MA 578: HW6

Benjamin Draves

11/25/2019

Exercise 1: BDA 14.1

In this exercise we look to fit a Bayesian linear regression model to the data in Table 7.3.

```r
#read in data
blue_earth <- c(rep(1, 14), rep(0, 41 - 14))
clay <- c(rep(0, 14),rep(1, 14),rep(0, 41 - 14 -14))
goodhue <- c(rep(0, 41-13), rep(1, 13))

first_floor <- c(0,0,0,0,0,1,0,0,0,1,0,0,0,0,
                 1,0,0,1,0,0,0,0,0,1,0,0,0,1,
                 0,1,0,1,0,0,0,0,0,0,0,0,0)
y <- log(c(5,13,7.2,6.8,12.8,5.8,9.5,6.0,3.8,14.3,1.8,6.9,4.7,9.5,
           0.9,12.9,2.6,3.5,26.6,1.5,13.0,8.8,19.5,2.5,9,13.1,3.6,6.9,
           14.3,6.9,7.6,9.8,2.6,43.5,4.9,3.5,4.8,5.6,3.5,3.9,6.7))
```

(a)

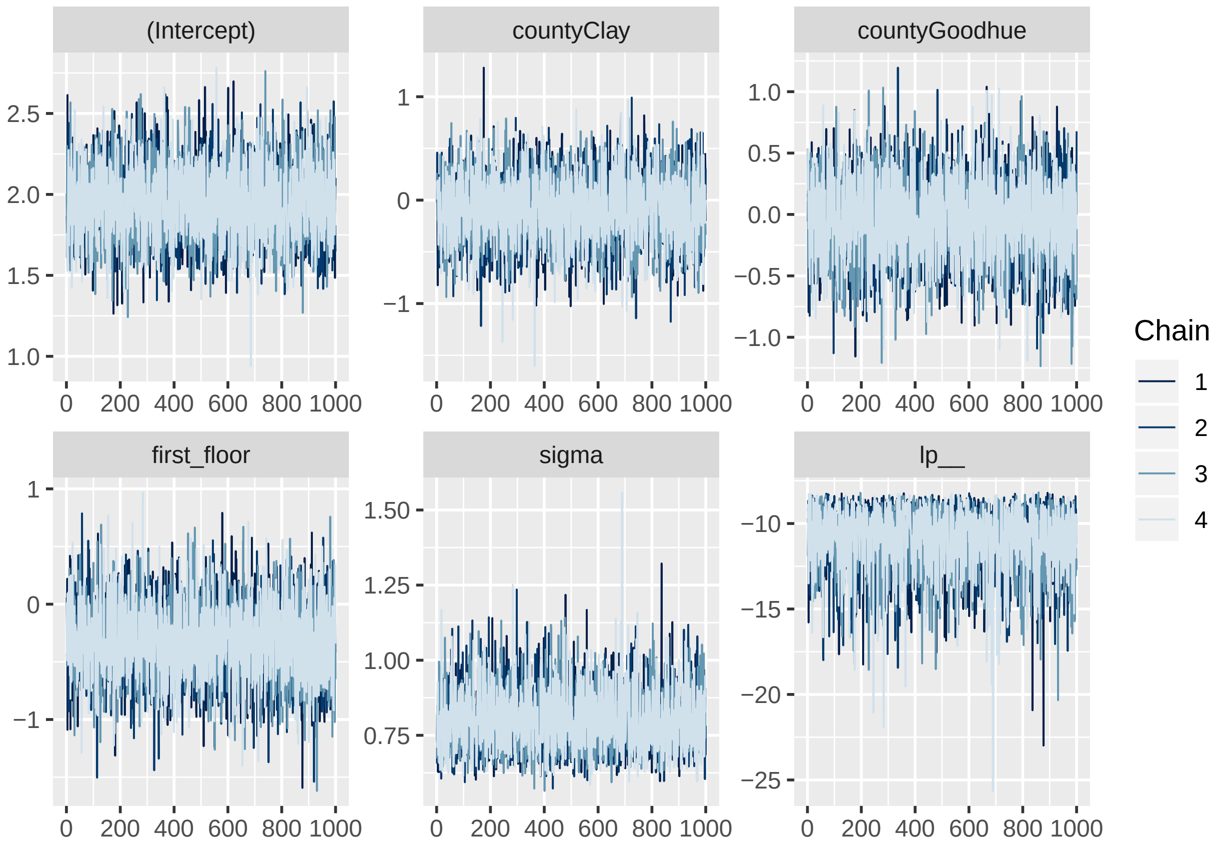
In this example we fit a Bayesian Linear Regression model to this data. As we intend to fit the model

$$\log(\text{Radon Concentrations}) = \beta_0 + \beta_1 I(\text{Blue Earth}) + \beta_2 I(\text{Clay}) + \beta_3 I(\text{Goodhue})$$

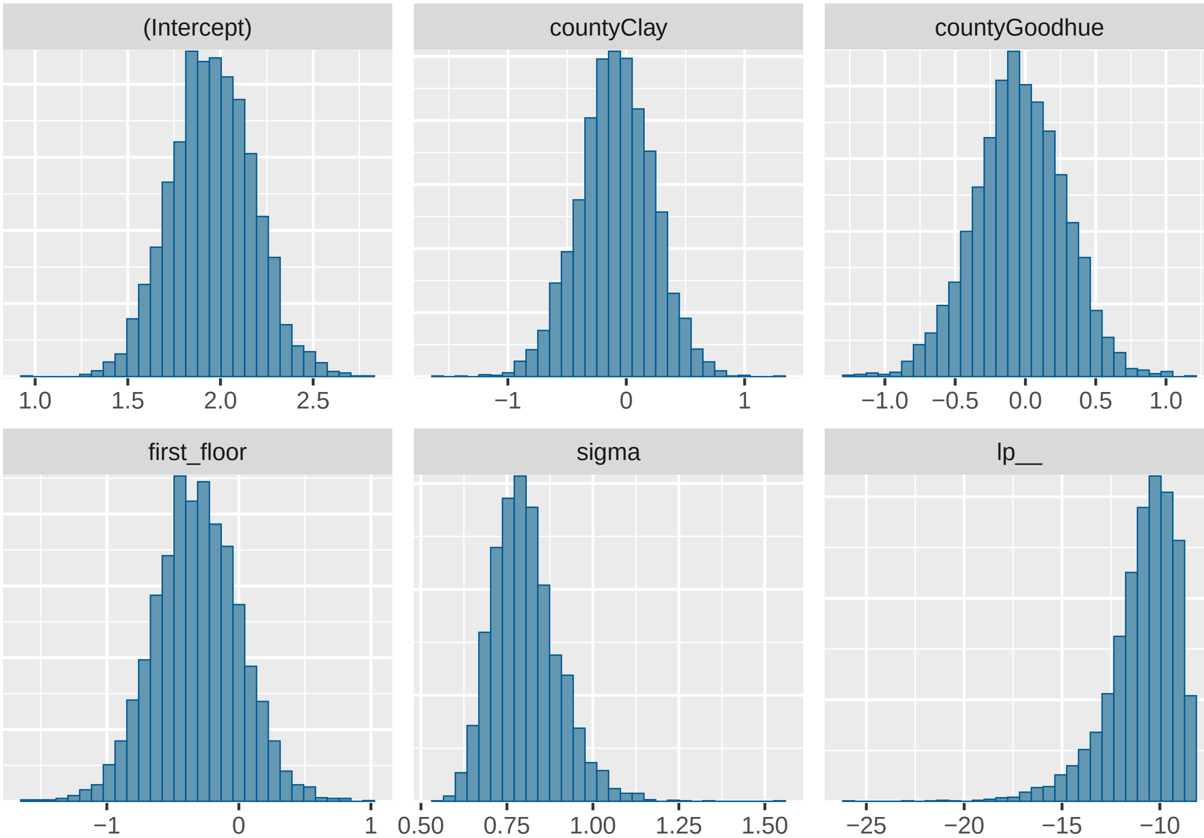
with a normal-inverse chi square prior, we can use the same code we built together in class to draw from the posterior $p(\beta, \sigma^2|Y)$.

We give posterior checks to ensure that our MCMC samples indeed comes from an approximation of the posterior distribution. Moreover, we provide a few posterior model checks for the Bayesian Linear Regression.

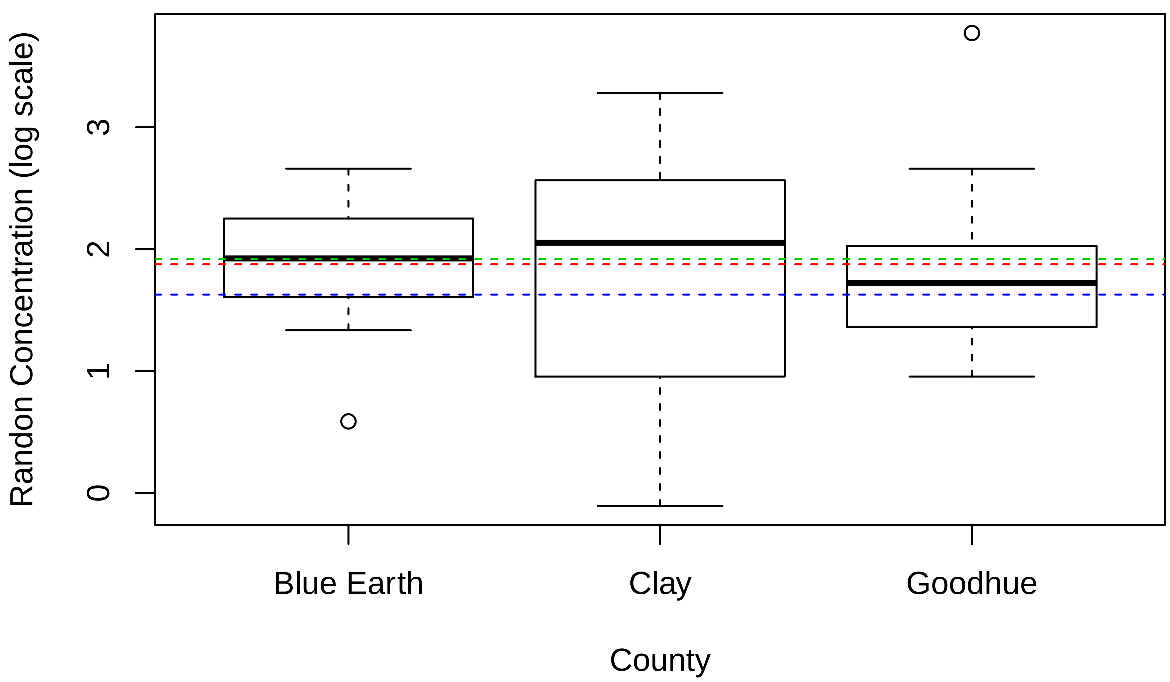
```r
#set up model matrix
county <- c(rep("Blue Earth", 14), rep("Clay", 14), rep("Goodhue", 13))
X <- cbind(model.matrix(~county), first_floor)
n <- length(y); p <- ncol(X)

#fit model
fit <- bslm_fit(y, X)
fit

#get simulations
library(rstan)
library(bayesplot)
library(beanplot)

sims <- bslm_sample(y, X)
mcmc_trace(sims)
```




```
mcmc_hist(sims)
```



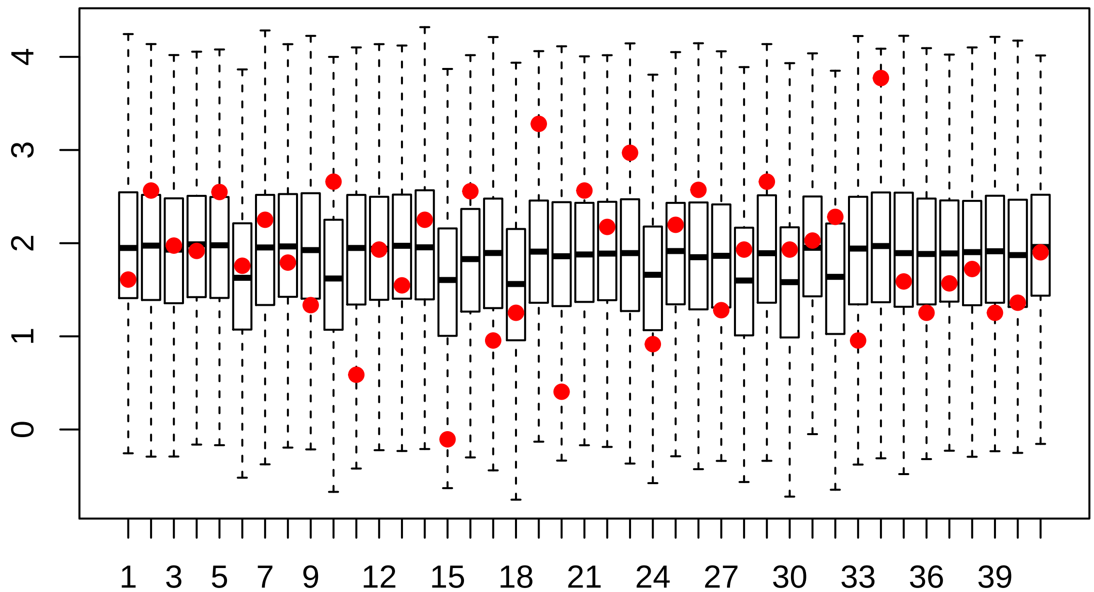

```r
# visualize
boxplot(y~county, ylab = "Randon Concentration (log scale)", xlab = "County")
for (i in 2:4) abline(h = fit$coef[1] + fit$coef[i], col = i, lty = "dashed")
```



In the above plot we plot the Randon concentration against the county. In addition we plot the MAP estimates for the mean of each of these counties. It does not apper that there is much disagreement between these groupings.

Next we complete a few posterior checks to assess goodness of fit.

```
ns <- dim(sims)[1]
y_rep <- matrix(nrow = ns, ncol = n)
for (is in 1:ns)
  y_rep[is,] <- rnorm(n, X %*% sims[is, 1, 1:p], sims[is, 1, p + 1])

#residual plot
boxplot(y_rep, outline = F); points(y, pch = 19, col = "red")
```

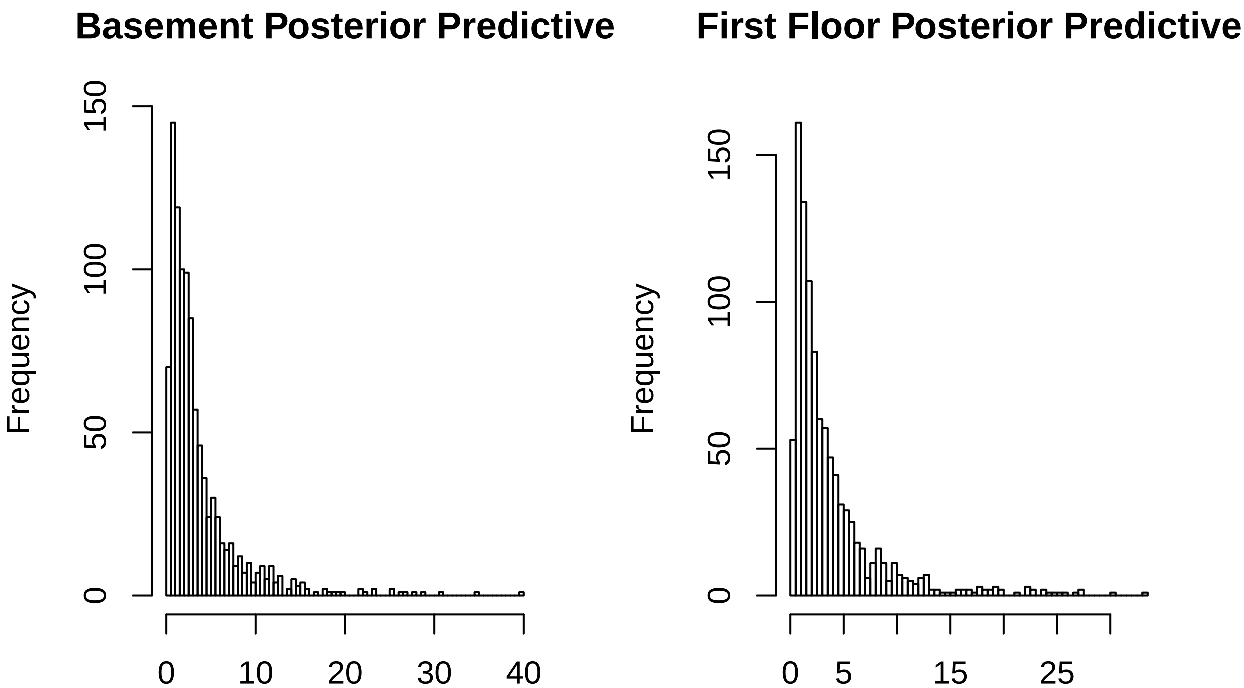


From the above residual plot it shows that the model may not fit very well. A few points fall outside the replicated posterior intervals suggesting that the infered posterior parameters and the model in general may not capture the variation in the dataset. This could be a function of the covariates not explaining the data well or a fundamentally incorrect assumption regarding the model.

(b)

In this exercise we look to build a posterior predictive distribution for a randon measurement in Blue Earth County. We can utilize the MC samples from part (a) to handle this task. Indeed we can sample $y_i^{(b)}|\Theta \sim N(\beta_0^{(b)}, (\sigma^2)^{(b)})$ for $b = 1, 2, \dots, B$. Notice as Blue Earth County was our baseline group that the mean corresponds only to β_0. Moreover, we plot two predictive distributions - one or a measurement made in the basement and for first floor measurements - where the mean is alterted to $N(\beta_0^{(b)} + \alpha^{(b)}, (\sigma^2)^{(b)})$ where $\alpha^{(b)}$ is the effect of the first floor measurement. To that end, consider the following histograms.

```
#basement measuremnts
y_base <- exp(rnorm(sims[,1,1], sims[,1,5]))
y_first <- exp(rnorm(sims[,1,1] + sims[,1,4], sims[,1,5]))

#plot posteriro predictives
par(mfrow = c(1,2))
hist(y_base, main = "Basement Posterior Predictive", breaks = 100, xlab = "")
hist(y_first, main = "First Floor Posterior Predictive", breaks = 100, xlab = "")
```



This plot suggests that on average we predict there to be less than a 20 concentration of Randon (measured in the basement or not) for a new homein Blue Earth County. The 95% quantiles are given by; basement 0.309, 15.179 and first floor 0.359, 18.401.

Exercise 2

BDA 14.11

Suppose that we can an errors-in-variables model with observations $\{(x_i, y_i)\}_{i=1}^n$ such that $(x_i \quad y_i)^T \sim N((u_i, a + bu_i), \Sigma)$. The goal of this exercise is to estimate the parameters (a, b) in the presence of errors in both variables.

(a)

Assume that $u_i \sim N(\mu, \tau^2)$. Then we can write the likehood of the data X, Y given the parameters $\Theta = \{a, b, \mu, \tau^2, \Sigma, u\}$ as follows

$$
\begin{aligned}
p(X, Y|\Theta) &\propto p(X, Y|u, a, b, \Sigma) \\
&= \prod_{i=1}^{n} \left[(2\pi)^{-2/2} |\Sigma|^{-1/2} \exp\left(-\frac{1}{2} \begin{bmatrix} x_i - u_i \\ y_i - (a + bu_i) \end{bmatrix}^T \Sigma^{-1} \begin{bmatrix} x_i - u_i \\ y_i - (a + bu_i) \end{bmatrix} \right) \right] \\
&\propto (|\Sigma|)^{-n/2} \exp\left(-\frac{1}{2} \sum_{i=1}^{n} \left(\begin{bmatrix} x_i - u_i \\ y_i - (a + bu_i) \end{bmatrix}^T \Sigma^{-1} \begin{bmatrix} x_i - u_i \\ y_i - (a + bu_i) \end{bmatrix} \right) \right)
\end{aligned}
$$

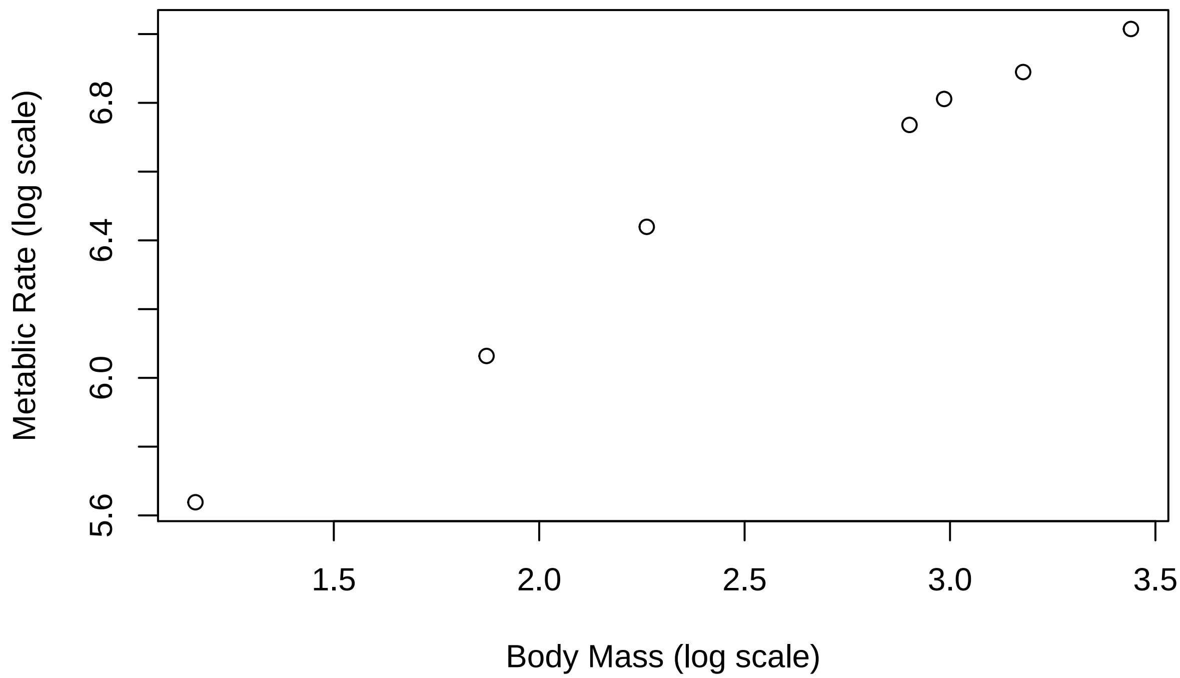
(b)

The paramters (a, b) are the parametes describing the mean behavior of y given x. As we have no reason to believe this value takes on any particular value reaonable priors might be $p(a, b) \propto 1$. For an empirical Bayes approach, we could instead assume that $p(a) \sim N(\bar{y}, \sigma^2)$ and $p(b) \sim N(\hat{b}_{OLS}, \sigma^2)$ which are informative priors given the training data. For the purpose of this exercise I will assume $p(\alpha, \beta) \propto 1$.

BDA 14.12

In this exercise we will model $y = \log(\text{Body Mass})$ as a linear function of $x = \log(\text{Metabolic rate})$. We import this data here.

```
#read in data
x <- log(c(31.2, 24.0, 19.8, 18.2, 9.6, 6.5, 3.2))
y <- log(c(1113, 982, 908, 842, 626, 430, 281))

#visualize relationship
plot(x, y, xlab = "Body Mass (log scale)", ylab = "Metablic Rate (log scale)")
```



(a)

For simplicity we assume that $p(a, b) \propto 1$ and that $\Sigma = \sigma^2 I$. Therefore the likelihood derived in exercise 14.11 can be rewritten as

$$p(X, Y|\Theta) \propto (\tau^2)^{-n/2}(\sigma^2)^{-n} \exp\left(-\frac{1}{2}\sum_{i=1}^{n}\left[\frac{1}{\sigma^2}(x_i - u_i)^2 + \frac{1}{\sigma^2}(y_i - (a + bu_i))^2 + \frac{1}{\tau^2}(u_i - \mu)^2\right]\right)$$

To derive the full posterior for Θ consider the following

$$
\begin{aligned}
p(\Theta|X,Y) &\propto p(X,Y|u,a,b,\sigma^2)p(u|\mu,\tau^2)p(a,b,\sigma^2.\tau^2,\mu) \\
&\propto (\tau^2\sigma^2)^{-n/2}\exp\left(-\frac{1}{2}\sum_{i=1}^{n}\left[\frac{1}{\sigma^2}(x_i-u_i)^2+\frac{1}{\sigma^2}(y_i-(a+bu_i))^2+\frac{1}{\tau^2}(u_i-\mu)^2\right]\right)
\end{aligned}
$$

To prerform posterior inference on this model it be necessary to turn to computational tools. In particular, we develop a Gibbs sampler to iterate through the paramters (a, b, σ^2, u) and sample from a Markov Chain with stationary distribution $p(\Theta|X, Y)$. In order to complete this, we will need access to the coniditional posteriors $p(u|\Theta \setminus u, X, Y)$, $p(a|\Theta \setminus a, X, Y)$, $p(b|\Theta \setminus b, X, Y)$, and $p(\sigma^2|\Theta \setminus \sigma^2, X, Y)$. We begin with a.

$$
\begin{aligned}
p(a|\Theta\setminus a,X,Y) &\propto \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(-2y_ia+a^2+2abu_i)\right) \\
&= \exp\left(\frac{(\bar{y}-b\bar{u})}{\sigma^2/n}a-\frac{1}{2\sigma^2/n}a^2\right)
\end{aligned}
$$

Therefore we see that $a|\Theta \setminus a, X, Y \sim N(\bar{y} - b\bar{u}, \sigma^2/n)$. Next we turn to the conditional update on b.

$$
\begin{aligned}
p(b|\Theta\setminus b,X,Y) &\propto \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(-2y_ibu_i+2abu_i+b^2u_i^2)\right) \\
&= \exp\left(\frac{\bar{yu}-\bar{u}a}{\sigma^2/n}b-\frac{\bar{u^2}}{2\sigma^2/n}b^2\right)
\end{aligned}
$$

Therefore we see that $b|\Theta \setminus b, X, Y \sim N\left(\frac{\bar{yu}-\bar{u}a}{\bar{u^2}}, \frac{\sigma^2/n}{\bar{u^2}}\right)$. Next we turn to updating σ^2. For ease of notation let $RSS(X) = \sum_{i=1}^{n}(x_i - u_i)^2$ and $RSS(Y) = \sum_{i=1}^{n}(y_i - (a + bu_i))^2$. Then we have

$$
p(\sigma^2|\Theta\setminus\sigma^2,X,Y) \propto (\sigma^2)^{-n}\exp\left(-\frac{1}{2\sigma^2}(RSS(X)+RSS(Y)\right)
$$

which we recogonize as $\sigma^2|\Theta\setminus\sigma^2, X, Y \sim \text{Inv-}\chi^2(2n, \frac{RSS(X)+RSS(Y)}{2n})$. Finally, we look to derive the conditional posterior distribution for u_i.

$$
\begin{aligned}
p(u_i|\Theta\setminus u_i,X,Y) &\propto \exp\left(-\frac{(x_i-u_i)^2}{2\sigma^2}-\frac{(y_i-(a+bu_i)^2)}{2\sigma^2}-\frac{(u_i-\mu)^2}{2\tau^2}\right) \\
&\propto \exp\left(-\frac{-2x_iu_i+u_i^2}{2\sigma^2}-\frac{-2(y_ib-ab)u_i+b^2u_i^2}{2\sigma^2}-\frac{-2\mu u_i+u_i^2}{2\tau^2}\right) \\
&= \exp\left(\left(\frac{x_i}{\sigma^2}+\frac{y_ib-ab}{\sigma^2}+\frac{\mu}{\tau^2}\right)u_i-\frac{1}{2}\left(\frac{1+b^2}{\sigma^2}+\frac{1}{\tau^2}\right)u_i^2\right)
\end{aligned}
$$

Thefore, we see that this distribution can be written as

$$
u_i|\Theta\setminus u_i,X,Y \sim N\left(\left[\frac{1}{\frac{1+b^2}{\sigma^2}+\frac{1}{\tau^2}}\right]\left(\frac{x_i+(y_i-a)b}{\sigma^2}+\frac{\mu}{\tau^2}\right),\frac{1}{\frac{1+b^2}{\sigma^2}+\frac{1}{\tau^2}}\right)
$$

With these derivations our Gibbs sampler tales the following form.

1. Sample $a \sim p(a|\Theta \setminus a, X, Y)$
2. Sample $b \sim p(b|\Theta \setminus b, X, Y)$
3. Sample $\sigma^2 \sim p(\sigma^2|\Theta \setminus \sigma^2, X, Y)$
4. For $i = 1, 2, \ldots n$ Do:

   i. Sample $u_i \sim p(u_i|\Theta \setminus u_i, X, Y)$

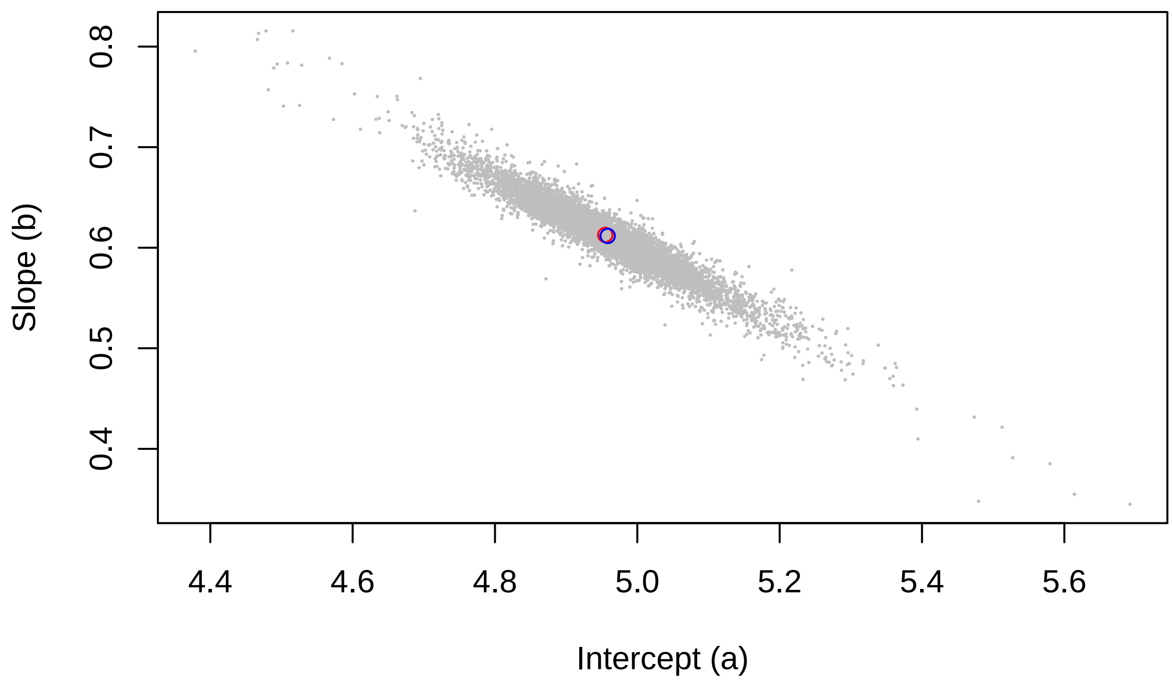
We implement this sampler below.

```
#--------------------------
#
#   Gibbs Sampler for
#   Errors in Variables
#     Model
#
#--------------------------

#set up storage + convergence
conv.crit <- FALSE
max.iters <- 20000
samps <- matrix(NA, nrow = 11, ncol = max.iters)
rownames(samps) <- c("iter", "a", "b", "sigma_2", paste0("u",1:length(x)))

#initialize values
a <- unname(coef(lm(y~x))[1])
b <- unname(coef(lm(y~x))[2])
sigma_2 <- 1
U <- x
tau_2 <- 10
mu <- coef(lm(y~x))[2]

#get sampler fro inverse chi square
rinvsquare <- function (ns, nu, nu_tau2) 1 / rgamma(ns, nu / 2, nu_tau2 / 2)

#compute useful values
iter <- 1
ybar <- mean(y)
n <- length(x)
set.seed(1985)
while(!conv.crit){

  #sample a
  a <- rnorm(1, mean  = ybar - b*mean(U), sd = sqrt(sigma_2/n))

  #sample b
  b <- rnorm(1, mean = (sum(y*U) - sum(U)*a) / (sum(U^2)) , sd = sqrt(sigma_2 / sum(U^2)))

  #sample sigma_2
  RSSX <- sum((x - U)^2)
  RSSY <- sum((y - (a  + b*U))^2)
  sigma_2 <- rinvsquare(1, 2*n, (RSSX + RSSY))

  #sample U's
  for(i in 1:length(x)){
    #set parameters
```

```
    var.here <- 1 / ( (1 + b^2)/sigma_2   +  1/tau_2 )
    mu.here <- var.here * ((x[i] + (y[i] - a)*b)/sigma_2 + mu/tau_2)

    #sample
    U[i] <- rnorm(1, mu.here, sqrt(var.here))
  }

  #store / update
  samps[, iter] <- c(iter, a, b, sigma_2, U)
  iter <- iter + 1


  #update convergence criterion
  conv.crit <- iter > max.iters

}
```

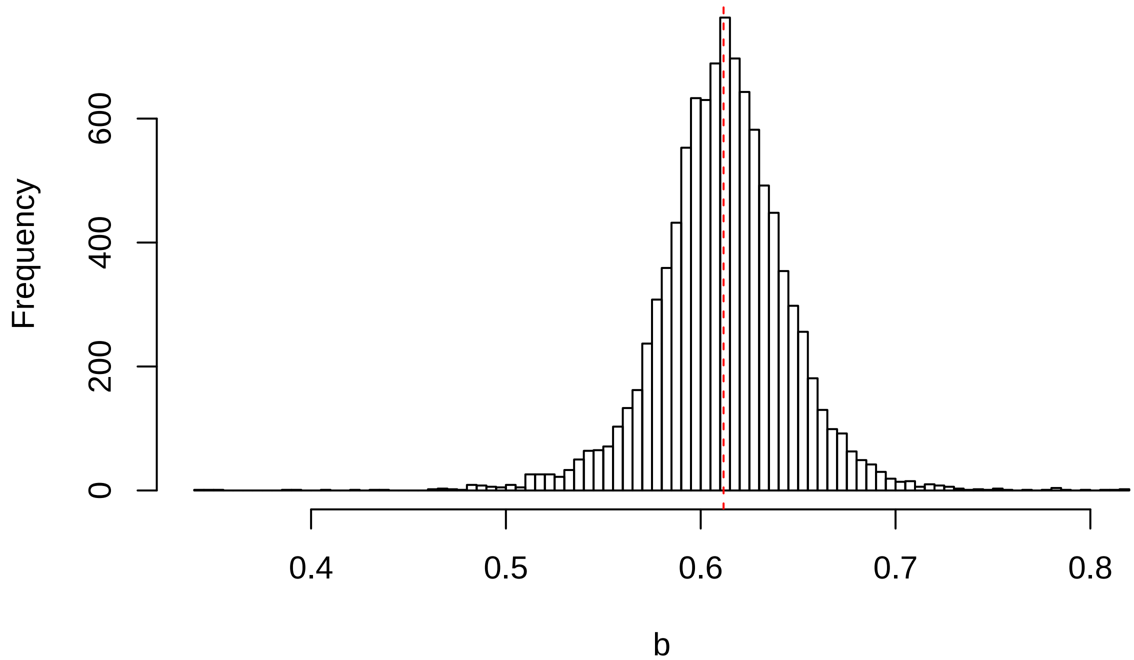
Using a warm up period of 10^4, we can visualize the fit of the model as follows. Moreover, we can plot the density of the MCMC estimates of the pair (a,b).

```
#visualize fit
warmup <- floor(max.iters/2)
#plot(x, y, xlab = "Body Mass (log scale)", ylab = "Metablic Rate (log scale)")
#abline(a = mean(samps[2, warmup:max.iters]), b = mean(samps[3,warmup:max.iters]), lwd = .1)

#Visualize estimates
plot(samps[2,warmup:max.iters], samps[3,warmup:max.iters],
     xlab = "Intercept (a)", ylab = "Slope (b)",
     cex = .1, col = "grey")
points(x = coef(lm(y~x))[1], y = coef(lm(y~x))[2], col = "red")
points(x = mean(samps[2,warmup:max.iters]), y = mean( samps[3,warmup:max.iters]), col = "blue")
```



b

```r
hist(samps[3,warmup:max.iters], xlab = "b", breaks = 100, main = "")
abline(v = mean(samps[3, warmup:max.iters]), col = "red", lty = "dashed")
```



From this we can see that on average we expect for a unit increase in Body Mass to see a 0.61 times increase in Metabolic Rate.

BDA 14.13

In this section we exercise we extend the model considered in 14.12 to the multiple linear regerssion setting. Assume that $\mathbb{E}[y_i|x_i^{(1)}, x_i^{(2))}] = a + bu_i + cw_i$ for $u_i, w_i \overset{i.i.d}{\sim} N(\mu, \tau^2)$ we can our posterior distribution to be of the form

$$\begin{aligned}
p(\Theta|X,Y) &\propto p(X,Y|u,a,b,c,\sigma^2)p(u|\mu,\tau^2)p(w|\mu,\tau^2)p(a,b,\sigma^2,\tau^2,\mu) \\
&\propto (\tau^2)^{-n}(\sigma^2)^{-3n/2}\exp\Big(-\frac{1}{2}\sum_{i=1}^n\Big[\frac{1}{\sigma^2}(x_i^{(1)}-u_i)^2 + \frac{1}{\sigma^2}(x_i^{(2)}-w_i)^2 + \frac{1}{\sigma^2}(y_i-(a+bu_i+cw_i))^2 \\
&\qquad + \frac{1}{\tau^2}(u_i-\mu)^2 + \frac{1}{\tau^2}(w_i-\mu)^2\Big]\Big)
\end{aligned}$$

From here we can devise a similar Gibbs sampling approach to posterior inference by cycling through the parameter set $\Theta = \{a, b, c, \sigma^2, u_i, w_i\}$. These derivations follow almost exactly from those derived above and we only expand on those that are substantively different. We summarize each conditional distribution below.

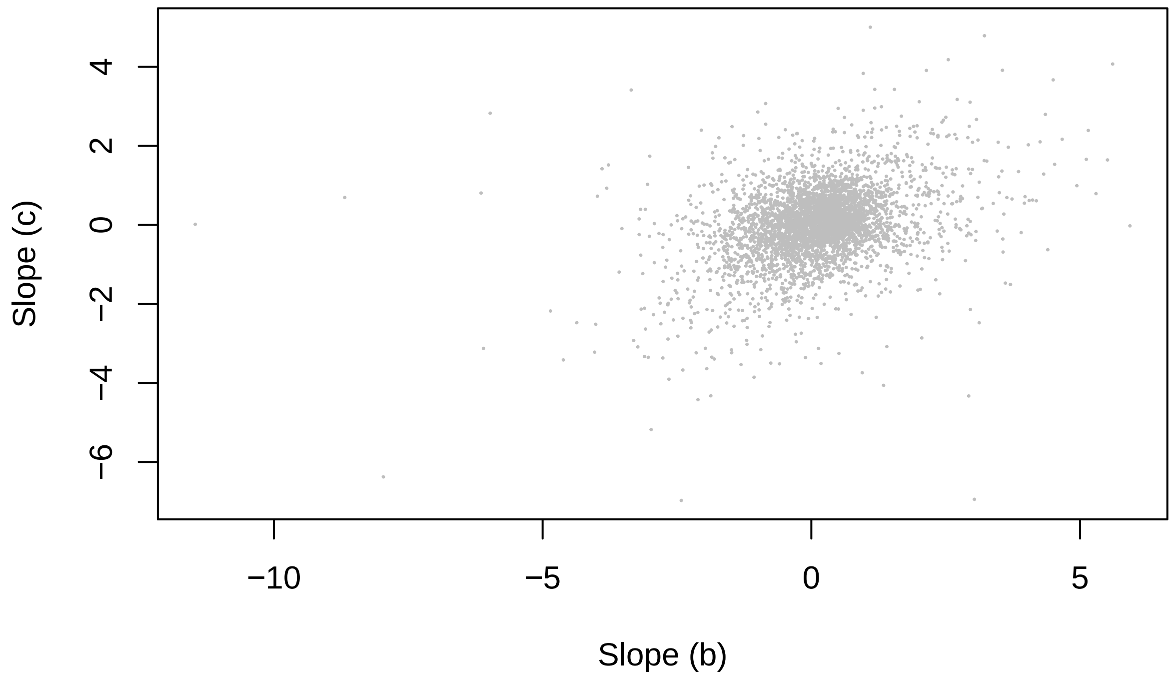
$$
\begin{aligned}
a|\Theta \setminus a, X, Y &\sim N\left(\bar{y} - b\bar{u} - c\bar{w}, \frac{\sigma^2}{n}\right) \\
b|\Theta \setminus b, X, Y &\sim N\left(\frac{\sum_{i=1}^n y_i u_i - a\sum_{i=1}^n u_i - c\sum_{i=1}^n w_i u_i}{\sum_{i=1}^n u_i^2}, \frac{\sigma^2}{\sum_{i=1}^n u_i^2}\right) \\
c|\Theta \setminus c, X, Y &\sim N\left(\frac{\sum_{i=1}^n y_i w_i - a\sum_{i=1}^n w_i - b\sum_{i=1}^n w_i u_i}{\sum_{i=1}^n w_i^2}, \frac{\sigma^2}{\sum_{i=1}^n w_i^2}\right) \\
\sigma^2|\Theta \setminus \sigma^2, X, Y &\sim \text{Inv-}\chi^2\left(3n, \frac{RSS(X_1) + RSS(X_2) + RSS(Y)}{3n}\right) \\
u_i|\Theta \setminus u_i, X, Y &\sim N\left(\left[\frac{1}{\frac{1+b^2}{\sigma^2} + \frac{1}{\tau^2}}\right]\left(\frac{x_i^{(1)} + (y_i - a - cw_i)b}{\sigma^2} + \frac{\mu}{\tau^2}\right), \frac{1}{\frac{1+b^2}{\sigma^2} + \frac{1}{\tau^2}}\right) \\
w_i|\Theta \setminus w_i, X, Y &\sim N\left(\left[\frac{1}{\frac{1+c^2}{\sigma^2} + \frac{1}{\tau^2}}\right]\left(\frac{x_i^{(2)} + (y_i - a - bu_i)c}{\sigma^2} + \frac{\mu}{\tau^2}\right), \frac{1}{\frac{1+c^2}{\sigma^2} + \frac{1}{\tau^2}}\right)
\end{aligned}
$$

We sample through each of these conditional and to sample from the posterior.

```r
#read in z values
z <- log(c(10750, 8805, 7500, 7662, 5286, 3724, 2423))

#set up storage + convergence
conv.crit <- FALSE
max.iters <- 20000
samps <- matrix(NA, nrow = 19, ncol = max.iters)
rownames(samps) <- c("iter", "a", "b", "c", "sigma_2",
                     paste0("u",1:length(x)), paste0("w",1:length(x)))

#initialize values
a <- unname(coef(lm(y~x+z))[1])
b <- unname(coef(lm(y~x+z))[2])
c <- unname(coef(lm(y~x+z))[3])
sigma_2 <- 1
U <- x
W <- z
tau_2 <- 10
mu <- mean(coef(lm(y~x+z))[2:3])

#get sampler fro inverse chi square
rinvsquare <- function (ns, nu, nu_tau2) 1 / rgamma(ns, nu / 2, nu_tau2 / 2)

#compute useful values
iter <- 1
ybar <- mean(y)
n <- length(x)
set.seed(1985)


while(!conv.crit){

  #sample a
  a <- rnorm(1, ybar - b *mean(U) - c*mean(W), sqrt(sigma_2/n))
```

```
  #sample b
  b <- rnorm(1,(sum(y*U) - a*sum(U) - c*sum(W*U)) / sum(U^2) , sqrt(sigma_2/sum(U^2)))

  #sample c
  c <- rnorm(1,(sum(y*W) - a*sum(W) - b*sum(W*U)) / sum(W^2) , sqrt(sigma_2/sum(W^2)))

  #sample sigma^2
  RSSX <- sum((x - U)^2)
  RSSY <- sum((y - (a + b*U + c*W))^2)
  RSSZ <- sum((z - W)^2)
  sigma_2 <- rinvsquare(1, 3*n, (RSSX + RSSY + RSSZ))

  #update U
  for(i in 1:length(x)){
    #set parameters
    var.here <- 1 / ( (1 + b^2)/sigma_2   +  1/tau_2 )
    mu.here <- var.here * ((x[i] + (y[i] - a - c*W[i])*b)/sigma_2 + mu/tau_2)

    #sample
    U[i] <- rnorm(1, mu.here, sqrt(var.here))
  }

  #Update W
  for(i in 1:length(x)){
    #set parameters
    var.here <- 1 / ( (1 + c^2)/sigma_2   +  1/tau_2 )
    mu.here <- var.here * ((z[i] + (y[i] - a - c*U[i])*c)/sigma_2 + mu/tau_2)

    #sample
    W[i] <- rnorm(1, mu.here, sqrt(var.here))
  }

  #store / update
  samps[, iter] <- c(iter, a, b, c, sigma_2, U, W)
  iter <- iter + 1


  #update convergence criterion
  conv.crit <- iter > max.iters

}
```

```
#visualize slope estimates
warmup <- floor(max.iters/2)
plot(samps[3, warmup:max.iters], samps[4, warmup:max.iters],
     xlab = "Slope (b)", ylab = "Slope (c)",
     cex = .1, col = "grey")
```



Due to the high colinearity of the features we see that the coefficient estimates are also highligh correlated. This is made apparent when plotting the posterior estimates of each of these parameters.

Exercise 3

Consider the following Bayesian linear regression model

$$\begin{aligned} y|\beta, \sigma^2 &\sim N(X\beta, \sigma^2 I) \\ \beta|\sigma^2 &\sim N(\beta_0, \sigma^2\Sigma_0) \\ \sigma^2 &\sim \text{Inv-}\chi^2(\nu, \tau^2) \end{aligned}$$

(a)

In this exercise we can derive the joint posterior. First notice we can write

$$\begin{aligned} p(\beta, \sigma^2|y) &\propto p(y|\beta, \sigma^2)p(\beta|\sigma^2)p(\sigma^2) \\ &\propto (\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2}(y - X\beta)^T(y - X\beta)\right)(\sigma^2)^{-p/2} \exp\left(-\frac{1}{2\sigma^2}(\beta - \beta_0)^T\Sigma_0^{-1}(\beta - \beta_0)\right)(\sigma^2)^{-\nu/2-1} \exp\left(-\frac{\nu\tau^2}{2\sigma^2}\right) \\ &\propto (\sigma^2)^{-\frac{n+p+\nu}{2}-1} \exp\left\{-\frac{1}{2\sigma^2}\left[\nu\tau^2 + RSS(\beta) + (\beta - \beta_0)^T\Sigma_0^{-1}(\beta - \beta_0)\right]\right\} \end{aligned}$$

From here it suffices to show

$$(y - X\beta)^T(y - X\beta) + (\beta - \beta_0)^T\Sigma_0^{-1}(\beta - \beta_0) = RSS(\hat{\beta}) + (\hat{\beta} - \beta_0)^T\Sigma_0^{-1}(\hat{\beta} - \beta_0) + (\beta - \hat{\beta})^T\Sigma_\beta^{-1}(\beta - \hat{\beta})$$

First notice by the centering trick for β^* the MLE we have

$$
\begin{aligned}
(y-X\beta)^T(y-X\beta)+(\beta-\beta_0)^T\Sigma_0^{-1}(\beta-\beta_0) &= (y-X\beta^*)^T(y-X\beta^*)+(X\beta^*-X\beta)^T(X\beta^*-X\beta)+(\beta-\beta_0)^T\Sigma_0^{-1}(\beta-\beta_0) \\
&= (y-X\beta^*)^T(y-X\beta^*)+(\beta^*-\beta)^T X^T X(\beta^*-\beta)+(\beta-\beta_0)^T\Sigma_0^{-1}(\beta-\beta_0)
\end{aligned}
$$

Next we can center this second term around $\hat{\beta}$

$$
\begin{aligned}
(\beta^*-\beta)X^T X(\beta^*-\beta) &= (\beta^*-\hat{\beta})X^T X(\beta^*-\hat{\beta})+(\hat{\beta}-\beta)X^T X(\hat{\beta}-\beta) \\
&+ 2(\hat{\beta}-\beta)^T X^T X(\hat{\beta}-\beta)
\end{aligned}
$$

Centering the third term around $\hat{\beta}$

$$
\begin{aligned}
(\beta-\beta_0)^T\Sigma_0^{-1}(\beta-\beta_0) &= (\hat{\beta}-\beta)^T\Sigma_0^{-1}(\hat{\beta}-\beta_0)+(\hat{\beta}-\beta_0)^T\Sigma_0^{-1}(\hat{\beta}-\beta_0) \\
&+ 2(\hat{\beta}-\beta)^T\Sigma_0^{-1}(\beta_0-\hat{\beta})
\end{aligned}
$$

Notice that these cross terms can be evaluated as follows.

$$
\begin{aligned}
(\hat{\beta}-\beta)^T X^T X(\beta^*-\hat{\beta})+(\hat{\beta}-\beta)^T\Sigma_0^{-1}(\beta_0-\hat{\beta}) &= (\hat{\beta}-\beta)^T\left(X^T X(\beta^*-\hat{\beta})+\Sigma_0^{-1}(\beta_0-\hat{\beta})\right) \\
&= (\hat{\beta}-\beta)^T\left(X^T X\beta^*-X^T X\hat{\beta}+\Sigma_0^{-1}\beta_0-\Sigma_0^{-1}\hat{\beta})\right) \\
&= (\hat{\beta}-\beta)^T\left(X^T X\beta^*-\Sigma_\beta\hat{\beta}+\Sigma_0^{-1}\beta_0)\right) \\
&= (\hat{\beta}-\beta)^T\left(X^T y+(X^T y-\Sigma_0^{-1}\beta_0)+\Sigma_0^{-1}\beta_0)\right) \\
&= 0
\end{aligned}
$$

Therefore, rewritting the following term we have

$$
\begin{aligned}
(\beta^*-\beta)^T X^T X(\beta^*-\beta)+(\beta-\beta_0)^T\Sigma_0^{-1}(\beta-\beta_0) &= (\hat{\beta}-\beta)^T(X^T X+\Sigma_0^{-1})(\hat{\beta}-\beta)+(\hat{\beta}-\beta_0)^T\Sigma_0^{-1}(\hat{\beta}-\beta_0) \\
&+ (\beta^*-\hat{\beta})^T X^T X(\beta^*-\hat{\beta})
\end{aligned}
$$

However, notice that we can write

$$
\begin{aligned}
RSS(\beta^*)+(\beta^*-\hat{\beta})^T X^T X(\beta^*-\hat{\beta}) &= (y-X\beta^*)^T y-X\beta^*)+(\beta^*-\hat{\beta})^T X^T X(\beta^*-\hat{\beta}) \\
&= (y-X\hat{\beta})^T(y-X\hat{\beta})+(X\hat{\beta}-X\beta^*)^T(X\hat{\beta}-X\beta^*) \\
&+ 2(X\hat{\beta}-X\beta^*)^T(y-X\hat{\beta})(\beta^*-\hat{\beta})^T X^T X(\beta^*-\hat{\beta}) \\
&= RSS(\hat{\beta})+2(\hat{\beta}-\beta^*)X^T X(\hat{\beta}^*)+2(\hat{\beta}-\beta^*)^T X^T(y-X\hat{\beta}) \\
&= RSS(\hat{\beta})+2(\hat{\beta}-\beta)^T X^T X(\beta^*-\hat{\beta})+(\hat{\beta}-\beta)^T\Sigma_0^{-1}(\beta_0-\hat{\beta}) \\
&= RSS(\hat{\beta})
\end{aligned}
$$

Combining this result with both centering tricks gives

$$
\begin{aligned}
RSS(\beta) + (\beta-\beta_0)^T\Sigma_0^{-1}(\beta-\beta_0) &= RSS(\beta^*) + (\beta^*-\hat{\beta})^T X^T X(\beta^*-\hat{\beta}) \\
&\quad + (\hat{\beta}-\beta_0)^T\Sigma_0^{-1}(\hat{\beta}-\beta_0) + (\beta-\hat{\beta})^T\Sigma_\beta^{-1}(\beta-\hat{\beta}) \\
&= RSS(\hat{\beta}) + (\hat{\beta}-\beta_0)^T\Sigma_0^{-1}(\hat{\beta}-\beta_0) + (\beta-\hat{\beta})^T\Sigma_\beta^{-1}(\beta-\hat{\beta})
\end{aligned}
$$

which is the desired result.

(b)

With the posterior we derived in part (a) we have

$$
\begin{aligned}
p(\sigma^2|y) &= \int p(\beta,\sigma^2|y)d\beta \\
&\propto (\sigma^2)^{-\frac{n+p+v}{2}-1}\exp\left\{-\frac{1}{2\sigma^2}\left[\nu\tau^2 + RSS(\hat{\beta}) + (\hat{\beta}-\beta_0)^T\Sigma_0^{-1}(\hat{\beta}-\beta_0)\right]\right\}\int_{\mathbb{R}}\exp\left(-\frac{1}{2\sigma^2}(\hat{\beta}-\beta)^T\Sigma_\beta^{-1}(\hat{\beta}-\beta)\right)d\beta \\
&= (\sigma^2)^{-\frac{n+p+v}{2}-1}\exp\left\{-\frac{1}{2\sigma^2}\left[\nu\tau^2 + RSS(\hat{\beta}) + (\hat{\beta}-\beta_0)^T\Sigma_0^{-1}(\hat{\beta}-\beta_0)\right]\right\}(2\pi\sigma^2)^{p/2}|\Sigma_\beta|^{1/2} \\
&\propto (\sigma^2)^{-\frac{n+v}{2}-1}\exp\left\{-\frac{1}{2\sigma^2}\left[\nu\tau^2 + RSS(\hat{\beta}) + (\hat{\beta}-\beta_0)^T\Sigma_0^{-1}(\hat{\beta}-\beta_0)\right]\right\}
\end{aligned}
$$

We recogonize this distribution as an inverse chi square with parameters

$$
\sigma^2|y \sim \text{Inv-}\chi^2\left(n+\nu, \frac{\nu\tau^2 + RSS(\hat{\beta}) + (\hat{\beta}-\beta_0)^T\Sigma_0^{-1}(\hat{\beta}-\beta_0)}{n+\nu}\right)
$$

(c)

Let $\beta_0 = 0$ so that $\hat{\beta} = \Sigma_\beta X^T y$. Consider the following expansion

$$
\begin{aligned}
RSS(\hat{\beta}) + \hat{\beta}^T\Sigma_0^{-1}\hat{\beta} &= (y - X\Sigma_\beta X^T y)^T(y - X\Sigma_\beta X^T y) + \hat{\beta}^T\Sigma_0^{-1}\hat{\beta} \\
&= y(I - X\Sigma_\beta X^T)(I - X\Sigma_\beta X^T)y + y^T X^T\Sigma_\beta\Sigma_0^{-1}\Sigma_\beta X^T y \\
&= y\left[I - X\Sigma_\beta X^T - X\Sigma_\beta X^T + X\Sigma_\beta X^T X\Sigma_\beta X^T + X\Sigma_\beta\Sigma_0^{-1}\Sigma_\beta X^T\right]y \\
&= y\left[I - X\left(\Sigma_\beta + \Sigma_\beta - \Sigma_\beta X^T X\Sigma_\beta - \Sigma_\beta\Sigma_0^{-1}\Sigma_\beta\right)X^T\right]y \\
&= y\left[I - X\left(2\Sigma_\beta - \Sigma_\beta(X^T X + \Sigma_0^{-1})\Sigma_\beta\right)X^T\right]y \\
&= y\left[I - X\left(2\Sigma_\beta - \Sigma_\beta\Sigma_\beta^{-1}\Sigma_\beta\right)X^T\right]y \\
&= y\left[I - X\left(2\Sigma_\beta - \Sigma_\beta\right)X^T\right]y \\
&= y\left[I - X\Sigma_\beta X^T\right]y \\
&= y\left[I - X(X^T X + \Sigma_0^{-1})X^T\right]y \\
&= y\left[I - H\right]y
\end{aligned}
$$