

CRCLM_2

Zhiyi Sun

2023-03-17

Data

```
CRCLM <- read_excel("CRCLM_final.xlsx")
```

```
str(CRCLM)
```

```
## tibble [28,628 x 22] (S3: tbl_df/tbl/data.frame)
## $ y      : num [1:28628] 1 0 0 0 0 0 1 0 0 1 ...
## $ cod    : num [1:28628] 1 0 1 1 1 1 1 1 1 1 ...
## $ gender : num [1:28628] 2 2 1 2 1 1 2 2 2 1 ...
## $ race   : num [1:28628] 2 1 1 1 3 2 1 2 1 1 ...
## $ age    : num [1:28628] 3 3 3 3 3 3 3 3 2 3 ...
## $ size   : num [1:28628] 3 1 1 3 1 3 2 2 2 1 ...
## $ marry  : num [1:28628] 1 2 3 2 3 2 2 1 2 3 ...
## $ income : num [1:28628] 3 3 3 3 3 3 3 3 3 3 ...
## $ year   : num [1:28628] 2018 2010 2010 2010 2010 ...
## $ site   : num [1:28628] 1 1 2 2 1 1 2 2 1 2 ...
## $ grade  : num [1:28628] 9 2 2 2 9 2 9 2 2 9 ...
## $ kind   : num [1:28628] 1 1 1 1 1 1 2 1 1 1 ...
## $ t      : num [1:28628] 3 2 2 2 2 1 1 2 2 3 ...
## $ N      : num [1:28628] 3 1 2 1 1 1 1 2 1 1 ...
## $ surgery: num [1:28628] 0 1 1 0 0 0 0 1 1 0 ...
## $ RX     : num [1:28628] 0 1 0 0 0 0 0 0 0 0 ...
## $ radiate: num [1:28628] 0 0 0 0 0 0 0 0 0 0 ...
## $ chem   : num [1:28628] 0 1 0 1 0 1 1 0 1 0 ...
## $ CEA    : num [1:28628] 1 2 0 1 1 2 1 1 1 1 ...
## $ bone   : num [1:28628] 0 0 0 0 0 0 0 0 0 0 ...
## $ brain  : num [1:28628] 0 0 0 0 0 0 0 0 0 0 ...
## $ lung   : num [1:28628] 1 0 0 0 1 1 0 0 0 1 ...
```

```
CRCLM_new <- CRCLM[, c('y', 'cod', 'gender', 'race', 'age', 'size', 'marry', 'income', 'site', 'grade',
```

```
CRCLM_new$y <- as.factor(CRCLM_new$y)
CRCLM_new$gender <- as.factor(CRCLM_new$gender)
CRCLM_new$race <- as.factor(CRCLM_new$race)
CRCLM_new$age <- as.factor(CRCLM_new$age)
CRCLM_new$size <- as.factor(CRCLM_new$size)
CRCLM_new$marry <- as.factor(CRCLM_new$marry)
CRCLM_new$income <- as.factor(CRCLM_new$income)
CRCLM_new$site <- as.factor(CRCLM_new$site)
```

```

CRCLM_new$grade <- as.factor(CRCLM_new$grade)
CRCLM_new$kind <- as.factor(CRCLM_new$kind)
CRCLM_new$t <- as.factor(CRCLM_new$t)
CRCLM_new$N <- as.factor(CRCLM_new$N)
CRCLM_new$surgery <- as.factor(CRCLM_new$surgery)
CRCLM_new$RX <- as.factor(CRCLM_new$RX)
CRCLM_new$radiate <- as.factor(CRCLM_new$radiate)
CRCLM_new$chem <- as.factor(CRCLM_new$chem)
CRCLM_new$CEA <- as.factor(CRCLM_new$CEA)
CRCLM_new$bone <- as.factor(CRCLM_new$bone)
CRCLM_new$brain <- as.factor(CRCLM_new$brain)
CRCLM_new$lung <- as.factor(CRCLM_new$lung)

CRCLM_0_1 <- CRCLM_new[CRCLM_new$cod == 0 | CRCLM_new$cod == 1,]
str(CRCLM_0_1)

```

```

## tibble [26,739 x 21] (S3: tbl_df/tbl/data.frame)
## $ y      : Factor w/ 2 levels "0","1": 2 1 1 1 1 1 2 1 1 2 ...
## $ cod     : num [1:26739] 1 0 1 1 1 1 1 1 1 1 ...
## $ gender  : Factor w/ 2 levels "1","2": 2 2 1 2 1 1 2 2 2 1 ...
## $ race    : Factor w/ 3 levels "1","2","3": 2 1 1 1 3 2 1 2 1 1 ...
## $ age     : Factor w/ 3 levels "1","2","3": 3 3 3 3 3 3 3 3 2 3 ...
## $ size    : Factor w/ 3 levels "1","2","3": 3 1 1 3 1 3 2 2 2 1 ...
## $ marry   : Factor w/ 3 levels "1","2","3": 1 2 3 2 3 2 2 1 2 3 ...
## $ income  : Factor w/ 3 levels "1","2","3": 3 3 3 3 3 3 3 3 3 3 ...
## $ site    : Factor w/ 2 levels "1","2": 1 1 2 2 1 1 2 2 1 2 ...
## $ grade   : Factor w/ 5 levels "1","2","3","4",...: 5 2 2 2 5 2 5 2 2 5 ...
## $ kind    : Factor w/ 2 levels "1","2": 1 1 1 1 1 1 2 1 1 1 ...
## $ t       : Factor w/ 3 levels "1","2","3": 3 2 2 2 2 1 1 2 2 3 ...
## $ N       : Factor w/ 3 levels "1","2","3": 3 1 2 1 1 1 1 2 1 1 ...
## $ surgery : Factor w/ 2 levels "0","1": 1 2 2 1 1 1 1 2 2 1 ...
## $ RX      : Factor w/ 2 levels "0","1": 1 2 1 1 1 1 1 1 1 1 ...
## $ radiate : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
## $ chem    : Factor w/ 2 levels "0","1": 1 2 1 2 1 2 2 1 2 1 ...
## $ CEA     : Factor w/ 3 levels "0","1","2": 2 3 1 2 2 3 2 2 2 2 ...
## $ bone    : Factor w/ 3 levels "0","1","2": 1 1 1 1 1 1 1 1 1 1 ...
## $ brain   : Factor w/ 3 levels "0","1","2": 1 1 1 1 1 1 1 1 1 1 ...
## $ lung    : Factor w/ 3 levels "0","1","2": 2 1 1 1 2 2 1 1 1 2 ...

```

```

CRCLM_0_1_2 <- CRCLM_new
str(CRCLM_0_1_2)

```

```

## tibble [28,628 x 21] (S3: tbl_df/tbl/data.frame)
## $ y      : Factor w/ 2 levels "0","1": 2 1 1 1 1 1 2 1 1 2 ...
## $ cod     : num [1:28628] 1 0 1 1 1 1 1 1 1 1 ...
## $ gender  : Factor w/ 2 levels "1","2": 2 2 1 2 1 1 2 2 2 1 ...
## $ race    : Factor w/ 3 levels "1","2","3": 2 1 1 1 3 2 1 2 1 1 ...
## $ age     : Factor w/ 3 levels "1","2","3": 3 3 3 3 3 3 3 3 2 3 ...
## $ size    : Factor w/ 3 levels "1","2","3": 3 1 1 3 1 3 2 2 2 1 ...
## $ marry   : Factor w/ 3 levels "1","2","3": 1 2 3 2 3 2 2 1 2 3 ...
## $ income  : Factor w/ 3 levels "1","2","3": 3 3 3 3 3 3 3 3 3 3 ...
## $ site    : Factor w/ 2 levels "1","2": 1 1 2 2 1 1 2 2 1 2 ...
## $ grade   : Factor w/ 5 levels "1","2","3","4",...: 5 2 2 2 5 2 5 2 2 5 ...

```

```
## $ kind : Factor w/ 2 levels "1","2": 1 1 1 1 1 1 2 1 1 1 ...
## $ t : Factor w/ 3 levels "1","2","3": 3 2 2 2 2 1 1 2 2 3 ...
## $ N : Factor w/ 3 levels "1","2","3": 3 1 2 1 1 1 1 2 1 1 ...
## $ surgery: Factor w/ 2 levels "0","1": 1 2 2 1 1 1 1 2 2 1 ...
## $ RX : Factor w/ 2 levels "0","1": 1 2 1 1 1 1 1 1 1 1 ...
## $ radiate: Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
## $ chem : Factor w/ 2 levels "0","1": 1 2 1 2 1 2 2 1 2 1 ...
## $ CEA : Factor w/ 3 levels "0","1","2": 2 3 1 2 2 3 2 2 2 2 ...
## $ bone : Factor w/ 3 levels "0","1","2": 1 1 1 1 1 1 1 1 1 1 ...
## $ brain : Factor w/ 3 levels "0","1","2": 1 1 1 1 1 1 1 1 1 1 ...
## $ lung : Factor w/ 3 levels "0","1","2": 2 1 1 1 2 2 1 1 1 2 ...
```

0+1+2 train/test

```
set.seed(2023)
sample <- sample(nrow(CRCLM_0_1_2), floor(nrow(CRCLM_0_1_2)*0.8))
train_0_1_2 <- CRCLM_0_1_2[sample,]
test_0_1_2 <- CRCLM_0_1_2[-sample,]

prevalence_0_1_2_train <- sum(train_0_1_2$cod == 1)/nrow(train_0_1_2)
prevalence_0_1_2_test <- sum(test_0_1_2$cod == 1)/nrow(test_0_1_2)

train_0_1_2 <- train_0_1_2[, c('y', 'gender', 'race', 'age', 'size', 'marry', 'income', 'site', 'grade',
str(train_0_1_2)
```

```
## tibble [22,902 x 20] (S3: tbl_df/tbl/data.frame)
## $ y : Factor w/ 2 levels "0","1": 1 1 2 1 1 1 1 1 2 1 ...
## $ gender : Factor w/ 2 levels "1","2": 2 1 2 1 1 2 2 1 2 1 ...
## $ race : Factor w/ 3 levels "1","2","3": 2 1 1 1 1 1 1 1 1 3 ...
## $ age : Factor w/ 3 levels "1","2","3": 2 2 3 3 2 3 1 2 3 3 ...
## $ size : Factor w/ 3 levels "1","2","3": 1 3 3 2 3 2 1 2 3 3 ...
## $ marry : Factor w/ 3 levels "1","2","3": 2 2 2 1 2 3 2 3 2 2 ...
## $ income : Factor w/ 3 levels "1","2","3": 2 2 3 3 3 2 2 3 3 2 ...
## $ site : Factor w/ 2 levels "1","2": 1 2 2 1 1 1 1 2 1 2 ...
## $ grade : Factor w/ 5 levels "1","2","3","4",...: 5 5 5 5 2 1 2 2 5 2 ...
## $ kind : Factor w/ 2 levels "1","2": 1 1 1 1 1 1 1 1 2 1 ...
## $ t : Factor w/ 3 levels "1","2","3": 2 2 3 2 2 1 2 3 2 3 ...
## $ N : Factor w/ 3 levels "1","2","3": 1 1 3 2 1 1 2 1 1 3 ...
## $ surgery: Factor w/ 2 levels "0","1": 1 1 1 1 1 2 2 1 1 1 ...
## $ RX : Factor w/ 2 levels "0","1": 2 1 1 1 1 1 1 1 1 1 ...
## $ radiate: Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
## $ chem : Factor w/ 2 levels "0","1": 2 1 2 2 2 1 2 2 1 1 ...
## $ CEA : Factor w/ 3 levels "0","1","2": 2 1 2 2 2 2 3 2 2 3 ...
## $ bone : Factor w/ 3 levels "0","1","2": 1 1 1 1 1 1 1 1 1 1 ...
## $ brain : Factor w/ 3 levels "0","1","2": 1 1 1 1 1 1 1 1 1 1 ...
## $ lung : Factor w/ 3 levels "0","1","2": 1 2 1 1 2 2 1 1 2 1 ...
```

```
test_0_1_2 <- test_0_1_2[, c('y', 'gender', 'race', 'age', 'size', 'marry', 'income', 'site', 'grade',
str(test_0_1_2)
```

```
## tibble [5,726 x 20] (S3: tbl_df/tbl/data.frame)
```

```
## $ y      : Factor w/ 2 levels "0","1": 1 1 1 1 2 1 1 1 2 2 ...
## $ gender : Factor w/ 2 levels "1","2": 1 2 1 2 1 2 2 1 1 2 ...
## $ race   : Factor w/ 3 levels "1","2","3": 1 1 2 2 1 3 1 3 3 2 ...
## $ age    : Factor w/ 3 levels "1","2","3": 3 3 3 3 3 2 2 3 3 2 ...
## $ size   : Factor w/ 3 levels "1","2","3": 1 3 3 2 1 2 3 2 3 2 ...
## $ marry  : Factor w/ 3 levels "1","2","3": 3 2 2 1 3 2 3 3 2 1 ...
## $ income : Factor w/ 3 levels "1","2","3": 3 3 3 3 3 3 3 3 3 3 ...
## $ site   : Factor w/ 2 levels "1","2": 2 2 1 2 2 1 2 2 2 1 ...
## $ grade  : Factor w/ 5 levels "1","2","3","4",...: 2 2 2 2 5 2 2 2 5 2 ...
## $ kind   : Factor w/ 2 levels "1","2": 1 1 1 1 1 1 1 1 2 1 ...
## $ t      : Factor w/ 3 levels "1","2","3": 2 2 1 2 3 2 1 2 2 2 ...
## $ N      : Factor w/ 3 levels "1","2","3": 2 1 1 2 1 1 1 2 2 1 ...
## $ surgery: Factor w/ 2 levels "0","1": 2 1 1 2 1 1 1 2 1 2 ...
## $ RX     : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
## $ radiate: Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
## $ chem   : Factor w/ 2 levels "0","1": 1 2 2 1 1 2 2 2 2 1 ...
## $ CEA    : Factor w/ 3 levels "0","1","2": 1 2 3 2 2 2 2 2 3 2 ...
## $ bone   : Factor w/ 3 levels "0","1","2": 1 1 1 1 1 1 1 1 1 1 ...
## $ brain  : Factor w/ 3 levels "0","1","2": 1 1 1 1 1 1 1 1 1 1 ...
## $ lung   : Factor w/ 3 levels "0","1","2": 1 1 2 1 2 1 1 1 1 1 ...
```

0+1+2 univariate class

```
u1_2 <- glm(y ~ gender, binomial(link='logit'), data = train_0_1_2)
summary(u1_2)
```

```
##
## Call:
## glm(formula = y ~ gender, family = binomial(link = "logit"),
##      data = train_0_1_2)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.6736  -0.6736  -0.6363  -0.6363   1.8422
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.36773    0.02515 -54.388  < 2e-16 ***
## gender2     -0.12661    0.03378  -3.748 0.000179 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 22383  on 22901  degrees of freedom
## Residual deviance: 22369  on 22900  degrees of freedom
## AIC: 22373
##
## Number of Fisher Scoring iterations: 4
```

```
u2_2 <- glm(y ~ race, binomial(link='logit'), data = train_0_1_2)
summary(u2_2)
```

```
##
## Call:
## glm(formula = y ~ race, family = binomial(link = "logit"), data = train_0_1_2)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.6921  -0.6501  -0.6501  -0.6078   1.8867
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.44674    0.01938 -74.661  <2e-16 ***
## race2        0.13980    0.04624   3.023  0.0025 **
## race3       -0.14835    0.05958  -2.490  0.0128 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 22383  on 22901  degrees of freedom
## Residual deviance: 22365  on 22899  degrees of freedom
## AIC: 22371
##
## Number of Fisher Scoring iterations: 4
```

```
u3_2 <- glm(y ~ age, binomial(link='logit'), data = train_0_1_2)
summary(u3_2)
```

```
##
## Call:
## glm(formula = y ~ age, family = binomial(link = "logit"), data = train_0_1_2)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.733  -0.733  -0.509  -0.509   2.107
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -2.1058    0.1223 -17.219  < 2e-16 ***
## age2         0.1273    0.1270   1.003   0.316
## age3         0.9287    0.1239   7.497 6.55e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 22383  on 22901  degrees of freedom
## Residual deviance: 21895  on 22899  degrees of freedom
## AIC: 21901
##
## Number of Fisher Scoring iterations: 4
```

```
u4_2 <- glm(y ~ size, binomial(link='logit'), data = train_0_1_2)
summary(u4_2)
```

```
##
## Call:
## glm(formula = y ~ size, family = binomial(link = "logit"), data = train_0_1_2)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.7774  -0.6416  -0.5542  -0.5542   1.9746
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.79594    0.02980 -60.267 < 2e-16 ***
## size2        0.31987    0.04354   7.346 2.04e-13 ***
## size3        0.75398    0.04020  18.755 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 22383  on 22901  degrees of freedom
## Residual deviance: 22023  on 22899  degrees of freedom
## AIC: 22029
##
## Number of Fisher Scoring iterations: 4
```

```
u5_2 <- glm(y ~ marry, binomial(link='logit'), data = train_0_1_2)
summary(u5_2)
```

```
##
## Call:
## glm(formula = y ~ marry, family = binomial(link = "logit"), data = train_0_1_2)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.7566  -0.6906  -0.5864  -0.5864   1.9211
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.31191    0.03383 -38.778 < 2e-16 ***
## marry2       -0.36142    0.04190  -8.625 < 2e-16 ***
## marry3        0.20734    0.04626   4.482 7.38e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 22383  on 22901  degrees of freedom
## Residual deviance: 22168  on 22899  degrees of freedom
## AIC: 22174
##
## Number of Fisher Scoring iterations: 4
```

```
u6_2 <- glm(y ~ income, binomial(link='logit'), data = train_0_1_2)
summary(u6_2)
```

```
##
## Call:
## glm(formula = y ~ income, family = binomial(link = "logit"),
##      data = train_0_1_2)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.6679  -0.6665  -0.6665  -0.6164   1.8731
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.386936   0.040055 -34.626 < 2e-16 ***
## income2      -0.004435   0.045966  -0.096 0.923135
## income3      -0.177342   0.051492  -3.444 0.000573 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 22383  on 22901  degrees of freedom
## Residual deviance: 22361  on 22899  degrees of freedom
## AIC: 22367
##
## Number of Fisher Scoring iterations: 4
```

```
u7_2 <- glm(y ~ site, binomial(link='logit'), data = train_0_1_2)
summary(u7_2)
```

```
##
## Call:
## glm(formula = y ~ site, family = binomial(link = "logit"), data = train_0_1_2)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.7253  -0.7253  -0.6082  -0.6082   1.8860
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.59358    0.02214 -71.97  <2e-16 ***
## site2        0.39246    0.03409  11.51  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 22383  on 22901  degrees of freedom
## Residual deviance: 22252  on 22900  degrees of freedom
## AIC: 22256
##
## Number of Fisher Scoring iterations: 4
```

```
u8_2 <- glm(y ~ grade, binomial(link='logit'), data = train_0_1_2)
summary(u8_2)
```

```
##
## Call:
## glm(formula = y ~ grade, family = binomial(link = "logit"), data = train_0_1_2)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.7967  -0.7256  -0.5355  -0.5355   2.0066
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.80615    0.10062 -17.950  < 2e-16 ***
## grade2      -0.06367    0.10435  -0.610  0.541763
## grade3       0.60611    0.10850   5.586  2.32e-08 ***
## grade4       0.54750    0.14400   3.802  0.000143 ***
## grade9       0.82135    0.10424   7.880  3.29e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 22383  on 22901  degrees of freedom
## Residual deviance: 21804  on 22897  degrees of freedom
## AIC: 21814
##
## Number of Fisher Scoring iterations: 4
```

```
u9_2 <- glm(y ~ kind, binomial(link='logit'), data = train_0_1_2)
summary(u9_2)
```

```
##
## Call:
## glm(formula = y ~ kind, family = binomial(link = "logit"), data = train_0_1_2)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.7857  -0.6413  -0.6413  -0.6413   1.8345
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.47700    0.01764 -83.725  < 2e-16 ***
## kind2        0.45987    0.05818   7.904  2.69e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 22383  on 22901  degrees of freedom
## Residual deviance: 22324  on 22900  degrees of freedom
## AIC: 22328
```



```
##
## Number of Fisher Scoring iterations: 4

u10_2 <- glm(y ~ t, binomial(link='logit'), data = train_0_1_2)
summary(u10_2)

##
## Call:
## glm(formula = y ~ t, family = binomial(link = "logit"), data = train_0_1_2)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.7895  -0.7151  -0.5602  -0.5602   1.9645
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.23321    0.04863 -25.358 < 2e-16 ***
## t2          -0.53942    0.05447  -9.903 < 2e-16 ***
## t3           0.22737    0.05553   4.094 4.23e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 22383  on 22901  degrees of freedom
## Residual deviance: 21922  on 22899  degrees of freedom
## AIC: 21928
##
## Number of Fisher Scoring iterations: 4

u11_2 <- glm(y ~ N, binomial(link='logit'), data = train_0_1_2)
summary(u11_2)

##
## Call:
## glm(formula = y ~ N, family = binomial(link = "logit"), data = train_0_1_2)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.8387  -0.6292  -0.6292  -0.5675   1.9523
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.51912    0.02221 -68.398 < 2e-16 ***
## N2          -0.22554    0.04435  -5.085 3.67e-07 ***
## N3           0.65515    0.04195  15.616 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 22383  on 22901  degrees of freedom
## Residual deviance: 22062  on 22899  degrees of freedom
```

```
## AIC: 22068
##
## Number of Fisher Scoring iterations: 4

u12_2 <- glm(y ~ surgery, binomial(link='logit'), data = train_0_1_2)
summary(u12_2)
```

```
##
## Call:
## glm(formula = y ~ surgery, family = binomial(link = "logit"),
##      data = train_0_1_2)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.7780  -0.7780  -0.5124  -0.5124   2.0471
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.03997    0.02134  -48.74  <2e-16 ***
## surgery1    -0.92412    0.03550  -26.04  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 22383  on 22901  degrees of freedom
## Residual deviance: 21665  on 22900  degrees of freedom
## AIC: 21669
##
## Number of Fisher Scoring iterations: 4
```

```
u13_2 <- glm(y ~ RX, binomial(link='logit'), data = train_0_1_2)
summary(u13_2)

##
## Call:
## glm(formula = y ~ RX, family = binomial(link = "logit"), data = train_0_1_2)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.6883  -0.6883  -0.6883  -0.4606   2.1431
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.31935    0.01783  -74.01  <2e-16 ***
## RX1         -0.87094    0.05553  -15.69  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 22383  on 22901  degrees of freedom
## Residual deviance: 22089  on 22900  degrees of freedom
```

```
## AIC: 22093
##
## Number of Fisher Scoring iterations: 4
```

```
u14_2 <- glm(y ~ radiate, binomial(link='logit'), data = train_0_1_2)
summary(u14_2)
```

```
##
## Call:
## glm(formula = y ~ radiate, family = binomial(link = "logit"),
##      data = train_0_1_2)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.6696  -0.6696  -0.6696  -0.5220   2.0301
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.38113    0.01763  -78.351  <2e-16 ***
## radiate1    -0.54330    0.05889   -9.226  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 22383  on 22901  degrees of freedom
## Residual deviance: 22288  on 22900  degrees of freedom
## AIC: 22292
##
## Number of Fisher Scoring iterations: 4
```

```
u15_2 <- glm(y ~ chem, binomial(link='logit'), data = train_0_1_2)
summary(u15_2)
```

```
##
## Call:
## glm(formula = y ~ chem, family = binomial(link = "logit"), data = train_0_1_2)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1473  -0.4597  -0.4597  -0.4597   2.1447
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.07118    0.02703  -2.633  0.00845 **
## chem1       -2.12292    0.03697 -57.422  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 22383  on 22901  degrees of freedom
## Residual deviance: 18939  on 22900  degrees of freedom
```

```
## AIC: 18943
##
## Number of Fisher Scoring iterations: 4
```

```
u16_2 <- glm(y ~ CEA, binomial(link='logit'), data = train_0_1_2)
summary(u16_2)
```

```
##
## Call:
## glm(formula = y ~ CEA, family = binomial(link = "logit"), data = train_0_1_2)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.7127  -0.6508  -0.6508  -0.4969   2.0751
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -2.02953    0.06509 -31.178  <2e-16 ***
## CEA1         0.58495    0.06835   8.558  <2e-16 ***
## CEA2         0.78865    0.07241  10.892  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 22383  on 22901  degrees of freedom
## Residual deviance: 22249  on 22899  degrees of freedom
## AIC: 22255
##
## Number of Fisher Scoring iterations: 4
```

```
u17_2 <- glm(y ~ bone, binomial(link='logit'), data = train_0_1_2)
summary(u17_2)
```

```
##
## Call:
## glm(formula = y ~ bone, family = binomial(link = "logit"), data = train_0_1_2)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.9023  -0.6337  -0.6337  -0.6337   1.8462
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.50344    0.01778 -84.551  < 2e-16 ***
## bone1        0.81514    0.06281  12.978  < 2e-16 ***
## bone2        0.49350    0.11278   4.376 1.21e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 22383  on 22901  degrees of freedom
```

```
## Residual deviance: 22215 on 22899 degrees of freedom
## AIC: 22221
##
## Number of Fisher Scoring iterations: 4
```

```
u18_2 <- glm(y ~ brain, binomial(link='logit'), data = train_0_1_2)
summary(u18_2)
```

```
##
## Call:
## glm(formula = y ~ brain, family = binomial(link = "logit"), data = train_0_1_2)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1274  -0.6438  -0.6438  -0.6438   1.8306
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -1.4684     0.0172 -85.383 < 2e-16 ***
## brain1         1.3495     0.1365   9.888 < 2e-16 ***
## brain2         0.5044     0.1057   4.772 1.83e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 22383 on 22901 degrees of freedom
## Residual deviance: 22274 on 22899 degrees of freedom
## AIC: 22280
##
## Number of Fisher Scoring iterations: 4
```

```
u19_2 <- glm(y ~ lung, binomial(link='logit'), data = train_0_1_2)
summary(u19_2)
```

```
##
## Call:
## glm(formula = y ~ lung, family = binomial(link = "logit"), data = train_0_1_2)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.7671  -0.6085  -0.6085  -0.6085   1.8856
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.59259     0.02060 -77.327 < 2e-16 ***
## lung1         0.51993     0.03679  14.132 < 2e-16 ***
## lung2         0.44621     0.11187   3.989 6.64e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
```

```
## Null deviance: 22383 on 22901 degrees of freedom
## Residual deviance: 22183 on 22899 degrees of freedom
## AIC: 22189
##
## Number of Fisher Scoring iterations: 4
```

0+1+2 univariate unclass

```
uc1_2 <- glm(y ~ unclass(gender), binomial(link='logit'), data = train_0_1_2)
summary(uc1_2)
```

```
##
## Call:
## glm(formula = y ~ unclass(gender), family = binomial(link = "logit"),
## data = train_0_1_2)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.6736  -0.6736  -0.6363  -0.6363   1.8422
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -1.24112    0.05512 -22.516 < 2e-16 ***
## unclass(gender) -0.12661    0.03378  -3.748 0.000179 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 22383 on 22901 degrees of freedom
## Residual deviance: 22369 on 22900 degrees of freedom
## AIC: 22373
##
## Number of Fisher Scoring iterations: 4
```

```
uc2_2 <- glm(y ~ unclass(race), binomial(link='logit'), data = train_0_1_2)
summary(uc2_2)
```

```
##
## Call:
## glm(formula = y ~ unclass(race), family = binomial(link = "logit"),
## data = train_0_1_2)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.6545  -0.6545  -0.6545  -0.6422   1.8332
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -1.41085    0.03868 -36.480 <2e-16 ***
## unclass(race) -0.02105    0.02603  -0.809   0.419
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 22383  on 22901  degrees of freedom
## Residual deviance: 22382  on 22900  degrees of freedom
## AIC: 22386
##
## Number of Fisher Scoring iterations: 4

uc3_2 <- glm(y ~ unclass(age), binomial(link='logit'), data = train_0_1_2)
summary(uc3_2)
```

```
##
## Call:
## glm(formula = y ~ unclass(age), family = binomial(link = "logit"),
##      data = train_0_1_2)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.7294  -0.7294  -0.5248  -0.5248   2.3265
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -3.3617     0.0979  -34.34  <2e-16 ***
## unclass(age)    0.7244     0.0355   20.41  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 22383  on 22901  degrees of freedom
## Residual deviance: 21915  on 22900  degrees of freedom
## AIC: 21919
##
## Number of Fisher Scoring iterations: 4
```

```
uc4_2 <- glm(y ~ unclass(size), binomial(link='logit'), data = train_0_1_2)
summary(uc4_2)
```

```
##
## Call:
## glm(formula = y ~ unclass(size), family = binomial(link = "logit"),
##      data = train_0_1_2)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.7724  -0.6534  -0.5495  -0.5495   1.9825
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -2.19281     0.04510  -48.62  <2e-16 ***
```

```
## unclass(size) 0.37865 0.02017 18.77 <2e-16 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 22383 on 22901 degrees of freedom
## Residual deviance: 22025 on 22900 degrees of freedom
## AIC: 22029
##
## Number of Fisher Scoring iterations: 4
```

```
uc5_2 <- glm(y ~ unclass(marry), binomial(link='logit'), data = train_0_1_2)
summary(uc5_2)
```

```
##
## Call:
## glm(formula = y ~ unclass(marry), family = binomial(link = "logit"),
## data = train_0_1_2)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.6882  -0.6515  -0.6515  -0.6165   1.8731
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -1.68638    0.05318 -31.708 < 2e-16 ***
## unclass(marry) 0.12218    0.02472  4.943 7.71e-07 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 22383 on 22901 degrees of freedom
## Residual deviance: 22358 on 22900 degrees of freedom
## AIC: 22362
##
## Number of Fisher Scoring iterations: 4
```

```
uc6_2 <- glm(y ~ unclass(income), binomial(link='logit'), data = train_0_1_2)
summary(uc6_2)
```

```
##
## Call:
## glm(formula = y ~ unclass(income), family = binomial(link = "logit"),
## data = train_0_1_2)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.6851  -0.6555  -0.6555  -0.6270   1.8566
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
```



```
## (Intercept)      -1.23154      0.05511 -22.346 < 2e-16 ***
## unclass(income) -0.09847      0.02507  -3.929 8.54e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 22383  on 22901  degrees of freedom
## Residual deviance: 22367  on 22900  degrees of freedom
## AIC: 22371
##
## Number of Fisher Scoring iterations: 4
```

```
uc7_2 <- glm(y ~ unclass(site), binomial(link='logit'), data = train_0_1_2)
summary(uc7_2)
```

```
##
## Call:
## glm(formula = y ~ unclass(site), family = binomial(link = "logit"),
##      data = train_0_1_2)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.7253  -0.7253  -0.6082  -0.6082   1.8860
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -1.98604    0.05131  -38.70 <2e-16 ***
## unclass(site)  0.39246    0.03409   11.51 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 22383  on 22901  degrees of freedom
## Residual deviance: 22252  on 22900  degrees of freedom
## AIC: 22256
##
## Number of Fisher Scoring iterations: 4
```

```
uc8_2 <- glm(y ~ unclass(grade), binomial(link='logit'), data = train_0_1_2)
summary(uc8_2)
```

```
##
## Call:
## glm(formula = y ~ unclass(grade), family = binomial(link = "logit"),
##      data = train_0_1_2)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.8083  -0.6358  -0.5613  -0.5613   2.0800
##
## Coefficients:
```

```
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -2.31352    0.04403  -52.54  <2e-16 ***
## unclass(grade) 0.27248    0.01208   22.56  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 22383  on 22901  degrees of freedom
## Residual deviance: 21875  on 22900  degrees of freedom
## AIC: 21879
##
## Number of Fisher Scoring iterations: 4
```

```
uc9_2 <- glm(y ~ unclass(kind), binomial(link='logit'), data = train_0_1_2)
summary(uc9_2)
```

```
##
## Call:
## glm(formula = y ~ unclass(kind), family = binomial(link = "logit"),
##      data = train_0_1_2)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.7857  -0.6413  -0.6413  -0.6413   1.8345
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -1.93687    0.06572 -29.474  < 2e-16 ***
## unclass(kind)  0.45987    0.05818   7.904 2.69e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 22383  on 22901  degrees of freedom
## Residual deviance: 22324  on 22900  degrees of freedom
## AIC: 22328
##
## Number of Fisher Scoring iterations: 4
```

```
uc10_2 <- glm(y ~ unclass(t), binomial(link='logit'), data = train_0_1_2)
summary(uc10_2)
```

```
##
## Call:
## glm(formula = y ~ unclass(t), family = binomial(link = "logit"),
##      data = train_0_1_2)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.7340  -0.6276  -0.6276  -0.5340   2.0092
##
```

```
## Coefficients:
##           Estimate Std. Error z value Pr(>|z|)
## (Intercept) -2.22682    0.06616  -33.66  <2e-16 ***
## unclass(t)   0.35099    0.02801   12.53  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 22383  on 22901  degrees of freedom
## Residual deviance: 22223  on 22900  degrees of freedom
## AIC: 22227
##
## Number of Fisher Scoring iterations: 4

uc11_2 <- glm(y ~ unclass(N), binomial(link='logit'), data = train_0_1_2)
summary(uc11_2)
```

```
##
## Call:
## glm(formula = y ~ unclass(N), family = binomial(link = "logit"),
##      data = train_0_1_2)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.7665  -0.6829  -0.6066  -0.6066   1.8886
##
## Coefficients:
##           Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.86201    0.03895  -47.81  <2e-16 ***
## unclass(N)   0.26249    0.02118   12.39  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 22383  on 22901  degrees of freedom
## Residual deviance: 22233  on 22900  degrees of freedom
## AIC: 22237
##
## Number of Fisher Scoring iterations: 4
```

```
uc12_2 <- glm(y ~ unclass(surgery), binomial(link='logit'), data = train_0_1_2)
summary(uc12_2)
```

```
##
## Call:
## glm(formula = y ~ unclass(surgery), family = binomial(link = "logit"),
##      data = train_0_1_2)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.7780  -0.7780  -0.5124  -0.5124   2.0471
```

```
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -0.11585    0.05124  -2.261   0.0238 *
## unclass(surgery) -0.92412    0.03550 -26.035   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##    Null deviance: 22383  on 22901  degrees of freedom
## Residual deviance: 21665  on 22900  degrees of freedom
## AIC: 21669
##
## Number of Fisher Scoring iterations: 4
```

```
uc13_2 <- glm(y ~ unclass(RX), binomial(link='logit'), data = train_0_1_2)
summary(uc13_2)
```

```
##
## Call:
## glm(formula = y ~ unclass(RX), family = binomial(link = "logit"),
##      data = train_0_1_2)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.6883  -0.6883  -0.6883  -0.4606   2.1431
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.44840    0.06353  -7.058 1.69e-12 ***
## unclass(RX) -0.87094    0.05553 -15.685 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##    Null deviance: 22383  on 22901  degrees of freedom
## Residual deviance: 22089  on 22900  degrees of freedom
## AIC: 22093
##
## Number of Fisher Scoring iterations: 4
```

```
uc14_2 <- glm(y ~ unclass(radiate), binomial(link='logit'), data = train_0_1_2)
summary(uc14_2)
```

```
##
## Call:
## glm(formula = y ~ unclass(radiate), family = binomial(link = "logit"),
##      data = train_0_1_2)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
```

```
## -0.6696 -0.6696 -0.6696 -0.5220 2.0301
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -0.83783    0.06633 -12.631  <2e-16 ***
## unclass(radiate) -0.54330    0.05889  -9.226  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 22383  on 22901  degrees of freedom
## Residual deviance: 22288  on 22900  degrees of freedom
## AIC: 22292
##
## Number of Fisher Scoring iterations: 4
```

```
uc15_2 <- glm(y ~ unclass(chem), binomial(link='logit'), data = train_0_1_2)
summary(uc15_2)
```

```
##
## Call:
## glm(formula = y ~ unclass(chem), family = binomial(link = "logit"),
##      data = train_0_1_2)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1473  -0.4597  -0.4597  -0.4597   2.1447
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   2.05173    0.05966  34.39  <2e-16 ***
## unclass(chem) -2.12292    0.03697 -57.42  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 22383  on 22901  degrees of freedom
## Residual deviance: 18939  on 22900  degrees of freedom
## AIC: 18943
##
## Number of Fisher Scoring iterations: 4
```

```
uc16_2 <- glm(y ~ unclass(CEA), binomial(link='logit'), data = train_0_1_2)
summary(uc16_2)
```

```
##
## Call:
## glm(formula = y ~ unclass(CEA), family = binomial(link = "logit"),
##      data = train_0_1_2)
##
## Deviance Residuals:
```

```
##      Min      1Q   Median      3Q      Max
## -0.7307 -0.6361 -0.6361 -0.5515  1.9791
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -2.11752    0.06746  -31.39  <2e-16 ***
## unclass(CEA)  0.31113    0.02954   10.53  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 22383  on 22901  degrees of freedom
## Residual deviance: 22271  on 22900  degrees of freedom
## AIC: 22275
##
## Number of Fisher Scoring iterations: 4
```

```
uc17_2 <- glm(y ~ unclass(bone), binomial(link='logit'), data = train_0_1_2)
summary(uc17_2)
```

```
##
## Call:
## glm(formula = y ~ unclass(bone), family = binomial(link = "logit"),
##      data = train_0_1_2)
##
## Deviance Residuals:
##      Min      1Q   Median      3Q      Max
## -0.9599 -0.6375 -0.6375 -0.6375  1.8403
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -1.96740    0.04987  -39.45  <2e-16 ***
## unclass(bone)  0.47721    0.04180   11.42  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 22383  on 22901  degrees of freedom
## Residual deviance: 22263  on 22900  degrees of freedom
## AIC: 22267
##
## Number of Fisher Scoring iterations: 4
```

```
uc18_2 <- glm(y ~ unclass(brain), binomial(link='logit'), data = train_0_1_2)
summary(uc18_2)
```

```
##
## Call:
## glm(formula = y ~ unclass(brain), family = binomial(link = "logit"),
##      data = train_0_1_2)
##
```

```
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.8918  -0.6459  -0.6459  -0.6459   1.8275
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -1.83355    0.05422  -33.82  < 2e-16 ***
## unclass(brain) 0.37226    0.04822   7.72 1.16e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 22383  on 22901  degrees of freedom
## Residual deviance: 22328  on 22900  degrees of freedom
## AIC: 22332
##
## Number of Fisher Scoring iterations: 4
```

```
uc19_2 <- glm(y ~ unclass(lung), binomial(link='logit'), data = train_0_1_2)
summary(uc19_2)
```

```
##
## Call:
## glm(formula = y ~ unclass(lung), family = binomial(link = "logit"),
##      data = train_0_1_2)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.8914  -0.6131  -0.6131  -0.6131   1.8783
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -2.00520    0.04591  -43.68  <2e-16 ***
## unclass(lung) 0.42915    0.03161   13.58  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 22383  on 22901  degrees of freedom
## Residual deviance: 22205  on 22900  degrees of freedom
## AIC: 22209
##
## Number of Fisher Scoring iterations: 4
```

0+1+2 multivariate glm

```
m2 <- glm(y ~ ., binomial(link='logit'), data = train_0_1_2)
summary(m2)
```

```

##
## Call:
## glm(formula = y ~ ., family = binomial(link = "logit"), data = train_0_1_2)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.3188  -0.5652  -0.3714  -0.2366   2.9570
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.979324   0.206341  -4.746 2.07e-06 ***
## gender2      -0.001589   0.040128  -0.040 0.968406
## race2         0.024469   0.054678   0.448 0.654512
## race3        -0.161009   0.068992  -2.334 0.019610 *
## age2          -0.139698   0.136934  -1.020 0.307640
## age3          0.241158   0.135106   1.785 0.074268 .
## size2         0.194269   0.049664   3.912 9.17e-05 ***
## size3         0.172809   0.051855   3.333 0.000861 ***
## marry2        -0.154456   0.049430  -3.125 0.001780 **
## marry3         0.049716   0.055740   0.892 0.372432
## income2       -0.075376   0.053653  -1.405 0.160059
## income3       -0.151825   0.060915  -2.492 0.012689 *
## site2         0.291654   0.041491   7.029 2.08e-12 ***
## grade2        0.170174   0.116721   1.458 0.144853
## grade3        0.853296   0.122695   6.955 3.54e-12 ***
## grade4        0.837604   0.164767   5.084 3.70e-07 ***
## grade9        0.732778   0.118251   6.197 5.76e-10 ***
## kind2         0.345567   0.068827   5.021 5.15e-07 ***
## t2            -0.206630   0.069656  -2.966 0.003013 **
## t3            -0.059734   0.065581  -0.911 0.362378
## N2            0.167738   0.059073   2.839 0.004519 **
## N3            0.221629   0.052649   4.210 2.56e-05 ***
## surgery1     -0.642876   0.063971 -10.049 < 2e-16 ***
## RX1          -0.158965   0.065524  -2.426 0.015263 *
## radiate1     -0.352094   0.071875  -4.899 9.65e-07 ***
## chem1        -2.068721   0.041160 -50.260 < 2e-16 ***
## CEA1         0.459978   0.077995   5.898 3.69e-09 ***
## CEA2         0.378222   0.082729   4.572 4.84e-06 ***
## bone1        0.627439   0.075156   8.348 < 2e-16 ***
## bone2       -0.108254   0.202943  -0.533 0.593742
## brain1       1.073278   0.164202   6.536 6.31e-11 ***
## brain2      -0.089284   0.188409  -0.474 0.635584
## lung1        0.337601   0.043880   7.694 1.43e-14 ***
## lung2       -0.012107   0.144907  -0.084 0.933413
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 22383  on 22901  degrees of freedom
## Residual deviance: 17367  on 22868  degrees of freedom
## AIC: 17435
##
## Number of Fisher Scoring iterations: 5

```



```
sqrt(rms::vif(m2))
```

```
## gender2 race2 race3 age2 age3 size2 size3 marry2
## 1.033770 1.038339 1.027432 3.125145 3.160853 1.161083 1.307303 1.280872
## marry3 income2 income3 site2 grade2 grade3 grade4 grade9
## 1.297348 1.383918 1.407103 1.061194 2.957587 2.408780 1.385841 2.973364
## kind2 t2 t3 N2 N3 surgery1 RX1 radiate1
## 1.019735 1.806471 1.659083 1.226631 1.135977 1.607891 1.060217 1.071951
## chem1 CEA1 CEA2 bone1 bone2 brain1 brain2 lung1
## 1.062254 1.921048 1.914153 1.039511 1.570650 1.038896 1.553694 1.045949
## lung2
## 1.149740
```

0+1 multivariate glm anova

```
anova(m2, test = 'Chisq')
```

```
## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: y
##
## Terms added sequentially (first to last)
##
##
##      Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
## NULL                                22901      22382
## gender  1      14.01      22900      22368 0.0001822 ***
## race    2      17.44      22898      22351 0.0001631 ***
## age     2     489.57      22896      21862 < 2.2e-16 ***
## size    2     346.05      22894      21516 < 2.2e-16 ***
## marry   2     142.34      22892      21373 < 2.2e-16 ***
## income  2      13.89      22890      21359 0.0009657 ***
## site    1      85.66      22889      21274 < 2.2e-16 ***
## grade   4     413.65      22885      20860 < 2.2e-16 ***
## kind    1      24.18      22884      20836 8.773e-07 ***
## t       2     150.63      22882      20685 < 2.2e-16 ***
## N       2      48.02      22880      20637 3.732e-11 ***
## surgery 1     162.53      22879      20475 < 2.2e-16 ***
## RX      1      33.90      22878      20441 5.794e-09 ***
## radiate 1      51.14      22877      20390 8.610e-13 ***
## chem    1    2770.94      22876      17619 < 2.2e-16 ***
## CEA     2      48.69      22874      17570 2.669e-11 ***
## bone    2      91.80      22872      17478 < 2.2e-16 ***
## brain   2      51.44      22870      17427 6.774e-12 ***
## lung    2      59.64      22868      17367 1.122e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

0+1+2 multivariate lrm

```
m2_lrm <- lrm(y ~ ., data = train_0_1_2, x=T, y=T)
print(m2_lrm, digits=3)
```

```
## Logistic Regression Model
```

```
##
```

```
## lrm(formula = y ~ ., data = train_0_1_2, x = T, y = T)
```

```
##
```

		Model Likelihood		Discrimination		Rank Discrim.	
		Ratio Test		Indexes		Indexes	
## Obs	22902	LR chi2	5015.51	R2	0.315	C	0.816
## 0	18512	d.f.	33	R2(33,22902)	0.196	Dxy	0.633
## 1	4390	Pr(> chi2)	<0.0001	R2(33,10645.5)	0.374	gamma	0.633
## max deriv	2e-08			Brier	0.118	tau-a	0.196

```
##
```

	Coef	S.E.	Wald Z	Pr(> Z)
## Intercept	-0.979	0.206	-4.75	<0.0001
## gender=2	-0.002	0.040	-0.04	0.9684
## race=2	0.024	0.055	0.45	0.6545
## race=3	-0.161	0.069	-2.33	0.0196
## age=2	-0.140	0.137	-1.02	0.3076
## age=3	0.241	0.135	1.78	0.0743
## size=2	0.194	0.050	3.91	<0.0001
## size=3	0.173	0.052	3.33	0.0009
## marry=2	-0.154	0.049	-3.12	0.0018
## marry=3	0.050	0.056	0.89	0.3724
## income=2	-0.075	0.054	-1.40	0.1601
## income=3	-0.152	0.061	-2.49	0.0127
## site=2	0.292	0.041	7.03	<0.0001
## grade=2	0.170	0.117	1.46	0.1449
## grade=3	0.853	0.123	6.95	<0.0001
## grade=4	0.838	0.165	5.08	<0.0001
## grade=9	0.733	0.118	6.20	<0.0001
## kind=2	0.346	0.069	5.02	<0.0001
## t=2	-0.207	0.070	-2.97	0.0030
## t=3	-0.060	0.066	-0.91	0.3624
## N=2	0.168	0.059	2.84	0.0045
## N=3	0.222	0.053	4.21	<0.0001
## surgery=1	-0.643	0.064	-10.05	<0.0001
## RX=1	-0.159	0.066	-2.43	0.0153
## radiate=1	-0.352	0.072	-4.90	<0.0001
## chem=1	-2.069	0.041	-50.26	<0.0001
## CEA=1	0.460	0.078	5.90	<0.0001
## CEA=2	0.378	0.083	4.57	<0.0001
## bone=1	0.627	0.075	8.35	<0.0001
## bone=2	-0.108	0.203	-0.53	0.5937
## brain=1	1.073	0.164	6.54	<0.0001
## brain=2	-0.089	0.188	-0.47	0.6356
## lung=1	0.338	0.044	7.69	<0.0001
## lung=2	-0.012	0.145	-0.08	0.9334

```
sqrt(rms::vif(m2_lrm))
```

```
## gender=2    race=2    race=3    age=2    age=3    size=2    size=3    marry=2
## 1.033770 1.038339 1.027432 3.125150 3.160858 1.161084 1.307303 1.280872
## marry=3    income=2    income=3    site=2    grade=2    grade=3    grade=4    grade=9
## 1.297348 1.383918 1.407103 1.061194 2.957588 2.408784 1.385843 2.973368
## kind=2      t=2      t=3      N=2      N=3    surgery=1      RX=1    radiate=1
## 1.019735 1.806471 1.659082 1.226633 1.135977 1.607891 1.060215 1.071951
## chem=1      CEA=1      CEA=2    bone=1    bone=2    brain=1    brain=2    lung=1
## 1.062254 1.921060 1.914164 1.039511 1.570650 1.038896 1.553694 1.045949
## lung=2
## 1.149740
```

0+1+2 ridge

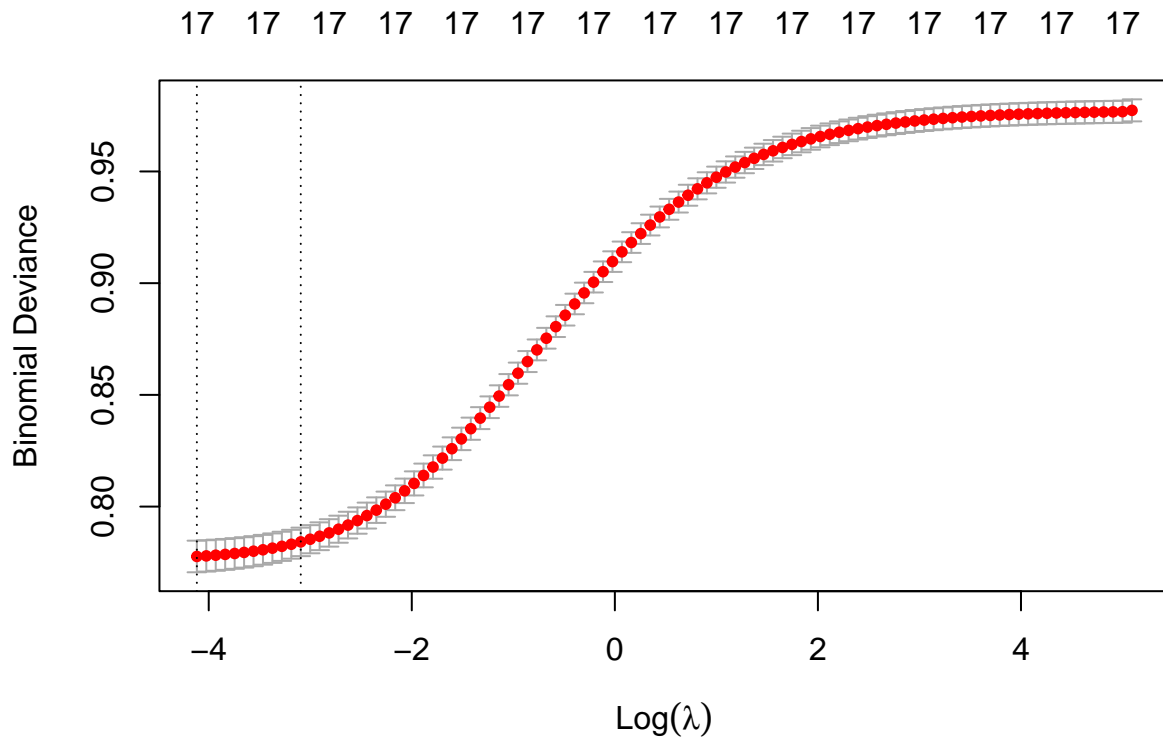
```
x <- data.matrix(train_0_1_2[, c('race','size','marry','income','site','grade','kind','t','N','surgery')])
y <- as.numeric(data.matrix(train_0_1_2$y))
```

```
#perform k-fold cross-validation to find optimal lambda value
set.seed(2023)
cv_model_rdg_2 <- cv.glmnet(x, y, alpha = 0, family = "binomial")

#find optimal lambda value that minimizes test MSE
best_lambda_rdg_2 <- cv_model_rdg_2$lambda.min
best_lambda_rdg_2
```

```
## [1] 0.01629584
```

```
#produce plot of test MSE by lambda value
plot(cv_model_rdg_2)
```



```
#find coefficients of best model
```

```
best_model_rdg_2 <- glmnet(x, y, alpha = 0, lambda = best_lambda_rdg_2, family = "binomial")
coef(best_model_rdg_2)
```

```
## 18 x 1 sparse Matrix of class "dgCMatrix"
##              s0
## (Intercept)  0.77150706
## race        -0.04169564
## size         0.10342592
## marry        0.06085393
## income      -0.07588702
## site         0.32108732
## grade        0.17268363
## kind         0.33127197
## t            0.01895845
## N            0.09864970
## surgery     -0.56870801
## RX          -0.21635031
## radiate     -0.20443955
## chem        -1.91281634
## CEA         0.08060105
## bone        0.22963553
## brain      -0.14121426
## lung        0.21995725
```

```
print(best_model_rdg_2)
```

```
##
## Call:  glmnet(x = x, y = y, family = "binomial", alpha = 0, lambda = best_lambda_rdg_2)
```

```
##
##   Df   %Dev Lambda
## 1 17 20.59 0.0163
```

0+1+2 multivariate Nomogram

```
m2_lrm_final <- lrm(y~race+size+marry+income+site+grade+kind+t+N+surgery+RX+radiate+chem+CEA+bone+brain)
print(m2_lrm_final, digits=3)
```

```
## Logistic Regression Model
```

```
##
## lrm(formula = y ~ race + size + marry + income + site + grade +
##      kind + t + N + surgery + RX + radiate + chem + CEA + bone +
##      brain + lung, data = train_0_1_2, x = T, y = T)
##
```

		Model Likelihood		Discrimination	Rank Discrim.
		Ratio Test		Indexes	Indexes
## Obs	22902	LR chi2	4944.05	R2	0.311
## 0	18512	d.f.	30	R2(30,22902)	0.193
## 1	4390	Pr(> chi2)	<0.0001	R2(30,10645.5)	0.370
## max deriv	1e-08			Brier	0.118
				tau-a	0.195

```
##
##      Coef   S.E. Wald Z Pr(>|Z|)
## Intercept -0.879 0.159 -5.52 <0.0001
## race=2     0.006 0.054  0.11 0.9110
## race=3    -0.169 0.069 -2.45 0.0143
## size=2     0.187 0.050  3.78 0.0002
## size=3     0.169 0.052  3.26 0.0011
## marry=2    -0.112 0.049 -2.30 0.0215
## marry=3     0.130 0.054  2.41 0.0161
## income=2   -0.079 0.054 -1.48 0.1394
## income=3   -0.147 0.061 -2.41 0.0157
## site=2     0.336 0.041  8.20 <0.0001
## grade=2    0.175 0.117  1.50 0.1325
## grade=3    0.849 0.123  6.92 <0.0001
## grade=4    0.838 0.165  5.09 <0.0001
## grade=9    0.740 0.118  6.26 <0.0001
## kind=2     0.343 0.069  4.99 <0.0001
## t=2        -0.202 0.069 -2.91 0.0036
## t=3        -0.052 0.065 -0.80 0.4254
## N=2         0.152 0.059  2.58 0.0098
## N=3         0.214 0.053  4.07 <0.0001
## surgery=1  -0.646 0.064 -10.13 <0.0001
## RX=1       -0.188 0.065 -2.88 0.0040
## radiate=1  -0.363 0.072 -5.07 <0.0001
## chem=1     -2.127 0.041 -52.36 <0.0001
## CEA=1       0.466 0.078  5.98 <0.0001
## CEA=2       0.393 0.083  4.76 <0.0001
## bone=1      0.621 0.075  8.28 <0.0001
## bone=2     -0.122 0.202 -0.60 0.5476
## brain=1     1.059 0.163  6.48 <0.0001
## brain=2    -0.076 0.188 -0.40 0.6858
```

```
## lung=1      0.342 0.044   7.81 <0.0001
## lung=2     -0.003 0.145  -0.02 0.9821
```

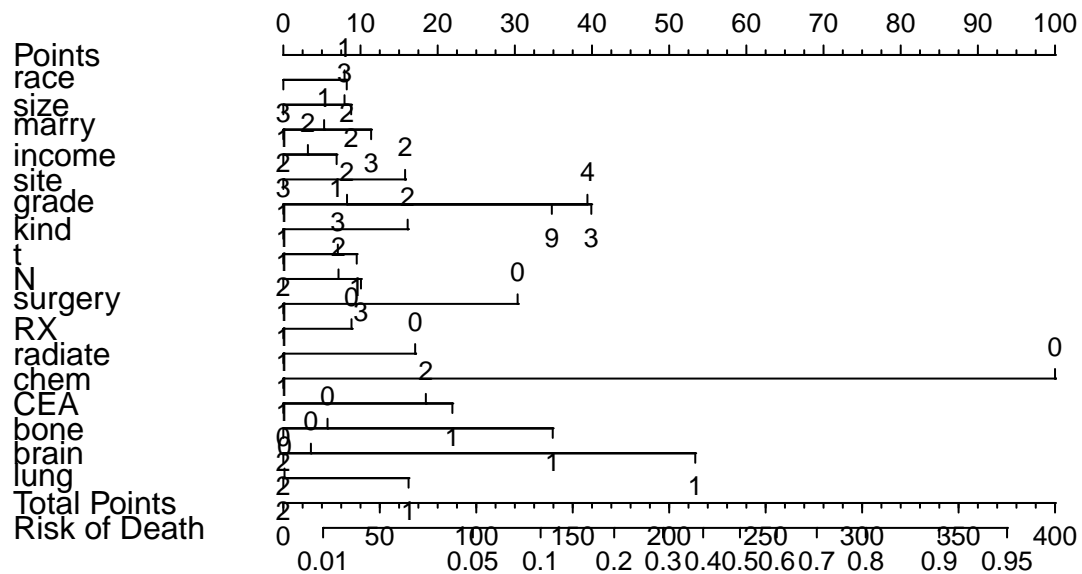
```
sqrt(rms::vif(m2_lrm_final))
```

```
##      race=2    race=3    size=2    size=3    marry=2    marry=3    income=2    income=3
## 1.037101 1.026930 1.159550 1.305122 1.268839 1.254502 1.383744 1.407164
##      site=2    grade=2    grade=3    grade=4    grade=9    kind=2      t=2      t=3
## 1.048815 2.960700 2.411862 1.386767 2.977385 1.019394 1.805169 1.658181
##      N=2      N=3    surgery=1      RX=1    radiate=1      chem=1      CEA=1      CEA=2
## 1.225244 1.135459 1.604990 1.057937 1.072228 1.050166 1.920347 1.913085
##      bone=1    bone=2    brain=1    brain=2    lung=1    lung=2
## 1.039625 1.568233 1.039064 1.550070 1.045776 1.149547
```

```
ddist <- datadist(train_0_1_2)
```

```
options(datadist='ddist')
```

```
nom2 <- nomogram(m2_lrm_final, fun=function(x)1/(1+exp(-x)),fun.at=c(.001, .01, .05, seq(.1,.9, by=.1),
plot(nom2)
```



0+1+2 multivariate calibrate

```
m2_final <- glm(y~race+size+marry+income+site+grade+kind+t+N+surgery+RX+radiate+chem+CEA+bone+brain+lun,
summary(m2_final)
```

```
##
## Call:
## glm(formula = y ~ race + size + marry + income + site + grade +
##      kind + t + N + surgery + RX + radiate + chem + CEA + bone +
##      brain + lung, family = binomial(link = "logit"), data = train_0_1_2)
##
```

```
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.3006  -0.5701  -0.3758  -0.2452   2.8859
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.878985   0.159336  -5.517 3.46e-08 ***
## race2        0.006089   0.054490   0.112 0.911020
## race3       -0.168672   0.068875  -2.449 0.014328 *
## size2        0.186981   0.049510   3.777 0.000159 ***
## size3        0.168558   0.051675   3.262 0.001107 **
## marry2       -0.112361   0.048874  -2.299 0.021504 *
## marry3        0.129538   0.053823   2.407 0.016096 *
## income2      -0.079145   0.053548  -1.478 0.139403
## income3      -0.146832   0.060805  -2.415 0.015744 *
## site2        0.335556   0.040940   8.196 2.48e-16 ***
## grade2       0.175470   0.116640   1.504 0.132485
## grade3       0.848744   0.122604   6.923 4.43e-12 ***
## grade4       0.837999   0.164506   5.094 3.51e-07 ***
## grade9       0.739894   0.118169   6.261 3.82e-10 ***
## kind2        0.342642   0.068681   4.989 6.07e-07 ***
## t2          -0.202253   0.069474  -2.911 0.003601 **
## t3          -0.052149   0.065420  -0.797 0.425369
## N2           0.152024   0.058894   2.581 0.009843 **
## N3           0.213913   0.052527   4.072 4.65e-05 ***
## surgery1    -0.645605   0.063741 -10.129 < 2e-16 ***
## RX1         -0.187806   0.065203  -2.880 0.003973 **
## radiate1    -0.363282   0.071705  -5.066 4.05e-07 ***
## chem1       -2.126724   0.040616 -52.362 < 2e-16 ***
## CEA1         0.465840   0.077837   5.985 2.17e-09 ***
## CEA2         0.392722   0.082556   4.757 1.96e-06 ***
## bone1        0.620662   0.074928   8.283 < 2e-16 ***
## bone2       -0.121764   0.202469  -0.601 0.547578
## brain1       1.059270   0.163427   6.482 9.07e-11 ***
## brain2      -0.075954   0.187756  -0.405 0.685818
## lung1        0.341889   0.043792   7.807 5.85e-15 ***
## lung2       -0.003252   0.144640  -0.022 0.982063
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 22383  on 22901  degrees of freedom
## Residual deviance: 17438  on 22871  degrees of freedom
## AIC: 17500
##
## Number of Fisher Scoring iterations: 5
```

```
sqrt(rms::vif(m2_final))
```

```
##      race2      race3      size2      size3      marry2      marry3      income2      income3
## 1.037101 1.026930 1.159550 1.305122 1.268839 1.254502 1.383744 1.407164
##      site2      grade2      grade3      grade4      grade9      kind2      t2      t3
## 1.048815 2.960699 2.411858 1.386765 2.977382 1.019394 1.805169 1.658182
```

```
##      N2      N3 surgery1      RX1 radiate1      chem1      CEA1      CEA2
## 1.225242 1.135459 1.604989 1.057938 1.072228 1.050166 1.920337 1.913075
##      bone1      bone2      brain1      brain2      lung1      lung2
## 1.039624 1.568233 1.039063 1.550070 1.045776 1.149547
```

test model

```
m2_final_test <- glm(y~race+size+marry+income+site+grade+kind+t+N+surgery+RX+radiate+chem+CEA+bone+brain
```

train

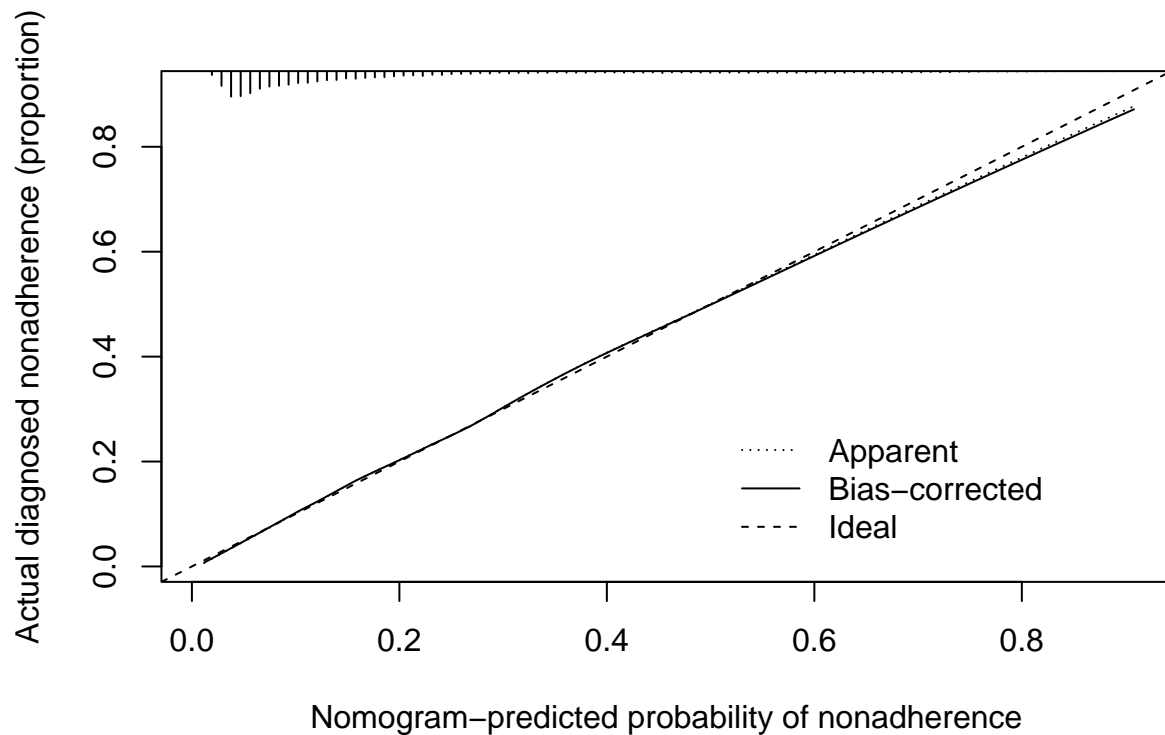
```
# hosmer-lemeshow
```

```
p.hoslem <- hoslem.test(m2_final$y, fitted(m2_final), g=10)$p.value
p.hoslem
```

```
## [1] 0.4653902
```

```
# plot 1
```

```
refit <- lrm(y~race+size+marry+income+site+grade+kind+t+N+surgery+RX+radiate+chem+CEA+bone+brain+lung, c
cal <- calibrate(refit, cmethod='hare', method='boot', B=1000, data=train_0_1_2)
plot(cal, xlab="Nomogram-predicted probability of nonadherence", ylab="Actual diagnosed nonadherence (p
```



```
##
## n=22902 Mean absolute error=0.004 Mean squared error=3e-05
## 0.9 Quantile of absolute error=0.007
```



```
# plot 2
```

```
plot(cal, xlim = c(0,1), ylim = c(0,1), xlab = "Prediced Probability", ylab = "Observed Probability", col = "black", lty = 1, lwd = 1)
```

```
##
```

```
## n=22902 Mean absolute error=0.004 Mean squared error=3e-05
```

```
## 0.9 Quantile of absolute error=0.007
```

```
lines(cal[,c("predy","calibrated.corrected")], type = 'l', lwd = 3, pch = 16, col = "#2166AC")
```

```
lines(cal[,c("predy","calibrated.orig")], type="l", pch=16, lwd=3, col="tomato")
```

```
abline(0, 1, lty = 2, lwd = 2, col = "#224444")
```

```
legend(0.6,0.2,
```

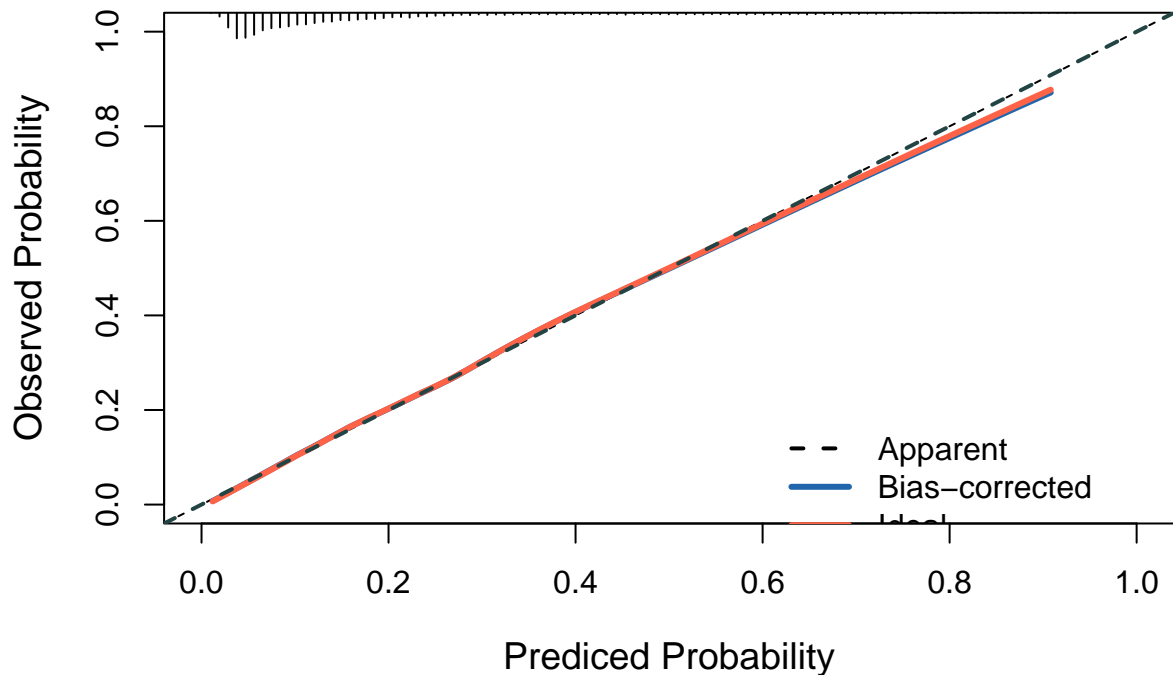
```
  c("Apparent","Bias-corrected","Ideal"),
```

```
  lty = c(2,1,1),
```

```
  lwd = c(2,3,3),
```

```
  col = c("black","#2166AC","tomato"),
```

```
  bty = "n")
```



B= 1000 repetitions, boot

Mean absolute error=0.004 n=22902

```
# plot 3
```

```
plot(cal, xlim = c(0,1), ylim = c(0,1), xlab = "Prediced Probability", ylab = "Observed Probability", col = "black", lty = 1, lwd = 1)
```

```
##
```

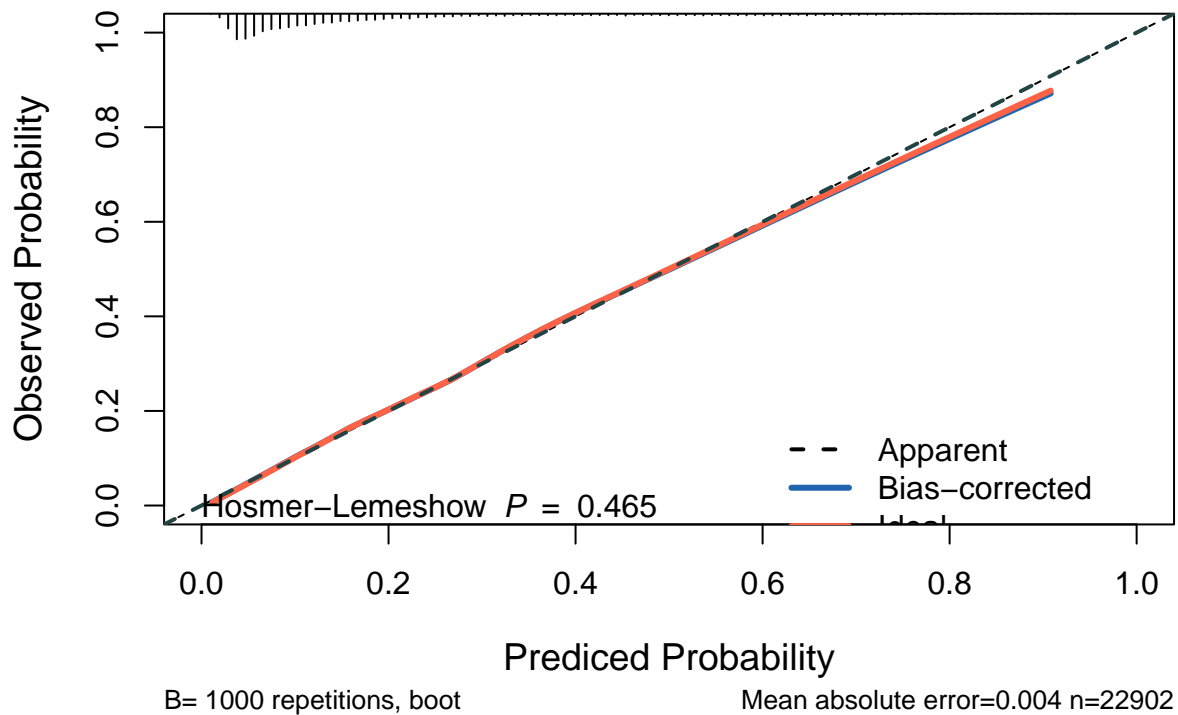
```
## n=22902 Mean absolute error=0.004 Mean squared error=3e-05
```

```
## 0.9 Quantile of absolute error=0.007
```

```
lines(cal[,c("predy","calibrated.corrected")], type = 'l', lwd = 3, pch = 16, col = "#2166AC")
```

```
lines(cal[,c("predy","calibrated.orig")], type="l", pch=16, lwd=3, col="tomato")
```

```
abline(0, 1, lty = 2, lwd = 2, col = "#224444")
legend(0.6,0.2,
      c("Apparent","Bias-corrected","Ideal"),
      lty = c(2,1,1),
      lwd = c(2,3,3),
      col = c("black","#2166AC","tomato"),
      bty = "n")
text(0,0,bquote("Hosmer-Lemeshow ~italic(P)~" = "~.(round(p.hoslem,3))),adj = 0)
```



test

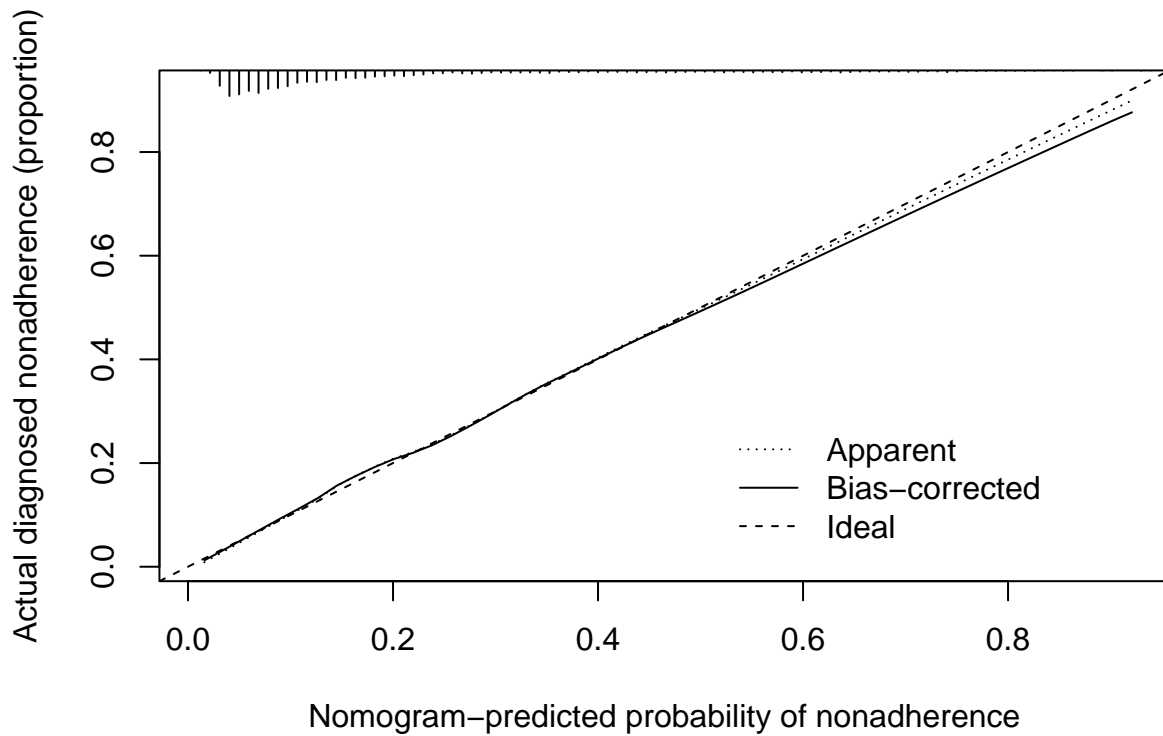
```
# hosmer-lemeshow
```

```
p.hoslem <- hoslem.test(m2_final_test$y, fitted(m2_final_test), g=10)$p.value
p.hoslem
```

```
## [1] 0.2301717
```

```
# plot 1
```

```
refit <- lrm(y~race+size+marry+income+site+grade+kind+t+N+surgery+RX+radiate+chem+CEA+bone+brain+lung,
cal <- calibrate(refit, cmethod='hare', method='boot', B=1000, data=test_0_1_2)
plot(cal, xlab="Nomogram-predicted probability of nonadherence", ylab="Actual diagnosed nonadherence (p
```



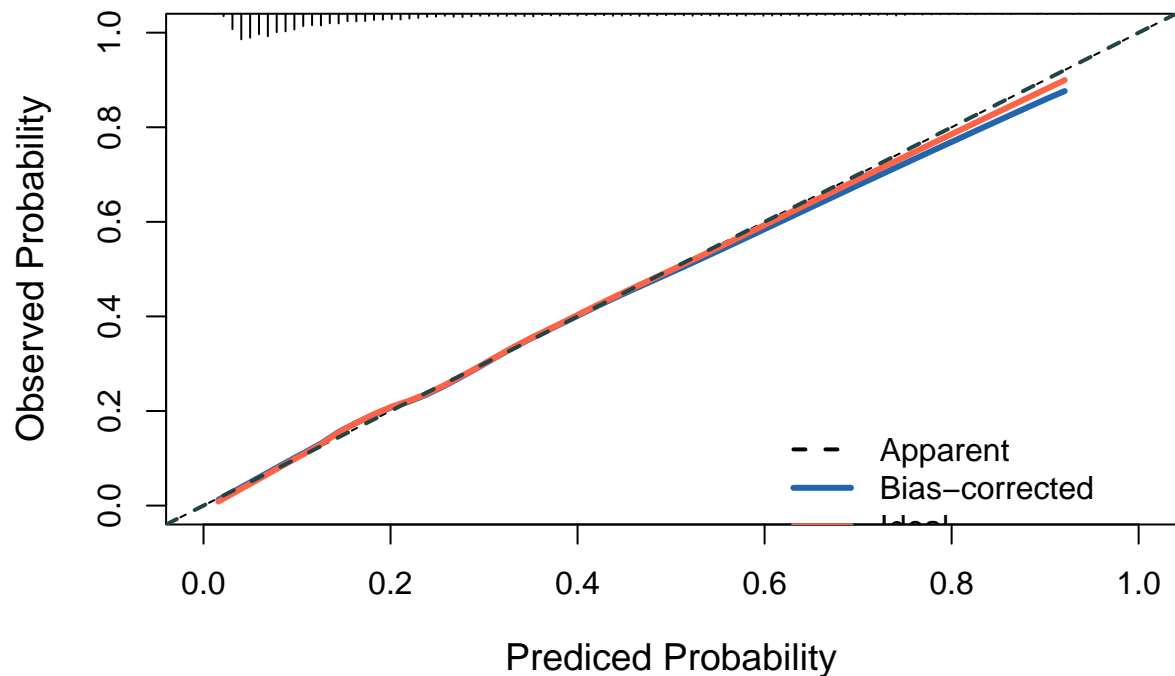
```
##
## n=5726   Mean absolute error=0.005   Mean squared error=7e-05
## 0.9 Quantile of absolute error=0.011
```

```
# plot 2
```

```
plot(cal, xlim = c(0,1), ylim = c(0,1), xlab = "Prediced Probability", ylab = "Observed Probability", col = "black", lty = 1, lwd = 3)
```

```
##
## n=5726   Mean absolute error=0.005   Mean squared error=7e-05
## 0.9 Quantile of absolute error=0.011
```

```
lines(cal[,c("predy","calibrated.corrected")], type = 'l', lwd = 3, pch = 16, col = "#2166AC")
lines(cal[,c("predy","calibrated.orig")], type="l", pch=16, lwd=3, col="tomato")
abline(0, 1, lty = 2, lwd = 2, col = "#224444")
legend(0.6,0.2,
      c("Apparent","Bias-corrected","Ideal"),
      lty = c(2,1,1),
      lwd = c(2,3,3),
      col = c("black","#2166AC","tomato"),
      bty = "n")
```



B= 1000 repetitions, boot

Mean absolute error=0.005 n=5726

```
# plot 3
```

```
plot(cal, xlim = c(0,1), ylim = c(0,1), xlab = "Predicted Probability", ylab = "Observed Probability", c
```

```
##
```

```
## n=5726 Mean absolute error=0.005 Mean squared error=7e-05
```

```
## 0.9 Quantile of absolute error=0.011
```

```
lines(cal[,c("predy","calibrated.corrected")], type = 'l', lwd = 3, pch = 16, col = "#2166AC")
```

```
lines(cal[,c("predy","calibrated.orig")], type="l", pch=16, lwd=3, col="tomato")
```

```
abline(0, 1, lty = 2, lwd = 2, col = "#224444")
```

```
legend(0.6,0.2,
```

```
  c("Apparent","Bias-corrected","Ideal"),
```

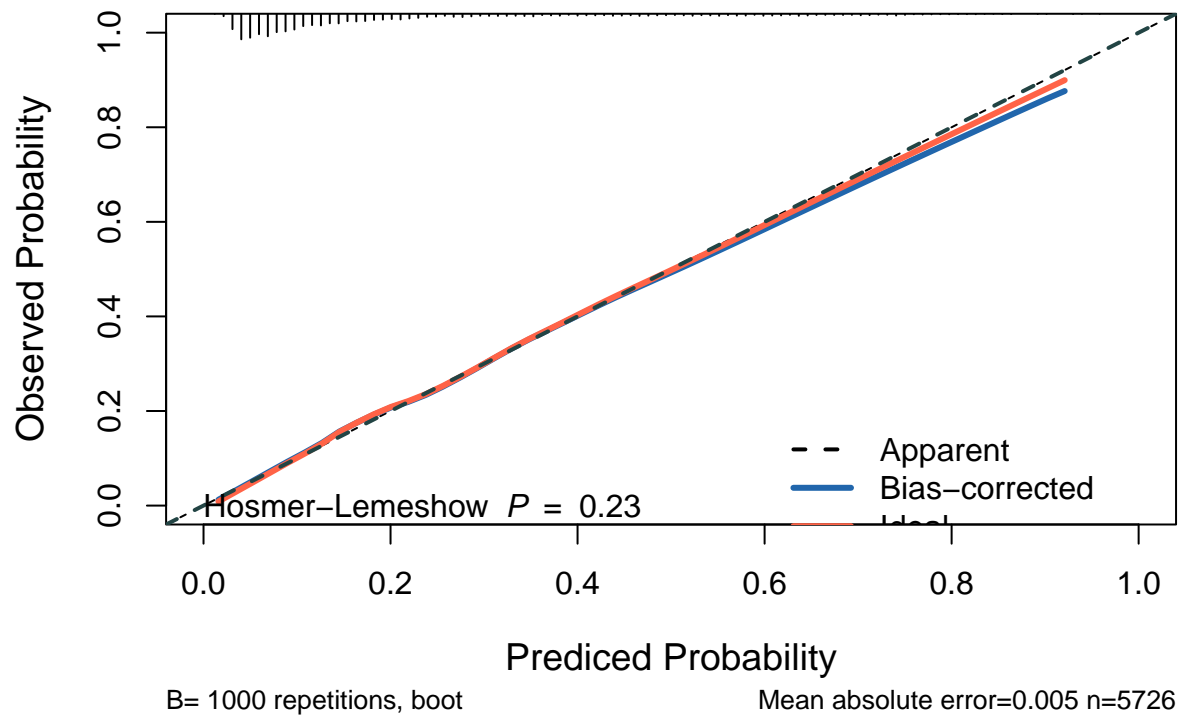
```
  lty = c(2,1,1),
```

```
  lwd = c(2,3,3),
```

```
  col = c("black","#2166AC","tomato"),
```

```
  bty = "n")
```

```
text(0,0,bquote("Hosmer-Lemeshow ~italic(P)~" = "~.(round(p.hoslem,3))),adj = 0)
```



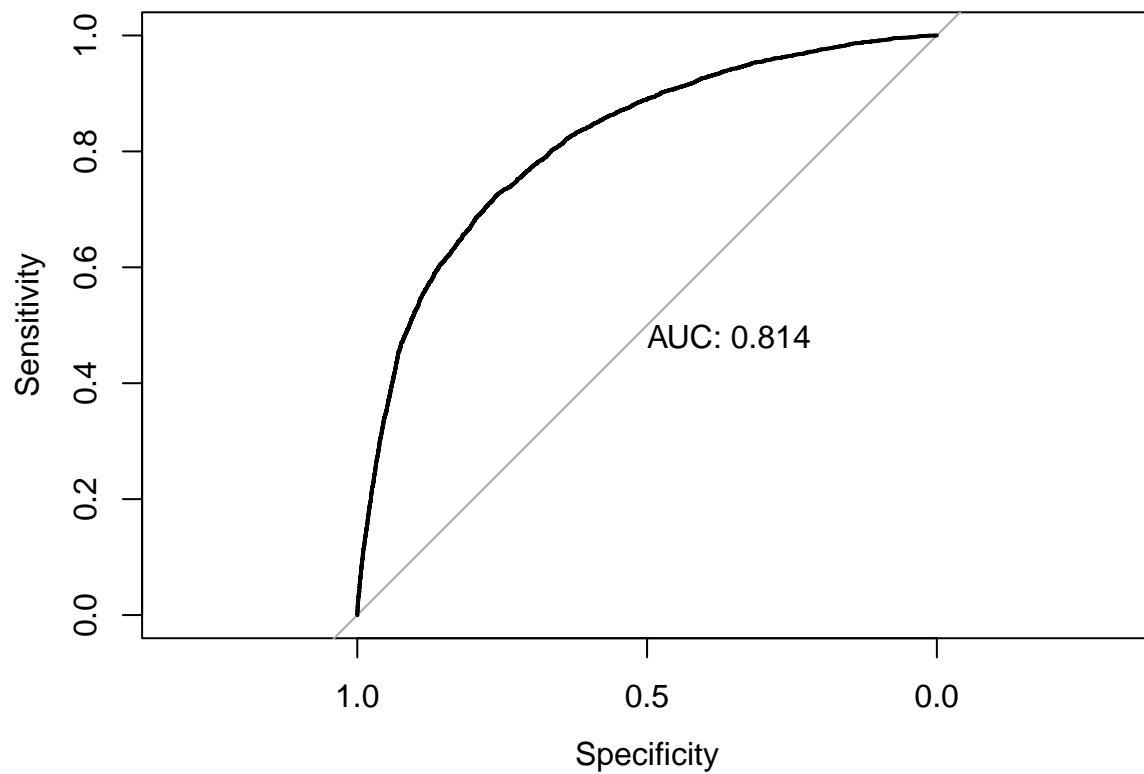
0+1+2 multivariate ROC

train

```
train_prob = predict(m2_final, newdata = train_0_1_2, type = "response")
train_roc = roc(train_0_1_2$y ~ train_prob, plot = TRUE, print.auc = TRUE)
```

```
## Setting levels: control = 0, case = 1
```

```
## Setting direction: controls < cases
```



```
as.numeric(train_roc$auc)
```

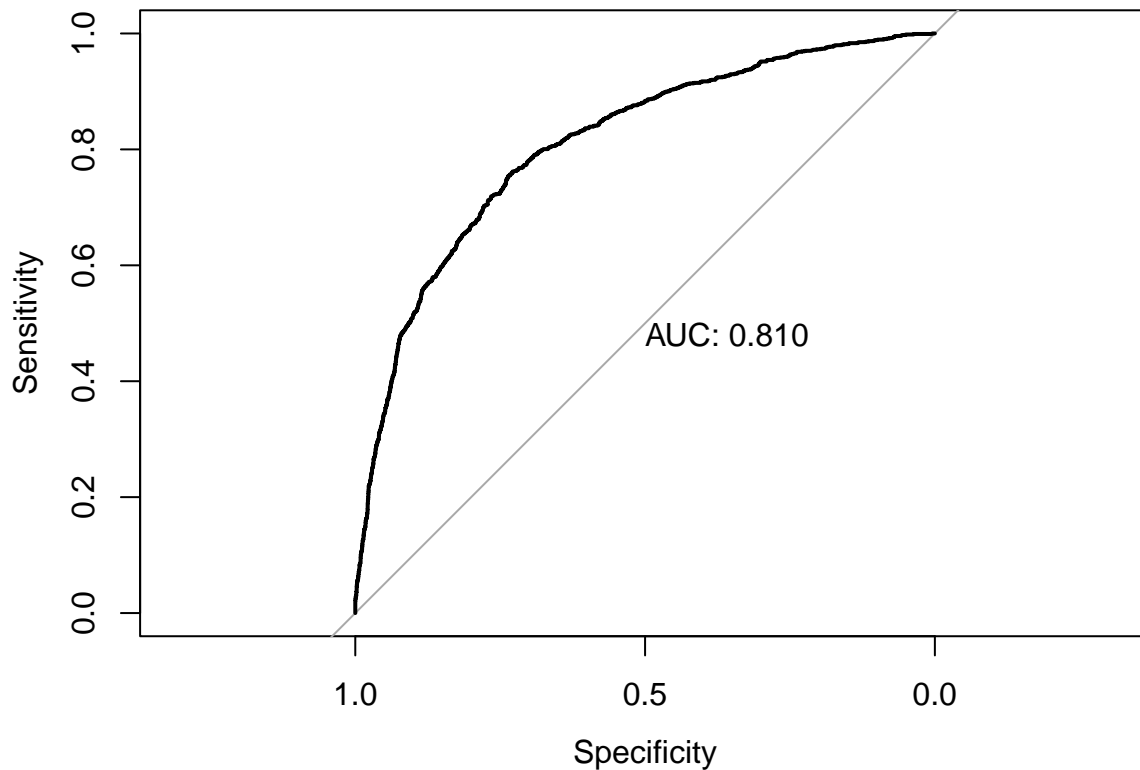
```
## [1] 0.8139692
```

```
test
```

```
test_prob = predict(m2_final, newdata = test_0_1_2, type = "response")  
test_roc = roc(test_0_1_2$y ~ test_prob, plot = TRUE, print.auc = TRUE)
```

```
## Setting levels: control = 0, case = 1
```

```
## Setting direction: controls < cases
```



```
as.numeric(test_roc$auc)
```

```
## [1] 0.8097144
```

0+1+2 multivariate DCA

train

```
train_0_1_2$y <- as.numeric(levels(train_0_1_2$y))[train_0_1_2$y]
str(train_0_1_2)
```

```
## tibble [22,902 x 20] (S3: tbl_df/tbl/data.frame)
## $ y      : num [1:22902] 0 0 1 0 0 0 0 0 1 0 ...
## $ gender : Factor w/ 2 levels "1","2": 2 1 2 1 1 2 2 1 2 1 ...
## $ race   : Factor w/ 3 levels "1","2","3": 2 1 1 1 1 1 1 1 1 3 ...
## $ age    : Factor w/ 3 levels "1","2","3": 2 2 3 3 2 3 1 2 3 3 ...
## $ size   : Factor w/ 3 levels "1","2","3": 1 3 3 2 3 2 1 2 3 3 ...
## $ marry  : Factor w/ 3 levels "1","2","3": 2 2 2 1 2 3 2 3 2 2 ...
## $ income : Factor w/ 3 levels "1","2","3": 2 2 3 3 3 2 2 3 3 2 ...
## $ site   : Factor w/ 2 levels "1","2": 1 2 2 1 1 1 1 2 1 2 ...
## $ grade  : Factor w/ 5 levels "1","2","3","4",...: 5 5 5 5 2 1 2 2 5 2 ...
## $ kind   : Factor w/ 2 levels "1","2": 1 1 1 1 1 1 1 1 2 1 ...
## $ t      : Factor w/ 3 levels "1","2","3": 2 2 3 2 2 1 2 3 2 3 ...
## $ N      : Factor w/ 3 levels "1","2","3": 1 1 3 2 1 1 2 1 1 3 ...
## $ surgery: Factor w/ 2 levels "0","1": 1 1 1 1 1 2 2 1 1 1 ...
## $ RX     : Factor w/ 2 levels "0","1": 2 1 1 1 1 1 1 1 1 1 ...
```

```
## $ radiate: Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
## $ chem : Factor w/ 2 levels "0","1": 2 1 2 2 2 1 2 2 1 1 ...
## $ CEA : Factor w/ 3 levels "0","1","2": 2 1 2 2 2 2 3 2 2 3 ...
## $ bone : Factor w/ 3 levels "0","1","2": 1 1 1 1 1 1 1 1 1 1 ...
## $ brain : Factor w/ 3 levels "0","1","2": 1 1 1 1 1 1 1 1 1 1 ...
## $ lung : Factor w/ 3 levels "0","1","2": 1 2 1 1 2 2 1 1 2 1 ...
```

```
model_1 <- decision_curve(y~race+size+marry+income+site+grade+kind+t+N+surgery+RX+radiate+chem+CEA+bone
```

```
## Calculating net benefit curves for case-control data. All calculations are done conditional on the o
```

```
## Note: The data provided is used to both fit a prediction model and to estimate the respective decis
```

```
summary(model_1)
```

```
##
```

```
## Standardized Net Benefit (95% Confidence Intervals):
```

```
##
```

risk threshold	cost:benefit ratio	percent high risk	All	y ~ race + size + marry + income + site + grade + kind + t + N + surgery + RX + radiate + chem + CEA + bone + brain + lung
-------------------	-----------------------	----------------------	-----	--

```
##
```

0	0:1	100 (100, 100)	1 (1, 1)	1 (1, 1)
---	-----	-------------------	-------------	-------------

0.01	1:99	100 (100, 100)	0.995 (0.995, 0.995)	0.995 (0.995, 0.995)
------	------	-------------------	-------------------------	-------------------------

0.02	1:49	100 (100, 100)	0.991 (0.991, 0.991)	0.991 (0.991, 0.991)
------	------	-------------------	-------------------------	-------------------------

0.03	3:97	100 (100, 100)	0.986 (0.986, 0.986)	0.986 (0.986, 0.986)
------	------	-------------------	-------------------------	-------------------------

0.04	1:24	100 (100, 100)	0.981 (0.981, 0.981)	0.981 (0.981, 0.981)
------	------	-------------------	-------------------------	-------------------------

0.05	1:19	100 (100, 100)	0.976 (0.976, 0.976)	0.976 (0.976, 0.976)
------	------	-------------------	-------------------------	-------------------------

0.06	3:47	100 (100, 100)	0.971 (0.971, 0.971)	0.971 (0.971, 0.971)
------	------	-------------------	-------------------------	-------------------------

0.07	7:93	100 (99.997, 100)	0.966 (0.966, 0.966)	0.966 (0.966, 0.966)
------	------	----------------------	-------------------------	-------------------------

0.08	2:23	100 (99.99, 100)	0.96 (0.96, 0.96)	0.96 (0.96, 0.96)
------	------	---------------------	----------------------	----------------------

0.09	9:91	99.998	0.955	0.955
------	------	--------	-------	-------

##			(99.956, 100)	(0.955, 0.955)	(0.955, 0.955)
##					
##	0.1	1:9	99.995	0.949	0.949
##			(99.91, 100)	(0.949, 0.949)	(0.949, 0.95)
##					
##	0.11	11:89	99.97	0.944	0.944
##			(99.84, 99.998)	(0.944, 0.944)	(0.944, 0.944)
##					
##	0.12	3:22	99.917	0.938	0.938
##			(99.749, 99.988)	(0.938, 0.938)	(0.938, 0.938)
##					
##	0.13	13:87	99.861	0.932	0.932
##			(99.627, 99.963)	(0.932, 0.932)	(0.932, 0.933)
##					
##	0.14	7:43	99.761	0.926	0.926
##			(99.458, 99.919)	(0.926, 0.926)	(0.926, 0.927)
##					
##	0.15	3:17	99.661	0.92	0.92
##			(99.227, 99.835)	(0.92, 0.92)	(0.919, 0.921)
##					
##	0.16	4:21	99.466	0.913	0.914
##			(98.972, 99.734)	(0.913, 0.913)	(0.913, 0.915)
##					
##	0.17	17:83	99.234	0.907	0.908
##			(98.697, 99.581)	(0.907, 0.907)	(0.907, 0.909)
##					
##	0.18	9:41	99.004	0.9	0.902
##			(98.329, 99.372)	(0.9, 0.9)	(0.9, 0.903)
##					
##	0.19	19:81	98.697	0.893	0.895
##			(97.928, 99.135)	(0.893, 0.893)	(0.893, 0.897)
##					
##	0.2	1:4	98.279	0.886	0.888
##			(97.483, 98.804)	(0.886, 0.886)	(0.887, 0.89)
##					
##	0.21	21:79	97.867	0.879	0.882
##			(96.95, 98.458)	(0.879, 0.879)	(0.88, 0.884)
##					
##	0.22	11:39	97.442	0.872	0.876
##			(96.305, 98.077)	(0.872, 0.872)	(0.872, 0.878)
##					
##	0.23	23:77	96.777	0.864	0.868
##			(95.663, 97.593)	(0.864, 0.864)	(0.866, 0.872)
##					
##	0.24	6:19	96.064	0.856	0.861
##			(94.964, 97.015)	(0.856, 0.856)	(0.858, 0.866)
##					
##	0.25	1:3	95.422	0.848	0.855
##			(94.264, 96.395)	(0.848, 0.848)	(0.851, 0.86)
##					
##	0.26	13:37	94.778	0.84	0.849
##			(93.538, 95.758)	(0.84, 0.84)	(0.844, 0.853)
##					
##	0.27	27:73	94.029	0.832	0.841

##			(92.687, 95.021)	(0.832, 0.832)	(0.837, 0.846)
##					
##	0.28	7:18	93.172	0.823	0.834
##			(91.929, 94.242)	(0.823, 0.823)	(0.829, 0.839)
##					
##	0.29	29:71	92.311	0.814	0.826
##			(91.189, 93.554)	(0.814, 0.814)	(0.822, 0.832)
##					
##	0.3	3:7	91.639	0.805	0.819
##			(90.422, 92.699)	(0.805, 0.805)	(0.814, 0.825)
##					
##	0.31	31:69	90.78	0.795	0.811
##			(89.64, 91.777)	(0.795, 0.795)	(0.807, 0.818)
##					
##	0.32	8:17	89.861	0.786	0.804
##			(88.865, 90.999)	(0.786, 0.786)	(0.799, 0.812)
##					
##	0.33	33:67	89.12	0.776	0.798
##			(88.082, 90.206)	(0.776, 0.776)	(0.793, 0.805)
##					
##	0.34	17:33	88.391	0.765	0.792
##			(87.372, 89.274)	(0.765, 0.765)	(0.785, 0.798)
##					
##	0.35	7:13	87.633	0.755	0.784
##			(86.678, 88.508)	(0.755, 0.755)	(0.777, 0.791)
##					
##	0.36	9:16	86.868	0.744	0.778
##			(85.965, 87.686)	(0.744, 0.744)	(0.77, 0.784)
##					
##	0.37	37:63	86.098	0.732	0.769
##			(85.197, 86.852)	(0.732, 0.732)	(0.762, 0.776)
##					
##	0.38	19:31	85.207	0.721	0.761
##			(84.354, 86.081)	(0.721, 0.721)	(0.754, 0.769)
##					
##	0.39	39:61	84.53	0.709	0.754
##			(83.639, 85.293)	(0.709, 0.709)	(0.746, 0.761)
##					
##	0.4	2:3	83.695	0.696	0.745
##			(82.867, 84.536)	(0.696, 0.696)	(0.738, 0.753)
##					
##	0.41	41:59	82.905	0.683	0.737
##			(82.063, 83.678)	(0.683, 0.683)	(0.729, 0.745)
##					
##	0.42	21:29	82.143	0.67	0.73
##			(81.233, 82.863)	(0.67, 0.67)	(0.721, 0.737)
##					
##	0.43	43:57	81.228	0.656	0.719
##			(80.454, 82.094)	(0.656, 0.656)	(0.713, 0.729)
##					
##	0.44	11:14	80.505	0.642	0.711
##			(79.657, 81.294)	(0.642, 0.642)	(0.704, 0.722)
##					
##	0.45	9:11	79.713	0.627	0.704

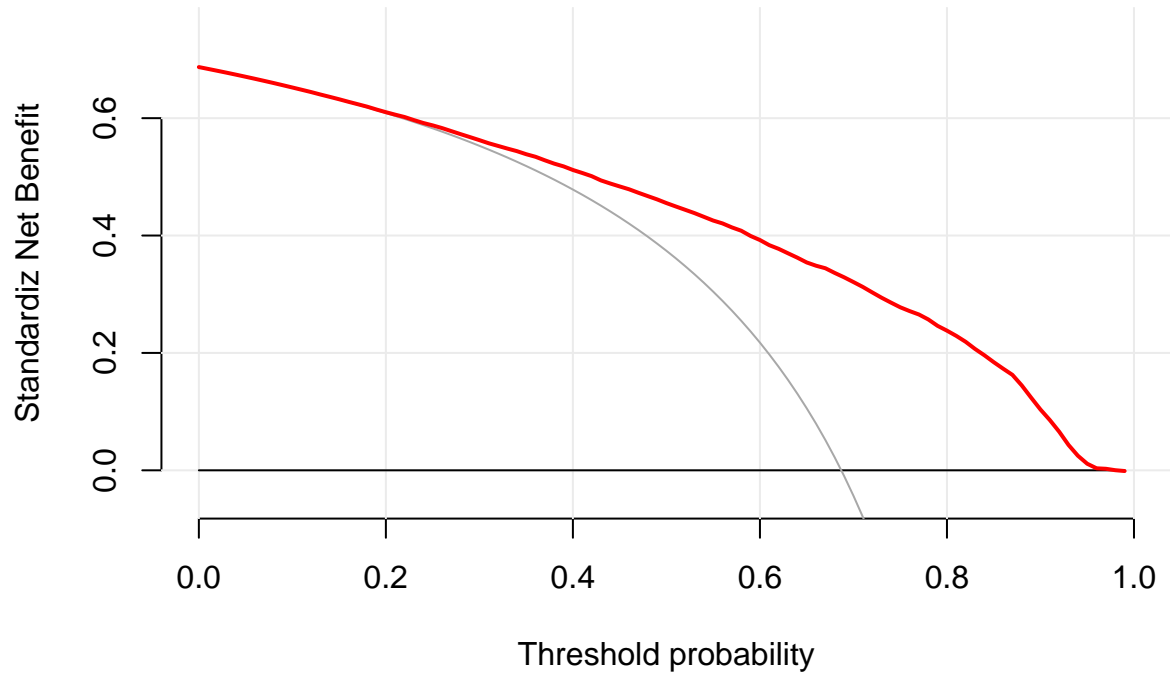
##			(78.812, 80.443)	(0.627, 0.627)	(0.696, 0.714)
##					
##	0.46	23:27	78.816	0.612	0.697
##			(77.938, 79.616)	(0.612, 0.612)	(0.688, 0.707)
##					
##	0.47	47:53	78.029	0.596	0.688
##			(77.071, 78.759)	(0.596, 0.596)	(0.68, 0.699)
##					
##	0.48	12:13	77.06	0.579	0.68
##			(76.189, 77.889)	(0.579, 0.579)	(0.67, 0.69)
##					
##	0.49	49:51	76.157	0.562	0.672
##			(75.295, 77.061)	(0.562, 0.562)	(0.662, 0.682)
##					
##	0.5	1:1	75.25	0.544	0.663
##			(74.416, 76.134)	(0.544, 0.544)	(0.654, 0.674)
##					
##	0.51	51:49	74.425	0.526	0.654
##			(73.536, 75.226)	(0.526, 0.526)	(0.646, 0.667)
##					
##	0.52	13:12	73.535	0.506	0.646
##			(72.549, 74.356)	(0.506, 0.506)	(0.637, 0.658)
##					
##	0.53	53:47	72.515	0.486	0.638
##			(71.633, 73.432)	(0.486, 0.486)	(0.628, 0.65)
##					
##	0.54	27:23	71.566	0.465	0.629
##			(70.669, 72.426)	(0.465, 0.465)	(0.619, 0.641)
##					
##	0.55	11:9	70.544	0.443	0.619
##			(69.693, 71.525)	(0.443, 0.443)	(0.611, 0.632)
##					
##	0.56	14:11	69.622	0.42	0.612
##			(68.662, 70.528)	(0.42, 0.42)	(0.601, 0.623)
##					
##	0.57	57:43	68.632	0.396	0.602
##			(67.634, 69.501)	(0.396, 0.396)	(0.592, 0.615)
##					
##	0.58	29:21	67.683	0.371	0.594
##			(66.534, 68.496)	(0.371, 0.371)	(0.582, 0.605)
##					
##	0.59	59:41	66.414	0.344	0.581
##			(65.444, 67.413)	(0.344, 0.344)	(0.572, 0.595)
##					
##	0.6	3:2	65.34	0.317	0.571
##			(64.396, 66.333)	(0.317, 0.317)	(0.56, 0.585)
##					
##	0.61	61:39	64.215	0.287	0.558
##			(63.255, 65.272)	(0.287, 0.287)	(0.55, 0.574)
##					
##	0.62	31:19	63.223	0.257	0.549
##			(62.096, 64.127)	(0.257, 0.257)	(0.539, 0.563)
##					
##	0.63	63:37	61.946	0.224	0.538

##			(61.071, 63.014)	(0.224, 0.224)	(0.527, 0.553)
##					
##	0.64	16:9	60.826	0.19	0.527
##			(59.898, 61.91)	(0.19, 0.19)	(0.516, 0.543)
##					
##	0.65	13:7	59.61	0.154	0.516
##			(58.685, 60.702)	(0.154, 0.154)	(0.505, 0.532)
##					
##	0.66	33:17	58.678	0.116	0.507
##			(57.612, 59.649)	(0.116, 0.116)	(0.496, 0.523)
##					
##	0.67	67:33	57.483	0.075	0.501
##			(56.45, 58.571)	(0.075, 0.075)	(0.485, 0.515)
##					
##	0.68	17:8	56.369	0.032	0.489
##			(55.387, 57.354)	(0.032, 0.032)	(0.475, 0.505)
##					
##	0.69	69:31	55.261	-0.014	0.479
##			(54.196, 56.185)	(-0.014, -0.014)	(0.464, 0.494)
##					
##	0.7	7:3	54.014	-0.063	0.467
##			(53.037, 55.046)	(-0.063, -0.063)	(0.452, 0.483)
##					
##	0.71	71:29	52.944	-0.115	0.455
##			(51.91, 53.777)	(-0.115, -0.115)	(0.441, 0.47)
##					
##	0.72	18:7	51.59	-0.171	0.441
##			(50.718, 52.413)	(-0.171, -0.171)	(0.427, 0.459)
##					
##	0.73	73:27	50.197	-0.232	0.428
##			(49.512, 51.211)	(-0.232, -0.232)	(0.415, 0.447)
##					
##	0.74	37:13	48.833	-0.297	0.416
##			(48.377, 50.074)	(-0.297, -0.297)	(0.404, 0.435)
##					
##	0.75	3:1	47.851	-0.367	0.404
##			(47.181, 48.996)	(-0.367, -0.367)	(0.391, 0.422)
##					
##	0.76	19:6	46.844	-0.443	0.395
##			(46.071, 48.005)	(-0.443, -0.443)	(0.381, 0.413)
##					
##	0.77	77:23	46.051	-0.525	0.386
##			(44.986, 47.053)	(-0.525, -0.525)	(0.369, 0.402)
##					
##	0.78	39:11	45.075	-0.615	0.374
##			(43.816, 46.113)	(-0.615, -0.615)	(0.356, 0.39)
##					
##	0.79	79:21	43.793	-0.714	0.358
##			(42.752, 45.23)	(-0.714, -0.714)	(0.343, 0.379)
##					
##	0.8	4:1	42.877	-0.822	0.346
##			(41.711, 44.277)	(-0.822, -0.822)	(0.329, 0.365)
##					
##	0.81	81:19	41.809	-0.942	0.333

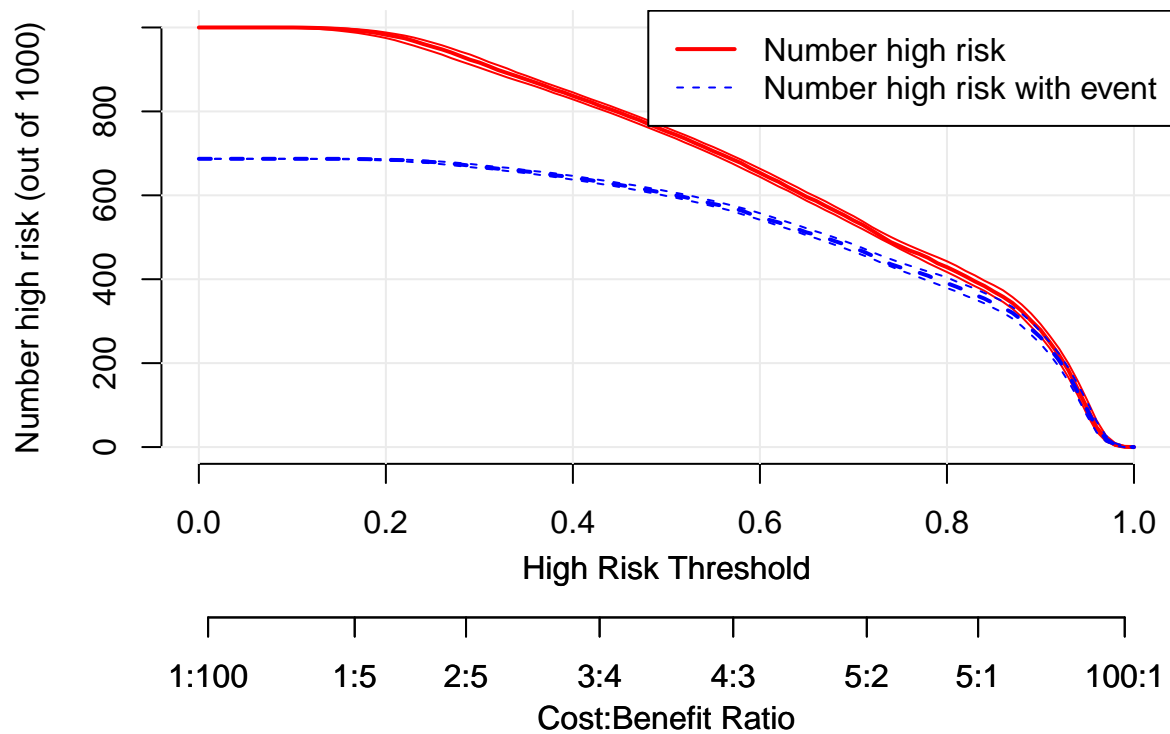
##			(40.627, 43.214)	(-0.942, -0.942)	(0.315, 0.352)
##					
##	0.82	41:9	40.692	-1.075	0.319
##			(39.446, 41.974)	(-1.075, -1.075)	(0.301, 0.337)
##					
##	0.83	83:17	39.469	-1.224	0.301
##			(38.323, 40.901)	(-1.224, -1.224)	(0.285, 0.322)
##					
##	0.84	21:4	38.39	-1.392	0.285
##			(37.243, 39.661)	(-1.392, -1.392)	(0.269, 0.306)
##					
##	0.85	17:3	37.05	-1.582	0.268
##			(35.872, 38.581)	(-1.582, -1.582)	(0.254, 0.29)
##					
##	0.86	43:7	35.812	-1.799	0.252
##			(34.584, 37.357)	(-1.799, -1.799)	(0.237, 0.273)
##					
##	0.87	87:13	34.338	-2.049	0.237
##			(32.972, 35.766)	(-2.049, -2.049)	(0.216, 0.254)
##					
##	0.88	22:3	32.439	-2.341	0.21
##			(31.005, 33.961)	(-2.341, -2.341)	(0.19, 0.232)
##					
##	0.89	89:11	30.103	-2.686	0.18
##			(28.669, 31.821)	(-2.686, -2.686)	(0.16, 0.204)
##					
##	0.9	9:1	27.923	-3.1	0.151
##			(26.393, 29.33)	(-3.1, -3.1)	(0.13, 0.174)
##					
##	0.91	91:9	24.966	-3.606	0.125
##			(23.542, 26.495)	(-3.606, -3.606)	(0.105, 0.145)
##					
##	0.92	23:2	21.881	-4.239	0.096
##			(20.275, 23.46)	(-4.239, -4.239)	(0.073, 0.115)
##					
##	0.93	93:7	18.253	-5.053	0.063
##			(16.537, 19.863)	(-5.053, -5.053)	(0.044, 0.084)
##					
##	0.94	47:3	13.759	-6.137	0.036
##			(12.297, 15.706)	(-6.137, -6.137)	(0.018, 0.055)
##					
##	0.95	19:1	9.232	-7.656	0.016
##			(7.677, 11.137)	(-7.656, -7.656)	(0, 0.035)
##					
##	0.96	24:1	4.697	-9.934	0.005
##			(3.667, 6.342)	(-9.934, -9.934)	(-0.008, 0.022)
##					
##	0.97	97:3	2.009	-13.73	0.004
##			(1.366, 2.756)	(-13.73, -13.73)	(-0.008, 0.014)
##					
##	0.98	49:1	0.527	-21.323	0
##			(0.302, 0.923)	(-21.323, -21.323)	(-0.006, 0.008)
##					
##	0.99	99:1	0.049	-44.102	-0.002

```
##          (0, 0.205)      (-44.102, -44.102)      (-0.007, 0.002)
##
##      1      Inf:1      0      NA      NA
##          (0, 0)      (NA, NA)      (NA, NA)
## -----
```

```
plot_decision_curve(model_1, curve.names = "Nomogram", xlim=c(0,1), xlab="Threshold probability", ylab=
```



```
##CIC
plot_clinical_impact(model_1, population.size= 1000,
  cost.benefit.axis = T,
  n.cost.benefits= 8, col =c('red', 'blue'),
  confidence.intervals=T,
  ylim=c(0,1000),
  legend.position="topright")
```



```
train_0_1_2$y <- as.factor(train_0_1_2$y)
str(train_0_1_2)
```

```
## tibble [22,902 x 20] (S3: tbl_df/tbl/data.frame)
## $ y      : Factor w/ 2 levels "0","1": 1 1 2 1 1 1 1 1 2 1 ...
## $ gender : Factor w/ 2 levels "1","2": 2 1 2 1 1 2 2 1 2 1 ...
## $ race   : Factor w/ 3 levels "1","2","3": 2 1 1 1 1 1 1 1 1 3 ...
## $ age    : Factor w/ 3 levels "1","2","3": 2 2 3 3 2 3 1 2 3 3 ...
## $ size   : Factor w/ 3 levels "1","2","3": 1 3 3 2 3 2 1 2 3 3 ...
## $ marry  : Factor w/ 3 levels "1","2","3": 2 2 2 1 2 3 2 3 2 2 ...
## $ income : Factor w/ 3 levels "1","2","3": 2 2 3 3 3 2 2 3 3 2 ...
## $ site   : Factor w/ 2 levels "1","2": 1 2 2 1 1 1 1 2 1 2 ...
## $ grade  : Factor w/ 5 levels "1","2","3","4",...: 5 5 5 5 2 1 2 2 5 2 ...
## $ kind   : Factor w/ 2 levels "1","2": 1 1 1 1 1 1 1 1 2 1 ...
## $ t      : Factor w/ 3 levels "1","2","3": 2 2 3 2 2 1 2 3 2 3 ...
## $ N      : Factor w/ 3 levels "1","2","3": 1 1 3 2 1 1 2 1 1 3 ...
## $ surgery: Factor w/ 2 levels "0","1": 1 1 1 1 1 2 2 1 1 1 ...
## $ RX     : Factor w/ 2 levels "0","1": 2 1 1 1 1 1 1 1 1 1 ...
## $ radiate: Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
## $ chem   : Factor w/ 2 levels "0","1": 2 1 2 2 2 1 2 2 1 1 ...
## $ CEA    : Factor w/ 3 levels "0","1","2": 2 1 2 2 2 2 3 2 2 3 ...
## $ bone   : Factor w/ 3 levels "0","1","2": 1 1 1 1 1 1 1 1 1 1 ...
## $ brain  : Factor w/ 3 levels "0","1","2": 1 1 1 1 1 1 1 1 1 1 ...
## $ lung   : Factor w/ 3 levels "0","1","2": 1 2 1 1 2 2 1 1 2 1 ...
```

```
test
```

```
test_0_1_2$y <- as.numeric(levels(test_0_1_2$y))[test_0_1_2$y]
str(test_0_1_2)
```

```
## tibble [5,726 x 20] (S3: tbl_df/tbl/data.frame)
## $ y      : num [1:5726] 0 0 0 0 1 0 0 0 1 1 ...
## $ gender : Factor w/ 2 levels "1","2": 1 2 1 2 1 2 2 1 1 2 ...
## $ race   : Factor w/ 3 levels "1","2","3": 1 1 2 2 1 3 1 3 3 2 ...
## $ age    : Factor w/ 3 levels "1","2","3": 3 3 3 3 3 2 2 3 3 2 ...
## $ size   : Factor w/ 3 levels "1","2","3": 1 3 3 2 1 2 3 2 3 2 ...
## $ marry  : Factor w/ 3 levels "1","2","3": 3 2 2 1 3 2 3 3 2 1 ...
## $ income : Factor w/ 3 levels "1","2","3": 3 3 3 3 3 3 3 3 3 3 ...
## $ site   : Factor w/ 2 levels "1","2": 2 2 1 2 2 1 2 2 2 1 ...
## $ grade  : Factor w/ 5 levels "1","2","3","4",...: 2 2 2 2 5 2 2 2 5 2 ...
## $ kind   : Factor w/ 2 levels "1","2": 1 1 1 1 1 1 1 1 2 1 ...
## $ t      : Factor w/ 3 levels "1","2","3": 2 2 1 2 3 2 1 2 2 2 ...
## $ N      : Factor w/ 3 levels "1","2","3": 2 1 1 2 1 1 1 2 2 1 ...
## $ surgery: Factor w/ 2 levels "0","1": 2 1 1 2 1 1 1 2 1 2 ...
## $ RX     : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
## $ radiate: Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
## $ chem   : Factor w/ 2 levels "0","1": 1 2 2 1 1 2 2 2 2 1 ...
## $ CEA    : Factor w/ 3 levels "0","1","2": 1 2 3 2 2 2 2 2 3 2 ...
## $ bone   : Factor w/ 3 levels "0","1","2": 1 1 1 1 1 1 1 1 1 1 ...
## $ brain  : Factor w/ 3 levels "0","1","2": 1 1 1 1 1 1 1 1 1 1 ...
## $ lung   : Factor w/ 3 levels "0","1","2": 1 1 2 1 2 1 1 1 1 1 ...
```

```
model_1 <- decision_curve(y~race+size+marry+income+site+grade+kind+t+N+surgery+RX+radiate+chem+CEA+bone
```

```
## Calculating net benefit curves for case-control data. All calculations are done conditional on the o
```

```
## Note: The data provided is used to both fit a prediction model and to estimate the respective decis
```

```
summary(model_1)
```

```
##
## Standardized Net Benefit (95% Confidence Intervals):
## -----
```

## risk	## cost:benefit	## percent	## All	## y ~ race + size + marry +
## threshold	## ratio	## high risk		## income + site + grade + kind +
##				## t + N + surgery + RX + radiate
##				## + chem + CEA + bone + brain +
##				## lung
## 0	## 0:1	## 100	## 1	## 1
		## (100, 100)	## (1, 1)	## (1, 1)
## 0.01	## 1:99	## 100	## 0.995	## 0.995
		## (100, 100)	## (0.995, 0.995)	## (0.995, 0.995)
## 0.02	## 1:49	## 100	## 0.991	## 0.991
		## (100, 100)	## (0.991, 0.991)	## (0.991, 0.991)

```
##
```


##	0.03	3:97	100	0.986	0.986
##			(100, 100)	(0.986, 0.986)	(0.986, 0.986)
##					
##	0.04	1:24	100	0.981	0.981
##			(100, 100)	(0.981, 0.981)	(0.981, 0.981)
##					
##	0.05	1:19	100	0.976	0.976
##			(99.993, 100)	(0.976, 0.976)	(0.976, 0.976)
##					
##	0.06	3:47	100	0.97	0.97
##			(99.979, 100)	(0.97, 0.97)	(0.97, 0.97)
##					
##	0.07	7:93	100	0.965	0.965
##			(99.966, 100)	(0.965, 0.965)	(0.965, 0.965)
##					
##	0.08	2:23	100	0.96	0.96
##			(99.933, 100)	(0.96, 0.96)	(0.96, 0.96)
##					
##	0.09	9:91	100	0.954	0.954
##			(99.89, 100)	(0.954, 0.954)	(0.954, 0.954)
##					
##	0.1	1:9	99.993	0.949	0.949
##			(99.807, 100)	(0.949, 0.949)	(0.949, 0.949)
##					
##	0.11	11:89	99.993	0.943	0.943
##			(99.671, 100)	(0.943, 0.943)	(0.943, 0.943)
##					
##	0.12	3:22	99.979	0.937	0.937
##			(99.503, 100)	(0.937, 0.937)	(0.936, 0.938)
##					
##	0.13	13:87	99.959	0.931	0.931
##			(99.305, 100)	(0.931, 0.931)	(0.93, 0.932)
##					
##	0.14	7:43	99.917	0.925	0.925
##			(99.085, 99.993)	(0.925, 0.925)	(0.923, 0.926)
##					
##	0.15	3:17	99.848	0.918	0.919
##			(98.705, 99.972)	(0.918, 0.918)	(0.917, 0.921)
##					
##	0.16	4:21	99.692	0.912	0.912
##			(98.28, 99.924)	(0.912, 0.912)	(0.91, 0.915)
##					
##	0.17	17:83	99.547	0.905	0.905
##			(97.852, 99.869)	(0.905, 0.905)	(0.903, 0.909)
##					
##	0.18	9:41	99.374	0.898	0.899
##			(97.376, 99.738)	(0.898, 0.898)	(0.896, 0.903)
##					
##	0.19	19:81	99.098	0.891	0.893
##			(96.905, 99.554)	(0.891, 0.891)	(0.89, 0.897)
##					
##	0.2	1:4	98.68	0.884	0.887
##			(96.351, 99.292)	(0.884, 0.884)	(0.882, 0.891)
##					

##	0.21	21:79	98.289	0.877	0.88
##			(95.739, 99.014)	(0.877, 0.877)	(0.875, 0.885)
##					
##	0.22	11:39	97.849	0.869	0.874
##			(95.137, 98.678)	(0.869, 0.869)	(0.867, 0.879)
##					
##	0.23	23:77	97.285	0.862	0.864
##			(94.638, 98.33)	(0.862, 0.862)	(0.86, 0.873)
##					
##	0.24	6:19	96.444	0.854	0.856
##			(93.896, 97.801)	(0.854, 0.854)	(0.853, 0.868)
##					
##	0.25	1:3	95.858	0.846	0.852
##			(93.217, 97.211)	(0.846, 0.846)	(0.845, 0.861)
##					
##	0.26	13:37	95.299	0.837	0.847
##			(92.657, 96.629)	(0.837, 0.837)	(0.839, 0.856)
##					
##	0.27	27:73	94.459	0.829	0.839
##			(91.847, 95.988)	(0.829, 0.829)	(0.831, 0.849)
##					
##	0.28	7:18	93.78	0.82	0.833
##			(91.053, 95.335)	(0.82, 0.82)	(0.824, 0.843)
##					
##	0.29	29:71	93.198	0.811	0.827
##			(90.281, 94.565)	(0.811, 0.811)	(0.816, 0.837)
##					
##	0.3	3:7	92.289	0.802	0.82
##			(89.536, 93.881)	(0.802, 0.802)	(0.808, 0.83)
##					
##	0.31	31:69	91.538	0.792	0.812
##			(88.696, 93.126)	(0.792, 0.792)	(0.801, 0.823)
##					
##	0.32	8:17	90.626	0.782	0.803
##			(87.842, 92.28)	(0.782, 0.782)	(0.793, 0.817)
##					
##	0.33	33:67	89.954	0.772	0.797
##			(87.103, 91.396)	(0.772, 0.772)	(0.785, 0.81)
##					
##	0.34	17:33	89.178	0.762	0.788
##			(86.339, 90.505)	(0.762, 0.762)	(0.777, 0.802)
##					
##	0.35	7:13	88.254	0.751	0.778
##			(85.399, 89.684)	(0.751, 0.751)	(0.768, 0.794)
##					
##	0.36	9:16	87.598	0.74	0.772
##			(84.646, 88.815)	(0.74, 0.74)	(0.759, 0.787)
##					
##	0.37	37:63	86.099	0.728	0.758
##			(83.737, 87.941)	(0.728, 0.728)	(0.75, 0.778)
##					
##	0.38	19:31	85.236	0.716	0.749
##			(82.978, 87.005)	(0.716, 0.716)	(0.741, 0.771)
##					

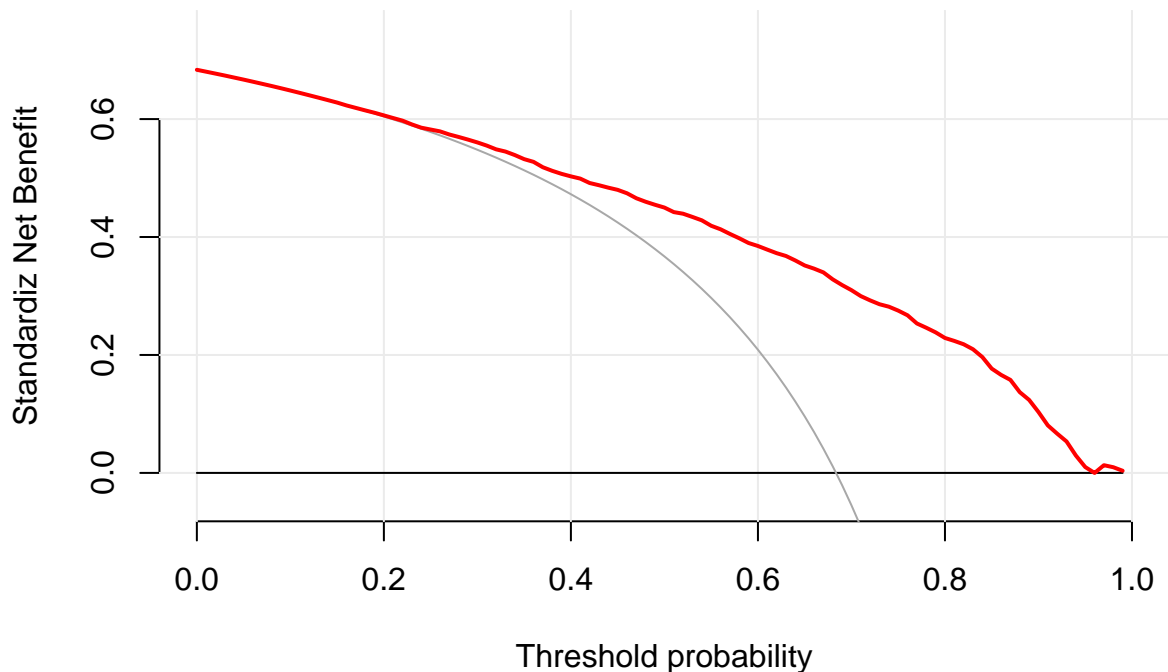
##	0.39	39:61	84.404	0.704	0.742
##			(82.1, 86.256)	(0.704, 0.704)	(0.732, 0.765)
##					
##	0.4	2:3	83.748	0.691	0.736
##			(81.239, 85.286)	(0.691, 0.691)	(0.724, 0.757)
##					
##	0.41	41:59	83.037	0.678	0.73
##			(80.386, 84.386)	(0.678, 0.678)	(0.716, 0.751)
##					
##	0.42	21:29	81.707	0.665	0.719
##			(79.414, 83.479)	(0.665, 0.665)	(0.708, 0.744)
##					
##	0.43	43:57	81.03	0.651	0.714
##			(78.515, 82.418)	(0.651, 0.651)	(0.697, 0.736)
##					
##	0.44	11:14	80.144	0.636	0.708
##			(77.697, 81.591)	(0.636, 0.636)	(0.69, 0.728)
##					
##	0.45	9:11	79.222	0.621	0.702
##			(76.721, 80.59)	(0.621, 0.621)	(0.68, 0.723)
##					
##	0.46	23:27	78.047	0.606	0.694
##			(75.928, 79.624)	(0.606, 0.606)	(0.673, 0.713)
##					
##	0.47	47:53	76.687	0.589	0.681
##			(74.923, 78.619)	(0.589, 0.589)	(0.666, 0.707)
##					
##	0.48	12:13	75.832	0.573	0.673
##			(73.971, 77.756)	(0.573, 0.573)	(0.656, 0.699)
##					
##	0.49	49:51	75.016	0.555	0.665
##			(73.128, 76.866)	(0.555, 0.555)	(0.648, 0.692)
##					
##	0.5	1:1	74.295	0.537	0.658
##			(72.215, 75.921)	(0.537, 0.537)	(0.64, 0.682)
##					
##	0.51	51:49	73.166	0.518	0.647
##			(71.281, 75.053)	(0.518, 0.518)	(0.63, 0.675)
##					
##	0.52	13:12	72.516	0.498	0.643
##			(70.421, 74.07)	(0.498, 0.498)	(0.621, 0.666)
##					
##	0.53	53:47	71.668	0.478	0.635
##			(69.431, 72.997)	(0.478, 0.478)	(0.612, 0.658)
##					
##	0.54	27:23	70.523	0.457	0.627
##			(68.429, 72.053)	(0.457, 0.457)	(0.603, 0.649)
##					
##	0.55	11:9	69.24	0.434	0.613
##			(67.367, 71.009)	(0.434, 0.434)	(0.595, 0.642)
##					
##	0.56	14:11	68.15	0.411	0.604
##			(66.54, 70.124)	(0.411, 0.411)	(0.583, 0.632)
##					

##	0.57	57:43	67.008	0.386	0.593
##			(65.463, 69.015)	(0.386, 0.386)	(0.572, 0.623)
##					
##	0.58	29:21	65.948	0.361	0.582
##			(64.435, 68.024)	(0.361, 0.361)	(0.564, 0.614)
##					
##	0.59	59:41	64.841	0.334	0.57
##			(63.217, 67.16)	(0.334, 0.334)	(0.553, 0.606)
##					
##	0.6	3:2	63.845	0.306	0.563
##			(62.054, 66.02)	(0.306, 0.306)	(0.54, 0.596)
##					
##	0.61	61:39	62.939	0.276	0.554
##			(60.866, 64.904)	(0.276, 0.276)	(0.53, 0.588)
##					
##	0.62	31:19	61.837	0.245	0.545
##			(59.747, 63.946)	(0.245, 0.245)	(0.521, 0.578)
##					
##	0.63	63:37	60.868	0.212	0.538
##			(58.853, 62.852)	(0.212, 0.212)	(0.511, 0.571)
##					
##	0.64	16:9	59.418	0.177	0.527
##			(57.74, 61.783)	(0.177, 0.177)	(0.503, 0.56)
##					
##	0.65	13:7	58.589	0.14	0.515
##			(56.551, 60.75)	(0.14, 0.14)	(0.493, 0.551)
##					
##	0.66	33:17	57.42	0.101	0.507
##			(55.431, 59.601)	(0.101, 0.101)	(0.482, 0.539)
##					
##	0.67	67:33	56.05	0.06	0.497
##			(54.462, 58.484)	(0.06, 0.06)	(0.471, 0.526)
##					
##	0.68	17:8	54.605	0.016	0.48
##			(53.38, 57.278)	(0.016, 0.016)	(0.459, 0.516)
##					
##	0.69	69:31	53.502	-0.03	0.466
##			(52.284, 56.018)	(-0.03, -0.03)	(0.445, 0.504)
##					
##	0.7	7:3	52.277	-0.08	0.453
##			(51.169, 54.763)	(-0.08, -0.08)	(0.431, 0.495)
##					
##	0.71	71:29	50.865	-0.133	0.439
##			(50.219, 53.643)	(-0.133, -0.133)	(0.419, 0.483)
##					
##	0.72	18:7	49.774	-0.19	0.428
##			(49.045, 52.631)	(-0.19, -0.19)	(0.408, 0.471)
##					
##	0.73	73:27	48.911	-0.252	0.418
##			(48.013, 51.549)	(-0.252, -0.252)	(0.397, 0.461)
##					
##	0.74	37:13	48.295	-0.318	0.412
##			(46.964, 50.672)	(-0.318, -0.318)	(0.386, 0.45)
##					

##	0.75	3:1	47.608	-0.389	0.403
##			(45.808, 49.786)	(-0.389, -0.389)	(0.374, 0.438)
##					
##	0.76	19:6	46.726	-0.466	0.391
##			(44.838, 48.917)	(-0.466, -0.466)	(0.362, 0.429)
##					
##	0.77	77:23	45.164	-0.55	0.371
##			(43.721, 48.061)	(-0.55, -0.55)	(0.351, 0.417)
##					
##	0.78	39:11	44.268	-0.641	0.36
##			(42.451, 47.204)	(-0.641, -0.641)	(0.339, 0.403)
##					
##	0.79	79:21	43.313	-0.742	0.349
##			(41.718, 46.152)	(-0.742, -0.742)	(0.326, 0.392)
##					
##	0.8	4:1	42.332	-0.852	0.335
##			(40.669, 45.284)	(-0.852, -0.852)	(0.313, 0.381)
##					
##	0.81	81:19	41.617	-0.974	0.327
##			(39.546, 44.297)	(-0.974, -0.974)	(0.3, 0.373)
##					
##	0.82	41:9	40.554	-1.109	0.319
##			(38.364, 43.05)	(-1.109, -1.109)	(0.286, 0.357)
##					
##	0.83	83:17	39.344	-1.26	0.307
##			(37.213, 42.043)	(-1.26, -1.26)	(0.27, 0.345)
##					
##	0.84	21:4	37.876	-1.431	0.287
##			(35.727, 40.901)	(-1.431, -1.431)	(0.254, 0.33)
##					
##	0.85	17:3	36.31	-1.623	0.259
##			(34.308, 39.676)	(-1.623, -1.623)	(0.237, 0.312)
##					
##	0.86	43:7	35.149	-1.844	0.243
##			(32.654, 38.293)	(-1.844, -1.844)	(0.211, 0.295)
##					
##	0.87	87:13	33.753	-2.098	0.231
##			(30.695, 36.74)	(-2.098, -2.098)	(0.189, 0.277)
##					
##	0.88	22:3	31.51	-2.395	0.2
##			(28.627, 34.942)	(-2.395, -2.395)	(0.166, 0.254)
##					
##	0.89	89:11	29.245	-2.746	0.181
##			(26.381, 32.858)	(-2.746, -2.746)	(0.138, 0.229)
##					
##	0.9	9:1	27.084	-3.167	0.151
##			(23.864, 30.648)	(-3.167, -3.167)	(0.118, 0.206)
##					
##	0.91	91:9	24.153	-3.681	0.118
##			(21.326, 27.905)	(-3.681, -3.681)	(0.092, 0.176)
##					
##	0.92	23:2	20.967	-4.324	0.097
##			(18.183, 25.021)	(-4.324, -4.324)	(0.067, 0.149)
##					

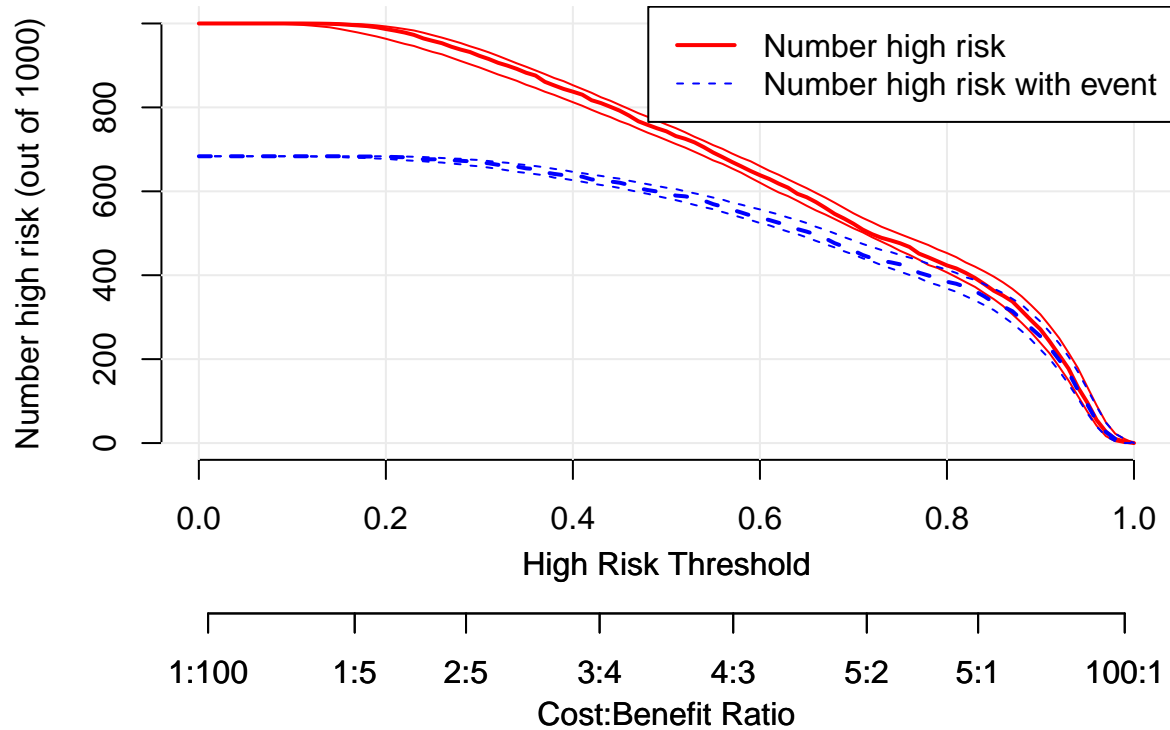
```
## 0.93      93:7      17.748      -5.151      0.078
##          (14.924, 21.527)      (-5.151, -5.151)      (0.041, 0.127)
##
## 0.94      47:3      13.295      -6.253      0.043
##          (11.218, 17.883)      (-6.253, -6.253)      (0.017, 0.096)
##
## 0.95      19:1      9.685      -7.796      0.015
##          (7.534, 13.425)      (-7.796, -7.796)      (-0.013, 0.07)
##
## 0.96      24:1      5.507      -10.111      0
##          (4.142, 8.992)      (-10.111, -10.111)      (-0.021, 0.047)
##
## 0.97      97:3      2.684      -13.969      0.019
##          (1.642, 4.934)      (-13.969, -13.969)      (-0.012, 0.039)
##
## 0.98      49:1      0.961      -21.685      0.014
##          (0.441, 2.222)      (-21.685, -21.685)      (-0.005, 0.028)
##
## 0.99      99:1      0.36      -44.832      0.005
##          (0, 0.961)      (-44.832, -44.832)      (0, 0.014)
##
## 1          Inf:1      0      NA      NA
##          (0, 0)      (NA, NA)      (NA, NA)
## -----
```

```
plot_decision_curve(model_1, curve.names = "Nomogram", xlim=c(0,1), xlab="Threshold probability", ylab=
```



```
##CIC
plot_clinical_impact(model_1, population.size= 1000,
  cost.benefit.axis = T,
  n.cost.benefits= 8, col =c('red', 'blue'),
```

```
confidence.intervals=T,
ylim=c(0,1000),
legend.position="topright")
```



```
test_0_1_2$y <- as.factor(test_0_1_2$y)
str(test_0_1_2)
```

```
## tibble [5,726 x 20] (S3: tbl_df/tbl/data.frame)
## $ y      : Factor w/ 2 levels "0","1": 1 1 1 1 2 1 1 1 2 2 ...
## $ gender : Factor w/ 2 levels "1","2": 1 2 1 2 1 2 2 1 1 2 ...
## $ race   : Factor w/ 3 levels "1","2","3": 1 1 2 2 1 3 1 3 3 2 ...
## $ age    : Factor w/ 3 levels "1","2","3": 3 3 3 3 3 2 2 3 3 2 ...
## $ size   : Factor w/ 3 levels "1","2","3": 1 3 3 2 1 2 3 2 3 2 ...
## $ marry  : Factor w/ 3 levels "1","2","3": 3 2 2 1 3 2 3 3 2 1 ...
## $ income : Factor w/ 3 levels "1","2","3": 3 3 3 3 3 3 3 3 3 3 ...
## $ site   : Factor w/ 2 levels "1","2": 2 2 1 2 2 1 2 2 2 1 ...
## $ grade  : Factor w/ 5 levels "1","2","3","4",...: 2 2 2 2 5 2 2 2 5 2 ...
## $ kind   : Factor w/ 2 levels "1","2": 1 1 1 1 1 1 1 1 2 1 ...
## $ t      : Factor w/ 3 levels "1","2","3": 2 2 1 2 3 2 1 2 2 2 ...
## $ N      : Factor w/ 3 levels "1","2","3": 2 1 1 2 1 1 1 2 2 1 ...
## $ surgery: Factor w/ 2 levels "0","1": 2 1 1 2 1 1 1 2 1 2 ...
## $ RX     : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
## $ radiate: Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
## $ chem   : Factor w/ 2 levels "0","1": 1 2 2 1 1 2 2 2 2 1 ...
## $ CEA    : Factor w/ 3 levels "0","1","2": 1 2 3 2 2 2 2 2 3 2 ...
## $ bone   : Factor w/ 3 levels "0","1","2": 1 1 1 1 1 1 1 1 1 1 ...
## $ brain  : Factor w/ 3 levels "0","1","2": 1 1 1 1 1 1 1 1 1 1 ...
## $ lung   : Factor w/ 3 levels "0","1","2": 1 1 2 1 2 1 1 1 1 1 ...
```

0+1+2 multivariate forest

```
fit.result<-summary(m2_final)
df1<-fit.result$coefficients
df2<-confint(m2_final)
```

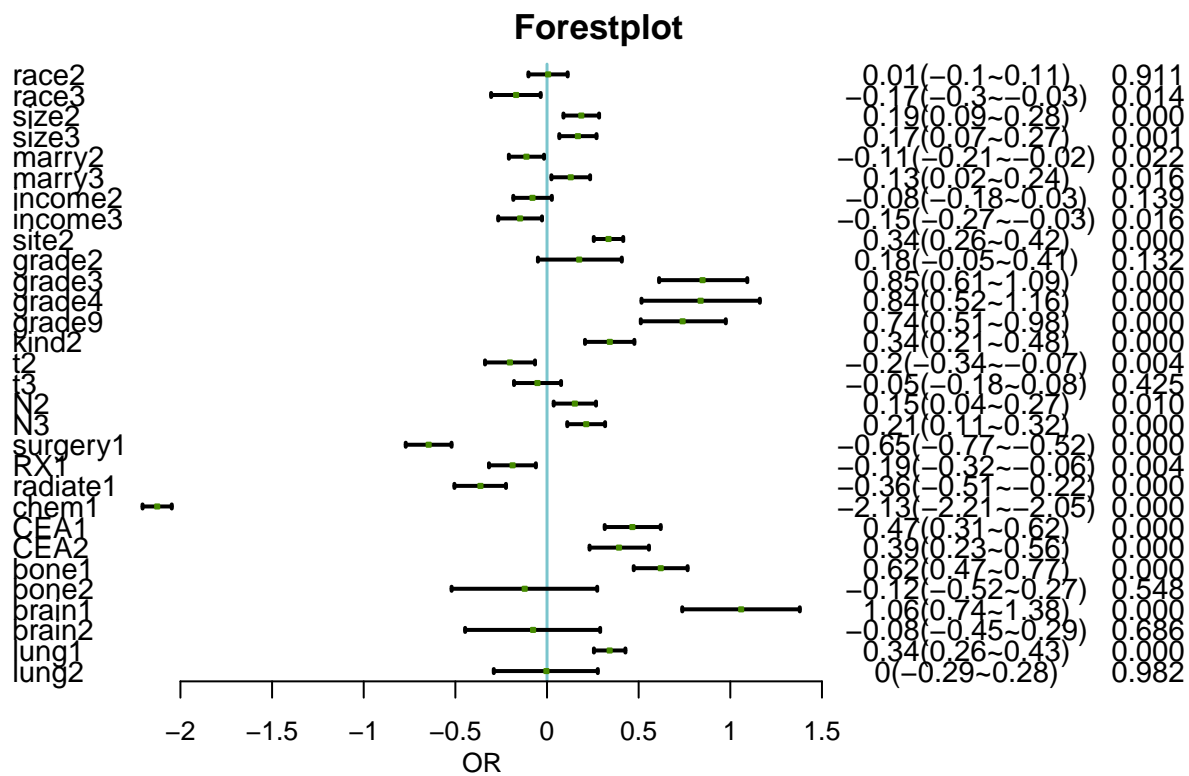
```
## Waiting for profiling to be done...
```

```
df3<-cbind(df1,df2)
df4<-data.frame(df3[-1,c(1,4,5,6)])
df4$Var<-rownames(df4)
colnames(df4)<-c("OR", "Pvalue", "OR_1", "OR_2", "Var")
df5<-df4[,c(5,1,2,3,4)]
df5$OR_mean<-df5$OR
df5$OR<-paste0(round(df5$OR,2),
               "(",
               round(df5$OR_1,2),
               "~",
               round(df5$OR_2,2),
               ")")
df5$Pvalue<-round(df5$Pvalue,3)
write.csv(df5,file = "forestplot_m2.csv",
          quote = F,row.names = F)

fp<-read.csv("forestplot_m2.csv",header=T)

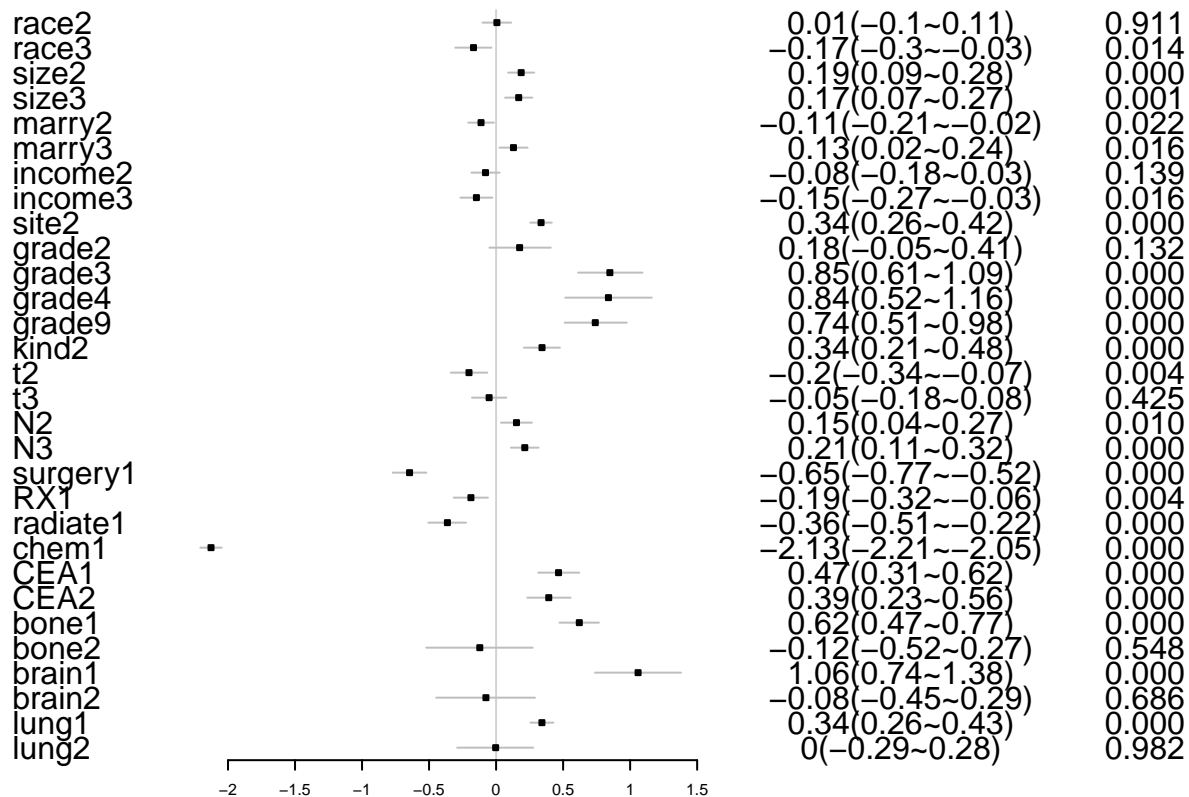
## plot 1

forestplot(labeltext=as.matrix(fp[,1:3]),
           mean=fp$OR_mean,
           lower=fp$OR_1,
           upper=fp$OR_2,
           zero=0,
           boxsize=0.2,
           lineheight = unit(7,'mm'),
           colgap=unit(2,'mm'),
           lwd.zero=1.5,
           lwd.ci=2,
           col=fpColors(box='#458B00',
                        summary='#8B008B',
                        lines = 'black',
                        zero = '#7AC5CD'),
           xlab="OR",
           lwd.xaxis =1,
           txt_gp = fpTxtGp(ticks = gpar(cex = 0.85),
                             xlab = gpar(cex = 0.8),
                             cex = 0.9),
           lty.ci = "solid",
           title = "Forestplot",
           line.margin = 0.08,
           graph.pos=2)
```

```
## plot 2

forestplot(labeltext=as.matrix(fp[,1:3]),
           mean=fp$OR_mean,
           lower=fp$OR_1,
           upper=fp$OR_2,
           zero=0,
           boxsize=0.2,
           graph.pos=2)
```



0+1+2 multivariate forward/backward/both AIC

```
fullmod_2 <- glm(y ~ ., family=binomial, data = train_0_1_2)
summary(fullmod_2)
```

```
##
## Call:
## glm(formula = y ~ ., family = binomial, data = train_0_1_2)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.3188  -0.5652  -0.3714  -0.2366   2.9570
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.979324   0.206341  -4.746 2.07e-06 ***
## gender2      -0.001589   0.040128  -0.040 0.968406
## race2         0.024469   0.054678   0.448 0.654512
## race3        -0.161009   0.068992  -2.334 0.019610 *
## age2         -0.139698   0.136934  -1.020 0.307640
## age3          0.241158   0.135106   1.785 0.074268 .
## size2         0.194269   0.049664   3.912 9.17e-05 ***
## size3         0.172809   0.051855   3.333 0.000861 ***
## marry2       -0.154456   0.049430  -3.125 0.001780 **
## marry3         0.049716   0.055740   0.892 0.372432
## income2      -0.075376   0.053653  -1.405 0.160059
```

```
## income3      -0.151825    0.060915   -2.492 0.012689 *
## site2        0.291654    0.041491    7.029 2.08e-12 ***
## grade2       0.170174    0.116721    1.458 0.144853
## grade3       0.853296    0.122695    6.955 3.54e-12 ***
## grade4       0.837604    0.164767    5.084 3.70e-07 ***
## grade9       0.732778    0.118251    6.197 5.76e-10 ***
## kind2        0.345567    0.068827    5.021 5.15e-07 ***
## t2           -0.206630    0.069656   -2.966 0.003013 **
## t3           -0.059734    0.065581   -0.911 0.362378
## N2           0.167738    0.059073    2.839 0.004519 **
## N3           0.221629    0.052649    4.210 2.56e-05 ***
## surgery1     -0.642876    0.063971  -10.049 < 2e-16 ***
## RX1          -0.158965    0.065524   -2.426 0.015263 *
## radiate1     -0.352094    0.071875   -4.899 9.65e-07 ***
## chem1        -2.068721    0.041160  -50.260 < 2e-16 ***
## CEA1          0.459978    0.077995    5.898 3.69e-09 ***
## CEA2          0.378222    0.082729    4.572 4.84e-06 ***
## bone1         0.627439    0.075156    8.348 < 2e-16 ***
## bone2        -0.108254    0.202943   -0.533 0.593742
## brain1        1.073278    0.164202    6.536 6.31e-11 ***
## brain2       -0.089284    0.188409   -0.474 0.635584
## lung1         0.337601    0.043880    7.694 1.43e-14 ***
## lung2        -0.012107    0.144907   -0.084 0.933413
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##    Null deviance: 22383  on 22901  degrees of freedom
## Residual deviance: 17367  on 22868  degrees of freedom
## AIC: 17435
##
## Number of Fisher Scoring iterations: 5
```

```
nothing_2 <- glm(y ~ 1, family=binomial, data = train_0_1_2)
summary(nothing_2)
```

```
##
## Call:
## glm(formula = y ~ 1, family = binomial, data = train_0_1_2)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.6524  -0.6524  -0.6524  -0.6524   1.8176
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.43909    0.01679  -85.73  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##    Null deviance: 22383  on 22901  degrees of freedom
```

```
## Residual deviance: 22383 on 22901 degrees of freedom
## AIC: 22385
##
## Number of Fisher Scoring iterations: 4
```

```
backwards_aic_2 = step(fullmod_2)
```

```
## Start: AIC=17435.02
## y ~ gender + race + age + size + marry + income + site + grade +
##      kind + t + N + surgery + RX + radiate + chem + CEA + bone +
##      brain + lung
```

```
##
##           Df Deviance   AIC
## - gender    1    17367 17433
## <none>      17367 17435
## - race      2    17373 17437
## - income    2    17373 17437
## - RX        1    17373 17439
## - t         2    17378 17442
## - size      2    17386 17450
## - marry     2    17388 17452
## - N         2    17391 17455
## - kind      1    17392 17458
## - radiate   1    17392 17458
## - CEA       2    17404 17468
## - brain     2    17409 17473
## - site      1    17416 17482
## - lung      2    17427 17491
## - bone      2    17436 17500
## - age       2    17438 17502
## - surgery   1    17470 17536
## - grade     4    17604 17664
## - chem      1    20077 20143
##
```

```
## Step: AIC=17433.02
## y ~ race + age + size + marry + income + site + grade + kind +
##      t + N + surgery + RX + radiate + chem + CEA + bone + brain +
##      lung
```

```
##
##           Df Deviance   AIC
## <none>      17367 17433
## - race      2    17373 17435
## - income    2    17373 17435
## - RX        1    17373 17437
## - t         2    17378 17440
## - size      2    17386 17448
## - marry     2    17389 17451
## - N         2    17391 17453
## - kind      1    17392 17456
## - radiate   1    17392 17456
## - CEA       2    17404 17466
## - brain     2    17409 17471
## - site      1    17417 17481
## - lung      2    17427 17489
```

```
## - bone      2      17436 17498
## - age       2      17438 17500
## - surgery   1      17470 17534
## - grade     4      17604 17662
## - chem      1      20079 20143
```

```
formula(backwards_aic_2)
```

```
## y ~ race + age + size + marry + income + site + grade + kind +
##      t + N + surgery + RX + radiate + chem + CEA + bone + brain +
##      lung
```

```
summary(backwards_aic_2)
```

```
##
## Call:
## glm(formula = y ~ race + age + size + marry + income + site +
##      grade + kind + t + N + surgery + RX + radiate + chem + CEA +
##      bone + brain + lung, family = binomial, data = train_0_1_2)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.3193  -0.5653  -0.3714  -0.2365   2.9569
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.98029    0.20489  -4.785 1.71e-06 ***
## race2        0.02453    0.05466   0.449 0.653651
## race3       -0.16093    0.06896  -2.334 0.019620 *
## age2        -0.13984    0.13689  -1.022 0.306980
## age3         0.24101    0.13505   1.785 0.074335 .
## size2        0.19418    0.04962   3.914 9.09e-05 ***
## size3        0.17278    0.05185   3.332 0.000861 ***
## marry2      -0.15446    0.04943  -3.125 0.001779 **
## marry3       0.05008    0.05497   0.911 0.362268
## income2     -0.07533    0.05364  -1.404 0.160215
## income3     -0.15178    0.06091  -2.492 0.012700 *
## site2        0.29180    0.04132   7.062 1.64e-12 ***
## grade2       0.17016    0.11672   1.458 0.144888
## grade3       0.85331    0.12270   6.955 3.53e-12 ***
## grade4       0.83770    0.16475   5.085 3.68e-07 ***
## grade9       0.73275    0.11825   6.197 5.77e-10 ***
## kind2        0.34564    0.06880   5.024 5.07e-07 ***
## t2          -0.20661    0.06965  -2.966 0.003015 **
## t3          -0.05971    0.06558  -0.911 0.362517
## N2           0.16775    0.05907   2.840 0.004516 **
## N3           0.22162    0.05265   4.209 2.56e-05 ***
## surgery1    -0.64289    0.06397 -10.050 < 2e-16 ***
## RX1         -0.15886    0.06547  -2.426 0.015249 *
## radiate1    -0.35211    0.07187  -4.899 9.63e-07 ***
## chem1       -2.06876    0.04115 -50.273 < 2e-16 ***
## CEA1         0.46003    0.07798   5.899 3.65e-09 ***
## CEA2         0.37827    0.08272   4.573 4.81e-06 ***
```

```
## bone1      0.62737    0.07514    8.350 < 2e-16 ***
## bone2     -0.10843    0.20290   -0.534 0.593066
## brain1     1.07341    0.16417    6.538 6.21e-11 ***
## brain2    -0.08913    0.18836   -0.473 0.636098
## lung1      0.33761    0.04388    7.694 1.43e-14 ***
## lung2     -0.01207    0.14490   -0.083 0.933615
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##    Null deviance: 22383  on 22901  degrees of freedom
## Residual deviance: 17367  on 22869  degrees of freedom
## AIC: 17433
##
## Number of Fisher Scoring iterations: 5
```

```
forwards_aic_2 = step(nothing_2, scope=list(lower=formula(nothing_2),upper=formula(fullmod_2)), direction="forward")
```

```
## Start:  AIC=22384.53
## y ~ 1
##
##           Df Deviance   AIC
## + chem      1    18939 18943
## + surgery    1    21665 21669
## + grade      4    21804 21814
## + age        2    21894 21900
## + t          2    21922 21928
## + size       2    22023 22029
## + N          2    22062 22068
## + RX         1    22090 22094
## + marry      2    22168 22174
## + lung       2    22183 22189
## + bone       2    22215 22221
## + CEA        2    22249 22255
## + site       1    22252 22256
## + brain      2    22274 22280
## + radiate    1    22288 22292
## + kind       1    22324 22328
## + income     2    22361 22367
## + race       2    22365 22371
## + gender     1    22368 22372
## <none>       22382 22384
##
## Step:  AIC=18943.4
## y ~ chem
##
##           Df Deviance   AIC
## + surgery    1    18246 18252
## + grade      4    18304 18316
## + t          2    18505 18513
## + N          2    18648 18656
## + size       2    18651 18659
## + lung       2    18747 18755
```

```

## + bone      2      18787 18795
## + RX        1      18797 18803
## + age       2      18826 18834
## + CEA       2      18862 18870
## + brain     2      18869 18877
## + site      1      18893 18899
## + marry     2      18894 18902
## + kind      1      18901 18907
## + race      2      18921 18929
## + radiate   1      18933 18939
## <none>             18939 18943
## + gender    1      18938 18944
## + income    2      18937 18945
##
## Step:  AIC=18251.99
## y ~ chem + surgery
##
##           Df Deviance   AIC
## + grade    4      17923 17937
## + site     1      18124 18132
## + age      2      18147 18157
## + bone     2      18156 18166
## + lung     2      18167 18177
## + N        2      18180 18190
## + kind     1      18192 18200
## + brain    2      18195 18205
## + CEA      2      18207 18217
## + marry    2      18211 18221
## + size     2      18215 18225
## + radiate  1      18224 18232
## + t        2      18223 18233
## + race     2      18227 18237
## + RX       1      18233 18241
## + income   2      18239 18249
## + gender   1      18242 18250
## <none>             18246 18252
##
## Step:  AIC=17936.69
## y ~ chem + surgery + grade
##
##           Df Deviance   AIC
## + age      2      17824 17842
## + site     1      17832 17848
## + lung     2      17835 17853
## + bone     2      17842 17860
## + CEA      2      17875 17893
## + brain    2      17875 17893
## + marry    2      17888 17906
## + N        2      17894 17912
## + kind     1      17897 17913
## + size     2      17900 17918
## + radiate  1      17904 17920
## + t        2      17903 17921
## + race     2      17906 17924

```

```

## + RX      1      17910 17926
## + income  2      17913 17931
## + gender  1      17920 17936
## <none>    17923 17937
##
## Step:  AIC=17841.87
## y ~ chem + surgery + grade + age
##
##           Df Deviance   AIC
## + lung    2      17737 17759
## + bone    2      17740 17762
## + site    1      17757 17777
## + brain   2      17774 17796
## + CEA     2      17777 17799
## + N       2      17793 17815
## + marry   2      17796 17818
## + kind    1      17798 17818
## + size    2      17800 17822
## + t       2      17806 17828
## + race    2      17807 17829
## + radiate 1      17809 17829
## + RX      1      17815 17835
## + income  2      17814 17836
## + gender  1      17822 17842
## <none>    17824 17842
##
## Step:  AIC=17758.58
## y ~ chem + surgery + grade + age + lung
##
##           Df Deviance   AIC
## + site    1      17661 17685
## + bone    2      17670 17696
## + brain   2      17698 17724
## + CEA     2      17699 17725
## + kind    1      17708 17732
## + N       2      17708 17734
## + marry   2      17709 17735
## + size    2      17713 17739
## + radiate 1      17719 17743
## + race    2      17719 17745
## + t       2      17719 17745
## + income  2      17726 17752
## + RX      1      17730 17754
## + gender  1      17734 17758
## <none>    17737 17759
##
## Step:  AIC=17684.61
## y ~ chem + surgery + grade + age + lung + site
##
##           Df Deviance   AIC
## + bone    2      17592 17620
## + CEA     2      17622 17650
## + brain   2      17622 17650
## + marry   2      17634 17662

```



```

## + kind      1      17636 17662
## + N         2      17635 17663
## + size      2      17636 17664
## + t         2      17647 17675
## + race      2      17649 17677
## + radiate   1      17652 17678
## + income    2      17651 17679
## + RX        1      17654 17680
## <none>      17661 17685
## + gender    1      17660 17686
##
## Step: AIC=17620.05
## y ~ chem + surgery + grade + age + lung + site + bone
##
##           Df Deviance   AIC
## + CEA      2      17554 17586
## + brain     2      17562 17594
## + marry     2      17564 17596
## + kind      1      17570 17600
## + N         2      17568 17600
## + size      2      17569 17601
## + radiate   1      17573 17603
## + t         2      17578 17610
## + race      2      17580 17612
## + income    2      17582 17614
## + RX        1      17586 17616
## <none>      17592 17620
## + gender    1      17591 17621
##
## Step: AIC=17586.03
## y ~ chem + surgery + grade + age + lung + site + bone + CEA
##
##           Df Deviance   AIC
## + brain     2      17524 17560
## + kind      1      17528 17562
## + marry     2      17527 17563
## + N         2      17530 17566
## + size      2      17531 17567
## + radiate   1      17537 17571
## + t         2      17540 17576
## + race      2      17542 17578
## + income    2      17544 17580
## + RX        1      17548 17582
## <none>      17554 17586
## + gender    1      17554 17588
##
## Step: AIC=17560.21
## y ~ chem + surgery + grade + age + lung + site + bone + CEA +
##      brain
##
##           Df Deviance   AIC
## + radiate   1      17496 17534
## + kind      1      17498 17536
## + marry     2      17497 17537

```

```

## + N      2      17500 17540
## + size   2      17500 17540
## + t      2      17511 17551
## + race   2      17512 17552
## + RX     1      17516 17554
## + income 2      17515 17555
## <none>    17524 17560
## + gender 1      17524 17562
##
## Step: AIC=17534.49
## y ~ chem + surgery + grade + age + lung + site + bone + CEA +
##      brain + radiate
##
##           Df Deviance   AIC
## + kind    1      17469 17509
## + marry   2      17469 17511
## + N       2      17473 17515
## + size    2      17473 17515
## + race    2      17484 17526
## + t       2      17486 17528
## + income  2      17487 17529
## + RX      1      17490 17530
## <none>     17496 17534
## + gender  1      17496 17536
##
## Step: AIC=17509.32
## y ~ chem + surgery + grade + age + lung + site + bone + CEA +
##      brain + radiate + kind
##
##           Df Deviance   AIC
## + marry   2      17443 17487
## + N       2      17446 17490
## + size    2      17448 17492
## + race    2      17458 17502
## + t       2      17458 17502
## + income  2      17460 17504
## + RX      1      17462 17504
## <none>     17469 17509
## + gender  1      17469 17511
##
## Step: AIC=17486.86
## y ~ chem + surgery + grade + age + lung + site + bone + CEA +
##      brain + radiate + kind + marry
##
##           Df Deviance   AIC
## + N       2      17418 17466
## + size    2      17422 17470
## + t       2      17432 17480
## + race    2      17434 17482
## + income  2      17434 17482
## + RX      1      17436 17482
## <none>     17443 17487
## + gender  1      17443 17489
##

```

```

## Step: AIC=17466.46
## y ~ chem + surgery + grade + age + lung + site + bone + CEA +
##      brain + radiate + kind + marry + N
##
##      Df Deviance  AIC
## + size    2    17400 17452
## + t        2    17407 17459
## + income   2    17409 17461
## + race     2    17410 17462
## + RX       1    17412 17462
## <none>      17418 17466
## + gender   1    17418 17468
##
## Step: AIC=17451.46
## y ~ chem + surgery + grade + age + lung + site + bone + CEA +
##      brain + radiate + kind + marry + N + size
##
##      Df Deviance  AIC
## + t        2    17388 17444
## + income   2    17390 17446
## + race     2    17391 17447
## + RX       1    17393 17447
## <none>      17400 17452
## + gender   1    17399 17453
##
## Step: AIC=17444.16
## y ~ chem + surgery + grade + age + lung + site + bone + CEA +
##      brain + radiate + kind + marry + N + size + t
##
##      Df Deviance  AIC
## + income   2    17379 17439
## + race     2    17379 17439
## + RX       1    17382 17440
## <none>      17388 17444
## + gender   1    17388 17446
##
## Step: AIC=17439.04
## y ~ chem + surgery + grade + age + lung + site + bone + CEA +
##      brain + radiate + kind + marry + N + size + t + income
##
##      Df Deviance  AIC
## + RX       1    17373 17435
## + race     2    17373 17437
## <none>      17379 17439
## + gender   1    17379 17441
##
## Step: AIC=17435.06
## y ~ chem + surgery + grade + age + lung + site + bone + CEA +
##      brain + radiate + kind + marry + N + size + t + income +
##      RX
##
##      Df Deviance  AIC
## + race     2    17367 17433
## <none>      17373 17435

```

```
## + gender 1 17373 17437
##
## Step: AIC=17433.02
## y ~ chem + surgery + grade + age + lung + site + bone + CEA +
## brain + radiate + kind + marry + N + size + t + income +
## RX + race
##
## Df Deviance AIC
## <none> 17367 17433
## + gender 1 17367 17435
```

```
formula(forwards_aic_2)
```

```
## y ~ chem + surgery + grade + age + lung + site + bone + CEA +
## brain + radiate + kind + marry + N + size + t + income +
## RX + race
```

```
summary(forwards_aic_2)
```

```
##
## Call:
## glm(formula = y ~ chem + surgery + grade + age + lung + site +
## bone + CEA + brain + radiate + kind + marry + N + size +
## t + income + RX + race, family = binomial, data = train_0_1_2)
##
## Deviance Residuals:
## Min 1Q Median 3Q Max
## -2.3193 -0.5653 -0.3714 -0.2365 2.9569
##
## Coefficients:
## Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.98029 0.20489 -4.785 1.71e-06 ***
## chem1 -2.06876 0.04115 -50.273 < 2e-16 ***
## surgery1 -0.64289 0.06397 -10.050 < 2e-16 ***
## grade2 0.17016 0.11672 1.458 0.144888
## grade3 0.85331 0.12270 6.955 3.53e-12 ***
## grade4 0.83770 0.16475 5.085 3.68e-07 ***
## grade9 0.73275 0.11825 6.197 5.77e-10 ***
## age2 -0.13984 0.13689 -1.022 0.306980
## age3 0.24101 0.13505 1.785 0.074335 .
## lung1 0.33761 0.04388 7.694 1.43e-14 ***
## lung2 -0.01207 0.14490 -0.083 0.933615
## site2 0.29180 0.04132 7.062 1.64e-12 ***
## bone1 0.62737 0.07514 8.350 < 2e-16 ***
## bone2 -0.10843 0.20290 -0.534 0.593066
## CEA1 0.46003 0.07798 5.899 3.65e-09 ***
## CEA2 0.37827 0.08272 4.573 4.81e-06 ***
## brain1 1.07341 0.16417 6.538 6.21e-11 ***
## brain2 -0.08913 0.18836 -0.473 0.636098
## radiate1 -0.35211 0.07187 -4.899 9.63e-07 ***
## kind2 0.34564 0.06880 5.024 5.07e-07 ***
## marry2 -0.15446 0.04943 -3.125 0.001779 **
## marry3 0.05008 0.05497 0.911 0.362268
```

```
## N2          0.16775    0.05907    2.840 0.004516 **
## N3          0.22162    0.05265    4.209 2.56e-05 ***
## size2       0.19418    0.04962    3.914 9.09e-05 ***
## size3       0.17278    0.05185    3.332 0.000861 ***
## t2          -0.20661    0.06965   -2.966 0.003015 **
## t3          -0.05971    0.06558   -0.911 0.362517
## income2     -0.07533    0.05364   -1.404 0.160215
## income3     -0.15178    0.06091   -2.492 0.012700 *
## RX1         -0.15886    0.06547   -2.426 0.015249 *
## race2        0.02453    0.05466    0.449 0.653651
## race3       -0.16093    0.06896   -2.334 0.019620 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##    Null deviance: 22383  on 22901  degrees of freedom
## Residual deviance: 17367  on 22869  degrees of freedom
## AIC: 17433
##
## Number of Fisher Scoring iterations: 5
```

```
bothways_aic_2 = step(nothing_2, list(lower=formula(nothing_2),upper=formula(fullmod_2)), direction="bo
```

```
## Start:  AIC=22384.53
## y ~ 1
##
##           Df Deviance   AIC
## + chem      1    18939 18943
## + surgery    1    21665 21669
## + grade      4    21804 21814
## + age        2    21894 21900
## + t          2    21922 21928
## + size       2    22023 22029
## + N          2    22062 22068
## + RX         1    22090 22094
## + marry      2    22168 22174
## + lung       2    22183 22189
## + bone       2    22215 22221
## + CEA        2    22249 22255
## + site       1    22252 22256
## + brain      2    22274 22280
## + radiate    1    22288 22292
## + kind       1    22324 22328
## + income     2    22361 22367
## + race       2    22365 22371
## + gender     1    22368 22372
## <none>       22382 22384
##
## Step:  AIC=18943.4
## y ~ chem
##
##           Df Deviance   AIC
## + surgery    1    18246 18252
```

```

## + grade      4      18304 18316
## + t          2      18505 18513
## + N          2      18648 18656
## + size       2      18651 18659
## + lung       2      18747 18755
## + bone       2      18787 18795
## + RX         1      18797 18803
## + age        2      18826 18834
## + CEA        2      18862 18870
## + brain      2      18869 18877
## + site       1      18893 18899
## + marry      2      18894 18902
## + kind       1      18901 18907
## + race       2      18921 18929
## + radiate    1      18933 18939
## <none>              18939 18943
## + gender     1      18938 18944
## + income     2      18937 18945
## - chem       1      22382 22384
##
## Step: AIC=18251.99
## y ~ chem + surgery
##
##           Df Deviance   AIC
## + grade    4      17923 17937
## + site      1      18124 18132
## + age       2      18147 18157
## + bone      2      18156 18166
## + lung      2      18167 18177
## + N         2      18180 18190
## + kind      1      18192 18200
## + brain     2      18195 18205
## + CEA       2      18207 18217
## + marry     2      18211 18221
## + size      2      18215 18225
## + radiate   1      18224 18232
## + t         2      18223 18233
## + race      2      18227 18237
## + RX        1      18233 18241
## + income    2      18239 18249
## + gender    1      18242 18250
## <none>              18246 18252
## - surgery   1      18939 18943
## - chem      1      21665 21669
##
## Step: AIC=17936.69
## y ~ chem + surgery + grade
##
##           Df Deviance   AIC
## + age       2      17824 17842
## + site      1      17832 17848
## + lung      2      17835 17853
## + bone      2      17842 17860
## + CEA       2      17875 17893

```

```

## + brain      2      17875 17893
## + marry      2      17888 17906
## + N          2      17894 17912
## + kind       1      17897 17913
## + size       2      17900 17918
## + radiate    1      17904 17920
## + t          2      17903 17921
## + race       2      17906 17924
## + RX         1      17910 17926
## + income     2      17913 17931
## + gender     1      17920 17936
## <none>              17923 17937
## - grade      4      18246 18252
## - surgery    1      18304 18316
## - chem       1      21358 21370
##
## Step:  AIC=17841.87
## y ~ chem + surgery + grade + age
##
##           Df Deviance   AIC
## + lung      2      17737 17759
## + bone      2      17740 17762
## + site      1      17757 17777
## + brain     2      17774 17796
## + CEA       2      17777 17799
## + N         2      17793 17815
## + marry     2      17796 17818
## + kind      1      17798 17818
## + size      2      17800 17822
## + t         2      17806 17828
## + race      2      17807 17829
## + radiate   1      17809 17829
## + RX        1      17815 17835
## + income    2      17814 17836
## + gender    1      17822 17842
## <none>              17824 17842
## - age       2      17923 17937
## - grade     4      18147 18157
## - surgery   1      18195 18211
## - chem      1      20910 20926
##
## Step:  AIC=17758.58
## y ~ chem + surgery + grade + age + lung
##
##           Df Deviance   AIC
## + site      1      17661 17685
## + bone      2      17670 17696
## + brain     2      17698 17724
## + CEA       2      17699 17725
## + kind      1      17708 17732
## + N         2      17708 17734
## + marry     2      17709 17735
## + size      2      17713 17739
## + radiate   1      17719 17743

```

```

## + race      2      17719 17745
## + t         2      17719 17745
## + income    2      17726 17752
## + RX        1      17730 17754
## + gender    1      17734 17758
## <none>      17737 17759
## - lung      2      17824 17842
## - age       2      17835 17853
## - surgery   1      18041 18061
## - grade     4      18069 18083
## - chem      1      20837 20857
##
## Step:  AIC=17684.61
## y ~ chem + surgery + grade + age + lung + site
##
##           Df Deviance   AIC
## + bone      2      17592 17620
## + CEA        2      17622 17650
## + brain      2      17622 17650
## + marry      2      17634 17662
## + kind       1      17636 17662
## + N          2      17635 17663
## + size       2      17636 17664
## + t          2      17647 17675
## + race       2      17649 17677
## + radiate    1      17652 17678
## + income     2      17651 17679
## + RX         1      17654 17680
## <none>      17661 17685
## + gender     1      17660 17686
## - age        2      17734 17754
## - site       1      17737 17759
## - lung       2      17757 17777
## - grade      4      17964 17980
## - surgery    1      17998 18020
## - chem       1      20714 20736
##
## Step:  AIC=17620.05
## y ~ chem + surgery + grade + age + lung + site + bone
##
##           Df Deviance   AIC
## + CEA        2      17554 17586
## + brain      2      17562 17594
## + marry      2      17564 17596
## + kind       1      17570 17600
## + N          2      17568 17600
## + size       2      17569 17601
## + radiate    1      17573 17603
## + t          2      17578 17610
## + race       2      17580 17612
## + income     2      17582 17614
## + RX         1      17586 17616
## <none>      17592 17620
## + gender     1      17591 17621

```



```

## - bone      2      17661 17685
## - age       2      17668 17692
## - lung      2      17670 17694
## - site      1      17670 17696
## - grade     4      17886 17906
## - surgery   1      17907 17933
## - chem      1      20642 20668
##
## Step: AIC=17586.03
## y ~ chem + surgery + grade + age + lung + site + bone + CEA
##
##           Df Deviance   AIC
## + brain    2      17524 17560
## + kind      1      17528 17562
## + marry     2      17527 17563
## + N         2      17530 17566
## + size      2      17531 17567
## + radiate   1      17537 17571
## + t         2      17540 17576
## + race      2      17542 17578
## + income    2      17544 17580
## + RX        1      17548 17582
## <none>      17554 17586
## + gender    1      17554 17588
## - CEA       2      17592 17620
## - bone      2      17622 17650
## - lung      2      17624 17652
## - age       2      17629 17657
## - site      1      17633 17663
## - grade     4      17855 17879
## - surgery   1      17856 17886
## - chem      1      20552 20582
##
## Step: AIC=17560.21
## y ~ chem + surgery + grade + age + lung + site + bone + CEA +
##      brain
##
##           Df Deviance   AIC
## + radiate   1      17496 17534
## + kind      1      17498 17536
## + marry     2      17497 17537
## + N         2      17500 17540
## + size      2      17500 17540
## + t         2      17511 17551
## + race      2      17512 17552
## + RX        1      17516 17554
## + income    2      17515 17555
## <none>      17524 17560
## + gender    1      17524 17562
## - brain     2      17554 17586
## - CEA       2      17562 17594
## - bone      2      17584 17616
## - lung      2      17586 17618
## - age       2      17601 17633

```

```

## - site      1      17603 17637
## - grade     4      17822 17850
## - surgery   1      17821 17855
## - chem      1      20508 20542
##
## Step: AIC=17534.49
## y ~ chem + surgery + grade + age + lung + site + bone + CEA +
##      brain + radiate
##
##           Df Deviance   AIC
## + kind      1      17469 17509
## + marry     2      17469 17511
## + N         2      17473 17515
## + size      2      17473 17515
## + race      2      17484 17526
## + t         2      17486 17528
## + income    2      17487 17529
## + RX        1      17490 17530
## <none>      17496 17534
## + gender    1      17496 17536
## - radiate   1      17524 17560
## - CEA       2      17532 17566
## - brain     2      17537 17571
## - lung      2      17558 17592
## - site      1      17560 17596
## - bone      2      17568 17602
## - age       2      17571 17605
## - grade     4      17792 17822
## - surgery   1      17798 17834
## - chem      1      20396 20432
##
## Step: AIC=17509.32
## y ~ chem + surgery + grade + age + lung + site + bone + CEA +
##      brain + radiate + kind
##
##           Df Deviance   AIC
## + marry     2      17443 17487
## + N         2      17446 17490
## + size      2      17448 17492
## + race      2      17458 17502
## + t         2      17458 17502
## + income    2      17460 17504
## + RX        1      17462 17504
## <none>      17469 17509
## + gender    1      17469 17511
## - kind      1      17496 17534
## - radiate   1      17498 17536
## - CEA       2      17509 17545
## - brain     2      17510 17546
## - site      1      17528 17566
## - lung      2      17533 17569
## - bone      2      17539 17575
## - age       2      17544 17580
## - grade     4      17741 17773

```

```

## - surgery 1 17776 17814
## - chem 1 20361 20399
##
## Step: AIC=17486.86
## y ~ chem + surgery + grade + age + lung + site + bone + CEA +
## brain + radiate + kind + marry
##
## Df Deviance AIC
## + N 2 17418 17466
## + size 2 17422 17470
## + t 2 17432 17480
## + race 2 17434 17482
## + income 2 17434 17482
## + RX 1 17436 17482
## <none> 17443 17487
## + gender 1 17443 17489
## - marry 2 17469 17509
## - kind 1 17469 17511
## - radiate 1 17471 17513
## - CEA 2 17481 17521
## - brain 2 17483 17523
## - site 1 17501 17543
## - lung 2 17506 17546
## - age 2 17513 17553
## - bone 2 17514 17554
## - grade 4 17715 17751
## - surgery 1 17744 17786
## - chem 1 20211 20253
##
## Step: AIC=17466.46
## y ~ chem + surgery + grade + age + lung + site + bone + CEA +
## brain + radiate + kind + marry + N
##
## Df Deviance AIC
## + size 2 17400 17452
## + t 2 17407 17459
## + income 2 17409 17461
## + race 2 17410 17462
## + RX 1 17412 17462
## <none> 17418 17466
## + gender 1 17418 17468
## - N 2 17443 17487
## - marry 2 17446 17490
## - kind 1 17445 17491
## - radiate 1 17447 17493
## - CEA 2 17456 17500
## - brain 2 17459 17503
## - site 1 17474 17520
## - lung 2 17480 17524
## - bone 2 17488 17532
## - age 2 17491 17535
## - grade 4 17660 17700
## - surgery 1 17656 17702
## - chem 1 20182 20228

```

```

##
## Step: AIC=17451.46
## y ~ chem + surgery + grade + age + lung + site + bone + CEA +
##      brain + radiate + kind + marry + N + size
##
##      Df Deviance  AIC
## + t      2    17388 17444
## + income  2    17390 17446
## + race    2    17391 17447
## + RX      1    17393 17447
## <none>      17400 17452
## + gender  1    17399 17453
## - size    2    17418 17466
## - N       2    17422 17470
## - marry   2    17426 17474
## - kind    1    17424 17474
## - radiate 1    17428 17478
## - CEA     2    17437 17485
## - brain   2    17441 17489
## - site    1    17456 17506
## - lung    2    17460 17508
## - bone    2    17469 17517
## - age     2    17473 17521
## - surgery 1    17584 17634
## - grade   4    17636 17680
## - chem    1    20139 20189
##
## Step: AIC=17444.16
## y ~ chem + surgery + grade + age + lung + site + bone + CEA +
##      brain + radiate + kind + marry + N + size + t
##
##      Df Deviance  AIC
## + income  2    17379 17439
## + race    2    17379 17439
## + RX      1    17382 17440
## <none>      17388 17444
## + gender  1    17388 17446
## - t       2    17400 17452
## - size    2    17407 17459
## - N       2    17412 17464
## - marry   2    17414 17466
## - kind    1    17413 17467
## - radiate 1    17414 17468
## - CEA     2    17426 17478
## - brain   2    17428 17480
## - site    1    17442 17496
## - lung    2    17449 17501
## - bone    2    17457 17509
## - age     2    17461 17513
## - surgery 1    17507 17561
## - grade   4    17625 17673
## - chem    1    20130 20184
##
## Step: AIC=17439.04

```

```

## y ~ chem + surgery + grade + age + lung + site + bone + CEA +
##   brain + radiate + kind + marry + N + size + t + income
##
##           Df Deviance   AIC
## + RX       1    17373 17435
## + race      2    17373 17437
## <none>      1    17379 17439
## + gender    1    17379 17441
## - income    2    17388 17444
## - t         2    17390 17446
## - size      2    17398 17454
## - marry     2    17404 17460
## - N         2    17404 17460
## - kind      1    17404 17462
## - radiate   1    17405 17463
## - CEA       2    17416 17472
## - brain     2    17419 17475
## - site      1    17433 17491
## - lung      2    17440 17496
## - bone      2    17448 17504
## - age       2    17453 17509
## - surgery   1    17499 17557
## - grade     4    17617 17669
## - chem      1    20107 20165
##
## Step: AIC=17435.06
## y ~ chem + surgery + grade + age + lung + site + bone + CEA +
##   brain + radiate + kind + marry + N + size + t + income +
##   RX
##
##           Df Deviance   AIC
## + race      2    17367 17433
## <none>      1    17373 17435
## + gender    1    17373 17437
## - RX        1    17379 17439
## - income    2    17382 17440
## - t         2    17384 17442
## - size      2    17392 17450
## - marry     2    17397 17455
## - N         2    17397 17455
## - radiate   1    17398 17458
## - kind      1    17398 17458
## - CEA       2    17410 17468
## - brain     2    17415 17473
## - site      1    17426 17486
## - lung      2    17432 17490
## - bone      2    17441 17499
## - age       2    17445 17503
## - surgery   1    17476 17536
## - grade     4    17611 17665
## - chem      1    20080 20140
##
## Step: AIC=17433.02
## y ~ chem + surgery + grade + age + lung + site + bone + CEA +

```

```
##      brain + radiate + kind + marry + N + size + t + income +
##      RX + race
##
##           Df Deviance   AIC
## <none>           17367 17433
## + gender      1      17367 17435
## - race        2      17373 17435
## - income      2      17373 17435
## - RX          1      17373 17437
## - t           2      17378 17440
## - size        2      17386 17448
## - marry       2      17389 17451
## - N           2      17391 17453
## - kind        1      17392 17456
## - radiate     1      17392 17456
## - CEA         2      17404 17466
## - brain       2      17409 17471
## - site        1      17417 17481
## - lung        2      17427 17489
## - bone        2      17436 17498
## - age         2      17438 17500
## - surgery     1      17470 17534
## - grade       4      17604 17662
## - chem        1      20079 20143
```

```
formula(bothways_aic_2)
```

```
## y ~ chem + surgery + grade + age + lung + site + bone + CEA +
##      brain + radiate + kind + marry + N + size + t + income +
##      RX + race
```

```
summary(bothways_aic_2)
```

```
##
## Call:
## glm(formula = y ~ chem + surgery + grade + age + lung + site +
##      bone + CEA + brain + radiate + kind + marry + N + size +
##      t + income + RX + race, family = binomial, data = train_0_1_2)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.3193  -0.5653  -0.3714  -0.2365   2.9569
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.98029    0.20489  -4.785 1.71e-06 ***
## chem1       -2.06876    0.04115 -50.273 < 2e-16 ***
## surgery1    -0.64289    0.06397 -10.050 < 2e-16 ***
## grade2       0.17016    0.11672   1.458 0.144888
## grade3       0.85331    0.12270   6.955 3.53e-12 ***
## grade4       0.83770    0.16475   5.085 3.68e-07 ***
## grade9       0.73275    0.11825   6.197 5.77e-10 ***
## age2        -0.13984    0.13689  -1.022 0.306980
```

```
## age3      0.24101    0.13505    1.785 0.074335 .
## lung1     0.33761    0.04388    7.694 1.43e-14 ***
## lung2    -0.01207    0.14490   -0.083 0.933615
## site2     0.29180    0.04132    7.062 1.64e-12 ***
## bone1     0.62737    0.07514    8.350 < 2e-16 ***
## bone2    -0.10843    0.20290   -0.534 0.593066
## CEA1      0.46003    0.07798    5.899 3.65e-09 ***
## CEA2      0.37827    0.08272    4.573 4.81e-06 ***
## brain1    1.07341    0.16417    6.538 6.21e-11 ***
## brain2   -0.08913    0.18836   -0.473 0.636098
## radiate1  -0.35211    0.07187   -4.899 9.63e-07 ***
## kind2     0.34564    0.06880    5.024 5.07e-07 ***
## marry2    -0.15446    0.04943   -3.125 0.001779 **
## marry3     0.05008    0.05497    0.911 0.362268
## N2        0.16775    0.05907    2.840 0.004516 **
## N3        0.22162    0.05265    4.209 2.56e-05 ***
## size2     0.19418    0.04962    3.914 9.09e-05 ***
## size3     0.17278    0.05185    3.332 0.000861 ***
## t2        -0.20661    0.06965   -2.966 0.003015 **
## t3        -0.05971    0.06558   -0.911 0.362517
## income2   -0.07533    0.05364   -1.404 0.160215
## income3   -0.15178    0.06091   -2.492 0.012700 *
## RX1       -0.15886    0.06547   -2.426 0.015249 *
## race2     0.02453    0.05466    0.449 0.653651
## race3    -0.16093    0.06896   -2.334 0.019620 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##    Null deviance: 22383  on 22901  degrees of freedom
## Residual deviance: 17367  on 22869  degrees of freedom
## AIC: 17433
##
## Number of Fisher Scoring iterations: 5
```

0+1+2 multivariate forward/backward/both BIC

```
backwards_bic_2 = step(fullmod_2, k=log(nrow(train_0_1_2)))
```

```
## Start:  AIC=17708.34
## y ~ gender + race + age + size + marry + income + site + grade +
##      kind + t + N + surgery + RX + radiate + chem + CEA + bone +
##      brain + lung
##
##           Df Deviance   AIC
## - race      2    17373 17694
## - income    2    17373 17695
## - gender    1    17367 17698
## - t         2    17378 17699
## - RX        1    17373 17704
## - size      2    17386 17707
```

```

## <none>          17367 17708
## - marry        2    17388 17710
## - N            2    17391 17712
## - kind         1    17392 17723
## - radiate      1    17392 17723
## - CEA          2    17404 17726
## - brain        2    17409 17730
## - site         1    17416 17748
## - lung         2    17427 17748
## - bone         2    17436 17757
## - age          2    17438 17760
## - surgery      1    17470 17801
## - grade        4    17604 17905
## - chem         1    20077 20409
##
## Step:  AIC=17694.31
## y ~ gender + age + size + marry + income + site + grade + kind +
##      t + N + surgery + RX + radiate + chem + CEA + bone + brain +
##      lung
##
##           Df Deviance   AIC
## - income   2    17382 17684
## - gender    1    17373 17684
## - t         2    17384 17685
## - RX        1    17379 17690
## - size      2    17392 17693
## <none>      17373 17694
## - marry     2    17397 17698
## - N         2    17397 17698
## - kind      1    17398 17709
## - radiate   1    17398 17709
## - CEA       2    17410 17711
## - brain     2    17415 17716
## - lung      2    17432 17734
## - site      1    17426 17737
## - bone      2    17441 17743
## - age       2    17445 17746
## - surgery   1    17476 17787
## - grade     4    17611 17892
## - chem      1    20078 20389
##
## Step:  AIC=17683.62
## y ~ gender + age + size + marry + site + grade + kind + t + N +
##      surgery + RX + radiate + chem + CEA + bone + brain + lung
##
##           Df Deviance   AIC
## - gender    1    17382 17674
## - t         2    17393 17674
## - RX        1    17388 17679
## - size      2    17401 17682
## <none>      17382 17684
## - N         2    17406 17687
## - marry     2    17407 17688
## - radiate   1    17407 17698

```



```

## - kind      1      17408 17699
## - CEA       2      17420 17701
## - brain     2      17424 17705
## - lung      2      17442 17723
## - site      1      17436 17727
## - bone      2      17451 17732
## - age       2      17453 17734
## - surgery   1      17485 17776
## - grade     4      17619 17880
## - chem      1      20102 20393
##
## Step:  AIC=17673.59
## y ~ age + size + marry + site + grade + kind + t + N + surgery +
##      RX + radiate + chem + CEA + bone + brain + lung
##
##           Df Deviance   AIC
## - t        2      17393 17664
## - RX       1      17388 17669
## - size     2      17401 17672
## <none>           17382 17674
## - N        2      17406 17677
## - marry    2      17408 17679
## - radiate  1      17407 17688
## - kind     1      17408 17689
## - CEA      2      17420 17691
## - brain    2      17424 17695
## - lung     2      17442 17713
## - site     1      17436 17717
## - bone     2      17451 17722
## - age      2      17453 17724
## - surgery  1      17485 17766
## - grade    4      17619 17870
## - chem     1      20103 20384
##
## Step:  AIC=17664.44
## y ~ age + size + marry + site + grade + kind + N + surgery +
##      RX + radiate + chem + CEA + bone + brain + lung
##
##           Df Deviance   AIC
## - RX       1      17400 17660
## - size     2      17412 17664
## <none>           17393 17664
## - N        2      17416 17667
## - marry    2      17419 17670
## - kind     1      17418 17679
## - CEA      2      17430 17681
## - radiate  1      17421 17682
## - brain    2      17436 17687
## - lung     2      17453 17704
## - site     1      17449 17710
## - bone     2      17462 17713
## - age      2      17464 17715
## - surgery  1      17551 17812
## - grade    4      17630 17860

```

```

## - chem      1      20112 20373
##
## Step: AIC=17660.48
## y ~ age + size + marry + site + grade + kind + N + surgery +
##       radiate + chem + CEA + bone + brain + lung
##
##           Df Deviance   AIC
## - size      2      17418 17659
## <none>                17400 17660
## - N          2      17422 17663
## - marry      2      17426 17666
## - kind       1      17424 17675
## - CEA        2      17437 17678
## - radiate    1      17428 17679
## - brain      2      17441 17682
## - lung       2      17460 17701
## - site       1      17456 17707
## - bone       2      17469 17710
## - age        2      17473 17714
## - surgery    1      17584 17835
## - grade      4      17636 17856
## - chem       1      20139 20390
##
## Step: AIC=17659.39
## y ~ age + marry + site + grade + kind + N + surgery + radiate +
##       chem + CEA + bone + brain + lung
##
##           Df Deviance   AIC
## <none>                17418 17659
## - N          2      17443 17664
## - marry      2      17446 17666
## - kind       1      17445 17676
## - CEA        2      17456 17677
## - radiate    1      17447 17678
## - brain      2      17459 17680
## - lung       2      17480 17700
## - site       1      17474 17705
## - bone       2      17488 17709
## - age        2      17491 17712
## - grade      4      17660 17860
## - surgery    1      17656 17886
## - chem       1      20182 20413

```

```
formula(backwards_bic_2)
```

```

## y ~ age + marry + site + grade + kind + N + surgery + radiate +
##       chem + CEA + bone + brain + lung

```

```
summary(backwards_bic_2)
```

```

##
## Call:
## glm(formula = y ~ age + marry + site + grade + kind + N + surgery +

```

```
##      radiate + chem + CEA + bone + brain + lung, family = binomial,
##      data = train_0_1_2)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.2349  -0.5672  -0.3750  -0.2420   2.9351
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.04127    0.19138  -5.441 5.30e-08 ***
## age2        -0.12229    0.13685  -0.894 0.371509
## age3         0.25838    0.13502   1.914 0.055667 .
## marry2      -0.17675    0.04867  -3.632 0.000281 ***
## marry3       0.04426    0.05468   0.809 0.418340
## site2        0.30689    0.04097   7.491 6.85e-14 ***
## grade2       0.15977    0.11651   1.371 0.170287
## grade3       0.84484    0.12242   6.901 5.16e-12 ***
## grade4       0.82979    0.16430   5.050 4.41e-07 ***
## grade9       0.73058    0.11804   6.189 6.04e-10 ***
## kind2        0.35800    0.06858   5.220 1.79e-07 ***
## N2           0.14462    0.05829   2.481 0.013101 *
## N3           0.22914    0.05150   4.449 8.62e-06 ***
## surgery1     -0.80805    0.05332 -15.153 < 2e-16 ***
## radiate1     -0.37137    0.07149  -5.195 2.05e-07 ***
## chem1        -2.07841    0.04095 -50.756 < 2e-16 ***
## CEA1          0.46295    0.07780   5.950 2.67e-09 ***
## CEA2          0.39736    0.08241   4.822 1.42e-06 ***
## bone1         0.63218    0.07500   8.429 < 2e-16 ***
## bone2        -0.10853    0.20190  -0.538 0.590904
## brain1        1.04800    0.16344   6.412 1.44e-10 ***
## brain2       -0.06232    0.18733  -0.333 0.739376
## lung1         0.34097    0.04379   7.786 6.90e-15 ***
## lung2        -0.01549    0.14488  -0.107 0.914849
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 22383  on 22901  degrees of freedom
## Residual deviance: 17418  on 22878  degrees of freedom
## AIC: 17466
##
## Number of Fisher Scoring iterations: 5
```

```
forwards_bic_2 = step(nothing_2, scope=list(lower=formula(nothing_2),upper=formula(fullmod_2)), direction="forward")
```

```
## Start:  AIC=22392.57
## y ~ 1
##
##           Df Deviance   AIC
## + chem      1    18939 18960
## + surgery   1    21665 21685
## + grade     4    21804 21854
## + age       2    21894 21925
```

```

## + t      2      21922 21952
## + size   2      22023 22053
## + N      2      22062 22092
## + RX     1      22090 22110
## + marry  2      22168 22198
## + lung   2      22183 22213
## + bone   2      22215 22245
## + site   1      22252 22272
## + CEA    2      22249 22279
## + brain  2      22274 22304
## + radiate 1      22288 22308
## + kind   1      22324 22344
## + gender 1      22368 22389
## + income 2      22361 22391
## <none>           22382 22393
## + race   2      22365 22395
##
## Step: AIC=18959.48
## y ~ chem
##
##           Df Deviance   AIC
## + surgery 1      18246 18276
## + grade   4      18304 18365
## + t       2      18505 18545
## + N       2      18648 18688
## + size    2      18651 18691
## + lung    2      18747 18787
## + RX      1      18797 18827
## + bone    2      18787 18828
## + age     2      18826 18866
## + CEA     2      18862 18902
## + brain   2      18869 18909
## + site    1      18893 18923
## + kind    1      18901 18931
## + marry   2      18894 18934
## <none>           18939 18960
## + race    2      18921 18961
## + radiate 1      18933 18963
## + gender  1      18938 18968
## + income  2      18937 18977
##
## Step: AIC=18276.11
## y ~ chem + surgery
##
##           Df Deviance   AIC
## + grade   4      17923 17993
## + site    1      18124 18165
## + age     2      18147 18197
## + bone    2      18156 18206
## + lung    2      18167 18218
## + N       2      18180 18230
## + kind    1      18192 18233
## + brain   2      18195 18245
## + CEA     2      18207 18257

```

```

## + marry      2      18211 18261
## + radiate    1      18224 18264
## + size       2      18215 18265
## + RX         1      18233 18273
## + t          2      18223 18273
## <none>       18246 18276
## + race       2      18227 18278
## + gender     1      18242 18282
## + income     2      18239 18289
##
## Step:  AIC=17992.96
## y ~ chem + surgery + grade
##
##           Df Deviance   AIC
## + site     1      17832 17913
## + age      2      17824 17914
## + lung     2      17835 17925
## + bone     2      17842 17932
## + CEA      2      17875 17965
## + brain    2      17875 17966
## + kind     1      17897 17977
## + marry    2      17888 17978
## + radiate  1      17904 17984
## + N        2      17894 17984
## + RX       1      17910 17990
## + size     2      17900 17990
## <none>     17923 17993
## + t        2      17903 17994
## + race     2      17906 17996
## + gender   1      17920 18000
## + income   2      17913 18004
##
## Step:  AIC=17912.69
## y ~ chem + surgery + grade + site
##
##           Df Deviance   AIC
## + lung     2      17734 17835
## + bone     2      17748 17848
## + age      2      17757 17857
## + CEA      2      17783 17884
## + brain    2      17784 17884
## + marry    2      17801 17901
## + kind     1      17812 17902
## + N        2      17806 17907
## + size     2      17809 17910
## + RX       1      17821 17912
## <none>     17832 17913
## + radiate  1      17825 17915
## + t        2      17817 17918
## + gender   1      17832 17922
## + race     2      17822 17922
## + income   2      17824 17924
##
## Step:  AIC=17834.75

```

```

## y ~ chem + surgery + grade + site + lung
##
##           Df Deviance   AIC
## + age      2    17661 17781
## + bone      2    17668 17788
## + CEA       2    17694 17815
## + brain     2    17698 17818
## + kind      1    17711 17821
## + marry     2    17703 17824
## + N         2    17711 17831
## + size      2    17712 17832
## <none>      17734 17835
## + radiate   1    17725 17835
## + RX        1    17726 17836
## + t         2    17720 17840
## + gender    1    17734 17844
## + race      2    17724 17844
## + income    2    17725 17846
##
## Step: AIC=17781.08
## y ~ chem + surgery + grade + site + lung + age
##
##           Df Deviance   AIC
## + bone      2    17592 17733
## + CEA       2    17622 17762
## + brain     2    17622 17763
## + kind      1    17636 17767
## + marry     2    17634 17774
## + N         2    17635 17776
## + size      2    17636 17777
## <none>      17661 17781
## + radiate   1    17652 17783
## + RX        1    17654 17785
## + t         2    17647 17787
## + race      2    17649 17790
## + gender    1    17660 17791
## + income    2    17651 17791
##
## Step: AIC=17732.59
## y ~ chem + surgery + grade + site + lung + age + bone
##
##           Df Deviance   AIC
## + CEA       2    17554 17715
## + kind      1    17570 17720
## + brain     2    17562 17723
## + radiate   1    17573 17724
## + marry     2    17564 17724
## + N         2    17568 17728
## + size      2    17569 17729
## <none>      17592 17733
## + RX        1    17586 17736
## + t         2    17578 17738
## + race      2    17580 17741
## + gender    1    17591 17742

```

```

## + income 2 17582 17743
##
## Step: AIC=17714.65
## y ~ chem + surgery + grade + site + lung + age + bone + CEA
##
##           Df Deviance  AIC
## + kind 1 17528 17698
## + brain 2 17524 17705
## + marry 2 17527 17707
## + radiate 1 17537 17708
## + N 2 17530 17711
## + size 2 17531 17712
## <none> 17554 17715
## + RX 1 17548 17719
## + t 2 17540 17721
## + race 2 17542 17723
## + gender 1 17554 17724
## + income 2 17544 17725
##
## Step: AIC=17698.2
## y ~ chem + surgery + grade + site + lung + age + bone + CEA +
## kind
##
##           Df Deviance  AIC
## + brain 2 17498 17688
## + radiate 1 17510 17691
## + marry 2 17501 17692
## + N 2 17504 17694
## + size 2 17506 17697
## <none> 17528 17698
## + RX 1 17522 17702
## + t 2 17513 17704
## + race 2 17516 17707
## + gender 1 17527 17708
## + income 2 17518 17709
##
## Step: AIC=17688.43
## y ~ chem + surgery + grade + site + lung + age + bone + CEA +
## kind + brain
##
##           Df Deviance  AIC
## + radiate 1 17469 17670
## + marry 2 17471 17682
## + N 2 17474 17685
## + size 2 17476 17686
## <none> 17498 17688
## + RX 1 17490 17691
## + t 2 17484 17694
## + race 2 17486 17697
## + gender 1 17498 17698
## + income 2 17488 17699
##
## Step: AIC=17670.1
## y ~ chem + surgery + grade + site + lung + age + bone + CEA +

```

```

##      kind + brain + radiate
##
##      Df Deviance   AIC
## + marry    2    17443 17664
## + N        2    17446 17666
## + size     2    17448 17668
## <none>      2    17469 17670
## + RX       1    17462 17673
## + race     2    17458 17679
## + t        2    17458 17679
## + gender   1    17469 17680
## + income   2    17460 17680
##
## Step: AIC=17663.72
## y ~ chem + surgery + grade + site + lung + age + bone + CEA +
##      kind + brain + radiate + marry
##
##      Df Deviance   AIC
## + N        2    17418 17659
## + size     2    17422 17663
## <none>      2    17443 17664
## + RX       1    17436 17667
## + t        2    17432 17673
## + gender   1    17443 17674
## + race     2    17434 17675
## + income   2    17434 17675
##
## Step: AIC=17659.39
## y ~ chem + surgery + grade + site + lung + age + bone + CEA +
##      kind + brain + radiate + marry + N
##
##      Df Deviance   AIC
## <none>      2    17418 17659
## + size     2    17400 17660
## + RX       1    17412 17664
## + t        2    17407 17668
## + gender   1    17418 17669
## + income   2    17409 17670
## + race     2    17410 17670

```

```
formula(forwards_bic_2)
```

```

## y ~ chem + surgery + grade + site + lung + age + bone + CEA +
##      kind + brain + radiate + marry + N

```

```
summary(forwards_bic_2)
```

```

##
## Call:
## glm(formula = y ~ chem + surgery + grade + site + lung + age +
##      bone + CEA + kind + brain + radiate + marry + N, family = binomial,
##      data = train_0_1_2)
##

```



```
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.2349  -0.5672  -0.3750  -0.2420   2.9351
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.04127    0.19138  -5.441 5.30e-08 ***
## chem1       -2.07841    0.04095 -50.756 < 2e-16 ***
## surgery1    -0.80805    0.05332 -15.153 < 2e-16 ***
## grade2       0.15977    0.11651   1.371 0.170287
## grade3       0.84484    0.12242   6.901 5.16e-12 ***
## grade4       0.82979    0.16430   5.050 4.41e-07 ***
## grade9       0.73058    0.11804   6.189 6.04e-10 ***
## site2        0.30689    0.04097   7.491 6.85e-14 ***
## lung1        0.34097    0.04379   7.786 6.90e-15 ***
## lung2       -0.01549    0.14488  -0.107 0.914849
## age2        -0.12229    0.13685  -0.894 0.371509
## age3         0.25838    0.13502   1.914 0.055667 .
## bone1        0.63218    0.07500   8.429 < 2e-16 ***
## bone2       -0.10853    0.20190  -0.538 0.590904
## CEA1         0.46295    0.07780   5.950 2.67e-09 ***
## CEA2         0.39736    0.08241   4.822 1.42e-06 ***
## kind2        0.35800    0.06858   5.220 1.79e-07 ***
## brain1       1.04800    0.16344   6.412 1.44e-10 ***
## brain2      -0.06232    0.18733  -0.333 0.739376
## radiate1    -0.37137    0.07149  -5.195 2.05e-07 ***
## marry2      -0.17675    0.04867  -3.632 0.000281 ***
## marry3       0.04426    0.05468   0.809 0.418340
## N2          0.14462    0.05829   2.481 0.013101 *
## N3          0.22914    0.05150   4.449 8.62e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 22383  on 22901  degrees of freedom
## Residual deviance: 17418  on 22878  degrees of freedom
## AIC: 17466
##
## Number of Fisher Scoring iterations: 5
```

```
bothways_bic_2 = step(nothing_2, list(lower=formula(nothing_2),upper=formula(fullmod_2)), direction="bo
```

```
## Start:  AIC=22392.57
## y ~ 1
##
##              Df Deviance   AIC
## + chem        1    18939 18960
## + surgery      1    21665 21685
## + grade        4    21804 21854
## + age          2    21894 21925
## + t            2    21922 21952
## + size         2    22023 22053
## + N            2    22062 22092
```

```

## + RX      1      22090 22110
## + marry   2      22168 22198
## + lung    2      22183 22213
## + bone    2      22215 22245
## + site    1      22252 22272
## + CEA     2      22249 22279
## + brain   2      22274 22304
## + radiate 1      22288 22308
## + kind    1      22324 22344
## + gender  1      22368 22389
## + income  2      22361 22391
## <none>    22382 22393
## + race    2      22365 22395
##
## Step: AIC=18959.48
## y ~ chem
##
##           Df Deviance   AIC
## + surgery 1      18246 18276
## + grade   4      18304 18365
## + t        2      18505 18545
## + N        2      18648 18688
## + size     2      18651 18691
## + lung     2      18747 18787
## + RX       1      18797 18827
## + bone     2      18787 18828
## + age      2      18826 18866
## + CEA      2      18862 18902
## + brain    2      18869 18909
## + site     1      18893 18923
## + kind     1      18901 18931
## + marry    2      18894 18934
## <none>     18939 18960
## + race     2      18921 18961
## + radiate  1      18933 18963
## + gender   1      18938 18968
## + income   2      18937 18977
## - chem     1      22382 22393
##
## Step: AIC=18276.11
## y ~ chem + surgery
##
##           Df Deviance   AIC
## + grade   4      17923 17993
## + site     1      18124 18165
## + age      2      18147 18197
## + bone     2      18156 18206
## + lung     2      18167 18218
## + N        2      18180 18230
## + kind     1      18192 18233
## + brain    2      18195 18245
## + CEA      2      18207 18257
## + marry    2      18211 18261
## + radiate  1      18224 18264

```

```

## + size      2      18215 18265
## + RX        1      18233 18273
## + t         2      18223 18273
## <none>      18246 18276
## + race      2      18227 18278
## + gender    1      18242 18282
## + income    2      18239 18289
## - surgery   1      18939 18960
## - chem      1      21665 21685
##
## Step:  AIC=17992.96
## y ~ chem + surgery + grade
##
##           Df Deviance   AIC
## + site     1     17832 17913
## + age      2     17824 17914
## + lung     2     17835 17925
## + bone     2     17842 17932
## + CEA      2     17875 17965
## + brain    2     17875 17966
## + kind     1     17897 17977
## + marry    2     17888 17978
## + radiate  1     17904 17984
## + N        2     17894 17984
## + RX       1     17910 17990
## + size     2     17900 17990
## <none>     17923 17993
## + t        2     17903 17994
## + race     2     17906 17996
## + gender   1     17920 18000
## + income   2     17913 18004
## - grade    4     18246 18276
## - surgery  1     18304 18365
## - chem     1     21358 21418
##
## Step:  AIC=17912.69
## y ~ chem + surgery + grade + site
##
##           Df Deviance   AIC
## + lung     2     17734 17835
## + bone     2     17748 17848
## + age      2     17757 17857
## + CEA      2     17783 17884
## + brain    2     17784 17884
## + marry    2     17801 17901
## + kind     1     17812 17902
## + N        2     17806 17907
## + size     2     17809 17910
## + RX       1     17821 17912
## <none>     17832 17913
## + radiate  1     17825 17915
## + t        2     17817 17918
## + gender   1     17832 17922
## + race     2     17822 17922

```

```

## + income 2 17824 17924
## - site 1 17923 17993
## - grade 4 18124 18165
## - surgery 1 18256 18326
## - chem 1 21168 21238
##
## Step: AIC=17834.75
## y ~ chem + surgery + grade + site + lung
##
## Df Deviance AIC
## + age 2 17661 17781
## + bone 2 17668 17788
## + CEA 2 17694 17815
## + brain 2 17698 17818
## + kind 1 17711 17821
## + marry 2 17703 17824
## + N 2 17711 17831
## + size 2 17712 17832
## <none> 17734 17835
## + radiate 1 17725 17835
## + RX 1 17726 17836
## + t 2 17720 17840
## + gender 1 17734 17844
## + race 2 17724 17844
## + income 2 17725 17846
## - lung 2 17832 17913
## - site 1 17835 17925
## - grade 4 18034 18095
## - surgery 1 18086 18176
## - chem 1 21078 21169
##
## Step: AIC=17781.08
## y ~ chem + surgery + grade + site + lung + age
##
## Df Deviance AIC
## + bone 2 17592 17733
## + CEA 2 17622 17762
## + brain 2 17622 17763
## + kind 1 17636 17767
## + marry 2 17634 17774
## + N 2 17635 17776
## + size 2 17636 17777
## <none> 17661 17781
## + radiate 1 17652 17783
## + RX 1 17654 17785
## + t 2 17647 17787
## + race 2 17649 17790
## + gender 1 17660 17791
## + income 2 17651 17791
## - age 2 17734 17835
## - site 1 17737 17847
## - lung 2 17757 17857
## - grade 4 17964 18044
## - surgery 1 17998 18109

```

```

## - chem      1      20714 20824
##
## Step:  AIC=17732.59
## y ~ chem + surgery + grade + site + lung + age + bone
##
##           Df Deviance   AIC
## + CEA      2      17554 17715
## + kind      1      17570 17720
## + brain     2      17562 17723
## + radiate   1      17573 17724
## + marry     2      17564 17724
## + N         2      17568 17728
## + size      2      17569 17729
## <none>      17592 17733
## + RX        1      17586 17736
## + t         2      17578 17738
## + race      2      17580 17741
## + gender    1      17591 17742
## + income    2      17582 17743
## - bone      2      17661 17781
## - age       2      17668 17788
## - lung      2      17670 17790
## - site      1      17670 17800
## - grade     4      17886 17986
## - surgery   1      17907 18037
## - chem      1      20642 20772
##
## Step:  AIC=17714.65
## y ~ chem + surgery + grade + site + lung + age + bone + CEA
##
##           Df Deviance   AIC
## + kind      1      17528 17698
## + brain     2      17524 17705
## + marry     2      17527 17707
## + radiate   1      17537 17708
## + N         2      17530 17711
## + size      2      17531 17712
## <none>      17554 17715
## + RX        1      17548 17719
## + t         2      17540 17721
## + race      2      17542 17723
## + gender    1      17554 17724
## + income    2      17544 17725
## - CEA       2      17592 17733
## - bone      2      17622 17762
## - lung      2      17624 17765
## - age       2      17629 17769
## - site      1      17633 17783
## - grade     4      17855 17976
## - surgery   1      17856 18006
## - chem      1      20552 20702
##
## Step:  AIC=17698.2
## y ~ chem + surgery + grade + site + lung + age + bone + CEA +

```

```

##      kind
##
##           Df Deviance   AIC
## + brain    2    17498 17688
## + radiate   1    17510 17691
## + marry     2    17501 17692
## + N         2    17504 17694
## + size      2    17506 17697
## <none>      17528 17698
## + RX        1    17522 17702
## + t         2    17513 17704
## + race      2    17516 17707
## + gender    1    17527 17708
## + income    2    17518 17709
## - kind      1    17554 17715
## - CEA       2    17570 17720
## - bone      2    17594 17744
## - lung      2    17600 17751
## - age       2    17603 17753
## - site      1    17601 17762
## - grade     4    17805 17935
## - surgery   1    17834 17994
## - chem      1    20518 20679
##
## Step:  AIC=17688.43
## y ~ chem + surgery + grade + site + lung + age + bone + CEA +
##      kind + brain
##
##           Df Deviance   AIC
## + radiate   1    17469 17670
## + marry     2    17471 17682
## + N         2    17474 17685
## + size      2    17476 17686
## <none>      17498 17688
## + RX        1    17490 17691
## + t         2    17484 17694
## + race      2    17486 17697
## - brain     2    17528 17698
## + gender    1    17498 17698
## + income    2    17488 17699
## - kind      1    17524 17705
## - CEA       2    17540 17710
## - bone      2    17556 17726
## - lung      2    17562 17732
## - age       2    17574 17745
## - site      1    17571 17752
## - grade     4    17772 17923
## - surgery   1    17799 17980
## - chem      1    20474 20655
##
## Step:  AIC=17670.1
## y ~ chem + surgery + grade + site + lung + age + bone + CEA +
##      kind + brain + radiate
##

```

```

##           Df Deviance   AIC
## + marry    2    17443 17664
## + N         2    17446 17666
## + size      2    17448 17668
## <none>      17469 17670
## + RX        1    17462 17673
## + race      2    17458 17679
## + t         2    17458 17679
## + gender    1    17469 17680
## + income    2    17460 17680
## - kind      1    17496 17687
## - radiate   1    17498 17688
## - CEA       2    17509 17689
## - brain     2    17510 17691
## - lung      2    17533 17714
## - site      1    17528 17719
## - bone      2    17539 17720
## - age       2    17544 17725
## - grade     4    17741 17902
## - surgery   1    17776 17967
## - chem      1    20361 20551
##
## Step: AIC=17663.72
## y ~ chem + surgery + grade + site + lung + age + bone + CEA +
##      kind + brain + radiate + marry
##
##           Df Deviance   AIC
## + N         2    17418 17659
## + size      2    17422 17663
## <none>      17443 17664
## + RX        1    17436 17667
## - marry     2    17469 17670
## + t         2    17432 17673
## + gender    1    17443 17674
## + race      2    17434 17675
## + income    2    17434 17675
## - kind      1    17469 17680
## - CEA       2    17481 17682
## - radiate   1    17471 17682
## - brain     2    17483 17684
## - lung      2    17506 17707
## - site      1    17501 17712
## - age       2    17513 17714
## - bone      2    17514 17715
## - grade     4    17715 17895
## - surgery   1    17744 17955
## - chem      1    20211 20422
##
## Step: AIC=17659.39
## y ~ chem + surgery + grade + site + lung + age + bone + CEA +
##      kind + brain + radiate + marry + N
##
##           Df Deviance   AIC
## <none>      17418 17659

```

```
## + size      2      17400 17660
## + RX        1      17412 17664
## - N         2      17443 17664
## - marry     2      17446 17666
## + t         2      17407 17668
## + gender    1      17418 17669
## + income    2      17409 17670
## + race      2      17410 17670
## - kind      1      17445 17676
## - CEA       2      17456 17677
## - radiate   1      17447 17678
## - brain     2      17459 17680
## - lung      2      17480 17700
## - site      1      17474 17705
## - bone      2      17488 17709
## - age       2      17491 17712
## - grade     4      17660 17860
## - surgery   1      17656 17886
## - chem      1      20182 20413
```

```
formula(bothways_bic_2)
```

```
## y ~ chem + surgery + grade + site + lung + age + bone + CEA +
##      kind + brain + radiate + marry + N
```

```
summary(bothways_bic_2)
```

```
##
## Call:
## glm(formula = y ~ chem + surgery + grade + site + lung + age +
##      bone + CEA + kind + brain + radiate + marry + N, family = binomial,
##      data = train_0_1_2)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.2349  -0.5672  -0.3750  -0.2420   2.9351
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.04127    0.19138  -5.441 5.30e-08 ***
## chem1       -2.07841    0.04095 -50.756 < 2e-16 ***
## surgery1    -0.80805    0.05332 -15.153 < 2e-16 ***
## grade2       0.15977    0.11651   1.371 0.170287
## grade3       0.84484    0.12242   6.901 5.16e-12 ***
## grade4       0.82979    0.16430   5.050 4.41e-07 ***
## grade9       0.73058    0.11804   6.189 6.04e-10 ***
## site2        0.30689    0.04097   7.491 6.85e-14 ***
## lung1        0.34097    0.04379   7.786 6.90e-15 ***
## lung2       -0.01549    0.14488  -0.107 0.914849
## age2        -0.12229    0.13685  -0.894 0.371509
## age3         0.25838    0.13502   1.914 0.055667 .
## bone1        0.63218    0.07500   8.429 < 2e-16 ***
## bone2       -0.10853    0.20190  -0.538 0.590904
```



```
## CEA1      0.46295    0.07780    5.950 2.67e-09 ***
## CEA2      0.39736    0.08241    4.822 1.42e-06 ***
## kind2     0.35800    0.06858    5.220 1.79e-07 ***
## brain1    1.04800    0.16344    6.412 1.44e-10 ***
## brain2    -0.06232    0.18733   -0.333 0.739376
## radiate1  -0.37137    0.07149   -5.195 2.05e-07 ***
## marry2    -0.17675    0.04867   -3.632 0.000281 ***
## marry3     0.04426    0.05468    0.809 0.418340
## N2        0.14462    0.05829    2.481 0.013101 *
## N3        0.22914    0.05150    4.449 8.62e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 22383  on 22901  degrees of freedom
## Residual deviance: 17418  on 22878  degrees of freedom
## AIC: 17466
##
## Number of Fisher Scoring iterations: 5
```

0+1+2 lasso

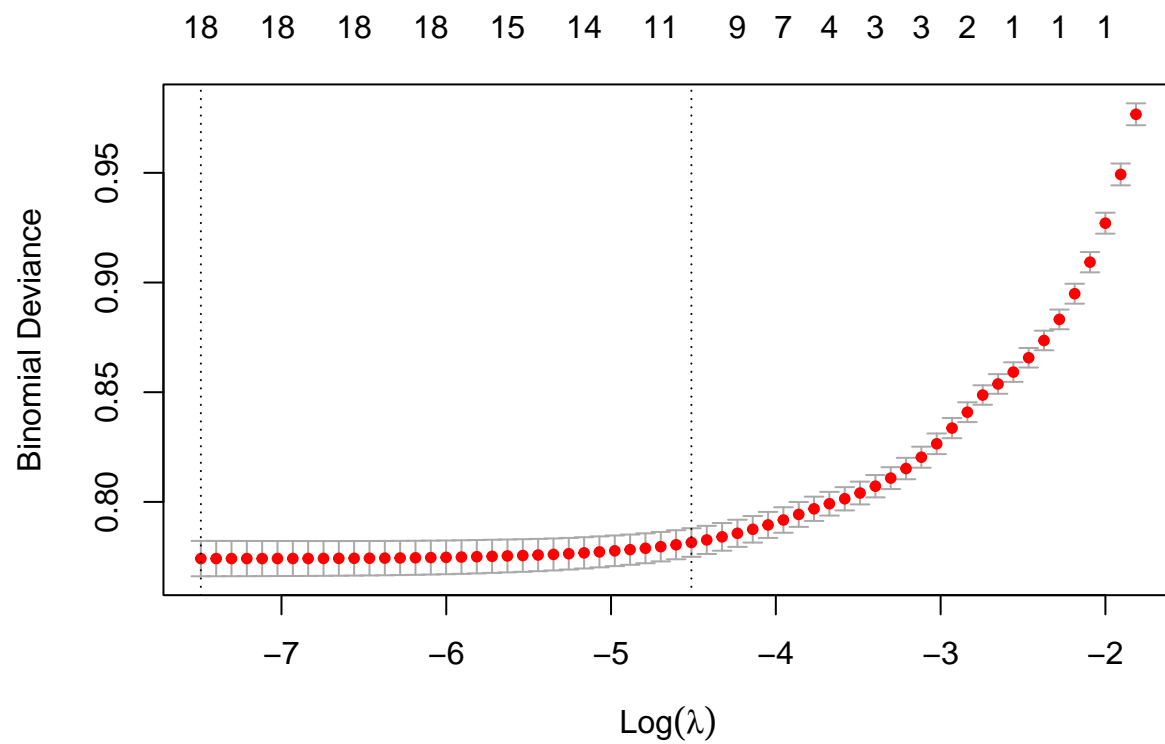
```
x <- data.matrix(train_0_1_2[, 2:ncol(train_0_1_2)])
y <- as.numeric(data.matrix(train_0_1_2$y))

#perform k-fold cross-validation to find optimal lambda value
set.seed(2023)
cv_model <- cv.glmnet(x, y, alpha = 1, family = "binomial")

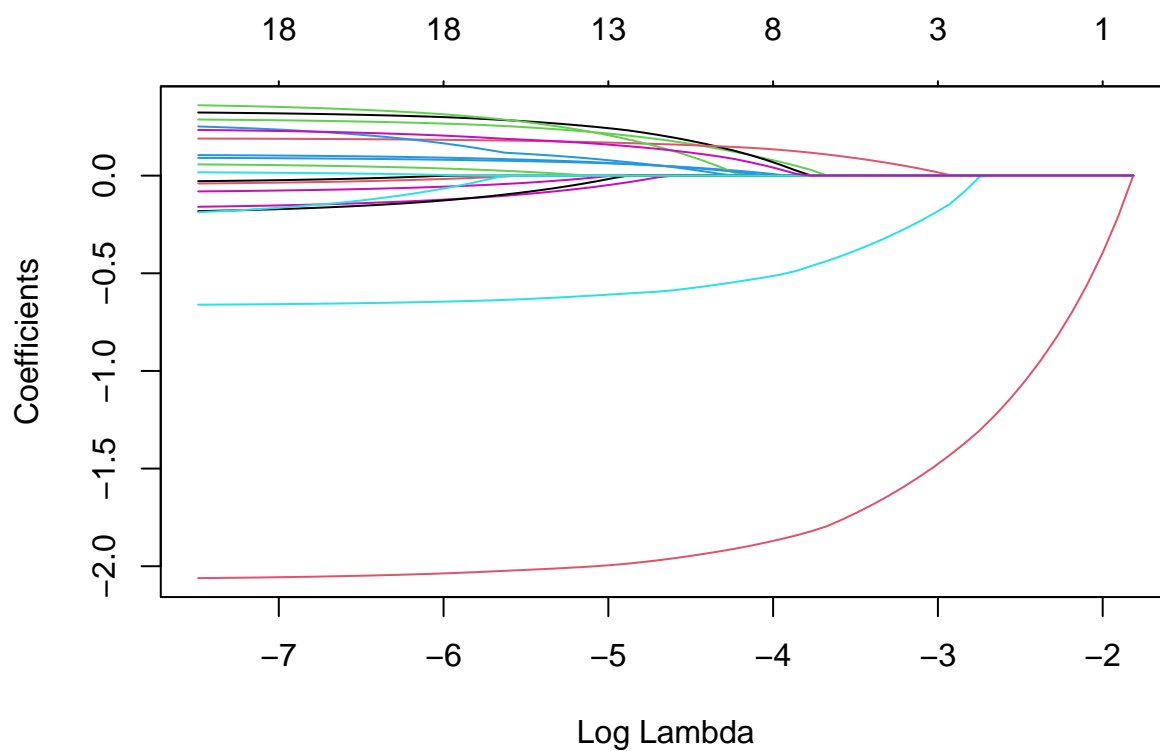
#find optimal lambda value that minimizes test MSE
best_lambda <- cv_model$lambda.min
best_lambda
```

```
## [1] 0.0005590239
```

```
#produce plot of test MSE by lambda value
plot(cv_model)
```



```
#Lasso path
plot(cv_model$glmnet.fit, "lambda", label=FALSE)
```



```
#find coefficients of best model
best_model <- glmnet(x, y, alpha = 1, lambda = best_lambda, family = binomial())
coef(best_model)
```

```
## 20 x 1 sparse Matrix of class "dgCMatrix"
##              s0
## (Intercept)  0.43503167
## gender      -0.02841015
## race        -0.04010444
## age         0.28752675
## size        0.09099798
## marry       0.01701699
## income      -0.08076250
## site        0.32313466
## grade       0.18998937
## kind        0.36036780
## t           .
## N           0.10530259
## surgery     -0.66081305
## RX          -0.15940880
## radiate     -0.18187862
## chem        -2.06076194
## CEA         0.05756372
## bone        0.25192533
## brain       -0.18717506
## lung        0.23391223
```

```
print(best_model)
```

```
##
## Call:  glmnet(x = x, y = y, family = binomial(), alpha = 1, lambda = best_lambda)
##
##      Df %Dev  Lambda
## 1 18    21 0.000559
```