

МИНИСТЕРСТВО СЕЛЬСКОГО ХОЗЯЙСТВА РОССИЙСКОЙ ФЕДЕРАЦИИ  
ДЕПАРТАМЕНТ НАУЧНО-ТЕХНОЛОГИЧЕСКОЙ ПОЛИТИКИ И ОБРАЗОВАНИЯ  
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ  
ВЫСШЕГО ОБРАЗОВАНИЯ «ДОНСКОЙ ГОСУДАРСТВЕННЫЙ АГРАРНЫЙ УНИВЕРСИТЕТ»  
(ФГБОУ ВО Донской ГАУ)

АЗОВО-ЧЕРНОМОРСКИЙ ИНЖЕНЕРНЫЙ ИНСТИТУТ – ФИЛИАЛ  
ФЕДЕРАЛЬНОГО ГОСУДАРСТВЕННОГО БЮДЖЕТНОГО ОБРАЗОВАТЕЛЬНОГО  
УЧРЕЖДЕНИЯ ВЫСШЕГО ОБРАЗОВАНИЯ  
«ДОНСКОЙ ГОСУДАРСТВЕННЫЙ АГРАРНЫЙ УНИВЕРСИТЕТ» В Г. ЗЕРНОГРАДЕ  
(Азово-Черноморский инженерный институт ФГБОУ ВО Донской ГАУ)

Кафедра высшей математики и механики

Н.А. Коптева, Н.М. Удинцова

**ДОПОЛНИТЕЛЬНЫЕ ГЛАВЫ МАТЕМАТИКИ**  
**Часть 1:**  
**Корреляционно-регрессионный анализ**  
**в научных исследованиях**

*Учебное пособие*

Зерноград – 2018

УДК 33

*Печатается по решению методической комиссии  
по направлениям подготовки 13.03.02 и 13.04.02  
«Электроэнергетика и электротехника»*

*Азово-Черноморского инженерного института – филиала  
федерального государственного бюджетного образовательного учреждения  
высшего образования «Донской государственный аграрный университет»  
в г. Зернограде*

**Рецензенты:**

*канд. техн. наук, доцент кафедры «Высшая математика и механика»*

**Середина М.Н.,**

*канд. техн. наук, доцент кафедры «Теплоэнергетика  
и информационно-управляющие системы» Руденко Н.Б.*

**Коптева, Н.А.** Дополнительные главы математики. Часть 1: Корреляционно-регрессионный анализ в научных исследованиях: учебное пособие / Н.А. Коптева, Н.М. Удинцова. – Зерноград: Азово-Черноморский инженерный институт ФГБОУ ВО Донской ГАУ, 2018. – 100 с.

Учебное пособие составлено в соответствии с ФГОС ВО по направлению подготовки 13.04.02 «Электроэнергетика и электротехника».

Учебное пособие содержит теоретические сведения и методические указания к выполнению лабораторно-практических работ по темам: парная регрессия и корреляция, множественная регрессия и корреляция в научных исследованиях, а также контрольные вопросы по изучаемым разделам и задания для самостоятельного выполнения.

Учебное пособие предназначено для студентов-магистрантов, обучающихся по направлению подготовки 13.04.02 «Электроэнергетика и электротехника».

Рассмотрено и одобрено на заседании кафедры  
высшей математики и механики.

Протокол № 8 от 24 апреля 2018 г.

Рассмотрено и одобрено методической комиссией по направлениям  
подготовки 13.03.02 и 13.04.02 «Электроэнергетика и электротехника».

Протокол № 5 от 17 мая 2018 г.

© Коптева Н.А., Удинцова Н.М., 2018

© Азово-Черноморский инженерный  
институт – филиал ФГБОУ ВО  
Донской ГАУ, 2018

## СОДЕРЖАНИЕ

ВВЕДЕНИЕ .....	4
ПРЕДИСЛОВИЕ. Основные понятия моделирования и корреляционно-регрессионного анализа .....	6
<b>Лабораторно-практические занятия № 1, 2.</b> Уравнение парной линейной регрессии. Оценка тесноты связи между параметрами. ....	12
<b>Лабораторно-практические занятия № 3, 4.</b> Уравнения парной нелинейной регрессии. Параболическая, гиперболическая регрессия.....	35
<b>Лабораторно-практические занятия № 5, 6.</b> Корреляция для нелинейной регрессии. Коэффициент эластичности .....	52
<b>Лабораторно-практические занятия № 7, 8.</b> Синтезирование моделей множественной регрессии .....	64
ЛИТЕРАТУРА .....	91
Приложение 1. Варианты заданий для самостоятельной работы .....	92
Приложение 2. Шкала Чеддока .....	95
Приложение 3. Таблица значений t-критерия Стьюдента .....	96
Приложение 4. Таблица значений F-критерия Фишера.....	97
Приложение 5. Примерный вариант контрольной работы .....	98

## ВВЕДЕНИЕ

Материал, изложенный в учебном пособии, соответствует федеральному государственному образовательному стандарту высшего образования дисциплины «Дополнительные главы математики (в том числе оптимизация)», изучаемой студентами-магистрантами, обучающимися по направлению подготовки 13.04.02 «Электроэнергетика и электротехника».

Учебное пособие содержит теоретические сведения и методические указания к выполнению лабораторно-практических работ по темам: парная регрессия и корреляция, множественная регрессия и корреляция в научных исследованиях, а также контрольные вопросы по изучаемым разделам, помогающие усвоить и закрепить изучаемый материал и задания для самостоятельного выполнения.

Учебное пособие может быть полезно как преподавателям при проведении занятий со студентами, обучающимися по указанному направлению, так и студентам для самостоятельного изучения соответствующего материала и является базой для подготовки к зачету.

Учебное пособие способствует формированию у студентов, обучающихся по направлению подготовки 13.04.02 «Электроэнергетика и электротехника»:

- знаний методов количественного анализа и моделирования (ОК-3);
- базовых математических и экономических моделей (ОПК-2);
- умений применять методы моделирования в профессиональной деятельности, анализировать и оценивать социально-экономические явления, события, процессы (ОК-3);
- адаптировать основные математические модели к конкретным задачам управления (ОПК-2);
- владений навыками использования основных методов количественного анализа и моделирования, теоретического и экспериментального исследования (ОК-3);

– навыками разработки экономико-математических моделей социально-экономических процессов и явлений и способностью адаптировать их к конкретным задачам управления (ОПК-2).

Изучение корреляционно-регрессионного анализа позволит обучающимся приобрести необходимые навыки, требуемые для дальнейшего обучения; развить способность к саморазвитию, самореализации, использованию творческого потенциала, анализу и восприятию информации, постановке цели и выбору путей ее достижения, а также способность применять современные методы исследования в соответствии с поставленной задачей, анализировать, оценивать и представлять результаты выполненной работы и обосновывать полученные выводы.

Всё это понадобится для успешной работы и ориентации в будущей профессиональной деятельности.

## ПРЕДИСЛОВИЕ

### Основные понятия моделирования и корреляционно-регрессионного анализа

В разделе «Дополнительные главы математики» мы будем изучать конкретные количественные и качественные взаимосвязи объектов и процессов с помощью математических и статистических методов и моделей.

*Главной задачей* исследования является количественная оценка имеющихся взаимосвязей между изучаемыми явлениями и процессами.

*Основным предметом изучения* является модель некоторого объекта. В основе любого научного исследования лежит построение модели, адекватной изучаемым реальным явлениям и процессам.

**Определение.** *Математическая модель* – это приближённое описание какого-либо класса явлений внешнего мира, выраженное с помощью математической символики. Чаще всего это уравнения и (или) неравенства между показателями (переменными), характеризующими функционирование моделируемой реальной системы.

**Определение.** *Вероятностная модель* – это математическая модель, имитирующая механизм функционирования гипотетического (не конкретного) реального явления (или системы) стохастической или вероятностной природы.

**Определение.** *Вероятностно-статистическая модель* – это вероятностная модель, значения отдельных характеристик (параметров) которой оцениваются по результатам наблюдений, то есть по имеющимся статистическим данным.

**Определение.** Вероятностно-статистическая модель, описывающая механизм функционирования реально существующей экономической или социально-экономической системы, называется *эконометрической моделью*.

Основным инструментом, используемым для построения эконометрических моделей, являются *методы корреляционного и регрессионного анализа*.

Корреляционный анализ ставит своей целью проверку наличия и значимости зависимости между переменными без разделения переменных на зависимые и объясняющие. Ответ на эти вопросы дается с помощью вычисления показателей (коэффициентов) корреляции.

Регрессионный анализ направлен на выражение изучаемой зависимости в виде аналитической формулы с предварительным выделением зависимых и объясняющих переменных.

**Определение.** *Факторными признаками (факторами)* называются признаки, обуславливающие изменения других, связанных с ними признаков. *Факторные* признаки называются также *независимыми* или *объясняющими* и обозначаются

$$x = (x_1, x_2, \dots, x_n).$$

**Определение.** *Результативными* называются признаки, изменяющиеся под действием факторных признаков.

Результативные признаки называются также *зависимыми* или *объясняемыми* или *выходными переменными* и обозначаются

$$y = (y_1, y_2, \dots, y_n).$$

**Определение.** *Корреляционной* называется связь, при которой изменение среднего значения результативного признака обусловлено изменением факторных признаков.

**Определение.** Уравнение корреляционной связи между фактором  $x$  и результатом  $y$ , выраженное в виде математической зависимости, называется *уравнением регрессии*.

В обобщенной форме эконометрическая модель, описывающая взаимосвязи между явлениями или закономерности их развития, представляется с помощью соотношения:

$$y = f(a, x) + \varepsilon, \tag{1}$$

где  $f(a, x)$  – функционал, выражающий вид и структуру взаимосвязей;

величина  $y$  – зависимая переменная или результативный признак;

величина  $x = (x_1, x_2, \dots, x_n)$  – вектор значений независимых переменных  $x_i$   
или факторных признаков (факторов);

$a = (a_0, a_1, a_2, \dots, a_n)$  – вектор некоторых произвольных констант,  
называемых параметрами модели;

$\varepsilon$  – ошибка модели.

Ошибка модели  $\varepsilon$  характеризует отличие наблюдаемых (экспериментальных) значений переменной  $y_i$  от вычисленных согласно соотношению (1) при конкретных значениях переменных факторов  $x_i$  и рассматривается как случайная величина, то есть

$$\varepsilon_i = y_i - \hat{y}_i,$$

где  $\hat{y}_i = f(a, x_i)$  – значения результативного признака, вычисленные согласно соотношения (1) при конкретных значениях переменных факторов  $x_i$ .

Для расчета численных значений параметров  $a_0, a_1, a_2, \dots, a_n$  используется статистический (экспериментальный) массив наблюдений за совместным проявлением изучаемого процесса и рассматриваемых факторов.

В зависимости от количества факторов, включенных в уравнение регрессии, принято различать *простую (парную)* и *множественную* регрессию.

**Определение.** *Парной регрессией* называется модель, выражающая зависимость среднего значения зависимой переменной  $y$  от одной независимой переменной  $x$ , то есть это модель вида

$$\hat{y} = f(x), \tag{2}$$

где  $\hat{y}$  – зависимая переменная (результативный признак);

$x$  – независимая, объясняющая переменная (факторный признак).

По виду аналитической зависимости различают *линейные* и *нелинейные* регрессии.

*Линейная парная регрессия* описывается уравнением:

$$\hat{y} = a_0 + a_1 \cdot x. \tag{3}$$



Если между изучаемыми явлениями существуют нелинейные соотношения, то они выражаются с помощью уравнений нелинейной регрессии.

Примеры наиболее часто используемых нелинейных регрессий:

- параболическая

$$\hat{y} = a_0 + a_1 \cdot x + a_2 \cdot x^2 \quad (4)$$

- гиперболическая

$$\hat{y} = a_0 + \frac{a_1}{x}, \quad (5)$$

- полулогарифмическая

$$\hat{y} = a_0 + a_1 \cdot \ln x, \quad (6)$$

- степенная

$$\hat{y} = a_0 \cdot x^{a_1}, \quad (7)$$

- показательная

$$\hat{y} = a_0 \cdot a_1^x. \quad (8)$$

**Определение.** *Множественной регрессией* называют модель, выражающую зависимость среднего значения зависимой переменной  $y$  от нескольких независимых переменных  $x_1, x_2, \dots, x_n$ , то есть модель вида

$$\hat{y} = f(x_1, x_2, \dots, x_n). \quad (9)$$

**Замечание.** Парная регрессия применяется, если имеется доминирующий фактор, обуславливающий большую долю изменения изучаемой объясняемой переменной, который и используется в качестве объясняющей переменной.

Множественная регрессия применяется в ситуациях, когда из множества факторов, влияющих на результативный признак, нельзя выделить один доминирующий фактор и необходимо учитывать одновременное влияние нескольких факторов.

Фактические значения результативного признака  $y$  отличаются от теоретических, рассчитанных по уравнению регрессии, то есть  $\hat{y}$ . Чем меньше эти отличия, тем ближе теоретические значения к эмпирическим данным, тем лучше качество модели.

Величина отклонений фактических и расчетных значений результативного признака по каждому наблюдению  $(y_i - \hat{y}_i)$  представляет собой *ошибку аппроксимации*.

Поскольку  $(y_i - \hat{y}_i)$  может быть величиной как положительной, так и отрицательной, ошибки аппроксимации для каждого наблюдения принято определять в процентах по модулю.

Отклонения  $(y_i - \hat{y}_i)$  можно рассматривать как *абсолютную ошибку аппроксимации*, а  $\left| \frac{y_i - \hat{y}_i}{y_i} \right| \cdot 100$  – как *относительную ошибку аппроксимации*.

Для того чтобы иметь общее суждение о качестве модели, из относительных отклонений по каждому наблюдению находят *среднюю ошибку аппроксимации*:

$$\bar{A} = \frac{1}{n} \cdot \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \cdot 100 \% . \quad (10)$$

Ошибка аппроксимации в пределах 5–7% свидетельствует о хорошем подборе модели к исходным данным.

В экономических исследованиях широкое применение находит такой показатель, как *коэффициент эластичности*.

Если зависимость между переменными  $x$  и  $y$  имеет вид  $y = f(x)$ , то *коэффициент эластичности*  $\mathcal{E}$  вычисляется по формуле

$$\mathcal{E} = f'(x) \cdot \frac{x}{y}, \quad (11)$$

где  $f'(x)$  – первая производная, характеризующая соотношение приростов результата и фактора для соответствующей формы связи.

*Коэффициент эластичности*  $\mathcal{E}$  показывает на сколько процентов в среднем изменится результативный признак  $y$  при изменении фактора  $x$  на 1% от своего номинального значения.

Парная регрессия может дать хороший результат при моделировании, если влиянием других факторов, воздействующих на объект исследования, можно пренебречь.

Если же этим влиянием пренебречь нельзя, то в этом случае следует попытаться выявить влияние других факторов, введя их в модель, то есть построить уравнение множественной регрессии

$$\hat{y} = f(x_1, x_2, \dots, x_n),$$

где  $y$  – зависимая переменная (результативный признак),

$x_i$  – независимые, или объясняющие, переменные (признаки-факторы).

Множественная регрессия широко используется в решении проблем спроса, доходности акций, при изучении функции издержек производства, в макроэкономических расчетах и целом ряде других вопросов эконометрики.

В настоящее время множественная регрессия – один из наиболее распространенных методов в эконометрике. Основная цель множественной регрессии – построить модель с большим числом факторов, определив при этом влияние каждого из них в отдельности, а также совокупное их воздействие на моделируемый показатель.

### **Контрольные вопросы**

1. Дайте определение математической модели, вероятностной модели, вероятностно-статистической модели, эконометрической модели.
2. Какие задачи решают корреляционный и регрессионный анализы?
3. Какие признаки называются факторными, а какие – результативными?
4. Какая связь называется корреляционной?
5. Дайте определение уравнения регрессии.
6. Какие виды аналитических зависимостей наиболее часто используются при построении моделей?
7. Объясните понятия парной и множественной регрессии.
8. Дайте определение средней ошибки аппроксимации.
9. Дайте определение коэффициента эластичности.
10. В каких случаях применяют модель парной регрессии, а в каких множественной?

## ЛАБОРАТОРНО-ПРАКТИЧЕСКИЕ ЗАНЯТИЯ № 1, 2

**Тема:** Уравнение парной линейной регрессии. Оценка тесноты связи между параметрами.

**Цель работы:** Познакомиться с уравнениями парной линейной регрессии, научиться находить параметры уравнения линейной регрессии с использованием Excel и среднюю ошибку аппроксимации, вычислять линейный коэффициент корреляции, детерминации, интерпретировать полученные результаты.

### 1. Построение уравнения парной линейной регрессии

По имеющимся данным  $n$  наблюдений за совместным изменением двух переменных показателей  $x$  и  $y$   $\{(x_i, y_i), i=1, 2, \dots, n\}$  необходимо определить аналитическую зависимость  $\hat{y} = f(x)$ , наилучшим образом описывающую данные наблюдений.

Результаты наблюдений удобно представлять в виде таблицы:

Таблица 1 – Данные наблюдений

	$x_i$	$y_i$
1	$x_1$	$y_1$
2	$x_2$	$y_2$
3	$x_3$	$y_3$
...	...	...
$n$	$x_n$	$y_n$

Каждая строка таблицы представляет собой результат одного наблюдения  $(x_i, y_i)$ .

Поясним понятие зависимости  $\hat{y} = f(x)$ , наилучшим образом описывающей данные наблюдений. Значения  $x_i, y_i$  из каждой строки можно рассматривать как координаты точки  $(x_i, y_i)$  на координатной плоскости  $Oxy$ .

**Определение.** Совокупность всех точек  $(x_i, y_i)$  на координатной плоскости  $Oxy$  называют *полем корреляции* (рисунок 1):

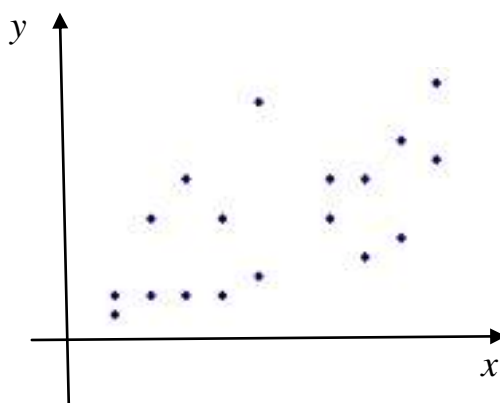


Рисунок 1 – Поле корреляции

Зависимости  $\hat{y} = f(x)$  соответствует некоторая кривая на плоскости. Чем ближе данная кривая подходит ко всем точкам поля корреляций, тем лучше зависимость  $\hat{y} = f(x)$  описывает исходные данные.

На рисунке 2 изображена лучшая линейная регрессия.

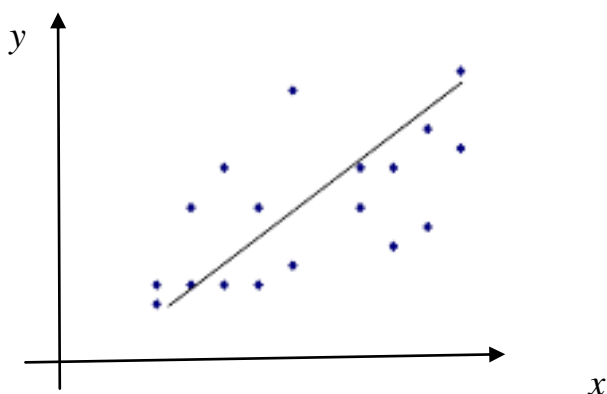


Рисунок 2 – Лучшая линейная регрессия

Линейная парная регрессия описывается уравнением

$$\hat{y} = a_0 + a_1 \cdot x. \quad (12)$$

Построение линейной регрессии сводится к оценке ее параметров  $a_0$  и  $a_1$ .

Классический подход к определению параметров линейной регрессии основан на *методе наименьших квадратов* (МНК). Суть его заключается в том, чтобы сумма квадратов отклонений фактических значений результативного признака  $y$  от расчетных (теоретических)  $\hat{y}$  была минимальной:

$$S = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \rightarrow \min, \quad (13)$$

где  $y_i$  – фактическое значение;

$\hat{y}_i$  – теоретическое или прогнозируемое значение:  $\hat{y}_i = \varphi(x, a_0, a_1)$ .

С учетом вида линейной парной регрессии (12) величина  $S$  является функцией неизвестных параметров  $a_0$  и  $a_1$ :

$$S(a_0, a_1) = \sum_{i=1}^n (y_i - a_0 - a_1 x_i)^2 \rightarrow \min. \quad (14)$$

Из математического анализа известно, что задача на нахождение минимума функции относится к классу экстремальных. Так как в данной задаче неизвестны  $a_0$  и  $a_1$ , то необходимо найти частные производные:

$$\begin{cases} \frac{\partial S}{\partial a_0} = 0 \\ \frac{\partial S}{\partial a_1} = 0 \end{cases}. \quad (15)$$

Выполняя соответствующие вычисления, получим для определения параметров  $a_0$  и  $a_1$  следующую систему уравнений:

$$\begin{cases} 2(-\sum_{i=1}^n y_i + \sum_{i=1}^n a_0 + \sum_{i=1}^n a_1 x_i) = 0 & / : 2 \\ 2(-\sum_{i=1}^n y_i x_i + \sum_{i=1}^n a_0 x_i + \sum_{i=1}^n a_1 x_i^2) = 0 & / : 2 \end{cases}. \quad (16)$$

Разделив каждое уравнение системы (17) на 2, получим:

$$\begin{cases} -\sum_{i=1}^n y_i + \sum_{i=1}^n a_0 + \sum_{i=1}^n a_1 x_i = 0 \\ -\sum_{i=1}^n y_i x_i + \sum_{i=1}^n a_0 x_i + \sum_{i=1}^n a_1 x_i^2 = 0 \end{cases}. \quad (17)$$

Рассмотрим второе слагаемое первого уравнения системы (18):

$$\sum_{i=1}^n a_0 = \underbrace{a_0 + a_0 + \dots + a_0}_{n \text{ раз}} = a_0 n. \quad (18)$$

Тогда окончательно получим систему уравнений для оценки параметров  $a_0$  и  $a_1$ :

$$\begin{cases} a_0 n + a_1 \sum_{i=1}^n x_i = \sum_{i=1}^n y_i \\ a_0 \sum_{i=1}^n x_i + a_1 \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i y_i \end{cases}. \quad (19)$$

Для решения указанной системы (19) используем формулы Крамера:

$$a_0 = \frac{\Delta a_0}{\Delta}, \quad a_1 = \frac{\Delta a_1}{\Delta}, \quad (20)$$

где  $\Delta$  – определитель основной матрицы системы:

$$\Delta = \begin{vmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{vmatrix}, \quad (21)$$

$\Delta a_0$  и  $\Delta a_1$  – определители, получаемые из  $\Delta$  заменой 1-го и 2-го столбца соответственно на столбец свободных членов (правых частей уравнений системы):

$$\Delta a_0 = \begin{vmatrix} \sum_{i=1}^n y_i & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i y_i & \sum_{i=1}^n x_i^2 \end{vmatrix}, \quad (22)$$

$$\Delta a_1 = \begin{vmatrix} n & \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i y_i \end{vmatrix}. \quad (23)$$

Для того чтобы иметь общее суждение о качестве модели, находят *среднюю ошибку аппроксимации*:

$$A = \frac{1}{n} \cdot \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \cdot 100 \%. \quad (24)$$

Ошибка аппроксимации в пределах 5–7% свидетельствует о хорошем подборе модели к исходным данным.

## 2. Пример построения уравнения линейной регрессии

Имеются данные об изменении расхода топлива  $Q$  (л/км) автомобиля при изменении скорости его движения  $V$  (км/ч) (таблица 2).

Получить уравнение линейной регрессии, отражающее зависимость расхода топлива от скорости автомобиля  $Q(V)$ . Оценить качество модели с помощью средней ошибки аппроксимации.

Таблица 2 – Исходные данные задачи

№	V(км/ч)	Q(л/ч)
1	33	0,15
2	36	0,18
3	54	0,26
4	56	0,29
5	79	0,45
6	57	0,29
7	68	0,42
8	73	0,41
9	57	0,28
10	48	0,26
11	61	0,34
12	59	0,32
13	65	0,37
14	47	0,23
15	58	0,31

Для определения коэффициентов линейной регрессии и средней ошибки аппроксимации в Excel необходимо рассчитать таблицу следующего вида (таблица 3):

Таблица 3 – Данные для определения коэффициентов линейной регрессии и средней ошибки аппроксимации

№	$X = V$	$Y = Q$	$X^2$	$XY$	$\hat{y}$	$\left  \frac{y_i - \hat{y}_i}{y_i} \right $
1	33	0,15	1089	4,950		
2	36	0,18	1296	6,480		
3	54	0,26	2916	14,040		
4	56	0,29	3136	16,240		
5	79	0,45	6241	35,550		
6	57	0,29	3249	16,530		
7	68	0,42	4624	28,560		
8	73	0,41	5329	29,930		
9	57	0,28	3249	15,960		
10	48	0,26	2304	12,480		
11	61	0,34	3721	20,740		
12	59	0,32	3481	18,880		
13	65	0,37	4225	24,050		
14	47	0,23	2209	10,810		
15	58	0,31	3364	17,980		
$\Sigma$	851	4,56	50433	273,18		



Две последние колонки таблицы заполним после того, как получим уравнение линейной регрессии.

Вычислим определители (21)–(23):

$$\Delta = \begin{vmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{vmatrix} = \begin{vmatrix} 15 & 851 \\ 851 & 50433 \end{vmatrix} = 32294;$$

$$\Delta a_0 = \begin{vmatrix} \sum_{i=1}^n y_i & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i y_i & \sum_{i=1}^n x_i^2 \end{vmatrix} = \begin{vmatrix} 4,56 & 258 \\ 273,18 & 14678 \end{vmatrix} = -2501,7;$$

$$\Delta a_1 = \begin{vmatrix} n & \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i y_i \end{vmatrix} = \begin{vmatrix} 15 & 4,56 \\ 851 & 273,18 \end{vmatrix} = 217,14.$$

На рисунке 3 изображен фрагмент листа Excel при вычислении определителей.

	A	B	C	D	E	F	G	H
7	5	79	0,45	6241	35,550			
8	6	57	0,29	3249	16,530			
9	7	68	0,42	4624	28,560			
10	8	73	0,41	5329	29,930			
11	9	57	0,28	3249	15,960			
12	10	48	0,26	2304	12,480			
13	11	61	0,34	3721	20,740			
14	12	59	0,32	3481	18,880			
15	13	65	0,37	4225	24,050			
16	14	47	0,23	2209	10,810			
17	15	58	0,31	3364	17,980			
18	Σ	851,000	4,560	50433	273,180			
19	Среднее	56,733	0,304	3362,2	18,212			
20								
21		15	851					
22	Δ=	851	50433 =					
23								
24		4,56	851 =					
25	Δa0=	273,18	50433					
26								
27	Δa1=	15	4,56 =					
28		851	273,18					

Рисунок 3 – Фрагмент листа Excel

**Замечание.** Для вычисления определителей в Excel необходимо использовать функцию **МОПРЕД**.

Например, если вы хотите вычислить определитель  $\Delta$ , поставьте курсор в пустую клетку правее определителя, как показано на рисунке 3, и выберите команду **Формулы – Вставить функцию** на панели в верхней части листа Excel (рисунок 4):

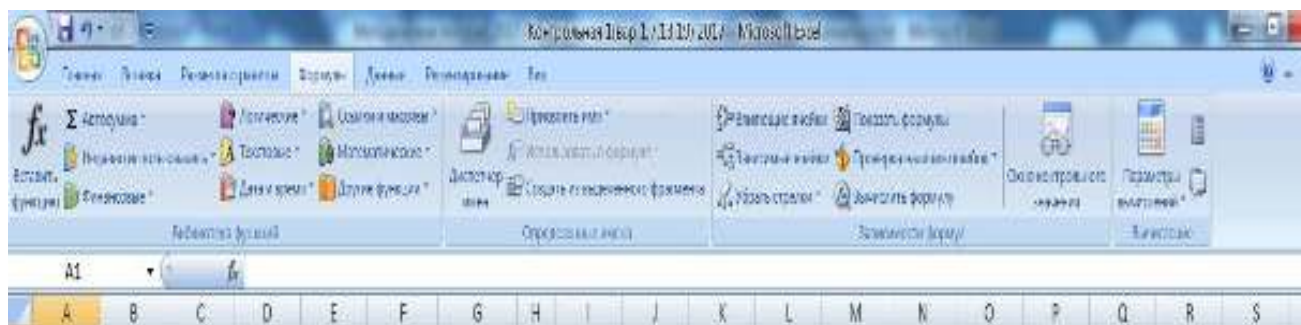


Рисунок 4 – Выбор команды Формулы – Вставить функцию

Вы вошли в Мастер функций (рисунок 5):

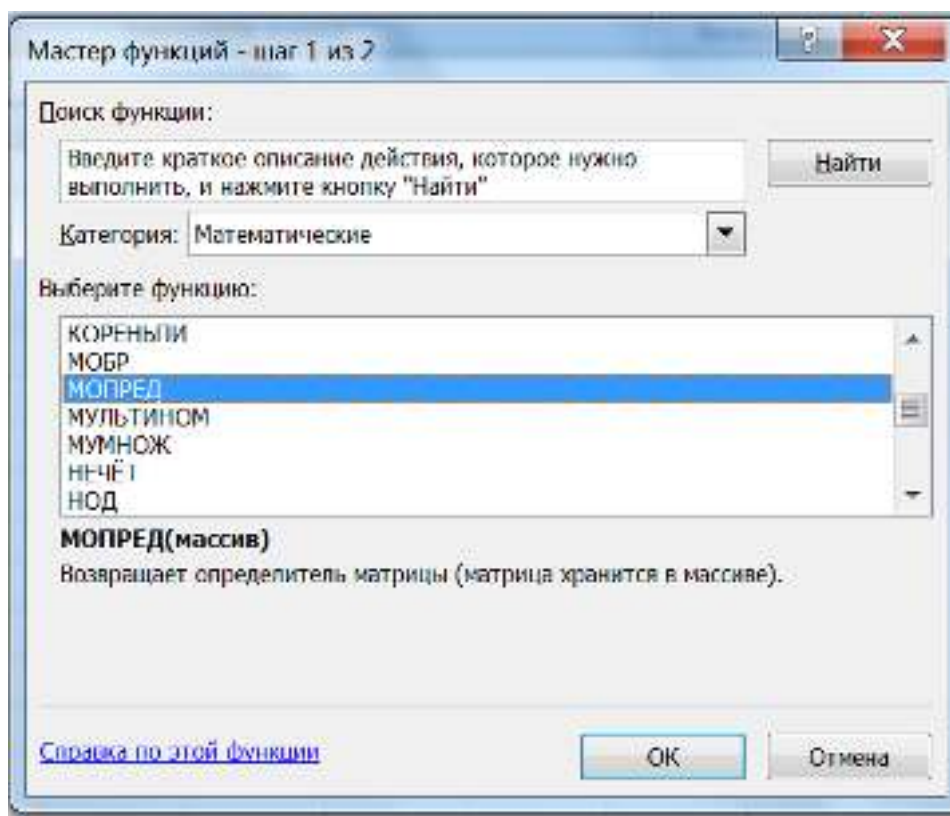
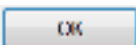


Рисунок 5 – Меню Мастера функций Excel

Найдите среди Математических функций функцию **МОПРЕД**, подтвердите выбор, нажав , появится окно (рисунок 6):

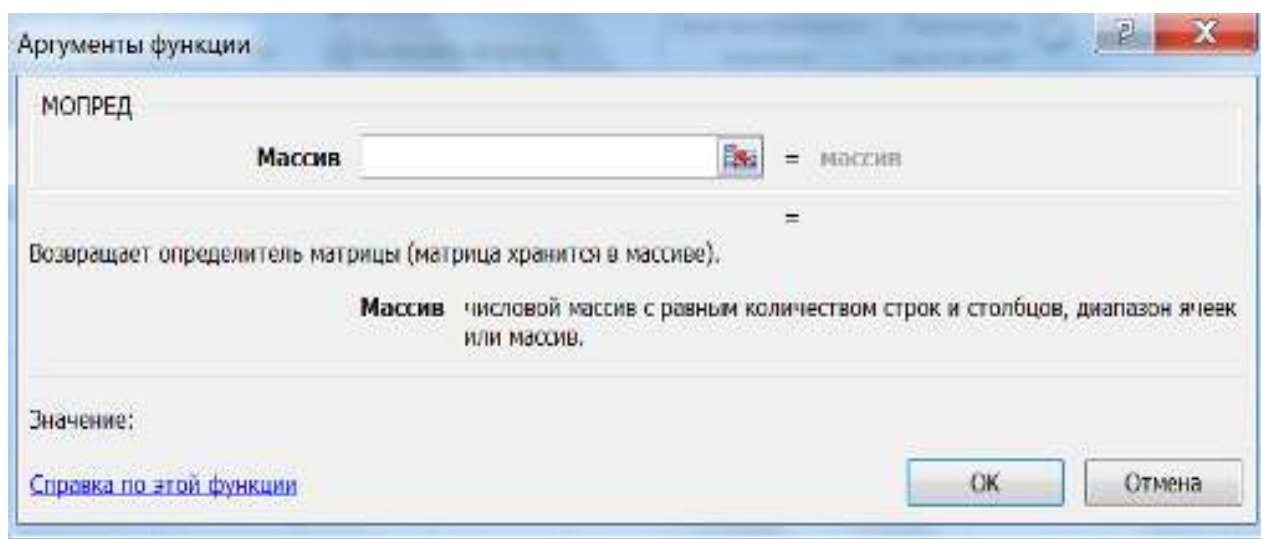



Рисунок 6 – Окно функции МОПРЕД

Выделите вычисляемый определитель, как показано на рисунке 7, и подтвердите, нажав , определитель будет посчитан.

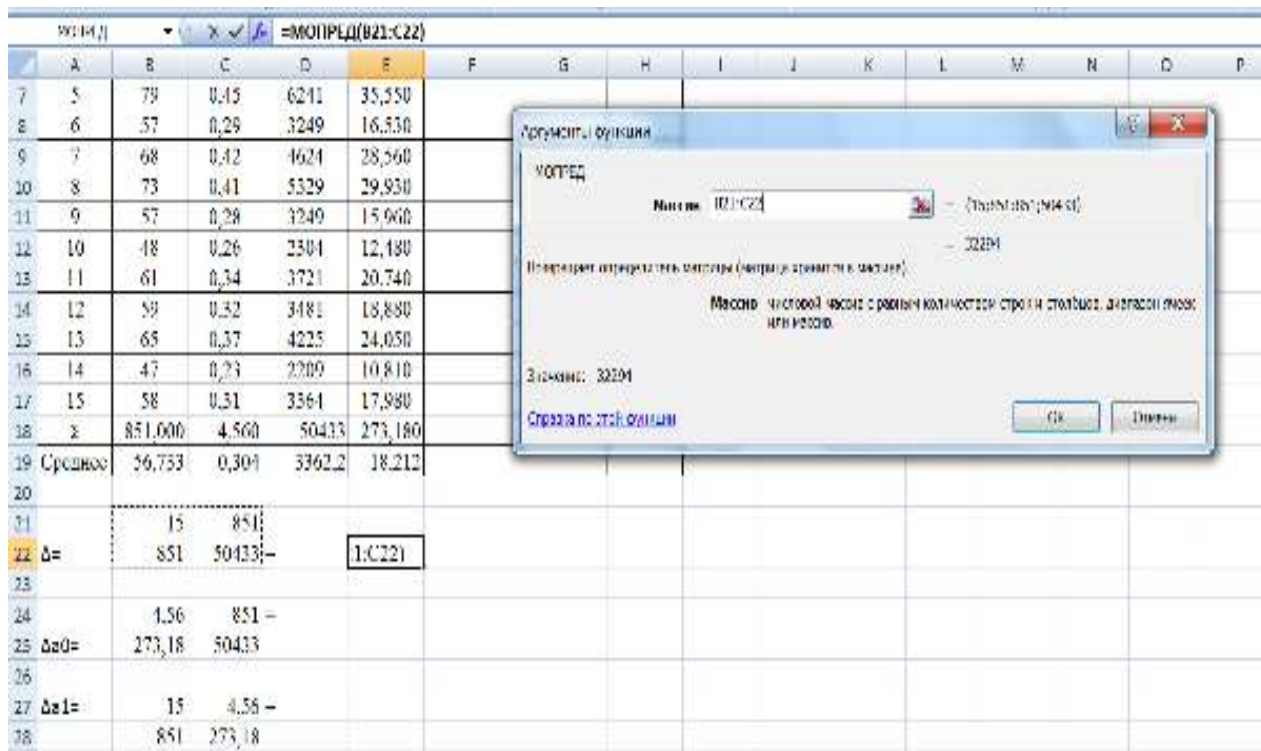


Рисунок 7 – Фрагмент листа Excel

Вычислим неизвестные коэффициенты уравнения линейной регрессии по формулам (20):

$$a_0 = \frac{\Delta a_0}{\Delta} = \frac{-2501,7}{32294} = -0,077;$$

$$a_1 = \frac{\Delta a_1}{\Delta} = \frac{217,14}{32294} = 0,007.$$

Полученные значения  $a_0$  и  $a_1$  подставим в уравнение линейной регрессии.

Найденное уравнение регрессии будет иметь вид:

$$\hat{y} = -0,077 + 0,007 \cdot x. \quad (25)$$

Чтобы проверить качество полученного уравнения линейной регрессии, заполним две последние колонки таблицы 3, подставив в  $\hat{y}$  значения  $a_0$  и  $a_1$ .

**Замечание.** При подстановке в таблицу значений  $a_0$  и  $a_1$  закрепить в ссылках адреса клеток, где находятся значения  $a_0$  и  $a_1$ , используя символ \$ (рисунок 8):

F3		f <sub>xy</sub>		=\$B\$30+\$B\$31*B3			
	A	B	C	D	E	F	G
1	Прямая (линейная регрессия)						
2	№	X	Y	X <sup>2</sup>	xy	$\hat{Y}$	$ (Y-\hat{Y})/Y $
3	1	33	0,15	1089	4,950	0,144	
4	2	36	0,18	1296	6,480		
5	3	54	0,26	2916	14,040		
6	4	56	0,29	3136	16,240		
7	5	79	0,45	6241	35,550		
8	6	57	0,29	3249	16,530		
9	7	68	0,42	4624	28,560		
10	8	73	0,41	5329	29,930		
11	9	57	0,28	3249	15,960		
12	10	48	0,26	2304	12,480		
13	11	61	0,34	3721	20,740		
14	12	59	0,32	3481	18,880		
15	13	65	0,37	4225	24,050		
16	14	47	0,23	2209	10,810		
17	15	58	0,31	3364	17,980		
18	Σ	851,000	4,560	50433	273,180		
19	Среднее	56,733	0,304	3362,2	18,212		

Рисунок 8 – Фрагмент листа Excel

Вычислите оставшиеся значения  $\hat{y}$  (таблица 4). Найдите сумму  $\hat{y}$  и убедитесь, что

$$\Sigma y = \Sigma \hat{y}.$$

Таблица 4 – Вычисление значений  $\hat{y}$

№	$X = V$	$Y = Q$	$XY$	$X^2$	$\hat{y}$	$\left  \frac{y_i - \hat{y}_i}{y_i} \right $
1	33	0,15	1089	4,950	0,144	
2	36	0,18	1296	6,480	0,165	
3	54	0,26	2916	14,040	0,286	
4	56	0,29	3136	16,240	0,299	
5	79	0,45	6241	35,550	0,454	
6	57	0,29	3249	16,530	0,306	
7	68	0,42	4624	28,560	0,380	
8	73	0,41	5329	29,930	0,413	
9	57	0,28	3249	15,960	0,306	
10	48	0,26	2304	12,480	0,245	
11	61	0,34	3721	20,740	0,333	
12	59	0,32	3481	18,880	0,319	
13	65	0,37	4225	24,050	0,360	
14	47	0,23	2209	10,810	0,239	
15	58	0,31	3364	17,980	0,313	
$\Sigma$	851,000	<b>4,560</b>	50433	273,180	<b>4,560</b>	

Для оценки качества модели найдем среднюю ошибку аппроксимации по формуле (24):

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

Для этого заполним последнюю колонку таблицы, вычислив  $\left| \frac{y_i - \hat{y}_i}{y_i} \right|$ .

**Замечание.** Чтобы вычислить *модуль*, воспользуйтесь функцией **ABS** (найти её можно среди Математических функций, выбрав команду **Формулы** – **Вставить функцию** в меню Excel) – рисунок 9.



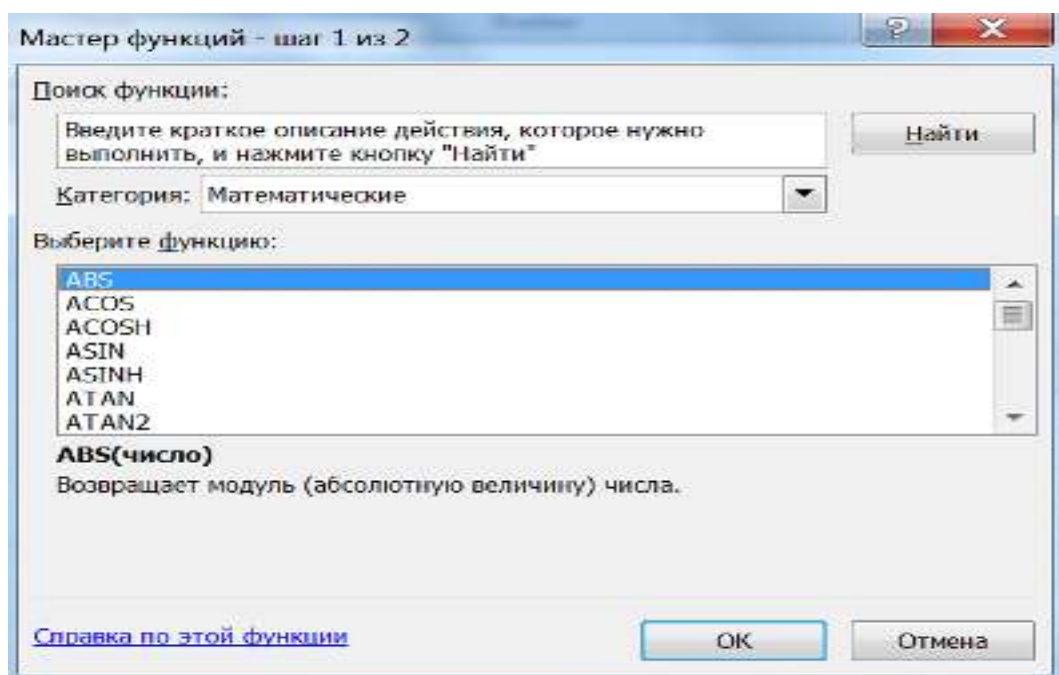


Рисунок 9 – Вызов функции ABS

ABS									
=ABS((C3-F3)/C3)									
	A	B	C	D	E	F	G	H	I
1	Прямая (линейная регрессия)								
2	№	X	Y	X <sup>2</sup>	XY	Ŷ	Y-Ŷ /Y		
3	1	33	0,15	1089	4,950	=ABS((C3-F3)/C3)			
4	2	36	0,18	1296	6,480	0,165			
5	3	54	0,26	2916	14,040	0,286			
6	4	56	0,29	3136	16,240	0,299			
7	5	79	0,45	6241	35,550	0,454			
8	6	57	0,29	3249	16,530	0,306			
9	7	68	0,42	4624	28,560	0,380			
10	8	73	0,41	5329	29,930	0,413			
11	9	57	0,28	3249	15,960	0,306			
12	10	48	0,26	2304	12,480	0,245			
13	11	61	0,34	3721	20,740	0,333			
14	12	59	0,32	3481	18,880	0,319			
15	13	65	0,37	4225	24,050	0,360			
16	14	47	0,23	2209	10,810	0,239			
17	15	58	0,31	3364	17,980	0,313			

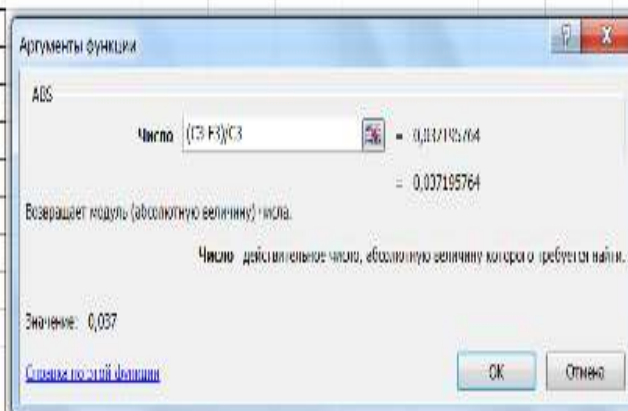


Рисунок 10 – Фрагмент листа Excel

Окончательный вид таблицы приведен ниже (таблица 5).

Таблица 5 – Вычисление значений  $\left| \frac{y_i - \hat{y}_i}{y_i} \right|$

$N$	$X = V$	$Y = Q$	$XY$	$X^2$	$\hat{y}$	$\left  \frac{y_i - \hat{y}_i}{y_i} \right $
1	33	0,15	1089	4,950	0,144	0,037
2	36	0,18	1296	6,480	0,165	0,086
3	54	0,26	2916	14,040	0,286	0,099
4	56	0,29	3136	16,240	0,299	0,031
5	79	0,45	6241	35,550	0,454	0,008
6	57	0,29	3249	16,530	0,306	0,054
7	68	0,42	4624	28,560	0,380	0,096
8	73	0,41	5329	29,930	0,413	0,008
9	57	0,28	3249	15,960	0,306	0,092
10	48	0,26	2304	12,480	0,245	0,057
11	61	0,34	3721	20,740	0,333	0,022
12	59	0,32	3481	18,880	0,319	0,002
13	65	0,37	4225	24,050	0,360	0,028
14	47	0,23	2209	10,810	0,239	0,037
15	58	0,31	3364	17,980	0,313	0,008
$\Sigma$	851,000	<b>4,560</b>	50433	273,180	<b>4,560</b>	0,665

Тогда средняя ошибка аппроксимации равна:

$$\bar{A} = \frac{1}{15} \cdot 0,665 \cdot 100\% = 4,44\%.$$

Ошибка аппроксимации меньше 7%, что говорит о хорошем подборе модели линейной регрессии к исходным данным.

Построим теперь *поле корреляции* и *график полученного уравнения линейной регрессии*.

Для этого выделим столбцы X, Y,  $\hat{y}$ , выберем в меню Excel **Вставка – Точечная** (рисунок 11).

Из предлагаемых форматов графиков выбираем первый (состоящий из точек).

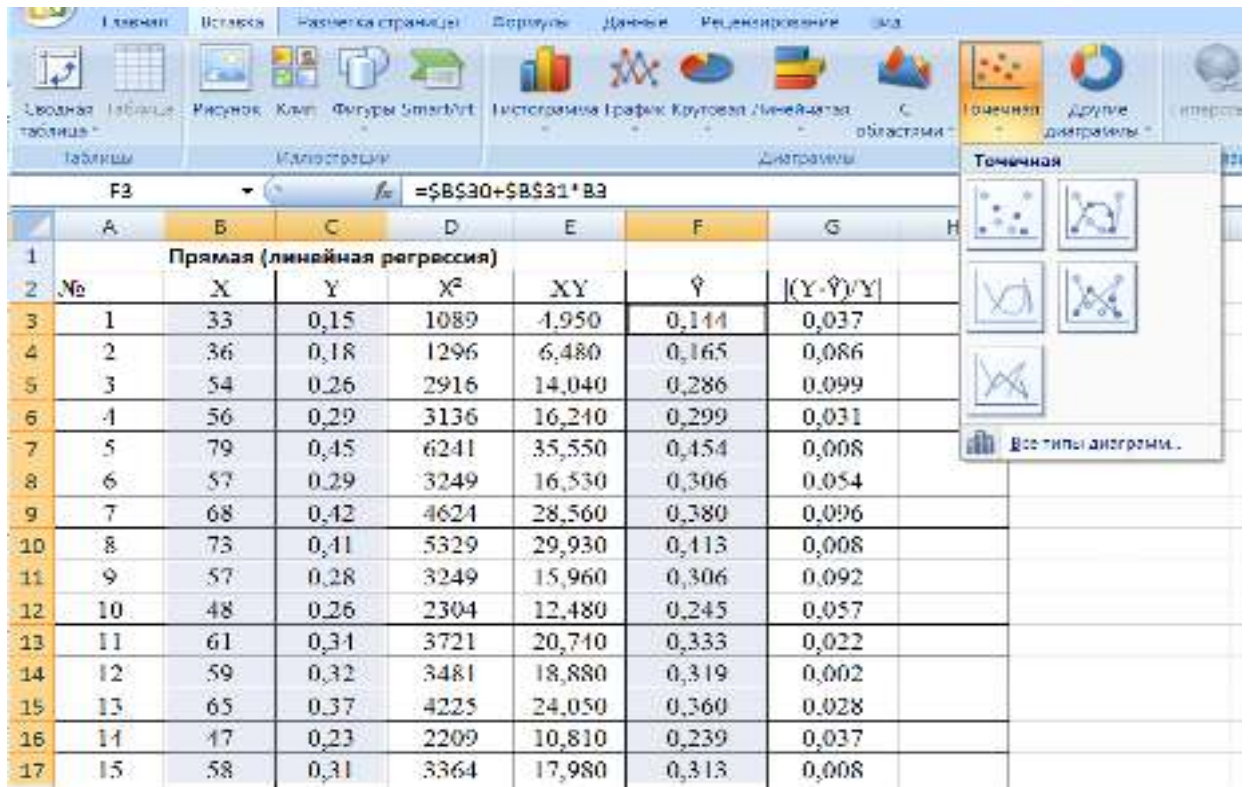


Рисунок 11 – Фрагмент построения графика в Excel

На экране появится график вида (рисунок 12):

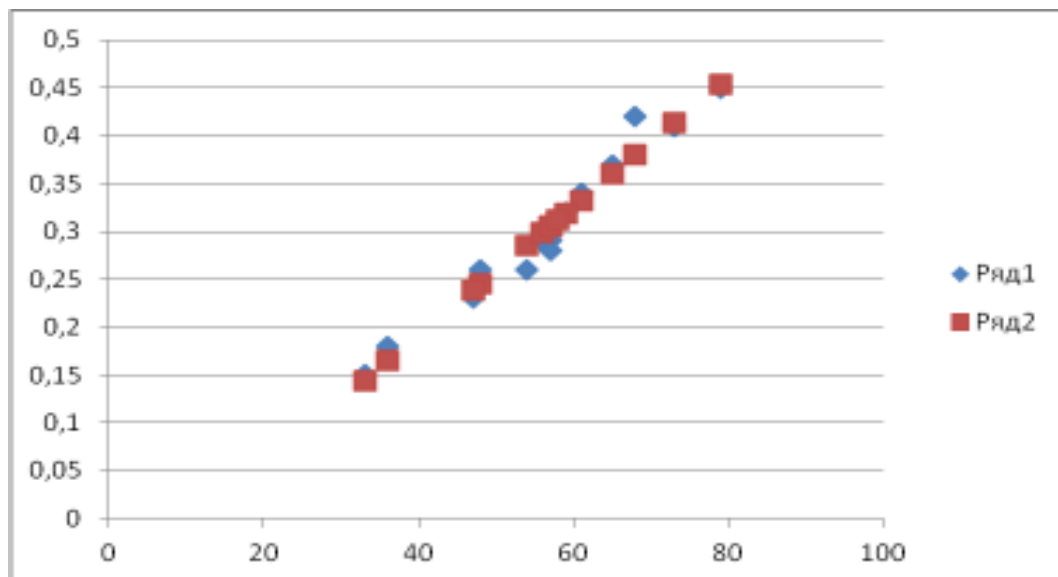


Рисунок 12 – Фрагмент построения графика в Excel

Здесь Ряд 1 – поле корреляции (его так и оставляем в виде точек), а Ряд 2 – график уравнения линейной регрессии (его необходимо преобразовать). Выбираем точку из Ряда 2, «щёлкаем» по ней правой кнопкой мыши.



В появившемся меню (рисунок 13) выбираем **«Изменить тип диаграммы»**:

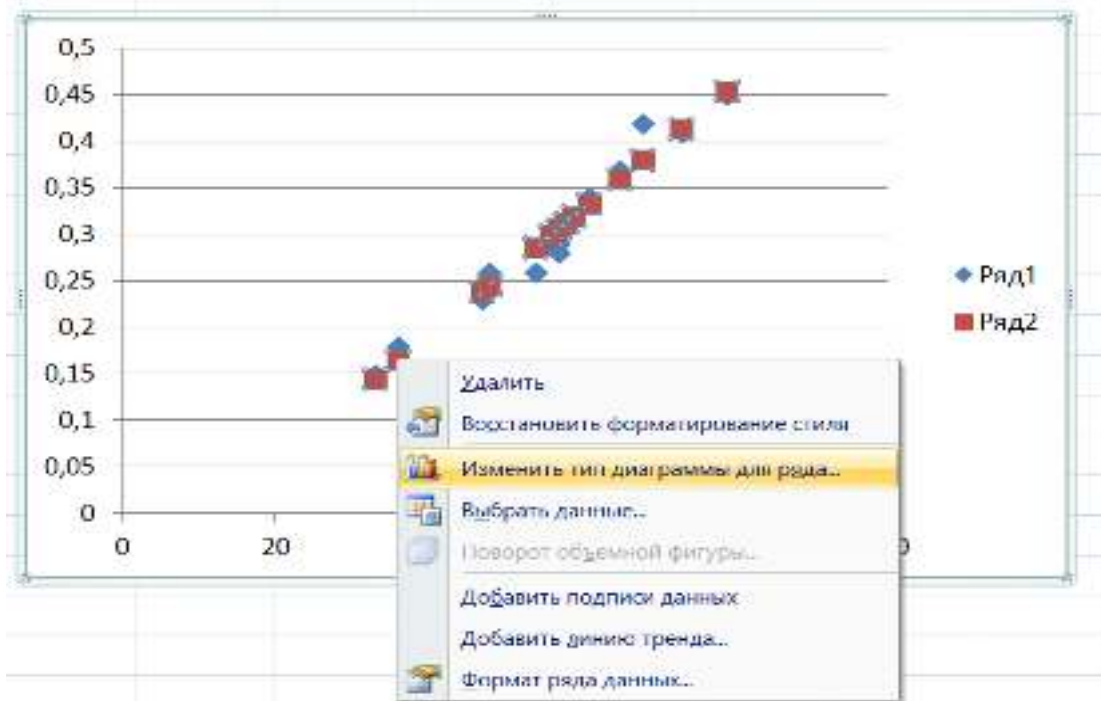


Рисунок 13 – Фрагмент построения графика в Excel

и из предлагаемых Точечных графиков выбираем третий (гладкая кривая без маркеров):

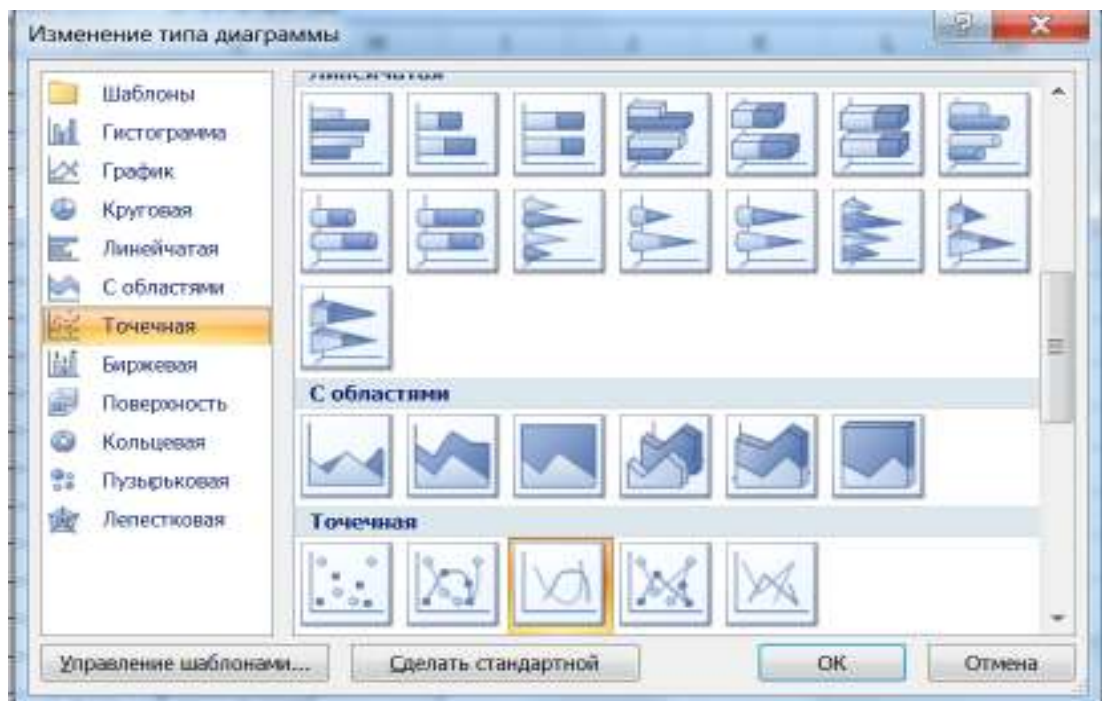


Рисунок 14 – Фрагмент построения графика в Excel

Окончательно график будет выглядеть следующим образом:

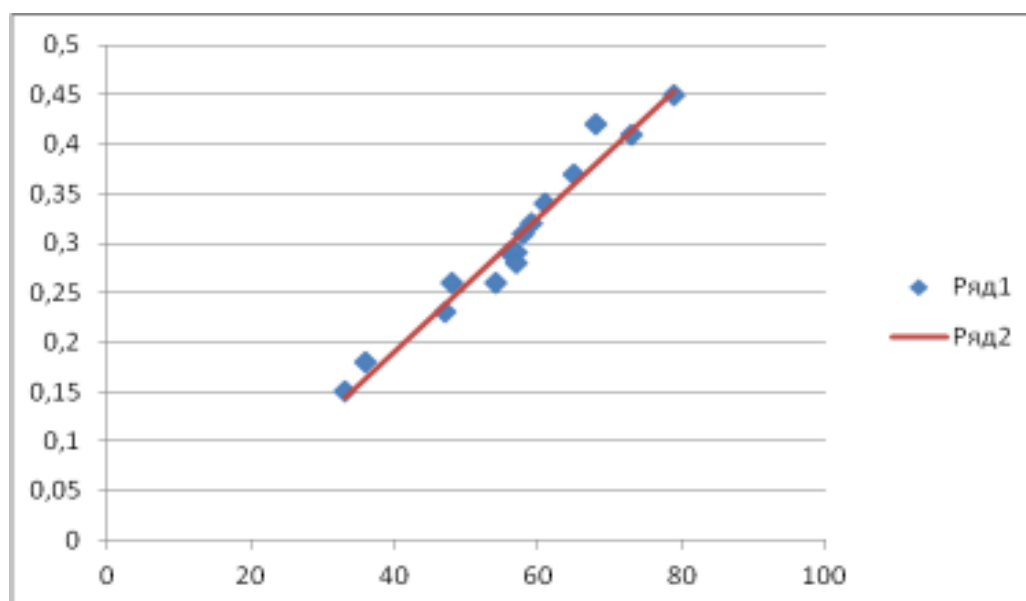


Рисунок 15 – Поле корреляции и график уравнения линейной регрессии

### 3. Оценка тесноты связи между параметрами

Современная наука исходит из взаимосвязи всех явлений природы и общества. Невозможно управлять явлениями, предсказывать их развитие без изучения характера, силы и других особенностей связей.

По силе связи между переменными различают *слабые* и *сильные* связи.

Задачи корреляционного анализа сводятся к изучению тесноты связи между варьирующими признаками, определению неизвестных причинных связей и оценке факторов, оказывающих наибольшее влияние на результативный признак.

Для измерения тесноты связи между факторами применяется несколько показателей.

#### *Линейный коэффициент корреляции*

Уравнение регрессии всегда дополняется показателем тесноты связи. При использовании линейной регрессии в качестве такого показателя высту-

пает *линейный коэффициент корреляции*. Иногда его называют *линейный коэффициент парной корреляции*.

Линейный коэффициент корреляции может быть вычислен по формуле

$$r_{xy} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x \cdot \sigma_y}, \quad (26)$$

где  $\bar{x}$  – среднее значение факторного признака  $x_i$ ,

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}; \quad (27)$$

$\bar{y}$  – среднее значение результативного признака  $y_i$ ,

$$\bar{y} = \frac{\sum_{i=1}^n y_i}{n}; \quad (28)$$

$\overline{xy}$  – среднее значение произведений  $x_i y_i$ ,

$$\overline{xy} = \frac{\sum_{i=1}^n x_i y_i}{n}; \quad (29)$$

$\sigma_x$  – среднее квадратическое отклонение факторного признака  $x_i$   
от общей средней  $\bar{x}$ ,

$$\sigma_x = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}} = \sqrt{\overline{x^2} - \bar{x}^2}; \quad (30)$$

$\sigma_y$  – среднее квадратическое отклонение результативного признака  $y_i$   
от общей средней  $\bar{y}$ ,

$$\sigma_y = \sqrt{\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n}} = \sqrt{\overline{y^2} - \bar{y}^2}. \quad (31)$$

Линейный коэффициент корреляции  $r_{xy}$  принимает значения в диапазоне

$$-1 \leq r_{xy} \leq 1.$$

Чем ближе величина  $|r_{xy}|$  к единице, тем теснее линейная связь и тем лучше линейная зависимость согласуется с данными наблюдений.

При  $r_{xy} > 0$  связь является *прямой*, при  $r_{xy} < 0$  – *обратной*.

При  $|r_{xy}| = 1$  связь становится функциональной.

Для оценки тесноты связи применяют шкалу Чеддока (таблица 6).

Таблица 6 – Шкала Чеддока

<i>Показатели</i>	<i>Характеристика</i>
0,1–0,3	слабая
0,3–0,5	умеренная
0,5–0,7	заметная
0,7–0,9	высокая
0,9–99	весьма высокая

Однако следует иметь в виду, что величина линейного коэффициента корреляции оценивает тесноту связи рассматриваемых признаков в ее *линейной форме*. Поэтому близость к нулю величины линейного коэффициента корреляции еще не означает *отсутствие связи* между признаками. При иной спецификации модели связь между признаками может оказаться достаточно тесной.

### ***Коэффициент детерминации***

Для оценки качества подбора линейной модели вычисляется квадрат линейного коэффициента корреляции  $r_{xy}^2$ , называемый *коэффициентом детерминации*.

Коэффициент детерминации характеризует долю дисперсии результативного признака  $y$ , объясняемую регрессией, в общей дисперсии результативного признака.

Иными словами, *коэффициент детерминации* показывает, какая часть общей вариации результативного признака  $y$  объясняется влиянием изучаемого фактора  $x$ .

Соответственно величина  $1 - r_{xy}^2$  характеризует долю дисперсии  $y$ , вызванную влиянием остальных не учтенных в модели факторов.

Например, если  $r_{xy}^2 = 0,96$ , то уравнением регрессии (изучаемым фактором  $x$ ) объясняется 96% дисперсии результативного признака  $y$ , а на долю прочих факторов (не включенных в уравнение регрессии) приходится 4% ее дисперсии.

### ***Оценка значимости линейного коэффициента корреляции***

Показатели тесноты связи, вычисленные по данным сравнительно небольшой статистической совокупности, могут искажаться под действием случайных причин. Это вызывает необходимость проверки их существенности.

Для оценки значимости линейного коэффициента корреляции  $r_{xy}$  применяется  $t$ -критерий Стьюдента, согласно которому выдвигается «нулевая» гипотеза  $H_0$  о статистической незначимости коэффициента линейной корреляции.

Эта гипотеза отвергается при выполнении условия

$$t_r > t_{\text{крит}},$$

где  $t_r$  – фактическое значение  $t$ -критерия Стьюдента, определяется как

$$t_r = r_{xy} \cdot \sqrt{\frac{n-2}{1-r_{xy}^2}}, \quad (32)$$

$t_{\text{крит}}$  – критическое (табличное) значение, определяется из таблицы

значений  $t$ -критерия Стьюдента (приложение 3) с учётом заданного уровня значимости  $\alpha$  и числа степеней свободы  $k = n - 2$ .

**Определение.** *Уровнем значимости* (обозначается  $\alpha$ ) в статистических гипотезах называется вероятность отвергнуть верную гипотезу (это так называемая ошибка первого рода). Уровень значимости  $\alpha$  обычно принимает значения 0,05 и 0,01, что соответствует вероятности совершения ошибки первого рода 5% и 1%.

Таким образом, вычисленное по формуле (32) значение  $t_r$  сравнивается с  $t_{\text{крит}}$ .

Если

$$t_r > t_{\text{крит}},$$

то величина линейного коэффициента корреляции признаётся существенной (*статистически значимой*).

Соответственно, если

$$t_r < t_{\text{крит}},$$

то подтверждается нулевая гипотеза  $H_0$  о *статистической незначимости* коэффициента линейной корреляции.

**Замечание.** Проверка гипотезы о значимости линейного коэффициента корреляции равносильна проверке гипотезы о значимости уравнения линейной регрессии.

#### 4. Пример вычисления линейного коэффициента корреляции

Используя данные рассмотренного в пункте 2 примера (таблица 2, стр. 16), вычислим линейный коэффициент корреляции  $r_{xy}$ , коэффициент детерминации  $r_{xy}^2$ . Оценим значимости линейного коэффициента корреляции  $r_{xy}$  с помощью  $t$ -критерия Стьюдента.

Чтобы вычислить линейный коэффициент корреляции  $r_{xy}$ , дополним таблицу 5 (стр. 22) колонкой  $Y^2$  и строкой средних значений, получим таблицу 7.

Средние значения, с учетом формул (27)–(29), определяются как сумма, деленная на  $n = 20$ .

Вычислим вначале по формуле (30) среднее квадратическое отклонение факторного признака  $x_i$  от общей средней  $\bar{x}$

$$\sigma_x = \sqrt{x^2 - \frac{\sum x^2}{n}} = \sqrt{3362,2 - 56,733^2} = 11,980$$

и по формуле (31) среднее квадратическое отклонение результативного признака  $y_i$  от общей средней  $\bar{y}$

$$\sigma_y = \sqrt{y^2 - \frac{\sum y^2}{n}} = \sqrt{0,099 - 0,304^2} = 0,082.$$

Таблица 7 – Данные для определения линейного коэффициента корреляции

№	$X = V$	$Y = Q$	$X^2$	$XY$	$\hat{y}$	$\left  \frac{y_i - \hat{y}_i}{y_i} \right $	$Y^2$
1	33	0,15	1089	4,950	0,144	0,037	0,023
2	36	0,18	1296	6,480	0,165	0,086	0,032
3	54	0,26	2916	14,040	0,286	0,099	0,068
4	56	0,29	3136	16,240	0,299	0,031	0,084
5	79	0,45	6241	35,550	0,454	0,008	0,203
6	57	0,29	3249	16,530	0,306	0,054	0,084
7	68	0,42	4624	28,560	0,380	0,096	0,176
8	73	0,41	5329	29,930	0,413	0,008	0,168
9	57	0,28	3249	15,960	0,306	0,092	0,078
10	48	0,26	2304	12,480	0,245	0,057	0,068
11	61	0,34	3721	20,740	0,333	0,022	0,116
12	59	0,32	3481	18,880	0,319	0,002	0,102
13	65	0,37	4225	24,050	0,360	0,028	0,137
14	47	0,23	2209	10,810	0,239	0,037	0,053
15	58	0,31	3364	17,980	0,313	0,008	0,096
<b>Сумма</b> $\Sigma$	851,000	<b>4,560</b>	50433	273,180	<b>4,560</b>	0,665	2,203
<b>Среднее</b> $\frac{\Sigma}{n}$	56,733	0,304	3362,2	18,212	0,304	0,044	0,099

Коэффициент корреляции  $r_{xy}$  вычислим по формуле (26):

$$r_{xy} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x \cdot \sigma_y} = \frac{18,212 - 56,733 \cdot 0,304}{11,980 \cdot 0,082} = 0,98.$$

По шкале Чеддока (приложение 2) связь между переменными весьма высокая.

Так как  $r_{xy} > 0$ , то связь между фактором  $x$  и результатом  $y$  *прямая*. То есть, с увеличением скорости автомобиля  $V$  увеличивается расход топлива  $Q$ .

Коэффициент детерминации равен:

$$r_{xy}^2 = 0,98^2 = 0,96,$$

то есть 96% вариации результативного признака  $y$  (расхода топлива) объясняется включенным в уравнение линейной регрессии (25) фактором  $x$  (изменением скорости автомобиля), а 4% приходится на долю факторов, не включенных в уравнение регрессии.

Оценим значимости линейного коэффициента корреляции  $r_{xy}$  с помощью  $t$ -критерия Стьюдента. Для этого по формуле (32) найдем фактическое значение  $t$ -критерия Стьюдента:

$$t_r = r_{xy} \cdot \sqrt{\frac{n-2}{1-r_{xy}^2}} = 0,98 \cdot \sqrt{\frac{15-2}{1-0,96}} = 17,73.$$

Определим критическое значение  $t_{крит}$  по таблице значений  $t$ -критерия Стьюдента (приложение 3) с учётом уровня значимости  $\alpha=0,05$  и числа степеней свободы  $k = n - 2 = 15 - 2 = 13$

$$t_{крит} = 2,16.$$

Так как  $17,73 > 2,16$  ( $t_r > t_{крит}$ ), то величина линейного коэффициента корреляции признаётся существенной. Следовательно, и уравнение линейной регрессии (25) статистически значимо.



### 5. Задания для самостоятельной работы

1. Получить у преподавателя индивидуальные задания (экспериментальные данные  $x, y$ ) – приложение 1.
2. Руководствуясь рассмотренным примером и в соответствии со своим заданием заполнить таблицу (у каждого студента свой вариант):

№	X	Y	XY	X <sup>2</sup>	$\hat{y}$	$\left  \frac{y_i - \hat{y}_i}{y_i} \right $
1						
2						
3						
...	...	...	...	...	...	...
15						
$\Sigma$						

3. Вычислить  $a_0$  и  $a_1$ , составить уравнение линейной регрессии  $\hat{y} = a_0 + a_1x$ .
4. Вычислить среднюю ошибку аппроксимации и сделать вывод о качестве уравнения линейной регрессии.
5. Построить поле корреляции и график полученного уравнения линейной регрессии.
6. Вычислить линейный коэффициент корреляции, оценить тесноту связи по шкале Чеддока.
7. Вычислить коэффициент детерминации, интерпретировать полученный результат.
8. Оценить значимость линейного коэффициента корреляции и уравнения линейной регрессии с помощью критерия Стьюдента.
9. Составить отчет по лабораторной работе.

### 6. Содержание отчета

Отчет должен содержать:

1. Определения и формулы, изложенные в теоретической части лабораторной работы.
2. Исходные данные и результаты выполняемых заданий.

### Контрольные вопросы

1. Дайте определение поля корреляции.
2. Запишите уравнение парной линейной регрессии.
3. Объясните суть метода наименьших квадратов.
4. Приведите систему для нахождения параметров линейной регрессии.
5. Приведите формулы Крамера и формулы определителей для вычисления параметров линейной регрессии.
6. Приведите формулу средней ошибки аппроксимации и поясните смысл этой величины.
7. Объясните, как в Excel построить графики.
8. Дайте определение линейного коэффициента корреляции.
9. Какие значения может принимать линейный коэффициент корреляции?
10. Для чего применяют шкалу Чеддока?
11. Дайте определение коэффициента детерминации.
12. Как оценить значимость линейного коэффициента корреляции и уравнения линейной регрессии?
13. Определите понятие «уровень значимости».
14. Определите понятие «число степеней свободы» в критерии Стьюдента.

## ЛАБОРАТОРНО-ПРАКТИЧЕСКИЕ ЗАНЯТИЯ № 3, 4

**Тема:** Уравнения парной нелинейной регрессии. Параболическая, гиперболическая регрессия.

**Цель работы:** Познакомиться с уравнениями парной нелинейной регрессии – уравнениями параболической, гиперболической регрессии, научиться находить параметры уравнений параболической, гиперболической регрессии с использованием Excel и соответствующие им средние ошибки аппроксимации.

### 1. Построение уравнения параболической регрессии

Если между явлениями существуют нелинейные соотношения, то они выражаются с помощью соответствующих нелинейных функций: параболы, гиперболы, логарифмической, степенной, показательной, экспоненциальной функций.

Если при равномерном возрастании  $x$  значения  $y$  возрастают или убывают ускоренно, то чаще всего в этом случае зависимость между коррелируемыми величинами может быть выражена в виде параболы

$$\hat{y} = a_0 + a_1 \cdot x + a_2 \cdot x^2, \quad (33)$$

параметры которой находят путем решения системы нормальных уравнений:

$$\begin{cases} a_0 n + a_1 \sum_{i=1}^n x_i + a_2 \sum_{i=1}^n x_i^2 = \sum_{i=1}^n y_i \\ a_0 \sum_{i=1}^n x_i + a_1 \sum_{i=1}^n x_i^2 + a_2 \sum_{i=1}^n x_i^3 = \sum_{i=1}^n x_i y_i \\ a_0 \sum_{i=1}^n x_i^2 + a_1 \sum_{i=1}^n x_i^3 + a_2 \sum_{i=1}^n x_i^4 = \sum_{i=1}^n x_i^2 y_i \end{cases} \quad (34)$$

Решить систему (34) относительно параметров  $a_0, a_1, a_2$  можно методом Крамера:

$$a_0 = \frac{\Delta a_0}{\Delta}, \quad a_1 = \frac{\Delta a_1}{\Delta}, \quad a_2 = \frac{\Delta a_2}{\Delta}, \quad (35)$$

где  $\Delta$  – определитель основной матрицы системы уравнений (34):

$$\Delta = \begin{vmatrix} n & \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i^3 \\ \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i^3 & \sum_{i=1}^n x_i^4 \end{vmatrix}, \quad (36)$$

а  $\Delta a_0$ ,  $\Delta a_1$ ,  $\Delta a_2$  – определители, получаемые из  $\Delta$  заменой 1-го, 2-го и 3-го столбца соответственно на столбец свободных членов (правых частей уравнений системы (34)):

$$\Delta a_0 = \begin{vmatrix} \sum_{i=1}^n y_i & \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \\ \sum_{i=1}^n x_i y_i & \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i^3 \\ \sum_{i=1}^n x_i^2 y_i & \sum_{i=1}^n x_i^3 & \sum_{i=1}^n x_i^4 \end{vmatrix}, \quad (37)$$

$$\Delta a_1 = \begin{vmatrix} n & \sum_{i=1}^n y_i & \sum_{i=1}^n x_i^2 \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i y_i & \sum_{i=1}^n x_i^3 \\ \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i^2 y_i & \sum_{i=1}^n x_i^4 \end{vmatrix}, \quad (38)$$

$$\Delta a_2 = \begin{vmatrix} n & \sum_{i=1}^n x_i & \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i y_i \\ \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i^3 & \sum_{i=1}^n x_i^2 y_i \end{vmatrix}. \quad (39)$$

Подставляя найденные величины  $a_0$ ,  $a_1$ ,  $a_2$  в уравнение (33), получим искомое уравнение параболической регрессии.

## 2. Пример построения уравнения параболической регрессии

Имеются данные об изменении расхода топлива  $Q$  (л/км) автомобиля при изменении скорости его движения  $V$  (км/ч) (таблица 2, лабораторная работа 1, стр. 17).

Необходимо получить уравнение параболической регрессии, отражающее зависимость расхода топлива от скорости автомобиля  $Q(V)$ . Оценить качество модели с помощью средней ошибки аппроксимации. Построить поле корреляции и график полученной параболической зависимости  $\hat{y}(x)$ .

Для определения коэффициентов параболической регрессии и средней ошибки аппроксимации в Excel необходимо рассчитать таблицу 8 следующего вида:

Таблица 8 – Данные для определения коэффициентов параболической регрессии и средней ошибки аппроксимации

$N^{\circ}$	$X = V$	$Y = Q$	$X^2$	$X^3$	$X^4$	$XY$	$X^2Y$	$\hat{y}$	$\left  \frac{y_i - \hat{y}_i}{y_i} \right $
1	33	0,15	1089	35937	1185921	4,95	163,35		
2	36	0,18	1296	46656	1679616	6,48	233,28		
3	54	0,26	2916	157464	8503056	14,04	758,16		
4	56	0,29	3136	175616	9834496	16,24	909,44		
5	79	0,45	6241	493039	38950081	35,55	2808,45		
6	57	0,29	3249	185193	10556001	16,53	942,21		
7	68	0,42	4624	314432	21381376	28,56	1942,08		
8	73	0,41	5329	389017	28398241	29,93	2184,89		
9	57	0,28	3249	185193	10556001	15,96	909,72		
10	48	0,26	2304	110592	5308416	12,48	599,04		
11	61	0,34	3721	226981	13845841	20,74	1265,14		
12	59	0,32	3481	205379	12117361	18,88	1113,92		
13	65	0,37	4225	274625	17850625	24,05	1563,25		
14	47	0,23	2209	103823	4879681	10,81	508,07		
15	58	0,31	3364	195112	11316496	17,98	1042,84		
$\Sigma$	851	4,56	50433	3099059	196363209	273,18	16943,84		

Две последние колонки таблицы заполним после того, как получим уравнение параболической регрессии.

Вычислим определители (36)–(39):

$$\Delta = \begin{vmatrix} n & \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i^3 \\ \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i^3 & \sum_{i=1}^n x_i^4 \end{vmatrix} = \begin{vmatrix} 15 & 851 & 50433 \\ 851 & 50433 & 3099059 \\ 50433 & 3099059 & 196363209 \end{vmatrix} = 16970671488,$$

$$\Delta a_0 = \begin{vmatrix} \sum_{i=1}^n y_i & \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \\ \sum_{i=1}^n x_i y_i & \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i^3 \\ \sum_{i=1}^n x_i^2 y_i & \sum_{i=1}^n x_i^3 & \sum_{i=1}^n x_i^4 \end{vmatrix} = \begin{vmatrix} 4,56 & 851 & 50433 \\ 273,18 & 50433 & 3099059 \\ 16943,84 & 3099059 & 196363209 \end{vmatrix} = -86064954,4,$$

$$\Delta a_1 = \begin{vmatrix} n & \sum_{i=1}^n y_i & \sum_{i=1}^n x_i^2 \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i y_i & \sum_{i=1}^n x_i^3 \\ \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i^2 y_i & \sum_{i=1}^n x_i^4 \end{vmatrix} = \begin{vmatrix} 15 & 4,56 & 50433 \\ 851 & 273,18 & 3099059 \\ 50433 & 16943,84 & 196363209 \end{vmatrix} = 67388267,88,$$

$$\Delta a_2 = \begin{vmatrix} n & \sum_{i=1}^n x_i & \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i y_i \\ \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i^3 & \sum_{i=1}^n x_i^2 y_i \end{vmatrix} = \begin{vmatrix} 15 & 851 & 4,56 \\ 851 & 50433 & 273,18 \\ 50433 & 3099059 & 16943,84 \end{vmatrix} = 422933,8.$$

**Замечание.** Для вычисления определителей в Excel необходимо использовать функцию **МОПРЕД**.

Вычислим неизвестные коэффициенты уравнения параболической регрессии по формулам (35):

$$a_0 = \frac{\Delta a_0}{\Delta} = \frac{-86064954,4}{16970671488} = -0,0051;$$

$$a_1 = \frac{\Delta a_1}{\Delta} = \frac{67388267,88}{16970671488} = 0,004,$$

$$a_2 = \frac{\Delta a_2}{\Delta} = \frac{422933,8}{16970671488} = 0,000025.$$

На рисунке 16 изображен фрагмент листа Excel при вычислении параметров уравнения параболической регрессии.

	A	B	C	D	E	F	G	H	I	J
1	№	x	y	x <sup>2</sup>	x <sup>3</sup>	x <sup>4</sup>	xy	x <sup>2</sup> y	$\hat{y}$	$(y-\hat{y})^2$
2	1	33	0,15	1089	35917	1185921	4,95	163,35		
3	2	36	0,18	1296	46656	1679616	6,48	233,28		
4	3	54	0,26	2916	157464	8503056	14,04	758,16		
5	4	56	0,29	3136	175616	9834496	16,24	909,44		
6	5	79	0,45	6241	493039	38950081	35,55	2808,45		
7	6	57	0,29	3249	185193	10556001	16,53	942,21		
8	7	68	0,42	4624	314432	21381376	28,56	1942,08		
9	8	73	0,41	5329	389017	28398241	29,93	2184,89		
10	9	57	0,28	3249	185193	10556001	15,96	909,72		
11	10	48	0,26	2304	110592	5308416	12,48	599,04		
12	11	61	0,34	3721	226981	13845841	20,74	1265,14		
13	12	59	0,32	3481	205379	12117361	18,88	1113,92		
14	13	65	0,37	4225	274625	17850625	24,05	1563,25		
15	14	47	0,23	2209	103823	4879681	10,81	508,07		
16	15	58	0,31	3364	195112	11316496	17,98	1042,84		
17	Σ	851	4,56	50433	3099059	196363209	273,18	16943,84		
18	среднее	56,733	0,304	3362,2	206603,933	13090880,6	18,212	1129,589		
19										
20	Δ=	15	851	50433						
21		851	50433	3099059	=	16970671488		a0=	-0,0051	
22		50433	3099059	196363209						
23										
24	Δa0=	4,56	851	50433						
25		273,18	50433	3099059	=	-86064954,4		a1=	0,0040	
26		16943,84	3099059	196363209						
27										
28	Δa1=	15	4,56	50433				a2=	0,000025	
29		851	273,18	3099059	=	67388267,88				
30		50433	16943,84	196363209						
31										
32	Δa2=	15	851	4,56						
33		851	50433	273,18	=	422933,8				
34		50433	3099059	16943,84						

Рисунок 16 – Фрагмент листа Excel

Полученные значения  $a_0$ ,  $a_1$  и  $a_2$  подставим в уравнение параболической регрессии.

Найденное уравнение параболической регрессии будет иметь вид:

$$\hat{y} = -0,0051 + 0,004 \cdot x + 0,000025 \cdot x^2. \quad (40)$$

Чтобы проверить качество полученного уравнения параболической регрессии, заполним две последние колонки таблицы 6, подставив в  $\hat{y}$  значения  $a_0$ ,  $a_1$  и  $a_2$ .

**Замечание.** При подстановке в таблицу значений  $a_0$ ,  $a_1$  и  $a_2$  закрепить в ссылках адреса клеток, где находятся значения  $a_0$ ,  $a_1$  и  $a_2$ , используя символ \$ (рисунок 17):

	A	B	C	D	E	F	G	H	I	J
	№	x	y	x <sup>2</sup>	x <sup>3</sup>	x <sup>4</sup>	xy	x <sup>2</sup> y	$\hat{y}$	(y- $\hat{y}$ )/y
2	1	33	0,15	1089	35937	1185921	4,95	163,35	0,153	
3	2	36	0,18	1296	46656	1679616	6,48	233,28		
4	3	54	0,26	2916	157464	8503056	14,04	758,16		
5	4	56	0,29	3136	175616	9834496	16,24	909,44		
6	5	79	0,45	6241	493039	38950081	35,55	2808,45		
7	6	57	0,29	3249	185193	10556001	16,53	942,21		
8	7	68	0,42	4624	314432	21381376	28,56	1942,08		
9	8	73	0,41	5329	389017	28398241	29,93	2184,89		
10	9	57	0,28	3249	185193	10556001	15,96	909,72		
11	10	48	0,26	2304	110592	5308416	12,48	599,04		
12	11	61	0,34	3721	226981	13845841	20,74	1265,14		
13	12	59	0,32	3481	205379	12117361	18,88	1113,92		
14	13	65	0,37	4225	274625	17850625	24,05	1563,25		
15	14	47	0,23	2209	103823	4879681	10,81	508,07		
16	15	58	0,31	3364	195112	11316496	17,98	1042,84		
17	Σ	851	4,56	50433	3099059	196363209	273,18	16943,84		
18	среднее	56,733	0,304	3362,2	206603,933	13090880,6	18,212	1129,589		

Рисунок 17 – Фрагмент листа Excel

Вычислите оставшиеся значения  $\hat{y}$  (таблица 9). Найдите сумму  $\hat{y}$  и убедитесь, что

$$\Sigma y = \Sigma \hat{y}.$$



Для оценки качества модели найдем среднюю ошибку аппроксимации по формуле (24):

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

Для этого заполним последнюю колонку таблицы 9, вычислив  $\left| \frac{y_i - \hat{y}_i}{y_i} \right|$ .

**Замечание.** Чтобы вычислить *модуль*, воспользуйтесь функцией **ABS** (найти её можно среди Математических функций, выбрав команду **Формулы – Вставить функцию** в меню Excel)

Таблица 9 – Вычисление значений  $\hat{y}$  и  $\left| \frac{y_i - \hat{y}_i}{y_i} \right|$

№	$X = V$	$Y = Q$	$X^2$	$X^3$	$X^4$	$XY$	$X^2Y$	$\hat{y}$	$\left  \frac{y_i - \hat{y}_i}{y_i} \right $
1	33	0,15	1089	35937	1185921	4,95	163,35	0,153	0,021
2	36	0,18	1296	46656	1679616	6,48	233,28	0,170	0,055
3	54	0,26	2916	157464	8503056	14,04	758,16	0,282	0,085
4	56	0,29	3136	175616	9834496	16,24	909,44	0,295	0,019
5	79	0,45	6241	493039	38950081	35,55	2808,45	0,464	0,031
6	57	0,29	3249	185193	10556001	16,53	942,21	0,302	0,042
7	68	0,42	4624	314432	21381376	28,56	1942,08	0,380	0,095
8	73	0,41	5329	389017	28398241	29,93	2184,89	0,418	0,019
9	57	0,28	3249	185193	10556001	15,96	909,72	0,302	0,079
10	48	0,26	2304	110592	5308416	12,48	599,04	0,243	0,066
11	61	0,34	3721	226981	13845841	20,74	1265,14	0,330	0,030
12	59	0,32	3481	205379	12117361	18,88	1113,92	0,316	0,013
13	65	0,37	4225	274625	17850625	24,05	1563,25	0,358	0,032
14	47	0,23	2209	103823	4879681	10,81	508,07	0,237	0,029
15	58	0,31	3364	195112	11316496	17,98	1042,84	0,309	0,003
$\Sigma$	851	<b>4,56</b>	50433	3099059	196363209	273,18	16943,84	<b>4,56</b>	0,616

Тогда средняя ошибка аппроксимации равна:

$$\bar{A} = \frac{1}{15} \cdot 0,616 \cdot 100 \% = 4,11\%.$$

Ошибка аппроксимации меньше 7%, что говорит о хорошем подборе модели параболической регрессии к исходным данным.

Построим теперь *поле корреляции* и *график полученного уравнения параболической регрессии*.

Для этого выделим столбцы X, Y,  $\hat{y}$ , выберем в меню Excel **Вставка – Точечная**. Из предлагаемых форматов графиков выбираем первый (состоящий из точек).

На экране появится график вида

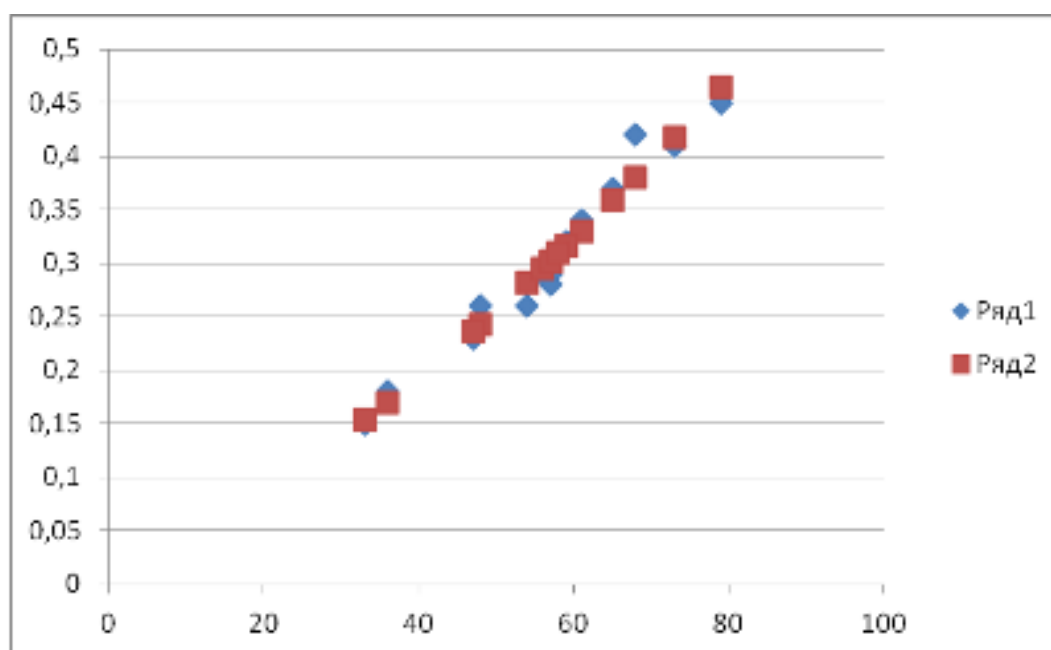


Рисунок 18 – Фрагмент построения графика в Excel

Здесь Ряд 1 – поле корреляции (его так и оставляем в виде точек), а Ряд 2 – график уравнения параболической регрессии (его необходимо преобразовать). Выбираем точку из Ряда 2, «щёлкаем» по ней правой кнопкой мыши.

В появившемся меню выбираем «Изменить тип диаграммы» и из предлагаемых Точечных графиков выбираем третий (гладкая кривая без маркеров).

Окончательно график будет выглядеть следующим образом:

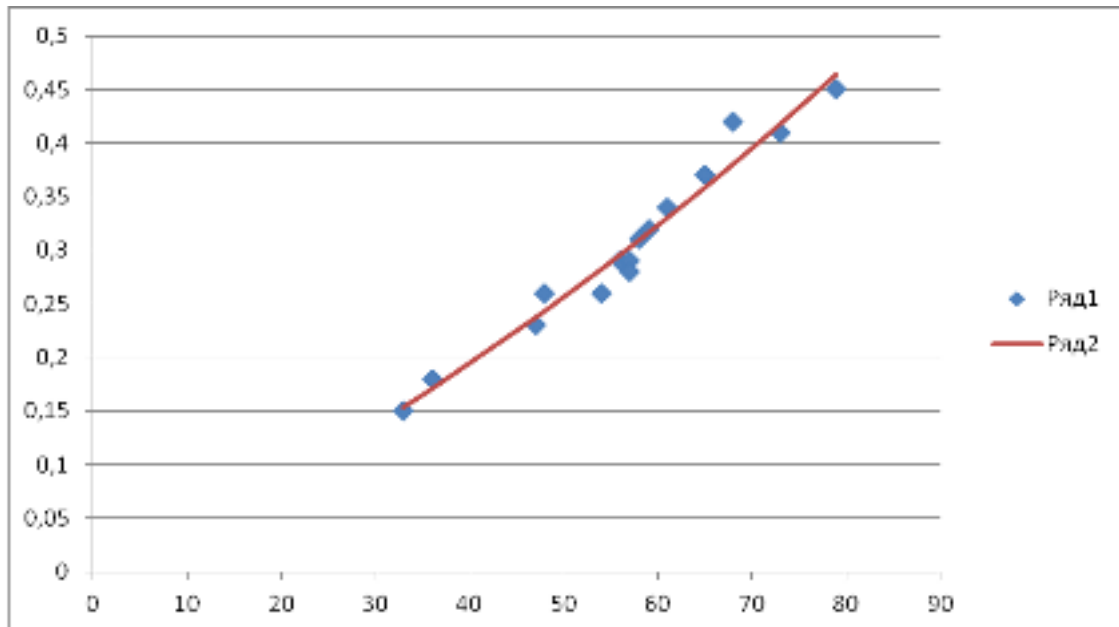


Рисунок 19 – Поле корреляции и график уравнения параболы регрессии

### 3. Построение уравнения гиперболической регрессии

Обратная зависимость между двумя величинами (когда с увеличением  $x$  уменьшается  $y$ ) может выражаться уравнением гиперболы

$$\hat{y} = a_0 + \frac{a_1}{x}, \quad (41)$$

параметры которого находят, решая систему уравнений

$$\begin{cases} a_0 n + a_1 \sum_{i=1}^n \frac{1}{x_i} = \sum_{i=1}^n y_i \\ a_0 \sum_{i=1}^n \frac{1}{x_i} + a_1 \sum_{i=1}^n \frac{1}{x_i^2} = \sum_{i=1}^n \frac{y_i}{x_i} \end{cases}. \quad (42)$$

Для решение системы (42) относительно неизвестных  $a_0, a_1$  также воспользуемся методом Крамера:

$$a_0 = \frac{\Delta a_0}{\Delta}, \quad a_1 = \frac{\Delta a_1}{\Delta}, \quad (43)$$

где  $\Delta$  – определитель основной матрицы системы уравнений (42)

$$\Delta = \begin{vmatrix} n & \sum_{i=1}^n \frac{1}{x_i} \\ \sum_{i=1}^n \frac{1}{x_i} & \sum_{i=1}^n \frac{1}{x_i^2} \end{vmatrix}, \quad (44)$$

а  $\Delta a_0$  и  $\Delta a_1$  – определители, получаемые из  $\Delta$  заменой 1-го и 2-го столбца соответственно на столбец свободных членов (правых частей уравнений системы (44)), то есть

$$\Delta a_0 = \begin{vmatrix} \sum_{i=1}^n y_i & \sum_{i=1}^n \frac{1}{x_i} \\ \sum_{i=1}^n \frac{y_i}{x_i} & \sum_{i=1}^n \frac{1}{x_i^2} \end{vmatrix}, \quad (45)$$

$$\Delta a_1 = \begin{vmatrix} n & \sum_{i=1}^n y_i \\ \sum_{i=1}^n \frac{1}{x_i} & \sum_{i=1}^n \frac{y_i}{x_i} \end{vmatrix}. \quad (46)$$

Подставляя найденные величины  $a_0, a_1$  в уравнение (41), получим иско-  
мое уравнение гиперболической регрессии.

#### 4. Пример построения уравнения гиперболической регрессии

Имеются данные об изменении расхода топлива  $Q$  (л/км) автомобиля при изменении скорости его движения  $V$  (км/ч) (лабораторная работа 1, стр. 17, таблица 2).

Необходимо получить уравнение гиперболической регрессии, отражающее зависимость расхода топлива от скорости автомобиля  $Q(V)$ . Оценить качество модели с помощью средней ошибки аппроксимации. Построить поле корреляции и график полученной гиперболической зависимости  $\hat{y}(x)$ .

Для определения коэффициентов гиперболической регрессии и средней ошибки аппроксимации в Excel необходимо рассчитать таблицу 10 следующего вида:

Таблица 10 – Данные для определения коэффициентов гиперболической регрессии и средней ошибки аппроксимации

$N_{\text{д}}$	$X = V$	$Y = Q$	$\frac{1}{X}$	$\frac{1}{X^2}$	$\frac{Y}{X}$	$\hat{y}$	$\left  \frac{y_i - \hat{y}_i}{y_i} \right $
1	33	0,15	0,0303	0,0009	0,0045		
2	36	0,18	0,0278	0,0008	0,005		
3	47	0,23	0,0213	0,0005	0,0049		
4	48	0,26	0,0208	0,0004	0,0054		
5	54	0,26	0,0185	0,0003	0,0048		
6	56	0,29	0,0179	0,0003	0,0052		
7	57	0,29	0,0175	0,0003	0,0051		
8	57	0,28	0,0175	0,0003	0,0049		
9	58	0,31	0,0172	0,0003	0,0053		
10	59	0,32	0,0169	0,0003	0,0054		
11	61	0,34	0,0164	0,0003	0,0056		
12	65	0,37	0,0154	0,0002	0,0057		
13	68	0,42	0,0147	0,0002	0,0062		
14	73	0,41	0,0137	0,0002	0,0056		
15	79	0,45	0,0127	0,0002	0,0057		
$\Sigma$	851	4,56	0,2787	0,0055	0,0794		

Две последние колонки таблицы заполним после того, как получим уравнение параболической регрессии.

Вычислим определители (44)–(46):

$$\Delta = \begin{vmatrix} n & \sum_{i=1}^n \frac{1}{x_i} \\ \sum_{i=1}^n \frac{1}{x_i} & \sum_{i=1}^n \frac{1}{x_i^2} \end{vmatrix} = \begin{vmatrix} 15 & 0,279 \\ 0,279 & 0,0055 \end{vmatrix} = 0,00496,$$

$$\Delta a_0 = \begin{vmatrix} \sum_{i=1}^n y_i & \sum_{i=1}^n \frac{1}{x_i} \\ \sum_{i=1}^n \frac{y_i}{x_i} & \sum_{i=1}^n \frac{1}{x_i^2} \end{vmatrix} = \begin{vmatrix} 4,56 & 0,279 \\ 0,0794 & 0,0055 \end{vmatrix} = 0,003,$$

$$\Delta a_1 = \begin{vmatrix} n & \sum_{i=1}^n y_i \\ \sum_{i=1}^n \frac{1}{x_i} & \sum_{i=1}^n \frac{y_i}{x_i} \end{vmatrix} = \begin{vmatrix} 15 & 0,279 \\ 0,279 & 0,0055 \end{vmatrix} = -0,080.$$

**Замечание.** Для вычисления определителей в Excel необходимо использовать функцию **МОПРЕД**.

Вычислим неизвестные коэффициенты уравнения гиперболической регрессии по формулам (43):

$$a_0 = \frac{\Delta a_0}{\Delta} = \frac{0,003}{0,00496} = 0,605;$$

$$a_1 = \frac{\Delta a_1}{\Delta} = \frac{-0,080}{0,00496} = -16,185.$$

Полученные значения  $a_0$  и  $a_1$  подставим в уравнение гиперболической регрессии.

Найденное уравнение гиперболической регрессии будет иметь вид:

$$\hat{y} = 0,605 - \frac{16,185}{x}. \quad (47)$$

Чтобы проверить качество полученного уравнения гиперболической регрессии, заполним две последние колонки таблицы 10, подставив в  $\hat{y}$  значения  $a_0$  и  $a_1$ .

**Замечание.** При подстановке в таблицу значений  $a_0$  и  $a_1$  закрепить в ссылках адреса клеток, где находятся значения  $a_0$ ,  $a_1$ , используя символ \$.

Вычислите оставшиеся значения  $\hat{y}$  (таблица 11). Найдите сумму  $\hat{y}$  и убедитесь, что

$$\Sigma y = \Sigma \hat{y}.$$

Для оценки качества модели найдем среднюю ошибку аппроксимации по формуле (24):

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

Для этого заполним последнюю колонку таблицы, вычислив  $\left| \frac{y_i - \hat{y}_i}{y_i} \right|$ .

Таблица 11 – Вычисление значений  $\hat{y}$  и  $\left| \frac{y_i - \hat{y}_i}{y_i} \right|$

№	$X = V$	$Y = Q$	$\frac{1}{X}$	$\frac{1}{X^2}$	$\frac{Y}{X}$	$\hat{y}$	$\left  \frac{y_i - \hat{y}_i}{y_i} \right $
1	33	0,15	0,0303	0,0009	0,0045	0,11	0,24
2	36	0,18	0,0278	0,0008	0,005	0,16	0,14
3	47	0,23	0,0213	0,0005	0,0049	0,26	0,13
4	48	0,26	0,0208	0,0004	0,0054	0,27	0,03
5	54	0,26	0,0185	0,0003	0,0048	0,30	0,17
6	56	0,29	0,0179	0,0003	0,0052	0,32	0,09
7	57	0,29	0,0175	0,0003	0,0051	0,32	0,11
8	57	0,28	0,0175	0,0003	0,0049	0,32	0,15
9	58	0,31	0,0172	0,0003	0,0053	0,33	0,05
10	59	0,32	0,0169	0,0003	0,0054	0,33	0,03
11	61	0,34	0,0164	0,0003	0,0056	0,34	0,00
12	65	0,37	0,0154	0,0002	0,0057	0,36	0,04
13	68	0,42	0,0147	0,0002	0,0062	0,37	0,13
14	73	0,41	0,0137	0,0002	0,0056	0,38	0,07
15	79	0,45	0,0127	0,0002	0,0057	0,40	0,11
$\Sigma$	851	4,56	0,2787	0,0055	0,0794	4,56	1,48

**Замечание.** Чтобы вычислить *модуль*, воспользуйтесь функцией **ABS** (найти её можно среди Математических функций, выбрав команду **Формулы – Вставить функцию** в меню Excel).

Тогда средняя ошибка аппроксимации равна:

$$\bar{A} = \frac{1}{15} \cdot 1,48 \cdot 100\% = 9,86\%.$$

Ошибка аппроксимации больше 7%, что говорит о плохом подборе модели гиперболической регрессии к исходным данным.

Построим теперь *поле корреляции* и *график полученного уравнения гиперболической регрессии*.

Для этого выделим столбцы X, Y,  $\hat{y}$ , выберем в меню Excel **Вставка – Точечная**. Из предлагаемых форматов графиков выбираем первый (состоящий из точек).

На экране появится график вида

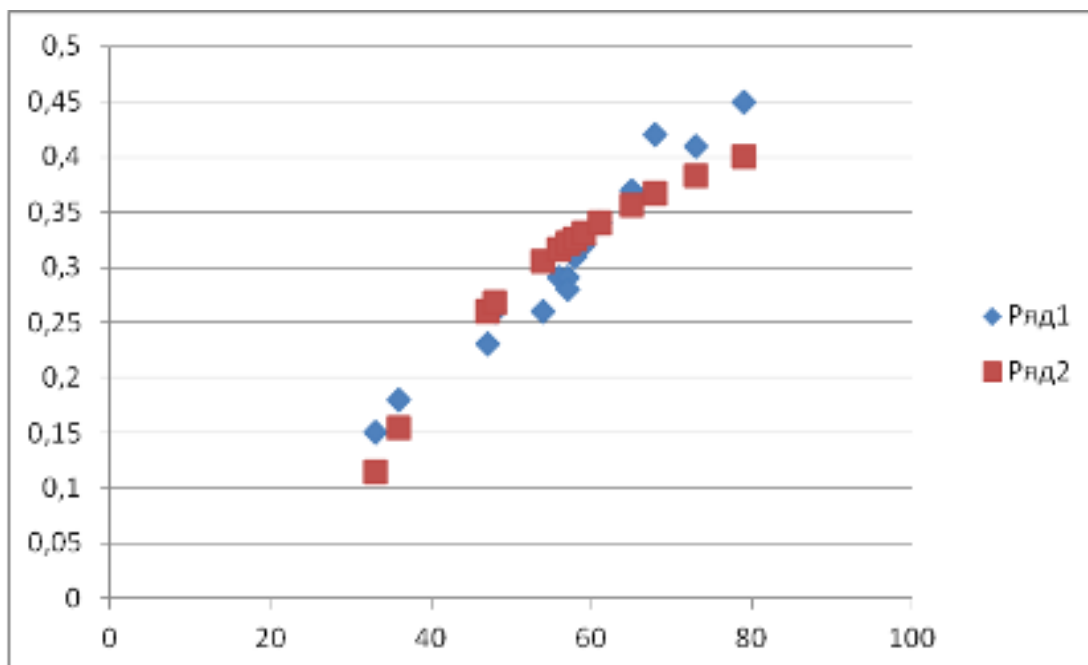


Рисунок 20 – Фрагмент построения графика в Excel



Здесь Ряд 1 – поле корреляции (его так и оставляем в виде точек), а Ряд 2 – график уравнения параболической регрессии (его необходимо преобразовать). Выбираем точку из Ряда 2, «щёлкаем» по ней правой кнопкой мыши.

В появившемся меню выбираем **«Изменить тип диаграммы»** и из предлагаемых Точечных графиков выбираем третий (гладкая кривая без маркеров).

Окончательно график будет выглядеть следующим образом:

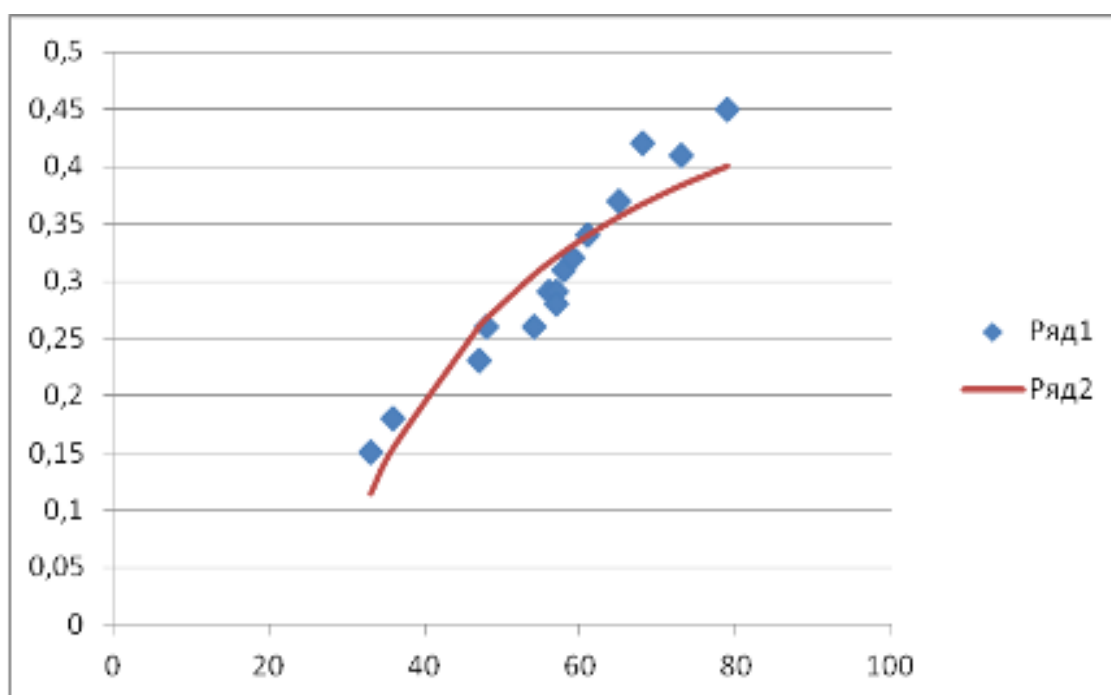


Рисунок 21 – Поле корреляции и график уравнения параболической регрессии

## 5. Задания для самостоятельной работы

1. Используйте те же исходные данные, что и для уравнения линейной регрессии (приложение 1).
2. Руководствуясь рассмотренным примером и в соответствии со своим заданием заполните таблицу:

$N\bar{o}$	$X$	$Y$	$X^2$	$X^3$	$X^4$	$XY$	$X^2Y$	$\hat{y}$	$\left  \frac{y_i - \hat{y}_i}{y_i} \right $
1									
2									
3									
4									
5									
...	...	...	...	...	...	...	...	...	...
15									
$\Sigma$									

3. Вычислите  $a_0, a_1, a_2$ , составьте уравнение параболической регрессии  $\hat{y} = a_0 + a_1 \cdot x + a_2 \cdot x^2$ .

4. Вычислите среднюю ошибку аппроксимации и сделайте вывод о качестве уравнения параболической регрессии.

5. Постройте поле корреляции и график полученного уравнения параболической регрессии.

6. Руководствуясь рассмотренным примером и в соответствии со своим заданием заполните таблицу:

$N\bar{o}$	$X = V$	$Y = Q$	$\frac{1}{X}$	$\frac{1}{X^2}$	$\frac{Y}{X}$	$\hat{y}$	$\left  \frac{y_i - \hat{y}_i}{y_i} \right $
1							
2							
3							
...	...	...	...	...	...	...	...
15							
$\Sigma$							

7. Вычислите  $a_0, a_1$ , составьте уравнение гиперболической регрессии

$$\hat{y} = a_0 + \frac{a_1}{x}.$$

8. Вычислите среднюю ошибку аппроксимации и сделайте вывод о качестве уравнения гиперболической регрессии.

9. Постройте поле корреляции и график полученного уравнения гиперболической регрессии.

10. Составьте отчет по лабораторной работе.

## **6. Содержание отчета**

Отчет должен содержать:

1. Определения и формулы, изложенные в теоретической части лабораторной работы.
2. Исходные данные и результаты выполняемых заданий.

### **Контрольные вопросы**

1. Запишите уравнение параболической регрессии.
2. Приведите систему для нахождения параметров параболической регрессии.
3. Приведите формулы Крамера и формулы определителей для вычисления параметров параболической регрессии.
4. Запишите уравнение параболической регрессии.
5. Запишите уравнение гиперболической регрессии.
6. Приведите систему для нахождения параметров гиперболической регрессии.
7. Приведите формулы Крамера и формулы определителей для вычисления параметров гиперболической регрессии.

## ЛАБОРАТОРНО-ПРАКТИЧЕСКИЕ ЗАНЯТИЯ № 5, 6

**Тема:** Корреляция для нелинейной регрессии. Коэффициент эластичности.

**Цель работы:** Познакомиться с корреляцией для нелинейной регрессии, научиться вычислять индекс детерминации, индекс корреляции для нелинейной регрессии, коэффициенты эластичности, интерпретировать полученные результаты.

### 1. Корреляция для нелинейной регрессии.

#### Индекс детерминации

Для статистической оценки тесноты связи в общем случае применяются следующие показатели вариации.

1. *Общая дисперсия* результативного признака  $\sigma_y^2$

$$\sigma_y^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n}, \quad (48)$$

отображающая совокупное влияние всех факторов.

Отклонение  $y_i - \bar{y}$  обусловлено тем, что сочетание значений факторов, влияющих на вариацию признака  $y$ , для каждой единицы анализируемой совокупности различно.

Здесь  $\bar{y}$  – среднее значение результата  $y$

$$\bar{y} = \frac{\sum_{i=1}^n y_i}{n}.$$

2. *Факторная дисперсия* результативного признака  $\sigma_{\hat{y}}^2$

$$\sigma_{\hat{y}}^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{n}, \quad (49)$$

отображающая вариацию  $y$  только от воздействия изучаемого фактора  $x$ .

Отклонения  $y_i - \bar{y}$  характеризуют разброс значений  $y_i$  относительно средней величины  $\bar{y}$ .

### 3. Остаточная дисперсия $\sigma_\varepsilon^2$

$$\sigma_\varepsilon^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}, \quad (50)$$

отображающая вариацию результативного признака  $y$  от всех прочих, кроме  $x$ , факторов.

Отклонения  $y_i - \hat{y}_i$  характеризуют колеблемость эмпирических (фактических) значений результативного признака  $y_i$  от их выравненных значений  $\hat{y}_i$ .

Общая дисперсия складывается из суммы факторной (объясненной уравнением регрессии) дисперсии и остаточной (обусловленной неучтенными факторами) дисперсии:

$$\sigma_y^2 = \sigma_{\hat{y}}^2 + \sigma_\varepsilon^2. \quad (51)$$

Формулу (51) называют «правилом сложения дисперсий».

Соотношение между факторной и общей дисперсиями характеризует меру тесноты связи между признаками  $x$  и  $y$ :

$$\frac{\sigma_{\hat{y}}^2}{\sigma_y^2} = R^2. \quad (52)$$

Показатель  $R^2$  называется *индексом детерминации*. Он выражает долю факторной дисперсии в общей дисперсии, то есть характеризует, какая часть общей вариации результативного признака  $y$  объясняется изучаемым фактором  $x$ .

Из «правила сложения дисперсий» (51) можно выразить

$$\sigma_{\hat{y}}^2 = \sigma_y^2 - \sigma_\varepsilon^2.$$

Тогда

$$R^2 = \frac{\sigma_y^2 - \sigma_\varepsilon^2}{\sigma_y^2}$$

или

$$R^2 = 1 - \frac{\sigma_\varepsilon^2}{\sigma_y^2}. \quad (53)$$

Таким образом, *индекс детерминации* может быть рассчитан также по формуле (53).

Также из формулы (52) с учетом формул (48) и (49) можно получить еще одну формулу для расчета *индекса детерминации*:

$$R^2 = \frac{\sigma_{\hat{y}}^2}{\sigma_y^2} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}. \quad (54)$$

**Замечание.** При функциональной связи значения  $\hat{y}_i$  полностью совпадают с соответствующими индивидуальными значениями  $y_i$ . Тогда  $\sigma_\varepsilon^2 = 0$ . В этом случае  $\sigma_{\hat{y}}^2 = \sigma_y^2$ , а при наличии корреляционной связи  $\sigma_{\hat{y}}^2 < \sigma_y^2$ .

## 2. Индекс корреляции

Уравнение нелинейной регрессии, так же как и в линейной зависимости, дополняется показателем тесноты связи, а именно *индексом корреляции*  $R_{xy}$ .

На основании формулы (52) определяется *индекс корреляции*  $R_{xy}$ :

$$R_{xy} = \sqrt{\frac{\sigma_{\hat{y}}^2}{\sigma_y^2}}. \quad (55)$$

Также индекс корреляции с учетом формул (53) и (54) может быть вычислен по формулам

$$R_{xy} = \sqrt{1 - \frac{\sigma_{\varepsilon}^2}{\sigma_y^2}} \quad (56)$$

или

$$R_{xy} = \sqrt{\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}}. \quad (57)$$

**Замечание.** По абсолютной величине линейный коэффициент корреляции  $r_{xy}$  равен индексу корреляции  $R_{xy}$  только при прямолинейной форме связи.

Индекс корреляции  $R_{xy}$  принимает значения в диапазоне

$$0 \leq R_{xy} \leq 1.$$

Чем ближе величина  $R_{xy}$  к единице, тем теснее связь между переменными и тем лучше уравнение регрессии согласуется с данными наблюдений.

При  $R_{xy} = 1$  связь становится функциональной.

Для характеристики тесноты связи, как и в случае линейной зависимости, применяют шкалу Чеддока (приложение 2).

### 3. Проверка качества уравнения нелинейной регрессии.

#### ***F*-критерий Фишера**

Индекс детерминации  $R^2$  используется для проверки статистической значимости в целом уравнения нелинейной регрессии по *F*-критерию Фишера.

Согласно *F*-критерию Фишера, выдвигается «нулевая» гипотеза  $H_0$  о статистической незначимости уравнения регрессии. Эта гипотеза отвергается при выполнении условия

$$F_R > F_{\text{крит}},$$

где  $F_{крит}$  определяется по таблицам  $F$ -критерия Фишера (приложение 4) при числах степеней свободы

$$k_1 = m, \quad k_2 = n - m - 1$$

и заданному уровню значимости  $\alpha$ .

Здесь  $m$  – число параметров при переменных  $x$ .

Фактическое значение критерия  $F_R$  определяется по формуле

$$F_R = \frac{R^2}{1 - R^2} \cdot \frac{n - m - 1}{m}. \quad (58)$$

Величина  $F_R$  сравнивается с критическим значением  $F_{крит}$ .

Если  $F_R > F_{крит}$ , то «нулевая» гипотеза о статистической незначимости уравнения регрессии отвергается, и уравнения регрессии признается статистически значимым с надежностью  $1 - \alpha$ .

В противном случае, если  $F_R < F_{крит}$ , принимается гипотеза о статистической незначимости уравнения регрессии.

#### **4. Пример вычисления индексов корреляции и детерминации для нелинейной регрессии и оценки качества уравнения нелинейной регрессии по критерию Фишера**

Имеются данные об изменении расхода топлива  $Q$  (л/км) автомобиля при изменении скорости его движения  $V$  (км/ч) (стр. 17, таблица 2).

Получено уравнение гиперболической регрессии, отражающее зависимость расхода топлива от скорости автомобиля  $Q(V)$  (стр. 44):

$$\hat{y} = 0,605 - \frac{16,186}{x}.$$

Найти общую, факторную и остаточную дисперсии. Проверить правило сложения дисперсий. Вычислить индекс детерминации  $R^2$ , индекс корреляции  $R_{xy}$ .



Оценить статистическую значимость уравнения гиперболической регрессии по  $F$ -критерию Фишера.

Для вычисления общей, факторной и остаточной дисперсии таблицу 11 дополним тремя новыми колонками (таблица 12):

Таблица 12 – Данные для вычисления общей, факторной и остаточной дисперсии

$N_{\text{д}}$	$X$	$Y$	$\frac{1}{X}$	$\frac{1}{X^2}$	$\frac{Y}{X}$	$\hat{y}$	$\left  \frac{y_i - \hat{y}_i}{y_i} \right $	$(y_i - \bar{y})^2$	$(\hat{y}_i - \bar{y})^2$	$(y_i - \hat{y}_i)^2$
1	33	0,15	0,0303	0,0009	0,0045	0,11	0,24	0,023716	0,036006	0,001278
2	36	0,18	0,0278	0,0008	0,005	0,16	0,14	0,015376	0,022166	0,000619
3	47	0,23	0,0213	0,0005	0,0049	0,26	0,13	0,005476	0,001906	0,000921
4	48	0,26	0,0208	0,0004	0,0054	0,27	0,03	0,001936	0,001331	5,65E-05
5	54	0,26	0,0185	0,0003	0,0048	0,30	0,17	0,001936	9,59E-07	0,002023
6	56	0,29	0,0179	0,0003	0,0052	0,32	0,09	0,000196	0,000137	0,00066
7	57	0,29	0,0175	0,0003	0,0051	0,32	0,11	0,000196	0,000281	0,000946
8	57	0,28	0,0175	0,0003	0,0049	0,32	0,15	0,000576	0,000281	0,001661
9	58	0,31	0,0172	0,0003	0,0053	0,33	0,05	3,6E-05	0,000469	0,000245
10	59	0,32	0,0169	0,0003	0,0054	0,33	0,03	0,000256	0,000696	0,000108
11	61	0,34	0,0164	0,0003	0,0056	0,34	0,00	0,001296	0,001251	3,92E-07
12	65	0,37	0,0154	0,0002	0,0057	0,36	0,04	0,004356	0,002673	0,000204
13	68	0,42	0,0147	0,0002	0,0062	0,37	0,13	0,013456	0,00393	0,002842
14	73	0,41	0,0137	0,0002	0,0056	0,38	0,07	0,011236	0,006239	0,00073
15	79	0,45	0,0127	0,0002	0,0057	0,40	0,11	0,021316	0,009183	0,002517
$\Sigma$	851	4,56	0,2787	0,0055	0,0794	4,56	1,48	0,10136	0,08655	0,01481

Общую дисперсию  $\sigma_y^2$  вычислим по формуле (48):

$$\sigma_y^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n} = \frac{0,10136}{15} = 0,006757.$$

Факторную дисперсию  $\sigma_{\hat{y}}^2$  вычислим по формуле (49):

$$\sigma_{\hat{y}}^2 = \frac{\sum_{i=1}^n (\bar{y}_i - \bar{y})^2}{n} = \frac{0,08655}{15} = 0,00577.$$

Остаточную дисперсию  $\sigma_{\varepsilon}^2$  вычислим по формуле (50):

$$\sigma_{\varepsilon}^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n} = \frac{0,01481}{15} = 0,000987.$$

Проверим правило сложения дисперсий (формула (51):

$$\sigma_y^2 = \sigma_{\hat{y}}^2 + \sigma_{\varepsilon}^2.$$

$$0,00577 + 0,000987 = 0,006757.$$

Правило сложения дисперсий выполняется.

Индекс детерминации  $R^2$  вычислим по формуле (52):

$$R^2 = \frac{\sigma_{\hat{y}}^2}{\sigma_y^2} = \frac{0,00577}{0,006757} = 0,854.$$

То есть 85,4% общей вариации результативного признака  $y$  (расхода топлива) объясняется изучаемым фактором  $x$  (изменением скорости движения автомобиля), а 14,6% – неучтенными в уравнении гиперболической регрессии факторами.

Индекс корреляции вычислим по формуле (55):

$$R_{xy} = \sqrt{\frac{\sigma_{\hat{y}}^2}{\sigma_y^2}} = \sqrt{0,854} = 0,924.$$

По шкале Чеддока (приложение 2) связь между признаками  $x$  и  $y$  весьма высокая.

Для оценки статистической значимости уравнения гиперболической регрессии применим  $F$ -критерий Фишера.

Фактическое значение критерия  $F_R$  определим по формуле (58):

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

где  $m = 1$  – число параметров при переменной  $x$ .

$$F_R = \frac{0,854}{1 - 0,854} \cdot \frac{15 - 1 - 1}{1} = 75,97.$$

Значение  $F_{крит}$  определим по таблице  $F$ -критерия Фишера (приложение 4) с учётом уровня значимости  $\alpha = 0,05$  и чисел степеней свободы

$$k_1 = m = 1, \quad k_2 = n - m - 1 = 15 - 1 - 1 = 13.$$

$$F_{крит} = 4,67.$$

Так как  $75,97 > 4,67$  ( $F_R > F_{крит}$ ), то уравнение гиперболической регрессии признается статистически значимым с надежностью 95%.

## 5. Коэффициент эластичности

В экономических исследованиях широкое применение находит такой показатель, как *коэффициент эластичности*.

Если зависимость между переменными  $x$  и  $y$  имеет вид  $y = f(x)$ , то *коэффициент эластичности*  $\mathcal{E}$  вычисляется по формуле

$$\mathcal{E} = f'(x) \cdot \frac{x}{y}, \quad (59)$$

где  $f'(x)$  – первая производная, характеризующая соотношение приростов результата и фактора для соответствующей формы связи.

*Коэффициент эластичности*  $\mathcal{E}$  показывает, на сколько процентов в среднем изменится результативный признак  $y$  при изменении фактора  $x$  на 1% от своего номинального значения.

Коэффициент эластичности  $\mathcal{E}$  в общем случае зависит от величины  $x$  и является величиной переменной.

Чтобы исключить эту зависимость применяется *средний коэффициент эластичности*

$$\bar{\mathcal{E}} = f'(\bar{x}) \cdot \frac{\bar{x}}{\bar{y}}, \quad (60)$$

который уже является величиной постоянной.

Средний коэффициент эластичности  $\bar{\mathcal{E}}$  показывает, на сколько процентов в среднем по совокупности значений фактора  $x$  изменится результативный признак  $y$  при изменении фактора  $x$  на 1%.

В таблице 13 приведены коэффициенты эластичности для различных видов уравнений регрессии.

Таблица 13 – Коэффициенты эластичности для ряда уравнений регрессии.

<b>Вид регрессии</b> $\hat{y}$	<b>Первая производная</b> $f'(x) = \hat{y}'$	<b>Коэффициент эластичности</b> $\mathcal{E} = f'(x) \cdot \frac{x}{y}$
Линейная $\hat{y} = a_0 + a_1 \cdot x$	$f'(x) = a_1$	$\mathcal{E} = \frac{a_1 \cdot x}{a_0 + a_1 \cdot x}$
Параболическая $\hat{y} = a_0 + a_1 \cdot x + a_2 \cdot x^2$	$f'(x) = a_1 + 2 \cdot a_2 \cdot x$	$\mathcal{E} = \frac{a_1 + 2 \cdot a_2 \cdot x}{a_0 + a_1 \cdot x + a_2 \cdot x^2}$
Гиперболическая $\hat{y} = a_0 + \frac{a_1}{x}$	$f'(x) = -\frac{a_1}{x^2}$	$\mathcal{E} = \frac{-a_1}{a_0 \cdot x + a_1}$

## 6. Пример вычисления коэффициента эластичности

Имеются данные об изменении расхода топлива  $Q$  (л/км) автомобиля при изменении скорости его движения  $V$  (км/ч) (стр. 17, таблица 2).

Вычислить средний коэффициент эластичности для уравнений линейной, параболической и гиперболической регрессии и пояснить его смысл.

Вычислим средний коэффициент эластичности для уравнений линейной регрессии (25):

$$\hat{y} = -0,077 + 0,007 \cdot x.$$

Из таблицы 7 (стр. 30) возьмем значение  $\bar{x} = 56,733$ .

Воспользуемся формулой среднего коэффициента эластичности для линейной регрессии из таблицы 13:

$$\Theta = \frac{a_1 \cdot \bar{x}}{a_0 + a_1 \cdot \bar{x}} = \frac{0,007 \cdot 56,733}{-0,077 + 0,007 \cdot 56,733} = 1,255.$$

Следовательно, при увеличении скорости движения автомобиля  $V$  на 1% расход топлива  $Q$  в среднем увеличится на 1,255%.

Вычислим теперь средний коэффициент эластичности для уравнений параболической регрессии (40):

$$\hat{y} = -0,0051 + 0,004 \cdot x + 0,000025 \cdot x^2.$$

Воспользуемся формулой среднего коэффициента эластичности для параболической регрессии из таблицы 13:

$$\Theta = \frac{a_1 + 2 \cdot a_2 \cdot \bar{x}}{a_0 + a_1 \cdot \bar{x} + a_2 \cdot \bar{x}^2} = \frac{(0,004 + 2 \cdot 0,000025 \cdot 56,733) \cdot 56,733}{-0,0051 + 0,004 \cdot 56,733 + 0,000025 \cdot 56,733^2} = 1,284.$$

Следовательно, при увеличении скорости движения автомобиля  $V$  на 1% расход топлива  $Q$  в среднем увеличится на 1,284%.

Вычислим теперь средний коэффициент эластичности для уравнений гиперболической регрессии (47):

$$\hat{y} = 0,605 - \frac{16,185}{x}.$$

Воспользуемся формулой среднего коэффициента эластичности для гиперболической регрессии из таблицы 13:

$$\varepsilon = \frac{-a_1}{a_0 \cdot x + a_1} = \frac{-(-16,185)}{0,605 \cdot 56,733 - 16,185} = 0,893.$$

Следовательно, при увеличении скорости движения автомобиля  $V$  на 1% расход топлива  $Q$  в среднем увеличится на 0,893%.

## 7. Задания для самостоятельной работы

1. По найденному ранее уравнению параболической регрессии рассчитать: общую, факторную, остаточную дисперсии. Проверить правило сложения дисперсий.
2. Для уравнения параболической регрессии вычислить индекс детерминации, индекс корреляции. Дать интерпретацию полученным результатам.
3. С помощью критерия Фишера оценить значимость уравнения параболической регрессии.
4. По найденному ранее уравнению гиперболической регрессии рассчитать: общую, факторную, остаточную дисперсии. Проверить правило сложения дисперсий.
5. Для уравнения гиперболической регрессии вычислить индекс детерминации, индекс корреляции. Дать интерпретацию полученным результатам.
6. С помощью критерия Фишера оценить значимость уравнения гиперболической регрессии.
7. Вычислить средний коэффициент эластичности для уравнения линейной регрессии. Объяснить полученный результат.
8. Вычислить средний коэффициент эластичности для уравнения параболической регрессии. Объяснить полученный результат.
9. Вычислить средний коэффициент эластичности для уравнения гиперболической регрессии. Объяснить полученный результат.
10. Составить отчет по лабораторной работе.

## 8. Содержание отчета

Отчет должен содержать:

1. Определения и формулы, изложенные в теоретической части лабораторной работы.
2. Исходные данные и результаты выполняемых заданий.

### Контрольные вопросы

1. Дайте определение общей, факторной и остаточной дисперсий.
2. В чем состоит «правило сложения дисперсий»?
3. Определите понятия индекса детерминации.
4. Дайте определение индекса корреляции.
5. Как проверяется значимость уравнения нелинейной регрессии?
6. Как вычисляется и как интерпретируется коэффициент эластичности?
7. Приведите формулы для вычисления коэффициентов эластичности для линейной, параболической и гиперболической регрессии.

## ЛАБОРАТОРНО-ПРАКТИЧЕСКИЕ ЗАНЯТИЯ № 7, 8

**Тема:** Синтезирование моделей множественной регрессии.

**Цель работы:** Познакомиться с уравнениями множественной регрессии. Научиться вычислять коэффициенты множественной регрессии, оценивать тесноту связи между переменными, оценивать качество полученного уравнения и значимость входящих в него факторов. Научиться вычислять коэффициенты эластичности для множественной регрессии и интерпретировать их значение.

### 1. Спецификация модели. Отбор факторов при построении уравнения множественной регрессии

Построение уравнения множественной регрессии начинается с решения вопроса о спецификации модели. Он включает в себя два круга вопросов: отбор факторов и выбор вида уравнения регрессии.

Факторы, включаемые во множественную регрессию, должны отвечать следующим требованиям:

1. Факторы должны быть количественно измеримы. Если необходимо включить в модель качественный фактор, не имеющий количественного измерения, то ему нужно придать количественную определенность.

2. Факторы не должны быть коррелированы между собой (*интеркоррелированы*) и тем более находиться в точной функциональной связи.

Если между факторами существует высокая корреляция, то нельзя определить их изолированное влияние на результативный показатель и параметры уравнения регрессии оказываются неинтерпретируемыми.

Насыщение модели лишними факторами не только не снижает величину остаточной дисперсии и не увеличивает показатель детерминации, но и приводит к статистической незначимости параметров регрессии.

Таким образом, хотя теоретически регрессионная модель позволяет учесть любое число факторов, практически в этом нет необходимости.



Считается, что две переменные явно *коллинеарны*, т.е. находятся между собой в линейной зависимости, если

$$r_{x_i x_j} \geq 0,7.$$

Если факторы явно коллинеарны, то они дублируют друг друга и один из них рекомендуется исключить из уравнения множественной регрессии.

Включение в модель коллинеарных факторов нежелательно в силу следующих последствий:

1. Затрудняется интерпретация параметров множественной регрессии как характеристик действия факторов в «чистом» виде, ибо факторы коррелированы; параметры линейной регрессии теряют экономический смысл.

2. Оценки параметров ненадежны, обнаруживают большие стандартные ошибки и меняются с изменением объема наблюдений (не только по величине, но и по знаку), что делает модель непригодной для анализа и прогнозирования.

## 2. Оценка параметров уравнения множественной линейной регрессии

### *Метод наименьших квадратов (МНК)*

Рассмотрим модель множественной регрессии с двумя факторами:

$$\hat{y} = a_0 + a_1 \cdot x_1 + a_2 \cdot x_2. \quad (61)$$

Для нахождения параметров  $a_0$ ,  $a_1$ ,  $a_2$  линейного уравнения множественной регрессии двухфакторной модели необходимо решить систему линейных уравнений

$$\begin{cases} a_0 \cdot n + a_1 \cdot \sum_{i=1}^n x_{1i} + a_2 \cdot \sum_{i=1}^n x_{2i} = \sum_{i=1}^n y_i \\ a_0 \cdot \sum_{i=1}^n x_{1i} + a_1 \cdot \sum_{i=1}^n x_{1i}^2 + a_2 \cdot \sum_{i=1}^n x_{2i} \cdot x_{1i} = \sum_{i=1}^n y_i \cdot x_{1i} \\ a_0 \cdot \sum_{i=1}^n x_{2i} + a_1 \cdot \sum_{i=1}^n x_{1i} \cdot x_{2i} + a_2 \cdot \sum_{i=1}^n x_{2i}^2 = \sum_{i=1}^n y_i \cdot x_{2i} \end{cases} \quad (62)$$

Приведем формулы Крамера для решения систем (62):

$$a_0 = \frac{\Delta a_0}{\Delta}, \quad a_1 = \frac{\Delta a_1}{\Delta}, \quad a_2 = \frac{\Delta a_2}{\Delta}, \quad (63)$$

где  $\Delta$  – определитель основной матрицы системы уравнений (62)

$$\Delta = \begin{vmatrix} n & \sum_{i=1}^n x_{1i} & \sum_{i=1}^n x_{2i} \\ \sum_{i=1}^n x_{1i} & \sum_{i=1}^n x_{1i}^2 & \sum_{i=1}^n x_{2i} \cdot x_{1i} \\ \sum_{i=1}^n x_{2i} & \sum_{i=1}^n x_{1i} \cdot x_{2i} & \sum_{i=1}^n x_{2i}^2 \end{vmatrix}, \quad (64)$$

а  $\Delta a_0$ ,  $\Delta a_1$ ,  $\Delta a_2$  – определители, получаемые из  $\Delta$  заменой 1-го, 2-го и 3-го столбца соответственно на столбец свободных членов (правых частей уравнений системы (62)), то есть

$$\Delta a_0 = \begin{vmatrix} \sum_{i=1}^n y_i & \sum_{i=1}^n x_{1i} & \sum_{i=1}^n x_{2i} \\ \sum_{i=1}^n y_i \cdot x_{1i} & \sum_{i=1}^n x_{1i}^2 & \sum_{i=1}^n x_{2i} \cdot x_{1i} \\ \sum_{i=1}^n y_i \cdot x_{2i} & \sum_{i=1}^n x_{1i} \cdot x_{2i} & \sum_{i=1}^n x_{2i}^2 \end{vmatrix}, \quad (65)$$

$$\Delta a_1 = \begin{vmatrix} n & \sum_{i=1}^n y_i & \sum_{i=1}^n x_{2i} \\ \sum_{i=1}^n x_{1i} & \sum_{i=1}^n y_i x_{1i} & \sum_{i=1}^n x_{2i} \cdot x_{1i} \\ \sum_{i=1}^n x_{2i} & \sum_{i=1}^n y_i \cdot x_{2i} & \sum_{i=1}^n x_{2i}^2 \end{vmatrix}, \quad (66)$$

$$\Delta_{a_2} = \begin{vmatrix} n & \sum_{i=1}^n x_{1i} & \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_{1i} & \sum_{i=1}^n x_{1i}^2 & \sum_{i=1}^n y_i \cdot x_{1i} \\ \sum_{i=1}^n x_{2i} & \sum_{i=1}^n x_{1i} \cdot x_{2i} & \sum_{i=1}^n y_i \cdot x_{2i} \end{vmatrix}. \quad (67)$$

### 3. Пример составления уравнения множественной линейной регрессии

Исследуется зависимость производительности автомобиля  $Q$  от скорости автомобиля  $V$  и его грузоподъемности  $q$  (таблица 14).

Таблица 14 – Исходные данные

№	V(км/ч)	q(т)	Q(т*км/ч)
1	28	0,57	15,96
2	29	0,73	21,17
3	32	0,6	19,2
4	33	0,51	16,83
5	36	0,73	26,28
6	37	0,62	22,94
7	39	0,79	30,81
8	40	0,48	19,2
9	44	0,85	37,4
10	46	0,94	43,24
11	47	0,79	37,13
12	49	0,6	29,4
13	53	0,53	28,09
14	54	0,61	32,94
15	55	0,75	41,25
16	55	0,8	44
17	57	0,62	35,34
18	57	0,62	35,34
19	59	0,58	34,22
20	61	0,49	29,89
21	63	0,54	34,02
22	66	0,39	25,74
23	66	0,69	45,54
24	66	0,58	38,28
25	68	0,75	51

Составить уравнение множественной линейной регрессии, отражающее указанную зависимость, используя метод наименьших квадратов.

Построим линейное уравнение множественной регрессии, отображающее зависимость  $Q(V,q)$ , обозначив производительность  $Q$  через  $y$ , скорость  $V$  через  $x_1$ , грузоподъемность  $q$  через  $x_2$ :

$$\hat{y} = a_0 + a_1 \cdot x_1 + a_2 \cdot x_2 .$$

Для определения коэффициентов  $a_0, a_1, a_2$  необходимо рассчитать таблицу значений:

Таблица 15 – Исходная информация для определения коэффициентов уравнения множественной регрессии

№	$V=x_1$	$q=x_2$	$Q=y$	$x_1^2$	$x_2^2$	$x_1 \cdot x_2$	$y \cdot x_1$	$y \cdot x_2$	$\hat{y}$	$\left  \frac{y - \hat{y}}{y} \right  \cdot 100$	$y^2$
1	28	0,57	15,96	784	0,325	15,96	446,88	9,097	14,66	8,161	254,72
2	29	0,73	21,17	841	0,533	21,17	613,93	15,454	23,26	9,850	448,17
3	32	0,6	19,2	1024	0,36	19,2	614,4	11,52	18,62	2,999	368,64
4	33	0,51	16,83	1089	0,260	16,83	555,39	8,583	14,75	12,341	283,25
5	36	0,73	26,28	1296	0,533	26,28	946,08	19,184	27,58	4,941	690,64
6	37	0,62	22,94	1369	0,384	22,94	848,78	14,223	22,71	1,003	526,24
7	39	0,79	30,81	1521	0,624	30,81	1201,59	24,340	32,42	5,238	949,26
8	40	0,48	19,2	1600	0,230	19,2	768	9,216	17,58	8,438	368,64
9	44	0,85	37,4	1936	0,723	37,4	1645,6	31,79	38,51	2,954	1398,76
10	46	0,94	43,24	2116	0,884	43,24	1989,04	40,646	44,23	2,286	1869,70
11	47	0,79	37,13	2209	0,624	37,13	1745,11	29,333	37,37	0,632	1378,64
12	49	0,6	29,4	2401	0,36	29,4	1440,6	17,64	29,12	0,940	864,36
13	53	0,53	28,09	2809	0,281	28,09	1488,77	14,888	28,10	0,045	789,05
14	54	0,61	32,94	2916	0,372	32,94	1778,76	20,093	32,71	0,697	1085,04
15	55	0,75	41,25	3025	0,563	41,25	2268,75	30,938	40,31	2,277	1701,56
16	55	0,8	44	3025	0,64	44	2420	35,2	42,80	2,717	1936
17	57	0,62	35,34	3249	0,384	35,34	2014,38	21,911	35,06	0,787	1248,92
18	57	0,62	35,34	3249	0,384	35,34	2014,38	21,911	35,06	0,787	1248,92
19	59	0,58	34,22	3481	0,336	34,22	2018,98	19,848	34,30	0,241	1171,01
20	61	0,49	29,89	3721	0,240	29,89	1823,29	14,646	31,05	3,876	893,41
21	63	0,54	34,02	3969	0,292	34,02	2143,26	18,371	34,78	2,227	1157,36
22	66	0,39	25,74	4356	0,152	25,74	1698,84	10,039	29,15	13,244	662,55
23	66	0,69	45,54	4356	0,476	45,54	3005,64	31,423	44,11	3,136	2073,89
24	66	0,58	38,28	4356	0,336	38,28	2526,48	22,202	38,63	0,902	1465,36
25	68	0,75	51	4624	0,563	51	3468	38,25	48,34	5,217	2601
Σ	1240	16,16	795,21	65322	10,86	795,21	41484,93	530,74	795,21	95,94	27435,08
Сред.	49,6	0,646	31,808	2612,88	0,434	31,808	1659,397	21,230	14,66	8,161	1097,40

Используя метод Крамера (формулы 63), найдём  $a_0, a_1$  и  $a_2$  :

$$a_0 = \frac{\Delta a_0}{\Delta},$$

$$a_1 = \frac{\Delta a_1}{\Delta},$$

$$a_2 = \frac{\Delta a_2}{\Delta},$$

где

$$\Delta = \begin{vmatrix} n & \sum_{i=1}^n x_{1i} & \sum_{i=1}^n x_{2i} \\ \sum_{i=1}^n x_{1i} & \sum_{i=1}^n x_{1i}^2 & \sum_{i=1}^n x_{2i} \cdot x_{1i} \\ \sum_{i=1}^n x_{2i} & \sum_{i=1}^n x_{1i} \cdot x_{2i} & \sum_{i=1}^n x_{2i}^2 \end{vmatrix} = \begin{vmatrix} 25 & 1240 & 16,16 \\ 1240 & 65322 & 795,21 \\ 16,16 & 795,21 & 10,86 \end{vmatrix} = 38475,37,$$

$$\Delta_{a_0} = \begin{vmatrix} \sum_{i=1}^n y_i & \sum_{i=1}^n x_{1i} & \sum_{i=1}^n x_{2i} \\ \sum_{i=1}^n y_i \cdot x_{1i} & \sum_{i=1}^n x_{1i}^2 & \sum_{i=1}^n x_{2i} \cdot x_{1i} \\ \sum_{i=1}^n y_i \cdot x_{2i} & \sum_{i=1}^n x_{1i} \cdot x_{2i} & \sum_{i=1}^n x_{2i}^2 \end{vmatrix} = \begin{vmatrix} 795,21 & 1240 & 16,16 \\ 41484,93 & 65322 & 795,21 \\ 530,74 & 795,21 & 10,86 \end{vmatrix} = -1195233,16,$$

$$\Delta_{a_1} = \begin{vmatrix} n & \sum_{i=1}^n y_i & \sum_{i=1}^n x_{2i} \\ \sum_{i=1}^n x_{1i} & \sum_{i=1}^n y_i x_{1i} & \sum_{i=1}^n x_{2i} \cdot x_{1i} \\ \sum_{i=1}^n x_{2i} & \sum_{i=1}^n y_i \cdot x_{2i} & \sum_{i=1}^n x_{2i}^2 \end{vmatrix} = \begin{vmatrix} 25 & 795,21 & 16,16 \\ 1240 & 41484,93 & 795,21 \\ 16,16 & 530,74 & 10,86 \end{vmatrix} = 23762,73,$$

$$\Delta_{a_2} = \begin{vmatrix} n & \sum_{i=1}^n x_{1i} & \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_{1i} & \sum_{i=1}^n x_{1i}^2 & \sum_{i=1}^n y_i \cdot x_{1i} \\ \sum_{i=1}^n x_{2i} & \sum_{i=1}^n x_{1i} \cdot x_{2i} & \sum_{i=1}^n y_i \cdot x_{2i} \end{vmatrix} = \begin{vmatrix} 25 & 1240 & 795,21 \\ 1240 & 65322 & 41484,93 \\ 16,16 & 795,21 & 530,74 \end{vmatrix} = 1919000,66.$$

$$a_0 = \frac{\Delta a_0}{\Delta} = \frac{-1195233,16}{38475,37} = -31,065,$$

$$a_1 = \frac{\Delta a_1}{\Delta} = \frac{23762,73}{38475,37} = 0,618,$$

$$a_2 = \frac{\Delta a_2}{\Delta} = \frac{1919000,66}{38475,37} = 49,876.$$

Таким образом, искомое уравнение множественной регрессии будет иметь вид:

$$\hat{y} = -31,065 + 0,618 \cdot x_1 + 49,876 \cdot x_2 \quad (68)$$

или

$$Q(V, q) = -31,065 + 0,618 \cdot V + 49,876 \cdot q.$$

#### 4. Уравнение в стандартизированном масштабе

Параметры  $a_0$ ,  $a_1$ ,  $a_2$  в линейной множественной регрессии называются коэффициентами «чистой» регрессии, а само уравнение

$$\hat{y} = a_0 + a_1 \cdot x_1 + a_2 \cdot x_2$$

называют уравнением в *натуральном масштабе* переменных.

Возможен и иной подход к определению параметров множественной регрессии, когда на основе матрицы парных коэффициентов корреляции строится уравнение регрессии в *стандартизированном масштабе*:

$$t_y = \beta_1 t_{x_1} + \beta_2 t_{x_2}, \quad (69)$$

где  $t_y, t_{x_1}, t_{x_2}$  – стандартизованные переменные, рассчитываемые по формулам:

$$t_y = \frac{y - \bar{y}}{\sigma_y}, \quad t_{x_i} = \frac{x_i - \bar{x}_i}{\sigma_{x_i}}, \quad (70)$$

$\beta_i$  – стандартизованные коэффициенты регрессии.

Стандартизованные коэффициенты регрессии показывают, на сколько единиц изменится в среднем результат, если соответствующий фактор  $x_i$  изменится на одну единицу при неизменном среднем уровне других факторов.

Коэффициенты «чистой» регрессии  $a_1, a_2$  связаны со стандартизованными коэффициентами регрессии  $\beta_i$  следующим образом:

$$a_i = \beta_i \frac{\sigma_y}{\sigma_{x_i}}, \quad (i=1, 2). \quad (71)$$

$$\beta_1 = \frac{r_{yx_1} - r_{yx_2} \cdot r_{x_1x_2}}{1 - r_{x_1x_2}^2}, \quad (72)$$

$$\beta_2 = \frac{r_{yx_2} - r_{yx_1} \cdot r_{x_1x_2}}{1 - r_{x_1x_2}^2}. \quad (73)$$

Параметр  $a_0$  определяется из уравнения

$$a_0 = \bar{y} - a_1 \cdot \bar{x}_1 - a_2 \cdot \bar{x}_2. \quad (74)$$

## 5. Проверка существенности факторов и показатели качества регрессии

### *Линейные коэффициенты парной корреляции*

Напомним, что линейный коэффициент корреляции показывает тесноту связи между переменными и может быть вычислен по формуле

$$r_{xy} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x \cdot \sigma_y},$$

где  $\bar{x}$  – среднее значение факторного признака  $x_i$ ;

$\bar{y}$  – среднее значение результативного признака  $y_i$ ;

$\overline{xy}$  – среднее значение произведений  $x_i y_i$ ;

$\sigma_x$  – среднее квадратическое отклонение факторного признака  $x_i$   
от общей средней  $\bar{x}$ :

$$\sigma_x = \sqrt{\overline{x^2} - \bar{x}^2};$$

$\sigma_y$  – среднее квадратическое отклонение результативного признака  $y_i$   
от общей средней  $\bar{y}$ :

$$\sigma_y = \sqrt{\overline{y^2} - \bar{y}^2}.$$

Для оценки тесноты связи применяют шкалу Чеддока (приложение 2).

Во множественной регрессии *парные коэффициенты корреляции* рассчитываются по аналогичным формулам:

$$r_{x_1 y} = \frac{\overline{x_1 y} - \bar{x}_1 \cdot \bar{y}}{\sigma_{x_1} \sigma_y}; \quad (75)$$

$$r_{x_2 y} = \frac{\overline{x_2 y} - \bar{x}_2 \cdot \bar{y}}{\sigma_{x_2} \sigma_y}; \quad (76)$$

$$r_{x_1 x_2} = \frac{\overline{x_1 x_2} - \bar{x}_1 \cdot \bar{x}_2}{\sigma_{x_1} \sigma_{x_2}}, \quad (77)$$

где

$$\sigma_{x_1} = \sqrt{\overline{x_1^2} - (\bar{x}_1)^2}, \quad (78)$$

$$\sigma_{x_2} = \sqrt{\overline{x_2^2} - (\bar{x}_2)^2}, \quad (79)$$

$$\sigma_y = \sqrt{\overline{y^2} - (\bar{y})^2}. \quad (80)$$

**Замечание:** Коэффициенты корреляции  $r_{x_i y}$  характеризуют тесноту связи между фактором  $x_i$  и результатом  $y$ , а коэффициенты  $r_{x_i x_j}$  – тесноту связи между факторами  $x_i$  и  $x_j$ .



Как уже говорилось ранее, две переменные явно *коллинеарны*, если находятся между собой в линейной зависимости. Это происходит, когда

$$r_{x_i x_j} \geq 0,7.$$

Если факторы явно коллинеарны, то они дублируют друг друга и один из них рекомендуется исключить из уравнения множественной регрессии.

### **6. Пример вычисления линейных коэффициентов парной корреляции. Составления уравнения множественной регрессии с использованием коэффициентов в стандартизированном масштабе**

Исследуется зависимость производительности автомобиля  $Q$  от скорости автомобиля  $V$  и его грузоподъемности  $q$  (таблица 16).

1. Составить уравнение линейной множественной регрессии в стандартизированном масштабе, отражающее указанную зависимость.

2. Вычислить коэффициенты «чистой» регрессии и записать соответствующее уравнение в *натуральном масштабе* переменных.

Таблица 16 – Исходные данные

№	V(км/ч)	q(т)	Q(т·км/ч)
1	33	2,47	81,51
2	36	1,65	59,40
3	54	1,84	99,36
4	56	2,37	132,72
5	79	2,13	168,27
6	57	1,09	62,13
7	68	2,4	163,20
8	73	1,92	140,16
9	57	1,35	76,95
10	48	1,78	85,44
11	61	1,49	90,89
12	59	1,76	103,84
13	65	2,31	150,15
14	47	1,67	78,49
15	58	2,24	129,92
16	77	1,79	137,83
17	62	2,28	141,36
18	61	2,50	152,50
19	74	2,38	176,12
20	56	2,31	129,36

1. Вычислим стандартизированные коэффициенты  $\beta_i$  по формулам (72)–(73):

$$\beta_1 = \frac{r_{yx_1} - r_{yx_2} \cdot r_{x_1x_2}}{1 - r_{x_1x_2}^2},$$

$$\beta_2 = \frac{r_{yx_2} - r_{yx_1} \cdot r_{x_1x_2}}{1 - r_{x_1x_2}^2},$$

где

$$r_{yx_1} = \frac{\overline{x_1 y} - \overline{x_1} \cdot \overline{y}}{\sigma_{x_1} \cdot \sigma_y},$$

$$r_{yx_2} = \frac{\overline{x_2 y} - \overline{x_2} \cdot \overline{y}}{\sigma_{x_2} \cdot \sigma_y},$$

$$r_{x_1x_2} = \frac{\overline{x_1 x_2} - \overline{x_1} \cdot \overline{x_2}}{\sigma_{x_1} \cdot \sigma_{x_2}},$$

$$\sigma_{x_1} = \sqrt{\overline{x_1^2} - (\overline{x_1})^2},$$

$$\sigma_{x_2} = \sqrt{\overline{x_2^2} - (\overline{x_2})^2},$$

$$\sigma_y = \sqrt{\overline{y^2} - (\overline{y})^2}.$$

Необходимые для данных формул значения удобно рассчитать в таблице (таблица 17).

Используя данные таблицы 17 и формулы (78)–(80), вычислим:

$$\sigma_{x_1} = \sqrt{\overline{x_1^2} - (\overline{x_1})^2} = \sqrt{3626,95 - 59,05^2} = \sqrt{140,048} = 11,834,$$

$$\sigma_{x_2} = \sqrt{\overline{x_2^2} - (\overline{x_2})^2} = \sqrt{4,105 - 1,987^2} = \sqrt{0,159} = 0,399,$$

$$\sigma_y = \sqrt{\overline{y^2} - (\overline{y})^2} = \sqrt{15197,57 - 117,98^2} = \sqrt{1278,29} = 35,75.$$

Таблица 17 – Информация для определения коэффициентов уравнения множественной регрессии

№	V=x <sub>1</sub>	q=x <sub>2</sub>	W=y	x <sub>1</sub> ·y	x <sub>2</sub> ·y	x <sub>1</sub> · x <sub>2</sub>	x <sub>1</sub> <sup>2</sup>	x <sub>2</sub> <sup>2</sup>	y <sup>2</sup>
1	33	2,47	81,51	2689,83	201,33	81,51	1089	6,101	6643,88
2	36	1,65	59,4	2138,4	98,01	59,4	1296	2,723	3528,36
3	54	1,84	99,36	5365,44	182,82	99,36	2916	3,386	9872,41
4	56	2,37	132,72	7432,32	314,55	132,72	3136	5,617	17614,60
5	79	2,13	168,27	13293,33	358,42	168,27	6241	4,537	28314,79
6	57	1,09	62,13	3541,41	67,72	62,13	3249	1,188	3860,14
7	68	2,4	163,2	11097,6	391,68	163,2	4624	5,760	26634,24
8	73	1,92	140,16	10231,68	269,11	140,16	5329	3,686	19644,83
9	57	1,35	76,95	4386,15	103,88	76,95	3249	1,823	5921,30
10	48	1,78	85,44	4101,12	152,08	85,44	2304	3,168	7299,99
11	61	1,49	90,89	5544,29	135,43	90,89	3721	2,220	8260,99
12	59	1,76	103,84	6126,56	182,76	103,84	3481	3,098	10782,75
13	65	2,31	150,15	9759,75	346,85	150,15	4225	5,336	22545,02
14	47	1,67	78,49	3689,03	131,08	78,49	2209	2,789	6160,68
15	58	2,24	129,92	7535,36	291,02	129,92	3364	5,018	16879,21
16	77	1,79	137,83	10612,91	246,72	137,83	5929	3,204	18997,11
17	62	2,28	141,36	8764,32	322,30	141,36	3844	5,198	19982,65
18	61	2,5	152,5	9302,5	381,25	152,5	3721	6,250	23256,25
19	74	2,38	176,12	13032,88	419,17	176,12	5476	5,664	31018,25
20	56	2,31	129,36	7244,16	298,82	129,36	3136	5,336	16734,01
Сумма	1181	39,73	2359,6	145889,04	4894,98	2359,6	72539	82,10	303951,46
Среднее	59,05	1,987	117,98	7294,452	244,75	117,98	3626,95	4,105	15197,57

Так же, используя данные таблицы 17 и формулы (75)–(77), вычислим парные коэффициенты корреляции:

$$r_{yx_1} = \frac{\overline{x_1 y} - \overline{x_1} \cdot \overline{y}}{\sigma_{x_1} \cdot \sigma_y} = \frac{7294,452 - 59,05 \cdot 117,98}{11,834 \cdot 35,75} = 0,775,$$

$$r_{yx_2} = \frac{\overline{x_2 y} - \overline{x_2} \cdot \overline{y}}{\sigma_{x_2} \cdot \sigma_y} = \frac{244,75 - 1,987 \cdot 117,98}{0,399 \cdot 35,75} = 0,728,$$

$$r_{x_1x_2} = \frac{\overline{x_1x_2} - \overline{x_1} \cdot \overline{x_2}}{\sigma_{x_1} \cdot \sigma_{x_2}} = \frac{117,98 - 59,05 \cdot 1,987}{11,834 \cdot 0,399} = 0,144.$$

Тогда

$$\beta_1 = \frac{r_{yx_1} - r_{yx_2} \cdot r_{x_1x_2}}{1 - r_{x_1x_2}^2} = \frac{0,775 - 0,728 \cdot 0,144}{1 - 0,144^2} = 0,684,$$

$$\beta_2 = \frac{r_{yx_2} - r_{yx_1} \cdot r_{x_1x_2}}{1 - r_{x_1x_2}^2} = \frac{0,728 - 0,775 \cdot 0,144}{1 - 0,144^2} = 0,630.$$

2. Построим линейное уравнение множественной регрессии в *натуральном масштабе* переменных:

$$\hat{y} = a_0 + a_1 \cdot x_1 + a_2 \cdot x_2.$$

Для нахождения коэффициентов  $a_1$  и  $a_2$  применим формулы (71), связывающие коэффициенты «чистой» регрессии и стандартизированные коэффициенты:

$$a_i = \beta_i \cdot \frac{\sigma_y}{\sigma_{x_i}} \quad (i=1,2).$$

Получаем:

$$a_1 = \beta_1 \cdot \frac{\sigma_y}{\sigma_{x_1}} = 0,684 \cdot \frac{35,75}{11,834} = 2,067,$$

$$a_2 = \beta_2 \cdot \frac{\sigma_y}{\sigma_{x_2}} = 0,630 \cdot \frac{35,75}{0,399} = 56,53.$$

Параметр  $a_0$  определим из уравнения (74):

$$a_0 = \overline{y} - a_1 \overline{x_1} - a_2 \overline{x_2}.$$

Получаем

$$a_0 = 117,98 - 2,067 \cdot 59,05 - 56,53 \cdot 1,987 = -116,4.$$

Таким образом, линейное уравнение множественной регрессии в *натуральном масштабе* переменных имеет вид:

$$\hat{y} = -116,40 + 2,07 \cdot x_1 + 56,53 \cdot x_2$$

или

$$Q(V, q) = -116,40 + 2,07 \cdot V + 56,53 \cdot q.$$

## 7. Частные коэффициенты корреляции

Наряду с линейными коэффициентами корреляции, характеризующими тесноту связи между переменными, во множественной регрессии широко используются *частные* показатели корреляции.

*Частные коэффициенты корреляции* характеризуют тесноту связи между результатом и соответствующим фактором при устранении влияния других факторов, включенных в уравнение регрессии.

Так, например, в модели с двумя факторами частный коэффициент корреляции  $r_{yx_1, x_2}$  (или  $r_{yx_1 \cdot x_2}$ ) характеризует тесноту связи между результатом  $y$  и фактором  $x_1$  при устранении влияния фактора  $x_2$ , а частный коэффициент корреляции  $r_{yx_2, x_1}$  (или  $r_{yx_2 \cdot x_1}$ ) характеризует тесноту связи между результатом  $y$  и фактором  $x_2$  при устранении влияния фактора  $x_1$ .

Частные коэффициенты корреляции вычисляют по формулам:

$$r_{yx_1, x_2} = \frac{r_{yx_1} - r_{yx_2} \cdot r_{x_1 x_2}}{\sqrt{(1 - r_{yx_2}^2)(1 - r_{x_1 x_2}^2)}}, \quad (81)$$

$$r_{yx_2, x_1} = \frac{r_{yx_2} - r_{yx_1} \cdot r_{x_1 x_2}}{\sqrt{(1 - r_{yx_1}^2)(1 - r_{x_1 x_2}^2)}}. \quad (82)$$

## 8. Совокупный коэффициент корреляции

Практическая значимость уравнения множественной регрессии оценивается с помощью показателя множественной корреляции и его квадрата – по-

казателя детерминации. Показатель множественной корреляции характеризует тесноту связи рассматриваемого набора факторов с исследуемым признаком или, иначе, оценивает тесноту совместного влияния факторов на результат.

При *линейной зависимости* признаков формула индекса множественной корреляции может быть представлена следующим выражением:

$$R_{yx_1x_2} = \sqrt{r_{yx_1} \cdot \beta_1 + r_{yx_2} \cdot \beta_2}. \quad (83)$$

Формула индекса множественной корреляции для *линейной регрессии* получила название *линейного коэффициента множественной корреляции*, или, что то же самое, *совокупного коэффициента корреляции*.

Зная частные коэффициенты корреляции, можно определить совокупный коэффициент корреляции по формуле

$$R_{yx_1x_2} = \sqrt{1 - (-r_{yx_1}^2) \cdot (-r_{yx_2, x_1}^2)} \quad (84)$$

или

$$R_{yx_1x_2} = \sqrt{1 - (-r_{yx_2}^2) \cdot (-r_{yx_1, x_2}^2)}. \quad (85)$$

Для линейной модели с двумя факторами совокупный коэффициент корреляции также может быть вычислен по формуле

$$R_{yx_1x_2} = \sqrt{\frac{r_{yx_1}^2 + r_{yx_2}^2 - 2 \cdot r_{yx_1} \cdot r_{yx_2} \cdot r_{x_1x_2}}{1 - r_{x_1x_2}^2}}. \quad (86).$$

## 9. Коэффициент множественной детерминации

Квадрат совокупного коэффициента корреляции  $R_{yx_1x_2}^2$  называется *коэффициентом множественной детерминации*.

Коэффициент множественной детерминации показывает, какая доля вариации результативного признака у объясняется факторами, включенными в уравнение множественной регрессии, а величина  $1 - R^2_{yx_1x_2}$  – долю влияния прочих, не учтенных моделью, факторов.

### **10. Пример вычисления частных коэффициентов корреляции, совокупного коэффициента корреляции, коэффициента множественной детерминации**

Используя данные предыдущего примера (пункт 6, стр. 73), вычислить:

- 1) частные коэффициенты корреляции;
- 2) совокупный коэффициент корреляции;
- 3) коэффициент множественной детерминации.

В предыдущем примере (стр. 75–76) были вычислены линейные коэффициенты парной корреляции:

$$r_{yx_1} = 0,775, \quad r_{yx_2} = 0,728, \quad r_{x_1x_2} = 0,144,$$

а также стандартизированные коэффициенты:

$$\beta_1 = 0,684, \quad \beta_2 = 0,630.$$

1. Вычислим частные коэффициенты корреляции, используя формулы (81)–(82):

$$r_{yx_1, x_2} = \frac{r_{yx_1} - r_{yx_2} \cdot r_{x_1x_2}}{\sqrt{(1 - r_{yx_2}^2)(1 - r_{x_1x_2}^2)}} = \frac{0,775 - 0,728 \cdot 0,144}{\sqrt{(1 - 0,728^2)(1 - 0,144^2)}} = 0,988,$$

$$r_{yx_2, x_1} = \frac{r_{yx_2} - r_{yx_1} \cdot r_{x_1x_2}}{\sqrt{(1 - r_{yx_1}^2)(1 - r_{x_1x_2}^2)}} = \frac{0,728 - 0,775 \cdot 0,144}{\sqrt{(1 - 0,775^2)(1 - 0,144^2)}} = 0,950.$$

Сравнивая частные коэффициенты корреляции, видим, что фактор  $x_1$  (скорость  $V$ ) имеет большую тесноту связи с результирующей функцией  $y$  (производительностью автомобиля  $W$ ), чем фактор  $x_2$  (масса перевозимого груза  $q$ ), а значит оказывает на нее большее влияние.

2. Вычислим совокупный коэффициент корреляции по формуле (83):

$$R_{yx_1x_2} = \sqrt{r_{yx_1}\beta_1 + r_{yx_2}\beta_2} = \sqrt{0,775 \cdot 0,684 + 0,728 \cdot 0,630} = \sqrt{0,989} = 0,994.$$

Вычислим теперь совокупный коэффициент корреляции, используя формулы (84):

$$R_{yx_1x_2} = \sqrt{1 - (r_{yx_1}^2 + r_{yx_2,x_1}^2 - 2 \cdot r_{yx_1} \cdot r_{yx_2,x_1})} = \sqrt{1 - (0,775^2 + 0,950^2 - 2 \cdot 0,775 \cdot 0,950)} = 0,994,$$

затем формулу (85):

$$R_{yx_1x_2} = \sqrt{1 - (r_{yx_2}^2 + r_{yx_1,x_2}^2 - 2 \cdot r_{yx_2} \cdot r_{yx_1,x_2})} = \sqrt{1 - (0,728^2 + 0,988^2 - 2 \cdot 0,728 \cdot 0,988)} = 0,994$$

и, наконец, формулу (86):

$$\begin{aligned} R_{yx_1x_2} &= \sqrt{\frac{r_{yx_1}^2 + r_{yx_2}^2 - 2 \cdot r_{yx_1} \cdot r_{yx_2} \cdot r_{x_1x_2}}{1 - r_{x_1x_2}^2}} = \\ &= \sqrt{\frac{0,775^2 + 0,728^2 - 2 \cdot 0,775 \cdot 0,728 \cdot 0,144}{1 - 0,144^2}} = 0,994. \end{aligned}$$

Очевидно, что результаты, полученные по формулам (83)–(86) – одинаковы. Поэтому вы можете использовать любую из них в своих расчетах.

3. Вычислим коэффициент множественной детерминации, возведя в квадрат совокупный коэффициент корреляции:

$$R_{yx_1x_2}^2 = 0,989.$$



Зависимость  $y$  от  $x_1$  и  $x_2$  характеризуется как весьма высокая, в которой 98,9% вариации производительности автомобиля  $W$  определяется вариацией учтенных в модели факторов: средней скорости и средней массы перевозимого груза.

Прочие факторы, не включенные в модель, составляют всего 1,1%.

## 11. Коэффициент эластичности

В научных исследованиях широкое применение находит также такой показатель, как *коэффициент эластичности*.

В *линейной множественной регрессии* для характеристики относительной силы влияния факторов  $x_1$  и  $x_2$  на результатив  $y$  рассчитываются *средние коэффициенты эластичности*:

$$\bar{\mathcal{E}}_{yx_i} = a_i \cdot \frac{\bar{x}_i}{\bar{y}}, \quad (i=1..n). \quad (87)$$

Величина среднего коэффициента эластичности  $\bar{\mathcal{E}}_{yx_i}$  показывает, на сколько процентов в среднем изменится результатив  $y$ , если значение фактора  $x_i$  изменится на 1%.

Причем, если коэффициент эластичности  $\bar{\mathcal{E}}_{yx_i} > 0$ , то с увеличением фактора  $x_i$  результатив  $y$  также увеличивается, а если  $\bar{\mathcal{E}}_{yx_i} < 0$ , то с увеличением фактора  $x_i$  результатив  $y$ , напротив, уменьшается.

## 12. Пример вычисления коэффициентов эластичности для множественной регрессии

Используя данные примера (пункт 6, стр. 73–76), вычислить средние коэффициенты эластичности  $\bar{\mathcal{E}}_{yx_1}$  и  $\bar{\mathcal{E}}_{yx_2}$ .

В таблице 17 (стр. 75) определены средние значения факторов  $x_1$  и  $x_2$  и результата  $y$ :

$$\overline{x_1} = 59,05, \quad \overline{x_2} = 1,987, \quad \overline{y} = 117,98.$$

На стр. 76 вычислены значения величин  $a_i$ :

$$a_1 = 2,067, \quad a_2 = 56,53.$$

Вычислим средние коэффициенты эластичности  $\overline{\mathcal{E}}_{yx_1}$  и  $\overline{\mathcal{E}}_{yx_2}$  по формулам (87):

$$\overline{\mathcal{E}}_{yx_1} = a_1 \cdot \frac{\overline{x_1}}{\overline{y}} = 2,067 \cdot \frac{59,05}{117,98} = 1,036,$$

$$\overline{\mathcal{E}}_{yx_2} = a_2 \cdot \frac{\overline{x_2}}{\overline{y}} = 56,53 \cdot \frac{1,987}{117,98} = 0,952.$$

Следовательно, с увеличением средней скорости  $V (x_1)$  на 1% от ее среднего уровня, средняя производительность автомобиля  $W$  увеличится на 1,036% (при неизменной массе перевозимого груза); а при увеличении средней массы перевозимого груза  $q (x_2)$  на 1%, средняя производительность автомобиля увеличивается на 0,952% от своего среднего уровня (при неизменной скорости).

Очевидно, что немного большее воздействие на среднюю производительность автомобиля  $W$  (то есть  $y$ ) оказывает размер средней скорости  $V (x_1)$ , что подтверждает ранее полученные результаты.

### 13. Оценка параметров множественной регрессии

#### *Оценка значимости уравнения множественной регрессии*

Значимость уравнения множественной регрессии в целом так же, как и в парной регрессии, оценивается с помощью  $F$ -критерия Фишера.

Согласно  $F$ -критерию Фишера, выдвигается «нулевая» гипотеза  $H_0$  о статистической незначимости уравнения множественной регрессии. Эта гипотеза отвергается при выполнении условия

$$F_R > F_{\text{крит}},$$

где  $F_{\text{крит}}$  определяется по таблицам  $F$ -критерия Фишера (приложение 4) при числах степеней свободы:

$$k_1 = m, \quad k_2 = n - m - 1$$

и заданному уровню значимости  $\alpha$ .

Здесь  $m$  – число параметров при переменных  $x$ ,  $n$  – число наблюдений.

Фактическое значение критерия  $F_R$  определяется по формуле

$$F_R = \frac{R_{yx_1x_2}^2}{1 - R_{yx_1x_2}^2} \cdot \frac{n - m - 1}{m}, \quad (88)$$

где  $R_{yx_1x_2}^2$  – коэффициент множественной детерминации;

$m$  – число параметров при переменных  $x$  (в линейной регрессии

совпадает с числом включенных в модель множественной регрессии факторов);

$n$  – число наблюдений.

Величина  $F_R$  сравнивается с критическим значением  $F_{\text{крит}}$ .

Если  $F_R > F_{\text{крит}}$ , то «нулевая» гипотеза о статистической незначимости уравнения регрессии отвергается, и уравнение регрессии признается статистически значимым с надежностью  $1 - \alpha$ .

В противном случае, если  $F_R < F_{\text{крит}}$ , принимается гипотеза о статистической незначимости уравнения регрессии.

### *Оценка значимости фактора, включенного в уравнение множественной регрессии*

Для множественной регрессии оценивается значимость не только уравнения в целом, но и каждого фактора, дополнительно включенного в регрессионную модель.

Необходимость такой оценки связана с тем, что не каждый фактор, вошедший в модель, может существенно увеличивать долю объясненной вариации результативного признака.

Кроме того, при наличии в модели нескольких факторов они могут вводиться в модель в разной последовательности.

Ввиду корреляции между факторами значимость одного и того же фактора может быть разной в зависимости от последовательности его введения в модель.

Мерой для оценки включения фактора в модель служит частный  $F$ -критерий, то есть  $F_{x_i}$ .

Частный  $F$ -критерий построен на сравнении прироста факторной дисперсии, обусловленного влиянием дополнительно включенного фактора, с остаточной дисперсией на одну степень свободы по регрессионной модели в целом.

Фактическое значение частного  $F$ -критерия сравнивается с табличным  $F_{крит}$  при уровне значимости  $\alpha$  и числе степеней свободы:

$$k_1 = 1, \quad k_2 = n - m - 1.$$

Если фактическое значение  $F_{x_i}$  превышает табличное

$$F_{x_i} > F_{крит},$$

то дополнительное включение фактора  $x_i$  в модель статистически оправдано и коэффициент «чистой» регрессии  $a_i$  при факторе  $x_i$  статистически значим.

Если же фактическое значение  $F_{x_i}$  меньше табличного

$$F_{x_i} < F_{\text{крит}},$$

то дополнительное включение в модель фактора  $x_i$  не увеличивает существенно долю объясненной вариации признака  $y$ , следовательно, нецелесообразно его включение в модель; коэффициент «чистой» регрессии  $a_i$  при данном факторе в этом случае статистически незначим.

Для двухфакторного уравнения частные  $F$ -критерии имеют вид:

$$F_{x_1} = \frac{R_{yx_1x_2}^2 - r_{yx_2}^2}{1 - R_{yx_1x_2}^2} \cdot \frac{n - m - 1}{1}, \quad (89)$$

$$F_{x_2} = \frac{R_{yx_1x_2}^2 - r_{yx_1}^2}{1 - R_{yx_1x_2}^2} \cdot \frac{n - m - 1}{1}, \quad (90)$$

где  $R_{yx_1x_2}^2$  – коэффициент множественной детерминации;

$r_{yx_i}^2$  – коэффициенты парной корреляции;

$n$  – число наблюдений;

$m$  – число факторов, включенных в модель множественной регрессии ( $m=2$ ).

С помощью частного  $F$ -критерия можно проверить значимость всех коэффициентов регрессии в предположении, что каждый соответствующий фактор  $x_i$  вводился в уравнение множественной регрессии последним.

#### **14. Пример оценки значимости уравнения множественной регрессии и факторов, включенных в уравнение**

Используя данные примера (пункт 6, стр. 73), оценить значимость построенного уравнения множественной регрессии и факторов, включенных в уравнение.

В примере (стр. 73–77) зависимость производительности автомобиля  $Q$  от скорости автомобиля  $V$  и его грузоподъемности  $q$  была описана уравнением линейной множественной регрессии:

$$\hat{y} = -116,40 + 2,07 \cdot x_1 + 56,53 \cdot x_2$$

или

$$Q(V, q) = -116,40 + 2,07 \cdot V + 56,53 \cdot q.$$

Здесь  $x_1$  – скорость автомобиля  $V$ ,  $x_2$  – грузоподъемности  $q$ .

Значимость уравнения множественной регрессии в целом оценим с помощью  $F$ -критерия Фишера.

Рассчитаем фактическое значение критерия по формуле (88):

$$F_R = \frac{R_{yx_1x_2}^2}{1 - R_{yx_1x_2}^2} \cdot \frac{n - m - 1}{m} = \frac{0,989}{1 - 0,989} \cdot \frac{20 - 2 - 1}{2} = 765,25,$$

где  $R_{yx_1x_2}^2$  – коэффициент множественной детерминации, вычисленный

в примере (пункт 10, стр. 79–80);

$n$  – число наблюдений;

$m$  – число факторов, включенных в модель.

Найденное фактическое значение  $F$ -критерия сравним с табличным (критическим) значением (приложение 4).

Табличное (критическое) значение  $F$ -критерия Фишера находится для заданных значений  $k_1$ ,  $k_2$  и  $\alpha$ , то есть  $F_{\text{крит}}(\alpha; k_1; k_2)$ ,

где  $k_1 = m = 2$  – число степеней свободы;

$k_2 = n - m - 1 = 20 - 2 - 1 = 17$  – число степеней свободы;

$\alpha$  – уровень значимости (допустимость ошибки).

Примем  $\alpha = 0,05$ , то есть 5%.

Тогда по таблице (приложение 4) определяем:

$$F_{\text{крит}}(\alpha; k_1; k_2) = F_{\text{крит}}(0,05; 2; 17) = 3,59.$$

Уравнение считается статистически значимым на уровне значимости  $1-\alpha$  в том случае, если фактическое значение  $F_R$  превышает критическое  $F_{крит}(\alpha; k1; k2)$ .

Получаем, что

$$F_R = 765,25 > F_{крит} = 3,59.$$

Следовательно, с вероятностью 0,95% делаем заключение о статистической значимости уравнения в целом и показателя тесноты связи  $R_{yx_1x_2}$ , которые сформировались под неслучайным воздействием факторов  $x_1$  и  $x_2$ .

Оценим значимость факторов, включенных в модель, поскольку каждый фактор, вошедший в модель, может существенно повлиять на качество модели.

Частные  $F$ -критерии  $F_{x_1}$  и  $F_{x_2}$  оценивают статистическую значимость присутствия факторов  $x_1$  и  $x_2$  в уравнении множественной регрессии, оценивают целесообразность включения в уравнение одного фактора после другого фактора, то есть  $F_{x_1}$  оценивает целесообразность включения в уравнение фактора  $x_1$  после того, как в него был включён фактор  $x_2$ . Соответственно  $F_{x_2}$  указывает на целесообразность включения в модель фактора  $x_2$  после фактора  $x_1$ .

Частные  $F$ -критерии  $F_{x_1}$  и  $F_{x_2}$  вычислим по формулам (89), (90):

$$F_{x_1} = \frac{R_{yx_1x_2}^2 - r_{yx_2}^2}{1 - R_{yx_1x_2}^2} \cdot \frac{n - m - 1}{1} = \frac{0,989 - 0,728^2}{1 - 0,989} \cdot \frac{20 - 2 - 1}{1} = 709,4,$$

$$F_{x_2} = \frac{R_{yx_1x_2}^2 - r_{yx_1}^2}{1 - R_{yx_1x_2}^2} \cdot \frac{n - m - 1}{1} = \frac{0,989 - 0,775^2}{1 - 0,989} \cdot \frac{20 - 2 - 1}{1} = 600,22.$$

Фактические значение  $F$ -критерия сравним с табличным при 5% уровне значимости ( $\alpha = 0,05$ ) и числах степеней свободы:

$$k1 = 1, \quad k2 = n - m - 1 = 20 - 2 - 1 = 17.$$

По таблице (приложение 4) определяем

$$F_{\text{крит}}(\alpha; k_1; k_2) = F_{\text{крит}}(0,05; 1; 17) = 4,45.$$

Получаем, что

$$F_{x_1} = 709,4 > F_{\text{крит}} = 4,45.$$

Следовательно, приходим к выводу о целесообразности включения в модель фактора  $x_1$  после того, как фактор  $x_2$  уже был включен, и коэффициент  $a_1$  при факторе  $x_1$  статистически значим.

Для фактора  $x_2$ :

$$F_{x_2} = 600,22 > F_{\text{крит}} = 4,45.$$

Поэтому дополнительное включение в модель фактора  $x_2$  после фактора  $x_1$  статистически оправдано, и коэффициент  $a_2$  при факторе  $x_2$  статистически значим.

### 15. Задания для самостоятельной работы

1. Получите у преподавателя исходные данные для множественной регрессии (приложение 1).
2. Постройте уравнение множественной линейной регрессии, используя МНК.
3. Вычислите парные коэффициенты корреляции. Интерпретируйте полученные значения.
4. Постройте уравнение множественной линейной регрессии, используя коэффициенты в стандартизированном масштабе.
5. Вычислите частные коэффициенты корреляции. Интерпретируйте полученные значения.
6. Вычислите совокупный коэффициент корреляции. Интерпретируйте полученное значение.



7. Вычислите коэффициент множественной детерминации. Интерпретируйте полученное значение.
8. Вычислите коэффициенты эластичности. Интерпретируйте полученные значения.
9. С помощью критерия Фишера оцените значимость уравнения множественной регрессии и каждого фактора, включенного в уравнение.
10. Составьте отчет по лабораторной работе.

## **16. Содержание отчета**

Отчет должен содержать:

1. Определения и формулы, изложенные в теоретической части лабораторной работы.
2. Исходные данные и результаты выполняемых заданий.

## **Контрольные вопросы**

1. Какие методы определения параметров уравнения множественной регрессии вы знаете?
2. Приведите формулы метода наименьших квадратов.
3. Объясните понятие «коэффициенты чистой регрессии».
4. Какое уравнение называют уравнением в натуральном масштабе переменных?
5. Приведите уравнение в стандартизированном масштабе.
6. Приведите формулы, связывающие «чистые» коэффициенты со стандартизированными.
7. Дайте определение и приведите формулы парных коэффициентов корреляции.
8. Что характеризуют парные коэффициенты корреляции и какие значения могут принимать?

9. Как вычисляются и что характеризуют частные коэффициенты корреляции?
10. Приведите формулы для вычисления совокупного коэффициента корреляции.
11. Дайте определение коэффициента множественной детерминации.
12. Что показывает коэффициент множественной детерминации?
13. Дайте определение коэффициента эластичности.
14. Как вычисляется средний коэффициент эластичности для линейной множественной регрессии?
15. Что показывает величина среднего коэффициента эластичности?
16. Что показывает знак среднего коэффициента эластичности?
17. Какой критерий применяют для оценки значимости уравнения множественной регрессии?
18. Приведите формулу вычисления фактического значения критерия Фишера.
19. Как определить критическое значение  $F$ -критерия Фишера?
20. При каком условии уравнение множественной регрессии статистически значимо?
21. Для чего применяют частный критерий Фишера?
22. Приведите формулы вычисления частных значений  $F$ -критерия Фишера  $F_{x_1}$  и  $F_{x_2}$  для уравнения с двумя факторами.

## ЛИТЕРАТУРА

1. Тимофеев, В.С. Эконометрика: учебник для академического бакалавриата / В.С. Тимофеев, А.В. Фадеев, В.Ю. Щеколдин. – 2-е изд., перераб. и доп. – М.: Юрайт, 2015. – 328 с.
2. Гмурман, В.Е. Теория вероятностей и математическая статистика: учебное пособие для бакалавров / В.Е. Гмурман. – 12-е изд. – М.: Юрайт, 2014. – 479 с.
3. Удинцова, Н.М. Эконометрика. Часть 1: Парная регрессия и корреляция в эконометрических исследованиях: учебное пособие / Н.М. Удинцова, Н.А. Коптева. – Зерноград: Азово-Черноморский инженерный институт ФГБОУ ВО Донской ГАУ, 2015. – 61 с.
4. Удинцова, Н.М. Эконометрика. Часть 2: Множественная регрессия и корреляция в эконометрических исследованиях: учебное пособие / Н.М. Удинцова, Н.А. Коптева. – Зерноград: Азово-Черноморский инженерный институт ФГБОУ ВО Донской ГАУ, 2016. – 73 с.
5. Калинина, В.Н. Теория вероятностей и математическая статистика: учебник для бакалавров / В.Н. Калинина. – 2-е изд., перераб. и доп. – М.: Юрайт, 2013. – 472 с.
6. Коптева, Н.А. Корреляционно-регрессионный и дисперсионный анализ: лабораторный практикум / Н.А. Коптева, Н.М. Удинцова, С.А. Коробской, И.П. Шульгина. – Зерноград: ФГБОУ ВПО АЧГАА, 2012. – 55 с.
7. Коптева, Н.А. Анализ и построение регрессионных моделей транспортных ситуаций. Решение транспортной задачи методами линейного программирования: учебно-методическое пособие к курсовой работе / Н.А. Коптева, Н.М. Удинцова, С.А. Коробской. – Зерноград: ФГБОУ ВПО АЧГАА, 2012. – 84 с.
8. Магнус, Я.Р. Эконометрика. Начальный курс: учебник / Я.Р. Магнус, П.Р. Катышев, А.А. Пересецкий. – 7-е изд., испр. – М.: Дело, 2005. – 504 с.
9. Эконометрика: учебник / И.И. Елисеева, С.В. Курышева, Т.В. Костева и др.; под ред. И.И. Елисеевой. – 2-е изд., перераб. и доп. – М.: Финансы и статистика, 2005. – 576 с.
10. Гусейнов, Ф.Г. Планирование эксперимента в задачах электроэнергетики / Ф.Г. Гусейнов, О.С. Мамедяров. – М.: Энергоатомиздат, 1988. – 151 с.

# **ПРИЛОЖЕНИЕ 1**

## **ВАРИАНТЫ ЗАДАНИЙ ДЛЯ САМОСТОЯТЕЛЬНОЙ РАБОТЫ**

<i>Вариант 1</i>				<i>Вариант 2</i>			<i>Вариант 3</i>			<i>Вариант 4</i>		
<i>№</i>	<i>Y</i>	<i>X1</i>	<i>X2</i>	<i>Y</i>	<i>X1</i>	<i>X2</i>	<i>Y</i>	<i>X1</i>	<i>X2</i>	<i>Y</i>	<i>X1</i>	<i>X2</i>
<i>1</i>	13,9	37,8	65,52	13,4	1,22	3,44	6,16	0,42	2,14	7,42	2,16	6,78
<i>2</i>	23,9	52,92	79,38	15,8	0,34	4,51	6,77	0,54	3,16	4,12	3,45	5,41
<i>3</i>	26,5	64,26	69,3	14,2	0,61	2,13	8,96	0,86	0,87	3,59	2,96	5,21
<i>4</i>	22,7	60,48	80,64	16,8	0,48	3,14	2,46	0,45	0,64	4,35	3,15	6,48
<i>5</i>	27,7	53,93	64,26	15,2	0,92	2,17	3,71	0,52	1,54	6,54	1,98	2,15
<i>6</i>	24,1	60,48	84,42	13,7	1,14	6,14	5,77	0,49	2,87	2,69	4,55	3,45
<i>7</i>	29	69,3	66,78	16,4	1,58	8,72	4,92	0,42	3,15	6,17	1,88	6,21
<i>8</i>	29	54,18	81,9	14,5	1,61	3,44	3,76	0,54	4,52	3,11	2,78	3,24
<i>9</i>	36,5	66,02	88,2	11,3	2,01	4,59	8,51	0,48	7,81	3,41	2,41	1,97
<i>10</i>	34	56,7	76,86	10,8	2,02	7,21	7,98	0,82	2,13	4,51	2,67	5,42
<i>11</i>	32,8	70,56	68,04	9,05	2,11	8,01	5,36	0,67	4,88	5,82	3,54	4,25
<i>12</i>	37,8	57,96	75,6	8,14	2,15	6,51	4,46	0,34	5,24	5,79	1,69	6,12
<i>13</i>	30,2	66,78	70,56	17,8	0,14	4,21	8,71	0,59	6,75	1,69	4,02	7,02
<i>14</i>	27,7	75,6	86,94	19,1	0,19	3,45	9,54	0,73	4,33	4,22	3,61	8,21
<i>15</i>	31,5	71,82	73,08	12,1	1,17	2,15	6,14	0,64	2,36	4,17	3,55	9,23

<i>Вариант 5</i>				<i>Вариант 6</i>			<i>Вариант 7</i>			<i>Вариант 8</i>		
<i>№</i>	<i>Y</i>	<i>X1</i>	<i>X2</i>	<i>Y</i>	<i>X1</i>	<i>X2</i>	<i>Y</i>	<i>X1</i>	<i>X2</i>	<i>Y</i>	<i>X1</i>	<i>X2</i>
<i>1</i>	9,01	4,59	2,35	10,7	0,21	4,98	21,05	1,11	0,54	22,1	2,31	0,31
<i>2</i>	7,51	6,17	3,24	9,01	0,64	6,38	28,31	1,39	0,98	34,6	5,61	0,98
<i>3</i>	6,38	5,31	5,61	8,97	0,97	2,35	40,5	2,84	0,36	29,4	4,86	0,14
<i>4</i>	9,48	2,8	2,31	9,28	0,45	4,68	38,14	5,91	1,25	68,1	6,45	0,65
<i>5</i>	10,5	3,16	8,64	4,56	1,31	5,26	41,25	4,38	0,57	98,2	10,35	0,87
<i>6</i>	8,29	5,23	5,21	5,98	1,44	7,65	65,34	9,75	0,98	34,5	4,84	0,25
<i>7</i>	8,74	4,1	6,45	9,87	0,57	8,94	57,31	6,28	0,45	28,1	3,64	0,64
<i>8</i>	8,61	6,02	7,36	6,87	0,87	1,23	80,14	7,98	1,2	20,4	2,01	0,12
<i>9</i>	8,35	3,23	9,01	7,98	0,35	7,56	44,21	5,21	1,35	76,4	9,12	0,39
<i>10</i>	9,51	3,87	5,21	8,59	0,76	8,64	30,21	3,64	1,71	88,1	10,54	0,54
<i>11</i>	4,83	5,02	3,65	7,89	0,69	4,25	29,43	1,59	0,45	79,9	12,02	0,89
<i>12</i>	10	4,11	4,56	5,39	0,64	8,32	46,21	2,78	1,68	19,3	2,05	0,78
<i>13</i>	6,72	4,11	8,95	3,84	1,09	4,21	20,46	0,98	1,36	57,3	4,39	0,65
<i>14</i>	12,35	1,86	4,32	2,91	1,76	6,38	60,15	4,56	4,57	68,3	5,27	0,26
<i>15</i>	9,21	5,01	7,23	4,89	1,55	8,21	50,21	5,02	0,98	79,5	8,12	0,98

<i>Вариант 9</i>				<i>Вариант 10</i>			<i>Вариант 11</i>			<i>Вариант 12</i>		
<i>№</i>	<i>Y</i>	<i>X1</i>	<i>X2</i>	<i>Y</i>	<i>X1</i>	<i>X2</i>	<i>Y</i>	<i>X1</i>	<i>X2</i>	<i>Y</i>	<i>X1</i>	<i>X2</i>
<i>1</i>	27	1,01	4,32	28,1	16,02	1,25	29,02	9,56	0,32	30,08	3,54	4,57
<i>2</i>	24,4	1,58	5,14	23	14,53	2,54	20,81	16,08	0,15	31,22	8,14	5,14
<i>3</i>	19,5	1,99	8,13	25,8	15,33	3,61	25,64	14,06	0,64	30,71	5,61	3,78
<i>4</i>	24,3	1,64	4,03	27,9	17,7	4,21	26,14	14,87	0,37	30,79	6,79	5,14
<i>5</i>	20,7	1,91	9,21	26,5	15,37	2,01	25,67	14,8	0,18	31,42	7,56	6,24
<i>6</i>	23,5	1,67	8,54	29,2	17,56	9,54	24,89	15,06	0,15	39,01	10,12	8,12
<i>7</i>	21,5	1,87	7,13	26,4	16,44	4,25	23,59	15,78	0,96	32,41	9,41	9,23
<i>8</i>	22,1	1,68	0,35	30,1	23,5	7,52	26,71	13,54	0,42	33,04	8,53	0,45
<i>9</i>	21,7	1,73	4,28	32	23,41	2,01	26,72	13,61	0,57	33,89	9,57	0,54
<i>10</i>	26	1,22	6,74	29,8	22,05	3,02	27,02	12,78	0,12	35,87	8,97	1,23
<i>11</i>	25,9	1,34	1,25	28,8	21,53	6,12	24,11	13,47	0,64	36,12	10,08	8,14
<i>12</i>	25,7	1,45	6,04	28,6	22,47	8,1	28,64	11,15	0,28	37,14	9,75	6,12
<i>13</i>	25	1,54	8,25	29,3	20,54	7,02	28,73	11,06	0,91	38,61	11,24	4,21
<i>14</i>	25	1,31	9,1	27	19,44	6,14	29,64	8,59	0,18	37,52	10,18	5,03
<i>15</i>	25,3	1,46	5,11	29,2	18,16	0,54	29,15	10,37	0,24	38,42	9,67	4,57

<i>Вариант 13</i>				<i>Вариант 14</i>			<i>Вариант 15</i>			<i>Вариант 16</i>		
<i>№</i>	<i>Y</i>	<i>X1</i>	<i>X2</i>	<i>Y</i>	<i>X1</i>	<i>X2</i>	<i>Y</i>	<i>X1</i>	<i>X2</i>	<i>Y</i>	<i>X1</i>	<i>X2</i>
<i>1</i>	5,02	0,24	1,54	6,12	0,48	6,54	7,58	1,02	5,64	8,16	2,57	4,95
<i>2</i>	6,01	0,34	2,94	3,4	0,21	4,82	6,91	1,65	2,15	9,1	2,69	1,25
<i>3</i>	4,97	0,65	4,38	5,31	0,64	4,12	4,38	1,74	7,91	5,2	1,68	8,34
<i>4</i>	3,51	0,12	5,65	3,72	0,31	5,01	7,59	1,69	5,24	8,99	5,15	5,03
<i>5</i>	4,38	0,46	2,31	4,65	0,15	9,22	8,64	1,35	7,96	8,24	2,98	4,31
<i>6</i>	6,2	0,31	4,68	6,42	0,82	4,21	5,29	0,98	1,35	12,3	4,46	8,9
<i>7</i>	6,73	0,76	5,24	3,98	0,71	8,47	10,11	0,92	4,82	3,73	1,98	7,21
<i>8</i>	5,44	0,64	2,64	8,1	0,93	2,35	11,02	0,47	8,21	9,13	2,06	6,81
<i>9</i>	4,89	0,11	4,9	2,99	0,11	6,48	7,61	1,54	4,36	10,45	2,78	6,33
<i>10</i>	6,04	0,56	6,14	3,4	0,13	9,41	4,91	1,67	7,54	6,53	3,46	4,63
<i>11</i>	5,78	0,49	7,25	6,57	0,24	2,17	5,64	1,35	9,31	10,8	4,58	9,14
<i>12</i>	5,12	0,32	1,32	5,96	0,41	3,54	7,86	0,87	1,25	11,88	3,78	7,21
<i>13</i>	4,96	0,46	7,95	7,03	0,51	9,25	6,31	1,43	1,84	12,64	5,97	2,13
<i>14</i>	3,75	0,21	2,35	6,81	0,81	8,1	5,84	1,15	2,64	5,68	2,49	4,03
<i>15</i>	5,11	0,17	4,65	5,99	0,29	5,31	9,62	0,69	9,01	11,88	7,02	7,09

<i>Вариант 17</i>				<i>Вариант 18</i>			<i>Вариант 19</i>			<i>Вариант 20</i>		
<i>№</i>	<i>Y</i>	<i>X1</i>	<i>X2</i>	<i>Y</i>	<i>X1</i>	<i>X2</i>	<i>Y</i>	<i>X1</i>	<i>X2</i>	<i>Y</i>	<i>X1</i>	<i>X2</i>
<i>1</i>	23,9	3,05	0,54	24,2	4,68	0,65	25,61	2,16	6,15	26,4	16,4	0,14
<i>2</i>	68,3	9,64	0,39	25,9	4,01	0,47	24,78	4,58	1,26	25,1	17,8	2,64
<i>3</i>	54,2	4,87	0,25	19,8	6,97	0,98	21,45	8,01	5,34	21,8	13,4	1,35
<i>4</i>	99,8	13,57	0,48	20,7	5,87	0,14	17,8	7,45	7,84	20,7	10,2	8,91
<i>5</i>	77,1	10,8	0,75	15,4	10,9	0,56	20,58	9,34	0,58	21,5	10,5	2,14
<i>6</i>	65,2	5,57	0,39	12,1	14,3	0,45	22,37	6,48	1,56	20,4	11,6	3,21
<i>7</i>	87,2	6,04	0,57	10,5	11,9	0,75	25	6,89	7,15	22,3	12,1	4,75
<i>8</i>	89,1	8,64	0,21	9,7	10,7	0,35	15	9,98	9,32	21,5	12,8	6,45
<i>9</i>	46,3	3,47	0,87	15,3	8,91	0,54	22,06	5,71	5,12	27,8	19,1	2,14
<i>10</i>	78,6	9,57	0,64	16,8	7,1	0,98	27,51	4,32	4,32	26,5	18,4	5,64
<i>11</i>	95,3	12,8	0,12	17,9	6,9	0,12	27,95	0,51	8,95	28,9	17,3	7,82
<i>12</i>	99,9	11,34	0,56	13,8	8,3	0,21	25,17	1,67	4,25	25,4	17,6	6,54
<i>13</i>	87,9	10,5	0,45	17,9	9,4	0,54	23,16	3,04	1,29	24,3	15,8	9,21
<i>14</i>	67,9	8,64	0,36	15,3	6,7	0,75	26,31	2,89	2,37	23,5	14,7	0,59
<i>15</i>	89,7	7,54	0,18	21,9	5,9	0,64	24,58	3,09	1,02	25,1	16,7	4,31

<i>Вариант 21</i>				<i>Вариант 22</i>			<i>Вариант 23</i>			<i>Вариант 24</i>		
<i>№</i>	<i>Y</i>	<i>X1</i>	<i>X2</i>	<i>Y</i>	<i>X1</i>	<i>X2</i>	<i>Y</i>	<i>X1</i>	<i>X2</i>	<i>Y</i>	<i>X1</i>	<i>X2</i>
<i>1</i>	31,47	11,54	1,41	32,15	16,1	0,21	33,07	7,64	4,51	34,01	0,12	5,41
<i>2</i>	30,46	10,27	2,04	41,51	9,8	0,84	34,12	7,05	1,21	43,51	7,48	6,34
<i>3</i>	30,78	11,51	3,45	32,71	15,7	1,11	33,21	5,74	6,31	42,38	6,54	1,28
<i>4</i>	39,45	17,84	4,02	40,78	10	1,54	34,02	4,42	2,15	35,21	1,78	7,36
<i>5</i>	38,12	16,73	2,1	40,17	10,4	0,97	35,74	4,11	4,1	34,87	0,67	5,12
<i>6</i>	37,99	17,41	5,41	39,57	10,8	1,23	36,24	3,91	2,11	41,37	5,98	4,67
<i>7</i>	36,75	16,45	6,41	34,12	15,1	0,84	36,98	2,97	7,51	36,51	1,28	1,25
<i>8</i>	36,12	15,42	8,01	35,76	14,3	0,97	36,41	2,41	8,12	34,61	2,06	5,31
<i>9</i>	35,23	14,64	7,31	36,21	14	1,87	37,15	1,97	6,31	37,21	2,78	4,91
<i>10</i>	32,15	12,47	1,23	36,72	14	1,12	42,78	0,54	5,32	37,05	3,08	1,08
<i>11</i>	33,41	13,57	2,45	37,05	12,8	1,36	38,02	1,74	9,01	37,77	3,17	1,12
<i>12</i>	34,11	12,91	9,12	37,51	12,6	0,94	39,47	1,62	5,03	38,01	3,54	7,45
<i>13</i>	35,47	12,89	4,12	37,46	12,8	0,57	38,14	0,94	4,21	38,98	4,61	3,12
<i>14</i>	33,78	13,01	5,61	38,04	12	1,02	40,54	1,21	7,11	42,71	5,55	2,54
<i>15</i>	36,08	14,01	7,84	38,97	10,6	1,15	41,81	0,63	6,81	41,15	6,41	1,91

**ПРИЛОЖЕНИЕ 2****Шкала Чеддока (характеристика связи)**

<i><b>Показатели</b></i>	<i><b>Характеристика</b></i>
<i><b>0,1 – 0,3</b></i>	слабая связь
<i><b>0,3 – 0,5</b></i>	умеренная связь
<i><b>0,5 – 0,7</b></i>	заметная связь
<i><b>0,7 – 0,9</b></i>	высокая связь
<i><b>0,9 – 0,99</b></i>	весьма высокая связь

### ПРИЛОЖЕНИЕ 3

**Таблица значений  $t$ -критерия Стьюдента  
на уровнях значимости 0,10; 0,05; 0,01**

k	$\alpha$			k	$\alpha$		
	0,10	0,05	0,01		0,10	0,05	0,01
1	6,3138	12,706	63,657	18	1,7341	2,1009	2,8784
2	2,9200	4,3027	9,9248	19	1,7291	2,0930	2,8609
3	2,3534	3,1825	5,8409	20	1,7247	2,0860	2,8453
4	2,1318	2,7764	4,6041	21	1,7207	2,0796	2,8314
5	2,0150	2,5706	4,0321	22	1,7171	2,0739	2,8188
6	1,9432	2,4469	3,7074	23	1,7139	2,0687	2,8073
7	1,8946	2,3646	3,4995	24	1,7109	2,0639	2,7969
8	1,8595	2,3060	3,3554	25	1,7081	2,0595	2,7874
9	1,8331	2,2622	3,2498	26	1,7056	2,0555	2,7787
10	1,8125	2,2281	3,1693	27	1,7033	2,0518	2,7707
11	1,7959	2,2010	3,1058	28	1,7011	2,0484	2,7633
12	1,7823	2,1788	3,0545	29	1,6991	2,0452	2,7564
13	1,7709	2,1604	3,0123	30	1,6973	2,0423	2,7500
14	1,7613	2,1448	2,9768	40	1,6839	2,0211	2,7045
15	1,7530	2,1315	2,9467	60	1,6707	2,0003	2,6603
16	1,7459	2,1199	2,9208	120	1,6577	1,9799	2,6174
17	1,7396	2,1098	2,8982	$\infty$	1,6449	1,9600	2,5758



## ПРИЛОЖЕНИЕ 4

**Таблица значений  $F$ -критерия Фишера  
на уровне значимости  $\alpha = 0,05$**

$k_1 \backslash k_2$	1	2	3	4	5	6	8	12	24	$\infty$
1	161,45	199,50	215,72	224,57	230,17	233,97	238,89	243,91	249,04	254,32
2	18,51	19,00	19,16	19,25	19,30	19,33	19,37	19,41	19,45	19,50
3	10,13	9,55	9,28	9,12	9,01	8,94	8,84	8,74	8,64	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,04	5,91	5,77	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,82	4,68	4,53	4,36
6	5,99	5,14	4,76	4,53	4,39	4,28	4,15	4,00	3,84	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,73	3,57	3,41	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,44	3,28	3,12	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,23	3,07	2,90	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,07	2,91	2,74	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	2,95	2,79	2,61	2,40
12	4,75	3,88	3,49	3,26	3,11	3,00	2,85	2,69	2,50	2,30
13	4,67	3,80	3,41	3,18	3,02	2,92	2,77	2,60	2,42	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,70	2,53	2,35	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,64	2,48	2,29	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,59	2,42	2,24	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,55	2,38	2,19	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,51	2,34	2,15	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,48	2,31	2,11	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,45	2,28	2,08	1,84
21	4,32	3,47	3,07	2,84	2,68	2,57	2,42	2,25	2,05	1,81
22	4,30	3,44	3,05	2,82	2,66	2,55	2,40	2,23	2,03	1,78
23	4,28	3,42	3,03	2,80	2,64	2,53	2,38	2,20	2,00	1,76
24	4,26	3,40	3,01	2,78	2,62	2,51	2,36	2,18	1,98	1,73
25	4,24	3,38	2,99	2,76	2,60	2,49	2,34	2,16	1,96	1,71
26	4,22	3,37	2,98	2,74	2,59	2,47	2,32	2,15	1,95	1,69
27	4,21	3,35	2,96	2,73	2,57	2,46	2,30	2,13	1,93	1,67
28	4,20	3,34	2,95	2,71	2,56	2,44	2,29	2,12	1,91	1,65
29	4,18	3,33	2,93	2,70	2,54	2,43	2,28	2,10	1,90	1,64
30	4,17	3,32	2,92	2,69	2,53	2,42	2,27	2,09	1,89	1,62
35	4,12	3,26	2,87	2,64	2,48	2,37	2,22	2,04	1,83	1,57
40	4,08	3,23	2,84	2,61	2,45	2,34	2,18	2,00	1,79	1,51
45	4,06	3,21	2,81	2,58	2,42	2,31	2,15	1,97	1,76	1,48
50	4,03	3,18	2,79	2,56	2,40	2,29	2,13	1,95	1,74	1,44
60	4,00	3,15	2,76	2,52	2,37	2,25	2,10	1,92	1,70	1,39
70	3,98	3,13	2,74	2,50	2,35	2,23	2,07	1,89	1,67	1,35
80	3,96	3,11	2,72	2,49	2,33	2,21	2,06	1,88	1,65	1,31
90	3,95	3,10	2,71	2,47	2,32	2,20	2,04	1,86	1,64	1,28
100	3,94	3,09	2,70	2,46	2,30	2,19	2,03	1,85	1,63	1,26
125	3,92	3,07	2,68	2,44	2,29	2,17	2,01	1,83	1,60	1,21
150	3,90	3,06	2,66	2,43	2,27	2,16	2,00	1,82	1,59	1,18
200	3,89	3,04	2,65	2,42	2,26	2,14	1,98	1,80	1,57	1,14
300	3,87	3,03	2,64	2,41	2,25	2,13	1,97	1,79	1,55	1,10
400	3,86	3,02	2,63	2,40	2,24	2,12	1,96	1,78	1,54	1,07
500	3,86	3,01	2,62	2,39	2,23	2,11	1,96	1,77	1,54	1,06
1000	3,85	3,00	2,61	2,38	2,22	2,10	1,95	1,76	1,53	1,03
$\infty$	3,84	2,99	2,60	2,37	2,21	2,09	1,94	1,75	1,52	1,00

## ПРИЛОЖЕНИЕ 5

### ПРИМЕРНЫЙ ВАРИАНТ КОНТРОЛЬНОЙ РАБОТЫ

Изучается зависимость потребления электроэнергии (тыс. кВт/ч) от производства продукции (тыс. ед.) и уровня механизации труда (%). Для этого по 10 производственным компаниям, выпускающим одноименную продукцию, были получены следующие данные:

Потребление электроэнергии, $y$	120	130	185	195	200	260	400	450	470	370
Производство продукции, $x_1$	12	16	19	22	23	26	38	41	45	39
Уровень механизации труда, $x_2$	19	25	28	30	35	47	58	60	65	53

1. Составить уравнение линейной зависимости потребления электроэнергии от производства продукции.
2. Найти ошибку аппроксимации.
3. Найти коэффициент эластичности и пояснить его значение.
4. Рассчитать коэффициент линейной корреляции и оценить тесноту связи между переменными по шкале Чеддока.
5. Найти индекс детерминации и объяснить его значение.
6. Проверить статистическую значимость коэффициентов с помощью критерия Стьюдента.
7. Составить уравнение гиперболической зависимости потребления электроэнергии от уровня механизации труда.
8. Найти ошибку аппроксимации.
9. Найти коэффициент эластичности и пояснить его значение.
10. Найти общую, факторную и остаточную дисперсии и проверить правило сложения дисперсий.
11. Найти индекс детерминации (объяснить его значение).
12. Рассчитать индекс корреляции и оценить тесноту связи между переменными по шкале Чеддока.

13. Проверить статистическую значимость индекса корреляции с помощью критерия Фишера.
14. Построить уравнение множественной регрессии зависимости потребления электроэнергии от производства продукции и от уровня механизации труда.
15. Найти погрешность аппроксимации, сделать вывод.
16. Найти коэффициенты эластичности и объяснить их смысл.
17. Рассчитать коэффициенты парной и частной корреляции и объяснить их смысл.
18. Рассчитать коэффициенты частной детерминации, объяснить их смысл.
19. Найти совокупный коэффициент множественной корреляции  $R$ , объяснить его смысл.
20. Найти коэффициент множественной детерминации  $R^2$  и объяснить его смысл.
21. Оценить статистическую значимость уравнения множественной регрессии.
22. Оценить статистическую значимость каждого фактора  $x_1, x_2$ .

**Коптева Нина Алексеевна**  
доктор технических наук, профессор

**Удинцова Надежда Михайловна**  
кандидат технических наук, доцент

**ДОПОЛНИТЕЛЬНЫЕ ГЛАВЫ МАТЕМАТИКИ**  
**Часть 1:**  
**Корреляционно-регрессионный анализ**  
**в научных исследованиях**

*Учебное пособие*

Редактор Н.П. Лучинкина  
Верстка Г.С. Кудрявцева  
Дизайн обложки С.П. Вдовикина

Подписано в печать 23.05.2018 г.  
Формат 60×84/16. Усл. п. л. 5,81. Тираж 20 экз. Заказ № 31.

Отдел информационных технологий и издательской деятельности  
Азово-Черноморского инженерного института – филиала  
ФГБОУ ВО Донской ГАУ  
347740, г. Зерноград Ростовской области, ул. Советская, 15.