**OM386 Marketing Analytics II**

**Assignment 1**

**Due: February 12th, 11:59pm**

**Khyathi Balusu**

**Kb42582**

**Linear Regression Analysis**

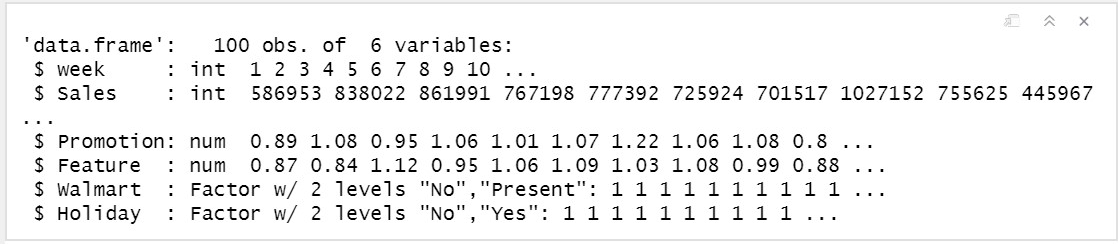
We will use a simple dataset to evaluate the impact of the opening of a new Walmart on the sales a local grocery store. Suppose that you have been hired as a consultant for the local grocery store. Store management is worried since Wal-Mart has entered the market by opening a "Wal-Mart Super-center" only 3 miles away. The management is interested in analyzing the impact on store sales after Wal-Mart's entry.

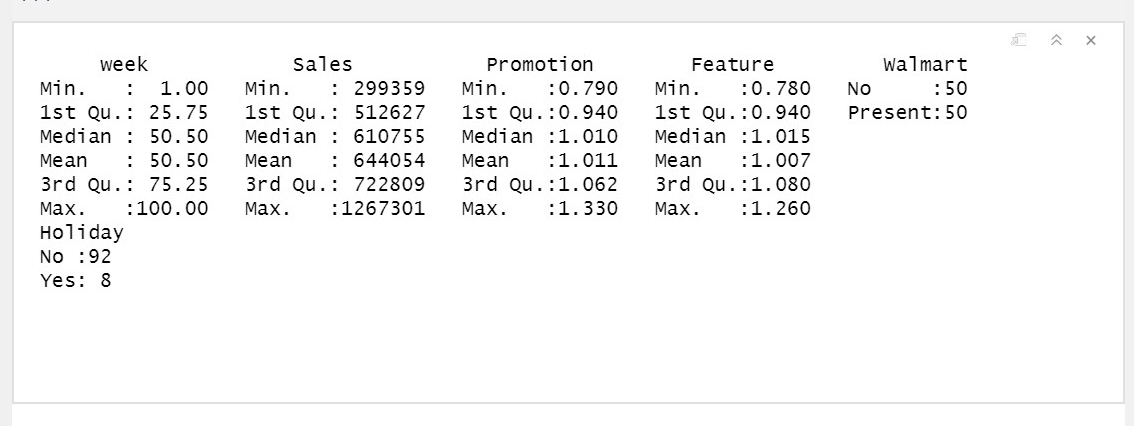
For the analysis, management has given you access to 50 weeks of sales data before the entry of Walmart and 50 weeks after. Please download and look at the data in “Walmart\_Data.csv” from Canvas.

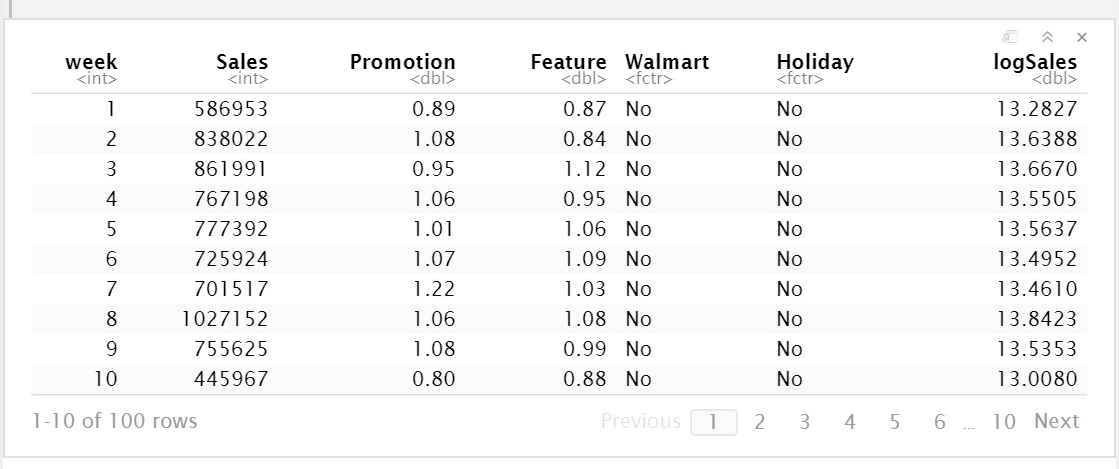
The dataset has the following variables:

|  |  |
| --- | --- |
| WEEK | Week number |
| Sales | weekly sales |
| Promotion | Index of weekly promotion activity –higher promotion index indicates more products on promotion in the store |
| Feature | Index of feature advertising activity – higher feature advertising index indicates more feature advertising |
| Walmart | A categorical variable = “No” in the weeks before the Walmart opens, and “Present” in the weeks before the Walmart opens |
| Holiday | Holiday = “Yes” during major holiday weeks, and “No” for non-holiday weeks |

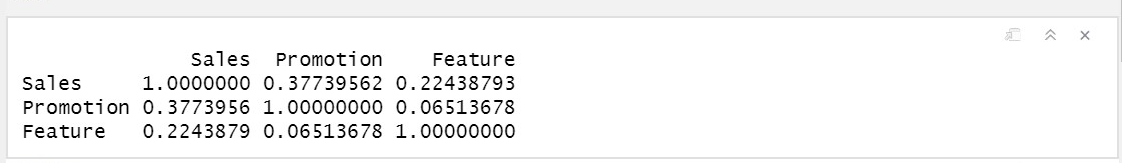
1. Pleas read the data into R. Use the function str( ) to find the structure of the data fame and the summary( ) to summarize the data. Please post the results here. Create a new variable called “logSales”, which is the logarithm of the variable “Sales” in the data frame “walmart”.



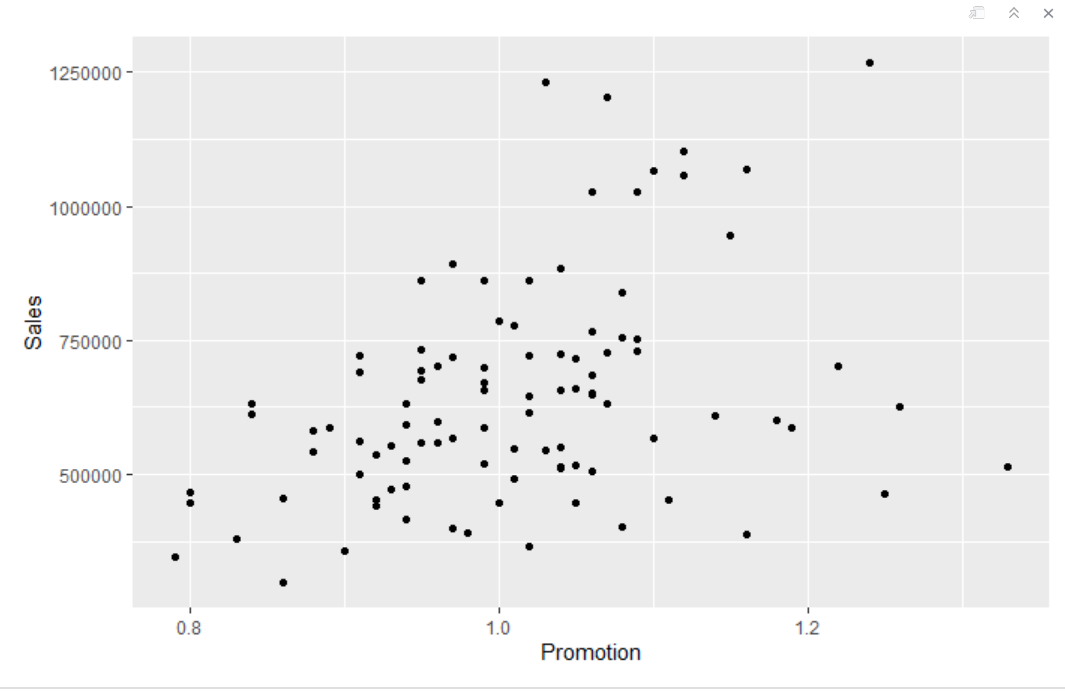




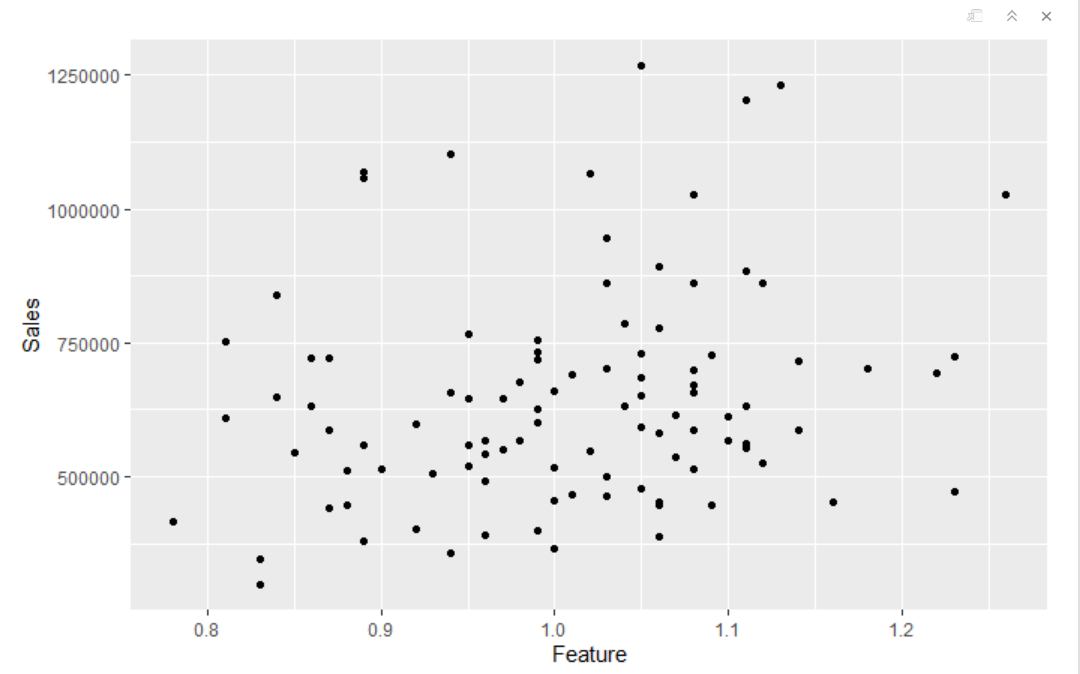
2). Use the correlation function cor( ) to find the pairwise correlation between the three variables, “Sales”, “Promotion” and “Feature”. Please post the resulting correlation matrix here. Create a scatter plot for “Sales” and “Promotion”. Make another scatter plot for “Sales” and “Feature”. Please post the plots here. Create histogram plots for both “Sales” and “logSales”. Please post the plots here.



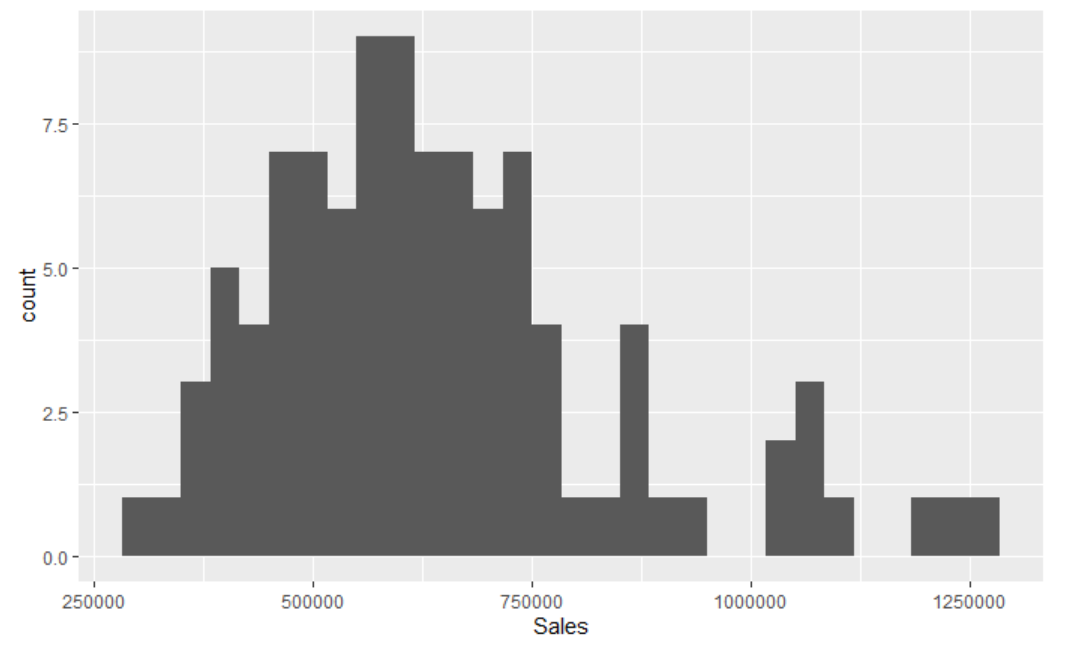
Sales vs Promotion:



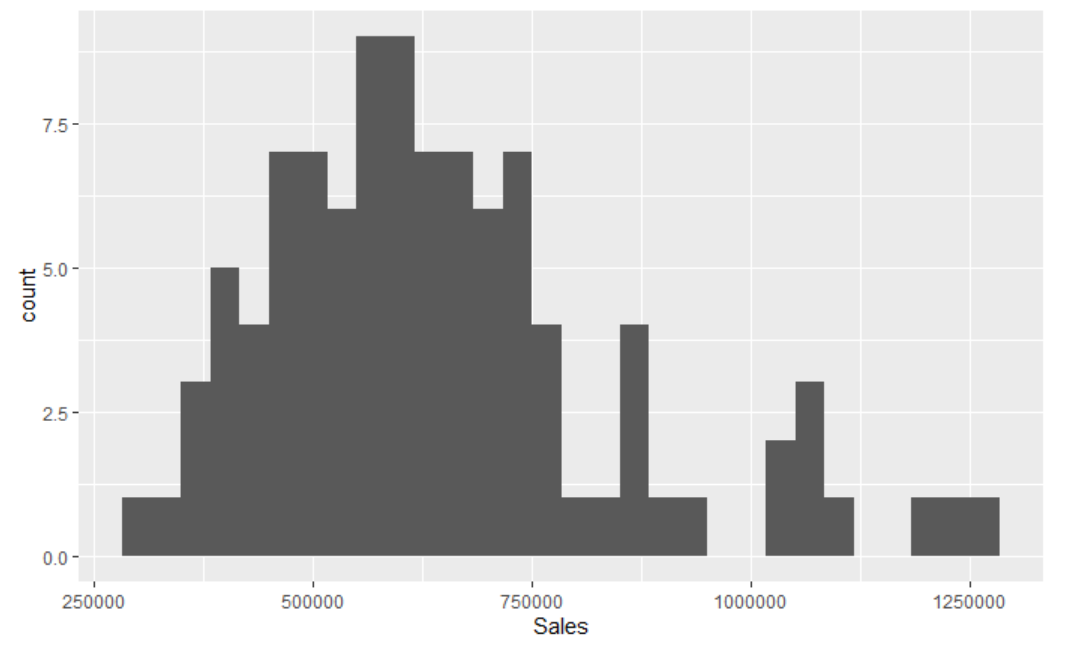
Sales Vs Feature:



Sales histogram:



logSales – histogram:



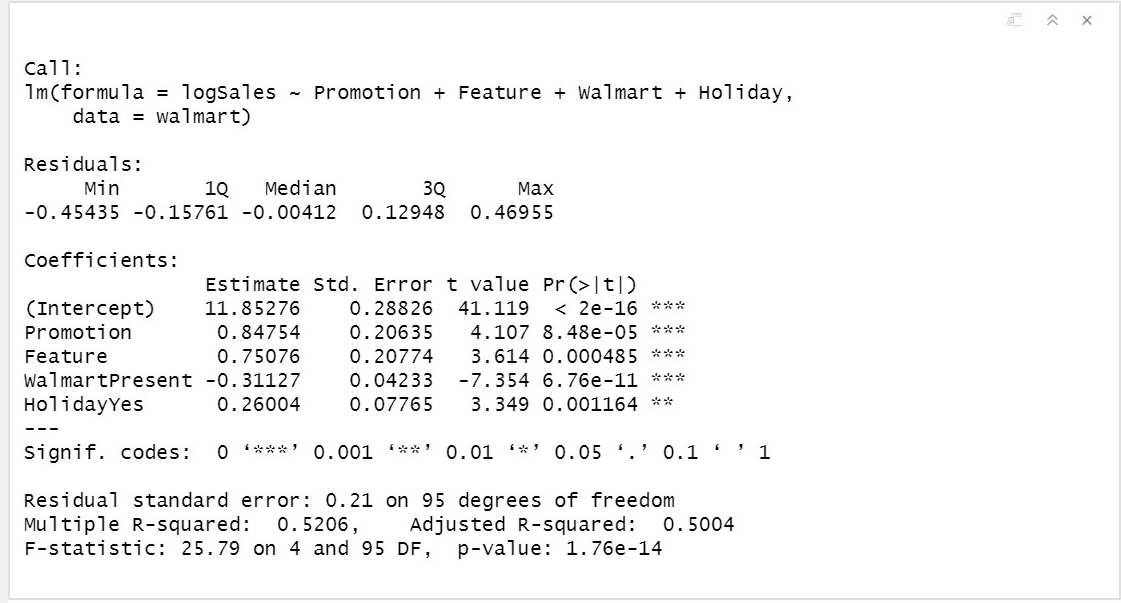
3). Estimate the following regression model using the functions lm ( ) and summary( )

*log(sales)* = *β*0 + *β*1×*Promotion* + *β*2×*Feature* + *β*3×*WalMart* + *β*4×*Holiday* + *error*

Paste the R regression output from summary( ) here.

Interpret the estimated coefficients *β*1, *β*2, *β*3, and *β*4.

Can we conclude the entry of Wal-mart affects the sales of the local store?



* **Exp(*β*):**  Increase the sales which occurs from a unit (one) increase in corresponding x variable.
* ***β*1:** Increase in sales of stores: exp(0.84754) – 1 = 1.333898 = 133.3898, this is due to unit (one) increase in promotion index
* ***β*2:** Increase in sales of stores: exp(0.75076) – 1 = 1.11861 = 111.861%, this is due to unit (one) increase in the feature index
* ***β*3:** Sales of stores having Walmart in vicinity: exp(-0.31127)-1 = -0.267 = 26.7% lesser than the sales of stores that don’t have Walmart nearby.
* ***β*4:** Increase in sales of stores: exp(0.26004) – 1 = 0.296982 = 29.6982% when it is a holiday

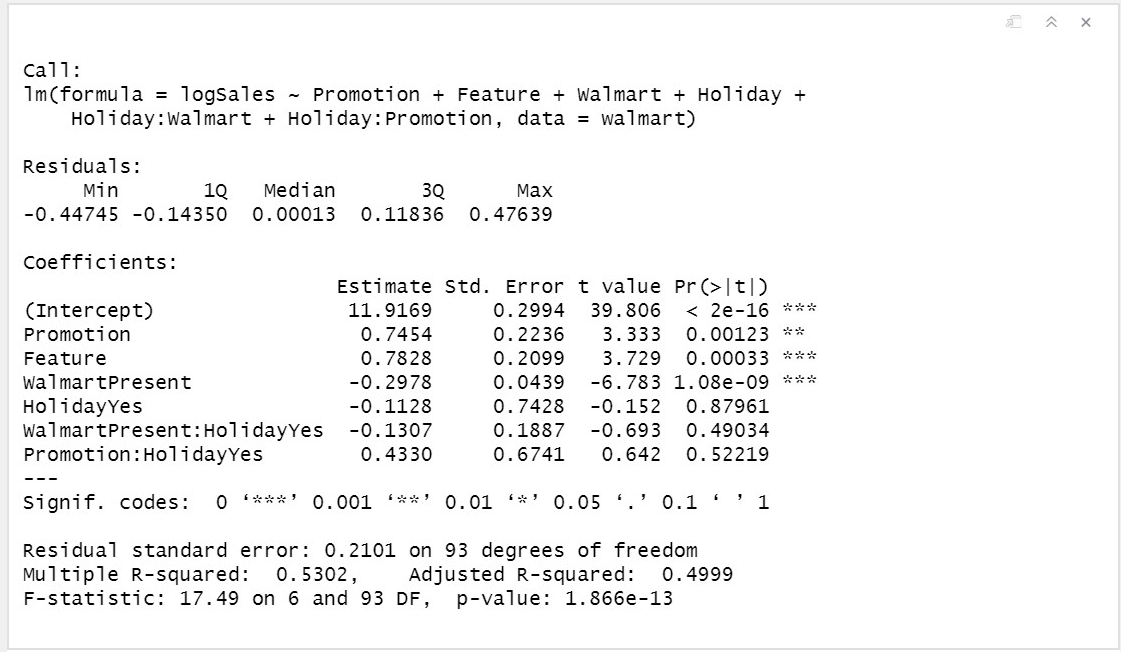
4). Estimate the following regression model using the functions lm( ) and summary( )

*log(sales)* = *β*0 + *β*1×*Promotion* + *β*2×*Feature* + *β*3×*WalMart* + *β*4×*Holiday* + *β*5×*Holiday*×*WalMart + β*6×*Holiday*×*Promotion + error*

Paste the R regression output from summary( ) here.

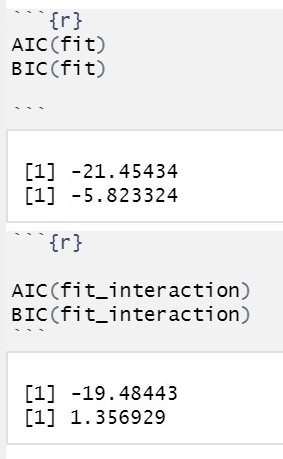
Interpret the estimated coefficients *β*1, *β*2, *β*3, *β*4, *β*5, and *β*6.

Compare this model with the one in Question (3) using AIC and BIC. Which is the better model?



Interpretation:

* ***β*1:** Increase in sales of stores: exp(0.7454) – 1 = 1.107284 = 110.7284% due to a one unit increase in the promotion index
* ***β*2:** Increase in sales of stores: exp(0.7828) – 1 = 1.187589 = 118.7589% due to a one unit increase in the feature index
* ***β*3:** Sales of stores having Walmart in vicinity: Exp(-0.2978)-1 = -0.2573 = 25.73% lesser than the sales of stores that don’t have Walmart in the vicinity.
* ***β*4:** Decrease in sales of stores: exp(-0.1128) – 1 = -0.1066707 = 10.667% when it is a holiday
* ***β*5:** additional decrease in the sales of stores : exp(-0.1307) – 1 = -0.122519 = 12.2519% when it is a holiday along with a Walmart in vicinity
* ***β*6:** additional increase in the sales of stores : exp(-0.4330) – 1 = 0.5418762 = 54.18762% when there is a one unit increase in promotional index when it is also a holiday



The AIC, BIC values show that the model in Q3 was much better than the other.

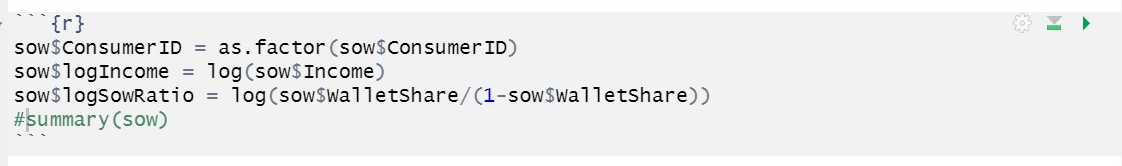
**Random Effects and Hierarchical Linear Models**

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)).

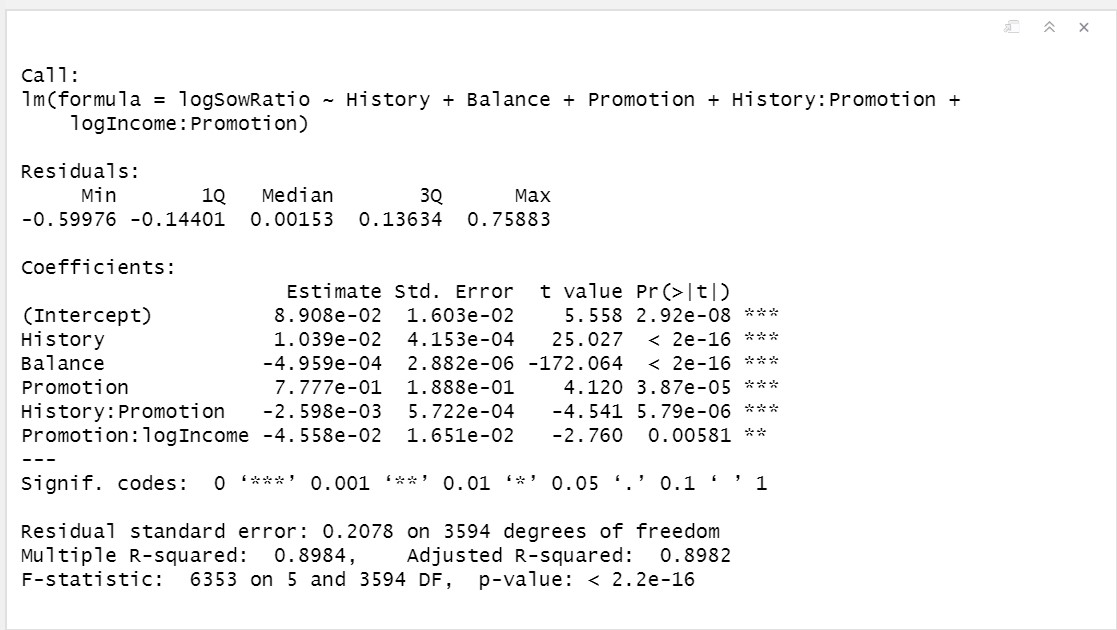


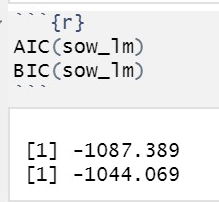
2). Use the function lm( ) to run the regression

*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*

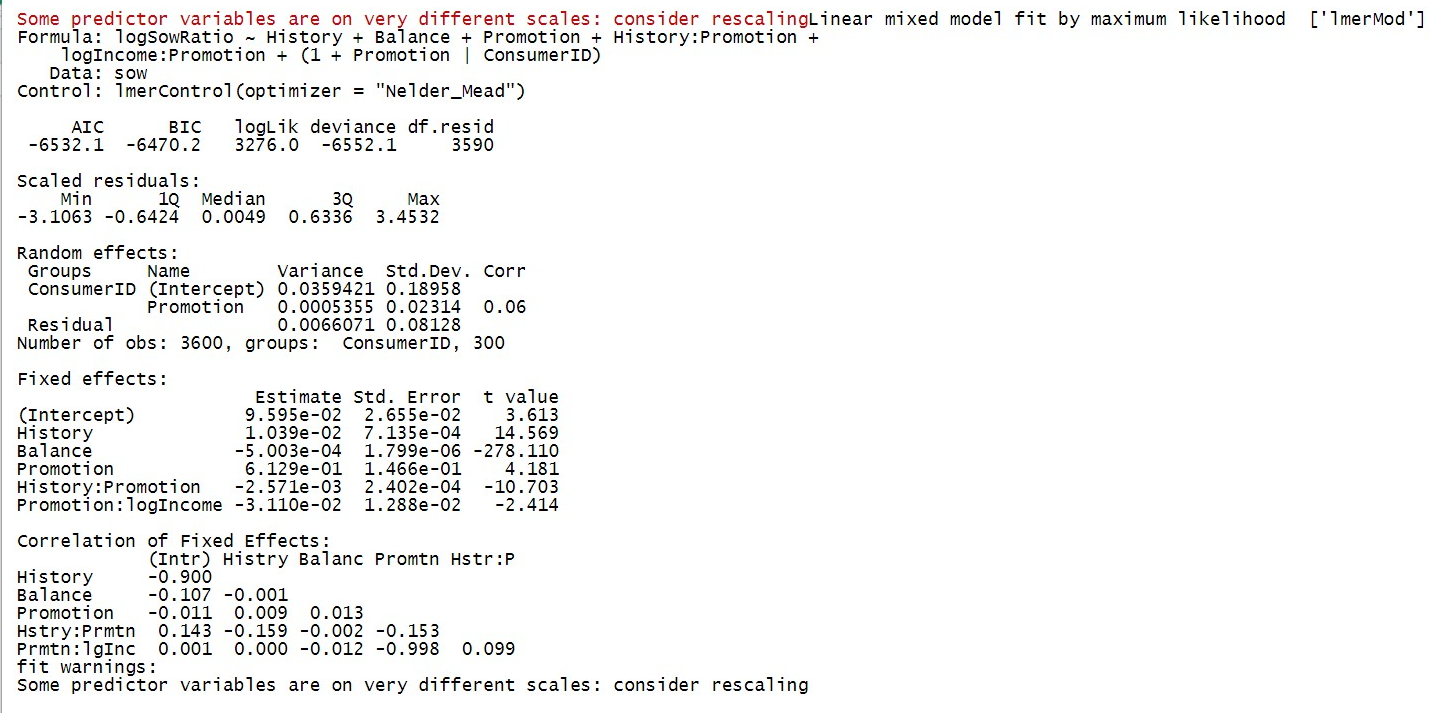
*β2i = γ0 +γ1×Historyi +γ2×logIncomei +ξi*

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

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.

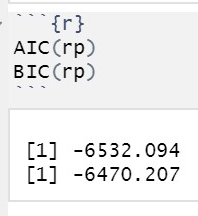
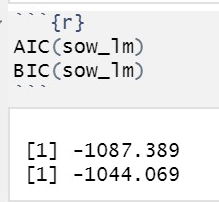
Please interpret the estimated fixed effects in the regression.

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



Interpretations:

* **History**: increase in the sow ratio :exp(0.01039) – 1 = 0.01044 = 1.044% due to a extra one month of using the card
* **Balance**: decrease in the sow ratio: exp(-0.0005003) – 1 = -0.00050017 = 0.050017% due to a one unit increase in unpaid balance
* **Promotion**: increase in the sow ratio exp(0.6129) – 1 = 0.8457764 = 84.57764% due to a one unit increase in promotion index
* **History; Promotion** : decrease in the sow ratio: exp(-0.002571) – 1 = -0.00257 = 0.257% due to a multiplicative change of history and promotion by 1.
* **Promotion;logIncome**: decrease in the sow ratio :exp(-0.0311) – 1 = -0.03062137 = 3.06% due to a multiplicative change of log income and promotion by 1.



Looking at AIC and BIC values, linear regression model performs much better (lower AIC, BIC values).