Тема 2. Регрессия МНК с одной объясняющей переменной.

Основные понятия и задачи этой части курса.

Уравнение регрессии представляет собой линейную зависимость объясняемой переменной (regressand) от объясняющих переменных (regressors). Для простоты мы сначала рассмотрим случай одной объясняющей переменной, тогда уравнение теоретической регрессии будет иметь вид:

(2.01) $$Y_i = \alpha + \beta \cdot X_i + \varepsilon_i$$

где α, β - параметры Y_i-объясняемая переменная, X_i - объясняющая переменная , ε_i - случайная ошибка, которая включает в себя не учтенные в модели факторы, которые оказывают влияние на Y_i. (2.01) называется уравнением теоретической регрессии, так как оно описывает зависимость, которая согласно нашим предположениям, имеет место в генеральной совокупности. То есть мы предполагаем, что изменение переменной Y_i можно представить, как сумму линейной функции зависящей от X_i и ряда несущественных или ненаблюдаемых факторов, которые представлены в модели в виде случайной величины ε_i.

Основные задачи, которые мы должны научиться решать, изучив эту главу, следующие. Во-первых, по имеющейся у нас выборке размера N, состоящей из Y_i и X_i $i = 1.....N$ нужно получить оценки неизвестных параметров α, β (мы будем обозначать их $\hat{\alpha}, \hat{\beta}$). Во-вторых, научиться проверять гипотезу о том, что модель (2.1) адекватно описывает наши данные, и гипотезы о значениях параметров α, β. В-третьих, научиться строить доверительные интервалы для истинных значений параметров α, β и для прогноза полученного по оцененной нами модели.

Для оценки параметров α и β в уравнении (2.01) мы будем использовать линейный метод наименьших квадратов. Обычно слово линейный опускают и называют его метод наименьших квадратов сокращенно МНК, в англо язычной литературе он называется (ordinary least squares сокращенно OLS). Линейным метод называется, так как уравнение (2.01) линейно по параметрам α и β.

Суть МНК состоит в том, что оценки $\hat{\alpha}$ и $\hat{\beta}$ выбираются так, чтобы сумма квадратов остатков минимальна:

(2.02) $$\sum_{i=1}^{N}\left(Y_i - \hat{\alpha} - \hat{\beta} \cdot X_i\right)^2 = \sum_{i=1}^{N}\left(Y_i - \hat{Y}_i\right)^2 = \sum_{i=1}^{N} e_i^2 \Rightarrow \min_{\hat{\alpha},\hat{\beta}}$$

Как видно из (2.02) под остатками $e_i = Y_i - \hat{Y}_i$ понимается разница между наблюдаемым нами значением объясняемой переменной Y_i и ее предсказанным значением $\hat{Y}_i = \hat{\alpha} + \hat{\beta} \cdot X_i$.

МНК является одним из трех основных методов, которые используются для оценки эконометрических уравнений. Другие два, это метод максимального правдоподобия и метод моментов. Широкое использование МНК объясняется тем, что полученные с его

помощью оценки обладают рядом хороших свойств. Рассмотрим эти свойства более подробно.

Математические свойства оценок МНК.

Свойства, которые мы сейчас рассмотрим, вытекают непосредственно из формул, по которым вычисляются оценки МНК. Поэтому для их выполнения не важно откуда мы получили X и Y, по которым мы оцениваем регрессию. То есть X и Y могут быть как просто набором детерминированных чисел, так и выборкой из случайных величин с произвольной функцией распределения.

Выведем формулы оценок МНК, для этого нужно решить задачу (2.02). Необходимые условия будут иметь вид:

$$(2.03)\quad \begin{cases} \dfrac{\partial \sum_{i=1}^{N}\left(Y_i-\widehat{\alpha}-\widehat{\beta}\cdot X_i\right)^2}{\partial \widehat{\alpha}} = -2\cdot\sum_{i=1}^{N}\left(Y_i-\widehat{\alpha}-\widehat{\beta}\cdot X_i\right)=\sum_{i=1}^{N}e_i=0 \\ \dfrac{\partial \sum_{i=1}^{N}\left(Y_i-\widehat{\alpha}-\widehat{\beta}\cdot X_i\right)^2}{\partial \widehat{\beta}} = -2\cdot X_i\sum_{i=1}^{N}\left(Y_i-\widehat{\alpha}-\widehat{\beta}\cdot X_i\right)=\sum_{i=1}^{N}X_i\cdot e_i=0 \end{cases}$$

(2.03) обычно называют системой нормальных уравнений. Решая эту систему, мы получаем формулы оценок коэффициентов

$$(2.04)\quad \begin{cases} \widehat{\beta}=\dfrac{\sum_{i=1}^{N}X_i\cdot Y_i-N\cdot\overline{X}\cdot\overline{Y}}{\sum_{i=1}^{N}X_i^2-N\cdot\overline{X}^2}=\dfrac{\sum_{i=1}^{N}\left(X_i-\overline{X}\right)\cdot\left(Y_i-\overline{Y}\right)}{\sum_{i=1}^{N}\left(X_i-\overline{X}\right)^2}=\dfrac{\sum x_i\cdot y_i}{\sum x_i^2}=\dfrac{\operatorname{cov}(X,Y)}{D(X)} \\ \widehat{\alpha}=\overline{Y}-\widehat{\beta}\cdot\overline{X} \end{cases}$$

где $y_i=Y_i-\overline{Y}\quad x_i=X_i-\overline{X}$.

Чтобы убедится, что найденная нами точка экстремума суммы квадратов ошибок является именно точкой минимума, а не максимума проверим достаточные условия, для этого вычислим вторые производные и проверим положительную определенность матрицы Гессе.

$$\frac{\partial^2\sum_{i=1}^{N}\left(Y_i-\widehat{\alpha}-\widehat{\beta}\cdot X_i\right)^2}{\partial\widehat{\alpha}^2}=2\cdot N;\qquad \frac{\partial^2\sum_{i=1}^{N}\left(Y_i-\widehat{\alpha}-\widehat{\beta}\cdot X_i\right)^2}{\partial\widehat{\beta}^2}=2\sum_{i=1}^{N}X_i^2;$$

$$\frac{\partial^2\sum_{i=1}^{N}\left(Y_i-\widehat{\alpha}-\widehat{\beta}\cdot X_i\right)^2}{\partial\widehat{\alpha}\cdot\partial\widehat{\beta}}=2\sum_{i=1}^{N}X_i$$

$$H=\begin{Bmatrix} 2\cdot N & 2\sum_{i=1}^{N}X_i \\ 2\sum_{i=1}^{N}X_i & 2\sum_{i=1}^{N}X_i^2 \end{Bmatrix}$$

$$2\cdot N\ge 0$$

$$\det H=4\cdot\left(N\cdot\sum_{i=1}^{N}X_i^2-\left(\sum_{i=1}^{N}X_i\right)^2\right)=4\cdot N\left(\sum_{i=1}^{N}X_i^2-N\cdot\overline{X}^2\right)=4\cdot N\sum_{i=1}^{N}\left(X_i-\overline{X}\right)^2\ge 0$$

Матрица Гессе положительно определена, мы действительно нашли минимум.

Используя оценки $\widehat{\alpha}$ и $\widehat{\beta}$, можно получить вектор МНК оценок объясняемой переменной или предсказанных значений:

(2.05) $$\widehat{Y}_i = \widehat{\alpha} + \widehat{\beta} \cdot X_i$$

и вектор остатков:

(2.06) $$e_i = Y_i - \widehat{Y}_i = Y_i - \widehat{\alpha} - \widehat{\beta} \cdot X_i$$

Уравнение (2.05) называется уравнением выборочной регрессии, так как в отличие от уравнения (2.01) в нем ~~все с крышками~~ используются не неизвестные нам значения параметров α и β, а полученные по выборке оценки $\widehat{\alpha}$ и $\widehat{\beta}$. Иногда уравнение выборочной регрессией записывают как $Y_i = \widehat{\alpha} + \widehat{\beta} \cdot X_i + e_i$.

Непосредственно из (2.04), (2.05) и (2.06) вытекают следующие свойства оценок МНК.

1) $\sum_{i=1}^{N} e_i = 0$ из (2.03)

2) $\sum_{i=1}^{N} X_i \cdot e_i = 0$ из (2.03)

3) $\sum_{i=1}^{N} \widehat{Y}_i \cdot e_i = 0$ из (2.05) и свойств 1 и 2.

4) $\frac{\sum_{i=1}^{N} \widehat{Y}_i}{N} = \frac{\sum_{i=1}^{N} Y_i}{N} = \overline{Y}$ - предсказанное среднее и выборочное среднее равны из (2.06) и свойства 1.

5) $\overline{Y} = \widehat{\alpha} + \widehat{\beta} \cdot \overline{X}$ то есть линия регрессии $\widehat{Y}_i = \widehat{\alpha} + \widehat{\beta} \cdot X_i$ всегда проходит через точку $\left(\overline{X}, \overline{Y}\right)$ из (2.04).

6) Сумма квадратов отклонений от среднего объясняемой переменной может быть представлена как:

(2.07) $$\sum_{i=1}^{N}\left(Y_i - \overline{Y}\right)^2 = \sum_{i=1}^{N}\left(\widehat{Y}_i - \overline{Y}\right)^2 + \sum_{i=1}^{N} e_i^2$$

из (2.06) и свойств 3 и 4.

Внимание! если из уравнения регрессии (2.05) исключить коэффициент α, то есть оценивать модель $\widehat{Y}_i = \widehat{\beta} \cdot X_i$, то справедливыми останутся только свойства 2 и 3. Все остальные будут не верны!

Задача на дом №1. Найдите формулу МНК оценки $\widehat{\beta}$ для модели $\widehat{Y}_i = \widehat{\beta} \cdot X_i$ и проверьте, что для этой модели выполняются только свойства 2 и 3.

Рассмотрим более подробно свойство 6, иногда его называют геометрической теоремой Пифагора. Сначала поймем, как его доказать. Из (2.06) $Y_i = \widehat{Y}_i + e_i$, соответственно $Y_i^2 = \widehat{Y}_i^2 + 2\widehat{Y} \cdot e_i + e_i^2$, суммируя по N, получаем:

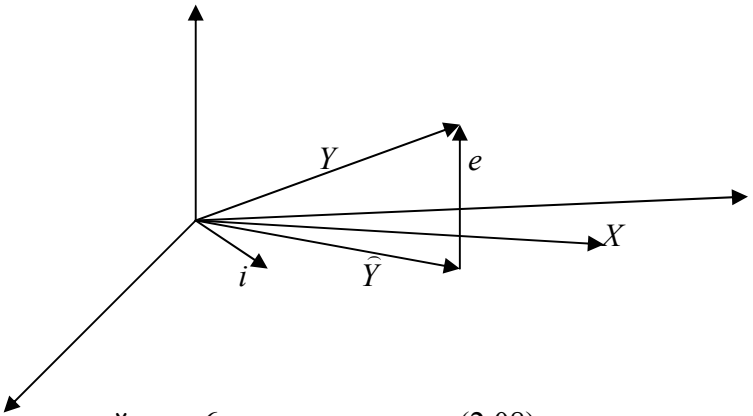
(2.08) $$\sum_{i=1}^{N} Y_i^2 = \sum_{i=1}^{N} \widehat{Y}_i^2 + \sum_{i=1}^{N} e_i^2$$

так как $\sum_{i=1}^{N} \widehat{Y}_i \cdot e_i = 0$. Рассмотрим, что это означает с точки зрения геометрии. У нас есть N-мерные векторы

$$Y=\begin{pmatrix} Y_1 \\ Y_i \\ Y_N \end{pmatrix} \hat{Y}=\begin{pmatrix} \hat{Y}_1 \\ \hat{Y}_i \\ \hat{Y}_N \end{pmatrix} X=\begin{pmatrix} X_1 \\ X_i \\ X_N \end{pmatrix} e=\begin{pmatrix} e_1 \\ e_i \\ e_N \end{pmatrix} i=\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

Из 2.05 мы знаем, что $\hat{Y}$ является линейной комбинацией векторов X и i: $\hat{Y}=\hat{\alpha}\cdot i+\hat{\beta}\cdot X$ поэтому вектор $\hat{Y}$ лежит в той же плоскости, что и вектора X и i. Так же знаем, что скалярные произведения векторов $\langle i,e\rangle=0$, так как $\sum_{i=1}^{N}e_i=0$, $\langle X,e\rangle=0$, так как $\sum_{i=1}^{N}X_i\cdot e_i=0$, значит вектор e ортогонален (или перпендикулярен) векторам X и i, а соответственно и плоскости, в которой эти два вектора лежат. Поэтому он также ортогонален вектору $\hat{Y}$, в этом и состоит свойство 3: $\sum_{i=1}^{N}\hat{Y}_i\cdot e_i=0$, то есть $\langle\hat{Y},e\rangle=0$. В свою очередь $Y=\hat{Y}+e$, поэтому вектора $Y,\hat{Y},e$ образуют прямоугольный треугольник и для их длин выполняется теорема Пифагора, что и записано в (2.08). (Если вы еще помните, длинна вектора это корень из суммы квадратов его компонент). Вектор $\hat{Y}$ является проекцией вектора Y на плоскость образованную векторами X и i. Соответственно вектор e это нормаль (перпендикуляр) проведенная к этой плоскости. Поэтому (2.03) и называют системой нормальных уравнений. Графически это изображено на рисунке 1.

Он мне все равно не нравится, хотя я его и долго рисовал, может попробую потом еще переделать его. У вас не должно складывается впечатление, что мы строим проекцию из трех мерного пространства в двумерное. Дело в том, что плоскость образованная векторами X и i никогда не будет совпадать с плоскостью образованной двумя осями координат, как у меня здесь получилось, да и пространство должно быть N-мерное.



Чтобы доказать свойство 6 нам осталось в (2.08) от суммы квадратов перейти к сумме квадратов отклонений от среднего. Так как $\sum_{i=1}^{N}e_i=0$ и средние значения наблюдаемых и предсказанных игреков у нас совпадают, то из (2.08) получаем: $\sum_{i=1}^{N}Y_i^2+N\cdot\overline{Y}^2=\sum_{i=1}^{N}\hat{Y}_i^2+N\cdot\overline{Y}^2-2\sum_{i=1}^{N}e_i+\sum_{i=1}^{N}e_i^2$,

$\sum_{i=1}^{N} Y_i^2 + N\cdot\overline{Y}^2 = \sum_{i=1}^{N}\widehat{Y}_i^2 + N\cdot\overline{Y}^2 - 2\sum_{i=1}^{N}\left(Y_i-\widehat{Y}_i\right)+\sum_{i=1}^{N} e_i^2$ выделяя нужные нам полные квадраты разностей $\sum_{i=1}^{N}\left(Y_i-\overline{Y}\right)^2 = \sum_{i=1}^{N}\left(\widehat{Y}_i-\overline{Y}\right)^2 + \sum_{i=1}^{N} e_i^2$ - это и есть свойство 6, которое мы доказывали.

Равенство (2.07) часто записывают, как $TSS = ESS + RSS$, где $TSS = \sum_{i=1}^{N}\left(Y_i-\overline{Y}\right)^2$ - total sum of squares, $ESS = \sum_{i=1}^{N}\left(\widehat{Y}_i-\overline{Y}\right)^2$ -explained sum of squares, $RSS = \sum_{i=1}^{N} e_i^2$ residual sum of squares. Некоторые полезные утверждение относительно $TSS, ESS, RSS, \widehat{\beta}$. Перейдем к $y_i = Y_i - \overline{Y}$, $x_i = X_i - \overline{X}$, тогда $TSS = \sum_{i=1}^{N} y_i^2$, $\widehat{\beta} = \frac{\sum x_i\cdot y_i}{\sum x_i^2}$,

$$RSS = \sum_{i=1}^{N} e_i^2 = \sum_{i=1}^{N}\left(Y_i - \widehat{\alpha} - \widehat{\beta}\cdot X_i\right)^2 = \left\{\widehat{\alpha} = \overline{Y} - \widehat{\beta}\cdot\overline{X}\right\} = \sum_{i=1}^{N}\left(Y_i - \overline{Y} + \widehat{\beta}\cdot\overline{X} - \widehat{\beta}\cdot X_i\right)^2 =$$

(2.09) $$= \sum_{i=1}^{N}\left(y_i - \widehat{\beta}\cdot x_i\right)^2 = \sum_{i=1}^{N} y_i^2 - 2\widehat{\beta}\cdot\sum_{i=1}^{N} y_i x_i + \widehat{\beta}^2\sum_{i=1}^{N} x_i^2 = \sum_{i=1}^{N} y_i^2 - 2\frac{\sum x_i\cdot y_i}{\sum x_i^2}\sum_{i=1}^{N} y_i x_i +$$

$$+\left(\frac{\sum x_i\cdot y_i}{\sum x_i^2}\right)^2\cdot\sum_{i=1}^{N} x_i^2 = \sum_{i=1}^{N} y_i^2 - \frac{\left(\sum x_i\cdot y_i\right)^2}{\sum x_i^2} = \sum_{i=1}^{N} y_i^2 - \widehat{\beta}\cdot\sum_{i=1}^{N} y_i x_i$$

(2.10) $$ESS = TSS - RSS = \sum_{i=1}^{N} y_i^2 - \sum_{i=1}^{N} y_i^2 + \widehat{\beta}\cdot\sum_{i=1}^{N} y_i x_i = \widehat{\beta}\cdot\sum_{i=1}^{N} y_i x_i$$

Коэффициент R^2 определяется как

(2.11) $$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$$

R^2 является мерой качества регрессии, он показывает долю общей дисперсии переменной Y, которую удалось объяснить с помощью переменной X. Он так же называется коэффициентом детерминации. Коэффициент детерминации показывает степень тесноты статистической связи между результирующим показателем и набором объясняющих факторов. R^2-это формула его вычисления в рассматриваемом нами случае, когда зависимость линейная. Коэффициент R^2 по определению меняется от 0 до 1.

Подставляя (2.10) и (2.04) в (2.11) получаем:

(2.12) $$R^2 = \frac{ESS}{TSS} = \widehat{\beta}\cdot\frac{\sum_{i=1}^{N} y_i x_i}{\sum_{i=1}^{N} y_i^2} = \frac{\left(\sum_{i=1}^{N} y_i x_i\right)^2}{\sum_{i=1}^{N} x_i^2\sum_{i=1}^{N} y_i^2} = r^2\left(X,Y\right),$$

где $r\left(Y,X\right) = \frac{\sum\left(X-\overline{X}\right)\cdot\left(Y-\overline{Y}\right)}{\sqrt{\sum\left(Y-\overline{Y}\right)^2\cdot\sum\left(X-\overline{X}\right)^2}}$-выборочный коэффициент корреляции между переменными X, Y. Так как $\widehat{Y}_i = \widehat{\alpha} + \widehat{\beta}\cdot X_i$, то $r\left(\widehat{Y},Y\right) = \frac{\widehat{\beta}}{\left|\widehat{\beta}\right|}\cdot r\left(X,Y\right)$. Соответственно

(2.13) $$R^2 = r^2\left(\widehat{Y},Y\right) = r^2\left(X,Y\right)$$

Задачка на дом №2. Покажите, что $r\left(\widehat{Y},Y\right) = +\sqrt{R^2}$, $r\left(X,Y\right) = \begin{cases} +\sqrt{R^2} & \widehat{\beta} > 0 \\ -\sqrt{R^2} & \widehat{\beta} < 0 \end{cases}$

Статистические свойства оценок МНК. Теорема Гаусса-Маркова.

В этой главе мы посмотрим, какими свойствами обладают МНК оценки модели регрессии, когда по имеющейся у нас выборке размера N мы оцениваем параметры существующей в генеральной совокупности линейной зависимости:

(2.14) $$Y_i = \alpha + \beta \cdot X_i + \varepsilon_i$$

Я начну с того, что приведу основной результат. **Теорема Гаусса-Маркова.**

Если

1) Модель правильно специфицирована, то есть зависимость вида (2.14) действительно существует.

2) X_i - являются детерминированными величинами и не равны все между собой.

3) ε_i -случайные величины причем

3.1) $E(\varepsilon_i) = 0, \forall i$

3.2) $\text{var}(\varepsilon_i) = \sigma_\varepsilon^2, \forall i$ (гомоскедастичность)

3.3) $\text{cov}(\varepsilon_i \varepsilon_j) = 0 \ \forall i \neq j$ (не коррелированность).

Тогда МНК оценки $\hat{\alpha}, \hat{\beta}$ являются

1) линейными по Y
2) несмещенными
3) эффективными

оценками параметров α, β

Весь наш дальнейший курс будет в основном состоять в том, что мы обобщим полученные результаты на случай нескольких объясняющих переменных, а потом будем смотреть, что изменится, если какая-нибудь из предпосылок теоремы Гаусса-Маркова не выполняется.

Рассмотрим, какими свойствами будет обладать случайная величина Y, если приведенные выше предположения справедливы.

1) $E(Y|X_i) = \alpha + \beta \cdot X_i$, так как $E(\varepsilon_i) = 0$. В модели линейной регрессии предполагается, что математическое ожидание Y зависит от X. Хотя X и не является случайной величиной, обычно пишут условное математическое ожидание. Это делается так потому, что в отличии от α и β, значение параметра X нам известно. Если отбросить предпосылку о том, что X является детерминированной величиной и предположить, что Y и X имеют двумерное нормальное распределение, тогда запись $E(Y|X_i) = \alpha + \beta \cdot X_i$ будет более уместна.

2) $\text{var}(Y|X) = \text{var}(Y) = \sigma_\varepsilon^2$, так как $\text{var}(\varepsilon_i) = \sigma_\varepsilon^2, \forall i$

3) $\text{cov}(Y_i Y_j) = 0 \ \forall i \neq j$, так как $\text{cov}(\varepsilon_i \varepsilon_j) = 0 \ \forall i \neq j$

Используя выборку из случайной величины Y и соответствующие ей значения детерминированной величины X, мы можем вычислить МНК оценки $\hat{\alpha}, \hat{\beta}$ и векторы оценок $\hat{Y}, e$. Все они являются функциями от выборочных значений случайной величины Y, поэтому тоже будут случайными величинами. Сейчас мы рассмотрим свойства этих четырех оценок, и при этом докажем теорему Гаусса-Маркова.

Вспомним, как вычисляется $\widehat{\beta}$ (формула 2.04) $\widehat{\beta}=\dfrac{\sum_{i=1}^{N}\left(X_i-\bar{X}\right)\cdot\left(Y_i-\bar{Y}\right)}{\sum_{i=1}^{N}\left(X_i-\bar{X}\right)^2}$, обозначим $w_i=\dfrac{\left(X_i-\bar{X}\right)}{\sum_{i=1}^{N}\left(X_i-\bar{X}\right)^2}=\dfrac{x_i}{\sum x_i^2}$, и обратим внимание, что

(2.15) $$\sum_{i=1}^{N}w_i=0;\ \sum_{i=1}^{N}w_i\cdot x_i=\sum_{i=1}^{N}w_i\cdot X_i=1\ \ \sum_{i=1}^{N}w_i^2=\frac{1}{\sum_{i=1}^{N}x_i^2},$$

соответственно $\widehat{\beta}=\sum_{i=1}^{N}w_i\cdot\left(Y_i-\bar{Y}\right)=\sum_{i=1}^{N}w_i\cdot Y_i-\bar{Y}\cdot\sum_{i=1}^{N}w_i=\sum_{i=1}^{N}w_i\cdot Y_i$ - линейность по Y $\widehat{\beta}$ доказана.

$E\left(\widehat{\beta}\right)=E\left(\sum_{i=1}^{N}w_i\cdot Y_i\right)=\sum_{i=1}^{N}w_i\cdot E\left(Y_i\right)=\sum_{i=1}^{N}w_i\cdot\left(\alpha+\beta\cdot X_i\right)=\alpha\sum_{i=1}^{N}w_i+\beta\cdot\sum_{i=1}^{N}w_i\cdot X_i=0+\beta\cdot1=\beta$ - несмещенность $\widehat{\beta}$ доказана.

Докажем эффективность.

(2.16) $$\operatorname{var}\left(\widehat{\beta}\right)=\operatorname{var}\left(\sum_{i=1}^{N}w_i\cdot Y_i\right)=\sum_{i=1}^{N}w_i^2\cdot\operatorname{var}\left(Y_i\right)=\sigma_\varepsilon^2\cdot\sum_{i=1}^{N}w_i^2=\frac{\sigma_\varepsilon^2}{\sum_{i=1}^{N}x_i^2}$$

пусть $b=\sum v_i\cdot Y_i$ - произвольная линейная несмещенная оценка параметра β найдем веса v_i, так чтобы ее дисперсия была минимальна, то есть чтобы оценка была эффективной в классе линейных несмещенных оценок. Если при этом мы получим, что v_i совпадают с w_i, значит $\widehat{\beta}$ - и есть единственная эффективная оценка β.

$E\left(b\right)=E\left(\sum_{i=1}^{N}v_i\cdot Y_i\right)=\sum_{i=1}^{N}v_i\cdot E\left(Y_i\right)=\sum_{i=1}^{N}w_i\cdot\left(\alpha+\beta\cdot X_i\right)=\alpha\sum_{i=1}^{N}v_i+\beta\cdot\sum_{i=1}^{N}v_i\cdot X_i=\beta$ получаем ограничения необходимые для того, чтобы оценка была несмещенной

$$\begin{cases}\sum_{i=1}^{N}v_i=0\\ \sum_{i=1}^{N}v_i\cdot X_i=1\end{cases}$$

$\operatorname{var}\left(b\right)=\operatorname{var}\left(\sum_{i=1}^{N}v_i\cdot Y_i\right)=\sum_{i=1}^{N}v_i^2\cdot\operatorname{var}\left(Y_i\right)=\sigma_\varepsilon^2\cdot\sum_{i=1}^{N}v_i^2$, (так как $\operatorname{cov}\left(Y_iY_j\right)=0$).

Задача поиска оценки с минимальной дисперсией будет иметь вид:

(2.17) $$\begin{cases}\sum_{i=1}^{N}v_i^2\Rightarrow\min\\ \sum_{i=1}^{N}v_i=0\\ \sum_{i=1}^{N}v_i\cdot X_i=1\end{cases}$$

Функция Лагранжа задачи (2.17) имеет вид: $L=\sum_{i=1}^{N}v_i^2-\lambda\cdot\sum_{i=1}^{N}v_i+\mu\cdot\left(1-\sum_{i=1}^{N}v_i\cdot X_i\right)$, где λ,μ множители Лагранжа. Берем производную по v_i:

(2.18) $$\frac{\partial L}{\partial v_i}=2\cdot v_i-\lambda-\mu\cdot X_i=0$$

суммируем условия первого порядка (2.18): $\sum\frac{\partial L}{\partial v_i}=2\cdot\sum v_i-N\cdot\lambda-\mu\cdot\sum X_i=0$, так как $\sum_{i=1}^{N}v_i=0$, то $-N\cdot\lambda-\mu\cdot\sum X_i=0\Rightarrow\lambda=-\mu\cdot\bar{X}$, подставляя в условие первого порядка (2.18) получаем:

(2.19) $$\frac{\partial L}{\partial v_i}=2\cdot v_i-\mu\cdot\left(X_i-\bar{X}\right)=0$$

Умножая (2.19) на X_i получаем $2\cdot v_i\cdot X_i-\mu\cdot\left(X_i-\bar{X}\right)\cdot X_i=0$, суммируем по всем i: $2\cdot\sum v_i\cdot X_i-\mu\cdot\sum\left(X_i-\bar{X}\right)\cdot X_i=0$, вспомним, что $\sum_{i=1}^{N}v_i\cdot X_i=1$ и что $\sum\left(X_i-\bar{X}\right)\cdot X_i=\sum\left(X_i-\bar{X}\right)\cdot X_i-\bar{X}\cdot\sum\left(X_i-\bar{X}\right)=\sum\left(X_i-\bar{X}\right)^2=\sum x_i^2$ поэтому $2\cdot1-\mu\cdot\sum x_i^2=0\Rightarrow\mu=\frac{2}{\sum x_i^2}$, подставляем это в (2.19) получаем $\frac{\partial L}{\partial v_i}=2\cdot v_i-\frac{2\cdot x_i}{\sum x_i^2}=0\Rightarrow v_i=\frac{x_i}{\sum x_i^2}$ $b=\sum v_i\cdot Y_i=\frac{\sum x_i\cdot Y_i}{\sum x_i^2}=\hat{\beta}$ - решая задачу поиска эффективной оценки параметра β в классе линейных несмещенных оценок мы получили формулу МНК оценки $\hat{\beta}$ - значит эта оценка и является единственной эффективной оценкой β в этом классе. Эффективность доказали. Хотя особо умные студенты могут возразить, что мы не можем утверждать, что, решая задачу поиска условного экстремума, мы нашли именно минимум дисперсии, а не максимум.

Задачка на дом №3. Предлагается проверить достаточные условия минимума в задаче (2.17). Надо будет выписать окаймленный Гессиан.

Разберемся теперь с $\hat{\alpha}$ по формуле (2.04) $\hat{\alpha}=\bar{Y}-\hat{\beta}\cdot\bar{X}=\frac{\sum Y_i}{N}-\bar{X}\cdot\sum_{i=1}^{N}w_i\cdot Y_i=\sum\left(\frac{1}{N}-\bar{X}\cdot w_i\right)\cdot Y_i$ - линейность по Y доказана.

$$E\left(\hat{\alpha}\right)=E\left(\sum\left(\frac{1}{N}-\bar{X}\cdot w_i\right)\cdot Y_i\right)=\sum\left(\frac{1}{N}-\bar{X}\cdot w_i\right)\cdot E\left(Y_i\right)=\sum\left(\frac{1}{N}-\bar{X}\cdot w_i\right)\cdot\left(\alpha+\beta\cdot X_i\right)=$$

$$=\sum\left(\frac{1}{N}-\bar{X}\cdot w_i\right)\cdot\left(\alpha+\beta\cdot X_i\right)=\alpha-\alpha\cdot\bar{X}\cdot\sum w_i+\beta\cdot\frac{\sum X_i}{N}-\beta\cdot\bar{X}\cdot\sum w_i\cdot X_i$$

используя соотношения (2.15), получаем $E\left(\hat{\alpha}\right)=\alpha-\alpha\cdot\bar{X}\cdot0+\beta\cdot\bar{X}-\beta\cdot\bar{X}\cdot1=\alpha$, несмещенность доказали.

$$\text{(2.20)}\qquad \begin{aligned}\operatorname{var}(\hat{\alpha}) &= \operatorname{var}\left(\sum\left(\frac{1}{N}-\bar{X}\cdot w_i\right)\cdot Y_i\right)=\sum\left(\frac{1}{N}-\bar{X}\cdot w_i\right)^2\cdot\operatorname{var}(Y_i)= \\ &= \sigma_\varepsilon^2\cdot\sum\left(\frac{1}{N}-\bar{X}\cdot w_i\right)^2=\sigma_\varepsilon^2\cdot\left(\frac{N}{N^2}-\frac{2\cdot\bar{X}\cdot\sum w_i}{N}+\bar{X}^2\cdot\sum w_i^2\right)= \\ &= \left(\frac{1}{N}+\frac{\bar{X}^2}{\sum x_i^2}\right)\cdot\sigma_\varepsilon^2\end{aligned}$$

для вывода формулы дисперсии использовались соотношения (2.15) и некоррелированность Y_i, Y_j для $i \neq j$.

Задачка на дом №4. Докажите эффективность $\hat{\alpha}$. Это делается точно также, как мы только что делали для $\hat{\beta}$.

И так мы доказали, что при выполнении условий перечисленных в теореме Гаусса-Маркова $\hat{\alpha}$ и $\hat{\beta}$ являются линейными по Y, несмещенными и эффективными оценками параметров α и β.

Возникает вопрос, являются ли они состоятельными. Ответ: да. Доказать это можно например, проверив, что они совпадают с оценками метода максимального правдоподобия, которые являются состоятельными.

Обобщенная теорема Гаусса-Маркова. Предпосылки теоремы остаются без изменений, а утверждение будет звучать так: для любых чисел c, d оценка $\hat{\theta} = c\cdot\hat{\alpha} + d\cdot\hat{\beta}$, где $\hat{\alpha}$ и $\hat{\beta}$ МНК оценки параметров α и β, будет линейной, несмещенной и эффективной оценкой параметра $\theta = c\cdot\alpha + d\cdot\beta$. Линейность и несмещенность $\hat{\theta}$, очевидно следуют из линейности и несмещенности $\hat{\alpha}$ и $\hat{\beta}$.

Задачка на дом №5. Покажите, что $\hat{\theta} = c\cdot\hat{\alpha} + d\cdot\hat{\beta}$ является линейной и несмещенной оценкой параметра $\theta = c\cdot\alpha + d\cdot\beta$.

Эффективность можно доказать аналогично тому, как мы это делали для $\hat{\beta}$. Но мы лучше сделаем это позже, когда будем рассматривать многомерную регрессию, так как там в векторно-матричной форме это все получается гораздо проще и компактнее. Пока просто посмотрим, чему равна дисперсия $\hat{\theta}$.

$$\text{(2.21)}\quad \operatorname{var}(\hat{\theta}) = \operatorname{var}\left(c\cdot\hat{\alpha}+d\cdot\hat{\beta}\right) = c^2\cdot\operatorname{var}(\hat{\alpha})+d^2\cdot\operatorname{var}(\hat{\beta})+2\cdot c\cdot d\cdot\operatorname{cov}(\hat{\alpha},\hat{\beta}),$$

$\operatorname{var}(\hat{\alpha})$, $\operatorname{var}(\hat{\beta})$- мы знаем (формулы 2.16 2.20), осталось вычислить $\operatorname{cov}(\hat{\alpha},\hat{\beta})$:

$$\text{(2.22)}\qquad \begin{aligned}\operatorname{cov}(\hat{\alpha},\hat{\beta}) &= \operatorname{cov}\left(\sum_{i=1}^N w_i\cdot Y_i;\sum\left(\frac{1}{N}-\bar{X}\cdot w_i\right)\cdot Y_i\right)=\sum_{i=1}^N\sum_{j=1}^N\operatorname{cov}\left(w_i\cdot Y_i;\left(\frac{1}{N}-\bar{X}\cdot w_j\right)\cdot Y_j\right)= \\ &= \sum_{i=1}^N\sum_{j=1}^N w_i\cdot\left(\frac{1}{N}-\bar{X}\cdot w_j\right)\operatorname{cov}(Y_i;Y_j)=\sigma_\varepsilon^2\sum_{i=1}^N w_i\cdot\left(\frac{1}{N}-\bar{X}\cdot w_i\right)=-\sigma_\varepsilon^2\cdot\bar{X}\sum_{i=1}^N w_i^2=-\frac{\sigma_\varepsilon^2\cdot\bar{X}}{\sum x_i^2}\end{aligned}$$

В доказательстве (2.22) я воспользовался свойствами (2.15) и тем, что $\operatorname{cov}(Y_i;Y_j)=\begin{cases}0 & i\neq j\\ \sigma_\varepsilon^2 & i=j\end{cases}$.Подставляя в (2.21) формулы (2.16), (2.20) и (2.22) имеем:

$$\text{(2.23)}\qquad \operatorname{var}(\hat{\theta}) = \sigma_\varepsilon^2\cdot\left[\frac{c^2}{N}+\frac{c^2\bar{X}^2+d^2-2\cdot c\cdot d\cdot\bar{X}}{\sum x_i^2}\right]=\sigma_\varepsilon^2\cdot\left[\frac{c^2}{N}+\frac{\left(c\cdot\bar{X}-d\right)^2}{\sum x_i^2}\right]$$

Оценка $\hat{\theta} = c\cdot\hat{\alpha} + d\cdot\hat{\beta}$ представляет для нас практический интерес, когда $c=1,\ d=X_i$. В этом случае, это будет МНК оценка $\hat{Y}_i = \hat{\alpha}+\hat{\beta}\cdot X_i$, которая соответственно

является линейной, несмещенной и эффективной оценкой $E\left(Y\middle|X_i\right)=\alpha+\beta\cdot X_i$. Мы можем вычислять эту оценку как для значений X_i $i=1.....N$ из массива $\left(Y,X\right)$, по которому мы оценивали регрессию, так и для произвольного значения X_0. Во втором случае мы получаем прогноз математического ожидания $E\left(Y\middle|X_0\right)$, для гипотетического случая, когда $X=X_0$. Подставляя в формулу (2.23) $c=1$, $d=X_i$, получаем дисперсию $\widehat{Y}_i$

$$\text{(2.24)}\qquad \operatorname{var}\left(\widehat{Y}_i\right)=\sigma_\varepsilon^2\cdot\left[\frac{1}{N}+\frac{\left(X_i-\overline{X}\right)^2}{\sum x_i^2}\right]$$

Из формулы (2.24) видно, что чем сильнее X_i (или X_0) отличается от $\overline{X}$ тем больше дисперсия $\widehat{Y}_i$ (или $\widehat{Y}_0$), а значит меньше точность предсказания с помощью линейной регрессии. Мы еще раз обратим на это внимание, когда будем строить доверительный интервал для прогноза.

Задачка на дом №6. Вычислите $\operatorname{cov}\left(\widehat{Y}_i,\widehat{Y}_j\right)$.

Мы также можем вычислить МНК оценки ошибок $e_i=Y_i-\widehat{Y}_i$, их называют остатками. Остатки обычно используют для проверки предпосылок теоремы Гаусса-Маркова о свойствах случайного возмущения ε_i, то есть их фактически рассматривают как выборку из этой случайной величины. Хотя это и не совсем правильно. Посмотрим насколько характеристики e_i совпадают с характеристиками ε_i

$$\text{(2.25)}\qquad E\left(e_i\right)=E\left(Y_i-\widehat{Y}_i\right)=E\left(Y_i-E\left(Y\middle|X_i\right)\right)=0$$

Из (2.25) видим, что математическое ожидание e_i и ε_i совпадают. Вычислим теперь дисперсию

$$\text{(2.26)}\qquad \operatorname{var}\left(e_i\right)=\operatorname{var}\left(Y_i-\widehat{Y}_i\right)=\operatorname{var}\left(Y_i\right)+\operatorname{var}\left(\widehat{Y}_i\right)-2\cdot\operatorname{cov}\left(Y_i,\widehat{Y}_i\right)$$

мы знаем, что $\operatorname{var}\left(Y_i\right)=\sigma_\varepsilon^2$, и $\operatorname{var}\left(\widehat{Y}_i\right)=\sigma_\varepsilon^2\cdot\left[\frac{1}{N}+\frac{\left(X_i-\overline{X}\right)^2}{\sum x_i^2}\right]$ (формула (2.24)), осталось вычислить $\operatorname{cov}\left(Y_i,\widehat{Y}_i\right)$:

$$\begin{aligned}&\operatorname{cov}\left(Y_i,\widehat{Y}_i\right)=\operatorname{cov}\left(Y_i;\widehat{\alpha}+\widehat{\beta}\cdot X_i\right)=\operatorname{cov}\left(Y_i;\widehat{\alpha}\right)+X_i\cdot\operatorname{cov}\left(Y_i;\widehat{\beta}\right)=\\ &=\operatorname{cov}\left(Y_i;\sum\left(\frac{1}{N}-\overline{X}\cdot w_j\right)\cdot Y_j\right)+X_i\cdot\operatorname{cov}\left(Y_i;\sum_{i=1}^{N}w_i\cdot Y_i\right)=\\ &=\sigma_\varepsilon^2\left(\frac{1}{N}-\overline{X}\cdot w_i+X_i\cdot w_i\right)=\sigma_\varepsilon^2\left(\frac{1}{N}+\frac{x_i^2}{\sum x_i^2}\right)\end{aligned}\qquad\text{(2.27)}$$

Внимательный читатель может обратить внимание, что $\operatorname{cov}\left(Y_i,\widehat{Y}_i\right)$, полученная в (2.27), совпадает с выражением для $\operatorname{var}\left(\widehat{Y}_i\right)$ (2.24), это не случайно. Значит, ее можно было вычислить более простым способом

$\operatorname{cov}\left(Y_i,\widehat{Y}_i\right)=\operatorname{cov}\left(\widehat{Y}_i+e_i,\widehat{Y}_i\right)=\operatorname{cov}\left(\widehat{Y}_i,\widehat{Y}_i\right)+\operatorname{cov}\left(e_i,\widehat{Y}_i\right)=\operatorname{var}\left(\widehat{Y}_i\right)$, так как $\operatorname{cov}\left(e_i,\widehat{Y}_i\right)=0$, почему это так будет объяснено ниже (смотри задачу №8).

Подставляя $\operatorname{var}\left(Y_i\right)=\sigma_\varepsilon^2$ (2.24) и (2.27) в (2.26) имеем:

(2.28) $\operatorname{var}(e_i) = \sigma_\varepsilon^2 \left(1 + \frac{1}{N} + \frac{x_i^2}{\sum x_i^2} - 2\left(\frac{1}{N} + \frac{x_i^2}{\sum x_i^2} \right) \right) = \sigma_\varepsilon^2 \cdot \left(1 - \frac{1}{N} - \frac{x_i^2}{\sum x_i^2} \right)$

Из формулы (2.28) видим, что в отличие от ε_i остатки не являются гомоскедастичными, то есть $\operatorname{var}(e_i) \neq \operatorname{var}(e_j)$.

Задачка на дом №7. Вычислите $\operatorname{cov}(e_i, e_j)$.

Для нас важно, что $\operatorname{cov}(e_i, e_j) \neq 0$, то есть корреляция остатков в отличие от ε_i не равна нулю. Это очевидно, так как остатки всегда будут линейно зависимы ведь для них должны выполнятся условия $\sum_{i=1}^{N} e_i = 0$ и $\sum_{i=1}^{N} X_i \cdot e_i = 0$. То есть независимыми среди e_i может быть только $N-2$ величины. Эти две степени свободы уходят к оценкам $\hat{\alpha}$ и $\hat{\beta}$, которые в свою очередь являются независимыми от остатков.

Задачка на дом №8. Покажите, что $\operatorname{cov}(e_i, \hat{\alpha}) = 0, \operatorname{cov}(e_i, \hat{\beta}) = 0 \; \forall i$ и $\operatorname{cov}(e_i, \hat{Y}_j) = 0 \; \forall i, j$

Существует метод, который позволяет получить $N-2$ оценки ошибок уравнения регрессии, которые в случае, если выполняются все предпосылки теоремы Гаусса-Маркова, по своим свойствам в точности совпадают с ε_i. То есть они имеют нулевое математическое ожидание, одинаковую дисперсию и не коррелированны. Это- метод рекурсивных остатков, он находится за рамками нашего курса. Желающие могут о нем прочитать в учебнике Johnston, в программе Eviews он реализован.

Остатки не нужно путать с ошибками, которые возникают при построении прогноза по модели регрессии. Между ними есть существенные различия. Пусть мы оценили уравнение регрессии $Y_i = \hat{\alpha} + \hat{\beta} \cdot X_i$ для $i = 1....N$ и хотим построить прогноз значения случайной величины Y, если $X = X_0$: $Y_0 = \alpha + \beta \cdot X_0 + \varepsilon_0$. Он будет вычисляться, как $\hat{Y}_0 = \hat{\alpha} + \hat{\beta} \cdot X_0$, ошибка прогноза будет определяться, как

(2.29) $$e_0 = Y_0 - \hat{Y}_0$$

Посмотрим, чем e_0 отличается от остатков регрессии. Математическое ожидание не изменилось: $E(e_0) = 0$ - такой прогноз называют несмещенным. Вычислим дисперсию

(2.30) $$\operatorname{var}(e_0) = \operatorname{var}(Y_0 - \hat{Y}_0) = \operatorname{var}(Y_0) + \operatorname{var}(\hat{Y}_0) - 2 \cdot \operatorname{cov}(Y_0, \hat{Y}_0)$$

Формула (2.30) внешне ничем не отличается от (2.26), даже значения $\operatorname{var}(Y_0)$ и $\operatorname{var}(\hat{Y}_0)$ будут вычисляться точно также $\operatorname{var}(Y_0) = \sigma_\varepsilon^2$, $\operatorname{var}(\hat{Y}_0) = \sigma_\varepsilon^2 \cdot \left[\frac{1}{N} + \frac{(X_0 - \bar{X})^2}{\sum x_i^2} \right]$ (2.24), однако теперь $\operatorname{cov}(Y_0, \hat{Y}_0) = 0$. Дело в том, что $\hat{Y}_0$ это МНК оценка, которая является линейной комбинацией Y_i для $i = 1....N$, а Y_0 среди них нет, так как $\operatorname{cov}(Y_i Y_0) = 0$ для $\forall i \neq 0$, то и $\operatorname{cov}(Y_0, \hat{Y}_0) = 0$. Подставляя все это в (2.30) получаем:

(2.31) $$\operatorname{var}(e_0) = \sigma_\varepsilon^2 \cdot \left(1 + \frac{1}{N} + \frac{(X_0 - \bar{X})^2}{\sum x_i^2} \right)$$

Сравнивая (2.31) и (2.28) можно заметить, что дисперсия ошибки прогноза всегда больше чем дисперсия любого из остатков. Так как $\operatorname{cov}(Y_0, \hat{Y}_0) = 0$, то из формулы (2.30)

следует, что дисперсию ошибки прогноза можно разложить на две составляющих. Первая – это дисперсия прогноза $\widehat{Y}_0$ (2.24), как мы ранее заметили она тем больше, чем сильнее X_0 отличается от среднего значения выборки, по которой мы оценивали регрессию. Второй составляющей дисперсии ошибки прогноза является дисперсия прогнозируемого нами значения Y_0, $\text{var}(Y_0) = \text{var}(\varepsilon_0) = \sigma_\varepsilon^2$ - она не зависит от объясняющей переменной X в силу предположения т. Гаусса-Маркова о гомоскедастичности ошибок.

Проверка гипотез и построение доверительных интервалов в регрессии с нормальными ошибками.

Чтобы проверять гипотезы относительно значений параметров модели регрессии (2.14), нужно дополнить предпосылки теоремы Гаусса-Маркова предположением о том, что ε_i имеют нормальное распределение. Его можно считать оправданным, так как нормально распределение описывает случайную величину значение, которой формируется под воздействием очень большого числа независимых случайных факторов, причем сила воздействия каждого отдельного фактора мала и не может превалировать среди остальных, а характер воздействия *аддитивный*. А ε_i у нас и представляют собой сумму всех не учтенных в модели факторов. Либо, если у нас большая выборка, можно сослаться на центральную предельную теорему.

И так мы предполагаем, что ε_i распределен $N(0; \sigma_i^2)$. Из условия 3.3 т. Гаусса-Маркова $\text{cov}(\varepsilon_i \varepsilon_j) = 0 \ \forall i \neq j$. Для нормального распределения из равенства нулю ковариации следует независимость случайных величин. Поэтому $\varepsilon_i; \varepsilon_j$ будут независимы. Y_i, Y_j тоже будут независимы и Y_i распределен $N(\alpha + \beta \cdot X_i; \sigma_\varepsilon^2)$.

Все МНК оценки $(\widehat{\alpha}, \widehat{\beta}, \widehat{Y}_i)$ и остатки e_i, являются линейными комбинациями независимых нормально распределенных величин Y_i, поэтому они тоже будут иметь нормальное распределение.

Теорема 2.1 Если выполняются все предпосылки т. Гаусса-Маркова и ε_i распределение нормально, то справедливы следующие утверждения :

$\widehat{\alpha}$ распределена $N\left(\alpha;\ \left(\frac{1}{N} + \frac{\overline{X}^2}{\sum x_i^2}\right) \cdot \sigma_\varepsilon^2\right)$, $\widehat{\beta}$ распределена $N\left(\beta;\ \sigma_\varepsilon^2 \Big/ \sum_{i=1}^{N} x_i^2\right)$,

$$\widehat{Y}_i - N\left(\alpha + \beta \cdot X_i; \left[\frac{1}{N} + \frac{\left(X_i - \overline{X}\right)^2}{\sum x_i^2}\right] \cdot \sigma_\varepsilon^2\right) \qquad e_i - N\left(0; \left[1 - \frac{1}{N} - \frac{\left(X_i - \overline{X}\right)^2}{\sum x_i^2}\right] \cdot \sigma_\varepsilon^2\right)$$

Причем случайные величины $\widehat{\alpha}, \widehat{\beta}$ и $\widehat{Y}_i \ \forall i$ независимы от остатков $e_j \ \forall j$ (последнее утверждение следует из результатов решения задачи №8). Ошибка прогноза e_0 (определена в 2.29) распределена $N\left(0; \left[1 + \frac{1}{N} + \frac{\left(X_0 - \overline{X}\right)^2}{\sum x_i^2}\right] \cdot \sigma_\varepsilon^2\right)$, причем случайные величины e_0 и e_i независимы для $\forall i = 1....N$.

Задачка на дом № 9. Докажите что $\text{cov}(e_0, e_i) = 0$, используйте результаты решения задачки №8.

Чтобы строить доверительные интервалы и проверять гипотезы нужно избавиться от неизвестного параметра σ_ε^2.

Теорема 2.2. Случайная величина $\sum_{i=1}^{N} e_i^2 \Big/ \sigma_\varepsilon^2$ имеет распределение $\chi^2(N-2)$ и независима от случайных величин $\hat{\alpha}$ и $\hat{\beta}$ (соответственно и от их линейной комбинации $\hat{Y}_i = \hat{\alpha} + \hat{\beta} \cdot X_i$)

Логика этой теоремы вполне понятна, мы уже отмечали, что остатки линейно зависимы, для них всегда выполняются соотношения $\sum_{i=1}^{N} e_i = 0$ и $\sum_{i=1}^{N} X_i \cdot e_i = 0$, поэтому среди N слагаемых суммы $\sum_{i=1}^{N} e_i^2$ независимых будет только $N-2$, поэтому у нас получаются такие степени свободы.

Следствия из теорем 2.1 и 2.2.

1) Оценка:

$$\hat{\sigma}_\varepsilon^2 = \frac{\sum_{i=1}^{N} e_i^2}{N-2} = \frac{RSS}{N-2} \text{ -} \tag{2.32}$$

является несмещенной оценкой σ_ε^2. Доказательство

$$\hat{\sigma}_\varepsilon^2 = \frac{\sum_{i=1}^{N} e_i^2}{N-2} = \frac{\sigma_\varepsilon^2}{N-2} \cdot \frac{\sum_{i=1}^{N} e_i^2}{\sigma_\varepsilon^2} = \frac{\sigma_\varepsilon^2}{N-2} \cdot \chi^2(N-2)$$

$$E\left(\hat{\sigma}_\varepsilon^2\right) = \frac{\sigma_\varepsilon^2}{N-2} \cdot E\left(\chi^2(N-2)\right) = \frac{\sigma_\varepsilon^2}{N-2} \cdot (N-2) = \sigma_\varepsilon^2$$

2) случайная величина $(\hat{\alpha} - \alpha) \Big/ \sqrt{\hat{\sigma}_\varepsilon^2 \cdot \left(\frac{1}{N} + \frac{\bar{X}^2}{\sum x_i^2}\right)}$ имеет распределение $t(N-2)$, где $\hat{\sigma}_\varepsilon^2$ определяется по формуле (2.32). Докажем это. По теореме 2.1 $\hat{\alpha}$ распределена $N\left(\alpha;\ \left(\frac{1}{N} + \frac{\bar{X}^2}{\sum x_i^2}\right) \cdot \sigma_\varepsilon^2\right)$, соответственно $a = \frac{\hat{\alpha} - \alpha}{\sqrt{\left(\frac{1}{N} + \frac{\bar{X}^2}{\sum x_i^2}\right) \cdot \sigma_\varepsilon^2}}$ будет распределена $N(0,1)$. По теореме 2.2 $\sum_{i=1}^{N} e_i^2 \Big/ \sigma_\varepsilon^2$ распределена $\chi^2(N-2)$ и независима от $\hat{\alpha}$ и соответственно независима от a, значит $t = \frac{a}{\sqrt{\sum_{i=1}^{N} e_i^2 \Big/ \sigma_\varepsilon^2} \Big/ N-2}$ будет иметь распределение $t(N-2)$. Осталось подставить a и заменить $\frac{\sum_{i=1}^{N} e_i^2}{N-2}$ на $\hat{\sigma}_\varepsilon^2$.

$$t=\frac{\frac{\hat{\alpha}-\alpha}{\sqrt{\left(\frac{1}{N}+\frac{\bar{X}^2}{\sum x_i^2}\right)\cdot\sigma_\varepsilon^2}}}{\sqrt{\frac{\sum_{i=1}^{N}e_i^2}{\sigma_\varepsilon^2}\Big/N-2}}=\frac{\hat{\alpha}-\alpha}{\sqrt{\frac{\sum_{i=1}^{N}e_i^2}{(N-2)\cdot\sigma_\varepsilon^2}\cdot\left(\frac{1}{N}+\frac{\bar{X}^2}{\sum x_i^2}\right)\cdot\sigma_\varepsilon^2}}=\frac{\hat{\alpha}-\alpha}{\sqrt{\hat{\sigma}_\varepsilon^2\cdot\left(\frac{1}{N}+\frac{\bar{X}^2}{\sum x_i^2}\right)}}$$ - это и есть величина распределение, которой мы доказываем.

Обозначим несмещенную оценку дисперсии $\hat{\alpha}$

(2.33) $$\hat{\sigma}_\alpha^2=\left(\frac{1}{N}+\frac{\bar{X}^2}{\sum x_i^2}\right)\cdot\hat{\sigma}_\varepsilon^2$$

где $\hat{\sigma}_\varepsilon^2$ определяется в (2.32).

Соответственно для проверки гипотезы $H_0\,\alpha=\alpha_0$ используется статистика:

(2.34) $$t_\alpha=\left(\hat{\alpha}-\alpha_0\right)\Big/\sqrt{\hat{\sigma}_\alpha^2}$$

Критическое множество будет зависеть от вида альтернативной гипотезы. Для $H_A\,\alpha\neq\alpha_0$, $H_0\,\alpha=\alpha_0$ будет отвергаться на уровне значимости γ, если t_α **не** попадает в интервал $\left[u_{\frac{\gamma}{2}}\left(t\left(N-2\right)\right),\ \ u_{1-\frac{\gamma}{2}}\left(t\left(N-2\right)\right)\right]$, где u_q - это квантиль порядка q определена в (1.23). Для $H_A\,\alpha<\alpha_0$, $H_0\,\alpha=\alpha_0$ отвергается на уровне значимости γ, если $t_\alpha<u_\gamma\left(t\left(N-2\right)\right)$. Соответственно, если $H_A\,\alpha>\alpha_0$, то $H_0\,\alpha=\alpha_0$ отвергается для $t_\alpha>u_\gamma\left(t\left(N-2\right)\right)$. Доверительный интервал для α с уровнем доверия $1-\gamma$ будет иметь вид:

(2.35) $$\hat{\alpha}+u_{\frac{\gamma}{2}}\left(t\left(N-2\right)\right)\cdot\sqrt{\hat{\sigma}_\alpha^2}<\alpha<\hat{\alpha}+u_{1-\frac{\gamma}{2}}\left(t\left(N-2\right)\right)\cdot\sqrt{\hat{\sigma}_\alpha^2}$$

так как в силу симметричности t распределения $u_{\frac{\gamma}{2}}\left(t\left(N-2\right)\right)=-u_{1-\frac{\gamma}{2}}\left(t\left(N-2\right)\right)$, в (2.35) удобнее отрицательные величины заменить на положительные со знаком минус и записать доверительный интервал:

(2.36) $$\hat{\alpha}-u_{1-\frac{\gamma}{2}}\left(t\left(N-2\right)\right)\cdot\sqrt{\hat{\sigma}_\alpha^2}<\alpha<\hat{\alpha}+u_{1-\frac{\gamma}{2}}\left(t\left(N-2\right)\right)\cdot\sqrt{\hat{\sigma}_\alpha^2}$$

3) Случайная величина $\left(\hat{\beta}-\beta\right)\Big/\sqrt{\frac{\hat{\sigma}_\varepsilon^2}{\sum x_i^2}}$ имеет распределение $t\left(N-2\right)$, где $\hat{\sigma}_\varepsilon^2$ определяется по формуле (2.32). Распределение доказывается так же, как и в свойстве 2. Для удобства обозначим несмещенную оценку дисперсии $\hat{\beta}$

(2.37) $$\hat{\sigma}_\beta^2=\frac{1}{\sum x_i^2}\hat{\sigma}_\varepsilon^2$$

где $\hat{\sigma}_\varepsilon^2$ определяется в (2.32).

Соответственно для проверки гипотезы $H_0\,\beta=\beta_0$ используется статистика:

(2.38) $$t_\beta=\left(\hat{\beta}-\beta_0\right)\Big/\sqrt{\hat{\sigma}_\beta^2}$$

Критические множества для различных альтернативных гипотез строятся точно так же как для t_α. Доверительный интервал для параметра β с уровнем доверия $1-\gamma$ имеет вид:

$$\hat{\beta} - u_{1-\frac{\gamma}{2}}\left(t\left(N-2\right)\right)\cdot\sqrt{\hat{\sigma}_{\beta}^{2}} < \beta < \hat{\beta} + u_{1-\frac{\gamma}{2}}\left(t\left(N-2\right)\right)\cdot\sqrt{\hat{\sigma}_{\alpha}^{2}} \tag{2.39}$$

где $\hat{\sigma}_{\beta}^{2}$ определяется в (2.37).

4) Случайная величина $\dfrac{\hat{Y}_0 - \left(\alpha + \beta\cdot X_0\right)}{\sqrt{\left[\dfrac{1}{N} + \dfrac{\left(X_0 - \bar{X}\right)^2}{\sum x_i^2}\right]\cdot\hat{\sigma}_{\varepsilon}^{2}}}$ имеет распределение $t\left(N-2\right)$

Распределение доказывается так же, как и в свойстве 2. Мы можем проверять гипотезы о истинном значении параметра $E\left(Y\middle|X_0\right) = \alpha + \beta\cdot X_0$, аналогично тому как это делалось в свойствах 2 и 3. Практический интерес для нас представляет доверительный интервал для математического ожидания $E\left(Y\middle|X_0\right)$.

$$\begin{aligned}\hat{Y}_0 - u_{1-\frac{\gamma}{2}}\left(t\left(N-2\right)\right)\cdot\sqrt{\left[\frac{1}{N} + \frac{\left(X_0 - \bar{X}\right)^2}{\sum x_i^2}\right]\cdot\hat{\sigma}_{\varepsilon}^{2}} < \alpha + \beta\cdot X_0 \\ < \hat{Y}_0 + u_{1-\frac{\gamma}{2}}\left(t\left(N-2\right)\right)\cdot\sqrt{\left[\frac{1}{N} + \frac{\left(X_0 - \bar{X}\right)^2}{\sum x_i^2}\right]\cdot\hat{\sigma}_{\varepsilon}^{2}}\end{aligned} \tag{2.40}$$

Из (2.40) мы видим, что чем сильнее значение X_0, для которого мы прогнозируем $E\left(Y\middle|X_0\right)$, отличается от среднего значения выборки, по которой мы оценили регрессию, тем шире будет доверительный интервал.

5) Случайная величина $\dfrac{e_0}{\sqrt{\left[1 + \dfrac{1}{N} + \dfrac{\left(X_0 - \bar{X}\right)^2}{\sum x_i^2}\right]\cdot\hat{\sigma}_{\varepsilon}^{2}}} = \dfrac{Y_0 - \hat{Y}_0}{\sqrt{\left[1 + \dfrac{1}{N} + \dfrac{\left(X_0 - \bar{X}\right)^2}{\sum x_i^2}\right]\cdot\hat{\sigma}_{\varepsilon}^{2}}}$ имеет распределение $t\left(N-2\right)$. Независимость e_0 и $\hat{\sigma}_{\varepsilon}^{2}$ следует из результатов решения задачи 9. Доверительный интервал для прогнозируемого значения Y_0 будет иметь вид:

$$\begin{aligned}\hat{Y}_0 - u_{1-\frac{\gamma}{2}}\left(t\left(N-2\right)\right)\cdot\sqrt{\left[1 + \frac{1}{N} + \frac{\left(X_0 - \bar{X}\right)^2}{\sum x_i^2}\right]\cdot\hat{\sigma}_{\varepsilon}^{2}} < Y_0 \\ < \hat{Y}_0 + u_{1-\frac{\gamma}{2}}\left(t\left(N-2\right)\right)\cdot\sqrt{\left[1 + \frac{1}{N} + \frac{\left(X_0 - \bar{X}\right)^2}{\sum x_i^2}\right]\cdot\hat{\sigma}_{\varepsilon}^{2}}\end{aligned} \tag{2.41}$$

Когда мы прогнозируем $E\left(Y\middle|X_0\right) = \alpha + \beta\cdot X_0$ неопределенность возникает из-за того, что мы не знаем истинных значений параметров α и β, а используем их оценки $\hat{\alpha}$ и $\hat{\beta}$. Когда мы переходим к прогнозу $Y_0 = \alpha + \beta\cdot X_0 + \varepsilon_0 = E\left(Y\middle|X_0\right) + \varepsilon_0$ возникает дополнительная неопределенность, так как значение ε_0 нам не известно. Поэтому, сравнивая (2.40) и (2.41), можно заметить, что доверительный интервал для Y_0 всегда шире чем для $E\left(Y\middle|X_0\right)$. Также можно отметить, что оба доверительных интервала тем шире, чем сильнее X_0 отличается от $\bar{X}$.

Обратите внимание, что в теореме 2.1 постулировалось нормальное распределение 5 величин (некоторые из них векторные) $\hat{\alpha}, \hat{\beta}, \hat{Y}_i, e_i, e_0$, а в следствиях из теоремы 2.1 мы построили t-распределение только для $\hat{\alpha}, \hat{\beta}, \hat{Y}_i, e_0$.

Задачка на дом №10. Почему величина $e_i \Big/ \sqrt{\left[1 - \frac{1}{N} - \frac{\left(X_i - \bar{X}\right)^2}{\sum x_i^2}\right] \cdot \hat{\sigma}_\varepsilon^2}$ не будет иметь распределение $t\left(N-2\right)$?

Из теоремы 2.1 мы знаем, что $\hat{\beta}$ распределена $N\left(\beta;\ \sigma_\varepsilon^2 \Big/ \sum_{i=1}^N x_i^2\right)$, соответственно $\dfrac{\hat{\beta} - \beta}{\sqrt{\sigma_\varepsilon^2 \Big/ \sum_{i=1}^N x_i^2}}$ - распределена $N\left(0,\ 1\right)$, а $\dfrac{\left(\hat{\beta} - \beta\right)^2}{\sigma_\varepsilon^2 \Big/ \sum_{i=1}^N x_i^2}$ распределена $\chi^2\left(1\right)$. Если справедлива $H_0\ \beta = 0$, то $\dfrac{\left(\hat{\beta} - \beta\right)^2}{\sigma_\varepsilon^2 \Big/ \sum_{i=1}^N x_i^2} = \dfrac{\hat{\beta}^2}{\sigma_\varepsilon^2 \Big/ \sum_{i=1}^N x_i^2} = \dfrac{\hat{\beta}^2 \cdot \sum_{i=1}^N x_i^2}{\sigma_\varepsilon^2} = \dfrac{\hat{\beta} \cdot \sum_{i=1}^N x_i \cdot y_i}{\sigma_\varepsilon^2} = \dfrac{ESS}{\sigma_\varepsilon^2}$ при выводе я воспользовался формулами (2.10) и (2.04). Мы показали, что при выполнении $H_0\ \beta = 0$ случайная величина ESS/σ_ε^2 распределена как $\chi^2\left(1\right)$. Из теоремы 2.2 случайная величина RSS/σ_ε^2 распределена $\chi^2\left(N-2\right)$ независима от $\hat{\beta}$ и поэтому независима от ESS/σ_ε^2. Соответственно, если $\beta = 0$ $\dfrac{TSS}{\sigma_\varepsilon^2} = \dfrac{ESS}{\sigma_\varepsilon^2} + \dfrac{RSS}{\sigma_\varepsilon^2} = \chi^2\left(1\right) + \chi^2\left(N-2\right) = \chi^2\left(N-1\right)$, так как слагаемые независимы.

Определим статистику:

$$F = \frac{ESS/1}{RSS/\left(N-2\right)} \tag{2.42}$$

Если справедлива $H_0\ \beta = 0$ статистика (2.42) будет иметь распределение $F\left(1,\ N-2\right)$, так $F = \dfrac{ESS/1}{RSS/\left(N-2\right)} = \dfrac{ESS/\hat{\sigma}_\varepsilon^2}{RSS/\hat{\sigma}_\varepsilon^2 \cdot \left(N-2\right)} = \dfrac{\chi^2\left(1\right)/1}{\chi^2\left(N-2\right)/\left(N-2\right)} = F\left(1; N-2\right)$. При проверке гипотезы $H_0\ \beta = 0$ против $H_A\ \beta \neq 0$ критическое множество будет правосторонним. То есть, если мы проверяем на уровне значимости γ, H_0 отвергается, если $F > u_{1-\gamma}\left(F\left(1, N-2\right)\right)$.

Когда мы прейдем к многофакторной регрессии, то F-статистика вида (2.42) будет нами использоваться для проверки гипотезы о том, что коэффициенты при всех объясняющих переменных равны нулю (обычно ее называют гипотезой о значимости регрессии). В случае однофакторной регрессии использование F-статистики не дает нам ничего нового по сравнению со статистикой t_β (определена в (2.38)). При проверке гипотезы $H_0\ \beta = 0$ использование обеих статистик дает одинаковые результаты.

Задачка на дом №11. Покажите что F и $t_\beta\left(\beta = 0\right)$ эквивалентны. Подсказка: докажите, что $F = t_\beta^2\left(\beta = 0\right)$ и вспомните, что $F\left(1,\left(N-2\right)\right) = t^2\left(N-2\right)$.

Рассмотренная выше F -статистика функционально выражается через коэффициент R^2. $F = \frac{ESS/1}{RSS/(N-2)} = \frac{\frac{ESS}{TSS}}{\frac{RSS}{TSS}/(N-2)} = \frac{\frac{ESS}{TSS}}{\left(1-\frac{ESS}{TSS}\right)/(N-2)} = \frac{R^2\cdot(N-2)}{(1-R^2)}$

(2.43) $$F = \frac{R^2\cdot(N-2)}{(1-R^2)}$$

Это вполне логично, так как R^2 тоже измеряет качество регрессии в целом.

Литература по однофакторной регрессии.

Я пользовался тремя учебниками, которые доступны в электронном виде: (1) Gujarati D.N. Basic Econometrics, 3e, 1995, (2) Maddala G.S. Introduction to Econometrics, 2e, 1992 (3) Johnston J., DiNardo J.J. Econometric methods, 4e. Еще заглядывал в конспекты лекций по эконометрике Канторовича Г. Г. и Ершова И. Б. В каждом из трех учебников можно найти почти все, что я написал (можно, но сложно). Я просто собрал материал из разных глав и изложил в краткой форме. Если вы не станете ограничиваться прочтением этой разработки, а обратитесь к первоисточникам, то многие вещи вы поймете лучше. Поэтому приведу более подробные ссылки.

Gujarati D.N- отличается тем, что в нем все изложено очень подробно, поэтому он раза в два больше двух других учебников. Он есть в двух видах папка Gujarati D.N. Basic Econometrics, 3e, 1995 и папка Gujarati D.N. Basic Econometrics, 3e, 1995, OF. Вторая лучше, так как в ней он лучшего качества, сделан весь в одном файле и на одной странице - одна страница, а не один разворот. Правда в версии OF нет оглавления, приходится его смотреть в первой версии. Страницы привожу по файлу Gujarati D.N. Basic Econometrics, 3e, 1995.pdf, 5,60 Mb, 1003 страницы. Материал по однофакторной регрессии содержится на стр. с 22 по 198. 22-42-что такое регрессионный анализ и чем он отличается от других видов исследования зависимостей. 42-57- что такое регрессия с одной объясняющей переменной. 57-70- МНК, 70-81-предпосылки т. Гаусса-Маркова (они там очень подробно разобраны, хорошо объясняется, в чем их смысл советую прочитать). 112-124-предпосылка о нормальности ошибок и ее следствия. 124-164- проверка гипотез и построение доверительных интервалов в модели регрессии. 164-198-дополнительный материал по однофакторной регрессии. Все написано подробно и просто, но боюсь, не у всех хватит времени и сил это прочитать.

Maddala G.S.-файл Maddala.pdf, 64,5Mb, 637 страниц. Разобраться в этом учебнике очень просто, оглавление есть, материал по однофакторной регрессии содержится в одной главе Chapter3 Simple Regression стр. 72-139. Вам нужны части 3.2, 3.4-3.7. Все кратко, прочитать вполне реально, остальные части 3 главы тоже вполне можете прочитать.

Johnston J., DiNardo J.J – файл Johnston J., DiNardo J.J. Econometric methods, 4e.pdf, 17,3 Mb, 514 страниц. Оглавление отсканировали все, кроме первой страницы, а она нам и нужна. Вам нужна глава 1 стр.11-51, регрессия начинается на странице 25. Во второй главе содержатся дополнительные материалы по однофакторной регрессии. Достаточно кратко, так как автора интересуют более сложные темы, но вполне понятно.

Платон.
platonhse@mail.ru