Микроэконометрика

Системы уравнений

Потанин Богдан Станиславович

доцент, кандидат экономических наук

2024-2025

Системы бинарных уравнений

Мотивация

- Системы бинарных уравнений предполагают одновременное оценивание параметров нескольких бинарных уравнений.

- Системы бинарных уравнений предполагают одновременное оценивание параметров нескольких бинарных уравнений.
- Например, можно одновременно оценивать модели, прогнозирующие, будет ли индивид курить и употреблять алкоголь. В таком случае эффективность оценок параметров каждого из уравнений может быть повышена за счет учета информации о связи между этими уравнениями: решениями о курении и употреблении алкоголя.

Системы бинарных уравнений

Мотивация

- Системы бинарных уравнений предполагают одновременное оценивание параметров нескольких бинарных уравнений.
- Например, можно одновременно оценивать модели, прогнозирующие, будет ли индивид курить и употреблять алкоголь. В таком случае эффективность оценок параметров каждого из уравнений может быть повышена за счет учета информации о связи между этими уравнениями: решениями о курении и употреблении алкоголя.
- Как правило предполагается, что уравнения связаны через случайные ошибки. Например, положительная корреляция случайных ошибок уравнений курения и употребления алкоголя может быть связана с наличием общих, не наблюдаемых характеристик, влияющих на принятие соответствующих решений (интеллект, психологические особенности и т.д.).

- Системы бинарных уравнений предполагают одновременное оценивание параметров нескольких бинарных уравнений.
- Например, можно одновременно оценивать модели, прогнозирующие, будет ли индивид курить и употреблять алкоголь. В таком случае эффективность оценок параметров каждого из уравнений может быть повышена за счет учета информации о связи между этими уравнениями: решениями о курении и употреблении алкоголя.
- Как правило предполагается, что уравнения связаны через случайные ошибки. Например, положительная корреляция случайных ошибок уравнений курения и употребления алкоголя может быть связана с наличием общих, не наблюдаемых характеристик, влияющих на принятие соответствующих решений (интеллект, психологические особенности и т.д.).
- С помощью иерархических систем можно в явном виде учитывать эндогенность одной бинарной переменной, влияющие на другую бинарную переменную. Например, можно оценивать влияние высшего образования на вероятность занятости с учетом эндогенности образования за счет оценивания дополнительного уравнения на наличие высшего образования.

- Рассмотрим систему из J бинарных уравнений:

$$y_{ji}^{*} = x_{ji}\beta_j + \varepsilon_{ji}$$

$$i \in \{1, ..., n\}, j \in \{1, ..., J\}$$

- Рассмотрим систему из J бинарных уравнений:

$$y_{ji}^{*} = x_{ji}\beta_j + \varepsilon_{ji}$$

$$y_{ji} = \begin{cases} 1, \text{ если } y_{ji}^{*} > 0 \\ 0, \text{ в противном случае} \end{cases}$$

$$i \in \{1, ..., n\}, j \in \{1, ..., J\}$$

Системы бинарных уравнений
Общая постановка

- Рассмотрим систему из J бинарных уравнений:

$$y_{ji}^{*} = x_{ji}\beta_j + \varepsilon_{ji}$$

$$y_{ji} = \begin{cases} 1, \text{ если } y_{ji}^{*} > 0 \\ 0, \text{ в противном случае} \end{cases}$$

$$i \in \{1, ..., n\}, j \in \{1, ..., J\}$$

- Уравнения системы различаются регрессорами x_{ji}, коэффициентами β_j и случайными ошибками ε_{ji}, не зависящими от x_{ji}.

Системы бинарных уравнений

Общая постановка

- Рассмотрим систему из J бинарных уравнений:

$$y_{ji}^* = x_{ji}\beta_j + \varepsilon_{ji}$$

$$y_{ji} = \begin{cases} 1, \text{ если } y_{ji}^* > 0 \\ 0, \text{ в противном случае} \end{cases}$$

$$i \in \{1, ..., n\}, j \in \{1, ..., J\}$$

- Уравнения системы различаются регрессорами x_{ji}, коэффициентами β_j и случайными ошибками ε_{ji}, не зависящими от x_{ji}.
- Случайные ошибки ε_i независимы между наблюдениями (по i), но могут быть зависимы между уравнениями (по j).

Системы бинарных уравнений

Общая постановка

- Рассмотрим систему из J бинарных уравнений:

$$y_{ji}^{*} = x_{ji}\beta_j + \varepsilon_{ji}$$

$$y_{ji} = \begin{cases} 1, \text{ если } y_{ji}^{*} > 0 \\ 0, \text{ в противном случае} \end{cases}$$

$$i \in \{1, ..., n\}, j \in \{1, ..., J\}$$

- Уравнения системы различаются регрессорами x_{ji}, коэффициентами β_j и случайными ошибками ε_{ji}, не зависящими от x_{ji}.
- Случайные ошибки ε_i независимы между наблюдениями (по i), но могут быть зависимы между уравнениями (по j).
- Из соображений идентифицируемости предполагается, что $E(\varepsilon_{ji}) = 0$ и $Var(\varepsilon_{ji}) = 1$.

- Рассмотрим систему из J бинарных уравнений:

$$y^*_{ji} = x_{ji}\beta_j + \varepsilon_{ji}$$

$$y_{ji} = \begin{cases} 1, \text{ если } y^*_{ji} > 0 \\ 0, \text{ в противном случае} \end{cases}$$

$$i \in \{1, ..., n\}, j \in \{1, ..., J\}$$

- Уравнения системы различаются регрессорами x_{ji}, коэффициентами β_j и случайными ошибками ε_{ji}, не зависящими от x_{ji}.
- Случайные ошибки ε_i независимы между наблюдениями (по i), но могут быть зависимы между уравнениями (по j).
- Из соображений идентифицируемости предполагается, что $E(\varepsilon_{ji}) = 0$ и $Var(\varepsilon_{ji}) = 1$.
- Для простоты без потери общности далее будем рассматривать двумерную систему бинарных уравнений $J = 2$.

Многомерная пробит модель

Общая постановка

- Предположим, что совместное распределение случайных ошибок является многомерным нормальным:

$$\begin{bmatrix}\varepsilon_{1i} \\ \varepsilon_{2i}\end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix}0 \\ 0\end{bmatrix}, \begin{bmatrix}1 & \rho \\ \rho & 1\end{bmatrix}\right),$$

- Предположим, что совместное распределение случайных ошибок является многомерным нормальным:
$$\begin{bmatrix}\varepsilon_{1i}\\ \varepsilon_{2i}\end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix} \begin{bmatrix}1 & \rho\\ \rho & 1\end{bmatrix}\right),$$
 где ρ является коэффициентом корреляции между случайными ошибками уравнений.

Многомерная пробит модель

Общая постановка

- Предположим, что совместное распределение случайных ошибок является многомерным нормальным:
$$\begin{bmatrix}\varepsilon_{1i}\\ \varepsilon_{2i}\end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix}0\\ 0\end{bmatrix} \begin{bmatrix}1 & \rho\\ \rho & 1\end{bmatrix}\right),$$
 где ρ является коэффициентом корреляции между случайными ошибками уравнений.
- Условная вероятности считаются следующим образом:

- Предположим, что совместное распределение случайных ошибок является многомерным нормальным:

$$\begin{bmatrix}\varepsilon_{1i}\\ \varepsilon_{2i}\end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}\begin{bmatrix}1 & \rho\\ \rho & 1\end{bmatrix}\right),$$

 где ρ является коэффициентом корреляции между случайными ошибками уравнений.
- Условная вероятности считаются следующим образом:

$$P(y_{1i}=1, y_{2i}=1|x_{1i}, x_{2i}) = P(-\varepsilon_{1i} < x_{1i}\beta_1, -\varepsilon_{2i} < x_{2i}\beta_2|x_{1i}, x_{2i}) = \Phi\left(x_{1i}\beta_1, x_{2i}\beta_2; \rho\right)$$

Многомерная пробит модель

Общая постановка

- Предположим, что совместное распределение случайных ошибок является многомерным нормальным:

$$\begin{bmatrix}\varepsilon_{1i}\\ \varepsilon_{2i}\end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}\begin{bmatrix}1 & \rho\\ \rho & 1\end{bmatrix}\right),$$

 где ρ является коэффициентом корреляции между случайными ошибками уравнений.
- Условная вероятности считаются следующим образом:

$$P(y_{1i}=1, y_{2i}=1|x_{1i}, x_{2i}) = P(-\varepsilon_{1i} < x_{1i}\beta_1, -\varepsilon_{2i} < x_{2i}\beta_2|x_{1i}, x_{2i}) = \Phi\left(x_{1i}\beta_1, x_{2i}\beta_2; \rho\right)$$

$$P(y_{1i}=1, y_{2i}=0|x_{1i}, x_{2i}) = P(-\varepsilon_{1i} < x_{1i}\beta_1, \varepsilon_{2i} < -x_{2i}\beta_2|x_{1i}, x_{2i}) = \Phi\left(x_{1i}\beta_1, -x_{2i}\beta_2; -\rho\right)$$

Многомерная пробит модель

Общая постановка

- Предположим, что совместное распределение случайных ошибок является многомерным нормальным:

$$\begin{bmatrix}\varepsilon_{1i}\\ \varepsilon_{2i}\end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}\begin{bmatrix}1 & \rho\\ \rho & 1\end{bmatrix}\right),$$

  где ρ является коэффициентом корреляции между случайными ошибками уравнений.
- Условная вероятности считаются следующим образом:

$$P(y_{1i}=1,y_{2i}=1|x_{1i},x_{2i}) = P(-\varepsilon_{1i}<x_{1i}\beta_1,-\varepsilon_{2i}<x_{2i}\beta_2|x_{1i},x_{2i}) = \Phi\left(x_{1i}\beta_1,x_{2i}\beta_2;\rho\right)$$

$$P(y_{1i}=1,y_{2i}=0|x_{1i},x_{2i}) = P(-\varepsilon_{1i}<x_{1i}\beta_1,\varepsilon_{2i}<-x_{2i}\beta_2|x_{1i},x_{2i}) = \Phi\left(x_{1i}\beta_1,-x_{2i}\beta_2;-\rho\right)$$

$$P(y_{1i}=0,y_{2i}=1|x_{1i},x_{2i}) = P(\varepsilon_{1i}<-x_{1i}\beta_1,-\varepsilon_{2i}<x_{2i}\beta_2|x_{1i},x_{2i}) = \Phi\left(-x_{1i}\beta_1,x_{2i}\beta_2;-\rho\right)$$

Многомерная пробит модель

Общая постановка

- Предположим, что совместное распределение случайных ошибок является многомерным нормальным:

$$\begin{bmatrix}\varepsilon_{1i}\\ \varepsilon_{2i}\end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}\begin{bmatrix}1 & \rho\\ \rho & 1\end{bmatrix}\right),$$

где ρ является коэффициентом корреляции между случайными ошибками уравнений.

- Условная вероятности считаются следующим образом:

$$P(y_{1i}=1, y_{2i}=1|x_{1i}, x_{2i}) = P(-\varepsilon_{1i} < x_{1i}\beta_1, -\varepsilon_{2i} < x_{2i}\beta_2|x_{1i}, x_{2i}) = \Phi\left(x_{1i}\beta_1, x_{2i}\beta_2; \rho\right)$$

$$P(y_{1i}=1, y_{2i}=0|x_{1i}, x_{2i}) = P(-\varepsilon_{1i} < x_{1i}\beta_1, \varepsilon_{2i} < -x_{2i}\beta_2|x_{1i}, x_{2i}) = \Phi\left(x_{1i}\beta_1, -x_{2i}\beta_2; -\rho\right)$$

$$P(y_{1i}=0, y_{2i}=1|x_{1i}, x_{2i}) = P(\varepsilon_{1i} < -x_{1i}\beta_1, -\varepsilon_{2i} < x_{2i}\beta_2|x_{1i}, x_{2i}) = \Phi\left(-x_{1i}\beta_1, x_{2i}\beta_2; -\rho\right)$$

$$P(y_{1i}=0, y_{2i}=0|x_{1i}, x_{2i}) = P(\varepsilon_{1i} < -x_{1i}\beta_1, \varepsilon_{2i} < -x_{2i}\beta_2|x_{1i}, x_{2i}) = \Phi\left(-x_{1i}\beta_1, -x_{2i}\beta_2; \rho\right)$$

- Предположим, что совместное распределение случайных ошибок является многомерным нормальным:

$$\begin{bmatrix}\varepsilon_{1i}\\ \varepsilon_{2i}\end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}\begin{bmatrix}1 & \rho\\ \rho & 1\end{bmatrix}\right),$$

где ρ является коэффициентом корреляции между случайными ошибками уравнений.

- Условная вероятности считаются следующим образом:

$$P(y_{1i}=1, y_{2i}=1|x_{1i},x_{2i}) = P(-\varepsilon_{1i} < x_{1i}\beta_1, -\varepsilon_{2i} < x_{2i}\beta_2|x_{1i},x_{2i}) = \Phi\left(x_{1i}\beta_1, x_{2i}\beta_2;\rho\right)$$

$$P(y_{1i}=1, y_{2i}=0|x_{1i},x_{2i}) = P(-\varepsilon_{1i} < x_{1i}\beta_1, \varepsilon_{2i} < -x_{2i}\beta_2|x_{1i},x_{2i}) = \Phi\left(x_{1i}\beta_1, -x_{2i}\beta_2;-\rho\right)$$

$$P(y_{1i}=0, y_{2i}=1|x_{1i},x_{2i}) = P(\varepsilon_{1i} < -x_{1i}\beta_1, -\varepsilon_{2i} < x_{2i}\beta_2|x_{1i},x_{2i}) = \Phi\left(-x_{1i}\beta_1, x_{2i}\beta_2;-\rho\right)$$

$$P(y_{1i}=0, y_{2i}=0|x_{1i},x_{2i}) = P(\varepsilon_{1i} < -x_{1i}\beta_1, \varepsilon_{2i} < -x_{2i}\beta_2|x_{1i},x_{2i}) = \Phi\left(-x_{1i}\beta_1, -x_{2i}\beta_2;\rho\right)$$

Где $\Phi(a,b;\rho)$ является функцией распределения стандартного двумерного нормального распределения с коэффициентом корреляции ρ.

- Для краткости обозначим $P_{j,k,i} = P(y_{1i} = j, y_{2i} = k)$, где $j, k \in \{0, 1\}$.

Многомерная пробит модель

Оценивание

- Для краткости обозначим $P_{j,k,i} = P(y_{1i} = j, y_{2i} = k)$, где $j, k \in \{0, 1\}$.
- Как правило, для оценивания используется метод максимального правдоподобия, предполагающий максимизацию следующей функции правдоподобия по β_1, β_2 и ρ:

$$L\left(\beta_1, \beta_2, \rho; x_{1i}, x_{2i}\right) = \prod_{i}^{n} P_{\mathbf{y_1}_i, \mathbf{y_2}_i, i} =$$

Многомерная пробит модель
Оценивание

- Для краткости обозначим $P_{j,k,i} = P(y_{1i} = j, y_{2i} = k)$, где $j, k \in \{0, 1\}$.
- Как правило, для оценивания используется метод максимального правдоподобия, предполагающий максимизацию следующей функции правдоподобия по β_1, β_2 и ρ:

$$L\left(\beta_1, \beta_2, \rho; x_{1i}, x_{2i}\right) = \prod_{i}^{n} P_{\mathbf{y_{1}}_i, \mathbf{y_{2}}_i, i} =$$

$$= \prod_{i=1}^{n} \Phi\left((2y_{1i} - 1)x_{1i}\beta_1, (2y_{2i} - 1)x_{2i}\beta_2; (2y_{1i} - 1)(2y_{2i} - 1)\rho\right)$$

- Для краткости обозначим $P_{j,k,i} = P(y_{1i} = j, y_{2i} = k)$, где $j, k \in \{0, 1\}$.
- Как правило, для оценивания используется метод максимального правдоподобия, предполагающий максимизацию следующей функции правдоподобия по β_1, β_2 и ρ:

$$L\left(\beta_1, \beta_2, \rho; x_{1i}, x_{2i}\right) = \prod_{i}^{n} P_{y_{\mathbf{1}i}, y_{\mathbf{2}i}, i} =$$

$$= \prod_{i=1}^{n} \Phi\left((2y_{1i} - 1)x_{1i}\beta_1, (2y_{2i} - 1)x_{2i}\beta_2; (2y_{1i} - 1)(2y_{2i} - 1)\rho\right)$$

- При $\rho \neq 0$ оценки данной модели более эффективны, чем оценки, полученные за счет оценивания каждого из бинарных уравнений по отдельности, поскольку мы учитываем дополнительную информацию о связи между уравнениями.

- Для краткости обозначим $P_{j,k,i} = P(y_{1i} = j, y_{2i} = k)$, где $j, k \in \{0, 1\}$.
- Как правило, для оценивания используется метод максимального правдоподобия, предполагающий максимизацию следующей функции правдоподобия по β_1, β_2 и ρ:

$$L\left(\beta_1, \beta_2, \rho; x_{1i}, x_{2i}\right) = \prod_{i}^{n} P_{y_{\mathbf{1}i}, y_{\mathbf{2}i}, i} =$$

$$= \prod_{i=1}^{n} \Phi\left((2y_{1i} - 1)x_{1i}\beta_1, (2y_{2i} - 1)x_{2i}\beta_2; (2y_{1i} - 1)(2y_{2i} - 1)\rho\right)$$

- При $\rho \neq 0$ оценки данной модели более эффективны, чем оценки, полученные за счет оценивания каждого из бинарных уравнений по отдельности, поскольку мы учитываем дополнительную информацию о связи между уравнениями.
- Для того, чтобы понять, есть ли смысл в оценивании системы, достаточно протестировать гипотезу $H_0 : \rho = 0$, например, LR-тестом. Если нулевая гипотеза не отвергается, то без потери в эффективности уравнения можно оценивать по отдельности.

- Для расчета предельных эффектов будем пользоваться следующими формулами:

$$\frac{\partial\Phi\left(a,b;\rho\right)}{\partial a}=\phi\left(a\right)\Phi\left(\frac{b-\rho a}{\sqrt{1-\rho^{2}}}\right),\quad\frac{\partial\Phi\left(a,b;\rho\right)}{\partial b}=\phi\left(b\right)\Phi\left(\frac{a-\rho b}{\sqrt{1-\rho^{2}}}\right)$$

- Для расчета предельных эффектов будем пользоваться следующими формулами:

$$\frac{\partial \Phi\left(a, b; \rho\right)}{\partial a} = \phi\left(a\right)\Phi\left(\frac{b-\rho a}{\sqrt{1-\rho^2}}\right), \quad \frac{\partial \Phi\left(a, b; \rho\right)}{\partial b} = \phi\left(b\right)\Phi\left(\frac{a-\rho b}{\sqrt{1-\rho^2}}\right)$$

- Предельные эффекты на маржинальные вероятности $P(y_{1i} = 1|x_{1i})$ и $P(y_{2i} = 1|x_{2i})$ рассчитываются точно так же, как в обычной пробит модели.

Многомерная пробит модель

Предельные эффекты

- Для расчета предельных эффектов будем пользоваться следующими формулами:

$$\frac{\partial \Phi\left(a, b; \rho\right)}{\partial a} = \phi\left(a\right)\Phi\left(\frac{b-\rho a}{\sqrt{1-\rho^2}}\right), \quad \frac{\partial \Phi\left(a, b; \rho\right)}{\partial b} = \phi\left(b\right)\Phi\left(\frac{a-\rho b}{\sqrt{1-\rho^2}}\right)$$

- Предельные эффекты на маржинальные вероятности $P(y_{1i}=1|x_{1i})$ и $P(y_{2i}=1|x_{2i})$ рассчитываются точно так же, как в обычной пробит модели.
- Рассмотрим предельный эффект регрессора v_i, входящего и в x_{1i} и в x_{2i} с коэффициентами β_{1v} и β_{2v} соответственно, на совместную вероятность:

$$\frac{\partial P(y_{1i}=1, y_{2i}=1|x_{1i}, x_{2i})}{\partial v} = \frac{\partial \Phi\left(x_{1i}\beta_1, x_{2i}\beta_2; \rho\right)}{\partial v} =$$

- Для расчета предельных эффектов будем пользоваться следующими формулами:

$$\frac{\partial\Phi\left(a,b;\rho\right)}{\partial a}=\phi\left(a\right)\Phi\left(\frac{b-\rho a}{\sqrt{1-\rho^2}}\right),\quad \frac{\partial\Phi\left(a,b;\rho\right)}{\partial b}=\phi\left(b\right)\Phi\left(\frac{a-\rho b}{\sqrt{1-\rho^2}}\right)$$

- Предельные эффекты на маржинальные вероятности $P(y_{1i}=1|x_{1i})$ и $P(y_{2i}=1|x_{2i})$ рассчитываются точно так же, как в обычной пробит модели.
- Рассмотрим предельный эффект регрессора v_i, входящего и в x_{1i} и в x_{2i} с коэффициентами β_{1v} и β_{2v} соответственно, на совместную вероятность:

$$\frac{\partial P(y_{1i}=1,y_{2i}=1|x_{1i},x_{2i})}{\partial v}=\frac{\partial\Phi\left(x_{1i}\beta_1,x_{2i}\beta_2;\rho\right)}{\partial v}=$$

$$=\beta_{1v}\phi\left(x_{1i}\beta_1\right)\Phi\left(\frac{x_{2i}\beta_2-\rho x_{1i}\beta_1}{\sqrt{1-\rho^2}}\right)+\beta_{2v}\phi\left(x_{2i}\beta_2\right)\Phi\left(\frac{x_{1i}\beta_1-\rho x_{2i}\beta_2}{\sqrt{1-\rho^2}}\right)$$

- Для расчета предельных эффектов будем пользоваться следующими формулами:

$$\frac{\partial\Phi\left(a,b;\rho\right)}{\partial a}=\phi\left(a\right)\Phi\left(\frac{b-\rho a}{\sqrt{1-\rho^2}}\right),\quad\frac{\partial\Phi\left(a,b;\rho\right)}{\partial b}=\phi\left(b\right)\Phi\left(\frac{a-\rho b}{\sqrt{1-\rho^2}}\right)$$

- Предельные эффекты на маржинальные вероятности $P(y_{1i}=1|x_{1i})$ и $P(y_{2i}=1|x_{2i})$ рассчитываются точно так же, как в обычной пробит модели.
- Рассмотрим предельный эффект регрессора v_i, входящего и в x_{1i} и в x_{2i} с коэффициентами β_{1v} и β_{2v} соответственно, на совместную вероятность:

$$\frac{\partial P(y_{1i}=1,y_{2i}=1|x_{1i},x_{2i})}{\partial v}=\frac{\partial\Phi\left(x_{1i}\beta_1,x_{2i}\beta_2;\rho\right)}{\partial v}=$$

$$=\beta_{1v}\phi\left(x_{1i}\beta_1\right)\Phi\left(\frac{x_{2i}\beta_2-\rho x_{1i}\beta_1}{\sqrt{1-\rho^2}}\right)+\beta_{2v}\phi\left(x_{2i}\beta_2\right)\Phi\left(\frac{x_{1i}\beta_1-\rho x_{2i}\beta_2}{\sqrt{1-\rho^2}}\right)$$

Если знаки β_{1v} и β_{2v} совпадают, то таким же будем и знак предельного эффекта.

- Предположим, что одна зависимая бинарная переменная в системе влияет на другую:

$$y_{1i}^{*} = x_{1i}\beta_1 + \alpha y_{2i} + \varepsilon_{1i}$$
$$y_{2i}^{*} = x_{2i}\beta_2 + \varepsilon_{2i},$$

- Предположим, что одна зависимая бинарная переменная в системе влияет на другую:

$$y_{1i}^{*} = x_{1i}\beta_1 + \alpha y_{2i} + \varepsilon_{1i}$$
$$y_{2i}^{*} = x_{2i}\beta_2 + \varepsilon_{2i},$$

где без потери общности так же можно было бы предположить, что y_{2i} входит в x_{1i}.

- Предположим, что одна зависимая бинарная переменная в системе влияет на другую:

$$y_{1i}^{*} = x_{1i}\beta_1 + \alpha y_{2i} + \varepsilon_{1i}$$
$$y_{2i}^{*} = x_{2i}\beta_2 + \varepsilon_{2i},$$

  где без потери общности так же можно было бы предположить, что y_{2i} входит в x_{1i}.

- Обычно y_{2i} именуют эндогенным бинарным регрессором. Ирерхические модели позволяют в явном виде учесть эндогенность y_{2i} по отношению к уравнению y_{1i}.

Многомерная пробит модель

Иерархическая модель

- Предположим, что одна зависимая бинарная переменная в системе влияет на другую:

$$y_{1i}^{*} = x_{1i}\beta_1 + \alpha y_{2i} + \varepsilon_{1i}$$
$$y_{2i}^{*} = x_{2i}\beta_2 + \varepsilon_{2i},$$

  где без потери общности так же можно было бы предположить, что y_{2i} входит в x_{1i}.

- Обычно y_{2i} именуют эндогенным бинарным регрессором. Ирерхические модели позволяют в явном виде учесть эндогенность y_{2i} по отношению к уравнению y_{1i}.
- В подобных системах представляет интерес оценивание условных вероятностей:

$$P(y_{1i} = k|x_{1i}, x_{2i}, y_{2i} = j), \text{ где } k, j \in \{0, 1\}$$

- Предположим, что одна зависимая бинарная переменная в системе влияет на другую:

$$\begin{gathered}y_{1i}^{*}=x_{1i}\beta_{1}+\alpha y_{2i}+\varepsilon_{1i}\\ y_{2i}^{*}=x_{2i}\beta_{2}+\varepsilon_{2i},\end{gathered}$$

  где без потери общности так же можно было бы предположить, что y_{2i} входит в x_{1i}.
- Обычно y_{2i} именуют эндогенным бинарным регрессором. Ирерхические модели позволяют в явном виде учесть эндогенность y_{2i} по отношению к уравнению y_{1i}.
- В подобных системах представляет интерес оценивание условных вероятностей:

$$P(y_{1i}=k|x_{1i},x_{2i},y_{2i}=j),\text{ где }k,j\in\{0,1\}$$

- Данные вероятности нетрудно рассчитать по формуле Байеса, например:

$$P(y_{1i}=1|x_{1i},x_{2i},y_{2i}=1)=\frac{P(y_{1i}=1,y_{2i}=1|x_{1i},x_{2i})}{P(y_{2i}=1|x_{1i},x_{2i})}=\frac{\Phi\left(x_{1i}\beta_{1}+\alpha,x_{2i}\beta_{2};\rho\right)}{\Phi\left(x_{2i}\beta_{2}\right)}$$

Многомерная пробит модель

Предельный эффекты в иерархической модели

- Предельный эффект условной вероятности по y_{2i} рассчитывается как разница условных вероятностей:

$$P(y_{1i}=1|x_{1i},x_{2i},y_{2i}=1)-P(y_{1i}=1|x_{1i},x_{2i},y_{2i}=0)=$$

Многомерная пробит модель

Предельный эффекты в иерархической модели

- Предельный эффект условной вероятности по y_{2i} рассчитывается как разница условных вероятностей:

$$P(y_{1i} = 1|x_{1i}, x_{2i}, y_{2i} = 1) - P(y_{1i} = 1|x_{1i}, x_{2i}, y_{2i} = 0) =$$
$$= \frac{\Phi\left(x_{1i}\beta_1 + \alpha, x_{2i}\beta_2; \rho\right)}{\Phi\left(x_{2i}\beta_2\right)} - \frac{\Phi\left(x_{1i}\beta_1, -x_{2i}\beta_2; -\rho\right)}{1 - \Phi\left(x_{2i}\beta_2\right)}$$

- Предельный эффект условной вероятности по y_{2i} рассчитывается как разница условных вероятностей:

$$P(y_{1i}=1|x_{1i},x_{2i},y_{2i}=1)-P(y_{1i}=1|x_{1i},x_{2i},y_{2i}=0)=$$
$$=\frac{\Phi\left(x_{1i}\beta_1+\alpha,x_{2i}\beta_2;\rho\right)}{\Phi\left(x_{2i}\beta_2\right)}-\frac{\Phi\left(x_{1i}\beta_1,-x_{2i}\beta_2;-\rho\right)}{1-\Phi\left(x_{2i}\beta_2\right)}$$

- Недостаток предельного эффекта с точки зрения содержательной интерпретации заключается в том, что он отражает влияние y_{2i} на y_{1i} не только непосредственно, то есть через α, но и косвенно – через условное распределение случайной ошибки $\varepsilon_{1i}|\varepsilon_{2i}>-x_{2i}\beta_2$.

- Предельный эффект условной вероятности по y_{2i} рассчитывается как разница условных вероятностей:

$$P(y_{1i}=1|x_{1i},x_{2i},y_{2i}=1)-P(y_{1i}=1|x_{1i},x_{2i},y_{2i}=0)=$$
$$=\frac{\Phi\left(x_{1i}\beta_1+\alpha,x_{2i}\beta_2;\rho\right)}{\Phi\left(x_{2i}\beta_2\right)}-\frac{\Phi\left(x_{1i}\beta_1,-x_{2i}\beta_2;-\rho\right)}{1-\Phi\left(x_{2i}\beta_2\right)}$$

- Недостаток предельного эффекта с точки зрения содержательной интерпретации заключается в том, что он отражает влияние y_{2i} на y_{1i} не только непосредственно, то есть через α, но и косвенно – через условное распределение случайной ошибки $\varepsilon_{1i}|\varepsilon_{2i}>-x_{2i}\beta_2$.
- Например, предельный эффект высшего образования y_{2i} на условную вероятность занятости y_{1i} может также отражать различия в вероятностях, обусловленные не только непосредственным эффектом образования (рост человеческого капитала и т.д.), но и тем, что высшее образование получают более мотивированные и способные индивиды.

- Для того, чтобы отделить непосредственный эффект от части, обусловленной неслучайным отбором, используется концепция потенциальных исходов:

$$y_{1i}^{(1)}=\begin{cases}1,\text{ если } x_{1i}\beta_1+\alpha+\varepsilon_{1i}\geq 0\\ 0,\text{ в противном случае}\end{cases}\qquad y_{1i}^{(0)}=\begin{cases}1,\text{ если } x_{1i}\beta_1+\varepsilon_{1i}\geq 0\\ 0,\text{ в противном случае}\end{cases}$$

Многомерная пробит модель
Потенциальные исходы

- Для того, чтобы отделить непосредственный эффект от части, обусловленной неслучайным отбором, используется концепция потенциальных исходов:

$$y_{1i}^{(1)} = \begin{cases} 1, \text{ если } x_{1i}\beta_1 + \alpha + \varepsilon_{1i} \geq 0 \\ 0, \text{ в противном случае} \end{cases} \qquad y_{1i}^{(0)} = \begin{cases} 1, \text{ если } x_{1i}\beta_1 + \varepsilon_{1i} \geq 0 \\ 0, \text{ в противном случае} \end{cases}$$

- **Потенциальный исходы** $y_{1i}^{(1)}$ и $y_{1i}^{(0)}$ могут, например, отражать факт занятости индивида в случаях, когда у него есть или нет высшего образования соответственно.

- Для того, чтобы отделить непосредственный эффект от части, обусловленной неслучайным отбором, используется концепция потенциальных исходов:

$$y_{1i}^{(1)} = \begin{cases} 1, \text{ если } x_{1i}\beta_1 + \alpha + \varepsilon_{1i} \geq 0 \\ 0, \text{ в противном случае} \end{cases} \qquad y_{1i}^{(0)} = \begin{cases} 1, \text{ если } x_{1i}\beta_1 + \varepsilon_{1i} \geq 0 \\ 0, \text{ в противном случае} \end{cases}$$

- **Потенциальный исходы** $y_{1i}^{(1)}$ и $y_{1i}^{(0)}$ могут, например, отражать факт занятости индивида в случаях, когда у него есть или нет высшего образования соответственно.
- **Наблюдаемый исход** y_{1i} зависит от значения **переменной воздействия** y_{2i}:

$$y_{1i} = \begin{cases} y_{1i}^{(1)}, \text{ если } y_{2i} = 1 \\ y_{1i}^{(0)}, \text{ если } y_{2i} = 0 \end{cases} = y_{1i}^{(1)} y_{2i} + y_{1i}^{(0)}(1 - y_{2i})$$

Многомерная пробит модель

Потенциальные исходы

- Для того, чтобы отделить непосредственный эффект от части, обусловленной неслучайным отбором, используется концепция потенциальных исходов:

$$y_{1i}^{(1)} = \begin{cases} 1, \text{ если } x_{1i}\beta_1 + \alpha + \varepsilon_{1i} \geq 0 \\ 0, \text{ в противном случае} \end{cases} \qquad y_{1i}^{(0)} = \begin{cases} 1, \text{ если } x_{1i}\beta_1 + \varepsilon_{1i} \geq 0 \\ 0, \text{ в противном случае} \end{cases}$$

- **Потенциальный исходы** $y_{1i}^{(1)}$ и $y_{1i}^{(0)}$ могут, например, отражать факт занятости индивида в случаях, когда у него есть или нет высшего образования соответственно.
- **Наблюдаемый исход** y_{1i} зависит от значения **переменной воздействия** y_{2i}:

$$y_{1i} = \begin{cases} y_{1i}^{(1)}, \text{ если } y_{2i} = 1 \\ y_{1i}^{(0)}, \text{ если } y_{2i} = 0 \end{cases} = y_{1i}^{(1)} y_{2i} + y_{1i}^{(0)}(1 - y_{2i})$$

- **Эффект воздействия** определяется как разница потенциальных исходов:

$$\text{TE}_i = y_{1i}^{(1)} - y_{1i}^{(0)}$$

Многомерная пробит модель

Потенциальные исходы

- Для того, чтобы отделить непосредственный эффект от части, обусловленной неслучайным отбором, используется концепция потенциальных исходов:

$$y_{1i}^{(1)} = \begin{cases} 1, \text{ если } x_{1i}\beta_1 + \alpha + \varepsilon_{1i} \geq 0 \\ 0, \text{ в противном случае} \end{cases} \qquad y_{1i}^{(0)} = \begin{cases} 1, \text{ если } x_{1i}\beta_1 + \varepsilon_{1i} \geq 0 \\ 0, \text{ в противном случае} \end{cases}$$

- **Потенциальный исходы** $y_{1i}^{(1)}$ и $y_{1i}^{(0)}$ могут, например, отражать факт занятости индивида в случаях, когда у него есть или нет высшего образования соответственно.
- **Наблюдаемый исход** y_{1i} зависит от значения **переменной воздействия** y_{2i}:

$$y_{1i} = \begin{cases} y_{1i}^{(1)}, \text{ если } y_{2i} = 1 \\ y_{1i}^{(0)}, \text{ если } y_{2i} = 0 \end{cases} = y_{1i}^{(1)} y_{2i} + y_{1i}^{(0)}(1 - y_{2i})$$

- **Эффект воздействия** определяется как разница потенциальных исходов:

$$\text{TE}_i = y_{1i}^{(1)} - y_{1i}^{(0)}$$

- **Проблема** – мы не можем оценить эффект воздействия, так как не способны одновременно наблюдать $y_{1i}^{(0)}$ и $y_{1i}^{(1)}$, поскольку, например, индивид либо получил, либо не получил высшее образование.

- Средний эффект воздействия (average treatment effect) определяется как:

$$\text{ATE} = \text{E}\left(y_{1i}^{(1)}\right) - \text{E}\left(y_{1i}^{(0)}\right) = \text{P}\left(y_{1i}^{(1)} = 1\right) - \text{P}\left(y_{1i}^{(0)} = 1\right) =$$
$$= \text{E}\left(\text{P}\left(y_{1i}^{(1)} = 1|x_{1i}\right) - \text{P}\left(y_{1i}^{(0)} = 1|x_{1i}\right)\right)$$

- Следовательно, в силу закона больших чисел этот эффект можно оценить как:

$$\widehat{\text{ATE}} = \frac{1}{n}\sum_{i=1}^{n}\hat{\text{P}}\left(y_{1i}^{(1)} = 1|x_{1i}\right) - \hat{\text{P}}\left(y_{1i}^{(0)} = 1|x_{1i}\right) = \frac{1}{n}\sum_{i=1}^{n}\Phi\left(x_{1i}\hat{\beta}_1 + \hat{\alpha}\right) - \Phi\left(x_{1i}\hat{\beta}\right)$$

- **Важно** – несмотря на то, что $\widehat{\text{ATE}}$ зависит лишь от $\hat{\beta}_1$ и $\hat{\alpha}$, при наличии корреляции между случайными ошибками $\rho \neq 0$ для получения состоятельных оценок этих параметров необходимо оценивать систему из двух уравнений: y_{1i} и y_{2i}.

- Иногда зависимая переменная может быть номинальной, то есть принимать одно из конечного числа взаимоисключающих значений.

- Иногда зависимая переменная может быть номинальной, то есть принимать одно из конечного числа взаимоисключающих значений.
- Часто рассматриваются номинальные переменные, отражающие выбор индивида, например, в отношении:

Модели мультиномиального выбора

Мотивация

- Иногда зависимая переменная может быть номинальной, то есть принимать одно из конечного числа взаимоисключающих значений.
- Часто рассматриваются номинальные переменные, отражающие выбор индивида, например, в отношении:
  - типа получаемого образования: техническое, гуманитарное или никакого

Модели мультиномиального выбора

Мотивация

- Иногда зависимая переменная может быть номинальной, то есть принимать одно из конечного числа взаимоисключающих значений.
- Часто рассматриваются номинальные переменные, отражающие выбор индивида, например, в отношении:
    - типа получаемого образования: техническое, гуманитарное или никакого
    - марки приобретаемого телефона

- Иногда зависимая переменная может быть номинальной, то есть принимать одно из конечного числа взаимоисключающих значений.
- Часто рассматриваются номинальные переменные, отражающие выбор индивида, например, в отношении:
  - типа получаемого образования: техническое, гуманитарное или никакого
  - марки приобретаемого телефона
  - типа используемого браузера

- Иногда зависимая переменная может быть номинальной, то есть принимать одно из конечного числа взаимоисключающих значений.
- Часто рассматриваются номинальные переменные, отражающие выбор индивида, например, в отношении:
  - типа получаемого образования: техническое, гуманитарное или никакого
  - марки приобретаемого телефона
  - типа используемого браузера
  - бронируемого отеля из числа доступных в городе альтернатив

Модели мультиномиального выбора

Мотивация

- Иногда зависимая переменная может быть номинальной, то есть принимать одно из конечного числа взаимоисключающих значений.
- Часто рассматриваются номинальные переменные, отражающие выбор индивида, например, в отношении:
  - типа получаемого образования: техническое, гуманитарное или никакого
  - марки приобретаемого телефона
  - типа используемого браузера
  - бронируемого отеля из числа доступных в городе альтернатив
  - музыкального сервиса

- Иногда зависимая переменная может быть номинальной, то есть принимать одно из конечного числа взаимоисключающих значений.
- Часто рассматриваются номинальные переменные, отражающие выбор индивида, например, в отношении:
  - типа получаемого образования: техническое, гуманитарное или никакого
  - марки приобретаемого телефона
  - типа используемого браузера
  - бронируемого отеля из числа доступных в городе альтернатив
  - музыкального сервиса
- Модели мультиномиального (множественного) выбора позволяет оценить вероятности выбора индивидом той или иной альтернативы, а также то, как различные факторы влияют на эти вероятности.

- Индивид выбирает одну из J альтернатив:

$$y_{ji}^{*} = x_i\beta_j + \varepsilon_{ji},$$

где y_{ji}^{*} отражает склонность к выбору j-й альтернативы i-м индивидом.

Модели мультиномиального выбора

Общая постановка

- Индивид выбирает одну из J альтернатив:

$$y_{ji}^{*} = x_i\beta_j + \varepsilon_{ji},$$

где y_{ji}^{*} отражает склонность к выбору j-й альтернативы i-м индивидом.

- Индивид выбирает альтернативу, к выбору которой он проявил наибольшую склонность:

$$y_i = \begin{cases} 1 \text{ если } y_{1i}^{*} > y_{2i}^{*}, y_{1i}^{*} > y_{3i}^{*}, ..., y_{1i}^{*} > y_{Ji}^{*} \\ 2 \text{ если } y_{2i}^{*} > y_{1i}^{*}, y_{2i}^{*} > y_{3i}^{*}, ..., y_{2i}^{*} > y_{Ji}^{*} \\ \vdots \\ J \text{ если } y_{Ji}^{*} > y_{1i}^{*}, y_{Ji}^{*} > y_{2i}^{*}, ..., y_{Ji}^{*} > y_{(J-1)i}^{*} \end{cases}$$

- Индивид выбирает одну из J альтернатив:

$$y_{ji}^{*} = x_i\beta_j + \varepsilon_{ji},$$

где y_{ji}^{*} отражает склонность к выбору j-й альтернативы i-м индивидом.

- Индивид выбирает альтернативу, к выбору которой он проявил наибольшую склонность:

$$y_i = \begin{cases} 1 \text{ если } y_{1i}^{*} > y_{2i}^{*}, y_{1i}^{*} > y_{3i}^{*}, ..., y_{1i}^{*} > y_{Ji}^{*} \\ 2 \text{ если } y_{2i}^{*} > y_{1i}^{*}, y_{2i}^{*} > y_{3i}^{*}, ..., y_{2i}^{*} > y_{Ji}^{*} \\ \vdots \\ J \text{ если } y_{Ji}^{*} > y_{1i}^{*}, y_{Ji}^{*} > y_{2i}^{*}, ..., y_{Ji}^{*} > y_{(J-1)i}^{*} \end{cases}$$

- На практике мы наблюдаем только y_i, то есть для каждого индивида наблюдается лишь наиболее предпочтительный для него выбор.

Мультиномиальная логит модель

Формулировка

- В мультиномиальной логит модели предполагается, что случайные ошибки ε_{ji} независимы и одинаково распределены в соответствии с распределением Гумбеля, имеющим следующие функцию распределения и функцию плотности:

$$F_{\varepsilon_{ji}}(t) = e^{-e^{-t}}, \qquad f_{\varepsilon_{ji}}(x) = e^{-t-e^{-t}}, \quad E(\varepsilon_{ji}) \approx 0.577$$

- В мультиномиальной логит модели предполагается, что случайные ошибки ε_{ji} независимы и одинаково распределены в соответствии с распределением Гумбеля, имеющим следующие функцию распределения и функцию плотности:

$$F_{\varepsilon_{ji}}(t) = e^{-e^{-t}}, \qquad f_{\varepsilon_{ji}}(x) = e^{-t-e^{-t}}, \quad E(\varepsilon_{ji}) \approx 0.577$$

- С помощью формулы свертки можно вывести выражение для условной вероятности выбора k-й альтернативы:

$$P(y_i = k|x_i) = P(y_{ik}^* > y_{i1}^*, .., y_{ik}^* > y_{iJ}^*) =$$

- В мультиномиальной логит модели предполагается, что случайные ошибки ε_{ji} независимы и одинаково распределены в соответствии с распределением Гумбеля, имеющим следующие функцию распределения и функцию плотности:

$$F_{\varepsilon_{ji}}(t) = e^{-e^{-t}}, \qquad f_{\varepsilon_{ji}}(x) = e^{-t-e^{-t}}, \quad E(\varepsilon_{ji}) \approx 0.577$$

- С помощью формулы свертки можно вывести выражение для условной вероятности выбора k-й альтернативы:

$$P(y_i = k|x_i) = P(y_{ik}^* > y_{i1}^*, .., y_{ik}^* > y_{iJ}^*) =$$
$$= P(x_i\beta_k + \varepsilon_{ki} > x_i\beta_1 + \varepsilon_{1i}, ..., x_i\beta_k + \varepsilon_{ki} > x_i\beta_J + \varepsilon_J|x_i) =$$

- В мультиномиальной логит модели предполагается, что случайные ошибки ε_{ji} независимы и одинаково распределены в соответствии с распределением Гумбеля, имеющим следующие функцию распределения и функцию плотности:

$$F_{\varepsilon_{ji}}(t) = e^{-e^{-t}}, \qquad f_{\varepsilon_{ji}}(x) = e^{-t-e^{-t}}, \quad E(\varepsilon_{ji}) \approx 0.577$$

- С помощью формулы свертки можно вывести выражение для условной вероятности выбора k-й альтернативы:

$$P(y_i = k|x_i) = P(y_{ik}^* > y_{i1}^*, .., y_{ik}^* > y_{iJ}^*) =$$
$$= P(x_i\beta_k + \varepsilon_{ki} > x_i\beta_1 + \varepsilon_{1i}, ..., x_i\beta_k + \varepsilon_{ki} > x_i\beta_J + \varepsilon_J|x_i) =$$
$$= P(\varepsilon_{1i} - \varepsilon_k < x_k\beta_k - x_i\beta_1, ..., \varepsilon_{Ji} - \varepsilon_k < x_i\beta_k - x_i\beta_J|x_i) = e^{x_i\beta_k} / \sum_{j=1}^{J} e^{x_i\beta_j}$$

- В мультиномиальной логит модели предполагается, что случайные ошибки ε_{ji} независимы и одинаково распределены в соответствии с распределением Гумбеля, имеющим следующие функцию распределения и функцию плотности:

$$F_{\varepsilon_{ji}}(t) = e^{-e^{-t}}, \qquad f_{\varepsilon_{ji}}(x) = e^{-t-e^{-t}}, \quad E(\varepsilon_{ji}) \approx 0.577$$

- С помощью формулы свертки можно вывести выражение для условной вероятности выбора k-й альтернативы:

$$P(y_i = k|x_i) = P(y^*_{ik} > y^*_{i1}, .., y^*_{ik} > y^*_{iJ}) =$$
$$= P(x_i\beta_k + \varepsilon_{ki} > x_i\beta_1 + \varepsilon_{1i}, ..., x_i\beta_k + \varepsilon_{ki} > x_i\beta_J + \varepsilon_J|x_i) =$$
$$= P(\varepsilon_{1i} - \varepsilon_k < x_k\beta_k - x_i\beta_1, ..., \varepsilon_{Ji} - \varepsilon_k < x_i\beta_k - x_i\beta_J|x_i) = e^{x_i\beta_k} / \sum_{j=1}^{J} e^{x_i\beta_j}$$

- Сформировав с помощью этих вероятностей функцию правдоподобия можно получить ММП оценки параметров $\beta_1, ..., \beta_J$.

- Вычитание из каждого из векторов регрессионных коэффициентов β_j любого вектора v соответствующих размеров не приведет к изменению вероятностей:

$$P(y_i = k|x_i) = \frac{e^{x_i(\beta_k - v)}}{\sum\limits_{j=1}^{J} e^{x_i\left(\beta_j - v\right)}} =$$

- Вычитание из каждого из векторов регрессионных коэффициентов β_j любого вектора v соответствующих размеров не приведет к изменению вероятностей:

$$P(y_i = k|x_i) = \frac{e^{x_i(\beta_k - v)}}{\sum\limits_{j=1}^{J} e^{x_i\left(\beta_j - v\right)}} = \frac{e^{-x_i v} e^{x_i \beta_k}}{e^{-x_i v}\sum\limits_{j=1}^{J} e^{x_i \beta_j}} =$$

- Вычитание из каждого из векторов регрессионных коэффициентов β_j любого вектора v соответствующих размеров не приведет к изменению вероятностей:

$$P(y_i = k|x_i) = \frac{e^{x_i(\beta_k - v)}}{\sum_{j=1}^{J} e^{x_i\left(\beta_j - v\right)}} = \frac{e^{-x_i v} e^{x_i\beta_k}}{e^{-x_i v}\sum_{j=1}^{J} e^{x_i\beta_j}} = \frac{e^{x_i\beta_k}}{\sum_{j=1}^{J} e^{x_i\beta_j}}$$

- Вычитание из каждого из векторов регрессионных коэффициентов β_j любого вектора v соответствующих размеров не приведет к изменению вероятностей:

$$P(y_i = k|x_i) = \frac{e^{x_i(\beta_k - v)}}{\sum\limits_{j=1}^{J} e^{x_i\left(\beta_j - v\right)}} = \frac{e^{-x_i v} e^{x_i\beta_k}}{e^{-x_i v}\sum\limits_{j=1}^{J} e^{x_i\beta_j}} = \frac{e^{x_i\beta_k}}{\sum\limits_{j=1}^{J} e^{x_i\beta_j}}$$

- Следовательно, максимизируя функцию правдоподобия по всем β_j мы получим бесконечное число решений, что порождает проблему неидентифицируемости. Поэтому без потери общности полагают $\beta_J = (0, ..., 0)$, откуда:

$$P(y_i = k|x_i) = \frac{e^{x_i\beta_k}}{1 + \sum\limits_{j=1}^{J-1} e^{x_i\beta_j}}$$

- Рассмотрим предельный эффект вероятности выбора k-й альтернативы по t-му регрессору:

$$\frac{\partial P(y_i = k|x_i)}{\partial x_{ti}} = P(y_i = k|x_i) \times \underbrace{\left[\beta_{kt} - \sum_{j=1}^{J} P(y_i = j|x_i)\beta_{jt}\right]}_{\text{во сколько раз возрастает вероятность}}$$

Мультиномиальная логит модель

Предельные эффекты

- Рассмотрим предельный эффект вероятности выбора k-й альтернативы по t-му регрессору:

$$\frac{\partial P(y_i = k|x_i)}{\partial x_{ti}} = P(y_i = k|x_i) \times \underbrace{\left[\beta_{kt} - \sum_{j=1}^{J} P(y_i = j|x_i)\beta_{jt}\right]}_{\text{во сколько раз возрастает вероятность}}$$

- Знак предельного эффекта не определяется знаком β_{kt}, за исключением двух случаев:

- Рассмотрим предельный эффект вероятности выбора k-й альтернативы по t-му регрессору:

$$\frac{\partial P(y_i = k|x_i)}{\partial x_{ti}} = P(y_i = k|x_i) \times \underbrace{\left[\beta_{kt} - \sum_{j=1}^{J} P(y_i = j|x_i)\beta_{jt}\right]}_{\text{во сколько раз возрастает вероятность}}$$

- Знак предельного эффекта не определяется знаком β_{kt}, за исключением двух случаев:
  - Если $\beta_{kt} = \max(\beta_{1t}, ..., \beta_{Jt})$, то предельный эффект положительный.

- Рассмотрим предельный эффект вероятности выбора k-й альтернативы по t-му регрессору:

$$\frac{\partial P(y_i = k|x_i)}{\partial x_{ti}} = P(y_i = k|x_i) \times \underbrace{\left[\beta_{kt} - \sum_{j=1}^{J} P(y_i = j|x_i)\beta_{jt}\right]}_{\text{во сколько раз возрастает вероятность}}$$

- Знак предельного эффекта не определяется знаком β_{kt}, за исключением двух случаев:
    - Если $\beta_{kt} = \max(\beta_{1t}, ..., \beta_{Jt})$, то предельный эффект положительный.
    - Если $\beta_{kt} = \min(\beta_{1t}, ..., \beta_{Jt})$, то предельный эффект отрицательный.

- Для осуществления предсказания необходимо рассчитать вероятность выбора индивидом каждой из альтернатив.

Мультиномиальная логит модель

Предсказание

- Для осуществления предсказания необходимо рассчитать вероятность выбора индивидом каждой из альтернатив.
- Предполагается, что индивид выбирает ту альтернативу, вероятность выбора которой является наибольшей.

- Для осуществления предсказания необходимо рассчитать вероятность выбора индивидом каждой из альтернатив.
- Предполагается, что индивид выбирает ту альтернативу, вероятность выбора которой является наибольшей.
- Это эквивалентно тому, чтобы присвоить индивиду ту альтернативу, склонность к выбору которой является максимальной.

- Для осуществления предсказания необходимо рассчитать вероятность выбора индивидом каждой из альтернатив.
- Предполагается, что индивид выбирает ту альтернативу, вероятность выбора которой является наибольшей.
- Это эквивалентно тому, чтобы присвоить индивиду ту альтернативу, склонность к выбору которой является максимальной.
- Например, Лаврентий выбирает между походом в кино, в театр и в оперу. Если в соответствии с нашими оценками он пойдет в кино с вероятностью 0.5, в театр с вероятностью 0.3 и в оперу с вероятностью 0.2, то мы предполагаем, что Лаврентий выберет кино.

- Предположим, что склонность формируется следующим образом:

$$y_{ji}^{*}=x_{i}\beta_{j}+x_{i}^{j}\beta^{*}+\varepsilon_{ji}$$

- Предположим, что склонность формируется следующим образом:

$$y_{ji}^{*} = x_i\beta_j + x_i^j\beta^* + \varepsilon_{ji}$$

где:

  - x_i – регрессоры, общие для всех альтернатив

- Предположим, что склонность формируется следующим образом:

$$y_{ji}^{*} = x_i\beta_j + x_i^j\beta^* + \varepsilon_{ji}$$

где:

  - x_i – регрессоры, общие для всех альтернатив
  - x_i^j – регрессоры, различающиеся между альтернативами

- Предположим, что склонность формируется следующим образом:
$$y_{ji}^{*} = x_i\beta_j + x_i^{j}\beta^{*} + \varepsilon_{ji}$$
где:
  - x_i – регрессоры, общие для всех альтернатив
  - x_i^{j} – регрессоры, различающиеся между альтернативами
  - β_j – регрессионные коэффициенты, учитывающие, что не различающиеся регрессоры могут по разному влиять на вероятность выбора той или иной альтернативы

- Предположим, что склонность формируется следующим образом:

$$y_{ji}^{*}=x_i\beta_j+x_i^j\beta^{*}+\varepsilon_{ji}$$

где:
  - x_i – регрессоры, общие для всех альтернатив
  - x_i^j – регрессоры, различающиеся между альтернативами
  - β_j – регрессионные коэффициенты, учитывающие, что не различающиеся регрессоры могут по разному влиять на вероятность выбора той или иной альтернативы
  - β^{*} – регрессионные коэффициенты, учитывающие, что разные регрессоры могут одинаковым образом влиять на вероятность выбора той или иной альтернативы

- Предположим, что склонность формируется следующим образом:

$$y_{ji}^{*} = x_i\beta_j + x_i^j\beta^{*} + \varepsilon_{ji}$$

  где:
  - x_i – регрессоры, общие для всех альтернатив
  - x_i^j – регрессоры, различающиеся между альтернативами
  - β_j – регрессионные коэффициенты, учитывающие, что не различающиеся регрессоры могут по разному влиять на вероятность выбора той или иной альтернативы
  - β^{*} – регрессионные коэффициенты, учитывающие, что разные регрессоры могут одинаковым образом влиять на вероятность выбора той или иной альтернативы
- Например, вы можете рассматривать модель, предсказывающую выбор индивидом средства поездки на работу: машина, такси или общественный транспорт. Тогда x_i будут отражать характеристики индивида (пол, возраст, доход), а x_i^j – характеристики альтернатив (стоимость и комфорт поездки на каждом из видов транспорта).

- Предположим, что склонность формируется следующим образом:

$$y_{ji}^{*} = x_i\beta_j + x_i^j\beta^* + \varepsilon_{ji}$$

  где:
  - x_i – регрессоры, общие для всех альтернатив
  - x_i^j – регрессоры, различающиеся между альтернативами
  - β_j – регрессионные коэффициенты, учитывающие, что не различающиеся регрессоры могут по разному влиять на вероятность выбора той или иной альтернативы
  - β^* – регрессионные коэффициенты, учитывающие, что разные регрессоры могут одинаковым образом влиять на вероятность выбора той или иной альтернативы
- Например, вы можете рассматривать модель, предсказывающую выбор индивидом средства поездки на работу: машина, такси или общественный транспорт. Тогда x_i будут отражать характеристики индивида (пол, возраст, доход), а x_i^j – характеристики альтернатив (стоимость и комфорт поездки на каждом из видов транспорта).
- Например, доход может по разному влиять на склонность к поездке на том или ином виде транспорта, а цена поездки – одинаково.

- Предельные эффекты по регрессорам x_i аналогичны рассмотренным ранее.

- Предельные эффекты по регрессорам x_i аналогичны рассмотренным ранее.
- Для краткости обозначим $P_{ki} = P(y_i = k|x_i, x_i^1, ..., x_i^J)$.

Мультиномиальная логит модель

Предельные эффекты в условной мультиномиальная логит модель

- Предельные эффекты по регрессорам x_i аналогичны рассмотренным ранее.
- Для краткости обозначим $P_{ki} = P(y_i = k|x_i, x_i^1, ..., x_i^J)$.
- Предельный эффект t-го регрессора из x_i^k на условную вероятность выбора k-й альтернативы:

$$\frac{\partial P_{ki}}{\partial x_{ti}^k} = P_{ki}\left(1 - P_{ki}\right)\beta_t^*$$

- Предельные эффекты по регрессорам x_i аналогичны рассмотренным ранее.
- Для краткости обозначим $P_{ki} = P(y_i = k|x_i, x_i^1, ..., x_i^J)$.
- Предельный эффект t-го регрессора из x_i^k на условную вероятность выбора k-й альтернативы:

$$\frac{\partial P_{ki}}{\partial x_{ti}^k} = P_{ki}\left(1 - P_{ki}\right)\beta_t^*$$

Данный предельный эффект может отражать как, например, изменение стоимости поездки на такси влияет на вероятность поездки на такси. Знак данного предельного эффекта совпадает со знаком коэффициента β_t^* при стоимости поездки на такси.

Мультиномиальная логит модель

Предельные эффекты в условной мультиномиальная логит модель

- Предельные эффекты по регрессорам x_i аналогичны рассмотренным ранее.
- Для краткости обозначим $P_{ki} = P(y_i = k|x_i, x_i^1, ..., x_i^J)$.
- Предельный эффект t-го регрессора из x_i^k на условную вероятность выбора k-й альтернативы:

$$\frac{\partial P_{ki}}{\partial x_{ti}^k} = P_{ki}\left(1 - P_{ki}\right)\beta_t^*$$

  Данный предельный эффект может отражать как, например, изменение стоимости поездки на такси влияет на вероятность поездки на такси. Знак данного предельного эффекта совпадает со знаком коэффициента β_t^* при стоимости поездки на такси.
- Предельный эффект t-го регрессора из x_i^j на условную вероятность выбора k-й альтернативы:

$$\frac{\partial P_{ki}}{\partial x_{ti}^j} = -P_{ki}P_{ji}\beta_t^*$$

- Предельные эффекты по регрессорам x_i аналогичны рассмотренным ранее.
- Для краткости обозначим $P_{ki} = P(y_i = k|x_i, x_i^1, ..., x_i^J)$.
- Предельный эффект t-го регрессора из x_i^k на условную вероятность выбора k-й альтернативы:

$$\frac{\partial P_{ki}}{\partial x_{ti}^k} = P_{ki}\left(1 - P_{ki}\right)\beta_t^*$$

  Данный предельный эффект может отражать как, например, изменение стоимости поездки на такси влияет на вероятность поездки на такси. Знак данного предельного эффекта совпадает со знаком коэффициента β_t^* при стоимости поездки на такси.

- Предельный эффект t-го регрессора из x_i^j на условную вероятность выбора k-й альтернативы:

$$\frac{\partial P_{ki}}{\partial x_{ti}^j} = -P_{ki}P_{ji}\beta_t^*$$

  Данный предельный эффект может отражать как, например, влияет изменение стоимости поездки на такси на вероятность поездки на машине. Знак данного предельного эффекта противоположен знаку коэффициента β_t^* при стоимости поездки на такси. Например, при $\beta_t^* < 0$ мы получаем, что повышение стоимости поездки на такси приводит, при прочих равных, к увеличению вероятности поездки на машине.

- Из-за того, что случайные ошибки независимы между альтернативами, отношения вероятностей выбора для двух альтернатив не зависит от вероятностей выбора других альтернатив:

$$\frac{P(y_i = k|x_i)}{P(y_i = j|x_i)} = \frac{e^{x_i\beta_k}}{e^{x_i\beta_j}}, \quad \forall k, j \in \{1, ..., J\}$$

- Из-за того, что случайные ошибки независимы между альтернативами, отношения вероятностей выбора для двух альтернатив не зависит от вероятностей выбора других альтернатив:

$$\frac{P(y_i = k|x_i)}{P(y_i = j|x_i)} = \frac{e^{x_i\beta_k}}{e^{x_i\beta_j}}, \quad \forall k, j \in \{1, ..., J\}$$

- Это свойство мультиномиальной логит модели именуют независимостью от посторонних альтернатив, поскольку выбор между двумя альтернативами не зависит от наличия других альтернатив (IIA - independence of irrelevant alternatives), что считается нереалистичным.

- Из-за того, что случайные ошибки независимы между альтернативами, отношения вероятностей выбора для двух альтернатив не зависит от вероятностей выбора других альтернатив:

$$\frac{P(y_i = k|x_i)}{P(y_i = j|x_i)} = \frac{e^{x_i\beta_k}}{e^{x_i\beta_j}}, \quad \forall k, j \in \{1, ..., J\}$$

- Это свойство мультиномиальной логит модели именуют независимостью от посторонних альтернатив, поскольку выбор между двумя альтернативами не зависит от наличия других альтернатив (IIA - independence of irrelevant alternatives), что считается нереалистичным.
- Для проверки соблюдения данного допущения применяется тест Хаусмана-Макфаддена. Нулевая гипотеза теста предполагает выполнение (IIA). Если нулевая гипотеза отклоняется, то оценки мультиномиальной логит модели могут оказаться несостоятельными.

Мультиномиальная пробит модель

Основные особенности

- Допущение о независимости от посторонних альтернатив можно ослабить за счет снятия допущения о независимости случайных ошибок.

Мультиномиальная пробит модель

Основные особенности

- Допущение о независимости от посторонних альтернатив можно ослабить за счет снятия допущения о независимости случайных ошибок.
- Мультиномиальная пробит модель предполагает, что случайные ошибки имеют многомерное нормальное распределение (для простоты и без потери общности положим $J = 3$):

$$\varepsilon_i = \begin{bmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \\ \varepsilon_{3i} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \rho_{13}\sigma_1\sigma_2 \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 & \rho_{23}\sigma_1\sigma_3 \\ \rho_{13}\sigma_1\sigma_2 & \rho_{23}\sigma_1\sigma_3 & \sigma_3^2 \end{bmatrix}\right)$$

Мультиномиальная пробит модель

Основные особенности

- Допущение о независимости от посторонних альтернатив можно ослабить за счет снятия допущения о независимости случайных ошибок.
- Мультиномиальная пробит модель предполагает, что случайные ошибки имеют многомерное нормальное распределение (для простоты и без потери общности положим $J = 3$):

$$\varepsilon_i = \begin{bmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \\ \varepsilon_{3i} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \rho_{13}\sigma_1\sigma_2 \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 & \rho_{23}\sigma_1\sigma_3 \\ \rho_{13}\sigma_1\sigma_2 & \rho_{23}\sigma_1\sigma_3 & \sigma_3^2 \end{bmatrix}\right)$$

- Из соображений идентифицируемости полагают $y_{3i}^* = 0$ и $\sigma_1 = 1$, что позволяет рассматривать совместное распределение случайных ошибок лишь первых двух альтернатив:

$$\begin{bmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix} \begin{bmatrix} 1 & \rho_{12}\sigma_2 \\ \rho_{12}\sigma_2 & \sigma_2^2 \end{bmatrix}\right)$$

- Максимизация функция правдоподобия является достаточно сложной технической задачей, особенно при большом числе альтернатив $J \geq 5$. Получение оценок параметров данной модели может занять несколько часов даже на достаточно продвинутых компьютерах.

Мультиномиальная пробит модель

Недостатки

- Максимизация функция правдоподобия является достаточно сложной технической задачей, особенно при большом числе альтернатив $J \geq 5$. Получение оценок параметров данной модели может занять несколько часов даже на достаточно продвинутых компьютерах.
- Формально параметры модели идентифицируются, однако (грубо говоря), параметры β_j коллинеарны параметрам ковариационной матрицы случайных ошибок, что осложняет получение точных оценок β_j. Однако, это не мешает получать точные предсказания вероятностей.