USING MULTINOMIAL LOGIT MODEL FOR ANALYSIS OF EDUCATIONAL SURVEY DATA

Wen Cheng, California Polytechnic University of Pomona

Hairui Tang, California Polytechnic University of Pomona

What Is Multinomial Logit Model?

- Establish a model with multiple level's output variables, and these predictive variables become a final predictive variable through a linear combination.
- Multinomial logit model in the dependent variable can take categorical values.
- An example is the film classification, which is "interesting", "so-so" or "boring".

Akaike's Information Criterion (AIC)

- Akaike's information criterion (AIC) proposed by Akaike is widely used for selecting the best model among the candidate models.
- The multinomial Logit model based on minimizing AIC to select the optimization model, rather than an ordinary linear model regarding the R square.
- A smaller AIC value indicates that the error of the competition model is lower and more reliable.

Data Description

TABLE 1. Multiple Student Traits Questions

Student Trait ID	Description of Each Trait
Q1	What involvement did you have in student clubs or organizations during your time at Cal Poly Pomona?
Q2	I attended one or more Pacific Southwest Regional ASCE Student Conferences
Q3	I attended one or more student-led professional conferences or seminars
Q4	I attended one or more professional society chapter meetings in the local area
Q5	I attended one or more non-student led professional conferences or seminars
Q6	I presented one or more papers or posters at a non-student led professional conferences or seminars
Q7	Full-time summer engineering job or internship
Q8	Part-time engineering job or internship during school year
Q9	Part-time non-engineering job during school year
Q10	Student research program at Cal Poly Pomona
Q11	Student research program at another university
Q12	During the academic year, on average how many hours a week were you employed?

TABLE 2. Related Abilities

Student Outcomes ID	Description of Student Learning Outcomes
a	Ability to apply knowledge of mathematics, science, and engineering.
b	Ability to design and conduct civil engineering experiments, as well as to analyze and interpret data.
c	Ability to design a system, component, or process to meet desired needs within realistic constraints such as economic, environmental, social, political, ethical, health and safety, manufacturability, and sustainability.
d	Ability to function on multidisciplinary teams.
e	Ability to identify, formulate, and solve engineering problems.
f	Understanding of professional and ethical responsibility.
g	Ability to communicate effectively.
h	Understanding of the impact of engineering solutions in a global, economic, environmental, and societal context.
i	Recognition of the need for, and an ability to engage in life-long learning.
j	Knowledge of contemporary issues and their importance to engineering systems.
k	Ability to use the techniques, skills, and modern engineering tools necessary for engineering practice.

TABLE 3. Descriptive Statistics of Responses to Self-Evaluation of Student Outcomes

	Count of Responces						Total Response Count
Student Outcomes	Excellent (5)	Good (4)	Average (3)	Fair (2)	Poor (1)	Rating Average *	
(a)	232	238	31	4	2	4.37	507
(b)	173	262	63	7	2	4.18	507
(c)	143	237	113	13	1	4.00	507
(d)	271	199	31	5	1	4.45	507
(e)	217	243	41	5	1	4.32	507
(f)	320	167	14	5	1	4.57	507
(g)	249	207	40	10	1	4.36	507
(h)	228	228	43	7	1	4.33	507
(i)	302	174	28	2	1	4.53	507
(j)	184	243	69	9	2	4.16	507
(k)	197	260	43	7	0	4.27	507

Data Description

- The survey was collected from a group of CPP graduating seniors of civil engineering, including two parts.
- The first part is the experience and performance of students during their school years as shown in Table 1. There are 12 questions which are considered as input data in the multinomial logit model.
- The second part represents the student's learning outcomes and covers 11 different abilities detailly presented in Table 2.
- Table 3 summarizes the frequency of responses for each outcome student, which will be used to develop the initial multinomial logit model.

Methodology of Multinomial Logit Model

- The probability of each type of outcome for a student can be recorded as:

$$P_n(i) = P(U_{in} \geq U_{jn}) \; \forall j \neq i \qquad (1)$$

- To estimate this possibility, U_{in}'s linear function can be expressed as:

$$U_{in} = \beta_i X_n + \varepsilon_{in} \qquad (2)$$

- Based on formular (1) and (2), the following equation can be written:

$$Pn(i) = P(\beta i Xn - \beta j Xn \geq \varepsilon jn - \varepsilon in) \; \forall j \neq i \qquad (3)$$

- Use generalized extreme value (GEV) distribution to produce a closed form model that can be readily estimated using standard maximum likelihood methods by using following equation:

$$P_n(i) = \exp(\beta_i X_n) / \sum_j \exp(\beta_i X_n) \qquad (4)$$

Methodology of Calculating AIC

- AIC is calculated by using the number of fitted parameters, including the intercept, in the model (k), and either the maximum likelihood estimate for the model (L) or the residual sum of squares of the model (RSS), two measures that are also easily derived from the output of any statistics package. In the case of least-squares regression analyses, the value of k must be increased by 1 to reflect the variance estimate as an extra model parameter. AIC is calculated as:

$$AIC = -2\ln(L)+2k \quad (5)$$

if using likelihood or

$$AIC= n[\ln(RSS /n)]+2k \quad (6)$$

Result

- This study selects five abilities that the students showed the lowest outcome grades overall to analyze related influential factors:
- b. Ability to design and conduct civil engineering experiments, as well as to analyze and interpret data.
- c. Ability to design a system, component, or process to meet desired needs within realistic constraints such as economic, environmental, social, political, ethical, health and safety, manufacturability, and sustainability.
- e. Ability to identify, formulate, and solve engineering problems.
- j. Recognition of the need for, and an ability to engage in life-long learning.
- k. Ability to use the techniques, skills, and modern engineering tools necessary for engineering practice.

Table 4 Influential Factors of Learning Outcome: b

Learning Out come	1	2	3	4	5	AIC Value without This Variable
Trait	Coefficients					
Q2	-	4.2867	4.2777	4.4024	5.0328	1067.520
Q7	-	5.9172	4.5564	5.4245	5.2074	1064.779
Q12	-	6.8721	6.3969	6.1953	6.2154	1067.520
Intercept	-	38.1320	38.0711	38.0669	38.0679	-
Final AIC	1063.689					

Table 5 Influential Factors of Learning Outcome: c

Learning Out come	1	2	3	4	5	AIC Value without This Variable
Trait	Coefficients					
Q4	-	85.52580	85.81096	86.70278	86.71185	1171.272
Q6	-	91.85218	92.29134	92.22124	93.06026	1165.882
Q8	-	-56.82973	-56.69983	-56.37731	-56.04308	1164.828
Q9	-	-50.54482	-50.34076	-50.72751	-50.86757	1162.804
Q12	-	-37.06390	-36.80031	-36.93123	-36.84164	1163.498
Intercept	-	196.7585	196.8121	196.8990	194.7223	-
Final AIC	1160.787					

Table 6 Influential Factors of Learning Outcome: e

Learning Out come	1	2	3	4	5	AIC Value without This Variable
Trait	Coefficients					
Q4	-	4.2867	4.2777	4.4024	5.0328	995.0973
Q7	-	5.9172	4.5564	5.4245	5.2074	986.2921
Intercept	-	9.4364	9.2790	9.2542	9.2553	-
Final AIC	986.1571					

Table 7 Influential Factors of Learning Outcome: j

Learning Out come	1	2	3	4	5	AIC Value without This Variable
Trait	Coefficients					
Q4	-	4.6952	5.7684	6.0314	6.5646	1103.903
Q10	-	18.6095	17.5167	18.7988	19.1424	1100.062
Intercept		10.8772	10.8088	10.7979	10.7989	-
Final AIC	1097.425					

Table 8 Influential Factors of Learning Outcome: k

Learning Out come	1	2	3	4	5	AIC Value without This Variable
Trait	Coefficients					
Q2	-	-	0.9275	0.8812	0.8823	993.3849
Q4	-	-	1.1873	1.1295	1.1309	992.706
Intercept	-	-	0.9879	2.6391	1.3211	-
Final AIC	989.5527					

Conclusion

- The five abilities chose to study are relatively lacking for Cal Poly Pomona students. Modeling through the Multinomial logistic regression and using AIC to gradually select the variables contained in the model, the result shows influential factors of each learning outcome. From these factors, Cal Poly Pomona can find the mean point to adjust teaching focus that further improve the quality of education. These results are expected to take additional insights for department managers and administrators as they design courses that meet students' needs. As well enterprises can also refer to these factors of generally low ability before interviewing students, so that they can have a better understanding of the focus of recruitment.

- In the field of statistical analysis, for the multiple categories data, multinomial logit model has been more and more popular. Because it does not assume normality, linearity, or homoscedasticity, and more data types are available. For this study, the deficiency is the accuracy of the original data, because this is a student self-evaluation survey, so it is hard to guarantee the objectivity and impartiality of the data. Some students may overestimate their abilities, while others may underestimate their abilities. If students' self-evaluation can be combined with teachers' evaluation of students, it is anticipated to yield more reliable findings.

Thank you