

# Classification of Acute Respiratory Distress Syndrome

*Robert Edwards*

## Data Preparation

### Scaling the continuous variables

### Dropping NAs from dataset

Dropping variables with >40% NAs (PreECMO\_Albumin)

### Train / Validation / Test

Call:

```
glm(formula = ECMO_Survival ~ . - Pt_ID, family = binomial(link = "logit"),
    data = train.ARDS)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.36813	0.01364	0.38692	0.69421	1.70684

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )	
(Intercept)	2.06639	0.61058	3.384	0.000714	***
Genderw	-0.06459	0.56055	-0.115	0.908268	
Indication	-0.09829	0.15104	-0.651	0.515206	
Duration_ECMO	-0.17506	0.25882	-0.676	0.498793	
Age	-0.33419	0.29714	-1.125	0.260712	
PreECMO_RR	-0.49082	0.24983	-1.965	0.049461	*
PreECMO_Vt	0.01795	0.26295	0.068	0.945568	
PreECMO_FiO2	0.52923	0.52941	1.000	0.317471	
PreECMO_Ppeak	-0.26456	0.35289	-0.750	0.453442	
PreECMO_Pmean	0.43753	0.48817	0.896	0.370111	
PreECMO_PEEP	0.20183	0.42513	0.475	0.634966	
PreECMO_PF	1.49079	0.83594	1.783	0.074527	.
PreECMO_SpO2	-0.91677	0.47902	-1.914	0.055640	.
PreECMO_PaCO2	0.44623	0.63171	0.706	0.479953	
PreECMO_pH	0.89219	0.79236	1.126	0.260171	
PreECMO_BE	-0.68108	0.65022	-1.047	0.294890	
PreECMO_Lactate	-0.61473	0.39306	-1.564	0.117827	
PreECMO_NAdose	-0.11791	0.25316	-0.466	0.641392	
PreECMO_MAP	0.07270	0.29291	0.248	0.803977	
PreECMO_Creatinine	0.02676	0.28414	0.094	0.924962	
PreECMO_Urea	-0.22805	0.34557	-0.660	0.509295	

```

PreECMO_CK      0.08294    0.24158    0.343 0.731348
PreECMO_Bilirubin 0.82880    0.55468    1.494 0.135127
PreECMO_CRP     -0.28476    0.35144   -0.810 0.417778
PreECMO_Fibrinogen -0.20003    0.31744   -0.630 0.528603
PreECMO_Ddimer  -0.13960    0.29423   -0.474 0.635175
PreECMO_ATIII   -0.04830    0.17355   -0.278 0.780781
PreECMO_Leukocytes 0.04779    0.24911    0.192 0.847868
PreECMO_Platelets 0.45217    0.35559    1.272 0.203515
PreECMO_TNFa     1.69276    1.10178    1.536 0.124443
PreECMO_IL6     -1.39933    0.75665   -1.849 0.064405 .
PreECMO_IL8      0.84094    0.46438    1.811 0.070158 .
PreECMO_siIL2   -0.52497    0.38638   -1.359 0.174240
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

(Dispersion parameter for binomial family taken to be 1)

```

Null deviance: 170.92 on 155 degrees of freedom
Residual deviance: 130.32 on 123 degrees of freedom
AIC: 196.32

```

Number of Fisher Scoring iterations: 7

A quick note about the `plogis` function: The `glm()` procedure with `family="binomial"` will build the logistic regression model on the given formula. When we use the `predict` function on this model, it will predict the  $\log(\text{odds})$  of the Y variable. This is not what we ultimately want because, the predicted values may not lie within the 0 and 1 range as expected. So, to convert it into prediction probability scores that is bound between 0 and 1, we use the `plogis()`. For more info see (blog on logistic regression)[<http://r-statistics.co/Logistic-Regression-With-R.html>].

## Optimal Prediction Probability Cutoff for the Model

The default cutoff prediction probability score is 0.5 or the ratio of — in the training data. But sometimes, tuning the probability cutoff can improve the accuracy in both the development and validation samples. The `InformationValue::optimalCutoff` function provides ways to find the optimal cutoff to improve the prediction of —, —, both — and — and o reduce the misclassification error. Let's compute the optimal score that minimizes the misclassification error for the above model.

## Model Diagnostics

```

      N      Y
0.2222222 0.9000000

```

## Logistic Regression LASSO

```
[1] "Iteration 1 of 11"  
[1] "Iteration 2 of 11"  
[1] "Iteration 3 of 11"  
[1] "Iteration 4 of 11"  
[1] "Iteration 5 of 11"  
[1] "Iteration 6 of 11"  
[1] "Iteration 7 of 11"  
[1] "Iteration 8 of 11"  
[1] "Iteration 9 of 11"  
[1] "Iteration 10 of 11"  
[1] "Iteration 11 of 11"
```

	nets.a	nets.mse
[1,]	0.0	0
[2,]	0.1	0
[3,]	0.2	0
[4,]	0.3	0
[5,]	0.4	0
[6,]	0.5	0
[7,]	0.6	0
[8,]	0.7	0
[9,]	0.8	0
[10,]	0.9	0
[11,]	1.0	0

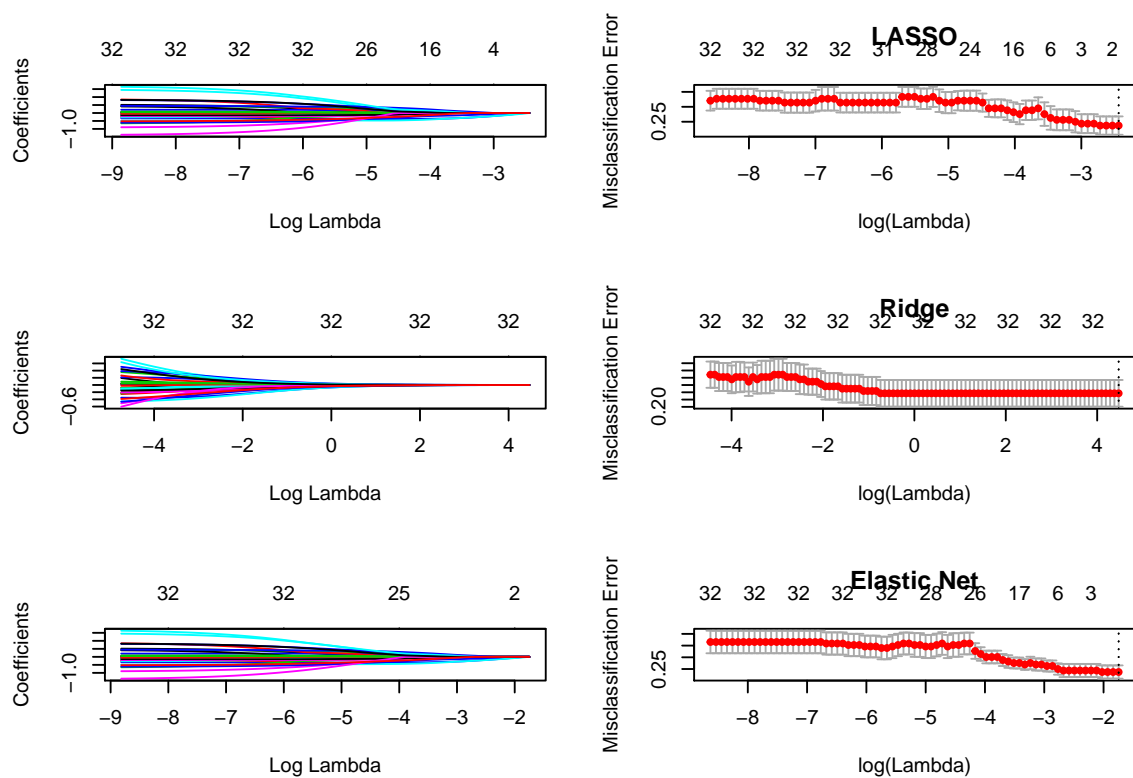


Figure 1: Lasso, Ridge, Elastic Net shrinkage and lambda cross-validation.

34 x 1 sparse Matrix of class "dgCMatrix"

```

      1
(Intercept) 1.168206
(Intercept) .
Genderw     .
Indication  .
Duration_ECMO .
Age         .
PreECMO_RR  .
PreECMO_Vt  .
PreECMO_FiO2 .
PreECMO_Ppeak .
PreECMO_Pmean .
PreECMO_PEEP .
PreECMO_PF  .
PreECMO_SpO2 .
PreECMO_PaCO2 .
PreECMO_pH  .
PreECMO_BE  .
PreECMO_Lactate .
PreECMO_NAdose .
PreECMO_MAP .
PreECMO_Creatinine .

```

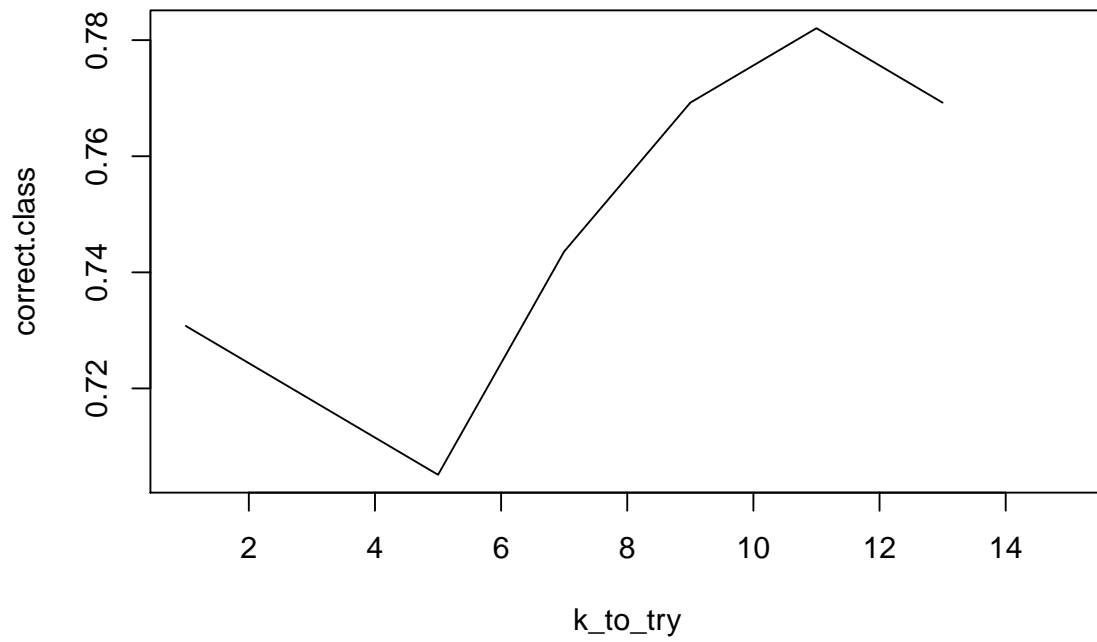
PreECMO_Urea	.
PreECMO_CK	.
PreECMO_Bilirubin	.
PreECMO_CRP	.
PreECMO_Fibrinogen	.
PreECMO_Ddimer	.
PreECMO_ATIII	.
PreECMO_Leukocytes	.
PreECMO_Platelets	.
PreECMO_TNFa	.
PreECMO_IL6	.
PreECMO_IL8	.
PreECMO_siIL2	.

## Model Diagnostics

N	Y
1.0	0.3

## K-Nearest Neighbors

```
knn.pred  1  2  
1  0  0  
2 18 60
```



## CVA

Call:

```
lda(train.data, grouping = train.class, prior = c(0.5, 0.5))
```

Prior probabilities of groups:

```
  1  2  
0.5 0.5
```

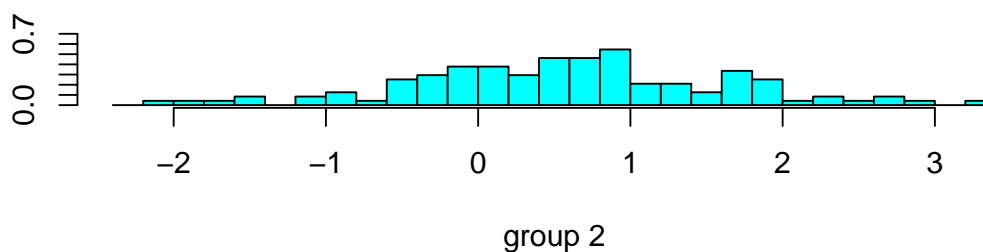
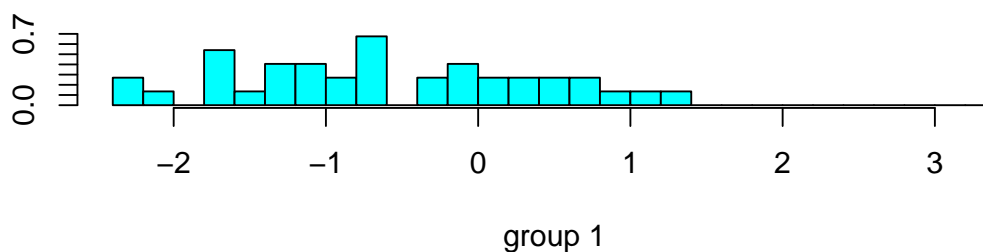
Group means:

	Duration_ECMO	Age	PreECMO_RR	PreECMO_Vt	PreECMO_FiO2
1	0.17011326	0.49130348	0.3613807	-0.14492842	0.1509813
2	0.07056348	0.04243535	-0.1421564	0.03175897	0.0658993
	PreECMO_Ppeak	PreECMO_Pmean	PreECMO_PEEP	PreECMO_PF	PreECMO_SpO2
1	0.04812984	-0.06664071	-0.07057527	-0.15494185	0.045583964
2	-0.14141295	-0.02782580	-0.03489447	-0.01992909	0.008969737
	PreECMO_PaCO2	PreECMO_pH	PreECMO_BE	PreECMO_Lactate	PreECMO_NAdose
1	0.04865712	-0.22205686	-0.08825413	0.25283223	0.07537849
2	-0.08852880	0.02194501	-0.05171488	-0.06990024	-0.06305472
	PreECMO_MAP	PreECMO_Creatinine	PreECMO_Urea	PreECMO_CK	PreECMO_Bilirubin
1	0.02183526	0.18529315	0.157008949	0.02465349	-0.07508043
2	0.09965918	0.09892448	0.005798352	0.09209170	0.05693255
	PreECMO_CRP	PreECMO_Fibrinogen	PreECMO_Ddimer	PreECMO_ATIII	
1	-0.04183654	0.0009114089	0.06401838	-0.05630207	
2	-0.05927163	-0.0642007345	-0.10890629	0.02885875	
	PreECMO_Leukocytes	PreECMO_Platelets	PreECMO_TNFa	PreECMO_IL6	
1	-0.07322120	-0.18851944	-0.12734461	-0.01846578	
2	-0.02137385	0.09582999	0.08328052	-0.06216531	
	PreECMO_IL8	PreECMO_siIL2			
1	-0.01379272	0.14233997			
2	0.09433882	-0.02785839			

Coefficients of linear discriminants:

	LD1
Duration_ECMO	0.007037672
Age	-0.296155173
PreECMO_RR	-0.483206778
PreECMO_Vt	0.094604126
PreECMO_FiO2	0.263373049
PreECMO_Ppeak	-0.241719223
PreECMO_Pmean	0.365599145
PreECMO_PEEP	0.208647282
PreECMO_PF	0.725524096
PreECMO_SpO2	-0.542669349
PreECMO_PaCO2	-0.047073672
PreECMO_pH	0.235062078
PreECMO_BE	-0.230177485
PreECMO_Lactate	-0.511091759
PreECMO_NAdose	-0.065984421
PreECMO_MAP	-0.013183977
PreECMO_Creatinine	0.102037474

PreECMO_Urea	-0.233670194
PreECMO_CK	0.061939746
PreECMO_Bilirubin	0.385837910
PreECMO_CRP	-0.032410050
PreECMO_Fibrinogen	-0.170151665
PreECMO_Ddimer	-0.141031272
PreECMO_ATIII	0.013650185
PreECMO_Leukocytes	0.071345021
PreECMO_Platelets	0.313399862
PreECMO_TNFa	0.246917241
PreECMO_IL6	-0.559264247
PreECMO_IL8	0.417083224
PreECMO_siIL2	-0.418665260



```
[1] "class"      "posterior" "x"
```

```
valid.class  1  2
              1  8 10
              2 22 38
```

	LD1
Duration_ECMO	0.007037672
Age	-0.296155173
PreECMO_RR	-0.483206778
PreECMO_Vt	0.094604126
PreECMO_FiO2	0.263373049



PreECMO_Ppeak	-0.241719223
PreECMO_Pmean	0.365599145
PreECMO_PEEP	0.208647282
PreECMO_PF	0.725524096
PreECMO_SpO2	-0.542669349
PreECMO_PaCO2	-0.047073672
PreECMO_pH	0.235062078
PreECMO_BE	-0.230177485
PreECMO_Lactate	-0.511091759
PreECMO_NAdose	-0.065984421
PreECMO_MAP	-0.013183977
PreECMO_Creatinine	0.102037474
PreECMO_Urea	-0.233670194
PreECMO_CK	0.061939746
PreECMO_Bilirubin	0.385837910
PreECMO_CRP	-0.032410050
PreECMO_Fibrinogen	-0.170151665
PreECMO_Ddimer	-0.141031272
PreECMO_ATIII	0.013650185
PreECMO_Leukocytes	0.071345021
PreECMO_Platelets	0.313399862
PreECMO_TNFa	0.246917241
PreECMO_IL6	-0.559264247
PreECMO_IL8	0.417083224
PreECMO_siIL2	-0.418665260

## LDA

Call:

```
lda(train.data, grouping = train.class)
```

Prior probabilities of groups:

	1	2
	0.2371795	0.7628205

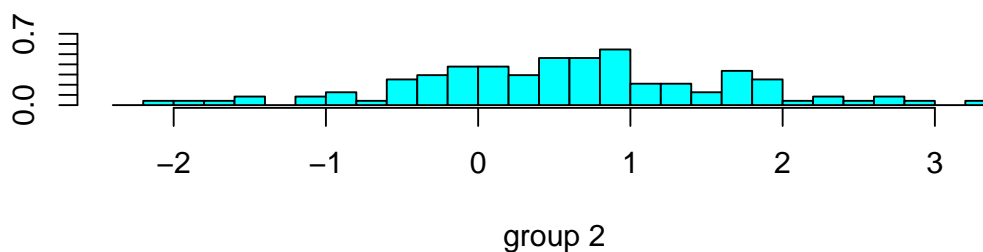
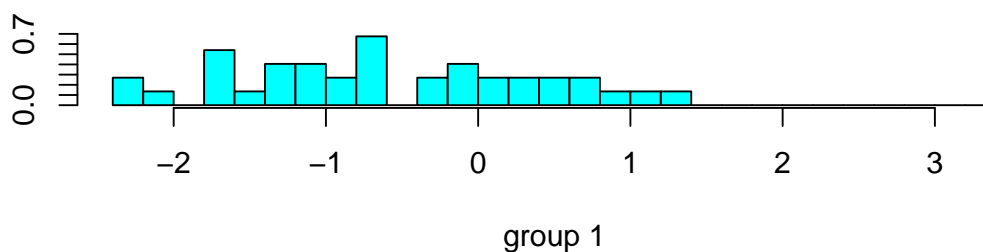
Group means:

	Duration_ECMO	Age	PreECMO_RR	PreECMO_Vt	PreECMO_FiO2
1	0.17011326	0.49130348	0.3613807	-0.14492842	0.1509813
2	0.07056348	0.04243535	-0.1421564	0.03175897	0.0658993
	PreECMO_Ppeak	PreECMO_Pmean	PreECMO_PEEP	PreECMO_PF	PreECMO_SpO2
1	0.04812984	-0.06664071	-0.07057527	-0.15494185	0.045583964
2	-0.14141295	-0.02782580	-0.03489447	-0.01992909	0.008969737
	PreECMO_PaCO2	PreECMO_pH	PreECMO_BE	PreECMO_Lactate	PreECMO_NAdose
1	0.04865712	-0.22205686	-0.08825413	0.25283223	0.07537849
2	-0.08852880	0.02194501	-0.05171488	-0.06990024	-0.06305472
	PreECMO_MAP	PreECMO_Creatinine	PreECMO_Urea	PreECMO_CK	PreECMO_Bilirubin
1	0.02183526	0.18529315	0.157008949	0.02465349	-0.07508043
2	0.09965918	0.09892448	0.005798352	0.09209170	0.05693255
	PreECMO_CRP	PreECMO_Fibrinogen	PreECMO_Ddimer	PreECMO_ATIII	
1	-0.04183654	0.0009114089	0.06401838	-0.05630207	
2	-0.05927163	-0.0642007345	-0.10890629	0.02885875	
	PreECMO_Leukocytes	PreECMO_Platelets	PreECMO_TNFa	PreECMO_IL6	
1	-0.07322120	-0.18851944	-0.12734461	-0.01846578	
2	-0.02137385	0.09582999	0.08328052	-0.06216531	
	PreECMO_IL8	PreECMO_siIL2			
1	-0.01379272	0.14233997			
2	0.09433882	-0.02785839			

Coefficients of linear discriminants:

	LD1
Duration_ECMO	0.007037672
Age	-0.296155173
PreECMO_RR	-0.483206778
PreECMO_Vt	0.094604126
PreECMO_FiO2	0.263373049
PreECMO_Ppeak	-0.241719223
PreECMO_Pmean	0.365599145
PreECMO_PEEP	0.208647282
PreECMO_PF	0.725524096
PreECMO_SpO2	-0.542669349
PreECMO_PaCO2	-0.047073672
PreECMO_pH	0.235062078
PreECMO_BE	-0.230177485
PreECMO_Lactate	-0.511091759
PreECMO_NAdose	-0.065984421
PreECMO_MAP	-0.013183977
PreECMO_Creatinine	0.102037474

PreECMO_Urea	-0.233670194
PreECMO_CK	0.061939746
PreECMO_Bilirubin	0.385837910
PreECMO_CRP	-0.032410050
PreECMO_Fibrinogen	-0.170151665
PreECMO_Ddimer	-0.141031272
PreECMO_ATIII	0.013650185
PreECMO_Leukocytes	0.071345021
PreECMO_Platelets	0.313399862
PreECMO_TNFa	0.246917241
PreECMO_IL6	-0.559264247
PreECMO_IL8	0.417083224
PreECMO_siIL2	-0.418665260



```
[1] "class"      "posterior" "x"
```

```
valid.class  1  2
              1  2 16
              2  6 54
```

	LD1
Duration_ECMO	0.007037672
Age	-0.296155173
PreECMO_RR	-0.483206778
PreECMO_Vt	0.094604126
PreECMO_FiO2	0.263373049

PreECMO_Ppeak	-0.241719223
PreECMO_Pmean	0.365599145
PreECMO_PEEP	0.208647282
PreECMO_PF	0.725524096
PreECMO_SpO2	-0.542669349
PreECMO_PaCO2	-0.047073672
PreECMO_pH	0.235062078
PreECMO_BE	-0.230177485
PreECMO_Lactate	-0.511091759
PreECMO_NAdose	-0.065984421
PreECMO_MAP	-0.013183977
PreECMO_Creatinine	0.102037474
PreECMO_Urea	-0.233670194
PreECMO_CK	0.061939746
PreECMO_Bilirubin	0.385837910
PreECMO_CRP	-0.032410050
PreECMO_Fibrinogen	-0.170151665
PreECMO_Ddimer	-0.141031272
PreECMO_ATIII	0.013650185
PreECMO_Leukocytes	0.071345021
PreECMO_Platelets	0.313399862
PreECMO_TNFa	0.246917241
PreECMO_IL6	-0.559264247
PreECMO_IL8	0.417083224
PreECMO_siIL2	-0.418665260

## QDA

Call:

```
qda(train.data, grouping = train.class)
```

Prior probabilities of groups:

```
      1      2
0.2371795 0.7628205
```

Group means:

```
      Duration_ECMO      Age PreECMO_RR PreECMO_Vt PreECMO_FiO2
1      0.17011326 0.49130348 0.3613807 -0.14492842 0.1509813
2      0.07056348 0.04243535 -0.1421564 0.03175897 0.0658993
      PreECMO_Ppeak PreECMO_Pmean PreECMO_PEEP PreECMO_PF PreECMO_SpO2
1      0.04812984 -0.06664071 -0.07057527 -0.15494185 0.045583964
2     -0.14141295 -0.02782580 -0.03489447 -0.01992909 0.008969737
      PreECMO_PaCO2 PreECMO_pH PreECMO_BE PreECMO_Lactate PreECMO_NAdose
1      0.04865712 -0.22205686 -0.08825413 0.25283223 0.07537849
2     -0.08852880 0.02194501 -0.05171488 -0.06990024 -0.06305472
      PreECMO_MAP PreECMO_Creatinine PreECMO_Urea PreECMO_CK PreECMO_Bilirubin
1      0.02183526 0.18529315 0.157008949 0.02465349 -0.07508043
2      0.09965918 0.09892448 0.005798352 0.09209170 0.05693255
      PreECMO_CRP PreECMO_Fibrinogen PreECMO_Ddimer PreECMO_ATIII
1     -0.04183654 0.0009114089 0.06401838 -0.05630207
2     -0.05927163 -0.0642007345 -0.10890629 0.02885875
      PreECMO_Leukocytes PreECMO_Platelets PreECMO_TNFa PreECMO_IL6
1      -0.07322120 -0.18851944 -0.12734461 -0.01846578
2      -0.02137385 0.09582999 0.08328052 -0.06216531
      PreECMO_IL8 PreECMO_siIL2
1     -0.01379272 0.14233997
2      0.09433882 -0.02785839

[1] "class" "posterior"
```

```
valid.class 1 2
      1 0 18
      2 2 58
```

```
      1      2
81 2.145894e-05 0.9999785411
101 5.025954e-86 1.0000000000
247 8.285293e-25 1.0000000000
143 9.998830e-01 0.0001170331
112 9.411744e-48 1.0000000000
379 5.754561e-02 0.9424543895
87 6.034413e-32 1.0000000000
438 1.564350e-78 1.0000000000
349 0.000000e+00 1.0000000000
108 3.564849e-38 1.0000000000
341 5.663892e-35 1.0000000000
353 2.028043e-24 1.0000000000
344 1.966669e-109 1.0000000000
```

442	5.437669e-26	1.0000000000
355	3.724644e-59	1.0000000000
315	3.794158e-10	0.9999999996
240	4.024391e-236	1.0000000000
260	1.125393e-11	1.0000000000
357	3.660473e-13	1.0000000000
406	8.530698e-01	0.1469302134

NULL

## Trees

n= 156

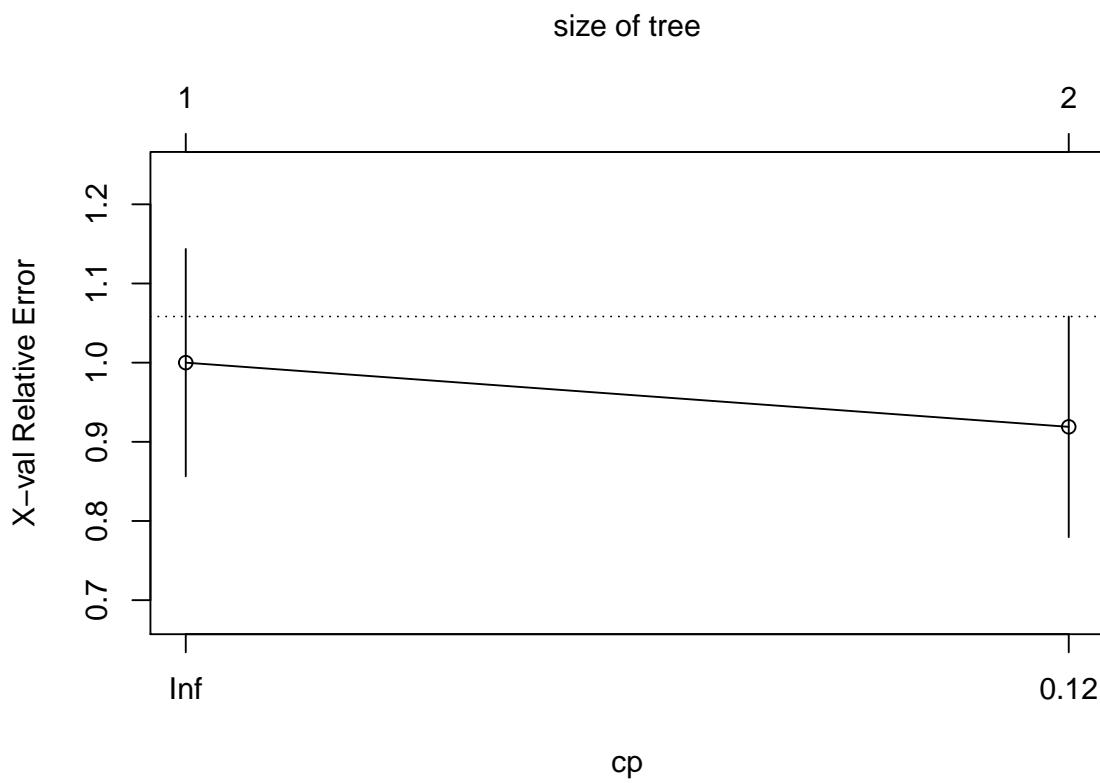
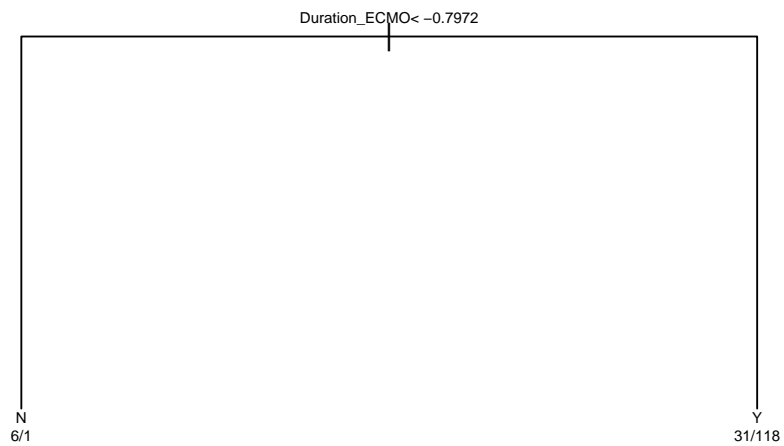
node), split, n, loss, yval, (yprob)

\* denotes terminal node

1) root 156 37 Y (0.2371795 0.7628205)

2) Duration\_ECMO< -0.7972145 7 1 N (0.8571429 0.1428571) \*

3) Duration\_ECMO>=-0.7972145 149 31 Y (0.2080537 0.7919463) \*





Classification tree:

```
rpart(formula = ECMO_Survival ~ ., data = train.ARDS[, -1], method = "class",
      control = rpart.control(minsplit = 1, minbucket = 1, cp = 0.1))
```

Variables actually used in tree construction:

```
[1] Duration_ECMO
```

Root node error: 37/156 = 0.23718

n= 156

	CP	nsplit	rel error	xerror	xstd
1	0.13514	0	1.00000	1.00000	0.14359
2	0.10000	1	0.86486	0.91892	0.13937

pred.valid.tree		
valid.class	N	Y
1	2	16
2	0	60

## Random Forest

Call:

```
randomForest(formula = ECMO_Survival ~ ., data = train.ARDS[, -1])
```

Type of random forest: classification

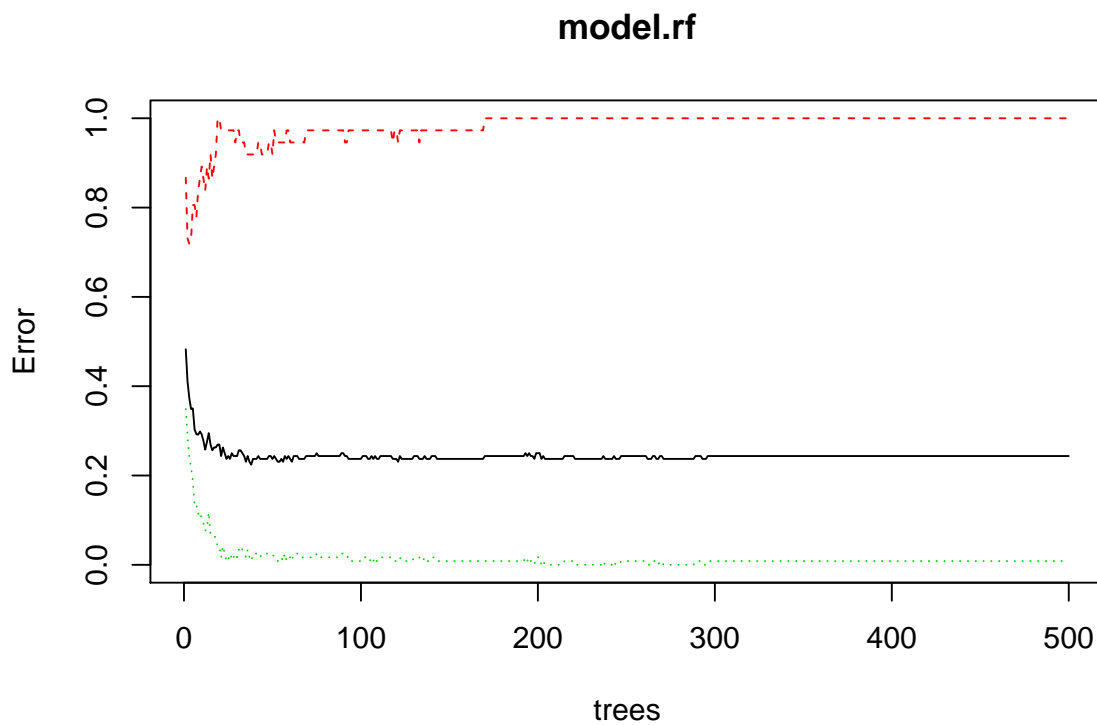
Number of trees: 500

No. of variables tried at each split: 5

OOB estimate of error rate: 24.36%

Confusion matrix:

	N	Y	class.error
N	0	37	1.000000000
Y	1	118	0.008403361



	pred.valid.rf		
valid.class	N	Y	
1	1	17	
2	0	60	

## Support Vector Machines

Call:

```
svm(formula = ECMO_Survival ~ ., data = train.ARDS[, -1], kernel = "polynomial",  
    cost = 10, scale = FALSE)
```

Parameters:

```
  SVM-Type:  C-classification  
 SVM-Kernel: polynomial  
      cost:  10  
    degree:  3  
     gamma:  0.03030303  
    coef.0:  0
```

Number of Support Vectors: 120

		pred.valid.svm	
valid.class	N	Y	
	1	3	15
	2	5	55

## Model Selection

```
      err.logit err.lasso  err.knn   err.cva   err.lda   err.qda   err.tree
[1,] 0.2564103 0.7692308 0.2307692 0.4102564 0.2820513 0.2564103 0.2051282
      err.rf    err.svm
[1,] 0.2179487 0.2564103
```