OM386
Assignment 1
Due: February 10th, 11:59pm

By: Callie Gilmore (cgg756)

Random Effects and Hierarchical Linear Models

In this exercise, we will use hierarchical linear models and regressions with random effects for an analytics problem from a credit card company. The credit card company would like to figure out whether offering more promotions (for example, gasoline rebates and coupons for using the credit card) to their existing customers can increase the share-of-wallet of the credit card (that is, the share of a consumer's monthly spending using the credit card in her total spending). The company would also like to figure out what customer characteristics make them more responsive to promotions.

The company conducted a field experiment by randomly selecting 300 customers and offering them different monthly promotions for 12 months. The share-of-wallet data were recorded in each month for every customer. The data set also included some consumer characteristics. Please download the data "CreditCard_SOW_Data.csv" from Canvas. It has the following variables:

ConsumerID	ID's of the sampled consumers
History	How long (number of months) the customer has been using the card before the experiment
Income	The customer's annual income
WalletShare	The card's share of wallet in the consumer's total monthly spending
Promotion	Index of monthly promotion activity –higher index indicates more pomotions
Balance	The customer's unpaid balance at the beginning of the month

1). Please read the data into R and create a data frame named "sow.data". Please convert consumer ID's to factors and create the following 2 variables in the data frame: logIncome = log(Income) and logSowRatio = log(WalletShare/(1-WalletShare)).

```
> sow.data <- read.csv(file = "CreditCard_SOW_data.csv", header=T)
> sow.data$ConsumerID = as.factor(sow.data$ConsumerID)
> sow.data$logIncome = log(sow.data$Income)
> sow.data$logSowRatio = log((sow.data$WalletShare/(1-sow.data$WalletShare)))
```

2). Use the function lm( ) to run the regression

$$logSowRatio_{ij} = \beta_0 + \beta_1 \times History_i + \beta_2 \times Balance_{ij} + \beta_3 \times Promotion_{ij} + \beta_4 \times History_i \times Promotion_{ij} + \beta_5 \times logIncome_i \times Promotion_{ij} + \varepsilon_{ij}$$

Copy and paste the results here.

```
> ## Part 2
> sow.lm = lm(logSowRatio~History+Balance+Promotion+History:Promotion+logIncome:Promotion, data=sow.data)
> summary(sow.lm)

Call:
lm(formula = logSowRatio ~ History + Balance + Promotion + History:Promotion +
    logIncome:Promotion, data = sow.data)

Residuals:
     Min       1Q   Median       3Q      Max
-0.59976 -0.14401  0.00153  0.13634  0.75883

Coefficients:
                      Estimate Std. Error  t value Pr(>|t|)
(Intercept)          8.908e-02  1.603e-02    5.558 2.92e-08 ***
History              1.039e-02  4.153e-04   25.027  < 2e-16 ***
Balance             -4.959e-04  2.882e-06 -172.064  < 2e-16 ***
Promotion            7.777e-01  1.888e-01    4.120 3.87e-05 ***
History:Promotion   -2.598e-03  5.722e-04   -4.541 5.79e-06 ***
Promotion:logIncome -4.558e-02  1.651e-02   -2.760  0.00581 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2078 on 3594 degrees of freedom
Multiple R-squared:  0.8984,    Adjusted R-squared:  0.8982
F-statistic:  6353 on 5 and 3594 DF,  p-value: < 2.2e-16

> AIC(sow.lm)
[1] -1087.389
> BIC(sow.lm)
[1] -1044.069
```

3). Estimate the following hierarchical linear model using the function lmer( ) in the R package "lme4"

$$logSowRatio_{ij} = \beta_{0i} + \beta_1 \times Balance_{ij} + \beta_{2i} \times Promotion_{ij} + \varepsilon_{ij}$$

$$\beta_{0i} = \mu_0 + \mu_1 \times History_i + \zeta_i$$

$$\beta_{2i} = \gamma_0 + \gamma_1 \times History_i + \gamma_2 \times logIncome_i + \xi_i$$

Following what we did in our class, please rewrite this hierarchical linear model as a one-level linear regression model with random effects.

logSowRatio = $\mu_0 + \zeta_i + \mu_1 History + \gamma_0 Promotion_{ij} + \xi_i Promotion_i + \gamma_1 History_i * Promotion_{ij} + \gamma_2 LogIncome_i * Promotion_{ij} + \beta_1 Balance_{ij} + \varepsilon_{ij}$

Which variables (and interactions) in the regression have fixed effects? Which ones have random effects? Specify the variables in lmer() and run the regression (please specify REML=F, control=lmerControl(optimizer ="Nelder_Mead") in lmer()). Please copy and paste the summary() of the regression here.

```
> summary(sow.lmer)
Linear mixed model fit by maximum likelihood  ['lmerMod']
Formula: logSowRatio ~ History + Balance + Promotion + History:Promotion +
    logIncome:Promotion + (1 + Promotion | ConsumerID)
   Data: sow.data
Control: lmerControl(optimizer = "Nelder_Mead")

     AIC      BIC   logLik deviance df.resid
 -6532.1  -6470.2   3276.0  -6552.1     3590

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.1063 -0.6424  0.0049  0.6336  3.4532

Random effects:
 Groups     Name        Variance  Std.Dev. Corr
 ConsumerID (Intercept) 0.0359421 0.18958
            Promotion   0.0005355 0.02314  0.06
 Residual               0.0066071 0.08128
Number of obs: 3600, groups:  ConsumerID, 300

Fixed effects:
                      Estimate Std. Error  t value
(Intercept)          9.595e-02  2.655e-02    3.613
History              1.039e-02  7.135e-04   14.569
Balance             -5.003e-04  1.799e-06 -278.110
Promotion            6.129e-01  1.466e-01    4.181
History:Promotion   -2.571e-03  2.402e-04  -10.703
Promotion:logIncome -3.110e-02  1.288e-02   -2.414

Correlation of Fixed Effects:
            (Intr) Histry Balanc Promtn Hstr:P
History     -0.900
Balance     -0.107 -0.001
Promotion   -0.011  0.009  0.013
Hstry:Prmtn  0.143 -0.159 -0.002 -0.153
Prmtn:lgInc  0.001  0.000 -0.012 -0.998  0.099
```

```
> AIC(sow.lmer)
[1] -6532.094
> BIC(sow.lmer)
[1] -6470.207
```

Fixed Effects: Intercept, History, Balance, Promotion, History:Promotion, and Promotion:logIncome

Random Effects: ConsumerID (Intercept), Promotion and Residual

Please interpret the estimated fixed effects in the regression.

History: increased the logSowRatio by .001039 per one month using the card

Balance: decreased the logSowRatio by .000503 per one unit increase

Promotion: increased the logSowRatio by .06129 per one unit increase

History:Promotion: decreased the logSowRatio by .002571

History:logIncome: decreased the logSowRatio by .0311

Compare model fit using AIC() and BIC() with the model in (2).

When comparing the AIC and BICs from the 2 models, the 2nd model performed much better because it had much lower AIC and BIC values. The AIC and BIC respectively for the first model were -1087.389 and -1044.069 while the second model had an AIC and BIC respectively of -6532.1 and -6470.2.

```
> AIC(sow.lm)
[1] -1087.389
> BIC(sow.lm)
[1] -1044.069
```

```
> AIC(sow.lmer)
[1] -6532.094
> BIC(sow.lmer)
[1] -6470.207
```

Linear and Hierarchical Linear Models: Bayesian Estimation

In this exercise, we will practice Bayesian estimation for hierarchical linear models and regressions with random effects using the same dataset "CreditCard_SOW_Data.csv".

1). Use the function MCMCregress() in the R package "MCMCpack" to estimate the linear regression

$$logSowRatio_{ij} = \beta_0 + \beta_1 \times History_i + \beta_2 \times Income_i + \beta_3 \times Balance_{ij} + \beta_4 \times Promotion_{ij} + \varepsilon_{ij}$$

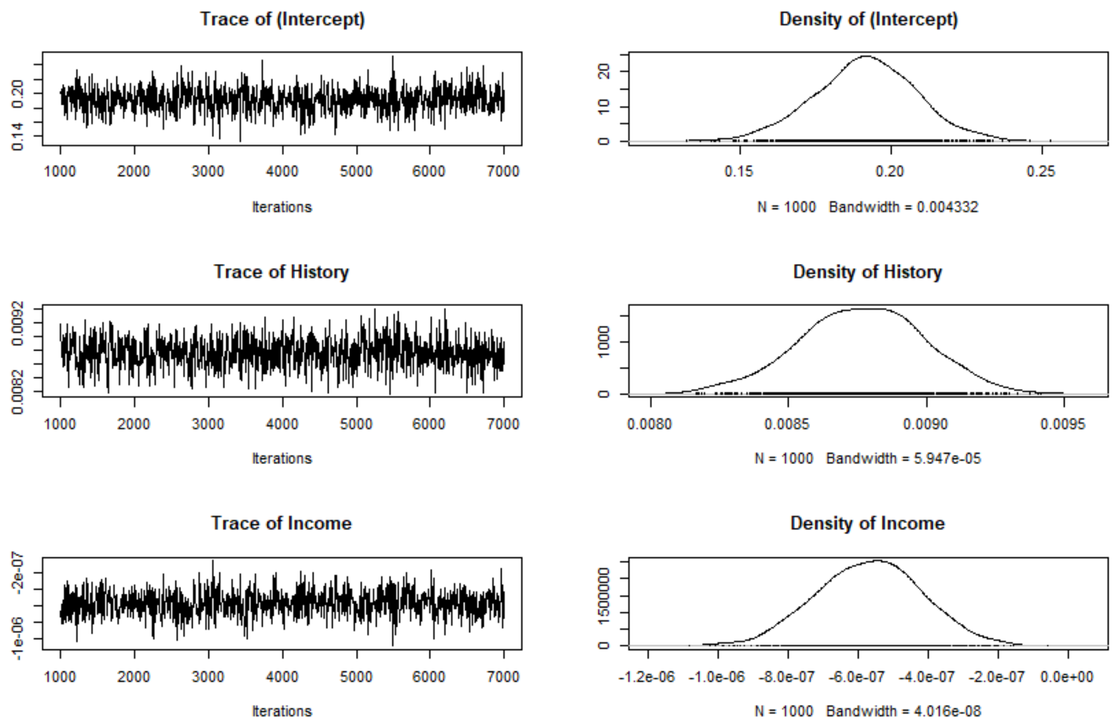
Use the summary() function to find the results of the estimation. Copy and pastes the results here.

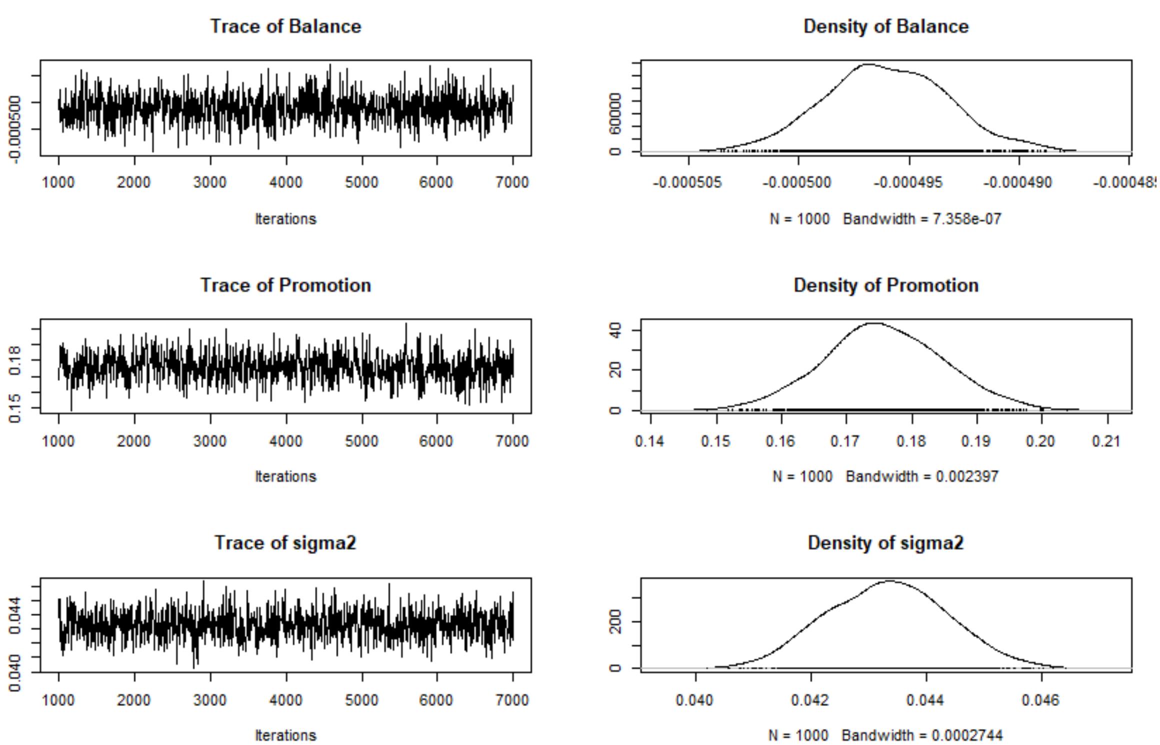
```
> sow.ba1 = MCMCregress(logSowRatio~History + Income + Balance + Promotion, mcm=6000, thin=6, data=sow.data)
> summary(sow.ba1)

Iterations = 1001:6995
Thinning interval = 6
Number of chains = 1
Sample size per chain = 1000

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

                  Mean        SD  Naive SE Time-series SE
(Intercept)  1.915e-01 1.699e-02 5.372e-04      5.372e-04
History      8.765e-03 2.233e-04 7.063e-06      7.063e-06
Income      -5.682e-07 1.508e-07 4.770e-09      5.041e-09
Balance     -4.960e-04 2.776e-06 8.780e-08      8.780e-08
Promotion    1.757e-01 9.001e-03 2.846e-04      2.998e-04
sigma2       4.332e-02 1.031e-03 3.259e-05      3.259e-05

2. Quantiles for each variable:

                  2.5%        25%        50%        75%      97.5%
(Intercept)  1.572e-01  1.812e-01  1.919e-01  2.030e-01  2.250e-01
History      8.293e-03  8.614e-03  8.769e-03  8.921e-03  9.189e-03
Income      -8.602e-07 -6.685e-07 -5.645e-07 -4.636e-07 -2.739e-07
Balance     -5.014e-04 -4.979e-04 -4.961e-04 -4.942e-04 -4.902e-04
Promotion    1.587e-01  1.697e-01  1.754e-01  1.819e-01  1.938e-01
sigma2       4.141e-02  4.256e-02  4.333e-02  4.400e-02  4.531e-02
```

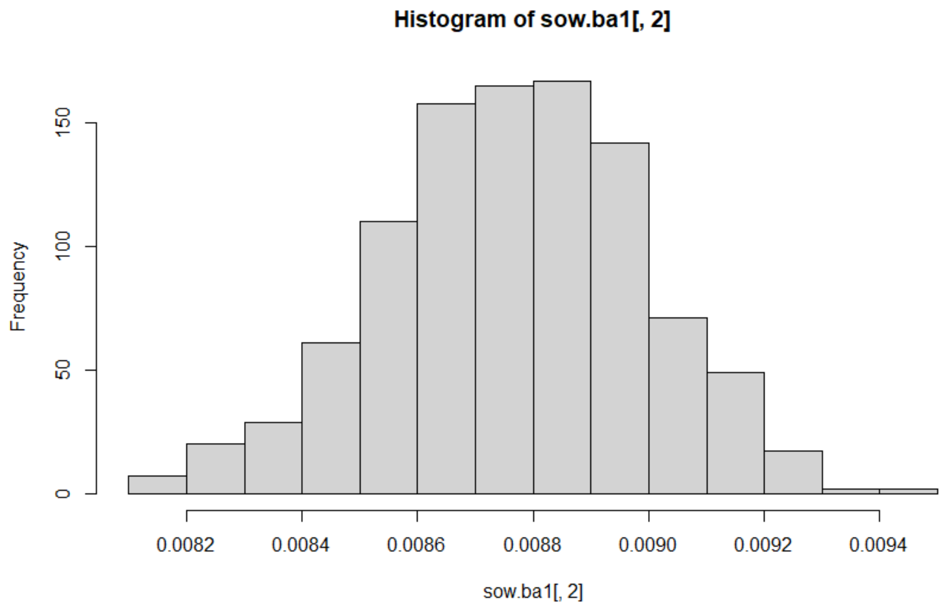
From the Bayesian posterior intervals (use 2.5% and 97.5% quantiles of the simulated posterior distributions), are regression coefficients significant at the 5% level?

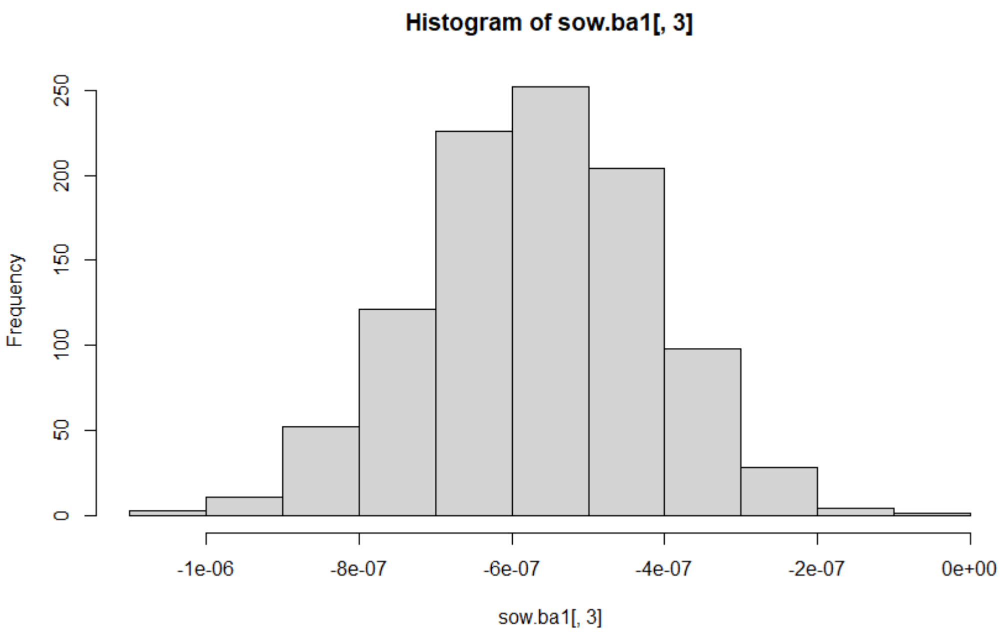
Yes, the regression coefficients are significant at the 5% level because none of the coefficients include 0 in the 2.5%-97.5% quantiles.

Use the plot() function to plot the posterior sampling chains and hist() to plot the posterior densities (histograms) for β_2 and β_3; copy and paste the results here.


Trace of (Intercept)
Iterations
Density of (Intercept)
N = 1000 Bandwidth = 0.004332
Trace of History
Iterations
Density of History
N = 1000 Bandwidth = 5.947e-05
Trace of Income
Iterations
Density of Income
N = 1000 Bandwidth = 4.016e-08


Trace of Balance
Iterations
Density of Balance
N = 1000 Bandwidth = 7.358e-07
Trace of Promotion
Iterations
Density of Promotion
N = 1000 Bandwidth = 0.002397
Trace of sigma2
Iterations
Density of sigma2
N = 1000 Bandwidth = 0.0002744





2). For the hierarchical linear model below,

$$logSowRatio_{ij} = \beta_{0i} + \beta_1 \times Balance_{ij} + \beta_{2i} \times Promotion_{ij} + \varepsilon_{ij}$$

$$\beta_{0i} = \mu_0 + \mu_1 \times History_i + \zeta_i$$

$\beta_{2i} = \gamma_0 + \gamma_1 \times History_i + \gamma_2 \times Income_i + \xi_i$

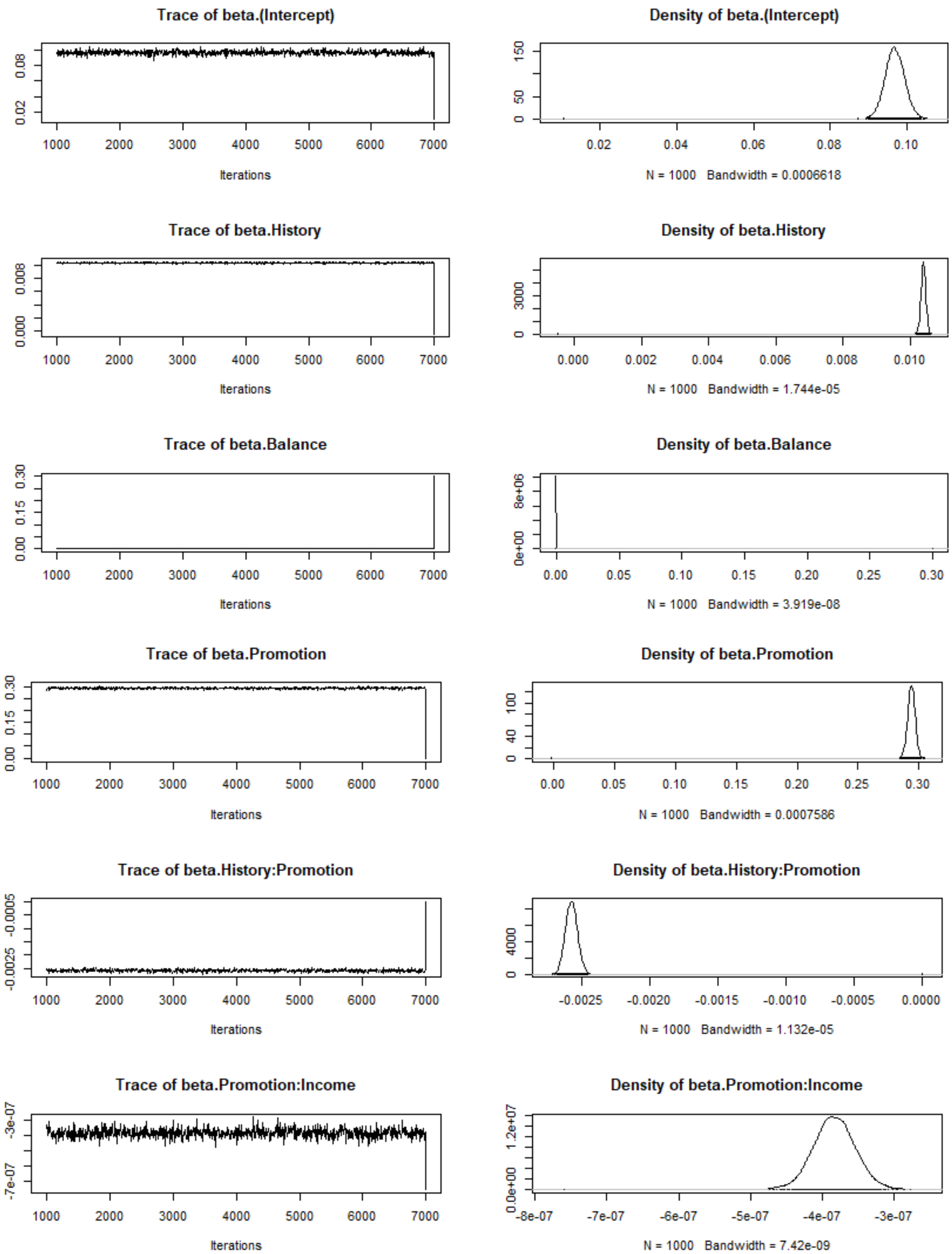
use the function MCMChregress( ) in the R package "MCMCpack" for its Bayesian estimation.

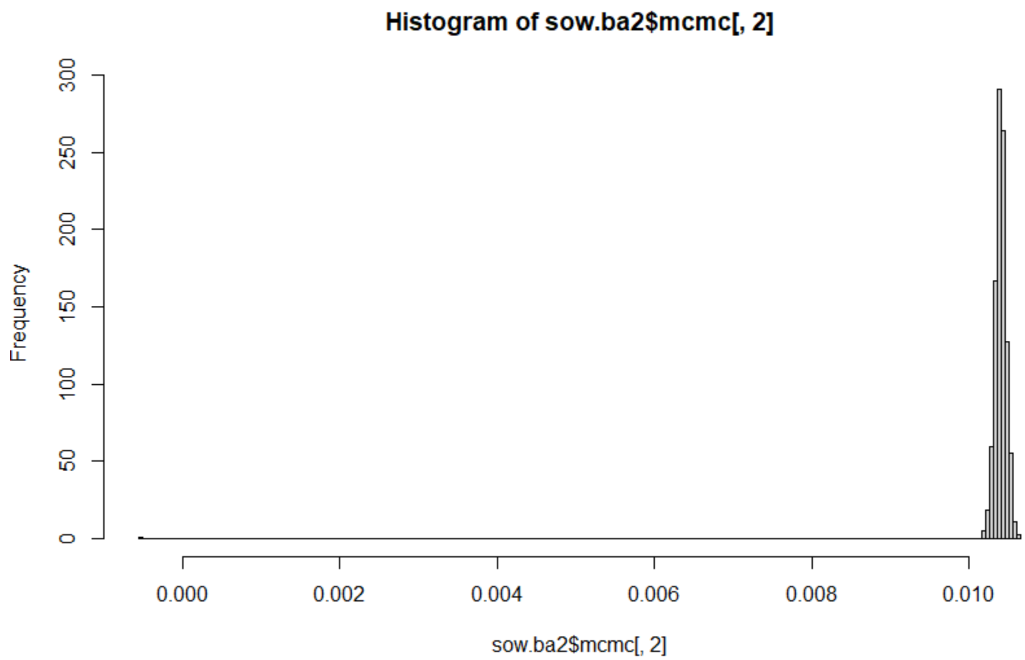
Please copy and paste the Bayesian estimation results of the fixed effects (same fixed effects as in (3)) in the model using summary("*yourBayesianModelName*"$mcmc[,1:6]). From the Bayesian posterior intervals, are the fixed effects significant at the 5% level?

```
> sow.ba2 = MCMChregress(fixed=logSowRatio~History+Balance+Promotion
+                       +History:Promotion+Income:Promotion,
+                       random=~Promotion, group="ConsumerID", mcmc=6000,
+                       thin=6, data=sow.data, r=3, R=diag(2))

Running the Gibbs sampler. It may be long, keep cool :)

**********:10.0%
**********:20.0%
**********:30.0%
**********:40.0%
**********:50.0%
**********:60.0%
**********:70.0%
**********:80.0%
**********:90.0%
**********:100.0%
> summary(sow.ba2$mcmc[,1:6])

Iterations = 1001:6995
Thinning interval = 6
Number of chains = 1
Sample size per chain = 1000

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

                             Mean        SD  Naive SE Time-series SE
beta.(Intercept)        9.677e-02 3.697e-03 1.169e-04      1.169e-04
beta.History            1.038e-02 3.513e-04 1.111e-05      1.111e-05
beta.Balance           -2.004e-04 9.499e-03 3.004e-04      3.004e-04
beta.Promotion          2.935e-01 9.836e-03 3.110e-04      3.110e-04
beta.History:Promotion -2.572e-03 9.146e-05 2.892e-06      2.892e-06
beta.Promotion:Income  -3.837e-07 3.131e-08 9.901e-10      9.901e-10

2. Quantiles for each variable:

                             2.5%        25%        50%        75%      97.5%
beta.(Intercept)        9.196e-02  9.521e-02  9.678e-02  9.854e-02  1.020e-01
beta.History            1.025e-02  1.035e-02  1.039e-02  1.044e-02  1.053e-02
beta.Balance           -5.011e-04 -5.009e-04 -5.008e-04 -5.007e-04 -5.005e-04
beta.Promotion          2.876e-01  2.919e-01  2.938e-01  2.958e-01  2.997e-01
beta.History:Promotion -2.656e-03 -2.604e-03 -2.575e-03 -2.547e-03 -2.489e-03
beta.Promotion:Income  -4.401e-07 -4.020e-07 -3.838e-07 -3.647e-07 -3.239e-07
```

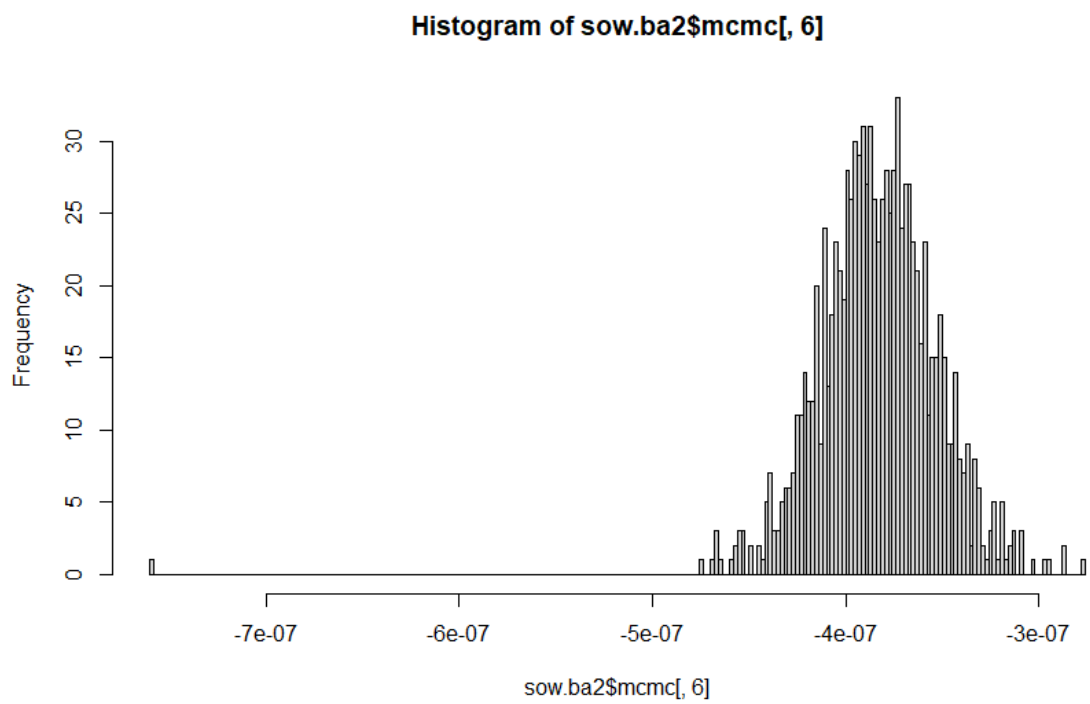
Yes, the fixed effects variables are significant at 5% because they do not include zero and contain the same signs between 2.5% and 97.5%.

Use the plot() and hist() function to plot the posterior sampling chains and posterior densities for μ_1 and γ_2; copy and paste the results here.


Trace of beta.(Intercept)
Iterations
Density of beta.(Intercept)
N = 1000 Bandwidth = 0.0006618
Trace of beta.History
Iterations
Density of beta.History
N = 1000 Bandwidth = 1.744e-05
Trace of beta.Balance
Iterations
Density of beta.Balance
N = 1000 Bandwidth = 3.919e-08
Trace of beta.Promotion
Iterations
Density of beta.Promotion
N = 1000 Bandwidth = 0.0007586
Trace of beta.History:Promotion
Iterations
Density of beta.History:Promotion
N = 1000 Bandwidth = 1.132e-05
Trace of beta.Promotion:Income
Iterations
Density of beta.Promotion:Income
N = 1000 Bandwidth = 7.42e-09



Histogram of sow.ba2$mcmc[, 2]
Frequency
0
50
100
150
200
250
300
0.000
0.002
0.004
0.006
0.008
0.010
sow.ba2$mcmc[, 2]



Histogram of sow.ba2$mcmc[, 6]
Frequency
0
5
10
15
20
25
30
-7e-07
-6e-07
-5e-07
-4e-07
-3e-07
sow.ba2$mcmc[, 6]