Lung Cancer

Jose Tamez

2023-04-28

1

Contents

1 Detailed Survival analyis of the Survival lung data.

	1.1 Exploring Raw Features with RRPlot 1.2 Reporting the Metrics 1.3 Modeling 1.4 Cox Model Performance 1.5 Cross-Validation	2 8 10 10 26
1	Detailed Survival analyis of the Survival lung data.	
	brary(survival) brary(FRESA.CAD)	
##	Loading required package: Rcpp	
##	Loading required package: stringr	
##	Loading required package: miscTools	
##	Loading required package: Hmisc	
## ##	Attaching package: 'Hmisc'	
## ##		
##	1 2 1 2 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3	
	Loading required package: pROC	
##	Type 'citation("pROC")' for a citation.	
## ##	Attaching package: 'pROC'	
	The following objects are masked from 'package:stats':	
##		
pa	<pre><- par(no.readonly = TRUE) nder::panderOptions('digits', 3) nder::panderOptions('keep.trailing.zeros',TRUE)</pre>	
da ⁻	ta(lung)	
##	Warning in data(lung): data set 'lung' not found	

```
lung$inst <- NULL
lung$status <- lung$status - 1
lung <- lung[complete.cases(lung),]
pander::pander(table(lung$status))</pre>
```

0	1
47	121

pander::pander(summary(lung\$time))

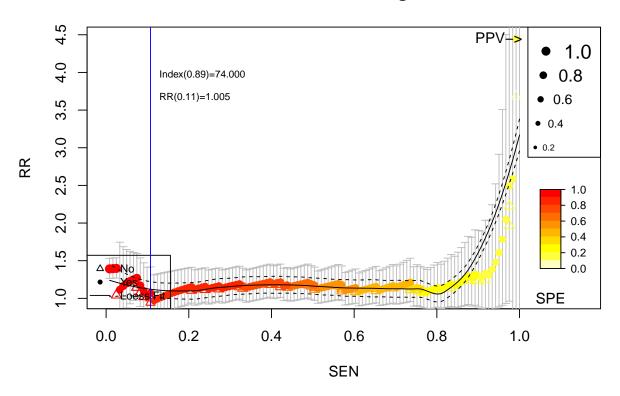
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
5	175	268	310	416	1022

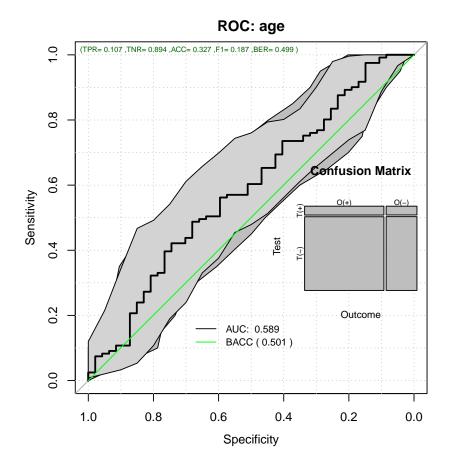
$1.1 \quad \text{Exploring Raw Features with RRPlot}$

```
convar <- colnames(lung)[lapply(apply(lung,2,unique),length) > 10]
convar <- convar[convar != "time"]
topvar <- univariate_BinEnsemble(lung[,c("status",convar)],"status")
pander::pander(topvar)</pre>
```

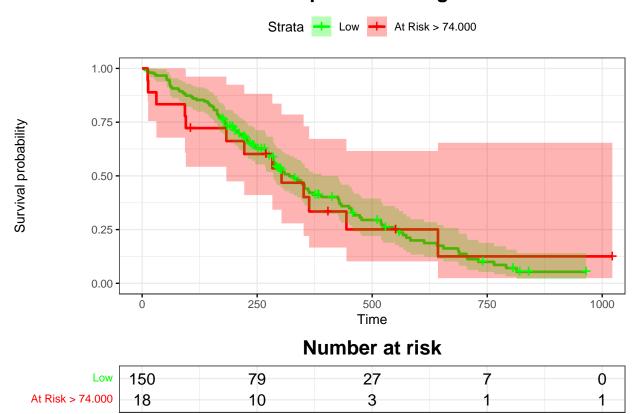
age	wt.loss
0.106	0.106

Relative Risk: age

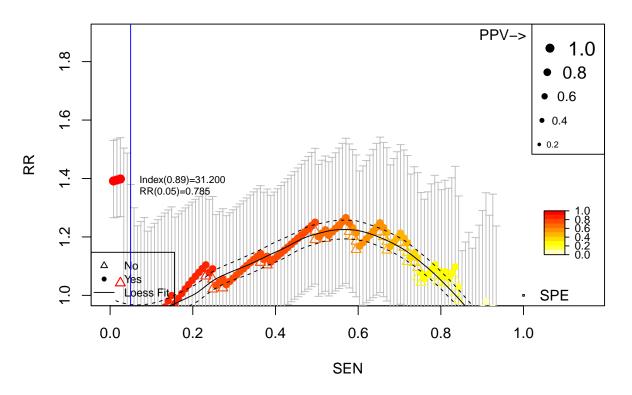


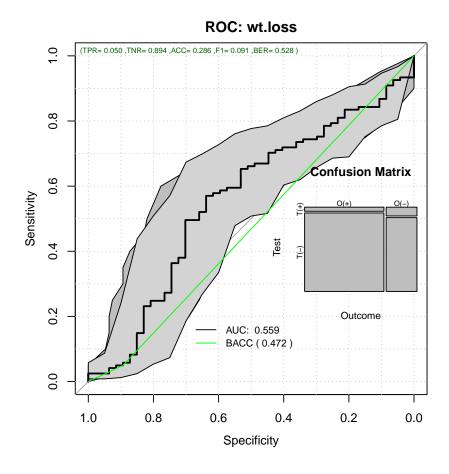


Kaplan-Meier: age



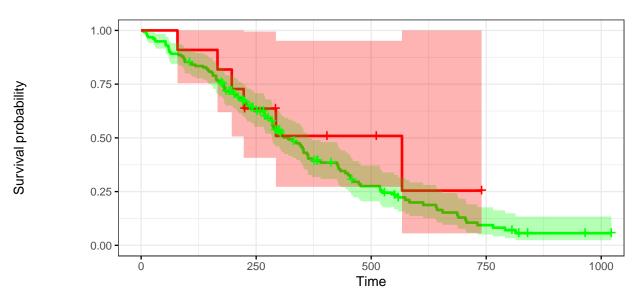
Relative Risk: wt.loss





Kaplan-Meier: wt.loss





Number at risk

Low	157	83	27	8	1
At Risk > 31.200	11	6	3	0	0

names(RRanalysis) <- topFive</pre>

1.2 Reporting the Metrics

```
ROCAUC <- NULL
CstatCI <- NULL
RRatios <- NULL
LogRangp <- NULL
Sensitivity <- NULL
Specificity <- NULL
for (topf in topFive)
  CstatCI <- rbind(CstatCI,RRanalysis[[topf]]$c.index$cstatCI)</pre>
  RRatios <- rbind(RRatios,RRanalysis[[topf]]$RR_atP)</pre>
  LogRangp <- rbind(LogRangp,RRanalysis[[topf]]$surdif$pvalue)</pre>
  Sensitivity <- rbind(Sensitivity,RRanalysis[[topf]]$ROCAnalysis$sensitivity)</pre>
  Specificity <- rbind(Specificity,RRanalysis[[topf]]$ROCAnalysis$specificity)</pre>
  ROCAUC <- rbind(ROCAUC,RRanalysis[[topf]]$ROCAnalysis$aucs)</pre>
rownames(CstatCI) <- topFive</pre>
rownames(RRatios) <- topFive</pre>
rownames(LogRangp) <- topFive</pre>
rownames(Sensitivity) <- topFive</pre>
rownames(Specificity) <- topFive</pre>
```

rownames(ROCAUC) <- topFive</pre>

pander::pander(ROCAUC)

	est	lower	upper
age	0.589	0.491	0.686
wt.loss	0.559	0.460	0.658

pander::pander(CstatCI)

	mean.C Index		lower	upper
age	0.558	0.557	0.494	0.618
wt.loss	0.513	0.513	0.449	0.570

pander::pander(RRatios)

	est	lower	upper
$egin{array}{c} { m age} \\ { m wt.loss} \end{array}$	$1.005 \\ 0.785$	$0.744 \\ 0.462$	1.36 1.33

pander::pander(LogRangp)

age	0.859
wt.loss	0.358

pander::pander(Sensitivity)

	est	lower	upper
$egin{array}{c} { m age} \\ { m wt.loss} \end{array}$	$0.1074 \\ 0.0496$	$0.0585 \\ 0.0184$	$0.177 \\ 0.105$

pander::pander(Specificity)

	est	lower	upper
$egin{array}{c} { m age} \\ { m wt.loss} \end{array}$	$0.894 \\ 0.894$	$0.769 \\ 0.769$	$0.965 \\ 0.965$

meanMatrix <- cbind(ROCAUC[,1],CstatCI[,1],Sensitivity[,1],Specificity[,1],RRatios[,1])
colnames(meanMatrix) <- c("ROCAUC","C-Stat","Sen","Spe","RR")
pander::pander(meanMatrix)</pre>

	ROCAUC	C-Stat	Sen	Spe	RR
age	0.589	0.558	0.1074	0.894	1.005

	ROCAUC	C-Stat	Sen	Spe	RR
wt.loss	0.559	0.513	0.0496	0.894	0.785

1.3 Modeling

Table 11: Table continues below

	Estimate	lower	HR	upper	u.Accuracy	r.Accuracy
${ m ph.ecog}$	4.32 e-01	1.194	1.541	1.988	0.679	0.649
\mathbf{sex}	-4.59e-01	0.456	0.632	0.876	0.649	0.679
pat.karno	-1.77e-03	0.997	0.998	1.000	0.506	0.720
ph.karno	-4.64e-07	1.000	1.000	1.000	0.577	0.720
age	4.57e-08	1.000	1.000	1.000	0.565	0.720

Table 12: Table continues below

	full.Accuracy	u.AUC	r.AUC	full.AUC	IDI	NRI
ph.ecog	0.601	0.601	0.620	0.600	0.0449	0.405
sex	0.601	0.620	0.601	0.600	0.0285	0.478
pat.karno	0.506	0.585	0.500	0.585	0.0292	0.342
ph.karno	0.577	0.570	0.500	0.570	0.0143	0.280
age	0.565	0.549	0.500	0.549	0.0162	0.195

	z.IDI	z.NRI	Delta.AUC	Frequency
ph.ecog	3.33	2.48	-0.02005	1.0
sex	2.76	2.85	-0.00167	1.0
pat.karno	2.44	2.24	0.08546	1.0
ph.karno	2.22	1.64	0.06998	0.8
age	1.97	1.14	0.04871	0.1

1.4 Cox Model Performance

Here we evaluate the model using the RRPlot() function.

1.4.1 The evaluation of the raw Cox model with RRPlot()

Here we will use the predicted event probability assuming a baseline hazard for events withing 5 years

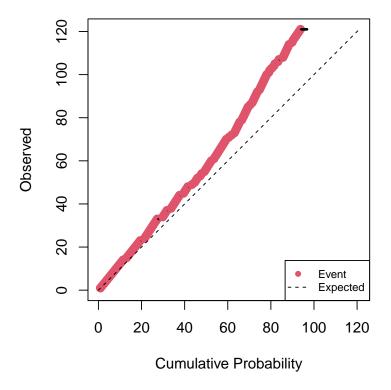
```
timeinterval <- 2*mean(subset(lung,status==1)$time)

h0 <- sum(lung$status & lung$time <= timeinterval)
h0 <- h0/sum((lung$time > timeinterval) | (lung$status==1))
pander::pander(t(c(h0=h0,timeinterval=timeinterval)),caption="Initial Parameters")
```

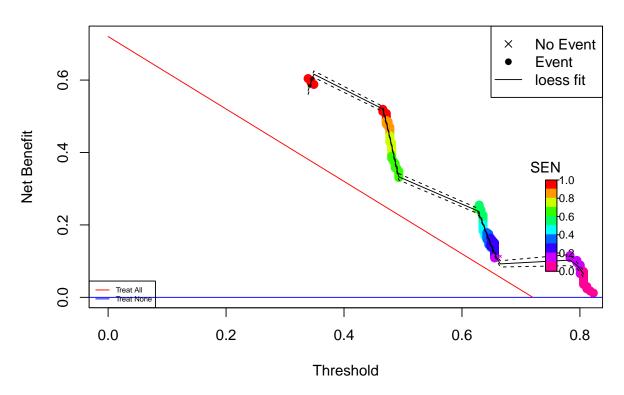
Table 14: Initial Parameters

h0	timeinterval
0.85	578

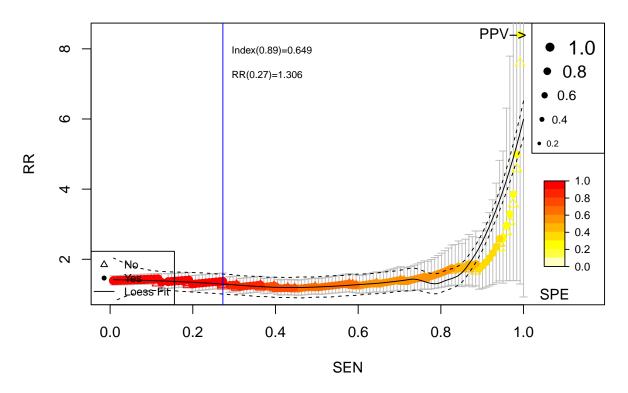
Cumulative vs. Observed: Raw Train: Lung Cancer



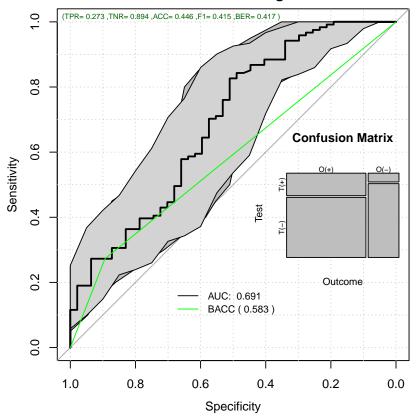
Decision Curve Analysis: Raw Train: Lung Cancer



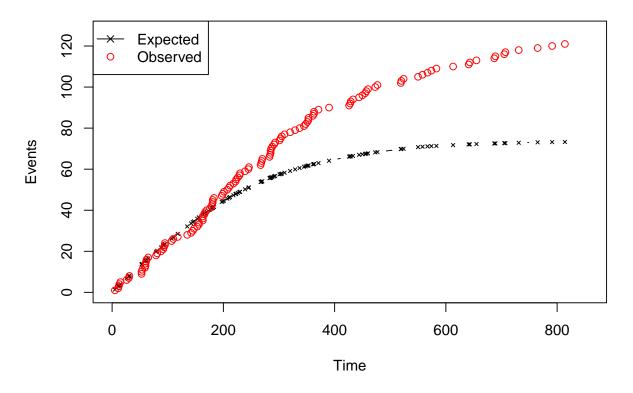
Relative Risk: Raw Train: Lung Cancer



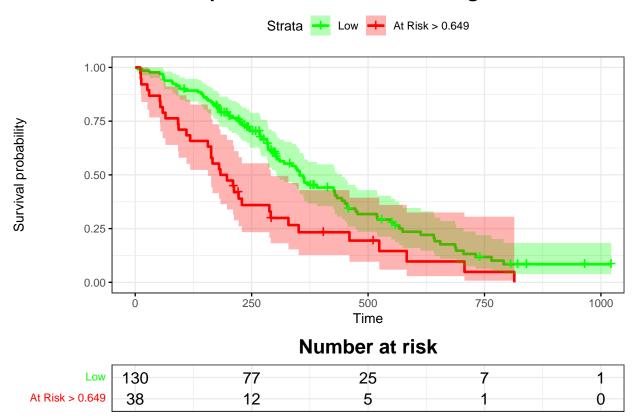
ROC: Raw Train: Lung Cancer



Time vs. Events: Raw Train: Lung Cancer



Kaplan-Meier: Raw Train: Lung Cancer



As we can see the Observed probability as well as the Time vs. Events are not calibrated.

1.4.2 Uncalibrated Performance Report

pander::pander(t(rrAnalysisTrain\$OERatio),caption="0/E Ratio")

Table 15: O/E Ratio

est	lower	upper
1.65	1.37	1.97

pander::pander(t(rrAnalysisTrain\$0E95ci),caption="0/E Ratio")

Table 16: O/E Ratio

mean	50%	2.5%	97.5%
1.21	1.21	1.16	1.25

pander::pander(t(rrAnalysisTrain\$OAcum95ci), caption="0/Acum Ratio")

Table 17: O/Acum Ratio

mean	50%	2.5%	97.5%
1.2	1.2	1.19	1.21

pander::pander(rrAnalysisTrain\$c.index\$cstatCI, caption="C. Index")

mean.C Index	median	lower	upper
0.651	0.651	0.586	0.714

pander::pander(t(rrAnalysisTrain\$ROCAnalysis\$aucs),caption="ROC AUC")

Table 19: ROC AUC

est	lower	upper
0.691	0.598	0.784

pander::pander((rrAnalysisTrain\$ROCAnalysis\$sensitivity), caption="Sensitivity")

Table 20: Sensitivity

est	lower	upper
0.273	0.196	0.361

pander::pander((rrAnalysisTrain\$ROCAnalysis\$specificity), caption="Specificity")

Table 21: Specificity

est	lower	upper
0.894	0.769	0.965

pander::pander(t(rrAnalysisTrain\$thr_atP),caption="Probability Thresholds")

Table 22: Probability Thresholds

90%	
0.649	

pander::pander(t(rrAnalysisTrain\$RR_atP),caption="Risk Ratio")

Table 23: Risk Ratio

est	lower	upper
1.31	1.11	1.54

pander::pander(rrAnalysisTrain\$surdif,caption="Logrank test")

Table 24: Logrank test Chisq = 10.879375 on 1 degrees of freedom, p = 0.000972

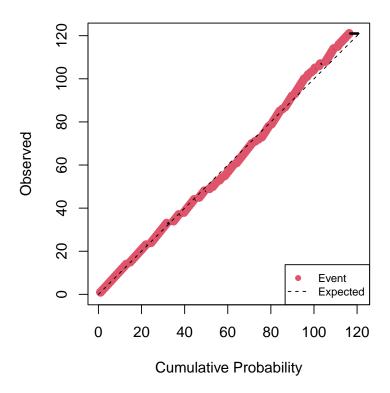
	N	Observed	Expected	$(O-E)^2/E$	$(O-E)^2/V$
$\begin{array}{c} { m class}{=}0 \\ { m class}{=}1 \end{array}$	130	88	101.3	1.76	10.9
	38	33	19.7	9.05	10.9

1.4.3 Cox Calibration

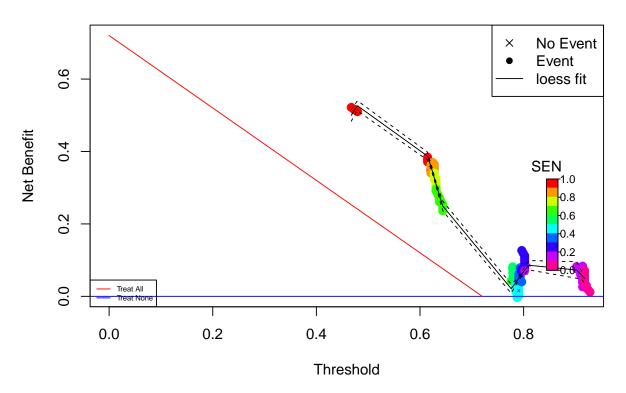
h0	Gain	DeltaTime
1.29	1.52	749

1.4.4 The RRplot() of the calibrated model

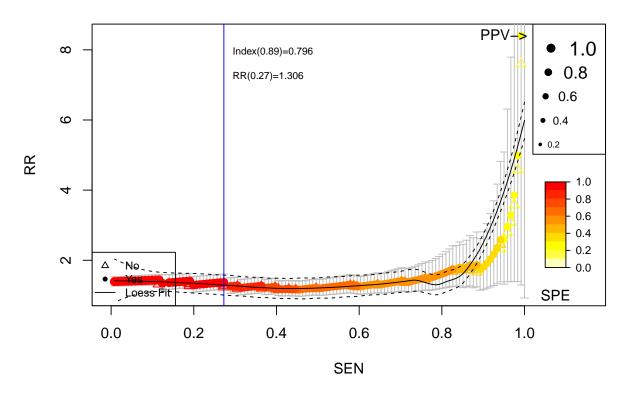
Cumulative vs. Observed: Train: Lung

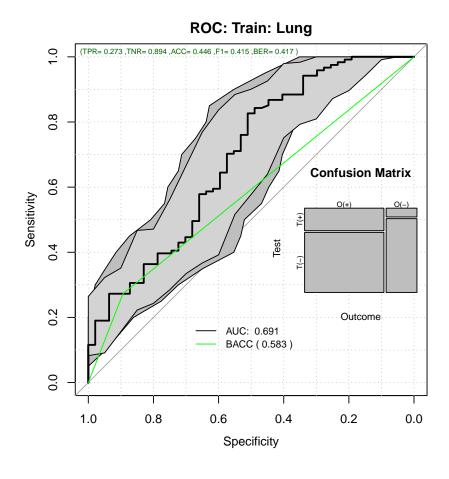


Decision Curve Analysis: Train: Lung

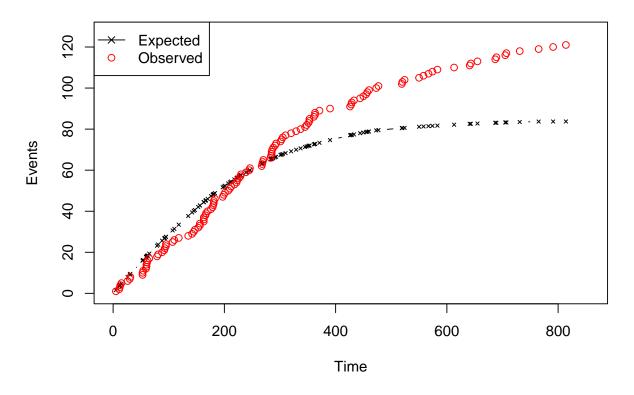


Relative Risk: Train: Lung

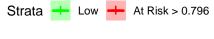


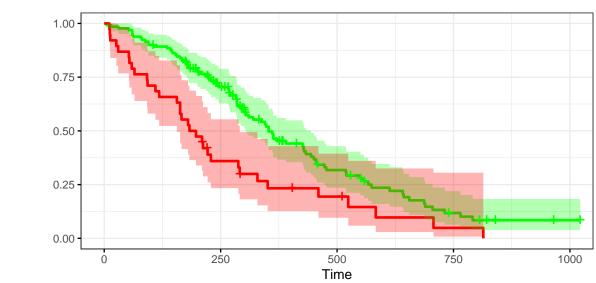


Time vs. Events: Train: Lung



Kaplan-Meier: Train: Lung





Number at risk

Low	130	77	25	7	1
At Risk > 0.796	38	12	5	1	0

1.4.5 Calibrated Train Performance

Survival probability

pander::pander(t(rrAnalysisTrain\$0ERatio), caption="0/E Ratio")

Table 26: O/E Ratio

est	lower	upper
1.45	1.2	1.73

pander::pander(t(rrAnalysisTrain\$0E95ci),caption="0/E Ratio")

Table 27: O/E Ratio

mean	50%	2.5%	97.5%
1.04	1.04	0.996	1.08

pander::pander(t(rrAnalysisTrain\$0Acum95ci),caption="0/Acum Ratio")

Table 28: O/Acum Ratio

mean	50%	2.5%	97.5%
1.01	1.01	0.999	1.01

pander::pander(rrAnalysisTrain\$c.index\$cstatCI, caption="C. Index")

mean.C Index	median	lower	upper
0.651	0.65	0.591	0.71

pander::pander(t(rrAnalysisTrain\$ROCAnalysis\$aucs),caption="ROC AUC")

Table 30: ROC AUC

est	lower	upper
0.691	0.598	0.784

pander::pander((rrAnalysisTrain\$ROCAnalysis\$sensitivity),caption="Sensitivity")

Table 31: Sensitivity

est	lower	upper
0.273	0.196	0.361

pander::pander((rrAnalysisTrain\$ROCAnalysis\$specificity), caption="Specificity")

Table 32: Specificity

est	lower	upper
0.894	0.769	0.965

pander::pander(t(rrAnalysisTrain\$thr_atP),caption="Probability Thresholds")

Table 33: Probability Thresholds

90%	
0.796	

pander::pander(t(rrAnalysisTrain\$RR_atP),caption="Risk Ratio")

Table 34: Risk Ratio

est	lower	upper
1.31	1.11	1.54

pander::pander(rrAnalysisTrain\$surdif,caption="Logrank test")

Table 35: Logrank test Chisq = 10.879375 on 1 degrees of freedom, p = 0.000972

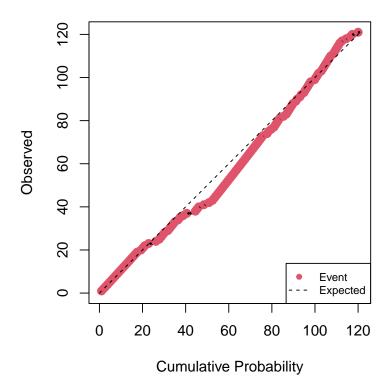
	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
class=0	130	88	101.3	1.76	10.9
class=1	38	33	19.7	9.05	10.9

1.5 Cross-Validation

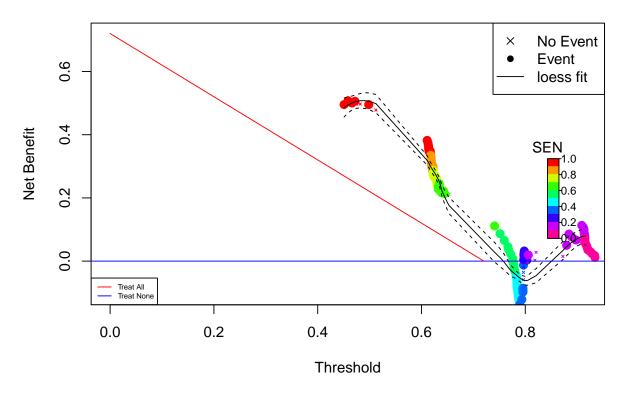
Tested: 117 Avg. Selected: 3.35 Min Tests: 1 Max Tests: 5 Mean Tests: 1.709402 . MAD: 0.4774963 [++].[+++].[+++].[+++].[+++].[+++].[+++].[+++].[+++].[+++].[+++] Tested: 141 Avg. Selected: 3.333333 Min]40 Tested: 152 Avg. Selected: 3.4 Min Tests: 1 Max Tests: 7 Mean Tests: 2.631579 . MAD: 0.4739485 $[60\ \mathrm{Tested}:\ 164\ \mathrm{Avg}.$ Selected: $3.416667\ \mathrm{Min}\ \mathrm{Tests}:\ 1\ \mathrm{Max}\ \mathrm{Tests}:\ 10\ \mathrm{Mean}\ \mathrm{Tests}:\ 3.658537$. Avg. Selected: 3.471429 Min Tests: 1 Max Tests: 10 Mean Tests: 4.216867 . MAD: 0.4761199 Tests: 1 Max Tests: 10 Mean Tests: 4.819277. MAD: 0.4762632. [+++]. [+++]. [+++]. [+++]. [++++]. [++++].].[+++].[+].[+].[+]90 Tested: 167 Avg. Selected: 3.444444 Min Tests: 1 Max Tests: 10 Mean Tests: Tested: 167 Avg. Selected: 3.51 Min Tests: 1 Max Tests: 12 Mean Tests: 5.988024 . Avg. Selected: 3.536364 Min Tests: 1 Max Tests: 15 Mean Tests: 6.586826 . MAD: 0.4749925 $\text{Min Tests: 17 Mean Tests: 18 Max Tests: 18 Max Tests: 19 Mean Tests: 19 Mean$ Tested: 168 Avg. Selected: 3.546154 Min Tests: 1 Max Tests: 17 Mean Tests: 7.738095 . MAD: 0.4752674 [++++], [+++], [+++], [+++], [+++], [+++], [+++], [+++], [++[.++].[++].[++-].[+++]150 Tested: 168 Avg. Selected: 3.526667 Min Tests: 1 Max Tests: 19 Mean Tested: 168 Avg. Selected: 3.54375 Min Tests: 1 Max Tests: 20 Mean Tests: 9.52381 . MAD: 0.474961 [+].[+++].[+++].[+++].[+++].[+++].[+++].[+++].[++++].[++++]170 Tested: 168 Avg. Selected: 3.529412 Min

```
Tested: 168 Avg. Selected: 3.544444 Min Tests: 1 Max Tests: 21 Mean Tests: 10.71429 . MAD: 0.4753713
Tested: 168 Avg. Selected: 3.56 Min Tests: 2 Max Tests: 22 Mean Tests: 11.90476 . MAD: 0.4755006
stp <- rcv$survTestPredictions</pre>
stp <- stp[!is.na(stp[,4]),]</pre>
bbx <- boxplot(unlist(stp[,1])~rownames(stp),plot=FALSE)</pre>
times <- bbx$stats[3,]</pre>
status <- boxplot(unlist(stp[,2])~rownames(stp),plot=FALSE)$stats[3,]</pre>
prob <- ppoisGzero(boxplot(unlist(stp[,4])~rownames(stp),plot=FALSE)$stats[3,],h0)</pre>
rdatacv <- cbind(status,prob)</pre>
rownames(rdatacv) <- bbx$names</pre>
names(times) <- bbx$names</pre>
rrAnalysisTest <- RRPlot(rdatacv,atProb=c(0.90),</pre>
                  timetoEvent=times,
                  title="Test: Lung Cancer",
                  ysurvlim=c(0.00,1.0),
                  riskTimeInterval=timeinterval)
```

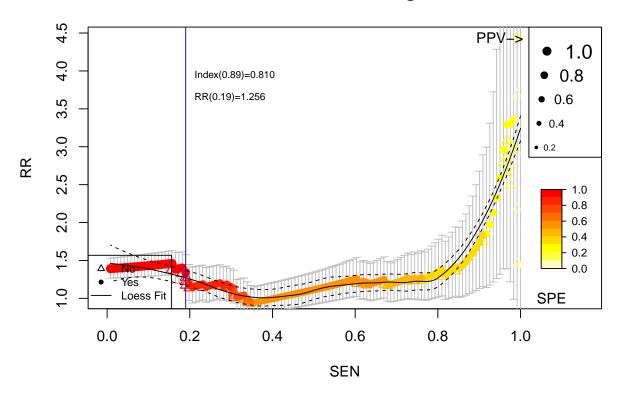
Cumulative vs. Observed: Test: Lung Cancer



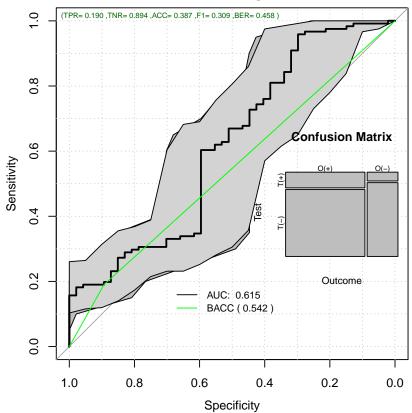
Decision Curve Analysis: Test: Lung Cancer



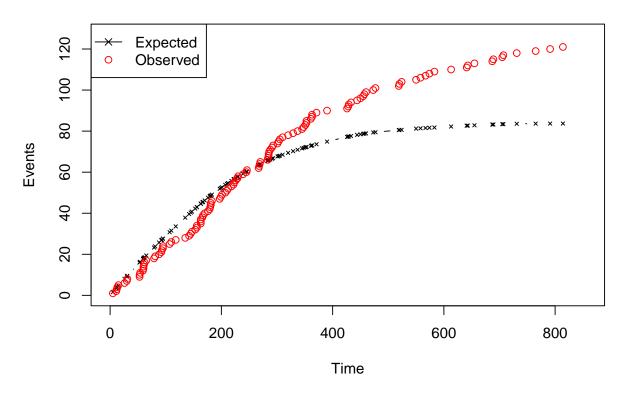
Relative Risk: Test: Lung Cancer



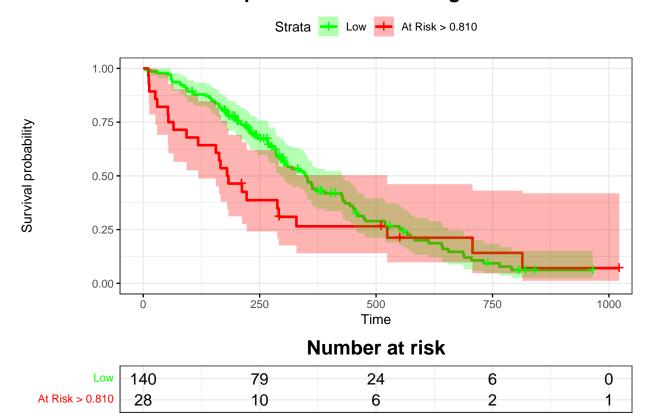




Time vs. Events: Test: Lung Cancer



Kaplan-Meier: Test: Lung Cancer



1.5.1 Cross-Validation Test Performance

pander::pander(t(rrAnalysisTest\$0ERatio), caption="0/E Ratio")

Table 36: O/E Ratio

est	lower	upper
1.45	1.2	1.73

pander::pander(t(rrAnalysisTest\$0E95ci),caption="0/E Ratio")

Table 37: O/E Ratio

mean	50%	2.5%	97.5%
1.03	1.03	0.996	1.07

pander::pander(t(rrAnalysisTest\$OAcum95ci), caption="0/Acum Ratio")

Table 38: O/Acum Ratio

mean	50%	2.5%	97.5%
0.961	0.962	0.95	0.972

pander::pander(rrAnalysisTest\$c.index\$cstatCI,caption="C. Index")

mean.C Index	median	lower	upper
0.601	0.601	0.536	0.66

pander::pander(t(rrAnalysisTest\$ROCAnalysis\$aucs),caption="ROC AUC")

Table 40: ROC AUC

est	lower	upper
0.615	0.515	0.715

pander::pander((rrAnalysisTest\$ROCAnalysis\$sensitivity), caption="Sensitivity")

Table 41: Sensitivity

est	lower	upper
0.19	0.124	0.271

pander::pander((rrAnalysisTest\$ROCAnalysis\$specificity),caption="Specificity")

Table 42: Specificity

est	lower	upper
0.894	0.769	0.965

pander::pander(t(rrAnalysisTest\$thr_atP),caption="Probability Thresholds")

Table 43: Probability Thresholds

6	90%
().81

pander::pander(t(rrAnalysisTest\$RR_atP),caption="Risk Ratio")

Table 44: Risk Ratio

est	lower	upper
1.26	1.05	1.51

pander::pander(rrAnalysisTest\$surdif,caption="Logrank test")

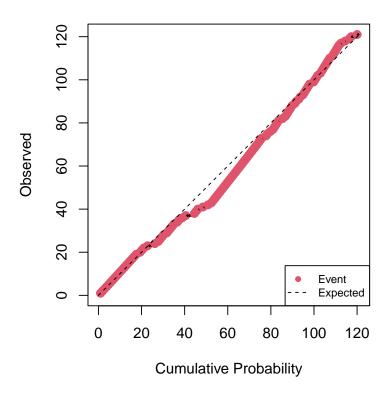
Table 45: Logrank test Chisq = 2.346444 on 1 degrees of freedom, p = 0.125569

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
class=0	140	98	103.8	0.328	2.35
class=1	28	23	17.2	1.981	2.35

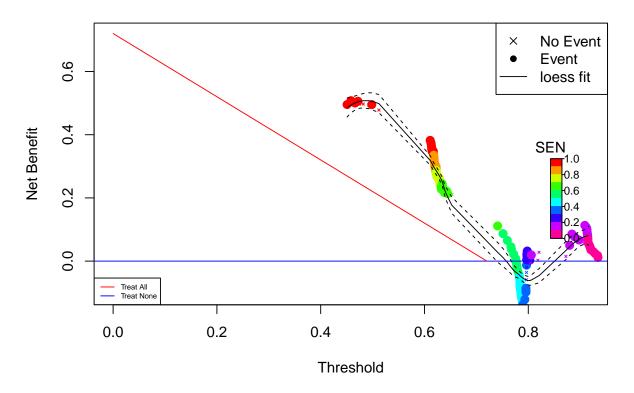
1.5.2 Calibrating the test results

h0	Gain	DeltaTime
0.85	1	752

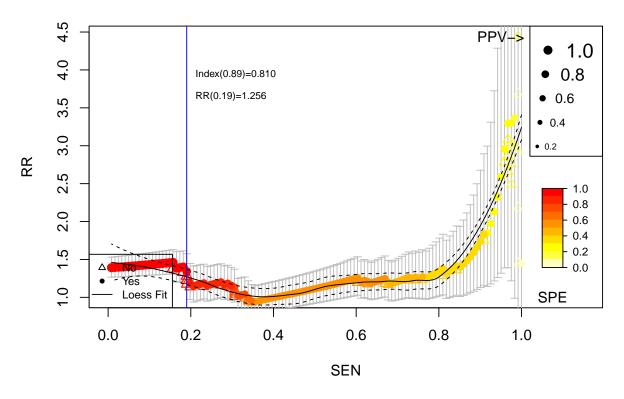
Cumulative vs. Observed: Calibrated Test: Lung



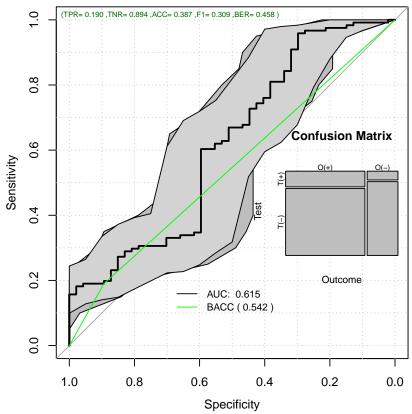
Decision Curve Analysis: Calibrated Test: Lung



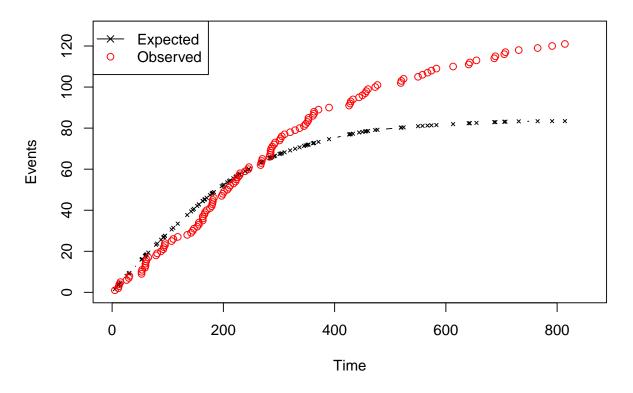
Relative Risk: Calibrated Test: Lung



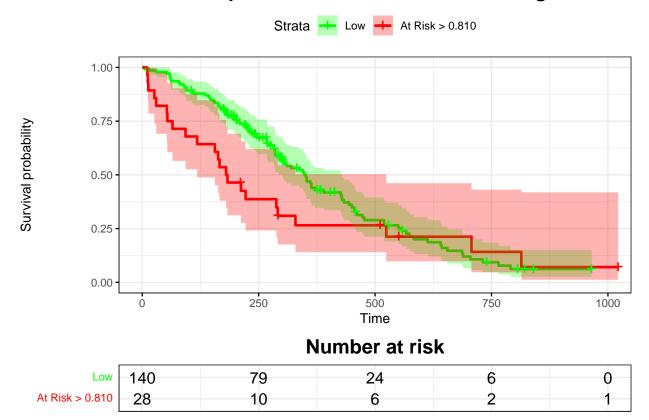




Time vs. Events: Calibrated Test: Lung



Kaplan-Meier: Calibrated Test: Lung



1.5.3 Calibrated Test Performance

pander::pander(t(rrAnalysisTest\$0ERatio), caption="0/E Ratio")

Table 47: O/E Ratio

est	lower	upper
1.45	1.2	1.73

pander::pander(t(rrAnalysisTest\$0E95ci),caption="0/E Ratio")

Table 48: O/E Ratio

mean	50%	2.5%	97.5%
1.04	1.04	0.998	1.08

pander::pander(t(rrAnalysisTest\$OAcum95ci), caption="0/Acum Ratio")

Table 49: O/Acum Ratio

mean	50%	2.5%	97.5%
0.961	0.961	0.951	0.972

pander::pander(rrAnalysisTest\$c.index\$cstatCI,caption="C. Index")

mean.C Index	median	lower	upper
0.601	0.602	0.536	0.664

pander::pander(t(rrAnalysisTest\$ROCAnalysis\$aucs),caption="ROC AUC")

Table 51: ROC AUC

est	lower	upper
0.615	0.515	0.715

pander::pander((rrAnalysisTest\$ROCAnalysis\$sensitivity),caption="Sensitivity")

Table 52: Sensitivity

est	lower	upper
0.19	0.124	0.271

pander::pander((rrAnalysisTest\$ROCAnalysis\$specificity),caption="Specificity")

Table 53: Specificity

est	lower	upper
0.894	0.769	0.965

pander::pander(t(rrAnalysisTest\$thr_atP),caption="Probability Thresholds")

Table 54: Probability Thresholds

90%

pander::pander(t(rrAnalysisTest\$RR_atP),caption="Risk Ratio")

Table 55: Risk Ratio

est	lower	upper
1.26	1.05	1.51

pander::pander(rrAnalysisTest\$surdif,caption="Logrank test")

Table 56: Logrank test Chisq = 2.346444 on 1 degrees of freedom, p = 0.125569

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
class=0	140	98	103.8	0.328	2.35
class=1	28	23	17.2	1.981	2.35