Manage Data and Estimate Classical/Frequentist Binary Regression

- Set your working directory and create a log file to divert your codes and results
- Read the external Stata data file gsscum7212Teach.dta into R.
- Keep six variables, include mental health (mntlhlth), age (age), sex (sex), race (race), education (educ), and income (inc1k)
- Explore these six variables using the the table and summary functions, or some other functions deemed appropriate. While presenting frequency distributions for continuous variables with too many categories (e.g., age or income), the value categories can be selectively presented.
- Dichotomize the mental health variable such that the new binary response variable is coded as one (1 = having poor mental health) if the number of days for poor mental health is greater than zero, otherwise it's coded as zero (0 = having excellent mental health). Label this new variable as mntlhlthc2
- Create a dummy variable for sex using male as the reference category (hint: for example, the new indicator variable can be called female with female coded as one and male coded as zero). Note that you can create a dummy variable for sex using female as the reference category (hint: for example the new indicator variable can be called male with male coded as one and female coded as zero). But when one enters the sex variable in a regression model, both dummy variables cannot be entered simultaneously, and usually either one has to be dropped. Please think about why.
- Create a set of dummy variables for the race variable. Note that the race variable has three categories, so one can create three dummy variables for race. Please be careful and clear about 1) how many of the three dummy variables, all measuring race, are usually used in a regression model and 2) how to interpret the results/corresponding coefficients (which group is the reference group?).
- Drop missing cases
- Run a classical (frequentist) binary logit regression of mental health (mntlhlth2) on age (age), sex (sex), race (race; coded as a factor variable or a set of dummy variables), education (educ), and income (inc1k), produce odds ratio coefficients, and sample one or two odds ratio coefficients for interpretation.

```
> rm(list=ls(all=TRUE))

> mygss <- read.dta("gsscum7212teach.dta", convert.factor=F)

> #MANAGE DATA AND RUN LOGIT
> #SELECT DATA
> useddta <- subset(mygss,
+                   select=c(mntlhlth, age, sex, race, educ, inc1k))

> #Create dummy variables female (male = 0)
> useddta$female <- as.numeric(useddta$sex==2)

> useddta$male <- as.numeric(useddta$sex == 1)

> #Create Binary Indicator Variables for Multi-Category Nomial Variables
>
> useddta$white <- ifelse(useddta$race == 1, 1, 0)

> useddta$black <- ifelse(useddta$race == 2, 1, 0)
```

```
> useddta$other <- ifelse(useddta$race == 3, 1, 0)

> nmdta <- useddta[complete.cases(useddta),] #no missing data

> #summarize data
> summary(nmdta$mntlhlth)
  Min.     1st Qu.     Median      Mean       3rd Qu.     Max.
  0.000    0.000       0.000       3.991      5.000       30.000

> summary(nmdta$inc1k)
  Min.     1st Qu.     Median      Mean       3rd Qu.     Max.
  0.245    15.921      28.157      39.404     48.475      155.140

> summary(nmdta$age)
  Min.     1st Qu.     Median      Mean       3rd Qu.     Max.
  18.00    31.00       42.00       41.81      51.00       84.00

> table(nmdta$female)

      0     1
      381   369

> table(nmdta$white)

      0     1
      160   590

> #Run OLS
> mntlhlth.model <- lm(formula = mntlhlth ~ age + female + black + other +
educ + inc1k, data = nmdta )

> (summary(mntlhlth.model))

Call:
lm(formula = mntlhlth ~ age + female + black + other + educ +
    inc1k, data = nmdta)

Residuals:
Min         1Q          Median             3Q          Max
-6.888      -4.069      -2.818             0.601       27.711

Coefficients:
             Estimate   Std. Error  t value     Pr(>|t|)
(Intercept)  6.911999   1.656997    4.171       3.38e-05 ***
age         -0.009268   0.020120    -0.461      0.64517
female       1.017865   0.528007    1.928       0.05427 .
black       -2.119554   0.790024    -2.683      0.00746 **
other        0.569797   0.955038    0.597       0.55094
educ        -0.188625   0.102889    -1.833      0.06716 .
inc1k       -0.004487   0.007959    -0.564      0.57309
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 7.105 on 743 degrees of freedom
Multiple R-squared:  0.02043, Adjusted R-squared:  0.01252
F-statistic: 2.582 on 6 and 743 DF,  p-value: 0.01753
```

```
> exp(mntlhlth.model$coefficients)
(Intercept)  age        female     black       other        educ         inc1k
1004.2528741 0.9907743  2.7672800  0.1200852   1.7679076    0.8280966    0.9955230

> exp(confint(mntlhlth.model))
                  2.5 %            97.5 %
(Intercept)       38.82429117      2.597662e+04
age               0.95240335       1.030691e+00
female            0.98147563       7.802373e+00
black             0.02546378       5.663126e-01
other             0.27114621       1.152698e+01
educ              0.67664131       1.013453e+00
inc1k             0.98008883       1.011200e+00
```

Task 1 Run Bayesian Binary Logit Model with Non-informative Priors

- Use the MCMClogit function from the MCMCpack package to run a Bayesian linear regression of mental health on age, sex (male as the reference category), race (white as the reference category), education, and income with non-informative priors (e.g., N (0, 1e6)). Note that the B0 argument in the MCMClogit function is for specifying priors for precision, the inverse of variance. Use the summary function to summarize the results from this Bayesian linear regression.

```
> #TASK 1: Bayesian Binary Logit Model With Non-Informative Priors
> library(MCMCpack)

> mcmc.model = MCMCregress(mntlhlth ~ age + female + black + other + educ +
inc1k,
+                          data = nmdta, burnin = 1000, mcmc = 10000, thin =
1,seed = 47304,b0 = 0, B0 = 1e-6,marginal.likelihood = "Chib95")

> summary(mcmc.model)

Iterations = 1001:11000
Thinning interval = 1
Number of chains = 1
Sample size per chain = 10000

1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:

                 Mean         SD  Naive   SE Time-series    SE
(Intercept)       6.905019    1.652458    1.652e-02         1.636e-02
age              -0.009368    0.020364    2.036e-04         2.036e-04
female            1.018804    0.523410    5.234e-03         5.311e-03
black            -2.133636    0.778937    7.789e-03         7.936e-03
other             0.578391    0.954541    9.545e-03         9.545e-03
educ             -0.187699    0.102528    1.025e-03         1.025e-03
inc1k            -0.004541    0.007927    7.927e-05         7.927e-05
sigma2           50.605190    2.605665    2.606e-02         2.665e-02

2. Quantiles for each variable:

                  2.5%        25%         50%         75%         97.5%
```

```
(Intercept)        3.65776     5.798106    6.901763    8.0250967   10.14658
age                -0.04969    -0.023026   -0.009389   0.0043824   0.03062
female             0.00458     0.664799    1.015528    1.3744351   2.06364
black              -3.65706    -2.655721   -2.131216   -1.6082498  -0.60899
other              -1.31836    -0.063896   0.582669    1.2131549   2.45338
educ               -0.38943    -0.257037   -0.187437   -0.1190083  0.01290
inc1k              -0.02015    -0.009885   -0.004572   0.0008091   0.01081
sigma2             45.79020    48.807843   50.525690   52.3088911  55.89927
```

Task 2 Interpret Results Using Credit Intervals

- Compare results from the Bayesian binary regression with those obtained from classical frequentist binary regression (e.g., are they different/similar, and if so, to what extent?). Use 95% credible intervals to interpret results for selected predictors, such as education and race.

```
> #TASK 2 Interpret Results Using Credit Intervals
> #Interpret data from summary
```

Interpretation:

For the traditional model, we had a coefficient for education of -0.188625. Whereas for the Bayesian model, the mean value for education was very close. In fact all of the estimates for the coefficients were within 0.01 units of the corresponding means using Bayesian methods.

```
2. Quantiles for each variable:

                   2.5%        25%         50%         75%         97.5%
(Intercept)        3.65776     5.798106    6.901763    8.0250967   10.14658
age                -0.04969    -0.023026   -0.009389   0.0043824   0.03062
female             0.00458     0.664799    1.015528    1.3744351   2.06364
black              -3.65706    -2.655721   -2.131216   -1.6082498  -0.60899
other              -1.31836    -0.063896   0.582669    1.2131549   2.45338
educ               -0.38943    -0.257037   -0.187437   -0.1190083  0.01290
inc1k              -0.02015    -0.009885   -0.004572   0.0008091   0.01081
sigma2             45.79020    48.807843   50.525690   52.3088911  55.89927
```

Credible Intervals:

There is a 95% probability that the true estimate of the coefficient of education would fall within the interval -0.3894 and 0.0129, given the evidence provided by the observed data. Since 0 falls within this evidence, there is no evidence that education has a statistically significant impact on mental health days.

There is a 95% probability that the true estimate for the coefficient of the variable black would fall between -3.6571 and -.60899, given the evidence provided by the observed data. Because 0 is not included within the interval we can conclude that being black has a significant negative impact on mental health.

There is a 95% probability that the true estimate for the coefficient of the variable other would fall between -1.3184 and 2.4534, given the evidence provided by the observed data.

Task 3 Run Bayesian Logit with Informative Priors

- Use the MCMClogit function from the MCMCpack package to run a Bayesian linear regression of mental health on age, sex (male as the reference category), race (white as the reference category), education, and income using informative prior (e.g., N (2, 0.001)). Then compare the results from this model with those from the first Bayesian binary regression model.

```
> #Task 3 Run Bayesian Logit With Informative Pairs
> #The informative priors is used for all the covariates
> mcmc.model2 = MCMCregress(mntlhlth ~ age + female + black + other + educ +
inc1k,data = nmdta, burnin = 1000, mcmc = 10000, thin = 1,seed = 47304,b0 =
1, B0 = .001,marginal.likelihood = "Chib95")

> summary(mcmc.model2)

Iterations = 1001:11000
Thinning interval = 1
Number of chains = 1
Sample size per chain = 10000

1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:

              Mean       SD          Naive SE          Time-series SE
(Intercept)  6.888026   1.650147    1.650e-02         1.633e-02
age         -0.009262   0.020356    2.036e-04         2.036e-04
female       1.019319   0.523323    5.233e-03         5.310e-03
black       -2.130668   0.778669    7.787e-03         7.934e-03
other        0.580484   0.954069    9.541e-03         9.541e-03
educ        -0.186841   0.102432    1.024e-03         1.024e-03
inc1k       -0.004557   0.007927    7.927e-05         7.927e-05
sigma2      50.604888   2.605638    2.606e-02         2.665e-02

2. Quantiles for each variable:

              2.5%       25%          50%          75%          97.5%
(Intercept)  3.644594   5.782807     6.884677     8.0060721    10.12558
age         -0.049572   -0.022921    -0.009287    0.0044762    0.03071
female       0.005248   0.665446     1.016017     1.3748670    2.06403
black       -3.653532   -2.652605    -2.128643    -1.6055065   -0.60681
other       -1.315529   -0.061339    0.584827     1.2151737    2.45434
educ        -0.388332   -0.256109    -0.186577    -0.1182027   0.01368
inc1k       -0.020160   -0.009902    -0.004587    0.0007921    0.01080
sigma2      45.788474   48.808543    50.525504    52.3091770   55.89620
> BF = BayesFactor(mcmc.model,mcmc.model2)

> BF
The matrix of Bayes Factors is:
                  mcmc.model  mcmc.model2
mcmc.model        1.00e+00    3.25e-11
mcmc.model2       3.08e+10    1.00e+00

The matrix of the natural log Bayes Factors is:
                  mcmc.model  mcmc.model2
mcmc.model        0.0         -24.2
mcmc.model2       24.2         0.0

 mcmc.model :
   call =
MCMCregress(formula = mntlhlth ~ age + female + black + other +
    educ + inc1k, data = nmdta, burnin = 1000, mcmc = 10000,
```

```
    thin = 1, seed = 47304, b0 = 0, B0 = 1e-06, marginal.likelihood =
"Chib95")

   log marginal likelihood =  -2602.745


 mcmc.model2 :
   call =
MCMCregress(formula = mntlhlth ~ age + female + black + other +
    educ + inc1k, data = nmdta, burnin = 1000, mcmc = 10000,
    thin = 1, seed = 47304, b0 = 1, B0 = 0.001, marginal.likelihood =
"Chib95")

   log marginal likelihood =  -2578.594
```

Interpretation:

Since the Bayes factor is greater than 150, there is strong evidence that the second informative model has greater predictive power than the noninformative model.

Task 4 Produce Prediction

- Using results from the first Bayesian logit (with non-informative priors) to predict the predicted probability of having poor mental health for a 35-year-old white female with college education and sample median income. Please report the mean of the prediction distribution and the 95% credible interval using empirical percentiles.

```
> #Task 4 Produce Prediction
> #35-Year-Old White female with college education and sample median income
> library(HDInterval)

> require(rjags)

> library(BEST)

> post.mat = as.matrix(mcmc.model)

> cut.mat = post.mat[,c("(Intercept)", "age",
'female','black','other','educ','inc1k')]

> pred.hyp <- 1*post.mat[,"(Intercept)"]+ 35*post.mat[,'age'] +
1*post.mat[,'female'] + 0*post.mat[,'black'] + 0*post.mat[,'other'] +
16*post.mat[,'educ'] + median(nmdta$inc1k)*post.mat[,'inc1k']

> summary(pred.hyp)
  Min.      1st Qu.     Median      Mean        3rd Qu.     Max.
  2.630     4.139       4.465       4.465       4.788       6.314

> quantile(pred.hyp, c(.025,.5,.975))
2.5%        50%         97.5%
3.527574    4.464752    5.400748
```

Interpretation:

The mean prediction for mental health days for a 35-year-old white female with a college degree and median income is 4.465.

There is a 95% probability that the actual number of mental health days for a 35-year-old female with a college degree and median income is between 3.5275 and 5.4007.

Task 5 Calculate Difference in Prediction

- Make the same prediction for an otherwise similar male. Then calculate the difference in the predictions and construct its 95% credible interval. Please interpret the results and answer if there is no difference between these two predictions using results from the credible interval.

```
> #Task 5 Calculate the Difference in Prediction
> #Prediction for 35-year-old white male with college education and median
income
> pred.hyp2 <- 1*post.mat[,"(Intercept)"]+ 35*post.mat[,'age'] +
0*post.mat[,'female'] + 0*post.mat[,'black'] + 0*post.mat[,'other'] +
16*post.mat[,'educ'] + median(nmdta$inc1k)*post.mat[,'inc1k']

> summary(pred.hyp2)
  Min.      1st Qu.     Median      Mean        3rd Qu.     Max.
  1.557     3.117       3.451       3.446       3.776       5.197

> #Differences in Predicted Outcomes
> quantile(pred.hyp-pred.hyp2, c(.025,.5,.975))
2.5%                50%                 97.5%
0.004579817         1.015528297         2.063644188
```

Interpretation:

The mean prediction for number of mental health days for a 35-year-old white male with a college education and median income is 3.446.

There is a 95% probability that the actual difference between mental health days for a 35-year-old female with a college education and median income and an otherwise similar male is between 0.0046 and 2.06364. Since 0 is not captured within the interval, the number of mental health days for 35-year-old female with a college education and median income is statistically significantly higher than for an otherwise similar male.

Close Out

- Close out the log file

R-Script

```
#
source("/Users/burrisfaculty/Desktop/DSCode/SOC686/Shepherd_Lab10_SOC686.r",
echo=T, max.deparse.length=10000)
library(foreign)


#Open Log and read in data
setwd("/Users/burrisfaculty/Desktop/DSCode/SOC686")
sink("Shepherd_asgn19.log", split=T)
rm(list=ls(all=TRUE))
mygss <- read.dta("gsscum7212teach.dta", convert.factor=F)

#MANAGE DATA AND RUN LOGIT
#SELECT DATA
useddta <- subset(mygss,
                  select=c(mntlhlth, age, sex, race, educ, inc1k))
```

```
#Create dummy variables female (male = 0)
useddta$female <- as.numeric(useddta$sex==2)
useddta$male <- as.numeric(useddta$sex == 1)

#Create Binary Indicator Variables for Multi-Category Nomial Variables

useddta$white <- ifelse(useddta$race == 1, 1, 0)
useddta$black <- ifelse(useddta$race == 2, 1, 0)
useddta$other <- ifelse(useddta$race == 3, 1, 0)

nmdta <- useddta[complete.cases(useddta),] #no missing data

#summarize data
summary(nmdta$mntlhlth)
summary(nmdta$inc1k)
summary(nmdta$age)
table(nmdta$female)
table(nmdta$white)

#Run OLS
mntlhlth.model <- lm(formula = mntlhlth ~ age + female + black + other + educ
+ inc1k, data = nmdta )
(summary(mntlhlth.model))
exp(mntlhlth.model$coefficients)
exp(confint(mntlhlth.model))
#TASK 1: Bayesian Binary Logit Model With Non-Informative Priors
library(MCMCpack)
mcmc.model = MCMCregress(mntlhlth ~ age + female + black + other + educ +
inc1k,
                         data = nmdta, burnin = 1000, mcmc = 10000, thin =
1,seed = 47304,b0 = 0, B0 = 1e-6,marginal.likelihood = "Chib95")
summary(mcmc.model)

#TASK 2 Interpret Results Using Credit Intervals
#Interpret data from summary

#Task 3 Run Bayesian Logit With Informative Pairs
#The informative priors is used for all the covariates
mcmc.model2 = MCMCregress(mntlhlth ~ age + female + black + other + educ +
inc1k,
                         data = nmdta, burnin = 1000, mcmc = 10000, thin =
1,seed = 47304,b0 = 1, B0 = .001,marginal.likelihood = "Chib95")
summary(mcmc.model2)
BF = BayesFactor(mcmc.model,mcmc.model2)
BF
#Task 4 Produce Prediction
#35-Year-Old White female with college education and sample median income
library(HDInterval)
require(rjags)
library(BEST)
post.mat = as.matrix(mcmc.model)
cut.mat = post.mat[,c("(Intercept)", "age",
'female','black','other','educ','inc1k')]
pred.hyp <- 1*post.mat[,"(Intercept)"]+ 35*post.mat[,'age'] +
1*post.mat[,'female'] +
  0*post.mat[,'black'] + 0*post.mat[,'other'] + 16*post.mat[,'educ'] +
median(nmdta$inc1k)*post.mat[,'inc1k']
```

```
summary(pred.hyp)
quantile(pred.hyp, c(.025,.5,.975))
#Task 5 Calculate the Difference in Prediction
#Prediction for 35-year-old white male with college education and median
income
pred.hyp2 <- 1*post.mat[,"(Intercept)"]+ 35*post.mat[,'age'] +
0*post.mat[,'female'] +
  0*post.mat[,'black'] + 0*post.mat[,'other'] + 16*post.mat[,'educ'] +
median(nmdta$inc1k)*post.mat[,'inc1k']
summary(pred.hyp2)

#Differences in Predicted Outcomes
quantile(pred.hyp-pred.hyp2, c(.025,.5,.975))

#Close log
save(nmdta, file = "Assignment_10.rdata")
sink()
```

Log:

```
> rm(list=ls(all=TRUE))

> mygss <- read.dta("gsscum7212teach.dta", convert.factor=F)

> #MANAGE DATA AND RUN LOGIT
> #SELECT DATA
> useddta <- subset(mygss,
+                   select=c(mntlhlth, age, sex, race, educ, inc1k))

> #Create dummy variables female (male = 0)
> useddta$female <- as.numeric(useddta$sex==2)

> useddta$male <- as.numeric(useddta$sex == 1)

> #Create Binary Indicator Variables for Multi-Category Nomial Variables
>
> useddta$white <- ifelse(useddta$race == 1, 1, 0)

> useddta$black <- ifelse(useddta$race == 2, 1, 0)

> useddta$other <- ifelse(useddta$race == 3, 1, 0)

> nmdta <- useddta[complete.cases(useddta),] #no missing data

> #summarize data
> summary(nmdta$mntlhlth)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.000   0.000   0.000   3.991   5.000  30.000

> summary(nmdta$inc1k)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.245  15.921  28.157  39.404  48.475 155.140

> summary(nmdta$age)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  18.00   31.00   42.00   41.81   51.00   84.00
```

```
> table(nmdta$female)

  0   1 
381 369 

> table(nmdta$white)

  0   1 
160 590 

> #Run OLS
> mntlhlth.model <- lm(formula = mntlhlth ~ age + female + black + other + 
educ + inc1k, data = nmdta )

> (summary(mntlhlth.model))

Call:
lm(formula = mntlhlth ~ age + female + black + other + educ + 
    inc1k, data = nmdta)

Residuals:
   Min     1Q Median     3Q    Max 
-6.888 -4.069 -2.818  0.601 27.711 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)  6.911999   1.656997   4.171 3.38e-05 ***
age         -0.009268   0.020120  -0.461  0.64517    
female       1.017865   0.528007   1.928  0.05427 .  
black       -2.119554   0.790024  -2.683  0.00746 ** 
other        0.569797   0.955038   0.597  0.55094    
educ        -0.188625   0.102889  -1.833  0.06716 .  
inc1k       -0.004487   0.007959  -0.564  0.57309    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 7.105 on 743 degrees of freedom
Multiple R-squared:  0.02043, Adjusted R-squared:  0.01252 
F-statistic: 2.582 on 6 and 743 DF,  p-value: 0.01753


> exp(mntlhlth.model$coefficients)
 (Intercept)          age       female        black        other         educ 
inc1k 
1004.2528741    0.9907743    2.7672800    0.1200852    1.7679076    0.8280966 
0.9955230 

> exp(confint(mntlhlth.model))
                 2.5 %       97.5 %
(Intercept) 38.82429117 2.597662e+04
age          0.95240335 1.030691e+00
female       0.98147563 7.802373e+00
black        0.02546378 5.663126e-01
other        0.27114621 1.152698e+01
educ         0.67664131 1.013453e+00
inc1k        0.98008883 1.011200e+00
```

```
> #TASK 1: Bayesian Binary Logit Model With Non-Informative Priors
> library(MCMCpack)

> mcmc.model = MCMCregress(mntlhlth ~ age + female + black + other + educ +
inc1k,
+                           data = nmdta, burnin = 1000, mcmc = 10000, thin =
1,seed = 47304,b0 = 0, B0 = 1e-6,marginal.likelihood = "Chib95")

> summary(mcmc.model)

Iterations = 1001:11000
Thinning interval = 1
Number of chains = 1
Sample size per chain = 10000

1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:

                 Mean       SD  Naive SE Time-series SE
(Intercept)  6.905019 1.652458 1.652e-02      1.636e-02
age         -0.009368 0.020364 2.036e-04      2.036e-04
female       1.018804 0.523410 5.234e-03      5.311e-03
black       -2.133636 0.778937 7.789e-03      7.936e-03
other        0.578391 0.954541 9.545e-03      9.545e-03
educ        -0.187699 0.102528 1.025e-03      1.025e-03
inc1k       -0.004541 0.007927 7.927e-05      7.927e-05
sigma2      50.605190 2.605665 2.606e-02      2.665e-02

2. Quantiles for each variable:

                2.5%       25%       50%        75%    97.5%
(Intercept)  3.65776  5.798106  6.901763  8.0250967 10.14658
age         -0.04969 -0.023026 -0.009389  0.0043824  0.03062
female       0.00458  0.664799  1.015528  1.3744351  2.06364
black       -3.65706 -2.655721 -2.131216 -1.6082498 -0.60899
other       -1.31836 -0.063896  0.582669  1.2131549  2.45338
educ        -0.38943 -0.257037 -0.187437 -0.1190083  0.01290
inc1k       -0.02015 -0.009885 -0.004572  0.0008091  0.01081
sigma2      45.79020 48.807843 50.525690 52.3088911 55.89927


> #TASK 2 Interpret Results Using Credit Intervals
> #Interpret data from summary
>
> #Task 3 Run Bayesian Logit With Informative Pairs
> #The informative priors is used for all the covariates
> mcmc.model2 = MCMCregress(mntlhlth ~ age + female + black + other + educ +
inc1k,
+                           data = nmdta, burnin = 1000, mcmc = 10000, thin =
1,seed = 47304,b0 = 1, B0 = .001,marginal.likelihood = "Chib95")

> summary(mcmc.model2)

Iterations = 1001:11000
Thinning interval = 1
Number of chains = 1
Sample size per chain = 10000
```

```
1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:

                 Mean       SD  Naive SE Time-series SE
(Intercept)  6.888026 1.650147 1.650e-02      1.633e-02
age         -0.009262 0.020356 2.036e-04      2.036e-04
female       1.019319 0.523323 5.233e-03      5.310e-03
black       -2.130668 0.778669 7.787e-03      7.934e-03
other        0.580484 0.954069 9.541e-03      9.541e-03
educ        -0.186841 0.102432 1.024e-03      1.024e-03
inc1k       -0.004557 0.007927 7.927e-05      7.927e-05
sigma2      50.604888 2.605638 2.606e-02      2.665e-02

2. Quantiles for each variable:

                 2.5%       25%       50%        75%    97.5%
(Intercept)  3.644594  5.782807  6.884677  8.0060721 10.12558
age         -0.049572 -0.022921 -0.009287  0.0044762  0.03071
female       0.005248  0.665446  1.016017  1.3748670  2.06403
black       -3.653532 -2.652605 -2.128643 -1.6055065 -0.60681
other       -1.315529 -0.061339  0.584827  1.2151737  2.45434
educ        -0.388332 -0.256109 -0.186577 -0.1182027  0.01368
inc1k       -0.020160 -0.009902 -0.004587  0.0007921  0.01080
sigma2      45.788474 48.808543 50.525504 52.3091770 55.89620

> BF = BayesFactor(mcmc.model,mcmc.model2)

> BF
The matrix of Bayes Factors is:
            mcmc.model mcmc.model2
mcmc.model    1.00e+00    3.25e-11
mcmc.model2   3.08e+10    1.00e+00

The matrix of the natural log Bayes Factors is:
            mcmc.model mcmc.model2
mcmc.model         0.0       -24.2
mcmc.model2       24.2         0.0

 mcmc.model :
   call = 
MCMCregress(formula = mntlhlth ~ age + female + black + other + 
    educ + inc1k, data = nmdta, burnin = 1000, mcmc = 10000, 
    thin = 1, seed = 47304, b0 = 0, B0 = 1e-06, marginal.likelihood = 
"Chib95")

   log marginal likelihood =  -2602.745 


 mcmc.model2 :
   call = 
MCMCregress(formula = mntlhlth ~ age + female + black + other + 
    educ + inc1k, data = nmdta, burnin = 1000, mcmc = 10000, 
    thin = 1, seed = 47304, b0 = 1, B0 = 0.001, marginal.likelihood = 
"Chib95")

   log marginal likelihood =  -2578.594
```

```
> #Task 4 Produce Prediction
> #35-Year-Old White female with college education and sample median income
> library(HDInterval)

> require(rjags)

> library(BEST)

> post.mat = as.matrix(mcmc.model)

> cut.mat = post.mat[,c("(Intercept)", "age",
'female','black','other','educ','inc1k')]

> pred.hyp <- 1*post.mat[,"(Intercept)"]+ 35*post.mat[,'age'] +
1*post.mat[,'female'] +
+   0*post.mat[,'black'] + 0*post.mat[,'other'] + 16*post.mat[,'educ'] +
median(nmdta$inc1k)*post.mat[,'inc1k']

> summary(pred.hyp)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  2.630   4.139   4.465   4.465   4.788   6.314

> quantile(pred.hyp, c(.025,.5,.975))
    2.5%      50%    97.5%
3.527574 4.464752 5.400748

> #Task 5 Calculate the Difference in Prediction
> #Prediction for 35-year-old white male with college education and median
income
> pred.hyp2 <- 1*post.mat[,"(Intercept)"]+ 35*post.mat[,'age'] +
0*post.mat[,'female'] +
+   0*post.mat[,'black'] + 0*post.mat[,'other'] + 16*post.mat[,'educ'] +
median(nmdta$inc1k)*post.mat[,'inc1k']

> summary(pred.hyp2)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  1.557   3.117   3.451   3.446   3.776   5.197

> #Differences in Predicted Outcomes
> quantile(pred.hyp-pred.hyp2, c(.025,.5,.975))
       2.5%         50%       97.5%
0.004579817 1.015528297 2.063644188

> #Close log
> save(nmdta, file = "Assignment_10.rdata")

> sink()
```