**Homework** (8 points, please pay attention to all the words in this orange box)

**Ex 22-1:** If the probability of getting an admission is  for the ith subject, which might be related to his/her gpa and gre scores. Considering the model



 and . Suppose the sample sizes are 30, 230, 430, and 630, respectively. Number of replications = 1000, and seed numbers from 1 to 1000, respectively. Please use simulations to evaluate the bias of maximum likelihood estimates (MLE) of the regression coefficients, and make box plots to see their distributions.

Hint:

1. Generate gpa and gre scores for n subjects.
2. Given gpa and gre, claculate .
3. sample(c(0,1),1,c(1-,),replace=F)
4. Use the Newton-Raphson method to find MLEs.

**Please note: Using the R built-in function “**glm(admit~gpa+gre,family=binomial)**” to answer this homework will be scored as 0, although you may use it to check your own answers.**

Bias of MLE

> bias

[,1] [,2] [,3]

n=30 -1.25114905 0.185444755 1.203350e-03

n=230 -0.13083763 0.022874614 1.121271e-04

n=430 -0.09723871 0.015521581 9.303015e-05

n=630 -0.03847831 0.007797278 2.778161e-05

> meanMLE

[,1] [,2] [,3]

n=30 -7.251149 1.185445 0.006203350

n=230 -6.130838 1.022875 0.005112127

n=430 -6.097239 1.015522 0.005093030

n=630 -6.038478 1.007797 0.005027782

> meanMLEglm

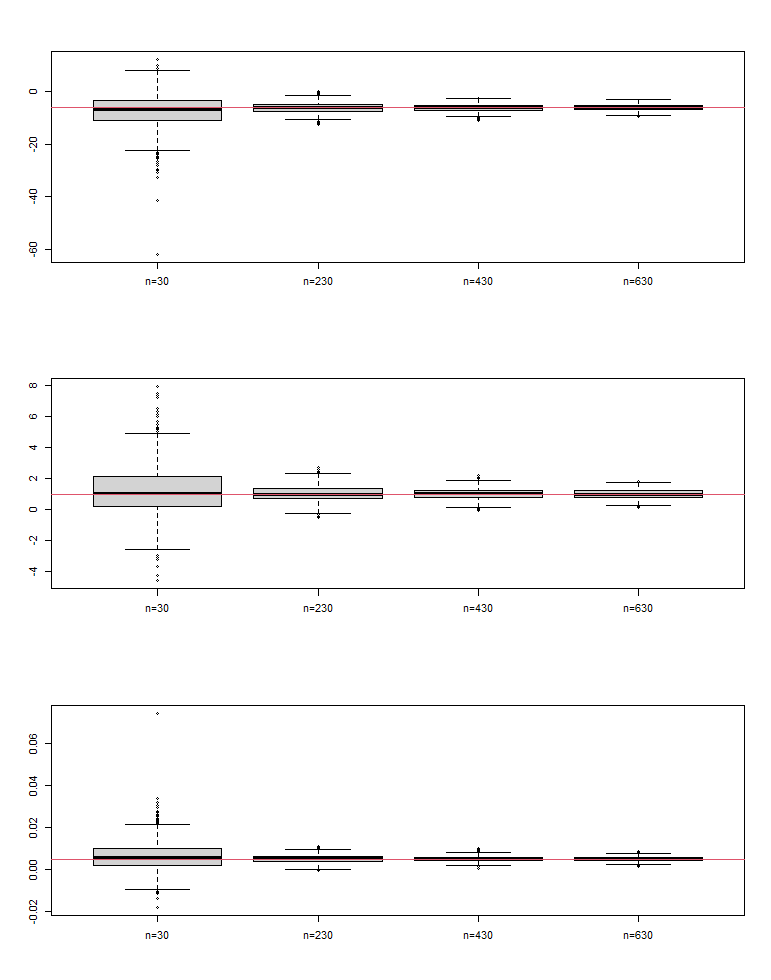
[,1] [,2] [,3]

n=30 -7.251149 1.185445 0.006203350

n=230 -6.130838 1.022875 0.005112127

n=430 -6.097239 1.015522 0.005093030

n=630 -6.038478 1.007797 0.005027782



Code:

# Ex 22-1

betaco <- c(-6,1.0,0.005)

n <- seq(30,630,200)

no.rep <- 1000

tol = 1e-9

n.max = 100

meanMLE <- matrix(NA,length(n),3)

meanMLEglm <- matrix(NA,length(n),3)

allMLE <- array(NA,c(no.rep,3,length(n)))

for(n.loop in 1: length(n)){

MLE <- matrix(NA,no.rep,3)

MLEglm <- matrix(NA,no.rep,3)

for(i in 1:no.rep){

set.seed(i)

gpa <- rnorm(n[n.loop],3.1,0.3)

gre <- rnorm(n[n.loop],580,80)

x.beta <- betaco[1]+betaco[2]\*gpa+betaco[3]\*gre

pi.admit <- exp(x.beta)/(1+exp(x.beta))

Y <- c()

for(j in 1:n[n.loop]){

Y[j] <- sample(c(0,1),1,c(1-pi.admit[j],pi.admit[j]),replace=F)

}

X <- cbind(rep(1,length(Y)),gpa,gre)

betacoef <- c(0,0,0)

pi1 <- exp(X %\*% betacoef)/(1+exp(X %\*% betacoef))

gradient <- t(X) %\*% (Y-pi1)

hessian <- -t(X) %\*% diag(c(pi1\*(1-pi1)),length(Y)) %\*% X

n.iter <- 0

while ((max(abs(gradient)) > tol) & (n.iter < n.max)) {

betacoef <- betacoef - solve(hessian) %\*% gradient

pi1 <- exp(X %\*% betacoef)/(1+exp(X %\*% betacoef))

gradient <- t(X) %\*% (Y-pi1)

hessian <- -t(X) %\*% diag(c(pi1\*(1-pi1)),length(Y)) %\*% X

n.iter <- n.iter + 1

}

if (n.iter == n.max) {

cat('newton failed to converge\n')

} else {

MLE[i,] <- betacoef

}

MLEglm[i,] <- glm(Y~gpa+gre,family=binomial)$coef

}

meanMLE[n.loop,] <- colSums(MLE)/no.rep

meanMLEglm[n.loop,] <- colSums(MLEglm)/no.rep

allMLE[,,n.loop] <- MLE

}

bias <- rbind(meanMLE[1,]-betaco, meanMLE[2,]-betaco, meanMLE[3,]-betaco, meanMLE[4,]-betaco)

rownames(meanMLE) <- c("n=30","n=230","n=430","n=630")

rownames(meanMLEglm) <- c("n=30","n=230","n=430","n=630")

rownames(bias) <- c("n=30","n=230","n=430","n=630")

meanMLE

meanMLEglm

bias

par(mfrow = c(3,1))

x <- seq(1,4,1)

boxplot(allMLE[,1,1],allMLE[,1,2],allMLE[,1,3],allMLE[,1,4], xaxt = 'n')

abline(h=betaco[1],col=2)

axis(1, at = x, labels = c("n=30","n=230","n=430","n=630"))

boxplot(allMLE[,2,1],allMLE[,2,2],allMLE[,2,3],allMLE[,2,4], xaxt = 'n')

abline(h=betaco[2],col=2)

axis(1, at = x, labels = c("n=30","n=230","n=430","n=630"))

boxplot(allMLE[,3,1],allMLE[,3,2],allMLE[,3,3],allMLE[,3,4], xaxt = 'n')

abline(h=betaco[3],col=2)

axis(1, at = x, labels = c("n=30","n=230","n=430","n=630"))