BIST P8110: Applied Regression II

16. Multinomial Logistic Regression

Qixuan Chen

Department of Biostatistics
Columbia University

Multinomial Logistic regression

- Multinomial logistic regression is also called generalized logit models.
- It generalizes logistic regression by allowing nominal outcome variables with more than two categories.
- It assumes that the outcome variable Y follows a multinomial distribution.
- It does not assume the ordering in different categories of Y.

Model

- Suppose that Y has K categories, the multinomial logit model is given by

$$log\left\{\frac{Pr(Y=k|X)}{Pr(Y=1|X)}\right\} = \alpha_k + \beta_{k1}X_1 + \cdots + \beta_{kp}X_p,\ k = 2, \ldots, K \quad (1)$$

where, the $K-1$ odds differ in not only the intercept α_k but also the slopes β_k.

Probability for Each Category

- From Model (1), we have

$$
\begin{aligned}
Pr(Y = 1|X) &= \frac{1}{1 + \sum_{k=2}^{K} e^{\alpha_k + \beta_{k1} X_1 + \cdots + \beta_{kp} X_p}} \\
Pr(Y = 2|X) &= \frac{e^{\alpha_2 + \beta_{21} X_1 + \cdots + \beta_{2p} X_p}}{1 + \sum_{k=2}^{K} e^{\alpha_k + \beta_{k1} X_1 + \cdots + \beta_{kp} X_p}} \\
&\vdots \\
Pr(Y = K|X) &= \frac{e^{\alpha_K + \beta_{K1} X_1 + \cdots + \beta_{Kp} X_p}}{1 + \sum_{k=2}^{K} e^{\alpha_k + \beta_{k1} X_1 + \cdots + \beta_{kp} X_p}}
\end{aligned}
$$

and we have $\sum_{k=1}^{K} Pr(Y = k|X) = 1$.

Model with $K = 3$

Model with $K = 4$

- For a four-level categorical outcome variable, three equations will be estimated. The equations are:

	Category	vs	Category
Equation 1:	2		1
Equation 2:	3		1
Equation 3:	4		1

  - Each equation models the odds of being in the category on the left versus the category on the right.
  - Three sets of coefficients for each X.

Interpretation of Parameters

- Both the intercepts and slopes can be interpreted similarly to dichotomous logistic regression
  - The α_k in Model (1) can be interpreted as the log odds of Y belonging to category k compared to the reference category when all $X = 0$.
  - A positive β_k coefficient indicates an increased chance that a subject with a higher score on X will be observed in category k than the reference category.

Case Study

- Over the course of one school year, third graders from three different schools exposed to three different styles of mathematics instruction are asked which style they prefer (Stokes, Davis, and Koch; 2000).
  - The variable style specifies three different styles of mathematics instruction (a self-paced computer-learning style, a team approach, and a traditional class approach).
  - The variable school specifies three different schools (1,2,3).
  - The variable program specifies two types of programs they are in (a regular school day versus a regular day supplemented with an afternoon school program).

School Program Data

Obs	school	program	style	count
1	1	regular	self	10
2	1	regular	team	17
3	1	regular	class	26
4	1	afternoo	self	5
5	1	afternoo	team	12
6	1	afternoo	class	50
7	2	regular	self	21
8	2	regular	team	17
9	2	regular	class	26
10	2	afternoo	self	16
11	2	afternoo	team	12
12	2	afternoo	class	36
13	3	regular	self	15
14	3	regular	team	15
15	3	regular	class	16
16	3	afternoo	self	12
17	3	afternoo	team	12
18	3	afternoo	class	20

Generalized Logit Model for Style Preference

- The generalized logit model is

$$log\left\{\frac{Pr(Y=\text{self})}{Pr(Y=\text{class})}\right\} = \beta_{0s} + \beta_{1s}X_1 + \beta_{2s}X_2 + \beta_{3s}X_3$$
$$log\left\{\frac{Pr(Y=\text{team})}{Pr(Y=\text{class})}\right\} = \beta_{0t} + \beta_{1t}X_1 + \beta_{2t}X_2 + \beta_{3t}X_3$$

where

$$X_1 = \begin{cases} 1 & \text{School} = 2 \\ 0 & \text{otherwise} \end{cases}$$

$$X_2 = \begin{cases} 1 & \text{School} = 3 \\ 0 & \text{otherwise} \end{cases}$$

$$X_3 = \begin{cases} 1 & \text{Program} = \text{afternoon} \\ 0 & \text{otherwise} \end{cases}$$

SAS Code

The SAS code is:

```
data school;
infile "C:\school.csv" delimiter = ',' MISSOVER DSD;
input school program $ style $ count;
run;

proc logistic data=school;
   freq count;
   class school(ref='1') program (ref='regular')/param=ref;
   model style(ref='class')=school program / link=glogit;
run;
```

The LOGISTIC Procedure Options

- The GLOGIT option forms the generalized logits.
- The REF option in the MODEL statement specifies the reference category for the generalized logit model.

SAS Output

Model Information

Data Set	WORK.SCHOOL
Response Variable	style
Number of Response Levels	3
Frequency Variable	count
Model	generalized logit
Optimization Technique	Newton-Raphson

Number of Observations Read	18
Number of Observations Used	18
Sum of Frequencies Read	338
Sum of Frequencies Used	338

Response Profile

Ordered Value	style	Total Frequency
1	class	174
2	self	79
3	team	85

Logits modeled use style='class' as the reference category.

SAS Output

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	699.404	682.934
SC	707.050	713.518
-2 Log L	695.404	666.934

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	28.4704	6	<.0001
Score	27.1190	6	0.0001
Wald	25.5881	6	0.0003

Type 3 Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
school	4	14.8424	0.0050
program	2	10.9160	0.0043

SAS Output

Analysis of Maximum Likelihood Estimates

Parameter		style	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		self	1	-1.2233	0.3154	15.0454	0.0001
Intercept		team	1	-0.5662	0.2586	4.7919	0.0286
school	2	self	1	1.0828	0.3539	9.3598	0.0022
school	2	team	1	0.1801	0.3172	0.3224	0.5702
school	3	self	1	1.3147	0.3839	11.7262	0.0006
school	3	team	1	0.6556	0.3395	3.7296	0.0535
program	afternoo	self	1	-0.7474	0.2820	7.0272	0.0080
program	afternoo	team	1	-0.7426	0.2706	7.5332	0.0061

Odds Ratio Estimates

Effect	style	Point Estimate	95% Wald Confidence Limits	
school 2 vs 1	self	2.953	1.476	5.909
school 2 vs 1	team	1.197	0.643	2.230
school 3 vs 1	self	3.724	1.755	7.902
school 3 vs 1	team	1.926	0.990	3.747
program afternoo vs regular	self	0.474	0.273	0.823
program afternoo vs regular	team	0.476	0.280	0.809

SAS Code for OR Calculation

The SAS code is:

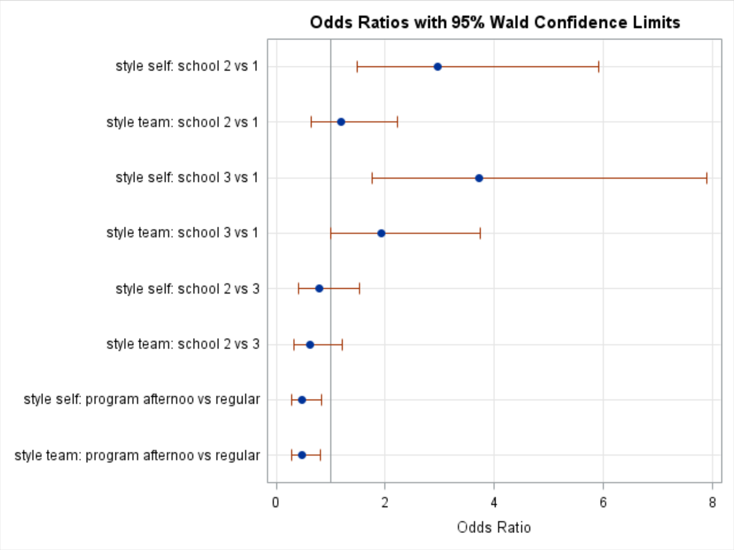
```
ods graphics on;
proc logistic data=school;
   freq count;
   class school(ref='1') program (ref='regular')/param=ref;
   model style(ref='class')=school program / link=glogit;
   oddsratio school;
   oddsratio program;
run;
ods graphics off;
```

SAS Output

Odds Ratio Estimates and Wald Confidence Intervals

Label	Estimate	95% Confidence Limits	
style self: school 2 vs 1	2.953	1.476	5.909
style team: school 2 vs 1	1.197	0.643	2.230
style self: school 3 vs 1	3.724	1.755	7.902
style team: school 3 vs 1	1.926	0.990	3.747
style self: school 2 vs 3	0.793	0.413	1.522
style team: school 2 vs 3	0.622	0.317	1.219
style self: program afternoo vs regular	0.474	0.273	0.823
style team: program afternoo vs regular	0.476	0.280	0.809

Plot for ORs



Odds Ratios with 95% Wald Confidence Limits
style self: school 2 vs 1
style team: school 2 vs 1
style self: school 3 vs 1
style team: school 3 vs 1
style self: school 2 vs 3
style team: school 2 vs 3
style self: program afternoo vs regular
style team: program afternoo vs regular
0
2
4
6
8
Odds Ratio

Predicted Probabilities for Style Preference

$$\widehat{Pr}(Y = \text{class}|X) = \frac{1}{1 + e^{\hat{\beta}_{0s}+\hat{\beta}_{1s}X_1+\hat{\beta}_{2s}X_2+\hat{\beta}_{3s}X_3} + e^{\hat{\beta}_{0t}+\hat{\beta}_{1t}X_1+\hat{\beta}_{2t}X_2+\hat{\beta}_{3t}X_3}}$$

$$\widehat{Pr}(Y = \text{self}|X) = \frac{e^{\hat{\beta}_{0s}+\hat{\beta}_{1s}X_1+\hat{\beta}_{2s}X_2+\hat{\beta}_{3s}X_3}}{1 + e^{\hat{\beta}_{0s}+\hat{\beta}_{1s}X_1+\hat{\beta}_{2s}X_2+\hat{\beta}_{3s}X_3} + e^{\hat{\beta}_{0t}+\hat{\beta}_{1t}X_1+\hat{\beta}_{2t}X_2+\hat{\beta}_{3t}X_3}}$$

$$\widehat{Pr}(Y = \text{team}|X) = \frac{e^{\hat{\beta}_{0t}+\hat{\beta}_{1t}X_1+\hat{\beta}_{2t}X_2+\hat{\beta}_{3t}X_3}}{1 + e^{\hat{\beta}_{0s}+\hat{\beta}_{1s}X_1+\hat{\beta}_{2s}X_2+\hat{\beta}_{3s}X_3} + e^{\hat{\beta}_{0t}+\hat{\beta}_{1t}X_1+\hat{\beta}_{2t}X_2+\hat{\beta}_{3t}X_3}}$$

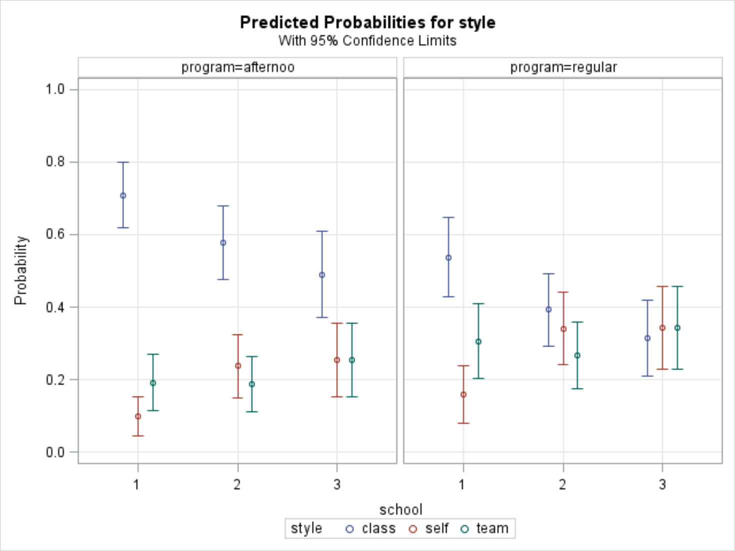
SAS Code for Prediction

```
ods graphics on;
proc logistic data=school;
   freq count;
   class school(ref='1') program (ref='regular')/param=ref;
   model style(ref='class')=school program / link=glogit;
   effectplot interaction(plotby=program) / clm noobs;
run;
ods graphics off;
```

The EFFECTPLOT statement creates a plot of the predicted values versus the levels of the School variable at each level of the Program variable.

- The CLM option adds confidence bars.
- The NOOBS option suppresses the display of the observations.

Plot for Predicted Values



Summary: Key Points

- How to write multinomial logistic regression?
- Understand the degrees of freedom associated with each test.
- How to interpret intercepts and slopes in multinomial logistic regression?
- How to code the multinomial logistic regression in SAS?
- How to calculate the predicted probability $\widehat{Pr}(Y = k|X)$?

Readings

Flom P.L “Multinomial and ordinal logistic regression using PROC LOGISTIC". NESUG 18.