

---

# Содержание

<b>1</b>	<b>Методы получения оценок</b>	<b>2</b>
<b>2</b>	<b>Свойства оценок</b>	<b>2</b>
2.1	Практическое напоминание об условном ожидании и дисперсии . . . . .	2
<b>3</b>	<b>Асимптотические методы</b>	<b>5</b>
<b>4</b>	<b>Святая троица тестов</b>	<b>5</b>
4.1	Чёрный трек . . . . .	5
4.2	Тройка тестов в матричной форме . . . . .	5
<b>5</b>	<b>ШБ — МНК</b>	<b>6</b>
5.1	Чёрный трек . . . . .	7
5.2	Задачи для ДЗ . . . . .	8
<b>6</b>	<b>Предпосылки о математическом ожидании и дисперсии</b>	<b>8</b>
6.1	Иерархия зависимостей случайных величин . . . . .	8
6.2	Ожидание и ковариационная матрица . . . . .	10
6.3	Теорема Гаусса — Маркова . . . . .	11
6.4	Статистические свойства остатков . . . . .	14
6.5	Оценивание дисперсии . . . . .	15
6.6	Неправильная спецификация модели . . . . .	16
6.7	Задачи для семинара: . . . . .	18
6.8	Компьютерные задачи для семинара: . . . . .	23
6.9	Домашнее задание: . . . . .	23
6.10	Чёрный трек: . . . . .	23
<b>7</b>	<b>Доверительные интервалы для коэффициентов</b>	<b>23</b>
7.1	Случай многомерного нормального распределения . . . . .	24
7.2	Независимость оценок $\beta$ и $\hat{\sigma}^2$ . . . . .	26
7.3	Проверка гипотез о параметрах . . . . .	26
<b>8</b>	<b>Бутстрэп</b>	<b>26</b>
<b>9</b>	<b>Выбор функциональной формы</b>	<b>27</b>
<b>10</b>	<b>Гетероскедастичность</b>	<b>27</b>
<b>11</b>	<b>Мультиколлинеарность и метод главных компонент</b>	<b>27</b>
<b>12</b>	<b>Эндогенность</b>	<b>27</b>
<b>13</b>	<b>Эффекты воздействия</b>	<b>27</b>
<b>14</b>	<b>Задачи</b>	<b>27</b>
<b>15</b>	<b>Логистическая регрессия: точечные оценки</b>	<b>27</b>

---

<b>16 Логистическая регрессия: доверительные интервал</b>	<b>27</b>
16.1 Смещение, цензурирование и ■■■■■■	28
16.2 Цензурирование	28
16.3 Усечение	28
16.4 Три осмысленных условных ожидания	29

## 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

- Найдите  $\mathbb{E}(y | x)$ ,  $\mathbb{V}\text{ar}(y | x)$ .
- Найдите  $\mathbb{E}(y)$ ,  $\mathbb{E}(x)$ ,  $\mathbb{C}\text{ov}(x, y)$ ,  $\mathbb{V}\text{ar}(x)$ .
- Найдите наилучшее линейное приближение  $\text{BestLin}(y | x)$ .

**Решение.** а) Условное математическое ожидание и дисперсия:

$$\mathbb{E}(y | x) = \begin{cases} 2.5, & x = 1, \\ 2, & x = 2, 4, \end{cases}$$

$$\mathbb{V}\text{ar}(y | x) = \begin{cases} 0.75, & x = 1, \\ 1, & x = 2, 4. \end{cases}$$

- Математические ожидания, ковариация и дисперсия:

$$\mathbb{E}(y) = 2.2, \quad \mathbb{E}(x) = 2.4, \quad \mathbb{C}\text{ov}(x, y) = -0.28, \quad \mathbb{V}\text{ar}(x) = 1.84.$$

- Наилучшее линейное приближение:

$$\text{BestLin}(y | x) \approx 2.57 - 0.15 \cdot x.$$

---

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  $\mathbb{E}(y \mid x)$ ,  $\text{Var}(y \mid x)$ .
- б) Найдите  $\mathbb{E}(y)$ ,  $\mathbb{E}(x)$ ,  $\text{Cov}(x, y)$ ,  $\text{Var}(x)$ .
- в) Найдите наилучшее линейное приближение  $\text{BestLin}(y \mid x)$ .

**Решение.**

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  $\mathbb{E}(y \mid x)$ ,  $\text{Var}(y \mid x)$ .
- б) Найдите  $\mathbb{E}(y)$ ,  $\mathbb{E}(x)$ ,  $\text{Cov}(x, y)$ ,  $\text{Var}(x)$ .
- в) Найдите наилучшее линейное приближение  $\text{BestLin}(y \mid x)$ .

**Решение.**

Здесь храбрый ассист напишет решение

Обратите внимание: для совместного нормального распределения условное ожидание  $\mathbb{E}(y \mid x)$  и наилучшее линейное приближение  $\text{BestLin}(y \mid x)$  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  $\mathbb{E}(y \mid x)$ ,  $\text{Var}(y \mid x)$ .
- б) Найдите  $\mathbb{E}(y)$ ,  $\mathbb{E}(x)$ ,  $\text{Cov}(x, y)$ ,  $\text{Var}(x)$ .
- в) Найдите наилучшее линейное приближение  $\text{BestLin}(y \mid x)$ .

**Решение.**

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). Наилучшее линейное приближение величины  $r$  с помощью величины  $s$  — это линейная функция от  $s$ ,

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

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

$$\beta_1 = \mathbb{E}(r) - \beta_2 \mathbb{E}(s), \quad \beta_2 = \frac{\text{Cov}(r, s)}{\text{Var}(s)}$$

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

Линейная независимость является симметричным явлением,  $\text{BestLin}(r \mid s) = \mathbb{E}(r)$ , если и только если  $\text{BestLin}(s \mid r) = \mathbb{E}(s)$ .

**Задача 5.** Выразите константы  $\beta_1$  и  $\beta_2$  в формуле для наилучшего линейного приближения

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

исходя из характеристик случайных величин  $r$  и  $s$ .

**Решение.** Выпишем целевую функцию в виде суммы

$$\mathbb{E}((r - \text{BestLin}(r, s))^2) = \text{Var}(r - \beta_1 - \beta_2 s) + (\mathbb{E}(r - \beta_1 - \beta_2 s))^2$$

Заметим, что  $\beta_1$  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  $\beta_2$  первое слагаемое.

$$\text{Var}(r - \beta_2 s) = \text{Var}(r) + \beta_2^2 \text{Var}(s) - 2\beta_2 \text{Cov}(r, s) \rightarrow \min_{\beta_2}.$$

Перед нами квадратичная функция от  $\beta_2$ , следовательно,

$$\beta_2 = \frac{\text{Cov}(r, s)}{\text{Var}(s)}.$$

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

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

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  $\text{Cov}(r, s) = 0$  равносильно условию  $\text{BestLin}(r \mid s) = \mathbb{E}(r)$  или условию  $\text{BestLin}(s \mid r) = \mathbb{E}(s)$ .

---

---

### 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. Чёрный 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  $i$ -е ограничение на вектор параметров  $\theta$ ,  $i = 1, \dots, r$ .

Введём следующие обозначения:

$$\begin{aligned} \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}, \quad \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} \end{aligned}$$

— 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  $\ell$  на множестве  $\Theta_{UR}$

---

---

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

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

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

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

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

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

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

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

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

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

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

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

Выразим вектор остатков в матричном виде:

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

Пусть  $s = (1 \ 1 \ \dots \ 1)^T$  — вектор размерности  $n \times 1$ , состоящий из единиц. Определим матрицу  $\pi = s^T (s^T s)^{-1} s^T$ . Матрица  $\pi$  — это матрица размерности  $n \times n$  вида

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

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:

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

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

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

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 RSS с ростом числа наблюдений

## 5.1. Чёрный трэк

**Определение 5.1** (LOOCV). Кросс-валидация с поочередным выкидыванием отдельных наблюдений. На английском языке она часто сокращается LOOCV (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** (связь обычных и кросс-валидационных остатков). 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 — меньше, 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:

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

□

## 5.2. Задачи для ДЗ

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

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

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** (наилучшая линейная аппроксимация). Наилучшее линейное приближение величины  $r$  с помощью величины  $s$  — это линейная функция от  $s$ ,

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

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

$$\beta_1 = \mathbb{E}(r) - \frac{\text{Cov}(r, s)}{\text{Var}(s)} \mathbb{E}(s), \quad \beta_2 = \frac{\text{Cov}(r, s)}{\text{Var}(s)}.$$

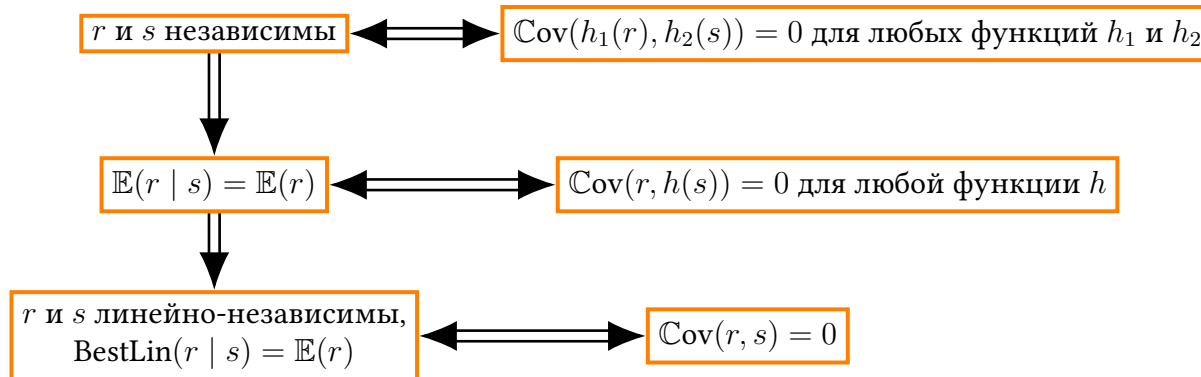
---



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

Некоторые авторы считают условие  $\text{Cov}(r, s) = 0$  определением линейной независимости.

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



Напомним, что определение независимых случайных величин,

**Определение 6.3** (независимость случайных величин). Случайные величины  $r$  и  $s$  называются независимыми если для любых<sup>1</sup> 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

**Решение.** Уже по формулировке подозреваем, что величины зависимые :).

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

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.

<sup>1</sup>Не совсем любых, требуется измеримость множеств. В рамках нашего курса мы не будем обращать внимания на данный нюанс.

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

**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  $r = s^2$ .

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

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

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

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

Найдём  $\mathbb{E}(s^3) = \int_{-1}^1 s^3 \text{pdf}(s) ds = \int_{-1}^1 s^3 \frac{1}{2} ds = \frac{1}{8} s^4 \Big|_{-1}^1 = 0$ . 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

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

Ковариационная матрица вектора  $r$  определяется следующим образом:

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

Ковариационная матрица векторов  $r$  и  $s$  определяется следующим образом:

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

Свойства вектора математических ожиданий и ковариационной матрицы:

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

б)  $\text{Cov}(r, s) = \mathbb{E}(rs^T) - \mathbb{E}(r) \mathbb{E}(s^T)$

$$\text{в) } \text{Cov}(Ar + b, s) = A \text{Cov}(r, s)$$

$$\text{г) } \text{Cov}(r, As + b) = \text{Cov}(r, s)A^T$$

$$\text{д) } \text{Var}(r) = \text{Cov}(r, r) = \mathbb{E}(rr^T) - \mathbb{E}(r) \mathbb{E}(r^T)$$

$$\text{е) } \text{Var}(Ar + b) = A \text{Var}(r) A^T$$

$$\text{ж) } \mathbb{E}(r^T Ar) = \text{trace}(A \text{Var}(r)) + \mathbb{E}(r^T) A \mathbb{E}(r)$$

$$\text{з) Если вектора } r \text{ и } s \text{ имеют одинаковый размер, то } \text{Var}(r + s) = \text{Var}(r) + \text{Var}(s) + \text{Cov}(r, s) + \text{Cov}(s, r)$$

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,

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

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 6.4** (Гаусс — Марков). 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,  $\text{Var}(u \mid X) = \sigma^2 I$ ;
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,  $\mathbb{E}(\hat{\beta} | X) = \beta$  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.

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

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  $w^T \hat{\beta}^{\text{alt}}$ :

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

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

Проверим условную несмещённость,

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

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

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

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

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

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

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

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

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

Разобьём 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  $\beta$  должно выполняться:

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

Переносим всё в левую сторону:

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

Получаем, что для любых  $\beta$  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 комбинации 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  $X$ .  $\square$

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,

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

Оценки ковариационных матриц этих оценок равны

$$\text{Var}(\hat{\beta} | X) = (X^T X)^{-1} \sigma^2 \text{ и } \text{Var}(\hat{\beta}_{\text{alt}} | X) = A^T A \sigma^2.$$

$\square$

Условие несмещённости альтернативной оценки имеет вид

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

То есть для несмещённости альтернативной оценки должно выполняться условие  $A^T X = I$ . Для простоты рассмотрим случай  $\sigma^2 = 1$ . Мы докажем, что разность 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 6.5** (положительно полуопределённая форма). Матрица  $D$  или квадратичная форма  $q(v) = v^T D v$  называется положительно полуопределённой, если  $q(v) \geq 0$  для любого вектора  $v$ .

**Теорема 6.6** (свойства положительно полуопределённой матрицы). 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  $B = A - X(X^T X)^{-1}$  и найдём  $B^T B$ :

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

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

$$B^T B = A^T A - (X^T X)^{-1} - (X^T X)^{-1} + (X^T X)^{-1} = A^T A - (X^T X)^{-1}.$$

Мы видим, что матрица  $D = A^T A - (X^T X)^{-1}$  оказалась разложенной в произведение  $D = B^T B$  и, следовательно, матрица  $D$  положительно полуопределена.  $\square$

## 6.4. Статистические свойства остатков

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

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

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

$$\text{Var}(\hat{u} | X) = \text{Var}(My | X) = M \text{Var}(y | X) M^T = M \sigma^2 I_n M^T = \sigma^2 M M^T = \sigma^2 M.$$

Вспомним, что у ковариационной матрицы ошибок  $\text{Var}(u | X) = \sigma^2 I$  на диагонали стоят одинаковые элементы, а вне диагонали стоят нули. А у ковариационной матрицы остатков  $\text{Var}(\hat{u} | X) = \sigma^2 M$  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  $\text{Var}(\hat{u}_i)$ . Например, при наличии константы в регрессии остатки обязательно удовлетворяют соотношению  $\sum \hat{u}_i = 0$ .

Посчитаем ковариационную матрицу вектора остатков и вектора прогнозов:

$$\begin{aligned} \text{Cov}(\hat{u}, \hat{y} | X) &= \text{Cov}(Mu, Py | X) = \text{Cov}(Mu, P(X\beta + u) | X) = \text{Cov}(Mu, X\beta + Pu | X) = \\ &= \text{Cov}(Mu, Pu | X) = M \text{Cov}(u, u | X) P = M \sigma^2 I_n P^T = \sigma^2 M P = 0, \text{ так как } P^T = P \text{ и } M P = 0. \end{aligned}$$

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  $\beta$ :

$$\begin{aligned} \text{Cov}(\hat{u}, \hat{\beta} | X) &= \text{Cov}(Mu, \beta + (X^T X)^{-1} X^T u | X) = \text{Cov}(Mu, (X^T X)^{-1} X^T u | X) = \\ &= M \text{Cov}(u, u | X) X (X^T X)^{-1} = M \text{Cov}(u, u | X) X (X^T X)^{-1} = \sigma^2 M X (X^T X)^{-1}, \text{ так как } M X = 0. \end{aligned}$$

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

## 6.5. Оценивание дисперсии

Метод наименьших квадратов позволяет оценить вектор параметров  $\beta$ , однако ~~власти скрывают~~ настоящую дисперсию никак не оценивает неизвестный параметр  $\sigma^2$ . 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  $\mathbb{E}(RSS | X)$ :

**Теорема 6.7** (ожидание суммы квадратов остатков). 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,  $\mathbb{E}(u | X) = 0$ ;
4. Условная ковариационная матрица ошибок пропорциональна единичной,  $\text{Var}(u | X) = \sigma^2 I$ ;
5. Оценка  $\hat{\beta}$  получена методом наименьших квадратов,  $\hat{\beta} = (X^T X)^{-1} X^T y$ ;

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

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

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

Продолжаем,

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

Подумаем о середине,

$$\mathbb{E}(\hat{u} \hat{u}^T | X) = \mathbb{E}(Mu(Mu)^T | X) = \mathbb{E}(Mu u^T M^T | X) = M \mathbb{E}(u u^T | X) M^T.$$

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

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

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

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

И мы легко строим несмещённую оценку,  $\hat{\sigma}^2 = RSS/(n - k)$ ,

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

□

---

## 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  $\text{Var}(y_i) = \sigma^2$  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  $\beta_j$  нам понадобятся оценки дисперсий  $\text{Var}(\hat{\beta}_j | X)$ . К счастью, у нас есть несмещённая оценка  $\hat{\sigma}^2$  для  $\sigma^2$ . Из неё мы легко построим оценку и для неизвестной ковариационной матрицы  $\text{Var}(\hat{\beta} | X) = \sigma^2 (X^T X)^{-1}$ . А именно, мы просто подставим оценку дисперсии вместо неизвестной дисперсии:

$$\widehat{\text{Var}}(\hat{\beta} | X) = \hat{\sigma}^2 (X^T X)^{-1} = \frac{RSS}{n - k} (X^T X)^{-1}.$$

Уточним, что эту оценку мы вывели из предпосылок теоремы Гаусса — Маркова. Если использовать другие предпосылки, то ковариационная матрица  $\text{Var}(\hat{\beta} | X)$  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?

Рассмотрим для начала случай, когда при оценивании модели мы пропускаем часть важных регрессоров. Истинная модель имеет вид

$$y = W\beta + V\gamma + u,$$

где  $W$  — матрица регрессоров размерности  $n \times k_1$ ,  $V$  — матрица регрессоров размерности  $n \times k_2$ . Обозначим через  $X = [W \ V]$   $n \times k$  матрицу всех регрессоров, где  $k = k_1 + k_2$ .

Вместо истинной модели оценивается следующая модель:

$$y = W\beta + \nu,$$

где  $\nu$  — вектор случайных ошибок в оцениваемой модели.

**Утверждение 6.1.** тут должно быть утверждение про смещённость

---



*Доказательство.* 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  $\beta = (W^T W)^{-1} W^T y$ . Мы бы всё равно хотели получать несмещённую оценку.

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

□

**Утверждение 6.2.** Пусть  $\text{Var}(\hat{\beta} \mid W, V)$  — ковариационная матрица вектора оценок  $\beta$ , полученного по полному набору регрессоров  $X = [W \ V]$ , а  $\text{Var}(\tilde{\beta} \mid W)$  — ковариационная матрица вектора оценок  $\beta$ , полученного по регрессорам из матрицы  $W$ . Тогда матрица  $\text{Var}(\hat{\beta} \mid W, V) - \text{Var}(\tilde{\beta} \mid W)$  является положительно полуопределённой матрицей.

Утверждение 6.2 означает, что на диагонали матрицы  $\text{Var}(\hat{\beta} \mid W, V) - \text{Var}(\tilde{\beta} \mid W)$  стоят неотрицательные значения. В свою очередь, диагональный элемент с индексами  $jj$  представляет собой разницу дисперсий оценок коэффициента  $\beta_j$ , полученных по полному и по сокращённому набору переменных. Это означает, что  $\text{Var}(\hat{\beta}_j \mid W, V) - \text{Var}(\tilde{\beta}_j \mid W) \geq 0$ , то есть оценка  $\tilde{\beta}_j$  имеет меньшую условную дисперсию. Из-за меньшей условной дисперсии оценка  $\tilde{\beta}_j$  может получиться более эффективной по сравнению с оценкой  $\hat{\beta}_j$ .

**Утверждение 6.3.** Оценка дисперсии случайной ошибки  $\tilde{\sigma}^2 = \frac{RSS}{n-k_1}$ , полученная по модели с пропущенными переменными, является смещённой,

$$\mathbb{E}(\tilde{\sigma}^2 \mid W) \neq \sigma^2$$

*Доказательство.* Вспомним матричное представление  $RSS$ :

$$RSS = y^T M y, \text{ где } M = I_n - W(W^T W)^{-1} W^T.$$

Рассчитаем математическое ожидание  $RSS$ , учитывая, что истинной моделью является модель по набору регрессоров  $X = [W \ V]$ :

$$\begin{aligned} \mathbb{E}(RSS \mid W, V) &= \mathbb{E}(y^T M y \mid W, V) = \mathbb{E}((W\beta + V\gamma + u)^T M (W\beta + V\gamma + u) \mid W, V) = \\ &= \mathbb{E}(u^T M u + 2\gamma^T V^T M u + \gamma^T V^T M V \gamma \mid W, V) = \sigma^2(n - k_1) + \gamma^T V^T M V \gamma. \end{aligned}$$

Выше мы использовали следующие результаты:

- $MW = 0$ ;
- $\mathbb{E}(2\gamma^T V^T M u \mid W, V) = 2\gamma^T V^T M \mathbb{E}(u \mid W, V) = 0$ ;
- $\mathbb{E}(u^T M u \mid W, V) = \sigma^2(n - k_1)$ .

Таким образом, получаем, что

$$\mathbb{E}(\tilde{\sigma}^2 \mid W, V) = \mathbb{E}\left(\frac{RSS}{n - k_1} \mid W, V\right) = \frac{1}{n - k_1}(\sigma^2(n - k_1) + \gamma^T V^T M V \gamma) = \sigma^2 + \frac{1}{n - k_1} \gamma^T V^T M V \gamma.$$

Оценка дисперсии  $\tilde{\sigma}^2$  будет несмещённой только, если  $\gamma = 0$ . Равенство  $\gamma = 0$  означает, что пропущенных переменных нет и  $X = W$ . Заметим также, что  $V^T M V = (M V)^T M V$ , что означает, что матрица  $V^T M V$  является положительно полуопределённой. Следовательно, смещение оценки  $\tilde{\sigma}^2$  в общем случае положительное. □

Далее проанализируем, что происходит со свойствами несмещённости и эффективности МНК-оценок в случае включения в модель лишних регрессоров.

Теперь истинной моделью является

$$y = X\beta + u.$$

Вместо истинной модели оценивается следующая модель:

$$y = X\beta + R\gamma + \nu,$$

где  $R$  — матрица лишних регрессоров.

**Утверждение 6.4.** При включении лишних регрессоров МНК-оценка  $\tilde{\beta}$ , полученная в модели с набором регрессоров  $(X \ R)$ , остаётся несмещённой, то есть  $\mathbb{E}(\tilde{\beta} \mid X, R) = \beta$ .

**Утверждение 6.5.** Пусть  $\text{Var}(\hat{\beta} \mid X)$  — ковариационная матрица вектора оценок  $\beta$ , полученного по истинному набору регрессоров  $X$ , а  $\text{Var}(\tilde{\beta} \mid X, R)$  — ковариационная матрица вектора оценок  $\beta$ , полученного по регрессорам из матрицы

$$(X \ R)$$

. Тогда матрица  $\text{Var}(\tilde{\beta} \mid X, R) - \text{Var}(\hat{\beta} \mid X)$  является положительно полуопределённой матрицей.

Утверждение 6.5 означает, что на диагонали матрицы  $\text{Var}(\tilde{\beta} \mid X, R) - \text{Var}(\hat{\beta} \mid X)$  стоят неотрицательные значения. В свою очередь, диагональный элемент с индексами  $j, j$  представляет собой разницу дисперсий оценок коэффициента  $\beta_j$ , полученных по расширенному и по истинному наборам переменных. Это означает, что  $\text{Var}(\tilde{\beta}_j \mid X, R) - \text{Var}(\hat{\beta}_j \mid X) \geq 0$ , то есть оценка  $\tilde{\beta}_j$  имеет большую условную дисперсию. Из-за большей условной дисперсии оценка  $\tilde{\beta}_j$  может получиться менее эффективной по сравнению с оценкой  $\hat{\beta}_j$ .

## 6.7. Задачи для семинара:

**Задача 9.** Исследовательница Мишель собрала данные по 20 студентам. Переменная  $y_i$  — количество решённых задач по эконометрике  $i$ -м студентом, а  $x_i$  — количество просмотренных серий любимого сериала за прошедший год. Оказалось, что  $\sum y_i = 10$ ,  $\sum x_i = 0$ ,  $\sum x_i^2 = 40$ ,  $\sum y_i^2 = 50$ ,  $\sum x_i y_i = 60$ .

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

**Решение.**

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  $\mathbb{E}(u | X) = 0$  и  $\text{Var}(u | X) = \sigma^2 I$ .

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  $\mathbb{E}(\hat{\sigma}^2 | X)$ ,  $\hat{\sigma}^2$ .

б) Найдите  $\text{Var}(u_1)$ ,  $\text{Var}(\beta_1)$ ,  $\text{Var}(\hat{\beta}_1 | X)$ ,  $\widehat{\text{Var}}(\hat{\beta}_1 | X)$ ,  $\mathbb{E}(\hat{\beta}_1^2 | X) - \beta_1^2$ ;

в) Найдите  $\text{Cov}(\hat{\beta}_2, \hat{\beta}_3 | X)$ ,  $\widehat{\text{Cov}}(\hat{\beta}_2, \hat{\beta}_3 | X)$ ,  $\text{Var}(\hat{\beta}_2 - \hat{\beta}_3 | X)$ ,  $\widehat{\text{Var}}(\hat{\beta}_2 - \hat{\beta}_3 | X)$ ;

г) Найдите  $\text{Var}(\beta_2 - \beta_3)$ ,  $\text{Corr}(\hat{\beta}_2, \hat{\beta}_3 | X)$ ,  $\widehat{\text{Corr}}(\hat{\beta}_2, \hat{\beta}_3 | X)$ ;

**Решение.**

replace 4 by  $\sigma^2$

check order of questions

а)  $\text{Var}(u_1) = \text{Var}(u)_{(1,1)} = 4 \cdot I_{(1,1)} = 4$

б)  $\text{Var}(\beta_1) = 0$ , так как  $\beta_1$  — детерминированная величина.

в)  $\text{Var}(\hat{\beta}_1) = \sigma^2 (X'X)^{-1}_{(1,1)} = 0.5\sigma^2 = 0.5 \cdot 4 = 2$

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

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

$$\mathbb{E}(\hat{\beta}_1) - \beta_1^2 = \mathbb{E}(\hat{\beta}_1) - (\mathbb{E}(\hat{\beta}_1))^2 = \widehat{\text{Var}}(\hat{\beta}_1) = 0.25$$

е)  $\text{Cov}(\hat{\beta}_2, \hat{\beta}_3) = \sigma^2 (X'X)^{-1}_{(2,3)} = 4 \cdot (-\frac{1}{2}) = -2$

ж)  $\widehat{\text{Cov}}(\hat{\beta}_2, \hat{\beta}_3) = \widehat{\text{Var}}(\hat{\beta})_{(2,3)} = \hat{\sigma}^2 (X'X)^{-1}_{(2,3)} = \frac{1}{2} \cdot (-\frac{1}{2}) = -\frac{1}{4}$

з)  $\text{Var}(\hat{\beta}_2 - \hat{\beta}_3) = \text{Var}(\hat{\beta}_2) + \text{Var}(\hat{\beta}_3) + 2\text{Cov}(\hat{\beta}_2, \hat{\beta}_3) = \sigma^2((X'X)^{-1}_{(2,2)} + (X'X)^{-1}_{(3,3)} + 2(X'X)^{-1}_{(2,3)}) = 4(1 + 1.5 + 2 \cdot (-0.5)) = 6$

и)  $\widehat{\text{Var}}(\hat{\beta}_2 - \hat{\beta}_3) = \widehat{\text{Var}}(\hat{\beta}_2) + \widehat{\text{Var}}(\hat{\beta}_3) + 2\widehat{\text{Cov}}(\hat{\beta}_2, \hat{\beta}_3) = \hat{\sigma}^2((X'X)^{-1}_{(2,2)} + (X'X)^{-1}_{(3,3)} + 2(X'X)^{-1}_{(2,3)}) = \frac{1}{2} \cdot 1.5 = 0.75$

к)  $\text{Var}(\beta_2 - \beta_3) = 0$

л)  $\text{Corr}(\hat{\beta}_2, \hat{\beta}_3) = \frac{\text{Cov}(\hat{\beta}_2, \hat{\beta}_3)}{\sqrt{\text{Var}(\hat{\beta}_2) \text{Var}(\hat{\beta}_3)}} = \frac{-2}{\sqrt{4 \cdot 6}} = -\frac{\sqrt{6}}{6}$

м)  $\widehat{\text{Corr}}(\hat{\beta}_2, \hat{\beta}_3) = \frac{\widehat{\text{Cov}}(\hat{\beta}_2, \hat{\beta}_3)}{\sqrt{\widehat{\text{Var}}(\hat{\beta}_2) \widehat{\text{Var}}(\hat{\beta}_3)}} = \frac{-\frac{1}{4}}{\sqrt{\frac{1}{2} \cdot \frac{3}{4}}} = -\frac{\sqrt{6}}{6}$

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

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

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

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

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

**Задача 11.** Рассмотрим классическую линейную модель  $y = X\beta + u$  с предпосылками Гаусса — Маркова:  $\mathbb{E}(u | X) = 0$  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  $\mathbb{E}(\cdot)$ ,  $\text{Var}(\cdot)$ ,  $\text{Cov}(\cdot, \cdot)$ .

**Решение.**

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  $\beta$  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?

**Решение.** а)  $\hat{\beta}_{\text{ols}} = (y_1 + 2y_2)/5$ ;

$$\text{б)} \text{Var}(\hat{\beta}_{\text{ols}} | x) = 1/5;$$

- 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  $u_1 = 1, u_2 = -1$ .

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

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

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

---

**Решение.**

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:  $\hat{\beta}$ ,  $V_{\hat{\beta}} = \text{Var}[\hat{\beta} | X]$ . Кажется, неплохой идеей будет оценить  $\theta$  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.

**Решение.**

Здесь нужны решения

**Задача 15.** Рассмотрим модель регрессии  $y_i = \beta_1 + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + u_i$ . Все предпосылки теоремы Гаусса — Маркова выполнены. Дополнительно предположим, что  $u_i \sim \mathcal{N}(0, \sigma^2)$ ,  $i = 1, \dots, n$ . Дополнительно известно, что на самом деле  $\beta_2 = \dots = \beta_k = 0$ .

- а) Найдите  $\mathbb{E}(R^2)$ .

**Решение:**

Модель без ограничений:

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

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  $Y = \frac{\frac{k_1}{k_2} X}{1 + \frac{k_1}{k_2} X} \sim \text{Beta}\left(\frac{k_1}{2}, \frac{k_2}{2}\right)$ .

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  $\text{Beta}\left(\frac{k-1}{2}, \frac{n-k}{2}\right)$ :

---

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

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  $\mathbb{E}(R_{\text{adj}}^2)$ .

**Решение:**

Скорректированный коэффициент детерминации имеет вид:

$$R_{\text{adj}}^2 = 1 - (1 - R^2) \frac{n-1}{n-k}.$$

Рассчитаем математическое ожидание:

$$\begin{aligned} E(R_{\text{adj}}^2) &= E\left(1 - (1 - R^2) \frac{n-1}{n-k}\right) = 1 - \frac{n-1}{n-k} + \frac{n-1}{n-k} E(R^2) = \\ &= 1 - \frac{n-1}{n-k} + \frac{n-1}{n-k} \frac{k-1}{n-1} = 0. \end{aligned}$$

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?

**Решение.** а) Ковариационная матрица будет содержать блоки  $B$  на диагонали

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

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, левый верхний блок пропорционален единичной матрицы, а все элементы правого нижнего блока равны  $\sigma^2$ :

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

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

---

## 6.8. Компьютерные задачи для семинара:

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. Домашнее задание:

## 6.10. 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  $\beta_1$  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 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  $\beta_1$  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.

---

## 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** (многомерное нормальное распределение). 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  $\text{Var}(v)$ . Никакие другие параметры в совместную функцию плотности не входят.

Для многомерного нормального распределения нет разницы между независимостью и некоррелированностью:

**Теорема 7.2** (некоррелированность и независимость для нормального вектора). Если нормальный вектор  $v$  состоит из двух подвекторов,  $v = (x, y)$ , то  $\text{Cov}(x, y) = 0$  если и только если подвекторы  $x$  и  $y$  независимы.

*Доказательство.* Докажем в одну сторону. Если подвекторы  $x$  и  $y$  независимы, то  $\text{Cov}(x, y) = 0$ . А теперь изящно докажем в обратную сторону. Если  $\text{Cov}(x, y) = 0$ , то вся ковариационная матрица  $\text{Var}(v)$  ровно такая же как и в случае независимых  $x$  и  $y$ . Остаётся лишь вспомнить, что  $\mathbb{E}(v)$  и  $\text{Var}(v)$  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  $\mathbb{E}(y \mid x)$ .
- в) Найдите условную дисперсию  $\text{Var}(y \mid x)$ .

**Решение.**

Здесь рассказать про определение bestlin в векторном случае?



- 
- а) Пусть  $\text{BestLin}(y \mid x) = a + Bx$ . Мы хотим, чтобы ошибка линейной аппроксимации  $r = y - \text{BestLin}(y \mid x)$  была некоррелирована с  $x$ ,

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

Другими словами,

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

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

$$\mathbb{Cov}(y, x) = \mathbb{Cov}(a + Bx, x) = \mathbb{Cov}(Bx, x) = B \mathbb{Cov}(x, x) = B \mathbb{Var}(x).$$

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

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

Получаем уравнение на  $a$ :

$$\mu_y = a + C_{yx}C_{xx}^{-1}\mu_x$$

И ответ,

$$\begin{cases} B = C_{yx}C_{xx}^{-1} \\ a = \mu_y - C_{yx}C_{xx}^{-1}\mu_x \end{cases}$$

- б) Для нормально распределённой пары векторов нулевая ковариация равносильная независимости. Следовательно, ошибка аппроксимации  $r = y - \text{BestLin}(y \mid x)$  и  $x$  независимы. Отсюда мы получаем, что для многомерного нормально распределённого вектора  $(x, y)$

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

- в) Условная дисперсия — это безусловная дисперсия ошибки прогноза,

$$\mathbb{Var}(y \mid x) = \mathbb{Var}(a + Bx + r \mid x) = \mathbb{Var}(r \mid x) = \mathbb{Var}(r).$$

Осталось вспомнить, что  $y = a + Bx + r$ , прогноз  $a + Bx$  и ошибка  $r$  некоррелированы,

$$\mathbb{Var}(y) = \mathbb{Var}(a + Bx) + \mathbb{Var}(r).$$

Значит,

$$\mathbb{Var}(y \mid x) = \mathbb{Var}(r) = \mathbb{Var}(y) - B \mathbb{Var}(x) B^T = C_{yy} - C_{yx}C_{xx}^{-1}C_{xx}C_{xx}^{-1}C_{xy} = C_{yy} - C_{yx}C_{xx}^{-1}C_{xy}.$$

Отметим, что для компонент  $x$  и  $y$  нормального вектора  $(x, y)$  условная дисперсия получилась постоянной и не зависящей от  $x$ .

Для ненормального распределения условное ожидание  $\mathbb{E}(y \mid x)$  и условная дисперсия  $\mathbb{Var}(y \mid x)$  вполне могут быть нелинейными.

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

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

---

---

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

МНК-оценка вектора коэффициентов  $\beta$  имеет вид

$$\hat{\beta} = (X^T X)^{-1} X^T y.$$

Несмещённая оценка дисперсии случайной ошибки:

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

Распишем

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

где  $A = (X^T X)^{-1} X^T$ .

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 будут и независимыми, что справедливо для нормально распределённых векторов:

$$\text{Cov}(\hat{\beta}, \hat{u}) = \text{Cov}(\beta + Au, Mu) = \text{Cov}(Au, Mu) = AM \text{Var}(u) = \sigma^2 AM = 0, \text{ так как } AM = 0.$$

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

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

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

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

Проверка гипотезы о незначимости модели в целом

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

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

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

## 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. Задачи

## 15. Логистическая регрессия: точечные оценки

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

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

## 16. Логистическая регрессия: доверительные интервал

Логистическая регрессия: доверительные интервалы и проверка гипотез.

Чёрный трек: разные хоббиты

---

## 16.1. 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.

## 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.

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

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

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

---

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 увидевший цен, полезность от ужина,

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

Предельный эффект для латентной переменной

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

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

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

Предельный эффект для цензурированной переменной

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

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

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

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

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

Предельный эффект для ожидания усечённой переменной

### Выборочное смещение

### Переключающиеся режимы

## List of Theorems

2.1	Определение (наилучшее линейное приближение)	4
2.2	Определение (линейно-независимые случайные величины)	4
5.1	Определение (LOOCV)	7
5.2	Теорема (связь обычных и кросс-валидационных остатков)	7
6.1	Определение (наилучшая линейная аппроксимация)	8
6.2	Определение (линейная независимость)	9
6.3	Определение (независимость случайных величин)	9
6.4	Теорема (Гаусс — Марков)	11
6.5	Определение (положительно полуопределённая форма)	14
6.6	Теорема (свойства положительно полуопределённой матрицы)	14
6.7	Теорема (ожидание суммы квадратов остатков)	15
7.1	Определение (многомерное нормальное распределение)	24
7.2	Теорема (некоррелированность и независимость для нормального вектора)	24

---