SCM 651 Fall 2018 Group Assignment 2 Solutions

Due Date: Midnight, Tuesday, 10/16/2018, Total Points = 40

Please use the programming language R to complete this assignment. Copy and paste relevant parts of the R output and/or screen shots into a Word file to prepare the answers.

1.(20 points) Please use the data set 651F18 Orange Juice Homework 2.csv to do all parts of question 1.

This data set provides, for a random sample of 5780 cases drawn from the Dominicks data base, made available by University of Chicago, Kilts Center:

- MOVE: Number of units sold for three brands of orange juice: Florida's Natural Home-squeezed (FLNAT), Tree Fresh (TF), and Tropicana Grove Stand (TROPICANA), at a store in a given week.
- PRICE: Unit price of the brand.
- logMOVE: Natural logarithm of MOVE.
- logPRICE: Natural logarithm of price.
- BRAND
- Season
- Feat (1 if product is on sale, 0 if not)
- Demographic variables at the store location: AGE9, AGE60, EDUC, ETHNIC, INCOME, HSIZEAV, HH3PLUS, HH4PLUS, HHSINGLE, HHLARGE, HVAL150, HVAL200, MORTGAGE, NOCAR, NWHITE, SINGLE, POVERTY, RETIRED, SINGLE, UNEMP, WORKWOM

1(a)(4+3+3 = 10 points) Fit a regression model with dependent variable logMOVE and the following independent variables:

- BRAND
- logPRICE
- Interaction between BRAND and logprice
- Feat
- Season
- Demographic variables given in the data set.

1(a)(i) From the estimated parameters, what are the price elasticities of demand of the three brands?

```
Coefficients:
                            Estimate Std. Error t value Pr(>|t|)
(Intercept)                -10.83734    4.36791  -2.481 0.013125 *
BRAND[T.TF]                 -0.54003    0.13661  -3.953 7.81e-05 ***
BRAND[T.TROPICANA]           1.06738    0.16127   6.619 3.95e-11 ***
logPRICE                    -3.31164    0.13166 -25.153  < 2e-16 ***
Feat                         0.23030    0.02409   9.559  < 2e-16 ***
Season[T.Spring]            -0.02445    0.02885  -0.847 0.396757
Season[T.Summer]            -0.02978    0.02849  -1.045 0.295895
Season[T.Winter]             0.04221    0.02909   1.451 0.146865
AGE9                        10.67244    2.23239   4.781 1.79e-06 ***
AGE60                        2.39892    1.45809   1.645 0.099972 .
EDUC                        -0.40258    0.34949  -1.152 0.249411
ETHNIC                      -1.50328    0.39754  -3.782 0.000157 ***
HH3PLUS                     -4.68900    3.52432  -1.330 0.183417
HH4PLUS                    -11.28363    3.63435  -3.105 0.001914 **
HHLARGE                    -32.06710    6.78987  -4.723 2.38e-06 ***

HHSINGLE                     3.96397    2.69468   1.471 0.141337
HSIZEAVG                     8.52585    2.13839   3.987 6.77e-05 ***
HVAL150                      0.94689    0.17724   5.343 9.52e-08 ***
HVAL200                     -0.08945    0.22792  -0.392 0.694721
INCOME                       0.06580    0.19917   0.330 0.741123
MORTGAGE                     0.15890    0.20936   0.759 0.447917
NOCAR                        1.70522    0.45465   3.751 0.000178 ***
NWHITE                       0.71943    0.39578   1.818 0.069153 .
POVERTY                     -1.87986    1.63822  -1.148 0.251221
RETIRED                     -1.41060    1.82679  -0.772 0.440043
SINGLE                       6.52714    0.91567   7.128 1.14e-12 ***
UNEMP                       -3.95899    2.67228  -1.482 0.138527
WORKWOM                     -4.82137    1.44372  -3.340 0.000845 ***
BRAND[T.TF]:logPRICE         0.40820    0.14540   2.807 0.005010 **
BRAND[T.TROPICANA]:logPRICE -0.33128    0.16793  -1.973 0.048577 *
```

In the model, FLNAT is the default brand, and dummy variables are created for TF and TROPICANA. The price elasticities of demand are given below.

Brand	Price Elasticity of Demand
Florida's Natural Home Squeezed	- 3.31164
Tree Fresh	- 3.31164 + .40820 = - 2.90344
Tropicana Grove Stand	- 3.31164 - .33128 = - 3.64292

1(a)(ii) Starting from the full model estimated, test the following hypothesis at a 99% level of confidence:

The price elasticity of demand is equal for the three brands.

WORKWOM	BRAND[T.TF]:	BRAND[T.TROP	Right-hand side
0	1	0	0
0	0	1	0

```
Linear hypothesis test

Hypothesis:
BRAND[T.TF]:logPRICE = 0
BRAND[T.TROPICANA]:logPRICE = 0

Model 1: restricted model
Model 2: logMOVE ~ BRAND + logPRICE + BRAND * logPRICE + Feat + Season +
    AGE9 + AGE60 + EDUC + ETHNIC + HH3PLUS + HH4PLUS + HHLARGE +
    HHSINGLE + HSIZEAVG + HVAL150 + HVAL200 + INCOME + MORTGAGE +
    NOCAR + NWHITE + POVERTY + RETIRED + SINGLE + UNEMP + WORKWOM

  Res.Df    RSS Df Sum of Sq      F       Pr(>F)
1   5752 3410.3
2   5750 3391.7  2    18.612 15.777 0.0000001469 ***
```

Since P value < .01, we reject the null hypothesis at a 99% level of confidence and conclude that at least one of three price elasticities is different from the others.

1(a)(iii) Starting from the full model estimated, test the following hypothesis at a 99% level of confidence:

The price elasticity of demand is equal for Tree Fresh and Tropicana.

SINGLE	UNEMP	WORKWOM	BRAND[T.TF]:	BRAND[T.TROP	Right-hand side
0	0	0	1	-1	0

```
Linear hypothesis test

Hypothesis:
BRAND[T.TF]:logPRICE - BRAND[T.TROPICANA]:logPRICE = 0

Model 1: restricted model
Model 2: logMOVE ~ BRAND + logPRICE + BRAND * logPRICE + Feat + Season +
    AGE9 + AGE60 + EDUC + ETHNIC + HH3PLUS + HH4PLUS + HHLARGE +
    HHSINGLE + HSIZEAVG + HVAL150 + HVAL200 + INCOME + MORTGAGE +
    NOCAR + NWHITE + POVERTY + RETIRED + SINGLE + UNEMP + WORKWOM

  Res.Df    RSS Df Sum of Sq      F        Pr(>F)
1   5751 3409.3
2   5750 3391.7  1    17.616 29.865 0.00000004828 ***
```

Since P value < .01, we reject the null hypothesis at a 99% level of confidence and conclude that the price elasticities of Tree Fresh and Tropicana are not equal.

1(b)(5 points) From the output from the full model in 1(a)(i), identify the demographic variables that are not significant at a 90% level of confidence. At a 99% level of confidence, test the null hypothesis that none of these variables is significant.

From the output for 1(a)(i), the demographic variables not significant at a 90% level of confidence (P value not less than 0.1) are the nine variables: EDUC, HH3PLUS, HHSINGLE, HVAL200, INCOME, MORTGAGE, POVERTY, RETIRED, UNEMP.

We used linear hypothesis test with the null hypothesis that all nine coefficients are zero.

AGE60	EDUC	ETHNIC	HH3PLUS	HH4PLUS	HHLARGE	HHSINGLE	HSIZEAVG	HVAL150	HVAL200	INCOME	MORTGAGE	NOCAR	NWHITE	POVERTY	RETIRED	SINGLE	UNEMP	WORKWOM
0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0

```
Linear hypothesis test

Hypothesis:
EDUC = 0
HH3PLUS = 0
HHSINGLE = 0
HVAL200 = 0
INCOME = 0
MORTGAGE = 0
POVERTY = 0
RETIRED = 0
UNEMP = 0

  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1   5759 3398.5
2   5750 3391.7  9    6.8022 1.2813 0.2413
```

Since P value > .01, we cannot reject the null hypothesis that all nine coefficients are zero at a 99% level of confidence.

1(c)(5 points) Based on the result of the hypothesis test in question 1(b), keep the full model or the restricted model as appropriate as the next model. Based on variance inflation factor (VIF), is there any problem with multi-collinearity with this model? If yes, then modify the model to reduce the effect of multi-collinearity. (To do this, identify groups of demographic variables that are strongly correlated with one another. Then, keep one variable from each group.) Fit the model and check if the problem of multi-collinearity is reduced.

Based on the result from 1(b), we continue with the reduced model where the insignificant demographic variables are dropped. The estimated coefficients of the model are given below.

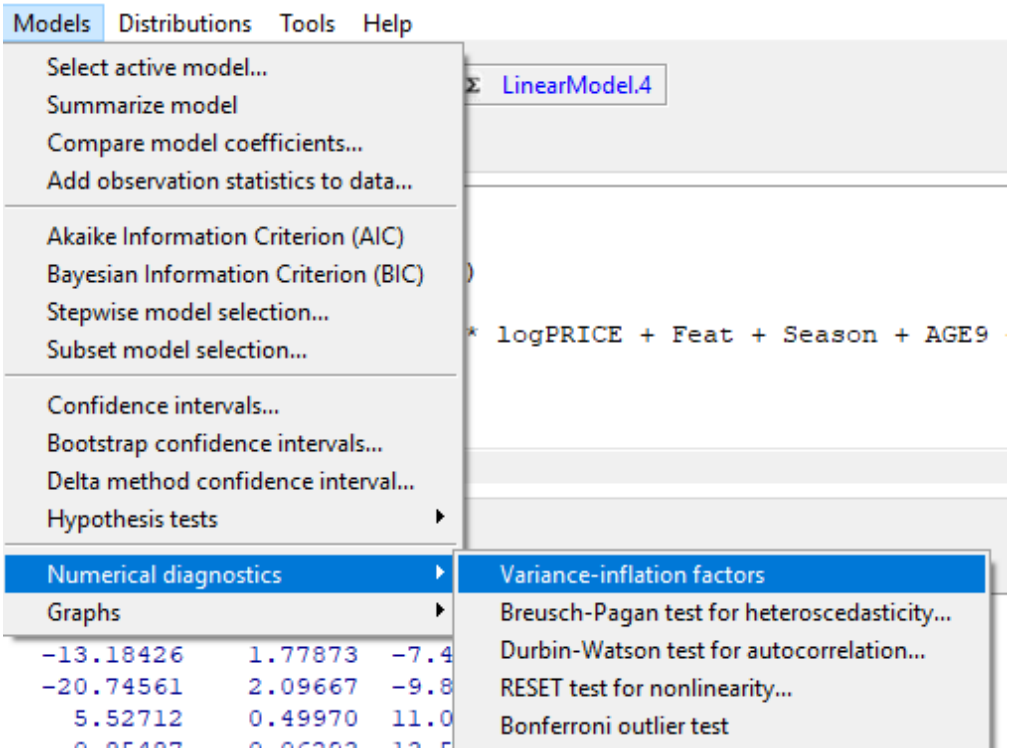
```
Call:
lm(formula = logMOVE ~ BRAND + logPRICE + BRAND * logPRICE +
    Feat + Season + AGE9 + AGE60 + ETHNIC + HH4PLUS + HHLARGE +
    HSIZEAVG + HVAL150 + NOCAR + NWHITE + SINGLE + WORKWOM, data = orange)

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)               -5.54517    1.08476  -5.112 3.29e-07 ***
BRAND[T.TF]               -0.53852    0.13646  -3.946 8.03e-05 ***
BRAND[T.TROPICANA]         1.05360    0.16110   6.540 6.68e-11 ***
logPRICE                  -3.31210    0.13137 -25.212  < 2e-16 ***
Feat                       0.23169    0.02405   9.635  < 2e-16 ***
Season[T.Spring]          -0.02554    0.02884  -0.885  0.37598
Season[T.Summer]          -0.03063    0.02848  -1.075  0.28221
Season[T.Winter]           0.04088    0.02908   1.406  0.15985
AGE9                       7.98205    1.44655   5.518 3.58e-08 ***
AGE60                      2.49530    0.63298   3.942 8.17e-05 ***
```

```
ETHNIC                        -1.79723    0.33612  -5.347 9.29e-08 ***
HH4PLUS                      -13.18426    1.77873  -7.412 1.42e-13 ***
HHLARGE                      -20.74561    2.09667  -9.895  < 2e-16 ***
HSIZEAVG                       5.52712    0.49970  11.061  < 2e-16 ***
HVAL150                        0.85487    0.06293  13.584  < 2e-16 ***
NOCAR                          1.37719    0.26946   5.111 3.31e-07 ***
NWHITE                         0.82815    0.31959   2.591  0.00959 **
SINGLE                         6.08374    0.78741   7.726 1.30e-14 ***
WORKWOM                       -1.96604    0.42518  -4.624 3.85e-06 ***
BRAND[T.TF]:logPRICE           0.40602    0.14530   2.794  0.00522 **
BRAND[T.TROPICANA]:logPRICE   -0.31680    0.16776  -1.888  0.05902 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7682 on 5759 degrees of freedom
Multiple R-squared:  0.5274, Adjusted R-squared:  0.5258
F-statistic: 321.3 on 20 and 5759 DF,  p-value: < 2.2e-16
```

The result of the test of multi-collinearity is given below.



```
                      GVIF Df GVIF^(1/(2*Df))
BRAND           903.788930  2        5.482981
logPRICE          8.440442  1        2.905244
Feat              1.289142  1        1.135404
Season            1.030213  3        1.004973
AGE9             12.214768  1        3.494963
AGE60            15.015506  1        3.874985
ETHNIC           39.862204  1        6.313652
HH4PLUS         112.764956  1       10.619085
HHLARGE          37.640121  1        6.135154
HSIZEAVG        155.551178  1       12.472016
HVAL150           2.314047  1        1.521199
NOCAR            11.901434  1        3.449846
NWHITE           36.361609  1        6.030059
SINGLE           26.021904  1        5.101167
WORKWOM           4.883809  1        2.209934
BRAND:logPRICE  675.547382  2        5.098166
```

As the VIF for AGE9, AGE60, ETHNIC, HH4PLUS, HHLARGE, HSIZEAVG, NOCAR, NWHITE and SINGLE are all greater than 10, multi-collinearity is a problem here. To reduce the problem, we reduce the model where we keep AGE9 out of AGE9 and AGE60, ETHNIC out of ETHNIC, NOCAR and NWHITE, HHLARGE out of HH4PLUS, HHLARGE, HSIZEAVG and SINGLE. The results for the reduced model are given below.

```
> vif(LinearModel.5)
                      GVIF Df GVIF^(1/(2*Df))
BRAND           900.554513  2        5.478069
logPRICE          8.413914  1        2.900675
Feat              1.282877  1        1.132642
Season            1.028090  3        1.004628
AGE9              3.784091  1        1.945274
ETHNIC            2.844471  1        1.686556
HHLARGE           4.328656  1        2.080542
HVAL150           1.785690  1        1.336297
NOCAR             2.932878  1        1.712565
WORKWOM           1.872903  1        1.368540
BRAND:logPRICE  673.408769  2        5.094126
```

Note that the multi-collinearity problem for demographic variables is removed.

2.(5 points) Please use the data set 651F18 Orange Juice Homework 2.csv to answer this question.

Fit a regression model with dependent variable log of move, and independent variables BRAND, Feat, logPRICE and BRAND*logPRICE. For the six cases given below, use R to construct 99% prediction intervals for logMOVE.

Case	BRAND	logPRICE	Feat
1	FLNAT	.9	1
2	FLNAT	1.0	0
3	TF	.55	1
4	TF	.75	0
5	TROPICANA	.80	1
6	TROPICANA	.95	0

Hint: Suppose the regression model is called lm1. Copy and paste the table above and save it as a csv file, say, newdata1.csv. Read this data set using R Commander with name newdata1. In the R Console, at the ready prompt > type

Predict(lm1,interval="prediction",level=.99, newdata = newdata1) <hit enter>

R will return with the 99% prediction intervals that include the estimate of logMOVE and lower and upper bounds for 99% prediction intervals.

```
> predict(lm1,interval="prediction",level = .99, newdata=newdata1)
       fit       lwr      upr
1 3.297963 1.1384684 5.457457
2 2.712472 0.5533087 4.871636
3 4.149094 1.9895548 6.308634
4 3.298166 1.1389990 5.457333
5 4.415150 2.2555732 6.574726
6 3.624230 1.4650688 5.783391
```

3.(20 points) Please use the data set 651F18 Bank Homework 2.csv to answer all parts of question 3.

This dataset provides, for 5000 individual customers of Universal Bank, the following information:

Personal Loan	Did this customer take a loan from the bank? 1 if yes, 0 if no.
Age	Age of the customer
Experience	Professional experience of the customer

Income	Income of customer
Family	Family size of customer
CCAvg	Average monthly credit card spending
Education	Three categories: 1: undergraduate, 2: graduate, 3: professional. This is also coded as two dummy variables: GRAD = 1 if graduate education, 0 if undergraduate or professional education PROF = 1 if professional education, 0 if undergraduate or graduate education
ED	1 if graduate or professional education, 0 otherwise (undergraduate education)
Mortgage	Size of mortgage
SecuritiesAccount	1 if yes, 0 if no
CDAccount	1 if yes, 0 if no
Online	1 if yes, 0 if no
CreditCard	1 if yes, 0 if no

3.(a)(4 points) Please perform logit analysis using Personal Loan as the dependent variable, and all the remaining variables as independent variables. (For education, include the two dummy variables GRAD and PROF.) Include **only main effects** in your model. Which variables are significant at a 90% level of confidence? Copy and paste screen shots from R analysis to support your answers.

```
Call:
glm(formula = PersonalLoan ~ Age + CCAvg + CDAccount + CreditCard +
    Experience + Family + GRAD + Income + Mortgage + Online +
    PROF + SecuritiesAccount, family = binomial(logit), data = Dataset)
```

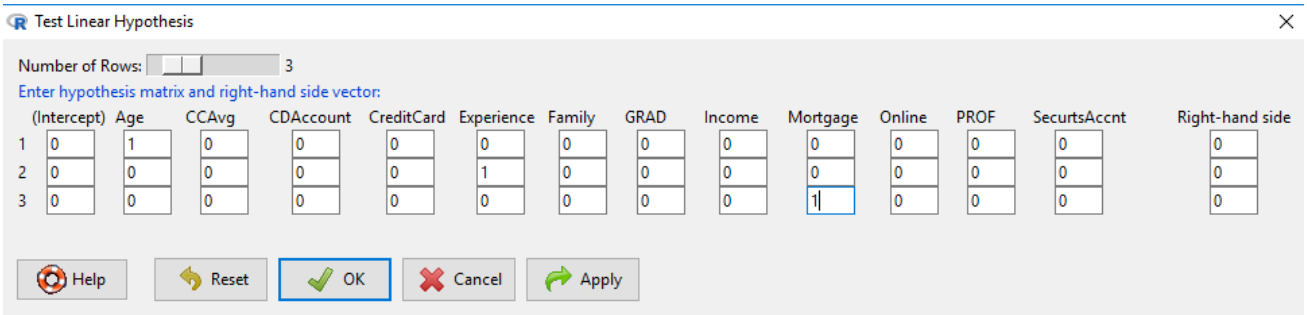
```
Coefficients:
                    Estimate  Std. Error z value Pr(>|z|)
(Intercept)       -12.3105489   1.8175249  -6.773 1.26e-11 ***
Age                -0.0359174   0.0672659  -0.534 0.593368
CCAvg               0.1633508   0.0440562   3.708 0.000209 ***
CDAccount           3.8389223   0.3415634  11.239  < 2e-16 ***
CreditCard         -1.0382002   0.2130955  -4.872 1.10e-06 ***
Experience          0.0450379   0.0668215   0.674 0.500309
Family              0.6181693   0.0770407   8.024 1.02e-15 ***
GRAD                3.9653781   0.2696001  14.708  < 2e-16 ***
Income              0.0601830   0.0029663  20.289  < 2e-16 ***
Mortgage            0.0007105   0.0005940   1.196 0.231641
Online             -0.7605294   0.1657435  -4.589 4.46e-06 ***
PROF                4.0640537   0.2669253  15.225  < 2e-16 ***
SecuritiesAccount  -0.8701362   0.3006910  -2.894 0.003806 **
```

```
    Null deviance: 3162.0  on 4999  degrees of freedom
Residual deviance: 1172.3  on 4987  degrees of freedom
AIC: 1198.3
```

The variables significant at a 90% level of confidence are CCAvg, CDAccount, CreditCard, Family, GRAD, Income, Online, PROF, and SecuritiesAccount.

3(b)(4 points) From your answer to question 3(a), list the variables that are not significant at a 90% level of confidence. Using test linear hypothesis, test the null hypothesis that none of these variables is significant using a 99% level of confidence. Copy and paste R screenshots to support your answer and clearly state your conclusion.

From 3(a), Age, Experience, and Mortgage have P values greater than .1 and are not significant at a 90% level of confidence. I performed linear hypothesis where the corresponding coefficients are set equal to zero.



```
Linear hypothesis test

Hypothesis:
Age = 0
Experience = 0
Mortgage = 0

Model 1: restricted model
Model 2: PersonalLoan ~ Age + CCAvg + CDAccount + CreditCard + Experience +
    Family + GRAD + Income + Mortgage + Online + PROF + SecuritiesAccount

  Res.Df Df  Chisq Pr(>Chisq)
1   4990
2   4987  3 3.5908     0.3092
```

Since P value = .3092 > .01, we cannot reject the null hypothesis that all three coefficients are zero at a 99% level of confidence.

3(c)(7 points) Based on your conclusion for 3(b), use the reduced model (if you cannot reject null hypothesis) or the original model (if you can reject the null hypothesis) to answer this question. If you are using the reduced model, estimate it.

Briefly discuss how the predictors in this model affect the probability of Personal Loan. Identify which one among the continuous predictors has the highest effect on the probability of Personal Loan. Identify which one among the 1/0 variables has the highest effect on the probability of personal loan.

(Hint: For a continuous predictor, compare how much the indicator function I changes if that predictor changes by one standard deviation. For a 1/0 variable, find how much I changes if that variable changes from 0 to 1.)

We use the reduced model after dropping Age, Experience, and Mortgage.

The estimated coefficients for the reduced model are given below.

```
Coefficients:
                    Estimate Std. Error z value Pr(>|z|)
(Intercept)       -12.970712   0.565806 -22.924  < 2e-16 ***
CCAvg               0.151415   0.043603   3.473 0.000515 ***
CDAccount           3.866880   0.340925  11.342  < 2e-16 ***
CreditCard         -1.039881   0.212728  -4.888 1.02e-06 ***
Family              0.618492   0.077058   8.026 1.00e-15 ***
GRAD                3.938590   0.268329  14.678  < 2e-16 ***
Income              0.060528   0.002942  20.575  < 2e-16 ***
Online             -0.752211   0.165311  -4.550 5.36e-06 ***
PROF                4.010641   0.262655  15.270  < 2e-16 ***
SecuritiesAccount  -0.866490   0.299702  -2.891 0.003838 **
```

To help interpret the coefficients, I computed the standard deviations of the quantitative predictors CCAvg, Family, and Income.

```
> numSummary(Dataset[,c("CCAvg", "Family", "Income"), drop=FALSE],
             mean        sd  IQR 0%  25%  50%  75% 100%    n
CCAvg    1.937938  1.747659  1.8  0  0.7  1.5  2.5   10 5000
Family   2.396400  1.147663  2.0  1  1.0  2.0  3.0    4 5000
Income  73.774200 46.033729 59.0  8 39.0 64.0 98.0  224 5000
```

The effects of the predictors are summarized below.

Variable	Coefficient	Standard Deviation	Coefficient*Standard Deviation
Continuous			
CCAvg	0.151415	1.747659	0.264622
Family	0.618492	1.147663	0.70982
Income	0.060528	46.033729	2.78633
Binary (1/0)			
CDAccount	3.866880		
CreditCard	-1.039881		
GRAD	3.938590		
Online	-0.752211		
PROF	4.010641		
SecuritiesAccount	-0.866490		

Thus, an increase in CCAvg, Family and Income all increase the probability of personal loan.

A person with a CD account is more likely to take a personal loan than a person without a CD account. A person with either a graduate degree or a professional degree is more likely to take a personal loan than a person with only an undergraduate degree. A person with a credit card, online account or a securities account is less likely to take personal loan than a person without the item.

Of the continuous variables, income has the highest effect on probability as a marginal change of one standard deviation in income changes I by 2.78633.

Out of the binary variables, GRAD has the highest effect on probability.

3(d)(5 points) Please perform a logit analysis with Personal Loan as the dependent variable and the following independent variables:

CCAvg, CDAccount , CreditCard, ED, Family, Income, Online, SecuritiesAccount,and the interaction of Income with each of ED, CD Account, Credit Card, Online and Securities Account.

(In this case, we will be able to check if the effect of income is "moderated," that is, affected by the level of ED, DC Account, etc. Here, ED, CD Account, etc., are called moderating variables.)

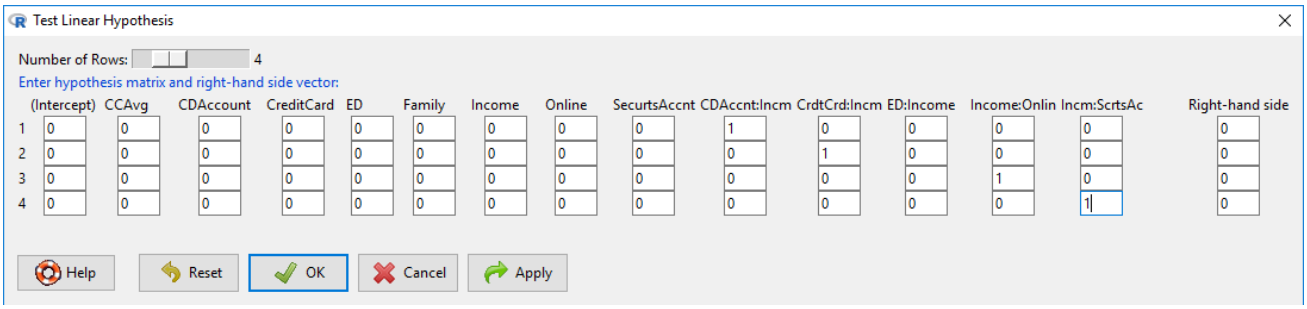
Focus on the interaction terms and identify the interactions that are not significant at a 90% level of confidence. At a 99% level of confidence, test the null hypothesis that none of these interaction terms is significant. Based on the result, keep the original model or the reduced model as appropriate. Briefly discuss how the effect of income is affected by the moderating variable.

```
Call:
glm(formula = PersonalLoan ~ CCAvg + CDAccount + CreditCard +
    ED + Family + Income + Online + SecuritiesAccount + Income *
    CDAccount + Income * CreditCard + Income * ED + Income *
    Online + Income * SecuritiesAccount, family = binomial(logit),
    data = Dataset)
```

```
    Null deviance: 3162.0  on 4999  degrees of freedom
Residual deviance:  849.4  on 4986  degrees of freedom
AIC: 877.4
```

```
Coefficients:
                           Estimate  Std. Error z value     Pr(>|z|)
(Intercept)              -7.8648725   0.6173288 -12.740      < 2e-16 ***
CCAvg                     0.2091331   0.0517929   4.038 0.0000539382 ***
CDAccount                 4.8925547   1.1567438   4.230 0.0000234115 ***
CreditCard               -1.6638308   0.8750384  -1.901       0.0572 .
ED                      -10.6892661   1.1041484  -9.681      < 2e-16 ***
Family                    0.8169216   0.0942480   8.668      < 2e-16 ***
Income                    0.0228901   0.0041195   5.557 0.0000000275 ***
Online                   -0.8347757   0.6479559  -1.288       0.1976
SecuritiesAccount        -1.8852091   1.0529051  -1.790       0.0734 .
CDAccount:Income         -0.0080032   0.0091572  -0.874       0.3821
CreditCard:Income         0.0037484   0.0069832   0.537       0.5914
ED:Income                 0.1259426   0.0103406  12.179      < 2e-16 ***
Income:Online            -0.0007825   0.0051085  -0.153       0.8783
Income:SecuritiesAccount  0.0078930   0.0085382   0.924       0.3553
```

From the output, the interactions of Income with CD Account, Credit Card, Online and Securities Account are not significant at a 90% level of confidence. The result of linear hypothesis test that these coefficients are all zero is given below.




```
Linear hypothesis test

Hypothesis:
CDAccount:Income = 0
CreditCard:Income = 0
Income:Online = 0
Income:SecuritiesAccount = 0

Model 1: restricted model
Model 2: PersonalLoan ~ CCAvg + CDAccount + CreditCard + ED + Family +
    Income + Online + SecuritiesAccount + Income * CDAccount +
    Income * CreditCard + Income * ED + Income * Online + Income *
    SecuritiesAccount

  Res.Df Df  Chisq Pr(>Chisq)
1   4990
2   4986  4 1.3001     0.8614
```

Since P value = .8614 > .01, we cannot reject the null hypothesis that all four coefficients are zero at a 99% level of confidence. The coefficients for the reduced model are given below.

```
Coefficients:
                    Estimate Std. Error z value Pr(>|z|)
(Intercept)        -7.822625   0.515658 -15.170  < 2e-16 ***
CCAvg               0.212874   0.051715   4.116 3.85e-05 ***
CDAccount           3.942669   0.384271  10.260  < 2e-16 ***
CreditCard         -1.214731   0.259601  -4.679 2.88e-06 ***
ED                -10.691213   1.102557  -9.697  < 2e-16 ***
Family              0.810191   0.093781   8.639  < 2e-16 ***
Income              0.022601   0.003089   7.316 2.56e-13 ***
Online             -0.922362   0.204460  -4.511 6.45e-06 ***
SecuritiesAccount  -0.962788   0.339646  -2.835  0.00459 **
ED:Income           0.125879   0.010327  12.189  < 2e-16 ***
```

The results show that the coefficient of income is .022601 for a person with undergraduate education only but is .022601+.125879 = .14848 for a person with a graduate or professional degree. Thus, an increase of one standard deviation in income increases I by

.022601*46.033729 = 1.04 for a person with only an undergraduate education, but by

.14848*46.033729 = 6.835 for a person with a graduate or professional degree. Thus, an increase in income has a much greater effect for a person with higher education.

Definitions of demographic variables in Questions 1 and 2

Variable Name	Description
AGE9	% Population under age 9
AGE60	% Population over age 60
ETHNIC	% Blacks & Hispanics
EDUC	% College Graduates
INCOME	Log of Median Income
HSIZEAV	Average Household Size
HH3PLUS	% of Households with 3 or more persons
HH4PLUS	% of households with 4 or more persons
HHSINGLE	% of households with1 person
HHLARGE	% of households with 5 or more persons
HVAL150	% of households with Value over $150,000
HVAL200	% of households with Value over $200,00
MORTGAGE	% of households with mortgages
NOCAR	% with No Vehicles
NWHITE	% of population that is non-white
POVERTY	% of population with income under $15,000
RETIRED	% of Retired
SINGLE	% of Singles
UNEMP	% of Unemployed
WORKWOM	% of Working Women with full-time jobs