

SolutionGood

Yu-Chen Xue

2018 年 6 月 4 日

MZines4You – Solutions

Dateset source:

<http://logisticregressionanalysis.com/303-what-a-logistic-regression-data-set-looks-like-an-example/>
Dataset introduction Household Income (Income; rounded to the nearest \$1,000.00) Gender (IsFemale = 1 if the person is female, 0 otherwise) Marital Status (IsMarried = 1 if married, 0 otherwise) College Educated (HasCollege = 1 if has one or more years of college education, 0 otherwise) Employed in a Profession (IsProfessional = 1 if employed in a profession, 0 otherwise) Retired (IsRetired = 1 if retired, 0 otherwise) Not employed (Unemployed = 1 if not employed, 0 otherwise) Length of Residency in Current City (ResLength; in years) Dual Income if Married (Dual = 1 if dual income, 0 otherwise) Children (Minors = 1 if children under 18 are in the household, 0 otherwise) Home ownership (Own = 1 if own residence, 0 otherwise) Resident type (House = 1 if residence is a single family house, 0 otherwise) Race (White = 1 if race is white, 0 otherwise) Language (English = 1 is the primary language in the household is English, 0 otherwise)

載入資料集

```
library(ggplot2)
library(readr)
KidCreative <- read_csv("D:/WORKSPACE/DATASETS/KidCreative.csv")
```

```
## Parsed with column specification:
## cols(
##   `Obs No.` = col_integer(),
##   Buy = col_integer(),
##   Income = col_integer(),
##   `Is Female` = col_integer(),
##   `Is Married` = col_integer(),
##   `Has College` = col_integer(),
##   `Is Professional` = col_integer(),
##   `Is Retired` = col_integer(),
##   Unemployed = col_integer(),
```

```
## `Residence Length` = col_integer(),
## `Dual Income` = col_integer(),
## Minors = col_integer(),
## Own = col_integer(),
## House = col_integer(),
## White = col_integer(),
## English = col_integer(),
## `Prev Child Mag` = col_integer(),
## `Prev Parent Mag` = col_integer()
## )
```

採用變數 Income 做回歸模型

```
KidCreative$Income <- as.numeric(KidCreative$Income)
model.income <- glm(Buy ~ Income, KidCreative, family = "binomial")
summary(model.income)
```

```
##
## Call:
## glm(formula = Buy ~ Income, family = "binomial", data = KidCreative)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.00280  -0.22466  -0.05466  -0.01655   2.87445
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -9.344e+00  8.315e-01  -11.24  <2e-16 ***
## Income       1.494e-04  1.336e-05   11.18  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 646.05  on 672  degrees of freedom
## Residual deviance: 249.32  on 671  degrees of freedom
## AIC: 253.32
##
## Number of Fisher Scoring iterations: 7
```

```
anova(model.income)
```

```
## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: Buy
##
## Terms added sequentially (first to last)
##
##
##          Df Deviance Resid. Df Resid. Dev
## NULL                      672      646.05
## Income  1    396.73      671      249.32
```

```
coef(model.income)
```

```
##      (Intercept)      Income
## -9.3441458339  0.0001494027
```

仿照資料集的分佈，生成數列

```
x.seq <- seq(0, 75000, 1000)
x.seq
```

```
## [1]      0 1000 2000 3000 4000 5000 6000 7000 8000 9000 10000
## [12] 11000 12000 13000 14000 15000 16000 17000 18000 19000 20000 21000
## [23] 22000 23000 24000 25000 26000 27000 28000 29000 30000 31000 32000
## [34] 33000 34000 35000 36000 37000 38000 39000 40000 41000 42000 43000
## [45] 44000 45000 46000 47000 48000 49000 50000 51000 52000 53000 54000
## [56] 55000 56000 57000 58000 59000 60000 61000 62000 63000 64000 65000
## [67] 66000 67000 68000 69000 70000 71000 72000 73000 74000 75000
```

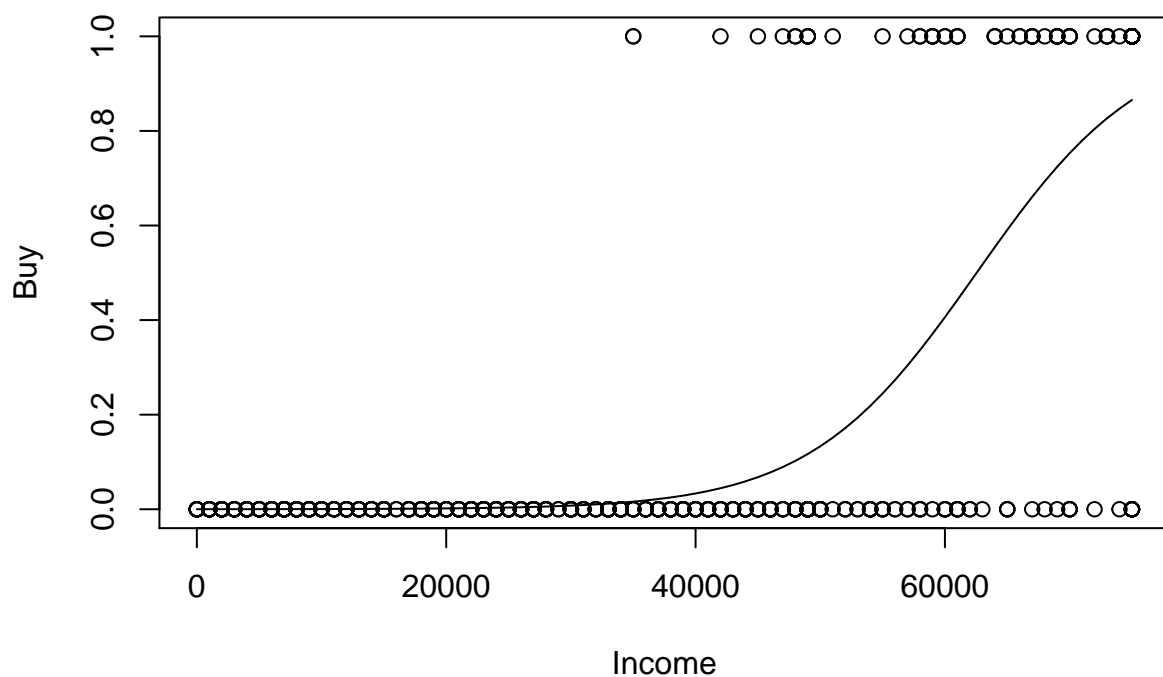
借助上述數列來做預測

```
pred.income <- predict(model.income, list(Income=x.seq), type="response")
pred.income
```

```
##           1           2           3           4           5
## 8.746837e-05 1.015616e-04 1.179254e-04 1.369254e-04 1.589861e-04
##           6           7           8           9          10
## 1.846005e-04 2.143407e-04 2.488711e-04 2.889627e-04 3.355107e-04
##          11          12          13          14          15
## 3.895540e-04 4.522985e-04 5.251439e-04 6.097143e-04 7.078944e-04
##          16          17          18          19          20
## 8.218711e-04 9.541815e-04 1.107768e-03 1.286045e-03 1.492970e-03
##          21          22          23          24          25
## 1.733131e-03 2.011847e-03 2.335280e-03 2.710568e-03 3.145976e-03
##          26          27          28          29          30
## 3.651069e-03 4.236912e-03 4.916294e-03 5.703989e-03 6.617051e-03
##          31          32          33          34          35
## 7.675142e-03 8.900910e-03 1.032041e-02 1.196354e-02 1.386463e-02
##          36          37          38          39          40
## 1.606289e-02 1.860313e-02 2.153628e-02 2.492017e-02 2.882009e-02
##          41          42          43          44          45
## 3.330948e-02 3.847050e-02 4.439445e-02 5.118205e-02 5.894341e-02
##          46          47          48          49          50
## 6.779763e-02 7.787183e-02 8.929958e-02 1.022185e-01 1.167667e-01
##          51          52          53          54          55
## 1.330785e-01 1.512788e-01 1.714759e-01 1.937538e-01 2.181640e-01
##          56          57          58          59          60
## 2.447160e-01 2.733697e-01 3.040280e-01 3.365321e-01 3.706605e-01
##          61          62          63          64          65
## 4.061313e-01 4.426093e-01 4.797171e-01 5.170500e-01 5.541936e-01
##          66          67          68          69          70
## 5.907416e-01 6.263137e-01 6.605705e-01 6.932251e-01 7.240505e-01
##          71          72          73          74          75
## 7.528826e-01 7.796192e-01 8.042153e-01 8.266765e-01 8.470510e-01
##          76
## 8.654204e-01
```

將預測結果用圖來呈現

```
plot(KidCreative$Income, KidCreative$Buy, pch = 1, xlab = "Income", ylab = "Buy")
lines(x.seq, pred.income)
```



收入大於 63000 的人群比較有可能購買雜誌

```
predict(model.income, list(Income=63000), type="response")
```

```
##          1
## 0.51705
```

Model2: 採用變數 Dual Income, Own 做回歸模型並進行預測

```
model2 <- glm(Buy ~ `Dual Income` * Own, KidCreative, family = "binomial", x=TRUE)
summary(model2)
```

```
##
## Call:
## glm(formula = Buy ~ `Dual Income` * Own, family = "binomial",
##      data = KidCreative, x = TRUE)
```

```
##
## Deviance Residuals:
##      Min        1Q    Median        3Q        Max
## -1.1945  -0.6945  -0.3320  -0.3320   2.4190
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)      -2.8707     0.2298 -12.490 < 2e-16 ***
## `Dual Income`       1.5715     0.3986   3.942 8.07e-05 ***
## Own                1.9834     0.2940   6.746 1.52e-11 ***
## `Dual Income`:Own  -0.6441     0.4822  -1.336   0.182
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 646.05  on 672  degrees of freedom
## Residual deviance: 526.57  on 669  degrees of freedom
## AIC: 534.57
##
## Number of Fisher Scoring iterations: 5
```

```
anova(model2)
```

```
## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: Buy
##
## Terms added sequentially (first to last)
##
##
##              Df Deviance Resid. Df Resid. Dev
## NULL                                672      646.05
## `Dual Income`      1    56.355      671      589.70
## Own                1    61.379      670      528.32
## `Dual Income`:Own  1     1.747      669      526.57
```

```
coef(model2)
```

```
##      (Intercept)      `Dual Income`      Own `Dual Income`:Own  
##      -2.8707358      1.5714528      1.9834326      -0.6441443
```

```
exp(coef(model2))
```

```
##      (Intercept)      `Dual Income`      Own `Dual Income`:Own  
##      0.05665722      4.81363635      7.26764705      0.52511170
```

```
predict(model2, list(`Dual Income`=1, Own=1), type="response")
```

```
##      1  
## 0.51
```

Model3: 採用變數 Dual Income, Is Married, Is Female, Prev Child Mag 做回歸模型並進行預測

```
model3 <- glm(Buy ~ `Dual Income` * `Is Married` * `Is Female` * `Prev Child Mag`, KidCreative, family =  
summary(model3)
```

```
##  
## Call:  
## glm(formula = Buy ~ `Dual Income` * `Is Married` * `Is Female` *  
##      `Prev Child Mag`, family = "binomial", data = KidCreative)  
##  
## Deviance Residuals:  
##      Min       1Q   Median       3Q      Max  
## -2.1899  -0.4355  -0.4355  -0.3132   2.4655  
##  
## Coefficients: (1 not defined because of singularities)  
##                                     Estimate  
## (Intercept)                       -2.99016  
## `Dual Income`                       2.07387  
## `Is Married`                        1.77377  
## `Is Female`                         0.68233  
## `Prev Child Mag`                   1.78619  
## `Dual Income`:`Is Married`         -1.30946
```

## `Dual Income`: `Is Female`	-0.45918	
## `Is Married`: `Is Female`	0.01277	
## `Dual Income`: `Prev Child Mag`	-4.36802	
## `Is Married`: `Prev Child Mag`	13.99627	
## `Is Female`: `Prev Child Mag`	-1.96326	
## `Dual Income`: `Is Married`: `Is Female`	-0.39703	
## `Dual Income`: `Is Married`: `Prev Child Mag`	-11.65560	
## `Dual Income`: `Is Female`: `Prev Child Mag`	19.80431	
## `Is Married`: `Is Female`: `Prev Child Mag`	-14.68420	
## `Dual Income`: `Is Married`: `Is Female`: `Prev Child Mag`	NA	
##	Std. Error	z value
## (Intercept)	0.34161	-8.753
## `Dual Income`	0.90371	2.295
## `Is Married`	0.52795	3.360
## `Is Female`	0.41781	1.633
## `Prev Child Mag`	0.74164	2.408
## `Dual Income`: `Is Married`	1.02794	-1.274
## `Dual Income`: `Is Female`	1.27458	-0.360
## `Is Married`: `Is Female`	0.64845	0.020
## `Dual Income`: `Prev Child Mag`	1081.13825	-0.004
## `Is Married`: `Prev Child Mag`	624.19440	0.022
## `Is Female`: `Prev Child Mag`	1.30047	-1.510
## `Dual Income`: `Is Married`: `Is Female`	1.41693	-0.280
## `Dual Income`: `Is Married`: `Prev Child Mag`	882.74586	-0.013
## `Dual Income`: `Is Female`: `Prev Child Mag`	624.19722	0.032
## `Is Married`: `Is Female`: `Prev Child Mag`	624.19638	-0.024
## `Dual Income`: `Is Married`: `Is Female`: `Prev Child Mag`	NA	NA
##	Pr(> z)	
## (Intercept)	< 2e-16	***
## `Dual Income`	0.02174	*
## `Is Married`	0.00078	***
## `Is Female`	0.10245	
## `Prev Child Mag`	0.01602	*
## `Dual Income`: `Is Married`	0.20271	
## `Dual Income`: `Is Female`	0.71865	
## `Is Married`: `Is Female`	0.98429	
## `Dual Income`: `Prev Child Mag`	0.99678	
## `Is Married`: `Prev Child Mag`	0.98211	
## `Is Female`: `Prev Child Mag`	0.13113	


```
## `Dual Income`: `Is Married`: `Is Female` 0.77932
## `Dual Income`: `Is Married`: `Prev Child Mag` 0.98947
## `Dual Income`: `Is Female`: `Prev Child Mag` 0.97469
## `Is Married`: `Is Female`: `Prev Child Mag` 0.98123
## `Dual Income`: `Is Married`: `Is Female`: `Prev Child Mag` NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 646.05  on 672  degrees of freedom
## Residual deviance: 525.53  on 658  degrees of freedom
## AIC: 555.53
##
## Number of Fisher Scoring iterations: 13
```

```
anova(model3)
```

```
## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: Buy
##
## Terms added sequentially (first to last)
##
##
##
```

	Df	Deviance
## NULL		
## `Dual Income`	1	56.355
## `Is Married`	1	30.498
## `Is Female`	1	2.190
## `Prev Child Mag`	1	9.959
## `Dual Income`: `Is Married`	1	5.187
## `Dual Income`: `Is Female`	1	0.503
## `Is Married`: `Is Female`	1	0.078
## `Dual Income`: `Prev Child Mag`	1	2.305
## `Is Married`: `Prev Child Mag`	1	0.382
## `Is Female`: `Prev Child Mag`	1	1.134
## `Dual Income`: `Is Married`: `Is Female`	1	0.018

## `Dual Income`: `Is Married`: `Prev Child Mag`	1	0.425
## `Dual Income`: `Is Female`: `Prev Child Mag`	1	9.909
## `Is Married`: `Is Female`: `Prev Child Mag`	1	1.583
## `Dual Income`: `Is Married`: `Is Female`: `Prev Child Mag`	0	0.000
##	Resid. Df	
## NULL	672	
## `Dual Income`	671	
## `Is Married`	670	
## `Is Female`	669	
## `Prev Child Mag`	668	
## `Dual Income`: `Is Married`	667	
## `Dual Income`: `Is Female`	666	
## `Is Married`: `Is Female`	665	
## `Dual Income`: `Prev Child Mag`	664	
## `Is Married`: `Prev Child Mag`	663	
## `Is Female`: `Prev Child Mag`	662	
## `Dual Income`: `Is Married`: `Is Female`	661	
## `Dual Income`: `Is Married`: `Prev Child Mag`	660	
## `Dual Income`: `Is Female`: `Prev Child Mag`	659	
## `Is Married`: `Is Female`: `Prev Child Mag`	658	
## `Dual Income`: `Is Married`: `Is Female`: `Prev Child Mag`	658	
##	Resid. Dev	
## NULL	646.05	
## `Dual Income`	589.70	
## `Is Married`	559.20	
## `Is Female`	557.01	
## `Prev Child Mag`	547.05	
## `Dual Income`: `Is Married`	541.86	
## `Dual Income`: `Is Female`	541.36	
## `Is Married`: `Is Female`	541.28	
## `Dual Income`: `Prev Child Mag`	538.98	
## `Is Married`: `Prev Child Mag`	538.59	
## `Is Female`: `Prev Child Mag`	537.46	
## `Dual Income`: `Is Married`: `Is Female`	537.44	
## `Dual Income`: `Is Married`: `Prev Child Mag`	537.02	
## `Dual Income`: `Is Female`: `Prev Child Mag`	527.11	
## `Is Married`: `Is Female`: `Prev Child Mag`	525.53	
## `Dual Income`: `Is Married`: `Is Female`: `Prev Child Mag`	525.53	

```
coef(model3)
```

```
##                (Intercept)
##                -2.99016123
##                `Dual Income`
##                2.07387050
##                `Is Married`
##                1.77376590
##                `Is Female`
##                0.68232678
##                `Prev Child Mag`
##                1.78618842
##                `Dual Income`:`Is Married`
##                -1.30946030
##                `Dual Income`:`Is Female`
##                -0.45918323
##                `Is Married`:`Is Female`
##                0.01277162
##                `Dual Income`:`Prev Child Mag`
##                -4.36802500
##                `Is Married`:`Prev Child Mag`
##                13.99627468
##                `Is Female`:`Prev Child Mag`
##                -1.96326063
##                `Dual Income`:`Is Married`:`Is Female`
##                -0.39703452
##                `Dual Income`:`Is Married`:`Prev Child Mag`
##                -11.65560016
##                `Dual Income`:`Is Female`:`Prev Child Mag`
##                19.80431216
##                `Is Married`:`Is Female`:`Prev Child Mag`
##                -14.68419991
## `Dual Income`:`Is Married`:`Is Female`:`Prev Child Mag`
##                NA
```

```
exp(coef(model3))
```

```
##                (Intercept)
##                5.027933e-02
```

```

##           `Dual Income`
##           7.955556e+00
##           `Is Married`
##           5.893004e+00
##           `Is Female`
##           1.978476e+00
##           `Prev Child Mag`
##           5.966667e+00
##           `Dual Income`:`Is Married`
##           2.699657e-01
##           `Dual Income`:`Is Female`
##           6.317995e-01
##           `Is Married`:`Is Female`
##           1.012854e+00
##           `Dual Income`:`Prev Child Mag`
##           1.267625e-02
##           `Is Married`:`Prev Child Mag`
##           1.198133e+06
##           `Is Female`:`Prev Child Mag`
##           1.403999e-01
##           `Dual Income`:`Is Married`:`Is Female`
##           6.723108e-01
##           `Dual Income`:`Is Married`:`Prev Child Mag`
##           8.670361e-06
##           `Dual Income`:`Is Female`:`Prev Child Mag`
##           3.989362e+08
##           `Is Married`:`Is Female`:`Prev Child Mag`
##           4.195010e-07
## `Dual Income`:`Is Married`:`Is Female`:`Prev Child Mag`
##                                     NA

```

```

predict(model3, list(`Dual Income`=1, `Is Married`=1, `Is Female`=1, `Prev Child Mag`=1),type="res

```

```

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type =
## ifelse(type == : prediction from a rank-deficient fit may be misleading

##           1
## 0.9090909

```

Model4: 採用變數 Income, Is Female, Prev Child Mag 做回歸模型並進行預測

```
model4 <- glm(Buy~ Income*`Is Female`*`Prev Child Mag`, KidCreative, family = "binomial")
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
summary(model4)
```

```
##
## Call:
## glm(formula = Buy ~ Income * `Is Female` * `Prev Child Mag`,
##      family = "binomial", data = KidCreative)
##
## Deviance Residuals:
##      Min        1Q      Median        3Q        Max
## -2.28758  -0.14823  -0.03060  -0.00535   2.36912
##
## Coefficients:
##                                Estimate Std. Error z value Pr(>|z|)
## (Intercept)                   -1.244e+01  2.078e+00  -5.986 2.16e-09
## Income                        1.835e-04  3.042e-05   6.032 1.61e-09
## `Is Female`                   2.967e+00  2.397e+00   1.238  0.216
## `Prev Child Mag`              -6.429e+02  4.670e+04  -0.014  0.989
## Income:`Is Female`            -2.335e-05  3.667e-05  -0.637  0.524
## Income:`Prev Child Mag`       1.030e-02  7.452e-01   0.014  0.989
## `Is Female`:`Prev Child Mag`  6.473e+02  4.670e+04   0.014  0.989
## Income:`Is Female`:`Prev Child Mag` -1.036e-02  7.452e-01  -0.014  0.989
##
## (Intercept)                  ***
## Income                      ***
## `Is Female`
## `Prev Child Mag`
## Income:`Is Female`
## Income:`Prev Child Mag`
## `Is Female`:`Prev Child Mag`
## Income:`Is Female`:`Prev Child Mag`
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 646.05  on 672  degrees of freedom
## Residual deviance: 221.84  on 665  degrees of freedom
## AIC: 237.84
##
## Number of Fisher Scoring iterations: 21
```

```
anova(model4)
```

```
## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: Buy
##
## Terms added sequentially (first to last)
##
##
```

	Df	Deviance	Resid. Df	Resid. Dev
## NULL			672	646.05
## Income	1	396.73	671	249.32
## `Is Female`	1	16.19	670	233.14
## `Prev Child Mag`	1	4.00	669	229.14
## Income:`Is Female`	1	1.47	668	227.67
## Income:`Prev Child Mag`	1	0.67	667	227.00
## `Is Female`:`Prev Child Mag`	1	0.64	666	226.36
## Income:`Is Female`:`Prev Child Mag`	1	4.52	665	221.84

```
coef(model4)
```

```
##              (Intercept)              Income
##      -1.243737e+01          1.834881e-04
##      `Is Female`          `Prev Child Mag`
##      2.967394e+00          -6.429264e+02
##      Income:`Is Female`      Income:`Prev Child Mag`
##      -2.334668e-05          1.030071e-02
##      `Is Female`:`Prev Child Mag` Income:`Is Female`:`Prev Child Mag`
##      6.472722e+02          -1.036354e-02
```

```
exp(coef(model4))
```

```
##              (Intercept)              Income
##          3.967534e-06          1.000184e+00
##          `Is Female`          `Prev Child Mag`
##          1.944119e+01          6.034324e-280
##          Income:`Is Female`          Income:`Prev Child Mag`
##          9.999767e-01          1.010354e+00
##          `Is Female`:`Prev Child Mag` Income:`Is Female`:`Prev Child Mag`
##          1.278582e+281          9.896900e-01
```

```
predict(model4, list(Income=63000, `Is Female`=1, `Prev Child Mag`=1),type="response")
```

```
##          1
## 0.732303
```

Model5: 採用變數 Income, Is Female, Prev Child Mag 做回歸模型並進行預測

```
model5 <- glm(Buy~ Income*`Prev Child Mag`, KidCreative, family = "binomial")
summary(model5)
```

```
##
## Call:
## glm(formula = Buy ~ Income * `Prev Child Mag`, family = "binomial",
##      data = KidCreative)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.23298  -0.20981  -0.05396  -0.01450   2.57702
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -9.779e+00  9.305e-01 -10.509  <2e-16 ***
## Income         1.546e-04  1.483e-05  10.425  <2e-16 ***
## `Prev Child Mag` 3.395e+00  2.032e+00  1.671   0.0948 .
## Income:`Prev Child Mag` -3.744e-05  3.465e-05  -1.081   0.2799
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 646.05 on 672 degrees of freedom
## Residual deviance: 243.61 on 669 degrees of freedom
## AIC: 251.61
##
## Number of Fisher Scoring iterations: 7
```

```
anova(model5)
```

```
## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: Buy
##
## Terms added sequentially (first to last)
##
##
##           Df Deviance Resid. Df Resid. Dev
## NULL                                672      646.05
## Income           1   396.73      671      249.32
## `Prev Child Mag` 1     4.72      670      244.61
## Income:`Prev Child Mag` 1     1.00      669      243.61
```

```
coef(model5)
```

```
##           (Intercept)           Income      `Prev Child Mag`
##          -9.778978e+00          1.546492e-04          3.395219e+00
## Income:`Prev Child Mag`
##          -3.744138e-05
```

```
exp(coef(model5))
```

```
##           (Intercept)           Income      `Prev Child Mag`
##          5.662963e-05          1.000155e+00          2.982118e+01
## Income:`Prev Child Mag`
##          9.999626e-01
```



```
predict(model5, list(Income=63000, `Prev Child Mag`=1), type="response")
```

```
##           1
```

```
## 0.7311244
```