Эконометрика

Шишкин Владимир Андреевич

Пермский государственный национальный исследовательский университет

Теорема Гаусса–Маркова

Предположим, что

1. $y = X\beta + \varepsilon$;
2. X — детерминированная $n \times (m+1)$-матрица ранга $m+1$;
3. $\mathsf{E}\,\varepsilon = 0$, $\mathsf{D}\,\varepsilon = \mathsf{E}\,(\varepsilon\varepsilon^\top) = \sigma^2 I_n$.

Тогда оценка метода наименьших квадратов является наиболее эффективной (в смысле наименьшей дисперсии) оценкой в классе линейных несмещённых оценок.

Best Linear Unbiased Estimator — BLUE

Проверка гипотезы о линейном ограничении общего вида $\mathrm{H}_0\colon H\beta = r$

H — $q \times (m+1)$-матрица, $q \leqslant m+1$

$$F_{\text{набл}} = \frac{(Hb-r)^\top \left(H(X^\top X)^{-1}H^\top\right)^{-1}(Hb-r)/q}{e^\top e/(n-m-1)} \sim$$
$$\sim F_{q,n-m-1}$$

Гипотеза H_0 отклоняется на уровне значимости α, если $F_{\text{набл}} > F_{\alpha,m,n-m-1}$

Проверка гипотезы о линейном ограничении общего вида $\mathrm{H}_0\colon H\beta = r$

H — $q \times (m+1)$-матрица, $q \leqslant m+1$

$$F_{\text{набл}} = \frac{(Hb-r)^\top \big(H(X^\top X)^{-1}H^\top)\big)^{-1}(Hb-r)/q}{e^\top e/(n-m-1)} \sim$$

$$\sim F_{q,n-m-1}$$

Гипотеза H_0 отклоняется на уровне значимости α, если $F_{\text{набл}} > F_{\alpha,m,n-m-1}$

Пример: условие

$$H = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \end{pmatrix} \quad r = \begin{pmatrix} 2 \\ 0 \end{pmatrix}$$

соответствует системе из двух линейных ограничений:

$$\beta_0 = 2$$

$$\beta_1 - \beta_2 = 0$$

Проверка гипотезы о линейном ограничении общего вида $\mathrm{H}_0: H\beta = r$

Пример: $H = I$

$$F_{\text{набл}} = \frac{(b-r)^\top (X^\top X)(b-r)/m}{e^\top e/(n-m-1)} \sim F_{m,n-m-1}$$

Доверительная область — эллипсоид в $(m+1)$-мерном пространстве коэффициентов β

Проверка гипотезы $H_0: \beta_{m-q} = \beta_{m-q+1} = \cdots = \beta_m = 0$

$X = [X_1, X_2]$
X_1 — $n \times (m+1-q)$ — матрица; X_2 — $(n \times q)$ — матрица

$$y = X\beta + \varepsilon, \quad y^* = X_1\beta_1 + \varepsilon^*$$
$$b = (X^\top X)^{-1} X^\top y, \quad b_1 = (X_1^\top X_1)^{-1} X_1^\top y$$
$$e = y - Xb, \quad e^* = y - X_1 b_1$$

$$F_{\text{набл}} = \frac{(e^{*\top}e^* - e^\top e)/q}{e^\top e/(n-m-1)} = \frac{(\mathrm{ESS_R} - \mathrm{ESS_{UR}})/q}{\mathrm{ESS_{UR}}/(n-m-1)} =$$
$$= \frac{(R^2_{\mathrm{UR}} - R^2_{\mathrm{R}})/q}{(1 - R^2_{\mathrm{UR}})/(n-m-1)} \sim F_{q,n-m-1}$$

Проверка гипотезы H_0: $\beta' = \beta''$, $\sigma' = \sigma''$ (тест Чоу)

$$X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$$

X_1 — $n_1 \times (m+1)$-матрица; X_2 — $n_2 \times (m+1)$-матрица

$y' = X_1 b' + \varepsilon'$, $\quad y'' = X_2 b'' + \varepsilon''$ $\qquad$ или $\qquad$ $y = Xb + \varepsilon$?

$$F_{\text{набл}} = \frac{(\mathrm{ESS_R} - \mathrm{ESS_{UR}})/(m+1)}{\mathrm{ESS_{UR}}/(n_1 + n_2 - 2m - 2)} \sim F_{m+1, n_1+n_2-2m-2}$$

Регрессионные модели

1. Модели, линейные по параметрам

$$y = \sum_{i=1}^{m} \beta_i \phi_i(x) + \varepsilon$$

2. Линеаризуемые модели

$$y = f(x, \varepsilon) \quad \Rightarrow \quad \tilde{y} = \sum_{i=1}^{m} \beta_i \phi_i(x) + \varepsilon$$

3. Существенно нелинейные модели

$$y = f(x) + \varepsilon$$

Модели с фиктивными переменными

Модели, имеющие в своём составе качественные независимые переменные.

Качественная переменная, имеющая k значений, заменяется $(k-1)$-й бинарной переменной (принимающей значения 0 или 1).

	x_1	x_2	x_3
зима	0	0	0
весна	1	0	0
лето	0	1	0
осень	0	0	1

Модели с фиктивными переменными

- *Аддитивная модель*:

$$y = \beta x + \tilde{\beta}\tilde{x} + \varepsilon$$

(сдвиг по вертикали на величину $\tilde{\beta}$)

- *Мультипликативная модель*:

$$y = \beta(1 + \tilde{\beta}\tilde{x})x + \varepsilon$$

(изменение угла наклона)

- *Смешанная модель*

Пример

Оценка вероятности. Биномиальное распределение

Выборка из $\mathcal{B}(n, p)$, $n = 10$: 7 единиц и 3 нуля.
Каково значение параметра p?

Пример

Оценка вероятности. Биномиальное распределение

Выборка из $\mathcal{B}(n, p)$, $n = 10$: 7 единиц и 3 нуля.
Вероятность 7 успехов в 10 испытаниях:

$$h(p) = C_{10}^{7} p^{7} (1-p)^{3}, \quad C_n^k = \frac{n!}{k!(n-k)!}$$

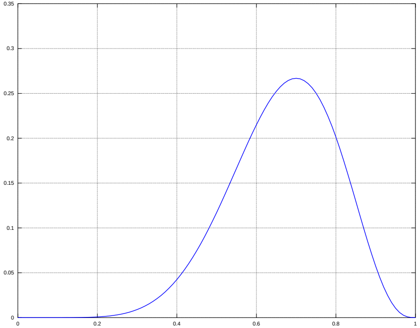
Пример

Оценка вероятности. Биномиальное распределение

Выборка из $\mathcal{B}(n,p)$, $n=10$: 7 единиц и 3 нуля.

Вероятность 7 успехов в 10 испытаниях:

$$h(p) = C_{10}^{7} p^{7} (1-p)^{3}, \quad C_n^k = \frac{n!}{k!(n-k)!}$$

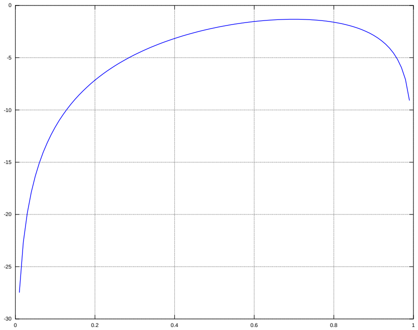


Пример

Оценка вероятности. Биномиальное распределение

Выборка из $\mathcal{B}(n, p)$, $n = 10$: 7 единиц и 3 нуля.

$$\ln\big(h(p)\big) = \ln(120) + 7\ln(p) + 3\ln(1-p)$$



Пример

Оценка вероятности. Биномиальное распределение

Выборка из $\mathcal{B}(n, p)$, $n = 10$: 7 единиц и 3 нуля.

$$\frac{d \ln h(p)}{dp} = \frac{7}{p} - \frac{3}{1-p}$$

Пример

Оценка вероятности. Биномиальное распределение

Выборка из $\mathcal{B}(n, p)$, $n = 10$: 7 единиц и 3 нуля.
При $\hat{p} = 0.7$ вероятность получения такой выборки максимальна

Пример

Оценка вероятности. Биномиальное распределение

Выборка из $\mathcal{B}(n, p)$.
Вероятность получения x успехов в n испытаниях:

$$C_n^x p^x (1-p)^{n-x}$$

Пример

Оценка вероятности. Биномиальное распределение

Выборка из $\mathcal{B}(n, p)$.
Вероятность получения x успехов в n испытаниях:

$$C_n^x p^x (1-p)^{n-x}$$

1. $f(x \mid n, p)$ — *функция распределения* distribution
2. $f(p \mid n, x)$ — *функция правдоподобия* likelihood function

Метод максимального правдоподобия

Последовательность случайных величин

$$\{y_1, y_2, \ldots, y_n\}$$

$h_n(\cdot, \theta_0)$ — совместная плотность распределения случайных величин $y = (y_1, y_2, \ldots, y_n)$
Требуется оценить вектор параметров $\theta_0 \in \Theta \subseteq \mathbb{R}^m$

Метод максимального правдоподобия

Для каждого (фиксированного) y вещественная функция

$$L_n(\theta) = L_n(\theta, y) = h_n(y, \theta), \quad \theta \in \Theta,$$

называется *функцией правдоподобия*.
$\ln L_n(\theta)$ — *логарифмическая функция правдоподобия*
Для фиксированного y любое значение $\hat{\theta}_n \in \Theta$, такое, что

$$L_n(\hat{\theta}_n(y), y) = \sup_{\theta \in \Theta} L_n(\theta, y),$$

называется *оценкой максимального правдоподобия* неизвестного параметра θ.

Метод максимального правдоподобия

Если

1. максимум

$$L_n(\hat{\theta}_n(y), y) = \sup_{\theta \in \Theta} L_n(\theta, y),$$

   достигается во внутренней точке пространства параметров Θ,
2. $L_n(\theta)$ — дифференцируемая функция

то

$$\frac{\partial \ln L_n(\theta)}{\partial \theta} = 0.$$

Оценка максимального правдоподобия параметров многомерного нормального распределения

$$(y_1, \ldots, y_n),\ y_i \sim \mathcal{N}(\mu, \sigma^2 I_m),\ y_i \in \mathbb{R}^m$$

Функция плотности вероятности для $y_i,\ i = 1, \ldots, n$

$$h(y_i) = (2\pi\sigma^2)^{-m/2} \exp\left(-\frac{(y_i - \mu)^\top (y_i - \mu)}{2\sigma^2}\right)$$

Функция правдоподобия

$$L = \prod_{i=1}^{n} h(y_i)$$

Оценка максимального правдоподобия параметров многомерного нормального распределения

$$(y_1, \dots, y_n),\ y_i \sim \mathcal{N}(\mu, \sigma^2 I_m),\ y_i \in \mathbb{R}^m$$

Логарифмическая функция правдоподобия

$$\ln L = -\frac{mn}{2}\ln(2\pi) - \frac{mn}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu)^\top (y_i - \mu)$$

Оценка максимального правдоподобия параметров многомерного нормального распределения

$$(y_1, \dots, y_n),\ y_i \sim \mathcal{N}(\mu, \sigma^2 I_m),\ y_i \in \mathbb{R}^m$$

Необходимые условия экстремума

$$\frac{\partial \ln L}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i - \mu) = 0,$$

$$\frac{\partial \ln L}{\partial \sigma^2} = -\frac{mn}{2\sigma^2} + \frac{1}{\sigma^4}\sum_{i=1}^{n}(y_i - \mu)^\top (y_i - \mu) = 0.$$

Оценка максимального правдоподобия параметров многомерного нормального распределения

$$(y_1, \ldots, y_n),\ y_i \sim \mathcal{N}(\mu, \sigma^2 I_m),\ y_i \in \mathbb{R}^m$$

Оценки максимального правдоподобия

$$\hat{\mu}_j = \bar{y}_j, \quad \hat{\sigma}^2 = \frac{1}{mn} \sum_{j=1}^{m} \sum_{i=1}^{n} (y_{j,i} - \bar{y}_j)^2$$

Свойства оценок максимального правдоподобия

1. *Инвариантность*: если $\hat{\theta}$ — оценка МП параметра θ и g — непрерывная функция, то $g(\hat{\theta})$ — оценка МП параметра $g(\theta)$.
2. *Состоятельность*: $p\lim\hat{\theta} = 0$, т.е.
$$\forall\varepsilon > 0\colon\ \mathsf{P}\big(|\hat{\theta} - \theta| > \varepsilon\big) \to 0 \text{ при } n \to \infty.$$
3. *Асимптотическая нормальность:*
$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} \mathcal{N}\big(0, F(\theta)^{-1}\big).$$
4. *Асимптотическая эффективность:* если $\hat{\theta}$ — оценка МП, то
$$\mathsf{D}\,\hat{\theta} \leqslant \mathsf{D}\,\tilde{\theta},$$
где $\tilde{\theta}$ — любая другая состоятельная и асимптотически нормальная оценка.

Оценка максимального правдоподобия в линейной модели

Стандартная линейная модель:

$$y = X\beta + u, \quad u \sim \mathcal{N}(0, \sigma^2 I_n)$$

Оценка максимального правдоподобия в линейной модели

Стандартная линейная модель:

$$y \sim \mathcal{N}(X\beta, \sigma^2 I_n)$$

Оценка максимального правдоподобия в линейной модели

Стандартная линейная модель:

$$y \sim \mathcal{N}(X\beta, \sigma^2 I_n)$$

Плотность распределения случайного вектора y:

$$h(y) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{(y - X\beta)^\top (y - X\beta)}{2\sigma^2}\right)$$

Оценка максимального правдоподобия в линейной модели

Стандартная линейная модель:

$$y \sim \mathcal{N}(X\beta, \sigma^2 I_n)$$

Логарифмическая функция правдоподобия:

$$\ln L(\beta, \sigma^2) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}(y - X\beta)^\top (y - X\beta)$$

Оценка максимального правдоподобия в линейной модели

Стандартная линейная модель:

$$y \sim \mathcal{N}(X\beta, \sigma^2 I_n)$$

Необходимые условия существования экстремума:

$$\frac{\partial \ln L}{\partial \beta^\top} = \frac{1}{\sigma^2}(y - X\beta)^\top X = 0,$$
$$\frac{\partial \ln L}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}(y - X\beta)^\top (y - X\beta) = 0.$$

Оценка максимального правдоподобия в линейной модели

Стандартная линейная модель:

$$y \sim \mathcal{N}(X\beta, \sigma^2 I_n)$$

Оценки максимального правдоподобия

$$\hat{\beta} = (X^\top X)^{-1} X^\top y, \quad \hat{\sigma}^2 = \frac{1}{n} e^\top e$$

Оценка МП для β совпадает с оценкой метода наименьших квадратов.
Оценка МП для σ^2 *не равна* оценке метода наименьших квадратов $s^2 = e^\top e/(n-m-1)$.

Модель бинарного выбора

Зависимая переменная принимает значения 0 или 1: $y \in \{0, 1\}$.

Модель бинарного выбора

Зависимая переменная принимает значения 0 или 1: $y \in \{0, 1\}$.

$$y_t = x_t^\top \beta + \varepsilon_t, \quad t = 1, \ldots, n$$

Модель бинарного выбора

Зависимая переменная принимает значения 0 или 1: $y \in \{0, 1\}$.

$$y_t = x_t^\top \beta + \varepsilon_t, \quad t = 1, \dots, n$$

$$\mathsf{E}\, y_t = 1 \cdot \mathsf{P}(y_t = 1) + 0 \cdot \mathsf{P}(y_t = 0) = \mathsf{P}(y_t = 1) = x_t^\top \beta$$

Модель бинарного выбора

Зависимая переменная принимает значения 0 или 1: $y \in \{0, 1\}$.

Линейная модель вероятности:

$$\mathsf{P}(y_t = 1) = x_t^\top \beta$$

Модель бинарного выбора

Probit- и logit-модели

$$\mathsf{P}(y_t = 1) = F(x_t^\top \beta)$$

$F(\cdot)$ — функция, значения которой лежат в отрезке $[0, 1]$

Модель бинарного выбора

Probit- и logit-модели

$$y_t^* = x_t^\top \beta + \varepsilon_t$$

$$y_t = \begin{cases} 1, & y_t^* \geqslant 0, \\ 0, & y_t^* < 0. \end{cases}$$

Модель бинарного выбора

Probit- и logit-модели

$$\mathsf{P}(y_t = 1) = \mathsf{P}(y_t^* \geqslant 0) = \mathsf{P}(x_t^\top \beta + \varepsilon_t \geqslant 0) =$$

$$= \mathsf{P}(\varepsilon_t \geqslant -x_t^\top \beta) = \mathsf{P}(\varepsilon_t \leqslant x_t^\top \beta) = F\left(\frac{x_t^\top \beta}{\sigma}\right)$$

Модель бинарного выбора

Probit- и logit-модели

$$\mathsf{P}(y_t = 1) = F\left(\frac{x_t^\top \beta}{\sigma}\right)$$

1. Probit-модель:

$$F(u) = \Phi(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u} e^{-\frac{z^2}{2}}\, dz$$

2. Logit-модель:

$$F(u) = \Lambda(u) = \frac{e^u}{1 + e^u}$$

Модель бинарного выбора

Оценивание модели

Функция правдоподобия:

$$L = L(y_1, \ldots, y_n) = \prod_{y_t=0} \left(1 - F(x_t^\top \beta)\right) \prod_{y_t=1} F(x_t^\top \beta)$$

Модель бинарного выбора

Оценивание модели

Функция правдоподобия:

$$L = \prod_{t=1}^{n} \big(1 - F(x_t^\top \beta)\big)^{1-y_t} F(x_t^\top \beta)^{y_t}$$

Логарифмическая функция правдоподобия:

$$\ln L = \sum_{t=1}^{n} \Big((1 - y_t) \ln\big(1 - F(x_t^\top \beta)\big) + y_t \ln F(x_t^\top \beta) \Big)$$

Модель бинарного выбора

Оценивание модели

$$\frac{\partial \ln L}{\partial \beta} = \sum_{t=1}^{n} \left(-\frac{(1 - y_t) p(x_t^\top \beta)}{(1 - F(x_t^\top \beta))} + \frac{y_t p(x_t^\top \beta)}{F(x_t^\top \beta)} \right) x_t = 0$$

$p(z) = F'(z)$ — плотность вероятности.

Модель бинарного выбора

Оценивание модели

Для logit-модели:

$$\frac{\partial \ln L}{\partial \beta} = \sum_{t=1}^{n} \big(y_t - \Lambda(x_t^\top \beta)\big) x_t = 0,$$

т.к. $\Lambda'(u) = \Lambda(u)\big(1 - \Lambda(u)\big)$.

Оценивание модели бинарного выбора в R:

1. probit-модель:

   glm(модель,family=binomial(link='probit'))

2. logit-модель:

   glm(модель,family=binomial(link='logit'))

Модель множественного выбора

Количество альтернатив больше двух

1. Номинальные зависимые переменные
2. Порядковые зависимые переменные

Модель множественного выбора

Номинальные зависимые переменные

m альтернатив

1. Дерево решений
2. Использование понятия случайной полезности

Модель множественного выбора

Номинальные зависимые переменные

Случайная полезность

Для индивидуума t альтернатива j имеет полезность $U_{t,j}$:

$$U_{t,j} = u_{t,j} + \varepsilon_{t,j}$$

Индивидуум t выберет альтернативу j, если $U_{t,j} > U_{t,k}$ для всех $k \neq j$:

$$P(y_t = j) = P(u_{t,j} + \varepsilon_{t,j} > u_{t,k} + \varepsilon_{t,k},\ \forall k \neq j, k = 1, \ldots, m)$$

Случайная полезность

Если ошибки $\varepsilon_{t,j}$ независимы и имеют распределение $F(x) = \exp\left(-e^{-x}\right)$, то

$$P(y_t = j) = \frac{\exp(u_{t,j})}{\exp(u_{t,1}) + \cdots + \exp(u_{t,m})}$$

Модель множественного выбора

Номинальные зависимые переменные

logit-модель множественного выбора

Если

$$u_{t,j} = x_{t,j}\beta,$$

то

$$P(y_t = j) = \frac{\exp(x_{t,j}\beta)}{\exp(x_{t,1}\beta) + \cdots + \exp(x_{t,m}\beta)}$$

Модель множественного выбора

Номинальные зависимые переменные

logit-модель множественного выбора

Если

$$u_{t,j} = x_{t,j}\beta,$$

то

$$P(y_t = j) = \frac{\exp(x_{t,j}\beta)}{\exp(x_{t,1}\beta) + \cdots + \exp(x_{t,m}\beta)}$$

Модель множественного выбора

Номинальные зависимые переменные

logit-модель множественного выбора

Если

$$u_{t,j} = x_{t,j}\beta_j,$$

то

$$P(y_t = j) = \frac{\exp(x_{t,j}\beta_j)}{\exp(x_{t,1}\beta_1) + \cdots + \exp(x_{t,m}\beta_m)}$$

Модель не идентифицируема.

Модель множественного выбора

Номинальные зависимые переменные

logit-модель множественного выбора

Если

$$u_{t,j} = x_{t,j}\beta_j,$$

то (с использованием нормировки $\beta_1 = 0$)

$$P(y_t = 1) = \frac{1}{1 + \exp(x_{t,2}\beta_1) + \dots + \exp(x_{t,m}\beta_m)}$$

$$P(y_t = j) = \frac{\exp(x_{t,j}\beta_j)}{1 + \exp(x_{t,1}\beta_2) + \dots + \exp(x_{t,m}\beta_m)}, \quad j \geqslant 2$$

Модель множественного выбора

Номинальные зависимые переменные

Ограничение logit-модели множественного выбора: предположение о статистической независимости полезностей $u_{t,j}$ по j
(независимость от посторонних альтернатив)

Модель множественного выбора

Порядковые зависимые переменные

Альтернативы: $a_1 \prec a_2 \prec \dots \prec a_m$

Модель множественного выбора

Порядковые зависимые переменные

Альтернативы: $a_1 \prec a_2 \prec \cdots \prec a_m$

$$y = \begin{cases} a_1, & y^* \leqslant c_1 \\ a_2, & c_1 < y^* \leqslant c_2 \\ a_3, & c_2 < y^* \leqslant c_3 \\ \ldots & \\ a_m, & y^* > c_{m-1} \end{cases}$$

y^* — скрытая переменная
c_j — фиксированные уровни

Модель множественного выбора

Порядковые зависимые переменные

Альтернативы: $a_1 \prec a_2 \prec \cdots \prec a_m$

$$
\begin{aligned}
y_t^* &= x_t\beta \\
P(y_t = a_1) &= F(c_1 - x_t\beta) \\
P(y_t = a_2) &= F(c_2 - x_t\beta) - F(c_1 - x_t\beta) \\
&\cdots \\
P(y_t = a_m) &= 1 - F(c_{m-1} - x_t\beta)
\end{aligned}
$$

Модель множественного выбора

Порядковые зависимые переменные

Альтернативы: $a_1 \prec a_2 \prec \cdots \prec a_m$
Функция правдоподобия:

$$L = F(c_1 - x_t\beta)\times \times \prod_{j=2,\dots,m-1} \big(F(c_j - x_t\beta) - F(c_{j-1} - x_t\beta)\big)\times \times \big(1 - F(c_{m-1} - x_t\beta)\big)$$