

Chapter 8. Logistic Regression

Daehyeog Lee

2023-11-05

Table of contents

8.1. What will this chapter tell me?	2
8.2. Background to logistic regression	2
8.3. What are the principles behind logistic regression?	2
8.3.1. Assessing the model: the log-likelihood statistic	2
8.3.2. Assessing the model: the deviance statistic	3
8.3.3. Assessing the model: R and R^2	3
8.3.4. Assessing the model: information criteria	3
8.3.5. Assessing the contribution of predictors: the z-statistic	4
8.3.6. The odds ratio	4
8.3.7. Methods of logistic regression	4
8.4. Assumptions and things that can go wrong	5
8.4.1. Assumptions	5
8.4.2. Incomplete information from the predictors	5
8.4.3. Complete separation	5
8.5. Packages used in this chapter	6
8.6. Binary logistic regression: an example that will make you feel eel	6
8.6.1. Preparing the data	6
8.6.2. The main logistic regression analysis	7
8.6.3. Basic logistic regression analysis using R	7
8.6.4. Interpreting a basic logistic regression	8
8.6.5. Model 1: Intervention only	9
8.6.6. Model 2: Intervention and Duration as predictors	12
8.6.7. Casewise diagnostics in logistic regression	13
8.6.8. Calculating the effect size	20
8.7. How to report logistic regression	20
8.8. Testing assumptions: another example	20
8.8.1 Testing for multicollinearity	24
8.8.2. Testing for linearity of the logit	26

8.9. Predicting several categories: multinomial logistic regression	31
8.9.1. Running multinomial logistic regression in R	31
8.9.2. Interpreting the multinomial logistic regression output	44
8.9.3. Reporting the results	47

8.1. What will this chapter tell me?

Logistic regression allows us to predict categorical outcomes based on predictor variables.

8.2. Background to logistic regression

Binary logistic regression predicts membership of only two categorical outcomes.

Multinomial (or polychotomous) logistic regression predict membership of more than two categories.

8.3. What are the principles behind logistic regression?

In logistic regression, instead of predicting the value of a variable Y from a predictor variable X_1 or several predictor variables (X_s), we predict the **probability** of Y occurring given known values of X_1 (or X_s).

We cannot apply linear regression directly to a situation in which the outcome variable is categorical. It is because one of the assumptions of linear regression is that the relationship between variables is linear. When the outcome variable is categorical, this assumption is violated.

8.3.1. Assessing the model: the log-likelihood statistic

We denote $P(Y_i)$ as the probability that Y occurs for the i th person, and denote Y_i as the actual outcome for the i th person. So, for a given person, Y will be either 0 or 1, and the predicted value $P(Y)$ will be a value between 0 and 1.

In logistic regression, we can use the observed and predicted values to assess the fit of the model. The measure we use is the **log-likelihood**. The log-likelihood is an indicator of how much unexplained information there is after the model has been fitted.

8.3.2. Assessing the model: the deviance statistic

The **deviance** is closely related to the log-likelihood.

$$\text{deviance} = -2 * \log\text{-likelihood}$$

The deviance is often referred to as **-2LL** because of the way it is calculated. It's actually rather convenient to always use the deviance rather than the log-likelihood because it has a chi-square distribution, which makes it easy to calculate the significance of the value.

Like multiple regression, our baseline model is the model that gives us the best prediction when we know nothing other than the values of the outcome. If we add one or more predictors to the model, we can compute the improvement of the model as follows:

$$\text{chi-square} = (-2\text{LL}(\text{baseline})) - (-2\text{LL}(\text{new}))$$

$$= 2\text{LL}(\text{new}) - 2\text{LL}(\text{baseline})$$

$$\text{df} = k_{\text{new}} - k_{\text{baseline}}$$

This difference is known as a likelihood ratio, and has a chi-square distribution with degrees of freedom equal to the number of parameters, k . The number of parameters in the baseline model will always be 1.

8.3.3. Assessing the model: R and R^2

The multiple correlation coefficient R and the corresponding R^2 are useful measures of how well the model fits the data. We can calculate a more literal version of the multiple correlation in logistic regression known as the R-statistic. Wald statistic, Hosmer and Lemeshow's R^2 , Cox and Snell's R^2 , Nagelkerke's R^2 take different formulas to calculate R^2 in logistic regression.

8.3.4. Assessing the model: information criteria

Akaike information criterion (AIC) and the Bayes information criterion (BIC) are used to judge model fit.

$$\text{AIC} = -2\text{LL} + 2k$$

$$\text{BIC} = -2\text{LL} + 2K * \log(n)$$

k is the number of predictors in the model. n is the number of cases in the model.

8.3.5. Assessing the contribution of predictors: the z-statistic

The **z-statistic**, which follows the **normal distribution**, tells the individual contribution of predictors. It tells us whether the b -coefficient for that predictor is significantly different from zero. If the coefficient is significantly different from zero then we can assume that the predictor is making a significant contribution to the prediction of the outcome (Y). z is the value of the regression coefficient (b) divided by its associated standard error. z -statistic should be used a little cautiously, because when b is large, the standard error tends to become inflated, resulting in the z -statistic being underestimated.

8.3.6. The odds ratio

The value of the **odds ratio** is crucial to the interpretation of logistic regression. Odds ratio is the exponential of B and is an indicator of the change in odds resulting from a unit change in the predictor. The **odds** of an event occurring are defined as the probability of an event occurring divided by the probability of that event not occurring.

$$\text{odds} = P(\text{event}) / P(\text{no event})$$

$$\Delta \text{ odds} = \text{odds after a unit change in the predictor} / \text{original odds}$$

If the value of the proportionate change in odds is greater than 1 then it indicates that as the predictor increases, the odds of the outcome occurring increase.

8.3.7. Methods of logistic regression

8.3.7.1. The forced entry method

It is a default method that simply places predictors into the regression model in one block, and estimate parameters for each predictor.

8.3.7.2. Stepwise methods

The forward method includes only a constant and then adds single predictors to the model based on the criterion that adding the variable must improve the AIC or BIC. The backward method begins the model with all predictors included. The computer then tests whether any of these predictors can be removed from the model without increasing the information criterion.

8.3.7.3. How do I select a method

Stepwise methods are considered to have no value for theory testing. However, they are defensible when used in situations where causality is not of interest and we merely wish to find a model to fit our data.

If we do decide to use a stepwise method then the backward method is preferable to the forward method. This is because of **suppressor effects**, which occur when a predictor has a significant effect but only when another variable is held constant.

8.4. Assumptions and things that can go wrong

8.4.1. Assumptions

- **Linearity:** In ordinary regression we assumed that the outcome had linear relationships with the predictors. In logistic regression, the outcome is categorical so this assumption is violated. The linearity assumption in logistic regression, therefore, is that there is a linear relationship between any continuous predictors and the logit of the outcome variable.
- **Independence of errors:** The cases of data should not be related. For example, we cannot measure the same people at different points in time.
- **Multicollinearity:** Predictors should not be too highly correlated.

8.4.2. Incomplete information from the predictors

Whenever samples are broken down into categories and one or more combinations are empty it creates problems. These will probably be signalled by coefficients that have unreasonably large standard errors.

8.4.3. Complete separation

Complete separation is when the outcome variable is perfectly predicted by one variable or a combination of variables.

8.5. Packages used in this chapter

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

Loading required package: carData

Loading required package: dfidx

Attaching package: 'dfidx'

The following object is masked from 'package:stats':

filter

8.6. Binary logistic regression: an example that will make you feel eel

- Outcome(dependent variable): **Cured** (cured / not cured)
- Predictor (independent variable): **Intervention** (intervention / no treatment)
- Predictor (independent variable): **Duration** (the number of days before treatment that the patient that had the problem)

8.6.1. Preparing the data

```
eelData <- read.delim("eel.dat", header = TRUE)
# The head() function shows the first six rows of the dataframe
head(eelData)
```

	Cured	Intervention	Duration
1	Not Cured	No Treatment	7
2	Not Cured	No Treatment	7
3	Not Cured	No Treatment	6
4	Cured	No Treatment	8
5	Cured	Intervention	7
6	Cured	No Treatment	6

When we read in data that are text strings like this, R helpfully converts them to factors. R creates levels of the factor by taking the text string in alphabetical order and assigning them ascending numerical values. So, *Cured* will be the baseline category because it is first. However, we might want *Not Cured* to be the baseline category. Fortunately, the `relevel()` function lets us specify the baseline category for a factor.

```
eelData$Cured <- factor(eelData$Cured, ordered = FALSE)
eelData$Cured <- relevel(eelData$Cured, "Not Cured")
eelData$Intervention <- factor(eelData$Intervention, ordered = FALSE)
eelData$Intervention <- relevel(eelData$Intervention, "No Treatment")
```

8.6.2. The main logistic regression analysis

8.6.2.1. Basic logistic regression analysis using R Commander

Setting the baseline category to one that we want: **Data -> Manage variables in active data set -> Reorder factor levels...**

To run binary logistic regression, choose **Statistics -> Fit models -> Generalized linear model...**

When we use a generalized linear model we need to specify a family and link function. The family relates to the type of distribution that we assume. For linear regression, we would choose normal distribution. For logistic regression, we would choose binomial. We also need to choose a link function. For logistic regression, we choose the logit.

8.6.3. Basic logistic regression analysis using R

To do logistic regression, we use the `glm()` function. *glm* stands for 'generalized linear model' and the general form of this function is:

```
newModel <- glm(outcome ~ predictor(s), data = dataframe, family = name of a
distribution, na.action = an action)
```

```
eelModel.1 <- glm(Cured ~ Intervention, data = eelData, family = binomial())
eelModel.2 <- glm(
  Cured ~ Intervention + Duration, data = eelData, family = binomial()
)
```

The first command created a model called `eelModel.1` in which **Cured** is predicted from only **Intervention** (*Cured ~ Intervention*) based on a logit function.

The second command created a model called `eelModel.2` in which **Cured** is predicted from both **Intervention** and **Duration** (*Cured ~ Intervention + Duration*)

8.6.4. Interpreting a basic logistic regression

To see the models that we have just generated, we need to execute the `summary()` function.

```
summary(eelModel.1)
```

Call:

```
glm(formula = Cured ~ Intervention, family = binomial(), data = eelData)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-0.2877	0.2700	-1.065	0.28671
InterventionIntervention	1.2287	0.3998	3.074	0.00212 **

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

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 154.08 on 112 degrees of freedom
Residual deviance: 144.16 on 111 degrees of freedom
AIC: 148.16

Number of Fisher Scoring iterations: 4

```
summary(eelModel.2)
```

Call:

```
glm(formula = Cured ~ Intervention + Duration, family = binomial(),  
     data = eelData)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-0.234660	1.220563	-0.192	0.84754
InterventionIntervention	1.233532	0.414565	2.975	0.00293 **
Duration	-0.007835	0.175913	-0.045	0.96447

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

(Dispersion parameter for binomial family taken to be 1)


```
Null deviance: 154.08  on 112  degrees of freedom
Residual deviance: 144.16  on 110  degrees of freedom
AIC: 150.16
```

```
Number of Fisher Scoring iterations: 4
```

8.6.5. Model 1: Intervention only

The overall fit of the model is assessed using the deviance statistic (-2LL). The larger values of the deviance statistic indicate poorer-fitting statistical models. **R** provides two deviance statistics: the null deviance and the residual deviance. The null deviance is the deviance of the model that contains no predictors other than the constant, which is -2LL(base). The residual deviance is the deviance for the model, which is -2LL(new).

For the null model, -2LL=154.08, but when **Intervention** has been included this value has been reduced to 144.16. This reduction tells us that the model is better at predicting whether someone was cured than it was before **Intervention** was added.

Model chi-square statistic assess how much better the model predicts the outcome variable. The value of the model chi-square statistic is equal to -2LL with **Intervention** included minus the value of -2LL when only the constant was in the model ($154.08 - 144.16 = 9.92$). The model chi-square is an analogue of the F-test for the linear regression. We can calculate this value and the degree of freedom associated with the chi-square statistic by executing:

```
modelChi <- eelModel.1$null.deviance - eelModel.1$deviance
modelChi
```

```
[1] 9.926201
```

```
chidf <- eelModel.1$df.null - eelModel.1$df.residual
chidf
```

```
[1] 1
```

As we can see, the change in degrees of freedom is 1, which reflects the fact that we have only one variable in the model.

To calculate the probability associated with this chi-square statistic we can use the `pchisq()` function. The probability we want is 1 minus the value of the `pchisq()` function, so we can execute:

```
chisq.prob <- 1 - pchisq(modelChi, chidf)
chisq.prob
```

```
[1] 0.001629425
```

In other words, the p-value is .002, so we can reject the null hypothesis that the model is not better than chance at predicting the outcome. This value is the likelihood ratio p-value of the model because we only had one predictor in the model. We can report that including **Intervention** produced a significant improvement in the fit of the model, $\chi^2(1) = 9.93, p = .002$

The coefficient in logistic regression is interpreted similarly as in linear regression, because it represents the change in the *logit* of the outcome variable associated with a one-unit change in the predictor variable. The z-statistic has a normal distribution and tells us whether the *b* coefficient for that predictor is significantly different from zero. We could say the **Intervention** was a significant predictor of being cured, $b = 1.23, z = 3.07, p < .002$.

Below is the various commands to calculate R^2 from -2LL and degree of freedom:

```
# Hosmer and Lemeshow's measure
R2.hl <- modelChi/eelModel.1$null.deviance
R2.hl
```

```
[1] 0.06442071
```

```
# Cox and Snell's R^2
R.cs <- 1 - exp((eelModel.1$deviance - eelModel.1$null.deviance) / 113)
R.cs
```

```
[1] 0.08409487
```

```
# Nagelkerke's estimate
R.n <- R.cs / (1-(exp(-eelModel.1$null.deviance / 113)))
R.n
```

```
[1] 0.112992
```

We can wrap up several commands into a function like below:

```

logisticPseudoR2s <- function(LogModel){
  dev <- LogModel$deviance
  nullDev <- LogModel$null.deviance
  modelN <- length(LogModel$fitted.values)
  R.l <- 1 - dev / nullDev
  R.cs <- 1 - exp(-(nullDev - dev) / modelN)
  R.n <- R.cs / (1-(exp(-(nullDev / modelN))))
  cat("Pseudo R^2 for logistic regression\n")
  cat("Hosmer and Lemeshow R^2", round(R.l, 3), "\n")
  cat("Cox and Snell R^2      ", round(R.cs, 3), "\n")
  cat("Nagelkerke R^2        ", round(R.n, 3), "\n")
}
logisticPseudoR2s(eelModel.1)

```

```

Pseudo R^2 for logistic regression
Hosmer and Lemeshow R^2 0.064
Cox and Snell R^2      0.084
Nagelkerke R^2        0.113

```

The `cat()` function is used to print the text in quotes.

To get the odds ratios, we must first calculate the odds of a patient being cured given that they didn't have the intervention. We then calculate the odds of a patient being cured given that they did have the intervention. The coefficients are stored in a variable called `coefficients`. We can access this variable as:

```
eelModel.1$coefficients
```

```

(Intercept) InterventionIntervention
-0.2876821      1.2286654

```

```
exp(eelModel.1$coefficients)
```

```

(Intercept) InterventionIntervention
0.750000    3.416667

```

```
exp(confint(eelModel.1))
```

Waiting for profiling to be done...

	2.5 %	97.5 %
(Intercept)	0.4374531	1.268674
InterventionIntervention	1.5820127	7.625545

The important thing about this confidence interval is that it doesn't cross 1. This is important because values greater than 1 mean that as the predictor variable increases, so do the odds of being cured. If the lower limit had been below 1 then it would tell us that there is a chance that in the population the direction of the relationship is the opposite to what we have observed.

8.6.6. Model 2: Intervention and Duration as predictors

```
summary(eelModel.2)
```

Call:

```
glm(formula = Cured ~ Intervention + Duration, family = binomial(),
     data = eelData)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-0.234660	1.220563	-0.192	0.84754
InterventionIntervention	1.233532	0.414565	2.975	0.00293 **
Duration	-0.007835	0.175913	-0.045	0.96447

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

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 154.08 on 112 degrees of freedom
 Residual deviance: 144.16 on 110 degrees of freedom
 AIC: 150.16

Number of Fisher Scoring iterations: 4

```
modelChi <- eelModel.1$deviance - eelModel.2$deviance
chidf <- eelModel.1$df.residual - eelModel.2$df.residual
chisq.prob <- 1 - pchisq(modelChi, chidf)
modelChi; chidf; chisq.prob
```

```
[1] 0.001983528
```

```
[1] 1
```

```
[1] 0.9644765
```

We can see that the b estimate for **Duration** is -0.008, and the probability value associated with that variable is not significant (0.964). Comparing model 1 with model 2, we can see that the deviance for the two models is the same (144.16), suggesting that model 2 is not an improvement over model 1. In addition, we can see that the AIC is higher in model 2 (150.16) than model 1 (148.16), indicating that model 1 is the better model. We can also find that the difference between the models (`modelChi`) is 0.00198, with one degree of freedom (`chidf`) and a p-value (`chisq.prob`) of 0.964. Since this value is greater than .05, we can conclude that model 2 is not a significant improvement over model 1.

With the `anova()` function, we can easily compare our two models by executing:

```
anova(eelModel.1, eelModel.2)
```

Analysis of Deviance Table

Model 1: Cured ~ Intervention

Model 2: Cured ~ Intervention + Duration

	Resid. Df	Resid. Dev	Df	Deviance
1	111	144.16		
2	110	144.16	1	0.0019835

8.6.7. Casewise diagnostics in logistic regression

8.6.7.1. Obtaining residuals

To obtain residuals, we can use the `resid()` function and include the model name within it.

The fitted values are the predicted probabilities of Y occurring given the values of each predictor for a given participant. Predicted probabilities are obtained with the `fitted()` function.

For example, as a basic set of diagnostic statistics we might execute:

```
eelData$predicted.proBABILITIES <- fitted(eelModel.1)
eelData$standardized.residuals <- rstandard(eelModel.1)
eelData$studentized.residuals <- rstudent(eelModel.1)
eelData$dfbeta <- dfbeta(eelModel.1)
eelData$dffit <- dffits(eelModel.1)
eelData$leverage <- hatvalues(eelModel.1)
```

8.6.7.2. Predicted probabilities

Let's take a look at the predicted probabilities. By using a `head()` function to look at the first few cases:

```
head(
  eelData[, c("Cured", "Intervention", "Duration", "predicted.probabilities")]
)
```

	Cured	Intervention	Duration	predicted.probabilities
1	Not Cured	No Treatment	7	0.4285714
2	Not Cured	No Treatment	7	0.4285714
3	Not Cured	No Treatment	6	0.4285714
4	Cured	No Treatment	8	0.4285714
5	Cured	Intervention	7	0.7192982
6	Cured	No Treatment	6	0.4285714

We can find from the model that the only significant predictor of being cured was having the intervention. Assuming that the model is accurate and the intervention has some substantive significance, we could conclude that our intervention is the single best predictor of getting better. Furthermore, the duration of the constipation pre-intervention and its interaction with the intervention did not significantly predict whether a person got better.

8.6.7.3. Interpreting residuals

It is important to examine the residuals to be sure that the model is good. The main purpose of examining residuals is to isolate points for which the model fits poorly (studentized residuals, standardized residuals, deviance statistics), and isolate points that exert an undue influence on the model (Cook's distance):

```
eelData[, c("leverage", "studentized.residuals", "dfbeta")]
```

	leverage	studentized.residuals	dfbeta.(Intercept)
1	0.01785714	-1.0643627	-3.886912e-02
2	0.01785714	-1.0643627	-3.886912e-02
3	0.01785714	-1.0643627	-3.886912e-02
4	0.01785714	1.3110447	4.782751e-02
5	0.01754386	0.8160435	-8.263528e-18
6	0.01785714	1.3110447	4.782751e-02
7	0.01754386	-1.6083171	1.622678e-17
8	0.01754386	0.8160435	-8.263528e-18

9	0.01785714	1.3110447	4.782751e-02
10	0.01785714	-1.0643627	-3.886912e-02
11	0.01754386	0.8160435	-8.263528e-18
12	0.01785714	1.3110447	4.782751e-02
13	0.01785714	1.3110447	4.782751e-02
14	0.01754386	-1.6083171	1.622678e-17
15	0.01785714	-1.0643627	-3.886912e-02
16	0.01785714	-1.0643627	-3.886912e-02
17	0.01754386	0.8160435	-8.263528e-18
18	0.01754386	-1.6083171	1.622678e-17
19	0.01754386	0.8160435	-8.263528e-18
20	0.01754386	0.8160435	-8.263528e-18
21	0.01785714	-1.0643627	-3.886912e-02
22	0.01754386	0.8160435	-8.263528e-18
23	0.01754386	0.8160435	-8.263528e-18
24	0.01785714	-1.0643627	-3.886912e-02
25	0.01754386	0.8160435	-8.263528e-18
26	0.01754386	0.8160435	-8.263528e-18
27	0.01785714	1.3110447	4.782751e-02
28	0.01785714	1.3110447	4.782751e-02
29	0.01754386	0.8160435	-8.263528e-18
30	0.01754386	0.8160435	-8.263528e-18
31	0.01785714	-1.0643627	-3.886912e-02
32	0.01785714	1.3110447	4.782751e-02
33	0.01754386	0.8160435	-8.263528e-18
34	0.01754386	-1.6083171	1.622678e-17
35	0.01754386	0.8160435	-8.263528e-18
36	0.01785714	1.3110447	4.782751e-02
37	0.01785714	1.3110447	4.782751e-02
38	0.01754386	-1.6083171	1.622678e-17
39	0.01785714	1.3110447	4.782751e-02
40	0.01754386	0.8160435	-8.263528e-18
41	0.01785714	1.3110447	4.782751e-02
42	0.01754386	-1.6083171	1.622678e-17
43	0.01754386	0.8160435	-8.263528e-18
44	0.01754386	-1.6083171	1.622678e-17
45	0.01785714	-1.0643627	-3.886912e-02
46	0.01785714	-1.0643627	-3.886912e-02
47	0.01785714	-1.0643627	-3.886912e-02
48	0.01754386	-1.6083171	1.622678e-17
49	0.01754386	-1.6083171	1.622678e-17
50	0.01785714	1.3110447	4.782751e-02
51	0.01754386	0.8160435	-8.263528e-18

52	0.01754386	0.8160435	-8.263528e-18
53	0.01754386	0.8160435	-8.263528e-18
54	0.01785714	1.3110447	4.782751e-02
55	0.01785714	-1.0643627	-3.886912e-02
56	0.01754386	0.8160435	-8.263528e-18
57	0.01754386	0.8160435	-8.263528e-18
58	0.01785714	-1.0643627	-3.886912e-02
59	0.01754386	0.8160435	-8.263528e-18
60	0.01754386	0.8160435	-8.263528e-18
61	0.01785714	1.3110447	4.782751e-02
62	0.01785714	-1.0643627	-3.886912e-02
63	0.01785714	1.3110447	4.782751e-02
64	0.01754386	-1.6083171	1.622678e-17
65	0.01754386	0.8160435	-8.263528e-18
66	0.01785714	-1.0643627	-3.886912e-02
67	0.01785714	-1.0643627	-3.886912e-02
68	0.01785714	1.3110447	4.782751e-02
69	0.01785714	-1.0643627	-3.886912e-02
70	0.01785714	-1.0643627	-3.886912e-02
71	0.01754386	0.8160435	-8.263528e-18
72	0.01785714	-1.0643627	-3.886912e-02
73	0.01754386	-1.6083171	1.622678e-17
74	0.01754386	0.8160435	-8.263528e-18
75	0.01754386	0.8160435	-8.263528e-18
76	0.01785714	1.3110447	4.782751e-02
77	0.01754386	0.8160435	-8.263528e-18
78	0.01754386	-1.6083171	1.622678e-17
79	0.01785714	1.3110447	4.782751e-02
80	0.01754386	-1.6083171	1.622678e-17
81	0.01785714	1.3110447	4.782751e-02
82	0.01785714	-1.0643627	-3.886912e-02
83	0.01754386	0.8160435	-8.263528e-18
84	0.01754386	0.8160435	-8.263528e-18
85	0.01754386	-1.6083171	1.622678e-17
86	0.01754386	0.8160435	-8.263528e-18
87	0.01785714	1.3110447	4.782751e-02
88	0.01754386	-1.6083171	1.622678e-17
89	0.01754386	0.8160435	-8.263528e-18
90	0.01785714	-1.0643627	-3.886912e-02
91	0.01754386	0.8160435	-8.263528e-18
92	0.01785714	-1.0643627	-3.886912e-02
93	0.01754386	0.8160435	-8.263528e-18
94	0.01754386	0.8160435	-8.263528e-18

95	0.01754386	-1.6083171	1.622678e-17
96	0.01754386	0.8160435	-8.263528e-18
97	0.01785714	1.3110447	4.782751e-02
98	0.01785714	-1.0643627	-3.886912e-02
99	0.01785714	-1.0643627	-3.886912e-02
100	0.01754386	0.8160435	-8.263528e-18
101	0.01785714	-1.0643627	-3.886912e-02
102	0.01785714	-1.0643627	-3.886912e-02
103	0.01754386	0.8160435	-8.263528e-18
104	0.01785714	1.3110447	4.782751e-02
105	0.01785714	-1.0643627	-3.886912e-02
106	0.01785714	1.3110447	4.782751e-02
107	0.01785714	-1.0643627	-3.886912e-02
108	0.01785714	-1.0643627	-3.886912e-02
109	0.01754386	0.8160435	-8.263528e-18
110	0.01754386	0.8160435	-8.263528e-18
111	0.01785714	-1.0643627	-3.886912e-02
112	0.01785714	-1.0643627	-3.886912e-02
113	0.01754386	0.8160435	-8.263528e-18

dfbeta.InterventionIntervention

1	3.886912e-02
2	3.886912e-02
3	3.886912e-02
4	-4.782751e-02
5	3.225994e-02
6	-4.782751e-02
7	-6.334765e-02
8	3.225994e-02
9	-4.782751e-02
10	3.886912e-02
11	3.225994e-02
12	-4.782751e-02
13	-4.782751e-02
14	-6.334765e-02
15	3.886912e-02
16	3.886912e-02
17	3.225994e-02
18	-6.334765e-02
19	3.225994e-02
20	3.225994e-02
21	3.886912e-02
22	3.225994e-02
23	3.225994e-02

24	3.886912e-02
25	3.225994e-02
26	3.225994e-02
27	-4.782751e-02
28	-4.782751e-02
29	3.225994e-02
30	3.225994e-02
31	3.886912e-02
32	-4.782751e-02
33	3.225994e-02
34	-6.334765e-02
35	3.225994e-02
36	-4.782751e-02
37	-4.782751e-02
38	-6.334765e-02
39	-4.782751e-02
40	3.225994e-02
41	-4.782751e-02
42	-6.334765e-02
43	3.225994e-02
44	-6.334765e-02
45	3.886912e-02
46	3.886912e-02
47	3.886912e-02
48	-6.334765e-02
49	-6.334765e-02
50	-4.782751e-02
51	3.225994e-02
52	3.225994e-02
53	3.225994e-02
54	-4.782751e-02
55	3.886912e-02
56	3.225994e-02
57	3.225994e-02
58	3.886912e-02
59	3.225994e-02
60	3.225994e-02
61	-4.782751e-02
62	3.886912e-02
63	-4.782751e-02
64	-6.334765e-02
65	3.225994e-02
66	3.886912e-02

67	3.886912e-02
68	-4.782751e-02
69	3.886912e-02
70	3.886912e-02
71	3.225994e-02
72	3.886912e-02
73	-6.334765e-02
74	3.225994e-02
75	3.225994e-02
76	-4.782751e-02
77	3.225994e-02
78	-6.334765e-02
79	-4.782751e-02
80	-6.334765e-02
81	-4.782751e-02
82	3.886912e-02
83	3.225994e-02
84	3.225994e-02
85	-6.334765e-02
86	3.225994e-02
87	-4.782751e-02
88	-6.334765e-02
89	3.225994e-02
90	3.886912e-02
91	3.225994e-02
92	3.886912e-02
93	3.225994e-02
94	3.225994e-02
95	-6.334765e-02
96	3.225994e-02
97	-4.782751e-02
98	3.886912e-02
99	3.886912e-02
100	3.225994e-02
101	3.886912e-02
102	3.886912e-02
103	3.225994e-02
104	-4.782751e-02
105	3.886912e-02
106	-4.782751e-02
107	3.886912e-02
108	3.886912e-02
109	3.225994e-02

110	3.225994e-02
111	3.886912e-02
112	3.886912e-02
113	3.225994e-02

Note that all cases have DFBetas less than 1, and leverage statistics are very close to the calculated expected value of 0.018. These imply that there are no influential cases having an effect on the model.

- **Leverage:** Lies between 0 (no influence) and 1 (complete influence). The expected leverage is $(k+1)/N$, where k is the number of predictors and N is the sample size.
- **Studentized, Standardized residual:** Only 5% should lie outside ± 1.96 , and about 1% should lie outside ± 2.58 .
- **DFBeta for the constant, DFBeta for the first predictor:** Should be less than 1

8.6.8. Calculating the effect size

We can use the odds ratio as an effect size measure.

8.7. How to report logistic regression

We have to report the beta values and their standard errors and significance value and some general statistics about the model. Reporting the odds ratio and its confidence interval is also highly recommended. If we include the constant or the variables that were not significant predictors, it will also help the readers.

8.8. Testing assumptions: another example

`penalty.dat` contains four variables, each in a separate column

- **Scored:** This is our outcome variable. 0 = Penalty missed, 1 = Penalty scored
- **PSWQ:** This is the first predictor variable and it gives a measure of the degree to which a player worries.
- **Previous:** This is the percentage of penalties scored by a particular player in their career (previous success)
- **Anxious:** This is the third predictor and it is a measure of state anxiety before taking the penalty.

```
# Self Test
penaltyData<-read.delim("penalty.dat", header = TRUE)
head(penaltyData)
```

	PSWQ	Anxious	Previous	Scored
1	18	21	56	Scored Penalty
2	17	32	35	Scored Penalty
3	16	34	35	Scored Penalty
4	14	40	15	Scored Penalty
5	5	24	47	Scored Penalty
6	1	15	67	Scored Penalty

```
#Create the two hierarchical models:
penaltyData$Scored <- factor(penaltyData$Scored, ordered = FALSE)
penaltyData$Scored <- relevel(penaltyData$Scored, "Missed Penalty")

penaltyModel.1 <-
  glm(Scored ~ Previous + PSWQ, data = penaltyData, family = binomial())
penaltyModel.2 <-
  glm(
    Scored ~ Previous + PSWQ + Anxious, data = penaltyData, family = binomial()
  )

summary(penaltyModel.1)
```

Call:

```
glm(formula = Scored ~ Previous + PSWQ, family = binomial(),
    data = penaltyData)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	1.28031	1.67017	0.767	0.44333
Previous	0.06480	0.02209	2.934	0.00335 **
PSWQ	-0.23009	0.07983	-2.882	0.00395 **

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

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 103.638 on 74 degrees of freedom

Residual deviance: 48.662 on 72 degrees of freedom
AIC: 54.662

Number of Fisher Scoring iterations: 6

```
summary(penaltyModel.2)
```

Call:

```
glm(formula = Scored ~ Previous + PSWQ + Anxious, family = binomial(),  
     data = penaltyData)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-11.49256	11.80175	-0.974	0.33016
Previous	0.20261	0.12932	1.567	0.11719
PSWQ	-0.25137	0.08401	-2.992	0.00277 **
Anxious	0.27585	0.25259	1.092	0.27480

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

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 103.638 on 74 degrees of freedom
Residual deviance: 47.416 on 71 degrees of freedom
AIC: 55.416

Number of Fisher Scoring iterations: 6

```
# Model 1  
modelChi <- penaltyModel.1>null.deviance - penaltyModel.1$deviance  
chidf <- penaltyModel.1$df.null - penaltyModel.1$df.residual  
chisq.prob <- 1 - pchisq(modelChi, chidf)  
modelChi; chidf; chisq.prob
```

```
[1] 54.97669
```

```
[1] 2
```

```
[1] 1.1533e-12
```

```
logisticPseudoR2s(penaltyModel.1)
```

```
Pseudo R^2 for logistic regression
Hosmer and Lemeshow R^2 0.53
Cox and Snell R^2      0.52
Nagelkerke R^2        0.694
```

```
exp(penaltyModel.1$coefficients)
```

```
(Intercept)    Previous      PSWQ
  3.5977600    1.0669482    0.7944584
```

```
exp(confint(penaltyModel.1))
```

```
Waiting for profiling to be done...
```

```
                2.5 %      97.5 %
(Intercept) 0.1379603 108.7408589
Previous    1.0263545  1.1216277
PSWQ        0.6614590  0.9115488
```

```
# Model 2
modelChi <- penaltyModel.1$deviance - penaltyModel.2$deviance
chidf <- penaltyModel.1$df.residual - penaltyModel.2$df.residual
chisq.prob <- 1 - pchisq(modelChi, chidf)
modelChi; chidf; chisq.prob
```

```
[1] 1.246267
```

```
[1] 1
```

```
[1] 0.2642667
```

```
logisticPseudoR2s(penaltyModel.2)
```

```
Pseudo R^2 for logistic regression
Hosmer and Lemeshow R^2 0.542
Cox and Snell R^2      0.527
Nagelkerke R^2         0.704
```

```
exp(penaltyModel.2$coefficients)
```

```
(Intercept)    Previous      PSWQ      Anxious
1.020573e-05 1.224593e+00 7.777351e-01 1.317649e+00
```

```
exp(confint(penaltyModel.2))
```

Waiting for profiling to be done...

```

                2.5 %      97.5 %
(Intercept) 1.876236e-16 6.425863e+04
Previous    9.607908e-01 1.612961e+00
PSWQ        6.420582e-01 8.988251e-01
Anxious     8.138225e-01 2.242252e+00
```

```
anova(penaltyModel.1, penaltyModel.2)
```

Analysis of Deviance Table

```
Model 1: Scored ~ Previous + PSWQ
Model 2: Scored ~ Previous + PSWQ + Anxious
  Resid. Df Resid. Dev Df Deviance
1      72      48.662
2      71      47.416  1    1.2463
```

8.8.1 Testing for multicollinearity

(From chapter 7.7.2.4.) Multicollinearity exists when there is a strong correlation between two or more predictors in a regression model. As collinearity increases, there are three problems that arise:

- **Untrustworthy *bs*:** The standard errors of the *b* coefficients increase as collinearity increases.

- **It limits the size of R:** If the two predictors are correlated, then the second predictor is likely to account for similar variance in the outcome to that accounted for by the first predictor. Since R is a measure of the multiple correlation between the predictors and the outcome and that R^2 indicates the variance in the outcome for which the predictors account, having uncorrelated predictors is beneficial.
- **Importance of predictors:** If the predictors are highly correlated, and each accounts for similar variance in the outcome, we can't know which of the two variables is important.

We can get the VIF and tolerance as we did in Chapter 7 by entering the model name into the `vif()` function from the *car* package:

```
vif(penaltyModel.2)
```

```
Previous      PSWQ    Anxious
35.227113  1.089767 35.581976
```

```
1/vif(penaltyModel.2)
```

```
Previous      PSWQ    Anxious
0.02838723 0.91762767 0.02810412
```

The first line gives us the VIF values and the second line gives us the tolerance. A VIF over 10 is usually considered problematic. Therefore, there is collinearity between state anxiety and previous experience of taking penalties, and this dependency results in the model becoming biased.

Let's figure out whether the predictors actually have a correlation:

```
# Self Test
cor(penaltyData[, c("Previous", "PSWQ", "Anxious")])
```

```
          Previous      PSWQ    Anxious
Previous  1.0000000 -0.6435448 -0.9928699
PSWQ      -0.6435448  1.0000000  0.6516416
Anxious   -0.9928699  0.6516416  1.0000000
```

The result shows that **Anxious** and **Previous** are highly negatively correlated ($r = -.99$). It is unclear which of the two variables predicts penalty success in the regression.

If there is a multicollinearity, one obvious solution is to omit one of the variables. However, there is no way of knowing which variable to omit. The better solution is to replace the predictor to another equally important predictor that does not have such strong multicollinearity. Also, we can collect more data to see whether the multicollinearity can be lessened. Or, we can run a factor analysis (Chapter 17) on these predictors and to use the resulting factor scores as a predictor.

8.8.2. Testing for linearity of the logit

We have to check that each variable is linearly related to the log of the outcome variable (**Scored**). To test this assumption, we need to run the logistic regression but include predictors that are the interaction between each predictor and the log of itself. To create the interaction terms of each of the variables with its log, we can execute:

```
penaltyData$logPSWQInt <- log(penaltyData$PSWQ) * penaltyData$PSWQ
```

This command creates a new variable called `logPSWQInt` in the `penaltyData` dataframe that is the variable `penaltyData$PSWQ` multiplied by the log of that variable `log(penaltyData$PSWQ)`.

Similarly,:

```
penaltyData$logAnxInt <- log(penaltyData$Anxious) * penaltyData$Anxious
penaltyData$logPrevInt <- log(penaltyData$Previous + 1) * penaltyData$Previous
# We added a constant (1) because 0 has no log value.
```

```
head(penaltyData)
```

	PSWQ	Anxious	Previous	Scored	logPSWQInt	logAnxInt	logPrevInt
1	18	21	56	Scored Penalty	52.02669	63.93497	226.41087
2	17	32	35	Scored Penalty	48.16463	110.90355	125.42316
3	16	34	35	Scored Penalty	44.36142	119.89626	125.42316
4	14	40	15	Scored Penalty	36.94680	147.55518	41.58883
5	5	24	47	Scored Penalty	8.04719	76.27329	181.94645
6	1	15	67	Scored Penalty	0.00000	40.62075	282.70702

To test the assumption we need to redo the analysis exactly the same as before, except that we should put all variables in a single block (not hierarchically). We create the model by executing:

```
penaltyTest.1 <-
  glm(
```

```

    Scored ~ PSWQ + Anxious + Previous + logPSWQInt + logAnxInt + logPrevInt,
    data = penaltyData, family = binomial()
  )
summary(penaltyTest.1)

```

Call:

```

glm(formula = Scored ~ PSWQ + Anxious + Previous + logPSWQInt +
    logAnxInt + logPrevInt, family = binomial(), data = penaltyData)

```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-3.57212	15.00782	-0.238	0.812
PSWQ	-0.42218	1.10255	-0.383	0.702
Anxious	-2.64804	2.79283	-0.948	0.343
Previous	1.66905	1.48005	1.128	0.259
logPSWQInt	0.04388	0.29672	0.148	0.882
logAnxInt	0.68151	0.65177	1.046	0.296
logPrevInt	-0.31918	0.31687	-1.007	0.314

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 103.638 on 74 degrees of freedom
 Residual deviance: 45.909 on 68 degrees of freedom
 AIC: 59.909

Number of Fisher Scoring iterations: 7

We're interested only in whether the interaction terms are significant. Any interaction that is significant indicates that the main effect has violated the assumption of linearity of the logit. Therefore, we can say that the assumption of linearity of the logit has been met for **PSWQ**, **Anxious**, and **Previous**.

Below is the real research:

```

# Labcoat Leni's Real Research
suicideData <- read.delim("Lacourse et al. (2001) Females.dat", header = TRUE)
head(suicideData)

```

	Age	Age_Group	Drug_Use	Father_Negligence	Gender	Isolation
1	15.83	14-16 Years Old	8		17 Female	6
2	14.92	14-16 Years Old	9		23 Female	8

3	15.33	14-16 Years Old	5	15 Female	18
4	15.83	14-16 Years Old	11	11 Female	9
5	14.92	14-16 Years Old	7	13 Female	5
6	14.58	14-16 Years Old	4	29 Female	15

	Marital_Status	Meaninglessness	Metal	Mother_Negligence	Normlessness
1	Together	10	4.837625	10	6
2	Together	26	6.000000	12	8
3	Together	19	6.000000	16	7
4	Separated or Divorced	13	4.000000	10	5
5	Together	13	8.000000	16	3
6	Separated or Divorced	18	7.000000	18	5

	Self_Estrangement	Suicide_Risk	Vicarious	Worshipping
1	15	Non-Suicidal	5	4
2	20	Non-Suicidal	4	6
3	17	Non-Suicidal	6	3
4	12	Non-Suicidal	3	3
5	6	Non-Suicidal	3	9
6	15	Suicidal	2	4

```

suicideData$Marital_Status <- factor(suicideData$Marital_Status, ordered = FALSE)
suicideData$Marital_Status <- relevel(suicideData$Marital_Status, "Together")
suicideData$Suicide_Risk <- factor(suicideData$Suicide_Risk, ordered = FALSE)
suicideData$Suicide_Risk <- relevel(suicideData$Suicide_Risk, "Non-Suicidal")
# To reorder, click Data -> Convert character variables to factors

suicideModel <-
  glm(
    Suicide_Risk ~ Age + Marital_Status + Mother_Negligence +
    Father_Negligence + Self_Estrangement + Isolation + Normlessness +
    Meaninglessness + Drug_Use + Metal + Worshipping + Vicarious,
    data = suicideData, family = binomial()
  )

summary(suicideModel)

```

Call:

```

glm(formula = Suicide_Risk ~ Age + Marital_Status + Mother_Negligence +
    Father_Negligence + Self_Estrangement + Isolation + Normlessness +
    Meaninglessness + Drug_Use + Metal + Worshipping + Vicarious,
    family = binomial(), data = suicideData)

```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-19.011598	6.208677	-3.062	0.00220	**
Age	0.692853	0.323445	2.142	0.03219	*
Marital_StatusSeparated or Divorced	0.183489	0.677262	0.271	0.78645	
Mother_Negligence	-0.019607	0.053244	-0.368	0.71268	
Father_Negligence	0.084632	0.047858	1.768	0.07699	.
Self_Estrangement	0.155153	0.064832	2.393	0.01670	*
Isolation	-0.005822	0.076152	-0.076	0.93906	
Normlessness	0.191307	0.108841	1.758	0.07880	.
Meaninglessness	-0.066610	0.061024	-1.092	0.27503	
Drug_Use	0.316855	0.103097	3.073	0.00212	**
Metal	0.135552	0.091719	1.478	0.13943	
Worshipping	0.158911	0.129474	1.227	0.21968	
Vicarious	-0.341831	0.196288	-1.741	0.08160	.

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

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 135.533 on 120 degrees of freedom
Residual deviance: 85.116 on 108 degrees of freedom
AIC: 111.12

Number of Fisher Scoring iterations: 6

```
modelChi <- suicideModel$null.deviance - suicideModel$deviance
chidf <- suicideModel$df.null - suicideModel$df.residual
chisq.prob <- 1 - pchisq(modelChi, chidf)
modelChi; chidf; chisq.prob
```

[1] 50.41736

[1] 12

[1] 1.179907e-06

```
exp(suicideModel$coefficients)
```

(Intercept)	5.538188e-09	Age	1.999411e+00
Marital_StatusSeparated or Divorced	1.201402e+00	Mother_Negligence	9.805836e-01
Father_Negligence	1.088317e+00	Self_Estrangement	1.167837e+00
Isolation	9.941947e-01	Normlessness	1.210831e+00
Meaninglessness	9.355597e-01	Drug_Use	1.372803e+00
Metal	1.145169e+00	Worshipping	1.172234e+00
Vicarious	7.104685e-01		

```
exp(confint(suicideModel))
```

Waiting for profiling to be done...

	2.5 %	97.5 %
(Intercept)	9.065656e-15	0.0004577666
Age	1.090546e+00	3.9372709912
Marital_StatusSeparated or Divorced	3.058015e-01	4.4898699168
Mother_Negligence	8.764185e-01	1.0837478703
Father_Negligence	9.932302e-01	1.2015111314
Self_Estrangement	1.035373e+00	1.3401633046
Isolation	8.532270e-01	1.1548163908
Normlessness	9.854717e-01	1.5210438866
Meaninglessness	8.242768e-01	1.0500790337
Drug_Use	1.132532e+00	1.7047218535
Metal	9.596213e-01	1.3819317960
Worshipping	9.084589e-01	1.5205816278
Vicarious	4.714224e-01	1.0272917992

We can report that including all the predictors produced a significant improvement in the fit of the model, $\chi^2(12) = 50.42, p < .0001$.

The most significant predictor was drug use.

8.9. Predicting several categories: multinomial logistic regression

In **Chat-Up Lines.dat**, there is one outcome variable (**Success**) with three categories (no response, phone number, go home with recipient) and four predictors: funniness of the chat-up line (**Funny**), sexual content of the chat-up line (**Sex**), degree to which the chat-up line reflects good characteristics (**Good_Mate**) and the gender of the person being chatted up (**Female** - Scored as 1 = female, 0 = male)

8.9.1. Running multinomial logistic regression in R

The outcome (**Success**) and **Gender** are stored as text, therefore **R** should have imported these variables as factors. We can check this by using `is.factor()` function.

```
chatData <- read.delim("Chat-Up Lines.dat", header = TRUE)
is.factor(chatData$Success)
```

```
[1] FALSE
```

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

```
[1] FALSE
```

```
head(chatData)
```

		Success	Funny	Sex	Good_Mate	Gender
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

Since the result is 'FALSE', we should convert them into factor by using `as.factor()` function.

```
as.factor(chatData$Success)
```

[1]	Get Phone Number	Go Home with Person	Get Phone Number
[4]	Go Home with Person	Get Phone Number	Get Phone Number
[7]	Go Home with Person	No response/Walk Off	Get Phone Number
[10]	Get Phone Number	No response/Walk Off	No response/Walk Off
[13]	Get Phone Number	Get Phone Number	No response/Walk Off
[16]	Get Phone Number	No response/Walk Off	Get Phone Number
[19]	Get Phone Number	Get Phone Number	Get Phone Number
[22]	Go Home with Person	No response/Walk Off	Get Phone Number
[25]	Go Home with Person	No response/Walk Off	Get Phone Number
[28]	Get Phone Number	Go Home with Person	Get Phone Number
[31]	Get Phone Number	Go Home with Person	Get Phone Number
[34]	Get Phone Number	Get Phone Number	Get Phone Number
[37]	Get Phone Number	Get Phone Number	Get Phone Number
[40]	Get Phone Number	No response/Walk Off	No response/Walk Off
[43]	No response/Walk Off	Go Home with Person	Get Phone Number
[46]	No response/Walk Off	Go Home with Person	Get Phone Number
[49]	No response/Walk Off	No response/Walk Off	Get Phone Number
[52]	Get Phone Number	Get Phone Number	Get Phone Number
[55]	Get Phone Number	Get Phone Number	Get Phone Number
[58]	Get Phone Number	No response/Walk Off	Go Home with Person
[61]	Get Phone Number	Get Phone Number	Go Home with Person
[64]	Get Phone Number	No response/Walk Off	No response/Walk Off
[67]	Go Home with Person	Get Phone Number	Get Phone Number
[70]	Get Phone Number	Get Phone Number	No response/Walk Off
[73]	Go Home with Person	Get Phone Number	Get Phone Number
[76]	Get Phone Number	Go Home with Person	No response/Walk Off
[79]	No response/Walk Off	Get Phone Number	Go Home with Person
[82]	No response/Walk Off	No response/Walk Off	Get Phone Number
[85]	Get Phone Number	Go Home with Person	Get Phone Number
[88]	Go Home with Person	Get Phone Number	Get Phone Number
[91]	Get Phone Number	Get Phone Number	Get Phone Number
[94]	Get Phone Number	Get Phone Number	Get Phone Number
[97]	Get Phone Number	Get Phone Number	Get Phone Number
[100]	Get Phone Number	Get Phone Number	Get Phone Number
[103]	No response/Walk Off	Get Phone Number	Get Phone Number
[106]	Get Phone Number	No response/Walk Off	No response/Walk Off
[109]	Get Phone Number	No response/Walk Off	Go Home with Person
[112]	No response/Walk Off	Get Phone Number	Go Home with Person
[115]	Get Phone Number	Go Home with Person	Get Phone Number
[118]	Go Home with Person	Get Phone Number	No response/Walk Off
[121]	Go Home with Person	No response/Walk Off	Get Phone Number
[124]	No response/Walk Off	No response/Walk Off	Get Phone Number
[127]	Get Phone Number	No response/Walk Off	Go Home with Person

[130]	Get Phone Number	Go Home with Person	No response/Walk Off
[133]	No response/Walk Off	Get Phone Number	Get Phone Number
[136]	No response/Walk Off	Get Phone Number	No response/Walk Off
[139]	Go Home with Person	Get Phone Number	Go Home with Person
[142]	Go Home with Person	Go Home with Person	Get Phone Number
[145]	No response/Walk Off	Get Phone Number	No response/Walk Off
[148]	Go Home with Person	No response/Walk Off	Get Phone Number
[151]	Get Phone Number	Get Phone Number	No response/Walk Off
[154]	No response/Walk Off	No response/Walk Off	Get Phone Number
[157]	No response/Walk Off	Get Phone Number	Get Phone Number
[160]	Go Home with Person	Go Home with Person	Get Phone Number
[163]	No response/Walk Off	Get Phone Number	Go Home with Person
[166]	Get Phone Number	Go Home with Person	No response/Walk Off
[169]	Get Phone Number	Go Home with Person	Get Phone Number
[172]	No response/Walk Off	Get Phone Number	Get Phone Number
[175]	Go Home with Person	Get Phone Number	No response/Walk Off
[178]	Get Phone Number	Get Phone Number	No response/Walk Off
[181]	Get Phone Number	Go Home with Person	No response/Walk Off
[184]	Get Phone Number	Get Phone Number	Go Home with Person
[187]	Get Phone Number	Get Phone Number	Get Phone Number
[190]	Get Phone Number	Go Home with Person	No response/Walk Off
[193]	Get Phone Number	Get Phone Number	Get Phone Number
[196]	Get Phone Number	Go Home with Person	Go Home with Person
[199]	Go Home with Person	No response/Walk Off	Get Phone Number
[202]	Go Home with Person	Get Phone Number	Get Phone Number
[205]	Go Home with Person	Get Phone Number	Get Phone Number
[208]	Get Phone Number	Get Phone Number	Get Phone Number
[211]	Go Home with Person	Get Phone Number	Get Phone Number
[214]	No response/Walk Off	Get Phone Number	Go Home with Person
[217]	No response/Walk Off	Get Phone Number	Get Phone Number
[220]	Get Phone Number	Get Phone Number	Get Phone Number
[223]	Get Phone Number	Go Home with Person	Get Phone Number
[226]	Get Phone Number	Get Phone Number	Get Phone Number
[229]	Get Phone Number	Get Phone Number	Get Phone Number
[232]	Get Phone Number	Go Home with Person	No response/Walk Off
[235]	Get Phone Number	Get Phone Number	Get Phone Number
[238]	Get Phone Number	Go Home with Person	Go Home with Person
[241]	Get Phone Number	Get Phone Number	Go Home with Person
[244]	Get Phone Number	Get Phone Number	No response/Walk Off
[247]	Get Phone Number	Go Home with Person	Get Phone Number
[250]	Get Phone Number	Get Phone Number	Get Phone Number
[253]	No response/Walk Off	No response/Walk Off	Get Phone Number
[256]	Get Phone Number	Go Home with Person	Get Phone Number

[259]	Get Phone Number	Go Home with Person	Go Home with Person
[262]	No response/Walk Off	Get Phone Number	No response/Walk Off
[265]	Get Phone Number	Get Phone Number	Get Phone Number
[268]	Get Phone Number	Get Phone Number	Go Home with Person
[271]	Get Phone Number	Get Phone Number	No response/Walk Off
[274]	Get Phone Number	Get Phone Number	No response/Walk Off
[277]	Go Home with Person	Get Phone Number	Get Phone Number
[280]	Go Home with Person	No response/Walk Off	No response/Walk Off
[283]	Get Phone Number	Get Phone Number	Get Phone Number
[286]	Go Home with Person	Get Phone Number	Get Phone Number
[289]	Go Home with Person	Get Phone Number	Get Phone Number
[292]	Get Phone Number	Go Home with Person	Get Phone Number
[295]	Go Home with Person	Go Home with Person	Get Phone Number
[298]	Go Home with Person	Go Home with Person	Get Phone Number
[301]	Get Phone Number	Go Home with Person	Go Home with Person
[304]	Go Home with Person	No response/Walk Off	Get Phone Number
[307]	Go Home with Person	Get Phone Number	Get Phone Number
[310]	Get Phone Number	Get Phone Number	Go Home with Person
[313]	Get Phone Number	Get Phone Number	Get Phone Number
[316]	No response/Walk Off	No response/Walk Off	Get Phone Number
[319]	Go Home with Person	No response/Walk Off	Go Home with Person
[322]	Go Home with Person	Go Home with Person	No response/Walk Off
[325]	No response/Walk Off	Get Phone Number	No response/Walk Off
[328]	Get Phone Number	Go Home with Person	Get Phone Number
[331]	Get Phone Number	Get Phone Number	Get Phone Number
[334]	No response/Walk Off	Get Phone Number	No response/Walk Off
[337]	Go Home with Person	Go Home with Person	Get Phone Number
[340]	Get Phone Number	Get Phone Number	Get Phone Number
[343]	No response/Walk Off	Get Phone Number	Get Phone Number
[346]	Get Phone Number	No response/Walk Off	Get Phone Number
[349]	Get Phone Number	No response/Walk Off	Get Phone Number
[352]	No response/Walk Off	Get Phone Number	Get Phone Number
[355]	Get Phone Number	Get Phone Number	No response/Walk Off
[358]	No response/Walk Off	Get Phone Number	Get Phone Number
[361]	No response/Walk Off	No response/Walk Off	Get Phone Number
[364]	Get Phone Number	No response/Walk Off	Get Phone Number
[367]	No response/Walk Off	No response/Walk Off	Get Phone Number
[370]	Get Phone Number	Get Phone Number	Go Home with Person
[373]	Get Phone Number	Get Phone Number	No response/Walk Off
[376]	No response/Walk Off	No response/Walk Off	Go Home with Person
[379]	Get Phone Number	Get Phone Number	Get Phone Number
[382]	No response/Walk Off	Get Phone Number	Get Phone Number
[385]	Get Phone Number	No response/Walk Off	Go Home with Person

[388]	No response/Walk Off	Get Phone Number	No response/Walk Off
[391]	Get Phone Number	Get Phone Number	Get Phone Number
[394]	No response/Walk Off	No response/Walk Off	Get Phone Number
[397]	No response/Walk Off	Get Phone Number	Get Phone Number
[400]	No response/Walk Off	Get Phone Number	No response/Walk Off
[403]	Go Home with Person	Get Phone Number	No response/Walk Off
[406]	No response/Walk Off	Go Home with Person	Get Phone Number
[409]	Get Phone Number	No response/Walk Off	No response/Walk Off
[412]	Get Phone Number	No response/Walk Off	No response/Walk Off
[415]	Go Home with Person	Get Phone Number	Get Phone Number
[418]	No response/Walk Off	Go Home with Person	No response/Walk Off
[421]	Get Phone Number	No response/Walk Off	Get Phone Number
[424]	Go Home with Person	No response/Walk Off	Get Phone Number
[427]	Go Home with Person	Get Phone Number	Get Phone Number
[430]	Get Phone Number	No response/Walk Off	No response/Walk Off
[433]	Get Phone Number	Get Phone Number	No response/Walk Off
[436]	Get Phone Number	No response/Walk Off	Get Phone Number
[439]	Go Home with Person	Get Phone Number	No response/Walk Off
[442]	No response/Walk Off	No response/Walk Off	No response/Walk Off
[445]	No response/Walk Off	No response/Walk Off	No response/Walk Off
[448]	Get Phone Number	No response/Walk Off	No response/Walk Off
[451]	No response/Walk Off	Get Phone Number	No response/Walk Off
[454]	Get Phone Number	Get Phone Number	Get Phone Number
[457]	No response/Walk Off	No response/Walk Off	Get Phone Number
[460]	Get Phone Number	No response/Walk Off	Go Home with Person
[463]	No response/Walk Off	No response/Walk Off	Get Phone Number
[466]	Get Phone Number	No response/Walk Off	Get Phone Number
[469]	No response/Walk Off	No response/Walk Off	No response/Walk Off
[472]	No response/Walk Off	No response/Walk Off	Go Home with Person
[475]	No response/Walk Off	No response/Walk Off	Get Phone Number
[478]	Get Phone Number	Get Phone Number	No response/Walk Off
[481]	No response/Walk Off	No response/Walk Off	No response/Walk Off
[484]	Get Phone Number	Get Phone Number	No response/Walk Off
[487]	No response/Walk Off	Get Phone Number	No response/Walk Off
[490]	Get Phone Number	Get Phone Number	Get Phone Number
[493]	No response/Walk Off	Get Phone Number	No response/Walk Off
[496]	Get Phone Number	No response/Walk Off	No response/Walk Off
[499]	Get Phone Number	No response/Walk Off	No response/Walk Off
[502]	No response/Walk Off	No response/Walk Off	Get Phone Number
[505]	Get Phone Number	Get Phone Number	Get Phone Number
[508]	No response/Walk Off	Go Home with Person	No response/Walk Off
[511]	Get Phone Number	Get Phone Number	Get Phone Number
[514]	No response/Walk Off	Get Phone Number	No response/Walk Off

[517]	No response/Walk Off	Get Phone Number	Get Phone Number
[520]	Go Home with Person	No response/Walk Off	No response/Walk Off
[523]	No response/Walk Off	No response/Walk Off	Get Phone Number
[526]	No response/Walk Off	No response/Walk Off	Get Phone Number
[529]	Get Phone Number	No response/Walk Off	Get Phone Number
[532]	No response/Walk Off	Get Phone Number	No response/Walk Off
[535]	No response/Walk Off	No response/Walk Off	No response/Walk Off
[538]	No response/Walk Off	No response/Walk Off	No response/Walk Off
[541]	No response/Walk Off	No response/Walk Off	Go Home with Person
[544]	Get Phone Number	Get Phone Number	No response/Walk Off
[547]	No response/Walk Off	No response/Walk Off	No response/Walk Off
[550]	Get Phone Number	Get Phone Number	Get Phone Number
[553]	Get Phone Number	Get Phone Number	No response/Walk Off
[556]	Get Phone Number	Get Phone Number	No response/Walk Off
[559]	Get Phone Number	Get Phone Number	Go Home with Person
[562]	No response/Walk Off	No response/Walk Off	No response/Walk Off
[565]	Get Phone Number	Get Phone Number	Go Home with Person
[568]	No response/Walk Off	No response/Walk Off	No response/Walk Off
[571]	No response/Walk Off	No response/Walk Off	Get Phone Number
[574]	Get Phone Number	No response/Walk Off	No response/Walk Off
[577]	Get Phone Number	Get Phone Number	Get Phone Number
[580]	No response/Walk Off	No response/Walk Off	No response/Walk Off
[583]	No response/Walk Off	Get Phone Number	No response/Walk Off
[586]	No response/Walk Off	Get Phone Number	Go Home with Person
[589]	Go Home with Person	No response/Walk Off	Get Phone Number
[592]	No response/Walk Off	Get Phone Number	Get Phone Number
[595]	No response/Walk Off	No response/Walk Off	No response/Walk Off
[598]	No response/Walk Off	Get Phone Number	No response/Walk Off
[601]	Get Phone Number	Get Phone Number	Go Home with Person
[604]	Go Home with Person	No response/Walk Off	Get Phone Number
[607]	Get Phone Number	Go Home with Person	Get Phone Number
[610]	No response/Walk Off	No response/Walk Off	Get Phone Number
[613]	No response/Walk Off	Go Home with Person	No response/Walk Off
[616]	No response/Walk Off	No response/Walk Off	No response/Walk Off
[619]	Get Phone Number	No response/Walk Off	Get Phone Number
[622]	No response/Walk Off	No response/Walk Off	Get Phone Number
[625]	Get Phone Number	Get Phone Number	Get Phone Number
[628]	Get Phone Number	Get Phone Number	Get Phone Number
[631]	No response/Walk Off	No response/Walk Off	No response/Walk Off
[634]	No response/Walk Off	No response/Walk Off	Get Phone Number
[637]	Get Phone Number	Go Home with Person	Get Phone Number
[640]	No response/Walk Off	Go Home with Person	Get Phone Number
[643]	Get Phone Number	Get Phone Number	No response/Walk Off

[646]	No response/Walk Off	No response/Walk Off	No response/Walk Off
[649]	Get Phone Number	Get Phone Number	No response/Walk Off
[652]	Go Home with Person	No response/Walk Off	No response/Walk Off
[655]	No response/Walk Off	Get Phone Number	Get Phone Number
[658]	Get Phone Number	Get Phone Number	No response/Walk Off
[661]	Get Phone Number	Get Phone Number	No response/Walk Off
[664]	No response/Walk Off	No response/Walk Off	Go Home with Person
[667]	Get Phone Number	No response/Walk Off	No response/Walk Off
[670]	Get Phone Number	Get Phone Number	No response/Walk Off
[673]	No response/Walk Off	Get Phone Number	Get Phone Number
[676]	No response/Walk Off	No response/Walk Off	Get Phone Number
[679]	Get Phone Number	No response/Walk Off	Get Phone Number
[682]	Go Home with Person	Get Phone Number	No response/Walk Off
[685]	Get Phone Number	No response/Walk Off	No response/Walk Off
[688]	Go Home with Person	No response/Walk Off	Get Phone Number
[691]	Go Home with Person	Get Phone Number	Go Home with Person
[694]	Get Phone Number	No response/Walk Off	No response/Walk Off
[697]	No response/Walk Off	No response/Walk Off	Get Phone Number
[700]	Get Phone Number	Go Home with Person	Get Phone Number
[703]	No response/Walk Off	No response/Walk Off	Go Home with Person
[706]	No response/Walk Off	Get Phone Number	Get Phone Number
[709]	No response/Walk Off	Get Phone Number	Go Home with Person
[712]	No response/Walk Off	No response/Walk Off	No response/Walk Off
[715]	No response/Walk Off	No response/Walk Off	Get Phone Number
[718]	No response/Walk Off	Go Home with Person	Get Phone Number
[721]	Get Phone Number	No response/Walk Off	No response/Walk Off
[724]	No response/Walk Off	Get Phone Number	Go Home with Person
[727]	Get Phone Number	Get Phone Number	No response/Walk Off
[730]	Get Phone Number	Get Phone Number	No response/Walk Off
[733]	Get Phone Number	Get Phone Number	No response/Walk Off
[736]	No response/Walk Off	Get Phone Number	No response/Walk Off
[739]	Go Home with Person	No response/Walk Off	No response/Walk Off
[742]	No response/Walk Off	Get Phone Number	No response/Walk Off
[745]	Get Phone Number	Get Phone Number	No response/Walk Off
[748]	Get Phone Number	No response/Walk Off	Get Phone Number
[751]	Get Phone Number	No response/Walk Off	No response/Walk Off
[754]	Get Phone Number	No response/Walk Off	No response/Walk Off
[757]	No response/Walk Off	Get Phone Number	No response/Walk Off
[760]	Get Phone Number	Get Phone Number	Get Phone Number
[763]	No response/Walk Off	No response/Walk Off	No response/Walk Off
[766]	Get Phone Number	No response/Walk Off	Get Phone Number
[769]	Get Phone Number	Get Phone Number	No response/Walk Off
[772]	Get Phone Number	Get Phone Number	Get Phone Number

[775]	No response/Walk Off	Get Phone Number	No response/Walk Off
[778]	Go Home with Person	No response/Walk Off	No response/Walk Off
[781]	Get Phone Number	No response/Walk Off	Get Phone Number
[784]	No response/Walk Off	No response/Walk Off	No response/Walk Off
[787]	No response/Walk Off	No response/Walk Off	Get Phone Number
[790]	No response/Walk Off	No response/Walk Off	Go Home with Person
[793]	Get Phone Number	Get Phone Number	No response/Walk Off
[796]	No response/Walk Off	No response/Walk Off	Go Home with Person
[799]	Go Home with Person	No response/Walk Off	No response/Walk Off
[802]	Get Phone Number	Go Home with Person	Get Phone Number
[805]	No response/Walk Off	No response/Walk Off	No response/Walk Off
[808]	No response/Walk Off	No response/Walk Off	No response/Walk Off
[811]	No response/Walk Off	Go Home with Person	Get Phone Number
[814]	Go Home with Person	Get Phone Number	No response/Walk Off
[817]	Get Phone Number	No response/Walk Off	Get Phone Number
[820]	Get Phone Number	Get Phone Number	No response/Walk Off
[823]	No response/Walk Off	No response/Walk Off	Go Home with Person
[826]	Get Phone Number	No response/Walk Off	Get Phone Number
[829]	No response/Walk Off	Get Phone Number	No response/Walk Off
[832]	Go Home with Person	No response/Walk Off	Get Phone Number
[835]	Go Home with Person	No response/Walk Off	No response/Walk Off
[838]	Get Phone Number	No response/Walk Off	No response/Walk Off
[841]	Get Phone Number	No response/Walk Off	No response/Walk Off
[844]	No response/Walk Off	Get Phone Number	Get Phone Number
[847]	No response/Walk Off	Go Home with Person	Get Phone Number
[850]	Get Phone Number	Go Home with Person	Get Phone Number
[853]	No response/Walk Off	No response/Walk Off	No response/Walk Off
[856]	No response/Walk Off	Get Phone Number	No response/Walk Off
[859]	No response/Walk Off	Get Phone Number	Get Phone Number
[862]	No response/Walk Off	No response/Walk Off	No response/Walk Off
[865]	Get Phone Number	Get Phone Number	Get Phone Number
[868]	Get Phone Number	No response/Walk Off	Get Phone Number
[871]	Get Phone Number	No response/Walk Off	No response/Walk Off
[874]	No response/Walk Off	Get Phone Number	Get Phone Number
[877]	No response/Walk Off	Get Phone Number	Get Phone Number
[880]	No response/Walk Off	No response/Walk Off	No response/Walk Off
[883]	No response/Walk Off	No response/Walk Off	No response/Walk Off
[886]	No response/Walk Off	No response/Walk Off	Get Phone Number
[889]	Get Phone Number	No response/Walk Off	No response/Walk Off
[892]	No response/Walk Off	No response/Walk Off	Go Home with Person
[895]	Get Phone Number	Get Phone Number	No response/Walk Off
[898]	No response/Walk Off	No response/Walk Off	Get Phone Number
[901]	Go Home with Person	No response/Walk Off	Get Phone Number

[904]	No response/Walk Off	Get Phone Number	Get Phone Number
[907]	No response/Walk Off	Get Phone Number	No response/Walk Off
[910]	Get Phone Number	No response/Walk Off	Get Phone Number
[913]	Get Phone Number	No response/Walk Off	Go Home with Person
[916]	No response/Walk Off	No response/Walk Off	Go Home with Person
[919]	Get Phone Number	No response/Walk Off	No response/Walk Off
[922]	Get Phone Number	No response/Walk Off	No response/Walk Off
[925]	Get Phone Number	No response/Walk Off	Get Phone Number
[928]	No response/Walk Off	No response/Walk Off	No response/Walk Off
[931]	No response/Walk Off	Get Phone Number	Get Phone Number
[934]	No response/Walk Off	Get Phone Number	Get Phone Number
[937]	No response/Walk Off	No response/Walk Off	No response/Walk Off
[940]	No response/Walk Off	Get Phone Number	No response/Walk Off
[943]	No response/Walk Off	Get Phone Number	No response/Walk Off
[946]	No response/Walk Off	No response/Walk Off	Get Phone Number
[949]	Get Phone Number	Get Phone Number	No response/Walk Off
[952]	No response/Walk Off	No response/Walk Off	Get Phone Number
[955]	No response/Walk Off	No response/Walk Off	Get Phone Number
[958]	No response/Walk Off	Get Phone Number	No response/Walk Off
[961]	Get Phone Number	Get Phone Number	No response/Walk Off
[964]	No response/Walk Off	Get Phone Number	Get Phone Number
[967]	Go Home with Person	Get Phone Number	Go Home with Person
[970]	No response/Walk Off	No response/Walk Off	Get Phone Number
[973]	Get Phone Number	No response/Walk Off	No response/Walk Off
[976]	Get Phone Number	Get Phone Number	Get Phone Number
[979]	Get Phone Number	Get Phone Number	Get Phone Number
[982]	No response/Walk Off	Get Phone Number	Get Phone Number
[985]	No response/Walk Off	Get Phone Number	Get Phone Number
[988]	Go Home with Person	Go Home with Person	No response/Walk Off
[991]	Get Phone Number	Get Phone Number	Get Phone Number
[994]	Get Phone Number	Get Phone Number	Get Phone Number
[997]	No response/Walk Off	No response/Walk Off	Get Phone Number
[1000]	Get Phone Number	Get Phone Number	No response/Walk Off
[1003]	Get Phone Number	No response/Walk Off	Get Phone Number
[1006]	Get Phone Number	No response/Walk Off	No response/Walk Off
[1009]	No response/Walk Off	Get Phone Number	Go Home with Person
[1012]	Get Phone Number	No response/Walk Off	Get Phone Number
[1015]	No response/Walk Off	No response/Walk Off	Get Phone Number
[1018]	Go Home with Person	Get Phone Number	Get Phone Number

Levels: Get Phone Number Go Home with Person No response/Walk Off

```
as.factor(chatData$Gender)
```

[illegible]


```

[831] Female Female Female Female Female Female Female Female Female Female
[841] Female Female Female Female Female Female Female Female Female Female
[851] Female Female Female Female Female Female Female Female Female Female
[861] Female Female Female Female Female Female Female Female Female Female
[871] Female Female Female Female Female Female Female Female Female Female
[881] Female Female Female Female Female Female Female Female Female Female
[891] Female Female Female Female Female Female Female Female Female Female
[901] Female Female Female Female Female Female Female Female Female Female
[911] Female Female Female Female Female Female Female Female Female Female
[921] Female Female Female Female Female Female Female Female Female Female
[931] Female Female Female Female Female Female Female Female Female Female
[941] Female Female Female Female Female Female Female Female Female Female
[951] Female Female Female Female Female Female Female Female Female Female
[961] Female Female Female Female Female Female Female Female Female Female
[971] Female Female Female Female Female Female Female Female Female Female
[981] Female Female Female Female Female Female Female Female Female Female
[991] Female Female Female Female Female Female Female Female Female Female
[1001] Female Female Female Female Female Female Female Female Female Female
[1011] Female Female Female Female Female Female Female Female Female Female
Levels: Female Male

```

The **Gender** is imported as a factor with ‘female’ as the baseline category. All of our predictions are based on females behaving differently than males, so it would be better to have ‘male’ as the baseline category by executing:

```

chatData$Gender <- factor(chatData$Gender, ordered = FALSE)
chatData$Success <- factor(chatData$Success, ordered = FALSE)
chatData$Gender <- relevel(chatData$Gender, ref = 2)
chatData$Success <- relevel(chatData$Success, ref = 3)
levels(chatData$Gender)

```

```

[1] "Male"    "Female"

```

```

levels(chatData$Success)

```

```

[1] "No response/Walk Off" "Get Phone Number"    "Go Home with Person"

```

This command resets the levels of the variable **Gender** such that the reference or baseline category is the category currently set as 2. Therefore, males become the reference category.

Instead of having one row per person, we need to have one row per person per category of the outcome variable. Each row will contain TRUE if the person was assigned to that category, and FALSE if they weren't. `mlogit.data()` function converts our data into the correct format:

```
mlChat <- mlogit.data(chatData, choice = "Success", shape = "wide")
head(mlChat)
```

~~~~~

first 10 observations out of 3060

~~~~~

	Success	Funny	Sex	Good_Mate	Gender	chid	alt	idx
1	TRUE	3	7	6	Male	1	Get Phone Number 1:mber	
2	FALSE	3	7	6	Male	1	Go Home with Person 1:rson	
3	FALSE	3	7	6	Male	1	No response/Walk Off 1: Off	
4	FALSE	5	7	2	Male	2	Get Phone Number 2:mber	
5	TRUE	5	7	2	Male	2	Go Home with Person 2:rson	
6	FALSE	5	7	2	Male	2	No response/Walk Off 2: Off	
7	TRUE	4	6	6	Male	3	Get Phone Number 3:mber	
8	FALSE	4	6	6	Male	3	Go Home with Person 3:rson	
9	FALSE	4	6	6	Male	3	No response/Walk Off 3: Off	
10	FALSE	3	7	5	Male	4	Get Phone Number 4:mber	

~~~ indexes ~~~~

|    | chid | alt                  |
|----|------|----------------------|
| 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     |

indexes: 1, 2

Now, we are ready to run the multinomial logistic regression, using the `mlogit()` function. The `mlogit()` function is similar to the `glm()` function, and takes the general form:

```
newModel <- mlogit(outcome ~ predictor(s), data = dataframe, na.action =
an action, reflevel = a number representing the baseline category for the
outcome)
```

In this example, `outcome` will be the variable **Success**. `reflevel` is a number representing the outcome category that we want to use as a baseline.

The basic idea of `mlogit()` is the same as the `lm()` and `glm()` functions, but one important difference is that we need to specify the reference or baseline category.

The next issue is what to include within the model. We are going to include the interaction terms **Sex x Gender** and **Funny x Gender** into the model. We can create the model by executing:

```
chatModel <- mlogit(
  Success ~ 1 | Good_Mate + Funny + Gender + Sex + Gender:Sex + Funny:Gender,
  data = mlChat,
)
```

### 8.9.2. Interpreting the multinomial logistic regression output

```
summary(chatModel)
```

Call:

```
mlogit(formula = Success ~ 1 | Good_Mate + Funny + Gender + Sex +
  Gender:Sex + Funny:Gender, data = mlChat, method = "nr")
```

Frequencies of alternatives:choice

|                  |                     |                      |
|------------------|---------------------|----------------------|
| Get Phone Number | Go Home with Person | No response/Walk Off |
| 0.47549          | 0.13235             | 0.39216              |

nr method

6 iterations, 0h:0m:0s

$g'(-H)^{-1}g = 0.00121$

successive function values within tolerance limits

Coefficients :

|                                  | Estimate   | Std. Error | z-value | Pr(> z )  |    |
|----------------------------------|------------|------------|---------|-----------|----|
| (Intercept):Go Home with Person  | -2.5032847 | 0.8300543  | -3.0158 | 0.0025630 | ** |
| (Intercept):No response/Walk Off | 1.7830697  | 0.6697724  | 2.6622  | 0.0077631 | ** |
| Good_Mate:Go Home with Person    | -0.0018205 | 0.0770265  | -0.0236 | 0.9811443 |    |
| Good_Mate:No response/Walk Off   | -0.1318400 | 0.0537261  | -2.4539 | 0.0141306 | *  |
| Funny:Go Home with Person        | 0.1790675  | 0.0897278  | 1.9957  | 0.0459693 | *  |
| Funny:No response/Walk Off       | -0.1393889 | 0.1101262  | -1.2657 | 0.2056135 |    |
| GenderFemale:Go Home with Person | -3.9801455 | 1.2106769  | -3.2875 | 0.0010107 | ** |

```

GenderFemale:No response/Walk Off  1.6462233  0.7962471  2.0675 0.0386891 *
Sex:Go Home with Person             0.1410768  0.1075727  1.3115 0.1897039
Sex:No response/Walk Off            -0.2762062  0.0891973 -3.0966 0.0019577 **
GenderFemale:Go Home with Person    -0.1283129  0.1496895 -0.8572 0.3913377
GenderFemale:No response/Walk Off   0.3483262  0.1058754  3.2900 0.0010020 **
Funny:Go Home with Person           0.6799621  0.1651761  4.1166 3.845e-05 ***
Funny:No response/Walk Off          -0.4924415  0.1399925 -3.5176 0.0004354 ***

```

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Log-Likelihood: -868.74

McFadden R<sup>2</sup>: 0.13816

Likelihood ratio test : chisq = 278.52 (p.value = < 2.22e-16)

From the output, we got a log-likelihood ratio of the overall model, which is a measure of how much unexplained variability there is in the data. Therefore, the difference or change in log-likelihood indicates how much new variance has been explained by the model. The chi-square test tests the decrease in unexplained variance from the baseline model to the final model, which is a difference of 139.26. By multiplying this by 2, we get the chi-square test, which is 278.52. This change is significant, which means that our final model explains a significant amount of the original variability. We also got a McFadden  $R^2$ , a measure of effect size.

For the interpretation, we can exponentiate the coefficients using the `exp()` function. These coefficients are stored in a variable called `coefficients` attached to the model:

```
exp(chatModel$coefficients)
```

```

(Intercept):Go Home with Person  (Intercept):No response/Walk Off
                                0.08181582                        5.94808709
Good_Mate:Go Home with Person     Good_Mate:No response/Walk Off
                                0.99818119                        0.87648126
Funny:Go Home with Person         Funny:No response/Walk Off
                                1.19610145                        0.86988969
GenderFemale:Go Home with Person  GenderFemale:No response/Walk Off
                                0.01868292                        5.18735182
Sex:Go Home with Person           Sex:No response/Walk Off
                                1.15151309                        0.75865651
GenderFemale:Go Home with Person  GenderFemale:No response/Walk Off
                                0.87957809                        1.41669436
Funny:Go Home with Person         Funny:No response/Walk Off
                                1.97380290                        0.61113251
attr(,"names.sup.coef")

```

```

character(0)
attr("fixed")
  (Intercept):Go Home with Person (Intercept):No response/Walk Off
              FALSE                FALSE
    Good_Mate:Go Home with Person   Good_Mate:No response/Walk Off
              FALSE                FALSE
      Funny:Go Home with Person     Funny:No response/Walk Off
              FALSE                FALSE
GenderFemale:Go Home with Person GenderFemale:No response/Walk Off
              FALSE                FALSE
      Sex:Go Home with Person       Sex:No response/Walk Off
              FALSE                FALSE
GenderFemale:Go Home with Person GenderFemale:No response/Walk Off
              FALSE                FALSE
      Funny:Go Home with Person     Funny:No response/Walk Off
              FALSE                FALSE
attr("sup")
character(0)

```

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

```

exp.chatModel.coefficients.
1          0.08181582
2          5.94808709
3          0.99818119
4          0.87648126
5          1.19610145
6          0.86988969
7          0.01868292
8          5.18735182
9          1.15151309
10         0.75865651
11         0.87957809
12         1.41669436
13         1.97380290
14         0.61113251

```

```
exp(confint(chatModel))
```

```
2.5 %      97.5 %
```

|                                   |             |            |
|-----------------------------------|-------------|------------|
| (Intercept):Go Home with Person   | 0.016080287 | 0.4162754  |
| (Intercept):No response/Walk Off  | 1.600538100 | 22.1049034 |
| Good_Mate:Go Home with Person     | 0.858310212 | 1.1608457  |
| Good_Mate:No response/Walk Off    | 0.788879872 | 0.9738104  |
| Funny:Go Home with Person         | 1.003209440 | 1.4260818  |
| Funny:No response/Walk Off        | 0.701010646 | 1.0794530  |
| GenderFemale:Go Home with Person  | 0.001741465 | 0.2004356  |
| GenderFemale:No response/Walk Off | 1.089378931 | 24.7008806 |
| Sex:Go Home with Person           | 0.932615946 | 1.4217883  |
| Sex:No response/Walk Off          | 0.636971899 | 0.9035873  |
| GenderFemale:Go Home with Person  | 0.001741465 | 0.2004356  |
| GenderFemale:No response/Walk Off | 1.089378931 | 24.7008806 |
| Funny:Go Home with Person         | 1.003209440 | 1.4260818  |
| Funny:No response/Walk Off        | 0.701010646 | 1.0794530  |

### 8.9.3. Reporting the results

We can report the results as with binary logistic regression using a table.