Assignment 3 - Question 2

Question 2

Taken from Albert (2007). Normal linear regression.

Dobson (2001) describes a birthweight regression study. One is interested in predicting a baby's birthweight (in grams) based on the gestational age (in weeks) and the gender of the baby. The data is presented in the file *birthweight.txt*. In the standard linear regression model, we assume that

$$\text{Birthweight (i)} = \beta_0 + \beta_1\, AGE\,(i) + \beta_2\, GENDER\,(i) + e(i)$$

where the *e(i)'s* are independent and Normally distributed random variables with mean **0** and variance σ^2.

(a) Use the R function lm to fit this model by Least-Squares. From the output, assess if the effects AGE and GENDER are significant, and if they are significant, describe the effects of each covariate on birthweight.

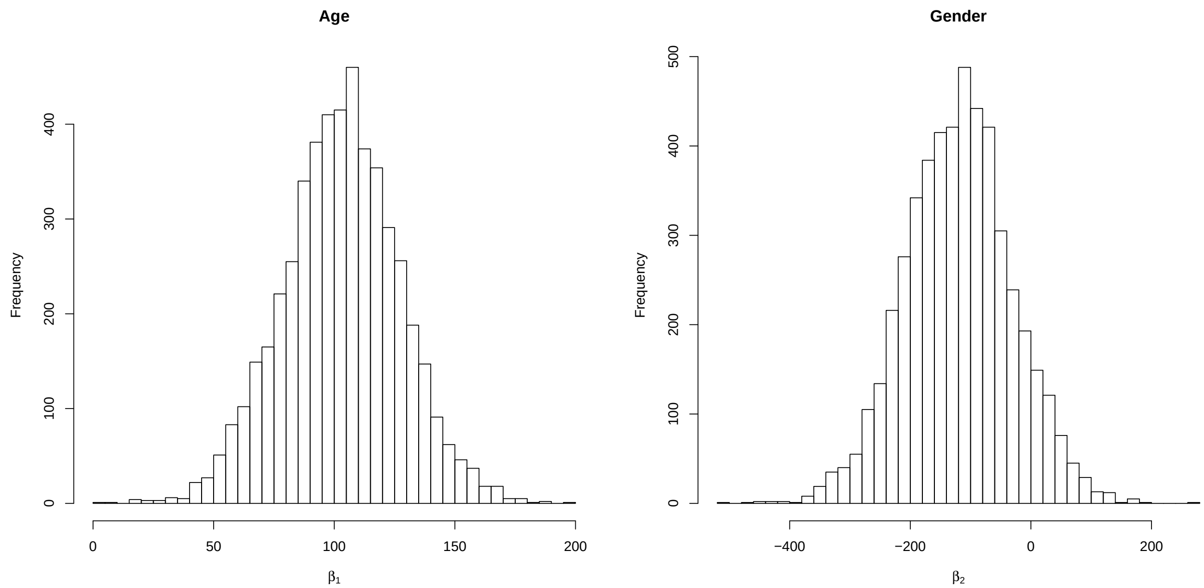
```
fit <- lm(weight ~ age + gender, data = birthweight, model = TRUE, x=TRUE,y=TRUE)
summary(fit)
##
## Call:
## lm(formula = weight ~ age + gender, data = birthweight, model = TRUE,
##     x = TRUE, y = TRUE)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -258.65 -155.48  -36.17  164.00  463.62
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -924.94     896.09  -1.032  0.31373
## age            103.02      23.32   4.417  0.00024 ***
## genderFemale  -121.25      84.97  -1.427  0.16827
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 208.1 on 21 degrees of freedom
## Multiple R-squared:  0.5032, Adjusted R-squared:  0.4558
## F-statistic: 10.63 on 2 and 21 DF,  p-value: 0.0006461
```

From the output of the linear model we can see that age is the only covariate that significantly effects birthweight ($\alpha = 0.05$). It has a positive effect on birthweight - i.e. as age increases, so does birthweight. Gender - which is not found to have a significant impact on birthweight - negatively effects birthweight if the child is female; it effects bw positively if the child is male. The value of the intercept is given as ≈ -924.94.

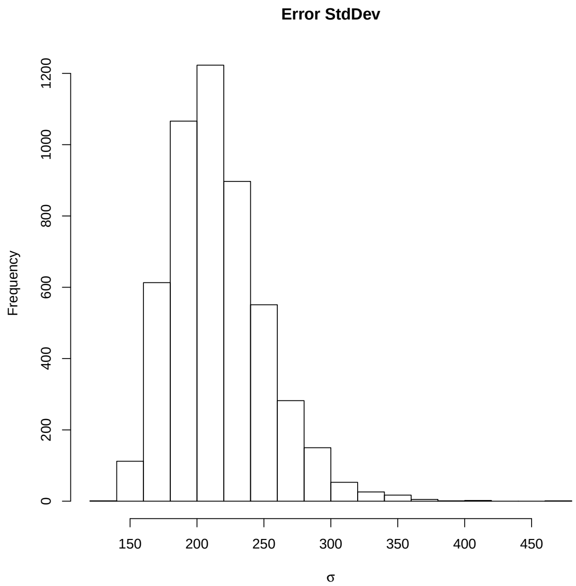
(b) Suppose a uniform prior is placed on the regression parameters $\beta = (\beta_0, \beta_1, \beta_2)$. Use the LearnBayes package in R (blinreg command) to simulate an MC sample of 5000 draws from the joint posterior distribution of (β, σ^2). From the simulated sample, compute posterior means and standard deviations of β_1 and β_2. Compare the results with the LS estimates obtained in (a).

```r
theta.sample <- blinreg(fit$y,fit$x,5000)
sumresids <- sum(fit$residual^2)
shape <- fit$df.residual/2
rate <- sumresids/2
sigma2 <- rigamma(1,shape,rate)
MSE <-  sum(fit$residuals^2)/fit$df.residual
vbeta <- vcov(fit)/MSE
beta <- rmnorm(1,mean=fit$coef,varcov=vbeta*sigma2)
```

```r
par(mfrow=c(1,2))
hist(theta.sample$beta[,2],main="Age",xlab=expression(beta[1]), 30)
hist(theta.sample$beta[,3],main="Gender", xlab=expression(beta[2]),30)
```




```r
hist(theta.sample$sigma,main="Error StdDev", xlab=expression(sigma))
# CIs for the Beta coefficients
apply(theta.sample$beta,2,quantile,c(.05,.5,.95))
##     X(Intercept)      Xage XgenderFemale
## 5%    -2401.7902  62.29281    -263.43666
## 50%    -936.1221 103.29552    -118.21295
## 95%     629.9070 141.24848      27.05655
# CIs for sigma
quantile(theta.sample$sigma,c(.05,.5,.95))
##       5%      50%      95%
## 167.1913 211.5657 280.4317
```

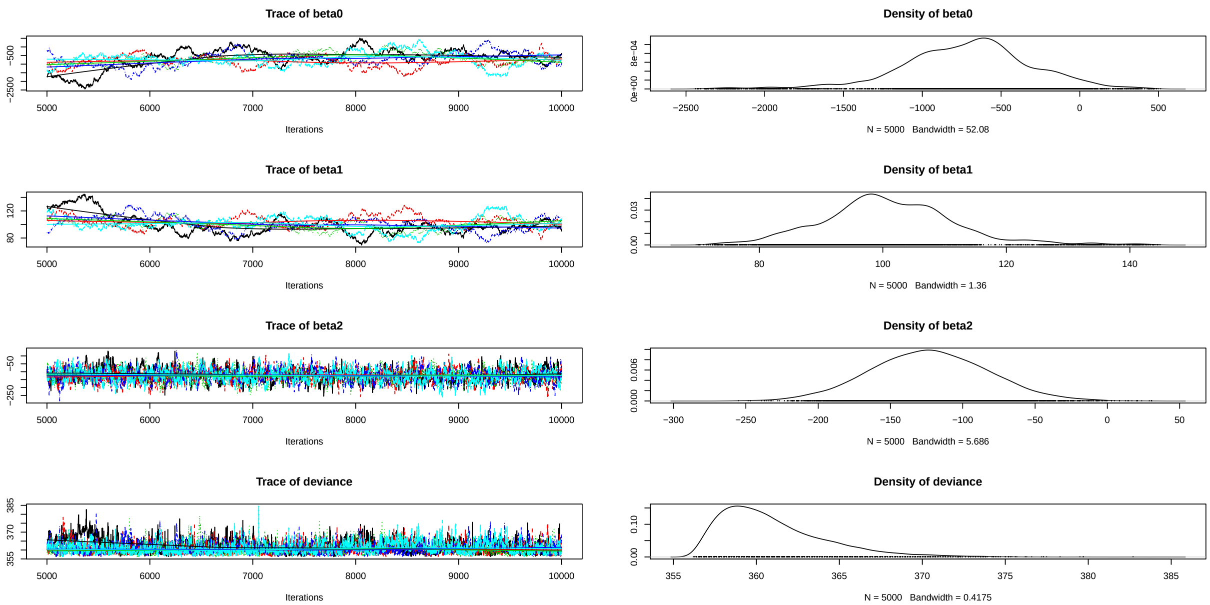


Here we are given values for the Intercept, Age coefficient and Gender coefficient of approximately $-907.2544, 102.49836$ and -119.5783 respectively, which are pretty close to the values obtained using the method in part (a).

(c) Repeat part (b) using WinBUGS. Compare the results with the LS estimates obtained in (a) and MC sampling in part (b).

```
require(R2OpenBUGS)
## Loading required package: R2OpenBUGS
require(coda)
## Loading required package: coda

#define the model
birthweightmodel <- function(){
  for (i in 1:n) {
    mu[i] <- beta0 + beta1*x1[i] + beta2*x2[i]
    yi[i] ~ dnorm(mu[i],tau)
  }
  beta0 ~  dflat()
  sigma ~ dunif(0.01, 100)
  tau <- 1/(sigma*sigma)
  beta1 ~  dflat()
  beta2 ~  dflat()
}
# write the model code out to a file
write.model(birthweightmodel, "birthweightmodel.txt")
model.file1 = paste(getwd(),"birthweightmodel.txt", sep="/")

#prepare the data for input into OpenBUGS
x1 <- fit$model[,2]
x2 <- as.integer(unclass(fit$model[,3]))
yi <- fit$model[,1]
n <- 24
```

```r
data <- list("x1","x2","yi","n")

#initialization of variables
inits <- function(){
  list(beta0=0, beta1=1,beta2=1,sigma=1)
}

WINE="/opt/local/bin/wine"
WINEPATH="/opt/local/bin/winepath"
OpenBUGS.pgm=paste0("/Users/benjamin/Applications/wine/",
                    "drive_c/ProgramFiles/OpenBUGS/OpenBUGS323/OpenBUGS.exe")

#these are the parameters to save
parameters = c("beta0", "beta1","beta2")

#run the model
birthweight.sim <- bugs(data,
                    inits,
                    model.file = model.file1,
                    parameters=parameters,
                    n.chains = 5,
                    n.iter = 10000,
                    OpenBUGS.pgm=OpenBUGS.pgm,
                    WINE=WINE,
                    WINEPATH=WINEPATH,
                    useWINE=T,
                    codaPkg = T,
                    working.directory = getwd(),
                    debug = F)

samples <- read.openbugs()
## Abstracting beta0 ... 5000 valid values
## Abstracting beta1 ... 5000 valid values
## Abstracting beta2 ... 5000 valid values
## Abstracting deviance ... 5000 valid values
## Abstracting beta0 ... 5000 valid values
## Abstracting beta1 ... 5000 valid values
## Abstracting beta2 ... 5000 valid values
## Abstracting deviance ... 5000 valid values
## Abstracting beta0 ... 5000 valid values
## Abstracting beta1 ... 5000 valid values
## Abstracting beta2 ... 5000 valid values
## Abstracting deviance ... 5000 valid values
## Abstracting beta0 ... 5000 valid values
## Abstracting beta1 ... 5000 valid values
## Abstracting beta2 ... 5000 valid values
## Abstracting deviance ... 5000 valid values
## Abstracting beta0 ... 5000 valid values
## Abstracting beta1 ... 5000 valid values
## Abstracting beta2 ... 5000 valid values
## Abstracting deviance ... 5000 valid values
plot(samples)
```




```
summary(samples)
##
## Iterations = 5001:10000
## Thinning interval = 1
## Number of chains = 5
## Sample size per chain = 5000
##
## 1. Empirical mean and standard deviation for each variable,
##    plus standard error of the mean:
##
##              Mean      SD Naive SE Time-series SE
## beta0    -718.8 418.164  2.64470        75.9269
## beta1     100.9  10.816  0.06841         1.9390
## beta2    -123.1  40.655  0.25712         1.1250
## deviance  361.0   3.191  0.02018         0.1361
##
## 2. Quantiles for each variable:
##
##              2.5%     25%    50%     75%  97.5%
## beta0    -1660.00 -966.12 -692.7 -467.17  61.33
## beta1       80.98   94.27  100.3  107.30 125.40
## beta2     -202.60 -150.80 -123.5  -95.72 -42.15
## deviance   356.80  358.60  360.3  362.60 369.00
```

WINBUGS reports values for the Intercept, Age coefficient and Gender coefficient of approximately $-718.8, 100.9$ and -123.1 respectively. The value given to the intercept is much lower than the values reported in parts (a) and (b), however the other coefficients are roughly the same as those calculated using the other methods.