**OM386 Advanced Data Analytics in Marketing**

**Assignment 1**

**Due: February 14th, 11:59pm**

**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("CreditCard\_SOW\_data.csv")*

*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))*

*head(sow.data)*

ConsumerID History Income WalletShare Promotion Balance logIncome logSowRatio

1 1 55 82000 0.643 0.5 836 11.31447 0.5884089

2 1 55 82000 0.628 0.2 467 11.31447 0.5236463

3 1 55 82000 0.567 1.0 1208 11.31447 0.2696216

4 1 55 82000 0.638 0.8 792 11.31447 0.5666941

5 1 55 82000 0.554 0.7 1215 11.31447 0.2168457

6 1 55 82000 0.573 1.1 1248 11.31447 0.2941017

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.

*reg1 = lm(logSowRatio ~ sow.data$History + Balance + Promotion + sow.data$History:Promotion + logIncome:Promotion, data = sow.data)*

*summary(reg1)*

Call:

lm(formula = logSowRatio ~ sow.data$History + Balance + Promotion +

sow.data$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 \*\*\*

sow.data$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 \*\*\*

sow.data$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

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.

*logSowRatioij = μ0 +μ1×Historyi +ζi + β1×Balanceij + (γ0 +γ1×Historyi +γ2×logIncomei +ξi ) X Promotionij +εij*

*logSowRatioij = μ0 +ζI + μ1×Historyi + γ0 Promotionij + ξi Promotionij + γ1Historyi X Promotionij + γ2×logIncomei X Promotionij + β1×Balanceij + εij*

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") or control=lmerControl(optimizer ="bobyqa") in lmer() and find out if the MLE method used by lmer() converge. Copy and paste the summary() of the model results here.

Fixed effect – History, Promotion, History\*Promotion,logIncome\*Promotion, Balance

Random effect – Promotion

MLE method used by lmer() converges.

Formula: logSowRatio ~ History + Promotion + History:Promotion + logIncome:Promotion +

Balance + (1 + Promotion | ConsumerID)

Data: sow.data

Control: lmerControl(optimizer = "bobyqa")

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) 0.0359422 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

Promotion 6.129e-01 1.466e-01 4.181

Balance -5.003e-04 1.799e-06 -278.110

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 Promtn Balanc Hstr:P

History -0.900

Promotion -0.011 0.009

Balance -0.107 -0.001 0.013

Hstry:Prmtn 0.143 -0.159 -0.153 -0.002

Prmtn:lgInc 0.001 0.000 -0.998 -0.012 0.099

Please interpret the estimated fixed effects in the regression.

Promotion is positively correlated with Wallet Share odds ratio with a coefficient of 61.29 but this reduces as the customer history increases; thus the older the customers the lesser they are influenced by promotions. Also, long-term customers have more wallet share. Wallet Share also reduces when balance increases.

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

AIC of first model -1087.389

AIC of second model -6532.1

The 2nd model performed better than the 1st.

**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.

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

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?

The coefficients are significant at the 5%; and as the signs do not change for each feature from 2.5% to 97.5%, the variables are significant

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.

1. Income

Diagram

Description automatically generated

Chart, histogram

Description automatically generated

1. Balance

A picture containing chart

Description automatically generated

Chart, histogram

Description automatically generated

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?

Iterations = 1001:6995

Thinning interval = 6

Number of chains = 1

Sample size per chain = 1000

Empirical mean and standard deviation for each variable,

plus standard error of the mean:

Mean SD Naive SE Time-series SE

beta.(Intercept) 9.673e-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

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

All variables seem significant as they do not change sign from 2.5% to 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.

1. History

Chart

Description automatically generated

Chart

Description automatically generated

1. Income and Promotion

Chart

Description automatically generated

Chart, histogram

Description automatically generated