Множественная Линейная Регрессия и Теорема Гаусса-Маркова I

Evgeny Sedashov, PhD
esedashov@hse.ru

3/02/2024

Ревью

- На прошлом занятии мы рассмотрели самую простую форму регрессионного анализа, а именно парную регрессию.
- Мы увидели, что МНК-оценки в парной регрессии обладают привлекательными статистическими свойствами, а именно состоятельностью и несмещённостью.
- Простота парной регрессии является и её проблемой – подавляющее большинство социальных феноменов невозможно объяснить только одной независимой переменной.
- Множественная регрессия позволяет моделировать более сложные процессы, в том числе такие, где есть переменные взаимодействия.

Множественная Регрессия и Двухфакторная (Factorial) ANOVA

- Множественная регрессия может служить полноценной заменной двухфакторной ANOVA.
- Для простых исследовательских дизайнов ANOVA всё равно является предпочитаемым методом анализа данных из-за концептуальной привлекательности данного подхода.
- Множественная регрессия подходит для более сложных проблем и, в особенности, для observational data.

Пример

- Предположим, что мы хотим понять факторы явки на выборы, используя данные на уровне УИКов; какие независимые переменные мы будем использовать?
- Образование (средний уровень по домохозяйствам на территории УИК).
- Возраст (средний по домохозяйствам на территории УИК).
- Цена недвижимости (в среднем за кв.м.).
- Погода в день выборов (1 – плохая, 0 – хорошая).
- Получаем следующее уравнение:

$$turnout = \beta_0 + \beta_1 educ + \beta_2 age + \beta_3 housprice + \beta_4 weather + u$$

- β_0 – интерсепт (константа), $\beta_1 - \beta_4$ – коэффициенты наклона; u – ошибка в генеральной совокупности (error term).

Интерпретация I

- Уравнение на предыдущем слайде не имеет простой визуализации, потому что оно соответствует гиперплоскости, заданной в 5 измерениях.
- Мы можем использовать интерпретацию "при прочих равных": при заданных значениях возраста, погодных условий и цены за кв.м., увеличение уровня образования на единицу даст изменение в явке на величину β_1.
- Идея "при прочих равных" – центральная в том числе для визуализации результатов (противоположной идеей является концепция общего равновесия).

Интерпретация II

- Интепретируя коэффициенты регрессии, полезно помнить о том, что интерпретация может иногда зависеть от функциональных форм зависимой и независимой переменных.
- Если зависимая и независимая переменные присутствуют в анализе в своих неизмененных формах, применяется стандартная интерпретация.
- Иногда нам требуется взять натуральный логариф зависимой, независимой, или обеих переменных сразу.

Интерпретация III

- Модель **level-log**: в рамках данной модели, один процент увеличения в x ведёт к $\beta/100$ единиц изменения в y.
- Модель **log-level**: в рамках данной модели, увеличение x на единицу ведёт к изменению y на $100 * \beta$%.
- Модель **log-log**: в рамках данной модели, увеличение x на один процент ведёт к изменению y на β%.

База I

- Мы обозначаем $\hat{\beta}_0$, $\hat{\beta}_1$ etc. МНК-оценки, полученные для конкретной выборки; параметры без "домиков" – параметры в генеральной совокупности.

- Мы обозначаем $\hat{y}_i$ предсказанные значения зависимой переменной (linear prediction или fitted value на английском):

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \ldots + \hat{\beta}_n x_{in}$$

- Мы обозначаем $\hat{u}_i$ остатки:

$$\hat{u}_i = y_i - \hat{y}_i$$

База II

- Остатки $\neq$ ошибки в генеральной совокупности!
- Во всех рассуждениях вектор – это вертикальный (column) вектор, а его транспонированная версия – это горизонтальный (raw) вектор.
- Жирным шрифтом обозначаются матрицы и векторы, если не оговорено обратное.
- Большие буквы ($\mathbf{X}$) – матрицы, малые ($\mathbf{y}$) – векторы.
- Для обозначения столбцов матриц (как правило, в контексте переменных) мы будем использовать мылые буквы с нижним индексом: $\mathbf{x_1}$ обозначает первую независимую переменную.

R^2

- R^2 для множественной регрессии определяется точно так же, как и для парной:

$$R^2 = \frac{\sum_{i=1}^{N}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} = \frac{SSE}{SST}$$

где SSE – the explained sum of squares, а SST – the total sum of squares.

- $R^2 \in [0; 1]$.

Введение

- Поскольку мы имеем дело с количеством параметров > 2, представление МНК-оценок в матричной форме является более удобным.
- Нам нужно два инструмента: частные производные и простейшие операции с матрицами и векторами.
- Мы рассмотрим условия первого порядка.

Условия Первого Порядка I

- Рассмотрим следующую регресссию для некоторой выборки:

$$y_i = \hat{\beta}_0 x_{i0} + \hat{\beta}_1 x_{i1} + \ldots + \hat{\beta}_n x_{in} + \hat{u}_i$$

- i указывает на номер строки в матрице, xs – независимые переменные.

- Заметьте, что x_{i0} – это не настоящая переменная, а искусственно созданный вектор с единицами в качестве элементов.

Условия Первого Порядка II

- Для нахождения условий первого порядка, необходимо минимизировать сумму квадратов остатков регрессии: $\hat{u}_i$:

$$\sum_{i=1}^{K}(y_i - \hat{\beta}_0 x_{i0} - \ldots - \hat{\beta}_n x_{in})^2$$

- Мы знаем, что в этом случае условия первого порядка задаются частными производными относительно всех параметров регрессии, приравненными к нулю.
- Мы получим $n+1$ частных производных, потому что мы имеем $n+1$ параметров регрессии.
- Заметьте, что общее количество наблюдений обозначено как K и суммирование по строкам массива!

Условия Первого Порядка III

- Получаем следующую систему уравнений:

$$-2\sum_{i=1}^{K} x_{i0}(y_i - \hat{\beta}_0 x_{i0} - \ldots - \hat{\beta}_n x_{in}) = 0$$

$$-2\sum_{i=1}^{K} x_{i1}(y_i - \hat{\beta}_0 x_{i0} - \ldots - \hat{\beta}_n x_{in}) = 0$$

$$\vdots$$

$$-2\sum_{i=1}^{K} x_{in}(y_i - \hat{\beta}_0 x_{i0} - \ldots - \hat{\beta}_n x_{in}) = 0$$

Условия Первого Порядка IV

- Заметьте, что мы получили, по сути, два вектора с $n+1$ элементов каждый: левая часть содержит вектор, состоящий из неизвестных чисел, правая часть – вектор с нулями.
- Перепишем условия первого порядка:

$$\sum_{i=1}^{K} x_{i0} y_i = \sum_{i=1}^{K} x_{i0}(\hat{\beta}_0 x_{i0} + ... + \hat{\beta}_n x_{in})$$

$$\vdots$$

$$\sum_{i=1}^{K} x_{in} y_i = \sum_{i=1}^{K} x_{in}(\hat{\beta}_0 x_{i0} + ... + \hat{\beta}_n x_{in})$$

Условия Первого Порядка V

- Для получения формулы МНК-оценок необходимо упростить систему уравнений на предыдущем слайде, используя матричные операции.
- Левая часть упрощается элементарно: это $\mathbf{X}^T\mathbf{y}$, то есть произведение транспонированной матрицы независимых переменных $\mathbf{X}$ на вектор значений зависимой переменной.
- Осталось привести к матричной форме правую часть.

Условия Первого Порядка VI

- Заметьте, что все суммы справа можно расписать как суммы нескольких сумм, похожих на суммы левой части, только вместо y_i будут стоять $x_{i0}...x_{in}$:

$$\sum_{j=0}^{n}\sum_{i=1}^{K} x_{i0}\hat{\beta}_j x_{ij}$$

$$\vdots$$

$$\sum_{j=0}^{n}\sum_{i=1}^{K} x_{in}\hat{\beta}_j x_{ij}$$

Условия Первого Порядка VII

- Формулы на предыдущем слайде приводятся к матричной форме:

$$\mathbf{X}^T\mathbf{X}\hat{\mathbf{b}}$$

- Запишем обе части уравнения:

$$\mathbf{X}^T\mathbf{y} = \mathbf{X}^T\mathbf{X}\hat{\mathbf{b}}$$

- Умножим обе части на обратную матрицу:

$$\hat{\mathbf{b}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$

Условия Гаусса-Маркова I

- **Условие I:** модель корректно определена и имеет только линейные параметры.
- **Условие II:** данные являются случайной выборкой из генеральной совокупности (ошибки не скоррелированы).
- **Условие III:** ни одна из независимых переменных не является константой (за исключением колонки с единицами) и нет прямой линейной связи между независимыми переменными ($x_i = 2x_j$ не позволяется, но x_i^2 можно включать в модель).

Условия Гаусса-Маркова II

- **Условие IV:**
$$\mathbb{E}[u|x_1...x_k] = 0$$

- **Условие V:**
$$\mathbb{V}[u|x_1...x_k] = \sigma^2$$

МНК – BLUE

Теорема I

Если условия I-V верны, МНК-оценка

$$\hat{\mathbf{b}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$

является Best Linear Unbiased Estimator (BLUE), то есть является несмещённой относительно параметров генеральной совокупности и имеет наиболее низкую дисперсию среди всех других несмещённых оценок.

МНК – Несмещённая Оценка I

- Мы берём за стартовую точку ту же самую идею, что и в парной регрессии – фиксируем значения независимых переменных.
- Заметьте, что
$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{u}$$
по определению.
- Подставляя выражение в формулу МНК-оценок, получаем:
$$\hat{\mathbf{b}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T(\mathbf{X}\mathbf{b} + \mathbf{u})$$
- Возьмём условное матожидание:
$$\mathbb{E}[\hat{\mathbf{b}}|\mathbf{X}] = \mathbb{E}[(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T(\mathbf{X}\mathbf{b} + \mathbf{u})|\mathbf{X}]$$

МНК – Несмещённая Оценка II

- Откроем скобки:

$$\mathbb{E}[(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}\mathbf{b}|\mathbf{X}] + \mathbb{E}[(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{u}|\mathbf{X}]$$

- Используя элементарные матричные операции, получаем:

$$\mathbb{E}[\mathbf{b}|\mathbf{X}] + \mathbb{E}[(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{u}|\mathbf{X}]$$

- Используя свойства матожидания, получаем:

$$\mathbb{E}[\mathbf{b}|\mathbf{X}] + (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbb{E}[\mathbf{u}|\mathbf{X}] = \mathbf{b}$$

Дисперсия МНК I

- Для начала введём несколько важных понятий.
- Ковариационная матрица для случайного вектора $\mathbf{z}$.
- Ковариационная матрица $\mathbb{V}[\mathbf{z}]$ для случайного вектора $\mathbf{z}$ определяется как квадратная матрица с дисперсиями случайных величин-компонентов на главной диагонали и ковариациями во всех других ячейках.

Дисперсия МНК II

- Докажем некоторые свойства ковариационных матриц.
- Из определения следует, что $\mathbb{V}[\mathbf{z}]$ as

$$\mathbb{V}[\mathbf{z}] = \mathbb{E}[(\mathbf{z} - \mathbb{E}[\mathbf{z}])(\mathbf{z} - \mathbb{E}[\mathbf{z}])^T]$$

- Первое важное свойство:

$$\mathbb{V}[\mathbf{z}] = \mathbb{E}[\mathbf{z}\mathbf{z}^T] - \mathbb{E}[\mathbf{z}]\mathbb{E}[\mathbf{z}]^T$$

Дисперсия МНК III

- Свойство, необходимое для дальнейших рассуждений о дисперсии МНК:

$$\mathbb{V}[m\mathbf{z}] = m\mathbb{V}[\mathbf{z}]m^T$$

где m – некоторая константная матрица,$\mathbf{z}$ – случайный вектор.

- Доказательство простое:

$$\mathbb{V}[m\mathbf{z}] = \mathbb{E}[m\mathbf{z}(m\mathbf{z})^T] - \mathbb{E}[m\mathbf{z}]\mathbb{E}[m\mathbf{z}]^T$$

$$\mathbb{V}[m\mathbf{z}] = m\mathbb{E}[\mathbf{z}\mathbf{z}^T]m^T - m\mathbb{E}[\mathbf{z}]\mathbb{E}[\mathbf{z}]^T m^T$$

Дисперсия МНК IV

- Вычислим дисперсию МНК:

$$\mathbb{V}[\hat{\mathbf{b}}] = \mathbb{V}[(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}] = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbb{V}[\mathbf{y}]((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T)^T$$

- Поскольку **y** – случайный вектор,

$$\mathbb{V}[\mathbf{y}]$$

является ковариационной матрицей.
- Условие Г-М II гарантирует, что данная ковариационная матрица имеет нули везде, кроме главной диагонали (содержащей дисперсии ошибок); Г-М V гарантирует, что все эти дисперсии равны σ^2.
- Таким образом, имеем:

$$(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbb{V}[\mathbf{y}] = \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$$

Дисперсия МНК V

- В свою очередь:

$$\mathbb{V}[\hat{\mathbf{b}}] = \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T)^T$$

- Применяя свойства транспонирования $(AB)^T = B^T A^T$ и $(A^{-1})^T = (A^T)^{-1}$, получаем:

$$\sigma^2(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T)^T =$$

$$\sigma^2((\mathbf{X}^T\mathbf{X})^{-1})^T = \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}$$

- Неизвестную величину σ^2 можно заменить несмещённой оценкой $\hat{\sigma^2} = \frac{\sum_{i=1}^{K} \hat{u_i^2}}{K-n-1}$, давая нам итоговую несмещённую оценку $\mathbb{V}[\hat{\mathbf{b}}]$.

МНК Имеет Наименьшую Дисперсию I

- Данное доказательство несколько сложнее несмещённости.
- Во-первых, нужно рассмотреть необходимые и достаточные условия несмещённости.
- Рассмотрим оценку

$$\tilde{\mathbf{b}} = \mathbf{P}\mathbf{y}$$

где

$$\mathbf{P} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T + \mathbf{W}$$

$\mathbf{W}$ – ненулевая матрица с $n+1$ строками and K колонками.

МНК Имеет Наименьшую Дисперсию II

- Для упрощения обозначений, на оставшихся слайдах всегда имеется в виду условная вероятность $\mathbb{E}[\cdot|\mathbf{X}]$.
- Возьмём матожидание $\tilde{\mathbf{b}}$:

$$\mathbb{E}[\tilde{\mathbf{b}}] = \mathbb{E}[\mathbf{P}\mathbf{y}] = \mathbb{E}[((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T + \mathbf{W})(\mathbf{X}\mathbf{b} + \mathbf{u})] =$$

$$\mathbb{E}[((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T + \mathbf{W})\mathbf{X}\mathbf{b}] + \mathbb{E}[((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T + \mathbf{W})\mathbf{u}]$$

- Первый компонент суммы – константа, потому что $\mathbf{b}$ – параметры в генеральной совокупности, а $\mathbf{X}$ фиксирована по определению.
- Второй компонент равено нулю:

$$\mathbb{E}[((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T + \mathbf{W})\mathbf{u}] = ((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T + \mathbf{W})\mathbb{E}[\mathbf{u}] = 0$$

в силу условия Г-М IV.

МНК Имеет Наименьшую Дисперсию III

- Получаем

  $$\mathbb{E}[\tilde{\mathbf{b}}] = ((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T + \mathbf{W})\mathbf{X}\mathbf{b} = I\mathbf{b} + \mathbf{W}\mathbf{X}\mathbf{b} = (I + \mathbf{W}\mathbf{X})\mathbf{b}$$

  где I – единичная матрица.

- Таким образом, $\tilde{\mathbf{b}}$ является несмещённой оценкой тогда и только тогда, когда $\mathbf{WX} = 0$.

МНК Имеет Наименьшую Дисперсию IV

- Теперь рассмотрим дисперсию уже знакомой нам оценки

$$\tilde{\mathbf{b}} = \mathbf{P}\mathbf{y}$$

- Перепишем

$$\mathbf{P} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T + \mathbf{W}$$

МНК Имеет Наименьшую Дисперсию V

- Заметьте, что $\mathbf{P}$ – константа:

$$\mathbb{V}[\mathbf{Py}] = \mathbf{P}\mathbb{V}[\mathbf{y}]\mathbf{P}^T$$

- Используя знакомую логику, получаем:

$$\mathbb{V}[\mathbf{Py}] = \sigma^2\mathbf{P}\mathbf{P}^T$$

- Подставим:

$$\mathbf{P}\mathbf{P}^T = ((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T + \mathbf{W}) * ((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T + \mathbf{W})^T$$

МНК Имеет Наименьшую Дисперсию VI

- Применяя свойства транспонирования, получаем:

$$((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T + \mathbf{W}) * (\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1} + \mathbf{W}^T)$$

- Раскроем скобки:

$$\mathbf{I}(\mathbf{X}^T\mathbf{X})^{-1} + (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}^T + \mathbf{W}\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1} + \mathbf{W}\mathbf{W}^T$$

- Очередной раз применив свойства транспонирования, получаем:

$$\mathbf{I}(\mathbf{X}^T\mathbf{X})^{-1} + (\mathbf{X}^T\mathbf{X})^{-1}(\mathbf{W}\mathbf{X})^T + \mathbf{W}\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1} + \mathbf{W}\mathbf{W}^T$$

- Вспомним, что в рамках Теоремы I мы рассматриваем только несмещённые оценки, поэтому $\mathbf{W}\mathbf{X} = 0$, что ведёт к:

$$(\mathbf{X}^T\mathbf{X})^{-1} + \mathbf{W}\mathbf{W}^T$$

МНК Имеет Наименьшую Дисперсию VII

- Но в этом случае

$$\mathbb{V}[\tilde{\mathbf{b}}] = \sigma^2((\mathbf{X}^T\mathbf{X})^{-1} + \mathbf{W}\mathbf{W}^T) = \sigma^2(\mathbf{X}^T\mathbf{X})^{-1} + \sigma^2\mathbf{W}\mathbf{W}^T =$$

$$\mathbb{V}[\hat{\mathbf{b}}] + \sigma^2\mathbf{W}\mathbf{W}^T$$

- Поскольку $\mathbf{W}\mathbf{W}^T$ – положительно полуопределённая матрица, дисперсия любой несмещённой оценки как минимум не меньше, чем дисперсия МНК! (см. Приложение)

Приложение

Сравнение Ковариационных Матриц I

- Мы определяем матрицу A как меньшую, чем B, если матрица $B - A$ является положительной полуопределённой.
- Легко доказать, что для любой матрицы $\mathbf{A}$ действительных чисел, $\mathbf{A}\mathbf{A}^T$ является положительной полуопределённой, т.е. $\mathbf{r}^T\mathbf{A}\mathbf{A}^T\mathbf{r} \geq 0$ для всех $\mathbf{r}$ векторов действительных чисел.

Сравнение Ковариационных Матриц II

- Мы имеем
$$(\mathbf{A}^T\mathbf{r})^T = \mathbf{r}^T\mathbf{A}$$
- Задав $\mathbf{y} = \mathbf{A}^T\mathbf{r}$, получаем:
$$\mathbf{r}^T\mathbf{A}\mathbf{A}^T\mathbf{r} = \mathbf{y}^T\mathbf{y} = \sum_{i=1}^{K} y_i^2 \geq 0$$
где K – количество строк в $\mathbf{A}$.
- Из этого следует, что
$$\mathbf{W}\mathbf{W}^T$$
положительная полуопределённая.

Несмещённая оценка σ^2

- Оценка
$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{K} \hat{u}_i^2}{K - n - 1}$$
является несмещённой оценкой дисперсии ошибок σ^2 в генеральной совокупности.
- След матрицы – сумма элементов главной диагонали.
- Необходимый результат для доказательства – циклическое свойство следа матрицы:
$$tr(\mathbf{AB}) = tr(\mathbf{BA})$$
то есть след матрицы одинаков для произведения двух матриц, если **A** имеет размерность n x m, а **B** имеет размерность m x n (иными словами, если операция умножения определена в левой и правой части тождества).

Несмещённая оценка σ^2

- Доказать несмещённость – значит доказать следующий результат:

$$\mathbb{E}[\hat{\sigma^2}] = \sigma^2$$

- Знаменатель $K - n - 1$ является константой и выносится за знак матожидания; таким образом, нужно разобрать матожидание суммы квадратов остатков.
- Запишем сумму в векторной форме:

$$\mathbb{E}[\hat{\mathbf{u}}^T \hat{\mathbf{u}}] = \sigma^2$$

Несмещённая оценка σ^2

- Выразим остатки, используя МНК-оценки:

$$\mathbb{E}[(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}})^T(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}})]$$

- Далее выразим $\mathbf{y}$ через ошибки и параметры генеральной совокупности:

$$\mathbb{E}[(\mathbf{X}\mathbf{b} + \mathbf{u} - \mathbf{X}\hat{\mathbf{b}})^T(\mathbf{X}\mathbf{b} + \mathbf{u} - \mathbf{X}\hat{\mathbf{b}})]$$

- Подставим формулу МНК-оценок и выразим $\mathbf{y}$ через ошибки и параметры генеральной совокупности:

$$\mathbf{X}\mathbf{b} + \mathbf{u} - \mathbf{X}\hat{\mathbf{b}} =$$

$$\mathbf{X}\mathbf{b} + \mathbf{u} - \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} =$$

$$\mathbf{X}\mathbf{b} + \mathbf{u} - \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T(\mathbf{X}\mathbf{b} + \mathbf{u}) =$$

Несмещённая оценка σ^2

- $$\mathbf{Xb} + \mathbf{u} - \mathbf{Xb} - \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{u} = (\mathbf{I} - \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T)\mathbf{u}$$

  где $\mathbf{I}$ – единичная матрица размерности KxK.
- Подставим результат обратно в матожидание:

  $$\mathbb{E}[((\mathbf{I} - \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T)\mathbf{u})^T((\mathbf{I} - \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T)\mathbf{u})]$$
- Применив свойства траснпонирования и раскрыв все скобки, получаем:

  $$\mathbb{E}[\mathbf{u}^T(\mathbf{I} - \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T)\mathbf{u}]$$

Несмещённая оценка σ^2

- Раскрыв скобки и применив линейность матожидания получаем:

$$\mathbb{E}[\mathbf{u}^T \mathbf{I} \mathbf{u}] - \mathbb{E}[\mathbf{u}^T \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{u}]$$

- Первый элемент равен $K\sigma^2$ (может быть представлен как сумма квадратов ошибок от одного до K, а матожидания квадратов ошибок равны σ^2 по условиям Г-М IV и V).
- Обозначим матрицу $\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ как $\mathbf{G}$ с произвольным элементом g_{ij}.
- Заметим, что выражение $\mathbf{u}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ – это вектор размерностью K, каждый элемент j которого может быть записан как $S_j = \sum_{i=1}^{K} \mu_i g_{ij}$.

Несмещённая оценка σ^2

- Финальная операция умножения на **u** эквивалента следующей сумме:

$$\sum_{j=1}^{K} \mu_j S_j = \sum_{j=1}^{K} \mu_j \sum_{i=1}^{K} \mu_i g_{ij}$$

- Применим матожидание, условия Г-М II, IV и V:

$$\mathbb{E}[\sum_{j=1}^{K} \mu_j \sum_{i=1}^{K} \mu_i g_{ij}] = \sigma^2 \sum_{j=1}^{K} g_{jj} = \sigma^2 tr(\mathbf{G})$$

- Используя циклическое свойство следа, имеем:

$$tr(\mathbf{G}) = tr(\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T) = tr(\mathbf{X}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}) = tr(\mathbf{I}_{n+1})$$

  где $\mathbf{I}_{n+1}$ – единичная матрица размерностью $(n+1)x(n+1)$, n – количество независимых переменных, а 1 – интерсепт.

Несмещённая оценка σ^2

- След единичной матрицы $\mathbf{I}_{n+1}$ равен $n+1$.
- Итого получаем $K\sigma^2 - \sigma^2(n+1) = \sigma^2(K-n-1)$.
- Умножим на знаменатель $K-n-1$, получаем σ^2.