Вступительное слово. *Эта лекция может оказаться трудной: сплошная теория, матрицы и никакой конкретики. Если математическая часть тяжёлая, пропустите выводы и доказательства. Главное — понять:*
а) какие предпосылки делаются в классической модели регрессии,
б) в чём смысл теоремы Гаусса-Маркова.

В конце лекции — краткая табличка с изложением основных фактов. Если сможете разобрать её содержимое, это уже замечательно.

Классическая линейная нормальная регрессионная модель и оценивание её параметров, часть I

Краткое содержание предыдущих лекций.
Имеется набор данных $(Y_1, X_{2,1}, X_{3,1}, \ldots, X_{k,1}), \ldots, (Y_n, X_{2,n}, X_{3,n}, \ldots, X_{k,n})$, по которому оценивается зависимость объясняемой переменной Y от объясняющих переменных (регрессоров) $X_2, \ldots, X_k$ вида $Y_i = \beta_1 + \beta_2 X_{2,i} + \ldots + \beta_k X_{k,i} + \epsilon_i,\ i = 1, \ldots, n.$.

Запись $Y_i = \beta_1 + \beta_2 X_{2,i} + \ldots + \beta_k X_{k,i} + \epsilon_i$ называется уравнением регрессии.
Часто удобно использовать матричную запись.

Введём вектор значений объясняемой переменной $y = \begin{pmatrix} Y_1 \\ Y_2 \\ \ldots \\ Y_n \end{pmatrix}$,

матрицу регрессоров $X = \begin{pmatrix} 1 & X_{2,1} & X_{3,1} & \ldots & X_{k,1} \\ 1 & X_{2,2} & X_{3,2} & \ldots & X_{k,2} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 1 & X_{2,n} & X_{3,n} & \ldots & X_{k,n} \end{pmatrix}$,

вектор коэффициентов $\beta = \begin{pmatrix} \beta_1 \\ \ldots \\ \beta_k \end{pmatrix}$,

вектор ошибок регрессии $\epsilon = \begin{pmatrix} \epsilon_1 \\ \ldots \\ \epsilon_n \end{pmatrix}$.

Уравнение регрессии в матричной форме записывается так: $y = X\beta + \epsilon$.

Оценками метода наименьших квадратов (МНК) для коэффициентов $\beta_1, \ldots, \beta_k$ называются значения $\hat{\beta}_1, \ldots, \hat{\beta}_k$, являющиеся решением задачи

$$\sum_{i=1}^{n} \left(Y_i - \beta_1 - \beta_2 X_{2,i} - \ldots - \beta_k X_{k,i}\right)^2 \to \min_{\beta_1, \ldots, \beta_k}.$$

Задача о наименьших квадратах в матричной форме:

$$(y - X\beta)'(y - X\beta) \to \min_{\beta}.$$

Задача эта имеет аналитическое решение. Итоговая формула для оценки МНК:

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

Классическая линейная нормальная регрессионная модель (КЛНРМ).
Теперь внесём вероятностное содержание в задачу исследования зависимости Y от $X_2, \ldots, X_k$.

Предпосылка 1. Случайные величины $Y_1, \ldots, Y_n$ связаны с объясняющими переменными соотношением $Y_i = \beta_1 + \beta_2 X_{2,i} + \ldots + \beta_k X_{k,i} + \epsilon_i,\ i = 1, \ldots, n$.

В матричной записи: $y = X\beta + \epsilon$.

Комментарий. Именно поэтому модель называется линейной, однако здесь не говорится, что зависимость между признаком Y и регрессорами должна быть линейной. Линейно должны входить коэффициенты $\beta_1, \ldots, \beta_k$. Например, модели $Y_i = \beta_1 + \beta_2 X_i + \beta_3 X_i^2 + \epsilon_i$ и $\ln Y_i = \beta_1 + \beta_2 \ln X_i + \epsilon_i$ — линейные по коэффициентам, а модель $Y_i = \beta_1 + X_i^{\beta_2} + \epsilon_i$ — нет, её коэффициенты не получится оценить обычным методом наименьших квадратов (для этого есть другая версия — нелинейный МНК).

Предпосылка 2. Объясняющие переменные $X_{2,i}, \ldots, X_{k,i}$ детерминированы (не случайны), между величинами $X_2, \ldots, X_k$ нет линейной зависимости.

В матричной записи: матрица регрессоров X детерминирована, $rank(X) = k$, где k — число столбцов матрицы X и коэффициентов $\beta_1, \ldots, \beta_k$.

Комментарий 1. Почему объясняющие переменные должны быть детерминированы, а не случайны? Это разумно в тех случаях, когда их назначает сам исследователь (он проводит контролируемый эксперимент). Но в действительности регрессионный анализ со случайными объясняющими переменными проходит так же, как и с детерминированными, а математический аппарат посложнее. Я не хочу сейчас вводить эти сложности.

Комментарий 2. Ранг матрицы — это максимальное число линейно независимых столбцов. Число столбцов матрицы X равно k , так что условие $rank(X) = k$ означает линейную независимость всех объясняющих переменных (всех столбцов матрицы). Нарушение этой предпосылки — случай $rank(X) < k$, линейная зависимость между какими-либо регрессорами — называется *строгой мультиколлинеарностью*. В этом случае оценка МНК $\hat{\beta} = (X'X)^{-1} X'y$ не существует, потому что матрица $X'X$ необратима.

Попытка пояснить, в чём проблема. Представьте, что вам захотелось оценить регрессию веса человека в килограммах Y на его рост в сантиметрах X и на его рост в метрах Z . Однозначно оценить коэффициенты уравнения $Y_i = \beta_1 + \beta_2 X_i + \beta_3 Z_i + \epsilon_i$ невозможно, потому что нельзя разделить зависимость веса от роста на зависимость Y от X и зависимость Y от Z .

Строгая мультиколлинеарность возникает редко, чаще по ошибке исследователя, который не замечает, что включил в модель линейно зависимые переменные. В прошлой лекции был рассмотрен пример — регрессия зарплаты на характеристики работника, включающие уровень образования. По образованию люди делились на *пять* категорий, а в модель включались дамми-переменные для любых *четырёх* категорий. Типичная студенческая ошибка — включить пять переменных, по одной на каждую категорию. И тут как раз появляется строгая мультиколлинеарность: эти переменные линейно зависимы, ведь в сумме они обязательно дают единицу.

Чаще встречается нестрогая мультиколлинеарность — зависимость между регрессорами, близкая к линейной. В этом случае оценка МНК существует, но может оказаться неудовлетворительной (например, она будет сильно меняться при небольших изменениях в данных). Подробнее об этом говорить сейчас не будем.

Предпосылка 3. Ошибки регрессии $\epsilon_1, \ldots, \epsilon_n$ — случайные величины, такие что:
3a. $E(\epsilon_i) = 0$, $i = 1, \ldots, n$;
3b. $D(\epsilon_i) = \sigma_\epsilon^2$, $Cov(\epsilon_i, \epsilon_j) = 0$, $i, j = 1, \ldots, n,\ i \neq j$;
3c. $\epsilon_1, \ldots, \epsilon_n$ имеют совместное нормальное распределение.

В матричной записи: ϵ — случайный вектор, такой что:

3a. $E(\epsilon)=0$ (тут 0 — это вектор из нулей);

3b. $V(\epsilon)=\sigma_\epsilon^2 I_n=\begin{pmatrix} \sigma_\epsilon^2 & 0 & \dots & 0 \\ 0 & \sigma_\epsilon^2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sigma_\epsilon^2 \end{pmatrix}$.

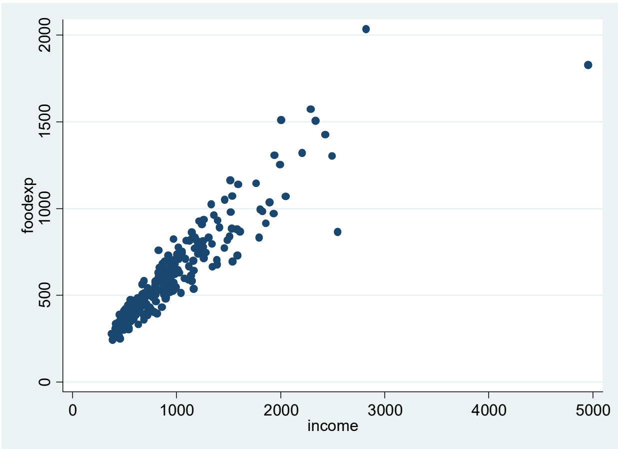
3c. $\epsilon \sim N(0, \sigma_\epsilon^2 I_n)$.

Прокомментирую все пункты по порядку.

Пункт **3a**. $E(\epsilon_i)=0$. Это формальность. В действительности та же самая модель получается, например, если уравнение регрессии записать без свободного члена: $Y_i=\beta_2 X_{2,i}+\dots+\beta_k X_{k,i}+\epsilon_i$, а удалённый оттуда коэффициент β_1 запихнуть в случайную ошибку: $E(\epsilon_i)=\beta_1$. То есть между математическим ожиданием ошибки и свободным членом разницы никакой нет, это одно и то же. Но нам удобнее будет явно включать в уравнение коэффициент β_1 , а математические ожидание ошибки назначить равным нулю.

Пункт **3b**, часть первая: $D(\epsilon_i)=\sigma_\epsilon^2$. У этой предпосылки есть причудливое название: *гомоскедастичность* («равноразбросие»). Смысл в том, что дисперсия случайных ошибок во всех наблюдениях одна и та же. Нарушение этой предпосылки, $D(\epsilon_i)\neq const$, называют *гетероскедастичностью*.

Ниже приведён график рассеяния для данных о 235 бельгийских семьях[1]. По горизонтали откладывается годовой доход семьи в бельгийских франках, по вертикали — затраты на питание:



Бедные семьи вынуждены большую часть дохода тратить на питание, а богатые могут выбирать, сколько им проесть, а сколько оставить на другие нужды, поэтому с ростом дохода увеличивается разброс затрат на еду. Если попытаться представить связь затрат с доходами в виде регрессионной модели, тут и возникнет гетероскедастичность.

А ещё на графике виднеется нетипичное наблюдение справа вверху: сверхбогачи, которые тратятся на еду будто обычные богачи. Сейчас оно отношения к делу не имеет, просто повод напомнить, что такие наблюдения могут оттягивать на себя линию регрессии, существенно влияя на результаты анализа.

Пункт **3b**, часть вторая: $Cov(\epsilon_i, \epsilon_j)=0, i\neq j$. Случайные ошибки в разных наблюдениях некоррелированы (а если выполняется **3c**, то и независимы). Нарушение этой предпосылки называется *автокорреляцией* случайной ошибки (или серийной корреляцией). Автокорреляция часто возникает в анализе временных рядов: отклонения объясняемой переменной Y от линии регрессии в близкие моменты времени часто зависимы — многим процессам свойственна инертность.

1 Файл «engel.gdt» из примеров, поставляемых вместе с программой gretl.

Пункт **3c**, $\epsilon_1, \dots, \epsilon_n$ имеют совместное нормальное распределение. Сегодня эта предпосылка нам не пригодится, а в следующих лекциях она будет использоваться при построении доверительных интервалов и статистических критериев. По-моему, нарушения этой предпосылки редко представляют значительную опасность, но, кажется, живут на свете и те, кто поспорил бы.

Предпосылки на этом заканчиваются. Дальше пойдёт то, что всегда идёт в математической статистике: будем оценивать параметры модели и проверять связанные с ними гипотезы.

Обратите внимание: параметры КЛНРМ — это не только коэффициенты уравнения регрессии $\beta_1, \dots, \beta_k$, это ещё и дисперсия случайной составляющей σ_ϵ^2. Пока что мы никак её не использовали — до сих пор ошибка регрессии вообще не рассматривалась как случайная величина. Она используется при расчёте доверительных интервалов и проверке гипотез.

Но сегодня обойдёмся коэффициентами $\beta_1, \dots, \beta_k$. Как их оценить? Наверное, методом наименьших квадратов — не зря же мы его проходили. И правда, не зря — оценки МНК обладают чудесными свойствами, которые утверждаются теоремой Гаусса–Маркова.

Теорема Гаусса–Маркова. Пусть выполнены предпосылки (1)–(3b) классической линейной нормальной регрессионной модели. Тогда оценка метода наименьших квадратов $\hat{\beta} = (X'X)^{-1}X'y$ есть эффективная в классе линейных (по вектору объясняемой переменной y) несмещённых оценок вектора коэффициентов β.

Подробнее о том, что это значит. Итак, оценка МНК 1) линейная по y, 2) несмещённая, 3) эффективная среди линейных несмещённых оценок. Разберёмся с этим по очереди. А для начала заметим, что мы впервые в курсе сталкиваемся с оценкой-вектором.

Линейность. Здесь имеется в виду, что оценка есть линейная функция от y, то есть представима в виде $\hat{\beta} = Ay$, где A — произвольная матрица подходящего размера. В случае МНК $A = (X'X)^{-1}X'$.
Сама по себе линейность не плоха и не хороша, если не считать того, что с линейными оценками обычно проще иметь дело.

Несмещённость. Это означает, что $E(\hat{\beta}) = \beta$. Каждая из случайных величин $\hat{\beta}_j$ есть несмещённая оценка соответствующего коэффициента β_j.

Эффективность. Пусть $\tilde{\beta}$ — линейная несмещённая оценка для β, то есть она представима в виде $\tilde{\beta} = Ay$ и $E(\tilde{\beta}) = \beta$. Тогда матрица $V(\tilde{\beta}) - V(\hat{\beta})$ неотрицательно определена.

Что это значит? Для начала вспомним, что на главной диагонали неотрицательно определённой матрицы обязательно стоят неотрицательные числа. А главные диагонали матриц $V(\tilde{\beta})$ и $V(\hat{\beta})$ содержат дисперсии оценок коэффициентов. Значит, дисперсии всех оценок из вектора $\tilde{\beta}$ не меньше дисперсий соответствующих оценок МНК.

Итак, оценки МНК имеют дисперсии не больше, чем у любых других линейных несмещённых оценок тех же коэффициентов по той же выборке.

Но это ещё не всё. Рассмотрим произвольную линейную комбинацию коэффициентов $c'\beta = c_1\beta_1 + \dots + c_k\beta_k$ и её оценки $c'\tilde{\beta}$ и $c'\hat{\beta}$ — они будут несмещёнными. Вот их дисперсии: $D(c'\tilde{\beta}) = c'V(\tilde{\beta})c$ и $D(c'\hat{\beta}) = c'V(\hat{\beta})c$. Разница дисперсий:

$$D(c'\tilde{\beta}) - D(c'\hat{\beta}) = c'V(\tilde{\beta})c - c'V(\hat{\beta})c = c'(V(\tilde{\beta}) - V(\hat{\beta}))c.$$

Вспомним определение: матрица A неотрицательно определена, если для любого вектора c подходящей размерности выполнено неравенство $c' Ac \geq 0$. А раз матрица $V(\tilde{\beta}) - V(\hat{\beta})$ неотрицательно определена, то $D(c'\tilde{\beta}) - D(c'\hat{\beta}) \geq 0$.

Итак, мораль: оценка МНК для любой линейной комбинации коэффициентов регрессии имеет дисперсию не больше, чем у любой другой линейной несмещённой оценки.

Конечно, свойства несмещённости и эффективности выполняются *для любых* значений вектора β.

В англоязычной литературе распространена аббревиатура BLUE (Best Linear Unbiased Estimator — лучшая линейная несмещённая оценка). Так вот, теорема Гаусса–Маркова утверждает, что оценки МНК для коэффициентов классической линейной регрессионной модели — BLUE.

О чём не говорит теорема Гаусса–Маркова. Она не говорит, что нет ничего лучше МНК. Ведь оценки могут быть смещёнными и нелинейными — можно попробовать найти что-то получше среди них. В некоторых случаях это удаётся.

Чего не требуют предпосылки теоремы. Они не включают предпосылку о нормальности случайной ошибки. Даже если она нарушается, оценки МНК всё равно BLUE.

Доказательство теоремы Гаусса–Маркова проводится в три этапа:
1) доказательство несмещённости;
2) вывод ковариационной матрицы оценки МНК;
3) доказательство эффективности.

Я изложу первые два этапа — они короткие и понять их легко. Они опираются на свойства вектора математических ожиданий и ковариационной матрицы (см. презентацию к лекции):

$$E(Ay+b) = AE(y)+b,$$
$$V(Ay+b) = AV(y)A'.$$

Этап 1. Докажем несмещённость.

$$E(\hat{\beta}) = E[(X'X)^{-1}X'y] = (X'X)^{-1}X'E(y).$$

Предпосылка 1 КЛНРМ говорит, что $y = X\beta + \epsilon$, поэтому

$$E(y) = E(X\beta+\epsilon) = E(X\beta)+E(\epsilon) = X\beta + E(\epsilon).$$

По предпосылке 3а $E(\epsilon)=0$, так что $E(y)=X\beta$. Значит,

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

Несмещённость доказана.

Этап 2. Теперь найдём ковариационную матрицу оценки.

$$V(\hat{\beta}) = V[(X'X)^{-1}X'y].$$

Чтобы вынести детерминированный множитель $(X'X)^{-1}X'$ за ковариационную матрицу, его нужно транспонировать. Тут пригодятся свойства транспонирования и обращения.

Напоминалка: некоторые свойства транспонирования и обращения матриц.

$$(A')'=A, \quad (AB)'=B'A', \quad (A^{-1})'=(A')^{-1}.$$

Из этих свойств следует, что матрица $X'X$ симметрична:

$$(X'X)'=X'(X')'=X'X.$$

Поэтому симметрична и $(X'X)^{-1}$:

$$((X'X)^{-1})'=((X'X)')^{-1}=(X'X)^{-1}.$$

Наконец, транспонируем $(X'X)^{-1}X'$:

$$[(X'X)^{-1}X']'=(X')'[(X'X)^{-1}]'=X(X'X)^{-1}.$$

Теперь уже можно вернуться к оценке МНК и её ковариационной матрице:

$$V(\hat{\beta})=V[(X'X)^{-1}X'y]=(X'X)^{-1}X'V(y)X(X'X)^{-1}.$$

Из предпосылок 1, 2 и 3b КЛНРМ следует, что $V(y)=V(X\beta+\epsilon)=V(\epsilon)=\sigma_\epsilon^2 I_n$. Значит,

$$V(\hat{\beta})=(X'X)^{-1}X'V(y)X(X'X)^{-1}=(X'X)^{-1}X'\sigma_\epsilon^2 I_n X(X'X)^{-1}.$$

Матрицу I_n в записи можно опустить — она на то и единичная, что никак не меняет произведение. Множитель σ_ϵ^2 — число, а не матрица, поэтому его можно переставить в любое место произведения. Например, в начало. Тогда получится вот что:

$$V(\hat{\beta})=(X'X)^{-1}X'\sigma_\epsilon^2 I_n X(X'X)^{-1}=\sigma_\epsilon^2(X'X)^{-1}X'X(X'X)^{-1}=\sigma_\epsilon^2(X'X)^{-1}.$$

Красотища. Это и есть ответ. Здесь второй этап доказательства заканчивается.

Третью часть — доказательство эффективности — вы можете разобрать самостоятельно по книге Я.Р. Магнуса, П.К. Катышева и А.А. Пересецкого «Эконометрика. Начальный курс». Хотя она чуть посложнее и подлиннее первых двух, ничего сверхъестественного в доказательстве нет. Некоторые преподаватели даже считают, что доказательство теоремы Гаусса–Маркова — это такая задача, которую можно дать самостоятельно решить на семинаре.

Основная часть лекции на этом подходит к концу. Остаётся подвести итог. Он на следующей странице.

Предпосылки классической линейной нормальной регрессионной модели
1. $y = X\beta + \epsilon$. 2. Матрица регрессоров X детерминирована, $rank(X) = k$, где k — число столбцов матрицы X. 3. ϵ — случайный вектор, такой что: 3a. $E(\epsilon) = 0$; 3b. $V(\epsilon) = \sigma_\epsilon^2 I_n = \begin{pmatrix} \sigma_\epsilon^2 & 0 & \dots & 0 \\ 0 & \sigma_\epsilon^2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sigma_\epsilon^2 \end{pmatrix}$. 3c. $\epsilon \sim N(0, \sigma_\epsilon^2 I_n)$.
Свойства оценки МНК $\hat{\beta} = (X'X)^{-1} X'y$ **(теорема Гаусса–Маркова)**
Если выполняются предпосылки (1)–(3b) КЛНРМ, то 1) $E(\hat{\beta}) = \beta$; 2) $V(\hat{\beta}) = \sigma_\epsilon^2 (X'X)^{-1}$; 3) для любой другой линейной несмещённой оценки $\tilde{\beta}$ верно, что $V(\tilde{\beta}) - V(\hat{\beta}) \geq 0$. В частности, $D(\tilde{\beta}_j) \geq D(\hat{\beta}_j)$.

Просмотрите внимательно эту таблицу. Если вы понимаете, о чём говорится в каждой строчке — это прекрасно, даже если в доказательстве вы пока не разобрались.

В следующий раз:

- оценивание дисперсии случайной ошибки и ковариационной матрицы оценок МНК;
- скорректированный коэффициент детерминации;
- теорема Фишера для КЛНРМ.