# Multinomial Logistic Regression

Code

Hide

```
library(car)
library(mlogit)
```

Hide

```
df<- read.delim('/home/atrides/Desktop/R/statistics_with_R/08_LogisticRegression/Data_Files/Chat-Up Lines.dat', header=TRUE)

head(df)
```

| | Success | Funny | Sex | Good_Mate | Gender |
|---|---|---|---|---|---|
| | <chr> | <int> | <int> | <int> | <chr> |
| 1 | Get Phone Number | 3 | 7 | 6 | Male |
| 2 | Go Home with Person | 5 | 7 | 2 | Male |
| 3 | Get Phone Number | 4 | 6 | 6 | Male |
| 4 | Go Home with Person | 3 | 7 | 5 | Male |
| 5 | Get Phone Number | 5 | 1 | 6 | Male |
| 6 | Get Phone Number | 4 | 7 | 5 | Male |

6 rows

Hide

```
# checking whether Success and Gender are factor vectors

is.factor(df$Success)
```

```
[1] FALSE
```

Hide

```
is.factor(df$Gender)
```

```
[1] FALSE
```

Hide

```
# converting Success and Gender into factor vectors
df$Success<- as.factor(df$Success)

# dont use ordered=TRUE in case of df$Gender, as it is just a nominal catego
rical variable
# use ordered=TRUE in case of ordinal categorical variables only
df$Gender<-  factor(df$Gender, levels = c('Male', 'Female'))
```

Hide

```
# to count the values of each factor
print(summary(df$Gender))
```

```
  Male Female
   348    672
```

Hide

```
print(summary(df$Success))
```

```
    Get Phone Number  Go Home with Person No response/Walk Off
                 485                  135                  400
```

Hide

```
mlchat<- mlogit.data(df, choice='Success', shape='wide')
head(mlchat , 15)
```

```
~~~~~~~
 first 15 observations out of 3060
~~~~~~~
```

| | Succ… | Fu… | Se | Good_M… | Gen… | c… | alt | idx |
|---|---|---|---|---|---|---|---|---|
| | <lgl> | <int> | <int> | <int> | <fctr> | <dbl> | <fctr> | <S3: idx> |
| 1 | TRUE | 3 | 7 | 6 | Male | 1 | Get Phone Number | <S3: idx> |
| 2 | FALSE | 3 | 7 | 6 | Male | 1 | Go Home with Person | <S3: idx> |
| 3 | FALSE | 3 | 7 | 6 | Male | 1 | No response/Walk Off | <S3: idx> |
| 4 | FALSE | 5 | 7 | 2 | Male | 2 | Get Phone Number | <S3: idx> |
| 5 | TRUE | 5 | 7 | 2 | Male | 2 | Go Home with Person | <S3: idx> |
| 6 | FALSE | 5 | 7 | 2 | Male | 2 | No response/Walk Off | <S3: idx> |
| 7 | TRUE | 4 | 6 | 6 | Male | 3 | Get Phone Number | <S3: idx> |

| | Succ... | Fu... | Se | Good_M... | Gen... | c... | alt | idx |
|---|---|---|---|---|---|---|---|---|
| | <lgl> | <int> | <int> | <int> | <fctr> | <dbl> | <fctr> | <S3: idx> |
| 8 | FALSE | 4 | 6 | 6 | Male | 3 | Go Home with Person | <S3: idx> |
| 9 | FALSE | 4 | 6 | 6 | Male | 3 | No response/Walk Off | <S3: idx> |
| 10 | FALSE | 3 | 7 | 5 | Male | 4 | Get Phone Number | <S3: idx> |

1-10 of 15 rows Previous **1** 2 Next

```
~~~ indexes ~~~~
```

| | chid | alt |
|---|---|---|
| | <dbl> | <fctr> |
| 1 | 1 | Get Phone Number |
| 2 | 1 | Go Home with Person |
| 3 | 1 | No response/Walk Off |
| 4 | 2 | Get Phone Number |
| 5 | 2 | Go Home with Person |
| 6 | 2 | No response/Walk Off |
| 7 | 3 | Get Phone Number |
| 8 | 3 | Go Home with Person |
| 9 | 3 | No response/Walk Off |
| 10 | 4 | Get Phone Number |

1-10 of 15 rows Previous **1** 2 Next

```
indexes:  1, 2
```

Hide

```
m01 <- mlogit(Success ~ 1 | Good_Mate + Funny + Gender + Sex + Gender:Sex +
Funny:Gender, data = mlchat, reflevel = "No response/Walk Off")
summary(m01)
```

```
Call:
mlogit(formula = Success ~ 1 | Good_Mate + Funny + Gender + Sex +
    Gender:Sex + Funny:Gender, data = mlchat, reflevel = "No response/Walk O
ff",
    method = "nr")

Frequencies of alternatives:choice
No response/Walk Off     Get Phone Number  Go Home with Person
             0.39216              0.47549              0.13235

nr method
6 iterations, 0h:0m:0s
g'(-H)^-1g = 0.00121
successive function values within tolerance limits

Coefficients :
                                  Estimate Std. Error z-value  Pr(>|z|)
(Intercept):Get Phone Number     -1.783070   0.669772 -2.6622 0.0077631 **
(Intercept):Go Home with Person  -4.286354   0.941398 -4.5532 5.284e-06 ***
Good_Mate:Get Phone Number        0.131840   0.053726  2.4539 0.0141306 *
Good_Mate:Go Home with Person     0.130019   0.083521  1.5567 0.1195351
Funny:Get Phone Number            0.139389   0.110126  1.2657 0.2056135
Funny:Go Home with Person         0.318456   0.125302  2.5415 0.0110376 *
GenderFemale:Get Phone Number    -1.646223   0.796247 -2.0675 0.0386891 *
GenderFemale:Go Home with Person -5.626369   1.328589 -4.2348 2.287e-05 ***
Sex:Get Phone Number              0.276206   0.089197  3.0966 0.0019577 **
Sex:Go Home with Person           0.417283   0.122083  3.4180 0.0006307 ***
GenderFemale:Get Phone Number    -0.348326   0.105875 -3.2900 0.0010020 **
GenderFemale:Go Home with Person -0.476639   0.163434 -2.9164 0.0035409 **
Funny:Get Phone Number            0.492441   0.139992  3.5176 0.0004354 ***
Funny:Go Home with Person         1.172404   0.199240  5.8844 3.996e-09 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Log-Likelihood: -868.74
McFadden R^2:  0.13816
Likelihood ratio test : chisq = 278.52 (p.value = < 2.22e-16)
```

Hide

```
# to see what can we access of m01 object
names(m01)
```

```
 [1] "coefficients"  "logLik"        "gradient"      "hessian"       "est.st
at"      "fitted.values" "probabilities"
 [8] "linpred"       "indpar"        "residuals"     "omega"         "rpar"
"nests"         "model"
[15] "freq"          "formula"       "call"
```

Hide

```
data.frame(exp(m01$coefficients))
```

| exp.m01.coefficients. <dbl> |
|---|
| 0.16812128 |
| 0.01375498 |
| 1.14092570 |
| 1.13885057 |
| 1.14957104 |
| 1.37500360 |
| 0.19277659 |
| 0.00360163 |
| 1.31811957 |
| 1.51783194 |

1-10 of 14 rows Previous **1** 2 Next

Hide

```
# Checking Some assumptions

# Assumption of Multicollinearity

# Collinearity is a property of the design matrix. Variance inflation factor
s
# are calculated based on linear regressions of the predictors against all o
ther predictors. That means you can just
# fit with glm instead of mlogit and pass that fit to vif to get the vif
m02 <- glm(Success ~Good_Mate + Funny + Gender + Sex, data = df, family=bino
mial())
car::vif(m02)
```

```
Good_Mate     Funny    Gender       Sex
 1.027076  1.304700  1.262266  1.017565
```

Hide

```
# also correlation value will show that there is no problem of multicollinea
rity.So, assumption of
# multicollinearity has been followed.
cor(df[, cbind('Good_Mate', 'Funny', 'Sex')])
```

```
           Good_Mate     Funny        Sex
Good_Mate 1.00000000 0.1632098 0.03794612
Funny     0.16320983 1.0000000 0.11560845
Sex       0.03794612 0.1156084 1.00000000
```

Hide

```


# Assumption of Linearity
# creating log transformed variables for continuous  variables
mlchat$logFunny<- log(mlchat$Funny)*mlchat$Funny
mlchat$logSex<- log(mlchat$Sex)*mlchat$Sex
mlchat$logGood_Mate<- log(mlchat$Good_Mate)*mlchat$Good_Mate

rough_model <- mlogit(Success ~ 1 | Good_Mate + Funny + Gender + Sex + Gende
r:Sex +  Funny:Gender + logFunny + logSex + logGood_Mate, data = mlchat, ref
level = "No response/Walk Off")
summary(rough_model)
```

```
Call:
mlogit(formula = Success ~ 1 | Good_Mate + Funny + Gender + Sex +
    Gender:Sex + Funny:Gender + logFunny + logSex + logGood_Mate,
    data = mlchat, reflevel = "No response/Walk Off", method = "nr")

Frequencies of alternatives:choice
No response/Walk Off     Get Phone Number  Go Home with Person
             0.39049              0.47572              0.13380

nr method
6 iterations, 0h:0m:0s
g'(-H)^-1g = 1.55E-07
gradient close to zero

Coefficients :
                                  Estimate Std. Error z-value  Pr(>|z|)
(Intercept):Get Phone Number     -2.486106   1.671547 -1.4873 0.1369333
(Intercept):Go Home with Person   0.587018   2.318365  0.2532 0.8001111
Good_Mate:Get Phone Number       -0.537751   0.500832 -1.0737 0.2829497
Good_Mate:Go Home with Person    -1.464688   0.670852 -2.1833 0.0290119 *
Funny:Get Phone Number            1.191819   0.600410  1.9850 0.0471434 *
Funny:Go Home with Person         2.057479   0.904690  2.2742 0.0229518 *
GenderFemale:Get Phone Number    -2.092638   0.878826 -2.3812 0.0172575 *
GenderFemale:Go Home with Person -7.150868   1.429676 -5.0017 5.681e-07 ***
Sex:Get Phone Number              0.502241   0.505971  0.9926 0.3208912
Sex:Go Home with Person          -2.094577   0.630662 -3.3212 0.0008962 ***
logFunny:Get Phone Number        -0.454160   0.253361 -1.7925 0.0730459 .
logFunny:Go Home with Person     -0.750941   0.369109 -2.0345 0.0419043 *
logSex:Get Phone Number          -0.081009   0.196931 -0.4114 0.6808098
logSex:Go Home with Person        0.963624   0.247381  3.8953 9.808e-05 ***
logGood_Mate:Get Phone Number     0.247234   0.188677  1.3104 0.1900759
logGood_Mate:Go Home with Person  0.604187   0.256594  2.3546 0.0185407 *
GenderFemale:Get Phone Number    -0.351405   0.117592 -2.9883 0.0028050 **
GenderFemale:Go Home with Person -0.448573   0.146268 -3.0668 0.0021637 **
Funny:Get Phone Number            0.583235   0.143020  4.0780 4.543e-05 ***
Funny:Go Home with Person         1.428894   0.230197  6.2073 5.391e-10 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Log-Likelihood: -847.89
McFadden R^2:  0.15096
Likelihood ratio test : chisq = 301.52 (p.value = < 2.22e-16)
```

Hide

```
# by seeing the log interaction term, its pretty clear that many values has
significance p<0.05,
# hence the assumption of linearity of logit has been violated
```