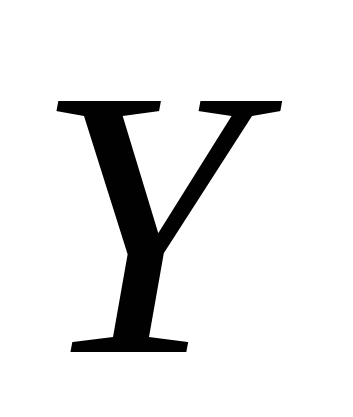
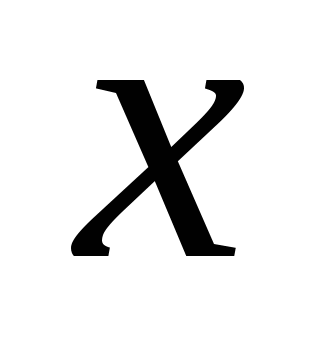
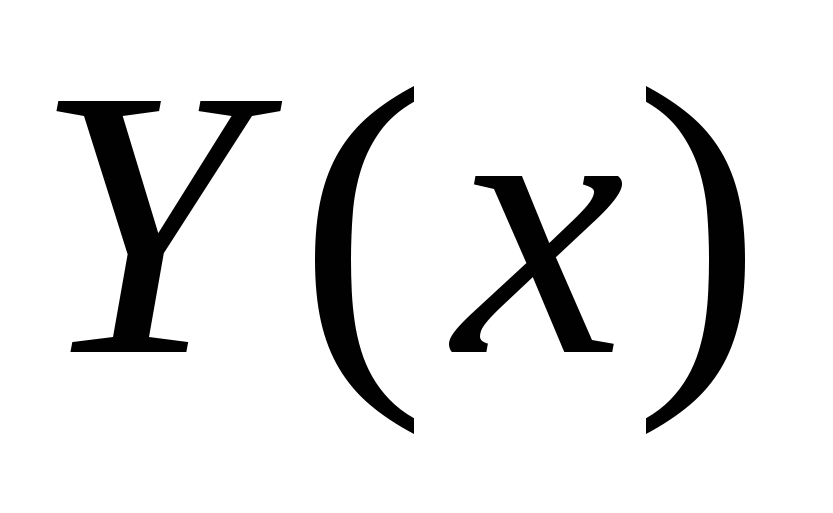
**Лекция №6-7.** *Корреляционный и регрессионный анализ.*

Понятие корреляции появилось в середине XIX века в работах английских статистиков Ф. Гальтона и К. Пирсона. Этот термин произошел от латинского "correlatio" - соотношение, взаимосвязь. Понятие регрессии (латинское "regressio" - движение назад) также введено Ф. Гальтоном, который, изучая связь между ростом родителей и их детей, обнаружил явление "регрессии к среднему" - рост детей очень высоких родителей имел тенденцию быть ближе к средней величине.

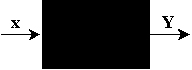
Теория и методы корреляционного анализа используются для выявления связи между случайными переменными и оценки ее тесноты. Основной задачей регрессионного анализа является установление формы и изучение зависимости между переменными.

В общем случае две величины могут быть связаны функциональной зависимостью, либо зависимостью другого рода, называемой статистической, либо быть независимыми. Статистической называется зависимость, при которой изменение одной из величин влечет изменение распределения другой. Статистическая зависимость, при которой изменение одной из величин влечет изменение среднего значения другой, называется корреляционной.

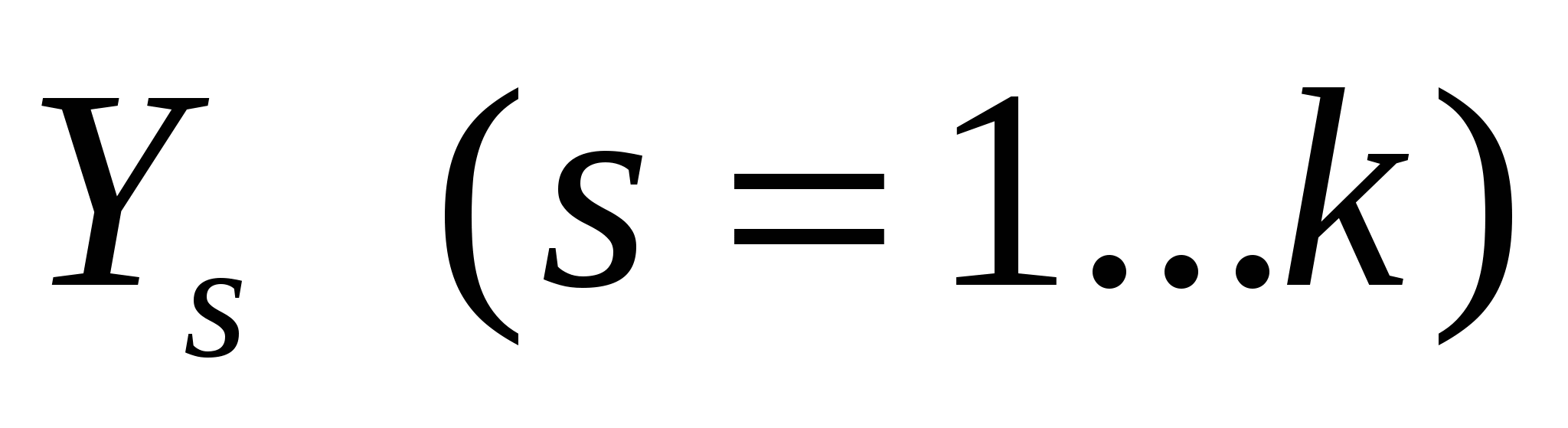
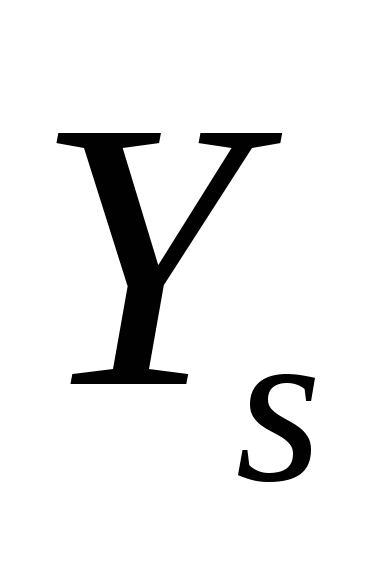
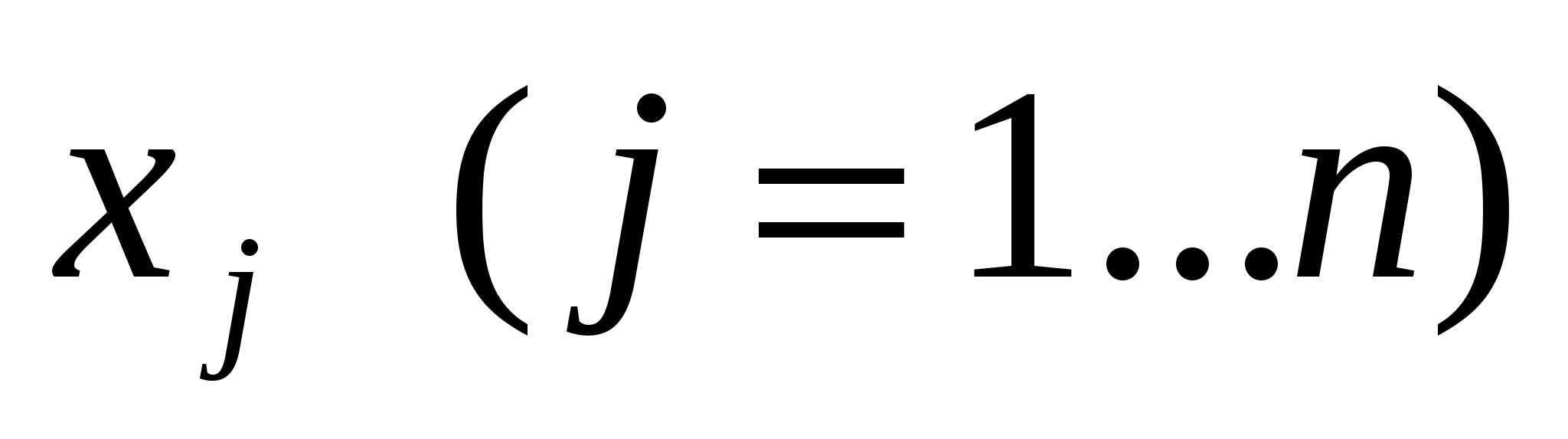
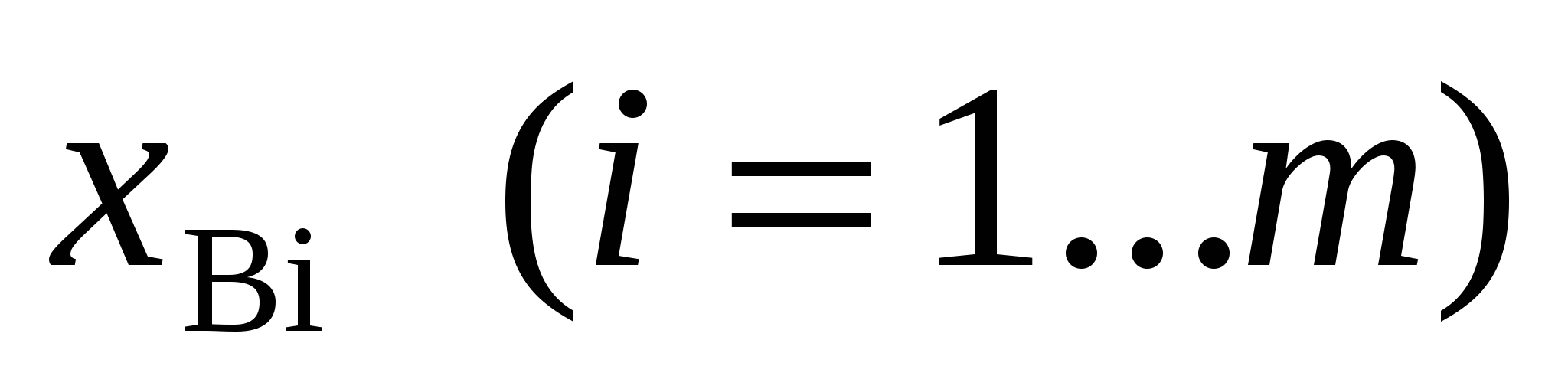
Между величинами, характеризующими явления менеджмента качества, в большинстве случаев существуют зависимости, отличные от функциональных зависимостей. Пусть, например, мы рассматриваем зависимость величины от величины–. Невозможность выявления строгой связи между двумя переменными объясняется тем, что значение зависимой переменной определяется не только значением переменной , но и другими (неконтролируемыми или неучтенными) факторами, а также тем, что измерение значений переменных неизбежно сопровождается некоторыми случайными ошибками. Вследствие этого корреляционный анализ широко используется при установлении взаимосвязи показателей качества и факторов влияющих на них.

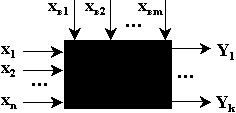
Если с увеличением значение зависимой переменной в среднем увеличивается, то такая зависимость называется прямой или положительной. Если среднее значение при увеличении уменьшается, имеет место отрицательная или обратная корреляция. Если с изменением значения в среднем не изменяются, то говорят, что корреляция – нулевая.

Часто при исследовании взаимосвязи между какими-либо показателями, представляют изучаемый объект в виде так называемого "черного (кибернетического) ящика". Самый простой случай – изучение связи между одной переменной , которую называют фактором (входной переменной, независимой переменной), и переменной , которую называют откликом (реакцией, показателем, зависимой переменной). Ситуации соответствует рисунок 6.1.



*Рис. 6.1. Исследуемая система в виде "черного ящика" (один фактор, один отклик)*

Однако в общем случае итогом функционирования системы является целый набор результирующих величин . При этом значения откликов определяются, с одной стороны, совокупностью факторов , а, с другой стороны, набором возмущений (случайных, неконтролируемых факторов). Такую ситуацию иллюстрирует рисунок 6.2.



*Рис. 6.2. Исследуемая система в виде "черного ящика" (общий случай)*

Основой статистических методов корреляционный и регрессионный анализ является один из семи простых инструментов контроля качества - диаграмма разброса (поле корреляции). Этот инструмент позволяет графически отобразить и в дальнейшем проанализировать вид и тесноту связи между исследуемыми факторами.

Чаще всего двумерную диаграмму разброса строят для выявления связей между следующими классами показателей:

1. характеристика качества и влияющий на нее фактор;
2. две различных характеристики качества;
3. два фактора, влияющие на одну характеристику качества.

Прежде чем начать исследование стохастической зависимости, необходимо убедится, что массив данных характеризует наличие только двух переменных, корреляционные связи которых надо раскрыть. То есть надо проанализировать собранную информацию на предмет расслоения данных измерения, проверить возможность вмешательства в одну из переменных дополнительного стратифицирующего фактора.

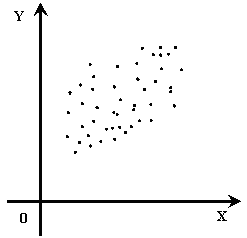
Построение поля корреляциисводится к следующим этапам:

1. Сбор не менее 25 пар данных исследуемых параметров в таблицу;
2. Нахождение максимальных и минимальных значений и . Выбор шкалы на горизонтальной и вертикальной оси так, чтобы длины рабочих областей были примерно равны.
3. Построение на отдельном листе координатной плоскости. Если исследуется влияние фактора на показатель качества, то фактор располагают по оси абсцисс, а показатель – по оси ординат; и нанесение собранных пар данных (в случае совпадения точек они либо располагаются максимально близко, либо обозначаются окружностями около первоначальной точки)
4. На диаграмму наносятся все необходимые обозначения:

* название диаграммы;
* интервал времени сбора данных;
* число пар данных;
* название и единицы для каждой оси;
* идентифицирующая информация составителя диаграммы.

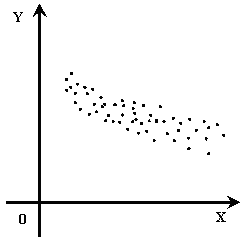
Анализ данной диаграммы начинают с формирования общего представления распределения совокупности исследуемых данных, затем проводится анализ на наличие выбросов (далеко отстоящих точек), которые, скорее всего, связанны либо с ошибками сбора данных, либо с изменениями условий работы. После анализа появления таких точек их можно исключить из диаграммы. После этого на поле корреляции распределение, скорее всего, будет соответствовать одному из типовых:

Если точки корреляционного поля образуют эллипс, главная диагональ которого имеет положительный угол наклона, то имеет место положительная корреляция (пример подобной ситуации можно видеть на рисунке 6.3).



*Рис. 6.3. Положительная корреляция*

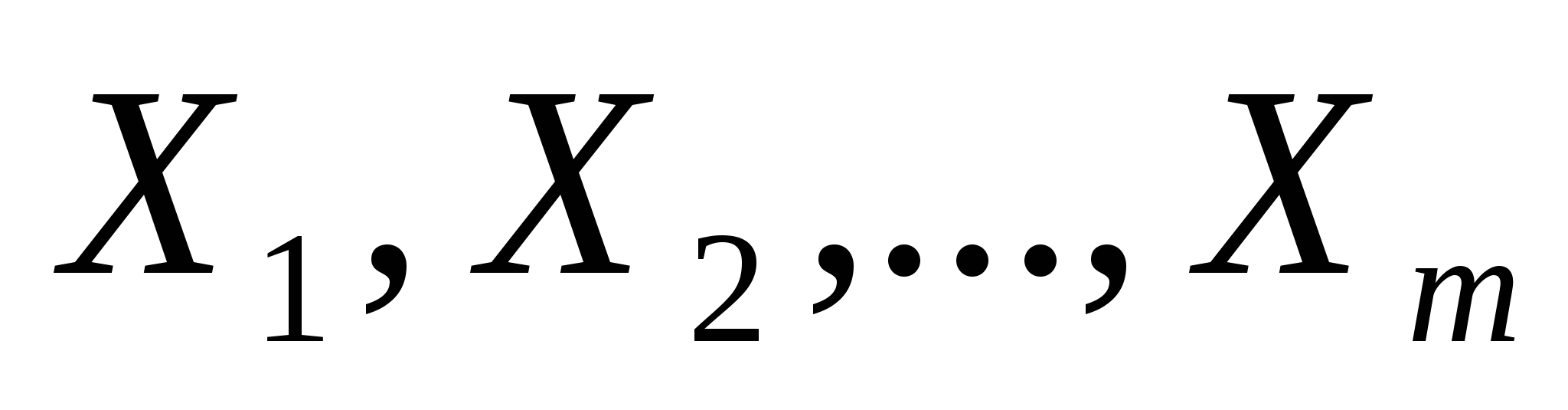
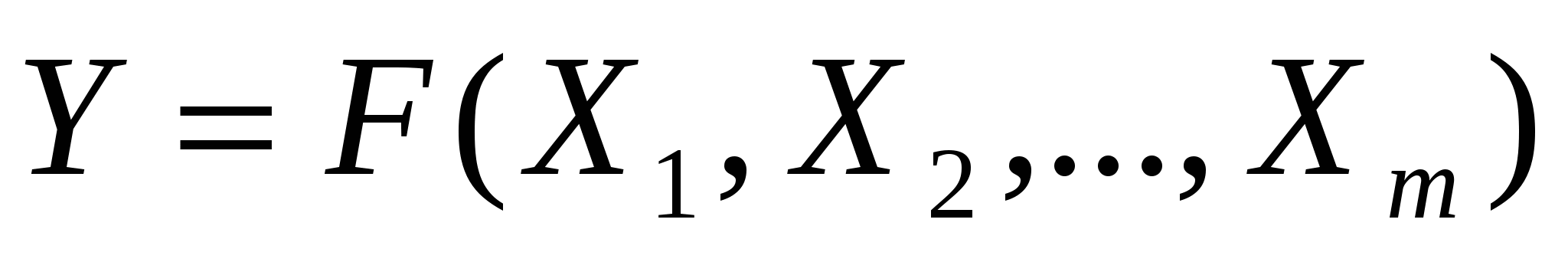
Если точки корреляционного поля образуют эллипс, главная диагональ которого имеет отрицательный угол наклона, то имеет место отрицательная корреляция (пример изображен на рисунке 6.4).



*Рис. 6.4. Отрицательная корреляция*

Если расположение точек по внешнему виду напоминает одну из нелинейных функций, то говорят, что наблюдается криволинейная корреляция.

Если же в расположении точек нет какой-либо закономерности, то говорят, что в этом случае наблюдается нулевая корреляция.

После визуального анализа распределения переходят к анализу, основанному на расчете корреляционных параметров. Таким образом, исследование зависимости с помощью поля корреляции и корреляционный анализ являются начальными этапами регрессионного анализа, целью которого является установление функциональной зависимости величины от величин , выраженной в виде уравнения регрессии (регрессионной модели):.

Полный регрессионный анализ включает следующие этапы:

1. Определение вида функции, описывающей функциональную связь между результативным признаком и факторными признаками (этап спецификации);

Выбор модели регрессии может производиться как на основе априорных исследований, так и на основе апостериорных исследований. Модели, в зависимости от вида функции , делятся на линейные модели и нелинейные модели; а также на однофакторные модели (парная модель регрессии) и многофакторные модели (модель множественной регрессии).

1. Определение коэффициентов регрессии (этап идентификации).

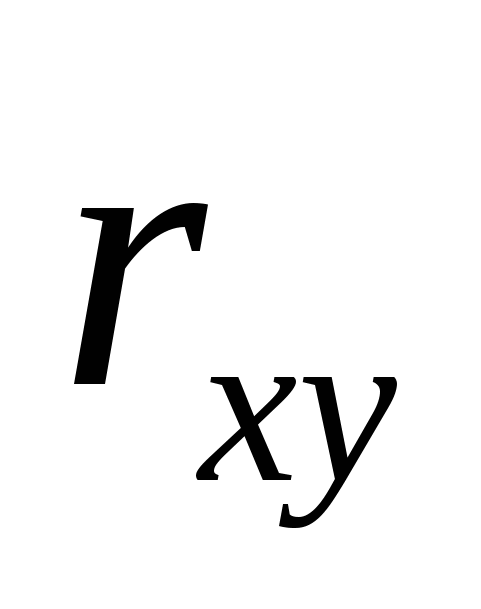
Параметры, входящие в модель регрессии, находятся с использованием методики аппроксимации по критерию наименьших квадратов.

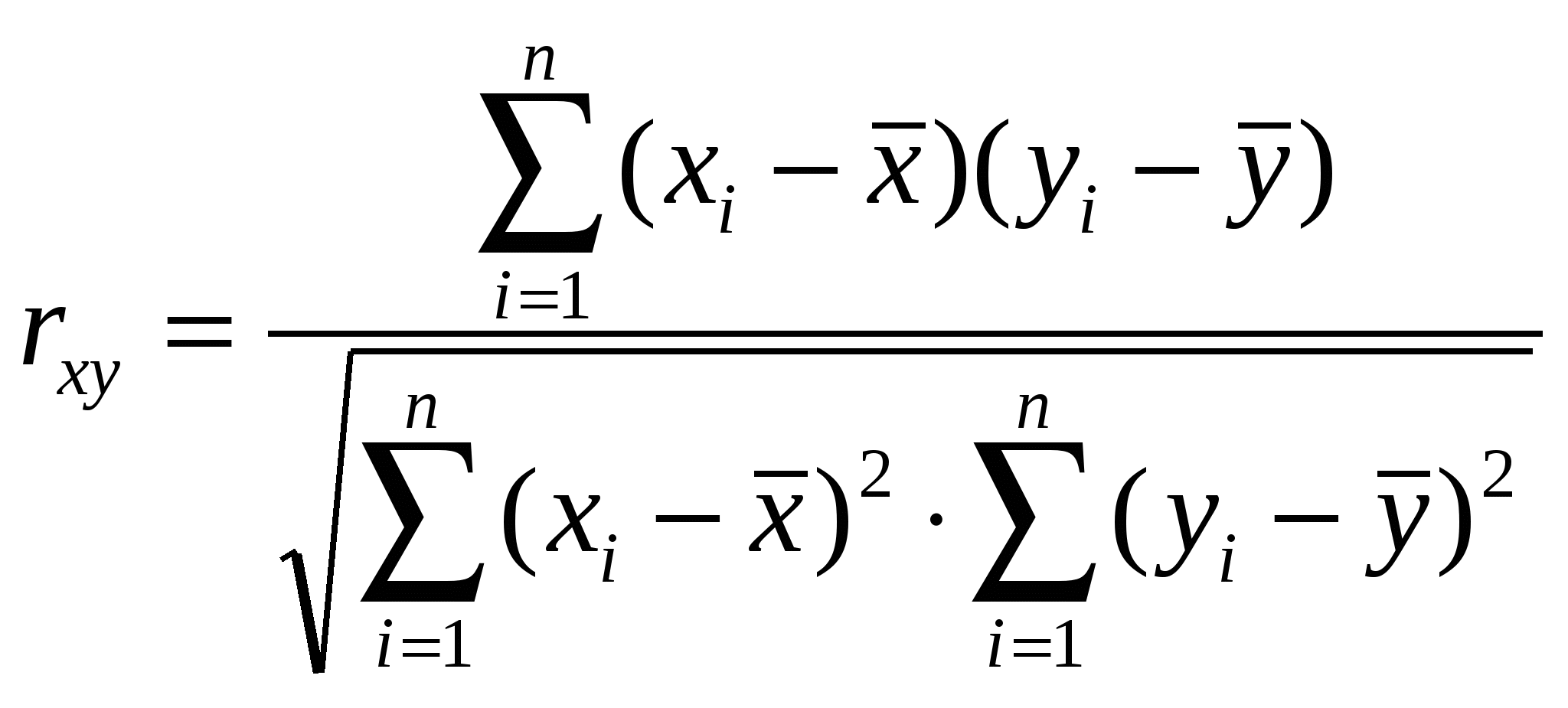
1. Расчет теоретических значений результативного признака для отдельных наборов значений факторов;
2. Исследование отклонений расчетных значений от эмпирических данных;
3. Оценка качества полученной модели и проверка соответствующих статистических гипотез о регрессии (этап верификации).

Оценка качества модели проводится на основе гипотез о значимости модели в целом и каждого ее параметра, доверительных интервалах и анализе остатков. Анализ остатков позволяет получить представление, насколько хорошо подобрана сама модель и насколько правильно выбран метод оценки коэффициентов. Согласно общим предположениям регрессионного анализа остатки должны вести себя как независимые случайные величины. Часто анализ графика остатков может показать на наличие тенденций или выбросов. Наличие определенных зависимостей говорит о неправильности выбора вида модели. В случае с выбросами, устранение их эффектов может производиться либо удалением таких точек (цензурированием), либо использованием методов, устойчивых к грубым отклонениям.

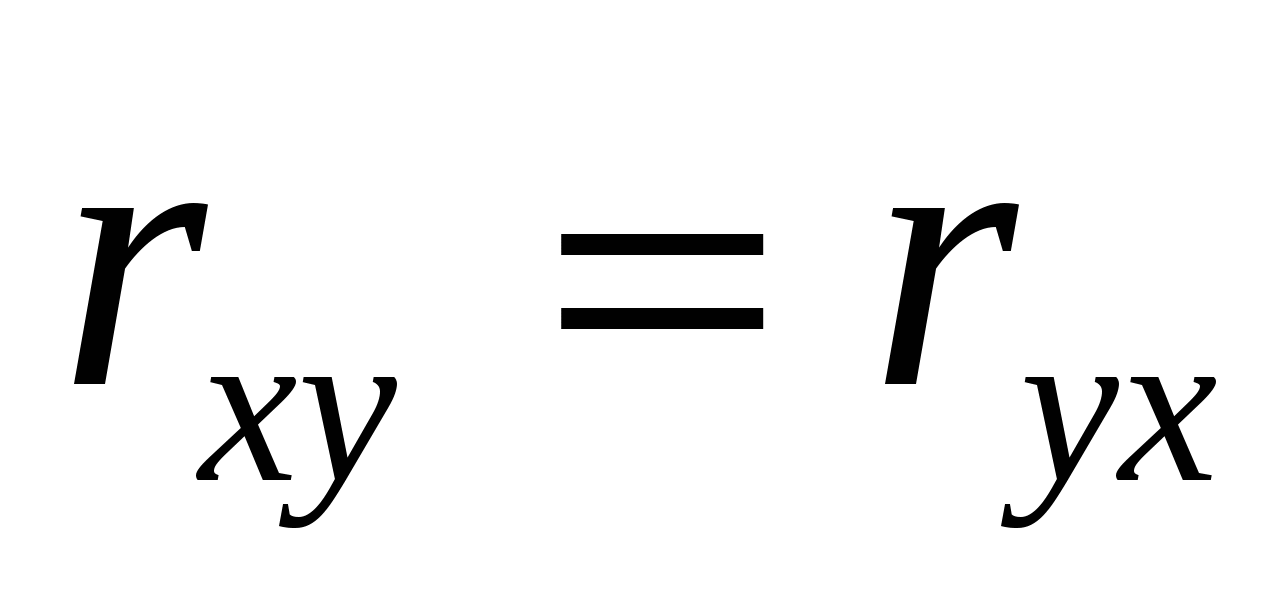
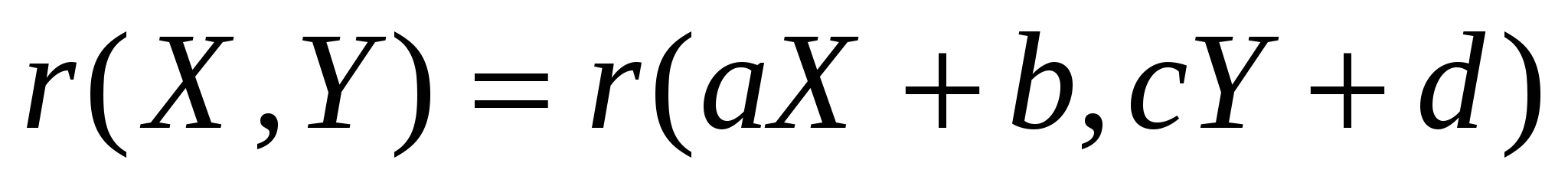
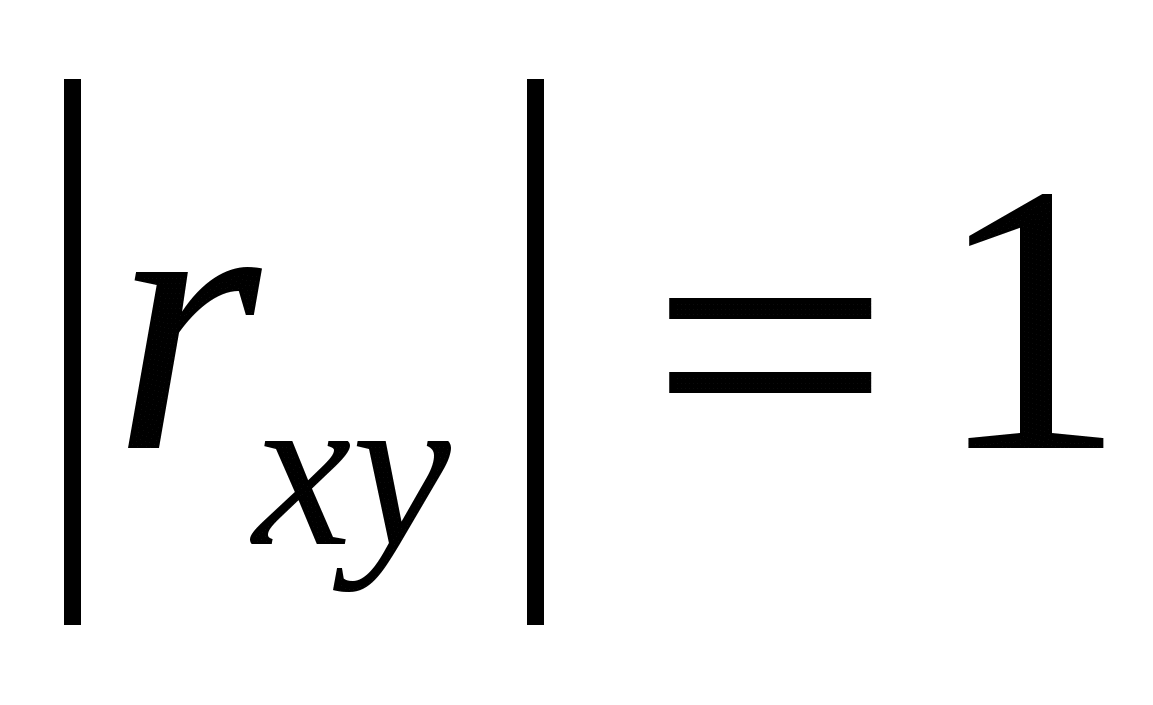
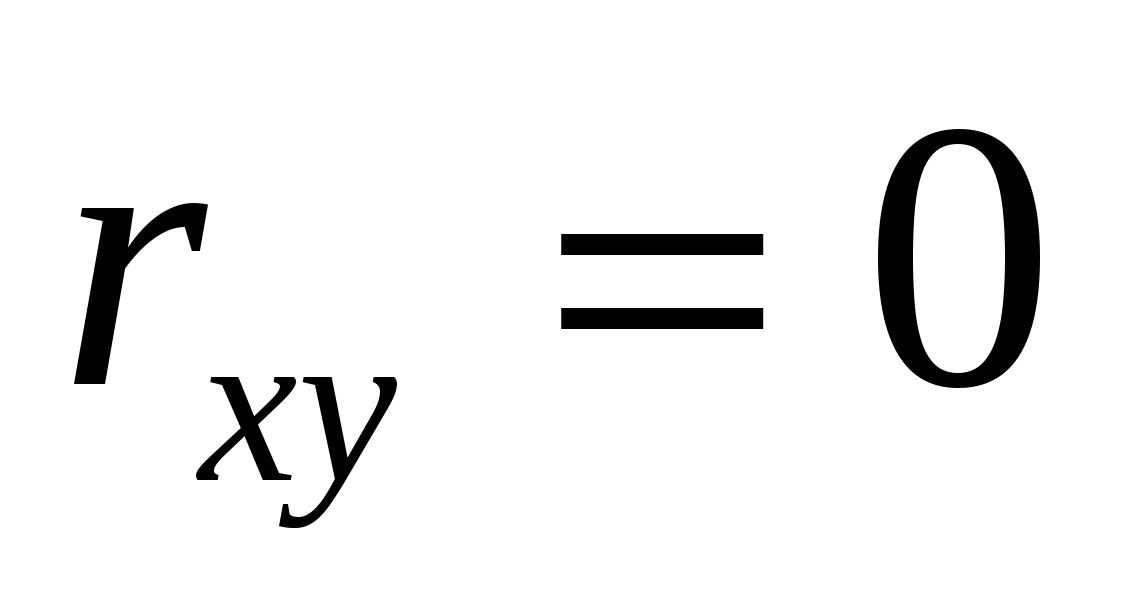
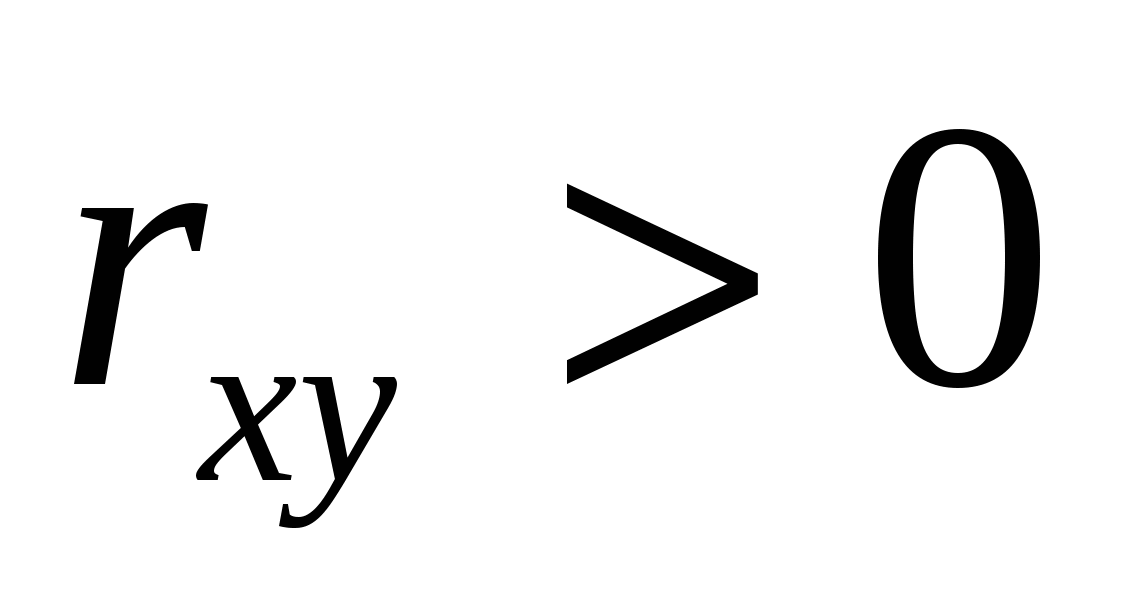
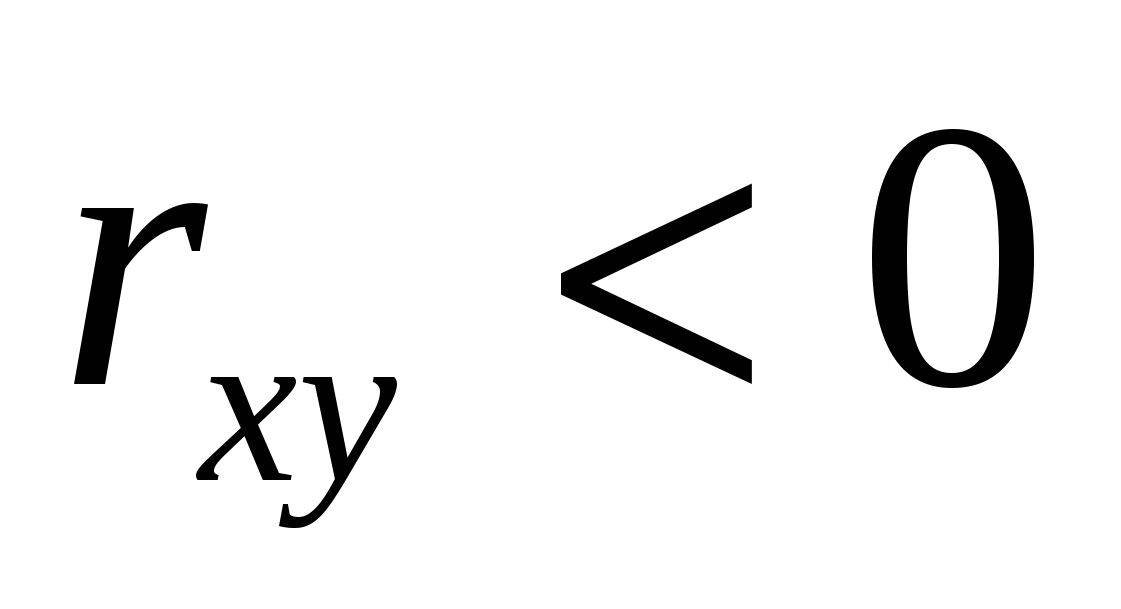
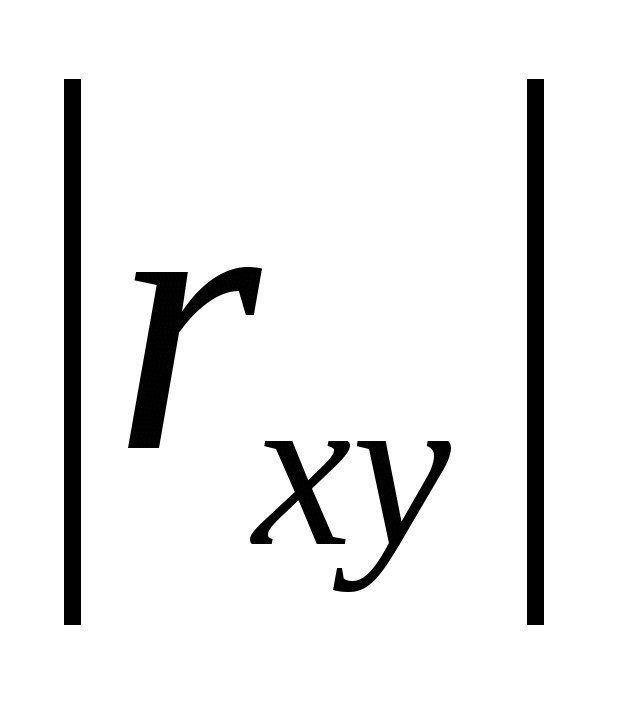
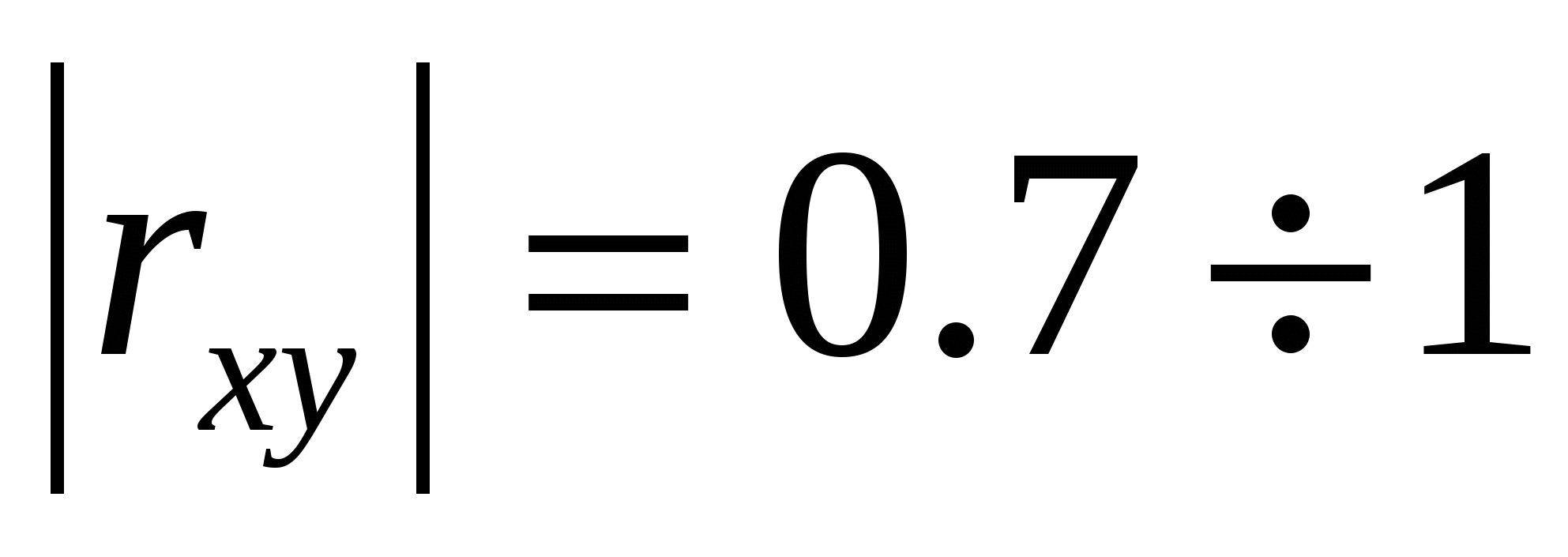
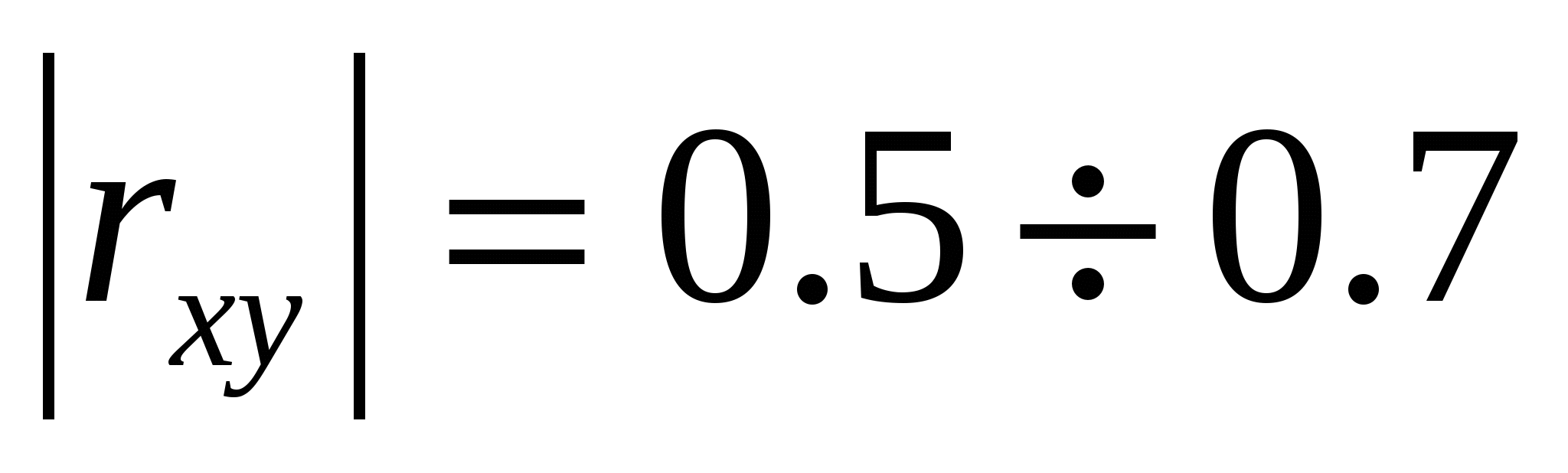
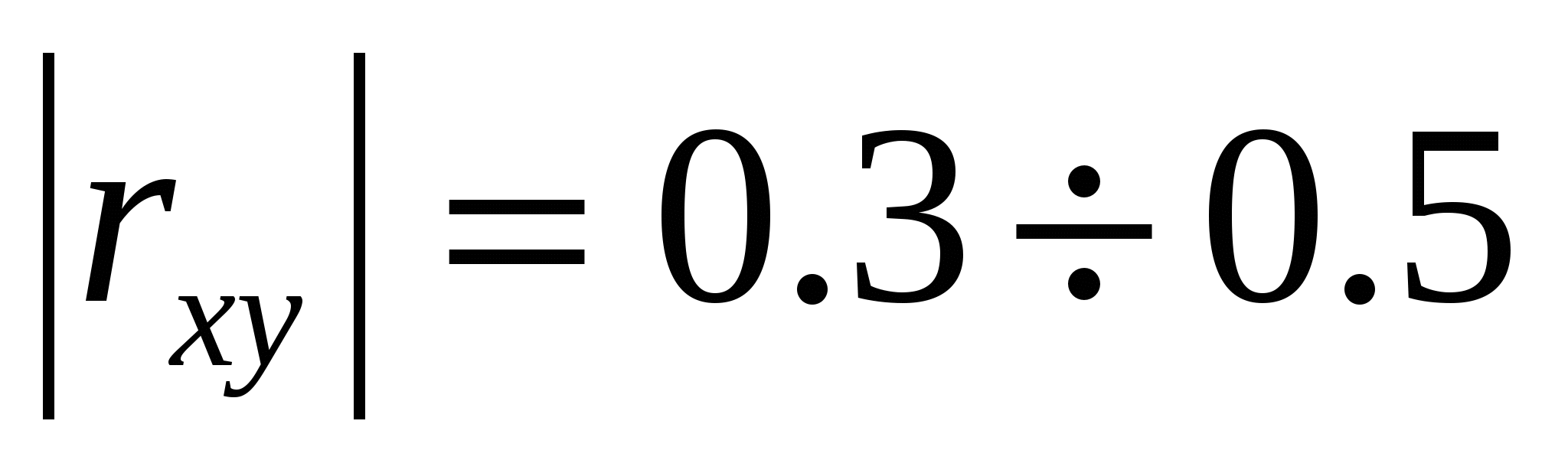
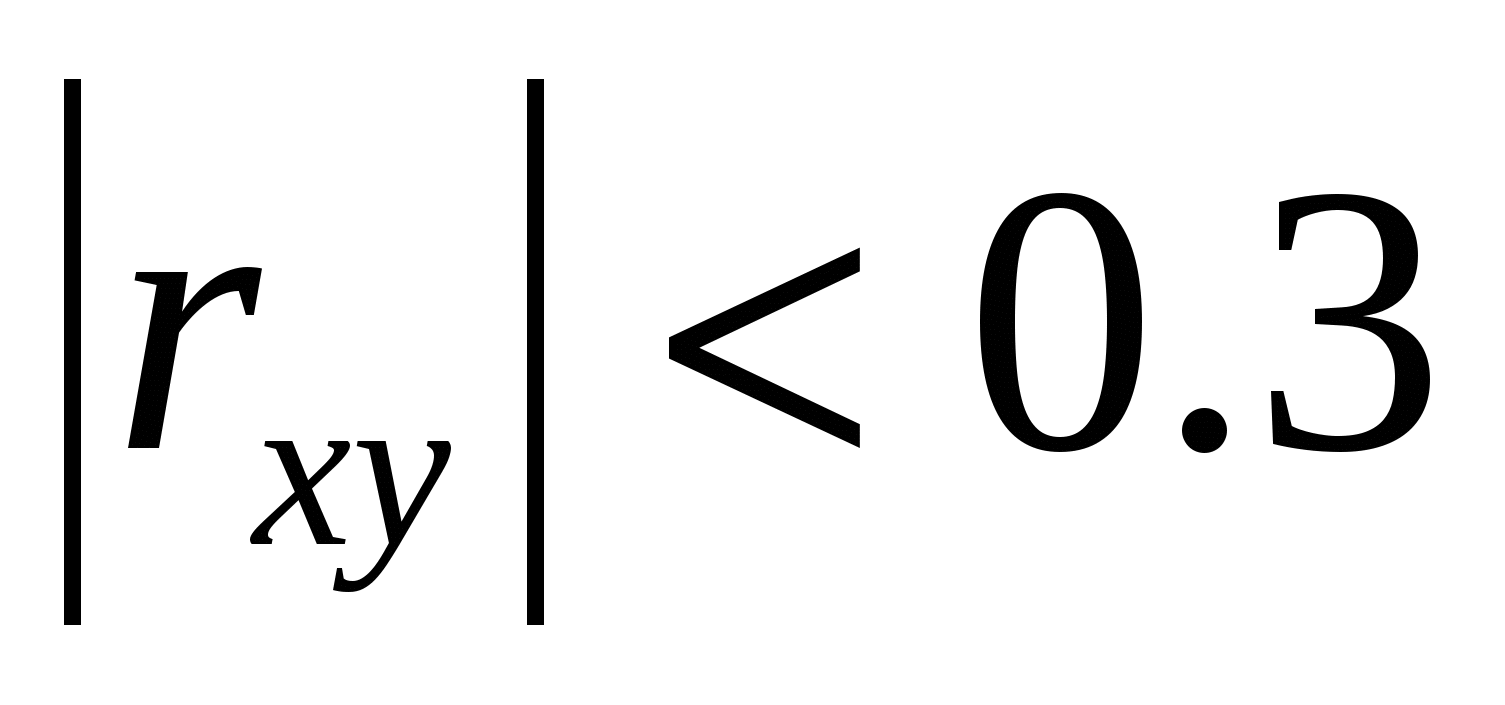
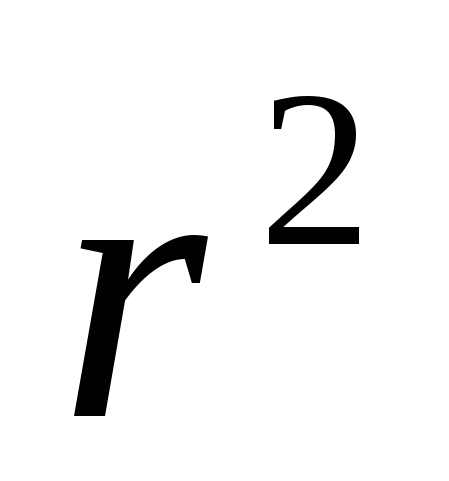
Однофакторные исследования

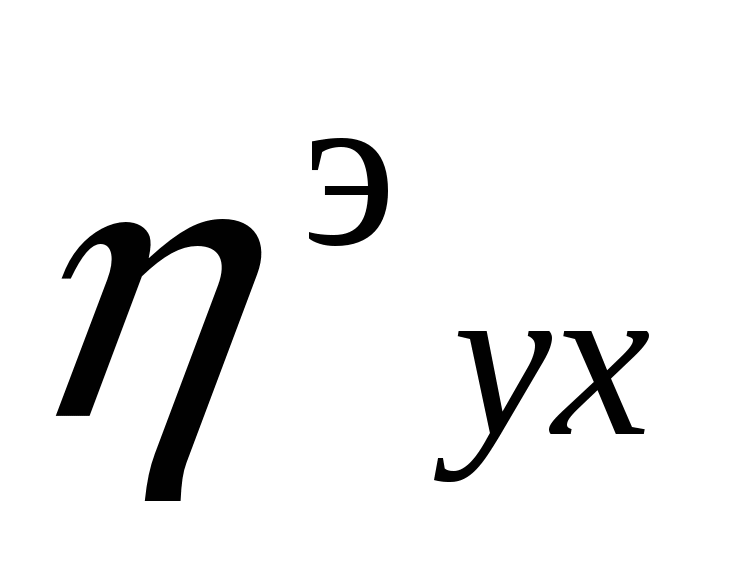
Корреляционный анализ сводиться к нахождению различных количественных показателей взаимосвязей. При исследовании однофакторных зависимостей самыми распространенными являются:

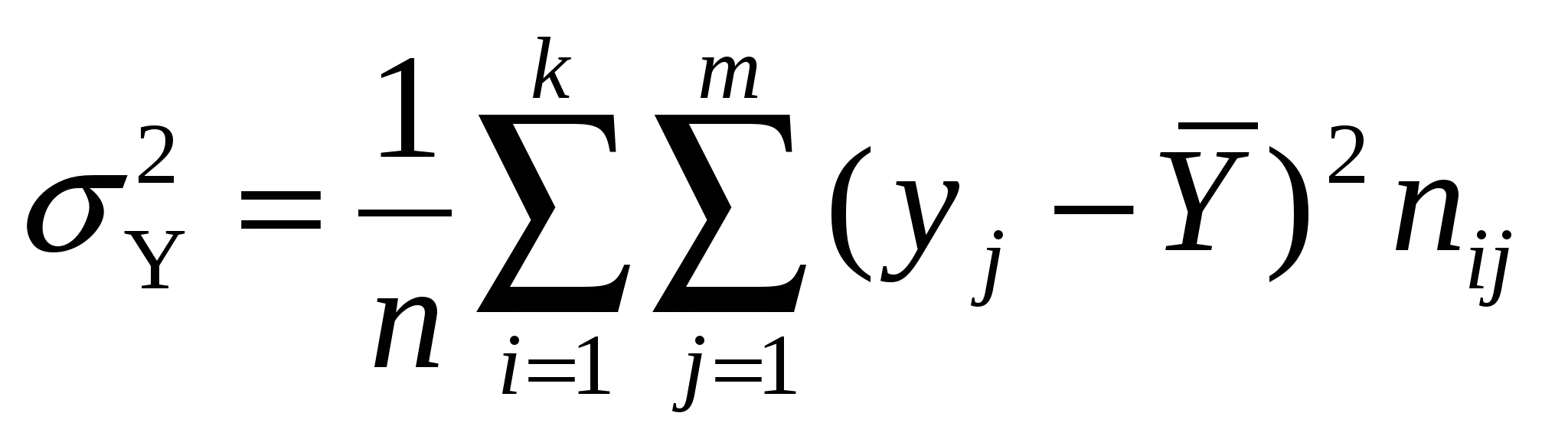
1. Коэффициент линейной корреляции выражает степень тесноты линейной связи между двумя случайными величинами и вычисляется по выборочным данным по формуле:

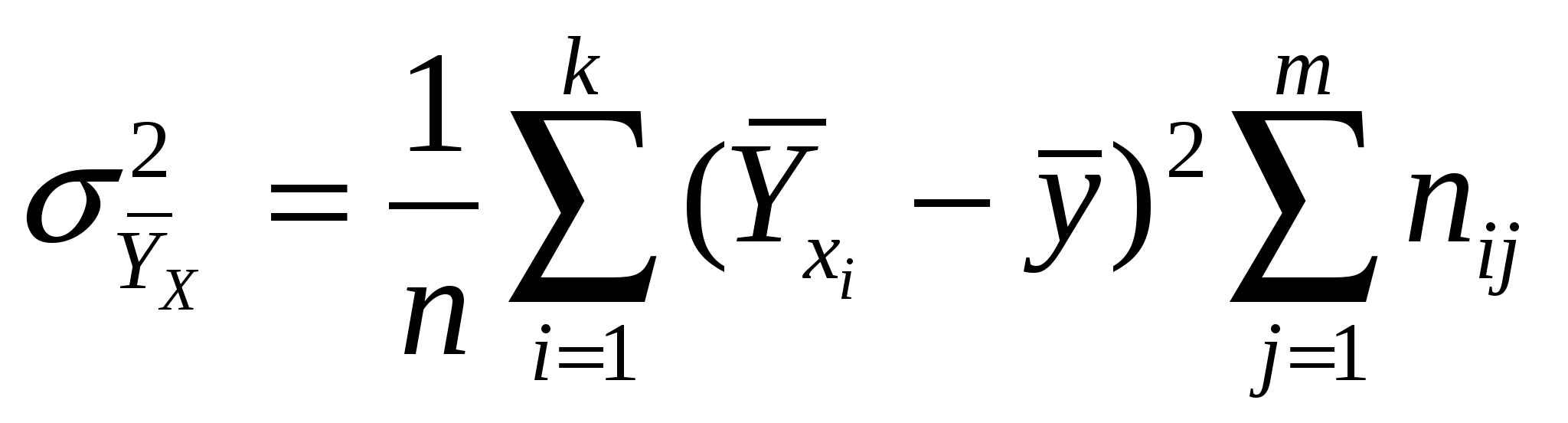
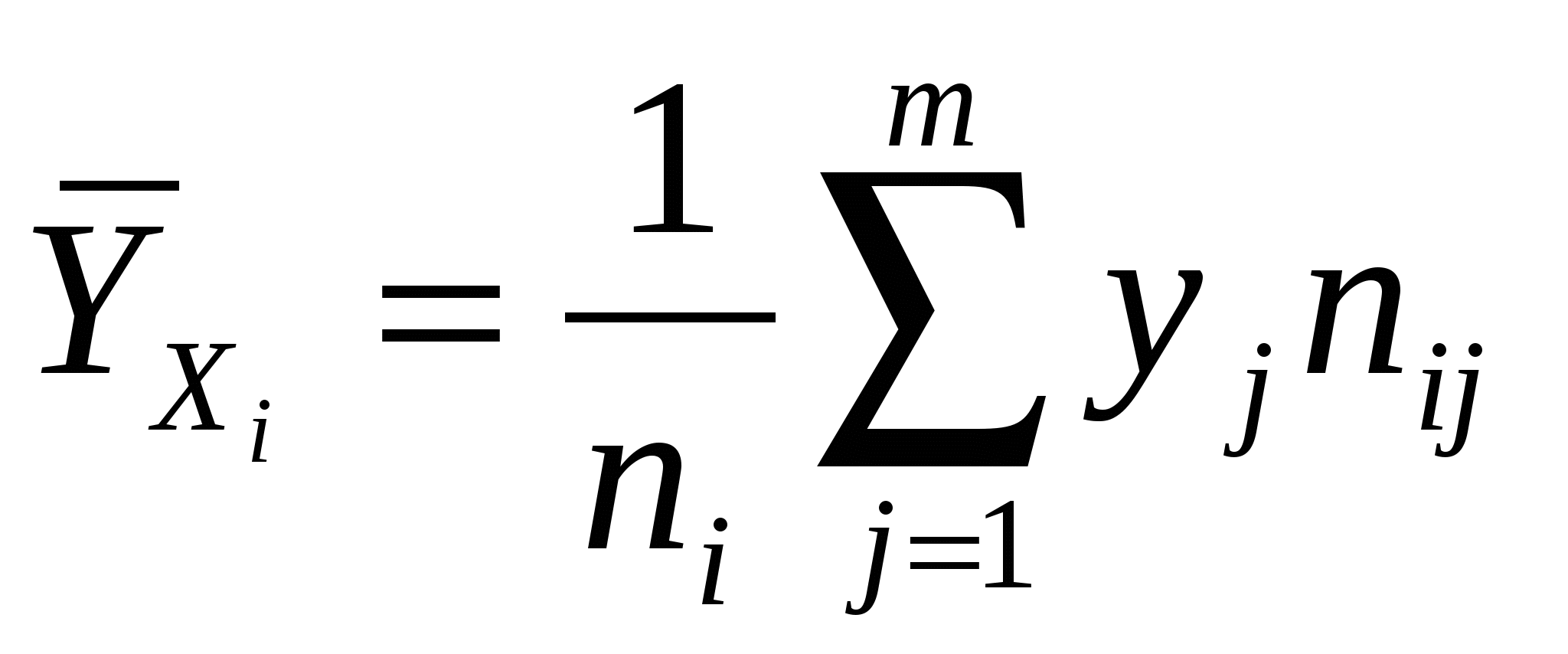


Линейный коэффициент корреляции обладает следующими свойствами и характеристиками:

* не имеет размерности, следовательно, сопоставим для величин различных порядков;
* 
* 
* если , то величины связаны линейной функциональной зависимостью;
* если , то между величинами нет линейной корреляционной зависимости, однако это не исключает существования другого вида корреляционной зависимости;
* Если , то связь – прямая корреляция, если- обратная корреляция;
* Чем больше , тем теснее зависимость. При этом связь сильная при; связь умеренная при; связь слабая при; связь практически отсутствует при;
* Величина называетсякоэффициентом детерминации. Он определяет долю вариации одной из переменных, которая объясняется вариацией другой переменной.

2. Эмпирическое корреляционное соотношение применяется для оценки тесноты нелинейной связи между случайными величинами и вычисляется с использованием общей и межгрупповой дисперсий:

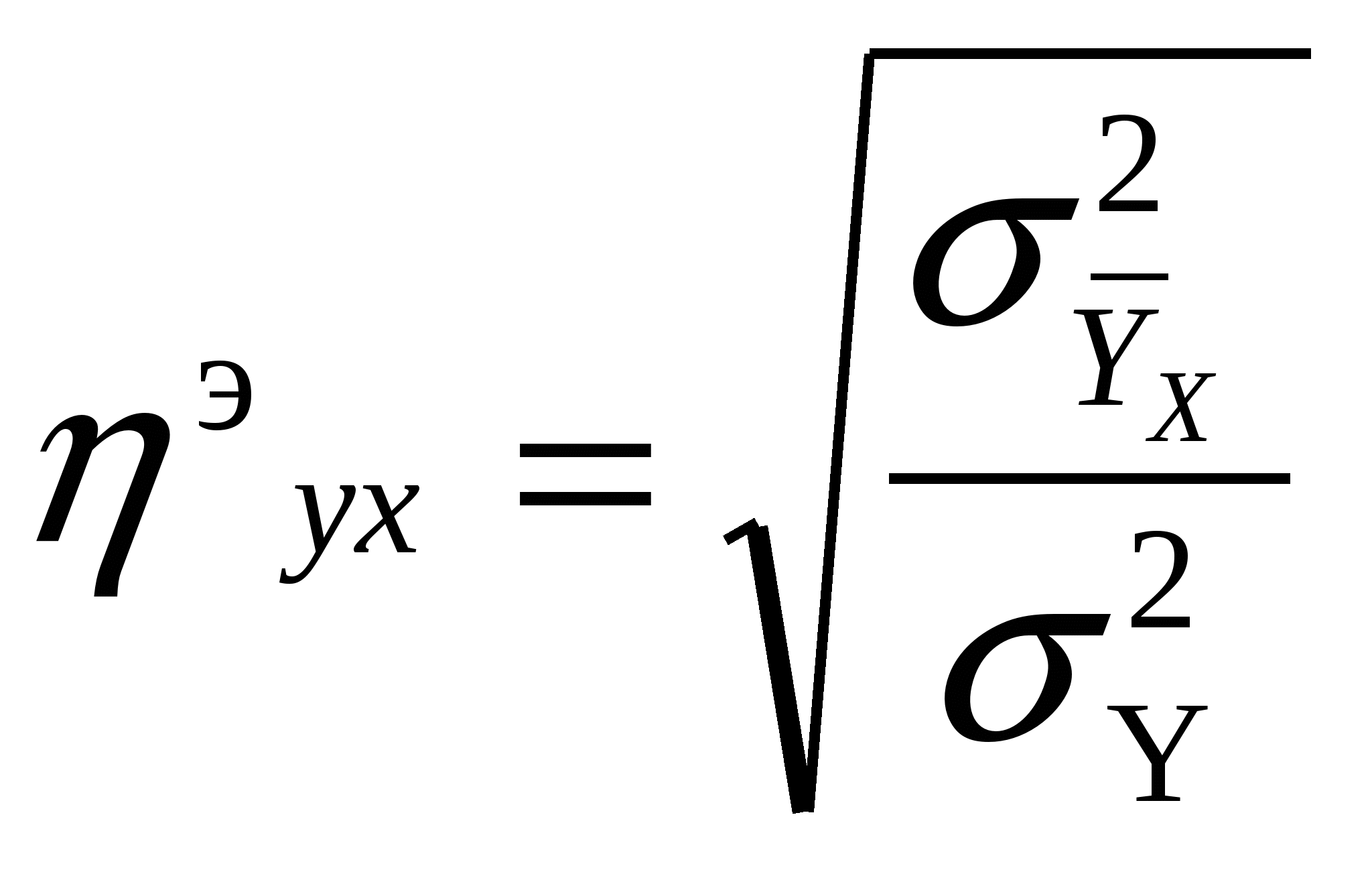
Общая дисперсия: 

Межгрупповая дисперсия: , где.

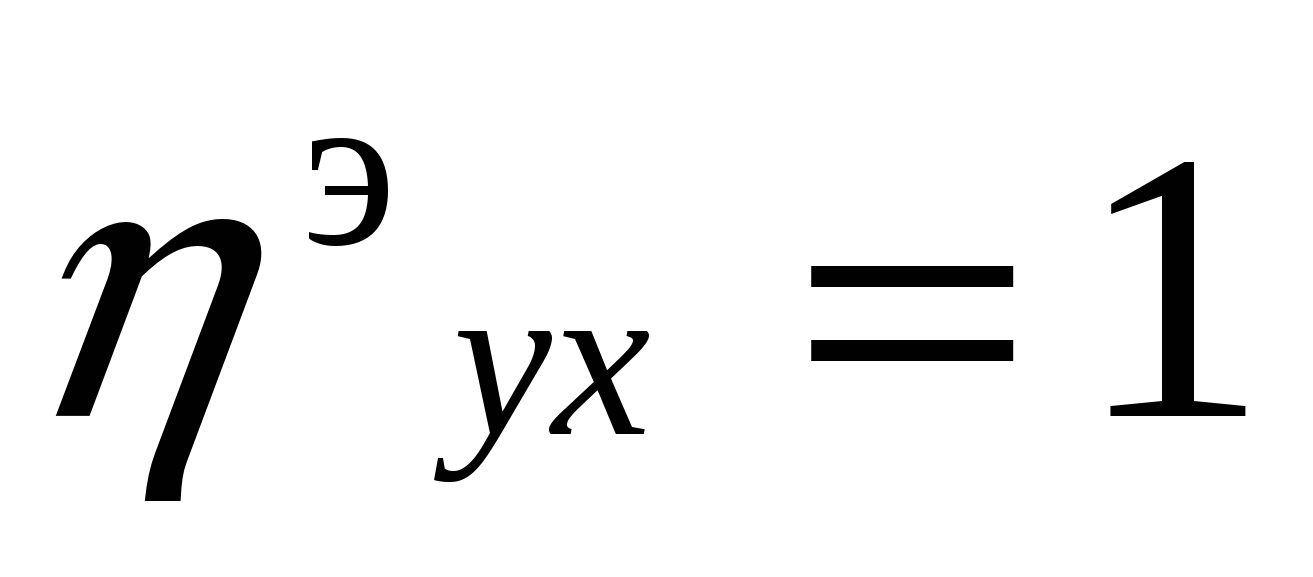
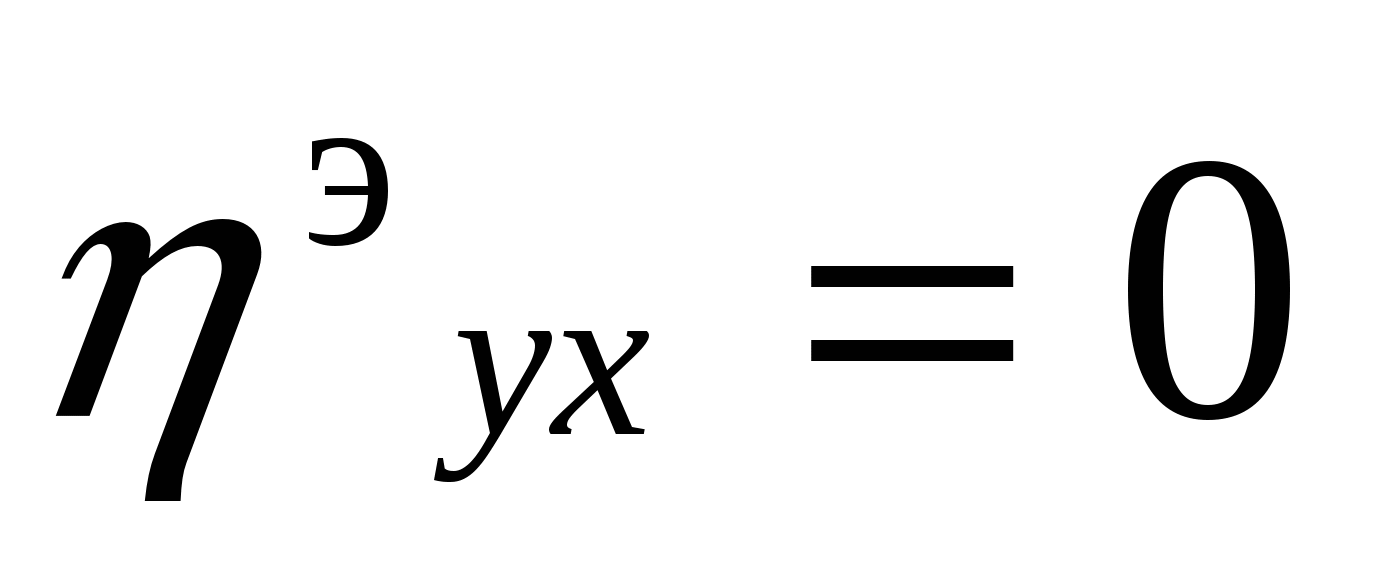
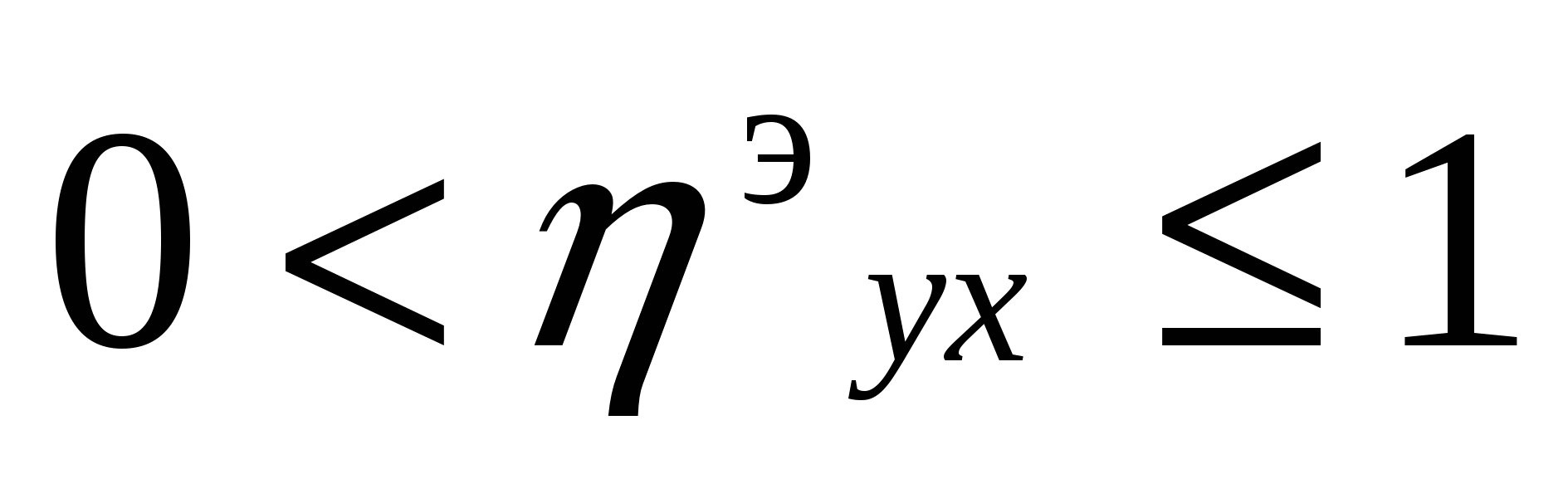
Для удобства расчетов данные могут быть представлены в виде таблицы:

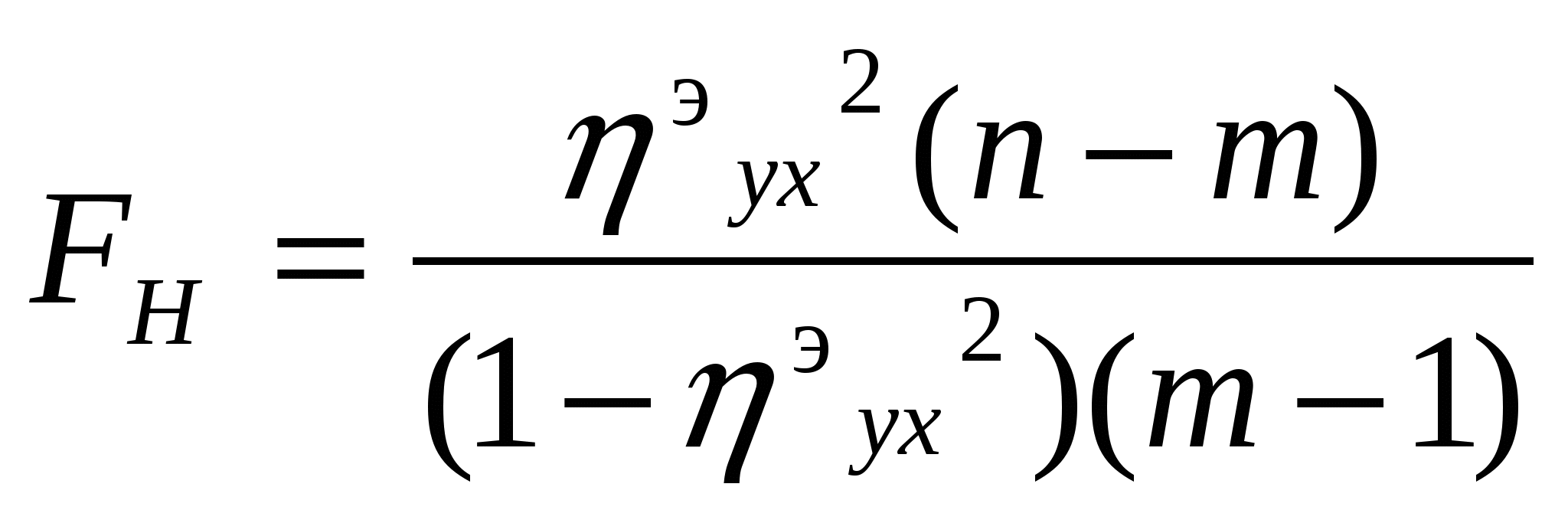
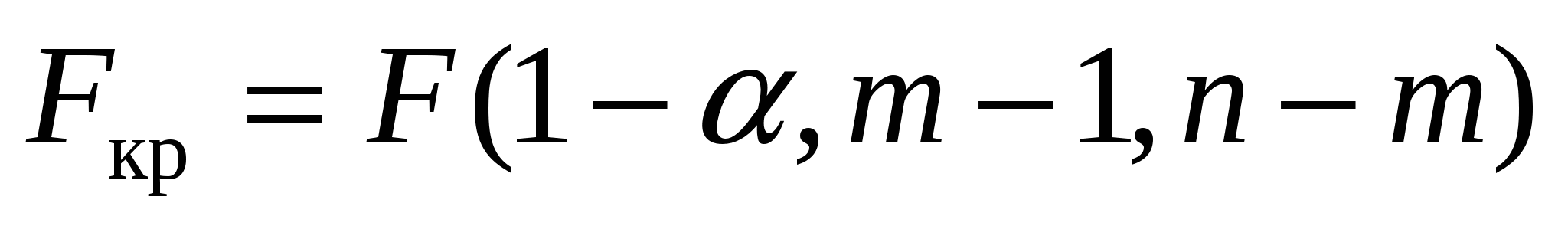
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| http://files.studfiles.ru/2706/176/html_Uir5GTcECm.pfrI/htmlconvd-IgQ3kx_html_6cb73b20.gif | http://files.studfiles.ru/2706/176/html_Uir5GTcECm.pfrI/htmlconvd-IgQ3kx_html_1650cc6f.gif | http://files.studfiles.ru/2706/176/html_Uir5GTcECm.pfrI/htmlconvd-IgQ3kx_html_183de386.gif | http://files.studfiles.ru/2706/176/html_Uir5GTcECm.pfrI/htmlconvd-IgQ3kx_html_2b2a08f4.gif |  | http://files.studfiles.ru/2706/176/html_Uir5GTcECm.pfrI/htmlconvd-IgQ3kx_html_2793a20a.gif |
| http://files.studfiles.ru/2706/176/html_Uir5GTcECm.pfrI/htmlconvd-IgQ3kx_html_m5a945c89.gif | http://files.studfiles.ru/2706/176/html_Uir5GTcECm.pfrI/htmlconvd-IgQ3kx_html_3dbb37e3.gif |  | http://files.studfiles.ru/2706/176/html_Uir5GTcECm.pfrI/htmlconvd-IgQ3kx_html_46af7212.gif |  | http://files.studfiles.ru/2706/176/html_Uir5GTcECm.pfrI/htmlconvd-IgQ3kx_html_554e02e1.gif |
|  |  |  |  |  |  |
| http://files.studfiles.ru/2706/176/html_Uir5GTcECm.pfrI/htmlconvd-IgQ3kx_html_5d4b5678.gif | http://files.studfiles.ru/2706/176/html_Uir5GTcECm.pfrI/htmlconvd-IgQ3kx_html_m8060e95.gif |  | http://files.studfiles.ru/2706/176/html_Uir5GTcECm.pfrI/htmlconvd-IgQ3kx_html_24d7c911.gif |  | http://files.studfiles.ru/2706/176/html_Uir5GTcECm.pfrI/htmlconvd-IgQ3kx_html_3e1e35ee.gif |
|  |  |  |  |  |  |
| http://files.studfiles.ru/2706/176/html_Uir5GTcECm.pfrI/htmlconvd-IgQ3kx_html_m1053a318.gif | http://files.studfiles.ru/2706/176/html_Uir5GTcECm.pfrI/htmlconvd-IgQ3kx_html_308a22f5.gif |  | http://files.studfiles.ru/2706/176/html_Uir5GTcECm.pfrI/htmlconvd-IgQ3kx_html_m96ee184.gif |  | http://files.studfiles.ru/2706/176/html_Uir5GTcECm.pfrI/htmlconvd-IgQ3kx_html_m74e291c8.gif |

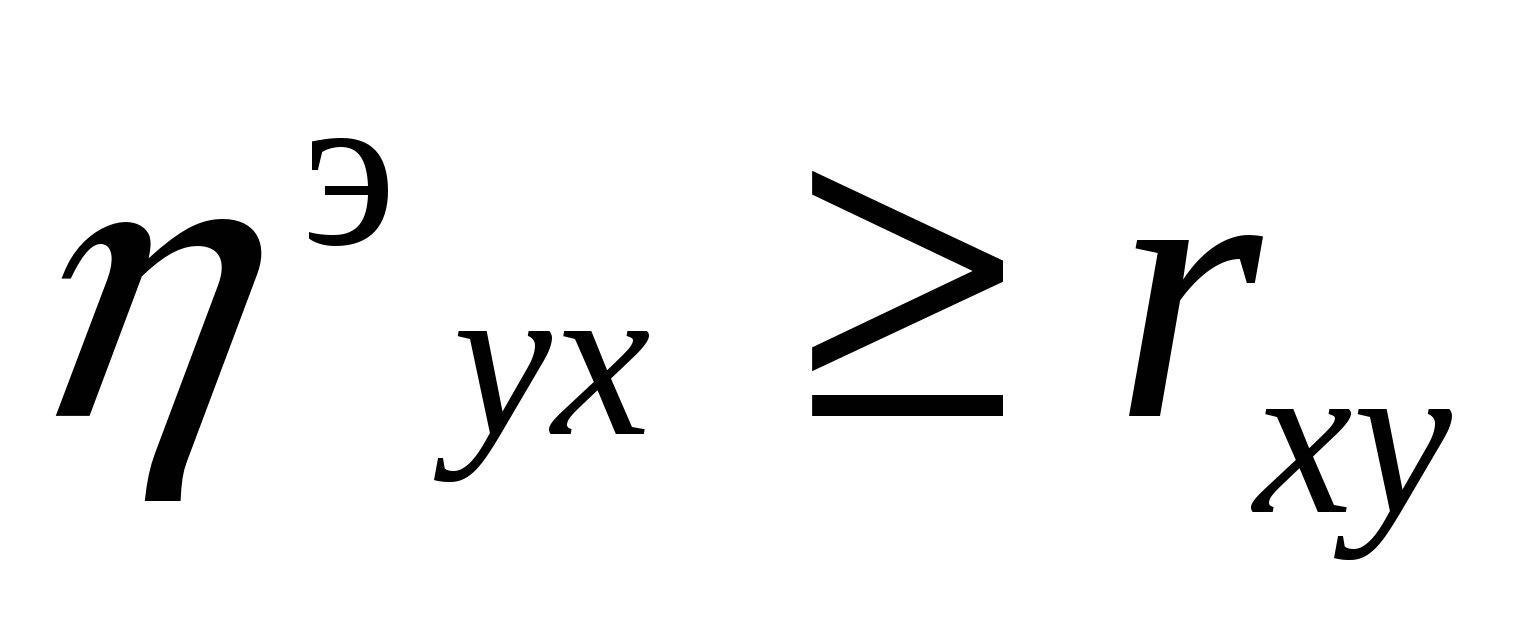
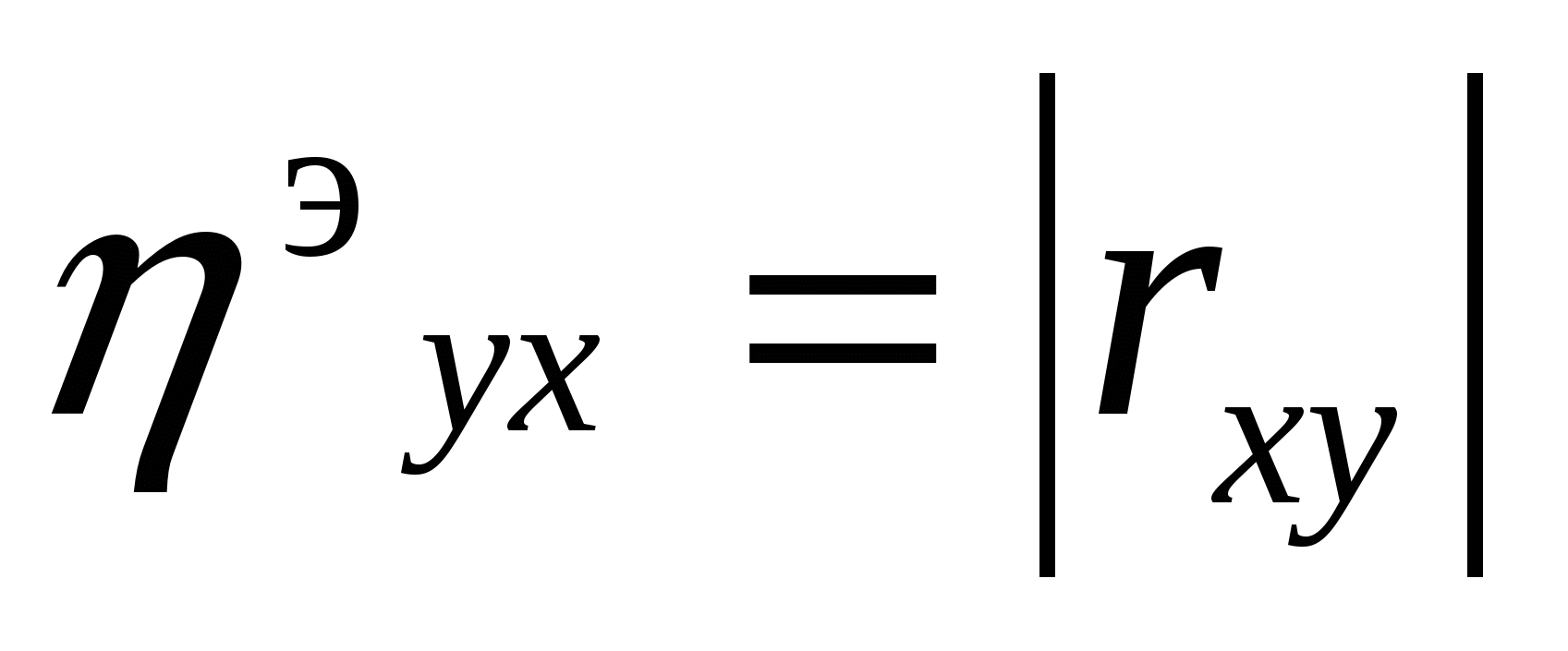
После всех дополнительных расчетов эмпирическое корреляционное соотношение можно рассчитать по формуле:

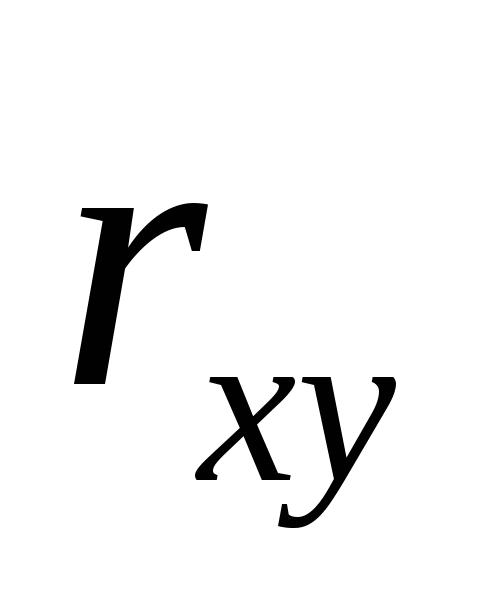


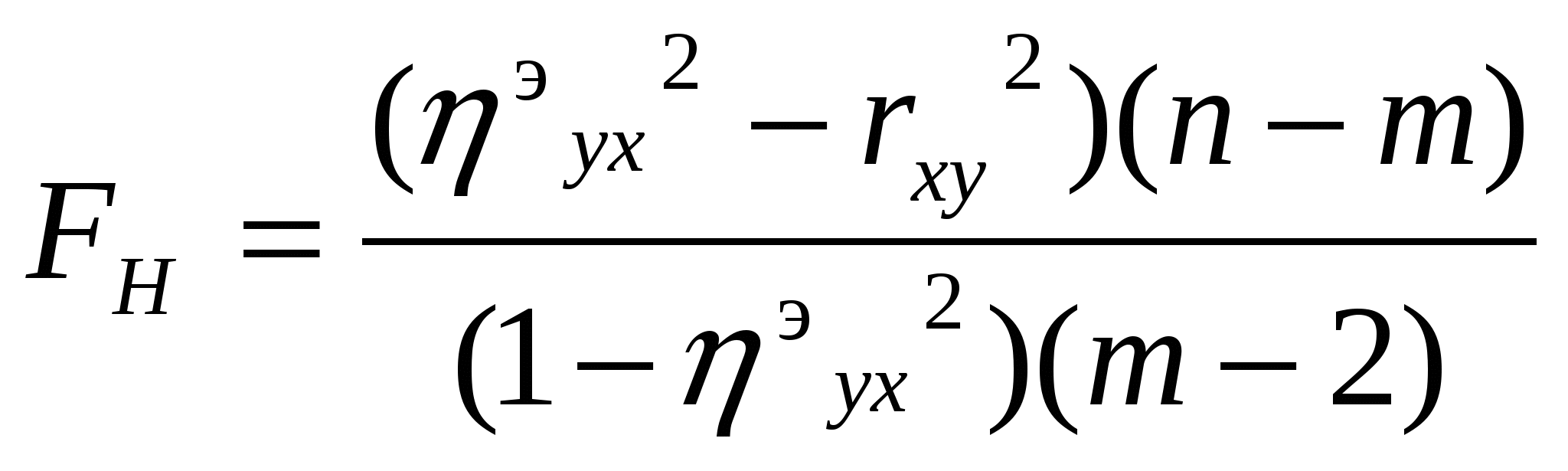
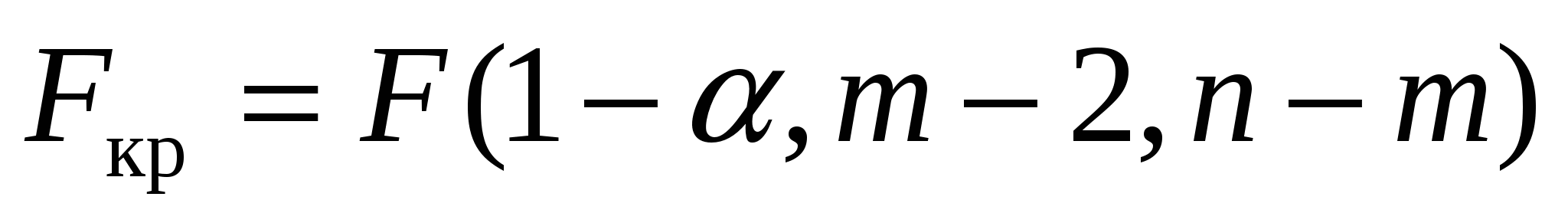
Эмпирическое корреляционное соотношение обладает следующими свойствами и характеристиками:

* Если , то исследуемые величины связаны функциональной связью, если же, то величины – независимы.
* 
* Проверка значимости эмпирического корреляционного отношения осуществляется по критерию:

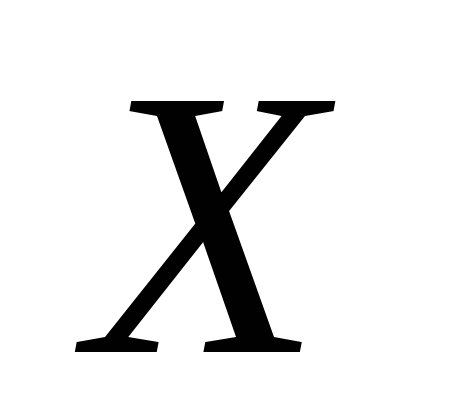
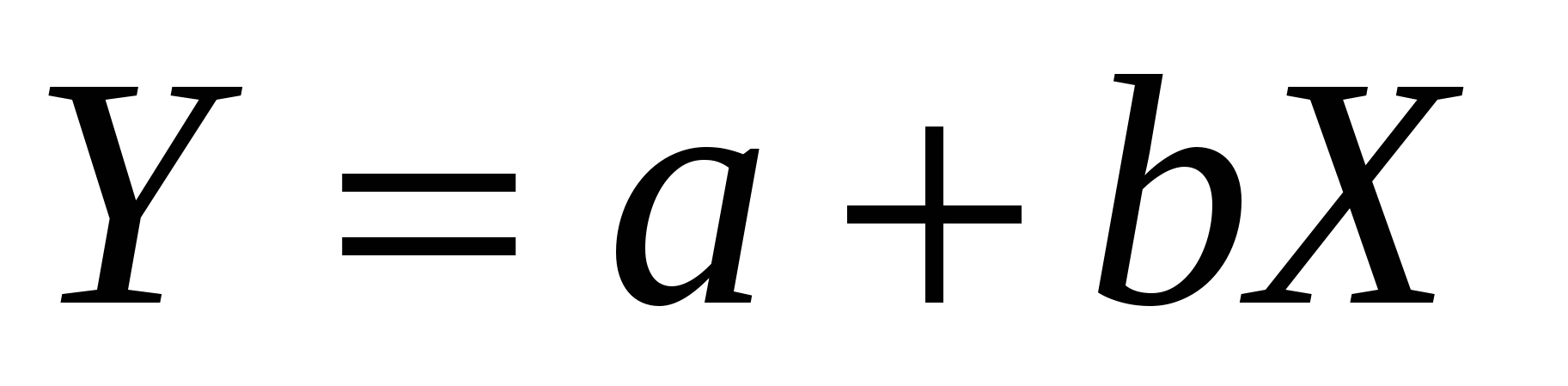
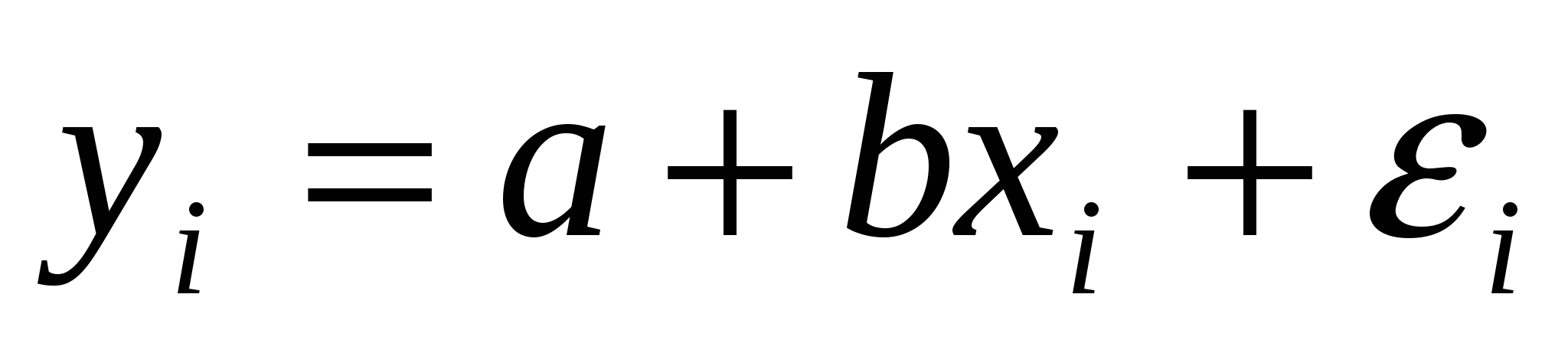
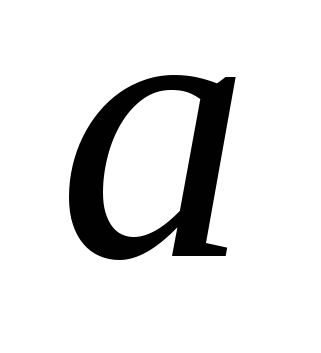
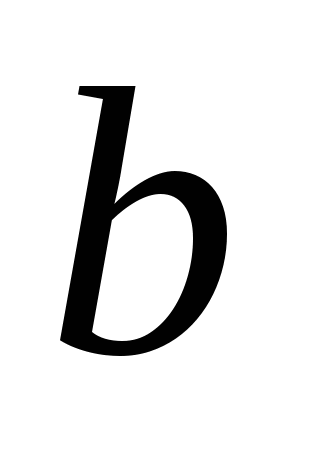
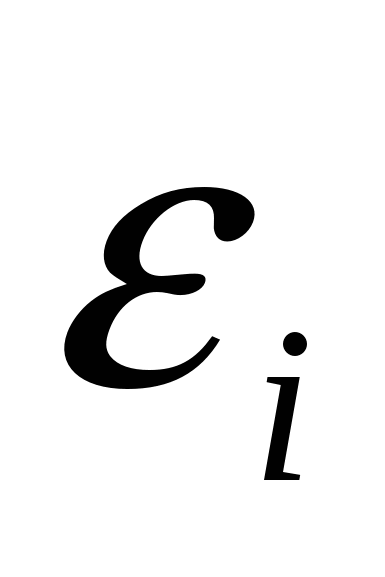
; 

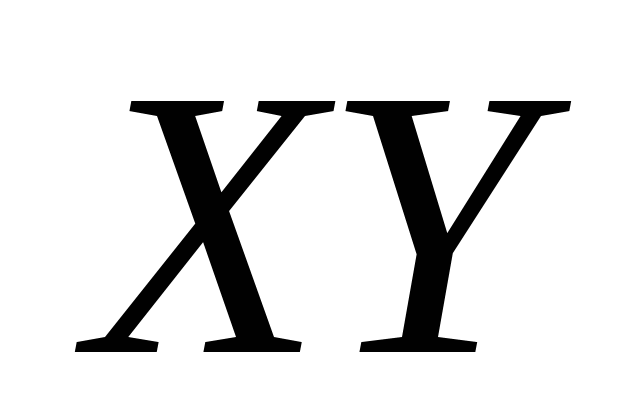
* 
* если , то связь между величинами является линейной.

Степень расхождения между величинами и может служить основанием для принятия гипотезы о линейности исследуемой связи. При этом используется критерий:

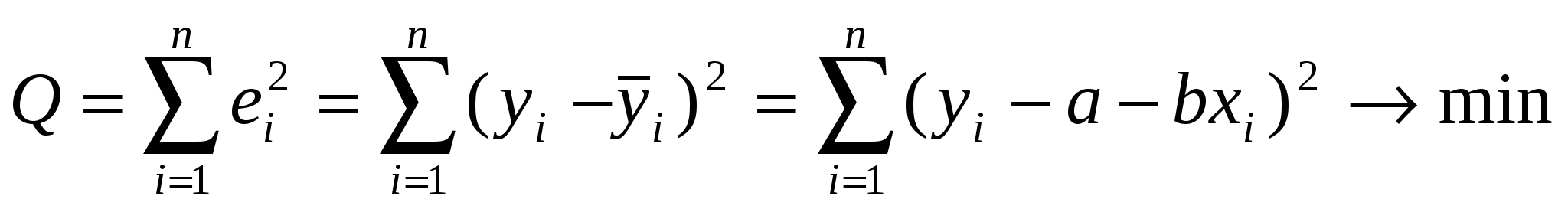
; 

*Парная линейная модель регрессии.*

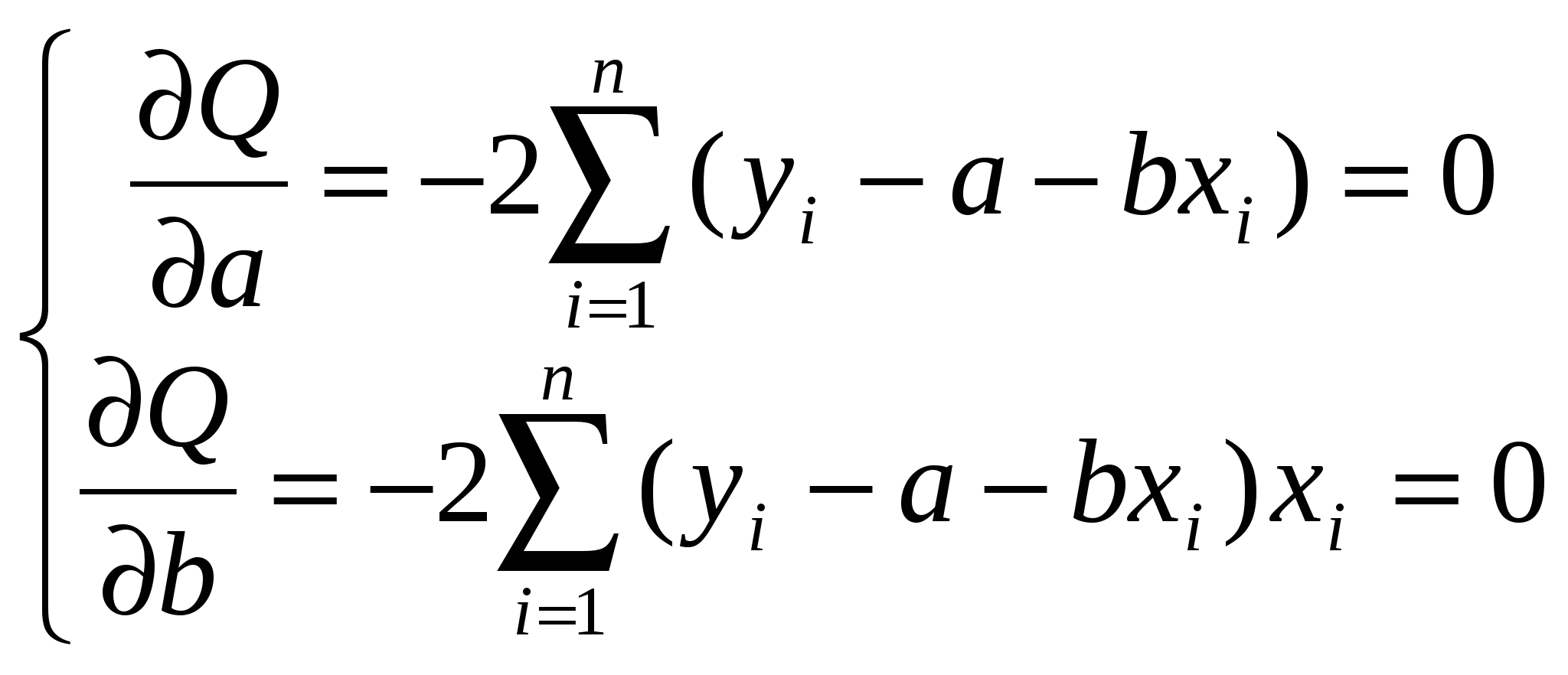
Однофакторная модель регрессии описывает зависимость между одной причиной и следствием с использованием линейной функции. Для каждого отдельного наблюдения соотношение выглядит следующим образом:, гдеи- коэффициенты регрессии;- независимая нормально распределенная величина-остаток с нулевым математическим ожиданием постоянной дисперсией.

Параметры и уравнения регрессии чаще всего оцениваются с помощью метода наименьших квадратов. Суть его состоит в том, чтобы, зная положение точек на плоскости , так провести линию регрессии, чтобы сумма квадратов отклонений этих точек по оси от проведенной прямой была минимальной.

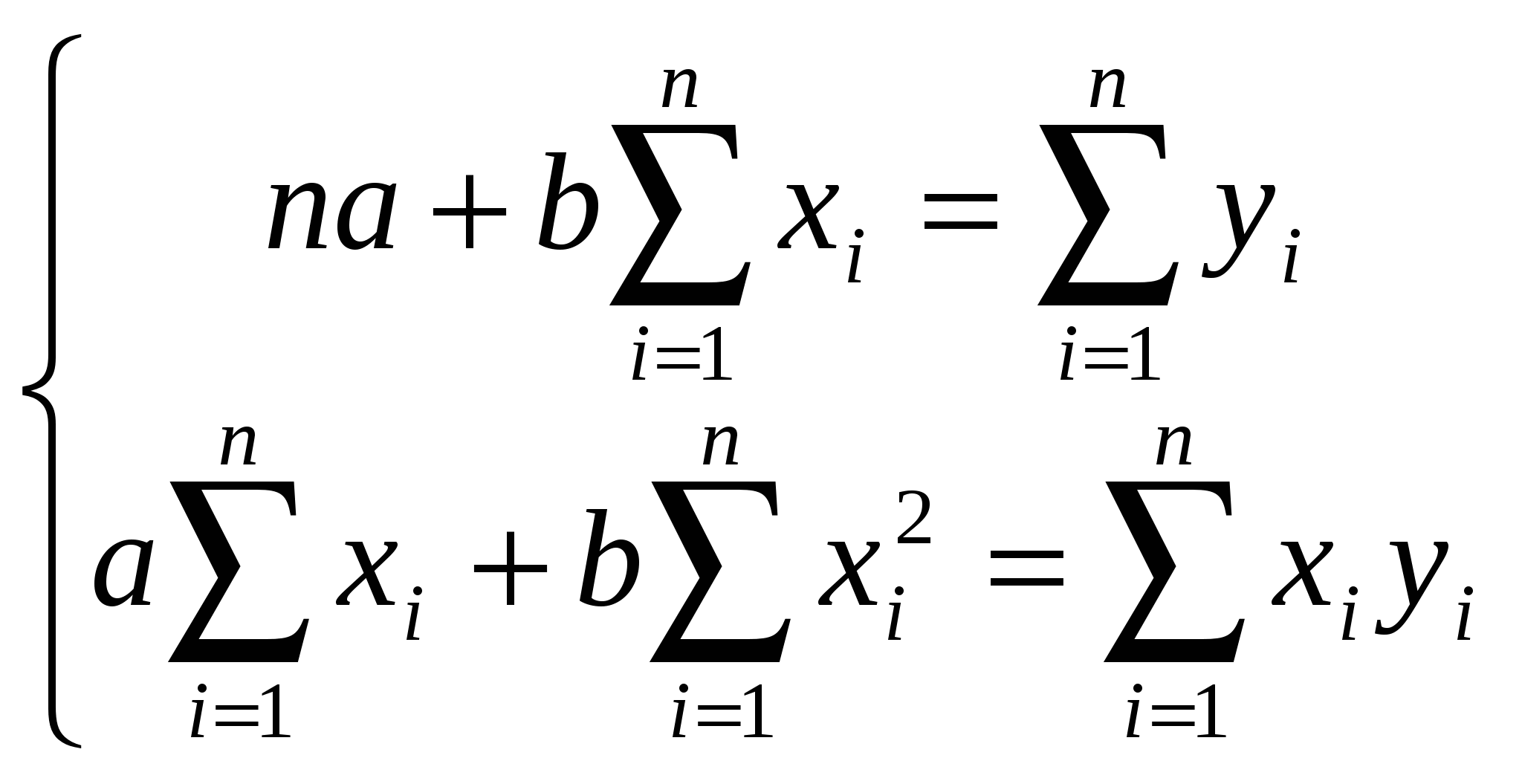
Математически критерий оценки параметров линейной парной регрессии записывается так:

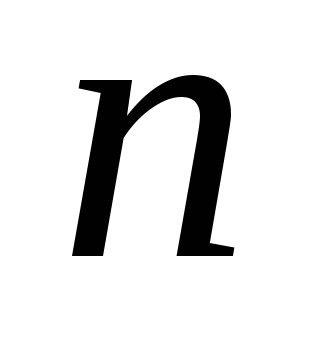
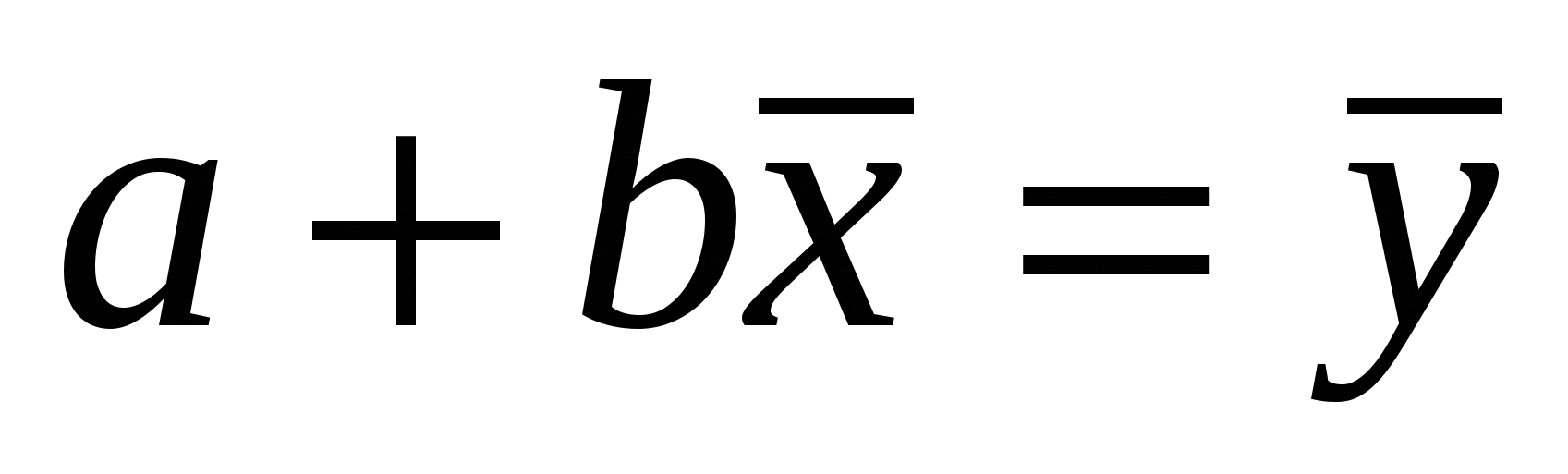


Условие существования экстремума функции – равенство нулю частных производных:

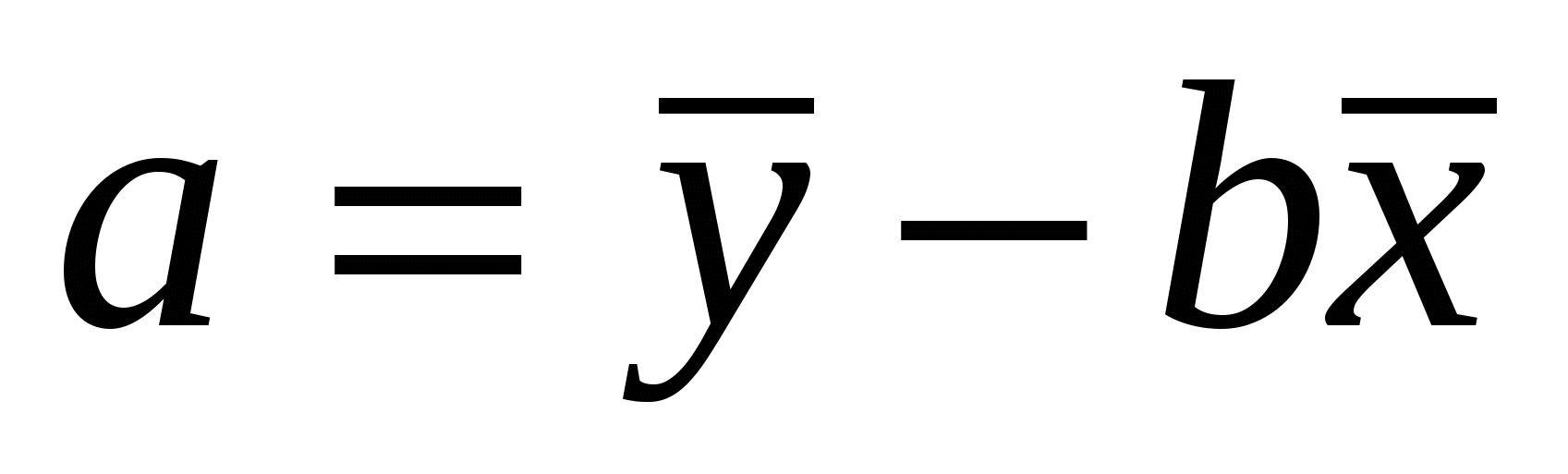
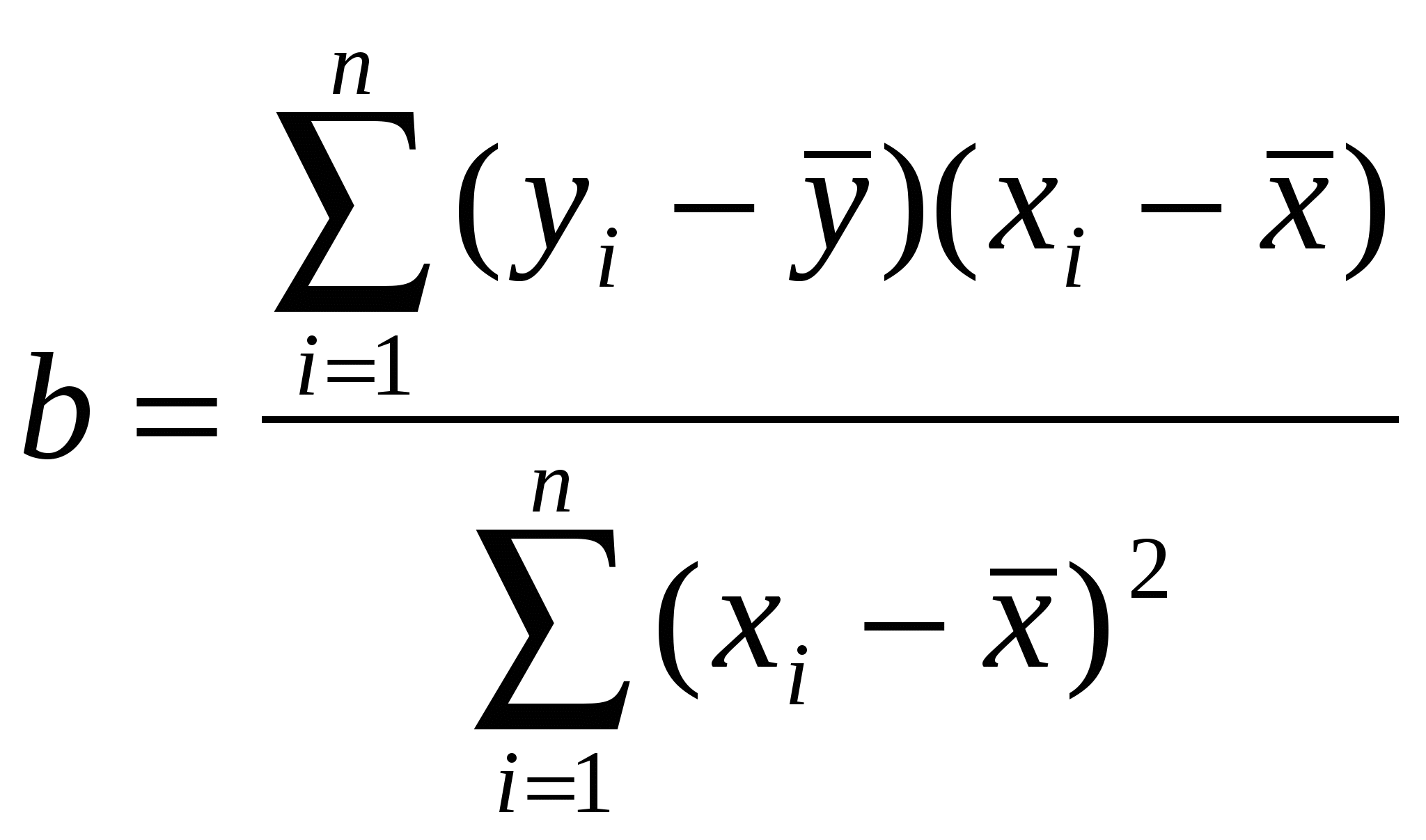


После раскрытия скобок и выполнения преобразования, получим систему из двух равнений с двумя неизвестными:

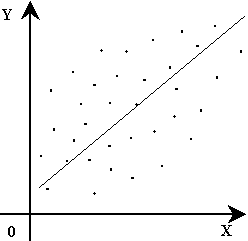


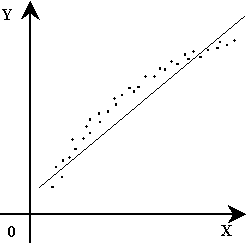
Разделив первое уравнение на , получим:

Решая систему, получим расчетные формулы для нахождения коэффициентов уравнения регрессии:



Качество полученной модели характеризуется определенными статистическими свойствами и точностью, т.е. степенью близости к фактическим данным. Модель считается хорошей со статистической точки зрения, если она адекватна и достаточно точна. Смысл используемых терминов характеризуют рисунки 6.6 и 6.7.

 *Рис. 6.5. Модель регрессии (модель адекватна, но не точна)*



*Рис. 6.6. Модель регрессии (модель точна, но не адекватна)*

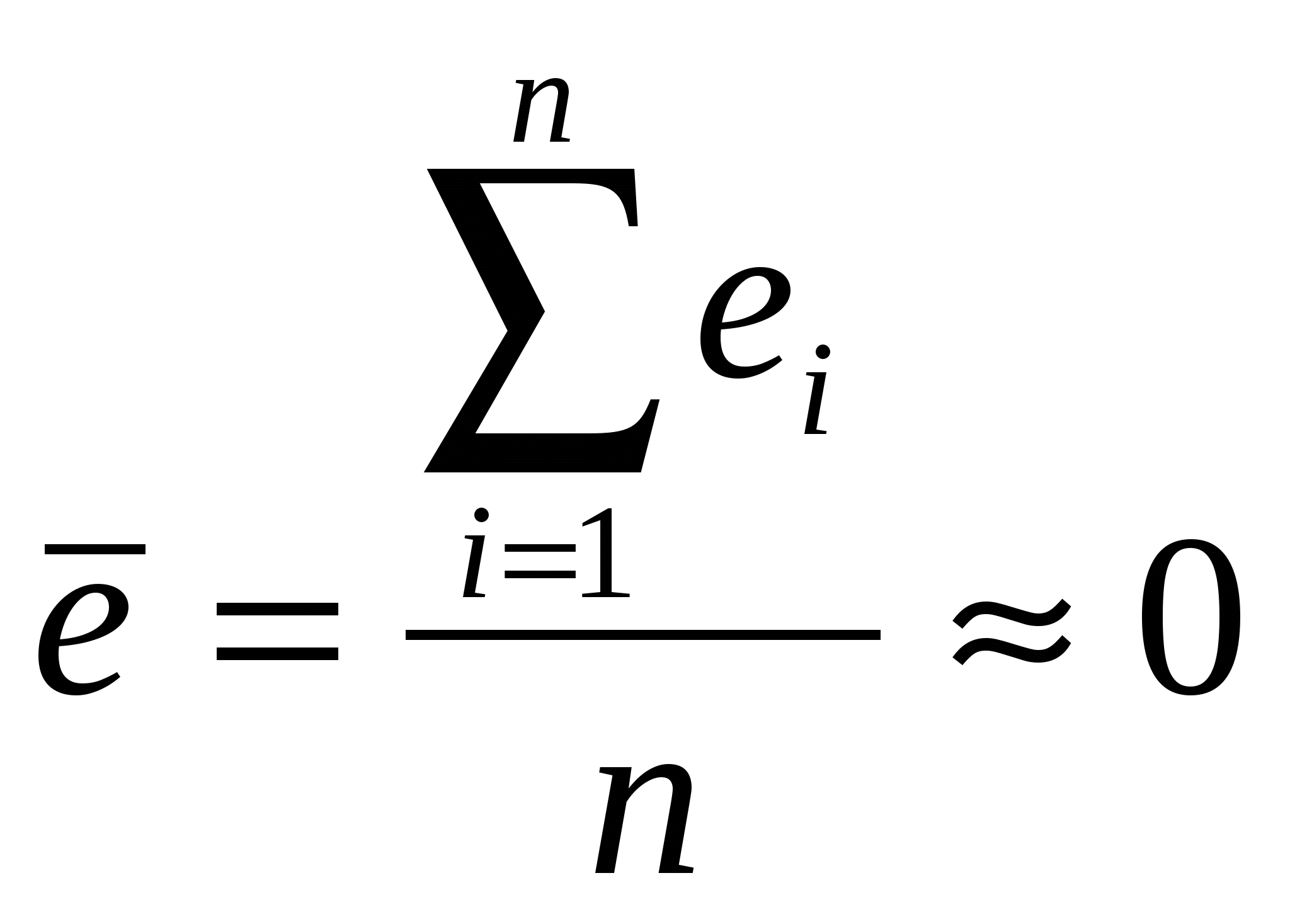
Для оценки качества линейного уравнения парной регрессии целесообразно:

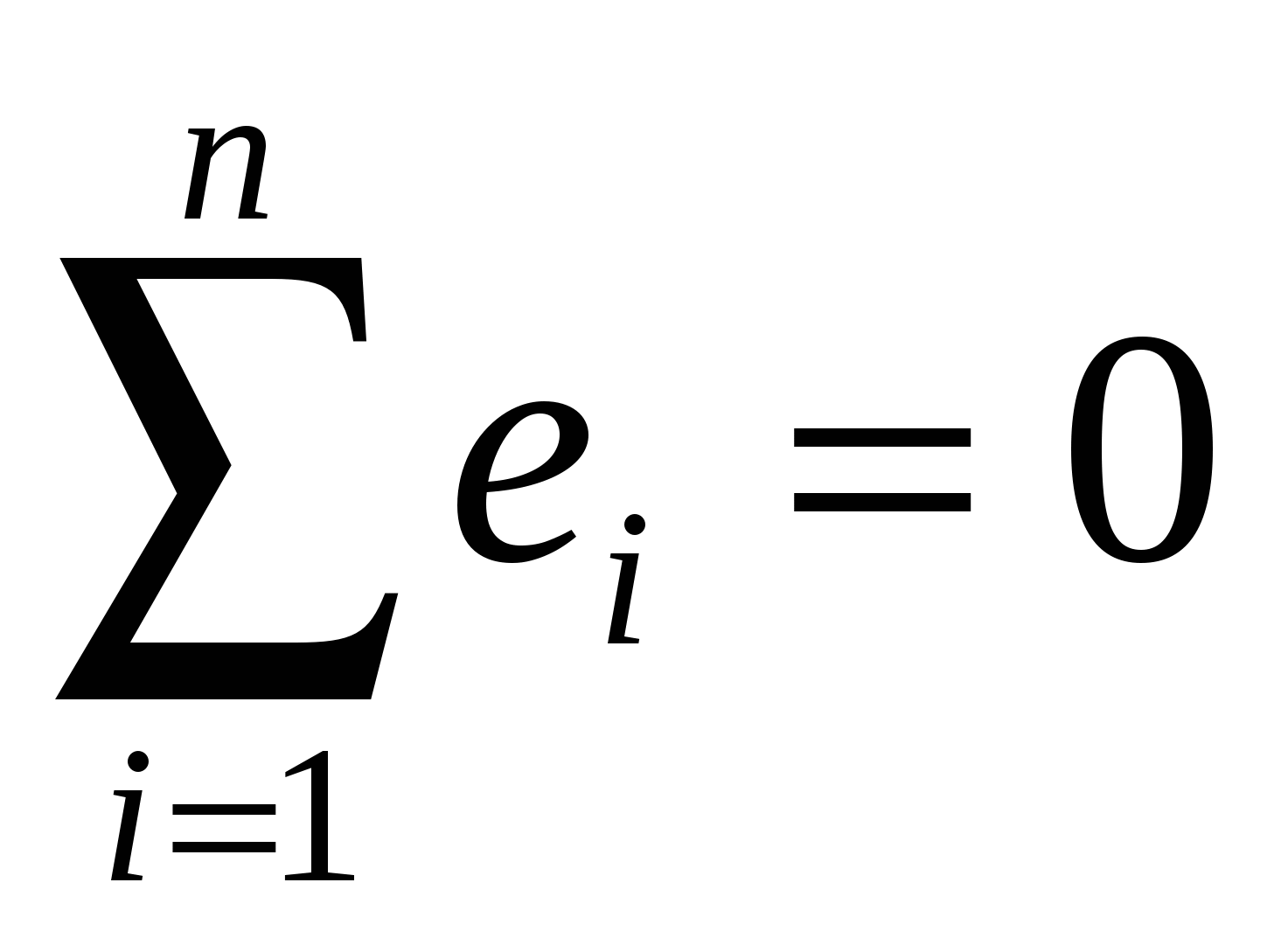
1. Проанализировать остаточный ряд. Модель считается адекватной исследуемому процессу, если:

* математическое ожидание значений остаточного ряда близко или равно нулю;
* значения остаточного ряда случайны;
* значения остаточного ряда независимы;
* значения остаточного ряда подчинены нормальному закону распределения.

Таким образом, анализ адекватности модели разбивается на несколько этапов:

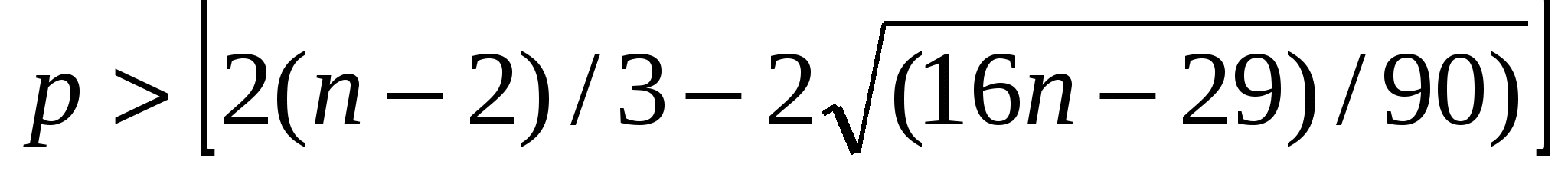
1. Равенство нулю математического ожидания ряда остатков означает выполнение следующего соотношения:

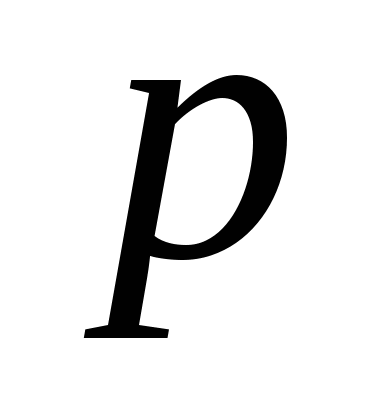


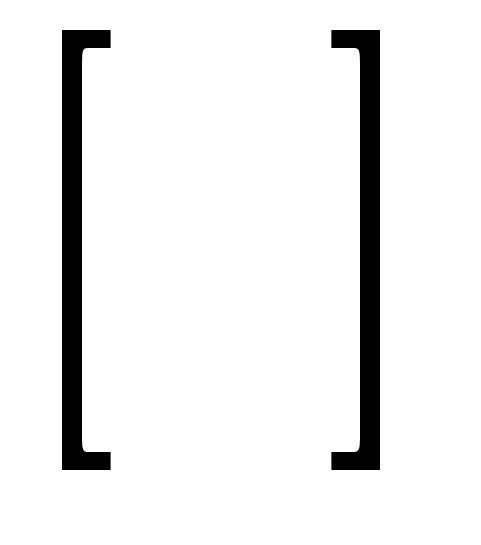
Однако в случае применения метода наименьших квадратов такая проверка является излишней, поскольку использование данного метода предполагает выполнение равенства , откуда безусловным образом следует равенство нулю математического ожидания значений остаточного ряда.

2. Проверка случайности последовательности проводится с помощью критерия пиков (поворотных точек). Каждое значение ряда сравнивается с двумя, рядом стоящими. Точка считается поворотной, если она либо больше и предыдущего и последующего значения, либо меньше и предыдущего и последующего значения.

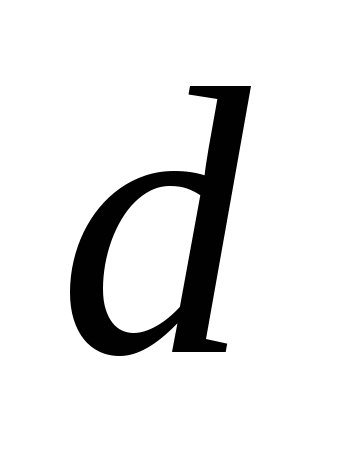
В случайном ряду должно выполняться строгое неравенство:

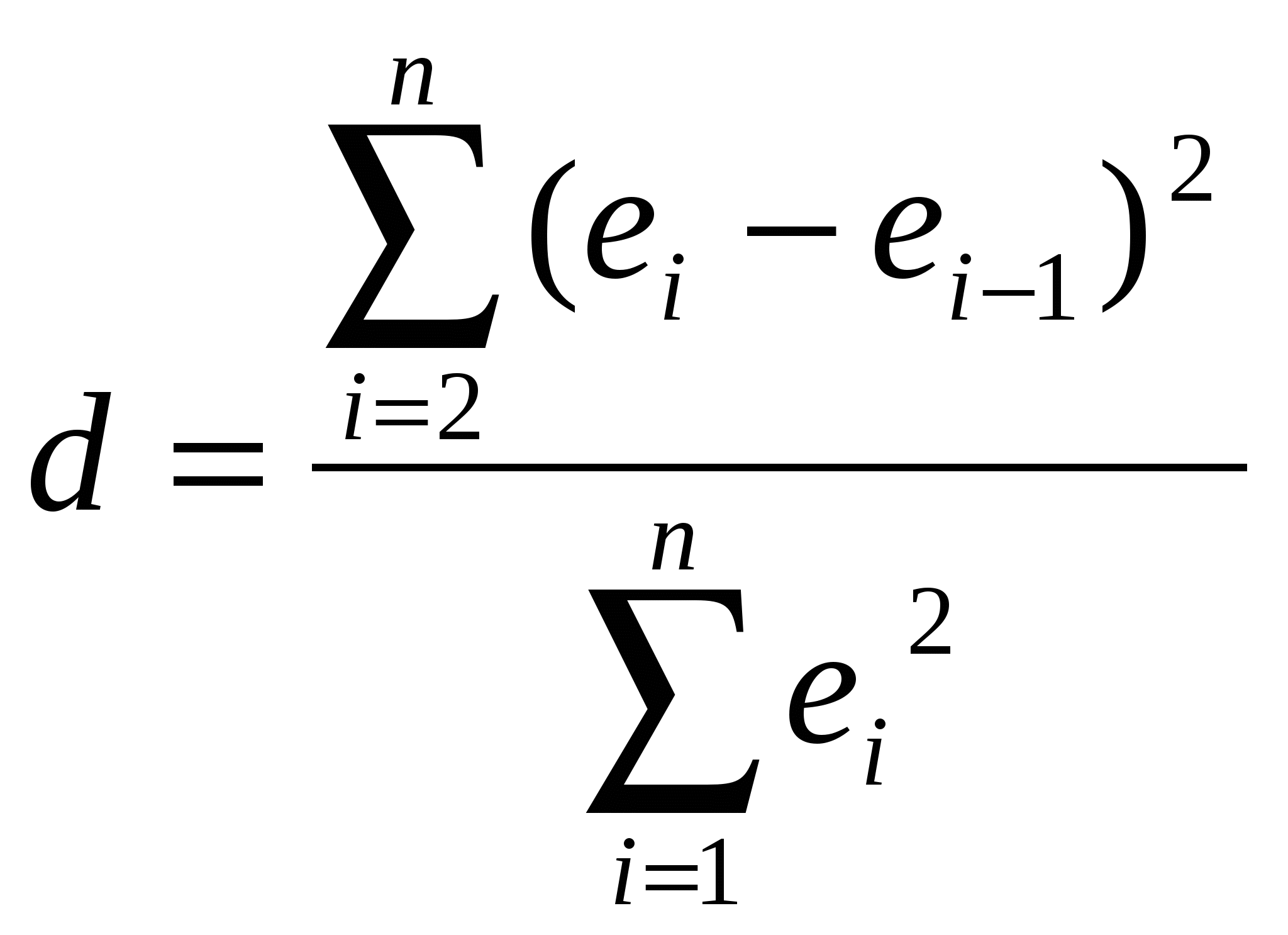
, где

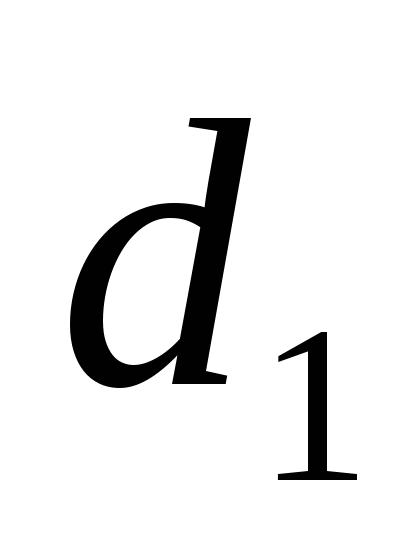
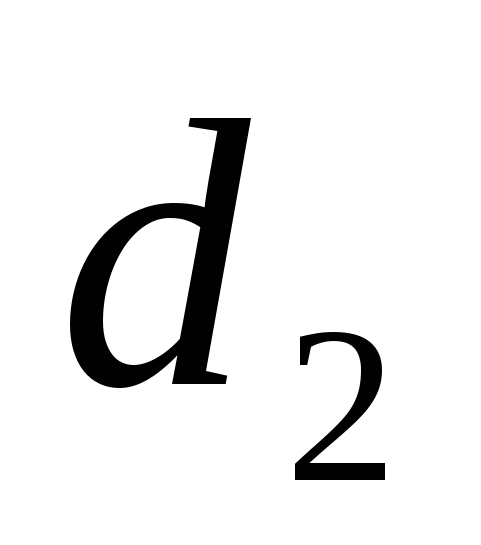
- число поворотных точек;

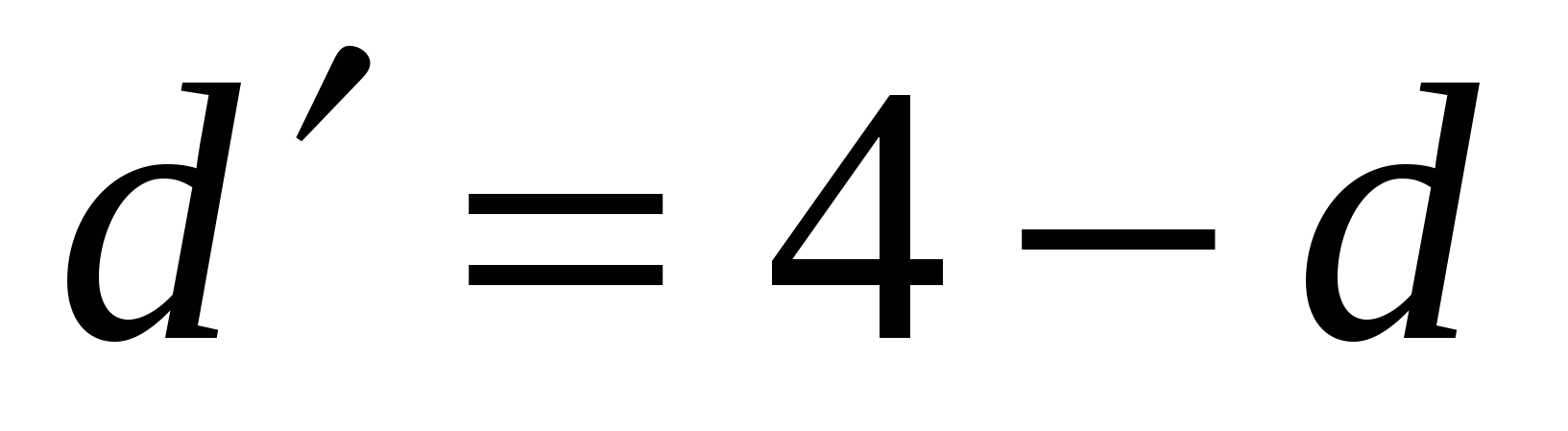
- целая часть результата вычислений.

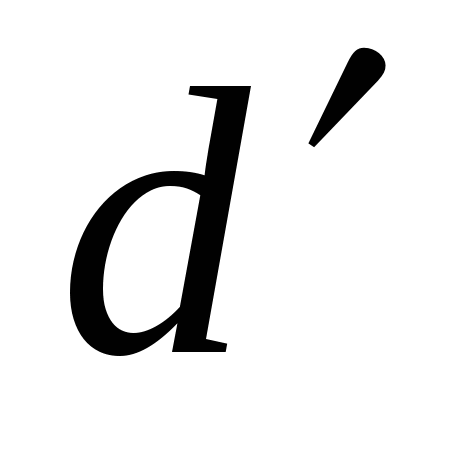
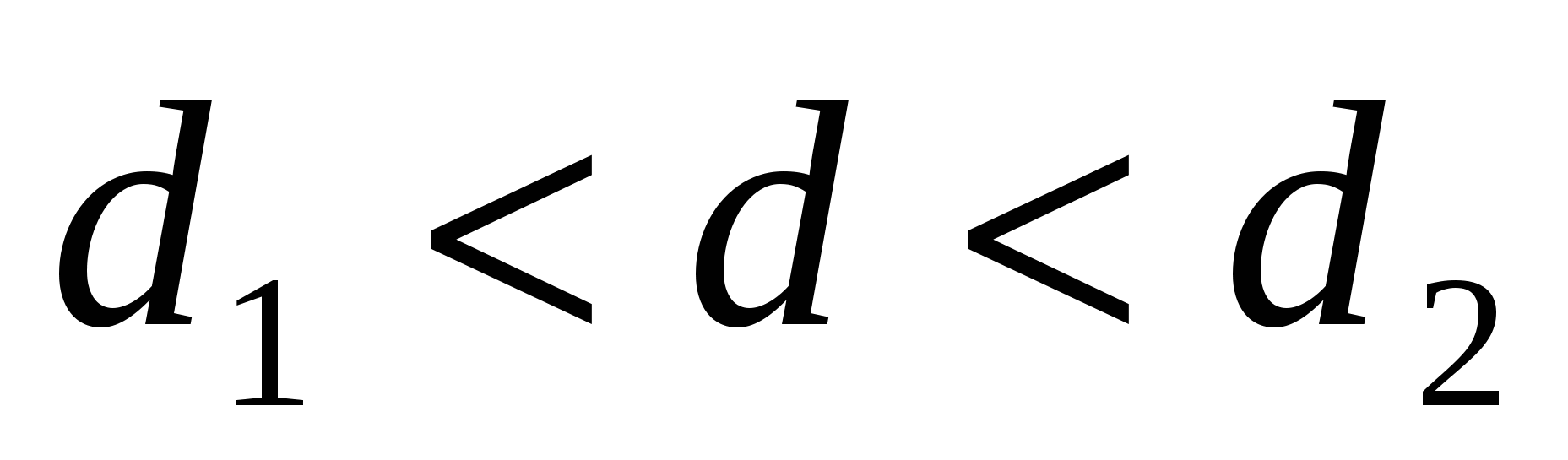
3. При проверке независимости значений определяется отсутствие в остаточном ряду автокорреляции, под которой понимается корреляция между элементами одного и того же числового ряда. Значительная автокорреляция говорит о том, что спецификация регрессии выполнена неправильно (неправильно определен тип зависимости).

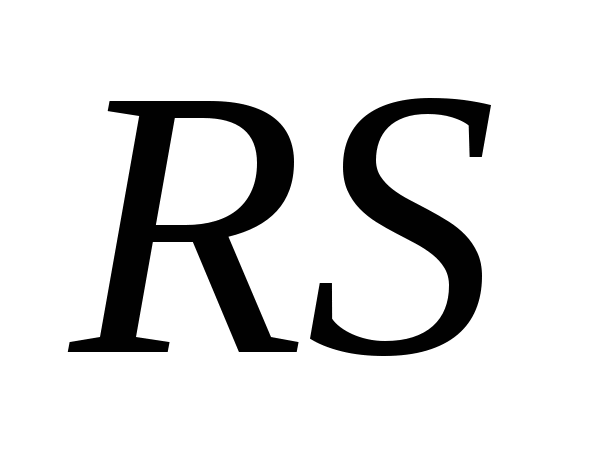
Наличие автокорреляции также может быть выявлено при помощи -критерия Дарбина-Уотсона. Значение критерия вычисляется по формуле:

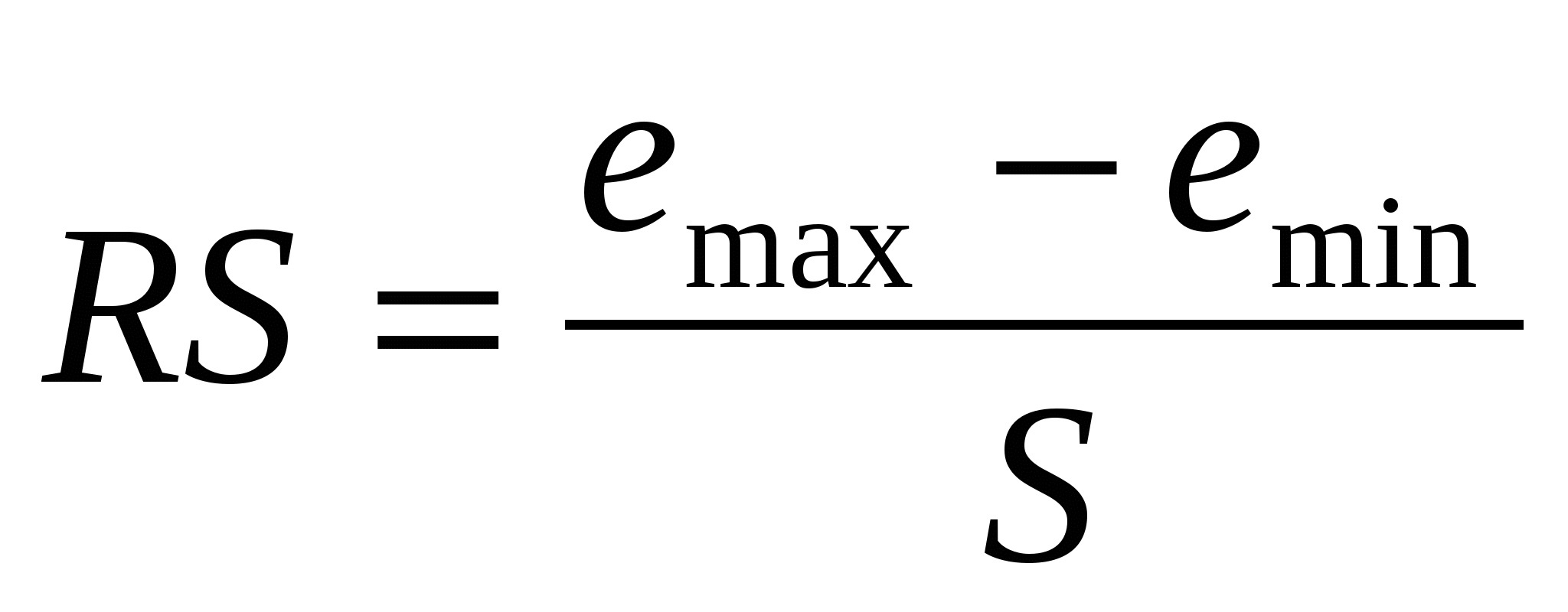


Эта величина сравнивается с двумя табличными уровнями (соответствующая таблица приведена в приложении к лекции): нижним значением - и верхним значением - . Если полученное значение больше двух, то перед сопоставлением его нужно преобразовать:



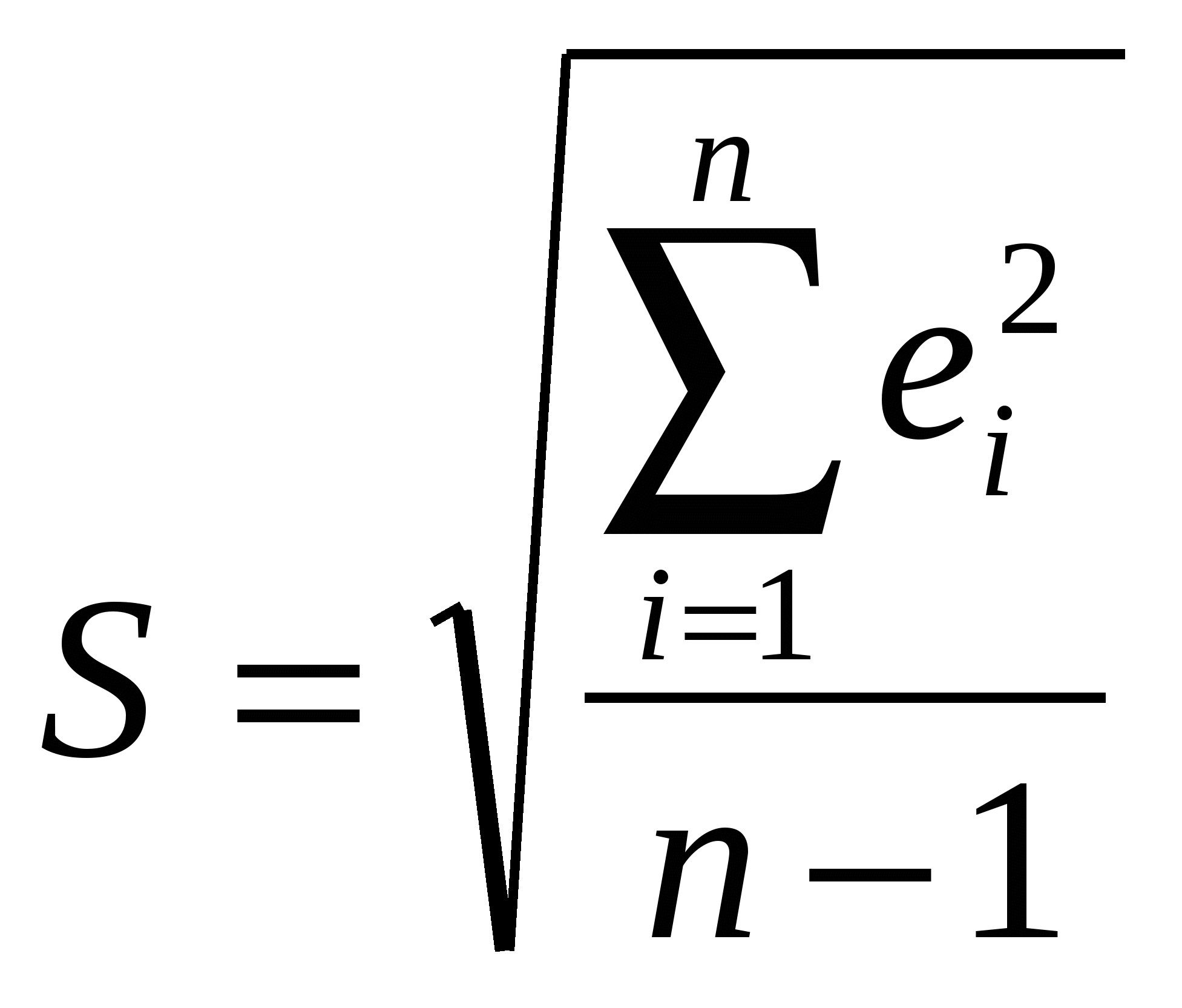
Если (или ) находится в интервале от нуля до , то значения остаточного ряда сильно автокоррелированы. Если значение -критерия попадает в интервал от до 2, то автокорреляция отсутствует. Если же - однозначного вывода об отсутствии или наличии автокорреляции сделать нельзя и необходимо использовать другой критерий, например, коэффициент автокорреляции первого порядка.

4. Соответствие остаточного ряда нормальному распределению проще всего проверить при помощи -критерия:

, где

- максимальное значение ряда остатков;

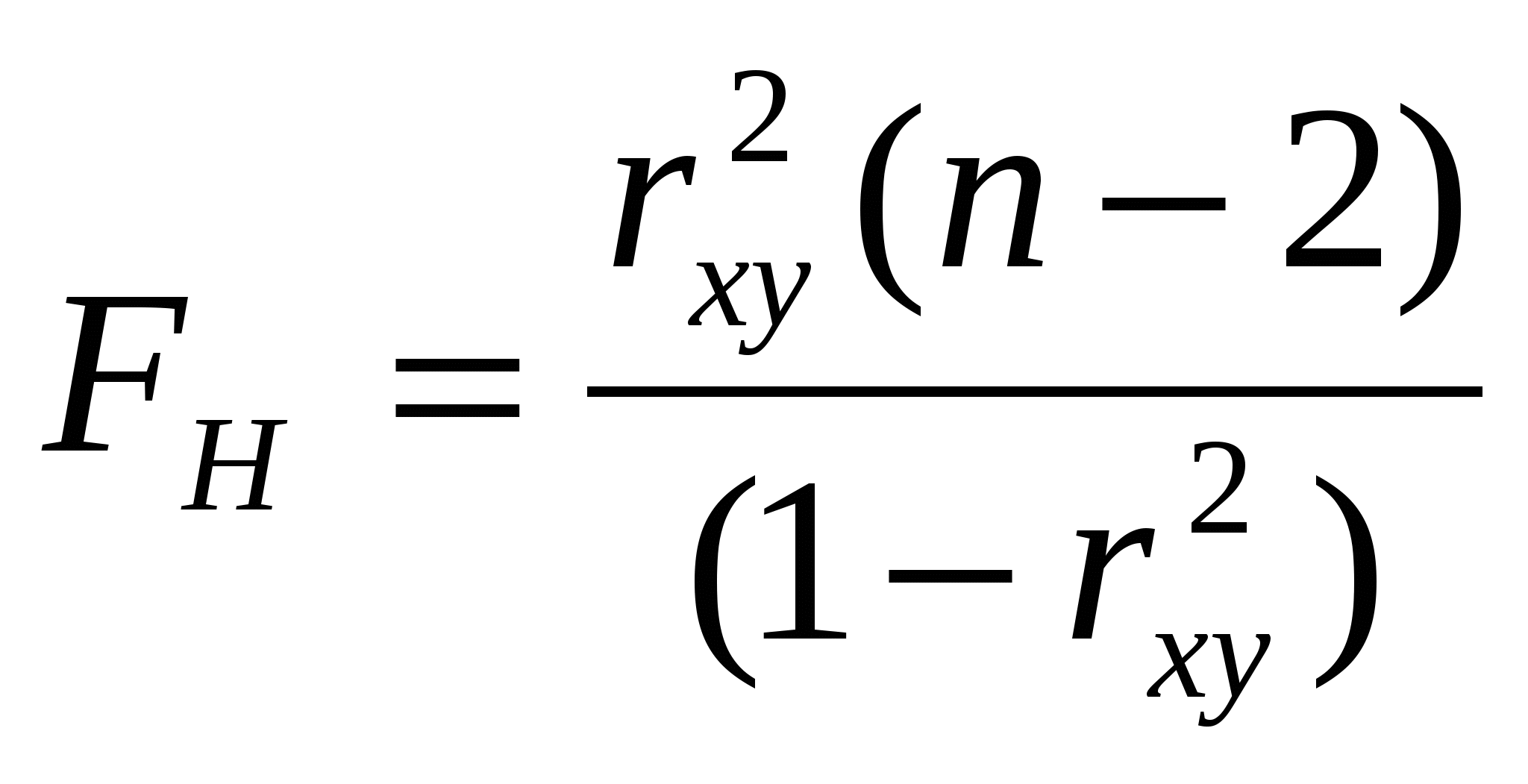
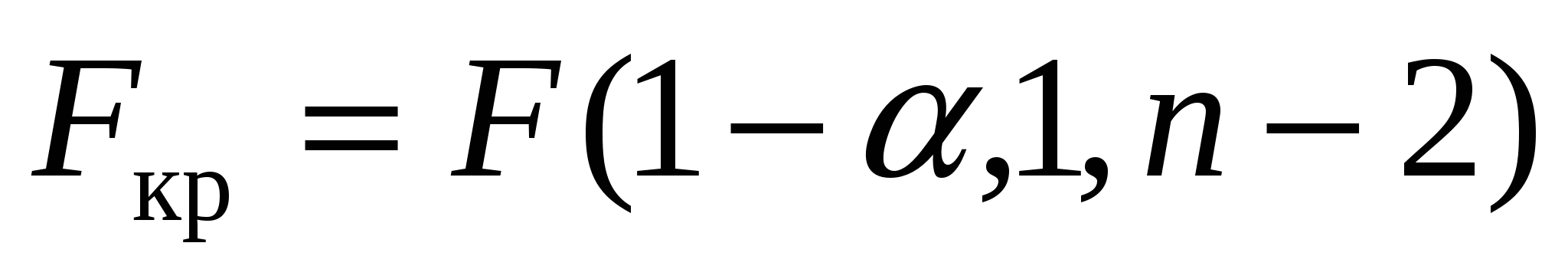
- минимальное значение ряда остатков;

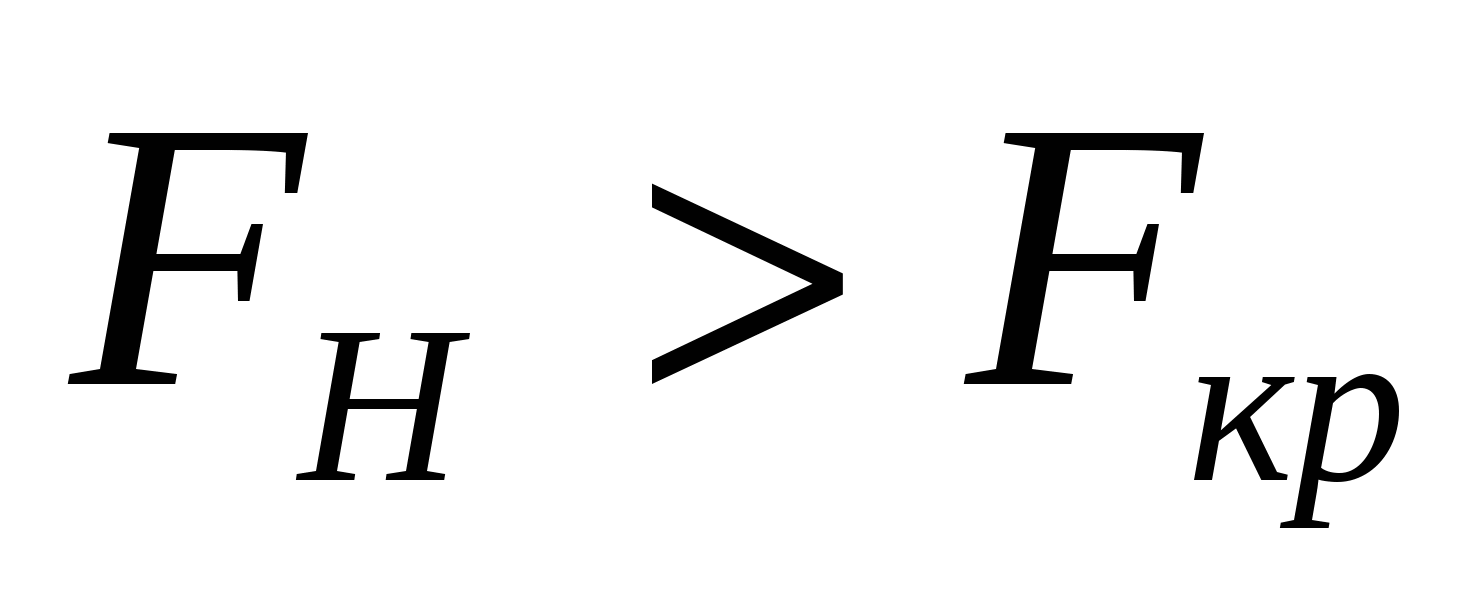
- среднеквадратическое отклонение значений остаточного ряда.

Если рассчитанное значение попадает между табулированными границами с заданным уровнем вероятности, то гипотеза о нормальном распределении принимается. Соответствующая статистическая таблица приведена в приложении к лекции.

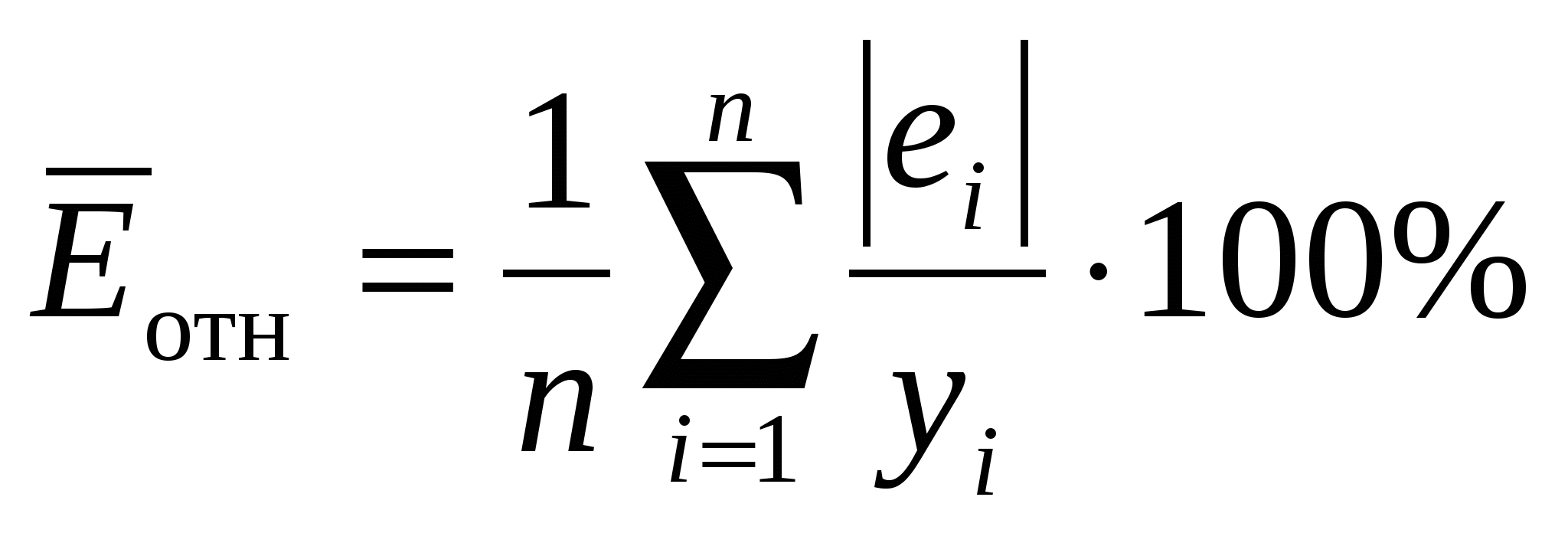
1. вычислить и оценить значимость коэффициента корреляции (если этого не было сделано на этапе выбора вида модели);
2. проверить адекватность (значимость) всей модели регрессии;

Для проверки значимости модели регрессии использую критерий Фишера:

; 

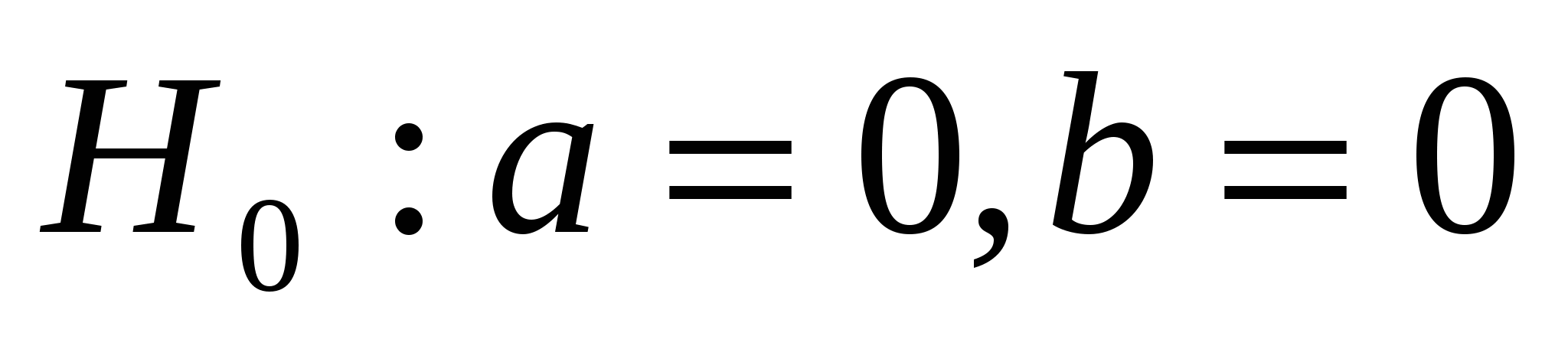
n – число наблюдений. Условие значимости модели: .

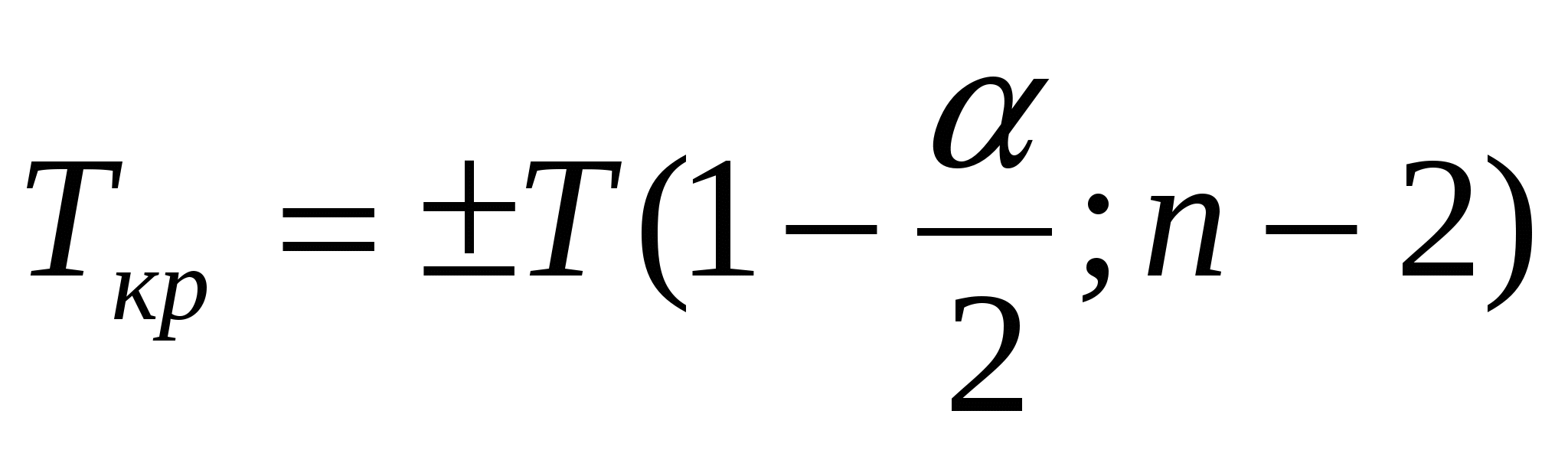
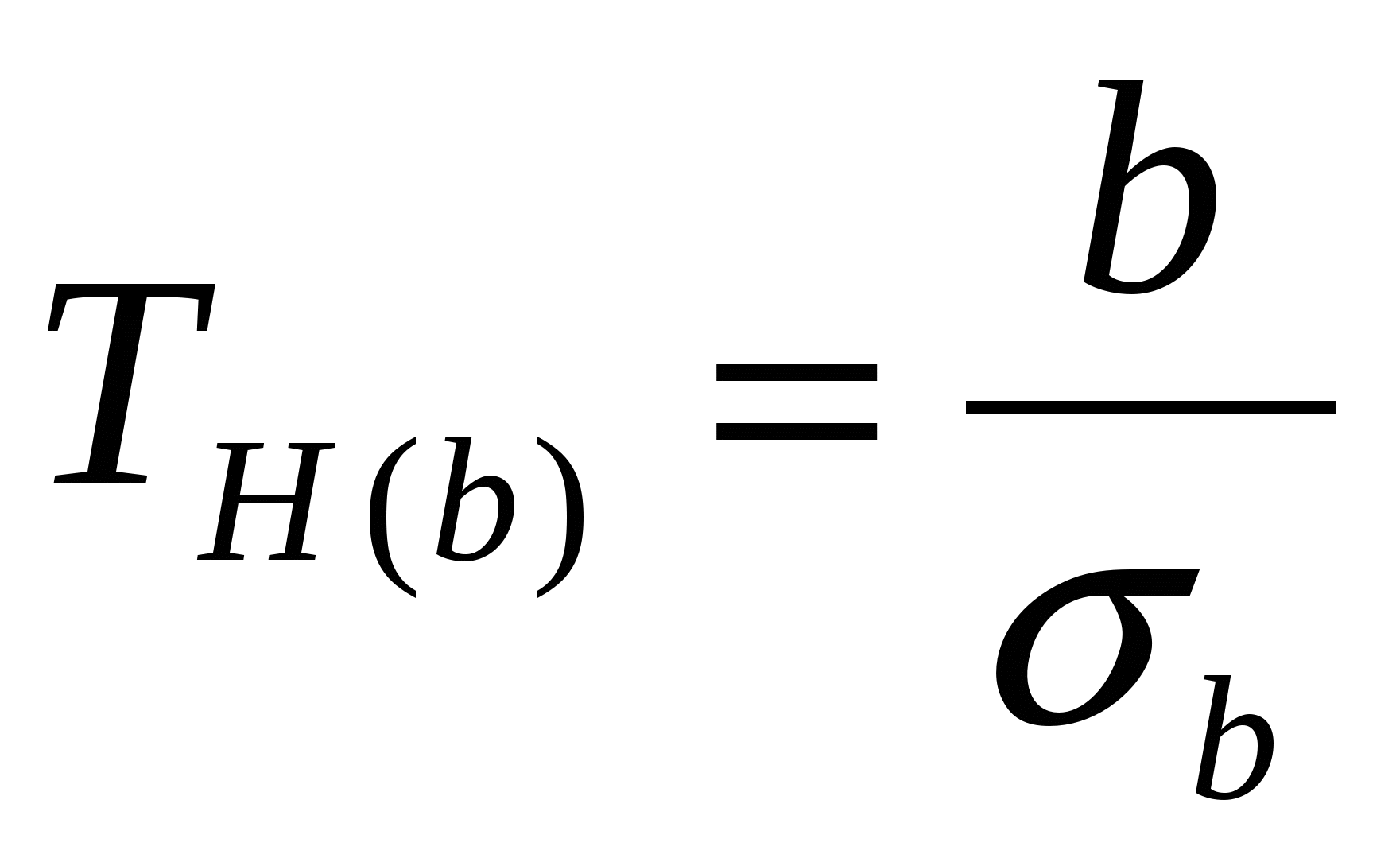
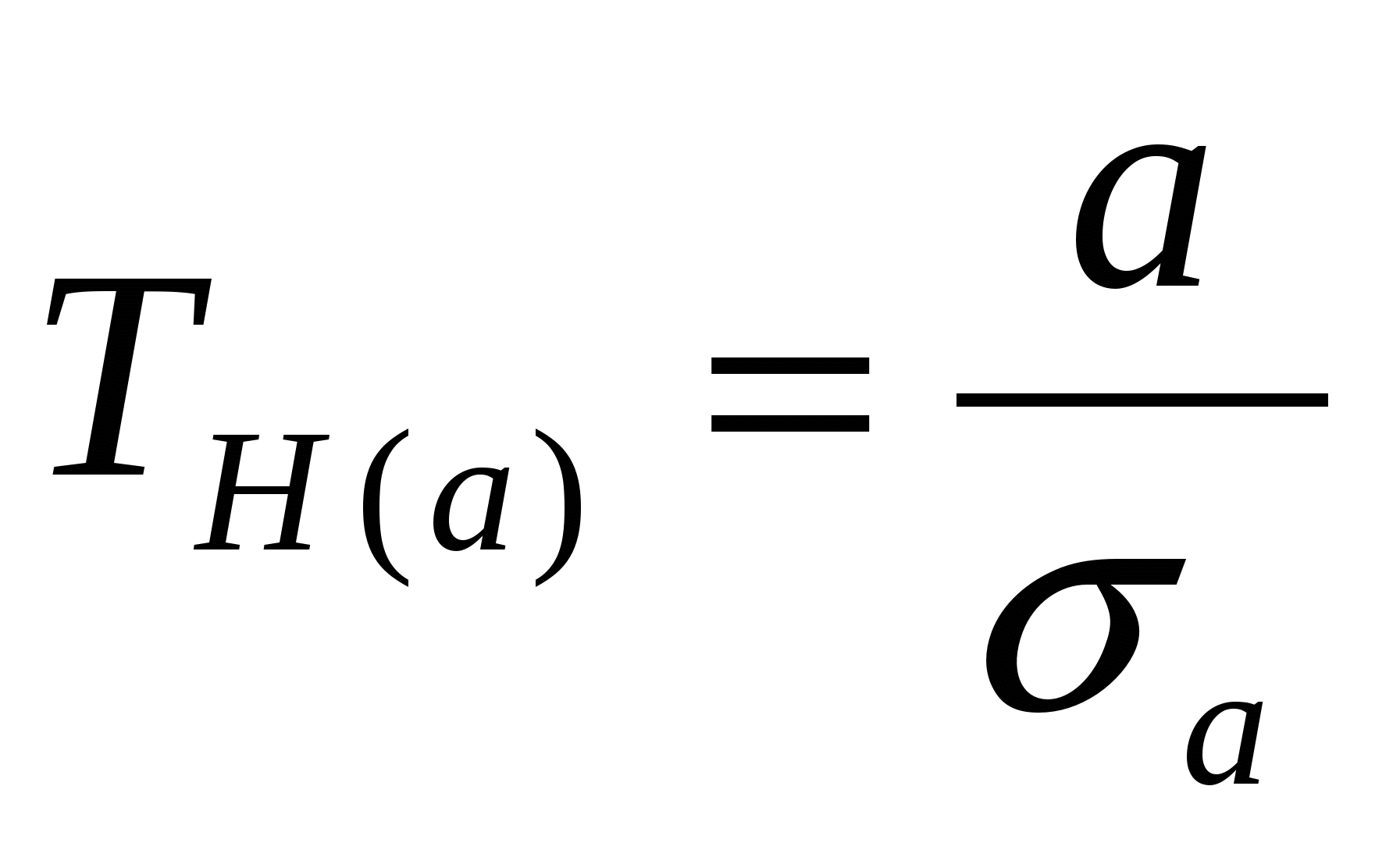
1. Оценить среднюю относительную ошибку:

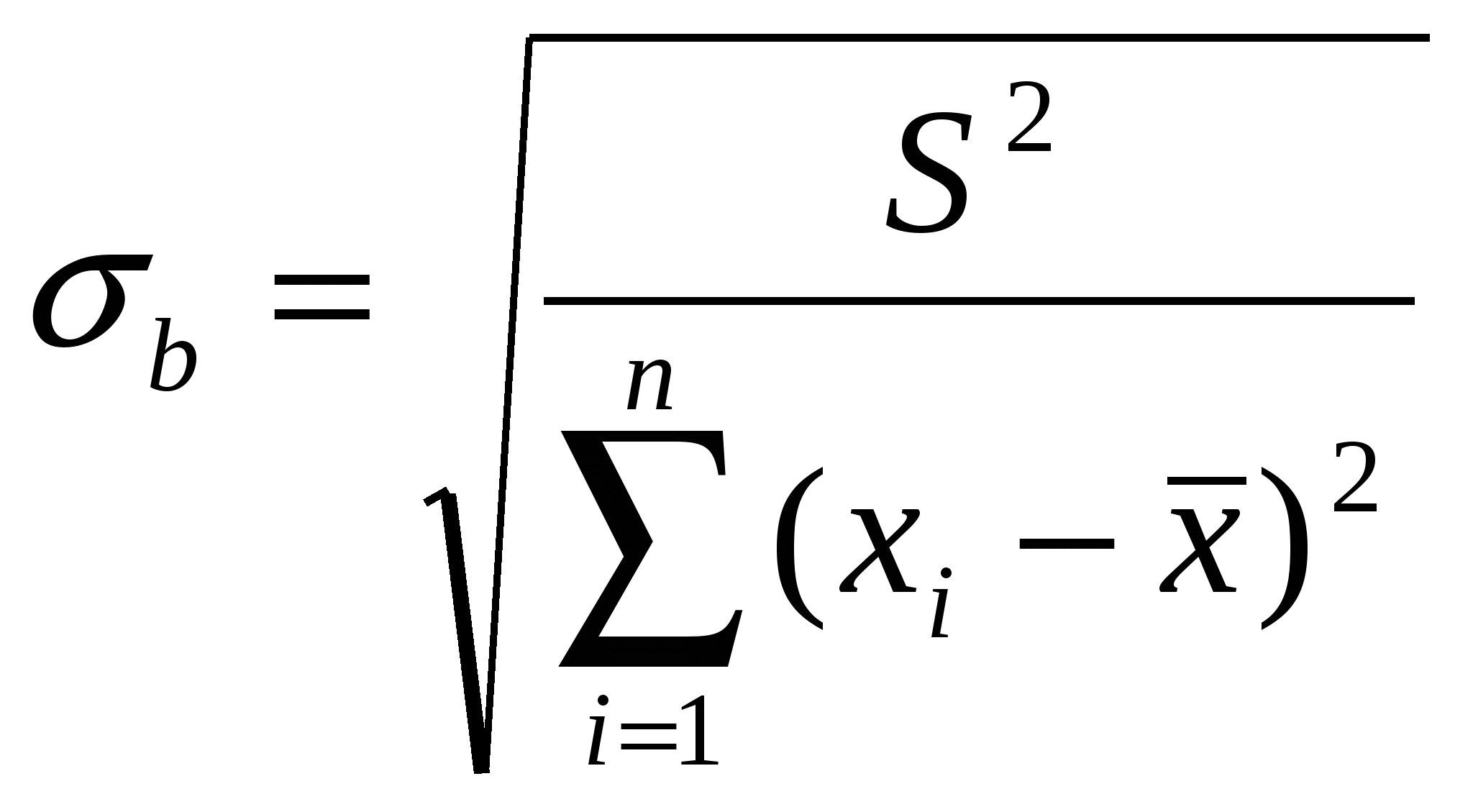
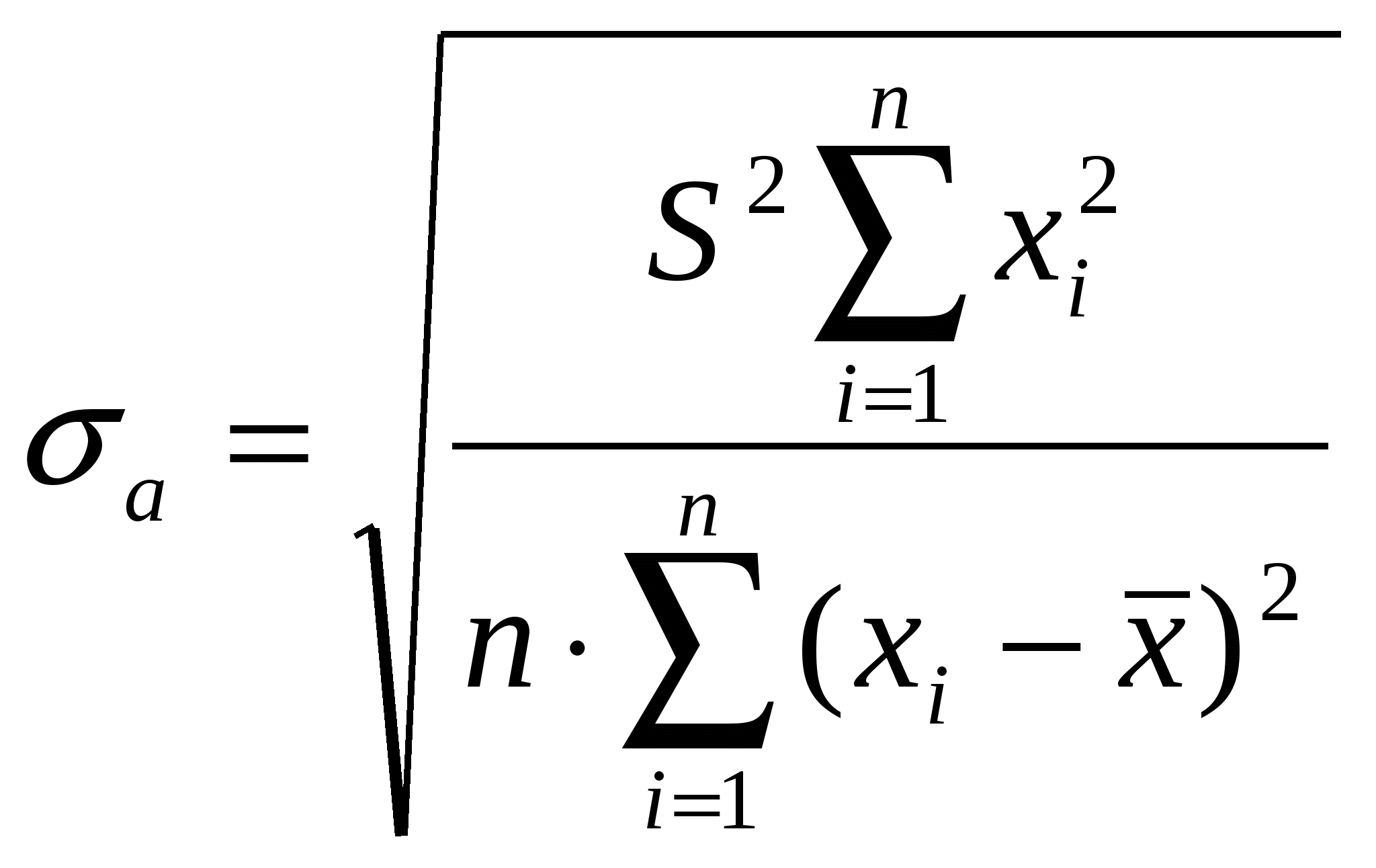


В отношении величины средней относительной ошибки, как правило, делают следующие выводы. Величина менее 5% свидетельствует о хорошем уровне точности, ошибка до 15% считается приемлемой.

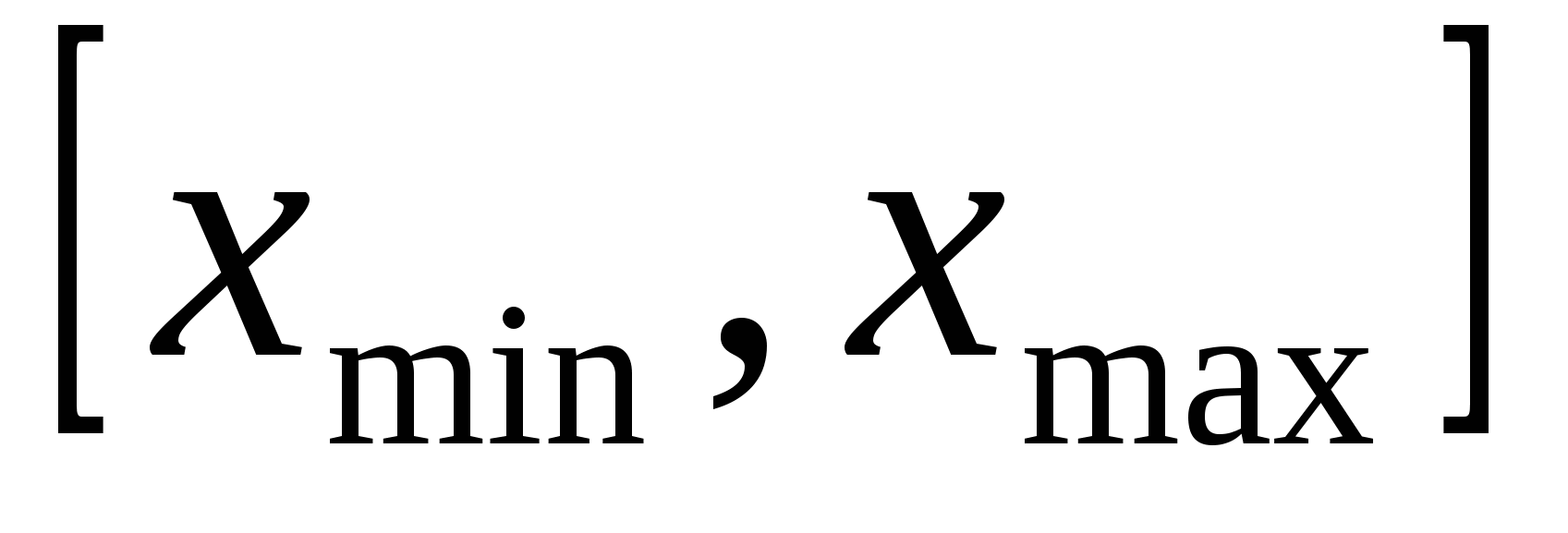
1. Проверить значимость параметров и модели регрессии;

Проверка значимости отдельных коэффициентов регрессии связана с определением наблюдаемых значений критерия Стьюдента для соответствующих коэффициентов регрессии. Нулевая гипотеза в данном случае имеет вид: .



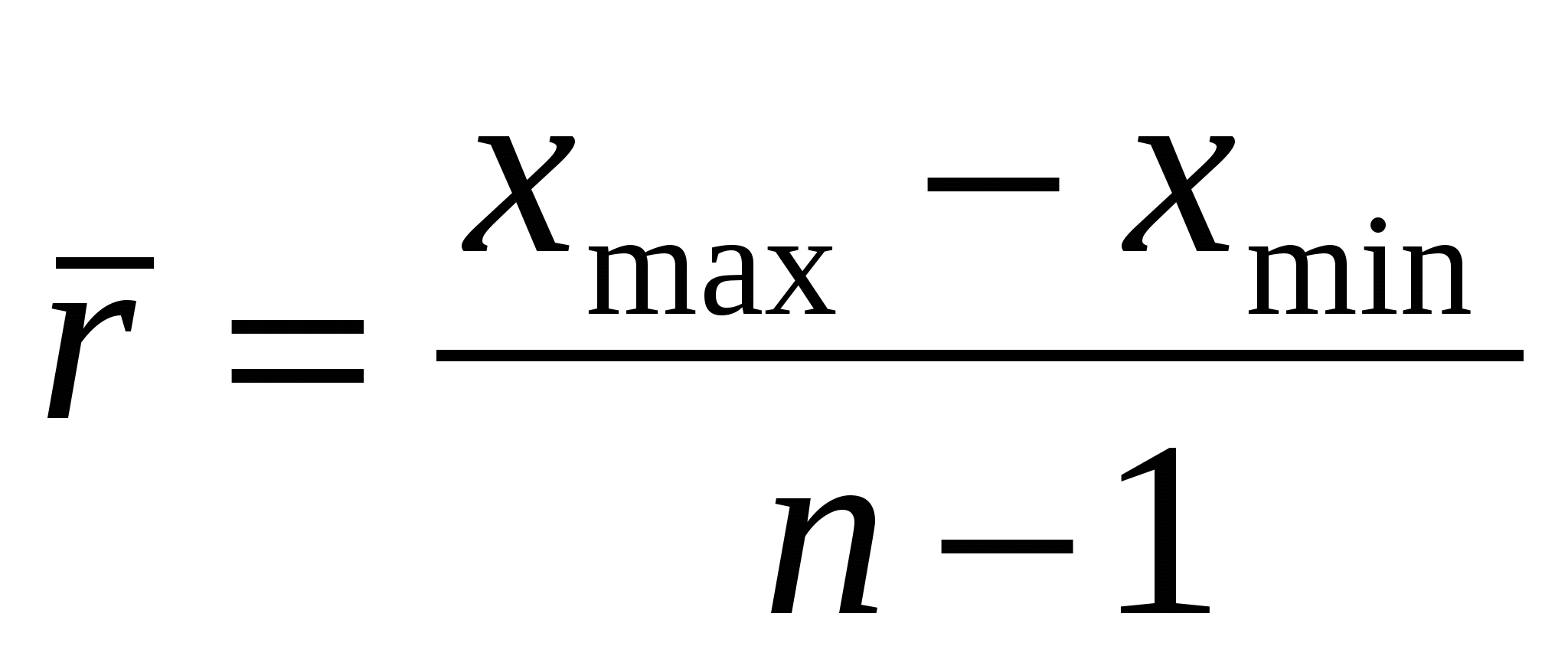


Если наблюдаемое значение критерия превосходит по модулю его табличное значение при заданном уровне значимости, то коэффициент регрессии считается значимым. В противном случае фактор, соответствующий этому коэффициенту, следует исключить из модели, при этом ее качество не ухудшиться.

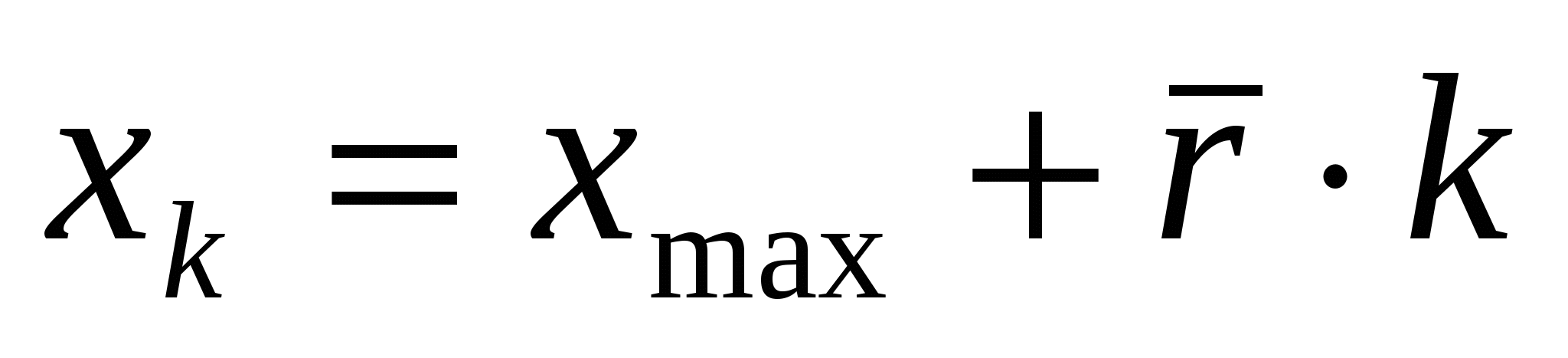
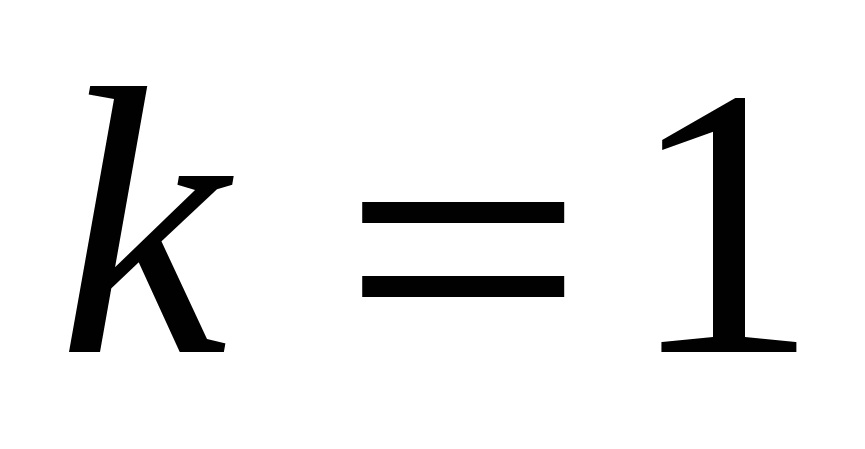
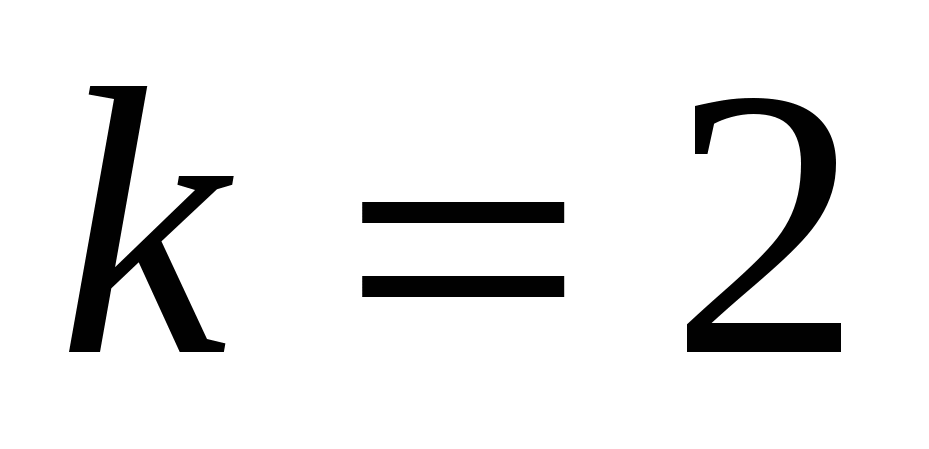
После проверка качества модели ее можно использовать для анализа процесса результатом, которого является значение исследуемого показателя качества, или для прогнозирования. Модель регрессии позволяет проводить как экстраполяцию, так и интерполяцию значений. Интерполяция - прогнозирование значений зависимой переменной для значений фактора, принадлежащих интервалу . Экстраполяция - прогнозирование значений зависимой переменной для значений фактора, выходящих за границы интервала .

Точечный прогноз получается путем простой подстановки соответствующих значений в уравнение регрессии.

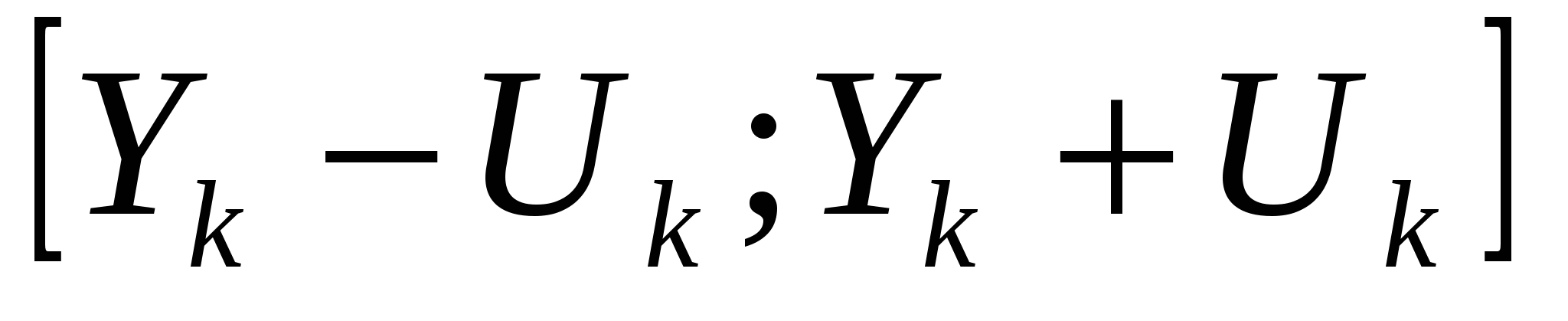
Зачастую значения факторов, для которых нужно сделать прогноз значения зависимой переменной, получают на основе среднего прироста значений фактора внутри выборочной совокупности:

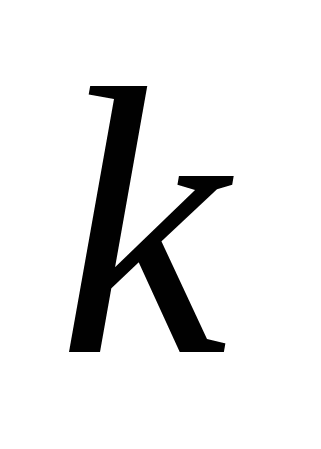


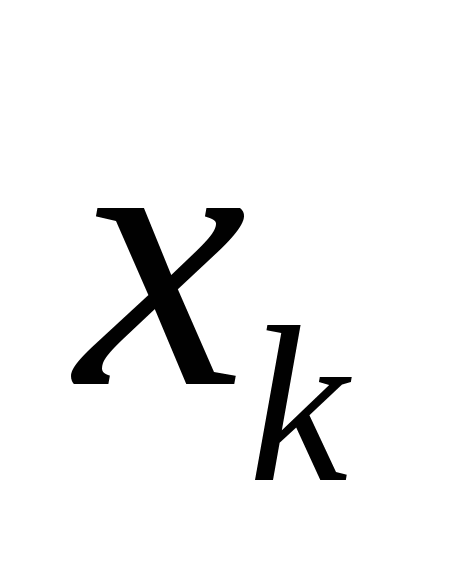
При выполнении экстраполяции для определения конкретного значения , используемого для расчета прогнозного значения , можно использовать формулу:

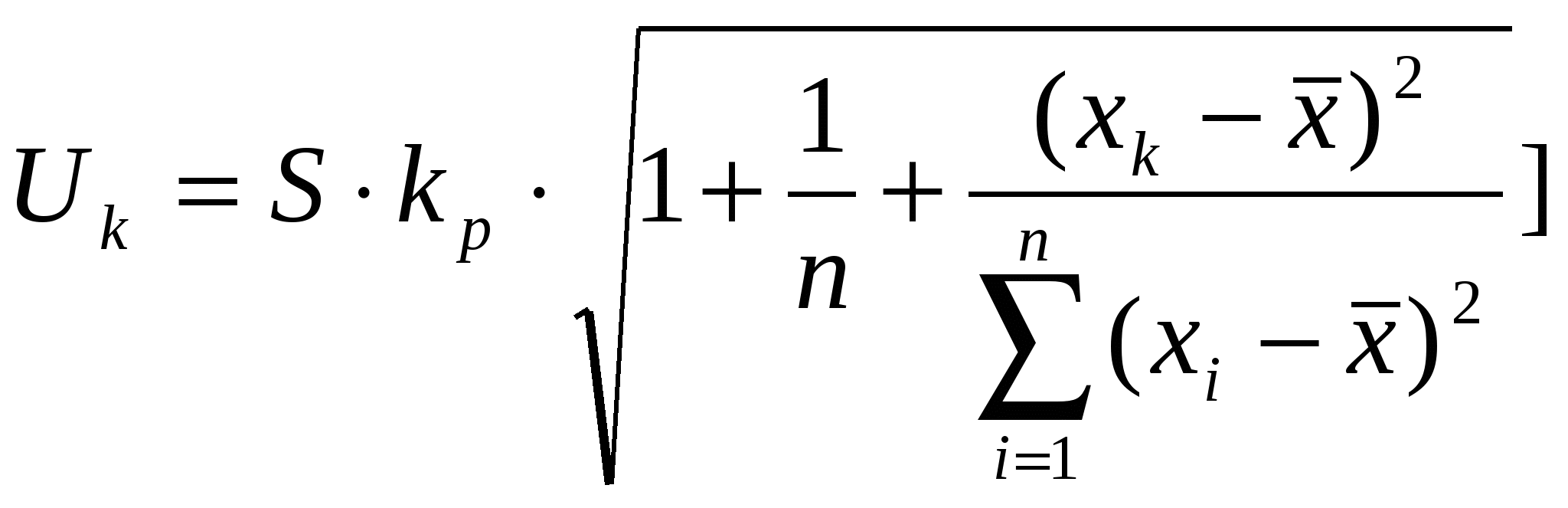
, при прогнозе на один шаг , на два шага и т.д.

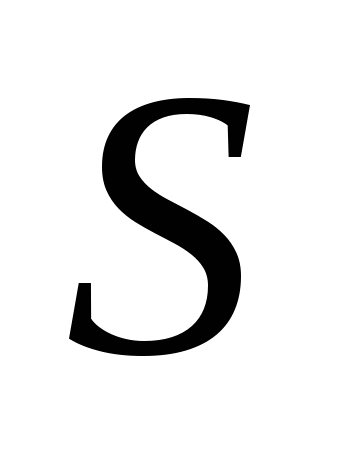
Подставляя полученное значение в уравнение регрессии, получим точечный прогноз величины . Однако вероятность точного "попадания" значения в эту точку достаточно мала. Поэтому представляет интерес вычисление перспективных оценок значений в виде доверительных интервалов:

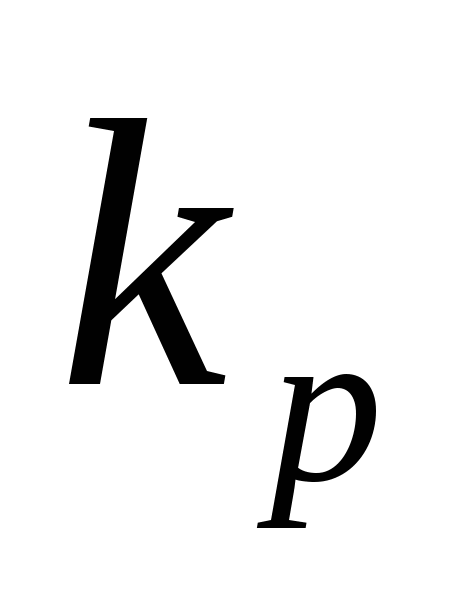
, где

- точечный прогноз величины для шага ,

- величина отклонения от точечного значения, соответствующая точке и заданному уровню вероятности. Величина для линейной модели рассчитывается по формуле:

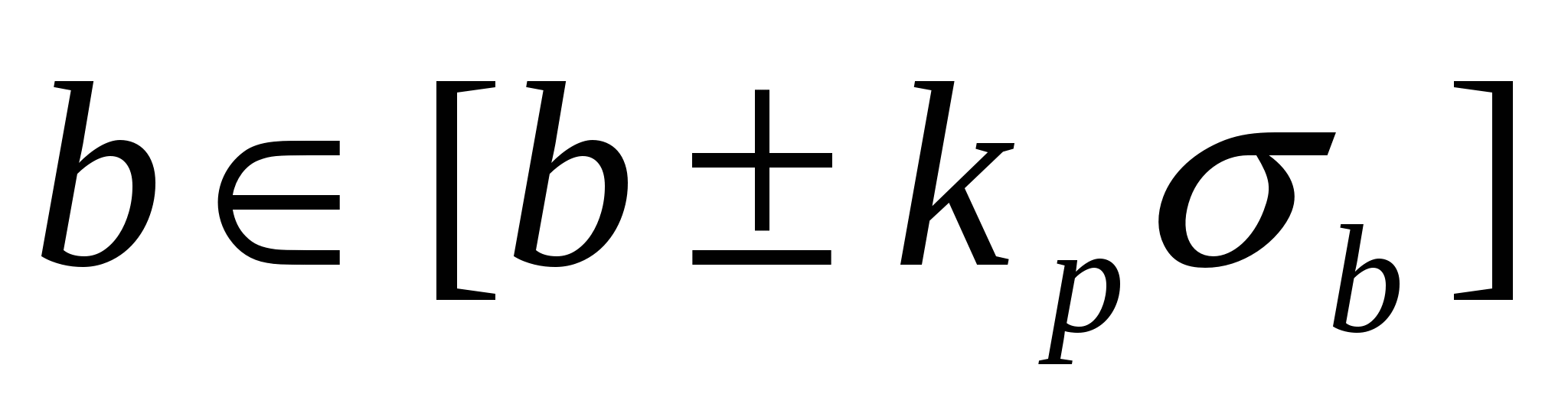
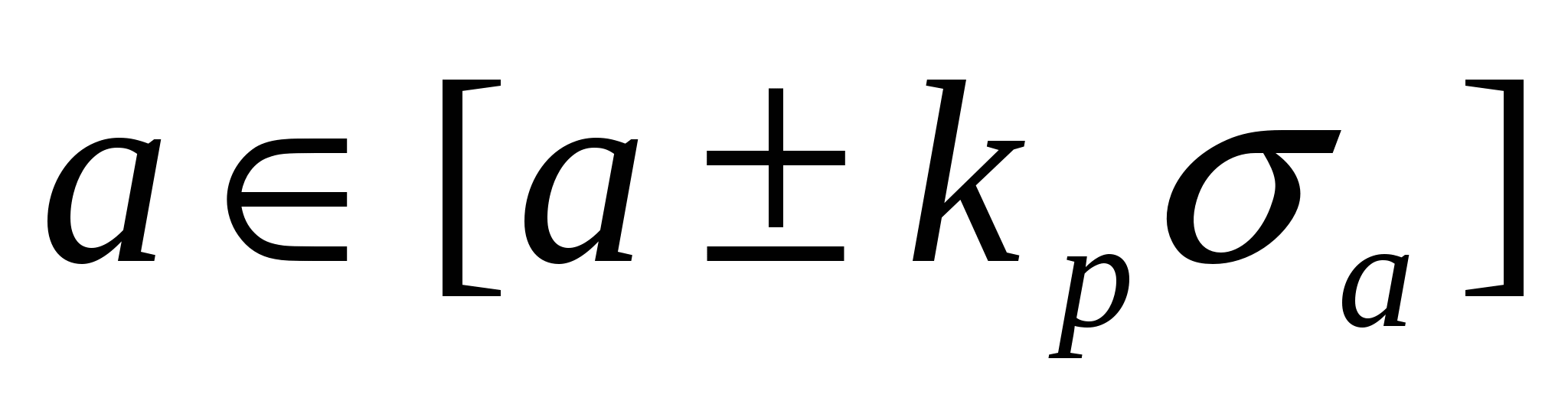
, где

- среднеквадратическое отклонение значений остаточного ряда;

- табличное значение t-статистики Стьюдента (соответствующая статистическая таблица приведена в приложении к лекции) для заданной вероятности попадания прогнозируемой величины внутрь доверительного интервала.

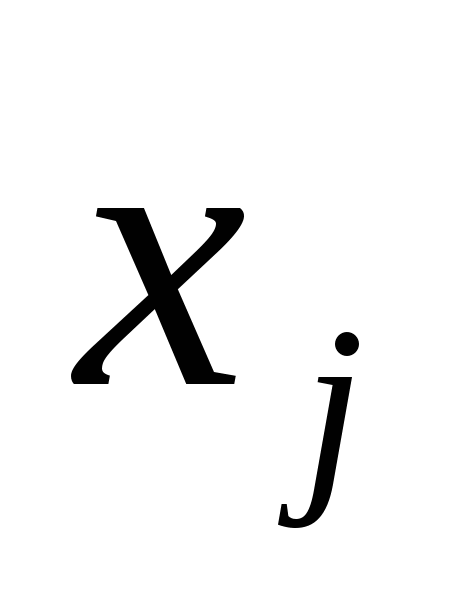
И если построенная модель регрессии адекватна, то с выбранной вероятностью можно утверждать, что при сохранении сложившихся закономерностей функционирования изучаемой системы прогнозируемая величина попадет в интервал, образованный нижней и верхней границами.

Иногда для значимого уравнения регрессии представляет интерес построение интервальных оценок параметров и :

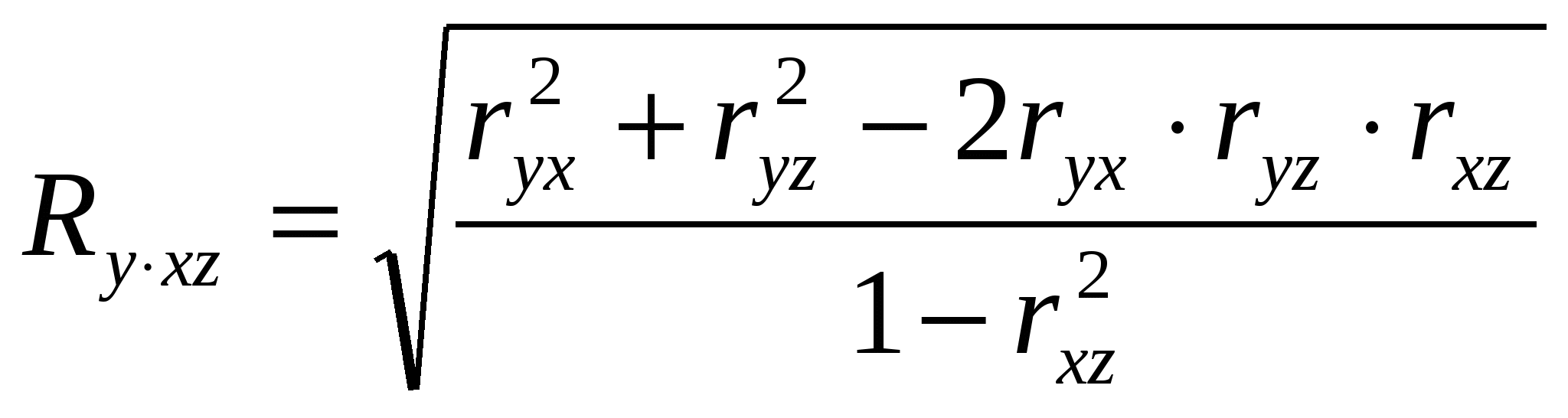


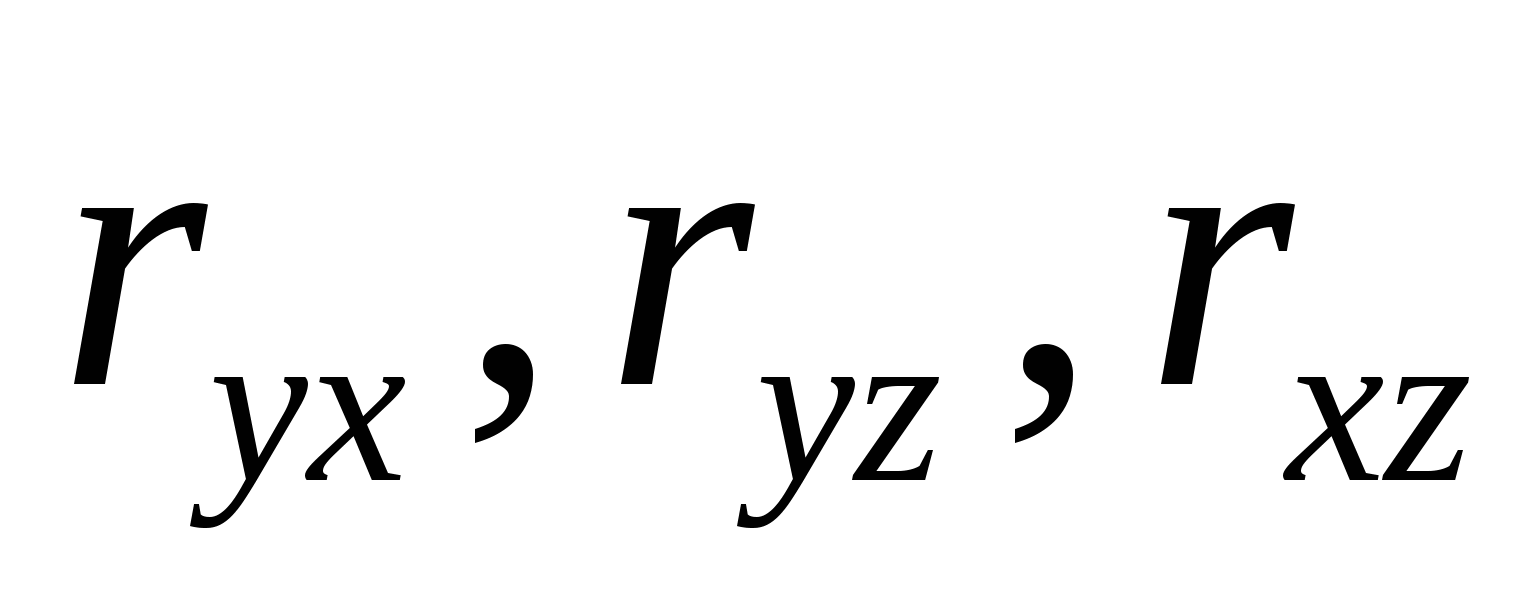
Множественная регрессия.

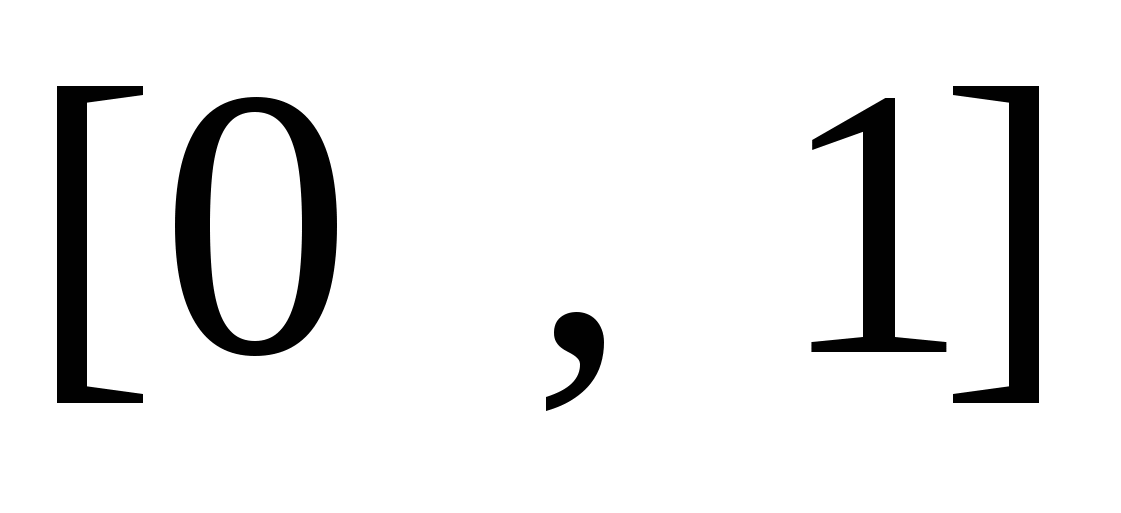
Экономические явления чаще всего адекватно описываются именно многофакторными моделями. Поэтому возникает необходимость обобщить этап корреляционного анализа на случай нескольких переменных.

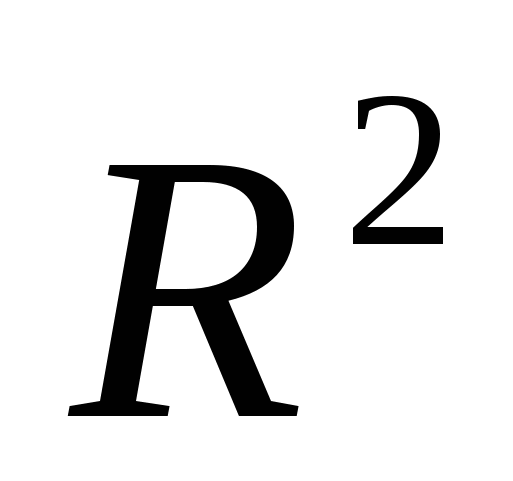
Теснота линейной взаимосвязи между переменной и рядом переменных , рассматриваемых в целом, может быть определена с помощьюкоэффициента множественной корреляции.

Предположим, что переменная испытывает влияние двух переменных - и . В этом случае коэффициент множественной корреляции может быть определен по формуле:

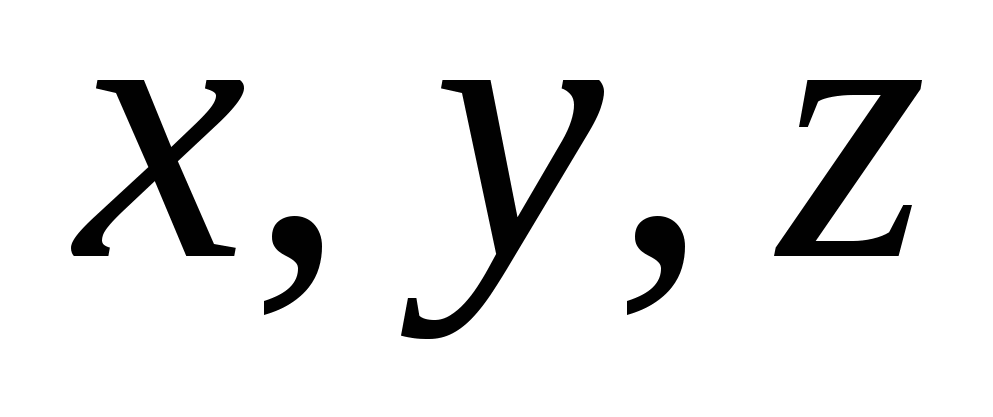


где - простые коэффициенты линейной парной корреляции.

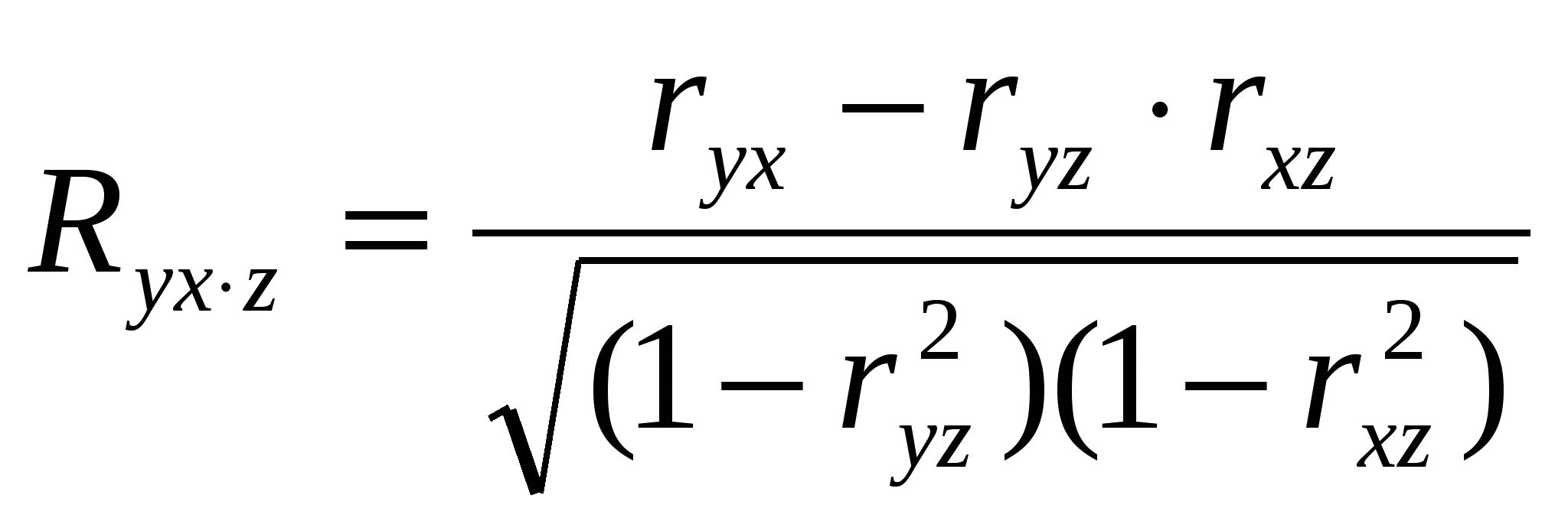
Коэффициент множественной корреляции заключен в пределах . Он не меньше, чем абсолютная величина любого парного или частного коэффициента корреляции с таким же первичным индексом.

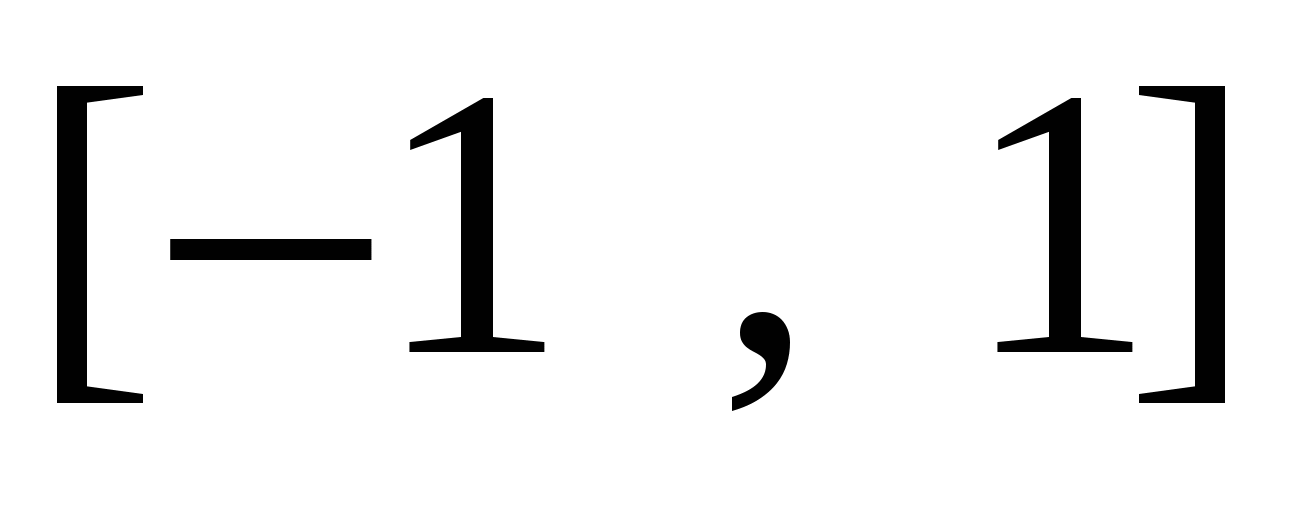
С помощью множественного коэффициента (по мере приближения 1) делается вывод о тесноте взаимосвязи, но не об ее направлении. Величина , называемая множественным коэффициентом детерминации, показывает, какую долю вариации исследуемой переменной объясняет вариация остальных учтенных переменных и .

Иногда представляет интерес измерение частных зависимостей между величинами и при условии, что воздействие других факторов, принимаемых во внимание, устранено. В качестве соответствующих измерителей приняты коэффициенты частной корреляции.

Рассмотрим порядок расчета коэффициента частной корреляции для случая, когда во взаимосвязи находятся три случайные переменные , для них могут быть получены простые коэффициенты линейной парной корреляции. Однако большая величина этих коэффициентов может быть обусловлена не только тем, что соответствующие величины действительно связаны между собой, но и в силу того, что обе переменные испытывают сильное действие третьего фактора.

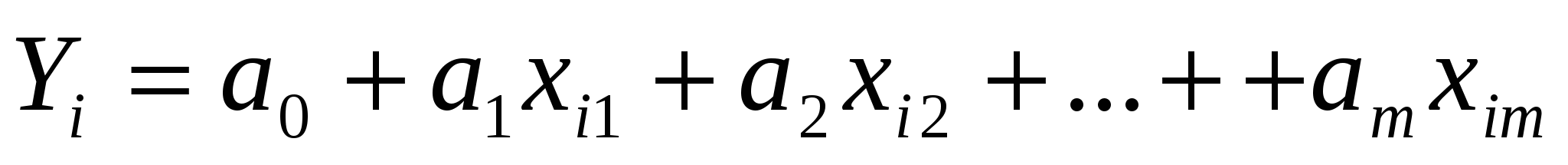
Коэффициент частной корреляции отличается от простого коэффициента линейной парной корреляции тем, что он измеряет парную корреляцию соответствующих признаков при условии, что влияние на них третьего фактора устранено. Соответствующая формула:

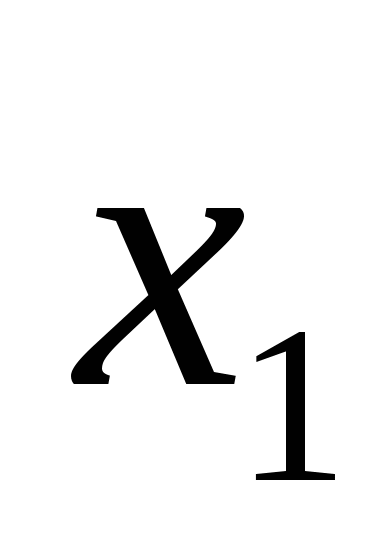
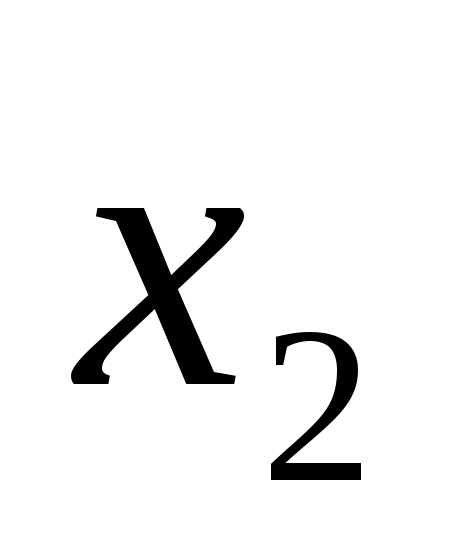


Частный коэффициент корреляции, так же как и парный коэффициент корреляции может принимать значения в диапазоне .

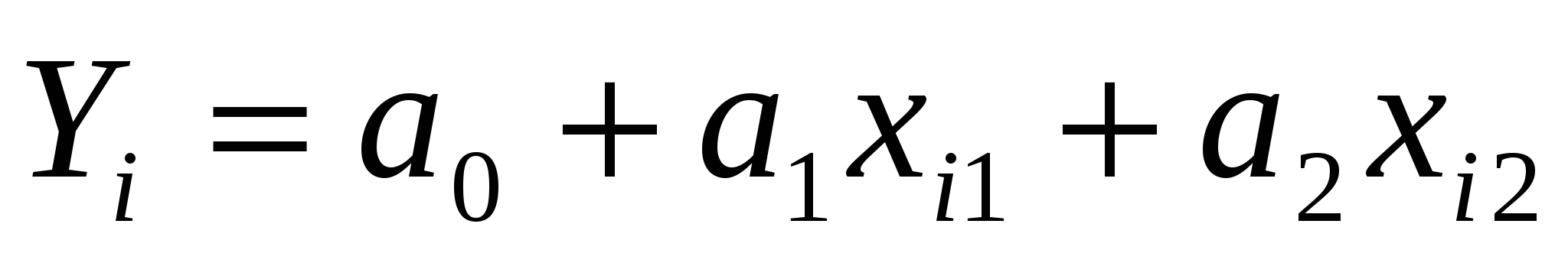
В тех случаях, когда необходимо оценить влияние нескольких факторов на исследуемую величину, строится уравнение множественной регрессии.

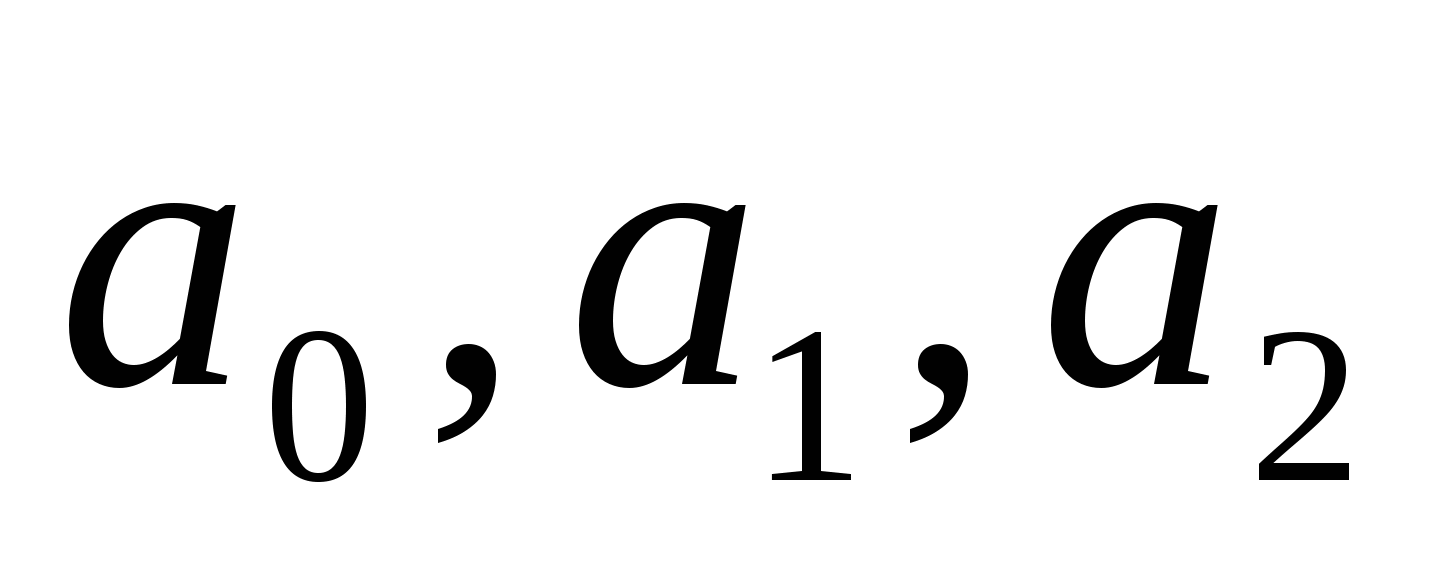
Если связь является линейной, то уравнение линейной множественной регрессии запишется в виде:



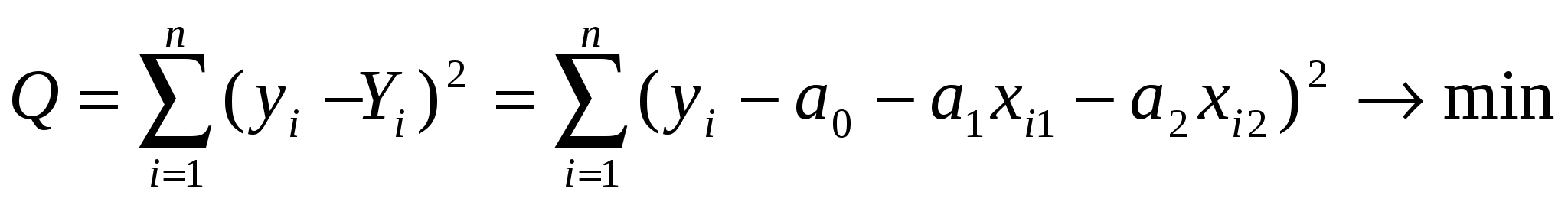
Рассмотрим случай, когда зависит от двух переменных и.

Уравнение с оцененными параметрами будет иметь вид:

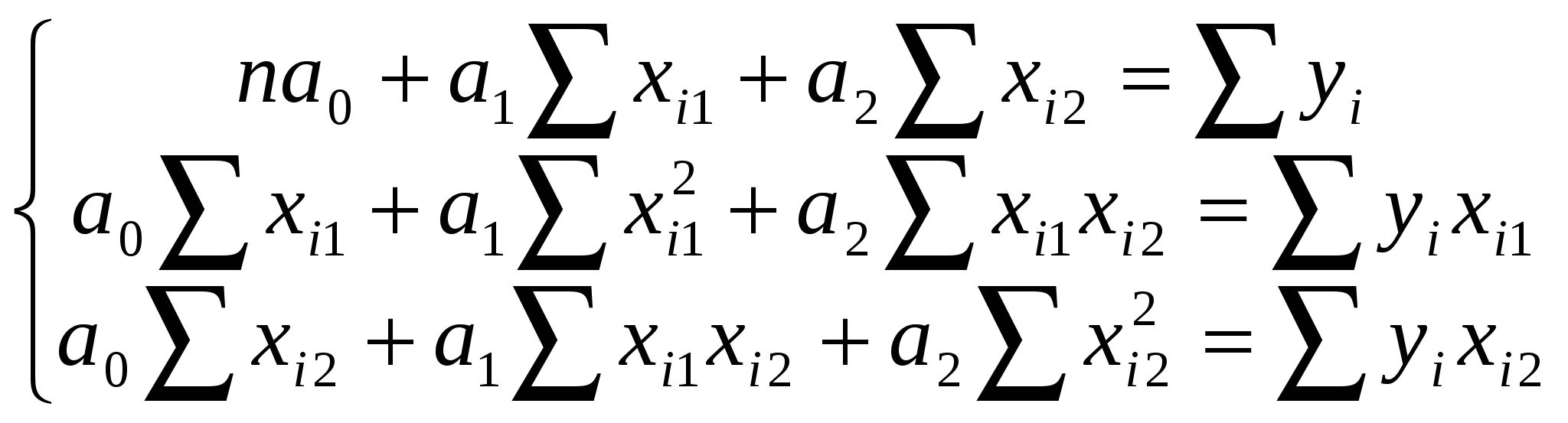


Чтобы определить значения коэффициентов , воспользуемся методом наименьших квадратов.

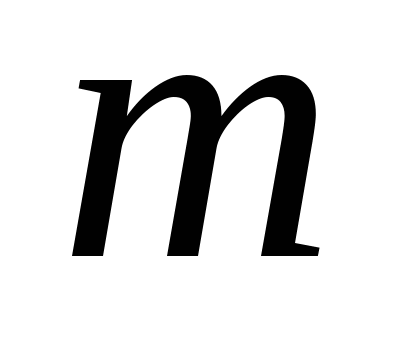
Как и ранее, задача формулируется следующим образом:

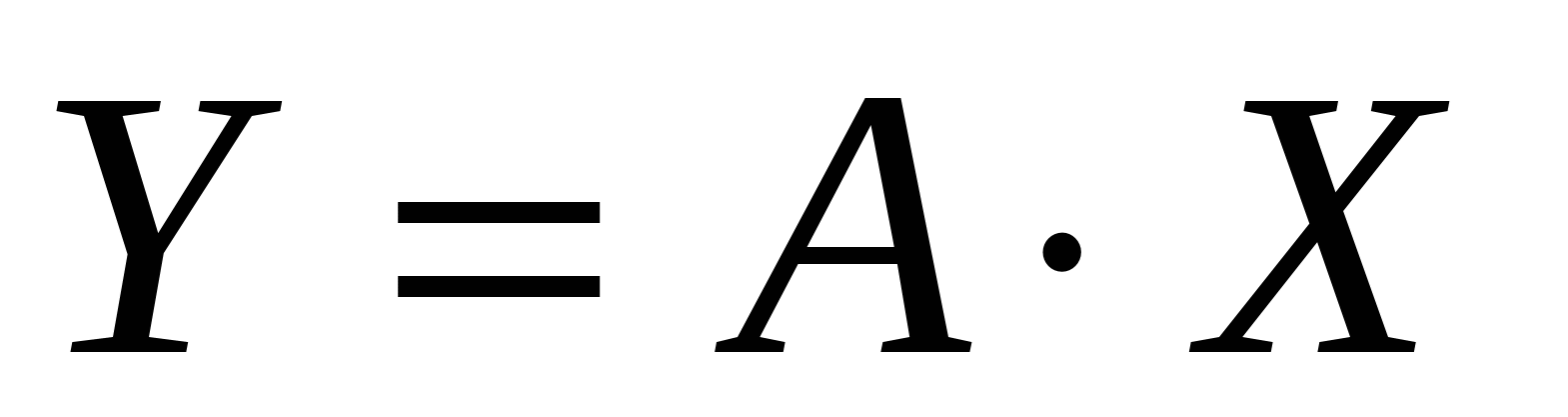


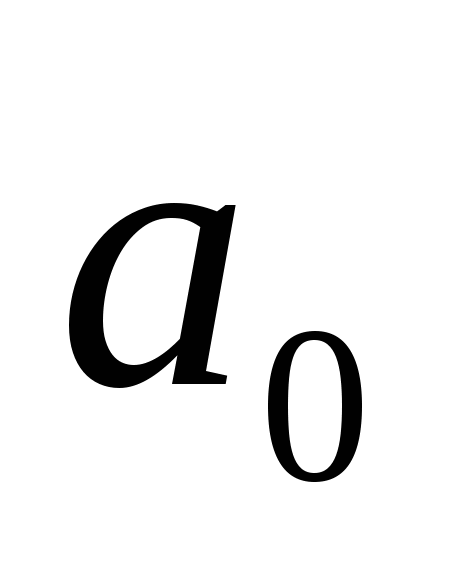
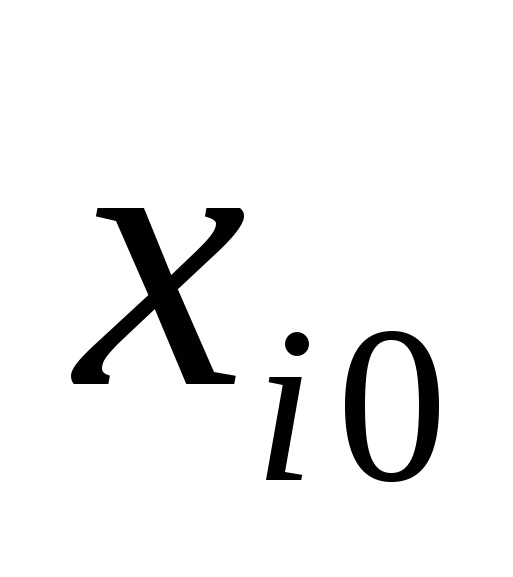
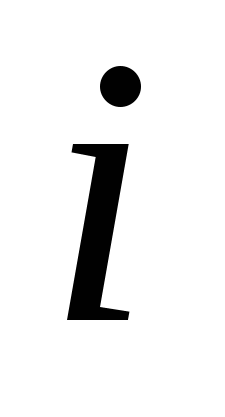
Приравняв частные производные нулю и выполнив преобразования, получим систему уравнений:



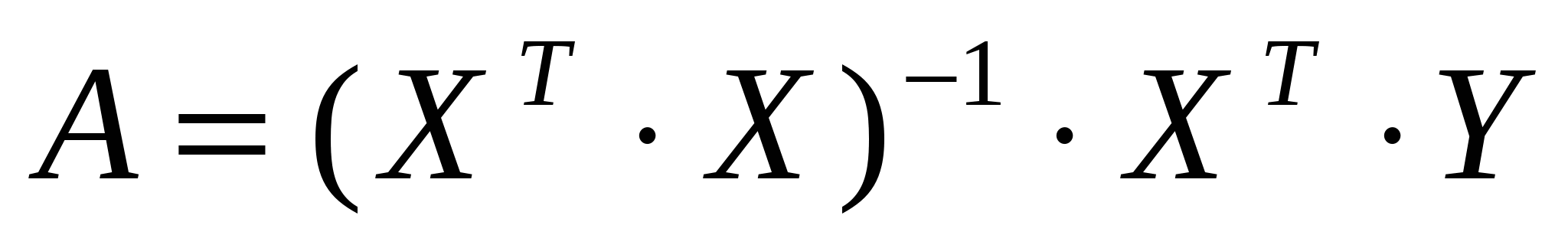
Решив систему, можно получить формулы для расчета коэффициентов уравнения множественной линейной регрессии.

Рассмотрим более общий случай - зависимость переменной от факторов, в матричном виде.



при этом в матрицу дополнительно введен столбец, все элементы которого равны 1, т.е. условно полагается, что свободный член умножается на фиктивную переменную, принимающую значение 1 для всех.

Можно показать, что для общего случая множественной линейной регрессии, коэффициенты уравнения могут быть определены из следующего соотношения:



Допустим, в результате анализа получено следующее уравнение регрессии:

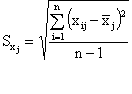
y = 2,4 + 0,8x1+ 3,2x2.

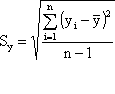
Если величины x1и x2являются соизмеримыми, то мы можем сопоставить влияние факторов x1и x2путем непосредственного сравнения соответствующих коэффициентов. В нашем примере можно сказать, что фактор x2воздействует на y в четыре раза сильнее.

В тех случаях, когда x1и x2измеряются в разных величинах для сравнения степени их влияния прибегают к нормированию коэффициентов регрессии и определяют так называемый бета-коэффициент (http://files.studfiles.ru/2706/176/html_Uir5GTcECm.pfrI/htmlconvd-IgQ3kx_html_m716acfc7.png):

j= aj∙http://files.studfiles.ru/2706/176/html_Uir5GTcECm.pfrI/htmlconvd-IgQ3kx_html_5e4a7133.png,

где aj- соответствующий коэффициент уравнения регрессии;

,  http://files.studfiles.ru/2706/176/html_Uir5GTcECm.pfrI/htmlconvd-IgQ3kx_html_6f5ebbb4.png- среднеквадратическое отклонение значений переменной xj(m – число учитываемых факторов);

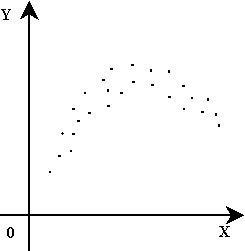
- среднеквадратическое отклонение значений переменной y.

Математически бета-коэффициент показывает, на какую часть величины среднеквадратического отклонения меняется среднее значение зависимой переменной с изменением независимой переменной на одно среднеквадратическое отклонение при фиксированном на постоянном уровне значении остальных независимых переменных.

Заметим, что некоторые авторы именуют бета-коэффициент стандартизированным коэффициентом регрессии.

Нелинейная регрессия

Пусть предварительный анализ исходной информации дает основание предполагать, что регрессионная зависимость носит нелинейный характер. Пример корреляционного поля, соответствующего нелинейной зависимости, представлен на рисунке 6.7.



***Рис. 6.7. Нелинейная зависимость***

Рассмотрим в качестве примера следующее уравнение регрессии:

http://files.studfiles.ru/2706/176/html_Uir5GTcECm.pfrI/htmlconvd-IgQ3kx_html_2da71370.png= a0+ a1x1+ a2http://files.studfiles.ru/2706/176/html_Uir5GTcECm.pfrI/htmlconvd-IgQ3kx_html_18cafb9a.png+ a3x2+ a4http://files.studfiles.ru/2706/176/html_Uir5GTcECm.pfrI/htmlconvd-IgQ3kx_html_12927fb6.png.

Пусть необходимо определить коэффициенты уравнения. В этом случае, как правило, выполняют линеаризующие преобразования переменных. Введем обозначения:

z1= x1;  z2= ;  z3= x2;   z4= .

Тогда исходное уравнение примет вид:

= a0+ a1z1+ a2z2+ a3z3+ a4z4.

Уравнение представляет собой уравнение линейной регрессии с четырьмя независимыми переменными. Коэффициенты последнего уравнения находятся по уже известной нам формуле.

A = (Zт∙Z)-1∙Zт∙Y.

После нахождения коэффициентов необходимо выполнить обратные преобразования для возврата к исходным переменным.