Эконометрика — А. А. Мамонтов

Последнее обновление имело быть 28 апреля 2012 г. в 05:19:39.

1 Семинар 1

Доугерти не так хорош, зато переведён. Первая глава звучит издевательски. Хороший учебник Вербика. Магнус, Катышев, Пересецкий — отодвинуть в сторону (сложен).

Не стесняйтесь спрашивать. На семинаре вы должны из Мамонтова выжать максимум. До Нового года Мамонтов будет до 21 вечера, но только если сидеть, а то исчезнет тихо, чтобы по дороге ещё никого не встретить. Бывает на Покре только по понедельникам и вторникам. Развлекательная процедура. Называйте это тестиками!

2 Семинар 2

Нельзя определять случайность через вероятность. А то получается, как у Мюнхгаузена. Были попытки через Ω всё сделать. Но будем различать два понятия: значение случайного события (a_k) и результат опыта (что получаем, когда бросаем, a_i). И a_i получаются из a_k. Нас интересует результат опыта. Что мы называем случайным? Не надо пугать народ Ω! «Репрезентативная выборка» командным голосом! Случайное событие — счётное множество. И в жизни мы никогда не встречаем непрерывных величин. В реальности всё конечно и не непрерывно. Может ли быть у случайного события конечное число исходов? Нет! Если нет бесконечного процесса, значит нет этой загадки. Проблема случайности — мы не можем предсказать каждый следующий исход, зная все предыдущие. И если это этому удовлетворяет, то мы на это смотрим как на случайное. Мы называем случайным, если не знаем закона. Чем величина от события отличается? Это число. Надо поставить любую числовую функцию на событие — появится случайная величина. Можно случайную величину определять как бесконечное число исходов и невозможность узнать следующее при всех известных предыдущих. Как проверить поток событий на случайность? Надо смотреть корреляции ближнего соседа. Корелляция через одного, двух, трёх. Если корреляция всегда 0 (мы проверяем гипотезу) ни уровне статистики, то случайность. Если рассматривать десятичное представление трансцендентных чисел (π, e), то он случаен. Но так ли это? Например, нам подсунули сотый знак π^2… IBM прикладывала софт для интегрирования и дифференцирования, и в 1960-е случился прокол. Люди начали пользоваться датчиком, и возникла проблема: на тройках вылезла корреляция. И кто-то доказал, что есть корреляция порядка 0,6. Вот вам и случайность! Монте-Карло никто не отменял. Было гораздо проще для разных начальных значений покидать разные числа. И ядерные реакторы обсчитывались на Монте-Карло. Для фондового рынка так же всё симулируется. Если происходит выброс, то мы занимаем место богов и повторяем случайность. Какие игры могут быть для богов? Все исходы уже известны, кому надо лезть под стол и петухом орать. Жизнь имеет смысл только потому, что мы процесса не знаем.

Самое первое свойство — матожидание. Что это? Если орёл или решка, то матожидание — это 0,5, или «ребро»? Не-е. Матожидание — это среднее по совокупности. Вредно запоминать матожидание как взвешенное по вероятности: $\mathbb{E}(a) = \sum a_k p_k$ — так как теряется смысл. Матожидание — это среднее! Это просто! Матожидание: $\mathbb{E}(a) = \operatorname*{plim}_{N\to\infty} \frac{\sum_i a_i}{N}$. Мощность множества выбросов меньше, поэтому среднее по генеральной совокупности работает. $\mathbb{E}(a) = \operatorname*{plim}_{N\to\infty} \frac{\sum a_k n_k}{N}$. $\frac{n_k}{N}$ — частота, которая в пределе становится вероятностью: $\operatorname{plim} \sum a_k f_k$.

Дисперсия: $\mathbb{E}\big(X - \mathbb{E}(X)\big)^2 = \operatorname{Var}(x)$ — случайная величина. Она показывает, как сильно величина X отклоняется от своего среднего. Обозначать будем $\operatorname{Var}(X)$. В Совке было $\mathsf{M}X$, $\mathsf{D}X$. Всё стало выпадать, и западные люди нас не поймут: дисперсия — это variance. Наука в Средневековье тяготела к тому, чтобы другие её не понимали. Но в науке сохраняется любовь к написанию наукообразные терминов для запугивания публики. А американцы для обозначения терминов берут бытовые слова: *variance* — изменяемость, *expectation* — ожидание. Слово «математическое» — для того, чтобы увести в сторону и придать ауру. Иногда даже матожидание пишется через μ_x, дисперсия — через σ_x^2. Для нас заглавные греческие буквы (M, P) близки, а для американцев это такая же экзотика, как грузинский алфавит. Идеи теорвера строятся на том, что у нас есть генератор одинаковости. Изначально мы ничего не делаем и знаем все вероятности. Если бы монетка была одинаковая, то был бы орёл. Можно по монетке молотком постучать, поэтому монетка априорно асимметрична, да она и так асимметрична!

Люди \ Пальцев	4	5	6
A	2	3	1
B	4	2	1

Пусть мы собрали людей с количеством пальцев на руках. $\mathbb{P}(4; A) = \frac{2}{13}$. А условное вероятность — это когда мы оставляем только то, что произошло в условии. Самое простое — перенормировать и поделить на эти $\frac{6}{13}$: $\mathbb{P}(4 \mid A) = \frac{\mathbb{P}(4;A)}{\mathbb{P}(A)}$. Другой пример: $\mathbb{P}(4 \mid X < 6) = \frac{6/13}{11/13}$.

«Но Борис Борисович был прав, потому что когда-то и я его тоже мучил».

Формула Байеса не позволяет вернуться на исходник, и тогда понимание исчезает. $\mathbb{E}(X \mid A) = \sum X_k \cdot \mathbb{P}(X \mid A)$. В дисперсии внимание: мы всё в условие должны запихать! $\operatorname{Var}(X \mid A) = \sum \big(X_k - \mathbb{E}(X \mid A)\big)^2 \cdot \mathbb{P}(X \mid A)$.

Когда у нас признаки независимы? Если X и A независимы, то $\mathbb{P}(X \mid A) = \mathbb{P}(X)$. Для совместной вероятности: $\mathbb{P}(X; A) = \mathbb{P}(X) \cdot \mathbb{P}(A)$.

Ковариация — это обобщение дисперсии: $\operatorname{Cov}(X; Y) = \mathbb{E}\Big(\big(X - \mathbb{E}(X)\big)\big(Y - \mathbb{E}(Y)\big)\Big) = \sigma^2_{XY}$. Из бытовых представлений: подбрасывая две монетки, мы знаем, что они независимы.

	A (0,5)	B (0,25)	C (0,25)
B_1	1/6		

Если строчки и столбцы пропорциональные, то независимость есть. Мы перешли в некоторое условие при зависимых данных, и появляется независимость.

У multiple choice есть забавная предыстория. Контрольная родилась из вопросов, которые задаются на сдаче первого задания! А потом надо было просто проверить. С Канторовичем 7 лет назад решили вернуть нас на фундамент. Вот понадобилась контрольная: что в голове осталась? И в multiple choice взяли ответы студентов. Иногда отвечали правильно, иногда неправильно. Некая генетическая память поколений о неправильном передаётся из года в год. $\mathbb{E}(\alpha X - \beta Y) = \alpha\mathbb{E}(X) - \beta\mathbb{E}(Y)$ — это применимо для всех, даже для зависимых, которым на свою зависимость наплевать. $\operatorname{Var}(\alpha X - \beta Y) = \alpha^2 \operatorname{Var}(X) + \beta^2 \operatorname{Var}(Y) - 2\alpha\beta \operatorname{Cov}(X; Y)$. Можно прибавить ноль, ничего от этого не изменится.

Чем статистика отличается от теорвера? В теорвере даётся априорная вероятность, функция распределения. Генеральной совокупностью называются все наблюдения и исходы. А функция распределения называется генеральной совокупностью, потому что информационно мы всё знаем. А в статистике случайность результирующая. Мы A не знаем, и для оценки этой величины используются X. Истинное значение подменяем оценкой. Оценка — это $\hat{A}$. В реальности есть какое-то реальное соотношение рубля и доллара в этих ваших экономиках, а оценка у брокера висит. В 1998 году каждый день твердили: «доллар стоит 6 рублей» (оригинальная оценка). Первое свойство — несмещённость: матожидание оценки совпадает с истинной величиной. $\mathbb{E}(X) = A$. Эффективность — когда поменьше дисперсия. Что лучше — эффективность или несмещённость? Критерий «лучше» вы сказали? И мы начинаем бодро обсуждать! Определяйте поле обсуждения! Вы как морская рыба, которая сразу крючок «ам». Лучше — это ближе попасть. Если делаем одно измерение, то лучше нам поближе попасть сразу к среднему. А при бесконечном измерении несмещённость лучше. Даже бесполезный предмет в краткосрочке хорош, в долгосрочке плох. Состоятельность — это когда у нашей величины есть ещё параметр X_n, причём $\mathbb{E}(X_n) = a_n$. $\lim\limits_{n\to\infty} a_n = a$, $\lim\limits_{n\to\infty} \operatorname{Var}(x_n) = 0$.

Что такая выборка? Выборка, которая не часть генеральной совокупности. Кидаем кубик, выбираем единицы. Это часть генеральной совокупности, но не выборка, даже если одна двойка затесалась. Выборка — это многомерная, или n-мерная случайная величина, состоящия из n НОРСВ случайных величин. Берём $\{X\}$, берём $X_1, X_2, \ldots, X_n$, следующей строкой берём $X_{n_1}, \ldots, X_{2n}$, $X_{2n+1}, \ldots, X_{3n}$ — следующего исхода мы не знаем. Потом берём первый столбец. Случайные величины независимы! И предугадать мы не знаем. И выбираем мы их одинаково, поэтому НОРСВ. В выборке первая строчка — это первый исход многомерной случайной величины. Если выпал орёл, то это не случайная величина, а исход. И эта выборка не бывает нерепрезентативной. Можно ряды по-разному писать. Закон отбора выборки не должен коррелировать с $\{X\}$, то это тоже выборка.

3 Семинар 3

Рациональных чисел куда меньше, чем иррациональных (вероятность 0), но мы просто берём то, что можем взять, и видим только это. Вокруг нас поэтому почти всегда нормальное распределение. Классический апофигей — это после Байеса давать Пуассона и ЦПТ. ЦПТ: мир нормален. Пусть даже величины чуть зависимы. Сумма случайных величин стремится к нормальному распределению. Прозрение наступило у Мамонтова позже, в 80-х, когда надо было оценить нормально распределённую случайную величину. Предлагалось 2 способа получить нормальное из равномерного. Идеи две: обернуть функцию, устроить интерполяцию и занять рабочий ресурс. Второй способ, быстрый: до безобразия примитивно. Берётся сумма 20 значений псевдослучайной величины. На этих 20 величинах точность получения — это 1 %. Это толщина оси на графике (с точностью до линии мы получаем нормальное распределение). Стоит только начать играть монеткой, как результат уже будет нормален. Плотность для нормального распределения — это $p(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$. Распределение Лапласа — это без двойки. Кроме того, $\mathbb{E}(z^2) = 1$. Появление хи-квадрата идёт через туалет, но запомните: от -1 до 1 лежит $\frac{2}{3}$, за двойками лежит 5 %, за тройками — полпроцента. Нормальное плюс нормальное равно нормальное. Пущено это в оборот в начале XIX века, тут же функция интеграла ошибок. А вот производные из этой функции появляются появляются в начале XX века.

χ^2_n — это целое семейство распределений с параметром. $\chi^2_n = z_1^2 + z_2^2 + \ldots,\ z_i \sim \mathcal{N}(0; 1)$. Причём все нормальные стандартные должны быть дружно независимыми. Плотность у χ^2 связана с гамма-функцией, что есть обобщение факториала. И чисто математически эта случайная величина χ^2 может оказаться распределена с $n \in \mathbb{R}$ степенями свободы. Причём если есть таблицы для 21 и 22 степенями, то и для 21,5 тоже можно найти. В очереди бывают вторые, третьи, но трёхсполовинных и питых не бывает (бывают испитые). С матожиданием примитивно до безобразия: $\mathbb{E}(\chi^2_n) = n$, $\operatorname{Var}(\chi^2_n) = 2n$. В контрольной была пилюлька, которую надо было описать, но не так, как у Сент-Экзюпери («удав, проглотивший слона»). Когда что-то где-то лежит, то это функция распределения чего. По оси абсцисс бегает случайная величина (значения случайной величины). Для каждого конкретного значения есть соответствующая точка (площадь под кривой единична). Если площадь — единичка, то что-то распределено для каждого значения. Поэтому этот график — график распределения. Если интеграл, то это вероятность. Интеграл чего есть вероятность? Плотность. Плотность вероятности! Если интегрируем, то получается вероятность. За ником «функция распределения» есть полное имя. Итак, по вертикали плотность вероятности p. Итак, это функция распределения плотности вероятности случайной величины, распределённой по χ^2_n. Всё прозрачно! Величина $\frac{\chi^2}{n}$, $\mathbb{E} = 1$, $\mathrm{Var} = \frac{2}{n}$. При $n \to \infty$ имеем матожидание 1, дисперсию 0. А что это такое? Это δ-функция Дирака в пределе. А ближе к 20 значениям (даже к 10) получим нормальное распределение!

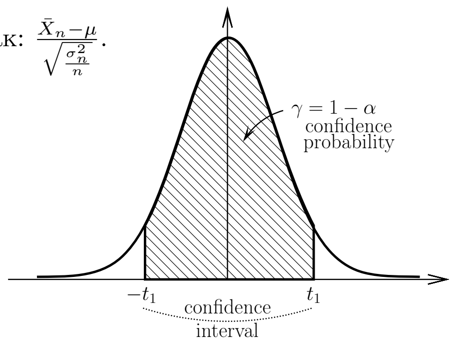
$t_n \sim \frac{\mathcal{N}(0;1)}{\sqrt{\chi^2_n/n}}$, матожидание у неё 0, а дисперсия — это $\frac{n}{n-2}$. После двадцатки идут похожие на нормальное распределения. Это начало XX века.

В 30-е годы появляется Фишер от двух параметров: $\mathrm{F}_{n;m} = \frac{\chi^2_n/n}{\chi^2_m/m}$, $\mathbb{E}(\mathrm{F}_{n;m}) = \frac{m}{m-2}$, причём при переворачивании получится тоже Фишер. И то, что лежало правее альфы, будет лежать левее. $\mathrm{F}_{n;m;\alpha} = \frac{1}{\mathrm{F}_{m;n;1-\alpha}}$, и считается F-статистика от хвоста. Хи-квадрат — это Фишер с бесконечностью в знаменателе, а Стьюдент в квадрате — это тоже Фишер.

Пусть есть СВ X и её матожидание μ, несмещённая оценка, которая годится для $\bar{X}_n$. Но чем отличаются они? $\mathbb{E}(X) = \mu$, $\mathbb{E}(\bar{X}_n) = \frac{1}{n} \ldots = \mu$. $\mathrm{Var}(X) = \sigma^2$, $\mathrm{Var}(\bar{X}_n) = \mathrm{Var}\left(\frac{\sigma X_i}{n}\right) = \frac{1}{n^2}\mathrm{Var}\left(\sum X_i\right) = \frac{1}{n^2}\sum \mathrm{Var}(X_i) = \frac{\sigma^2}{n}$. В выборке величины независимы, и среднее по выборке даёт дисперсию меньше. Дисперсия — это квадрат отклонения от среднего. $\frac{\sum(X_i - \bar{X}_n)^2}{\ldots}$. Если X_i — полный набор, то тогда просто (всех мы знаем). А если это выборка, то набор можно повторять. А наша оценка $\frac{\sum(X_i - \bar{X}_n)^2}{n-1}$. Если набор конечный, то делим на n.

Мнемоническое правило, которое пока работает без исключений. Если надо найти, на что надо делить, надо делить на степени свободы (количество независимых). Если X_i, то надо брать n. Но $\xi_i = X_i - \bar{X}_n$ — по свойствам линейной комбинации у нас их будет $\sum \xi_i = \sum X_i - \sum \bar{X}$. Если убрать 1, то будет χ^2_{n-1} с другими степенями свободы.

$\frac{X_i - \mu}{\sqrt{\sigma^2_n}} \sim \mathcal{N}(0;1)$, $\frac{X_i - \mu}{\sqrt{\hat{\sigma}^2_n}} \sim t_{n-1}$. А если брать среднее, то надо писать так: $\frac{\bar{X}_n - \mu}{\sqrt{\frac{\sigma^2_n}{n}}}$.



Надо делать проверку не через доверительный интервал, а через t-статистику. В чём связь между доверительным интервалом и доверительной вероятностью? Во-первых, ось по центру. $\mathbb{P}(|t| < t_1) = 1 - \alpha$.

Проверка гипотез. $\mathcal{H}_0$ — это то, что мы хотим проверить. Например, $\mu = 1$. $\mathcal{H}_1\colon \mu \neq 3,\ \mu > 3,\ \mu = 2$. Это то, что мы принимаем, если наша гипотеза ошибочна, — нельзя так говорить! Потому что это мы понимаем как $\mu \neq 3 \Rightarrow \forall \mu$.

«Но так как я не имею права выжигать вас калёным железом,
то я применю гуманный метод — дам подумать до следующего раза».

Раньше в книжках было боже упаси про хвост распределения написать. При переводах хвост вырвался на волю, а на научных семинарах хвосты звучали как нечего делать. Когда мы попадали в наш интервал, то гипотеза не отвергается. Если мы сравниваем $\mu \neq 3$ и $\mu > 3$, то мы просто говорим: меняется интервал отвержения. $\mathcal{H}_1$ даёт нам область, в которой будем отвергать гипотезу. $\mathcal{H}_0$ — что, $\mathcal{H}_1$ — как. Это как подлежащее и сказуемое.

Пусть мы попали на фабрику, где золото в мешочках по 3 кг взвешивают. На входе тара пустая, на выходе заполненная. Мы проверяем допуск. Если пересыпать, то это плохо, поэтому мы проверяем $\mu > 3$. Для сбыта важно, чтобы не было рекламации от клиентов. Тогда вопрос: а для них что критично: $\mu < 3$. Пришли мы к руководству, а им хорошо, чтобы не было рекламации и чтобы своё золото не пересыпать. Для них критично $\mu \neq 3$. Если суд дороже, чем пересып своего золота, то тогда мы будем иметь более тяжёлый левый хвост, куда нельзя попадать. Пишут тогда так: $\frac{2}{3}\alpha$, $\frac{\alpha}{3}$. Когда нет детализированной ситуации, то мы просто поровну раскидываем. Надеюсь, больше вы не будете мне принимать альтернативные гипотезы? А чтобы в памяти оставалось то, что надо, надо просто давать чуть больнее: экзистенциализм никто не отменял.

Ошибка первого рода — отвержение $\mathcal{H}_0$. Вероятность такой ошибки равна уровню значимости. Ошибка второго рода не так критична, поэтому заталкивается во фразу: «Мы не отвергаем гипотезы». Ошибка первого рода в Библии: если Бог, то Бог поможет. Признали, что не Бог, так как Бог не помог, а потом выяснилось, что это ошибка первого рода (самая ранняя проверка статистической гипотезы).

К домашке: отношение S^2_n/S^2_m — статистика. Сумма равна единичке, так как мы выворачиваем. Делать в Экселе. На выходе должно быть не меньше трёх значащих цифр (не нули). Будьте внимательны на вводе данных. К сожалению, печатать обязательно. Промежуточные значения добавляйте, чтобы понять, где мы ошиблись.

11 октября с утра контрольная. А 18 числа будет сдача 2 задания, и тут будет штрафбалл за первое. На каком-то этапе можно попытаться оптимизировать; но главное, чтобы эта оптимизация не коснулась интересов преподавателя.

4 Семинар 4

Dies iræ! Dies illa
Solvet sæclum in favilla:
Teste David cum Sibylla!

Весь семинар было побиение младенцев, с первого раза не сдал никто!

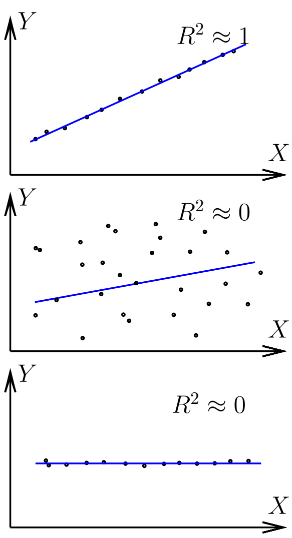
5 Семинар 5

На начальном этапе наши представления об эконометрике — это регрессионый анализ: точки соединяем в прямые и кривые. Тогда мы сможем сопоставляться с теорией. Но всегда присутствует наше желание заглянуть в будущее. Слово «регрессия» есть во всех европейских языках. Попытка поменять термин приводит к фиаско. До 90-х годов корень из оценки дисперсии называется среднеквадратичным отклонением. Этот термин неудобен. Применили «стандартную ошибку», и это сейчас моднѣе. Был ещё термин — deviation, но он не прижился. Буква «Х» была хѣром. Аз буки веди глаголь добро — это «я букву знаю и говорю правильно». Слово «похерить» совершенно естественно. Убрали букву — пошло восприятие как жаргона. Регрессия — это какое-то упрощение. В начале были наблюдения. Это n-мерное случайное пространство. А на выходе имеем два коэффициента, описывающих кривую. n-мерное пространство просто проецируется на пространство, менее мерное в порядки разов. Регрессия — это в основном МНК. Когда не было МНК (когда мы не умели точки превращать в прямые), науки не было: мы расширяем представление абстрактно. Всего 200 лет мы знание имеем!

История МНК занятная, это была борьба приоритетов — Лежандра и Гаусса. В 1805 году Лежандр печатает статью, в которой описывает МНК и называет МНК «МНК». А Гаусс скоро написал, что метод банален до безобразия, Гаусс его применяет в 1802 году, а самому ему он известен с 1797 года. А тогда была целая дискуссия. Лаплас пытался их безуспешно развести. Лежандр — член Парижской академии наук, они считают себя богами, Гаусс живёт в захолустном Берлине. Гаусс много моложе Лежандра, и он ссылается на то, что знал его в 18–19 лет. Мы сейчас пишем письмо в центр Вселенной о том, что академик фигнёй страдает, что это ничего особенного. И это письмо пишем из Верхнепупска, где нет ни железной, ни асфальтовой дороги. Лежандр был чистым математиком для математики. Гаусс шёл и интересовался влиянием ошибок измерения на ошибки результатов. Подходы у него были прикладными.

Вот есть у нас некий набор точек. И мы его меняем на прямую или кривую. $Y = a + bx$ — механизм перевода точек в кривую. Можно опустить перпендикуляр. Но надо опустить перпендикуляр. А искать третий знак в n наблюдениях — это очень уныло. Поэтому будем считать расстояние при фиксированном X_i, и расстояние будет равно $e_i = Y_i - (a + bX_i)$.

Как называется результат вычитания? «„Частное“ — это хорошо, в этом что-то есть!»



Гаусс выбрал функцию и просуммировал квадраты, хотя можно было и модуль кубов брать, и функцию специальную клепать. В общем, минимизмиуется по a и b. Условие таково: $\frac{\partial I}{\partial a} = 0$, $\frac{\partial I}{\partial b} = 0$, $b^* = \frac{\sum x_i y_i}{\sum x_i^2}$, и обязательно проходит через точку $(\bar{X}, \bar{Y})$. $x_i = X_i - \bar{X}$, $y_i = Y_i - \bar{Y}$. $\mathrm{Var}(b^*) = \frac{\sigma^2}{\sum x_i^2}$, $\mathrm{Var}(a^*) = \sigma^2 \left(\frac{1}{n} + \frac{\bar{X}^2}{\sum x_i^2}\right)$, $\mathrm{Cov}(a^*, b^*) = -\frac{\sigma^2 \bar{X}}{\sum x_i^2}$. Кроме того, дисперсия набора точек распадается на два слагаемых (факторизация): $\sum(Y_i - \bar{Y})^2 = \sum(\hat{Y}_i - \bar{Y})^2 + \sum(Y_i - \hat{Y}_i)^2$ — теорема Пифагора в n-мерном пространстве. Так и называют: $TSS = ESS + RSS$. Мы называем RSS остатком. Появляется и величина $R^2 = \frac{ESS}{RSS}$ — доля объяснённой дисперсии ко всей дисперсии. Этот критерий позволяет сопоставлять полезное с шумом, который появляется. Остальные методы такого критерия не имеют. Но мы упрощались дважды. Во-первых, мы потеряли прямые, перпендикулярные оси X, во-вторых, мерили расстояния по-своему. Такая регрессия называется парной (один регрессор).

Если точки близко стоят, то $R^2 \approx 1$. Если разбросаны хаотично, то $R^2 \approx 0$. Внимание: критерий R^2 и наш критерий разные. Мы сравниваем с расстоянием до оси, а R^2 считает ESS. Поэтому большой R^2 — это хорошо. Маленький может ничего не значить — вот наша расплата за упрощение.

Через 100 лет на это надевается название модели для данных. В статистике есть прямая $Y = \alpha + \beta X$, а мы наблюдаем какие-то разбросанные точки. Поэтому мы можем по точкам построить регрессию с крышками: $\hat{\alpha} + \hat{\beta} X = \hat{Y}$. Она, конечно же, может отличаться от нашей прямой. Нам нужна несмещённость (дотянуться до истины) и эффективность (дотянуться побыстрее). То, как мы хотим чтобы данные были устроены[1] Мы хотим, чтобы данные были так устроены: $Y_i = \alpha + \beta X_i + \varepsilon_i$. При этом получается, что X неслучаен и зафиксирован, а ошибки переводятся на Y. Это представление называется моделью данных.

Теорема Гаусса—Маркова. Теорема экзотична по многим пунктам. Ни Гаусс, ни Марков её не формулировали. Это теорема памяти обоих авторов. Гаусс вложил МНК, Марков — модельность, причём Гаусс и Марков не

[1] Запятая не нужна! Ваш буквоед-кун.

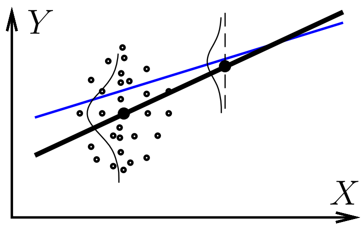
пересекались, жили в Берлине и в Питере, Марков родился через 3 года после смерти Гаусса. До середины 1970-х теорема Гаусса—Маркова была просто идеей. На входе у нас речь об одном, а в выходах — совсем другое.

Теорема Гаусса—Маркова. Если наши данные обладают следующими свойствами:

1. Модель правильно специфицирована;
2. Все X_i детерминированные (неслучайные) и не все равны между собой;
3. Ошибки не носят систематического характера;
4. Дисперсии ошибок одинаковы;
5. Все ошибки независимы между собой;

— то оценки МНК эффективны в классе линейных несмещённых оценок. Кроме того, вдруг $X_i = \varkappa_i Y_i$?

1а. Модель представляет собою детерминированную часть (функциональные особенности) и аддитивную ошибку, которая отделена от функциональной части. **1б.** Модель линейна по α и β. **1в.** В модели отсутствует как недоопределённость, так и переопределённость. Появление невлияющего фактора или фактора, который повторно учитывается, называется переопределённостью модели. А неучёт фактора — это недоопределённость. Всё равно дисперсия таких коэффициентов будет больше, чем в правильной модели. **1г.** Модель данных адекватна устройству данных, то есть имеет ту же самую функциональную форму. Положим, наши данные таковы: $Y = \alpha + \frac{\beta}{X} + \varepsilon$. А мы берём стандартную. Поэтому дисперсия в таком случае всё равно будет больше. Можно использовать $|e|$ или e^4, но будет больше дисперсия. Выигрыш МНК: у нас и так квадраты максимизируются, а дисперсия — это и есть квадрат.



Рассмотрим модель глазами нас и олимпийцев. Боги смотрят на точку, а мы получаем вертикальный ряд, матожидание — 0. Наблюдения-то связаны только с ошибкой! Бесконечно меривши, берём матожидание и получаем истинную точку. Пусть теперь $Y = f(\alpha + \beta X + \varepsilon)$. Садимся на место богов. Как они раскидают точки для смертных? Они честно ведут игру и получают какую-то окрестность. Теперь мы смертные — смотрим на точку. И бог его знает, где ей соответствует значение. Мы можем всё измерить, получить серию точек, измерить все распределения, получить прямую, однако того, что за синей прямой есть чёрная, мы и знать-то не знаем. А познаваем ли мир? Нет, непознаваем. И это проблема позитивизма: мы прогнозируем синее — мир в ощущениях. Мы видим то, что можем видеть, и знаем это. Подход материалистов: мир всегда такой — $Y_i = \alpha + \beta X_i + \varepsilon_i$ — и мы его, следственно, знаем.

Общий вывод из теоремы: если данные совпали с нашими, то МНК — лучшая оценка. Иначе мы и знать не будем, что видим не то, что надо. Нас сильно волнует наличие разума, но если он не влияет на получение зачёта по эконометрике и наличие сегодняшнего ужина, то и не надо переживать.

По заданию на дом. Пусть мы получили $\hat{\alpha} + \hat{\beta} X$. Надо проверить коэффициенты на незначимость (число без знака — это 0). Надо проверить $\mathcal{H}_0\colon \beta = 0$ при $\mathcal{H}_1 \neq 0$. Надо не забыть вычесть 0: $\frac{\hat{\beta}-0}{\sqrt{\hat{\sigma}_{\hat{\beta}}}}$. В итоге у ESS только 1 степень свободы, а у TSS их $n-1$, а $RSS \sim \chi^2_{n-2}$, поэтому $\frac{\hat{\beta}-0}{\sqrt{\hat{\sigma}_{\hat{\beta}}}} \sim t_{n-2}$. Если мы попали в хвост далеко, то гипотеза отвергается, утверждение о том, что коэффициент 0, не проходит, и мы будем использовать оценку. Это делается, чтобы понять зависимость. И чем дальше мы улетаем в хвост, то тем корректнее утверждение о значимости. Получаем p-value. И для каждого конкретного значения он будет выступать ошибкой первого рода. Ошибка первого рода: коэффициент значим.

Цена может зависеть от инфляции, сезонности и действий с промоушеном. Доходы и прочее — это более тонкая ситуация. Пусть мы исследовали и говорим: цена растёт с темпом 5 копеек в неделю, и вероятность ошибки — 3,7 % — это и есть p-value. А если попадаем в тело, то не отвергаем гипотезы. И модель будет такова: $Y_i = \alpha + \varepsilon_i$. И тут вступает в силу принцип Оккама: выбирайте более простую модель. И тогда гипотеза принимается. То, за что нас ругали, работает: это эконометрика. Оккам: не надо преумножать сущность без необходимости. Английская версия: be easy. Дома нужно будет 17-е значение. Можно сразу записать: $\mathrm{Var}(\hat{Y}) = \sigma^2 \left(\frac{1}{n} + \frac{(X_{17} - \bar{X})^2}{\sum x_i^2}\right)$. Будет так? $\hat{Y} - Y_{17}$, и у неё в дисперсии добавится $+1$. Рекомендации: делать в Экселе в электронной таблице. В «Анализе данных» есть линейная регрессия. При распечатке избавьте Мамонтова от комментариев от себя. Выводов никаких. No comments.

6 Семинар 6

Модель CAPM. $r_a = r_t + \beta(r_m - r_t) + \varepsilon$.

Когда Маркович вязал матожидание с дисперсией, то поначалу всё было очевидно, например, идея о связи матожидания с дисперсией. Например, на машине Чернышевкому получится добраться до Вышки за 30 минут, если он встал в 4 часа, и 3 часа, если он встал в 7:30. И в зависимости от дисперсии мы сдвигаем матожидание. Преодолеть первые 2 километра — это час, а потом всё нормально. Корень из дисперсии мы тоже перекладываем на матожидание, когда мы идём в магазин. 5 лет назад безрисковым активом были Treasury Bills. Оказалось, что это самый большой спекулянт. А сейчас ни еврооблигации, ни американские бумаги не суть безрисковые.

Записывается задача в форме $r_a - r_f = \beta(r_m - r_f) + \varepsilon$, $Y_i = \beta X_i + \varepsilon_i$. Линия регрессии проходит через начало координат. Если у нас незначимое β, мы должны его выбросить. Идея упрощения — это то, что зависимости от

X нет. И мы вправе упроститься. Так что $\alpha = 0$ — это не упрощение. Это усложнение: наша прямая должна пройти через начало координат. И даже если коэффициент α незначим, то это не значит, что мы должны к более строгой модели перебежать. Мы делаем две модели: сначала строим со свободным членом, проверяем его на незначимость, а потом переходим к модели без свободного члена. α выделена не только этим: вычислительно β сильно похожи. Считая оценку дисперсии, получим немножко другое: в $\alpha + \beta X + \varepsilon$ $n-2$ степеней свободы, так как $\bar{Y}$ забрало одну степень свободы, а в $\beta X + \varepsilon$ $n-1$ степень свободы, так как опосредованно через среднее степень свободы забрана не была. И мы рассматриваем единицу в качестве обычного регрессора. Без свободного члена не работает теорема Пифагора, и у нас исчезает одно условие минимума. Производная по α давала нулевую сумму остатков, и поэтому среднее по наблюдениям и вычисленным значениям не совпадают. Поэтому приходится кем-то жертвует. Мамонтову нравится TSS сохранять, а Эксель считает $RSS + ESS = TSS$, но тогда TSS не вяжется. R^2 — вещь полезная (больше у МНК нечего рассматривать), но он не самодостаточен. Если точек мало, то тогда R^2 такой же, как и когда их много, но когда их много, то β существенно значимей, дисперсия падает, а R^2 в силу геометрии не меняется.

Итак, запомните: сначала считайте с α, проверяйте на незначимость и переходите к модели без оной.

Зачем нам нужна нормальность наших ошибок? Если они нормальны, то нормальны константы и Y. Если нормальны Y, то суммы Y распределены Y. Нормальность β позволяет использовать известные оценки. Если мы потеряем нормальность, то не так страшно: у Гаусса—Маркова требуется совсем немногого — одинаковая дисперсия, и то для удобства, чтобы было доказывать теорему проще. Итак, мы получили всё, кроме лоска для проверки гипотезы. Но ненормальность ошибок всё равно асимптотически приведёт к нормальности β. Ведущий знак на 30 наблюдениях уже хорош. Даже если β ненормальны, то всё равно-то данные считаются по какой-то t-статистике, получается p-value 4,3, а на самом деле 4,2. Есть ряд математиков, которые говорят, что без нормальности жизни нет, но на самом деле какая разница, если мы в асимптотике живём?

Тест Харки—Бира методологически полезен: $JB = n\left[\frac{S^2}{6} + \frac{(k-3)^2}{24}\right]$, $S = \frac{\sum e_i^3}{n\hat{\sigma}^3_{ML}}$, $k = \frac{\sum e_i^4}{n\hat{\sigma}^4_{ML}}$. Проверяется $\mathcal{H}_0\colon S = 0,\ k = 3$, $\mathcal{H}_1\colon S \neq 0,\ k \neq 3$. Нормальные величины возводим в квадрат, их вдвоём складываем, получаем χ^2_2. И асимптотика бодро идёт. Чтобы тест пожёстче сделать, можно степеней подбавить, но это избыток. Куда корректнее делать тест для модели с альфа, так как мы изначально полагали: $e_i = 0$, и более корректная проверка производится для модели. И помните: отсутствие нормальности не так страшно, когда наблюдений достаточно много. Если у нас $\varepsilon \sim \mathcal{N}$, то $\varepsilon^2 \sim \chi^2_1$, $(\varepsilon + a)^2 = a^1 + 2a\varepsilon + \varepsilon^2$ Если есть не нулевое значение, то получается константа и хи-квадрат: $a^2 + \chi^2_1$. В модели без свободного члена, скорее всего, асимметрия будет давать увеличение.

Природа так устроена, что вокруг нас неслучайных величин нет. И здесь мы будем рассматривать, когда у нас есть ошибки измерения: ошибки X и Y могут зависеть от чего угодно. Рассмотрим цену некоторого товара. Для такой-то массы товара цена такая. Произвольную точку мы взяли, например, 1, и значение Y для $X = 1$ может быть посчитано неточно. Выверить ошибки измерения легко, просто чуть увеличится дисперсия Y. А у X случайность может не зависеть от ε. Тогда $X_i = Z_i + w_i$. Когда случайность X не зависит от ε, то происходит такая вещь: само значение β оказывается смещённым. $\beta = \mathbb{E}(\hat{\beta}) = \beta_0\left(1 - \frac{\sigma^2_w}{\sigma^2_X}\right)$. X — это конечный набор, мы можем посчитать. σ^2_w тоже получим. Если мы бесконечно много мерим, то мы получим чистую β. И $\beta - \beta_0 = -\beta_0\frac{\sigma^2_w}{\sigma^2_X}$. Если мы мерим в килограммах, а отклонение в граммах, то это нормально. А если недосыпали в килограммах, то это уже вопиюще. Поэтому хорошо, когда X стоят подальше. β — то, что можем получить β и досчитать β_0 для оценки смещения. Пусть $\beta_0 = \hat{\beta}$. А $\frac{\sigma^2_w}{\sigma^2_X}$. Это позволит нам оценить тяжесть ошибки. А есть у нас ещё $\hat{\sigma}_{\hat{\beta}}$, и эта ошибка сопоставляется с полученной на вычислениях. А когда соизмеримы сигмы, то порядок оценивания близок, значит, оценка неточна. И понятно, почему у нас рушится модель, когда катаклизмы на рынке происходит, так как неизвестно, что мы в тот момент мерим. Нам даётся ошибка измерения r_m, и если мы напишем не 4,3, А 4,2.

Когда складываются или вычитаются две неточные величины, то складываются (всегда) абсолютные ошибки. А при умножении и делении складываются относительные ошибки. % — это относительная ошибка. Кстати, $\hat{\beta}\frac{\sigma^2_w}{\sigma^2_x}$ уже относится к задаче без альфы. В Экселе поиграйте галочками, напишите, что константа 0. β_0 возникает из-за наличия ошибки. Ели из большого числа вычитать большое, то относительная ошибка катастрофически растёт. Итак, надо сделать две регрессии, помните! И точность коэффициентов вам важнее, чем какой-то эфемерный R^2.

Δx, Δy (сидят абсолютные на X и Y). Свойства: $\delta X = \frac{\Delta X}{X}$, $\delta Y = \frac{\Delta Y}{Y}$, $\Delta(X \pm Y) = \Delta X + \Delta Y$, $\delta XY = \delta X + \delta y$. Когда данные вводите, тогда сразу и распечатывайте. На бумаге ошибки быстрее находятся.

7 Семинар 7

$Y = \alpha + \beta x + \varepsilon$, $Y = \alpha + \beta f(x) + \varepsilon$. Это просто замена переменных. Нечто иное происходит, когда $g(y)$. Проблема — это то, что функция, и появляется другая зависимость от ε.

Функция Кобба—Дугласа: $Y = AK^{\beta_1}L^{\beta_2}$, $\ln Y = \ln A + \beta_1 \ln K + \beta_2 \ln L$, $Y = \alpha + \beta_1 X_1 + \beta_2 X^2 + \varepsilon$. Но для последнего ε необходимо в модели e^{ε}, хотя вероятнее, что в модели была $+\delta$, которая пропадает при логарифмировании. Если ошибки большие, то нам не дано увидеть функцию, которая за ними стоит. Поэтому любые уравнения не подойдут, линейные или нелинейные. Пример: попробуйте шёпотом в метро разговаривать.

Надо просто подождать, пока шума не будет. Обратная ситуация — когда ошибки малы в разы или на порядок. Запишем это иначе: $Y = AK^{\beta_1}L^{\beta_2}\left(1+\frac{\delta}{AK^{\beta_1}L^{\beta_2}}\right)$, и в логарифме будет $+\ln\left(1+\frac{\delta}{AK^{\beta_1}L^{\beta_2}}\right)$. Поскольку она невелика, то логарифм эквивалентен $\frac{\delta}{AK^{\beta_1}L^{\beta_2}}$. Когда ошибка так маячит в уравнении, очень часто происходит упрощение модели. Помогает линеаризация: первый, линейный, член даёт ошибку, поэтому этому даже не Тейлор, а Макларен. Надо тогда проверить гипотезу о незначимости коэффициентов: $\beta_1+\beta_2=1$ при альтернативе $\neq 1$. Для проверки гипотезы мы берём оценки, $\frac{\hat\beta_1+\hat\beta_2-1}{\hat\sigma_{\hat\beta_1+\hat\beta_2}} \sim t_{n-3}$. $\hat\sigma^2_{\hat\beta_1+\hat\beta_2} = \hat\sigma^2_{\hat\beta_1} + \hat\sigma^2_{\hat\beta_2} + 2\operatorname{Cov}(\hat\beta_1,\hat\beta_2)$. Эксель ковариацию не хочет считать. Но достаточно всего лишь вспомнить, что $\operatorname{Cov}(\hat\beta_1,\hat\beta_2) = \hat\sigma_{\hat\beta_1}\hat\sigma_{\hat\beta_1}\cdot r(\hat\beta_1,\hat\beta_2)$, $r(\hat\beta_1,\hat\beta_2) = -r(X_1,X_2)$.

Рекомендация считающим в экселе. Поставьте галочку в пункте «Метки». Тогда первая строчка будет трактоваться как названия, и эти названия будут переноситься в отчёт. Тогда надо будет и с названиями выделять.

Пусть мы проверили t-статистику, попали в правый хвост. Что значит, что эта сумма больше единицы? Тогда будет возрастающая отдача от масштаба. Положительный коэффициент масштаба в экономике означает что? Функция Кобба—Дугласа означает вот что: каждая точка — это одно предприятие. Поэтому она собирается для отрасли. Если 7 предприятий, то это мало. Можно взять панельные данные и через год взять 7 точек (но надо помнить, что функция сменится). Что значит возрастающая отдача от масштаба в отрасли? Представьте, что работаете в инвестиционном банке. Все инвестиционные проекты делаются так: деньги дают так, когда вам не надо, а когда надо, то не дают (а то не вернёте). В инвестиционном проекте будет куча пунктов, таких, как наличие рабочей силы, отношение администрации. Инвестиционная компания бюрократичная. Начинается процедура прохождения по инстанциям. Все инстанции должны сказать «да», иначе его рассматривать не будут. При выхождении наверх ещё не факт, что дали добро. При положительном масштабе пришли три человека. Первый хочет открыть предприятие в этой отрасли. Второй хочет докупить предприятие в этой отрасли и расшириться. Третий хочет что-то новое в этой отрасли создать. Объём выпуска никогда просто так не увеличится, если нет спроса. Только один путь — положительный эффект масштаба капитала — отрасль беременна монополией, и она схлопнется в монополию. Следующий вопрос — стоимость компании вырастет. Подрастают на 5–10 % и поглощаемые, и поглощающие. Просто это объективная предпосылка. Гораздо проще сделать так, чтобы фирма захотела, чтобы её купили. У большой фирмы есть запас, поэтому мы опускаем у неё цены; у нас запас есть, а маленькая будет задыхаться. Для приходящих в отрасль надо опускать цены, чтобы отвоевать покупателей. Новое предприятие сдохнет. А прикупить кого-то — это пожалуйста. При отрицательном эффекте масштаба же добро дадим инноватору, так как большие фирмы сдохнут, а спасать их — пустое занятие. А как это вычислять? Выпуск — это денежное выражение, и при текущей технологии нужен дефлятор. Если фирма продаёт ещё вовне, то будем считать валютный курс. Капитал — это основной и привлечённый. Для учёта всего нужны настроечные коэффициенты. Труд тоже учитывается непросто. Если учитывать по головам, то уборщица окажет важное место. Если зарплата, то тогда топ-менеджеры учтутся сильно. p-value будет большим, но хотя бы тренды можно просчитать. Японцы это делали в середине 1980-х годов.

$\ln Y = \alpha + \beta_1\ln K + \beta_1\ln L + \varepsilon$, $Y = \alpha + \beta_1 K + \beta_2 L + \varepsilon$. Как выбрать из двух регрессий? $Y = \alpha + \beta f(x) + \varepsilon$, $Y = \alpha + \beta g(x) + \varepsilon$. Когда одинаковые Y, то TSS одинаковые. Когда мы выбираем, то RSS могут несильно различаться. Для проверки значимости различия вводится тест Бокса—Кокса. $\mathcal{H}_0\colon ① \sim ②$, то есть регрессии эквивалентны: $\frac{n}{2}\left|\ln\frac{RSS_1}{RSS_2}\right| \overset{\text{as}}{\sim} \chi^2_1$. Если p-value малое, то тогда первая модель лучше, а если большое, то тогда обе модели безразличны.

В различиях при моделях с $\ln Y$ и Y принято нормировать: $Y^* = \sqrt[n]{\prod\limits_{i=1}^{n} Y_i}$, $\ln Y^* = \frac{1}{n}\sum \ln Y_i = \overline{\ln Y_i}$. Это уже добавил Зарембка. $\tilde Y = \frac{Y}{Y^*}$. Это не значит, что мы мы будем пользоваться изменённой регрессией. Надо просто проверить, как одно или другое объясняет RSS.

Первое — сделать регрессию, проверку тестом Бокса—Кокса, потом проверку $\beta_1+\beta_2=1$.

Новая тема. Линейная регрессия с ограничениями.

$Y = \alpha + \beta_1 X_1 + \ldots + \beta_k X_k + \beta_{k+1}X_{k+1} + \ldots + \beta_{k+l} + \varepsilon$. Проверяется гипотеза: $\beta_{k+1;\ldots;k+l} = 0$, что то же, что $\sum_{i=k+1}^{k+l}\beta_i^2 = 0$.

$Y = \alpha + \beta_1 X_1 + \ldots + \beta_k X_k = \varepsilon$, $TSS = ESS + RSS$, поэтому в этой модели ошибки ползут из ESS в RSS. RSS прирастает за счёт этих самых l штук слагаемых. У unrestricted RSS имеет $n-k-l-1$ степеней, у restricted — $n-k-1$. Поэтому соотносится $\frac{(RSS_{\text{restr}} - RSS_{\text{unrestr}})/l}{RSS_{\text{unrestr}}/(n-k-l-1)} \sim \mathrm{F}_{l;n-k-l-1}$. Если хоть кто-то из β не ноль, то мы быстро уходим в хвост.

Частный случай. $Y = \alpha + \beta_1 X_1 + \ldots + \beta_k X_k + \beta_{k+1}X_{k+1} + \varepsilon$ — unrestricted. $\mathcal{H}_o\colon \beta_{k+1} = 0$ против $\neq 0$. $\frac{(RSS_{\text{restr}} - RSS_{\text{unrestr}})/1}{RSS_{\text{unrestr}}/(n-k-1)} \sim \mathrm{F}_{1;n-k-2}$. $\frac{\hat\beta_{k+1}}{\hat\sigma_{\hat\beta_{k+1}}} \sim t_{n-k-2}$.

Второй случай. $k=0$, $Y = \alpha + \beta_1 X_1 + \ldots + \beta_l X_L + \varepsilon$. $\mathcal{H}_0\colon \beta_1^2 + \ldots + \beta_l^2 = 0$ и что $Y = \alpha + \varepsilon$. $\frac{(RSSr - RSS_{\text{unrestr}})/l}{RSS_{\text{unrestr}}/(n-l-1)} = \frac{ESSu/l}{RSS_{\text{unrestr}}/(n-l-1)}$ — эта гипотеза проверки неадекватности регрессии, или гипотеза проверки, что $R^2 = 0$.

Третий случай. $Y = \alpha + \beta_1 X^1 + \beta_1 X_2 + \varepsilon$. $\mathcal{H}_0\colon \beta_1 + \beta_2 = 1$. $Y = \alpha + \beta_1 X_1 + (1-\beta_1)X_2 + \varepsilon$. Тогда $\underbrace{Y - X_2}_{\tilde Y} = \alpha + \beta_1(X_1 - X_2) + \varepsilon$, и это регрессия с ограничением. $\frac{(RSS_{\text{restr}} - RSS_{\text{unrestr}})/1}{RSS_{\text{unrestr}}/(n-3)} \sim \mathrm{F}_{1;n-3}$.

Когда нет постоянной отдачи от масштаба, то Фишер не показывает, как она отличается. А t-статистика показывает куда, поэтому она удобна, когда гипотеза не выполняется. Посчитать двумя способами первую задачу.

Фиктивные переменные: $D = \begin{cases} 1; \\ 0. \end{cases}$ $P = \alpha + \beta_1 X_1 + \beta_2 D + \varepsilon$. Выбор определяется базовой кривой. Есть $P = \alpha + \beta Q$, $Q = \alpha + \beta P$. В эконометрике в модели есть ошибка. Количество фиксированное, а количество меняется. Поэтому и берётся $P = f(Q)$. И именно поэтому Маршалл так думал: и спрос, и предложения мы видим обратными. А прямой спрос находится внутри домохозяйства.

Пусть собираем цену по бананам. Звоним, спрашиваем: почём? — А сколько вам надо? По цене за тонну будет 1,09 \$, за 5 тонн — 1,07 \$, за 50 тонн — 1,05 \$, за тысячу тонн — 1,02 \$. Функция обратная. Это спрос или предложение? Это равновесное состояние? Банальная ситуация, а не потому, что бананы. Три года изучения экономики впустую. Это не оценка. Это фирма нам подсовывает, как она хотела бы видеть спрос. Реальный спрос где-то там. Но получается предложение! Это действительное предложение! Итак, это предложение в виде подхода на спрос. Предложение в виде желаемого спроса.

Если вы спрашиваете про два банана, то это вы с девушкой заигрываете на том конце провода.

8 Семинар 8

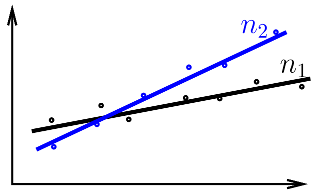
На каждое задание надо писать резюме. Это должно быть напечатано (если можете от руки писать гарнитурой Таймс, ради бога). Объём — одна страница. Учимся писать резюме. Остаётся право подхихикивать над ним. Мы пишем, что мы делали и как мы делали.

Можно будет сдавать каждое следующее домашнее задание, не сдав предыдущее. Будет ещё предэкзаменационная работа. Скоро надо проверить опись стандартных и характерных ошибок.

По поводу задания. Множественная регрессия в матричной форме: $\vec{Y} = \boldsymbol{X}\vec{\beta} + \boldsymbol{\varepsilon}$, k регрессоров, n наблюдения. $(n \times 1) = n \times (k \times 1)(k \times 1) + (n \times 1)$. $\boldsymbol{X} = (1, X_1, \ldots, X_k)$. $\hat{\vec{\beta}} = (X'X)^{-1}X'\vec{Y}$, $\mathrm{Var}(\hat{\vec{\beta}}) = \sigma^2(X'X)^{-1}$, $\hat{\beta} = \frac{\sum x_i y_i}{\sum x_i^2}$, $\mathrm{Var}(\hat{\beta}) = \frac{\sigma^2}{\sum x_i^2}$. $\hat{\sigma}^2 = \frac{RSS}{d.\,f.} = \frac{e'e}{n-k-1}$. Требуется, чтобы матрица $(X'X)$ была невырожденной. X независимы между собой.

Часто будем записывать в одномерном виде, так как всё равно мы всё засовываем в пакет.

Тест Чоу (Чау) на устойчивость коэффициентов. Даны два набора наблюдений: $n_1 + n_2 = n$, k регрессоров. Надо проверить, что они совпадают: $\mathcal{H}_0 \vec{\beta}^{(1)} = \vec{\beta}^{(2)}$, $k+1$ равенств. У общей регрессии $n-k-1$ степеней свободы. Но пока мы единицу пишем единицей, надо считать, что коэффициентов $k+1$, регрессоров k. Тест не получается, если n_1 или n_2 меньше, чем количество регрессоров.



$$\frac{(RSS_{1+2} - RSS_1 - RSS_2)/(k+1)}{(RSS_1 + RSS_2)/(n - 2(k+1))}$$

Выяснилось, что мы тексты читать не умеем. Если будет три набора, то в числителе $k+1$ удваивается, в знаменателе $-3(k+1)$. То есть в знаменателе количество связей.

Тест на предсказательную силу. Мы отбираем какие-то точки на конце и смотрим, хорошо ли основная часть описывает концевые. $\mathcal{H}_0\colon \sigma_n^2 = \sigma_{n-p}^2$ (примечание: гомоскедастичность — это если сравниваем p и $n-p$). Тогда

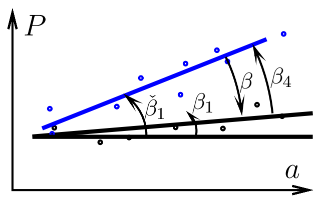
$$\frac{(RSS_n - RSS_{n-p})/p}{RSS_{n-p}/(n-p-k-1)}$$

Когда-то нас просили написать доверительный интервал для следующего. Формально мы оказывались в рамках теста для $n = 1$. А тут общий вид. Можно проверять и внутренние точки. Этот тест более общий. Берут $p \ll n$.

Мы, ища зависимости, вываливаем все X дружно и думаем, кого выбросить и кого оставить. Это для случая независимости или слабой зависимости. Легко выбросить все незначимые. Но из-за мультиколлинеарности мы действуем по одному и из помидоров выбрасываем тот, который тухлее всех. Чем мы можем наблюдать? Текущее состояние t-статистик. Кроме того, можно при $\beta_{i_{1,2}} = 0$ взять $\frac{(RSS_{\mathrm{restr}} - RSS_{\mathrm{unrestr}})/\ldots}{RSS_{\mathrm{unrestr}}/(n-k-1)}$. Смотрят $R^2 = 1 - \frac{RSS}{TSS}$ и $\bar{R}^2 = 1 - \frac{RSS/(n-k[-1])}{TSS/(n-1)}$. Нормированный используется только при сопоставлении регрессии. Если выбросить значимый коэффициент, то нормированный уменьшается. А если выбрасываем незначимый, а нормированный может остаться таким же или подрасти. Достоинство: чтобы посчитать F-статистику, надо посчитать, а тут проще. Наигрыш: не зависит от статистики. Проигрыш: нормированный $\bar{R}^2$ отчасти дебиловат, так как он даёт не количественную, а качественную характеристику при $t = 1$ и вероятности 20 %. Это спорно. И вообще он анахроничен. Мы через F-статистику получаем ряд изменений $\bar{R}^2$.

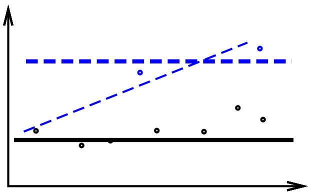
Попробуйте ради удовольствия выбросить что-то значимое. Нужно, чтобы галочка на метках стояла уж наверняка. Не корёжьте начальной таблицы, делайте копии.

Модель: $P = \beta_0 + \beta_1 Q + \beta_2 I + \beta_3 D + \beta_4 DQ + \beta_5 DI + \varepsilon$. Двумя тестами ЧОУ проверить на дамми. Может первый тест не получиться, так как наблюдений будет меньше, чем регрессоров.



До этого говорилось, что от выбора дамми ничего не зависит. Зависит! Вас обманывали. Сначала били за то, что принимали гипотезу, а потом бьют за неотвержение. А представьте, что мы работаем с наклоном. Пусть состояние наклона меньше на окраине, больше на центре. Выбираем так: 0 — окраина, 1 — центр. Тогда коэффициент на окраине может быть незначимым. Меняем дамми: 1 для окраины, 0 для центра. А незначима сумма $\beta_1 + \beta_4$. Но мы же не видим суммы! А сразу проверять суммы — это странно. Получается, что при одном подходе коэффициент β_1 выкинут, при втором живёт. Но мы не знаем, какой коэффициент правильный. Поэтому проще перевернуть дамми и сразу посчитать.

«А принцип Оккама? XV век никого не отменял!»



Пусть для центра мало наблюдений, только две точки. Регрессия целиком не получится. Будет либо параллельный сдвиг, либо другой угол. Или — или. Незначимы β_3 или β_4. Более честно сдвинуться, а не наклониться. Не выбрасывайте чистых дамми; начинайте с углов наклона.

Средняя цена в корзинке зависит от того, сколько лежит в корзинке. Даже не надо спрашивать, что это — спрос или предложение. Это психологическая функция группы покупателей. Это происходит на кассе; это вообще не спрос, не ходите в плену иллюзий. Если покупок мало, то бета незначимый. А значимость проявляется, когда существенно большое количество.

Мысленный эксперимент. Берёте продовольственный магазин, и надо купить не батон хлеба и коробку масла, а громадный список в голове целый формируется. И мы, полные решимости, берём телегу (ещё у нас позитивное настроение потому, что совершите ЭТО). По себе знаем: доставать ручку неудобно. Идём по списку снизу вверх. Делаем второй круг и обращаем внимание, что никто не сгруппировал это по отделам. Ещё хуже, когда по памяти двигаемся, и тогда блуждание по магазину ещё более активное. Список наполнился, терпение подходит к концу, уже напряжённо. Добираемся третий раз в молочный отдел. Написано: «Домик в деревне». А мы его не видим. Берём какое попало. Выходим из магазина, у нас окрылённость, вы совершили подвиг. На каком-то этапе нам придётся тащить, мы вваливаемся в дом, гордые, довольные, и слышим: «Мы пьём только „Домик в деревне“!» — «Не было!» — «Там на соседней полке всегда стоит „Домик в деревне“!» — «Сама сходи!» Если бы товар в магазине был более дорогой, мы бы определились. Но мы устаём от покупок, и правильно, расставив товары, взять деньги, которые мы готовы ему отдать. Мы готовы отдать больше за то, что мы устали. Как дотянуться до больше? Поставить поближе дорогое. Кроме того, почти то же самое должно стоить подороже. Можно спросить, почему такая ерунда, что всё разно стоит. Пока мы не озверели, мы можем проявить реакцию на цену. Вроде другая цена была. Возвращаемся — другая цена. Потом начинаем сравнивать их. Мы занимаемся не покупкой, а сравнением, и боже упаси, если дата будет одинаковой. Будет третий с другой ценой — мы будем в безумном восторге. Был «homo покупатель», стал «homo исследователь». И не исключено, что мы пойдём в соседний магазин сравнивать. У Мамонтова была ситуация, когда одно было дороже в аптеке, чем другое. Разница — поставщики разные. Чтобы нам подтолкнуть нечто такое же в новой цене, нам надо видоизменить. Потребитель попадается на это на больших покупках, и надо подставить бету так, чтобы было удобно взять, когда мы выходим из магазина. Также очень важны товары-дополнители. Мы устаём от покупок, но при покупке дорогого у нас наступает эйфория. Пусть мы взяли много денег, купили шубу, осталась сумма, но она до дома и не дойдёт: счас мы её и кончим! И додумаемся только тогда, когда выйдем из этого состояния. Если узко рассматривать, то узко и получается. К женским шубам очень хорошо продавать дорогие мужские галстуки. Есть «чувство вины»: «Дорогой, смотри, какой галстук я тебе купила!» — «А это что?!»

Средства по уходу за шубам нельзя продавать с шубами, так как это напоминание о недолговечности. «Представьте, что оформляете младенцев, а рядом покойников записывают — memento mori!»

Когда Никсон был вице-президентом у Эйзенхауэра, был плакат: деньги всегда есть, только в чьём кармане?

С тележкой и ребёнком мы постоянно двигаемся и не отвлекаемся на «Купи! Хочу!». У кассы мы будем стоять. Стоим с ребёнком. Это комплементарный товар к этому стоянию. Ребёнок входит в стояние, и у нас нет возможности оттащить его от угла. А здесь чуть-чуть подороже.

Комплементарность можно рассматривать ещё шире. В СССР не было круглосуточных магазинов. В 1991 даже запчасти круглосуточные были. Появились продовольственные магазины, где бакалея и зубные щётки, а сейчас стало пошире только на подгузники. Мамонтов понял всю гениальность: круглосуточный магазин работает ночью. Маленькая стоечка, где продаются мужские носки. Взял бутылку вина, конфеты, а целый день на ногах… Это товар под определённое время. И цена его совсем другая, но мы готовы платить эту цену, потому что это наша β_1. Человек попёрся не за носками. Мы на кассе это собираем. У большого магазина более тысячи транзакций, а за год за миллион заваливает. С точки зрения данных у нас море разливенное: выбирай — не хочу.

Но цены нельзя часто менять. Любое изменение цены всегда приводит к уменьшению выручки. Кто приходил, тот и приходит, везде ведь есть ограниченность информации. Новых никого пока нет. Если увеличить цену, то часть клиентов может уйти к соседям. А как менять цены? Торговый центр грабанули — хороший повод поменять в цене. Цены на ЖКХ выросли, клиент приходит и бурчит: «Ах, и у них тоже выросли». Переставлять товары часто нельзя, но надо. Для чего нужны дисконтные карты? Дают 3 % за что-то? Для чего нам нужна дисконтная карта? Чтобы определить изменяемость предпочтений во времени, так как мы можем единственным способом привязать изменение перестановок к определённой покупке. Мы видим динамику во времени.

> «Ехать через пол-Москвы, чтобы получить пенки с дерьма, — разориться!»

Ребята, думайте в эту сторону, как можно делать. В магазинах можно развернуться. Доход хорошо иметь в чеке, а не в опросе. А как иметь доход в чеке? Когда транзакций много, наблюдений много, и ошибка в иксах неважна, дисперсии маленькие. Пусть будет три — четыре градации. В нормальных магазинах кассир, кроме выбивания, отмечает воров. Он информирует о том, что есть проблема, и охрана этим занимается. Продавцы за некоторым бегают стаей, а к некоторому не подойдут. Есть критерий, за который можно зацепиться. И информация ложится в базу. Можно возраст проверять, пол. Развести по возрасту легко. Можно по полу определять. Нужно пощупать контингент при помощи реальности.

В России, простите, у нас пилят бюджетные деньги. Зачем же такой тонкостью заниматься? Бери больше, кидай дальше, наливай полнее.

Всё это на информационном уровне. Лежит база — проверяете это. Кассиров можно настраивать, это настраиваемый инструмент. Можно знакомых пропускать через одного кассира. Можно отфильтровать. Почему постоянно кассир этого мальчика обзывает старушкой? Потому что он тыкает абы попало. Вот это живая эконометрика.

Если всем становится всё по фигу, то $\beta_1 > 0$.

9 Семинар 9

Сегодня мы поговорим о мультиколлинеарности. Это когда появляется зависимость между иксами, и подход тут иной. Вот так вот: наберём зависимостей, выбрасываем незначимые, и значимые становятся незначимыми. Одно цепляем — другое становится нензначимым. Может, выбрасываем значимые, а незначимые становятся значимыми. Выбросили кучку незначимых — выбросили что-то полезное. Ещё признак: есть t-статистики, p-value, а можно проверить на $R^2 = 0$. Если t-статистики хуже, чем F, то это мультиколлинеарность. Непонятно, что делать с моделью. Зацепиться не за что.

Пусть $Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$. Что нам требуется от регрессии? Нам требуются только коэффициенты и их точность. И если точность нас устраивает, то всё нормально. Надо, чтобы коэффициенты были значимыми. Мы за значимость боремся. Если p-value большие, то тогда ошибка может быть больше, чем порядок наблюдения. Чуть глубже копнём. Посчитаем коэффициенты парных корреляций между иксами. r_{X_1,X_2}, r_{X_2,X_3}, r_{X_1,X_3}. Попарно они могут друга ничтожить; надо смотреть на корреляцию β.

Пусть истинная дисперсия первого равна $\sigma^2_{\hat{\beta}_1}$. Это представимо в форме произведения: $\sigma^2_{\hat{\beta}_1} = \frac{\sigma^2}{\sum x^2_{1i}} \cdot \frac{1}{1-R^2_{1;2,3}}$. Строится другая регрессия и частичный R^2: $X_2 = \gamma_0 + \gamma_1 X_2 + \gamma_2 X_3 + \delta$, и для этой регрессии строится $R^2_{1;2,3}$, то есть множественный коэффициент корреляции первого от всех. Итак, дисперсия первого коэффициента зависит от первого и своих соседей. Множитель $\frac{1}{1-R^2_{1;2,3}}$ называется VIF (множитель, раздувающий инфляцию). VIF от трёх до десяти — это серьёзная мультиколлинеарность.

Пусть изначально был p-value равен 2 % значению для t-статистики 2,5, а потом стало из-за VIF, равного 3, уже 20 %.

«Откуда ноги растут? Где та задница?»

$\hat{\vec{\beta}} = (X'X)^{-1}X'\vec{Y}$. Главное условие — это когда определитель не равен нулю. Вышка от метро далеко? Если из Питера смотреть, то ближе, чем из Нью-Йорка. Определитель можно сравнивать только с самим собой. Если у нас есть матрица $a_{i,j}$ размера $n \times n$, если мы умножаем $a_{i_1,j_1} \cdot a_{i_2,j_2} \cdot \ldots \cdot a_{i_n,j_n}$, где $i_1, \ldots, i_n$, $j_1, \ldots, j_n$ берётся различными способами, потом учитывается произведение с $(-1)^{i+j}$, и определитель получается примерно таким: взяли первый столбец, берём самый крупный элемент, взяли второй столбец, самый крупный элемент, и так по всем столбцам. Примитивный пример:

$$\begin{vmatrix} 4 & 2 \\ 2 & 1 \end{vmatrix} = 0, \quad \begin{vmatrix} 4 & 2 \\ 2 & -1 \end{vmatrix} = -8, \quad \begin{vmatrix} 4 & 2 \\ 2 & 1{,}001 \end{vmatrix} = 0{,}004$$

В первом случае мы ожидали определитель порядка 4, а получили в последнем случае о-малое.

Это называется малым определителем. Это называется плохо обусловленными матрицами (bad conditioned). Проблема — большое различие собственных значений $\frac{\lambda_{\max}}{\lambda_{\min}} = N$, $\sqrt{N} = I$. Это Bad Conditioned Index. Пакеты современные его считают.

Давно у нас была линейная алгебра. Она нам нужна, чтобы понять, что такое собственное значение. Но из-за разложения мы смысла не видим. Если мы в любой проблеме находим собственное решение (это может быть и нелинейная проблема), то мы начинаем воспринимать проблему изнутри.

Идиотская задача. Берут два типа серых животных. Их два типа. Начинают их взвешивать. Получают суммарный вес. Надо по суммарному весу решить обратную задачу: сколько весит один, сколько другой. Серые были слоны и мыши. Есть понятие ошибки измерения (весы), а есть случайность как внешний фактор. Слоны терпеть не будут вечно на протяжении эксперимента, и это будет в сумме тяжелее, чем все мыши. А можно рассмотреть отношение длины носа к длине хвоста, и тут ситуацию можно разгрести, даже если одной мыши отгрызли хвост и она стала похожей на слона.

Физики, конечно, представили 16-разрядную АЦП, но у нас мерили только 14 разрядов. А стандартная экономика — это 3–4 цифры. Четвёртый порядок может быть уже никаким. λ — это мы слоновье дерьмо с мышами пытаемся ~~сменивать~~ сравнивать не в пользу одних. А вот если слонов с тапирами сопоставлять, это куда ни шло.

Пусть коэффициенты значимы, точность хорошая, и тогда суетиться не надо. Но ничего не делать тоже нельзя. Если точность от Вышки до метро 854 ± 2 метра, то это нормально. 800 ± 500 — это ненормально. 10 орехов — это куча? А один орех? Чёткое определение: куча — это когда много. А что такое много? Чёткое определение: много — это когда съел и больше не хочется.

Для уменьшения первого компонента дисперсии надо ~~добавить~~ в разы увеличить число наблюдений. Мультиколлинеарность — это ещё и micronumerocity. Идея простая и со вторым множителем: переспецифицировать модель. Надо выбирать там, где значимые коэффициенты, и нам наплевать на R^2 и на F-статистику.

Метод главных компонент: нормируем, центрируем (минус среднее делить на дисперсии). В матрице $\boldsymbol{Z}'\boldsymbol{Z}$ считаем собственные значения, считаем векторы, оставляем только значимые, которые составляют 0,7–0,8 суммы. Маленький изъян: если мы меняем набор наблюдений, то векторы плывут. Мы получили $\boldsymbol{Z}$ — линейная комбинация иксов. А если мы не можем объяснить физический или биологический смысл этих векторов, то это схоластика и бессмыслица. Можно подойти от истинного понимания собственных решений. На небольшом наборе будет не очень ортогонально, а потом она вылезет. Гребневый анализ был модным в начале 1960-х, когда не было большого набора точек. Если брать Advanced учебники (Johnson), то там написано только то, что мультиколлинеарность — это нетрудная проблема, которая решается только увеличением количества точек или переосмыслением модели. Это как насморк: если сопли, как хобот мамонта, висят, то это плохо, но с насморком любой степени можно эффективно бороться: высморкаться.

$Q_B = \alpha + \beta_1 P_B + \beta_2 P_H + \beta_3 I + \varepsilon$. Зависимость потребления водки от цен выгодна производителям. А если все перестанут пить, то Минздрав сдохнет. Поэтому он ходит и проверяет, ведь во время проверок можно много чего получить, ибо это бюрократия. Представьте экономику директора. Он знает, как торговать, а мы говорим, что коэффициенты незначимые, зависимости ку от пэ нет, но эр квадрат большой, и дисперсия большая. А директор интерпретирует всё перпендикулярно и нас выгоняет. Хорошо, незначимые, но почему? Исследователи любят, когда всё в экономике качается. Цены меняются, всё выпукло, всё заметно. В противном случае надо сидеть двое суток, чтобы поймать карасика.

Мы выясним: цены сильно коррелируют, доход — слабо. Что значит собственное решение в задаче о водке и наливке? Товары можно рассматривать как комплементы и субституты. А одновременно пара товаров всегда является и комплементом, и субститутом. Есть публика, для которой или это, или это, так как их интересует процесс один: нажраться и подешевле. Есть публика другого рода, которая эстетическая: надо по очереди запивать одно другим. Тот эстет, у которого остался один напиток, допивает его и нажирается, переходя в другое состояние. И те, кто нажирается, выбирают как-нибудь то или другое, чтобы смешать (а как оно вместе). Это как будто бы эластичность.

Цены достаточно сильно зависимы. Цена водки почти параллельна цене наливки. Комплемент лежит между ними, субститут ортогонален ему. Будем потом для упрощения считать, что комплемент — это водка, и надо только получить ортогональную составляющую. $P_H = \gamma_0 + \gamma_1 P_B + \delta$, $P_\perp = P_H - \gamma_1 P_B$, $Q_B = \alpha + \beta_1' P_B + \beta_2' P_H^\perp + \varepsilon$, а смысл γ_1 — в каком количестве одно уравновешивает другое по силе эффекта. А $\beta_2' P_H^\perp$ — это из задачи потребителя доля наливки у тех, кто эстетствует и выбирает по очереди.

Идём дальше. $I = \tilde\gamma_0 + \tilde\gamma_1 P_B + \tilde\gamma_2 P_H^\perp + \delta$. $\beta_3' I^\perp$ — это доля тех, кто смотрит не на цены, а на доход, когда пьёт. Вот она, сущность проблемы: одна проблема не заменяет другой регрессии. Решение задачи трёх тел всегда сложнее на порядок, поэтому мы сначала раскладываем на цену и доход.

Что нам предстоит делать.

1. Сделали начальную регрессию.
2. Посчитали коэффициенты корреляции.
3. Получили VIF'ы из-за больших корреляций.
4. Изменяем модели для достижения значимых коэффициентов.
5. Строим ортогональные псевдособственные решения и делим составляющую.
6. Строим регрессию со значимым.

Это ползадачи.

Модель данных у всех одна и та же, поэтому мы вправе взять 2–5 соседских наборов, добавить свои. Проверить тестом Чоу одинаковость наборов данных. Каждый данные, кстати, норовит по-своему внести. А при объединении мы вперёд погнали. Хорошо, что некоторые данные имеют откровенную нестыковку. И с этим расширенным набором надо делать всё то же самое. Когда построим ортогональные, надо их проверить на ортогональность (коэффициенты парных корреляций). И в этот момент можно посмотреть на то, чтобы выбор нашей модели не изменялся. На большом количестве наблюдений модель может поплыть. Если модель плывёт, то надо ещё увеличивать набор. Резюме уместить на полстраницы. Кроме нашего того, что творилось, можно добавить немного экономического смысла. И не нужен примитивный смысл (если это увеличим в 2 раза, то это увеличится в 2,7). Кстати, в модели нас интересует уже $\beta_1' P_B$ и $\beta_2' P_H^{\perp}$, то есть компоненты.

Нехилое замечание: если разбираем модель и говорим директору о группах потребителей, то говорим директору то, что он понял. И мы говорим ему то, что мы поняли. Он начинает ощущать себя не последним в мире вычислений.

10 Семинар 10

Сегодня разговор будет об очередном нарушении теоремы Гаусса—Маркова. Итак, гетероскедастичность. А что у нас будет меняться с оценками, если мы закроем глаза на гетероскедастичность? Да несмещённые они останутся, а эффективность может в любую сторону измениться.

$\vec{Y} = X\vec{\beta} + \vec{\varepsilon}$, $\hat{\vec{\beta}} = (X'X)^{-1}X'Y$. Было так: $\boldsymbol{\Omega} = \sigma^2 \boldsymbol{I}_n$, было OLS, а формально самый худший случай, когда $\boldsymbol{\Omega} = \begin{pmatrix} \sigma_1^2 & \sigma_{1;2}^2 & \\ \sigma_{1;2}^2 & \ddots & \\ \sigma_{1;n}^2 & \cdots & \sigma_n^2 \end{pmatrix}$ — симметричная положительно определённая матрица. Тогда $\hat{\vec{\beta}} = (X'\Omega^{-1}X)^{-1}X'\boldsymbol{\Omega}^{-1}\vec{Y}$. Это называется обобщённый метод наименьший квадратов, GLS.

Скалярное произведение в ортогональном пространстве равно $x_1x_2 + y_1y_2$. А в неортогональном мы между произведениями вставляем ещё матрицу перехода (общий вид — билинейная форма). $\boldsymbol{\Omega}^{-1}$ — тензор второго ранга, и у нас меняется метрика. От одних векторов мы переходим к другим. А дальше мы можем преобразованиями вернуть её в ортогональный вид — $\begin{pmatrix} \sigma_1^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_n^2 \end{pmatrix}$, — затем собственными векторами пронормировать до единичной. Выражение это красивое, но 20 лет назад на него смотрели с ужасом. Тогда это было непосильной задачей, поэтому было желание убежать в OLS. Любую симметричную матрицу можно представить в виде произведения $\boldsymbol{\Omega}^{-1} = \boldsymbol{P}'\boldsymbol{P}$. Подставим: $\hat{\vec{\beta}} = \boldsymbol{(PX)'(PX)}^{-1}\boldsymbol{(PX)'P}\vec{Y}$, $\boldsymbol{PX} = \tilde{\boldsymbol{X}}$, $\boldsymbol{P}\vec{Y} = \tilde{\vec{Y}}$, $\hat{\vec{\beta}} = (\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}'\tilde{\vec{Y}}$.

И в это тильдованное пространство мы ходим только вычислять. Тогда $\boldsymbol{\Omega}^{-1} = \begin{pmatrix} 1/\sigma_1^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1/\sigma_n^2 \end{pmatrix}$, $\boldsymbol{P} = \begin{pmatrix} 1/\sigma_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1/\sigma_n \end{pmatrix}$. А что это собой представляет?

Пусть $Y_i = \beta_0 + \beta_1 X_{i1} + \ldots + \beta_k X_{ik} + \varepsilon_i$. Поделим каждое наблюдение на σ_i. $\frac{Y_i}{\sigma_i} = \beta_0 \frac{1}{\sigma_i} + \beta_1 \frac{X_{i1}}{\sigma_i} + \ldots + \beta_k \frac{X_{ik}}{\sigma_i} + \frac{e_i}{\sigma_i}$, $\tilde{Y}_i = \beta_0 \cdot \tilde{1}_i + \beta_1 \tilde{X}_{i1} + \ldots + \beta_k \tilde{X}_{ik} + \nu_i$. В метод $P'P$ стали ходить сейчас с редкостью.

> «По-моему, совсем несложно представить тильдованную единичку!»

Существенное замечание: мы могли ещё вытащить общий множитель: $\hat{\sigma}^2 = \frac{RSS}{\mathrm{df}}$. Мы строим регрессию, $k+2$ наблюдений отбрасываем. С одной стороны, информации осталось много, а с другой, дисперсий мало, так как надо хотя бы их $\frac{n(n-1)}{2}$. Из наблюдений мы не можем вытащить дисперсии. Поэтому нужны либо априорные дисперсии, либо вычисленные. Третий путь — если гетероскедастичность имеет довольно упрощённую функциональность (на уровне линейной), то для линии нужно две точки.

Здесь везде расписано про дисперсию и дисперсию ошибок.

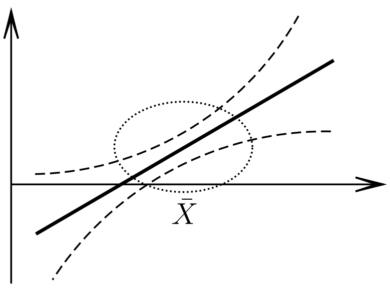
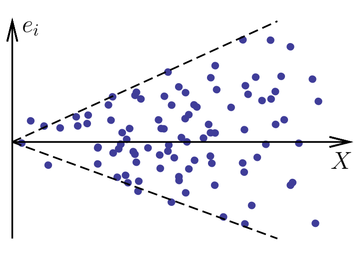


Рис. 1. Средняя часть

> «Ну-ка мы будем среднее звено вышибать: поднялся, с вещами и сюда. Следующий шаг — в угол. А высшая мера…» — «Ремень?»

Ошибок у нас нет, есть только остатки. А вот сказать, что сказать, что оценки — ошибка ошибок, — это некорректно. Оно ведь всё случайно. Только лишь в совокупности своей остатки отчасти объясняют поведение в совокупности. Самый маленький остаток в среднем, а на краях они большие, поэтому массив остатков описывает ошибки в средней части. $\{e_i\} \sim \widehat{\{\varepsilon_i\}}$. Далее, $\varepsilon_i^2 \sim \sigma_i^2 \sim \hat{e}_i^2$. Пусть у нас есть натыканные e_i, когда проблем нет. Гетероскедастичность — это функциональная зависимость. Но плохо подрисовывать рога функции, так как зависимости никакой может и не быть. Информацию о наклоне мы можем выявить, всего две точки.



Как выявить гетероскедастичность, если мы умеем строить регрессию? Если взять зависимость между e_i и X_i, то мы получим ось икс. Мы кофе заварили, получили гущу, на ней можно гадать, но сделать из неё кофе мы уже не можем. Чайный пакетик можно заварить около семи раз, а на восьмой пакетики всплывают, чтобы посмотреть на этого жлоба.

Поэтому мы берём модуль остатков или квадрат, что есть одно и то же. $|e_i| = \alpha + \beta X_i + \delta_i$. Если β значимый, то это гетероскедастичность. Мы сможем даже оценить будущие ошибки. Идея теста Глейзера (1963) лежит на поверхности: если что-то надо сделать, то самое глупое — это интересоваться, можно ли так делать. Глейзер делает такое: он пытается вместо X подставлять разные функции, e_i в квадрат возводить. Но его изначально волновало избавление от гетероскедастичности. Глейзер — человек от сохи, который влез в науку только потому, что пришлось это сделать. Потом приходят обычные математики, которые: «Фи! Почему здесь пользуетесь t-статистиками?» 9 лет математики воздыхали по поводу правильности. Годлфельд и Квант придумывают в 1968 тест. Он находит гетероскедастичность, но не говорит, как исправлять. Берите говёный Глейзер, и там исправлять будете. У Глейзера просто не было богатого опыта отбиваться от тех, которые фикают до полного пуризма. Мы будем гетероскедастичность тестировать на 5 %. А если она меньше 1 %, то нам на неё наплевать. Гетероскедастичность смещает нам оценку дисперсии коэффициента, меняет доверительный интервал. Если гетероскедастичность 5 %, то тогда доверительный интервал поменялся так (прикиньте на глаз изменение ширины окошка):

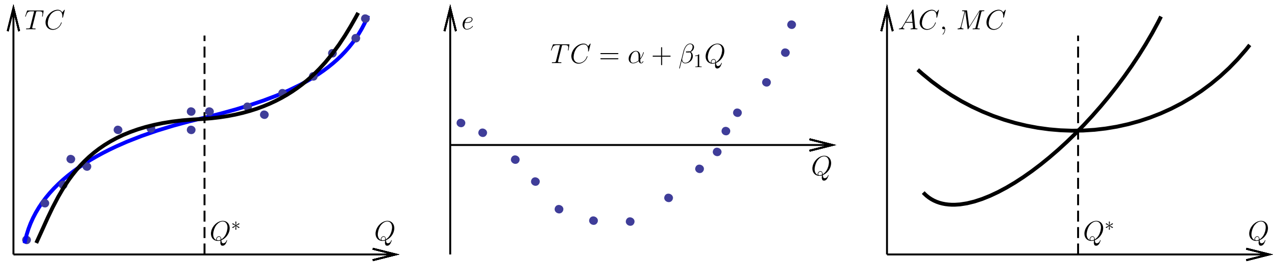
$$| \qquad\quad | \;\Rightarrow\; | \qquad\quad | \quad \text{или} \quad | \qquad\quad | \;\Rightarrow\; | \qquad\quad |$$

Даже если у нас ярковыраженная гетероскедастичность, то у нас поменяется нормальность на одну десятую, поэтому переходят к асимптотике. А если она не есть ярковыраженная, то ничего страшного.

Тест Парка. $e_i^2 = AX^\beta$. Здесь изначально шли, что были логарифмические преобразования. Но самое главное только одно: не бояться делать. $\ln e_i^2 = \alpha + \beta \ln X_i + \delta_i$. Если же гетероскедастичность пропорциональна иксу, то тогда можно просто поделить на икс. Но давайте по земле ходить. А кушать мы по земле любим. А потом думаем, как боги на небе элем питаются: они до сытости его едят или нет?

Вторая задача. (1) Поставьте галочку, чтобы были графики остатков. Сделайте его нормального размера. Это визуализация. В идеале соотношение должно быть 2,35 : 1, потому что у Мамонтова в гараже кинотеатр имеет именно такое соотношение. (2) Glejser. По X, X^2, $\frac{1}{X}$, $\sqrt{X}$. (3) Парк. (4) Голдфельд—Квандт, 6–4–6. (5) Бройш—Пейган—Годфри. Первая версия: $Z_1 = \hat{Y}$, $Z_2 = \hat{Y}^2/$ ВТорая версия: $Z_1 = X_1$, $Z_2 = X_2$. (6) Уайт, X, X^2, все квадратичные члены. Если два теста выявляют, исправляйте.

Но эти все тесты никто ничего не делает. Сейчас есть процедура Уайта. Печатается значение дисперсии без учёта и с учётом гетероскедастичности. И значение теста через хи-квадрат. Если учитываем, то исправляем дисперсии коэффициентов. А можно закрыть глаза в 95 % случаев. В остальных мы заведомо знаем, какого рода функция должна вылезать в гетероскедастичности. Исправляем в OLS взвешенным методом.



Первая задача. Имеется кубоподобная $TC(Q)$. Если сделать регрессию, то на остатках будет остаточность какого-то такого типа, как на рисунке выше. Здесь неправильная функциональная форма. Да всегда она такая. Модель данных неадекватна устройству данных. Матожидание эпсилонов не 0, дисперсии разные, так как постоянное смещение, ковариация между собой всегда есть, а чего там живое остаётся? Только что иксы независимы. Естественно, это требует исправления. Поэтому надо выходить на $TC = \alpha + \beta_1 Q + \beta_2 Q^2 + \beta_3 Q^3 + \varepsilon$ — полиномиальная регрессия. Q и Q^2 разные там, где у функций производные нулевые. Уходим подальше от особенностей — всё становится похожим. Особенности — это перегибы, нули. Если они совпадают, то это одна и та же функция.

> «В чём особенность того, что вы сидите на стуле? Стул такой (показывает седло), а оно такое (показывает два бугра)».

Здесь VIF'ы у всех будут одинаковыми: 150, 300, 900. А то тут торгуются: 13, 12… А был разговор о том, что VIF 3 — это уже серьёзно. $\widehat{TC} = \hat{\alpha} = \hat{\beta}_1 Q + \hat{\beta}_2 Q^2 + \hat{\beta}_2 Q^3$. А те, у кого будут маркеры на прямой или не будет их на наблюдениях-точечках, вне зависимости от того, были они на семинаре или нет, уже знают, что они будут делать. $\widehat{MC} = \widehat{TC}' = \hat{\beta}_1 + 2\hat{\beta}_1 Q + 3\hat{\beta}_3 Q^2$. $\widehat{AC} = \frac{\widehat{TC}}{Q}$, $Q^* = -\frac{2\hat{\beta}_2}{6\hat{\beta}_3}$. Мамонтов понял, что вершина параболы — это задница, потому что у нас проблемы с ментальностью, и мы это поймём. Вот получаем мы AC, MC, и тогда $TC = \alpha' + \beta'(Q - Q^*)$ будет обладать более точными альфой и бетой. Квадратичный член во второй регрессии будет почти ноль, и надо будет его туда ручками положить. Вот тут получился ноль — убрали симметрию. А в первой регрессии его убирать нельзя. А убирание того, кого нет, ещё уменьшает дисперсию. Кота уже нет — улыбка осталась. Коэффициента нет, а дисперсия ещё осталась от него. Если мы возьмём линейную комбинацию, то получим наигрыш на значениях производной. Вот за эту производную можно было бороться. Но на этом задача не заканчивается.

Нам упростили жизнь и не стали давать сотню наблюдений. Нам одинаковые наблюдения слили в одно, посчитали среднее. А легче было бы по одному считать. Дисперсия среднего: $\sigma^2_{\bar X} = \frac{\sigma^2}{n}$. На седьмом задании у нас это спросят. Значит, это связано с объёмом выборки. Поэтому $\sigma^2_i \sim \frac{1}{n_i}$, и последнюю модель надо пересчитать взвешенную модель. Получается, каждое значение наблюдения получено по своей выборке — n_i. И из этого мы оцениваем поведение дисперсии. Мы твёрдое не знаем, мы знаем только изменение. А весы будут как корень: $w_i = \sqrt{n_i}$.

Следующая тема будет после отгуляния сессии. Это будет 2-го числа, будет не 1-го, это очень важный момент. Вот это задание начнётся приниматься до сессии, будет один день после сессии, а 9-го числа будет уже поздно. Текущее задание (7-е) будет приниматься без штрафа всю следующую неделю.

11 Семинар 11

Зависимость ошибок между собой. Наши наблюдения коррелируют между собой. Зависимость такого рода: линейная; чаще всего зависимость первого порядка, то есть только с соседями. Дальние порядки рассматривают гораздо реже. $Y = \alpha + \beta X + \varepsilon$. Будет иметься в виду индекс t. Временной ряд предполагает упорядоченность по переменным. Неявно она бывает в индексе, явно иногда в переменную входит. Когда $\varepsilon = f(\varepsilon_{t-1}, \varepsilon_{t-1}, \ldots)$,то это автокорреляция. Если $Y = \alpha + \beta_1 X_t + \beta_2 Y_{t-1} + \varepsilon_t$, то это авторегрессионная схема. Если $Y_t = \alpha + \beta_1 X_t + \beta_2 X_{t-1} + \beta_3 X_{t-2} + \varepsilon_t$, то это регрессия с лаговыми переменными.

Автокорреляция сильно пересекается в временными рядами. Оные сделали эконометрику эконометрикой: денежные проблемы решались через временные ряды — самая хлебная часть, средства для освоения ближе всего. Мы рассматриваем автокорреляцию. Проблема разгребать случайность в правой части тоже серьёзная. Независимость икса от ошибки — это нам уже не снится. Поэтому не всё здесь можно вчистую МНК проработать. Но чаще его и используют. Мы рассмотрим автокорреляцию от ε_{t-1}. Можно так записать: $\varepsilon_t = \rho\varepsilon_{t-1} + \nu_t$, $\mathbb{E}(\nu_t) = 0$. Нет свободного члена, возникает гетероскедастичность. Какой смысл у ρ?

$$\mathbb{E}(\varepsilon_t\varepsilon_{t-1}) = \rho\mathbb{E}(\varepsilon^2_{t-1}) + \underbrace{\mathbb{E}(\nu_t\varepsilon_{t-1})}_{=0}, \quad \rho = \frac{\mathrm{Cov}(\varepsilon_t;\varepsilon_{t-1})}{\sigma^2}$$

ρ — коэффициент автокорреляции первого порядка. Схема называется AR(1), или марковская авторегрессионная схема первого порядка. Автокорреляция называется авторегрессией, потому что $\varepsilon_t = f(\varepsilon_{t-1})$. Если авторегрессионная схема второго порядка, то тогда будет ε_{t-1}. Неявная схема: $\varepsilon_t = \nu_t + \lambda\nu_{t-1}$, где ν — независимые случайные величины с матожиданием 0. Она неявная, и она называется скользящей средней первого порядка: MA(1) (Moving Average). Бывают смеси: ARMA(1,2) = AR(1) MA(2).

Автокорреляция влияет только на дисперсию. Появляется проблема более ярковыраженная с эффективностью, если сравнивать с гетероскедастичностью. Здесь сильно может увести в сторону. При появлении автокорреляции мы пытаемся её исправить. Из-за чего она может возникнуть? Лаговый тип переменных: то, что сегодня, зависит от вчера. Во временных рядах эти зависимости очень характерны.

Появление автокорреляции означает: если рассматривать матрицу $\mathbf{\Omega}$ — дисперсионно-ковариационную матрицу наблюдений — то тогда можно вытащить σ^2 за скобку, когда дисперсия одинаковая. **** МАТРИЦА С помощью GLS получаем оценки $\vec\beta = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\vec Y$. Матрица оборачиваемая, задача решаемая, но это не сразу слепить удастся.

Чтобы избежать этого и решить при помощи OLS, можно провести такой подход. $Y_t = \alpha + \beta X_t + \varepsilon_t$, $Y_{t-1} + \alpha + \beta X_{t-1} + \varepsilon_{t-1}$. Второе уравнение домножим на ρ и вычтем. $Y_t - \rho Y_{t-1} = \alpha(1-\rho) + \beta(X_t - \rho X_{t-1}) + \underbrace{\varepsilon_t - \rho\varepsilon_{t-1}}_{\nu_t}$, $\tilde Y = \alpha\tilde 1 + \beta\tilde X_t + \nu_t$, а далее OLS. Это квазиразностное уравнение (а если $\rho = 1$, то это чисто разность на уровне первого порядка в производных). Мы перевалили в OLS и расплатились за это первым наблюдением. Когда ρ большие, то оценка производной в разностных схемах очень ненадёжная. Что точнее считается — производная или интеграл? Вычислительно интеграл — это сумма, и ошибка то туда, то туда, и ошибка засредняется. Если считать с округлением, то ошибка растёт, как $\sqrt{n}$. А когда считать производную… Если считать производную разностно, то болтанка будет несусветная. На полиноме зацепляются все точки разом. А на производную накладывается случайность. При больших ρ даже потеря одной точки — это плохо. Терять точку не хотелось, поэтому было предложено добавлять одну псевдоточку. Поправка Прейса—Уинстона: было $t = 2; \ldots; n$, а $\tilde Y_1 = \sqrt{1-\rho^2}Y_1$, $\tilde X_1 = \sqrt{1-\rho^2}X_1$, $\tilde 1_1 = \sqrt{1-\rho^2}$, $1_2 = 1 - \rho$. Информационно мы ничего не добавляем, а вычислительно наигрыш есть. Техника рождается из идеи согласованного импеданса, чтобы уравнения для производной мягко старотовали.

Это всё касается случая, когда мы ρ знаем, когда утром какой-то голос нашептал это, как домысел. Квазиразностно очень легко вторым способом устранить автокорреляцию. Мы теряем две точки и ещё радуемся, что нам повезло. В финансовых уравнениях стандартно до четвёртого порядка доходят, когда мы годом трясём над квартальными данными.

Просьба, напоминание, рекомендация: преобразование Койка немножко похоже на это. Но если это называть преобразованием Койка, то будут лупить по рукам. Как это правильно называть? *Квазиразностное уравнение*!

Поехали по тестам. Когда были проблемы мультиколлинеарности, была проблема значимость. Гетероскедастичность — смотрели остатки. Автокорреляция ожидается, и будем пытаться наблюдать в остатках.

Гетероскедастичность — $e(X)$, и неважен порядок X. Тут существенен порядок. Поведение такого типа — положительная автокорреляция. Такого (*****) — отрицательная. А чем это отличатся от того случая, когда X Были по оси? Как понять, что зависимость не функциональная, а именно автокорреляция? Тесты, которые ловят автокорреляцию, могут что-то ловить на упорядоченных иксах. Если с обувью мы посмотрели в другой набор, то там то же самое. Устойчивость поведения гетероскедастичности функциональна. У одного фаза одна, у другого фаза другая. А-ля синусоида плывёт. Автокорреляция может начать изгибаться волной с разной фазы. Если рассматривать с точка зрения матожидания ошибки, то оно равно нулю. А если функциональная ошибка, то ожидание не ноль. То, что фаза плывёт, обеспечивает то, что фаза равна нулю. Чаще рассматривают график вида AR(1) для остатков. $e_t = \rho e_{t-1} + \nu_t$. В идеале должно быть равномерное облако.

***

ОБратите внимание: это одна и так же переменная. Они переходят одна в другую. По осям — квадрат, точность — 5 %. У кого будет не квадрат, учтитк, у Мамонтова прессовка и оквадрачивание всегда получались. Это будет первое ρ, которое мы можем получить, — ρ из схемы AR(1).

Теперь тест Дарбина—Уотсона, а ещё лучше — статистика.

$$dw = \frac{\sum (e_t - e_{t-1})^2}{\sum e_t^2}$$

По-французски — это Дюрбан. Мы не одиноки в обзываниях Дёрбина по-своему. Эта статистика необычна и нетривиальна по всем статьям сразу. Когда $\rho = 1$, то будет 0. Когда $\rho = -1$, то будет 4. Какая ещё статистика так меняется? На устном экзамене это была последняя соломинка: в каких пределах меняется статистика Дарбина—Уотсона? Никому и в голову не придёт, что от 0 до 4! Там ладно от -1 до 1 или от 0 до 1. Далее, этот тест реально проверяет 4 гипотезы, которые мы записываем как две: $\mathcal{H}_0^+ : \rho > 0$, $\mathcal{H}_1^+ : \cancel{\rho > 0}$, $\mathcal{H}_0^- : \rho < 0$, $\mathcal{H}_1^- : \cancel{\rho < 0}$. Если мы попадаем не в край, то гипотеза об отсутствии или наличии автокорреляции принимается. Справа принимается отрицательная, слева принимается отсутствие, а в зоне неопределённости мы проводим дополнительный анализ.

Ограничение теста Дарбина—Уотсона: применяется, когда автокорреляция только первого порядка, когда есть свободный член, когда отсутствует авторегрессия, когда отсутствует лаговая переменная, когда нормальность и гомоскедастичность. Надо, чтобы не было провалов в данных. Можно работать и на обычных иксах, и он будет ловить не гетероскедастичность, а функциональную зависимость.

По тестам. Тест серий — это разобрать самим. Он же run test. Первое задание: AR(1), Дарбин—Уотсон, график, остатки от t, остатки от остатков. Тест серий. Тест Бройша—Годфри. Он очень универсальный, позволяет ловить даже скользящее среднее. Идея БПГ: квадрат остатков относительно функций. Тут остатки относительно всех-всех иксов плюс остаток вчерашний, позавчерашний… Рекомендация: сделайте до пятого порядка (не включая), потом переходите на четвёртый, третий. Смотрите, как себя ведёт статистика с R^2, а также что происходит со значимостью коэффициентов перед ρ.

Схема исправления. Итерационная процедура Кокрейна—Оркутта. $Y = \alpha + \beta X_t + \varepsilon_t$, получили $\hat{\alpha}$, $\hat{\beta}$, считаем вычисленные $\hat{Y} = \hat{\alpha} + \hat{\beta} X$, потом считаем $e_t = Y_t - \hat{Y}_t$, по схеме AR(1) считаем $e_t = \rho e_{t-1} + \nu_t$. С помощью ρ при помощи псевдоразностной схемы получаем $\tilde{Y} = \alpha \tilde{1} + \beta \tilde{X} + v_t$ с поправкой. Получаем новые $\hat{\alpha}$, $\hat{\beta}$, снова вычисляем $\hat{Y}$ новое. Прерываемся, когда $\Delta\rho = 0{,}01$ или $n \leqslant 10$. В пакете тоже можно задавать это ограничение. Двухшаговый: найти ρ и новые коэффициенты.

Двухшаговый Дарбин — сделать, применить. И ещё. Решётку Хилдрета— Лу. Брать разное $\rho \in [-1; 1]$ с шагом 0,1, выбираем RSS минимальное, потом берём шаг 0,01, а потом рисуем график. Примерно такой: ***

Вторая задача — степень освоения инвестиций. $k_t = \alpha + \beta_1 I_{t-0} + \beta_2 I_{t-1} + \beta_3 I_{t-2}$. Освоенный капитал — оборудование приобретено, запущено, начало возвращать средства. Где прерваться? До какой глубины лаги считать? Тинберген говорит: строить регрессию, пока какая-то из переменных станет незначимой. Самое плохое — когда из-за мультиколлинеарности не станет незначимой какая-то бета. Отрицательный знак — одно и то же. Средний лаг: $\frac{\beta_1 \cdot 0 + \beta_1 \cdot 1 + \beta_3 \cdot 3}{\beta_1 + \beta_2 + \beta_3}$. Спросят, как средний лаг мерится по преобразованию Койка.