

## Автокорреляция

Наряду с мультиколлинеарностью и гетероскедастичностью в регрессионных моделях может иметь место *автокорреляция остатков*.

Само по себе понятие **автокорреляции** представляет собой зависимость последовательных элементов временного и пространственного ряда данных.

Данное явление имеет место, когда нарушается предположение о независимости случайных величин, т.е. когда

$$\text{cov}(u_j, u_l) \neq 0, \quad j \neq l.$$

Наиболее часто автокорреляция остатков наблюдается, когда выборка берется во временном разрезе, что связано и инерционностью и цикличностью многих экономических процессов и явлений. Другой причиной автокорреляции может быть неверная спецификация модели, когда в уравнение не включена существенно влияющая переменная и ее влияние отражается на величинах остатков.

В простейшем виде автокорреляция остатков может быть представлена уравнением

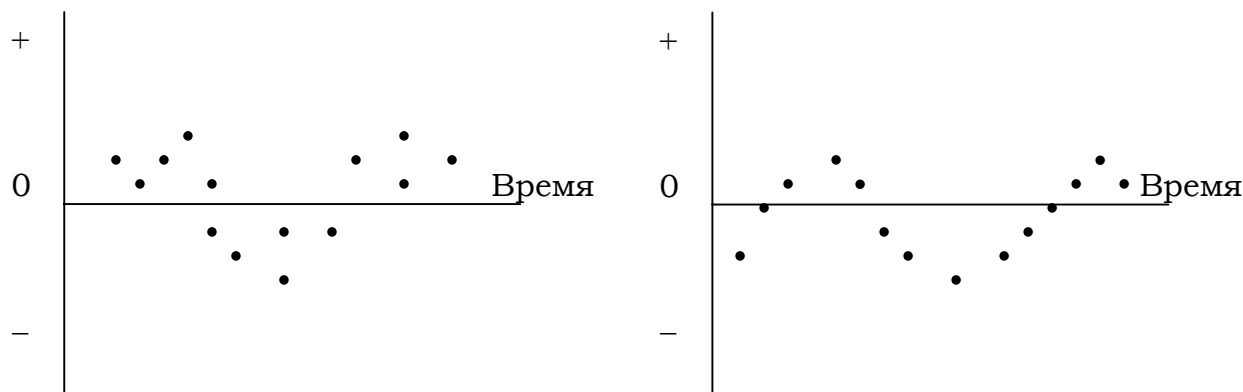
$$u_t = \rho u_{t-1} + \varepsilon_t,$$

где  $|\rho| < 1$  – автокорреляционный коэффициент первого порядка,  $t$  – период времени, а ошибки  $\varepsilon_t$  обладают свойствами

$$\begin{cases} M(\varepsilon_t) = 0, \\ \left\{ \begin{array}{l} \text{Var}(\varepsilon_t) = \sigma_\varepsilon^2 = \text{const} \\ \text{cov}(\varepsilon_t, \varepsilon_{t+s}) = 0, \quad s \neq 0 \end{array} \right. \end{cases}$$

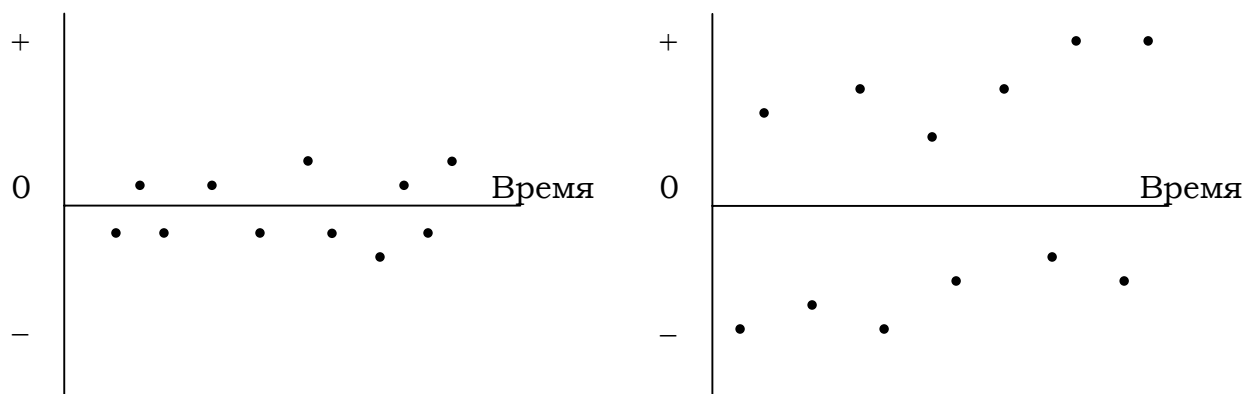
Величина  $\rho$  определяет силу автокорреляции в уравнении. Если  $\rho = 0$ , то автокорреляция отсутствует, т.к.  $\varepsilon$  будет равно  $u$ , и остатки  $u$  в этом случае будут отвечать классическим условиям применения 1МНК. Если  $\rho$  будет приближаться к единице по абсолютной величине, то это будет означать, что значение предыдущего наблюдения остатка  $\varepsilon$  становится более важным в определении текущего значения  $\varepsilon_t$ . Величина  $\rho$  большая, чем единица, не имеет смысла.

Автокорреляция может быть положительной и отрицательной. Это определяется знаком величины  $\rho$ . Ее положительные значения свидетельствуют о том, что остатки имеют тенденцию сохранять свой знак от наблюдения к наблюдению.



Графическая иллюстрация положительной автокорреляции

Отрицательная автокорреляция предполагает наличие тенденции к постоянной смене знака величины остатка. Такая ситуация встречается довольно редко.



Графическая иллюстрация отрицательной автокорреляции

Автокорреляция может учитывать зависимость последующего значения величины остатка от предыдущего. Могут иметь место и более длительные запаздывания. Так, например, остаток, рассчитанный для настоящего периода времени (года), может быть функцией от величины остатка годом раньше. Если наблюдения собраны по кварталам, описанная ситуация представится в виде уравнения

$$u_t = \rho u_{t-4} + \varepsilon_t.$$

Кроме того, зависимость может быть определена от величин остатков сразу нескольких периодов времени. Например

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \varepsilon_t.$$

Такая формулировка носит название автокорреляции второго порядка.

Автокорреляционное уравнение первого порядка может быть преобразовано следующим образом

$$\begin{aligned} u_t &= \rho u_{t-1} + \varepsilon_t = \rho(\rho u_{t-2} + \varepsilon_{t-1}) + \varepsilon_t = \dots = \\ &= \varepsilon_t + \rho \varepsilon_{t-1} + \rho^2 \varepsilon_{t-2} + \dots \end{aligned}$$

так как  $|\rho^\infty| \rightarrow 0$  и, следовательно,  $\rho^r u_{t-r} = 0$ .

Другими словами,

$$u_t = \sum_{r=0}^{\infty} \rho^r \varepsilon_{t-r}.$$

Учитывая свойства остатков  $\varepsilon_t$  ( $M(\varepsilon_t) = 0$ ), можем заключить, что  $M(u_t) = 0$ . Далее рассмотрим математическое ожидание квадратов остатков (или дисперсию остатков):

$$\text{Var}(u_t) = M(u_t^2) = M(\varepsilon_t^2) + \rho^2 M(\varepsilon_{t-1}^2) + \rho^4 M(\varepsilon_{t-2}^2) + \dots$$

Независимость последовательных значений  $\varepsilon_t$  и постоянство их дисперсий позволяет представить вариацию остатков  $u_t$  в виде

$$\text{Var}(u_t) = (1 + \rho^2 + \rho^4 + \dots) \sigma_\varepsilon^2.$$

А так как  $|\rho| < 1$ , то выражение в скобках представляет собой бесконечно убывающую геометрическую прогрессию с первым членом, равным единице, и знаменателем  $\rho^2$ . С учетом этого имеем

$$\sigma_u^2 = \frac{\sigma_\varepsilon^2}{1 - \rho^2}.$$

Ковариация последовательных значений остатков  $u_t$  определяется следующим образом:

$$\begin{aligned} \text{cov}(u_t, u_{t-1}) &= \rho \sigma_u^2, \\ \text{cov}(u_t, u_{t-2}) &= \rho^2 \sigma_u^2, \\ &\vdots \\ \text{cov}(u_t, u_{t-s}) &= \rho^s \sigma_u^2, \end{aligned}$$

что означает невыполнение условия независимости последовательных значений остатков.

Таким образом, при наличии автокорреляции остатков имеет место выражение

$$Var(u) = \sigma_u^2 Z,$$

где  $Z$  – матрица коэффициентов автокорреляции  $s$ -го порядка для временного ряда  $u_t$  или

$$Var(u) = V,$$

что может быть представлено следующим образом:

$$Var(u) = V = \sigma_u^2 \cdot \begin{pmatrix} 1 & \rho & \rho^2 & \dots & \rho^{n-1} \\ \rho & 1 & \rho & \dots & \rho^{n-2} \\ \rho^2 & \rho & 1 & \dots & \rho^{n-3} \\ \dots & \dots & \dots & \dots & \dots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \dots & 1 \end{pmatrix}$$

Последствиями автокорреляции являются

1. Неэффективность оценок параметров модели (выборочные дисперсии вектора оценок  $\hat{A}$  могут быть неоправданно большими);
2. Невозможность использования статистических критериев Стьюдента и Фишера для проверки значимости параметров модели, так как выборочные дисперсии рассчитываются по неуточненным формулам;
3. Неэффективность прогнозов (с большой выборочной дисперсией), получаемых на основе неэффективных оценок параметров.

## **Проверка наличия автокорреляции**

### **Критерий Дарбина-Уотсона**

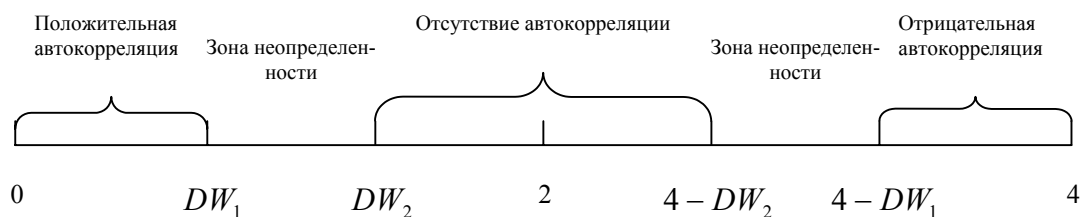
Данный критерий используется для проверки наличия автокорреляции остатков. Он является наиболее часто используемым тестом. Проверка гипотезы о наличии автокорреляции осуществляется за несколько шагов, при этом имеют место зоны неопределенности, где оценить указанное явление не удается.

1-й шаг. Рассчитываем статистику Дарбина-Уотсона ( $DW$ ) по формуле:

$$DW = \frac{\sum_{t=2}^n (u_t - u_{t-1})^2}{\sum_{t=1}^n u_t^2}.$$

Значение данной статистики может быть в интервале от 0 до 4.

2-й шаг. Для заданного уровня значимости  $\alpha$ , числа степеней свободы, равного числу факторов, включенных в модель, и числа наблюдений находим значения  $DW_1$  и  $DW_2$  – верхнюю и нижнюю границы. Если рассчитанное значение критерия расположено в районе двух, то автокорреляция остатков отсутствует. Положительная автокорреляция имеет место, когда  $DW < 2$ , отрицательная – когда  $DW > 2$ . При этом необходимо учитывать найденные верхнюю и нижнюю границы критерия. Окончательный вывод о наличии или отсутствии автокорреляции можно сделать на основе сопоставления рассчитанного значения с приведенной ниже шкалой.



Зоны автокорреляции

При  $\rho = 1$  значение критерия  $DW = 0$ , при  $\rho = 0$   $DW = 2$  и при  $\rho = -1$   $DW = 4$ .

Рассмотрим **пример**. Пусть имеется выборка из 12 наблюдений.

Номер наблюдения (магазин)	Зависимая переменная (товарооборот), $y$	Независимая переменная (торговая площадь), $x$
1	2,93	0,31
2	5,27	0,98
3	6,85	1,21
4	7,01	1,29
5	7,02	1,12
6	8,35	1,49
7	4,33	0,78
8	5,77	0,94
9	7,68	1,29
10	3,16	0,48
11	1,52	0,24
12	3,15	0,55

Необходимо оценить наличие автокорреляции остатков.

Построим модель, найдем остатки и рассчитаем значения критерия  $DW$ .

Уравнение имеет вид

$$\hat{y} = 0,6057 + 5,222x.$$

Значение критерия составляет

$$DW = \frac{\sum_{t=2}^n (u_t - u_{t-1})^2}{\sum_{t=1}^n u_t^2} = \frac{3,425}{1,666} = 2,056.$$

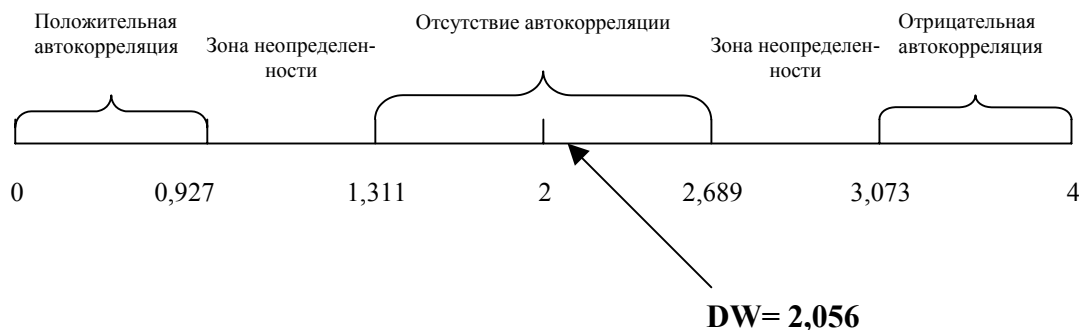
Найдем табличные значения верхней и нижней границ критерия для одной независимой переменной, 12 наблюдений и уровня значимости  $\alpha = 5\%$ :

$$DW_1 = 0,927, \quad DW_2 = 1,311.$$

Находим значения

$$4 - DW_2 = 4 - 1,311 = 2,689 \quad \text{и} \quad 4 - DW_1 = 4 - 0,927 = 3,073.$$

Как видим, полученное значение попадает в интервал, где автокорреляция отсутствует.



### **Критерий фон Неймана**

Наряду с критерием Дарбина-Уотсона, довольно близким к нему для проверки автокорреляции остатков является критерий фон Неймана. Расчет критерия осуществляется по формуле

$$Q = \frac{\sum_{t=2}^n (u_t - u_{t-1})^2}{\frac{n-1}{\sum_{t=1}^n u_t^2}}.$$

Сопоставляя данную формулу с формулой для критерия Дарбина-Уотсона, получаем

$$Q = DW \frac{n}{n-1}.$$

Полученное значение критерия фон Неймана сравнивается с табличным для заданного уровня значимости  $\alpha$  и числа наблюдений. Если

$$Q_{расч.} < Q_{табл.},$$

то имеет место положительная автокорреляция.

### **Нециклический и циклический коэффициенты автокорреляции**

С помощью нециклического коэффициента устанавливается степень взаимосвязи каждого последующего значения величины остатка с предыдущим, т.е. между двумя рядами чисел

$$u_1, u_2, \dots, u_{n-1}$$

и

$$u_2, u_3, \dots, u_n$$

по формуле:

$$r_{нц} = \frac{\sum_{t=2}^n u_t u_{t-1} - \frac{1}{n-1} \left( \sum_{t=1}^n u_t \right) \left( \sum_{t=2}^n u_{t-1} \right)}{\sqrt{\left[ \sum_{t=1}^n u_t^2 - \frac{1}{n-1} \left( \sum_{t=1}^n u_t \right)^2 \right] \left[ \sum_{t=2}^n u_{t-1}^2 - \frac{1}{n-1} \left( \sum_{t=2}^n u_{t-1} \right)^2 \right]}}.$$

Данный коэффициент изменяется в пределах от  $-1$  до  $+1$ . Значение коэффициента, близкое к нулю, свидетельствует об отсутствии автокорреляции остатков.

На практике чаще рассчитывают циклический коэффициент автокорреляции, так как вероятностное распределение нециклического коэффициента установить сложно.

Циклический коэффициент устанавливает степень связи между следующими рядами чисел:

$$u_1, u_2, \dots, u_{n-1}, u_n$$

и

$$u_2, u_3, \dots, u_n, u_1.$$

Для этого используется формула:

$$r_u = \frac{\sum_{t=2}^n u_t u_{t-1} + u_n u_1 - \frac{1}{n} \left( \sum_{t=1}^n u_t \right)^2}{\sum_{t=1}^n u_t^2 - \frac{1}{n} \left( \sum_{t=1}^n u_t \right)^2}.$$

Если имеет место равенство  $u_1 = u_n$ , то нециклический и циклический коэффициенты автокорреляции совпадают.

Для проверки наличия автокорреляции остатков рассчитанное значение циклического коэффициента сравнивают с табличным при заданном уровне значимости  $\alpha$  и числе наблюдений  $n$ . Если выполняется соотношение

$$r_{\text{расч.}} \geq r_{\text{табл.}},$$

то автокорреляция существует.

Если принять следующее допущение, что

$$\sum_{t=1}^n u_t \approx \sum_{t=2}^n u_{t-1} \approx 0,$$

то циклический коэффициент автокорреляции можно рассчитать по формуле:

$$r_u = \frac{\sum_{t=2}^n u_t u_{t-1}}{\sum_{t=1}^n u_t^2}.$$

Часто используется другая формула для расчетов:

$$r_u = \frac{\frac{1}{n-1} \sum_{t=2}^n u_t u_{t-1}}{\frac{1}{n} \sum_{t=1}^n u_t^2}.$$

### **Оценка параметров модели с автокоррелированными остатками (метод Эйткена)**

Если в эконометрической модели

$$y_t = a_0 + a_1 x_t + u_t$$

остатки обладают свойством автокорреляции и изменяются в соответствии с моделью



$$u_t = \rho u_{t-1} + \varepsilon_t,$$

где величины  $\varepsilon_t$  распределены нормально, то для того чтобы избежать автокорреляции, исходная модель преобразуется к виду, когда ее остатки представлены величинами  $\varepsilon_t$ . Остатки  $\varepsilon_t$  могут быть выражены через остальные элементы модели:

$$\varepsilon_t = u_t - \rho u_{t-1}.$$

Запишем уравнение регрессии для  $t-1$  наблюдения:

$$y_{t-1} = a_0 + a_1 x_{t-1} + u_{t-1}.$$

Умножим обе части уравнения на автокорреляционный коэффициент  $\rho$  и вычтем из уравнения для периода времени  $t$ . В результате получим:

$$y_t - \rho y_{t-1} = a_0(1 - \rho) + a_1(x_t - \rho x_{t-1}) + (u_t - \rho u_{t-1}).$$

К преобразованным таким способом данным можно применить метод наименьших квадратов (МНК).

Параметр  $\rho$  приблизительно равен циклическому коэффициенту автокорреляции с поправкой на величину смещения.

Применение метода Эйткена предполагает нахождение матриц  $Z$  или  $V$ .

Так, матрица  $Z$  имеет вид:

$$Z = \begin{pmatrix} 1 & \rho & \rho^2 & \dots & \rho^{n-1} \\ \rho & 1 & \rho & \dots & \rho^{n-2} \\ \rho^2 & \rho & 1 & \dots & \rho^{n-3} \\ \dots & \dots & \dots & \dots & \dots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \dots & 1 \end{pmatrix}.$$

В данной матрице коэффициенты  $\rho^s$  определяют порядок автокорреляции остатков  $u_t$ . Для автокорреляции нулевого порядка  $\rho^0 = 1$ .

Обратная матрица имеет вид:

$$Z^{-1} = \frac{1}{1 - \rho^2} \begin{pmatrix} 1 & -\rho & 0 & \dots & 0 \\ -\rho & 1 + \rho^2 & -\rho & \dots & 0 \\ 0 & -\rho & 1 + \rho^2 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix}.$$

Наличие большого количества нулей в матрице объясняется тем, что при порядке, большем, чем 2 автокорреляционные коэффициенты  $\rho^s$  (или ковариации остатков) приближаются к нулю.

Для расчета величины  $\rho$  циклический коэффициент корректируется на величину смещения. В результате имеем следующие формулы для расчетов:

$$r_y^{кор.} = \frac{n}{n-1} \cdot \frac{\sum_{t=2}^n u_t u_{t-1}}{\sum_{t=1}^n u_t^2} + \frac{m+1}{n},$$

где  $\frac{m+1}{n}$  – величина смещения ( $m$  – число независимых переменных),

или

$$r_y^{кор.} = r_y + \frac{r_y + \lambda}{n-1 - \frac{1+r_y\lambda}{1-r_y\lambda}}.$$

Параметр  $\lambda$  рассчитывается с помощью формулы

$$\lambda = \frac{\sum_{t=2}^n x_t x_{t-1}}{\sum_{t=2}^n x_{t-1}^2},$$

где  $x_t$  – отклонение переменной  $x$  от своего среднего значения.

Для матрицы  $V = \sigma_u^2 Z$  остаточная дисперсия вычисляется как

$$\sigma_u^2 = \frac{u'u}{n-k},$$

а с учетом смещения как

$$\sigma_u^2 = \frac{u'u}{n-k} \left[ n - \frac{1+\lambda\rho}{1-\lambda\rho} \right].$$

Алгоритм расчета параметров модели методом Эйткена распадается на пять шагов:

1-й шаг. Расчет параметров уравнения регрессии методом 1МНК и расчет остатков.

2-й шаг. Исследование остатков на наличие автокорреляции.

3-й шаг. Формирование матрицы ковариации остатков  $Z$  или  $V$ .

4-й шаг. Обращение матрицы  $Z$  или  $V$ .

5-й шаг. Оценка параметров модели методом Эйткена, используя выражения

$$\hat{A} = (X'Z^{-1}X)^{-1}X'Z^{-1}Y$$

или

$$\hat{A} = (X'V^{-1}X)^{-1}X'V^{-1}Y.$$

### **Метод преобразования исходной информации**

Рассмотрим альтернативный подход к расчету параметров уравнения регрессии для случая, когда имеет место автокорреляция остатков первого порядка. Алгоритм метода распадается на два шага:

1-й шаг. Преобразование исходной информации с помощью параметра  $\rho$ .

2-й шаг. Использование метода 1МНК для получения оценок параметров уравнения регрессии.

Матрица преобразования исходной информации  $T$  выглядит следующим образом:

$$T = \begin{pmatrix} \sqrt{1-\rho^2} & 0 & 0 & \dots & 0 \\ -\rho & 1 & 0 & \dots & 0 \\ 0 & -\rho & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix}.$$

С учетом данной матрицы уравнение регрессии преобразовывается к виду

$$TY = TX\hat{A} + Tu.$$

В этом случае дисперсионная матрица будет отвечать необходимым требованиям постоянства дисперсий остатков

$$M(Tuu'T') = \sigma_u^2 E.$$

Преобразованные данные имеют вид:

$$TY = \begin{pmatrix} \sqrt{1-\rho^2} y_1 \\ y_2 - \rho y_1 \\ y_3 - \rho y_2 \\ y_4 - \rho y_3 \\ \vdots \\ y_n - \rho y_{n-1} \end{pmatrix},$$

$$TX = \begin{pmatrix} \sqrt{1-\rho^2} & \sqrt{1-\rho^2}x_1^1 & \dots & \sqrt{1-\rho^2}x_1^m \\ 1-\rho & x_2^1-\rho x_1^1 & \dots & x_2^m-\rho x_1^m \\ \dots & \dots & \dots & \dots \\ 1-\rho & x_n^1-\rho x_{n-1}^1 & \dots & x_n^m-\rho x_{n-1}^m \end{pmatrix}.$$

В ряде случаев используется матрица  $T_1$  размерностью  $(n-1) \times n$ , которая получается из матрицы  $T$  путем вычеркивания из нее первой строки:

$$T_1 = \begin{pmatrix} -\rho & 1 & 0 & \dots & 0 \\ 0 & -\rho & 1 & \dots & 0 \\ 0 & 0 & -\rho & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

В общем случае, когда порядок автокорреляционной модели не известен, указанные методы не используются для расчета параметров уравнения регрессии.