

STA 601 Homework 8

Lingyun Shao

Nov. 07, 2018

8.3

Hierarchical modeling: The files `school11.dat` through `school18.dat` give weekly hours spent on homework for students sampled from eight different schools. Obtain posterior distributions for the true means for the eight different schools using a hierarchical normal model with the following prior parameters:

$$\mu_0 = 7, \gamma_0^2 = 5, \tau_0^2 = 10, \eta_0 = 2, \sigma_0^2 = 15, \nu_0 = 2$$

- a) Run a Gibbs sampling algorithm to approximate the posterior distribution of $\{\theta, \sigma^2, \mu, \tau^2\}$. Assess the convergence of the Markov chain, and find the effective sample size for $\{\sigma^2, \mu, \tau^2\}$. Run the chain long enough so that the effective sample sizes are all above 1,000.

According to the result from textbook, we can easily derive the full conditional distributions:

Suppose we have m groups

Full conditional distribution of μ

$$\{\mu | \theta_1, \dots, \theta_m, \tau^2\} \sim \text{normal}\left(\frac{m\bar{\theta}/\tau^2 + \mu_0/\gamma_0^2}{m/\tau^2 + 1/\gamma_0^2}, \frac{1}{m/\tau^2 + 1/\gamma_0^2}\right)$$

Full conditional distribution of τ^2

$$\{1/\tau^2 | \theta_1, \dots, \theta_m, \mu\} \sim \text{gamma}\left(\frac{\eta_0 + m}{2}, \frac{\eta_0\tau_0^2 + \sum(\theta_j - \mu)^2}{2}\right)$$

Full conditional distribution of θ_j

$$\{\theta_j | \mu, y_{1,j}, \dots, y_{n_j,j}, \sigma^2\} \sim \text{normal}\left(\frac{n_j\bar{y}_j/\sigma^2 + \mu/\tau^2}{n_j/\sigma^2 + 1/\tau^2}, \frac{1}{n_j/\sigma^2 + 1/\tau^2}\right)$$

Full conditional distribution of μ

$$\{1/\sigma^2 | \theta, y_1, \dots, y_n\} \sim \text{gamma}\left(\frac{\nu_0 + \sum_{j=1}^m n_j}{2}, \frac{\nu_0\sigma_0^2 + \sum_{j=1}^m \sum_{i=1}^{n_j} (y_{i,j} - \theta_j)^2}{2}\right)$$

```
dt_url = paste0('http://www2.stat.duke.edu/~pdh10/FCBS/Exercises/school', 1:8, '.dat')
y = map(dt_url, read.csv, header = FALSE)
```

```
mu0 = 7; gam0_2 = 5; tau0_2 = 10
eta0 = nu0 = 2; sig0_2 = 15
```

```
# define updating functions
```

```
update_mu = function(theta, tau2) {
```

```

m = length(theta)
Var = 1/(m/tau2+1/gam0_2)
Mean = Var*(m*mean(theta)/tau2+mu0/gam0_2)
rnorm(1, Mean, sqrt(Var))
}

update_tau2 = function(theta, mu) {
  m = length(theta)
  a = (eta0+m)/2
  b = (eta0*tau0_2+sum((theta-mu)^2))/2
  1/rgamma(1, a, b)
}

update_theta = function(mu, tau2, sig2, y) {
  m = length(y)
  y.bar = map_dbl(1:m, function(i) y[[i]][,1] %>% mean)
  y.n = map_dbl(1:m, function(i) y[[i]][,1] %>% length)
  Var = 1/(y.n/sig2 + 1/tau2)
  Mean = Var*(y.n*y.bar/sig2 + mu/tau2)
  rnorm(m, Mean, sqrt(Var))
}

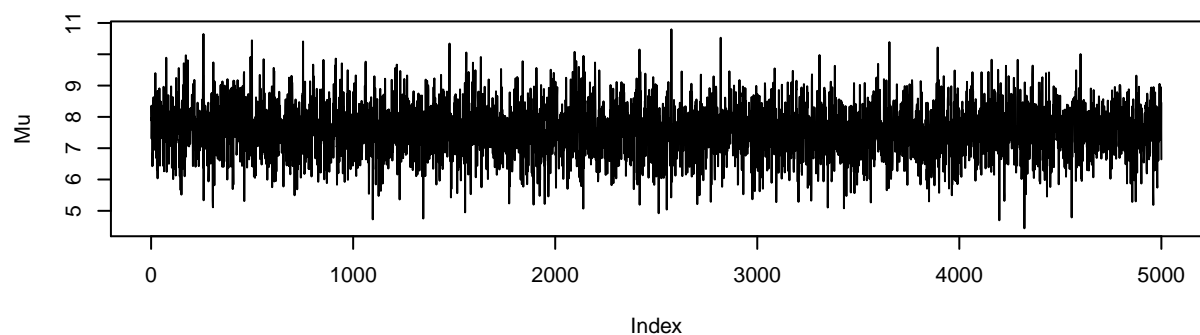
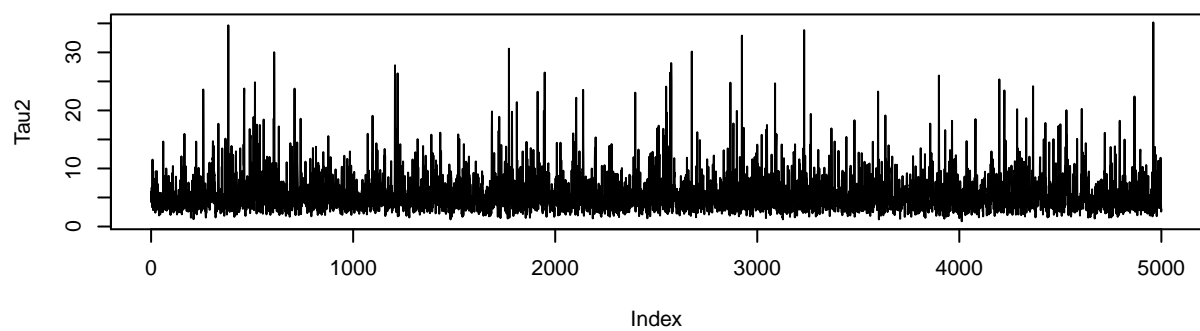
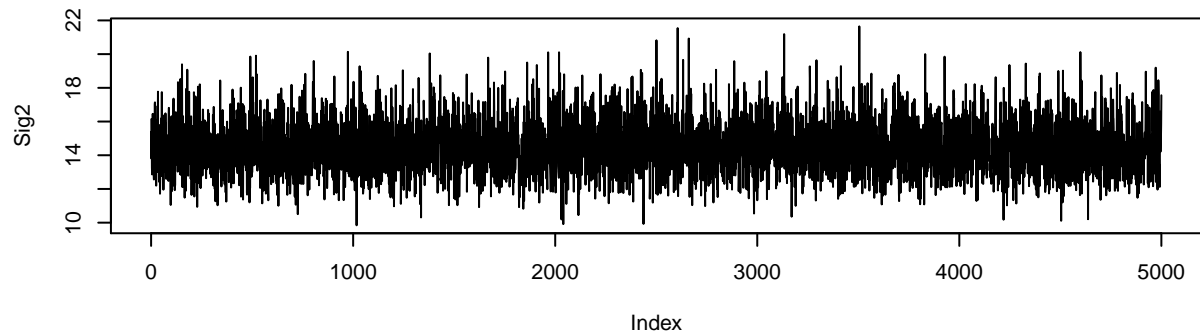
update_sig2 = function(theta, y) {
  m = length(y)
  n = y %>% unlist() %>% length()
  s2 = map_dbl(1:m, function(i) {(y[[i]][,1] - theta[i])^2 %>% sum}) %>% sum
  a = (nu0+n)/2
  b = (nu0*sig0_2 + s2)/2
  1/rgamma(1, a, b)
}

# initial values
m = length(y)
mu = mean(y %>% unlist)
sig2 = tau2 = var(y %>% unlist)
theta = map_dbl(1:m, ~y[[,]][,1] %>% mean)

ite = 5000
Mu = rep(NA, ite)
Tau2 = rep(NA, ite)
Theta = matrix(NA, nrow = ite, ncol = m)
Sig2 = rep(NA, ite)
for(i in 1:ite) {
  mu = update_mu(theta,tau2)
  tau2 = update_tau2(theta, mu)
  theta = update_theta(mu, tau2, sig2, y)
  sig2 = update_sig2(theta, y)
  Mu[i] = mu
  Tau2[i] = tau2
  Theta[i,] = theta
  Sig2[i] = sig2
}

```

```
# traceplot
par(mfrow = c(3,1))
plot(Sig2, type = 'l')
plot(Tau2, type = 'l')
plot(Mu, type = 'l')
```

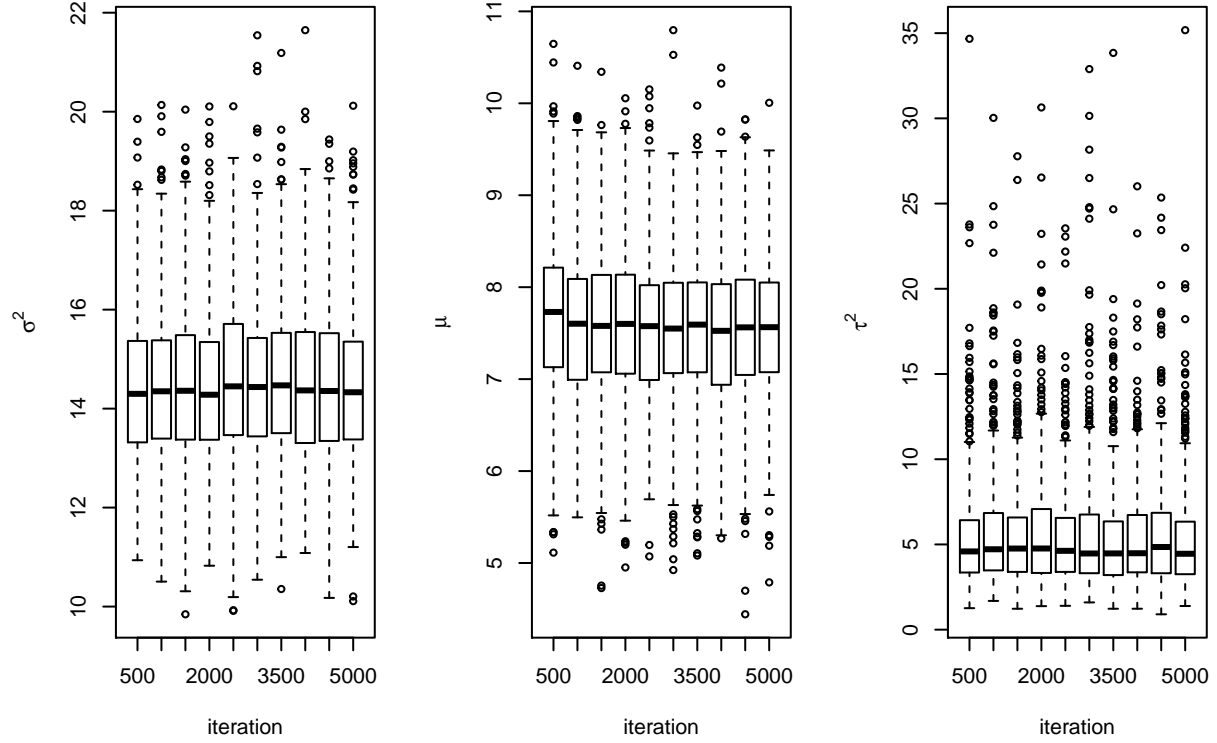


```
# stationary plot
stationarity.plot = function(x,...){
  S = length(x)
  scan = 1:S
  ng = min( round(S/100),10)
```

```

group = S*ceiling( ng*scan/S) /ng
boxplot(x~group,...) }
par(mfrow = c(1, 3))
stationarity.plot(Sig2,xlab="iteration",ylab=expression(sigma^2))
stationarity.plot(Mu,xlab="iteration",ylab=expression(mu))
stationarity.plot(Tau2,xlab="iteration",ylab=expression(tau^2))

```



```

sam.MCMC = list(Sig2, Tau2, Mu)
res.a = map_dbl(sam.MCMC, effectiveSize) %>% as.data.frame()
rownames(res.a) = c('$\\sigma^2$', '$\\tau^2$', '$\\mu$')
colnames(res.a) = 'Effective Sample'
kable(res.a, caption = 'Effective Sample Size (5000 MCMC samples)')

```

Table 1: Effective Sample Size (5000 MCMC samples)

Effective Sample	
σ^2	4584.717
τ^2	3522.533
μ	3836.632

Judging from the traceplot and stationary plot, we can clear see that our Markov Chain is stationary since the is no sign of our MCMC samples going in one direction or jumping up and down within parameter spaces. Besides, the relatively large effective sample sizes also tell us that our Markov Chain has a good mixing.

- b) Compute posterior means and 95% confidence regions for σ^2, μ, τ^2 . Also, compare the posterior densities to the prior densities, and discuss what was learned from the data.

```
res.b = cbind(map_dbl(sam.MCMC, mean),
               map_dbl(sam.MCMC, ~quantile(., 0.025)),
               map_dbl(sam.MCMC, ~quantile(., 0.975)))
rownames(res.b) = c('$\\sigma^2$', '$\\tau^2$', '$\\mu$')
colnames(res.b) = c('Mean', '2.5\\%', '97.5\\%')
res.b %>% kable(caption = 'Poerior Mean and CI of parameters')
```

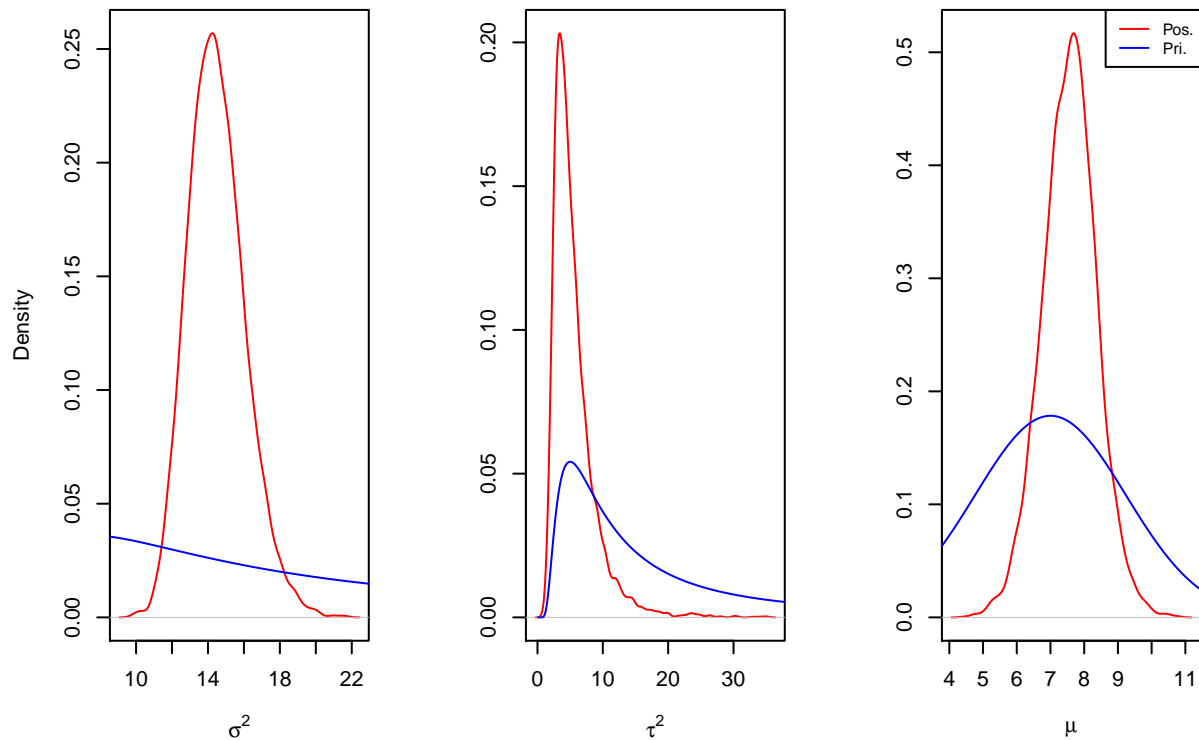
Table 2: Poerior Mean and CI of parameters

	Mean	2.5%	97.5%
σ^2	14.486598	11.740553	17.927292
τ^2	5.513436	1.960314	14.253303
μ	7.563180	5.944212	9.150816

```
par(mfrow = c(1,3))
set.seed(100)
plot(density(Sig2), xlab = expression(sigma^2), main = '', col = 2)
seq1 = seq(0.1, 100, by = 0.1)
lines(seq1, dinvgamma(seq1, nu0/2, nu0*sig0_2/2),
      col = 4)

plot(density(Tau2), xlab = expression(tau^2), main = '', col = 2, ylab = '')
seq2 = seq(0.1, 100, by = 0.1)
lines(seq2, dinvgamma(seq2, eta0/2, eta0*tau0_2/2),
      col = 4)

plot(density(Mu), xlab = expression(mu), main = '', col = 2, ylab = '')
seq3 = seq(0.1, 100, by = 0.1)
lines(seq3, dnorm(seq3, mu0, sqrt(gam0_2)), col = 4)
legend('topright', lty = 1, col = c(2, 4), legend = c('Pos.', 'Pri.'), cex = 0.75)
```



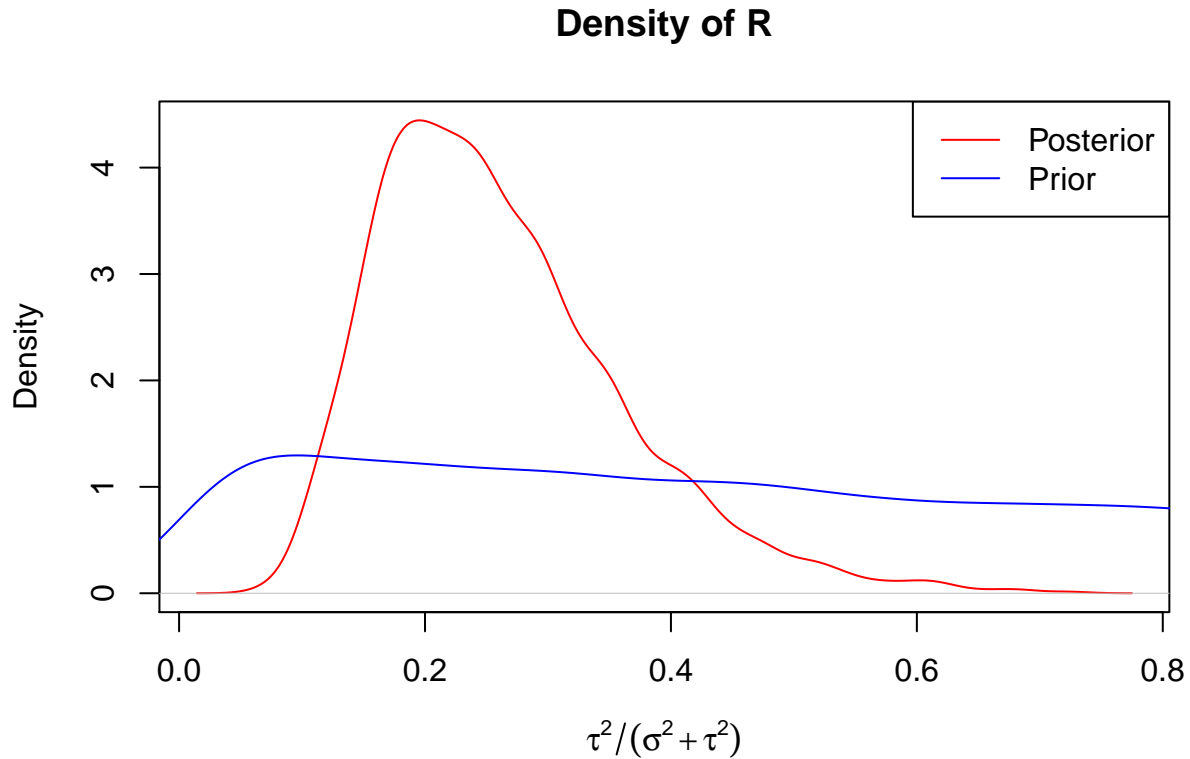
Comment:

The posteriors here for all parameters are much more peaked, i.e. concentrated than the prior distributions. Judging from the plot, we can see that our prior beliefs for these 3 parameters are with a huge degree of uncertainty. The prior distributions are widely spread. With the data, we updated our prior beliefs and get a more certain posterior belief as is shown in the posterior distributions.

Besides, we can see that the distribution of σ^2 has changed a lot before and after we observe the data. While the change in the other two distributions is relatively moderate.

- c) Plot the posterior density of $R = \frac{\tau^2}{\sigma^2 + \tau^2}$ and compare it to a plot of the prior density of R. Describe the evidence for between-school variation.

```
set.seed(100)
R.pos = Tau2/(Sig2+Tau2)
tau0_2.sam = 1/rgamma(ite, eta0/2, eta0*tau0_2/2)
sig0_2.sam = 1/rgamma(ite, nu0/2, nu0*sig0_2/2)
R.pri = tau0_2.sam/(sig0_2.sam+tau0_2.sam)
par(mfrow=c(1,1))
plot(density(R.pos), col = 2,
     xlab = expression(tau^2/(sigma^2 + tau^2)),
     main = 'Density of R')
lines(density(R.pri), col = 4)
legend('topright', lty = 1, col = c(2, 4), legend = c('Posterior', 'Prior'))
```



```
res.c = rbind(c(mean(R.pri), sd(R.pri)),
              c(mean(R.pos), sd(R.pos)))
colnames(res.c) = c('Mean', 'Sd')
rownames(res.c) = c('Prior', 'Posterior')
res.c %>% kable(caption = 'Statistics of R', digits = 3)
```

Table 3: Statistics of R

	Mean	Sd
Prior	0.439	0.285
Posterior	0.261	0.100

Comment:

$R = \frac{\tau^2}{\sigma^2 + \tau^2}$ is the proportion of between-school variation to the total variation. We have a prior belief that this proportion has a expectation of 0.439 with a large degree of uncertainty. However, once we observed the data, we can update our prior belief and would conclude that R is approximately round 0.261 with less uncertainty (with a standard deviation of 0.1).

So combining our prior belief and data, there is clear evidence that there is between school variation, which is about 0.261 of the total variation.

- d) Obtain the posterior probability that θ_7 is smaller than θ_6 , as well as the posterior probability that θ_7 is the smallest of all the θ 's.

```
set.seed(100)
res.d = cbind(mean(Theta[,7] < Theta[,6]),
```

```

mean(apply(Theta, 1, which.min) == 7))
colnames(res.d) <- c("$\\text{Pr}\\left(\\theta_7 < \\theta_6 \\mid \\boldsymbol{y}\\right)$",
"$\\text{Pr}\\left(\\theta_7 < \\text{min}\\left(\\theta_{-7}\\right) \\mid \\boldsymbol{y}\\right)$")
kable(res.d, caption = 'Inference about $\\theta$')

```

Table 4: Inference about θ

$\Pr(\theta_7 < \theta_6 \mathbf{y})$	$\Pr(\theta_7 < \min(\theta_{-7}) \mathbf{y})$
0.52	0.3234

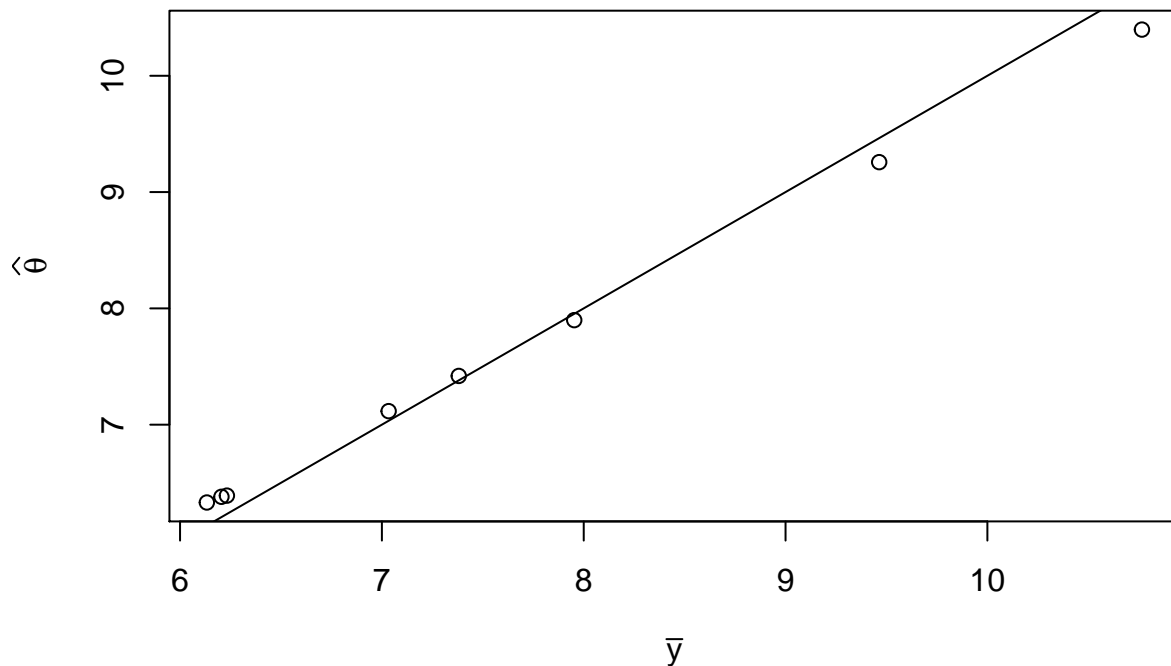
- e) Plot the sample averages $\bar{y}_1, \dots, \bar{y}_8$ against the posterior expectations of $\theta_1, \dots, \theta_8$, and describe the relationship. Also compute the sample mean of all observations and compare it to the posterior mean of μ .

```

y.bar = map_dbl(1:8, function(i) y[[i]][,1] %>% mean)
Theta.mean = colMeans(Theta)
plot(y.bar, Theta.mean, xlab = expression(bar(y)),
     ylab = expression(widehat(theta)),
     main = 'Posterior Expectations vs Sample Averages')
abline(a = 0, b = 1)

```

Posterior Expectations vs Sample Averages



```

y.mean = y %>% unlist %>% mean
mu.mean = mean(Mu)
kable(cbind("$\\bar{y}$" = y.mean,
            "$\\hat{\\mu}$" = mu.mean),
      caption = 'Sample Mean and Posterior mean of $\\mu$')

```


Table 5: Sample Mean and Posterior mean of μ

\bar{y}	$\hat{\mu}$
7.691278	7.56318

Comment:

From the plot above We can know that a school with a sample average more close to the global mean, i.e. posterior mean of μ , would have a less difference between its group mean θ_j and group sample average \bar{y}_j .

However, for those schools whose sample averages are away from the global mean μ , their group mean θ_j will be pulled away from \bar{y}_j towards μ , leading to a bigger difference between \bar{y}_j and θ_j .