Simulated ROC

jdt

12/13/2020

Set seed for the simulation

```
  # clear the environment and set seed
rm(list = ls())
set.seed(12345)
```

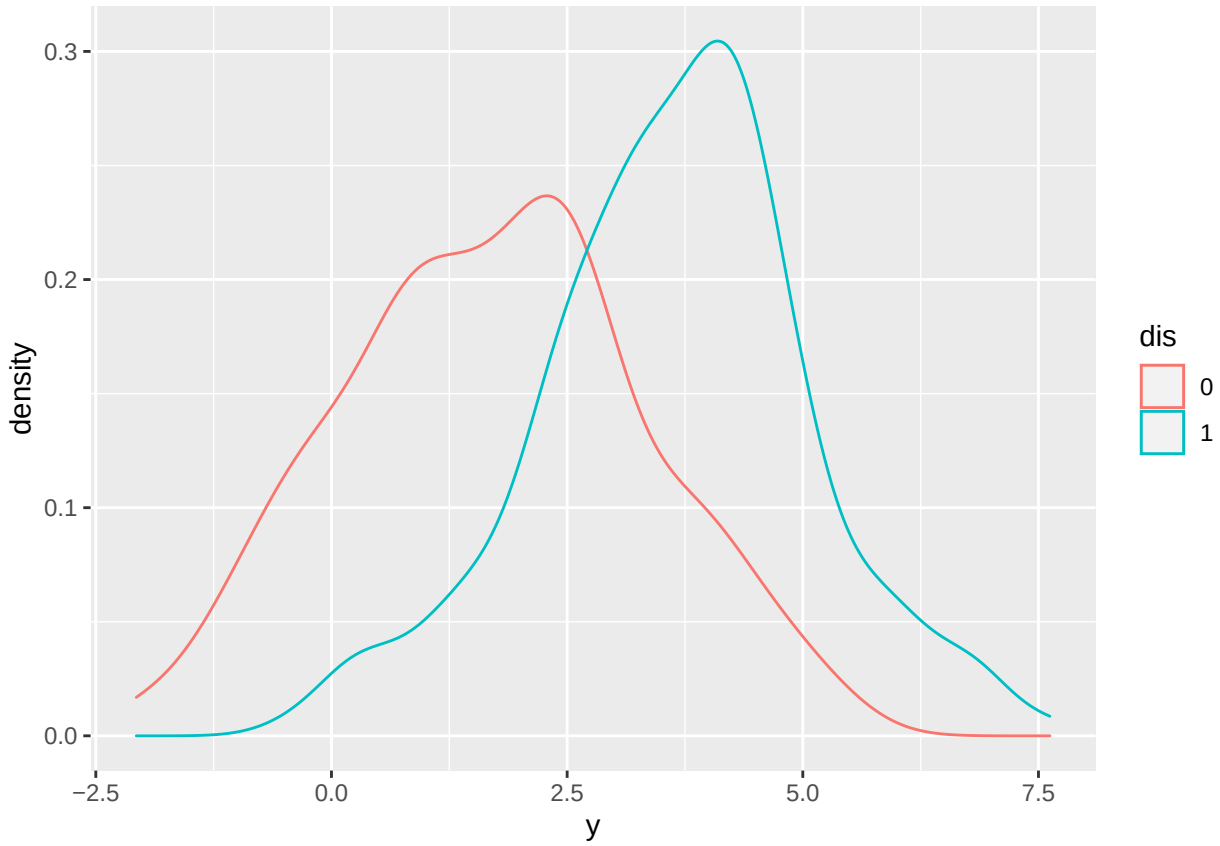
Function to generate normal data. For dis=0 ("Control") and dis=1 ("Disease"). The separation between the two groups is controlled by one's choice of c0, c1, sd_e0, and sd_e1. n0 and n1 are the sample sizes for the groups.

```
library(ggplot2)

gen_Norm_data = function(c0,sd_e0,n0,c1,sd_e1,n1){
  eps0 = rnorm(n0,0,sd_e0)
  eps1 = rnorm(n1,0,sd_e1)
  y0 = c0 + eps0
  y1 = c1 + eps1

  data.frame(y = c(y0, y1),
             dis = as.factor(c(rep(0,n0), rep(1,n1))))
}
dat1 = gen_Norm_data(1.5,1.5,200,3.5,1.5,200)

#plot density functions
ggplot(dat1, aes(color = dis, y)) + geom_density()
```



Create discrete variables for Y

```
y=dat1[,1]
summary(y)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  -2.071   1.395   2.737   2.672   4.024   7.621
```

```
case=dat1[,2]
```

Create binary table with cutoff y > 2.5

```
high_y = y > 2.5
tex1 = table(case,high_y)
addmargins(tex1)
```

```
##      high_y
## case  FALSE TRUE Sum
##   0     137   63 200
##   1      40  160 200
##   Sum   177  223 400
```

```
prop.out = prop.table(tex1,1)
specificity = prop.out[1,1]
sensitivity = prop.out[2,2]
prop.out
```

```
##     high_y
## case FALSE  TRUE
##    0 0.685 0.315
##    1 0.200 0.800
```

```
sensitivity
```

```
## [1] 0.8
```

```
specificity
```

```
## [1] 0.685
```

```
TPR = sensitivity
FPR = 1 - specificity
TPR
```

```
## [1] 0.8
```

```
FPR
```

```
## [1] 0.315
```

Create binary table with cutoff y > 3.0

```
high_y = y > 3.0
tex1 = table(case,high_y)
addmargins(tex1)
```

```
##      high_y
## case  FALSE TRUE Sum
##   0     161   39 200
##   1      60  140 200
##   Sum   221  179 400
```

```
prop.out = prop.table(tex1,1)
specificity = prop.out[1,1]
sensitivity = prop.out[2,2]
prop.out
```

```
##     high_y
## case FALSE  TRUE
##    0 0.805 0.195
##    1 0.300 0.700
```

```
sensitivity
```

```
## [1] 0.7
```

```
specificity
```

```
## [1] 0.805
```

```
TPR = sensitivity
FPR = 1 - specificity
TPR
```

```
## [1] 0.7
```

```
FPR
```

```
## [1] 0.195
```

Instead of creating a table at each cutoff point one can construct a ROC plot for a continuous variable Y. Which is constructing using the pairs (TPR, FPR) at each cutoff point (in R). Sometimes (as in SAS) this curve is smoothed.

```
library(ROCR)
pred = prediction(y,case)
perf=performance(pred, "tpr", "fpr")
plot(perf)
```