Hands-on ROC analysis marker assessment illustrations and R practice

Assessment of diagnostic markers

Continuous-scaled marker

We will employ part of the CD4 marker dataset in order to illustrate the construction of the ROC curve for a continuous-scaled marker.

CD4 measurements for the 15 Controls were:

$$\{59, 66, 45, 62, 51, 50, 49, 58, 53, 42, 50, 47, 51, 62, 48\}$$

while respective measurements for the 12 Acute Brucellosis Cases were:

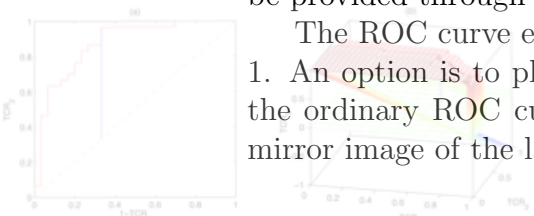
$$\{72, 70, 69, 82, 68, 59, 76, 61, 59, 73, 49, 77\}.$$

Suppose that we choose $c = 61$, thus considering measurements larger than or equal to 61 as diseased and measurements smaller than 61 as healthy. Then, the sensitivity is equal to 75% (i.e. 9/12), while the specificity is 80% (i.e. 12/15). Also, $LR^+ = 3.75$ and $LR^- = 0.31$.

Calculation for PPV and NPV follows next: Suppose that the prevalence of disease in the target population is 2%. Then, $PPV = 7.11\%$, while $NPV = 99.37\%$. Predictive values are calculated based on the assumption that the prior probability of a positive test result based on the reference standard is just 2%, equal to the prevalence of disease. Correct estimation of the prevalence in the target population is crucial for the correct estimation of PPV, NPV in turn. For example, if the prevalence is actually 5% then, $PPV = 16.48\%$, while $NPV = 98.38\%$. (PPV, NPV were calculated from the online calculator at `http://vassarstats.net`).

Estimates of prevalence are not possible in designed case-control studies (as in the Brucellosis example). In cohort studies, where a cohort from the population of interest is followed and population characteristics and disease status are recorded, the estimation is straightforward (as number of diseased over number of subjects in the study). Valid estimates of prevalence can also be provided through epidemiological studies in the target population.

The ROC curve estimated from the data at hand is illustrated in Figure 1. An option is to plot the graph of sensitivity versus specificity instead of the ordinary ROC curve. Both are depicted in Figure 1. The former is a mirror image of the latter. The AUC is equal to 0.886.



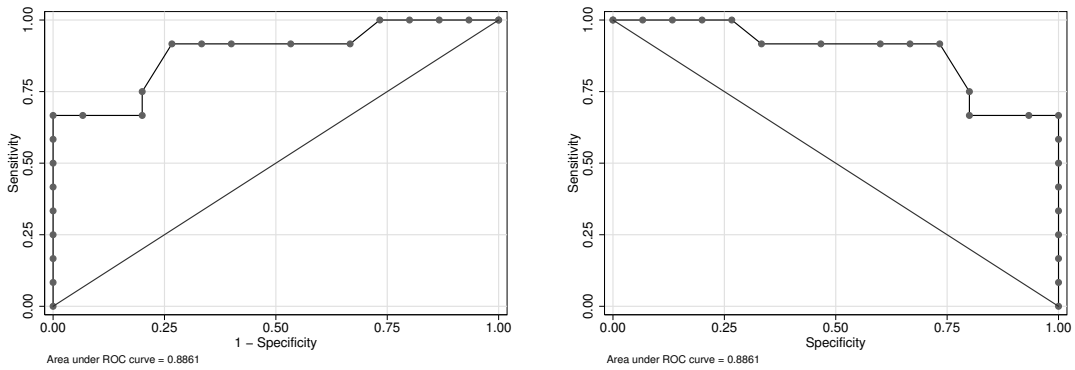


Figure 1: ROC curve for CD4 (left panel) and respective sens vs spec graph (right panel).

Ordinal-scaled marker

The ordinal-scaled measurements of the MoCA examination (Parkinson disease data) will be used for this illustration. Data are given in Table 1. With MoCA scores the higher the measurement, the healthier the brain of the subject under study. As a result $X_1 > X_2$ in general. To deal with this inconvenience we actually assign value '1' to 'Intact', '2' to 'Minimal impairment', '3' to 'Mild dementia', and '4' to 'Moderate dementia' and not the other way around. This strategy will result in an ROC curve above the main diagonal and in an AUC larger than 0.5.

Figure 2 illustrates the conventional and inverted ROC curves that can be produced based on the MoCA data. The left panel illustrates the conventional ROC where higher scores are representative of diseased subjects, while the right panel is the ROC curve that corresponds to scores assigned the other way around, i.e. '1' to 'Moderate dementia' etc. Notice that there were no subjects with severe dementia.

We continue with the conventional ROC curve. If we choose $c = 2$ as the cut-off point (i.e. patients with a score equal to or larger than 2 will be considered as diseased), the resulting sensitivity is $46/60 = 76.67\%$ $(65.96\%, 87.37\%)$, while the specificity is $71/80 = 88.75\%$ $(81.83\%, 95.67\%)$. 95% confidence intervals are given in parentheses. The AUC is 0.844.

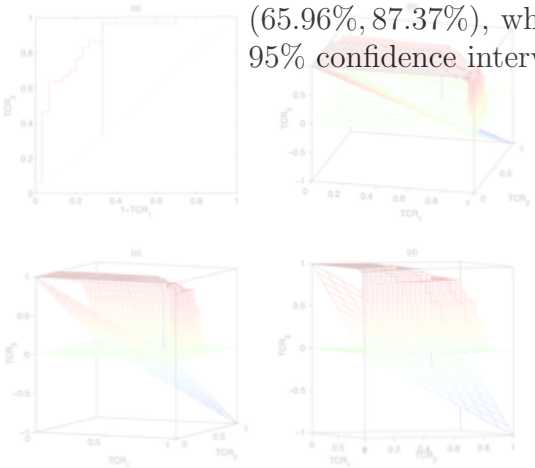


Table 1: Contingency table for MoCA test by true disease state. Numbers represent frequencies.

	'Intact'	'Minimal impairment'	'Mild dementia'	'Moderate dementia'
D−	71	9	0	0
D+	14	28	11	7

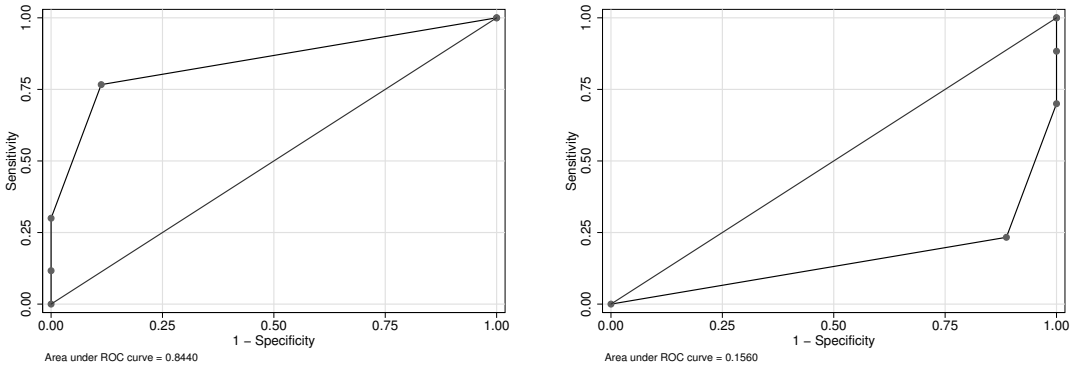


Figure 2: Conventional ROC curve for MoCA scores (left-hand panel) and inverted ROC curve (right-hand panel). For the right-hand panel, larger values are representative of non-diseased subjects in general.

Binary-scaled marker

Table 2 shows the results of a binary diagnostic marker in the form of a 2×2 contingency table. Then, sensitivity is 83.33% or 50/60 (73.90%, 92.76%), while specificity is $76/83 = 91.57\%$ (85.59%, 97.54%). 95% confidence intervals are given in parentheses.

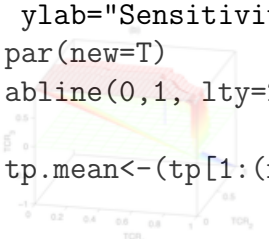
Table 2: Contingency table for a binary diagnostic marker. Numbers represent frequencies.

	Negative test: T−	Positive test: T+
D−	76	7
D+	10	50

R Lab section

The following set of commands may be used for the implementation in R of the CD4 data application:

```
  controls<-c(59, 66, 45, 62, 51, 50, 49, 58, 53, 42, 50,
47, 51, 62, 48)
  cases<-c(72, 70, 69, 82, 68, 59, 76, 61, 59, 73, 49, 77)

  n.con<-length(controls)
  n.cas<-length(cases)
  subj<-unique(sort(c(controls,cases)))
  subj.con<-unique(sort(controls))
  subj.cas<-unique(sort(cases))
  n.tot<-length(subj)

  tp<-rep(0,n.tot)
  fp<-rep(0,n.tot)
  for (i in 1:n.tot){
    tp[i]<-sum(cases >= subj[i])/n.cas
    fp[i]<-sum(controls >= subj[i])/n.con
  }
  tp<-c(1,tp,0)
  fp<-c(1,fp,0)

  plot(fp,tp, type='l', xlab="1-Specificity",
   ylab="Sensitivity")
  par(new=T)
  abline(0,1, lty=2)

  tp.mean<-(tp[1:(n.tot+1)] + tp[2:(n.tot+2)])/2
```

```
fp.diff<--diff(fp)
auc<-sum(fp.diff*tp.mean)
auc
```

In short, the code above asks R to estimate sensitivity (tp) and 1-specificity (fp) and plot the results, while the AUC is estimated according to the trapezoidal rule based on the (fp, tp) pairs, where fp corresponds to the x-axis and tp to the y-axis.

R package `asbio` offers a nice GUI for the depiction of an ROC curve. Figure 3 depicts the output given by using the function `see.roc.tck()`.

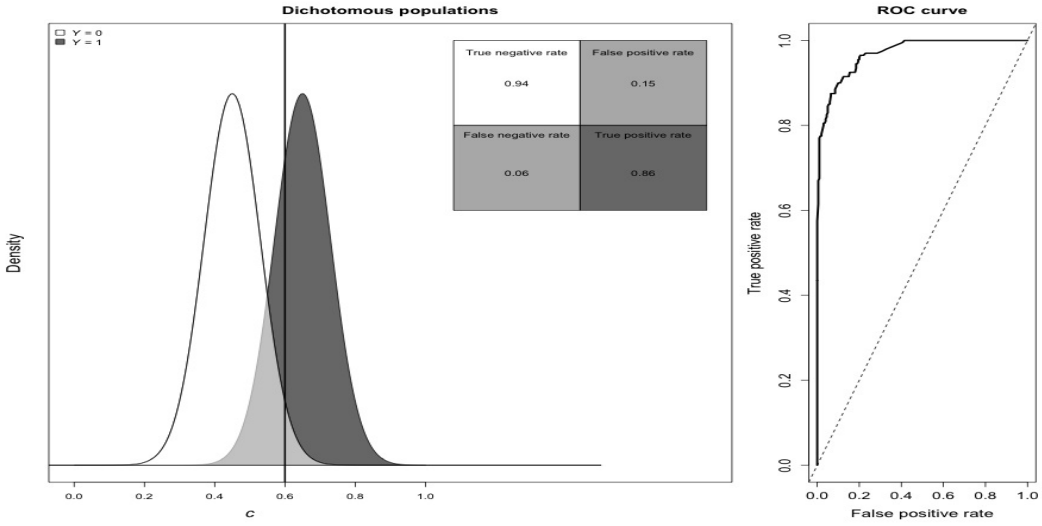
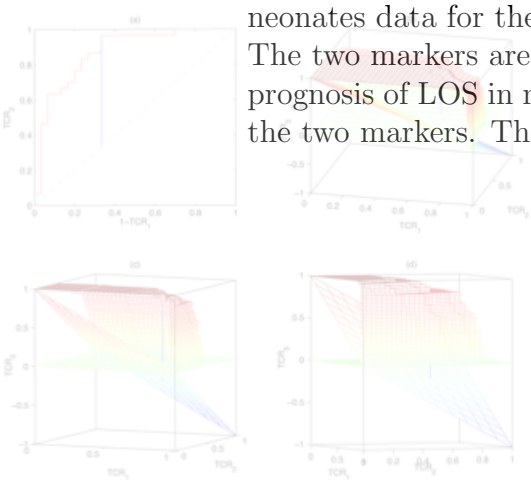


Figure 3: ROC curve illustration using the `asbio` package in R. $Y = 0$ refers to the control group, while $Y = 1$ refers to the case group.

Comparison of diagnostic markers

We illustrate the methods presented using the Late-onset sepsis (LOS) in neonates data for the comparison of diagnostic markers sTREM-1 and IL-6. The two markers are compared in terms of their diagnostic accuracy for the prognosis of LOS in neonates. Figure 4 depicts the empirical ROC curves for the two markers. The DeLong test results in $p = 0.053$.



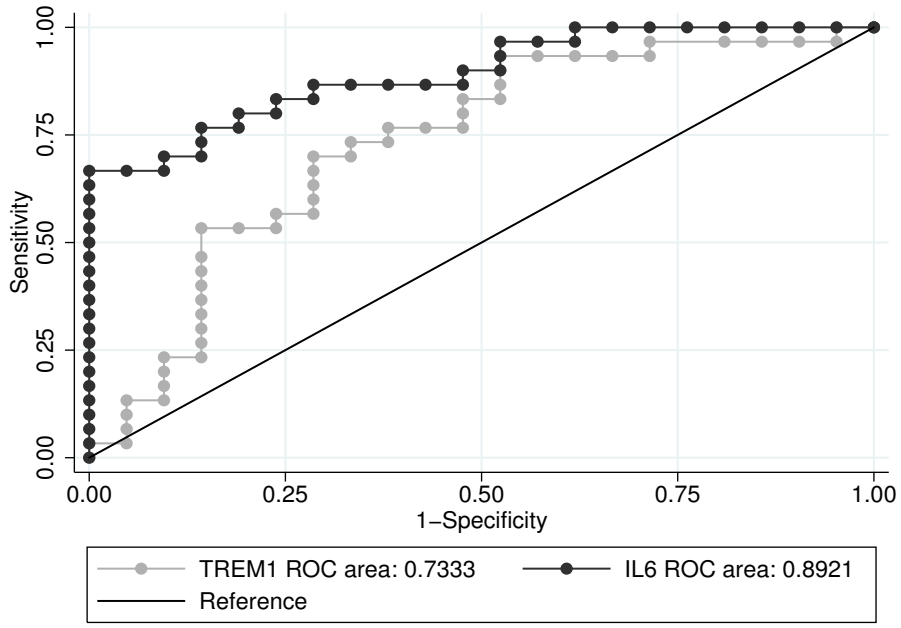


Figure 4: Illustration of sTREM-1 vs. IL-6 for the prediction of LOS in neonates, empirical ROC estimates are shown.

The empirical AUC for sTREM-1 is 0.733 (0.585, 0.882), while for IL-6 it is 0.892 (0.808, 0.976).

Equivalent results can be produced using the `pROC` package. Specifically, one may add the following to the respective code in the Lab Section:

```
library('pROC')
Zt<-trem2[,2]
Zi<-trem2[,3]
roctrem<-roc(Gold,Zt)
rocil6<-roc(Gold,Zi)
roc.test(roctrem,rocil6,method="delong")
roc.test(roctrem,rocil6,method="venkatraman")
roc.test(roctrem,rocil6,method="bootstrap")
```

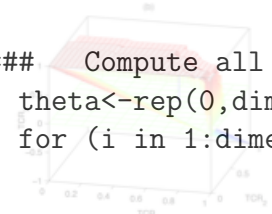
We get the same results for the `"delong"` choice in `method` and a $p = 0.084$ using `"venkatraman"` in `method` for the Venkatraman, Begg procedure, even though the latter is not a test for the AUCs but rather for the equality of

the empirical ROC curves. The `"bootstrap"` option resulted in $p = 0.055$. Based on U-statistics theory, it is expected in general that the bootstrap and U-statistics approaches will perform similarly.

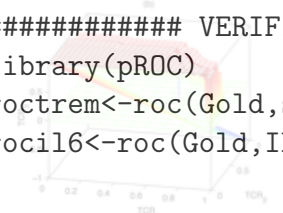
Lab section

`R` code for the nonparametric comparison of two paired AUCs (De Long test).

```
tremdata<-read.csv(file.choose())
#### choose tremdatabasic.csv
#### second column is the gold/reference standard
#### third and fourth columns trem1 & il6 measurements
#### respectively

trem2<-tremdata[,2:4]
Gold<-trem2[,1]
Z<-trem2[,2:3]
contrast<-t(c(1,-1))

#### R code for this section adapted from
#### http://www.ccs.miami.edu/~hishwaran/software.html
#### ustat.con is a function for AUC calculation
#### this was given in previous sessions.
#### u.contrast is a function implementing
#### the DeLong method for AUC comparison
u.contrast<-function(Gold,Z,contrast)
{
  dimen<-dim(Z)[2]
  Gold.uniq<-unique(sort(Gold))
  contrast<-as.matrix(contrast)
  n.neg<-length(Gold[Gold==Gold.uniq[1]])
  n.pos<-length(Gold)-n.neg
  Y<-Z[(Gold==Gold.uniq[1]),c(1:dimen)]   # Nondiseased group
  X<-Z[(Gold==Gold.uniq[2]),c(1:dimen)]   # Diseased group

###   Compute all areas (there are dimen number)
  theta<-rep(0,dimen)
  for (i in 1:dimen){
```

```
    theta[i]<-ustat.con(Y[,i],X[,i])
          }

###  Compute X,Y variance matrices
  V01<-matrix(0,n.neg,dimen,byrow=T)
  V10<-matrix(0,n.pos,dimen,byrow=T)

  for (row in 1:n.neg){
    for (col in 1:dimen){
      V01[row,col]<-ustat.con(Y[row,col],X[,col])
    }
  }
  for (row in 1:n.pos){
    for (col in 1:dimen){
      V10[row,col]<-ustat.con(Y[,col],X[row,col])
    }
  }
  S01<-(t(V01)%*%V01-n.neg*outer(theta,theta))/(n.neg-1)
  S10<-(t(V10)%*%V10-n.pos*outer(theta,theta))/(n.pos-1)
  Svar<-S10/n.pos+S01/n.neg
  Var<-contrast%*%Svar%*%t(contrast)
  test.L<-contrast%*%theta  #Test value
  test<-t(test.L)%*%solve(Var)%*%test.L #Chisquare value
  deg.f<-min(dim(contrast))
  chi.p<-1-pchisq(test,deg.f)

###     Return promised object:

  return(list(area=theta,testvalue=c(test.L),
                      chisquare=c(test,format(deg.f),chi.p),
                      Var=Var)) }

u.contrast(Gold,Z,contrast)

############ VERIFICATION (CHECK) ############
library(pROC)
roctrem<-roc(Gold,sTREM1.BC)
rocil6<-roc(Gold,IL6.BC)
```

```
roc.test(roctrem,rocil6,method="delong")
roc.test(roctrem,rocil6,method="venkatraman")
roc.test(roctrem,rocil6,method="bootstrap")
```