Assignment 3

Yufei Yin

Exercise 1 (4 points):

Chapter 3, Exercise 3 from Bayesian Data Analysis, 3rd Edition (Gelman et al.), which is linked to from the course Canvas page.

Chapter 3, Exercise 3

Estimation from two independent experiments: an experiment was performed on the effects of magnetic fields on the flow of calcium out of chicken brains. Two groups of chickens were involved: a control group of 32 chickens and an exposed group of 36 chickens. One measurement was taken on each chicken, and the purpose of the experiment was to measure the average flow μ_c in untreated (control) chickens and the average flow μ_t in treated chickens. The 32 measurements on the control group had a sample mean of 1.013 and a sample standard deviation of 0.24. The 36 measurements on the treatment group had a sample mean of 1.173 and a sample standard deviation of 0.20.

(a)

Assuming the control measurements were taken at random from a normal distribution with mean μ_c and variance σ_c^2 , what is the posterior distribution of μ_c? Similarly, use the treatment group measurements to determine the marginal posterior distribution of μ_t. Assume a uniform prior distribution on (μ_c, μ_t, $\log\sigma_c$, $\log\sigma_t$).

joint prior distribution

$P(\mu_c, log\sigma_c, \mu_t, log\sigma_t) \propto 1$

joint posterior distribution

$P(\mu_c, log\sigma_c, \mu_t, log\sigma_t | y_c, y_t) \propto P(y_c, y_t | \mu_c, log\sigma_c, \mu_t, log\sigma_t)$

Because these two experiments are independent

$P(\mu_c, log\sigma_c, \mu_t, log\sigma_t | y_c, y_t) = P(\mu_c, log\sigma_c | y_c) P(\mu_t, log\sigma_t | y_t) = \prod_{i=1}^{n_c} N(y_{ci} | \mu_c, \sigma_c^2) \prod_{i=1}^{n_t} N(y_{ti} | \mu_t, \sigma_t^2)$

change of variable method

$\frac{dlog\sigma_c}{d\sigma_c^2} = \frac{d(\frac{1}{2}log\sigma_c^2)}{d\sigma_c^2} = \frac{1}{2\sigma_c^2}$

$P(\mu_c, \sigma_c^2 | y_c) = P(\mu_c, log\sigma_c | y_c) |detJ| \propto P(\mu_c, log\sigma_c | y_c) \cdot \frac{1}{\sigma_c^2} = \prod_{i=1}^{n} N(y_{ci} | \mu_c, \sigma_c^2) \cdot \frac{1}{\sigma_c^2}$

marginal posterior distribution

$P(\mu_c|y_c) \propto [1 + \frac{n(\mu-\bar{y})}{(n-1)s^2}]^{-\frac{n}{2}} \sim t_{n-1}(\bar{y}, \frac{s^2}{n})$

So

$P(\mu_c|y_c) \sim t_{32-1}(1.013, \frac{0.24^2}{32})$

similarly

$P(\mu_t|y_t) \sim t_{36-1}(1.173, \frac{0.20^2}{36})$

(b)

What is the posterior distribution for the difference, $\mu_t - \mu_c$? To get this, you may sample from the independent t distributions you obtained in part (a) above. Plot a histogram of your samples and give an approximate 95% posterior interval for $\mu_t - \mu_c$.

The problem of estimating two normal means with unknown ratio of variances is called the Behrems-Fisher problem.

$\frac{\mu - \bar{y}}{s/\sqrt{n}} | y \sim t_{n-1}(location = 0, scale = 1)$

```
# control
n.c <- 32; sd.c <- 0.24; ybar.c <- 1.013
set.seed(460)
mu.c <- ybar.c + (sd.c/sqrt(n.c)) * rt(n = 1000, df = n.c-1)
# treatment
n.t <- 36; sd.t <- 0.20; ybar.t <- 1.173
mu.t <- ybar.t + (sd.t/sqrt(n.t)) * rt(n = 1000, df = n.t-1)
```

```
# histogram
hist(x = mu.t-mu.c,
    breaks = 25,
    probability = TRUE,
    xlab = expression(paste(mu,"t - ",mu, "c")),
    main = expression(paste("Histogram of ",mu,"t - ",mu, "c")))
```


Histogram of μt − μc
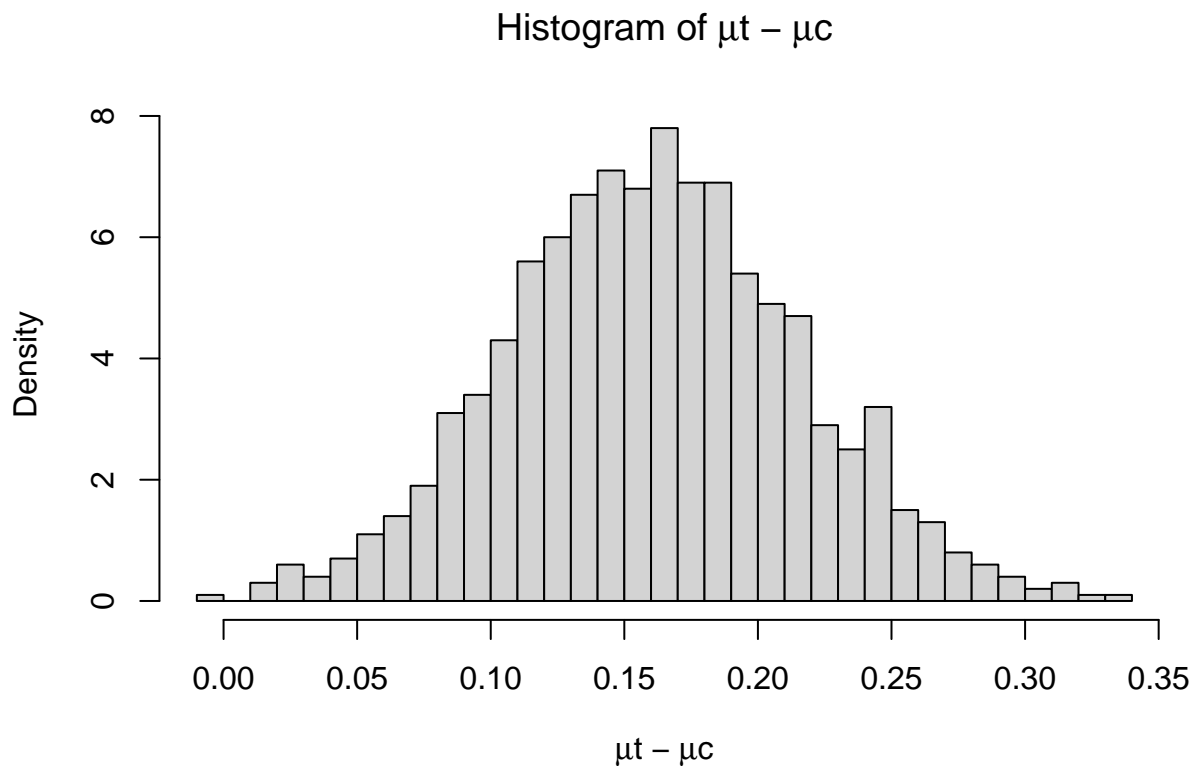
```r
# 95% posterior interval
lower <- sort(mu.t-mu.c)[1000*0.025]
upper <- sort(mu.t-mu.c)[1000*0.975]
round(c(lower, upper),4)
```

```
## [1] 0.0543 0.2682
```

The approximate 95% posterior interval is [0.0543, 0.2682].

Exercise 2 (6 points total):

For this exercise, we will use Monte Carlo simulation to redo the "Estimating the speed of light" example from class (and BDA3 Chapter 3). The data is contained in the file light.txt.

Part 1 (0.5 points):

The data in light.txt are Simon Newcomb's measurements of the amount of time required for light to travel a distance of 7442 meters. Make a histogram of these measurements, being sure to use a small enough binwidth to properly visualize the distribution. (E.g., in R you might use breaks=50 as an argument to hist().)

```
dataframe = read.table("light.txt",col.names = "speed of light")
hist(dataframe$speed.of.light,
     probability = TRUE,
     breaks = 50,
     main = "Histogram of Simon Newcomb's measurements\n for estimating the speed of light",
     xlab = "speed of light")
```
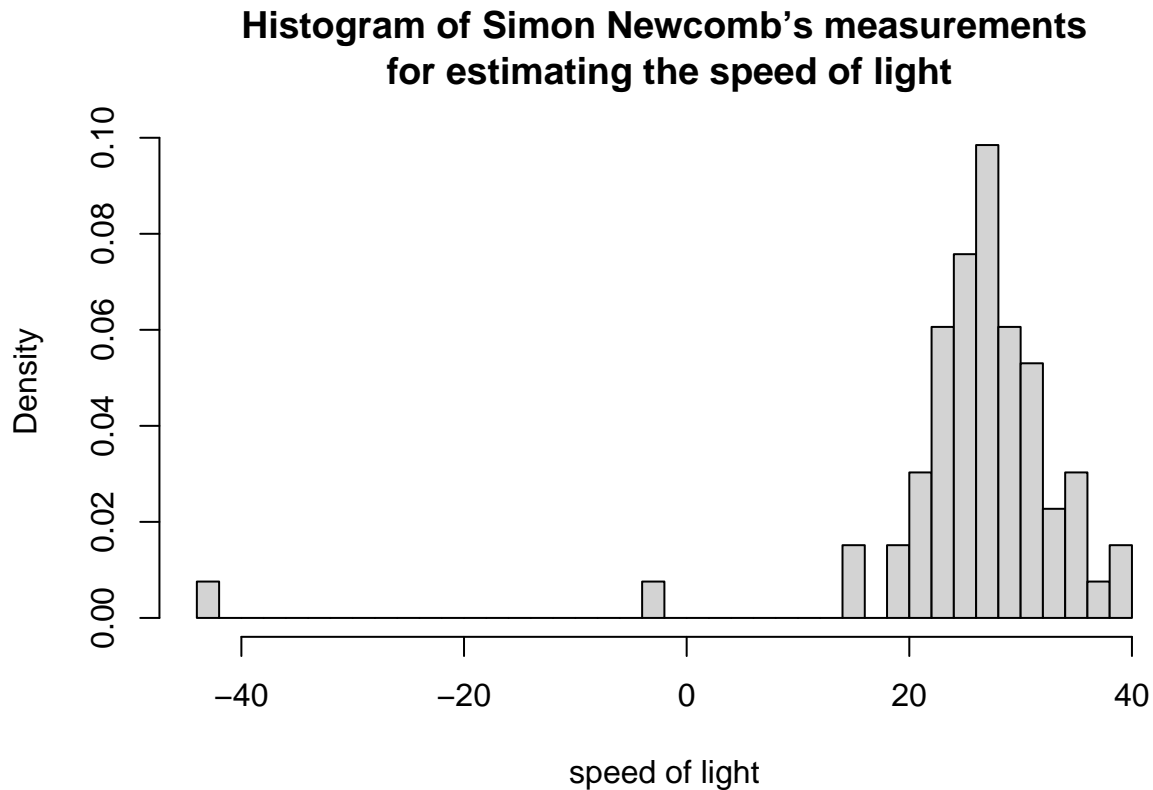
Part 2 (2.5 points):

We (inappropriately) model the data using a normal distribution with mean μ and variance σ^2, both unknown. (Include the outlying measurements for this exercise.) Use Monte Carlo to obtain 2000 draws from the joint posterior distribution of μ and σ^2. Use these draws to visualize:

(i) the joint posterior distribution of μ and σ^2(you may wish to use transparency to improve visualization),

(ii) the marginal posterior distribution of μ, and

(iii) the marginal posterior distribution of σ^2.

For this, also include your code to perform Monte Carlo simulation. Be sure to comment it sufficiently so that it is understandable, particularly if you use a language other than R.

Monte Carlo to draw samples from $P(\mu, \sigma^2 | y)$:

- For s = 1, ... ,S:

  1. Draw $(\sigma^2)^{(s)} \sim Inv - \chi^2(n-1, s^2)$

  2. Draw $\mu^{(s)} \sim N(\bar{y}, \frac{(\sigma^2)^{(s)}}{n})$

- The result of each iteration s is a sample from the posterior distribution: $(\mu^{(s)}, (\sigma^2)^{(s)}) \sim p(\mu, \sigma^2 | y)$

```r
# Use LaplacesDemon for rinvchisq, and set seed
library(LaplacesDemon)
set.seed(460)

# observed data
y <- dataframe$speed.of.light

# Create vectors to store posterior draws
sig2 <- NULL; mu <- NULL

# Draw 2000 MC samples from the joint posterior
for(ii in 1:2000)
{
    sig2[ii] <- rinvchisq(n=1, df=length(y)-1, scale=var(y))
    mu[ii] <- rnorm(n=1, mean=mean(y), sd=sqrt(sig2[ii]/length(y)))
}
```
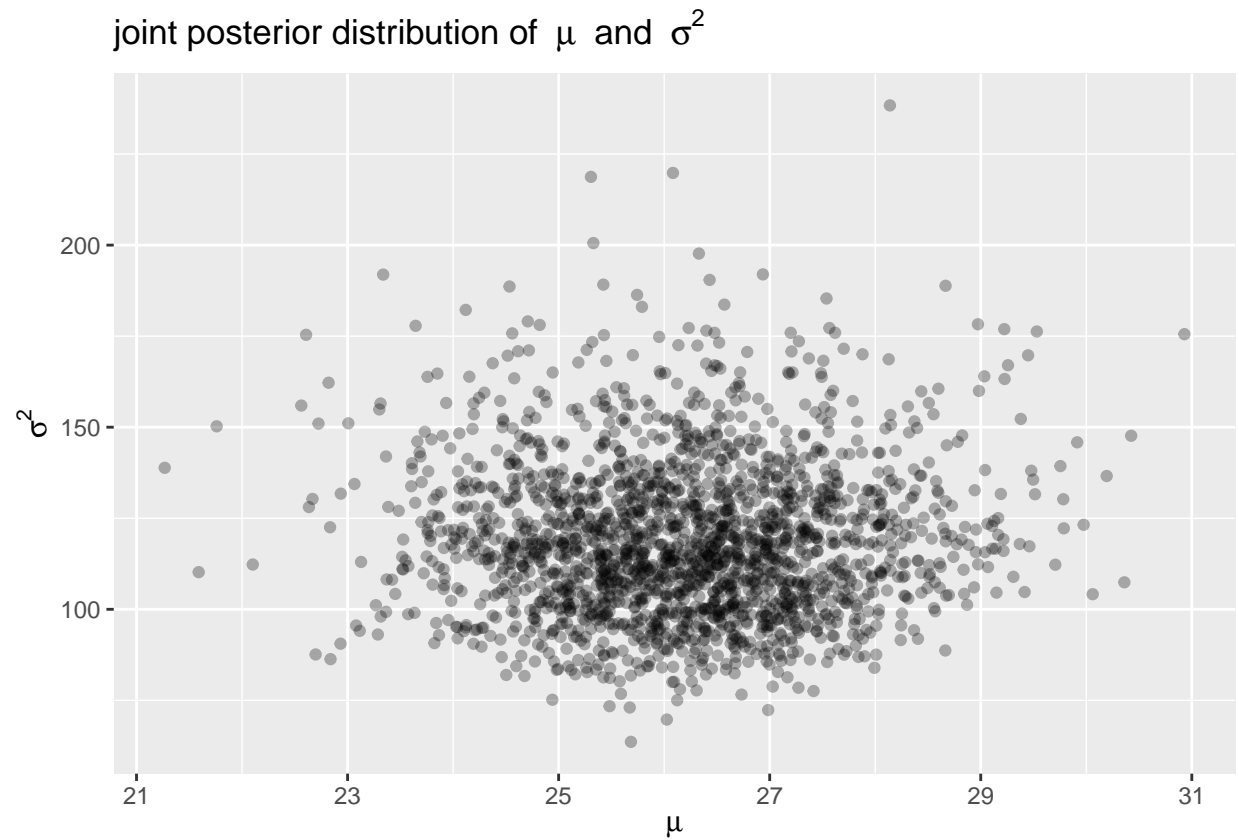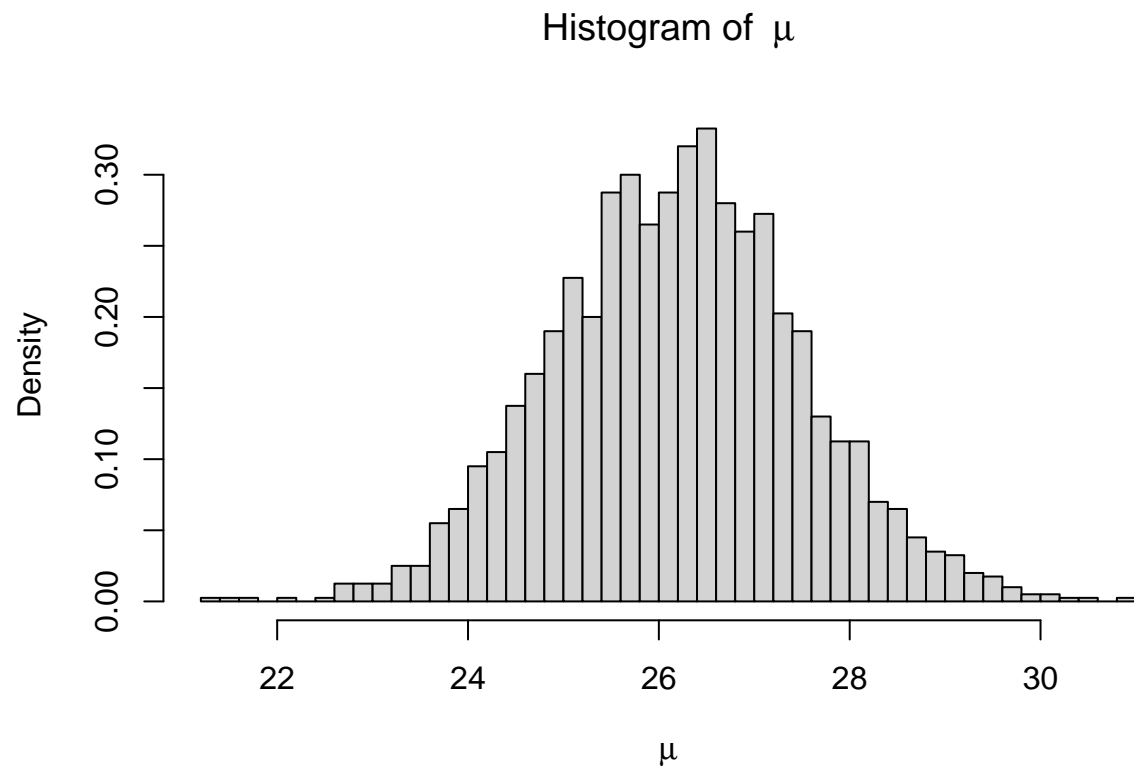
```
# plot
library(ggplot2)
data = data.frame(mu,sig2)


ggplot(data, aes(x = mu, y = sig2)) +
    geom_point(alpha=0.3) +
    labs(title = expression("joint posterior distribution of "~mu~" and "~sigma^2)) +
    xlab(expression(mu)) +
    ylab(expression(sigma^2))
```
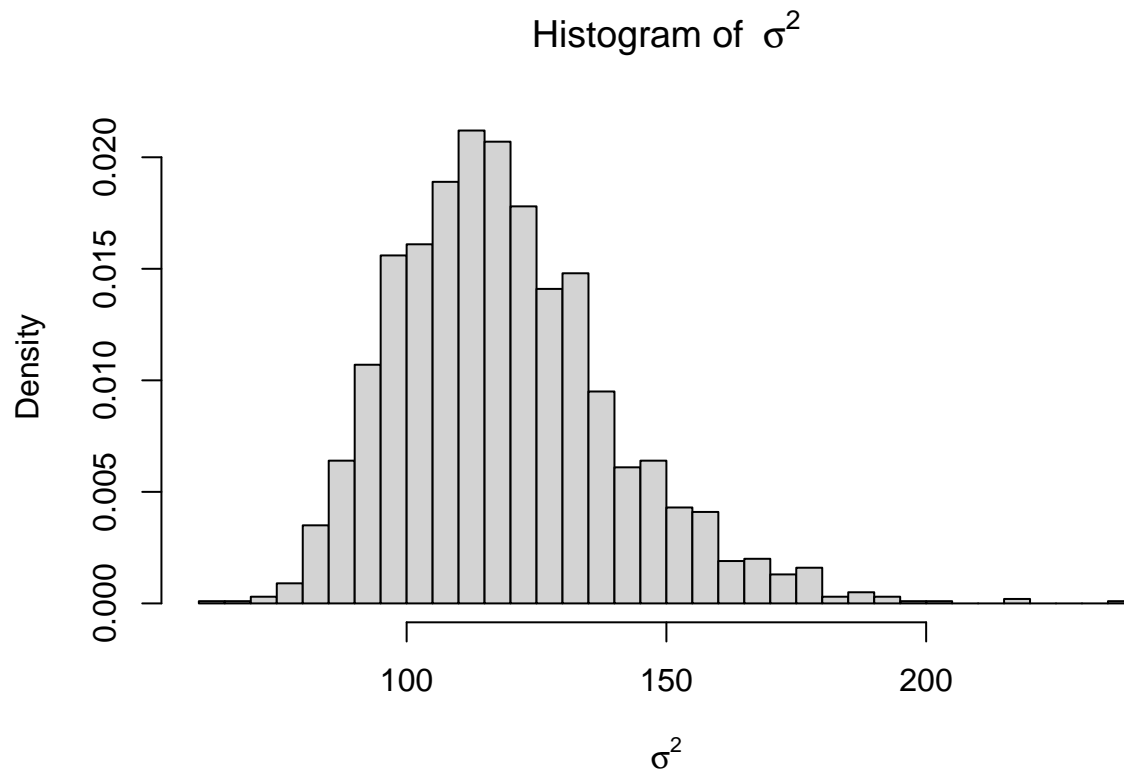
joint posterior distribution of μ and σ^2

```
hist(mu,prob=TRUE,
     breaks = 50,
     xlab = expression(mu),
     main = expression("Histogram of "~mu))
```

Histogram of μ

```
hist(sig2,prob=TRUE,
    breaks = 50,
    xlab = expression(sigma^2),
    main = expression("Histogram of "~sigma^2))
```

Histogram of σ^2

Part 3 (1 point):

Obtain a point estimate and equal-tailed 95% posterior interval for μ using the Monte Carlo draws. For the point estimate, use the Monte Carlo draws to estimate the posterior mean for μ: $E(\mu|y)$.

```r
# point estimate, posterior means for mu
mean(mu)
```

```
## [1] 26.19659
```

```r
# equal-tailed 95% posterior interval for mu
lower_mu <- sort(mu)[1000*0.025]
upper_mu <- sort(mu)[1000*0.975]
round(c(lower_mu,upper_mu),4)
```

```
## [1] 23.3145 26.1729
```

Part 4 (2 points).

Suppose we're interested in posterior inference for the precision: $\tau^2 = \frac{1}{\sigma^2}$. Obtain a Monte Carlo sample from the marginal posterior distribution of τ^2 and use a histogram to visualize the distribution. Then, give a point estimate and equal-tailed 95% posterior interval for τ^2.

```r
# we can obtain samples of tau2 directly by transforming the samples of sig2.
tau2 <- 1/sig2
```

```r
# histogram
hist(tau2,
     probability = TRUE,
     breaks = 50,
     xlab = expression(tau^2),
     main = expression("Histogram of "~tau^2))
```

Histogram of τ^2
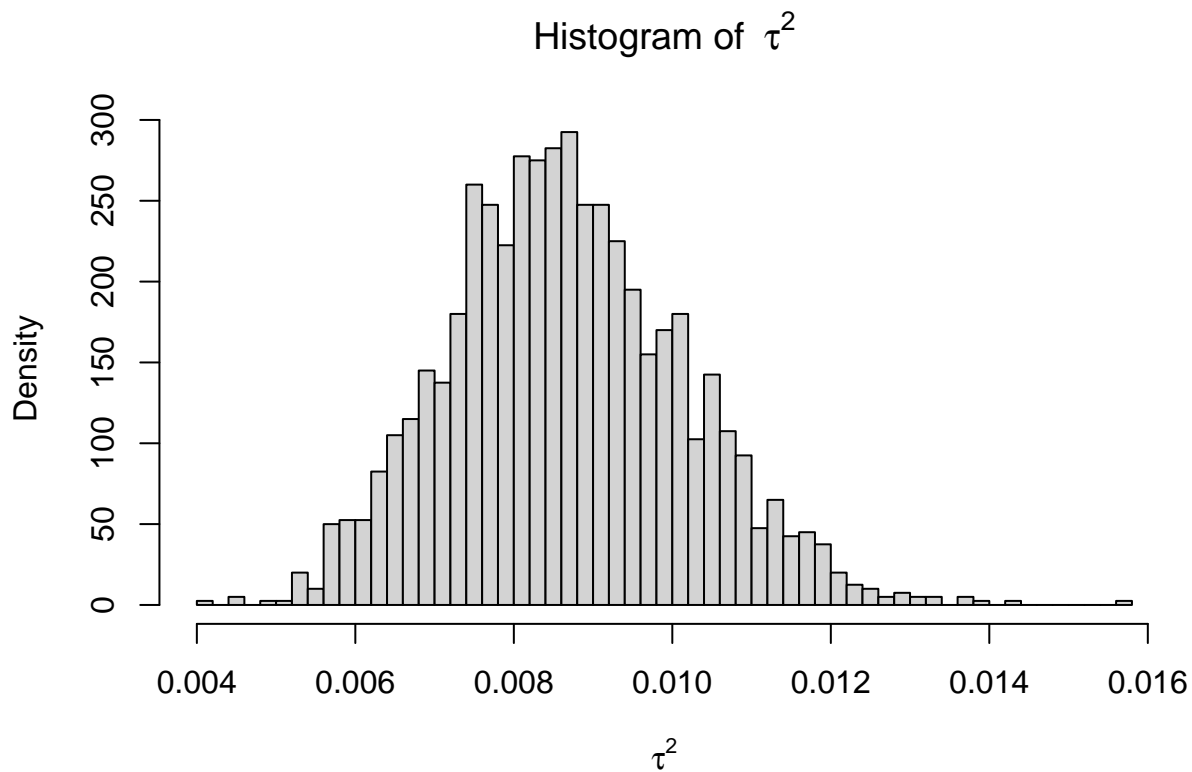

```r
# point estimate, posterior means for tau2
mean(tau2)
```

```
## [1] 0.008662175
```

```r
# equal-tailed 95% posterior interval for tau2
lower_tau2 <- sort(tau2)[1000*0.025]
upper_tau2 <- sort(tau2)[1000*0.975]
round(c(lower_tau2,upper_tau2),4)
```

```
## [1] 0.0057 0.0085
```