Sensitivity, specificity and ROC

PhD Flavio Lichtenstein

2071, July 22th

Two distributions - t-student test (t.test)

Here we will present 2 distributions and makey them close togheter or far appart. The statistics that measures the distance between two normal distribuitons is t-student.
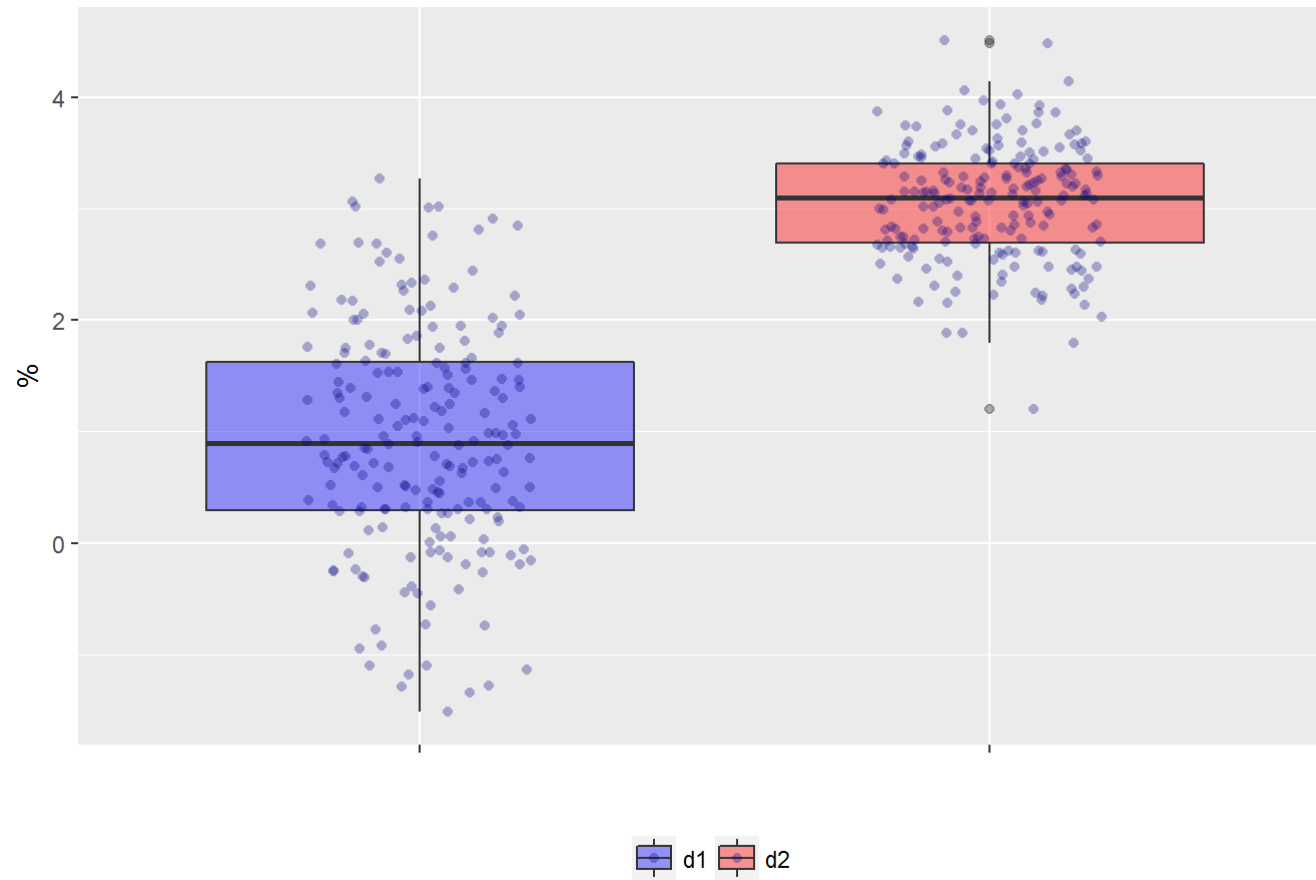
```
source("c:/colaboracoes/losac/R/support.R")

d1 = rnorm(200, mean=1, sd=1)
d2 = rnorm(200, mean=3, sd=.5)

data = data.frame(d1, d2)
options(digits=5)
head(data)
```

```
##          d1      d2
## 1   0.72062 3.0741
## 2   0.98587 2.3686
## 3   1.60485 3.2819
## 4  -0.23900 2.8000
## 5   0.51608 3.5503
## 6   1.17288 3.5979
```
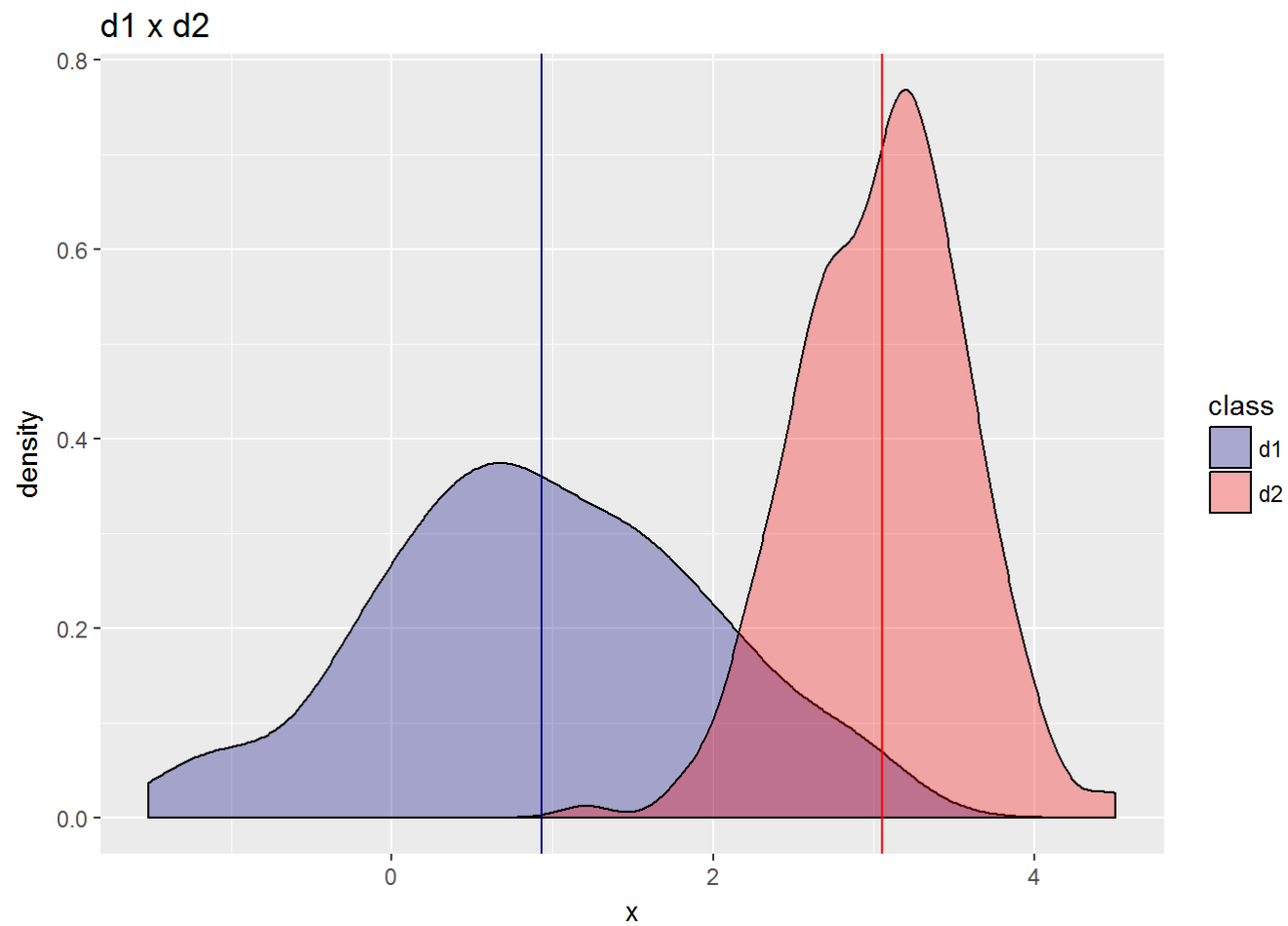
```
my.boxplot(data, classX=c("one","two"), cols=c(1,2), title="normal distribution", ylab="%", colors=c("blue", "red"),
is.log=F,ylim=NA)
```

normal distribution

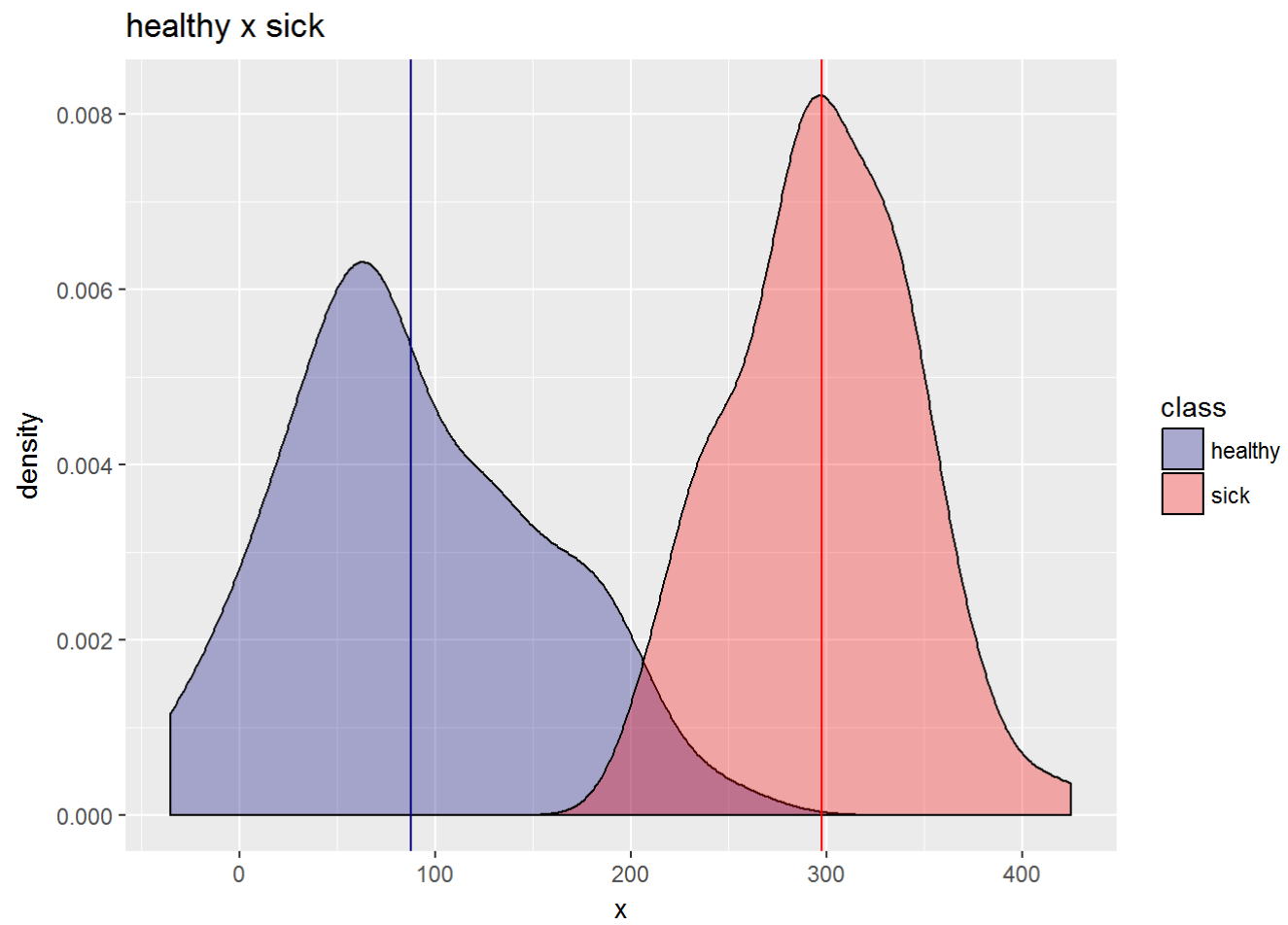Can we define if the distributions are sufficient far appart?

```
print_2densities(d1, d2)
```

d1 x d2


```
healthy = rnorm(100, mean=1, sd=.7)
sick    = rnorm(100, mean=3, sd=.5)

healthy = round(healthy*100)
sick = round(sick*100)

print_2densities(healthy, sick, title="healthy x sick", classes=c("healthy", "sick"))
```
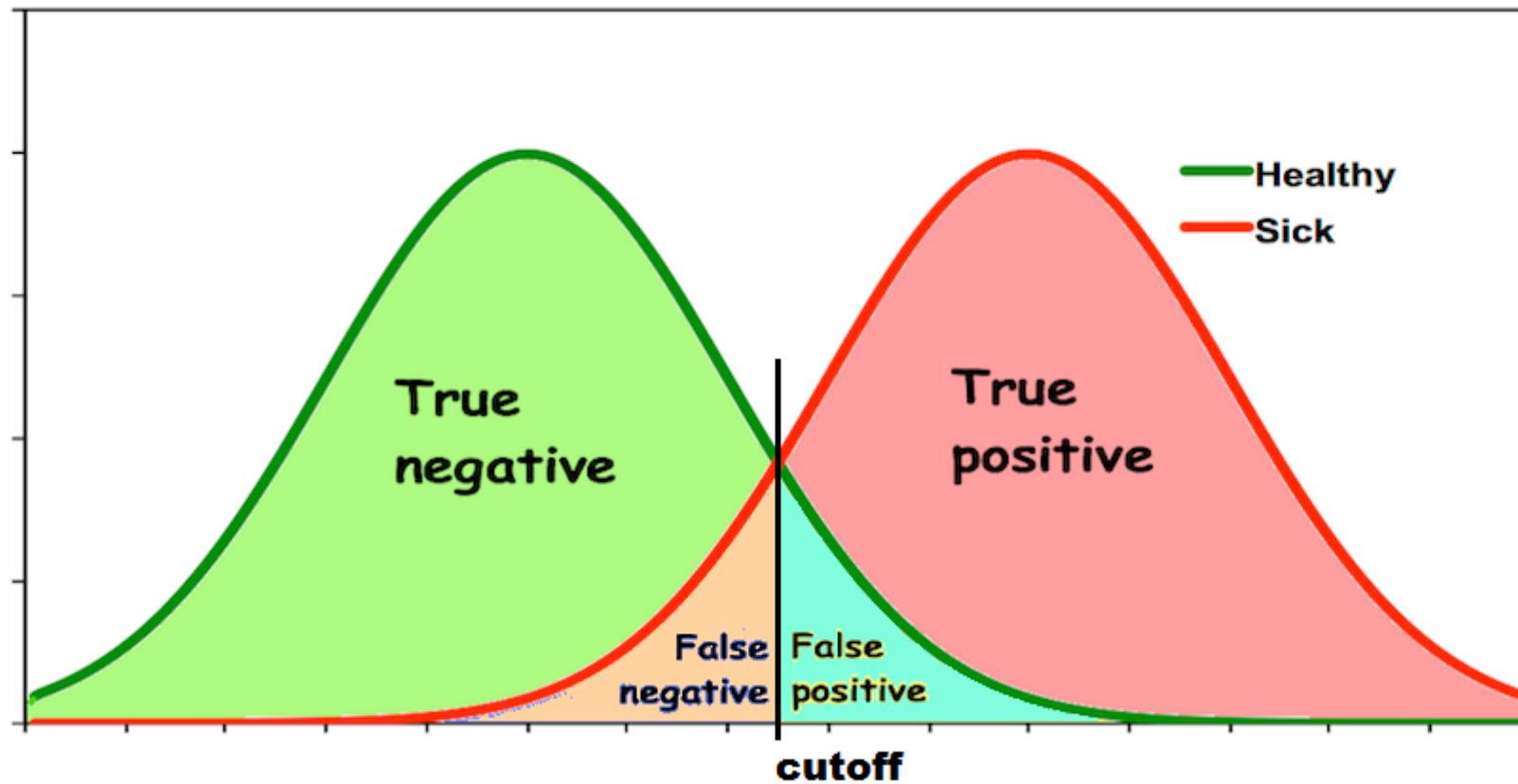
healthy x sick

GOLD STANDARD

	Sick	Healthy
Test+	True Positives (TP)	False Positives (FP)
Test-	False Negatives (FN)	True Negatives (TN)

Classification Outcome

Conditional: truth table

Sensitity

Sensitivity, recall, hit rate, or true positive rate (TPR)

$$Sensitivity = TPR = \frac{TP}{P} = \frac{TP}{TP + FN}$$

	Sick	Healthy
Test+	True Positives (TP)	False Positives (FP)
Test-	False Negatives (FN)	True Negatives (TN)

Conditional: sensitivity

Especifity

specificity or true negative rate (TNR)

$$Especifity = TNR = \frac{TN}{N} = \frac{TN}{TN + FP}$$

Conditional: specificity

Precision

precision or positive predictive value (PPV)

$$Precision = PPV = \frac{TP}{TP + FP}$$

ROC - Receiver Operating Curve

In statistics, a receiver operating characteristic curve, i.e. ROC curve, is a graphical plot that illustrates the diagnostic ability of a binary classifier system as its discrimination threshold is varied.

The ROC curve is created by plotting the true positive rate (TPR) against the false positive rate (FPR) at various threshold settings. The true-positive rate is also known as sensitivity, recall or probability of detection[1] in machine learning.

https://en.wikipedia.org/wiki/Receiver_operating_characteristic (https://en.wikipedia.org/wiki/Receiver_operating_characteristic)
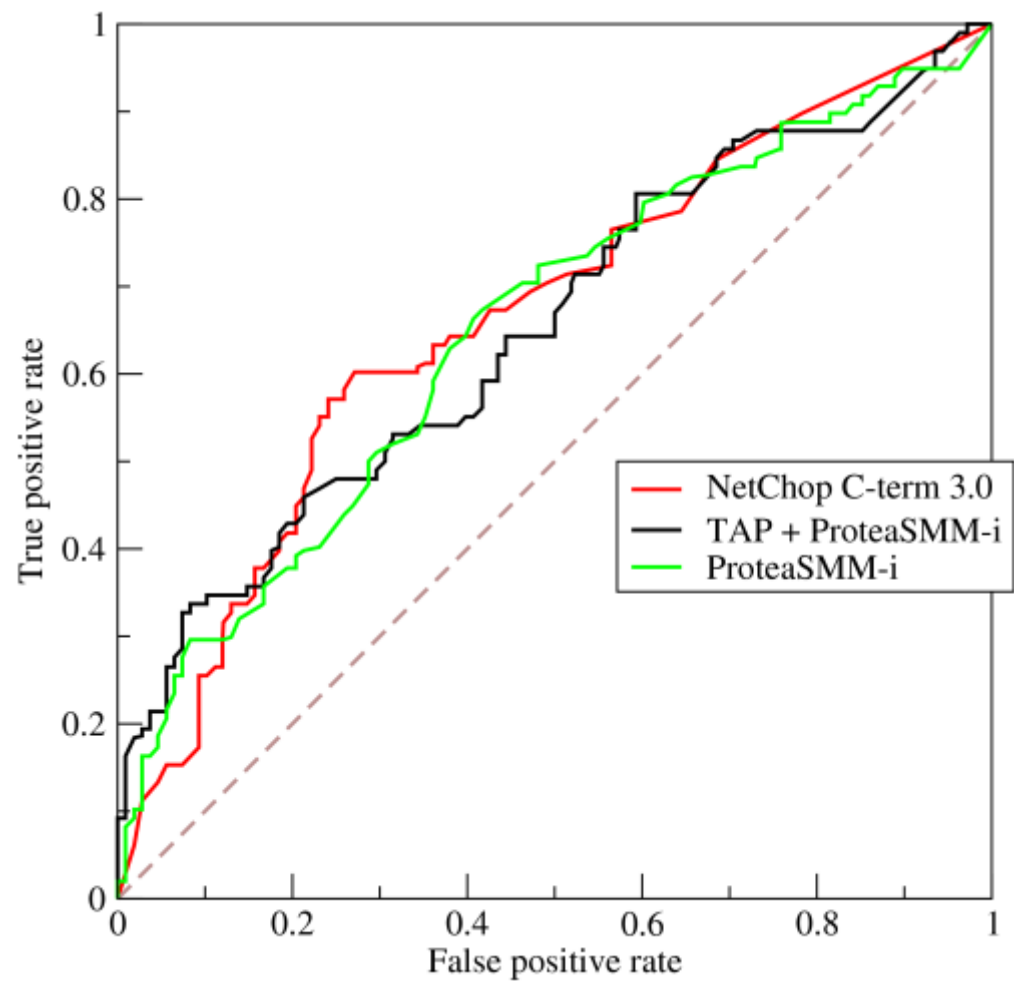
```
# lista = contigency_table(healthy, sick)
# names(lista)

healthy = rnorm(100, mean=1, sd=.3)
sick    = rnorm(100, mean=2, sd=.3)

lista = doRoc(healthy, sick, "healthy", "sick", "healthy", "sick", xlab="measure", isLog=F, titleAux="")
names(lista)
```
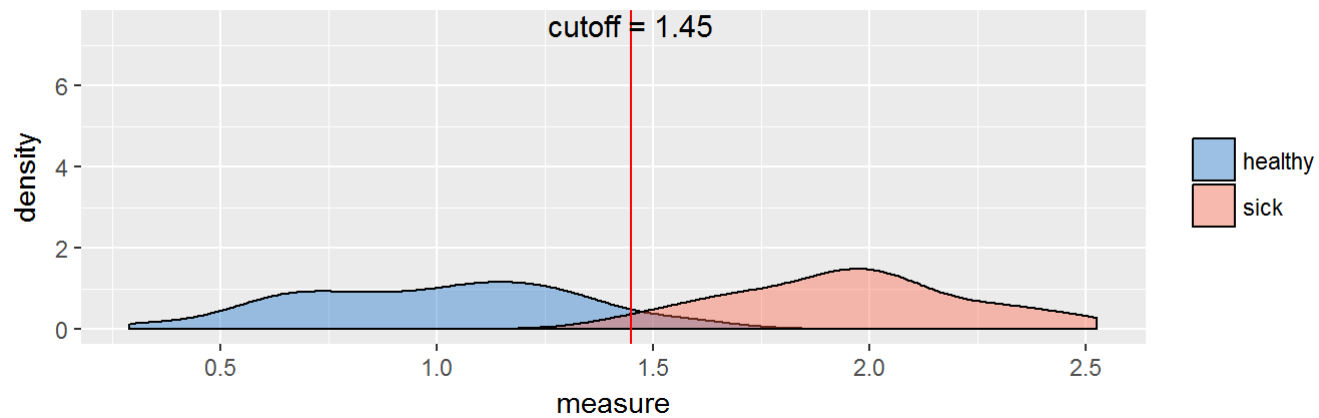
```
## [1] "p1"        "p2"        "best_sens" "best_spec" "cutoff"
```

```
p1 = lista[["p1"]]
p2 = lista[["p2"]]

multiplot(p1,p2,layout=matrix(c(1,2),ncol=1, byrow=F))
```