536
## SBP                    -1.974349e+00 1.585706e+00  -1.245091864
## DBP                    -3.434262e+00 2.897902e+00  -1.185085774
## MAP                     3.326616e+00 3.681494e+00   0.903604967
## eGFR                    2.247017e-01 7.329239e-01   0.306582619
## Chol                   -1.604051e+01 1.153231e+01  -1.390918681
## HDL                     2.311659e+01 3.596337e+01   0.642781546
## StatinUser              1.871798e+01 2.941891e+01   0.636256766
## FRS                     7.074804e+00 5.040616e+00   1.403559387
```

9

```
## Age:Gender                         1.033990e-01  2.873407e-01   0.359848159
## Age:Ethnicity1                     -1.210495e+02  1.886771e+02  -0.641569692
## Age:Ethnicity2                      1.382094e+00  1.299384e+00   1.063653297
## Age:Ethnicity3                      2.347109e+00  3.605845e+00   0.650917772
## Age:EducationLowerSecond            2.868488e-02  3.009536e-01   0.095313276
## Age:EducationPrimary                3.343405e-01  4.427175e-01   0.755200451
## Age:EducationUniv                  -4.300514e-01  3.065756e-01  -1.402758060
## Age:VAT                             4.105510e-02  5.596711e-02   0.733557685
## Age:CVDYes                          4.689378e-01  5.588308e-01   0.839140832
## Age:DMYes                           3.370157e-01  5.512473e-01   0.611369378
## Age:Smoking                         1.318882e-01  2.870204e-01   0.459507956
## Age:Hypertension                   -4.146626e-01  3.052286e-01  -1.358531362
## Age:BMI                            -2.618183e-03  2.703091e-02  -0.096858852
## Age:SBP                             1.195408e-02  1.830602e-02   0.653013687
## Age:DBP                             5.420015e-02  3.377549e-02   1.604718361
## Age:MAP                            -3.395715e-02  4.408060e-02  -0.770342311
## Age:eGFR                           -2.268805e-03  8.077380e-03  -0.280883741
## Age:Chol                            1.085520e-01  1.049510e-01   1.034310896
## Age:HDL                            -1.599860e-01  3.053570e-01  -0.523931011
## Age:StatinUser                      5.768524e-02  3.096622e-01   0.186284409
## Age:FRS                            -1.665827e-02  1.674112e-02  -0.995050753
## Gender:Ethnicity1                  -2.549341e+03  4.012275e+03  -0.635385432
## Gender:Ethnicity2                  -1.163152e+01  1.324981e+01  -0.877863735
## Gender:Ethnicity3                   9.041398e+00  2.933392e+01   0.308223317
## Gender:EducationLowerSecond        -2.136748e+00  3.706283e+00  -0.576520420
## Gender:EducationPrimary            -9.079867e+00  5.560602e+00  -1.632892881
## Gender:EducationUniv                1.345465e+00  3.372872e+00   0.398907785
## Gender:VAT                         -8.387690e-02  6.426367e-01  -0.130519924
## Gender:CVDYes                       9.029943e+00  6.320174e+00   1.428749106
## Gender:DMYes                        3.538045e+00  7.434494e+00   0.475895895
## Gender:Smoking                      2.729352e+00  4.638966e+00   0.588353407
## Gender:Hypertension                 5.582271e+00  4.187289e+00   1.333147000
## Gender:BMI                         -3.420405e-01  3.393404e-01  -1.007956668
## Gender:SBP                          1.106340e-01  2.309079e-01   0.479126094
## Gender:DBP                         -1.602419e-02  4.003890e-01  -0.040021556
## Gender:MAP                         -9.963502e-02  5.242418e-01  -0.190055464
## Gender:eGFR                        -1.801105e-01  1.009495e-01  -1.784165116
## Gender:Chol                         1.382659e+00  1.699403e+00   0.813614497
## Gender:HDL                         -5.137143e+00  4.791918e+00  -1.072043072
## Gender:StatinUser                  -1.002811e+00  3.790347e+00  -0.264569808
## Gender:FRS                         -7.149778e-01  8.364921e-01  -0.854733373
## Ethnicity1:EducationLowerSecond    -8.806499e+02  1.310192e+03  -0.672153345
## Ethnicity2:EducationLowerSecond     1.575702e+01  1.352232e+01   1.165260020
## Ethnicity3:EducationLowerSecond     3.072104e+01  2.341224e+01   1.312178663
## Ethnicity1:EducationPrimary        -4.624646e+03  7.183544e+03  -0.643783353
## Ethnicity2:EducationPrimary         1.621046e+01  1.650017e+01   0.982442038
## Ethnicity3:EducationPrimary         1.781356e+01  5.905802e+01   0.301628144
## Ethnicity2:EducationUniv            7.007989e+00  1.754195e+01   0.399498876
```

```
## Ethnicity3:EducationUniv             3.617251e+01 2.609248e+01  1.386319340
## Ethnicity1:VAT                       -3.191566e+02 4.791449e+02 -0.666096078
## Ethnicity2:VAT                        3.736583e+00 2.700866e+00  1.383476054
## Ethnicity3:VAT                        1.847980e-01 4.502523e+00  0.041043216
## Ethnicity2:CVDYes                     8.161266e+00 2.491384e+01  0.327579625
## Ethnicity2:DMYes                      3.109235e+01 2.257423e+01  1.377338138
## Ethnicity1:Smoking                   -1.425511e+03 2.232553e+03 -0.638511276
## Ethnicity2:Smoking                    3.195638e+01 2.111962e+01  1.513113755
## Ethnicity3:Smoking                    2.463541e+01 3.877001e+01  0.635424462
## Ethnicity1:Hypertension              -4.881370e+02 7.373127e+02 -0.662048788
## Ethnicity2:Hypertension               7.889963e+00 1.804404e+01  0.437261458
## Ethnicity3:Hypertension               1.740020e+01 1.695790e+01  1.026082458
## Ethnicity1:BMI                       -2.343054e+02 3.620533e+02 -0.647157306
## Ethnicity2:BMI                        2.255645e+00 1.944100e+00  1.160251585
## Ethnicity3:BMI                       -6.487684e-01 2.103713e+00 -0.308392032
## Ethnicity1:SBP                        1.875128e+02 2.883665e+02  0.650258501
## Ethnicity2:SBP                       -3.520249e-01 1.182915e+00 -0.297590951
## Ethnicity3:SBP                        2.807858e+00 1.998076e+00  1.405280573
## Ethnicity1:DBP                       -6.205942e+02 9.766706e+02 -0.635418115
## Ethnicity2:DBP                        6.042689e-01 1.834779e+00  0.329341613
## Ethnicity3:DBP                       -8.832747e-01 2.393253e+00 -0.369068603
## Ethnicity1:MAP                        4.167603e+01 7.948252e+01  0.524342054
## Ethnicity2:MAP                       -2.400492e-01 2.377678e+00 -0.100959468
## Ethnicity3:MAP                       -1.691562e+00 3.414574e+00 -0.495394760
## Ethnicity1:eGFR                      -7.276683e+01 1.119708e+02 -0.649873291
## Ethnicity2:eGFR                       4.528849e-01 3.468574e-01  1.305680397
## Ethnicity3:eGFR                      -1.102656e-01 9.507508e-01 -0.115977388
## Ethnicity1:Chol                       3.292252e+02 5.362196e+02  0.613974572
## Ethnicity2:Chol                       3.938439e+00 5.825789e+00  0.676035266
## Ethnicity3:Chol                      -5.000474e+00 9.034042e+00 -0.553514537
## Ethnicity1:HDL                       -3.774516e+03 5.863789e+03 -0.643699214
## Ethnicity2:HDL                       -5.238210e+00 2.388075e+01 -0.219348654
## Ethnicity3:HDL                       -2.842191e+01 2.956274e+01 -0.961409804
## Ethnicity1:StatinUser                 2.937818e+03 4.582701e+03  0.641066949
## Ethnicity2:StatinUser                -1.153305e+01 1.241632e+01 -0.928862556
## Ethnicity3:StatinUser                 8.328196e+00 2.724260e+01  0.305704925
## Ethnicity2:FRS                       -3.739105e+00 3.991298e+00 -0.936814106
## Ethnicity3:FRS                       -4.803536e+00 8.429941e+00 -0.569818443
## EducationLowerSecond:VAT              2.591377e-01 6.994795e-01  0.370472148
## EducationPrimary:VAT                  9.370504e-01 9.387876e-01  0.998149463
## EducationUniv:VAT                    -2.835126e-01 7.105101e-01 -0.399026779
## EducationLowerSecond:CVDYes          -1.799017e+00 5.697897e+00 -0.315733477
## EducationPrimary:CVDYes              -9.820030e-01 6.805160e+00 -0.144302715
## EducationUniv:CVDYes                  1.783723e-01 6.384950e+00  0.027936371
## EducationLowerSecond:DMYes           -1.103668e+01 7.867442e+00 -1.402829783
## EducationPrimary:DMYes               -8.131117e-01 9.460603e+00 -0.085947134
## EducationUniv:DMYes                  -8.535860e+00 8.755069e+00 -0.974962026
## EducationLowerSecond:Smoking         -1.299044e+00 4.478923e+00 -0.290034999
```

```
## EducationPrimary:Smoking        3.231934e+00 6.545467e+00  0.493766717
## EducationUniv:Smoking           2.088726e+00 4.500921e+00  0.464066393
## EducationLowerSecond:Hypertension -4.684247e-01 4.192875e+00 -0.111719222
## EducationPrimary:Hypertension   5.120413e+00 5.986651e+00  0.855305079
## EducationUniv:Hypertension      4.118682e+00 4.392089e+00  0.937749993
## EducationLowerSecond:BMI        7.721972e-01 3.662592e-01  2.108335516
## EducationPrimary:BMI            4.176480e-01 4.743563e-01  0.880451966
## EducationUniv:BMI               3.398570e-01 3.803969e-01  0.893427172
## EducationLowerSecond:SBP       -6.777053e-02 2.658515e-01 -0.254918744
## EducationPrimary:SBP           -6.087362e-02 3.365353e-01 -0.180883318
## EducationUniv:SBP              -3.672890e-02 2.832085e-01 -0.129688542
## EducationLowerSecond:DBP       -3.812738e-01 4.735101e-01 -0.805207380
## EducationPrimary:DBP           -7.387669e-01 6.318019e-01 -1.169301466
## EducationUniv:DBP               9.969854e-02 4.786547e-01  0.208289060
## EducationLowerSecond:MAP        4.154393e-02 6.361500e-01  0.065305245
## EducationPrimary:MAP            6.571952e-01 7.764670e-01  0.846391624
## EducationUniv:MAP              -3.963973e-01 6.589044e-01 -0.601600657
## EducationLowerSecond:eGFR      -1.633519e-01 1.035416e-01 -1.577645159
## EducationPrimary:eGFR          -3.384724e-01 1.529756e-01 -2.212590402
## EducationUniv:eGFR             -2.257475e-01 1.058294e-01 -2.133125617
## EducationLowerSecond:Chol      -2.186946e-01 1.765646e+00 -0.123860937
## EducationPrimary:Chol           1.263009e+00 2.502212e+00  0.504756764
## EducationUniv:Chol             -2.298418e+00 1.748447e+00 -1.314548105
## EducationLowerSecond:HDL        1.592504e+00 5.315825e+00  0.299577950
## EducationPrimary:HDL            2.760666e+00 7.258300e+00  0.380346136
## EducationUniv:HDL               1.240729e+01 4.994312e+00  2.484283677
## EducationLowerSecond:StatinUser -6.213706e-01 3.706243e+00 -0.167655092
## EducationPrimary:StatinUser     5.774008e+00 5.098468e+00  1.132498607
## EducationUniv:StatinUser       -8.425704e+00 3.966149e+00 -2.124404160
## EducationLowerSecond:FRS        3.979203e-01 8.784549e-01  0.452977521
## EducationPrimary:FRS           -1.405144e+00 1.326815e+00 -1.059035428
## EducationUniv:FRS               1.273911e+00 8.778137e-01  1.451231472
## VAT:CVDYes                      7.170998e-01 9.493000e-01  0.755398499
## VAT:DMYes                       1.377142e+00 1.205247e+00  1.142621978
## VAT:Smoking                     1.259611e+00 8.512187e-01  1.479773931
## VAT:Hypertension                3.953641e-01 7.817249e-01  0.505758675
## VAT:BMI                        -2.789124e-03 7.041614e-02 -0.039609159
## VAT:SBP                         2.841036e-02 4.562710e-02  0.622664213
## VAT:DBP                         5.408908e-03 8.742641e-02  0.061868129
## VAT:MAP                         3.415108e-02 1.117139e-01  0.305701341
## VAT:eGFR                       -8.612004e-04 1.962267e-02 -0.043888032
## VAT:Chol                        2.963944e-01 3.226496e-01  0.918626381
## VAT:HDL                         2.199294e-01 9.478312e-01  0.232034333
## VAT:StatinUser                  1.246185e-01 6.732994e-01  0.185086253
## VAT:FRS                        -2.460170e-01 1.660497e-01 -1.481586749
## CVDYes:DMYes                   -1.390140e+00 8.470671e+00 -0.164112192
## CVDYes:Smoking                  4.736504e+00 7.986857e+00  0.593037307
## CVDYes:Hypertension             3.913326e+00 6.544155e+00  0.597988036
```

```
## CVDYes:BMI                     1.773477e-01 5.271339e-01  0.336437707
## CVDYes:SBP                     4.305497e-01 3.586928e-01  1.200330161
## CVDYes:DBP                     1.251730e-01 6.035188e-01  0.207405248
## CVDYes:MAP                    -5.363251e-01 7.934693e-01 -0.675924222
## CVDYes:eGFR                    2.012738e-01 1.640932e-01  1.226581908
## CVDYes:Chol                    2.078064e-01 2.817657e+00  0.073751498
## CVDYes:HDL                     8.498952e+00 1.052643e+01  0.807392076
## CVDYes:StatinUser             -2.155453e+00 4.998409e+00 -0.431227852
## CVDYes:FRS                    -2.407808e-01 1.602073e+00 -0.150293301
## DMYes:Smoking                 -2.884223e-01 9.058662e+00 -0.031839399
## DMYes:Hypertension           -1.412459e+01 7.921644e+00 -1.783037966
## DMYes:BMI                     -4.043731e-01 6.244859e-01 -0.647529604
## DMYes:SBP                      5.645276e-01 4.212669e-01  1.340071033
## DMYes:DBP                      6.781367e-01 7.607340e-01  0.891424128
## DMYes:MAP                     -1.050833e+00 9.676407e-01 -1.085974424
## DMYes:eGFR                    -2.586532e-01 2.004542e-01 -1.290335830
## DMYes:Chol                    -2.869825e+00 2.773899e+00 -1.034581709
## DMYes:HDL                      1.422307e+00 1.022441e+01  0.139109028
## DMYes:StatinUser               5.132526e+00 7.120006e+00  0.720859786
## DMYes:FRS                      1.313069e+00 1.535085e+00  0.855372361
## Smoking:Hypertension          -7.367108e+00 5.022771e+00 -1.466741670
## Smoking:BMI                    4.450879e-02 4.263670e-01  0.104390779
## Smoking:SBP                    2.214943e-01 3.038938e-01  0.728854370
## Smoking:DBP                    5.346905e-01 5.341204e-01  1.001067277
## Smoking:MAP                   -2.614716e-01 7.093156e-01 -0.368625158
## Smoking:eGFR                   2.482337e-02 1.232288e-01  0.201441270
## Smoking:Chol                   3.543017e-01 1.859122e+00  0.190574769
## Smoking:HDL                   -5.173444e-01 5.394801e+00 -0.095896832
## Smoking:StatinUser             9.299143e-02 4.929177e+00  0.018865507
## Smoking:FRS                   -2.286754e-01 8.167263e-01 -0.279990255
## Hypertension:BMI               6.190770e-01 4.002473e-01  1.546736269
## Hypertension:SBP              -1.519656e-01 3.030800e-01 -0.501404320
## Hypertension:DBP              -1.189542e-01 5.355424e-01 -0.222119057
## Hypertension:MAP               2.034854e-01 7.209618e-01  0.282241548
## Hypertension:eGFR            -2.190576e-01 1.207880e-01 -1.813570455
## Hypertension:Chol            -1.349868e+00 1.853545e+00 -0.728262555
## Hypertension:HDL              4.719640e-02 5.652124e+00  0.008350207
## Hypertension:StatinUser        8.709709e+00 3.992987e+00  2.181251408
## Hypertension:FRS               1.378483e+00 8.982879e-01  1.534566851
## BMI:SBP                        4.649984e-03 2.374813e-02  0.195804165
## BMI:DBP                       -2.425919e-02 4.316116e-02 -0.562060605
## BMI:MAP                        5.722126e-03 5.511013e-02  0.103830754
## BMI:eGFR                       1.531921e-02 1.061103e-02  1.443705865
## BMI:Chol                      -4.243285e-02 1.616431e-01 -0.262509554
## BMI:HDL                       -1.672067e-01 4.660276e-01 -0.358791406
## BMI:StatinUser                -5.672212e-01 3.576686e-01 -1.585884954
## BMI:FRS                       -4.625516e-02 7.841176e-02 -0.589900852
## SBP:DBP                        1.222398e-02 1.514688e-02  0.807029199
```

```
## SBP:MAP                  -8.938427e-03 7.669719e-03 -1.165417769
## SBP:eGFR                  3.657722e-03 7.338953e-03  0.498398391
## SBP:Chol                  1.771141e-01 1.156305e-01  1.531724712
## SBP:HDL                  -1.393505e-01 3.397911e-01 -0.410106463
## SBP:StatinUser           -1.746585e-01 2.513319e-01 -0.694931616
## SBP:FRS                  -8.600512e-03 4.917767e-02 -0.174886542
## DBP:MAP                   4.659323e-03 1.360322e-02  0.342516322
## DBP:eGFR                 -9.978448e-03 1.275081e-02 -0.782573556
## DBP:Chol                  2.647336e-01 2.134027e-01  1.240535245
## DBP:HDL                  -3.829780e-01 5.998418e-01 -0.638465100
## DBP:StatinUser            3.724398e-01 4.613567e-01  0.807270881
## DBP:FRS                  -1.156631e-01 9.847262e-02 -1.174571145
## MAP:eGFR                  1.059158e-03 1.711533e-02  0.061883594
## MAP:Chol                 -3.247238e-01 2.789641e-01 -1.164034460
## MAP:HDL                   1.577025e-01 8.017957e-01  0.196686623
## MAP:StatinUser            9.308591e-02 5.988348e-01  0.155445060
## MAP:FRS                   4.516647e-02 1.227162e-01  0.368056332
## eGFR:Chol                -2.701137e-02 4.728199e-02 -0.571282439
## eGFR:HDL                  6.728119e-02 1.384693e-01  0.485892635
## eGFR:StatinUser          -5.911040e-03 1.068447e-01 -0.055323663
## eGFR:FRS                  7.943281e-03 2.379125e-02  0.333874066
## Chol:HDL                  2.780818e-01 1.600537e+00  0.173742812
## Chol:StatinUser          -9.133423e-01 1.730133e+00 -0.527902858
## Chol:FRS                 -3.934882e-02 2.582370e-01 -0.152374841
## HDL:StatinUser           -6.492150e+00 5.200683e+00 -1.248326363
## HDL:FRS                   1.001086e+00 8.227178e-01  1.216803708
## StatinUser:FRS           -8.399464e-01 8.938081e-01 -0.939739042
##                          Pr(>|t|)
## (Intercept)              0.03778238
## Age                      0.06015434
## Gender                   0.50150621
## Ethnicity1               0.51773423
## Ethnicity2               0.18035731
## Ethnicity3               0.55291663
## EducationLowerSecond     0.73207595
## EducationPrimary         0.63432756
## EducationUniv            0.04056854
## VAT                      0.13719745
## CVDYes                   0.05141356
## DMYes                    0.57996078
## Smoking                  0.04204218
## Hypertension             0.56591774
## BMI                      0.94044080
## SBP                      0.21326392
## DBP                      0.23614371
## MAP                      0.36632871
## eGFR                     0.75919739
## Chol                     0.16442610
```

```
## HDL                                    0.52044968
## StatinUser                             0.52469190
## FRS                                     0.16062670
## Age:Gender                              0.71900391
## Age:Ethnicity1                          0.52123625
## Age:Ethnicity2                          0.28763187
## Age:Ethnicity3                          0.51518463
## Age:EducationLowerSecond                0.92407686
## Age:EducationPrimary                    0.45022988
## Age:EducationUniv                       0.16086557
## Age:VAT                                 0.46331617
## Age:CVDYes                              0.40150437
## Age:DMYes                               0.54103412
## Age:Smoking                             0.64592626
## Age:Hypertension                        0.17446943
## Age:BMI                                 0.92284953
## Age:SBP                                 0.51383284
## Age:DBP                                 0.10873541
## Age:MAP                                 0.44120037
## Age:eGFR                                0.77883262
## Age:Chol                                0.30113310
## Age:HDL                                 0.60039257
## Age:StatinUser                          0.85224323
## Age:FRS                                 0.31984852
## Gender:Ethnicity1                       0.52525976
## Gender:Ethnicity2                       0.38013763
## Gender:Ethnicity3                       0.75794893
## Gender:EducationLowerSecond             0.56433728
## Gender:EducationPrimary                 0.10267060
## Gender:EducationUniv                    0.69000960
## Gender:VAT                              0.89617002
## Gender:CVDYes                           0.15325400
## Gender:DMYes                            0.63420768
## Gender:Smoking                          0.55637068
## Gender:Hypertension                     0.18265630
## Gender:BMI                              0.31361398
## Gender:SBP                              0.63190852
## Gender:DBP                              0.96808049
## Gender:MAP                              0.84928761
## Gender:eGFR                             0.07456926
## Gender:Chol                             0.41597601
## Gender:HDL                              0.28384783
## Gender:StatinUser                       0.79137190
## Gender:FRS                              0.39281523
## Ethnicity1:EducationLowerSecond         0.50157440
## Ethnicity2:EducationLowerSecond         0.24407184
## Ethnicity3:EducationLowerSecond         0.18963116
## Ethnicity1:EducationPrimary             0.51979990
```

```
## Ethnicity2:EducationPrimary        0.32601731
## Ethnicity3:EducationPrimary        0.76297123
## Ethnicity2:EducationUniv           0.68957418
## Ethnicity3:EducationUniv           0.16582520
## Ethnicity1:VAT                     0.50543708
## Ethnicity2:VAT                     0.16669459
## Ethnicity3:VAT                     0.96726610
## Ethnicity2:CVDYes                  0.74326853
## Ethnicity2:DMYes                   0.16858306
## Ethnicity1:Smoking                 0.52322409
## Ethnicity2:Smoking                 0.13043051
## Ethnicity3:Smoking                 0.52523432
## Ethnicity1:Hypertension            0.50802673
## Ethnicity2:Hypertension            0.66197545
## Ethnicity3:Hypertension            0.30499388
## Ethnicity1:BMI                     0.51761463
## Ethnicity2:BMI                     0.24610393
## Ethnicity3:BMI                     0.75782058
## Ethnicity1:SBP                     0.51561022
## Ethnicity2:SBP                     0.76605054
## Ethnicity3:SBP                     0.16011455
## Ethnicity1:DBP                     0.52523845
## Ethnicity2:DBP                     0.74193673
## Ethnicity3:DBP                     0.71212103
## Ethnicity1:MAP                     0.60010676
## Ethnicity2:MAP                     0.91959415
## Ethnicity3:MAP                     0.62038341
## Ethnicity1:eGFR                    0.51585898
## Ethnicity2:eGFR                    0.19183210
## Ethnicity3:eGFR                    0.90768371
## Ethnicity1:Chol                    0.53931157
## Ethnicity2:Chol                    0.49910718
## Ethnicity3:Chol                    0.57998152
## Ethnicity1:HDL                     0.51985446
## Ethnicity2:HDL                     0.82640393
## Ethnicity3:HDL                     0.33647837
## Ethnicity1:StatinUser              0.52156275
## Ethnicity2:StatinUser              0.35308784
## Ethnicity3:StatinUser              0.75986552
## Ethnicity2:FRS                     0.34898286
## Ethnicity3:FRS                     0.56887365
## EducationLowerSecond:VAT           0.71107536
## EducationPrimary:VAT               0.31834425
## EducationUniv:VAT                  0.68992194
## EducationLowerSecond:CVDYes        0.75224229
## EducationPrimary:CVDYes            0.88527797
## EducationUniv:CVDYes               0.97771607
## EducationLowerSecond:DMYes         0.16084418
```

```
## EducationPrimary:DMYes              0.93151823
## EducationUniv:DMYes                 0.32971323
## EducationLowerSecond:Smoking        0.77182367
## EducationPrimary:Smoking            0.62153262
## EducationUniv:Smoking               0.64265764
## EducationLowerSecond:Hypertension  0.91105879
## EducationPrimary:Hypertension       0.39249883
## EducationUniv:Hypertension          0.34850172
## EducationLowerSecond:BMI            0.03514315
## EducationPrimary:BMI                0.37873486
## EducationUniv:BMI                   0.37175074
## EducationLowerSecond:SBP            0.79881568
## EducationPrimary:SBP                0.85648001
## EducationUniv:SBP                   0.89682768
## EducationLowerSecond:DBP            0.42080891
## EducationPrimary:DBP                0.24244071
## EducationUniv:DBP                   0.83502740
## EducationLowerSecond:MAP            0.94793839
## EducationPrimary:MAP                0.39744950
## EducationUniv:MAP                   0.54751756
## EducationLowerSecond:eGFR           0.11482706
## EducationPrimary:eGFR               0.02705398
## EducationUniv:eGFR                  0.03305257
## EducationLowerSecond:Chol           0.90143954
## EducationPrimary:Chol               0.61379295
## EducationUniv:Chol                  0.18883329
## EducationLowerSecond:HDL            0.76453452
## EducationPrimary:HDL                0.70373444
## EducationUniv:HDL                   0.01307373
## EducationLowerSecond:StatinUser     0.86687391
## EducationPrimary:StatinUser         0.25757931
## EducationUniv:StatinUser            0.03377558
## EducationLowerSecond:FRS            0.65062083
## EducationPrimary:FRS                0.28972913
## EducationUniv:FRS                   0.14689383
## VAT:CVDYes                          0.45011111
## VAT:DMYes                           0.25335116
## VAT:Smoking                         0.13911276
## VAT:Hypertension                    0.61308948
## VAT:BMI                             0.96840922
## VAT:SBP                             0.53358596
## VAT:DBP                             0.95067488
## VAT:MAP                             0.75986825
## VAT:eGFR                            0.96499864
## VAT:Chol                            0.35841705
## VAT:HDL                             0.81653835
## VAT:StatinUser                      0.85318273
## VAT:FRS                             0.13862950
```

```
## CVDYes:DMYes                      0.86966169
## CVDYes:Smoking                    0.55323247
## CVDYes:Hypertension               0.54992494
## CVDYes:BMI                        0.73658097
## CVDYes:SBP                        0.23017286
## CVDYes:DBP                        0.83571741
## CVDYes:MAP                        0.49917766
## CVDYes:eGFR                       0.22014403
## CVDYes:Chol                       0.94121651
## CVDYes:HDL                        0.41954985
## CVDYes:StatinUser                 0.66635558
## CVDYes:FRS                        0.88055046
## DMYes:Smoking                     0.97460374
## DMYes:Hypertension                0.07475262
## DMYes:BMI                         0.51737379
## DMYes:SBP                         0.18039543
## DMYes:DBP                         0.37282366
## DMYes:MAP                         0.27763919
## DMYes:eGFR                        0.19710377
## DMYes:Chol                        0.30100659
## DMYes:HDL                         0.88937993
## DMYes:StatinUser                  0.47109163
## DMYes:FRS                         0.39246160
## Smoking:Hypertension              0.14262520
## Smoking:BMI                       0.91687115
## Smoking:SBP                       0.46618783
## Smoking:DBP                       0.31693205
## Smoking:MAP                       0.71245153
## Smoking:eGFR                      0.84037685
## Smoking:Chol                      0.84888076
## Smoking:HDL                       0.92361344
## Smoking:StatinUser                0.98495054
## Smoking:FRS                       0.77951792
## Hypertension:BMI                  0.12210684
## Hypertension:SBP                  0.61614938
## Hypertension:DBP                  0.82424700
## Hypertension:MAP                  0.77779152
## Hypertension:eGFR                 0.06991432
## Hypertension:Chol                 0.46654987
## Hypertension:HDL                  0.99333852
## Hypertension:StatinUser           0.02929693
## Hypertension:FRS                  0.12507020
## BMI:SBP                           0.84478608
## BMI:DBP                           0.57414638
## BMI:MAP                           0.91731550
## BMI:eGFR                          0.14899990
## BMI:Chol                          0.79295938
## BMI:HDL                           0.71979423
```

```
## BMI:StatinUser                       0.11294525
## BMI:FRS                               0.55533293
## SBP:DBP                               0.41975883
## SBP:MAP                               0.24400803
## SBP:eGFR                              0.61826562
## SBP:Chol                              0.12577030
## SBP:HDL                               0.68177782
## SBP:StatinUser                        0.48718999
## SBP:FRS                               0.86118892
## DBP:MAP                               0.73200331
## DBP:eGFR                              0.43398297
## DBP:Chol                              0.21494302
## DBP:HDL                               0.52325413
## DBP:StatinUser                        0.41961964
## DBP:FRS                               0.24032541
## MAP:eGFR                              0.95066257
## MAP:Chol                              0.24456800
## MAP:HDL                               0.84409551
## MAP:StatinUser                        0.87648831
## MAP:FRS                               0.71287554
## eGFR:Chol                             0.56788123
## eGFR:HDL                              0.62710385
## eGFR:StatinUser                       0.95588690
## eGFR:FRS                              0.73851441
## Chol:HDL                              0.86208759
## Chol:StatinUser                       0.59763338
## Chol:FRS                              0.87890878
## HDL:StatinUser                        0.21207779
## HDL:FRS                               0.22384224
## StatinUser:FRS                        0.34748053
```

e) Fit both a single ridge regression (**ridge1**) and a single LASSO regression model (**lasso1**) using `glmnet` with $\lambda = 1$ for each based on the **complexModel**. Provide a visual to compare the resulting $\hat{\beta}$ coefficients for the 3 models fit so far. It is OK if you look at just a subset of these predictors.
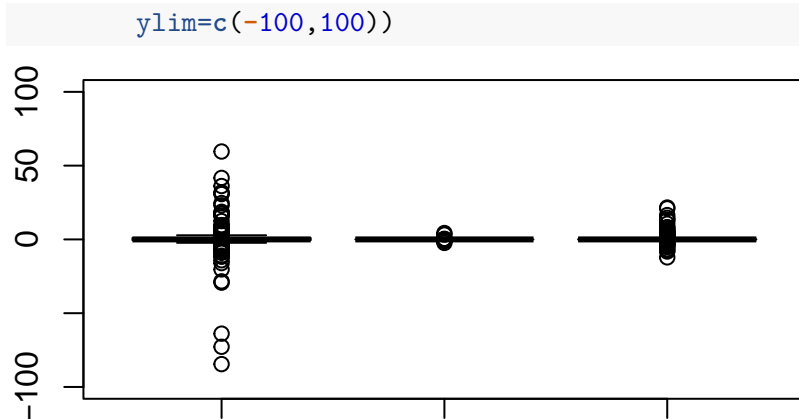
```
library(glmnet)
```

```
## Loading required package: Matrix
```

```
## Loaded glmnet 4.1-8
```

```
X = model.matrix(complexModel)[,-1]
dim(X)
```

```
## [1] 2000  247
```

```
ridge1 = glmnet(X,prevend.train$RFFT,alpha=0,lambda=1)
lasso1 = glmnet(X,prevend.train$RFFT,alpha=1,lambda=1)
```

```
boxplot(complexModel$coef[-1],as.numeric(lasso1$beta), as.numeric(ridge1$beta),
```

```
        ylim=c(-100,100))
```


f) Compare the coefficient estimates for these 3 models in text. Explain why the results of this comparison make sense.

g) Compare the accuracy of each of these models on the test set you created. Which model performed best? Which model performed worst? What did you expect to find?

```
yhat.lm = predict(complexModel,new=prevend.test)
```

```
## Warning in predict.lm(complexModel, new = prevend.test): prediction from
## rank-deficient fit; attr(*, "non-estim") has doubtful cases
```

```
X.test = model.matrix(complexModel,data=prevend.test)[,-1]
yhat.lasso1 = predict(lasso1,newx =X.test)
yhat.ridge1 = predict(ridge1,newx =X.test)

MSE = function(y,yhat){
  return(mean((y-yhat)^2))
}

MSE(prevend.test$RFFT,yhat.lm)
```

```
## [1] 177012.7
```
```
MSE(prevend.test$RFFT,yhat.lasso1)
```

```
## [1] 401.558
```
```
MSE(prevend.test$RFFT,yhat.ridge1)
```

```
## [1] 418.078
```

Question 2: Tuning λ

In this question we will be using k-fold cross-validation via `cv.glmnet` on the training data to identify optimal λ values, then calculate RMSE on the test data.

a) Ridge regression:

20

i. Using a k of 5, perform k-fold cross validation with ridge regression to identify the value of λ that produces the lowest mean RMSE based on the **complexModel**. Label the resulting model **tunedRidge**.

```r
ridges = cv.glmnet(X,prevend.train$RFFT,alpha=0,nfolds=5,
                   lambda=2^((-10:10)))
plot(ridges)
```
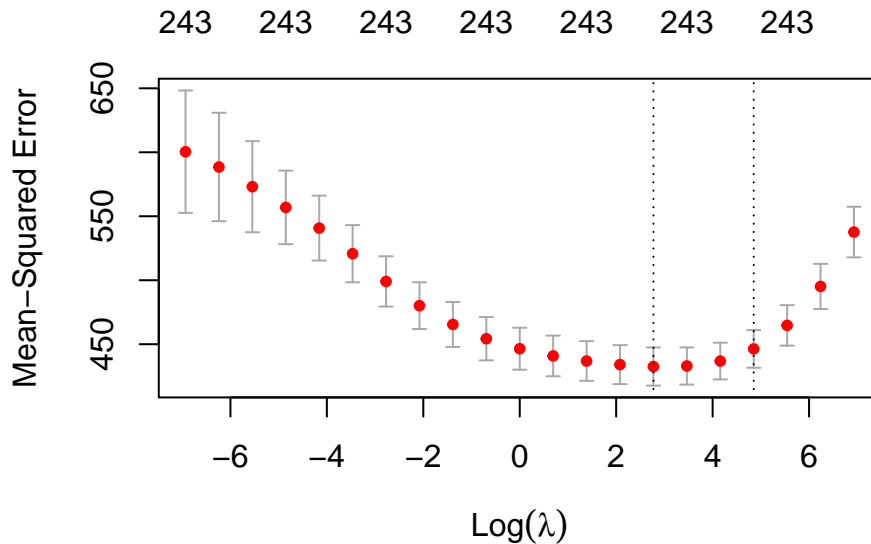

ii. Provide a visual to aid in the choice of optimal λ in the previous part.

iii. From the optimal λ., calculate the RMSE on the test set.

```r
names(ridges)
```

```
##  [1] "lambda"     "cvm"        "cvsd"       "cvup"       "cvlo"
##  [6] "nzero"      "call"       "name"       "glmnet.fit" "lambda.min"
## [11] "lambda.1se" "index"
```

```r
ridges
```

```
##
## Call:  cv.glmnet(x = X, y = prevend.train$RFFT, lambda = 2^((-10:10)),    nfolds = 5, alpl
##
## Measure: Mean-Squared Error
##
##     Lambda Index Measure    SE Nonzero
## min     16     7   432.5 14.91     243
## 1se    128     4   446.3 14.71     243
```

b) LASSO:

i. Using a k of 5, perform k-fold cross validation with LASSO regression to identify the value of λ that produces the lowest mean RMSE based on the **complexModel**. Label the resulting model **tunedLasso**.

ii. Provide a visual to aid in the choice of optimal λ in the previous part.

iii. From the optimal λ from part i., calculate the RMSE on the test set.

c) How do these well-tuned ridge and LASSO perform relative to the full **complexModel**? What do the results suggest about the nature of the variables in the data set for predicting RFFT score (e.g., relative number of strong versus weak predictors)?

d) How could the LASSO and ridge models be improved to predict RFFT score in the data set?

Question 3: Model selection

We will be considering four nested models for predicting RFFT score (`RFFT`). The variables of interest are statin use (`Statin`), age (`Age`), educational attainment (`Education`), presence of diabetes (`DM`), and presence of cardiovascular disease (`CVD`).

- Model A: statin use (`Statin`), age (`Age`)
- Model B: statin use, age, educational attainment (`Education`)
- Model C: statin use, age, education, diabetes (`DM`)
- Model D: statin use, age, education, diabetes, cardiovascular disease (`CVD`)

a) Fit the models and look at the summary output. Calculate AIC and BIC for all four models. Which model is the best according to AIC? For BIC? Are the results what you would expect based on the summary output? Explain.

Sequential variable selection

b) Fit an initial model, **model1**, that has only main effects.

c) Use backward selection from **model1** according to AIC and print the summary of the final model; name it **model2**.

d) Use a combined stepwise procedure according to AIC, starting with **model1** and setting the intercept-only model as the lower-limit model and the full model with all main effects and all 2-way interaction terms from predictors in **model2** as the upper-limit model. Print the summary of the final model and name it **model3**.

e) Compare the BIC and AIC for all three sequentially chosen models.

f) Determine which of the three models performs best on `prevend.test` based on RMSE. How do these results compare to the well-tuned LASSO and ridge models?

Question 4: Cross-validation

a) *Random subsets.* Using an 80-20 train-validation split over 100 iterations, determine which of the four models has the lowest average RMSE (over the validation sets):

i. **complexModel**
ii. **tunedRidge**
iii. **tunedLasso**
iv. the best sequentially chosen model (from **model1**, **model2**, and **model3**)

b) Why might it be reasonable to generally expect that RMSEs from the validation sets are lower than RMSEs from the test sets? Is this the case here?