Assignment 4

Kishan Sarpangala

**Question**:

Aneurysmal subarachnoid hemorrhage (aSAH) is a serious medical condition with bleeding occurring in the brain. There are several causes. Head trauma is one of them. Neurosurgeons, to begin with, focus on stopping the bleeding and preventing its reoccurrence. Even if successful in stopping bleeding, brain damage with cognitive impairment could occur subsequently. In the class I attempted to develop a diagnostic test to predict brain damage based on a single protein biomarker s100b. There are other ways to develop a prediction model. I want you to pursue these other prediction models.

**Answer**:

1. Write a short note on the protein biomarker s100b from internet sources (not more than 5 sentences). 2 points

- associated with Alzheimer's disease.

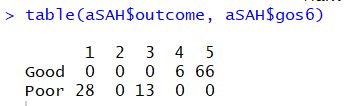
-Calcium binding protein

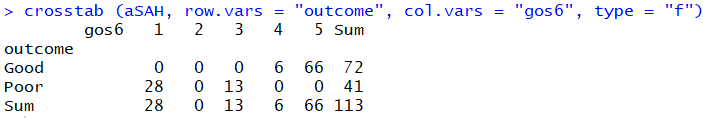
- s100b belongs to s100 family of proteins,

- localized in a variety of cells in the nucleus and cytoplasm,

- involved in the regulation of cell cycle progression and differentiation

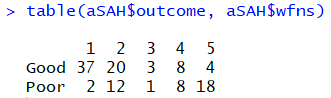
2. Cross-tabulate ‘outcome’ and ‘gos6.’ Comment on the table. 2 points

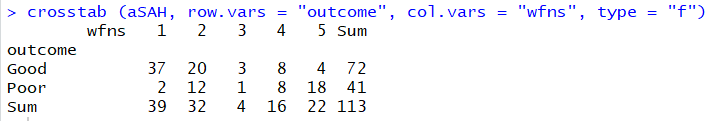


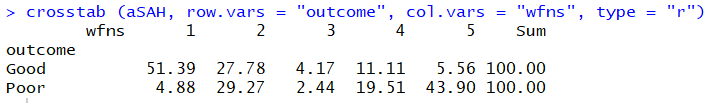


Good outcomes are predominantly in group 4 and 5, while 1 and 3 have the poor putcomes with respect to gos6.

3. Cross-tabulate ‘outcome’ and ‘wfns.’ Comment on the table. 2 points

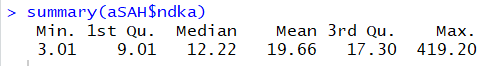


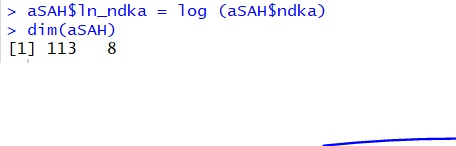


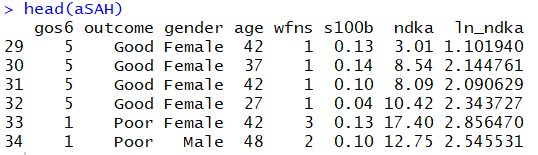


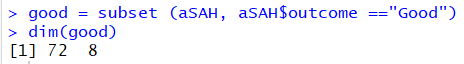
Wfns groups 1 and 2 have more good outcomes, while the poor ones are spread across gps 2, 4, 5.

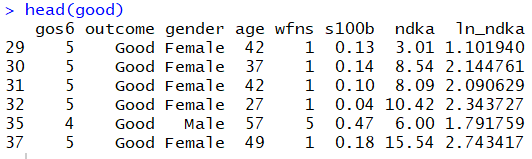
4. I want you to explore whether the protein ‘ndka’ could be a good biomarker. There is enormous variation in the variable. Take ln transformation of the variable. I want you to work on log(ndka). Segregate the into ‘good’ or ‘bad’ outcomes. Plot the density curves of log(ndka) for ‘good’ and ‘bad’ outcomes in the same graph. Comment on the graph. 5 points

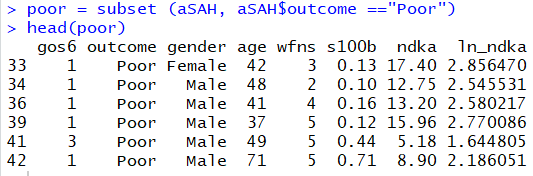








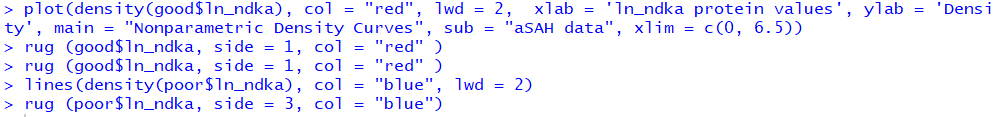


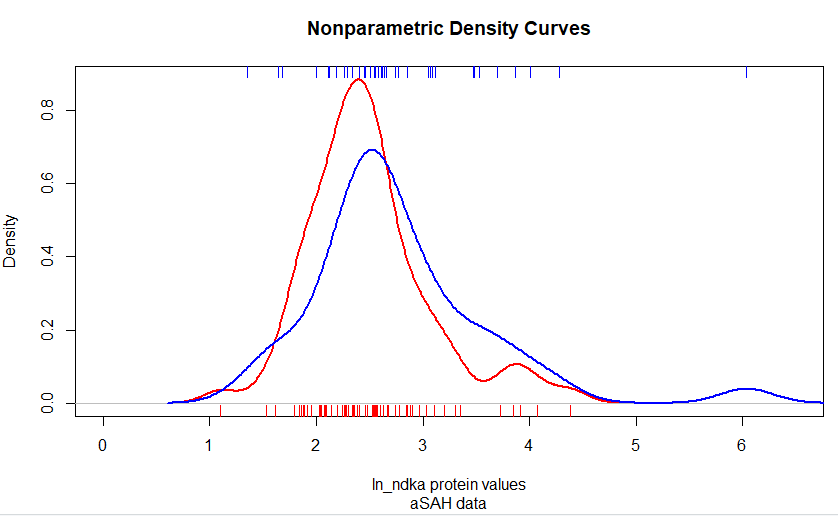




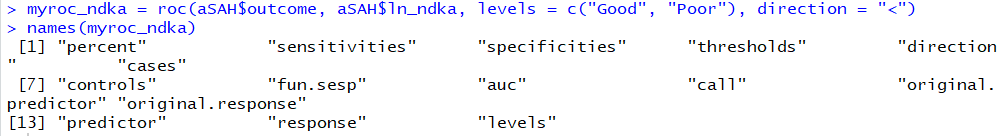


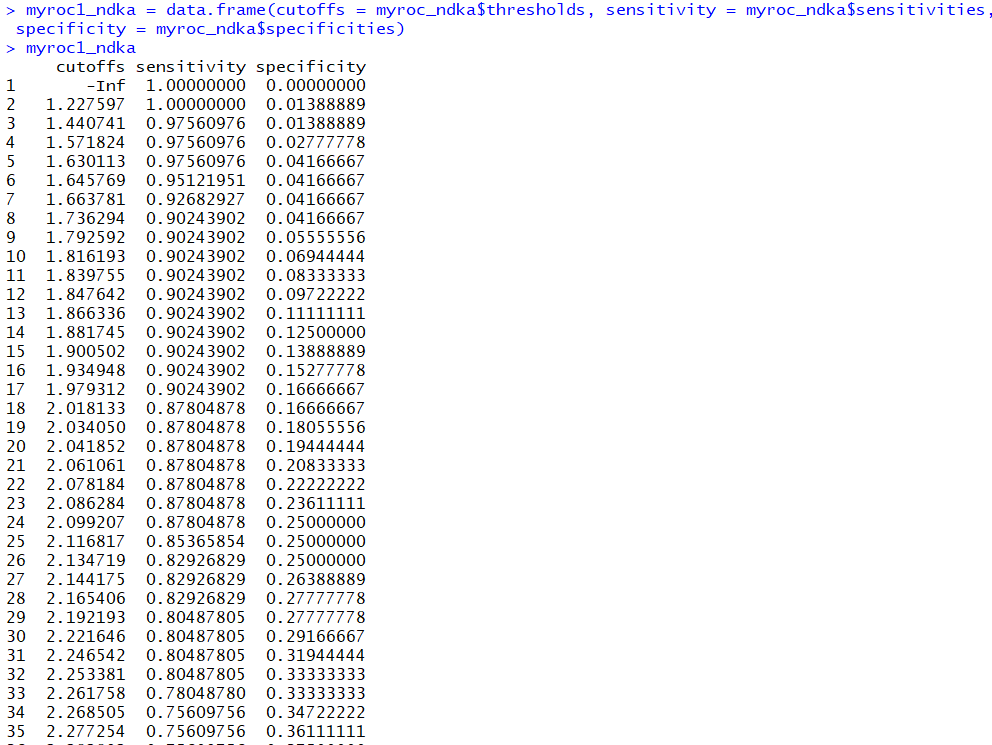


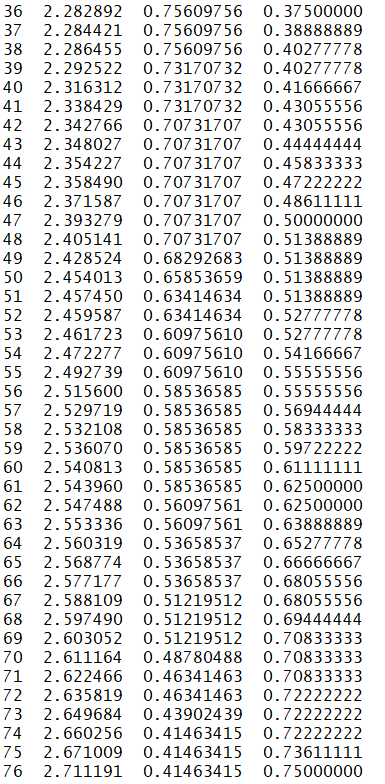


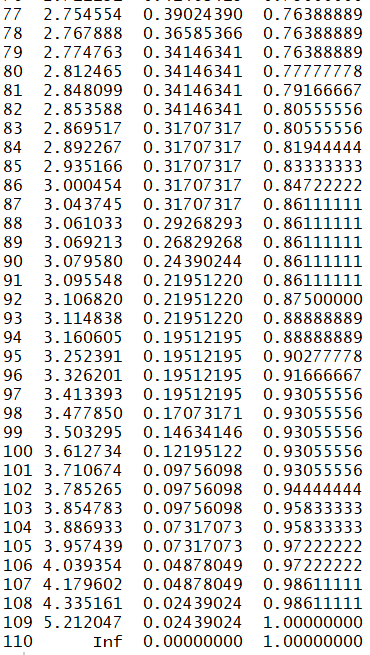


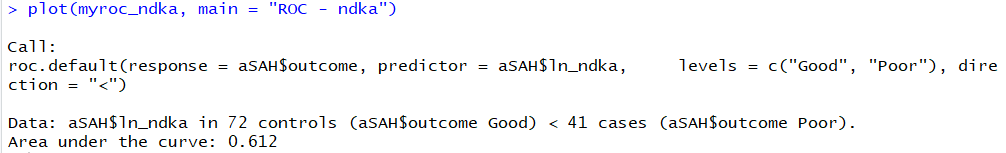
5. Invoke the ‘roc’ function with log(ndka) as a biomarker. Plot the ROC curve that comes with it. 5 points

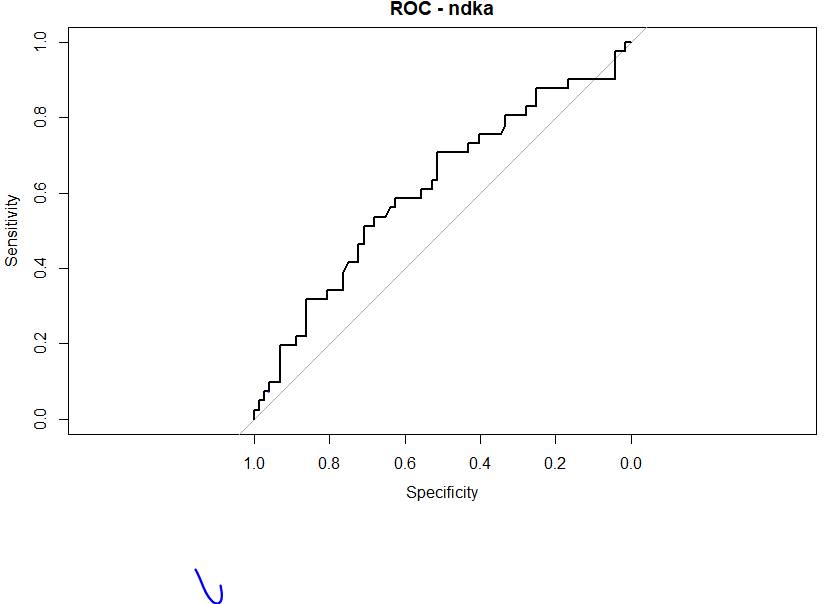




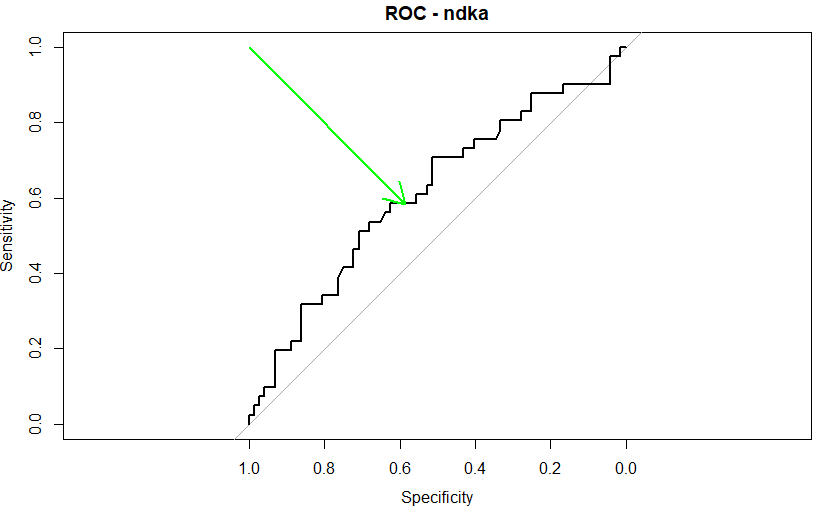








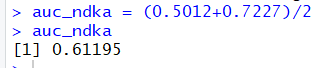
After including 

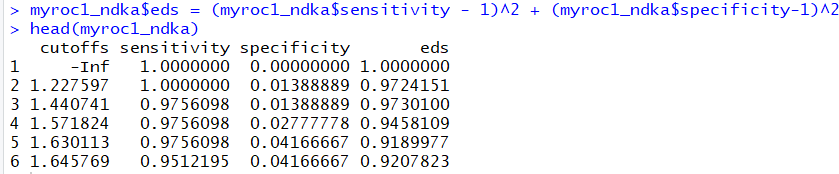


6. Find a 95% confidence interval for AUC. Comment on the utility of the biomarker ‘log(ndka).’ 4 points

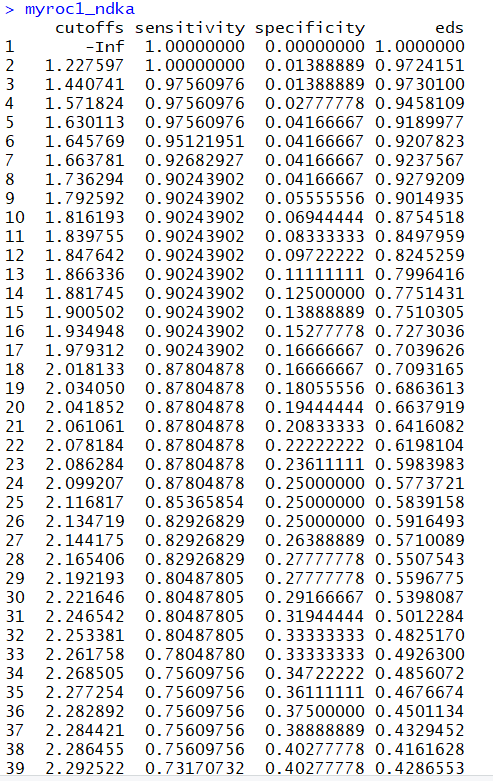


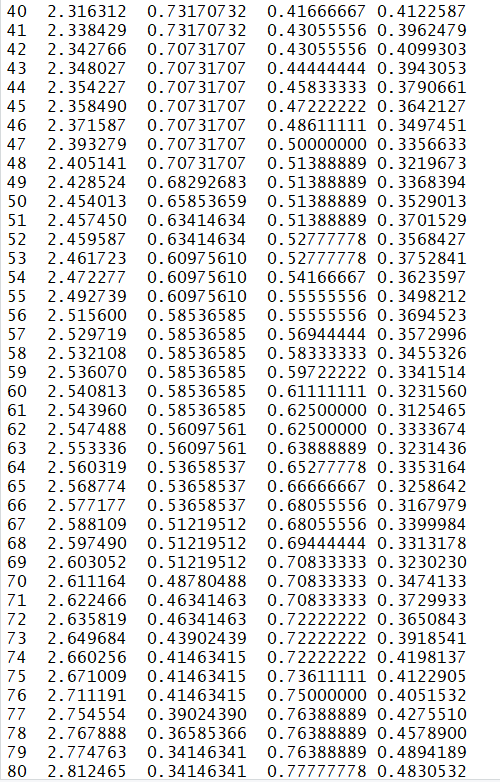
7. Develop a diagnostic test based on log(ndka). 5 points

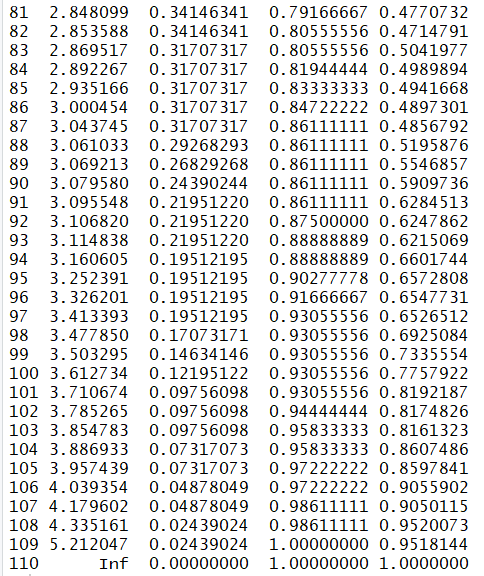


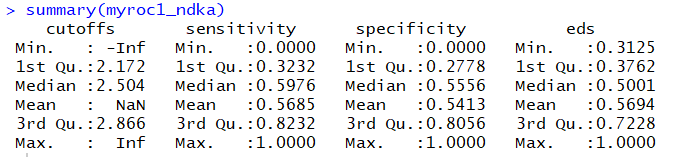






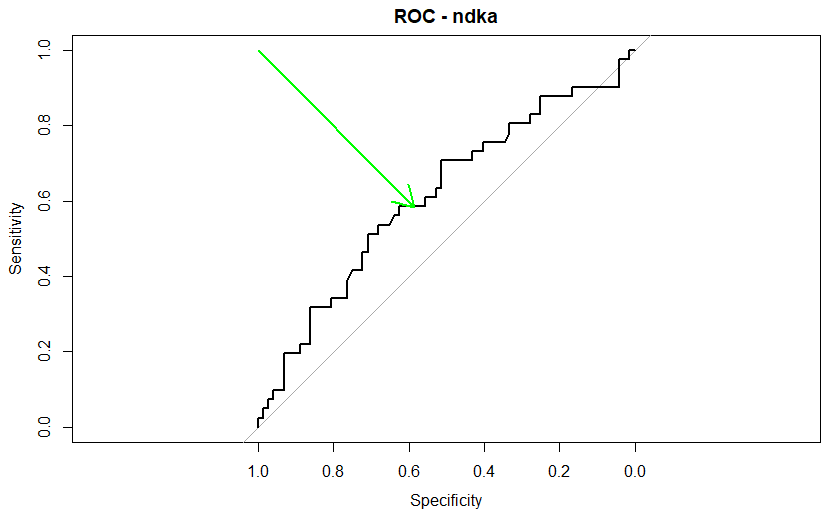




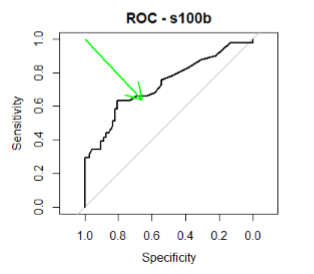


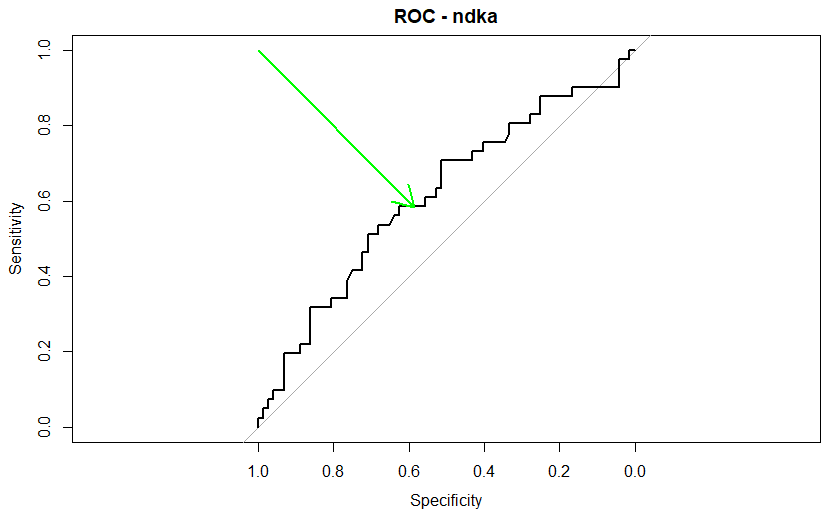






8. Compare the biomarkers s100b and log(ndka). 5 points





S100b is a better biomarker for a diagnostic test than ln ndka is.

**Code:**

library(mlbench)

require(foreign)

require(ggplot2)

require(MASS)

library(reshape2)

require(reshape2)

require(nnet)

library(caret)

install.packages("pROC")

library(pROC)

data(aSAH)

head(aSAH)

summary(aSAH)

data(aSAH)

table(aSAH$outcome, aSAH$gos6)

source("http://pcwww.liv.ac.uk/~william/R/crosstab.r")

crosstab (aSAH, row.vars = "outcome", col.vars = "gos6", type = "f")

table(aSAH$outcome, aSAH$wfns)

crosstab (aSAH, row.vars = "outcome", col.vars = "wfns", type = "f")

crosstab (aSAH, row.vars = "outcome", col.vars = "wfns", type = "r")

summary(aSAH$ndka)

aSAH$ln\_ndka = log (aSAH$ndka)

dim(aSAH)

head(aSAH)

good = subset (aSAH, aSAH$outcome =="Good")

dim(good)

head(good)

poor = subset (aSAH, aSAH$outcome =="Poor")

head(poor)

dim(poor)

summary(good$ln\_ndka)

summary(poor$ln\_ndka)

plot(density(good$ln\_ndka), col = "red", lwd = 2, xlab = 'ln\_ndka protein values', ylab = 'Density', main = "Nonparametric Density Curves", sub = "aSAH data", xlim = c(0, 6.5))

rug (good$ln\_ndka, side = 1, col = "red" )

lines(density(poor$ln\_ndka), col = "blue", lwd = 2)

rug (poor$ln\_ndka, side = 3, col = "blue")

myroc\_ndka = roc(aSAH$outcome, aSAH$ln\_ndka, levels = c("Good", "Poor"), direction = "<")

names(myroc\_ndka)

myroc1\_ndka = data.frame(cutoffs = myroc\_ndka$thresholds, sensitivity = myroc\_ndka$sensitivities, specificity = myroc\_ndka$specificities)

myroc1\_ndka

plot(myroc\_ndka, main = "ROC - ndka")

ci.auc(myroc\_ndka)

auc\_ndka = (0.5012+0.7227)/2

auc\_ndka

myroc1\_ndka$eds = (myroc1\_ndka$sensitivity - 1)^2 + (myroc1\_ndka$specificity-1)^2

head(myroc1\_ndka)

dim(myroc1\_ndka)

myroc1\_ndka

summary(myroc1\_ndka)

min(myroc1\_ndka$eds)

arrows(1,1,0.58536585, 0.58333333, col = "green", lwd = 2)