Qualitative Choice Models

Jerome Dumortier

06 March 2024

Overview

Packages and Files

Required packages:

- AER
- erer
- glmmML
- MASS
- mlogit
- nnet

Required files:

```
data("Fishing",package="mlogit")
data("TravelMode",package="AER")
```

Types of Categorical Dependent Variables

Binary

- Outcome variable is either 0 or 1

Categorical outcome and naturally ordered

- Happiness of a person: Very happy, happy, okay, or sad

Categorical outcome and not ordered

- Commute to campus: Bus, bike, walk, or car

Ordered Logit Model

Theoretical Aspects I

Assumption of a latent variable y^* that is unobserved by the researcher:

$$y_i^* = \beta_0 + \beta_1 \cdot x_i + \epsilon_i$$

In the case of a happiness model, this may be a measure of "happiness." What the researcher does measure is an m-alternative ordered model:

$$y_i = j \quad \text{if} \quad \alpha_{j-1} < y_i^* \leq \alpha_j \quad \text{for} \quad j = 1, \ldots, m$$

where $\alpha_0 = -\infty$ and $\alpha_m = \infty$.

Theoretical Aspects II

In this case, we have

$$\begin{aligned} Pr(y_i = j) &= Pr(\alpha_{j-1} < y_i^* \leq \alpha_j) \\ &= Pr(\alpha_{j-1} < \beta_0 + \beta_1 \cdot x_i + \epsilon_i \leq \alpha_j) \\ &= Pr(\alpha_{j-1} - \beta_0 - \beta_1 \cdot x_i < \epsilon_i \leq \alpha_j - \beta_0 - \beta_1 \cdot x_i) \\ &= F(\alpha_j - \beta_0 - \beta_1 \cdot x_i) - F(\alpha_{j-1} - \beta_0 - \beta_1 \cdot x_i) \end{aligned}$$

Ordered logit:

$$F(z) = \frac{e^z}{1 + e^z}$$

Example using `fpdata`

Survey on the purchase frequency of organic tomatoes and organic strawberries:

- Never (1), rarely (2), once per month (3), every 2 weeks (4), 1-2 times a week (5), almost daily (6)

Independent variables are

- Age and female
- Education: High school (1), some college (2), bachelor (3), master (4), technical school diploma (5), doctorate (6)

Setup and Calculation in R

```
fpdata$strawberriesorg    = as.factor(fpdata$strawberriesorg)
strawdata                 = fpdata[c("strawberriesorg","age",
                                   "education","female",
                                   "kidsunder12")]
strawdata                 = na.omit(strawdata)
bhat                      = polr(strawberriesorg~age+education+
                                    female+kidsunder12,
                                data=strawdata,Hess=TRUE)
```

Results

```
## Call:
## polr(formula = strawberriesorg ~ age + education + female + kidsunder12,
##     data = strawdata, Hess = TRUE)
##
## Coefficients:
##                Value Std. Error t value
## age         -0.02034   0.009838 -2.0676
## education    0.01596   0.112028  0.1425
## female      -0.41533   0.280485 -1.4808
## kidsunder12  0.28560   0.321778  0.8876
##
## Intercepts:
##     Value   Std. Error t value
## 0|1 -1.4958  0.6497    -2.3022
## 1|2 -0.4381  0.6434    -0.6810
## 2|3  0.2084  0.6394     0.3259
## 3|4  0.8352  0.6442     1.2964
## 4|5  1.6314  0.6699     2.4353
##
## Residual Deviance: 526.4547
## AIC: 544.4547
```

Interpretation

For the organic purchases data, the cuts are under "Intercepts" and thus, we have (rounded coefficients):

$$z = -0.020 \cdot age + 0.0160 \cdot education - 0.415 \cdot female + 0.286 \cdot kidsunder12$$

The cutoff points can be interpreted as follows:

$$\begin{aligned}
Pr(y = 1) &= P(z + \epsilon_i \leq -1.4958) \\
Pr(y = 2) &= P(-1.4958 < z + \epsilon_i \leq -0.4381) \\
Pr(y = 3) &= P(-0.4381 < z + \epsilon_i \leq 0.2084) \\
Pr(y = 4) &= P(0.2084 < z + \epsilon_i \leq 0.8352) \\
Pr(y = 4) &= P(0.8352 < z + \epsilon_i \leq 1.6314) \\
Pr(y = 6) &= P(1.6314 \leq z + \epsilon_i)
\end{aligned}$$

Predicted Probabilities and Marginal Effects

Predicted probabilities:

```
bhat.pred = predict(bhat,type="probs")
```

Marginal Effects:

```
x = ocME(bhat)
x$out$ME.all
```

```
##              effect.0 effect.1 effect.2 effect.3 effect.4 effect.5
## age             0.005    0.000   -0.001   -0.001   -0.001   -0.001
## education      -0.004    0.000    0.001    0.001    0.001    0.001
## female          0.098   -0.003   -0.023   -0.024   -0.023   -0.026
## kidsunder12    -0.067    0.001    0.015    0.017    0.016    0.018
```

Multinomial Logit Model

Data

Data format:

- Categorical data with no natural ordering
- Choice of concentrations at the O'Neill School: Environmental Policy and Sustainability, Homeland Security and Emergency Management, Innovation and Social Change, Nonprofit Management, Policy Analysis, Public Management, or Urban and Regional Governance.

Great resource on all aspects related to categorical models: mlogit

Models and Considerations

Various models

- Multinomial logit: Only individual-specific variables
- Conditional logit: Only alternative-specific variables
- Mixed logit: Individual- and alternative-specific variables

Independence of irrelevant alternatives (IIA)

- Example: Pie, ice cream, and cheese cake for dessert

Applications

Revealed preferences:

- Observed choices of individuals

Stated preference

- Hypothetical choice situations

Economists' modelling of choice

- Utility/happiness/satisfaction associated with multiple choice situations

Theoretical Aspects

Travel choice model dependent on cost (x) and time (z):

$$V_j = \alpha_j + \beta_1 \cdot x_j + \beta_2 \cdot z_j$$

Probability of choosing alternative j (assuming three choices)

$$P(1) = \frac{e^{V_1}}{e^{V_1} + e^{V_2} + e^{V_3}}$$
$$P(2) = \frac{e^{V_2}}{e^{V_1} + e^{V_2} + e^{V_3}}$$
$$P(3) = \frac{e^{V_3}}{e^{V_1} + e^{V_2} + e^{V_3}}$$

Note that $P(1) + P(2) + P(3) = 1$

Data Management I

Long versus wide data

- Long: One row for each alternative
- Wide: One row for each choice situation

Wide format: `Fishing` (Description)

- Fishing modes: beach, pier, private, and charter
- Alternative-specific variables: price and catch
- Individual-specific variables: income
- Suitability of the “wide” format to store individual-specific variables

Long format: `TravelMode` (Description)

Data Management for Fishing and TravelMode

Fishing (wide format)

```
Fishing = mlogit.data(Fishing,shape="wide",varying=2:9,
                      choice="mode")
```

- Designation of alternative specific variables with "varying"

Travel Mode (long format)

```
TravelMode1 = mlogit.data(TravelMode,choice="choice",shape="long",
                          alt.levels=c("air","train","bus","car"))
TravelMode2 = mlogit.data(TravelMode,choice="choice",shape="long",
                          alt.var="mode")
```

Error message when attempting to visualize the data.

Estimation

Individual-specific independent variables only

```
bhat           = mlogit(mode~0|income,Fishing)
summary(bhat)
fishing.fitted = fitted(bhat,outcome=FALSE)
effects(bhat,covariate="income")
```

Estimation

Individual- and alternative-specific independent variables

```
bhat           = mlogit(mode~catch+price|income,
                        data=Fishing)
summary(bhat)
fishing.fitted = fitted(bhat,outcome=FALSE)
effects(bhat,covariate="income")
```

Travel Mode I

```
bhat1 = mlogit(choice~gcost+wait|income+size,
               data=TravelMode1,reflevel="car")
summary(bhat1)
```

Travel Mode II

```
bhat2 = mlogit(choice~gcost+wait|income+size,
               data=TravelMode2,reflevel="car")
summary(bhat2)
```

Electric Vehicle Data

```
evdata          = mlogit.data(evdata,shape="wide",
                              choice="choice")
bhat            = mlogit(choice~0|age+female+level2+
                         numcars+edu+income+politics,
                         data=evdata)
summary(bhat)
evdata.fitted   = fitted(bhat,outcome=FALSE)
```