
1. Методы получения оценок

Методы получения оценок: метод максимального правдоподобия, метод моментов, метод наименьших квадратов.

2. Свойства оценок

Свойства оценок: несмещённость, состоятельность, эффективность в классе.

3. Асимптотические методы

Центральная предельная теорема. Лемма Слуцкого. Дельта-метод. Построение асимптотических доверительных интервалов.

4. Святая троица тестов

Три классических теста: LM, LR, Wald.

Чёрный трек: тесты в матричной форме для вектора параметров?

5. ШБ — МНК

МНК в скалярной и матричной форме без статистических свойств. Строгая мультиколлинеарность. Рассказать, что коэффициенты при стандартизации всех переменных называют частными корреляциями.

Коммент: Здесь первый раз говорим слова "строгая мультиколлинеарность".

Чёрный трек: нелинейный мнк численно?

Задачи для доски:

МНК и R^2 руками на доске

Задачи для колаба:

МНК и R^2

Рост R^2 с ростом числа регрессоров

Рост RSS с ростом числа наблюдений

6. Предпосылки о математическом ожидании и дисперсии

МНК со статистическими предпосылками на ожидание и дисперсию. Теорема Гаусса-Маркова.

6.1. Ожидание и ковариационная матрица

Пусть r — случайный вектор размерности $n \times 1$, s — случайный вектор размерности $k \times 1$. A и b — неслучайные матрица и вектор соответственно, имеющие подходящие размерности.

Математическим ожиданием случайного вектора r называется вектор

$$\mathbb{E}(r) = \begin{pmatrix} \mathbb{E}(r_1) \\ \mathbb{E}(r_2) \\ \dots \\ \mathbb{E}(r_n) \end{pmatrix}.$$

Ковариационная матрица вектора r определяется следующим образом:

$$\mathbb{V}\text{ar}(r) = \begin{pmatrix} \text{Cov}(r_1, r_1) & \text{Cov}(r_1, r_2) & \dots & \text{Cov}(r_1, r_n) \\ \text{Cov}(r_2, r_1) & \text{Cov}(r_2, r_2) & \dots & \text{Cov}(r_2, r_n) \\ \dots & \dots & \dots & \dots \\ \text{Cov}(r_n, r_1) & \text{Cov}(r_n, r_2) & \dots & \text{Cov}(r_n, r_n) \end{pmatrix}.$$

Ковариационная матрица векторов r и s определяется следующим образом:

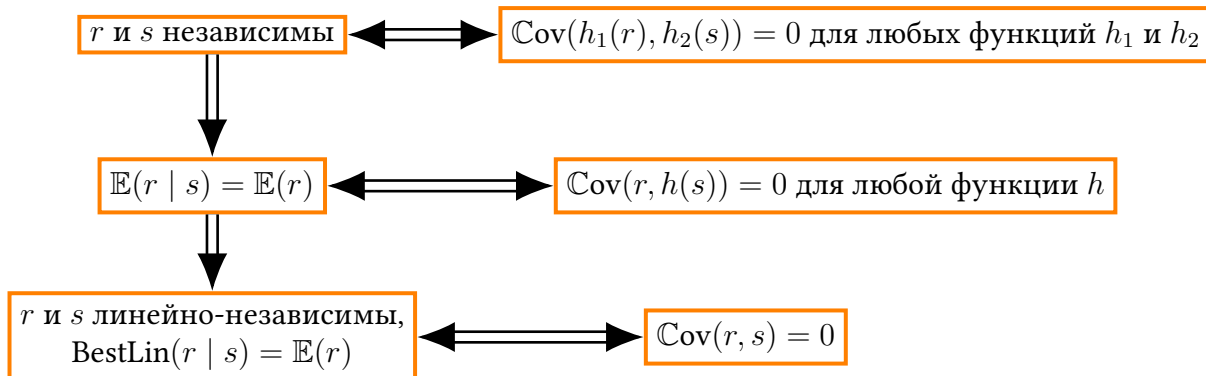
$$\text{Cov}(r, s) = \begin{pmatrix} \text{Cov}(r_1, s_1) & \text{Cov}(r_1, s_2) & \dots & \text{Cov}(r_1, s_k) \\ \text{Cov}(r_2, s_1) & \text{Cov}(r_2, s_2) & \dots & \text{Cov}(r_2, s_k) \\ \dots & \dots & \dots & \dots \\ \text{Cov}(r_n, s_1) & \text{Cov}(r_n, s_2) & \dots & \text{Cov}(r_n, s_k) \end{pmatrix}.$$

Далее рассмотрим свойства для вектора математических ожиданий и ковариационной матрицы:

- а) $\mathbb{E}(Ar + b) = A \mathbb{E}(r) + b$
- б) $\text{Cov}(r, s) = \mathbb{E}(rs^T) - \mathbb{E}(r) \mathbb{E}(s^T)$
- в) $\text{Cov}(Ar + b, s) = A \text{Cov}(r, s)$
- г) $\text{Cov}(r, As + b) = \text{Cov}(r, s) A^T$
- д) $\mathbb{V}\text{ar}(r) = \text{Cov}(r, r) = \mathbb{E}(rr^T) - \mathbb{E}(r) \mathbb{E}(r^T)$
- е) $\mathbb{V}\text{ar}(Ar + b) = A \mathbb{V}\text{ar}(r) A^T$
- ж) $\mathbb{E}(r^T Ar) = \text{trace}(A \mathbb{V}\text{ar}(r)) + \mathbb{E}(r^T) A \mathbb{E}(r)$

6.2. Иерархия зависимостей случайных величин

Можно выделить три степени независимости случайных величин. Рассмотрим их на примере пары произвольных величин r и s .



Напомним, что случайные величины r и s называются независимыми, если для любых числовых множеств A и B независимы события $\{r \in A\}$ и $\{s \in B\}$:

$$\mathbb{P}(r \in A, s \in B) = \mathbb{P}(r \in A) \cdot \mathbb{P}(s \in B)$$

Определить линейную независимость величин r и s можно по-разному. Некоторые авторы считают условие $\text{Cov}(r, s) = 0$ определением линейной независимости, в этом случае нижняя эквивалентность на графике тривиальна. Нам кажется более логичным другой подход. Величины r и s линейно независимы, если наилучшее линейное приближение r с помощью s не зависит от s .

Задача 6.1. Покажем, что из равенства условного и безусловного математических ожиданий не следует независимость случайных величин:

	1/3	1/3	1/3
r	-1	1	0
s	0	0	1

Определение 6.2. Наилучшее линейное приближение величины r с помощью величины s — это линейная функция от s ,

$$\text{BestLin}(r \mid s) = \alpha + \beta s,$$

где константы α и β находятся из решения задачи оптимизации $\mathbb{E}((r - \text{BestLin}(r, s))^2) \rightarrow \min_{\alpha, \beta}$.

6.3. Теорема Гаусса — Маркова

Чтобы исследовать свойства полученной точечной оценки $\hat{\beta}$ нам потребуются предпосылки о математическом ожидании и ковариационной матрице вектора u .

Мы предположим, что случайные ошибки в среднем равны нулю, а именно,

$$\mathbb{E}(u \mid X) = 0.$$

Предпосылку о математическом ожидании можно записать и в скалярном виде,

$$\mathbb{E}(u_i \mid X) = 0, \quad \text{при } \forall i \in \{1, \dots, n\}.$$

Теорема 6.3. Если

- а) Модель линейна по параметрам: $y = X\beta + u$;
- б) Матрица X размера $[n \times k]$ имеет полный ранг k .
- в) Условное ожидание ошибок равно нулю, $\mathbb{E}(u \mid X) = 0$;
- г) Условная ковариационная матрица ошибок пропорциональна единичной, $\text{Var}(u \mid X) = \sigma^2 I$;
- д) Оценка $\hat{\beta}$ получена методом наименьших квадратов, $\hat{\beta} = (X^T X)^{-1} X^T y$;
- е) Альтернативная оценка $\hat{\beta}^{\text{alt}}$ является условно несмещёнными $\mathbb{E}(\hat{\beta}^{\text{alt}} \mid X) = \beta$ и линейной по y ;

то

- а) Оценка $\hat{\beta}$ является линейной по y ;
- б) Оценка $\hat{\beta}$ является условно несмещённой, $\mathbb{E}(\hat{\beta} \mid X) = \beta$ и несмещённой, $\mathbb{E}(\hat{\beta}) = \beta$;

в) Оценка любого коэффициента $\hat{\beta}_j$ является более эффективной, чем альтернативная оценка $\hat{\beta}_j^{\text{alt}}$:

$$\text{Var}(\hat{\beta}_j | X) \leq \text{Var}(\hat{\beta}_j^{\text{alt}} | X).$$

Вывод теоремы можно усилить, для любой линейной комбинации коэффициентов $w^T \beta$ МНК-оценка $w^T \hat{\beta}$ эффективнее альтернативной оценки $w^T \hat{\beta}^{\text{alt}}$:

$$\text{Var}(w^T \hat{\beta}_j | X) \leq \text{Var}(w^T \hat{\beta}_j^{\text{alt}} | X).$$

6.4. Задачи для доски:

Задача 6.4. Исследовательница Мишель собрала данные по 20 студентам. Переменная y_i — количество решённых задач по эконометрике i -ым студентом, а x_i — количество просмотренных серий любимого сериала за прошедший год. Оказалось, что $\sum y_i = 10$, $\sum x_i = 0$, $\sum x_i^2 = 40$, $\sum y_i^2 = 50$, $\sum x_i y_i = 60$.

- а) Найдите МНК-оценки коэффициентов парной регрессии.
- б) В рамках предположения $\mathbb{E}(u_i | X) = 0$ найдите $\mathbb{E}(y_i | X)$, $\mathbb{E}(\hat{\beta}_j | X)$, $\mathbb{E}(\hat{u}_i | X)$, $\mathbb{E}(\hat{y}_i | X)$.
- в) Предположим дополнительно, что $\text{Var}(u_i | X) = \sigma^2$ и u_i при фиксированных X независимы. Найдите $\text{Var}(y_i | X)$, $\text{Var}(y_i(x_i - \bar{x}) | X)$, $\text{Var}(\sum y_i(x_i - \bar{x}) | X)$, $\text{Var}(\hat{\beta}_2 | X)$.

Задача 6.5. Рассмотрим классическую линейную модель $y = X\beta + u$ с предпосылками Гаусса — Маркова: $\mathbb{E}(u | X) = 0$ и $\text{Var}(u | X) = \sigma^2 I$. Для всех случайных векторов $(y, \hat{y}, \hat{\beta}, u, \hat{u}, \bar{y})$ найдите все возможные ожидания и ковариационные матрицы $\mathbb{E}(\cdot)$, $\text{Var}(\cdot)$, $\text{Cov}(\cdot, \cdot)$.

Задача 6.6. Рассмотрим модель $y_i = \beta x_i + u_i$ с двумя наблюдениями, $x_1 = 1$, $x_2 = 2$. Величины u_1 и u_2 независимы и равновероятно равны $+1$ или -1 .

- а) Найдите оценку $\hat{\beta}_{\text{ols}}$ для β с помощью метода наименьших квадратов.
- б) Чему равна дисперсия $\text{Var}(\hat{\beta}_{\text{ols}} | x)$ и ожидание $\mathbb{E}(\hat{\beta}_{\text{ols}} | x)$?
- в) Постройте несмещённую оценку $\hat{\beta}_{\text{best}}$ с наименьшей дисперсией.
- г) Чему равна дисперсия $\text{Var}(\hat{\beta}_{\text{best}} | x)$?
- д) А как же теорема Гаусса — Маркова? Почему в данном примере удаётся построить оценку с дисперсией меньше, чем у оценки методом наименьших квадратов?
- а) $\hat{\beta}_{\text{ols}} = (y_1 + 2y_2)/5$;
- б) $\text{Var}(\hat{\beta}_{\text{ols}} | x) = 1/5$;
- в) Заметим, что по величине $2y_1 - y_2$ можно однозначно восстановить величины ошибок u_1 и u_2 . Например, если $2y_1 - y_2 = 3$, то $u_1 = 1$, $u_2 = -1$.

$$\hat{\beta}_{\text{best}} = \begin{cases} y_1 + 1, & \text{если } 2y_1 - y_2 < 0, \\ y_1 - 1, & \text{если } 2y_1 - y_2 > 0. \end{cases}$$

- г) Шок контент, $\text{Var}(\hat{\beta}_{\text{best}} | x) = 0$.
-

- д) Построенная оценка $\hat{\beta}_{\text{best}}$ является нелинейной по y , а теорема Гаусса — Маркова гарантирует только, что метод наименьших квадратов порождает несмещённую оценку с наименьшей дисперсией среди линейных по y оценок.

Задача 6.7. (Hansen 4.14)

Задана модель $y = X\beta + u$, для которой выполняются предпосылки теоремы Гаусса — Маркова. Вас интересует величина $\theta = \beta^2$. Пусть получены МНК-оценки коэффициентов: $\hat{\beta}$, $V_{\hat{\beta}} = \text{Var}[\hat{\beta}|X]$. Тогда кажется неплохой идеей оценить θ как $\hat{\theta} = \hat{\beta}^2$.

- а) Найдите $\mathbb{E}[\hat{\theta}|X]$. Является ли $\hat{\theta}$ смещённой?
- б) Предложите способ коррекции смещения для получения несмещённой оценки $\hat{\theta}^*$, используя результаты предыдущего пункта.

Задача 6.8. Рассмотрим модель регрессии $y_i = \beta_1 + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + u_i$. Пусть все предпосылки теоремы Гаусса — Маркова выполнены. Дополнительно предположим, что $u_i \sim N(0, \sigma^2)$, $i = 1, \dots, n$. Дополнительно известно, что на самом деле $\beta_2 = \dots = \beta_k = 0$.

- а) Найдите $\mathbb{E}(R^2)$.
- б) Найдите $\mathbb{E}(R_{\text{adj}}^2)$.

6.5. Задачи для колаб:

Генерация R2 для вывода распределения

Генерация смещения

Генерация лишних регрессоров

Реальный пример с лишним регрессорами (тип знаки зодиака и ретроградный)

Какая-то длинная задача, которую из темы в тему и в ней находить потом нарушения предпосылок?

<https://colab.research.google.com/drive/1wFrLyGcVVEtx96jS93I4z8asgAQwqldw?usp=sharing>

6.6. Чёрный трэк:

Умножение блочных матриц. Если размеры блоков допускают операцию умножения, то:

$$\left[\begin{array}{c|c} A & B \\ \hline C & D \end{array} \right] \cdot \left[\begin{array}{c|c} E & F \\ \hline G & H \end{array} \right] = \left[\begin{array}{c|c} AE + BG & AF + BH \\ \hline CE + DG & CF + DH \end{array} \right].$$

Формула Фробениуса (блочное обращение).

$$\left[\begin{array}{c|c} A & B \\ \hline C & D \end{array} \right]^{-1} = \left[\begin{array}{c|c} A^{-1} + A^{-1}BH^{-1}CA^{-1} & -A^{-1}BH^{-1} \\ \hline -H^{-1}CA^{-1} & H^{-1} \end{array} \right],$$

где A — невырожденная квадратная матрица размерности $n \times n$, D — квадратная матрица размерности $k \times k$, $H = D - CA^{-1}B$.

Задача 6.9. Пусть истинной является модель $y = X_1\beta_1 + X_2\beta_2 + u$, где X_1 , X_2 — матрицы признаков размерностей $n \times k_1$ и $n \times k_2$ соответственно. Вместо истинной модели вы оцениваете модель вида $y = X_1\beta_1 + v$, где v — вектор случайной ошибки, удовлетворяющий предпосылкам теоремы Гаусса — Маркова.

-
- а) Будет ли МНК-оценка вектора параметров β_1 несмещённой?
 - б) Будет ли несмещённой МНК-оценка дисперсии случайной ошибки?
 - в) Рассчитайте $\text{Var}(\hat{\beta}_1)$. Не противоречит ли полученный результат теореме Гаусса — Маркова?

Задача 6.10. Пусть истинной является модель $y = X_1\beta_1 + u$, где X_1 — матрица признаков размерности $n \times k_1$. Вместо истинной модели вы оцениваете модель вида $y = X_1\beta_1 + X_2\beta_2 + v$, где X_2 — матрица признаков размерности $n \times k_2$, v — вектор случайной ошибки, удовлетворяющий предпосылкам теоремы Гаусса — Маркова.

- а) Будет ли МНК-оценка вектора параметров β_1 несмещённой?
- б) Будет ли несмещённой МНК-оценка дисперсии случайной ошибки?
- в) Рассчитайте $\text{Var}(\hat{\beta}_1)$. Не противоречит ли полученный результат теореме Гаусса — Маркова?

7. Доверительные интервалы для коэффициентов

Построение доверительных интервалов для МНК оценок. Проверка гипотез. Асимптотика без нормальности ошибок. Нормальность ошибок.

8. Бутстрэп

Бутстрэп. Классический бутстрэп до регрессии и бутстрэп в регрессии. Метод наименьших модулей. Чёрный трек: возможно, разные варианты бутстрэпа в регрессии? ВСА-бутстрэп до регрессии?

9. Выбор функциональной формы

Дамми-переменные и их интерпретация. Функциональные формы: полиномы, логарифмы, интерпретация коэффициентов. Информационные критерии.

Чёрный трек: Структурные сдвиги. Тест Чоу. Локально-линейная регрессия (LOESS).

10. Гетероскедастичность

Гетероскедастичность. Тестирование гетероскедастичности. Робастные оценки. Доступный обобщённый МНК.

Задачи для доски:

Хансен: во сколько раз может быть недооценена дисперсия из-за гетероскедастичности

Коммент: акцент на робастных ошибках, тестирование и обобщённый МНК — кратко.

11. Мультиколлинеарность и метод главных компонент

Мультиколлинеарность и метод главных компонент.

Чёрный трек: несколько взглядов на метод главных компонент? LASSO?

12. Эндогенность

Эндогенность. Инструментальные переменные. Ошибка измерения регрессора. Двухшаговый МНК.

13. Эффекты воздействия

Оценка эффектов воздействия. ATE. LATE. Четкий (sharp) и нечеткий (fuzzy) разрывный регрессионный дизайн (RDD).

Чёрный трек: Метод разность разностей (DiD). Динамический метода разность разностей (Event Study).

14. Логистическая регрессия: точечные оценки

Логистическая регрессия: Бинарный и упорядоченный логит. Точечные оценки, прогнозы. Интерпретация предельных эффектов.

Чёрный трек: Множественные логиты. Неупорядоченные, условные, смешанные логиты.

15. Логистическая регрессия: доверительные интервал

Логистическая регрессия: доверительные интервалы и проверка гипотез.

Чёрный трек: разные хоббиты

15.1. Смещение, цензурирование и ■■■■■■

Представим себе ситуацию, в которой зависимая количественная не всегда наблюдаема. Для моделирования этой ситуации мы введём скрытую латентная переменная y_i^* , которая линейно зависит от предиктора x_i , как обычно,

$$y_i^* = x_i^T \beta + u_i, \quad y^* = X^T \beta + u$$

Бинарная переменная $z_i \in \{0, 1\}$ равна 1 в случае, если мы наблюдаем y_i^* .

Возможно несколько случаев:

	наблюдаемость y^*	наблюдаемость x	наблюдаемость
Цензурирование censored model	зависит от y^*	всегда	
Усечение truncated model	зависит от y^*	если наблюдаем y^*	
Выборочное смещение sample selection	зависит от w	всегда	всегда
Переключающиеся режимы switching regimes	всегда, w переключает тип зависимости	всегда	всегда

Представим себе, что мы открыли дорогой ресторан. К нам заглядывают клиенты. Часть клиентов ужасаются от ценника и убегают, $y_i^* < 0$. Часть клиентов остаются и ужинают у нас, $y_i^* > 0$. Вместо нуля можно выбрать другой порог, но с нулём чуть-чуть удобнее.

15.2. Цензурирование

Рассмотрим самый распространённый вариант цензурирования: вместо отрицательных значений латентной переменной y_i^* мы видим нули.

Эта модель известна как тобит модель типа I, type I Tobit model.

$$\begin{cases} y_i^* = x_i^T \beta + u_i, & y^* = X^T \beta + u \\ (u \mid X) \sim \mathcal{N}(0, \sigma^2 I) \\ y_i = \max\{y_i^*, 0\} \\ (x_i, y_i) \text{ наблюдаемы при любых } i \end{cases}$$

Лог-функция правдоподобия равна

$$\ell(\beta, \sigma) = \sum_{y_i=0} \ln F(-x_i^T \beta / \sigma) + \sum_{y_i>0} \ln f((y_i - x_i^T \beta) / \sigma) - \sum_{y_i>0} \ln \sigma$$

15.3. Усечение

$$\begin{cases} y_i^* = x_i^T \beta + u_i, & y^* = X^T \beta + u \\ (u \mid X) \sim \mathcal{N}(0, \sigma^2 I) \\ y_i = \max\{y_i^*, 0\} \\ (x_i, y_i) \text{ наблюдаемы, если } y_i > 0 \end{cases}$$

Лог-функция правдоподобия равна

$$\ell(\beta, \sigma) = \sum_{y_i>0} \ln f((y_i - x_i^T \beta) / \sigma) - \sum_{y_i>0} \ln F(x_i^T \beta / \sigma) - \sum_{y_i>0} \ln \sigma$$

15.4. Три осмысленных условных ожидания

Ожидание латентной переменной показывает, сколько в среднем планирует потратить гость ресторана на ужин, ещё не видевший цен, полезность от ужина,

$$m^*(x_i) = \mathbb{E}(y_i^* \mid x_i) = x_i^T \beta$$

Предельный эффект для латентной переменной

$$\partial \mathbb{E}(y_i^* \mid x_{ij}) / \partial x_{ij} = \beta_j$$

Ожидание цензурированной переменной, $y_i = \max\{y_i^*, 0\}$, сколько в среднем потратит человек, заглянувший в ресторан, с учётом того, что часть уйдёт испугавшись ценника

$$m(x_i) = \mathbb{E}(y_i \mid x_i) = x_i^T \beta F(x_i^T \beta / \sigma) + \sigma f(x_i^T \beta / \sigma)$$

Предельный эффект для цензурированной переменной

$$\partial \mathbb{E}(y_i \mid x_{ij}) / \partial x_{ij} =$$

Условное ожидание усечённой переменной, $(y_i \mid y_i^* > 0)$, средний чек в ресторане

$$m^\#(x_i) = \mathbb{E}(y_i \mid x_i, y_i^* > 0) = x_i^T \beta + \sigma \lambda(x_i^T \beta / \sigma),$$

где $\lambda(s)$ — обратное отношение Миллса, inverse Mills ratio,

$$\lambda(s) = \mathbb{E}(v \mid v + s > 0) = f(s) / F(s), \quad v \sim \mathcal{N}(0; 1)$$

Предельный эффект для ожидания усечённой переменной

Выборочное смещение
Переключающиеся режимы