Generalized Logistic Regression

1 Multinomial Logistic Regression

Logistic regression can easily be extended to outcomes with multiple categories. Initially consider an outcome Y with values 0, 1, ..., J. We will consider Y=0 our referent or non-case group, but beyond that do not need to make any assumptions about order of severity for the remaining outcome categories. For simplicity most notation and examples will assume that J=2 or a total of 3 possible outcomes. The model is based on the generalized logit function:

$$g_1(x) = log\left[\frac{p(y=1|x)}{p(y=0|x)}\right] = \beta_{01} + \beta_{11}X \tag{1}$$

$$g_2(x) = log\left[\frac{p(y=2|x)}{p(y=0|x)}\right] = \beta_{02} + \beta_{12}X \tag{2}$$

The conditional probabilities for each outcome category are:

$$p(y=0|x) = \frac{1}{1 + exp(g_1(x)) + exp(g_2(x))} \tag{3}$$

$$p(y=0|x) = \frac{exp(g_1(x))}{1 + exp(g_1(x)) + exp(g_2(x))} \tag{4}$$

$$p(y=0|x) = \frac{exp(g_2(x))}{1 + exp(g_1(x)) + exp(g_2(x))} \tag{5}$$

The coefficients are interpreted as log odds ratios, just as in the binary logistic model; hypothesis tests and confidence intervals are constructed similarly. In some circumstances, it may further support the analysis to compare the magnitude of the two (J-1) estimated odds ratios. For example to test that the coefficients (i.e., log odds ratios) for a variable X are the same for outcomes 1 and 2, we would test the hypothesis 11 = 21, or equivalently, 11 - 21 = 0. The point estimate is the difference in the estimated values, the variance can be calculated as:

$$Var(\hat{\beta_1}1 - \hat{\beta_1}2) = Var(\hat{\beta_1}1) + Var(\hat{\beta_1}2) - 2Cov(\hat{\beta_1}1, \hat{\beta_1}2) \tag{6}$$

Lets fit this model with an nominal three level outcome, in PROC LOGISTIC with the link=glogit option:

```
proc logistic data=meexp descending order=data;
model me=hist/link=glogit ;
run;
```

2 Ordinal Regression

When the scale of a multiple category outcome is ordinal one could use the multinomial logistic model described above. This analysis would not take into account the ordinal nature of the outcome and hence the estimated odds ratios may not address the questions asked of the analysis. We will now discuss a number of different logistic regression models that take into account the ordering of the outcomes.

Recall the multinomial model for the i^{th} group:

$$g_i(x) = log\Big[\frac{p(y=i|x)}{p(y=0|x)}\Big] = \beta_{0i} + \beta_{1i}X \tag{7}$$

There are other ways to model this. For example the Adjacent-Category Logistic model:

$$g_i(x) = log\Big[\frac{p(y=i|x)}{p(y=i-1|x)}\Big] = \beta_0 + \beta_1 X \tag{8}$$

The Continuation-ratio logistic model:

$$g_i(x) = log\Big[\frac{p(y=i|x)}{p(y<i|x)}\Big] = \beta_{0i} + \beta_{1i}X \tag{9}$$

And the Proportional Odds Model:

$$g_i(x) = log\Big[\frac{p(y\leq i|x)}{p(y>i|x)}\Big] = \beta_{0i} + \beta_{1i}X \tag{10}$$

Right now we'll only discuss the Proportional Odds and the Multinomial models. Probably the most frequently used ordinal logistic regression model in practice is the proportional odds model. The other models mentioned compare a single outcome response to one or more reference responses. The proportional odds model describes a less than or equal versus more comparison. For example if the outcome is extent of disease the model gives the log odds of no more severe outcome versus a more severe outcome. The constraint placed on the model is that the log odds does not depend on the outcome category. Thus inferences from the models lend themselves to a general discussion of direction of response and do not have to focus on specific outcome categories.

Lets fit this in SAS:

```
proc logistic data=temp descending;
model bwt4=smoke/link=glogit;
run;


proc logistic data=temp descending;
model bwt4=smoke;
output out=da pred=p predprobs=i c;
run;
```