# Lung Cancer

Jose Tamez

2023-05-11

### 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	1 2 8 10 10 26
1	Detailed Survival analyis of the Survival lung data.	
1.0	0.1 Libraries	
	ibrary(survival) ibrary(FRESA.CAD)	
##	# Loading required package: Rcpp	
##	# Loading required package: stringr	
##	# Loading required package: miscTools	
##	# Loading required package: Hmisc	
	# Attaching package: 'Hmisc'	
## ##	# The following objects are masked from 'package:base': #	
##	format.pval, units	
##	# Loading required package: pROC	
##	# Type 'citation("pROC")' for a citation.	
## ##	# # Attaching package: 'pROC'	
##		
##		
pai	<pre>p &lt;- par(no.readonly = TRUE) ander::panderOptions('digits', 3) ander::panderOptions('keep.trailing.zeros',TRUE)</pre>	

#### 1.0.2 Libraries

```
data(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))

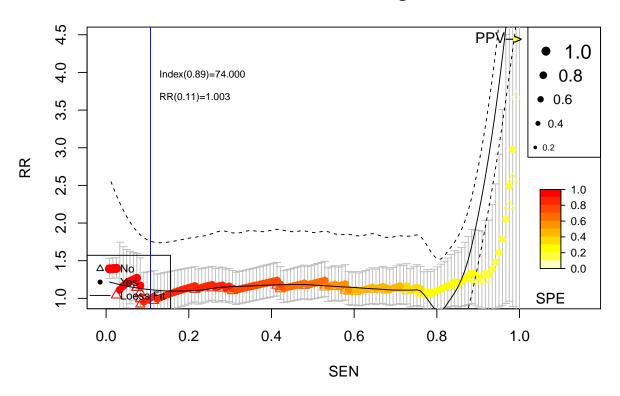
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
5	175	268	310	416	1022

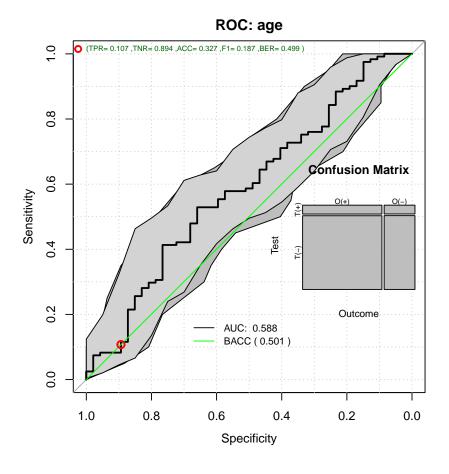
#### 1.1 Exploring Raw Features with RRPlot

```
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>
```

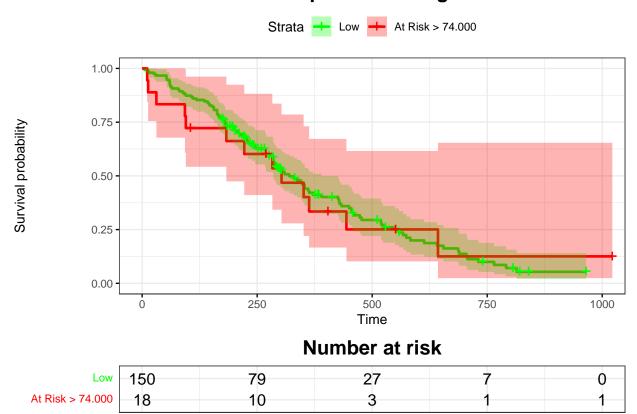
age	wt.loss
0.106	0.106

## Relative Risk: age

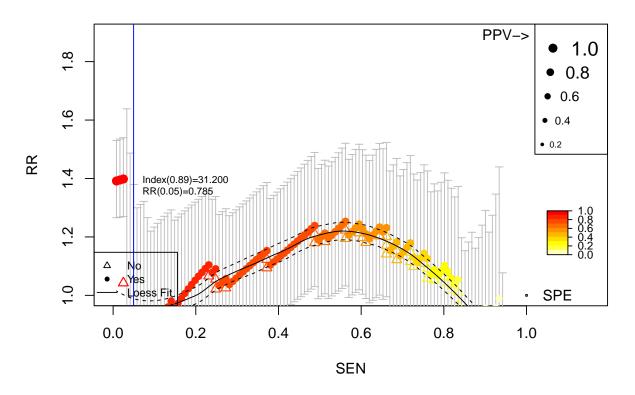


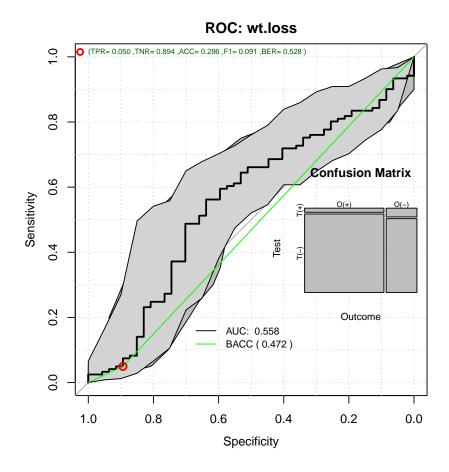


# Kaplan-Meier: age



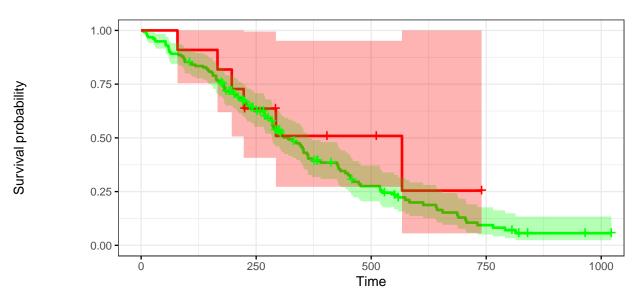
### Relative Risk: wt.loss





### 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	0.588	0.490	0.686
wt.loss	0.558	0.459	0.656

#### pander::pander(CstatCI)

	mean.C Index	median	lower	upper
age	0.559	0.559	0.498	0.620
wt.loss	0.518	0.516	0.454	0.576

#### pander::pander(RRatios)

	est	lower	upper
$egin{array}{c} { m age} \\ { m wt.loss} \end{array}$	1.003 0.785	$0.741 \\ 0.462$	1.36 1.33

#### pander::pander(LogRangp)

age	0.857
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	0.588	0.559	0.1074	0.894	1.003

	ROCAUC	C-Stat	Sen	Spe	RR
wt.loss	0.558	0.518	0.0496	0.894	0.785

### 1.3 Modeling

Table 11: Table continues below

	Estimate	lower	HR	upper	u.Accuracy	r.Accuracy
ph.ecog	4.32e-01	1.194	1.541	1.988	0.679	0.649
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	-4.06e-07	1.000	1.000	1.000	0.577	0.720
age	9.13e-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	0.7
age	1.97	1.14	0.04871	0.2

#### 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

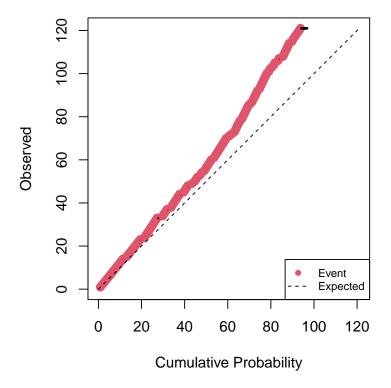
```
timeinterval <- 2*mean(subset(lung,status==1)$time)

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")
```

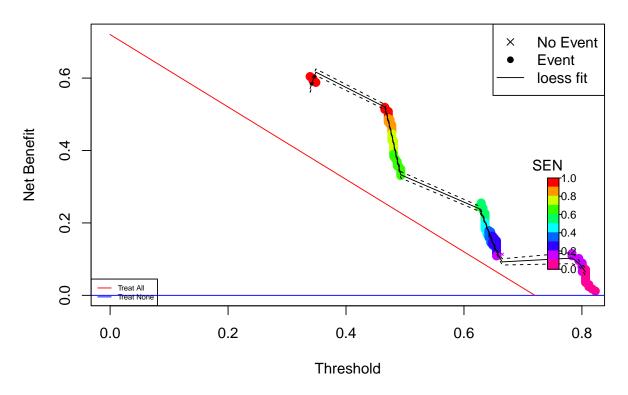
Table 14: Initial Parameters

h0	timeinterval
0.85	578

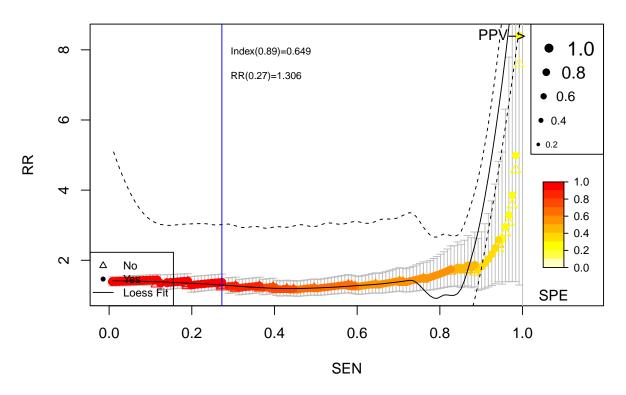
### Cumulative vs. Observed: Raw Train: Lung Cancer



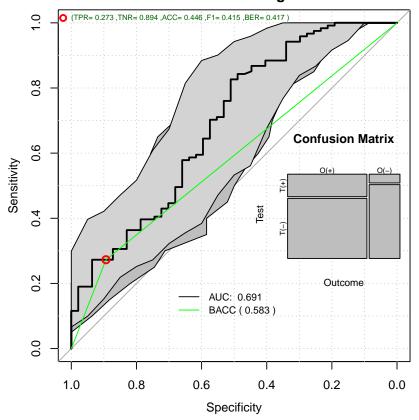
## **Decision Curve Analysis: Raw Train: Lung Cancer**



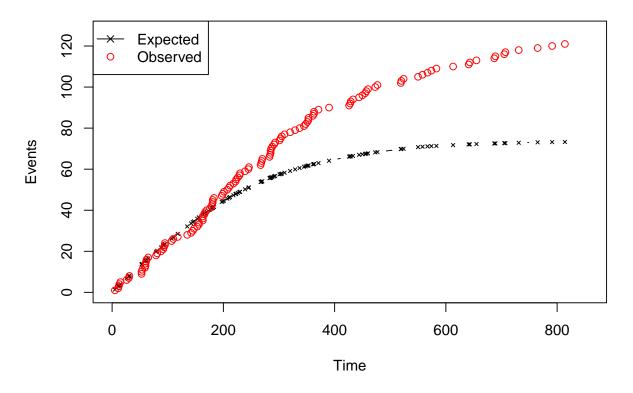
## Relative Risk: Raw Train: Lung Cancer



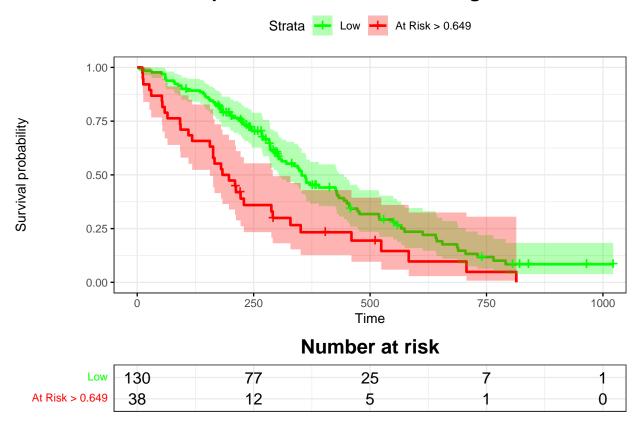
**ROC: Raw Train: Lung Cancer** 



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\$keyPoints),caption="Threshold values")

Table 15: Threshold values

	@:0.9	@MAX_BACC	@MAX_RR	@SPE100	p(0.5)
$\mathbf{Thr}$	0.649	0.478	0.339	0.339	0.493
$\mathbf{R}\mathbf{R}$	1.240	1.742	68.491	68.491	1.270
SEN	0.273	0.826	1.000	1.000	0.612
$\mathbf{SPE}$	0.872	0.511	0.191	0.191	0.596
$\mathbf{BACC}$	0.573	0.669	0.596	0.596	0.604

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

Table 16: O/E Ratio

O/E	Low	Upper	p.value
1.65	1.37	1.97	3.16e-07

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

Table 17: O/E Mean

mean	50%	2.5%	97.5%
1.23	1.23	1.19	1.27

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

Table 18: O/Acum Mean

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	0.65	0.584	0.707

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

Table 20: ROC AUC

est	lower	upper
0.691	0.598	0.784

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

Table 21: Sensitivity

est	lower	upper
0.273	0.196	0.361

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

Table 22: Specificity

est	lower	upper
0.894	0.769	0.965

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

Table 23: Probability Thresholds

90%	
0.649	

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

Table 24: Risk Ratio

est	lower	upper
1.31	1.11	1.54

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

Table 25: Logrank test Chisq = 10.879375 on 1 degrees of freedom, p = 0.000972

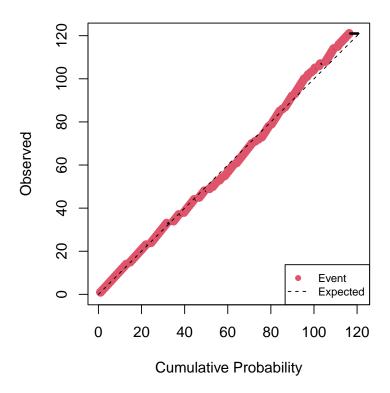
	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
class=0	130	88	101.3	1.76	10.9
class=1	38	33	19.7	9.05	10.9

#### 1.4.3 Cox Calibration

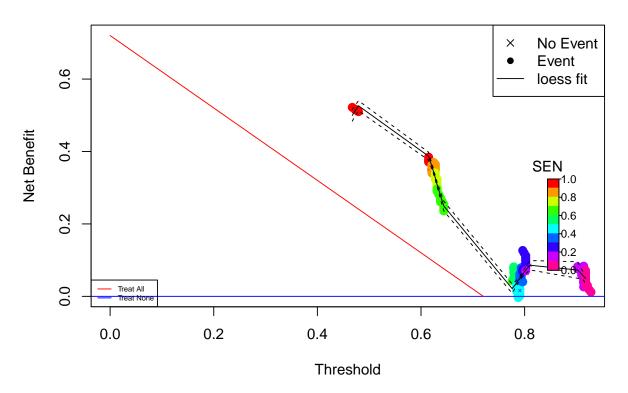
h0	Gain	DeltaTime
1.29	1.52	749

#### 1.4.4 The RRplot() of the calibrated model

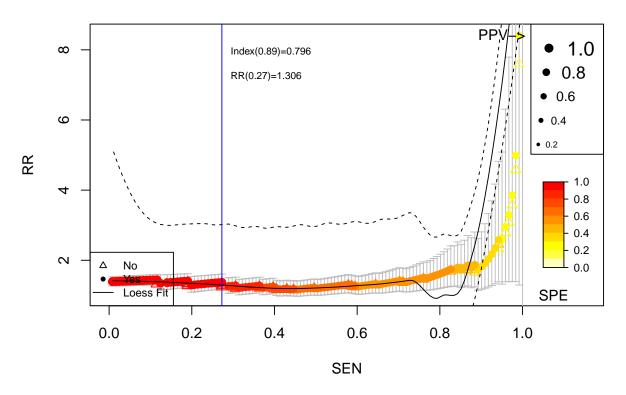
# Cumulative vs. Observed: Train: Lung

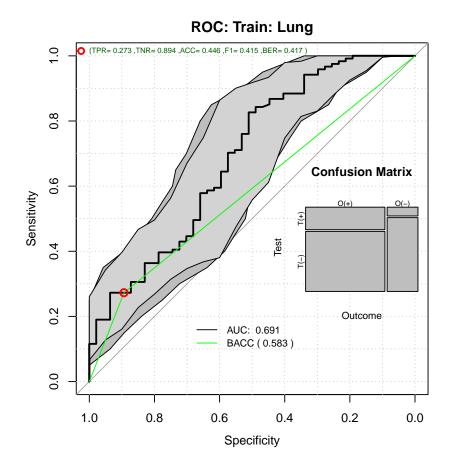


# **Decision Curve Analysis: Train: Lung**

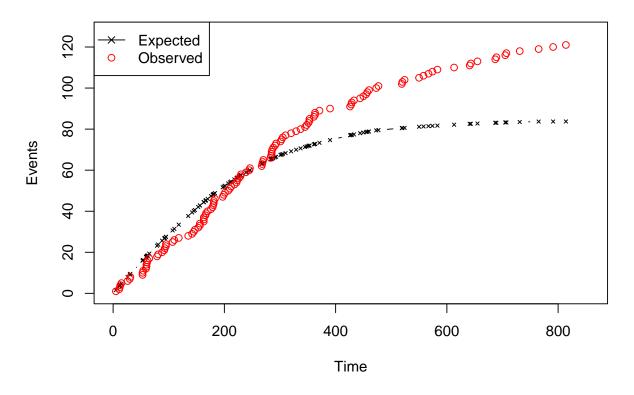


## Relative Risk: Train: Lung



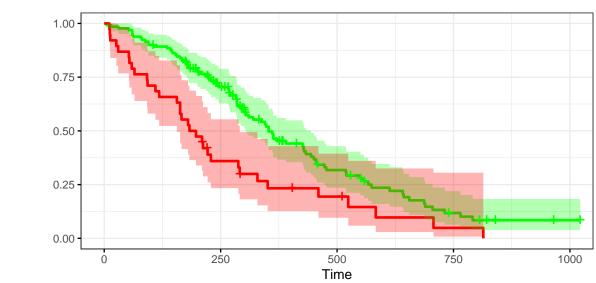


Time vs. Events: Train: Lung



### 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\$keyPoints), caption="Threshold values")

Table 27: Threshold values

	@:0.9	@MAX_BACC	@MAX_RR	@SPE100	p(0.5)
$\mathbf{Thr}$	0.796	0.628	0.467	0.467	0.479
$\mathbf{R}\mathbf{R}$	1.240	1.742	68.491	68.491	2.784
$\mathbf{SEN}$	0.273	0.826	1.000	1.000	0.959
$\mathbf{SPE}$	0.872	0.511	0.191	0.191	0.277
$\mathbf{BACC}$	0.573	0.669	0.596	0.596	0.618

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

Table 28: O/E Ratio

O/E	Low	Upper	p.value
1.45	1.2	1.73	0.000124

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

Table 29: O/E Mean

mean	50%	2.5%	97.5%
1.06	1.06	1.02	1.09

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

Table 30: O/Acum Mean

mean	50%	2.5%	97.5%
1	1	0.996	1.01

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

mean.C Index	median	lower	upper
0.651	0.651	0.589	0.712

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

Table 32: ROC AUC

est	lower	upper
0.691	0.598	0.784

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

Table 33: Sensitivity

est	lower	upper
0.273	0.196	0.361

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

Table 34: Specificity

est	lower	upper
0.894	0.769	0.965

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

Table 35: Probability Thresholds

90%	
0.796	

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

Table 36: Risk Ratio

est	lower	upper
1.31	1.11	1.54

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

Table 37: Logrank test Chisq = 10.879375 on 1 degrees of freedom, p = 0.000972

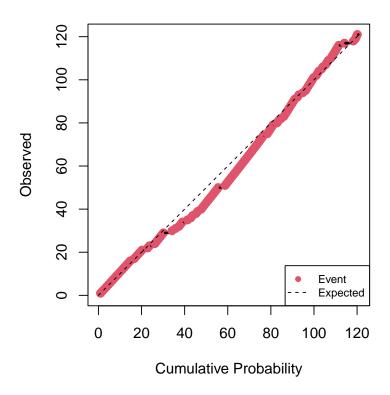
	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
class=0	130	88	101.3	1.76	10.9
class=1	38	33	19.7	9.05	10.9

#### 1.5 Cross-Validation

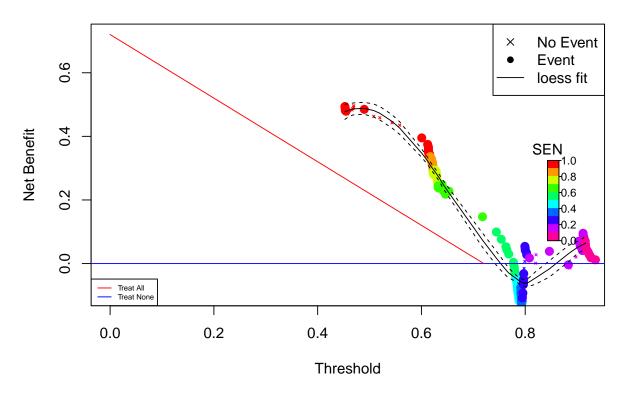
Min Tests: 1 Max Tests: 5 Mean Tests: 1.388889 . MAD: 0.4886257 .[++].[+++].[+++].[+++].[+-+]. Tested: 140 Avg. Selected: 3.533333 Min Tests: 1 Max Tests: 5 Mean Tests: 2.142857 . MAD: 0.4767379[++].[+++]Selected: 3.6 Min Tests: 1 Max Tests: 7 Mean Tests: 2.580645 . MAD: 0.4774606 [++], [+++], [+++], [+++], [+++], [+++], [+++], [++++], [++++] Tested: 158 Avg. Selected: 3.58 Min Tests: Tested: 166 Avg. Selected: 3.616667 Min Tests: 1 Max Tests: 10 Mean Tests: 3.614458 . MAD: [.+++].++].++Tested: 167 Avg. Selected: 3.644444 Min Tests: 1 Max Tests: 14 Mean Tests: 5.389222 . MAD:  $Selected: \ 3.65 \ Min \ Tests: \ 1 \ Max \ Tests: \ 16 \ Mean \ Tests: \ 5.952381 \ . \ MAD: \ 0.4784858$ Tests: 1 Max Tests: 16 Mean Tests: 6.547619 . MAD: 0.4778259 .[+++].[+++].[+++].[+++]. |.[++++].[+++].[+++].[+++].[+++].[++++] Tested: 168 Avg. Selected: 3.616667 Min Tests: 2 Max Tests:

```
Selected: 3.646154 Min Tests: 2 Max Tests: 16 Mean Tests: 7.738095 .
Tested: 168 Avg.
Avg. Selected: 3.642857 Min Tests: 2 Max Tests: 16 Mean Tests: 8.333333 . MAD: 0.4774368
168 Avg.
3.633333 Min Tests: 2 Max Tests: 16 Mean Tests: 8.928571 . MAD: 0.4772352 .[+++].[+++].[+++].
Selected: 3.629412 Min Tests: 3 Max Tests: 20 Mean Tests: 10.11905 .
Tested: 168 Avg.
Avg. Selected: 3.627778 Min Tests: 3 Max Tests: 21 Mean Tests: 10.71429 . MAD: 0.477139
Tested: 168 Avg. Selected: 3.625 Min Tests: 3 Max Tests: 23 Mean Tests: 11.90476 . MAD: 0.4768734
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),
             timetoEvent=times,
             title="Test: Lung Cancer",
             ysurvlim=c(0.00,1.0),
             riskTimeInterval=timeinterval)
```

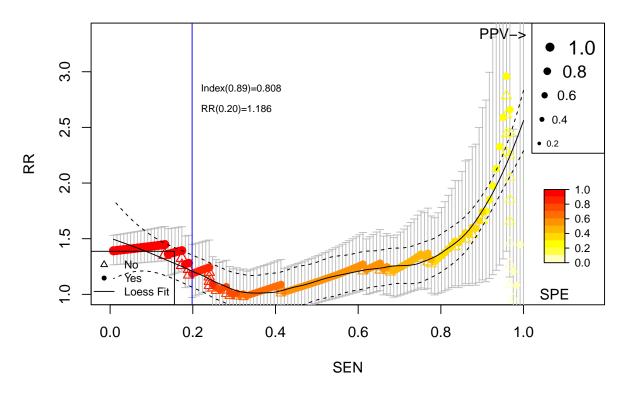
# **Cumulative vs. Observed: Test: Lung Cancer**



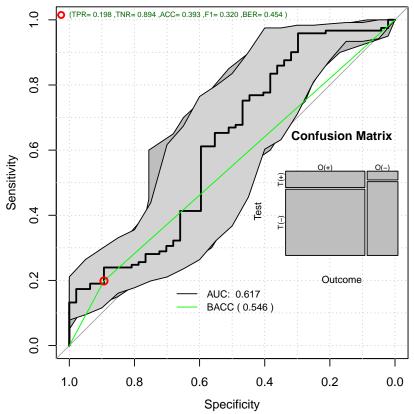
## **Decision Curve Analysis: Test: Lung Cancer**



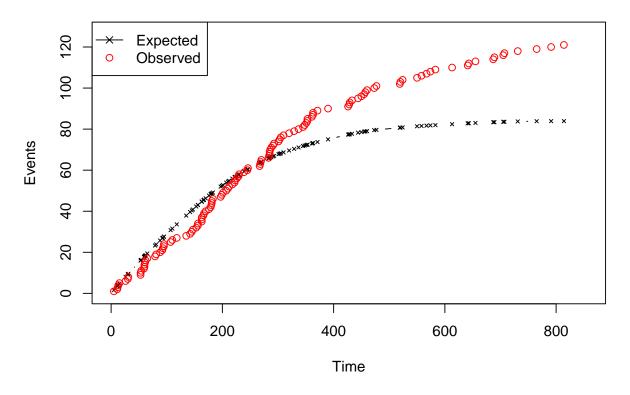
# Relative Risk: Test: Lung Cancer



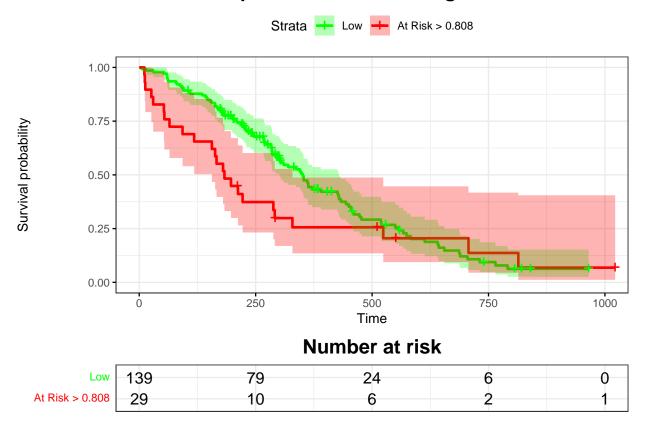




Time vs. Events: Test: Lung Cancer



### Kaplan-Meier: Test: Lung Cancer



#### 1.5.1 Cross-Validation Test Performance

pander::pander(t(rrAnalysisTest\$keyPoints), caption="Threshold values")

Table 38: Threshold values

	@:0.9	@MAX_BACC	@MAX_RR	@SPE100	p(0.5)
Thr	0.808	0.601	0.601	0.4525	0.508
$\mathbf{R}\mathbf{R}$	1.186	2.958	2.958	7.2455	2.275
SEN	0.198	0.959	0.959	1.0000	0.959
$\mathbf{SPE}$	0.894	0.298	0.298	0.0213	0.213
$\mathbf{BACC}$	0.546	0.628	0.628	0.5106	0.586

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

Table 39: O/E Ratio

O/E	Low	Upper	p.value
1.44	1.2	1.72	0.000128

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

Table 40: O/E Mean

mean	50%	2.5%	97.5%
1.05	1.05	1.01	1.09

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

Table 41: O/Acum Mean

mean	50%	2.5%	97.5%
0.955	0.955	0.945	0.965

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

mean.C Index	median	lower	upper
0.606	0.607	0.543	0.667

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

Table 43: ROC AUC

est	lower	upper
0.617	0.517	0.717

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

Table 44: Sensitivity

est	lower	upper
0.198	0.131	0.281

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

Table 45: Specificity

est	lower	upper
0.894	0.769	0.965

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

Table 46: Probability Thresholds

90%	
0.808	

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

Table 47: Risk Ratio

est	lower	upper
1.19	0.972	1.45

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

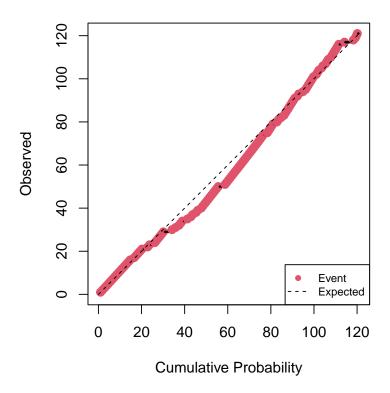
Table 48: Logrank test Chisq = 2.869716 on 1 degrees of freedom, p = 0.090261

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
class=0	139	97	103.5	0.409	2.87
class=1	29	24	17.5	2.417	2.87

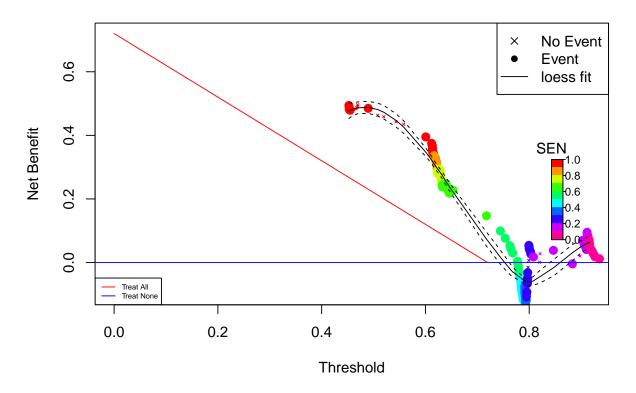
#### 1.5.2 Calibrating the test results

h0	Gain	DeltaTime
0.85	1	754

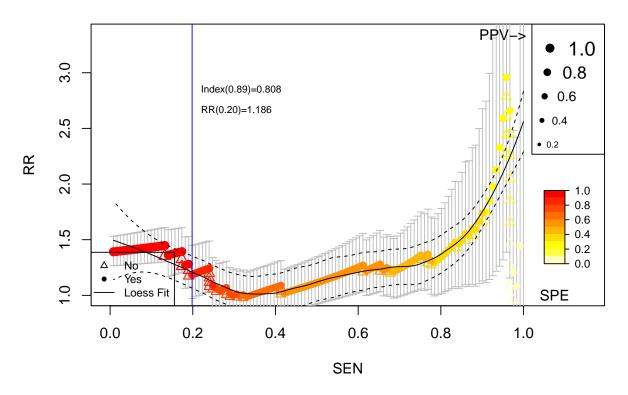
## **Cumulative vs. Observed: Calibrated Test: Lung**



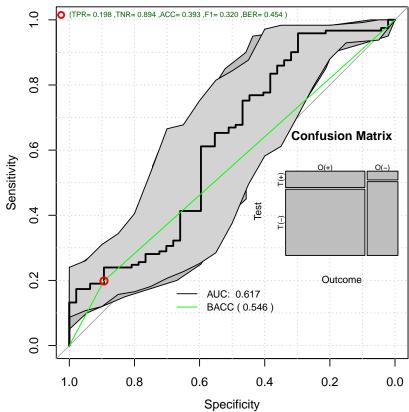
## **Decision Curve Analysis: Calibrated Test: Lung**



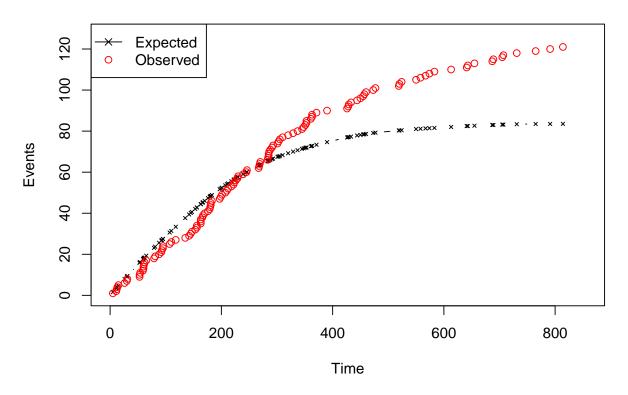
# Relative Risk: Calibrated Test: Lung



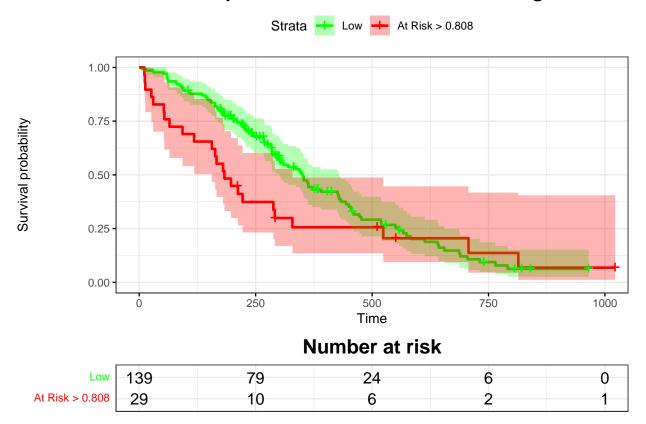




Time vs. Events: Calibrated Test: Lung



### Kaplan-Meier: Calibrated Test: Lung



#### 1.5.3 Calibrated Test Performance

pander::pander(t(rrAnalysisTest\$keyPoints), caption="Threshold values")

Table 50: Threshold values

	@:0.9	@MAX_BACC	@MAX_RR	@SPE100	p(0.5)
Thr	0.808	0.601	0.601	0.4525	0.508
$\mathbf{R}\mathbf{R}$	1.186	2.958	2.958	7.2455	2.275
SEN	0.198	0.959	0.959	1.0000	0.959
$\mathbf{SPE}$	0.894	0.298	0.298	0.0213	0.213
$\mathbf{BACC}$	0.546	0.628	0.628	0.5106	0.586

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

Table 51: O/E Ratio

O/E	Low	Upper	p.value
1.45	1.2	1.73	0.000122

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

Table 52: O/E Mean

mean	50%	2.5%	97.5%
1.06	1.06	1.02	1.1

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

Table 53: O/Acum Mean

mean	50%	2.5%	97.5%
0.955	0.955	0.945	0.964

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

mean.C Index	median	lower	upper
0.606	0.606	0.54	0.673

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

Table 55: ROC AUC

est	lower	upper
0.617	0.517	0.717

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

Table 56: Sensitivity

est	lower	upper
0.198	0.131	0.281

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

Table 57: Specificity

est	lower	upper	
0.894	0.769	0.965	

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

Table 58: Probability Thresholds

90%	
0.808	

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

Table 59: Risk Ratio

est	lower	upper
1.19	0.972	1.45

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

Table 60: Logrank test Chisq = 2.869716 on 1 degrees of freedom, p = 0.090261

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
$\begin{array}{c} {\rm class}{=}0 \\ {\rm class}{=}1 \end{array}$	139	97	103.5	0.409	2.87
	29	24	17.5	2.417	2.87