Bayesian Analysis

Client name

4/22/2022

importing the dataset and libraries

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.6     v dplyr   1.0.7
## v tidyr   1.1.4     v stringr 1.4.0
## v readr   2.1.1     v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(runjags)
```

```
##
## Attaching package: 'runjags'

## The following object is masked from 'package:tidyr':
##
##     extract
```

```
library(MCMCpack)
```

```
## Loading required package: coda

## Loading required package: MASS

##
## Attaching package: 'MASS'

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

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

## ## Copyright (C) 2003-2022 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)
## ##
```

```r
Data=read.table("HWhours5schools.csv",sep = ",",header = TRUE)
Data
```

```
##    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
## 7       1 16.38
## 8       1  4.54
## 9       1  8.86
## 10      1 11.94
## 11      1 12.47
## 12      1 11.11
## 13      1 11.65
## 14      1 14.53
## 15      1  9.61
## 16      1  7.38
## 17      1  3.34
## 18      1  9.06
## 19      1  9.45
## 20      1  5.98
## 21      1  7.44
## 22      1  8.50
## 23      1  1.55
## 24      1 11.45
## 25      1  9.73
## 26      2  0.29
## 27      2  1.13
## 28      2  6.52
## 29      2 11.72
## 30      2  6.54
## 31      2  5.63
## 32      2 14.59
## 33      2 11.74
## 34      2  9.12
## 35      2  9.43
## 36      2 10.64
## 37      2 12.28
## 38      2  9.50
## 39      2  0.63
## 40      2 15.35
## 41      2  5.31
## 42      2  8.49
## 43      2  3.04
## 44      2  3.77
## 45      2  6.22
## 46      2  2.14
## 47      2  6.58
## 48      2  1.11
## 49      3  4.33
## 50      3  7.77
```

```
## 51        3  4.15
## 52        3  5.64
## 53        3  7.69
## 54        3  5.04
## 55        3 10.01
## 56        3 13.43
## 57        3 13.63
## 58        3  9.90
## 59        3  5.72
## 60        3  5.16
## 61        3  4.33
## 62        3 12.90
## 63        3 11.27
## 64        3  6.05
## 65        3  0.95
## 66        3  6.02
## 67        3 12.22
## 68        3 12.85
## 69        4 12.46
## 70        4  6.42
## 71        4  5.96
## 72        4  0.92
## 73        4 11.43
## 74        4  2.27
## 75        4  1.54
## 76        4  6.55
## 77        4  2.30
## 78        4  0.57
## 79        4  7.40
## 80        4  6.63
## 81        4  7.02
## 82        4  2.95
## 83        4  4.44
## 84        4  7.78
## 85        4  8.36
## 86        4 13.32
## 87        4  8.81
## 88        4  2.06
## 89        4 14.17
## 90        4  0.88
## 91        4 10.36
## 92        4  4.97
## 93        5 12.97
## 94        5 13.60
## 95        5 13.54
## 96        5  5.49
## 97        5 11.52
## 98        5  8.23
## 99        5  8.98
## 100       5  6.42
## 101       5 12.01
## 102       5 15.08
## 103       5  7.16
## 104       5 10.84
```

```
## 105      5  8.15
## 106      5  4.27
## 107      5 14.21
## 108      5 15.93
## 109      5  8.99
## 110      5 10.12
## 111      5  5.65
## 112      5 14.94
## 113      5 14.20
## 114      5  8.43
## 115      5 10.18
## 116      5 17.47
```

Question 3

To compare weekly hours spent on homework by students, data is collected from a sample of five different schools.Explore the weekly hours spent on homework by students from the five schools. Do the school specific mean seem significantly different from each other? What about their variances?

```r
school1=Data%>%filter(school==1)
summary(school1)
```

```
##      school      hours
##  Min.   :1   Min.   : 1.550
##  1st Qu.:1   1st Qu.: 7.440
##  Median :1   Median : 9.610
##  Mean   :1   Mean   : 9.464
##  3rd Qu.:1   3rd Qu.:11.650
##  Max.   :1   Max.   :16.380
```

```r
count(school1)
```

```
##    n
## 1 25
```

```r
school2=Data%>%filter(school==2)
summary(school2)
```

```
##      school      hours
##  Min.   :2   Min.   : 0.290
##  1st Qu.:2   1st Qu.: 3.405
##  Median :2   Median : 6.540
##  Mean   :2   Mean   : 7.033
##  3rd Qu.:2   3rd Qu.:10.070
##  Max.   :2   Max.   :15.350
```

```r
count(school2)
```

```
##    n
## 1 23
```

```r
school3=Data%>%filter(school==3)
summary(school3)
```

```
##      school      hours
##  Min.   :3   Min.   : 0.950
##  1st Qu.:3   1st Qu.: 5.130
```

```
##  Median :3   Median : 6.870
##  Mean   :3   Mean   : 7.953
##  3rd Qu.:3   3rd Qu.:11.508
##  Max.   :3   Max.   :13.630
```

```
count(school3)
```

```
##    n
## 1 20
```

```
school4=Data%>%filter(school==4)
summary(school4)
```

```
##      school      hours
##  Min.   :4   Min.   : 0.570
##  1st Qu.:4   1st Qu.: 2.292
##  Median :4   Median : 6.485
##  Mean   :4   Mean   : 6.232
##  3rd Qu.:4   3rd Qu.: 8.473
##  Max.   :4   Max.   :14.170
```

```
count(school4)
```

```
##    n
## 1 24
```

```
school5=Data%>%filter(school==5)
summary(school5)
```

```
##      school      hours
##  Min.   :5   Min.   : 4.27
##  1st Qu.:5   1st Qu.: 8.21
##  Median :5   Median :10.51
##  Mean   :5   Mean   :10.77
##  3rd Qu.:5   3rd Qu.:13.75
##  Max.   :5   Max.   :17.47
```

```
count(school5)
```

```
##    n
## 1 24
```

Means for each schools

```
mean(school1$hours)
```

```
## [1] 9.464
```

```
mean(school2$hours)
```

```
## [1] 7.033478
```

```
mean(school3$hours)
```

```
## [1] 7.953
```

```
mean(school4$hours)
```

```
## [1] 6.232083
```

```
mean(school5$hours)
```

```
## [1] 10.76583
```

Variances for each schools

```
var(school1$hours)

## [1] 15.09637

var(school2$hours)

## [1] 20.11328

var(school3$hours)

## [1] 14.30271

var(school4$hours)

## [1] 16.83032

var(school5$hours)

## [1] 13.38859
```

Set up a hierarchical normal model with common and unknown variance in their likelihood. Write out the likelihood, the prior distribution and the hyper-prior distributions

```
modelString="
model{
for (i in 1:N){
y[i]~dnorm(mu_j[DataIndex[i]],invsigma2)
}
## priors
for(j in 1:J){
mu_j[j]~dnorm(mu,invtau2)
}
invsigma2~dgamma(a_s,b_s)
sigma=sqrt(pow(invsigma2,-1))
## Hyperpriors
mu~dnorm(mu0,g0)
invtau2~dgamma(a_t,b_t)
tau=sqrt(pow(invtau2,-1))
}
"
```

define the data and prior parameters

The data contains a list of schools, the number of hours spent also contains normal hyperparameters mu0 and g0 and two sets of gamma hyperparameters(a__t and b__t) for invtau2, and(a__s and b__s) for invsigma2

```
y=Data$hours
DataIndex=Data$school
N=length(y)
J=length(unique(DataIndex))
Data2=list("y"=y,"DataIndex"=DataIndex,
           "N"=N,"J"=J,"mu0"=3,"g0"=1,
           "a_t"=1,"b_t"=1,"a_s"=1,"b_s"=1)
Data2

## $y
##   [1]  2.11  9.75 13.88 11.30  8.93 15.66 16.38  4.54  8.86 11.94 12.47 11.11
```

```
##  [13] 11.65 14.53  9.61  7.38  3.34  9.06  9.45  5.98  7.44  8.50  1.55 11.45
##  [25]  9.73  0.29  1.13  6.52 11.72  6.54  5.63 14.59 11.74  9.12  9.43 10.64
##  [37] 12.28  9.50  0.63 15.35  5.31  8.49  3.04  3.77  6.22  2.14  6.58  1.11
##  [49]  4.33  7.77  4.15  5.64  7.69  5.04 10.01 13.43 13.63  9.90  5.72  5.16
##  [61]  4.33 12.90 11.27  6.05  0.95  6.02 12.22 12.85 12.46  6.42  5.96  0.92
##  [73] 11.43  2.27  1.54  6.55  2.30  0.57  7.40  6.63  7.02  2.95  4.44  7.78
##  [85]  8.36 13.32  8.81  2.06 14.17  0.88 10.36  4.97 12.97 13.60 13.54  5.49
##  [97] 11.52  8.23  8.98  6.42 12.01 15.08  7.16 10.84  8.15  4.27 14.21 15.93
## [109]  8.99 10.12  5.65 14.94 14.20  8.43 10.18 17.47
##
## $DataIndex
##   [1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2
##  [38] 2 2 2 2 2 2 2 2 2 2 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 4 4 4 4 4 4
##  [75] 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5
## [112] 5 5 5 5 5
##
## $N
## [1] 116
##
## $J
## [1] 5
##
## $mu0
## [1] 3
##
## $g0
## [1] 1
##
## $a_t
## [1] 1
##
## $b_t
## [1] 1
##
## $a_s
## [1] 1
##
## $b_s
## [1] 1
```

#Define the data and prior parameters

```
posterior=run.jags(modelString,n.chains=5,data=Data2,monitor=c("mu","tau","mu_j","sigma"))
```

```
## Warning: No initial values were provided - JAGS will use the same initial values
## for all chains
```

```
## Calling the simulation...
## Welcome to JAGS 4.3.0 on Sun Jun 19 17:44:26 2022
## JAGS is free software and comes with ABSOLUTELY NO WARRANTY
## Loading module: basemod: ok
## Loading module: bugs: ok
## . . Reading data file data.txt
## . Compiling model graph
##    Resolving undeclared variables
##    Allocating nodes
```

```
## Graph information:
##    Observed stochastic nodes: 116
##    Unobserved stochastic nodes: 8
##    Total graph size: 254
## . Initializing model
## . Adaptation skipped: model is not in adaptive mode.
## . Updating 4000
## -------------------------------------------------| 4000
## ************************************************** 100%
## . . . . . Updating 10000
## -------------------------------------------------| 10000
## ************************************************** 100%
## . . . . . . . . Updating 0
## . Deleting model
## .
## Note: the model did not require adaptation
## Simulation complete.  Reading coda files...
## Coda files loaded successfully
## Calculating summary statistics...
## Calculating the Gelman-Rubin statistic for 8 variables....
## Finished running the simulation
```

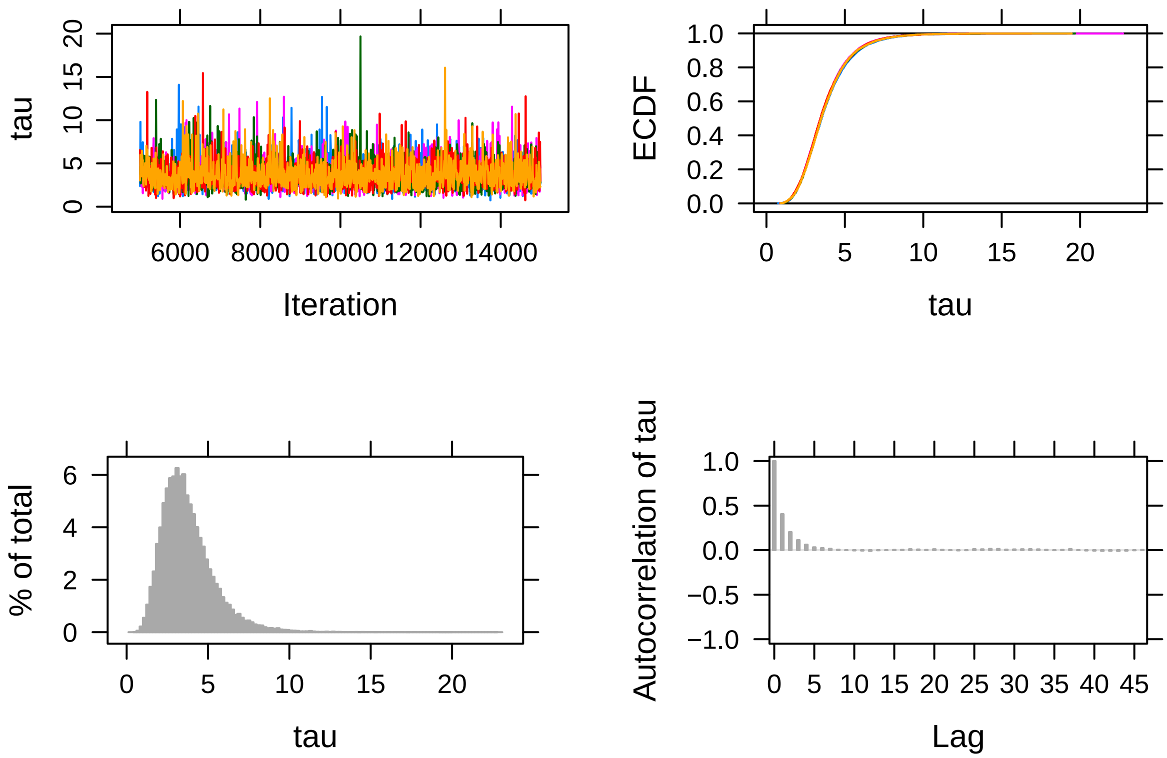
Use the JAGS to obtain posterior samples of the parameters in the hierarchical model.Perform appropriate MCMC diagnostics

```r
# Five significant figure
print(posterior,digits =5)
```

```
##
## JAGS model summary statistics from 50000 samples (chains = 5; adapt+burnin = 5000):
##
##         Lower95 Median Upper95   Mean      SD Mode     MCerr MC%ofSD SSeff
## mu       2.4107   4.59  6.8354 4.5886  1.1366   -- 0.0089697     0.8 16057
## tau       1.191 3.4746  7.0126 3.7675  1.6377   --  0.011962     0.7 18743
## mu_j[1]  7.5885 9.1713  10.713 9.1722 0.79643   -- 0.0037636     0.5 44780
## mu_j[2]  5.3327 6.9035  8.4767 6.9053 0.80266   -- 0.0035992     0.4 49734
## mu_j[3]  6.0706 7.7271  9.4552 7.7301 0.86482   -- 0.0039928     0.5 46913
## mu_j[4]  4.5863 6.1598  7.6971 6.1595 0.78977   -- 0.0035452     0.4 49629
## mu_j[5]  8.7615 10.386   12.03 10.382 0.82965   -- 0.0042457     0.5 38185
## sigma    3.4809  3.978  4.5312 3.9928  0.2703   -- 0.0012726     0.5 45114
##
##               AC.10    psrf
## mu        0.0021313  1.0002
## tau       -0.003501  1.0002
## mu_j[1]   0.0032592       1
## mu_j[2]   0.0041246       1
## mu_j[3] -0.00044277       1
## mu_j[4]  -0.0043681       1
## mu_j[5]  -0.0025857  1.0001
## sigma    -0.0030898 0.99999
##
## Total time taken: 5.8 seconds
```

```r
# Plot of the posterior
plot(posterior,vars ="tau")
```

```
## Generating plots...
```



MCMC diagnosis A 95% credible interval

```
tau_draws=as.mcmc(posterior,vars='tau')
```

```
## Warning in as.mcmc.runjags(posterior, vars = "tau"): Combining the 5 mcmc chains
## together
```

```
sigma_draws <- as.mcmc(posterior, vars = "sigma")
```

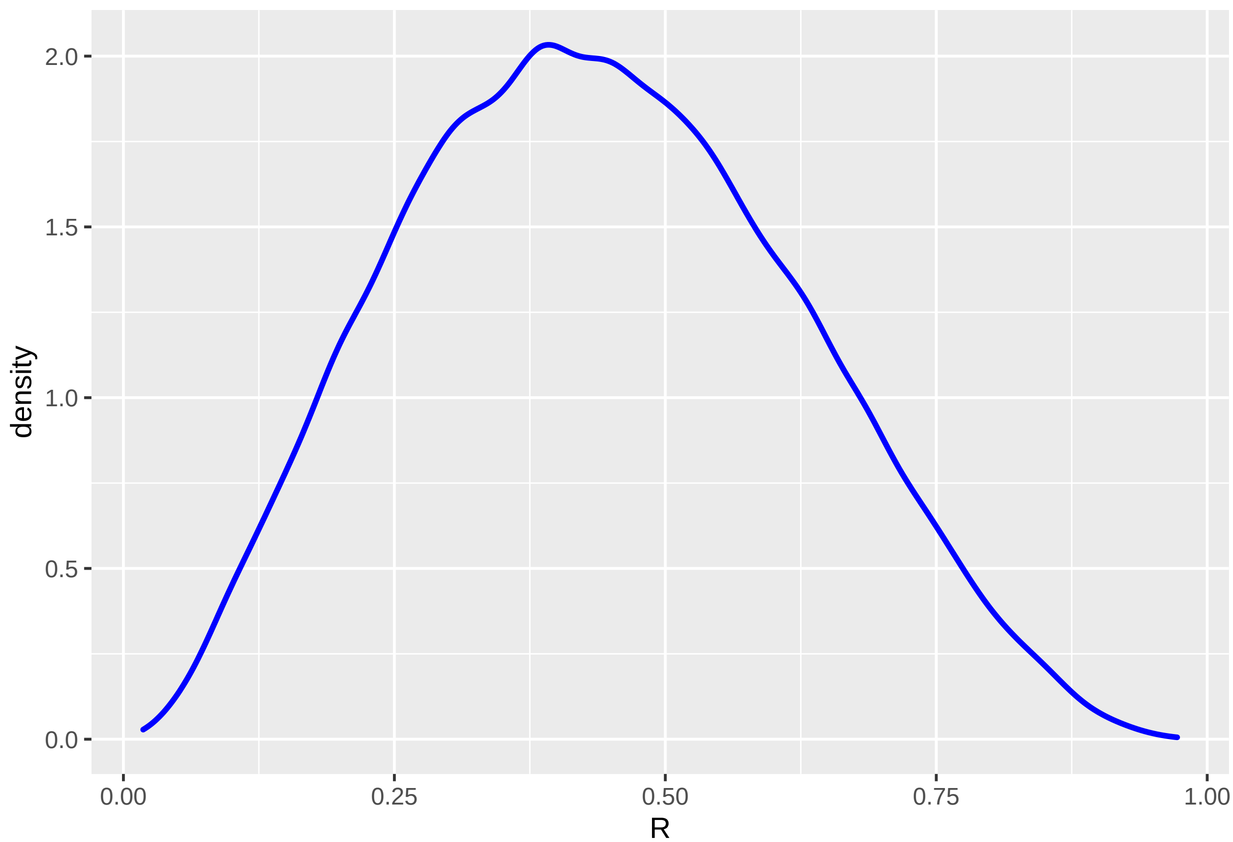
```
## Warning in as.mcmc.runjags(posterior, vars = "sigma"): Combining the 5 mcmc
## chains together
```

```
R <- (tau_draws^2)/(tau_draws^2+sigma_draws^2)
quantile(R, c(0.025, 0.975))
```

```
##      2.5%     97.5%
## 0.1183648 0.7964592
```

```
df=as.data.frame(R)
library(ggplot2)
ggplot(data=df,mapping =aes(x=R))+geom_density(size = 1,color="blue")
```

```
## Don't know how to automatically pick scale for object of type mcmc. Defaulting to continuous.
```



https://miktex.org/download