simpop\_with\_impaired.R

jonathanunderwood

Tue Apr 4 22:19:22 2017

# A comparison of two methods of diagnosing HIV-associated cognitive impairment
# with a new, multivariate approach inspired by the work of Huizenga et al
#
# Jonathan Underwood (v1.0 December 2016)
#
# For research purposes only
# If this is used in your research please cite our paper - thanks

require(MASS)
require(psych)
sigma <-matrix(c(1,0.210754443741029,0.145349600511535,0.528406352623827,0.367212674407502,0.377429548420058,
 0.210754443741029,1,0.601043770264608,0.233912234063957,0.156447757547805,0.27289232693919,
 0.145349600511535,0.601043770264608,1,0.0580374512388783,0.166080151414515,0.357473226464959,
 0.528406352623827,0.233912234063957,0.0580374512388783,1,0.210596654179787,0.222291349676154,
 0.367212674407502,0.156447757547805,0.166080151414515,0.210596654179787,1,0.100172102842648,
 0.377429548420058,0.27289232693919,0.357473226464959,0.222291349676154,0.100172102842648,1), nrow = 6)
syn\_n=100000
mean=50
mean2=30
sd=10
sd2=10

sample\_n <- 290 # study n
reps <- 10000 #replicates
covar <- sigma \* sd^2
means <- rep(50,6)
alpha <- 0.1 # one sided 0.05
p <- nrow(sigma)
critical <- -3.249432
# -sqrt((qf(p = 1- alpha, df1 = p, df2 = sample\_n-p) \* (sample\_n+1) \* p\*(sample\_n-1))/((sample\_n-p)\* sample\_n))
# -sqrt(qbeta(1-alpha,p/2, (sample\_n-p-1)/2 ) \* ((sample\_n-1)^2)/sample\_n)

results\_HAND <- NULL
results\_GDS <-NULL
results\_mahal <- NULL
results\_HAND.kappa <- NULL
results\_GDS.kappa <- NULL
results\_mahal.kappa <- NULL
results\_HAND.sens <- NULL
results\_HAND.spec <- NULL
results\_HAND.acc <- NULL
results\_HAND.ppv <- NULL
results\_HAND.npv <- NULL
results\_GDS.sens <- NULL
results\_GDS.spec <- NULL
results\_GDS.acc <- NULL
results\_GDS.ppv <- NULL
results\_GDS.npv <- NULL
results\_mahal.sens <- NULL
results\_mahal.spec <- NULL
results\_mahal.acc <- NULL
results\_mahal.ppv <- NULL
results\_mahal.npv <- NULL
results\_all <- NULL
results\_all.sens <- NULL
results\_all.spec <- NULL
results\_all.acc <- NULL
results\_all.ppv <- NULL
results\_all.npv <- NULL

for(a in seq(5,40,5)){
 print(paste("Now simulating ", reps, " repititons of ", sample\_n, " patients with a population prevalence of impairment of ", a, "%", sep = ""))
 syn\_n\*(a/100) -> n2
 syn\_n-n2 ->n
 data.frame((mvrnorm(n, mu = rep(0,nrow(sigma)), Sigma = sigma,empirical = TRUE) \* sd) +mean) -> synthetic\_pop2
 rep(1,n) -> synthetic\_pop2$label

 data.frame((mvrnorm(n2, mu = rep(0,nrow(sigma)), Sigma = sigma,empirical = TRUE) \* sd2) +mean2) ->synthetic\_pop2.imp
 rep(2,n2) -> synthetic\_pop2.imp$label

 rbind(synthetic\_pop2, synthetic\_pop2.imp) ->synthetic\_pop2

 for(i in 1:reps) {
 synthetic\_pop2[sample(nrow(synthetic\_pop2),size = sample\_n, replace = F),] -> test
 apply(test[1:6],1,function(x) sum(x<40)>1) -> HAND # generates frascati
 apply(test[1:6],1, function(i) sign(mean(i)-mean)\*sqrt(mahalanobis(i,center = means, cov = covar))) -> mahal\_dist # generates signed mahal distance

 apply(test[1:6],2,function(x)
 ifelse(x>19,
 ifelse(x>24,
 ifelse(x>29,
 ifelse(x>34,
 ifelse(x>40,0,1),2),3),4),5)
 ) -> test2
 apply(test2,1,function(x) mean(x)>=0.5) ->GDS # generates global deficit score

 rbind(results\_HAND, sum(HAND)/sample\_n \* 100) -> results\_HAND
 rbind(results\_GDS, sum(GDS)/sample\_n \* 100) -> results\_GDS
 rbind(results\_mahal, sum(mahal\_dist<critical)/sample\_n \* 100) -> results\_mahal
 rbind(results\_all, sum(HAND==1 & GDS ==1 & mahal\_dist<critical)/sample\_n \* 100) -> results\_all
 rbind(results\_HAND.kappa, cohen.kappa(table( as.matrix(HAND)==1, test$label==2))$kappa) -> results\_HAND.kappa
 rbind(results\_GDS.kappa, cohen.kappa(table( as.matrix(GDS)==1, test$label==2))$kappa) -> results\_GDS.kappa
 rbind(results\_mahal.kappa, cohen.kappa(table( as.matrix(mahal\_dist<critical)==1, test$label==2))$kappa) -> results\_mahal.kappa

 rbind(results\_HAND.sens, sum((HAND)==1 & test$label==2)/sum(test$label==2)) ->results\_HAND.sens
 rbind(results\_HAND.spec, sum((HAND)==0 & test$label==1)/sum(test$label==1)) ->results\_HAND.spec
 rbind(results\_HAND.acc, (sum(HAND==1 & test$label==2) + sum(HAND==0 & test$label==1))/sample\_n) -> results\_HAND.acc
 rbind(results\_HAND.ppv, sum(HAND==1 & test$label==2)/sum(HAND==1)) -> results\_HAND.ppv
 rbind(results\_HAND.npv, sum(HAND==0 & test$label==1)/sum(HAND==0)) -> results\_HAND.npv

 rbind(results\_GDS.sens, sum((GDS)==1 & test$label==2)/sum(test$label==2)) ->results\_GDS.sens
 rbind(results\_GDS.spec, sum((GDS)==0 & test$label==1)/sum(test$label==1)) ->results\_GDS.spec
 rbind(results\_GDS.acc, (sum(GDS==1 & test$label==2) + sum(GDS==0 & test$label==1))/sample\_n) -> results\_GDS.acc
 rbind(results\_GDS.ppv, sum(GDS==1 & test$label==2)/sum(GDS==1)) -> results\_GDS.ppv
 rbind(results\_GDS.npv, sum(GDS==0 & test$label==1)/sum(GDS==0)) -> results\_GDS.npv

 rbind(results\_mahal.sens, sum(mahal\_dist<critical & test$label==2)/sum(test$label==2)) ->results\_mahal.sens
 rbind(results\_mahal.spec, sum(mahal\_dist>critical & test$label==1)/sum(test$label==1)) ->results\_mahal.spec
 rbind(results\_mahal.acc, (sum(mahal\_dist<critical & test$label==2) + sum(mahal\_dist>critical & test$label==1))/sample\_n) -> results\_mahal.acc
 rbind(results\_mahal.ppv, sum(mahal\_dist<critical & test$label==2)/sum(mahal\_dist<critical)) -> results\_mahal.ppv
 rbind(results\_mahal.npv, sum(mahal\_dist>critical & test$label==1)/sum(mahal\_dist>critical)) -> results\_mahal.npv

 rbind(results\_all.sens, sum((HAND==1 & GDS ==1 & mahal\_dist<critical) & test$label==2)/sum(test$label==2)) ->results\_all.sens
 rbind(results\_all.spec, sum((HAND==0 | GDS ==0 | mahal\_dist>critical) & test$label==1)/sum(test$label==1)) ->results\_all.spec
 rbind(results\_all.acc, (sum((HAND==1 & GDS ==1 & mahal\_dist<critical) & test$label==2) + sum((HAND==0 | GDS ==0 | mahal\_dist>critical) & test$label==1))/sample\_n) -> results\_all.acc
 rbind(results\_all.ppv, sum((HAND==1 & GDS ==1 & mahal\_dist<critical) & test$label==2)/sum((HAND==1 & GDS ==1 & mahal\_dist<critical))) -> results\_all.ppv
 rbind(results\_all.npv, sum((HAND==0 | GDS ==0 | mahal\_dist>critical) & test$label==1)/sum((HAND==0 | GDS ==0 | mahal\_dist>critical))) -> results\_all.npv
 }
}

for(i in 0:7){
 i\*reps ->i1
 (i+1)\*reps ->i2
 print(paste("Population prevalence of impairment of ", (i+1)\*5, "%", sep = ""))
 print(apply(data.frame(results\_HAND[i1:i2],results\_HAND.sens[i1:i2],results\_HAND.spec[i1:i2], results\_HAND.acc[i1:i2], results\_HAND.ppv[i1:i2], results\_HAND.npv[i1:i2]), 2, function(x) {
 (c(mean(x), sort(x)[.05\*reps], sort(x)[.95\*reps]))
 }))

 print(apply(data.frame(results\_GDS[i1:i2],results\_GDS.sens[i1:i2],results\_GDS.spec[i1:i2], results\_GDS.acc[i1:i2], results\_GDS.ppv[i1:i2], results\_GDS.npv[i1:i2]), 2, function(x) {
 (c(mean(x), sort(x)[.05\*reps], sort(x)[.95\*reps]))
 }))

 print(apply(data.frame(results\_mahal[i1:i2],results\_mahal.sens[i1:i2],results\_mahal.spec[i1:i2], results\_mahal.acc[i1:i2], results\_mahal.ppv[i1:i2], results\_mahal.npv[i1:i2]), 2, function(x) {
 (c(mean(x), sort(x)[.05\*reps], sort(x)[.95\*reps]))
 }))

 print(apply(data.frame(results\_all[i1:i2],results\_all.sens[i1:i2],results\_all.spec[i1:i2], results\_all.acc[i1:i2], results\_all.ppv[i1:i2], results\_all.npv[i1:i2]), 2, function(x) {
 (c(mean(x), sort(x)[.05\*reps], sort(x)[.95\*reps]))
 }))
}