

PHÂN TÍCH HỒI QUI LOGISTIC ĐA THỨC

Mô hình hồi qui logistic nhị thức (binary logistic) để xem xét mối quan hệ giữa biến phụ thuộc là biến nhị phân (ví dụ: sống/chết) và biến độc lập có thể là biến số (định lượng) hoặc biến định tính và phương trình liên hệ có dạng:

$$\text{Log}\left(\frac{p}{1-p}\right) = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots \quad (\text{trong đó } x_1, x_2, \dots \text{ là các biến độc}$$

lập). $\text{Odds} = \frac{p}{1-p}$ (p là xác suất tử vong, $1-p$ là xác suất còn sống)

20.1 Mô hình hồi qui logistic đa thức

Mô hình hồi qui logistic đa thức (Multinomial logistic regression) tương tự như mô hình hồi qui logistic nhị thức nhưng biến phụ thuộc là biến định tính có nhiều hơn 2 trạng thái (hoặc mức). Ví dụ (khỏi bệnh, khỏi với dư chứng, tử vong) hoặc (tốt, trung bình, xấu).

Mô hình hồi qui logistic đa thức phát biểu:

$$\text{Log}\left(\frac{p_i}{p_j}\right) = \alpha_{ij} + \beta_{ij1} X_1 + \beta_{ij2} X_2 + \dots + \varepsilon_{ij}$$

Gọi: p_0 là xác suất khỏi bệnh

p_1 là xác suất khỏi với dư chứng

p_2 xác suất tử vong

Ta có 3 phương trình sau:

$$\text{Log}\left(\frac{p_1}{p_0}\right) = \alpha_{10} + \beta_{101} X_1 + \beta_{102} X_2 + \dots \quad (1)$$

$$\text{Log}\left(\frac{p_2}{p_0}\right) = \alpha_{20} + \beta_{201} X_1 + \beta_{202} X_2 + \dots \quad (2)$$

$$\text{Log}\left(\frac{p_2}{p_1}\right) = \alpha_{21} + \beta_{211} X_1 + \beta_{212} X_2 + \dots \quad (3)$$

20.2 Phân tích hồi qui logistic đa thức trong SPSS

Tìm sự liên hệ giữa cân nặng, triệu chứng thiếu máu của các bà mẹ mang thai và trẻ sinh ra nhẹ cân. Biến phụ thuộc (kết cục) sinh nhẹ cân là biến định tính, được chia làm 3 mức: bình thường ($>2500\text{g}$), nhẹ cân trung bình ($1500-2500\text{g}$) và nhẹ cân nặng (< 1500).

Bảng 20.1 Số liệu 30 bà mẹ sinh con

Mabn	Canme	Thieumau	Nhecan	Mabn	Canme	Thieumau	Nhecan
1	44	0	0	16	52	0	1
2	46	0	0	17	55	1	1
3	48	0	0	18	54	1	1
4	50	0	0	19	55	1	1
5	52	0	0	20	60	1	1
6	54	0	0	21	50	0	2
7	56	0	0	22	47	0	2
8	58	0	0	23	56	0	2
9	60	1	0	24	49	1	2
10	62	1	0	25	55	1	2
11	47	0	1	26	51	1	2
12	48	0	1	27	45	1	2
13	51	0	1	28	53	1	2
14	50	0	1	29	47	1	2
15	53	0	1	30	46	1	2

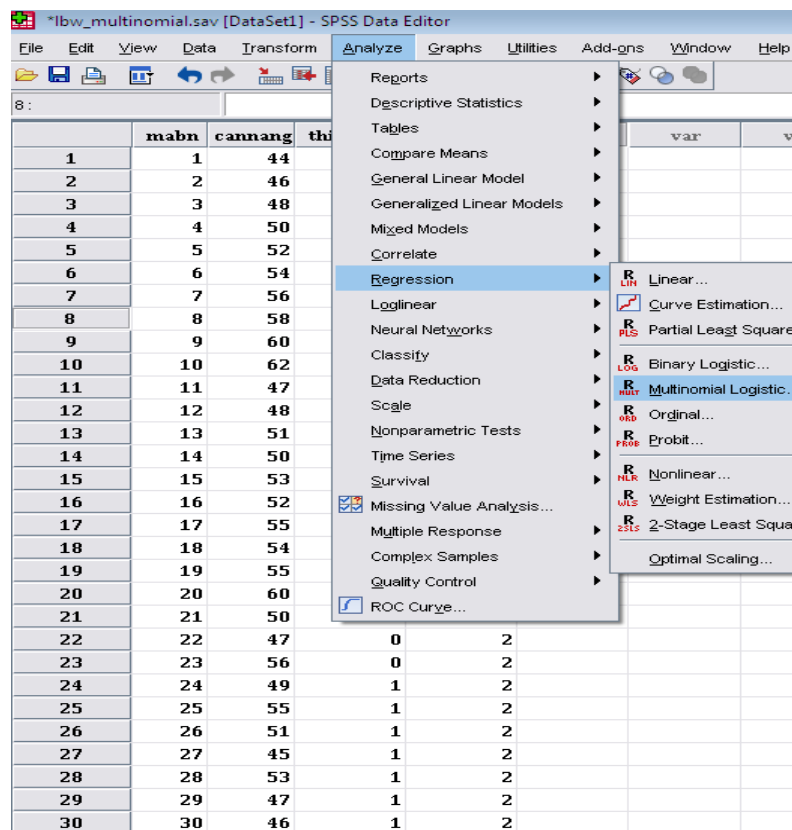
Cột 1: mã BN

Cột 2: Cân nặng của mẹ (biến số)

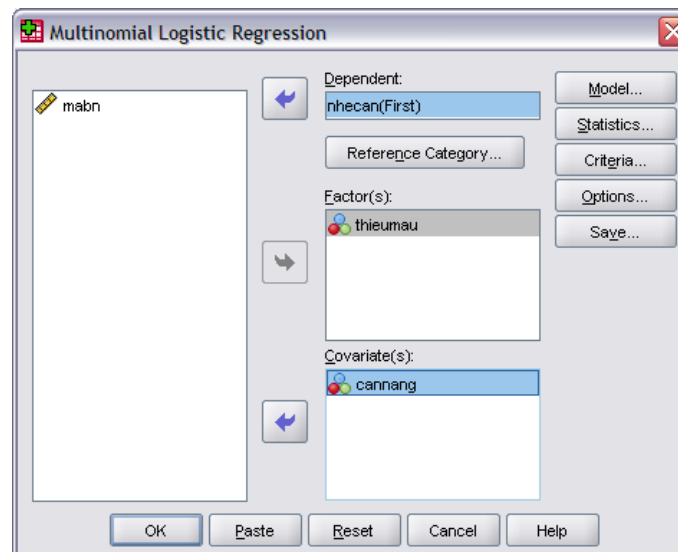
Cột 3: Triệu chứng thiếu máu của mẹ (biến nhị phân 0 và 1)

Cột 4 : Biến kết cục sinh nhẹ cân có 3 mức (0: bình thường; 1: nhẹ cân trung bình; 2: nhẹ cân mức độ nặng)

Vào menu: **Analyze> Regression>Multinomial Logistic**

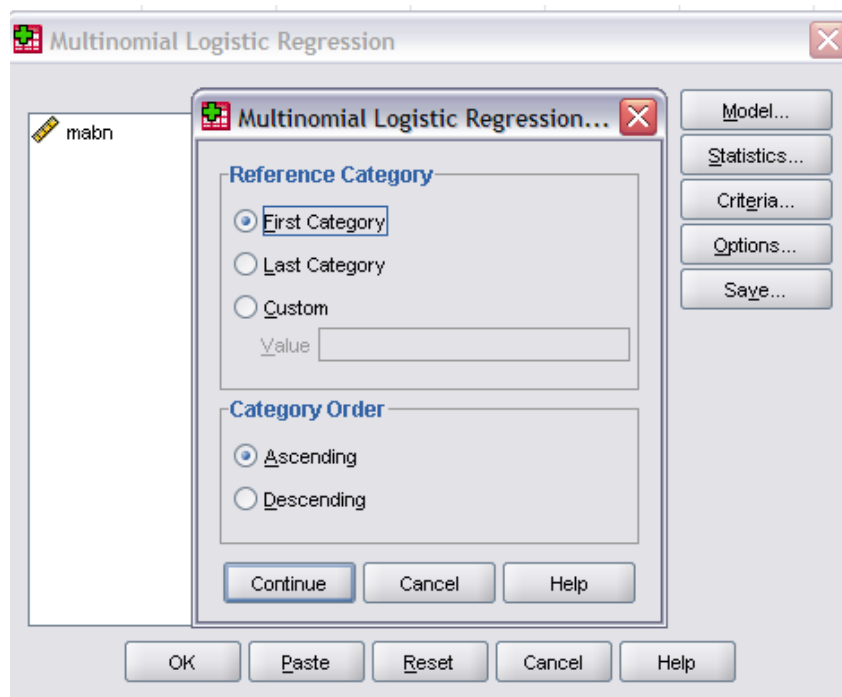


Vào màn hình Multinomial Logistic Regression



Nhập chuyển biến nhecan vào ô biến phụ thuộc (Dependent), biến thieumau(biến nhị phân) vào ô Factor(s) và biến cannang vào ô Covariate

Nhấn vào hộp thoại Reference Category, chọn mức tham khảo là First Category, có mã nhecan=0 (cân nặng bình thường) như hình sau:



Nhấp Continue, nhấn OK cho kết quả sau:

Nominal Regression

[DataSet1] F:\BAIVIET\BAITAP\lbw_multinomial.sav

Case Processing Summary

	N	Marginal Percentage
nhecan 0	10	33.3%
1	10	33.3%
2	10	33.3%
thieumau 0	17	56.7%
1	13	43.3%
Valid	30	100.0%
Missing	0	
Total	30	
Subpopulation	21 ^a	

a. The dependent variable has only one value observed in 14 (66.7%) subpopulations.

Model Fitting Information

Model	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	53.205			
Final	41.099	12.105	4	.017

Pseudo R-Square

Cox and Snell	.332
Nagelkerke	.374
McFadden	.184

Likelihood Ratio Tests

Effect	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood of Reduced Model	Chi-Square	df	Sig.
Intercept	41.099 ^a	.000	0	.
cannang	47.836	6.737	2	.034
thieumau	50.474	9.375	2	.009

Các bảng trên mô tả sự phù hợp của mô hình, nói chung trị số -2 Log likelihood càng nhỏ sự phù hợp mô hình càng lớn. Còn trị pseudo- R^2 gần tương tự như R^2 trong mô hình hồi qui tuyến tính.

Bảng 20.2 Kết quả so sánh giữa nhóm nhẹ cân nặng (hoặc trung bình) so với nhóm sơ sinh cân nặng bình thường (reference category: 0)

Parameter Estimates								
		B	Std. Error	Wald	df	Sig.	Exp(B)	95% Confidence Interval for Exp (B)
nhecan ^a								Lower Bound Upper Bound
1	Intercept	6.771	7.077	.915	1	.339		
	cannang	-.107	.123	.758	1	.384	.898	.705 1.144
	[thieumau=0]	-1.563	1.263	1.531	1	.216	.210	.018 2.491
	[thieumau=1]	0 ^b			0			
2	Intercept	19.125	8.505	5.056	1	.025		
	cannang	-.329	.153	4.631	1	.031	.719	.533 .971
	[thieumau=0]	-3.683	1.478	6.205	1	.013	.025	.001 .456
	[thieumau=1]	0 ^b			0			
a. The reference category is: 0.								
b. This parameter is set to zero because it is redundant.								

Bảng kết quả gồm 2 hàng 9 cột.

Hàng có số 1 (nhẹ cân trung bình) các biến dự đoán (cân nặng mẹ, thiếu máu) được so sánh với mức tham khảo là 0 (cân bình thường)

Hàng có số 2 (nhẹ cân mức độ nặng) các biến dự đoán (cân nặng mẹ, thiếu máu) được so sánh với mức tham khảo là 0 (cân bình thường)

Cột 1: gồm các biến dự đoán (cân nặng mẹ và thiếu máu)

Cột 2 : B là hệ số hồi qui

Cột 3: Sai số chuẩn

Cột 4: Trị của Wald test tương tự χ^2 (chi square) trong mô hình hồi qui tuyến tính.

Cột 5: df (bậc tự do)

Cột 6: Sig. (ý nghĩa thống kê)

Cột 7: Exponent của hệ số hồi qui B ở cột 2 (e^B) chính là odds ratio

Cột 8-9: khoảng tin cậy 95% của OR

Diễn dịch kết quả trên như sau:

Nếu gọi p_0 là xác suất trẻ có cân nặng bình thường

p_1 là xác suất trẻ nhẹ cân trung bình

$$\text{Log} \left(\frac{p_1}{p_0} \right) = 6,77 - 0,10x \text{ Cannangme} - 1.56x \text{ Thieumau}$$

Nếu mẹ không bị thiếu máu thì log(odds) giảm được 1,56 hoặc giảm được 79% nguy cơ sinh nhẹ cân trung bình (OR=0,21; KTC 95%:0,01-2,49), tuy nhiên sự khác biệt này không có ý nghĩa thống kê ($p=0,21$). Nếu mẹ có cân nặng bình thường thì log(odds) giảm được 0,10 hoặc giảm được 11% nguy cơ sinh nhẹ cân trung bình (OR=0,89; KTC95%: 0,70 - 1,14), tuy nhiên sự khác biệt này cũng không có ý nghĩa thống kê ($p=0,38$). Như vậy, mẹ có cân nặng thấp và thiếu máu chưa phải là yếu tố nguy cơ sinh con nhẹ cân mức độ trung bình (1500-2500g).

Nếu gọi p_0 là xác suất trẻ có cân nặng bình thường

p_2 là xác suất trẻ nhẹ cân nặng

$$\text{Log} \left(\frac{p_2}{p_0} \right) = 19,1 - 0,32 \times \text{Cannangme} - 3,68 \times \text{Thieumau}$$

Nếu mẹ không bị thiếu máu thì log(odds) giảm được 3,68 hoặc giảm được 99,5% nguy cơ sinh nhẹ cân mức độ nặng (OR=0,025; KTC 95%: 0,001-0,456; $p=0,013$).

Nếu mẹ có cân nặng bình thường thì log(odds) giảm được 0,32 hoặc giảm được 28% nguy cơ sinh nhẹ cân mức độ nặng (OR=0,719; KTC95%: 0,53 - 0,97; $p=0,031$). Như vậy, mẹ nhẹ cân và thiếu máu là yếu tố nguy cơ sinh con nhẹ cân mức độ nặng (<1500g).

Để so sánh các yếu tố nguy cơ dẫn đến sinh nhẹ cân nặng với nhẹ cân trung bình, ta khai loại tham khảo (reference category) là 1 trong SPSS.

Bảng 20.3 Kết quả so sánh giữa sinh nhẹ cân nặng với nhẹ cân trung bình (reference category: 1)

Parameter Estimates								
nhecan ^a		B	Std. Error	Wald	df	Sig.	Exp(B)	95% Confidence Interval for Exp (B)
								Lower Bound Upper Bound
0	Intercept	-6.771	7.077	.915	1	.339		
	cannang	.107	.123	.758	1	.384	1.113	.874 1.418
	[thieumau=0]	1.563	1.263	1.531	1	.216	4.772	.402 56.724
	[thieumau=1]	0 ^b			0			
2	Intercept	12.354	7.068	3.055	1	.080		
	cannang	-.222	.131	2.860	1	.091	.801	.619 1.036
	[thieumau=0]	-2.120	1.189	3.180	1	.075	.120	.012 1.234
	[thieumau=1]	0 ^b			0			

a. The reference category is: 1.
b. This parameter is set to zero because it is redundant.

$$\text{Log} \left(\frac{p_2}{p_1} \right) = 12,3 - 0,22 \times \text{Cannangme} - 2,12 \times \text{Thieumau}$$

So sánh giữa nhóm trẻ sơ sinh nhẹ cân nặng với trẻ sơ sinh nhẹ cân trung bình, thì sự khác biệt về cân nặng mẹ ($p=0,091$) và thiếu máu ($p=0,075$) chưa có ý nghĩa thống kê.

Tóm lại mẹ có thiếu máu ($p=0,013$) và cân nặng thấp ($p=0,03$) làm tăng nguy cơ sinh con nhẹ cân ở mức độ nặng.

Tài liệu tham khảo:

Germán Rodríguez. The Multinomial Logit Model, Princeton University, website: <http://data.princeton.edu/>, truy cập 02/2010