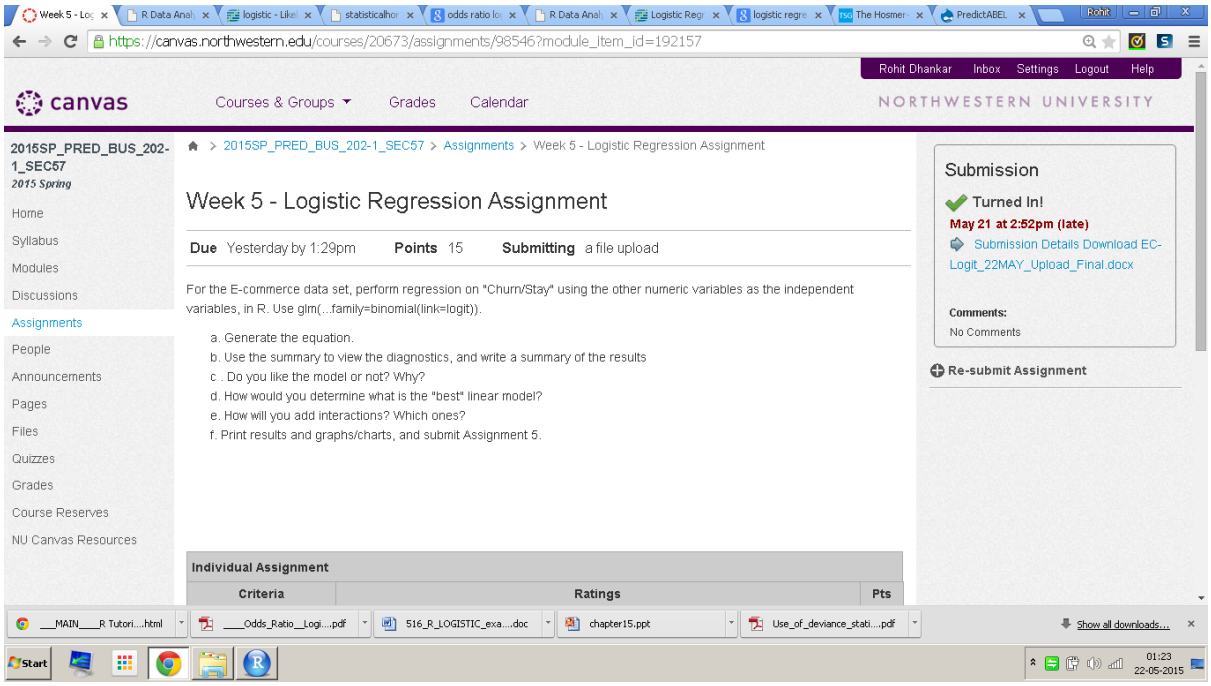


Week 5 - Logistic Regression Assignment

```
# For the E-commerce data set, perform regression on "Churn/Stay" using the other numeric variables as the independent variables, in R. Use glm(...family=binomial(link=logit)).


# a. Generate the equation.
```

Four equations generated –

```
m1<-glm(e$st~e$sls+e$sco+e$eco+e$csec+e$osec+e$qcec+e$spec+e$ad, data =e, family = "binomial")

m2<-glm(e$st~e$sls+e$eco+e$qcec+e$ad, data =e, family = "binomial")

m3<-glm(e$st~e$sls*e$sco+e$eco*e$csec+e$osec*e$qcec+e$spec*e$ad, data =e, family = "binomial")

m4<-glm(e$st~e$sls+e$sco*e$eco*e$csec+e$qcec*e$spec*e$ad, data =e, family = "binomial")

# b. Use the summary to view the diagnostics, and write a summary of the results
```

Ans-B- As seen below.

```
# Model - glm(m3) - AIC: 10803
# Model - glm(m4) -AIC: 10761


anova(object=m2, test="Chisq")

# Null- Resid deviance ==13842 at Df==9999
# e$ad- Resid deviance ==10863 at Df==9995


anova(object=m3, test="Chisq")

# Null- Resid deviance ==13842 at Df==9999
# e$spec:e$ad- Resid deviance ==10777 at Df==9987


anova(object=m4, test="Chisq")

# Null- Resid deviance ==13842 at Df==9999
# e$qcec:e$spec:e$ad- Resid deviance ==10729 at Df==9984


# c . Do you like the model or not? Why?
```

Ans-C-Model – glm(m4) is liked.

```
# d. How would you determine what is the "best" linear model?
```

Ans-D-Best model chosen basis various parameters of Confusion Matrix , anva(Chisq) and Hosmer-Lemeshow test.

```
# e. How will you add interactions? Which ones?
```

Interactions added in Model –glm(m3)

```
# f. Print results and graphs/charts, and submit Assignment 5.

#Reading Data
e<-read.csv("C:/STAT/Wk-5-EC-Logistic/ec.csv")
# First model -all independent variables.
attach(e)
m1<-glm(e$st~e$sls+e$sco+e$eco+e$csec+e$osec+e$qcec+e$spec+e$ad,
data =e, family = "binomial")
summary(m1)

##
## Call:
## glm(formula = e$st ~ e$sls + e$sco + e$eco + e$csec + e$osec +
##     e$qcec + e$spec + e$ad, family = "binomial", data = e)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.7229  -0.8522   0.1636   0.9333   1.6499
##
```

```
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.109e+00  3.419e-02 -32.442  < 2e-16 ***
## e$sls        1.479e-05  2.753e-06   5.373 7.75e-08 ***
## e$sco       -4.030e-04  1.413e-03  -0.285    0.775
## e$eco        1.664e-03  2.344e-04   7.099 1.26e-12 ***
## e$csec      -1.628e-02  9.958e-03  -1.635    0.102
## e$osec       8.666e-03  1.003e-02   0.864    0.388
## e$qcec      -2.414e-03  2.641e-04  -9.139  < 2e-16 ***
## e$spec      -4.492e-03  2.877e-03  -1.561    0.118
## e$ad         7.214e-02  3.971e-03  18.169  < 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: 13842  on 9999  degrees of freedom
## Residual deviance: 10789  on 9991  degrees of freedom
## AIC: 10807
##
## Number of Fisher Scoring iterations: 6

# From Summary- FOUR undermentioned variables, not Significant to Churn Status:-
# e$sco - p-Value==0.775
# e$csec- p-Value==0.102
# e$osec- p-Value==0.388
# e$spec- p-Value==0.118
## Remove FOUR variables - check if model is more Significant.
#
plot(predict(m1, type="response"),residuals(m1, type= "deviance"))
```

```
plot(hatvalues(m1))
```

```
plot(rstudent(m1))
```

```
plot(cooks.distance(m1))
```

Diagnostic plots helpful when response variable takes on many values.

 # Here response variable is Binomial - thus ignoring Diagnostic Plots.

```
# Create Confusion Matrix for this glm().
# Compare Classification capability with other glm()'s.
ModelFit_m1<-m1$fit
ProbabModelFit_m1<-m1$fit>0.5
P1<-as.data.frame(ProbabModelFit_m1)
# "if" loop to convert TRUE and FALSE into 1 and 0.
v1<-vector()
for (i in 1:nrow(P1))
{
  if(P1[i,1]=="TRUE")
  {
    v1[i]<-1
  }
  if(P1[i,1]=="FALSE")
  {
    v1[i]<-0
  }
}
V1<-as.data.frame(cbind(e$st,v1))
colnames(V1)<-c("yhat","m1")
# write.table(V1,"C:/STAT/V1.csv", sep=",")
```

Confusion Matrix – Model – glm(m1)

TN(f[0,0])	FP (f[0,1])		
3904	869	4773	Total ACTUAL NO- 0
1897	3330	5227	Total ACTUAL YES - 1
FN (f[1,0])	TP(f[1,1])		
5801	4199		
Total [PRED] NO- 0 CHURNED Count	Total [[PRED] YES-1 STAYED Count		

Percentage of Misclassification OR - Misclassification Rate:	28%	//=(FP+FN)/n	Less is better.
Sensitivity OR - True Positive Rate	64%	//=TP/(TP+FN)	More is better.
Specificity OR - True Negative rate (1- FalsePositive Rate)	82%	//=TN/(TN+FP)	More is better.

Confusion Matrix – Model – glm(m2)

TN(f[0,0])	FP (f[0,1])		
3929	844	4773	Total ACTUAL NO- 0
1961	3266	5227	Total ACTUAL YES - 1
FN (f[1,0])	TP(f[1,1])		
5890	4110		
Total [PRED] NO- 0 CHURNED Count	Total [[PRED] YES-1 STAYED Count		

Percentage of Misclassification OR - Misclassification Rate:	28%	//=(FP+FN)/n	Less is better.
Sensitivity OR - True Positive Rate	62%	//=TP/(TP+FN)	More is better.
Specificity OR - True Negative rate (1- FalsePositive Rate)	82%	//=TN/(TN+FP)	More is better.

Confusion Matrix – Model – glm(m3)

TN(f[0,0])	FP (f[0,1])		
3905	868	4773	Total ACTUAL NO- 0
1899	3328	5227	Total ACTUAL YES - 1
FN (f[1,0])	TP(f[1,1])		
5804	4196		
Total [PRED] NO- 0 CHURNED Count	Total [[PRED] YES-1 STAYED Count		

Percentage of Misclassification OR - Misclassification Rate:	28%	//=(FP+FN)/n	Less is better.
Sensitivity OR - True Positive Rate	64%	//=TP/(TP+FN)	More is better.
Specificity OR - True Negative rate (1- FalsePositive Rate)	82%	//=TN/(TN+FP)	More is better.

Confusion Matrix – Model – glm(m4)

TN(f[0,0])	FP (f[0,1])		
3843	930	4773	Total ACTUAL NO- 0
1826	3401	5227	Total ACTUAL YES - 1
FN (f[1,0])	TP(f[1,1])		
5669	4331		
Total [PRED] NO- 0 CHURNED Count	Total [[PRED] YES-1 STAYED Count		

Percentage of Misclassification OR - Misclassification Rate:	28%	//=(FP+FN)/n	Less is better.
Sensitivity OR - True Positive Rate	65%	//=TP/(TP+FN)	More is better.
Specificity OR - True Negative rate (1- FalsePositive Rate)	81%	//=TN/(TN+FP)	More is better.

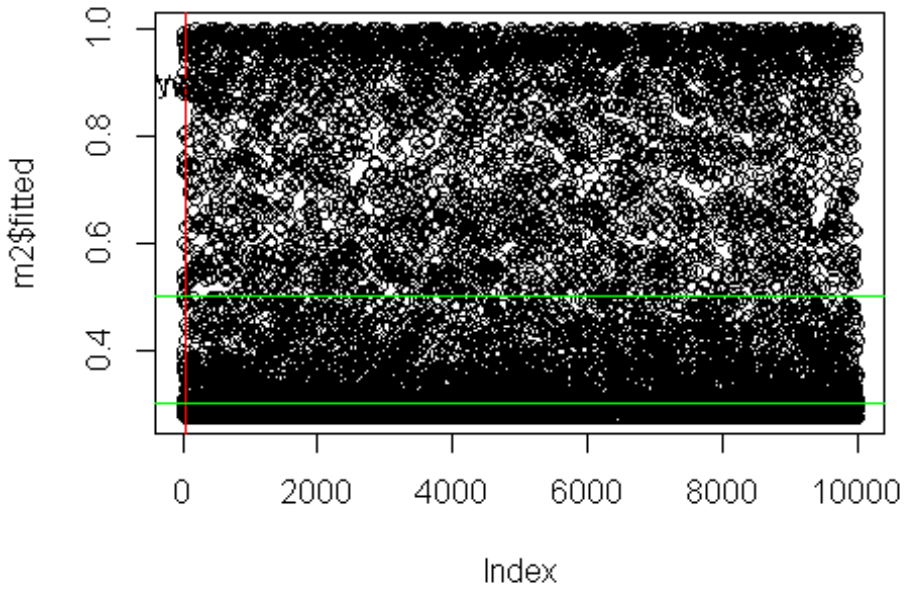
As seen from the Conf Matrix – Model – glm(m1) , which has all the Independent variables is preferred as Significant.

Also as seen below -**Prefer a glm()- with Smaller Mean Deviance / Residual Deviance-m1==Residual deviance:10789 .Whereas- m2==Residual deviance: 10863**

```
# Second glm()-as seen below -
m2<-glm(e$st~e$sls+e$eco+e$qcec+e$ad, data =e, family = "binomial")
summary(m2)

##
## Call:
## glm(formula = e$st ~ e$sls + e$eco + e$qcec + e$ad, family = "binomial",
##     data = e)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.6332  -0.8610   0.1405   0.9674   1.6083
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.035e+00  3.193e-02 -32.419  < 2e-16 ***
## e$sls        6.453e-06  2.583e-06   2.498   0.0125 *
## e$eco        5.969e-04  1.842e-04   3.240   0.0012 **
## e$qcec      -9.973e-04  1.966e-04  -5.071 3.95e-07 ***
## e$ad         6.181e-02  3.122e-03  19.798  < 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: 13842  on 9999  degrees of freedom
## Residual deviance: 10863  on 9995  degrees of freedom
## AIC: 10873
##
## Number of Fisher Scoring iterations: 6

# Independent Variables - all FOUR have Significant p-Values.
# Comparison between models- AIC - AIC of 2nd Model -m2 Higher at- AIC:
10873
# Prefer a glm() - with Smaller Mean Deviance / Residual Deviance -
m1==Residual deviance: 10789 # Whereas - m2==Residual deviance: 10863
# Confusion Matrix created in Excel as seen below -
ModelFit_m2<-m2$fit
ProbabModelFit_m2<-m2$fit>0.5
P2<-as.data.frame(ProbabModelFit_m2)
```

```r
# "if" loop to convert TRUE and FALSE into 1 and 0.

v2<-vector()
for (i in 1:nrow(P2))
{
 if(P2[i,1]=="TRUE")
 {
  v2[i]<-1
 }
 if(P2[i,1]=="FALSE")
 {
  v2[i]<-0
 }
}

V2<-as.data.frame(cbind(e$st,v2))
colnames(V2)<-c("yhat","m2")
write.table(V2,"C:/STAT/Wk-5-EC-Logistic/V2.csv", sep=",")
## CIs using profiled log-likelihood
confint(m2)

## Waiting for profiling to be done...

##                  2.5 %       97.5 %
## (Intercept) -1.097900e+00 -9.727387e-01
## e$sls        1.463126e-06  1.157731e-05
## e$eco        2.391078e-04  9.606299e-04
## e$qcec      -1.376291e-03 -6.050908e-04
## e$ad         5.571897e-02  6.795780e-02

# Analysis of the Residual Deviance from the ANOVA summary ...Analysis of Deviance Table
anova(object=m1, test="Chisq")

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

## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: e$st
##
## Terms added sequentially (first to last)
##
##
##       Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
## NULL                  9999     13842
## e$sls   1  2190.16     9998     11652 < 2.2e-16 ***
## e$sco   1   405.70     9997     11246 < 2.2e-16 ***
## e$eco   1     4.77     9996     11242   0.02902 *
## e$csec  1    24.05     9995     11218 9.386e-07 ***
## e$osec  1     4.01     9994     11214   0.04510 *
## e$qcec  1   101.70     9993     11112 < 2.2e-16 ***
## e$spec  1     0.35     9992     11112   0.55221
## e$ad    1   322.37     9991     10789 < 2.2e-16 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# Null- Resid deviance ==13842 at Df==9999
# e$ad- Resid deviance ==10789 at Df==9991
# The deviance was reduced by points on 8 degrees of freedom, for a p-value of...
x<-13842-10789
x

## [1] 3053

1 - pchisq(3053, df=8)

## [1] 0

plot(m2$fitted)
abline(v=30.5,col="red")
abline(h=.3,col="green")
abline(h=.5,col="green")
text(15,.9,"Churned = 0")
text(40,.9,"Stayed = 1")
```




```
# Hosmer - The Hosmer-Lemeshow test...
library("ResourceSelection", lib.loc="~/R/win-library/3.1")

## ResourceSelection 0.2-4   2014-05-19

hl <- hoslem.test(m1$y, fitted(m1), g=10)
hl

##
##  Hosmer and Lemeshow goodness of fit (GOF) test
```

```
##
## data:  m1$y, fitted(m1)
## X-squared = 66.6066, df = 8, p-value = 2.32e-11
```

```
cbind(hl$observed,hl$expected)
```

```
##                y0  y1     yhat0    yhat1
## [0.255,0.264] 793 207 738.41580 261.5842
## (0.264,0.279] 754 246 727.79106 272.2089
## (0.279,0.305] 729 271 708.95382 291.0462
## (0.305,0.344] 682 318 677.52100 322.4790
## (0.344,0.417] 554 446 623.44816 376.5518
## (0.417,0.525] 470 530 532.75843 467.2416
## (0.525,0.679] 380 620 397.52050 602.4795
## (0.679,0.844] 276 724 239.10715 760.8929
## (0.844,0.943] 106 894  99.11548 900.8845
## (0.943,1]      29 971  28.36861 971.6314
```

```
pihat1 <- m1$fitted
pihat1cat <- cut(pihat1, breaks=c(0,quantile(pihat1, probs=seq(0.1,0.9,0.1)),1),
labels=FALSE)
```

```
m3<-glm(e$st~e$sls*e$sco+e$eco*e$csec+e$osec*e$qcec+e$spec*e$ad, data
=e, family = "binomial")
summary(m3)
```

```
##
## Call:
## glm(formula = e$st ~ e$sls * e$sco + e$eco * e$csec + e$osec *
##     e$qcec + e$spec * e$ad, family = "binomial", data = e)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.3791  -0.8512   0.1592   0.9319   2.3418
##
## Coefficients:
##                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -1.099e+00  3.449e-02 -31.870  < 2e-16 ***
## e$sls          1.752e-05  3.374e-06   5.192 2.08e-07 ***
## e$sco          5.631e-04  2.055e-03   0.274  0.78406
## e$eco          1.553e-03  2.487e-04   6.243 4.28e-10 ***
## e$csec        -1.662e-02  1.006e-02  -1.651  0.09866 .
## e$osec         9.403e-03  1.015e-02   0.927  0.35402
## e$qcec        -2.268e-03  3.075e-04  -7.376 1.63e-13 ***
## e$spec        -2.338e-02  7.698e-03  -3.037  0.00239 **
## e$ad           6.722e-02  4.299e-03  15.638  < 2e-16 ***
## e$sls:e$sco   -1.633e-08  9.037e-09  -1.807  0.07079 .
## e$eco:e$csec   7.816e-08  2.485e-07   0.314  0.75315
## e$osec:e$qcec -1.949e-08  3.640e-07  -0.054  0.95729
## e$spec:e$ad    6.444e-04  2.618e-04   2.461  0.01386 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
```

```
##
##     Null deviance: 13842  on 9999  degrees of freedom
## Residual deviance: 10777  on 9987  degrees of freedom
## AIC: 10803
##
## Number of Fisher Scoring iterations: 7
```

```
# Confusion Matrix created in Excel as seen below -
ModelFit_m3<-m3$fit
ProbabModelFit_m3<-m3$fit>0.5
P3<-as.data.frame(ProbabModelFit_m3)
# "if" loop to convert TRUE and FALSE into 1 and 0.

v3<-vector()
for (i in 1:nrow(P3))
{
  if(P3[i,1]=="TRUE")
  {
    v3[i]<-1
  }
  if(P3[i,1]=="FALSE")
  {
    v3[i]<-0
  }
}

V3<-as.data.frame(cbind(e$st,v3))
colnames(V3)<-c("yhat","m2")
# write.table(V3,"C:/STAT/Wk-5-EC-Logistic/V3.csv", sep=",")



m4<-glm(e$st~e$sls+e$sco*e$eco*e$csec+e$qcec*e$spec*e$ad, data =e,
family = "binomial")

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

summary(m4)

##
## Call:
## glm(formula = e$st ~ e$sls + e$sco * e$eco * e$csec + e$qcec *
##     e$spec * e$ad, family = "binomial", data = e)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -4.3529  -0.8376   0.1294   0.9022   2.7156
##
## Coefficients:
##                    Estimate Std. Error z value Pr(>|z|)
## (Intercept)       -1.160e+00  3.632e-02 -31.933  < 2e-16 ***
## e$sls              1.468e-05  2.798e-06   5.245 1.56e-07 ***
## e$sco              2.071e-03  2.612e-03   0.793 0.427819
## e$eco              2.078e-03  2.673e-04   7.772 7.71e-15 ***
## e$csec            -4.906e-03  1.708e-03  -2.873 0.004067 **
## e$qcec            -1.874e-03  4.361e-04  -4.298 1.73e-05 ***
```

```
## e$spec           -3.242e-02  9.544e-03  -3.397 0.000682 ***
## e$ad             5.823e-02  5.445e-03  10.693  < 2e-16 ***
## e$sco:e$eco      -5.575e-06  1.414e-06  -3.943 8.05e-05 ***
## e$sco:e$csec      2.554e-06  6.066e-06   0.421 0.673768
## e$eco:e$csec     -2.172e-06  4.496e-07  -4.830 1.36e-06 ***
## e$qcec:e$spec     8.738e-06  1.579e-05   0.553 0.579968
## e$qcec:e$ad       1.018e-05  1.125e-05   0.905 0.365387
## e$spec:e$ad       8.784e-04  3.314e-04   2.650 0.008040 **
## e$sco:e$eco:e$csec  7.860e-09  1.565e-09   5.021 5.14e-07 ***
## e$qcec:e$spec:e$ad  1.555e-07  4.964e-07   0.313 0.754070
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 13842  on 9999  degrees of freedom
## Residual deviance: 10729  on 9984  degrees of freedom




## AIC: 10761
##
## Number of Fisher Scoring iterations: 10
```

```
# Model - glm(m3) - AIC: 10803
# Model - glm(m4) -AIC: 10761
ModelFit_m4<-m4$fit
ProbabModelFit_m4<-m4$fit>0.5
P4<-as.data.frame(ProbabModelFit_m4)
# "if" loop to convert TRUE and FALSE into 1 and 0.

v4<-vector()
for (i in 1:nrow(P4))
{
 if(P4[i,1]=="TRUE")
 {
  v4[i]<-1
 }
 if(P4[i,1]=="FALSE")
 {
  v4[i]<-0
 }
}

V4<-as.data.frame(cbind(e$st,v4))
colnames(V4)<-c("yhat","m4")
# write.table(V4,"C:/STAT/V4.csv", sep=",")



# Analysis of the Residual Deviance from the ANOVA summary ...Analysis of
Deviance Table
anova(object=m2, test="Chisq")
```

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

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

```
## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: e$st
##
## Terms added sequentially (first to last)
##
##
##       Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
## NULL                  9999      13842
## e$sls   1  2190.16     9998      11652 < 2.2e-16 ***
## e$eco   1   129.00     9997      11523 < 2.2e-16 ***
## e$qcec  1   236.79     9996      11286 < 2.2e-16 ***
## e$ad    1   423.51     9995      10863 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Null- Resid deviance ==13842 at Df==9999
# e$ad- Resid deviance ==10863 at Df==9995
```

```
anova(object=m3, test="Chisq")
```

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

```
## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: e$st
##
## Terms added sequentially (first to last)
##
##
##             Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
## NULL                        9999      13842
## e$sls        1  2190.16     9998      11652 < 2.2e-16 ***
## e$sco        1   405.70     9997      11246 < 2.2e-16 ***
## e$eco        1     4.77     9996      11242  0.029025 *
## e$csec       1    24.05     9995      11218 9.386e-07 ***
## e$osec       1     4.01     9994      11214  0.045101 *
## e$qcec       1   101.70     9993      11112 < 2.2e-16 ***
## e$spec       1     0.35     9992      11112  0.552213
## e$ad         1   322.37     9991      10789 < 2.2e-16 ***
## e$sls:e$sco    1     4.80     9990      10784  0.028393 *
## e$eco:e$csec   1     0.11     9989      10784  0.744254
## e$osec:e$qcec  1     0.03     9988      10784  0.862922
## e$spec:e$ad    1     7.54     9987      10777  0.006033 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Null- Resid deviance ==13842 at Df==9999
# e$spec:e$ad- Resid deviance ==10777 at Df==9987

anova(object=m4, test="Chisq")

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

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

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

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

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

## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: e$st
##
## Terms added sequentially (first to last)
##
##
##                 Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
## NULL                           9999     13842
## e$sls            1  2190.16     9998     11652 < 2.2e-16 ***
## e$sco            1   405.70     9997     11246 < 2.2e-16 ***
## e$eco            1     4.77     9996     11242   0.02902 *
## e$csec           1    24.05     9995     11218 9.386e-07 ***
## e$qcec           1   101.15     9994     11116 < 2.2e-16 ***
## e$spec           1     0.30     9993     11116   0.58452
## e$ad             1   326.30     9992     10790 < 2.2e-16 ***
## e$sco:e$eco      1     0.90     9991     10789   0.34220
## e$sco:e$csec     1     0.80     9990     10788   0.37057
## e$eco:e$csec     1     1.37     9989     10787   0.24186
## e$qcec:e$spec    1     3.83     9988     10783   0.05043 .
## e$qcec:e$ad      1     1.29     9987     10782   0.25608
## e$spec:e$ad      1     4.78     9986     10777   0.02884 *
## e$sco:e$eco:e$csec  1    47.76     9985     10729 4.828e-12 ***
## e$qcec:e$spec:e$ad  1     0.10     9984     10729   0.75348
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# Null- Resid deviance ==13842 at Df==9999
# e$qcec:e$spec:e$ad- Resid deviance ==10729 at Df==9984









############################################################

# Tried to Combine Binomial variables of EComm data and glm()- create
Confusion Matrix within R. Could not be done.
```

```
# install.packages('caret', dependencies = TRUE)
## The R inbuilt caret Package - Confusion Matrix is not used as the code is not clear...
# library("caret", lib.loc="~/R/win-library/3.1")
# lvs <- c("Stayed", "Churned")
# truth <- factor(
#   c(
#     rep(lvs, times = c(4199,5801))),
#                levels = rev(lvs))
# pred <- factor(
#   c(
#     rep(lvs, times = c(5227,4773))),
#   levels = rev(lvs))
#
# xtab <- table(pred, truth)
# xtab
# confusionMatrix(xtab)
# confusionMatrix(pred, truth)
# confusionMatrix(xtab, prevalence = 0.25)
#
#
#
# lvs <- c("normal", "abnormal")
# truth <- factor(rep(lvs, times = c(86, 258)),
#                levels = rev(lvs))
# pred <- factor(
#   c(
#     rep(lvs, times = c(54, 32)),
#     rep(lvs, times = c(27, 231))),
#   levels = rev(lvs))
#
# xtab <- table(pred, truth)
#
# confusionMatrix(xtab)
# confusionMatrix(pred, truth)
# confusionMatrix(xtab, prevalence = 0.25)
```