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.

2.1. Практическое напоминание об условном ожидании и дисперсии

Вспомним условное ожидание и условную дисперсию в дискретном случае:

Задача 1. Совместный закон распределения пары случайных величин (x, y) задан таблицей:

	$y = 1$	$y = 3$
$x = 1$	0.1	0.3
$x = 2$	0.1	0.1
$x = 4$	0.2	0.2

а) Найдите $\mathbb{E}(y \mid x)$, $\mathbb{V}\mathrm{ar}(y \mid x)$.

б) Найдите $\mathbb{E}(y)$, $\mathbb{E}(x)$, $\mathbb{C}\mathrm{ov}(x, y)$, $\mathbb{V}\mathrm{ar}(x)$.

в) Найдите наилучшее линейное приближение $\mathrm{BestLin}(y \mid x)$.

Здесь одарённый ассист напишет решение

Теперь вспомним, как считать условные характеристики случайных величин при наличии совместной плотности:

Задача 2. Пара случайных величин (x, y) имеет функцию плотности

$$f(x, y) = \begin{cases} (2x + 4y)/3, & \text{если } x \in [0, 1], y \in [0, 1], \\ 0, & \text{иначе.} \end{cases}$$

а) Найдите $\mathbb{E}(y \mid x)$, $\mathbb{V}\mathrm{ar}(y \mid x)$.

б) Найдите $\mathbb{E}(y)$, $\mathbb{E}(x)$, $\mathbb{C}\mathrm{ov}(x, y)$, $\mathbb{V}\mathrm{ar}(x)$.

в) Найдите наилучшее линейное приближение $\mathrm{BestLin}(y \mid x)$.

Здесь мудрый ассист напишет решение

Особо обратим внимание на случай двумерного нормального распределения:

Задача 3. Пара случайных величин (x, y) имеет совместное нормальное распределение

$$\begin{pmatrix} x \\ y \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 3 \\ 2 \end{pmatrix}, \begin{pmatrix} 10 & -2 \\ -2 & 20 \end{pmatrix}\right)$$

а) Найдите $\mathbb{E}(y \mid x)$, $\mathbb{V}\mathrm{ar}(y \mid x)$.

б) Найдите $\mathbb{E}(y)$, $\mathbb{E}(x)$, $\mathbb{C}\mathrm{ov}(x, y)$, $\mathbb{V}\mathrm{ar}(x)$.

в) Найдите наилучшее линейное приближение $\mathrm{BestLin}(y \mid x)$.

Здесь храбрый ассист напишет решение

Обратите внимание: для совместного нормального распределения условное ожидание $\mathbb{E}(y \mid x)$ и наилучшее линейное приближение $\mathrm{BestLin}(y \mid x)$ идеально совпадают. Условное ожидание и линейное приближение совпадут и в том случае, если величина x принимает всего два значения. Убедимся в этом с помощью простой задачи

Задача 4. На первом шаге Илон Маск случайным образом выбирает одно из двух значений случайной величины x, $\mathbb{P}(x = 1) = 0.4$, $\mathbb{P}(x = 2) = 0.6$. На втором шаге Шивон Зилис выбирает значение y из экспоненциального распределения x с интенсивностью x.

а) Найдите $\mathbb{E}(y \mid x)$, $\mathbb{V}\mathrm{ar}(y \mid x)$.

б) Найдите $\mathbb{E}(y)$, $\mathbb{E}(x)$, $\mathbb{C}\mathrm{ov}(x, y)$, $\mathbb{V}\mathrm{ar}(x)$.

в) Найдите наилучшее линейное приближение $\mathrm{BestLin}(y \mid x)$.

Здесь неотразимый ассист напишет решение

Теперь найдём условное ожидание и условную дисперсию для совместного нормального распределения в общем виде.

Определение 2.1. Наилучшее линейное приближение величины r с помощью величины s — это линейная функция от s,

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

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

$$\beta_1 = \ldots, \beta_2 = \ldots.$$

Определение 2.2. Величины r и s называются линейно-независимыми, если $\mathrm{BestLin}(r \mid s) = \mathbb{E}(r)$

Задача 5. Выразите константы β_1 и β_2 в формуле для наилучшего линейного приближения

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

исходя из характеристик случайных величин r и s.

Выпишем целевую функцию в виде суммы

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

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 $\mathrm{BestLin}(r \mid s) = \mathbb{E}(r)$ или условию $\mathrm{BestLin}(s \mid r) = \mathbb{E}(s)$.

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?

4.1. Черный трек

4.2. Тройка тестов в матричной форме

Рассмотрим применение тестов W (тест Вальда), LR (тест отношения правдоподобия) и LM (тест множителей Лагранжа) для тестирования гипотез о параметрах модели.

Пусть требуется протестировать систему ограничений относительно вектора неизвестных параметров

$$H_0 : \begin{cases} g_1(\theta) = 0 \\ g_2(\theta) = 0 \\ \dots \\ g_r(\theta) = 0 \end{cases}$$

где $g_i(\theta)$ — функция, которая задаёт i-ое ограничение на вектор параметров θ, $i = 1, \dots, r$.

Введём следующие обозначения:

$$\frac{\partial g}{\partial \theta^T} = \begin{pmatrix} \partial g_1 / \partial \theta^T \\ \partial g_2 / \partial \theta^T \\ \vdots \\ \partial g_r / \partial \theta^T \end{pmatrix} = \begin{pmatrix} \frac{\partial g_1}{\partial \theta_1} & \frac{\partial g_1}{\partial \theta_2} & \dots & \frac{\partial g_1}{\partial \theta_k} \\ \frac{\partial g_2}{\partial \theta_1} & \frac{\partial g_2}{\partial \theta_2} & \dots & \frac{\partial g_2}{\partial \theta_k} \\ \dots & \dots & \dots & \dots \\ \frac{\partial g_r}{\partial \theta_1} & \frac{\partial g_r}{\partial \theta_2} & \dots & \frac{\partial g_r}{\partial \theta_k} \end{pmatrix}$$

$$\frac{\partial g^T}{\partial \theta} = \begin{pmatrix} \frac{\partial g_1^T}{\partial \theta} & \frac{\partial g_2^T}{\partial \theta} & \dots & \frac{\partial g_r^T}{\partial \theta} \end{pmatrix} = \begin{pmatrix} \frac{\partial g_1}{\partial \theta_1} & \frac{\partial g_2}{\partial \theta_1} & \dots & \frac{\partial g_r}{\partial \theta_1} \\ \frac{\partial g_1}{\partial \theta_2} & \frac{\partial g_2}{\partial \theta_2} & \dots & \frac{\partial g_r}{\partial \theta_2} \\ \dots & \dots & \dots & \dots \\ \frac{\partial g_1}{\partial \theta_k} & \frac{\partial g_2}{\partial \theta_k} & \dots & \frac{\partial g_r}{\partial \theta_k} \end{pmatrix}, \frac{\partial \ell}{\partial \theta} = \begin{pmatrix} \frac{\partial \ell}{\partial \theta_1} \\ \frac{\partial \ell}{\partial \theta_2} \\ \vdots \\ \frac{\partial \ell}{\partial \theta_k} \end{pmatrix}$$

$$I(\theta) = -E\left(\frac{\partial^2 \ell}{\partial \theta \partial \theta^T}\right) = -\mathbb{E} \begin{pmatrix} \frac{\partial^2 \ell}{\partial \theta_1 \partial \theta_1} & \frac{\partial^2 \ell}{\partial \theta_1 \partial \theta_2} & \dots & \frac{\partial^2 \ell}{\partial \theta_1 \partial \theta_k} \\ \frac{\partial^2 \ell}{\partial \theta_2 \partial \theta_1} & \frac{\partial^2 \ell}{\partial \theta_2 \partial \theta_2} & \dots & \frac{\partial^2 \ell}{\partial \theta_2 \partial \theta_k} \\ \dots & \dots & \dots & \dots \\ \frac{\partial^2 \ell}{\partial \theta_k \partial \theta_1} & \frac{\partial^2 \ell}{\partial \theta_k \partial \theta_2} & \dots & \frac{\partial^2 \ell}{\partial \theta_k \partial \theta_k} \end{pmatrix}$$

— информационная матрица Фишера

$\Theta_{UR} := \Theta$ — множество допустимых значений вектора неизвестных параметров без учёта ограничений

$\Theta_R := \{\theta \in \Theta : g(\theta) = 0\}$ — множество допустимых значений вектора неизвестных параметров с учётом ограничений

$\hat{\theta}_{UR} \in \Theta_{UR}$ — точка максимума функции ℓ на множестве Θ_{UR}

$\hat{\theta}_R \in \Theta_R$ — точка максимума функции ℓ на множестве Θ_R

Тогда для тестирования гипотезы H_0 можно воспользоваться одной из следующих ниже статистик:

$LR := -2(\ell(\hat{\theta}_R) - \ell(\hat{\theta}_{UR})) \overset{as.}{\sim} \chi_r^2$ — статистика отношения правдоподобия

$W := g^T(\hat{\theta}_{UR}) \cdot \left[\frac{\partial g}{\partial \theta^T}(\hat{\theta}_{UR}) \cdot I^{-1}(\hat{\theta}_{UR}) \cdot \frac{\partial g^T}{\partial \theta}(\hat{\theta}_{UR})\right]^{-1} g(\hat{\theta}_{UR}) \overset{as.}{\sim} \chi_r^2$ — статистика Вальда

$LM := \left[\frac{\partial \ell}{\partial \theta}(\hat{\theta}_R)\right]^T \cdot I^{-1}(\hat{\theta}_R) \cdot \left[\frac{\partial \ell}{\partial \theta}(\hat{\theta}_R)\right] \overset{as.}{\sim} \chi_r^2$ — статистика множителей Лагранжа

5. ШБ — МНК

МНК в скалярной и матричной форме без статистических свойств. Строгая мультиколлинеарность.

Определим матрицу оператора ортогонального проектирования на подпространство, порожденное векторами x_j, $j = 1, \dots, k$:

$$H = X(X^T X)^{-1} X^T.$$

Матрица–проектор (hat-matrix) H является

- симметричной, то есть $H^T = H$

$$H^T = X(X^TX)^{-1}X^T = H$$

- идемпотентной, то есть как $H^2 = H$

$$H^2 = X(X^TX)^{-1}(X^TX)(X^TX)^{-1}X^T = X(X^TX)^{-1}X^T = H$$

- $\text{rank}\, H = \text{trace}\, H = k$

$$\text{trace}\, H = \text{trace}(X(X^TX)^{-1}X^T) = \text{trace}((X^TX)(X^TX)^{-1}) = \text{trace}\, I_k = k.$$

  Здесь мы использовали свойство следа: $\text{trace}(ABC) = \text{trace}(CAB)$, A, B, C — матрицы.

Определим матрицу $M = I - H$. Несложно убедиться, что матрица M так же, как и матрица H симметричная и идемпотентная. При этом $\text{trace}\, M = \text{trace}(I_n - H) = \text{trace}\, I_n - \text{trace}\, H = n - k$.

Из геометрического смысла матриц H и M следует, что

$$HX = X,\ MX = 0.$$

Выразим вектор остатков в матричном виде:

$$\hat{u} = y - \hat{y} = y - Hy = (I - H)y = My = M(X\beta + u) = Mu.$$

Пусть $s = \begin{pmatrix} 1 & 1 & \cdots & 1 \end{pmatrix}^T$ — вектор размерности $n \times 1$, состоящий из единиц. Определим матрицу $\pi = s^T(s^Ts)^{-1}s^T$. Матрица π — это матрица размерности $n \times n$ вида

$$\pi = \frac{1}{n}\begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \dots & \dots & \dots & \dots \\ 1 & 1 & \cdots & 1 \end{pmatrix}.$$

В качестве домашнего упражнения покажите, что $\pi c = \bar{c}$, где c — произвольный вектор размерности $n \times 1$.

Используя введенные обозначения, запишем TSS, ESS и RSS в матричном виде:

$$TSS = (y - \bar{y})^T(y - \bar{y}) = (y - \pi y)^T(y - \pi y) = (y(I - \pi))^T(y(I - \pi)) = (I - \pi)^Ty^Ty(I - \pi) = y^T(I - \pi)y$$

$$ESS = (\hat{y} - \bar{y})^T(\hat{y} - \bar{y})) = (Hy - \pi y)^T(Hy - \pi y) = (y(H - \pi))^T(y(H - \pi)) = (H - \pi)^Ty^Ty(H - \pi) = y^T(H - \pi)y$$

$$RSS = (y - \hat{y})^T(y - \hat{y}) = (y - Hy)^T(y - Hy) = (y(I - H))^T(y(I - H)) = (I - H)^Ty^Ty(I - H) = y^T(I - H)y$$

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 RSSc 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 $H = X(X^T X)^{-1} X^T$, $\hat{u}$ - 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:

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

□

5.2. Задачи для ДЗ

Задача 6. Пусть $s = \begin{pmatrix} 1 & 1 & \cdots & 1 \end{pmatrix}^T$ — вектор размерности $n \times 1$, состоящий из единиц. Определим матрицу $\pi = s^T(s^T s)^{-1} s^T$. Матрица π — это матрица размерности $n \times n$ вида

$$\pi = \frac{1}{n} \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \cdots & \cdots & \cdots & \cdots \\ 1 & 1 & \cdots & 1 \end{pmatrix}.$$

Покажите, что $\pi c = \bar{c}$, где c — произвольный вектор размерности $n \times 1$.

6. Предпосылки о математическом ожидании и дисперсии

В этой главе мы познакомимся с понятиями независимости и линейной независимости; расчётом математических ожиданий, ковариаций и дисперсий в матричном виде.

Добавим в метод наименьших квадратов ряд статистических предпосылок на ожидание и дисперсию.

Сформулируем и докажем теорему Гаусса - Маркова (которая пообещает, что МНК-оценки будут обладать свойствами несмещённости и эффективности).

6.1. Иерархия зависимостей случайных величин

Напомним определение наилучшей линейной аппроксимации

Определение 6.1. Наилучшее линейное приближение величины r с помощью величины s — это линейная функция от s,

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

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

$$\beta_1 = ..., \beta_2 = ...$$

Определение 6.2. Величины r и s называются линейно-независимыми, если $\mathrm{BestLin}(r \mid s) = \mathbb{E}(r)$

Некоторые авторы считают условие $\mathbb{C}\mathrm{ov}(r, s) = 0$ определением линейной независимости.

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 $\{s \in B\}$:

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

Из независимости величин r и s следует, что информация, известная об s, никак не помогает угадывать значение r. Поэтому условное математическое ожидание для r равно безусловному. Точно также из независимости r и s следует $\mathbb{E}(s \mid r) = \mathbb{E}(s)$. Обратное утверждение неверно, что показывает контрпример ниже.

Задача 7. Покажем, что из равенства условного и безусловного математических ожиданий не следует независимость случайных величин. Пусть дискретные случайные величины r характеризуют погоду (-1 снег, 1 солнце, 0 дождь), s – наличие зонта (0 нет или 1 есть) и ниже приведена таблица их совместного распределения.

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

Уже по формулировке подозреваем, что величины зависимые :).

Найдём условное ожидание зонта при условии, что мы видим погоду на улице: $\mathbb{E}(s \mid r) = \begin{cases} 0 \text{ если } r \in \{-1, \\ 1, \text{ если } r = 0. \end{cases}$

Получается, что информация о погоде помогает предсказать наличие зонтика, события не являются независимыми.

Найдём ожидания о погоде за окном, если вы можете наблюдать наличие или отсутствие зонта у человека: $\mathbb{E}(r \mid s) = \begin{cases} (-1) \times 1/6 + 1 \times 1/6 = 0, \text{ если } s = 0, \\ 0, \text{ если } s = 1. \end{cases}$.

Обычное безусловное ожидание погоды на улице: $\mathbb{E}(r) = (-1) \times 1/3 + 1 \times 1/3 + 0 \times 1/3 = 0$

Получается, что $\mathbb{E}(r \mid s) = \mathbb{E}(r) = 0$, но события зависимы.

Вернемся к тому факту, что из равенства условного и безусловного математических ожиданий следует нулевая ковариация. Используя закон повторных математических ожиданий $\mathbb{C}\mathrm{ov}(r, s) = \mathbb{E}(rs) - \mathbb{E}(r)\,\mathbb{E}(s) = \mathbb{E}(\mathbb{E}(rs \mid s)) - \mathbb{E}(r)\,\mathbb{E}(s) = \mathbb{E}(s\,\mathbb{E}(r \mid s)) - \mathbb{E}(r)\,\mathbb{E}(s) = \mathbb{E}(r)\,\mathbb{E}(s) - \mathbb{E}(r)\,\mathbb{E}(s) = 0$.

Задача 8. Из нулевой ковариации не следует равенство условного и безусловного математических ожиданий (и тем более не следует независимость). Пусть случайная величина s имеет равномерное распределение на отрезке $[-1; 1]$, а $r = s^2$.

Напоминаем, что для равномерно распределённой случайной величины $\mathbb{E}(s) = \frac{-1+1}{2} = 0, pdf(s) = \frac{1}{1-(-1)} = \frac{1}{2}$.

$\mathbb{E}(r \mid s) = \mathbb{E}(s^2 \mid s) = s^2 \neq 0$ в общем случае.

При этом $\mathbb{C}\mathrm{ov}(r, s) = \mathbb{E}(rs) - \mathbb{E}(r)\,\mathbb{E}(s) = \mathbb{E}(s^3) - \mathbb{E}(s^2) \times 0 = \mathbb{E}(s^3)$.

Математическое ожидание сложной функции $\mathbb{E}(g(x)) = \int_b^a g(x) pdf(x) dx$, если $x \in [a, b]$.

Найдём $\mathbb{E}(s^3) = \int_{-1}^1 s^3 pdf(s) ds = \int_{-1}^1 s^3 \frac{1}{2} ds = \frac{1}{8} s^4 |_{-1}^1 = 0$. Значит, мы получили нулевую ковариацию у зависимых случайных величин.

Вывод

Существуют независимые случайные величины, но на ???

6.2. 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

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

Ковариационная матрица вектора r определяется следующим образом:

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

Ковариационная матрица векторов r и s определяется следующим образом:

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

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

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

Условные ожидание и дисперсия определяются аналогично и обладают аналогичными свойствами. Главное — не забывать ставить вертикальную палочку!

Например,

написать пример

6.3. Теорема Гаусса — Маркова

$$y = X\beta + u$$

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,

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

Важно пояснить смысл введённой предпосылки. При оценивании связи между регрессорами X и переменной y мы не предполагаем, что величины u и X независимы. В ошибки модели попадают все те факторы, которые мы забыли включить в регрессию. Эти факторы могут быть взаимосвязаны с тем, что в регрессию всё же попало. Мы делаем более слабое предположение лишь о бесполезности всей собранной в X информации для угадывания u (и следующей из неё линейной независимости между ошибками и регрессорами, в том числе о нулевой ковариации).

Теорема 6.3. Если

1. Модель линейна по параметрам: $y = X\beta + u$;
2. Матрица X размера $[n \times k]$ имеет полный ранг k.
3. Условное ожидание ошибок равно нулю, $\mathbb{E}(u \mid X) = 0$;
4. Условная ковариационная матрица ошибок пропорциональна единичной, $\mathbb{V}\mathrm{ar}(u \mid X) = \sigma^2 I$;
5. Оценка $\hat{\beta}$ получена методом наименьших квадратов, $\hat{\beta} = (X^T X)^{-1} X^T y$;

то

(a) Оценка $\hat{\beta}$ является линейной по y;

(b) Оценка $\hat{\beta}$ является условно несмещённой, $\mathbb{E}(\hat{\beta} \mid X) = \beta$ и несмещённой, $\mathbb{E}(\hat{\beta}) = \beta$;

(c) Оценка любого коэффициента $\hat{\beta}_j$ является наиболее эффективной в классе линейных несмещённых оценок.

Что означает «эффективная в классе линейных несмещённых оценок»? Это означает, что у любой другой линейной по y несмещённой оценки $\hat{\beta}_j^{\text{alt}}$ дисперсия не меньше, чем у МНК-оценки.

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

В иностранной литературе для простоты запоминания используется аббревиатура BLUE – best linear unbiased estimator. То есть при выполнении условий теоремы Гаусса-Маркова мы получаем несмещённые и наилучшие (в терминах эффективности) оценки в классе всех линейных оценок.

Хорошие оценки подобны хорошему подвенечному платью, Something Olde, Something New, Something Borrowed, Something Blue, A Sixpence in your Shoe.

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

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

Доказательство. Линейность оценки по y видна прямо из её формулы, $\hat\beta = (X^TX)^{-1}X^Ty$.

Проверим условную несмещённость,

$$\mathbb{E}(\hat\beta \mid X) = \mathbb{E}((X^TX)^{-1}X^Ty \mid X) = (X^TX)^{-1}X^T\,\mathbb{E}(y \mid X).$$

Для удобства посчитаем $\mathbb{E}(y \mid X)$ отдельно,

$$\mathbb{E}(y \mid X) = \mathbb{E}(X\beta + u \mid X) = X\beta + \mathbb{E}(u \mid X) = X\beta.$$

И теперь завершаем вычисление $\mathbb{E}(\hat\beta \mid X)$:

$$\mathbb{E}(\hat\beta \mid X) = (X^TX)^{-1}X^T\,\mathbb{E}(y \mid X) = (X^TX)^{-1}X^TX\beta = \beta.$$

Мы доказали условную несмещённость оценки, $\mathbb{E}(\hat\beta \mid X) = \beta$. Безусловная несмещённость следует из свойства условного ожидания,

$$\mathbb{E}(\hat\beta) = \mathbb{E}(\mathbb{E}(\hat\beta \mid X)) = \mathbb{E}(\beta) = \beta.$$

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, $\hat\beta_1 = a^Ty$ и $\hat\beta_1^{\text{alt}} = a_{\text{alt}}^Ty$.

Замечаем, что $\mathbb{V}\text{ar}(\hat\beta_1 \mid X) = \sigma^2 a^Ta$, и $\mathbb{V}\text{ar}(\hat\beta_1^{\text{alt}}) = \sigma^2 a_{\text{alt}}^T a_{\text{alt}}$. 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.

здесь простая картинка с теоремой Пифагора!

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 эффективности, которое можно найти во многих учебниках, не замечает связи с теоремой Пифагора и исследует разницу ковариационных матриц. Приведём его здесь для демонстрации альтернативной техники!

Доказательство. У нас есть две линейных по y оценки: МНК-оценка и оценка-конкурент,

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

Оценки ковариационных матриц этих оценок равны

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

□

Условие несмещённости альтернативной оценки имеет вид

$$\mathbb{E}(\hat{\beta}_{\text{alt}} \mid X) = \mathbb{E}(A^T y \mid X) = A^T X\beta = \beta.$$

То есть для несмещённости альтернативной оценки должно выполняться условие $A^T X = I$. Для простоты рассмотрим случай $\sigma^2 = 1$. Мы докажем, что разница этих матриц $D = A^T A - (X^T X)^{-1}$ является положительно полуопределённой матрицы.

Вспомним из линейной алгебры определение и свойства положительно полуопределённой матрицы.

Определение 6.4. Матрица D или квадратичная форма $q(v) = v^T D v$ называется положительно полуопределённой, если $q(v) \geq 0$ для любого вектора v.

Теорема 6.5. Матрица D является положительно полуопределённой, если и только если её можно записать в виде произведения $D = B^T B$.

У положительно полуопределённой матрицы D на диагонали находятся неотрицательные числа.

Если $D = A^T A - (X^T X)^{-1}$ — положительно полуопределена, то $d_{ii} \geq 0$ и, следовательно, $[A^T A]_{ii} \geq [(X^T X)^{-1}]_{ii}$, то есть, дисперсии альтернативных оценок не меньше дисперсий МНК-оценок.

Перейдём к доказательству положительной полуопределённости D:

Доказательство. Возьмём $B = A - X(X^TX)^{-1}$ и найдём B^TB:

$$B^TB = (A-X(X^TX)^{-1})^T(A-X(X^TX)^{-1}) = A^TA-A^TX(X^TX)^{-1}-(X^TX)^{-1}X^TA+(X^TX)^{-1}X^TX(X^TX)^{-1}$$

В силу несмещённости $A^TX = I$ или $X^TA = I$, поэтому

$$B^TB = A^TA - (X^TX)^{-1} - (X^TX)^{-1} + (X^TX)^{-1} = A^TA - (X^TX)^{-1}.$$

Мы видим, что матрица $D = A^TA - (X^TX)^{-1}$ оказалась разложенной в произведение $D = B^TB$ и, следовательно, матрица D положительно полуопределена. □

6.4. Статистические свойства остатков

Используя матричное представление для остатков $\hat{u} = My = Mu$, вычислим вектор математических ожиданий и ковариационной матрицы:

$$\mathbb{E}(\hat{u} \mid X) = \mathbb{E}(My \mid X) = M\,\mathbb{E}(y \mid X) = MX\beta = 0, \text{ так как } MX = 0.$$

Ожидаемое значение остатков равно нулю, также как и ожидаемое значение ошибок, $\mathbb{E}(\hat{u} \mid X) = \mathbb{E}(u \mid X) = 0$.

$$\mathbb{V}\text{ar}(\hat{u} \mid X) = \mathbb{V}\text{ar}(My \mid X) = M\,\mathbb{V}\text{ar}(y \mid X)M^T = M\sigma^2I_nM^T = \sigma^2MM^T = \sigma^2M.$$

Вспомним, что у ковариационной матрицы ошибок $\mathbb{V}\text{ar}(u \mid X) = \sigma^2I$ на диагонали стоят одинаковые элементы, а вне диагонали стоят нули. А у ковариационной матрицы остатков $\mathbb{V}\text{ar}(\hat{u} \mid X) = \sigma^2M$ на диагоналях находятся разные элементы и вне диагонали элементы в общем случае не равны нулю.

Другими словами, остатки $\hat{u}_i$ зависимы между собой и имеют разную дисперсию $\mathbb{V}\text{ar}(\hat{u}_i)$. Например, при наличии константы в регрессии остатки обязательно удовлетворяют соотношению $\sum \hat{u}_i = 0$.

Посчитаем ковариационную матрицу вектора остатков и вектора прогнозов:

$$\mathbb{C}\text{ov}(\hat{u}, \hat{y} \mid X) = \mathbb{C}\text{ov}(Mu, Py \mid X) = \mathbb{C}\text{ov}(Mu, P(X\beta + u) \mid X) = \mathbb{C}\text{ov}(Mu, X\beta + Pu \mid X) =$$

$$= \mathbb{C}\text{ov}(Mu, Pu \mid X) = M\,\mathbb{C}\text{ov}(u, u \mid X)P = M\sigma^2I_nP^T = \sigma^2MP = 0, \text{так как } P^T = P \text{ и } MP = 0.$$

Следовательно, вектор остатков и вектор прогнозов линейно независимы. Метод наименьших квадратов даёт наилучший линейный прогноз, то есть даже зная прогнозные значения $\hat{y}$ нет возможности уменьшить остатки модели.

Посчитаем ковариационную матрицу вектора остатков и МНК-оценки вектора параметров β:

$$\mathbb{C}\text{ov}(\hat{u}, \hat{\beta} \mid X) = \mathbb{C}\text{ov}(Mu, \beta + (X^TX)^{-1}X^Tu \mid X) = \mathbb{C}\text{ov}(Mu, (X^TX)^{-1}X^Tu \mid X) =$$

$$= M\,\mathbb{C}\text{ov}(u, u \mid X)X(X^TX)^{-1} = M\,\mathbb{C}\text{ov}(u, u \mid X)X(X^TX)^{-1} = \sigma^2MX(X^TX)^{-1}, \text{так как } MX = 0.$$

Следовательно, вектор остатков и вектор МНК-оценок параметров модели.

6.5. Оценивание дисперсии

Метод наименьших квадратов позволяет оценить вектор параметров β, однако ~~власти скрывают настоящую дисперсию~~ никак не оценивает неизвестный параметр σ^2. Интуиция говорит, что высокая дисперсия ошибок u_i должна проявляться в высоком разбросе $\hat{u}_i$, поэтому разумно попробовать построить оценку $\hat{\sigma}^2$ на базе $RSS = \sum \hat{u}_i^2$.

Для построения оценки $\hat{\sigma}^2$ найдём ожидание $\mathbb{E}(RSS \mid X)$:

Теорема 6.6. Если выполнены предпосылки теоремы Гаусса — Маркова,

1. Модель линейна по параметрам: $y = X\beta + u$;
2. Матрица X размера $[n \times k]$ имеет полный ранг k.
3. Условное ожидание ошибок равно нулю, $\mathbb{E}(u \mid X) = 0$;
4. Условная ковариационная матрица ошибок пропорциональна единичной, $\mathbb{V}\mathrm{ar}(u \mid X) = \sigma^2 I$;
5. Оценка $\hat{\beta}$ получена методом наименьших квадратов, $\hat{\beta} = (X^T X)^{-1} X^T y$;

то $\mathbb{E}(RSS \mid X) = \mathbb{E}(\sum \hat{u}_i^2 \mid X) = (n-k)\sigma^2$.

Из этой теоремы следует, что оценка $\hat{\sigma}^2 = RSS/(n-k)$ — несмещённая оценка для неизвестной дисперсии σ^2.

Доказательство. На помощь нам придёт след матрицы! След матрицы прекрасен двумя свойствами. Во-первых, его можно менять местами с математическим ожиданием, $\mathbb{E}(\operatorname{trace} W)) = \operatorname{trace} \mathbb{E}(W)$. Во-вторых, внутри следа можно переставлять местами перемножаемые матрицы, $\operatorname{trace}(AB) = \operatorname{trace}(BA)$. Кроме того, на скалярную величину след можно навесить совершенно бесплатно! Если величина R — не вектор, а скаляр, то $\operatorname{trace} R = R$.

Продолжаем,

$$\mathbb{E}(\hat{u}^T \hat{u} \mid X) = \mathbb{E}(\operatorname{trace}(\hat{u}^T \hat{u}) \mid X) = \mathbb{E}(\operatorname{trace}(\hat{u}\hat{u}^T) \mid X) = \operatorname{trace} \mathbb{E}(\hat{u}\hat{u}^T \mid X).$$

Подумаем о середине,

$$\mathbb{E}(\hat{u}\hat{u}^T \mid X) = \mathbb{E}(Mu(Mu)^T \mid X) = \mathbb{E}(Muu^T M^T \mid X) = M\,\mathbb{E}(uu^T \mid X)M^T.$$

Вспомним, что матрица M — проектор, поэтому $M^T = M$, $M^2 = M$. У матрицы uu^T на диагонали стоят u_i^2, вне диагонали — $u_i u_j$. Поэтому $\mathbb{E}(uu^T \mid X) = \sigma^2 I$. Завершаем вычисления,

$$\mathbb{E}(\hat{u}\hat{u}^T \mid X) = M\,\mathbb{E}(uu^T \mid X)M^T = M \cdot \sigma^2 I \cdot M^T = \sigma^2 M^2 = \sigma^2 M$$

След проектора равен размерности пространства, на которое он проецирует, поэтому $\operatorname{trace} M = n - k$ и

$$\mathbb{E}(RSS \mid X) = \operatorname{trace}(\sigma^2 M) = (n-k)\sigma^2$$

□

Выборочная дисперсия при случайной выборке

Заметим, что данная теорема обобщает старый факт про выборочную дисперсию! Вспомним, что для выборки из независимых y_i с ожиданием $\mathbb{E}(y_i) = \mu$ и дисперсией $\mathbb{V}\mathrm{ar}(y_i) = \sigma^2$ несмещённая оценка дисперсии имеет вид

$$\hat{\sigma}^2 = \frac{\sum (y_i - \bar{y})^2}{n-1}.$$

В данном случае величины y_i можно представить в виде $y_i = \mu + u_i$. Тогда предпосылки теоремы Гаусса — Маркова выполнены, матрица регрессоров X — это просто единственный столбец-регрессор из единиц, $k = 1$, $\beta = \mu$. В этом случае $\hat{\beta} = \bar{y}$, все прогнозы равны $\hat{u}_i = \bar{y}$ и $RSS = \sum (y_i - \hat{y}_i)^2 = \sum (y_i - \bar{y})^2$. И мы видим, что новая оценка совпадает в этом случае со старой:

$$\hat{\sigma}^2 = \frac{RSS}{n-k} = \frac{\sum (y_i - \hat{y}_i)^2}{n-1} = \frac{\sum (y_i - \bar{y})^2}{n-1}$$

Оценка дисперсии оценок коэффициентов

Для построения доверительных интервалов для коэффициентов β_j нам понадобятся оценки дисперсий $\mathbb{V}\mathrm{ar}(\hat{\beta}_j \mid X)$. К счастью, если у нас есть несмещённая оценка $\hat{\sigma}^2$ для σ^2, то мы легко из неё построим оценку и для неизвестной ковариационной матрицы $\mathbb{V}\mathrm{ar}(\hat{\beta} \mid X) = \sigma^2 (X^T X)^{-1}$, просто подставив оценку вместо неизвестного параметра:

$$\widehat{\mathbb{V}\mathrm{ar}}(\hat{\beta} \mid X) = \hat{\sigma}^2 (X^T X)^{-1} = \frac{RSS}{n-k} (X^T X)^{-1}.$$

Уточним, что эту оценку мы вывели из предпосылок теоремы Гаусса — Маркова. Если использовать другие предпосылки, то ковариационная матрица $\mathbb{V}\mathrm{ar}(\hat{\beta} \mid X)$ перестанет быть равной $\sigma^2 (X^T X)^{-1}$ и нам потребуется другой способ оценивания.

6.6. Неправильная спецификация модели

Одной из предпосылок теоремы Гаусса — Маркова является правильный выбор спецификации, при котором мы регрессируем y в точности на набор истинных регрессоров. В реальности такое условие вряд ли выполнимо, так как не до всех регрессоров мы способны догадаться. А если догадаемся, то не все сможем измерить или собрать. Можно ли допустить неполную спецификацию модели, но получить BLUE-оценки (несмещённые и эффективные в классе линейных) для собранных регрессоров?

Пусть X — истинный набор регрессоров, а W — собранный датасет. При этом $X = [W\ V]$. Тогда новая МНК-оценка получается из изменившейся предпосылки о правильности спецификации $\tilde{\beta} = (W^T W)^{-1} W^T y$. Мы бы всё равно хотели получать несмещённую оценку.

$$\mathbb{E}(\tilde{\beta} \mid W) =$$

Что будет с эффективностью оценок, если пропущена часть важных регрессоров? Ковариационная матрица оценок коэффициентов $\tilde{\beta}$, полученная по сокращенному набору регрессоров W, имеет вид:

$$\mathbb{V}\mathrm{ar}(\hat{\beta}) = \sigma^2 (X^T X)^{-1}.$$

Вычислим ковариационную матрицу оценок коэффициентов $[\beta \quad \gamma]^T$, полученную по полному набору регрессоров:

$$\mathbb{V}\mathrm{ar}([\beta \quad \gamma]^T) = \sigma^2 ([W\ V]^T [W\ V]).$$

Применяя формулу блочного обращения матриц, получаем

$$\mathbb{V}\mathrm{ar}(\tilde{\beta}) = \sigma^2 (W^T W - W^T V (V^T V)^{-1} V^T W)^{-1}.$$

Рассмотрим разность двух ковариационных матриц:

$$\mathbb{V}\mathrm{ar}(\hat{\beta}) - \mathbb{V}\mathrm{ar}(\tilde{\beta}) =$$

6.7. Задачи для доски:

Задача 9. 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 $\mathbb{V}\mathrm{ar}(u_i \mid X) = \sigma^2$ и u_i при фиксированных X независимы. Найдите $\mathbb{V}\mathrm{ar}(y_i \mid X)$, $\mathbb{V}\mathrm{ar}(y_i(x_i - \bar{x}) \mid X)$, $\mathbb{V}\mathrm{ar}(\sum y_i(x_i - \bar{x}) \mid X)$, $\mathbb{V}\mathrm{ar}(\hat{\beta}_2 \mid X)$.

Здесь нужны решения

Задача 10. Рассмотрим модель $y_i = \beta_1 + \beta_2 x_i + \beta_3 z_i + u_i$, где

$$X = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}, \quad y = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix}, \quad u = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \end{pmatrix}.$$

Случайные ошибки u_i независимы и нормально распределены с $\mathbb{E}(u \mid X) = 0$ и $\mathbb{V}\mathrm{ar}(u \mid X) = \sigma^2 I$.

Для удобства расчётов даны матрицы: $X'X$ и $(X'X)^{-1}$

$$X'X = \begin{pmatrix} 5 & 3 & 1 \\ 3 & 3 & 1 \\ 1 & 1 & 1 \end{pmatrix}, \quad (X'X)^{-1} = \begin{pmatrix} 0.5 & -0.5 & 0 \\ -0.5 & 1 & -0.5 \\ 0 & -0.5 & 1.5 \end{pmatrix}.$$

а) Найдите $\mathbb{E}(\hat{\sigma}^2 \mid X)$, $\hat{\sigma}^2$.

б) Найдите $\mathbb{V}\mathrm{ar}(u_1)$, $\mathbb{V}\mathrm{ar}(\beta_1)$, $\mathbb{V}\mathrm{ar}(\hat{\beta}_1 \mid X)$, $\widehat{\mathbb{V}\mathrm{ar}}(\hat{\beta}_1 \mid X)$, $\mathbb{E}(\hat{\beta}_1^2 \mid X) - \beta_1^2$;

в) Найдите $\mathbb{C}\mathrm{ov}(\hat{\beta}_2, \hat{\beta}_3 \mid X)$, $\widehat{\mathbb{C}\mathrm{ov}}(\hat{\beta}_2, \hat{\beta}_3 \mid X)$, $\mathbb{V}\mathrm{ar}(\hat{\beta}_2 - \hat{\beta}_3 \mid X)$, $\widehat{\mathbb{V}\mathrm{ar}}(\hat{\beta}_2 - \hat{\beta}_3 \mid X)$;

г) Найдите $\mathbb{V}\mathrm{ar}(\beta_2 - \beta_3)$, $\mathbb{C}\mathrm{orr}(\hat{\beta}_2, \hat{\beta}_3 \mid X)$, $\widehat{\mathbb{C}\mathrm{orr}}(\hat{\beta}_2, \hat{\beta}_3 \mid X)$;

Дополнительно предположим, что случайные ошибки имеют нормальное распределение при фиксированных X.

д) Постройте 95% доверительный интервал для β_2.

е) Проверьте гипотезу H_0: $\beta_2 = 0$ против H_a: $\beta_2 \neq 0$;

ж) Проверьте гипотезу H_0: $\beta_2 = \beta_3$ против H_a: $\beta_2 \neq \beta_3$;

replace 4 by σ^2

check order of questions

а) $\mathbb{V}\mathrm{ar}(u_1) = \mathbb{V}\mathrm{ar}(u)_{(1,1)} = 4 \cdot I_{(1,1)} = 4$

б) $\mathbb{V}\mathrm{ar}(\beta_1) = 0$, так как β_1 — детерминированная величина.

в) $\mathbb{V}\mathrm{ar}(\hat{\beta}_1) = \sigma^2 (X'X)^{-1}_{(1,1)} = 0.5\sigma^2 = 0.5 \cdot 4 = 2$

г) $\widehat{\mathbb{V}\mathrm{ar}}(\hat{\beta}_1) = \hat{\sigma}^2 (X'X)^{-1}_{(1,1)} = 0.5\hat{\sigma}^2_{(1,1)} = 0.5\frac{RSS}{5-3} = 0.25RSS = 0.25y'(I - X(X'X)^{-1}X')y = 0.25 \cdot 1 = 0.25$
$\hat{\sigma}^2 = \frac{RSS}{n-k} = \frac{1}{2}$.

д) Так как оценки МНК являются несмещёнными, то $\mathbb{E}(\hat{\beta}) = \beta$, значит:

$$\mathbb{E}(\hat{\beta}_1) - \beta_1^2 = \mathbb{E}(\hat{\beta}_1) - (\mathbb{E}(\hat{\beta}_1))^2 = \widehat{\mathbb{V}\mathrm{ar}}(\hat{\beta}_1) = 0.25$$

е) $\mathbb{C}\mathrm{ov}(\hat{\beta}_2, \hat{\beta}_3) = \sigma^2 (X'X)^{-1}_{(2,3)} = 4 \cdot \left(-\frac{1}{2}\right) = -2$

ж) $\widehat{\mathbb{C}\mathrm{ov}}(\hat{\beta}_2, \hat{\beta}_3) = \widehat{\mathbb{V}\mathrm{ar}}(\hat{\beta})_{(2,3)} = \hat{\sigma}^2 (X'X)^{-1}_{(2,3)} = \frac{1}{2} \cdot \left(-\frac{1}{2}\right) = -\frac{1}{4}$

з) $\mathbb{V}\mathrm{ar}(\hat{\beta}_2 - \hat{\beta}_3) = \mathbb{V}\mathrm{ar}(\hat{\beta}_2) + \mathbb{V}\mathrm{ar}(\hat{\beta}_3) + 2\,\mathbb{C}\mathrm{ov}(\hat{\beta}_2, \hat{\beta}_3) = \sigma^2((X'X)^{-1}_{(2,2)} + (X'X)^{-1}_{(3,3)} + 2(X'X)^{-1}_{(2,3)} = 4(1 + 1.5 + 2 \cdot (-0.5)) = 6$

и) $\widehat{\mathbb{V}\mathrm{ar}}(\hat{\beta}_2 - \hat{\beta}_3) = \widehat{\mathbb{V}\mathrm{ar}}(\hat{\beta}_2) + \widehat{\mathbb{V}\mathrm{ar}}(\hat{\beta}_3) + 2\,\widehat{\mathbb{C}\mathrm{ov}}(\hat{\beta}_2, \hat{\beta}_3) = \hat{\sigma}^2((X'X)^{-1}_{(2,2)} + (X'X)^{-1}_{(3,3)} + 2(X'X)^{-1}_{(2,3)} = \frac{1}{2} \cdot 1.5 = 0.75$

к) $\mathbb{V}\mathrm{ar}(\beta_2 - \beta_3) = 0$

л) $\mathbb{C}\mathrm{orr}(\hat{\beta}_2, \hat{\beta}_3) = \frac{\mathbb{C}\mathrm{ov}(\hat{\beta}_2, \hat{\beta}_3)}{\sqrt{\mathbb{V}\mathrm{ar}(\hat{\beta}_2)\,\mathbb{V}\mathrm{ar}(\hat{\beta}_3)}} = \frac{-2}{\sqrt{4 \cdot 6}} = -\frac{\sqrt{6}}{6}$

м) $\widehat{\mathbb{C}\mathrm{orr}}(\beta_2, \beta_3) = \frac{\widehat{\mathbb{C}\mathrm{ov}}(\hat{\beta}_2, \hat{\beta}_3)}{\sqrt{\widehat{\mathbb{V}\mathrm{ar}}(\hat{\beta}_2)\,\widehat{\mathbb{V}\mathrm{ar}}(\hat{\beta}_3)}} = \frac{-\frac{1}{4}}{\sqrt{\frac{1}{2} \cdot \frac{3}{4}}} = -\frac{\sqrt{6}}{6}$

н) $(n-k)\frac{\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-k}$.

$$\mathbb{E}\left((n-k)\frac{\hat{\sigma}^2}{\sigma^2}\right) = n - k$$

$$\mathbb{E}\left(\frac{\hat{\sigma}^2}{2}\right) = 1$$

$$\mathbb{E}(\hat{\sigma}^2) = 2$$

о) $\hat{\sigma}^2 = \frac{RSS}{n-k} = \frac{1}{2}$

Задача 11. Рассмотрим классическую линейную модель $y = X\beta + u$ с предпосылками Гаусса — Маркова: $\mathbb{E}(u \mid X) = 0$ и $\mathbb{V}\mathrm{ar}(u \mid X) = \sigma^2 I$. Для всех случайных векторов (y, $\hat{y}$, $\hat{\beta}$, u, $\hat{u}$, $\bar{y}$) найдите все возможные ожидания и ковариационные матрицы $\mathbb{E}(\cdot)$, $\mathbb{V}\mathrm{ar}(\cdot)$, $\mathbb{C}\mathrm{ov}(\cdot, \cdot)$.

Здесь нужны решения

Задача 12. 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 $\mathbb{V}\mathrm{ar}(\hat{\beta}_{\text{ols}} \mid x)$ 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 $\mathbb{V}\mathrm{ar}(\hat{\beta}_{\text{best}} \mid x)$?

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 13. Предположим, что все предпосылки теоремы Гаусса — Маркова выполнены. Вычислите математические ожидания для TSS, ESS и RSS, используя их матричные представления.

Здесь нужны решения

Задача 14. (Hansen 4.14)

Задана модель $y = X\beta + u$, для которой выполняются предпосылки теоремы Гаусса — Маркова. Вас интересует величина $\theta = \beta^2$. Получены МНК-оценки коэффициентов: $\hat{\beta}$, $V_{\hat{\beta}} = \mathbb{V}\text{ar}[\hat{\beta} \mid X]$. Кажется, неплохой идеей будет оценить θ как $\hat{\theta} = \hat{\beta}^2$.

а) Найдите $\mathbb{E}[\hat{\theta} \mid X]$. Является ли $\hat{\theta}$ смещённой?

б) Предложите способ коррекции смещения для получения несмещённой оценки $\hat{\theta}^*$, используя результаты предыдущего пункта.

Здесь нужны решения

Задача 15. Рассмотрим модель регрессии $y_i = \beta_1 + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + u_i$. 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 $\beta_2 = \cdots = \beta_k = 0$.

а) Найдите $\mathbb{E}(R^2)$.

Решение:

Модель без ограничений:

$$y_i = \beta_1 + \beta_2 x_{i1} + \cdots + \beta_k x_{ik} + \varepsilon_i.$$

Модель с ограничениями (истинная модель!):

$$y_i = \beta_1 + \varepsilon_i.$$

Тогда F-статистика имеет следующий вид:

$$F = \frac{R^2(k-1)}{(1-R^2)/(n-k)} \sim F(k-1, n-k).$$

Выразим R^2:

$$R^2(n-k) = F(1-R^2)(n-k)$$

Факт дня №1: Если $X \sim F(k_1, k_2)$, то $Y = \frac{\frac{k_1}{k_2}X}{1+\frac{k_1}{k_2}X} \sim Beta\left(\frac{k_1}{2}, \frac{k_2}{2}\right)$.

Используя факт дня №1, получаем:

$$R^2 = \frac{(k-1)F}{(n-k)+(k-1)F} = \frac{\frac{k-1}{n-k}F}{1+\frac{k-1}{n-k}F} \sim \text{Beta}\left(\frac{k-1}{2}, \frac{n-k}{2}\right).$$

Тогда чтобы посчитать математическое ожидание R^2, надо вспомнить, чему равно математическое ожидание для $Beta\left(\frac{k-1}{2}, \frac{n-k}{2}\right)$:

$$E(R^2) = \frac{\frac{k-1}{2}}{\frac{k-1}{2}+\frac{n-k}{2}} = \frac{k-1}{n-1}.$$

Что нам даёт полученный результат? Математическое ожидание коэффициента детерминации линейно по k. То есть даже при включении в модель лишних факторов R^2 все равно продолжает линейно расти!

б) Найдите $\mathbb{E}(R^2_{\text{adj}})$.
Решение:
Скорректированный коэффициент детерминации имеет вид:

$$R^2_{adj.} = 1 - \left(1-R^2\right)\frac{n-1}{n-k}.$$

Рассчитаем математическое ожидание:

$$E(R^2_{adj.}) = E\left(1 - \left(1-R^2\right)\frac{n-1}{n-k}\right) = 1 - \frac{n-1}{n-k} + \frac{n-1}{n-k}E(R^2) =$$

$$= 1 - \frac{n-1}{n-k} + \frac{n-1}{n-k}\frac{k-1}{n-1} = 0.$$

Скорректированный R^2 помог решить проблему линейного роста по k!

Задача 16. 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

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

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 $S_{ij} = \sigma^2$.

6.8. 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

6.9. 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 17. 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 $\mathbb{V}\text{ar}(\hat{\beta}_1)$. Не противоречит ли полученной результат теореме Гаусса — Маркова?

Задача 18. 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 $\mathbb{V}\text{ar}(\hat{\beta}_1)$. 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.

7.1. Случай многомерного нормального распределения

сопроводить оценкой правдоподобия и показать, что она совпадает с МНК

Напомним несколько фактов про многомерное нормальное распределение.

Начнём с классического определения:

Определение 7.1. Вектор v имеет многомерное невырожденное нормальное распределение, $v \sim \mathcal{N}(\mu, C)$, если его совместная функция плотности равна

$$f(v) = (2\pi)^{-n/2} \det(C)^{-1/2} \exp\left(-\frac{1}{2}(v-\mu)^T C^{-1}(v-\mu)\right),$$

где n — размерность вектора v.

Заметим, что совместный закон распределения нормального вектора v полностью определён его ожиданием $\mathbb{E}(v)$ и его ковариационной матрицей $\mathbb{V}\text{ar}(v)$. Никакие другие параметры в совместную функцию плотности не входят.

Для многомерного нормального распределения нет разницы между независимостью и некоррелированностью:

Теорема 7.2. Если нормальный вектор v состоит из двух подвекторов, $v = (x, y)$, то $\mathbb{C}\text{ov}(x, y) = 0$ если и только если подвекторы x и y независимы.

Доказательство. Докажем в одну сторону. Если подвекторы x и y независимы, то $\mathbb{C}\text{ov}(x, y) = 0$. А теперь изящно докажем в обратную сторону. Если $\mathbb{C}\text{ov}(x, y) = 0$, то вся ковариационная матрица $\mathbb{V}\text{ar}(v)$ ровно такая же как и в случае независимых x и y. Остаётся лишь вспомнить, что $\mathbb{E}(v)$ и $\mathbb{V}\text{ar}(v)$ полностью определяют закон распределения нормального вектора v, а значит компоненты обязаны быть независимы. $\square$

Также для многомерного нормального распределения нет разницы между условным ожиданием $\mathbb{E}(y \mid x)$ и наилучшим линейным приближением $\text{BestLin}(y \mid x)$, другими словами функция $\mathbb{E}(y \mid x)$ линейна по x.

Задача 19. Рассмотрим совместное нормальное распределение

$$\begin{pmatrix} x \\ y \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}, \begin{pmatrix} C_{xx} & C_{xy} \\ C_{yx} & C_{yy} \end{pmatrix}\right)$$

а) Найдите наилучшее линейное приближение $\text{BestLin}(y \mid x)$.

б) Найдите условное ожидание $\mathbb{E}(y \mid x)$.

Здесь рассказать про определение bestlin в векторном случае?

Пусть $\text{BestLin}(y \mid x) = a + Bx$. Мы хотим, чтобы ошибка линейной аппроксимации не зависела от x, то есть,

$$\mathbb{C}\text{ov}(y - \text{BestLin}(y \mid x), x) = 0.$$

Подставим $\text{BestLin}(y \mid x) = a + Bx$.

$$\mathbb{C}\text{ov}(y, x) = \mathbb{C}\text{ov}(Bx, x)$$

Отсюда $C_{yx} = BC_{xx}$ и $B = C_{xx}^{-1}C_{yx}$. Кроме того, ошибка линейной аппроксимации должна иметь нулевое ожидание, следовательно,

$$\mathbb{E}(y) = \mathbb{E}(\text{BestLin}(y \mid x)) = a + B\,\mathbb{E}(x).$$

Следовательно,

дописать доказательство

И напоследок отметим, что для компонент x и y нормального вектора (x, y) условная дисперсия постоянна и не зависит от x.

Введём дополнительную предпосылку $(u|X) \sim \mathcal{N}(0, \sigma^2 I)$. Учитывая, что $\hat{\beta} = (X^TX)^{-1}X^Ty = \beta + (X^TX)^{-1}X^Tu$, получаем

$$(\hat{\beta} \mid X) \sim \mathcal{N}(\beta, \sigma^2(X^TX)^{-1}).$$

7.2. Независимость оценок β и $\hat{\sigma}^2$

МНК-оценка вектора коэффициентов β имеет вид

$$\hat{\beta} = (X^TX)^{-1}X^Ty.$$

Несмещенная оценка дисперсии случайной ошибки:

$$\hat{\sigma}^2 = \frac{RSS}{n-k} = \frac{\hat{u}^T\hat{u}}{n-k}.$$

Распишем

$$\hat{\beta} = (X^TX)^{-1}X^Ty = (X^TX)^{-1}X^T(X\beta + u) = \beta + (X^TX)^{-1}X^Tu = \beta + Au,$$

где $A = (X^TX)^{-1}X^T$.

В случае, когда случайный вектор ошибок u является нормальным, можно показать, что оценки $\hat{\beta}$ и $\hat{u}$ будут независимыми.

При $u \sim \mathcal{N}(0, \sigma^2 I_n)$ случайные векторы $\hat{\beta}$ и $\hat{u}$ имеют совместное многомерное нормальное распределение. Покажем, что $\hat{\beta}$ и $\hat{u}$ являются некоррелированными, из чего следует, что они также будет и независимыми, что справедливо для нормально распределенных векторов:

$$\mathbb{C}\text{ov}(\hat{\beta}, \hat{u}) = \mathbb{C}\text{ov}(\beta + Au, Mu) = \mathbb{C}\text{ov}(Au, Mu) = AM\,\mathbb{V}\text{ar}(u) = \sigma^2 AM = 0, \textbf{ так как } AM = 0.$$

Так как $\hat{\sigma}^2$ есть функция от случайного вектора $\hat{u}$, то оценки $\hat{\beta}$ и $\hat{\sigma}^2$ также независимы.

7.3. Проверка гипотез о параметрах

$$H_0 : \beta_j = \beta_j^0$$
$$H_1 : \beta_j \neq \beta_j^0$$

$$t = \frac{\hat{\beta}_j - \beta_j^0}{\text{se}(\hat{\beta}_j)} \overset{H_0}{\sim} t(n-k)$$

Проверка гипотезы о незначимости модели в целом

$$H_0 : \beta_2 = \cdots = \beta_k = 0$$

$$H_1 : \sum_{j=2}^{k} \beta_j^2 > 0$$

$$F = \frac{R^2/(k-1)}{(1-R^2)/(n-k} \overset{H_0}{\sim} F(k-1, n-k).$$

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. Задачи

15. Логистическая регрессия: точечные оценки

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

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

16. Логистическая регрессия: доверительные интервал

Логистическая регрессия: доверительные интервалы и проверка гипотез.

Чёрный трэк: разные хоббиты

16.1. 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.

16.2. 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

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

16.3. 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

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

16.4. 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

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

где $\mathrm{IMR}(s)$ — обратное отношение Миллса, inverse Mills ratio,

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

Предельный эффект для ожидания усечённой переменной

Выборочное смещение

Переключающиеся режимы