

ĐẠI HỌC QUỐC GIA TP HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
NGÀNH KHOA HỌC DỮ LIỆU



Mô hình tuyến tính nâng cao
Nghiên cứu các vấn đề mở rộng của
Mô hình hồi quy logistic đa thức

GVHD: TS. Tô Đức Khánh

HV: Lê Thị Mai Thảo - 23C01041

Tp. Hồ Chí Minh - 2/2025

1 Mô hình hồi quy logistic đa thức (multinomial logistic)

1.1 Định nghĩa

Hồi quy logistic đa thức (Multinomial Logistic Regression) là một mô hình thống kê mở rộng của hồi quy logistic nhị phân, được sử dụng khi biến phản hồi Y có nhiều hơn hai nhóm. Trong mô hình này, có hai dạng biến. Đầu tiên, các biến phụ thuộc và có thứ bậc (ordinal response); ví dụ như các bậc tăng dần về lương, nhóm tuổi, hay các mức độ hài lòng của khách hàng,... Thứ hai, các biến độc lập và không có thứ tự (multinomial response) ; ví dụ như các loại xe, tình trạng chẩn đoán bệnh,... Mô hình hồi quy logistic đa thức được định nghĩa:

$$\log \left(\frac{\pi_j(x)}{\pi_1(x)} \right) = \beta_{j0} + \beta_{j1}x_1 + \beta_{j2}x_2 + \dots + \beta_{jp}x_p$$

trong đó:

- $\pi_j(x) = P(Y = j|X = x)$ là xác suất của nhóm j với tập hợp biến độc lập $X = (x_1, x_2, \dots, x_p)$.
- $\pi_1(x) = P(Y = 1|X = x)$ là xác suất của nhóm tham chiếu.
- $\beta_{j0}, \beta_{j1}, \dots, \beta_{jp}$ là các hệ số hồi quy cần ước lượng cho từng nhóm j .

1.2 Mô hình tỷ số cược tỷ lệ

Mô hình tỷ số cược tỷ lệ (proportional odds model) được sử dụng cho biến phản hồi có thứ bậc (ordinal response variable). Mục tiêu là dự đoán xác suất của một quan sát thuộc vào một nhóm bất kỳ trong tập hợp các nhóm có thứ bậc. Phương trình của mô hình có dạng:

$$\log \left(\frac{\theta_j(x)}{\theta_1(x)} \right) = \beta_{j0} + \beta_1x_1 + \beta_2x_2 + \dots + \beta_px_p$$

Trong đó: - $\theta_j(x) = P(Y \leq j|X = x)$ là xác suất tích lũy của hạng mục j . - $\theta_1(x)$ là xác suất tích lũy của hạng mục đầu tiên. - β_{j0} là hằng số chặn (intercept) tương ứng với hạng mục j . - $\beta_1, \beta_2, \dots, \beta_p$ là hệ số hồi quy, giống nhau cho mọi mức của biến phản hồi, đảm bảo giả định tỷ số cược tỷ lệ.

2 Điều chỉnh ước lượng hệ số của mô hình

Trong hồi quy logistic đa biến có thứ bậc (ordinal logistic regression), ước lượng hợp lý cực đại có thể bị thiên lệch. Để giảm thiên lệch, ta có thể điều chỉnh tham số.

Phương trình score function trong ước lượng hợp lý cực đại là:

$$U_t = \sum_{r=1}^n w_r d_r (y_r - \mu_r) x_{rt}.$$

trong đó,

- $w_r = \frac{d_r^2}{\sigma^2}$
- $d_r = \frac{d\mu_r}{d\eta_r}$
- $x_{rt} = \frac{\partial \eta_r}{\partial \beta_t}$

Để giảm thiên lệch, Firth (1993) đề xuất cực đại hóa hàm hợp lý có phạt:
[2] [1] [3]

$$\log L^*(\beta) = \log L(\beta) + \frac{1}{2} \log |I(\beta)|$$

trong đó $I(\beta)$ là ma trận thông tin Fisher.

Phương trình điều chỉnh theo Kosmidis và Firth (2009):

$$U_t = \sum_{r=1}^n w_r d_r (y_r^* - \mu_r) x_{rt}.$$

trong đó, biến phản hồi điều chỉnh:

$$y_r^* = y_r + \frac{1}{2} h_r \frac{d'_r}{w_r} + d_r \text{tr}(F^{-1} D^2(\eta_r; \beta)).$$

trong đó:

- $d_r = \frac{d\mu_r}{d\eta_r}$, đạo hàm của kỳ vọng.
- $d'_r = \frac{d^2 \mu_r}{d\eta_r^2}$, đạo hàm bậc hai. h_r là phần tử đường chéo thứ r của ma trận $H = XF^{-1}X^TW$.
- $F = X^TWX$, ma trận thông tin Fisher.

Thay thế y_r bằng y_r^* trong phương pháp bình phương tối thiểu có trọng số lặp (iterative reweighted least squares - IWLS) giúp giảm thiên lệch.

Thuật toán sẽ được triển khai như sau:

- Bước 1: Khởi tạo
 - Chọn giá trị ban đầu $\beta^{(0)}$.
 - Tính ma trận trọng số $W^{(0)}$ từ mô hình logistic có thứ bậc.
- Bước 2: Cập nhật tham số. Tại vòng lặp j , cập nhật:

$$\tilde{\beta}^{(j+1)} = (X^TW^{(j)}X)^{-1}X^TW^{(j)}(\zeta^{(j)} - \xi^{(j)}).$$

trong đó:

$$- \zeta_r = \sum_{t=1}^p \beta_t x_{rt} + \frac{y_r - \mu_r}{d_r} \text{ (quan sát làm việc của ML).}$$

$$- \xi_r = -\frac{d'_r h_r}{2w_r d_r} - \frac{\text{tr}(F^{-1} D^2(\eta_r; \beta))}{2}.$$

- Bước 3: Lặp cho đến hội tụ. Dừng lại khi $|\tilde{\beta}^{(j+1)} - \tilde{\beta}^{(j)}|$ nhỏ hơn một ngưỡng nhất định.

3 Kết quả

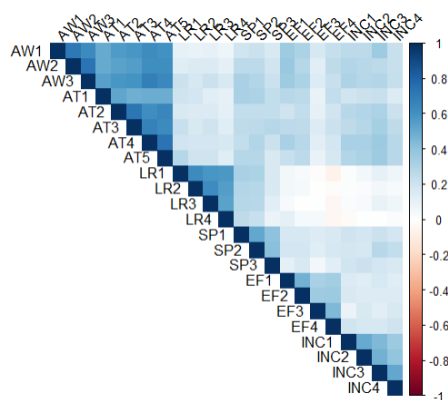
3.1 Mô tả biến

Danh mục	Mô tả
Demographics	
Gender	Giới tính (1-Nam; 2-Nữ)
Age	Tuổi (0=15-22; 1=23-45)
Education.level	Trình độ học vấn (1= Trung học phổ thông; 2= Đại học; 3=Sau đại học)
Income	Thu nhập (1=Không có thu nhập; 2= <4 triệu; 3= 4-10 triệu; 4= >10 triệu)
Nhận thức về rác thải điện tử	
AW1	Thu gom và tái chế chất thải điện tử làm giảm việc khai thác tài nguyên làm nguyên liệu thô.
AW2	Thu gom và tái chế rác thải điện tử giúp cải thiện chất lượng môi trường.
AW3	Tái chế chất thải điện tử cho phép thu hồi các vật liệu có giá trị.
Thái độ đối với việc thu gom và tái chế chất thải	
AT1	Thu gom rác thải điện tử đóng góp đáng kể cho xã hội.
AT2	Tham gia thu gom và tái chế rác thải điện tử hỗ trợ sự phát triển bền vững.
AT3	Tổ chức các hoạt động thu gom, tái chế rác thải điện tử là một giải pháp hữu ích.
AT4	Mọi người đều có trách nhiệm tái chế rác thải điện tử.
AT5	Tôi quan tâm đến ý tưởng thu gom rác thải điện tử.
Luật và quy định	
LR1	Pháp luật Việt Nam quy định trách nhiệm rõ ràng của công dân trong việc tái chế rác thải điện tử.
LR2	Chính sách của chính phủ sẽ tác động đến hành vi tái chế chất thải điện tử.
LR3	Tôi sẽ tuân thủ luật pháp và quy định liên quan đến tái chế chất thải điện tử.
LR4	Tôi sẽ tham gia các khóa học về luật và quy định về thu gom và tái chế rác thải điện tử nếu có.
Áp lực xã hội	

SP1	Tôi sẽ tham gia thu gom và tái chế rác thải điện tử nếu gia đình và bạn bè tôi tham gia.
SP2	Truyền thông ảnh hưởng đến việc thu gom và tái chế rác thải điện tử.
SP3	Cộng đồng địa phương tác động đến sự tham gia của tôi trong việc thu gom và tái chế rác thải điện tử.
Yếu tố kinh tế	
EF1	Chi phí vận chuyển cao đối với việc tái chế rác thải điện tử.
EF2	Phí thu gom và tái chế rác thải điện tử cao.
EF3	Các chương trình tái chế có chi phí đắt đỏ.
EF4	Lợi ích khi tham gia thu gom và tái chế rác thải điện tử: khuyến mãi, ưu đãi, v.v.
Sự bất tiện	
INC1	Tôi thấy khó phân loại rác thải điện tử để tái chế.
INC2	Tôi không có thời gian mang rác thải điện tử đến điểm thu gom.
INC3	Tôi thấy việc vận chuyển rác thải điện tử đến điểm thu gom rất bất tiện.
INC4	Các chương trình thu gom rác thải điện tử hiện tại thiếu các điểm thu gom.
Ý định tái chế	
RI1	Tôi sẵn sàng tham gia các mô hình thu gom rác thải điện tử khi có sẵn.

Thang đo (likert scale 5): Hoàn toàn không đồng ý, không đồng ý, trung lập, đồng ý, hoàn toàn đồng ý.

3.2 Tính factor



Hình 1: Tương quan của các indicator

Kết quả cho thấy các biến ý thức (Awareness- AW) và thái độ (Attitude- AT) có tương quan khá cao (> 60%). Các biến khảo sát cho 1 chủ đề cũng sẽ có sự tương quan cao với nhau. Do vậy, các biến có kết quả tương quan cao sẽ được nhóm nhau (Bảng 1.

Sau khi khảo sát mô hình nhóm các biến bằng như mô tả ở trên bằng phân tích nhân tố khẳng định (Confirmatory Factor Analysis - CFA), kết quả như sau:

```

1      lavaan 0.6-19 ended normally after 49 iterations
2
3      Estimator                      ML
4      Optimization method            NLMINB
5      Number of model parameters      52
6
7      Number of observations          1020
8
9      Model Test User Model:
10
11     Test statistic                    1658.505
12     Degrees of freedom                224
13     P-value (Chi-square)             0.000
14
15     Model Test Baseline Model:
16
17     Test statistic                    10909.040
18     Degrees of freedom                253
19     P-value                          0.000
20
21     User Model versus Baseline Model:
22
23     Comparative Fit Index (CFI)       0.865
24     Tucker-Lewis Index (TLI)         0.848
25
26     Loglikelihood and Information Criteria:
27
28     Loglikelihood user model (H0)      -17456.715
29     Loglikelihood unrestricted model (H1) -16627.463
30
31     Akaike (AIC)                      35017.430
32     Bayesian (BIC)                    35273.663
33     Sample-size adjusted Bayesian (SABIC) 35108.506
34
35     Root Mean Square Error of Approximation:
36
37     RMSEA                             0.079
38     90 Percent confidence interval - lower 0.076
39     90 Percent confidence interval - upper 0.083
40     P-value H_0: RMSEA <= 0.050         0.000
41     P-value H_0: RMSEA >= 0.080         0.367
42
43     Standardized Root Mean Square Residual:
44
45     SRMR                              0.063
46
47     Parameter Estimates:
48

```

49	Standard errors				Standard	
50	Information				Expected	
51	Information saturated (h1) model				Structured	
52						
53	Latent Variables:					
54		Estimate	Std.Err	z-value	P(> z)	Std.
		lv	Std.all			
55	Awareness_Attitude = ~					
56	AW1	0.547	0.019	28.834	0.000	
		0.547	0.774			
57	AW2	0.518	0.018	29.458	0.000	
		0.518	0.786			
58	AW3	0.461	0.015	31.276	0.000	
		0.461	0.818			
59	AT1	0.351	0.016	22.159	0.000	
		0.351	0.637			
60	AT2	0.434	0.015	28.257	0.000	
		0.434	0.764			
61	AT3	0.419	0.014	29.269	0.000	
		0.419	0.782			
62	AT4	0.429	0.013	32.251	0.000	
		0.429	0.834			
63	AT5	0.471	0.015	31.070	0.000	
		0.471	0.814			
64	Law_Regulations = ~					
65	LR1	0.583	0.021	28.183	0.000	
		0.583	0.790			
66	LR2	0.528	0.018	28.767	0.000	
		0.528	0.802			
67	LR3	0.474	0.018	26.641	0.000	
		0.474	0.758			
68	LR4	0.478	0.020	24.421	0.000	
		0.478	0.710			
69	Economic_Factors = ~					
70	SP1	0.364	0.020	18.038	0.000	
		0.364	0.591			
71	SP2	0.403	0.022	18.544	0.000	
		0.403	0.605			
72	SP3	0.284	0.020	14.078	0.000	
		0.284	0.476			
73	EF1	0.369	0.025	14.827	0.000	
		0.369	0.498			
74	EF2	0.345	0.023	15.027	0.000	
		0.345	0.504			
75	EF3	0.210	0.018	11.406	0.000	
		0.210	0.393			
76	EF4	0.263	0.020	13.337	0.000	
		0.263	0.453			
77	Inconveniences = ~					
78	INC1	0.423	0.020	20.800	0.000	
		0.423	0.652			
79	INC2	0.430	0.020	21.796	0.000	
		0.430	0.678			
80	INC3	0.469	0.020	23.949	0.000	
		0.469	0.732			
81	INC4	0.389	0.020	19.660	0.000	
		0.389	0.623			

```

82
83 Covariances:
84
85      Estimate Std.Err z-value P(>|z|) Std.
86      lv Std.all
87
88 Awareness_Attitude ~~
89   Law_Regulatings    0.279    0.033    8.516    0.000
90     0.279    0.279
91   Economic_Fctrs    0.605    0.027   22.022    0.000
92     0.605    0.605
93   Inconveniences    0.537    0.028   19.018    0.000
94     0.537    0.537
95 Law_Regulations ~~
96   Economic_Fctrs    0.339    0.037    9.243    0.000
97     0.339    0.339
98   Inconveniences    0.081    0.038    2.106    0.035
99     0.081    0.081
100 Economic_Factors ~~
101   Inconveniences    0.516    0.034   15.018    0.000
102     0.516    0.516
103
104 Variances:
105
106      Estimate Std.Err z-value P(>|z|) Std.lv
107      Std.all
108
109   .AW1    0.200    0.010   20.081    0.000    0.200
110     0.400
111   .AW2    0.167    0.008   19.885    0.000    0.167
112     0.383
113   .AW3    0.105    0.005   19.204    0.000    0.105
114     0.331
115   .AT1    0.180    0.008   21.444    0.000    0.180
116     0.594
117   .AT2    0.135    0.007   20.247    0.000    0.135
118     0.417
119   .AT3    0.112    0.006   19.946    0.000    0.112
120     0.388
121   .AT4    0.081    0.004   18.752    0.000    0.081
122     0.304
123   .AT5    0.113    0.006   19.291    0.000    0.113
124     0.337
125   .LR1    0.205    0.013   15.959    0.000    0.205
126     0.376
127   .LR2    0.155    0.010   15.399    0.000    0.155
128     0.357
129   .LR3    0.166    0.010   17.229    0.000    0.166
130     0.425
131   .LR4    0.224    0.012   18.586    0.000    0.224
132     0.495
133   .SP1    0.247    0.013   18.782    0.000    0.247
134     0.651
135   .SP2    0.281    0.015   18.489    0.000    0.281
136     0.634
137   .SP3    0.276    0.013   20.526    0.000    0.276
138     0.774
139   .EF1    0.413    0.020   20.258    0.000    0.413
140     0.752
141   .EF2    0.349    0.017   20.182    0.000    0.349
142     0.746

```


114	.EF3	0.241	0.011	21.305	0.000	0.241
	0.846					
115	.EF4	0.267	0.013	20.768	0.000	0.267
	0.795					
116	.INC1	0.241	0.013	18.099	0.000	0.241
	0.574					
117	.INC2	0.217	0.012	17.414	0.000	0.217
	0.540					
118	.INC3	0.190	0.012	15.544	0.000	0.190
	0.463					
119	.INC4	0.239	0.013	18.769	0.000	0.239
	0.612					
120	Awareness_Attd	1.000				1.000
	1.000					
121	Law_Regulations	1.000				1.000
	1.000					
122	Economic_Fctrs	1.000				1.000
	1.000					
123	Inconveniences	1.000				1.000
	1.000					

Kết quả:

- CFI và TLI đo lường mức độ cải thiện của mô hình so với mô hình cơ sở bằng cách so sánh độ phù hợp giữa hai mô hình. Giá trị CFI dao động từ 0 đến 1. Với ≥ 0.90 : phù hợp tốt, $0.85 - 0.89$: có thể chấp nhận nhưng cần cải thiện. **Kết quả cho CFI = 0.865, TLI = 0.848 cho thấy mô hình có thể chấp nhận.**
- RMSEA (Root Mean Square Error of Approximation) đo lường sai số trung bình của mô hình so với dữ liệu thực tế. Giá trị RMSEA càng nhỏ càng tốt: ≤ 0.05 : Rất tốt; $0.05 - 0.08$: Chấp nhận được; > 0.08 : Không phù hợp. Với **RMSEA = 0.079, mô hình ở ngưỡng chấp nhận được.**
- SRMR đo lường sự khác biệt giữa ma trận hiệp phương sai quan sát và mô hình hóa (giá trị càng nhỏ, sự khác biệt càng ít). ≤ 0.08 : Mô hình phù hợp tốt. ≤ 0.05 : Mô hình rất tốt. Với **SRMR = 0.063, mô hình đạt mức phù hợp tốt.**

3.3 Ước lượng mô hình GLM

```

1  Call:
2  polr(formula = RI1 ~ Age + Education.level + Income + Awareness_
      Attitude +
3      Law_Regulations + Economic_Factors + Inconveniences, data =
      data,
4      Hess = TRUE)
5
6  Coefficients:
7              Value Std. Error t value
8  Age           0.01579   0.13348  0.1183
9  Education.level 0.86470   0.16359  5.2858
10 Income        -0.03206   0.12813 -0.2502
11 Awareness_Attd -0.75595   0.12960 -5.8331

```

```

12 Law_Regulations      0.08696      0.08562      1.0156
13 Economic_Factors     0.49065      0.12901      3.8033
14 Inconveniencies      0.60261      0.10709      5.6269
15
16 Intercepts:
17      Value Std. Error t value
18 1|2  0.9482  0.3244      2.9229
19 2|3  4.8036  0.3647     13.1722
20
21 Residual Deviance: 1555.022
22 AIC: 1573.022

```

Kết quả cho thấy trình độ học vấn, các yếu tố kinh tế, sự bất tiện có ảnh hưởng tích cực, mang ý nghĩa thống kê cao. Trái lại, ý thức và thái độ có ảnh hưởng tiêu cực về thống kê. Tuổi, thu nhập và luật thì cho kết quả không có ý nghĩa thống kê.

Biến	Hệ số Odds Ratio
Tuổi	1.0159
Trình độ học vấn	2.3743
Thu nhập	0.9684
Nhận thức và thái độ	0.4696
Luật và quy định	1.0908
Yếu tố kinh tế	1.6334
Sự bất tiện	1.8269

Bảng 2: Hệ số Odds Ratio từ mô hình hồi quy

Khi tuổi lên 1 nhóm mới, khả năng có ý định tái chế tăng 1.59%, nhưng mức ảnh hưởng này không đáng kể. Trong khi đó, trình độ học vấn có tác động mạnh hơn: khi tăng 1 bậc học vấn, khả năng có ý định tái chế tăng gấp 2.37 lần (Bảng 2).

Thu nhập lại có tác động ngược chiều nhưng khá nhỏ. Khi thu nhập tăng 1 bậc, khả năng có ý định tái chế giảm 3.16%. Ngược lại, nhận thức và thái độ có ảnh hưởng đáng kể, khi mức độ nhận thức và thái độ tăng, khả năng không tái chế giảm 53.04%, nghĩa là ý định tái chế tăng mạnh (Bảng 2).

Khi cảm nhận về luật pháp được cải thiện thêm 1 mức độ, khả năng có ý định tái chế tăng 9.08%. Yếu tố kinh tế, có tác động lớn hơn—tăng 1 đơn vị giúp khả năng tái chế tăng 63.34%. Đáng chú ý, ngay cả khi sự bất tiện tăng lên, khả năng có ý định tái chế vẫn tăng 82.69% (Bảng 2).

3.4 Mô hình stepwise

```

1      Start:  AIC=1573.02
2  RI1 ~ Age + Education.level + Income + Awareness_Attitude + Law_
      Regulations +
3      Economic_Factors + Inconveniencies
4
5              Df      AIC

```

```

6 - Age 1 1571.0
7 - Income 1 1571.1
8 - Law_Regulations 1 1572.1
9 <none> 1573.0
10 - Economic_Factors 1 1585.6
11 - Education.level 1 1599.6
12 - Inconveniences 1 1603.9
13 - Awareness_Attitude 1 1606.3
14
15 Step: AIC=1571.04
16 RI1 ~ Education.level + Income + Awareness_Attitude + Law_
    Regulations +
17 Economic_Factors + Inconveniences
18
19 Df AIC
20 - Income 1 1569.1
21 - Law_Regulations 1 1570.1
22 <none> 1571.0
23 + Age 1 1573.0
24 - Economic_Factors 1 1583.7
25 - Education.level 1 1597.6
26 - Inconveniences 1 1601.9
27 - Awareness_Attitude 1 1604.3
28
29 Step: AIC=1569.1
30 RI1 ~ Education.level + Awareness_Attitude + Law_Regulations +
    Economic_Factors + Inconveniences
31
32 Df AIC
33 - Law_Regulations 1 1568.1
34 <none> 1569.1
35 + Income 1 1571.0
36 + Age 1 1571.1
37 - Economic_Factors 1 1581.7
38 - Inconveniences 1 1600.0
39 - Education.level 1 1603.5
40 - Awareness_Attitude 1 1612.2
41
42 Step: AIC=1568.06
43 RI1 ~ Education.level + Awareness_Attitude + Economic_Factors +
    Inconveniences
44
45 Df AIC
46 <none> 1568.1
47 + Law_Regulations 1 1569.1
48 + Age 1 1570.1
49 + Income 1 1570.1
50 - Economic_Factors 1 1586.6
51 - Inconveniences 1 1598.1
52 - Education.level 1 1601.7
53 - Awareness_Attitude 1 1610.2
54
55 Call:
56 polr(formula = RI1 ~ Education.level + Awareness_Attitude +
    Economic_Factors +
57 Inconveniences, data = data, Hess = TRUE)
58
59 Coefficients:
60

```

```

61           Value Std. Error t value
62 Education.level      0.8284    0.1405    5.896
63 Awareness_Attitude -0.7520    0.1160   -6.482
64 Economic_Factors     0.5373    0.1196    4.492
65 Inconveniences       0.5757    0.1034    5.567
66
67 Intercepts:
68   Value Std. Error t value
69 1|2  0.9267  0.2962    3.1289
70 2|3  4.7925  0.3418   14.0222
71
72 Residual Deviance: 1556.057
73 AIC: 1568.057

```

Kết quả cho thấy:

- Mô hình cuối cùng được chọn có 4 biến quan trọng nhất cho dự đoán biến phụ thuộc RI1 gồm Education.level, Awareness Attitude , Economic Factors, Inconveniences.
- Residual Deviance = 1556.057: Độ lệch phần dư thấp hơn mô hình ban đầu (tốt hơn).
- AIC = 1568.057: Giá trị AIC giảm từ 1573.02 đến 1568.06 (tốt hơn).
- Awareness Attitude (Nhận thức và thái độ) có hệ số âm. Điều này cho thấy, khi khách hàng có độ nhận thức về rác thải điện tử và thái độ về thu gom và tái chế càng cao thì khả năng chấp nhận tham gia chương trình tái chế càng thấp. Có thể đến từ việc họ cần có nhiều suy xét hơn trước khi tham gia.
- Đối với các biến về học vấn, các yếu tố về kinh tế và sự bất tiện lợi càng cao thì khả năng khách hàng chấp nhận càng cao.

3.5 Mô hình điều chỉnh tham số

```

1 brac1(formula = RI1 ~ Age + Education.level + Income + Awareness_
      Attitude +
2       Law_Regulations + Economic_Factors + Inconveniences, data =
      data,
3       method = "brglmFit")
4
5 Coefficients:
6           Estimate Std. Error z value Pr(>|z|)
7 1:(Intercept)      1.0611    0.3451    3.08  0.00211 **
8 2:(Intercept)      2.6745    0.7586    3.53  0.00042 ***
9 1:Age              -0.0314    0.1447   -0.22  0.82845
10 2:Age              0.1578    0.2901    0.54  0.58639
11 1:Education.level -0.9066    0.1741   -5.21  1.9e-07 ***
12 2:Education.level  0.0435    0.3943    0.11  0.91216
13 1:Income           0.0722    0.1376    0.52  0.59984
14 2:Income          -0.1513    0.2808   -0.54  0.59004
15 1:Awareness_Attitude 0.7808    0.1402    5.57  2.5e-08 ***
16 2:Awareness_Attitude 0.1072    0.2837    0.38  0.70557

```

```

17 1:Law_Regulations      -0.1790      0.0977      -1.83      0.06686 .
18 2:Law_Regulations      0.2455      0.1656      1.48      0.13810
19 1:Economic_Factors     -0.4582      0.1429      -3.21      0.00134 **
20 2:Economic_Factors     -0.1524      0.2581      -0.59      0.55501
21 1:Inconveniencies      -0.5270      0.1167      -4.52      6.3e-06 ***
22 2:Inconveniencies      -0.6086      0.2242      -2.71      0.00663 **
23 ---
24 Signif. codes:  0      ***      0.001      **      0.01      *      0.05      .
                0.1      1
25
26 Residual Deviance: 1545.765
27 Log-likelihood: -772.8824
28 AIC: 1577.765
29
30
31 Type of estimator: AS_mixed (mixed bias-reducing adjusted score
    equations)
32 Number of Fisher Scoring iterations: 21

```

Kết quả cho thấy:

- Hệ số chặn: Intercept 1 (ngưỡng (cut-point) đầu tiên, phân tách giữa nhóm 1 và nhóm 2- 3) và Intercept 2 (ngưỡng (cut-point) đầu tiên, phân tách giữa nhóm 1-2 và nhóm 3) có ý nghĩa thống kê.
- Sự bất tiện (Inconveniencies) càng cao thì sẽ càng ảnh hưởng đến mong muốn tham gia ($p < 0.05$).
- Thái độ và ý thức (Awareness Attitude) càng cao thì càng ảnh hưởng đến người có ý định trung lập ($p < 0.001$), nhưng không có ảnh hưởng đến nhóm người có ý định cao hơn.
- Trình độ học vấn càng cao và các yếu tố kinh tế càng chi phối nhiều (education level, Economic Factors) thì càng ảnh hưởng đến nhóm người có ý định trung lập ($p < 0.05$).

3.6 So sánh giữa model stepwise và model điều chỉnh tham số

Mô hình	AIC	Log-Likelihood
modp_step	1568.057	-778.0283 (df = 6)
modp_firth	1577.765	-772.8824 (df = 16)

Bảng 3: So sánh hai mô hình hồi quy

Khi so sánh giữa mô hình stepwise và mô hình đã điều chỉnh tham số, kết quả cho thấy mô hình stepwise có AIC thấp hơn ($AIC_{stepwise} = 1568.057 < AIC_{firth} = 1577.765$) và số tham số thấp hơn ($df_{stepwise} = 6 < df_{firth} = 16$). Điều này cho thấy mô hình stepwise đơn giản, ít tham số hơn. Tuy nhiên, khi xét Log-Likelihood, mô hình điều chỉnh tham số lại cho thấy phù hợp hơn với dữ liệu với giá trị Log-likelihood lớn hơn (Bảng 3).

4 Kết luận

Kết quả cho thấy, việc giảm thiên lệch bằng tham số có thể mang đến một mô hình phù hợp, tuy nhiên mô hình sẽ phức tạp hơn so với mô hình stepwise. Mô hình stepwise lại tốt hơn so với mô hình tuyến tính ban đầu.

Tài liệu

- [1] DAVID FIRTH. Bias reduction of maximum likelihood estimates. *Biometrika*, 1993.
- [2] David Firth Ioannis Kosmidis. Bias reduction in generalized nonlinear models. https://www.ikosmidis.com/files/ikosmidis_JSM2009.pdf.
- [3] I. Kosmidis and D. Firth. Multinomial logit bias reduction via the poisson log-linear model. *Biometrika*, 2011.