

WWS 509 Generalized Linear Models: Precept 8

Multinomial and Sequential Logit Models

Kristin E. Bietsch

Office of Population Research, Princeton University

November 2012

Introducing the Data

This data comes from Germán's website. We are looking at how people rate their level of satisfaction in their housing. The options are high, medium, and low. Since there are more than 2 options, we can use multinomial logistic regression to predict responses. Controls we will consider include contact with neighbors (high or low), type of housing (tower, apartment, atrium, or terrace), and influence (on a scale of 0-2).

Multinomial Logit Model

- There are $J-1$ equations. One of the outcomes will serve as the baseline.
- Like the logit model, the log-odds is a linear function of the predictor.
- It makes no difference which category we pick as the reference cell, because we can convert from one formulation to another
- $\log\left(\frac{\pi_{i1}}{\pi_{i2}}\right) = \log\left(\frac{\pi_{i1}}{\pi_{i3}}\right) - \log\left(\frac{\pi_{i2}}{\pi_{i3}}\right)$

Sequential Logit Model

- Also known as the hierarchical logit model
- We are now looking at nested comparisons. For example, I have to be ok with my housing not to rate my satisfaction level as “low,” and then I choose between medium or high.
 - Can you think of other data which might benefit from this model?
 - * **Marital status: single, monogamous, polygamous**

Interpretation

General

1. Why did I change contact and influence with “-1”?
 - **To make the constant interpretable. Before, neither of them had 0 in their range, now they do.**
2. Which housing type is the reference category? Where besides the logit output can you find this?
 - **When I make the local, I am excluding tower.**

Multinomial Logit

1. Looking at the model only including housing type, what seems to be the best housing? The worst?
 - **People seem to be very happy if they live in a tower, and not satisfied if they live a terrace apartment.**
2. Does this first model fit the data?
 - **No, the Likelihood-ratio test is highly significant.**
3. Is influence, net of housing, predictive of satisfaction?
 - **Yes**
4. What assumption am I making my modeling influence as continuous and not categorical?
 - **That it is linear, 0-1 is that same as 1-2.**
5. What is the odds ratio for a person with an influence level of 2 of having a satisfaction level of medium compared to low in this model? Of high compared to low?
 - **Odds ratio for medium to low:** $e^{(2 \cdot .3207971)} = 1.89950666$
 - **Odds ratio for high to low:** $e^{(2 \cdot .7736902)} = 4.699144167$
6. What about in the model that also controls for neighbor contact?
 - **Odds ratio for medium to low:** $e^{(2 \cdot .3487541)} = 2.008741085$
 - **Odds ratio for high to low:** $e^{(2 \cdot .8103557)} = 5.056686364$
7. What about contact with neighbors?
 - **Odds ratio for medium to low:** $e^{(.3635188)} = 1.438381898$
 - **Odds ratio for high to low:** $e^{(.4794532)} = 1.61590974$
8. Does the model with neighbors fit the data? How do you know?
 - **Yes, the likelihood-ratio test is not significant.**

Sequential Logit

1. Do you see anything interesting here that we did not see in the multinomial model?
 - **Neighbor contact does not matter in choosing between medium and high satisfaction.**
2. Which model do you prefer for this data: multinomial or sequential?
 - **Sequential makes more sense to me in terms of decision making.**