Breast Cancer: Wisconsin

Jose Tamez

2023-05-11

Contents

1 Showcasing RRPlots

1.0.1 Libraries

```
library(survival)
library(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'

## 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)
pander::panderOptions('digits', 3)
#pander::panderOptions('table.split.table', 400)
```

```
pander::panderOptions('keep.trailing.zeros',TRUE)
layout(matrix(1:1, nrow=1))
```

1.0.2 Wisconsin Data Set

```
dataBreast <- read.csv("~/GitHub/RISKPLOTS/DATA/wpbc.data", header=FALSE)
table(dataBreast$V2)
```

```
##
##   N   R
## 151  47
```

```
rownames(dataBreast) <- dataBreast$V1
dataBreast$V1 <- NULL
dataBreast$status <- 1*(dataBreast$V2=="R")
dataBreast$V2 <- NULL
dataBreast$time <- dataBreast$V3
dataBreast$V3 <- NULL
dataBreast <- sapply(dataBreast,as.numeric)
```

```
## Warning in lapply(X = X, FUN = FUN, ...): NAs introduced by coercion
```

```
dataBreast <- as.data.frame(dataBreast[complete.cases(dataBreast),])
table(dataBreast$status)
```

```
##
##   0   1
## 148  46
```

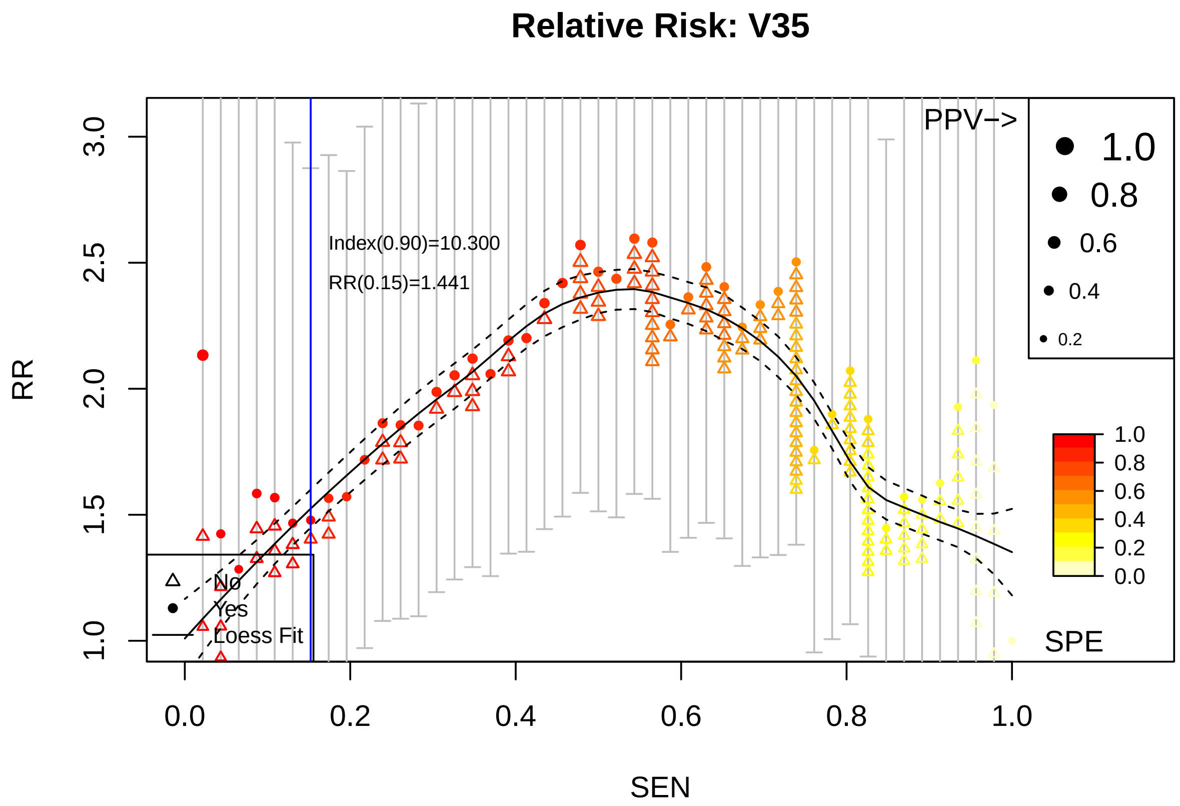
1.1 Exploring Raw Features with RRPlot

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

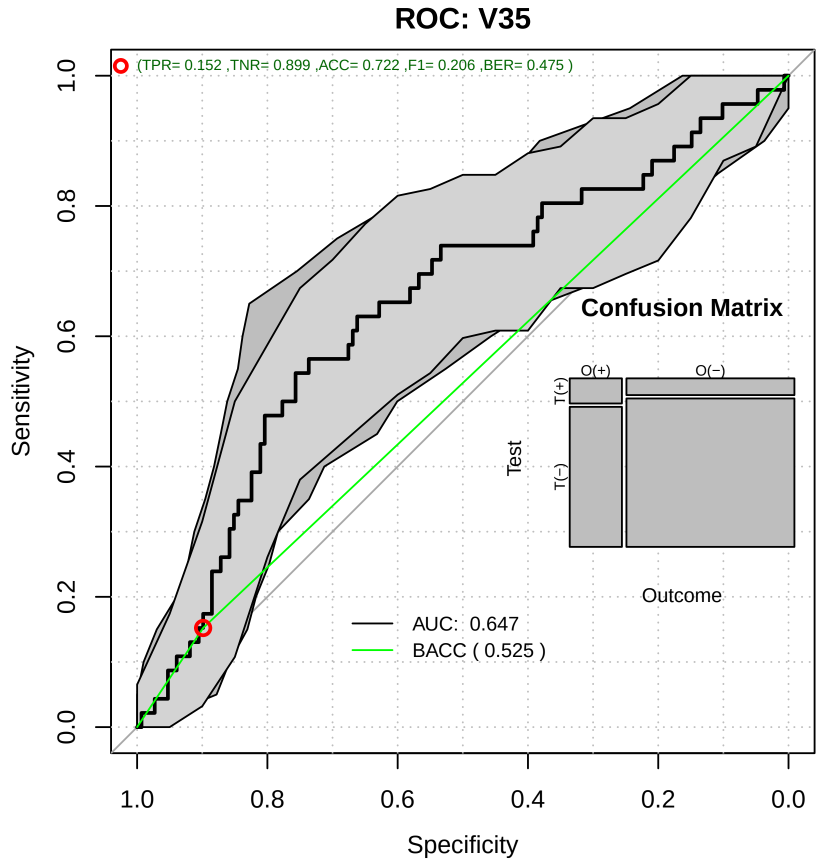
V35	V24	V34	V7	V16	V14	V17
0.0261	0.0261	0.0261	0.0623	0.126	0.126	0.126

```
topv <- min(5,length(topvar))
topFive <- names(topvar)[1:topv]
RRanalysis <- list();
idx <- 1
topf <- topFive[1]
for (topf in topFive)
{
  RRanalysis[[idx]] <- RRPlot(cbind(dataBreast$status,dataBreast[,topf]),
                              atProb=c(0.90,0.80),
                  timetoEvent=dataBreast$time,
                  title=topf,
#                  plotRR=FALSE
                  )
```

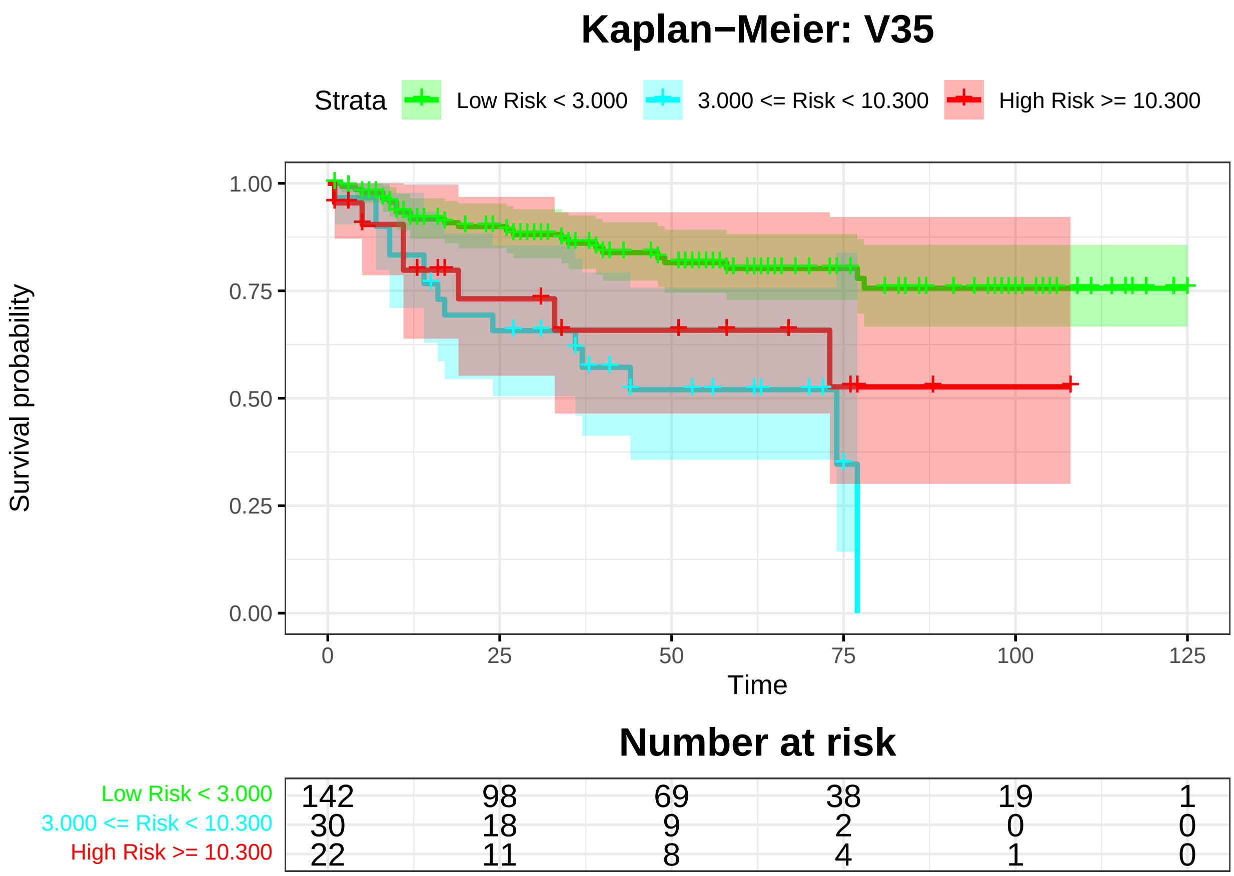
```
  idx <- idx + 1
}
```

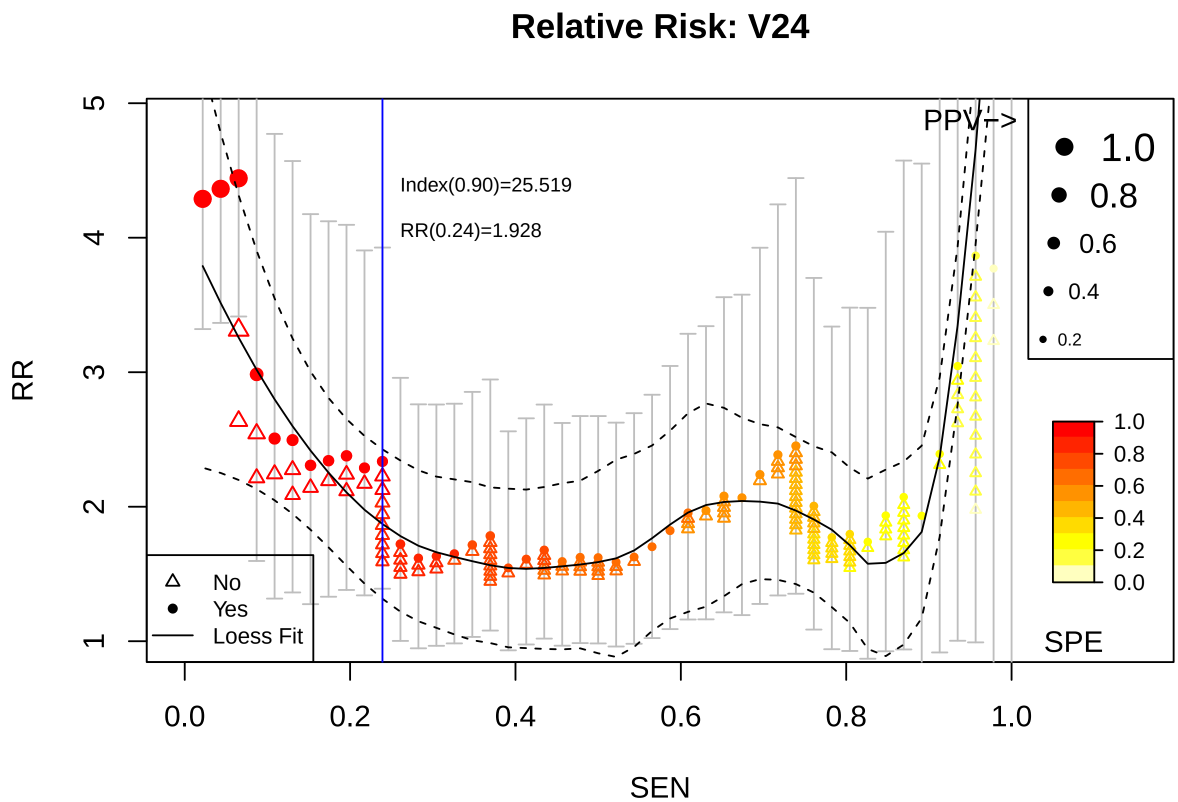


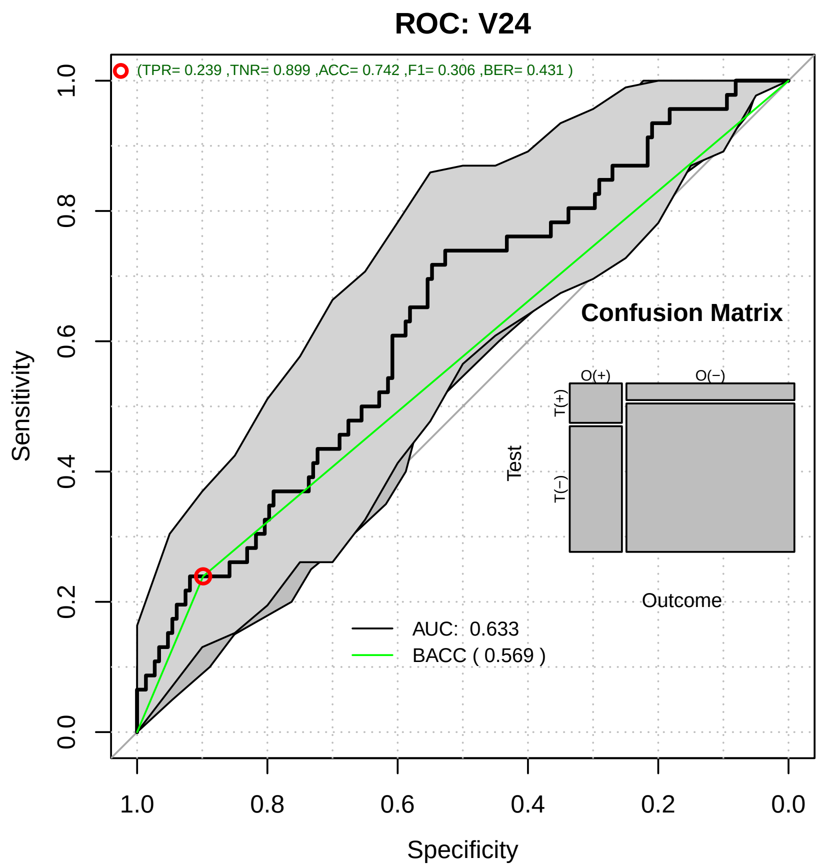
Relative Risk: V35
RR
SEN
PPV->
1.0
0.8
0.6
0.4
0.2
SPE
Index(0.90)=10.300
RR(0.15)=1.441
No
Yes
Loess Fit

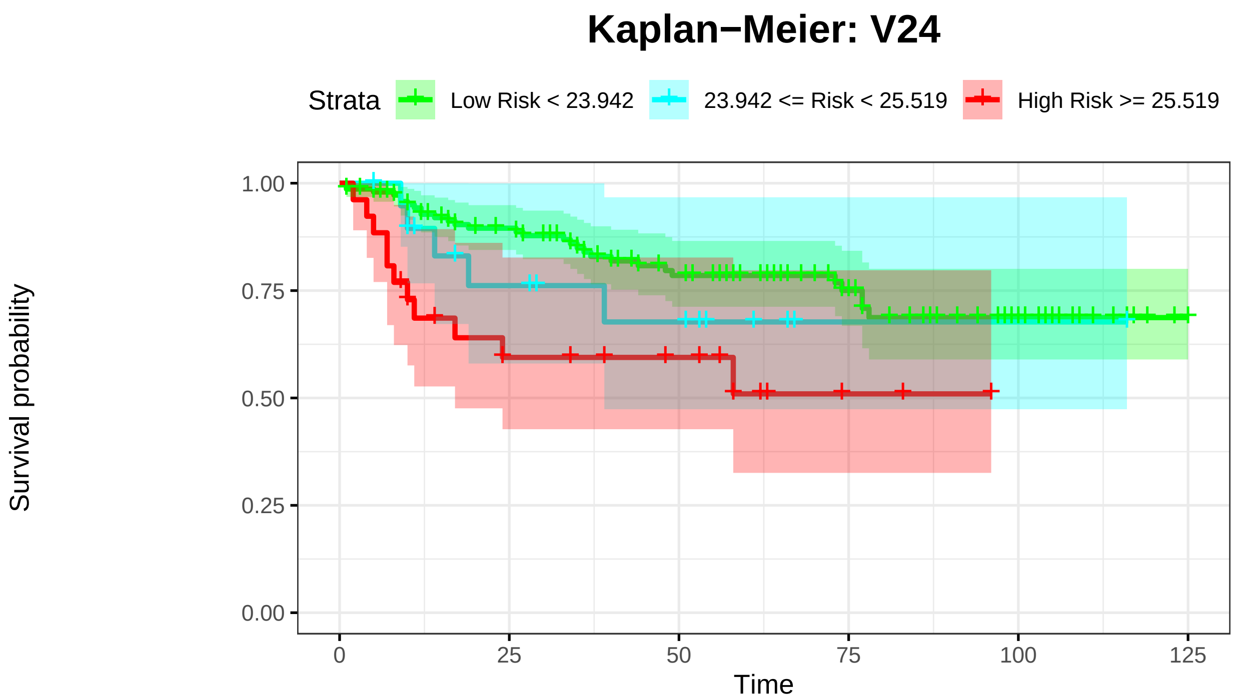


ROC: V35
(TPR= 0.152 ,TNR= 0.899 ,ACC= 0.722 ,F1= 0.206 ,BER= 0.475)
Sensitivity
Specificity
1.0
0.8
0.6
0.4
0.2
0.0
Confusion Matrix
O(+)
O(−)
T(+)
T(−)
Test
Outcome
AUC: 0.647
BACC (0.525)


Kaplan−Meier: V35
Strata
Low Risk < 3.000
3.000 <= Risk < 10.300
High Risk >= 10.300
Survival probability
1.00
0.75
0.50
0.25
0.00
0
25
50
75
100
125
Time
Number at risk
Low Risk < 3.000 142 98 69 38 19 1
3.000 <= Risk < 10.300 30 18 9 2 0 0
High Risk >= 10.300 22 11 8 4 1 0

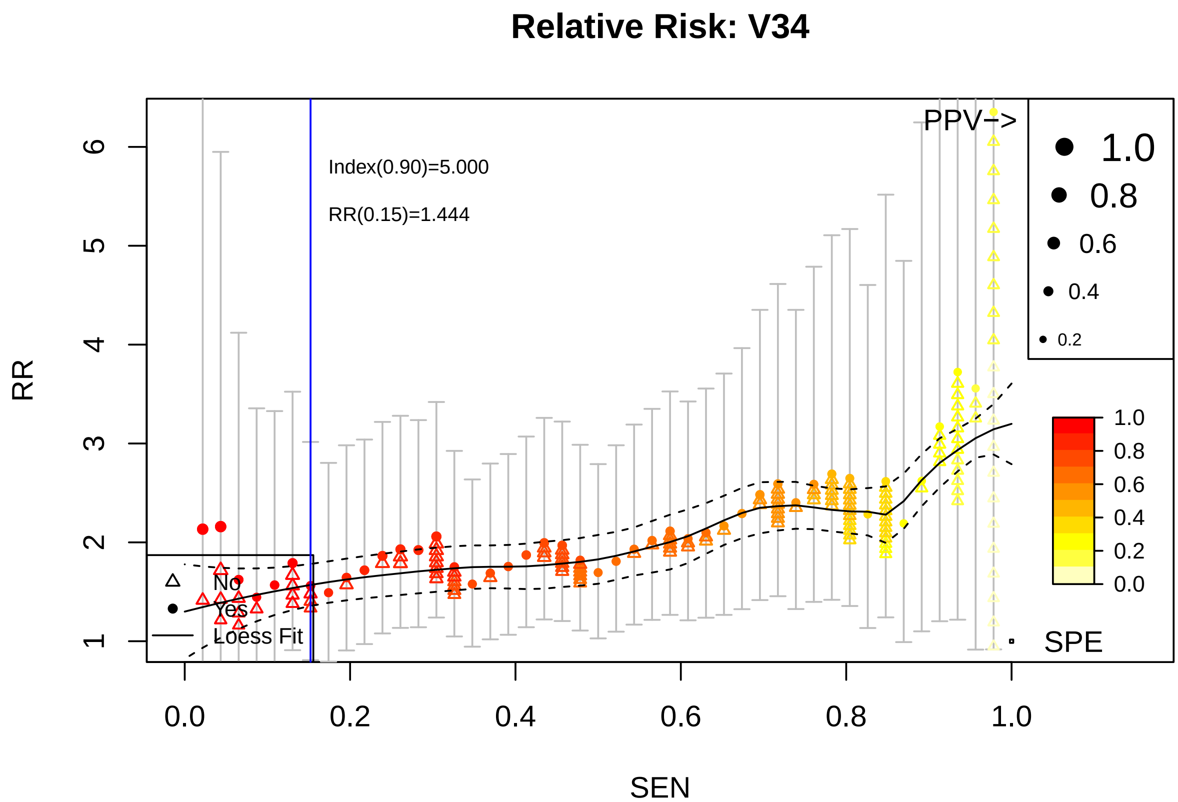

Relative Risk: V24
Index(0.90)=25.519
RR(0.24)=1.928
PPV-->
1.0
0.8
0.6
0.4
0.2
1.0
0.8
0.6
0.4
0.2
0.0
SPE
No
Yes
Loess Fit
RR
1
2
3
4
5
0.0
0.2
0.4
0.6
0.8
1.0
SEN


ROC: V24
(TPR= 0.239 ,TNR= 0.899 ,ACC= 0.742 ,F1= 0.306 ,BER= 0.431)
Sensitivity
Specificity
1.0
0.8
0.6
0.4
0.2
0.0
Confusion Matrix
O(+)
O(−)
T(+)
T(−)
Test
Outcome
AUC: 0.633
BACC (0.569)

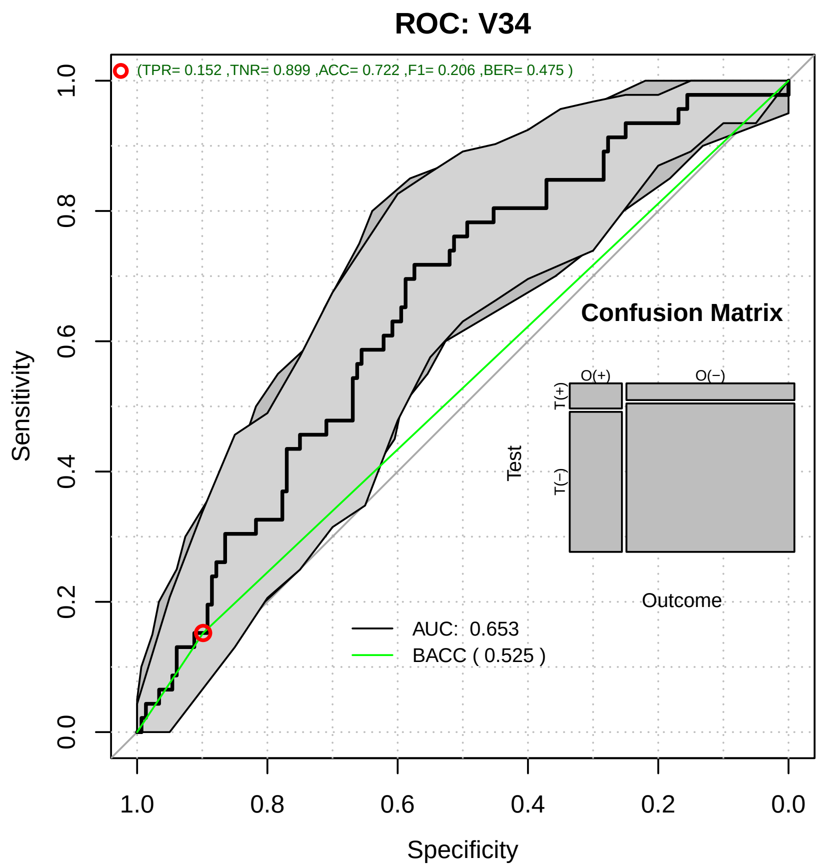


Number at risk

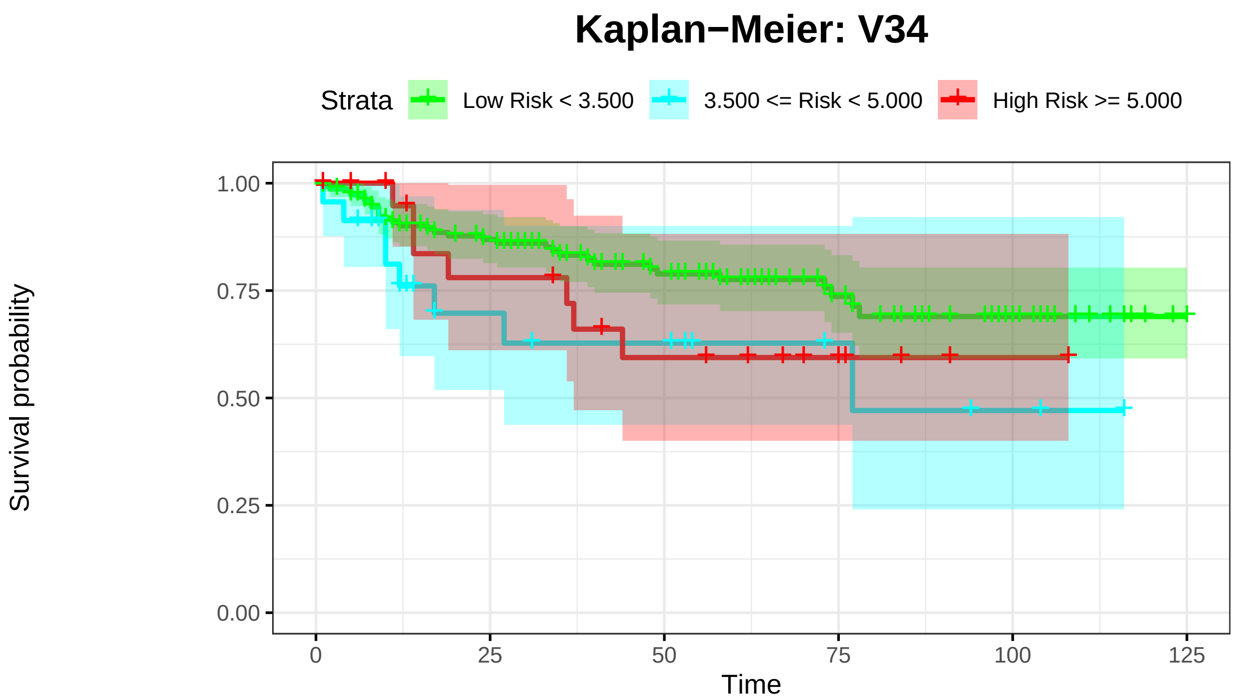
	0	25	50	75	100	125
Low Risk < 23.942	148	104	69	41	19	1
23.942 <= Risk < 25.519	20	11	8	1	1	0
High Risk >= 25.519	26	12	9	2	0	0



Relative Risk: V34
RR
SEN
PPV−>
SPE
Index(0.90)=5.000
RR(0.15)=1.444
No
Yes
Loess Fit
1.0
0.8
0.6
0.4
0.2
0.0

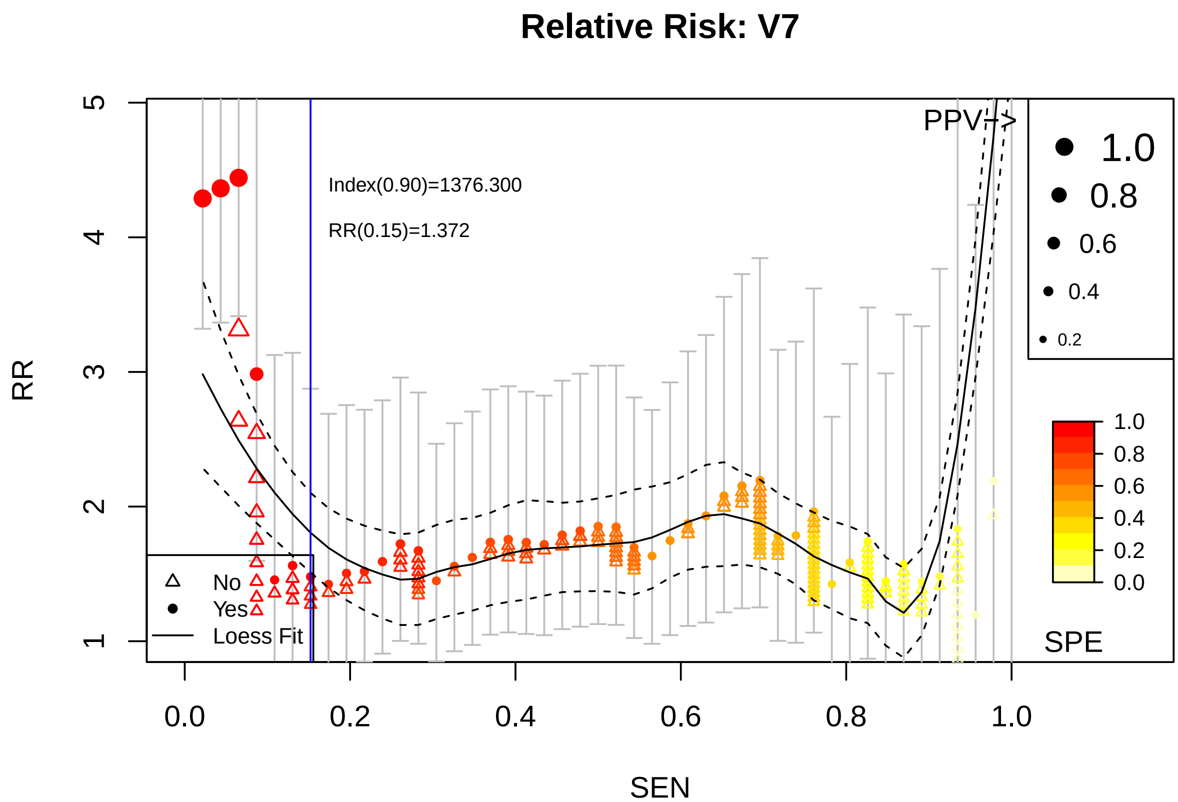


ROC: V34
(TPR= 0.152 ,TNR= 0.899 ,ACC= 0.722 ,F1= 0.206 ,BER= 0.475)
Sensitivity
1.0
0.8
0.6
0.4
0.2
0.0
Specificity
1.0
0.8
0.6
0.4
0.2
0.0
Confusion Matrix
O(+)
O(−)
T(+)
T(−)
Test
Outcome
AUC: 0.653
BACC (0.525)

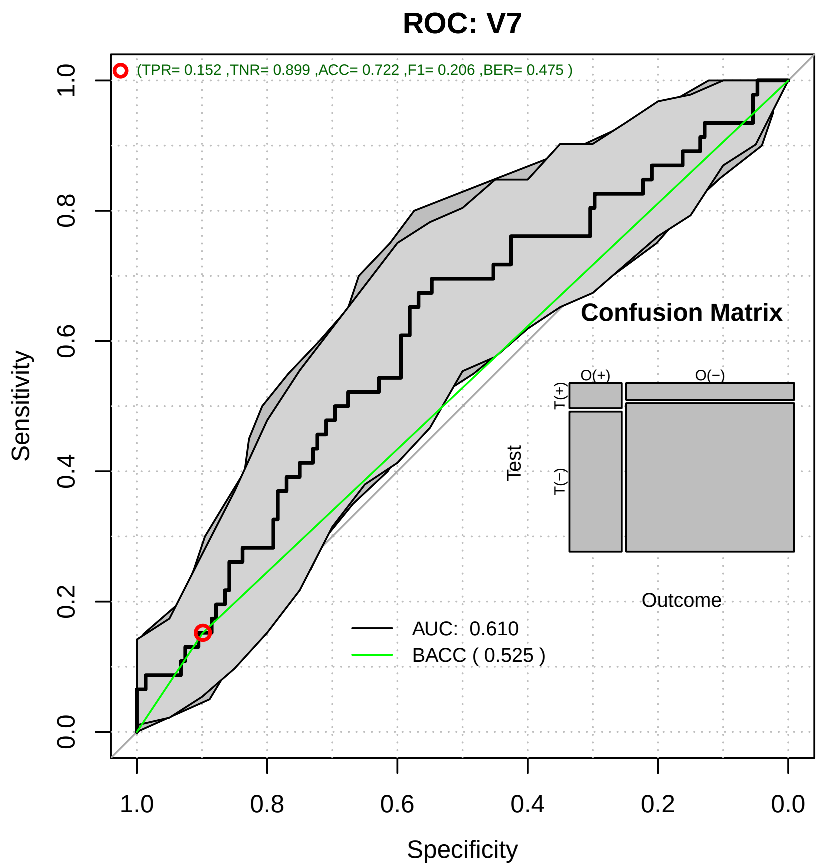


Number at risk

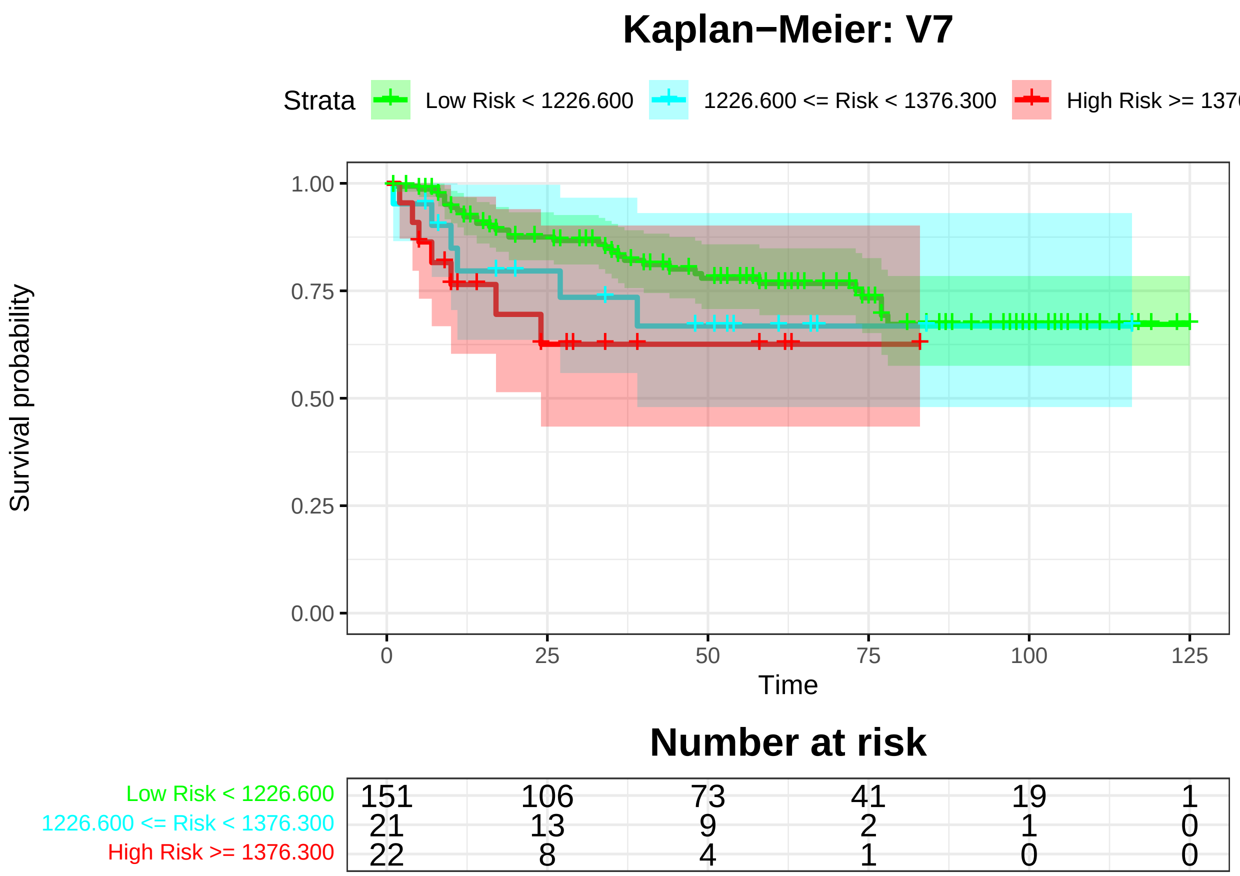
	0	25	50	75	100	125
Low Risk < 3.500	149	103	69	35	17	1
3.500 <= Risk < 5.000	23	10	8	4	2	0
High Risk >= 5.000	22	14	9	5	1	0

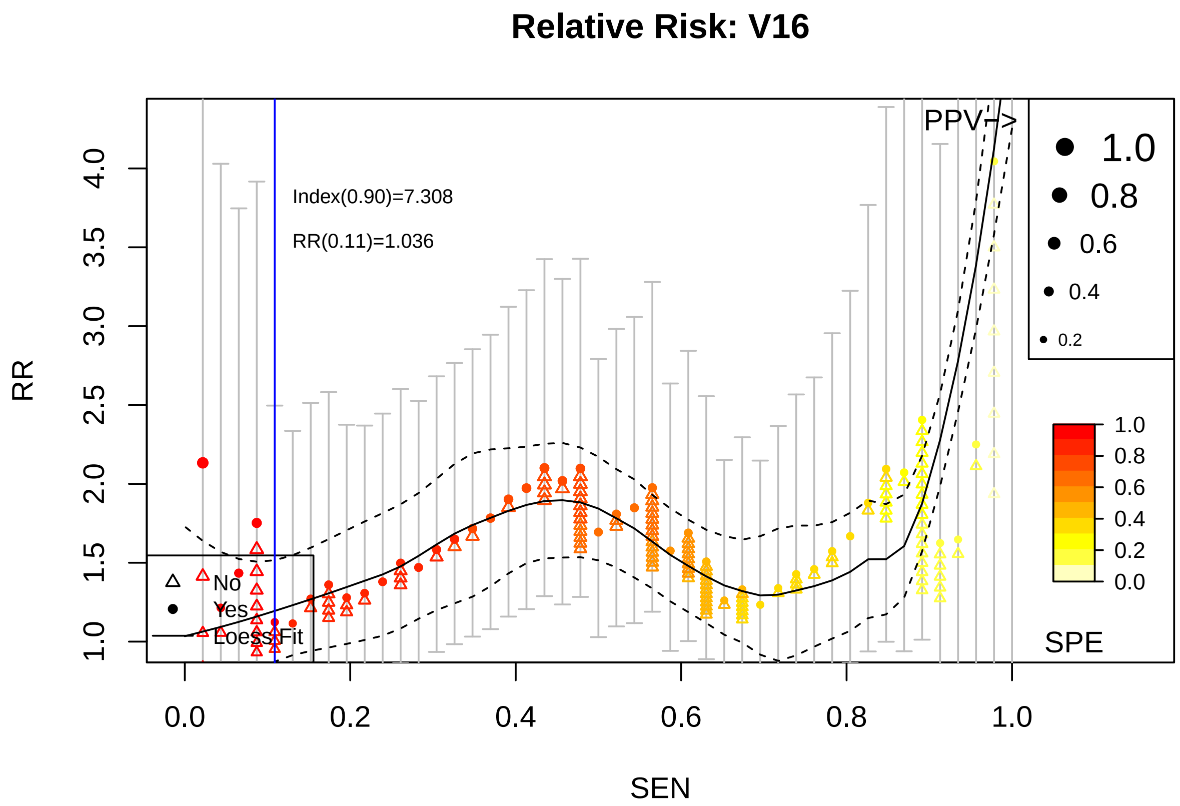


Relative Risk: V7
Index(0.90)=1376.300
RR(0.15)=1.372
PPV-->
1.0
0.8
0.6
0.4
0.2
1.0
0.8
0.6
0.4
0.2
0.0
SPE
No
Yes
Loess Fit
RR
1
2
3
4
5
0.0
0.2
0.4
0.6
0.8
1.0
SEN

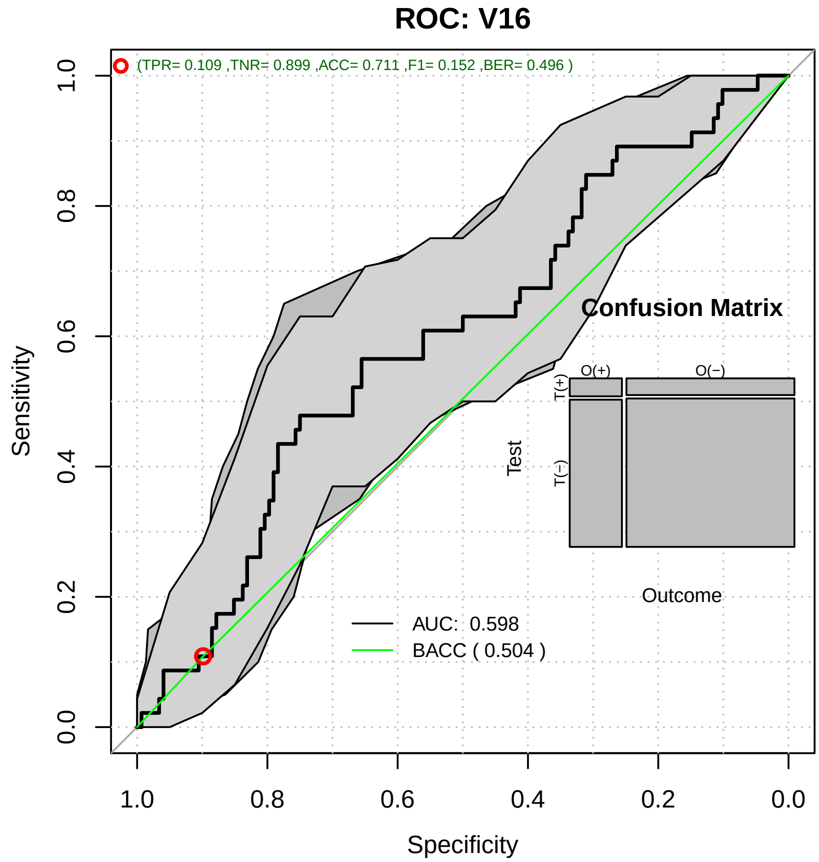


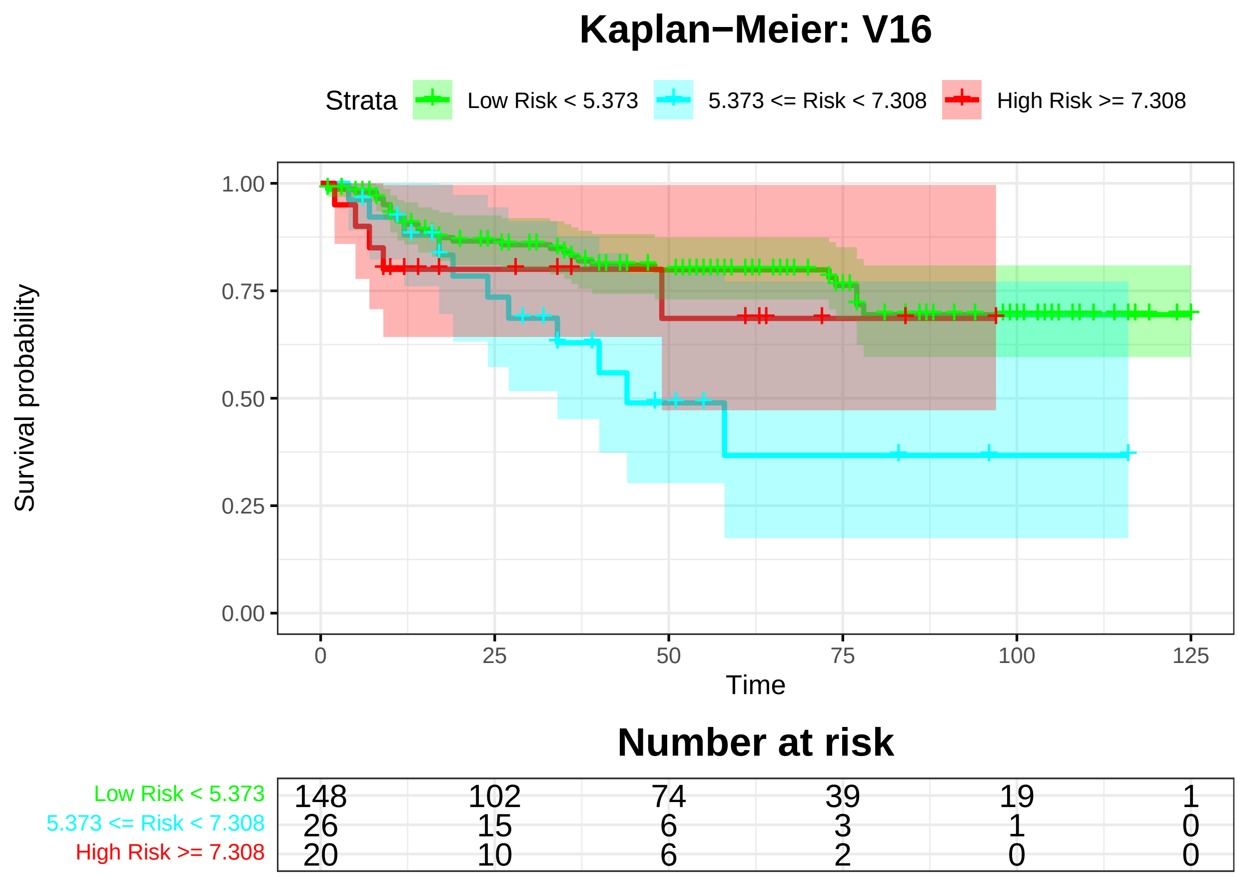
ROC: V7
(TPR= 0.152 ,TNR= 0.899 ,ACC= 0.722 ,F1= 0.206 ,BER= 0.475)
Sensitivity
Specificity
1.0
0.8
0.6
0.4
0.2
0.0
Confusion Matrix
O(+)
O(−)
T(+)
T(−)
Test
Outcome
AUC: 0.610
BACC (0.525)


Kaplan−Meier: V7
Strata
Low Risk < 1226.600
1226.600 <= Risk < 1376.300
High Risk >= 1376
Survival probability
1.00
0.75
0.50
0.25
0.00
0
25
50
75
100
125
Time
Number at risk
Low Risk < 1226.600
1226.600 <= Risk < 1376.300
High Risk >= 1376.300
151 106 73 41 19 1
21 13 9 2 1 0
22 8 4 1 0 0



Relative Risk: V16
Index(0.90)=7.308
RR(0.11)=1.036
PPV-->
1.0
0.8
0.6
0.4
0.2
SPE
No
Yes
Loess Fit
RR
SEN


ROC: V16
(TPR= 0.109 ,TNR= 0.899 ,ACC= 0.711 ,F1= 0.152 ,BER= 0.496)
Sensitivity
Specificity
1.0
0.8
0.6
0.4
0.2
0.0
Confusion Matrix
O(+)
O(−)
T(+)
T(−)
Test
Outcome
AUC: 0.598
BACC (0.504)




```
names(RRanalysis) <- topFive
```

1.2 Reporting the Metrics

```
pander::pander(RRanalysis[[1]]$keyPoints,caption=topFive[1])
```

Table 2: V35

	Thr	RR	SEN	SPE	BACC
@:0.9	1.0e+01	1.57	0.174	0.89865	0.536
@:0.8	3.0e+00	2.50	0.478	0.79730	0.638
@MAX_BACC	2.0e+00	2.58	0.565	0.73649	0.651
@MAX_RR	2.0e+00	2.60	0.543	0.75676	0.650
@SPE100	-9.6e-09	1.00	1.000	0.00676	0.503

```
pander::pander(RRanalysis[[2]]$keyPoints,caption=topFive[2])
```

Table 3: V24

	Thr	RR	SEN	SPE	BACC
@:0.9	25.4	1.94	0.239	0.8919	0.566
@:0.8	24.0	1.72	0.348	0.7973	0.573
@MAX_BACC	20.3	2.45	0.739	0.5270	0.633

	Thr	RR	SEN	SPE	BACC
@MAX_RR	16.6	3.87	0.957	0.1824	0.569
@SPE100	15.5	30.33	1.000	0.0811	0.541

```
RRanalysis[[2]]$keyPoints["@MAX_BACC",c("BACC","RR")]
```

```
           BACC        RR
```

@MAX_BACC 0.6330787 2.451923

```
ROCAUC <- NULL
CstatCI <- NULL
RRatios <- NULL
LogRangp <- NULL
Sensitivity <- NULL
Specificity <- NULL
MAXBACC <- NULL

for (topf in topFive)
{
  CstatCI <- rbind(CstatCI,RRanalysis[[topf]]$c.index$cstatCI)
  RRatios <- rbind(RRatios,RRanalysis[[topf]]$RR_atP)
  LogRangp <- rbind(LogRangp,RRanalysis[[topf]]$surdif$pvalue)
  Sensitivity <- rbind(Sensitivity,RRanalysis[[topf]]$ROCAnalysis$sensitivity)
  Specificity <- rbind(Specificity,RRanalysis[[topf]]$ROCAnalysis$specificity)
  ROCAUC <- rbind(ROCAUC,RRanalysis[[topf]]$ROCAnalysis$aucs)
  MAXBACC <- rbind(MAXBACC,RRanalysis[[topf]]$keyPoints["@MAX_BACC",c("BACC")])
}
rownames(CstatCI) <- topFive
rownames(RRatios) <- topFive
rownames(LogRangp) <- topFive
rownames(Sensitivity) <- topFive
rownames(Specificity) <- topFive
rownames(ROCAUC) <- topFive
rownames(MAXBACC) <- topFive

pander::pander(ROCAUC)
```

	est	lower	upper
V35	0.647	0.552	0.742
V24	0.633	0.542	0.724
V34	0.653	0.565	0.740
V7	0.610	0.515	0.705
V16	0.598	0.504	0.692

```
pander::pander(CstatCI)
```

	mean.C Index	median	lower	upper
V35	0.633	0.634	0.549	0.713
V24	0.677	0.678	0.593	0.757
V34	0.649	0.650	0.581	0.719

	mean.C Index	median	lower	upper
V7	0.666	0.668	0.588	0.747
V16	0.614	0.615	0.530	0.705

```
pander::pander(RRatios)
```

	est	lower	upper
V35	1.44	0.739	2.81
V24	1.93	1.122	3.31
V34	1.44	0.741	2.82
V7	1.37	0.700	2.69
V16	1.04	0.462	2.32

```
pander::pander(LogRangp)
```

V35	9.35e-05
V24	9.38e-03
V34	1.01e-01
V7	7.33e-02
V16	2.13e-02

```
pander::pander(Sensitivity)
```

	est	lower	upper
V35	0.152	0.0634	0.289
V24	0.239	0.1259	0.388
V34	0.152	0.0634	0.289
V7	0.152	0.0634	0.289
V16	0.109	0.0362	0.236

```
pander::pander(Specificity)
```

	est	lower	upper
V35	0.899	0.838	0.942
V24	0.899	0.838	0.942
V34	0.899	0.838	0.942
V7	0.899	0.838	0.942
V16	0.899	0.838	0.942

```
pander::pander(MAXBACC)
```

V35	0.651
V24	0.633
V34	0.646
V7	0.621

V16	0.614

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

	ROCAUC	C-Stat	Sen	Spe	RR	MAX_BACC
V35	0.647	0.633	0.152	0.899	1.44	0.651
V24	0.633	0.677	0.239	0.899	1.93	0.633
V34	0.653	0.649	0.152	0.899	1.44	0.646
V7	0.610	0.666	0.152	0.899	1.37	0.621
V16	0.598	0.614	0.109	0.899	1.04	0.614

1.3 Modeling

```
ml <- BSWiMS.model(Surv(time,status)~1,data=dataBreast,NumberofRepeats = 10)
```

[++]…..

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

Table 12: Table continues below

	Estimate	lower	HR	upper	u.Accuracy	r.Accuracy
V24	5.85e-02	1.02	1.06	1.10	0.598	0.237
V26	4.47e-03	1.00	1.00	1.01	0.593	0.336
V27	1.85e-04	1.00	1.00	1.00	0.608	0.324
V34	1.22e-02	1.00	1.01	1.02	0.634	0.297
V7	6.03e-08	1.00	1.00	1.00	0.588	0.237
V35	4.47e-03	1.00	1.00	1.01	0.727	0.599
V6	1.07e-07	1.00	1.00	1.00	0.577	0.237

Table 13: Table continues below

	full.Accuracy	u.AUC	r.AUC	full.AUC	IDI	NRI	z.IDI
V24	0.598	0.609	0.500	0.609	0.0619	0.437	2.87
V26	0.598	0.598	0.529	0.602	0.0622	0.398	2.75
V27	0.609	0.608	0.525	0.606	0.0561	0.434	2.75
V34	0.630	0.618	0.517	0.616	0.0309	0.464	2.38
V7	0.588	0.595	0.500	0.595	0.0487	0.380	2.30
V35	0.616	0.641	0.603	0.607	0.0281	0.551	2.25
V6	0.577	0.588	0.500	0.588	0.0459	0.353	2.19

	z.NRI	Delta.AUC	Frequency
V24	2.67	0.10914	1.0
V26	2.41	0.07310	1.0

	z.NRI	Delta.AUC	Frequency
V27	2.64	0.08104	1.0
V34	2.81	0.09871	1.0
V7	2.30	0.09489	0.8
V35	3.41	0.00427	1.0
V6	2.13	0.08813	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()

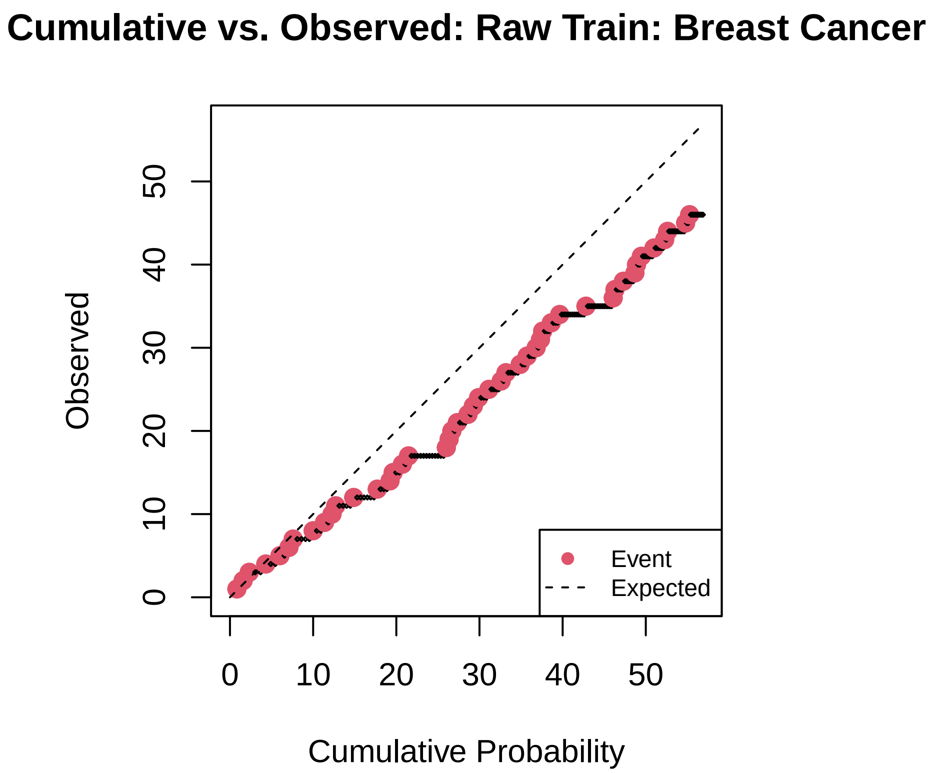
Here we will use the predicted event probability assuming a baseline hazard for events withing 5 years

```
index <- predict(ml,dataBreast)
timeinterval <- 2*mean(subset(dataBreast,status==1)$time)

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

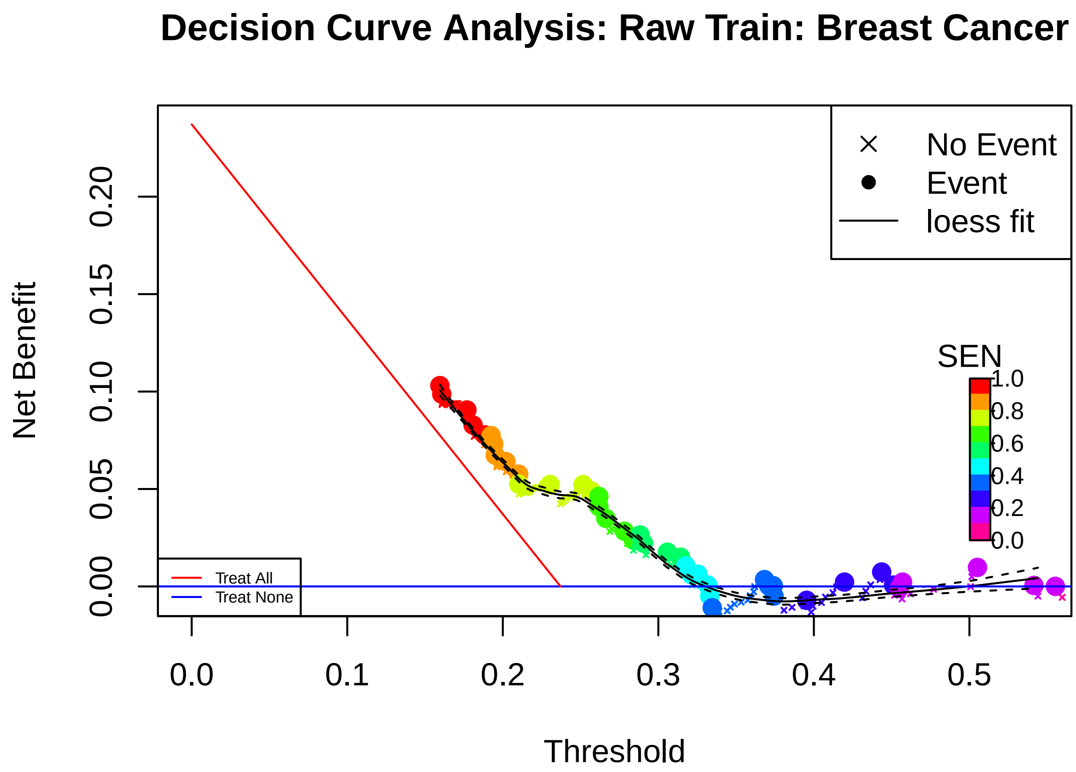
Table 15: Initial Parameters

h0	timeinterval
0.323	51.1

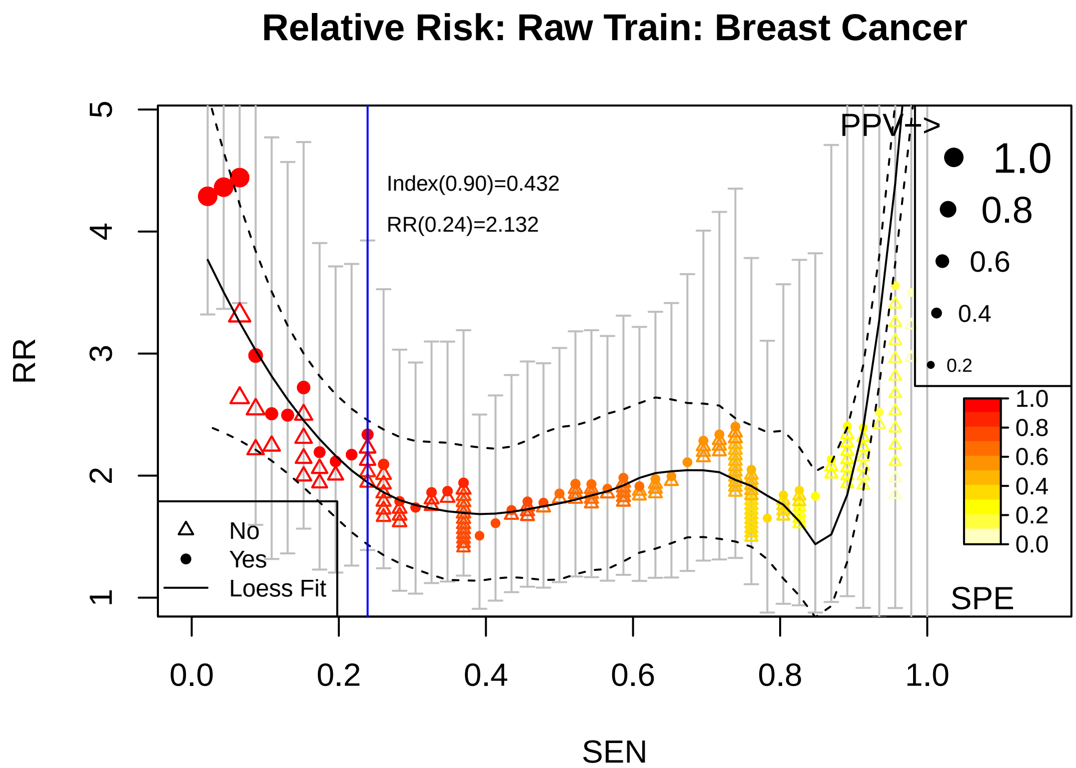
```
rdata <- cbind(dataBreast$status,ppoisGzero(index,h0))
rownames(rdata) <- rownames(dataBreast)

rrAnalysisTrain <- RRPlot(rdata,atProb=c(0.90,0.80),
                     timetoEvent=dataBreast$time,
                     title="Raw Train: Breast Cancer",
                     ysurvlim=c(0.00,1.0),
                     riskTimeInterval=timeinterval)
```

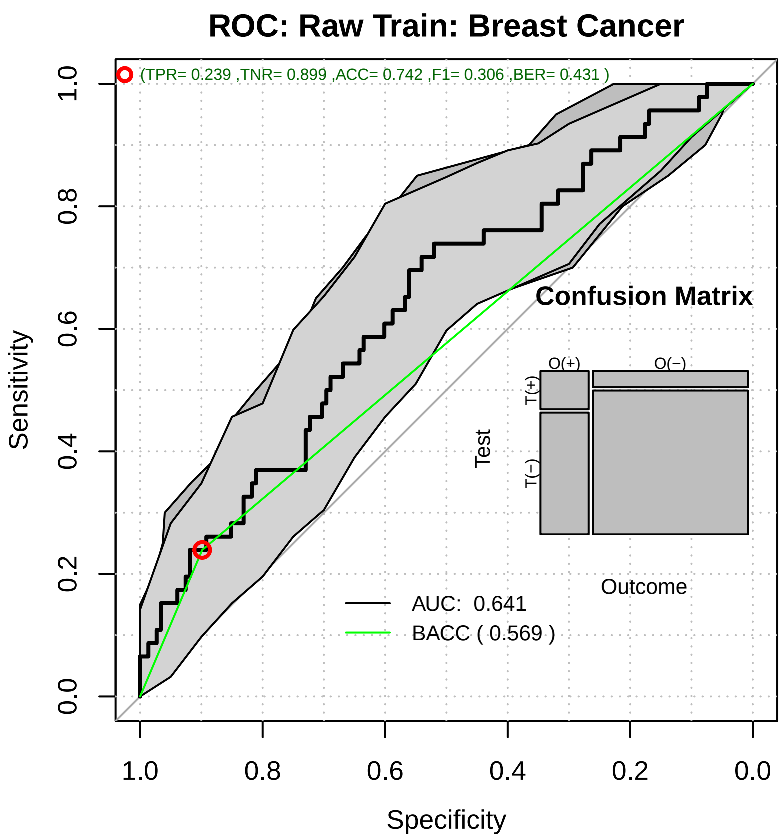


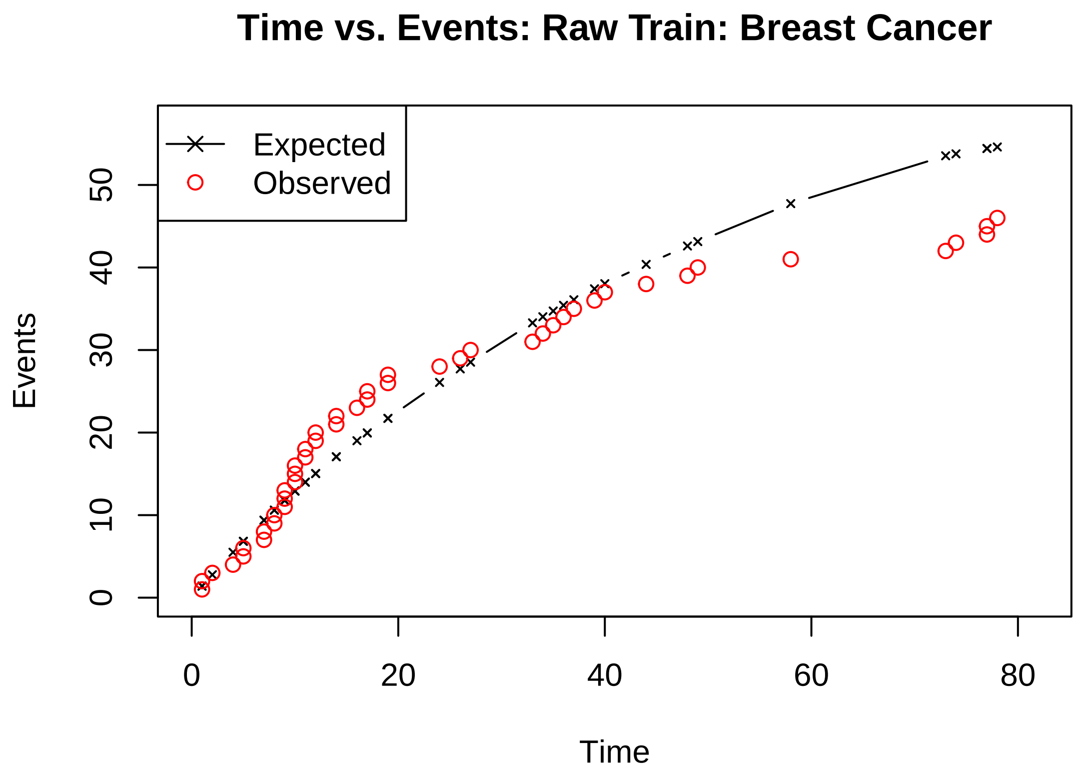
Cumulative vs. Observed: Raw Train: Breast Cancer
Observed
0
10
20
30
40
50
Event
Expected
0
10
20
30
40
50
Cumulative Probability



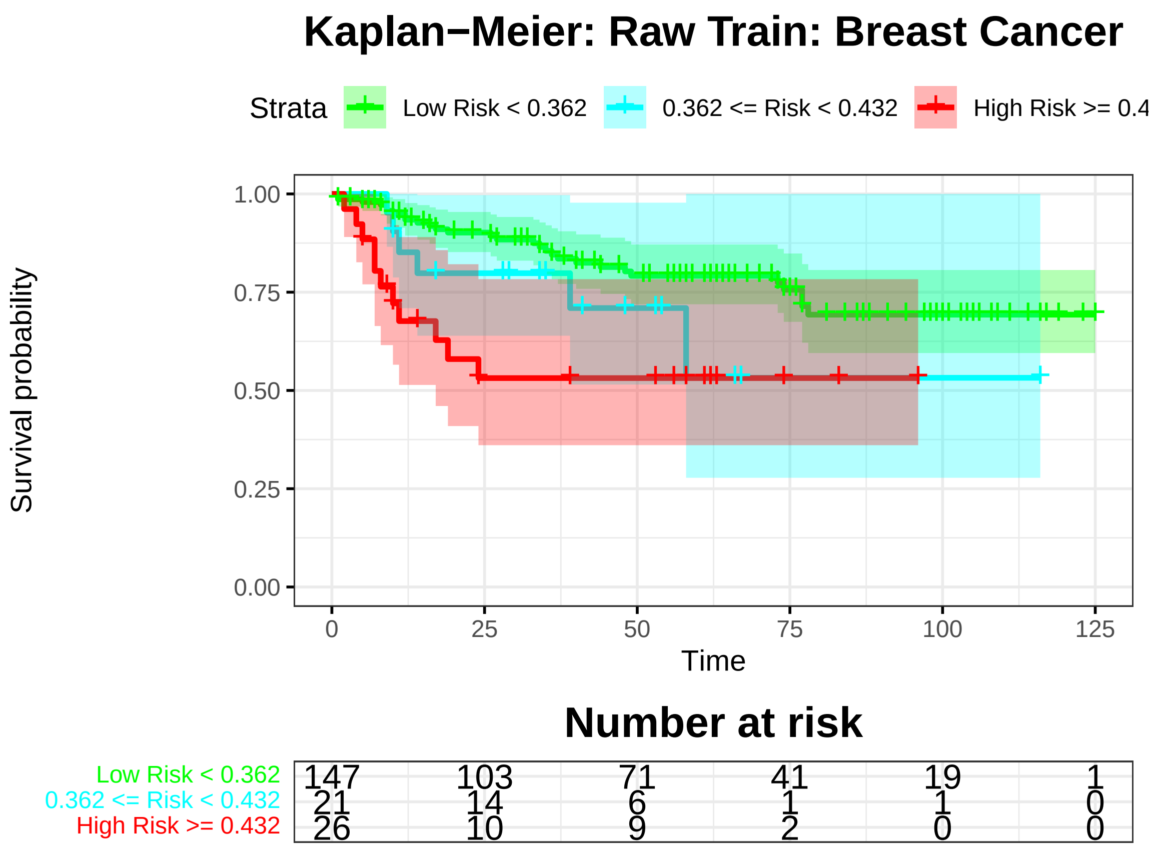
Decision Curve Analysis: Raw Train: Breast Cancer
Net Benefit
0.00
0.05
0.10
0.15
0.20
0.0
0.1
0.2
0.3
0.4
0.5
Threshold
No Event
Event
loess fit
SEN
1.0
0.8
0.6
0.4
0.2
0.0
Treat All
Treat None


Relative Risk: Raw Train: Breast Cancer
Index(0.90)=0.432
RR(0.24)=2.132
PPV-->
1.0
0.8
0.6
0.4
0.2
1.0
0.8
0.6
0.4
0.2
0.0
SPE
No
Yes
Loess Fit
RR
1
2
3
4
5
0.0
0.2
0.4
0.6
0.8
1.0
SEN


ROC: Raw Train: Breast Cancer
(TPR= 0.239 ,TNR= 0.899 ,ACC= 0.742 ,F1= 0.306 ,BER= 0.431)
Sensitivity
Specificity
1.0
0.8
0.6
0.4
0.2
0.0
Confusion Matrix
O(+)
O(−)
T(+)
T(−)
Test
Outcome
AUC: 0.641
BACC (0.569)



Time vs. Events: Raw Train: Breast Cancer
Expected
Observed
Events
0
10
20
30
40
50
0
20
40
60
80
Time



1.4.2 Uncalibrated Performance Report

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

Table 16: Threshold values

	@:0.9	@:0.8	@MAX_BACC	@MAX_RR	@SPE100	p(0.5)
Thr	0.431	0.362	0.252	0.177	0.1596	0.501
RR	1.944	1.833	2.402	3.557	27.6503	2.308
SEN	0.239	0.370	0.739	0.957	1.0000	0.152
SPE	0.892	0.797	0.520	0.169	0.0743	0.953
BACC	0.566	0.583	0.630	0.563	0.5372	0.552

```
pander::pander(t(rrAnalysisTrain$OERatio$estimate),caption="O/E Test")
```

Table 17: O/E Test

O/E	Low	Upper	p.value
0.842	0.617	1.12	0.278

```
pander::pander(t(rrAnalysisTrain$OE95ci),caption="O/E Mean")
```

Table 18: O/E Mean

mean	50%	2.5%	97.5%
1.02	1.02	0.969	1.07

```
pander::pander(t(rrAnalysisTrain$OARatio$estimate),caption="O/Acum Test")
```

Table 19: O/Acum Test

O/A	Low	Upper	p.value
0.809	0.592	1.08	0.163

```
pander::pander(t(rrAnalysisTrain$OAcum95ci),caption="O/Acum Mean")
```

Table 20: O/Acum Mean

mean	50%	2.5%	97.5%
0.793	0.793	0.787	0.799

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

Table 21: C. Index

mean.C Index	median	lower	upper
0.683	0.683	0.6	0.76

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

Table 22: ROC AUC

est	lower	upper
0.641	0.55	0.733

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

Table 23: Sensitivity

est	lower	upper
0.239	0.126	0.388

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

Table 24: Specificity

est	lower	upper
0.899	0.838	0.942

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

Table 25: Probability Thresholds

90%	80%
0.432	0.362

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

Table 26: Risk Ratio

est	lower	upper
2.13	1.25	3.62

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

Table 27: Logrank test Chisq = 11.354656 on 2 degrees of freedom, p = 0.003423

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
class=0	147	29	36.99	1.725	8.974
class=1	21	6	4.41	0.573	0.643
class=2	26	11	4.60	8.896	9.982