

Лекции по эконометрическому моделированию

Автор конспекта: Булыгин М.Е.

Лектор: Нерадовская Ю.В.

СПбГЭУ

Санкт-Петербург

1 июня 2021 г.

Содержание

1	Принципы и методы эконометрического моделирования	2
1.1	Эконометрическое моделирование как метод моделирования социальноэкономических процессов .	2
1.2	Этапы эконометрического моделирования	4
1.3	Сбор и анализ данных	5
1.4	Принципы эконометрического моделирования	5
1.5	Методы эконометрического моделирования	6
2	Обзор математико-статистического инструментария эконометрического моделирования	7
2.1	Методы предварительного анализа данных	7
2.2	Отбор факторов в модель регрессии	15
2.3	Классическая нормальная линейная модель	16
2.4	Анализ качества случайных остатков	17
2.5	Оценка значимости параметров регрессии (тренда)	19
2.6	Односторонние t-тесты	26
2.6.1	Практическое использование односторонних t-тестов	31
2.7	Статистические характеристики оценок параметров уравнения регрессии	33
2.8	Оценка значимости регрессии (тренда)	37
2.9	Последствия неправильной спецификации модели	39
2.10	Оценка существенности включения независимых переменных	45
2.11	Тест на функциональную форму	46
2.12	Выбор наилучшей функции регрессии (тренда)	46

1 Принципы и методы эконометрического моделирования

1.1 Эконометрическое моделирование как метод моделирования социальноэкономических процессов

Модель (из словаря) – образец для изготовления чего-нибудь; тип конструкции; манекенщик или манекенщица; уменьшенное (или в натуральную величину) воспроизведение или макет чего-нибудь; схема какого-нибудь физического объекта или явления.

Требования к модели: «Модели должны быть настолько простыми, насколько это возможно, но не проще» (приписывают А.Эйнштейну).

Виды моделей:

- Физические;
- Аналоговые;
- Символические.

Математическая модель экономического объекта – это его гомоморфное отображение в виде совокупности уравнений, неравенств, логических отношений, графиков.

Экономическая (экономико-математическая) модель – математическая модель, описывающая механизм функционирования экономической или социально-экономической системы.

Классификация экономических моделей:

- Макро- и микроэкономические;
- Теоретические и прикладные;
- Статические и динамические;
- Детерминированные и стохастические.

Моделирование – формирование условного образа реальной системы, отражающего лишь некоторые существенные стороны процесса ее функционирования.

Принципы моделирования:

- Информационная достаточность;
- Параметризация отдельных подсистем;
- Агрегирование;
- Осуществимость цели исследования;
- Множественность моделей.

Этапы построения экономической модели:

1. Формулировка предмета и цели исследования;
2. Выделение элементов системы и их характеристик;
3. Дескриптивное описание связей между элементами или характеристиками;

4. Формирование математической модели;
5. Расчеты по модели и анализ решения.

Характеристики системы:

- Внешняя среда (эндогенная);
- Управляющие воздействия (эндогенная);
- Внутренние характеристики (эндогенная);
- Выходные характеристики (экзогенная).

Экзогенные переменные:

- Управляемые и неуправляемые;
- Внешние и внутренние;
- Определенные и неопределенные.

Неопределенность бывает стохастическая и нестохастическая.

Нестохастическая неопределенность:

- Поведенческая;
- Гносеологическая;
- Метрологическая;
- Целевая.

Адекватность модели – причины неадекватности:

- Ошибки в моделировании;
- Нестабильность неопределенных факторов;
- Недостоверность исходных данных;
- Нерепрезентативность выборки.

Оценка адекватности модели:

- Ретроспективный анализ;
- Логико-математический анализ;
- Экспертная оценка.

Оценка точности модели:

- $$d_l = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{n}$$

- $d_s = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}}$
- $\bar{A} = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|$
- $R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$

Эконометрическая модель – вероятностно-статистическая модель, описывающая механизм функционирования экономической или социально-экономической системы.

Паутинообразная модель (в такой форме еще не является эконометрической моделью):

$$\begin{cases} Q_{\text{предл},t} = f(p_{t-1}) \\ Q_{\text{спрос},t} = g(p_t) \\ \lim_{t \rightarrow \infty} f(p_{t-1}) = \lim_{t \rightarrow \infty} g(p_{t-1}) \end{cases}$$

Паутинообразная модель в форме эконометрической модели:

$$\begin{cases} Q_{\text{предл},t} = a_1 + b_1 p_{t-1} + e_{1t} \\ Q_{\text{спрос},t} = a_2 + b_2 p_t + e_{2t} \\ \lim_{t \rightarrow \infty} f(p_{t-1}) = \lim_{t \rightarrow \infty} g(p_{t-1}) \end{cases}$$

Объект эконометрического моделирования – социально-экономические явления и процессы.

Предмет эконометрического моделирования – статистические закономерности, присущие исследуемым явлениям и процессам.

1.2 Этапы эконометрического моделирования

1. Постановка задачи:

1.1. Цель исследования:

- Анализ объекта (процесса);
- Прогнозирование;
- Имитация развития объекта (процесса);
- Подготовка управленческих решений;

1.2. Показатели и их роль в модели

2. Анализ сущности изучаемого объекта;
3. Сбор и анализ данных;
4. Построение теоретической модели;
5. Статистический анализ модели и оценка ее параметров;
6. Верификация модели.

1.3 Сбор и анализ данных

- Сбор данных;
- Анализ качества данных;
- Отбор показателей в модель.

Сбор данных – Федеральная служба государственной статистики **www.gks.ru**.

Подсайты ФСГС:

- Единая межведомственная информационно-статистическая система (ЕМИСС);
- Показатели муниципальных образований.

Анализ качества данных:

- «Аномальные значения»;
- Однородность совокупности;
- Периодизация временных рядов;
- Полнота данных (цензурированные выборки);
- Вид переменных.

Данные с пропусками:

- Восстановление:
 - Заменить пропущенные данные на средние по уровню ряда (совокупности);
 - Построить обратную зависимость независимой переменной (x) от зависимой переменной (y), рассчитывая пропущенные значения;
 - Построить регрессию независимой переменной (x) от других экзогенных параметров, рассчитывая пропущенные значения;
- Исключение переменной, по которой отсутствует информация;
- Особая категория – «умышленные пропуски».

1.4 Принципы эконометрического моделирования

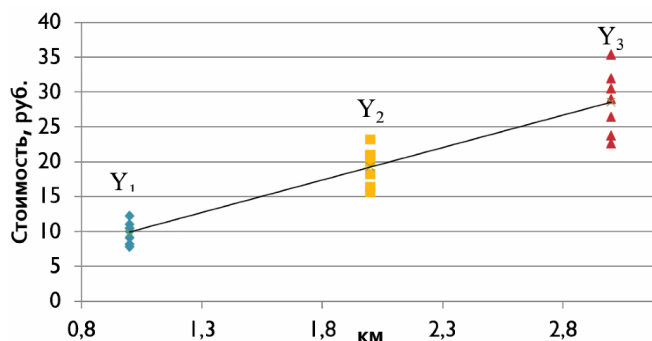
Принципы бывают априорные и апостериорные.

Априорные принципы:

- Теоретическая основа модели;
- Достаточное число наблюдений ($n \geq 7m$);
- Наблюдения являются выборочными данными;
- Количественная измеримость;
- Стохастические связи;

- Однородность связи;
- Зависимая переменная (всегда детерминированная часть и стохастическая часть): $y = \hat{y} + \varepsilon$
- Регрессионная модель (математическое ожидание СВ-н зависимой переменной лежит на линии регрессии):

$$MY_i = f(x_1, x_2, \dots, x_p) = \hat{y}_i$$



- Независимые переменные (стохастические, детерминированные).

Апостериорные принципы:

- Множественность моделей: неуниверсальность конкретной числовой модели;
- Оценки параметров как случайные величины;
- Критерии выбора модели в условиях теоретической неопределенности;
- Постулирование характеристик связи на основе выбранной модели

1.5 Методы эконометрического моделирования

Методы эконометрического моделирования:

- Метод наименьших квадратов;
- Обобщенный метод наименьших квадратов;
- Метод максимального правдоподобия;
- Методы построения моделей временных рядов;
- Методы анализа и решения систем эконометрических уравнений.

Другие методы математической и прикладной статистики:

- Статистическая проверка гипотез;
- Типологизация и кластеризация социально-экономических объектов: методы дискриминантного анализа, кластерного анализа.

2 Обзор математико-статистического инструментария эконометрического моделирования

2.1 Методы предварительного анализа данных

Методы предварительного анализа данных:

- Графический – расположение данных на графике;
- Логический (исторический) – предполагает понимание возможных изменений данных связанных с изменениями в социально-экономических системах;
- Аналитические методы:
 - Критерии «выбросов» (аномальности);
 - Кластеризация в случае пространственных данных, периодизация в случае временных рядов;
 - Объединение данных – смыкание рядов;
 - Устранение несопоставимости отдельных уровней временного ряда (сначала нужно понять, что какие то уровни временного ряда несопоставимы);
 - Выявление тенденции, периодических колебаний.

Проблемы выявления «аномальных» точек (аномальность понимается по разному в двух случаях):

- $MY_i = const$ – нет регрессии, такое может встречаться только во временных рядах (постоянство математического ожидания \Rightarrow стационарность ряда);
- $MY_i = f(x_1, x_2, \dots, x_p)$ – есть регрессия.

Первый случай – постоянство математического ожидания:

$$\begin{aligned}MY_i &= const \\DY_i &= const \\v &= \frac{\sigma_y}{\bar{y}} \cdot 100\%\end{aligned}$$

Где v – коэффициент вариации (должен быть небольшим). Не совсем удачный коэффициент: часто может быть высоким при однородности совокупности, зависит от единиц измерения.

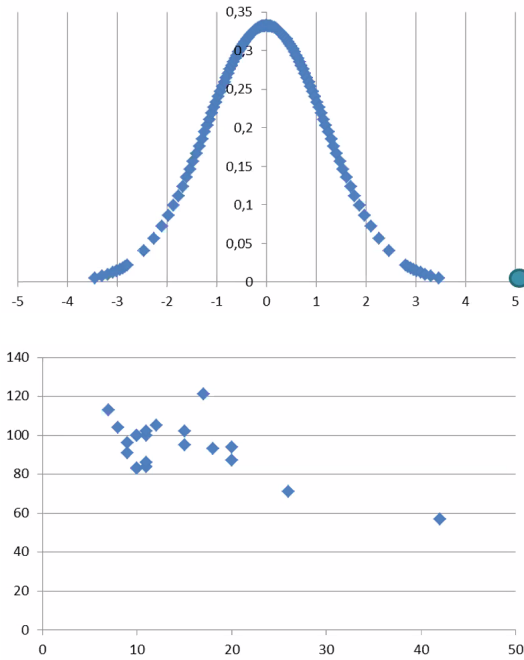
Вместо коэффициента вариации предлагается проверить нормальность распределения признака y по правилу трех сигм – если СВ признака y распределена нормально, то точка, являющаяся аномальной будет выходить за границы следующего доверительного интервала (так как у нормальной случайной величины в этом доверительном интервале лежит 99.72% значений случайной величины):

$$(\bar{y} - 3\sigma_y; \bar{y} + 3\sigma_y)$$

Второй случай – непостоянство математического ожидания:

$$\begin{aligned}MY_i &= f(x_1, x_2, \dots, x_p) \\DY_i &= const\end{aligned}$$

В таком случае возможны две ситуации: или точка является «влиятельной», или точка является «выбросом».



Синяя точка существенно отклонилась относительно среднего значения, в данном случае эта точка – *выброс* (далеко от математического ожидания).

МНК посчитает, что тут три «облака» точек. Точки, являющиеся вторым и третьим облаком, называют *влиятельными*.

Влиятельность:

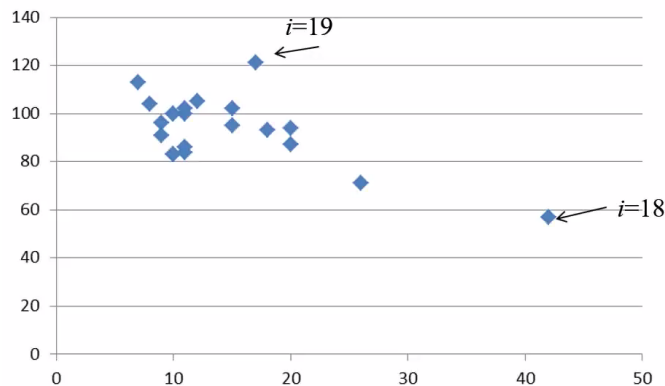
$$se(y_{\text{пр}}) = \sqrt{\frac{SS_{\text{ост}}}{n-2} \left(1 + \frac{1}{n} + \frac{(x_{\text{пр}} - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right)}$$

$$se(y_{\text{пр}}) = \sqrt{\frac{SS_{\text{ост}}}{n-m-1} 1 + X_{\text{пр}}^T \cdot (X^T \cdot X)^{-1} \cdot X_{\text{пр}}}$$

$$h_i = X_i^T \cdot (X_{(i)}^T \cdot X_{(i)})^{-1} \cdot X_i$$

Где h_i – характеристика влиятельности, а $X_{(i)}$ – матрица без исследуемого элемента i , X_i – вектор, который содержит компоненты точки i , которую проверяем на влиятельность: считаем точку влиятельной, если $h_i > \frac{2m}{n}$

Для нашего примера:



$$h_{18} = 1.87, h_{кр} = 0.1$$

$$h_{19} = 0.056, h_{кр} = 0.1$$

Таким образом, точка 18 является влиятельной, а 19, которая больше походила на выброс, не является влиятельной – просто выброс.

Выбросы:

$$B = (X^T \cdot X)^{-1} \cdot X^T \cdot Y$$

$$\hat{Y} = X \cdot B$$

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

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

Матрица H (размера $n \times n$) является основой для определения выбросов.

$$\hat{Y} = H \cdot Y$$

Можно показать, что элементы матрицы H являются частью формулы ковариации случайных остатков: если остатки не связаны $cov(\varepsilon_i, \varepsilon_j) = 0$, то $h_{ij} = 0$ (либо близок к нулю), а если $cov(\varepsilon_i, \varepsilon_i) \neq 0$, то $h_{ii} \neq 0$.

$$\hat{y}_i = y_i \cdot h_{ii} + \sum_{j \neq i} y_j h_{ij}$$

При нулевой ковариации случайных остатков (в случае независимости случайных остатков по следствию Гаусса-Маркова) $cov(\varepsilon_i, \varepsilon_j) = 0$ элементы $h_{ij} = 0$ (либо близки к нулю), поэтому основная «нагрузка» на отклонение \hat{y}_i от y_i лежит на элементе h_{ii} , который называют *элементом влияния*. На его основе рассчитываются меры выбросов.

Стандартизованные (студентизованные) остатки:

$$\hat{\varepsilon}(i) = \frac{\hat{\varepsilon}_i}{\sqrt{1 - h_{ii}}} / \sqrt{\frac{SS_{\varepsilon} - \frac{\hat{\varepsilon}_i^2}{1 - h_{ii}}}{n - m - 2}}$$

$\hat{\varepsilon}_i$ – остаток, рассчитанный по уравнению регрессии, параметры которого рассчитаны по всем наблюдениям (получается, выброс сюда входит), h_{ii} – элемент диагональный из матрицы H , которая тоже строится по всем наблюдениям, включая предполагаемый выброс, SS_{ε} по той же регрессии, включая проверяемое наблюдение.

Элемент считается выбросом, если $|\hat{\varepsilon}(i)| > 2$

Для нашего примера:

$$\hat{\varepsilon}(19) = 3.42$$

$$\hat{\varepsilon}(кр) = 2$$

$$\hat{\varepsilon}(18) = -0.29$$

Таким образом, предположения подтвердились: 19 точка оказалась выбросом, а 18 не оказалась выбросом (она была признана влиятельной ранее).

Кластеризация (периодизация):

- Модели с фиктивными переменными – ANOVA (модели дисперсионного анализа)

$$y = a + c \cdot z + \varepsilon$$

- Модели ANCOVA (модели ковариационного анализа)

$$y = a + b \cdot x + c_{11} \cdot z + c_{12} \cdot z \cdot x + \varepsilon$$

В случае незначимости коэффициентов при фиктивных переменных мы говорим об однородности совокупности (абсолютная однородность в случае незначимости всех коэффициентов перед фиктивными переменными).

Тест Чоу – оценка неоднородности совокупности:

$$y = a + b_1 \cdot x_1 + \dots + b_p \cdot x_p + \varepsilon$$

- Регрессия общая: $n = n_1 + n_2$ наблюдений, $SS_{\text{ост}}^{(0)}$
- Регрессия (1): n_1 наблюдений, $SS_{\text{ост}}^{(1)}$
- Регрессия (2): n_2 наблюдений, $SS_{\text{ост}}^{(2)}$

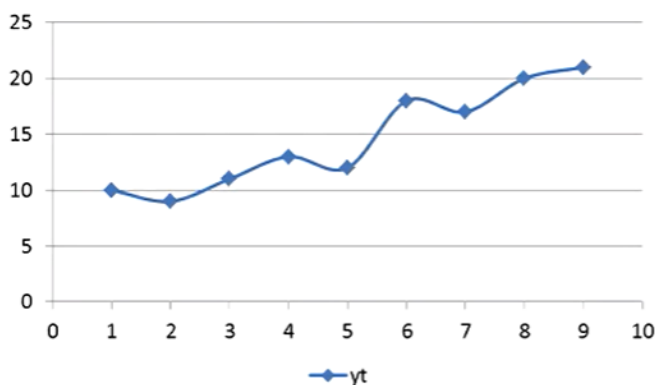
Тестирование с помощью критерия Фишера:

$$F = \frac{SS_{\text{ост}}^{(0)} - (SS_{\text{ост}}^{(1)} + SS_{\text{ост}}^{(2)})}{SS_{\text{ост}}^{(1)} + SS_{\text{ост}}^{(2)}} \cdot \frac{n - m_1 - m_2 - 2}{m_1 + m_2 + 1 - m}$$

$$F_{\text{табл}}(\alpha; df_1 = m_1 + m_2 + 1 - m; df_2 = n - m_1 - m_2 - 2)$$

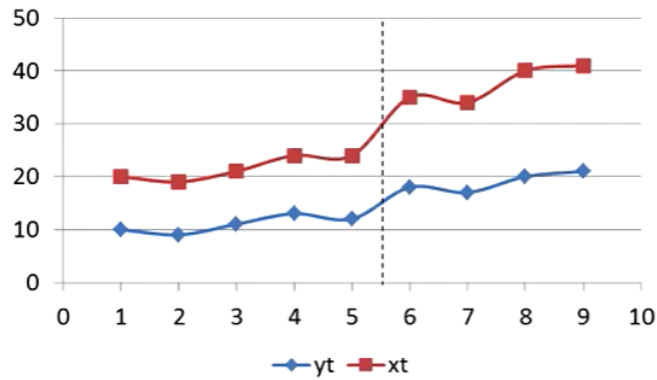
$F \geq F_{\text{табл}}$ означает неоднородность совокупности (то есть описывается разными тенденциями), в противном случае – наоборот.

Метод параллельной периодизации: рассмотрим следующий ряд, о котором мы предполагаем неоднородность процесса – то есть, что нужна периодизация временного ряда.



Метод параллельной периодизации состоит в том, что мы подбираем вторую переменную (y – первая), которая, с одной стороны, тесно связана с y , а, с другой стороны, имеет более четкий период. Например, переменная x для примера:

Объединение данных – проблема смыкания рядов динамики. Предположим, что у нас есть следующий временной ряд, в котором с $t = 1$ по $t = 5$ показатель рассчитывался по одной методологии – y_t^1 , а с $t = 5$ по $t = 9$ рассчитывался по другой методологии – y_t^2 :



	1	2	3	4	5	6	7	8	9
y_t^1	10	9	11	13	12				
y_t^2					15	22	20	24	25

В момент времени $t = 5$ показатель был рассчитан по обеим методологиям – в таком случае можем использовать коэффициент пересчета $K = \frac{y_5^2}{y_5^1}$. В данном случае $K = \frac{15}{12} = 1.25$, по этому коэффициенту пересчитываются старые уровни ряда $y_t^2 = K \cdot y_t^1$:

	1	2	3	4	5	6	7	8	9
y_t^1	10	9	11	13	12				
y_t^2	12.5	11.25	13.75	16.25	15	22	20	24	25

Устранение несопоставимости отдельных уровней временного ряда:

- Пересчет уровней ряда по единой методике;
- В единых территориальных и организационных рамках;
- В единых ценах (дефлянтирование).

Индексы – отношение значений показателя за два периода времени (отчетный уровень к базисному).

Индивидуальные индексы:

$$i = \frac{p_1}{p_0}$$

Сводные индексы:

$$I = \frac{\sum T_1}{\sum T_0}$$

Агрегатные индексы:

$$I_p = \frac{\sum p_1 q_1}{\sum p_0 q_0}$$

Индексы цен (так как нас интересуют дефлянтированные) бывают:

- Ласпейреса:

$$I_p = \frac{\sum p_1 q_0}{\sum p_0 q_0}$$

- Пааше:

$$I_p = \frac{\sum p_1 q_1}{\sum p_0 q_1}$$

- Фишера:

$$I_p = \sqrt{\frac{\sum p_1 q_1}{\sum p_0 q_1} \cdot \frac{\sum p_1 q_0}{\sum p_0 q_0}}$$

Цепные индексы:

$$i_{1/0} = \frac{p_1}{p_0}, i_{2/1} = \frac{p_2}{p_1}, i_{3/2} = \frac{p_3}{p_2}, i_{4/3} = \frac{p_4}{p_3}$$

Базисный индекс:

$$i_{4/0} = \frac{p_1}{p_0} \cdot \frac{p_2}{p_1} \cdot \frac{p_3}{p_2} \cdot \frac{p_4}{p_3}$$

Единицы измерения индексов:

- Доли – для расчетов;
- Проценты – официальная отчетность и информация.

Индекс потребительских цен (ИПЦ) рассчитывается на основе выборочного наблюдения цен, выборка эта по нескольким признакам:

- По населенным пунктам;
- По торговым предприятиям;
- По товарам и услугам.

Официальные данные, как правило содержат два вида индексов:

- ИПЦ i/j :

$$I_{i(t)/(i-1)(t)} \text{ или } I_{1(t)/12(t-1)}, i = 1, \dots, 12$$

- ИПЦ декабрь к декабрю предыдущего года:

$$I_{12(t)/12(t-1)}$$

Обозначение $i(t)$: i – номер месяца, (t) – номер года.

Необходимы годовые ИПЦ и квартальные ИПЦ – для их нахождения необходимо преобразовывать месячные цепные индексы. ИПЦ, исходные данные:

	1991	1992	1993
к концу предыдущего месяца			
январь	1.062	3.453	1.258
февраль	1.048	1.38	1.247
март	1.063	1.299	1.201
апрель	1.635	1.217	1.187
май	1.93	1.119	1.181
июнь	1.012	1.191	1.199
июль	1.006	1.106	1.2239
август	1.005	1.086	1.26
сентябрь	1.011	1.115	1.23
октябрь	1.035	1.229	1.195
ноябрь	1.089	1.261	1.1639
декабрь	1.121	1.252	1.125
к декабрю предыдущего года:			
декабрь	2.604	26.0884	9.399

Методика пересчета цепных месячных ИПЦ в цепные годовые:

1. ИПЦ текущего месяца к декабрю предыдущего года:

$$I_{i(t)/12(t-1)} = \left(\prod_{j=2}^i I_{j(t)/(j-1)(t)} \right) I_{1(t)/12(t-1)}$$

Для 1991 года:

$$I_{\text{янв } 91/\text{дек } 90} = 1.062$$

$$I_{\text{фев } 91/\text{дек } 90} = 1.062 \cdot 1.048 = 1.112976$$

$$I_{\text{март } 91/\text{дек } 90} = 1.062 \cdot 1.048 \cdot 1.063 = 1.183093$$

...

$$I_{\text{дек } 91/\text{дек } 90} = 1.062 \cdot \dots \cdot 1.121 = 2.604016$$

В результате получаем:

	1991	1992	1993
к декабрю предыдущего года			
январь	1.062	3.453	1.258
февраль	1.112976	4.76514	1.568726
март	1.183093	6.189917	1.88404
апрель	1.934358	7.533129	2.236355
май	1.992389	8.429571	2.641136
июнь	2.016297	10.03962	3.166722
июль	2.028395	11.10382	3.875751
август	2.038537	12.05876	4.883446
сентябрь	2.060961	13.4455	6.006638
октябрь	2.133095	16.52452	7.177933
ноябрь	2.32294	20.83742	8.354396
декабрь	2.604016	26.08845	9.398696

2. ИПЦ текущего месяца к декабрю года, предшествующего предыдущему:

$$I_{i(t)/12(t-2)} = I_{i(t)/12(t-1)} \cdot I_{12(t-1)/12(t-2)}$$

Например, для 1991 – нет данных (декабрь 1989), для 1992 – декабрь 1990, для 1993 – декабрь 1991. 1992 к декабрю 1990:

$$I_{\text{янв } 92/\text{дек } 90} = I_{\text{янв } 92/\text{дек } 91} \cdot I_{\text{дек } 91/\text{дек } 90}$$

$$I_{\text{фев } 92/\text{дек } 90} = I_{\text{фев } 92/\text{дек } 91} \cdot I_{\text{дек } 91/\text{дек } 90}$$

...

Для примера:

$$\text{январь: } 3.453 \cdot 2.604016 = 8.99166$$

$$\text{февраль: } 4.76514 \cdot 2.604016 = 12.4085$$

$$\text{март: } 6.189917 \cdot 2.604016 = 16.11864$$

...

В результате получаем:

	1991	1992	1993
к декабрю предыдущего года			
январь		8.991666	32.81928
февраль		12.4085	40.92564
март		16.11864	49.15169
апрель		19.61639	58.34306
май		21.95074	68.90315
июнь		26.14333	82.61488
июль		28.914333	101.1123
август		31.40117	127.4016
сентябрь		35.0123	156.7039
октябрь		43.03012	187.2612
ноябрь		54.26098	217.9533
декабрь		67.93475	245.1974

3. ИПЦ текущего года к предыдущему:

$$I_{t/(t-1)} = \frac{\sum_{i=1}^{12} I_{i(t)/12(t-2)}}{\sum_{i=1}^{12} I_{i(t-1)/12(t-2)}}$$

Для нашего примера:

$$I_{1992/1991} = \frac{I_{\text{январь } 92/\text{дек } 90} + I_{\text{фев } 92/\text{дек } 90} + \dots + I_{\text{дек } 92/\text{дек } 90}}{I_{\text{январь } 91/\text{дек } 90} + I_{\text{фев } 91/\text{дек } 90} + \dots + I_{\text{дек } 91/\text{дек } 90}} =$$

$$= \frac{8.991666 + 12.4085 + \dots + 67.93475}{1.062 + 1.112976 + \dots + 2.604016} \approx 16.26$$

4. Базисные ИПЦ

Первый год = «база», $I_{1/1} = 1.00$, а следующие годы получаются следующим образом:

$$I_{t/(t-k)} = \prod_{j=t-k+1}^t I_{j/(j-1)}$$

Для нашего примера:

	1991	1992	1993	1994
ИПЦ цепн	1.00	16.26494	9.741572	4.076914
ИПЦ баз	1.00	16.26494	158.4461	645.9709

Методика пересчета цепных месячных ИПЦ в цепные квартальные:

1. То же самое;
2. То же самое;
3. ИПЦ текущего квартала к предыдущему:

$$I_{t/(t-1)} = \frac{\sum_{i=4}^6 I_{i(t)/12(t-2)}}{\sum_{i=1}^3 I_{i(t)/12(t-2)}}$$

Исключение составляет квартальный индекс первого квартала по сравнению с четвертым кварталом предыдущего года:

$$I_{t/(t-1)} = \frac{\sum_{i=1}^3 I_{i(t)/12(t-2)}}{\sum_{i=10}^{12} I_{i(t-1)/12(t-2)}}$$

Квартал к предыдущему кварталу для нашего примера:

$$\begin{aligned} I_{\text{вт.кв. 1992/перв.кв.1992}} &= \frac{I_{\text{апр 92/дек 90}} + I_{\text{май 92/дек 90}} + I_{\text{июнь 92/дек 90}}}{I_{\text{янв 91/дек 90}} + I_{\text{фев 91/дек 90}} + I_{\text{март 91/дек 90}}} = \\ &= \frac{19.61639 + 21.95074 + 26.14333}{8.991666 + 12.4085 + 16.11864} \approx 1.805 \end{aligned}$$

Выявление тенденции:

- Графический метод;
- Тест «восходящих» и «нисходящих» серий;
- Проверка постоянства средних и дисперсий;
- Анализ автокорреляционной функции;
- Оценка значимости уравнения тренда.

2.2 Отбор факторов в модель регрессии

Требования к факторам:

- Влияние независимой x переменной на зависимую переменную y :

$$|r_{yx_i}| \rightarrow 1$$

- Отсутствие интеркорреляции (мультиколлинеарность):

$$\Delta_{r_{xx}} = \begin{vmatrix} 1 & r_{x_1x_2} & \cdots & r_{x_1x_p} \\ r_{x_1x_2} & 1 & \cdots & r_{x_2x_p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{x_1x_p} & r_{x_2x_p} & \cdots & 1 \end{vmatrix} \rightarrow 1$$

- Выполнение условия того, что факторы должны быть более тесно связаны с результатом, чем друг с другом:

$$\begin{cases} |r_{yx_i}| > |r_{x_ix_j}| \\ |r_{yx_j}| > |r_{x_ix_j}| \end{cases}$$

Выявление особенностей факторов:

- Наличие случайной составляющей;
- Зависимость от других показателей;
- Ложная корреляция с результатом и факторами.

Отбор факторов в уравнение множественной регрессии:

- Метод исключения;
- Метод включения;
- Шаговый регрессионный анализ.

Метод исключения:

1. Строится модель со всеми факторами (бывают ограничения на включение в модель всех факторов: например, невыполнение $n > 7m$ – должен быть достаточный объем наблюдений);
2. Оценка значимости включения переменных с помощью F_x (частного F -критерия, который показывает значимость включения дополнительной переменной, который основывается на значимости сокращения остаточной суммы квадратов);
3. Исключение переменной x_s при $F_{x_s} < F_{\text{крит}}$ и $F_{x_s} = \min F_x$
4. Пересчет модели;
5. Переход к шагу 2.

Шаговый регрессионный анализ:

1. Включение первой переменной при условии, что $|R_{yx_1}| = \max_j (|R_{yx_j}|)$ и $F_{x_1} > F_{\text{крит}}$;
2. Включение последующих переменных – максимальный $F_{x_j(x_1 \dots x_{j-1})} > F_{\text{крит}}$;
3. Пересчет F_x для новой модели;
4. Если $F_{x_s} < F_{\text{крит}}$ и $F_{x_s} = \min F_x$, то исключение переменной x_s и переход к пункту 3, иначе переход к пункту 2.

2.3 Классическая нормальная линейная модель

Анализ модели:

- Определить виды переменных, входящих в модель: в простейшей модели все переменные количественные;
- Функциональная форма модели;
- Гипотезы о характере случайных остатков;
- Идентифицируемость системы эконометрических уравнений.

Исходя из анализа модели происходит выбор метода оценки параметров модели.

Предпосылки построения классической нормальной линейной модели (КНЛМ):

1. $MY_i = f(x_1, \dots, x_p) = \hat{y}_i$, где $f(x_1, \dots, x_p) = b_0 + b_1 \cdot x_1 + \dots + b_p \cdot x_p + \varepsilon$
2. $\sigma_{Y_i}^2 = \sigma_{Y_j}^2 = \sigma_Y^2 = \text{const} \quad \forall i, j$
3. Y_i имеет нормальное распределение $N(\hat{y}_i, \sigma_{Y_i}^2) \quad \forall i$
4. $r_{Y_i Y_j} = 0 \quad \forall i, j$

5. x_i – неслучайная $\forall i$

Требования к случайным остаткам КНЛМ (условия Гаусса-Маркова):

1. $M_{\varepsilon_i} = M(Y_i - MY_i) = 0$
2. $\sigma_{\varepsilon_i}^2 = \sigma_{\varepsilon_j}^2 = \sigma_{\varepsilon}^2 = const \quad \forall i, j$
3. $\varepsilon_i \in N(0, \sigma_{\varepsilon}^2) \quad \forall i$
4. $r_{\varepsilon_i \varepsilon_j} = 0 \quad \forall i, j$
5. $r_{\varepsilon_i t} = 0 \quad \forall i$

Если данные требования к случайным остаткам выполняются, то оценки МНК являются эффективными в классе несмещенных оценок.

Свойства МНК-оценок при соблюдении предпосылок КНЛМ:

- Несмещенность:

$$M\hat{b}_j = b_j$$

- Эффективность:

$$M(\hat{b}_j - b_j)^2 = \sigma_{\hat{b}_j}^2 = \min \sigma_{b_j}^2$$

- Состоятельность:

$$\hat{b}_j \xrightarrow{n \rightarrow \infty} b_j$$

Проверка качества модели:

- Проверка случайных остатков на постоянство дисперсии;
- Проверка значимости параметров;
- Проверка значимости функции.

2.4 Анализ качества случайных остатков

$M_{\varepsilon_i} = M(Y_i - MY_i) = 0$ – выполняется по МНК

$$\sigma_{\varepsilon_i}^2 = \sigma_{\varepsilon_j}^2 = \sigma_{\varepsilon}^2 = const \quad \forall i, j$$

Гомоскедастичность тесты:

- Гольдфельда-Квандта
- Уайта: $\varepsilon_i^2 = a + b_{11} \cdot x_1 + b_{12} \cdot x_1^2 + \dots + b_{p1} \cdot x_p + b_{p2} \cdot x_p^2 + [c_{12} \cdot x_1 \cdot x_2 + \dots + c_{p-1,p} \cdot x_{p-1} \cdot x_p] + \delta$
- Парка: $\ln \varepsilon^2 = a + b \cdot \ln x_j + \delta$
- Глейзера: $|\varepsilon| = a + b \cdot x_j^k + \delta \quad k = -2, -1, -0.5, 0.5, 1, 2$
- Ранговой корреляции Спирмена: $\rho = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2-1)}$

$$r_{\varepsilon_i \varepsilon_j} = 0 \quad \forall i, j$$

Проверка независимости случайных остатков:

- Коэффициент автокорреляции
- Критерий Дарбина-Уотсона:

$$dw = \frac{\sum_{i=2}^n (\varepsilon_t - \varepsilon_{t-1})^2}{\sum_{i=1}^n \varepsilon_t}$$

$$\varepsilon_i \in N(0, \sigma_\varepsilon^2) \forall i$$

Проверка нормальности распределения случайных остатков:

- Коэффициенты асимметрии и эксцесса:

$$|As| < 1.5 \sqrt{\frac{6(n-2)}{(n+1)(n+3)}}$$

$$\left| Ex + \frac{6}{n+1} \right| < 1.5 \sqrt{\frac{24n(n-2)(n-3)}{(n+1)^2(n+3)(n+5)}}$$

- RS-критерий
- критерий Жака-Бера

RS-критерий:

$$R = \varepsilon_{max} - \varepsilon_{min}$$

$$S = \sqrt{\frac{\sum \varepsilon_t^2}{n-1}}$$

$$RS = \frac{R}{S}$$

Табличные значения RS-критерия:

n	RS min	RS max
10	2,67	3,69
15	2,96	4,14
20	3,18	4,49
25	3,34	4,71
30	3,47	4,89

Критерий Жака-Бера:

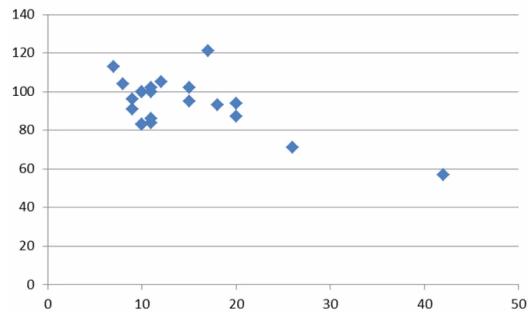
$$H_0 : As = 0; Ex = 0$$

$$H_1 : As \neq 0; Ex \neq 0$$

$$JB = \frac{n-m}{6} \cdot (As^2 + \frac{Ex^2}{4})$$

$$\chi^2(\alpha; 2)$$

$$JB < \chi^2$$



Пример:

Регрессия по всей совокупности:

$$y = 109.9 - 1.13 \cdot x + \hat{\varepsilon}$$

Нормальность остатков ($n = 20$):

$$As = 0.73; \quad Ex = 1.16; \quad \left| Ex + \frac{6}{n+1} \right| = 1.43$$

$$se(As) = 0.70; \quad se(Ex) = 1.13$$

$$RS = 4.27$$

$$RS_{max} = 4.49$$

$$JB = 2.89$$

$$\chi^2(0.05; 2) = 5.99$$

Первый критерий не показал нормальность, а 2 и 3 показали нормальность – в таком случае, ориентируемся на RS-критерий и критерий Жака-Бера и говорим о нормальности распределения случайных остатков.

$$r_{\varepsilon_{it}} = 0 \quad \forall i \quad - \text{гипотеза об отсутствии закономерной составляющей}$$

Делаем тесты на стационарность ряда (если временной ряд).

2.5 Оценка значимости параметров регрессии (тренда)

Оценки параметров – случайные величины:

$$y = b_0 + b_1 \cdot x_1 + \dots + b_j \cdot x_j + \dots + b_p \cdot x_p + \varepsilon$$

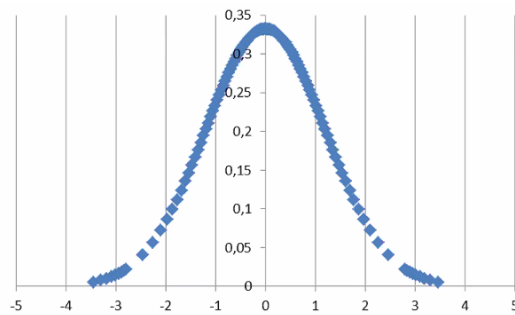
По выборочным данным оцениваем параметры модели:

$$\hat{b}_j \rightarrow b_j$$

Так как оценка параметра – случайная величина, возникает вопрос: какой у нее закон распределения? Ответ: оценка параметра имеет нормальный закон распределения, так как по ЦПТ сумма независимых одинаково распределенных случайных величин имеет распределение, близкое к нормальному.

$$\hat{b}_j \in N(b_j, \sigma_{b_j}^2)$$

$$b_j = ?$$



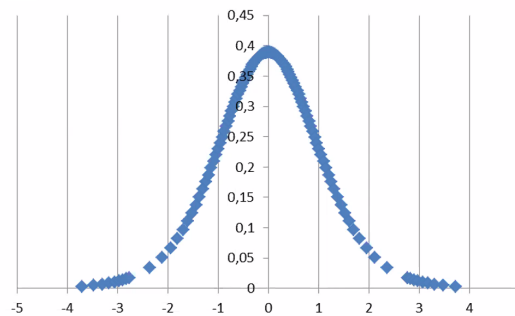
$$\sigma_{b_j}^2 = ?$$

Введем $b_j = B_j$ (B_j – число, которое мы предполагаем истинным значением параметра), а $\sigma_{b_j} \rightarrow se(\hat{b}_j)$.

Тогда:

$$\frac{\hat{b}_j - B_j}{se(\hat{b}_j)} \in t(n - m - 1)$$

Распределение Стьюдента (для $df = 10$):



Например:

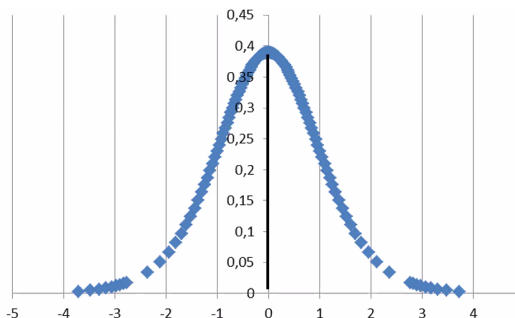
$$y = 2 + 3 \cdot x + \hat{\varepsilon}, \quad n = 12$$

$$\hat{b}_1 = 3, \quad se(\hat{b}_1) = 0.5$$

Гипотеза H_0 :

$$b_1 = B_1 = 3.0$$

$$t_{b_1} = \frac{3 - 3}{0.5} = 0$$



$$P(\hat{b}_1 = b_1) = ?$$

$$P(\hat{b}_1 = b_1) = P\left(\frac{\hat{b}_1 - B_1}{se(\hat{b}_1)} = \frac{b_1 - B_1}{se(\hat{b}_1)}\right) = P\left(t_{b_1} = \frac{b_1 - B_1}{se(\hat{b}_1)}\right) = P(t_{b_1} = 0) = 0$$

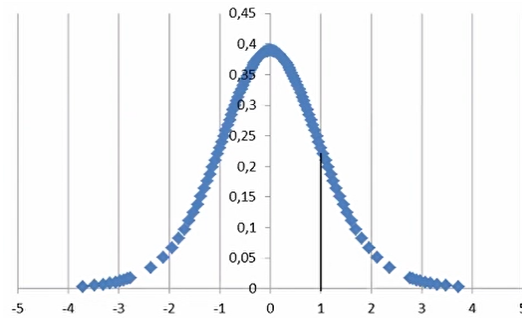
$$P(\hat{b}_1 < b_1) = P\left(\frac{\hat{b}_1 - B_1}{se(\hat{b}_1)} < \frac{b_1 - B_1}{se(\hat{b}_1)}\right) = P(t_{b_1} < 0) = 0.5$$

Гипотеза H_0 :

$$b_1 = B_1 = 2.5$$

$$t_{b_1} = \frac{3 - 2.5}{0.5} = 1$$

$$t_{b_1} = 1, df = 10$$



Можно определить:

$$P(t_{b_1} < 1) = F(t_{b_1}) = 0.8296$$

То есть:

$$\begin{aligned} P(t_{b_1} < 1) &= P\left(\frac{\hat{b}_1 - b_1}{se(\hat{b}_1)} < \frac{\hat{b}_1 - B_1}{se(\hat{b}_1)}\right) = P(b_1 > B_1) = 0.8296 \Rightarrow \\ &\Rightarrow P(b_1 \leq B_1) = 1 - 0.8296 = 0.1704 \end{aligned}$$

Продолжение примера:

$$y = 2 + 2 \cdot x + \hat{\varepsilon}, \quad n = 12$$

$$\hat{b}_1 = 2, se(\hat{b}_1) = 0.5$$

Гипотеза H_0 :

$$b_1 = B_1 = 2.5$$

$$t_{b_1} = \frac{2 - 2.5}{0.5} = -1$$

Можно определить:

$$P(t_{b_1} < -1) = F(t_{b_1}) = 0.1704$$

То есть:

$$\begin{aligned} P(t_{b_1} < -1) &= P\left(\frac{\hat{b}_1 - b_1}{se(\hat{b}_1)} < \frac{\hat{b}_1 - B_1}{se(\hat{b}_1)}\right) = P(b_1 > B_1) = 0.1704 \Rightarrow \\ &\Rightarrow P(b_1 \leq B_1) = 1 - 0.1704 = 0.8296 \end{aligned}$$

Когда говорим об отклонении расчетного значения \hat{b}_j от гипотетического b_j нас не волнует, в какую сторону произошло отклонение, поэтому, так как распределение симметрично, а $P(t_{b_j} = 0)$ – идеальный вариант, то нас интересует модуль отклонения расчетного значения $|\hat{b}_j|$ от гипотетического b_j .

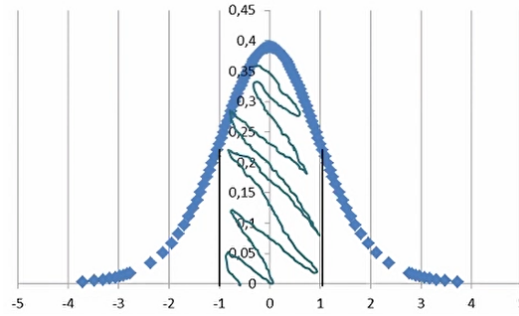
Обобщим два предыдущих случая:

$$\begin{aligned} P\left(\frac{\hat{b}_j - b_j}{se(\hat{b}_j)} < \left|\frac{\hat{b}_j - B_j}{se(\hat{b}_j)}\right|\right) &= P\left(\frac{\hat{b}_j - b_j}{se(\hat{b}_j)} < |t_{b_j}|\right) = P\left(-t_{b_j} < \frac{\hat{b}_j - b_j}{se(\hat{b}_j)} < t_{b_j}\right) = \\ &= F(t_{b_j}) - F(-t_{b_j}) = P(\hat{b}_j - t_{b_j} \cdot se(\hat{b}_j) < b_j < \hat{b}_j + t_{b_j} \cdot se(\hat{b}_j)) \end{aligned}$$

Для нашего примера – для $|t_{b_j}| = 1$:

$$P(\hat{b}_j - t_{b_j} \cdot se(\hat{b}_j) < b_j < \hat{b}_j + t_{b_j} \cdot se(\hat{b}_j)) = 0.8296 - 0.1704 = 0.6591$$

$$|t_{b_j}| = 1 \quad P(-1 < t_{b_j} < 1) = 0.6591$$



Пусть задано значение B_j , тогда $t_{b_j} = f(\hat{b}_j)$.

Пусть $|t_{b_j}| = 1$:

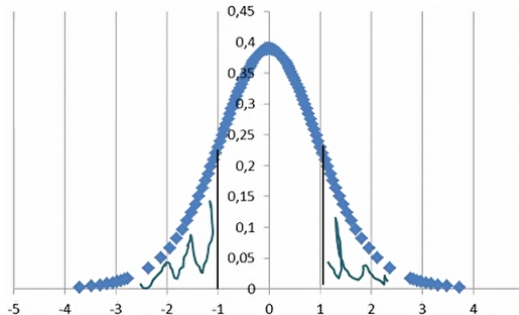
$$P(-1 < t_{b_j} < 1) = 0.6591$$

$$P(t_{b_j} > 1) = \frac{1 - 0.6591}{2} = 0.17045$$

$$P(t_{b_j} < -1) = \frac{1 - 0.6591}{2} = 0.17045$$

$$0.17045 + 0.17045 = 0.3409 \equiv \alpha_{b_j}$$

$$P(|t_{b_j}| > 1) = 0.3409$$



Идея: если верна нулевая гипотеза и коэффициент, оцениваемый, по нашим данным, варьируется вокруг истинного значения, то оставлять такие большие крылья – неправильно, для этого можем задать большую вероятность попадания в интервал, расширив его. Зададим:

$$P(\hat{b}_j - t_{b_j} \cdot se(\hat{b}_j) < b_j < \hat{b}_j + t_{b_j} \cdot se(\hat{b}_j)) = 0.95$$

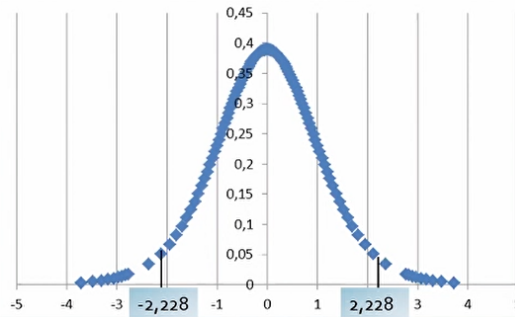
Или кратко:

$$P(\hat{b}_{j,\text{мин}} < b_j < \hat{b}_{j,\text{макс}}) = 0.95$$

Будем считать, что если $B_j \in (\hat{b}_{j,\text{мин}}; \hat{b}_{j,\text{макс}})$, то $b_j \approx B_j$, а $(\hat{b}_{j,\text{мин}}; \hat{b}_{j,\text{макс}})$ называется **областью принятия гипотезы**.

Для нашего примера:

$$P(-2.228 < t_{b_j} < 2.228) = 0.95$$



Вероятности хвостов: $\alpha = 1 - 0.95 = 0.025 + 0.025 = 0.05$. Гипотеза $b_j = B_j$ отвергается с вероятностью $\alpha = 0.05$, параметр α называется **уровнем значимости**.

Незначимость параметра b_j : $b_j = 0 \in (\hat{b}_{j,\text{мин}}; \hat{b}_{j,\text{макс}})$

Если:

$$b_j = 0$$

$$\sigma_{\hat{b}_j} \rightarrow se(\hat{b}_j)$$

Тогда:

$$t_{b_j} = \frac{\hat{b}_j - 0}{se(\hat{b}_j)} = \frac{\hat{b}_j}{se(\hat{b}_j)} \in t(n - m - 1)$$

Гипотеза о статистической незначимости параметра b_j :

$$H_0 : b_j = 0$$

$$H_1 : b_j \neq 0$$

α - уровень значимости, который задан

Ошибки первого и второго рода:

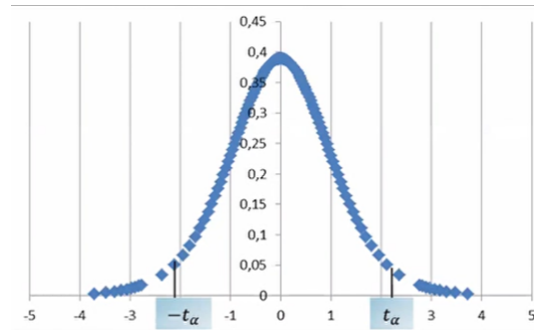
Гипотеза H_0	верна	не верна
принята	нет ошибки	ошибка второго рода
отвергнута	ошибка первого рода	нет ошибки

α – вероятность ошибки первого рода, β – вероятность ошибки второго рода.

Критерий Стьюдента для оценки значимости параметра b :

$$t_{\text{табл}} = t_{\alpha} = f(df; \alpha) > 0$$

$$P(-t_{\alpha} < t < t_{\alpha}) = 1 - \alpha$$



Отвергаем нулевую гипотезу о незначимости:

$$|t_b| \geq t_{\text{табл}} (= t_{\alpha})$$

$$|t_b| \rightarrow \alpha_b$$

$$\alpha_b \leq \alpha$$

Пусть:

$$H_0 : b_j = B_j$$

Отвергаем нулевую гипотезу:

$$\left| \frac{\hat{b}_j - B_j}{se(\hat{b}_j)} \right| \geq t_{\alpha}$$

Каковы возможные значения b_j – каков доверительный интервал?

$$\left| \frac{\hat{b}_j - B_j}{se(\hat{b}_j)} \right| \geq t_{\alpha}$$

$$-t_{\alpha} < \frac{\hat{b}_j - B_j}{se(\hat{b}_j)} < t_{\alpha}$$

$$\hat{b}_j - t_{\alpha} \cdot se(\hat{b}_j) < b_j < \hat{b}_j + t_{\alpha} \cdot se(\hat{b}_j)$$

$$b_{j,\text{мин}} < b_j < b_{j,\text{макс}}$$

$$P(b_{j,\text{мин}} < b_j < b_{j,\text{макс}}) = 1 - \alpha$$

Пример: x – сумма счета в ресторане, евро; y – чаевые, евро. Получено уравнение регрессии:

$$y = 0.492 + 0.126x + \hat{\varepsilon}, \quad n = 20$$

$$(se) \quad (0.698) \quad (0.020)$$

$$\alpha = 0.05; df = 20 - 2 = 18 \Rightarrow t_\alpha = 2.1$$

Выдвигаем гипотезы (нулевые и альтернативные) по двум параметрам:

$$H_0 : b_0 = 0$$

$$H_1 : b_0 \neq 0$$

$$H_0 : b_1 = 0$$

$$H_1 : b_1 \neq 0$$

$$t_{b_0} = \frac{0.492 - 0}{0.698} = 0.7$$

$$t_{b_1} = \frac{0.126 - 0}{0.020} = 6.3$$

Свободный член незначим, коэффициент регрессии значим.

$$b_{0,\text{мин}} = 0.492 - 2.1 \cdot 0.698 \approx -0.97$$

$$b_{0,\text{макс}} = 0.492 + 2.1 \cdot 0.698 \approx 1.96$$

$$-0.97 \leq b_0 \leq 1.96$$

$$b_{1,\text{мин}} = 0.126 - 2.1 \cdot 0.020 \approx 0.084$$

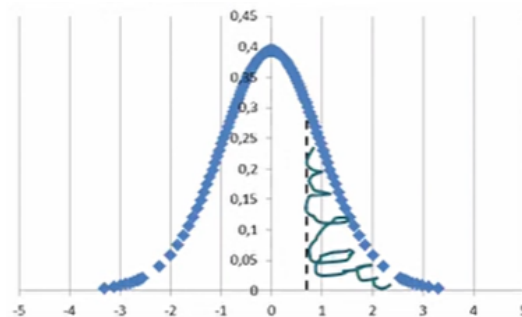
$$b_{1,\text{макс}} = 0.126 + 2.1 \cdot 0.020 \approx 0.168$$

$$0.084 \leq b_1 \leq 0.168$$

С вероятностью 0.95 истинное значение b_0 (b_1) находится в соответствующих границах.

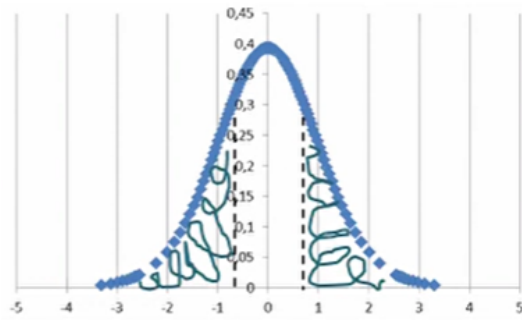
Интерпретация коэффициента регрессии: с изменением суммы счета в ресторане на один евро, чаевые в среднем составят 0.126 евро, с вероятностью 0.95 они будут находиться в границах от 0.084 евро до 0.168 евро.

$$P(t \geq t_{b_0}) = P(t \geq 0.7) = 1 - P(t < 0.7) = 1 - F(0.7) \approx 1 - 0.7536 = 0.2464$$



$$P(|t| \geq t_{b_0}) = P(t \geq 0.7) + P(t \leq -0.7) = 1 - F(0.7) + F(-0.7) \approx 1 - 0.7536 + 0.2464 = 0.4928$$

$$0.4928 = \text{p-value}(\alpha_{b_0})$$



Если $\alpha_{b_0} > \alpha$ – незначимость свободного члена. В общем случае $\alpha_{b_0} > \alpha$ – принимается $H_0 : b_0 = 0$ либо $H_0 : b_0 = B_0$

Аналогично для коэффициента регрессии: $\alpha_{b_1} < \alpha$ – значимость коэффициента регрессии. В примере $\alpha_{b_1} = 7 \cdot 10^{-6} < 0.05$ – отвергается нулевая гипотеза о незначимости.

$$b_1 = ?$$

$$0.084 \leq b_1 \leq 0.168$$

Предположим:

$$H_0 : b_1 = 0.13$$

$$H_1 : b_1 \neq 0.13$$

$$t_{b_1} = \frac{0.126 - 0.13}{0.020} = -0.2$$

Нет оснований отвергнуть нулевую гипотезу.

Вывод:

$$H_0 : b_j = B_j$$

$$H_1 : b_j \neq B_j$$

Частный случай – гипотеза о незначимости, в случае $B_j = 0$:

$$H_0 : b_j = 0 = B_j$$

$$H_1 : b_j \neq 0 = B_j$$

2.6 Односторонние t-тесты

Раньше было:

$$H_0 : b_j = B$$

$$H_1 : b_j \neq B$$

Однако, возможны другие вопросы:

$$H_0 : b_j = B$$

$$H_1 : b_j = C$$

или:

$$H_0 : b_j = B$$

$$H_1 : b_j > (<) B$$

Пусть:

$$H_0 : b_j = B$$

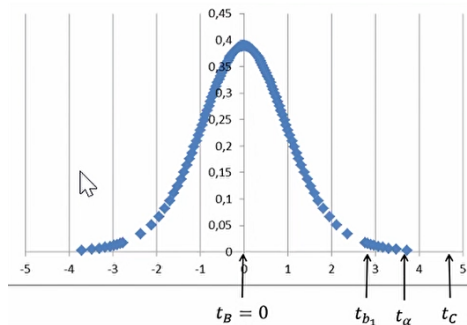
$$H_1 : b_j = C$$

$$B < C$$

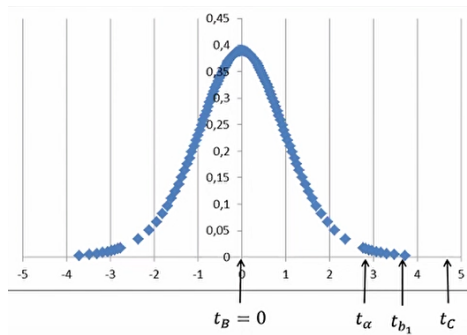
$$y = \hat{b}_0 + \hat{b}_1 \cdot x + \hat{\varepsilon}$$

Вариант 1:

$$B < \hat{b}_1 < C$$



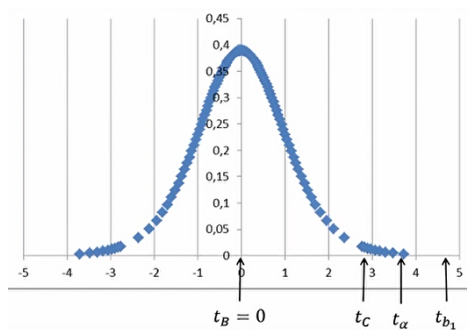
Так как $|t_{b_1}| < t_{\alpha}$, то мы принимаем нулевую гипотезу с вероятностью $1 - \alpha$.



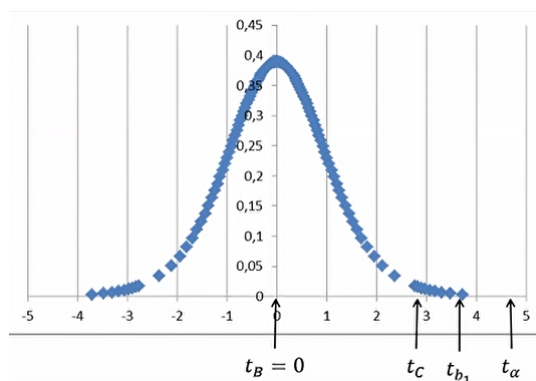
Так как $|t_{b_1}| > t_{\alpha}$, то мы отклоняем нулевую гипотезу и склоняемся в сторону альтернативной с вероятностью $1 - \alpha$.

Вариант 2:

$$B < C < \hat{b}_1$$



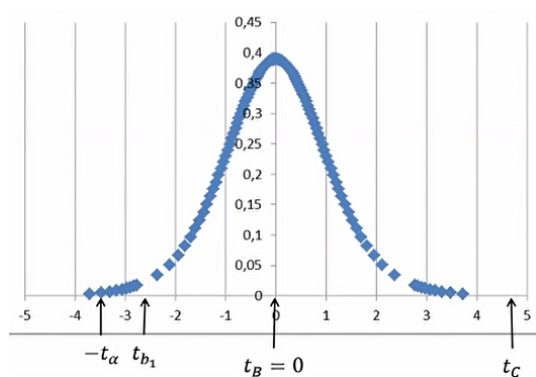
Так как $|t_{b_1}| > t_{\alpha}$, то мы отклоняем нулевую гипотезу и склоняемся в сторону альтернативной с вероятностью $1 - \alpha$.



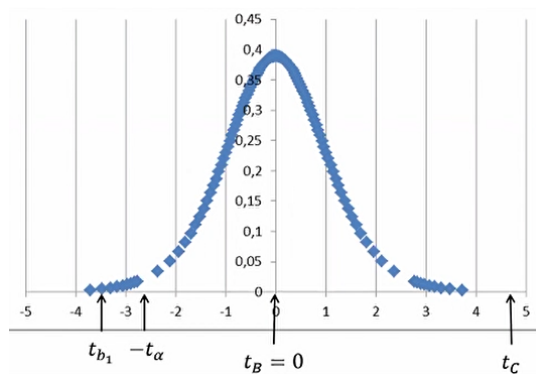
Так как $|t_{b_1}| < t_{\alpha}$, то мы принимаем нулевую гипотезу с вероятностью $1 - \alpha$, что странно, так как b_1 ближе к C , чем к B – проблема в том, что мы выбрали слишком близкие друг к другу C и B .

Вариант 3:

$$\hat{b}_1 < B < C$$



Так как $|t_{b_1}| < t_{\alpha}$, то мы принимаем нулевую гипотезу с вероятностью $1 - \alpha$.



В этой ситуации, как ни странно, принимаем нулевую гипотезу с вероятностью $1 - \alpha$, несмотря на то, что $|t_{b_1}| > t_{\alpha}$, так как при любых \hat{b}_1 : $\hat{b}_1 < B < C$ значение \hat{b}_1 будет ближе к B , чем к C . В этом проявляется односторонность t-теста. То есть, при $\hat{b}_1 < B < C$, если $t_{b_1} < 0$, то мы всегда принимаем нулевую гипотезу.

Обобщим полученное:

$$H_0 : b_j = B$$

$$H_1 : b_j = C$$

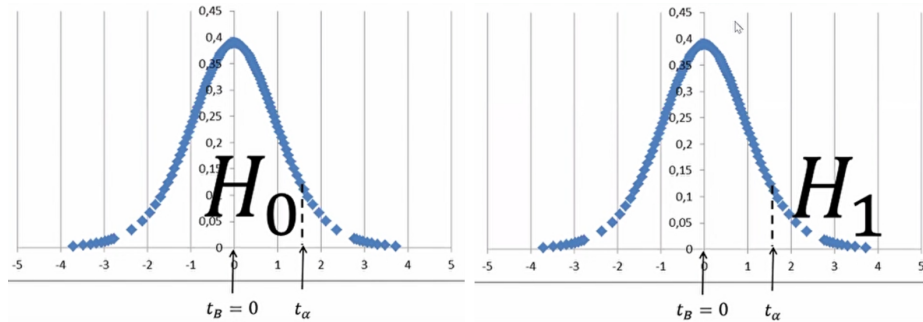
$$(B < C)$$

Если $t_\alpha < t_{b_j}$, то мы отвергаем нулевую гипотезу, а если $t_\alpha > t_{b_j}$, то нет оснований отвергнуть нулевую гипотезу. Из за такого направления (либо до t_α , либо после) такой тест называется односторонним t-тестом.

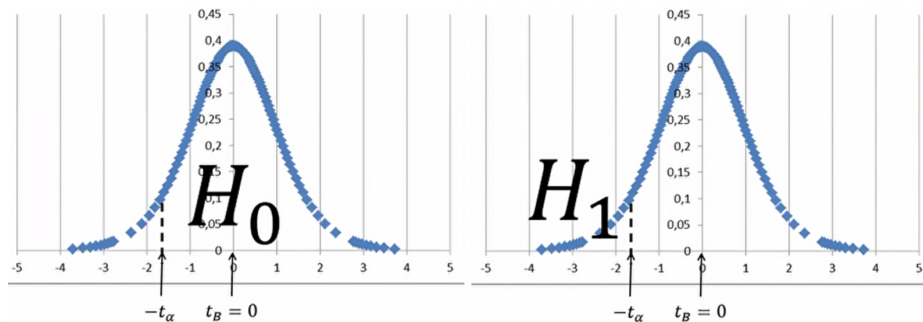
$$(C < B)$$

Если $t_{b_j} < -t_\alpha$, то мы отвергаем нулевую гипотезу, а если $-t_\alpha < t_{b_j}$, то нет оснований отвергнуть нулевую гипотезу.

$$B < C$$



$$C < B$$



Аналитически:

$$H_0 : b_j = B$$

$$H_1 : b_j = C$$

Если $B < C$, то $H_1 : b_j = C \Rightarrow H_1 : b_j > B$, а если $B > C$, то $H_1 : b_j = C \Rightarrow H_1 : b_j < B$.

Уровень значимости α :

- Двусторонний t-критерий: $P(t < -t_\alpha) = P(t > t_\alpha) = \frac{\alpha}{2}$
- Односторонний t-критерий: или $B < C$ и $P(t > t_\alpha) = \alpha$, или $B > C$ и $P(t < -t_\alpha) = \alpha$.

Поэтому, если хотим взять t_α для одностороннего критерия, пользуясь таблицей двустороннего t-критерия, то нужно взять α в два раза больше (так как такое значение отсекает левые и правые области по 0.05).

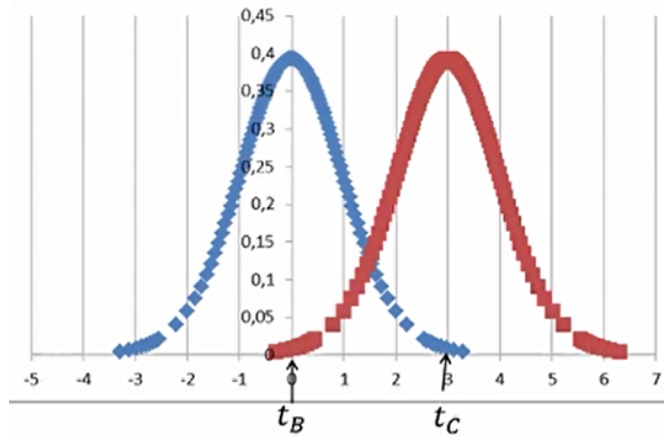
Пусть $B < \hat{b}_j < C$:

$$H_0 : b_j = B$$

$$H_1 : b_j = C$$

Если $t_b < t_\alpha$, то принимаем H_0 . В таком случае возможна ошибка второго рода – принятие неверной нулевой гипотезы. В односторонних тестах можно определить вероятность такой ошибки, ошибки второго рода.

$$B < C$$



$$B < C$$

$$t_c = ?$$

$$H_0 : b_j = B$$

Пусть $\hat{b}_j = C$:

$$t_c = \frac{C - B}{se(\hat{b}_j)} > 0$$

Или, в другой системе координат, сдвинутой с началом вправо по оси абсцисс на t_c :

$$H_0 : b_j = C$$

$$t_c = \frac{C - C}{se(\hat{b}_j)} = 0$$

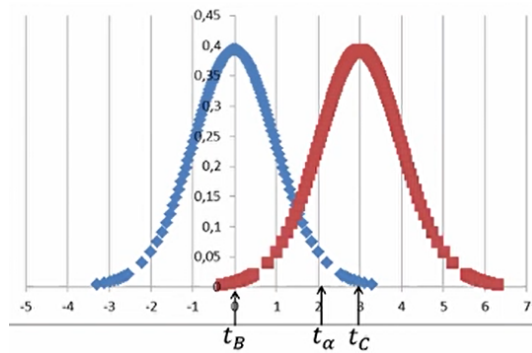
В данном случае думаем про нулевую гипотезу как $H_0 : b_j = B$.

$$B < C$$

У синего распределения t_α отсекает маленький хвост, соответствующий вероятности α . Видно, что, если мы примем при таком раскладе нулевую гипотезу, то, возможно, если нулевой гипотезой было бы $H_0 : b_j = C$, то, возможно, мы отказались бы от такой нулевой гипотезы, и это было бы лучше. Возможно, мы примем неверную нулевую гипотезу, то есть совершим ошибку второго рода.

$$H_0 : b_j = B$$

$$t_\alpha = \frac{b_\alpha - B}{se(\hat{b}_j)}$$



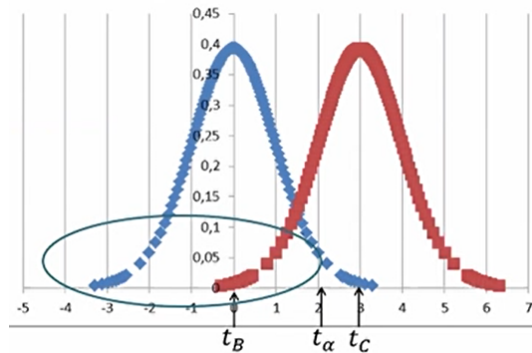
$$b_{\alpha} = B + t_{\alpha} \cdot se(\hat{b}_j)$$

Если нулевая гипотеза является обратной (для $B < C$):

$$H_0 : b_j = C$$

$$t_{b_{\alpha}(c)} = \frac{b_{\alpha} - C}{se(\hat{b}_j)} < 0$$

Хвост, отсекаемый $t_{b_{\alpha}(c)}$, у красного распределения, будет большим на графике в нашем случае (но в общем случае не факт).



Для того, чтобы выяснить, насколько велика вероятность выбрать неверную нулевую гипотезу: находим t_{α} , после рассчитываем $t_{b_{\alpha}(c)}$. Тогда, вероятность ошибки второго рода γ ($B < C$):

$$\gamma = P(t < t_{b_{\alpha}(c)})$$

Обратная величина, $1 - \gamma$, называется *мощностью критерия*. Чем больше разность $C - B$, тем дальше друг от друга расположены распределения, тем меньше будет ошибка второго рода.

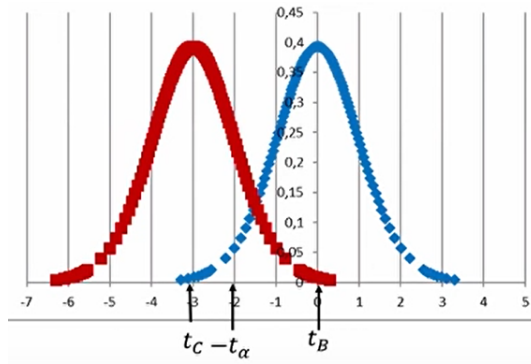
$$B > C$$

Вероятность ошибки второго рода γ ($B > C$):

$$\gamma = P(t \geq t_{b_{\alpha}(c)}) = 1 - P(t < t_{b_{\alpha}(c)})$$

2.6.1 Практическое использование односторонних t-тестов

Оценка стационарности временного ряда. Стационарный ряд:



- $MY_t = const$
- $DY_t = const$
- $cov(Y_t, Y_{t-\tau}) = f(\tau) \forall t, \tau$

Стационарные процессы:

- Процессы МА:

$$y_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}$$

- Процессы AR:

$$y_t = a + c_1 y_{t-1} + \dots + c_p y_{t-p} + \varepsilon_t$$

Частные случаи процессов:

- «Белый шум» – частный случай стационарного процесса, когда $cov(Y_t, Y_{t-1}) = 0$
- Случайное блуждание: $y_t = y_{t-1} + \varepsilon_t$
- Условие стационарности: корни уравнения (включая комплексные)

$$1 - c_1 z - c_2 z^2 - \dots - c_p z^p = 0$$

должны лежать вне «единичного корня», то есть $|z| > 1$

Нестационарные процессы:

- С детерминированным трендом: $y_t = b_0 + b_1 t + \varepsilon_t$
- Со стохастическим трендом: $y_t = c_1 y_{t-1} + \varepsilon_t$ или $y_t = b_0 + c_1 y_{t-1} + \varepsilon_t$
- С детерминированным и стохастическим трендом: $y_t = b_0 + b_1 t + c_1 y_{t-1} + \varepsilon_t$

Условие стационарности для $y_t = c_1 y_{t-1} + \varepsilon_t$ при $c_1 < 1$, если $c_1 = 1$, то случайное блуждание, если $c_1 > 1$, то «взрывной» характер динамики.

По факту, проверяем, либо $c_1 = 1$, либо $c_1 > 1$. То есть, полная аналогия с односторонним t-тестом. Возникает проблема, которая заключается в том, что, если коэффициент c_1 близок к единице, то есть, если гипотеза нулевая верна, то t-статистика не распределена по закону Стюдента. Эта проблема решается – проведем преобразование:

$$y_t = c_1 y_{t-1} + \varepsilon_t$$

$$y_t - y_{t-1} = c_1 y_{t-1} + \varepsilon_t - y_{t-1}$$

$$\Delta y_t = (c_1 - 1)y_{t-1} + \varepsilon_t$$

$$\Delta y_t = \delta y_{t-1} + \varepsilon_t$$

Это преобразование привело к тому, что в нестационарном процессе коэффициент δ при y_{t-1} будет равен не единице, а нулю. Тогда:

$$H_0 : \delta = 0$$

$$H_0 : \delta < 0$$

Однако, распределение такой случайной величины, по прежнему, не подчиняется распределению Стьюдента, но для этого распределения были рассчитаны критические значения Дики-Фуллера.

Тест Дики-Фуллера.

Критические значения: DF статистика. Стационарность:

$$t < DF_t$$

Тест Дики-Фуллера проверяет во временном ряду наличие единичного корня – то есть он проверяет нестационарность (единичный корень \approx нестационарность), а если гипотеза отвергается, то ряд считается стационарным. Если хотим получить стационарный ряд, то p -value должно быть достаточно маленьким, чтобы мы отвергли нулевую гипотезу о нестационарности. Стационарность:

$$t < DF$$

Расширенный тест Дики-Фуллера (ADF).

$$\Delta y_t = \delta \cdot y_{t-1} + \sum_{j=1}^{p-1} (\alpha_j \cdot \Delta y_{t-j}) + \varepsilon_t$$

Где $\delta = c_1 + c_2 + \dots + c_p - 1$

$$H_0 : \delta = 0$$

$$H_1 : \delta < 0$$

2.7 Статистические характеристики оценок параметров уравнения регрессии

Для изучения этих характеристик требуется матричная запись уравнения регрессии:

$$Y = X \cdot \hat{B} + \hat{E}$$

Вектор \hat{B} является вектором оценок истинных значений, поэтому они подвержены случайным колебаниям, то есть являются случайными величинами:

$$M\hat{B} = ?$$

$$D\hat{B} = \sigma_{\hat{B}}^2 = ?$$

Матожидание оценок использовалось в препосылках классической нормальной линейной модели, а дисперсия оценок использовалась в расчете критерия Стьюдента.

Из МНК получаем вектор оценок параметров:

$$\hat{B} = (X^T \cdot X)^{-1} \cdot X^T \cdot Y$$

$$\hat{B} = (X^T \cdot X)^{-1} \cdot X^T \cdot (X \cdot B + E)$$

Предполагаем, что мы правильно угадали параметры истинной модели (и состав переменных, и функциональную форму):

$$\begin{aligned} M\hat{B} &= M((X^T \cdot X)^{-1} \cdot X^T \cdot (X \cdot B + E)) = M((X^T \cdot X)^{-1} \cdot (X^T \cdot X) \cdot B) + \\ &+ M((X^T \cdot X)^{-1} \cdot X^T \cdot E) = MB + (X^T \cdot X)^{-1} \cdot X^T \cdot ME = B + 0 = B \end{aligned}$$

Оценка дисперсии:

$$D\hat{B} = \sigma_B^2 = ?$$

$$\begin{aligned} cov(\hat{B}) &= M\left((\hat{B} - M\hat{B}) \cdot (\hat{B} - M\hat{B})^T\right) = \\ &= M\left(((X^T \cdot X)^{-1} \cdot X^T \cdot (X \cdot B + E) - B) \cdot ((X^T \cdot X)^{-1} \cdot X^T \cdot (X \cdot B + E) - B)^T\right) = \\ &= M((B + (X^T \cdot X)^{-1} \cdot X^T \cdot E - B) \cdot (B + (X^T \cdot X)^{-1} \cdot X^T \cdot E - B)^T) = \\ &= M(((X^T \cdot X)^{-1} \cdot X^T \cdot E) \cdot ((X^T \cdot X)^{-1} \cdot X^T \cdot E)^T) = \\ &= (X^T \cdot X)^{-1} \cdot X^T \cdot M(E \cdot E^T) \cdot X \cdot (X^T \cdot X)^{-1} \end{aligned}$$

$$M(E \cdot E^T) = ?$$

$$M(E \cdot E^T) = M((E - ME) \cdot (E - ME)^T) = cov(E) = \Omega$$

Матрица ковариации случайных остатков Ω :

$$\Omega = \begin{pmatrix} \sigma_{\varepsilon_1}^2 & cov(\varepsilon_1, \varepsilon_2) & \dots & cov(\varepsilon_1, \varepsilon_n) \\ cov(\varepsilon_2, \varepsilon_1) & \sigma_{\varepsilon_2}^2 & \dots & cov(\varepsilon_2, \varepsilon_n) \\ \vdots & \vdots & \ddots & \vdots \\ cov(\varepsilon_n, \varepsilon_1) & cov(\varepsilon_n, \varepsilon_2) & \dots & \sigma_{\varepsilon_n}^2 \end{pmatrix}$$

При гомоскедастичности и некоррелированности случайных остатков:

$$\Omega = \begin{pmatrix} \sigma_{\varepsilon}^2 & 0 & \dots & 0 \\ 0 & \sigma_{\varepsilon}^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_{\varepsilon}^2 \end{pmatrix} = \sigma_{\varepsilon}^2 \cdot I$$

$$\begin{aligned} cov(\hat{B}) &= (X^T \cdot X)^{-1} \cdot X^T \cdot M(E \cdot E^T) \cdot X \cdot (X^T \cdot X)^{-1} = (X^T \cdot X)^{-1} \cdot X^T \cdot \sigma_{\varepsilon}^2 \cdot I \cdot X \cdot (X^T \cdot X)^{-1} = \\ &= \sigma_{\varepsilon}^2 \cdot (X^T \cdot X)^{-1} \cdot X^T \cdot X \cdot (X^T \cdot X)^{-1} = \sigma_{\varepsilon}^2 \cdot (X^T \cdot X)^{-1} \end{aligned}$$

Таким образом:

$$\begin{aligned} cov(\hat{B}) &= (X^T \cdot X)^{-1} \cdot \sigma_{\varepsilon}^2 \\ cov(\hat{b}_j, \hat{b}_j) &= \sigma_{\hat{b}_j}^2 \\ \sigma_{\varepsilon}^2 &= ? \end{aligned}$$

Для того, чтобы иметь дисперсию коэффициентов регрессии, необходимо найти истинную дисперсию остатков.

$$\hat{E} = Y - X \cdot \hat{B} = Y - X \cdot (X^T \cdot X)^{-1} \cdot X^T \cdot Y =$$

$$\begin{aligned}
&= (X \cdot B + E) - X \cdot ((X^T \cdot X)^{-1} \cdot X^T \cdot (X \cdot B + E)) = \\
&= X \cdot B + E - X \cdot (X^T \cdot X)^{-1} \cdot X^T \cdot X \cdot B - X \cdot (X^T \cdot X)^{-1} \cdot X^T \cdot E = \\
&= X \cdot B + E - X \cdot B - X \cdot (X^T \cdot X)^{-1} \cdot X^T \cdot E = \\
&= E - X \cdot (X^T \cdot X)^{-1} \cdot X^T \cdot E
\end{aligned}$$

Теперь найдем матожидание произведения векторов случайных остатков:

$$\begin{aligned}
M(\hat{E}^T \cdot \hat{E}) &= M((E - X \cdot (X^T \cdot X)^{-1} \cdot X^T \cdot E)^T \cdot (E - X \cdot (X^T \cdot X)^{-1} \cdot X^T \cdot E)) = \\
&= M(E^T \cdot E - E^T \cdot X \cdot (X^T \cdot X)^{-1} \cdot X^T \cdot E - E^T \cdot X \cdot (X^T \cdot X)^{-1} \cdot X^T \cdot E + \\
&\quad + E^T \cdot X \cdot (X^T \cdot X)^{-1} \cdot X^T \cdot X \cdot (X^T \cdot X)^{-1} \cdot X^T \cdot E) = \\
&= M(E^T \cdot E - E^T \cdot X \cdot (X^T \cdot X)^{-1} \cdot X^T \cdot E) =
\end{aligned}$$

Уменьшаемое:

$$M(E^T \cdot E) = M((E - ME)^T \cdot (E - ME)) = n \cdot \sigma_\varepsilon^2$$

Вычитаемое:

$$\begin{aligned}
&X \cdot (X^T \cdot X)^{-1} \cdot X^T \equiv C(n \times n) \\
M(E^T \cdot C \cdot E) &= M\left(\sum_{i,j=1}^n (c_{ij} \cdot \varepsilon_j \cdot \varepsilon_i)\right) = \sum_{i,j=1}^n c_{ij} \cdot M(\varepsilon_i \cdot \varepsilon_j) = \\
&= \sum_{i,j=1}^n c_{ij} \cdot \text{cov}(\varepsilon_i, \varepsilon_j) = \sum_{i=1}^n c_{ii} \cdot \text{cov}(\varepsilon_i, \varepsilon_i) = \sum_{i=1}^n c_{ii} \cdot \sigma_\varepsilon^2 = \sigma_\varepsilon^2 \sum_{i=1}^n c_{ii} = \sigma_\varepsilon^2 \cdot \text{tr}(C) = \\
&= \sigma_\varepsilon^2 \cdot \text{tr}(X \cdot (X^T \cdot X)^{-1} \cdot X^T) = \sigma_\varepsilon^2 \cdot \text{tr}((X^T \cdot X)^{-1} X^T \cdot X) = \sigma_\varepsilon^2 \cdot \text{tr}(E_{m+1}) = (m+1) \cdot \sigma_\varepsilon^2
\end{aligned}$$

Таким образом:

$$M(\hat{E}^T \cdot \hat{E}) = n \cdot \sigma_\varepsilon^2 - (m+1) \cdot \sigma_\varepsilon^2 = (n-m-1) \cdot \sigma_\varepsilon^2$$

С другой стороны:

$$M(\hat{E}^T \cdot \hat{E}) = \sum_{i=1}^n \hat{\varepsilon}_i^2 = SS_{\text{ост}}$$

Тогда получаем оценку дисперсии случайных остатков:

$$\hat{\sigma}_\varepsilon^2 = \frac{\sum_{i=1}^n \hat{\varepsilon}_i^2}{n-m-1} = \frac{SS_{\text{ост}}}{n-m-1} = s_e^2$$

Вывели раньше:

$$\begin{aligned}
\text{cov}(\hat{B}) &= (X^T \cdot X)^{-1} \cdot \hat{\sigma}_\varepsilon^2 \\
\hat{\sigma}_{\hat{b}_j} &= \hat{\sigma}_\varepsilon^2 \cdot [(X^T \cdot X)^{-1}]_{jj} \quad j = 0, 1, \dots, m
\end{aligned}$$

Например, для $y = \hat{b}_0 + \hat{b}_1 \cdot x + \hat{\varepsilon}$:

$$X^T \cdot X = \begin{pmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{pmatrix}$$

Диагональные элементы матрицы $(X^T \cdot X)^{-1}$:

$$jj = 11 \Rightarrow \frac{n}{|X^T \cdot X|} = \frac{n}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}$$

$$\begin{aligned}
&= \frac{n}{n^2 \cdot \sigma_x^2} = \frac{1}{\sum_{i=1}^n (x_i - \bar{x})^2} \\
&[(X^T \cdot X)^{-1}]_{11} = \frac{1}{\sum_{i=1}^n (x_i - \bar{x})^2} \\
&\hat{\sigma}_{b_1}^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2 / (n-2)}{\sum_{i=1}^n (x_i - \bar{x})^2} \\
&\hat{\sigma}_{b_1} = se(\hat{b}_1) = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2 / (n-2)}{\sum_{i=1}^n (x_i - \bar{x})^2}}
\end{aligned}$$

Диагональные элементы матрицы $(X^T \cdot X)^{-1}$:

$$\begin{aligned}
jj = 00 &\Rightarrow \frac{\sum_{i=1}^n x_i^2}{|X^T \cdot X|} = \frac{\sum_{i=1}^n x_i^2}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \\
&= \frac{\sum_{i=1}^n x_i^2}{n^2 \cdot \sigma_x^2} = \frac{\sum_{i=1}^n x_i^2}{n \cdot \sum_{i=1}^n (x_i - \bar{x})^2} \\
&[(X^T \cdot X)^{-1}]_{00} = \frac{\sum_{i=1}^n x_i^2}{n \cdot \sum_{i=1}^n (x_i - \bar{x})^2} \\
&\hat{\sigma}_{b_0}^2 = \frac{\sum_{i=1}^n x_i^2 \cdot \sum_{i=1}^n (y_i - \hat{y}_i)^2 / (n-2)}{n \cdot \sum_{i=1}^n (x_i - \bar{x})^2} \\
&\hat{\sigma}_{b_0} = se(\hat{b}_0) = \sqrt{\frac{\sum_{i=1}^n x_i^2 \cdot \sum_{i=1}^n (y_i - \hat{y}_i)^2 / (n-2)}{n \cdot \sum_{i=1}^n (x_i - \bar{x})^2}}
\end{aligned}$$

К вопросу об использовании t-статистики для проверки гипотез о значениях параметров регрессии.

$$\hat{b}_j \sim N(b_j, \sigma_{b_j}^2)$$

Где $\sigma_{b_j}^2$ – истинная дисперсия \hat{b}_j , а $se^2(\hat{b}_j)$ – оценка дисперсии \hat{b}_j .

$$\frac{\hat{b}_j - B_j}{\sigma_{b_j}} \sim N(0, 1)$$

Можно показать, что:

$$\frac{SS_{\text{ост}}}{\sigma_\varepsilon^2} = \frac{\hat{\sigma}_\varepsilon^2 \cdot (n - m - 1)}{\sigma_\varepsilon^2} \sim \chi^2(n - m - 1)$$

Тогда:

$$\begin{aligned}
&\frac{\hat{b}_j - B_j}{\sigma_{b_j}} \bigg/ \sqrt{\frac{\hat{\sigma}_\varepsilon^2 \cdot (n - m - 1)}{\sigma_\varepsilon^2 \cdot (n - m - 1)}} \sim t(n - m - 1) \\
&\frac{(\hat{b}_j - B_j) \cdot \sigma_\varepsilon}{\sigma_{b_j} \cdot \hat{\sigma}_\varepsilon} \sim t(n - m - 1)
\end{aligned}$$

Так как:

$$\frac{\sigma_{b_j}}{\sigma_\varepsilon} = \frac{se(b_j)}{\hat{\sigma}_\varepsilon} \Rightarrow \frac{\hat{b}_j - B_j}{se(b_j)} \sim t(n - m - 1)$$

2.8 Оценка значимости регрессии (тренда)

Проверка значимости уравнения регрессии (тренда):

$$H_0 : b_1 = b_2 = \dots = b_p = 0$$

$$H_1 : \exists b_j \neq 0$$

Суммы квадратов при выполнении нулевой гипотезы H_0 :

$$SS_{\text{ост}} = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - b_0)^2 = \sum_{i=1}^n (y_i - (\bar{y} - b_1 \cdot \bar{x}))^2 = \sum_{i=1}^n (y_i - \bar{y})^2 = SS_{\text{общ}}$$

$$SS_{\text{факт}} = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 = \sum_{i=1}^n (b_0 - \bar{y})^2 = \sum_{i=1}^n (\bar{y} - \bar{y})^2 = 0$$

Если $H_0 : b_1 = b_2 = \dots = b_p = 0$, то

$$\frac{SS_{\text{ост}}}{\sigma_\varepsilon^2} \sim \chi^2(n - m - 1)$$

$$\frac{SS_{\text{факт}}}{\sigma_\varepsilon^2} \sim \chi^2(m)$$

Тогда:

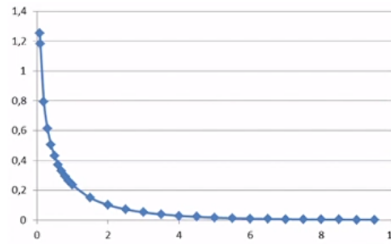
$$\frac{SS_{\text{факт}} / m}{SS_{\text{ост}} / (n - m - 1)} = \frac{SS_{\text{факт}} \cdot (n - m - 1)}{SS_{\text{ост}} \cdot m} = \frac{SS_{\text{факт}}}{SS_{\text{ост}}} \cdot \frac{n - m - 1}{n} \sim F(m, n - m - 1)$$

Критерий Фишера:

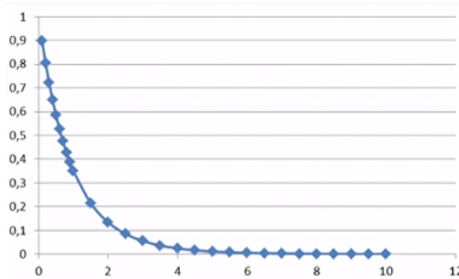
$$F = \frac{DS_{\text{факт}}}{DS_{\text{ост}}} = \frac{SS_{\text{факт}} \cdot (n - m - 1)}{SS_{\text{ост}} \cdot m} = \frac{SS_{\text{факт}}}{SS_{\text{ост}}} \cdot \frac{n - m - 1}{n}$$

Если нулевая гипотеза выполняется, то F -статистика имеет распределение Фишера, поэтому в качестве табличного значения мы берем $F_{\text{табл}}(\alpha; m; n - m - 1)$ и смотрим, насколько полученное распределение похоже на распределение Фишера.

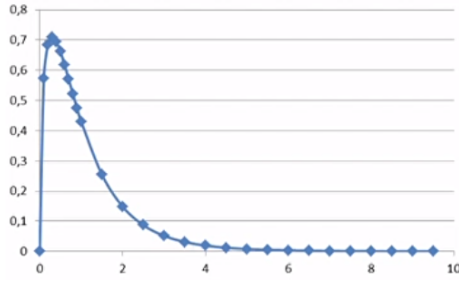
Плотность распределения Фишера при степенях свободы ($m = 1, n - m - 1 = 20$):



Плотность распределения Фишера при степенях свободы ($m = 3, n - m - 1 = 20$):



Плотность распределения Фишера при степенях свободы ($m = 18, n - m - 1 = 20$):



Так как справедливы следующие выражения:

$$R^2 = \frac{SS_{\text{факт}}}{SS_{\text{общ}}} = 1 - \frac{SS_{\text{ост}}}{SS_{\text{общ}}}$$

$$SS_{\text{факт}} = R^2 \cdot SS_{\text{общ}}$$

$$SS_{\text{ост}} = (1 - R^2) \cdot SS_{\text{общ}}$$

То, подставляя эти выражения в формулу F -статистики, получаем, что:

$$F = \frac{R^2 \cdot SS_{\text{общ}}}{(1 - R^2) \cdot SS_{\text{общ}}} \cdot \frac{n - m - 1}{m}$$

$$F = \frac{R^2}{1 - R^2} \cdot \frac{n - m - 1}{m}$$

Таким образом, если $b_1 = b_2 = \dots = b_p = 0$, то $y = b_0 + 0 \cdot x_1 + \dots + 0 \cdot x_p + \varepsilon = b_0 + \varepsilon$ (то есть уравнение выглядит как стационарный ряд, незначимо влияние x на y). Тогда:

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = \frac{\sum_{i=1}^n (b_0 - b_0)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = 0$$

То есть, можно сказать, что нулевая гипотеза о равенстве в истинной модели всех параметров нулю эквивалентна нулевой гипотезе о незначимости коэффициента детерминации:

$$H_0 : b_1 = \dots = b_p = 0 \iff H_0 : R^2 = 0$$

Для $y = \hat{b}_0 + \hat{b}_1 \cdot x + \hat{\varepsilon}$:

$$t_{b_1} = \frac{\hat{b}_1}{se(b_1)}$$

Так как

$$se(b_1) = \frac{SS_{\text{ост}}}{(n - 2) \cdot \sum_{i=1}^n (x_i - \bar{x})^2}$$

то:

$$t_{b_1}^2 = \frac{\hat{b}_1^2 \cdot (n - 2) \cdot \sum_{i=1}^n (x_i - \bar{x})^2}{SS_{\text{ост}}} = \frac{\hat{b}_1^2 \cdot (n - 2) \cdot \sigma_x^2 \cdot n}{SS_{\text{ост}}}$$

Также, мы знаем, что:

$$\hat{r} = \hat{b}_1 \cdot \frac{\sigma_x}{\sigma_y}, \quad \hat{b}_1 = \hat{r} \cdot \frac{\sigma_y}{\sigma_x}$$

Откуда, подставляя \hat{b}_1 , получим

$$t_{b_1}^2 = \frac{\hat{r}^2 \cdot \sigma_y^2 \cdot (n-2) \cdot \sigma_x^2 \cdot n}{\sigma_x^2 \cdot SS_{\text{ост}}} = \frac{\hat{r}^2 \cdot \sigma_y^2 \cdot (n-2) \cdot n}{SS_{\text{ост}}} = \frac{R^2 \cdot \sigma_y^2 \cdot (n-2) \cdot n}{SS_{\text{ост}}} = \frac{SS_{\text{факт}} \cdot \sigma_y^2 \cdot (n-2) \cdot n}{SS_{\text{общ}} \cdot SS_{\text{ост}}} =$$

$$= \frac{SS_{\text{факт}} \cdot \sigma_y^2 \cdot (n-2) \cdot n}{\sigma_y^2 \cdot n \cdot SS_{\text{ост}}} = \frac{SS_{\text{факт}} \cdot (n-2)}{SS_{\text{ост}}} = \frac{SS_{\text{факт}}}{SS_{\text{ост}}} \cdot \frac{n-1-1}{1} = F$$

2.9 Последствия неправильной спецификации модели

Спецификация модели предполагает:

1. Выбор независимых переменных;
2. Выбор функциональной формы модели.

Каковы последствия, если функциональная форма не соответствует реальному экономическому явлению? Будет:

1. Ошибки интерпретации и прогнозирования;
2. Нарушения требований к случайным остаткам;
3. Недостоверность оценок параметров.

Каковы последствия, если будет несоответствие состава независимых переменных «реальной» модели:

1. Невключение существенных переменных;
2. Включение несущественных переменных.

Начнем рассматривать вопрос со случая не включения существенных переменных. «Реальная» модель:

$$y = b_0 + b_1 \cdot x_1 + \dots + a_1 \cdot z_1 + \dots + \varepsilon$$

$$Y = X \cdot B + Z \cdot A + E$$

Построена модель:

$$y = \hat{b}_0 + \hat{b}_1 \cdot x_1 + \dots + \hat{\varepsilon}$$

$$Y = X \cdot \hat{B} + \hat{E}$$

Получаем МНК-оценку параметров коэффициентов «неверной» модели:

$$\hat{B} = (X^T \cdot X)^{-1} \cdot X^T \cdot Y = (X^T \cdot X)^{-1} \cdot X^T \cdot (X \cdot B + Z \cdot A + E)$$

$$M\hat{B} = B + (X^T \cdot X)^{-1} \cdot X^T \cdot Z \cdot MA + 0 = B + (X^T \cdot X)^{-1} \cdot X^T \cdot Z \cdot A$$

Поскольку $M\hat{B} = B + \dots$, то оценка коэффициентов является смещенной, если $(X^T \cdot X)^{-1} \cdot X^T \cdot Z \cdot A \neq 0$.

Оценка \hat{B} несмещена, если:

1. $Z = 0$ (то есть «реальная» модель соответствовала построенной)
2. $X^T \cdot Z = 0$

$$D\hat{B} = \sigma_{\hat{B}}^2 = ?$$

$$\text{cov}(\hat{B}) = M \left((\hat{B} - M\hat{B}) \cdot (\hat{B} - M\hat{B})^T \right)$$

$$\begin{aligned} \hat{B} - M\hat{B} &= (X^T \cdot X)^{-1} \cdot X^T \cdot (X \cdot B + Z \cdot A + E) - B - (X^T \cdot X)^{-1} \cdot X^T \cdot Z \cdot A = \\ &= (X^T \cdot X)^{-1} \cdot X^T \cdot X \cdot B + (X^T \cdot X)^{-1} \cdot X^T \cdot Z \cdot A + (X^T \cdot X)^{-1} \cdot X^T \cdot E - B - (X^T \cdot X)^{-1} \cdot X^T \cdot Z \cdot A = \\ &= (X^T \cdot X)^{-1} \cdot X^T \cdot E \end{aligned}$$

Тогда, подставив $\hat{B} - M\hat{B}$, получим, что:

$$\begin{aligned} \text{cov}(\hat{B}) &= M \left((\hat{B} - M\hat{B}) \cdot (\hat{B} - M\hat{B})^T \right) = M \left(((X^T \cdot X)^{-1} \cdot X^T \cdot E) \cdot ((X^T \cdot X)^{-1} \cdot X^T \cdot E)^T \right) = \\ &= (X^T \cdot X)^{-1} \cdot M(E \cdot E^T) = (X^T \cdot X)^{-1} \cdot \sigma_{\varepsilon}^2 \end{aligned}$$

Нужно изучить поведение дисперсии случайных остатков построенной модели, поэтому посмотрим на:

$$\hat{E} = Y - X \cdot \hat{B} = Y - X \cdot (X^T \cdot X)^{-1} \cdot X^T \cdot Y = (I - X \cdot (X^T \cdot X)^{-1} \cdot X^T) \cdot Y$$

Положим $I - X \cdot (X^T \cdot X)^{-1} \cdot X^T \equiv S$. Можно показать, что такая матрица S является симметричной (то есть $S^T = S$) и идемпотентной (то есть $S^2 = S$). Также, она обладает другим свойством - произведение матрицы S на X равно нулевой матрице:

$$S \cdot X = X - X \cdot (X^T \cdot X)^{-1} \cdot X^T \cdot X = X - X = 0$$

Тогда, получаем, что:

$$\hat{E} = S \cdot Y$$

$$\begin{aligned} M(\hat{E}^T \cdot \hat{E}) &= M(Y^T \cdot S^T \cdot S \cdot Y) = M(Y^T \cdot S \cdot S \cdot Y) = M(Y^T \cdot S^2 \cdot Y) = M(Y^T \cdot S \cdot Y) = \\ &= M((X \cdot B + Z \cdot A + E)^T \cdot S \cdot (X \cdot B + Z \cdot A + E)) = \\ &= M(B^T \cdot X^T \cdot S \cdot X \cdot B + B^T \cdot X^T \cdot S \cdot Z \cdot A + B^T \cdot X^T \cdot S \cdot E + A^T \cdot Z^T \cdot S \cdot X \cdot B + \\ &+ A^T \cdot Z^T \cdot S \cdot Z \cdot A + A^T \cdot Z^T \cdot S \cdot E + E^T \cdot S \cdot X \cdot B + E^T \cdot S \cdot Z \cdot A + E^T \cdot S \cdot E) = \end{aligned}$$

Поскольку $S \cdot X = 0$, то выражение упрощается:

$$\begin{aligned} M(\hat{E}^T \cdot \hat{E}) &= M(A^T \cdot Z^T \cdot S \cdot Z \cdot A + A^T \cdot Z^T \cdot S \cdot E + E^T \cdot S \cdot Z \cdot A + E^T \cdot S \cdot E) = \\ &= M(A^T \cdot Z^T \cdot S \cdot Z \cdot A + E^T \cdot S \cdot E) = M(E^T \cdot S \cdot E) + M(A^T \cdot Z^T \cdot S \cdot Z \cdot A) \end{aligned}$$

Положим $S \equiv I - X \cdot (X^T \cdot X)^{-1} \cdot X^T \equiv I - C$, тогда:

$$M(\hat{E}^T \cdot \hat{E}) = M(\hat{E}^T \cdot I \cdot \hat{E}) - M(\hat{E}^T \cdot C \cdot \hat{E}) + M(A^T \cdot Z^T \cdot S \cdot Z \cdot A)$$

Преобразования с $M(\hat{E}^T \cdot I \cdot \hat{E})$ и $M(\hat{E}^T \cdot C \cdot \hat{E})$ были рассмотрены в вопросе 2.7, поэтому, опустив преобразования из вопроса 2.7, мы можем записать:

$$M(\hat{E}^T \cdot \hat{E}) = n \cdot \sigma_{\varepsilon}^2 - \sigma_{\varepsilon}^2 \cdot (m + 1) + A^T \cdot Z^T \cdot S \cdot Z \cdot A = \sigma_{\varepsilon}^2 \cdot (n - m - 1) + A^T \cdot Z^T \cdot S \cdot Z \cdot A = SS_{\text{ост}}$$

Тогда:

$$\hat{\sigma}_{\varepsilon}^2 = \frac{SS_{\text{ост}}}{n - m - 1} - \frac{1}{n - m - 1} \cdot A^T \cdot Z^T \cdot S \cdot Z \cdot A$$

Мы не знаем A и Z , поэтому для полученной модели получаем следующую оценку дисперсии случайных остатков:

$$\hat{\sigma}_\varepsilon^2 = \frac{SS_{\text{ост}}}{n - m - 1}$$

Нужно понять, насколько неверная получилась оценка дисперсии случайных остатков.

Для «реальной» модели матрица независимых переменных $(X \ Z)$ имеет вид:

$$X \ Z = \begin{pmatrix} 1 & x_{11} & \dots & z_{11} & \dots \\ 1 & x_{12} & \dots & z_{12} & \dots \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & x_{1n} & \dots & z_{1n} & \dots \end{pmatrix}$$

Оценки коэффициентов «реальной модели»:

$$\begin{bmatrix} \hat{B}^* \\ \hat{A} \end{bmatrix} = \begin{pmatrix} \hat{b}_0^* \\ \hat{b}_1^* \\ \vdots \\ \hat{a}_0^* \\ \hat{a}_1^* \\ \vdots \end{pmatrix}$$

Матрица ковариации оценок коэффициентов «реальной» модели имеет вид:

$$\text{cov} \begin{bmatrix} \hat{B}^* \\ \hat{A} \end{bmatrix} = \left(\begin{bmatrix} X & Z \end{bmatrix}^T \cdot \begin{bmatrix} X & Z \end{bmatrix} \right)^{-1} \cdot \sigma_\varepsilon^2 = \left(\begin{bmatrix} X^T \\ Z^T \end{bmatrix} \cdot \begin{bmatrix} X & Z \end{bmatrix} \right)^{-1} \cdot \sigma_\varepsilon^2 = \begin{bmatrix} X^T \cdot X & X^T \cdot Z \\ Z^T \cdot X & Z^T \cdot Z \end{bmatrix}^{-1} \cdot \sigma_\varepsilon^2$$

Поскольку мы строили модель, оценивая \hat{B}^* , то в $\text{cov} \begin{bmatrix} \hat{B}^* \\ \hat{A} \end{bmatrix}$ нас интересует блок матрицы с \hat{B}^* .

По правилу обращения блочных матриц:

$$\text{cov}(\hat{B}^*) = (X^T \cdot X - X^T \cdot Z \cdot (Z^T \cdot Z)^{-1} \cdot Z^T \cdot X)^{-1} \cdot \sigma_\varepsilon^2$$

Чтобы избавиться от -1 степени матрицы, появляющейся в разложении $\text{cov}(\hat{B}^*)$, рассмотрим обратные матрицы ковариаций:

$$\begin{aligned} (\text{cov}(\hat{B}))^{-1} - (\text{cov}(\hat{B}^*))^{-1} &= ((X^T \cdot X)^{-1} \cdot \sigma_\varepsilon^2)^{-1} - ((X^T \cdot X - X^T \cdot Z \cdot (Z^T \cdot Z)^{-1} \cdot Z^T \cdot X)^{-1} \cdot \sigma_\varepsilon^2)^{-1} = \\ &= \frac{1}{\sigma_\varepsilon^2} (X^T \cdot X - X^T \cdot X + X^T \cdot Z \cdot (Z^T \cdot Z)^{-1} \cdot Z^T \cdot X) = \frac{1}{\sigma_\varepsilon^2} (X^T \cdot Z \cdot (Z^T \cdot Z)^{-1} \cdot Z^T \cdot X) \geq 0 \end{aligned}$$

Таким образом, ковариация «реальной модели» больше, чем ковариация «короткой» модели, при условии, если мы знаем истинную дисперсию случайных остатков.

Невключение существенной переменной:

1. Оценки параметров смещены;
2. Оценка дисперсии случайных остатков смещена;
3. Дисперсии оценок параметров «короткой» модели меньше, чем дисперсии оценок параметров «реальной» модели.

Второй случай – наоборот, включение несущественных переменных.

«Реальная модель»:

$$y = b_0 + b_1 \cdot x_1 + \dots + \varepsilon$$

$$Y = X \cdot B + E$$

Построена модель:

$$y = \hat{b}_0 + \hat{b}_1 \cdot x_1 + \dots + \hat{a}_1 \cdot z_1 + \dots + \hat{\varepsilon}$$

$$Y = X \cdot \hat{B} + Z \cdot \hat{A} + \hat{E}$$

МНК-оценка:

$$\begin{bmatrix} \hat{B} \\ \hat{A} \end{bmatrix} = \left(\begin{bmatrix} X & Z \end{bmatrix}^T \cdot \begin{bmatrix} X & Z \end{bmatrix} \right)^{-1} \cdot \begin{bmatrix} X & Z \end{bmatrix}^T \cdot Y = \left(\begin{bmatrix} X & Z \end{bmatrix}^T \cdot \begin{bmatrix} X & Z \end{bmatrix} \right)^{-1} \cdot \begin{bmatrix} X & Z \end{bmatrix}^T \cdot (X \cdot B + E)$$

$$\begin{bmatrix} \hat{B} \\ \hat{A} \end{bmatrix} = \begin{pmatrix} X^T \cdot X & X^T \cdot Z \\ Z^T \cdot X & Z^T \cdot Z \end{pmatrix}^{-1} \cdot \begin{pmatrix} X^T \\ Z^T \end{pmatrix} \cdot Y$$

Пусть:

$$\begin{pmatrix} X^T \cdot X & X^T \cdot Z \\ Z^T \cdot X & Z^T \cdot Z \end{pmatrix}^{-1} = \begin{pmatrix} L & D \\ E & F \end{pmatrix}$$

Тогда:

$$\hat{B} = (L \cdot X^T + D \cdot Z^T) \cdot Y$$

Нужно выразить L, D :

$$L = (X^T \cdot X - X^T \cdot Z \cdot (Z^T \cdot Z)^{-1} \cdot Z^T \cdot X)^{-1}$$

$$D = -L \cdot X^T \cdot Z \cdot (Z^T \cdot Z)^{-1}$$

Еще раз рассмотрим L :

$$L = (X^T \cdot X - X^T \cdot Z \cdot (Z^T \cdot Z)^{-1} \cdot Z^T \cdot X)^{-1} = (X^T \cdot (I - Z \cdot (Z^T \cdot Z)^{-1} \cdot Z^T) \cdot X)^{-1}$$

Положим S_z (матрицу S_z является идемпотентной и симметричной):

$$S_z \equiv I - Z \cdot (Z^T \cdot Z)^{-1} \cdot Z^T$$

$$L = (X^T \cdot S_z \cdot X)^{-1}$$

Тогда рассмотрим матрицу D :

$$D = -(X^T \cdot S_z \cdot X)^{-1} \cdot X^T \cdot Z \cdot (Z^T \cdot Z)^{-1}$$

Подставим L, D в вектор \hat{B} :

$$\begin{aligned} \hat{B} &= (L \cdot X^T + D \cdot Z^T) \cdot Y = ((X^T \cdot S_z \cdot X)^{-1} \cdot X^T - (X^T \cdot S_z \cdot X)^{-1} \cdot X^T \cdot Z \cdot (Z^T \cdot Z)^{-1} \cdot Z^T) \cdot Y = \\ &= ((X^T \cdot S_z \cdot X)^{-1} \cdot X^T \cdot (I - Z \cdot (Z^T \cdot Z)^{-1} \cdot Z^T) \cdot Y) = ((X^T \cdot S_z \cdot X)^{-1} \cdot X^T \cdot S_z \cdot Y) \end{aligned}$$

Найдем математическое ожидание вектора \hat{B} :

$$\begin{aligned} M\hat{B} &= M(((X^T \cdot S_z \cdot X)^{-1} \cdot X^T \cdot S_z \cdot Y)) = M(((X^T \cdot S_z \cdot X)^{-1} \cdot X^T \cdot S_z \cdot (X \cdot B + E))) = \\ &= M(((X^T \cdot S_z \cdot X)^{-1} \cdot X^T \cdot S_z \cdot X \cdot B)) + M(((X^T \cdot S_z \cdot X)^{-1} \cdot X^T \cdot S_z \cdot E)) = \end{aligned}$$

$$= ((X^T \cdot S_z \cdot X)^{-1} \cdot X^T \cdot S_z \cdot X) \cdot MB + 0 = I \cdot B = B$$

$$\text{cov}(\hat{B}) = M \left((\hat{B} - M\hat{B}) \cdot (\hat{B} - M\hat{B})^T \right)$$

Рассмотрим $\hat{B} - M\hat{B}$:

$$\hat{B} - M\hat{B} = ((X^T \cdot S_z \cdot X)^{-1} \cdot X^T \cdot S_z \cdot X \cdot B) + ((X^T \cdot S_z \cdot X)^{-1} \cdot X^T \cdot S_z \cdot E) - B$$

$$\text{cov}(\hat{B}) = M \left((\hat{B} - M\hat{B}) \cdot (\hat{B} - M\hat{B})^T \right)$$

Рассмотрим $\hat{B} - M\hat{B}$:

$$\hat{B} - M\hat{B} = ((X^T \cdot S_z \cdot X)^{-1} \cdot X^T \cdot S_z \cdot X \cdot B) + ((X^T \cdot S_z \cdot X)^{-1} \cdot X^T \cdot S_z \cdot E) - B$$

Тогда:

$$\begin{aligned} \text{cov}(\hat{B}) &= M \left(((X^T \cdot S_z \cdot X)^{-1} \cdot X^T \cdot S_z \cdot E) \cdot ((X^T \cdot S_z \cdot X)^{-1} \cdot X^T \cdot S_z \cdot E)^T \right) = \\ &= M \left((X^T \cdot S_z \cdot X)^{-1} \cdot X^T \cdot S_z \cdot E \cdot E^T \cdot S_z \cdot X \cdot (X^T \cdot S_z \cdot X)^{-1} \right) = \\ &= \sigma_\varepsilon^2 \cdot (X^T \cdot S_z \cdot X)^{-1} \cdot X^T \cdot S_z \cdot S_z \cdot X \cdot (X^T \cdot S_z \cdot X)^{-1} = \sigma_\varepsilon^2 \cdot (X^T \cdot S_z \cdot X)^{-1} \end{aligned}$$

Мы будем оценивать дисперсии по формуле $\text{cov}(\hat{B}) = \sigma_\varepsilon^2 \cdot (X^T \cdot S_z \cdot X)^{-1}$. Если бы была истинная модель, то в формуле бы не было элемента S_z . Нужно понять, как будут соотноситься две оценки для истинной и полученной модели. Для этого проведем ряд преобразований:

$$\begin{aligned} \text{cov}(\hat{B}) &= \sigma_\varepsilon^2 \cdot (X^T \cdot S_z \cdot X)^{-1} = \sigma_\varepsilon^2 \cdot (X^T \cdot (I - Z \cdot (Z^T \cdot Z)^{-1} \cdot Z^T) \cdot X)^{-1} = \\ &= \sigma_\varepsilon^2 \cdot (X^T \cdot X - X^T \cdot Z \cdot (Z^T \cdot Z)^{-1} \cdot Z^T \cdot X)^{-1} = \\ &= \sigma_\varepsilon^2 \cdot (X^T \cdot X - I \cdot X^T \cdot Z \cdot (Z^T \cdot Z)^{-1} \cdot Z^T \cdot X)^{-1} = \\ &= \sigma_\varepsilon^2 \cdot (X^T \cdot X - X^T \cdot X \cdot (X^T \cdot X)^{-1} \cdot X^T \cdot Z \cdot (Z^T \cdot Z)^{-1} \cdot Z^T \cdot X)^{-1} = \\ &= \sigma_\varepsilon^2 \cdot (I - (X^T \cdot X)^{-1} \cdot X^T \cdot Z \cdot (Z^T \cdot Z)^{-1} \cdot Z^T \cdot X) \cdot (X^T \cdot X)^{-1} = \dots \end{aligned}$$

Имеет место алгебраическое тождество:

$$(I + A \cdot B)^{-1} = I - A \cdot (I + B \cdot A)^{-1} \cdot B$$

Тогда:

$$\begin{aligned} \text{cov}(\hat{B}) &= \sigma_\varepsilon^2 \cdot (I - (X^T \cdot X)^{-1} \cdot X^T \cdot Z \cdot (Z^T \cdot Z)^{-1} \cdot Z^T \cdot X) \cdot (X^T \cdot X)^{-1} = \\ &= \sigma_\varepsilon^2 \cdot (I + (X^T \cdot X)^{-1} \cdot X^T \cdot Z \cdot (I - (Z^T \cdot Z)^{-1} \cdot Z^T \cdot X \cdot (X^T \cdot X)^{-1} \cdot X^T \cdot Z)^{-1} \cdot (Z^T \cdot Z)^{-1} \cdot Z^T \cdot X) \cdot (X^T \cdot X)^{-1} = \\ &= \sigma_\varepsilon^2 \cdot (I + (X^T \cdot X)^{-1} \cdot X^T \cdot Z \cdot ((Z^T \cdot Z)^{-1} \cdot Z^T \cdot Z - (Z^T \cdot Z)^{-1} \cdot Z^T \cdot X \cdot (X^T \cdot X)^{-1} \cdot X^T \cdot Z)^{-1} \cdot (Z^T \cdot Z)^{-1} \cdot Z^T \cdot X \cdot (X^T \cdot X)^{-1} = \\ &= \sigma_\varepsilon^2 \cdot (I + (X^T \cdot X)^{-1} \cdot X^T \cdot Z \cdot (Z^T \cdot Z - Z^T \cdot X \cdot (X^T \cdot X)^{-1} \cdot X^T \cdot Z)^{-1} \cdot Z^T \cdot Z \cdot (Z^T \cdot Z)^{-1} \cdot Z^T \cdot X) \cdot (X^T \cdot X)^{-1} = \\ &= \sigma_\varepsilon^2 \cdot (I + (X^T \cdot X)^{-1} \cdot X^T \cdot Z \cdot (Z^T \cdot (I - X \cdot (X^T \cdot X)^{-1} \cdot X^T) \cdot Z)^{-1} \cdot Z^T \cdot X) \cdot (X^T \cdot X)^{-1} = \dots \end{aligned}$$

Вспоминаем, что $I - X \cdot (X^T \cdot X)^{-1} \cdot X^T \equiv S$. Тогда имеем:

$$\begin{aligned} \text{cov}(\hat{B}) &= \sigma_\varepsilon^2 \cdot (I + (X^T \cdot X)^{-1} \cdot X^T \cdot Z \cdot (Z^T \cdot S \cdot Z)^{-1} \cdot Z^T \cdot X) \cdot (X^T \cdot X)^{-1} = \\ &= \sigma_\varepsilon^2 \cdot (X^T \cdot X)^{-1} + \sigma_\varepsilon^2 \cdot (X^T \cdot X)^{-1} \cdot X^T \cdot Z \cdot (Z^T \cdot S \cdot Z)^{-1} \cdot Z^T \cdot X \cdot (X^T \cdot X)^{-1} \end{aligned}$$

Первое слагаемое $\sigma_\varepsilon^2 \cdot (X^T \cdot X)^{-1}$ соответствует ковариационной матрице для «истинной» (короткой) модели. Построив длинную модель, мы прибавили дополнительное слагаемое (можно показать, что это число положительное). Тогда

$$\text{cov}(\hat{B}) > \sigma_\varepsilon^2 \cdot (X^T \cdot X)^{-1}$$

Поймем, как ведет себя истинная дисперсия остатков:

$$\sigma_\varepsilon^2 = ?$$

$$Y = X \cdot \hat{B} + Z \cdot \hat{A} + \hat{E}$$

$$Y = X * \cdot \hat{B}^* + \hat{E}$$

$$\hat{E} = Y - X * \cdot \hat{B}^*$$

Введем S^* :

$$E = S^* \cdot Y$$

$$S^* \equiv I - X^* \cdot (X^{*T} \cdot X^*)^{-1} \cdot X^{*T}$$

$$M(\hat{E}^T \cdot \hat{E}) = M(Y^T \cdot S^* \cdot Y)$$

Причем $S^* \cdot X^* = 0$.

Так как $X^* = \begin{bmatrix} X & Z \end{bmatrix}$, то:

$$S^* \cdot X^* = 0 \Rightarrow \begin{bmatrix} S^* \cdot X & S^* \cdot Z \end{bmatrix} = 0$$

Тогда $S^* \cdot X = 0$, $S^* \cdot Z = 0$. Теперь:

$$\begin{aligned} M(Y^T \cdot S^* \cdot Y) &= M((X \cdot B + E)^T \cdot S^* \cdot (X \cdot B + E)) = \\ &= M(B^T \cdot X^T \cdot S^* \cdot X \cdot B + B^T \cdot X^T \cdot S^* \cdot E + E^T \cdot S^* \cdot X \cdot B + E^T \cdot S^* \cdot E) = M(E^T \cdot S^* \cdot E) = \\ &= M(E^T \cdot I \cdot E) - M(E^T \cdot X^* \cdot (X^{*T} \cdot X^*)^{-1} \cdot X^{*T} \cdot E) = \sigma_\varepsilon^2 \cdot (n - m - 1 - k) \end{aligned}$$

В этом же вопросе было написано, что оценкой матрицы $M(\hat{E}^T \cdot \hat{E})$ является остаточная сумма квадратов $SS_{\text{ост}}$. То есть $M(\hat{E}^T \cdot \hat{E}) = SS_{\text{ост}}$. Тогда оценкой дисперсии случайных остатков является следующее выражение:

$$\sigma_\varepsilon^2 = \frac{SS_{\text{ост}}}{n - m - 1 - k}$$

Где m – количество слагаемых с переменными x , k – количество слагаемых с переменными z .

Получается, что в таком случае, оценка дисперсии случайных остатков будет корректной, то есть оценка дисперсии случайных остатков «длинной» модели является несмещенной.

Включение несущественных переменных:

1. Оценки параметров несмещенны;
2. Оценка дисперсии случайных остатков несмещенна;
3. Дисперсии оценок параметров «короткой» («реальной») модели меньше, чем дисперсии оценок параметров «длинной» модели.

2.10 Оценка существенности включения независимых переменных

Есть «короткая» и «длинная» регрессии. «Короткая» регрессия:

$$y = b_0 + b_1 \cdot x_1 + \dots + \varepsilon$$

«Длинная» регрессия:

$$y = b_1 + b_1 \cdot x_1 + \dots + a_1 \cdot z_1 + \dots + \varepsilon$$

$$H_0 : a_1 = a_2 = \dots = a_k = 0$$

$$H_1 : \exists a_j : a_j \neq 0$$

Иначе:

$$H_0 : SS_{\text{ост}(0)} - SS_{\text{ост}(a)} = 0$$

$$H_1 : SS_{\text{ост}(0)} - SS_{\text{ост}(a)} \neq 0$$

Для этого используем F-критерий:

$$F_{(0)} = \frac{SS_{\text{ост}(0)} - SS_{\text{ост}(a)}}{SS_{\text{ост}(a)}} \cdot \frac{n - m - k - 1}{k}$$

$$F_{(0)} = \frac{R_{(a)}^2 - R_{(0)}^2}{1 - R_{(a)}^2} \cdot \frac{n - m - k - 1}{k}$$

Сравниваем $F_{(0)}$ с $F_{\text{крит}} = F_{\alpha} = F(\alpha; k; n - m - k - 1)$. Если $F_{(0)} > F_{\alpha}$, то отвергаем нулевую гипотезу в пользу альтернативной – стоило включать независимые переменные.

Для $k = 1$:

$$F_{(0)} = F_{x_j} = \frac{SS_{\text{ост}(0)} - SS_{\text{ост}(x_j)}}{SS_{\text{ост}(x_j)}} \cdot \frac{n - m - 1 - 1}{1}$$

$$F_{\alpha} = F(\alpha; k; n - m - k - 1)$$

$$F_{x_j} = t_{a_j}^2$$

Пример.

$$y = -7.7 + 1.2 \cdot x_1 - 4.1 \cdot x_2 + 0.1 \cdot x_3 + \hat{\varepsilon}$$

$$SS_{\text{ост}(a)} = 295.5, \quad n = 30$$

$$y = -46.5 + 1.7 \cdot x_1 + \hat{\varepsilon}$$

$$SS_{\text{ост}(a)} = 701.0$$

$$F_{(0)} = \frac{701 - 295.5}{295.5} \cdot \frac{30 - 1 - 2 - 1}{2} = 17.8$$

$$F_{\alpha} = F(0.05; 2; 26) = 3.37$$

Так как $F_{(0)} > F_{\alpha}$, то включение x_2 и x_3 является статистически значимым.

2.11 Тест на функциональную форму

H_0 : линейная модель $y = b_0 + b_1 \cdot x + \dots + \varepsilon$

H_1 : нелинейная модель

RESET-тест Рамсея – строим регрессию:

$$y = b_0 + b_1 \cdot x_1 + \dots + c_1 \cdot \hat{y}^2 + c_2 \cdot \hat{y}^3 + \dots + \varepsilon$$

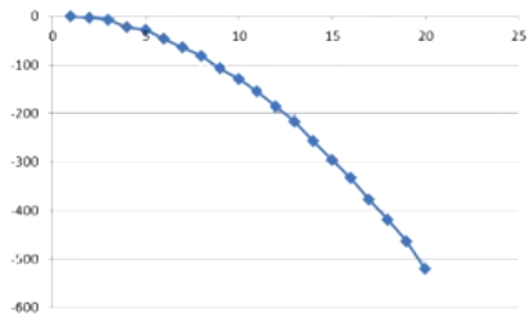
Тогда нулевую и альтернативную гипотезу можно переформулировать следующим образом:

$$H_0 : c_1 = c_2 = \dots = 0$$

$$H_1 : \exists c_j : c_j \neq 0$$

Для проверки нулевой и альтернативной гипотезы используется, по аналогии с предыдущим вопросом, F-критерий для сравнения «короткой» (линейной модели) и «длинной» (нелинейной модели).

$F > F_\alpha$ – отвергаем нулевую гипотезу, функциональная форма нелинейна с вероятностью 0.95. Пример.



$$y = 101.4 - 27.4 \cdot x + \hat{\varepsilon}$$

$$SS_{\text{ост}} = 28769.7$$

$$y = 20.0 - 9.8 \cdot x - 0.002 \cdot \hat{y}^2 - 1.04 \cdot 10^{-7} \cdot \hat{y}^3 + \hat{\varepsilon}$$

$$SS_{\text{ост}} = 89.7$$

$$F = \frac{28769.7 - 89.7}{89.7} \cdot \frac{20 - 1 - 2 - 1}{2} = 2557.9$$

$$F_\alpha = F(0.05; 2; 16) = 3.63$$

Таким образом, функциональная форма является нелинейной с вероятностью 0.95.

2.12 Выбор наилучшей функции регрессии (тренда)

Критирии выбора:

1. Экономическая обоснованность;
2. Качество случайных остатков;
3. Значимость;
4. Теснота связи (качество аппроксимации);

5. Простота.

Качество аппроксимации:

$$R^2$$
$$R_{\text{скорр}}^2 = 1 - (1 - R^2) \cdot \frac{n - 1}{n - m - 1}$$

Информационные критерии.

Критерий Акаике:

$$AIC = \ln \left(\frac{SS_{\text{ост}}}{n} \right) + \frac{2 \cdot (m + 1)}{n} + \ln(2 \cdot \pi)$$

Критерий Шварца:

$$SC = \ln \left(\frac{SS_{\text{ост}}}{n} \right) + \frac{(m + 1) \cdot \ln n}{n} + \ln(2 \cdot \pi)$$

Оба критерия, чем меньше - тем лучше.

Средняя ошибка аппроксимации:

$$MAPE = \bar{A} = \frac{1}{n} \cdot \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \cdot 100\%$$

Стандартная ошибка регрессии (оценка дисперсии остатков):

$$\sigma_{\varepsilon} = SEE = \sqrt{\frac{SS_{\text{ост}}}{n - m - 1}}$$

Средняя квадратическая ошибка расчетного значения (RMSE – root mean squared error):

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}}$$

Среднее значение модуля ошибки (MAE – mean absolute error):

$$MAE = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{n}$$