

Привет, дорогой студент!

Я буду помогать тебе быть в курсе всех новостей с семинаров Б. Б. Демешева и лекций Демидовой О.А. по эконометрике в 2017-2018 году. Конечно же я не претендую на авторское право, потому что все нижеизложенное - это просто мой взгляд на эконометрику, который создан Демешевым, Демидовой и Доугерти. Всем мира и опенсурса!

Давай же начнем наше увлекательное приключение в мир регрессий, тестов и красивых графиков! :)

Так как записи семинаров ведутся не с первого сентября, а с 27 ноября 2017 года, то переписывать сюда всё с самого начала трудно, поэтому я прикладываю краткий summary всего самого важного, что было пройдено на лекциях и семинарах за пропущенный период. Если ты нашел неточность или ошибку, то скорее пиши Борису Борисовичу или мне.

Конвенция об обозначениях

Во всех последующих записях мы используем следующие обозначения.

1. a, b - это векторы $(n \times 1)$ из констант,
2. A, B - это матрицы $(n \times k)$ из констант,
3. α, γ - это скаляры, в виды констант,
4. r, s - это тоже векторы $(n \times 1)$, но состоят они из случайных величин,
5. R, S - это матрицы $(n \times k)$, каждый компонент которых теперь случайная величина,
6. t - это скаляр в виде случайной величины.

Чтобы быть успешным эконометристом (эконометрессой), нужно все-таки выучить вспомнить школьную программу, линейную алгебру и курс ТВ и МС.

Начнем со старой доброй геометрии.

1. Теорема Пифагора
2. Теорема Фалеса
3. Теорема о трех перпендикулярах
4. Теорема косинусов и синусов
5. Скалярное произведение векторов и кое-что еще

Линейная алгебра

1. Матрицы

1.1 Сложение и вычитание

1.2 Умножение матриц

1.3 Обращение матриц

1.4 Транспонирование

1.5 Определитель матрицы

1.6 След матрицы

2. Собственные числа и векторы матриц

ТВИМС

1. Вероятности

1.1 Безусловные вероятности

1.2 Условные вероятности

2. Характеристики распределений и зависимостей

2.1 Математическое ожидание

2.2 Дисперсия

2.3 Ковариация

2.4 Корреляция

3. Проверка гипотез

4. p-value

Еще немного теории

Уже не терпится строить регрессии? Всё, всё, это последний ~~но это неточно~~ ликбез. Давай вспомним про нашего старого друга с первого курса, про дифференциал.

Правила дифференцирования векторов и матриц

1. $d(A \times R \times B) = A \times d(R) \times B$

2. $d(\alpha \times a) = 0(n \times 1)$

3. $d(\alpha \times A) = 0(n \times k)$

4. $d(R + S) = dR + dS$

5. $d(R)' = (dR)'$, где ' означает транспонирование

6. d сам знает, как дифференцировать сложную функцию.

7. $d\left(\frac{R}{t}\right) = \frac{d(R) \times t - R \times d(t)}{t^2}$
 8. $d(r' \times A \times r) = r' \times (A' + A)d(r)$ (настоятельно рекомендую проверить этот результат, а то мало ли что)
 9. $d(R^{-1}) = -R^{-1} \times dR \times R^{-1}$, где R^{-1} это обратная матрица для R
-

Вот теперь мы готовы к изучению эконометрики!

Начнем с самой главной лошадки эконометриста - МНК

Что такое регрессия?

Начать нужно с того, а что такое регрессия и зачем она нужна. Много чего в жизни можно описать уравнением, но чаще это дает нам статический результат вот здесь и сейчас. Примером такого статического и 100% верного описания будет любая теорема, вот у меня любимая теорема - теорема косинусов, которая утверждает, что если ты знаешь длины двух сторон и косинус угла между ними, то ты без особого труда найдешь третью сторону в треугольнике. Такую зависимость очень просто выразить в виде уравнения, которое описывает зависимость, но является ли оно регрессией? Нет, не является. Почему? Потому что регрессия это не просто описание зависимости между объясняемой и объясняющими переменными, это ожидание значения зависимой переменной при заданных объясняющих переменных, то есть регрессия - это условное математическое ожидание от некоторого процесса, содержащего случайную величину. Заметим важную вещь, что пока мы работаем только с линейными связями. Вот мы и получили первый важный вывод:

$$E(y_i|x_i) = E((\beta_0 + \sum_{i=1}^n \beta_i \times x_i + \epsilon_i)|x_i) = \hat{\beta}_0 + \sum_{i=1}^n \hat{\beta}_i \times x_i = \hat{y}$$

OLS без матриц

В чем смысл метода МНК? Считая, что вся случайность i -го наблюдения и непредсказуемость находится в ϵ_i , то мы можем минимизировать ошибку нашего прогноза, сложив и возведя в квадрат разницу между истинным результатом модели и его оценкой. Зачем нужен квадрат? Что бы не "схлопывались" ошибки прогноза.

Формализация МНК:

$$\sum_{i=1}^N (y - \hat{y})^2 \rightarrow \min_{\hat{\beta}_j, j \in Z}$$

Ф.О.С. дают нам оценки для всех $\hat{\beta}_j$

Если регрессия парная и содержит константу, тогда справедливы следующие результаты минимизации

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \times \bar{x}$$

$$\hat{\beta}_1 = \frac{\sum_{i=0}^N (x_i - \bar{x}) \times (y_i - \bar{y})}{\sum_{i=0}^N (x_i - \bar{x})^2} = \frac{cov(\hat{x}_i, y_i)}{var(\hat{x})}$$

У МНК есть красивая геометрическая интерпретация, которая верна как для парного, так и для многомерного случая, которую я нарисую внизу :)

#Тут будет график. Скоро.

В трехмерном пространстве есть 4 n-мерных вектора: y, x, z и 1 , где $y \notin Lin(x, z, 1)$

Если спроектировать y на $Lin(x, z, 1)$, то получим

\hat{y} , а вектор, который соединяет y и \hat{y} является вектором ошибок прогноза — RSS . Теперь спроектируем \hat{y} на единичный вектор, тем самым мы получим вектор средних значений — \bar{y} — через вычитание из вектора \hat{y} объясняемых квадратов отклонений — ESS . Получается, что, так как $RSS \perp \hat{y}$, а $ESS \perp \bar{y}$, тогда проекция y на 1 дает в результате вектор средних значений y и равен $ESS + RSS$

Вот так мы получили еще три столба для эконометрики

$RSS = \sum_{i=1}^N (y - \hat{y})^2$ - это необъясненная сумма квадратов остатков регрессии, мы ее минимизируем

$ESS = \sum_{i=1}^N (\hat{y} - \bar{y})^2$ - это объясненная сумма квадратов отклонений

$TSS = \sum_{i=1}^N (y - \bar{y})^2$ - это общая сумма квадратов отклонений

И соотношения между ними:

$$TSS = RSS + ESS$$

$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$, $R^2 \in [0; 1]$ - это коэффициент детерминации, то есть показатель качества подгонки или (культурным языком) соответствия нашей модели к реальности. При добавлении в уравнение регрессии любого фактора, R^2 не уменьшится.

Интересно заметить, что $R^2 = \frac{\sum_{i=1}^N (\hat{y} - \bar{y})^2}{\sum_{i=1}^N (y - \bar{y})^2} = \rho_{y, \hat{y}}^2$. Это также имеет красивую геометрическую интерпретацию, которую я привожу ниже.

#тут будет график. Скоро

То же самое пространство, но только оно перевернуто с ног на голову. Заметим, что если $R^2 = \frac{ESS}{TSS}$, то $R^2 = \cos^2(\phi)$, то косинус можно расписать немного иначе, а именно

$$\cos(\phi) = \frac{(\sum_{i=0}^n (y - \bar{y}) \times (\hat{y} - \bar{y}))^2}{\sum_{i=0}^n (y - \bar{y})^2 \times \sum_{i=0}^n (\hat{y} - \bar{y})^2} = \frac{cov(y, \hat{y})}{\sigma_y^2 \times \sigma_{\hat{y}}^2} = \rho_{y, \hat{y}}$$

Заметим еще, что задачи $RSS \rightarrow \min$ и $R^2 \rightarrow \max$ эквивалентны.

Еще раз повторяю. Это справедливо только в том случае, если есть константа в построенном прогнозе!

Почему пропал ϵ ?

Действительно, а почему я так смело убрала случайную величину? Магия, жульничество, тайна за семью печатями? Нет, все проще. Пришло время знакомиться с OLS и теоремой Гаусса-Маркова, в рамках которой мы будем работать довольно долго.

Теорема Гаусса-Маркова (ТГМ)

Работает для регрессионной модели, если выполняются следующие предпосылки

- Модель правильно специфицирована, то есть в нее включен только необходимый x , объясняющий y
- $\exists x_i$, что $x_i \neq x_j, i, j \in Z$
- $\forall \epsilon_i, E(\epsilon_i) = 0$
- $\forall \epsilon_i \text{ Var}(\epsilon_i) = \sigma_{\epsilon_i}^2$
- $\forall \epsilon_i \epsilon_i \sim N(0, \sigma_{\epsilon_i}^2)$ - случайный член ~~гомоген~~гомоскедастичен
- $\text{cov}(\epsilon_i, \epsilon_j) = 0$ - отсутствие автокорреляции
- Объясняющая переменная содержит некоторую вариацию. Это значит, что x не является константой, иначе мы бы не смогли рассчитать коэффициенты регрессии, так как напомним, что

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \times \bar{x}$$

$$\hat{\beta}_1 = \frac{\sum_{i=0}^N (x_i - \bar{x}) \times (y_i - \bar{y})}{\sum_{i=0}^N (x_i - \bar{x})^2} = \frac{\text{cov}(x_i, y_i)}{\text{var}(x)}$$

А это значит, что $\bar{x} = x$, а следовательно $\hat{\beta}_1$ была бы неопределена, так как числитель и знаменатель дроби были бы равны 0

Это гарантирует, что стохастический компонент содержится только в ϵ .

Для такой модели оценки МНК:

- Best

$$\forall \hat{\beta}_i, \text{MSE}(\hat{\beta}_i) \leq \text{MSE}(\tilde{\beta}_i)$$

- Linear

Модель линейна по y

- Unbiased

$$E(\hat{\beta}_i) = \beta_i$$

- Estimator

С английского - оценка параметра

На веру ничего и никогда принимать нельзя, поэтому докажем каждый из пунктов последовательно, но перед этим сделаем подготовительные упрощения:

$$\begin{aligned} \text{var}(\hat{\beta}_0) &= \text{var}(\bar{y} - \hat{\beta}_1 \times \bar{x}) = \text{var}(\hat{\beta}_1 \times \bar{x}) = \\ \bar{x}^2 \times \text{var}\left(\frac{\sum_{i=0}^N (x_i - \bar{x}) \times (y_i - \bar{y})}{\sum_{i=0}^N (x_i - \bar{x})^2}\right) &= \sigma_\epsilon^2 \left(\frac{\sum_{i=1}^n x_i^2 + n \times \bar{x}^2}{n \times \sum_{i=1}^n (x_i - \bar{x})^2}\right) = \sigma_\epsilon^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2}\right) \end{aligned}$$

Тут надо, кое что вспомнить: $\hat{\beta}_1 = \beta_1 + \frac{\epsilon_i \times (x_i - \bar{x})}{\sum_{j=1}^n (x_j - \bar{x})^2}$

$$\begin{aligned} \text{var}(\beta_1) &= E([\hat{\beta}_1 - E(\hat{\beta}_1)]^2) = E(\hat{\beta}_1 - \beta_1)^2 = \\ E\left(\left[\sum_{i=1}^n \frac{\epsilon_i \times (x_i - \bar{x})}{\sum_{j=1}^n (x_j - \bar{x})^2}\right]^2\right) &= E\left(\sum_{i=1}^n \frac{\epsilon_i^2 \times (x_i - \bar{x})^2}{\sum_{j=1}^n (x_j - \bar{x})^4}\right) + \sum_{j=1}^n \sum_{i=1}^n \frac{\epsilon_i \times (x_i - \bar{x})}{\sum_{j=1}^n (x_j - \bar{x})^2} \times \frac{\epsilon_j \times (x_j - \bar{x})}{\sum_{j=1}^n (x_j - \bar{x})^2} = \\ \left(\sum_{i=1}^n \frac{E(\epsilon_i^2) \times (x_i - \bar{x})^2}{\sum_{j=1}^n (x_j - \bar{x})^4}\right) &+ \sum_{j=1}^n \sum_{i=1}^n \frac{E(\epsilon_i) \times (x_i - \bar{x})}{\sum_{j=1}^n (x_j - \bar{x})^2} \times \frac{E(\epsilon_j) \times (x_j - \bar{x})}{\sum_{j=1}^n (x_j - \bar{x})^2} \end{aligned}$$

С учетом, что $E(\epsilon_i)^2 = \sigma_{\epsilon_i}^2$, $E(\epsilon_i \times \epsilon_j) = 0$ получаем, что

$$\text{var}(\beta_1) = \frac{\sigma_\epsilon^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

Ты герой, если дочитал до этого места, а не пропустил доказательство.

Давайте будем считать, что $w_i = \frac{x_i - \bar{x}}{\sum_{j=1}^n (x_j - \bar{x})^2}$, тогда будут справедливы следующие свойства:

$$1. \sum_{i=0}^n w_i = \frac{\sum_{i=0}^n (x_i - \bar{x})}{\sum_{j=1}^n (x_j - \bar{x})^2} = 0$$

$$2. \sum_{i=0}^n x_i \times w_i = 1$$

$$3. \sum_{i=0}^n w_i^2 = \frac{1}{\sum_{j=1}^n (x_j - \bar{x})^2}$$

Перепишем наши оценки параметров в новом виде

$$\hat{\beta}_0 = \sum_{i=1}^n \left(\frac{1}{n} - w_i \times \bar{x} \right) \times y_i$$

$$\hat{\beta}_1 = \sum_{i=1}^n w_i \times y_i$$

Best

Рассмотрим $\widetilde{\beta}_1 = \sum_{i=1}^n \widetilde{w}_i \times y_i$, которая является несмещенной оценкой для β_1 , ф значит справедливо $\sum_{i=1}^n \widetilde{w}_i = 0$, $\sum_{i=1}^n \widetilde{w}_i \times x_i = 1$.

$$var(\widetilde{\beta}_1) = var\left(\sum_{i=1}^n \widetilde{w}_i \times y_i\right) = \sigma_\epsilon^2 \times \sum_{i=1}^n \widetilde{w}_i^2 \rightarrow \min$$

Рассмотрим поближе $\sum_{i=1}^n \widetilde{w}_i^2 = \sum_{i=1}^n (\widetilde{w}_i - w_i + w_i)^2 =$

$$\sum_{i=1}^n (\widetilde{w}_i - w_i)^2 + 2 \sum_{i=1}^n (\widetilde{w}_i - w_i) w_i + \sum_{i=1}^n w_i^2 =$$

$$\sum_{i=1}^n (\widetilde{w}_i - w_i)^2 + 2 \sum_{i=1}^n (\widetilde{w}_i w_i - w_i^2) + \sum_{i=1}^n w_i^2 =$$

$$\sum_{i=1}^n (\widetilde{w}_i - w_i)^2 + 2 \sum_{i=1}^n \left(\widetilde{w}_i \times \frac{(x_i - \bar{x})}{\sum_{j=1}^n (x_j - \bar{x})^2} \right) - \frac{1}{\sum_{j=1}^n (x_j - \bar{x})^2} =$$

$$\sum_{i=1}^n (\widetilde{w}_i - w_i)^2 + 2 \left(\frac{\sum_{i=1}^n (\widetilde{w}_i \times x_i) - \bar{x} \times \sum_{i=1}^n (\widetilde{w}_i)}{\sum_{j=1}^n (x_j - \bar{x})^2} \right) - \frac{1}{\sum_{j=1}^n (x_j - \bar{x})^2} =$$

$$\sum_{i=1}^n (\widetilde{w}_i - w_i)^2 + \frac{1}{\sum_{j=1}^n (x_j - \bar{x})^2}$$

В таком случае, только при $\widetilde{w}_i = w_i$, $var(\hat{\beta}_1) = \frac{\sigma_\epsilon^2}{\sum_{i=1}^n x_i^2}$

Linear

Линейность оценки следует из детерминированности x и линейности $\hat{c} \hat{v}(x, \epsilon)$

Unbiased

$$E(\hat{\beta}_1) = E\left(\beta_1 + \frac{\text{cov}(\hat{\epsilon}, x)}{\text{var}(\hat{x})}\right) = \beta_1 + \frac{E(\sum_{i=1}^n (\epsilon - \hat{\epsilon}) \times (x - \hat{x}))}{(n-1) \times \text{var}(\hat{x})} = \beta_1$$

$$E(\hat{\beta}_0) = E(\bar{y} - \hat{\beta}_1 \times \bar{x}) = \beta_0$$

Полезные результаты МНК без матриц

В этой главе мы закинем удочку в тему проверки гипотез, так как именно это задача любого анализа модели: доказать, что построенная модель - это не просто набор из случайно напиханных факторов, которые каким-то образом попались на глаза в процессе исследования, а обдуманые и очень тщательно выбранные факторы, адекватно отображающие интересующий процесс в жизни.

Важно помнить следующие выводы из парной регрессии в условиях ТГМ

1. $\bar{\hat{\epsilon}} = 0, \hat{\epsilon} = y - \hat{y}$
2. $\bar{y} = \bar{\hat{y}}$
3. $\sum_{i=0}^n x_i \times \hat{\epsilon} = 0$
4. $\sum_{i=0}^n \hat{y} \times \hat{\epsilon} = 0$
5. Несмещенная оценка дисперсии ошибки прогноза выражается $\hat{\sigma}_{\epsilon}^2 = \frac{RSS}{n-2}$

Давай подумаем, какая из всех возможных прямых будет ближе всего ко всем точкам - выборке наблюдений?

Конечно, регрессионная, просто по ее определению, ведь разброс остатков для нее минимальный, а значит среднее квадратичное отклонение для разброса еще меньше. Зная это, выведем еще одну несмещенную оценку для $\sigma_{\epsilon_i}^2$.

$$E(MSD(\epsilon_i)) = E\left(\frac{1}{n} \sum_{i=1}^n (\epsilon_i - \bar{\epsilon})^2\right) = \frac{1}{n} E\left(\sum_{i=1}^n (\epsilon_i)^2\right) = \frac{n-2}{n} \times \sigma_{\epsilon}^2$$

$$s_{\epsilon}^2 = \frac{n}{n-2} \times MSD(\epsilon) = \frac{n}{(n-2) \times n} \times \sum_{i=1}^n (\epsilon_i)^2 = \frac{RSS}{n-2} = \sigma_{\epsilon}^2$$

-
6. $\frac{RSS}{\sigma_{\epsilon}^2} \sim \chi_{n-2}^2$

Тут надо вспомнить, из чего же складывается распределение χ_n^2

Если $\forall z_i \sim N(0; 1)$, $z_j \sim N(0; 1)$, $i \in (1, n)$ и при условии, что $cov(z_i, z_j) = 0$, верно $\sum_{i=1}^n z_i^2 \sim \chi_n^2$

Используем метод пристального взгляда и... очевидно!

7. Нормальность случайных возмущений

Последнее неочевидно, поэтому требует доказательства.

Из ТГМ нам известно, что $\epsilon_i \sim N(0, \sigma_{\epsilon_i}^2)$, $cov(\epsilon_i, \epsilon_j) = 0$, это дает нам право считать, что

$$\hat{\beta}_1 = \frac{cov(x_i, y_i)}{var(x)} = \frac{cov(x_i, \beta_0 + \beta_1 \times x_i + \epsilon_i)}{var(x)} = \frac{cov(x_i, \beta_0)}{var(x)} + \beta_1 \times \frac{cov(x_i, x_i)}{var(x)} + \frac{cov(x_i, \epsilon_i)}{var(x)} = \beta_1 + \frac{cov(x_i, \epsilon_i)}{var(x)},$$

поэтому $\hat{\beta}_1 \sim N(\beta_1, \sigma_{\beta_1}^2)$. Аналогичный анализ справедлив и для $\hat{\beta}_0$

8. $\forall i \in [1, 2]$ оценки $\hat{\beta}_i$ и $\hat{\sigma}_{\epsilon}^2$ являются независимыми

Это доказывается через ковариацию

Тестирование гипотез в парной регрессии

Ну вот наконец-то я добралась до второго модуля.

Гипотезы!

Для начала обозначим план работ :)

Всего нам нужно знать 2 гипотезы для случая парной регрессии.

1. Гипотеза о конкретном значении коэффициентов

Рассмотрим только случай двусторонней гипотезы:

$$H_0: \beta_1 = \beta_1^0$$

$$H_a: \beta_1 \neq \beta_1^0$$

Как мы уже знаем $\beta_1 \sim N(\beta_1, \sigma_{\beta_1}^2)$, тогда наша тестовая статистика принимает вид

$\frac{\hat{\beta}_1 - \beta_1^0}{\sigma_{\hat{\beta}_1}} \sim N(0,1)$, но вот же грусть-печаль, нам никак не узнать $\sigma_{\hat{\beta}_1}^2$, так как в нее входит неизвестный никому параметр $\sigma_{\epsilon_i}^2$. Тогда поступим классическим образом и используем оценку неизвестного параметра: $\hat{\sigma}_{\hat{\beta}_1}^2$

В таком случае $\frac{\hat{\beta}_1 - \beta_1^0}{\hat{\sigma}_{\hat{\beta}_1}} \sim t_{n-2}$, тогда доверительный интервал для проверки будет иметь вид:

$$\hat{\beta}_1 - t_{n-2, 1-\frac{\alpha}{2}} \times \hat{\sigma}_{\hat{\beta}_1} \leq \beta_1^0 \leq \hat{\beta}_1 + t_{n-2, \frac{\alpha}{2}} \times \hat{\sigma}_{\hat{\beta}_1}$$

Если заданное в нулевой гипотезе значение беты попадает в доверительный интервал, тогда гипотеза не отвергается.

2. Гипотеза о значимости коэффициента

Рассмотрим только случай двусторонней гипотезы:

$$H_0: \beta_1 = 0$$

$$H_a: \beta_1 \neq 0$$

В таком случае $\frac{\hat{\beta}_1}{\hat{\sigma}_{\hat{\beta}_1}} \sim t_{n-2}$, тогда доверительный интервал для проверки будет иметь вид:

$$\hat{\beta}_1 - t_{n-2, 1-\frac{\alpha}{2}} \times \hat{\sigma}_{\hat{\beta}_1} \leq \beta_1^0 \leq \hat{\beta}_1 + t_{n-2, \frac{\alpha}{2}} \times \hat{\sigma}_{\hat{\beta}_1}$$

Если 0 попадает в доверительный интервал, тогда гипотеза не отвергается или иными словами, коэффициент НЕ значим.

МНК в матрицах.

Тут будет линал, поэтому ~~слабонервным не смотреть~~ нужно кое-что дополнительно ввести.

Да, если бы все в мире описывалось простыми парными регрессиями, то жизнь была бы сказкой, но ~~к сожалению~~ реальность многогранная и парные регрессии используются редко, так как плохо отображают действительность, вместо них используют многофакторные модели или множественные регрессии, поэтому визуализация становится сложной, а взаимосвязи между факторами и регрессантом - не очевидными. Что же нам делать? Вспоминать линейную алгебру конечно!

Не пугайся, мой друг, больно не будет. :)

Для начала определимся с обозначениями, следуя Конвенции:

1. $y_{n \times 1}$ - это вектор регрессант
2. $X_{n \times (k+1)}$ - это матрица констант
3. $\beta_{(k+1) \times 1}$ - это вектор из $k + 1$ коэффициентов
4. $u_{n \times 1}$ - это вектор случайных величин

Поэтому справедливо следующее

1. $y = X\beta + u$
 2. $\hat{y} = X\hat{\beta}$
 3. $\hat{\beta} = (X'X)^{-1}X'y$
 4. $\hat{u} = y - \hat{y}$
 5. $E(u) = 0_{n \times 1}$
 6. $Var(u) = \sigma^2 \times I_{n \times n}$
 7. $RSS = u'u = (y - X\hat{\beta})'(y - X\hat{\beta}) = y'y - 2\hat{\beta}'X'y + \hat{\beta}'X'X\hat{\beta}$
-

Посмотрим теперь на показатели качества подгонки множественной регрессии

В общем-то коэффициент множественной детерминации будет определяться по той же самой формуле, как и раньше $R^2 = \frac{ESS}{TSS}$ и почти все свойства сохраняются с небольшими изменениями.

1. $R^2 = 1 - \frac{RSS}{TSS}$, но опять же оговоримся о надобности константы в модели для работы этого свойства.
 2. $R^2 = \frac{var(\hat{y})}{var(y)} = \sum_{i=1}^k \frac{cov(x_i, y)}{var(x_i)}$
 3. $R^2 = \frac{\sum_{i=1}^N (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^N (y_i - \bar{y})^2} = \rho_{y_i, \hat{y}_i}^2$
 4. $RSS \rightarrow \min \Leftrightarrow R^2 \rightarrow \max$
 5. При добавлении фактора R^2 не уменьшается
-

$$R^2 = 1 - \frac{\sum_{i=1}^n \hat{u}_i (\hat{\beta}_0 \dots \beta_k)}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

$$R^2 = 1 - \frac{\sum_{i=1}^n \hat{u}(\hat{\beta}_0 \dots \beta_{k+1})}{\sum_{i=1}^n (y - \bar{y})^2}$$

Если решить задачу максимизации для двух таких R^2 , то окажется, что второй R^2 не уменьшится. Проверь это сам, а то вдруг я тебя обманываю.

А пока поговорим про факторное разложение R^2 . Сейчас мы будем работать в матрицах, напомним, чтобы получить вектор, который состоит из квадратов каких-то величин, нужно транспонированный вектор умножить на себя.

$$R^2 = \frac{ESS}{TSS} =$$

$$\frac{TSS - RSS}{TSS} = \frac{y'y - u'u}{y'y} =$$

$$\frac{y'y - (y - \hat{y})' * (y - \hat{y})}{y'y} =$$

$$\frac{y'y - (y - X\hat{\beta})' * (y - X\hat{\beta})}{y'y} =$$

$$\frac{y'y - (y'y - \hat{\beta}'X'y - y'X\hat{\beta} + \hat{\beta}' * X'X\hat{\beta})}{y'y} =$$

$$\frac{\hat{\beta}'(X'y - X'X\hat{\beta}) + y'X\hat{\beta}}{y'y} =$$

Тут магия линейной алгебры (Просто в лоб и последовательно совмещаются арифметические действия с матрицами, а потом числитель и знаменатель делятся на $(n - 1)$. Зачем? А вот смотри...)

$$\frac{(n - 1) \sum_{i=1}^k ((y - \bar{y})(x_i - \bar{x}) \hat{\beta}_i)}{(n - 1) \sum_{i=1}^k (y - \bar{y})^2} = \frac{\sum_{i=1}^k \text{cov}(x_i, y)}{\text{var}(y)}$$

Поэтому наш фактор x_i объясняет долю $\frac{\text{cov}(x_i, y)}{\text{var}(y) R^2}$ в регрессии

Как мы уже знаем, чем больше R^2 , тем по идее наша модель лучше описывает реальность, то есть напихать ~~до~~ очень много факторов, то даже при единичном значении R^2 модель будет плохой, поэтому умные люди придумали корректировать коэффициент детерминации регрессии на число степеней свободы и назвали его R_{adj} .

$$R_{adj} = \overline{R^2} = 1 - \frac{RSS/(n - k - 1)}{TSS/(n - 1)}$$

Свойства:

1. Связь коэффициентов R^2 и R_{adj}^2

$$R_{adj}^2 = 1 - (1 - R^2) \times \frac{n-1}{n-k-1}, \text{ это следует из простой арифметики } R^2 = 1 - \frac{RSS}{TSS} \rightarrow \frac{RSS}{TSS} = 1 - R^2$$

$$2. R_{adj}^2 \leq R^2$$

Замечу еще, что последнее свойство справедливо только для регрессий с одинаковыми зависимыми переменными, но можно взять разные наборы Независимых переменных.

3. Лучше модель с большим R_{adj}^2

Вот так незаметно закончилась глава, а ты боялся. :)

Теорема Гаусса-Маркова в матрицах

В каком-то смысле использование упрощает нашу жизнь, например, теперь нужно писать только одну формулу для оценок коэффициентов.

Из-за того что когда-то давно мы предположили, что ошибки - это случайная величина, мы имеем право считать оценки коэффициентов случайными величинами. Это работает и в матрицах, так как матрица X детерминирована, а вектор u случаен.

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

Теорема Гаусса-Маркова для множественной регрессии

Если в $y = X\beta + u$ выполнены предпосылки

- Модель правильно специфицирована, то есть в нее включены все необходимые факторы, а лишних факторов нет и выбрана правильная функциональная форма модели
- Ранг матрицы X максимален, то есть матрица невырожденная
- $E(u) = 0$
- $Var(u) = \sigma_u^2 \times I_{n \times n}$
- $\forall \epsilon_i, \epsilon_j \in u, cov(\epsilon_i, \epsilon_j) = 0$

тогда оценки МНК являются BLUE

Если ТГМ выполняется, то верно

$$\hat{\sigma}_u^2 = \frac{RSS}{n-k-1}, \text{ где } k - \text{это количество факторов в модели.}$$

Ура! Тестируем гипотезы в рамках множественной регрессии!

Я искренне верю, что ты умеешь считать односторонние доверительные интервалы, поэтому распишу только двусторонние. Анализ проверки гипотезы немного усложняется, так как мы работаем не с одной бетой, а с множеством бет.

Каков наш план, капитан? А вот, собственно:

1. Гипотеза о конкретном значении коэффициента

$$H_0: \beta_i = \beta_i^0$$

$$H_a: \beta_i \neq \beta_i^0$$

В случае множественной регрессии $\frac{\hat{\beta}_i - \beta_i^0}{\sigma_{\hat{\beta}_i}} \sim t_{n-k-1}$, тогда доверительный интервал для проверки будет иметь вид:

$$\hat{\beta}_i - t_{n-k-1, \frac{\alpha}{2}} \times \sigma_{\hat{\beta}_i} \leq \beta_i^0 \leq \hat{\beta}_i + t_{n-k-1, 1-\frac{\alpha}{2}} \times \sigma_{\hat{\beta}_i}$$

Если заданное в нулевой гипотезе значение беты попадает в доверительный интервал, тогда гипотеза не отвергается. Похоже на что-то?

2. Гипотеза о значимости коэффициента

Рассмотрим только случай двусторонней гипотезы:

$$H_0: \beta_i = 0$$

$$H_a: \beta_i \neq 0$$

В таком случае $\frac{\hat{\beta}_i}{\sigma_{\hat{\beta}_i}} \sim t_{n-k-1}$, тогда доверительный интервал для проверки будет иметь вид:

$$\hat{\beta}_i - t_{n-k-1, \frac{\alpha}{2}} \times \sigma_{\hat{\beta}_i} \leq \beta_i^0 \leq \hat{\beta}_i + t_{n-k-1, 1-\frac{\alpha}{2}} \times \sigma_{\hat{\beta}_i}$$

Если 0 попадает в доверительный интервал, тогда гипотеза не отвергается или иными словами, коэффициент НЕ значим.

3. Проверка гипотез об адекватности регрессии

Эта гипотеза необходима для проверки наличия линейной зависимости между x и y . Если раньше достаточно было проверить значимость x , то сейчас надо поступить хитрее. Так как в множественной регрессии существует совокупное влияние нескольких $x_{i,j}$ на y , то давай проверим гипотезу о том, что все коэффициенты равны 0.

$$H_0: \beta_0 = \dots = \beta_k = 0$$

$$H_a: \exists \beta_i \neq 0, \forall i = 1 \dots k$$

Иными словами нулевая гипотеза гласит, что выбранный набор независимых переменных HE оказывает линейного воздействия на y .

Тестовая статистика для проверки гипотезы имеет вид:

$$\frac{ESS/k}{RSS/(n-k-1)} = \frac{R^2/k}{(1-R^2)/(n-k-1)} = F_{k, n-k-1}$$

Грустно тут придумывать доверительный интервал, поэтому просто нужно сравнить полученное значение F-теста с критическим для обозначенного количества степеней свободы

Гипотеза *отвергается*, если $F_{k, n-k-1} > F_{\alpha(k, n-k-1)}^{cr}$

А теперь разовею мысль: а что если в y добавить скажем $m - k$ переменных, а чего мелочиться, то каков будет вклад новых факторов в регрессии?

Для этого уже есть гипотеза и статистика.

$$H_0: \beta_{k+1} = \dots = \beta_m = 0$$

$$H_a: \exists \beta_i \neq 0, \forall i = k + 1 \dots m$$

$$\frac{(RSS - RSS_m)/m - k}{RSS_m/(n - m)} = F_{m-k, n-m}$$

Если гипотеза отвергается, то вклада новые факторы не приносят.

4. Проверка общей гипотезы о наличии линейных соотношений между коэффициентами

Что такое линейные ограничения? Это когда несколько бет при разных факторах равны между собой или просто равны 0.

Строго говоря, есть три варианта, как проверить такую гипотезу: громоздким t-тестом, тестом Вальда и F-тестом, о последнем, как о наиболее удобном мы и поговорим ниже.

Пусть Q это некоторая матрица размера $(q \times k + 1)$, где q -это количество линейно независимых ограничений на коэффициенты (то есть количество не выражаемых друг из друга ограничений).

$$H_0: Q\beta = q$$

$$H_a: Q\beta \neq q$$

$F = \frac{(RSS_r - RSS_{ur})/q}{RSS_{ur}/(n-k-1)} = \frac{(R_{ur}^2 - R_r^2)/q}{(1 - R_{ur}^2)/(n-k-1)}$, где RSS_r - модель с учетом всех ограничений RSS_{ur} - модель без ограничений.

Если гипотеза не отвергается, то лучше использовать короткую модель, потому что она удобнее для анализа. И знаешь, для объяснения такой статистики есть геометрическая интерпретация.

#Тут будет график. Скоро. А вообще, если ты знаешь, как тут рисовать 3д модели, помоги мне, пожалуйста.

Функциональна форма регрессии и еще немножко о тестах

Всегда ли удобно использовать регрессию для выборки в лоб? Если я задаю такой вопрос, то очевидно, что ответ “нет”, и это так. Действительно, а зачем тогда люди придумывали логарифмы и всякие разные преобразования? Конечно, чтобы облегчить нам жизнь. ~~ если нет, то гуманитария ответ ~~. Давай разбираться, как можно преобразовывать переменные в регрессиях.

Что делать, если зависимость задана $\hat{y}_i = \hat{\beta}_0 + \frac{\hat{\beta}_1}{x_i}$ или $y = \hat{\beta}_1 X^{\hat{\beta}_2}$? Ну не плакать же, ведь в конце концов к обеим зависимостям можно применить линейный регрессионный анализ :)

Заметим опять же, что для метрики важна линейность по параметрам, а не по переменным, поэтому можно заменить $z_i = \frac{1}{x_i}$ и радоваться жизни дальше. А что делать с $y = \hat{\beta}_1 X^{\hat{\beta}_2}$, ведь преобразование вида $z_i = \frac{\hat{\beta}_2}{\sqrt{x_i}}$ приведет в ужас эконометриста, потому что модель мамочки не линейна и по параметрам, и по переменным. Но все же можно к ней применять линейный регрессионный анализ, если ты владеешь темной-темной магией логарифми-и-ирования. Но если серьезно, то сейчас ты поймешь, что без логарифмов никуда, ведь он подобен утюжку для волос, который из кудельков делает прямые волосы, но только для функций.

:sparkles: А теперь магия

$$y = \hat{\beta}_1 X^{\hat{\beta}_2}$$

$$\ln y = \ln \hat{\beta}_1 + \hat{\beta}_2 \ln x$$

Вот и все :)

Такое преобразование дает нам логлинейную зависимость, ибо она линейна относительно логарифмов. Как интерпретировать такую зависимость? В нашем случае трудно дать интерпретацию $\hat{\beta}_1$, но для $\hat{\beta}_2$ она есть, бэта при факторе показывает, на сколько процентов изменится наш y при изменении на 1% x . Ничего не напоминает? Да, эластичность даже тут нас нашла

Иногда природные процессы бывают сложными и описываются так: $y = \hat{\beta}_1 e^{x\hat{\beta}_2}$. Что же делать? Так, ну-ка утер слезы! Зовем логарифм!

$$y = \hat{\beta}_1 e^{x\hat{\beta}_2}$$

$$\ln y = \ln \hat{\beta}_1 + \ln e^{x \hat{\beta}_2}$$

$$\ln y = \ln \hat{\beta}_1 + x \hat{\beta}_2$$

Вот так мы снова справились с нелинейностью! Такие модели называют полулогарифмическими моделями. Но как же интерпретировать нашу модель? Теперь β_2 означает изменение y в *процентах* при изменении x на одну *абсолютную* величину.

А теперь вопрос на миллион. Как выбрать нужную модель, если они все классенькие и линейные, да еще и все адекватно описывают. Ответ: надо тестировать.

Как выбрать функциональную форму модели, если все они тебе нравятся?

Умение линеаризовать модель в некоем роде формализует линейный регрессионный анализ, что на руку исследователю, но одновременно это здорово усложняет ему жизнь, так как при непосредственном выборе формы модели нельзя сравнивать сумму квадратов отклонений и R^2 , как мы делали в случае наличия ограничений в модели, потому что теперь эти показатели не измеряют одно и то же, а если еще и безразмерный R^2 в двух регрессиях почти одинаковый, то все равно надо держать себя в руках и не плакать!

Есть три теста, которые надо знать:

1. Тест Бера-МакАлера или как выбрать за четыре шага между полулогарифмической и линейной моделью.

Шаг 1:

Найдем оцененные значения зависимой переменной в каждой модели \hat{y} , $\ln \hat{y}$

Шаг 2:

Составим вспомогательные регрессии

$$e^{\hat{y}} = \beta_0 + \sum_{i=1}^k \beta_i x_i + \epsilon_{1,i}$$

$$\ln \hat{y} = \beta_0 + \sum_{i=1}^k \beta_i x_i + \epsilon_{2,i}$$

Шаг 3: Оценим совсем новые регрессии

$$\ln y = \beta_0 + \sum_{i=1}^k \beta_i x_i + \theta_1 \epsilon_{1,i}^{\wedge} + u_i$$

$$y = \beta_0 + \sum_{i=1}^k \beta_i x_i + \theta_1 \epsilon_{2,i}^{\wedge} + u_i$$

Шаг 4: Сравним $\hat{\theta}_1$ и $\hat{\theta}_2$, если $\hat{\theta}_1$ значим, то выбираем полулогарифмическую модель, если значим $\hat{\theta}_2$ - то линейную, а если оба коэффициента одновременно значимы или незначимы, то можно наконец-то пустить скупую слезу, потому что придется применить другой тест.

2. РЕ-тест МакКиннона для выбора между полулогарифмической и линейной моделью.

Шаг 1:

Найдем оцененные значения зависимой переменной в каждой модели \hat{y} , $\ln \hat{y}$

Шаг 2: Оценим вспомогательные регрессии

$$\ln \hat{y} = \beta_0 + \sum_{i=1}^k \beta_i x_i + \gamma_1 (\hat{y} - e^{\ln \hat{y}}) + \epsilon_1$$

$$\hat{y} = \beta_0 + \sum_{i=1}^k \beta_i x_i + \gamma_2 (\ln \hat{y} - \ln \hat{y}) + \epsilon_2$$

Шаг 3: Оцениваем эту модель и сравниваем значимость γ

Если $\hat{\gamma}_1$ **НЕ**значим, то выбираем полулогарифмическую модель, если **НЕ**значим $\hat{\gamma}_2$ - то линейную, а если оба коэффициента одновременно значимы или незначимы, то вообще есть повод беспокоиться.

2. РЕ-тест МакКиннона для выбора между логарифмической и линейной моделью.

Шаг 1:

Найдем оцененные значения зависимой переменной в каждой модели

$$\hat{y} = \beta_0 + \sum_{i=1}^k \beta_i x_i + \epsilon_{1,i}$$

$$\ln \hat{y} = \beta_0 + \sum_{i=1}^k \beta_i \ln x_i + \epsilon_{1,i}$$

Шаг 2: Оценим вспомогательные регрессии

$$\ln \hat{y} = \beta_0 + \sum_{i=1}^k \beta_i \ln x_i + \delta_1 (\hat{y} - e^{\ln \hat{y}}) + \epsilon_1$$

$$\hat{y} = \beta_0 + \sum_{i=1}^k \beta_i x_i + \delta_2 (\ln \hat{y} - \ln \hat{y}) + \epsilon_2$$

Шаг 3: Оцениваем эту модель и сравниваем значимость δ

Если $\hat{\delta}_1$ НЕзначим, то выбираем логарифмическую модель, если НЕзначим $\hat{\delta}_2$ - то линейную, а если оба коэффициента одновременно значимы или незначимы, то есть серьезный такой повод беспокоиться.

3. Универсальный тест Бокса-Кокса №1

Это волшебный тест, где мы меняем масштаб наблюдений y , чтобы сравнивать RSS непосредственно.

Шаг 1: Вспомним, что среднее геометрическое должно быть равно экспоненте среднего арифметического

$$e^{\frac{1}{n} \sum_{i=1}^n \ln(y)} = e^{\frac{1}{n} \ln(\prod_{i=1}^n y)} = (\prod_{i=1}^n y)^{\frac{1}{n}}$$

Шаг 2: Масштабируем наблюдения

$$y_i^* = \frac{y_i}{(\prod_{i=1}^n y)^{\frac{1}{n}}}$$

$$\ln(y_i^*) = \ln\left(\frac{y_i}{(\prod_{i=1}^n y)^{\frac{1}{n}}}\right)$$

Шаг 3:

Оценим модели с использованием y_i^* и $\ln(y_i^*)$, а потом просто оценим RSS двух моделей

3. Тест Бокса-Кокса №2 интуитивно труднопонимаемый, но легко заучиваемый

По сути это такой же Бокс-Кокс как и раньше, но с неочевидной трансформацией переменных.

Шаг 1: Введем саму трансформацию

$y^{(\lambda)} = \frac{y^\lambda - 1}{\lambda}$, $\lambda \neq 0$, чтобы избежать неопределенности. Если устремить лямбду к нулю и воспользоваться правилом Лопиталя, то получим логарифм игрека

В общем виде

Тогда линейная форма принимает интересный вид

$$y^{(\lambda)} = \beta_0 + \sum_{i=1}^k \beta_i x_i^{(\theta)} + \epsilon_{1,i}$$

В оцененной модели нас теперь интересуют сразу значимость и лямбды, и теты.

$\theta = \lambda = 1$ выбираем линейную модель

$\theta = \lambda = 0$ выбираем логлинейную модель

$\theta = 1$ без сомнений используем линейную модель

$\lambda = 0$ использовать полулогарифмическую модельку

$$\theta = -1, \lambda = 1, \text{ то } y_i = \beta_0 + \beta_1 \times \frac{1}{x_i} + u_i$$

Фиктивные переменные и зачем они нужны

Фиктивные переменные или их еще называют дамми - это такие классы переменные, которые принимают только два значения: 0 и 1. Их называют качественными, потому что они не измеримы по какой-либо шкале. Они переменные нужны для “разветвления” регрессий на два или больше путей развития событий.

Классическим примером всегда является уравнение минцеровского типа (да-да, экономика труда) для заработной платы.

$$\ln(\hat{W}) = \hat{\beta}_0 + \hat{\psi}_0 \times D_1(Male) + \hat{\beta}_1 \times Edu + \hat{\beta}_2 \times City$$

Пусть

Заметим, что постановка задачи нетолерантная, тут рассматривается не только два гендера, но еще и женщинам определен 0. Грусть и боль, но мы идем далее.

Так вот, в нашей нетолерантной задаче респонденты делятся на два вида, и для каждого существует своя регрессия.

$$\text{Для мужчин } \ln(\hat{W}) = \hat{\beta}_0 + \hat{\psi}_0 + \hat{\beta}_1 \times Edu + \hat{\beta}_2 \times City$$

$$\text{Для женщин } \ln(\hat{W}) = \hat{\beta}_0' + \hat{\beta}_1' \times Edu + \hat{\beta}_2' \times City$$

очевидно, что зарплаты мужчин в среднем на $\hat{\psi}_0$ процентов выше, чем у женщин.

Но можно еще поиграть с моделью.

$$\ln(\hat{W}) = \hat{\beta}_0 + \hat{\psi}_0 \times D_1(Male) + \hat{\beta}_1 \times \hat{\psi}_1 \times D_1Edu + \hat{\beta}_2 \times City$$

Пусть

Пусть

Интерпретация дамми переменной все такая же не толерантная, даже хуже. Теперь мы не просто мужчин от женщин разделяем, а еще высокообразованных людей от не очень образованных. Наши регрессии принимают вид

Образованные

Для мужчин

$$\ln(\hat{W}) = \hat{\beta}_0 + \hat{\psi}_0 + \hat{\beta}_1 \times \hat{\psi}_1 Edu + \hat{\beta}_2 \times City$$

Для женщин

$$\ln(\hat{W}) = \hat{\beta}_0' + \hat{\beta}_1' \times \hat{\psi}_1 \times Edu + \hat{\beta}_2' \times City$$

Youtube бьюти блоггеры и те, кто ушел после 9-го класса.

Для мужчин

$$\ln(\hat{W}) = \hat{\beta}_0 + \hat{\psi}_0 + \hat{\beta}_2 \times City$$

Для женщин

$$\ln(\hat{W}) = \hat{\beta}_0^* + \hat{\beta}_2^* \times City$$

В этом случае у нас 4 регрессии, и если рассматривать их отдельно по группам, то можно заметить, что кроме константы для мужчин и женщин меняется угол наклона для образованных и не очень (простите меня за нетолерантность и оскорбление чувств выбравших иной путь... хотя.. надеюсь, что вторая группа в выборке никогда этого не прочитает)

Но если быть честным, то где ты видел страну, в которой существует только два типа образования? Я осмелюсь предположить, что таких стран вообще нет.

В России, например, существуют школы, техникумы, колледжи, военные училища, вузы, программы дополнительного профессионального образования, а каждый из перечисленных уровней образования в свою очередь делится на свои подуровни, поэтому деление на 2 ветки не корректно. Давай соберем мужчин и женщин в одну выборку обратно, а вот фактор *Edu* размножим до вот такого набора: *School, Tech, College, Uni*. Как думаешь, сколько дамми переменных нам нужно? Верно 3 штуки, чтобы избежать коварной ловушки дамми-переменных. Ведь как говорит один всем известный и великий Пэр:

Настоящие джентельмены и благородные дамы никогда не включают дамми-переменные на все значения факторной переменной и константу в свою регрессию, потому что знают о недопустимости существования жесткой линейной зависимости между регрессорами, ведь иначе нарушается неприменимый закон ТГМ о линейной независимости регрессоров.

Пусть основной регрессией, относительно которой мы будем проводить сравнение, будет та, что выделяет людей с дипломом вуза от остальных. В общем виде оцененная модель имеет вид:

$$\ln(\hat{W}) = \hat{\beta}_0 + \hat{\psi}_1 \times D_1Sch + \hat{\psi}_2 \times D_2Tech + \hat{\psi}_3 \times D_3Coll + \hat{\beta}_1 \times City$$

Она разветвляется на

$$\text{Тех, кто закончил 11 классов } \ln(\hat{W}) = \hat{\beta}_0 + \hat{\psi}_1 \times Sch + \hat{\beta}_2 \times City$$

$$\text{Тех, кто закончил 9 классов и техникум } \ln(\hat{W}) = \hat{\beta}_0 + \hat{\psi}_2 \times Tech + \hat{\beta}_2 \times City$$

$$\text{Тех, кто закончил 11 классов и колледж } \ln(\hat{W}) = \hat{\beta}_0 + \hat{\psi}_3 \times Coll + \hat{\beta}_2 \times City$$

$$\text{Тех, кто закончил 11 классов и вуз } \ln(\hat{W}) = \hat{\beta}_0 + \hat{\beta}_2 \times City$$

Вот так мы с помощью дамми-переменной развили наш анализ на 4 ветки и избежали ловушки дамми. Замечу, что выбор базовой категории никак не влияет на наклон уравнения, но меняет показатель стандартной ошибки для бэты.

Наверное в процессе чтения у тебя возник законный вопрос, а почему перед дамми стоит какая-нибудь другая свободная греческая буква, но не бэта?

А вот почему. На самом деле каждую из регрессий можно построить без использования дамми, просто заранее разбив выборки нужным способом. Но дело это прямо сказать, мучительное муторное, поэтому было придумано вот что.

Умные, но уставшие ученые предложили найти разницу между бэтами в регрессиях для женщин и мужчин и положить ее равной некоторой γ , которую надо ставить перед особенной бинарной переменной (дамми). Так вот для женщин будет исходная маленькая бета ноль, а для мужчин бэта ноль плюс гамма, то есть их родное большое значение бэты ноль штрих. Так и была положена эра благоразумия дамми переменных. Хотя никто не отрицает, что бэта штрих может быть меньше бэты 0 из-за отрицательной гаммы, но я человек позитивный и отрицательные числа не очень люблю, но признаю их существование. :)

Тесты, тесты и еще раз тесты!

Как ты уже понял, главное правило эконометриста “доверяй, но проверяй”, поэтому в этой главе мы снова будем тестировать наши гипотезы, но теперь уже с дамми-переменными.

Начнем с самой простой гипотезы, которая состоит в том, что разница γ равна нулю, её проверяют стандартным t -тестом в случае одной фиктивной переменной, а что делать, если таких “пустых” переменных много?

Предположим, что тестируется наша гипотеза

$$\ln(\hat{W}) = \hat{\beta}_0 + \hat{\psi}_1 \times D_1Sch + \hat{\psi}_2 \times D_2Tech + \hat{\psi}_3 \times D_3Coll + \hat{\beta}_1 \times City$$

$$H_0: \psi_1 = \psi_2 = \psi_3 = 0$$

$$H_a: \exists \psi_i \neq 0 \quad \forall i \in [1, 2, 3]$$

Как мы уже знаем, такие гипотезы проверяют F-тестом

Тестируемая статистика имеет вид

$$F = \frac{(RSS_{NO \ DUMMIES} - RSS_{WITH \ DUMMIES}) / (k + 1)}{(RSS_{WITH \ DUMMIES}) / (n - 2k - 2)}$$

Тест Чоу

Или как исследовать структурную устойчивость коэффициентов модели. Бывает, что существует выборка с 2-мя сильно различными подвыборками, и исследователь стоит перед выбором: оценивать единую выборку или же каждую подвыборку отдельно. Для того, чтобы делать правильный выбор, был придуман тест Чоу. Для примера мы продолжаем использовать уравнение Минцера.

$$\ln(W_{pooled}) = \hat{\beta}_0 + \hat{\psi}_1 \times D_1Sch + \hat{\psi}_2 \times D_2Tech + \hat{\psi}_3 \times D_3Coll + \hat{\beta}_1 \times City$$

Единую зависимость в регрессии с дамми-переменными можно проверить иначе. Ведь мы знаем, как выглядит оцененная модель, когда все дамми равны 0.

$$\ln(\hat{W}_2) = \hat{\beta}_0 + \hat{\beta}_2 \times City$$

Предположим, что все дамми равны 1, это конечно же, невозможно, но вдруг... на самом деле в таком случае, мы просто используем факт, что у человека есть какое-то образование, что просто соответствует фактору *Edu*, тогда вспомним историю про усталых ученых и гамму.

$\beta_i' = \beta_i + \psi_i$, тогда регрессия будет вида

$$\ln(W_1) = (\beta_0 + \psi_0) + (\psi_1 + \beta_1) \times Sch + (\psi_2 + \beta_2)Tech + (\psi_3 + \beta_3)Coll + (\psi_4 + \beta_4)City$$

И наша нулевая гипотеза принимает вид

$$H_0: \psi_i = \beta_i'$$

$$H_a: \exists \psi_i \neq \beta_i' \quad \forall i \in [1,2,3]$$

А теперь сама статистика

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

Если посмотреть внимательно, то будет ясно, что эта статистика очень похожа на статистику, которая проверяет наличие линейных ограничений, а значит $RSS_{pooled} = RSS_R$, а $RSS_1 + RSS_2 = RSS_{ur}$, поэтому можно сделать вывод, что $(RSS_1 + RSS_2) \leq RSS_{pooled}$

Последняя тема к экзамену! Пропущенные переменные и избыток факторов.

В районе темы ТГМ я уже говорила, что пихать в регрессию все факторы, которые тебе попались на глаза, - это **очень** плохая идея. Но теперь мы от императивного правила перейдем к осознанию проблемы излишка или нехватки факторов в регрессии, а якорным словом для нас станет "**смещение**".

Пусть истинная модель имеет вид $y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \beta_3 x_{3,i} + u$

Но по каким-то неведомым причинам, скажем, потому что решили оценивать модель в 2 часа ночи из-за бессоницы, наша оцененная модель имеет вид:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1,i} + \hat{\beta}_2 x_{2,i}$$

Тогда бэта 2 будет одновременно играть две роли, а именно отражать влияние второго фактора и мнимый эффект третьего фактора, который определяется способностью имитировать поведение третьего фактора вторым фактором и, собственно, чистым эффектом влияния третьего фактора на объясняемую переменную. Все это порождает смещение оценок бэточек, а те в свою очередь делают некорректным результат оценивания стандартной ошибки, что уже в свою очередь ведет к смещению Т-статистик, а это черевато тем, что какой-нибудь важный в реальности фактор при

проверке гипотез станет незначимым, что может привести к очень большим проблемам мирового масштаба.

Формально это записывается таким образом при “неучете важного фактора” $\hat{\beta}_i = \beta_i + \beta_j \times \frac{cov(x_i, x_j)}{var(x_j)}$ и если ковариация между иксами или сама бета пропущенного фактора равны нулю, то оценка будет несмещенной, потому что мнимого следа не будет.

Поэтому, чтобы даже в 2 ночи правильно выбирать факторы для регрессий, был разработан ряд тестов.

Давай рассмотрим самый популярный тест - тест Рамсея или RESET-test.

Этот тест отвечает на вопрос, а надо ли включать в регрессию еще факторы или может уже все нужные учтены?

Тут матрицы и векторы! Осторожно!

Исходная модель $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u$

Шаг 1

Оцениваем модельки и внимательно глядим на вектор оцененной объясняемой переменной.

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3$$

Шаг 2

Сделаем новую модель, где включаем после всех k начальных факторов еще m в виде оцененного игрока, который возводится в степень равную порядковому номеру своего коэффициента - дельты- плюс 1. И оцениваем модель.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \delta_1 \times \hat{y}^2 + \delta_2 \times \hat{y}^3 + \delta_3 \times \hat{y}^4 + \delta_4 \times \hat{y}^5 + u$$

Шаг 3

Теперь гипотеза об отсутствии пропущенных переменных может быть задана следующим образом

$$H_0: \forall \delta_i, \delta_i = 0, \quad i \in [1, m]$$

$$H_a: \exists \delta_i \neq 0$$

Тогда снова прибегнем к F-тесту вида. Если статистика больше критического значения, то мы молодцы, потому что включили все значимые переменные.

$$F = \frac{(RSS_r - RSS_{ur}) / (m - 1)}{RSS_{ur} / (n - k - m + 1)}$$

Вот и все.

Модуль 3

Весна идет, весне дорогу, но сессия уже совсем близко. Пора начать готовиться :)

Мультиколлинеарность данных и способы борьбы с ней.

Мультиколлинеарность - звучит это страшно, будем разбираться по порядку.

Мультиколлинеарность - это ситуация, когда между регрессорами существует линейная связь. Это первая ситуация, когда мы нарушим предпосылки на регрессоры ТГМ, но как говорила Гермiona Гренджер, нарушать правила - это так весело :)

Существует два вида мультиколлинеарности:

1. Строгая мультиколлинеарность, она же идеальная, - ситуация, когда между регрессорами существует строгая линейная зависимость, что несет за собой $\det(X) = 0$. Это очень плохой вид мультиколлинеарности, так как он возникает чаще всего из-за невнимательности исследователя. Классическим примером такой ошибки является dummy trap, когда исследователь использует и дамми для мужчин, и дамми для женщин в одной регрессии.

$$y_i = \beta_0 + \beta_1 x_i + \beta_3 \text{male}_i + \beta_4 \text{female}_i$$

И если записать матрицу X в явном виде:

$$\begin{pmatrix} 1 & x_1 & 0 & 1 \\ 1 & x_2 & 1 & 0 \\ 1 & x_3 & 1 & 0 \\ 1 & x_4 & 0 & 1 \end{pmatrix}$$

То видно, что если сложить последние два столбца матрицы, то получается первый столбец. Вывод: есть строгая линейная зависимость, что автоматически гарантирует нам неединственность оценок МНК, потому что ранг матрицы X не равен k, а потому матрицы $(X'X)^{-1}$ нет. Как с этим бороться? Ну, смотри, можно ложиться спать не с рассветом, удалить Dota2 и не сидеть в ВК, пока делаешь исследования, потому что все это мешает быть внимательным. Во всех остальных случаях я буду под мультиколлинеарность иметь в виду условную мультиколлинеарность.

2. Условная мультиколлинеарность, она же естественная, нестрогая или квазимultiколлинеарность, - это ситуация, когда зависимость вроде бы есть, а вроде бы и нет, т.е. она примерная, И (приятности) такая мультиколлинеарность НЕ нарушает предпосылки ТГМ.

Так получается потому, что есть в модели регрессоры-родственники, которые либо измеряют почти одно и то же, либо просто исторически между ними есть естественные соотношения. Это неприятность для нас, если выборка маленькая, так как оценки неустойчивые, но все-таки мультиколлинеарность - это самая нестрашная проблема, которая к тому же просто решается.

Что такого страшного в мультиколлинеарности?

Из-за того, что один регрессор хорошо объясняет другой/другие (читай: RSS_j маленький), смещаются оценки дисперсии коэффицентов,

$$Var(\hat{\beta}_j) = \frac{\hat{\sigma}_j^2}{RSS_j} = \frac{1}{1 - R_j^2} \times \frac{\hat{\sigma}_j^2}{TSS_j} \gg 0$$

Где R_j^2 - это показатель детерминации в регрессии x_j на все остальные регрессоры.

что приводит к тому, что доверительные интервалы, где используются стандартные ошибки, расширяются поэтому вероятность принять значимый коэффициент за незначимый становится ненулевой, из-за чего модель становится неустойчивой и чувствительной к добавлению новых переменных.

А так ли страшен черт, как его малюют? Честно? Нет, если нам не очень важна интерпретация коэффициентов, а нужны только \hat{y}_i , то можно забить со спокойной душой на нестрогую мультиколлинеарность, потому что ТГМ работает и оценки BLUE.

Теперь не страшно? Правильно, но как же выявить мультиколлинеарность?

Эта задача не так проста, как может показаться на первый взгляд, потому что точного критерия, который сказал бы нам, что мультиколлинеарность есть или нет, не существует, но ученые не сдаются, они нашли несколько способов заподозрить наличие этого явления в модели.

Способ №1 Визуальный

Самый простой способ заподозрить мультиколлинеарность - это просто оценить модель. Если исследователь видит группу незначимых по отдельности коэффициентов, а гипотеза о том, что они незначимы одновременно не отвергается, - мультиколлинеарность есть.

Способ №2 Показатель вздутия дисперсии

Надо отметить, что примерно линейная связь может быть между несколькими факторами, и чтобы найти такие связи для каждого фактора, используют показатель вздутия дисперсии. Это, конечно, требует много времени, чтобы рассчитать для каждого j -го регрессора $VIF(X_j)$, так как ну регрессоров аж k штук, но если хотя бы один виф будет больше 6, то можно задуматься о том, что мультиколлинеарность в регрессии есть.

$$VIF(X_j) = \frac{1}{1 - R_j^2} > 6$$

Способ №3 Параметр обусловленности матрицы $X'X$

Когда мультиколлинеарность строгая, то $\det(X'X) = 0$, а если она нестрогая, то определитель будет близок к нулю. Значение определителя будет зависеть от единиц измерения переменных, вот и вопрос: если мои переменные в нанометрах, а выражены в метрах, то это мультиколлинеарность или значение определителя маленькое из-за единиц измерения? Ответ нам даст метод поиска мультиколлинеарности через параметр обусловленности.

$$CN = \sqrt{\frac{\lambda_{max}}{\lambda_{min}}} > 30$$

Где $\lambda_{min}, \lambda_{max}$ - это наименьшее и наибольшее значения собственных значений матрицы $X'X$.

Если этот показатель условно больше 30, то это сигнал о том, что мультиколлинеарность есть.

Как с мультиколлинеарность бороться?

Мультиколлинеарность особенно страшна для малого количества наблюдений, потому что с ростом наблюдений или исходов эксперимента ее влияние слабее, так как дисперсия будет уменьшаться. Поэтому первое, что приходит в голову, давайте увеличим количество наблюдений или проведем серию дополнительных экспериментов. Утопично, нереально, но в теории должно работать.

Другой вариант - это корректировка модели с жертвованием несмещенности оценок, но делать это нужно осторожно, чтобы не причинить вреда здоровью (преподавателя).

Способ №1. По-тихому выкинуть из модели одну из переменных, которая (по нашему мнению) порождает эту квазимультиколлинеарность, но тогда мы жертвуем несмещенностью оставшихся коэффициентов. Но лучше это делать в крайних случаях.

Почему должно сработать?

1. Мы же не бездумно выкидываем регрессоры, верно? Пусть мы выкидываем $r < k$ регрессоров, которые мы подозреваем на порождение мультиколлинеарности, но перед этим нужно проверить гипотезу о том, что эти r регрессоров незначимы одновременно, чтобы убедиться, что она отвергается. Если мы запишем наблюдаемую F-статистику для гипотезы и она окажется меньше 1, то при удалении группы регрессоров R_{adj} должен увеличиться.
2. Необходимым условием для увеличения R_{adj} при удалении группы регрессоров является неравенство $|t| < \sqrt{r}$ (для одного регрессора - это будет достаточным условием)

Способ №2. Ввести штраф в МНК. Вот это уже по-научному. Есть 3 варианта штрафа МНК за то, что $\hat{\beta}_i \gg 0$, но принцип один:

$$\min_{\hat{\beta}_j, j \in Z} (RSS + \text{penalty}(\hat{\beta}))$$

1. Ridge-regression

$$\min_{\hat{\beta}_j, j \in Z} (RSS + \lambda (\sum_{i=1}^k \hat{\beta}_i^2))$$

Оценки для $\hat{\beta} = (X'X + \lambda I)^{-1} X'Y$

2. LASSO-regression

$$\min_{\hat{\beta}_j, j \in Z} (RSS + \lambda (\sum_{i=1}^k |\hat{\beta}_i|))$$

3. Elastic net method

$$\min_{\hat{\beta}_j, j \in Z} (RSS + \lambda_1 (\sum_{i=1}^k \hat{\beta}_i^2) + \lambda_2 (\sum_{i=1}^k |\hat{\beta}_i|))$$

Способ №3 Метод главных компонент

Суть МГК состоит в том, что с помощью простой замены переменных мы можем “забрать” в новую переменную всю изменчивость из самых волатильных старых переменных. С помощью МГК мы создаем k новых некоррелированных между собой регрессоров, которые представляют собой средневзвешанные исходные переменные.

Алгоритм создания новых переменных: Шаг 1: Центрирование переменных. Это одна из самых простых замен.

Пусть изначально был набор переменных $(a_1 \dots a_n)$, где a_i - вектор из n наблюдений.

Найдем среднее значение по каждой переменной и сделаем простенькую замену:

$$x_i = a_i - \bar{a}_i$$

Так вот мы незаметно перешли к новым переменным, но это еще не все.

Шаг 2: Поиск главных компонент

Тут все просто, чисто теоритически, наши главные компоненты - это просто все средневзвешанные иксы с максимально возможной дисперсией

$$\begin{cases} \sum_{i=1}^k w_i^2 \\ Var(pc_1) \rightarrow \max_{w_i \forall i=1, \dots, k} \end{cases}$$

Где pc_1 - это наша первая главная компонента с самым большим разбросом.

А теперь хинт: почему мы центрировали переменные? А потому что максимальная дисперсия для pc_1 равнозначна максимальной длине вектора pc_1 . То есть задача сводится к безусловной максимизации

$$|pc_1|^2 = \sum_{j=1}^n (\sum_{i=1}^k (w_i \times x_i))_j^2 \rightarrow \max_{w_i \forall i=1, \dots, k}$$

С точки зрения геометрии мы сделали так, чтобы вокруг pc_1 был максимальный разброс наблюдений.

Свойства МГК

В общем случае мы получаем, что МНК создает из старых k регрессоров создает k новых регрессоров.

$$\begin{cases} pc_1 = \sum_{i=1}^k w_{1i} x_{1i} \\ \dots \\ pc_k = \sum_{i=1}^k w_{ki} x_{ki} \end{cases}$$

Которые

1. $\rho(\hat{pc}_i, \hat{pc}_j) = 0 \forall i, j \in (1, \dots, k)$
2. $\sum_{i=1}^k \hat{Var}(x_i) = \sum_{j=1}^k \hat{Var}(pc_j)$

А так как мы пытаемся сделать так, что $\hat{Var}(pc_1)$ была максимальна из всех, то получается, что почти вся вариативность в $\sum_{i=1}^k \hat{Var}(pc_j)$ заложен в pc_1 , а поэтому оказывается, что все k переменных можно заменить если не одной, то заметно меньше, чем k ? набором переменных, а потому меньшее число переменных несет в себе почти всю информацию о переменных в МГК.

3. В терминах линейной алгебры МГК

Если x_i переменные центрированные, то $pc_j = Xv_j$, $|pc_j|^2 = \lambda_j$, где X - матрица v_j - собственные вектора матрицы $X'X$, а λ_j - собственное число матрицы $X'X$ для j -го собственного вектора.

Трудности применения методов на практике:

1. Разброс одной переменной не соотносим с другой переменной, если переменные с разными единицами измерениями
2. МГК выбирает несколько наиболее изменчивых переменных, но не дает гарантий, что будет выбрана переменная, которая сильнее других объясняет регрессант лучше других.

Как решать такие проблемы?

Действительно, а почему МГК, примененный в лоб, должен должен выбрать самую информативный регрессор? Такой волатильный регрессор вполне может быть просто шумным и мало влиять на объясняемую переменную. Чтобы МГК не ловил белый шум и переменные с мелкими единицами измерения, нужно еще скорректировать данные. А точнее не только центрировать, а нормировать исходные данные.

Гетероскедастичеость

К сожалению, мир не идеален, он полон риска и неопределенности, поэтому теоритические модели, которые хороши для мирка без силы трения, арбитража, с рациональными жителями и полным предвидением, нужно корректировать для нашей

планеты. И если мультиколлинеарность не внушает страха исследователям, то вот такие звери как гетероскедастичность и авторреляция заставляют понервничать.

Теперь мы предполагаем, что предпосылки ТГМ нарушаются. Но не все, а пока только одна.

Выше мы считали, что дисперсия ошибок всех ошибок наблюдений была одинакова. То есть $Var(\epsilon|X) = \sigma^2 I = const$ или ситуация гомоскедастичности.

Но нам бы и не знать, что верное решение задачи одно, а вот пути решения и ошибки на нем у всех разные: кто-то сочтет, что $2+2$ равно 5, а кто-то тоже самое примет равным 3. В общем это зависит от состояния мира. А чем же это нарушение плохо для нас?

Введем формальное определение гетероскедастичности

Ситуацию, когда $E(\epsilon_i^2|X) \neq const$ называют условной гетероскедастичностью, т.е. дисперсии ошибок двух наблюдений не равны между собой, потому что формула дисперсии задается функцией от σ^2 . При такой гетероскедастичности ошибки вполне себе могут быть одинаково распределенными и условно независимыми.

Ситуацию, когда $E(\epsilon_i^2|X) \neq \sigma^2$ называют безусловной гетероскедастичностью, то есть дисперсии просто не равны друг друга и какой-либо связи между ними нет.

Ситуация условной гетероскедастичности - это частое явление нетолько для экономических исследований, а если быть честным, то искать гетероскедастичность не нужно. Заранее можно считать, что в данных условная гетероскедастичность есть, да хотя бы потому что в любой базе данные бывают разного размера - сложно представить, что относительное изменение на 1% ВВП России и ВВП Кубы в абсолютном значении будет примерно равным. Мы, конечно, разберем способы поиска гетероскедастичности, но я лично разделяю позицию Демешева Б.Б. о том, что лучше сразу бороться с ней.

А все-таки, что же плохого в том, что дисперсии ошибок не равны между собой?

Линейность модели сохраняется, несмещенность оценок сохраняется. Ну что ж, оценки неэффективны, но это не мешает же им быть несмещенными... Ну пока ничего страшного. Так и должно быть, ведь только одна предпосылка нарушается. Но на самом деле, гетероскедастичность на не друг, так как именно из-за нее мы не можем строить доверительные интервалы и проверять гипотезы.

Пусть ошибки распределены нормально, тогда при наличии условной гетероскедастичности НЕВЕРНО:

$\frac{\hat{\beta}_i - \beta^0}{se(\hat{\beta}_i)} \sim t_{n-k}$. Из-за нарушение одной единственной предпосылки мы потеряли возможность строить доверительные интервалы для проверки гипотезы о значимости коэффициента.

$\frac{RSS}{\sigma^2} \sim \chi^2_{n-k}$ из чего автоматом следует, что

$\frac{RSS_R - RSS_{UN}/q}{RSS_{UN}/(n-k)} \sim F_{q, n-k}$, поэтому гипотезу о наличии линейно независимых ограничений тоже проверять нельзя.

Всегда есть последняя надежда - а может при большом количестве наблюдений условная гетероскедастичность не так страшна? Асимптотические свойства же...

Если у нас много наблюдений, то оценка коэффициента по вероятности, действительно, стремится к настоящему значению, а $\frac{RSS}{n-k}$ стремится к σ^2 , но не верно теперь $\frac{\hat{\beta}_i - \beta^0}{se(\hat{\beta}_i)} \sim N(0,1)$

Как заподозрить или убедиться в наличии гетероскедастичности в регрессии?

Способ №1 Графический

Для этого нужно просто оценить модель обычным МНК и найти остатки. Если построить график между модулем (или квадратом) остатка регрессии и подозрительного на порождение гетероскедастичности регрессора и заметить какую-то зависимость, то нужно корректировать модель, потому что гетероскедастичность точно есть.

Способ №2 тест Уайта. Самый старый из формальных тестов.

Этот тест рассматривает в общем виде зависимость между дисперсией случайного члена и объясняющими переменными, и поскольку нам не известна дисперсия случайного члена в i -м наблюдении, то за прокси берется квадрат отклонения (остаток регрессии) для этого наблюдения. Тест заключается в оценивании регрессии квадратов отклонений на объясняющие переменные модели, их квадраты и попарные произведения, исключая повторения.

Гипотеза Уайта

$$H_0: E(\epsilon_i^2 | X) = \text{const}$$

$$H_a: E(\epsilon_i^2 | X) \neq \text{const}$$

Алгоритм:

Шаг 1: Оценить основную регрессию и найти остатки

Шаг 2: Построить и оценить вспомогательную регрессию.

$$\hat{\epsilon}_i^2 = \gamma_0 + \sum_{j=1}^k \gamma_j \times x_{i,j} + \sum_{j=1}^k \gamma_{j+k} x_{i,j}^2 + \sum_{l=1}^k \sum_{j=1}^k \gamma_{j+2k} x_{i,j} \times x_{i,l} + u_i \quad \forall l \neq j$$

Шаг 3: считаем статистику Уайта

$$LM = nR_{aux}^2 \sim \chi_{m-1}^2,$$

где R_{aux}^2 - это коэффициент детерминации для вспомогательной модели, а m - это число регрессоров во вспомогательной модели.

Тест интуитивно понятный, Уайт считал, что все факторы тем или иным образом влияют на значение остатки в каждом наблюдении, поэтому при гомоскедастичности

подразумевается отсутствие аутлайнеров. При такой большой абстракции тест Уайта имеет большой минус, а именно он плохо работает на малых выборках и требует предпосылки о нормальности остатков и не показывает функциональную форму гетероскедастичности.

Способ №3: Тест Голдфреда-Квандта

Это самый популярный тест, потому что он хорошо справляется со своей задачей и на малых выборках, и на больших. Помимо предпосылок для регрессионного анализа он требует лишь нормальность от остатков регрессии. Его проводят, если подразумевают, что дисперсии остатков пропорциональны некоторой переменной x_j , т.е. условную гетероскедастичность.

Гипотеза этого теста совпадает с гипотезой в тесте Уайта.

Шаг 1 Оцениваем регрессию и находим остатки модели

Шаг 2 Ищем графически регрессор, с увеличением которого растет дисперсия ошибок.

Шаг 3 Упорядочиваем все наблюдения по модулю этой переменной и делим этот вариационный ряд на три примерно равные части

Шаг 4 Выкидываем из анализа середину этого вариационного ряда

Шаг 5 Оцениваем по первой и третьей группе отдельно регрессию

Шаг 6 Считаем статистику для проверки гипотезы о равенстве дисперсий

$$\frac{RSS_2/(n_2 - k)}{RSS_1/(n_1 - K)} \sim F_{n_2-k, n_1-k}$$

Однако зависимость дисперсии ошибок от x_j не обязана быть линейной, если функциональная форма зависимости сложнее, то применяют тест Глейзера.

Алгоритм:

Шаг 1: Оцениваем регрессию и находим остатки модели

Шаг 2: Ищем графически регрессор, с увеличением которого растет дисперсия ошибок.

Шаг 3: Оцениваем три вспомогательных регрессии

$$|\hat{\epsilon}| = \alpha + \beta \times X_j + u_i, |\hat{\epsilon}| = \alpha + \beta \times \sqrt{X_j} + u_i, |\hat{\epsilon}| = \alpha + \frac{\beta}{X_j} + u_i$$

Шаг 4: Проверка гипотезы

$$H_0: \beta = 0 \quad H_a: \beta \neq 0$$

Если коэффициент значим хотя бы в одной из регрессий, то имеет место быть гетероскедастичность.

Способ №5 Тест Бройша-Пагана

А что если все настолько плохо, что в наших данных гетероскедастичность есть, но от параметров, которые мы не включили в модель? Так, так, так! Не плакать! Эта проблема изящно решена в тесте Бройша-Пагана. К слову, Тревор Бройш и Адриан Паган живут в Австралии...

Этот тест заимствует гипотезу у теста Уайта, но меняет альтернативную гипотезу.

$$H_0: E(\epsilon_i^2|X) = \text{const}$$

$H_a: E(\epsilon_i^2|X) \sim f(\alpha_0 + \sum_{i=1}^p Z_i \times \alpha_i)$, где Z_i - это невключенный в модель i -й фактор, предполагается, что f обязательно гладкая функция.

Алгоритм:

Шаг 1 Оцениваем регрессию и находим остатки модели

Шаг 2 Ищем графически регрессор, с увеличением которого растет дисперсия ошибок.

Шаг 3 Строим оценку для нормировки квадратов остатков регрессии

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n \hat{\epsilon}_i^2$$

Шаг 4 Оцениваем скорректированную регрессию и вычисляем для нее ESS

$$\frac{\epsilon_i^2}{\hat{\sigma}^2} = \gamma_0 + \sum_{j=1}^p \gamma_j \times Z_{j,i} + u_i$$

При выполнении нулевой гипотезы - наличие гомоскедастичности - тестовая статистика $BP = \frac{ESS}{2} \sim \chi_p^2$ не превышает значение критическое.

Методы оценки параметров линейной регрессии в условиях гетероскедастичности.

Предположим, что мы убедились, гетескедастичность есть в наших данных, но тогда старые формулы оценки нельзя использовать. А что делать, если доверительный интервал построить нужно и гипотезы прямо срочно нужно проверить?

Возможно два сценария развития событий:

1. Утопический, если нам известна функциональная форма гетероскедастичности;
2. Реалистичная, когда нам про строение гетероскедастичности вообще ничего не известно.

Случай №1 Взвешенный МНК

Ну пусть случилось так, что у нас есть доступ к сов. секретным документам и мы знаем, как выглядит функция $f(\sigma^2)$, которая порождает условную гетероскедастичность,

Пусть она, например, выглядит вот так $f(\sigma^2) = x_i \times \sigma^2$, а регрессия имеет очень просто вид $y_i = \beta_0 + \beta_1 \times x_i + \epsilon_i$, то есть дисперсия ошибки наблюдения определяется значением параметра x для каждого i -го наблюдения. Самым простым и хитрым

способом вернуть гомоскедастичность - разделить всю регрессию на x_i и сделать простую замену вида $\tilde{y}_i = \frac{y_i}{x_i}$, чтобы не нарушать линейность модели.

Способ №2 Обобщенный МНК

Ну это простенько в скалярах, а как это выглядит в матрицах? Достаточно дружелюбно.

$$\hat{\beta} = (\tilde{X}' \Omega^{-1} \tilde{X})^{-1} (\tilde{X}' \Omega^{-1} \tilde{Y}), \text{ где } \Omega =$$

Не трудно догадаться, что в качестве весов выступают величины, обратные дисперсии, и чем больше волатильность наблюдения, тем меньше его вес в модели.

Для построения доверительных интервалов нам тоже нужно знать стандартное отклонение для оценки коэффициента, которое получают из дисперсии, просто извлекая квадратный корень. Но еще с первого семестра нам известно, что дисперсию нам вряд ли как-то можно получить, но оценки - вполне, да и по ТГМ они состоятельные.

$\text{Var}(\hat{\beta}) = (X' \Omega^{-1} X)^{-1}$, а оценка ковариационной матрицы имеет вид

$$\text{Var}(\hat{\beta}) = \frac{1}{n} \left(\frac{1}{n} X' X \right)^{-1} \times \left(\frac{1}{n} \sum_{i=1}^n \hat{\epsilon}_i^2 x_i \times x_i' \right) \left(\frac{1}{n} X' X \right)^{-1}, \text{ где } x_i - \text{это } i\text{-я строка в матрице } X.$$

Тогда как будет оценка стандартного отклонения? $\sigma_i \times I = U D^{\frac{1}{2}} U'$, где U - матрица из собственных векторов единичной длины для F D - диагональная матрица из собственных чисел F , где F - это матрица функциональной формы зависимости (читай, той самой функции $f(\sigma^2)$)

Алгоритм:

Шаг 1 Найти матрицу функциональной зависимости F , ее же в степени -0.5

Шаг 2 Сделать переход к новым переменным

$$y = X\beta + \epsilon \text{ нужно домножить на } F, Fy = FX\beta + F\epsilon$$

Шаг 3 Используй МНК

Автокорреляция

Смотри временные ряды.

Метод максимального правдоподобия.

Самый удобный метод оценивания. Пусть у нас есть параметр θ , который нужно оценить, еще есть выборка, по которой можно оценить параметр θ . Так вот ММП говорит, что нужно взять за оценку такой $\hat{\theta}$, чтобы по доступным данным вероятность была максимально. Сам метод максимального правдоподобия прост и интуитивно (да, мне нравится это слово) понятный. Нам просто нужно максимизировать вероятность события/исхода. Зачем нам второй раз проходить ММП, спросишь ты меня? Так вот, проверять гипотезы, на основе результатов ММП проще, быстрее и понятнее. А почему так? А потому что смотри на свойства оценок ММП :)

Свойства оценок максимального правдоподобия

0. Инвариантность оценок

1. Оценки максимального правдоподобия состоятельные

$$\text{plim}_{n \rightarrow \infty}(\hat{\theta}_{ML}) = \theta$$

2. Оценки максимального правдоподобия асимптотически несмещенные

$$E_{n \rightarrow \infty}(\hat{\theta}_{ML}) = \theta$$

3. Оценки максимального правдоподобия асимптотически эффективные

$$MSE(\hat{\theta}_{ML}) \leq MSE(\tilde{\theta})$$

4. Оценки максимального правдоподобия асимптотически нормально распределены

$$\hat{\theta}_{ML} \sim N(\mu, \sigma^2), \text{ где } \sigma^2 = I_{n \times n}^{-1} = -E\left(\frac{d^2 l}{d\theta^2}\right)$$

5. Оценки максимального правдоподобия просто замечательные.

6. Оценки максимального правдоподобия совпадают с оценками МНК, если ошибки наблюдений распределены нормально.

Но есть у ММП один минус, который следует из решения задачи максимизации функции правдоподобия. Если не полениться и пожалеть меня в два часа ночи, да и просто взять и решить задачу для, скажем нормального распределения, где участвует дисперсия в явном виде, тогда можно получить забавный результат, что оценки коэффициента бэты у мнк и ммп совпадают, а оценка дисперсии бэты в последнем случае будет смещенной, но состоятельной, смещенность которой будет пропадать с ростом наблюдений.

$$\hat{\sigma}_{ML}^2 = \frac{RSS}{n}$$

В матричном виде оценки принимают вид

$$\hat{\beta} = (X'X)^{-1}X'y, \hat{\sigma}^2 = \frac{(y - X\hat{\beta})'(y - X\hat{\beta})}{n}$$

Проверка гипотез при помощи тестов

На основе метода максимального правдоподобия существует три теста. Я приведу гипотезу, распределение и статистику. Все три используются для проверки наличия линейных и нелинейных ограничений.

$$H_0: \forall i \in (0, \dots, p) \theta = \theta_0 \text{ где } p \leq k$$

LM-test

$$LM = \text{grad}(\theta_0)' \times I(\theta_0) \times \text{grad}(\theta_0) \sim \chi_p^2$$

LR-test

$$LR = -2(l(\theta_0) - l(\hat{\theta}_{ML})) \sim \chi_p^2$$

Wald test

$$W = (\hat{\theta} - \theta_0)' \times I^{-1} \times (\hat{\theta} - \theta_0) \sim \chi_p^2$$

Модели бинарного выбора

Раньше мы работали с непрерывными моделями, которые не имели строгих ограничений по значению объясняющей переменной, но бывают случаи, когда модель предполагает строгий ответ - Да или Нет. В таких случаях нам нужны бинарные модели.

Но чем плоха линейная модель, спросишь ты. Давайте зададим простенькую регрессию

$$P(y_i \leq 0) = \beta_0 + \beta_1 \times x_i + u_i$$

Что мешает правой части быть больше 1 или меньше 0? Откровенного говоря, ничего, поэтому линейная модель нам не подойдет. Даже если поставить жесткие ограничения на значение объясняемой переменной, то мы потеряем нормальность ошибок - с вероятностью p_i ошибки принимают вид $\beta_0 + \beta_1 \times x_i$ и с вероятностью p_i $1 - \beta_0 + \beta_1 \times x_i$ - так как ошибки не будут нормальными, мы не сможем проверять гипотезы с помощью t-статистики.

Что же делать? Как же быть? Метрику учить в течение семестра, а не в последнюю ночь, как ты. :)

Ну да ладно, введем определение модели бинарного выбора. Как мы поняли, главная проблема у мнк состоит в том, что наши ошибки распределены не нормально и дисперсия этих ошибок зависит от x_i - о боже, гетероскедастичность! Ну и если первую проблему можно решить используя ММП, то для решения проблемы с гетероскедастичностью введем особую s-видную функцию $F(\beta_0 + \beta_1 \times x_i)$

Предположим, что у нас есть некая ненаблюдаемая функция, своего рода серый кардинал, интерпретация которой невозможна ни сейчас в 5 утра, когда я это пишу, ни в 11 утра, когда мое брэнное тело поплетется на экзамен. Но эта функция очень важна потому что она определяет наш выбор в сторону единицы или нуля для объясняющей переменной.

Y^* - это ненаблюдаемая функция

$$Y^* = \beta_0 + \beta_1 \times x_i + u_i$$

Y - это наша объясняемая функция, она связана с ненаблюдаемой функцией следующим образом.

$$\begin{cases} y_i = 1, y_i^* \geq 0 \\ y_i = 0, y_i^* < 0 \end{cases}$$

Тогда можно развернуть модель бинарного выбора следующим образом (как Галя Вордпресс на Азуре)

$$P(y_i = 1) = P(y_i^* \geq 0) = P(\beta_0 + \beta_1 \times x_i + u_i \geq 0) = P(u_i \geq -(\beta_0 + \beta_1 \times x_i)) = P(u_i \leq \beta_0 + \beta_1 \times x_i) = F(\beta_0 + \beta_1 \times x_i)$$

в силу того, что ненаблюдаемая функция симметрична относительно нуля. Красиво, да?

Осталось лишь придумать это сигмовидную функцию. Ну за нас это уже сделали.

Логит и пробит модели

Логит и пробит - это две неразлучные сестрички, среди которых нет лучшей или отстающей, ну... правда логит попышнее в 1.6 раз, чем пробит, но это мелочи. Единственное различие между ними - это спецификация формы распределения ϵ

У пробит $\epsilon \sim N(0,1)$

У логит $\epsilon \sim \text{logistic}$, где $\text{logistic}, f(t) = \frac{e^{-t}}{(1+e^{-t})^2} = \frac{e^t}{(1+e^t)^2}$

Ну и приготовились брать интегральчик для мини-доказательства того, что $P(y_i = 1) = F(t), t = \beta_0 + \beta_1 \times x_i$

$$\begin{aligned} F(t) = P(\epsilon_i \leq t) &= \int_{-\infty}^t f(t) dt = \int_{-\infty}^t \frac{e^t}{(1+e^t)^2} dt = \int_{-\infty}^t \frac{d(e^t)}{(1+e^t)^2} = \left(-\frac{1}{1+e^t}\right) \Big|_{-\infty}^t \\ &= -\frac{1}{1+e^t} + 1 = \frac{e^t}{(1+e^t)^2} \end{aligned}$$

Ну все это классно, но что-то совсем непонятно, что такое t . Функция t теперь имеет красивую интерпретацию отношения шансов, которая правда существует только в логит модели.

$$\ln\left(\frac{P(y_i = 1)}{P(y_i = 0)}\right) = \ln\left(\frac{\frac{e^t}{(1+e^t)^2}}{1 - \frac{e^t}{(1+e^t)^2}}\right) = \ln(e^t) = t = \beta_0 + \beta_1 \times x_i$$

то есть с ростом x на 1 единицу отношение победы к проигрышу (условно) растет на $\beta_1 \times 100\%$

Оценка параметров и качества подгонки модели

Где-то я упомянула, что решают проблему ненормальности оценок с помощью ММП. ну так вот, чтобы искать оценки β нужно использовать метод ММП.

Так как мы где-то очень далеко наверху файла предполагали независимость наблюдений, то справедливо вот что:

$$L((y_i, X|\beta)) = \prod_{i=0}^n (F[t(x_i)]) \times \prod_{i=0}^n (1 - F[t(x_i)])$$

В логарифмический вид переведи сам плиз

$$\frac{dl}{d\beta} = F(t) \times (1 - F(t))$$

$$\frac{d^2 l}{d\beta^2} = F(t) \times (1 - F(t))(1 - 2F(t)) = H = -I_{n \times n}$$

$$Var(\hat{\beta}) = I_{n \times n}^{-1}$$

Для проверки гипотез использовать тесты МП, множителей Лагранжа и Вальда.

Для оценки качества регрессии тут не пойдет просто R^2 уже по понятным причинам :0

Поэтому существует три аналога для оценки качества подгонки модели

$$1. \text{ Коэффициент детерминации МакФаддена } R_{MF}^2 = 1 - \frac{\hat{l}}{l_0}$$

$$2. \quad R_p^2 = \frac{1}{1 + \frac{2}{n}(\hat{l} - l_0)} \text{ Псевдо } R^2$$

$$3. \quad \text{Сравнение точности прогнозирования } R^2 = 1 - \frac{W_1}{W_0}$$

Где $W_1 = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$ - доля неверных прогнозов, $W_0 =$

$$\begin{cases} 1 - \hat{p}, & \text{if } \hat{p} > 0.5 \\ \hat{p}, & \text{if } \hat{p} \leq 0.5 \end{cases}$$

- доля неверных прогнозов в модели из константы

ЭЭЭЭххх, по-хорошему надо еще писать про предельные эффекты, то уже полтретьего дня, а я еще не гугу в задачи, поэтому поговорим о ROC кривой.

ROC curve

1. чувствительность - на сколько хорошо модель предсказывает 1, в зависимости от cutoff- значения. Cutoff - это вот то значение ненаблюдаемой функции, которое мы выбираем для признания объясняющей переменной за единичку.
2. Специфичность - тоже самое, что и чувствительность, но относительно 0
3. ROC-кривая - это на сколько хорошо модель предсказывает 1 в координатах чувствительность и 1- специфичность.

Если получаем просто прямую - товарищ, да ты на кофейной гуще гадаешь Если получаем выпуклую кривую - ну норм, на правду похоже Если получаем вогнутую кривую - все очень плохо (Картинка с уткой)

Модуль 4

Эндогенность

Ну понеслась... мне тоже страшно, а еще на меня столько всего плохого свалилось, что очень хочется заработать все плохое, потому что ну пусть мне плохо, так хотя бы тебе, читатель, помогу. Говорят, что добрые дела возвращаются, может вернется и мне? Странный эпиграф, но это не покровка11, поэтому пишу, как ~~могу~~ хочу, так что допустим.

Вернемся к любимым матрицам и векторам. Раньше мы считали, что все регрессоры статичны и просто даны, но в жизни это не так, потому что ну где это видано, чтобы

нам заранее был дан уровень безрисковой ставки заранее? Поэтому мы смело меняем предпосылки модели из ТГМ во что-то более масштабное. (читать голосом Билла Шифера)

1. Теперь все x_i стохастические, то есть случайные.
2. При любой реализации матрица X имеет ранг k , по количеству x , а не наблюдений.
3. $\exists \text{plim}_{n \rightarrow \infty} \frac{1}{n} (X^T X)$
4. $\text{plim}_{n \rightarrow \infty} \frac{1}{n} X^T \epsilon = 0$

В этом случае МНК оценки все еще синие, но хм... последнее условие сильное и далеко не всегда сбывается в жизни, поэтому возникает умное слово “эндогенность”, которое по своей сути описывает долгое и продолжительное “Б#\$%&B” эконометристов, потому что при эндогенности оценки МНК несостоятельные и даже асимптотически не являются несмещенными. Так происходит потому, что часть или все регрессоры коррелируют с ошибкой, а значит, что роль неопределенности увеличивается, что серьезно ухудшает прогнозы. Ну сами посмотрите.

$$\hat{\beta} = \beta + \left(\frac{1}{n} X^T X\right)^{-1} \times \left(\frac{1}{n} X^T \epsilon\right)$$

$$\text{plim}_{n \rightarrow \infty} \hat{\beta} = \beta + \text{plim}_{n \rightarrow \infty} \left(\frac{1}{n} X^T X\right)^{-1} \text{plim}_{n \rightarrow \infty} \left(\frac{1}{n} X^T \epsilon\right) \neq \beta$$

Но дорогу осилит идущий, поэтому были придуманы инструментальные переменные, не то что бы прямо приятный способ, но это лучше, чем ничего.

Инструментальные переменные - это такие переменные, которые

1. Хорошо коррелируют с изначальными переменными.
2. Совсем не коррелируют с ошибкой.

$$\text{plim}_{n \rightarrow \infty} \frac{1}{n} Z_i \epsilon = 0$$

В общем-то сократительный смысл в том, чтобы просто сделать замену переменных. Вот пусть есть плохой X в $Y = \beta_0 + \beta_1 X + \epsilon$, коррелирующий с ошибкой, да еще и сам какой-то неуравновешенный, вот давай его заменим на хороший Z . Ну класс, $Y = \beta_0^{IV} + \beta_1^{IV} Z + \epsilon$.

$$\text{При таком раскладе получим, что } \beta^{IV} = \frac{c\hat{v}(Z, Y)}{c\hat{v}(Z, X)} = \frac{c\hat{v}(Z, \beta_0 + \beta_1 X + \epsilon)}{c\hat{v}(Z, X)} = 0 + \beta + \frac{c\hat{v}(Z, \epsilon)}{c\hat{v}(Z, X)}$$

стремно, но давай возьмем предел по вероятности этой гадости.

$$\text{plim}_{n \rightarrow \infty} \beta + \text{plim}_{n \rightarrow \infty} \frac{c\hat{v}(Z, \epsilon)}{c\hat{v}(Z, X)} = \beta + \frac{\text{plim}_{n \rightarrow \infty} \frac{1}{n} Z^T \epsilon}{\text{plim}_{n \rightarrow \infty} \frac{1}{n} Z^T X} = \beta$$

Чего-то хорошего о распределении бэти инструментальных переменных сказать нечего, но в больших выборках оно стремится к нормальному распределению с дисперсией

$$\sigma_{\beta^{IV}}^2 = \frac{\sigma_{\epsilon}^2}{\sum_{i=1}^2 (X_i - \bar{X})^2} \times \frac{1}{\hat{\rho}_{XZ}}$$

осторожно, хардкор!

Вот есть $\hat{\beta}^{IV} = (Z^T)^{-1} Z^T Y$, а как же наша первая бэточка будет оцениваться?

Пусть количество инструментальных переменных меньше количества наших стохастических регрессоров, тогда можно получить состоятельную оценку для первоначальной бэты через двухшаговый МНК.

Шаг1: Проецируем каждый вектор X в пространство векторов Z , вспоминая про матрицу-шляпку, да-да, первый семестр вспомнить придется...чуть-чуть.

$$X = \alpha Z + \epsilon$$

Где оцененный(прогнозируемый) $\hat{X} = Z(Z^T Z)^{-1} Z^T X$

Шаг2: Берем вектор оценок X из первого шага и запикиваем вместо изначального X в исходную регрессию.

$$Y = \beta \hat{X}$$

Шаг3 (ироничный для двухшагового МНК): Оцениваем МНК как обычно

$$\hat{\beta} = (\hat{X}^T \hat{X})^{-1} \hat{X}^T Y = (Z(Z^T Z)^{-1} Z^T X)^{-1} (Z(Z^T Z)^{-1} Z^T Y)$$

Самый главный вопрос, а где взять эти инструментальные переменные. Ну... это открытый вопрос, пальцем в небо можно попробовать тыкнуть.

Инструментальные переменные - метод в теории хороший и дает состоятельные оценки, но они все-таки неэффективные, а поэтому лучше его использовать, когда в модели проблема эндогенности стоит остро, как у меня сдать метрику, например.

Для этого был придуман тест Дарбина Ву Хаусмана.

$$H_0: \forall \lim_{n \rightarrow \infty} \frac{1}{n} X_j^T \epsilon = 0$$

$$H_a: \exists \lim_{n \rightarrow \infty} \frac{1}{n} X_j^T \epsilon \neq 0$$

$$\text{Тестовая статистика } H = (\hat{\beta}^{IV} - \hat{\beta})^T \times (\hat{V}(\hat{\beta}^{IV}) - \hat{V}(\hat{\beta}))^{-1} \times (\hat{\beta}^{IV} - \hat{\beta})$$

$$H > \chi_{k-1, 1-\alpha}^2$$

Системы однородных уравнений

Предыстория. Жил был экономист Дж.Кейнс и был он очень умным, и еще был экономист Хигс, который идеи Кейнса формализовал и перевел на язык уравнений, а

еще был экономист Филлипс, который заметил связь между инфляцией и безработицей. А еще были прикольные чиновники, которые включили в Ad Нос модели кривую Филлипса и сломали экономику нафиг. Так о чем это я... ну во-первых, если ты не умеешь в метрику, то не суйся в ЦБ, сломаешь последнее, во-вторых, коли ты дочитал до этого момента, то ты почти уже умеешь в метрику и ты априори крут, в-третьих, смещение в одновременных уравнениях - это не только в теории плохо, но и может разломать экономику страны.

А что же такого плохого сделали чиновники, когда приняли на веру идею Филлипса? Во-первых, связь инфляции и безработицы сложнее простой гиперболки, а во-вторых ну вот как так можно? Пусть есть индекс цен потребительской корзины (ИЦПК), связанный с темпом прироста зарплаты, и темп прироста зарплаты, выраженный через ИЦПК и безработицу.

$$\begin{cases} p = \beta_0 + \beta_1 w + u_p \\ w = \alpha_0 + \alpha_1 p + \alpha_2 U + u_w \end{cases}$$

Такая система одновременных уравнений включает двусторонние связи, и казалось бы, вырази темп прироста зп, впиши его в первое уравнение и живи спокойно. Но вспомни, что в нашем мире все неопределенное, стремное и непонятное и больше так не думай, а еще не смей лезть в стохастическое исчисление, а то узнаешь, что теорема Пифагора не всегда верна и сломаешь себе психику. Не повторяй моих ошибок, рил. К чему тут стохан, спросишь ты меня? Ибо будет два источника неопределенности будет.

$$p = \beta + \beta_1(\alpha_0 + \alpha_1 p + \alpha_2 U + u_w) + u_p = \beta + \beta_1 \alpha_0 + \beta_1 \alpha_1 p + \beta_1 \alpha_2 U + \beta_1 u_w + u_p$$

$$p = \frac{\beta_0 + \beta_1(\alpha_0 + \alpha_2 U) + \beta_1 u_w + u_p}{1 - \beta_1 \alpha_1}$$

Пришло время разобраться с понятиями эндогенный и экзогенный, представим, что вы этого не знаете, а курсы Оксаны Анатольевны сдали с магией и письменным разрешением на её применение из Хогвартса.

Эндогенные переменные - это те переменные, которые определяются взаимодействием соотношений в модели, это наши p и w .

Экзогенные - это переменные, определяющиеся внешним миром по законам мироздания, которые только боженьке известны, это наша U .

Математические зависимости в приведенной форме - это зависимости, которые эндогенные переменные выражают через экзогенные. Если в системе есть уравнение, которое эндогенную переменную выражает через эндогенную, то это называют рекурсия уравнением в структурной форме, которое нужно приводить к приведенной.

Вернемся к нашей системе. Когда мы привели нашу систему к виду $p = p(U, u_p, u_w)$, мы приобрели новый источник неопределенности, что сделало оценки МНК неочень хорошими. Если каждое уравнение оценивать отдельно в этой системе через МНК, то для $\hat{\beta}_1 = \beta_1 + \frac{\sum(u_{p,i} - \bar{u}_p)(w_i - \bar{w})}{\sum(w_i - \bar{w})^2}$ появляется стохастический регрессор.

$$\begin{aligned}
\text{plim } \hat{\beta}_1 &= \hat{\beta}_1 + \frac{\text{plim } \frac{1}{n} \sum_{i=1}^n (u_{p,i} - \bar{u}_p)(w_i - \bar{w})}{\text{plim } \frac{1}{n} \sum_{i=1}^n (w_i - \bar{w})^2} = \beta_2 + \frac{\text{cov}(u_p, w)}{\text{var}(w)} \\
&= \beta_1 + (1 - \alpha_1 \beta_1) \frac{\alpha_1 \sigma_{u_p}^2}{\alpha_2^2 \sigma_U^2 + \sigma_{u_w}^2 + \alpha_1^2 \sigma_{u_p}^2} \neq \beta_1
\end{aligned}$$

Значит МНК оценка смещена, а направление смещения определяется знаком $(1 - \alpha_1 \beta_1)$. В нашем случае оценка завышена из экономических соображений. Значение плима можно считать прокси для смещения

Что делать?

Использовать метод инструментальных переменных. Нужно найти такой регрессор, чтобы она коррелировала с w , не коррелировала с шоками. Тут далеко ходить не нужно, обратим внимание на экзогенную переменную U как инструмент для w .

$$\hat{\beta}^{IV} = \frac{\sum_{i=1}^n (U_i - \bar{U})(p_i - \bar{p})}{\sum_{i=1}^n (U_i - \bar{U})(w_i - \bar{w})} = \beta_1 + \frac{\sum_{i=1}^n (U_i - \bar{U})(u_{i,p} - \bar{u}_{i,p})}{\sum_{i=1}^n (U_i - \bar{U})(w_i - \bar{w})}$$

Берем предел по вероятности от выражения выше и плачем от счастья, но не долго. Для одного уравнения мы смогли придумать инструментальную переменную. а что делать со вторым? Если использовать U для индекса цен как инструментальную переменную, то появится проблема совершенной мультиколлинеарности, что сделает оценки коэффициентов бессмысленными и невозможными. Для этого была придумана такая процедура, как индентификация модели. Цимусидеи можно уложить в одно предложение:

“Если число экзогенных переменных больше или равно числу эндогенных переменных за вычетом одного, то модель индентифицируема.”

По-научному это звучит следующим образом:

“Пусть i -е поведенческое уравнение модели индентифицируемо. Тогда справедливо неравенство:

$$N_i \geq G - M_i$$

N_i - количество экзогенных переменных, которые НЕ включены в итом уравнении, G - количество эндогенных переменных, M_i - количество эндогенных переменных, которые НЕ включены в итое уравнение”.

Индентифицирована - это значит, что все коэффициенты можно однозначно оценить. Все было бы хорошо, если бы не очередное “но”. Если в системе уравнений есть одно уравнение, которое определяет эндогнную переменную только через эндогенные, и хотя бы одно уранение, определяющее эндогенную переменную через экзогенные переменные так, что суммарно число эндогенных и экзогенных переменных равно, то нет, вот те уравнения, которые описывают связи через эндогенные переменные неиндентифицируется. Кислотный хнык, дамы и господа, а оценивать все-таки нужно и такие системы, котрых, просто поверьте моему опыту работы, много.

Поэтому используют двухшаговый метод МНК.

Шаг 1: Берем все экзогенные переменные в переопределенном уравнении и запишем их в единую линейную функцию. Говорим, что это прогнозы переопределенных эндогенных переменных и запишем это “прогнозы” в неидентифицируемое уравнение.

Шаг 2: Оцениваем через МНК это новое уравнение

Из всего этого приходит на ум логичное обобщение, что в системе с G эндогенными переменными, каждое уравнение должно содержать хотя бы $G - 1$ экзогенных переменных, чтобы оно идентифицировалось. Вот такое вот правило размерности или ранга. Радует.

Еще немного науки о системах уравнений. Поговорим немного о косвенном МНК.

Косвенный метод наименьших квадратов - еще одна симпатичная процедура для получения состоятельных оценок в одновременных системах уравнений на уровне 7-го класса. Вот продолжим мутузить нашу систему.

$$\begin{cases} p = \beta_0 + \beta_1 w + u_p \\ w = \alpha_0 + \alpha_1 p + \alpha_2 U + u_w \end{cases}$$

Что если сделать уравнения экзогенными по принципу подставь одну штуку в другую штуку.

$$\begin{cases} p = \frac{\beta_0 + \alpha_0 \beta_1 + \alpha_2 \beta_1 U + u_p + \beta_1 u_w}{1 - \alpha_1 \beta_1} \\ w = \frac{\alpha_0 + \alpha_1 \beta_0 + \alpha_2 U + u_w + \alpha_1 u_p}{1 - \alpha_1 \beta_1} \end{cases}$$

Теперь наши уравнения имеют вид

$$\begin{cases} p = \gamma_0 + \gamma_1 U + \epsilon_1 \\ w = \gamma_3 + \gamma_4 U + \epsilon_2 \end{cases}$$

Найдем прогнозы эндогенных параметров.

$$\begin{cases} \hat{p} = \hat{\gamma}_0 + \hat{\gamma}_1 U \\ \hat{w} = \hat{\gamma}_3 + \hat{\gamma}_4 U \end{cases}$$

Ну и по классике снова делаем систему уравнений

$$\begin{cases} \frac{\hat{\beta}_0 + \hat{\alpha}_0 \hat{\beta}_1}{1 - \hat{\alpha}_1 \hat{\beta}_1} = \hat{\gamma}_0 \\ \frac{\hat{\alpha}_2 \hat{\beta}_1}{1 - \hat{\alpha}_1 \hat{\beta}_1} = \hat{\gamma}_1 \\ \frac{\hat{\alpha}_0 + \hat{\alpha}_1 \hat{\beta}_0}{1 - \hat{\alpha}_1 \hat{\beta}_1} = \hat{\gamma}_2 \\ \frac{\hat{\alpha}_2}{1 - \hat{\alpha}_1 \hat{\beta}_1} = \hat{\gamma}_3 \end{cases}$$

То, что идентифицируемо, мы найдем, а на нет и суда нет.

Но без ответа остались два вопроса:

1. Как слету определить, какие идентифицируются уравнения?
2. Как быстро найти неидентифицируемые переменные?

Начнем со второго вопроса - на него отвечает правило порядка.

$$(a_{i1} \quad \dots \quad a_{ii} \quad \dots \quad a_{Gi}) \times \begin{pmatrix} y_{t1} \\ y_{t2} \\ \dots \\ y_{ti} \\ \dots \\ y_{Gt} \end{pmatrix} + (b_{i1} \quad \dots \quad b_{ii} \quad \dots \quad b_{iK}) \begin{pmatrix} x_{1t} \\ x_{it} \\ \dots \\ x_{Kt} \end{pmatrix}$$

G - количество эндогенных переменных в модели, K - количество экзогенных переменных в модели, $a_{ii} = 1$ - условие нормализации.

Ответ на первый вопрос - в условии идентификации поставить “не” везде, где справедливо

Time series

Моя любимая тема, потому что временной ряд - это прям почти классическая база данных, теперь наблюдения не являются независимыми, теперь данные - это последовательность, которая является частью бесконечной последовательности, имя которой время. Но пока мы не играемся с панельными данными, поэтому весь хард впереди. :)

Почти в течение всей этой темы надо понимать одну важную вещь - мы работаем с рядом, с прямой, с вектором, а не с перекрестными данными. Иными словами, забудьте нафиг все, что я писала наверху, на некоторое время, потому что временные ряды - это особый мир, с особыми правилами, с особенными регрессиями. ~~они не прощают ошибок~~

Статические и динамические модели

Статические модели отличаются от динамических тем, что в них нет лагов. Если по-честному, то мы почти весь год ещё кхе-кхе... занимаемся статическими моделями. В

принципе в статических моделях все очень хорошо, можно использовать МНК, но только с логарифмизацией показателей, чтобы убрать тенденцию к нарастанию или резкому убыванию показателей, так как множество процессов в жизни - это "накопительные" процессы. Динамические модели намного интереснее, потому что они лагают.:D

Вообще странно предполагать, что наблюдения независимы во временных рядах, ну вот если ты мне скажешь, что значение реального ВВП в прошлом квартале не зависит от позапрошлого я буду долго смеяться, еще как зависит! Для того, чтобы это зависимость отобразить были придуманы лаги и лаговая структура, ну или по простому включение в регрессию значения регрессанта за прошлые периоды... это правда порождает проблему мультиколлинеарности со всеми вытекающими, которая решается введением простой лаговой структуры с распределением Койка (коэффициентами перед в лаговой структуре - это убывающая геометрическая прогрессия, хз было ли это на лекции.)

Автокорреляция

Уже есть понимание, что наши объясняемые переменные скорее всего как-то связаны друг с другом, а потому нужно силу взаимосвязей как-то померить, вот для этих целей придумали автокорреляцию.

Допущения для временных рядов

1. Модель линейна
2. Временные ряды для регрессоров слабо устойчивы (наблюдаемые иксы случайно получаются). Сомнительное допущение
3. Нет строгой взаимосвязи между регрессорами в выборке
4. Мат.ожидание случайного члена равно 0, а еще он гомоскедастичен
5. Значения случайного члена распределены независимо друг от друга
6. Случайный член независимо распределен от регрессоров и условное мат.ожидание его тоже равно 0
7. Случайный член имеет нормальное распределение.

Вот теперь можно говорить серьезно про автокорреляцию. Начнем с определения. Ну все просто же, автокорреляция или серийная корреляция - это когда значение случайного члена устанавливается НЕ независимо от значений случайного члена в остальных наблюдениях, а значит $cov(u_t, u_{t-a}) \neq 0 \forall a \in Z$. Автокорреляция во многом похожа на гетероскедастичность, поэтому если в лоб применить МНК ко временным рядам, то о проверке гипотез можно забыть, хотя в общем-то оценки коэффициентов остаются несмещенными и состоятельными, но они неэффективные из-за очень высокой дисперсии.

Главные причины, почему такое явление вообще есть, это:

1. Систематические ошибки измерения, аля вчера мерили в кг, а сегодня в граммах гречу.
2. Пропущенные переменные. Тут просто, влияние этой переменной никто не отменял, но так как оно не наблюдаемое, то оно сидит в ошибке и продолжает оказывать влияние на наш регрессант и коррелировать с прошлым значением.
3. Неправильная форма модели
4. Инерция в модели, то есть если наши изначальные переменные коррелируют, то и остатки тоже будут.

Автокорреляция характерна больше всего для временных рядов, потому что чаще всего в таких моделях забывают про лаг, который в таком случае автоматом включается в случайный член. В общем основной причиной автокорреляции является пропуск этого самого лага, чаще всего она положительна, так как накопительная особенность.

Взглянем на автокорреляцию поближе. Она бывает разного порядка в зависимости от того, сколько случайных членов из разных периодов мы включим в модель. Так $AR(1)$ - это обозначение автокорреляции первого порядка, которая имеет вид:

$$\begin{cases} y_t = \beta_0 + \beta_1 x_t + u_t \\ u_t = \rho u_{t-1} + \epsilon_t \end{cases}$$

Где $u_t \sim WN(0; \sigma^2), \rho \neq 0$

Основной альтернативой автокорреляционному процессу является процесс скользящего среднего, вот внизу $MA(1)$

$$\begin{cases} y_t = \beta_0 + \beta_1 x_t + u_t \\ u_t = \lambda_0 \epsilon_t + \lambda_1 \epsilon_{t-1} \end{cases}$$

Чтобы обнаружить автокорреляцию, проводят тест Дарбина-Уотсона, которая рассчитывается по величинам отклонений с помощью выражения.

$$d = \frac{\sum_{i=2}^T (e_t - e_{t-1})^2}{\sum_{i=1}^T e_t^2}$$

Которая на больших выборках стремиться к $d \rightarrow 2(1 - \rho)$, где ρ - это корреляция между остатками.

У нее нет какого-то особенного распределения, поэтому для нее это просто условия для отрицательной и (внимание) неотрицательной автокорреляции.

Несложно догадаться, что она принимает значения от 0 до 4, но как проверять гипотезы, если нет определенного распределения. Дарбин и Уотсон придумали целую таблицу значений статистики DW^l и DW^u , которые определяются в зависимости от количества наблюдений и уровня значимости. Есть три варианта расположения нашей тестовой статистики относительно этого интервала:

1. $DW > 4 - DW^l$ - имеет место отрицательная корреляция
2. $DW \in [4 - DW^u; 4 - DW^l]$ - неопределенность, тест не дает результата

3. $DW < 4 - DW^u$ - автокорреляция отсутствует

r(картинку сюда)

Тест Дарбина Уотсона достаточно старый, поэтому есть ряд ограничений на его использование.

Так, например, его НЕ ИСПОЛЬЗУЮТ, если

1. Есть лаг регрессантов в качестве регрессора
2. Нет свободного члена
3. Маленькая выборка
4. Распределение отличается от нормального
5. Есть гетероскедастичность
6. Есть авторегрессия вышних порядков

И поэтому необходимо было разрабатывать тесты на автокорреляцию с ослаблением предпосылок, так и появился тест Бройша-Годфри.

Этот тест делается в 4 шага.

1. Оцениваем оригинальную модель (а то китайские неоч) и рассчитываем для нее вектор ошибок прогноза $e = y - \hat{y}$
2. Оцениваем вспомогательную регрессию

$$e_t = \beta_0 + \beta_1 e_{t-1} + \dots + \beta_{t-p} + u_t$$

3. Проверяем гипотезу об адекватности модели

$$H_0: \forall \beta_i = 0 \quad H_a: \exists \beta_i \neq 0$$

4. Проверяем расчетную статистику

$T \times R^2 \sim F_{p, T-p-1}$, где T - количество наблюдений, R^2 - коэффициент детерминации из вспомогательной регрессии, p - количество лагов во вспомогательной регрессии.

Ну ок, убедились мы, что автокорреляция, как и гетероскедастичность - это плохо, что делать?

1. Проверить модель на пропуск переменных, через тест Рамсея например. (ну мало ли, на старуху тоже бывает проруха)
2. Переход к взвешенным разностям.

Чтобы это сделать, нужно найти оценку ρ , например, если мы уверены, что имеем дело с автокорреляцией первого порядка, то можно использовать МНК из $e_i = \rho e_{i-1} + u_i$ или $\hat{\rho} = 1 - \frac{d}{2}$ из теста Дарбина Уотсона.

Но вся жесьь впереди, молодой падаван.

Есть уравнение

$$y_t = \beta_0 + \beta_1 x_{1,t} + \beta_2 x_{2,t} + \dots + \beta_m x_{m,t}$$

Сдвинем его на шаг назад

$$y_{t-1} = \beta_0 + \beta_1 x_{1,t-1} + \beta_2 x_{2,t-1} + \dots + \beta_m x_{m,t-1}$$

Домножим на ρ

$$\rho y_{t-1} = \rho \beta_0 + \rho \beta_1 x_{1,t-1} + \rho \beta_2 x_{2,t-1} + \dots + \rho \beta_m x_{m,t-1}$$

Вычитаем последнее уравнение из первого

$$y_t - \rho y_{t-1} = \beta_0 - \rho \beta_0 + \beta_1 x_{1,t} - \rho \beta_1 x_{1,t-1} + \beta_2 x_{2,t} - \rho \beta_2 x_{2,t-1} + \dots + \beta_m x_{m,t} - \rho \beta_m x_{m,t-1}$$

Чтож... мы потеряли одно наблюдение, но решили проблему автокорреляции первого порядка. Нужно решить проблему автокорреляции m порядка? Вычитай из первого уравнения m вспомогательных уравнений, каждый раз сдвигаясь на 1 шаг назад. Последнее из таких уравнений должно содержать m лагов и домножаться на коэффициент корреляции текущего момента t на момент $t - m$.

Но есть еще один вариант избавления от автокорреляции. В конце концов, автокорреляция ну очень похожа на гетероскедастичность, и коли точный прогноз у нас неплохой, то ну и почему бы просто не оценить ковариационную матрицу остатков с поправкой и продолжать радоваться жизни? Действительно!

Вот так, дети мои, придумали оценку ковариационной матрицы остатков регрессии в форме Ньюи-Веста. Сырбор в том, что такая оценка является обобщением на случай наличия гетероскедастичности и автокорреляции одновременно, поэтому отмотайте назад на тему гетероскедастичности в раздел ОМНК и ошибок Уайта.

Стационарные и нестационарные временные ряды

Теперь поговорим о самом важном за этот модуль, о том что такое стационарный ряд и почему он невероятно крут. Возможно, у тебя сложилось впечатление, что временные ряды - это совсем просто, ну рилли, что сложного в модели, где регрессоры - это регрессанты прошлых периодов и несколько ошибок из прошлых наблюдений?

Только посмотрите на эту милоту $ARMA(1,1)$:3

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 u_{t-1} + u_t$$

Но в тихом омуте, как говорится, черти водятся...

Так вот, стационарный ряд - это ряд, у которого теоритическое мат.ожидание и дисперсия не зависят от времени, а ковариационная матрица между значениями в разные моменты времени не зависит от количества лагов в модели.

Сложно... а есть прям на пальцах, то этот ряд не должно так сильно штормить, как парочку моих знакомых после знатной пьянки на утро.

Если ряд нестационарный, то прогнозирование и вообще какая-либо оценка бесполезны.

А как проверить ряд на стационарность? Ну... Если у тебя уже был курс диффуриков, то радуйся, не было - гугли.

Нам потребуются разностные уравнения, уиии. Берем наш ряд, оцениваем его по “рабочекрестьянски” бобриным способом через МНК. Ииии...

$$\hat{y}_t = \hat{\beta} y_{t-1}$$

Составляем характеристическое уравнение.

$$0 = \hat{\beta} \lambda$$

$$\lambda = 0$$

Если эта $|\lambda| \leq 1$, то ряд стационарный и не содержит лагов из будущего.

Ряд нестационарен чаще всего потому, что

1. Есть тренд, то есть мат.ожидание процесса изменяется во времени в определенном направлении.
2. Лаговый полином ошибок регрессии необратим, то есть с ростом выборки ошибки очень старых наблюдений продолжают оказывать значимое влияние на существующий регрессант.
3. Нестационарные остатки.

Интересности про лаговый полином. Есть такая интересная штука - случайное блуждание, которое сохраняет влияние шока прошлых периодов на сегодняшний игрек, в общем случае, когда $\beta_2 = 1$ в регрессии $y_t = \beta_0 + y_{t-1} + \epsilon_t$, то каждое следующее наблюдение будет сохранять влияние прошлых шоков с поправкой на нынешний шок, random walk - это прокс макроэкономических шоков в модели с наивными ожиданиями, когда лучший прогноз на завтра - это значение переменной сегодня. В таком случае матожидание и дисперсия регрессора не имеют безусловных значений, они меняются во времени и зависят от анализируемого момента. А это значит что? Правильно случайное блуждание - это стохастически тренд.

Детерминированный тренд имеет вид:

$$y_t = \beta_0 + \beta_1 t + u_t$$

Если взять разности от детерминированного тренда, то неожиданно получим стационарный ряд. Вообще, если от ряда можно взять бесконечное число разностей и получить стационарный ряд, то его называют *процессом, стационарным в разностях*, коим примером является наше случайное блуждание. Если нестационарный ряд может быть преобразован в стационарный ряд взятием всего одной разности, то его называют интегрированным процессом первого порядка, или в простонародье - *ARIMA(p, q, 1)*. Классно же ну?

Каковы последствия нестационарности?

По-хорошему все очень плохо с регрессиями, которые основаны на анализе нестационарных рядов, потому что ухудшаются свойства оценок и повышается риск оценки (прости, Господи) мнимой регрессии.

По порядку

1. Оценки совсем плохи, потому что они даже не несмещенные.
2. Мнимые регрессии - это плохо. В модели, где временные ряды не стационарны, может быть такая ситуация, что обе переменные содержат временной тренд одного вида, но они, например, связаны друг с другом как-то не напрямую. Вот, например, есть я, которая ну очень любит фрукты и овощи, но при этом еще и метрику много решаю, ибо будет плохо, если делать я этого не буду. Я не придерживаюсь какого-то плана в еде или графика в решении задач, а делаю оба действия как бог на душу положит, но есть тренд, что перед летней сессией я много решаю задачек и хомячу черешню. Так вот, если оценить влияние поедания черешни на количество верно решенных задачек в листках к семинарам, то R^2 будет очень даже приличный, что приводит к наивной мысли, что количество черешни влияет на мою учебу, но это не так, я просто до 20% не много читаю, решаю и нервничаю. Но иногда мнимая связь не так и очевидно мнимая. Например, так часто бывает в макроэкономике, и Гренджер и Ньюболд показали, что такое явление может быть и в интегрированных временных рядах (ARIMA), и даже в при использовании стационарных рядов, если в последних игнорируется автокорреляция в случайном члене.

Как же эту нестационарность искать?

Вот действительно, анализ временных рядов - это тихий ужас, потому что, когда ты получаешь этот стремный ряд, то уже на уровне визуализации даже мне порой хочется громко рыдать. В макре, например, большая часть временных рядов подчинено долгосрочной тенденции и уже априори нестационарны, поэтому надо проводить дополнительные процедуры для поиска ответов на бесконечное количество побочных вопросов аля "стационарен ли ряд в разностях или трендово стационарен?", "Если стационарен в разностях, то каков у него порядок интегрирования" и прочее. Это нудно, но все нужно проверить перед анализом, чтобы сам анализ не был пустой тратой времени.

Наиболее приятный и красивый с точки зрения графичков метод обнаружения нестационарности является коррелограмма. Коррелограмма - это график автокорреляционной функции, который показывает, как каждое из прошлых наблюдений коррелирует с текущим. Она используется для идентификации порядков p, d, q в ARIMA.

Сама автокорреляционная функция имеет вид

$$\rho_k = \frac{E(x_t - \mu_x)(x_{t+k} - \mu_t)}{\sqrt{E(x_t - \mu_x)^2 E(x_{t+k} - \mu_t)^2}} \quad \forall k = 1, \dots$$

И для некоторых функций она имеет специфический и узнаваемый вид.

Для AR(1) - гипербола в первом квадранте

Для $MA(q)$ - это q первых лагов отличны от нуля, а потом все лаги равны 0.

Для случайного блуждания - это что-то отдаленно похожее на логарифм.

Возвращаясь в плоскость нестационарных рядов - для них коэффициенты автокорреляции не определены, но можно оценить математическое ожидание этих коэффициентов. В теории для длинных рядов эти коэффициенты убывают и затухают.

К сожалению, метод коррелограммы достаточно простой, а потому существуют проблемы для его использования, а именно коррелограммы математического ожидания для коэффициентов в нестационарном ряду ну вот до полного смешения похожи на стационарный процесс $AR(1)$, а так же в коротких нестационарных рядах коэффициенты очень быстро убывают.

Второй и более формальный с точки зрения иерархического способ является поиск единичного корня или тест Дики-Фуллера.

Модели Бокса-Дженкинса

Регрессионные динамические модели. Модели с распределенными лагами

Их не будет, а я ОЧЕНЬ хочу спать

Модели панельных данных