

Модели бинарного выбора

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 две модели:

линейную:
$$EMP = 0.337 + 0.029S \quad (1)$$

(0.047) (0.003)

и логит – модель:

$$P\{EMP_i = 1\} = \frac{1}{1 + \exp\{-Z_i\}}, \quad Z_i = -1.006 + 0.148S, \quad (2)$$

(0.242) (0.018)

а) Как оценивается логит-модель?

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

Решение:

а) Логит-модель оценивается с помощью метода максимального правдоподобия.

б) $z(\bar{s}) = -1.006 + 0.148 \cdot 13.5 = 0.992$

$$\exp(-z(\bar{s})) = \exp(-0.992) = 0.371$$

$$f(z) = \frac{e^{-z}}{(1 + e^{-z})^2}$$

$$f(z(\bar{s})) = \frac{0.371}{(1 + 0.371)^2} = 0.198$$

$$\frac{\partial p(\bar{s})}{\partial s} = \frac{\partial p}{\partial z} \cdot \frac{\partial z}{\partial s} = f(z(\bar{s})) \cdot \hat{\beta}_j = 0.198 \cdot 0.148 = 0.03$$

Интерпретация: при увеличении длительности обучения (S) на 1 год для индивида со «средними» способностями вероятность найти работу увеличивается на 3%.

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:

Линейная:
$$Y = 0.308 + 0.012X,$$

(0.058) (0.00114)

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

(0.865) (0.021)

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

Решение:

Найдем предельный эффект в логит модели:

$$z(\bar{x}) = -5,004 + 0,1666 \cdot 50,15 = 3,351$$

$$\exp(-z(\bar{s})) = \exp(-3.351) = 0,0351$$

$$f(z) = \frac{e^{-z}}{(1 + e^{-z})^2}$$

$$f(z(\bar{x})) = \frac{0,0351}{(1 + 0,0351)^2} = 0,033$$

$$\frac{\partial p(\bar{x})}{\partial x} = \frac{\partial p}{\partial z} \cdot \frac{\partial z}{\partial x} = f(z(\bar{x})) \cdot \hat{\beta}_j = 0,033 \cdot 0,1666 \approx 0,0055$$

При увеличении обобщенной оценки результатов тестов на 1 балл для индивида с обобщенной оценкой равной 50,15 вероятность получения аттестата увеличивается на 0,55%.

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.

Решение:

Для начала найдем средний доход просителя из линейной модели.

$$\bar{Y} = 0.5 + 1.5\bar{X}$$

$Y_i = 1$, если ссуда была получена.

$Y_i = 0$, иначе.

$Y_i = 0$ в 250 случаях $\Rightarrow Y_i = 1$ в 500 случаях.

$$\text{Отсюда } \bar{Y} = \frac{500}{750} = \frac{2}{3}.$$

Подставим это значение в $\bar{Y} = 0.5 + 1.5\bar{X}$:

$$\frac{2}{3} = \frac{1}{2} + 1,5 \cdot \bar{X} \Rightarrow \bar{X} = 0,111$$

Теперь найдем предельный эффект в пробит модели:

$$z(\bar{x}) = 0,45 + 3 \cdot 0,111 = 0,783333$$

$$f(z(\bar{x})) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} = 0,2935$$

$$\frac{\partial p(\bar{x})}{\partial x} = \frac{\partial p}{\partial z} \cdot \frac{\partial z}{\partial x} = f(z(\bar{x})) \cdot \hat{\beta}_j \approx 0,2935 \cdot 3 \approx 0,88$$

При увеличении дохода на 1 единицу (в данной задаче доход измеряется, скорее всего в крупных единицах, например в млн. руб.) для индивида со средним доходом вероятность получения ссуды увеличивается на 88%.

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 – результаты теста в первом семестре, X_2 – средний балл по остальным предметам, D – равную 1, если студент обучался по новой методике и 0, если по старой. Оценили две модели:

$$\text{линейную } Y = -1.498 + 0.01X_1 + 0.464X_2 + 0.379D \quad (1)$$

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

$$Z = -7.452 + 0.052X_1 + 1.626X_2 + 1.426D.$$

а) Дайте экономическую интерпретацию коэффициентам модели 1. Каковы недостатки этой модели?

б) Найдите предельный эффект переменной D при средних значениях $\bar{X}_1 = 21.938$,

$\bar{X}_2 = 3.117$ (разность вероятностей улучшения результата при $D = 1$ и $D = 0$).

Решение:

Пункт б.

Рассчитаем предельный эффект переменной D при средних значениях:

$$p(z(\bar{x}_1, \bar{x}_2; D = 1)) - p(z(\bar{x}_1, \bar{x}_2; D = 0)).$$

$$z(\bar{x}_1, \bar{x}_2; D = 1) = -7,452 + 0,052 \cdot 21,938 + 1,626 \cdot 3,117 + 1,426 \cdot 1 = 0,183$$

$$z(\bar{x}_1, \bar{x}_2; D = 0) = -7,452 + 0,052 \cdot 21,938 + 1,626 \cdot 3,117 + 1,426 \cdot 0 = -1,243$$

$$p(z(\bar{x}_1, \bar{x}_2; D = 1)) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0,183} e^{-\frac{t^2}{2}} dt = \frac{1}{\sqrt{2\pi}} e^{-\frac{0,183^2}{2}} = 0,392$$

$$p(z(\bar{x}_1, \bar{x}_2; D = 0)) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-1,243} e^{-\frac{t^2}{2}} dt = \frac{1}{\sqrt{2\pi}} e^{-\frac{-1,243^2}{2}} = 0,184$$

$$p(z(\bar{x}_1, \bar{x}_2; D = 1)) - p(z(\bar{x}_1, \bar{x}_2; D = 0)) = 0,392 - 0,184 \approx 0,2$$

Переход на обучение по новой методике увеличивает вероятность улучшения результата студента на 20%.

5. Для анализа аудитории, использующей Интернет для учебы по данным для 1314 индивидов были оценены линейная и пробит модели (последняя с предельными

эффектами), в которых $\text{intlear} = 1$ при использовании индивидом Интернета для учебы и 0 в противном случае, $\text{male} = 1$ для мужчин и 0 для женщин, income – заработная плата индивида по основному месту работы, age – возраст. В скобках указаны соответственно t или z - статистики. В чем состоят недостатки линейной модели? Дайте интерпретацию полученным результатам.

$$\text{INTLEAR} = -0.78 - .013 \text{ AGE} - 4.53 \cdot 10^{-10} \text{ INCOME} - 0.073 \text{ MALE}$$

(18.48) (-11.45) (-0.96) (-3.05)

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

$$Z = 1.01 - 0.044 \text{ AGE} - 1.51 \cdot 10^{-9} \text{ INCOME} - 0.23 \text{ MALE}$$

(7.25) (-10.82) (-0.98) (-3.08)

Marginal effects after probit

variable	dy/dx	Std. Err.	z	P>z	[95% C.I.]	
age	-.0147444	.00133	-11.08	0.000	-.017353	-.012136
income	-5.07e-10	.00000	-0.98	0.328	-1.5e-09	5.1e-10
male*	-.0783057	.02536	-3.09	0.002	-.128015	-.028597

(*) dy/dx is for discrete change of dummy variable from 0 to 1

Решение:

Переменная INCOME незначима, значит, доход не влияет на пользование Интернетом.

Каждый дополнительный год снижает вероятность использования Интернета для учебы на 1,47%. Мужчины используют Интернет для учебы реже женщин на 7,8%.