**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)).

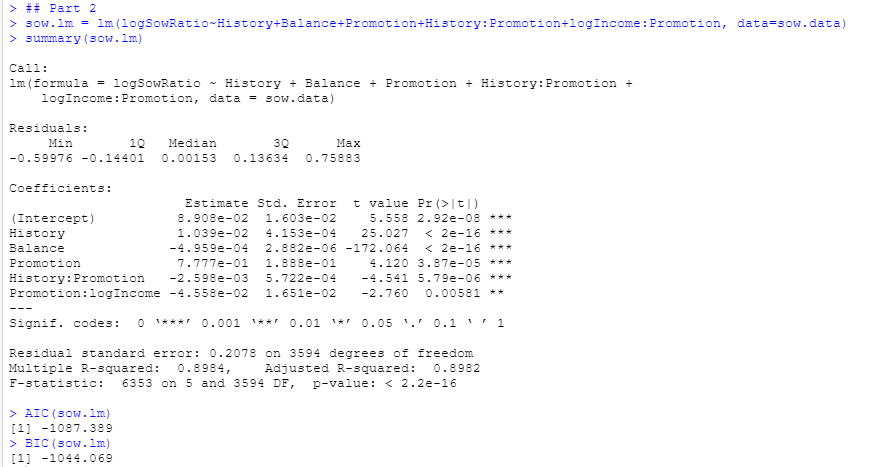


2). Use the function lm( ) to run the regression

*logSowRatioij = β0 + β1×Historyi + β2×Balanceij + β3×Promotionij +*

*β4×Historyi×Promotionij + β5×logIncomei×Promotionij + εij*

Copy and paste the results here.



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

*logSowRatioij = β0i + β1×Balanceij + β2i×Promotionij + εij*

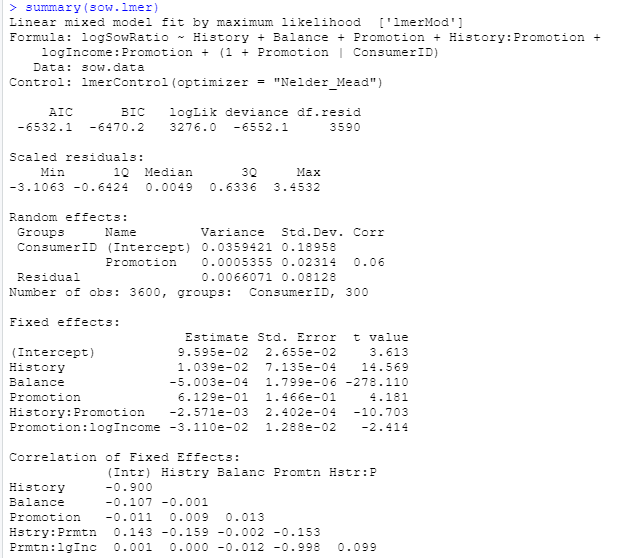
*β0i = μ0 +μ1×Historyi +ζi*

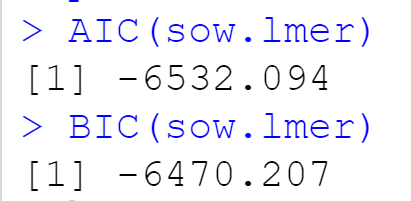
*β2i = γ0 +γ1×Historyi +γ2×logIncomei +ξi*

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

logSowRatio =

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.





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

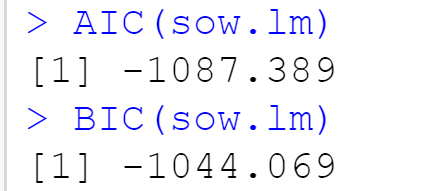
Promotion: increased the logSowRatio by .06129 per one unit increase

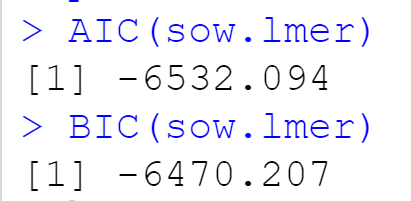
History:Promotion: decreased the logSowRatio by .002571

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.





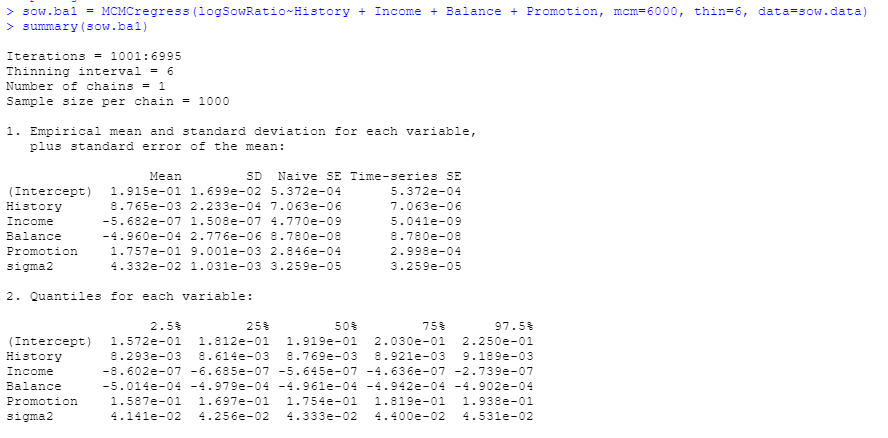
**Linear and Hierarchical Linear Models: Bayesian Estimation**

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

*logSowRatioij = β0 + β1×Historyi +β2×Incomei +β3×Balanceij +* *β4×Promotionij + εij*

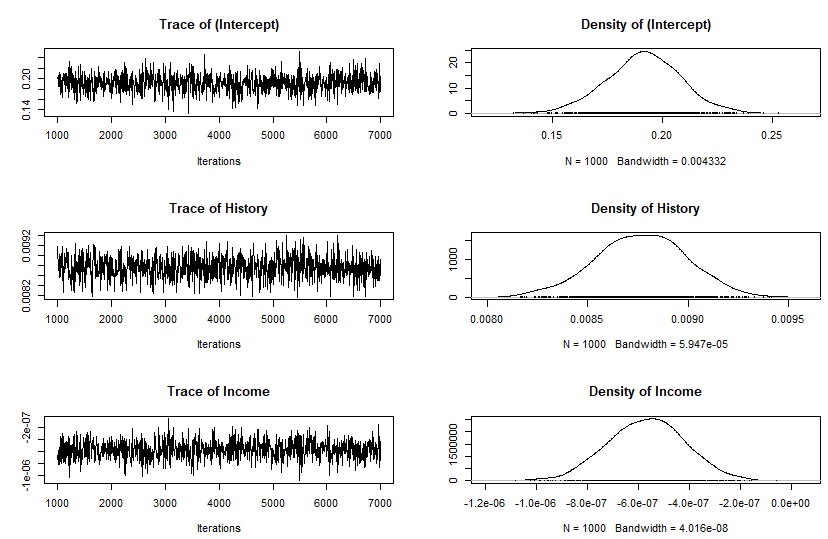
Use the summary() function to find the results of the estimation. Copy and pastes the results here.

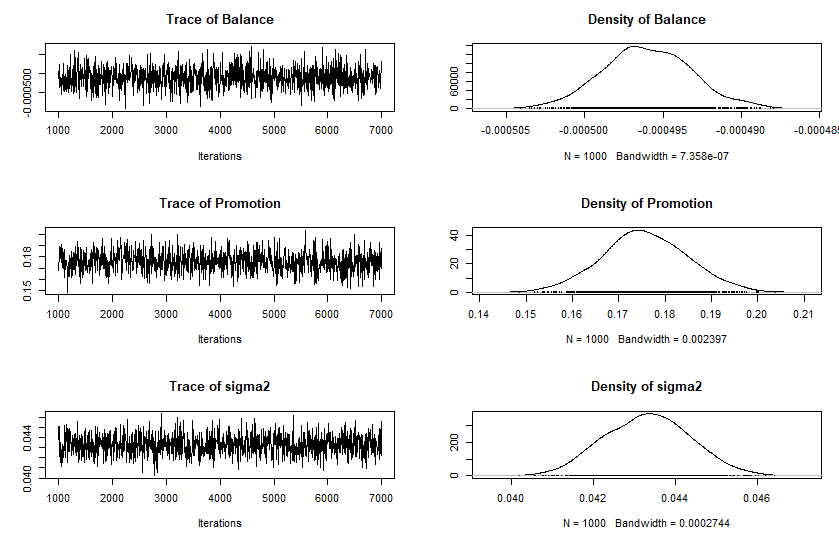


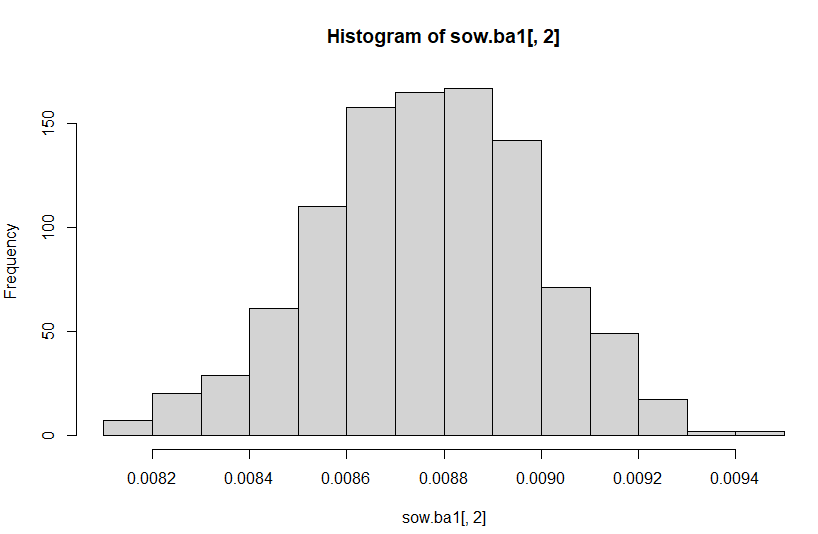
From the Bayesian posterior intervals (use 2.5% and 97.5% quantiles of the simulated posterior distributions), are regression coefficients significant at the 5% level?

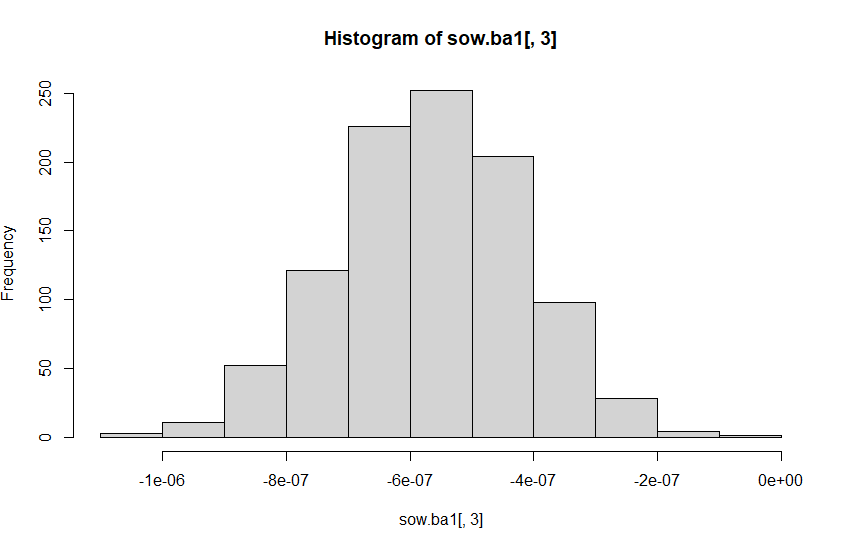
Yes, the regression coefficients are significant at the 5% level because none of the coefficients include 0 in the 2.5%-97.5% quantiles.

Use the plot() function to plot the posterior sampling chains and hist() to plot the posterior densities (histograms) for *β2* and *β3*; copy and paste the results here.









2).For the hierarchical linear model below,

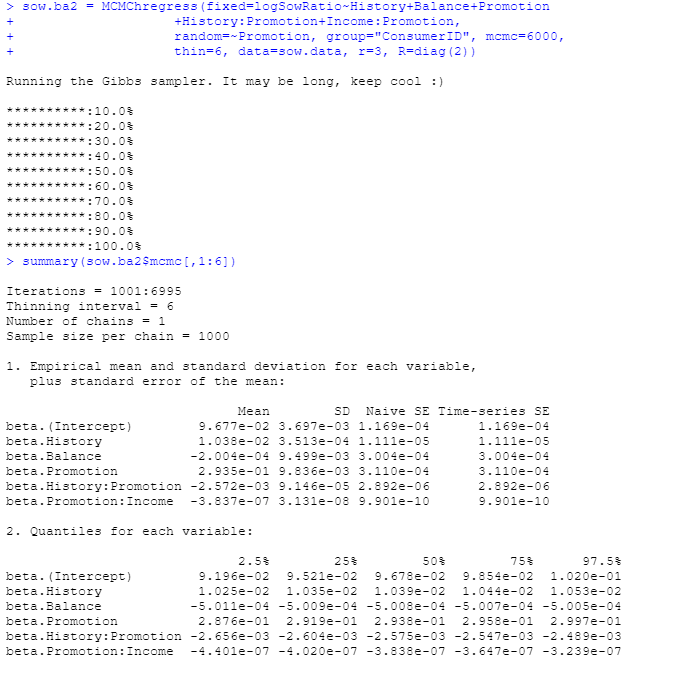
*logSowRatioij = β0i + β1×Balanceij + β2i×Promotionij + εij*

*β0i = μ0 +**μ1×Historyi +ζi*

*β2i = γ0 +γ1×Historyi +**γ2×Incomei +ξi*

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?



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 *μ1* and *γ2*; copy and paste the results here.

