LAB2

Name: Huiyu Hu

```
library(R2jags)
```

```
## Loading required package: rjags
```

```
## Loading required package: coda
```

```
## Linked to JAGS 4.3.0
```

```
## Loaded modules: basemod,bugs
```

```
##
## Attaching package: 'R2jags'
```

```
## The following object is masked from 'package:coda':
##
##     traceplot
```

```
library(lattice) # Needed for scatterplot matrix
# Set working directory
setwd("/Users/huiyuhu/Desktop/Study/UCLA_Biostat/BIOSTAT234/lab/Lab 2")
getwd()
```

```
## [1] "/Users/huiyuhu/Desktop/Study/UCLA_Biostat/BIOSTAT234/lab/Lab 2"
```

```
housing <- read.table("housingdata2.txt")
#Give the columns useful names
colnames(housing) <- c("cost", "eaves", "windows", "yard", "roof")

#SEPARATE X & Y
y <- housing[,1]
x <- as.matrix(housing[,2:5])
```

- Define the model.

```
# sink("housingmodel.txt")
# cat("
# model
# {
#    for(i in 1:N) {
#        y[i] ~ dnorm( mu[i] , tau )
#        mu[i] <- beta0 + inprod(x[i,] , beta[] )
#       }
#
#    beta0 ~ dnorm( mbeta0 , precbeta0)
#
# for (j in 1:K) {
#    beta[j] ~ dnorm( m[j] , prec[j] )
#       }
#      tau ~ dgamma( tau.a , tau.b )
#      sigma <- 1 / sqrt( tau )
#   }
#   ",fill = TRUE)
# sink()
```

- Define the prior: Based on the influence we wish to permit the prior data to have, we might consider several priors.

```
#DIFFERENT PRIORS TO TRY
dataA<-list(N=21, K=4, m=c(1.6053, 1.2556, 2.3413, 3.6771),
            prec=c(.2164, .1105, .2061, .1337), tau.a=17,
            tau.b=1128, mbeta0= -5.682, precbeta0=.05464, x=x, y=y)


dataB<-list(N=21, K=4, m=c(1.6053, 1.2556, 2.3413, 3.6771),
            prec=c(.02774, .014160, .02642, .01714), tau.a=2.1795,
            tau.b=144.6, mbeta0= -5.682, precbeta0=.007005, x=x, y=y)

dataC<-list(N=21, K=4, m=c(1.6053, 1.2556, 2.3413, 3.6771),
            prec=c(.005549, .002832, .005284, .003428), tau.a=.4359,
            tau.b=28.92, mbeta0= -5.682, precbeta0=.00140, x=x, y=y)
```

- Define the initial values and the parameters to monitor

```
#SET UP INITAL VALUES
inits <- rep(list(list(beta0=0, beta=c(1,1,1,1),tau=1)),5)

#DEFINE PARAMETERS TO MONITOR
parameters <- c("beta0", "beta" , "tau")
```

- Fit the model.

```
#RUN THE JAGS PROGRAM, SAVING DATA TO LAB2.SIM
lab2.sim.a <- jags (dataA, inits, parameters, "housingmodel.txt", n.chains=5,
    n.iter=5100, n.burnin=100, n.thin=1, DIC=FALSE)
```

```
## module glm loaded
```

```
## module dic loaded
```

```
## Compiling model graph
##    Resolving undeclared variables
##    Allocating nodes
## Graph information:
##    Observed stochastic nodes: 21
##    Unobserved stochastic nodes: 6
##    Total graph size: 190
##
## Initializing model
```

```
#n.chain: do this with 5 diff initial value

lab2.sim.b <- jags (dataB, inits, parameters, "housingmodel.txt", n.chains=5,
    n.iter=5100, n.burnin=100, n.thin=1, DIC=FALSE)
```

```
## Compiling model graph
##    Resolving undeclared variables
##    Allocating nodes
## Graph information:
##    Observed stochastic nodes: 21
##    Unobserved stochastic nodes: 6
##    Total graph size: 190
##
## Initializing model
```

```
lab2.sim.c <- jags (dataC, inits, parameters, "housingmodel.txt", n.chains=5,
    n.iter=5100, n.burnin=100, n.thin=1, DIC=FALSE)
```

```
## Compiling model graph
##    Resolving undeclared variables
##    Allocating nodes
## Graph information:
##    Observed stochastic nodes: 21
##    Unobserved stochastic nodes: 6
##    Total graph size: 190
##
## Initializing model
```

1. Summarize briefly the effects on all parameters of changing from prior A to B to C. (Briefly = one sentence total; two only if really necessary).

```
knitr::kable(lab2.sim.a$BUGSoutput$summary[,c("mean", "sd")],format = "pandoc", digits =
 3,
             caption = "Parameter estimates from Model A")
```

Parameter estimates from Model A

	mean	sd
beta[1]	1.769	1.358
beta[2]	1.789	1.975
beta[3]	2.230	1.312
beta[4]	2.580	2.252
beta0	-6.957	3.789
tau	0.021	0.004

```
knitr::kable(lab2.sim.b$BUGSoutput$summary[,c("mean", "sd")],format = "pandoc", digits =
 3,
             caption = "Parameter estimates from Model B")
```

Parameter estimates from Model B

	mean	sd
beta[1]	1.913	1.402
beta[2]	3.190	2.326
beta[3]	2.498	1.258
beta[4]	1.153	5.021
beta0	-8.845	9.658
tau	0.046	0.014

```
knitr::kable(lab2.sim.c$BUGSoutput$summary[,c("mean", "sd")],format = "pandoc", digits =
 3,
             caption = "Parameter estimates from Model C")
```

Parameter estimates from Model C

	mean	sd
beta[1]	1.908	1.165
beta[2]	3.478	1.962
beta[3]	2.536	1.036
beta[4]	0.921	10.624
beta0	-9.211	21.221
tau	0.074	0.025

- The beta 1-4 changing from prior A to B to C were closer to the classical regression's beta result.

2. Give a table of inferences for the coefficient of roofs for the three priors. Briefly explain why it comes out as it does.

```
roof.a <- lab2.sim.a$BUGSoutput$summary["beta[4]",c("mean", "sd")]
roof.b <- lab2.sim.b$BUGSoutput$summary["beta[4]",c("mean", "sd")]
roof.c <- lab2.sim.c$BUGSoutput$summary["beta[4]",c("mean", "sd")]
roof <- rbind(roof.a,roof.b, roof.c)

knitr::kable(roof,format = "pandoc", caption = "Coefficient of roofs for the three prior
s")
```

Coefficient of roofs for the three priors

	mean	sd
roof.a	2.5803552	2.251904
roof.b	1.1530075	5.021021
roof.c	0.9210331	10.624076

- Since there is 0 variance for roof, therefore the estimate gets more centered on zero according to the less and less emphasis on prior.

3. For one of the three priors:

a. Show summaries of the futurefit, futureobs, futuretail in a properly formatted table for the house in perfect condition.

- The first definition calculates the parameter beta0 + xi*beta for xi = (1,1,2,2), which corresponds to a house in perfect condition
- Add future part in JAGS model

```
# sink("housingmodel_pred.txt")
# cat("
# model
# {
#    for(i in 1:N) {
#        y[i] ~ dnorm( mu[i] , tau )
#        mu[i] <- beta0 + inprod(x[i,] , beta[] )
#       }
#
#    beta0 ~ dnorm( mbeta0 , precbeta0)
#
# for (j in 1:K) {
#    beta[j] ~ dnorm( m[j] , prec[j] )
#       }
# tau ~ dgamma( tau.a , tau.b )
# sigma <- 1 / sqrt( tau )
# futurefit <- beta0 + beta[1] + beta[2] + beta[3]*2 + beta[4]*2
# futureobs ~ dnorm(futurefit, tau)
# futuretail <- beta0 + beta[1] + beta[2] + beta[3]*2 + beta[4]*2 + 1.645*sigma
#   }
#   ",fill = TRUE)
# sink()
```

```
inits<-rep(list(list(beta0=0, beta=c(1,1,1,1),tau=1, futureobs=10)),5)
parameters <- c("beta0", "beta" , "tau", "sigma", "futurefit", "futureobs", "futuretail"
)
lab2.sim.pred <- jags (dataA , inits, parameters, "housingmodel_pred.txt", n.chains=5,
                       n.iter=5100, n.burnin=100, n.thin=1, DIC=FALSE)
```

```
## Compiling model graph
##    Resolving undeclared variables
##    Allocating nodes
## Graph information:
##    Observed stochastic nodes: 21
##    Unobserved stochastic nodes: 7
##    Total graph size: 198
##
## Initializing model
```

```
knitr::kable(lab2.sim.pred$BUGSoutput$summary[c("futurefit", "futureobs", "futuretail"),
c("mean", "sd")],format = "pandoc", digits = 3,
             caption = "Summary of predictions for a perfect house")
```

Summary of predictions for a perfect house

	mean	sd
futurefit	6.244	2.959
futureobs	6.258	7.583

	mean	sd
futuretail	17.654	3.247

- Therefore, the estimate of the cost for a perfect house is about $6200.

b. Which house is in the worst condition? Calculate the three futurefit, futureobs and futuretail variables for this house and provide a formatted table.

```
x <- as.data.frame(x)
x$total <-rowSums(x)
x[14,]
```

```
##    eaves windows yard roof total
## 14     3    3.33    4    2 12.33
```

- Based on the result above (the sum of each row), the maximum of score is #14 observation (3.00 3.33 4 2)

```
# sink("housingmodel_pred1.txt")
# cat("
# model
# {
#    for(i in 1:N) {
#        y[i] ~ dnorm( mu[i] , tau )
#        mu[i] <- beta0 + inprod(x[i,] , beta[] )
#       }
#
#    beta0 ~ dnorm( mbeta0 , precbeta0)
#
# for (j in 1:K) {
#    beta[j] ~ dnorm( m[j] , prec[j] )
#       }
# tau ~ dgamma( tau.a , tau.b )
# sigma <- 1 / sqrt( tau )
# futurefit <- beta0 + beta[1]*3.00 + beta[2]*3.33 + beta[3]*4 + beta[4]*2
# futureobs ~ dnorm(futurefit, tau)
# futuretail <- beta0 + beta[1]*3.00 + beta[2]*3.33 + beta[3]*4 + beta[4]*2 + 1.645*sigm
a
#   }
#   ",fill = TRUE)
# sink()
```

```
inits<-rep(list(list(beta0=0, beta=c(1,1,1,1),tau=1, futureobs=10)), 5)
parameters <- c("beta0", "beta" , "tau", "sigma", "futurefit", "futureobs", "futuretail"
)
lab2.sim.pred <- jags (dataA , inits, parameters, "housingmodel_pred1.txt", n.chains=5,
                       n.iter=5100, n.burnin=100, n.thin=1, DIC=FALSE)
```

```
## Compiling model graph
##    Resolving undeclared variables
##    Allocating nodes
## Graph information:
##    Observed stochastic nodes: 21
##    Unobserved stochastic nodes: 7
##    Total graph size: 203
##
## Initializing model
```

```
knitr::kable(lab2.sim.pred$BUGSoutput$summary[c("futurefit", "futureobs", "futuretail"),
c("mean", "sd")],format = "pandoc", digits = 3,
             caption = "Summary of predictions for the worst house")
```

Summary of predictions for the worst house

	mean	sd
futurefit	18.382	2.726
futureobs	18.303	7.421
futuretail	29.794	2.923

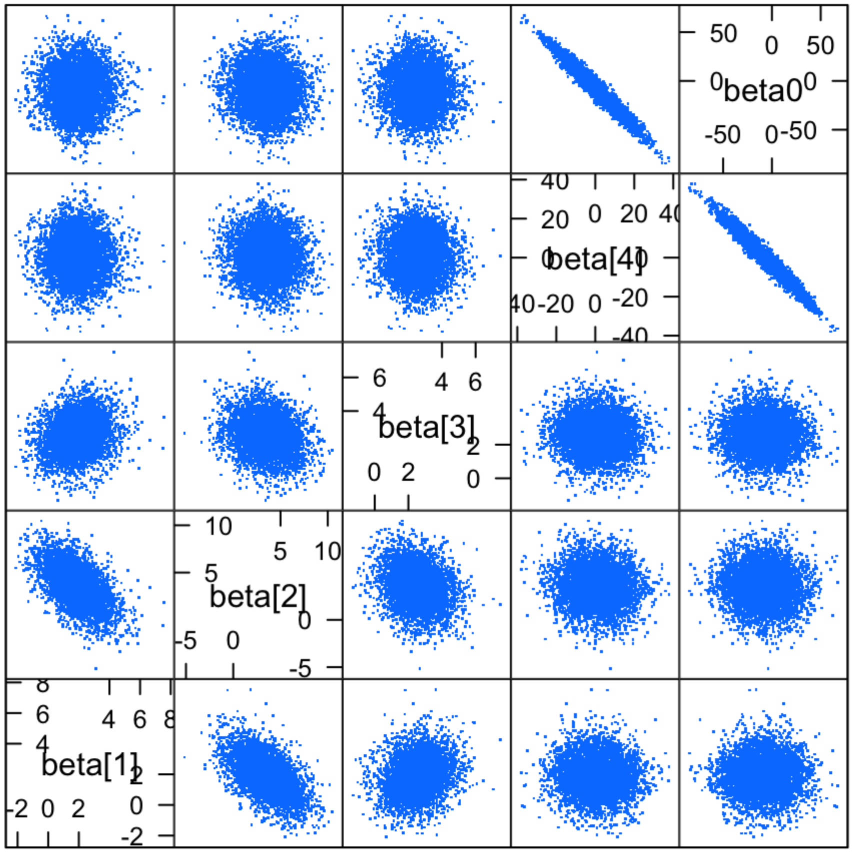
- Therefore, the estimate of the cost for worst house is about \$18400, less than the cost from original data (\$25000)

4. For prior (C), what two coefficients (including the intercept) have the highest posterior correlation? Briefly explain why.

```
temp2=lab2.sim.c$BUGSoutput$sims.matrix
head(temp2)
```

```
##        beta[1]   beta[2]  beta[3]    beta[4]     beta0        tau
## [1,] 2.7680209 1.8874418 2.408444  11.991038 -28.98574 0.09192612
## [2,] 1.5222685 3.2908617 1.538133   9.288328 -23.15509 0.11885125
## [3,] 1.5699872 0.8041684 3.289667   6.889867 -13.70479 0.07892607
## [4,] 2.4010517 1.9021043 3.384738 -17.035328  25.95643 0.05886793
## [5,] 0.4613515 3.7873459 2.288952   9.904509 -24.29799 0.06962394
## [6,] 1.8595733 4.6156942 2.710769   3.371132 -16.92674 0.07710152
```

```
splom(temp2[1:5000,1:5],pch=".")    #   Scatterplot matrix of correlation plots
```



Scatter Plot Matrix

- Based on result above, that beta0 (intercept) and beta[4] (roof) are highly correlated for Prior C. Since the variation of roof data is 0, it is like a constant number (2) in this process. Therefore, both beta[4]*roof and beta0 are zero and there will be strong correlation between them.

5. Briefly interpret the three variables futurefit, futureobs, futuretail in your own words.

```
# futurefit <- beta0 + beta[1] + beta[2] + beta[3]*2 + beta[4]*2
# futureobs ~ dnorm(futurefit, tau)
# futuretail <- beta0 + beta[1] + beta[2] + beta[3]*2 + beta[4]*2 + 1.645*sigma
```

- Futurefit: the expectation of mean of the distribution and can be used to estimate the cost based on the condition of the house.
- Futureobs: a random sample from the distribution of the predicted cost estimate.
- Futuretail: upper limit of 90% interval of the expectation of mean of the distribution.

6. Suppose we pool the two data sets after the inflation correction. Also, the expert at the housing department told you he thought each unit increase in any rating scale ought to increase the cost by around $1000. You're not sure that all coefficients should be positive. Suggest priors (all regression coefficients and for sigma^2) to use now. Write one or two sentences justifying your priors.

- Since each unit increase in any rating scale ought to increase the cost by around $1000, 1 as prior mean for beta 1-4 will be the coeffients. I also decrease the precision, then the coeffient estimate for regression could be positive and negative. For tau, I didn't change and still assumed that it is Gamma(17, 1128), with mean 17/1128 = 1/66.35. The number 17 = 34/2 is one half the degrees of freedom for estimating. I am not sure how to justify the intercept, I still assumed N0 ~ (-5.682, .007005)

- EAVES: beta[1] ~ N_0(1, 0.001)
- WINDOWS: beta[2] ~ N_0(1, 0.001)
- YARDS: beta[3] ~ N_0 (1, 0.001)
- ROOF: beta[4] ~ N_0(1, 0.001)
- Intercept: beta0 ~ N_0(-5.682, .007005)
- Tau: tau ~ Gamma(17, 1128)