Untitled

Nathaniel Brown, In Hee Ho, Sarah Zimmermann

October 19, 2017

Introduction

The data are from a study of time to critical neurological assessment for patients with stroke-like symptoms who are admitted to the emergency room. We are interested in the factors predictive of the time to assessment following admission to the ED for n=335 patients with mild to moderate motor impairment. The goal of the analysis is to perform inferences on the impact of clinical presentation, gender, and race (Black, Hispanic, and others) on time to neurological assessment, where clinical presentation is measured as the number of the four major stroke symptoms: headache, loss of motor skills or weakness, trouble talking or understanding, and vision problems. However, as discussed in our previous report, we group Blacks and Hispanics together, and number of symptoms of 3 and 4 together, due to their small sample size.

Methods

The team has cleaned, understood, and modeled these time to critical neurological assessment for patients with stroke-like symptoms data in order to solve the scientific problem of exploring if gender, race/ethnicity, and clinical presentation have an affect on wait list to assessment. To do so the team has approached the problem as such: *

Data Exploration

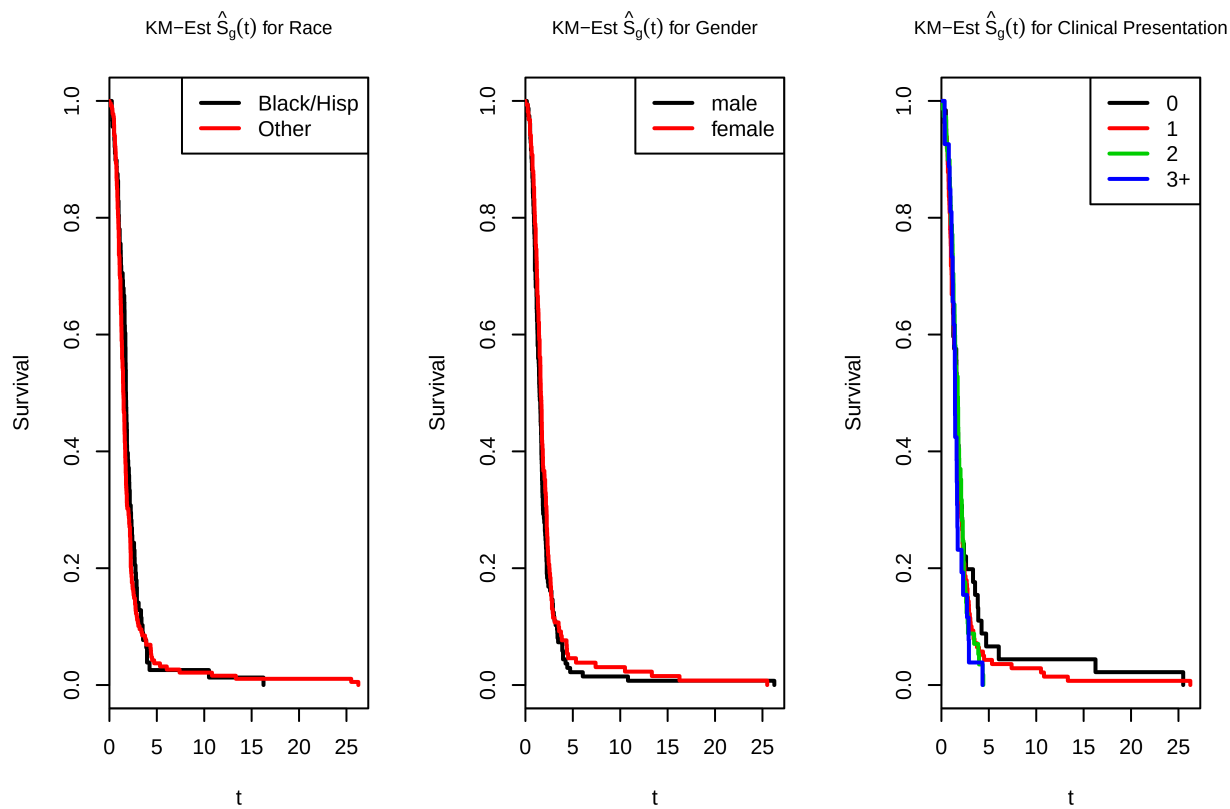
Variables

The original data set contains 335 observations across 9 variables. They are defined as:

Variable Name	Short Description	Type
nctdel	min of neurologist time to assessment & CT scan from arrival at ER	continous
fail	1 if got neurologist/CT scan & 0 otherwise	categorical
male	1 if male, 0 if female	categorical
black	1 if black, 0 if not black	categorical
hisp	1 if hispanic, 0 if not hispanic	categorical
sn1	0/1 indicator 1 main symptom	categorical
sn2	0/1 indicator 2 main symptoms	categorical
sn3	0/1 indicator 3 main symptoms	categorical

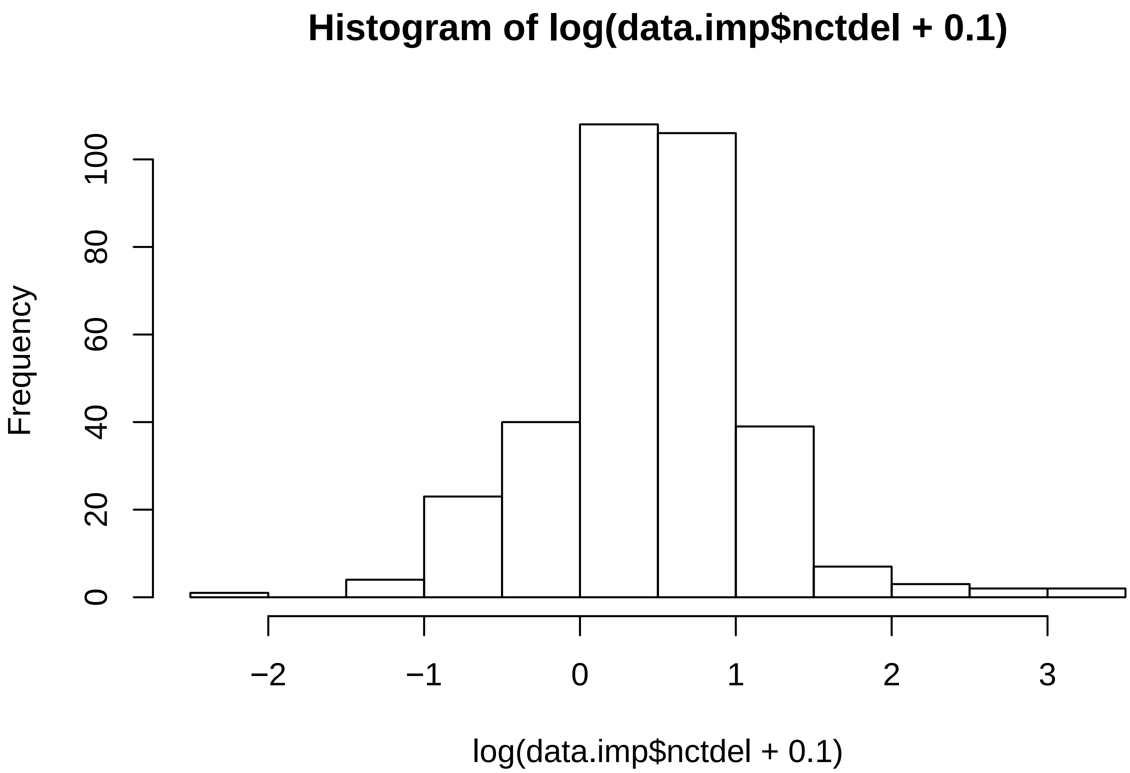
Variable Name	Short Description	Type
all4	0/1 indicator all main sumptoms	categorical

Exploratory Data Analysis

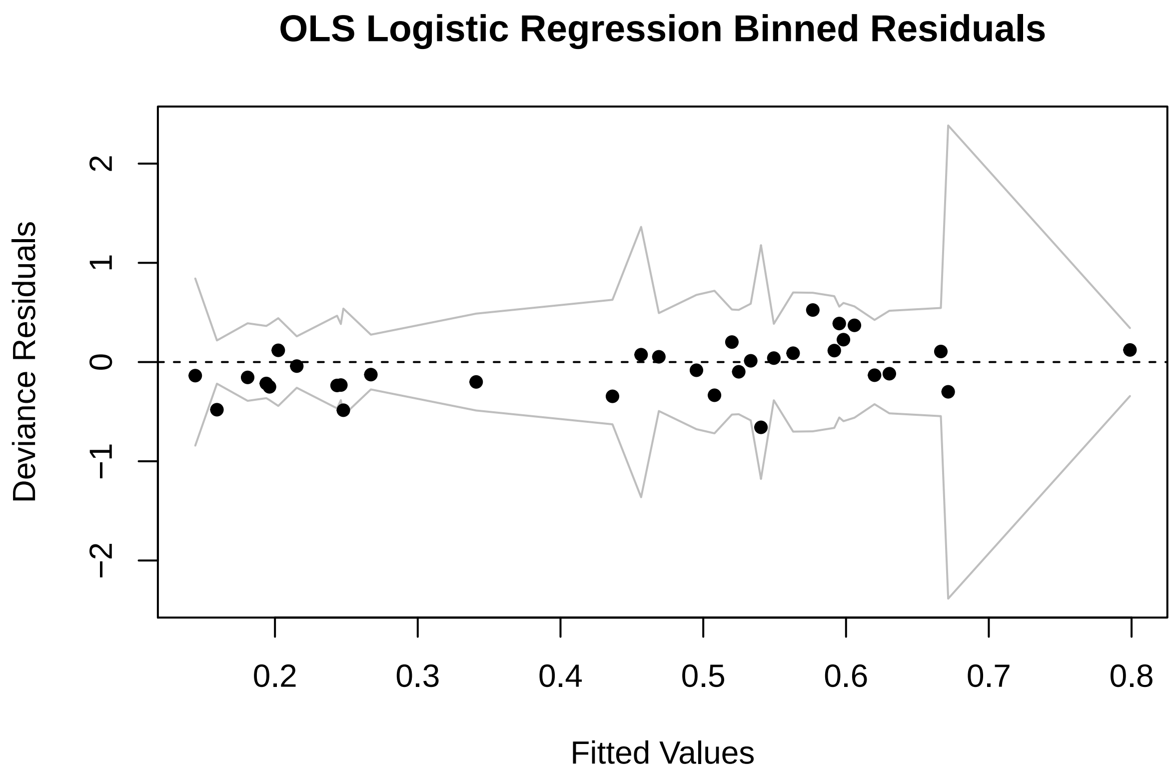


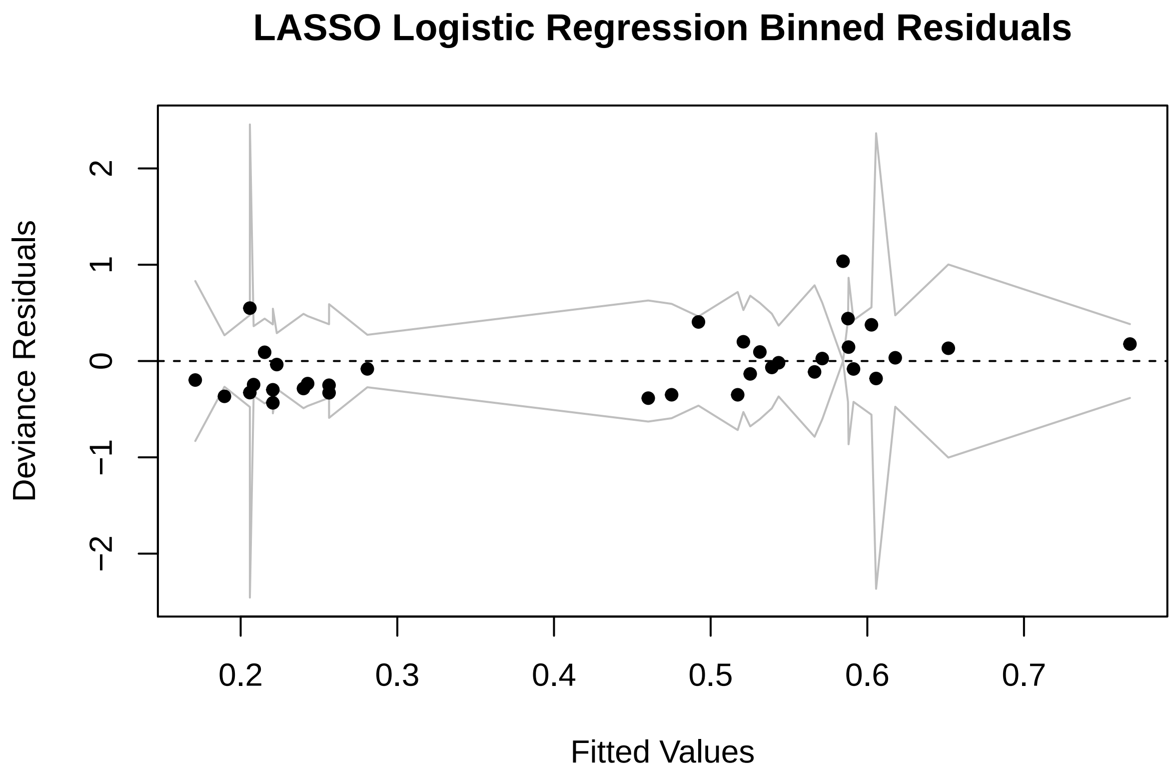
Initial Model Exploration

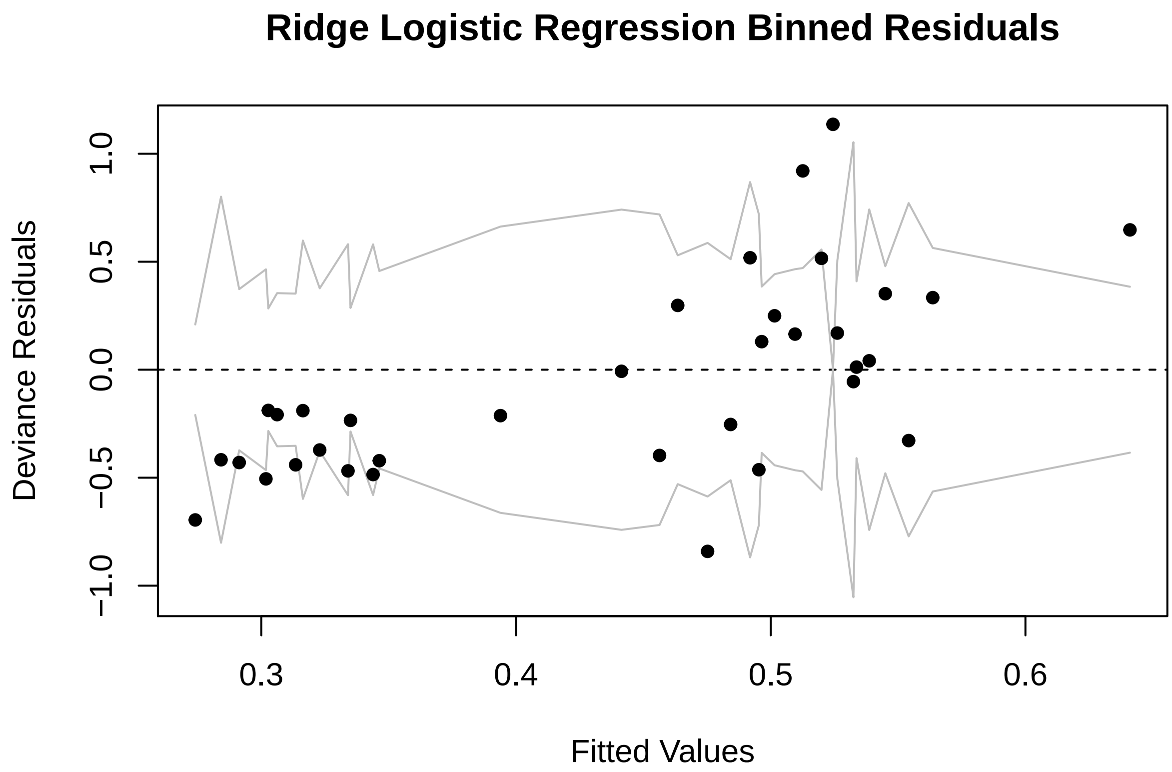
describe process of lasso, ols, and ridge

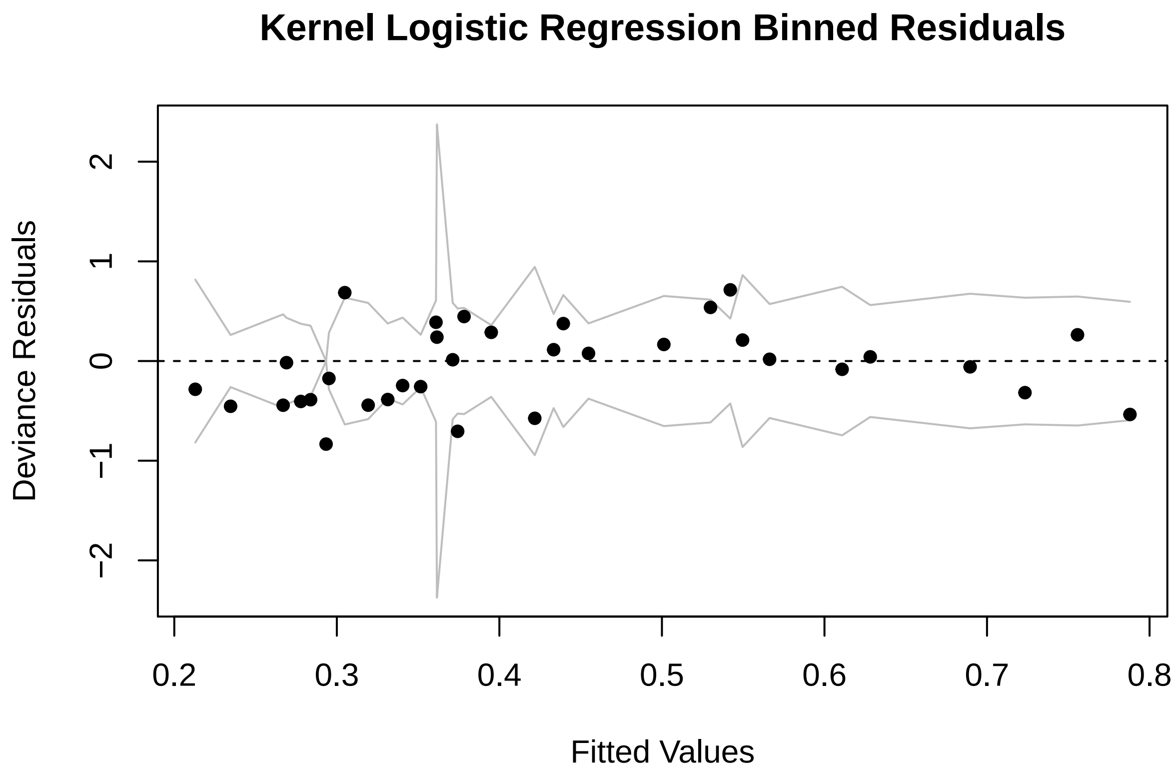

Histogram of log(data.imp$nctdel + 0.1)
Frequency
0
20
40
60
80
100
−2
−1
0
1
2
3
log(data.imp$nctdel + 0.1)

Diagnostic

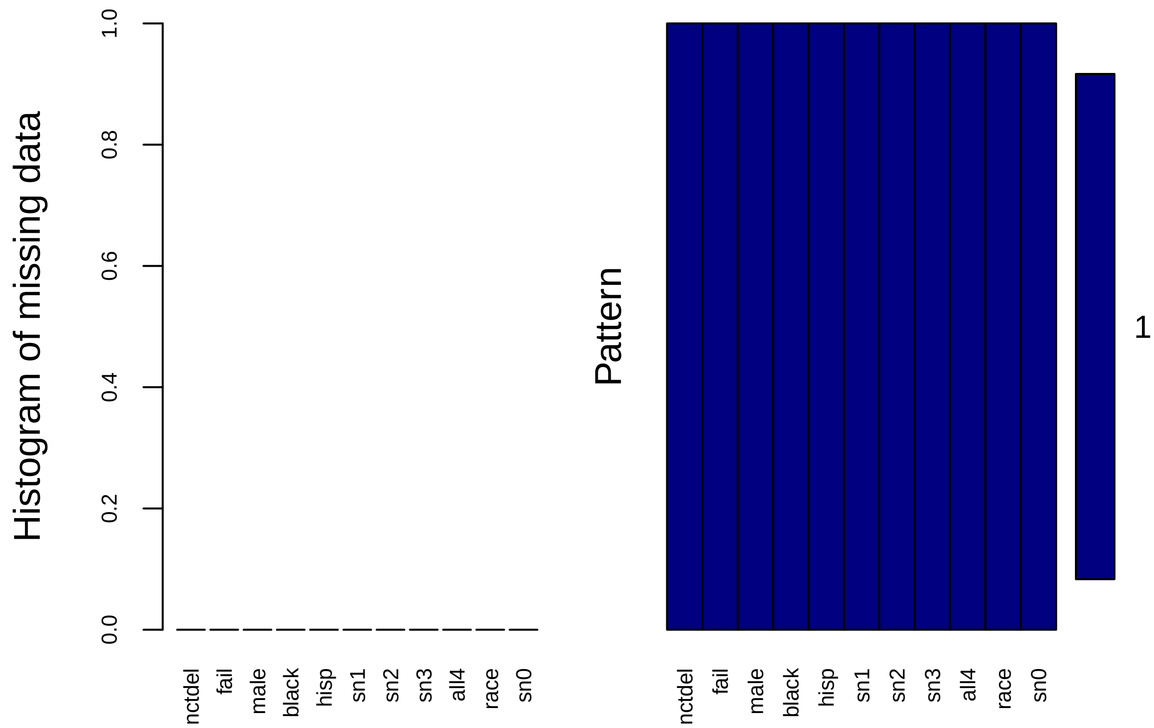

OLS Logistic Regression Binned Residuals
Deviance Residuals
2
1
0
−1
−2
0.2
0.3
0.4
0.5
0.6
0.7
0.8
Fitted Values


LASSO Logistic Regression Binned Residuals
Deviance Residuals
2
1
0
−1
−2
0.2
0.3
0.4
0.5
0.6
0.7
Fitted Values

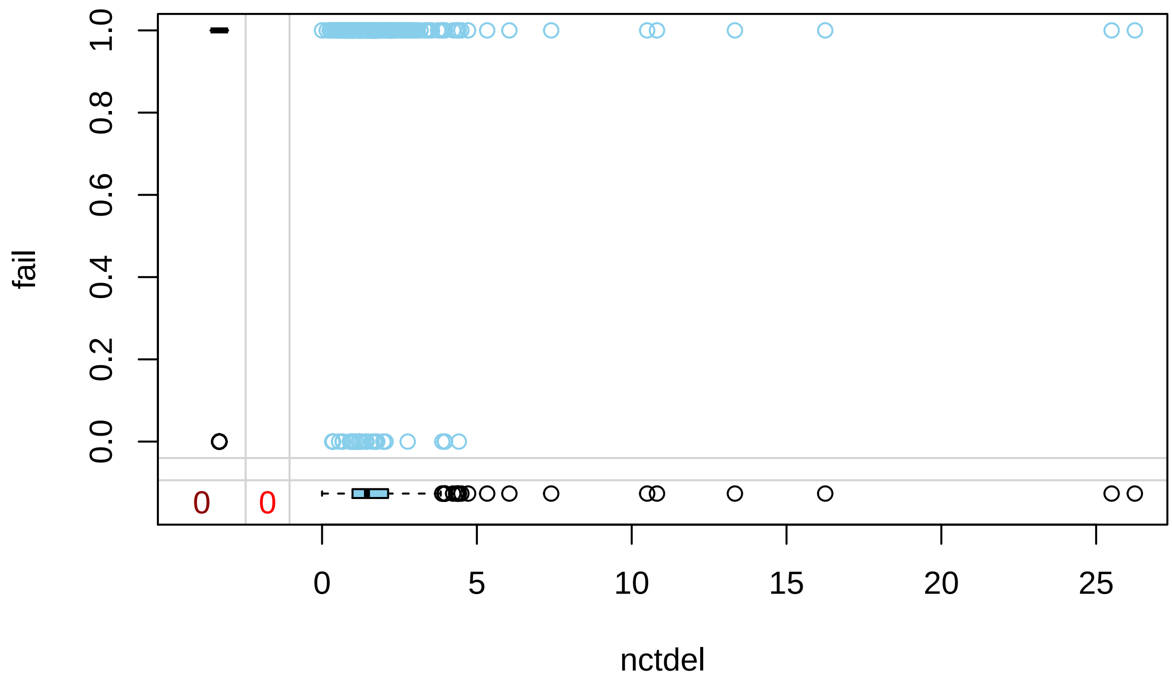

Ridge Logistic Regression Binned Residuals
Deviance Residuals
−1.0
−0.5
0.0
0.5
1.0
0.3
0.4
0.5
0.6
Fitted Values

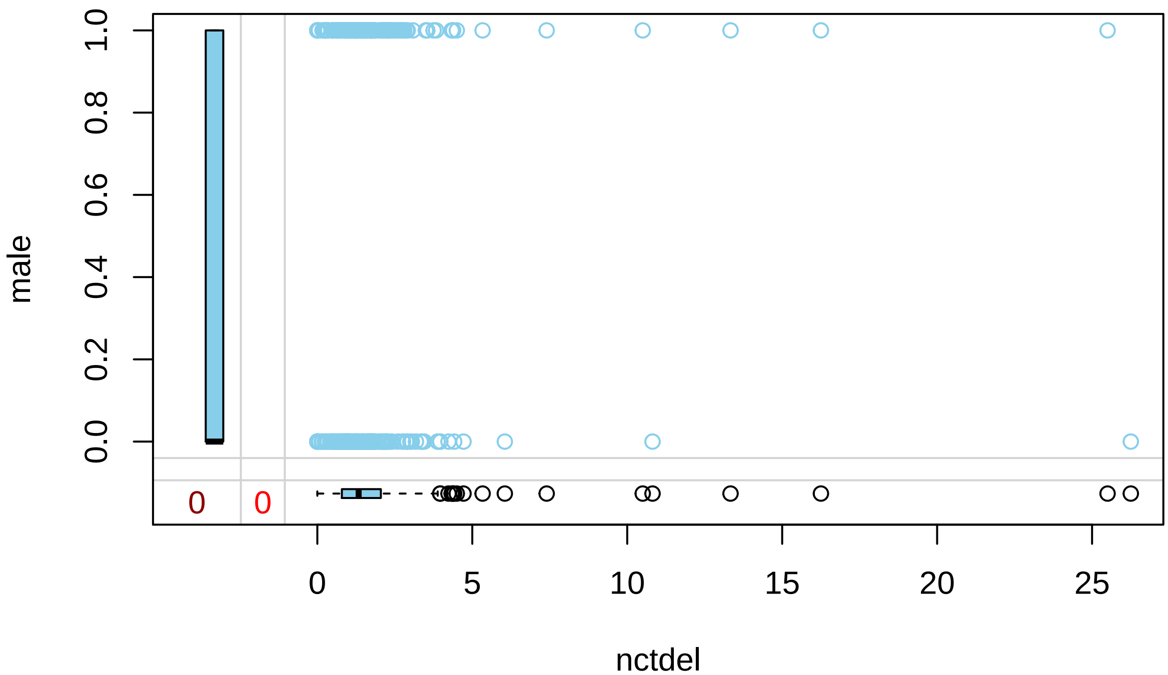



```
##       nctdel fail male black hisp sn1 sn2 sn3 all4 race sn0
## [1,]       1    1    1     1    1   1   1   1    1    1   1 0
## [2,]       0    0    0     0    0   0   0   0    0    0   0 0
```

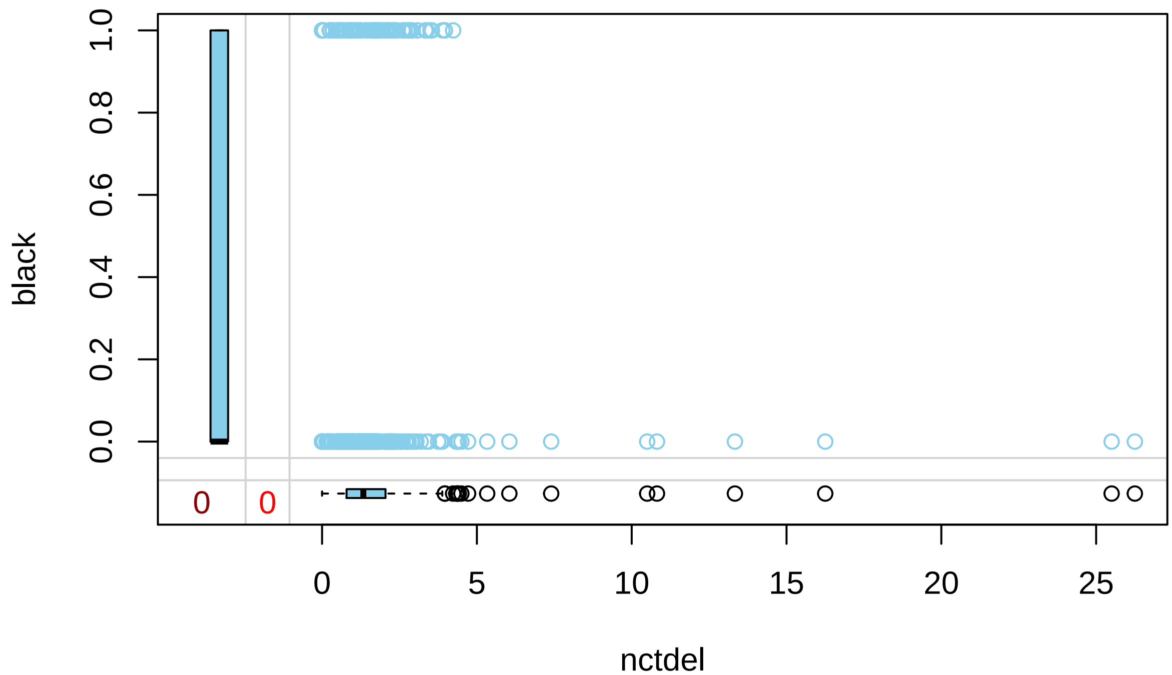


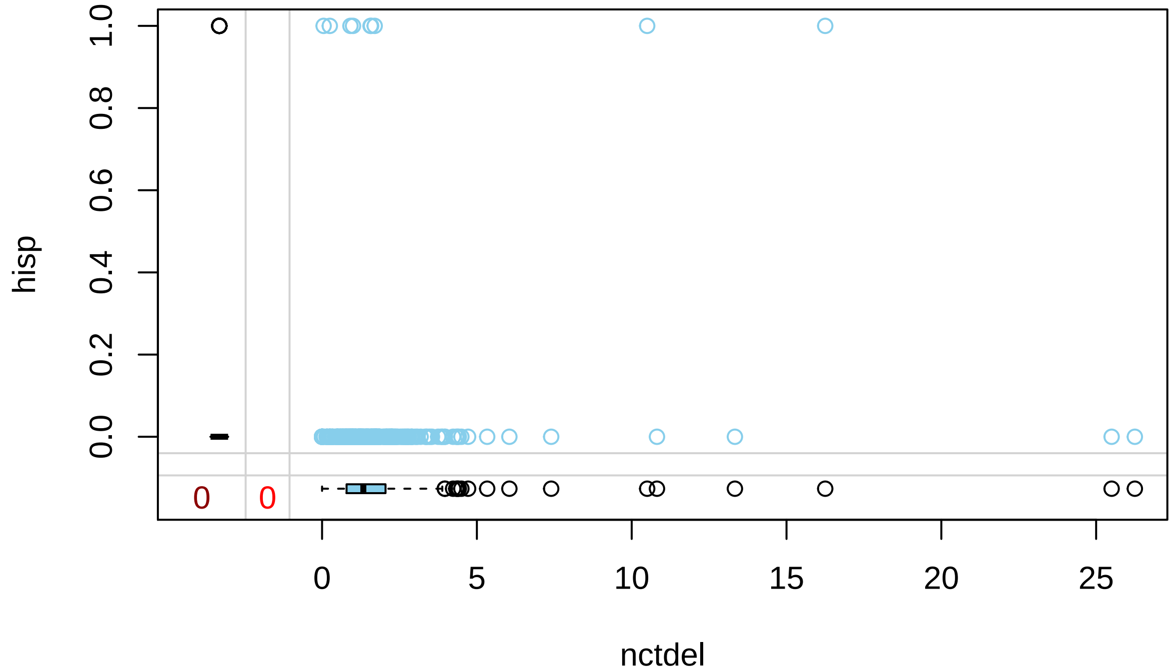

```
##
##  Variables sorted by number of missings:
##  Variable Count
##    nctdel     0
##      fail     0
##      male     0
##     black     0
##      hisp     0
##       sn1     0
##       sn2     0
##       sn3     0
##      all4     0
##      race     0
##       sn0     0
```


fail
0.0
0.2
0.4
0.6
0.8
1.0
0
0
0
5
10
15
20
25
nctdel

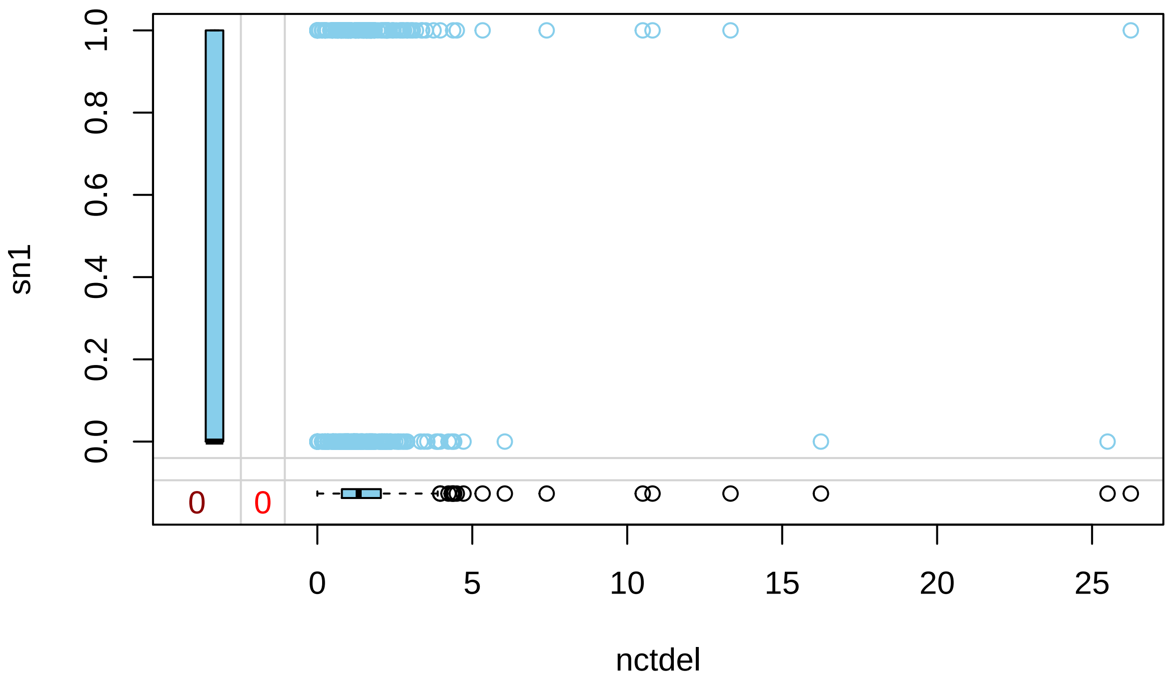


male
1.0
0.8
0.6
0.4
0.2
0.0
0
0
0
5
10
15
20
25
nctdel

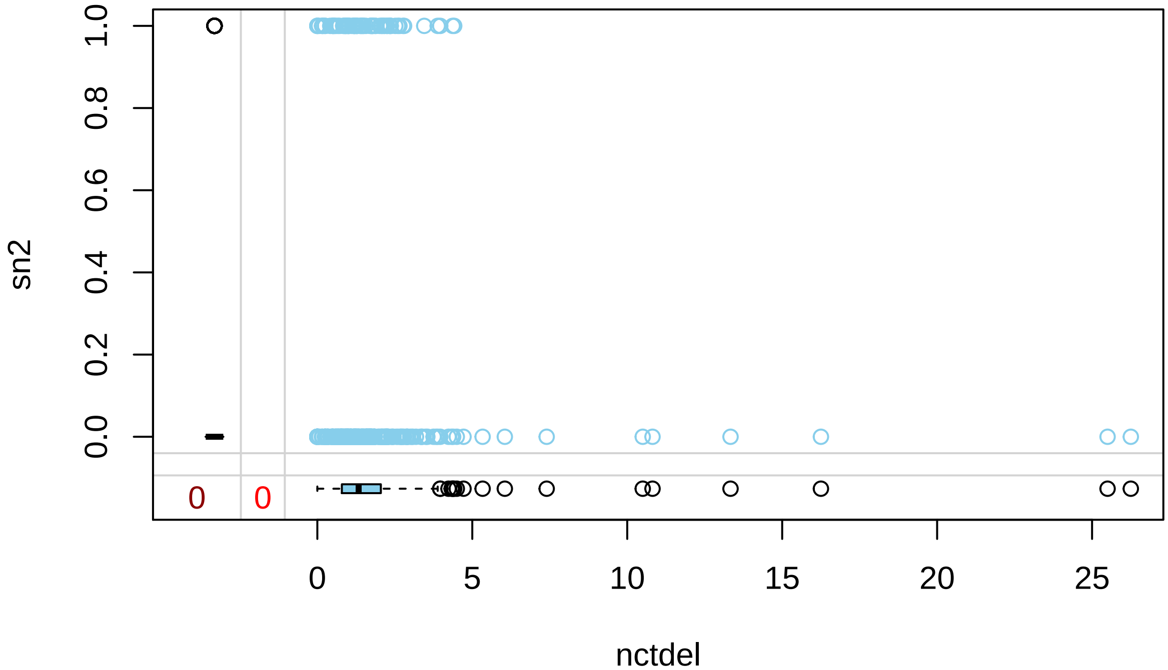

black
0.0
0.2
0.4
0.6
0.8
1.0
0
0
0
5
10
15
20
25
nctdel

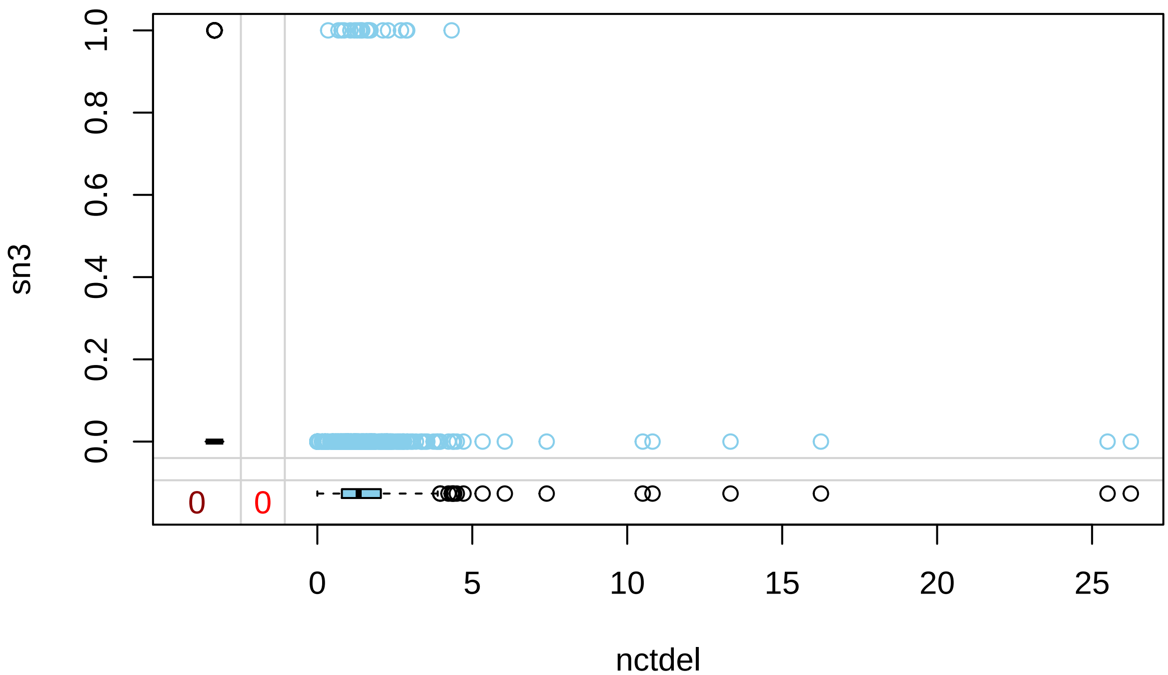


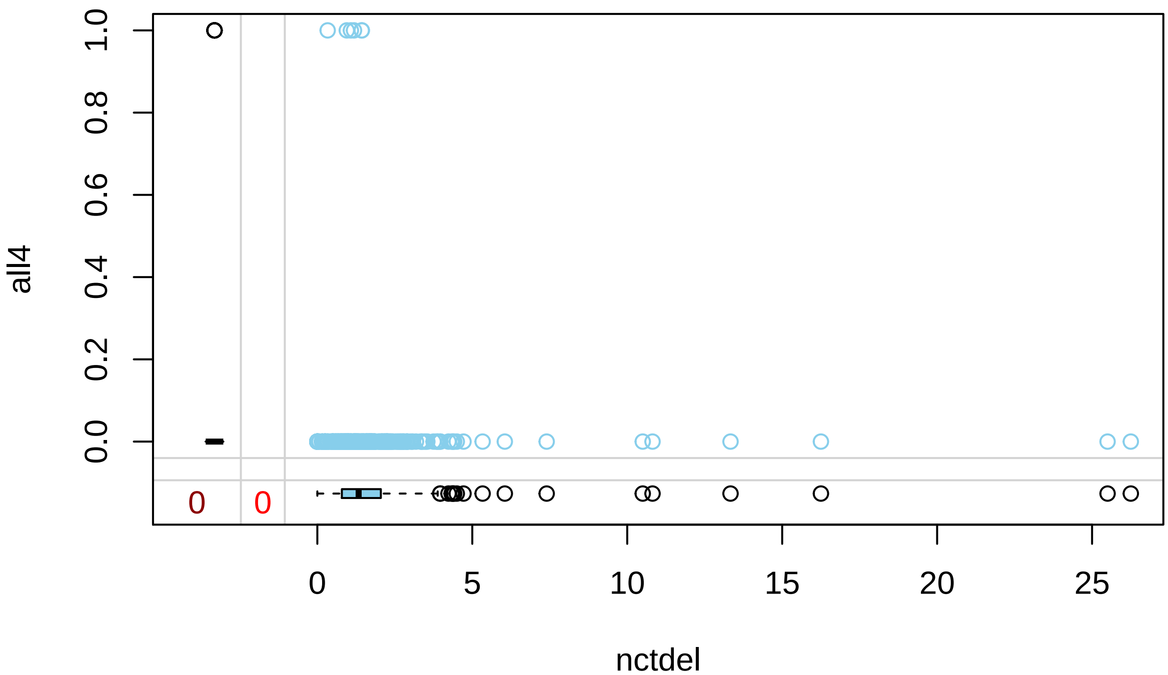
hisp
1.0
0.8
0.6
0.4
0.2
0.0
0
0
0
5
10
15
20
25
nctdel



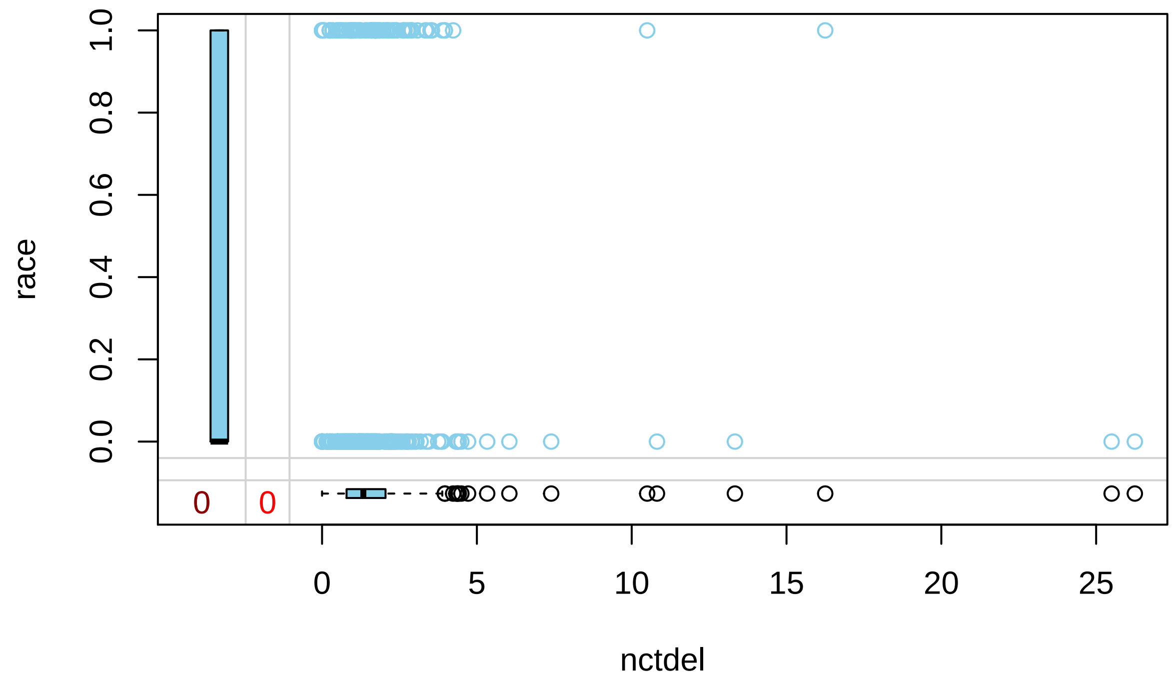
sn1
1.0
0.8
0.6
0.4
0.2
0.0
0
0
0
5
10
15
20
25
nctdel


sn2
1.0
0.8
0.6
0.4
0.2
0.0
0
0
0
5
10
15
20
25
nctdel


sn3
0.0
0.2
0.4
0.6
0.8
1.0
0
0
0
5
10
15
20
25
nctdel



all4
1.0
0.8
0.6
0.4
0.2
0.0
0
0
0
5
10
15
20
25
nctdel



##Interpretation

	Deviance p-value
OLS	2e-04
LASSO Penalty	3e-04
Ridge Penalty	0e+00
Kernels	0e+00

	Lower	Upper
symptom0	-1.2348	0.1283
symptom1	-0.8128	0.4192
symptom2	-0.9683	0.3673
raceother	-0.2452	0.4814
male	-0.6261	0.0439
X1	-1.6083	-0.2653
X2	-0.1101	1.2464
X3	0.1159	1.6606
X4	-0.6474	1.2667
X5	-1.0553	1.3369
X6	-926.4905	958.4814

	LASSO Estimate
(Intercept)	0.0000

	LASSO Estimate
symptom0	0.0000
symptom1	0.0000
symptom2	0.0000
raceother	0.0000
male	0.0000
X1	-1.0788
X2	0.0347
X3	0.1736
X4	0.0000
X5	0.0000
X6	0.9557

	Ridge Estimate
(Intercept)	0.0000
symptom0	-0.1646
symptom1	-0.0401
symptom2	-0.0893
raceother	-0.0588
male	-0.1393
X1	-0.5499
X2	0.2039
X3	0.3068
X4	0.0513
X5	-0.0139
X6	0.9175

	Lower	Upper
symptom0	-1.3827	-0.0953
symptom1	-0.9360	0.2222
symptom2	-1.0734	0.1903
raceother	-0.2915	0.4013
male	-0.5736	0.0674
k1	-5.6259	-0.1256
k2	5.8663	13.2598

Discussion

why nothing is significant:

```
## # A tibble: 4 x 6
##   symptom     mean     n        sd    lower    upper
##     <chr>    <dbl> <int>     <dbl>    <dbl>    <dbl>
## 1       0 1.560370    45 0.8675425 1.306892 1.813849
## 2       1 1.547995   133 0.7804779 1.415350 1.680640
## 3       2 1.618750    56 0.7784150 1.414871 1.822629
## 4      3+ 1.493333    25 0.6746227 1.228881 1.757785
```

```
## # A tibble: 2 x 3
##   gender     mean   median
##    <chr>    <dbl>    <dbl>
## 1 female 1.516541 1.433333
## 2   male 1.606217 1.566667
```

```
## # A tibble: 2 x 3
##                 race     mean   median
##                <chr>    <dbl>    <dbl>
## 1 Black or Hispanic 1.727556 1.716667
## 2             Other 1.491938 1.383333
```

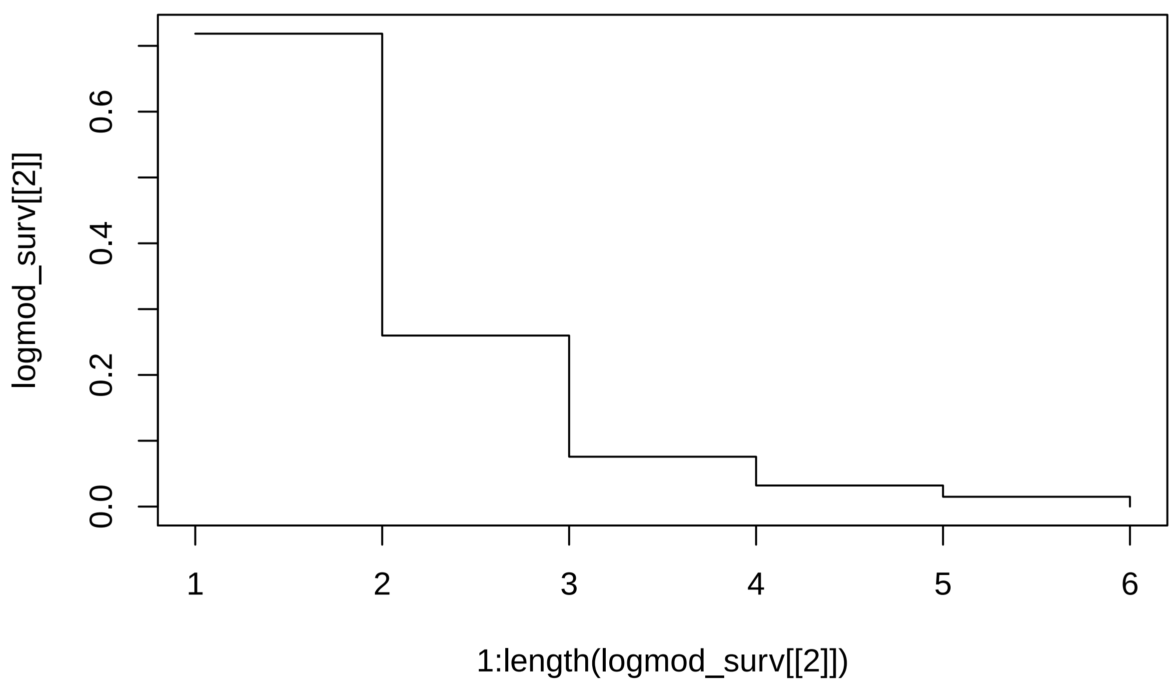
References

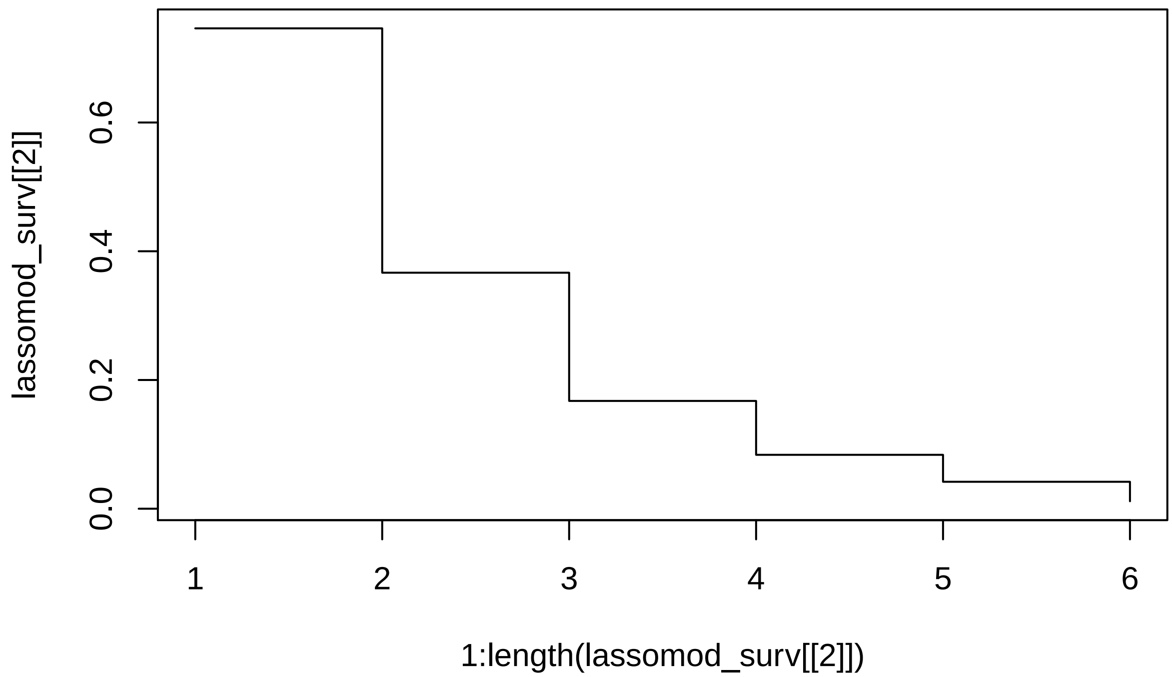
https://www.r-bloggers.com/imputing-missing-data-with-r-mice-package/

Credits

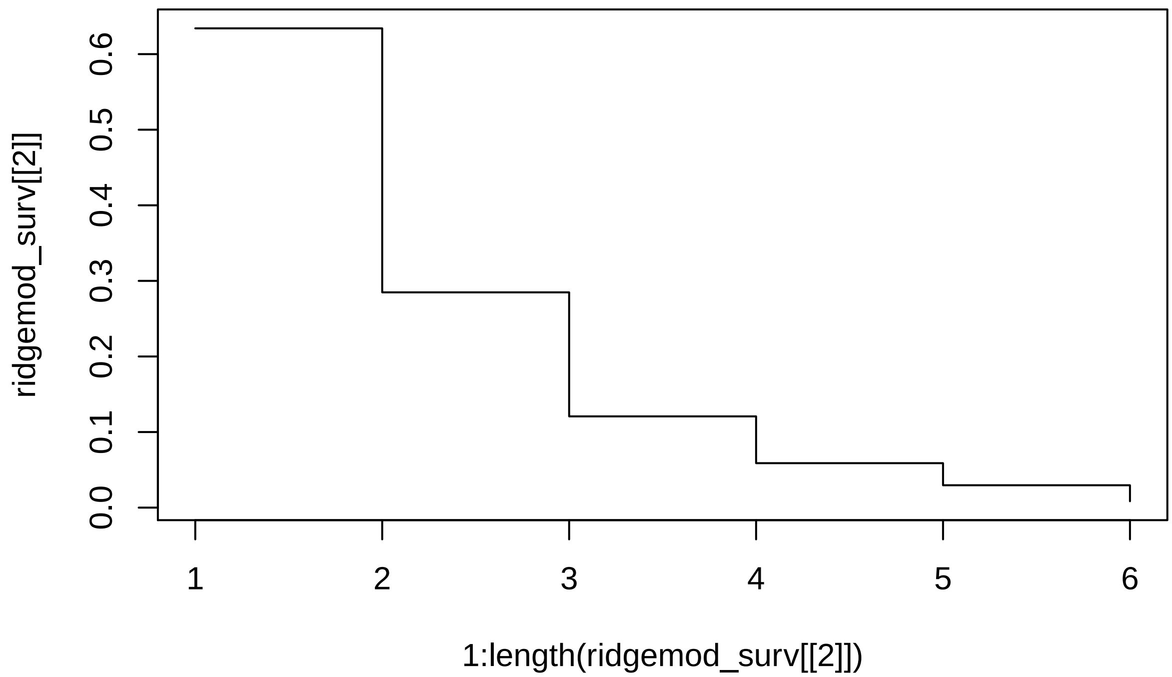
Survival Curves:

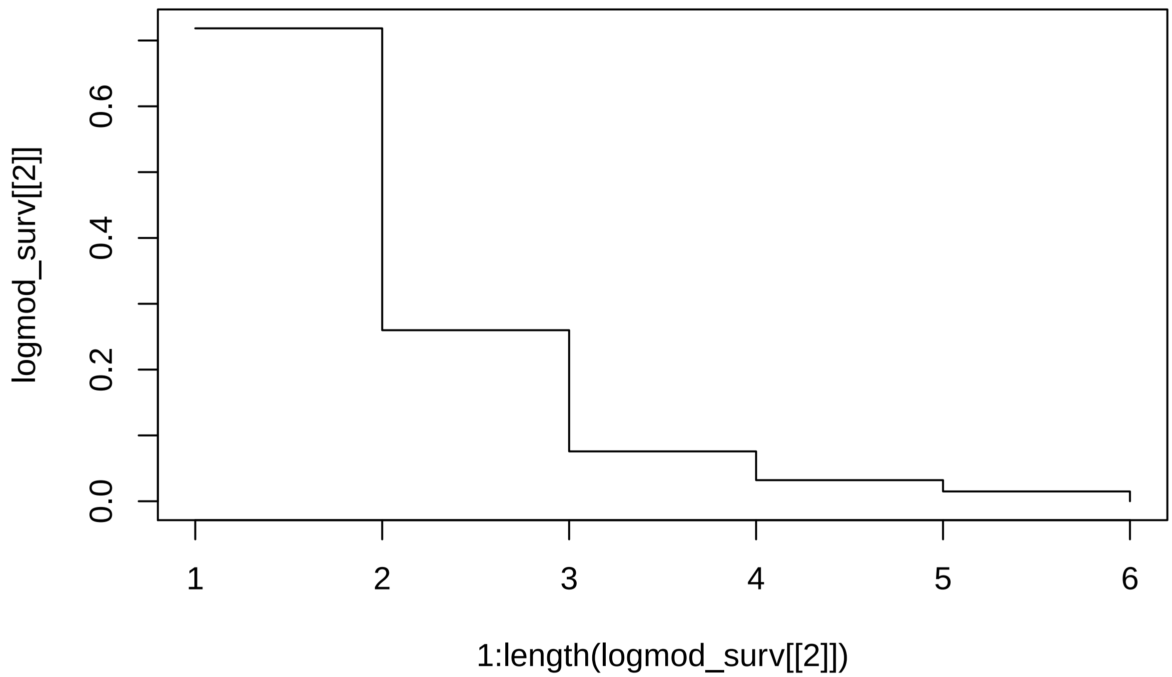
Question for Jonathan: how do we plot a survival curve from glm???



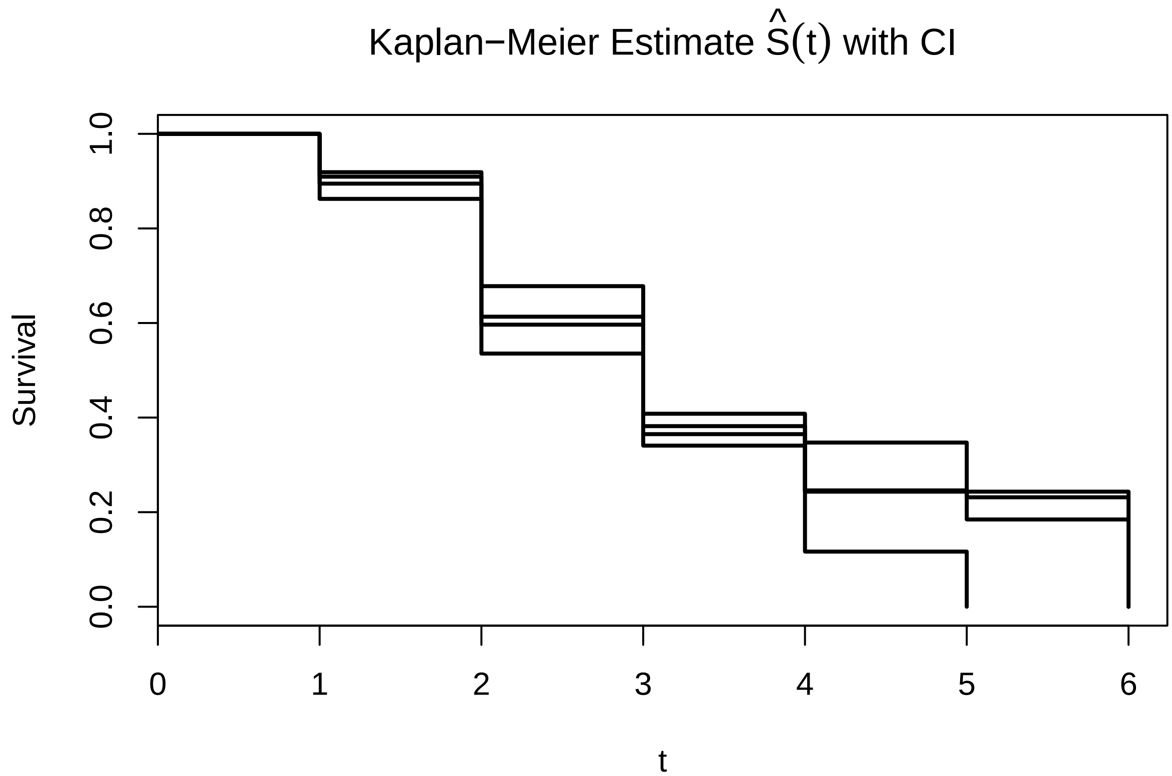


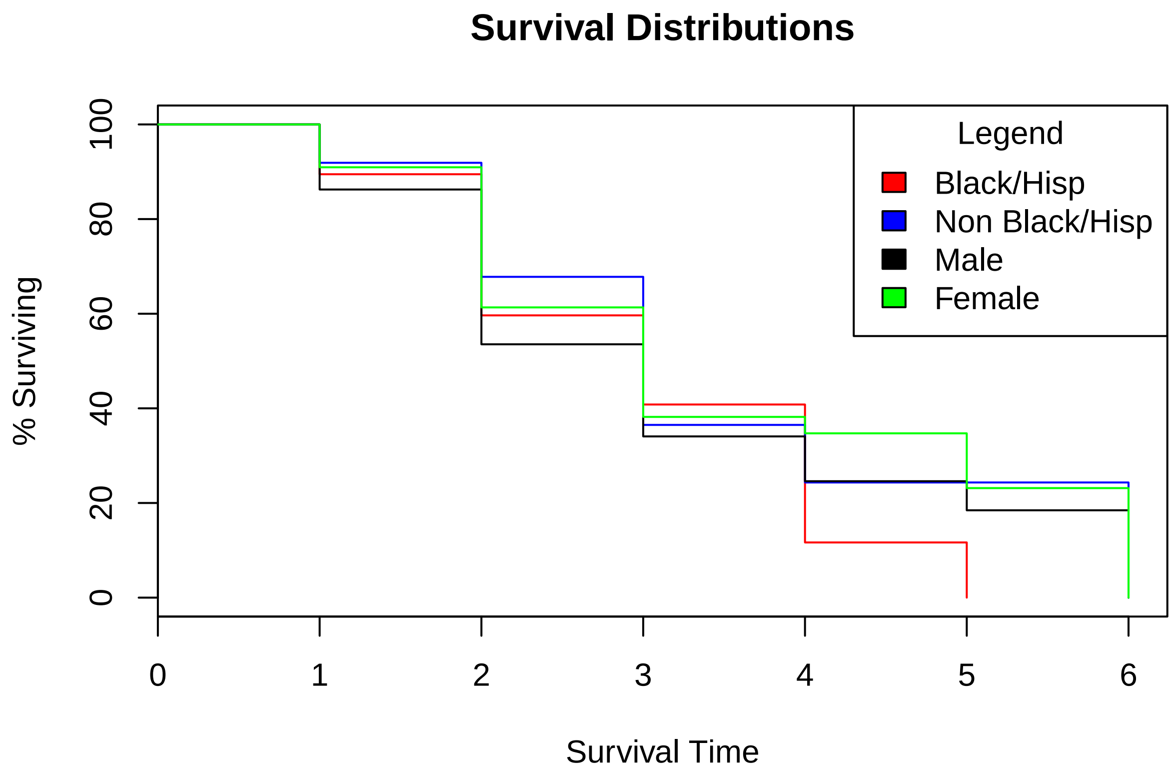
lassomod_surv[[2]]
0.0
0.2
0.4
0.6
1
2
3
4
5
6
1:length(lassomod_surv[[2]])


0.0
0.1
0.2
0.3
0.4
0.5
0.6
ridgemod_surv[[2]]
1
2
3
4
5
6
1:length(ridgemod_surv[[2]])



logmod_surv[[2]]
0.0
0.2
0.4
0.6
1
2
3
4
5
6
1:length(logmod_surv[[2]])


Kaplan−Meier Estimate Ŝ(t) with CI
Survival
0.0
0.2
0.4
0.6
0.8
1.0
0
1
2
3
4
5
6
t




```
## Call:
## survdiff(formula = Surv(timecat, fail) ~ raceother + male, data = datcat_X)
##
##                          N Observed Expected (O-E)^2/E (O-E)^2/V
## raceother=0, male=0  95       38     35.6     0.164      0.26
## raceother=0, male=1 111       42     47.7     0.675      1.17
## raceother=1, male=0 240      103     91.2     1.517      3.20
## raceother=1, male=1 243       94    102.5     0.706      1.60
##
##  Chisq= 4.4  on 3 degrees of freedom, p= 0.225
```