Diabetes Bayesian

Rory Quinlan

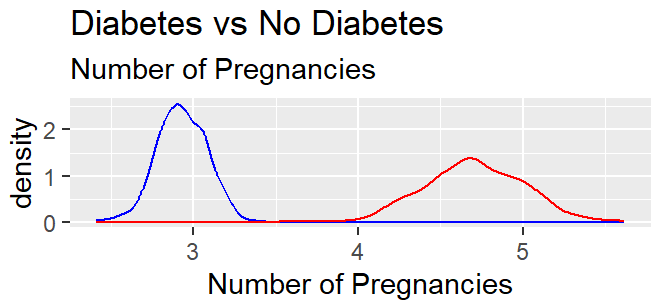
```
library(MASS)
library(ggplot2)
library(gridExtra)
table <- read.table("C:\\Users\\roryq\\Downloads\\azdiabetes.dat", header = T)
dia <- table[table$diabetes == "Yes", 1: dim(table)[2]-1]
ndia <- table[table$diabetes == "No",1:dim(table)[2]-1]
colnames(dia)
```

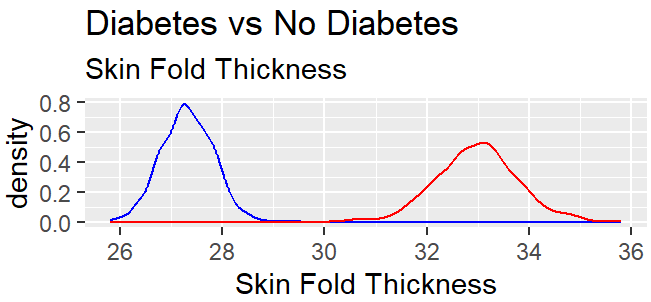
```
## [1] "npreg" "glu"   "bp"    "skin"  "bmi"   "ped"   "age"
```

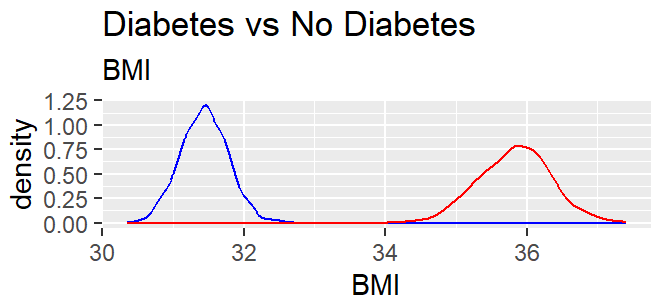
```
# MC estimates for diabetes parameters
#prior parameters
ybar <- apply(dia, 2, mean)
Sigma <- cov(dia)
n <- dim(dia)[1]
mu0 <- ybar
nu0 <- 9
L0 <- S0 <- Sigma
theta_ad<-sigma_ad <- NULL
set.seed(11000)
for(s in 1:1000){

#Update theta
 Ln <- solve(solve(L0) + n * solve(Sigma))
 mun <- Ln %*% (solve(L0) %*% mu0 + n * solve(Sigma) %*% ybar)
 theta <- mvrnorm(1, mun, Ln)

#update Simga
 Sn <- S0 + (t(dia) - c(theta)) %*% t( t(dia) - c(theta))
 Sigma <- solve( rWishart(1, nu0 + n, solve(Sn))[,,1])

# save results

 theta_ad <- rbind(theta_ad, theta) ; sigma_ad <- rbind(sigma_ad, c(Sigma))
}
```

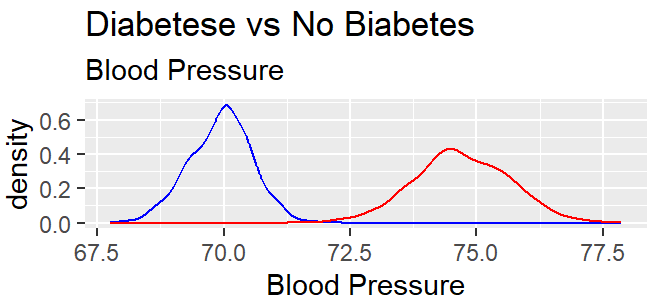
```r
# MC estimates for non-diabetes parameters
#prior parameters
ybar <- apply(ndia, 2, mean)
Sigma <- cov(ndia)
n <- dim(ndia)[1]
mu0 <- ybar
nu0 <- 9
L0 <- S0 <- Sigma
theta_an<-sigma_an <- NULL
set.seed(10000)
for(s in 1:1000){

#Update theta
 Ln <- solve(solve(L0) + n * solve(Sigma))
 mun <- Ln %*% (solve(L0) %*% mu0 + n * solve(Sigma) %*% ybar)
 theta <- mvrnorm(1, mun, Ln)

 #update Simga
 Sn <- S0 + (t(ndia) - c(theta)) %*% t( t(ndia) - c(theta))
 Sigma <- solve( rWishart(1, nu0 + n, solve(Sn))[,,1])

 # save results

 theta_an <- rbind(theta_an, theta) ; sigma_an <- rbind(sigma_an, c(Sigma))
}
```

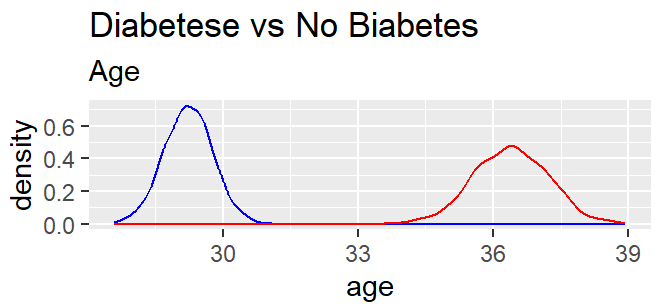
```
# Create data frame from MC
theta_an <- data.frame(theta_an)
sigma_an <- data.frame(sigma_an)
theta_ad <- data.frame(theta_ad)
sigma_ad <- data.frame(sigma_ad)
theta_an$diabetes <- sigma_an$diabetes <- "no"
theta_ad$diabetes <- sigma_ad$diabetes <- "yes"
theta <- rbind(theta_an, theta_ad)
sigma <- rbind(sigma_an, sigma_ad)

# Plot thetas for yes and no diabetes on number of pregnancies
p1<-ggplot() + geom_density(aes(x = npreg), data = theta_an, color = "blue", ) + geom_density(ae
s(x = npreg), data = theta_ad, color = "red")+labs(x="Number of Pregnancies", title= "Diabetes v
s No Diabetes", subtitle="Number of Pregnancies")

p2<- ggplot() + geom_density(aes(x = skin), data = theta_an, color = "blue", ) + geom_density(ae
s(x = skin), data = theta_ad, color = "red")+labs(x="Skin Fold Thickness",title= "Diabetes vs No
Diabetes", subtitle="Skin Fold Thickness")

p3<- ggplot() + geom_density(aes(x = bmi), data = theta_an, color = "blue", ) + geom_density(aes
(x = bmi), data = theta_ad, color = "red")+labs(x="BMI", title= "Diabetes vs No Diabetes", subti
tle="BMI")

p4<-ggplot() + geom_density(aes(x = bp), data = theta_an, color = "blue", ) + geom_density(aes(x
= bp), data = theta_ad, color = "red")+labs(x="Blood Pressure", title= "Diabetese vs No Biabete
s", subtitle="Blood Pressure")

p5<-ggplot() + geom_density(aes(x = age), data = theta_an, color = "blue", ) + geom_density(aes
(x = age), data = theta_ad, color = "red")+labs(x="age", title= "Diabetese vs No Biabetes", subt
itle="Age")

p6<-ggplot() + geom_density(aes(x = glu), data = theta_an, color = "blue", ) + geom_density(aes
(glu), data = theta_ad, color = "red")+labs(x="Glucose Level", title= "Diabetese vs No Biabete
s", subtitle="Glucose Level")

grid.arrange(p1, p2, p3,p4,p5,p6, ncol = 2)
```

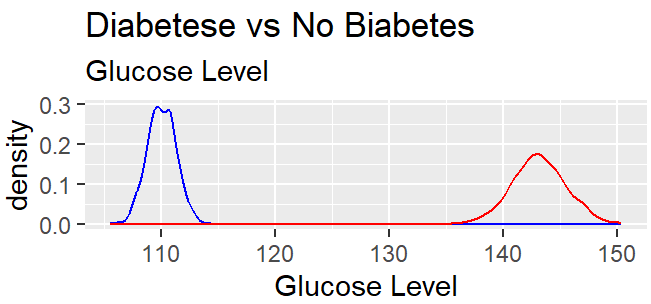




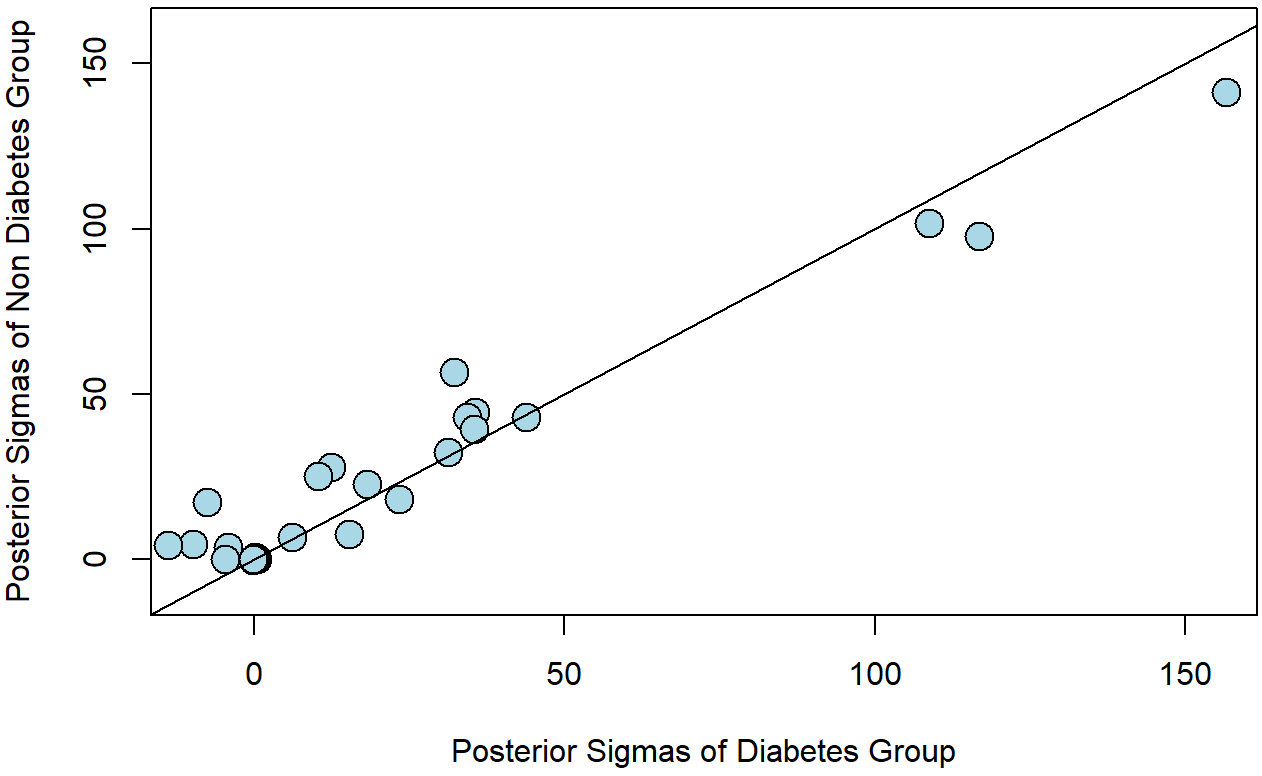









```
post.sigd <- as.numeric(apply(sigma_ad[,1:49], 2, mean))
post.sign <- as.numeric(apply(sigma_an[,1:49], 2, mean))
plot(x = post.sigd, y = post.sign, xlim = c(-10, 155), pch=21,ylim=c(-10,160),col="black", bg="l
ightblue", cex=2,xlab= "Posterior Sigmas of Diabetes Group", ylab="Posterior Sigmas of Non Diabe
tes Group")
abline(coef = c(0,1))
```


Posterior Sigmas of Non Diabetes Group
0
50
100
150
0
50
100
150
Posterior Sigmas of Diabetes Group