FAQ: How do I interpret odds ratios in logistic regression?

Introduction

When a binary outcome variable is modeled using logistic regression, it is assumed that the logit transformation of the outcome variable has a linear relationship with the predictor variables. This makes the interpretation of the regression coefficients somewhat tricky. In this page, we will walk through the concept of odds ratio and try to interpret the logistic regression results using the concept of odds ratio in a couple of examples.

From probability to odds to log of odds

Everything starts with the concept of probability. Let's say that the probability of success of some event is .8. Then the probability of failure is $1 - .8 = .2$. The odds of success are defined as the ratio of the probability of success over the probability of failure. In our example, the odds of success are $.8/.2 = 4$. That is to say that the odds of success are 4 to 1. If the probability of success is .5, i.e., 50-50 percent chance, then the odds of success is 1 to 1.
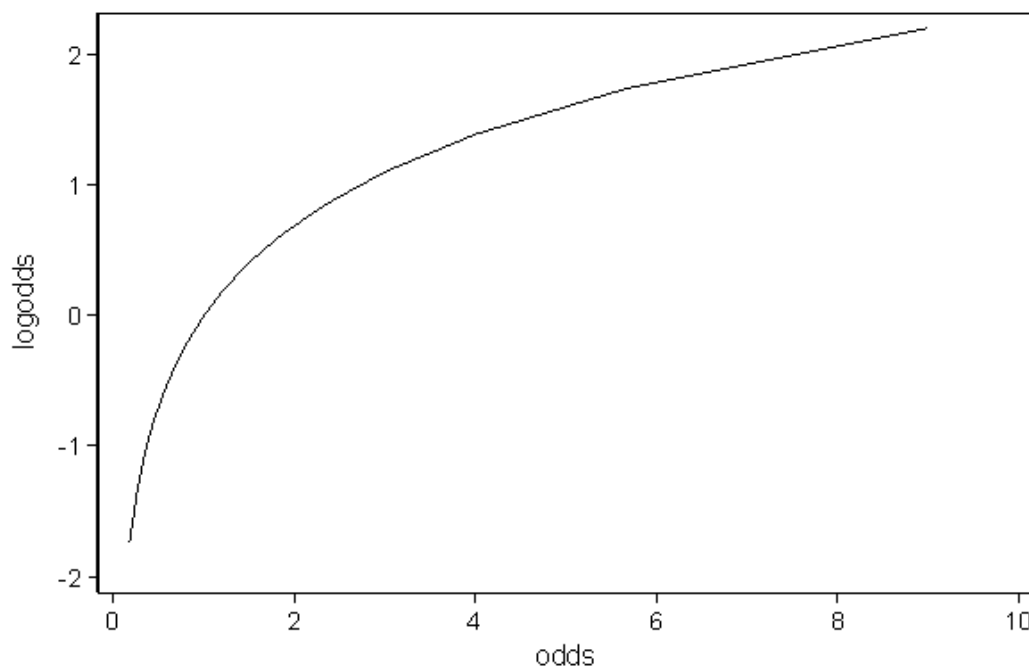
The transformation from probability to odds is a monotonic transformation, meaning the odds increase as the probability increases or vice versa. Probability ranges from 0 and 1. Odds range from 0 and positive infinity. Below is a table of the transformation from probability to odds and we have also plotted for the range of p less than or equal to .9.

```
  p      odds
 .001   .001001
 .01    .010101
 .15   .1764706
 .2       .25
 .25   .3333333
 .3    .4285714
 .35   .5384616
 .4    .6666667
 .45   .8181818
 .5       1
 .55   1.222222
 .6       1.5
 .65   1.857143
 .7    2.333333
 .75      3
 .8       4
 .85   5.666667
 .9       9
 .999     999
 .9999    9999
```
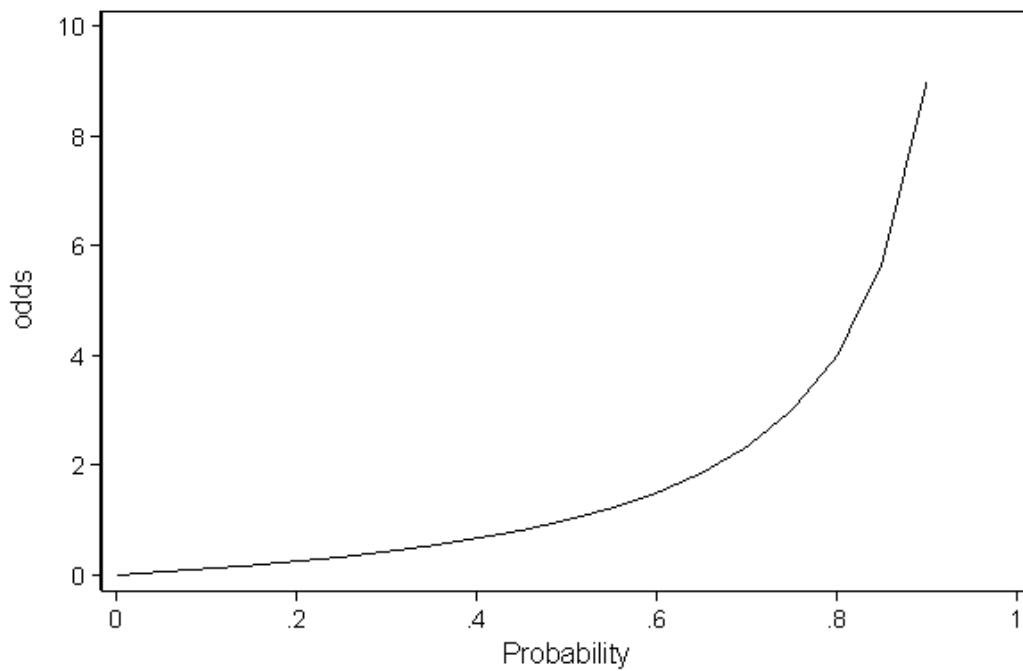
The transformation from odds to log of odds is the log transformation. Again this is a monotonic transformation. That is to say, the greater the odds, the greater the log of odds and vice versa. The table below shows the relationship among the probability, odds and log of odds. We have also shown the plot of log odds against odds.

```
  p       odds    logodds
 .001   .001001 -6.906755
 .01    .010101  -4.59512
 .15   .1764706 -1.734601
 .2        .25 -1.386294
 .25   .3333333 -1.098612
 .3    .4285714 -.8472978
 .35   .5384616 -.6190392
 .4    .6666667 -.4054651
 .45   .8181818 -.2006707
 .5        1        0
 .55   1.222222  .2006707
 .6        1.5  .4054651
 .65   1.857143  .6190392
 .7    2.333333  .8472978
 .75        3  1.098612
 .8         4  1.386294
 .85   5.666667  1.734601
 .9         9  2.197225
 .999      999  6.906755
 .9999     9999   9.21024
```

Why do we take all the trouble doing the transformation from probability to log odds? One reason is that it is usually difficult to model a variable which has restricted range, such as probability. This transformation is an attempt to get around the restricted range problem. It maps probability ranging between 0 and 1 to log odds ranging from negative infinity to positive infinity. Another reason is that among all of the infinitely many choices of transformation, the log of odds is one of the easiest to understand and interpret. This transformation is called logit transformation. The other common choice is the probit transformation, which will not be covered here.

A logistic regression model allows us to establish a relationship between a binary outcome variable and a group of predictor variables. It models the logit-transformed probability as a linear relationship with the predictor variables. More formally, let Y be the binary outcome variable indicating failure/success with $\{0,1\}$ and p be the probability of y to be 1, $p = P(Y=1)$. Let $x_1, \cdots, x_k$ be a set of predictor variables. Then the logistic regression of Y on $x_1, \cdots, x_k$ estimates parameter values for $\beta_0, \beta_1, \cdots, \beta_k$ via maximum likelihood method of the following equation

$$\text{logit}(p) = \log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k.$$

Exponentiate and take the multiplicative inverse of both sides,

$$\frac{1-p}{p} = \frac{1}{\exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k)}.$$

Partial out the fraction on the left-hand side of the equation and add one to both sides,

$$\frac{1}{p} = 1 + \frac{1}{\exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k)}.$$

Change 1 to a common denominator,

$$\frac{1}{p} = \frac{\exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k) + 1}{\exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k)}.$$

Finally, take the multiplicative inverse again to obtain the formula for the probability $P(Y=1)$,

$$p = \frac{\exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k)}{1 + \exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k)}.$$

We are now ready for a few examples of logistic regressions. We will use a sample dataset, https://stats.idre.ucla.edu/wp-content/uploads/2016/02/sample.csv, for the purpose of illustration. The data set has 200 observations and the outcome variable used will be **hon**, indicating if a student is in an honors class or not. So our p = prob(**hon**=1). We will purposely ignore all the significance tests and focus on the meaning of the regression coefficients. The output on this page was created using Stata with some editing.

Logistic regression with no predictor variables

Let's start with the simplest logistic regression, a model without any predictor variables. In an equation, we are modeling

$\text{logit}(p) = \beta_0$

```
Logistic regression                       Number of obs   =       200
                                          LR chi2(0)      =      0.00
                                          Prob > chi2     =         .
Log likelihood = -111.35502                   Pseudo R2       =    0.0000


-----------------------------------------------------------------------
      hon |    Coef.  Std. Err.    z    P>|z|   [95% Conf. Interval]
------------+----------------------------------------------------------
  intercept |  -1.12546   .1644101   -6.85   0.000   -1.447697   -.8032217
-----------------------------------------------------------------------
```

This means $\log(p/(1-p)) = -1.12546$. What is p here? It turns out that p is the overall probability of being in honors class (**hon** = 1). Let's take a look at the frequency table for **hon**.

```
     hon |   Freq.    Percent      Cum.
------------+-----------------------------------
       0 |    151     75.50      75.50
       1 |     49     24.50     100.00
------------+-----------------------------------
   Total |    200    100.00
```

So p = 49/200 = .245. The odds are .245/(1-.245) = .3245 and the log of the odds (logit) is log(.3245) = -1.12546. In other words, the intercept from the model with no predictor variables is the estimated log odds of being in honors class for the whole population of interest. We can also transform the log of the odds back to a probability: p = exp(-1.12546)/(1+exp(-1.12546)) = .245, if we like.

Logistic regression with a single dichotomous predictor variables

Now let's go one step further by adding a binary predictor variable, **female**, to the model. Writing it in an equation, the model describes the following linear relationship.

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

```
Logistic regression                          Number of obs  =       200
                                             LR chi2(1)    =      3.10
                                             Prob > chi2   =    0.0781
Log likelihood = -109.80312                  Pseudo R2     =    0.0139


-----------------------------------------------------------------------
     hon |    Coef.  Std. Err.    z   P>|z|    [95% Conf. Interval]
-------------+---------------------------------------------------------
  female |  .5927822  .3414294   1.74  0.083  -.0764072   1.261972
intercept | -1.470852  .2689555  -5.47  0.000  -1.997995  -.9437087
-----------------------------------------------------------------------
```

Before trying to interpret the two parameters estimated above, let's take a look at the crosstab of the variable **hon** with **female**.

```
        |     female
   hon |   male   female |   Total
-----------+--------------------+----------
     0 |    74      77 |    151
     1 |    17      32 |     49
-----------+--------------------+----------
  Total |    91     109 |    200
```

In our dataset, what are the odds of a male being in the honors class and what are the odds of a female being in the honors class? We can manually calculate these odds from the table: for males, the odds of being in the honors class are $(17/91)/(74/91) = 17/74 = .23$; and for females, the odds of being in the honors class are $(32/109)/(77/109) = 32/77 = .42$. The ratio of the odds for female to the odds for male is $(32/77)/(17/74) = (32*74)/(77*17) = 1.809$. So the odds for males are 17 to 74, the odds for females are 32 to 77, and the odds for female are about 81% higher than the odds for males.

Now we can relate the odds for males and females and the output from the logistic regression. The intercept of -1.471 is the log odds for males since male is the reference group (**female** = 0). Using the odds we calculated above for males, we can confirm this: $\log(.23) = -1.47$. The coefficient for **female** is the log of odds ratio between the female group and male group: $\log(1.809) = .593$. So we can get the odds ratio by exponentiating the coefficient for female. Most statistical packages display both the raw regression coefficients and the exponentiated coefficients for logistic regression models. The table below is created by Stata.

```
Logistic regression                          Number of obs  =       200
                                             LR chi2(1)    =      3.10
                                             Prob > chi2   =    0.0781
Log likelihood = -109.80312                  Pseudo R2     =    0.0139


-----------------------------------------------------------------------
     hon | Odds Ratio  Std. Err.    z   P>|z|    [95% Conf. Interval]
-------------+---------------------------------------------------------
  female |  1.809015  .6176508   1.74  0.083   .9264389   3.532379
-----------------------------------------------------------------------
```

Logistic regression with a single continuous predictor variable

Another simple example is a model with a single continuous predictor variable such as the model below. It describes the relationship between students' math scores and the log odds of being in an honors class.

```
logit(p) = β₀ + β₁*math

Logistic regression                        Number of obs   =      200
                                           LR chi2(1)      =    55.64
                                           Prob > chi2     =   0.0000
Log likelihood = -83.536619                Pseudo R2       =   0.2498


------------------------------------------------------------------------
     hon |    Coef.  Std. Err.    z   P>|z|   [95% Conf. Interval]
---------+--------------------------------------------------------------
    math | .1563404  .0256095   6.10  0.000   .1061467    .206534
intercept| -9.793942 1.481745  -6.61  0.000   -12.69811  -6.889775
------------------------------------------------------------------------
```

In this case, the estimated coefficient for the intercept is the log odds of a student with a math score of zero being in an honors class. In other words, the odds of being in an honors class when the math score is zero is exp(-9.793942) = .00005579. These odds are very low, but if we look at the distribution of the variable **math**, we will see that no one in the sample has math score lower than 30. In fact, all the test scores in the data set were standardized around mean of 50 and standard deviation of 10. So the intercept in this model corresponds to the log odds of being in an honors class when **math** is at the *hypothetical* value of zero.

How do we interpret the coefficient for math? The coefficient and intercept estimates give us the following equation:

$\log(p/(1\text{-}p)) = \text{logit}(p) = -9.793942 + .1563404*\textbf{math}$

Let's fix **math** at some value. We will use 54. Then the conditional logit of being in an honors class when the math score is held at 54 is

$\log(p/(1\text{-}p))(\textbf{math}=54) = -9.793942 + .1563404 *\textbf{54}.$

We can examine the effect of a one-unit increase in math score. When the math score is held at 55, the conditional logit of being in an honors class is

$\log(p/(1\text{-}p))(\textbf{math}=55) = -9.793942 + .1563404*\textbf{55}.$

Taking the difference of the two equations, we have the following:

$\log(p/(1\text{-}p))(\textbf{math}=55) - \log(p/(1\text{-}p))(\textbf{math} = 54) = .1563404.$

We can say now that the coefficient for **math** is the difference in the log odds. In other words, for a one-unit increase in the math score, the expected change in log odds is .1563404.

Can we translate this change in log odds to the change in odds? Indeed, we can. Recall that logarithm converts multiplication and division to addition and subtraction. Its inverse, the exponentiation converts addition and subtraction back to multiplication and division. If we exponentiate both sides of our last equation, we have the following:

$\exp[\log(p/(1\text{-}p))(\textbf{math}=55) - \log(p/(1\text{-}p))(\textbf{math} = 54)] = \exp(\log(p/(1\text{-}p))(\textbf{math}=55)) / \exp(\log(p/(1\text{-}p))$
$(\textbf{math} = 54)) = \text{odds}(\textbf{math}=55)/\text{odds}(\textbf{math}=54) = \exp(.1563404) = 1.1692241.$

So we can say for a one-unit increase in math score, we expect to see about 17% increase in the odds of being in an honors class. This 17% of increase does not depend on the value that math is held at.

Logistic regression with multiple predictor variables and no interaction terms

In general, we can have multiple predictor variables in a logistic regression model.

$$logit(p) = log(p/(1-p)) = \beta_0 + \beta_1 * x1 + \ldots + \beta_k * xk$$

Applying such a model to our example dataset, each estimated coefficient is the expected change in the log odds of being in an honors class for a unit increase in the corresponding predictor variable holding the other predictor variables constant at certain value. Each exponentiated coefficient is the ratio of two odds, or the change in odds in the multiplicative scale for a unit increase in the corresponding predictor variable holding other variables at certain value. Here is an example.

$$logit(p) = log(p/(1-p)) = \beta_0 + \beta_1 * \textbf{math} + \beta_2 * \textbf{female} + \beta_3 * \textbf{read}$$

```
Logistic regression                        Number of obs   =      200
                                  LR chi2(3)       =     66.54
                                  Prob > chi2    =    0.0000
Log likelihood = -78.084776              Pseudo R2      =    0.2988


-----------------------------------------------------------------------------
     hon |    Coef.   Std. Err.     z   P>|z|   [95% Conf. Interval]
-------------+---------------------------------------------------------------
    math |  .1229589  .0312756    3.93  0.000    .0616599   .1842578
  female |   .979948  .4216264    2.32  0.020    .1535755   1.80632
    read |  .0590632  .0265528    2.22  0.026    .0070207   .1111058
 intercept | -11.77025  1.710679   -6.88  0.000   -15.12311  -8.417376
-----------------------------------------------------------------------------
```

This fitted model says that, holding **math** and **reading** at a fixed value, the odds of getting into an honors class for females (**female** = 1) over the odds of getting into an honors class for males (**female** = 0) is exp(.979948) = 2.66. In terms of percent change, we can say that the odds for females are 166% higher than the odds for males. The coefficient for **math** says that, holding **female** and **reading** at a fixed value, we will see 13% increase in the odds of getting into an honors class for a one-unit increase in math score since exp(.1229589) = 1.13.

Logistic regression with an interaction term of two predictor variables

In all the previous examples, we have said that the regression coefficient of a variable corresponds to the change in log odds and its exponentiated form corresponds to the odds ratio. This is only true when our model does not have any interaction terms. When a model has interaction term(s) of two predictor variables, it attempts to describe how the effect of a predictor variable depends on the level/value of another predictor variable. The interpretation of the regression coefficients become more involved.

Let's take a simple example.

$$logit(p) = log(p/(1-p)) = \beta_0 + \beta_1 * \textbf{female} + \beta_2 * \textbf{math} + \beta_3 * \textbf{female*math}$$

```
Logistic regression                      Number of obs   =        200
                                         LR chi2(3)     =      62.94
                                         Prob > chi2    =     0.0000
Log likelihood = -79.883301                 Pseudo R2      =     0.2826


------------------------------------------------------------------------
      hon |    Coef.   Std. Err.    z    P>|z|    [95% Conf. Interval]
-------------+----------------------------------------------------------
    female | -2.899863  3.094186  -0.94  0.349   -8.964357   3.164631
      math |  .1293781  .0358834   3.61  0.000    .0590479   .1997082
femalexmath |  .0669951    .05346   1.25  0.210   -.0377846   .1717749
  intercept | -8.745841   2.12913  -4.11  0.000  -12.91886  -4.572823
------------------------------------------------------------------------
```

In the presence of interaction term of **female** by **math**, we can no longer talk about the effect of **female**, holding all other variables at certain value, since it does not make sense to fix **math** and **femalexmath** at certain value and still allow female change from 0 to 1!

In this simple example where we examine the interaction of a binary variable and a continuous variable, we can think that we actually have two equations: one for males and one for females. For males (**female**=0), the equation is simply

$$\text{logit}(p) = \log(p/(1-p)) = \beta_0 + \beta_2 * \textbf{math.}$$

For females, the equation is

$$\text{logit}(p) = \log(p/(1-p)) = (\beta_0 + \beta_1) + (\beta_2 + \beta_3) * \textbf{math.}$$

Now we can map the logistic regression output to these two equations. So we can say that the coefficient for math is the effect of math when **female** = 0. More explicitly, we can say that for male students, a one-unit increase in math score yields a change in log odds of 0.13. On the other hand, for the female students, a one-unit increase in math score yields a change in log odds of (.13 + .067) = 0.197. In terms of odds ratios, we can say that for male students, the odds ratio is exp(.13) = 1.14 for a one-unit increase in math score and the odds ratio for female students is exp(.197) = 1.22 for a one-unit increase in math score. The ratio of these two odds ratios (female over male) turns out to be the exponentiated coefficient for the interaction term of **female** by **math**: 1.22/1.14 = exp(.067) = 1.07.