ROC and Reclassification analysis in R

R Ísland meeting

Thor Aspelund, Icelandic Heart Association, University of Iceland
Public Health

October 30th 2014

CHD risk models

- Icelandic Heart Association Risk Score
  - `http://www.hjarta.is`
- European HeartSCORE
  - `http://www.heartscore.org/Pages/welcome.aspx`
- Framingham Risk Score (Circulation 1998)
  - `http://circ.ahajournals.org/content/97/18/1837.long`

The ongoing search for risk markers

- There is an existing risk score in use
- We would like to introcude a new marker to improve the score
- Is the new score better than the old score? **That is the question**

Introducing a new marker

Measuring improvement using statistcal models

- Model 1: Basic risk score model
- Model 2: Basic risk score model + new marker
- Is risk score 2 better than risk score 1?
- In other words: Is Model 2 an improvment of than Model 1?

How is a risk score evaluated?

- In categorical data analysis we have met various concordance measures, such as:
  - Kendall's tau_a
  - Somer's D
  - The C index

Concordance measures - binary outcome

View Y as a binary outcome, either $Y = 1$ or $Y = 0$. Let Z be a continous risk score. The concordance between the risk score and the outcome can be measured by Kendall's τ_a

$$\tau_a(Y, Z) = E(\text{sign}(Z_1 - Z_2)\text{sign}(Y_1 - Y_2)))$$

where the pairs (Y_1, Z_1) and (Y_2, Z_2) are chosen at random.

Somer's D is an adaption of τ_a

$$D(Z, Y) = \tau_a(Y, Z)/\tau_a(Y, Y)$$

Concordance and the C statistic

If there are no ties in Z it can be shown that

$$D(Z, Y) = 2 \cdot P(Z_i > Z_j | Y_i > Y_j) - 1 = 2 \cdot C(Z, Y) - 1$$

Where

$$C(Z, Y) = P(Z_i > Z_j | Y_i > Y_j)$$

- The C statistic is the probability that a risk score for a case $Y_i = 1$ is greater than the risk score for a control $Y_j = 0$.
- We want this probability to be high.
- It can be shown (via integration) that $0.5 <= C(Z, Y) <= 1$.

ROC analysis, AUC and the C statistic

- ROC analysis is an analysis of sensitivity and specificity
- Let Z be a continuous risk score and z be an arbitrary cutoff value. Choose a random case or a control (Denote by Y). Compute the risk score Z for each.
- Assume you are blinded to the case control status. Declare the subject to be a case if Z>z, otherwise a control.
- Then $P(Z > z|Y = 1)$ is the sensitivity or the True Positive Probability (TP).
- $P(Z > z|Y = 0) = 1 - P(Z < z|Y = 0)$ is 1-specificity or the False Positive Probability (FP).
- A graph of $TP = P(Z > z|Y = 1)$ vs. $FP = P(Z > z|Y = 0)$ is the ROC curve for the diagnostic test Z.
- It can be shown that the area under the curve (AUC) equals the C statistic. AUC=C!
- Common values for C for risk models are in the range 0.70 to 0.75.

Frank Harrell and the rms package

- Frank Harell provides many diagnostic functions in the *rms* package. C and D are displayed as elements in *lrm* (logistic regression models) objects. However, C with D and a standard error for D are provided with *rcorr.cens*. Note that the standard error of C is half the standard error of D.
- Recall: $C = 0.5 \cdot (D + 1)$

The example data

- WCGS data - Western Collaborative Group Study
- Prospective study of heart disease among men in California, initiated in 1960
- `http://clinicaltrials.gov/show/NCT00005174`
- N $=$ 3154, age 39 to 59, free of heart disease
- Follow-up for 10 years
- Data (*wcgs*) available via the *epitools* package
- N $=$ 3141 with complete data on risk factors used for analysis

C and D results from the rms package

- Consider Model 0: chd $\sim$ age

```
fit0 <- lrm(chd69 ~ age0,data=wcgs.s,x=T,y=T)
data.frame(C=fit0$stats[6],D=fit0$stats[7])
```

```
##          C      D
## C 0.6212 0.2424
```

```
rc0<-rcorr.cens(predict(fit0),fit0$y)
data.frame(Estimate=rc0[1],SE=rc0[3]/2,
           Lower=rc0[1]-1.96*rc0[3]/2,Upper=rc0[1]+1.96*rc0
```

```
##           Estimate      SE  Lower  Upper
## C Index     0.6212 0.01873 0.5845 0.6579
```

We have $C = 0.6212$ for Model 0. There is room for improvement.

Are under the ROC curve for Model 0 and threshold

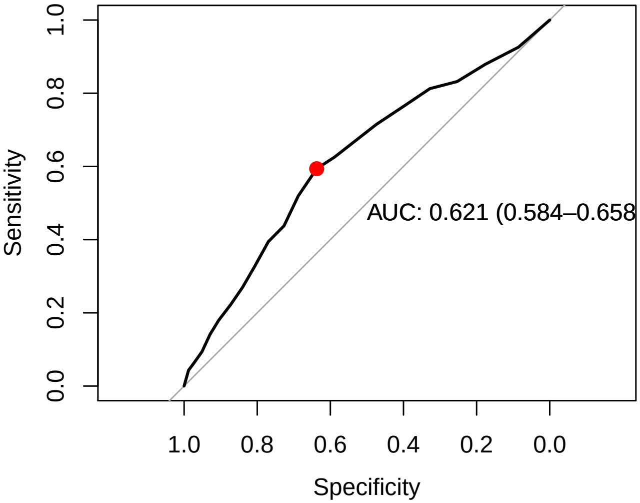
```r
roc0<-roc(fit0$y,predict(fit0),ci=T)
roc0.c <- coords(roc0,x="best",best.method=c("closest.tople
roc0
```

```
##
## Call:
## roc.default(response = fit0$y, predictor = predict(fit0)
##
## Data: predict(fit0) in 2885 controls (fit0$y 0) < 256 ca
## Area under the curve: 0.621
## 95% CI: 0.584-0.658 (DeLong)
```

```r
roc0.c
```

```
##   threshold specificity sensitivity
##     -2.4045      0.6374      0.5938
```

ROC curve for Model 0 and threshold



The threshold on the risk scale

```r
plogis(roc0.c[1])
```

```
## threshold
##   0.08283
```

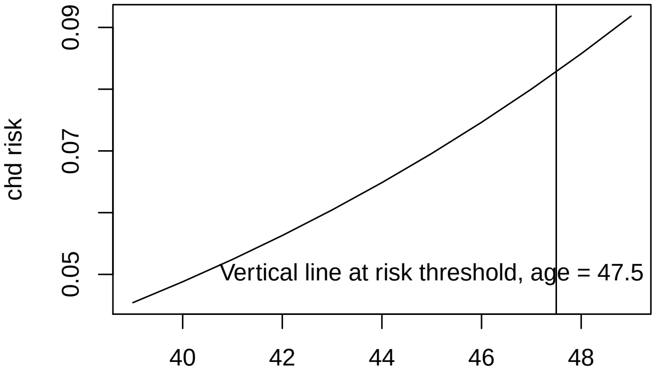
The age where the threshold is reached

```r
data.frame(age=(roc0.c[1]-coef(fit0)[1])/coef(fit0)[2])
```

```
##                age
## threshold 47.5
```

Risk as a function of age

```
plot(39:49,predict(fit0,newdata=data.frame(age0=39:49),type
abline(v=(roc0.c[1]-coef(fit0)[1])/coef(fit0)[2])
text(45,0.05,paste("Vertical line at risk threshold, age ='
```



Model 1 - add cholesterol, blood pressure, bmi, and smoking

```
fit1<-update(fit0,.~.+cholmmol + sbp0 + bmi + smoker)
data.frame(C=fit1$stats[6],D=fit1$stats[7])
```

```
##          C      D
## C 0.7323 0.4646
```

```
rc1<-rcorr.cens(predict(fit1),fit1$y)
data.frame(Estimate=rc1[1],SE=rc1[3]/2,
           Lower=rc1[1]-1.96*rc1[3]/2,Upper=rc1[1]+1.96*rc1
```

```
##         Estimate      SE  Lower  Upper
## C Index   0.7322 0.01562 0.7016 0.7628
```

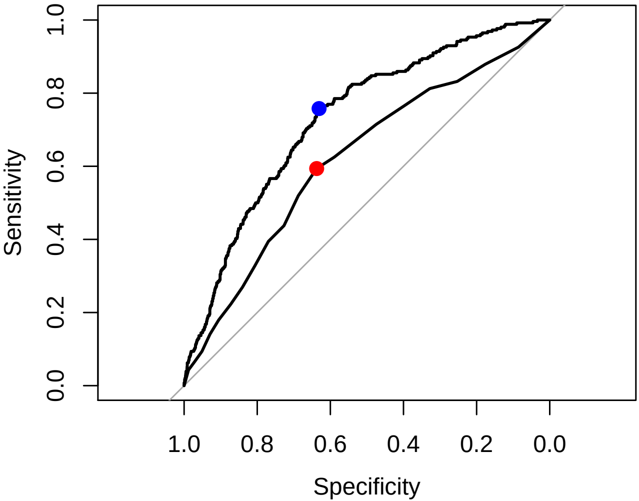
Model 1 is an improvment

```
##
## Call:
## roc.default(response = fit1$y, predictor = predict(fit1)
##
## Data: predict(fit1) in 2885 controls (fit1$y 0) < 256 ca
## Area under the curve: 0.732
## 95% CI: 0.702-0.763 (DeLong)
```

```
##   threshold specificity sensitivity
##     -2.5167      0.6308      0.7578
```

```
## threshold
##    0.0747
```

ROC curve for Model 1 and Model 0 and threshold



Test of improvement

```r
roc.test(roc0,roc1)
```

```
##
##  DeLong's test for two correlated ROC curves
##
## data:  roc0 and roc1
## Z = -6.517, p-value = 7.156e-11
## alternative hypothesis: true difference in AUC is not e
## sample estimates:
## AUC of roc1 AUC of roc2
##      0.6212      0.7322
```

The hunt for a new marker

- We have our basic risk model (Model 1)
- The C statistic is 0.7322
- This is a typical value for a chd risk model
- We would still like to improve it
- The hunt is on for a new marker
- We add the new marker to Model 1 and measure the improvement

Our new marker - Personality A vs B

Personality is associated with CHD. The OR > 2.

```
oddsratio.wald(wcgs$dibpat0f,wcgs$chd69)
```

```
## $data
##           Outcome
## Predictor    0   1 Total
##      B      1486  79  1565
##      A      1411 178  1589
##      Total 2897 257  3154
##
## $measure
##           odds ratio with 95% C.I.
## Predictor estimate lower upper
##         B    1.000    NA    NA
##         A    2.373 1.803 3.123
##
## $p.value
```

Models

- Model 0: 0 chd ∼ age
- Model 1: chd ∼ age + bmi + chol + systolic + smoker
- Model 2: chd ∼ age + bmi + chol + systolic + smoker + personality

–> personality with 2 levels (A,B) is our *new* marker

Model 2 - Add the personality marker

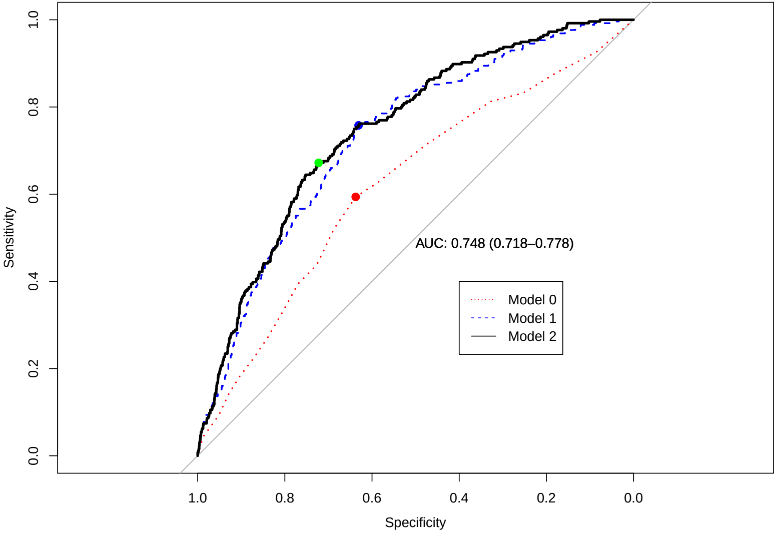
```r
fit2 <- update(fit1,.~.+dibpat0f)
```

The adjusted OR is

```r
data.frame(OR=exp(coef(fit2)[7]),Lower=exp(confint.default
```

```
##                OR Lower Upper
## dibpat0f=A 2.007 1.512 2.663
```

ROC curves 0 1 and 2



Likelihood ratio test comparing Model 1 & 2

```r
lrtest(fit1,fit2)
```

```
##
## Model 1: chd69 ~ age0 + cholmmol + sbp0 + bmi + smoker
## Model 2: chd69 ~ age0 + cholmmol + sbp0 + bmi + smoker
##
## L.R. Chisq       d.f.          P
##  2.453e+01  1.000e+00  7.305e-07
```

- This means that the marker is highly significant.
- Recall that the OR ≈ 2

The Pepe 2004 paper

- Limitations of the Odds Ratio in Gauging the Performance of a Diagnostic, Prognostic, or Screening Marker
- Margaret Sullivan Pepe, Holly James, Gary Longton, Wendy Leisenring, and Polly Newcomb
- American Journal of Epidemiology 2004

Message

- Tells us about the limitations of the OR as a measure of diagnostic capacity and that ROC curves and sensitivity and specificity must be studied.
- Also demonstrates how difficult it is to see a change in ROC curves between models using 1 new marker

Formally comparing ROC curves 1 and 2

```r
roc.test(roc1,roc2)
```

```
##
##  DeLong's test for two correlated ROC curves
##
## data:  roc1 and roc2
## Z = -2.364, p-value = 0.01806
## alternative hypothesis: true difference in AUC is not e
## sample estimates:
## AUC of roc1 AUC of roc2
##      0.7322      0.7481
```

- Statisticsally signficant
- The increment is less than 0.02!

The NEJM paper from the Icelandic Heart Association

- C-reactive protein and other circulating markers of inflammation in the prediction of coronary heart disease.
- Danesh J1, Wheeler JG, Hirschfield GM, Eda S, Eiriksdottir G, Rumley A, Lowe GD, Pepys MB, Gudnason V.
- N Engl J Med. 2004 Apr 1;350(14):1387-97.

Message

- CRP is a statistically significant marker
- Adding CRP to a risk model using traditional risk factors increase the ROC area by 0.01
- *C-reactive protein is a relatively moderate predictor of coronary heart disease. Recommendations regarding its use in predicting the likelihood of coronary heart disease may need to be reviewed*

Frustration among CRP advocates

What are we going to do about these small increments in AUC?

Nancy R Cook paper 2006 - Reclassification

- The effect of including C-reactive protein in cardiovascular risk prediction models for women.
- Cook NR1, Buring JE, Ridker PM.
- Ann Intern Med. 2006 Jul 4;145(1):21-9.

Message

- Introduced the concept of **reclassification**
- Do subjects move between risk categories after adding the predictor?
- A global risk prediction model that includes hsCRP improves cardiovascular risk classification in women, particularly among those with a 10-year risk of 5% to 20%. In models that include age, blood pressure, and smoking status, hsCRP improves prediction at least as much as do lipid measures.

Nancy R Cook paper 2006 - Reclassification

- Did I mention the conflict of interest?

Potential conflict of interest reported in Cook's paper

Dr. Ridker is listed as a co-inventor on patents held by the Brigham and Women's Hospital that relate to the use of inflammatory biomarkers in cardiovascular disease.

Example of Cook's approach

```
tblc
```

```
##           pred2c
## pred1c     (0,10] (10,20] (20,100]
##   (0,10]     2099     190        0
##   (10,20]     189     384       80
##   (20,100]      0      61      138
```

- This shows that many participants are reclassified.
- For example: 190 are reclassfied from 0 to 10% risk into 10-20% risk

Peninca 2008

- Have to consider reclassification of people who develop and who do not develop the events **separately**
- Defines the net reclassification improvement NRI based on risk categories

	(0,10]	(10,20]	(20,100]	(0,10]	(10,20]	(20,100]
StatusCHD		CHD=0			CHD=1	
(0,10]	2018	160	0	81	30	0
(10,20]	175	313	60	14	71	20
(20,100]	0	54	105	0	7	33

- Inroduces statistical inference about reclassification (NRI) and Integrated discrimination improvement (IDI)

Test of net reclassification NRI

Asymptotic test of

$$NRI = (\hat{p}_{up,events} - \hat{p}_{down,events}) - (\hat{p}_{up,nonevents} - \hat{p}_{down,nonevents})$$

- Notice the retrospective definition
- Doesn't really apply to **case control** data unless we can adjust the risk estimates to be meaningful

NRI estimate

- Using *reclass* from
- `http://www.ucr.uu.se/en/index.php/epistat/program-code/306-nri-and-idi`

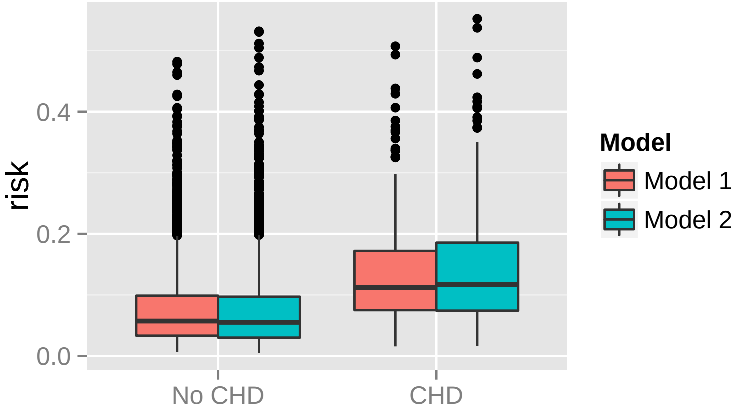
```
rcls<-reclass(chd69 ~ age0 + cholmmol + sbp0 + bmi + smoker
              + dibpat0f,lim=c(0.1,0.2),wcgs.s,1,TRUE)
```

the estimate was 0.1164 with 95% CI as (0.0503,0.1825). In fact -0.0031 in without event and 0.1133 in with outcome.

Test of IDI

Asymptotic test of the difference in difference between risk of non-cases and cases

$IDI = (\bar{\hat{p}}_{new,events} - \bar{\hat{p}}_{new,nonevents}) - (\bar{\hat{p}}_{down,events} - \bar{\hat{p}}_{down,events})$.
Estimate $= 0.009$ with SE $= 0.0023$.



Adding a new marker - Statistics to report

- JAMA 2009 Review paper:
- Assessment of Claims of Improved Prediction Beyond the Framingham Risk Score
- Ioanna Tzoulaki, PhD George Liberopoulos, MD John P. A. Ioannidis, MD

-> Set standard

- Akaike Information Criteria (AIC)
- AUCs with and without the new predictor
- Difference in AUC with a confidence interval
- Calibration with and without the new marker with a goodness of fit test
- Documentation of **reclassification**