Logistic Regression

Understanding the likelihood of having or experiencing a particular condition or event

Logistic regressions allow the use of dummy variables as the dependent variable. Because it is using dummies, it is a non-parametric test and not as powerful as an OLS regression. It is harder for it to find relationships.

Logistic regressions determine the likelihood that someone will have the particular condition associated with the dependent variable. For instance, it determines the likelihood that someone will complete treatment or call in sick. It does this by creating the best set of b coefficients (converted into odds ratios) to represent the highest likelihood associated with the dependent variable.

R^2 and Model Significance

Logistic regressions provide two pseudo R^2 values – the Cox and Snell and Nagelkerke. One typically goes with whichever one is larger. Logistic regression use the -2LogL and chi-square to determine significance. The R^2 and significance are read the same way as in linear regressions.

Odds Ratios (Exp (B) or e^b)

The odds ratio (or e^b) expresses the likelihood that someone with a particular characteristic will have the dependent variable condition or not. That is, they determine the odds ratio associated with a unit change in X (for I/R variables) or presence of characteristics (for 0/1 variables) controlling for all other variables in the model.

You read the odds ratio by stating how much more or less likely someone is to be in the dependent variable condition. The odds ratio, as provided in output, refers to the number of times <u>as likely</u> someone is to be in the condition compared to others. You convert it to be able to say whether someone is more or less likely.

For I/R variables, the odds ratio is comparison between those with a value 1 to those with a value of 0. For dummy variables, the odds ratio is a comparison between those who have a value of 1 (or are in the condition) and those who have a value of 0 (or not in the condition).

If the odds ratios are not provided, one can take the exponential of the b coefficient to determine it by hand.

$e^b = 1$

There is equal likelihood of being in the condition regardless of the value associated with an I/R variable or the group in which someone is for a dummy variable.

$e^b > 1$

Someone is more likely to be in the condition than someone else.

To determine how much more likely, subtract 1 from the odds ratio. That is, you are getting rid of the equal likelihood. If you do not subtract 1, then you read the odds ratio as "as likely." You can convert the number to a percentage by multiplying it by 100.

Example: $e^b = 9.34$

I/R
Someone with 1 child is predicted to be 9.34 times (or 934%) as likely to be in poverty compared to those with no kids.

Someone with 1 child is predicted to be 8.34 times (834%) more likely to be in poverty than those with no kids. OR, for each additional child in the family, the family is predicted to be 834% more likely to be in poverty.

Dummy
Someone with a drug problem is predicted to be 9.34 times (or 934%) as likely to be in prison compared to someone with no drug problem.

Someone with a drug problem is predicted to be 8.34 times (834%) more likely to be in prison than someone with no drug problem. OR, the odds of someone with a drug problem being in prison are predicted to be 834% higher than someone who does not have a drug problem.

$e^b < 1$

Someone is less likely to be in the condition than someone else.

To determine how much less likely, subtract e^b from 1. That is, you finding out how much below equal likelihood. If you do not subtract from 1, then you read the e^b as "as likely." You can convert this number to a percentage by multiplying by 100.

Example: $e^b = .85$

I/R
Someone with 1 child is predicted to be .85 times (or 85%) as likely to be in poverty compared to those with no kids.

Someone with 1 child is .15 times (or 15%) less likely to be in poverty than those with no kids. Or, for each additional child in the family, the family is 15% less likely to be in poverty.

Dummy
Someone with a high school degree is .85 times (or 85%) as likely to have poor health as someone without a high school degree.

Someone with a high school degree is .15 times (or 15%) less likely to have poor health compared to someone without a high school degree. Or, the odds of someone with a high school degree having poor health are predicted to 15% lower than those who do not have high school degrees.

Interpreting the b and making predictions

The b coefficients make it possible to make predictions about the likelihood that individuals with particular characteristics will have the DV condition.

Determining the odds ratio of one individual relative to another with a different X value

The formula for determining this odds ratio is:

$$\frac{e^{bx}}{e^{bx}}$$

Where you multiply the coefficient by the value of each X and take the exponential for each individual. You then divide the numerator by the denominator. You read the odds ratio as you would the odds ratio provided in the table, comparing the numerator individual with the denominator individual.

Example:

You want to determine the likelihood that child will complete a support group for children with incarcerated parents when her mother has been in prison for 5 years, compared to a child whose mother has been in prison for 10 years.

b= .765

$$\frac{e^{bx}}{e^{bx}} = \frac{e^{.764(5)}}{e^{.764(10)}} = \frac{e^{3.82}}{e^{7.64}} = \frac{45.6}{2079.74} = .02$$

1 - .02 = .98 = 98%

A child whose mother has been in prison 5 years is 98% less likely to complete the support group, compared to someone whose mother has been in prison for 10 years.

NOTE: If you put a 1 for the X in the numerator and a 0 in the X for the denominator, you get the odds ratio, or e^b.

Probability estimates

Probabilities are the likelihood that an individual with a particular set of characteristics will be in the condition of the DV variable. It is not relative to another group. This probability value has to be between 0 and 1.

The formula for determining the probabilities:

$$\frac{e^{a + b1x1 + b2x2 + b3x3 + bkxk}}{1 + e^{a + b1x1 + b2x2 + b3x3 + bkxk}}$$

Where you first calculate the regression equation using the a and b coefficients (not the odds ratios) provided in the table. You then take the exponential of the answer and divide it by 1 plus the exponential of the answer of regression equation. You multiply the answer by 100 to determine the percent likelihood.

Examples of interpreting the results:

- The likelihood of an individual with these characteristics [being in the condition] is predicted to be [probability in percentage form].

OR

- It is predicted that there is a [probability in percentage form] likelihood that an individual with these characteristics will [be in the condition].

Example:

The likelihood of someone with 5 kids, employed and with a college degree feeling stress is predicted to be 50%.

It is predicted that there is a 135% chance that a 16 year old child living in a dangerous neighborhood and scoring 10 on a behavioral problem scale will have witnessed violence.