Multinomial Logistic Regression

Andrea Arias Ll. Luis Sandoval-Mejía Yangmei (Emily) Wang

Texas Tech University
ISQS 5349: Regression Analysis

April 28, 2016

Agenda

Introduction

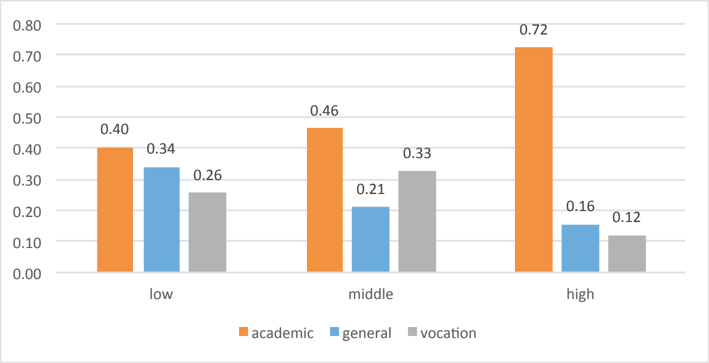
Let's consider a data set

- A data set with n observations where the response variable can take one of several discrete values $(1,2,\ldots,J)$
- Let Y be program type[1]:
  - General
  - Academic
  - Vocation

[1] IDRE, 2016

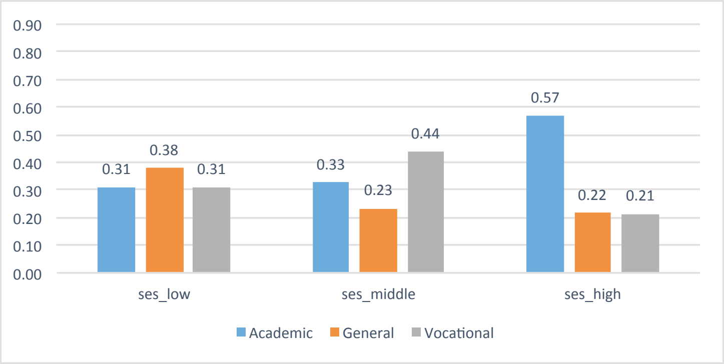
Some data analysis

Probability of choosing one program by SES in general



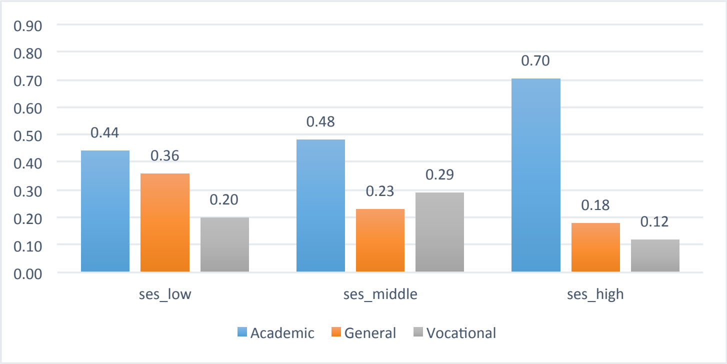
Some data analysis (cont.)

Probability of choosing one program by SES at the 1^{st} quartile of writing score



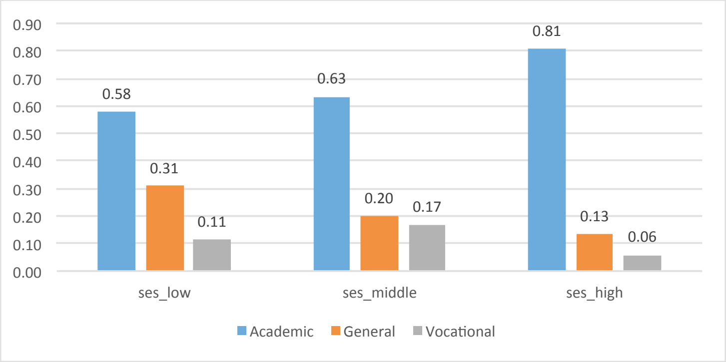
Some data analysis (cont.)

Probability of choosing one program by SES at the mean writing score



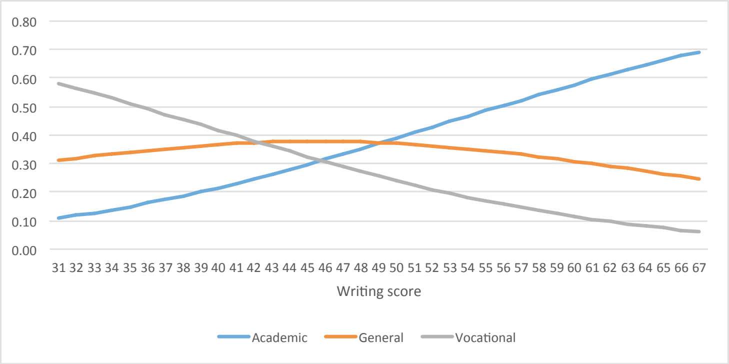
Some data analysis (cont.)

Probability of choosing one program by SES at the 3^{rd} quartile of writing score



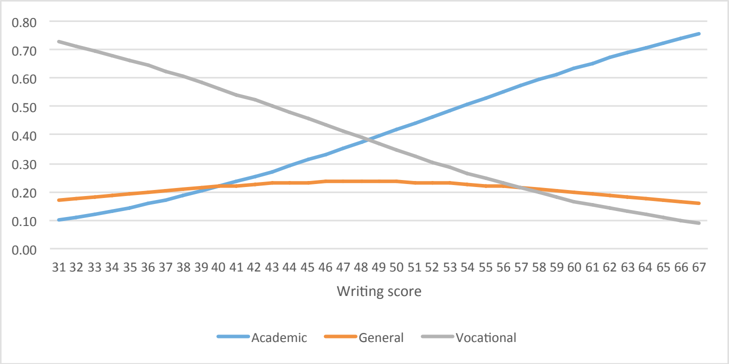
Some data analysis (cont.)

Probability of choosing one program by writing score at SES low



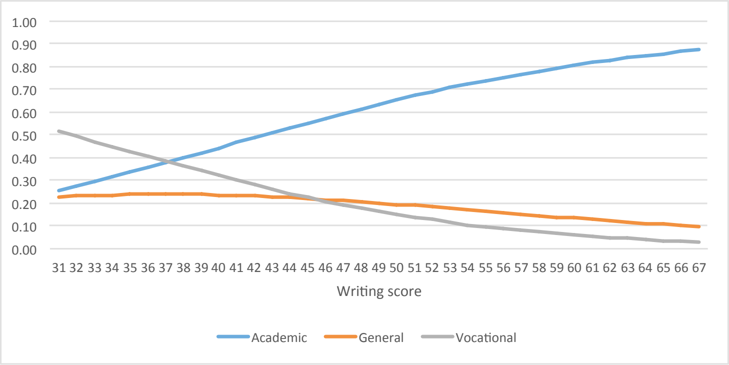
Some data analysis (cont.)

Probability of choosing one program by writing score at SES middle



Some data analysis (cont.)

Probability of choosing one program by writing score at SES high



When do we use Multinomial Logit?

- When we need to model nominal (unordered choices) outcome variables!
  - The individual is choosing between more than two alternatives.
    - Undergraduate major choice
    - Food choices
    - Car manufacturer choice
- **Important!!** The multinomial logit model applies only when the predictor variables are **individual specific**.
  - Income.
  - Age.
  - Social economic status.
  - ... and so on.

The Multinomial Distribution

- Consider a random variable Y_i that may take one of several discrete values (1,2,...,J)
- Let:

$$Pr\{Y_i = j\} = \pi_{ij} \quad i = 1,..,n \quad j = 1,..,J$$

Where π_{ij} is the probability that the i-th individual choose the j-th category[2].

- And

$$\sum_{j=1}^{J} \pi_{ij} = 1 \quad i = 1,..,n$$

[2] Rodriguez,2016

The Multinomial Logit Model

- A model for the probabilities where the probabilities depend on a vector X_i.
- Nominate one of the response categories as baseline.
- Calculate the logits for all other categories. **with respect to the baseline**.
- Let the logits be a **linear function** of the predictors.

The Multinomial Logit model (cont.)

$$\log\left(\frac{\pi_{ij}}{\pi_{iJ}}\right) = X_i'\beta_j \quad j = 1, .., J-1$$

The Multinomial Logit model (cont.)

- Modeling the probabilities[3]:

$$Prob(Y_i = j|X_i) = \pi_{ij} = \frac{exp(X_i^{'}\beta_j)}{1+\sum_{j=1}^{J-1} exp(X_i^{'}\beta_j)}, \quad j = 1, 2, ..., J-1$$

$$Prob(Y_i = J|X_i) = \pi_{iJ} = \frac{1}{1+\sum_{j=1}^{J-1} exp(X_i^{'}\beta_j)},$$

- **Note:** The binomial logit model is a special case when $J = 2$.

[3] Greene, 2012

Model assumptions

- Linearity
- Independence of observations
- Independence of irrelevant alternatives
  (Not as important as the first two!!)

Maximum Likelihood Estimation

- Recall that the Likelihood function for a parameter vector θ resulting from an **independent** sample is[4]:

$$L(\theta|y_1, y_2, ..., y_n) = p(y_1|\theta) \times p(y_2|\theta) \times ... \times p(y_n|\theta)$$

- In our case, since for observation i the probability is given by:

$$Pr\{Y_i = j\} = \pi_{ij}$$

- We have that the Likelihood function is the following:

$$L(\pi|data) = \prod_{i \in choice1} \pi_{i1} \prod_{i \in choice2} \pi_{i2} ... \prod_{i \in choiceJ} \pi_{iJ}$$

- We choose the $\hat{\beta}$ that maximizes

$$L(\pi|data)$$

[4] Westfall and Henning, 2013

Estimated probabilities

- Once we have estimated the parameters $(\hat{\beta})$, we can estimate the probabilities for each particular **cohort**.

$$
\begin{aligned}
Pr\left(Y=O_0|X_1=x_1,X_2=x_2\right) &= \frac{1}{1+e^{(\hat{\beta}_{10}+\hat{\beta}_{11}x_1+\hat{\beta}_{12}x_2)}+e^{(\hat{\beta}_{20}+\hat{\beta}_{21}x_1+\hat{\beta}_{22}x_2)}} \\
Pr\left(Y=O_1|X_1=x_1,X_2=x_2\right) &= \frac{e^{(\hat{\beta}_{10}+\hat{\beta}_{11}x_1+\hat{\beta}_{12}x_2)}}{1+e^{(\hat{\beta}_{10}+\hat{\beta}_{11}x_1+\hat{\beta}_{12}x_2)}+e^{(\hat{\beta}_{20}+\hat{\beta}_{21}x_1+\hat{\beta}_{22}x_2)}} \\
Pr\left(Y=O_2|X_1=x_1,X_2=x_2\right) &= \frac{e^{(\hat{\beta}_{20}+\hat{\beta}_{21}x_1+\hat{\beta}_{22}x_2)}}{1+e^{(\hat{\beta}_{10}+\hat{\beta}_{11}x_1+\hat{\beta}_{12}x_2)}+e^{(\hat{\beta}_{20}+\hat{\beta}_{21}x_1+\hat{\beta}_{22}x_2)}}
\end{aligned}
$$

Simulation
Accounting example

- **Response variable:** Tax service provider.
  Three nominal choices:
  - B - Big4
  - N - Non Big4
  - S - Self preparer

- **Predictor variable:** Company's tax preparation budget.
  (lbudget)

Simulation
Accounting example

- The Model:

$$\log\left(\frac{Pr(Y=N|lbudget)}{Pr(Y=B|lbudget)}\right) = \beta_{01}+\beta_{11}lbudget$$
$$\log\left(\frac{Pr(Y=S|lbudget)}{Pr(Y=B|lbudget)}\right) = \beta_{02}+\beta_{12}lbudget$$

- The probabilities of choosing each of the three choices:

$$Pr\left(Y=B|lbudget\right) = \frac{1}{1+e^{(\beta_{01}+\beta_{11}lbudget)}+e^{(\beta_{02}+\beta_{12}lbudget)}}$$
$$Pr\left(Y=N|lbudget\right) = \frac{e^{(\beta_{01}+\beta_{11}lbudget)}}{1+e^{(\beta_{01}+\beta_{11}lbudget)}+e^{(\beta_{02}+\beta_{12}lbudget)}}$$
$$Pr\left(Y=S|lbudget\right) = \frac{e^{(\beta_{02}+\beta_{12}lbudget)}}{1+e^{(\beta_{01}+\beta_{11}lbudget)}+e^{(\beta_{02}+\beta_{12}lbudget)}}$$

References

- McKee, D. 2015. An intuitive introduction to the multinomial logit. Available at: `https://www.youtube.com/watch?v=n7zHXjuE6PE&t=2732s`
- Institute for Digital Research and Education, UCLA. R Data analysis examples: Multinomial Logistic Regression. Available at: `http://www.ats.ucla.edu/stat/r/dae/mlogit.htm`
- Greene, W, H. 2012. Econometric Analysis, 7th Edition. Prentice Hall, NJ.
- Rodriguez, G. 2016. A note on interpreting multinomial logit coefficients. Princeton University. Available at: `http://data.princeton.edu/wws509/stata/mlogit.html`
- Westfall, P., Henning, K. S. (2013). Understanding advanced statistical methods. CRC Press.

THANK YOU !!!

Questions?