Contents

1 Multiple Logistic Regression

So far we have looked at the case where our dependent variable is binary, i.e. it has just two categories. However, as you know there are many nominal variables with more than two categories. If we look at the example we gave earlier about staying on in education post-16, you can imagine that we might want a little more detail, such as knowing whether students have stayed in mainstream education, are in vocational training, or have left education altogether. This would be a three category nominal variable.

We therefore need methods that can model this type of dependent variable as well. Luckily we can extend the logistic regression model to do this. We call this extended method multinomial logistic regression, and refer to logistic regression for dichotomous dependent variables as binary logistic regression.

The basic principle of multinomial logistic regression is similar to that for binomial logistic regression, in that it is based on the probability of membership of each category of the dependent variable. So, in our three category example, the focus is on what the probability is of being in education, in training or not in education.

The way multinomial logistic regression deals with the variables in this case is somewhat similar to the concept of dummy variables we have seen earlier, in that it compares the probability of being in each of n-1 categories compared to a baseline or reference category. In a way we can say that we are fitting n-1 separate binary logistic models, where we compare category 1 to the baseline category, then category 2 to the baseline and so on. In practice software algorithms allow us to model the comparisons to the baseline simultaneously using maximum likelihood estimation, which is better as doing it sequentially could lead to misestimation of the standard errors. Therefore, multinomial logistic regression is basically an extension of binary logistic regression for nominal variables with more than two categories.