Bayesian simple linear regression

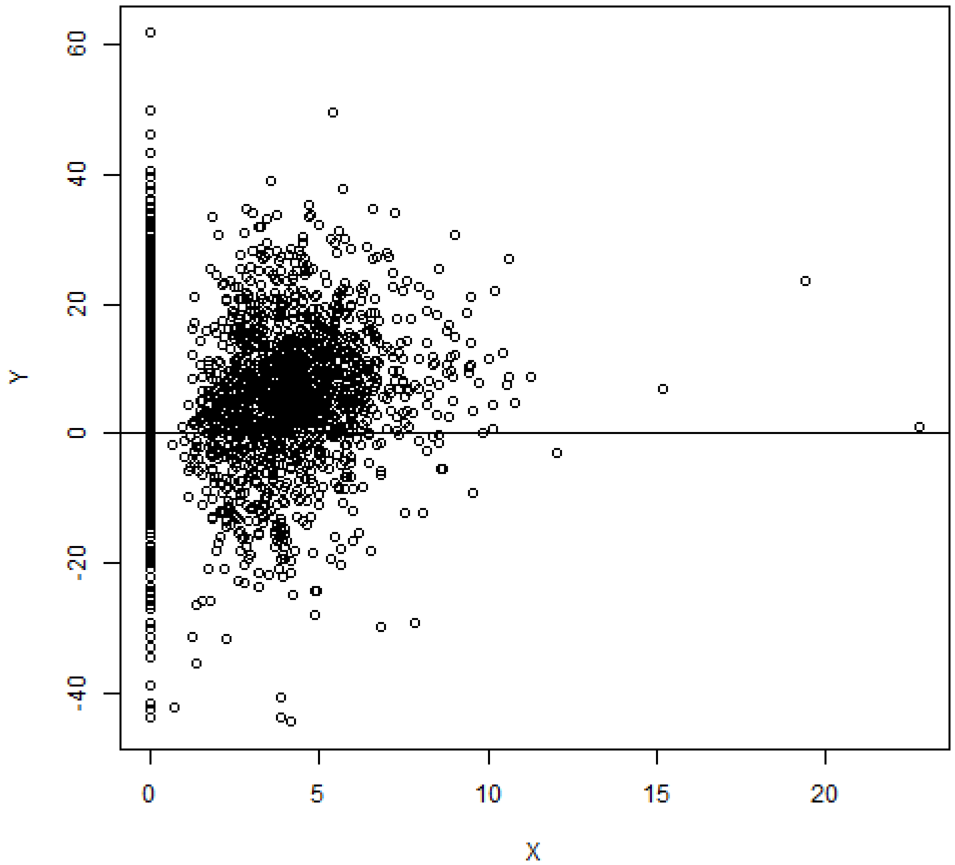
In this exercise we will regress Y_i, the percent increase in support for GOP, onto X_i, the square root of per capita manufacturing shipments. The model is

$$Y_i \sim \text{Normal}(\alpha + \beta X_i, \sigma^2).$$

We select priors

$$\alpha, \beta \sim \text{Normal}(\mu_0, \sigma_0^2) \quad \sigma^2 \sim \text{InvGamma}(a, b).$$

```
load("S:\\Documents\\www\\ABA\\code\\election_2008_2016.RData")
X    <- sqrt(X[,14])
junk <- is.na(X+Y)
Y    <- Y[!junk]
X    <- X[!junk]

n     <- length(Y)

plot(X,Y)
abline(0,0)
```



```
OLS <- lm(Y~X)
summary(OLS)
```

```
##
## Call:
## lm(formula = Y ~ X)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -51.091  -5.194  -0.176   5.705  55.024
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 6.676417   0.246885  27.043   <2e-16 ***
## X           0.003592   0.075514   0.048    0.962
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.49 on 3109 degrees of freedom
## Multiple R-squared:  7.279e-07,  Adjusted R-squared:  -0.0003209
## F-statistic: 0.002263 on 1 and 3109 DF,  p-value: 0.9621
```

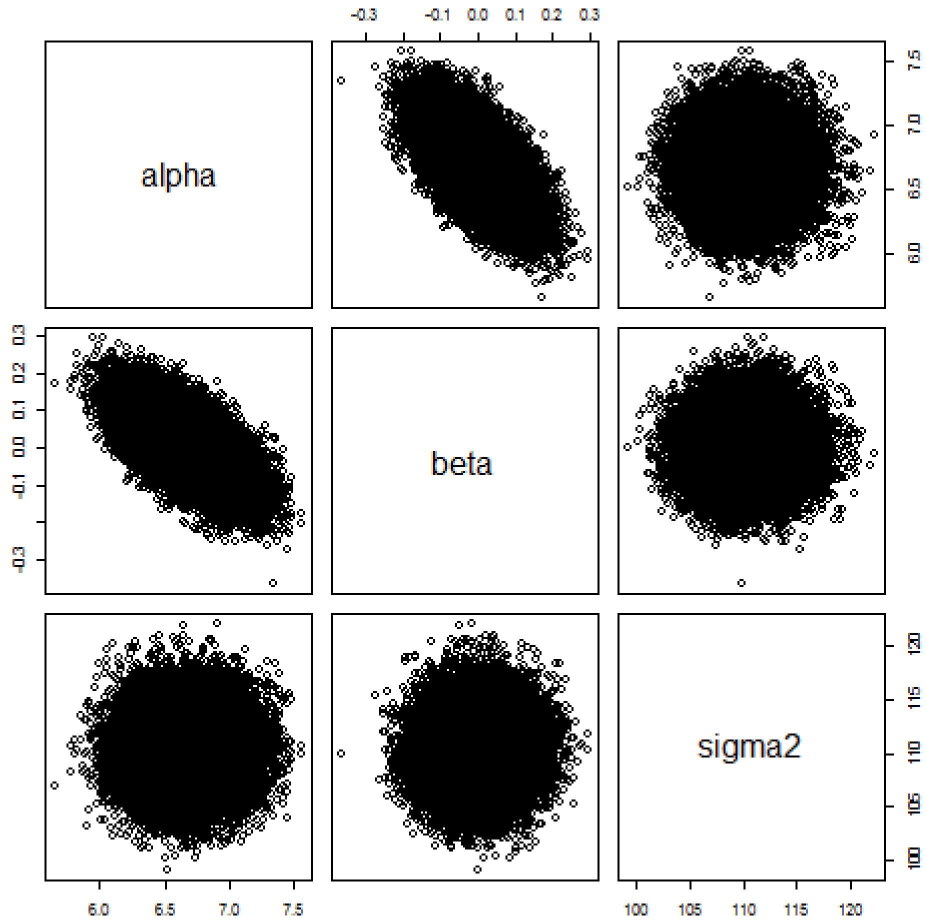
```
### Priors

  mu0 <- 0
  s20 <- 1000
  a   <- 0.01
  b   <- 0.01
```

MCMC!

```
n.iters <- 30000
keepers <- matrix(0,n.iters,3)
colnames(keepers)<-c("alpha","beta","sigma2")

# Initial values
alpha       <- OLS$coef[1]
beta        <- OLS$coef[2]
s2          <- var(OLS$residuals)
keepers[1,] <- c(alpha,beta,s2)

for(iter in 2:n.iters){

  # sample alpha

   V     <- n/s2+mu0/s20
   M     <- sum(Y-X*beta)/s2+1/s20
   alpha <- rnorm(1,M/V,1/sqrt(V))

  # sample beta

   V     <- sum(X^2)/s2+mu0/s20
   M     <- sum(X*(Y-alpha))/s2+1/s20
   beta  <- rnorm(1,M/V,1/sqrt(V))

  # sample s2|mu,Y,Z

   A  <- n/2 + a
   B  <- sum((Y-alpha-X*beta)^2)/2 + b
   s2 <- 1/rgamma(1,A,B)

  # keep track of the results
   keepers[iter,] <- c(alpha,beta,s2)

}
```

Plots of the joint posterior distribution.

```
pairs(keepers)
```



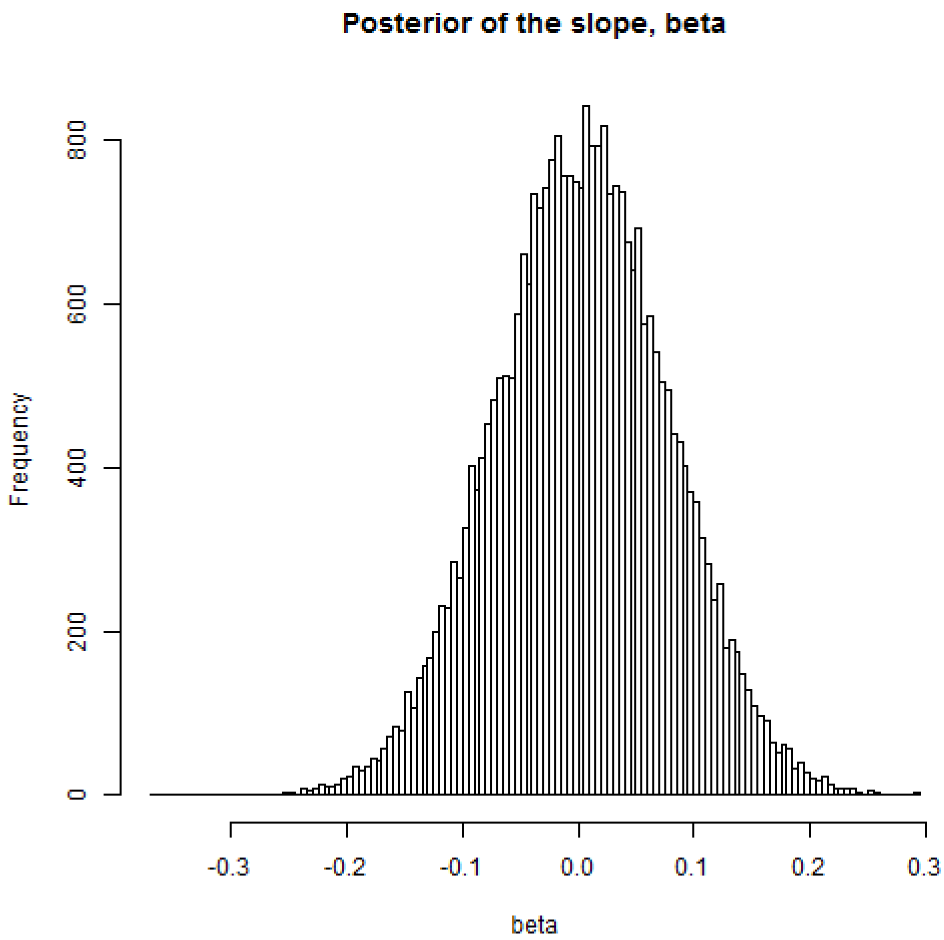
Summarize the marginal distributions in a table

```
output <- matrix(0,3,4)
rownames(output) <- c("Intercept","manufacturing","sigma2")
colnames(output) <- c("Mean","SD","Q025","Q975")

output[,1] <- apply(keepers,2,mean)
output[,2] <- apply(keepers,2,sd)
output[,3] <- apply(keepers,2,quantile,0.025)
output[,4] <- apply(keepers,2,quantile,0.975)

kable(output,digits=3)
```

	Mean	SD	Q025	Q975
Intercept	6.677	0.245	6.196	7.159
manufacturing	0.004	0.075	-0.143	0.151
sigma2	110.151	2.795	104.895	115.787

Plot the marginal posterior $f(\beta|Y)$.

```
beta <- keepers[,2]
hist(beta,main="Posterior of the slope, beta",breaks=100)
```



The results are similar to least squares; there appears to be a no (marginal) relationship between these two variables.