ЛЕКЦИЯ 20

МОДЕЛИ БИНАРНОГО ВЫБОРА

Часть I

1. Когда целесообразно применять модели бинарного выбора? Несколько примеров
2. Линейная модель вероятности. Проблемы, связанные с ее использованием
3. Logit-модель
4. Probit-модель

Когда целесообразно применять модели бинарного выбора? Несколько примеров

Ранее в качестве зависимой переменной Y мы предполагали некий количественный признак, который может принимать непрерывное множество значений.
Кроме того, в классической регрессионной модели предполагалось, что ошибки имеют стандартное нормальное распределение.

В то же время часто бывает, что зависимая переменная является дискретной.
Примеры:

- Покупать ли автомобиль;
- Идти ли на выборы;

- Способ попадания из дома на работу (пешком, на метро, наземным общественным транспортом или на личном автомобиле);
- Выбрать ли для отдыха авиа – или автобусный тур;
- Переехать ли на постоянное место жительства в другой регион и т.п.

Если есть только два возможных значения зависимой переменной, то такие модели будут называться моделями бинарного выбора.

Для таких моделей зависимая переменная, которая может иметь нечисловую переменную, принимает значения 0 или 1.

Формально применение МНК для таких моделей возможно, но есть несколько «подводных камней».

Рассмотрим линейную вероятностную модель и те проблемы, которые возникают при ее интерпретации.

Линейная вероятностная модель

Linear probability model

В рамках этой модели:

$$Y_t = \begin{cases} 1, & \textit{если событие произошло}; \\ 0, & \textit{если не произошло}. \end{cases}$$

$$Y_i = \beta_1 + \beta_2 X_i + u_i, \ E(u_i) = 0$$

$$Y_i = E(Y_i | X_i) + u_i$$

Таким образом, мы представляем Y как сумму детерминированной части и случайного возмущения. Пока формально все как и прежде.
Но: Y может принимать только два значения 0 и 1.

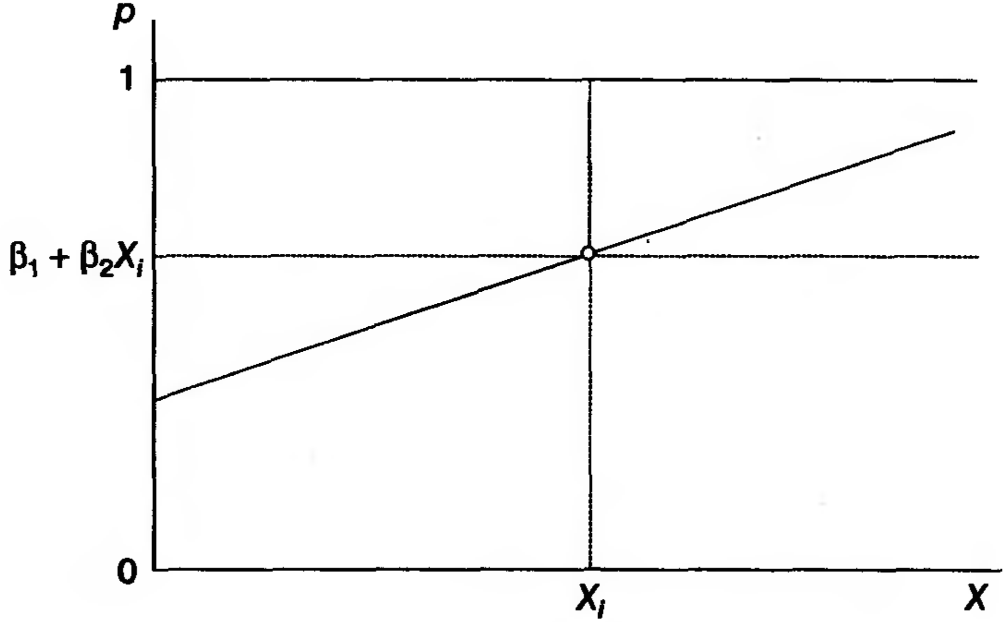
Найдем математическое ожидание Y:

$$E(Y_i) = 0 \cdot P(Y_i = 0) + 1 \cdot P(Y_i = 1) =$$
$$= 0 \cdot (1 - p_i) + 1 \cdot p_i = p_i$$

Откуда, $p_i = \beta_1 + \beta_2 X_i$

Значит, выражение $\beta_1 + \beta_2 X_i$ имеет смысл вероятности, и, следовательно, значения выражения должны лежать в пределах $[0;\ 1]$

Графическое представление:



Основные <u>недостатки</u> линейной вероятностной модели:

- Остатки распределены не нормально: их распределение и вовсе является дискретным

Если $Y_i = 1$, то $u_i = 1 - \beta_1 - \beta_2 X_i$

Если же $Y_i = 0$, то $u_i = -\beta_1 - \beta_2 X_i$

- Дисперсия остатков зависит от значений регрессора:

$$Var(u_i) = (\beta_1 + \beta_2 X_i)(1 - \beta_1 - \beta_2 X_i)$$

!!!! Выведите формулу для дисперсии в качестве упражнения самостоятельно, составив предварительно закон распределения для u_i

Это обозначает, что у нас есть проблема гетероскедастичности, следовательно, оценки параметров, полученные обычным МНК, больше не будут эффективными, и желательно тогда пользоваться ОМНК.

- Коэффициент β_1 практически всегда является неинтерпретируемым.

- Поскольку $\beta_1 + \beta_2 X_i$ имеет смысл вероятности, то мы не должны выходить за пределы $[0;\ 1]$, а напрямую в эту модель не заложены никакие ограничения. Следовательно, необходимо наложить либо ограничения на регрессор, либо на параметры, чтобы остаться в рамках $[0;\ 1]$

В силу перечисленных недостатков, линейная вероятностная модель редко находит применение на практике.

Основные предположения, лежащие в основе logit- и probit-моделей

Обойти проблемы линейной модели можно, сделав несколько предположений:

Предположим, что

$$P(Y_i = 1) = F(\beta_1 + \beta_2 X_i),$$

где F есть некоторая функция, у которой область значений лежит в пределах $[0;\ 1]$

Естественно предположить, что в качестве такой функции может быть использована <u>функция распределения</u> некоторой случайной величины.

Зададим эту функцию путем введения <u>латентной</u> (ненаблюдаемой) <u>переменной</u>

$$Y_i^* = \beta_1 + \beta_2 X_i + \varepsilon_i,\ \ \varepsilon_i \sim iid(0; \sigma^2), i = 1,\ldots,n \qquad \textbf{(*)}$$

Причем:

$$Y_i = 1, \ \textit{если} \ \ Y_i^* \geq 0, \quad i = 1, \ldots, n;$$

$$Y_i = 0, \ \textit{если} \ \ Y_i^* < 0, \quad i = 1, \ldots, n$$

В модели (*) мы ввели, что ошибки независимы и одинаково распределены.

Определим в дальнейшем, что F будет функцией распределения нормированной случайной ошибки, причем функция плотности является симметричной.

Тогда

$$P(Y_i = 1) = P(Y_i^* \geq 0) = P(\beta_1 + \beta_2 X_i + \varepsilon_i \geq 0) =$$

$$= P(\varepsilon_i \geq -\beta_1 - \beta_2 X_i) = P\left({\varepsilon_i}/{\sigma} \leq {(\beta_1 + \beta_2 X_i)}/{\sigma} \right) =$$

$$= F\left({(\beta_1 + \beta_2 X_i)}/{\sigma} \right)$$

В описанной модели параметры $\beta_{1,2}$ и σ участвуют только в отношении, так что по отдельности не могут быть идентифицированы. Поэтому без ограничения общности можно считать, что σ=1.

В зависимости от того, какую именно функцию F мы выберем, получим логит- или пробит-модели.

Logit-модель

В качестве функции F выбирается функция логистического распределения (S-образная кривая):

$$p = F(Z) = \frac{1}{1+e^{-Z}}, \ Z = \beta_1 + \beta_2 X$$

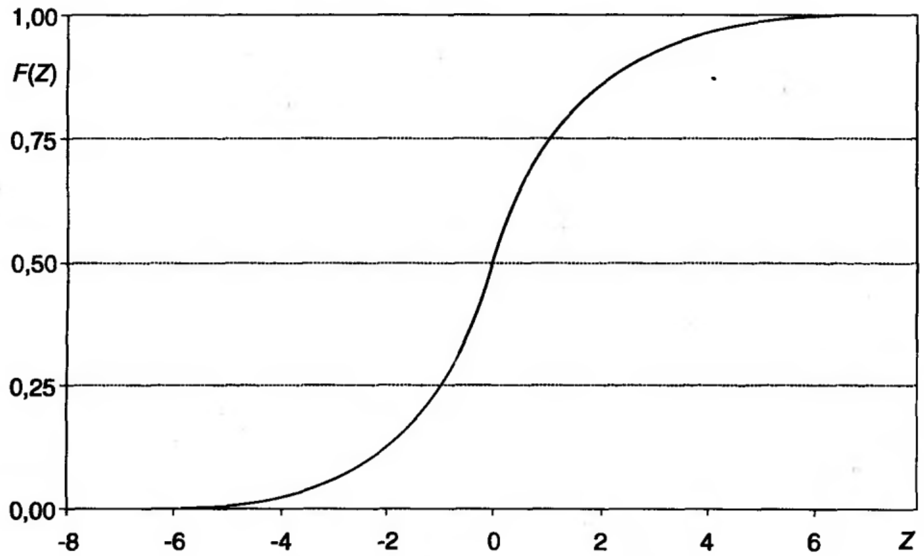


График логистической функции

Производная функции F является функцией плотности:

$$f\left(Z\right) = F'\left(Z\right) = \frac{e^{-Z}}{\left(1 + e^{-Z}\right)^2}$$

и характеризует предельный эффект объясняющего фактора.

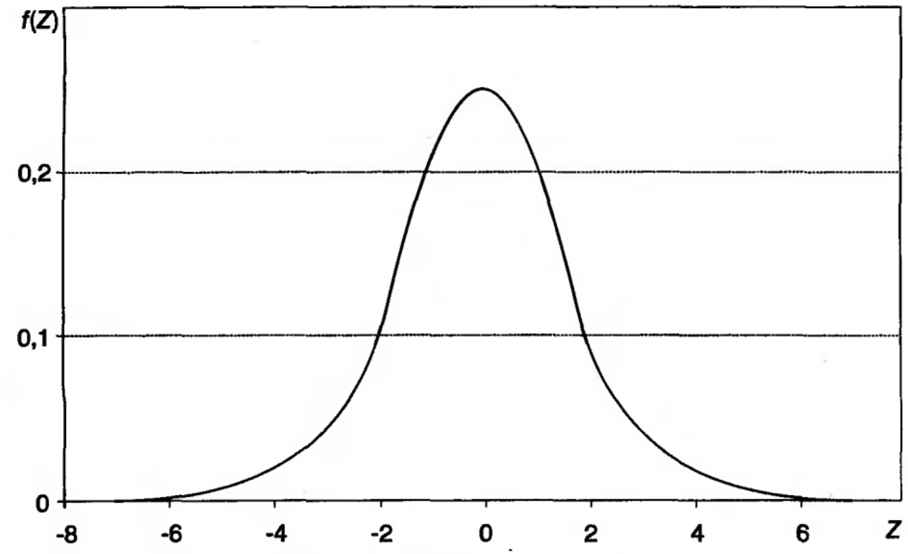


График функции плотности логистического распределения

Для больших по модулю значений Z воздействие на вероятность мало. Это воздействие возрастает для небольших значений Z.

Функция F нелинейно зависит от параметров $\beta_1, \ \beta_2$
Для оценки этих параметров, как правило, используется метод максимального правдоподобия (ММП).

Об этом мы поговорим позже. Пока остановимся на интерпретации параметров модели.

Здесь, в отличие от классической регрессионной модели, смысл оценок коэффициентов иной.

Для моделей бинарного выбора говорят о **<u>предельных эффектах</u>** объясняющих факторов.

Рассмотрим в общем случае модель, в которой Y определяется множеством факторов X.

$$p = F(Z) = \frac{1}{1+e^{-Z}},$$

$$Z = \beta_1 + \beta_2 X_2 + \ldots + \beta_k X_k$$

Предельным эффектом фактора X_i называется частная производная функции F по этому фактору:

$$\frac{\partial p}{\partial X_i} = \frac{\partial p}{\partial Z} \cdot \frac{\partial Z}{\partial X_i} = f(Z) \cdot \beta_i = \frac{e^{-Z}}{\left(1 + e^{-Z}\right)^2} \cdot \beta_i$$

Предельный эффект <u>не является константой</u>, а зависит от других факторов.

Предельный эффект при среднем значении фактора очень мал. Это связано с тем, что вероятность события Y=1 при средних результатах и так достаточно велика.

Probit-модель

Для этой модели в качестве функции F выбирается функция стандартного нормального распределения:

$$p = F(Z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{Z} e^{-\frac{u^2}{2}} du,$$

$$Z = \beta_1 + \beta_2 X_2 + \ldots + \beta_k X_k$$

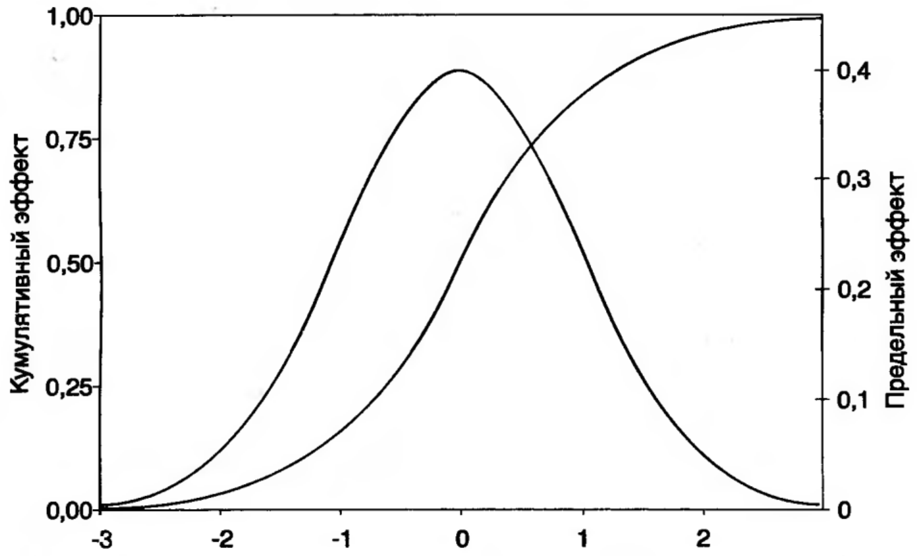
Функция плотности стандартного нормального распределения:

$$f(Z) = F'(Z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{Z^2}{2}}$$

Предельный эффект *i*-го фактора для пробит-модели:

$$\frac{\partial p}{\partial X_i} = \frac{\partial p}{\partial Z} \cdot \frac{\partial Z}{\partial X_i} = f\left(Z\right) \cdot \beta_i = \frac{1}{\sqrt{2\pi}} e^{-\frac{Z^2}{2}} \cdot \beta_i$$

Кумулятивный и предельный эффекты для пробит-модели представлены на рисунке.



Проблемы, связанные с оцениванием ММП, а также статистические тесты и вопросы, связанные с качеством оценивания мы рассмотрим в следующей лекции.

Пока рассмотрим **пример**.

Имеется выборка из 540 респондентов.
Изучаем зависимость окончания школы для индивидов в зависимости от ряда факторов.

Определим бинарную зависимую переменную
GRAD=1, если индивид закончил школу;
GRAD=0, в противном случае.

Для этих респондентов фиксировались также:

- совокупные результаты тестирования познавательных способностей: ASVABC (по 100-балльной шкале);
- число лет обучения матери респондента SF;
- число лет обучения отца SM;
- пол респондента MALE (фиктивная переменная, 1 – для мужчин, 0 – для женщин).

В таблице приведены результаты оценивания логит-модели. Использовался ММП, несколько итераций.

Также в нижней части таблицы рассчитаны предельные эффекты.

. logit GRAD ASVABC SM SF MALE
Logit estimates Number of obs = 540
LR chi2(4) = 43.75
Prob > chi2 = 0.0000
Log likelihood = -96.804844 Pseudo R2 = 0.1843

GRAD	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
ASVABC	.1329127	.0245718	5.41	0.000	.0847528	.1810726
SM	-.023178	.0868122	-0.27	0.789	-.1933267	.1469708
SF	.0122663	.0718876	0.17	0.865	-.1286307	.1531634
MALE	.1279654	.3989345	0.32	0.748	-.6539318	.9098627
_cons	-3.252373	1.065524	-3.05	0.002	-5.340761	-1.163985

Таблица 10.4. Логит–оценивание: зависимая переменная: *GRAD*

Переменная	Среднее	b	Среднее × b	$f(Z)$	$bf(Z)$
ASVABC	51,36	0,1329	6,8257	0,0281	0,0037
SM	11,58	–0,0231	–0,2687	0,0281	–0,0007
SF	11,84	0,0123	0,1456	0,0281	0,0003
MALE	0,50	0,1280	0,0640	0,0281	0,0036
Постоянный член	1,000	–3,2524	–3,2524		
Итого			3,5143		

Некоторые пояснения к построенной модели:

$$Z = \beta_1 + \beta_2 ASVABC + \beta_3 SM + \beta_4 SF + \beta_5 MALE.$$

Средние значения по каждой переменной приведены в нижней части таблицы.
С их помощью рассчитывается среднее значение Z по приведенной выше формуле и имеющимся оценкам параметров, оцененных ММП.
Затем считается

$$f\left(\overline{Z}\right) = \frac{e^{-\overline{Z}}}{\left(1+e^{-\overline{Z}}\right)^2} = \frac{e^{-3,5143}}{\left(1+e^{-3,5143}\right)} = 0,0281$$

Теперь считаем предельные эффекты.
Для примера рассчитаем предельный эффект для ASVABC:

$$\frac{\partial p}{\partial ASVABC} = f\left(\overline{Z}\right) \cdot \beta_{ASVABC} = 0{,}0281 \cdot 0{,}1329 = 0{,}0037$$

Полученное значение говорит о том, что при увеличении результата теста индивида на 1 балл, вероятность окончить школу возрастет на 0,37% пункта.
Из таблицы также видно, что от уровня образования родителей эффект незначителен.

И еще пара формул в качестве комментария к таблице результатов. Эти статистики мы детально рассмотрим в следующей лекции:

Псевдо-R^2:

$$\text{псевдо-}R^2 = 1 - \frac{\log L}{\log L_0}.$$

Отношение правдоподобия (LR-статистика):

$$2\log\frac{L}{L_0} = 2(\log L - \log L_0),$$

имеющее хи-квадрат распределение с числом степеней свободы (k-1).