STA 602. HW11

Yicheng Shen

16 November, 2022

[1] 8.1

(a) I expect the $Var(y_{i,j}|\mu,\tau^2)$ to be bigger. The variability of first sampling a group, then sampling a unit from within the group include both group level as well as within group variance, which is larger than only within-group variance just in a fixed group.

(b) $Cov(y_{i1,j}, y_{i2,j}|\theta_j, \sigma^2)$ should be zero because if knowing the group's mean and variance, the individuals within the group are i.i.d.

$Cov(y_{i1,j}, y_{i2,j}|\mu, \tau^2)$ would be positive. Now the θ_j and σ^2 are unknown, which means we only know how groups are different from each other, but we should expect observations from the same group j should be positively correlated.

(c) First we compute the first three quantities, given the group-level mean and variance:

$$\text{Within each group:} \quad \text{Var}(y_{i,j}|\theta_j, \sigma^2) = \sigma^2$$
$$\text{Var}(\bar{y}_{\cdot,j}|\theta_j, \sigma^2) = \sigma^2/n_j$$
$$\text{Cov}(y_{i_1,j}, y_{i_2,j}|\theta_j, \sigma^2) = E(y_{i_1,j}y_{i_2,j}) - E(y_{i_1,j})E(y_{i_2,j})$$
$$\text{i.i.d if given } \theta_j, \sigma^2 \quad = E(y_{i_1,j})E(y_{i_2,j}) - E(y_{i_1,j})E(y_{i_2,j}) = 0$$

Then we compute the next three, given the overall mean and variance and using the law of total variance: :

$$\text{Var}(y_{i,j}|\mu, \tau^2) = \text{Var}(E(y_{i,j}|\theta_j, \sigma^2)|\mu, \tau^2) + E(\text{Var}(y_{i,j}|\theta_j, \sigma^2)|\mu, \tau^2)$$
$$= \text{Var}(\theta_j|\mu, \tau^2) + E(\sigma^2|\mu, \tau^2) = \tau^2 + \sigma^2$$
$$\text{Var}(\bar{y}_{\cdot,j}|\mu, \tau^2) = \text{Var}(E(\bar{y}_{\cdot,j}|\theta_j, \sigma^2)|\mu, \tau^2) + E(\text{Var}(\bar{y}_{\cdot,j}|\theta_j, \sigma^2)|\mu, \tau^2)$$
$$= \text{Var}(\theta_j|\mu, \tau^2) + E(\frac{\sigma^2}{n_j}|\mu, \tau^2) = \tau^2 + (\sigma^2/n_j)$$
$$\text{Cov}(y_{i_1,j}, y_{i_2,j}|\mu, \tau^2) = E(\text{Cov}(y_{i_1,j}, y_{i_2,j}|\theta_j, \sigma^2)|\mu, \tau^2) + \text{Cov}(E(y_{i_1,j}|\theta_j, \sigma^2), E(y_{i_2,j}|\theta_j, \sigma^2))$$
$$= E(0|\mu, \tau^2) + \text{Cov}(E(y_{i_1,j}|\theta_j, \sigma^2), E(y_{i_2,j}|\theta_j, \sigma^2))$$
$$= \text{Cov}(\theta_j, \theta_j) = \tau^2$$

My guesses for the quantities in a and b are right.

(d) Basically, we want to show the full conditional of μ, overall mean, only depends on $\theta_1...\theta_m, \tau^2$.

To get the full conditional for μ, we need to marginalize it out of the joint distributions:

$$p(\mu \mid \boldsymbol{y}_1, \ldots, \boldsymbol{y}_m, \theta_1, \ldots, \theta_m, \sigma^2, \tau^2) = \frac{p(\mu, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_m, \theta_1, \ldots, \theta_m, \sigma^2, \tau^2)}{\int p(\mu, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_m, \theta_1, \ldots, \theta_m, \sigma^2, \tau^2) \, d\mu}$$

$$\text{Expand into priors and sampling model} = \frac{p(\mu)p(\tau^2)p(\sigma^2)p(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_m \mid \theta_1, \ldots, \theta_m, \sigma^2)p(\theta_1, \ldots, \theta_m \mid \mu, \tau^2)}{\int p(\mu)p(\tau^2)p(\sigma^2)p(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_m \mid \theta_1, \ldots, \theta_m, \sigma^2)p(\theta_1, \ldots, \theta_m \mid \mu, \tau^2) \, d\mu}$$

$$\text{Get constants out} = \frac{p(\mu)p(\tau^2)p(\sigma^2)p(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_m \mid \theta_1, \ldots, \theta_m, \sigma^2)p(\theta_1, \ldots, \theta_m \mid \mu, \tau^2)}{p(\tau^2)p(\sigma^2)p(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_m \mid \theta_1, \ldots, \theta_m, \sigma^2) \int p(\mu)p(\theta_1, \ldots, \theta_m \mid \mu, \tau^2) \, d\mu}$$

$$\text{Remove those do not depend on } \mu = \frac{p(\mu)p(\theta_1, \ldots, \theta_m \mid \mu, \tau^2)}{\int p(\mu)p(\theta_1, \ldots, \theta_m \mid \mu, \tau^2) \, d\mu}$$

$$\text{Bayes rule } P(A|B) = \frac{P(A)P(B|A)}{\int P(A)P(B|A)dA}$$

$$= p(\mu \mid \theta_1, \ldots, \theta_m, \tau^2)$$

This could be interpreted as: if given all group means $\theta_1, \ldots, \theta_m$ and the between group variance, τ^2 the distribution of overall mean, μ, does not depend on the with-in group variance σ^2 or the data $y_1 \ldots y_m$.

[2] 8.2

a. Here I assume the problem means $\mu \sim N(mean = 75, var = 100, sd = 10)$.

```
S = 10000
mu0 = 10
gamma0_2 = 100
nu0 = 2
sigma0_2 = 100

# initialization
n1 = n2 = 16
y1bar <- 75.2; y2bar <- 77.5
sy1 <- 7.3; sy2 <- 8.1
s2.pool <- ((n1-1)*sy1 + (n2)*sy2)/(n1+n2-2)

mu<- (y1bar+y2bar)/2
delta <- (y1bar-y2bar)/2
sigma2 <- s2.pool

THETA <- matrix(NA,nrow=S,ncol=3)
result <- matrix(NA, nrow = 20, ncol = 7)

set.seed(8848)
DELTA_0 <- c(-4,-2,0,2,4)
TAU_0 <- c(10, 50, 100, 500)

for (j in 1:5){
    delta0=DELTA_0[j]
for (k in 1:4){
    tau0_2=TAU_0[k]

for (i in 1:S){
  # update mu
```

```r
  y1.tilde <- y1bar - delta
  y2.tilde <- y2bar + delta
  gamman_2 = (1/gamma0_2+(n1+n2)/sigma2)^(-1)
  mu = rnorm(1,
             mean = gamman_2*(mu0/gamma0_2 + (mean(c(y1.tilde,y2.tilde))*(n1+n2))/sigma2 ),
             sd = sqrt(gamman_2))

  # update sigma2
  sigma2 = 1 / rgamma(1,
                   shape = (nu0 + n1 + n2) / 2,
                   rate = (nu0 * sigma0_2 + sy1^2*(n1-1) + sy2^2*(n2-1) +
                              n1 * (y1bar-mu-delta)^2 + n2*(y2bar-mu+delta)^2 ) / 2)

  # update delta
  y1.hat <- y1bar - mu
  y2.hat <- mu - y2bar
  taun_2 = (1/tau0_2+(n1+n2)/sigma2)^(-1)
  delta = rnorm(1,
             mean = taun_2*(delta0/tau0_2 + (mean(c(y1.hat, y2.hat))*(n1+n2))/sigma2 ),
             sd = sqrt(taun_2))

  THETA[i,] <- c(mu, delta, sigma2)
}
  n=4*(j-1)+k
  result[n,1]=delta0
  result[n,2]=tau0_2
  result[n,3]=mean(THETA[,2]<0)
  result[n,4:5]=quantile(THETA[,2],c(0.025,0.975))
  result[n,6]=(gamma0_2-tau0_2)/(gamma0_2+tau0_2)
  result[n,7]=cor(THETA[,1]+THETA[,2],THETA[,1]-THETA[,2])
}
}
colnames(result)=c("delta0","tau0_2","Pr(delta<0|Y)","2.5%","97.5%","Prior CORR","Post CORR")
knitr::kable(result)
```

delta0	tau0_2	Pr(delta<0|Y)	2.5%	97.5%	Prior CORR	Post CORR
-4	10	0.8995	-4.269096	0.8768264	0.8181818	0.1179072
-4	50	0.8116	-4.032660	1.6177327	0.3333333	0.0535327
-4	100	0.8034	-4.109361	1.6304537	0.0000000	0.0384597
-4	500	0.7934	-4.013868	1.7085617	-0.6666667	0.0381162
-2	10	0.8386	-3.874220	1.2583260	0.8181818	0.1253856
-2	50	0.7970	-4.073308	1.6767307	0.3333333	0.0263373
-2	100	0.7999	-4.008059	1.6413830	0.0000000	0.0335324
-2	500	0.7887	-4.040606	1.7033512	-0.6666667	0.0253436
0	10	0.7645	-3.533728	1.6321705	0.8181818	0.1218687
0	50	0.7767	-3.915173	1.8044740	0.3333333	0.0366395
0	100	0.7921	-3.964032	1.7083975	0.0000000	0.0433433
0	500	0.7876	-4.032699	1.7441424	-0.6666667	0.0062273
2	10	0.6790	-3.108309	2.0369814	0.8181818	0.1351237
2	50	0.7678	-3.805642	1.7749934	0.3333333	0.0574130
2	100	0.7828	-4.024533	1.7004375	0.0000000	0.0266031
2	500	0.7881	-3.925319	1.7422380	-0.6666667	0.0259188

delta0	tau0_2	Pr(delta<0|Y)	2.5%	97.5%	Prior CORR	Post CORR
4	10	0.5924	-2.858209	2.5080061	0.8181818	0.1094054
4	50	0.7533	-3.740377	1.8966698	0.3333333	0.0483243
4	100	0.7631	-3.855258	1.8645280	0.0000000	0.0358796
4	500	0.7807	-3.934465	1.7330692	-0.6666667	0.0411359

$$\rho_{\text{prior}} = \frac{Cov(\mu + \delta_0, \mu - \delta_0)}{\sqrt{Var(\mu + \delta_0)Var(\mu - \delta_0)}}$$

$$= \frac{\gamma_0 - \tau_0^2}{\sqrt{(\gamma_0 + \tau_0^2)(\gamma_0 + \tau_0^2)}} = \frac{\gamma_0 - \tau_0^2}{\gamma_0 + \tau_0^2}$$

b. $\theta_A < \theta_B$ means that δ must be negative.

If people are really confident with the belief that δ is negative, then the posterior is highly likely to give $\theta_A < \theta_B$. But if for the prior people are really sure δ is positive; or people are really unsure about the prior opinion, then the probability of $\theta_A < \theta_B$ will be getting lower.

[3] 8.3

a. First read in all data.

```
school <- lapply(1:8, function(i)
{
  data.frame(hour = read.table(paste0("http://www2.stat.duke.edu/~pdh10/FCBS/Exercises/school",
                                      i, ".dat")),
             school = i)
})
school <- do.call(rbind, school)
colnames(school) <- c("hour", "school")
```

The Gibbs sampler is done below.

```
set.seed(8848)
S <- 5000
mu0 <- 7
gamma0_2 <- 5
tau0_2 <- 10
eta0 <- 2
sigma0_2 <- 15
nu0 <- 2

m <- length(unique(school$school))
n <- sapply(1:m, function(i){nrow(school[school$school == i,])})

THETA <- matrix(NA, nrow = S, ncol = m)

SIGMA2 <- MU <- TAU2 <- c()
nun <- nu0 + nrow(school)
ybar <- sapply(1:m, function(i){mean(school[school$school == i, "hour"])})
```

```r
# initial
theta = ybar
mu = mean(theta)

for (i in 1:S)
{ # tau
  tau2 = 1 / rgamma(1, (eta0 + m) / 2, (eta0 * tau0_2 + sum((theta - mu)^2)) / 2)

  # mu
  vmu = 1 / (1 / gamma0_2 + m / tau2)
  mu = rnorm(1, vmu * (mu0 / gamma0_2 + m / tau2 * mean(theta)), sqrt(vmu))

  # sigma2
  ss = sum(sapply(1:m, function(i) {
    sum((school[school$school == i, "hour"] - theta[i])^2)
  }))
  sigma2 = 1 / rgamma(1, nun / 2, (nu0 * sigma0_2 + ss) / 2)

  # theta
  vtheta = 1 / (n / sigma2 + 1 / tau2)
  theta = rnorm(m, vtheta * (n / sigma2 * ybar + 1 / tau2 * mu), sqrt(vtheta))

  # save
  THETA[i,] = theta
  SIGMA2 = c(SIGMA2, sigma2)
  MU = c(MU, mu)
  TAU2 = c(TAU2, tau2)
}

colnames(THETA) = paste0("THETA", 1:m)
```
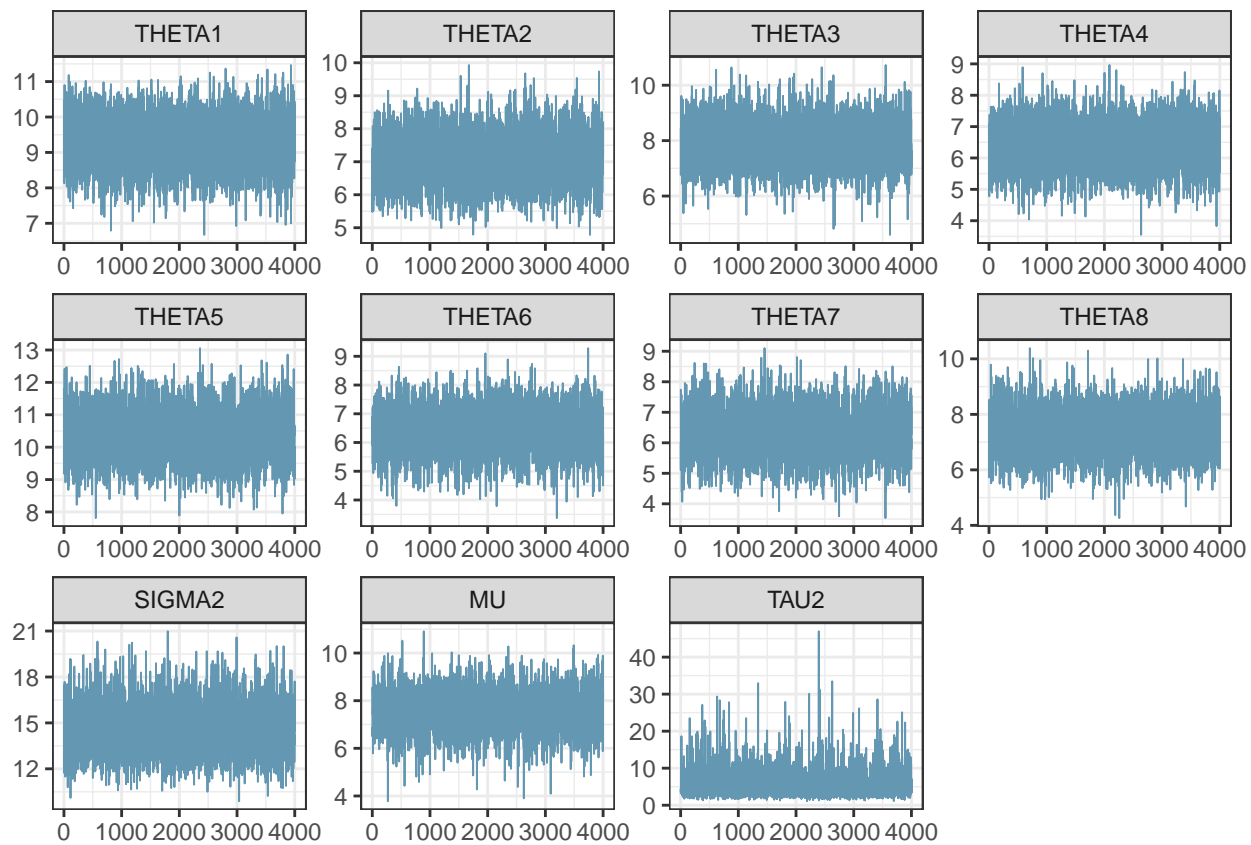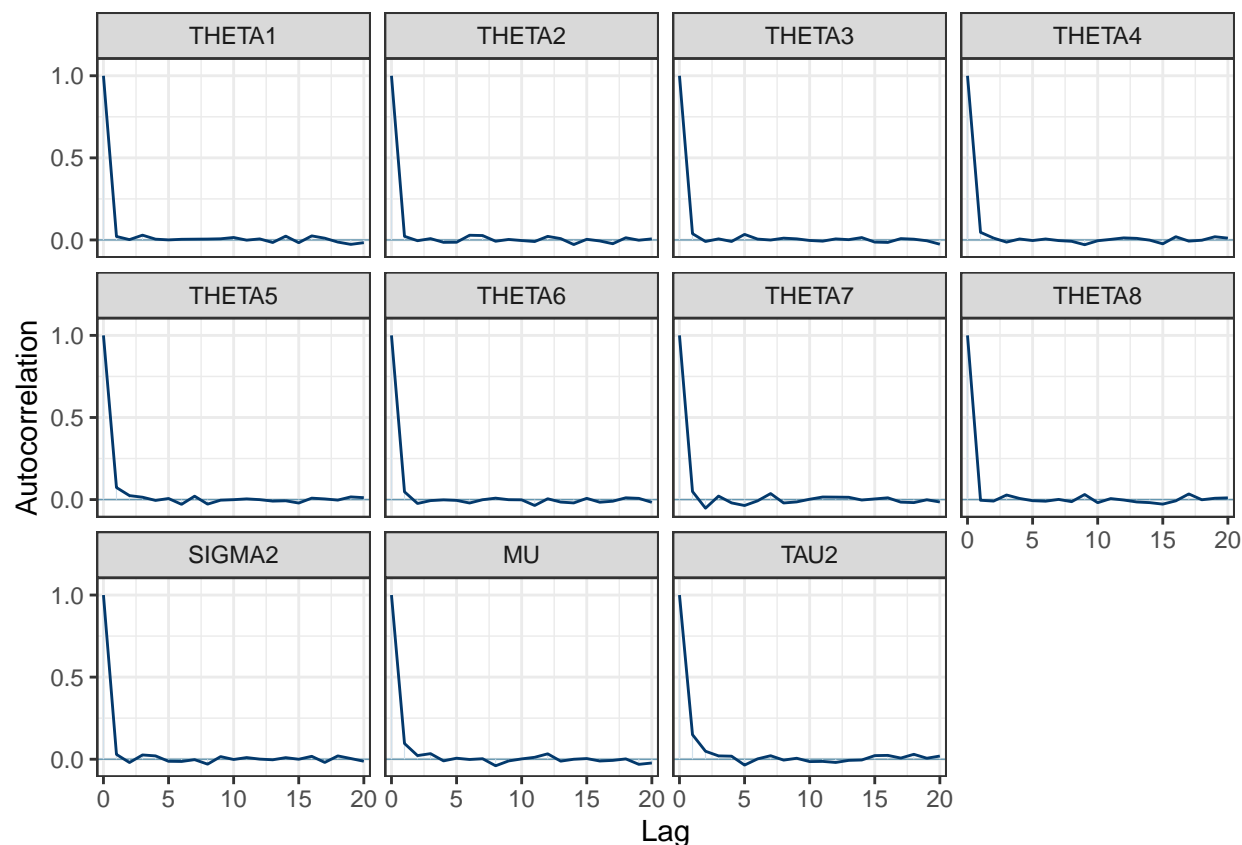
The MCMC diagnostics are done below, which suggest that the chain has converged well, maybe a bit concerning τ^2

```r
mcmc_trace(cbind(THETA,SIGMA2,MU,TAU2)[1000:S,])
```

```
mcmc_acf(cbind(THETA,SIGMA2,MU,TAU2)[1000:S,])
```

```r
effectiveSize(cbind(SIGMA2,MU,TAU2)[1000:S,])
```

```
##    SIGMA2       MU     TAU2
## 3725.341 3023.110 2915.720
```

The effective sample sizes are good enough since they are all over 1000.

b. The mean and 95% CI for the three parameters are shown below.

```r
t(apply(cbind(SIGMA2,MU,TAU2), MARGIN = 2, FUN = mean))
```

```
##        SIGMA2       MU     TAU2
## [1,] 14.4708 7.576328 5.516008
```

```r
t(apply(cbind(SIGMA2,MU,TAU2), MARGIN = 2, FUN = quantile, probs = c(0.025, 0.975)))
```

```
##             2.5%     97.5%
## SIGMA2 11.755500 17.863702
## MU      5.940992  9.146964
## TAU2    1.904357 15.021225
```

The density comparison of prior and posterior are shown below.

```
par(mfrow = c(2,2))
plot(seq(10, 22, by = 0.1), dinvgamma(seq(10, 22, by = 0.1), nu0 / 2, nu0 * sigma0_2 / 2),
     type = "l", col = 1, ylim = c(0, 0.27), ylab = "", main = "Density of SIGMA2")
lines(density(SIGMA2[1000:S]), col = 2)
legend(x = "topright", legend=c("Posterior","Prior"),
       col = c(2,1), lty = 1, cex = 0.7)
plot(seq(0, 12, by = 0.1), dnorm(seq(0, 12, by = 0.1), mean = mu0, sd = sqrt(gamma0_2)),
     type = "l", col = 1, ylim = c(0, 0.55), ylab = "", main = "Density of MU")
lines(density(MU[1000:S]), col = 2)
legend(x = "topright", legend=c("Posterior","Prior"),
       col = c(2,1), lty = 1, cex = 0.7)
plot(seq(0, 30, by = 0.1), dinvgamma(seq(0, 30, by = 0.1), eta0 / 2, eta0 * tau0_2 / 2),
     type = "l", col = 1, ylim = c(0, 0.2), ylab = "", main = "Density of TAU2")
lines(density(TAU2[1000:S]), col = 2)
legend(x = "topright", legend=c("Posterior","Prior"),
       col = c(2,1), lty = 1, cex = 0.7)
```

Density of SIGMA2

Density of MU

Density of TAU2
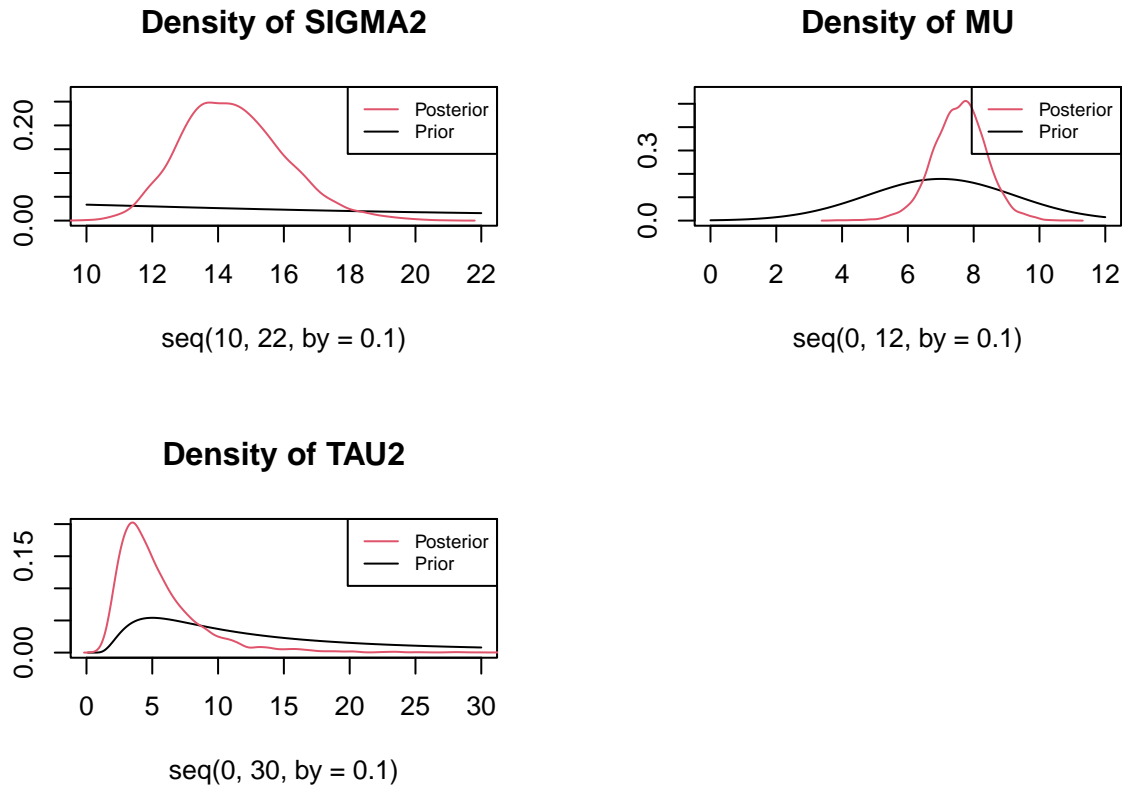
It seems that the prior for σ^2 is too diffused and the data strengths the posterior. The prior choice for μ is not bad, and the posterior only strengthens somewhere near it. The prior for τ^2 is also pretty diffused, but roughly having the right shape according to its posterior.

c. We know that τ^2 and σ^2 both have inverse gamma priors. And the posterior R can be obtained via calculating the posterior draws:

```
t2_prior = (1 / rgamma(1001, eta0 / 2, eta0 * tau0_2 / 2))
s2_prior = (1 / rgamma(1001, nu0 / 2, nu0 * sigma0_2 / 2))
plot(density(t2_prior/(t2_prior+s2_prior)), type = "l", ylab = "", main = "Density of R", ylim = c(0,4.4
lines(density(TAU2/ (TAU2+SIGMA2)), col = 2)
legend(x = "topright", legend=c("Posterior","Prior"),
       col = c(2,1), lty = 1, cex = 0.7)
```

Density of R

N = 1001   Bandwidth = 0.06446
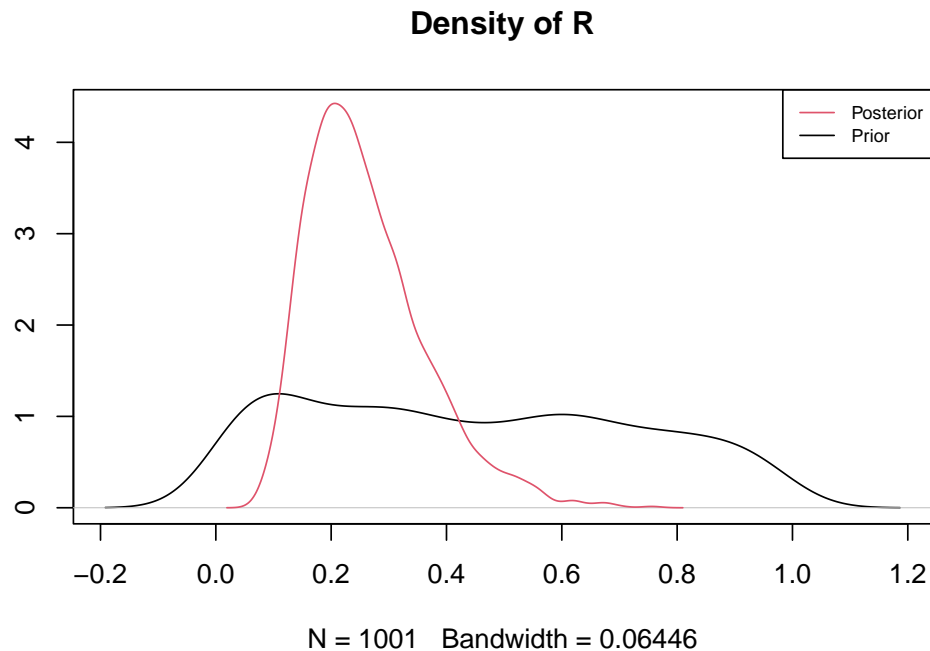
d. The posterior probability that $\theta_7 < \theta_6$ is 0.5204.

```
mean(THETA[, 7] < THETA[, 6])
```

```
## [1] 0.5204
```

The posterior probability that θ_7 is the smallest among all θ is 0.3202.

```
min_index<- c()
for (i in 1:nrow(THETA))
{ min_index[i] <- which.min(THETA[i,]) }
mean(min_index == 7)
```

```
## [1] 0.3202
```

e. The red line is the sample grand mean of 7.691278, the blue line is the overall posterior mean of
7.576328.

Overall in most groups the sample mean of the group dominates the posterior mean. But if the sample is at
the tail (particularly high or low), then the posterior mean for that group gets more shrinkage towards the
overall mean.

```
dat <- cbind(ybar, colMeans(THETA))
colnames(dat) <- c("sample averages", "posterior expectations")

dat %>% as.data.frame() %>% mutate(index = 1:8) %>%
ggplot(aes(x = `sample averages`, y = `posterior expectations`, label = index)) +
  geom_point() + geom_text(hjust=0, vjust=0, size = 4) +
  geom_abline(slope = 1, intercept = 0) +
  geom_hline(yintercept = mean(school$hour), color = "red") +
  geom_hline(yintercept = mean(MU), color = "blue")
```