Содержание

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 трэк: нелинейный мнк численно?

Задачи для доски:

МНК и R2 руками на доске

Задачи для колаба:

МНК и R2

Рост R2 с ростом числа регрессоров

Рост RSS с ростом числа наблюдений

5.1. Чёрный трэк

Определение 5.1. Кросс-валидация с поочередным выкидыванием отдельных наблюдений. Leave one out cross validation.

Рассмотрим модель $y = X\beta + u$.

Оценим модель без первого наблюдения. Получим МНК-оценки $\hat{\beta}^{(-1)}$. С помощью этих оценок спрогнозируем первое наблюдение, получим прогноз $\hat{y}_1^{CV}$ и ошибку прогноза $\hat{u}_1^{CV}$.

Вернём первое наблюдение в выборку и удалим второе наблюдение. Получим МНК-оценки $\hat{\beta}^{(-2)}$. С помощью этих оценок спрогнозируем второе наблюдение, получим прогноз $\hat{y}_2^{CV}$ и ошибку прогноза $\hat{u}_2^{CV}$.

Поступим так с каждым наблюдением. На выходе получим вектор кросс-валидационных прогнозов $\hat{y}^{CV}$ и вектор кросс-валидационных ошибок прогнозов $\hat{u}^{CV}$.

Теорема 5.2. Если модель $y = X\beta + u$ оценивается с помощью МНК и проводится кросс-валидации с поочередным выкидыванием отдельных наблюдений, то:

$$\hat{u}_i = (1 - H_{ii}) \cdot \hat{u}_i^{CV},$$

где H — матрица-шляпница $H = X(X^TX)^{-1}X^T$, $\hat{u}$ — остатки регрессии, а $\hat{u}^{CV}$ — кросс-валидационные ошибки прогнозов.

Заметим, что сомножитель $(1 - H_{ii}) \in (0; 1)$. Другими словами, теорема численно формализует интуитивно ожидаемый результат: кросс-валидационные остатки по знаку совпадают с обычными остатками, а по абсолютной величине — больше, так как соответствующее наблюдение не используется при оценивании коэффициента.

Доказательство. Оценим модель без последнего наблюдения, $\hat{y}^- = X^-\hat{\beta}^-$.

Создадим вектор y^*, который будет отличаться от y только последним, n-м элементом: вместо настоящего y_n там будет стоять прогноз по модели без последнего наблюдения $\hat{y}_n^-$.

Раз уж мы добавили новую точку лежащую ровно на выборочной регрессии, то при оценки модели $\hat{y}^* = X\hat{\beta}^*$ мы получим в точности старые оценки $\hat{\beta}^* = \hat{\beta}^-$. Следовательно, и прогнозы эти две модели дают одинаковые, $\hat{y}_i^* = \hat{y}_i^-$.

А теперь посмотрим на последний элемент вектора $v = H(y^* - y)$.

С одной стороны, он равен последней строке матрицы H умножить на вектор $(y^* - y)$. В векторе $(y^* - y)$ только последний элемент ненулевой, поэтому $v_n = H_{nn}(\hat{y}_n^- - y_n)$.

С другой стороны, мы можем раскрыть скобки, и заметить, что $v = Hy^* - Hy$. И окажется, что $v_n = \hat{y}_n^* - \hat{y}_n = \hat{y}_n^- - \hat{y}_n$.

Отсюда

$$\hat{y}_n^- - \hat{y}_n = H_{nn}(\hat{y}_n^- - y_n)$$

Приводим подобные слагаемые и добавляем слева и справа y_n, получаем как раз то, что нужно:

$$y_n - \hat{y}_n = (1 - H_{ii})(y_n - \hat{y}_n^-)$$

□

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:

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

Ковариационная матрица векторов r и s определяется следующим образом:

$$\mathbb{C}\mathrm{ov}(r,s)=\begin{pmatrix}\mathbb{C}\mathrm{ov}(r_1,s_1) & \mathbb{C}\mathrm{ov}(r_1,s_2) & \dots & \mathbb{C}\mathrm{ov}(r_1,s_k)\\ \mathbb{C}\mathrm{ov}(r_2,s_1) & \mathbb{C}\mathrm{ov}(r_2,s_2) & \dots & \mathbb{C}\mathrm{ov}(r_2,s_k)\\ \dots & \dots & \dots & \dots\\ \mathbb{C}\mathrm{ov}(r_n,s_1) & \mathbb{C}\mathrm{ov}(r_n,s_2) & \dots & \mathbb{C}\mathrm{ov}(r_n,s_k)\end{pmatrix}.$$

Свойства вектора математических ожиданий и ковариационной матрицы:

а) $\mathbb{E}(Ar+b)=A\,\mathbb{E}(r)+b$

б) $\mathbb{C}\mathrm{ov}(r,s)=\mathbb{E}(rs^T)-\mathbb{E}(r)\,\mathbb{E}(s^T)$

в) $\mathbb{C}\mathrm{ov}(Ar+b,s)=A\,\mathbb{C}\mathrm{ov}(r,s)$

г) $\mathbb{C}\mathrm{ov}(r,As+b)=\mathbb{C}\mathrm{ov}(r,s)A^T$

д) $\mathbb{V}\mathrm{ar}(r)=\mathbb{C}\mathrm{ov}(r,r)=\mathbb{E}(rr^T)-\mathbb{E}(r)\,\mathbb{E}(r^T)$

е) $\mathbb{V}\mathrm{ar}(Ar+b)=A\,\mathbb{V}\mathrm{ar}(r)A^T$

ж) $\mathbb{E}(r^TAr)=\mathrm{trace}(A\,\mathbb{V}\mathrm{ar}(r))+\mathbb{E}(r^T)A\,\mathbb{E}(r)$

з) Если вектора r и s имеют одинаковый размер, то $\mathbb{V}\mathrm{ar}(r+s)=\mathbb{V}\mathrm{ar}(r)+\mathbb{V}\mathrm{ar}(s)+\mathbb{C}\mathrm{ov}(r,s)+\mathbb{C}\mathrm{ov}(s,r)$

6.2. Иерархия зависимостей случайных величин

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

$$\begin{tikzcd}
\boxed{r \text{ и } s \text{ независимы}} \arrow[r, Leftrightarrow] \arrow[d, Rightarrow] & \boxed{\mathbb{C}\mathrm{ov}(h_1(r),h_2(s))=0 \text{ для любых функций } h_1 \text{ и } h_2} \\
\boxed{\mathbb{E}(r\mid s)=\mathbb{E}(r)} \arrow[r, Leftrightarrow] \arrow[d, Rightarrow] & \boxed{\mathbb{C}\mathrm{ov}(r,h(s))=0 \text{ для любой функции } h} \\
\boxed{\begin{array}{c} r \text{ и } s \text{ линейно-независимы,} \\ \mathrm{BestLin}(r\mid s)=\mathbb{E}(r)\end{array}} \arrow[r, Leftrightarrow] & \boxed{\mathbb{C}\mathrm{ov}(r,s)=0}
\end{tikzcd}$$

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 $\mathbb{C}\mathrm{ov}(r,s)=0$ 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,

$$\operatorname{BestLin}(r \mid s) = \beta_1 + \beta_2 s,$$

где константы β_1 и β_2 находятся из решения задачи оптимизации $\mathbb{E}((r - \operatorname{BestLin}(r, s)^2) \to \min_{\beta_1, \beta_2}$.

Задача 6.3. Выразите константы β_1 и β_2 в формуле для наилучшего линейного приближения

$$\operatorname{BestLin}(r \mid s) = \beta_1 + \beta_2 s,$$

исходя из характеристик случайных величин r и s.

Выпишем целевую функцию в виде суммы

$$\mathbb{E}((r - \operatorname{BestLin}(r, s)^2) = \mathbb{V}\mathrm{ar}(r - \beta_1 - \beta_2 s) + (\mathbb{E}(r - \beta_1 - \beta_2 s))^2$$

Заметим, что β_1 не влияет на первое слагаемое, так как дисперсия константы равна нулю. И при этом, выбрав $\beta_1 = \mathbb{E}(r - \beta_2 s) = \mathbb{E}(r) - \beta_2 \mathbb{E}(s)$ мы добьёмся того, что второе слагаемое будет равно нулю, своему наименьшему возможному значению.

Остаётся минимизировать с помощью β_2 первое слагаемое.

$$\mathbb{V}\mathrm{ar}(r - \beta s) = \mathbb{V}\mathrm{ar}(r) + \beta_2^2 \mathbb{V}\mathrm{ar}(s) - 2\beta_2 \mathbb{C}\mathrm{ov}(r, s) \to \min_{\beta_2}.$$

Перед нами квадратичная функция от β_2, следовательно,

$$\beta_2 = \frac{\mathbb{C}\mathrm{ov}(r, s)}{\mathbb{V}\mathrm{ar}(s)}.$$

Обратите внимание, эта формула — родная «теоретическая» сестра «выборочной» формулы для парной регресии

$$\hat{\beta}_2 = \frac{S_{xy}}{S_{xx}}.$$

Аналогия между оценкой и истинным коэффициентом действует и для первого коэффициента,

$$\beta_1 = \mathbb{E}(r) - \beta_2 \mathbb{E}(s), \quad \hat{\beta}_1 = \bar{y} - \hat{\beta}_2 \bar{x}.$$

И, попутно, мы замечаем, что условие $\mathbb{C}\mathrm{ov}(r, s) = 0$ равносильно условию $\operatorname{BestLin}(r \mid s) = \mathbb{E}(r)$ или условию $\operatorname{BestLin}(s \mid r) = \mathbb{E}(s)$.

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,

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

Теорема 6.4. 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, $\mathbb{V}\mathrm{ar}(u \mid X) = \sigma^2 I$;

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 $\hat{\beta}_j^{\text{alt}}$:

$$\mathbb{V}\mathrm{ar}(\hat{\beta}_j \mid X) \leq \mathbb{V}\mathrm{ar}(\hat{\beta}_j^{\text{alt}} \mid X).$$

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

$$\mathbb{V}\mathrm{ar}(w^T \hat{\beta}_j \mid X) \leq \mathbb{V}\mathrm{ar}(w^T \hat{\beta}_j^{\text{alt}} \mid X).$$

Доказательство. Эффективность МНК-оценок — это реинкарнация теоремы Пифагора. Мы увидим, что дисперсия МНК-оценки — это квадрат длины катета, дисперсия альтернативной несмещённой оценки — квадрат длины гипотенузы.

Для примера рассмотрим оценку первого коэффициента бета, $\hat{\beta}_1$. Доказательство не меняется ни капли, если рассмотреть оценку другого коэффициента, скажем, $\hat{\beta}_7$ или даже оценку произвольной линейной комбинацию коэффициентов бета, например, $\hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_3$.

Итак, у нас есть две оценки, $\hat{\beta}_1$ и $\hat{\beta}_1^{\text{alt}}$. Обе они линейны по y, следовательно, $\hat{\beta}_1 = a^T y$ и $\hat{\beta}_1^{\text{alt}} = a_{\text{alt}}^T y$.

Замечаем, что $\mathbb{V}\mathrm{ar}(\hat{\beta}_1 \mid X) = \sigma^2 a^T a$, и $\mathbb{V}\mathrm{ar}(\hat{\beta}_1^{\text{alt}}) = \sigma^2 a_{\text{alt}}^T a_{\text{alt}}$. То есть дисперсии пропорциональны квадратам длин векторов a и a^{alt}. Осталось доказать, что вектор a не длиннее вектора a^{alt} :)

Для этого мы докажем, что вектор a^{alt} — это гипотенуза, а вектор a — катет. Нам нужно доказать, что вектор $a - a^{\text{alt}}$ перпендикулярен вектору a.

Разобъём доказательство перпендикулярности a и $a - a^{\text{alt}}$ на два шага:

Шаг 1. Вектор $a - a^{\text{alt}}$ перпендикулярен любому столбцу матрицы X.

Шаг 2. Вектор a является линейной комбинацией столбцов матрицы X.

TODO: здесь картинка!

Приступаем к шагу 1. Обе оценки несмещённые, поэтому для любых β должно выполняться:

$$\mathbb{E}(\hat{\beta}_1 \mid X) = \mathbb{E}(\hat{\beta}_1^{\text{alt}} \mid X)$$

Переносим всё в левую сторону:

$$\mathbb{E}((a^T - a_{\text{alt}}^T)(X\beta + u) \mid X) = 0$$

Получаем, что для любых β должно быть выполнено условие:

$$(a - a_{\text{alt}})^T X\beta = 0$$

Это возможно только, если вектор $(a - a_{\text{alt}})^T X$ равен нулю. Следовательно, вектор $(a - a_{\text{alt}})$ перпендикулярен любому столбцу X.

Приступаем к шагу 2.

Вспоминаем, что $\hat{\beta} = (X^T X)^{-1} X^T y$. Следовательно, нужная строка весов a^T — это первая строка в матрице $(X^T X)^{-1} X^T$. Замечаем, что выражение имеет вид $A \cdot X^T$.

Вспоминаем из линейной алгебры, что при умножении матриц AB получается матрица C, на которую можно взглянуть несколькими способами! Можно считать, что C — это разные линейные комбинации столбцов левой матрицы A. Можно считать, что C — это разные линейные комбинаций строк правой матрицы B.

Применим второй взгляд :) Получаем, что строка a^T — линейная комбинация строк матрицы X^T. Или, другими словами, столбец a — линейная комбинация столбцов матрицы X. □

Классическое доказательство, которое можно найти во многих учебниках, не замечает связи с теоремой Пифагора и исследует разницу ковариационных матриц.

Доказательство. У нас есть две линейных по y оценки:

$$\hat{\beta} = (X^T X)^{-1} X^T y \text{ и } \hat{\beta}_{\text{alt}} = A_{\text{alt}}^T y.$$

Оценки ковариационных матриц этих оценок равны

$$\mathbb{V}\text{ar}(\hat{\beta} \mid X) = (X^T X)^{-1}\sigma^2 \text{ и } \mathbb{V}\text{ar}(\hat{\beta}_{\text{alt}} \mid X) = A^T A\sigma^2.$$

□

6.4. Задачи для доски:

Задача 6.5. 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 $\mathbb{E}(u_i \mid X) = 0$ найдите $\mathbb{E}(y_i \mid X)$, $\mathbb{E}(\hat{\beta}_j \mid X)$, $\mathbb{E}(\hat{u}_i \mid X)$, $\mathbb{E}(\hat{y}_i \mid X)$.

в) Предположим дополнительно, что $\mathbb{V}\text{ar}(u_i \mid X) = \sigma^2$ и u_i при фиксированных X независимы. Найдите $\mathbb{V}\text{ar}(y_i \mid X)$, $\mathbb{V}\text{ar}(y_i(x_i - \bar{x}) \mid X)$, $\mathbb{V}\text{ar}(\sum y_i(x_i - \bar{x}) \mid X)$, $\mathbb{V}\text{ar}(\hat{\beta}_2 \mid X)$.

Задача 6.6. Рассмотрим классическую линейную модель $y = X\beta + u$ с предпосылками Гаусса — Маркова: $\mathbb{E}(u \mid X) = 0$ и $\mathbb{V}\text{ar}(u \mid X) = \sigma^2 I$. Для всех случайных векторов (y, $\hat{y}$, $\hat{\beta}$, u, $\hat{u}$, $\bar{y}$) найдите все возможные ожидания и ковариационные матрицы $\mathbb{E}(\cdot)$, $\mathbb{V}\text{ar}(\cdot)$, $\mathbb{C}\text{ov}(\cdot, \cdot)$.

Задача 6.7. 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 $\mathbb{V}\text{ar}(\hat{\beta}_{\text{ols}} \mid x)$ и ожидание $\mathbb{E}(\hat{\beta}_{\text{ols}} \mid x)$?

в) Постройте несмещённую оценку $\hat{\beta}_{\text{best}}$ с наименьшей дисперсией.

г) Чему равна дисперсия $\mathbb{V}\text{ar}(\hat{\beta}_{\text{best}} \mid x)$?

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) $\mathbb{V}\text{ar}(\hat{\beta}_{\text{ols}} \mid x) = 1/5$;

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, $\mathbb{V}\text{ar}(\hat{\beta}_{\text{best}} \mid x) = 0$.

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 6.8. (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: $\hat{\beta}$, $V_{\hat{\beta}} = \mathbb{V}\text{ar}[\hat{\beta}|X]$. 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 6.9. Рассмотрим модель регрессии $y_i = \beta_1 + \beta_2 x_{i2} + \ldots + \beta_k x_{ik} + u_i$. Пусть все предпосылки теоремы Гаусса — Маркова выполнены. Дополнительно предположим, что $u_i \sim N(0, \sigma^2), i = 1, \ldots, n.$. Дополнительно известно, что на самом деле $\beta_2 = \ldots = \beta_k = 0$.

а) Найдите $\mathbb{E}(R^2)$.

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

Задача 6.10. У овечки Долли был набор данных из n наблюдений для которого были выполнены предпосылки теоремы Гаусса — Маркова. Овечка Долли клонировала каждое наблюдение по одному разу и дописала каждое наблюдение-клон сразу после исходного наблюдения.

а) Как выглядит ковариационная матрица ошибок для нового набора данных?

б) Как изменится ответ на (а), если Долли клонирует только последнее наблюдение n раз?

а) Ковариационная матрица будет содержать блоки B на диагонали

$$\mathbb{V}\mathrm{ar}(u) = \begin{pmatrix} B & 0 & 0 & \ldots \\ 0 & B & 0 & \ldots \\ 0 & 0 & B & \ldots \\ \ldots & & & \end{pmatrix},$$

где каждый блок равен $B = \begin{pmatrix} \sigma^2 & \sigma^2 \\ \sigma^2 & \sigma^2 \end{pmatrix}$.

б) Ковариационная матрица будет состоять из четырех блоков: два блока нулевые, левые верхний блок пропорционален единичной матрицы, а все элементы правого нижнего блока равны σ^2:

$$\mathbb{V}\mathrm{ar}(u) = \begin{pmatrix} \sigma^2 \cdot I & 0 \\ 0 & S \end{pmatrix},$$

где I — единичная матрица, а все $S_{ij} = \sigma^2$.

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/1wFrLyGcVVETx96jS93I4z8asgAQwqIdw?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 6.11. 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 $\mathbb{V}\mathrm{ar}(\hat{\beta}_1)$. Не противоречит ли полученной результат теореме Гаусса — Маркова?

Задача 6.12. 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 $\mathbb{V}\mathrm{ar}(\hat{\beta}_1)$. Не противоречит ли полученной 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 трэк: возможно, разные варианты бутстрэпа в регрессии? BCA-бутстрэп до регрессии?

9. Выбор функциональной формы

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

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

15. Логистическая регрессия: доверительные интервал

Логистическая регрессия: доверительные интервалы и проверка гипотез.

Чёрный трэк: разные хоббиты

15.1. Смещение, цензурирование и [redacted]

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.

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

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

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

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