ESC_hw3

Eunsung CHOI

2019 11 11

####install packages before starting HW3

```
library(knitr)
library(ggplot2)
```

```
## Registered S3 methods overwritten by 'ggplot2':
##   method         from
##   [.quosures     rlang
##   c.quosures     rlang
##   print.quosures rlang
```

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(tidyr)
library(coda)
```

```
## Warning: package 'coda' was built under R version 3.6.1
```

```
library(MCMCpack)
```

```
## Warning: package 'MCMCpack' was built under R version 3.6.1
```

```
## Loading required package: MASS
```

```
##
## Attaching package: 'MASS'
```

```
## The following object is masked from 'package:dplyr':
##
##     select
```

```
## ##
## ## Markov Chain Monte Carlo Package (MCMCpack)
```

```
## ## Copyright (C) 2003-2019 Andrew D. Martin, Kevin M. Quinn, and Jong Hee Park
```

```
## ##
## ## Support provided by the U.S. National Science Foundation
```

```
## ## (Grants SES-0350646 and SES-0350613)
## ##
```

#HW3 ###8-3(a)

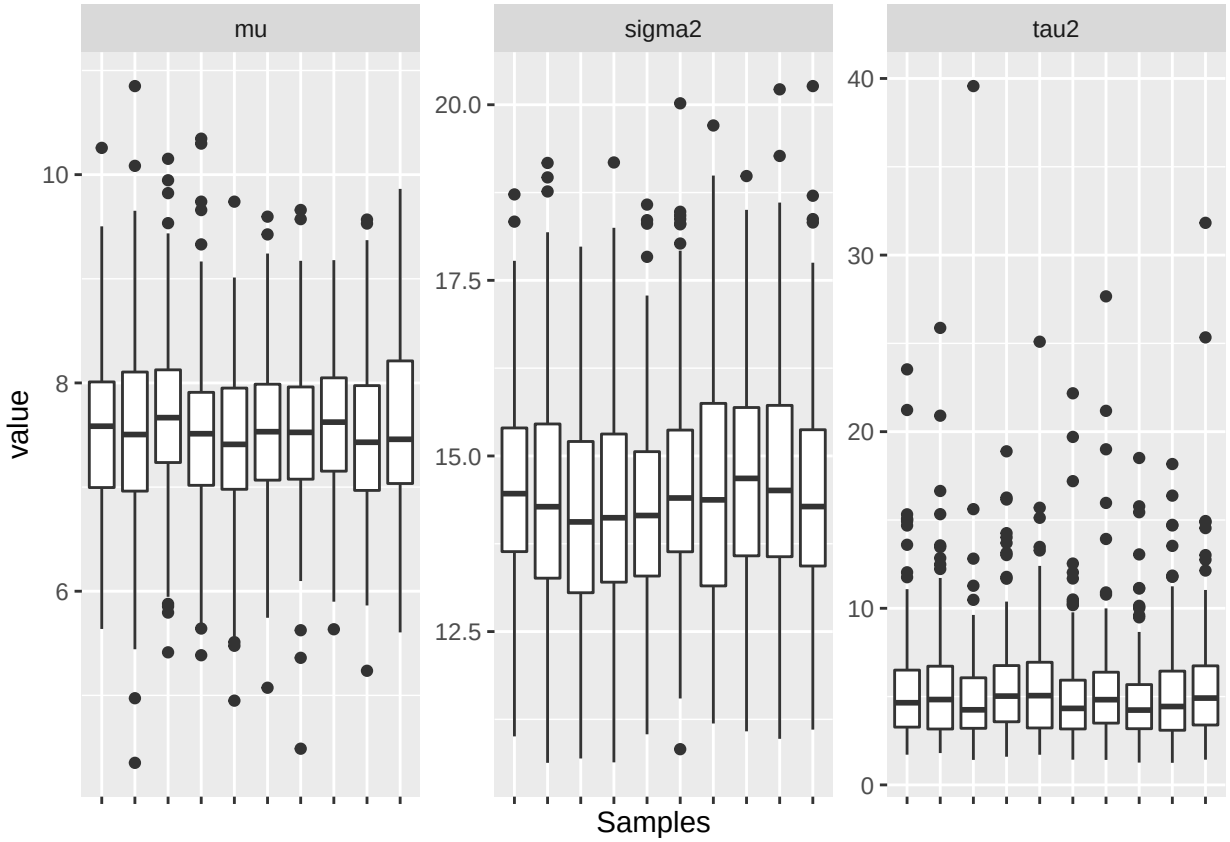
```r
schools <- lapply(1:8, function(i) {
raw_data <- paste0('http://www.stat.washington.edu/people/pdhoff/Book/Data/hwdata/school', i, '.dat') %>%
  url %>%
  read.table

data.frame(
  school = i,
  hours = raw_data[, 1] %>% as.numeric
  )
})

schools_data <- do.call(rbind, schools)
Y <- schools_data
head(Y)
```

```
##   school hours
## 1      1  2.11
## 2      1  9.75
## 3      1 13.88
## 4      1 11.30
## 5      1  8.93
## 6      1 15.66
```

```r
mu0 <- 7
g20 <- 5
t20 <- 10
eta0 <- 2
s20 <- 15
nu0 <- 2

m = length(unique(Y[, 1]))

n = sv = ybar = rep(NA, m)
for (j in 1:m) {
  Y_j <- Y[Y[, 1] == j, 2]
  ybar[j] <- mean(Y_j)
  sv[j] <- var(Y_j)
  n[j] <- length(Y_j)
}

theta <- ybar
sigma2 <- mean(sv)
mu <- mean(theta)
tau2 <- var(theta)

S <- 1500
THETA <- matrix(nrow = S, ncol = m)

SMT <- matrix(nrow = S, ncol = 3)
colnames(SMT) <- c('sigma2', 'mu', 'tau2')
for (s in 1:S) {

  for (j in 1:m) {
    vtheta = 1 / (n[j] / sigma2 + 1 / tau2)
```

```
    etheta = vtheta * (ybar[j] * n[j] / sigma2 + mu / tau2)
    theta[j] = rnorm(1, etheta, sqrt(vtheta))
  }


  nun = nu0 + sum(n) # TODO: Could cache this
  ss = nu0 * s20

  for (j in 1:m) {
    ss <- ss + sum((Y[Y[, 1] == j, 2] - theta[j])^2)
  }
  sigma2 <- 1 / rgamma(1, nun / 2, ss / 2)

  vmu <- 1 / (m / tau2 + 1 /g20)
  emu <- vmu * (m * mean(theta) / tau2 + mu0 / g20)
  mu <- rnorm(1, emu, sqrt(vmu))

  etam <- eta0 + m
  ss <- eta0 * t20 + sum((theta - mu)^2)
  tau2 <- 1 / rgamma(1, etam / 2, ss / 2)

  THETA[s, ] <- theta
  SMT[s, ] <- c(sigma2, mu, tau2)
}
```

```
smt_df <- data.frame(SMT)
colnames(smt_df) <- c('sigma2', 'mu', 'tau2')
smt_df$s <- 1:S
cut <- 10
smt_df <- smt_df %>%
  tbl_df %>%
  mutate(scut = cut(s, breaks = cut)) %>%
  gather('variable', 'value', sigma2:tau2)
ggplot(smt_df, aes(x=scut, y=value))+
  facet_wrap(~ variable, scales = 'free_y')+
  geom_boxplot()+
  theme(axis.text.x = element_blank())+
  xlab('Samples')
```




```
effectiveSize(SMT[, 1])
```

```
## var1
## 1500
```

```
effectiveSize(SMT[, 2])
```

```
##     var1
## 1334.834
```

```
effectiveSize(SMT[, 3])
```

```
##    var1
## 1120.48
```

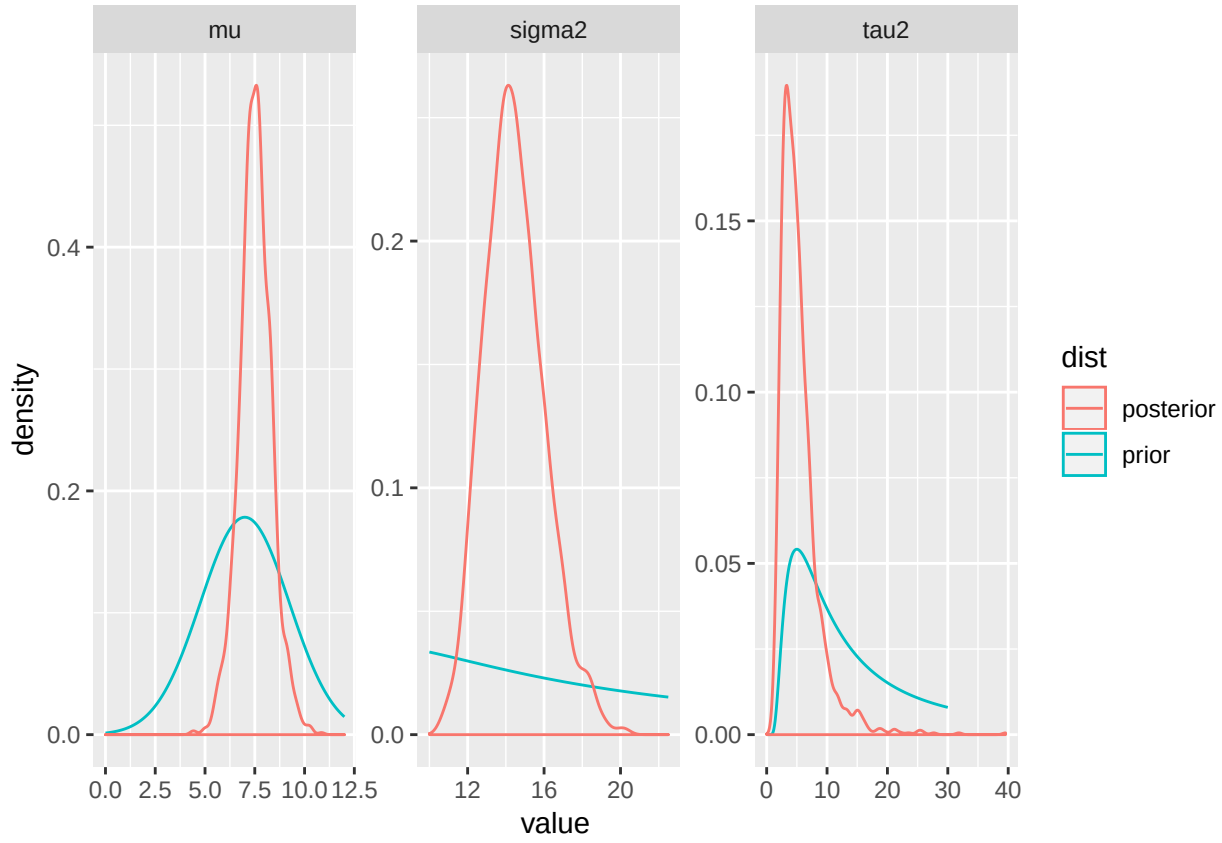
###8-3(b)

```
t(apply(SMT, MARGIN = 2, FUN = quantile, probs = c(0.025, 0.5, 0.975)))
```

```
##             2.5%       50%     97.5%
## sigma2 11.766842 14.342422 17.977436
## mu      5.877219  7.524336  9.219914
## tau2    1.885579  4.595894 14.405324
```

```
sigma2_prior <- data.frame(
value = seq(10, 22.5, by = 0.1),
density = dinvgamma(seq(10, 22.5, by = 0.1), nu0 / 2, nu0 * s20 / 2),
variable = 'sigma2'
)
tau2_prior <- data.frame(
```

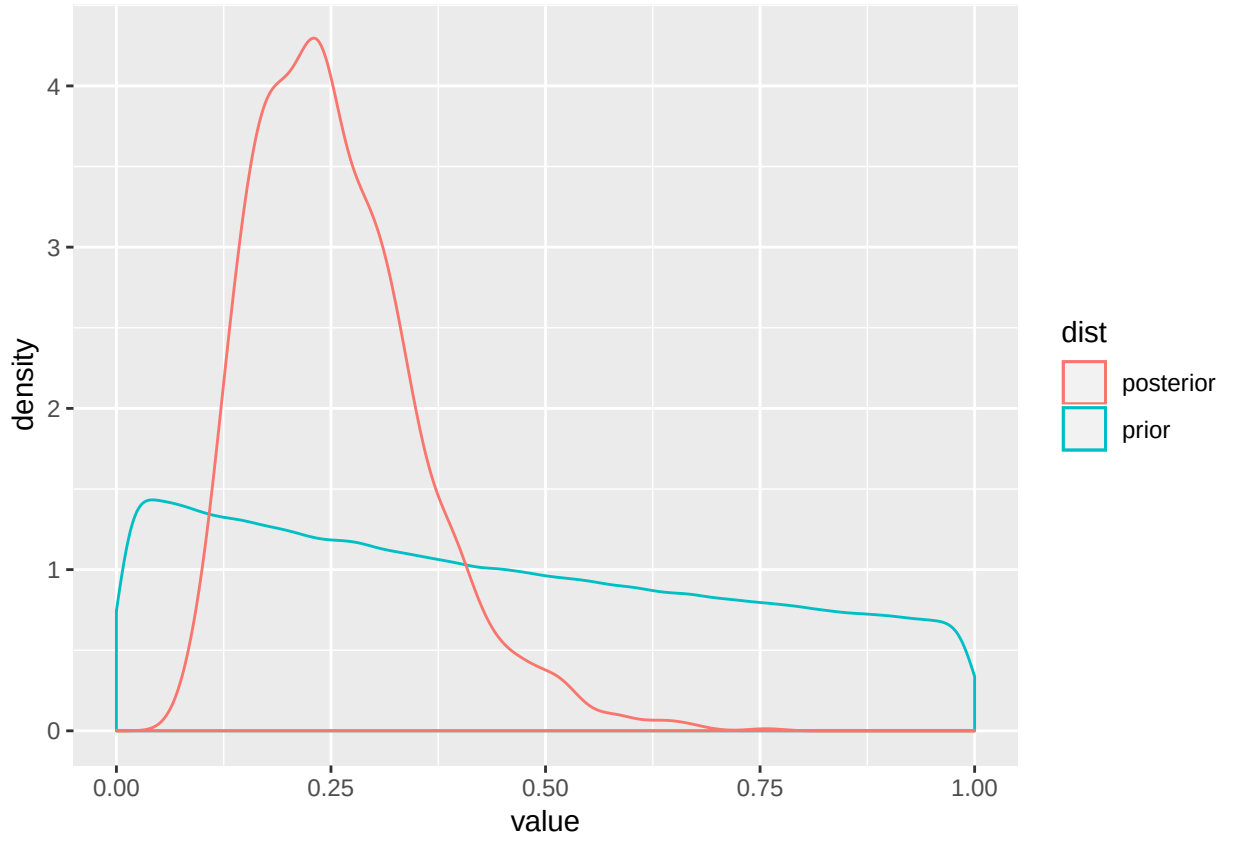
```
value = seq(0, 30, by = 0.1),
density = dinvgamma(seq(0, 30, by = 0.1), eta0 / 2, eta0 * t20 / 2),
variable = 'tau2'
)
mu_prior <- data.frame(
value = seq(0, 12, by = 0.1),
density = dnorm(seq(0, 12, by = 0.1), mu0, sqrt(g20)),
variable = 'mu'
)
priors <- rbind(sigma2_prior, tau2_prior, mu_prior)
priors$dist <- 'prior'
smt_df$dist <- 'posterior'
ggplot(priors, aes(x=value, y=density, color=dist))+
geom_line()+
geom_density(data=smt_df, mapping = aes(x=value, y=..density..))+
facet_wrap(~variable, scales = 'free')
```



###8-3(c)

```
t20_prior <- (1/rgamma(1e6, eta0/2, eta0*t20/2))
s20_prior <- (1/rgamma(1e6, nu0/2, nu0*s20/2))
R_prior <- data.frame(
value = (t20_prior)/(t20_prior+s20_prior),
dist = 'prior'
)
R_post <- data.frame(
```

```
value = SMT[,'tau2']/(SMT[,'tau2'] + SMT[,'sigma2']),
dist = 'posterior'
)
ggplot(R_prior, aes(x=value, y=..density.., color=dist))+
geom_density(data=R_prior)+
geom_density(data=R_post)
```




```
mean(R_post$value)
```

```
## [1] 0.257322
```

###8-3(d)

```
theta7_lt_6 <- THETA[, 7] < THETA[, 6]
mean(theta7_lt_6)
```

```
## [1] 0.5093333
```

```
theta7_smallest <- (THETA[, 7] < THETA[, -7]) %>%
  apply(MARGIN = 1, FUN = all)
mean(theta7_smallest)
```

```
## [1] 0.3193333
```

###8-3(e)

```
relationship = data.frame(
  sample_average = ybar,
  post_exp = colMeans(THETA),
```

```
  school = 1:length(ybar)
)

ggplot(relationship, aes(x = sample_average, y = post_exp, label = school)) +
  geom_text() +
  geom_abline(slope = 1, intercept = 0) +
  geom_hline(yintercept = mean(schools_data[, 'hours']), lty = 2) +
  annotate('text', x = 10, y = 7.9, label = paste0("Pooled sample mean ", round(mean(schools_data[, 'hou
  geom_hline(yintercept = mean(SMT[, 'mu']), color = 'blue') +
  annotate('text', x = 10, y = 7.4, label = paste0("Posterior exp.mu ", round(mean(SMT[, 'mu']), 2)), co
```