Микроэконометрика

Модели бинарного выбора

Потанин Богдан Станиславович

старший преподаватель, кандидат экономических наук

2021-2022

- В базовых регрессионных моделях зависимая переменная является непрерывной: зарплата, прибыль, стоимость квартиры и т.д.

- В базовых регрессионных моделях зависимая переменная является непрерывной: зарплата, прибыль, стоимость квартиры и т.д.
- Например, при помощи метода наименьших квадратов можно оценить, как образование и стаж работы влияют на зарплату индивида.

Модели бинарного выбора

Мотивация

- В базовых регрессионных моделях зависимая переменная является непрерывной: зарплата, прибыль, стоимость квартиры и т.д.
- Например, при помощи метода наименьших квадратов можно оценить, как образование и стаж работы влияют на зарплату индивида.
- На практике зависимая переменная может быть измерена не только в непрерывной, но и в некоторой иной шкале, в том числе в бинарной.

Модели бинарного выбора
Мотивация

- В базовых регрессионных моделях зависимая переменная является непрерывной: зарплата, прибыль, стоимость квартиры и т.д.
- Например, при помощи метода наименьших квадратов можно оценить, как образование и стаж работы влияют на зарплату индивида.
- На практике зависимая переменная может быть измерена не только в непрерывной, но и в некоторой иной шкале, в том числе в бинарной.
- Например, можно изучать, как характеристики банка (структура активов, размер и т.д.) влияют на возможность дефолта, который может либо произойти, либо не произойти.

Модели бинарного выбора

Мотивация

- В базовых регрессионных моделях зависимая переменная является непрерывной: зарплата, прибыль, стоимость квартиры и т.д.
- Например, при помощи метода наименьших квадратов можно оценить, как образование и стаж работы влияют на зарплату индивида.
- На практике зависимая переменная может быть измерена не только в непрерывной, но и в некоторой иной шкале, в том числе в бинарной.
- Например, можно изучать, как характеристики банка (структура активов, размер и т.д.) влияют на возможность дефолта, который может либо произойти, либо не произойти.
- **Бинарные переменные** принимают два возможных значения, кодируемые как 0 и 1.

- В базовых регрессионных моделях зависимая переменная является непрерывной: зарплата, прибыль, стоимость квартиры и т.д.
- Например, при помощи метода наименьших квадратов можно оценить, как образование и стаж работы влияют на зарплату индивида.
- На практике зависимая переменная может быть измерена не только в непрерывной, но и в некоторой иной шкале, в том числе в бинарной.
- Например, можно изучать, как характеристики банка (структура активов, размер и т.д.) влияют на возможность дефолта, который может либо произойти, либо не произойти.
- **Бинарные переменные** принимают два возможных значения, кодируемые как 0 и 1.

Зависимая переменная	1	0	Возможные регрессоры
Дефолт банка	Произошел	Не произошел	Характеристики банков
Занятость индивида	Работает	Не работает	Характеристики индивидов
Результат стартапа	Успех	Неудача	Характеристики стартапов
Выплата дивидендов	Произошла	Не произошла	Характеристики фирм

Модели бинарного выбора
Формализация

- Обозначим через y_i бинарную переменную, относящуюся к i-му наблюдению, где $i \in \{1, ..., n\}$ и n – объем выборки.

- Обозначим через y_i бинарную переменную, относящуюся к i-му наблюдению, где $i \in \{1, ..., n\}$ и n – объем выборки.
- Через $x_i = (x_{1i}, ..., x_{mi})$ обозначим вектор (строку) значений независимых переменных i-го наблюдения, где m – число регрессоров.

Модели бинарного выбора

Формализация

- Обозначим через y_i бинарную переменную, относящуюся к i-му наблюдению, где $i \in \{1, ..., n\}$ и n – объем выборки.
- Через $x_i = (x_{1i}, ..., x_{mi})$ обозначим вектор (строку) значений независимых переменных i-го наблюдения, где m – число регрессоров.
- Предположим, что наблюдения, определяемые как случайные векторы (y_i, x_i), независимы между собой (по i) и одинаково распределены.

Модели бинарного выбора

Формализация

- Обозначим через y_i бинарную переменную, относящуюся к i-му наблюдению, где $i \in \{1, ..., n\}$ и n – объем выборки.
- Через $x_i = (x_{1i}, ..., x_{mi})$ обозначим вектор (строку) значений независимых переменных i-го наблюдения, где m – число регрессоров.
- Предположим, что наблюдения, определяемые как случайные векторы (y_i, x_i), независимы между собой (по i) и одинаково распределены.
- Допустим, что распределение $(y_i|x_i)$ задается как:

$$P(y_i = 1|x_i) = F^*(x_i; \beta), \qquad P(y_i = 0|x_i) = 1 - F^*(x_i; \beta),$$

где F^* – это некоторая функция с вектором параметров β.

- Обозначим через y_i бинарную переменную, относящуюся к i-му наблюдению, где $i \in \{1, ..., n\}$ и n – объем выборки.
- Через $x_i = (x_{1i}, ..., x_{mi})$ обозначим вектор (строку) значений независимых переменных i-го наблюдения, где m – число регрессоров.
- Предположим, что наблюдения, определяемые как случайные векторы (y_i, x_i), независимы между собой (по i) и одинаково распределены.
- Допустим, что распределение $(y_i|x_i)$ задается как:

$$P(y_i = 1|x_i) = F^*(x_i; \beta), \qquad P(y_i = 0|x_i) = 1 - F^*(x_i; \beta),$$

  где F^* – это некоторая функция с вектором параметров β.
- На практике обычно рассматривают модели с функцими вида $F^*(x_i; \beta) = F(x_i\beta)$, где $x_i\beta$ именуют **линейным индексом**, а $\beta = (\beta_1, ..., \beta_m)$ – вектором (столбцом) регрессионных коэффициентов.

Модели бинарного выбора

Формализация

- Обозначим через y_i бинарную переменную, относящуюся к i-му наблюдению, где $i \in \{1, ..., n\}$ и n – объем выборки.
- Через $x_i = (x_{1i}, ..., x_{mi})$ обозначим вектор (строку) значений независимых переменных i-го наблюдения, где m – число регрессоров.
- Предположим, что наблюдения, определяемые как случайные векторы (y_i, x_i), независимы между собой (по i) и одинаково распределены.
- Допустим, что распределение $(y_i|x_i)$ задается как:

$$P(y_i = 1|x_i) = F^*(x_i;\beta), \qquad P(y_i = 0|x_i) = 1 - F^*(x_i;\beta),$$

  где F^* – это некоторая функция с вектором параметров β.
- На практике обычно рассматривают модели с функциями вида $F^*(x_i;\beta) = F(x_i\beta)$, где $x_i\beta$ именуют **линейным индексом**, а $\beta = (\beta_1, ..., \beta_m)$ – вектором (столбцом) регрессионных коэффициентов.
- Для того, чтобы изучить, как x_i влияют на y_i, необходимо подобрать функцию $F(.)$ и оценить β.

- Предположим, что:

$$P(y_i = 1|x_i) = F(x_i\beta) = x_i\beta$$

- Предположим, что:

$$P(y_i = 1|x_i) = F(x_i\beta) = x_i\beta$$

- Рассчитаем условное математическое ожидание:

$$E(y_i|x_i) = 1 \times x_i\beta + 0 \times (1 - x_i\beta) = x_i\beta$$

- Предположим, что:

$$P(y_i = 1|x_i) = F(x_i\beta) = x_i\beta$$

- Рассчитаем условное математическое ожидание:

$$E(y_i|x_i) = 1 \times x_i\beta + 0 \times (1 - x_i\beta) = x_i\beta$$

- В результате получаем регрессионное уравнение:

$$y_i = x_i\beta + \varepsilon_i, \text{ где } \varepsilon_i = y_i - E(y_i|x_i)$$

Линейно-вероятностная модель

Оценивание с помощью метода нименьших квадратов

- Предположим, что:

$$P(y_i = 1|x_i) = F(x_i\beta) = x_i\beta$$

- Рассчитаем условное математическое ожидание:

$$E(y_i|x_i) = 1 \times x_i\beta + 0 \times (1 - x_i\beta) = x_i\beta$$

- В результате получаем регрессионное уравнение:

$$y_i = x_i\beta + \varepsilon_i, \text{ где } \varepsilon_i = y_i - E(y_i|x_i)$$

- Обратим внимание, что:

$$E(\varepsilon_i|x_i) = E(y_i|x_i) - E\left(\underbrace{E(y_i|x_i)}_{\text{константа}}\right) = x_i\beta - x_i\beta = 0$$

- Предположим, что:

$$P(y_i = 1|x_i) = F(x_i\beta) = x_i\beta$$

- Рассчитаем условное математическое ожидание:

$$E(y_i|x_i) = 1 \times x_i\beta + 0 \times (1 - x_i\beta) = x_i\beta$$

- В результате получаем регрессионное уравнение:

$$y_i = x_i\beta + \varepsilon_i,\ \text{где } \varepsilon_i = y_i - E(y_i|x_i)$$

- Обратим внимание, что:

$$E(\varepsilon_i|x_i) = E(y_i|x_i) - E\left(\underbrace{E(y_i|x_i)}_{\text{константа}}\right) = x_i\beta - x_i\beta = 0$$

Поскольку $E(\varepsilon_i|x_i) = 0$, то с помощью метода наименьших квадратов (МНК) можно получить состоятельную оценку $\hat{\beta}$ вектора β.

Линейно-вероятностная модель

Оценивание с помощью метода нименьших квадратов

- Предположим, что:
$$P(y_i = 1|x_i) = F(x_i\beta) = x_i\beta$$
- Рассчитаем условное математическое ожидание:
$$E(y_i|x_i) = 1 \times x_i\beta + 0 \times (1 - x_i\beta) = x_i\beta$$
- В результате получаем регрессионное уравнение:
$$y_i = x_i\beta + \varepsilon_i, \text{ где } \varepsilon_i = y_i - E(y_i|x_i)$$
- Обратим внимание, что:
$$E(\varepsilon_i|x_i) = E(y_i|x_i) - E\left(\underbrace{E(y_i|x_i)}_{\text{константа}}\right) = x_i\beta - x_i\beta = 0$$

  Поскольку $E(\varepsilon_i|x_i) = 0$, то с помощью метода наименьших квадратов (МНК) можно получить состоятельную оценку $\hat{\beta}$ вектора β.
- Оценки условных вероятностей считаются как $\hat{P}(y_i = 1|x_i) = x_i\hat{\beta}$.

- Для того, чтобы измерить влияние x_i на y_i, можно рассчитать предельные эффекты регрессоров на условную вероятность принятия бинарной переменной единичного значения.

- Для того, чтобы измерить влияние x_i на y_i, можно рассчитать предельные эффекты регрессоров на условную вероятность принятия бинарной переменной единичного значения.
- Предельный эффект x_{ki} на $P(y_i = 1|x_i)$ рассчитывается как:

$$\frac{\partial P(y_i = 1|x_i)}{\partial x_{ki}} = \frac{\partial x_i\beta}{\partial x_{ki}} = \frac{\partial\left(x_{1i}\beta_1 + ... + x_{mi}\beta_m\right)}{\partial x_{ki}} = \beta_k$$

- Для того, чтобы измерить влияние x_i на y_i, можно рассчитать предельные эффекты регрессоров на условную вероятность принятия бинарной переменной единичного значения.
- Предельный эффект x_{ki} на $P(y_i = 1|x_i)$ рассчитывается как:

$$\frac{\partial P(y_i = 1|x_i)}{\partial x_{ki}} = \frac{\partial x_i \beta}{\partial x_{ki}} = \frac{\partial\left(x_{1i}\beta_1 + ... + x_{mi}\beta_m\right)}{\partial x_{ki}} = \beta_k$$

- В результате оценка предельного эффекта совпадает с оценкой соответствующего регрессионного коэффициента $\hat{\beta}_k$.

- Для того, чтобы измерить влияние x_i на y_i, можно рассчитать предельные эффекты регрессоров на условную вероятность принятия бинарной переменной единичного значения.
- Предельный эффект x_{ki} на $P(y_i = 1|x_i)$ рассчитывается как:

$$\frac{\partial P(y_i = 1|x_i)}{\partial x_{ki}} = \frac{\partial x_i\beta}{\partial x_{ki}} = \frac{\partial\left(x_{1i}\beta_1 + ... + x_{mi}\beta_m\right)}{\partial x_{ki}} = \beta_k$$

- В результате оценка предельного эффекта совпадает с оценкой соответствующего регрессионного коэффициента $\hat{\beta}_k$.
- Допустим, вы оценили влияние ряда характеристик индивида, в том числе дохода (в тысячах рублей), на вероятность наличия у него страхового полиса. Вы получили, что $\hat{\beta}_{\text{доход}} = 0.02$.

Линейно-вероятностная модель

Предельные эффекты

- Для того, чтобы измерить влияние x_i на y_i, можно рассчитать предельные эффекты регрессоров на условную вероятность принятия бинарной переменной единичного значения.
- Предельный эффект x_{ki} на $P(y_i = 1|x_i)$ рассчитывается как:

$$\frac{\partial P(y_i = 1|x_i)}{\partial x_{ki}} = \frac{\partial x_i \beta}{\partial x_{ki}} = \frac{\partial \left(x_{1i}\beta_1 + ... + x_{mi}\beta_m\right)}{\partial x_{ki}} = \beta_k$$

- В результате оценка предельного эффекта совпадает с оценкой соответствующего регрессионного коэффициента $\hat{\beta}_k$.
- Допустим, вы оценили влияние ряда характеристик индивида, в том числе дохода (в тысячах рублей), на вероятность наличия у него страхового полиса. Вы получили, что $\hat{\beta}_{\text{доход}} = 0.02$. В соответствии с вашей оценкой при прочих равных при увеличении (уменьшении) дохода индивида на 5 тысяч рублей вероятность наличия полиса для него возрастет (снизится) на $0.02 * 5 = 0.1$.

Проблема интерпретации вероятностей

- Поскольку $P(y_i = 1|x_i) = x_i\beta \in [0, 1]$, то необходимо ввести ограничение на носитель x_i, то есть $\text{supp}(x_i) = \{t \in R^m : 0 \leq t\beta \leq 1\}$.

Проблема интерпретации вероятностей

- Поскольку $P(y_i = 1|x_i) = x_i\beta \in [0, 1]$, то необходимо ввести ограничение на носитель x_i, то есть $\text{supp}(x_i) = \{t \in R^m : 0 \leq t\beta \leq 1\}$.
- Даже при соблюдении данного весьма нереалистичного ограничения оценки вероятностей могут быть отрицательными или превышать единицу.

Проблема интерпретации вероятностей

- Поскольку $P(y_i = 1|x_i) = x_i\beta \in [0, 1]$, то необходимо ввести ограничение на носитель x_i, то есть $\text{supp}(x_i) = \{t \in R^m : 0 \leq t\beta \leq 1\}$.
- Даже при соблюдении данного весьма нереалистичного ограничения оценки вероятностей могут быть отрицательными или превышать единицу.
- Например, если $x_i = (5, -2)$ и $\hat{\beta} = (2, 4)^T$, то:

$$x_i\hat{\beta} = 5 \times 2 - 2 \times 4 = 2 \implies \hat{P}(y_i = 1|x_i) = 2 > 1$$

Проблема интерпретации вероятностей

- Поскольку $P(y_i = 1|x_i) = x_i\beta \in [0, 1]$, то необходимо ввести ограничение на носитель x_i, то есть $\text{supp}(x_i) = \{t \in R^m : 0 \leq t\beta \leq 1\}$.
- Даже при соблюдении данного весьма нереалистичного ограничения оценки вероятностей могут быть отрицательными или превышать единицу.
- Например, если $x_i = (5, -2)$ и $\hat{\beta} = (2, 4)^T$, то:

$$x_i\hat{\beta} = 5 \times 2 - 2 \times 4 = 2 \implies \hat{P}(y_i = 1|x_i) = 2 > 1$$

Проблема гетероскедастичности

- Обратим внимание, что $(y_i|x_i)$ имеет распределение Бернулли с параметром $x_i\beta$.

Проблема интерпретации вероятностей

- Поскольку $P(y_i = 1|x_i) = x_i\beta \in [0, 1]$, то необходимо ввести ограничение на носитель x_i, то есть $\text{supp}(x_i) = \{t \in R^m : 0 \leq t\beta \leq 1\}$.
- Даже при соблюдении данного весьма нереалистичного ограничения оценки вероятностей могут быть отрицательными или превышать единицу.
- Например, если $x_i = (5, -2)$ и $\hat{\beta} = (2, 4)^T$, то:

$$x_i\hat{\beta} = 5 \times 2 - 2 \times 4 = 2 \implies \hat{P}(y_i = 1|x_i) = 2 > 1$$

Проблема гетероскедастичности

- Обратим внимание, что $(y_i|x_i)$ имеет распределение Бернулли с параметром $x_i\beta$.
- Случайные ошибки гетероскедастичны, поскольку:

$$Var(\varepsilon_i|x_i) = Var(y_i - E(y_i|x_i)|x_i) = Var(y_i|x_i) = x_i\beta(1 - x_i\beta)$$

Проблема интерпретации вероятностей

- Поскольку $P(y_i = 1|x_i) = x_i\beta \in [0, 1]$, то необходимо ввести ограничение на носитель x_i, то есть $\text{supp}(x_i) = \{t \in R^m : 0 \leq t\beta \leq 1\}$.
- Даже при соблюдении данного весьма нереалистичного ограничения оценки вероятностей могут быть отрицательными или превышать единицу.
- Например, если $x_i = (5, -2)$ и $\hat{\beta} = (2, 4)^T$, то:

$$x_i\hat{\beta} = 5 \times 2 - 2 \times 4 = 2 \implies \hat{P}(y_i = 1|x_i) = 2 > 1$$

Проблема гетероскедастичности

- Обратим внимание, что $(y_i|x_i)$ имеет распределение Бернулли с параметром $x_i\beta$.
- Случайные ошибки гетероскедастичны, поскольку:

$$Var(\varepsilon_i|x_i) = Var(y_i - E(y_i|x_i)|x_i) = Var(y_i|x_i) = x_i\beta(1 - x_i\beta)$$

- Из-за гетероскедастичности МНК оценки окажутся неэффективны.

- Черех X обозначим матрицу регрессоров, у которой в i-й строке располагается вектор x_i.

- Черех X обозначим матрицу регрессоров, у которой в i-й строке располагается вектор x_i.
- Из-за гетероскедастичности асимптотическую ковариационную матрицу оценок необхоидмо оценивать при помощи бутстрапа, либо с использованием поправки:

$$\widehat{As.Cov}(\hat{\beta}) = (X^T X)^{-1}\left(\sum_{i=1}^{n} x_i\hat{\beta}(1 - x_i\hat{\beta})x_i^T x_i\right)(X^T X)^{-1}$$

- Черех X обозначим матрицу регрессоров, у которой в i-й строке располагается вектор x_i.
- Из-за гетероскедастичности асимптотическую ковариационную матрицу оценок необхоидмо оценивать при помощи бутстрапа, либо с использованием поправки:

$$\widehat{As.Cov}(\hat{\beta}) = (X^TX)^{-1}\left(\sum_{i=1}^{n}x_i\hat{\beta}(1-x_i\hat{\beta})x_i^Tx_i\right)(X^TX)^{-1}$$

- Тестирование гипотез о регрессионных коэффициентах β осуществляется стандартным образом.

- Через X обозначим матрицу регрессоров, у которой в i-й строке располагается вектор x_i.
- Из-за гетероскедастичности асимптотическую ковариационную матрицу оценок необхоидмо оценивать при помощи бутстрапа, либо с использованием поправки:

$$\widehat{As.Cov}(\hat{\beta}) = (X^T X)^{-1}\left(\sum_{i=1}^{n} x_i\hat{\beta}(1 - x_i\hat{\beta})x_i^T x_i\right)(X^T X)^{-1}$$

- Тестирование гипотез о регрессионных коэффициентах β осуществляется стандартным образом.
- Тестирование гипотез об условных вероятностях сводится к тестирование гипотез о линейной комбинации регрессионных коэффициентов. Например, если $x_i = (1, 2)$, то $H_0 : P(y_i = 1|x_i) = 0.5$ можно сформулировать как $H_0 : \beta_1 + 2\beta_2 = 0.5$ и тестировать соответствующую гипотезу стандартным образом.

- Поскольку функция $F(x_i\beta)$ задает вероятность, то желательно, чтобы при любых возможных значениях x_i она принимала значения от 0 до 1.

- Поскольку функция $F(x_i\beta)$ задает вероятность, то желательно, чтобы при любых возможных значениях x_i она принимала значения от 0 до 1.
- Для этого предположим, что $F(x_i\beta)$ является функцией распределения некоторой непрерывной случайной величины. Чаще всего рассматривают функцию распределения нормального или логистического распределения.

- Поскольку функция $F(x_i\beta)$ задает вероятность, то желательно, чтобы при любых возможных значениях x_i она принимала значения от 0 до 1.
- Для этого предположим, что $F(x_i\beta)$ является функцией распределения некоторой непрерывной случайной величины. Чаще всего рассматривают функцию распределения нормального или логистического распределения.
- Через $\Phi(t)$ обозначим функцию распределения стандартного нормального распределения. Если $F(x_i\beta) = \Phi(x_i\beta)$, то мы получаем **пробит модель**.

- Поскольку функция $F(x_i\beta)$ задает вероятность, то желательно, чтобы при любых возможных значениях x_i она принимала значения от 0 до 1.
- Для этого предположим, что $F(x_i\beta)$ является функцией распределения некоторой непрерывной случайной величины. Чаще всего рассматривают функцию распределения нормального или логистического распределения.
- Через $\Phi(t)$ обозначим функцию распределения стандартного нормального распределения. Если $F(x_i\beta)=\Phi(x_i\beta)$, то мы получаем **пробит модель**.
- Через $\Lambda(t)=\frac{1}{1+e^{-t}}$ обозначим функцию распределения стандартного логистического распределения. Если $F(x_i\beta)=\Lambda(x_i\beta)$, то мы получаем **логит модель**.

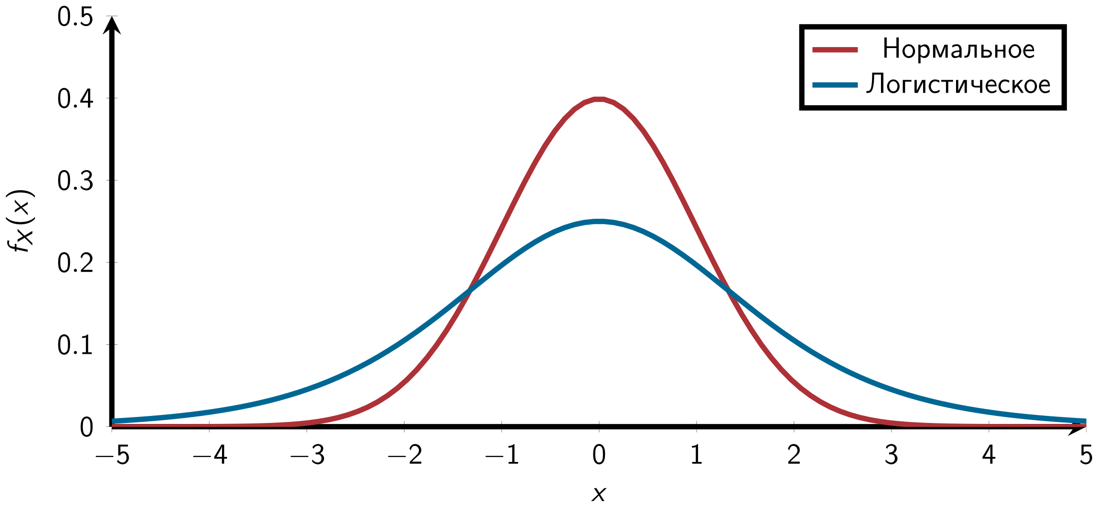
- Поскольку функция $F(x_i\beta)$ задает вероятность, то желательно, чтобы при любых возможных значениях x_i она принимала значения от 0 до 1.
- Для этого предположим, что $F(x_i\beta)$ является функцией распределения некоторой непрерывной случайной величины. Чаще всего рассматривают функцию распределения нормального или логистического распределения.
- Через $\Phi(t)$ обозначим функцию распределения стандартного нормального распределения. Если $F(x_i\beta)=\Phi(x_i\beta)$, то мы получаем **пробит модель**.
- Через $\Lambda(t)=\frac{1}{1+e^{-t}}$ обозначим функцию распределения стандартного логистического распределения. Если $F(x_i\beta)=\Lambda(x_i\beta)$, то мы получаем **логит модель**.
- Нормальное и логистическое распределения симметричны вокруг нуля, поэтому:

$$\Phi(-x_i\beta)=1-\Phi(x_i\beta),\qquad \Lambda(-x_i\beta)=1-\Lambda(x_i\beta)$$
$$\phi(-x_i\beta)=\phi(x_i\beta),\qquad \lambda(-x_i\beta)=\lambda(x_i\beta)$$

Где $\phi(.)$ и $\lambda(.)$ это функции плотности стандартного нормального и логистического распределений соответственно.

Классические модели бинарного выбора

Сравнение нормального и логистического распределений



Классические модели бинарного выбора

Оценивание

- В силу сложной нелинейной формы функции $F(x_i\beta)$ в логит и пробит моделях мы не можем привести регрессионное уравнение к виду, при котором оценки β могут быть получены при помощи метода наименьших квадратов.

- В силу сложной нелинейной формы функции $F(x_i\beta)$ в логит и пробит моделях мы не можем привести регрессионное уравнение к виду, при котором оценки β могут быть получены при помощи метода наименьших квадратов.
- Для того, чтобы получить состоятельные и асимптотически эффективные оценки β можно воспользоваться методом максимального правдоподобия.

- В силу сложной нелинейной формы функции $F(x_i\beta)$ в логит и пробит моделях мы не можем привести регрессионное уравнение к виду, при котором оценки β могут быть получены при помощи метода наименьших квадратов.
- Для того, чтобы получить состоятельные и асимптотически эффективные оценки β можно воспользоваться методом максимального правдоподобия.
- Функция правдоподобия (условная на X) будет иметь вид:

$$L(\beta; y|X) = \prod_{i=1}^{n} F(x_i\beta)^{y_i}\left(1 - F(x_i\beta)\right)^{1-y_i} = \prod_{i:y_i=1} F(x_i\beta) \prod_{i:y_i=0} 1 - F(x_i\beta)$$

- В силу сложной нелинейной формы функции $F(x_i\beta)$ в логит и пробит моделях мы не можем привести регрессионное уравнение к виду, при котором оценки β могут быть получены при помощи метода наименьших квадратов.
- Для того, чтобы получить состоятельные и асимптотически эффективные оценки β можно воспользоваться методом максимального правдоподобия.
- Функция правдоподобия (условная на X) будет иметь вид:

$$L(\beta;y|X)=\prod_{i=1}^{n}F(x_i\beta)^{y_i}\left(1-F(x_i\beta)\right)^{1-y_i}=\prod_{i:y_i=1}F(x_i\beta)\prod_{i:y_i=0}1-F(x_i\beta)$$

- Можно показать, что в случае пробит и логит моделей функция правдоподобия будет иметь единственный максимум, который, однако, не может быть записан в аналитическом виде, вследствие чего его поиск осуществляется с помощью численных методов оптимизации (Newton, BFGS, BHHH и т.д.).

- Для удобства процесс генерации данных в моделях бинарного выбора обычно формулируют с использованием латентной переменной:

$$y_i^* = x_i\beta + \varepsilon_i$$

$$y_i = \begin{cases} 1, \text{ если } y_i^* \geq 0 \\ 0, \text{ если } y_i^* < 0 \end{cases} = \begin{cases} 1, \text{ если } \varepsilon_i \geq -x_i\beta \\ 0, \text{ если } \varepsilon_i < -x_i\beta \end{cases}$$

Где случайные ошибки ε_i одинаково распределены, независимы и не зависят от x_i.

Классические модели бинарного выбора

Латентная переменная

- Для удобства процесс генерации данных в моделях бинарного выбора обычно формулируют с использованием латентной переменной:

$$y_i^* = x_i\beta + \varepsilon_i$$

$$y_i = \begin{cases} 1, \text{ если } y_i^* \geq 0 \\ 0, \text{ если } y_i^* < 0 \end{cases} = \begin{cases} 1, \text{ если } \varepsilon_i \geq -x_i\beta \\ 0, \text{ если } \varepsilon_i < -x_i\beta \end{cases}$$

  Где случайные ошибки ε_i одинаково распределены, независимы и не зависят от x_i.
- Образно говоря латентная переменная y_i^* отражает склонность зависимой бинарной переменной y_i к тому, чтобы принять единичное значение.

- Для удобства процесс генерации данных в моделях бинарного выбора обычно формулируют с использованием латентной переменной:

$$y_i^* = x_i\beta + \varepsilon_i$$

$$y_i = \begin{cases} 1, \text{ если } y_i^* \geq 0 \\ 0, \text{ если } y_i^* < 0 \end{cases} = \begin{cases} 1, \text{ если } \varepsilon_i \geq -x_i\beta \\ 0, \text{ если } \varepsilon_i < -x_i\beta \end{cases}$$

Где случайные ошибки ε_i одинаково распределены, независимы и не зависят от x_i.

- Образно говоря латентная переменная y_i^* отражает склонность зависимой бинарной переменной y_i к тому, чтобы принять единичное значение.
- Данная формулировка эквивалентна рассмотренной ранее, поскольку:

$$P(y_i = 1|x_i) = P(\varepsilon_i \geq -x_i\beta|x_i) = 1 - P(\varepsilon_i < -x_i\beta|x_i) = 1 - F_{\varepsilon_i}(-x_i\beta)$$

Где через $F_{\varepsilon_i}(-x_i\beta)$ обозначается функция распределения случайной ошибки ε_i.

- Для удобства процесс генерации данных в моделях бинарного выбора обычно формулируют с использованием латентной переменной:

$$y_i^* = x_i\beta + \varepsilon_i$$

$$y_i = \begin{cases} 1, \text{ если } y_i^* \geq 0 \\ 0, \text{ если } y_i^* < 0 \end{cases} = \begin{cases} 1, \text{ если } \varepsilon_i \geq -x_i\beta \\ 0, \text{ если } \varepsilon_i < -x_i\beta \end{cases}$$

  Где случайные ошибки ε_i одинаково распределены, независимы и не зависят от x_i.

- Образно говоря латентная переменная y_i^* отражает склонность зависимой бинарной переменной y_i к тому, чтобы принять единичное значение.
- Данная формулировка эквивалентна рассмотренной ранее, поскольку:

$$P(y_i = 1|x_i) = P(\varepsilon_i \geq -x_i\beta|x_i) = 1 - P(\varepsilon_i < -x_i\beta|x_i) = 1 - F_{\varepsilon_i}(-x_i\beta)$$

  Где через $F_{\varepsilon_i}(-x_i\beta)$ обозначается функция распределения случайной ошибки ε_i. Поскольку случайные ошибки одинаково распределены получаем $1 - F_{\varepsilon_i}(-x_i\beta) = F(x_i\beta)$.

Классические модели бинарного выбора

Латентная переменная

- Для удобства процесс генерации данных в моделях бинарного выбора обычно формулируют с использованием латентной переменной:

$$y_i^* = x_i\beta + \varepsilon_i$$

$$y_i = \begin{cases} 1, \text{ если } y_i^* \geq 0 \\ 0, \text{ если } y_i^* < 0 \end{cases} = \begin{cases} 1, \text{ если } \varepsilon_i \geq -x_i\beta \\ 0, \text{ если } \varepsilon_i < -x_i\beta \end{cases}$$

  Где случайные ошибки ε_i одинаково распределены, независимы и не зависят от x_i.
- Образно говоря латентная переменная y_i^* отражает склонность зависимой бинарной переменной y_i к тому, чтобы принять единичное значение.
- Данная формулировка эквивалентна рассмотренной ранее, поскольку:

$$P(y_i = 1|x_i) = P(\varepsilon_i \geq -x_i\beta|x_i) = 1 - P(\varepsilon_i < -x_i\beta|x_i) = 1 - F_{\varepsilon_i}(-x_i\beta)$$

  Где через $F_{\varepsilon_i}(-x_i\beta)$ обозначается функция распределения случайной ошибки ε_i. Поскольку случайные ошибки одинаково распределены получаем $1 - F_{\varepsilon_i}(-x_i\beta) = F(x_i\beta)$.
- Если ε_i имеет стандартное нормальное распределение, то мы получаем пробит модель, а если стандартное логистическое – логит модель.

Пробит модель

Поскольку $\varepsilon_i \sim \mathcal{N}(0,1)$, то:

$$P(y_i = 1|x_i) = 1 - F_{\varepsilon_i}(-x_i\beta) = 1 - \Phi(-x_i\beta) = 1 - (1 - \Phi(x_i\beta)) = \Phi(x_i\beta)$$

Пробит модель

Поскольку $\varepsilon_i \sim \mathcal{N}(0,1)$, то:

$$P(y_i = 1|x_i) = 1 - F_{\varepsilon_i}(-x_i\beta) = 1 - \Phi(-x_i\beta) = 1 - (1 - \Phi(x_i\beta)) = \Phi(x_i\beta)$$

В результате логарифм функции правдоподобия имеет вид:

$$\ln L(\theta; y|X) = \sum_{i=1}^{n} y_i \ln\left(\Phi\left(x_i\beta\right)\right) + (1 - y_i)\ln\left(1 - \Phi\left(x_i\beta\right)\right)$$

Пробит модель

Поскольку $\varepsilon_i \sim \mathcal{N}(0,1)$, то:

$$P(y_i = 1|x_i) = 1 - F_{\varepsilon_i}(-x_i\beta) = 1 - \Phi(-x_i\beta) = 1 - (1 - \Phi(x_i\beta)) = \Phi(x_i\beta)$$

В результате логарифм функции правдоподобия имеет вид:

$$\ln L(\theta; y|X) = \sum_{i=1}^{n} y_i \ln\left(\Phi\left(x_i\beta\right)\right) + (1 - y_i)\ln\left(1 - \Phi\left(x_i\beta\right)\right)$$

Логит модель

Поскольку $\varepsilon_i \sim \text{Logistic}\,(0,1)$, то:

$$P(y_i = 1|x_i) = 1 - F_{\varepsilon_i}(-x_i\beta) = 1 - \Lambda(-x_i\beta) = 1 - (1 - \Lambda(x_i\beta)) = \Lambda(x_i\beta) = \frac{1}{1 + e^{-x_i\beta}}$$

Пробит модель

Поскольку $\varepsilon_i \sim \mathcal{N}(0,1)$, то:

$$P(y_i = 1|x_i) = 1 - F_{\varepsilon_i}(-x_i\beta) = 1 - \Phi(-x_i\beta) = 1 - (1 - \Phi(x_i\beta)) = \Phi(x_i\beta)$$

В результате логарифм функции правдоподобия имеет вид:

$$\ln L(\theta; y|X) = \sum_{i=1}^{n} y_i \ln\left(\Phi\left(x_i\beta\right)\right) + (1 - y_i)\ln\left(1 - \Phi\left(x_i\beta\right)\right)$$

Логит модель

Поскольку $\varepsilon_i \sim \text{Logistic}\,(0,1)$, то:

$$P(y_i = 1|x_i) = 1 - F_{\varepsilon_i}(-x_i\beta) = 1 - \Lambda(-x_i\beta) = 1 - (1 - \Lambda(x_i\beta)) = \Lambda(x_i\beta) = \frac{1}{1 + e^{-x_i\beta}}$$

В результате логарифм функции правдоподобия имеет вид:

$$\ln L(\theta; y|X) = \sum_{i=1}^{n} y_i \ln\left(\frac{1}{1 + e^{-x_i\beta}}\right) + (1 - y_i)\ln\left(1 - \frac{1}{1 + e^{-x_i\beta}}\right)$$

- Поскольку в пробит и логит моделях β оценивается при помощи метода максимального правдоподобия, то ММП оценка $\hat{\beta}$ будет асимптотически нормальной.

- Поскольку в пробит и логит моделях β оценивается при помощи метода максимального правдоподобия, то ММП оценка $\hat{\beta}$ будет асимптотически нормальной.
- Следовательно, для тестирования гипотез достаточно оценить асимптотическую ковариационную матрицу $\hat{\beta}$ используя любую стандартную технику, например, обратный гессиан, произведение якобианов, сэндвич или бутстрап.

- Поскольку в пробит и логит моделях β оценивается при помощи метода максимального правдоподобия, то ММП оценка $\hat{\beta}$ будет асимптотически нормальной.
- Следовательно, для тестирования гипотез достаточно оценить асимптотическую ковариационную матрицу $\hat{\beta}$ используя любую стандартную технику, например, обратный гессиан, произведение якобианов, сэндвич или бутстрап.
- Обозначим через $\hat{V}=\widehat{As.Cov}\left(\hat{\beta}\right)$ оценку асимптотической ковариационной матрицы, а через $\hat{V}_{ij}$ – ее элемент, расположенный в i-й строке и j-м столбце, в частности $\widehat{As.Var}(\hat{\beta}_k)=\hat{V}_{kk}$.

- Поскольку в пробит и логит моделях β оценивается при помощи метода максимального правдоподобия, то ММП оценка $\hat{\beta}$ будет асимптотически нормальной.
- Следовательно, для тестирования гипотез достаточно оценить асимптотическую ковариационную матрицу $\hat{\beta}$ используя любую стандартную технику, например, обратный гессиан, произведение якобианов, сэндвич или бутстрап.
- Обозначим через $\hat{V} = \widehat{As.Cov}\left(\hat{\beta}\right)$ оценку асимптотической ковариационной матрицы, а через $\hat{V}_{ij}$ – ее элемент, расположенный в i-й строке и j-м столбце, в частности $\widehat{As.Var}(\hat{\beta}_k) = \hat{V}_{kk}$. Тогда для тестирования гипотезы $H_0 : \beta_k = \beta_k^0$ против альтернативы $H_0 : \beta_k \neq \beta_k^0$ можно использовать следующую статистику:

$$T = \frac{\hat{\beta}_k - \beta_k^0}{\sqrt{\hat{V}_{kk}}}, \quad T|H_0 \xrightarrow{d} \mathcal{N}(0,1) \implies \text{p-value} = 2\min(\Phi(T), 1 - \Phi(T))$$

- Через A обозначим детермированную (не случайную) матрицу с r строками (число ограничений) и m столбцами (число регрессоров). Через 0_r обозначим вектор столбец, состоящий из r нулей.

- Через A обозначим детермированную (не случайную) матрицу с r строками (число ограничений) и m столбцами (число регрессоров). Через 0_r обозначим вектор столбец, состоящий из r нулей.
- Для тестирования гипотез вида $H_0 : A\beta_k = 0_r$ против альтернативы $H_0 : A\beta_k \neq 0_r$ можно использовать LR-тест, тест Вальда или LM-тест.

- Через A обозначим детермированную (не случайную) матрицу с r строками (число ограничений) и m столбцами (число регрессоров). Через 0_r обозначим вектор столбец, состоящий из r нулей.
- Для тестирования гипотез вида $H_0 : A\beta_k = 0_r$ против альтернативы $H_0 : A\beta_k \neq 0_r$ можно использовать LR-тест, тест Вальда или LM-тест.
- Во всех трех тестах при верной нулевой гипотезе тестовая статистика будет иметь асимптотическое распределение $\chi^2(r)$.

- Через A обозначим детермированную (не случайную) матрицу с r строками (число ограничений) и m столбцами (число регрессоров). Через 0_r обозначим вектор столбец, состоящий из r нулей.
- Для тестирования гипотез вида $H_0 : A\beta_k = 0_r$ против альтернативы $H_0 : A\beta_k \neq 0_r$ можно использовать LR-тест, тест Вальда или LM-тест.
- Во всех трех тестах при верной нулевой гипотезе тестовая статистика будет иметь асимптотическое распределение $\chi^2(r)$.
- LR-тест имеет самую простую тестовую статистику, но для его применения необходимо оценить и модель с ограничениями (накладываемыми нулевой гипотезой), и модель без ограничений.

- Через A обозначим детермированную (не случайную) матрицу с r строками (число ограничений) и m столбцами (число регрессоров). Через 0_r обозначим вектор столбец, состоящий из r нулей.
- Для тестирования гипотез вида $H_0 : A\beta_k = 0_r$ против альтернативы $H_0 : A\beta_k \neq 0_r$ можно использовать LR-тест, тест Вальда или LM-тест.
- Во всех трех тестах при верной нулевой гипотезе тестовая статистика будет иметь асимптотическое распределение $\chi^2(r)$.
- LR-тест имеет самую простую тестовую статистику, но для его применения необходимо оценить и модель с ограничениями (накладываемыми нулевой гипотезой), и модель без ограничений.
- Для проведения теста Вальда достаточно оценить лишь полную модель.

- Через A обозначим детермированную (не случайную) матрицу с r строками (число ограничений) и m столбцами (число регрессоров). Через 0_r обозначим вектор столбец, состоящий из r нулей.
- Для тестирования гипотез вида $H_0 : A\beta_k = 0_r$ против альтернативы $H_0 : A\beta_k \neq 0_r$ можно использовать LR-тест, тест Вальда или LM-тест.
- Во всех трех тестах при верной нулевой гипотезе тестовая статистика будет иметь асимптотическое распределение $\chi^2(r)$.
- LR-тест имеет самую простую тестовую статистику, но для его применения необходимо оценить и модель с ограничениями (накладываемыми нулевой гипотезой), и модель без ограничений.
- Для проведения теста Вальда достаточно оценить лишь полную модель.
- Для осуществления LM-теста достаточно оценить модель с ограничениями.

- Через x_{ki} обозначим k-й элемент (регрессор) вектора независимых переменных x_i. Предположим, что он измерен в непрерывной шкале.

- Через x_{ki} обозначим k-й элемент (регрессор) вектора независимых переменных x_i. Предположим, что он измерен в непрерывной шкале.
- Предельный эффект переменной x_{ki} на условную вероятность рассчитывается как:

$$\frac{\partial P(y_i = 1|x_i)}{\partial x_{ki}} = \frac{\partial\left(1 - F_{\varepsilon_i}(-x_i\beta)\right)}{\partial x_{ki}} = \beta_k f_{\varepsilon_i}(-x_i\beta)$$

Где $f_{\varepsilon_i}(.)$ это функция плотности ε_i.

- Через x_{ki} обозначим k-й элемент (регрессор) вектора независимых переменных x_i. Предположим, что он измерен в непрерывной шкале.
- Предельный эффект переменной x_{ki} на условную вероятность рассчитывается как:
$$\frac{\partial P(y_i = 1|x_i)}{\partial x_{ki}} = \frac{\partial\left(1 - F_{\varepsilon_i}(-x_i\beta)\right)}{\partial x_{ki}} = \beta_k f_{\varepsilon_i}(-x_i\beta)$$
Где $f_{\varepsilon_i}(.)$ это функция плотности ε_i.
- Знак предельного эффекта совпадает со знаком β_k, однако его величина зависит от линейного индекса $x_i\beta$.

- Через x_{ki} обозначим k-й элемент (регрессор) вектора независимых переменных x_i. Предположим, что он измерен в непрерывной шкале.
- Предельный эффект переменной x_{ki} на условную вероятность рассчитывается как:
$$\frac{\partial P(y_i=1|x_i)}{\partial x_{ki}}=\frac{\partial\left(1-F_{\varepsilon_i}(-x_i\beta)\right)}{\partial x_{ki}}=\beta_k f_{\varepsilon_i}(-x_i\beta)$$
Где $f_{\varepsilon_i}(.)$ это функция плотности ε_i.
- Знак предельного эффекта совпадает со знаком β_k, однако его величина зависит от линейного индекса $x_i\beta$.
- В случае пробит и логит моделей получаем:

Пробит модель: $$\frac{\partial P(y_i=1|x_i)}{\partial x_{ki}}=\beta_k\phi(-x_i\beta)=\beta_k\phi(x_i\beta)=\beta_k\frac{1}{\sqrt{2\pi}}e^{-\frac{(x_i\beta)^2}{2}}$$

- Через x_{ki} обозначим k-й элемент (регрессор) вектора независимых переменных x_i. Предположим, что он измерен в непрерывной шкале.
- Предельный эффект переменной x_{ki} на условную вероятность рассчитывается как:
$$\frac{\partial P(y_i = 1|x_i)}{\partial x_{ki}} = \frac{\partial\left(1 - F_{\varepsilon_i}(-x_i\beta)\right)}{\partial x_{ki}} = \beta_k f_{\varepsilon_i}(-x_i\beta)$$
Где $f_{\varepsilon_i}(.)$ это функция плотности ε_i.
- Знак предельного эффекта совпадает со знаком β_k, однако его величина зависит от линейного индекса $x_i\beta$.
- В случае пробит и логит моделей получаем:

Пробит модель: $$\frac{\partial P(y_i = 1|x_i)}{\partial x_{ki}} = \beta_k\phi(-x_i\beta) = \beta_k\phi(x_i\beta) = \beta_k\frac{1}{\sqrt{2\pi}}e^{-\frac{(x_i\beta)^2}{2}}$$

Логит модель: $$\frac{\partial P(y_i = 1|x_i)}{\partial x_{ki}} = \partial\left(\frac{1}{1 + e^{-x_i\beta}}\right)/\partial x_{ki} = \beta_k\frac{e^{-x_i\beta}}{(1 + e^{-x_i\beta})^2}$$

- Через $x_{(-k)i}$ обозначим вектор независимых переменных без k-й компоненты. По аналогии через $\beta_{(-k)}$ обозначим вектор регрессионных коэффициентов без k-го элемента.

- Через $x_{(-k)i}$ обозначим вектор независимых переменных без k-й компоненты. По аналогии через $\beta_{(-k)}$ обозначим вектор регрессионных коэффициентов без k-го элемента.
- Допустим, что x_{ki} является бинарной переменной. Тогда ее предельный эффект на условную вероятность рассчитывается как разница в условных вероятностях при единичном и нулевом значении бинарного регрессора (при прочих равных):

$$P(y_i = 1|x_{ki} = 1, x_{(-k)i}) - P(y_i = 1|x_{ki} = 0, x_{(-k)i})$$

- Через $x_{(-k)i}$ обозначим вектор независимых переменных без k-й компоненты. По аналогии через $\beta_{(-k)}$ обозначим вектор регрессионных коэффициентов без k-го элемента.
- Допустим, что x_{ki} является бинарной переменной. Тогда ее предельный эффект на условную вероятность рассчитывается как разница в условных вероятностях при единичном и нулевом значении бинарного регрессора (при прочих равных):

$$P(y_i = 1|x_{ki} = 1, x_{(-k)i}) - P(y_i = 1|x_{ki} = 0, x_{(-k)i})$$

- В случае пробит и логит моделей получаем:

$$\textbf{Пробит модель: } \Phi\left(x_{(-k)i}\beta_{(-k)} + \beta_k\right) - \Phi\left(x_{(-k)i}\beta_{(-k)}\right)$$

- Через $x_{(-k)i}$ обозначим вектор независимых переменных без k-й компоненты. По аналогии через $\beta_{(-k)}$ обозначим вектор регрессионных коэффициентов без k-го элемента.
- Допустим, что x_{ki} является бинарной переменной. Тогда ее предельный эффект на условную вероятность рассчитывается как разница в условных вероятностях при единичном и нулевом значении бинарного регрессора (при прочих равных):

$$P(y_i = 1|x_{ki} = 1, x_{(-k)i}) - P(y_i = 1|x_{ki} = 0, x_{(-k)i})$$

- В случае пробит и логит моделей получаем:

Пробит модель: $\Phi\left(x_{(-k)i}\beta_{(-k)} + \beta_k\right) - \Phi\left(x_{(-k)i}\beta_{(-k)}\right)$

Логит модель: $\Lambda\left(x_{(-k)i}\beta_{(-k)} + \beta_k\right) - \Lambda\left(x_{(-k)i}\beta_{(-k)}\right)$

- Через $x_{(-k)i}$ обозначим вектор независимых переменных без k-й компоненты. По аналогии через $\beta_{(-k)}$ обозначим вектор регрессионных коэффициентов без k-го элемента.
- Допустим, что x_{ki} является бинарной переменной. Тогда ее предельный эффект на условную вероятность рассчитывается как разница в условных вероятностях при единичном и нулевом значении бинарного регрессора (при прочих равных):

$$P(y_i = 1|x_{ki} = 1, x_{(-k)i}) - P(y_i = 1|x_{ki} = 0, x_{(-k)i})$$

- В случае пробит и логит моделей получаем:

$$\textbf{Пробит модель: } \Phi\left(x_{(-k)i}\beta_{(-k)} + \beta_k\right) - \Phi\left(x_{(-k)i}\beta_{(-k)}\right)$$

$$\textbf{Логит модель: } \Lambda\left(x_{(-k)i}\beta_{(-k)} + \beta_k\right) - \Lambda\left(x_{(-k)i}\beta_{(-k)}\right)$$

- Для расчета предельных эффектов по порядковым и номинальным переменным можно по аналогии рассчитывать разницу в вероятностях при различных значениях соответствующих переменных и фиксированных значениях остальных.

- Допустим вы оценили как стаж (в годах) и факт наличия высшего образования (бинарный регрессор) влияют на вероятность того, что индивид работает (бинарная зависимая переменная). Для краткости присвоим константе, а также переменным на стаж и образование индексы 1, 2 и 3 соответственно.

- Допустим вы оценили как стаж (в годах) и факт наличия высшего образования (бинарный регрессор) влияют на вероятность того, что индивид работает (бинарная зависимая переменная). Для краткости присвоим константе, а также переменным на стаж и образование индексы 1, 2 и 3 соответственно.
- По результатам оценивания с помощью пробит модели оказалось, что $\hat{\beta} = (1, 0.1, 0.2)$.

Классические модели бинарного выбора

Пример оценивания вероятностей и предельных эффектов

- Допустим вы оценили как стаж (в годах) и факт наличия высшего образования (бинарный регрессор) влияют на вероятность того, что индивид работает (бинарная зависимая переменная). Для краткости присвоим константе, а также переменным на стаж и образование индексы 1, 2 и 3 соответственно.
- По результатам оценивания с помощью пробит модели оказалось, что $\hat{\beta} = (1, 0.1, 0.2)$.
- У Лаврентия имеется высшее образование и пятилетний стаж работы, то есть $x_i = (1, 5, 1)$. Поэтому оценка линейного индекса для него равна $x_i\hat{\beta} = 1 \times 1 + 0.1 \times 5 + 0.2 \times 1 = 1.7$, а оценка вероятности занятости составит $\hat{P}(y_i = 1|x_i) = \Phi(1.7) \approx 0.955$.

- Допустим вы оценили как стаж (в годах) и факт наличия высшего образования (бинарный регрессор) влияют на вероятность того, что индивид работает (бинарная зависимая переменная). Для краткости присвоим константе, а также переменным на стаж и образование индексы 1, 2 и 3 соответственно.
- По результатам оценивания с помощью пробит модели оказалось, что $\hat{\beta} = (1, 0.1, 0.2)$.
- У Лаврентия имеется высшее образование и пятилетний стаж работы, то есть $x_i = (1, 5, 1)$. Поэтому оценка линейного индекса для него равна $x_i\hat{\beta} = 1 \times 1 + 0.1 \times 5 + 0.2 \times 1 = 1.7$, а оценка вероятности занятости составит $\hat{P}(y_i = 1|x_i) = \Phi(1.7) \approx 0.955$.
- Предельный эффект стажа на условную вероятность занятости окажется равен $0.1 \times \phi(1.7) \approx 0.0955$. То есть при прочих равных при увеличении стажа на 1 год вероятность занятости Лаврентия возрастет, приблизительно, на 0.0955.

- Допустим вы оценили как стаж (в годах) и факт наличия высшего образования (бинарный регрессор) влияют на вероятность того, что индивид работает (бинарная зависимая переменная). Для краткости присвоим константе, а также переменным на стаж и образование индексы 1, 2 и 3 соответственно.
- По результатам оценивания с помощью пробит модели оказалось, что $\hat{\beta} = (1, 0.1, 0.2)$.
- У Лаврентия имеется высшее образование и пятилетний стаж работы, то есть $x_i = (1, 5, 1)$. Поэтому оценка линейного индекса для него равна $x_i\hat{\beta} = 1 \times 1 + 0.1 \times 5 + 0.2 \times 1 = 1.7$, а оценка вероятности занятости составит $\hat{P}(y_i = 1|x_i) = \Phi(1.7) \approx 0.955$.
- Предельный эффект стажа на условную вероятность занятости окажется равен $0.1 \times \phi(1.7) \approx 0.0955$. То есть при прочих равных при увеличении стажа на 1 год вероятность занятости Лаврентия возрастет, приблизительно, на 0.0955.
- Поскольку предельные эффекты непрерывных переменных считаются как производные, то интерпретация в терминах изменения на единицу будет точна лишь при достаточно мелких единицах измерения регрессора.

- Допустим вы оценили как стаж (в годах) и факт наличия высшего образования (бинарный регрессор) влияют на вероятность того, что индивид работает (бинарная зависимая переменная). Для краткости присвоим константе, а также переменным на стаж и образование индексы 1, 2 и 3 соответственно.
- По результатам оценивания с помощью пробит модели оказалось, что $\hat{\beta} = (1, 0.1, 0.2)$.
- У Лаврентия имеется высшее образование и пятилетний стаж работы, то есть $x_i = (1, 5, 1)$. Поэтому оценка линейного индекса для него равна $x_i\hat{\beta} = 1 \times 1 + 0.1 \times 5 + 0.2 \times 1 = 1.7$, а оценка вероятности занятости составит $\hat{P}(y_i = 1|x_i) = \Phi(1.7) \approx 0.955$.
- Предельный эффект стажа на условную вероятность занятости окажется равен $0.1 \times \phi(1.7) \approx 0.0955$. То есть при прочих равных при увеличении стажа на 1 год вероятность занятости Лаврентия возрастет, приблизительно, на 0.0955.
- Поскольку предельные эффекты непрерывных переменных считаются как производные, то интерпретация в терминах изменения на единицу будет точна лишь при достаточно мелких единицах измерения регрессора.
- При прочих равных при условии наличия высшего образования Лаврентий будет работать с вероятностью на $\Phi(1.7) - \Phi(1.7 - 0.2) \approx 0.022$ большей, чем в случае, если у Лаврентия не будет высшего образования.

- В силу инвариантности ММП оценок для того, чтобы получить асимптотически эффективную и асимптотически нормальную оценку предельного эффекта достаточно подставить $\hat{\beta}$ вместо β в формулы для предельных эффектов.

- В силу инвариантности ММП оценок для того, чтобы получить асимптотически эффективную и асимптотически нормальную оценку предельного эффекта достаточно подставить $\hat{\beta}$ вместо β в формулы для предельных эффектов.
- Благодаря асимптотической нормальности оценок предельных эффектов для тестирования гипотез достаточно оценить асимптотическую дисперсию оценки предельного эффекта с помощью дельта метода.

- В силу инвариантности ММП оценок для того, чтобы получить асимптотически эффективную и асимптотически нормальную оценку предельного эффекта достаточно подставить $\hat{\beta}$ вместо β в формулы для предельных эффектов.
- Благодаря асимптотической нормальности оценок предельных эффектов для тестирования гипотез достаточно оценить асимптотическую дисперсию оценки предельного эффекта с помощью дельта метода.
- Например, поскольку $\partial\phi(t)/\partial t = -t\phi(t)$, то в пробит модели получаем:

$$\widehat{As.Var}\left(\frac{\partial P(\widehat{y_i = 1}|x_i)}{\partial x_{ki}}\right) = g_i^T \widehat{As.Cov}\left(\hat{\beta}\right) g_i$$

- В силу инвариантности ММП оценок для того, чтобы получить асимптотически эффективную и асимптотически нормальную оценку предельного эффекта достаточно подставить $\hat{\beta}$ вместо β в формулы для предельных эффектов.
- Благодаря асимптотической нормальности оценок предельных эффектов для тестирования гипотез достаточно оценить асимптотическую дисперсию оценки предельного эффекта с помощью дельта метода.
- Например, поскольку $\partial\phi(t)/\partial t = -t\phi(t)$, то в пробит модели получаем:

$$\widehat{As.Var}\left(\frac{\partial P(\widehat{y_i = 1}|x_i)}{\partial x_{ki}}\right) = {g_i}^T \widehat{As.Cov}\left(\hat{\beta}\right) g_i$$

$${g_i}^T = \left(\triangle_{\hat{\beta}}\phi(x_i\hat{\beta})\right)^T = \left(\frac{\partial\beta_k\phi(x_i\hat{\beta})}{\partial\beta_1}, \ldots, \frac{\partial\beta_k\phi(x_i\hat{\beta})}{\partial\hat{\beta}_m}\right) =$$

- В силу инвариантности ММП оценок для того, чтобы получить асимптотически эффективную и асимптотически нормальную оценку предельного эффекта достаточно подставить $\hat{\beta}$ вместо β в формулы для предельных эффектов.
- Благодаря асимптотической нормальности оценок предельных эффектов для тестирования гипотез достаточно оценить асимптотическую дисперсию оценки предельного эффекта с помощью дельта метода.
- Например, поскольку $\partial\phi(t)/\partial t = -t\phi(t)$, то в пробит модели получаем:

$$\widehat{As.Var}\left(\frac{\partial P(\widehat{y_i = 1}|x_i)}{\partial x_{ki}}\right) = g_i^T \widehat{As.Cov}\left(\hat{\beta}\right) g_i$$

$$g_i^T = \left(\triangle_{\hat{\beta}}\phi(x_i\hat{\beta})\right)^T = \left(\frac{\partial\beta_k\phi(x_i\hat{\beta})}{\partial\beta_1}, \ldots, \frac{\partial\beta_k\phi(x_i\hat{\beta})}{\partial\hat{\beta}_m}\right) =$$

$$= \left(-x_{1i}\hat{\beta}_k\phi(x_i\hat{\beta}), \ldots, (1 - x_{ki}\hat{\beta}_k)\phi(x_i\hat{\beta}), \ldots., -x_{mi}\hat{\beta}_k\phi(x_m\hat{\beta})\right)$$

- Поскольку предельные эффекты зависят от линейного индекса, то для каждого наблюдения в выборке предельные эффекты будут различаться. Существуют два основных подхода к тому, как агрегировать информацию о предельных эффектах.

- Поскольку предельные эффекты зависят от линейного индекса, то для каждого наблюдения в выборке предельные эффекты будут различаться. Существуют два основных подхода к тому, как агрегировать информацию о предельных эффектах.
- **Средний предельный эффект** (СПЭ) оценивается как среднее значение по всем оценкам предельных эффектов соответствующей переменной:

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\partial P(\widehat{y_i = 1}|x_i)}{\partial x_{ki}} = \frac{\hat{\beta}_k}{n}\sum_{i=1}^{n} f_{\varepsilon_i}(-x_i\hat{\beta})$$

- Поскольку предельные эффекты зависят от линейного индекса, то для каждого наблюдения в выборке предельные эффекты будут различаться. Существуют два основных подхода к тому, как агрегировать информацию о предельных эффектах.
- **Средний предельный эффект** (СПЭ) оценивается как среднее значение по всем оценкам предельных эффектов соответствующей переменной:

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\partial P(\widehat{y_i = 1}|x_i)}{\partial x_{ki}} = \frac{\hat{\beta}_k}{n}\sum_{i=1}^{n} f_{\varepsilon_i}(-x_i\hat{\beta})$$

- **Предельный эффект для среднего наблюдения** (ПЭС) предполагает, что предельный эффект оценивается в точке, соответствующей некоторому среднему значению регрессоров.

- Поскольку предельные эффекты зависят от линейного индекса, то для каждого наблюдения в выборке предельные эффекты будут различаться. Существуют два основных подхода к тому, как агрегировать информацию о предельных эффектах.
- **Средний предельный эффект** (СПЭ) оценивается как среднее значение по всем оценкам предельных эффектов соответствующей переменной:

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\partial P(\widehat{y_i = 1}|x_i)}{\partial x_{ki}} = \frac{\hat{\beta}_k}{n}\sum_{i=1}^{n} f_{\varepsilon_i}(-x_i\hat{\beta})$$

- **Предельный эффект для среднего наблюдения** (ПЭС) предполагает, что предельный эффект оценивается в точке, соответствующей некоторому среднему значению регрессоров.
- Для этого сперва считают среднее значение для каждого регрессора, измеренного в непрерывной шкале $\overline{x}_k$. Для порядковых регрессоров либо считают среднее значение, либо берут наиболее часто встречающееся значение (моду). Для номинальных регрессоров либо берут наиболее часто встречающееся значение, либо считают ПЭС по подвыборкам для каждого значения этого регрессора (например, отдельно считают предельные эффекты на вероятность занятости для представителей различных профессий).

- Поскольку предельные эффекты зависят от линейного индекса, то для каждого наблюдения в выборке предельные эффекты будут различаться. Существуют два основных подхода к тому, как агрегировать информацию о предельных эффектах.
- **Средний предельный эффект** (СПЭ) оценивается как среднее значение по всем оценкам предельных эффектов соответствующей переменной:

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\partial P(\widehat{y_i = 1}|x_i)}{\partial x_{ki}} = \frac{\hat{\beta}_k}{n}\sum_{i=1}^{n}f_{\varepsilon_i}(-x_i\hat{\beta})$$

- **Предельный эффект для среднего наблюдения** (ПЭС) предполагает, что предельный эффект оценивается в точке, соответствующей некоторому среднему значению регрессоров.
- Для этого сперва считают среднее значение для каждого регрессора, измеренного в непрерывной шкале $\overline{x}_k$. Для порядковых регрессоров либо считают среднее значение, либо берут наиболее часто встречающееся значение (моду). Для номинальных регрессоров либо берут наиболее часто встречающееся значение, либо считают ПЭС по подвыборкам для каждого значения этого регрессора (например, отдельно считают предельные эффекты на вероятность занятости для представителей различных профессий).
- Для тестирования гипотез о СПЭ и ПЭС также можно применить дельта-метод.

- Модели бинарного выбора (например, логит и пробит) могут сравниваться по информационным критериям, в частности, по AIC и BIC.

$$\text{AIC} = 2\left(m - \ln L(\hat{\beta}, y|X)\right)$$

$$\text{BIC} = \ln(n)m - 2\ln L(\hat{\beta}, y|X))$$

Чем меньше значение соответствующих критериев, тем более качественной считается модель.

- Модели бинарного выбора (например, логит и пробит) могут сравниваться по информационным критериям, в частности, по AIC и BIC.

$$\text{AIC} = 2\left(m - \ln L(\hat{\beta}, y|X)\right)$$

$$\text{BIC} = \ln(n)m - 2\ln L(\hat{\beta}, y|X))$$

  Чем меньше значение соответствующих критериев, тем более качественной считается модель.

- Для предсказания значений бинарных переменных как правило предполагают, что предсказанное значение принимает единичное значение, если оценка его условной вероятности превышает некоторое пороговое значение p^*. Обычно предполагают $p^* = 0.5$:

$$\hat{y}_i = \begin{cases} 1, \text{ если } \hat{P}\left(y_i = 1|x_i\right) \geq p^* \\ 0, \text{ в противном случае} \end{cases}$$

- Модели бинарного выбора (например, логит и пробит) могут сравниваться по информационным критериям, в частности, по AIC и BIC.

$$\text{AIC} = 2\left(m - \ln L(\hat{\beta}, y|X)\right)$$

$$\text{BIC} = \ln(n)m - 2\ln L(\hat{\beta}, y|X))$$

  Чем меньше значение соответствующих критериев, тем более качественной считается модель.

- Для предсказания значений бинарных переменных как правило предполагают, что предсказанное значение принимает единичное значение, если оценка его условной вероятности превышает некоторое пороговое значение p^*. Обычно предполагают $p^* = 0.5$:

$$\hat{y}_i = \begin{cases} 1, \text{ если } \hat{P}\left(y_i = 1|x_i\right) \geq p^* \\ 0, \text{ в противном случае} \end{cases}$$

- Качество моделей можно также сравнивать по точности прогнозов, рассчитываемой как доля верных предсказаний:

$$\frac{1}{n}\sum_{i=1}^{n} I(\hat{y}_i = y_i), \qquad I(\hat{y}_i = y_i) = \begin{cases} 1, \text{ если } \hat{y}_i = y_i \\ 0, \text{ в противном случае} \end{cases}$$

Особые случаи предельных эффектов

Переменные взаимодействия

- Для простоты рассмотрим случай с двумя переменными, отражающими стаж работы x_1 (в годах) и пол индивида x_2 (1 - мужчина, 0 - женщина), объясняющими занятость индивида y_i (1 - работает, 0 - не работает). Запишем выражение для латентной переменной, отражающей склонность к занятости:

$$y_i^* = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{1i} x_{2i} + \varepsilon_i = \beta_0 + \beta_2 x_{2i} + (\beta_1 + \beta_3 x_{2i})\, x_{1i} + \varepsilon_i$$

- Для простоты рассмотрим случай с двумя переменными, отражающими стаж работы x_1 (в годах) и пол индивида x_2 (1 - мужчина, 0 - женщина), объясняющими занятость индивида y_i (1 - работает, 0 - не работает). Запишем выражение для латентной переменной, отражающей склонность к занятости:

$$y_i^* = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{1i} x_{2i} + \varepsilon_i = \beta_0 + \beta_2 x_{2i} + (\beta_1 + \beta_3 x_{2i})\, x_{1i} + \varepsilon_i$$

- Запишем предельный эффект стажа x_{1i} на условную вероятность занятости:

$$\frac{\partial P(y_i = 1|x_i)}{\partial x_{1i}} = (\beta_1 + \beta_3 x_{2i})\, f_{\varepsilon_i}(-x_i\beta)$$

Особые случаи предельных эффектов

Переменные взаимодействия

- Для простоты рассмотрим случай с двумя переменными, отражающими стаж работы x_1 (в годах) и пол индивида x_2 (1 - мужчина, 0 - женщина), объясняющими занятость индивида y_i (1 - работает, 0 - не работает). Запишем выражение для латентной переменной, отражающей склонность к занятости:

$$y_i^* = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{1i} x_{2i} + \varepsilon_i = \beta_0 + \beta_2 x_{2i} + (\beta_1 + \beta_3 x_{2i})\, x_{1i} + \varepsilon_i$$

- Запишем предельный эффект стажа x_{1i} на условную вероятность занятости:

$$\frac{\partial P(y_i = 1|x_i)}{\partial x_{1i}} = (\beta_1 + \beta_3 x_{2i})\, f_{\varepsilon_i}(-x_i\beta)$$

- Таким образом, для мужчин $(x_{2i} = 1)$ предельный эффект стажа на занятость составляет $(\beta_1 + \beta_3) f_{\varepsilon_i}(-x_i\beta)$, а для женщин $(x_{2i} = 0)$ он будет равен $\beta_1 f_{\varepsilon_i}(-x_i\beta)$. Если $\beta_1 > 0$ и $\beta_3 > 0$, то это можно интерпретировать как то, что при прочих равных влияние стажа на вероятность занятости для мужчин сильнее, чем для женщин, что, например, может быть обусловлено дискриминацией на рынке труда.

Особые случаи предельных эффектов

Нелинейный эффект

- Для простоты рассмотрим случай с двумя переменными, отражающими возраст x_1 (в годах) и доход индивида x_2 (в рублях), объясняющими вероятность занятий спортом y_i (1 - занимается, 0 - не занимается). Запишем выражение для латентной переменной, отражающей склонность к занятиям спортом:

$$y_i^* = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{1i}^2 + \beta_3 x_{2i} + \varepsilon_i$$

Особые случаи предельных эффектов

Нелинейный эффект

- Для простоты рассмотрим случай с двумя переменными, отражающими возраст x_1 (в годах) и доход индивида x_2 (в рублях), объясняющими вероятность занятий спортом y_i (1 - занимается, 0 - не занимается). Запишем выражение для латентной переменной, отражающей склонность к занятиям спортом:

$$y_i^* = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{1i}^2 + \beta_3 x_{2i} + \varepsilon_i$$

- Запишем предельный эффект возраста x_{1i} на условную вероятность занятий спортом:

$$\frac{\partial P(y_i = 1|x_i)}{\partial x_{1i}} = (\beta_1 + 2\beta_2 x_{1i})\, f_{\varepsilon_i}(-x_i\beta)$$

Особые случаи предельных эффектов

Нелинейный эффект

- Для простоты рассмотрим случай с двумя переменными, отражающими возраст x_1 (в годах) и доход индивида x_2 (в рублях), объясняющими вероятность занятий спортом y_i (1 - занимается, 0 - не занимается). Запишем выражение для латентной переменной, отражающей склонность к занятиям спортом:

$$y_i^* = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{1i}^2 + \beta_3 x_{2i} + \varepsilon_i$$

- Запишем предельный эффект возраста x_{1i} на условную вероятность занятий спортом:

$$\frac{\partial P(y_i = 1|x_i)}{\partial x_{1i}} = (\beta_1 + 2\beta_2 x_{1i}) f_{\varepsilon_i}(-x_i\beta)$$

- Таким образом, знак предельного эффекта возраста может изменяться в зависимости от возраста индивида. Например, если $\beta_1 = 100$ и $\beta_2 = -1$, то, при прочих равных, вероятность занятий спортом будет увеличиваться (поскольку предельный эффект положительный) с возрастом вплоть до достижения индивидом 50-ти лет, после чего – убывать (так как предельный эффект отрицательный).

Классические модели бинарного выбора

Тестирование гипотезы о возможности оценивания общей модели

- Предположим, что исследователь оценивает влияние различных характеристик на вероятность того, что индивид работает.

- Предположим, что исследователь оценивает влияние различных характеристик на вероятность того, что индивид работает.
- Необходимо проверить, можно ли оценивать общую модель для мужчин и женщин, то есть $H_0 : \beta_M = \beta_F$, где β_M и β_F отражают влияние различных характеристик на вероятность занятости для мужчин и для женщин соответственно.

- Предположим, что исследователь оценивает влияние различных характеристик на вероятность того, что индивид работает.
- Необходимо проверить, можно ли оценивать общую модель для мужчин и женщин, то есть $H_0 : \beta_M = \beta_F$, где β_M и β_F отражают влияние различных характеристик на вероятность занятости для мужчин и для женщин соответственно.
- Во-первых, можно ввести переменные взаимодействия между полом и каждым регрессором, а затем протестировать совместную значимость соответствующих переменных взаимодействия, например, тестом Вальда.

- Предположим, что исследователь оценивает влияние различных характеристик на вероятность того, что индивид работает.
- Необходимо проверить, можно ли оценивать общую модель для мужчин и женщин, то есть $H_0 : \beta_M = \beta_F$, где β_M и β_F отражают влияние различных характеристик на вероятность занятости для мужчин и для женщин соответственно.
- Во-первых, можно ввести переменные взаимодействия между полом и каждым регрессором, а затем протестировать совместную значимость соответствующих переменных взаимодействия, например, тестом Вальда.
- Во-вторых, можно, сперва, оценить модель бинарного выбора по всем индивидам, то есть рассчитать $\hat{\beta}$ предполагая, что $\beta = \beta_M = \beta_F$. Затем можно оценить по отдельности модели по выборкам из мужчин и женщин, что позволит получить оценки $\hat{\beta}_M$ и $\hat{\beta}_F$.

Классические модели бинарного выбора

Тестирование гипотезы о возможности оценивания общей модели

- Предположим, что исследователь оценивает влияние различных характеристик на вероятность того, что индивид работает.
- Необходимо проверить, можно ли оценивать общую модель для мужчин и женщин, то есть $H_0 : \beta_M = \beta_F$, где β_M и β_F отражают влияние различных характеристик на вероятность занятости для мужчин и для женщин соответственно.
- Во-первых, можно ввести переменные взаимодействия между полом и каждым регрессором, а затем протестировать совместную значимость соответствующих переменных взаимодействия, например, тестом Вальда.
- Во-вторых, можно, сперва, оценить модель бинарного выбора по всем индивидам, то есть рассчитать $\hat{\beta}$ предполагая, что $\beta = \beta_M = \beta_F$. Затем можно оценить по отдельности модели по выборкам из мужчин и женщин, что позволит получить оценки $\hat{\beta}_M$ и $\hat{\beta}_F$.
- Таким образом, логарифм правдоподобия ограниченной модели будет равен $\ln L(\hat{\beta})$, а в полной, то есть учитывающей различие в коэффициентах для мужчин и женщин $\ln L(\hat{\beta}_M) + \ln L(\hat{\beta}_F)$. Для тестирования гипотезы достаточно воспользоваться LR-тестом, тестовая статистика которого при верной нулевой гипотезе будет иметь Хи-квадрат распределение с числом степеней свободы, совпадающим с числом регрессионных коэффициентов m (поскольку в ограниченной модели мы накладываем ограничение на все коэффициенты, включая константу).