

Эконометрика-2 ММАЭ

Семинар 19

Лекции: А.А. Пересецкий
Семинары: Е.С. Вакуленко

Multinomial models

Multinomial logit

Пусть есть $j = 1, \dots, m$ неупорядоченных альтернатив

"случайная полезность"

$$P(y = j | x) = P(u_j + \varepsilon_j > u_k + \varepsilon_k)$$

Цель: нужно знать плотность совместного распределения случайных ошибок ε , имеет простое аналитическое представление для распределения экстремальных значений:

$$F(z) = \exp(-e^{-z})$$

Тогда

$$P(y = j | x) = \frac{\exp(u_j)}{\exp(u_1) + \dots + \exp(u_m)}$$

Предполагаем

$$u_j = x' \beta_j$$

Нормировка (для идентифицируемости модели)

$$u_1 = x' \beta_1 = 0$$

Следовательно

$$P(y = 1 | x) = \frac{1}{1 + \exp(u_2) + \dots + \exp(u_m)}$$

$$P(y = j | x) = \frac{\exp(u_j)}{1 + \exp(u_2) + \dots + \exp(u_m)}$$

Problem 1. (Wooldridge – CrossSection&Panel).

The data in PENSION.dta are a subset of data used by Papke (1998) in assessing the impact of allowing individuals to choose their own allocations on asset allocation in pension plans. Initially, Papke codes the responses “mostly bonds,” “mixed,” and “mostly stocks” as 0, 50, and 100, and uses a linear regression model estimated by OLS. The binary explanatory variable choice is unity if the person has choice in how his or her pension fund is invested.

id – family identifier

pyears – years in pension plan

prftshr – =1 if profit sharing plan (employee receives a percentage of those profits based on the company's earnings)

choice – =1 if can choose method invest

female – =1 if female
 married – =1 if married
 age – age in years
 educ – highest grade completed
 finc25 – \$15,000 < faminc92 <= \$25,000
 finc35 – \$25,000 < faminc92 <= \$35,000
 finc50 – \$35,000 < faminc92 <= \$50,000
 finc75 – \$50,000 < faminc92 <= \$75,000
 finc100 – \$75,000 < faminc92 <= \$100,000
 finc101 – \$100,000 < faminc92
 wealth89 – net worth, 1989, \$1000
 black – =1 if black
 stckin89 – =1 if owned stock in 1989
 irain89 – =1 if had IRA in 1989 (individual retirement savings program)
 pctstck – 0=mstbnds, 50=mixed, 100=mststcks

OLS

reg pctstck choice age educ wealth89 prftshr

Controlling for age, education, gender, race, marital status, income (via a set of dummy variables), wealth, and whether the plan is profit sharing, gives the OLS estimate $\hat{\beta}_{choice} = 12.05$. This result means that, other things equal, a person having choice has about 12 percentage points more assets in stocks.

mlogit pctstck choice age educ wealth89 prftshr, baseoutcome(0)

Mlogit: the independence of irrelevant alternatives (IIA) property from rational choice theory!

The ratios of the probabilities (odds), ex. $\frac{P(y=j)}{P(y=0)}$, are independent of one another. Otherwise: use **nested models** or **multinomial probit**.

Interpretation: $\hat{\beta}_j$ – parameters of a binary choice logit model between alternative j and the base alternative 0. So a positive coefficient from *mlogit* means that as the regressor increases, we are more likely to choose alternative j than the base alternative 0.

mlogit pctstck choice age educ wealth89 prftshr, rr baseoutcome(0)

Relative-risk ratio (relative odds) for mlogit models:

$$\frac{P(y_i = j)}{P(y_i = 1)} = \exp(x_i \beta_j)$$

or

Interpretation: $\exp(\beta_j)$ – gives proportionate change in the relative risk of choosing alternative j rather than alternative 0 when x_i changes by 1 unit.

Marginal Effects

margins, predict(outcome(50)) dydx(choice age educ wealth89 prftshr)

margins, predict(outcome(0)) dydx(choice age educ wealth89 prftshr)

Problem 2. Самостоятельная работа.

Загрузите файл с данными fish.dta. Данные содержат информацию о количестве пойманной рыбы людьми, находящимся на отдыхе. Переменные в файле:

Camper – 1, есть палатка, 0, иначе.

Child – количество детей, которых взяли с собой на рыбалку

Count – количество пойманной рыбы

Persons – количество людей в группе

Задание:

1. Постройте бинарную переменную fish – поймали рыбу или нет.
2. Изучите описательные статистики выборки. Постройте гистограмму распределения переменной count.
3. Оцените логит модель с зависимой переменной fish И объясняющими всеми остальными переменными в модели. Что можно сказать о качестве построенной модели?
4. Оцените кривую ROC. Постройте кривые чувствительности и специфичности.
5. Найдите предельные эффекты в точке средних значений и средний предельный эффект. Найдите предельный эффект для группы из 3-х человек с палаткой и одним ребенком.
6. Оцените модель Пуассона. Найдите предельный эффект в точке средних значений. Найдите предельный эффект для группы из 3-х человек с палаткой и одним ребенком.

Problem 3. Задача 12.14 из учебника Катыхшева, Магнуса и Пересецкого.

Пусть $y_t^* = x_t' \beta + \varepsilon_t$, где ошибки ε_t имеют плотность распределения $f(x)$ и

$$y_t = \begin{cases} \alpha_1, & \text{если } y_t^* \leq \alpha_1, \\ y_t^*, & \text{если } \alpha_1 < y_t^* < \alpha_2, \\ \alpha_2 & \text{если } y_t^* \geq \alpha_2. \end{cases}$$

- а) Найдите распределение y_t .
- б) Найдите логарифмическую функцию правдоподобия для оценивания вектора β .
- в) Найдите $(\partial E y) / (\partial x)$.

Problem 4. (Задача составлена на основе задачи 12.5 из Сборника задач к начальному курсу эконометрики Катыхшева, Магнуса, Пересецкого и Головань).

Рассмотрим модель бинарного выбора $\mathbb{P}\{Y_t = 1\} = F(\alpha + \beta d_t)$, где d – фиктивная переменная (принимая значения 0 или 1). Ниже представлены результаты 100 наблюдений:

	$y = 0$	$y = 1$
$d = 0$	20	30
$d = 1$	30	20

- А.** Используя пробит-модель, оцените параметры α и β .

- В.** При помощи теста отношения правдоподобия протестируйте гипотезу $H_0 : \beta = 0$ на уровне значимости 5%.
- С.** *Протестируйте гипотезу $H_0 : \beta = 0$ на уровне значимости 5% с помощью теста Вальда и теста множителей Лагранжа.