# Análise Dataset de Cárdio

José Elvano Moraes

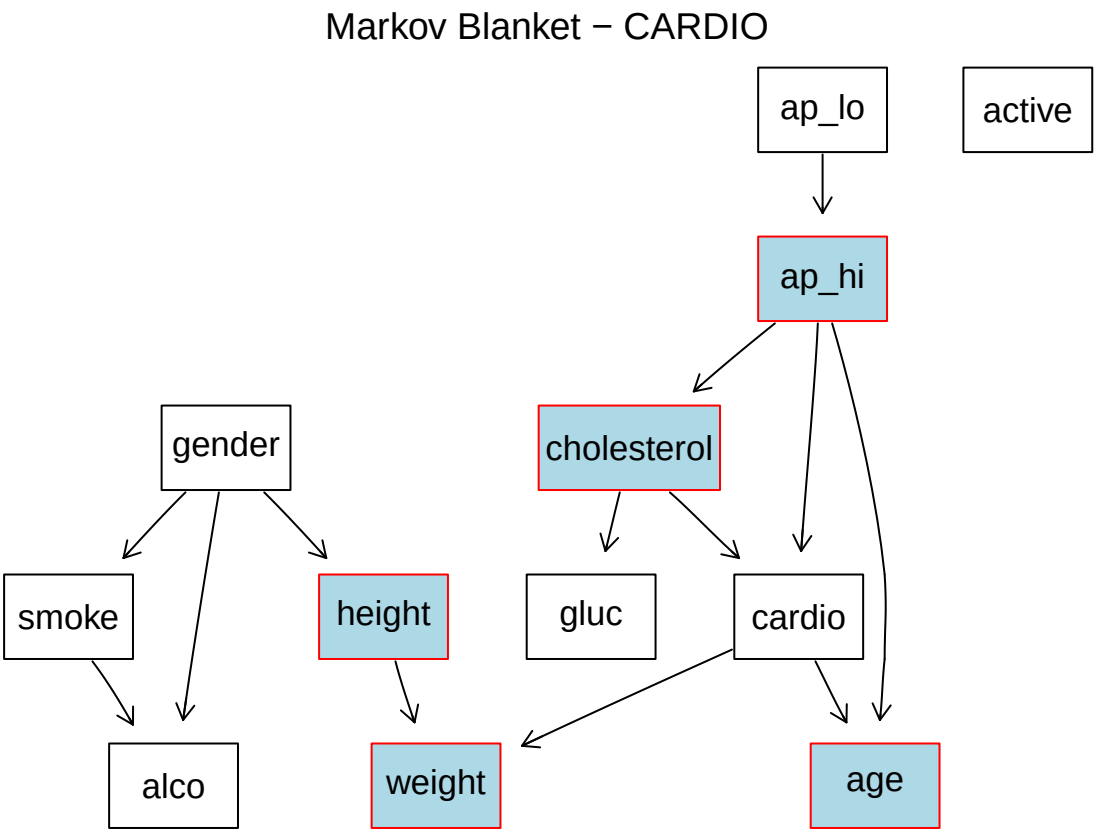
4/27/2021

## Variáveis

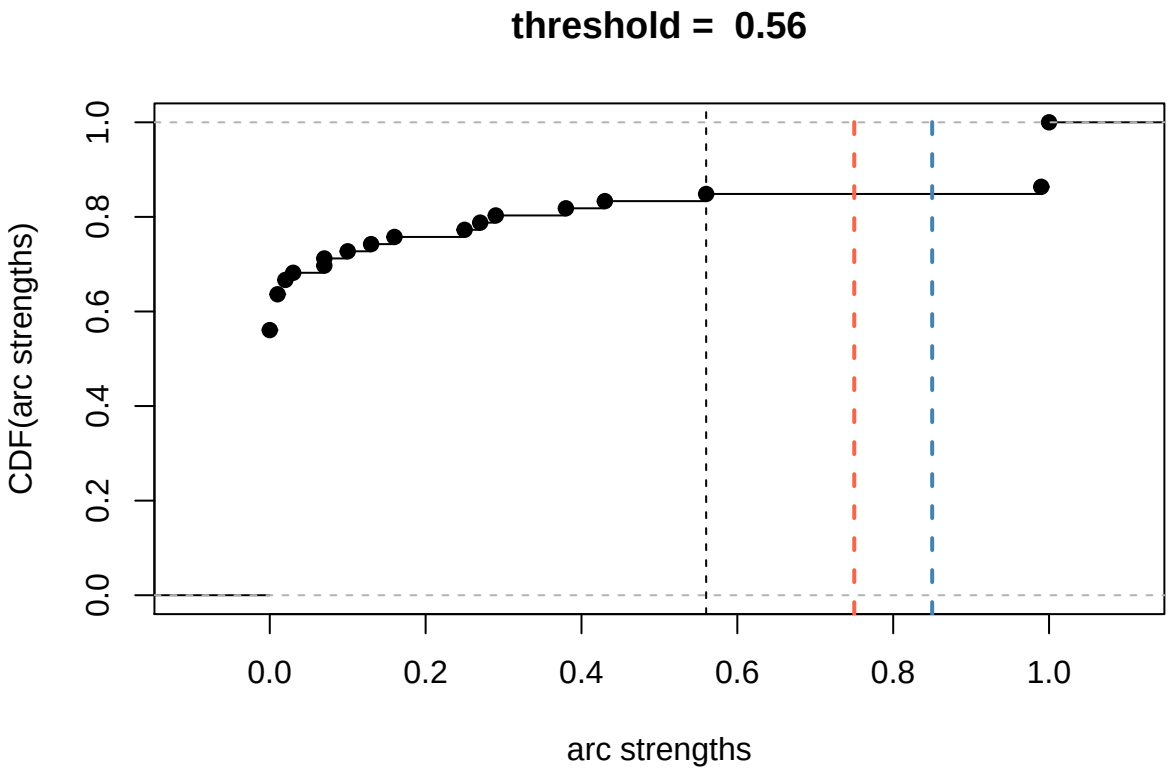
- Idade
- Sexo
- Altura
- Peso
- Pressão arterial sistólica
- Pressão arterial diastólica
- Colesterol
- Glicemia
- Tabagismo
- Consumo de álcool
- Sedentarismo
- **Variável alvo: cardio** apresenta doença cardíaca ou não

```
##  gender   cholesterol gluc     smoke    alco     active   cardio
##  1:4780   1:5495      1:6227   0:6700   0:6951   0:1497   0:3693
##  2:2571   2:1012      2: 578   1: 651   1: 400   1:5854   1:3658
##           3: 844      3: 546
##
##                 age             height          weight          ap_lo
##  (10878,17320.5]  :1544   (76,159] :1936   (30,57] : 613   (10,70] :1385
##  (17320.5,19108.9]:1617   (159,164]:1631   (57,64] :1145   (70,80] :3801
##  (19108.9,21169.4]:2131   (164,170]:2371   (64,80] :3575   (80,90] :1621
##  (21169.4,23670]  :2059   (170,184]:1413   (80,117]:2018   (90,100]: 544
##        ap_hi
##  (70,110] :1334
##  (110,120]:3029
##  (120,140]:2034
##  (140,180]: 954
```

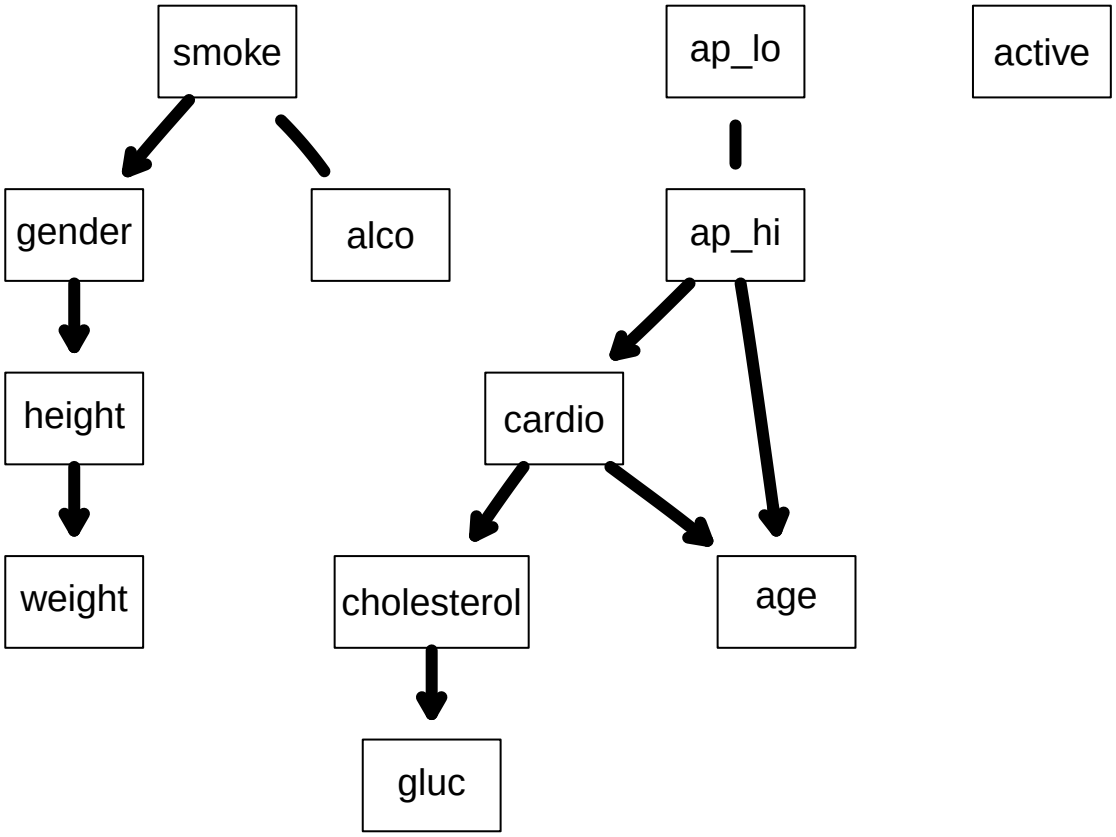
**Grafo mostrando como nodos coloridos as variáveis das quais a variavel CARDIO tem dependência estatística (*Markov blanket*)**



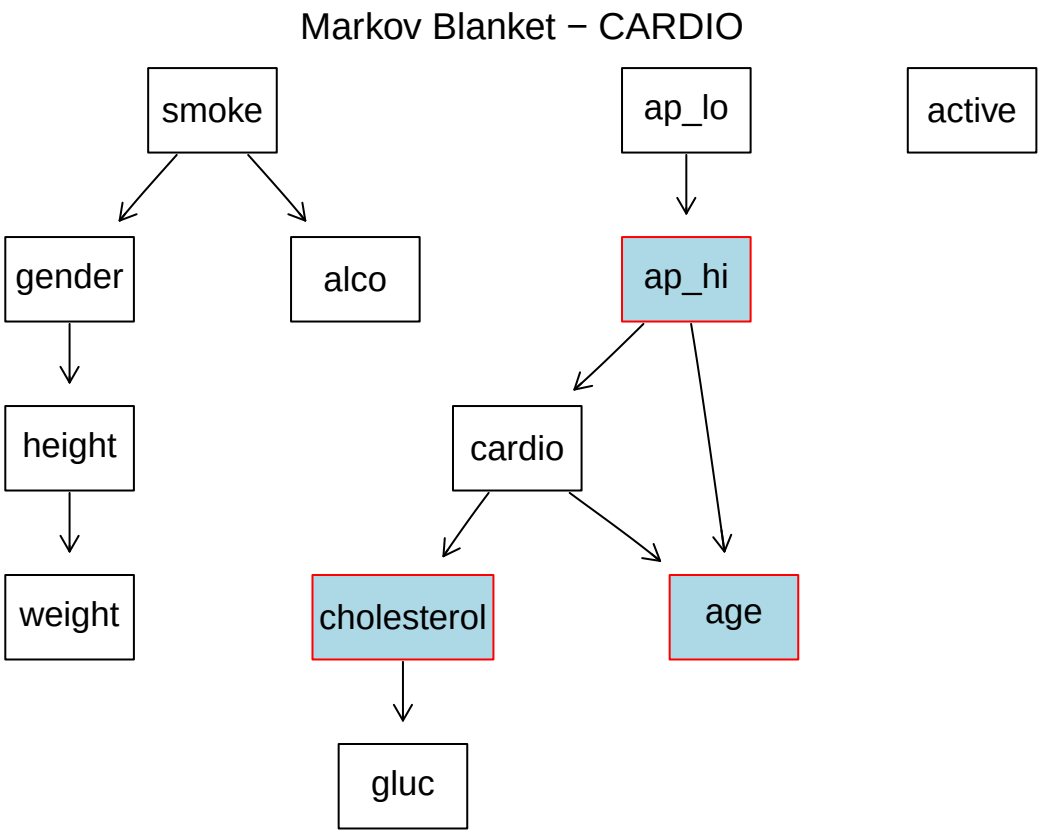
**CDF da correlação (correlação de Pearson) entre os nodos do grafo**



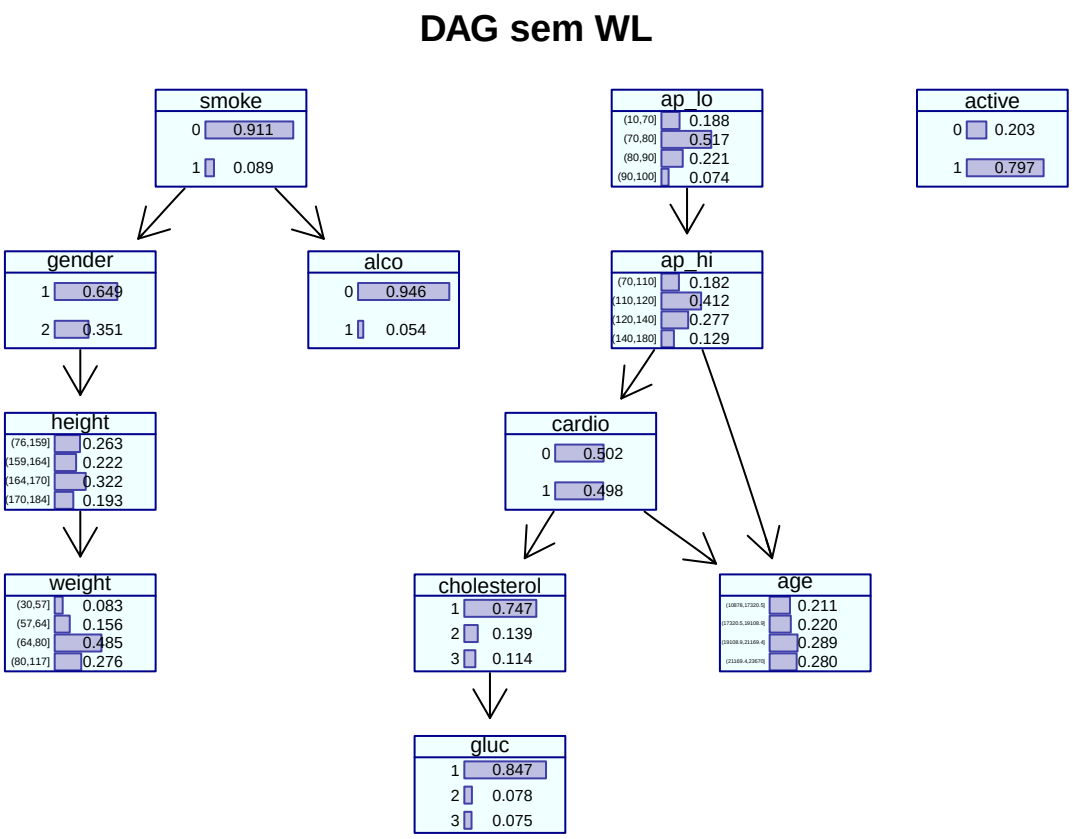
**Grafo médio após 100 iterações**



**Grafo médio tornado um *DAG***



**Distribuição de probabilidade (PDF) de cada variável**



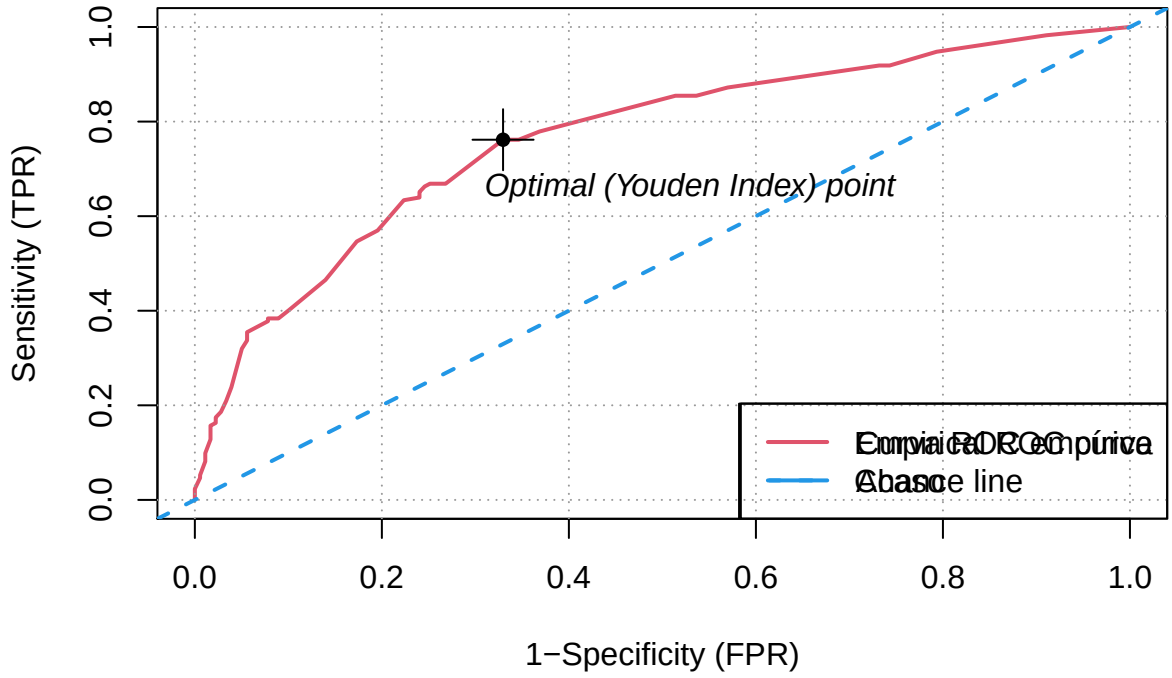

---

# Predição

**Desempenho estatístico da rede**

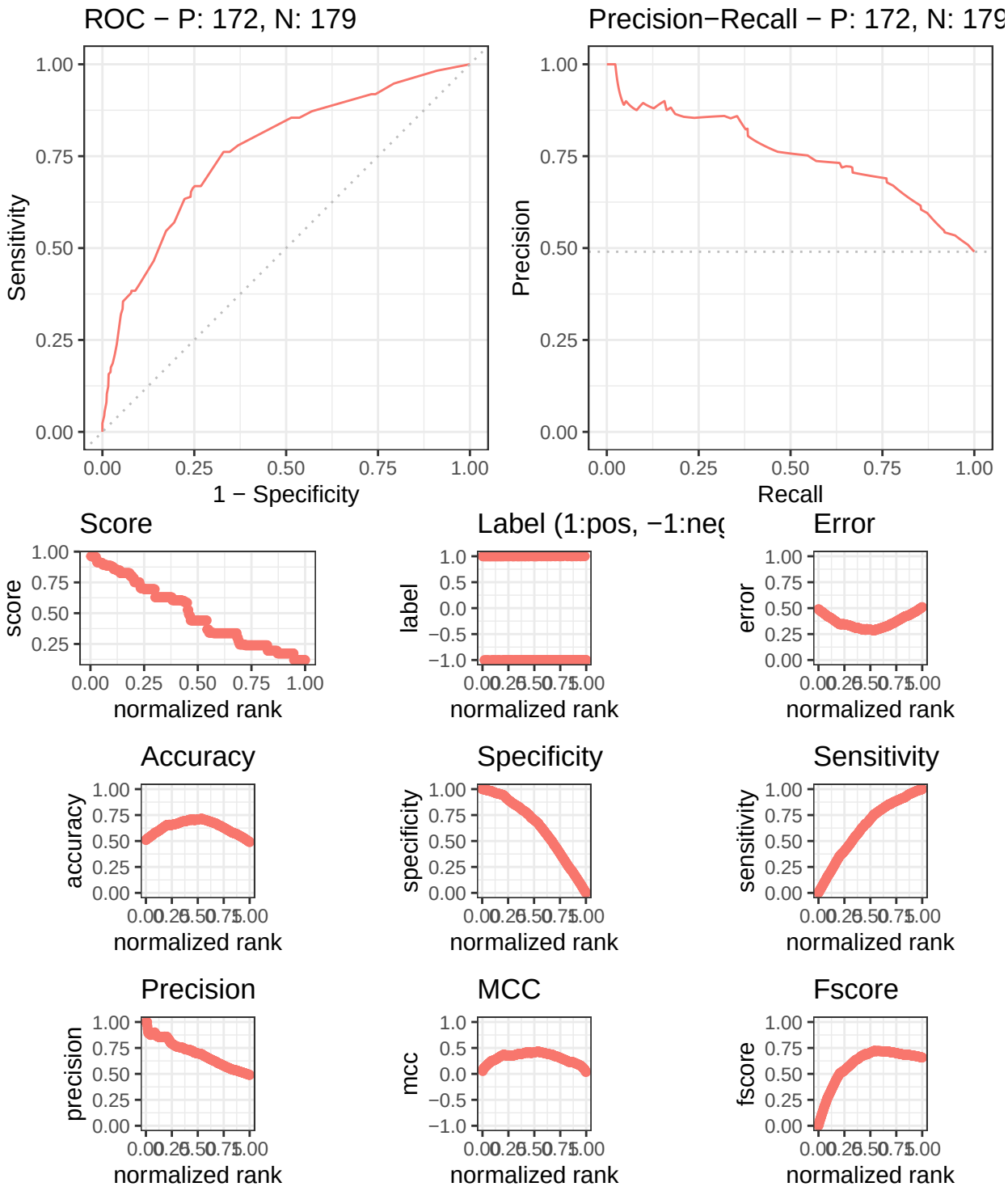
```
##           Outcome +    Outcome -      Total
## Test +          134           45        179
## Test -           57          115        172
## Total           191          160        351
##
## Point estimates and 95 % CIs:
## -------------------------------------------------------
## Apparent prevalence                   0.51 (0.46, 0.56)
## True prevalence                       0.54 (0.49, 0.60)
## Sensitivity                           0.70 (0.63, 0.77)
## Specificity                           0.72 (0.64, 0.79)
## Positive predictive value             0.75 (0.68, 0.81)
## Negative predictive value             0.67 (0.59, 0.74)
## Positive likelihood ratio             2.49 (1.91, 3.25)
## Negative likelihood ratio             0.42 (0.33, 0.53)
## -------------------------------------------------------
```

**Curva ROC**



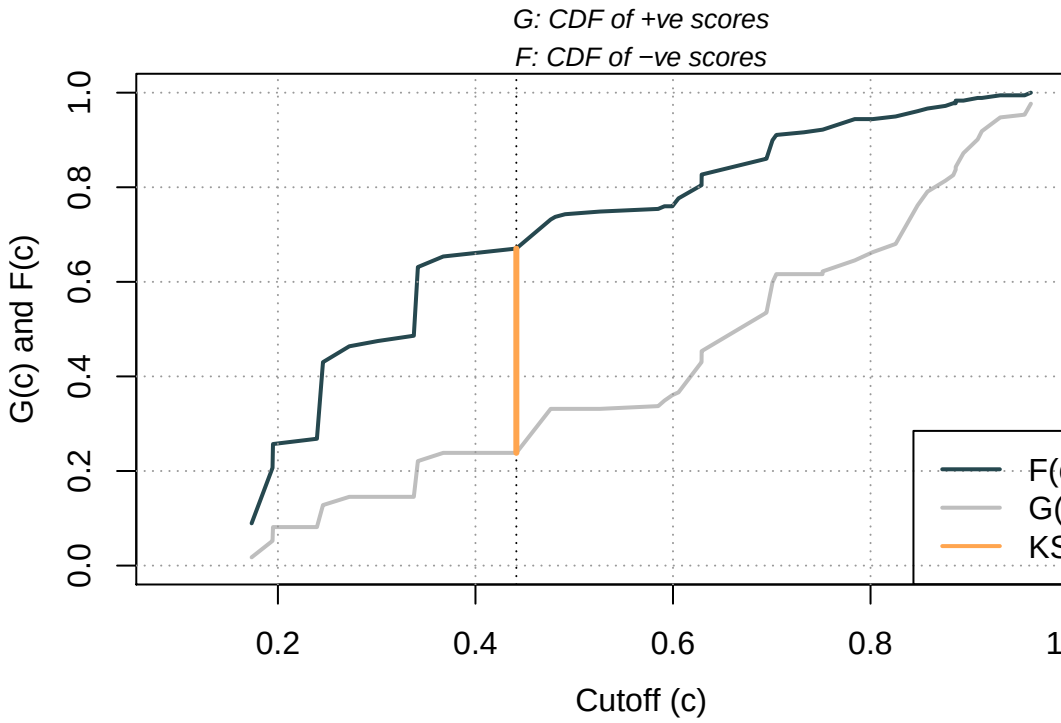

```
##
##  Method used: empirical
##  Number of positive(s): 172
##  Number of negative(s): 179
##  Area under curve: 0.7644
##
## 1     Method used: empirical
## 2 Number of positive(s): 172
## 3 Number of negative(s): 179
## 4   Area under curve: 0.7644
```

**Outra implementação de curva ROC e outras estatísticas**




**KS plot**

KS plot shows the cumulative density functions F(c) and G(c) in the positive and negative populations. If the positive population have higher value, then negative curve (F(c)) ramps up quickly. The KS statistic is the maxi-




mum difference of F(c) and G(c).

```
## KS Stat (empirical) : 0.432018968429258
```

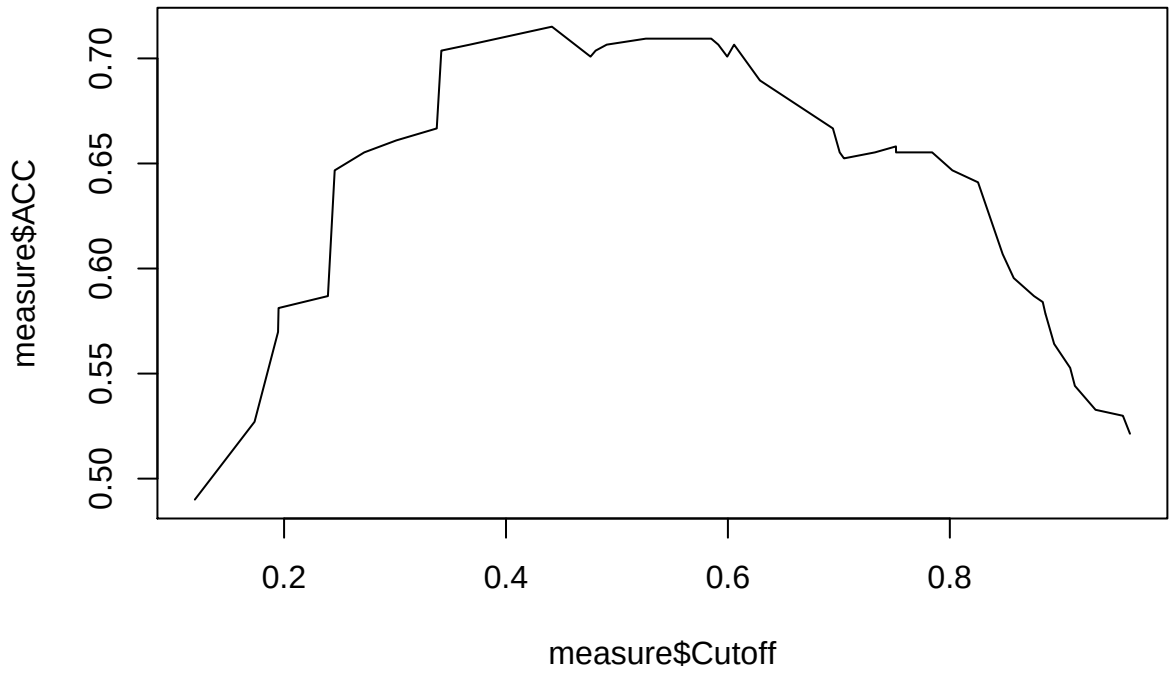
```
## KS Stat (empirical) cutoff : 0.441399200963467
```

# TODO

- ACC: Overall accuracy of classification.
- MIS: Misclassification rate.
- SENS: Sensitivity.
- SPEC: Specificity.
- PREC: Precision.
- REC: Recall. Same as sensitivity.
- PPV: Positive predictive value.
- NPV: Negative predictive value.
- TPR: True positive rate.
- FPR: False positive rate.
- TNR: True negative rate.
- FNR: False negative rate.
- pDLR: Positive diagnostic likelihood ratio.
- nDLR: Negative diagnostic likelihood ratio.
- FSCR: F-score

```
measure <- measureit(score = score, class = class,
                     measure = c("ACC", "SENS", "FSCR"))

plot(measure$ACC~measure$Cutoff, type = "l")
```


measure$ACC
0.50
0.55
0.60
0.65
0.70
0.2
0.4
0.6
0.8
measure$Cutoff