Homework 5a

homework 5a

Exercise 5.3 of BDA

```
library(rstan)
```

```
## Loading required package: ggplot2
```

```
## Loading required package: StanHeaders
```

```
## rstan (Version 2.17.4, GitRev: 2e1f913d3ca3)
```

```
## For execution on a local, multicore CPU with excess RAM we recommend calling
## options(mc.cores = parallel::detectCores()).
## To avoid recompilation of unchanged Stan programs, we recommend calling
## rstan_options(auto_write = TRUE)
```

```
rstan_options(auto_write = TRUE)
options(mc.cores = parallel::detectCores())
```

The model is:

$$y_i \sim N(\theta_j, \sigma_j)$$

$$\theta_j \sim N(\mu, \tau)$$

$$p(\mu, \tau) \propto 1$$

Thus, we can get that:

$$\frac{\theta_j - \mu}{\tau} = \eta \sim N(0, 1)$$

(a)

```
# input the data
y <- c(28,8,-3,7,-1,1,18,12)
s <- c(15,10,16,11,9,11,10,18)
J <- 8
school_dat <- list(J=J,y=y,sigma=s)
fit <- stan('homework5a.stan',data=school_dat)
```

```
## Warning in readLines(file, warn = TRUE): incomplete final line found on '/
## Users/yi/Desktop/study/books and classes/columbia/bayesian-data-analysis/
## homework/homework 8/homework5a.stan'
```

```
## Warning: There were 2 divergent transitions after warmup. Increasing adapt_delta a
bove 0.8 may help. See
## http://mc-stan.org/misc/warnings.html#divergent-transitions-after-warmup
```

```
## Warning: Examine the pairs() plot to diagnose sampling problems
```

```
print(fit,digits=1)
```

```
## Inference for Stan model: homework5a.
## 4 chains, each with iter=2000; warmup=1000; thin=1;
## post-warmup draws per chain=1000, total post-warmup draws=4000.
##
##           mean se_mean  sd  2.5%  25%  50%  75% 97.5% n_eff Rhat
## mu        8.0     0.1 4.9  -1.3  4.7  8.0 11.1  17.8  1602    1
## tau       6.8     0.1 5.8   0.2  2.5  5.4  9.4  21.7  1575    1
## eta[1]    0.4     0.0 0.9  -1.5 -0.2  0.4  1.0   2.1  4000    1
## eta[2]    0.0     0.0 0.9  -1.7 -0.6  0.0  0.5   1.8  4000    1
## eta[3]   -0.2     0.0 0.9  -2.1 -0.8 -0.2  0.4   1.6  2924    1
## eta[4]    0.0     0.0 0.9  -1.7 -0.6  0.0  0.5   1.7  4000    1
## eta[5]   -0.4     0.0 0.8  -2.0 -0.9 -0.4  0.2   1.3  3416    1
## eta[6]   -0.2     0.0 0.9  -1.8 -0.8 -0.2  0.3   1.5  3543    1
## eta[7]    0.4     0.0 0.9  -1.4 -0.2  0.4  0.9   2.1  4000    1
## eta[8]    0.0     0.0 0.9  -1.8 -0.6  0.0  0.7   1.9  3144    1
## theta[1] 11.8     0.2 8.6  -1.4  6.1 10.4 15.9  33.2  2924    1
## theta[2]  7.8     0.1 6.4  -5.3  3.9  7.7 11.5  20.7  4000    1
## theta[3]  6.0     0.2 7.8 -12.5  2.0  6.5 10.7  20.2  2380    1
## theta[4]  7.8     0.1 6.4  -5.3  4.0  7.8 11.6  20.5  4000    1
## theta[5]  5.1     0.1 6.3  -9.4  1.4  5.7  9.4  16.0  4000    1
## theta[6]  6.1     0.1 6.6  -8.7  2.5  6.6 10.2  18.3  4000    1
## theta[7] 10.8     0.1 6.7  -0.4  6.3 10.2 14.7  25.8  4000    1
## theta[8]  8.6     0.1 8.1  -6.5  4.1  8.2 12.7  26.4  2982    1
## lp__     -4.7     0.1 2.6 -10.3 -6.3 -4.5 -2.9  -0.2  1286    1
##
## Samples were drawn using NUTS(diag_e) at Tue Oct  2 16:01:45 2018.
## For each parameter, n_eff is a crude measure of effective sample size,
## and Rhat is the potential scale reduction factor on split chains (at
## convergence, Rhat=1).
```

stan has already simulate 4000 thetas for us. we can directly use these.

```r
# probability of best
prob_best <- c()
for (i in 1:8){
  prob_best <- c(prob_best,mean(extract(fit)$theta[,i]==apply(extract(fit)$theta,1,ma
x)))
}

result <- data.frame('school'=letters[1:8],'prob_best'=prob_best)

par_best <- matrix(NA,8,8)
for (i in 1:8){
  for (j in 1:8){
      par_best[i,j] <- mean(extract(fit)$theta[,i] - extract(fit)$theta[,j] > 0)
  }
}


for (i in 1:8){
  result[,i+2] <- par_best[,i]
}
colnames(result) <- c('school','prob_best',letters[1:8])
is.num <- sapply(result, is.numeric)
result[is.num] <- lapply(result[is.num], round, 3)
print(result,digits = 2)
```

```
##   school prob_best    a    b    c    d    e    f    g    h
## 1      a     0.264 0.00 0.64 0.69 0.65 0.75 0.70 0.52 0.61
## 2      b     0.104 0.36 0.00 0.57 0.50 0.62 0.58 0.36 0.48
## 3      c     0.078 0.31 0.43 0.00 0.43 0.54 0.50 0.30 0.42
## 4      d     0.098 0.35 0.50 0.57 0.00 0.62 0.57 0.36 0.48
## 5      e     0.045 0.25 0.38 0.46 0.38 0.00 0.45 0.26 0.36
## 6      f     0.065 0.30 0.42 0.50 0.43 0.55 0.00 0.30 0.41
## 7      g     0.201 0.48 0.64 0.70 0.64 0.74 0.70 0.00 0.60
## 8      h     0.145 0.39 0.52 0.58 0.52 0.64 0.59 0.40 0.00
```

(b)

if τ is infinite, the school effect θ_j are independent in their posterior distribution $\theta_j|y \sim N(y_j, \sigma_j^2)$. Thus,we have

$$P(\theta_i > \theta_j|y) = \Phi(\frac{y_i - y_j}{\sqrt{\sigma_i^2 + \sigma_j^2}})$$

And the best probability is:

$$P(\theta_j = max) = \int \prod_{i \neq j} \Phi(\frac{\theta_i - y_j}{\sigma_j}) \phi(\theta_i | y_i, \sigma_i)) d\theta_i$$

Since the integral is not easy to compute, I choose to esimate this probability through simulation

```
par_best <- matrix(NA,8,8)
for (i in 1:8){
  for (j in 1:8){
      this <- pnorm((y[i]-y[j])/sqrt(s[i]^2 + s[j] ^2))
      if (this == 0.5){
         par_best[i,j] =0
      }else{
         par_best[i,j] = this
      }
  }
}

## simulate the theta
thetas <- matrix(NA,ncol = 8,nrow = 10000)
for (i in 1:1000){
  for (j in 1:8){
    thetas[i,j] <- rnorm(1,mean = y[j],sd=s[j])
  }
}
prob_best <- c()
for (i in 1:8){
  prob_best <- c(prob_best, mean(thetas[,i]==apply(thetas,1,max),na.rm = T))
}
result <- data.frame('school'=letters[1:8],'prob_best'=prob_best)
for (i in 1:8){
  result[,i+2] <- par_best[,i]
}
colnames(result) <- c('school','prob_best',letters[1:8])
is.num <- sapply(result, is.numeric)
result[is.num] <- lapply(result[is.num], round, 3)
print(result,digits = 2)
```

```
##   school prob_best     a    b    c    d    e    f     g    h
## 1      a     0.536 0.000 0.87 0.92 0.87 0.95 0.93 0.710 0.75
## 2      b     0.034 0.134 0.00 0.72 0.53 0.75 0.68 0.240 0.42
## 3      c     0.029 0.079 0.28 0.00 0.30 0.46 0.42 0.133 0.27
## 4      d     0.037 0.129 0.47 0.70 0.00 0.71 0.65 0.230 0.41
## 5      e     0.004 0.049 0.25 0.54 0.29 0.00 0.44 0.079 0.26
## 6      f     0.014 0.073 0.32 0.58 0.35 0.56 0.00 0.126 0.30
## 7      g     0.176 0.290 0.76 0.87 0.77 0.92 0.87 0.000 0.61
## 8      h     0.170 0.247 0.58 0.73 0.59 0.74 0.70 0.385 0.00
```

(c)

In accordance with the result in BDA 5.5, as the value of τ increasing the difference between differen school becoming more obvious. As we can see the probability of A school to be the best school increase from 0.27 to 0.52. At the same time, most of pairwise probability tends to be larger. This is because, for a larger τ, different school could have a bigger or smaller value of θ_j . In this way, the mean of estimated treatment effect can be more different from each other.

(d)

If the value of $\tau = 0$. Clearly, $\theta_j = \mu$ for all schools. Thus, difference between different schools are 0 and the data's variance is only comes from the variance within the group(school). Thus, every shcool have the equal probability to be the best and on one is more likely to be better than the other,