

Эконометрика, 2018-2019, 3 модуль
Семинары 6-7
18-25.02.19
Для Группы Э_Б2016_Э_3
Семинарист О.А.Демидова

Модели бинарного выбора

1) Демидова, Малахов, 11.5

Задание 11.5.

База данных rlms14 (файл rlms14.dta)

В файле rlms14.dta приведены данные о заработной плате, возрасте, образовании, профессиональной принадлежности россиян из базы данных RLMS, 14 раунд опроса, проводившийся в 2005 г.

Переменные:

income – ответ на вопрос: «Сколько денег в течение последних 30 дней Вы получили по основному месту работы после вычета налогов и отчислений?»

wage - ответ на вопрос: «Ваша среднемесячная зарплата за последние 12 месяцев?»

rsu – первичная единица отбора (1 – Санкт – Петербург, ..., 38 – Амурская обл.),

marst – семейное положение (1 – никогда в браке не состояли, 2- состоите в зарегистрированном браке, 3 – живете вместе, но не зарегистрированы, 4 – разведены и в браке не состоите, 5 – вдовец/вдова),

occup – профессиональная группа (1- военнослужащие, 2 – законодатели, крупные чиновники, управленцы, 3 – профессионалы с высшим образованием, 4 – конторские служащие, 5 – занятые в сфере обслуживания, 6 – квалифицированные с/х работники, 7 – ремесленники, 8 – промышленные рабочие),

educ – образование (1 – 0 классов школы, 23 – аспирантура и т.п. с дипломом),

diplom – законченное образование (1 – 0-6 классов, 1-2 – незаконченное среднее образование, 3 - законченное среднее образование, 4 - законченное среднее образование, 5 - законченное среднее специальное образование, 6 - законченное высшее образование и выше),

gender – пол респондента (1 – мужской, 2 – женский),

birth_year – год рождения,

industry – ответ на вопрос: «в какой отрасли Вы работаете?» (1 – легкая и пищевая промышленность, ...),

beginning – ответ на вопрос: «с какого года Вы работаете на этом предприятии?»

boss – ответ на вопрос: «У Вас есть подчиненные на этой работе?» (1 – да, 2 – нет),

subordinates – ответ на вопрос: «Сколько у Вас подчиненных?»

duration_weekh - ответ на вопрос: «Сколько часов в среднем продолжается Ваша обычная рабочая неделя?»

life_sat - ответ на вопрос: «Насколько Вы удовлетворены или не удовлетворены Вашей жизнью в целом?» (1 - «Полностью удовлетворены», ..., 5 – «Совсем не удовлетворены»),

job_sat - ответ на вопрос: «Насколько Вы удовлетворены или не удовлетворены Вашей работой в целом?» (1 - «Полностью удовлетворены», 2 - «Скорее удовлетворены», 3 – «И да, и нет», 4 – «Не очень удовлетворены», 5 – «Совсем не удовлетворены»),

job_career - ответ на вопрос: «Насколько Вы удовлетворены или не удовлетворены возможностями Вашего профессионального роста?» (1 - «Полностью удовлетворены», ..., 5 - «Совсем не удовлетворены»),

children - ответ на вопрос: «Сколько у Вас детей?»),

child_less18 - ответ на вопрос: «Сколько у Вас детей моложе 18 лет?»),

По данным файла rlms14.dta с помощью статистического пакета STATA попытайтесь выявить факторы, влияющие на удовлетворенность россиян жизнью в целом, работой, возможностями профессионального роста.

Например, для выявления факторов, влияющих на удовлетворенность россиян жизнью в целом

- 1) Создайте зависимую переменную life, равную 1 для тех, кто на вопрос «Насколько Вы удовлетворены или не удовлетворены Вашей жизнью в целом?» ответили «Полностью удовлетворены» или «Скорее удовлетворены» (т.е. для кого значения переменной life_sat равно 1 или 2).
- 2) Оцените логит и пробит модели с зависимой переменной life, объясняющие переменные выберите самостоятельно (при необходимости создайте новые, например возраст age = 2005 - birth_year и т.п.), Вычислите предельные эффекты объясняющих факторов в выбранной Вами точке. Дайте интерпретацию полученным результатам. Используйте статистический пакет STATA.

Методические указания.

Полезные команды:

```
set more off
```

```
gen life = 0
```

```
replace life = 1 if job_sat < 3
```

```
gen age = 2005 - birth_year
```

```
tab educ
```

```
tab diplom
```

```
logit life gender income children age i.diplom
```

```
label list psu
```

```
logit life gender income children age i.diplom if psu == 2
```

```
gen higheduc = 0
```

```
replace higheduc = 1 if diplom == 6
```

```
probit life gender income children age higheduc if psu == 2
```

```
mfxx
```

```
sum gender income children age higheduc
```

```
mfxx, at(1 5000 0 25 1)
```

```
probit life income children age higheduc if psu == 2 & gender == 2
```

Дополнительные задания.

- 3) Оцените качество оценки моделей

Полезные команды:

```
estat classification
```

```
estat gof
```

```
estat classification, cutoff(0.4)
```

- 4) Постройте график кривых чувствительности и специфичности

Полезная команда:

```
lsens
```

- 5) Постройте график ROC – кривой и вычислите площадь под ней

Полезная команда:

```
lroc
```

2) Демидова, Малахов, задачи 11.1-11.4, задание 11.3

Задача 11.1.

Исследователя интересует зависимость вероятности найти работу от уровня образования индивидуума. Введя в качестве зависимой переменную EMP, равную 1 для работающих и 0 для неработающих и S – количество лет обучения в качестве объясняющей, он оценил логит – модель:

$$P\{EMP_i = 1\} = \frac{1}{1 + \exp\{-Z_i\}}, \quad Z_i = -1.006 + 0.148S$$

(0.242) (0.018)

Оцените предельный эффект объясняющего фактора для среднего значения переменной $S = 13.5$.

Задача 11.2.

По наблюдениям для 570 индивидуумов оценена зависимость получения школьником аттестата от обобщенной оценки результатов тестов X. Переменная Y принимает значение 1, если аттестат был получен и 0 в противном случае.

Оцененные модели имеют следующий вид:

Логит: $P\{Y_i = 1\} = \frac{1}{1 + \exp\{-Z_i\}}, \quad Z_i = -5.004 + 0.1666X$

(0.865) (0.021)

Пробит: $P\{Y_i = 1\} = F(Z_i), \quad F(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-t^2/2} dt, \quad Z = -2.7 + 0.53X$

(0.083) (0.0117)

Дайте экономическую интерпретацию полученным результатам для логит и пробит моделей. Найдите предельный эффект объясняющего фактора в точке $\bar{X} = 50.15$.

Задача 11.3.

Из 750 обратившихся за ссудой в банк 250 было в ней отказано. Для оценки вероятности получения ссуды были оценены линейная и пробит модели:

$$Y = 0.5 + 1.5X,$$

$$P\{Y_i = 1\} = F(Z_i), \quad F(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-t^2/2} dt$$

$$Z = 0.45 + 3X$$

где $Y_i = 1$ для получивших ссуду и 0 иначе, X – доход просителя.

По пробит модели найти предельный эффект дохода в среднем.

Задача 11.4.

Для того, чтобы определить, эффективна ли новая методика преподавания микроэкономики, провели следующий эксперимент: протестировали всех студентов по микроэкономике в конце первого и второго семестра. Часть студентов во втором семестре обучали по новой методике, часть по старой. После этого в качестве объясняющей выбрали переменную Y, равную 1, если результат студента улучшился и 0 в противном случае, а в качестве объясняющих переменных X_1 –

