Interpreting Odd Ratios in Logistic Regression

OmaymaS

May 20, 2016

Contents

Introduction

Interpreting the logistic regression's coefficients is somehow tricky. Looking at some examples beside doing the math helps getting the concept of odds, odds ratios and consequently getting more familiar with the meaning of the regression coefficients. The following examples are mainly taken from IDRE UCLE FAQ Page and they are recreated with R.

Probability, Odds and Log of Odds

Let's say that the probability of success is $p = 0.8$, then the probability of failure is $1 - p = 0.2$. The odds of success is $\frac{p}{1-p} = \frac{0.8}{1-0.8} = 4$, i.e. the odds of success is 4 to 1 and the odds of failure is 0.25 to 1.

Note that:

- **Probability** ranges from 0 to 1

- **Odds** range from 0 to ∞

- **Log Odds** range from $-\infty$ to ∞

That is why the log odds are used to avoid modeling a variable with a restricted range such as probability.

Logistic Regression

Logistic regression models the the logit-transformed probability as a linear relationship with the predictor variables as follows:

$logit(p) = \frac{p}{1-p} = \beta_0 + \beta_1 x_1 + ... + \beta_x x_x$

or in terms of probabilities:

$p = \frac{exp(\beta_0 + \beta_1 x_1 + ... + \beta_x x_x)}{1 + exp(\beta_0 + \beta_1 x_1 + ... + \beta_x x_x)}$

Examples

The following examples use a dataset that contains 200 observations about students. The binary outcome variable we will use is **hon** which indicates if a student is an honor class or not.

Loading Libraries And Data

```r
library(dplyr)
library(knitr)
library(pander)
```

```r
hd<-read.csv("honordata.csv",sep=",")

head(hd)
```

```
##   female read write math hon femalexmath
## 1      0   57    52   41   0           0
## 2      1   68    59   53   0          53
## 3      0   44    33   54   0           0
## 4      0   63    44   47   0           0
## 5      0   47    52   57   0           0
## 6      0   44    52   51   0           0
```

Logistic Regression With No Predictor Variables

Here we will start with a simple model without any predictors:

$$logit(p) = \beta_0$$

Fitting The Model

```r
#fit the model without any predictor
f0<-glm(hon~1,data = hd,family = binomial)

summary(f0)$coeff
```

```
##               Estimate Std. Error   z value     Pr(>|z|)
## (Intercept) -1.12546    0.16441 -6.845446 7.623779e-12
```

As we can see:

- The intercept= **-1.12546** which corresponds to **the log odds of the probability of being in an honor class** p .

- We can go from the log odds to the odds by exponentiating the coefficient which gives us the odds **O=0.3245**.

- We can go backwards to the probability by calculating $p = \frac{O}{1+O} = \mathbf{0.245}$.

Calculating The Log Odds Manually

Let's do the math with the original data step by step to see the transformation from probablity to odds to log odds

```
honor<-hd %>%
        group_by(hon) %>%
        summarise(freq=n(),prob=(n()/nrow(.)),odds=prob/(1-prob), logodds=log(odds)) %>%
        round(.,5)

hprob<-honor$prob[honor$hon==1]
hodds<-honor$odds[honor$hon==1]
hlogodds<-honor$logodds[honor$hon==1]

pander(honor)
```

hon	freq	prob	odds	logodds
0	151	0.755	3.08163	1.12546
1	49	0.245	0.32450	-1.12546

We can see that:

- **The probability of being in an honor class $p = 0.245$**

- **The odds of the probability of being in an honor class $O = \frac{0.245}{0.755} =$ hodds**

- **The log odds of the probability of being in an honor class $log(O) =$ -1.12546** which is the intercept value we got from fitting the logistic regression model.

Logistic Regression With a Single Dichotomous Predictor Variable

Here we will use a binary predictor variable **female** in our model:

$$logit(p) = \beta_0 + \beta_1 * female$$

Fitting The Model

```
#
#fit the model with female as the predictor
f1<-glm(hon~female,data = hd,family = binomial)

summary(f1)$coeff
```

```
##               Estimate Std. Error   z value     Pr(>|z|)
## (Intercept) -1.4708517  0.2689554 -5.468756 4.532047e-08
## female       0.5927822  0.3414293  1.736178 8.253231e-02
```

We can see that:

- The intercept= **-1.47085** which corresponds to **the log odds for males being in an honor class** (since male is the reference group, female=0).

- The coefficient for female= **0.59278** which corresponds to **the log of odds ratio between the female group and male group**. The odds ratio equals **1.81** which means **the odds for females are about 81% higher than the odds for males**.

Calculating The Log Odds Manually

Here we can see the probabilites, odds and log odds for each case to compare with the logistic regression results

```
honor1<-hd %>%
        group_by(female,hon) %>%
        summarise(freq=n()) %>%
        mutate(all=sum(freq),prob=freq/all,odds=prob/(1-prob),logodds=log(odds)) %>%
        round(.,5)

pander(honor1)
```

female	hon	freq	all	prob	odds	logodds
0	0	74	91	0.81319	4.35294	1.47085
0	1	17	91	0.18681	0.22973	-1.47085
1	0	77	109	0.70642	2.40625	0.87807
1	1	32	109	0.29358	0.41558	-0.87807

Logistic Regression With a Single Continuous Predictor Variable

Here we will use a single continuous predictor variable **math** in our mode:

$$logit(p) = \beta_0 + \beta_1 * math$$

Fitting The Model

```
#fit the model with math as the predictor
f2<-glm(hon~math,data = hd,family = binomial)

summary(f2)$coeff
```

```
##               Estimate Std. Error   z value     Pr(>|z|)
## (Intercept) -9.7939421 1.48174484 -6.609736 3.850061e-11
## math         0.1563404 0.02560948  6.104784 1.029399e-09
```

In this case:

- The intercept= **-9.79394** which is interpreted as **the log odds of a student with a math score of zero being in an honors class**.

- The coefficient for math= **0.15634** which is interpreted as **the expected change in log odds for a one-unit increase in the math score**. The odds ratio can be calculated by exponentiating this value to get **1.16922** which means **we expect to see about 17% increase in the odds of being in an honors class, for a one-unit increase in math score**

We can also confirm this interpretation by looking at the predicted values using the estimated coefficients, i.e. the equation:

$logit(p) = \frac{p}{1-p} = -9.79394 + 0.15634 * math$

First we will add a column with the predicted values to our original data frame.

```
hd$predicted<-predict(f2)
```

Then we will examine the effect of a one-unit increase in math score by subtracting the corresponding log odds. For example, we will look at the math scores at 54 and 53 and calculate the difference in the estimated log odds. Then exponentiate it to get the odds ratio.

```
s1<-hd$predicted[hd$math==53][1]
s2<-hd$predicted[hd$math==54][1]

logodd_diff<-s2-s1
odd_ratio<-exp(logodd_diff)
```

We can see that:

- $logit(p)_{(math=54)} - logit(p)_{(math=53)} = 0.1563404$

- $Odds_{(math=54)}/Odds_{(math=53)} = 1.1692241$

Which goes with the interpretation mentioned earlier.