

CRCLM

Zhiyi Sun

2023-03-16

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 train/test

```
set.seed(2023)
sample <- sample(nrow(CRCLM_0_1), floor(nrow(CRCLM_0_1)*0.8))
train_0_1 <- CRCLM_0_1[sample,]
test_0_1 <- CRCLM_0_1[-sample,]

prevalence_0_1_train <- sum(train_0_1$cod == 1)/nrow(train_0_1)
prevalence_0_1_test <- sum(test_0_1$cod == 1)/nrow(test_0_1)

train_0_1 <- train_0_1[, c('y', 'gender', 'race', 'age', 'size', 'marry', 'income', 'site', 'grade', 'kind', 't', 'N', 'surgery', 'RX', 'radiate', 'chem', 'CEA', 'bone', 'brain', 'lung')]
str(train_0_1)
```

```
## tibble [21,391 x 20] (S3: tbl_df/tbl/data.frame)
## $ y : Factor w/ 2 levels "0","1": 2 1 1 1 1 1 2 1 2 1 ...
## $ gender : Factor w/ 2 levels "1","2": 1 2 2 2 1 2 1 2 1 2 ...
## $ race : Factor w/ 3 levels "1","2","3": 1 1 1 1 1 1 1 2 2 1 ...
## $ age : Factor w/ 3 levels "1","2","3": 3 3 1 2 2 3 3 2 3 2 ...
## $ size : Factor w/ 3 levels "1","2","3": 2 3 3 1 1 2 2 1 1 3 ...
## $ marry : Factor w/ 3 levels "1","2","3": 3 2 2 2 2 3 3 1 3 2 ...
## $ income : Factor w/ 3 levels "1","2","3": 2 2 1 3 2 2 2 3 2 2 ...
## $ site : Factor w/ 2 levels "1","2": 2 1 1 1 1 1 2 2 1 1 ...
## $ grade : Factor w/ 5 levels "1","2","3","4",...: 5 2 2 2 2 5 5 5 5 5 ...
## $ kind : Factor w/ 2 levels "1","2": 1 1 1 1 1 1 1 1 1 1 ...
## $ t : Factor w/ 3 levels "1","2","3": 2 3 3 2 2 3 2 3 1 2 ...
## $ N : Factor w/ 3 levels "1","2","3": 1 1 1 1 1 1 1 3 3 1 ...
## $ surgery: Factor w/ 2 levels "0","1": 1 2 1 1 2 1 2 1 2 1 ...
## $ RX : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 2 1 1 1 ...
## $ radiate: Factor w/ 2 levels "0","1": 1 2 1 1 1 1 1 1 2 1 ...
## $ chem : Factor w/ 2 levels "0","1": 2 2 2 2 2 1 2 2 2 2 ...
## $ CEA : Factor w/ 3 levels "0","1","2": 2 3 2 2 3 3 2 2 2 3 ...
## $ bone : Factor w/ 3 levels "0","1","2": 1 1 1 1 1 1 1 1 2 1 ...
## $ brain : Factor w/ 3 levels "0","1","2": 1 1 1 1 1 1 1 1 2 1 ...
## $ lung : Factor w/ 3 levels "0","1","2": 1 1 2 1 1 1 1 2 2 1 ...
```

```
test_0_1 <- test_0_1[, c('y', 'gender', 'race', 'age', 'size', 'marry', 'income', 'site', 'grade', 'kind', 't', 'N', 'surgery', 'RX', 'radiate', 'chem', 'CEA', 'bone', 'brain', 'lung')]
str(test_0_1)
```

```
## tibble [5,348 x 20] (S3: tbl_df/tbl/data.frame)
```

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

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 univariate class

```
u1 <- glm(y ~ gender, binomial(link='logit'), data = train_0_1)
summary(u1)
```

```
##
## Call:
## glm(formula = y ~ gender, family = binomial(link = "logit"),
## data = train_0_1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.6625  -0.6625  -0.6260  -0.6260   1.8581
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.40478    0.02624 -53.539 < 2e-16 ***
## gender2      -0.12549    0.03533  -3.552 0.000383 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
```

```
##
## Null deviance: 20561 on 21390 degrees of freedom
## Residual deviance: 20548 on 21389 degrees of freedom
## AIC: 20552
##
## Number of Fisher Scoring iterations: 4

u2 <- glm(y ~ race, binomial(link='logit'), data = train_0_1)
summary(u2)
```

```
##
## Call:
## glm(formula = y ~ race, family = binomial(link = "logit"), data = train_0_1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.6761  -0.6396  -0.6396  -0.6082   1.8862
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.48302    0.02027  -73.161  <2e-16 ***
## race2        0.12354    0.04895   2.524   0.0116 *
## race3       -0.11087    0.06104  -1.817   0.0693 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 20561 on 21390 degrees of freedom
## Residual deviance: 20550 on 21388 degrees of freedom
## AIC: 20556
##
## Number of Fisher Scoring iterations: 4
```

```
u3 <- glm(y ~ age, binomial(link='logit'), data = train_0_1)
summary(u3)
```

```
##
## Call:
## glm(formula = y ~ age, family = binomial(link = "logit"), data = train_0_1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.7245  -0.7245  -0.4972  -0.4972   2.1745
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -2.2654    0.1323  -17.117  < 2e-16 ***
## age2         0.2371    0.1371   1.729   0.0838 .
## age3         1.0618    0.1339   7.927  2.25e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 20561 on 21390 degrees of freedom
## Residual deviance: 20080 on 21388 degrees of freedom
## AIC: 20086
##
## Number of Fisher Scoring iterations: 4
```

```
u4 <- glm(y ~ size, binomial(link='logit'), data = train_0_1)
summary(u4)
```

```
##
## Call:
## glm(formula = y ~ size, family = binomial(link = "logit"), data = train_0_1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.7576  -0.6368  -0.5504  -0.5504   1.9810
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.81084    0.03088 -58.637 < 2e-16 ***
## size2        0.31806    0.04507   7.058 1.69e-12 ***
## size3        0.70936    0.04211  16.844 < 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: 20561 on 21390 degrees of freedom
## Residual deviance: 20273 on 21388 degrees of freedom
## AIC: 20279
##
## Number of Fisher Scoring iterations: 4
```

```
u5 <- glm(y ~ marry, binomial(link='logit'), data = train_0_1)
summary(u5)
```

```
##
## Call:
## glm(formula = y ~ marry, family = binomial(link = "logit"), data = train_0_1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.7439  -0.6747  -0.5810  -0.5810   1.9300
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.36416    0.03576 -38.150 < 2e-16 ***
## marry2      -0.32950    0.04398  -7.492 6.81e-14 ***
## marry3       0.22081    0.04881   4.524 6.06e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 20561  on 21390  degrees of freedom
## Residual deviance: 20381  on 21388  degrees of freedom
## AIC: 20387
##
## Number of Fisher Scoring iterations: 4

u6 <- glm(y ~ income, binomial(link='logit'), data = train_0_1)
summary(u6)

##
## Call:
## glm(formula = y ~ income, family = binomial(link = "logit"),
##      data = train_0_1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.6567  -0.6567  -0.6466  -0.6114   1.8811
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.45892    0.04267  -34.193  <2e-16 ***
## income2      0.03441    0.04875   0.706   0.4803
## income3     -0.12347    0.05427  -2.275   0.0229 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 20561  on 21390  degrees of freedom
## Residual deviance: 20546  on 21388  degrees of freedom
## AIC: 20552
##
## Number of Fisher Scoring iterations: 4

u7 <- glm(y ~ site, binomial(link='logit'), data = train_0_1)
summary(u7)

##
## Call:
## glm(formula = y ~ site, family = binomial(link = "logit"), data = train_0_1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.7208  -0.7208  -0.5941  -0.5941   1.9086
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.64485    0.02327  -70.68  <2e-16 ***
## site2       0.42964    0.03564  12.05  <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: 20561  on 21390  degrees of freedom
## Residual deviance: 20417  on 21389  degrees of freedom
## AIC: 20421
##
## Number of Fisher Scoring iterations: 4
```

```
u8 <- glm(y ~ grade, binomial(link='logit'), data = train_0_1)
summary(u8)
```

```
##
## Call:
## glm(formula = y ~ grade, family = binomial(link = "logit"), data = train_0_1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.7851  -0.7245  -0.5248  -0.5248   2.0251
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.87180    0.10533 -17.771 < 2e-16 ***
## grade2      -0.04108    0.10925  -0.376  0.70694
## grade3       0.66810    0.11346   5.888 3.90e-09 ***
## grade4       0.49282    0.15247   3.232  0.00123 **
## grade9       0.85282    0.10911   7.816 5.44e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 20561  on 21390  degrees of freedom
## Residual deviance: 20014  on 21386  degrees of freedom
## AIC: 20024
##
## Number of Fisher Scoring iterations: 4
```

```
u9 <- glm(y ~ kind, binomial(link='logit'), data = train_0_1)
summary(u9)
```

```
##
## Call:
## glm(formula = y ~ kind, family = binomial(link = "logit"), data = train_0_1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.7591  -0.6323  -0.6323  -0.6323   1.8484
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.50846    0.01843 -81.854 < 2e-16 ***
```

```
## kind2          0.41157    0.06161    6.681 2.38e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 20561  on 21390  degrees of freedom
## Residual deviance: 20519  on 21389  degrees of freedom
## AIC: 20523
##
## Number of Fisher Scoring iterations: 4
```

```
u10 <- glm(y ~ t, binomial(link='logit'), data = train_0_1)
summary(u10)
```

```
##
## Call:
## glm(formula = y ~ t, family = binomial(link = "logit"), data = train_0_1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.7706  -0.6897  -0.5590  -0.5590   1.9665
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.31495    0.05184 -25.365  < 2e-16 ***
## t2          -0.46232    0.05770  -8.012 1.13e-15 ***
## t3           0.25290    0.05902   4.285 1.83e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 20561  on 21390  degrees of freedom
## Residual deviance: 20199  on 21388  degrees of freedom
## AIC: 20205
##
## Number of Fisher Scoring iterations: 4
```

```
u11 <- glm(y ~ N, binomial(link='logit'), data = train_0_1)
summary(u11)
```

```
##
## Call:
## glm(formula = y ~ N, family = binomial(link = "logit"), data = train_0_1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.8231  -0.6181  -0.6181  -0.5636   1.9587
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.55824    0.02327 -66.971  < 2e-16 ***
```

```
## N2          -0.20116    0.04615   -4.359 1.31e-05 ***
## N3          0.64992    0.04397   14.781 < 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: 20561  on 21390  degrees of freedom
## Residual deviance: 20282  on 21388  degrees of freedom
## AIC: 20288
##
## Number of Fisher Scoring iterations: 4
```

```
u12 <- glm(y ~ surgery, binomial(link='logit'), data = train_0_1)
summary(u12)
```

```
##
## Call:
## glm(formula = y ~ surgery, family = binomial(link = "logit"),
##      data = train_0_1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.7648  -0.7648  -0.5072  -0.5072   2.0565
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.07960    0.02239  -48.21  <2e-16 ***
## surgery1    -0.90629    0.03701  -24.49  <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: 20561  on 21390  degrees of freedom
## Residual deviance: 19928  on 21389  degrees of freedom
## AIC: 19932
##
## Number of Fisher Scoring iterations: 4
```

```
u13 <- glm(y ~ RX, binomial(link='logit'), data = train_0_1)
summary(u13)
```

```
##
## Call:
## glm(formula = y ~ RX, family = binomial(link = "logit"), data = train_0_1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.6765  -0.6765  -0.6765  -0.4597   2.1446
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
```

```
## (Intercept) -1.35830    0.01868   -72.73   <2e-16 ***
## RX1         -0.83569    0.05735   -14.57   <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: 20561  on 21390  degrees of freedom
## Residual deviance: 20310  on 21389  degrees of freedom
## AIC: 20314
##
## Number of Fisher Scoring iterations: 4
```

```
u14 <- glm(y ~ radiate, binomial(link='logit'), data = train_0_1)
summary(u14)
```

```
##
## Call:
## glm(formula = y ~ radiate, family = binomial(link = "logit"),
##      data = train_0_1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.6600  -0.6600  -0.6600  -0.5068   2.0572
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.41323    0.01845  -76.612   <2e-16 ***
## radiate1     -0.57448    0.06170   -9.311   <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: 20561  on 21390  degrees of freedom
## Residual deviance: 20464  on 21389  degrees of freedom
## AIC: 20468
##
## Number of Fisher Scoring iterations: 4
```

```
u15 <- glm(y ~ chem, binomial(link='logit'), data = train_0_1)
summary(u15)
```

```
##
## Call:
## glm(formula = y ~ chem, family = binomial(link = "logit"), data = train_0_1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1385  -0.4557  -0.4557  -0.4557   2.1523
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
```

```
## (Intercept) -0.09222    0.02846   -3.24  0.00119 **
## chem1      -2.12016    0.03865  -54.85  < 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: 20561  on 21390  degrees of freedom
## Residual deviance: 17441  on 21389  degrees of freedom
## AIC: 17445
##
## Number of Fisher Scoring iterations: 4
```

```
u16 <- glm(y ~ CEA, binomial(link='logit'), data = train_0_1)
summary(u16)
```

```
##
## Call:
## glm(formula = y ~ CEA, family = binomial(link = "logit"), data = train_0_1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.6950  -0.6417  -0.6417  -0.5027   2.0646
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -2.00489    0.06593 -30.407  < 2e-16 ***
## CEA1         0.52901    0.06944   7.618 2.57e-14 ***
## CEA2         0.70730    0.07398   9.561  < 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: 20561  on 21390  degrees of freedom
## Residual deviance: 20460  on 21388  degrees of freedom
## AIC: 20466
##
## Number of Fisher Scoring iterations: 4
```

```
u17 <- glm(y ~ bone, binomial(link='logit'), data = train_0_1)
summary(u17)
```

```
##
## Call:
## glm(formula = y ~ bone, family = binomial(link = "logit"), data = train_0_1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.8981  -0.6226  -0.6226  -0.6226   1.8634
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
```

```
## (Intercept) -1.54224    0.01863 -82.800 < 2e-16 ***
## bone1       0.84262    0.06507  12.949 < 2e-16 ***
## bone2       0.50681    0.11906   4.257 2.07e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##    Null deviance: 20561  on 21390  degrees of freedom
## Residual deviance: 20395  on 21388  degrees of freedom
## AIC: 20401
##
## Number of Fisher Scoring iterations: 4
```

```
u18 <- glm(y ~ brain, binomial(link='logit'), data = train_0_1)
summary(u18)
```

```
##
## Call:
## glm(formula = y ~ brain, family = binomial(link = "logit"), data = train_0_1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.0815  -0.6341  -0.6341  -0.6341   1.8456
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.50202    0.01798 -83.537 < 2e-16 ***
## brain1       1.27216    0.14314   8.888 < 2e-16 ***
## brain2       0.46535    0.11192   4.158 3.22e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##    Null deviance: 20561  on 21390  degrees of freedom
## Residual deviance: 20475  on 21388  degrees of freedom
## AIC: 20481
##
## Number of Fisher Scoring iterations: 4
```

```
u19 <- glm(y ~ lung, binomial(link='logit'), data = train_0_1)
summary(u19)
```

```
##
## Call:
## glm(formula = y ~ lung, family = binomial(link = "logit"), data = train_0_1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.7554  -0.5986  -0.5986  -0.5986   1.9014
##
## Coefficients:
```

```
##           Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.62846    0.02156 -75.525 < 2e-16 ***
## lung1       0.51287    0.03850  13.321 < 2e-16 ***
## lung2       0.52027    0.11513   4.519 6.21e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##    Null deviance: 20561  on 21390  degrees of freedom
## Residual deviance: 20379  on 21388  degrees of freedom
## AIC: 20385
##
## Number of Fisher Scoring iterations: 4
```

0+1 univariate unclass

```
uc1 <- glm(y ~ unclass(gender), binomial(link='logit'), data = train_0_1)
summary(uc1)
```

```
##
## Call:
## glm(formula = y ~ unclass(gender), family = binomial(link = "logit"),
##      data = train_0_1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.6625  -0.6625  -0.6260  -0.6260   1.8581
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -1.27929    0.05756 -22.224 < 2e-16 ***
## unclass(gender) -0.12549    0.03533  -3.552 0.000383 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##    Null deviance: 20561  on 21390  degrees of freedom
## Residual deviance: 20548  on 21389  degrees of freedom
## AIC: 20552
##
## Number of Fisher Scoring iterations: 4
```

```
uc2 <- glm(y ~ unclass(race), binomial(link='logit'), data = train_0_1)
summary(uc2)
```

```
##
## Call:
## glm(formula = y ~ unclass(race), family = binomial(link = "logit"),
##      data = train_0_1)
```

```
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.6432  -0.6432  -0.6432  -0.6355   1.8434
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -1.45749    0.04027  -36.193  <2e-16 ***
## unclass(race) -0.01318    0.02700   -0.488    0.626
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 20561  on 21390  degrees of freedom
## Residual deviance: 20561  on 21389  degrees of freedom
## AIC: 20565
##
## Number of Fisher Scoring iterations: 4
```

```
uc3 <- glm(y ~ unclass(age), binomial(link='logit'), data = train_0_1)
summary(uc3)
```

```
##
## Call:
## glm(formula = y ~ unclass(age), family = binomial(link = "logit"),
##      data = train_0_1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.7215  -0.7215  -0.5104  -0.5104   2.3644
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -3.49159    0.10327  -33.81  <2e-16 ***
## unclass(age)  0.75952    0.03739   20.32  <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: 20561  on 21390  degrees of freedom
## Residual deviance: 20093  on 21389  degrees of freedom
## AIC: 20097
##
## Number of Fisher Scoring iterations: 4
```

```
uc4 <- glm(y ~ unclass(size), binomial(link='logit'), data = train_0_1)
summary(uc4)
```

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



```
##      data = train_0_1)
##
## Deviance Residuals:
##      Min        1Q      Median        3Q        Max
## -0.7543  -0.6442  -0.5474  -0.5474   1.9861
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -2.17792    0.04684  -46.50  <2e-16 ***
## unclass(size)  0.35550    0.02111   16.84  <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: 20561  on 21390  degrees of freedom
## Residual deviance: 20274  on 21389  degrees of freedom
## AIC: 20278
##
## Number of Fisher Scoring iterations: 4
```

```
uc5 <- glm(y ~ unclass(marry), binomial(link='logit'), data = train_0_1)
summary(uc5)
```

```
##
## Call:
## glm(formula = y ~ unclass(marry), family = binomial(link = "logit"),
##      data = train_0_1)
##
## Deviance Residuals:
##      Min        1Q      Median        3Q        Max
## -0.6792  -0.6409  -0.6409  -0.6044   1.8921
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -1.73637    0.05603  -30.988  < 2e-16 ***
## unclass(marry)  0.12903    0.02605   4.953 7.29e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 20561  on 21390  degrees of freedom
## Residual deviance: 20536  on 21389  degrees of freedom
## AIC: 20540
##
## Number of Fisher Scoring iterations: 4
```

```
uc6 <- glm(y ~ unclass(income), binomial(link='logit'), data = train_0_1)
summary(uc6)
```

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

```
## glm(formula = y ~ unclass(income), family = binomial(link = "logit"),
##     data = train_0_1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.6663  -0.6443  -0.6443  -0.6230   1.8628
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -1.31792    0.05802  -22.72  < 2e-16 ***
## unclass(income) -0.07434    0.02626   -2.83  0.00465 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 20561  on 21390  degrees of freedom
## Residual deviance: 20553  on 21389  degrees of freedom
## AIC: 20557
##
## Number of Fisher Scoring iterations: 4
```

```
uc7 <- glm(y ~ unclass(site), binomial(link='logit'), data = train_0_1)
summary(uc7)
```

```
##
## Call:
## glm(formula = y ~ unclass(site), family = binomial(link = "logit"),
##     data = train_0_1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.7208  -0.7208  -0.5941  -0.5941   1.9086
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -2.07448    0.05381  -38.55  <2e-16 ***
## unclass(site)  0.42964    0.03564   12.05  <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: 20561  on 21390  degrees of freedom
## Residual deviance: 20417  on 21389  degrees of freedom
## AIC: 20421
##
## Number of Fisher Scoring iterations: 4
```

```
uc8 <- glm(y ~ unclass(grade), binomial(link='logit'), data = train_0_1)
summary(uc8)
```

```
##
```

```
## Call:
## glm(formula = y ~ unclass(grade), family = binomial(link = "logit"),
##      data = train_0_1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.7969  -0.6255  -0.5516  -0.5516   2.0966
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -2.35409    0.04606  -51.11  <2e-16 ***
## unclass(grade) 0.27400    0.01262   21.71  <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: 20561  on 21390  degrees of freedom
## Residual deviance: 20091  on 21389  degrees of freedom
## AIC: 20095
##
## Number of Fisher Scoring iterations: 4
```

```
uc9 <- glm(y ~ unclass(kind), binomial(link='logit'), data = train_0_1)
summary(uc9)
```

```
##
## Call:
## glm(formula = y ~ unclass(kind), family = binomial(link = "logit"),
##      data = train_0_1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.7591  -0.6323  -0.6323  -0.6323   1.8484
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -1.92003    0.06938  -27.672  < 2e-16 ***
## unclass(kind) 0.41157    0.06161   6.681 2.38e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 20561  on 21390  degrees of freedom
## Residual deviance: 20519  on 21389  degrees of freedom
## AIC: 20523
##
## Number of Fisher Scoring iterations: 4
```

```
uc10 <- glm(y ~ unclass(t), binomial(link='logit'), data = train_0_1)
summary(uc10)
```

```
##
## Call:
## glm(formula = y ~ unclass(t), family = binomial(link = "logit"),
##      data = train_0_1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.7213  -0.6178  -0.6178  -0.5267   2.0218
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -2.25083    0.06942  -32.42  <2e-16 ***
## unclass(t)   0.34576    0.02941   11.76  <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: 20561  on 21390  degrees of freedom
## Residual deviance: 20420  on 21389  degrees of freedom
## AIC: 20424
##
## Number of Fisher Scoring iterations: 4
```

```
uc11 <- glm(y ~ unclass(N), binomial(link='logit'), data = train_0_1)
summary(uc11)
```

```
##
## Call:
## glm(formula = y ~ unclass(N), family = binomial(link = "logit"),
##      data = train_0_1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.7546  -0.6720  -0.5967  -0.5967   1.9044
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.89771    0.04078  -46.53  <2e-16 ***
## unclass(N)   0.26237    0.02218   11.83  <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: 20561  on 21390  degrees of freedom
## Residual deviance: 20424  on 21389  degrees of freedom
## AIC: 20428
##
## Number of Fisher Scoring iterations: 4
```

```
uc12 <- glm(y ~ unclass(surgery), binomial(link='logit'), data = train_0_1)
summary(uc12)
```

```
##
## Call:
## glm(formula = y ~ unclass(surgery), family = binomial(link = "logit"),
##      data = train_0_1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.7648  -0.7648  -0.5072  -0.5072   2.0565
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -0.17331    0.05361  -3.233  0.00123 **
## unclass(surgery) -0.90629    0.03701 -24.490 < 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: 20561  on 21390  degrees of freedom
## Residual deviance: 19928  on 21389  degrees of freedom
## AIC: 19932
##
## Number of Fisher Scoring iterations: 4
```

```
uc13 <- glm(y ~ unclass(RX), binomial(link='logit'), data = train_0_1)
summary(uc13)
```

```
##
## Call:
## glm(formula = y ~ unclass(RX), family = binomial(link = "logit"),
##      data = train_0_1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.6765  -0.6765  -0.6765  -0.4597   2.1446
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -0.52261    0.06584  -7.937 2.07e-15 ***
## unclass(RX)  -0.83569    0.05735 -14.572 < 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: 20561  on 21390  degrees of freedom
## Residual deviance: 20310  on 21389  degrees of freedom
## AIC: 20314
##
## Number of Fisher Scoring iterations: 4
```

```
uc14 <- glm(y ~ unclass(radiate), binomial(link='logit'), data = train_0_1)
summary(uc14)
```

```
##
## Call:
## glm(formula = y ~ unclass(radiate), family = binomial(link = "logit"),
##      data = train_0_1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.6600  -0.6600  -0.6600  -0.5068   2.0572
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -0.83875    0.06948  -12.072  <2e-16 ***
## unclass(radiate) -0.57448    0.06170   -9.311  <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: 20561  on 21390  degrees of freedom
## Residual deviance: 20464  on 21389  degrees of freedom
## AIC: 20468
##
## Number of Fisher Scoring iterations: 4
```

```
uc15 <- glm(y ~ unclass(chem), binomial(link='logit'), data = train_0_1)
summary(uc15)
```

```
##
## Call:
## glm(formula = y ~ unclass(chem), family = binomial(link = "logit"),
##      data = train_0_1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1385  -0.4557  -0.4557  -0.4557   2.1523
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    2.02794    0.06264   32.37  <2e-16 ***
## unclass(chem) -2.12016    0.03865  -54.85  <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: 20561  on 21390  degrees of freedom
## Residual deviance: 17441  on 21389  degrees of freedom
## AIC: 17445
##
## Number of Fisher Scoring iterations: 4
```

```
uc16 <- glm(y ~ unclass(CEA), binomial(link='logit'), data = train_0_1)
summary(uc16)
```

```
##
## Call:
## glm(formula = y ~ unclass(CEA), family = binomial(link = "logit"),
##      data = train_0_1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.7122  -0.6279  -0.6279  -0.5519   1.9784
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -2.08594    0.07010  -29.756  <2e-16 ***
## unclass(CEA)   0.28112    0.03083   9.119   <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: 20561  on 21390  degrees of freedom
## Residual deviance: 20477  on 21389  degrees of freedom
## AIC: 20481
##
## Number of Fisher Scoring iterations: 4
```

```
uc17 <- glm(y ~ unclass(bone), binomial(link='logit'), data = train_0_1)
summary(uc17)
```

```
##
## Call:
## glm(formula = y ~ unclass(bone), family = binomial(link = "logit"),
##      data = train_0_1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.9601  -0.6266  -0.6266  -0.6266   1.8573
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -2.02495    0.05218  -38.81  <2e-16 ***
## unclass(bone)  0.49657    0.04370   11.36  <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: 20561  on 21390  degrees of freedom
## Residual deviance: 20443  on 21389  degrees of freedom
## AIC: 20447
##
## Number of Fisher Scoring iterations: 4
```

```
uc18 <- glm(y ~ unclass(brain), binomial(link='logit'), data = train_0_1)
summary(uc18)
```

```
##
## Call:
## glm(formula = y ~ unclass(brain), family = binomial(link = "logit"),
##      data = train_0_1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.8609  -0.6360  -0.6360  -0.6360   1.8427
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -1.84239    0.05720 -32.210 < 2e-16 ***
## unclass(brain)  0.34693    0.05099   6.804 1.02e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 20561  on 21390  degrees of freedom
## Residual deviance: 20519  on 21389  degrees of freedom
## AIC: 20523
##
## Number of Fisher Scoring iterations: 4
```

```
uc19 <- glm(y ~ unclass(lung), binomial(link='logit'), data = train_0_1)
summary(uc19)
```

```
##
## Call:
## glm(formula = y ~ unclass(lung), family = binomial(link = "logit"),
##      data = train_0_1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.8815  -0.6026  -0.6026  -0.6026   1.8951
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -2.04883    0.04803 -42.65  <2e-16 ***
## unclass(lung)  0.43467    0.03303  13.16  <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: 20561  on 21390  degrees of freedom
## Residual deviance: 20395  on 21389  degrees of freedom
## AIC: 20399
##
## Number of Fisher Scoring iterations: 4
```


0+1 multivariate glm

```
m1 <- glm(y ~ ., binomial(link='logit'), data = train_0_1)
summary(m1)
```

```
##
## Call:
## glm(formula = y ~ ., family = binomial(link = "logit"), data = train_0_1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.0990  -0.5589  -0.3710  -0.2374   2.9458
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -1.13217    0.21752  -5.205 1.94e-07 ***
## gender2      -0.01054    0.04175  -0.252 0.800702
## race2         0.02422    0.05774   0.419 0.674888
## race3        -0.11628    0.07089  -1.640 0.100958
## age2         -0.03198    0.14643  -0.218 0.827118
## age3         0.35181    0.14451   2.434 0.014913 *
## size2         0.21639    0.05129   4.219 2.46e-05 ***
## size3         0.16808    0.05395   3.115 0.001837 **
## marry2       -0.14332    0.05174  -2.770 0.005608 **
## marry3        0.02727    0.05849   0.466 0.641115
## income2      -0.03092    0.05652  -0.547 0.584337
## income3      -0.09560    0.06394  -1.495 0.134878
## site2         0.33178    0.04325   7.671 1.70e-14 ***
## grade2        0.14760    0.12159   1.214 0.224801
## grade3        0.86660    0.12762   6.791 1.12e-11 ***
## grade4        0.70851    0.17388   4.075 4.61e-05 ***
## grade9        0.75913    0.12309   6.167 6.94e-10 ***
## kind2         0.25424    0.07278   3.493 0.000477 ***
## t2           -0.10205    0.07350  -1.389 0.164965
## t3           -0.03713    0.06925  -0.536 0.591842
## N2            0.17271    0.06135   2.815 0.004873 **
## N3            0.23125    0.05489   4.213 2.52e-05 ***
## surgery1     -0.67583    0.06649 -10.165 < 2e-16 ***
## RX1          -0.11327    0.06757  -1.676 0.093643 .
## radiate1     -0.39118    0.07534  -5.192 2.08e-07 ***
## chem1        -2.06344    0.04321 -47.751 < 2e-16 ***
## CEA1          0.38690    0.07932   4.878 1.07e-06 ***
## CEA2          0.26743    0.08473   3.156 0.001599 **
## bone1         0.67445    0.07825   8.619 < 2e-16 ***
## bone2         0.06867    0.21305   0.322 0.747208
## brain1        1.09543    0.17356   6.311 2.76e-10 ***
## brain2       -0.27908    0.19847  -1.406 0.159666
## lung1         0.33646    0.04589   7.332 2.26e-13 ***
## lung2         0.02545    0.15075   0.169 0.865912
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
```

```
##
## Null deviance: 20561 on 21390 degrees of freedom
## Residual deviance: 16014 on 21357 degrees of freedom
## AIC: 16082
##
## Number of Fisher Scoring iterations: 5
```

```
sqrt(rms::vif(m1))
```

```
## gender2 race2 race3 age2 age3 size2 size3 marry2
## 1.031636 1.038400 1.029077 3.195516 3.229750 1.158242 1.295888 1.286012
## marry3 income2 income3 site2 grade2 grade3 grade4 grade9
## 1.298958 1.397235 1.422765 1.061211 2.950419 2.416176 1.367010 2.967779
## kind2 t2 t3 N2 N3 surgery1 RX1 radiate1
## 1.018611 1.827362 1.674100 1.226231 1.131576 1.605541 1.059733 1.076858
## chem1 CEA1 CEA2 bone1 bone2 brain1 brain2 lung1
## 1.067226 1.869983 1.865947 1.041671 1.570547 1.042671 1.558724 1.046734
## lung2
## 1.166541
```

0+1 multivariate glm anova

```
anova(m1, 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			21390	20561	
## gender	1	12.58	21389	20548	0.0003894 ***
## race	2	10.89	21387	20537	0.0043072 **
## age	2	481.77	21385	20056	< 2.2e-16 ***
## size	2	279.42	21383	19776	< 2.2e-16 ***
## marry	2	115.73	21381	19660	< 2.2e-16 ***
## income	2	10.17	21379	19650	0.0061880 **
## site	1	94.20	21378	19556	< 2.2e-16 ***
## grade	4	410.26	21374	19146	< 2.2e-16 ***
## kind	1	14.88	21373	19131	0.0001147 ***
## t	2	108.09	21371	19023	< 2.2e-16 ***
## N	2	45.33	21369	18978	1.433e-10 ***
## surgery	1	168.08	21368	18810	< 2.2e-16 ***
## RX	1	26.41	21367	18783	2.762e-07 ***
## radiate	1	51.85	21366	18731	5.988e-13 ***
## chem	1	2484.44	21365	16247	< 2.2e-16 ***
## CEA	2	35.84	21363	16211	1.646e-08 ***

```
## bone      2      93.63      21361      16117 < 2.2e-16 ***
## brain     2      50.00      21359      16067 1.392e-11 ***
## lung      2      53.67      21357      16014 2.216e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

0+1 multivariate lrm

```
m1_lrm <- lrm(y ~ ., data = train_0_1, x=T, y=T)
print(m1_lrm, digits=3)
```

```
## Logistic Regression Model
##
## lrm(formula = y ~ ., data = train_0_1, x = T, y = T)
##
##                               Model Likelihood      Discrimination      Rank Discrim.
##                               Ratio Test      Indexes      Indexes
## Obs          21391      LR chi2      4547.25      R2          0.310      C          0.814
## 0             17409      d.f.          33      R2(33,21391)0.190      Dxy        0.628
## 1             3982      Pr(> chi2) <0.0001      R2(33,9722.2)0.371      gamma     0.628
## max |deriv| 2e-08      Brier          0.116      tau-a     0.190
##
##          Coef   S.E.  Wald Z Pr(>|Z|)
## Intercept -1.132 0.218  -5.20 <0.0001
## gender=2   -0.011 0.042  -0.25 0.8007
## race=2      0.024 0.058   0.42 0.6749
## race=3     -0.116 0.071  -1.64 0.1010
## age=2       -0.032 0.146  -0.22 0.8271
## age=3       0.352 0.145   2.43 0.0149
## size=2      0.216 0.051   4.22 <0.0001
## size=3      0.168 0.054   3.12 0.0018
## marry=2     -0.143 0.052  -2.77 0.0056
## marry=3      0.027 0.058   0.47 0.6411
## income=2    -0.031 0.057  -0.55 0.5843
## income=3    -0.096 0.064  -1.50 0.1349
## site=2      0.332 0.043   7.67 <0.0001
## grade=2     0.148 0.122   1.21 0.2248
## grade=3     0.867 0.128   6.79 <0.0001
## grade=4     0.709 0.174   4.07 <0.0001
## grade=9     0.759 0.123   6.17 <0.0001
## kind=2      0.254 0.073   3.49 0.0005
## t=2         -0.102 0.073  -1.39 0.1650
## t=3         -0.037 0.069  -0.54 0.5918
## N=2         0.173 0.061   2.82 0.0049
## N=3         0.231 0.055   4.21 <0.0001
## surgery=1   -0.676 0.066 -10.16 <0.0001
## RX=1        -0.113 0.068  -1.68 0.0936
## radiate=1   -0.391 0.075  -5.19 <0.0001
## chem=1      -2.063 0.043 -47.75 <0.0001
## CEA=1       0.387 0.079   4.88 <0.0001
## CEA=2       0.267 0.085   3.16 0.0016
## bone=1      0.674 0.078   8.62 <0.0001
```

```
## bone=2      0.069 0.213   0.32 0.7472
## brain=1     1.095 0.174   6.31 <0.0001
## brain=2    -0.279 0.198  -1.41 0.1597
## lung=1      0.336 0.046   7.33 <0.0001
## lung=2      0.025 0.151   0.17 0.8659
```

```
sqrt(rms::vif(m1_lrm))
```

```
##  gender=2    race=2    race=3    age=2    age=3    size=2    size=3    marry=2
##  1.031636  1.038400  1.029077  3.195532  3.229766  1.158243  1.295889  1.286012
##   marry=3  income=2  income=3    site=2    grade=2    grade=3    grade=4    grade=9
##  1.298958  1.397235  1.422764  1.061211  2.950421  2.416181  1.367012  2.967785
##    kind=2      t=2      t=3      N=2      N=3  surgery=1      RX=1  radiate=1
##  1.018611  1.827361  1.674100  1.226233  1.131575  1.605542  1.059731  1.076858
##    chem=1    CEA=1    CEA=2    bone=1    bone=2    brain=1    brain=2    lung=1
##  1.067226  1.869994  1.865958  1.041671  1.570547  1.042671  1.558724  1.046734
##    lung=2
##  1.166542
```

0+1 ridge

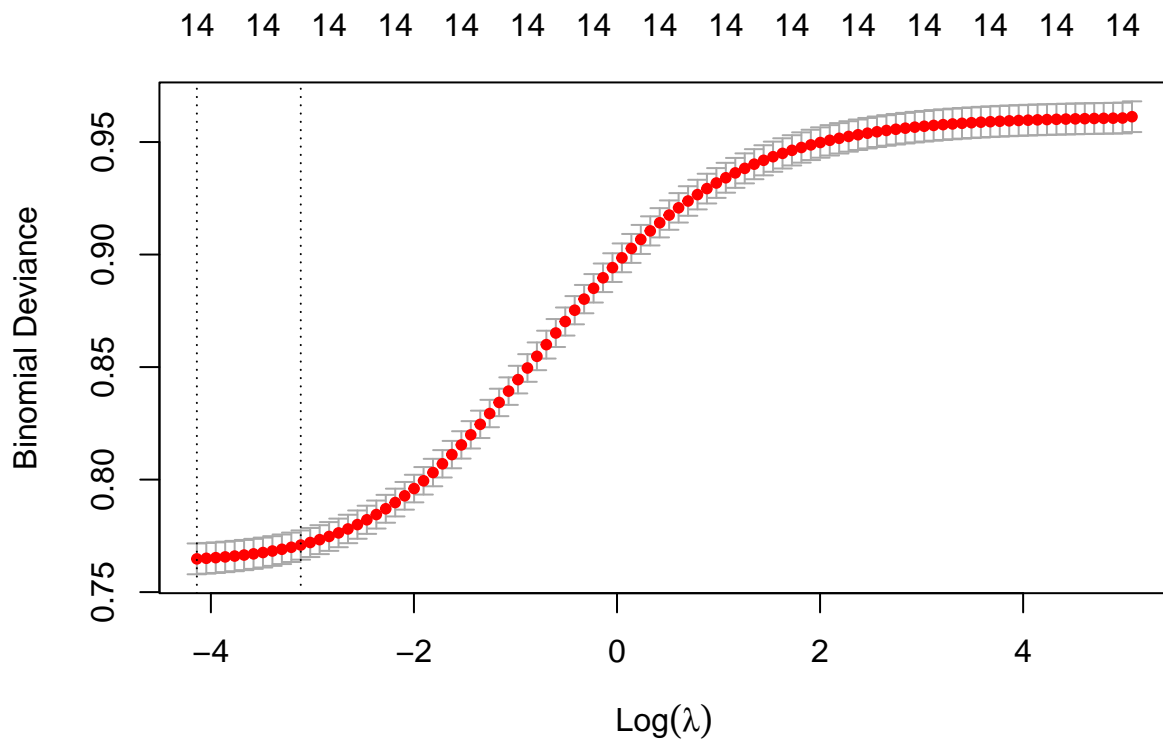
```
x <- data.matrix(train_0_1[, c('age','size','marry','site','grade','kind','N','surgery','radiate','chem')])
y <- as.numeric(data.matrix(train_0_1$y))
```

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

#find optimal lambda value that minimizes test MSE
best_lambda_rdg <- cv_model_rdg$lambda.min
best_lambda_rdg
```

```
## [1] 0.01595253
```

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



```
#find coefficients of best model
best_model_rdg <- glmnet(x, y, alpha = 0, lambda = best_lambda_rdg, family = "binomial")
coef(best_model_rdg)
```

```
## 15 x 1 sparse Matrix of class "dgCMatrix"
##              s0
## (Intercept) -0.30391216
## age         0.29916658
## size        0.09904501
## marry       0.02028276
## site        0.32607117
## grade       0.18016664
## kind        0.26983456
## N           0.11441871
## surgery     -0.57655514
## radiate     -0.22085790
## chem       -1.87230200
## CEA         0.04036846
## bone        0.29141453
## brain      -0.21157419
## lung        0.22328920
```

```
print(best_model_rdg)
```

```
##
## Call:  glmnet(x = x, y = y, family = "binomial", alpha = 0, lambda = best_lambda_rdg)
##
##   Df %Dev Lambda
## 1 14 20.58 0.01595
```

0+1 multivariate Nomogram

```
m1_lrm_final <- lrm(y~age+size+marry+site+grade+kind+N+surgery+radiate+chem+CEA+bone+brain+lung, data =
print(m1_lrm_final, digits=3)
```

```
## Logistic Regression Model
##
## lrm(formula = y ~ age + size + marry + site + grade + kind +
##      N + surgery + radiate + chem + CEA + bone + brain + lung,
##      data = train_0_1, x = T, y = T)
##
##                               Model Likelihood      Discrimination      Rank Discrim.
##                               Ratio Test          Indexes          Indexes
## Obs          21391      LR chi2      4534.90          R2          0.309      C          0.814
## 0            17409      d.f.          25          R2(25,21391)0.190      Dxy          0.627
## 1             3982      Pr(> chi2) <0.0001      R2(25,9722.2)0.371      gamma        0.628
## max |deriv| 9e-09          Brier          0.116      tau-a        0.190
##
##      Coef   S.E.  Wald Z Pr(>|Z|)
## Intercept -1.246 0.205  -6.08 <0.0001
## age=2      -0.028 0.146  -0.19 0.8493
## age=3       0.360 0.144   2.49 0.0127
## size=2      0.210 0.051   4.11 <0.0001
## size=3      0.176 0.053   3.30 0.0010
## marry=2     -0.155 0.051  -3.05 0.0023
## marry=3      0.028 0.058   0.49 0.6263
## site=2      0.344 0.043   8.04 <0.0001
## grade=2     0.144 0.121   1.19 0.2343
## grade=3     0.860 0.127   6.75 <0.0001
## grade=4     0.710 0.174   4.09 <0.0001
## grade=9     0.754 0.123   6.13 <0.0001
## kind=2      0.255 0.073   3.51 0.0005
## N=2         0.165 0.061   2.71 0.0067
## N=3         0.227 0.054   4.22 <0.0001
## surgery=1   -0.739 0.059 -12.60 <0.0001
## radiate=1   -0.403 0.075  -5.37 <0.0001
## chem=1      -2.067 0.043 -48.02 <0.0001
## CEA=1       0.389 0.079   4.91 <0.0001
## CEA=2       0.276 0.085   3.27 0.0011
## bone=1      0.675 0.078   8.63 <0.0001
## bone=2      0.067 0.213   0.31 0.7531
## brain=1     1.087 0.173   6.29 <0.0001
## brain=2    -0.271 0.198  -1.36 0.1724
## lung=1      0.337 0.046   7.36 <0.0001
## lung=2      0.020 0.151   0.13 0.8971
```

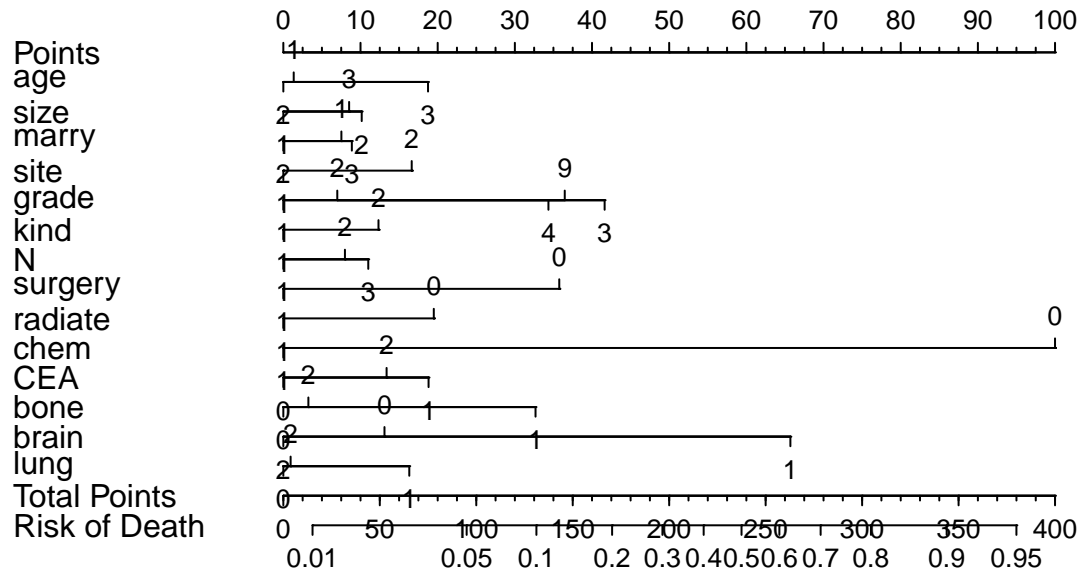
```
sqrt(rms::vif(m1_lrm_final))
```

```
##      age=2      age=3      size=2      size=3      marry=2      marry=3      site=2      grade=2
## 3.191693 3.225575 1.152737 1.282667 1.267494 1.278542 1.049828 2.947364
##      grade=3      grade=4      grade=9      kind=2      N=2      N=3      surgery=1      radiate=1
## 2.412455 1.365628 2.963661 1.017619 1.216355 1.112216 1.416621 1.072399
```

```
##      chem=1      CEA=1      CEA=2      bone=1      bone=2      brain=1      brain=2      lung=1
## 1.063092 1.869172 1.864437 1.041242 1.571584 1.039564 1.560201 1.045359
##      lung=2
## 1.165897
```

```
ddist <- datadist(train_0_1)
options(datadist='ddist')
```

```
nom1 <- nomogram(m1_lrm_final, fun=function(x)1/(1+exp(-x)),fun.at=c(.001, .01, .05, seq(.1,.9, by=.1)),
plot(nom1)
```



0+1 multivariate calibrate

```
m1_final <- glm(y~age+size+marry+site+grade+kind+N+surgery+radiate+chem+CEA+bone+brain+lung, binomial(1),
summary(m1_final)
```

```
##
## Call:
## glm(formula = y ~ age + size + marry + site + grade + kind +
##      N + surgery + radiate + chem + CEA + bone + brain + lung,
##      family = binomial(link = "logit"), data = train_0_1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.1118  -0.5562  -0.3734  -0.2397   2.9465
##
## Coefficients:
##      Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.24561    0.20492  -6.078 1.21e-09 ***
## age2        -0.02778    0.14622  -0.190 0.849316
## age3         0.35967    0.14428   2.493 0.012674 *
## size2        0.20989    0.05104   4.112 3.92e-05 ***
```

```
## size3      0.17618    0.05339    3.300 0.000967 ***
## marry2     -0.15543    0.05098   -3.049 0.002299 **
## marry3      0.02803    0.05757    0.487 0.626349
## site2      0.34376    0.04277    8.037 9.21e-16 ***
## grade2     0.14443    0.12143    1.189 0.234286
## grade3     0.86035    0.12741    6.753 1.45e-11 ***
## grade4     0.70993    0.17359    4.090 4.32e-05 ***
## grade9     0.75361    0.12288    6.133 8.63e-10 ***
## kind2      0.25483    0.07270    3.505 0.000456 ***
## N2         0.16494    0.06083    2.711 0.006700 **
## N3         0.22744    0.05395    4.215 2.49e-05 ***
## surgery1   -0.73887    0.05863  -12.602 < 2e-16 ***
## radiate1   -0.40290    0.07501   -5.371 7.81e-08 ***
## chem1      -2.06653    0.04303  -48.021 < 2e-16 ***
## CEA1       0.38894    0.07926    4.907 9.26e-07 ***
## CEA2       0.27648    0.08463    3.267 0.001088 **
## bone1      0.67532    0.07821    8.635 < 2e-16 ***
## bone2      0.06704    0.21315    0.315 0.753125
## brain1     1.08700    0.17272    6.294 3.10e-10 ***
## brain2     -0.27083    0.19850   -1.364 0.172440
## lung1      0.33727    0.04583    7.360 1.84e-13 ***
## lung2      0.01951    0.15083    0.129 0.897086
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##    Null deviance: 20561  on 21390  degrees of freedom
## Residual deviance: 16026  on 21365  degrees of freedom
## AIC: 16078
##
## Number of Fisher Scoring iterations: 5
```

```
sqrt(rms::vif(m1_final))
```

```
##      age2      age3      size2      size3      marry2      marry3      site2      grade2
## 3.191680 3.225563 1.152736 1.282667 1.267494 1.278542 1.049829 2.947362
##      grade3      grade4      grade9      kind2      N2      N3      surgery1      radiate1
## 2.412450 1.365627 2.963656 1.017619 1.216353 1.112217 1.416619 1.072399
##      chem1      CEA1      CEA2      bone1      bone2      brain1      brain2      lung1
## 1.063092 1.869162 1.864427 1.041242 1.571584 1.039564 1.560201 1.045359
##      lung2
## 1.165897
```

test model

```
m1_final_test <- glm(y~age+size+marry+site+grade+kind+N+surgery+radiate+chem+CEA+bone+brain+lung, binom
```


train

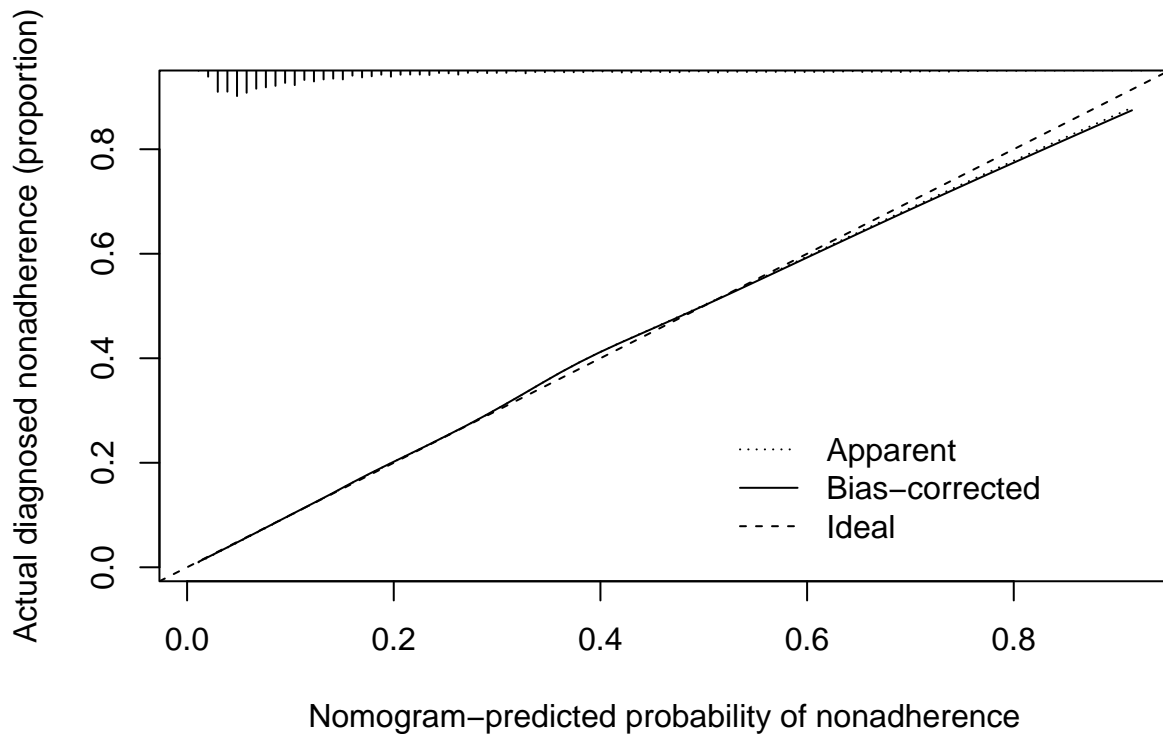
```
# hosmer-lemeshow
```

```
p.hoslem <- hoslem.test(m1_final$y, fitted(m1_final), g=10)$p.value  
p.hoslem
```

```
## [1] 0.4859891
```

```
# plot 1
```

```
refit <- lrm(y~age+size+marry+site+grade+kind+N+surgery+radiate+chem+CEA+bone+brain+lung, data = train_0_1)  
cal <- calibrate(refit, cmethod='hare', method='boot', B=1000, data=train_0_1)  
plot(cal, xlab="Nomogram-predicted probability of nonadherence", ylab="Actual diagnosed nonadherence (p
```



```
##  
## n=21391 Mean absolute error=0.002 Mean squared error=2e-05  
## 0.9 Quantile of absolute error=0.008
```

```
# plot 2
```

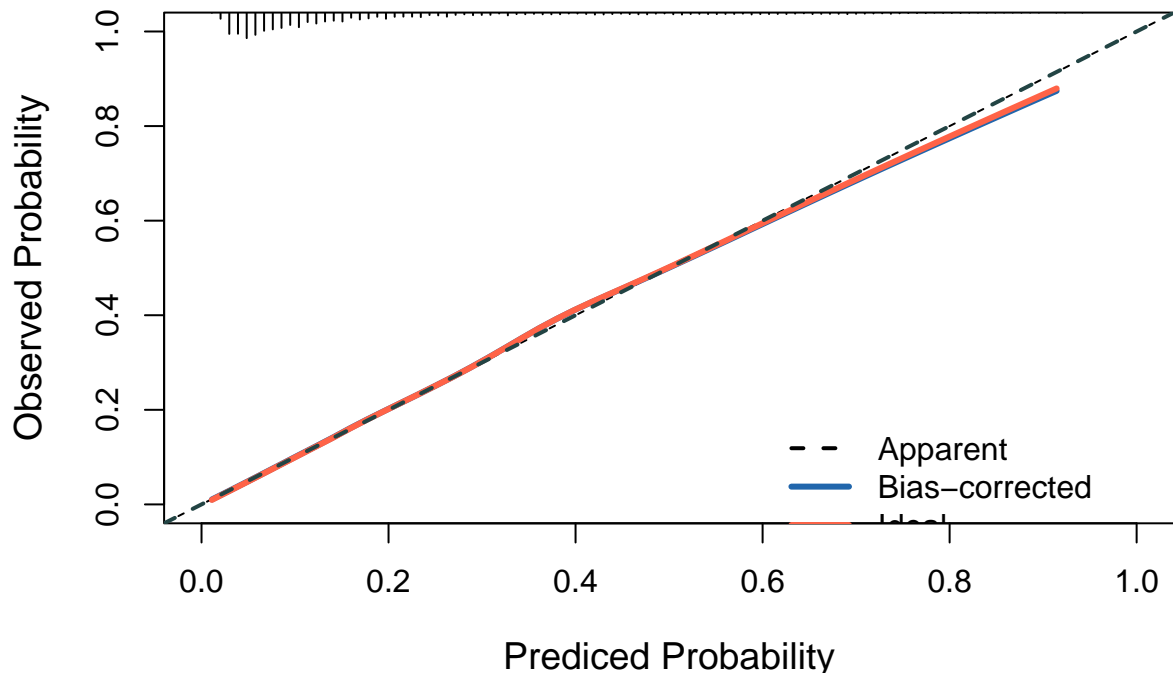
```
plot(cal, xlim = c(0,1), ylim = c(0,1), xlab = "Prediced Probability", ylab = "Observed Probability", co
```

```
##  
## n=21391 Mean absolute error=0.002 Mean squared error=2e-05  
## 0.9 Quantile of absolute error=0.008
```

```

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=0.002 n=21391

plot 3

```

plot(cal, xlim = c(0,1), ylim = c(0,1), xlab = "Prediced Probability", ylab = "Observed Probability", col = "black", lty = 2, lwd = 2)

```

##

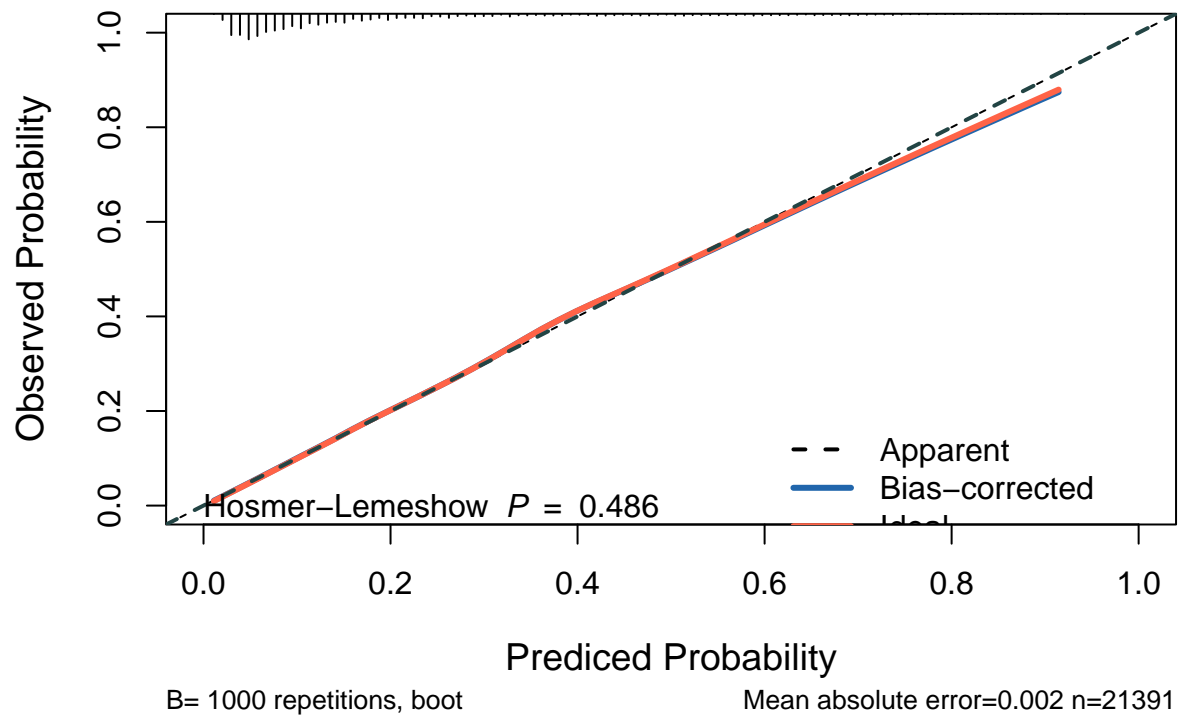
n=21391 Mean absolute error=0.002 Mean squared error=2e-05

0.9 Quantile of absolute error=0.008

```

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

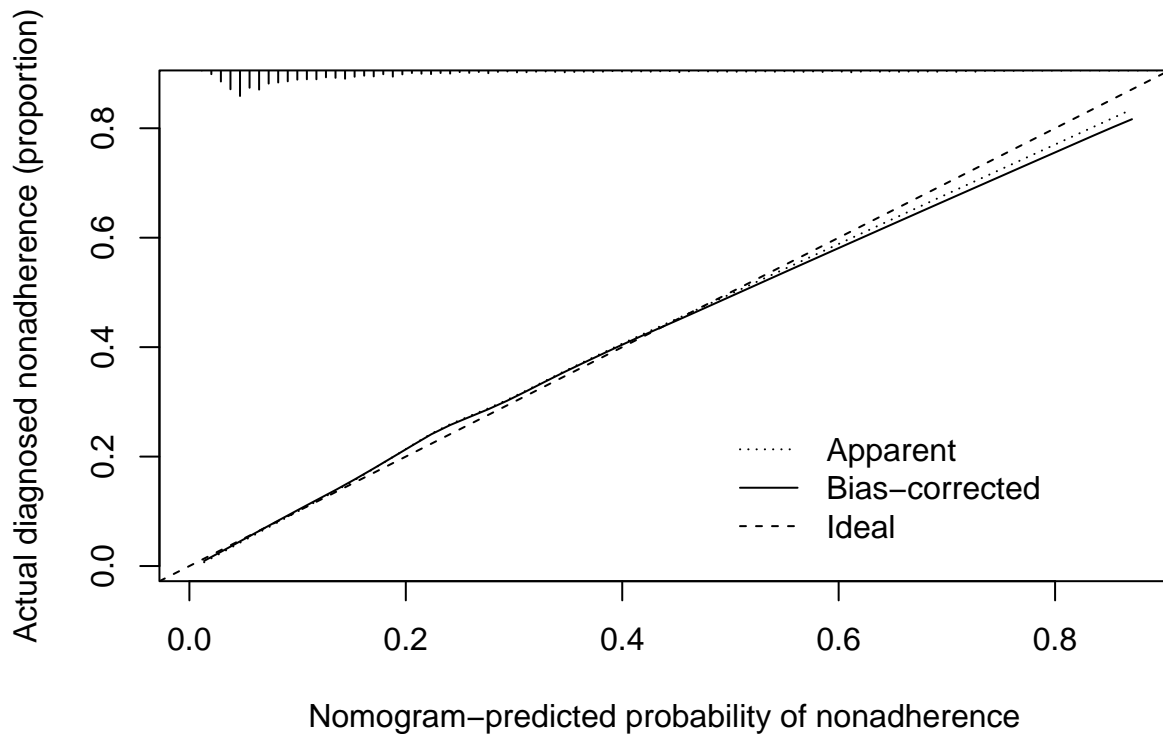
```
# hosmer-lemeshow
```

```
p.hoslem <- hoslem.test(m1_final_test$y, fitted(m1_final_test), g=10)$p.value
p.hoslem
```

```
## [1] 0.7720614
```

```
# plot 1
```

```
refit <- lrm(y~age+size+marry+site+grade+kind+N+surgery+radiate+chem+CEA+bone+brain+lung, data = test_0_1)
cal <- calibrate(refit, cmethod='hare', method='boot', B=1000, data=test_0_1)
plot(cal, xlab="Nomogram-predicted probability of nonadherence", ylab="Actual diagnosed nonadherence (p
```



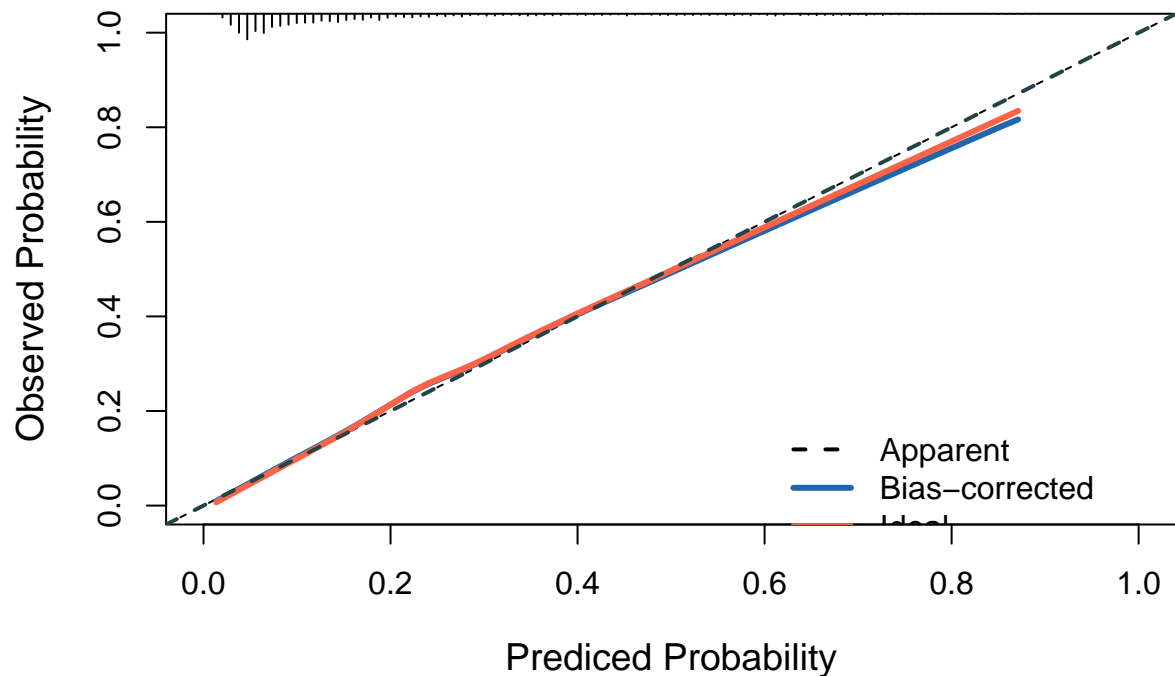
```
##
## n=5348   Mean absolute error=0.006   Mean squared error=0.00011
## 0.9 Quantile of absolute error=0.017
```

```
# 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=5348   Mean absolute error=0.006   Mean squared error=0.00011
## 0.9 Quantile of absolute error=0.017
```

```
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=0.006 n=5348

```
# plot 3
```

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

```
##
```

```
## n=5348 Mean absolute error=0.006 Mean squared error=0.00011
```

```
## 0.9 Quantile of absolute error=0.017
```

```
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"),
```

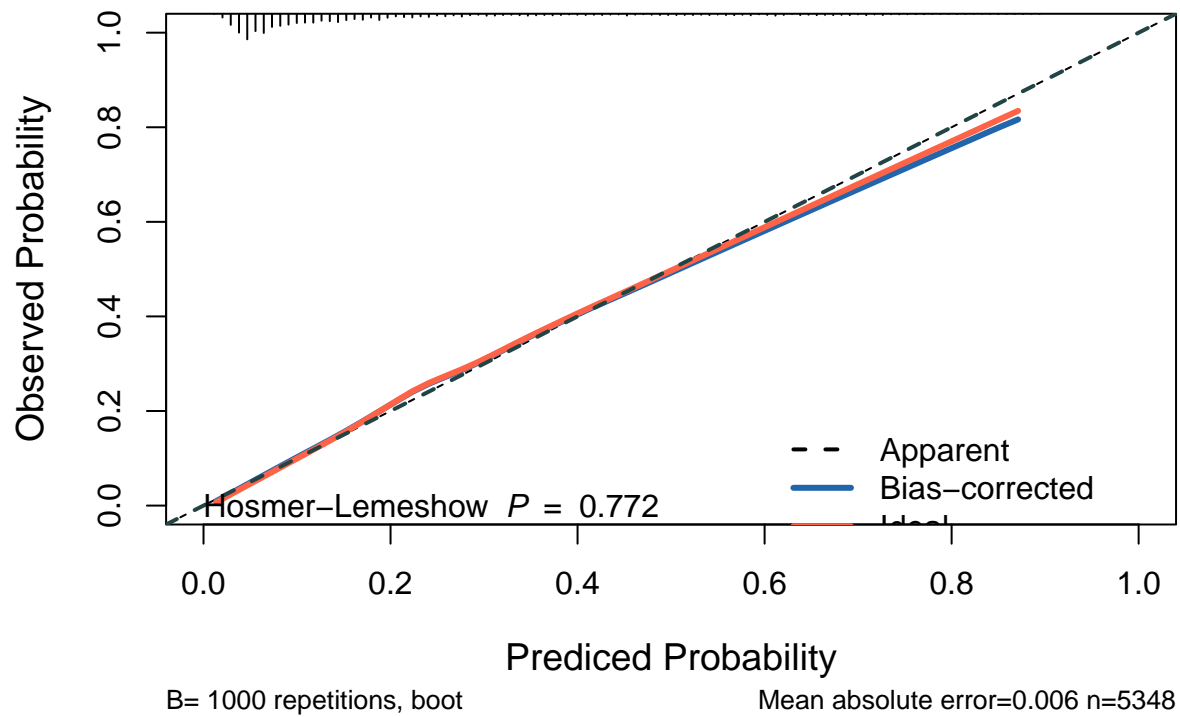
```
  lty = c(2,1,1),
```

```
  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)
```



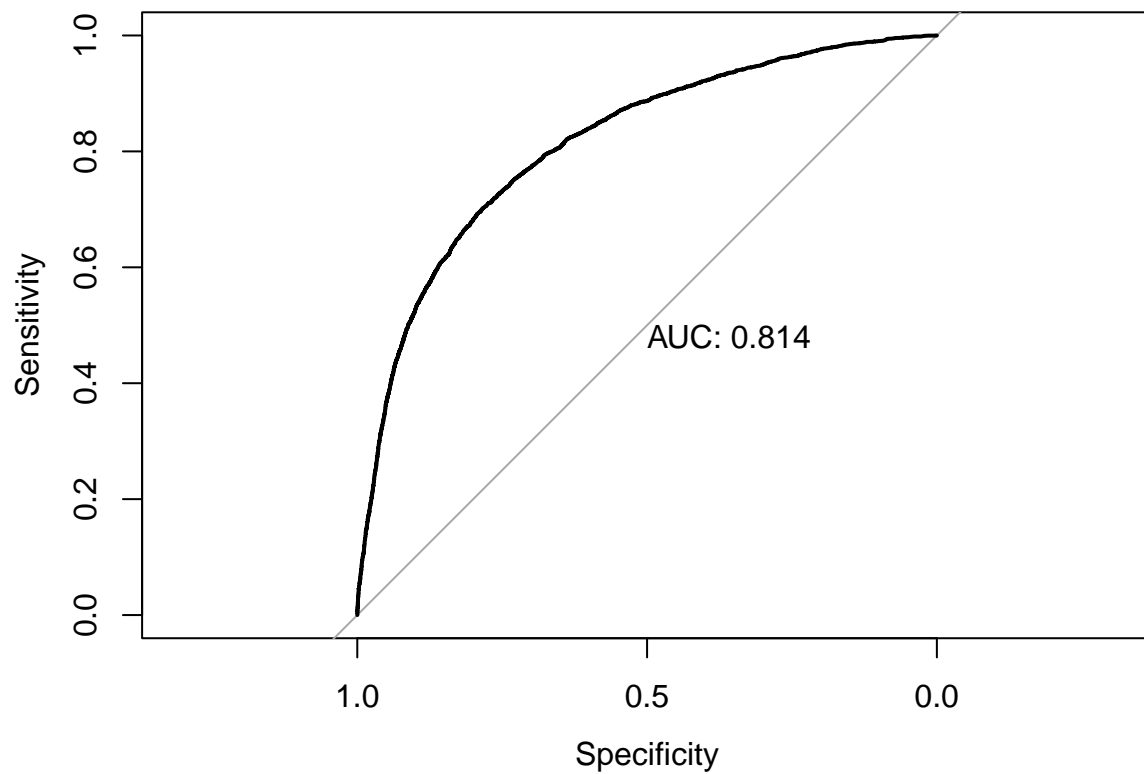
0+1 multivariate ROC

train

```
train_prob = predict(m1_final, newdata = train_0_1, type = "response")
train_roc = roc(train_0_1$y ~ train_prob, plot = TRUE, print.auc = TRUE)
```

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

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



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

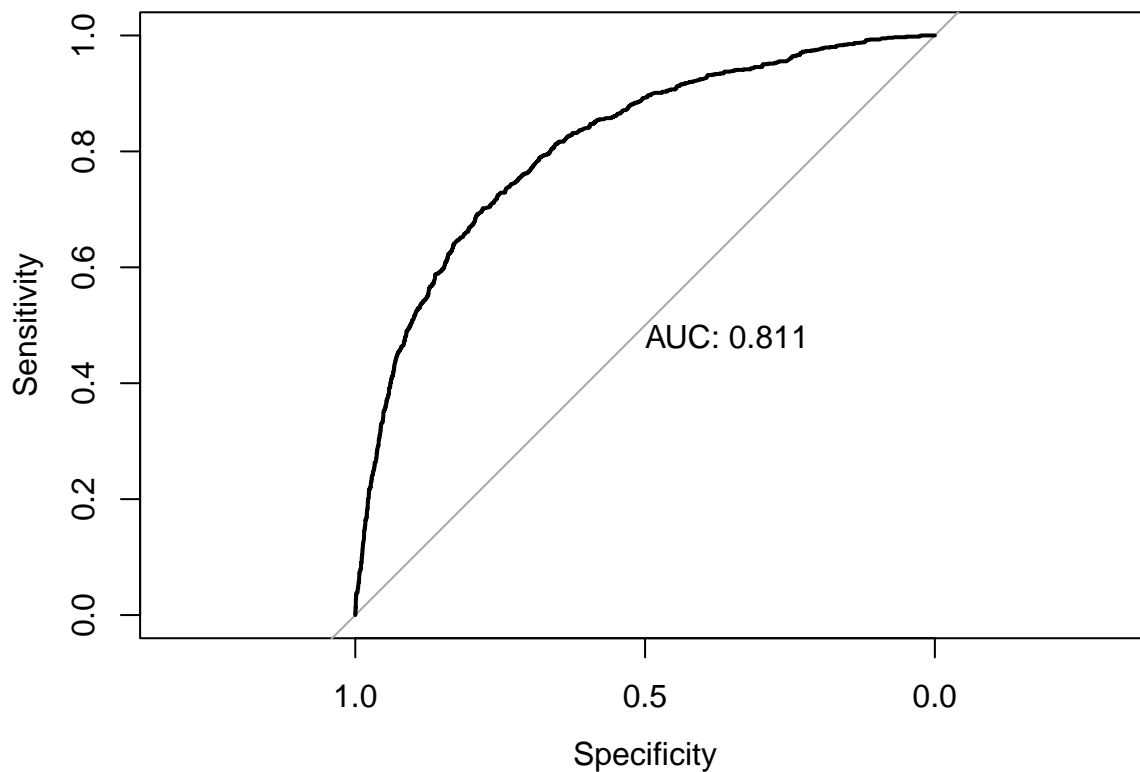
```
## [1] 0.8136104
```

```
test
```

```
test_prob = predict(m1_final, newdata = test_0_1, type = "response")
test_roc = roc(test_0_1$y ~ test_prob, plot = TRUE, print.auc = TRUE)
```

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

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



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

```
## [1] 0.8110304
```

0+1 multivariate DCA

train

```
train_0_1$y <- as.numeric(levels(train_0_1$y))[train_0_1$y]
str(train_0_1)
```

```
## tibble [21,391 x 20] (S3: tbl_df/tbl/data.frame)
## $ y      : num [1:21391] 1 0 0 0 0 0 1 0 1 0 ...
## $ gender : Factor w/ 2 levels "1","2": 1 2 2 2 1 2 1 2 1 2 ...
## $ race   : Factor w/ 3 levels "1","2","3": 1 1 1 1 1 1 1 2 2 1 ...
## $ age    : Factor w/ 3 levels "1","2","3": 3 3 1 2 2 3 3 2 3 2 ...
## $ size   : Factor w/ 3 levels "1","2","3": 2 3 3 1 1 2 2 1 1 3 ...
## $ marry  : Factor w/ 3 levels "1","2","3": 3 2 2 2 2 3 3 1 3 2 ...
## $ income : Factor w/ 3 levels "1","2","3": 2 2 1 3 2 2 2 3 2 2 ...
## $ site   : Factor w/ 2 levels "1","2": 2 1 1 1 1 1 2 2 1 1 ...
## $ grade  : Factor w/ 5 levels "1","2","3","4",...: 5 2 2 2 2 5 5 5 5 5 ...
## $ kind   : Factor w/ 2 levels "1","2": 1 1 1 1 1 1 1 1 1 1 ...
## $ t      : Factor w/ 3 levels "1","2","3": 2 3 3 2 2 3 2 3 1 2 ...
## $ N      : Factor w/ 3 levels "1","2","3": 1 1 1 1 1 1 1 3 3 1 ...
## $ surgery: Factor w/ 2 levels "0","1": 1 2 1 1 2 1 2 1 2 1 ...
## $ RX     : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 2 1 1 1 ...
```



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

```
model_1 <- decision_curve(y~age+size+marry+site+grade+kind+N+surgery+radiate+chem+CEA+bone+brain+lung, c
```

```
## 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 ~ age + size + marry + site + grade + kind + N + surgery + 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)	0.996 (0.996, 0.996)	0.996 (0.996, 0.996)
0.02	1:49	100 (100, 100)	0.993 (0.993, 0.993)	0.993 (0.993, 0.993)
0.03	3:97	100 (100, 100)	0.989 (0.989, 0.989)	0.989 (0.989, 0.989)
0.04	1:24	100 (100, 100)	0.985 (0.985, 0.985)	0.985 (0.985, 0.985)
0.05	1:19	100 (100, 100)	0.981 (0.981, 0.981)	0.981 (0.981, 0.981)
0.06	3:47	100 (100, 100)	0.977 (0.977, 0.977)	0.977 (0.977, 0.977)
0.07	7:93	100 (100, 100)	0.973 (0.973, 0.973)	0.973 (0.973, 0.973)
0.08	2:23	100 (99.998, 100)	0.969 (0.969, 0.969)	0.969 (0.969, 0.969)
0.09	9:91	100 (99.995, 100)	0.964 (0.964, 0.964)	0.964 (0.964, 0.964)

##					
##	0.1	1:9	100	0.96	0.96
##			(99.974, 100)	(0.96, 0.96)	(0.96, 0.96)
##					
##	0.11	11:89	100	0.955	0.955
##			(99.957, 100)	(0.955, 0.955)	(0.955, 0.955)
##					
##	0.12	3:22	99.98	0.951	0.951
##			(99.916, 100)	(0.951, 0.951)	(0.951, 0.951)
##					
##	0.13	13:87	99.977	0.946	0.946
##			(99.882, 100)	(0.946, 0.946)	(0.946, 0.946)
##					
##	0.14	7:43	99.937	0.941	0.941
##			(99.814, 99.997)	(0.941, 0.941)	(0.941, 0.942)
##					
##	0.15	3:17	99.915	0.936	0.936
##			(99.715, 99.983)	(0.936, 0.936)	(0.936, 0.937)
##					
##	0.16	4:21	99.889	0.931	0.931
##			(99.608, 99.97)	(0.931, 0.931)	(0.931, 0.932)
##					
##	0.17	17:83	99.718	0.926	0.927
##			(99.47, 99.933)	(0.926, 0.926)	(0.926, 0.927)
##					
##	0.18	9:41	99.665	0.921	0.921
##			(99.323, 99.872)	(0.921, 0.921)	(0.921, 0.922)
##					
##	0.19	19:81	99.628	0.915	0.916
##			(99.152, 99.768)	(0.915, 0.915)	(0.915, 0.917)
##					
##	0.2	1:4	99.387	0.91	0.911
##			(98.938, 99.654)	(0.91, 0.91)	(0.91, 0.912)
##					
##	0.21	21:79	99.254	0.904	0.905
##			(98.551, 99.538)	(0.904, 0.904)	(0.904, 0.907)
##					
##	0.22	11:39	99.05	0.898	0.9
##			(98.219, 99.376)	(0.898, 0.898)	(0.898, 0.901)
##					
##	0.23	23:77	98.839	0.892	0.895
##			(97.902, 99.175)	(0.892, 0.892)	(0.892, 0.896)
##					
##	0.24	6:19	98.259	0.886	0.889
##			(97.511, 98.935)	(0.886, 0.886)	(0.887, 0.891)
##					
##	0.25	1:3	98.038	0.88	0.883
##			(97.068, 98.612)	(0.88, 0.88)	(0.881, 0.886)
##					
##	0.26	13:37	97.782	0.873	0.878
##			(96.59, 98.187)	(0.873, 0.873)	(0.874, 0.88)
##					
##	0.27	27:73	96.832	0.866	0.87
##			(96.161, 97.909)	(0.866, 0.866)	(0.868, 0.874)

##					
##	0.28	7:18	96.626	0.86	0.864
##			(95.717, 97.409)	(0.86, 0.86)	(0.861, 0.868)
##					
##	0.29	29:71	96.35	0.853	0.858
##			(95.166, 96.989)	(0.853, 0.853)	(0.855, 0.863)
##					
##	0.3	3:7	95.81	0.845	0.853
##			(94.689, 96.497)	(0.845, 0.845)	(0.848, 0.857)
##					
##	0.31	31:69	95.014	0.838	0.847
##			(94.058, 95.988)	(0.838, 0.838)	(0.842, 0.851)
##					
##	0.32	8:17	94.687	0.83	0.841
##			(93.452, 95.515)	(0.83, 0.83)	(0.836, 0.845)
##					
##	0.33	33:67	94.251	0.822	0.834
##			(92.793, 94.891)	(0.822, 0.822)	(0.829, 0.838)
##					
##	0.34	17:33	93.481	0.814	0.827
##			(92.159, 94.292)	(0.814, 0.814)	(0.822, 0.832)
##					
##	0.35	7:13	92.582	0.806	0.82
##			(91.468, 93.716)	(0.806, 0.806)	(0.815, 0.825)
##					
##	0.36	9:16	91.839	0.797	0.813
##			(90.817, 93.055)	(0.797, 0.797)	(0.807, 0.819)
##					
##	0.37	37:63	91.399	0.788	0.805
##			(90.165, 92.372)	(0.788, 0.788)	(0.8, 0.812)
##					
##	0.38	19:31	90.785	0.779	0.798
##			(89.417, 91.551)	(0.779, 0.779)	(0.793, 0.805)
##					
##	0.39	39:61	89.931	0.769	0.792
##			(88.8, 90.873)	(0.769, 0.769)	(0.785, 0.797)
##					
##	0.4	2:3	88.975	0.759	0.783
##			(88.091, 90.169)	(0.759, 0.759)	(0.777, 0.79)
##					
##	0.41	41:59	88.405	0.749	0.774
##			(87.336, 89.426)	(0.749, 0.749)	(0.769, 0.783)
##					
##	0.42	21:29	87.948	0.739	0.767
##			(86.673, 88.696)	(0.739, 0.739)	(0.761, 0.776)
##					
##	0.43	43:57	86.825	0.728	0.76
##			(85.892, 87.883)	(0.728, 0.728)	(0.754, 0.77)
##					
##	0.44	11:14	85.93	0.716	0.752
##			(85.223, 87.052)	(0.716, 0.716)	(0.745, 0.762)
##					
##	0.45	9:11	85.3	0.705	0.745
##			(84.529, 86.337)	(0.705, 0.705)	(0.738, 0.755)

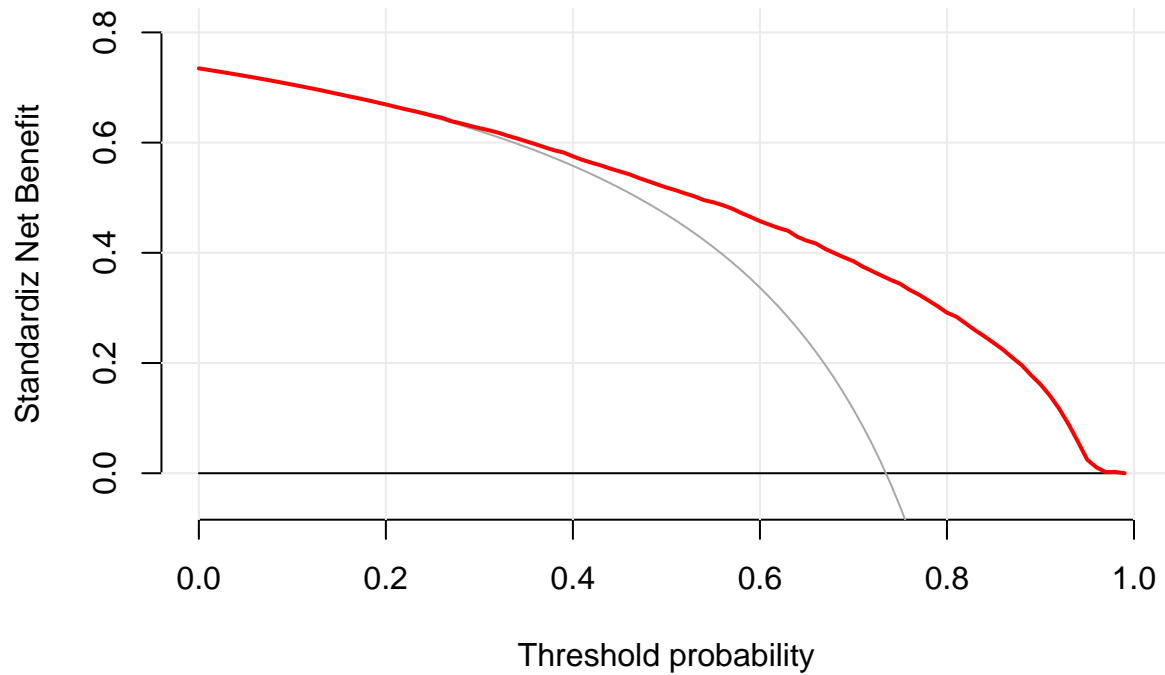
##					
##	0.46	23:27	84.944	0.692	0.739
##			(83.879, 85.554)	(0.692, 0.692)	(0.73, 0.747)
##					
##	0.47	47:53	84.02	0.68	0.73
##			(83.053, 84.78)	(0.68, 0.68)	(0.721, 0.741)
##					
##	0.48	12:13	82.83	0.667	0.722
##			(82.316, 83.996)	(0.667, 0.667)	(0.714, 0.734)
##					
##	0.49	49:51	82.267	0.653	0.714
##			(81.674, 83.235)	(0.653, 0.653)	(0.706, 0.727)
##					
##	0.5	1:1	81.779	0.639	0.706
##			(81.02, 82.533)	(0.639, 0.639)	(0.698, 0.718)
##					
##	0.51	51:49	81.19	0.624	0.699
##			(80.214, 81.809)	(0.624, 0.624)	(0.69, 0.712)
##					
##	0.52	13:12	80.342	0.609	0.691
##			(79.397, 81.081)	(0.609, 0.609)	(0.681, 0.703)
##					
##	0.53	53:47	79.418	0.593	0.684
##			(78.627, 80.313)	(0.593, 0.593)	(0.674, 0.696)
##					
##	0.54	27:23	78.389	0.576	0.675
##			(77.813, 79.45)	(0.576, 0.576)	(0.666, 0.688)
##					
##	0.55	11:9	77.995	0.559	0.669
##			(76.923, 78.689)	(0.559, 0.559)	(0.658, 0.681)
##					
##	0.56	14:11	77.187	0.541	0.662
##			(76.02, 77.757)	(0.541, 0.541)	(0.649, 0.672)
##					
##	0.57	57:43	76.119	0.521	0.654
##			(75.098, 76.845)	(0.521, 0.521)	(0.641, 0.664)
##					
##	0.58	29:21	74.824	0.501	0.643
##			(74.13, 75.889)	(0.501, 0.501)	(0.632, 0.655)
##					
##	0.59	59:41	73.944	0.48	0.633
##			(73.107, 74.984)	(0.48, 0.48)	(0.621, 0.646)
##					
##	0.6	3:2	73.078	0.458	0.623
##			(72.098, 74.014)	(0.458, 0.458)	(0.612, 0.637)
##					
##	0.61	61:39	72.031	0.435	0.614
##			(71.043, 72.974)	(0.435, 0.435)	(0.602, 0.628)
##					
##	0.62	31:19	70.935	0.411	0.606
##			(69.911, 71.97)	(0.411, 0.411)	(0.592, 0.618)
##					
##	0.63	63:37	70.106	0.385	0.599
##			(68.918, 70.91)	(0.385, 0.385)	(0.581, 0.608)

##					
##	0.64	16:9	68.927	0.358	0.584
##			(67.905, 69.979)	(0.358, 0.358)	(0.572, 0.599)
##					
##	0.65	13:7	67.776	0.33	0.574
##			(66.771, 68.835)	(0.33, 0.33)	(0.561, 0.589)
##					
##	0.66	33:17	66.841	0.299	0.567
##			(65.519, 67.778)	(0.299, 0.299)	(0.551, 0.579)
##					
##	0.67	67:33	65.332	0.267	0.554
##			(64.421, 66.624)	(0.267, 0.267)	(0.54, 0.569)
##					
##	0.68	17:8	64.352	0.233	0.543
##			(63.267, 65.626)	(0.233, 0.233)	(0.53, 0.56)
##					
##	0.69	69:31	63.436	0.196	0.533
##			(62.319, 64.442)	(0.196, 0.196)	(0.519, 0.55)
##					
##	0.7	7:3	62.507	0.158	0.524
##			(61.072, 63.247)	(0.158, 0.158)	(0.507, 0.538)
##					
##	0.71	71:29	60.965	0.116	0.51
##			(59.998, 62.265)	(0.116, 0.116)	(0.495, 0.529)
##					
##	0.72	18:7	59.824	0.072	0.499
##			(58.894, 61.031)	(0.072, 0.072)	(0.484, 0.517)
##					
##	0.73	73:27	58.87	0.024	0.488
##			(57.68, 59.886)	(0.024, 0.024)	(0.474, 0.506)
##					
##	0.74	37:13	57.741	-0.027	0.477
##			(56.474, 58.699)	(-0.027, -0.027)	(0.462, 0.496)
##					
##	0.75	3:1	56.323	-0.083	0.468
##			(55.286, 57.374)	(-0.083, -0.083)	(0.45, 0.485)
##					
##	0.76	19:6	55.066	-0.143	0.453
##			(54.084, 56.159)	(-0.143, -0.143)	(0.438, 0.472)
##					
##	0.77	77:23	53.868	-0.209	0.441
##			(52.689, 54.877)	(-0.209, -0.209)	(0.424, 0.458)
##					
##	0.78	39:11	52.469	-0.28	0.427
##			(51.253, 53.382)	(-0.28, -0.28)	(0.41, 0.444)
##					
##	0.79	79:21	50.778	-0.358	0.413
##			(49.832, 52.039)	(-0.358, -0.358)	(0.394, 0.431)
##					
##	0.8	4:1	49.616	-0.444	0.397
##			(48.568, 50.712)	(-0.444, -0.444)	(0.379, 0.417)
##					
##	0.81	81:19	48.418	-0.539	0.387
##			(47.146, 49.494)	(-0.539, -0.539)	(0.364, 0.404)

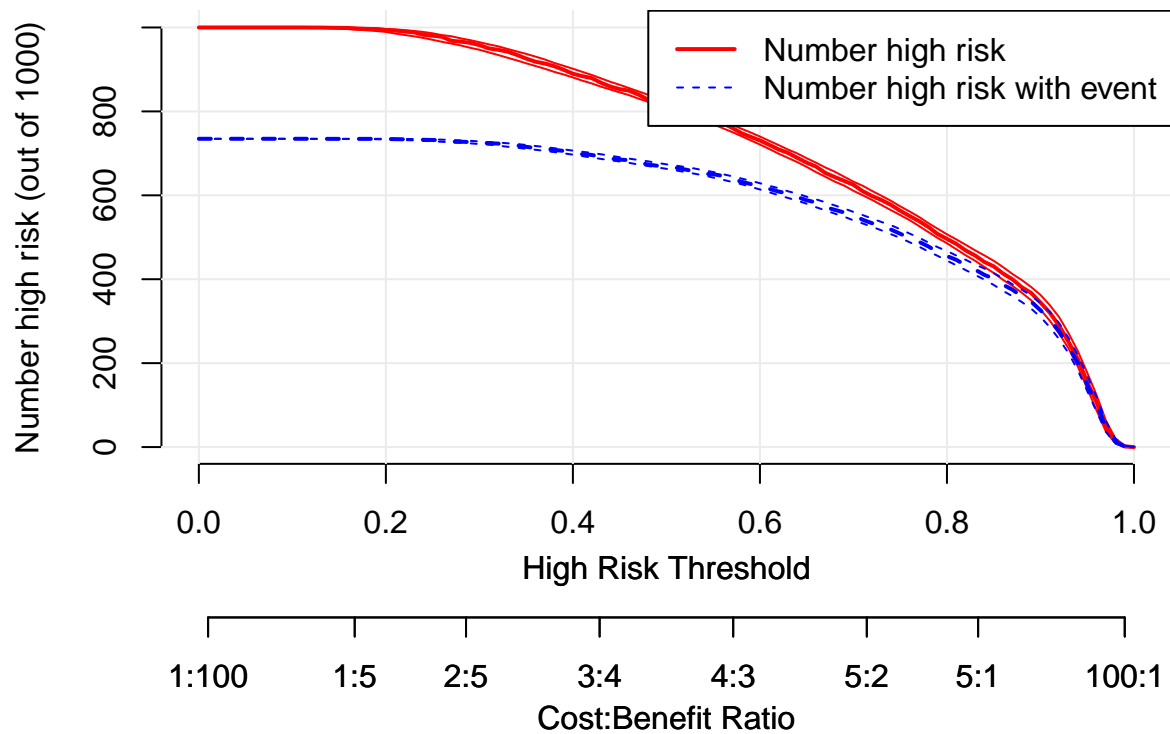
##					
##	0.82	41:9	46.866	-0.645	0.37
##			(45.706, 48.217)	(-0.645, -0.645)	(0.35, 0.388)
##					
##	0.83	83:17	45.659	-0.763	0.353
##			(44.41, 46.901)	(-0.763, -0.763)	(0.335, 0.375)
##					
##	0.84	21:4	44.121	-0.895	0.338
##			(42.92, 45.68)	(-0.895, -0.895)	(0.321, 0.36)
##					
##	0.85	17:3	43.063	-1.046	0.322
##			(41.497, 44.387)	(-1.046, -1.046)	(0.305, 0.344)
##					
##	0.86	43:7	41.343	-1.218	0.305
##			(39.851, 42.89)	(-1.218, -1.218)	(0.287, 0.325)
##					
##	0.87	87:13	39.984	-1.416	0.285
##			(38.412, 41.249)	(-1.416, -1.416)	(0.268, 0.306)
##					
##	0.88	22:3	38.11	-1.647	0.267
##			(36.784, 39.802)	(-1.647, -1.647)	(0.247, 0.287)
##					
##	0.89	89:11	36.737	-1.921	0.242
##			(34.986, 38.152)	(-1.921, -1.921)	(0.224, 0.266)
##					
##	0.9	9:1	34.436	-2.249	0.219
##			(32.783, 36.258)	(-2.249, -2.249)	(0.201, 0.242)
##					
##	0.91	91:9	32.022	-2.65	0.192
##			(30.115, 33.819)	(-2.65, -2.65)	(0.172, 0.215)
##					
##	0.92	23:2	28.906	-3.152	0.159
##			(27.024, 30.949)	(-3.152, -3.152)	(0.138, 0.183)
##					
##	0.93	93:7	25.064	-3.796	0.121
##			(23.514, 27.138)	(-3.796, -3.796)	(0.1, 0.145)
##					
##	0.94	47:3	20.951	-4.656	0.078
##			(19.181, 22.852)	(-4.656, -4.656)	(0.057, 0.104)
##					
##	0.95	19:1	15.487	-5.859	0.033
##			(13.956, 17.726)	(-5.859, -5.859)	(0.019, 0.059)
##					
##	0.96	24:1	10.682	-7.664	0.014
##			(8.734, 12.37)	(-7.664, -7.664)	(-0.003, 0.033)
##					
##	0.97	97:3	4.769	-10.673	0.003
##			(3.703, 6.298)	(-10.673, -10.673)	(-0.01, 0.022)
##					
##	0.98	49:1	1.313	-16.69	0.003
##			(0.855, 2.075)	(-16.69, -16.69)	(-0.006, 0.014)
##					
##	0.99	99:1	0.149	-34.74	0
##			(0.04, 0.358)	(-34.74, -34.74)	(-0.004, 0.005)

```
##
##      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$y <- as.factor(train_0_1$y)
str(train_0_1)
```

```
## tibble [21,391 x 20] (S3: tbl_df/tbl/data.frame)
## $ y      : Factor w/ 2 levels "0","1": 2 1 1 1 1 1 2 1 2 1 ...
## $ gender : Factor w/ 2 levels "1","2": 1 2 2 2 1 2 1 2 1 2 ...
## $ race   : Factor w/ 3 levels "1","2","3": 1 1 1 1 1 1 1 2 2 1 ...
## $ age    : Factor w/ 3 levels "1","2","3": 3 3 1 2 2 3 3 2 3 2 ...
## $ size   : Factor w/ 3 levels "1","2","3": 2 3 3 1 1 2 2 1 1 3 ...
## $ marry  : Factor w/ 3 levels "1","2","3": 3 2 2 2 2 3 3 1 3 2 ...
## $ income : Factor w/ 3 levels "1","2","3": 2 2 1 3 2 2 2 3 2 2 ...
## $ site   : Factor w/ 2 levels "1","2": 2 1 1 1 1 1 2 2 1 1 ...
## $ grade  : Factor w/ 5 levels "1","2","3","4",...: 5 2 2 2 2 5 5 5 5 5 ...
## $ kind   : Factor w/ 2 levels "1","2": 1 1 1 1 1 1 1 1 1 1 ...
## $ t      : Factor w/ 3 levels "1","2","3": 2 3 3 2 2 3 2 3 1 2 ...
## $ N      : Factor w/ 3 levels "1","2","3": 1 1 1 1 1 1 1 3 3 1 ...
## $ surgery: Factor w/ 2 levels "0","1": 1 2 1 1 2 1 2 1 2 1 ...
## $ RX     : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 2 1 1 1 ...
## $ radiate: Factor w/ 2 levels "0","1": 1 2 1 1 1 1 1 1 2 1 ...
## $ chem   : Factor w/ 2 levels "0","1": 2 2 2 2 2 1 2 2 2 2 ...
## $ CEA    : Factor w/ 3 levels "0","1","2": 2 3 2 2 3 3 2 2 2 3 ...
## $ bone   : Factor w/ 3 levels "0","1","2": 1 1 1 1 1 1 1 1 2 1 ...
## $ brain  : Factor w/ 3 levels "0","1","2": 1 1 1 1 1 1 1 1 2 1 ...
## $ lung   : Factor w/ 3 levels "0","1","2": 1 1 2 1 1 1 1 2 2 1 ...
```

```
test
```



```
test_0_1$y <- as.numeric(levels(test_0_1$y))[test_0_1$y]
str(test_0_1)
```

```
## tibble [5,348 x 20] (S3: tbl_df/tbl/data.frame)
## $ y      : num [1:5348] 0 0 0 1 1 1 0 0 0 0 ...
## $ gender : Factor w/ 2 levels "1","2": 1 1 2 1 2 2 2 1 1 2 ...
## $ race   : Factor w/ 3 levels "1","2","3": 1 2 2 1 1 3 1 1 1 2 ...
## $ age    : Factor w/ 3 levels "1","2","3": 3 3 3 3 3 3 2 3 3 3 ...
## $ size   : Factor w/ 3 levels "1","2","3": 1 3 2 1 2 1 3 3 1 1 ...
## $ marry  : Factor w/ 3 levels "1","2","3": 3 2 1 3 2 2 2 1 2 2 ...
## $ 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 1 2 2 2 1 1 1 1 1 ...
## $ grade  : Factor w/ 5 levels "1","2","3","4",...: 2 2 2 5 2 2 2 5 2 3 ...
## $ kind   : Factor w/ 2 levels "1","2": 1 1 1 1 1 1 1 1 1 1 ...
## $ t      : Factor w/ 3 levels "1","2","3": 2 1 2 3 2 2 2 3 2 2 ...
## $ N      : Factor w/ 3 levels "1","2","3": 2 1 2 1 2 1 1 3 1 2 ...
## $ surgery: Factor w/ 2 levels "0","1": 2 1 2 1 2 1 2 1 2 2 ...
## $ RX     : Factor w/ 2 levels "0","1": 1 1 1 1 2 1 1 1 1 1 ...
## $ radiate: Factor w/ 2 levels "0","1": 1 1 1 1 1 2 1 1 1 1 ...
## $ chem   : Factor w/ 2 levels "0","1": 1 2 1 1 2 2 2 2 2 2 ...
## $ CEA    : Factor w/ 3 levels "0","1","2": 1 3 2 2 2 2 1 2 2 1 ...
## $ 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 2 1 1 1 1 1 1 ...
```

```
model_1 <- decision_curve(y~age+size+marry+site+grade+kind+N+surgery+radiate+chem+CEA+bone+brain+lung, o
```

```
## 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 ~ age + size + marry + site
## threshold	## ratio	## high risk		+ grade + kind + N + surgery +
##				radiate + chem + CEA + bone +
##				brain + lung
## 0	## 0:1	## 100	## 1	## 1
		## (100, 100)	## (1, 1)	## (1, 1)
## 0.01	## 1:99	## 100	## 0.996	## 0.996
		## (100, 100)	## (0.996, 0.996)	## (0.996, 0.996)
## 0.02	## 1:49	## 100	## 0.993	## 0.993
		## (100, 100)	## (0.993, 0.993)	## (0.993, 0.993)
## 0.03	## 3:97	## 100	## 0.989	## 0.989

##			(100, 100)	(0.989, 0.989)	(0.989, 0.989)
##					
##	0.04	1:24	100	0.985	0.985
##			(100, 100)	(0.985, 0.985)	(0.985, 0.985)
##					
##	0.05	1:19	100	0.981	0.981
##			(99.994, 100)	(0.981, 0.981)	(0.981, 0.981)
##					
##	0.06	3:47	100	0.977	0.977
##			(99.988, 100)	(0.977, 0.977)	(0.977, 0.977)
##					
##	0.07	7:93	100	0.973	0.973
##			(99.982, 100)	(0.973, 0.973)	(0.973, 0.973)
##					
##	0.08	2:23	100	0.969	0.969
##			(99.969, 100)	(0.969, 0.969)	(0.969, 0.969)
##					
##	0.09	9:91	100	0.964	0.964
##			(99.951, 100)	(0.964, 0.964)	(0.964, 0.964)
##					
##	0.1	1:9	100	0.96	0.96
##			(99.914, 100)	(0.96, 0.96)	(0.96, 0.96)
##					
##	0.11	11:89	100	0.955	0.955
##			(99.829, 100)	(0.955, 0.955)	(0.955, 0.956)
##					
##	0.12	3:22	100	0.951	0.951
##			(99.792, 100)	(0.951, 0.951)	(0.951, 0.951)
##					
##	0.13	13:87	99.994	0.946	0.946
##			(99.676, 100)	(0.946, 0.946)	(0.946, 0.947)
##					
##	0.14	7:43	99.982	0.941	0.941
##			(99.566, 100)	(0.941, 0.941)	(0.941, 0.942)
##					
##	0.15	3:17	99.963	0.936	0.936
##			(99.426, 100)	(0.936, 0.936)	(0.936, 0.938)
##					
##	0.16	4:21	99.884	0.931	0.932
##			(99.304, 99.994)	(0.931, 0.931)	(0.931, 0.933)
##					
##	0.17	17:83	99.866	0.926	0.927
##			(99.133, 99.994)	(0.926, 0.926)	(0.926, 0.928)
##					
##	0.18	9:41	99.78	0.921	0.922
##			(98.987, 99.976)	(0.921, 0.921)	(0.92, 0.923)
##					
##	0.19	19:81	99.67	0.915	0.916
##			(98.816, 99.951)	(0.915, 0.915)	(0.915, 0.919)
##					
##	0.2	1:4	99.585	0.91	0.911
##			(98.566, 99.896)	(0.91, 0.91)	(0.909, 0.914)
##					
##	0.21	21:79	99.456	0.904	0.906

##			(98.218, 99.829)	(0.904, 0.904)	(0.903, 0.909)
##					
##	0.22	11:39	99.371	0.898	0.901
##			(97.949, 99.75)	(0.898, 0.898)	(0.898, 0.904)
##					
##	0.23	23:77	99.133	0.892	0.895
##			(97.59, 99.634)	(0.892, 0.892)	(0.892, 0.899)
##					
##	0.24	6:19	98.779	0.886	0.89
##			(97.26, 99.505)	(0.886, 0.886)	(0.886, 0.894)
##					
##	0.25	1:3	98.578	0.88	0.884
##			(96.851, 99.378)	(0.88, 0.88)	(0.88, 0.888)
##					
##	0.26	13:37	98.316	0.873	0.877
##			(96.339, 99.225)	(0.873, 0.873)	(0.874, 0.883)
##					
##	0.27	27:73	97.993	0.867	0.871
##			(95.663, 98.943)	(0.867, 0.867)	(0.867, 0.878)
##					
##	0.28	7:18	97.591	0.86	0.864
##			(95.085, 98.675)	(0.86, 0.86)	(0.861, 0.873)
##					
##	0.29	29:71	97.383	0.853	0.859
##			(94.425, 98.365)	(0.853, 0.853)	(0.854, 0.867)
##					
##	0.3	3:7	96.742	0.846	0.855
##			(93.859, 97.998)	(0.846, 0.846)	(0.848, 0.862)
##					
##	0.31	31:69	95.816	0.838	0.848
##			(93.103, 97.609)	(0.838, 0.838)	(0.84, 0.857)
##					
##	0.32	8:17	95.578	0.83	0.843
##			(92.578, 97.11)	(0.83, 0.83)	(0.834, 0.852)
##					
##	0.33	33:67	94.585	0.822	0.834
##			(91.783, 96.66)	(0.822, 0.822)	(0.826, 0.845)
##					
##	0.34	17:33	93.708	0.814	0.825
##			(91.008, 95.895)	(0.814, 0.814)	(0.818, 0.839)
##					
##	0.35	7:13	93.202	0.806	0.821
##			(90.338, 95.235)	(0.806, 0.806)	(0.809, 0.832)
##					
##	0.36	9:16	92.708	0.797	0.814
##			(89.684, 94.608)	(0.797, 0.797)	(0.803, 0.826)
##					
##	0.37	37:63	91.441	0.788	0.806
##			(88.932, 93.949)	(0.788, 0.788)	(0.795, 0.82)
##					
##	0.38	19:31	90.606	0.779	0.8
##			(88.304, 93.089)	(0.779, 0.779)	(0.787, 0.814)
##					
##	0.39	39:61	90.119	0.77	0.792

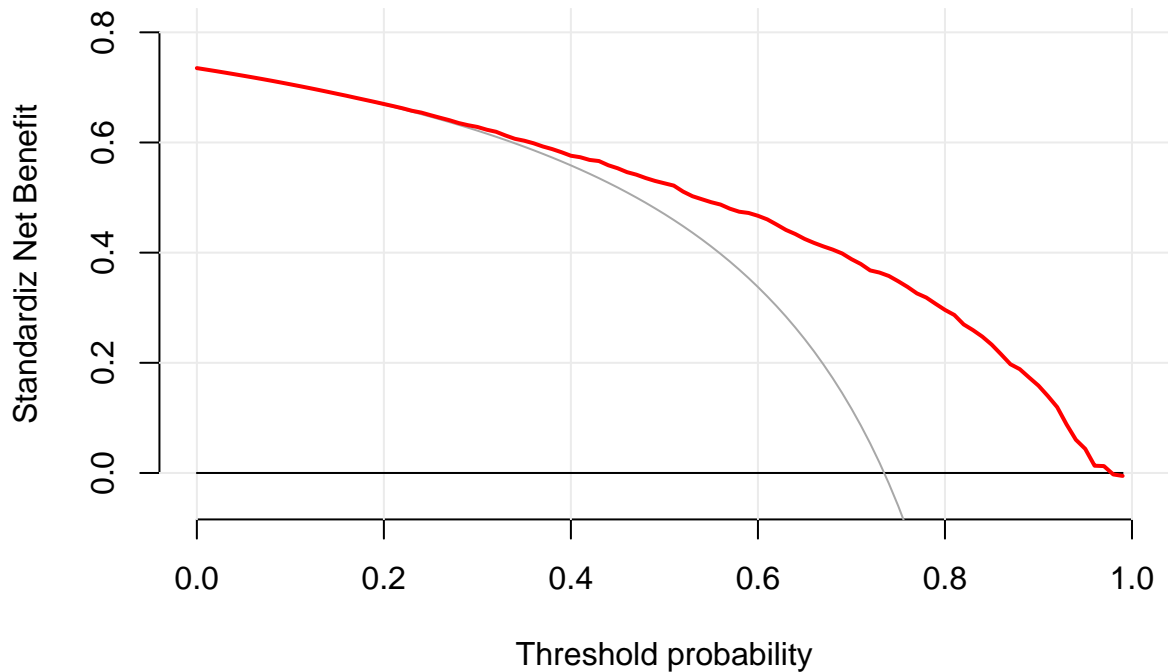
##			(87.599, 92.141)	(0.77, 0.77)	(0.78, 0.807)
##					
##	0.4	2:3	89.182	0.76	0.784
##			(87.083, 91.257)	(0.76, 0.76)	(0.773, 0.801)
##					
##	0.41	41:59	88.803	0.75	0.78
##			(86.406, 90.484)	(0.75, 0.75)	(0.766, 0.795)
##					
##	0.42	21:29	87.859	0.739	0.773
##			(85.629, 89.641)	(0.739, 0.739)	(0.758, 0.789)
##					
##	0.43	43:57	87.243	0.728	0.77
##			(85.031, 88.879)	(0.728, 0.728)	(0.75, 0.783)
##					
##	0.44	11:14	86.525	0.717	0.76
##			(84.411, 88.152)	(0.717, 0.717)	(0.742, 0.777)
##					
##	0.45	9:11	85.38	0.705	0.753
##			(83.486, 87.451)	(0.705, 0.705)	(0.737, 0.77)
##					
##	0.46	23:27	84.443	0.693	0.743
##			(82.611, 86.428)	(0.693, 0.693)	(0.729, 0.762)
##					
##	0.47	47:53	84.017	0.68	0.737
##			(81.915, 85.644)	(0.68, 0.68)	(0.724, 0.756)
##					
##	0.48	12:13	83.305	0.667	0.728
##			(81.272, 84.891)	(0.667, 0.667)	(0.716, 0.749)
##					
##	0.49	49:51	82.647	0.654	0.721
##			(80.604, 84.086)	(0.654, 0.654)	(0.706, 0.742)
##					
##	0.5	1:1	81.898	0.64	0.716
##			(79.514, 83.233)	(0.64, 0.64)	(0.697, 0.735)
##					
##	0.51	51:49	81.015	0.625	0.71
##			(78.761, 82.559)	(0.625, 0.625)	(0.688, 0.727)
##					
##	0.52	13:12	79.927	0.609	0.695
##			(77.977, 81.865)	(0.609, 0.609)	(0.679, 0.719)
##					
##	0.53	53:47	79.028	0.594	0.684
##			(77.357, 81.053)	(0.594, 0.594)	(0.671, 0.713)
##					
##	0.54	27:23	78.017	0.577	0.676
##			(76.659, 80.088)	(0.577, 0.577)	(0.66, 0.706)
##					
##	0.55	11:9	77.542	0.559	0.669
##			(75.809, 79.321)	(0.559, 0.559)	(0.652, 0.697)
##					
##	0.56	14:11	76.727	0.541	0.663
##			(74.977, 78.449)	(0.541, 0.541)	(0.642, 0.689)
##					
##	0.57	57:43	75.906	0.522	0.653

##			(74.091, 77.689)	(0.522, 0.522)	(0.633, 0.682)
##					
##	0.58	29:21	75.321	0.502	0.645
##			(73.076, 76.931)	(0.502, 0.502)	(0.623, 0.674)
##					
##	0.59	59:41	74.511	0.481	0.642
##			(72.201, 76.074)	(0.481, 0.481)	(0.612, 0.665)
##					
##	0.6	3:2	74.036	0.459	0.635
##			(71.221, 75.295)	(0.459, 0.459)	(0.604, 0.656)
##					
##	0.61	61:39	72.692	0.436	0.626
##			(70.128, 74.423)	(0.436, 0.436)	(0.594, 0.647)
##					
##	0.62	31:19	71.781	0.412	0.614
##			(69.225, 73.335)	(0.412, 0.412)	(0.586, 0.639)
##					
##	0.63	63:37	70.431	0.386	0.6
##			(68.14, 72.358)	(0.386, 0.386)	(0.575, 0.63)
##					
##	0.64	16:9	69.185	0.359	0.59
##			(67.033, 71.287)	(0.359, 0.359)	(0.565, 0.622)
##					
##	0.65	13:7	68.152	0.331	0.578
##			(65.975, 70.308)	(0.331, 0.331)	(0.556, 0.611)
##					
##	0.66	33:17	66.699	0.3	0.568
##			(64.84, 69.042)	(0.3, 0.3)	(0.545, 0.603)
##					
##	0.67	67:33	65.768	0.268	0.56
##			(63.667, 67.908)	(0.268, 0.268)	(0.532, 0.594)
##					
##	0.68	17:8	64.716	0.234	0.552
##			(62.536, 66.777)	(0.234, 0.234)	(0.523, 0.585)
##					
##	0.69	69:31	63.203	0.198	0.542
##			(61.279, 65.717)	(0.198, 0.198)	(0.513, 0.575)
##					
##	0.7	7:3	62.274	0.159	0.528
##			(59.936, 64.362)	(0.159, 0.159)	(0.499, 0.565)
##					
##	0.71	71:29	60.663	0.117	0.516
##			(58.728, 63.365)	(0.117, 0.117)	(0.489, 0.554)
##					
##	0.72	18:7	59.485	0.073	0.5
##			(57.365, 62.192)	(0.073, 0.073)	(0.479, 0.544)
##					
##	0.73	73:27	58.123	0.025	0.495
##			(56.244, 61.017)	(0.025, 0.025)	(0.466, 0.534)
##					
##	0.74	37:13	57.163	-0.026	0.487
##			(55.051, 59.714)	(-0.026, -0.026)	(0.454, 0.523)
##					
##	0.75	3:1	55.644	-0.081	0.473

##			(53.717, 58.555)	(-0.081, -0.081)	(0.441, 0.51)
##					
##	0.76	19:6	54.927	-0.141	0.46
##			(52.513, 57.232)	(-0.141, -0.141)	(0.431, 0.5)
##					
##	0.77	77:23	52.923	-0.207	0.444
##			(51.273, 55.864)	(-0.207, -0.207)	(0.416, 0.487)
##					
##	0.78	39:11	52.224	-0.278	0.433
##			(50.015, 54.535)	(-0.278, -0.278)	(0.402, 0.474)
##					
##	0.79	79:21	50.427	-0.356	0.418
##			(48.765, 53.416)	(-0.356, -0.356)	(0.39, 0.459)
##					
##	0.8	4:1	49.364	-0.442	0.403
##			(47.381, 52.163)	(-0.442, -0.442)	(0.375, 0.445)
##					
##	0.81	81:19	48.168	-0.537	0.39
##			(45.863, 51.316)	(-0.537, -0.537)	(0.36, 0.428)
##					
##	0.82	41:9	46.177	-0.642	0.367
##			(44.496, 49.775)	(-0.642, -0.642)	(0.342, 0.41)
##					
##	0.83	83:17	45.369	-0.76	0.353
##			(42.736, 48.468)	(-0.76, -0.76)	(0.321, 0.397)
##					
##	0.84	21:4	43.718	-0.892	0.337
##			(41.467, 47.461)	(-0.892, -0.892)	(0.303, 0.38)
##					
##	0.85	17:3	42.637	-1.043	0.317
##			(39.98, 46.308)	(-1.043, -1.043)	(0.282, 0.361)
##					
##	0.86	43:7	41.004	-1.214	0.293
##			(37.983, 44.448)	(-1.214, -1.214)	(0.263, 0.344)
##					
##	0.87	87:13	38.959	-1.412	0.269
##			(36.435, 43.102)	(-1.412, -1.412)	(0.243, 0.321)
##					
##	0.88	22:3	37.374	-1.643	0.256
##			(35.104, 41.541)	(-1.643, -1.643)	(0.221, 0.303)
##					
##	0.89	89:11	36.209	-1.917	0.236
##			(33.297, 39.368)	(-1.917, -1.917)	(0.199, 0.284)
##					
##	0.9	9:1	34.303	-2.244	0.216
##			(31.537, 37.862)	(-2.244, -2.244)	(0.177, 0.262)
##					
##	0.91	91:9	31.966	-2.645	0.19
##			(29.249, 35.933)	(-2.645, -2.645)	(0.15, 0.237)
##					
##	0.92	23:2	30.049	-3.145	0.163
##			(26.3, 33.275)	(-3.145, -3.145)	(0.118, 0.21)
##					
##	0.93	93:7	26.099	-3.789	0.12

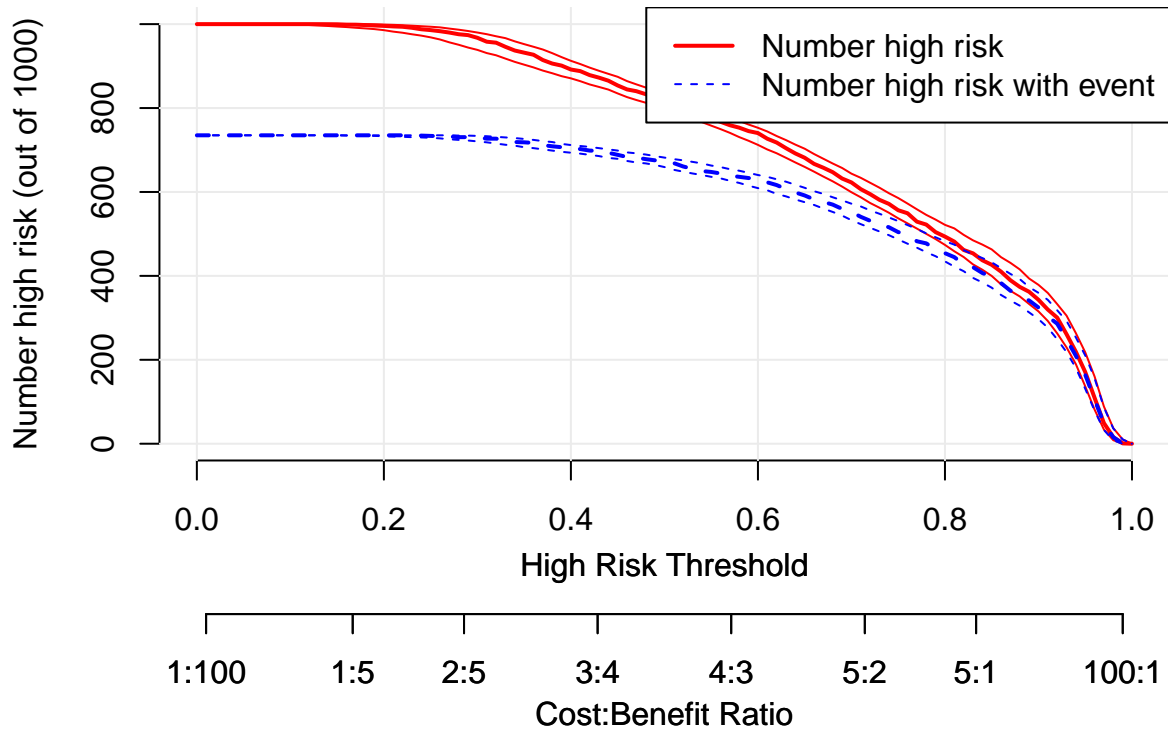
```
##          (22.962, 30.527)    (-3.789, -3.789)          (0.085, 0.178)
##
## 0.94      47:3          21.912          -4.647          0.082
##          (18.958, 26.352)    (-4.647, -4.647)          (0.05, 0.14)
##
## 0.95      19:1          17.083          -5.849          0.06
##          (13.673, 21.778)    (-5.849, -5.849)          (0.019, 0.104)
##
## 0.96      24:1          10.937          -7.651          0.018
##          (8.128, 16.008)    (-7.651, -7.651)          (-0.013, 0.074)
##
## 0.97      97:3          4.913          -10.655          0.017
##          (3.373, 8.856)    (-10.655, -10.655)          (-0.024, 0.054)
##
## 0.98      49:1          1.565          -16.663          -0.004
##          (0.837, 3.718)    (-16.663, -16.663)          (-0.021, 0.027)
##
## 0.99      99:1          0.079          -34.686          -0.007
##          (0, 1.025)    (-34.686, -34.686)          (-0.024, 0.01)
##
## 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$y <- as.factor(test_0_1$y)
str(test_0_1)
```

```
## tibble [5,348 x 20] (S3: tbl_df/tbl/data.frame)
## $ y      : Factor w/ 2 levels "0","1": 1 1 1 2 2 2 1 1 1 1 ...
## $ gender : Factor w/ 2 levels "1","2": 1 1 2 1 2 2 2 1 1 2 ...
## $ race   : Factor w/ 3 levels "1","2","3": 1 2 2 1 1 3 1 1 1 2 ...
## $ age    : Factor w/ 3 levels "1","2","3": 3 3 3 3 3 3 2 3 3 3 ...
## $ size   : Factor w/ 3 levels "1","2","3": 1 3 2 1 2 1 3 3 1 1 ...
## $ marry  : Factor w/ 3 levels "1","2","3": 3 2 1 3 2 2 2 1 2 2 ...
## $ 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 1 2 2 2 1 1 1 1 1 ...
## $ grade  : Factor w/ 5 levels "1","2","3","4",...: 2 2 2 5 2 2 2 5 2 3 ...
## $ kind   : Factor w/ 2 levels "1","2": 1 1 1 1 1 1 1 1 1 1 ...
## $ t      : Factor w/ 3 levels "1","2","3": 2 1 2 3 2 2 2 3 2 2 ...
## $ N      : Factor w/ 3 levels "1","2","3": 2 1 2 1 2 1 1 3 1 2 ...
## $ surgery: Factor w/ 2 levels "0","1": 2 1 2 1 2 1 2 1 2 2 ...
## $ RX     : Factor w/ 2 levels "0","1": 1 1 1 1 2 1 1 1 1 1 ...
## $ radiate: Factor w/ 2 levels "0","1": 1 1 1 1 1 2 1 1 1 1 ...
## $ chem   : Factor w/ 2 levels "0","1": 1 2 1 1 2 2 2 2 2 2 ...
## $ CEA    : Factor w/ 3 levels "0","1","2": 1 3 2 2 2 2 1 2 2 1 ...
## $ 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 2 1 1 1 1 1 1 ...
```


0+1 multivariate forest

```
fit.result<-summary(m1_final)
df1<-fit.result$coefficients
df2<-confint(m1_final)
```

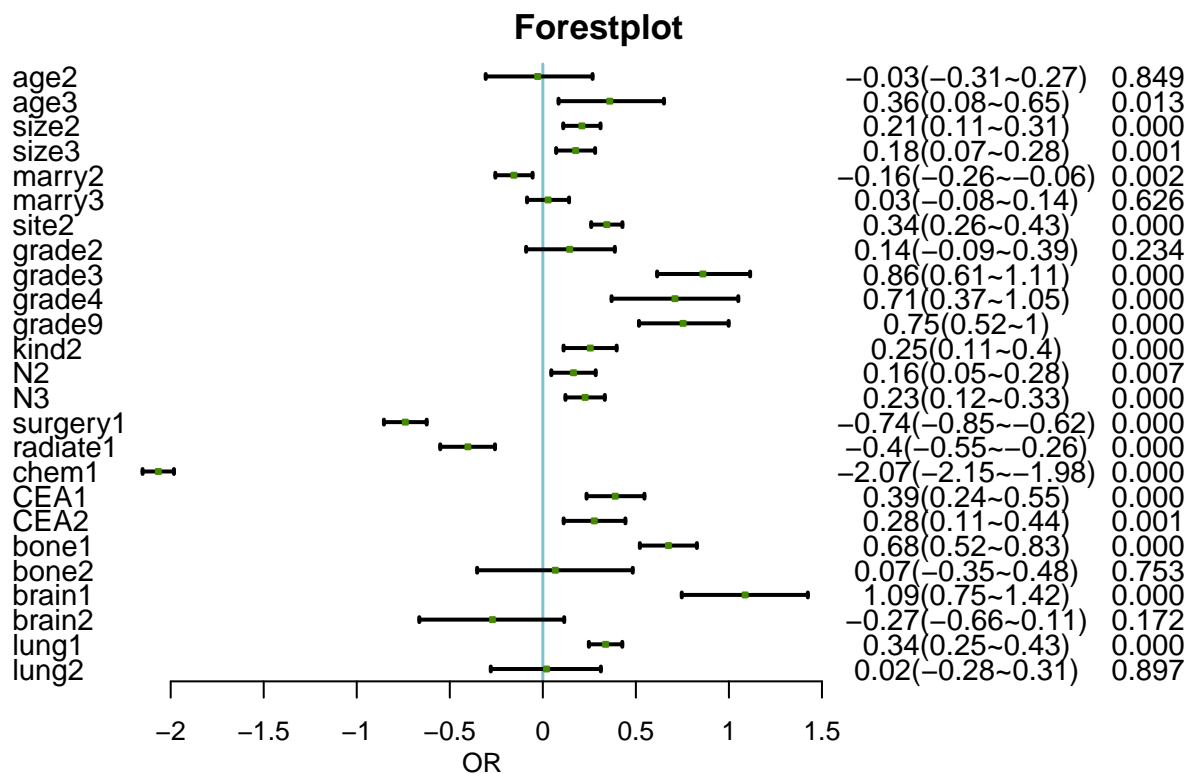
```
## Waiting for profiling to be done...
```

```
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_m1.csv",
          quote = F,row.names = F)

fp<-read.csv("forestplot_m1.csv",header=T)

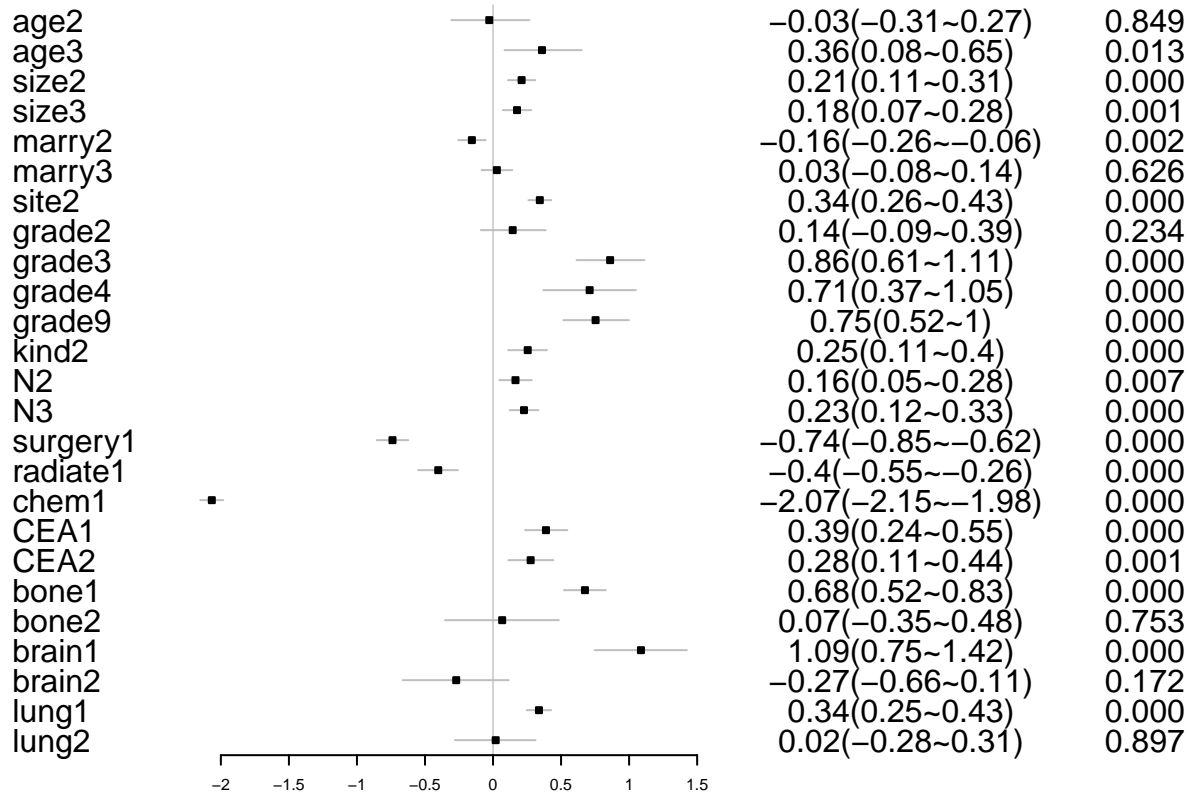
## plot 1

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 multivariate forward/backward/both AIC

```
fullmod <- glm(y ~ ., family=binomial, data = train_0_1)
summary(fullmod)
```

```
##
## Call:
## glm(formula = y ~ ., family = binomial, data = train_0_1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.0990  -0.5589  -0.3710  -0.2374   2.9458
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.13217    0.21752  -5.205 1.94e-07 ***
## gender2      -0.01054    0.04175  -0.252 0.800702
## race2         0.02422    0.05774   0.419 0.674888
## race3        -0.11628    0.07089  -1.640 0.100958
## age2          -0.03198    0.14643  -0.218 0.827118
## age3          0.35181    0.14451   2.434 0.014913 *
## size2         0.21639    0.05129   4.219 2.46e-05 ***
## size3         0.16808    0.05395   3.115 0.001837 **
## marry2        -0.14332    0.05174  -2.770 0.005608 **
## marry3         0.02727    0.05849   0.466 0.641115
## income2       -0.03092    0.05652  -0.547 0.584337
```

```
## income3      -0.09560    0.06394   -1.495  0.134878
## site2        0.33178    0.04325    7.671  1.70e-14 ***
## grade2       0.14760    0.12159    1.214  0.224801
## grade3       0.86660    0.12762    6.791  1.12e-11 ***
## grade4       0.70851    0.17388    4.075  4.61e-05 ***
## grade9       0.75913    0.12309    6.167  6.94e-10 ***
## kind2        0.25424    0.07278    3.493  0.000477 ***
## t2           -0.10205    0.07350   -1.389  0.164965
## t3           -0.03713    0.06925   -0.536  0.591842
## N2           0.17271    0.06135    2.815  0.004873 **
## N3           0.23125    0.05489    4.213  2.52e-05 ***
## surgery1     -0.67583    0.06649  -10.165  < 2e-16 ***
## RX1          -0.11327    0.06757   -1.676  0.093643 .
## radiate1     -0.39118    0.07534   -5.192  2.08e-07 ***
## chem1        -2.06344    0.04321  -47.751  < 2e-16 ***
## CEA1          0.38690    0.07932    4.878  1.07e-06 ***
## CEA2          0.26743    0.08473    3.156  0.001599 **
## bone1         0.67445    0.07825    8.619  < 2e-16 ***
## bone2         0.06867    0.21305    0.322  0.747208
## brain1        1.09543    0.17356    6.311  2.76e-10 ***
## brain2       -0.27908    0.19847   -1.406  0.159666
## lung1         0.33646    0.04589    7.332  2.26e-13 ***
## lung2         0.02545    0.15075    0.169  0.865912
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 20561  on 21390  degrees of freedom
## Residual deviance: 16014  on 21357  degrees of freedom
## AIC: 16082
##
## Number of Fisher Scoring iterations: 5
```

```
nothing <- glm(y ~ 1, family=binomial, data = train_0_1)
summary(nothing)
```

```
##
## Call:
## glm(formula = y ~ 1, family = binomial, data = train_0_1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.6419  -0.6419  -0.6419  -0.6419   1.8337
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.47520    0.01757  -83.98  <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: 20561  on 21390  degrees of freedom
```

```
## Residual deviance: 20561  on 21390  degrees of freedom
## AIC: 20563
##
## Number of Fisher Scoring iterations: 4
```

```
backwards_aic = step(fullmod)
```

```
## Start:  AIC=16081.63
## 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    16014 16080
## - t          2    16016 16080
## - income     2    16016 16080
## - race       2    16017 16081
## <none>       16014 16082
## - RX         1    16016 16082
## - kind       1    16026 16092
## - marry      2    16028 16092
## - size       2    16034 16098
## - N          2    16038 16102
## - CEA        2    16041 16105
## - radiate    1    16042 16108
## - brain      2    16055 16119
## - lung       2    16067 16131
## - site       1    16072 16138
## - age        2    16081 16145
## - bone       2    16085 16149
## - surgery    1    16119 16185
## - grade      4    16256 16316
## - chem       1    18452 18518
##
```

```
## Step:  AIC=16079.7
## y ~ race + age + size + marry + income + site + grade + kind +
##      t + N + surgery + RX + radiate + chem + CEA + bone + brain +
##      lung
```

```
##
##           Df Deviance   AIC
## - t          2    16016 16078
## - income     2    16016 16078
## - race       2    16017 16079
## <none>       16014 16080
## - RX         1    16016 16080
## - kind       1    16026 16090
## - marry      2    16029 16091
## - size       2    16034 16096
## - N          2    16038 16100
## - CEA        2    16041 16103
## - radiate    1    16042 16106
## - brain      2    16055 16117
## - lung       2    16067 16129
## - site       1    16073 16137
```

```

## - age      2      16081 16143
## - bone     2      16085 16147
## - surgery  1      16119 16183
## - grade    4      16256 16314
## - chem     1      18453 18517
##
## Step: AIC=16077.89
## y ~ race + age + size + marry + income + site + grade + kind +
##      N + surgery + RX + radiate + chem + CEA + bone + brain +
##      lung
##
##           Df Deviance   AIC
## - income   2      16019 16077
## - race     2      16019 16077
## <none>      16016 16078
## - RX       1      16019 16079
## - kind     1      16028 16088
## - marry    2      16031 16089
## - size     2      16036 16094
## - N        2      16040 16098
## - CEA      2      16044 16102
## - radiate  1      16046 16106
## - brain    2      16058 16116
## - lung     2      16070 16128
## - site     1      16076 16136
## - age      2      16084 16142
## - bone     2      16088 16146
## - surgery  1      16158 16218
## - grade    4      16258 16312
## - chem     1      18454 18514
##
## Step: AIC=16076.69
## y ~ race + age + size + marry + site + grade + kind + N + surgery +
##      RX + radiate + chem + CEA + bone + brain + lung
##
##           Df Deviance   AIC
## <none>      16019 16077
## - race     2      16023 16077
## - RX       1      16022 16078
## - kind     1      16030 16086
## - marry    2      16034 16088
## - size     2      16038 16092
## - N        2      16042 16096
## - CEA      2      16046 16100
## - radiate  1      16048 16104
## - brain    2      16060 16114
## - lung     2      16072 16126
## - site     1      16079 16135
## - age      2      16086 16140
## - bone     2      16090 16144
## - surgery  1      16161 16217
## - grade    4      16260 16310
## - chem     1      18467 18523

```

```
formula(backwards_aic)
```

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

```
summary(backwards_aic)
```

```
##  
## Call:  
## glm(formula = y ~ race + age + size + marry + site + grade +  
##      kind + N + surgery + RX + radiate + chem + CEA + bone + brain +  
##      lung, family = binomial, data = train_0_1)  
##  
## Deviance Residuals:  
##      Min       1Q   Median       3Q      Max  
## -2.1147  -0.5608  -0.3716  -0.2375   2.9285  
##  
## Coefficients:  
##              Estimate Std. Error z value Pr(>|z|)  
## (Intercept) -1.23032    0.20557  -5.985 2.16e-09 ***  
## race2        0.03531    0.05722   0.617 0.537232  
## race3       -0.13348    0.06991  -1.909 0.056215 .  
## age2        -0.03268    0.14637  -0.223 0.823307  
## age3         0.35016    0.14444   2.424 0.015340 *  
## size2        0.20992    0.05106   4.112 3.93e-05 ***  
## size3        0.17687    0.05342   3.311 0.000929 ***  
## marry2      -0.14288    0.05171  -2.763 0.005724 **  
## marry3       0.03118    0.05775   0.540 0.589258  
## site2       0.33484    0.04300   7.786 6.90e-15 ***  
## grade2      0.14552    0.12149   1.198 0.231001  
## grade3      0.86213    0.12751   6.761 1.37e-11 ***  
## grade4      0.70957    0.17378   4.083 4.44e-05 ***  
## grade9      0.75636    0.12296   6.151 7.68e-10 ***  
## kind2       0.25288    0.07272   3.477 0.000506 ***  
## N2          0.16324    0.06091   2.680 0.007363 **  
## N3          0.22782    0.05397   4.221 2.43e-05 ***  
## surgery1    -0.71313    0.06049 -11.789 < 2e-16 ***  
## RX1         -0.11303    0.06749  -1.675 0.093961 .  
## radiate1    -0.39869    0.07506  -5.312 1.08e-07 ***  
## chem1      -2.06540    0.04316 -47.854 < 2e-16 ***  
## CEA1        0.38764    0.07929   4.889 1.01e-06 ***  
## CEA2        0.27228    0.08467   3.216 0.001301 **  
## bone1       0.67451    0.07820   8.626 < 2e-16 ***  
## bone2       0.06703    0.21297   0.315 0.752968  
## brain1      1.10387    0.17337   6.367 1.92e-10 ***  
## brain2     -0.27154    0.19830  -1.369 0.170890  
## lung1       0.33500    0.04586   7.306 2.76e-13 ***  
## lung2       0.02373    0.15072   0.157 0.874914  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## (Dispersion parameter for binomial family taken to be 1)
```

```
##
## Null deviance: 20561 on 21390 degrees of freedom
## Residual deviance: 16019 on 21362 degrees of freedom
## AIC: 16077
##
## Number of Fisher Scoring iterations: 5
```

```
forwards_aic = step(nothing, scope=list(lower=formula(nothing),upper=formula(fullmod)), direction="forw
```

```
## Start: AIC=20562.88
## y ~ 1
##
##           Df Deviance  AIC
## + chem      1    17441 17445
## + surgery    1    19928 19932
## + grade      4    20014 20024
## + age        2    20080 20086
## + t          2    20199 20205
## + size       2    20273 20279
## + N          2    20282 20288
## + RX         1    20310 20314
## + lung       2    20379 20385
## + marry      2    20381 20387
## + bone       2    20395 20401
## + site       1    20417 20421
## + CEA        2    20460 20466
## + radiate    1    20464 20468
## + brain      2    20475 20481
## + kind       1    20519 20523
## + income     2    20546 20552
## + gender     1    20548 20552
## + race       2    20550 20556
## <none>      20561 20563
##
```

```
## Step: AIC=17445.22
## y ~ chem
##
##           Df Deviance  AIC
## + surgery    1    16836 16842
## + grade      4    16837 16849
## + t          2    17101 17109
## + N          2    17183 17191
## + size       2    17211 17219
## + lung       2    17263 17271
## + bone       2    17297 17305
## + RX         1    17329 17335
## + age        2    17333 17341
## + brain      2    17378 17386
## + CEA        2    17379 17387
## + site       1    17384 17390
## + marry      2    17411 17419
## + kind       1    17420 17426
## + radiate    1    17433 17439
## + race       2    17431 17439
```



```

## <none>          17441 17445
## + gender      1    17440 17446
## + income      2    17440 17448
##
## Step:  AIC=16842.16
## y ~ chem + surgery
##
##           Df Deviance   AIC
## + grade    4    16517 16531
## + site     1    16706 16714
## + age      2    16740 16750
## + bone     2    16749 16759
## + lung     2    16764 16774
## + N        2    16767 16777
## + brain    2    16789 16799
## + kind     1    16803 16811
## + size     2    16808 16818
## + CEA      2    16810 16820
## + radiate  1    16813 16821
## + marry    2    16814 16824
## + t        2    16825 16835
## + RX       1    16827 16835
## + race     2    16826 16836
## + gender   1    16832 16840
## <none>          16836 16842
## + income    2    16834 16844
##
## Step:  AIC=16530.59
## y ~ chem + surgery + grade
##
##           Df Deviance   AIC
## + site     1    16415 16431
## + age      2    16422 16440
## + lung     2    16435 16453
## + bone     2    16438 16456
## + brain    2    16472 16490
## + CEA      2    16483 16501
## + N        2    16486 16504
## + marry    2    16493 16511
## + radiate  1    16496 16512
## + size     2    16496 16514
## + kind     1    16502 16518
## + race     2    16505 16523
## + RX       1    16508 16524
## + t        2    16510 16528
## + gender   1    16513 16529
## <none>          16517 16531
## + income    2    16513 16531
##
## Step:  AIC=16431.08
## y ~ chem + surgery + grade + site
##
##           Df Deviance   AIC
## + lung     2    16326 16346

```

```

## + bone      2      16331 16351
## + age       2      16344 16364
## + brain     2      16371 16391
## + CEA       2      16379 16399
## + N         2      16388 16408
## + size      2      16393 16413
## + marry     2      16394 16414
## + kind      1      16405 16423
## + radiate   1      16407 16425
## + RX        1      16408 16426
## + race      2      16409 16429
## + t         2      16411 16431
## <none>      16415 16431
## + income    2      16412 16432
## + gender    1      16414 16432
##
## Step: AIC=16345.71
## y ~ chem + surgery + grade + site + lung
##
##           Df Deviance   AIC
## + age      2      16258 16282
## + bone     2      16260 16284
## + brain    2      16293 16317
## + CEA      2      16297 16321
## + N        2      16301 16325
## + size     2      16304 16328
## + marry    2      16305 16329
## + kind     1      16314 16336
## + radiate  1      16316 16338
## + RX       1      16321 16343
## + race     2      16320 16344
## <none>     16326 16346
## + t        2      16322 16346
## + income   2      16322 16346
## + gender   1      16325 16347
##
## Step: AIC=16281.54
## y ~ chem + surgery + grade + site + lung + age
##
##           Df Deviance   AIC
## + bone     2      16190 16218
## + brain    2      16222 16250
## + CEA      2      16228 16256
## + N        2      16232 16260
## + size     2      16235 16263
## + marry    2      16239 16267
## + kind     1      16246 16272
## + radiate  1      16249 16275
## + race     2      16251 16279
## + RX       1      16254 16280
## + income   2      16253 16281
## <none>     16258 16282
## + t        2      16254 16282
## + gender   1      16257 16283

```

```

##
## Step: AIC=16217.5
## y ~ chem + surgery + grade + site + lung + age + bone
##
##           Df Deviance   AIC
## + CEA      2    16161 16193
## + brain     2    16161 16193
## + N         2    16165 16197
## + size      2    16166 16198
## + radiate   1    16169 16199
## + marry     2    16170 16202
## + kind      1    16178 16208
## + race      2    16183 16215
## + RX        1    16186 16216
## + income    2    16185 16217
## <none>      16190 16218
## + t         2    16186 16218
## + gender    1    16189 16219
##
## Step: AIC=16192.87
## y ~ chem + surgery + grade + site + lung + age + bone + CEA
##
##           Df Deviance   AIC
## + brain     2    16132 16168
## + N         2    16137 16173
## + size      2    16138 16174
## + radiate   1    16142 16176
## + marry     2    16143 16179
## + kind      1    16148 16182
## + race      2    16154 16190
## + RX        1    16158 16192
## + income    2    16156 16192
## <none>      16161 16193
## + t         2    16157 16193
## + gender    1    16160 16194
##
## Step: AIC=16168.47
## y ~ chem + surgery + grade + site + lung + age + bone + CEA +
##      brain
##
##           Df Deviance   AIC
## + radiate   1    16103 16141
## + N         2    16108 16148
## + size      2    16108 16148
## + marry     2    16114 16154
## + kind      1    16119 16157
## + race      2    16126 16166
## + RX        1    16128 16166
## + income    2    16128 16168
## <none>      16132 16168
## + t         2    16129 16169
## + gender    1    16132 16170
##
## Step: AIC=16140.58

```

```

## y ~ chem + surgery + grade + site + lung + age + bone + CEA +
##      brain + radiate
##
##           Df Deviance   AIC
## + N       2    16078 16120
## + size    2    16079 16121
## + marry   2    16084 16126
## + kind    1    16089 16129
## + race    2    16096 16138
## + RX      1    16099 16139
## + income  2    16098 16140
## <none>          16103 16141
## + gender  1    16102 16142
## + t       2    16101 16143
##
## Step:  AIC=16120.29
## y ~ chem + surgery + grade + site + lung + age + bone + CEA +
##      brain + radiate + N
##
##           Df Deviance   AIC
## + size    2    16056 16102
## + marry   2    16059 16105
## + kind    1    16064 16108
## + race    2    16071 16117
## + RX      1    16075 16119
## + income  2    16073 16119
## <none>          16078 16120
## + gender  1    16078 16122
## + t       2    16076 16122
##
## Step:  AIC=16102.29
## y ~ chem + surgery + grade + site + lung + age + bone + CEA +
##      brain + radiate + N + size
##
##           Df Deviance   AIC
## + marry   2    16038 16088
## + kind    1    16044 16092
## + race    2    16049 16099
## + RX      1    16053 16101
## + income  2    16052 16102
## <none>          16056 16102
## + gender  1    16056 16104
## + t       2    16054 16104
##
## Step:  AIC=16088
## y ~ chem + surgery + grade + site + lung + age + bone + CEA +
##      brain + radiate + N + size + marry
##
##           Df Deviance   AIC
## + kind    1    16026 16078
## + RX      1    16035 16087
## + race    2    16033 16087
## + income  2    16034 16088
## <none>          16038 16088

```

```

## + t      2      16036 16090
## + gender 1      16038 16090
##
## Step: AIC=16077.98
## y ~ chem + surgery + grade + site + lung + age + bone + CEA +
##      brain + radiate + N + size + marry + kind
##
##           Df Deviance   AIC
## + RX      1      16023 16077
## + race    2      16022 16078
## + income  2      16022 16078
## <none>           16026 16078
## + t      2      16024 16080
## + gender  1      16026 16080
##
## Step: AIC=16077.15
## y ~ chem + surgery + grade + site + lung + age + bone + CEA +
##      brain + radiate + N + size + marry + kind + RX
##
##           Df Deviance   AIC
## + race    2      16019 16077
## + income  2      16019 16077
## <none>           16023 16077
## + t      2      16021 16079
## + gender  1      16023 16079
##
## Step: AIC=16076.69
## y ~ chem + surgery + grade + site + lung + age + bone + CEA +
##      brain + radiate + N + size + marry + kind + RX + race
##
##           Df Deviance   AIC
## <none>           16019 16077
## + income  2      16016 16078
## + t      2      16016 16078
## + gender  1      16019 16079

```

```
formula(forwards_aic)
```

```

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

```

```
summary(forwards_aic)
```

```

##
## Call:
## glm(formula = y ~ chem + surgery + grade + site + lung + age +
##      bone + CEA + brain + radiate + N + size + marry + kind +
##      RX + race, family = binomial, data = train_0_1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.1147  -0.5608  -0.3716  -0.2375   2.9285
##

```

```
## Coefficients:
##           Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.23032    0.20557  -5.985 2.16e-09 ***
## chem1       -2.06540    0.04316 -47.854 < 2e-16 ***
## surgery1    -0.71313    0.06049 -11.789 < 2e-16 ***
## grade2       0.14552    0.12149   1.198 0.231001
## grade3       0.86213    0.12751   6.761 1.37e-11 ***
## grade4       0.70957    0.17378   4.083 4.44e-05 ***
## grade9       0.75636    0.12296   6.151 7.68e-10 ***
## site2        0.33484    0.04300   7.786 6.90e-15 ***
## lung1        0.33500    0.04586   7.306 2.76e-13 ***
## lung2        0.02373    0.15072   0.157 0.874914
## age2        -0.03268    0.14637  -0.223 0.823307
## age3         0.35016    0.14444   2.424 0.015340 *
## bone1        0.67451    0.07820   8.626 < 2e-16 ***
## bone2        0.06703    0.21297   0.315 0.752968
## CEA1         0.38764    0.07929   4.889 1.01e-06 ***
## CEA2         0.27228    0.08467   3.216 0.001301 **
## brain1       1.10387    0.17337   6.367 1.92e-10 ***
## brain2      -0.27154    0.19830  -1.369 0.170890
## radiate1     -0.39869    0.07506  -5.312 1.08e-07 ***
## N2           0.16324    0.06091   2.680 0.007363 **
## N3           0.22782    0.05397   4.221 2.43e-05 ***
## size2        0.20992    0.05106   4.112 3.93e-05 ***
## size3        0.17687    0.05342   3.311 0.000929 ***
## marry2      -0.14288    0.05171  -2.763 0.005724 **
## marry3       0.03118    0.05775   0.540 0.589258
## kind2        0.25288    0.07272   3.477 0.000506 ***
## RX1         -0.11303    0.06749  -1.675 0.093961 .
## race2        0.03531    0.05722   0.617 0.537232
## race3       -0.13348    0.06991  -1.909 0.056215 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##    Null deviance: 20561  on 21390  degrees of freedom
## Residual deviance: 16019  on 21362  degrees of freedom
## AIC: 16077
##
## Number of Fisher Scoring iterations: 5
```

```
bothways_aic = step(nothing, list(lower=formula(nothing),upper=formula(fullmod)), direction="both")
```

```
## Start:  AIC=20562.88
## y ~ 1
##
##           Df Deviance   AIC
## + chem      1    17441 17445
## + surgery    1    19928 19932
## + grade      4    20014 20024
## + age        2    20080 20086
## + t          2    20199 20205
## + size       2    20273 20279
```

```

## + N      2      20282 20288
## + RX      1      20310 20314
## + lung    2      20379 20385
## + marry   2      20381 20387
## + bone    2      20395 20401
## + site    1      20417 20421
## + CEA     2      20460 20466
## + radiate 1      20464 20468
## + brain   2      20475 20481
## + kind    1      20519 20523
## + income  2      20546 20552
## + gender  1      20548 20552
## + race    2      20550 20556
## <none>    20561 20563
##
## Step:  AIC=17445.22
## y ~ chem
##
##           Df Deviance   AIC
## + surgery  1     16836 16842
## + grade    4     16837 16849
## + t        2     17101 17109
## + N        2     17183 17191
## + size     2     17211 17219
## + lung     2     17263 17271
## + bone     2     17297 17305
## + RX       1     17329 17335
## + age      2     17333 17341
## + brain    2     17378 17386
## + CEA      2     17379 17387
## + site     1     17384 17390
## + marry    2     17411 17419
## + kind     1     17420 17426
## + radiate  1     17433 17439
## + race     2     17431 17439
## <none>     17441 17445
## + gender   1     17440 17446
## + income   2     17440 17448
## - chem     1     20561 20563
##
## Step:  AIC=16842.16
## y ~ chem + surgery
##
##           Df Deviance   AIC
## + grade    4     16517 16531
## + site     1     16706 16714
## + age      2     16740 16750
## + bone     2     16749 16759
## + lung     2     16764 16774
## + N        2     16767 16777
## + brain    2     16789 16799
## + kind     1     16803 16811
## + size     2     16808 16818
## + CEA      2     16810 16820

```

```

## + radiate 1 16813 16821
## + marry 2 16814 16824
## + t 2 16825 16835
## + RX 1 16827 16835
## + race 2 16826 16836
## + gender 1 16832 16840
## <none> 16836 16842
## + income 2 16834 16844
## - surgery 1 17441 17445
## - chem 1 19928 19932
##
## Step: AIC=16530.59
## y ~ chem + surgery + grade
##
##           Df Deviance   AIC
## + site 1 16415 16431
## + age 2 16422 16440
## + lung 2 16435 16453
## + bone 2 16438 16456
## + brain 2 16472 16490
## + CEA 2 16483 16501
## + N 2 16486 16504
## + marry 2 16493 16511
## + radiate 1 16496 16512
## + size 2 16496 16514
## + kind 1 16502 16518
## + race 2 16505 16523
## + RX 1 16508 16524
## + t 2 16510 16528
## + gender 1 16513 16529
## <none> 16517 16531
## + income 2 16513 16531
## - grade 4 16836 16842
## - surgery 1 16837 16849
## - chem 1 19626 19638
##
## Step: AIC=16431.08
## y ~ chem + surgery + grade + site
##
##           Df Deviance   AIC
## + lung 2 16326 16346
## + bone 2 16331 16351
## + age 2 16344 16364
## + brain 2 16371 16391
## + CEA 2 16379 16399
## + N 2 16388 16408
## + size 2 16393 16413
## + marry 2 16394 16414
## + kind 1 16405 16423
## + radiate 1 16407 16425
## + RX 1 16408 16426
## + race 2 16409 16429
## + t 2 16411 16431
## <none> 16415 16431

```



```

## + income 2 16412 16432
## + gender 1 16414 16432
## - site 1 16517 16531
## - grade 4 16706 16714
## - surgery 1 16776 16790
## - chem 1 19421 19435
##
## Step: AIC=16345.71
## y ~ chem + surgery + grade + site + lung
##
##           Df Deviance   AIC
## + age      2    16258 16282
## + bone      2    16260 16284
## + brain     2    16293 16317
## + CEA       2    16297 16321
## + N         2    16301 16325
## + size      2    16304 16328
## + marry     2    16305 16329
## + kind      1    16314 16336
## + radiate   1    16316 16338
## + RX        1    16321 16343
## + race      2    16320 16344
## <none>      16326 16346
## + t         2    16322 16346
## + income    2    16322 16346
## + gender    1    16325 16347
## - lung      2    16415 16431
## - site      1    16435 16453
## - grade     4    16624 16636
## - surgery   1    16619 16637
## - chem      1    19340 19358
##
## Step: AIC=16281.54
## y ~ chem + surgery + grade + site + lung + age
##
##           Df Deviance   AIC
## + bone      2    16190 16218
## + brain     2    16222 16250
## + CEA       2    16228 16256
## + N         2    16232 16260
## + size      2    16235 16263
## + marry     2    16239 16267
## + kind      1    16246 16272
## + radiate   1    16249 16275
## + race      2    16251 16279
## + RX        1    16254 16280
## + income    2    16253 16281
## <none>      16258 16282
## + t         2    16254 16282
## + gender    1    16257 16283
## - age       2    16326 16346
## - lung      2    16344 16364
## - site      1    16342 16364
## - surgery   1    16543 16565

```

```

## - grade      4      16557 16573
## - chem       1      18983 19005
##
## Step:  AIC=16217.5
## y ~ chem + surgery + grade + site + lung + age + bone
##
##           Df Deviance   AIC
## + CEA      2      16161 16193
## + brain     2      16161 16193
## + N         2      16165 16197
## + size      2      16166 16198
## + radiate   1      16169 16199
## + marry     2      16170 16202
## + kind      1      16178 16208
## + race      2      16183 16215
## + RX        1      16186 16216
## + income    2      16185 16217
## <none>      16190 16218
## + t         2      16186 16218
## + gender    1      16189 16219
## - bone      2      16258 16282
## - lung      2      16258 16282
## - age       2      16260 16284
## - site      1      16277 16303
## - surgery   1      16455 16481
## - grade     4      16480 16500
## - chem      1      18908 18934
##
## Step:  AIC=16192.87
## y ~ chem + surgery + grade + site + lung + age + bone + CEA
##
##           Df Deviance   AIC
## + brain     2      16132 16168
## + N         2      16137 16173
## + size      2      16138 16174
## + radiate   1      16142 16176
## + marry     2      16143 16179
## + kind      1      16148 16182
## + race      2      16154 16190
## + RX        1      16158 16192
## + income    2      16156 16192
## <none>      16161 16193
## + t         2      16157 16193
## + gender    1      16160 16194
## - CEA       2      16190 16218
## - lung      2      16224 16252
## - bone      2      16228 16256
## - age       2      16231 16259
## - site      1      16250 16280
## - surgery   1      16413 16443
## - grade     4      16457 16481
## - chem      1      18843 18873
##
## Step:  AIC=16168.47

```

```

## y ~ chem + surgery + grade + site + lung + age + bone + CEA +
##   brain
##
##           Df Deviance   AIC
## + radiate  1    16103 16141
## + N        2    16108 16148
## + size     2    16108 16148
## + marry    2    16114 16154
## + kind     1    16119 16157
## + race     2    16126 16166
## + RX       1    16128 16166
## + income   2    16128 16168
## <none>           16132 16168
## + t        2    16129 16169
## + gender   1    16132 16170
## - brain    2    16161 16193
## - CEA      2    16161 16193
## - lung     2    16188 16220
## - bone     2    16193 16225
## - age      2    16205 16237
## - site     1    16221 16255
## - surgery  1    16383 16417
## - grade    4    16426 16454
## - chem     1    18811 18845
##
## Step:  AIC=16140.58
## y ~ chem + surgery + grade + site + lung + age + bone + CEA +
##   brain + radiate
##
##           Df Deviance   AIC
## + N        2    16078 16120
## + size     2    16079 16121
## + marry    2    16084 16126
## + kind     1    16089 16129
## + race     2    16096 16138
## + RX       1    16099 16139
## + income   2    16098 16140
## <none>           16103 16141
## + gender   1    16102 16142
## + t        2    16101 16143
## - CEA      2    16130 16164
## - radiate  1    16132 16168
## - brain    2    16142 16176
## - lung     2    16158 16192
## - age      2    16173 16207
## - site     1    16174 16210
## - bone     2    16176 16210
## - surgery  1    16356 16392
## - grade    4    16394 16424
## - chem     1    18699 18735
##
## Step:  AIC=16120.29
## y ~ chem + surgery + grade + site + lung + age + bone + CEA +
##   brain + radiate + N

```

```

##
##           Df Deviance   AIC
## + size      2    16056 16102
## + marry      2    16059 16105
## + kind       1    16064 16108
## + race       2    16071 16117
## + RX         1    16075 16119
## + income     2    16073 16119
## <none>       1    16078 16120
## + gender     1    16078 16122
## + t          2    16076 16122
## - N          2    16103 16141
## - CEA        2    16105 16143
## - radiate    1    16108 16148
## - brain      2    16118 16156
## - lung       2    16132 16170
## - site       1    16146 16186
## - age        2    16150 16188
## - bone       2    16150 16188
## - surgery    1    16282 16322
## - grade      4    16336 16370
## - chem       1    18674 18714
##
## Step: AIC=16102.29
## y ~ chem + surgery + grade + site + lung + age + bone + CEA +
##      brain + radiate + N + size
##
##           Df Deviance   AIC
## + marry      2    16038 16088
## + kind       1    16044 16092
## + race       2    16049 16099
## + RX         1    16053 16101
## + income     2    16052 16102
## <none>       1    16056 16102
## + gender     1    16056 16104
## + t          2    16054 16104
## - size       2    16078 16120
## - N          2    16079 16121
## - CEA        2    16083 16125
## - radiate    1    16086 16130
## - brain      2    16097 16139
## - lung       2    16109 16151
## - site       1    16125 16169
## - bone       2    16128 16170
## - age        2    16130 16172
## - surgery    1    16220 16264
## - grade      4    16308 16346
## - chem       1    18630 18674
##
## Step: AIC=16088
## y ~ chem + surgery + grade + site + lung + age + bone + CEA +
##      brain + radiate + N + size + marry
##
##           Df Deviance   AIC

```

```

## + kind      1      16026 16078
## + RX        1      16035 16087
## + race      2      16033 16087
## + income    2      16034 16088
## <none>      16038 16088
## + t         2      16036 16090
## + gender    1      16038 16090
## - marry     2      16056 16102
## - size      2      16059 16105
## - N         2      16061 16107
## - CEA       2      16064 16110
## - radiate   1      16068 16116
## - brain     2      16079 16125
## - lung      2      16091 16137
## - age       2      16108 16154
## - site      1      16106 16154
## - bone      2      16111 16157
## - surgery   1      16199 16247
## - grade     4      16292 16334
## - chem      1      18509 18557
##
## Step: AIC=16077.98
## y ~ chem + surgery + grade + site + lung + age + bone + CEA +
##      brain + radiate + N + size + marry + kind
##
##           Df Deviance   AIC
## + RX      1      16023 16077
## + race     2      16022 16078
## + income   2      16022 16078
## <none>     16026 16078
## + t       2      16024 16080
## + gender   1      16026 16080
## - kind     1      16038 16088
## - marry    2      16044 16092
## - size     2      16046 16094
## - N        2      16050 16098
## - CEA      2      16054 16102
## - radiate  1      16056 16106
## - brain    2      16067 16115
## - lung     2      16080 16128
## - site     1      16090 16140
## - age      2      16096 16144
## - bone     2      16098 16146
## - surgery  1      16189 16239
## - grade    4      16266 16310
## - chem     1      18489 18539
##
## Step: AIC=16077.15
## y ~ chem + surgery + grade + site + lung + age + bone + CEA +
##      brain + radiate + N + size + marry + kind + RX
##
##           Df Deviance   AIC
## + race     2      16019 16077
## + income   2      16019 16077

```

```

## <none>          16023 16077
## - RX            1    16026 16078
## + t             2    16021 16079
## + gender        1    16023 16079
## - kind          1    16035 16087
## - marry         2    16041 16091
## - size          2    16043 16093
## - N             2    16046 16096
## - CEA           2    16050 16100
## - radiate       1    16053 16105
## - brain         2    16065 16115
## - lung          2    16076 16126
## - site          1    16087 16139
## - age           2    16091 16141
## - bone          2    16095 16145
## - surgery       1    16166 16218
## - grade         4    16264 16310
## - chem          1    18468 18520
##
## Step:  AIC=16076.69
## y ~ chem + surgery + grade + site + lung + age + bone + CEA +
##      brain + radiate + N + size + marry + kind + RX + race
##
##           Df Deviance   AIC
## <none>          16019 16077
## - race          2    16023 16077
## - RX            1    16022 16078
## + income        2    16016 16078
## + t             2    16016 16078
## + gender        1    16019 16079
## - kind          1    16030 16086
## - marry         2    16034 16088
## - size          2    16038 16092
## - N             2    16042 16096
## - CEA           2    16046 16100
## - radiate       1    16048 16104
## - brain         2    16060 16114
## - lung          2    16072 16126
## - site          1    16079 16135
## - age           2    16086 16140
## - bone          2    16090 16144
## - surgery       1    16161 16217
## - grade         4    16260 16310
## - chem          1    18467 18523

```

```
formula(bothways_aic)
```

```

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

```

```
summary(bothways_aic)
```

```
##
```

```

## Call:
## glm(formula = y ~ chem + surgery + grade + site + lung + age +
##      bone + CEA + brain + radiate + N + size + marry + kind +
##      RX + race, family = binomial, data = train_0_1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.1147  -0.5608  -0.3716  -0.2375   2.9285
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.23032    0.20557  -5.985 2.16e-09 ***
## chem1       -2.06540    0.04316 -47.854 < 2e-16 ***
## surgery1    -0.71313    0.06049 -11.789 < 2e-16 ***
## grade2       0.14552    0.12149   1.198 0.231001
## grade3       0.86213    0.12751   6.761 1.37e-11 ***
## grade4       0.70957    0.17378   4.083 4.44e-05 ***
## grade9       0.75636    0.12296   6.151 7.68e-10 ***
## site2        0.33484    0.04300   7.786 6.90e-15 ***
## lung1        0.33500    0.04586   7.306 2.76e-13 ***
## lung2        0.02373    0.15072   0.157 0.874914
## age2        -0.03268    0.14637  -0.223 0.823307
## age3         0.35016    0.14444   2.424 0.015340 *
## bone1        0.67451    0.07820   8.626 < 2e-16 ***
## bone2        0.06703    0.21297   0.315 0.752968
## CEA1         0.38764    0.07929   4.889 1.01e-06 ***
## CEA2         0.27228    0.08467   3.216 0.001301 **
## brain1       1.10387    0.17337   6.367 1.92e-10 ***
## brain2      -0.27154    0.19830  -1.369 0.170890
## radiate1     -0.39869    0.07506  -5.312 1.08e-07 ***
## N2           0.16324    0.06091   2.680 0.007363 **
## N3           0.22782    0.05397   4.221 2.43e-05 ***
## size2        0.20992    0.05106   4.112 3.93e-05 ***
## size3        0.17687    0.05342   3.311 0.000929 ***
## marry2       -0.14288    0.05171  -2.763 0.005724 **
## marry3        0.03118    0.05775   0.540 0.589258
## kind2        0.25288    0.07272   3.477 0.000506 ***
## RX1         -0.11303    0.06749  -1.675 0.093961 .
## race2        0.03531    0.05722   0.617 0.537232
## race3       -0.13348    0.06991  -1.909 0.056215 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 20561  on 21390  degrees of freedom
## Residual deviance: 16019  on 21362  degrees of freedom
## AIC: 16077
##
## Number of Fisher Scoring iterations: 5

```

0+1 multivariate forward/backward/both BIC

```
backwards_bic = step(fullmod, k=log(nrow(train_0_1)))
```

```
## Start: AIC=16352.64
## y ~ gender + race + age + size + marry + income + site + grade +
##      kind + t + N + surgery + RX + radiate + chem + CEA + bone +
##      brain + lung
##
##           Df Deviance   AIC
## - t         2    16016 16335
## - income    2    16016 16335
## - race      2    16017 16336
## - gender    1    16014 16343
## - RX        1    16016 16346
## - marry     2    16028 16347
## <none>      16014 16353
## - size      2    16034 16353
## - kind      1    16026 16355
## - N         2    16038 16357
## - CEA       2    16041 16360
## - radiate   1    16042 16371
## - brain     2    16055 16374
## - lung      2    16067 16386
## - age       2    16081 16400
## - site      1    16072 16401
## - bone      2    16085 16404
## - surgery   1    16119 16448
## - grade     4    16256 16555
## - chem      1    18452 18780
##
## Step: AIC=16334.89
## y ~ gender + race + age + size + marry + income + site + grade +
##      kind + N + surgery + RX + radiate + chem + CEA + bone + brain +
##      lung
##
##           Df Deviance   AIC
## - income    2    16019 16318
## - race      2    16019 16318
## - gender    1    16016 16325
## - RX        1    16019 16328
## - marry     2    16031 16330
## - size      2    16036 16335
## <none>      16016 16335
## - kind      1    16028 16337
## - N         2    16040 16339
## - CEA       2    16043 16342
## - radiate   1    16046 16355
## - brain     2    16058 16357
## - lung      2    16069 16368
## - age       2    16084 16383
## - site      1    16076 16385
## - bone      2    16088 16387
```



```

## - surgery 1 16158 16467
## - grade 4 16258 16537
## - chem 1 18453 18762
##
## Step: AIC=16317.77
## y ~ gender + race + age + size + marry + site + grade + kind +
## N + surgery + RX + radiate + chem + CEA + bone + brain +
## lung
##
## Df Deviance AIC
## - race 2 16023 16302
## - gender 1 16019 16308
## - RX 1 16022 16311
## - marry 2 16034 16313
## - size 2 16038 16318
## <none> 16019 16318
## - kind 1 16030 16320
## - N 2 16042 16321
## - CEA 2 16046 16325
## - radiate 1 16048 16338
## - brain 2 16060 16340
## - lung 2 16072 16351
## - age 2 16086 16365
## - site 1 16078 16367
## - bone 2 16090 16370
## - surgery 1 16161 16450
## - grade 4 16260 16519
## - chem 1 18466 18755
##
## Step: AIC=16302.3
## y ~ gender + age + size + marry + site + grade + kind + N + surgery +
## RX + radiate + chem + CEA + bone + brain + lung
##
## Df Deviance AIC
## - gender 1 16023 16292
## - RX 1 16026 16295
## - marry 2 16040 16299
## - size 2 16043 16302
## <none> 16023 16302
## - kind 1 16035 16304
## - N 2 16046 16306
## - CEA 2 16050 16310
## - radiate 1 16053 16322
## - brain 2 16065 16324
## - lung 2 16076 16335
## - age 2 16091 16350
## - bone 2 16095 16354
## - site 1 16086 16356
## - surgery 1 16166 16435
## - grade 4 16264 16503
## - chem 1 18467 18736
##
## Step: AIC=16292.36
## y ~ age + size + marry + site + grade + kind + N + surgery +

```

```

##      RX + radiate + chem + CEA + bone + brain + lung
##
##           Df Deviance   AIC
## - RX      1    16026 16285
## - marry   2    16041 16290
## - size    2    16043 16292
## <none>           16023 16292
## - kind    1    16035 16294
## - N       2    16046 16296
## - CEA     2    16050 16300
## - radiate 1    16053 16312
## - brain   2    16065 16314
## - lung    2    16076 16326
## - age     2    16091 16340
## - bone    2    16095 16344
## - site    1    16087 16346
## - surgery 1    16166 16425
## - grade   4    16264 16493
## - chem    1    18468 18728
##
## Step: AIC=16285.22
## y ~ age + size + marry + site + grade + kind + N + surgery +
##      radiate + chem + CEA + bone + brain + lung
##
##           Df Deviance   AIC
## - marry   2    16044 16283
## - size    2    16046 16285
## <none>           16026 16285
## - kind    1    16038 16287
## - N       2    16050 16289
## - CEA     2    16054 16293
## - radiate 1    16056 16306
## - brain   2    16067 16306
## - lung    2    16080 16319
## - age     2    16096 16335
## - bone    2    16098 16337
## - site    1    16090 16340
## - surgery 1    16189 16438
## - grade   4    16266 16486
## - chem    1    18489 18738
##
## Step: AIC=16283.08
## y ~ age + size + site + grade + kind + N + surgery + radiate +
##      chem + CEA + bone + brain + lung
##
##           Df Deviance   AIC
## <none>           16044 16283
## - size    2    16064 16284
## - kind    1    16056 16286
## - N       2    16066 16286
## - CEA     2    16072 16292
## - radiate 1    16074 16303
## - brain   2    16085 16304
## - lung    2    16098 16318

```

```
## - bone      2      16114 16334
## - age       2      16117 16336
## - site      1      16109 16338
## - surgery   1      16209 16439
## - grade     4      16283 16482
## - chem      1      18608 18837
```

```
formula(backwards_bic)
```

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

```
summary(backwards_bic)
```

```
##
## Call:
## glm(formula = y ~ age + size + site + grade + kind + N + surgery +
##      radiate + chem + CEA + bone + brain + lung, family = binomial,
##      data = train_0_1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.1550  -0.5632  -0.3694  -0.2419   2.9262
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.29341    0.20391  -6.343 2.25e-10 ***
## age2         -0.04462    0.14585  -0.306 0.759677
## age3          0.34689    0.14319   2.423 0.015411 *
## size2         0.21485    0.05098   4.214 2.51e-05 ***
## size3         0.18055    0.05334   3.385 0.000712 ***
## site2         0.34501    0.04271   8.079 6.55e-16 ***
## grade2        0.14556    0.12142   1.199 0.230609
## grade3        0.85588    0.12739   6.719 1.84e-11 ***
## grade4        0.71620    0.17344   4.129 3.64e-05 ***
## grade9        0.75358    0.12287   6.133 8.61e-10 ***
## kind2         0.25968    0.07260   3.577 0.000347 ***
## N2            0.16150    0.06078   2.657 0.007878 **
## N3            0.22336    0.05392   4.142 3.44e-05 ***
## surgery1     -0.74395    0.05858 -12.699 < 2e-16 ***
## radiate1     -0.40184    0.07499  -5.359 8.38e-08 ***
## chem1        -2.09057    0.04270 -48.956 < 2e-16 ***
## CEA1          0.39548    0.07922   4.992 5.97e-07 ***
## CEA2          0.28052    0.08459   3.316 0.000913 ***
## bone1         0.66965    0.07813   8.571 < 2e-16 ***
## bone2         0.07187    0.21330   0.337 0.736161
## brain1        1.08575    0.17275   6.285 3.27e-10 ***
## brain2       -0.27125    0.19858  -1.366 0.171963
## lung1         0.33792    0.04580   7.378 1.61e-13 ***
## lung2         0.01143    0.15088   0.076 0.939606
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```

## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 20561  on 21390  degrees of freedom
## Residual deviance: 16044  on 21367  degrees of freedom
## AIC: 16092
##
## Number of Fisher Scoring iterations: 5

forwards_bic = step(nothing, scope=list(lower=formula(nothing),upper=formula(fullmod)), direction="forward")

## Start:  AIC=20570.85
## y ~ 1
##
##           Df Deviance   AIC
## + chem      1    17441 17461
## + surgery   1    19928 19948
## + grade     4    20014 20064
## + age       2    20080 20110
## + t         2    20199 20229
## + size      2    20273 20303
## + N         2    20282 20312
## + RX        1    20310 20330
## + lung      2    20379 20409
## + marry     2    20381 20411
## + bone      2    20395 20425
## + site      1    20417 20437
## + radiate   1    20464 20484
## + CEA       2    20460 20490
## + brain     2    20475 20504
## + kind      1    20519 20539
## + gender    1    20548 20568
## <none>      20561 20571
## + income    2    20546 20576
## + race      2    20550 20580
##
## Step:  AIC=17461.16
## y ~ chem
##
##           Df Deviance   AIC
## + surgery   1    16836 16866
## + grade     4    16837 16897
## + t         2    17101 17141
## + N         2    17183 17223
## + size      2    17211 17251
## + lung      2    17263 17303
## + bone      2    17297 17336
## + RX        1    17329 17359
## + age       2    17333 17373
## + site      1    17384 17414
## + brain     2    17378 17418
## + CEA       2    17379 17419
## + kind      1    17420 17450
## + marry     2    17411 17451
## <none>      17441 17461

```

```

## + radiate 1 17433 17463
## + gender 1 17440 17470
## + race 2 17431 17471
## + income 2 17440 17480
##
## Step: AIC=16866.07
## y ~ chem + surgery
##
##           Df Deviance  AIC
## + grade 4 16517 16586
## + site 1 16706 16746
## + age 2 16740 16789
## + bone 2 16749 16798
## + lung 2 16764 16814
## + N 2 16767 16817
## + brain 2 16789 16838
## + kind 1 16803 16842
## + radiate 1 16813 16853
## + size 2 16808 16858
## + CEA 2 16810 16860
## + marry 2 16814 16864
## <none> 16836 16866
## + RX 1 16827 16867
## + gender 1 16832 16872
## + t 2 16825 16875
## + race 2 16826 16876
## + income 2 16834 16884
##
## Step: AIC=16586.38
## y ~ chem + surgery + grade
##
##           Df Deviance  AIC
## + site 1 16415 16495
## + age 2 16422 16511
## + lung 2 16435 16525
## + bone 2 16438 16528
## + brain 2 16472 16562
## + CEA 2 16483 16573
## + N 2 16486 16575
## + radiate 1 16496 16576
## + kind 1 16502 16582
## + marry 2 16493 16582
## + size 2 16496 16586
## <none> 16517 16586
## + RX 1 16508 16588
## + gender 1 16513 16593
## + race 2 16505 16595
## + t 2 16510 16599
## + income 2 16513 16602
##
## Step: AIC=16494.84
## y ~ chem + surgery + grade + site
##
##           Df Deviance  AIC

```

```

## + lung      2      16326 16425
## + bone      2      16331 16431
## + age       2      16344 16444
## + brain     2      16371 16471
## + CEA       2      16379 16479
## + N         2      16388 16488
## + size      2      16393 16493
## + marry     2      16394 16494
## + kind      1      16405 16495
## <none>      16415 16495
## + radiate   1      16407 16497
## + RX        1      16408 16498
## + gender    1      16414 16504
## + race      2      16409 16509
## + t         2      16411 16510
## + income    2      16412 16512
##
## Step:  AIC=16425.42
## y ~ chem + surgery + grade + site + lung
##
##           Df Deviance   AIC
## + age      2      16258 16377
## + bone      2      16260 16379
## + brain     2      16293 16412
## + CEA       2      16297 16416
## + N         2      16301 16421
## + kind      1      16314 16423
## + size      2      16304 16424
## + marry     2      16305 16424
## <none>      16326 16425
## + radiate   1      16316 16426
## + RX        1      16321 16430
## + gender    1      16325 16435
## + race      2      16320 16439
## + t         2      16322 16441
## + income    2      16322 16442
##
## Step:  AIC=16377.19
## y ~ chem + surgery + grade + site + lung + age
##
##           Df Deviance   AIC
## + bone      2      16190 16329
## + brain     2      16222 16362
## + CEA       2      16228 16368
## + N         2      16232 16371
## + size      2      16235 16374
## + kind      1      16246 16375
## <none>      16258 16377
## + radiate   1      16249 16378
## + marry     2      16239 16379
## + RX        1      16254 16384
## + gender    1      16257 16387
## + race      2      16251 16391
## + income    2      16253 16393

```

```

## + t          2      16254 16394
##
## Step: AIC=16329.09
## y ~ chem + surgery + grade + site + lung + age + bone
##
##           Df Deviance   AIC
## + radiate  1      16169 16319
## + CEA      2      16161 16320
## + brain    2      16161 16321
## + N        2      16165 16325
## + size     2      16166 16326
## + kind     1      16178 16328
## <none>      16190 16329
## + marry    2      16170 16330
## + RX       1      16186 16336
## + gender   1      16189 16338
## + race     2      16183 16342
## + income   2      16185 16345
## + t        2      16186 16345
##
## Step: AIC=16318.9
## y ~ chem + surgery + grade + site + lung + age + bone + radiate
##
##           Df Deviance   AIC
## + brain    2      16130 16299
## + CEA      2      16142 16312
## + N        2      16145 16314
## + size     2      16146 16316
## + kind     1      16158 16318
## <none>      16169 16319
## + marry    2      16150 16320
## + RX       1      16167 16326
## + gender   1      16169 16328
## + race     2      16163 16332
## + income   2      16165 16335
## + t        2      16167 16337
##
## Step: AIC=16299.1
## y ~ chem + surgery + grade + site + lung + age + bone + radiate +
##      brain
##
##           Df Deviance   AIC
## + CEA      2      16103 16292
## + N        2      16105 16295
## + size     2      16106 16295
## + kind     1      16118 16298
## <none>      16130 16299
## + marry    2      16110 16300
## + RX       1      16126 16305
## + gender   1      16129 16309
## + race     2      16123 16313
## + income   2      16125 16315
## + t        2      16128 16317
##

```

```

## Step: AIC=16292.02
## y ~ chem + surgery + grade + site + lung + age + bone + radiate +
##      brain + CEA
##
##      Df Deviance  AIC
## + N      2    16078 16288
## + size    2    16079 16288
## + kind    1    16089 16288
## <none>      16103 16292
## + marry   2    16084 16294
## + RX      1    16099 16299
## + gender  1    16102 16302
## + race    2    16096 16306
## + income  2    16098 16308
## + t       2    16101 16310
##
## Step: AIC=16287.67
## y ~ chem + surgery + grade + site + lung + age + bone + radiate +
##      brain + CEA + N
##
##      Df Deviance  AIC
## + kind    1    16064 16284
## + size    2    16056 16286
## <none>      16078 16288
## + marry   2    16059 16288
## + RX      1    16075 16295
## + gender  1    16078 16297
## + race    2    16071 16301
## + income  2    16073 16303
## + t       2    16076 16306
##
## Step: AIC=16283.88
## y ~ chem + surgery + grade + site + lung + age + bone + radiate +
##      brain + CEA + N + kind
##
##      Df Deviance  AIC
## + size    2    16044 16283
## <none>      16064 16284
## + marry   2    16046 16285
## + RX      1    16061 16291
## + gender  1    16064 16294
## + race    2    16058 16297
## + income  2    16060 16299
## + t       2    16062 16302
##
## Step: AIC=16283.08
## y ~ chem + surgery + grade + site + lung + age + bone + radiate +
##      brain + CEA + N + kind + size
##
##      Df Deviance  AIC
## <none>      16044 16283
## + marry   2    16026 16285
## + RX      1    16041 16290
## + gender  1    16043 16293

```



```
## + race      2      16037 16296
## + income    2      16039 16298
## + t         2      16041 16301
```

```
formula(forwards_bic)
```

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

```
summary(forwards_bic)
```

```
##
## Call:
## glm(formula = y ~ chem + surgery + grade + site + lung + age +
##      bone + radiate + brain + CEA + N + kind + size, family = binomial,
##      data = train_0_1)
##
## Deviance Residuals:
##      Min        1Q      Median        3Q        Max
## -2.1550   -0.5632   -0.3694   -0.2419    2.9262
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -1.29341    0.20391  -6.343 2.25e-10 ***
## chem1        -2.09057    0.04270 -48.956 < 2e-16 ***
## surgery1     -0.74395    0.05858 -12.699 < 2e-16 ***
## grade2        0.14556    0.12142   1.199 0.230609
## grade3        0.85588    0.12739   6.719 1.84e-11 ***
## grade4        0.71620    0.17344   4.129 3.64e-05 ***
## grade9        0.75358    0.12287   6.133 8.61e-10 ***
## site2         0.34501    0.04271   8.079 6.55e-16 ***
## lung1         0.33792    0.04580   7.378 1.61e-13 ***
## lung2         0.01143    0.15088   0.076 0.939606
## age2          -0.04462    0.14585  -0.306 0.759677
## age3          0.34689    0.14319   2.423 0.015411 *
## bone1         0.66965    0.07813   8.571 < 2e-16 ***
## bone2         0.07187    0.21330   0.337 0.736161
## radiate1     -0.40184    0.07499  -5.359 8.38e-08 ***
## brain1        1.08575    0.17275   6.285 3.27e-10 ***
## brain2       -0.27125    0.19858  -1.366 0.171963
## CEA1          0.39548    0.07922   4.992 5.97e-07 ***
## CEA2          0.28052    0.08459   3.316 0.000913 ***
## N2            0.16150    0.06078   2.657 0.007878 **
## N3            0.22336    0.05392   4.142 3.44e-05 ***
## kind2         0.25968    0.07260   3.577 0.000347 ***
## size2         0.21485    0.05098   4.214 2.51e-05 ***
## size3         0.18055    0.05334   3.385 0.000712 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 20561  on 21390  degrees of freedom
```

```
## Residual deviance: 16044 on 21367 degrees of freedom
## AIC: 16092
##
## Number of Fisher Scoring iterations: 5
```

```
bothways_bic = step(nothing, list(lower=formula(nothing),upper=formula(fullmod)), direction="both", k=1)
```

```
## Start: AIC=20570.85
## y ~ 1
##
##           Df Deviance   AIC
## + chem      1    17441 17461
## + surgery    1    19928 19948
## + grade      4    20014 20064
## + age        2    20080 20110
## + t          2    20199 20229
## + size       2    20273 20303
## + N          2    20282 20312
## + RX         1    20310 20330
## + lung       2    20379 20409
## + marry      2    20381 20411
## + bone       2    20395 20425
## + site       1    20417 20437
## + radiate    1    20464 20484
## + CEA        2    20460 20490
## + brain      2    20475 20504
## + kind       1    20519 20539
## + gender     1    20548 20568
## <none>              20561 20571
## + income     2    20546 20576
## + race       2    20550 20580
##
```

```
## Step: AIC=17461.16
## y ~ chem
##
##           Df Deviance   AIC
## + surgery    1    16836 16866
## + grade      4    16837 16897
## + t          2    17101 17141
## + N          2    17183 17223
## + size       2    17211 17251
## + lung       2    17263 17303
## + bone       2    17297 17336
## + RX         1    17329 17359
## + age        2    17333 17373
## + site       1    17384 17414
## + brain      2    17378 17418
## + CEA        2    17379 17419
## + kind       1    17420 17450
## + marry      2    17411 17451
## <none>              17441 17461
## + radiate    1    17433 17463
## + gender     1    17440 17470
## + race       2    17431 17471
```

```

## + income  2    17440 17480
## - chem    1    20561 20571
##
## Step:  AIC=16866.07
## y ~ chem + surgery
##
##           Df Deviance   AIC
## + grade    4    16517 16586
## + site     1    16706 16746
## + age      2    16740 16789
## + bone     2    16749 16798
## + lung     2    16764 16814
## + N        2    16767 16817
## + brain    2    16789 16838
## + kind     1    16803 16842
## + radiate  1    16813 16853
## + size     2    16808 16858
## + CEA      2    16810 16860
## + marry    2    16814 16864
## <none>      16836 16866
## + RX       1    16827 16867
## + gender   1    16832 16872
## + t        2    16825 16875
## + race     2    16826 16876
## + income   2    16834 16884
## - surgery  1    17441 17461
## - chem     1    19928 19948
##
## Step:  AIC=16586.38
## y ~ chem + surgery + grade
##
##           Df Deviance   AIC
## + site     1    16415 16495
## + age      2    16422 16511
## + lung     2    16435 16525
## + bone     2    16438 16528
## + brain    2    16472 16562
## + CEA      2    16483 16573
## + N        2    16486 16575
## + radiate  1    16496 16576
## + kind     1    16502 16582
## + marry    2    16493 16582
## + size     2    16496 16586
## <none>      16517 16586
## + RX       1    16508 16588
## + gender   1    16513 16593
## + race     2    16505 16595
## + t        2    16510 16599
## + income   2    16513 16602
## - grade    4    16836 16866
## - surgery  1    16837 16897
## - chem     1    19626 19686
##
## Step:  AIC=16494.84

```

```

## y ~ chem + surgery + grade + site
##
##           Df Deviance   AIC
## + lung      2    16326 16425
## + bone      2    16331 16431
## + age       2    16344 16444
## + brain     2    16371 16471
## + CEA       2    16379 16479
## + N         2    16388 16488
## + size      2    16393 16493
## + marry     2    16394 16494
## + kind      1    16405 16495
## <none>      16415 16495
## + radiate   1    16407 16497
## + RX        1    16408 16498
## + gender    1    16414 16504
## + race      2    16409 16509
## + t         2    16411 16510
## + income    2    16412 16512
## - site      1    16517 16586
## - grade     4    16706 16746
## - surgery   1    16776 16846
## - chem      1    19421 19491
##
## Step: AIC=16425.42
## y ~ chem + surgery + grade + site + lung
##
##           Df Deviance   AIC
## + age      2    16258 16377
## + bone     2    16260 16379
## + brain    2    16293 16412
## + CEA      2    16297 16416
## + N        2    16301 16421
## + kind     1    16314 16423
## + size     2    16304 16424
## + marry    2    16305 16424
## <none>     16326 16425
## + radiate  1    16316 16426
## + RX       1    16321 16430
## + gender   1    16325 16435
## + race     2    16320 16439
## + t        2    16322 16441
## + income   2    16322 16442
## - lung     2    16415 16495
## - site     1    16435 16525
## - grade    4    16624 16684
## - surgery  1    16619 16709
## - chem     1    19340 19430
##
## Step: AIC=16377.19
## y ~ chem + surgery + grade + site + lung + age
##
##           Df Deviance   AIC
## + bone     2    16190 16329

```

```

## + brain      2      16222 16362
## + CEA        2      16228 16368
## + N          2      16232 16371
## + size       2      16235 16374
## + kind       1      16246 16375
## <none>              16258 16377
## + radiate    1      16249 16378
## + marry      2      16239 16379
## + RX         1      16254 16384
## + gender     1      16257 16387
## + race       2      16251 16391
## + income     2      16253 16393
## + t          2      16254 16394
## - age        2      16326 16425
## - lung       2      16344 16444
## - site       1      16342 16452
## - grade      4      16557 16637
## - surgery    1      16543 16653
## - chem       1      18983 19092
##
## Step: AIC=16329.09
## y ~ chem + surgery + grade + site + lung + age + bone
##
##           Df Deviance   AIC
## + radiate  1      16169 16319
## + CEA      2      16161 16320
## + brain    2      16161 16321
## + N        2      16165 16325
## + size     2      16166 16326
## + kind     1      16178 16328
## <none>              16190 16329
## + marry    2      16170 16330
## + RX       1      16186 16336
## + gender   1      16189 16338
## + race     2      16183 16342
## + income   2      16185 16345
## + t        2      16186 16345
## - bone     2      16258 16377
## - lung     2      16258 16378
## - age      2      16260 16379
## - site     1      16277 16407
## - grade    4      16480 16579
## - surgery  1      16455 16585
## - chem     1      18908 19037
##
## Step: AIC=16318.9
## y ~ chem + surgery + grade + site + lung + age + bone + radiate
##
##           Df Deviance   AIC
## + brain    2      16130 16299
## + CEA      2      16142 16312
## + N        2      16145 16314
## + size     2      16146 16316
## + kind     1      16158 16318

```

```

## <none>          16169 16319
## + marry        2    16150 16320
## + RX           1    16167 16326
## + gender       1    16169 16328
## - radiate      1    16190 16329
## + race         2    16163 16332
## + income       2    16165 16335
## + t            2    16167 16337
## - age          2    16238 16368
## - lung         2    16239 16369
## - bone         2    16249 16378
## - site         1    16242 16382
## - grade        4    16458 16567
## - surgery      1    16438 16577
## - chem         1    18818 18958
##
## Step: AIC=16299.1
## y ~ chem + surgery + grade + site + lung + age + bone + radiate +
##      brain
##
##           Df Deviance   AIC
## + CEA      2    16103 16292
## + N         2    16105 16295
## + size      2    16106 16295
## + kind      1    16118 16298
## <none>      16130 16299
## + marry    2    16110 16300
## + RX       1    16126 16305
## + gender   1    16129 16309
## + race     2    16123 16313
## + income   2    16125 16315
## + t        2    16128 16317
## - brain    2    16169 16319
## - radiate  1    16161 16321
## - lung     2    16191 16340
## - age      2    16200 16349
## - bone     2    16204 16354
## - site     1    16198 16358
## - grade    4    16414 16544
## - surgery  1    16396 16556
## - chem     1    18759 18919
##
## Step: AIC=16292.02
## y ~ chem + surgery + grade + site + lung + age + bone + radiate +
##      brain + CEA
##
##           Df Deviance   AIC
## + N         2    16078 16288
## + size      2    16079 16288
## + kind      1    16089 16288
## <none>      16103 16292
## + marry    2    16084 16294
## + RX       1    16099 16299
## - CEA      2    16130 16299

```

```

## + gender    1    16102 16302
## + race      2    16096 16306
## + income    2    16098 16308
## + t         2    16101 16310
## - brain     2    16142 16312
## - radiate   1    16132 16312
## - lung      2    16158 16328
## - age       2    16173 16342
## - bone      2    16176 16346
## - site      1    16174 16353
## - surgery   1    16356 16536
## - grade     4    16394 16543
## - chem      1    18699 18878
##
## Step:  AIC=16287.67
## y ~ chem + surgery + grade + site + lung + age + bone + radiate +
##      brain + CEA + N
##
##           Df Deviance   AIC
## + kind      1    16064 16284
## + size      2    16056 16286
## <none>             16078 16288
## + marry     2    16059 16288
## - N         2    16103 16292
## + RX        1    16075 16295
## - CEA       2    16105 16295
## + gender    1    16078 16297
## + race      2    16071 16301
## + income    2    16073 16303
## + t         2    16076 16306
## - brain     2    16118 16307
## - radiate   1    16108 16308
## - lung      2    16132 16321
## - age       2    16150 16339
## - bone      2    16150 16340
## - site      1    16146 16346
## - surgery   1    16282 16481
## - grade     4    16336 16506
## - chem      1    18674 18874
##
## Step:  AIC=16283.88
## y ~ chem + surgery + grade + site + lung + age + bone + radiate +
##      brain + CEA + N + kind
##
##           Df Deviance   AIC
## + size      2    16044 16283
## <none>             16064 16284
## + marry     2    16046 16285
## - kind      1    16078 16288
## - N         2    16089 16288
## + RX        1    16061 16291
## - CEA       2    16094 16293
## + gender    1    16064 16294
## + race      2    16058 16297

```

```
## + income 2 16060 16299
## + t 2 16062 16302
## - brain 2 16104 16304
## - radiate 1 16095 16304
## - lung 2 16120 16319
## - bone 2 16135 16335
## - age 2 16136 16336
## - site 1 16129 16338
## - surgery 1 16270 16480
## - grade 4 16308 16488
## - chem 1 18650 18860
##
## Step: AIC=16283.08
## y ~ chem + surgery + grade + site + lung + age + bone + radiate +
## brain + CEA + N + kind + size
##
## Df Deviance AIC
## <none> 16044 16283
## - size 2 16064 16284
## + marry 2 16026 16285
## - kind 1 16056 16286
## - N 2 16066 16286
## + RX 1 16041 16290
## - CEA 2 16072 16292
## + gender 1 16043 16293
## + race 2 16037 16296
## + income 2 16039 16298
## + t 2 16041 16301
## - radiate 1 16074 16303
## - brain 2 16085 16304
## - lung 2 16098 16318
## - bone 2 16114 16334
## - age 2 16117 16336
## - site 1 16109 16338
## - surgery 1 16209 16439
## - grade 4 16283 16482
## - chem 1 18608 18837
```

```
formula(bothways_bic)
```

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

```
summary(bothways_bic)
```

```
##
## Call:
## glm(formula = y ~ chem + surgery + grade + site + lung + age +
## bone + radiate + brain + CEA + N + kind + size, family = binomial,
## data = train_0_1)
##
## Deviance Residuals:
## Min 1Q Median 3Q Max
```



```
## -2.1550 -0.5632 -0.3694 -0.2419 2.9262
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.29341    0.20391  -6.343 2.25e-10 ***
## chem1       -2.09057    0.04270 -48.956 < 2e-16 ***
## surgery1    -0.74395    0.05858 -12.699 < 2e-16 ***
## grade2       0.14556    0.12142   1.199 0.230609
## grade3       0.85588    0.12739   6.719 1.84e-11 ***
## grade4       0.71620    0.17344   4.129 3.64e-05 ***
## grade9       0.75358    0.12287   6.133 8.61e-10 ***
## site2        0.34501    0.04271   8.079 6.55e-16 ***
## lung1        0.33792    0.04580   7.378 1.61e-13 ***
## lung2        0.01143    0.15088   0.076 0.939606
## age2        -0.04462    0.14585  -0.306 0.759677
## age3         0.34689    0.14319   2.423 0.015411 *
## bone1        0.66965    0.07813   8.571 < 2e-16 ***
## bone2        0.07187    0.21330   0.337 0.736161
## radiate1     -0.40184    0.07499  -5.359 8.38e-08 ***
## brain1       1.08575    0.17275   6.285 3.27e-10 ***
## brain2      -0.27125    0.19858  -1.366 0.171963
## CEA1         0.39548    0.07922   4.992 5.97e-07 ***
## CEA2         0.28052    0.08459   3.316 0.000913 ***
## N2           0.16150    0.06078   2.657 0.007878 **
## N3           0.22336    0.05392   4.142 3.44e-05 ***
## kind2        0.25968    0.07260   3.577 0.000347 ***
## size2        0.21485    0.05098   4.214 2.51e-05 ***
## size3        0.18055    0.05334   3.385 0.000712 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 20561 on 21390 degrees of freedom
## Residual deviance: 16044 on 21367 degrees of freedom
## AIC: 16092
##
## Number of Fisher Scoring iterations: 5
```

0+1 lasso

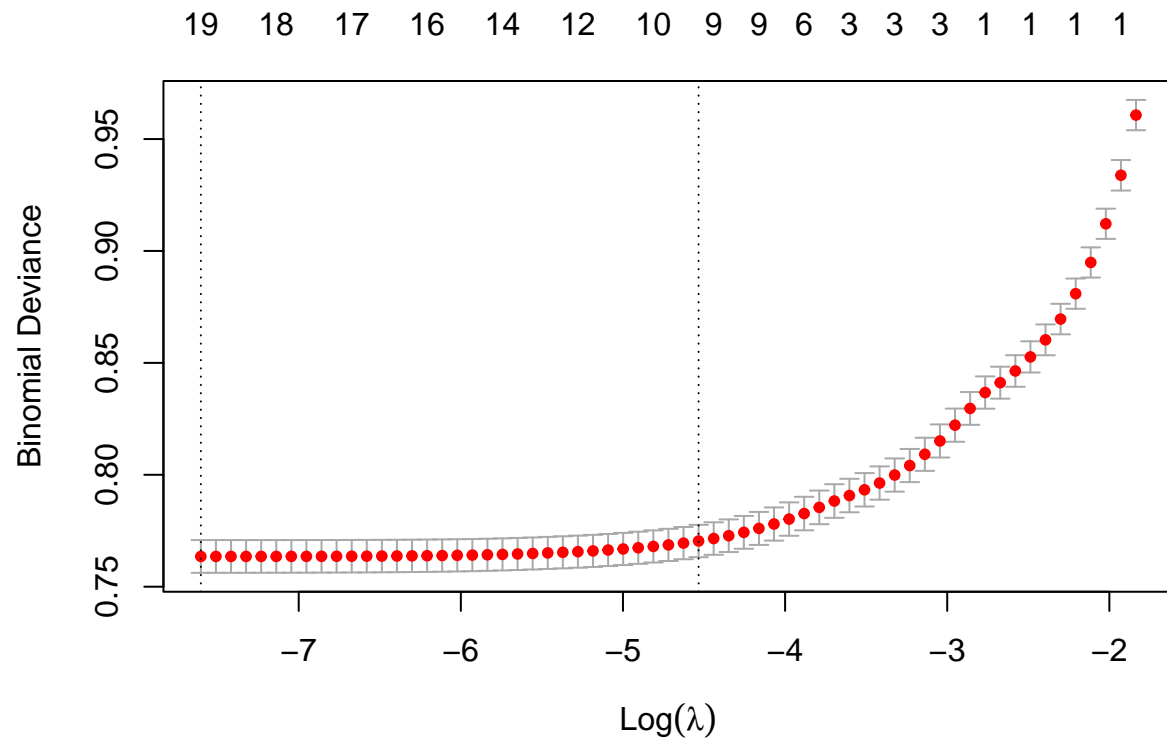
```
x <- data.matrix(train_0_1[, 2:ncol(train_0_1)])
y <- as.numeric(data.matrix(train_0_1$y))

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

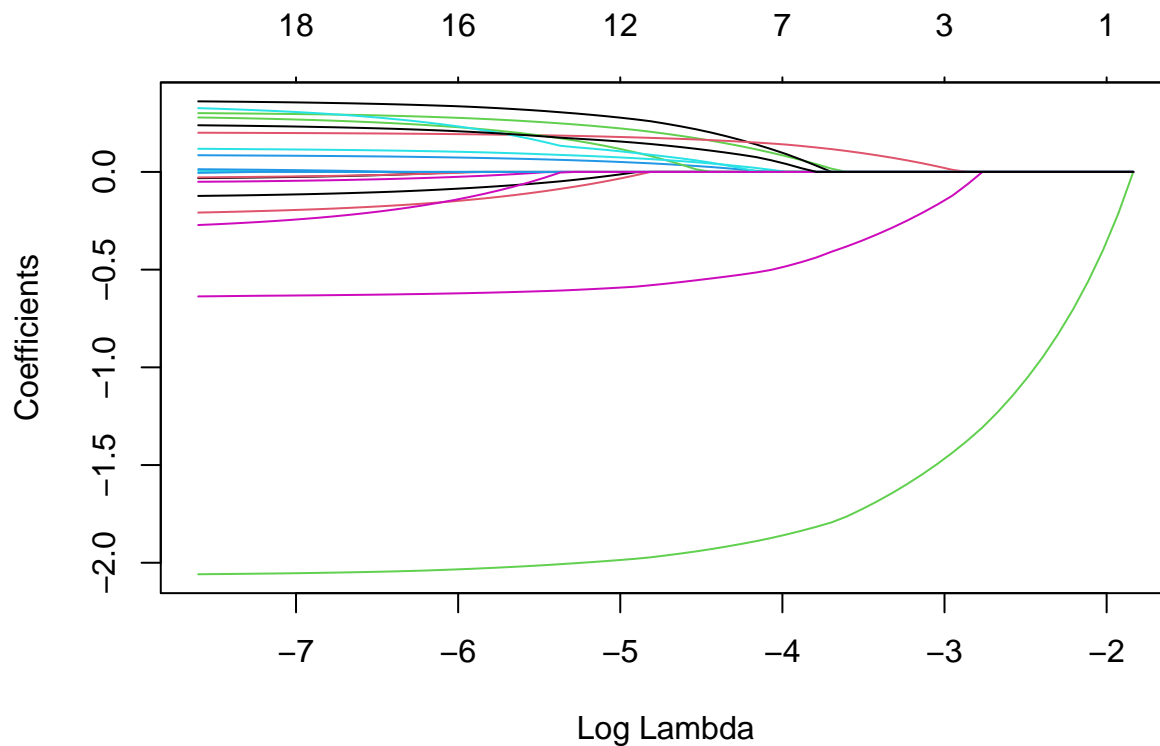
#find optimal lambda value that minimizes test MSE
best_lambda <- cv_model$lambda.min
best_lambda
```

```
## [1] 0.0004986309
```

```
#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)  0.378266999
## gender      -0.031699473
## race        -0.027928531
## age         0.301090006
## size        0.085547659
## marry       0.005554809
## income      -0.050357937
## site        0.361166556
## grade       0.200656051
## kind        0.278727760
## t           -0.004669078
## N           0.118484341
## surgery     -0.636793155
## RX          -0.122754381
## radiate     -0.207935059
## chem        -2.058528363
## CEA         0.013796220
## bone        0.326453458
## brain       -0.271595099
## lung        0.238815385
```

```
print(best_model)
```

```
##
```

```
## Call:  glmnet(x = x, y = y, family = binomial(), alpha = 1, lambda = best_lambda)
##
##      Df  %Dev   Lambda
## 1 19 20.78 0.0004986
```