УЧЕТНО-СТАТИСТИЧЕСКИЙ ФАКУЛЬТЕТ

ФЕДЕРАЛЬНОЕ АГЕНТСТВО ПО ОБРАЗОВАНИЮ

ГОСУДАРСТВЕННОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ

ВСЕРОССИЙСКИЙ ЗАОЧНЫЙ ФИНАНСОВО-ЭКОНОМИЧЕСКИЙ
ИНСТИТУТ

СТАТИСТИКА

Методические указания
по выполнению лабораторной работы
№ 2

Для самостоятельной работы студентов III курса
всех специальностей (первое и второе высшее образование)

МОСКВА 2006

ФЕДЕРАЛЬНОЕ АГЕНТСТВО ПО ОБРАЗОВАНИЮ

ГОСУДАРСТВЕННОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ

**ВСЕРОССИЙСКИЙ ЗАОЧНЫЙ
ФИНАНСОВО-ЭКОНОМИЧЕСКИЙ ИНСТИТУТ**

СТАТИСТИКА

**Методические указания
по выполнению лабораторной работы № 2**

Автоматизированный корреляционно-регрессионный анализ взаимосвязи статистических данных в среде *MS Excel*

Для самостоятельной работы студентов III курса всех специальностей
(первое и второе высшее образование)

**Москва
ВУЗОВСКИЙ УЧЕБНИК
2006**

ББК 65.290-93

Методические указания по выполнению лабораторной работы № 2 «Автоматизированный корреляционно-регрессионный анализ взаимосвязи статистических данных в среде *MS Excel*» подготовили:
д-р физ.-мат. наук, проф. ***Г.П. Кожевникова***;
канд. техн. наук, доц. ***А.В. Голикова***

Ответственный редактор проф. ***Г.П. Кожевникова***

Методические указания по выполнению лабораторной работы № 2 «Автоматизированный корреляционно-регрессионный анализ взаимосвязи статистических данных в среде *MS Excel*» одобрены на заседании Научно-методического совета ВЗФЭИ

Проректор по УМР, председатель НМС профессор ***Д.М. Дайитбегов***

Статистика. Компьютерные лабораторные работы: Методические указания к лабораторной работе № 2 «Автоматизированный корреляционно-регрессионный анализ взаимосвязи статистических данных в среде *MS Excel*». — М.: Вузовский учебник, 2006. — 70 с.

ББК 65.290-93

Лабораторная работа № 2 Автоматизированный корреляционно-регрессионный анализ взаимосвязи статистических данных в среде *MS Excel*

I. Цели, содержание и организация выполнения лабораторной работы

1. Цель и задачи работы

Цель работы — освоение методики корреляционно-регрессионного анализа взаимосвязи социально-экономических явлений с применением компьютерных средств.

Изучение взаимосвязей явлений и процессов — одна из важнейших задач статистических исследований.

Методы и приемы корреляционно-регрессионного анализа позволяют:

- выявить наличие корреляционной связи признаков (показателей) и оценить ее тесноту;
- найти аналитическое выражение связи в виде уравнения регрессии;
- оценить качество найденной модели связи.

Табличный процессор *Microsoft Excel* и его надстройка **Пакет анализа** предоставляют ряд программных средств для автоматизированного решения вышеперечисленных трех задач.

2. Компьютерные средства

Операционная среда:	*Windows* 2000/*XP*.
Программное средство:	Пакет программ для работы с электронными таблицами *Microsoft Excel*.
Версии *MS Excel*:	*Excel* 97, *Excel* 2000.
Технологическая среда:	Программная надстройка *MS Excel* **Пакет анализа** и библиотека из 78 статистических функций, встроенных в *Excel*.

3. Требования к уровню подготовленности студента к лабораторной работе

Лабораторная работа проводится после изучения студентами следующих тем курса статистики: *статистическое наблюдение, сводка и группировка статистических данных, ряды распределения, средние величины и показатели вариации, выборочный метод, статистическое изучение взаимосвязей явлений и процессов.*

Для выполнения работы студент ***должен знать***:

- цель и содержание работы, порядок ее выполнения и отчетности (*раздел I «Методических указаний»);*
- статистическую сущность задач и методику корреляционно-регрессионного анализа статистических данных *(раздел II «Методических указаний»)*
- основные теоретические положения выполняемых заданий (*из раздела III «Методических указаний» пункты «Краткие теоретические сведения к заданиям»).*

Студент ***должен обладать навыками*** работы в среде *Microsoft Excel*:

- строить электронные таблицы;
- составлять и копировать расчетные формулы;
- использовать статистические и математические функции инструмента **Мастер функций**;
- строить статистические графики с использованием инструмента **Мастер диаграмм**;
- использовать инструменты **Корреляция** и **Регрессия** надстройки **Пакет анализа.**

Перед выполнением лабораторной работы студенту следует ***ознакомиться с технологией выполнения каждого задания.***

4. Содержание и структура лабораторной работы

4.1. Постановка задачи

Корреляционно-регрессионный анализ (КР-анализ) взаимосвязи признаков является составной частью проводимого статистического исследования двух экономических показателей статистической совокупности 32 предприятий и частично использует результаты Лабораторной работы № 1.

В Лабораторной работе № 2 изучается взаимосвязь между факторным признаком *Среднегодовая стоимость основных производственных фондов* (признак *X*) и результативным признаком *Выпуск продукции* (признак *Y*), значениями которых являются исходные данные Лабораторной работы № 1 после исключения из них аномальных значений.

В процессе статистического исследования необходимо решить ряд задач.

1. Установить наличие ***статистической связи*** между факторным признаком *X* и результативным признаком *Y*:

а) графическим методом;

б) методом сопоставления параллельных рядов.

2. Установить наличие ***корреляционной связи*** между признаками *X* и *Y* методом аналитической группировки.

3. Оценить тесноту связи признаков *X* и *Y* на основе:

а) эмпирического корреляционного отношения η;

б) линейного коэффициента корреляции *r*.

4. Построить однофакторную линейную регрессионную модель связи признаков *X* и *Y*, используя инструмент **Регрессия** надстройки **Пакет анализа.**

5. Оценить адекватность и практическую пригодность построенной линейной регрессионной модели, указав:

а) доверительные интервалы коэффициентов a_0, a_1;

б) степень тесноты связи признаков *X* и *Y*;

в) погрешность регрессионной модели.

6. Дать экономическую интерпретацию:

а) коэффициента регрессии a_1;

б) коэффициента эластичности $K_э$;

в) остаточных величин ε_i.

7. Найти наиболее адекватное нелинейное уравнение регрессии с помощью средств инструмента **Мастер диаграмм.** Построить для этого уравнения теоретическую кривую регрессии.

4.2. Структура лабораторной работы

Лабораторная работа состоит из трех этапов — подготовительного, расчетного и заключительного.

На ***подготовительном этапе*** формируется индивидуальная рабочая среда проведения вычислений по исходным данным варианта.

На ***расчетном этапе*** выполняются три задания.

Задание 1. Построение аналитической группировки для выявления корреляционной зависимости результативного признака от факторного и оценка тесноты взаимосвязи этих признаков.

Задание 2. Построение однофакторной линейной регрессионной модели связи изучаемых признаков с помощью инструмента **Регрессия** надстройки **Пакет анализа**.

Задание 3. Построение однофакторных нелинейных регрессионных моделей связи признаков с помощью инструмента **Мастер диаграмм** и выбор наиболее адекватного уравнения регрессии.

Каждое задание имеет следующую структуру:

1. Краткие теоретические сведения;
2. Технология выполнения задания;
3. Алгоритмы выполнения задания.

Краткие теоретические сведения необходимы для понимания студентом статистической сущности задания.

В ***технологической части*** излагаются особенности применения инструментов **Пакет анализа, Мастер диаграмм** и других средств *Excel* при автоматизированном решении статистических задач, указанных в заданиях.

В ***алгоритмической части*** представлены алгоритмы действий в среде *Excel*, выполнение которых реализует технологические процессы решения статистических задач.

На ***заключительном (аналитическом) этапе*** анализируются полученные статистические показатели, таблицы и графики, делаются выводы о виде и тесноте взаимосвязи признаков *X* и *Y*, анализируются построенные модели взаимосвязи, выполняется экономическая интерпретация параметров модели.

В методических указаниях к выполнению заданий используются три вида таблиц:

- результативные таблицы рассматриваемых показателей (макеты таблиц приведены в Приложении 2.2);
- результативные таблицы демонстрационного примера *«Методических указаний»*;
- таблицы собственно *«Методических указаний»*.

Во избежание коллизий при ссылке на различные виды таблиц к номерам таблиц второго и третьего вида добавляются соответственно идентификаторы «ДП» (демонстрационный пример) и «М» (методические указания).

4.3. Отчетность по работе

По результатам выполнения лабораторной работы студент представляет отчет. Отчет должен содержать следующие разделы:

1. Титульный лист (образец дан в Приложении 2.1, электронная копия — в файле **Формат отчета.doc**);
2. Постановка задачи корреляционно-регрессионного анализа, включая исходные данные варианта (электронная копия постановки задачи — в файле **Формат отчета.doc**);
3. Распечатка рабочего файла с результативными таблицами и графиками (*Лист 2* Рабочего файла);
4. Выводы по результатам выполнения лабораторной работы.

Выводы излагаются в текстовой форме в порядке, соответствующем перечню 7 задач п. 1.1 — **Постановка задачи**, и сопровождаются ссылками на соответствующие результативные таблицы и графики.

Структура отчета по Лабораторной работе № 2 дана в файле **Формат отчета.doc** и копируется в отчетный файл персональной папки студента на подготовительном этапе.

Подготовка отчета выполняется вне рамок времени, отведенного на выполнение лабораторной работы. Защита отчета студентом производится у преподавателя, ведущего лабораторное занятие.

II. Теоретические основы и методика проведения корреляционно-регрессионного анализа данных

1. Корреляционная связь как разновидность стохастических статистических связей

Изучение объективно существующих связей между социально-экономическими явлениями и процессами — одна из важнейших задач статистической науки.

Среди многих форм связей, имеющих количественный характер и изучаемых количественными методами, особое место занимают *факторные связи*, для исследований которых применяются методы корреляционно-регрессионного анализа.

По своему характеру факторные связи относятся к *причинно-следственным связям*, суть которых заключается в том, что одни явления (причины), протекая в определенных условиях, порождают другие явления (следствия).

Факторные связи между явлениями (причинами, условиями, следствиями) отражаются во взаимосвязях признаков (показателей), характеризующих эти явления.

При изучении факторных связей среди взаимосвязанных признаков (показателей) выделяют факторные и результативные. К *факторным признакам* относят те, которые характеризуют явления-причины и явления-условия и при проведении статистического исследования рассматриваются как независимые. *Результативные признаки* характеризуют явления-следствия и являются зависимыми от факторных в том смысле, что изменение величины факторных признаков ведет к изменению величины результативного признака.

Существуют различные виды и формы факторных связей. Их можно классифицировать по различным критериям — *харак-*

теру, степени тесноты, направлению, виду аналитического выражения связи, количеству факторов в модели связи (рис. 1).

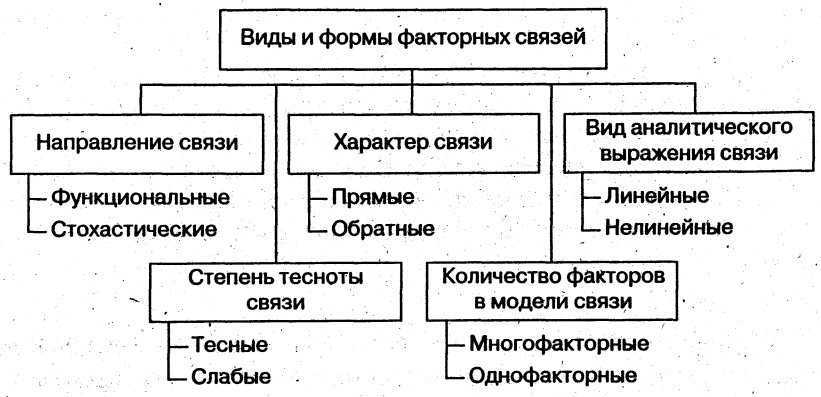


Рис.1. Классификация факторных связей

Функциональные связи

Связь результативного признака Y с факторным признаком X называется ***функциональной***, если каждому возможному значению x_i признака X соответствует одно (или несколько) ***однозначно определенных*** значений y_i признака Y.

Математической моделью однофакторной функциональной связи служит уравнение

$$y_i = f(x_i), \quad (i = 1, 2, ..., n), \qquad (1)$$

где x_i, y_i — факторный и результативный признак соответственно;

f — функция, определяющая зависимость результативного признака от факторного.

В случае зависимости признака Y от нескольких факторных признаков $X_1, X_2, ..., X_m$ модель связи имеет вид:

$$y_i = f(x_{i_1}, x_{i_2}, ..., x_{i_m}) \quad (i = 1, 2, ..., n).$$

Характерная особенность функциональной связи состоит в том, что проявляется в каждом отдельном случае наблюдения и для каждой единицы исследуемой совокупности. При этом известен ***полный перечень*** всех факторов, влияющих на результативный признак Y, а также точный механизм их влияния, выраженный формулой функции $f(x)$. Ввиду этого функциональные связи характеризуются как *полные*, *жесткие*, *детерминированные*, *строго определенные*.

Стохастические связи

В области социально-экономических явлений факторные связи редко носят жестко детерминированный характер. Это объ-

ясняется тем, что наряду с существенными факторами, оказывающими основное, главное влияние на величину результативного признака, на него воздействуют и многие другие, в том числе случайные факторы, причем ***механизм влияния всех факторов в совокупности точно определить невозможно*** и появление каждого конкретного значения y_i носит ***случайный характер***. Связи, учитывающие случайный характер зависимости признаков, относят к числу стохастических (вероятностных).

Стохастическая связь признаков — это связь, при которой одному и тому же значению x_i фактора X (случайному или неслучайному) могут соответствовать ***различные случайные*** значения y_{i_1}, y_{i_2}, ..., y_{i_k} результативного признака Y:

$$x_i \rightarrow \{y_{i_1}, y_{i_2}, ..., y_{i_k}\}. \qquad (2)$$

Возможность появления для одного и того же значения x_i различных результативных значений y_{i_1}, y_{i_2}, ..., y_{i_k} обусловлена тем, что на признак Y помимо учтенного фактора x одновременно воздействуют многие другие неучтенные и неконтролируемые (***случайные***) факторы, которые в каждом конкретном наблюдении могут менять и силу, и направление своего воздействия. Значения фактора X также могут зависеть от случайных обстоятельств. Случайный характер носят и ошибки измерения признаков X и Y, возникающие при проведении статистических наблюдений. Ввиду всех этих обстоятельств значения результативного признака Y, отвечающие факторному значению x_i, оказываются подверженными ***случайному разбросу*** $y_{i_1}, y_{i_2}, ..., y_{i_k}$, причем появление того или иного значения y_i (в силу его случайного характера) не может быть определено точно, а лишь предсказано с некоторой вероятностью.

Математическая модель однофакторной стохастической связи имеет вид уравнения

$$y_i = \varphi(x_i) + \varepsilon_i, \qquad (3)$$

где x_i, y_i — значения факторного и результативного признаков соответственно;

φ — функция, определяющая ту часть значения признака y_i, которая формируется под воздействием учтенного в модели фактора X;

ε_i — часть значения признака y_i, которая возникает вследствие действия неучтенных или неконтролируемых случайных факторов, а также возможных ошибок измерения признаков X, Y.

Если в модели учитывается зависимость признака Y от ряда факторов, то модель имеет вид

$$y_i = \varphi(x_{i_1}, x_{i_2}, ..., x_{i_m}) + \varepsilon_i. \qquad (4)$$

Характерной особенностью стохастических связей является то, что они обнаруживаются не в каждом отдельном случае наблюдения, как при функциональных связях, а лишь при достаточно большом числе наблюдений. При стохастических связях не известен ни полный перечень факторных признаков, ни точное правило их взаимодействия с результативным признаком Y, поэтому эти связи характеризуются как *неполные*, *нежесткие*, *случайные*, *недетерминированные*, *неопределенные*.

Примерами однофакторной стохастической связи являются зависимости потребления семьей продуктов питания от дохода семьи, оценок на экзаменах — от сложности учебных дисциплин, торговой выручки — от затрат на рекламу, себестоимости продукции — от объема производства.

Разновидности стохастических связей представляет классификационная схема на рис. 2.

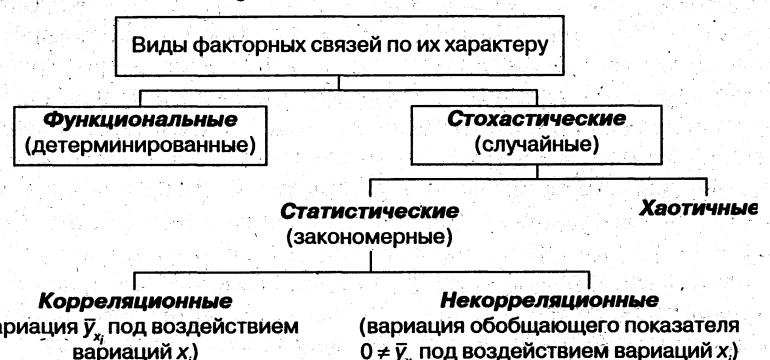


Рис. 2. Классификация факторных связей по их характеру

Корреляционные связи, их свойства и формы выражения

В статистических исследованиях рассматривается особая разновидность стохастических связей — статистические связи, важным частым случаем которых является корреляционная связь (см. рис. 2).

В статистике для описания стохастических соответствий вида (2) множество $\{y_{i_1}, y_{i_2}, ..., y_{i_k}\}$ представляют в виде ряда распределения (т.е. с учетом частоты повторения каждого из значений y_{i_j}) и затем характеризуют построенный ряд *обобщающими статистическими показателями* — средними значениями, показателями вариации, некоторыми относительными показателями и т.д.

Стохастическую связь между ***случайными*** значениями признаков X и Y называют ***статистической***, если с изменением значений x_i фактора X ***закономерным образом изменяется*** какой-либо из обобщающих статистических показателей распределения

$y_{i_1}, y_{i_2}, ..., y_{i_k}$ признака Y. Если при изменении x_i имеет место ***закономерное изменение средних арифметических значений*** $\bar{y}_i$ распределения признака Y, то статистическая связь называется ***корреляционной***. Если же средние $\bar{y}_i$ изменяются незакономерно, но имеет место закономерность изменения ***каких-либо других обобщающих показателей*** распределений признака Y (например, показателей вариации), связь между признаками является статистической, хотя и не корреляционной (см. рис. 2).

Пусть, например, первичные данные статистического наблюдения представлены в табл. 1, где каждому значению x_i фактора X соответствуют несколько значений результативного признака Y.

Таблица 1

Первичные данные статистического наблюдения

x_i	10	9	11	8	9	10	9	11	8	10	9	10	8	9	11
y_i	24	20	27	18	20	24	20	27	20	27	24	27	20	27	30

Для стохастической связи признаков X, Y, заданной табл. 1, имеют место следующие зависимости вида (2), в которых признак Y представлен дискретными рядами распределения:

$$x_1 = 8 \rightarrow \begin{array}{c|c|c} y_1 & 18 & 20 \\ \hline f_1 & 1 & 2 \end{array} \qquad x_3 = 10 \rightarrow \begin{array}{c|c|c} y_3 & 24 & 27 \\ \hline f_3 & 2 & 2 \end{array}$$

$$x_2 = 9 \rightarrow \begin{array}{c|c|c|c} y_2 & 20 & 24 & 27 \\ \hline f_2 & 3 & 1 & 1 \end{array} \qquad x_4 = 11 \rightarrow \begin{array}{c|c|c} y_4 & 27 & 30 \\ \hline f_4 & 2 & 1 \end{array} \qquad (5)$$

Выбирая в качестве обобщающего показателя этих рядов распределения среднее арифметическое значение $\bar{y}$, зависимость Y от X можно представить в виде соответствий

$$x_1 = 8 \rightarrow \bar{y}_{x_1} = \frac{18 \cdot 1 + 20 \cdot 2}{3} = 19{,}3; \qquad x_3 = 10 \rightarrow \bar{y}_{x_3} = 25{,}5;$$

$$x_2 = 9 \rightarrow \bar{y}_{x_2} = \frac{20 \cdot 3 + 24 \cdot 1 + 27 \cdot 1}{5} = 22{,}2; \quad x_3 = 10 \rightarrow \bar{y}_{x_4} = 28{,}0.$$

Так как с ростом значений x_i фактора X систематически возрастают и средние значения $\bar{y}_{x_i}$ результативного признака Y (табл. 2), то связь между этими признаками носит закономерный характер и, следовательно, является статистической. Поскольку закономерно меняется средняя величина $\bar{y}_{x_i}$, то эта статистическая связь корреляционная.

Средние арифметические значения $\bar{y}_{x_i}$ распределений признака Y, вычисленные при условии, что фактор X принимает фиксированное значение x_i, называют ***условными средними.***

Для рассмотренного примера зависимость между значениями x_i и условными средними $\bar{y}_{x_i}$ определяет табл. 2.

Таблица 2

Корреляционная взаимосвязь признаков

x_i	8	9	10	11
$\bar{y}_{x_i}$	19,3	22,2	25,5	28,0

Корреляционные взаимосвязи признаков имеют разную форму выражения, различаясь по степени тесноты, направлению, количеству факторов в модели связи (см. рис. 1).

Теснота корреляционной связи характеризует степень ее приближения к функциональной связи. Если значению x_i признака X соответствуют ***близкие друг другу***, тесно расположенные около своей средней $\bar{y}_{x_i}$ значения $y_{i_1}, y_{i_2}, ..., y_{i_k}$, то связь ***тесная (сильная)***. Если же эти результативные значения существенно отклоняются от $\bar{y}_{x_i}$, связь менее тесная (она может быть слабой, умеренной, заметной).

Таким образом, степень тесноты связи зависит от степени варьирования результативного признака Y при фиксированном значении факторного признака X.

В **зависимости от направления изменения результативного признака** различаются прямые и обратные связи. Если результативный признак Y изменяется в том же направлении, что и факторный признак X (т.е. с ростом X признак Y также возрастает, а при уменьшении X — уменьшается), то ***связь прямая***. Если же результативный признак меняется в противоположном направлении, то ***связь обратная***.

По количеству факторов, действующих на результативный признак, различают связи однофакторные и многофакторные. Если исследуется связь между одним признаком-фактором X и результативным признаком Y (при абстрагировании от влияния на Y всех других факторов), то говорят ***об однофакторной связи и парной корреляции*** (рассматривается *пара признаков*). Если же изучается воздействие на Y нескольких факторных признаков $X_1, X_2, ..., X_m$, то связь называют ***многофакторной***, а корреляцию — ***множественной***.

В случае многофакторной связи имеется в виду, что все влияющие факторы действуют в комплексе — *одновременно и во взаимосвязи*. Если же изучается зависимость между результативным и одним из факторных признаков при фиксированных значениях других факторных признаков, то говорят о ***частной корреляции.***

Для корреляционной связи характерны **следующие свойства.**

1. Будучи стохастическими, корреляционные связи проявляются не в единичных наблюдениях, а ***в общем и среднем*** при достаточно большом числе наблюдений. Поэтому для своего исследования они требуют ***эмпирических статистических данных, полученных на основе массовых наблюдений***.

2. Эмпирические статистические данные отображают, как правило, совокупное действие на результативный показатель всех имеющих место причин и условий, однако в корреляционных связях учитываются лишь некоторые из них. Наличие прочих «неучтенных» факторов проявляется в том, что корреляционные связи, даже обнаруженные на основе массового материала (где случайные факторы нивелируются), оказываются ***неполными***. По силе связи они никогда не достигают связи функциональной — полной и однозначной.

3. Корреляционные связи являются ***необратимыми***: наличие зависимости результативного признака Y от фактора X не означает наличия обратной связи — зависимости X от Y (так, производительность труда зависит от уровня автоматизации производства, но обратной зависимости нет).

2. Табличное и графическое представление однофакторных корреляционных связей

При изучении однофакторных корреляционных связей удобной формой представления зависимости признака Y от фактора X являются корреляционные и аналитические статистические таблицы, а также точечные и линейные графики в декартовой системе координат (X, Y).

1. Табличное представление корреляционных связей

При построении корреляционной таблицы значения признаков X и Y ранжируются в порядке возрастания, факторные значения x_i располагаются, как правило, в строках таблицы, результативные — в столбцах (графах), а на пересечении строк и столбцов проставляются числа, указывающие *частоту появления* различных результативных значений y_j при фиксированном факторном значении x_i. При таком построении таблицы каждая i-я строка представляет *распределение признака Y при условии $X = x_i$*. В итоговой строке таблицы проставляются частоты n_{y_j} повторения в эмпирических данных результативного значения y_j, а в итоговом столбце — частоты n_{x_i} повторения значений x_i. Примером корреляционной таблицы служит табл. 3, представляющая зависимость признаков, заданную табл. 1. Легко видеть, что табл. 3 — это компактное выражение стохастических зависимостей (5) между значениями x_i

и соответствующими распределениями результативного признака $y_{i_1}, y_{i_2}, ..., y_{i_k}$.

Таблица 3

Корреляционная таблица взаимосвязи признаков

X \ Y	18	20	24	27	30	n_x
8	1	2				3
9		3	1	1		5
10			2	2		4
11				2	1	3
n_y	1	5	3	5	1	15

В статистической практике изучение взаимосвязей явлений проводится, как правило, по достаточно большому числу наблюдений, а значения наблюдаемых признаков X и Y представляются в сгруппированном виде. При этом в корреляционной таблице строки соответствуют группировке факторных значений x_i, столбцы — группировке результативных значений y_j. На пересечении j-й строки и k-го столбца указывается численность единиц совокупности, факторные значения которых принадлежат j-му интервалу группировки признака X, а результативные — k-му интервалу группировки признака Y. Примером корреляционной таблицы для сгруппированных исходных данных служит табл. 4.

Таблица 4

Распределение предприятий по величине среднегодовой стоимости основных фондов и выпуску продукции

Среднегодовая стоимость основных производственных фондов, млн руб.	Выпуск продукции, млн руб.					
	80,00–108,80	108,80–137,60	137,60–166,40	166,40–195,20	195,20–224,00	n_x
94,00–134,80	1	2				3
134,80–175,60	4	2	2			8
175,60–216,40	1	3	4	1	1	10
216,40–257,20			3	2		5
257,20–298,00				2	2	4
n_y	6	7	9	5	3	30

Наиболее удобной формой представления корреляционных зависимостей при большом числе наблюдений являются ***групповые аналитические таблицы***, отражающие результаты ***аналитической группировки совокупности по факторному признаку***.

При построении аналитической таблицы для каждой выделенной j-й группы подсчитывается численность составляющих ее

факторных значений x, а также суммарное и среднее $\bar{y}_j$ групповые значения результативного признака. Примером аналитической таблицы является табл. 5.

Таблица 5

Зависимость выпуска продукции от среднегодовой стоимости основных производственных фондов

Номер группы	Группы предприятий по стоимости основных фондов, млн руб	Число предприятий	Выпуск продукции, млн руб.	
			Всего по группе	В среднем на одно предприятие, $\bar{y}_j$
1	94,00–134,80	3	331,00	110,33
2	134,80–175,60	8	887,00	110,88
3	175,60–216,40	10	1461,00	146,10
4	216,40–257,20	5	824,00	164,80
5	257,20–298,00	4	806,00	201,50
Итого		30	4309,00	143,63

Графическое представление корреляционной связи

Для графического представления парных корреляционных связей применяются два вида графиков — ***поле корреляции*** и эмпирическая линия связи, называемая также ***эмпирической линией регрессии.***

Поле корреляции — это точечный график, используемый для изображения связи признаков в совокупностях небольшого объема. При построении графика в декартовой системе координат по оси абсцисс в определенном масштабе наносятся значения факторного признака, а по оси ординат — результативного. На пересечении абсцисс и ординат отмечаются точки (x_i, y_i), совокупность которых и представляет корреляционное поле (рис. 3).

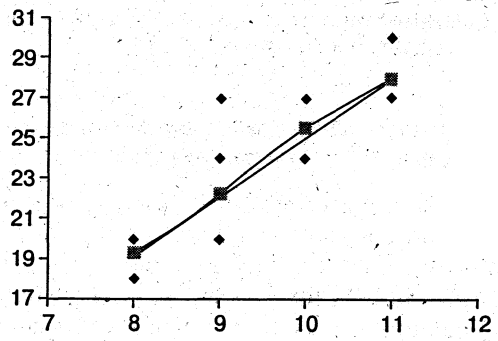


Рис. 3. Корреляционное поле и эмпирическая линия связи для условных средних $\bar{y}_X$ (по данным табл. 1, 2)

Эмпирическая линия связи представляют собой ломаную линию, изображающую изменение средних значений признака Y в зависимости от изменения значений фактора X.

При несгруппированных значениях признака X по оси абсцисс откладываются значения x_i, а по оси ординат — условные средние $\bar{y}_{x_i}$. Нанеся на поле графика точки $(x_i, \bar{y}_{x_i})$ и соединив их последовательно отрезками прямых, получают ломаную линию, которая и является эмпирической линией связи — ***графиком условных средних*** $\bar{y}_{x_i}$ результативного признака. Пример такого графика дан на рис. 3 для несгруппированных данных табл. 1, 2.

В случае сгруппированных факторных значений по оси абсцисс откладываются ***середины*** x'_j интервалов группировки, а по оси ординат — соответствующие средние групповые значения $\bar{y}_j$ результативного признака. Отметив точки $(x'_j, \bar{y}_j)$ и соединив их отрезками прямых, получают эмпирическую линию связи — ***график групповых средних*** $\bar{y}_j$ результативного признака. Пример графика для групповых средних табл. 5 дан на рис. 4.

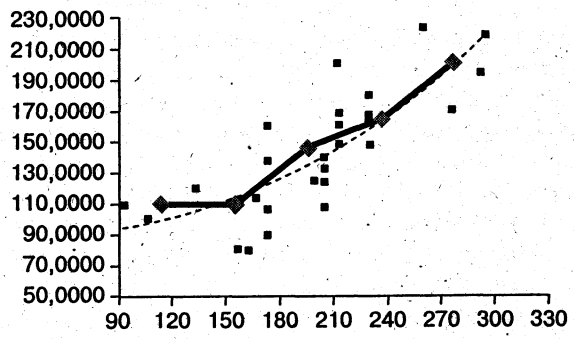


Рис. 4. Корреляционное поле и эмпирическая линия связи для групповых средних $\bar{y}_j$ по данным табл. 5 (пунктирная линия — параболическая теоретическая линия регрессии)

3. Моделирование однофакторных корреляционных связей на основе функциональных зависимостей

Воздействие на результативный признак Y фактора X осуществляется в условиях сложного взаимодействия факторов, отражающих различные причины и условия формирования результативных значений. Ввиду этого в каждом конкретном случае наблюдения (x_i, y_i) величина y_i зависит не только от величины x_i, но и от того, как именно сложатся в этом случае все прочие факторы, влияющие на Y.

В модели стохастической связи (3) такое случайное стечение обстоятельств отражается в случайностной компоненте ε_i, а на графике корреляционного поля — в наличии разных точек y_i, отвечающих точке x_i (см. рис. 3, 4).

Переходя при построении корреляционной зависимости к средним значениям признака Y (условным $\bar{y}_{x_i}$ или групповым $\bar{y}_j$), тем самым отчасти элиминируют влияние случайных факторов. Иными словами, ***за счет усреднения*** результативных значений y_{i_1}, y_{i_1}, ..., y_{i_k} корреляционная связь, будучи по своей природе стохастической (неоднозначной), выражается ***в форме однозначных зависимостей***:

$$\bar{y}_{x_i} = f(x_i), \qquad (i = 1, 2, ..., n), \tag{6}$$

изображаемых графически некоторой эмпирической линией связи.

Эмпирическая линия связи является обычно ломаной линией, имеющей более или менее значительные изломы (см. рис. 3, 4). Наличие таких изломов объясняется тем, что влияние на признак Y факторов, неучтенных в модели, ***погашается в средних величинах*** $\bar{y}_{x_i}$ ***не полностью*** в силу недостаточно большого (ограниченного) количества наблюдений (x_i, y_i).

Для того чтобы полностью абстрагироваться (отвлечься) от влияния на Y всех иных (кроме X) факторов и установить подлинную закономерность взаимосвязи признаков X и Y, в статистике прибегают к выравниванию эмпирической ломаной линии связи по некоторой плавной, «сглаженной» кривой, около которой группируются или к которой тяготеют точки $(x_i, \bar{y}_{x_i})$ (на рис. 4 сглаживающая прямая обозначена пунктирной линией).

Линию, сглаживающую эмпирическую ломаную линию связи, называют ***теоретической линией регрессии Y на X***, или просто ***линией регрессии***. Эта линия отражает теоретическую форму связи признаков X и Y, т.е. закономерность изменения ***средних значений*** признака Y в зависимости от изменения фактора X при условии ***полного взаимопоглощения всех прочих случайных по отношению к фактору X причин***. Иначе говоря, теоретическая линия регрессии определяет основную тенденцию взаимосвязи признаков X и Y.

Уравнение

$$\hat{y}_x = f(x), \tag{7}$$

описывающее математически теоретическую линию регрессии, называют ***уравнением регрессии***. В уравнении (7) переменная $\hat{y}_x$ — ***это средняя величина признака Y***, меняющаяся по мере изменения фактора X, а функция $f(x)$ устанавливает аналитический вид ***однозначной зависимости между вариациями x и $\hat{y}_x$***.

Таким образом, уравнение регрессии ***аппроксимирует*** (приближенно характеризует) корреляционную связь признаков X и Y, представляя ее ***в форме функциональной зависимости*** (7). При этом значения $\bar{y}_{x_i}$ выступают в качестве приближенных значений условных средних $\bar{y}_{x_i}$ (или значений групповых средних $\bar{y}_j$), полученных в предположении, что x_i является единственной причиной изменения y_i, а случайная возмущающая переменная ε_i отсутствует ($\varepsilon_i = 0$).

Уравнение регрессии (7), являясь математической моделью изучаемой корреляционной связи и выражая среднюю величину признака Y как функцию признака X, при правильном построении модели будет выявлять главнейшие свойства взаимосвязи признаков X и Y, исключая отдельные «возмущения», вызванные случайными, не характерными для изучаемого явления факторами.

Абстрагирование в регрессионной (функциональной) модели связи от того обстоятельства, что кроме фактора X на признак Y воздействуют и многие другие факторы, приводит, конечно, к некоторому упрощению действительного механизма взаимосвязи признаков, однако позволяет сконцентрировать внимание на ***закономерности зависимости признака Y от фактора X.***

Правомерность моделирования стохастической корреляционной связи на основе функциональной зависимости (7) будет оправданной лишь в тех случаях, если корреляционная связь не столь значительно отстоит от функциональной, т.е. не дает значительной погрешности в отклонениях ($y_i - \bar{y}_{x_i}$).

Это требование порождает в теории корреляционной связи две главные задачи:

- ***определить теоретическую форму связи*** — подыскать такую форму функциональной зависимости (7), которая в наилучшей степени отвечает сущности обнаруженной корреляционной связи признаков;
- ***измерить тесноту связи*** — оценить, в какой мере изучаемая корреляционная связь приближается по своей силе к связи изучаемой функциональной.

В однофакторных регрессионных моделях взаимосвязи социально-экономических явлений наиболее часто используются следующие ***типы математических функций***, описывающих теоретическую линию регрессии и характеризующих механизм взаимодействия факторного и результативного признаков:

$\hat{y}_x = a_0 + a_1 x$ — линейная,

$\hat{y}_x = a_0 + a_1 \frac{1}{x}$ — гиперболическая,

$\hat{y}_x = a_0 + a_1 \lg x$ — логарифмическая,

$\hat{y}_x = a_0 \cdot x^{a_1}$ — степенная, (8)

$\hat{y}_x = a_0 + a_1 x + a_2 x^2$ — параболическая,

$\hat{y}_x = a_0 + a_1^x$ — показательная.

Коэффициенты уравнений регрессии a_0, a_1, a_2, ... называют ***параметрами связи.***

Функциональные зависимости (8) описывают типы кривых, применяемых для сглаживания ломаных эмпирических линий связи, причем операция сглаживания сводится, по существу, к нахождению численных значений параметров a_k.

Наиболее простой регрессионной моделью однофакторой корреляционной связи является линейная модель

$$\hat{y}_x = a_0 + a_1 x, \qquad (9)$$

изображаемая графически прямой линией. Модель отражает ***линейную взаимосвязь признаков X и Y,*** когда с возрастанием значений X происходит непрерывное, более или менее равномерное возрастание или убывание средних значений Y (см. рис. 3).

Все прочие модели (8) отражают тот или иной вид нелинейной взаимосвязи признаков, когда изменение средних значений Y в зависимости от X происходит неравномерно (с ускорением, замедлением или изменением направления связи). В этих случаях сглаживающие теоретические линии регрессии представляют собой соответствующие нелинейные кривые — гиперболы, параболы 2-го порядка (как на рис. 4) и др.

Разброс фактических значений y_i вокруг теоретических значений $\hat{y}_x$, рассчитанных по избранному для моделирования уравнению регрессии, обусловлен влиянием множества случайных факторов. Разности

$$\varepsilon_i = y_i - \hat{y}_{x_i}, \qquad (10)$$

называемые ***остаточными величинами*** (или ***остатками***), оценивают отклонения расчетных значений $\bar{y}_{x_i}$ от фактических значений y_i.

Следовательно, при построении регрессионной модели численные значения коэффициентов a_k выбранного типового уравнения регрессии (8) необходимо искать так, чтобы обеспечить ***наименьшие возможные остатки*** ε_i для всех случаев наблюдения (x_i, y_i).

Для этой цели используется ***метод наименьших квадратов (МНК)***, который позволяет рассчитать параметры a_k выбранного типового уравнения регрессии таким образом, чтобы теоретическая линия регрессии была бы ***в среднем наименее удалена от всех точек (x_i, y_i)*** по сравнению с любой другой теоретической линией регрессии, отвечающей выбранному типу функции связи (8).

Согласно МНК задача поиска значений параметров a_k, минимизирующих сумму погрешностей (10), имеет вид

$$S = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} \left(y_i - \hat{y}_{x_i} \right)^2 \to \min \qquad (11)$$

и решается как задача на экстремум — путем приравнивания к нулю первых частных производных функции S по каждому искомому параметру a_k уравнения регрессии. Это приводит к системе уравнений, называемой ***нормальной***, решение которой дает численные значения параметров a_k, минимизирующие функцию S.

Таким образом, параметры связи a_k в силу их расчета по МНК являются ***усредненными по всей совокупности наблюдений*** (x_i, y_i). Они отражают взаимосвязь признаков X и Y только ***в общем итоге, по всей совокупности в целом*** (для каждой индивидуальной пары (x_i, y_i) значения a_k остаются неизвестными).

Вследствие усреднения параметров связи a_k результативные значения $\hat{y}_x$ также являются ***усредненными*** по всей совокупности наблюдений, откуда вытекает важная ***особенность уравнений регрессии***: будучи похожими ***по своей форме*** на уравнения функциональной зависимости (справедливые для каждого отдельного i-го наблюдения), они имеют другой ***содержательный смысл*** — показывают ***типичное для всей совокупности в целом*** соотношение между величинами факторного и результативного признаков.

При изучении ***многофакторных*** корреляционных связей ***методология их моделирования уравнениями регрессии*** аналогична рассмотренной. Уравнения многофакторной регрессии имеют вид

$$\hat{y}_{x_1, x_2, \ldots, x_m} = f(x_1, x_2, \ldots, x_m)$$

и позволяют приближенно оценить меру влияния на результативный признак Y каждого из включенных в модель факторов X при фиксированных (на среднем уровне) значениях остальных факторов, а также оценить влияние на Y различных сочетаний рассматриваемых факторов.

4. Методика корреляционно-регрессионного анализа (КРА)

Исследование связи между признаками требует прежде всего проведения ***теоретического анализа существа изучаемого явления,*** включая формулировку задачи исследования, отбор факторных признаков X, влияющих на результативный признак Y, выдвижение гипотезы о наличии корреляционной связи между результативным и факторными признаками.

По завершении теоретического анализа проводится ***анализ свойств совокупности единиц наблюдения*** (x_i, y_i). Необходимость такого анализа обусловлена тем, что для практического применения методов КРА должны выполняться определенные требования в отношении отбора единиц наблюдения:

1) ***однородность изучаемой статистической совокупности*** (например, для совокупности предприятий это однородность выпускаемой продукции, одинаковый характер технологического процесса, одинаковый тип используемого оборудования);

2) ***репрезентативность выборки*** единиц наблюдаемой совокупности, так как при малой выборке может быть «затушевано» действие случайных факторов, взаимопогашение которых происходит при расчете условных средних $\bar{y}_x$;

3) ***достаточность объема эмпирических данных*** для выявления закономерности связи (число факторных признаков должно быть в 5–6 раз меньше объема изучаемой совокупности);

4) ***независимость включаемых в регрессионную модель факторов-признаков*** $X_1, X_2, \ldots, X_m$, так как наличие связи между ними свидетельствует о том, что они характеризуют одни и те же стороны изучаемого явления и в значительной мере дублируют друг друга;

5) ***нормальный характер распределения*** изучаемого признака Y при фиксированных значениях признаков $X_1, X_2, \ldots, X_m$.

В статистических исследованиях часто приходится сталкиваться с теми или иными отклонениями от указанных требований, однако практика показывает, что незначительные отклонения не являются препятствием к применению методов КРА.

Корреляционно-регрессионный анализ взаимосвязей признаков (показателей) включает следующие этапы:

1) установление факта наличия корреляционной связи изучаемых признаков, определение направления связи и эмпирическая оценка ее тесноты;

2) проверка статистической значимости (неслучайности) выявленной корреляционной связи;

3) выбор аналитической формы связи и построение математической модели связи в виде уравнения регрессии;

4) оценка статистической значимости коэффициентов построенного уравнения регрессии и определение их доверительных интервалов;

5) анализ адекватности построенной регрессионной модели связи;

6) экономическая интерпретация регрессионной модели связи.

На каждом из этапов КРА применяются соответствующие статистические методы и числовые характеристики.

Содержание этапов КРА рассматривается ниже на примере ***парной корреляции признаков X и Y.***

1. ***Для установления факта наличия корреляционной связи*** факторного и результативного признаков используются методы:

- сопоставления рядов значений признаков X и Y;
- графического представления взаимосвязи признаков;
- корреляционных таблиц;
- аналитической группировки.

При использовании метода аналитической группировки оценивается (на основе данных аналитической таблицы) ***степень тесноты*** корреляционной связи признаков X и Y, для чего рассчитываются специальные показатели:

r — ***линейный коэффициент корреляции***, измеряющий тесноту связи в предположении линейности взаимосвязи признаков X и Y;

η — ***эмпирическое корреляционное отношение***, выступающее как универсальный показатель тесноты связи при любой форме связи (как линейной, так и нелинейной);

η^2 — ***эмпирический коэффициент детерминации*** (***причинности***), определяющий силу связи, т.е. оценивающий, насколько вариация результативного признака Y объясняется вариацией фактора X.

Расчет показателей производится по формулам

$$r = \frac{\sum x_i y_i - \dfrac{\sum x_i \times \sum y_i}{n}}{\sqrt{\left(\sum x_i^2 - \dfrac{\left(\sum x_i\right)^2}{n}\right) \times \left(\sum y_i^2 - \dfrac{\left(\sum y_i\right)^2}{n}\right)}}; \qquad (12)$$

$$\eta^2 = \frac{\delta^2_{\text{факт}}}{\sigma_y^2}, \quad \eta = \sqrt{\frac{\delta^2_{\text{факт}}}{\sigma_y^2}},$$

где $\delta^2_{\text{факт}}$ — межгрупповая дисперсия результативного признака Y, обусловленная влиянием только фактора X;

δ^2_y — общая дисперсия признака Y, обусловленная влиянием на Y всех факторов, включая X;

n — число единиц наблюдения (т.е. число пар (x_i, y_i)), суммирование в показателе r производится по всем n наблюдаемым признакам.

Межгрупповая дисперсия признака Y определяется на основе данных аналитической таблицы по формуле

$$\delta^2_{\text{факт}} = \frac{\sum_{j=1}^{m} \left(\bar{y}_j - \bar{y}\right)^2 n_j}{\sum_{j=1}^{m} n_j}, \qquad (13)$$

где $\bar{y}$ — общая средняя признака Y для всей совокупности;

$\bar{y}_j$ — среднее значение признака Y в j-й группе;

n_j — численность j-й группы;

m — число выделенных групп.

Общая дисперсия признака Y вычисляется по формулам

$$\sigma_y^2 = \frac{\sum_{i=1}^{n} \left(y_i - \bar{y}\right)^2}{n} \quad \text{или} \quad \sigma_y^2 = \overline{y^2} - \bar{y}^2, \qquad (14)$$

где $\overline{y^2} = \dfrac{\sum_{i=1}^{n} y_i^2}{n}$, $\bar{y}^2 = \left(\dfrac{\sum_{i=1}^{n} y_i}{n}\right)$.

Для показателей силы и тесноты корреляционной связи характерны следующие свойства.

А. Значения показателей изменяются в пределах:

$$-1 \le r \le 1, \quad 0 \le \eta \le 1, \quad 0 \le \eta^2 \le 1.$$

Чем ближе значения показателей к единице, тем теснее связь и больше сила связи.

Знак при r указывает на направление связи: знак «+» соответствует прямой линейной зависимости, знак «–» — обратной.

Б. Для качественной оценки тесноты связи используется ***шкала Чэддока***:

Значение показателей тесноты связи $\lvert r\rvert$, η	0,1–0,3	0,3–0,5	0,5–0,7	0,7–0,9	0,9–0,99
Характеристика связи	Слабая	Умеренная	Заметная	Высокая	Весьма высокая

В. Если $|r| = 1$ или $\eta = 1$, то корреляционная зависимость становится полной, т.е. функциональной (равенство показателей единице достигается при $\delta^2_{\text{факт}} = \delta^2_y$, что означает отсутствие влияния на Y любых иных, кроме X, факторов).

Г. Если $\eta = 0$, то между признаками X и Y нет корреляционной связи (равенство $\eta = 0$ имеет место только при $\delta^2_{\text{факт}} = 0$, что означает независимость признака Y от фактора X).

Если $r = 0$, то между изучаемыми признаками ***нет линейной корреляции***, что не исключает, однако, существования какого-либо другого вида корреляционной зависимости (параболической, показательной или др.).

Д. Факт совпадения или несовпадения значений показателей η и r используется для оценки формы связи: $\eta = r$ ***только при наличии прямолинейной связи***. Несовпадение этих показателей означает, что связь между признаками криволинейная. Установлено, что если

$$|\eta^2 - r^2| \le 0{,}1, \qquad (15)$$

то зависимость признака Y от фактора X можно считать прямолинейной.

2. ***Проверка статистической значимости корреляционной связи признаков*** позволяет количественно оценить, насколько выявленная связь между факторным и результативным признаками носит ***неслучайный характер***, т.е. насколько она является ***типичной***, ***существенной*** для изучаемого явления (для генеральной совокупности).

Необходимость в такой оценке обусловлена тем, что корреляционному анализу обычно подвергается сравнительно небольшие (по составу единиц) совокупности и возникает вопрос, насколько оценки тесноты связи, сделанные по ограниченным выборочным данным, правомерны для всей генеральной совокупности. Если связь признается существенной, она моделируется и исследуется на последующих этапах методами КРА. Если же связь оценена как несущественная, это означает, что выбор факторного признака произведен недостаточно корректно и следует попытаться отыскать иную зависимость признаков.

Для оценки существенности связи используются ***критерии***, известные из математической статистики (F-критерий Фишера, t-критерий Стьюдента, критерий «χ-квадрат» и др.).

В случае малых выборок ($n \le 30$) проверка существенности связи проводится путем сравнения расчетного значения $\eta^2_{\text{расч}}$ с его ***критическим значением*** $\eta^2_{\text{табл}}(k_1, k_2)$, заданным в таблице распределения Стьюдента для уровня значимости α (обычно $\alpha = 0{,}05$ или $\alpha = 0{,}01$*) и числа степеней свободы $k_1 = m - 1$, $k_2 = n - m$, где m — число групп в аналитической таблице, n — объем наблюдаемой совокупности. Если $\eta^2_{\text{расч}} > \eta^2_{\text{табл}}$, то связь признается неслучайной (типичной для изучаемого явления).

* Уровень значимости α связан с доверительной вероятностью P соотношением $\alpha = 1 - P$. Поскольку в экономических исследованиях обычно используются уровни надежности $P = 0{,}954$ или $P = 0{,}997$, то наиболее часто применяются уровни значимости $\alpha = 0{,}05$ или $\alpha = 0{,}01$.

3. ***Построение математической модели*** $\hat{y}_x = f(x)$ ***корреляционной зависимости признаков*** осуществляется в два шага.

Первый шаг заключается в том, чтобы по виду корреляционного поля или эмпирической линии регрессии (построенным по фактически наблюдаемым данным (x_i, y_i)) установить ***основную тенденцию взаимосвязи признаков*** и выразить ее в форме соответствующей математической функции связи вида (8). При этом для выбора типа функции связи важен лишь ***общий вид*** функции $f(x)$, без конкретизации значений входящих в нее параметров связи a_k ($k = 0, 1, 2, \ldots$). Выбор того или иного типа функции связи означает лишь выдвижение и принятие некоторой (теоретически обоснованной или практически приемлемой) гипотезы о механизме взаимодействия изучаемых признаков.

На **втором шаге** определяются численные значения параметров связи a_k выбранной типовой функции $f(x)$. Для этой цели применяется ***метод наименьших квадратов*** (МНК), основанный на использовании критерия минимизации остатков (11). Применение МНК приводит к ***системе нормальных уравнений*** с неизвестными параметрами a_k, причем система содержит столько уравнений, сколько параметров связи имеется в типовой функции $f(x)$. В результате решения системы нормальных уравнений параметры a_k типовой функции связи $f(x)$ получают конкретные числовые значения и модель приобретает вид уравнения регрессии $\hat{y}_x = f(x)$, в котором значения a_k ($k = 0, 1, 2, \ldots$) являются числовыми коэффициентами при k-й степени фактора X^k. Это уравнение и является искомой математической моделью изучаемой корреляционной связи. На ее основе в дальнейшем рассчитываются теоретические значения $\hat{y}_x$ результативного признака.

Часто для выражения формы связи подходит одновременно несколько типовых функций $f(x)$, поэтому окончательный выбор вида функции связи должен быть обоснован путем рассмотрения и оценки альтернативных вариантов регрессионных моделей.

4. ***Оценка статистической значимости коэффициентов уравнения регрессии и определение их доверительных интервалов***. При построении уравнения регрессии $\hat{y}_x = f(x)$ параметры a_k рассчитываются по ограниченному числу эмпирических данных (x_i, y_i) и, следовательно, являются лишь ***приближенными оценками*** фактических параметров связи. Поэтому необходимо вычислить ***средние ошибки*** μ_{a_k} найденных параметров a_k и с заданной доверительной вероятностью P ***определить пределы, в которых могут находиться фактические значения*** a_k. Кроме того, найденные параметры a_k необходимо ***проверить на статистическую значимость*** (неслучайность).

Расчет ошибок параметров a_k основан на использовании ***остаточной дисперсии*** $\sigma^2_{ост}$ (обозначаемой также σ^2_ε), которая характеризует колеблемость эмпирических значений y_i около их выровненных значений $\hat{y}_{x_i}$ (т.е. около теоретической линии регрессии). Иными словами, остаточная дисперсия оценивает вариацию остатков ε_i, определяемых соотношением (10). Расчет остаточной дисперсии производится по формуле

$$\sigma^2_{ост} = \frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_{x_i}\right)^2}{n}. \qquad (16)$$

В случае линейной корреляционной связи $\hat{y}_x = a_0 + a_1 x$ средние ошибки параметров a_0 и a_1 вычисляются по формулам

$$\mu_{a_0} = \sqrt{\frac{\sigma^2_{ост}}{n-2}}, \quad \mu_{a_1} = \sqrt{\frac{\sigma^2_{ост}}{\sigma^2_x (n-2)}}, \qquad (17)$$

где σ^2_x — дисперсия факторного признака X.

Доверительный интервал для каждого параметра a_k рассчитывается как величина

$$\Delta_{a_k} = \pm t \cdot \mu_{a_k}, \qquad (18)$$

где t — коэффициент доверия, соответствующий заданному уровню надежности P.

Статистическая значимость параметра a_k (т.е. неслучайность найденного значения a_k, его типичность для всей генеральной совокупности) проверяется путем сопоставления величины a_k со средней ошибкой μ_{a_k} исходя из t-критерия Стьюдента

$$t_{a_k} = \frac{a_k}{\mu_{a_k}}.$$

При большом числе наблюдений ($n > 30$) параметр a_k считается значимым, если $t_{a_k} > 3$.

Если выборка мала ($n \le 30$), рассчитанная величина t_{a_k} сопоставляется с табличным (критическим) значением t-критерия Стьюдента для числа степеней свободы $n - 2$ и уровня значимости α ($\alpha = 0{,}05$ или $\alpha = 0{,}01$ в зависимости от заданного уровня надежности P). Если $t_{a_k} > t_{табл}(\alpha, n{-}2)$, то параметр a_k считается значимым.

Проверка значимости производится для каждого параметра связи a_k построенного уравнения регрессии. Если параметр a_k является значимым, то практически невероятно, что его значение обусловлено только стечением случайных обстоятельств.

5. ***Анализ адекватности регрессионной модели*** преследует цель оценить, насколько построенная теоретическая модель взаимосвязи признаков отражает фактическую зависимость между этими признаками, и тем самым оценить ***практическую пригодность*** синтезированной модели связи. Такая оценка необходима, в частности, для сравнительного анализа качества альтернативных вариантов моделей с целью выбора наилучшей из них.

Проверка адекватности регрессионной модели фактическим данным наблюдения (x_i, y_i) осуществляется путем оценки тесноты связи между факторными значениями x_i и выровненными (теоретическими) результативными значениями $\hat{y}_{x_i}$, рассчитанными по уравнению регрессии. При этом используются следующие показатели вариации признака Y:

- ***общая дисперсия*** σ^2_y, вычисляемая по формуле (14) и оценивающая вариацию *эмпирических значений* y_i под влиянием всех действующих на признак Y факторов;
- ***факторная дисперсия*** $\delta^2_{факт}$, оценивающая вариацию *расчетных (теоретических) значений* Y под воздействием фактора X:

$$\delta^2_{факт} = \frac{\sum_{i=1}^{n}\left(\hat{y}_{x_i} - \bar{y}\right)^2}{n}, \qquad (20)$$

где отклонения $\left(\hat{y}_{x_i} - \bar{y}\right)$ характеризуют колеблемость выровненных значений $\hat{y}_{x_i}$ от их средней величины $\bar{y}^*$;

- ***остаточная дисперсия*** $\sigma^2_{ост}$, вычисляемая по формуле (16) и характеризующая вариацию *расчетных значений* Y под воздействием всех иных, кроме X, факторов.

Анализ адекватности модели выполняется в три этапа:

- измерение тесноты связи признаков в уравнении регрессии;
- определение по величине показателей тесноты связи практической пригодности построенной модели связи;
- проверка статистической значимости показателей тесноты связи.

1. Для построенной регрессионной модели измерение ***тесноты связи*** признаков X и Y осуществляется на основе следующих показателей:

R^2 — индекс детерминации (называемый также ***теоретическим коэффициентом детерминации*** и обозначаемый $\eta^2_{теор}$), показывающий, какая часть общей вариации *расчетных (теоретических) значений* признака Y объясняется вариацией фактора X;

R — индекс корреляции (называемый также ***теоретическим корреляционным отношением*** и обозначаемый $\eta_{теор}$), оценивающий

* Средняя величина расчетных значений $\hat{y}_{x_i}$ совпадает с общей средней $\bar{y}$ эмпирических значений y_i, поскольку при применении метод наименьших квадратов суммы теоретических и эмпирических значений признака Y совпадают.

степень тесноты связи между факторными значениями x_i и расчетными результативными значениями $\hat{y}_{x_i}$;

r — линейный коэффициент корреляции, используемый для измерения тесноты связи признаков в регрессионной модели в случае линейной функции связи $f(x)$.

Расчет этих показателей определяется следующими формулами:

$$R^2 = \frac{\delta^2_{\text{факт}}}{\sigma^2_y}; \qquad R = \sqrt{\frac{\delta^2_{\text{факт}}}{\sigma^2_y}}; \qquad r = a_1 \frac{\sigma_x}{\sigma_y}, \tag{21}$$

где a_1 — коэффициент регрессии в регрессионной модели связи.

Используя соотношения (14) и (20) для вычисления соответствующих дисперсий, а также известное из математической статистики ***правило сложения дисперсий***

$$\sigma^2_y = \delta^2_{\text{факт}} + \sigma^2_{\text{ост}},$$

для расчета показателей R^2 и R применяют формулы

$$R^2 = \frac{\sum_{i=1}^{n}\left(\hat{y}_{x_i} - \bar{y}\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2} = 1 - \frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_{x_i}\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2}, \tag{22}$$

$$R = \sqrt{\frac{\sum_{i=1}^{n}\left(\hat{y}_{x_i} - \bar{y}\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2}} = \sqrt{1 - \frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_{x_i}\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2}}. \tag{23}$$

2. ***Практическая пригодность синтезированной регрессионной модели связи*** оценивается по величине показателей r (в случае линейности модели), R или R^2.

Из формул (21)–(23) следует, что значения всех трех показателей R^2, R и r зависят от того, какая типовая форма (8) уравнения регрессии была выбрана, поэтому эти показатели можно рассматривать ***как измерители степени близости выбранной теоретической линии регрессии к фактическим данным***. Качественный характер такой близости может оцениваться в соответствии со шкалой Чэддока.

Индекс корреляции R принимает значения в пределах $0 \le R \le 1$. При этом:

- близость R к единице означает, что связь между признаками достаточно хорошо описывается избранным уравнением корреляционной зависимости (8);
- если R равен или близок к нулю, это означает, что между фактором X и *теоретическими значениями* признака Y либо нет связи, либо если она и существует, то не может быть охарактеризована выбранным для моделирования типовым аналитическим выражением связи (8).

Аналогичные утверждения имеют место и для линейного коэффициента корреляции r, принимающего значения в пределах $-1 \le r \le 1$:

- близость $|r|$ к единице свидетельствует о хорошей аппроксимации фактических данных полученной линейной функцией связи $\hat{y}_x = a_0 + a_1 x$;
- близость $|r|$ к нулю, означает, что уравнение регрессии не может быть линейным.

Пригодность построенной регрессионной модели для практического использования можно оценить и по величине индекса детерминации R^2:

- неравенству $R^2 > 0{,}5$ отвечают значения $R > 0{,}7$ (или $|r| > 0{,}7$), что означает высокую степень тесноты связи признаков в уравнении регрессии. При этом более 50% вариации расчетных значений признака Y объясняется влиянием фактора X, что позволяет считать применение синтезированного уравнения регрессии правомерным;
- при $R \le 0{,}7$ (или $|r| \le 0{,}7$) величина R^2 всегда будет меньше 50%. Это означает, что на долю вариации фактора X приходится меньшая часть по сравнению с прочими признаками, влияющими на вариацию расчетных значений Y. При таких условиях построенная математическая модель связи практического значения не имеет.

В тех случаях, когда рассматриваются альтернативные регрессионные модели, индекс детерминации R^2 используется в качестве ***критерия предпочтительности*** того или иного уравнения регрессии: ***наилучшей считается модель с наибольшим значением R^2***.

3. Так как показатели тесноты связи R или r рассчитываются на основе ограниченной совокупности наблюдаемых эмпирических данных (x_i, y_i), значения которых могли быть искажены влиянием случайных факторов, то найденные по уравнению регрессии показатели тесноты связи r, R проверяются на их неслучайность (значимость).

Для оценки значимости линейного коэффициента корреляции r применяется *t-критерий Стьюдента*, фактическое значение которого рассчитывается по формуле

$$t_r = |r|\sqrt{\frac{n-1}{1-r^2}}. \tag{24}$$

Расчетное значение критерия t_r сравнивается с критическим $t_{табл}$, определяемым по таблице значений t-критерия Стьюдента с учетом заданного уровня значимости α и числа степеней свободы $k = n - 2$. Если $t_{расч} > t_{табл}$, то величина коэффициента корреляции признается значимой.

Для оценки значимости индекса корреляции R применяется ***F-критерий Фишера*** F_R, фактическое значение которого определяется по формуле

$$F_R = \frac{R^2}{1 - R^2} \cdot \frac{n - m}{m - 1}, \qquad (25)$$

где m — число параметров уравнения регрессии.

Расчетная величина F_R сравнивается с критическим табличным значением $F_{табл}(\alpha, k_1, k_2)$, которое определяется по таблице F-критерия с учетом принятого уровня значимости α и числа степеней свободы $k_1 = m - 1$ и $k_2 = n - m$. Если $F_{расч} > F_{табл}$, то величина найденного индекса корреляции R признается значимой.

Значимость показателя тесноты связи R или r означает, что зависимость между признаками X и Y регрессионной модели является статистически существенной, т.е. построенная регрессионная модель в целом адекватна исследуемому процессу. Следовательно, выводы, сделанные на основе регрессионной модели, построенной по данным ограниченной выборки, можно с достаточной вероятностью распространить на всю генеральную совокупность.

В качестве критерия адекватности регрессионной модели в практике экономико-статистического анализа помимо показателя тесноты связи r, R и R^2 используются также следующие показатели:

- ***средняя квадратическая ошибка уравнения регрессии*** σ_ε, представляющая собой среднее квадратическое отклонение эмпирических значений признака Y от теоретических:

$$\sigma_\varepsilon = \sqrt{\frac{\sum_{i=1}^{n} \left(y_i - \hat{y}_{x_i}\right)^2}{n - m}};$$

- ***средняя ошибка аппроксимации*** $\bar{\varepsilon}$, выраженная в процентах:

$$\bar{\varepsilon} = \frac{1}{n} \sum_{i=1}^{n} \frac{\left|y_i - \hat{y}_{x_i}\right|}{y_i} \cdot 100.$$

В адекватных моделях ошибки $\frac{\sigma_\varepsilon}{\bar{y}} \cdot 100$ и $\bar{\varepsilon}$ не должны превышать 12–15%.

Показатели σ_ε и $\bar{\varepsilon}$ могут использоваться при подборе наилучшей модели функции связи: ***более адекватной считается та модель, в которой меньше величина*** σ_ε ***(или*** $\bar{\varepsilon}$***).***

6. ***Экономическая интерпретация построенной регрессионной модели взаимосвязи признаков*** — это, по существу, перевод модели с языка статистики и математики на язык экономики.

Интерпретация уравнения регрессии используется для целей анализа и прогноза взаимосвязей признаков.

1. В процессе анализа прежде всего выясняется, как факторный признак влияет на величину результативного признака. ***Чем больше величина коэффициентов регрессии*** a_k ***при k-й степени фактора X, тем значительнее влияние данного признака на результативный.***

В случае линейного уравнения регрессии $\hat{y}_x = a_0 + a_1 x$ величина коэффициента регрессии a_1 показывает, насколько ***в среднем*** (в абсолютном выражении) изменяется значение результативного признака Y при изменении фактора X на единицу его измерения. Знак при a_1 показывает направление этого изменения.

Если в соответствии с экономической теорией факторный признак должен иметь положительное значение, а коэффициент a_k имеет знак «–», то необходимо проверить расчеты параметров связи, а также возможность ошибок при сборе и обработке информации.

2. С целью расширения возможностей экономического анализа используется коэффициент эластичности $Э = a_1 \frac{\bar{x}}{\bar{y}}$, который показывает, на сколько процентов изменяется в среднем результативный признак при изменении факторного признака на 1%.

3. Анализируя остатки $\varepsilon_i = y_i - \hat{y}_{x_i}$, характеризующие отклонения i-х наблюдений от значений $\hat{y}_{x_i}$, которые следует ожидать в среднем, можно сделать ряд практических выводов об эффективности экономической деятельности рассматриваемых хозяйствующих субъектов и выявить скрытые резервы их развития и повышения деловой активности. При этом наиболее значительный экономический интерес представляют наибольшие и наименьшие положительные и отрицательные отклонения ε_i.

4. Уравнение регрессии $\hat{y}_x = f(x)$ может использоваться для краткосрочного прогнозирования ожидаемых значений результативного признака Y в зависимости от тех или иных значений фактора X.

Прогнозные значения результативного признака Y получают путем подстановки в уравнение регрессии $\hat{y}_x = f(x)$ ожидаемых значений признака X. Уравнение позволяет определять в рассмат-

риваемом временном периоде возможные значения признака Y при значениях X из интервала $[x_{min} - x_{max}]$.

При прогнозировании значений Y за рассмотренными пределами изменения фактора X необходимо соблюдать следующее ограничение: ***нельзя подставлять в уравнение регрессии факторные значения x_i, существенно отличающиеся от тех, на основе которых это уравнение было получено.***

Для выполнения этого требования при выборе факторных производных значений x рекомендуется не выходить за пределы $^1/_3$ размаха вариации фактора x ($x_{max} - x_{min}$) — как сверх минимального (x_{min}), так и сверх максимального (x_{max}) значений, которые имеются в исходных эмпирических данных.

III. Порядок выполнения лабораторной работы

Для выполнения Лабораторной работы № 2 выделяется *Лист 2* рабочего файла, сформированного в персональной папке студента при выполнении Лабораторной работы № 1, и используется следующая информация из Лабораторной работы № 1:

- исходные данные — таблица 1, полученная после удаления аномальных значений (**A4:C33**);
- интервальный ряд распределения единиц совокупности по факторному признаку X — *Среднегодовая стоимость основных производственных фондов* из табл. 7 (**A102:B106**);
- диаграмма рассеяния, расположенная начиная с ячейки **F4**.

1. Подготовительный этап

На данном этапе студент должен проделать следующие обязательные действия, связанные с организацией индивидуальной рабочей среды выполнения Лабораторной работы № 2:

- скопировать необходимую информацию из Лабораторной работы № 1 на *Лист 2 рабочего файла* персональной папки **ФИО**;
- записать в *отчетный файл* Лабораторной работы № 2 индивидуальный вариант исходных данных.

На *Листе 2* рабочего файла персональной папки студента заготовлены макеты таблиц, используемые при выполнении Лабораторной работы № 2 (Приложение 2.1).

Для записи необходимой информации в рабочий и отчетный файлы персональной папки необходимо выполнить следующие действия.

1. Скопировать данные из *Листа 1* в *Лист 2* рабочего файла в соответствии с нижеследующей таблицей:

Лист 1

Номер таблицы	Содержимое таблицы	Адресация содержимого
Табл. 1	Исходные данные	**B4:C33**
–	Диаграмма рассеяния	Начиная с ячейки **F4**
Табл. 7	Интервальный ряд распределения факторного признака X	A102: B106

↓ Копировать в ↓

Лист 2

Номер таблицы	Содержимое таблицы	Адресация содержимого
Табл. 2.1	Исходные данные	**B4:C33**
–	Диаграмма рассеяния	Начиная с ячейки **E4** и **E20**
Табл. 2.2 и табл. 2.3	Интервальный ряд распределения факторного признака X	**B41:C45** и **B52:C56**

2. Скопировать исходные данные варианта из табл. 2.1 *Листа 2* рабочего файла в отчетный файл **Отчет1.doc** в выделенное для этой цели место (в разделе отчета **Постановка задач**).

2. Этап выполнения статистических расчетов

Задание 1. Построение аналитической группировки для выявления корреляционной зависимости результативного признака от факторного и оценка тесноты взаимосвязи признаков

Выполнение *Задания 1* заключается в решении трех задач:

1. Построение аналитической группировки предприятий по факторному признаку *Среднегодовая стоимость основных производственных фондов*.

2. Оценка тесноты связи изучаемых признаков на основе эмпирического корреляционного отношения.

3. Оценка тесноты связи изучаемых признаков на основе линейного коэффициента корреляции (в предположении, что взаимосвязь признаков линейная).

Краткие теоретические сведения

При наличии многочисленных факторов, оказывающих влияние на социально-экономические явления, для исследования связей между ними необходимо ***выделить главные, существенные***

факторы, игнорируя (элиминируя) влияние всех прочих, несущественных факторов

В естественных науках исключение влияния несущественных факторов производится путем эксперимента. При статистическом изучении социально-экономических явлений проведение всеохватывающего эксперимента невозможно ввиду ***массовости данных***, поэтому в статистике элиминирование влияния на изучаемое явление несущественных факторов осуществляется путем применения специальных статистических методов и приемов.

Наличие взаимосвязей признаков устанавливается прежде всего на основе теоретического анализа. При этом выдвигается ***гипотеза (предположение) о наличии корреляционной связи*** между явлениями. Применение к эмпирическим данным ряда статистических методов позволяет подтвердить или опровергнуть выдвинутую гипотезу.

Для выявления наличия связей между признаками применяются самые разнообразные статистические методы — как элементарные, не требующие привлечения математического аппарата, так и более сложные, связанные с проведением математических расчетов (дисперсионный анализ, применение критерия «χ-квадрат»).

К ***элементарным статистическим методам*** выявления взаимосвязей признаков относятся: методы сопоставления параллельных рядов, применения аналитических и корреляционных таблиц, графический метод.

Метод сопоставления взаимосвязанных параллельных рядов является простейшим приемом обнаружения связи между признаками.

Метод заключается в выявлении статистической связи признаков путем простого параллельного сравнения факторных и результативных значений y отдельных единиц совокупности. Для этого значения $x_1, x_2, ..., x_n$ фактора X ранжируется, т.е. располагаются в порядке возрастания (или убывания). Затем строится ряд соответствующих значений результативного признака Y, и путем сопоставления двух построенных рядов выявляется либо наличие (и направление) связи, либо ее отсутствие. Если, например, с возрастанием значений признака X значения признака Y также в целом возрастают при *наличии некоторых отклонений* от этой общей тенденции, то между признаками X и Y ***возможно наличие*** прямой корреляционной связи. Такое заключение имеет место, например, для табл. 6 с ранжированными факторными значениями (построенной по данным табл. 1).

Таблица 6

Взаимосвязанные параллельные ряды

x_i	8	8	8	9	9	9	9	9	10	10	10	10	11	11	11
y_i	18	20	20	20	20	20	24	27	24	24	27	27	27	27	30

К недостаткам метода следует отнести прежде всего невозможность определения количественной меры связи между изучаемыми признаками. Кроме того, при большом числе различных значений y_i, соответствующих одному и тому же значению x_i, восприятие таких параллельных рядов затруднительно, особенно для больших по объему статистических совокупностей. В таких случаях для выявления наличия связи признаков целесообразно пользоваться статистическими таблицами — аналитическими или корреляционными (см. раздел II — «Теоретические основы и методика проведения корреляционно-регрессионного анализа данных», п. 2 «Табличное и графическое представление однофакторных корреляционных связей»).

При выявлении наличия связи ***методом аналитической группировки*** формируется группировка единиц совокупности по факторному признаку X, а затем для каждой выделенной j-й группы рассчитываются средние значения $\bar{y}_j$ результативного признака Y. Если при переходе от одной группы к другой средние значения $\bar{y}_j$ будут изменяться с определенной закономерностью — возрастать или убывать, то между признаками X и Y существует корреляционная связь (как, например, в табл. 5).

При использовании ***метода корреляционных таблиц***, охватывающих два интервальных ряда распределения — факторного и результативного признаков, прослеживают визуально, как именно расположена в таблице основная масса частот повторения в эмпирических данных сочетаний (x_i, y_j). Концентрация частот вдоль диагонали от левого верхнего угла таблицы к правому нижнему (т.е. большему значению X соответствует большее значение Y) означает наличие ***прямой*** корреляционной связи между признаками (как, например, в табл. 4). Если же частоты концентрируются около диагонали от левого нижнего угла к правому верхнему (когда большему значению X соответствует меньшее значение Y), то связь между признаками X и Y ***обратная***.

Интенсивная концентрация частот около диагонали таблицы указывает на факт ***тесной*** корреляционной связи. Так, в корреляционной табл. 4 наблюдается тесная связь между среднегодовой стоимостью основных производственных фондов и выпуском продукции.

Корреляционная таблица дает более правильную характеристику тесноты связи в случае, если число выделенных групп одинаково для обоих изучаемым признаков X и Y (см. табл. 4).

Графический метод состоит в построении корреляционного поля — множества точек (x_i, y_j) в декартовой системе координат (X, Y) (см. рис. 3, 4). По характеру расположения точек корреляционного поля можно сделать вывод о наличии или отсутствии стохастической связи и о характере связи (линейная или нелинейная, а если связь линейная — то прямая или обратная).

При отсутствии тесной связи имеет место беспорядочное расположение точек на графике (рис. 5, *а*). Чем сильнее связь между признаками, тем теснее будут группироваться точки вокруг некоторой определенной линии, выражающей форму связи, т.е. возле линии регрессии (см. рис. 3, 4). Если имеется тенденция равномерного изменения значения результативного признака Y, то можно предположить наличие прямолинейной корреляционной связи, в случае неравномерного изменения Y — наличие криволинейной корреляционной связи.

На рис. 5 представлено графическое изображение некоторых типов парной корреляции признаков.

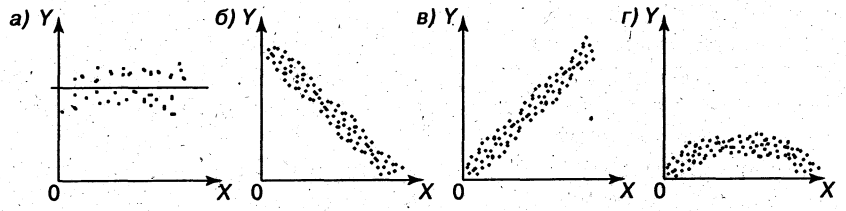


Рис. 5. Разновидности корреляционного поля признаков X и Y:
а — связь между X и Y отсутствует; *б* — связь между X и Y линейная обратная;
в — связь между X и Y линейная прямая; *г* — связь нелинейная

Наличие корреляционной взаимосвязи признаков может быть также установлено по виду графика эмпирической линии связи для условных или групповых средних признака Y (см. рис. 3, 4)

Если бы зависимость Y от X была функциональной, то все точки (x_i, y_j) корреляционного поля были бы расположены на определенной эмпирической линии связи, представляющей графически функциональную зависимость Y от X. При корреляционной связи вследствие различных случайных факторов точки (x_i, y_j) не лежат на одной линии, но все же их расположение обнаруживает определенную тенденцию, которая выражается видом эмпирической линии связи. Так, например, эмпирическая линия связи на рис. 3 по своему виду приближается к прямой линии, а на рис. 4 — к параболической кривой 2-го порядка.

Технология выполнения Задания 1

Задача 1. Построение аналитической группировки предприятий по признаку *Среднегодовая стоимость основных производственных фондов*

Построение аналитической группировки в среде *MS Excel* производится с использованием инструмента **Сортировка** и встроенной функции **СУММ.**

Результаты выполнения аналитической группировки представляются в табл. 2.2, макет которой приведен на рис. 2.1.

Номер группы	Группы предприятий по среднегодовой стоимости основных фондов, млн руб.	Число предприятий	Выпуск продукции, млн руб.	
			Всего	В среднем на одно предприятие
1				
2				
...				
5				
Итого				

Рис. 2.1. Макет таблицы 2.2

Первые три графы этой таблицы были заполнены на **Подготовительном этапе** соответствующими данными, полученными при выполнении Лабораторной работы № 1. Четвертая и пятая графы таблицы являются расчетными.

Для расчета групповых суммарных и средних значений результативного признака Y необходимо знать, какие конкретно единицы наблюдений (предприятия) входят в каждую из сформированных пяти групп. Это достигается ***путем ранжирования (сортировки)*** предприятий по значению факторного признака X. При построении ранжированного ряда предприятий вместе со значением факторного признака x_i перемещаются в соответствующую позицию и номер предприятия, и значение результативного признака y_i.

По отсортированному ряду с учетом известного распределения частот (гр. 3 табл. 2.2) легко установить те значения признака Y, которые попадают в каждую из групп. На рис. 2.2 приведена в качестве примера схема ранжирования и разбиения на три группы ряда девяти предприятий при заданных групповых частотах (2; 4; 3).

а)

№ пр-тия	x_i	y_i
1	25	71
2	32	68
3	12	75
4	17	60
5	34	71
6	39	77
7	22	70
8	28	64
9	26	70

→

б)

№ пр-тия	x_i	y_i
3	12	75
4	17	60
7	22	70
1	25	71
9	26	70
8	28	64
2	32	68
5	34	71
6	39	77

→

в)

№ пр-тия	x_i	y_i
3	12	75
4	17	60
7	22	70
1	25	71
9	26	70
8	28	64
2	32	68
5	34	71
6	39	77

Рис. 2.2. Схема распределения значений признака *Y* по группам:
а — первичный ряд; *б* — ранжированный ряд; *в* — ранжированный ряд с группами, выделенными цветовой заливкой

С использованием описанного алгоритма распределения значений y_i по группам задача построения аналитической группировки решается в три этапа:

1. Ранжирование единиц совокупности по возрастанию факторного признака *Среднегодовая стоимость основных производственных фондов*.

2. Распределение предприятий по группам.

3. Расчет суммарных и средних групповых значений результативного признака *Y*.

Этап 1. Ранжирование единиц совокупности по возрастанию факторного признака *Среднегодовая стоимость основных производственных фондов*.

Для построения ранжированного ряда предприятий применяется инструмент *Excel* **Сортировка**, запуск которого осуществляется последовательностью двух действий:

1. Выделить исходные данные (табл. 2.1).

2. Данные⇒Сортировка.

Задание управляющей информации в диалоговом окне инструмента Сортировка

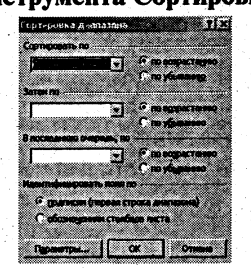


Рис. 2.3. Диалоговое окно инструмента **Сортировка**

В появившемся диалоговом окне инструмента **Сортировка** (рис. 2.3) задаются необходимые параметры.

1. Поле **Сортировать по** — указывается заголовок столбца, по которому будут упорядочиваться (сортироваться) данные, или его обозначение *A*, *B*, *C* и т.д.

2. Переключатель **по возрастанию/по убыванию** — устанавливается в положение, соответствующее направлению сортировки.

3. Поля **Затем по** и **В последнюю очередь по** — активизируются, если необходимо сортировать по двум или трем признакам.

4. Переключатель **Идентифицировать поля по подписям/обозначениям столбцов листа** — устанавливается в зависимости от содержания поля **Сортировать по**: если оно содержит названия признаков (*Среднегодовая стоимость основных производственных фондов* и *Выпуск продукции*), то выбирается положение **по подписям**; если поле содержит обозначение столбцов (*A*, *B*, *C*,...), выбирается положение **по обозначениям столбцов листа.**

5. OK.

Этап 2. Распределение предприятий по группам

Число предприятий в каждой группе определяется распределением частот группировки предприятий по факторному признаку *X* (табл. 2.2, гр. 3). Выбирая поочередно значения частот f_j и отсчитывая в ранжированном ряду соответствующее количество предприятий, легко выделить визуально каждую *j*-ю группу. Для наглядности и удобства работы на следующем этапе целесообразно использовать цветовую заливку групп, выделяя каждую группу различным цветом (желательно контрастным).

Этап 3. Расчет суммарных и средних групповых значений результативного признака *Y*

Для расчета суммарных групповых значений результативного признака *Y* используется функция инструмента **Мастер функций**: **СУММ(Диапазон ячеек)** — математическая функция, вычисляющая сумму значений величин, содержащихся в указанном диапазоне ячеек.

Имя функции (**СУММ**) занесено в макет табл. 2.2. Для вычисления значений функции необходимо в качестве аргументов функции указать диапазон ячеек *j*-й группы, для которой выполняется вычисление указанной функции.

При выполнении этапа 3 для функции **СУММ** ***диапазон ячеек группы распознается по соответствующей цветовой заливке данной группы!***

Для рассмотренного выше примера (рис. 2.2 в) при заданной на рис. 2.4 адресации данных диапазоны ячеек для функции СУММ приведены в таблице на рис. 2.5.

	A	B	C
1	№ пр-тия	x_i	y_i
2	3	12	75
3	4	17	60
4	7	22	70
5	1	25	71
6	9	26	70
7	8	28	64
8	2	32	68
9	5	34	71
10	6	39	77

Рис. 2.4. Ранжированный ряд с группами, выделенными цветовой заливкой

Номер группы	Группы предприятий по среднегодовой стоимости основных фондов, млн руб.	Число предприятий	Выпуск продукции, млн руб.	
			Всего	В среднем на одно предприятие
1	10–20	2	**=СУММ(C2:C3)**	
2	20–30	4	**=СУММ(C4:C7)**	
3	30–40	3	**=СУММ(C8:C10)**	
Итого		9		

Рис. 2.5. Формульный шаблон выходной таблицы аналитической группировки (для рассматриваемого примера)

Задача 2. Оценка тесноты связи изучаемых признаков на основе эмпирического корреляционного отношения.

Для анализа тесноты связи между факторным и результативным признаками рассчитывается показатель η — эмпирическое корреляционное отношение, задаваемое формулой (12):

$$\eta = \sqrt{\frac{\delta^2_{\text{факт}}}{\sigma^2_y}},$$

где σ^2_y — общая дисперсия признака Y;

$\delta^2_{\text{факт}}$ — факторная дисперсия признака Y.

Величина общей дисперсия σ^2_y рассчитывается с помощью функции **ДИСПР** инструмента **Мастер функций**.

Для расчета факторной дисперсии $\delta^2_{\text{факт}}$ используется правило сложения дисперсий

$$\sigma^2_y = \delta^2_{\text{факт}} + \overline{\sigma^2_j},$$

согласно которому

$$\delta^2_{\text{факт}} = \sigma^2_y - \overline{\sigma^2_j},$$

где σ^2_j — внутригрупповая дисперсия j-й группы ***результативных значений*** ($j = 1, 2, ..., 5$).

Внутригрупповые дисперсии σ^2_j для каждой группы рассчитываются с помощью функции **ДИСПР** инструмента **Мастер функций**.

Результаты выполненных расчетов представляются табл. 2.3, макет которой приведен на рис. 2.6.

Номер группы	Группы предприятий по среднегодовой стоимости основных фондов, млн руб.	Число предприятий	Внутригрупповые дисперсии признака Y
1			
2			
...			
5			
Итого			

Рис. 2.6. Макет таблицы 2.3

Поскольку *Excel* не содержит встроенных функций для расчета взвешенных средних, то вычисление средней величины $\overline{\sigma^2_j}$ внутригрупповых дисперсий σ^2_j (гр. 4 табл. 2.3) производится по формуле

$$\overline{\sigma^2_j} = \frac{\sum_{j=1}^{k} \sigma^2_j n_j}{\sum_{j=1}^{k} n_j},$$

где σ^2_j — внутригрупповая дисперсия j-й группы;

n_j — количество предприятий в j-й группе;

k — количество групп ($k = 5$).

При этом для вычисления числителя $\sum_{j=1}^{k} \sigma^2_j n_j$ используется функция **СУММПРОИЗВ**.

Результаты выполненных расчетов представляются в табл. 2.4, макет которой приведен на рис. 2.7.

Общая дисперсия σ_y^2	Дисперсия средняя из внутригрупповых σ_j^2	Факторная дисперсия $\delta^2_{\text{факт}}$	Эмпирическое корреляционное отношение η

Рис. 2.7. Макет таблицы 2.4

Для расчета общей дисперсии σ_y^2, средней $\overline{\sigma_j^2}$ из внутригрупповых дисперсий σ_j^2 и показателя η используются функции инструмента **Мастер функций**:

1. **ДИСПР (Диапазон ячеек)** — статистическая функция, оценивающая дисперсию σ^2.

2. **СУММПРОИЗВ (Диапазон ячеек1, Диапазон ячеек2)** — математическая функция, вычисляющая скалярное произведение

$$a_1 \cdot b_1 + a_2 \cdot b_2 + \ldots + a_k \cdot b_k,$$

где a_i — значение из **Диапазона ячеек1**;

b_i — значение из **Диапазона ячеек2** ($i = 1, 2, \ldots, k$).

3. **КОРЕНЬ (Число)** — математическая функция, вычисляющая квадратный корень из числа, введенного в качестве аргумента.

Имена функций (**ДИСПР**, **СУММПРОИЗВ** и **КОРЕНЬ**) занесены в макеты табл. 2.3 и 2.4.

Для вычисления значения внутригрупповых дисперсий в функции **ДИСПР** (см. табл. 2.3) необходимо в качестве аргумента функции указать диапазон ячеек j-й группы, для которой выполняется вычисление указанной функции (***диапазон ячеек j-й группы распознается по соответствующей цветовой заливке данной группы***).

Для вычисления значения общей дисперсий в функции **ДИСПР** (см. табл. 2.4) в качестве аргумента функции указан диапазон C4:C33 (***диапазон ячеек из табл. 2.1 со значениями y_i признака Y***).

Для вычисления значения средней из внутригрупповых дисперсий в функции **СУММПРОИЗВ** (см. табл. 2.4) в качестве аргумента функции M1 указан диапазон ячеек, содержащих значения внутригрупповых дисперсий (**D52:D56**), а в качестве аргумента функции M2 — диапазон ячеек, содержащих значения частот ряда распределения в группах (**C52:C56**).

Для вычисления значения функции **КОРЕНЬ** (см. табл. 2.4) в качестве аргумента функции введена формула **C63/A63** для расчета отношения $\dfrac{\delta^2_{\text{факт}}}{\sigma_y^2}$.

Для выполнения вычислений следует ввести знак равенства «=» перед именами функций и формулами в табл. 2.3, 2.4.

Задача 3. Оценка тесноты связи изучаемых признаков на основе линейного коэффициента корреляции (в предположении, что взаимосвязь признаков линейная)

В случае линейной связи факторного и результативного признаков оценить тесноту связи можно не только с помощью значения корреляционного отношения η, но и используя линейный коэффициент корреляции r.

Для определения тесноты связи на основе коэффициента r в *Excel* используется инструмент **Корреляция** надстройки **Пакет анализа**, запуск которого осуществляется следующим образом:

Сервис⇒Анализ данных⇒Корреляция⇒OK.

Задание управляющей информации в диалоговом окне инструмента Корреляция

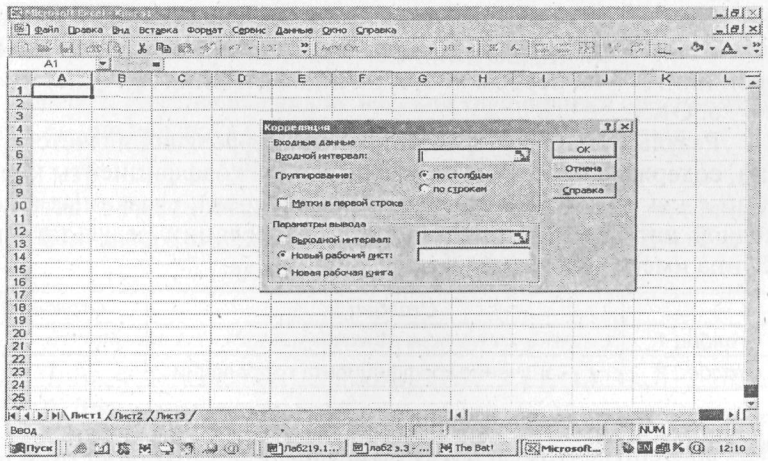


Рис. 2.8. Диалоговое окно инструмента **Корреляция**

В появившемся диалоговом окне инструмента **Корреляция** (рис. 2.8) задаются необходимые параметры.

1. Поле **Входной интервал** — вводится ссылка на диапазоны ячеек, содержащих значения признаков, для которых оценивается теснота линейной связи.

Примечание. В качестве входного интервала может быть указан диапазон, который содержит ряды значений сразу нескольких (более двух) анализируемых признаков. В таком случае показатели корреляции будут рассчитаны между парами всех исследуемых признаков и представлены в единой таблице в виде матрицы.

2. Переключатель **Группирование: по столбцам/строкам** — устанавливается в положение **по столбцам** или **по строкам** в зависимости от того, в каком направлении располагаются анализиру-

емые данные во входном диапазоне — вертикально (по столбцам) или горизонтально (по строкам).

3. Флажок **Метки в первой строке** — устанавливается в активное состояние, если первая строка во входном диапазоне содержит заголовки. Если заголовки отсутствуют, поле не активизируется. В этом случае для данных выходного диапазона будут автоматически созданы стандартные названия — *столбец 1*, *столбец 2* и т.д.

4. Поле **Выходной интервал** — вводится ссылка на ячейку заголовка первого столбца выходной результативной таблицы. Размер выходного диапазона ячеек определяется автоматически.

В случае возможного наложения выходного диапазона на другие данные на экране появится соответствующее сообщение.

5. Переключатели **Новый рабочий лист** и **Новая рабочая книга** — устанавливаются в активное положение при необходимости открытия соответственно нового листа или новой книги. В новом листе результаты анализа располагаются начиная с ячейки A1, в новой книге — на первом листе, начиная с ячейки A1.

6. OK.

Результатом работы инструмента **Корреляция** является таблица, содержащая рассчитанные линейные коэффициенты корреляции *r* для всех комбинаций столбцов (строк), указанных в поле **Входной интервал**. В случае парной корреляции результативная таблица имеет вид:

	Столбец 1	*Столбец 2*
Столбец 1	1	
Столбец 2	*r* — значение коэффициента корреляции	1

Расположение данных на рабочем листе *Excel*

Исходные данные и их статистические характеристики располагаются в таблицах рабочего файла персональной папки студента на ***Листе 2*** в соответствии с табл. 2.1–М.

Таблица 2.1–М

Расположение данных на рабочем листе *Excel*

Данные	Адреса ячеек таблиц на Листе 2 рабочего файла
Исходные данные после удаления «аномальных» значений – таблица 2.1	B4:C33
Интервальный ряд распределения предприятий по факторному признаку: – таблица 2.2 – таблица 2.3	 B41:B45 B52:B56

Данные	Адреса ячеек таблиц на Листе 2 рабочего файла
Групповые значения результативного признака — таблица 2.2: – графа «всего по группе» – графа «в среднем на одно предприятие»	 D41:D45 E41:E45
Значения внутригрупповых дисперсий σ_j^2 — в таблице 2.3	D52:D56
Средняя из внутригрупповых дисперсий $\overline{\sigma_j^2}$ — в таблице 2.4	B63
Значение факторной дисперсии $\delta^2_{факт}$ — в таблице 2.4	C63
Значение эмпирического корреляционного отношения η — в таблице 2.4	D53
Значение линейного коэффициента корреляции *r* — в таблице 2.5	B70

Алгоритмы выполнения Задания 1

Задача 1. Построение аналитической группировки предприятий по признаку *Среднегодовая стоимость основных производственных фондов*

Этап 1. Ранжирование единиц совокупности по возрастанию факторного признака *Среднегодовая стоимость основных производственных фондов*

Алгоритм 1.1. Ранжирование исходных данных.

1. Выделить исходные данные табл. 2.1 (**A4:C33**);
2. **Данные⇒Сортировка**;
3. **Сортировать по⇐**заголовок столбца, по которому выполняется сортировка, т.е. *Среднегодовая стоимость основных производственных фондов*;
4. **по возрастанию/по убыванию** — устанавливается в положение **по возрастанию**;
5. **Затем** и **В последнюю очередь по** — не активизируются;
6. **Идентифицировать поля по подписям/обозначениям столбцов листа** — устанавливается в положение **подписям**;
7. OK.

В результате указанных действий в табл. 2.1 размещаются данные, ранжированные по возрастанию признака *Среднегодовая стоимость основных производственных фондов*.

Этап 2. Распределение предприятий по группам

Алгоритм 1.2. Выделение групп предприятий с помощью заливки контрастным цветом

1. Из всего диапазона отсортированных данных A4:C33 выделить мышью диапазон ячеек ***первой группы***, для чего необходимо отсчитать в ранжированном ряду количество строк, соответствующее числу предприятий ***первой группы*** (гр. 3 табл. 2.2),

2. Нажать на панели инструментов кнопку ;

3. Выбрать цвет по собственному усмотрению;

4. Выполнить действия 1–3 для ***всех групп***, выбирая контрастные цвета для цветовой заливки очередной группы.

Результат работы алгоритмов 1.1 и 1.2 для демонстрационного примера дан в табл. 2.1–ДП.

Таблица 2.1–ДП

Исходные данные

	A	B	C
3	Номер предприятия	Среднегодовая стоимость основных производственных фондов, млн руб.	Выпуск продукции, млн руб.
4	1	94,00	110,00
5	2	107,00	101,00
6	3	134,00	120,00
7	4	157,00	81,00
8	5	163,00	80,00
9	6	167,00	114,00
10	29	167,00	114,00
11	7	173,00	161,00
12	8	173,00	90,00
13	9	177,00	178,00
14	10	179,00	107,00
15	11	200,00	125,00
16	12	201,00	108,00
17	13	205,00	133,00
18	30	205,00	133,00
19	14	208,00	124,00
20	15	212,00	201,00
21	16	213,00	161,00
22	17	214,00	151,00
23	18	216,00	169,00

	A	B	C
24	19	218,00	149,00
25	20	230,00	180,00
26	21	234,00	148,00
27	22	237,00	162,00
28	23	241,00	166,00
29	24	248,00	168,00
30	32	260,00	224,00
31	26	276,00	171,00
32	27	290,00	191,00
33	28	298,00	220,00

Этап 3. Расчет суммарных и средних групповых значений результативного признака *Y — Выпуск продукции*

Алгоритм 1.3. Расчет суммарных групповых значений результативного признака

1. В ячейке (**D41**), выделенной согласно табл. 2.1–М, для суммарного значения результативного признака *Выпуск продукции* ***первой группы***, перед формулой поставить знак равенства «=»;

2. В качестве аргумента функции указать диапазон ячеек из табл. 2.1 с результативными значениями y_i ***первой группы*** (визуально легко определяется по цвету заливки диапазона);

3. **Enter;**

4. Выполнить действия 1–3 поочередно для ***всех групп***, используя цветовые заливки диапазонов.

Алгоритм 1.4. Расчет средних групповых значений результативного признака

1. В ячейке (**E41**), выделенной согласно табл. 2.1–М, для среднего значения признака *Выпуск продукции*, относящихся к ***первой группе***, перед формулой поставить знак равенства «=»;

2. **Enter;**

3. Выполнить действия 1–2 поочередно для ***всех групп.***

Для расчета итоговых сумм в табл. 2.2 (в ячейках **C46, D46** и **E46**) перед формулами необходимо поставить знак равенства «=».

Результат работы алгоритмов 1.3 и 1.4 для демонстрационного примера дан в табл. 2.2–ДП.

Таблица 2.2–ДП

Зависимость выпуска продукции от среднегодовой стоимости основных фондов

	A	B	C	D	E
39	Номер группы	Группы предприятий по стоимости основных фондов	Число предприятий	Выпуск продукции	
40				Всего	В среднем на одно предприятие
41	1	94 – 134,8	3	331,00	110,33
42	2	134,8 – 175,6	6	640,00	106,67
43	3	175,6 – 216,4	11	1590,00	144,55
44	4	216,4 – 257,2	6	973,00	162,17
45	5	257,2 – 298	4	806,00	201,50
46	Итого		30	4340,00	144,67

Задача 2. Оценка тесноты связи изучаемых признаков на основе эмпирического корреляционного отношения

Алгоритм 2.1. Расчет внутригрупповых дисперсий результативного признака

1. В ячейке, выделенной согласно табл. 2.1–М, для внутригрупповых дисперсий ***первой группы*** (**D52**), перед формулой поставить знак равенства «=»;

2. В качестве аргумента функции указать диапазон ячеек из табл. 2.1 с ранжированными значениями y_i ***первой группы*** — визуально легко определяется по цвету заливки диапазона;

3. **Enter**;

4. Выполнить действия 1–3 поочередно для ***всех групп***, используя цветовые заливки диапазонов.

Для расчета итоговых сумм в табл. 2.3 (в ячейках **C57** и **D57**) перед формулами необходимо поставить знак равенства «=».

Результат работы алгоритма 2.1 для демонстрационного примера дан в табл. 2.3–ДП.

Таблица 2.3–ДП

Показатели внутригрупповой вариации

	A	B	C	D
51	Номер группы	Группы предприятий по стоимости основных фондов	Число предприятий	Внутригрупповая дисперсия
52	1	94 — 134,8	3	60,22
53	2	134,8 — 175,6	6	784,56
54	3	175,6 — 216,4	11	821,16
55	4	216,4 — 257,2	6	123,47
56	5	257,2 — 298	4	472,25
57	Итого		30	2261,66

Алгоритм 2.2. Расчет общей, средней из внутригрупповых и факторной дисперсий

В ячейках, выделенных согласно табл. 2.1–М, для общей дисперсии (**A63**), для средней из внутригрупповых дисперсий (**B63**) и для значения факторной дисперсии (**C63**) перед формулами необходимо поставить знак равенства «=».

Алгоритм 2.3. Расчет эмпирического корреляционного отношения

1. В ячейке, выделенной согласно табл. 2.1–М, для эмпирического корреляционного отношения (**D63**) перед формулой поставить знак равенства «=»;

2. **Enter**.

В результате работы алгоритмов 2.2–2.3 *Excel* осуществляет вывод результатов расчета показателей (для демонстрационного примера табл.2.4–ДП).

Таблица 2.4–ДП

Показатели дисперсии и эмпирического корреляционного отношения

	A	B	C	D
62	Общая дисперсия σ_y^2	Средняя из внутригрупповых $\overline{\sigma_j^2}$	Факторная дисперсия $\delta^2_{факт}$	Эмпирическое корреляционное отношение η
63	1450,288889	551,6852535	898,6035354	0,787148735

Задача 3. Оценка тесноты связи изучаемых признаков на основе линейного коэффициента корреляции

Алгоритм 3.1. Расчет линейного коэффициента корреляции

1. **Сервис⇒Анализ данных⇒Корреляция⇒ОК**.

2. **Входной интервал**⇐диапазон ячеек табл. 2.1 со значениями факторного и результативного признаков (**B4:C33**);

3. **Группирование** — по столбцам;

4. **Метки в первой строке** — не активизировать;

5. **Выходной интервал** — адрес ячейки заголовка первого столбца выходной табл. 2.5 (**A68**);

6. **Новый рабочий лист** и **Новая рабочая книга** — не активизировать;

7. ОК.

В результате работы алгоритма 3.1 *Excel* выдает оценку тесноты связи факторного и результативного признаков (для демонстрационного примера табл. 2.5–ДП).

Таблица 2.5–ДП

Линейный коэффициент корреляции признаков

	A	B	C
68		*Столбец1*	*Столбец2*
69	*Столбец1*	1	
70	*Столбец2*	0,753661673	1

Задание 2. Построение однофакторной линейной регрессионной модели связи изучаемых признаков с помощью инструмента Регрессия *надстройки* Пакет анализа

Краткие теоретические сведения

Простейшей формой корреляционной связи признаков является парная линейная корреляция, представляющая собой линейную зависимость результативного признака Y от факторного признака X.

Ее практическое значение состоит в том, что при исследовании взаимосвязи социально-экономических явлений во многих случаях среди всех факторов, влияющих на результативный признак, выделяют один важнейший фактор, который в основном определяет вариацию результативного признака.

Уравнение парной линейной корреляционной связи имеет следующий вид:

$$\hat{y}_x = a_0 + a_1 x,$$

где $\hat{y}_x$ — расчетное теоретическое значение результативного признака Y, полученное по уравнению регрессии;

a_0 — среднее значение признака Y в точке $x = 0$;

a_0, a_1 — коэффициенты уравнения регрессии (параметры связи).

Гипотеза о линейной зависимости между признаками X и Y выдвигается в том случае, если значения обоих признаков возрастают (или убывают) одинаково, примерно в арифметической прогрессии.

Уравнение парной линейной корреляции показывает среднее изменение результативного признака Y при изменении фактора X на одну единицу его измерения, т.е. вариацию признака Y, которая приходится на единицу вариации фактора X. Знак параметра указывает направление этого изменения.

Коэффициенты уравнения a_0, a_1 отыскиваются ***методом наименьших квадратов*** (МНК). Как изложено в разделе II «Теоретические основы и методика корреляционно-регрессионного анализа данных» (п. 3 «Моделирование однофакторных корреляционных связей на основе функциональных зависимостей»), в основу МНК положено требование минимальности сумм квадратов отклонений эмпирических значений y_i от выровненных $\hat{y}_{x_i}$. При линейной зависимости критерий минимизации (11) принимает вид

$$S = \sum_{i=1}^{n}\left(y_i - \hat{y}_{x_i}\right)^2 = \sum_{i=1}^{n}\left(y_i - a_0 - a_1 x_i\right)^2 \to \min.$$

Для нахождения значений параметров a_0, a_1, при которых функция двух переменных S может достигнуть минимума, приравнивают к нулю частные производные S по a_0, a_1 и тем самым получают систему двух уравнений с двумя неизвестными a_0, a_1:

$$\begin{cases} \dfrac{\partial S}{\partial a_0} = 2\sum\limits_{i=1}^{n}\left(y_i - a_0 - a_1 x_i\right)(-1) = 0; \\ \dfrac{\partial S}{\partial a_1} = 2\sum\limits_{i=1}^{n}\left(y_i - a_0 - a_1 x_i\right)(-x) = 0. \end{cases}$$

Сократив каждое уравнение на –2, раскрыв скобки и перенеся члены с x в одну строку, а с y — в другую, для определения a_0, a_1 получают систему:

$$\begin{cases} n a_0 + a_1 \sum\limits_{i=1}^{n} x_i = \sum\limits_{i=1}^{n} y_i; \\ a_0 \sum\limits_{i=1}^{n} x_i + a_1 \sum x_i^2 = \sum\limits_{i=1}^{n} x_i y_i. \end{cases}$$

Эта система называется ***системой нормальных уравнений МНК*** для линейного уравнения регрессии.

Все суммы, необходимые для конкретизации нормальных уравнений, определяют по эмпирическим данным (x_i, y_i).

Решая полученную систему, находят искомые параметры a_0, a_1 — коэффициенты линейного уравнения регрессии.

Расчет коэффициента может быть выполнен по формулам

$$a_0 = \frac{\sum_{i=1}^{n} y_i \sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} y_i x_i \sum_{i=1}^{n} x_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)}; \quad a_1 = \frac{n\sum_{i=1}^{n} y_i x_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}.$$

Иногда эти коэффициенты удобнее вычислять по формулам

$$a_1 = \frac{\overline{xy} - \overline{x}\,\overline{y}}{\overline{x^2} - \overline{x}^2}; \qquad a_0 = \overline{y} - a_1\overline{x}.$$

Построив линейное уравнение регрессии, следует проанализировать качество синтезированной регрессионной модели, оценить адекватность и практическую пригодность модели, дать ее экономическую интерпретацию. Необходимый для этих целей теоретический аппарат КРА изложен в разделе II «Теоретические основы и методика корреляционно-регрессионного анализа данных» (п. 4 «Методика КРА»).

Технология выполнения Задания 2

Регрессионный анализ заключается в определении аналитического выражения связи между факторным признаком *X* и результативным признаком *Y*.

В случае линейной формы связи построение модели средствами *Excel* осуществляется с помощью инструмента **Регрессия** надстройки **Пакет анализа.**

В результате работы инструмента **Регрессия** производится расчет параметров a_0 и a_1 уравнения линейной регрессии $y = a_0 + a_1 x$ и проверка его адекватности исследуемым фактическим данным:

x	y
x_1	y_1
x_2	y_2
...	...
x_n	y_n

Запуск инструмента **Регрессия** осуществляется последовательностью действий:

Сервис⇒Анализ данных⇒Регрессия.

Задание управляющей информации в диалоговом окне инструмента Регрессия

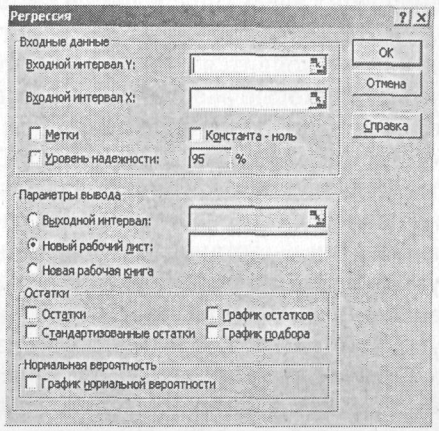


Рис. 2.9. Диалоговое окно инструмента **Регрессия**

В появившемся диалоговом окне инструмента **Регрессия** (рис. 2.9) для построения однофакторной модели связи задаются следующие параметры.

1. Поле **Входной интервал *Y*** — вводится ссылка на диапазон ячеек, содержащих значения ***результативного признака Y.*** Диапазон должен состоять из одного столбца.

2. Поле **Входной интервал *X*** — вводится ссылка на диапазон ячеек, содержащих значения ***факторного признака X.***

3. Флажок **Метки в первой строке/Метки в первом столбце** — устанавливается в активное состояние, если первая строка во входном диапазоне содержит заголовки. Если заголовки отсутствуют, поле не активизируется. В этом случае для данных выходного диапазона будут автоматически созданы стандартные названия — *столбец 1, столбец 2* и т.д.

4. Флажок **Уровень надежности** — устанавливается в активное состояние, если в поле, расположенное напротив флажка, необходимо ввести уровень надежности, ***отличный от уровня 95%, который применяется по умолчанию.*** Установленный уровень надежности используется для проверки значимости (неслучайности) коэффициента детерминации R^2 и коэффициентов регрессии a_0 и a_1, а также для формирования доверительных интервалов с заданным уровнем надежности.

5. Флажок **Константа-ноль** — устанавливается в активное состояние, если требуется, чтобы свободный член a_0 уравнения регрессии был равен нулю (в этом случае линия регрессии проходит через начало координат).

6. Поле **Выходной интервал** — вводится ссылка на ячейку заголовка первого столбца выходной результативной таблицы. Размер выходного диапазона ячеек определяется автоматически.

В случае возможного наложения выходного диапазона на другие данные на экране появится соответствующее сообщение.

7. Переключатели **Новый рабочий лист** и **Новая рабочая книга** — устанавливаются в активное положение при необходимости открытия соответственно нового листа или новой книги. В новом листе результаты анализа располагаются начиная с ячейки A1, в новой книге — на первом листе, начиная с ячейки A1.

8. Флажок **Остатки** — устанавливается в активное состояние, если требуется сформировать выходную таблицу остатков $(y_i - \hat{y}_i)$, представляющую собой разности между фактическими y_i и расчетными $\hat{y}_i$ значениями результативного признака Y.

9. Флажок **Стандартизованные остатки** — устанавливается в активное состояние, если требуется включить в выходную таблицу остатков столбец стандартизованных остатков.

10. Флажок **График остатков** — устанавливается в активное состояние, если требуется вывести на рабочий лист точечные графики зависимости остатков от факторных признаков x_i.

11. Флажок **График подбора** — устанавливается в активное состояние, если требуется вывести на рабочий лист точечные графики зависимости результативных расчетных значений $\hat{y}_i$ от факторных признаков x_i.

12. Флажок **График нормальной вероятности** — устанавливается в активное состояние, если требуется вывести на рабочий лист точечные графики зависимости фактических значений y_i от автоматически формируемых интервалов персентилей. График строится на основе генерируемой таблицы **«Вывод вероятности»**.

13. **ОК**.

В результате работы инструмента **Регрессия** *Excel* формирует следующий набор пяти таблиц.

1. Таблица **Регрессионная статистика** — содержит линейный коэффициент корреляции r, коэффициент детерминации R^2, количество наблюдений n и другие параметры:

Регрессионная статистика

Множественный *R*		= r
R-квадрат		= R^2
Нормированный *R*-квадрат		
Стандартная ошибка		= σ_ε
Наблюдения		= n

2. Таблица **Дисперсионный анализ** — содержит значения факторной и остаточной дисперсий (графа *MS*) и другие параметры дисперсионного анализа:

Дисперсионный анализ

	df	*SS*	*MS*	*F*	Значимость *F*
Регрессия					
Остаток					
Итого					

3. **Результативная таблица** — содержит значения параметров a_0 и a_1 уравнения регрессии и их статистические оценки, включая границы доверительных интервалов для коэффициентов уравнения регрессии:

	Коэффициенты	*Стандартная ошибка*	*t-статистика*	*P-значение*	*Нижние 95%*	*Верхние 95%*	*Нижние 68,3%*	*Верхние 68,3%*
Y-пересечение								
Переменная *X* 1								

4. Таблица **Вывод остатка** — содержит рассчитанные (сглаженные, предсказанные) значения $\hat{y}_i$ результативного признака и значения остатков $\varepsilon_i = y_i - \hat{y}_i$:

Вывод остатка

Наблюдение	*Предсказанное Y*	*Остатки*	*Стандартные остатки*

5. Таблица **Вывод вероятности** — содержит интервалы персентилей и соответствующие им фактические значения y_i результативного признака Y.

Примечание. В анализе результатов выполнения **Задания 2** таблица **Вывод вероятности** не используются. В остальных выходных таблицах используются только отдельные графы, указанные в постановке задачи **Задания 2** и выделенные в приведенных форматах таблиц заливкой.

Между терминологией инструмента **Регрессия** и терминами, принятыми в отечественной статистике, имеется ряд расхождений. Согласование терминологии приводится в табл. 2.2–М.

Таблица 2.2–М

Статистическая интерпретация параметров инструмента Регрессия

Параметр инструмента **Регрессия**	Статистический показатель	Обозначение
Множественный *R*	Линейный коэффициент корреляции	r
R-квадрат	Коэффициент детерминации	R^2
Стандартная ошибка	Среднее квадратическое отклонение расчетных значений от фактических	$\sigma_{\varepsilon e}$
Наблюдения	Число наблюдений	n
MS	Дисперсия факторная и остаточная	$\delta^2_{факт} - \sigma^2_{\varepsilon}$
Y-пересечение	Свободный член регрессии	a_0
Переменная *X* 1	Коэффициент регрессии	a_1
Коэффициенты	Значение коэффициентов уравнения регрессии	a_i
Нижние 95% и Верхние 95%	Соответственно нижние и верхние границы доверительных интервалов для коэффициентов регрессии a_0 и a_1, рассчитанные для уровня надежности $P = 0{,}95$	–
Нижние 68,3% и Верхние 68,3%	Соответственно нижние и верхние границы доверительных интервалов для коэффициентов регрессии a_0 и a_1, рассчитанные для уровня надежности $P = 0{,}683$	–
Предсказанное *Y*	Расчетные значения результативного признака	$\hat{y}_i$
Остатки	Отклонения расчетных значений $\hat{y}_i$ от фактических $(y_i - \hat{y}_i)$	ε_i

Алгоритмы выполнения Задания 2

Алгоритм 1. Расчет параметров уравнения линейной регрессии и проверка его адекватности фактическим данным

1. **Сервис⇒Анализ данных⇒Регрессия⇒ОК;**
2. **Входной интервал** *Y*⇐диапазон ячеек таблицы со значениями признака *Y — Выпуск продукции* **(C4:C33);**
3. **Входной интервал** *X*⇐диапазон ячеек таблицы со значениями признака *X — Стоимость основных фондов* **(B4:B33);**
4. **Метки в первой строке/Метки в первом столбце** — не активизировать;
5. **Уровень надежности**⇐68,3;
6. **Константа–ноль** — не активизировать;

7. **Выходной интервал**⇐адрес ячейки заголовка первого столбца первой выходной результативной таблицы **(A75);**
8. **Новый рабочий лист** и **Новая рабочая книга** — не активизировать;
9. **Остатки** — активизировать;
10. **Стандартизованные остатки** — не активизировать;
11. **График остатков** — не активизировать;
12. **График подбора** — активизировать;
13. **График нормальной вероятности** — не активизировать;
14. **ОК;**
15. Полученный график необходимо расположить после выходных таблиц, начиная с ячейки **A135.**

В результате указанных действий осуществляется вывод в заданный диапазон рабочего файла четырех выходных таблиц и одного графика, начиная с ячейки, указанной в поле **Выходной интервал** (для демонстрационного примера они имеют следующий вид).

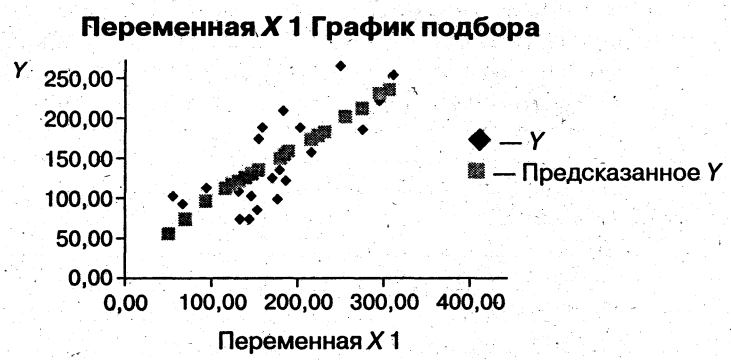
	A	B
77	*Регрессионная статистика*	
78	Множественный *R*	0,753661673
79	*R*-квадрат	0,568005917
80	Нормированный *R*-квадрат	0,552577557
81	Стандартная ошибка	25,90882817
82	Наблюдения	30

	A	B	C	D	E	F
84	Дисперсионный анализ					
85		*df*	*SS*	*MS*	*F*	*Значимость F*
86	Регрессия	1	24713,1801	24713,1801	36,81570256	1,52606E-06
87	Остаток	28	18795,48657	18795,48657		
88	Итого	29	43508,66667			

	A	B	C	D	E
90		Коэффициенты	Стандартная ошибка	*t*-статистика	*P*-значение
91	*Y*-пересечение	21,64454934	21,64454934	1,039615992	0,307412837
92	Переменная *X* 1	0,605324507	0,605324507	6,067594462	1,52606E-06

	F	G	H	I
90	*Нижние 95%*	*Верхние 95%*	*Нижние 68,3%*	*Верхние 68,3%*
91	–21,0028	64,29193	0,432468	42,85664
92	0,400968	0,809681	0,503681	0,706968

	A	B	C
96	Вывод остатка		
97			
98	*Наблюдение*	*Предсказанное Y*	*Остатки*
99	1	78,54505301	31,45494699
100	2	86,4142716	14,5857284
101	3	102,7580333	17,24196671
102	4	116,680497	-35,68049696
	...	...	...
128	30	202,0312525	17,96874754



Интерпретация терминов таблицы в принятых статистических терминах приведена выше в табл. 2.1–М.

Задание 3. Построение однофакторных нелинейных регрессионных моделей связи признаков с помощью инструмента Мастер диаграмм *и выбор наиболее адекватного уравнения регрессии*

Краткие теоретические сведения

В изучении корреляционных связей важным этапом корреляционно-регрессионного анализа является выбор адекватного (наиболее подходящего) эмпирическим данным уравнения регрессии. Именно от адекватности примененной регрессионной модели зависит правильность выводов корреляционно-регрессионного анализа.

В однофакторных регрессионных нелинейных моделях взаимосвязи социально-экономических явлений наиболее часто используются следующие типы математических функций, характеризующих изменение средних значений результативного признака Y в зависимости от изменения факторного признака X:

$\hat{y}_x = a_0 + a_1 \frac{1}{x}$ — гиперболическая,

$\hat{y}_x = a_0 + a_1 \lg x$ — логарифмическая,

$\hat{y}_x = a_0 \cdot x^{a_1}$ — степенная,

$\hat{y}_x = a_0 + a_1 x + a_2 x^2$ — параболическая,

$\hat{y}_x = a_0 + a_1^x$ — показательная.

Каждая из этих моделей отражает определенный вид нелинейной взаимосвязи признаков, когда изменение средних значений Y в зависимости от X происходит неравномерно — с ускорением, замедлением или изменением направления связи.

Задача заключается в том, чтобы из множества альтернативных (конкурирующих) вариантов функций связи $f(x)$ выбрать для моделирования такую, которая лучше других выражает реально существующие связи между изучаемыми признаками, обеспечивает наилучшую аппроксимацию (приближение) и достаточную статистическую достоверность и надежность.

Выбор для регрессионной модели $\hat{y}_x = f(x)$ типа математической функции связи $f(x)$ может опираться на теоретические знания об изучаемом явлении, опыт предыдущих исследований или осуществляться эмпирически — последовательным перебором и оценкой функции различных типов.

В качестве критерия подбора адекватной математической функции связи $f(x)$ используются показатели:

R^2 — индекс детерминации, показывающий, какая доля вариации расчетных значений $\hat{y}_x$ признака Y объясняется влиянием фактора X;

σ_ε^2 — остаточная дисперсия, оценивающая среднее отклонение расчетных значений Y от эмпирических и вычисляемая по формуле (16);

$\bar{\varepsilon}$ — средняя ошибка аппроксимации, выражающая в процентах меру отклонения расчетных значений Y от фактических.

Наилучшей является модель с наибольшим значением показателя R^2 и наименьшим значением показателя σ_ε^2 или $\bar{\varepsilon}$.

Технология выполнения Задания 3

Возможности инструмента **Мастер диаграмм** позволяют быстро производить построение и анализ адекватности регрессионных моделей, базирующихся на использовании различного рода зависимостей: линейной, логарифмической, степенной, экспоненциальной, полиномиальной (2–6 степеней). Для этой цели ис-

пользуется пункт **Добавить линию тренда** меню **Диаграмма**. Будучи ориентированный на построение трендов рядов динамики, пункт **Добавить линию тренда** может быть использован и для построения регрессионных моделей.

Построение моделей осуществляется непосредственно на диаграмме рассеяния, перенесенной из Лабораторной работы № 1 и расположенной на *Листе 2* начиная с ячейки **E4**. Для обращения к пункту **Добавить линию тренда** необходимо выполнить последовательность действий:

1. Выделить мышью диаграмму рассеяния;
2. **Диаграмма⇒Добавить линию тренда**.

В появившемся диалоговом окне **Линия тренда** (рис. 2.10) на вкладке **Тип** задается вид регрессионной модели: линейный, логарифмический и др.

На вкладке **Параметры** (рис. 2.11) задаются параметры тренда, из которых при построении регрессионной модели используются только два последних.

1. Флажок **Показывать уравнение на диаграмме** — устанавливается в активное состояние, если требуется показать на диаграмме рассеяния уравнение регрессии.

2. Флажок **Поместить на диаграмму величину достоверности аппроксимации R^2** — устанавливается в активное состояние, если требуется показать на диаграмме значение коэффициента детерминации R^2.

3. **OK.**

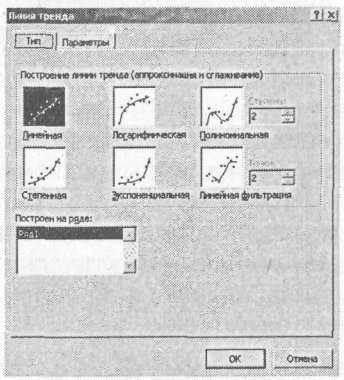


Рис. 2.10. Диалоговое окно **Линия тренда**, вкладка **Тип**

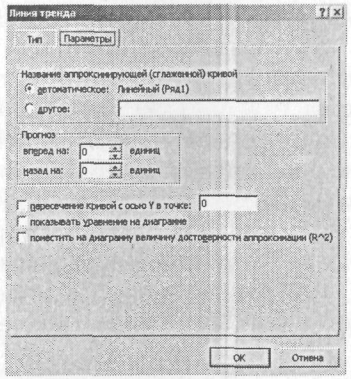


Рис. 2.11. Диалоговое окно **Линия тренда**, вкладка **Параметры**

В результате выполнения указанных действий на диаграмме рассеяния отображается линия и уравнение регрессии, а также коэффициент детерминации R^2.

Уравнение и значение коэффициента детерминация можно перемещать по полю графика, используя «захват мышью».

В Задании 3 нелинейные уравнения регрессии и их графики строятся для следующих видов корреляционной зависимости:

- полиномиальной 2-й степени;
- полиномиальной 3-й степени;
- степенной;
- экспоненциальной.

Выбор наиболее адекватного уравнения нелинейной регрессии определяется максимальным значением коэффициента R^2.

Построение различных моделей регрессии можно осуществлять на одной и той же диаграмме рассеяния. При этом для каждой линии регрессии целесообразно выбирать различный цвет. Для этого необходимо выполнить следующую последовательность действий.

1. Установить курсор на полученную на диаграмме рассеяния линию регрессии.

2. Щелкнуть правой кнопкой мыши на линии регрессии и в появившемся контекстном меню выбрать пункт **Формат линии тренда**.

В появившемся диалоговом окне **Формат линии тренда** (рис. 2.12) на вкладке **Вид** задается тип, цвет и толщина линии тренда.

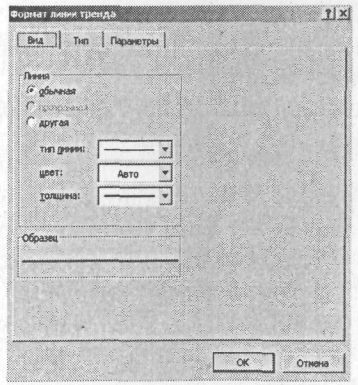


Рис. 2.12. Диалоговое окно **Формат линии тренда**, вкладка **Вид**

Максимальное значение коэффициента детерминации R^2 определяет вид искомого уравнения регрессии и его график, которые следует расположить на отдельной диаграмме рассеяния.

Для этой цели используется вторая скопированная из Лабораторной работы № 1 диаграмма рассеяния, находящаяся на *Листе 2* начиная с ячейки **E20**.

Алгоритмы выполнения Задания 3

Алгоритм 1. Построение уравнений регрессионных моделей для различных видов зависимости признаков с использованием средств инструмента Мастер диаграмм

1. Выделить мышью диаграмму рассеяния, расположенную начиная с ячейки **E4**, и увеличить масштаб диаграммы на весь экран;

2. **Диаграмма⇒Добавить линию тренда;**

3. Выбрать вкладку **Тип**, задать вид регрессионной модели — *полином 2-го порядка*;

4. Выбрать вкладку **Параметры** и выполнить действия:

1. Переключатель **Название аппроксимирующей кривой: автоматическое/другое** — установить в положение **автоматическое**;
2. Поле **Прогноз вперед на** — не активизировать;
3. Поле **Прогноз назад на** — не активизировать;
4. Флажок **Пересечение кривой с осью *Y* в точке** — не активизировать;
5. Флажок **Показывать уравнение на диаграмме** — активизировать;
6. Флажок **Поместить на диаграмму величину достоверности аппроксимации R^2** — активизировать;
7. **OK;**
8. Установить курсор на линию регрессии и щелкнуть правой клавишей мыши;
9. В появившемся диалоговом окне **Формат линии тренда** выбрать тип, цвет и толщину линии;
10. **OK;**
11. Вынести уравнение и коэффициент R^2 за корреляционное поле. При необходимости уменьшить размер шрифта.

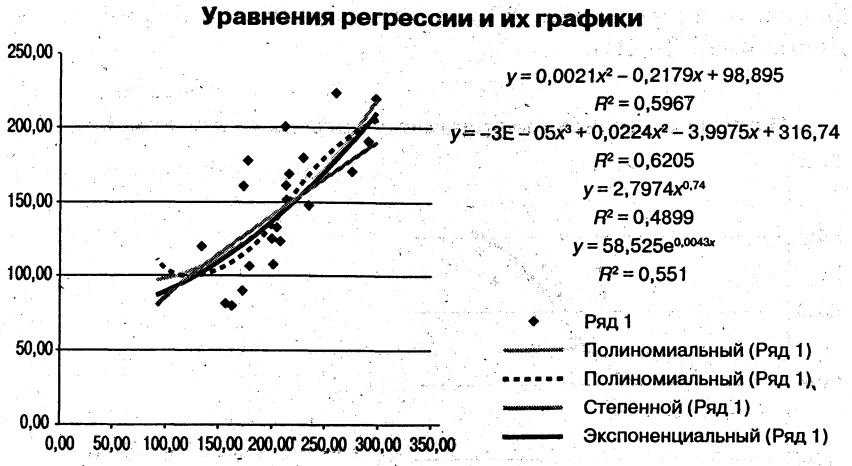
5. Действия 3–4 (в п. 4 — шаги 1–11) выполнить поочередно для следующих видов регрессионных моделей:

- полином 3-го порядка;
- степенная;
- экспоненциальная.

Переместить Диаграмму 2.1 в конец рабочего файла, начиная с ячейки **A155**.

В результате указанных действий для выбранных видов моделей регрессии осуществляется вывод на диаграмму рассеяния четырех уравнений регрессии, их графиков и значений соответствующих коэффициентов детерминации R^2 (для демонстрационного примера это диаграмма 2.1–ДП).

Диаграмма 2.1–ДП



Алгоритм 2. Построение наиболее адекватного уравнения регрессии

1. Путем визуального анализа значений R^2 выбрать по максимальной величине R^2 наиболее адекватное уравнение регрессии;

2. Выделить диаграмму рассеяния, расположенную начиная с ячейки **E20**;

3. **Диаграмма⇒Добавить линию тренда;**

4. Выбрать вкладку **Тип** и задать вид наиболее адекватной нелинейной регрессионной модели;

5. Выбрать вкладку **Параметры:**

1. Переключатель **Название аппроксимирующей кривой: автоматическое/другое** — установить в положение **автоматическое**;
2. Поле **Прогноз вперед на** — не активизировать;
3. Поле **Прогноз назад на** — не активизировать;
4. Флажок **Пересечение кривой с осью *Y* в точке** — не активизировать;
5. Флажок **Показывать уравнение на диаграмме** — активизировать;
6. Флажок **Поместить на диаграмму величину достоверности аппроксимации R^2** — активизировать;
7. **OK.**

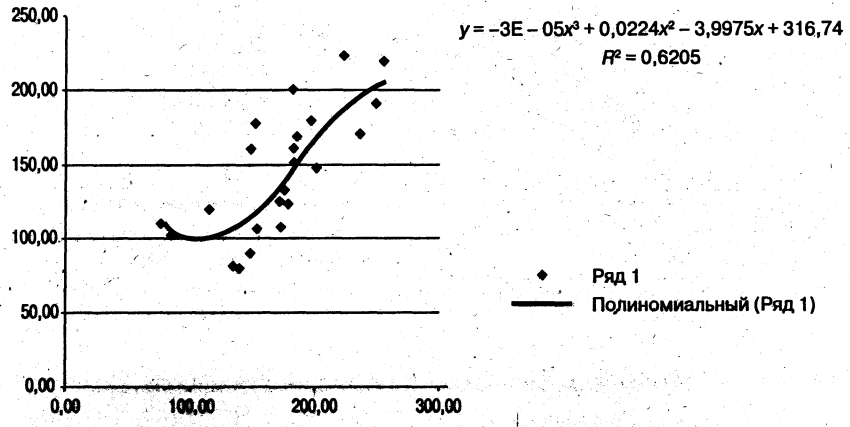
Переместить Диаграмму 2.2 в конец рабочего файла, начиная с ячейки **A190**.

В результате указанных действий осуществляется вывод на диаграмму рассеяния уравнения регрессии для выбранной наи-

более адекватной модели регрессии, ее графика и значения коэффициента детерминации R^2 (для демонстрационного примера это Диаграмма 2.2–ДП).

Диаграмма 2.2–ДП

Наиболее адекватное уравнение регрессии и его график



Приложение 2.1

ВСЕРОССИЙСКИЙ ЗАОЧНЫЙ ФИНАНСОВО-ЭКОНОМИЧЕСКИЙ ИНСТИТУТ

КАФЕДРА СТАТИСТИКИ

ОТЧЕТ

о результатах выполнения
компьютерной лабораторной работы № 2

Автоматизированный корреляционно-регрессионный анализ взаимосвязи статистических данных в среде MS Excel

Вариант №____

Выполнил: ст. III курса гр._____

_____
Ф.И.О.

Проверил:_____
Должность Ф.И.О.

Москва, 2006 г.

Приложение 2.2

Расположение макетов результативных таблиц в рабочем файле персональной папки студента

	A	B	C	D	E
37					Таблица 2.2
39	Зависимость выпуска продукции от среднегодовой стоимости основных фондов				
39	Номер группы	Группы предприятий по стоимости основных фондов	Число предприятий	Выпуск продукции	
40				Всего	В среднем на одно предприятие
41	1				D41/C41
42	2				D42/C42
43	3				D43/C43
44	4				D44/C44
45	5				D45/C45
46	Итого		СУММ (C41:C45)	СУММ (D41:D45)	D46/C46
47					
48					
49				Таблица 2.3	
50	Показатели внутригрупповой вариации				
51	Номер группы	Группы предприятий по стоимости основных фондов	Число предприятий	Внутригрупповая дисперсия	
52	1			ДИСПР()	
53	2			ДИСПР()	
54	3			ДИСПР()	
55	4			ДИСПР()	
56	5			ДИСПР()	
57	Итого		СУММ (C52:C56)	СУММ(D52:D56)	
58					
59					
60				Таблица 2.4	
61	Показатели дисперсии и эмпирического корреляционного отношения				
62	Общая дисперсия	Средняя из внутригрупповых дисперсий	Межгрупповая дисперсия	Эмпирическое корреляционное отношение	
63	ДИСПР (C4:C33)	СУММПРОИЗВ (D52:D56;C52:C56)/C46	A63-B63	КОРЕНЬ (C63/A63)	
64					
65					

66			Таблица 2.5		
67	Линейный коэффициент корреляции признаков				
68		Столбец1	Столбец2		
69	Столбец1				
70	Столбец2				
71					
72					

	A	B	C	D	E	F
73	Выходные таблицы					
74						
75	Вывод итогов					
76						
77	*Регрессионная статистика*					
78	Множественный *R*					
79	*R*-квадрат					
80	Нормированный *R*-квадрат					
81	Стандартная ошибка					
82	Наблюдения					
83						
84	Дисперсионный анализ					
85		*df*	*SS*	*MS*	*F*	Значимость *F*
86	Регрессия					
87	Остаток					
88	Итого					
89						

	A	B	C	D	E	F	G	H	I
90		***Коэффициенты***	***Стандартная ошибка***	***t-статистика***	***P-значение***	***Нижние 95%***	***Верхние 95%***	***Нижние 68,3%***	***Верхние 68,3%***
91	***Y-пересечение***								
92	***Переменная X 1***								
93									
94									
95									
96	Вывод остатка								
97									
98	*Наблюдение*	*Предсказанное Y*	*Остатки*						
99									
...									

Литература

1. Громыко Г.Л. Теория статистики: Учебник. — М.: ИНФРА-М, 2005.
2. Ефимова М.Р., Петрова Е.В., Румянцев В.Н. Общая теория статистики: Учебник. — М.: ИНФРА-М, 2004.
3. Козлов А.Ю., Мхитарян В.С., Шишов В.Ф. Статистические функции MS Excel в экономико-статистических расчетах: Учеб. пособие для вузов/Под ред. проф. В.С. Мхитаряна. — М.: ЮНИТИ-ДАНА, 2003.
4. Козлов А.Ю., Шишов В.Ф. Пакет анализа MS Excel в экономико-статистических расчетах: Учеб. пособие для вузов/Под ред. проф. В.С. Мхитаряна. — М.: ЮНИТИ-ДАНА, 2003.
5. Макарова Н.В., Трофимец В.Я. Статистика в Excel: Учеб. пособие. — М.: Финансы и статистика, 2002.
6. Общая теория статистики: Статистическая методология в изучении коммерческой деятельности: Учебник/Под ред. О.Э. Башиной, А.А. Спирина. — М.: Финансы и статистика, 2005.
7. Практикум по статистике: Учеб. пособие для вузов/Под ред. В.М. Симчеры; ВЗФЭИ. — М.: Финстатинформ, 1999.
8. Салин В.Н., Медведев В.А., Кудряшова С.И., Шпаковская Е.П. Макроэкономическая статистика: Учеб. пособие. — М.: Дело, 2000.
9. Статистика: Учебник/Под ред. И.И. Елисеевой. — М.: ТК Велби: Проспект, 2002.
10. Теория статистики: Учебник/Под ред. Р.А. Шмойловой. — М.: Финансы и статистика, 2004.

Содержание

Статистика. Компьютерные лабораторные работы: Методические указания к лабораторной работе № 2 «Автоматизированный корреляционно-регрессионный анализ взаимосвязи статистических данных в среде *MS Excel*». — М.: Вузовский учебник, 2006.

Сдано в набор 30.07.2006. Подписано в печать 15.08.2006.
Формат 60x88/16. Бумага типографская № 2. Гарнитура «*Newton*».
Печать офсетная. Усл. печ. л. 4,41. Уч.-изд. л. 4,5.
Тираж 10 000 экз. Заказ № 1/22-06.

Издательский Дом «Вузовский учебник»
127247, Москва, ул. С. Ковалевской, д. 1, стр. 52

Отпечатано в полном соответствии
с качеством предоставленных диапозитивов в ОАО «Домодедовская типография»
142001, г. Домодедово, Каширское шоссе, 4, корп. 1.

Заказ 965.