Differences Between Schools Bayesian Analysis

Rory Quinlan

```
library(kableExtra)
library(dplyr)
# Read Data
s1 <- scan("school1.dat")
s2 <- scan("school2.dat")
s3 <- scan("school3.dat")
```

Posterior Means

School 1

```
# Using monte carlo simulation for all three schools

# Mu0, var0, k0, v0 given
mu0 <- 5
var0 <- 4
k0 <- 1
v0 <- 2
#posterior mean for s1
n1 <- length(s1)
y_bar1 <- mean(s1)
var1 <- var(s1)
mu_1 <- (k0 * mu0 + n1 * y_bar1)/(k0 + n1)
```

School 2

```
#posterior mean for s2
n2 <- length(s2)
y_bar2 <- mean(s2)
var2 <- var(s2)
mu_2 <- (k0 * mu0 + n2 * y_bar2)/(k0 + n2)
cat("Posterior mean for school 2 is:", mu_2)
```

```
## Posterior mean for school 2 is: 6.94875
```

School 3

```
#posterior mean for s3
n3 <- length(s3)
y_bar3 <- mean(s3)
var3 <- var(s3)
mu_3 <- (k0 * mu0 + n3 * y_bar3)/(k0 + n3)
cat("Posterior mean for school 3 is:", mu_3)
```

```
## Posterior mean for school 3 is: 7.812381
```

Results of Posterioir Means

```
Schools<- c("School 1", "School 2", "School 3")
Posterior_Means<- c(mu_1,mu_2,mu_3)

dfm<-as.data.frame(cbind(Schools,Posterior_Means))

dfm %>% kbl %>% kable_styling(full_width=F)
```

Schools	Posterior_Means
School 1	9.29230769230769
School 2	6.94875
School 3	7.81238095238095

95% Confidence Intervals

School 1

```
# 95% CI for s1 theta (mean)
set.seed(100)
vn1 <- v0 + n1
kn1 <- k0 + n1
s1n1 <- (1/vn1) * (v0 * var0 + (n1-1) * var1 + ((k0 * n1 )/kn1) * (y_bar1 - mu0)^2 )
s1_postsample <- 1/rgamma(10000, vn1/2, vn1 * s1n1 / 2)
theta1_postsample <- rnorm(10000, mu_1, sqrt(s1_postsample/(n1 + k0)))
```

```
# 95% CI for s1 sd
sd1_postsample <- sqrt(s1_postsample)
```

School 2

```
# 95% CI for s2 theta (mean)
set.seed(100)
vn2 <- v0 + n2
kn2 <- k0 + n2
s2n2 <- (1/vn2) * (v0 * var0 + (n2-1) * var2 + ((k0 * n2 )/kn2) * (y_bar2 - mu0)^2 )
s2_postsample <- 1/rgamma(10000, vn2/2, vn2 * s2n2 / 2)
theta2_postsample <- rnorm(10000, mu_2, sqrt(s2_postsample/(n2 + k0)))
```

```
# 95% CI for s2 sd
sd2_postsample <- sqrt(s2_postsample)
```

School 3

```
# 95% CI for s3 theta (mean)
set.seed(100)
vn3 <- v0 + n3
kn3 <- k0 + n3
s3n3 <- (1/vn3) * (v0 * var0 + (n3-1) * var3 + ((k0 * n3 )/kn3) * (y_bar3 - mu0)^2 )
s3_postsample <- 1/rgamma(10000, vn3/2, vn3 * s3n3 / 2)
theta3_postsample <- rnorm(10000, mu_2, sqrt(s2_postsample/(n2 + k0)))
```

```
# 95% CI for s3 sd
sd3_postsample <- sqrt(s3_postsample)
```

Results of Posterioir Confidence Intervals

```
`95% CI for Mean (Lower)`<-c(quantile(theta1_postsample,.025),quantile(theta2_postsample,.025),quantile(thet
a3_postsample,.025))

`95% CI for Mean (Upper)`<-c(quantile(theta1_postsample,.975),quantile(theta2_postsample,.975),quantile(thet
a3_postsample,.975) )

`95% CI for SD (Lower)`<- c(quantile(sd1_postsample,.025),quantile(sd2_postsample,.025),quantile(sd3_postsam
ple,.025))

`95% CI for SD (Upper)`<- c(quantile(sd1_postsample,.975) ,quantile(sd2_postsample,.975),quantile(sd3_postsa
mple,.975))
```

```
dfm<-as.data.frame(cbind(Schools,Posterior_Means,`95% CI for Mean (Lower)`,`95% CI for Mean (Upper)`,`95% CI
for SD (Lower)`, `95% CI for SD (Upper)`))

dfm %>% kbl %>% kable_styling(full_width=F)
```

	Schools	Posterior_Means	95% CI for Mean (Lower)	95% CI for Mean (Upper)	95% CI for SD (Lower)	95% CI for SD (Upper)
X2.5.	School 1	9.29230769230769	7.760371785724	10.816246659127	2.99810285901479	5.15591510551911
X2.5..1	School 2	6.94875	5.15875984580521	8.71322970325739	3.3350928249688	5.87349814387442
X2.5..2	School 3	7.81238095238095	5.15166518703728	8.7530067257975	2.79129585708162	5.11478032051347

Probability that School 1 means are less than eachother

```
# Create parameter variables for function

small <- theta1_postsample
medium <- theta2_postsample
large <- theta3_postsample


# Create function that will find P(theta1 < theta2 < theta3)
# Using loop for each integer from 1 to 10000 check if medium is greater than small
# If it is then check if large is bigger than medium, if it is add it to sum variable
# then divide that amount by the total possible (10000) values

prob = function(small, medium, large){
 sum = 0
 for(i in 1:10000){
 if(medium[i] > small[i]){
 if(large[i] > medium[i]){
 sum = sum + 1
 }
 }
 }
 sum/10000
}
```

```
# Now that function is defined rearrange all possible iterations of small, medium, large which is 3! or 6 po
ssibilities

`Theta Differences`<- c("P(theta1 < theta3 < theta2)","P(theta2 < theta1 < theta3)","P(theta2 < theta3 < the
ta1)","P(theta3 < theta2 < theta1)","P(theta3 < theta1 < theta2)","P(theta1 < theta2 < theta3)")

Probaility<- c(prob(small, large, medium),prob(medium, small, large),prob(medium, large, small),prob(large,
medium, small),prob(large, small, medium),prob(small, medium, large))

df<- as.data.frame(cbind(`Theta Differences`,Probaility))

df %>% kbl() %>% kable_styling(full_width=F)
```

Theta Differences	Probaility
P(theta1 < theta3 < theta2)	0.0012
P(theta2 < theta1 < theta3)	0.0215
P(theta2 < theta3 < theta1)	0.4719
P(theta3 < theta2 < theta1)	0.4814
P(theta3 < theta1 < theta2)	0.0222
P(theta1 < theta2 < theta3)	0.0018