The Multinomial Model

Christopher Weber

Invalid Date

The Nominal Model

- These notes follow Long (1997), Chapter 6
- Often, dependent variables don't have a natural ordering
- If we have multi-category nominal data, we will again violate the assumptions of the classical linear regression model
- In the case of nominal data, we again can use the intuition of logit and probit with binary variables
- The *Multinomial Logit* and *Multinomial Probit*

An Example

- Voting (1=Democrat; 2=Republican; 3=Libertarian)
- Run one logit model predicting the probability of Democrat relative to Republican voting
- Run a second model predicting Democrat versus Libertarian
- Run a third model predicting Republican versus Libertarian

Intuition

- Long (1997), Chapter 6
- Assume $y_{obs} \in (R, D, L)$.

$$ln(\frac{pr(D|x)}{pr(R|x)} = \beta_{0,D|R} + \beta_{1,D|R}x$$

$$ln(\frac{pr(D|x)}{pr(L|x)} = \beta_{0,D|L} + \beta_{1,D|L}x$$

$$ln(\frac{pr(R|x)}{pr(L|x)} = \beta_{0,R|L} + \beta_{1,R|L}x$$

Intuition

$$\frac{pr(D|x)}{pr(R|x)} = exp(\beta_{0,D|R} + \beta_{1,D|R}x)$$

$$\frac{pr(D|x)}{pr(L|x)} = exp(\beta_{0,D|L} + \beta_{1,D|L}x)$$

$$\frac{pr(R|x)}{pr(L|x)} = exp(\beta_{0,R|L} + \beta_{1,R|L}x)$$

Intuition

- However, the sum of the first two equations equals the third equation. We need not estimate each model; it's redundant (and not identified)
- Calculate the probability of being in the kth category

$$pr(y = K|x) = \frac{exp(X\beta_k)}{\sum_k exp(X\beta_k)}$$

- Multiply the above expression by τ, $exp(x\tau)/exp(x\tau)$
- The probabilities will stay the same, but $\beta = \beta + \tau$

Instead

$$y_{obs} = \begin{cases} D, 1/(1 + \sum_{k=2}^{K} exp(XB_k)) \\ R, exp(XB_R)/(1 + \sum_{k=2}^{K} exp(XB_k)) \\ L, exp(XB_L)/(1 + \sum_{k=2}^{K} exp(XB_k)) \end{cases}$$

- We estimate $k-1$ unique equations, where one category serves as the baseline, reference category

The Likelihood

- The probability of being in the kth category for the ith subject is,

$$\frac{pr(y_i = K|x_i) = exp(XB)}{\sum exp(XB)}$$

- Calculate the joint parameter space,
$pr(y_i = 1|X_i) \times pr(y_i = 2|X_i) \times pr(y_i = 3|X_i) \times ....pr(y_i = K|X_i)$

- This is just the joint probability for category membership, for each subject, so

The Likelihood

$$\frac{pr(y_i|X_i) = \prod_{k=1}^{K} exp(XB)}{\sum exp(XB)}$$

$$pr(y|X) = \prod_{i=1}^{N} \prod_{k=1}^{K} \frac{exp(XB)}{\sum exp(XB)}$$

The Log Likelihood

$$Loglik(\beta|y,X) = \sum_{i=1}^{N}\sum_{k=1}^{K} log[\frac{exp(XB)}{\sum exp(XB)}]$$

Interpretation

- With k categories, there are $k-1$ unique equations in the multinomial logit model. In other words, if we include 2 covariates and there are 3 categories, we would estimate six parameters
- The partial derivative is different at levels of x

$$\frac{\partial pr(y=k|x)}{\partial x} = \sum_{j=1}^{J} \beta_{j,m} pr(y=k|x)$$

Interpretation

- The key to understand here is that one category serves as the baseline and we interpret the results of the $k-1$ categories

$$H_0 = \beta_{k,1|r} = \beta_{k,2|r} = ...\beta_{k,J|r}$$

Interpretation

- Likewise, we may also test the probability of being in the kth category, given a particular value of x.

$$pr(y = k|x) = exp(xB_k) / \sum_{j=1}^{J} exp(xB_k)$$

Independence of Irrelevant Alternatives

- The multinomial models make a relatively strong assumption about the choice process
- It is called the **Independence of Irrelevant Alternatives** (IIA) assumption
- The probability of odds contrasting two choices are unaffected by additional alternatives
- McFadden (cited on Long 1997, p. 182) introduces the now classic **Red Bus/Blue Bus** example

Transportation

- The logic.....
- Say there are two forms of transportation available in a city: The city bus and driving one's car.
- If an individual is indifferent to these approaches, taking advantage of both about equally, assume that $p(car) = 0.5$ and $p(bus) = 0.5$,
- The odds of taking the bus relative to the car is 1:1. The buses in the city are all red

Irrelevance?

- The city introduces a bus on this individual's route
- The only difference is that the bus is blue
- Because the blue bus is identical (with the exception of the color), the individual probably doesn't prefer it over the red bus
- The only way that IIA holds is if the probability of $p(car) = 0.33, p(Red) = 0.33, p(Blue) = 0.33$

Irrelevance?

- This doesn't make much sense; it implies that the individual will ride the bus over driving – the probability of taking *a bus* is $2/3$
- Logically, what we should observe is that $p(drive) = 0.5, p(red) = 0.25, p(blue) = 0.25$. This involves a violation of IIA
- The only way for IIA to hold is if the associated probabilities change and $p(car) = p(red)$
- But we are unlikely to observe this if we logically think about the problem

Tests

- The odds of selecting the red bus, relative to the car should be the same regardless of whether blue buses are available
- We need to make the IIA in both the multinomial and conditional logit models
- Voting (Bush and Clinton 1992)
- The assumption holds that the odds (i.e., the coefficients) should be the same in both models. This can be tested by using a “Hausman test”

The Hausman Test

- Conceptually, the test involves comparing the full multinomial model to one where outcome categories are dropped from the analysis
- The test is distributed χ^2 and relies on the change in coefficients weighted by the inverse of the variance-covariance matrix of the full and restricted multinomial models
- See Long (1997, p 184) for the exact calculation. This is often called a Hausman test, or a Hausman-McFadden test of IIA