Stepwise

Zelin Wang

2023-09-08

Load data

```
os_data<-read.csv("OS.csv")
```

Set random seed for subsequent random selection and assignment operations

```
set.seed(2023)
```

Partition data and create index matrix of selected values

```
library(caret)
```

```
## Loading required package: ggplot2
```

```
## Loading required package: lattice
```

```
index <- createDataPartition(os_data$os, p=.8, list=FALSE, times=1)
```

Create test and train data frames

```
train_df <- os_data[index,]
test_df <- os_data[-index,]
```

Verify number of rows (cases) in each data frame

```r
nrow(train_df)
```

```
## [1] 17230
```

```r
nrow(test_df)
```

```
## [1] 4307
```

```r
#define response variable
```

```r
y_nom <- train_df$os
```

```r
#define matrix of predictor variables
```

```r
x_nom <- data.matrix(train_df[, c(2:21)])
data_nom <- data.frame(x_nom,y_nom)
```

Generate full model, exclude all cod related variables, and numeric variables.

```r
full_model<-glm(y_nom~as.factor(gender)+as.factor(race)+as.factor(age)+as.factor(size)+as.factor(marry)
summary(full_model)
```

```
##
## Call:
## glm(formula = y_nom ~ as.factor(gender) + as.factor(race) + as.factor(age) +
##     as.factor(size) + as.factor(marry) + as.factor(income) +
##     as.factor(site) + as.factor(grade) + as.factor(kind) + as.factor(T) +
##     as.factor(N) + as.factor(surgery_pri) + as.factor(RX_Summ) +
##     as.factor(radiate) + as.factor(chem) + as.factor(CEA) + as.factor(bone) +
##     as.factor(brain) + as.factor(lung) + as.factor(group), family = binomial(link = "logit"),
##     data = data_nom)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.4171  -0.5892  -0.3870  -0.2092   2.9038
##
## Coefficients:
##                     Estimate Std. Error z value Pr(>|z|)
## (Intercept)        -1.432314   0.271214  -5.281 1.28e-07 ***
## as.factor(gender)2 -0.036637   0.044459  -0.824 0.409907
## as.factor(race)2   -0.050801   0.060652  -0.838 0.402267
## as.factor(race)3   -0.106288   0.076542  -1.389 0.164949
## as.factor(age)2    -0.081604   0.164054  -0.497 0.618891
## as.factor(age)3     0.328238   0.161644   2.031 0.042293 *
## as.factor(size)2    0.232612   0.055054   4.225 2.39e-05 ***
## as.factor(size)3    0.158097   0.057971   2.727 0.006388 **
```

```
## as.factor(marry)2         -0.173036   0.055402  -3.123 0.001789 **
## as.factor(marry)3          0.004135   0.061776   0.067 0.946634
## as.factor(income)2        -0.075571   0.058795  -1.285 0.198676
## as.factor(income)3        -0.156826   0.067522  -2.323 0.020202 *
## as.factor(site)2           0.351713   0.045831   7.674 1.67e-14 ***
## as.factor(grade)2          0.155902   0.123262   1.265 0.205945
## as.factor(grade)3          0.821228   0.129585   6.337 2.34e-10 ***
## as.factor(grade)4          0.865420   0.173929   4.976 6.50e-07 ***
## as.factor(grade)9          0.615826   0.125303   4.915 8.89e-07 ***
## as.factor(kind)2           0.384829   0.073464   5.238 1.62e-07 ***
## as.factor(T)2             -0.119951   0.078183  -1.534 0.124969
## as.factor(T)3              0.060590   0.072126   0.840 0.400879
## as.factor(N)2              0.032137   0.084618   0.380 0.704105
## as.factor(N)3              0.230162   0.060060   3.832 0.000127 ***
## as.factor(surgery_pri)1   -0.408569   0.128248  -3.186 0.001444 **
## as.factor(RX_Summ)1       -0.195512   0.077579  -2.520 0.011730 *
## as.factor(radiate)1       -0.287744   0.080417  -3.578 0.000346 ***
## as.factor(chem)1          -2.159389   0.045577 -47.379  < 2e-16 ***
## as.factor(CEA)1            0.444658   0.090000   4.941 7.79e-07 ***
## as.factor(CEA)2            0.345084   0.094730   3.643 0.000270 ***
## as.factor(bone)1           0.672075   0.079640   8.439  < 2e-16 ***
## as.factor(bone)2          -0.056800   0.206117  -0.276 0.782878
## as.factor(brain)1          1.178483   0.168305   7.002 2.52e-12 ***
## as.factor(brain)2         -0.206901   0.194482  -1.064 0.287394
## as.factor(lung)1           0.359183   0.048137   7.462 8.54e-14 ***
## as.factor(lung)2           0.112689   0.145929   0.772 0.439983
## as.factor(group)2          0.413219   0.094256   4.384 1.17e-05 ***
## as.factor(group)9          0.500428   0.137884   3.629 0.000284 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 18341  on 17229  degrees of freedom
## Residual deviance: 13921  on 17194  degrees of freedom
## AIC: 13993
##
## Number of Fisher Scoring iterations: 5
```

Perform stepwise regression

create null model

```
null_model<-glm(y_nom~1,data=data_nom,family=binomial(link="logit"))
```

Perform forward selection

AIC threshold (Akaike Information Criterion)

```
model_step1 = step(null_model,
                    scope = list(lower=null_model, upper=full_model),
                    direction = "forward", k= 2, trace=0)

###summary model
summary(model_step1)
```

```
##
## Call:
## glm(formula = y_nom ~ as.factor(chem) + as.factor(group) + as.factor(grade) +
##     as.factor(site) + as.factor(lung) + as.factor(bone) + as.factor(age) +
##     as.factor(brain) + as.factor(kind) + as.factor(CEA) + as.factor(size) +
##     as.factor(radiate) + as.factor(N) + as.factor(marry) + as.factor(surgery_pri) +
##     as.factor(RX_Summ) + as.factor(T) + as.factor(income), family = binomial(link = "logit"),
##     data = data_nom)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.4390  -0.5888  -0.3873  -0.2092   2.9182
##
## Coefficients:
##                      Estimate Std. Error z value Pr(>|z|)
## (Intercept)          -1.47441    0.26913  -5.478 4.29e-08 ***
## as.factor(chem)1     -2.15775    0.04552 -47.398  < 2e-16 ***
## as.factor(group)2     0.41362    0.09423   4.390 1.14e-05 ***
## as.factor(group)9     0.49938    0.13782   3.623 0.000291 ***
## as.factor(grade)2     0.15701    0.12322   1.274 0.202604
## as.factor(grade)3     0.82402    0.12951   6.362 1.99e-10 ***
## as.factor(grade)4     0.87497    0.17377   5.035 4.77e-07 ***
## as.factor(grade)9     0.61753    0.12527   4.930 8.23e-07 ***
## as.factor(site)2      0.35607    0.04534   7.853 4.06e-15 ***
## as.factor(lung)1      0.35814    0.04811   7.444 9.74e-14 ***
## as.factor(lung)2      0.11378    0.14595   0.780 0.435669
## as.factor(bone)1      0.66938    0.07961   8.408  < 2e-16 ***
## as.factor(bone)2     -0.06360    0.20609  -0.309 0.757632
## as.factor(age)2      -0.08706    0.16401  -0.531 0.595555
## as.factor(age)3       0.32581    0.16159   2.016 0.043764 *
## as.factor(brain)1     1.18009    0.16814   7.019 2.24e-12 ***
## as.factor(brain)2    -0.20216    0.19445  -1.040 0.298497
## as.factor(kind)2      0.38942    0.07340   5.306 1.12e-07 ***
## as.factor(CEA)1       0.44360    0.08996   4.931 8.18e-07 ***
## as.factor(CEA)2       0.34543    0.09471   3.647 0.000265 ***
## as.factor(size)2      0.23087    0.05502   4.197 2.71e-05 ***
## as.factor(size)3      0.15795    0.05795   2.726 0.006415 **
## as.factor(radiate)1  -0.28772    0.08040  -3.579 0.000345 ***
## as.factor(N)2         0.03429    0.08458   0.405 0.685183
## as.factor(N)3         0.23062    0.06004   3.841 0.000122 ***
## as.factor(marry)2    -0.16999    0.05468  -3.109 0.001879 **
```

```
## as.factor(marry)3          0.01533     0.06071    0.253 0.800649
## as.factor(surgery_pri)1  -0.41198     0.12816   -3.214 0.001307 **
## as.factor(RX_Summ)1       -0.19371     0.07755   -2.498 0.012491 *
## as.factor(T)2             -0.11782     0.07815   -1.508 0.131618
## as.factor(T)3              0.06085     0.07209    0.844 0.398636
## as.factor(income)2        -0.07456     0.05830   -1.279 0.200951
## as.factor(income)3        -0.16281     0.06594   -2.469 0.013546 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 18341  on 17229  degrees of freedom
## Residual deviance: 13924  on 17197  degrees of freedom
## AIC: 13990
##
## Number of Fisher Scoring iterations: 5
```

```
cat("AIC = ", AIC(model_step1), sep = "")
```

```
## AIC = 13990.13
```

BIC threshold (Bayesian information criterion)

```
model_step2 = step(null_model,
                   scope = list(lower=null_model, upper=full_model),
                   direction = "forward", k= log(nrow(os_data)), trace=0)
```

```
#summary model
summary(model_step2)
```

```
##
## Call:
## glm(formula = y_nom ~ as.factor(chem) + as.factor(group) + as.factor(grade) +
##     as.factor(site) + as.factor(lung) + as.factor(bone) + as.factor(age) +
##     as.factor(brain) + as.factor(kind) + as.factor(CEA) + as.factor(surgery_pri) +
##     as.factor(radiate) + as.factor(size), family = binomial(link = "logit"),
##     data = data_nom)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.3829  -0.5886  -0.3917  -0.2109   2.8409
##
## Coefficients:
##                          Estimate Std. Error z value Pr(>|z|)
## (Intercept)              -1.68785    0.25500  -6.619 3.62e-11 ***
## as.factor(chem)1         -2.19080    0.04507 -48.606  < 2e-16 ***
## as.factor(group)2         0.43233    0.07447   5.806 6.41e-09 ***
## as.factor(group)9         0.60663    0.13536   4.482 7.41e-06 ***
## as.factor(grade)2         0.15161    0.12299   1.233 0.217689
## as.factor(grade)3         0.81246    0.12917   6.290 3.18e-10 ***
```

```
## as.factor(grade)4         0.86614    0.17338   4.996 5.87e-07 ***
## as.factor(grade)9         0.64614    0.12475   5.179 2.23e-07 ***
## as.factor(site)2          0.36395    0.04514   8.063 7.47e-16 ***
## as.factor(lung)1          0.36465    0.04795   7.604 2.87e-14 ***
## as.factor(lung)2          0.11511    0.14579   0.790 0.429764
## as.factor(bone)1          0.66706    0.07940   8.401  < 2e-16 ***
## as.factor(bone)2         -0.04538    0.20628  -0.220 0.825885
## as.factor(age)2          -0.09965    0.16335  -0.610 0.541849
## as.factor(age)3           0.31779    0.16009   1.985 0.047140 *
## as.factor(brain)1         1.17174    0.16673   7.028 2.10e-12 ***
## as.factor(brain)2        -0.15739    0.19439  -0.810 0.418130
## as.factor(kind)2          0.38795    0.07324   5.297 1.18e-07 ***
## as.factor(CEA)1           0.44328    0.08975   4.939 7.85e-07 ***
## as.factor(CEA)2           0.36244    0.09447   3.837 0.000125 ***
## as.factor(surgery_pri)1  -0.48249    0.12656  -3.812 0.000138 ***
## as.factor(radiate)1      -0.31389    0.07995  -3.926 8.64e-05 ***
## as.factor(size)2          0.22508    0.05473   4.113 3.91e-05 ***
## as.factor(size)3          0.19944    0.05716   3.489 0.000484 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 18341  on 17229  degrees of freedom
## Residual deviance: 13979  on 17206  degrees of freedom
## AIC: 14027
##
## Number of Fisher Scoring iterations: 5
```

```
cat("BIC = ", BIC(model_step2), sep = "")
```

```
## BIC = 14213.23
```

Perform backward elimination

AIC threshold (Akaike Information Criterion)

```
model_step3 = step(full_model,
                   direction = "backward", k= 2, trace=0)
```

```
#summary model
summary(model_step3)
```

```
##
## Call:
## glm(formula = y_nom ~ as.factor(age) + as.factor(size) + as.factor(marry) +
##     as.factor(income) + as.factor(site) + as.factor(grade) +
##     as.factor(kind) + as.factor(T) + as.factor(N) + as.factor(surgery_pri) +
##     as.factor(RX_Summ) + as.factor(radiate) + as.factor(chem) +
##     as.factor(CEA) + as.factor(bone) + as.factor(brain) + as.factor(lung) +
##     as.factor(group), family = binomial(link = "logit"), data = data_nom)
```

```
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.4390  -0.5888  -0.3873  -0.2092   2.9182
##
## Coefficients:
##                        Estimate Std. Error z value Pr(>|z|)
## (Intercept)            -1.47441    0.26913  -5.478 4.29e-08 ***
## as.factor(age)2        -0.08706    0.16401  -0.531 0.595555
## as.factor(age)3         0.32581    0.16159   2.016 0.043764 *
## as.factor(size)2        0.23087    0.05502   4.197 2.71e-05 ***
## as.factor(size)3        0.15795    0.05795   2.726 0.006415 **
## as.factor(marry)2      -0.16999    0.05468  -3.109 0.001879 **
## as.factor(marry)3       0.01533    0.06071   0.253 0.800649
## as.factor(income)2     -0.07456    0.05830  -1.279 0.200951
## as.factor(income)3     -0.16281    0.06594  -2.469 0.013546 *
## as.factor(site)2        0.35607    0.04534   7.853 4.06e-15 ***
## as.factor(grade)2       0.15701    0.12322   1.274 0.202604
## as.factor(grade)3       0.82402    0.12951   6.362 1.99e-10 ***
## as.factor(grade)4       0.87497    0.17377   5.035 4.77e-07 ***
## as.factor(grade)9       0.61753    0.12527   4.930 8.23e-07 ***
## as.factor(kind)2        0.38942    0.07340   5.306 1.12e-07 ***
## as.factor(T)2          -0.11782    0.07815  -1.508 0.131618
## as.factor(T)3           0.06085    0.07209   0.844 0.398636
## as.factor(N)2           0.03429    0.08458   0.405 0.685183
## as.factor(N)3           0.23062    0.06004   3.841 0.000122 ***
## as.factor(surgery_pri)1 -0.41198   0.12816  -3.214 0.001307 **
## as.factor(RX_Summ)1    -0.19371    0.07755  -2.498 0.012491 *
## as.factor(radiate)1    -0.28772    0.08040  -3.579 0.000345 ***
## as.factor(chem)1       -2.15775    0.04552 -47.398  < 2e-16 ***
## as.factor(CEA)1         0.44360    0.08996   4.931 8.18e-07 ***
## as.factor(CEA)2         0.34543    0.09471   3.647 0.000265 ***
## as.factor(bone)1        0.66938    0.07961   8.408  < 2e-16 ***
## as.factor(bone)2       -0.06360    0.20609  -0.309 0.757632
## as.factor(brain)1       1.18009    0.16814   7.019 2.24e-12 ***
## as.factor(brain)2      -0.20216    0.19445  -1.040 0.298497
## as.factor(lung)1        0.35814    0.04811   7.444 9.74e-14 ***
## as.factor(lung)2        0.11378    0.14595   0.780 0.435669
## as.factor(group)2       0.41362    0.09423   4.390 1.14e-05 ***
## as.factor(group)9       0.49938    0.13782   3.623 0.000291 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 18341  on 17229  degrees of freedom
## Residual deviance: 13924  on 17197  degrees of freedom
## AIC: 13990
##
## Number of Fisher Scoring iterations: 5
```

```r
cat("AIC = ", AIC(model_step3), sep = "")
```

```
## AIC = 13990.13
```

BIC threshold (Bayesian information criterion)

```
model_step4 = step(full_model,
                   direction = "backward", k= log(nrow(os_data)), trace=0)

#summary model
summary(model_step4)
```

```
##
## Call:
## glm(formula = y_nom ~ as.factor(age) + as.factor(size) + as.factor(site) +
##     as.factor(grade) + as.factor(kind) + as.factor(surgery_pri) +
##     as.factor(radiate) + as.factor(chem) + as.factor(CEA) + as.factor(bone) +
##     as.factor(brain) + as.factor(lung) + as.factor(group), family = binomial(link = "logit"),
##     data = data_nom)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.3829  -0.5886  -0.3917  -0.2109   2.8409
##
## Coefficients:
##                         Estimate Std. Error z value Pr(>|z|)
## (Intercept)             -1.68785    0.25500  -6.619 3.62e-11 ***
## as.factor(age)2         -0.09965    0.16335  -0.610 0.541849
## as.factor(age)3          0.31779    0.16009   1.985 0.047140 *
## as.factor(size)2         0.22508    0.05473   4.113 3.91e-05 ***
## as.factor(size)3         0.19944    0.05716   3.489 0.000484 ***
## as.factor(site)2         0.36395    0.04514   8.063 7.47e-16 ***
## as.factor(grade)2        0.15161    0.12299   1.233 0.217689
## as.factor(grade)3        0.81246    0.12917   6.290 3.18e-10 ***
## as.factor(grade)4        0.86614    0.17338   4.996 5.87e-07 ***
## as.factor(grade)9        0.64614    0.12475   5.179 2.23e-07 ***
## as.factor(kind)2         0.38795    0.07324   5.297 1.18e-07 ***
## as.factor(surgery_pri)1 -0.48249    0.12656  -3.812 0.000138 ***
## as.factor(radiate)1     -0.31389    0.07995  -3.926 8.64e-05 ***
## as.factor(chem)1        -2.19080    0.04507 -48.606  < 2e-16 ***
## as.factor(CEA)1          0.44328    0.08975   4.939 7.85e-07 ***
## as.factor(CEA)2          0.36244    0.09447   3.837 0.000125 ***
## as.factor(bone)1         0.66706    0.07940   8.401  < 2e-16 ***
## as.factor(bone)2        -0.04538    0.20628  -0.220 0.825885
## as.factor(brain)1        1.17174    0.16673   7.028 2.10e-12 ***
## as.factor(brain)2       -0.15739    0.19439  -0.810 0.418130
## as.factor(lung)1         0.36465    0.04795   7.604 2.87e-14 ***
## as.factor(lung)2         0.11511    0.14579   0.790 0.429764
## as.factor(group)2        0.43233    0.07447   5.806 6.41e-09 ***
## as.factor(group)9        0.60663    0.13536   4.482 7.41e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 18341  on 17229  degrees of freedom
```

8

```
## Residual deviance: 13979  on 17206  degrees of freedom
## AIC: 14027
##
## Number of Fisher Scoring iterations: 5
```

```r
cat("bIC = ", BIC(model_step3), sep = "")
```

```
## bIC = 14246.03
```

Perform stepwise selection (both)

AIC threshold (Akaike Information Criterion)

```r
model_step5 = step(null_model,
                   scope = list(lower=null_model, upper=full_model),
                   direction = "both", k= 2, trace=0)

#summary model
summary(model_step5)
```

```
##
## Call:
## glm(formula = y_nom ~ as.factor(chem) + as.factor(group) + as.factor(grade) +
##     as.factor(site) + as.factor(lung) + as.factor(bone) + as.factor(age) +
##     as.factor(brain) + as.factor(kind) + as.factor(CEA) + as.factor(size) +
##     as.factor(radiate) + as.factor(N) + as.factor(marry) + as.factor(surgery_pri) +
##     as.factor(RX_Summ) + as.factor(T) + as.factor(income), family = binomial(link = "logit"),
##     data = data_nom)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.4390  -0.5888  -0.3873  -0.2092   2.9182
##
## Coefficients:
##                       Estimate Std. Error z value Pr(>|z|)
## (Intercept)           -1.47441    0.26913  -5.478 4.29e-08 ***
## as.factor(chem)1      -2.15775    0.04552 -47.398  < 2e-16 ***
## as.factor(group)2      0.41362    0.09423   4.390 1.14e-05 ***
## as.factor(group)9      0.49938    0.13782   3.623 0.000291 ***
## as.factor(grade)2      0.15701    0.12322   1.274 0.202604
## as.factor(grade)3      0.82402    0.12951   6.362 1.99e-10 ***
## as.factor(grade)4      0.87497    0.17377   5.035 4.77e-07 ***
## as.factor(grade)9      0.61753    0.12527   4.930 8.23e-07 ***
## as.factor(site)2       0.35607    0.04534   7.853 4.06e-15 ***
## as.factor(lung)1       0.35814    0.04811   7.444 9.74e-14 ***
## as.factor(lung)2       0.11378    0.14595   0.780 0.435669
## as.factor(bone)1       0.66938    0.07961   8.408  < 2e-16 ***
## as.factor(bone)2      -0.06360    0.20609  -0.309 0.757632
## as.factor(age)2       -0.08706    0.16401  -0.531 0.595555
## as.factor(age)3        0.32581    0.16159   2.016 0.043764 *
## as.factor(brain)1      1.18009    0.16814   7.019 2.24e-12 ***
```

9

```
## as.factor(brain)2          -0.20216     0.19445  -1.040 0.298497
## as.factor(kind)2            0.38942     0.07340   5.306 1.12e-07 ***
## as.factor(CEA)1             0.44360     0.08996   4.931 8.18e-07 ***
## as.factor(CEA)2             0.34543     0.09471   3.647 0.000265 ***
## as.factor(size)2            0.23087     0.05502   4.197 2.71e-05 ***
## as.factor(size)3            0.15795     0.05795   2.726 0.006415 **
## as.factor(radiate)1        -0.28772     0.08040  -3.579 0.000345 ***
## as.factor(N)2               0.03429     0.08458   0.405 0.685183
## as.factor(N)3               0.23062     0.06004   3.841 0.000122 ***
## as.factor(marry)2          -0.16999     0.05468  -3.109 0.001879 **
## as.factor(marry)3           0.01533     0.06071   0.253 0.800649
## as.factor(surgery_pri)1    -0.41198     0.12816  -3.214 0.001307 **
## as.factor(RX_Summ)1        -0.19371     0.07755  -2.498 0.012491 *
## as.factor(T)2              -0.11782     0.07815  -1.508 0.131618
## as.factor(T)3               0.06085     0.07209   0.844 0.398636
## as.factor(income)2         -0.07456     0.05830  -1.279 0.200951
## as.factor(income)3         -0.16281     0.06594  -2.469 0.013546 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 18341  on 17229  degrees of freedom
## Residual deviance: 13924  on 17197  degrees of freedom
## AIC: 13990
##
## Number of Fisher Scoring iterations: 5
```

```r
cat("AIC = ", AIC(model_step5), sep = "")
```

```
## AIC = 13990.13
```

BIC threshold (Bayesian information criterion)

```r
model_step6 = step(null_model,
                   scope = list(lower=null_model, upper=full_model),
                   direction = "both", k= log(nrow(os_data)), trace=0)

#summary model
summary(model_step6)
```

```
##
## Call:
## glm(formula = y_nom ~ as.factor(chem) + as.factor(group) + as.factor(grade) +
##     as.factor(site) + as.factor(lung) + as.factor(bone) + as.factor(age) +
##     as.factor(brain) + as.factor(kind) + as.factor(CEA) + as.factor(surgery_pri) +
##     as.factor(radiate) + as.factor(size), family = binomial(link = "logit"),
##     data = data_nom)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
```

```
## -2.3829  -0.5886  -0.3917  -0.2109   2.8409
##
## Coefficients:
##                          Estimate Std. Error z value Pr(>|z|)
## (Intercept)              -1.68785    0.25500  -6.619 3.62e-11 ***
## as.factor(chem)1         -2.19080    0.04507 -48.606  < 2e-16 ***
## as.factor(group)2         0.43233    0.07447   5.806 6.41e-09 ***
## as.factor(group)9         0.60663    0.13536   4.482 7.41e-06 ***
## as.factor(grade)2         0.15161    0.12299   1.233 0.217689
## as.factor(grade)3         0.81246    0.12917   6.290 3.18e-10 ***
## as.factor(grade)4         0.86614    0.17338   4.996 5.87e-07 ***
## as.factor(grade)9         0.64614    0.12475   5.179 2.23e-07 ***
## as.factor(site)2          0.36395    0.04514   8.063 7.47e-16 ***
## as.factor(lung)1          0.36465    0.04795   7.604 2.87e-14 ***
## as.factor(lung)2          0.11511    0.14579   0.790 0.429764
## as.factor(bone)1          0.66706    0.07940   8.401  < 2e-16 ***
## as.factor(bone)2         -0.04538    0.20628  -0.220 0.825885
## as.factor(age)2          -0.09965    0.16335  -0.610 0.541849
## as.factor(age)3           0.31779    0.16009   1.985 0.047140 *
## as.factor(brain)1         1.17174    0.16673   7.028 2.10e-12 ***
## as.factor(brain)2        -0.15739    0.19439  -0.810 0.418130
## as.factor(kind)2          0.38795    0.07324   5.297 1.18e-07 ***
## as.factor(CEA)1           0.44328    0.08975   4.939 7.85e-07 ***
## as.factor(CEA)2           0.36244    0.09447   3.837 0.000125 ***
## as.factor(surgery_pri)1  -0.48249    0.12656  -3.812 0.000138 ***
## as.factor(radiate)1      -0.31389    0.07995  -3.926 8.64e-05 ***
## as.factor(size)2          0.22508    0.05473   4.113 3.91e-05 ***
## as.factor(size)3          0.19944    0.05716   3.489 0.000484 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 18341  on 17229  degrees of freedom
## Residual deviance: 13979  on 17206  degrees of freedom
## AIC: 14027
##
## Number of Fisher Scoring iterations: 5
```

```
cat("BIC = ", BIC(model_step6), sep = "")
```

```
## BIC = 14213.23
```