Risk-Evaluation: Breast Cancer Royston-Altman

Jose Tamez

2023-05-01

Contents

1	Eva	luation of RISK survival models
	1.1	The libraries
	1.2	Breast Cancer Royston-Altman data
		Cox Modeling
		Cox Model Performance
	1.5	Performance on the external data set
	1.6	Logistic Model
	1.7	Logistic Model Performance
	1.8	Logistic Model Poisson Calibration
	1.9	Comparing the COX and Logistic Models on the Independent Data

1 Evaluation of RISK survival models

This document highlights the use of

- RRPlot(),
- CoxRiskCalibration(), and
- CalibrationProbPoissonRisk(),

for the evaluation (RRPlot), and calibration of cox models (CoxRiskCalibration) or logistic models (CalibrationProbPoissonRisk) of survival data.

Furthermore, it can be used to evaluate any Risk index that reruns the probability of a future event on external data-set.

This document will use the survival::rotterdam, and survival::gbsg data-sets to train and predict the risk of cancer recurrence after surgery. Both Cox and Logistic models will be trained and evaluated.

Here are some sample plots returned by the evaluated functions:

1.1 The libraries

```
library(survival)
library(FRESA.CAD)

## Loading required package: Rcpp

## Loading required package: stringr
```

Loading required package: Hmisc

Loading required package: miscTools

```
##
## 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'
## The following objects are masked from 'package:stats':
##
##
       cov, smooth, var
#source("~/GitHub/FRESA.CAD/R/RRPlot.R")
#source("~/GitHub/FRESA.CAD/R/PoissonEventRiskCalibration.R")
op <- par(no.readonly = TRUE)</pre>
pander::panderOptions('digits', 3)
#pander::panderOptions('table.split.table', 400)
pander::panderOptions('keep.trailing.zeros',TRUE)
```

1.2 Breast Cancer Royston-Altman data

1.2.1 data(gbsg, package="survival") and data(rotterdam, package="survival")

```
gbsgdata <- gbsg
rownames(gbsgdata) <- gbsgdata$pid</pre>
gbsgdata$pid <- NULL</pre>
odata <-rotterdam
rownames(odata) <- odata$pid
odata$pid <- NULL
odata$rfstime <- odata$rtime
odata$status <- odata$recur
odata$rtime <- NULL
odata$recur <- NULL
odata <- odata[,colnames(odata) %in% colnames(gbsgdata)]</pre>
odata$size <- 10*(odata$size=="<=20") +
  35*(odata\$size=="20-50") +
  60*(odata$size==">50")
data <- as.data.frame(model.matrix(Surv(rfstime,status)~.*.,odata))</pre>
data$`(Intercept)` <- NULL</pre>
dataBrestCancerTrain <- cbind(time=odata[rownames(data), "rfstime"], status=odata[rownames(data), "status"]</pre>
colnames(dataBrestCancerTrain) <-str_replace_all(colnames(dataBrestCancerTrain),":","_")</pre>
colnames(dataBrestCancerTrain) <-str_replace_all(colnames(dataBrestCancerTrain)," ","")</pre>
colnames(dataBrestCancerTrain) <-str_replace_all(colnames(dataBrestCancerTrain),"\\.","_")</pre>
colnames(dataBrestCancerTrain) <-str_replace_all(colnames(dataBrestCancerTrain),"-","_")</pre>
```

```
colnames(dataBrestCancerTrain) <-str_replace_all(colnames(dataBrestCancerTrain),">","_")
dataBrestCancerTrain$time <- dataBrestCancerTrain$time/365 ## To years

pander::pander(table(odata[rownames(data),"status"]),caption="rotterdam")</pre>
```

Table 1: rotterdam

0	1
1464	1518

1.2.2 data(gbsg, package="survival") data conditioning

```
gbsgdata <- gbsgdata[,colnames(odata)]
data <- as.data.frame(model.matrix(Surv(rfstime,status)~.*.,gbsgdata))

data$`(Intercept)` <- NULL

dataBrestCancerTest <- cbind(time=gbsgdata[rownames(data),"rfstime"],status=gbsgdata[rownames(data),"st

colnames(dataBrestCancerTest) <-str_replace_all(colnames(dataBrestCancerTest),":","_")
colnames(dataBrestCancerTest) <-str_replace_all(colnames(dataBrestCancerTest),"","")
colnames(dataBrestCancerTest) <-str_replace_all(colnames(dataBrestCancerTest),"\\.","_")
colnames(dataBrestCancerTest) <-str_replace_all(colnames(dataBrestCancerTest),"\\.","_")
colnames(dataBrestCancerTest) <-str_replace_all(colnames(dataBrestCancerTest),"-","_")
dataBrestCancerTest$time <- dataBrestCancerTest$time/365

pander::pander(table(odata[rownames(data),"status"]), caption="gbsg")</pre>
```

Table 2: gbsg

0	1
499	183

1.3 Cox Modeling

```
ml <- BSWiMS.model(Surv(time,status)~.,data=dataBrestCancerTrain,loops=1,NumberofRepeats = 5)</pre>
```

```
sm <- summary(ml)
pander::pander(sm$coefficients)</pre>
```

Table 3: Table continues below

	Estimate	lower	$_{ m HR}$	upper	u.Accuracy	r.Accuracy
age_nodes	0.000716	1.001	1.001	1.001	0.626	0.600
$\mathbf{size} \mathbf{_grade}$	0.005649	1.005	1.006	1.006	0.598	0.623
\mathbf{nodes}	0.086582	1.082	1.090	1.099	0.637	0.642
${f size}$	0.006888	1.005	1.007	1.009	0.595	0.641
${\bf size_nodes}$	-0.000378	1.000	1.000	1.000	0.624	0.643

	Estimate	lower	HR	upper	u.Accuracy	r.Accuracy
age_size	-0.000149	1.000	1.000	1.000	0.567	0.627
${f grade}$	0.204934	1.146	1.227	1.314	0.565	0.637
age	-0.003113	0.996	0.997	0.998	0.513	0.628
${\it grade_nodes}$	-0.013784	0.981	0.986	0.992	0.635	0.645

Table 4: Table continues below

	full.Accuracy	u.AUC	r.AUC	full.AUC	IDI	NRI
age_nodes	0.632	0.630	0.601	0.634	0.03040	0.4594
\mathbf{size} grade	0.632	0.599	0.626	0.634	0.01868	0.3914
\mathbf{nodes}	0.643	0.640	0.643	0.644	0.00745	0.0564
${f size}$	0.643	0.595	0.642	0.644	0.01447	0.3587
${f size_nodes}$	0.643	0.629	0.644	0.644	0.00346	0.3430
$\mathbf{age}\mathbf{_size}$	0.632	0.568	0.630	0.634	0.00635	0.1935
${f grade}$	0.643	0.561	0.638	0.644	0.00926	0.2069
age	0.643	0.513	0.628	0.644	0.00416	0.0917
${\bf grade_nodes}$	0.643	0.639	0.646	0.644	0.00207	-0.0910

	z.IDI	z.NRI	Delta.AUC	Frequency
age_nodes	12.81	14.37	0.033056	1
${f size_grade}$	9.82	11.29	0.007947	1
nodes	8.33	1.66	0.000148	1
\mathbf{size}	8.05	9.97	0.001322	1
${f size_nodes}$	7.25	9.57	-0.000377	1
$\mathbf{age}\mathbf{_size}$	5.95	5.36	0.004078	1
${f grade}$	5.88	6.31	0.005344	1
age	5.27	2.51	0.015465	1
${f grade_nodes}$	5.03	-2.55	-0.002609	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()

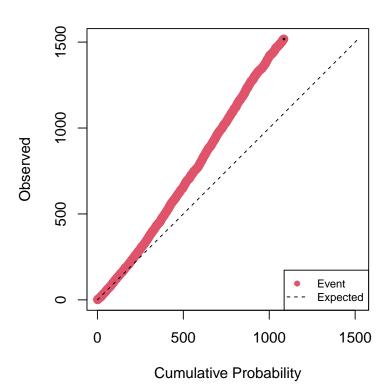
Here we will use the predicted event probability assuming a baseline hazard for events withing 5 years timeinterval <- 5 # Five years</pre>

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

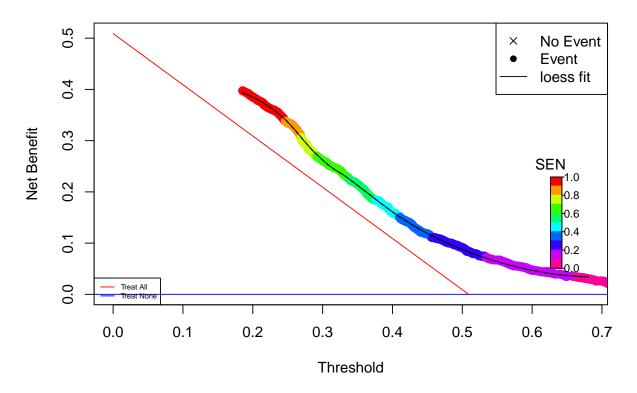
Table 6: Initial Parameters

h0	timeinterval
0.429	5

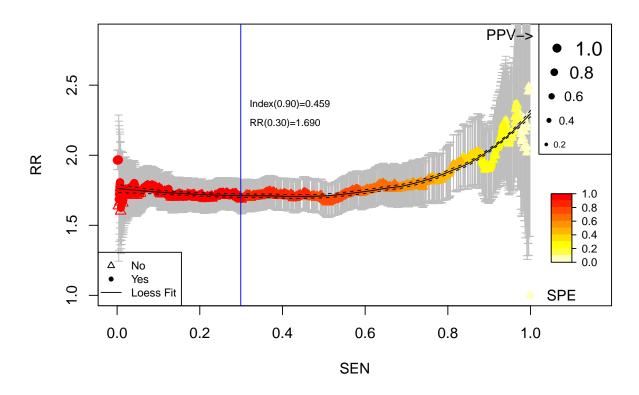
Cumulative vs. Observed: Train: Breast Cancer



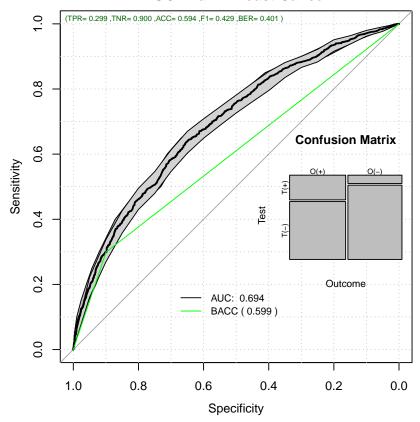
Decision Curve Analysis: Train: Breast Cancer



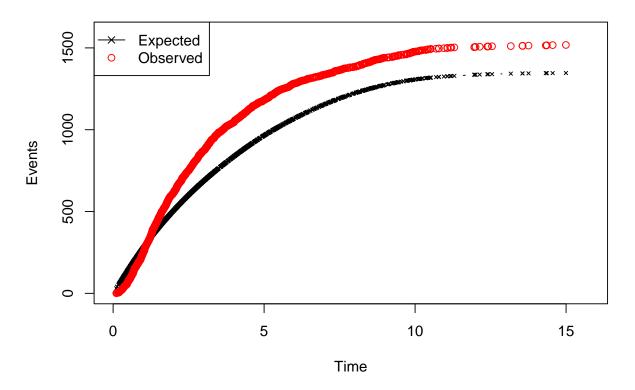
Relative Risk: Train: Breast Cancer



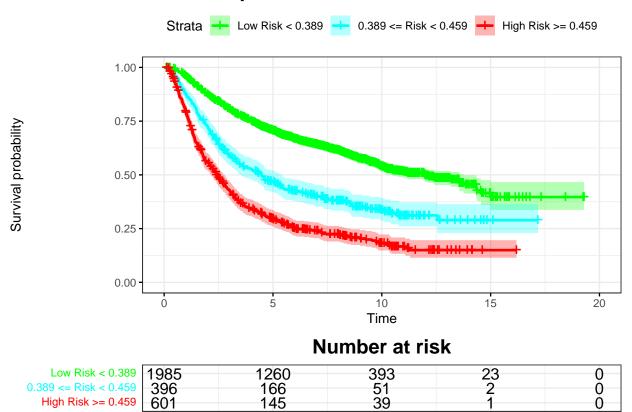
ROC: Train: Breast Cancer



Time vs. Events: Train: Breast Cancer



Kaplan-Meier: Train: Breast 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 7: Threshold values

	@:0.9	@:0.8	@MAX_BACC	@MAX_RR	@SPE100
${f Thr}$	0.459	0.389	0.320	0.214	0.18549
$\mathbf{R}\mathbf{R}$	1.690	1.713	1.799	2.376	1.00000
\mathbf{SEN}	0.299	0.462	0.644	0.965	1.00000
\mathbf{SPE}	0.900	0.798	0.646	0.125	0.00137
BACC	0.599	0.630	0.645	0.545	0.50068

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

Table 8: O/E Ratio

est	lower	upper
1.13	1.07	1.19

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

Table 9: O/E Ratio

mean	50%	2.5%	97.5%
1.13	1.13	1.12	1.14

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

Table 10: O/Acum Ratio

mean	50%	2.5%	97.5%
1.34	1.34	1.34	1.34

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

mean.C Index	median	lower	upper
0.676	0.677	0.662	0.69

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

Table 12: ROC AUC

est	lower	upper
0.694	0.675	0.713

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

Table 13: Sensitivity

est	lower	upper
0.299	0.276	0.323

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

Table 14: Specificity

est	lower	upper
0.9	0.883	0.915

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

Table 15: Probability Thresholds

90%	80%	at_max_BACC	at_max_RR	atSPE100
0.459	0.389	0.32	0.214	0.185

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

Table 16: Risk Ratio

est	lower	upper
1.69	1.59	1.8

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

Table 17: Logrank test Chisq = 465.079317 on 2 degrees of freedom, p = 0.000000

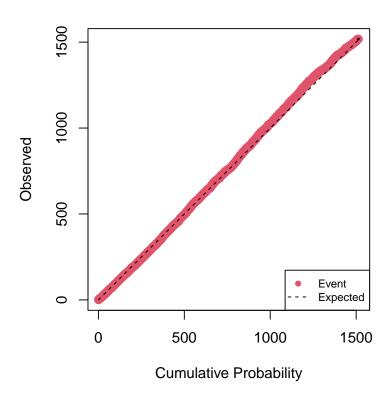
	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
class=0	1985	816	1144	93.9	385.7
class=1	396	248	177	28.0	31.8
class=2	601	454	197	336.3	391.3

1.4.3 Cox Calibration

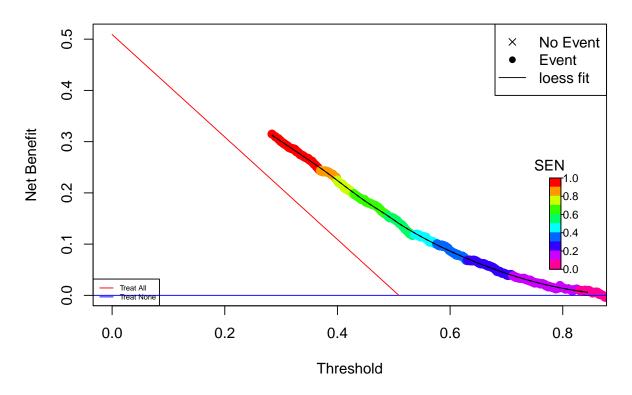
h0	Gain	DeltaTime
0.698	1.35	6.97

1.4.4 The RRplot() of the calibrated model

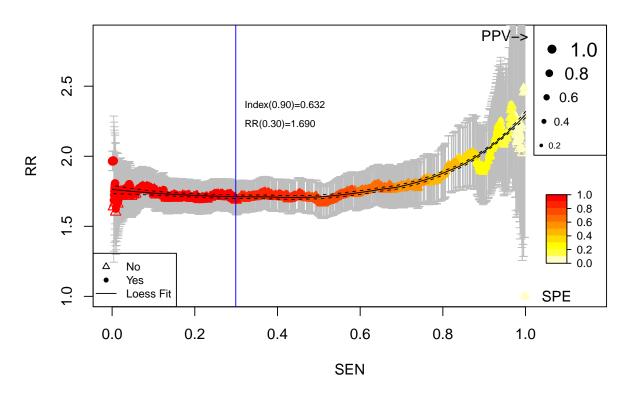
Cumulative vs. Observed: Cal. Train: Breast Cancer



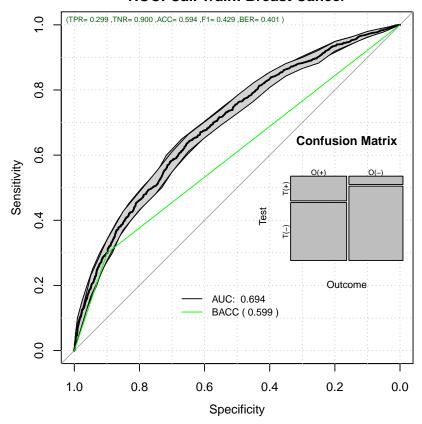
Decision Curve Analysis: Cal. Train: Breast Cancer



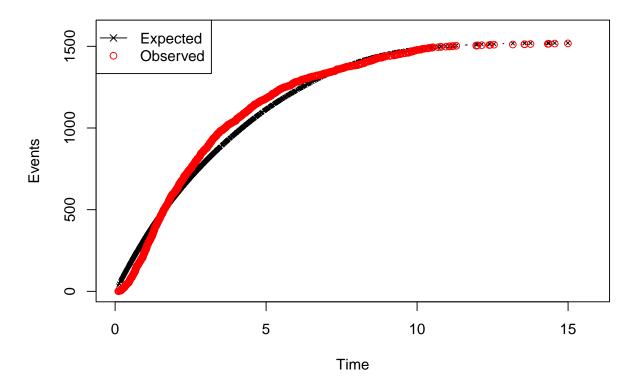
Relative Risk: Cal. Train: Breast Cancer



ROC: Cal. Train: Breast Cancer

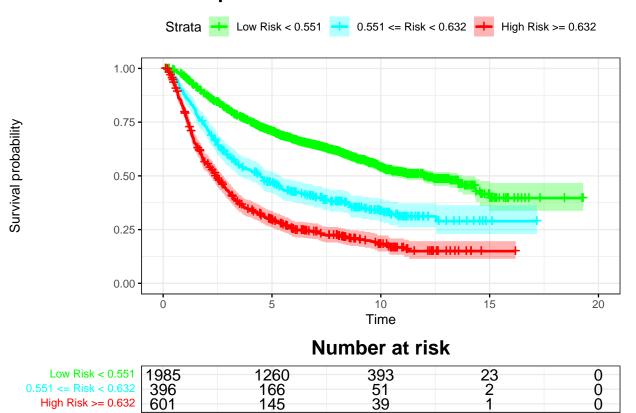


Time vs. Events: Cal. Train: Breast Cancer



Kaplan-Meier: Cal. Train: Breast Cancer

0



1.4.5 Calibrated Train Performance

396

601

0.551 <= Risk < 0.632

High Risk >= 0.632

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

Table 19: Threshold values

	@:0.9	0:0.8	@MAX_BACC	@MAX_RR	@SPE100
Thr	0.632	0.551	0.466	0.324	0.28381
$\mathbf{R}\mathbf{R}$	1.690	1.713	1.799	2.376	1.00000
\mathbf{SEN}	0.299	0.462	0.644	0.965	1.00000
\mathbf{SPE}	0.900	0.798	0.646	0.125	0.00137
\mathbf{BACC}	0.599	0.630	0.645	0.545	0.50068

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

Table 20: O/E Ratio

est	lower	upper
0.998	0.949	1.05

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

Table 21: O/E Ratio

mean	50%	2.5%	97.5%
0.977	0.977	0.969	0.985

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

Table 22: O/Acum Ratio

mean	50%	2.5%	97.5%
1.01	1.01	1.01	1.01

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

mean.C Index	median	lower	upper
0.676	0.676	0.663	0.69

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

Table 24: ROC AUC

est	lower	upper
0.694	0.675	0.713

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

Table 25: Sensitivity

est	lower	upper
0.299	0.276	0.323

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

Table 26: Specificity

est	lower	upper
0.9	0.883	0.915

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

Table 27: Probability Thresholds

90%	80%	at_max_BACC	at $_{\rm max}$ RR	atSPE100
0.632	0.551	0.466	0.324	0.284

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

Table 28: Risk Ratio

est	lower	upper
1.69	1.59	1.8

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

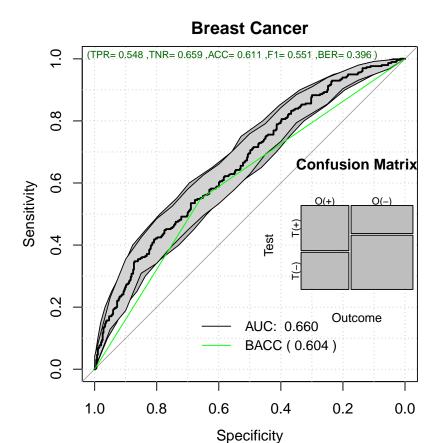
Table 29: Logrank test Chisq = 465.079317 on 2 degrees of freedom, p = 0.000000

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
class=0	1985	816	1144	93.9	385.7
class=1	396	248	177	28.0	31.8
class=2	601	454	197	336.3	391.3

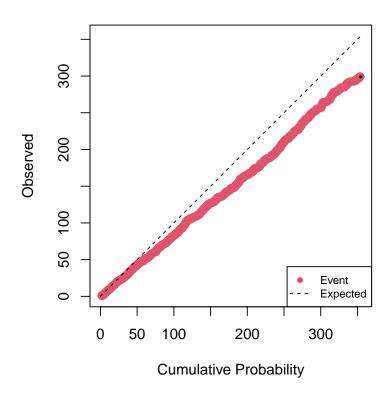
1.5 Performance on the external data set

index <- predict(ml,dataBrestCancerTest)</pre>

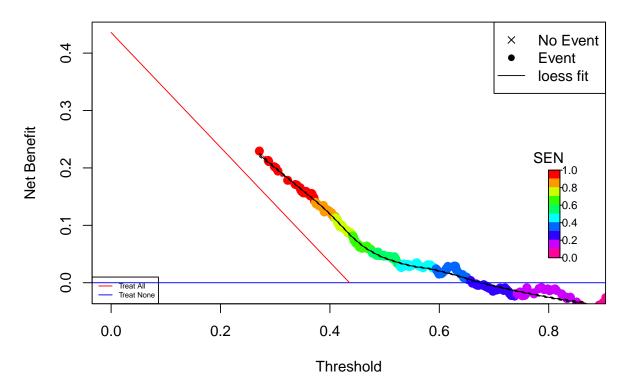
pp <- predictionStats_binary(cbind(dataBrestCancerTest\$status,index),plotname="Breast Cancer")</pre>



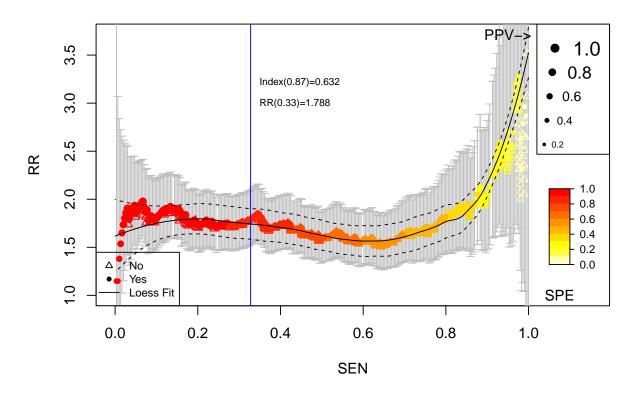
Cumulative vs. Observed: Test: Breast Cancer



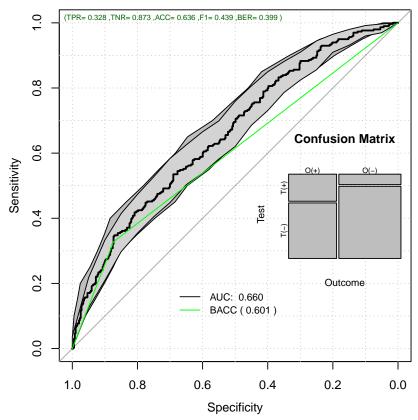
Decision Curve Analysis: Test: Breast Cancer



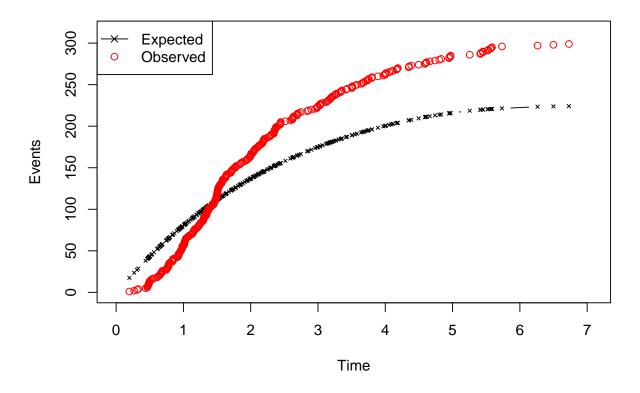
Relative Risk: Test: Breast Cancer



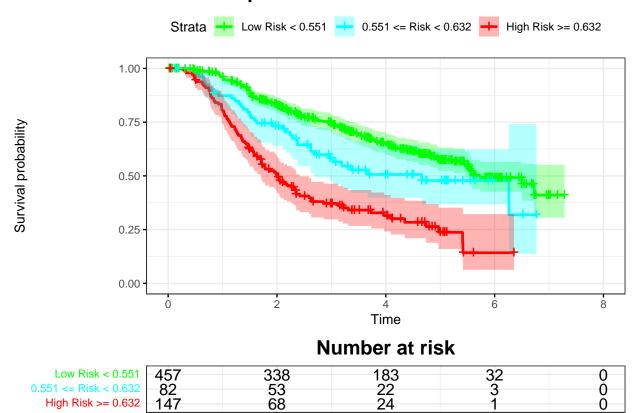




Time vs. Events: Test: Breast Cancer



Kaplan-Meier: Test: Breast Cancer



par(op)

1.5.1 External Data Report

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

Table 30: Threshold values (continued below)

	@: 0.631504143125695	@: 0.551358546734404	@: 0.466097007699806
\mathbf{Thr}	0.631	0.552	0.466
$\mathbf{R}\mathbf{R}$	1.799	1.643	1.589
\mathbf{SEN}	0.331	0.452	0.622
\mathbf{SPE}	0.873	0.757	0.579
\mathbf{BACC}	0.602	0.604	0.600

Table 31: Table continues below

	@: 0.323887654054491	@: 0.283809391129424	@MAX_BACC	@MAX_RR
\mathbf{Thr}	0.3227	0.2846	0.583	0.337
$\mathbf{R}\mathbf{R}$	2.6254	4.4083	1.758	3.279
\mathbf{SEN}	0.9799	0.9967	0.418	0.977
\mathbf{SPE}	0.0749	0.0233	0.809	0.111

	@:0.323887654054491	@:0.283809391129424	@MAX_BACC	@MAX_RR
BACC	0.5274	0.5100	0.613	0.544

	@SPE100
Thr	0.2710
$\mathbf{R}\mathbf{R}$	26.3824
\mathbf{SEN}	1.0000
\mathbf{SPE}	0.0155
\mathbf{BACC}	0.5078

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

Table 33: O/E Ratio

est	lower	upper
1.33	1.19	1.49

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

• C Index: 0.664

Dxy: 0.328S.D.: 0.0311

• n: 686

• missing: θ

• uncensored: 299

• Relevant Pairs: 266144

• Concordant: 176737

• Uncertain: 203702

• cstatCI:

mean.C Index	median	lower	upper
0.664	0.664	0.632	0.698

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

Table 35: ROC AUC

est	lower	upper
0.66	0.619	0.7

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

Table 36: Sensitivity

est	lower	upper
0.328	0.275	0.384

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

Table 37: Specificity

est	lower	upper
0.873	0.836	0.905

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

Table 38: Probability Thresholds (continued below)

90%	80%	at_max_BACC	at_max_RR	atSPE100	at_max_BACC	at_max_RR
0.632	0.551	0.466	0.324	0.284	0.583	0.337

atSPE100 0.271

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

Table 40: Risk Ratio

est	lower	upper
1.79	1.53	2.09

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

Table 41: Logrank test Chisq = 81.471750 on 2 degrees of freedom, p = 0.000000

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
class=0	457	164	221.4	14.888	58.181
class=1	82	37	33.2	0.438	0.494
class=2	147	98	44.4	64.710	77.254

1.5.2 Calibrating the index on the test data

calprob <- CoxRiskCalibration(ml,dataBrestCancerTest,"status","time")</pre>

pander::pander(c(h0=calprob\$h0,

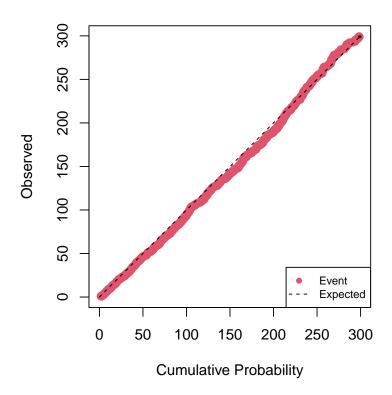
Gain=calprob\$hazardGain,

DeltaTime=calprob\$timeInterval),

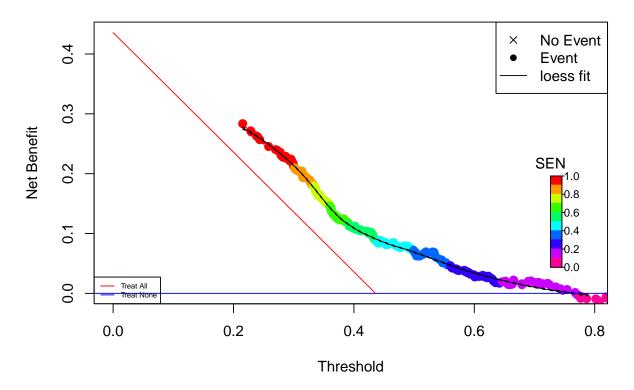
caption="Cox Calibration Parameters")

h0	Gain	DeltaTime
0.535	0.925	4.87

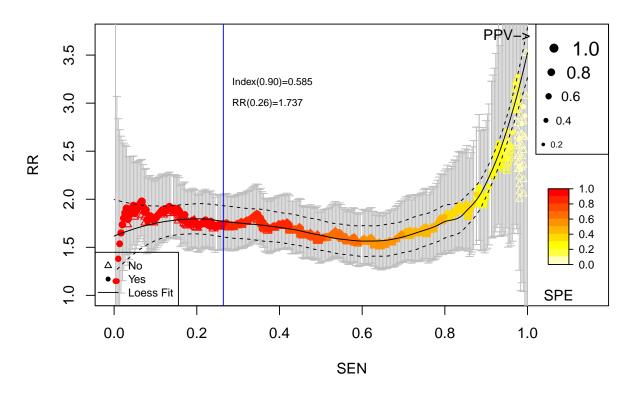
Cumulative vs. Observed: Cal. Test: Breast Cancer



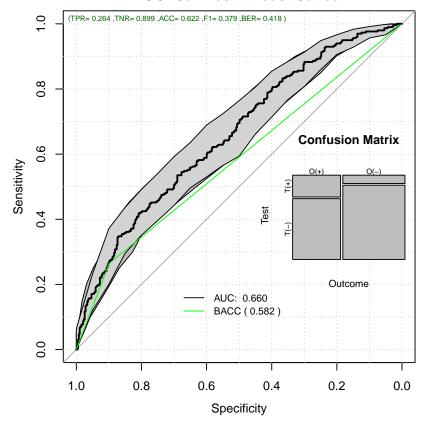
Decision Curve Analysis: Cal. Test: Breast Cancer



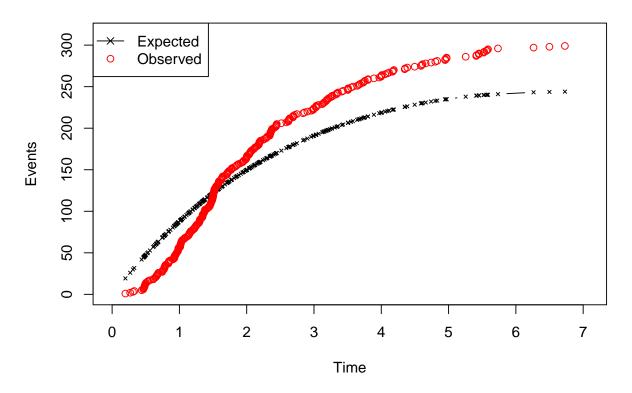
Relative Risk: Cal. Test: Breast Cancer



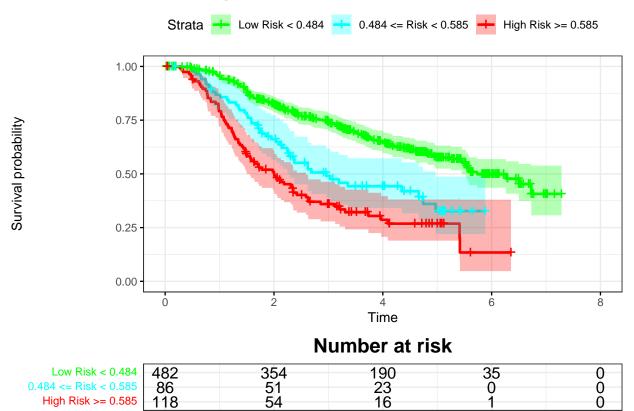
ROC: Cal. Test: Breast Cancer



Time vs. Events: Cal. Test: Breast Cancer



Kaplan-Meier: Cal. Test: Breast Cancer



1.5.3 After Calibration Report

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

Table 43: Threshold values

	@:0.9	@:0.8	@MAX_BACC	@MAX_RR	@SPE100
Thr	0.584	0.484	0.489	0.270	0.2152
$\mathbf{R}\mathbf{R}$	1.741	1.721	1.758	3.279	26.3824
\mathbf{SEN}	0.268	0.421	0.418	0.977	1.0000
\mathbf{SPE}	0.899	0.798	0.809	0.111	0.0155
\mathbf{BACC}	0.583	0.610	0.613	0.544	0.5078

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

Table 44: O/E Ratio

est	lower	upper
1.23	1.09	1.37

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

• C Index: 0.664

• **Dxy**: 0.328

• **S.D.**: 0.0311

• n: 686

• missing: θ

• uncensored: 299

Relevant Pairs: 266144
 Concordant: 176737
 Uncertain: 203702

• cstatCI:

mean.C Index	median	lower	upper
0.664	0.664	0.634	0.692

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

Table 46: ROC AUC

est	lower	upper
0.66	0.619	0.7

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

Table 47: Sensitivity

est	lower	upper
0.264	0.215	0.318

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

Table 48: Specificity

est	lower	upper
0.899	0.865	0.927

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

Table 49: Probability Thresholds

90%	80%	at_max_BACC	at_max_RR	atSPE100
0.585	0.484	0.489	0.27	0.215

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

Table 50: Risk Ratio

est	lower	upper
1.74	1.48	2.05

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

Table 51: Logrank test Chisq = 80.835092 on 2 degrees of freedom, p = 0.000000

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
class=0	482	173	232.4	15.20	69.5
class=1	86	47	32.0	7.02	7.9
class=2	118	79	34.6	57.14	65.4

1.6 Logistic Model

Here we train a logistic model on the same data set

```
## Only label subjects that present event withing five years

dataBrestCancerR <- subset(dataBrestCancerTrain, time>=5 | status==1)

dataBrestCancerR$status <- dataBrestCancerR$status * (dataBrestCancerR$time < 5)

dataBrestCancerR$time <- NULL

#ml <- BSWiMS.model(status~1, data=dataBrestCancerR, loops=20, NumberofRepeats = 5)

mlog <- BSWiMS.model(status~1, data=dataBrestCancerR, loops=1, NumberofRepeats = 5)</pre>
```

sm <- summary(mlog)
pander::pander(sm\$coefficients)</pre>

Table 52: Table continues below

	Estimate	lower	OR	upper	u.Accuracy	r.Accuracy
size_nodes	1.05e-03	1.001	1.001	1.001	0.669	0.571
\mathbf{nodes}	4.33e-02	1.040	1.044	1.048	0.676	0.634
${\it grade_nodes}$	1.50e-02	1.014	1.015	1.016	0.682	0.637
age_nodes	1.06e-03	1.001	1.001	1.001	0.678	0.653
${f size_grade}$	1.75 e-03	1.001	1.002	1.002	0.632	0.682
$\mathbf{age}\mathbf{_size}$	8.73e-05	1.000	1.000	1.000	0.608	0.682
grade	2.27e-01	1.168	1.254	1.347	0.571	0.683
age_meno	-6.04e-03	0.992	0.994	0.996	0.571	0.676
age_pgr	-5.42e-06	1.000	1.000	1.000	0.571	0.686
${f age_grade}$	-1.65e-03	0.997	0.998	0.999	0.574	0.690
${ m meno_grade}$	1.02e-01	1.045	1.107	1.173	0.571	0.683
$nodes_hormon$	-1.38e-02	0.979	0.986	0.994	0.587	0.688
${f size}$	3.94e-03	1.002	1.004	1.006	0.611	0.693
${ m meno_pgr}$	3.19e-04	1.000	1.000	1.001	0.571	0.687

	Estimate	lower	OR	upper	u.Accuracy	r.Accuracy
pgr	-1.07e-04	1.000	1.000	1.000	0.571	0.689
$meno_nodes$	-2.60e-02	0.955	0.974	0.994	0.640	0.686
${f grade_pgr}$	-3.51e-05	1.000	1.000	1.000	0.571	0.669
$meno_size$	2.34e-03	1.000	1.002	1.004	0.604	0.691

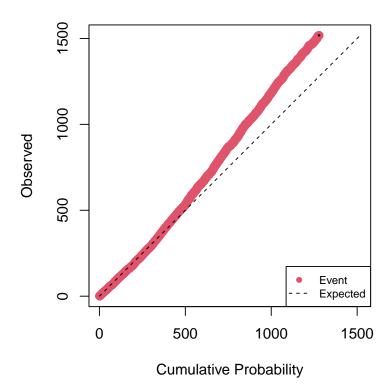
Table 53: Table continues below

	full.Accuracy	u.AUC	r.AUC	full.AUC	IDI
size_nodes	0.668	0.627	0.500	0.628	0.11233
${f nodes}$	0.690	0.639	0.621	0.662	0.07110
${f grade_nodes}$	0.686	0.649	0.624	0.655	0.06580
${f age_nodes}$	0.686	0.642	0.621	0.657	0.03346
${f size_grade}$	0.686	0.626	0.646	0.655	0.01787
$\mathbf{age}\mathbf{_size}$	0.686	0.577	0.649	0.657	0.01534
${f grade}$	0.690	0.500	0.653	0.662	0.01340
age_meno	0.686	0.500	0.645	0.657	0.00782
${f age_pgr}$	0.686	0.500	0.656	0.657	0.00512
${f age_grade}$	0.690	0.507	0.661	0.662	0.00454
${f meno_grade}$	0.686	0.500	0.652	0.657	0.00425
${f nodes_hormon}$	0.686	0.526	0.658	0.655	0.00280
${f size}$	0.690	0.618	0.663	0.662	0.00507
${f meno_pgr}$	0.686	0.500	0.657	0.657	0.00316
\mathbf{pgr}	0.686	0.500	0.659	0.655	0.00257
${f meno_nodes}$	0.686	0.595	0.656	0.657	0.00264
${f grade_pgr}$	0.668	0.500	0.627	0.628	0.00241
${f meno_size}$	0.690	0.578	0.663	0.662	0.00185

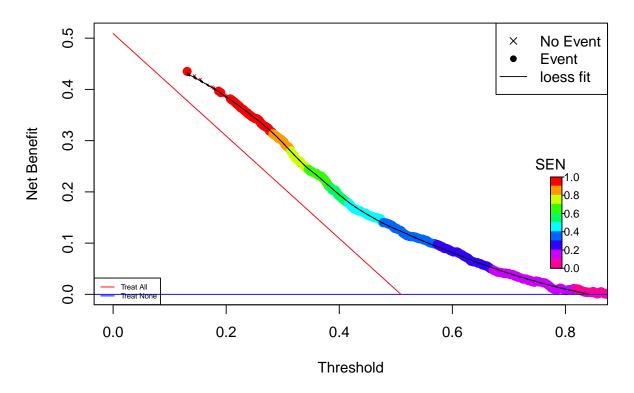
	NRI	z.IDI	z.NRI	Delta.AUC	Frequency
size_nodes	0.63654	17.86	18.870	0.128490	1
${f nodes}$	0.57106	14.13	16.179	0.040494	1
${f grade_nodes}$	0.54866	13.66	15.650	0.031087	1
${f age_nodes}$	0.21312	9.39	5.710	0.035896	1
${f size_grade}$	0.29411	6.74	7.728	0.008648	1
$\mathbf{age}\mathbf{_size}$	0.29152	6.41	7.652	0.007600	1
${f grade}$	0.19036	6.20	4.983	0.008461	1
age_meno	0.08057	4.76	2.337	0.012065	1
${f age_pgr}$	0.00745	4.11	0.194	0.000417	1
${f age_grade}$	0.11372	3.60	2.960	0.000315	1
${f meno_grade}$	0.20428	3.47	5.343	0.004441	1
${f nodes_hormon}$	0.45522	3.44	12.150	-0.002853	1
${f size}$	0.21050	3.42	5.600	-0.001075	1
${f meno_pgr}$	0.05977	3.35	1.558	-0.000429	1
\mathbf{pgr}	0.19759	2.64	5.745	-0.004123	1
${f meno_nodes}$	-0.06329	2.59	-1.645	0.000631	1
${f grade_pgr}$	0.17471	2.55	5.058	0.001252	1
meno_size	0.10227	2.43	2.662	-0.001378	1

1.7 Logistic Model Performance

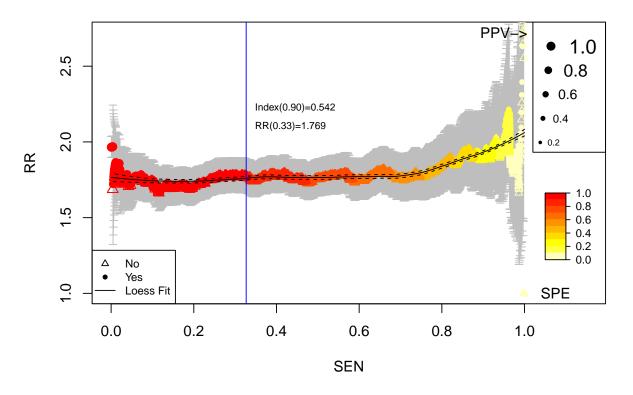
Cumulative vs. Observed: Logistic Train: Breast Cancer



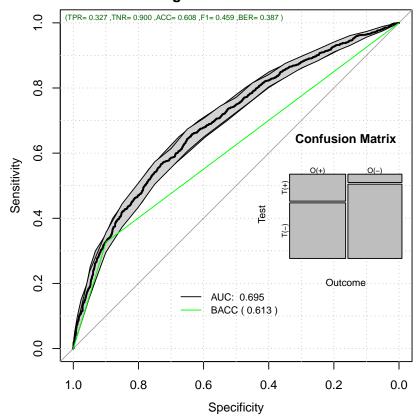
Decision Curve Analysis: Logistic Train: Breast Cancer



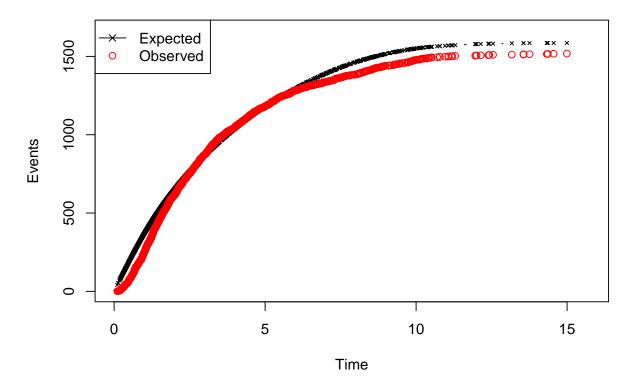
Relative Risk: Logistic Train: Breast Cancer



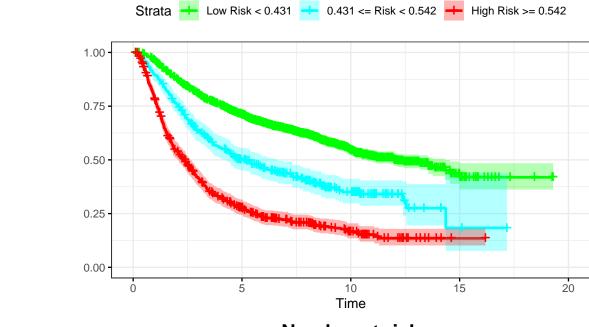
ROC: Logistic Train: Breast Cancer



Time vs. Events: Logistic Train: Breast Cancer



Kaplan-Meier: Logistic Train: Breast Cancer



Number at risk

Low Risk < 0.431	1975	1268	399	23	0
0.431 <= Risk < 0.542	364	160	47	2	0
High Risk $>= 0.542$	643	143	37	1	0

par(op)

Survival probability

1.7.1 Training Report

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

Table 55: Threshold values

	@:0.9	@:0.8	@MAX_BACC	@MAX_RR	@SPE100
Thr	0.542	0.431	0.394	0.255	0.130969
$\mathbf{R}\mathbf{R}$	1.765	1.739	1.799	2.213	1.000000
\mathbf{SEN}	0.327	0.470	0.566	0.962	1.000000
\mathbf{SPE}	0.900	0.799	0.731	0.125	0.000683
\mathbf{BACC}	0.613	0.635	0.648	0.543	0.500342

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

Table 56: O/E Ratio

est	lower	upper
0.957	0.91	1.01

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

• C Index: 0.68

• **Dxy**: 0.36

• **S.D.**: 0.014

• n: 2982

• missing: θ

• uncensored: 1518

• Relevant Pairs: 6184528

• Concordant: 4206590

• Uncertain: 2703838

• cstatCI:

mean.C Index	median	lower	upper
0.68	0.68	0.667	0.694

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

Table 58: ROC AUC

est	lower	upper
0.695	0.677	0.714

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

Table 59: Sensitivity

est	lower	upper
0.327	0.303	0.351

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

Table 60: Specificity

est	lower	upper
0.9	0.883	0.915

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

Table 61: Probability Thresholds

90%	80%	at_max_BACC	at_max_RR	atSPE100
0.542	0.431	0.394	0.255	0.131

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

Table 62: Risk Ratio

est	lower	upper
1.77	1.66	1.88

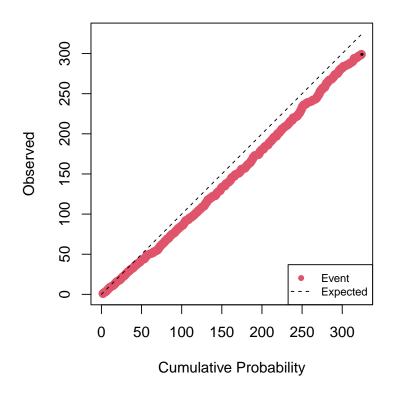
pander::pander(rrAnalysisTrain\$surdif,caption="Logrank test")

Table 63: Logrank test Chisq = 543.347175 on 2 degrees of freedom, p = 0.000000

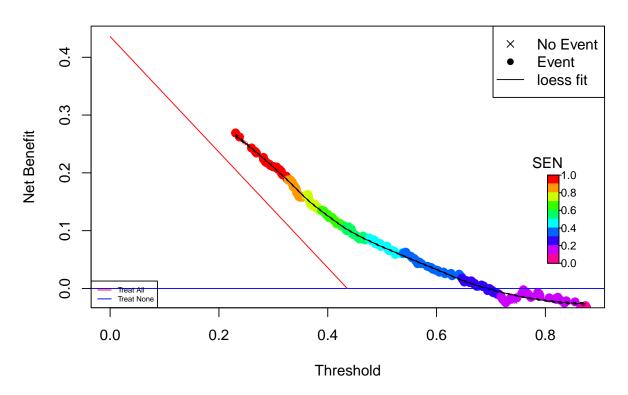
	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
class=0	1975	804	1145	101.5	418.9
class=1	364	218	169	14.1	15.9
class=2	643	496	204	418.2	490.7

1.7.2 Results on the validation set using Logistic model

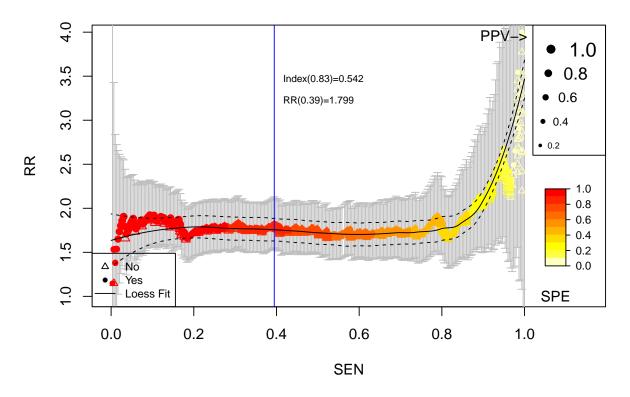
Cumulative vs. Observed: Logistic Test: Breast Cancer



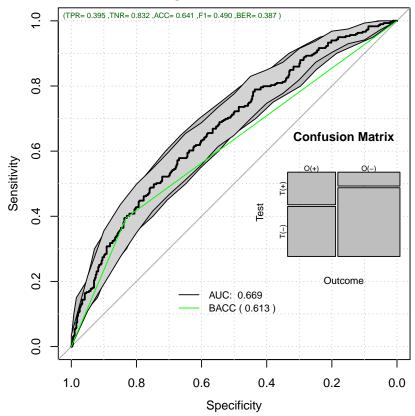
Decision Curve Analysis: Logistic Test: Breast Cancer



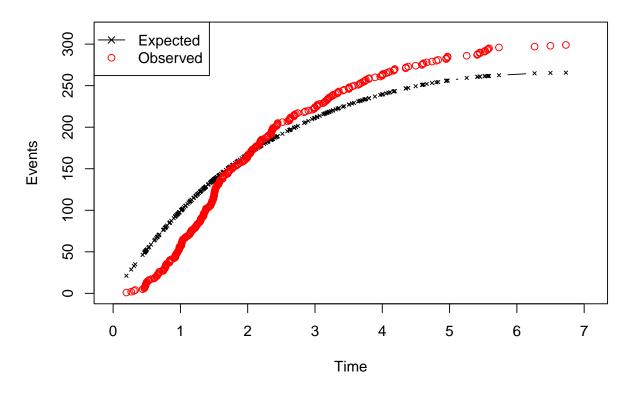
Relative Risk: Logistic Test: Breast Cancer



ROC: Logistic Test: Breast Cancer

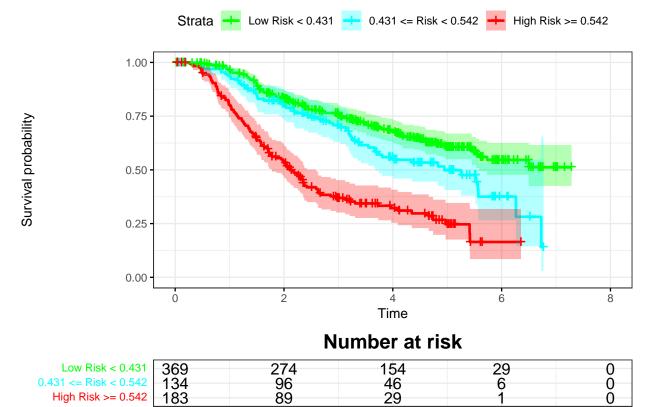


Time vs. Events: Logistic Test: Breast Cancer



Kaplan-Meier: Logistic Test: Breast Cancer

0



par(op)

1.7.3 Validation Report

0.431 <= Risk < 0.542

High Risk >= 0.542

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

Table 64: Threshold values (continued below)

	@: 0.541514635560692	@: 0.431407973270606	@: 0.394073801020846
Thr	0.542	0.431	0.394
$\mathbf{R}\mathbf{R}$	1.792	1.702	1.746
SEN	0.395	0.595	0.686
\mathbf{SPE}	0.832	0.638	0.545
\mathbf{BACC}	0.613	0.617	0.615

Table 65: Table continues below

	@: 0.255194040798287	@: 0.130968771182643	@MAX_BACC	@MAX_RR
Thr	0.255	0.2306	0.439	0.306
$\mathbf{R}\mathbf{R}$	3.094	21.9530	1.756	2.678
\mathbf{SEN}	0.993	1.0000	0.579	0.950
\mathbf{SPE}	0.031	0.0129	0.669	0.181

	@:0.255194040798287	@:0.130968771182643	@MAX_BACC	@MAX_RR
BACC	0.512	0.5065	0.624	0.565

	@SPE100
Thr	0.2306
$\mathbf{R}\mathbf{R}$	21.9530
\mathbf{SEN}	1.0000
\mathbf{SPE}	0.0129
\mathbf{BACC}	0.5065

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

Table 67: O/E Ratio

est	lower	upper
1.13	1	1.26

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

• C Index: 0.669

Dxy: 0.338S.D.: 0.0309

• n: 686

• missing: θ

• uncensored: 299

• Relevant Pairs: 266144

Concordant: 178115Uncertain: 203702

• cstatCI:

mean.C Index	median	lower	upper
0.669	0.67	0.639	0.699

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

Table 69: ROC AUC

est	lower	upper
0.669	0.628	0.709

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

Table 70: Sensitivity

est	lower	upper
0.395	0.339	0.453

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

Table 71: Specificity

est	lower	upper
0.832	0.791	0.868

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

Table 72: Probability Thresholds (continued below)

90%	80%	at_max_BACC	at_max_RR	atSPE100	at_max_BACC	at_max_RR
0.542	0.431	0.394	0.255	0.131	0.439	0.306

atSPE100	
0.231	

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

Table 74: Risk Ratio

est	lower	upper
1.8	1.54	2.11

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

Table 75: Logrank test Chisq = 92.507991 on 2 degrees of freedom, p = 0.000000

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
class=0	369	121	181.7	20.2997	52.3868
class=1	134	60	61.7	0.0479	0.0604
class=2	183	118	55.5	70.2342	88.0195

1.8 Logistic Model Poisson Calibration

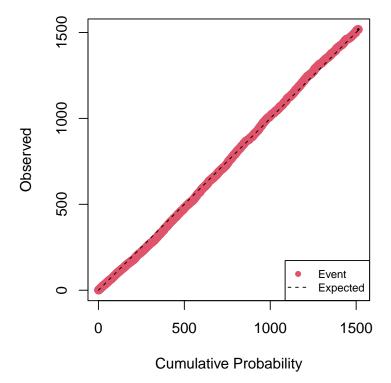
riskdata <- cbind(dataBrestCancerTrain\$status,predict(mlog,dataBrestCancerTrain,type="prob"),dataBrestCalprob <- CalibrationProbPoissonRisk(riskdata)</pre>

pander::pander(c(h0=calprob\$h0,

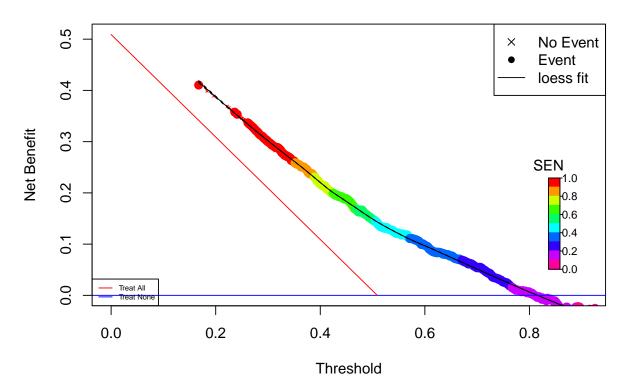
```
Gain=calprob$hazardGain,
   DeltaTime=calprob$timeInterval),
caption="Logistic Calibration Parameters")
```

1-0	O-:	D-14- Ti
no	Gain	DeltaTime
0.676	1.31	7.14

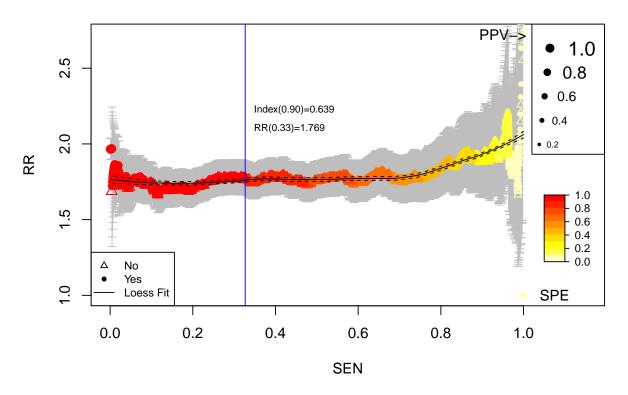
Cumulative vs. Observed: Cal. Logistic Train: Breast Cancer



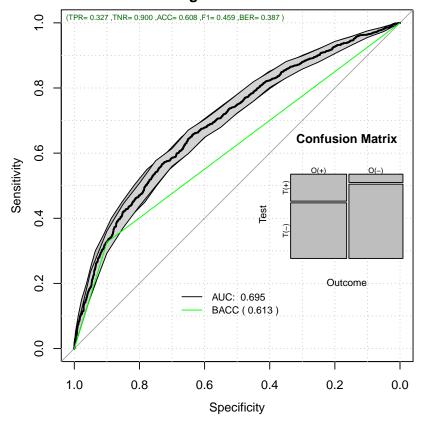
Decision Curve Analysis: Cal. Logistic Train: Breast Cancer



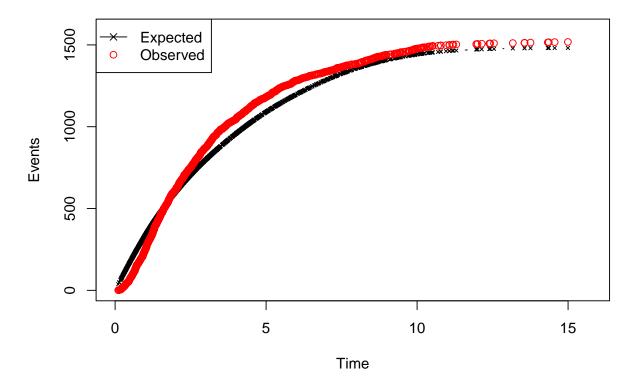
Relative Risk: Cal. Logistic Train: Breast Cancer



ROC: Cal. Logistic Train: Breast Cancer

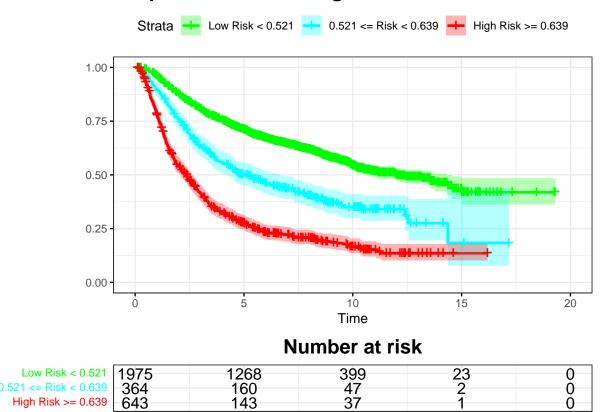


Time vs. Events: Cal. Logistic Train: Breast Cancer



Kaplan-Meier: Cal. Logistic Train: Breast Cancer

0



par(op)

0.521 <= Risk < 0.639

High Risk >= 0.639

Survival probability

1.8.1 Report of the calibrated logistic: training

364 643

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

Table 77: Threshold values

	@:0.9	@:0.8	@MAX_BACC	@MAX_RR	@SPE100
Thr	0.639	0.521	0.480	0.319	0.167426
$\mathbf{R}\mathbf{R}$	1.765	1.739	1.799	2.213	1.000000
\mathbf{SEN}	0.327	0.470	0.566	0.962	1.000000
\mathbf{SPE}	0.900	0.799	0.731	0.125	0.000683
\mathbf{BACC}	0.613	0.635	0.648	0.543	0.500342

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

Table 78: O/E Ratio

est	lower	upper
1.02	0.974	1.08

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

• C Index: 0.68

• **Dxy**: 0.36

• **S.D.**: 0.014

• n: 2982

• missing: θ

• uncensored: 1518

• Relevant Pairs: 6184528

• Concordant: 4206588

• Uncertain: 2703838

• cstatCI:

mean.C Index	median	lower	upper
0.68	0.68	0.666	0.693

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

Table 80: ROC AUC

est	lower	upper
0.695	0.677	0.714

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

Table 81: Sensitivity

est	lower	upper
0.327	0.303	0.351

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

Table 82: Specificity

est	lower	upper
0.9	0.883	0.915

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

Table 83: Probability Thresholds

90%	80%	at_max_BACC	at $_{\rm max}$ RR	atSPE100
0.639	0.521	0.48	0.319	0.167

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

Table 84: Risk Ratio

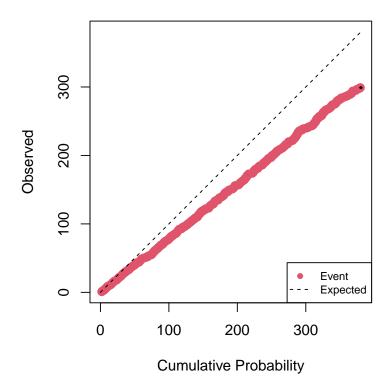
est	lower	upper
1.77	1.66	1.88

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

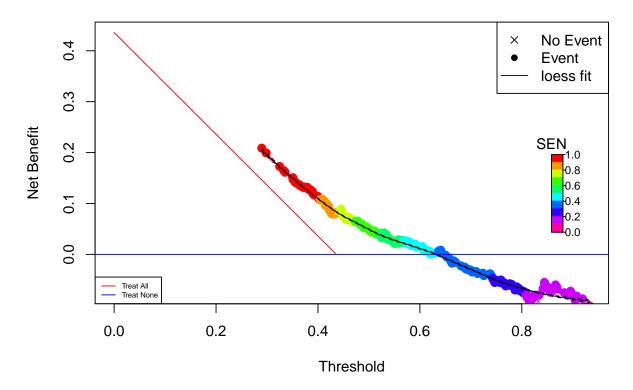
Table 85: Logrank test Chisq = 543.347175 on 2 degrees of freedom, p = 0.000000

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
class=0	1975	804	1145	101.5	418.9
class=1	364	218	169	14.1	15.9
class=2	643	496	204	418.2	490.7

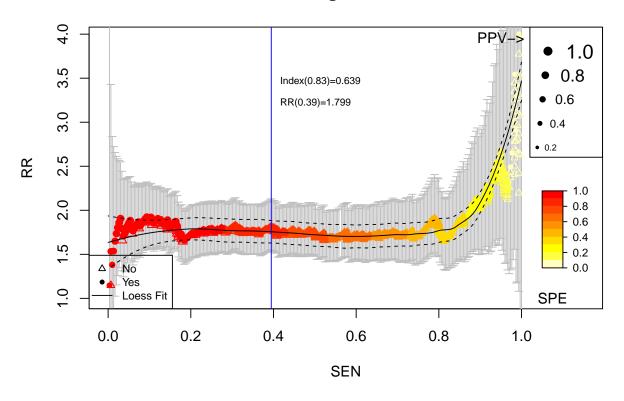
Cumulative vs. Observed: Cal. Logistic Test: Breast Cancer



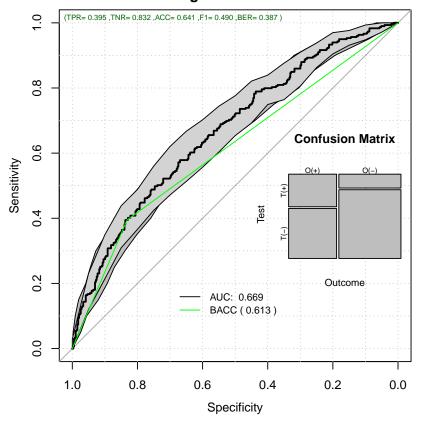
Decision Curve Analysis: Cal. Logistic Test: Breast Cancer



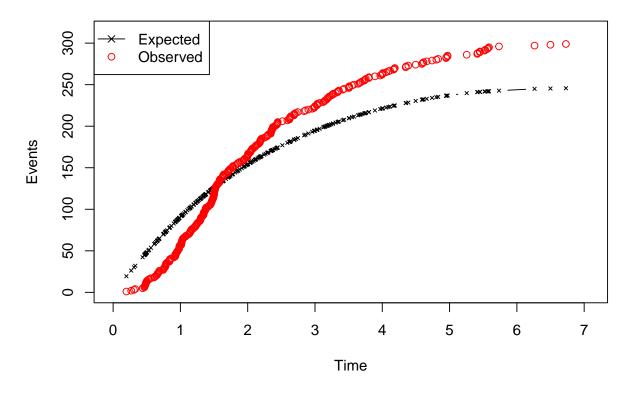
Relative Risk: Cal. Logistic Test: Breast Cancer



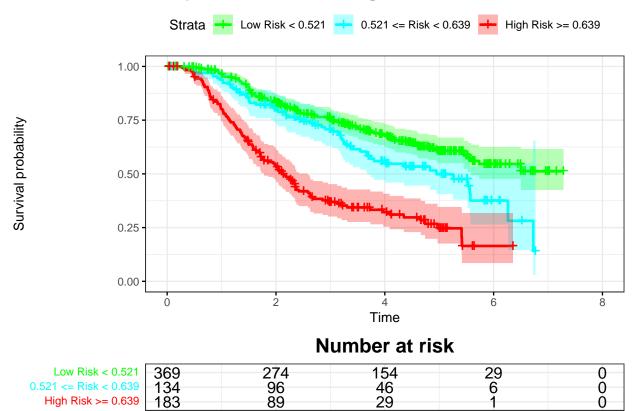
ROC: Cal. Logistic Test: Breast Cancer



Time vs. Events: Cal. Logistic Test: Breast Cancer



Kaplan-Meier: Cal. Logistic Test: Breast Cancer



par(op)

1.8.2 Report of the calibrated validation

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

Table 86: Threshold values (continued below)

	@: 0.63864953482035	@: 0.521434578375582	@: 0.480013144733203
Thr	0.639	0.521	0.480
$\mathbf{R}\mathbf{R}$	1.792	1.702	1.746
SEN	0.395	0.595	0.686
\mathbf{SPE}	0.832	0.638	0.545
\mathbf{BACC}	0.613	0.617	0.615

Table 87: Table continues below

	@: 0.319265361359366	@: 0.167425979697821	@MAX_BACC	@MAX_RR
Thr	0.320	0.2897	0.529	0.379
$\mathbf{R}\mathbf{R}$	3.094	21.9530	1.756	2.678
\mathbf{SEN}	0.993	1.0000	0.579	0.950
\mathbf{SPE}	0.031	0.0129	0.669	0.181

	@:0.319265361359366	@:0.167425979697821	@MAX_BACC	@MAX_RR
BACC	0.512	0.5065	0.624	0.565

@SPE100
0.2897
21.9530
1.0000
0.0129
0.5065

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

Table 89: O/E Ratio

est	lower	upper
1.22	1.08	1.36

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

• C Index: 0.669

Dxy: 0.338S.D.: 0.0309

• **n**: 686

• missing: θ

• uncensored: 299

• Relevant Pairs: 266144

• Concordant: 178115

• Uncertain: 203702

• cstatCI:

mean.C Index	median	lower	upper
0.669	0.67	0.639	0.701

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

Table 91: ROC AUC

est	lower	upper
0.669	0.628	0.709

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

Table 92: Sensitivity

est	lower	upper
0.395	0.339	0.453

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

Table 93: Specificity

est	lower	upper
0.832	0.791	0.868

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

Table 94: Probability Thresholds (continued below)

90%	80%	at_max_BACC	at_max_RR	atSPE100	at_max_BACC	at_max_RR
0.639	0.521	0.48	0.319	0.167	0.529	0.379

atSPE100	
0.29	

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

Table 96: Risk Ratio

est	lower	upper
1.8	1.54	2.11

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

Table 97: Logrank test Chisq = 92.507991 on 2 degrees of freedom, p = 0.000000

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
class=0	369	121	181.7	20.2997	52.3868
class=1	134	60	61.7	0.0479	0.0604
class=2	183	118	55.5	70.2342	88.0195

1.9 Comparing the COX and Logistic Models on the Independent Data

pander::pander(t(rrCoxTestAnalysis\$0Acum95ci))

mean	50%	2.5%	97.5%
0.841	0.841	0.839	0.842

pander::pander(t(rrAnalysisTestLogistic\$0Acum95ci))

mean	50%	2.5%	97.5%
0.791	0.791	0.791	0.792

pander::pander(t(rrCoxTestAnalysis\$0E95ci))

mean	50%	2.5%	97.5%
1.07	1.07	1.03	1.1

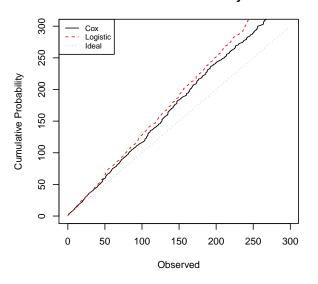
pander::pander(t(rrAnalysisTestLogistic\$0E95ci))

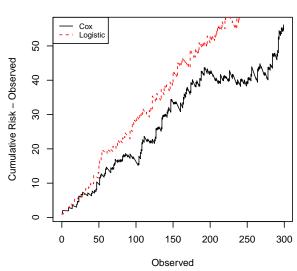
mean	50%	2.5%	97.5%
0.955	0.955	0.926	0.984

```
maxobs <- sum(dataBrestCancerTest$status)</pre>
par(mfrow=c(1,2),cex=0.75)
plot(rrCoxTestAnalysis$CumulativeOvs,type="l",lty=1,
    main="Cumulative Probability",
     xlab="Observed",
     ylab="Cumulative Probability",
    ylim=c(0,maxobs),
    xlim=c(0,maxobs))
lines(rrAnalysisTestLogistic$CumulativeOvs,lty=2,col="red")
lines(x=c(0,maxobs),y=c(0,maxobs),lty=3,col="gray")
legend("topleft",legend = c("Cox","Logistic","Ideal"),
       col=c("black","red","gray"),
       lty=c(1,2,3),
       cex=0.75
)
plot(rrCoxTestAnalysis$CumulativeOvs$Observed,
     rrCoxTestAnalysis$CumulativeOvs$Cumulative-
       rrCoxTestAnalysis$CumulativeOvs$Observed,
     main="Cumulative Risk Difference",
     xlab="Observed",
     ylab="Cumulative Risk - Observed",
     type="1",
     lty=1)
lines(rrAnalysisTestLogistic$CumulativeOvs$Observed,
    rrAnalysisTestLogistic$CumulativeOvs$Cumulative-
       rrAnalysisTestLogistic$CumulativeOvs$Observed,
```

Cumulative Probability

Cumulative Risk Difference

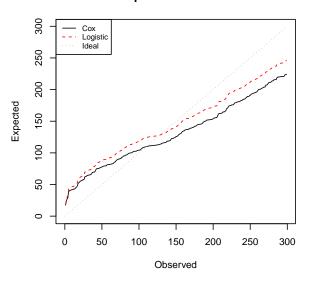


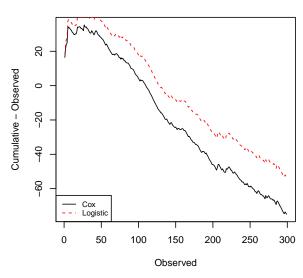


```
plot(rrCoxTestAnalysis$0EData[,2:3],type="1",lty=1,
     main="Expected over Time",
     xlab="Observed",
     ylab="Expected",
     ylim=c(0,maxobs),
     xlim=c(0,maxobs))
lines(rrAnalysisTestLogistic$OEData[,2:3],lty=2,col="red")
lines(x=c(0,maxobs),y=c(0,maxobs),lty=3,col="gray")
legend("topleft",legend = c("Cox","Logistic","Ideal"),
       col=c("black","red","gray"),
       lty=c(1,2,3),
       cex=0.75
)
plot(rrCoxTestAnalysis$0EData$0bserved,
     rrCoxTestAnalysis$OEData$Expected-
       rrCoxTestAnalysis$0EData$0bserved,
     main="Expected vs Observed Difference",
     xlab="Observed",
     ylab="Cumulative - Observed",
     type="1",
     lty=1)
lines(rrAnalysisTestLogistic$0EData$0bserved,
     rrAnalysisTestLogistic$OEData$Expected-
       rrAnalysisTestLogistic$OEData$Observed,
    lty=2,col="red")
```

Expected over Time

Expected vs Observed Difference





par(op)