

Mikroekonometria

9

Mikołaj Czajkowski
Wiktor Budziński

Wielomianowy model logitowy

- ▶ Uogólnienie modelu binarnego
- ▶ Wybór pomiędzy 2 lub większą liczbą alternatyw
 - ▶ Np. wybór środka transportu, głos w wyborach, wybór marki produktu
- ▶ Podobnie jak dla modelu binarnego dane mogą być w dwóch formatach
 - ▶ Format krótki
 - ▶ Po jednym wierszu dla respondenta
 - ▶ Zmienna objaśniana przyjmuje wartości od 1 do liczby alternatyw
 - ▶ Obserwujemy tylko atrybuty wybranej alternatywy
 - ▶ Format długi
 - ▶ Dla każdego respondenta po kilka wierszy (tyle ile alternatyw w wyborze)
 - ▶ Zmienna objaśniana przyjmuje wartości 0 albo 1, 1 tylko dla wybranej alternatywy
 - ▶ Obserwujemy atrybuty wszystkich alternatyw

Wielomianowy model logitowy

- ▶ W zależności od dostępności danych wyróżnia się dwa typy modelu
 - ▶ Format krótki – wielomianowy model logitowy
 - ▶ Zmiennymi objaśniającymi są zwykle charakterystyki konsumentów (wchodzą do modelu jako interakcje z ASC)
 - ▶ Format długi – wielomianowy (warunkowy) model logitowy
 - ▶ Zmiennymi objaśniającymi mogą być zarówno atrybuty jak i charakterystyki konsumentów
 - Zmienne specyficzne dla respondenta muszą, jak zwykle, wchodzić do modelu jako interakcje z ASC lub atrybutami
- ▶ W praktyce to dość sztuczny podział (wynika z formatu danych, a nie postaci modelu). W literaturze używa się nazwy wielomianowy model logitowy dla obu przypadków

Wielomianowy model logitowy

- ▶ Użyteczność konsumenta i z wyboru alternatywy j spośród J_j alternatyw

$$U_{ij} = \mathbf{X}_{ij}\boldsymbol{\beta} + \mathbf{Z}_i\boldsymbol{\gamma}_j + \varepsilon_{ij}$$

- ▶ \mathbf{X} – wektor cech (*atrybutów*) danej alternatywy
- ▶ \mathbf{Z} – wektor cech (*charakterystyk*) konsumenta
- ▶ ε – składnik losowy (np. nieobserwowalne skłonności konsumenta)
- ▶ Parametry atrybutów zwykle są generyczne
 - ▶ Takie same dla każdej alternatywy, bez indeksu j
- ▶ Przykładowo, dla wyboru środka transportu:
 - ▶ \mathbf{X}_{ij} będzie zawierało takie zmienne jak czas i koszt przejazdu (atrybuty). Mogą przyjmować różne wartości dla różnych alternatyw. Możliwe tylko przy długim formacie danych.
 - ▶ \mathbf{Z}_i będzie zawierało takie zmienne jak wiek i dochód. Takie same dla różnych alternatyw (ale różne dla różnych respondentów)

Wielomianowy model logitowy

- ▶ Zmienne Z_i muszą mieć parametry specyficzne dla alternatywy
 - ▶ Liczą się tylko różnice w poziomach użyteczności – jeśli do użyteczności dla wszystkich alternatyw dodamy 10, to użyteczności z alternatyw (a więc prawdopodobieństwa ich wyboru) się nie zmieniają
 - ▶ Parametry dla jednej alternatywy są ustalone na poziomie 0 – poziom referencyjny
 - ▶ Interpretacja parametrów względem poziomu referencyjnego
 - ▶ Z_i może zawierać stałą, parametry dla stałej nazywamy stałymi specyficznymi dla alternatywy (ASC – *alternative specific constants*)

Wielomianowy model logitowy

- ▶ ε_{ij} – niezależne składniki losowe o jednakowych rozkładach (ang. *independent, identically distributed*, IID)
 - ▶ Jeśli ε_{ij} mają rozkład wartości ekstremalnych I rodzaju to

$$P(Y_i = j) = \frac{\exp(\mathbf{x}_{ij}\boldsymbol{\beta} + \mathbf{z}_i\boldsymbol{\nu}_j)}{\sum_{k=1}^{J_i} \exp(\mathbf{x}_{ik}\boldsymbol{\beta} + \mathbf{z}_i\boldsymbol{\nu}_k)}$$

- ▶ Zmienna objaśniana przyjmuje wartość 1 dla wybranej alternatywy i 0 wpp (d_{ij})

$$\ln L = \sum_{i=1}^N \sum_{j=1}^{J_i} d_{ij} \log \frac{\exp(\mathbf{x}_{ij}\boldsymbol{\beta} + \mathbf{z}_i\boldsymbol{\nu}_j)}{\sum_{k=1}^{J_i} \exp(\mathbf{x}_{ik}\boldsymbol{\beta} + \mathbf{z}_i\boldsymbol{\nu}_k)}$$

- ▶ Gradient i Hesjan są stosunkowo proste – estymacja przebiega standardowo

Wielomianowy model logitowy

- ▶ Ponieważ składniki losowe IID, a parametry takie same dla każdego respondenta – panelowa struktura danych nie ma znaczenia (możemy ją na razie zignorować)
- ▶ Miary dopasowania – te same problemy co dla modeli wyboru binarnego
- ▶ Wyniki (parametry) wymagają uważnej interpretacji
 - ▶ Są relatywne względem poziomu referencyjnego
- ▶ Osadzenie w ramach modelu użyteczności losowej pozwala na wygodną interpretację miar zmian dobrobytu
 - ▶ Gotowość do zapłaty (ang. *willingness to pay*, WTP)
 - ▶ Np. krańcowe stopy substytucji, zmiana kompensująca

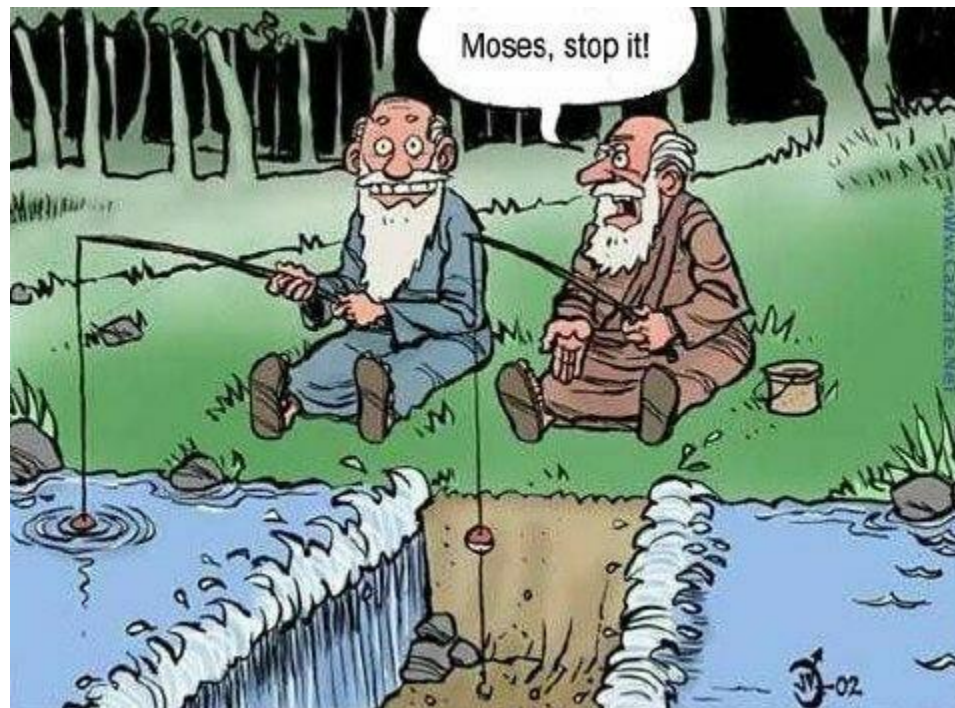
Zadanie 1. Wybór sposobu wędkowania

1. Wczytaj zbiór danych `me.fishmode1.dta`
2. Skonstruuj model wyjaśniający wybór sposobu wędkowania (z brzegu, z pomostu, z własnej łodzi, z wyczarterowanej łodzi) za pomocą charakterystyk respondentów (ASC i dochodu)
3. Zinterpretuj wyniki
4. Wczytaj zbiór danych `me.fishmode2.dta`
5. Skonstruuj model wyjaśniający wybór sposobu wędkowania za pomocą atrybutów wyboru (ASC, ceny czarteru i oczekiwanej wielkości połowu)



Przykład – wybór sposobu wędkowania

6. Uzupełnij powyższy model o charakterystyki respondentów
7. Zinterpretuj uzyskane wyniki



Efekty krańcowe

- ▶ Efekty krańcowe trochę bardziej skomplikowane niż w modelu binarnym
 - ▶ Zmiana poziomu atrybutu jednej alternatywy będzie miała wpływ na prawdopodobieństwa wyboru wszystkich alternatyw
- 8. Policz efekty krańcowy dla ceny czarteru i oczekiwanej wielkości połowu
- 9. Policz efekt krańcowy dla dochodu

Obserwowalna heterogeniczność preferencji

- ▶ Krańcowa użyteczność dla poszczególnych atrybutów może się różnić między respondentami
 - ▶ Można to uwzględnić w modelu wprowadzając interakcje między atrybutami a charakterystykami respondenta
 - ▶ Różni respondenci będą mieli różne gotowości do zapłaty
10. Sprawdź czy respondenci z różnym dochodem mają różne preferencje
 11. Policz WTP dla wszystkich kwartyli dochodów

Niezależność od niewybranych alternatyw

- ▶ Wielomianowy model logitowy jest prosty, ale narzuca dość restrykcyjne założenia
 - ▶ Składniki losowe IID
 - ▶ Użyteczność z wyboru różnych alternatyw nieskorelowana
 - ▶ Iloraz prawdopodobieństw wyboru 2 alternatyw nie zależy od atrybutów pozostałych alternatyw (ang. *independence from irrelevant alternatives*, IIA)
- ▶ Jeśli dwie alternatywy podobne, to spodziewamy się, że ich składniki losowe będą skorelowane
 - ▶ Przykład – wybór między autobusem, czerwonym samochodem i czarnym samochodem
 - ▶ Spodziewamy się, że możliwość wyboru czarnego samochodu nie zmieni znacząco wyboru autobusu, a zmniejszy częstość wyboru czerwonego samochodu
 - ▶ Założenia MNL oznaczają, że wpływ na wybór każdej alternatywy będzie jednakowy.
 - ▶ Np. jeśli wyrzucimy z próby czarny samochód to prawdopodobieństwa pozostałych alternatyw wzrosną proporcjonalnie, co jest mało realistyczne

Niezależność od niewybranych alternatyw

▶ Test IIA – test Hausmana i McFaddena

- ▶ Jeśli jakiś podzbiór alternatyw faktycznie nie ma znaczenia dla wyborów, to po ich usunięciu ze zbioru nie zmieni systematycznie oszacowań parametrów
 - ▶ Jeśli założenia spełnione – pomijanie alternatyw jest nieefektywne, ale nie powinno powodować niezgodności

▶ Statystyka testu

$$H = (\boldsymbol{\beta}_r - \boldsymbol{\beta}_f)' (\mathbf{V}_r - \mathbf{V}_f)^{-1} (\boldsymbol{\beta}_r - \boldsymbol{\beta}_f)$$

- ▶ r – estymator dla ograniczonego zbioru wyborów
- ▶ f – estymator dla pełnego zbioru wyborów
- ▶ Statystyka ma rozkład chi-kwadrat z liczbą stopni swobody równą liczbie parametrów

Niezależność od niewybranych alternatyw

12. Wykorzystaj `me.fishmode2.dta` do przetestowania właściwości IIA modelu MNL
 - ▶ Test może nie działać dla modeli ze stałymi specyficznymi dla alternatyw (porównywane modele muszą mieć tyle samo parametrów)
 - ▶ Kolejność ma znaczenie

Case study – segregacja odpadów

- ▶ **Badanie DCE z 2011 roku, mieszkańcy Podkowy Leśnej**
 - ▶ Kontekst – wprowadzenie ustawy śmieciowej
 - ▶ Jakie czynniki mają znaczenie dla wyboru optymalnego kontraktu wywozu odpadów?
 - ▶ Liczba kategorii sortowania
 - ▶ Częstotliwość odbioru
 - ▶ Koszt

Case study – segregacja odpadów

▶ Przykładowa sytuacja wyboru

Sytuacja 1.	Wariant 1	Wariant 2	Wariant 3
Sposób sortowania w gospodarstwie domowym	Brak	Brak	Na 5 kategorii
Częstotliwość odbioru	Raz na 1 tydzień	Raz na 4 tygodnie	Raz na 2 tygodnie
Miesięczny koszt dla Pana(i) gospodarstwa domowego	75 zł	50 zł	25 zł
Pana(i) wybór:	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

- ▶ 311 gospodarstw domowych
- ▶ 6 sytuacji wyboru na respondenta

Praca domowa ME.9 (grupy 2 lub 3-osobowe)

1. Wczytaj zbiór danych me.recycling.dta
2. Skonstruuj model wyjaśniający jakie czynniki mają znaczenie przy wyborze optymalnego kontraktu
 - ▶ Czy model powinien zawierać stałe specyficzne dla alternatywy?
 - ▶ Przetestuj czy założenie o stałych krańcowych użytecznościach atrybutów jest uzasadnione
 - ▶ Wprowadź nieliniowe przekształcenia do modelu dla atrybutów innych niż koszt
 - ▶ Przetestuj założenie o IIA
 - ▶ Oblicz gotowości do zapłaty za poszczególne poziomy atrybutów
 - ▶ Czy w danych występuje heterogeniczność preferencji? Jakie zmienne mogą pomóc ją wyjaśnić?

```
set seed 10+"Nr indeksu"  
gen seq = int((_n-1)/3) +1  
sample2 90, cluster(seq)
```

