TRENDS
Technology To Transcend

HOW TO INTERPRET ODDS RATIOS IN LOGISTIC REGRESSION

April 26, 2018

LOGISTIC REGRESSION FOR MACHINE LEARNING

TRENDS

- Logistic regression is another technique borrowed by machine learning from the field of statistics.
- It is the go-to method for binary classification problems (problems with two class values).
- Used for predictive modeling in applied machine learning.
- Formally, it is a statistical technique capable of predicting a binary outcome.
- It's a well-known strategy, widely used in disciplines ranging from credit and finance to medicine to criminology and other social sciences.
- It is fairly intuitive and very effective.

ODDS RATIO

- Measures association between an exposure and an outcome
- 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.

PROBABILITY

Let’s say that the probability of success is .8, thus

p = .8

Then the probability of failure is

q = 1 – p = .2

ODDS

Odds are determined from probabilities and range between 0 and infinity.

Odds are defined as the ratio of the probability of success and the probability of failure. The odds of success are

odds(success) = p/(1-p) or p/q = .8/.2 = 4

that is, the odds of success are 4 to 1.

The odds of failure would be

odds(failure) = q/p = .2/.8 = .25

Or, that the odds of failure are 1 to 4. The odds of success and the odds of failure are just reciprocals of one another, i.e., 1/4 = .25 and 1/.25 = 4.

Next, we will add another variable to the equation so that we can compute an odds ratio.

ODDS RATIO

Using data from Pedhazur (1997), a great example in estimating STEM participation for women.

Suppose that seven out of 10 males are admitted to an engineering school while three of 10 females are admitted.

The probabilities for admitting a male are?

p = 7/10 = .7

q = 1 – .7 = .3

If you are male, the probability of being admitted is 0.7 and the probability of not being admitted is 0.3.

ODDS RATIO

Here are the same probabilities for females

p = 3/10 = .3

q = 1 – .3 = .7

If you are female it is just the opposite, the probability of being admitted is 0.3 and the probability of not being admitted is 0.7.

ODDS RATIO

We can now use the probabilities to compute the odds of admission for both males and females

odds(male) = .7/.3 = 2.33333

odds(female) = .3/.7 = .42857

Next, we compute the odds ratio for admission

OR = 2.3333/.42857 = 5.44

Thus, for a male, the odds of being admitted are 5.44 times larger than the odds for a female being admitted.

ASIDE: ABOUT LOGITS

There is a direct relationship between the coefficients produced by **logit** and the odds ratios produced by **logistic**.

A logit is defined as the log base e (log) of the odds:

logit(p) = log(odds) = log(p/q)

The range is negative infinity to positive infinity. In regression it is easiest to model unbounded outcomes.

Logistic regression is in reality an ordinary regression using the logit as the response variable. The logit transformation allows for a linear relationship between the response variable and the coefficients:

logit(p) = a + bX

log(p/q) = a + bX

TRENDS
Technology To Transcend

Thank you!

As of April 18, 2018