МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное бюджетное образовательное
учреждение высшего образования
«ПЕРМСКИЙ ГОСУДАРСТВЕННЫЙ
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ»

ЭКОНОМЕТРИКА
в схемах и таблицах

*Допущено методическим советом Пермского
государственного национального исследовательского
университета в качестве учебно-методического пособия
для студентов, обучающихся по направлениям подготовки
бакалавров «Экономика» и «Менеджмент», «Бизнес-информатика»,
«Прикладная математика в экономике»*

Пермь 2018

Составители: канд. физ-мат. наук, доц. ***М.В. Радионова***;
канд. физ-мат. наук, доц. ***Н.В. Фролова***

УДК 330.43 (075.8)
ББК 65в611я73
Э 40

Рецензенты:
д-р экон. наук, профессор кафедры экономики и управления
промышленным производством Пермского национального
исследовательского политехнического университета ***Ж.А. Мингалева***;
кафедра высшей математики НИУ ВШЭ – Пермь
(заведующий кафедрой – канд. физ.-мат. наук,
профессор ***А.П. Иванов***)

Э40 **Эконометрика**: учеб.-метод. пособие / сост. М.В. Радионова, Н.В. Фролова. – Пермь: Изд-во Перм. нац. исслед. политехн. ун-та, 2018. **–** 187 с.

ISBN 978-5-398-02140-0

Учебно-методическое пособие подготовлено авторами на основе опыта преподавания эконометрики студентам экономического факультета Пермского государственного национального исследовательского университета и факультета экономики, менеджмента и бизнес-информатики Национального исследовательского университета «Высшая школа экономики» (Пермь).

Изложены основные сведения по разделам курса «Эконометрика» в доступной табличной форме. Помимо необходимого теоретического материала приведено много примеров практического применения теоретических результатов. Большое количество практических примеров, приложений и статистических таблиц, а также заданий для самостоятельной работы студентов призвано помочь им освоить ключевые понятия и идеи статистического анализа, научиться использовать их при решении прикладных задач.

Пособие предназначено для студентов экономического факультета, обучающихся по направлениям подготовки бакалавров «Экономика» и «Менеджмент», «Бизнес-информатика», «Прикладная математика в экономике», а также может быть использовано студентами, аспирантами, преподавателями экономических вузов, менеджерами предприятий.

УДК 330.115 (075.8)
ББК 65в611я73

Печатается по решению редакционно-издательского совета
Пермского государственного национального исследовательского университета

ISBN 978-5-398-02140-0

СОДЕРЖАНИЕ

ВВЕДЕНИЕ

Эконометрика является областью знаний, которая охватывает вопросы применения статистических методов к теоретическим моделям, описывающим реальные экономические процессы. С помощью моделей можно получить обширную информацию об экономических системах, объяснить те или иные явления или процессы, но никогда не удается получить всю информацию и однозначно определить истинный механизм экономического явления или процесса.

Цель изучения курса – помочь студентам составить представление о содержании эконометрики как научной дисциплины и как области практической деятельности, позволяющей анализировать социально-экономические процессы.

В настоящем пособии даются основные понятия, модели и методы эконометрики, приводятся примеры. Пособие предназначено для студентов очно-заочного отделения экономического факультета Пермского государственного национального исследовательского университета и факультета экономики, менеджмента и бизнес-информатики Национального исследовательского университета Высшая школа экономики – Пермь.

Задачами изучения дисциплины являются:

- знакомство с современным представлением об эконометрике;
- овладение основными приемами выборочных исследований;
- изучение теории и алгоритмов метода наименьших квадратов;
- освоение методов корреляционного и регрессионного анализа, применяемых для построения различных эконометрических моделей;
- овладение приемами практического анализа временных рядов;
- овладение основными приемами оценивания систем одновременных эконометрических уравнений;
- приобретение навыков использования результатов экономического анализа для прогноза и принятия обоснования экономических решений;

- формирование навыков проведения сложных компьютерных расчетов с использованием эконометрических моделей.

Основные темы, рассмотренные в пособии:

- Основные понятия эконометрики
- Однофакторная линейная регрессионная модель
- Общая линейная модель наблюдений при классических предположениях
- Анализ линейной модели наблюдений при отклонениях от классических предположений
- Модели стационарных и нестационарных временных рядов

Для работы с данным пособием необходимы базовые знания некоторых разделов следующих учебных дисциплин: высшая математика, теория вероятностей и математическая статистика, общая и экономическая статистика.

По курсу «Эконометрика» студент выполняет одну контрольную работу. В данном издании приводится 25 вариантов контрольной работы (номера вариантов с 1 по 25). Обязательными требованиями к ее оформлению являются следующие:

1) при решении каждой задачи необходимо полностью приводить ее условие;

2) решение задачи должно сопровождаться необходимыми формулами, таблицами, графиками, положениями и выводами;

3) при выполнении работы с использованием персонального компьютера следует указывать название и версию программного обеспечения, которое используется студентом;

4) в тексте работы желательно приводить результаты промежуточных расчетов (за исключением работ, расчеты в которых выполнены на персональных компьютерах и сопровождаются распечатками);

5) в конце выполненной работы указать все информационные источники, используемые при написании.

Глава 1. ОСНОВНЫЕ ПОНЯТИЯ ЭКОНОМЕТРИКИ

Введение в эконометрическое моделирование. Цели и задачи эконометрики. Виды данных. Классификация переменных в эконометрических моделях. Характеристика переменных, входящих в модели. Основные виды эконометрических моделей. Этапы эконометрического исследования. Основные виды моделей зависимости. Статистическая и функциональная зависимость. Корреляционный анализ. Проверка статистических гипотез. Общая схема решения задачи проверки статистических гипотез. Основы построения доверительных интервалов.

Основные понятия эконометрики

<table>
<tr><td colspan="3">Эконометрика (Econometrics) – это наука, дающая обоснованный анализ и прогноз социально-экономических процессов и явлений на основе статистических данных
Эконометрика – «измерения в экономике» (Рагнар Фриш, норвежский экономист, 1926)
Термин «эконометрика» имеет в своей основе два слова: «экономика» и «метрика» (от греч. metron – «метод расчета определения расстояния между двумя точками в пространстве»). В общем случае эконометрику можно определить как науку об экономических измерениях</td></tr>
<tr><th>Методы</th><th>Где используются</th><th>Эконометрика базируется на</th></tr>
<tr><td>Корреляционно-регрессионный анализ;

Анализ временных рядов;

Системы одновре-</td><td>макроуровень (модели национальной экономики, межстрановой анализ);
мезоуровень (модели региональной экономики, отраслей, секторов);
микроуровень (модели</td><td>Экономическая теория;

Экономическая статистика;

Теория вероятностей и математическая</td></tr>
</table>

менных уравнений; Статистические методы классификации снижения размерности.	поведения потребителя, домашних хозяйств, предприятий, фирм).	статистика; Экономическая кибернетика.

Цель эконометрики	количественный анализ экономических объектов, на основе статистических данных
Задачи эконометрики	
Спецификация модели и параметризация модели	Построение эконометрической модели для эконометрического анализа
Верификация модели	Проверка качества параметров модели и самой модели
Прогнозирование	Нахождение оценки будущего значения зависимой переменной при заданном значении независимой переменной

Прикладные цели эконометрики	
моделирование различных возможных сценариев социально-экономического развития анализируемой системы (многовариантные сценарные расчеты, ситуационное моделирование)	прогноз социально-экономических показателей, характеризующих состояние и развитие анализируемой системы

Типы данных, используемых в эконометрическом моделировании	
Перекрестные данные (cross-section data)	это данные по какому-либо экономическому показателю, полученные для однотипных объектов и относящиеся к одному периоду времени (либо временной промежуток не имеет значения).

Временные ряды (time series)	данные об одном объекте, процессе за несколько последовательных моментов времени, т.е. характеризуется динамика развития изучаемого объекта, процесса.
Панельные данные	данные по разным объектам за последовательные периоды времени (т.е. перекрестные данные + временные ряды)

Переменные в эконометрических моделях и их виды	
Переменная	обозначение какого-либо показателя. Совокупность переменных (показателей) определяет состояние экономического объекта.
Характеристика переменных, входящих в модели	
Y	основной экономический показатель, поведение которого изучается (зависимая переменная, эндогенная переменная, отклик, результативный признак)
$X_1, X_2, \ldots X_k$	экономические показатели (независимые переменные, экзогенные переменные, факторы, факторные признаки, предикторы), которые оказывают (а возможно, и нет, исследователь тоже может ошибаться) влияние на поведение Y. В том случае, если в модели будет только одна независимая переменная, вместо X_1 в целях упрощения будет использоваться обозначение X
t	независимая переменная, интерпретируемая как время
ε	случайная компонента регрессионной модели (используется для описания эффекта влияния на Y тех экономических показателей и внешних обстоятельств, учесть влияние которых не представляется возможным или целесообразным)

Основные виды эконометрических моделей	
Регрессионные модели с одним уравнением	определяют статистическую зависимость между переменной Y и влияющими факторами $X_1, X_2, ... X_k$
Модели временных рядов	учитывают зависимость результативной переменной Y от фактора времени
Системы одновременных эконометрических уравнений	это системы взаимозависимых регрессионных уравнений

Основные этапы эконометрического моделирования	
Постановочный этап	Определение конечных целей моделирования, набора участвующих в модели переменных, их роли
Априорный этап	Предмодельный анализ экономической сущности изучаемого явления, формирование и формализация априорной информации
Информационный этап	Сбор необходимой статистической информации, т.е. измерение значений участвующих в модели факторов и показателей изучаемого явления
Спецификация модели	Выбор общего вида модели, в том числе состава и формы входящих в нее связей
Идентификация модели	Статистический анализ модели и в первую очередь статистическое оценивание неизвестных параметров модели
Верификация модели	Сопоставление модельных и реальных данных, проверка адекватности модели, оценка точности модели

Статистическая зависимость признаков

Виды зависимостей между признаками	
Функциональная зависимость	когда каждому значению одной переменной соответствует точно определенное значение другой переменной
Статистическая (или стохастическая, вероятностная) зависимость	каждому значению одной переменной соответствует определенное (условное) распределение другой переменной, т.е. когда каждому значению одной переменной соответствует не какое-то определенное, а множество возможных значений другой переменной
Корреляционная зависимость Частный случай статистической зависимости	связь, при которой каждому значению независимой переменной X соответствует *определенное математическое ожидание (среднее значение)* зависимой переменной Y

Корреляционный анализ	
Сущность корреляционного анализа	это совокупность методов оценивания степени тесноты статистической связи между анализируемыми переменными
Виды корреляции	
Парная	характеризует взаимосвязь двух переменных на фоне действия остальных показателей и является самым распространенным показателем тесноты связи при статистическом анализе данных

Частная	характеризует взаимосвязь несколькими переменными с исключением влияния на их связь всех остальных переменных
Множественная	характеризует взаимосвязь между результативным и несколькими факторными признаками

Регрессионный анализ	
Сущность регрессионного анализа	это совокупность методов для определения аналитической формы зависимости результативного признака от одного или нескольких факторных признаков
Функции регрессии (синоним – **уравнение регрессии**)	это функция, характеризующая зависимость среднего значение зависимой переменной Y от независимых переменных $X_1, \ldots, X_k$
Обозначение функции регрессии $M(Y \mid X_1, X_2, ... X_k) =$ $= f(X_1, X_2, ... X_k, \boldsymbol{\beta}) \equiv Y_X$	$f(X_1, X_2, ... X_k, \boldsymbol{\beta})$ – это некоторая неслучайная функция, известная с точностью до одномерного или многомерного параметра $\boldsymbol{\beta}$. Неизвестный параметр (одномерный или многомерный) $\boldsymbol{\beta} = (b_0, b_1, ..., b_k)^T$ Вместо обозначения среднего уровня будем использовать запись Y_X.

Виды регрессионных моделей	
Аддитивная регрессионная модель	$Y = Y_X + \varepsilon$
Мультипликативная регрессионная модель	$Y = Y_X \cdot \delta$
Регрессионная ошибка (погрешность) – случайные компоненты регрессионной модели	ε, δ
По количеству переменных в модели	
Парная	отражают зависимость одного результирующего признака Y от одной независимой переменной X

Множест-венная	отражают зависимость одного результирующего признака Y от нескольких независимых переменных $X_1, X_2, \dots X_k$
По форме	
Линейные по параметрам	Парная линейная регрессионная модель $Y = a + b \cdot X + \varepsilon$
	Парная гиперболическая регрессионная модель $Y = a + \dfrac{b}{X} + \varepsilon$
	Парная квадратичная регрессионная модель $Y = a + bX + cX^2 + \varepsilon$
	Множественная линейная регрессионная модель $Y = b_0 + b_1X_1 + b_2X_2 + \ldots b_kX_k + \varepsilon$
Нелинейные по парамет-рам	Показательная регрессионная модель $Y = a \cdot b^X \cdot \varepsilon$
	Экспоненциальная регрессионная модель $Y = a \cdot e^{bX} \cdot \varepsilon$
	Степенная регрессионная модель $Y = a \cdot X^b \cdot \varepsilon$

Проверка статистических гипотез	
Статистическая гипотеза	любое утверждение о виде или свойствах распределения исследуемой случайной величины (величин).
Задача проверки статистической гипотезы	заключается в подтверждении или опровержении этого предположения на основании выборочных (экспериментальных) данных. Проверка статистической гипотезы означает проверку соответствия выборочных данных выдвинутой гипотезе

Основная (нулевая) гипотеза	утверждение, подлежащее проверке. Обозначение: H_0.
Конкурирующая (альтернативная) гипотеза	утверждение, отличное от нулевой гипотезы. Обозначение: H_1
Статистический критерий (тест, решающее правило)	правило, с помощью которого решается задача проверки гипотез Проверка гипотез завершается принятием одного из двух решений: • *принять нулевую гипотезу* (имеющиеся результаты наблюдений согласуются с утверждением нулевой гипотезы) • *отклонить нулевую гипотезу* (имеющиеся результаты наблюдений не согласуются с утверждением нулевой гипотезы) Подбирается ***критерий проверки гипотезы H*₀** $K = g\left(X_1, X_2, ..., X_n\right)$, являющийся мерой расхождения между предполагаемыми, гипотетическими и опытными, полученными по выборке значениями или характеристиками исследуемой случайной величины. Критерий $K = g\left(X_1, X_2, ..., X_n\right)$ – функция элементов выборки
Статистика критерия	$K_{выч} = K\left(x_1, x_2, \ldots, x_n\right)$ в качестве аргументов этой функции подставлены произвольные результаты наблюдений $X_1, X_2, \ldots, X_n$
Ошибка первого рода	отклонить верную $\boldsymbol{H_0}$
Ошибка второго рода	принять неверную $\boldsymbol{H_0}$

Уровень значимости	наибольшее значение вероятности, несовместимое с признанием случайности экспериментально вычисленного значения статистики критерия Обозначение: α Предпочтительные значения α : 0,2, 0,1, 0,05
Мощность критерия	вероятность отклонения ложной гипотезы Обозначение: $1-\beta$
р-значение статистики критерия	эмпирический уровень значимости критерия

Общая схема решения задачи проверки статистических гипотез	
Шаг 1	Задаемся некоторым уровнем значимости α – максимально возможным значением вероятности ошибки первого рода
Шаг 2	Исходя из заданного уровня значимости с помощью статистики критерия $K(\cdot)$ множество всех возможных результатов наблюдений разбивается на область принятия и критическую область
Шаг 3	Находят вычисленное значение статистики критерия
Шаг 4	Если окажется, что вычисленное значение статистики критерия принадлежит области принятия, то принимается нулевая гипотеза, если окажется, что это значение принадлежит критической области, то нулевая гипотеза отклоняется

Теоретические основы построения доверительных интервалов

Пусть $X_1, X_2, \ldots, X_n$ – независимая повторная выборка из совокупности X с неизвестным параметром распределения θ	
Доверительный интервал	это интервал с границами $U_1(X_1, \ldots, X_n)$ и $U_2(X_1, \ldots, X_n)$, в который попадает неизвестный параметр θ с заданной доверительной вероятностью γ, т.е. $P(U_1 < \theta < U_2) = \gamma$

$U_1(X_1,\ldots,X_n)$ и $U_2(X_1,\ldots,X_n)$	доверительные пределы
γ	доверительная вероятность (надежность) задается исследователем обычно 0,9, 0,95 или 0,98
Виды доверительных интервалов	
Двусторонний	(U_1, U_2)
Правосторонний	$(-\infty, U_2)$
Левосторонний	(U_1, ∞)

Тесты для проверки знаний

1. Эконометрика – это

1) специальный раздел математики, посвященный анализу экономической информации.

2) это наука, дающая обоснованный анализ и прогноз социально-экономических процессов и явлений на основе статистических данных

3) наука, которая осуществляет качественный анализ взаимосвязей экономических явлений и процессов.

4) раздел экономической теории, связанный с анализом статистической информацией;

5) среди приведенных ответов нет правильного.

2. Экзогенные переменные – это

1) переменные, значения которых задаются извне модели;

2) переменные, которые являются предметом объяснения в эконометрической модели;

3) эндогенные переменные, измеренные в прошлые моменты времени;

4) лаговые эндогенные переменные;

5) среди приведенных ответов нет правильного.

3. Эндогенные переменные – это

1) переменные, значения которых задаются извне модели;

2) переменные, которые являются предметом объяснения в эконометрической модели;

3) экзогенные переменные, измеренные в прошлые моменты времени;

4) лаговые экзогенные переменные;

5) среди приведенных ответов нет правильного.

4. Переменные, описывающие количественные характеристики экономического явления или процесса, называются …

1) Фиктивными;

2) Несущественными;

3) Параметрами;

4) Объясняющими;

5) Среди приведенных ответов нет правильного.

5. Расставьте основные этапы эконометрического моделирования в правильном порядке:

1) идентификация модели;

2) информационный этап;

3) параметризация;

4) априорный этап;

5) постановочный этап;

6) верификация модели.

6. Верификация модели – это

1) выбор общего вида модели, выявление входящих в нее связей;

2) статистическое оценивание неизвестных параметров модели;

3) сопоставление реальных и модельных данных, проверка адекватности модели, оценка точности модельных данных;

4) регистрация значений участвующих в модели факторов и показателей;

5) среди приведенных ответов нет правильного.

7. Идентификация модели – это

1) формирование цели исследования, набора участвующих в модели экономических переменных;

2) статистическое оценивание неизвестных параметров модели;

3) выбор общего вида модели, выявление входящих в нее связей;

4) регистрация значений участвующих в модели факторов и показателей.

8. Параметризация модели – это

1) выбор общего вида модели, выявление входящих в нее связей;

2) регистрация значений участвующих в модели факторов и показателей;

3) статистическое оценивание неизвестных параметров модели;

4) сопоставление реальных и модельных данных, проверка адекватности модели, оценка точности модельных данных.

9. Временной ряд характеризует…

1) совокупность последовательных моментов времени

2) данные, описывающие один объект за ряд последовательных моментов времени

3) данные, описывающие совокупность различных объектов в определенный момент времени

4) зависимость последовательных моментов времени

10. Объясняемые, зависимые переменные в моделях любого типа называются …

1) Экзогенными;

2) Лаговыми;

3) Эндогенными;

4) Предопределенными

5) Среди приведенных ответов нет правильного

11. В уравнении парной линейном регрессии $Y = a + b \cdot X + \varepsilon$ параметрами являются

1) Переменные $Y,\ X$

2) Параметры $a,\ b$

3) Величина ε

4) Среди приведенных ответов нет правильно

12. В классической модели парной линейной регрессии $Y = a + b \cdot X + \varepsilon$

1) y, ε – детерминированные величины, x – случайная величина.

2) x – детерминированная величина, y, ε – случайные величины.

3) y – детерминированная величина, x, ε – случайные величины.

4) ε – детерминированная величина, x, y – случайные величины.

5) Среди приведенных ответов нет правильно

13. Относительно формы зависимости различают…

1) линейную и нелинейную регрессии

2) положительную и отрицательную регрессии

3) простую и множественную регрессии

4) линейную и множественную регрессии

14. Какое из уравнений регрессий является нелинейным по параметрам

1) $Y_X = ab^X$ 2) $Y_X = a + bX + cX^2$

3) $Y_X = a + \dfrac{b}{X} + \varepsilon$ 4) $Y_X = aX_1 + b\ln X_2$

15. Переменная X в модели $Y = a + b \cdot X + \varepsilon$ называется

1) регрессором

2) зависимой переменной

3) регрессантом

4) случайной ошибкой

5) прогнозом

16. Если Y – зависимая переменная, X – регрессор, ε – случайное возмущение, a,b – параметры модели, то парной линейной регрессионной моделью является:

1) $Y = a + bX + \varepsilon$

2) $Y = a \cdot X^b \cdot \varepsilon$

3) $Y = a + \frac{b}{X} + \varepsilon$

4) $Y = a \cdot b^X \cdot \varepsilon$

5) $Y = a \cdot e^{bX} \cdot \varepsilon$

17. Статистической гипотезой называется

1) любое утверждение о виде или свойствах распределения исследуемой случайной величины (величин)

2) основная (нулевая) гипотеза

3) конкурирующая (альтернативная) гипотеза

4) подтверждении или опровержении предположения на основании выборочных (экспериментальных) данных

Глава 2. КОРРЕЛЯЦИОННЫЙ АНАЛИЗ

Виды корреляционной зависимости. Парный коэффициент корреляции. Его свойства. Проверка гипотезы о значимости парного коэффициента корреляции. Корреляционная матрица. Частный коэффициент корреляции. Его свойства. Проверка гипотезы о значимости частного коэффициента корреляции. Множественный коэффициент корреляции, Его свойства. Проверка гипотезы о значимости множественного коэффициента корреляции.

Виды корреляционной зависимости	
В зависимости от направления действия	***Прямая*** корреляционная связь заключается в том, что с увеличением (уменьшением) значений объясняющей переменной увеличивается (уменьшается) значение зависимой переменной
	Обратная корреляционная связь заключается в том, что с увеличением (уменьшением) значений объясняющей переменной уменьшается (увеличивается) значение зависимой переменной
По аналитическому выражению	**Прямолинейная** зависимость: с возрастанием величины факторного признака происходит равномерное возрастание (или убывание) величин результативного признака (выражаются уравнением прямой линии)
	Нелинейная (криволинейная) зависимость: с возрастанием величины факторного признака возрастание (или убывание) результативного признака происходит неравномерно (выражаются уравнениями кривых линий)
В зависимости от количества признаков, включенных в модель	***Однофакторные (парные)*** – зависимость между одной объясняющей переменной и зависимой переменной
	Многофакторные (множественные) **–** зависимость между несколькими объясняющими переменными и зависимой переменной

Парный коэффициент корреляции

Выборочный коэффициент корреляции Пирсона (sample correlation coefficient) (или просто коэффициента корреляции)	Показывает тесноту линейной связи между количественными случайными переменными X и Y
Формула парного коэффициента корреляции	$r_*(X,Y)=\dfrac{K_*(X,Y)}{S_X\cdot S_Y}=\dfrac{K_*(X,Y)}{\sqrt{S_X^2\cdot S_Y^2}}$
$\overline{X}=\dfrac{\sum_{i=1}^{n}x_i}{n}$, $\overline{Y}=\dfrac{\sum_{i=1}^{n}y_i}{n}$, $\overline{XY}=\dfrac{\sum_{i=1}^{n}x_iy_i}{n}$, $\overline{X^2}=\dfrac{\sum_{i=1}^{n}x_i^2}{n}$, $\overline{Y^2}=\dfrac{\sum_{i=1}^{n}y_i^2}{n}$ выборочные средние	
$S_X^2=\overline{X^2}-\left(\overline{X}\right)^2$	*выборочная дисперсия (sample variances*) переменной X
$S_Y^2=\overline{Y^2}-\left(\overline{Y}\right)^2$	*выборочная дисперсия (sample variances*) переменной X
$K_*\left(X,Y\right)=$ $=\overline{XY}-\overline{X}\cdot\overline{Y}$	*выборочная ковариация* или выборочный ковариационный момент
Свойства коэффициента корреляции	1) он принимает значения от –1 до +1; 2) если $\left\|r_*\left(X,Y\right)\right\|\geq 0,7$, то линейная связь между переменными X и Y считается сильной, при $\left\|r_*\left(X,Y\right)\right\|<0,3$ линейная связь слабая; 3) если $r_*(X,Y)=\pm 1$, то корреляционное поле наблюдений представляет собой совокупность точек, которые можно расположить на одной прямой; знак «+» свидетельствует о прямой линейной зависимости между переменными X и Y, а знак «–» – об обратной линейной зависимости; 4) при $r_*\left(X,Y\right)=0$ линейная корреляционная связь отсутствует

<table>
<tr><th colspan="2">Проверка гипотезы о значимости парного
коэффициента корреляции</th></tr>
<tr><td>Назначение</td><td>Используется для проверки гипотезы о значимости коэффициента корреляции между двумя переменными</td></tr>
<tr><td>Данные</td><td>Имеются результаты n наблюдений $(x_1, y_1),\ldots,(x_n, y_n)$ переменных X и Y</td></tr>
<tr><td>Допущения</td><td>Признаки X и Y в генеральной совокупности имеют нормальное распределение и указанные данные представляют выборку из этой генеральной совокупности</td></tr>
<tr><td colspan="2">Алгоритм проверки гипотезы</td></tr>
<tr><td colspan="2">Шаг 1. Выдвинуть основную и альтернативную гипотезы</td></tr>
<tr><td>Основная (нулевая) гипотеза</td><td>$H_0 : r(X,Y) = 0$ – генеральный коэффициент корреляции равен нулю (или иначе: коэффициент корреляции генеральной совокупности незначим) (statistically unsignificant)</td></tr>
<tr><td>Конкурирующая гипотеза</td><td>$H_1 : r(X,Y) \neq 0$ о том, что генеральный коэффициент корреляции значим (statistically significant)</td></tr>
<tr><td colspan="2">Шаг 2. Найти критическое (табличное) значение статистики критерия</td></tr>
<tr><td>Критическое значение критерия</td><td>$K_{таб} = t_{1-\alpha}[St(n-2)]$ двусторонний квантиль уровня $1-\alpha$ распределения Стьюдента с $n-2$ степенями свободы) находится по табл. А3</td></tr>
<tr><td colspan="2">Шаг 3. Найти вычисленное (фактическое) значение статистики критерия</td></tr>
<tr><td>Статистика критерия (фактическое значение)</td><td>$t = \dfrac{r_*(X,Y)\sqrt{n-2}}{\sqrt{1-r_*^2(X,Y)}}$</td></tr>
<tr><td colspan="2">Шаг 4. Сравнить вычисленное значение с критическим</td></tr>
<tr><td>Область принятия нулевой гипотезы</td><td>Гипотеза о значимости коэффициента корреляции отклоняется (т.е. не отвергаем гипотезу H_0), если $\lvert K_{выч} \rvert \le K_{таб}$</td></tr>
</table>

Матрица парных коэффициентов корреляции

Парные коэффициенты корреляции могут быть вычислены для любых двух переменных X_j и Y, а также для X_j и X_i. На основе парных коэффициентов корреляции можно построить выборочную корреляционную матрицу	$\mathbf{Q}=\begin{pmatrix} 1 & r_*(Y,X_1) & \dots & r_*(Y,X_m) \\ r_*(Y,X_1) & 1 & \dots & \dots \\ \dots & \dots & \dots & \dots \\ r_*(Y,X_m) & \dots & \dots & 1 \end{pmatrix}$
Данная матрица симметрична относительно главной диагонали и положительно определена.	

Пример. Рассмотрим методы корреляционного анализа на примере данных о цене товара (тыс. руб.) – X_1, интенсивности потока покупателей (тыс. чел. в день) – X_2 и объеме продаж (тыс. шт.) – Y. Данные представлены в следующей таблице:

X_1	X_2	Y
1	2,3	7
2	2,5	7
3	3,5	6
4	2,4	5
5	2	3

Найти парный коэффициент корреляции между ценой товара и объемом продаж. На уровне значимости 0,05 проверить гипотезу о значимости генерального коэффициента корреляции. Найти все парные коэффициенты корреляции. Выписать корреляционную матрицу.

Решение. Предположим, что рассматриваемые признаки Y, X_1, X_2 в генеральной совокупности имеют нормальное распределение и указанные данные представляют выборку из этой генеральной совокупности.

Определим коэффициент корреляции между Y и X_1, для этого вычислим суммы

$$\sum_{i=1}^{n} X_{1i} = 15, \sum_{i=1}^{n} Y_i = 28, \sum_{i=1}^{n} X_{1i}^2 = 55, \sum_{i=1}^{n} Y_i^2 = 168, \sum_{i=1}^{n} X_i Y_i = 74$$

и средние

$$\overline{X_1} = \frac{1}{n}\sum_{i=1}^{n} X_{1i} = \frac{15}{5} = 3,\ \overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i = \frac{28}{5} = 5{,}6,$$

$$\overline{X_1^2} = \frac{1}{n}\sum_{i=1}^{n} X_{1i}^2 = \frac{55}{5} = 11,\ \overline{Y^2} = \frac{1}{n}\sum_{i=1}^{n} Y_i^2 = \frac{168}{5} = 33{,}6,$$

$$\overline{YX_1} = \frac{1}{n}\sum_{i=1}^{n} X_{1i} Y_i = \frac{74}{5} = 14{,}8.$$

Выборочные дисперсии будут равны

$$S_X^2 = \overline{X^2} - \left(\overline{X}\right)^2 = 11 - 3^2 = 2,\ S_{X1} = \sqrt{S_{X1}^2} = \sqrt{2} = 1{,}41,$$

$$S_Y^2 = \overline{Y^2} - \left(\overline{Y}\right)^2 = 33{,}6 - 5{,}6^2 = 2{,}24,\ S_Y = \sqrt{S_Y^2} = \sqrt{2{,}24} = 1{,}49.$$

Выборочная ковариация будет равна

$$K_*\left(Y, X_1\right) = \overline{YX_1} - \overline{X_1} \cdot \overline{Y} = 14{,}8 - 3 \cdot 5{,}6 = -2.$$

Выборочный коэффициент корреляции между ценой товара и объемом продаж равен

$$r_*\left(Y, X_1\right) = \frac{K_*\left(Y, X_1\right)}{S_{X1} \cdot S_Y} = \frac{-2}{1{,}41 \cdot 1{,}49} = -0{,}95,$$

что говорит о сильной линейной обратной взаимосвязи между переменными.

Проверим нулевую гипотезу $H_0 : r(X, Y) = 0$ о том, что генеральный коэффициент корреляции незначим, против конкурирующей гипотезы $H_1 : r(X, Y) \neq 0$ о том, что генеральный коэффициент корреляции значим.

Фактическое значение ($K_{\text{выч}}$) t-статистики критерия

$$K_{\text{выч}} = \frac{r_*\left(Y, X_1\right)\cdot\sqrt{n-2}}{\sqrt{1-r_*^2\left(Y, X_1\right)}} = \frac{-0,95\cdot\sqrt{5-2}}{\sqrt{1-0,95^2}} = 5,27.$$

Найдем критическое значение статистики по табл. А3:

$$K_{таб} = t_{1-\alpha}[St(n-2)] = t_{0,95}\left[St\left(3\right)\right] = x_{0,975}\left[St\left(3\right)\right] = 3,18.$$

Гипотеза о незначимости коэффициента корреляции отклоняется, поскольку $K_{\text{выч}} > K_{\text{таб}}$, т.е. на уровне значимости 0,05 можно утверждать, что коэффициент корреляции в генеральной совокупности значим.

Далее, парные коэффициенты корреляции между остальными переменными $\mathrm{r}_*\left(Y, X_1\right) = -0,95, r_*\left(Y, X_2\right) = 0,39,\ r_*\left(X_1, X_2\right) = -0,19$, то корреляционная матрица будет иметь вид

$$\mathbf{Q} = \begin{pmatrix} 1 & r_*\left(X_1, X_2\right) & r_*\left(Y, X_1\right) \\ r_*\left(X_1, X_2\right) & 1 & r_*\left(Y, X_2\right) \\ r_*\left(Y, X_1\right) & r_*\left(Y, X_2\right) & 1 \end{pmatrix} =$$

$$= \begin{pmatrix} 1 & -0,19 & -0,95 \\ -0,19 & 1 & 0,39 \\ -0,95 & 0,39 & 1 \end{pmatrix}.$$

Частный коэффициент корреляции

Частный коэффициент корреляции	показывает силу между переменными Y и X_j при исключении влияния на их связь всех остальных независимых переменных
Формула частного коэффициента корреляции	$r_*(Y, X_j \mid X_1, \ldots, X_k) = \dfrac{-Q_{Y, X_j}}{\sqrt{Q_{X_j, X_j}\cdot Q_{Y,Y}}}$

<table>
<tr><td colspan="2">Частный коэффициент корреляции используется для измерения силы «очищенной» линейной связи между двумя заданными переменными. Такое применение частного коэффициента корреляции теоретически обосновано лишь в том случае, когда переменные Y, X_1, …, X_k имеют совместное нормальное распределение. Во всех остальных случаях подобное использование носит эвристический характер.</td></tr>
<tr><td>Q_{Y,X_j}</td><td>алгебраическое дополнение коэффициента корреляции $r_*\left(Y,X_j\right)$ в выборочной корреляционной матрице Q переменных $Y,X_1,\ldots.X_k$</td></tr>
<tr><td>Q_{X_j,X_j}</td><td>алгебраическое дополнение коэффициента корреляции $r_*\left(X_j,X_j\right)=1$ в матрице Q</td></tr>
<tr><td>$Q_{Y,Y}$</td><td>алгебраическое дополнение коэффициента корреляции $r_*\left(Y,Y\right)=1$ в матрице Q</td></tr>
<tr><td>для трех переменных Y,X_1,X_2</td><td>$r_*\left(Y,X_1;X_2\right)=\dfrac{r_*\left(Y,X_1\right)-r_*\left(Y,X_2\right)\cdot r_*\left(X_1,X_2\right)}{\sqrt{\left(1-r_*^2\left(Y,X_2\right)\right)\left(1-r_*^2\left(X_1,X_2\right)\right)}}$.</td></tr>
<tr><td colspan="2">Порядок частного коэффициента корреляции определяется количеством факторов, влияние которых исключается.</td></tr>
<tr><td colspan="2">Например, $r_*\left(Y,X_1;X_2\right)$ – коэффициент частной корреляции первого порядка, а $r_*\left(Y,X_1;X_2,\ldots,X_k\right)$ – коэффициент частной корреляции порядка k</td></tr>
<tr><td>Свойства коэффициента корреляции</td><td>1) он принимает значения от –1 до +1;
2) если $\left|r_*(Y,X_j\mid X_1,\ldots,X_k)\right|\geq 0,7$, то линейная связь между переменными X и Y считается сильной, при $\left|r_*(Y,X_j\mid X_1,\ldots,X_k)\right|<0,3$ линейная связь слабая;</td></tr>
</table>

	3) если $r_*(Y, X_j \mid X_1, \ldots, X_k) = \pm 1$, то корреляционное поле наблюдений представляет собой совокупность точек, которые можно расположить на одной прямой; знак «+» свидетельствует о прямой линейной зависимости между переменными X_j и Y, а знак «–» – об обратной линейной зависимости; 4) при $r_*(Y, X_j \mid X_1, \ldots, X_k) = 0$ линейная корреляционная связь отсутствует

Проверка гипотезы о значимости частного коэффициента корреляции	
Назначение	Используется для проверки гипотезы о значимости частного коэффициента корреляции между двумя переменными
Данные	Имеются результаты n наблюдений $(x_1, y_1), \ldots, (x_n, y_n)$ переменных X и Y
Допущения	переменные $Y, X_1, \ldots. X_k$ имеют совместное нормальное распределение
Алгоритм проверки гипотезы	
Шаг 1. Выдвинуть основную и альтернативную гипотезы	
Основная (нулевая) гипотеза	$H_0 : r(Y, X_j \mid X_1, \ldots, X_k) = 0$ – генеральный частный коэффициент корреляции равен нулю (или иначе: коэффициент корреляции генеральной совокупности незначим) *(statistically unsignificant)*
Конкурирующая гипотеза	$H_1 : r(Y, X_j \mid X_1, \ldots, X_k) \neq 0$ о том, что генеральный частный коэффициент корреляции значим *(statistically significant)*
Шаг 2. Найти критическое (табличное) значение статистики критерия	
Критическое значение критерия	$K_{таб} = t_{1-\alpha}[St(n-k-2)]$ двусторонний квантиль уровня $1-\alpha$ распределения Стьюдента с $n-k-2$ степенями свободы) находится по табл. А3

Шаг 3. Найти вычисленное (фактическое) значение статистики критерия	
Статистика критерия (фактическое значение)	$t = \dfrac{r_*(Y, X_j \mid X_1, \ldots, X_k)\sqrt{n-k-2}}{\sqrt{1 - r_*^2(Y, X_j \mid X_1, \ldots, X_k)}}$
Шаг 4. Сравнить вычисленное значение с критическим	
Область принятия нулевой гипотезы	Гипотеза о значимости частного коэффициента корреляции отклоняется (т.е. не отвергаем гипотезу Y), если $b = 0$

Пример. *По данным предыдущего примера найти частный коэффициент корреляции* $r_*\left(Y, X_1 | X_2\right)$.

Решение. Поскольку корреляционная матрица имеет вид

$$Y = \begin{pmatrix} 1 & r_*\left(X_1, X_2\right) & r_*\left(Y, X_1\right) \\ r_*\left(X_1, X_2\right) & 1 & r_*\left(Y, X_2\right) \\ r_*\left(Y, X_1\right) & r_*\left(Y, X_2\right) & 1 \end{pmatrix} =$$

$$= \begin{pmatrix} 1 & -0,19 & -0,95 \\ -0,19 & 1 & 0,39 \\ -0,95 & 0,39 & 1 \end{pmatrix}$$

а алгебраические дополнения

$$Q_{Y,X_1} = \left(-1\right)^{1+3} \cdot \begin{vmatrix} -0,19 & 1 \\ -0,95 & 0,39 \end{vmatrix} = -0,19 \cdot 0,39 + 0,95 \cdot 1 = 0,8759,$$

$$Q_{X_1,X_1} = \left(-1\right)^{1+1} \cdot \begin{vmatrix} 1 & 0,39 \\ 0,39 & 1 \end{vmatrix} = 1 - 0,39^2 = 0,8479,$$

$$Q_{Y,Y} = \left(-1\right)^{3+3} \cdot \begin{vmatrix} 1 & -0,19 \\ -0,19 & 1 \end{vmatrix} = 1 - 0,19^2 = 0,9639,$$

частный коэффициент корреляции будет равен

$$r_*\left(Y, X_1 | X_2\right) = \frac{-Q_{Y,X_1}}{\sqrt{Q_{X_1,X_1} \cdot Q_{Y,Y}}} = \frac{-0,8759}{\sqrt{0,8479 \cdot 0,9639}} = -0,9689,$$

что говорит об очень тесной обратной линейной связи между ценой товара и объемами продаж при фиксированной интенсивности потока покупателей.

Или по упрощенной формуле

$$r_*\left(Y, X_1 | X_2\right) = \frac{r_*\left(Y, X_1\right) - r_*\left(Y, X_2\right) \cdot r_*\left(X_1, X_2\right)}{\sqrt{\left(1 - r_*^2\left(Y, X_1\right)\right) \cdot \left(1 - r_*^2\left(X_1, X_2\right)\right)}} =$$

$$= \frac{-0,95 - \left(-0,19\right) \cdot 0,39}{\sqrt{\left(1 - 0,19^2\right) \cdot \left(1 - 0,39^2\right)}} = -0,9689.$$

Порядок частного коэффициента корреляции будет равен единице (k = 1), т.к. количество зафиксированных переменных равно 1 (это переменная X_2).

На уровне значимости 0,05 проверить гипотезу о значимости частного коэффициента корреляции $r_*\left(Y, X_1 | X_2\right)$.

Решение. Проверим гипотезу $H_0 : r\left(Y, X_1 | X_2\right) = 0$ (частный коэффициент корреляции в генеральной совокупности незначим) против конкурирующей $H_0 : r\left(Y, X_1 | X_2\right) \neq 0$ (частный коэффициент корреляции в генеральной совокупности значим).

Вычисленное значение статистики критерия равно

$$K_{\text{выч}} = \frac{r_*\left(Y, X_1 | X_2\right)\sqrt{n - k - 2}}{\sqrt{1 - r_*^2\left(Y, X_1 | X_2\right)}} = \frac{-0,9689\sqrt{5 - 1 - 2}}{\sqrt{1 - 0,9689^2}} = -5,54.$$

Критическое значение статистики находится по табл. А3:

$$K_{\text{таб}} = t_{1-0,05}\left[St\left(5 - 1 - 2\right)\right] = t_{0,95}\left[St\left(2\right)\right] = 4,3.$$

Поскольку $\left|K_{\text{выч}}\right| > K_{\text{выч}}$, то гипотезу о значимости частного коэффициента корреляции принимаем на уровне значимости 0,05.

Множественный коэффициент корреляции

Множественный коэффициент корреляции	используется в качестве измерителя статистической связи между результирующим показателем Y и набором объясняющих переменных $X_1,\ldots X_k$ (линейная форма зависимости Y от набора $X_1,\ldots X_k$). Он показывает долю дисперсии рассматриваемой величины Y, обусловленную влиянием остальных переменных $X_1,\ldots X_k$, включенных в корреляционную модель
Формула множественного коэффициента корреляции	$R_{Y.X} = \sqrt{1 - \frac{\lvert \mathbf{Q} \rvert}{Q_{YY}}}$
$\lvert \mathbf{Q} \rvert$	определитель матрицы корреляций $\mathbf{Q}$
$Q_{Y,Y}$	алгебраическое дополнение коэффициента корреляции $r_*(Y,Y)=1$ в матрице $\mathbf{Q}$
Свойства множественного коэффициента корреляции	он принимает значения от 0 до +1; если $R_{Y,X} \geq 0,7$, то линейная связь между переменными $X_1,\ldots,X_k$ и Y считается сильной, при $R_{Y,X} < 0,3$ линейная связь слабая

Проверка гипотезы о значимости множественного коэффициента корреляции

Назначение	Используется для проверки гипотезы о значимости множественного коэффициента корреляции между показателем Y и набором объясняющих переменных $X_1,\ldots X_k$
Данные	Имеются результаты n наблюдений Y и $X_1,\ldots X_k$
Допущения	переменные $Y, X_1,\ldots X_k$ имеют совместное нормальное распределение

Алгоритм проверки гипотезы	
Шаг 1. Выдвинуть основную и альтернативную гипотезы	
Основная (нулевая) гипотеза	$H_0:\ R_{y.X}=0$ о том, что коэффициент множественной корреляции генеральной совокупности равен нулю (т.е. статистические незначим)
Конкурирующая гипотеза	$H_1:\ R_{y.X}\neq 0$ о том, что коэффициент множественной корреляции генеральной совокупности не равен нулю (значим)
Шаг 2. Найти критическое (табличное) значение статистики критерия	
Критическое значение критерия	$K_{таб}=x_{1-\alpha}[F(k,\ n-k-1)]$ односторонний квантиль уровня $1-\alpha$ распределение Фишера с $v_1=k$ и $v_2=n-k-1$ степенями свободы, находится по табл. А5 и 6
Шаг 3. Найти вычисленное (фактическое) значение статистики критерия	
Статистика критерия (фактическое значение)	$F=\dfrac{R^2_{y,X}}{1-R^2_{y,X}}\cdot\dfrac{n-k-1}{k}$,
Шаг 4. Сравнить вычисленное значение с критическим	
Область принятия нулевой гипотезы	Гипотеза о значимости множественного коэффициента корреляции отклоняется (т.е. не отвергаем гипотезу H_0), если $K_{выч}<K_{таб}$

Пример. *Определить силу влияния цены товара и интенсивности потока покупателей на объем продаж с помощью множественного коэффициента корреляции. На уровне значимости 0,05 проверить значимость множественного коэффициента корреляции.*

Решение. Поскольку корреляционная матрица имеет вид

$$\mathbf{Q} = \begin{pmatrix} 1 & r_*(X_1, X_2) & r_*(Y, X_1) \\ r_*(X_1, X_2) & 1 & r_*(Y, X_2) \\ r_*(Y, X_1) & r_*(Y, X_2) & 1 \end{pmatrix} =$$

$$= \begin{pmatrix} 1 & -0,19 & -0,95 \\ -0,19 & 1 & 0,39 \\ -0,95 & 0,39 & 1 \end{pmatrix},$$

определитель будет $|\boldsymbol{Q}| = 1 + (-0,95)\cdot(-0,19)\cdot 0,39 + (-0,95)\times$ $\times(-0,19)\cdot 0,39 - (-0,95)^2 - 0,39^2 - (-0,19)^2 = 0,05,$

а алгебраическое дополнение

$$Q_{Y,Y} = (-1)^{3+3}\cdot\begin{vmatrix} 1 & -0,19 \\ -0,19 & 1 \end{vmatrix} = 1 - 0,19^2 = 0,9639.$$

Тогда множественный коэффициент корреляции будет равен

$$R_{Y.X} = \sqrt{1 - \frac{|\boldsymbol{Q}|}{Q_{Y,Y}}} = \sqrt{1 - \frac{0,05}{0,9639}} = 0,974.$$

На уровне значимости 0,05 проверим гипотезу о незначимости множественного коэффициента корреляции $H_0:\ R_{y.X} = 0$ против конкурирующей $H_1:\ R_{y.X} \neq 0$.

Вычисленное значение статистики для проверки гипотезы будет равно

$$K_{\text{выч}} = \frac{R^2_{Y.X}}{1 - R^2_{Y.X}}\cdot\frac{n-k-1}{k} = \frac{0,974^2}{1-0,974^2}\cdot\frac{5-2-1}{2} = 18,48,$$

где k = 2 – число факторных переменных (в нашем случае это X_1, X_2).

Критическое значение статистики находится по табл. А6 и равно

$$K_{\text{табл}} = x_{1-0,05}\left[F(2,2)\right] = 19,0.$$

Поскольку $K_{выч} < K_{таб}$, то гипотезу о незначимости модели следует принять.

Тесты для проверки знаний

1. Коэффициент парной корреляции характеризует…
1) тесноту нелинейной связи между несколькими переменными
2) тесноту линейной связи между двумя переменными
3) тесноту линейной связи между несколькими переменными
4) тесноту нелинейной связи между двумя переменными
5) среди приведенных ответов нет правильного

2. Корреляция подразумевает наличие связи между:
1) случайными факторами
2) параметрами
3) переменными
4) результатом и случайными факторами
5) среди приведенных ответов нет правильного

3. Коэффициент парной корреляции изменяется в пределах
1) [0, 1)
2) [–1, 1]
3) (0, 1)
4) [0, 1]
5) среди приведенных ответов нет правильного

4. Какое из этих значений может принимать коэффициент парной корреляции при прямой связи?
1) – 0,6
2) 0,6
3) 1,2
4) – 1,2
5) среди приведенных ответов нет правильного

5. Какое из этих значений может принимать коэффициент парной корреляции при обратной связи?

1) – 0,8

2) 0,7

3) 1,2

4) – 1,2

5) среди приведенных ответов нет правильного

6. Значение коэффициента корреляции равно 0,85. Можно сделать вывод о том, что связь между результативным признаком и факторами является……

1) достаточно тесной и прямой
2) слабой и обратной
3) не тесной, но прямой
4) функциональной
5) среди приведенных ответов нет правильного

7. Коэффициент парной корреляции между величинами X и Y равен нулю, тогда какое утверждение верно

1) величины независимы
2) между величинами нет никакой функциональной зависимости
3) между величинами нет линейной зависимости
4) все три предыдущих ответа неверны
5) среди приведенных ответов нет правильного

8. Рассчитанный по выборке коэффициент парной корреляции оказался равным –1. Это означает, что

1) между изучаемыми переменными есть слабая отрицательная линейная связь
2) между изучаемыми переменными есть связь, но она не является линейной
3) между изучаемыми переменными есть функциональная линейная отрицательная связь;

4) между изучаемыми переменными отсутствует связь
5) полученное число никак не интерпретируется

9. Чем ближе значение модуля линейного коэффициента корреляции к единице, тем … между изучаемыми признаками.

1) теснее линейная связь
2) ярко выражена нелинейная связь
3) слабее линейная связь
4) отсутствует какая-либо зависимость
5) среди приведенных ответов нет правильного

10. Известны следующие результаты наблюдений: $n = 20$, $\overline{X} = 15$, $\overline{Y} = 12$, $\sum_{i=1}^{n} X_i Y_i = 3900$, $\sum_{i=1}^{n} Y_i^2 = 3600$, $\sum_{i=1}^{n} X_i^2 = 5000$.

Выборочный коэффициент корреляции равен:

1) 0,75 2) 0,5 3) –0,5 4) –0,75 5) 0,25

11. По данным 17 наблюдений найден выборочный коэффициент корреляции $r^*(X,Y) = 0{,}8.$ Найти вычисленную статистику критерия для проверки гипотезы о значимости коэффициента корреляции в генеральной совокупности.

1) 5,164
2) 6,928
3) 3,20
4) 5,333
5) 3,098

12. Линейный коэффициент корреляции между признаками X и Y, заданными таблицей

X	1	2	3
Y	3	2	1

1) равен 0
2) меньше нуля, но больше –1

3) равен 1

4) равен –1

5) больше нуля, но меньше 1

13. Матрица коэффициентов корреляции вектора (X_1, X_2, Y), где Y – зависимая переменная, равна $\begin{pmatrix} 1 & * & -0,6 \\ 0,6 & 1 & -0,5 \\ * & * & 1 \end{pmatrix}$. Множественный коэффициент корреляции $R_{y.x1,x2}$

1) 0,45

2) 0,55

3) 0,625

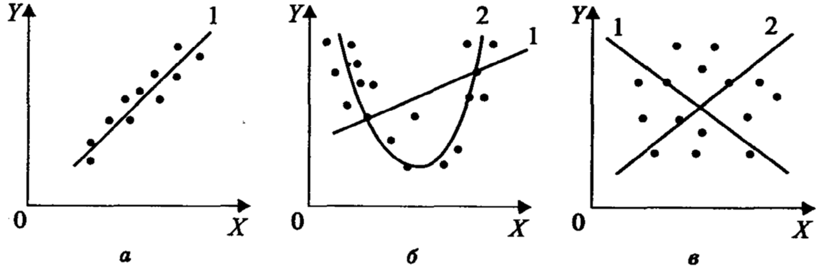
4) 0,6

5) Среди приведенных нет ответа

Глава 3. ПАРНЫЙ РЕГРЕССИОННЫЙ АНАЛИЗ

Простейшая линейная регрессионная модель (ПЛРМ). Природа случайной ошибки. Корреляционное поле наблюдений и его применение к выбору формы регрессии. Оценки методом наименьших квадратов коэффициентов ПЛРМ. Интерпретация коэффициентов ПЛРМ. Коэффициент детерминации и его свойства. Теорема Гаусса–Маркова. Доверительные интервалы для коэффициентов регрессии и проверка гипотез об их значимости (t – тест). Проверка значимости всей регрессии на основе критерия Фишера. Прогнозирование значения зависимой переменной по ПЛРМ, точность прогноза. Функциональные преобразования в линейной регрессионной модели. Линеаризация нелинейной регрессионной модели.

Парная линейная регрессионная модель	
Вид парной линейной регрессионной модели	$Y = a + bX + \varepsilon$
Функция регрессии	$Y_X = a + bX$
Случайная ошибка модели (***error***, или ***disturbance***)	ε Наличие случайного члена ε связано с воздействием на зависимую переменную Y других факторов, не учтенных в функции регрессии, а также с возможной нелинейностью модели и ошибками измерения
Неизвестные параметры модели	a, b
Спецификация модели	
Диаграммы рассеяния или ***корреляционное поле наблюдений***	это график, на котором результаты n наблюдений $(x_1, y_1), \ldots, (x_n, y_n)$ переменных X и Y изображаются в виде точек в декартовой системе координат

<table>
<tr><td colspan="2">


а) Взаимосвязь между X и Y близка к линейной: $Y_X = a + bX$
б) Взаимосвязь близка к квадратичной: $Y_X = a + bX + cX^2$
в) Взаимосвязь между X и Y отсутствует. Какую бы мы ни выбрали форму связи, результаты проверки ее качества будут неудачными
</td></tr>
<tr><td colspan="2">Метод наименьших квадратов
для оценки параметров модели (МНК)</td></tr>
<tr><td colspan="2">Служит для нахождения оценок неизвестных параметров a, b модели $Y = a + bX + \varepsilon$ по результатам наблюдений $(x_i,\ y_i),\ i = 1,...,n$.</td></tr>
<tr><td colspan="2">Результаты наблюдений связаны моделью наблюдений: $y_i = a + bx_i + \varepsilon_i$, где ε_i – случайная ошибка i -го наблюдения</td></tr>
<tr><td>Суть МНК</td><td>определить такие значения a, b , при которых достигает своего минимального значения сумма квадратов ошибок модели в n наблюдениях:

$$Q = Q(a,b) = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - a - bx_i)^2$$</td></tr>
<tr><td>Система нормальных уравнений служит для определения оценок неизвестных параметров линейной регрессии</td><td>

$$\begin{cases} an + b\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i, \\ a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} y_i x_i. \end{cases}$$</td></tr>
<tr><td>Оценки параметров модели, найденные МНК</td><td>

$$\begin{cases} \hat{b} = \dfrac{K_*(X,Y)}{S_X^2} = r_*(X,Y) \cdot \dfrac{S_Y}{S_X} \\ \hat{a} = \overline{Y} - \hat{b}\overline{X} \end{cases}$$</td></tr>
</table>

<table>
<tr><td colspan="2">Статистики $\hat{a}, \hat{b}$ называются *оценками наименьших квадратов (least squares estimates)* неизвестных параметров a и b, или *LS-оценками*.</td></tr>
<tr><td>***Оценка функции регрессии*** (эмпирическая регрессия, выборочная регрессия)</td><td>$\hat{Y}_X = \hat{a} + \hat{b}X$</td></tr>
<tr><td colspan="2">Интерпретация найденных параметров</td></tr>
<tr><td>Оценка $\hat{a}$</td><td>дает оценку среднего значения зависимой переменной при $X = 0$. Эта интерпретация возможна или невозможна в зависимости от того, насколько далеко находится $X = 0$ от выборочных значений X.</td></tr>
<tr><td>Оценка $\hat{b}$</td><td>угловой коэффициент регрессии. Он показывает, на сколько единиц в среднем изменится переменная Y при увеличении независимой переменной X на единицу</td></tr>
<tr><td>Регрессионные остатки (оценки погрешности)</td><td>$e_i = y_i - \hat{y}_i$, где y_i – реальные значения зависимой переменной, $\hat{y}_i$ – расчетные значения</td></tr>
</table>

Предпосылки парного регрессионного анализа и теорема Гаусса-Маркова

<table>
<tr><td colspan="2">Рассмотрим модель наблюдений: $y_i = a + bx_i + \varepsilon_i$, где ε_i – случайная ошибка i-го наблюдения</td></tr>
<tr><td colspan="2">Условия, определяющие классическую парную линейную регрессионную модель</td></tr>
<tr><td>Значения $x_1, x_2, ..., x_n$ являются неслучайными величинами</td><td>В регрессионном анализе часто вместо условия о неслучайности значений объясняющей переменной используется более слабое условие о независимости (некоррелированности) объясняющей переменной и случайной ошибки.</td></tr>
<tr><td>Математическое ожидание случайной ошибки в каждом наблюдении равно нулю $M(\varepsilon_i) = 0,\ i = \overline{1, n}$</td><td>Регрессионные ошибки в среднем не оказывают влияния на зависимую переменную</td></tr>
</table>

Дисперсия случайной ошибки постоянна для всех наблюдений $D(\varepsilon_i) = M(\varepsilon_i^2) = \sigma^2$, $i = \overline{1, n}$.	Условие *постоянства дисперсии* случайной ошибки называют также гомоскедастичностью. *Зависимость* дисперсии случайной ошибки от номера наблюдения называется гетероскедастичностью
Случайные ошибки различных наблюдений статистически не связаны (некоррелированы) между собой, т е. $M(\varepsilon_i \varepsilon_j) = 0$ *при* $i \neq j$	Это условие указывает на некоррелированность случайных ошибок для разных наблюдений. Условие часто нарушается, когда исходные данные являются временными рядами. В этом случае говорят об автокорреляции случайных ошибок
$\varepsilon_i \sim N(0; \sigma^2)$	Случайные ошибки имеют совместное нормальное распределение
Теорема Гаусса – Маркова	
Если для модели $y_i = a + bx_i + \varepsilon_i$ выполняются условия 1–4, то оценки $\hat{a}, \hat{b}$, найденные с помощью МНК, являются наилучшими линейными несмещенными оценками (*Best Linear Unbiased Estimator*, или *BLUE*)) параметров a и b. При этом для любых действительных значений c_1, c_2 оценка $\hat{a}c_1 + \hat{b}c_2$ является наилучшей линейной несмещенной оценкой параметрической функции $ac_1 + bc_2$	

Основные статистики парного регрессионного анализа

$TSS = n \cdot S_Y^2 = \sum_{i=1}^{n} (y_i - \overline{y})^2$	*Полная сумма квадратов (total sum of squares)* характеризует общую изменчивость зависимой переменной
$ESS = \sum_{i=1}^{n} (\hat{y}_i - \overline{y})^2$	*Сумма квадратов, объясненной моделью* (*explained sum of squares), или факторной суммой квадратов* определяет ту часть TSS, которую позволяет объяснить предложенная модель

$RSS = \sum_{i=1}^{n} e_i^2$	*Остаточная сумма квадратов* (*residual sum of squares*) определяет ту часть вариации, которую не позволяет объяснить предложенная модель
$TSS = ESS + RSS$	Основное тождество дисперсионного анализа
$\hat{\sigma}^2 = \frac{RSS}{n-m}$	*Оценка дисперсии (ошибки) модели* – несмещенная оценка дисперсии ошибки модели
$\hat{\sigma} = \sqrt{\hat{\sigma}^2}$	*Стандартная ошибка модели*
$S_{\hat{a}} = \sqrt{\hat{\sigma}^2 \cdot \frac{\overline{X^2}}{n \cdot S_X^2}}$	*Стандартная ошибка оценки параметра* a *(Std. Error)* – это среднеквадратичное отклонение оценки параметра модели
$S_{\hat{b}} = \sqrt{\hat{\sigma}^2 \cdot \frac{1}{n \cdot S_X^2}}$	*Стандартная ошибкаоценки параметра* b *(Std. Error)* – это среднеквадратичное отклонение оценки параметра модели
$R^2 = 1 - \frac{RSS}{TSS} = \frac{ESS}{TSS}$	*Коэффициент детерминации* определяет долю дисперсии зависимой переменной, объясненную эмпирической функцией регрессии. По этой причине коэффициент детерминации является основной мерой качества подгонки регрессионной модели к данным с помощью выборочного уравнения регрессии
Свойства коэффициента детерминации	
1.	$R^2 \in [0; 1]$
2.	Чем ближе R^2 к единице, тем выше качество подгонки зависимой переменной с помощью выбранной регрессионной модели и тем более точно функция $\hat{Y}_X$ аппроксимирует поведение Y с учетом изменения независимой переменной X
3.	Для простейшей парной линейной регрессионной модели (какую мы и рассматриваем) коэффициент детерминации равен квадрату коэффициента корреляции: $R^2 = r_*^2(X, Y)$
4.	Вычисление R^2 корректно, если константа a включена в уравнение регрессии

Проверка гипотезы о незначимости параметра модели (критерий Стьюдента)

<table>
<tr><th colspan="2">Проверка гипотезы при наличии двусторонней альтернативы</th></tr>
<tr><td>Назначение</td><td>Используется для проверки гипотезы о незначимости параметра модели в генеральной совокупности</td></tr>
<tr><td>Данные</td><td>Имеются результаты n наблюдений $(x_1, y_1),\ldots,(x_n, y_n)$ переменных X и Y</td></tr>
<tr><td>Допущения</td><td>Модель наблюдений: $y_i = a + bx_i + \varepsilon_i$, где ε_i – случайные величины, имеющие нормальное распределение</td></tr>
<tr><td colspan="2">Алгоритм проверки гипотезы</td></tr>
<tr><td colspan="2">Шаг 1. Выдвинуть гипотезы основную и альтернативную на уровне значимости α</td></tr>
<tr><td>Основная (нулевая) гипотеза</td><td>$H_0 : b = 0$ (коэффициент незначим), генеральный коэффициент регрессии равен нулю (или иначе: коэффициент регрессии генеральной совокупности незначим) (statistically unsignificant).
При $b = 0$ можно говорить об отсутствии влияния независимой переменной X на результат – значение зависимой переменной Y. Это означает, что функция регрессии не содержит независимую переменную. Поэтому в случае парной линейной регрессионной модели проверка гипотезы $b = 0$ равносильна проверке гипотезы о незначимости парной линейной регрессионной модели</td></tr>
<tr><td>Конкурирующая (альтернативная) гипотеза</td><td>$H_1 : b \neq 0$ (коэффициент значим) о том, что генеральный коэффициент регрессии значим (statistically significant)</td></tr>
</table>

Шаг 2. Найти критическое (табличное) значение статистики критерия Стьюдента	
Критическое значение критерия	$K_{таб} = t_{1-\alpha}[St(n-m)]$ двусторонний квантиль уровня $1-\alpha$ распределения Стьюдента с $n-m$ степенями свободы) находится по табл. А3
Шаг 3. Найти вычисленное (фактическое) значение статистики критерия Стьюдента	
Статистика критерия (фактическое значение)	$t - statistics = \dfrac{\hat{b}}{S_{\hat{b}}}$
Шаг 4. Сравнить вычисленное значение с критическим	
Область принятия нулевой гипотезы	Гипотеза о значимости коэффициента регрессии отклоняется (т.е. не отвергаем гипотезу Y), если $b = 0$
Проверка гипотез о незначимости параметра a осуществляется таким же образом, как и о незначимости параметра b	

Проверка гипотезы о значимости модели (критерий Фишера)

Назначение	Используется для проверки гипотезы о значимости *линейной регрессионной модели* в генеральной совокупности
Данные	Имеются результаты n наблюдений $(x_1, y_1), \ldots, (x_n, y_n)$ переменных X и Y
Допущения	Модель наблюдений: $y_i = a + bx_i + \varepsilon_i$, где ε_i – случайные величины, имеющие нормальное распределение
Алгоритм проверки гипотезы	
Шаг 1. Выдвинуть гипотезы основную и альтернативную на уровне значимости α	
Основная (нулевая) гипотеза	$H_0 : R^2_Г = 0$ (*коэффициент детерминации генеральной совокупности равен нулю, модель регрессии незначима*)

Конкурирующая гипотеза	$H_0 : R^2_\Gamma \neq 0$ (*модель регрессии значима*)
Шаг 2. Найти критическое (табличное) значение статистики критерия	
Критическое значение критерия	$K_{таб} = x_{1-\alpha}\left[F(m-1,\, n-m)\right]$ квантиль уровня $1-\alpha$ распределения Фишера с $\nu_1 = m-1$ и $\nu_2 = n-m$ степенями свободы находится по табл. А5 и А6
Шаг 3. Найти вычисленное (фактическое) значение статистики критерия Фишера	
Статистика критерия (фактическое значение)	$F - statistics = \frac{R^2}{1-R^2} \cdot \frac{n-m}{m-1}$, где m – число параметров модели
Шаг 4. Сравнить вычисленное значение с критическим	
Область принятия нулевой гипотезы	Если $K_{выч} < K_{таб}$, то гипотезу H_0 о незначимости модели следует принять, т.е. нет оснований отклонить эту гипотезу

Доверительные интервалы для параметров линейного уравнения парной регрессии

Двусторонние доверительные интервалы	
$\hat{a} - t_\gamma[St(n-m)] \cdot S_{\hat{a}} < a < \hat{a} + t_\gamma[St(n-m)] \cdot S_{\hat{a}}$	для параметра a
$\hat{b} - t_\gamma[St(n-m)] \cdot S_{\hat{b}} < b < \hat{b} + t_\gamma[St(n-m)] \cdot S_{\hat{b}}$	для параметра b
Правосторонние доверительные интервалы	
$a < \hat{a} + x_\gamma[St(n-m)] \cdot S_{\hat{a}}$	для параметра a
$b < \hat{b} + x_\gamma[St(n-m)] \cdot S_{\hat{b}}$	для параметра b
Левосторонние доверительные интервалы	
$\hat{a} - x_\gamma[St(n-m)] \cdot S_{\hat{a}} < a,$	для параметра a
$\hat{b} - x_\gamma[St(n-m)] \cdot S_{\hat{b}} < b$	для параметра b

$t_\gamma[St(n-m)]$	двусторонний квантиль уровня γ распределения Стьюдента с $n-m$ степенями свободы находится по табл. А3
$x_\gamma[St(n-m)]$	односторонний квантиль уровня γ распределения Стьюдента с $n-m$ степенями свободы находится по табл. А3

Прогнозирование и интервальное оценивание в парной линейной регрессионной модели

<table>
<tr><td colspan="2">Под прогнозированием в эконометрике понимается построение оценки будущего значения зависимой переменной Y при заданном значении независимой переменной $X = x_*$.</td></tr>
<tr><td colspan="2">Прогнозируемое значение обозначим</td></tr>
<tr><td>y_*</td><td>индивидуальное значение зависимой переменной</td></tr>
<tr><td>$M[y_*]$</td><td>среднее значение зависимой переменной</td></tr>
<tr><td colspan="2">Различают</td></tr>
<tr><td>точечное прогнозирование</td><td>оценка истинного значения прогноза (некоторое число)</td></tr>
<tr><td>$\hat{y}_* = \hat{a} + \hat{b}x_*$</td><td>точечный прогноз индивидуального и среднего значений зависимой переменной</td></tr>
<tr><td>интервальное прогнозирование</td><td>интервал, в котором находится прогнозируемая величина с заданным уровнем надежности γ</td></tr>
<tr><td colspan="2">Интервальный прогноз среднего значений зависимой переменной</td></tr>
<tr><td colspan="2">$$\hat{y}_* - t_\gamma[St(n-m)] \cdot S_{Y_X} < M[y_*] < \hat{y}_* + t_\gamma[St(n-m)] \cdot S_{Y_X}$$</td></tr>
<tr><td>$S_{Y_X} = \sqrt{\hat{\sigma}^2 \cdot \left(\frac{1}{n} + \frac{(x_* - \bar{X})^2}{n \cdot S_X^2} \right)}$</td><td>стандартная ошибка оценки среднего значения зависимой переменной $M[y_*]$</td></tr>
<tr><td colspan="2">Интервальный прогноз индивидуального значений зависимой переменной</td></tr>
<tr><td colspan="2">$$\hat{y}_* - t_\gamma[St(n-m)] \cdot S_{y_*} < y_* < \hat{y}_* + t_\gamma[St(n-m)] \cdot S_{y_*}$$</td></tr>
<tr><td>$S_{y_*} = \sqrt{\hat{\sigma}^2 \cdot \left(1 + \frac{1}{n} + \frac{(x_* - \bar{X})^2}{n \cdot S_X^2} \right)}$</td><td>стандартная ошибка оценки индивидуального значения зависимой переменной y_*</td></tr>
</table>

Пример. Коммерческий директор исследует зависимость спроса на известную торговую марку от количества проведенных мероприятий по ее продвижению (например, рекламные ролики, участие в выставочных конгрессах и др.). Для этого собрана следующая статистика (табл. 1)

Таблица 1

Районы сбыта	Количество проводимых маркетинговых мероприятий по сбыту (ед.)	Товарооборот марки в разных районах сбыта (млн руб.)
1	109	27,9
2	114	28,1
3	109	28,3
4	103	27,2
5	118	28,8
6	116	29,6
7	124	28,1
8	133	28,3
9	132	27,2
10	138	28,8
11	143	31,1
12	148	32,7
13	134	29,5
14	145	30,2
15	150	31,5
16	159	33,6
17	153	31,1

Директору необходимо для анализа ответить на следующие вопросы:

1) Определить, какой фактор исследуется (зависимый) и какой влияющий (независимый).

В соответствии с условиями исследования директор должен проанализировать показатель «спрос на продукцию известной торговой марки», обозначим эту переменную Y. В качестве влияющей (независимой, объясняющей) переменной возьмем количество проведенных мероприятий по продвижению продукции данной торговой марки и обозначим переменную – X.

2) Построить поле корреляции результата и фактора и сформулировать гипотезу о форме связи между товарооборотом и количеством проводимых мероприятий (записать предполагаемую регрессионную модель).

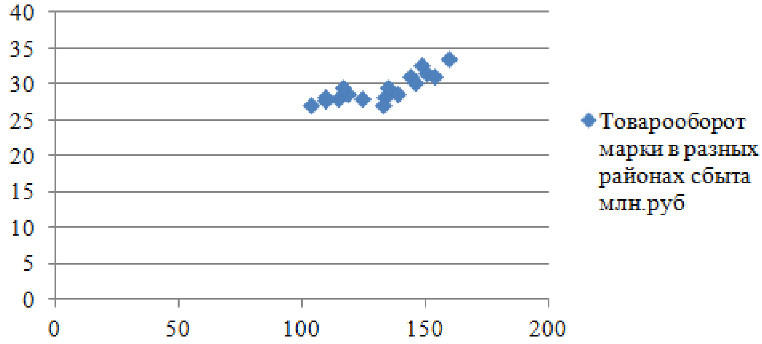


Рис. Поле корреляции

На основании поля корреляции и эмпирических знаний о связи между товарооборотом и продвижением товара можно выдвинуть гипотезу о линейной зависимости спроса на товар от политики продвижения данной торговой марки. Вытянутость облака точек на графике «Поле корреляции» вдоль предполагаемой прямой показывает, что в среднем, с увеличением количества мероприятий увеличивается объем товарооборота.

Предполагаемая регрессионная модель будет иметь вид: $y_i = a + bx_i + \varepsilon_i, i = 1 ... n$.

3) Для более полного исследования исходных данных следующий этап будет связан с получением описательных статистик. *Используя выборочные данные, заполним табл. 2.*

По полученным статистикам вычислим средние значения по Y и X, дисперсии по Y и X, средние квадратичные отклонения по Y и X, коэффициент ковариации.

Таблица 2

№ п/п	Количество проводимых маркетинговых мероприятий по сбыту (ед.) X	Товарооборот марки в разных районах сбыта (млн руб.) Y	X^2	Y^2	$X \cdot Y$
1	109	27,9	11881	778,41	3041,1
2	114	28,1	12996	789,61	3203,4
3	109	28,3	11881	800,89	3084,7
4	103	27,2	10609	739,84	2801,6
5	118	28,8	13924	829,44	3398,4
6	116	29,6	13456	876,16	3433,6
7	124	28,1	15376	789,61	3484,4
8	133	28,3	17689	800,89	3763,9
9	132	27,2	17424	739,84	3590,4
10	138	28,8	19044	829,44	3974,4
11	143	31,1	20449	967,21	4447,3
12	148	32,7	21904	1069,29	4839,6
13	134	29,5	17956	870,25	3953
14	145	30,2	21025	912,04	4379
15	150	31,5	22500	992,25	4725
16	159	33,6	25281	1128,96	5342,4
17	153	31,1	23409	967,21	4758,3
Сумма	2228	502	296804	14881,34	66220,5

Результаты расчетов и соответствующие формулы поместим в табл. 3.

Таблица 3

Формула	Значение	Единицы измерения
$\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n}$	$\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{2228}{17} = 131,06$	ед. (количество мероприятий)
$\overline{Y} = \frac{\sum_{i=1}^{n} y_i}{n}$	$\overline{Y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{502}{17} = 29,53$	млн руб.

Окончание табл. 3

Формула	Значение	Единицы измерения
Формула для дисперсии по X $S_X^2 = \overline{X^2} - \left(\overline{X}\right)^2$	$S_X^2 = \overline{X^2} - \left(\overline{X}\right)^2 = = \dfrac{296804}{17} - 131{,}06^2 = 282{,}34$	(ед.^{2})
$S_Y^2 = \overline{Y^2} - \left(\overline{Y}\right)^2$	$S_Y^2 = \overline{Y^2} - \left(\overline{Y}\right)^2 = = \dfrac{14881{,}34}{17} - 29{,}53^2 = 3{,}39$	(млн руб.)^2
$S_X = \sqrt{S_X^2}$	$S_X = \sqrt{S_X^2} = \sqrt{282{,}39} = 16{,}81$	ед.
$S_Y = \sqrt{S_Y^2}$	$S_Y = \sqrt{S_Y^2} = \sqrt{3{,}39} = 1{,}84$	млн руб.
$K_*(X,Y) = = \overline{XY} - \overline{X} \cdot \overline{Y}$	$K_*(X,Y) = \overline{XY} - \overline{X} \cdot \overline{Y} = = \dfrac{66220{,}5}{17} - 131{,}06 \cdot 29{,}53 = 25{,}23$	(ед.)*(млн руб.)

4) Определить выборочный коэффициент корреляции и поясните его смысл. Сделать вывод о силе линейной зависимости между переменными Y и X .

Коэффициент ковариации и средние квадратичные отклонения были найдены в пункте 3). Подставим найденные статистики в выражение для коэффициента корреляции.

$$r_*(X,Y) = \frac{K_*(X,Y)}{S_X S_Y} = \frac{K_*(X,Y)}{\sqrt{S_X^2 \cdot S_Y^2}} = \frac{25{,}23}{16{,}81 \cdot 1{,}84} = 0{,}816 \; .$$

Значение коэффициента корреляции, также, как и поле корреляции указывает на то, что существует прямая линейная зависимость между изучаемыми показателями.

5) Оценить значимость коэффициента корреляции на уровне значимости 0,05.

Если признаки X и Y в генеральной совокупности имеют нормальное распределение, то появляется возможность оценки незначимости коэффициента корреляции.

Проверим нулевую гипотезу $H_0 : r(X,Y)=0$ о том, что генеральный коэффициент корреляции незначим, против конкурирующей гипотезы $H_1 : r(X,Y)\neq 0$ о том, что генеральный коэффициент корреляции значим.

Фактическое значение ($K_{выч}$) *t*-статистики критерия

$$K_{выч} = \frac{r_*(X,Y)\sqrt{n-2}}{\sqrt{1-r_*^2(X,Y)}} = \frac{0,816\cdot\sqrt{17-2}}{\sqrt{1-0,816^2}} = 5,45.$$

Табличное значение критерия Стьюдента найдем по табл. 3:

$$K_{таб} = t_{1-\alpha}\left[St(n-2)\right] = t_{0,95}\left[St\left(17-2\right)\right] = x_{0,975}\left[St\left(15\right)\right] = 2,13.$$

Гипотеза о незначимости коэффициента корреляции можно отвергнуть, поскольку $\left|K_{выч}\right| > K_{таб}$. Таким образом, наблюдается достаточно сильная и значимая линейная зависимость между мероприятиями по продвижению известной марки и товарооборотом.

6) *Оценить параметры уравнения парной линейной регрессии и дать интерпретацию коэффициента регрессии b.*

7) Все промежуточные вычисления представлены в таблице выше:

Найдем оценки параметров по формулам

$$\begin{cases} \hat{b} = \dfrac{K_*(X,Y)}{S_X^2} = \dfrac{25,23}{282,34} = 17,83, \\ \hat{a} = \overline{Y} - \hat{b}\overline{X} = 25,53 - 17,829\cdot 131,06 = 0,09. \end{cases}$$

Оценка функции регрессии зависимости спроса на продукцию известной торговой марки от количества проведенных мероприятий по продвижению продукции данной торговой марки имеет следующий вид:

$$\hat{Y}_X = \hat{a} + \hat{b}X = 17,83 + 0,09 \cdot X$$

Связь между Y и X определяет знак коэффициента регрессии b (если коэффициент b *положителен, то* прямая связь между переменными, иначе – обратная). В нашем примере связь прямая, т.е. с увеличением количества мероприятий товарооборот также увеличивается.

Интерпретация найденных параметров модели.

Коэффициент b показывает среднее изменение результативного показателя (в единицах измерения Y) с повышением величины фактора X на единицу его измерения. Для сквозного примера интерпретация будет следующая: с увеличением мероприятий по продвижению марки на одно мероприятие товарооборот в среднем увеличится на 0,09 млн. рублей.

Параметр a формально определяет прогнозируемый уровень Y, но только в том случае, если X=0 находится близко к выборочным наблюдениям. Но если X=0 находится далеко от выборочных значений x, то буквальная интерпретация может привести к неверным результатам, и даже если линия регрессии довольно точно описывает значения наблюдаемой выборки, нет гарантий, что также будет при экстраполяции влево или вправо. В данном случае, если даже мероприятия по продвижению марки не будут проводиться, то и тогда в среднем товарооборот составит около 18 млн. руб.

8) Для верификации модели, вычислим регрессионные остатки.

Подставив в уравнение регрессии значения X из выборки, можно определить выровненные (средние) значения результативного показателя для каждого наблюдения. Для рассматриваемого примера, подставляя в полученное оценочное уравнение регрессии значения переменной X (количество мероприятий), получим для каждого значения X выровненное (среднее) в соответствии с уравнением оценочное значение товарооборота (табл. 4).

Таблица 4

№ п/п	Товарооборот марки в разных районах сбыта (млн руб.) Y	Расчетные значения зависимой переменной (оценка товарооборота марки в разных районах сбыта (млн руб.) $\hat{Y}$	$e = Y - \hat{Y}$
1	27,9	27,56006684	0,3399
2	28,1	28,00645169	0,0935
3	28,3	27,56006684	0,7399
4	27,2	27,02440502	0,1756
5	28,8	28,36355957	0,4364
6	29,6	28,18500563	1,4150
7	28,1	28,89922139	-0,7992
8	28,3	29,70271412	-1,4027
9	27,2	29,61343715	-2,4134
10	28,8	30,14909897	-1,3491
11	31,1	30,59548382	0,5045
12	32,7	31,04186866	1,6581
13	29,5	29,79199109	-0,2920
14	30,2	30,77403776	-0,5740
15	31,5	31,2204226	0,2796
16	33,6	32,02391533	1,5761
17	31,1	31,48825351	-0,3883
			$\sum_{i=1}^{n} e_i = 0$

9) Заполним таблицу дисперсионного анализа.

Общая сумма квадратов *TSS (Total Sum of Squares)*	$TSS = nS_Y^2 = \sum_{i=1}^{n} \left(y_i - \overline{y} \right)^2$	57,57
Объясненная регрессией сумма квадратов *ESS (Explained Sum of Squares)*	$ESS = \sum_{i=1}^{n} \left(\hat{y}_i - \overline{y} \right)^2$	38,29
Необъясненная уравнением регрессии сумма квадратов *RSS (Residual Sum of Squares)*	$RSS = \sum e_i^2$, где $e_i = y_i - \hat{y}_i$	19,28
Основное тождество дисперсионного анализа	*TSS=ESS+RSS*	57,57 = =38,29 + 19,28

10) Определим стандартные ошибки модели и стандартные ошибки оценок параметров модели. Сделайте выводы.

Стандартная ошибка модели $\hat{\sigma}$	Стандартная ошибка оценки параметра b $S_{\hat{b}}$	Стандартная ошибка оценки параметра a $S_{\hat{a}}$
$\hat{\sigma} = \sqrt{\hat{\sigma}^2} = \sqrt{\dfrac{RSS}{n-m}}$	$S_{\hat{b}} = \sqrt{\hat{\sigma}^2 \cdot \dfrac{1}{n \cdot S_X^2}}$	$S_{\hat{a}} = \sqrt{\hat{\sigma}^2 \cdot \dfrac{\overline{X^2}}{n \cdot S_X^2}}$
0,64	0,02	2,16

Стандартная ошибка оценки параметра регрессии используется для проверки существенности коэффициента регрессии и для расчета его доверительных интервалов. Стандартная ошибка остаточной компоненты служит мерой точности модели, чем меньше (ближе к нулю), тем лучше.

11) Определите коэффициент детерминации и дайте его интерпретацию.

Формула	Значение	Интерпретация
$R^2 = 1 - \dfrac{RSS}{TSS} = \dfrac{ESS}{TSS}$	0,67	Доля дисперсии зависимой переменной Y, объясненная построенной линейной моделью составляет 67%, остальные 33% дисперсии объяснить не удалось, то есть, 67%-я доля вариации товарооборота учтена в модели и обусловлена влиянием мероприятий, связанных с продвижением марки.

12) На уровне значимости 0,05 оцените статистическую значимость коэффициента регрессии b. Сделайте выводы.

Выдвинем гипотезы: основную о незначимости параметра b и альтернативную.

$H_0: b = 0$ (*коэффициент незначим*), генеральный коэффициент регрессии равен нулю (или иначе: коэффициент регрессии генеральной совокупности незначим) *(statistically unsignificant).*

$H_1: b \neq 0$ (*коэффициент значим*) о том, что генеральный коэффициент регрессии значим *(statistically significant).*

Проверка гипотезы осуществляется с помощью t-статистики следующего вида: $t - statistics = \dfrac{\hat{b}}{S_{\hat{b}}}$.

Вычислим значение t-статистики $K_{выч} = \dfrac{\hat{b}}{S_b} = \dfrac{0,09}{0,02} = 4,5.$

Критическое значение $K_{таб}$ найдем по таблицам Стьюдента (табл. А3):

$$K_{таб} = t_{1-0,05}\left[St(17-2)\right] = t_{0,95}\left[St(15)\right] =$$
$$= x_{\frac{1+0,95}{2}}\left[St(15)\right] = x_{0,975}\left[St(15)\right] = 2,13.$$

Поскольку $|K_{выч}| > K_{таб}$, то нулевую гипотезу следует отвергнуть на уровне значимости 0,05.

13) На уровне значимости 0,05 оцените статистическую значимость уравнения регрессии в целом. Сделайте выводы.

Один из первых шагов, связанных с оценкой качества модели, состоит в нахождении коэффициента детерминации, при этом важным моментом является проверка значимости уравнения регрессии в целом.

Проверим гипотезу о значимости модели с помощью критерия Фишера, для этого будем проверять основную гипотезу о незначимости коэффициента детерминации генеральной совокупности:

$H_0: R^2_Г = 0$ (*модель регрессии незначима*);

$H_0: R^2_Г \neq 0$ (*модель регрессии значима*).

Вычислим значение статистики F-критерия по формуле:

$$K_{выч} = \frac{R^2}{1-R^2} \cdot \frac{n-m}{m-1} = \frac{0,67}{1-0,67} \cdot \frac{17-2}{2-1} = 29,8.$$

Критическое значение статистики критерия находим по табл. А6:

$$K_{таб} = x_{1-\alpha}[F(m-1,\ n-m)] =$$
$$= x_{1-0,05}[F(2-1, 17-2)] = x_{0,95}[F(1,15)] = 4,54.$$

Так как $K_{выч} \geq K_{таб}$, то гипотезу о незначимости H_0 следует отклонить.

14) Постройте 95%-ые доверительные интервалы для параметров модели a и b. Сделайте выводы.

$$\hat{a} - t_\gamma[St(n-m)] \cdot S_{\hat{a}} < a < \hat{a} + t_\gamma[St(n-m)] \cdot S_{\hat{a}} \rightarrow$$
$$\rightarrow 17,83\ -\ 2,13 \cdot 2,16 < a <\ 17,83\ +\ 2,13 \cdot 2,16,$$

$$\hat{b} - t_\gamma[St(n-m)] \cdot S_{\hat{b}} < b < \hat{b} + t_\gamma[St(n-m)] \cdot S_{\hat{b}} \rightarrow$$
$$\rightarrow 0,09\ - 2,13 \cdot\ 0,02\ < b <\ 0,09\ +\ 2,13 \cdot\ 0,02.$$

Следовательно, границы доверительного интервала будут такими:

Параметр	*Левая граница 95%-го доверительного интервала*	*Правая граница 95%-го доверительного интервала*
a	13,22	22,44
b	0,05	0,12

Вывод. С вероятностью 0,95 можно утверждать, что значение параметра a находится в пределах от 13,22 до 22,44, а параметра b – в пределах от 0,05 до 0,12.

15) С вероятностью 0,95 постройте доверительный интервал ожидаемого значения результативного признака, если факторный признак увеличится на 5% от своего среднего значения.

Сначала определим точечный прогноз, для этого найдем прогнозное значение факторного признака: $x_* = 1,05 \cdot \overline{X} = 1,05 \cdot\ 131,06 =$ $= 137,61$.

Точечный прогноз среднего значения зависимой переменной определяется статистикой: $\hat{y}_* = \hat{a} + \hat{b}x_*$. Подставим в уравнение $\hat{Y}_X = 17,83 + 0,09 \cdot X$ прогнозное значение x_*, получим $\hat{y}_* = 17,83 + 0,09 \cdot 131,06 = 29,63$, т.е. если количества проведенных мероприятий по продвижению продукции данной торговой марки увеличится на 5 % от своего среднего уровня, то спрос на продукцию будет в среднем 29,63 млн руб.

Для построение интервального прогноза найдем стандартную ошибку ожидаемого значения зависимой переменной
$$S_{Y_X} = \sqrt{\hat{\sigma}^2 \cdot \left(\frac{1}{n} + \frac{(x_* - \overline{X})^2}{n \cdot S_X^2}\right)} = \sqrt{0,41 \cdot \left(\frac{1}{17} + \frac{(137,61 - 131,06)^2}{7 \cdot 282,64}\right)} = 0,18.$$

Подставим все найденные значения в выражение для интервального прогноза среднего значения зависимой переменной: $\hat{y}_* - t_\gamma[St(n-m)] \cdot S_{Y_X} < \mathrm{M}\left[y_*\right] < \hat{y}_* + t_\gamma[St(n-m)] \cdot S_{Y_X}$, получим $29,63 - 2,13 \cdot 0,18 < M[y_*] < 29,63 + 2,13 \cdot 0,18$ и, следовательно, границы доверительного интервала будут такими:

Левая граница 95%-го доверительного интервала	*Правая граница 95%-го доверительного интервала*
29,25	30,01

С вероятностью 0,95 можно утверждать, что среднее значение товарооборота известной торговой марки при проведении рекламных мероприятий в количестве 131,06 единиц будет находиться в интервале от 29,25 до 30,01 (млн. руб.)

16) С вероятностью 0,95 постройте доверительный интервал индивидуального значения результативного признака, если факторный признак увеличится на 5 % от своего среднего значения.

Доверительный интервал для индивидуального значения зависимой переменной имеет вид:

$$\hat{y}_* - t_\gamma[St(n-m)]\cdot S_{y_*} < y_* < \hat{y}_* + t_\gamma[St(n-m)]\cdot S_{y_*}$$

Значения всех статистик для данного неравенства уже рассчитаны выше, кроме стандартной ошибки $S_{y_*} = \sqrt{\hat{\sigma}^2 \cdot \left(1 + \frac{1}{n} + \frac{(x_* - \overline{X})^2}{n \cdot S_X^2}\right)} =$

$$= \sqrt{0,41 \cdot \left(1 + \frac{1}{17} + \frac{(137,61 - 131,06)^2}{7 \cdot 282,64}\right)} = 0,67 .$$

Подставим все найденные значения в выражение для интервального прогноза индивидуального значения зависимой переменной:

29,63 – 2,13·0,67 < $M[y_*]$ < 29,63 + 2,13·0,67, и, следовательно, границы доверительного интервала будут такими:

Левая граница 95%-го доверительного интервала	*Правая граница 95%-го доверительного интервала*
28,20	31,06

С вероятностью 0,95 можно утверждать, что индивидуальное значение товарооборота известной торговой марки при проведении рекламных мероприятий в количестве 131,06 единиц будет находиться в интервале от 28,20 до 31,06 (млн. руб.)

Заметим, что интервал для индивидуального прогнозного значения зависимой переменной всегда шире по сравнению с интервалом для ожидаемого (среднего) значения зависимой переменной.

Сделаем общие выводы по модели:

Если для рассмотренной парной линейной регрессионной модели выполняются условия Гаусса-Маркова, то прогнозирование по данной модели возможно в краткосрочном периоде.

Тесты для проверки знаний

1. МНК используется для оценивания
1) параметров линейной регрессии
2) величины коэффициента детерминации
3) средней ошибки аппроксимации
4) величины коэффициента корреляции
5) среди приведенных нет правильного ответа

2. Величина коэффициента регрессии показывает
1) среднее изменение результата при увеличении объясняющей переменной на единицу
2) тесноту связи между объясняющей переменной и результатом
3) характер связи между объясняющей переменной и результатом
4) тесноту связи между исследуемыми переменными
5) среди приведенных нет правильного ответа

3. Какое из приведенных значений может быть коэффициентом детерминации
1) 0,99 2) –1,2 3) 1,2 4) 2 5) –0,7

4. Гипотеза о значимости в целом уравнения линейной регрессии проверяется с помощью критерия…
1) Фишера
2) Дарбина-Уотсона
3) Стьюдента
4) Пирсона
5) среди приведенных нет правильного ответа

5. Какое из следующих утверждений истинно.
1) $0 \le R^2 \le 1$
2) Для парной линейной регрессии коэффициент корреляции превосходит коэффициент детерминации

3) Оценки параметров парной линейной регрессионной модели могут принимать только положительные значения

4) RSS = ESS + TSS

5) среди приведенных нет правильного ответа

6. Значение коэффициента корреляции равно 0,81. Можно сделать вывод о том, что связь между результативным признаком и факторами является……

1) достаточно тесной
2) слабой
3) не тесной
4) функциональной
5) среди приведенных нет правильного ответа

7. Метод наименьших квадратов используется для оценивания

1) параметров линейной регрессии
2) средней ошибки аппроксимации
3) величины коэффициента детерминации
4) величины коэффициента корреляции
5) среди приведенных нет правильного ответа

8. Суть МНК состоит в:

1) минимизации суммы квадратов значений зависимой переменной;

2) минимизации суммы квадратов отклонений точек наблюдений от уравнения регрессии;

3) минимизация суммы квадратов отклонений точек эмпирического уравнения регрессии от точек теоретического уравнения регрессии

4) минимизации суммы квадратов коэффициентов регрессии

5) среди приведенных нет правильного ответа

9. Коэффициент детерминации равен нулю при:

1) ESS=0
2) RSS=0
3) TSS=0
4) Нет правильного ответа

10. Независимые переменные в регрессионных моделях называются:

1) откликами;
2) возмущениями;
3) регрессорами;
4) остатками
5) среди приведенных ответов нет правильного

11. Предпосылкой МНК является то, что…

1) остаточные величины имеют неслучайный характер
2) остаточные величины имеют случайный характер
3) при уменьшении моделируемых значений результативного признака значение остатка уменьшается
4) при увеличении моделируемых значений результативного признака значение остатка увеличивается

12. Относительно формы зависимости различают…

1) линейную и нелинейную регрессии
2) положительную и отрицательную регрессии
3) простую и множественную регрессии
4) линейную и множественную регрессии
5) среди приведенных ответов нет правильного

13. Если коэффициент регрессии является несущественным, то его значение приравнивается к…

1) нулю и соответствующий фактор включается из модели
2) к табличному значению и соответствующий фактор не включается в модель

3) нулю и соответствующий фактор не включается в модель
4) к единице и не влияет на результат
5) среди приведенных ответов нет правильного

14. Доля остаточной дисперсии зависимой переменной в ее общей дисперсии составила 30%, следовательно, величина
1) Коэффициента детерминации равна 0,7;
2) Коэффициента детерминации равна 0,3;
3) Коэффициента регрессии равна 0,3;
4) Множественный коэффициент регрессии равен 0,7
5) Среди приведенных ответов нет правильного

15. Критерий Фишера в экономических моделях служит
1) Для оценки параметров регрессии
2) Показателем линейной связи между переменными
3) Показателем преимущества выбранной модели перед другими
4) Для проверки статистической модели
5) Среди приведенных ответов нет правильного

16. Коэффициент регрессии считается статистически значимым, если справедливо следующее утверждение
1) Доверительный интервал для этого коэффициента не содержит 0;
2) Фактическое значение статистики Стьюдента для этого коэффициента по модулю меньше критического (табличного);
3) Вероятность того, что этот коэффициент равен 0 близка к 1;
4) Доверительный интервал для этого коэффициента содержит 1
5) Среди приведенных ответов нет правильного

17. Значение индекса корреляции, рассчитанное для нелинейного уравнения регрессии, характеризует тесноту … связи.
1) Нелинейной;
2) Линейной;

3) Случайной;
4) Обратно пропорциональной
5) Среди приведенных ответов нет правильного

18. Как связаны между собой коэффициент корреляции и коэффициент парной линейной регрессии
1) всегда имеют одинаковые знаки
2) всегда имеют разные знаки
3) не связаны ни при каких условиях
4) в некоторых случаях связаны, а в некоторых не связаны
5) среди приведенных ответов нет правильного

19. Методом наименьших квадратов по 12 наблюдениям получена оценка функции регрессии $Y_X = 1 + 0,4X$, $R^2 = 0,6$. Проверяется гипотеза о существенности (значимости) множественной регрессии на уровне значимости $\alpha = 0,05$. Значение F-статистики для проверки этой гипотезы равно:
1) 15
2) 20
3) 4
4) 6
5) Среди приведенных ответов нет правильного

20. По данным 10 фирм получено уравнение регрессии для объема реализации товарной продукции *Y* в зависимости от затрат на рекламу *X*: $\hat{Y}_X = 6 - 4X$ в предположении, что результаты этих наблюдений связаны моделью $y_i = a + bx_i + \varepsilon_i$. Были найдены также стандартные ошибки оценок параметров регрессии: $S_a = 0.2, S_b = 0.5.$ Значение *t*-статистики при проверке гипотезы о значимости (существенности) коэффициента *b* оказалось равным
1) 1 2) –8 3) 12 4) 3 5) Среди приведенных ответов нет правильного

21. В результате анализа данных найдены $\overline{X}=1$, $\overline{Y}=4$, $\overline{XY}=1$, $S_X^2=4$, $S_Y^2=1$. В этом случае точечная оценка функции регрессии $Y_X=a+bX$ при $X=1$ равна

1) 4,0 2) 2,5 3) 1,5 4) 3 5) 4,5

22. По результатам 10 наблюдений исследовалась зависимость Y от X. Найдены значения следующих статистик: $\overline{X}=1$, $\overline{Y}=4$, $\overline{XY}=1$, $S_X^2=4$, $S_Y^2=10$, $RSS=4$. В этом случае с вероятностью 0,8 разница между верхней и нижней границами интервальной оценки функции регрессии $Y_X=a+bX$ при $X=3$ равна:

1) 0,89 2) 0,18 3) 2,48 4) 1,78 5) Среди приведенных ответов нет правильного

23. Остаточная сумма квадратов RSS равна 10. Выборочная дисперсия зависимой переменной по 10 наблюдениям равна 2. Тогда коэффициент детерминации R^2 равен

1) 0.5 2) 0.975 3) 0.95 4) 0.125 5) Среди приведенных ответов нет правильного

24. Индекс к корреляции равен 0,7. Чему равен коэффициент детерминации

1) 0,49 2) 0,7 3) 0,09 4) 0,1 5)Среди приведенных ответов нет правильного

25. По данным 10 фирм получено уравнение регрессии для объема реализации товарной продукции Y в зависимости от затрат на рекламу X: $\hat{Y}_X=6+3X$ в предположении, что результаты этих наблюдений связаны моделью $y_i=a+bx_i+\varepsilon_i$. Были найдены также стандартные ошибки оценок параметров регрессии: $S_a=2, S_b=3$. Значение t-статистики при проверке гипотезы о значимости (существенности) коэффициента b оказалось равным

1) 1 2) 3 3) 1,5 4) 2

26. В результате анализа данных найдены $\overline{X}=1$, $\overline{Y}=4$, $\overline{XY}=1$, $S_X^2=4$, $S_Y^2=1$. В этом случае точечная оценка функции регрессии $Y_X=a+bX$ при $X=1$ равна

1) 4,0 2) 2,5 3) 1,5 4) 3 5) 4,5

27. Оцененное уравнение зависимости переменной Y от переменной X имеет вид $\widehat{Y}=3.4X-5.1$. Тогда коэффициент эластичности переменной Y по переменной X при $X=2$ равен

1) 4
2) 6,8
3) 3,4
4) –4,2
5) Среди приведенных ответов нет правильного

Глава 4. МНОЖЕСТВЕННЫЙ ЛИНЕЙНЫЙ РЕГРЕССИОННЫЙ АНАЛИЗ

Множественный регрессионный анализ: особенности спецификации модели, отбор факторов при построении множественной регрессии. Классическая нормальная линейная модель множественной регрессии, оценка параметров методом МНК и их свойства. Проверка гипотез о значимости параметров модели (критерий Стьюдента) и значимости всей модели (критерий Фишера). Оценка влияния объясняющих переменных на зависимую переменную. Определение интервальных прогнозов в множественной линейной регрессионной модели.

Множественная линейная регрессионная модель	
Модель множественной линейной регрессии – модель с k объясняющими переменными	$Y = b_0 + b_1 X_1 + \ldots + b_k X_k + \varepsilon$
Функция регрессии	$Y_X = b_0 + b_1 X_1 + \ldots + b_k X_k$
$b_0, b_1, \ldots, b_k$	неизвестные параметры модели
Случайная ошибка модели (***error***, или ***disturbance***)	ε Наличие случайного члена ε связано с воздействием на зависимую переменную Y других факторов, не учтенных в функции регрессии, а также с возможной нелинейностью модели и ошибками измерения
Моделью наблюдений множественной линейной регрессии $y_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + \ldots b_k x_{ik} + \varepsilon_i$, $i = 1 \ldots n$.	Имеются результаты n наблюдений переменных $X_1, \ldots, X_k, Y$, которые обычно обозначаются следующим образом: $\left(x_{i,1}, \ldots, x_{i,k}, y_i\right)$, $i = 1, 2, \ldots, n$. Здесь $x_{i,j}$ – результат i-го наблюдения переменной X_j; y_i – результат i-го наблюдения переменной Y

Матричная форма записи модели	
$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$	*Модель множественной линейной регрессии с k объясняющими переменными*
$\mathbf{Y} = \left(y_1, \ldots, y_n\right)^T$	вектор значений зависимой переменной
$\boldsymbol{\beta} = \left(b_0, \ldots, b_k\right)^T$	вектор неизвестных параметров, m – число параметров модели, $m \le n$
$\mathbf{X} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1k} \\ 1 & x_{21} & \cdots & x_{2k} \\ \cdots & \cdots & \cdots & \cdots \\ 1 & x_{n1} & \cdots & x_{nk} \end{pmatrix}$	матрица данных с неслучайными элементами. Кроме того, будем считать матрицу $\mathbf{X}^\mathsf{T}\mathbf{X}$ невырожденной. Это предположение выполняется, если ранг матрицы $\mathbf{X}$ равен m, $m \le n$

Метод наименьших квадратов для оценки параметров модели (МНК)	
Имеются результаты n наблюдений переменных $X_1, \ldots, X_k, Y$, которые обычно обозначаются следующим образом: $\left(x_{i,1}, \ldots, x_{i,k}, y_i\right)$, $x_{i,j}$ – результат i-го наблюдения переменной X_j; y_i – результат i-го наблюдения переменной, $i = 1,...,n$, $j = 1,...,k$	
Результаты наблюдений связаны моделью наблюдений: $y_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + \ldots b_k x_{ik} + \varepsilon_i, i = 1 \ldots n$, где ε_i – случайная ошибка i-го наблюдения Y. МНК служит для нахождения оценок неизвестных параметров $b_0, b_1, \ldots, b_k$ модели	
Суть МНК	определить такие значения $b_0, b_1, \ldots, b_k$, при которых достигает своего минимального значения сумма квадратов ошибок модели в n наблюдениях: $Q = Q(a,b) = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - b_0 - b_1 x_{i1} - b_2 x_{i2} - \ldots b_k x_{ik})^2$
Оценки параметров модели, найденные МНК	$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\mathsf{T}\mathbf{X})^{-1}\mathbf{X}^\mathsf{T}\mathbf{Y}$

<table>
<tr><td>Точечные оценки параметров модели</td><td colspan="2">$\hat{\boldsymbol{\beta}} = \left(\hat{b}_0, \ldots, \hat{b}_k\right)^T$</td></tr>
<tr><td colspan="3">

$$\mathbf{X}^{\mathsf{T}}\mathbf{X} = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_{11} & x_{21} & \cdots & x_{n1} \\ \cdots & \cdots & \cdots & \cdots \\ x_{1k} & x_{2k} & \cdots & x_{nk} \end{pmatrix} \cdot \begin{pmatrix} 1 & x_{11} & \cdots & x_{1k} \\ 1 & x_{21} & \cdots & x_{2k} \\ \cdots & \cdots & \cdots & \cdots \\ 1 & x_{n1} & \cdots & x_{nk} \end{pmatrix} =$$

$$= \begin{pmatrix} n & \sum x_{i1} & \cdots & \sum x_{ik} \\ \sum x_{i1} & \sum x_{i1}^2 & \cdots & \sum x_{i1}x_{ik} \\ \cdots & \cdots & \cdots & \cdots \\ \sum x_{ik} & \sum x_{i1}x_{ik} & \cdots & \sum x_{ik}^2 \end{pmatrix} = n \begin{pmatrix} 1 & \bar{X}_1 & \cdots & \bar{X}_k \\ \bar{X}_1 & \overline{X_1^2} & \cdots & \overline{X_1 X_k} \\ \cdots & \cdots & \cdots & \cdots \\ \bar{X}_k & \overline{X_1 X_k} & \cdots & \overline{X_1^2} \end{pmatrix}$$

</td></tr>
<tr><td colspan="3">

$$\mathbf{X}^{\mathsf{T}}\mathbf{Y} = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_{11} & x_{21} & \cdots & x_{n1} \\ \cdots & \cdots & \cdots & \cdots \\ x_{1k} & x_{2k} & \cdots & x_{nk} \end{pmatrix} \cdot \begin{pmatrix} y_1 \\ y_2 \\ \cdots \\ y_n \end{pmatrix} = \begin{pmatrix} \sum y_i \\ \sum y_i x_{i1} \\ \cdots \\ \sum y_i x_{ik} \end{pmatrix} = n \begin{pmatrix} \bar{Y} \\ \overline{X_1 Y} \\ \cdots \\ \overline{X_k Y} \end{pmatrix}.$$

</td></tr>
<tr><td>Оценка функции регрессии (эмпирическая регрессия, выборочная регрессия)</td><td colspan="2">$\hat{Y}_X = \hat{b}_0 + \hat{b}_1 X_1 + \ldots + \hat{b}_k X_k$</td></tr>
<tr><td colspan="3">Интерпретация найденных параметров</td></tr>
<tr><td>$\hat{b}_0$</td><td colspan="2">дает оценку среднего значения зависимой переменной при $X_1 = 0, \ldots, X_k = 0$</td></tr>
<tr><td>$\hat{b}_i$</td><td colspan="2">показывает, на сколько единиц в среднем изменится переменная Y при увеличении независимой переменной X_i на единицу при прочих равных условиях</td></tr>
<tr><td>Регрессионные остатки (оценки погрешности)</td><td colspan="2">$e_i = y_i - \hat{y}_i$, где y_i – реальные значения зависимой переменной, $\hat{y}_i$ – расчетные значения</td></tr>
</table>

Предпосылки в регрессионном анализе и теорема Гаусса-Маркова

Рассмотрим линейную модель наблюдений: $y_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + + \ldots b_k x_{ik} + \varepsilon_i$, $i = 1 \ldots n$, где ε_i – случайная ошибка i-го наблюдения	
Условия, определяющие классическую линейную регрессионную модель	
1. *Правильная спецификация модели*	модель линейная по параметрам
2. **X** – детерминированная матрица, имеет максимальны ранг m, $m \le n$	переменные X_j являются детерминированными величинами, значения которых можно задать
3. *Математическое ожидание случайной ошибки в каждом наблюдении равно нулю* $M(\varepsilon_i) = 0$, $i = \overline{1,n}$	ошибки в среднем не оказывают влияния на зависимую переменную
4. *Дисперсия случайной ошибки постоянна для всех наблюдений* $D(\varepsilon_i) = M(\varepsilon_i^2) = \sigma^2$, $i = \overline{1,n}$.	гомоскедастичность модели
5. *Случайные ошибки различных наблюдений статистически не связаны (некоррелированы) между собой, т.е.* $M(\varepsilon_i \varepsilon_j) = 0$ *при* $i \ne j$	некоррелированность случайных ошибок для разных наблюдений
6. $\varepsilon_i \sim N(0; \sigma^2)$	случайные ошибки имеют совместное нормальное распределение
Теорема Гаусса – Маркова	
Если для модели $y_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + \ldots b_k x_{ik} + \varepsilon_i, i = 1 \ldots n,\ j = 1 ... k$ выполняются условия 1-5, то оценки $\hat{b}_0, \ldots, \hat{b}_k$, найденные с помощью МНК, являются наилучшими линейными несмещенными оценками (*Best Linear Unbiased Estimator*, или *BLUE*)) параметров $b_0, b_1, \ldots, b_k$.	

Основные статистики регрессионного анализа

$TSS = n \cdot S_Y^2 = \sum_{i=1}^{n}(y_i - \overline{y})^2$	*Полная сумма квадратов (total sum of squares)* характеризует общую изменчивость зависимой переменной.
$ESS = \sum_{i=1}^{n}(\hat{y}_i - \overline{y})^2$	*Сумма квадратов, объясненная моделью* (*explained sum of squares*), *или факторная сумма квадратов* определяет ту часть TSS, которую позволяет объяснить предложенная модель
$RSS = \sum e_i^2$	*Остаточная сумма квадратов* (*residual sum of squares*) определяет ту часть вариации, которую не позволяет объяснить предложенная модель.
$TSS = ESS + RSS$	Основное тождество дисперсионного анализа
$\hat{\sigma}^2 = \dfrac{RSS}{n-m}$	*Оценка дисперсии (ошибки) модели* – несмещенная оценка дисперсии ошибки модели
$\hat{\sigma} = \sqrt{\hat{\sigma}^2}$	*Стандартная ошибка модели*
$\hat{\mathbf{D}}(\hat{\boldsymbol{\beta}}) = \hat{\sigma}^2 \cdot (\mathbf{X}^{\mathsf{T}}\mathbf{X})^{-1}$	*Несмещенная оценка дисперсионной матрицы* вектора оценок параметров регрессии
$S_{\hat{b}_i} = \sqrt{\hat{\sigma}^2 \cdot \left(\mathbf{X}^{\mathsf{T}}\mathbf{X}\right)_{ii}^{-1}}$	*Стандартные ошибки параметров регрессии.* Двойной индекс ii в формуле указывает на то, что элемент находится на $i+1$ - м (при $k = m+1$) или на i -м (при $k = m$) месте главной диагонали матрицы $(\mathbf{X}^{\mathsf{T}}\mathbf{X})^{-1}$
Коэффициент детерминации	
$R^2 = 1 - \dfrac{RSS}{TSS} = \dfrac{ESS}{TSS}$	*Коэффициент детерминации* определяет долю дисперсии зависимой переменной, объясненную множественной функцией регрессии: $Y_X = b_0 + b_1X_1 + b_2X_2 + \ldots + b_kX_k$.
Свойства коэффициента детерминации	
1.	$R^2 \in [0; 1]$
2.	Чем ближе R^2 к единице, тем выше качество подгонки зависимой переменной с помощью выбранной регрессионной модели и тем более точно функция $\hat{Y}_X$ аппроксимирует поведение Y с учетом изменения независимых переменных

3.	$R^2 = r_*^2(Y,\hat{Y})$ – коэффициент детерминации равен квадрату выборочного коэффициента корреляции $r_*(Y,\hat{Y})$ между зависимой переменной Y и ее оценкой $\hat{Y} \equiv \hat{Y}_X = \hat{b}_0 + \hat{b}_1 X_1 + \hat{b}_2 X_2 + \ldots + \hat{b}_k X_k$
4.	Вычисление R^2 корректно, если константа b_0 включена в уравнение регрессии
5.	Величину $R = r_*(Y,\hat{Y}) = \sqrt{R^2}$ принято называть *множественным коэффициентом корреляции (multiple-R)*. Чем ближе его значение к единице, тем в большей степени можно считать, что имеющиеся данные связаны множественной линейной регрессионной зависимостью
6.	Известно, что включение в модель одной или нескольких новых переменных может привести только к увеличению R^2. Коэффициент детерминации не может уменьшиться даже при включении в модель несущественных переменных. Более того, включением в модель достаточно большого числа переменных можно добиться того, что значение коэффициента детерминации станет равным единице.

Скорректированный коэффициент детерминации	
$R^2_{adj} = 1 - \frac{n-1}{n-m} \cdot \frac{RSS}{TSS} =$ $= 1 - \frac{n-1}{n-m} \cdot (1 - R^2)$	*это коэффициент детерминации* с поправкой на число степеней свободы

Свойства скорректированного коэффициента детерминации	
1.	$R^2_{adj} \in [0;\, 1]$
2.	Если увеличение доли объясненной регрессии при добавлении новой переменной мало, то скорректированный коэффициент детерминации может уменьшиться, следовательно, добавлять переменную нецелесообразно
3.	$R^2_{adj} < R^2$ при $n > m$ и $m > 1$

4.	При использовании коэффициента R^2_{adj} для выбора между конкурирующими моделями лучшей признается та из них, для которой этот коэффициент принимает максимальное значение

Информационный критерий Акайке (Akaike's information criterion – AIC**)**	
$AIC = \ln\left(\dfrac{RSS}{n}\right) + \dfrac{2k}{n}$	Применяется для выбора между несколькими альтернативными моделями, а также «штрафует» за увеличение количества объясняющих переменных в модели. Среди нескольких альтернативных моделей предпочтение отдается модели с наименьшим значением AIC, в которой достигается определенный компромисс между величиной остаточной суммы квадратов и количеством объясняющих переменных
Информационный критерий Шварца (Schwarz's information criterion – SC, SIC**)**	
$SC = \ln\left(\dfrac{RSS}{n}\right) + \dfrac{k \cdot \ln n}{n}$	Среди нескольких альтернативных моделей (полной и редуцированных) предпочтение отдается модели с наименьшим значением SC

Проверка гипотезы о незначимости параметра модели (критерий Стьюдента)

Проверка гипотезы при наличии двусторонней альтернативы	
Назначение	Используется для проверки гипотезы о незначимости параметра модели в генеральной совокупности
Данные	Имеются результаты n наблюдений переменных $X_1,\ldots,X_k,Y,$ которые обычно обозначаются следующим образом: $\left(x_{i,1},\ldots,x_{i,k},y_i\right)$, $x_{i,j}$ – результат i-го наблюдения переменной X_j; y_i – результат i-го наблюдения переменной

Допущения	Модель наблюдений: $y_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + \dots b_k x_{ik} + \varepsilon_i, i = 1 \dots n$, где ε_i – случайные величины, имеющие нормальное распределение
Алгоритм проверки гипотезы	
Шаг 1. Выдвинуть гипотезы основную и альтернативную на уровне значимости α	
Основная (нулевая) гипотеза	$H_0 : b_i = 0$ (*коэффициент незначим*), – генеральный параметр регрессии равен нулю (или иначе: коэффициент регрессии генеральной совокупности незначим) *(statistically unsignificant)*
Конкурирующая (альтернативная) гипотеза	$H_1 : b_i \neq 0$ (*коэффициент значим*) о том, что генеральный коэффициент регрессии значим *(statistically significant)*
Шаг 2. Найти критическое (табличное) значение статистики критерия Стьюдента	
Критическое значение критерия	$K_{таб} = t_{1-\alpha}[St(n-m)]$ двусторонний квантиль уровня $1-\alpha$ распределения Стьюдента с $n-m$ степенями свободы) находится по табл. А3
Шаг 3. Найти вычисленное (фактическое) значение статистики критерия Стьюдента	
Статистика критерия (фактическое значение)	$K_{выч} = t - statistics = \dfrac{\hat{b}_i}{S_{\hat{b}_i}}$
Шаг 4. Сравнить вычисленное значение с критическим	
Область принятия нулевой гипотезы	Гипотеза о значимости коэффициента регрессии отклоняется (т.е. не отвергаем гипотезу Y), если $b = 0$

Проверка гипотезы о значимости модели (критерий Фишера)

Назначение	Используется для проверки гипотезы о значимости *линейной регрессионной модели* в генеральной совокупности	
Данные	Имеются результаты n наблюдений переменных $X_1,\ldots,X_k,Y$, которые обычно обозначаются следующим образом: $\left(x_{i,1},\ldots,x_{i,k},y_i\right)$, $x_{i,j}$ – результат i-го наблюдения переменной X_j; y_i – результат i-го наблюдения переменной	
Допущения	Модель наблюдений: $y_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + \ldots b_k x_{ik} + \varepsilon_i, i = 1\ldots n$, где ε_i – случайные величины, имеющие нормальное распределение	
Алгоритм проверки гипотезы		
Шаг 1. Выдвинуть гипотезы основную и альтернативнуюна уровне значимости α		
Основная (нулевая) гипотеза	$H_0 : R^2_{\Gamma} = 0$ (*коэффициент детерминации генеральной совокупности равен нулю, модель регрессии незначима*)	
Конкурирующая гипотеза	$H_0 : R^2_{\Gamma} \neq 0$ (*модель регрессии значима*)	
Шаг 2. Найти критическое (табличное) значение статистики критерия		
Критическое значение критерия	$K_{таб} = x_{1-\alpha}\left[F(m-1,\, n-m)\right]$ квантиль уровня $1-\alpha$ распределения Фишера с $\nu_1 = m-1$ и $\nu_2 = n-m$ степенями свободы находится по табл. А5 и А6	
Шаг 3. Найти вычисленное (фактическое) значение статистики критерия Фишера		

Статистика критерия (фактическое значение)	$K_{выч} = F - statistics = \frac{R^2}{1-R^2} \cdot \frac{n-m}{m-1}$, где m – число параметров модели
Шаг 4. Сравнить вычисленное значение с критическим	
Область принятия нулевой гипотезы	Если $K_{выч} < K_{таб}$, то гипотезу H_0 о незначимости модели следует принять, т.е. нет оснований отклонить эту гипотезу

Доверительные интервалы для параметров линейного уравнения множественной регрессии

$\hat{b}_i - t_\gamma[St(n-m)]\ S_{\hat{b}_i} \le b_i \le$ $\le \hat{b} + t_\gamma[St(n-m)]\ S_{\hat{b}_i}$	Двусторонние доверительные интервалы
$b_i < \hat{b} + x_\gamma[St(n-m)] \cdot S_{\hat{b}_i}$	Правосторонние доверительные интервалы
$\hat{b}_i - x_\gamma[St(n-m)] \cdot S_{\hat{b}_i} < b_i$	Левосторонние доверительные интервалы

$\hat{b}_i$	МНК-оценка коэффициента b_i
$S_{\hat{b}_i}$	стандартная ошибка коэффициента регрессии $\hat{b}_i$
γ	доверительная вероятность, задается исследователем обычно 0,9; 0,95; 0,98
$t_\gamma[St(n-m)]$	двусторонний квантиль уровня γ распределения Стьюдента с $n-m$ степенями свободы находится по табл. А3
$x_\gamma[St(n-m)]$	односторонний квантиль уровня γ распределения Стьюдента с $n-m$ степенями свободы находится по табл. А3

Оценка влияния объясняющих переменных на зависимую переменную

<table>
<tr><td colspan="2">По коэффициентам регрессии обычно невозможно определить, какой из факторов оказывает наибольшее влияние на зависимую переменную, так как коэффициенты регрессии между собой несопоставимы (они, как правило, имеют различную размерность)</td></tr>
<tr><td>Парный коэффициент корреляции</td><td>показывает тесноту линейной взаимосвязи между переменными Y и X_j (формулы для расчетов приведены в главе 2)</td></tr>
<tr><td>Частный коэффициент корреляции</td><td>показывает тесноту линейной взаимосвязи между переменными Y и X_j, исключая влияние на их связь всех остальных независимых переменных (формулы для расчетов приведены в главе 2)</td></tr>
<tr><td>Частный средний коэффициент эластичности</td><td>показывает на сколько процентов в среднем изменится зависимая переменная с увеличением на 1 % определенного фактора X_j при фиксированном значении других факторов.
Рассчитывается по формуле $\bar{Э}_j = \hat{b}_j \frac{\bar{X}_j}{\bar{Y}}$,
где $\bar{X}_j$ – среднее значение изучаемого фактора X_j</td></tr>
<tr><td>Стандартизованный коэффициент регрессии (или бэта-коэффициент)</td><td>показывает на сколько величин S_Y изменится в среднем зависимая переменная Y при увеличении только одной j-й объясняющей переменной на S_{X_j}.
Рассчитывается по формуле $\hat{\beta}_j = \hat{b}_j \frac{S_{Xj}}{S_Y}$
Бэта-коэффициенты определены только для модели со свободным членом и могут быть найдены с помощью следующего соотношения:</td></tr>
</table>

	$\hat{\beta} = \begin{pmatrix} a_1 \\ \dots \\ a_k \end{pmatrix} = \mathbf{Q}_X^{-1} \cdot \begin{pmatrix} r_*(Y, X_1) \\ \dots \\ r_*(Y, X_k \end{pmatrix} = \mathbf{Q}_X^{-1} \cdot \mathbf{r}_{Y,X,}$ в котором $\mathbf{Q}_X$ – выборочная корреляционная матрица независимых переменных $X_1, X_2, ..., X_k$

Прогнозирование и интервальное оценивание в множественной линейной регрессионной модели

<table>
<tr><td colspan="3">Под прогнозированием в эконометрике понимается построение оценки будущего значения зависимой переменной Y при заданном значении независимых переменных $\mathbf{x}_* = \left(1, x_{*1}, \ldots, x_{*k}\right)^{\mathsf{T}}$.</td></tr>
<tr><td colspan="3">Прогнозируемое значение обозначим</td></tr>
<tr><td>$y_* = b_0 + \sum_{j=1}^{k} b_j x_{*j} + \varepsilon_*$</td><td colspan="2">индивидуальное значение зависимой переменной</td></tr>
<tr><td>$\mathbf{M}\left[y_*\right] = b_0 + \sum_{j=1}^{k} b_j x_{*j}$</td><td colspan="2">среднее значение зависимой переменной</td></tr>
<tr><td colspan="3">Различают</td></tr>
<tr><td>точечное прогнозирование</td><td colspan="2">оценка истинного значения прогноза (некоторое число)</td></tr>
<tr><td>$\hat{y}_* = \hat{b}_0 + \sum_{j=1}^{k} \hat{b}_j x_{*j}$</td><td colspan="2">точечный прогноз (точечная оценка) индивидуального и среднего значений зависимой переменной</td></tr>
<tr><td>интервальное прогнозирование</td><td colspan="2">интервал, в котором находится прогнозируемая величина с заданным уровнем надежности γ</td></tr>
<tr><td colspan="3">Интервальный прогноз ожидаемого (среднего) значения зависимой переменной</td></tr>
<tr><td colspan="3">$\hat{y}_* - t_\gamma[St(n-m)] \cdot S_{Y_X} < M[y_*] < \hat{y}_* + t_\gamma[St(n-m)] \cdot S_{Y_X}$</td></tr>
<tr><td colspan="2">$S_{\hat{Y}_X} = \sqrt{\widehat{\sigma^2}\left[\mathbf{x}_*\left(\mathbf{X}^{\mathsf{T}}\mathbf{X}\right)^{-1}\mathbf{x}_*^{\mathsf{T}}\right]}$</td><td>стандартная ошибка оценки среднего значения зависимой переменной $M[y_*]$</td></tr>
<tr><td colspan="3">Интервальный прогноз индивидуального значения зависимой переменной</td></tr>
</table>

$\hat{y}_* - t_\gamma[St(n-m)] \cdot S_{y_*} < y_* < \hat{y}_* + t_\gamma[St(n-m)] \cdot S_{y_*}$	
$S_{y_*} = \sqrt{\hat{\sigma}^2 \cdot \left[1 + \mathbf{x}_* \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1} \mathbf{x}_*^{\mathrm{T}}\right]}$	стандартная ошибка оценки индивидуального значения зависимой переменной y_*

Линейные регрессионные модели с фиктивными переменными

Фиктивная переменная (манекен, DUMMY)	вспомогательная переменная, используемая либо вместо реальной содержательной неколичественной переменной, либо для описания каких-либо содержательных обстоятельств, принимающая только два значения 0 и 1
Ситуация 1 **(применение фиктивных переменных к описанию влияния качественных переменных)**	
Пример: построить модель, описывающую зависимость размера заработной платы Y работника от стажа его работы X_1, возраста работника X_2, а также наличия высшего образования X_3. В данном случае переменная X_3 является качественной. Она может быть учтена в модели опосредованно через фиктивную переменную D, определяемую следующим образом: $D = \begin{cases} 0, & \text{если работник не имеет высшего образования,} \\ 1, & \text{если работник имеет высшее образование.} \end{cases}$ Если предположить, что зависимость носит линейный характер, то искомая модель будет определяться следующим выражением: $Y = b_0 + b_1 X_1 + b_2 X_2 + b_3 D + \varepsilon$. Важно отметить, что коэффициент при фиктивной переменной в подобных моделях, как правило, имеет определенную содержательную интерпретацию. В данном случае коэффициент b_3 показывает, насколько работник с высшим образованием получает в среднем «больше» работника без высшего образования при прочих равных обстоятельствах.	

Ситуация 2 **(применение фиктивных переменных к описанию обстоятельств, носящих сезонный или периодический характер)**
Пример: построить модель, описывающую зависимость объема потребляемого продукта Y от времени года. При этом предполагается, что никакие другие обстоятельства не сказываются на величине потребления. Поскольку известны 4 времени года (зима, весна, лето, осень), следует применить 3 фиктивные переменные: $D_1 = \begin{cases} 1, & \text{если определяется величина потребления зимой,} \\ 0 - & \text{в остальных случаях.} \end{cases}$ $D_2 = \begin{cases} 1, & \text{если определяется величина потребления весной,} \\ 0 - & \text{в остальных случаях,} \end{cases}$ $D_3 = \begin{cases} 1, & \text{если определяется величина потребления летом,} \\ 0 - & \text{в остальных случаях}. \end{cases}$ Поскольку функция $Y_D = b_0 + b_1 D_1 + b_2 D_2 + b_3 D_3$ в данном случае показывает средний объем потребления в любое время года, ее коэффициенты имеют определенный содержательный смысл: значение b_0 равно среднему объему потребления при $D_1 = D_2 = D_3 = 0,$ т.е. осенью; значение $b_0 + b_1$ равно среднему объему потребления при $D_1 = 1, D_2 = D_3 = 0,$ т.е. зимой, поэтому коэффициент b_1 показывает, насколько величина потребления зимой «больше», чем осенью. Аналогичную интерпретацию имеют коэффициенты b_2 и b_3. Таким образом, каждый из коэффициентов b_1, b_2, b_3 показывает средние сезонные отклонения в объеме потребления.
Ситуация 3 **(применение фиктивных переменных к построению кусочно-линейных регрессионных моделей)**
Пример: пусть x_t – размер основных фондов некоторого предприятия в период t, а y_t – объем продукции, выпущенной за этот же период. Предполагаем, что в любой момент времени зависимость между этими показателями является линейной, однако в известный момент времени t_0 происходит изменение уравнения регрессии. На формальном уровне такая зависимость может быть описана следующим образом:

$$y_t = b_0 + b_1 x_t + b_2 \left(x_t - x_{t_0} \right) D_t + \varepsilon_t, \quad t = 1, 2, \ldots, n.$$

В данном случае $D_t = 0$ при $t \le t_0$ и $D_t = 1$ при $t > t_0$. При этом линейная функция регрессии имеет коэффициент наклона b_1 при $t \le t_0$ и $b_1 + b_2$ при $t > t_0$. Проверка гипотезы $b_2 = 0$ в этом случае равносильна проверке предположения о том, что функция регрессии не изменяется с течением времени, т.е. *в модели отсутствуют структурные изменения.*

Пример. Имеются следующие данные о динамике объема продаж услуг Y (в млн руб.) некоторого сервисного предприятия от количества работающих X_1 (в тыс. чел.) и объема основных производственных средств X_2 (в млн руб.). Данные за 12 лет приведены в табл. 5, представленной ниже.

Предполагая, что между переменными Y, X_1 и X_2 существует линейная регрессионная зависимость, найти ее аналитическое выражение (уравнение регрессии Y по X_1 и X_2).

Таблица 5

Год	Y	X_1	X_2
2003	11	2,1	11
2004	13	2,2	11
2005	16	2,3	12
2006	24	2,4	13
2007	24	2,5	15
2008	26	2,5	17
2009	27	2,6	17
2010	28	2,7	18
2011	29	2,6	18
2012	33	2,7	19
2013	33	2,7	19
2014	36	2,7	22

По условию исходные переменные связаны соотношением $Y = b_0 + b_1 X_1 + b_2 X_2 + \varepsilon$. Вектор $\mathbf{Y}$ и матрица $\mathbf{X}$ в данном случае имеют следующий вид:

$$\mathbf{Y}=\begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix}=\begin{pmatrix} 11 \\ 13 \\ \dots \\ 36 \end{pmatrix},\quad \mathbf{X}=\begin{pmatrix} 1 & 2,1 & 11 \\ 1 & 2,2 & 11 \\ \dots & \dots & \dots \\ 1 & 2,7 & 22 \end{pmatrix}.$$

Для удобства результаты промежуточных вычислений поместим во вспомогательную табл. 6.

Получим следующую матрицу:

$$\mathbf{X}^{\mathrm{T}}\mathbf{X}=\begin{pmatrix} 1 & 1 & \cdots & 1 \\ 2,1 & 2,2 & \cdots & 2,7 \\ 11 & 11 & \cdots & 22 \end{pmatrix}\cdot\begin{pmatrix} 1 & 2,1 & 11 \\ 1 & 2,2 & 11 \\ \dots & \dots & \dots \\ 1 & 2,7 & 22 \end{pmatrix}=\begin{pmatrix} 12 & 30 & 192 \\ 30 & 75,48 & 487,7 \\ 192 & 487,7 & 32,12 \end{pmatrix}$$

(см. суммы в итоговой строке таблицы).

$$\mathbf{X}^{\mathrm{T}}\mathbf{Y}=\begin{pmatrix} 1 & 1 & \cdots & 1 \\ 2,1 & 2,2 & \cdots & 2,7 \\ \dots & \dots & \dots & \dots \\ 11 & 11 & \cdots & 22 \end{pmatrix}\cdot\begin{pmatrix} 11 \\ 13 \\ \dots \\ 36 \end{pmatrix}=\begin{pmatrix} 300 \\ 767,7 \\ 5101 \end{pmatrix}.$$

Таблица 6

Год	Y	X_1	X_2	Y^2	X_1^2	X_2^2	X_1X_2	X_1Y	X_2Y
2003	11	2,1	11	121	4,41	121	23,1	23,1	121
2004	13	2,2	11	169	4,84	121	24,2	28,6	143
2005	16	2,3	12	256	5,29	144	27,6	36,8	192
2006	24	2,4	13	576	5,76	169	31,2	57,6	312
2007	24	2,5	15	576	6,25	225	37,5	60	360
2008	26	2,5	17	676	6,25	289	42,5	65	442
2009	27	2,6	17	729	6,76	289	44,2	70,2	459
2010	28	2,7	18	784	7,29	324	48,6	75,6	504
2011	29	2,6	18	841	6,76	324	46,8	75,4	522
2012	33	2,7	19	1089	7,29	361	51,3	89,1	627
2013	33	2,7	19	1089	7,29	361	51,3	89,1	627
2014	36	2,7	22	1296	7,29	484	59,4	97,2	792
Суммы	300	30	192	8202	75,48	3212	487,7	767,7	5101

Матрицу $\mathbf{A}^{-1} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}$ определим по формуле $\mathbf{A}^{-1} = \dfrac{1}{\det \mathbf{A}} \mathbf{A}^{*}$, где $\det \mathbf{A}$ – определитель матрицы $\mathbf{X}^{\mathrm{T}}\mathbf{X}$, а $\mathbf{A}^{*}$ – матрица, присоединенная к матрице $\mathbf{X}^{\mathrm{T}}\mathbf{X}$. Получим

$$\mathbf{A}^{-1} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1} = \frac{1}{94{,}92}\begin{pmatrix} 4590{,}426 & -2721{,}64 & 138{,}868 \\ -2721{,}64 & 1679{,}989 & 92{,}3572 \\ 138{,}868 & -92{,}3572 & 5{,}79012 \end{pmatrix} =$$

$$= \begin{pmatrix} 48{,}361 & -28{,}673 & 1{,}463 \\ -28{,}673 & 17{,}699 & -0{,}973 \\ 1{,}463 & -0{,}973 & 0{,}061 \end{pmatrix}.$$

Умножая эту матрицу на вектор $\mathbf{X}^{\mathrm{T}}\mathbf{Y}$, получим

$$\hat{\boldsymbol{\beta}} = \begin{pmatrix} 48{,}361 & -28{,}673 & 1{,}463 \\ -28{,}673 & 17{,}699 & -0{,}973 \\ 1{,}463 & -0{,}973 & 0{,}061 \end{pmatrix} \cdot \begin{pmatrix} 300 \\ 767{,}7 \\ 5101 \end{pmatrix} = \begin{pmatrix} -42{,}23 \\ 20{,}27 \\ 1{,}04 \end{pmatrix}.$$

Поэтому выборочное уравнение множественной регрессии, характеризующее зависимость объема продаж услуг от количества работающих и объема основных производственных средств, имеет следующий вид:

$$\hat{Y}_X = -42{,}23 + 20{,}27 X_1 + 1{,}04 X_2 .$$

Данное уравнение показывает, что при увеличении количества работающих (X_1) на 1 тыс. чел. (при неизменном X_2) средний объем продаж услуг Y увеличится на 20,27 млн руб., а при увеличении объема основных производственных средств (X_2) на 1 млн руб. (при неизменном X_1) средний объем продаж услуг увеличится на 1,04 млн руб.

2) Найти оценку дисперсии модели, стандартную ошибку регрессии и стандартные ошибки коэффициентов регрессии. Построить 90%-й доверительный интервал для дисперсии модели.

Для нахождения оценки дисперсии модели необходимо найти сумму квадратов остатков модели

$RSS = \sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2 = \sum_{i=1}^{n} e_i^2$, где $\hat{y}_i$ – расчетные значения зависимой переменной. Используя результаты промежуточных вычислений, помещенные в приводимую далее табл. 7, получим

$$RSS = \sum_{i=1}^{n} e_i^2 = (-0,72)^2 + (-0,74)^2 + \cdots + 0,84^2 + 0,73^2 = 31,65.$$

Учитывая, что число параметров модели (b_0, b_1, b_2) равно $m = 3$, а объем выборки $n = 12$, найдем оценку дисперсии модели

$$\hat{\sigma}^2 = \frac{RSS}{n-m} = \frac{31,65}{12-3} = 3,52.$$

Стандартная ошибка модели будет равна $\hat{\sigma} = \sqrt{\hat{\sigma}^2} =$ $= \sqrt{3,52} = 1,88$.

Таблица 7

Год	y_i	$\hat{y}_i = -42,23 + 20,27X_{i1} + 1,04X_{i2}$	$e_i = y_i - \hat{y}_i$
2003	11	11,72	– 0,72
2004	13	13,74	– 0,74
2005	16	16,81	– 0,81
2006	24	19,87	4,13
2007	24	23,96	0,04
2008	26	26,04	– 0,04
2009	27	28,06	– 1,06
2010	28	31,12	– 3,12
2011	29	29,10	– 0,10
2012	33	32,16	0,84
2013	33	32,16	0,84
2014	36	35,27	0,73

Оценка ковариационной матрицы будет иметь вид

$$\hat{\mathbf{D}}(\hat{\boldsymbol{\beta}}) = 3{,}52 \cdot \begin{pmatrix} 48{,}361 & -28{,}673 & 1{,}463 \\ -28{,}673 & 17{,}699 & -0{,}973 \\ 1{,}463 & -0{,}973 & 0{,}061 \end{pmatrix} =$$

$$= \begin{pmatrix} 170{,}23 & -100{,}93 & 5{,}15 \\ -100{,}93 & 62{,}30 & -3{,}42 \\ 5{,}15 & -3{,}42 & 0{,}21 \end{pmatrix}.$$

Стандартные ошибки коэффициентов регрессии имеют следующие значения:

$$S_{b_0} = \sqrt{170{,}23} = 13{,}05,\ S_{b_1} = \sqrt{62{,}30} = 7{,}89,\ S_{b_2} = \sqrt{0{,}21} = 0{,}46.$$

3) На уровне значимости 0,05 проверить гипотезы о значимости коэффициентов регрессии.

Рассмотрим последовательно решение задачи для каждого из трех коэффициентов.

1. $H_0 : b_0 = 0$ (коэффициент регрессии незначим),

$H_1 : b_0 \neq 0$ (коэффициент регрессии значим).

Расчетное значение t-статистики равно:

$$K_{выч} = \frac{\hat{b}_0}{S_{b_0}} = \frac{-42{,}23}{13{,}05} = -3{,}24.$$

Критическое значение статистики определяем по табл. А3:

$$K_{таб} = t_{1-0{,}05}[St(12-3)] = t_{0{,}95}[St(9)] =$$
$$= x_{\frac{1+0{,}95}{2}}[St(9)] = x_{0{,}975}[St(9)] = 2{,}26.$$

Поскольку $|K_{выч}| > K_{таб}$, гипотезу о значимости параметра регрессии b_0 следует принять.

2. $H_0 : b_1 = 0$ (коэффициент регрессии незначим),

$H_1 : b_1 \neq 0$ (коэффициент регрессии значим).

Расчетное значение t-статистики равно:

$K_{выч} = \frac{\hat{b}_1}{S_{b_1}} = \frac{20,27}{7,89} = 2,57$. Критическое значение статистики то же, что и в предыдущем случае. Поскольку $|K_{выч}| > K_{таб}$, параметр модели b_1 значим на уровне значимости 0,05.

3. $H_0 : b_2 = 0$ (коэффициент регрессии незначим),

$H_1 : b_2 \neq 0$ (коэффициент регрессии значим).

Расчетное значение t-статистики равно: $K_{выч} = \frac{\hat{b}_2}{S_{b_2}} = \frac{1,04}{0,46} = 2,27$. Поскольку $|K_{выч}| > K_{таб}$, параметр модели b_2 значим на уровне значимости 0,05.

4) Найти частные коэффициенты эластичности и стандартизованные коэффициенты регрессии. Сделать вывод о силе влияния факторов X_1 и X_2 на результирующую переменную Y.

Проведем вспомогательные вычисления:

$$\bar{X}_1 = \frac{1}{n}\sum_{i=1}^{n} x_{i1} = \frac{30}{12} = 2,5, \quad X_2 = \frac{1}{n}\sum_{i=1}^{n} x_{i2} = \frac{192}{12} = 16,$$

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{300}{12} = 25, \quad \overline{X_1^2} = \frac{1}{n}\sum_{i=1}^{n} x_{i1}^2 = \frac{75,48}{12} = 6,29,$$

$$\overline{X_2^2} = \frac{1}{n}\sum_{i=1}^{n} x_{i2}^2 = \frac{3212}{12} = 267,67,$$

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} y_i^2 = \frac{8202}{12} = 683,5,$$

$$S_1^2 = \overline{X_1^2} - (\bar{X}_1)^2 = 6,29 - (2,5)^2 = 0,04,$$

$$S_2^2 = \overline{X_2^2} - (\bar{X}_2)^2 = 267,67 - (16)^2 = 11,67,$$

$$S_Y^2 = \overline{Y_1^2} - (\bar{Y}_1)^2 = 683,5 - (25)^2 = 58,5.$$

Найдем частные коэффициенты эластичности:

$$\overline{Э}_1 = \hat{b}_1 \cdot \frac{\overline{X}_1}{\overline{Y}} = 20,27 \cdot \frac{2,5}{25} = 2,03, \quad \overline{Э}_2 = \hat{b}_2 \cdot \frac{\overline{X}_2}{\overline{Y}} = 1,04 \cdot \frac{16}{25} = 0,67.$$

Вывод. При увеличении только количества работающих (X_1) на 1% и неизменном X_2 средний объем продаж услуг (Y) увеличится на 2,03%, а при увеличении объема основных производственных средств (X_2) на 1% и неизменном X_1 средний объем продаж услуг (Y) увеличится на 0,67 %.

Определим стандартизованные коэффициенты регрессии:

$$\hat{\beta}_1 = \hat{b}_1 \cdot \frac{S_{X_1}}{S_Y} = 20,27 \cdot \frac{\sqrt{0,04}}{\sqrt{58,5}} = 0,53, \ \hat{\beta}_2 = \hat{b}_2 \cdot \frac{S_{X_2}}{S_Y} = 1,04 \cdot \frac{\sqrt{11,67}}{\sqrt{58,5}} = 0,46.$$

Вывод. При увеличении только количества работающих (X_1) на величину S_{X_1} (при неизменном X_2) средний объем продаж услуг (Y) увеличится на 0,53 величин S_Y, а при увеличении объема основных производственных средств (X_2) на величину S_{X_2} (при неизменном X_1) средний объем продаж услуг (Y) увеличится на 0,46 величин S_Y.

На основании рассчитанных коэффициентов можно сделать вывод о том, что относительный эффект от влияния количества рабочих X_1 на объем продаж Y выше, чем от влияния объема основных производственных средств X_2.

5) Построить 98%-е доверительные интервалы для коэффициентов регрессии.

По доверительной вероятности $\gamma = 0,98$ вычислим квантиль:

$$t_\gamma[St(n-m)] = t_{0,98}[St(12-3)] = x_{\frac{1+0,98}{2}}[St(9)] = x_{0,99}[St(9)] = 2,82.$$

Доверительный интервал для параметра b_0 примет вид

$$-42,23-2,82\cdot 13,05<b_0<-42,23+2,82\cdot 13,05,\ \ 79,03<b_0<5,43.$$

Доверительный интервал для параметра b_1 определяется неравенствами:

$$20,27-2,82\cdot 7,89<b_1<20,27+2,82\cdot 7,89,\ \ -1,89<b_1<42,52.$$

Доверительный интервал для параметра b_2 имеет вид

$$1,04-2,82\cdot 0,46<b_2<1,04+2,82\cdot 0,46,\ \ -0,26<b_2<2,34.$$

Вывод. С вероятностью 0,98 можно утверждать, что значение параметра b_0 находится в пределах от 79,03 до 5,43; параметра b_1 – в пределах от –1,89 до 42,52; а параметра b_2 – в пределах от –0,26 до 2,34.

6) Определить коэффициент детерминации и скорректированный коэффициент детерминации. Сделать выводы.

Для нахождения коэффициента детерминации воспользуемся формулой:

$$R^2=1-\frac{RSS}{n\cdot S_Y^2}=1-\frac{31,65}{12\cdot 58,5}=0,95.$$

Вывод. Доля дисперсии зависимой переменной Y, объясненная построенной линейной моделью, составляет 95%, остальные 5% дисперсии объяснить не удалось.

Скорректированный коэффициент детерминации:

$$R_{adj}^2=1-\frac{n-1}{n-m}\cdot(1-R^2)=1-\frac{12-1}{12-3}\cdot(1-0,95)=0,94.$$

Для интерпретации скорректированного коэффициента детерминации оценим модель без включения в нее независимой переменной X_1, т.е. рассмотрим зависимость объема продаж услуг Y (в млн руб.) некоторого сервисного предприятия от объема основных производственных средств X_2 (в млн руб.).

Поэтому выборочное уравнение регрессии, характеризующее зависимость объема продаж услуг от объема основных производственных средств, имеет вид

$$\widehat{Y_X} = -9,4 + 2,15X_2.$$

Коэффициент детерминации равен: $R^2 = 0,92$. Скорректированный коэффициент детерминации равен: $R^2_{adj} = 0,91$.

Вывод. Поскольку скорректированный коэффициент детерминации для парной модели меньше, чем для множественной, то лучшей признается множественная линейная регрессионная модель с включенными в нее переменными X_1 и X_2.

7) На уровне значимости 0,05 проверить гипотезу о значимости модели.

Проверим гипотезу о значимости модели $Y = b_0 + b_1X_1 + b_2X_2 + \varepsilon$.

$H_0 : R^2_{\Gamma} = 0$ (модель незначима).

$H_1 : R^2_{\Gamma} \neq 0$ (модель значима).

Вычислим значение F-статистики:

$$K_{выч} = \frac{R^2}{1-R^2} \cdot \frac{n-m}{k} = \frac{0,95}{1-0,95} \cdot \frac{12-3}{2} = 85,5.$$

Критическое значение статистики находим по табл. А6:

$$K_{таб} = x_{1-\alpha}[F(k,\ n-m)] = x_{1-0,05}[F(2,12-3)] = x_{0,95}[F(2,9)] = 4,26.$$

Так как $K_{выч} = 85,5 > K_{таб} = 4,26$, гипотезу H_0 о незначимости модели следует отвергнуть. На уровне значимости 0,05 можно утверждать, что модель значима.

Тесты для проверки знаний

1. Метод наименьших квадратов используется для оценивания

1) параметров линейной регрессии
2) средней ошибки аппроксимации
3) величины коэффициента детерминации
4) величины коэффициента корреляции
5) среди приведенных нет правильного ответа

2. Коэффициент детерминации рассчитывается для оценки качества

1) мультиколлинеарных факторов
2) факторов, не включенных в уравнение регрессии
3) параметров уравнения регрессии
4) подбора уравнения регрессии
5) среди приведенных ответов нет правильного

3. Критические значения критерия Стьюдента определяются по…

1) уровню значимости и одной степени свободы
2) двум степеням свободы
3) уровню незначимости
4) трем и более степеням свободы
5) среди приведенных ответов нет правильного

4. Какая гипотеза для множественной линейной регрессионной модели проверяется в тесте Стьюдента?

1) $H_0 : b_i = 0$
2) $H_0 : R_{Г}^2 = 0$
3) $H_0 : b_i = b_{i+1}$
4) $H_0 : D(\varepsilon_i) = \sigma^2$
5) среди приведенных ответов нет правильного

6. Какая гипотеза для множественной линейной регрессионной модели проверяется в тесте Фишера?

1) $H_0 : b_i = 0$
2) $H_0 : R_Г^2 = 0$
3) $H_0 : b_i = 8$
4) $H_0 : D(\varepsilon_i) = \sigma^2$
5) среди приведенных ответов нет правильного

7. Какие статистические гипотезы выдвигаются при проверке статистической значимости построенной модели? (возможно несколько вариантов ответов)

1) зависимая о статистической зависимости
2) нулевая – о статистической незначимости
3) независимая о статистической независимости
4) альтернативная о статистической значимости
5) среди приведенных ответов нет правильного

8. При выполнении предпосылок МНК оценки параметров регрессии обладают свойствами (возможно несколько вариантов ответов)

1) состоятельность
2) несмещенность
3) достоверность
4) эффективность
5) среди приведенных ответов нет правильного

9. Оценка значимости уравнения регрессии осуществляется с помощью

1) t-критерия Стьюдента
2) коэффициента корреляции
3) коэффициента детерминации
4) F-критерия Фишера
5) среди приведенных ответов нет правильного

10. При обсуждении существенности параметра регрессии рассматривается нулевая статистическая гипотеза о (об)_________ этого параметра

1) положительном значении
2) отрицательном значении
3) отличии от нуля
4) равенстве нулю
5) среди приведенных ответов нет правильного

11. Для проверки значимости коэффициента детерминации используется статистика с распределением

1) Стьюдента
2) Гаусса
3) Фишера
4) Хи-квадрат
5) Среди приведенных ответов нет правильного

12. Доля остаточной дисперсии зависимой переменной в ее общей дисперсии составила 30%, следовательно, величина

1) Коэффициента детерминации равна 0,7
2) Коэффициента детерминации равна 0,3
3) Коэффициента регрессии равна 0,3
4) Множественный коэффициент регрессии равен 0,7
5) Среди приведенных ответов нет правильного

13. Критерий Фишера в экономических моделях служит

1) Для оценки параметров регрессии
2) Показателем линейной связи между переменными
3) Показателем преимущества выбранной модели перед другими
4) Для проверки статистической значимости уравнения регрессии
5) Среди приведенных ответов нет правильного

14. Коэффициент регрессии считается статистически значимым, если справедливо следующее утверждение

1) Доверительный интервал для этого коэффициента не содержит 0

2) Фактическое значение статистики Стьюдента для этого коэффициента по модулю меньше критического (табличного)

3) Вероятность того, что этот коэффициент равен 0, близка к 1

4) Доверительный интервал для этого коэффициента содержит 1

5) Среди приведенных ответов нет правильного

15. Четырехфакторное уравнение регрессии оценено по 22 наблюдениям. Сколько степеней свободы в этом случае имеет t-статистика?

1) 17

2) 4

3) 22

4) 20

5) Среди приведенных ответов нет правильного

16. Укажите сумму номеров неверных утверждений

1) Значение коэффициента детерминации не может увеличиться при включении в множественную линейную регрессионную модель переменной, не оказывающей влияние на зависимую переменную

2) Корреляционная форма зависимости является частным случаем функциональной зависимости

3) В соответствии с теоремой Гаусса- Маркова точечная оценка функции регрессии является эффективной оценкой

4) Тест, проверяющий качество множественной регрессионной модели в целом предполагает использование распределения Стьюдента

5) Значение RSS увеличивается при включении в множественную регрессионную модель еще одного фактора

17. Фиктивными переменными в уравнении множественной регрессии являются

1) Комбинация исключенных в уравнении регрессии факторов, повышающая адекватность модели

2) Дополнительные количественные переменные, улучшающие решение

3) Переменные, представляющие простейшие функции от уже включенных в модель переменных

4) Качественные переменные, преобразованные в количественные

5) Среди приведенных ответов нет правильного

18. В регрессию $\hat{Y}_X = 1,5 + 0,3X_1 + X_2 + 0,5X_3$ добавили переменную X_4. Переменная оказалась незначимой. Как изменился коэффициент детерминации (R^2)?

1) Коэффициент детерминации увеличился
2) Коэффициент детерминации уменьшился
3) Коэффициент детерминации не изменился
4) Среди приведенных ответов нет правильного

19. Построена модель $Y_X = 1 - \underset{(-2,4)}{0,4} \cdot X_1 + \underset{(3,0)}{1,5} \cdot X_2 - \underset{(-3,75)}{0,75} \cdot X_3$, $R^2 = 0,86$. Определить, что находится в скобках.

1) P-значение
2) Стандартные ошибки коэффициентов регрессии
3) Коэффициенты вздутия дисперсий
4) t – статистики
5) Среди приведенных ответов нет правильного

20. Методом наименьших квадратов по 16 наблюдениям получена оценка функции регрессии $Y_X = 1 - 0,4X_1 + X_2 + 0,75X_3$, $R^2 = 0,8$. Проверяется гипотеза о существенности (значимости)

множественной регрессии на уровне значимости $\alpha = 0{,}05$. Значение F-статистики для проверки этой гипотезы равно

1) 16
2) 20
3) 12
4) 18
5) 24

21. Исследовалась зависимость заработной платы от стажа работы (по 30 наблюдениям). В модель введена фиктивная переменная, принимающая значение 1, если работник – женщина, и значение 0, если работник – мужчина. Получена оценка: $\hat{Y} = 8000 + 1500 * X - 1500 * D$, $R^2 = 0.35$. Проверяется, является ли модель значимой (α = 5 %). Разница между вычисленным и табличным значением равна:

1) 1,22
2) 3,92
3) -1,64
4) 4,62
5) Среди приведенных ответов нет правильного

22. При исследовании зависимости оборота розничной торговли (Y, млрд руб.) от трех факторов:

X_1 – денежные доходы населения, млрд руб.;

X_2 – численность безработных, млн чел.;

X_3 – официальный курс рубля по отношению к доллару США, получена следующая модель: $\hat{Y}_X = 64{,}1 - 0{,}4X_1 - 3{,}2X_2 + 2{,}6X_3$. Как интерпретируется коэффициент при факторном признаке X_2?

1) при увеличении численности безработных на 1 % оборот розничной торговли в среднем будет уменьшаться на 3,2 %

2) при увеличении численности безработных на 1 млн чел. оборот розничной торговли в среднем будет уменьшаться на 3,2 млрд руб.

3) при увеличении только численности безработных на 1 млн чел. оборот розничной торговли в среднем будет увеличиваться на 3,2 млрд руб.

4) при увеличении только численности безработных на 1 млн чел. оборот розничной торговли в среднем будет уменьшаться на 3,2 млрд руб.

5) среди приведенных ответов нет правильного

23. Оцененное уравнение зависимости переменной Y от переменной X имеет вид $\widehat{Y} = 3.4X - 5.1$. Тогда коэффициент эластичности переменной Y по переменной X при $X = 2$ равен

1) 4

2) 6.8

3) 3.4

4) –4.2

5) Среди приведенных ответов нет правильного

Глава 5. РАЗЛИЧНЫЕ АСПЕКТЫ МНОЖЕСТВЕННОЙ РЕГРЕССИИ

Возможные отклонения от предположений классической множественной линейной регрессионной модели: автокорреляция, гетероскедастичность различных наблюдений; закон распределения отличный от нормального. Неформальные методы их обнаружения, возможные экономические причины возникновения. Природа проблемы гетероскедастичности. Виды гетероскедастичности. Последствия гетероскедастичности. Способы выявления гетероскедастичности. Методы преодоления гетероскедастичности. Обнаружение автокорреляции с помощью различных критериев.

Проверка основных предположений регрессионного анализа
При работе с реальными статистическими данными важно проверить, действительно ли желаемые ограничения (условия, определяющие классическую линейную регрессионную модель) имеют место. Основу тестов, ориентированных на проверку классических предположений, составляют остатки: $e_i = y_i - \hat{y}_i$, $i = \overline{1,n}$.

Проверка нормальности ошибок модели

Критерий Жарка – Бера (Jarque – Bera)	
Назначение	Используется для проверки гипотезы о нормальности регрессионных остатков
Данные	Модель наблюдений: $Y_i = b_0 + b_1 X_{1i} + \ldots + b_k X_{ki} + \varepsilon_i$, где ε_i – случайные величины, имеющие нормальное распределение, $e_i = y_i - \hat{y}_i$, – регрессионные остатки, $i = \overline{1,n}$.
Допущения	Критерий применим лишь при достаточно большом объеме выборки n.

Алгоритм проверки гипотезы	
Шаг 1. Выдвинуть гипотезы основную и альтернативную на уровне значимости α	
Основная (нулевая) гипотеза	H_0 : регрессионные остатки имеют нормальное распределение
Шаг 2. Найти критическое (табличное) значение статистики критерия	
Критическое значение критерия	$K_{таб} = x_{1-\alpha}\left[\chi^2(2)\right]$ квантиль уровня $1-\alpha$ распределения Хи-квадрат с 2 степенями свободы находится по табл. А4
Шаг 3. Найти вычисленное (фактическое) значение статистики критерия	
Статистика критерия (фактическое значение)	$JB = n\left[\frac{Sk^2}{6} + \frac{Ku^2}{24}\right]$
Шаг 4. Сравнить вычисленное значение с критическим	
Область принятия нулевой гипотезы	Гипотеза о нормальном распределении остатков не отвергаем, если $K_{выч} < K_{таб}$

Проверка адекватности линейной модели анализируемым данным

Критерий Рамсея ($RESET$ **-тест)**	
Назначение	Используется для проверки гипотезы о линейности модели
Данные	Имеются результаты n наблюдений $(x_1, y_1),\ldots,(x_n, y_n)$ переменных X и Y
Допущения	Модель наблюдений: $Y_i = b_0 + b_1 X_{1i} + \ldots + b_k X_{ki} + \varepsilon_i$, где ε_i – случайные величины, имеющие нормальное распределение
Алгоритм проверки гипотезы	
Шаг 1. Выдвинуть гипотезы основную и альтернативную на уровне значимости α	

Основная (нулевая) гипотеза	H_0: модель правильно специфицирована
Конкурирующая (альтернативная) гипотеза	H_1: модель неверно специфицирована
Шаг 2. Найти вычисленное (фактическое) значение статистики критерия	
Статистика критерия (фактическое значение)	$K_{выч} = n \cdot R^2$, где коэффициент детерминации R^2 вычисляется по вспомогательной регрессионной модели $e_i = A_0 + A_1\left[\hat{y}_i\right]^2 + \ldots + A_{p-1}\left[\hat{y}_i\right]^p + u_i$, а $e_i, i = \overline{1,n}$ – регрессионные остатки исходной модели $Y_i = b_0 + b_1 X_{1i} + \ldots + b_k X_{ki} + \varepsilon_i$
Шаг 3. Найти критическое (табличное) значение статистики критерия	
Критическое значение критерия	$K_{таб} = x_{1-\alpha}\left[\chi^2\left(p-1\right)\right]$ – квантиль уровня $1-\alpha$ распределения Хи-квадрат с $p-1$ степенями свободы) находится по табл. А4
Шаг 4. Сравнить вычисленное значение с критическим	
Область принятия нулевой гипотезы	Гипотезу о правильной спецификации не отвергаем, если $K_{выч} < K_{таб}$

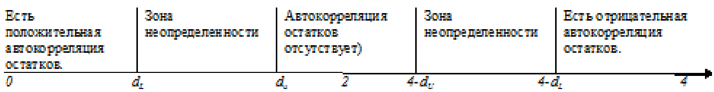
Проверка остатков на автокорреляцию

Одно из предположений регрессионного анализа заключается в некоррелированности результатов различных наблюдений: $\mathbf{M}\left[\varepsilon_i \varepsilon_j\right] = 0$ при $i \neq j$. Если данное условие не выполняется, то говорят о том, что случайный член подвержен *автокорреляции* *(autocorrelation, serial correlation)*	
Модель авторегрессии первого порядка	$\varepsilon_i = \rho\varepsilon_{i-1} + u_i$, $i = 1,\ldots,n$, где $\left|\rho\right| < 1$, а u_i, $i = 1,2,\ldots,n$ – независимые в совокупности случайные величины, имеющие одно и то же нормальное распределение $N\left(0,\sigma^2\right)$

Коэффициент автокорреляции первого порядка	$\rho(1)=\dfrac{\sum_{i=2}^{n} e_i e_{i-1}}{\sum_{i=1}^{n} e_i^2}$ если $\rho > 0$, то автокорреляция *положительная* если $\rho < 0$, то автокорреляция *отрицательная* если $\rho = 0$, то автокорреляция *отсутствует*

Проверка гипотезы об автокорреляции остатков

Критерий Дарбина-Уотсона	
Назначение	Используется для проверки гипотезы об автокорреляции регрессионных остатков
Данные	Модель наблюдений: $Y_i = b_0 + b_1 X_{1i} + \ldots + b_k X_{ki} + \varepsilon_i$, где ε_i – случайные величины, имеющие нормальное распределение, $e_i = y_i - \hat{y}_i$, – регрессионные остатки, $i = \overline{1,n}$.
Допущения	Критерий применим лишь при достаточно большом объеме выборки n.
Алгоритм проверки гипотезы	
Шаг 1. Выдвинуть гипотезы основную и альтернативную на уровне значимости α	
Основная (нулевая) гипотеза	$H_0 : \rho = 0$ автокорреляция первого порядка отсутствует
Конкурирующая (альтернативная) гипотеза	$H_1 : \rho \neq 0$ есть автокорреляция первого порядка
Шаг 2. Найти критическое (табличное) значение статистики критерия	
Критическое значение критерия	С помощью табл. А8 находят критические значения критерия Дарбина – Уотсона: $d_U = d_U(\alpha, k, n)$ и $d_L = d_L(\alpha, k, n)$, определяемые по значениям входных параметров α, k, n.

<table>
<tr><td colspan="2">Шаг 3. Найти вычисленное (фактическое) значение статистики критерия</td></tr>
<tr><td>Статистика критерия (фактическое значение)</td><td>$$DW = \frac{\sum_{i=2}^{n}\left(e_i - e_{i-1}\right)^2}{\sum_{i=1}^{n} e_i^2}$$ Известно, что с ростом объема выборки $DW \xrightarrow{P} 2 - 2\rho$.
Если автокорреляция остатков отсутствует, то $DW \approx 0$. При положительной автокорреляции имеем $0 < DW \le 2$, при отрицательной – $2 < DW \le 4$.</td></tr>
<tr><td colspan="2">Шаг 4. Сравнить вычисленное значение с критическим</td></tr>
<tr><td>Область принятия нулевой гипотезы</td><td>Гипотеза об отсутствии автокорреляции первого порядка не отклоняется, если выполняется неравенство $d_U\left(\frac{\alpha}{2}, k, n\right) < DW_{выч} < 4 - d_U\left(\frac{\alpha}{2}, k, n\right)$; гипотеза отклоняется в пользу альтернативной, если
$DW_{выч} > 4 - d_L\left(\frac{\alpha}{2}, k, n\right)$ или $DW_{выч} < d_L\left(\frac{\alpha}{2}, k, n\right)$.


Если фактически наблюдаемое значение DW попадает в зону неопределенности, то на практике предполагают существование
автокорреляции остатков и отклоняют гипотезу H_0.</td></tr>
<tr><td colspan="2">Недостатком критерия является наличие области неопределенности критерия, также методика расчета направлена только на выявление автокорреляции остатков первого порядка, а также то, что критические значения d-статистики определены для объемов выборки не менее 15</td></tr>
</table>

Пример. На основе помесячных данных о количестве разводов (тыс.) в регионе была построена модель временного ряда: $\hat{Y}_t = 2,5 + 0,3t$. В таблице указаны остатки множественной линейной регрессионной модели:

Месяц	Остатки	Месяц	Остатки	Месяц	Остатки
Январь	-1,0	Май	-2,0	Сентябрь	2,5
Февраль	2,0	Июнь	-1,1	Октябрь	1,0
Март	-0,5	Июль	3,0	Ноябрь	-3,0
Апрель	0,5	Август	1,0	Декабрь	2,0

Вычислить коэффициент автокорреляции и проверить гипотезу об автокорреляции ошибок с помощью критерия Дарбин – Уотсона на уровне значимости 0,05.

Решение. Для нахождения коэффициента автокорреляции воспользуемся следующей формулой:

$$\rho(1) = \frac{\sum_{i=2}^{n} e_i e_{i-1}}{\sum_{i=1}^{n} e_i^2} = \frac{2\cdot(-1) + (-0,5)\cdot 2 + \cdots + 2\cdot(-3)}{(-1)^2 + 2^2 + (-0,5)^2 + \cdots 2^2} = -0,16.$$

Такое значение коэффициента автокорреляции говорит о достаточно слабой зависимости между остатками e_i и e_{i-1} .

Проверим нулевую гипотезу об отсутствии автокорреляции первого порядка, т. е. $H_0 : \rho = 0$.

Для этого вычислим статистику Дарбина – Уотсона:

$$DW_{выч} = \frac{\sum_{i=2}^{n} \left(e_i - e_{i-1}\right)^2}{\sum_{i=1}^{n} e_i^2} = \frac{(2+1)^2 + (-0,5-2)^2 + \cdots + (2+3)^2}{(-1)^2 + 2^2 + (-0,5)^2 + \cdots 2^2} = 2,16.$$

С помощью табл. А8 найдем критические значения критерия Дарбина – Уотсона: $d_U = d_U(\alpha, k, n) = 1{,}33$ и $d_L = d_L(\alpha, k, n) = 0{,}97$, где $\alpha = 0{,}05$, $k = 1$ (одна объясняющая переменная) и $n = 12$. Поскольку $1{,}33 < DW_{выч} < 4 - 1{,}33 = 2{,}67$, то гипотеза об отсутствии автокорреляции первого порядка принимается.

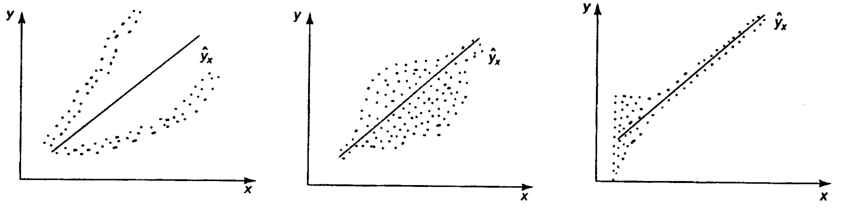
Проверка гипотезы об автокорреляции остатков

Критерий Бреуша – Годфри	
Назначение	Используется для проверки гипотезы об автокорреляции регрессионных остатков
Данные	Модель наблюдений: $Y_i = b_0 + b_1 X_{1i} + \ldots + b_k X_{ki} + \varepsilon_i$, где ε_i – случайные величины, имеющие нормальное распределение, $e_i = y_i - \hat{y}_i$, – регрессионные остатки, $i = \overline{1,n}$.
Допущения	Критерий применим лишь при достаточно большом объеме выборки n.
Алгоритм проверки гипотезы	
Шаг 1. Выдвинуть гипотезы основную и альтернативную на уровне значимости α	
Основная (нулевая) гипотеза	$H_1 : \rho_1 = \rho_2 = ... = \rho_k = 0,$ автокорреляция отсутствует в авторегрессии $e_t = \rho_1 e_{t-1} + \rho_2 e_{t-2} + ... + \rho_k e_{t-k}$,
Конкурирующая (альтернативная) гипотеза	$H_1 : хотя\,бы\,один\,\rho_i \neq 0,$ т.е. есть автокорреляция
Шаг 2. Найти критическое (табличное) значение статистики критерия	
Критическое значение критерия	$K_{табл} = x_{1-\alpha}[\chi^2(k)]$ – квантиль уровня $1-\alpha$ распределения Хи-квадрат с k степенями свободы, находится по табл. А4

Шаг 3. Найти вычисленное (фактическое) значение статистики критерия	
Статистика критерия (фактическое значение)	$K_{выч} = (n-k)\cdot R^2$, где n – число наблюдений для исходной модели, R^2 – коэффициент детерминации вспомогательной модели случайных отклонений от всех экзогенных переменных, а также лаговых переменных отклонений до порядка k включительно: $e_t = \delta_0 + \delta_1 x_{1t} + ... + \delta_p x_{pt} + \alpha_1 e_{t-1} + \alpha_2 e_{t-2}$
Шаг 4. Сравнить вычисленное значение с критическим	
Область принятия нулевой гипотезы	Если $K_{выч} > K_{таб}$, то H_0 отклоняется при этом уровне значимости α.

Проверка гетероскедастичности случайных ошибок

Гомоскедастичность	Дисперсия случайной ошибки постоянна для всех наблюдений $D(\varepsilon_i) = M(\varepsilon_i^2) = \sigma^2,\ i = \overline{1,n}.$
Гетероскедастичность *(heteroscedasticity)*	Дисперсия случайной ошибки постоянна для всех наблюдений $D(\varepsilon_i) = M(\varepsilon_i^2) = \sigma_i^2,\ i = \overline{1,n},$ т.е. «модель с неравноточными наблюдениями». В случае гетероскедастичной модели МНК-оценки коэффициентов регрессии оказываются хотя и несмещенными, но неэффективными
Графический способ обнаружения гетероскедастичности	



дисперсия остатков растет по мере увеличения X	дисперсия остатков достигает максимальной величины при средних значениях переменной X и уменьшается при минимальных и максимальных значениях X	дисперсия остатков однородна по мере увеличения значений X.
Нулевая гипотеза для проверки гетероскедастичности	$H_0 : \mathrm{D}[\varepsilon_1] = \mathrm{D}[\varepsilon_2] = \ldots = \mathrm{D}[\varepsilon_n] = \sigma^2$ – гипотеза о равноточности наблюдений (или о гомоскедастичности модели)	
Критерии для проверки	тест Уайта (White), тест ранговой корреляции Спирмена, тесты Парка (Park), Глейзера (Glejser), Голдфелда – Куандта (Goldfeld – Quandt), Бреуша – Пагана (Breusch – Pagan)	
Коррекция гетероскедастичности линейной регрессионной модели		
После установления гетероскедастичности модель следует попробовать преобразовать с целью устранения этой проблемы. Для этого используется *взвешенный МНК*.		

Проверка гипотезы о гетероскедастичности

Критерий Спирмена	
Назначение	Критерий предназначен для обнаружения наличия зависимости значения дисперсии случайной ошибки от значения определенной независимой переменной с использованием рангового подхода.
Данные	Модель наблюдений: $Y_i = b_0 + b_1 X_{1i} + \ldots + b_k X_{ki} + \varepsilon_i$, где ε_i – случайные величины, имеющие нормальное распределение, $e_i = y_i - \hat{y}_i$, – регрессионные остатки, $i = \overline{1, n}$.
Допущения	Критерий применим лишь при достаточно большом объеме выборки n.

Алгоритм проверки гипотезы	
Шаг 1. Выдвинуть гипотезы основную и альтернативную на уровне значимости α	
Основная (нулевая) гипотеза	$H_0: \mathrm{D}[\varepsilon_1] = \mathrm{D}[\varepsilon_2] = \ldots = \mathrm{D}[\varepsilon_n] = \sigma^2$ – гипотеза о равноточности наблюдений (или о гомоскедастичности модели)
Конкурирующая (альтернативная) гипотеза	H_1: гетероскедастичность модели
Шаг 2. Найти критическое (табличное) значение статистики критерия	
Критическое значение критерия	$K_{таб} = t_{1-\alpha}[N]$ – двусторонний квантиль уровня $1-\alpha$ нормального распределения, находится по табл. А2
Шаг 3. Найти вычисленное (фактическое) значение статистики критерия	
Статистика критерия (фактическое значение)	$K_{выч} = \lvert\rho_s\rvert\sqrt{n-1}$, где $\rho_s = 1 - \frac{6}{n^3 - n} \cdot \sum_{i=1}^{n} \left[R(x_{ij}) - R(\lvert e_i \rvert) \right]^2$ где $R(z_i)$ – ранг z_i в выборке $z_1, z_2, \ldots, z_n$. *Ранг элемента выборки* – это порядковый номер, под которым это значение располагается в вариационном ряду. Если имеется несколько равных по значению элементов выборки, то их ранг равен среднему значению номеров.
Шаг 4. Сравнить вычисленное значение с критическим	
Область принятия нулевой гипотезы	Если $\lvert K_{выч} \rvert \le K_{таб}$, то H_0 об отсутствии гетероскедастичности не отвергаем при этом уровне значимости α.

Мультиколлинеарность

Мультиколлинеарность	линейная взаимосвязь двух или нескольких объясняющих переменных
Совершенная мультиколлинеарность	если объясняющие переменные связаны строгой функциональной зависимостью, не позволяет однозначно определить коэффициенты регрессии и разделить вклады объясняющих переменных в их влиянии на зависимую переменную Y
Несовершенная мультиколлинеарность	сильная корреляционная зависимостью между объясняющими переменными, характеризуется высоким коэффициентом корреляции между соответствующими объясняющими переменными.
Последствия мультиколлинеарности	
1.	Большие дисперсии (стандартные ошибки) оценок. Это затрудняет нахождение истинных значений определяемых величин и расширяет интервальные оценки, ухудшая их точность
2.	Уменьшаются t-статистики коэффициентов, что может привести к неоправданному выводу о существенности влияния соответствующей объясняющей переменной на зависимую переменную
3.	Оценки коэффициентов по МНК и их стандартные ошибки становятся очень чувствительными к малейшим изменениям данных, т. е. они становятся неустойчивыми.
4.	Затрудняется определение вклада каждой из объясняющих переменных в объясняемую уравнением регрессии дисперсию зависимой переменной
5.	Возможно получение неверного знака у коэффициента регрессии.
Определение мультиколлинеарности	
1.	Коэффициент детерминации R^2 достаточно высок, но некоторые из коэффициентов регрессии статистически незначимы, т.е. они имеют низкие t-статистики.

2.	Парная корреляция между малозначимыми объясняющими переменными достаточно высока. Однако данный признак будет надежным лишь в случае двух объясняющих переменных. При большем их количестве целесообразнее использование частных коэффициентов корреляции
3.	Высокие частные коэффициенты корреляции. Частные коэффициенты корреляции определяют силу линейной зависимости между двумя переменными без учета влияния на них других переменных
4.	Сильная вспомогательная (дополнительная) регрессия. Мультиколлинеарность может иметь место вследствие того, что какая-либо из объясняющих переменных является линейной (или близкой к линейной) комбинацией других объясняющих переменных. Для данного анализа строятся уравнения регрессии каждой из объясняющих переменных X_j, $j=1,\ldots,m$ на оставшиеся объясняющие переменные вспомогательной регрессии. Вычисляются соответствующие коэффициенты детерминации R_j^2 и рассчитывается их статистическая значимость на основе F-статистики: $F_j = \frac{R_j^2}{1-R_j^2} \cdot \frac{n-m}{m-1}$. Здесь n − число наблюдений, m − число объясняющих переменных в первоначальном уравнении регрессии. Статистика F имеет распределение Фишера с $v_1 = m-1$ и $v_2 = n-m$ степенями свободы. Если коэффициент R_j^2 статистически незначим, то X_j не является линейной комбинацией других переменных и ее можно оставить в уравнении регрессии. В противном случае есть основания считать, что X_j существенно зависит от других объясняющих переменных и имеет место мультиколлинеарность.
Методы устранения мультиколлинеарности	
1	Исключение переменной(ых) из модели
2	Получение дополнительных данных или новой выборки

3	Изменение спецификации модели
4	Использование предварительной информации о некоторых параметрах
5	Преобразование переменных

Пример. Исследуется регрессионная зависимость выпуска продукции деревообрабатывающей промышленности на душу населения Y от валового внутреннего продукта на душу населения X в том же году для 18 стран по исходным данным, приведенным в табл. 8.

Таблица 8

Страна	***X***	***Y***	**Страна**	***X***	***Y***
1	3	18	10	24	100
2	6	27	11	25	63
3	7	18	12	26	130
4	9	45	13	27	135
5	13	55	14	28	60
6	15	68	15	35	70
7	18	51	16	37	80
8	21	84	17	44	180
9	22	85	18	48	110

Если предположить, что зависимость между переменными Y и X является линейной вида $Y = a + bX + \varepsilon,$ то оценка функции регрессии, найденная по методу наименьших квадратов, имеет следующий вид: $\hat{Y}_X = \underset{(13,9)}{19,28} + \underset{(0,54)}{2,53}\,X, \quad R^2 = 0,76.$

В скобках указаны стандартные ошибки оценок соответствующих коэффициентов регрессии. Корреляционное поле, соответствующее приведенным в таблице данным, представлено на рис. 1.

Анализируя график, можно предположить наличие гетероскедастичности в модели, поскольку при увеличении значений переменной X разброс значений переменной Y увеличивается.

На уровне значимости 0,05 проверим гипотезу о гетероскедастичности остатков по тесту Спирмена.

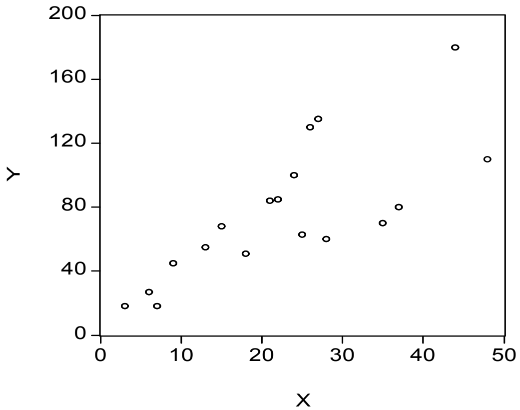


Рис. 1. Зависимость выпуска продукции
на душу населения от ВВП

Решение. Проверим гипотезу $H_0 : \mathrm{D}\left[\varepsilon_1\right] = \mathrm{D}\left[\varepsilon_2\right] = \ldots = = \mathrm{D}\left[\varepsilon_n\right] = \sigma^2$ против альтернативной $H_1 : \mathrm{D}\left[\varepsilon_i\right] \neq \sigma^2,\ i = \overline{1,n}$. В табл. 9 приведены данные о значениях переменной X, регрессионных остатков $|e|$, а также ранги этих величин.

Таблица 9

X	$\lvert e\rvert$	$R(X)$	$R(\lvert e\rvert)$	$R(X)$ - -$R(\lvert e\rvert)$	X	$\lvert e\rvert$	$R(X)$	$R(\lvert e\rvert)$	$R(X)$ - -$R(\lvert e\rvert)$
3	8,87	1	4	–3	24	20,02	10	11	–1
6	7,46	2	3	–1	25	19,51	11	10	1
7	18,99	3	9	–6	26	44,96	12	16	–4
9	2,96	4	2	2	27	47,43	13	17	–4
13	2,84	5	1	4	28	30,10	14	12	2
15	10,78	6	6	0	35	37,80	15	15	0
18	13,81	7	8	–1	37	32,86	16	14	2
21	11,60	8	7	1	44	49,43	17	18	–1
22	10,08	9	5	4	48	30,68	18	13	5

Таким образом, ранговый коэффициент корреляции Спирмена

$$\rho_s = 1 - \frac{6}{n^3 - n}\sum_{i=1}^{n}\left[R(x_i) - R(|e_i|)\right]^2 =$$

$$1 - \frac{6}{18^3 - 18}\cdot\left((-3)^2 + (-1)^2 + (-6)^2 + \cdots + (-1)^2 + 5^2\right) =$$

$$= 1 - \frac{6}{5814}\cdot 152 = 0,84.$$

Фактическое значение статистики критерия

$$K_{выч} = |\rho_s|\sqrt{n-1} = 0,84\cdot\sqrt{18-1} = 3,48.$$

Критическое значение статистики критерия находим по табл. А2:

$$K_{таб} = t_{1-0,05}[N] = t_{0,95}[N] = x_{\frac{1+0,95}{2}}[N] = x_{0,975}[N] = 1,96.$$

Поскольку $|K_{выч}| > K_{таб}$, гипотезу о гомоскедастичности отвергаем. На уровне значимости 0,05 можно утверждать, что модель гетероскедастична.

Тесты для проверки знаний

1. Укажите сумму номеров неверных утверждений

1) Значение коэффициента детерминации не может увеличиться при включении в множественную линейную регрессионную модель переменной, не оказывающей влияние на зависимую переменную

2) Тест Спирмена используют при проверке условия гомоскедастичности модели

3) Корреляционная форма зависимости является частным случаем функциональной зависимости

4) График уравнения простейшей линейной регрессии всегда проходит через точку $(\bar{X}, \bar{Y})$

5) Среди приведенных ответов нет верного

2. Выявление гетероскедастичности в остатках удается устранить при помощи ОМНК

1) введением в выражение для дисперсии остатков коэффициента пропорциональности

2) введением в модель фиктивных переменных

3) преобразованием переменных

4) критерий Дарбина-Уотсона

5) среди приведенных ответов нет верного

3. Нарушением предпосылок МНК является случай …. остатков:

1) наличие автокорреляции

2) случайного характера

3) гомоскедастичности

4) нормального распределения

5) среди приведенных ответов нет верного

4. ОМНК подразумевает

1) преобразование переменных

2) переход от множественной регрессии к парной

3) двухэтапное применение МНК

4) введение в выражение для дисперсии остатков коэффициента пропорциональности

5) среди приведенных ответов нет верного

5. Несмещенность оценки характеризует

1) ее зависимость от объема выборки

2) наименьшую дисперсию остатков

3) равенство нулю математического ожидания остатков

4) увеличение точности ее вычисления с увеличением объема выборки

5) среди приведенных ответов нет верного

6. С помощью теста Гольдфельда-Квандта проверяется

1) наличие мультиколлинеарности в модели

2) наличие автокорреляции в остатках модели

3) наличие гетероскедастичности в остатках модели

4) наличие зависимости между исследуемыми переменными

5) среди приведенных ответов нет верного

7. Значение коэффициента автокорреляции рассчитывается по аналогии с …

1) нелинейным коэффициентом корреляции

2) линейным коэффициентом корреляции

3) линейным коэффициентом детерминации

4) линейным коэффициентом регрессии

5) среди приведенных ответов нет верного

8. Если известна матрица ковариации ошибок, в которой на диагонали все элементы разные, то выполняется условие …

1) состоятельности

2) гомоскедастичности

3) гетероскедастичности

4) некоррелированности

5) среди приведенных ответов нет верного

9. Матрица парных коэффициентов корреляции имеет следующий вид:

	y	x_1	x_2
y	1		
x_1	0,2	1	
x_2	0,6	0,1	1

На основе матрицы парных коэффициентов корреляции определить, какое уравнение регрессии лучше строить:

1) парную линейную регрессию y на x_1

2) парную линейную регрессию y на x_2

3) множественную линейную регрессию y на x_1 и x_2?

4) парную линейную регрессию x_1 на x_2

5) среди приведенных ответов нет верного

10. Критерий Дарбина-Уотсона используется для проверки

1) гетероскедастичности остатков

2) гомоскедастичности остатков

3) автокоррелированности остатков

4) наличия нормального распределения у случайной ошибки регрессионной модели

5) среди приведенных ответов нет верного

11. Выберите номера верных утверждений

1) Корреляционная зависимость является частным случаем функциональной зависимости

2) Параметр b в модели множественной линейной регрессии характеризует среднее изменение Y только вследствие единичного прироста соответствующей переменной X

3) В тесте Гольдфельда-Квандта нулевая гипотеза утверждает, что модель гетероскедастична

1) 1 2) 2 3) 3 4) 1, 2, 3 5) среди приведенных ответов нет верного

12. Выберите номера верных утверждений

1) Число степеней свободы t-статистики для проверки значимости параметров модели $Y = a + bX + \varepsilon$ равно $n-1$

2) Для проверки значимости коэффициента ранговой корреляции используется статистика распределения Фишера

3) $\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$ – необъясненная уравнением регрессии сумма квадратов отклонений

1) 1 2) 2 3) 3 4) 2, 3 5) среди приведенных ответов нет верного

13. Выберите номера верных утверждений

1) В общем случае каждый уровень временного ряда формируется под воздействием тенденции, сезонных колебаний и случайных факторов.

2) Гомоскедастичность остатков подразумевает рост дисперсии остатков с увеличением значения фактора

3) Корреляция подразумевает наличие связи между случайными переменными

1) 1, 2, 3 2) 3 3) 2 4) 2, 3 5) среди приведенных ответов нет верного

14. Выберите номера верных утверждений

1) Один из путей преодоления проблемы мультиколлинеарности состоит в исключении одной из двух объясняющих переменных, связанных пропорциональной зависимостью

2) Для обнаружения автокорреляции первого порядка используется критерий Чоу

3) Обобщенный метод наименьших квадратов может использоваться для корректировки гетероскедастичности остатков

1) 1 2) 2 3) 3 4) 2, 3 5) среди приведенных ответов нет верного

15. Выберите номера верных утверждений

1) Регрессионная модель производственной функции Кобба-Дугласа $Y = AK^{\alpha}L^{\beta}\varepsilon,$ является квадратичной

2) Значение коэффициента автокорреляции рассчитывается по аналогии с нелинейным коэффициентом корреляции

3) С помощью теста Спирмена проверяется наличие автокорреляции в остатках модели

1) 1 2) 2 3) 3 4) 2, 1 5) среди приведенных ответов нет верного

16. Выберите номера верных утверждений

1) Коэффициент детерминации рассчитывается для оценки качества мультиколлинеарных факторов

2) Тест Спирмена используют при проверке условия гомоскедастичности модели

3) Тест Уайта используют при проверке условия автокорреляции в остатках

1) 1 2) 2 3) 3 4) 2, 3 5) среди приведенных ответов нет верного

17. Проявление гетероскедастичности в остатках удается устранить с помощью метода ОМНК путем…

1) расчета критерия Дарбина-Уотсона

2) преобразования переменных

3) введения в выражение для дисперсии остатков коэффициента пропорциональности

4) введения в модель фиктивных переменных

5) среди приведенных ответов нет верного

18. Факторы являются коллинеарными, если…

1) теснота связи между ними строго больше 1

2) теснота связи между ними не превышает по абсолютной величине 0,7

3) теснота связи между ними превышает по абсолютной величине 0,7

4) теснота связи между ними равна 0

5) среди приведенных ответов нет верного

19. Гомоскедастичность подразумевает…

1) рост дисперсии остатков с увеличением значения фактора

2) уменьшение дисперсии остатков с уменьшением значения фактора

3) одинаковую дисперсию остатков при каждом значении фактора

4) максимальную дисперсию остатков при средних значениях фактора

5) среди приведенных ответов нет верного

20. С помощью теста Уайта проверяется…

1) наличие зависимости между исследуемыми переменными

2) наличие мультиколлинеарности в модели

3) наличие гетероскедастичности в остатках модели

4) наличие автокорреляции в остатках модели

5) среди приведенных ответов нет верного

21. Гетероскедастичность остатков подразумевает…

1) зависимость дисперсии остатков от значения фактора

2) независимость математического ожидания остатков от значения фактора

3) зависимость математического ожидания остатков от значения фактора

4) постоянство дисперсии остатков независимо от значения фактора

5) среди приведенных ответов нет верного

22. Одним из методов, используемых в тесте на гетероскедастичность Спирмена, является…

1) нахождение среднего значения

2) ранжирование

3) выравнивание числовых значений выборки по убыванию

4) деление выборки на три равные части

5) среди приведенных ответов нет верного

23. Выберите номера верных утверждений

1) По мере удаления значения зависимой переменной от среднего по выборке длина доверительного интервала для зависимой переменной уменьшается

2) При наличии отрицательной автокорреляции значение статистики Дарбина-Уотсона находится в окрестности числа 4

3) Модель разложения на детерминированную и случайную составляющую может иметь мультипликативную форму

1) 1, 2, 3 2) 1, 2 3) 2, 3 4) 1,3 5) среди приведенных ответов нет верного

24. Для обнаружения автокорреляции первого порядка используется критерий …

1) Чоу
2) Фишера
3) Дарбина-Уотсона
4) Стьюдента
5) среди приведенных ответов нет верного

25. С помощью теста Гольдфельда-Квандта проверяется

1) наличие гетероскедастичности в остатках модели
2) наличие автокорреляции в остатках модели
3) наличие мультиколлинеарности в модели
4) наличие зависимости между исследуемыми переменными
5) среди приведенных ответов нет верного

Глава 6. АНАЛИЗ ОДНОМЕРНЫХ ВРЕМЕННЫХ РЯДОВ

Понятие временного ряда. Стационарные и нестационарные временные ряды. Виды нестационарных временных рядов. Компоненты регрессионной модели аддитивного типа. Коэффициенты автокорреляции, коррелограмма. Моделирование тенденции временного ряда на основе методов скользящего среднего и аналитического выравнивания. Оценка точности выбранных моделей прогнозирования.

<table>
<tr><td colspan="4">Временным рядом («time series») называется последовательность значений случайного показателя $Y(t)=y_t$, упорядоченная в хронологическом порядке $Y(1)=y_1,\ Y(2)=y_2,...,\ Y(n)=y_n$. Отдельные наблюдения временного ряда называются уровнями этого ряда. Временной ряд является частным случаем случайного процесса, который также называется стохастическим процессом.</td></tr>
<tr><td colspan="3">Случайный процесс называется стационарным, если он находится в определенном смысле в статистическом равновесии, т.е. его свойства с вероятностной точки зрения не зависят от времени</td><td>Процесс нестационарен, если эти свойства нарушаются</td></tr>
<tr><td colspan="4" align="center">Виды нестационарных временных рядов</td></tr>
<tr><td colspan="2">регрессионная модель аддитивного типа
$y_t=\psi_1(t)+\psi_2(t)+\psi_3(t)+\varepsilon_t$,
$t=1,...,n$</td><td colspan="2">регрессионная модель мультипликативного типа
$y_t=\psi_1(t)\cdot\psi_2(t)\cdot\psi_3(t)\cdot\varepsilon_t$,
$t=1,...,n$</td></tr>
<tr><td colspan="4" align="center">Компоненты регрессионной модели аддитивного типа</td></tr>
<tr><td>$\psi_1(t)$</td><td colspan="3">тренд временного ряда (некоторая неслучайная функция, отражающая долговременную тенденцию в изменении показателя y_t); чаще всего эта тенденция описывается с помощью монотонной функции полиномиального типа</td></tr>
<tr><td>$\psi_2(t)$</td><td colspan="3">сезонная компонента временного ряда (некоторая неслучай-</td></tr>
</table>

	ная функция, отражающая периодически повторяющиеся через определенные промежутки времени колебания анализируемого показателя внутри года); эта компонента описывается с помощью периодических функций, в аналитическом представлении которых присутствуют гармоники (тригонометрические функции); период T гармоники определяется содержательной сущностью задачи; в описании модели могут присутствовать несколько сезонных компонент. Чаще всего описание таких компонент осуществляется с помощью частичных сумм ряда Фурье следующего вида: $$\psi_2(t) = \sum_{j=1}^{r}\left[A_j \cos\frac{2\pi jt}{T} + B_j \sin\frac{2\pi jt}{T}\right],$$ Также для анализа сезонности используется метод скользящего среднего и фиктивные переменные.	
$\psi_3(t)$	периодическая компонента временного ряда (некоторая неслучайная функция, отражающая периодически повторяющиеся через определенные промежутки времени колебания анализируемого показателя); эта компонента описывается, так же, как и сезонная, с помощью периодических функций	
ε_t	случайная ошибка модели	
Автокорреляция во временном ряду	зависимость значений уровней временного ряда от предыдущего (сдвиг на один временной интервал), предпредыдущего (сдвиг на два временных интервала)	
Коэффициент автокорреляции первого порядка	выборочный коэффициент корреляции между уровнями исходного временного ряда и уровнями того же ряда, сдвинутыми на один момент времени	
Формула	$$\rho(1) = \frac{(n-1)\sum_{t=1}^{n-1} y_t y_{t+1} - \sum_{t=1}^{n-1} y_t \sum_{t=1}^{n-1} y_{t+1}}{\sqrt{(n-1)\sum_{t=1}^{n-1} y_t^2 - \left(\sum_{t=1}^{n-1} y_t\right)^2} \cdot \sqrt{(n-1)\sum_{t=1}^{n-1} y_{t+1}^2 - \left(\sum_{t=1}^{n-1} y_{t+1}\right)^2}}$$	

Свойства	1) коэффициент автокорреляции первого порядка принимает значения от –1 до +1 2) если он близок к единице, то можно выявить наличие во временном ряду тенденции, т.е. установить тесную связь между текущим и непосредственно предшествующим уровнями временного ряда 3) он строится по аналогии с коэффициентом корреляции и, таким образом, характеризует тесноту только линейной связи текущего и предыдущего уровней ряда. Поэтому по коэффициенту автокорреляции можно судить о наличии линейной (или близкой к линейной) тенденции. Для некоторых временных рядов, имеющих сильную нелинейную тенденцию (например, параболу второго порядка или экспоненту), коэффициент автокорреляции уровней исходного ряда может приближаться к нулю 4) по знаку коэффициента автокорреляции нельзя делать вывод о возрастающей или убывающей тенденции в уровнях ряда. Большинство временных рядов экономических данных содержат положительную автокорреляцию уровней, однако при этом могут иметь убывающую тенденцию
Коэффициент автокорреляции порядка τ	выборочный коэффициент корреляции между уровнями исходного временного ряда и уровнями того же ряда, сдвинутыми на τ моментов времени
Формула	$\rho(\tau)=\dfrac{(n-\tau)\sum_{t=1}^{n-\tau} y_t y_{t+\tau} - \sum_{t=1}^{n-\tau} y_t \sum_{t=1}^{n-\tau} y_{t+\tau}}{\sqrt{(n-\tau)\sum_{t=1}^{n-\tau} y_t^2 - \left(\sum_{t=1}^{n-\tau} y_t\right)^2}\cdot\sqrt{(n-\tau)\sum_{t=1}^{n-\tau} y_{t+\tau}^2 - \left(\tau\sum_{t=1}^{n-\tau} y_{t+\tau}\right)^2}}$ обычно порядок коэффициента выбирается $\tau < n/4$
Автокорреляционная функция	это последовательность выборочных коэффициентов автокорреляции уровней первого, второго и т.д. порядков

Коррело-грамма	график зависимости значений временного ряда от величины лага (порядка τ коэффициента автокорреляции)

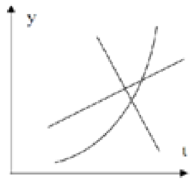
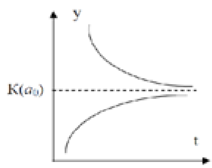
Анализ автокорреляционной функции и коррелограммы позволяет определить лаг, при котором автокорреляция наиболее высокая, следовательно, и лаг, при котором связь между уровнями ряда наиболее тесная, т.е. при помощи анализа автокорреляционной функции и коррелограммы можно выявить структуру ряда:	
1	Если наиболее высоким оказался коэффициент автокорреляции первого порядка, исследуемый ряд может содержать только тенденцию.
2	Если наиболее высоким оказался коэффициент автокорреляции порядка τ $(\tau>1)$, то ряд может содержать циклические колебания с периодичностью в τ моментов времени.
3	Если ни один из коэффициентов автокорреляции не является значимым, можно сделать следующие предположения: ряд не содержит тенденции и циклических колебаний и имеет структуру случайной компоненты; ряд может содержать сильную нелинейную тенденцию, для выявления которой нужно провести дополнительный анализ.

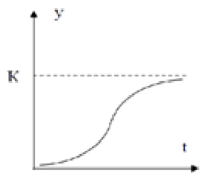
Моделирование тенденции временного ряда

Моделирование тенденции (тренда) временного ряда является важнейшей классической задачей анализа экономических временных рядов	
Сглаживание временного ряда по методу скользящего среднего	
Цель сглаживания временного ряда заключается в получении ряда с меньшим разбросом уровней, что в ряде случаев позволяет на основе визуального анализа сделать вывод о наличии тенденции и ее характерных особенностях	Сглаживание заключается в замене исходных уровней ряда y_t сглаженными значениями y'_t, которые получаются как среднее значение определенного числа уровней исходного ряда, симметрично окружающих значение y_t

<table>
<tr><td colspan="3">Для вычисления сглаженных значений y'_t по методу скользящего среднего используются следующие формулы:</td></tr>
<tr><td colspan="2">при нечетном количестве исходных уровней ряда g = 2p+1
$$y'_t = \frac{\sum_{l=t-p}^{t+p} y_i}{2p+1} = $$ $$= \frac{y_{t-p} + y_{t-p+1} + \ldots + y_{t+p-1} + y_{t+p}}{2p+1}$$</td><td>при четном количестве исходных уровней ряда g = 2p
$$y'_t = \frac{\frac{1}{2} y_{t-p} + y_{t-p+1} + \ldots + y_{t+p-1} + \frac{1}{2} y_{t+p}}{2p}.$$</td></tr>
<tr><td colspan="3">где y_t – фактическое значение уровня исходного ряда в момент времени t, y'_t – значение скользящей средней в момент времени t.</td></tr>
<tr><td colspan="2">Формула при $g = 3$ и $g = 5$ имеет вид $y'_t = \frac{y_{t-1} + y_t + y_{t+1}}{3}$, и
$$y'_t = \frac{y_{t-2} + y_{t-1} + y_t + y_{t+1} + y_{t+2}}{5}.$$</td><td>Формула при $g = 2$ и $g = 4$ принимает вид $y'_t = \frac{\frac{1}{2} y_{t-1} + y_t + \frac{1}{2} y_{t+1}}{2}$, и
$$y'_t = \frac{\frac{1}{2} y_{t-2} + y_{t-1} + y_t + y_{t+1} + \frac{1}{2} y_{t+2}}{4}.$$</td></tr>
<tr><td colspan="3">При использовании скользящего среднего при g = 2p+1 первые и последние p уровней ряда сгладить нельзя. Иногда можно использовать следующий прием:</td></tr>
<tr><td>а)</td><td colspan="2">Вычисляется средний прирост Δ_y на последнем активном участке
$(y_{n-g}, \ldots, y_n)$ $\Delta_y = \frac{y_n - y_{n-g}}{g-1}$, где g – количество уровней ряда</td></tr>
<tr><td>б)</td><td colspan="2">Определяются значения последних p = (g–1)/2 уровней сглаженного временного ряда с помощью последовательного прибавления среднего абсолютного прироста Δ_y к последнему сглаженному значению y'_{n-p}:
$$y'_{n-p+1} = y'_{n-p} + \Delta_y,\ y'_{n-p+2} = y'_{n-p+1} + \Delta_y, \ldots, y'_n = y'_{n-1} + \Delta_y.$$</td></tr>
<tr><td colspan="3">Отметим, что важным свойством процедуры сглаживания является полное устранение сезонных и периодических компонент из временного ряда, если длина интервала сглаживания берется равной или кратной периоду колебаний.</td></tr>
</table>

<table>
<tr><td colspan="3">Метод аналитического выравнивания</td></tr>
<tr><td colspan="3">Аналитический подход основан на допущении, что исследователь может задать общий вид функции, описывающей регулярную, неслучайную составляющую.</td></tr>
<tr><td colspan="3">Аналитическим выравниванием временного ряда называют нахождение аналитической функции $\hat{y}_t = \psi_1(t)$, характеризующей основную тенденцию изменения уровней ряда с течением времени.</td></tr>
<tr><td colspan="2">Кривая роста</td><td>функция $\psi_1(t)$</td></tr>
<tr><td colspan="2">линейная</td><td>$\psi_1(t) = a_0 + a_1 t$</td></tr>
<tr><td colspan="2">парабола второго и более высоких порядков</td><td>$\psi_1(t) = a_0 + a_1 t + a_2 t^2$,</td></tr>
<tr><td colspan="2">гиперболическая</td><td>$\psi_1(t) = a_0 + a_1 / t$</td></tr>
<tr><td colspan="2">экспонента</td><td>$\psi_1(t) = e^{a_0 + a_1 t}$</td></tr>
<tr><td colspan="2">показательная</td><td>$\psi_1(t) = a_0 \cdot a_1^{\ t}$</td></tr>
<tr><td colspan="2">степенная</td><td>$\psi_1(t) = a_0 \cdot t^{a_1}$</td></tr>
<tr><td colspan="3">Для определения вида тенденции (тренда) применяются следующие методы:</td></tr>
<tr><td></td><td colspan="2">качественный анализ изучаемого процесса</td></tr>
<tr><td></td><td colspan="2">построение и визуальный анализ графика зависимости уровней ряда от времени</td></tr>
<tr><td></td><td colspan="2">расчет и анализ показателей динамики временного ряда</td></tr>
<tr><td></td><td colspan="2">анализ автокорреляционной функции исходного и преобразованного временного ряда</td></tr>
<tr><td></td><td colspan="2">метод перебора, при котором строятся кривые роста различного вида с последующим выбором наилучшей на основании различных показателей оценки точности моделей</td></tr>
<tr><td colspan="3">Выбор вида тенденции на основе качественного анализа</td></tr>
</table>

Процессы с монотонным характером развития и отсутствием пределов роста. Эти условия справедливы для поведения многих экономических показателей, например, для большинства натуральных показателей промышленного производства. В этом случае для моделирования тенденции могут использоваться: линейная функция $\psi_1(t) = a_0 + a_1 t$, параболическая функция $\psi_1(t) = a_0 + a_1 t + a_2 t^2$, экспоненциальная функция $\psi_1(t) = e^{a_0 + a_1 t}$, степенная функция $\psi_1(t) = a_0 \cdot t^{a_1}$.	 *а*) I класс
II) *Процессы, которые имеют предел роста (падения) в исследуемом периоде, так называемые процессы с «насыщением».* Развитие процесса происходит под влиянием некоторых ограничивающих факторов, величина воздействия которых растет вместе с ростом достигнутого уровня. В этом случае для моделирования тенденции используются гиперболическая функция $\psi_1(t) = a_0 + a_1 / t$ модифицированная экспонента $\psi_1(t) = K + a_0 \cdot a_1^{\,t}$ с параметром a_1, удовлетворяющим условию $0 < a_1 < 1$. В случае гиперболы параметр a_0 равен пределу роста, к которому значение уровня процесса приближается (при росте *t*) снизу в случае $a_1 < 0$ либо сверху при $a_1 > 0$. В случае модифицированной экспоненты параметр *K* равен пределу роста, к которому значение уровня процесса приближается (при росте *t*) снизу в случае $a_1 < 0$ либо сверху при $a_1 > 0$.	 *б*) II класс

III) *S-образные процессы*. Такие процессы представляют как бы два последовательных лавинообразных процесса (когда прирост зависит от уже достигнутого уровня): один с ускорением развития, а другой – с замедлением. С такими процессами часто сталкиваются в демографических исследованиях, в страховых расчетах, при решении задач прогнозирования научно-технического прогресса, при определении спроса на новый вид продукции. Для моделирования тенденции *S*-образных процессов следует использовать логистическую функцию $\psi_1(t)=\dfrac{K}{1+a_0\cdot e^{-a_1 t}}$ (с параметром $a_1<1$), кривую Гомперца $\psi_1(t)=K\cdot a_0^{a_1^t}$ с параметрами, удовлетворяющими условиям $a_0>0$, $a_1<1$. Предел роста в обоих случаях равен параметру *K*.	 *в*) III класс
Выбор вида тенденции на основе анализа показателей динамики временного ряда	
Если исходный временной ряд содержит тенденцию, а временной ряд последовательных разностей первого порядка $\Delta y_t = y_t - y_{t-1}$ не содержит тенденцию, то можно сделать вывод, что тенденция линейно зависит от времени $\psi_1(t)=a_0+a_1 t$.	Коэффициент a_1 в данном случае численно равен среднему абсолютному приросту уровня явления за единицу измерения временного параметра *t* (за сутки, неделю, месяц, год и т.д.).
Если исходный временной ряд и временной ряд последовательных разностей первого порядка содержат тенденцию, а временной ряд последовательных разностей второго порядка	Относительно тенденции в виде полинома от *t* более высокой степени вывод делается аналогично.

$\Delta^2 y_t = \Delta y_t - \Delta y_{t-1}$ не содержит тенденцию, то можно сделать вывод, что тенденция задается полиномом второго порядка от времени $\psi_1(t) = a_0 + a_1 t + a_2 t^2$.	
Если исходный временной ряд содержит тенденцию, а временной ряд коэффициентов роста $k_t = y_t / y_{t-1}$ не содержит тенденцию, то можно сделать вывод, что тенденция экспоненциально зависит от времени. Иными словами, тенденция имеет вид экспоненциальной функции $\psi_1(t) = e^{a_0 + a_1 t}$.	Величина a_1 в данном случае численно равна среднему коэффициенту роста уровня явления за единицу измерения временного параметра t (за сутки, неделю, месяц, год и т. д.).

Оценка точности выбранных моделей прогнозирования

Показатели оценки точности характеризуют качество модели. Ошибка прогноза – величина, характеризующая расхождение между фактическим y_t и прогнозным значением показателя $\hat{y}_t$		
Показатель	***Формула***	***Свойства***
Абсолютная ошибка прогноза	$\Delta_t = \hat{y}_t - y_t$	Имеет ту же размерность, что и прогнозируемый показатель и зависит от масштаба измерения уровней ряда
Относительная ошибка прогноза	$\delta_t = \frac{\hat{y}_t - y_t}{y_t} \cdot 100\%$	Показывает, насколько велика ошибка по сравнению с фактическим значением уровня исходного ряда. Если относительная ошибка больше нуля , то прогнозная оценка завышена и наоборот

Средняя абсолютная ошибка (Mean Absolute Error, MAE)	$\left\lvert\overline{\Delta}\right\rvert = \dfrac{\sum_{t=1}^{n} \lvert\Delta_t\rvert}{n} = \dfrac{\sum_{t=1}^{n} \lvert\hat{y}_t - y_t\rvert}{n}$	Используют в случае, когда оценку нужно получить в тех же единицах, в которых измерены уровни ряда
Среднеквадратичная ошибка (Mean Squared Error, MSE)	$\hat{\sigma} = \sqrt{\dfrac{\sum_{t=1}^{n} (\hat{y}_t - y_t)^2}{n}}$ или $\hat{\sigma} = \sqrt{\dfrac{\sum_{t=1}^{n} (\hat{y}_t - y_t)^2}{n-m}}$, где m – число параметров в модели	При сравнении двух моделей наилучшей признается та, у которой средневкадратичная ошибка наименьшая
Средняя абсолютная процентная ошибка (Mean Absolute Percentage Error, MAPE)	$\left\lvert\overline{\delta}\right\rvert = \dfrac{\sum_{t=1}^{n} \lvert\delta_t\rvert}{n} \cdot 100\%$	Если $\left\lvert\overline{\delta}\right\rvert < 10\%$ – высокая точность модели; $10\% < \left\lvert\overline{\delta}\right\rvert < 20\%$ – точность признать хорошей; $20\% < \left\lvert\overline{\delta}\right\rvert < 50\%$ – точность удовлетворительная

Пример. Временной ряд, определяющий динамику товарооборота одного из филиала некоторой торговой сети (%), задан в табл. 10.

Таблица 10

Динамика товарооборота (%) по кварталам

Год	КВАРТАЛ			
	первый	второй	третий	четвертый
2012	31,7	34,5	37	39,8
2013	42,2	44,6	46,9	48,8
2014	51	53,9	56,9	59,8

Определить коэффициент автокорреляции первого порядка и дайте его интерпретацию.

Решение. Построим графическое изображение временного ряда по данным за кварталы. На графике (рис. 2) наглядно видно наличие возрастающей тенденции. Естественно предположить, что товарооборот филиала в текущий момент времени зависит от товарооборота в предыдущий момент времени. Возможно существование линейного, степенного и полиномиального тренда.

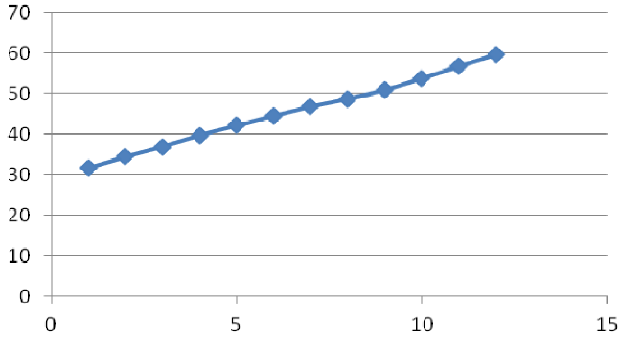


Рис. 2. Поле корреляции

Для расчета коэффициентов автокорреляции 1-го и 2-го порядков вычислим необходимые суммы, для этого воспользуемся вспомогательной таблицей (*табл.* 11):

Таблица 11

Расчет коэффициентов автокорреляции

t	y_t	y_{t+1}	$y_t y_{t+1}$	y_t^2	y_{t+1}^2
1	31,7	34,5	1093,65	1004,89	1190,25
2	34,5	37	1276,5	1190,25	1369
3	37	39,8	1472,6	1369	1584,04
4	39,8	42,2	1679,56	1584,04	1780,84
5	42,2	44,6	1882,12	1780,84	1989,16
6	44,6	46,9	2091,74	1989,16	2199,61
7	46,9	48,8	2288,72	2199,61	2381,44
8	48,8	51	2488,8	2381,44	2601

Окончание табл. 11

t	y_t	y_{t+1}	$y_t y_{t+1}$	y_t^2	y_{t+1}^2
9	51	53,9	2748,9	2601	2905,21
10	53,9	56,9	3066,91	2905,21	3237,61
11	56,9	59,8	3402,62	3237,61	3576,04
12	59,8				
Суммы	487,3	515,4	23492,12	22243,05	24814,2

Таким образом, коэффициент автокорреляции первого порядка будет равен:

$$r(1)=\frac{(n-1)\sum_{t=1}^{n-1}y_t y_{t+1}-\sum_{t=1}^{n-1}y_t\sum_{t=1}^{n-1}y_{t+1}}{\sqrt{(n-1)\sum_{t=1}^{n-1}y_t^2-\left(\sum_{t=1}^{n-1}y_t\right)^2}\cdot\sqrt{(n-1)\sum_{t=1}^{n-1}y_{t+1}^2-\left(\sum_{t=1}^{n-1}y_{t+1}\right)^2}}=$$

$$\frac{(12-1)\cdot 23492{,}12-487{,}3\cdot 515{,}4}{\sqrt{(12-1)\cdot 22243{,}05-487{,}3^2}\cdot\sqrt{(12-1)\cdot 24814{,}2-515{,}4^2}}=0{,}99.$$

Вывод. Такое значение коэффициента автокорреляции первого порядка говорит о том, что во временном ряду существует ярко выраженная линейная тенденция.

Тесты для проверки знаний

1. Под уровнем временного ряда подразумевается…
1) значение показателя в определенный момент времени
2) усредненное значение временного ряда
3) наименьшее значение временного ряда
4) наибольшее значение временного ряда
5) среди приведенных ответов нет верного

2. Временной ряд характеризует…
1) совокупность последовательных моментов времени
2) данные, описывающие один объект за ряд последовательных моментов времени

3) данные, описывающие совокупность разлисчных объектов в определенный момент времени

4) зависимость последовательных моментов времени

5) среди приведенных ответов нет верного

3. Укажите методы уменьшения (устранения) автокорреляции во временных рядах:

1) авторегрессионных преобразований;

2) построения коррелограммы;

3) включения дополнительного фактора;

4) последовательных разностей

5) среди приведенных ответов нет верного

4. Временной ряд – это совокупность значений экономического показателя …

1) за несколько последовательных моментов (периодов) времени

2) по однотипным объектам

3) за несколько непоследовательных моментов (периодов) времени

4) независящих от времени

5) среди приведенных ответов нет верного

5. Компонента временного ряда, отражающая повторяемость экономических процессов в течение не очень длительного периода (года, квартала, месяца и т.д.), называется …

1) сезонной компонентой

2) трендом

3) циклической компонентой

4) случайной компонентой

5) среди приведенных ответов нет верного

6. Долгосрочную тенденцию изменения признака называют …

1) сезонной компонентой

2) циклической компонентой

3) случайной компонентой

4) трендом

5) среди приведенных ответов нет верного

7. В стационарном временном ряде трендовая компонента …

1) присутствует

2) отсутствует

3) имеет линейную зависимость от времени

4) имеет нелинейную зависимость от времени

5) среди приведенных ответов нет верного

8. Значение коэффициента автокорреляции рассчитывается по аналогии с …

1) нелинейным коэффициентом корреляции

2) линейным коэффициентом корреляции

3) линейным коэффициентом детерминации

4) линейным коэффициентом регрессии

5) среди приведенных ответов нет верного

9. В общем случае каждый уровень временного ряда формируется под воздействием …

1) тенденции и случайных факторов

2) тенденции, сезонных колебаний и случайных факторов.

3) случайных временных воздействий

4) сезонных колебаний и случайных факторов

5) среди приведенных ответов нет верного

10. Модель временного ряда предполагает …

1) отсутствие последовательности моментов (периодов) времени, в течении которых рассматривается поведение экономического показателя

2) независимость значений экономического показателя от времени

3) зависимость значений экономического показателя от времени

4) пренебрежение временными характеристиками ряда

5) среди приведенных ответов нет верного

11. Если временной ряд представлен в виде произведения соответствующих компонент, то полученная модель носит название …

1) аддитивной

2) смешанной

3) мультипликативной

4) степенной

5) среди приведенных ответов нет верного

12. Стационарность временного ряда означает отсутствие …

1) наблюдений по уровням временного ряда

2) временной характеристики

3) значений уровней ряда

4) тренда

5) среди приведенных ответов нет верного

13. Под автокорреляцией уровней временного ряда подразумевается

1) корреляционная зависимость между последовательными уровнями ряда

2) функциональная зависимость между последовательными уровнями ряда

3) функциональная зависимость между двумя временными рядами

4) корреляционно-функциональная зависимость между последовательными уровнями ряда

5) среди приведенных ответов нет верного

14. Аддитивная модель содержит компоненты в виде

1) сомножителей
2) отношений
3) слагаемых
4) комбинации слагаемых и сомножителей
5) среди приведенных ответов нет верного

15. Пусть Y_t – временной ряд, T_t- трендовая, S_t- сезонная, а E_t- случайная его составляющие. В принятых обозначениях мультипликативная временная модель выглядит следующим образом:

1) $Y_t=T_t \cdot S_t+E_t$
2) $Y_t=T_t \cdot S_t \cdot E_t$
3) $Y_t=T_t+S_t+E_t$
4) $Y_t=T_t+S_t \cdot E_t$
5) среди приведенных ответов нет верного

16. Пусть X_t – значения временного ряда с квартальными наблюдениями, S_t – мультипликативная сезонная компонента, причем для второго квартала года $S_t = S_2 = 1/6$, для третьего года $S_t = S_3 = 3$, для четвертого квартала года $S_t = S_4 = 2$. Определите оценку сезонной компоненты для первого квартала года $S_t = S_1 = ...$

1) 1
2) –1
3) 0
4) 6/5
5) среди приведенных ответов нет верного

ДОМАШНЯЯ КОНТРОЛЬНАЯ РАБОТА

Задание 1 «Парная линейная регрессионная модель»

1) Выполнить экономический анализ задачи и сделать выводы, что вы выбираете в качестве изучаемого показателя (Y), и что в качестве влияющего (X). Постройте поле корреляции результата и фактора и сформулируйте гипотезу о форме связи.

2) Используя выборочные данные, заполнить следующую таблицу:

№п/п	X	Y	X^2	Y^2	$X \cdot Y$
1					
2					
…..					…
n					
Сумма					
Среднее					

3) По выборочным данным найти средние значения по Y и X, дисперсии по Y и X, средние квадратичные отклонения по Y и X, результаты расчетов и соответствующие формулы поместить в следующую таблицу:

Формула для среднего значения X	Среднее значение X	Единицы измерения
….	…..	….
Формула для среднего значения Y	Среднее значение Y	Единицы измерения
….	…	…
Формула для дисперсии по X	Дисперсия по X	Единицы измерения
…..	……..	…….
Формула для дисперсии по Y	Дисперсия по Y	Единицы измерения
……	…..	…..
Формула для среднего квадратичного по X	Среднее квадратичное по X	Единицы измерения
….	……	……
Формула для среднего квадратичного по Y	Среднее квадратичное по Y	Единицы измерения

4) Определить выборочный коэффициент корреляции и поясните его смысл. Сделать вывод о силе линейной зависимости между переменными Y и X. На уровне значимости 0,05 проверить гипотезу о значимости коэффициента корреляции в генеральной совокупности.

5) Оценить параметры парной линейной регрессионной модели $Y = a + bX + \varepsilon$ методом наименьших квадратов. Дать экономическую интерпретацию найденных коэффициентов.

6) Вычислить регрессионные остатки. Определить *RSS, TSS, ESS*. Проверить основное тождество дисперсионного анализа. Найти оценку дисперсии ошибки модели.

7) Определить стандартные ошибки оценок параметров регрессии и стандартную ошибку модели.

8) Определить коэффициент детерминации и дать его интерпретацию.

9) Построить 95%-ые доверительные интервалы для коэффициентов регрессии.

10) На уровне значимости 0,05 оцените статистическую значимость коэффициента регрессии b. Сделайте выводы.

11) На уровне значимости 0,05 оцените статистическую значимость уравнения регрессии в целом. Сделайте выводы.

12) Найти точечный прогноз для зависимой переменной и дать интерпретацию полученного результата, если факторный признак увеличится на 5% от своего среднего значения.

13) С вероятностью 0,95 постройте доверительный интервал ожидаемого значения результативного признака, если факторный признак увеличится на 5% от своего среднего значения.

14) С вероятностью 0,95 постройте доверительный интервал индивидуального значения результативного признака, если факторный признак увеличится на 5 % от своего среднего значения.

Вариант 1. По территориям Центрального района известны данные за 2014г. О доли денежных доходов, направленных на прирост сбережений во вкладах, займах, сертификатах и на покупку валюты, в общей сумме среднедушевого денежного дохода, % и среднемесячной заработной платы, тыс. руб.

Район	Среднедушевой денежный доход	Среднемесячная зарплата
Брянская обл.	6,9	289
Владимирская обл.	8,7	334
Ивановская обл.	6,4	300
Калужская обл.	8,4	343
Костромская обл.	6,1	356
Орловская обл.	9,4	289
Рязанская обл.	11,0	341
Смоленская обл.	6,4	327
Тверская обл.	9,3	357
Тульская обл.	8,2	352
Ярославская обл.	8,6	381

Вариант 2. По территориям Центрального района известны данные за 2015г.:

Район	Средний размер назначенных ежемесячных пенсий, тыс. руб.	Прожиточный минимум в среднем на одного пенсионера в месяц, тыс. руб.,
Брянская обл.	24	17,8
Владимирская обл.	22	20,2
Ивановская обл.	22	19,7
Калужская обл.	21	20,1
Костромская обл.	22	18,9
г. Москва	25	30,2
Московская обл.	23	21,5
Орловская обл.	23	16,6
Рязанская обл.	21	19,9
Смоленская обл.	22	18,0
Тверская обл.	22	18,1
Тульская обл.	23	18,6
Ярославская обл.	22	17,8

Вариант 3. Рассматривается зависимость производительности труда от возраста рабочих

№ п/п	Производительность труда рабочих, тыс. руб.,	Возраст рабочих	№ п/п	Производительность труда рабочих, тыс. руб.,	Возраст рабочих
1	12	20	6	11	19
2	8	30	7	12	30
3	13	40	8	9	18
4	15	35	9	11	29
5	16	26	10	9	21

Вариант 4. Рассматривается зависимость между месячными затратами на рекламу и соответствующими объемами продаж

Месяц	Расходы на рекламу (100 ф.ст.)	Объемы продаж (100 ф.ст.)
Январь	4.1	15.6
Февраль	6.2	16.8
Март	5.8	15.9
Апрель	7.9	16.6
Май	8.6	16.4
Июнь	3.0	15.9
Июль	5.0	15.8
Август	7.2	17.0
Сентябрь	8.4	16.9
Октябрь	10.6	18.2
Ноябрь	11.0	17.5
Декабрь	7.0	15.9

Вариант **5.** На крупном промышленном предприятии было проведено исследование зависимости времени подготовки работника, необходимое для выработки у него навыков в определенной области, от возраста. Имеются следующие данные:

Показатели	Работник									
	А	Б	В	Г	Д	Е	Ж	З	И	К
Возраст (лет)	18	19	20	21	22	23	29	38	40	41
Время подготовки (часов)	4	3	4	6	5	8	6	7	10	8

Вариант 6. Во время недавних переговоров между работниками и руководством представители профсоюза пожаловались на слабость управления, выражающуюся в потерях времени из-за нехватки материалов и выхода оборудования из строя. В настоящее время в компании действует система оплаты труда, согласно которой до 25 % заработной платы работника формируется за счет дополнительных начислений по результатам производительности труда. Участники переговоров со стороны профсоюза сделали упор на то, что работники теряют в заработной плате не по своей вине. Для подтверждения этого они представили данные по средним суммам начислений за производительность труда группе из 50 работников в сравнении с временными потерями за период в 10 недель.

Неделя	Средняя сумма начислений за производительность труда (ф. ст.)	Потери производственного времени (%)
1	40	8
2	35	6
3	20	10
4	25	11
5	45	5
6	60	4
7	75	4
8	40	6
9	20	12
10	50	8

Вариант 7. Бюджетное обследование 10 случайным образом отобранных семей дало следующие результаты:

Номер семьи	1	2	3	4	5	6	7	8	9	10
Реальный доход семьи (т.руб.)	5.0	4.5	4.2	7.5	3.5	6.2	7.7	6.0	5.9	3.8
Реальный расход семьи на продовольственные товары (т.руб.)	3.0	2.6	1.5	3.4	1.8	5.0	5.2	4.3	3.6	2.1

Вариант 8. Изучается влияние изменения объема промышленного производства на товарооборот. Для этого по 10 регионам РФ были получены следующие данные:

№ п/п	Розничный товарооборот (в % к пред. году)	Объем промышленного производства (в % к пред. году)
1	80	85
2	75	73
3	81	86
4	95	80
5	90	97
6	92	79
7	89	92
8	84	88
9	89	86
10	87	77

Вариант 9. Изучается влияние изменения среднедушевого дохода на товарооборот. Для этого по 10 регионам РФ были получены следующие данные:

№ п/п	Розничный товарооборот (в % к пред. году)	Среднедушевой денежный доход (в % к пред. году)
1	89	88
2	75	85
3	82	81
4	84	87
5	91	87
6	92	110
7	89	102
8	107	105
9	89	94
10	87	92

Вариант 10. Имеются следующие данные о потреблении электроэнергии владельцами индивидуальных домов:

№ п/п	1	2	3	4	5	6	7	8	9	10
Число совместно проживающих членов семьи	2	3	3	4	4	5	5	6	7	7
Годовое потребление электроэнергии, тыс.квт.-час.	15	14	16	19	20	22	23	25	24	22

Вариант 11. Имеются следующие данные:

Чистый доход, млрд долл. США	-0,9	1,3	2	0,6	0,7	0,9	1,1	1,9	2,6	1,3
Оборот капитала, млрд долл. США	12,7	21,4	13,5	13,4	4,2	15	15,5	17,9	16,5	20

Вариант 12. Имеются следующие данные:

Чистый доход, млрд долл. США	-0,9	1,3	2	0,6	0,7	0,9	1,1	1,9	2,6	1,3
Использованный капитал, млрд долл. США	11,9	13,7	11,6	14,2	12,8	15	13	5,8	6	2,5

Вариант 13. Администрация страховой компании приняла решение о введении нового вида услуг – страхования на случай пожара. С целью определения тарифов по выборке 10 случаев пожара анализируется зависимость стоимости ущерба, нанесенного пожаром, от расстояния до ближайшей пожарной станции:

№ п/п	1	2	3	4	5	6	7	8	9	10
Общая сумма ущерба (тыс.руб.)	20	15	10	18	15	9	10	12	14	19
Расстояние до ближайшей пожарной станции (км)	3,6	1,7	2,9	3,0	2,2	3,2	3,6	3,9	4,2	4,8

Вариант 14. Имеются данные по 10 предприятиям:

№	Производительность труда, млн. руб. на 1 рабочего	Энерговооруженность квт час на 1 рабочего
1	9,8	4,8
2	6,7	2,6
3	12,4	7,0
4	6,9	3,8
5	11,8	5,5
6	7,3	3,0
7	8,4	3,4
8	10,7	5,2
9	11,1	5,4
10	7,3	2,9

Вариант 15. Имеются данные по 10 предприятиям:

№	Производительность труда, млн. руб. на 1 рабочего	Доля рабочих ручного труда в общей численности рабочих
1	9,8	40
2	6,7	59
3	12,4	38
4	6,9	57
5	11,8	31
6	7,3	56
7	8,4	45
8	10,7	35
9	11,1	32
10	7,3	54

Вариант 16. Имеются данные о рекламной компании в магазинах с демонстрацией антисептических качеств своего нового моющего средства. Через 10 недель компания решила проанализировать эффективность этого вида рекламы, сопоставив еженедельные:

Объем продаж, тыс. руб.	72	76	78	70	68	80	79	78	69	71
Расходы на рекламу, тыс. руб.	5	8	6	5	3	9	8	6	7	4

Вариант 17. Получены данные для предприятий машиностроения:

№	Рентабельность (прибыль в $)	Произв. труда, млн. руб. на 1 работника
1	7	7
2	8	10
3	7	9
4	9	11
5	9	11
6	8	11
7	11	17
8	11	14
9	16	13
10	14	12

Вариант 18. Получены данные для предприятий машиностроения:

№	Рентабельность (прибыль в $)	Средний возраст производственного оборудования, лет
1	7	20
2	8	19
3	7	21
4	9	17
5	9	16
6	8	18
7	11	15
8	11	14
9	16	10
10	14	13

Вариант 19. Имеются данные по группам предприятий за отчетный период о зависимости себестоимости единицы продукции от величины выпуска продукции:

Себестоимость, тыс. руб.	2	3	4	5	6	7	8	6	8	9
Выпуск продукции, тыс. шт.	1.9	1.7	1.8	1.6	1.4	1.6	1.7	1.5	1.3	1.9

Вариант 20. Компанию по прокату автомобилей интересует зависимость между пробегом автомобилей (тыс. км) и стоимостью технического обслуживания (тыс. руб.). Для выяснения этой связи было отобрано 10 автомобилей:

X	6	7	8	9	10	12	9	8	12	15
У	13	16	15	20	19	14	10	9	11	18

Вариант 21. Имеются следующие данные о бюджетном обследовании семей

Номер семьи	1	2	3	4	5	6	7	8	9	10
Реальный доход семьи (тыс. руб.)	5.0	4.5	4.2	7.5	3.5	6.2	7.7	6.0	5.9	3.8
Реальный расход семьи на продовольственные товары (тыс. руб)	3.0	2.6	1.5	3.4	1.8	5.0	5.2	4.3	3.6	2.1

Вариант 22. Имеются следующие данные об обследовании семей

Номер семьи	1	2	3	4	5	6	7	8	9	10
Число совместно проживающих членов семьи, чел.	3	3	2	4	4	4	5	6	7	7
Годовое потребление электроэнергии, тыс. кв.-час	15	12	18	21	20	24	26	29	31	28

Вариант 23. Имеются следующие данные о бюджетном обследовании семей

Номер семьи	1	2	3	4	5	6	7	8	9	10
Реальный доход семьи (т.руб.)	6.0	3	5	6	4	7	7	7	6	4
Реальный расход семьи на продовольственные товары (т.руб.)	3,5	3	2	4	1.8	2,2	6,2	3,3	3.6	2,3

Вариант 24. Имеются следующие по предприятиям

Потреблено материалов на единицу продукции, кг	9	6	5	4	3,7	3,5	6	7	3,5	3,6
Выпуск продукции, тыс.ед.	100	200	300	400	500	600	700	150	120	250

Вариант 25. Имеются данные о пробеге подержанного автомобиля и его цене.

Пробег, тыс.км	Цена, долл
5	15360
10	13700
15	12180
17	12000
20	12500
20	12630
25	12000
30	10000
30	11000
31	10500
65	4000
70	2600
75	1790
92	500

Задание 2 «Множественная линейная регрессионная модель»

1) Определите эндогенные и экзогенные переменные задачи. Выдвинете гипотезу о виде связи между зависимой и независимыми переменными и запишите соответствующую модель.

2) Найдите оценки параметров множественной линейной модели методом наименьших квадратов. Запишите полученное уравнение множественной регрессии и поясните экономический смысл его параметров при переменных. (Задание можно решить с помощью Пакета анализа данных MS Excel).

3) Определите парные коэффициенты корреляции. Между какими показателями коэффициент корреляции наибольший? Сделайте выводы. Выясните возможную мультиколлинеарность в модели. (Задание можно решить с помощью инструмента Корреляция MS Excel).

4) Используя найденные парные коэффициенты корреляции, вычислить частные коэффициенты корреляции. Сделайте выводы.

5) Найдите коэффициенты эластичности по всем переменным. Проинтерпретируйте все коэффициенты эластичности. Определите, какой фактор оказывает наибольшее влияние на Y.

6) Определите стандартизованные коэффициенты регрессии. Запишите уравнение регрессии в стандартизованной форме.

7) Определить множественный коэффициент корреляции, коэффициент детерминации, скорректированный коэффициент детерминации, сделайте выводы.

8) На уровне значимости 0,05 проверить значимость параметров модели.

9) Постройте 95%-ые доверительные интервалы для параметров множественной регрессии.

10) На уровне значимости 0,05 оцените статистическую значимость уравнения регрессии в целом.

11) На уровне значимости 0,05 проверьте гипотезу о гетероскедастичности остатков модели с помощью критерия Спирмена.

12) На уровне значимости 0,1 проверьте предположение об автокорреляции остатков, используя критерий Дарбина-Уотсона.

13) Подведите общий итог: можно ли использовать данную модель для прогноза? Если нет, то, как следует изменить модель для ее практического использования?

Вариант 1. Изучается влияние изменения объема промышленного производства и среднедушевого дохода на товарооборот. Для этого по 10 регионам РФ были получены следующие данные:

№ п/п	Розничный товарооборот (в % к пред. году)	Объем промышленного производства (в % к пред. году)	Среднедушевой денежный доход (в % к пред. году)
1	89	85	88
2	75	70	85
3	82	86	81
4	84	80	87
5	91	97	87
6	92	79	110
7	89	92	102
8	107	99	105
9	89	83	94
10	87	77	92

Вариант 2. .По выборке из 10 почтовых отправлений изучается зависимость отправки корреспонденции экспресс почтой от веса конверта и дальности перевозки:

№	Стоимость доставки, тыс.руб.	Вес конверта, г	Дальность перевозки, тыс.км
1	26	590	0,5
2	39	320	1,5
3	80	440	2,0
4	92	660	1,6
5	44	75	2,8
6	15	70	0,8
7	145	650	2,4
8	19	450	0,5
9	10	60	1,0
10	140	750	1,9

Вариант 3. Имеются данные по 10 предприятиям:

№	Производительность труда, млн. руб. на 1 рабочего	Энерговооруженность квт час на 1 рабочего	Доля рабочих ручного труда в общей численности рабочих
1	9,8	4,8	40
2	6,7	2,6	59
3	12,4	7,0	38
4	6,9	3,8	57
5	11,8	5,5	31
6	7,3	3,0	56
7	8,4	3,4	45
8	10,7	5,2	35
9	11,1	5,4	32
10	7,3	2,9	54

Вариант 4. Получены данные для предприятий машиностроения:

№	Рентабельность (прибыль в $ стоимость основных и оборотных фондов)	Произв. труда, млн руб. на 1 работника	Средний возраст производств. оборудования, лет
1	7	7	20
2	8	10	19
3	7	9	21
4	9	11	17
5	9	11	16
6	8	11	18
7	11	17	15
8	11	14	14
9	16	13	10
10	14	12	13

Вариант 5. Получены данные о ценах и дивидендах по обыкновенным акциям, а также о доходности капитала компании:

№	Цена акции, $ США	Доходность дивидендов, %	Уровень дивидендов, %
1	25	15.2	2.6
2	20	13.9	2.1
3	15	15.8	1.5
4	34	12.8	3.1
5	20	6.9	2.5
6	33	14.6	3.1
7	28	15.4	2.9
8	30	17.3	2.8
9	23	13.7	2.4
10	24	12.7	2.4

Вариант 6. Статистические данные рынка однокомнатных квартир

№	Цена квартиры (руб.)	Общая площадь ($м^2$)	Удаленность от метро мин. пешком
1	4392500	36,3	40
2	6272000	44	15
3	3797500	37	40
4	5500000	39,5	10
5	4150000	37,6	10
6	5000000	40	15
7	4900000	39	5
8	4750000	39	1
9	4750000	39	5
10	3773000	32,6	2
11	4410000	37,6	10
12	5145000	32	2
13	4532500	37	7

Вариант 7. Имеются следующие данные.

№ п/п год	Объем сбыта, шт.	Цена товара, тыс. долл.	Цена конкурента, тыс. долл.
1	197280	20	21,1
2	160300	20,2	21
3	167400	20,6	21,5
4	155000	20,8	21,6
5	153200	20,6	21,5
6	178000	20,6	21,5
7	180000	20,5	21,6
8	162200	20,4	21,4
9	170000	20,3	21,4
10	157000	20,2	21,4
11	169200	20,3	21,2
12	157000	20,8	21,2

Вариант 8. По выборке из 10 почтовых отправлений изучается зависимость стоимости отправки корреспонденции экспресс-почтой от веса конверта и дальности перевозки:

№	Стоимость доставки, тыс. руб.	Вес конверта, г	Дальность перевозки, тыс. км
1	26	590	0,5
2	39	320	1,5
3	80	440	2
4	92	660	1,6
5	44	75	2,8
6	15	70	0,8
7	145	650	2,4
8	19	450	0,5
9	10	60	1
10	140	750	1,8

Вариант 9. Имеются следующие данные.

№	Производит. труда, млн. руб. на 1 рабочего	Энерговооруженность квт-час на 1 рабочего	Доля рабочих ручного труда в общей числ. рабочих (%)
1	9,8	4,8	40
2	6,7	2,8	59
3	12,4	7	38
4	6,9	3,8	57
5	11,8	5,5	31
6	7,3	3	56
7	8,4	3,4	45
8	10,7	5,2	35
9	11,1	5,4	32
10	7,3	2,9	54

Вариант 10. Имеются следующие данные.

Цена за 1 кв.м в, долл.	Площадь офиса, кв.м	Удаленность от метро, мин. пешком
4189	931	10
2800	300	15
2150	2260	10
3000	450	10
1500	1000	10
3404	1880	5
3100	300	5
2500	2900	10
3660	71	10
2500	1891	10

Вариант 11. Имеются следующие данные.

Цена, долл.	Пробег, км	Возраст, лет
5750	170000	9
5900	138000	10
6000	200000	10
6200	150000	10
6300	150000	10
6400	157000	9
6450	134000	9
6550	134000	9
6700	170000	8
6800	122000	7

Вариант 12. Менеджер новой чебуречной не уверен в правильности выбранной цены на чебуреки, поэтому в течение 10 недель он варьирует цену и записывает количество проданных чебуреков. После финансового кризиса спрос на чебуреки упал, и менеджер был вынужден тратить часть своих средств на рекламу. Для изучения зависимости объема продаж от цены и расходов на рекламу менеджер использует парную линейную регрессионную модель.

Полученные данные приведены в таблице (q_t – количество проданных чебуреков, p_t – цена на чебуреки (руб.), t – номер недели, a_t – затраты на рекламу (100 руб.)):

t	q_t	p_t	a_t
1	525	5,92	4,79
2	567	6,50	3,61
3	396	6,54	5,49
4	726	6,11	2,78
5	265	6,62	5,74
6	615	5,15	1,34
7	370	5,02	5,81
8	789	5,02	3,39
9	513	6,77	3,74
10	661	5,57	3,59

Вариант 13. Получены данные для предприятий машиностроения:

№	Рентабельность (прибыль в $ стоимость основных и оборотных фондов)	Произв. труда, млн. руб. на 1 работника	Средний возраст работающих на предприятии, лет
1	7	7	30
2	8	10	39
3	7	9	45
4	9	11	36
5	9	11	40
6	8	11	50
7	11	17	38
8	11	14	30
9	16	13	29
10	14	12	47

Вариант 14. В кейнсианской теории спрос на деньги зависит от доходов и процентных ставок. Рассмотрим следующую модель: $m_t = b_0 + b_1 y_t + b_2 i_t + \varepsilon_t$, где m_t – агрегат денежной массы (млрд. долл), y_t – валовой национальный продукт (ВНП) (млрд. долл.), i_t процентные ставки по 6-месячным гос. облигациям США. Данные в таблице:

t	y_t	m_t	i_t
1961	565,0	149,2	2,908
1962	637,7	161,8	3,686
1963	756,0	173,7	5,082
1964	873,4	199,4	5,470
1965	992,7	216,5	6,562
1966	1185,9	251,9	4,466
1967	1434,2	277,5	7,926
1968	1718,0	310,4	5,266
1969	2163,9	363,2	7,572
1970	2631,7	414,1	11,374

Вариант 15. В таблице заданы 3 временных ряда: первый из них представляет нарастающую прибыль по кварталам коммерческого банка y_t, второй и третий ряд – процентные ставки этого банка по кредитованию юридических лиц x_{1t} и депозитным вкладам x_{2t} за этот же период:

№ набл.	1	2	3	4	5	6	7	8	9	10
y_t	15	20	22	14	25	28	25	28	30	31
x_{1t}	32	34	41	38	42	48	50	52	54	51
x_{2t}	32	28	26	24	25	23	19	27	22	20

Вариант 16. Рассмотреть зависимость перевозок пассажиров железнодорожным транспортом в зависимости от перевозок на автобусах и метро (млн. чел.).

Год	Железная дорога	Автобус	Метро
1970	2500	15053	2047
1980	2971	23356	3036
1990	3143	28626	3695
1995	1833	22817	4150
2000	1419	22033	4186
2001	1306	20883	4205
2002	1271	19620	4200
2003	1704	17898	4205
2004	1335	16552	4211
2005	1339	11297	3574

Вариант 17. Имеются следующие данные

Заработная плата, долл/год	Пол	Возраст, лет	Стаж работы, лет
32368	ж	42	3
53174	м	54	10
52722	м	47	10
53423	м	47	1
50602	м	44	5
49033	м	42	10
24395	м	30	5
24395	ж	52	6
43124	м	48	8
23975	ж	58	4

Вариант 18. В таблице заданы 3 временных ряда: первый из них представляет нарастающую прибыль по кварталам коммерческого банка y_t, второй и третий ряд – процентные ставки этого банка по кредитованию юридических лиц x_{1t} и депозитным вкладам x_{2t} за этот же период:

№ набл.	1	2	3	4	5	6	7	8	9	10
y_t	110	88	78	89	82	80	76	78	76	70
x_{1t}	15	20	22	14	25	28	25	28	30	31
x_{2t}	42	47	50	48	67	57	61	59	65	54

Вариант 19. Определение стоимости офисных помещений

№	Цена 1 кв.м. в долл	Удаленность от центра города (км)	Класс офиса
1	750	13	2
2	1300	4,5	2
3	791	10,36	3
4	1884	4,3	1
5	663	8,5	4
6	895	12	2
7	2700	3	1
8	576	11	3
9	751	15	3
10	1128	6,5	3

Вариант 20. Имеются следующие данные.

№ набл.	1	2	3	4	5	6	7	8	9	10
Чистый доход, млрд долл. США, y_t	0,9	1,7	0,7	1,7	2,6	1,3	4,1	1,6	6,9	0,4
Числ. служащих, тыс. чел., x_{1t}	43	64,7	24	50,2	106	96,6	347	85,6	745	4,1
Рыночная капитализация компании, млрд. долл. США, x_{2t}	40,9	40,5	38,9	38,5	37,3	26,5	37	36,8	36,3	35,3

Вариант 21. Имеются следующие данные.

№ набл.	1	2	3	4	5	6	7	8	9	10
Чистый доход, млрд долл. США, y_t	–0,9	1,3	2	0,6	0,7	0,9	1,1	1,9	2,6	1,3
Оборот капитала, млрд долл. США, x_{1t}	12,7	21,4	13,5	13,4	4,2	15	15,5	17,9	16,5	20
Использованный капитал, млрд долл. США, x_{2t}	11,9	13,7	11,6	14,2	12,8	15	13	5,8	6	2,5

Вариант 22. Имеются следующие данные.

Цена, долл	Пробег, км	возраст, лет
7100	165000	7
7400	97000	8
7500	140000	7
7600	135000	9
7800	100000	8
7850	110500	7
7900	99000	7
7900	118000	7
7950	115000	5
8000	82000	7

Вариант 23. Имеются данные о ценах и дивидендах по обыкновенным акциям, а также о доходности капитала компании:

№	Цена акции, $США	Доходность дивидендов, %	Уровень дивидендов,%
1	25	15,2	2,6
2	20	13,9	2,1
3	15	15,8	1,5
4	34	12,9	3,1
5	20	6,9	2,5
6	33	14,6	3,1
7	28	15,4	2,9
8	30	17,3	2,8
9	23	13,7	2,4
10	24	12,7	2,4

Вариант 24. Для анализа эффективности работы предприятий машиностроения были получены следующие данные:

№	Рентабельность (прибыль в $ стоимость основных и оборотных фондов)	Производит. труда, млн. руб. на 1 раб.	Средний возраст производственного оборудования
1	7	7	20
2	8	10	19
3	7	9	21
4	9	11	17
5	9	11	16
6	8	11	18
7	11	17	15
8	11	14	14
9	16	13	10
10	15	18	10

Вариант 25. Изучается влияние изменения объема промышленного производства и среднедушевого дохода на товарооборот:

№	Розничный товарооборот (в % к предыдущему году)	Объем промышл. производства (в % к предыд. году)	Среднедушевой денежный доход (в % к предыд. году)
1	92	95	86
2	80	73	89
3	86	76	86
4	94	86	89
5	98	99	88
6	92	82	108
7	87	96	101
8	103	96	103
9	98	89	99
10	96	83	96

СПИСОК ЛИТЕРАТУРЫ

Основная

1. Эконометрика : учебник для бакалавриата и магистратуры / И. И. Елисеева [и др.] ; под ред. И. И. Елисеевой. – М. : Издательство Юрайт, 2018.

2. Эконометрика (продвинутый уровень): конспект лекций / Крянев А.В. – М.:КУРС, НИЦ ИНФРА-М, 2017.

3. Герасимов, А.Н. Эконометрика: продвинутый уровень : учебное пособие / А.Н. Герасимов, Е.И. Громов, Ю.С. Скрипниченко ; Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования Ставропольский государственный аграрный университет. – Ставрополь : Ставропольский государственный аграрный университет, 2016.

4. Невежин В.П. Практическая эконометрика в кейсах : учеб. пособие / В.П. Невежин, Ю.В. Невежин. М. : ИД «ФОРУМ» : ИНФРА-М, 2016. – 317 с. + Доп. материалы [Электронный ресурс; URL: http://www.znanium.com]. – (Высшее образование).

5. Айвазян С.А., Мхитарян В.С. Прикладная статистика и основы эконометрики. М.: ЮНИТИ,1998.

6. Айвазян С.А., Мхитарян В.С. Практикум по прикладной статистике и эконометрике. М.: МЭСИ, 2000.

7. Доугерти К. Введение в эконометрику. М.: МГУ,1999.

8. Бабешко О.Л. Основы эконометрического моделирования: учеб. пособие. М.: Комкнига, 2006.

9. Методические указания по практике для магистрантов очной и очно-заочной форм обучения / С.М. Бычкова, И.И. Костусенко, В.В. Скобара и др. ; Министерство сельского хозяйства РФ, Санкт-Петербургский государственный аграрный университет, Кафедра бухгалтерского учета и аудита. – Санкт-Петербург : СПбГАУ, 2018. – 38 с. : табл. ; То же [Электронный ресурс]. – URL: http://biblioclub.ru/index.php?page=book&id=491716 (16.11.2018).

10. Кремер Н.Ш. Эконометрика [Электронный ресурс] : учебник для студентов вузов / Н.Ш. Кремер, Б.А. Путко. –3-е изд. – Электрон. текстовые данные. – М. : ЮНИТИ-ДАНА, 2017.

11. Магнус Я.Р. и др. Эконометрика. Начальный курс. М.: Дело, 1999.

Дополнительная

1. Айвазян С.А., Мхитарян В.С. Практикум по прикладной статистике и эконометрике: учебник / Моск. гос. ун-т. экономики, статистики и информатики. М.: 1998.

2. Бородич С.А. Вводный курс эконометрики: учеб. пособие. Минск: БГУ, 2000.

3. Дуброва Т.А. Статистические методы прогнозирования. М.: ЮНИТИ-ДАНА, 2003.

4. Дуброва Т.А.,Аархипова М.Ю. Статистические методы прогнозирования в экономике: учеб. пособие, практикум, тесты, программа курса / Моск. гос. ун-т экономики, статистики и информатики. М., 2004. 136 с.

5. Дуброва Т.А., Бакуменко Л.П. и др. Анализ временных рядов и прогнозирование в системе «Statistica». М.: МЭСИ, 2002.

6. Канторович Г.Г. Эконометрика // Методические материалы по экономическим дисциплинам для преподавателей средних школ и вузов. Экономическая статистика. Эконометрика: программы, тесты, задачи, решения / под ред. Л.С. Гребнева. М.: ГУ-ВШЭ, 2000.

7. Катышев П.К. и др. Сборник задач к начальному курсу эконометрики. М.: Дело, 2002.

8. Коломак Е.А. Эконометрический анализ панельных данных. Новосибирск: НГУ, 2007.

9. Кулинич Е.И. Эконометрия. М.: Финансы и статистика, 2001.

10. Лукашин Ю.П. Адаптивные методы краткосрочного прогнозирования временных рядов. М.: Финансы и статистика, 2003.

11. Мардас А.Н. Эконометрика. СПб.: Питер, 2001.

12. Носко В.П. Эконометрика для начинающих (Дополнительные главы). М.: ИЭПП, 2005. 379с.

13. Практикум по эконометрике / под ред. Н.И. Елисеевой. М.: Финансы и статистика, 2001.

14. Радионова М.В., Чичагов В.В. Руководство к решению задач по эконометрике. Пермь, 2008.

15. Радионова М.В., Фролова Н.В., Чичагов В.В. Эконометрика. Учебное пособие. Пермь, 2008, 2015.

16. Ратникова Т.А. Введение в анализ панельных данных, ГУ-ВШЭ, 2004.

17. Сборник задач по эконометрике: учеб. пособие для студ. экон. вузов / сост. Е.Ю. Дорохина, Л.Ф. Преснякова, Н.П. Тихомиров. М.: Экзамен, 2003.

18. Статистическое моделирование и прогнозирование: учеб. пособие / под ред. А. Г. Гранберга. М.: Финансы и статистика, 1990.

19. Уотшем Т.Дж., Паррамоу К. Количественные методы в финансах / пер. с англ.; под. ред. М.Р. Ефимовой. М.: Финансы, ЮНИТИ, 1999.

20. Четыркин Е. Н. Статистические методы прогнозирования. М.: Статистика, 1975.

21. Эконометрика: учеб. пособие для вузов / А.И. Орлов. М.: Экзамен, 2002.

22. Эконометрика: учебник / Е.Ю.Дорохина, Н.П. Тихомиров. М.: Экзамен, 2003.

23. Экономико-математические методы и прикладные модели / под ред. В.В. Федосеева. М.: ЮНИТИ,1999.

24. Baltagi, B.H. Econometric Analysis of Panel Data (3rd edition). New York: John Wiley & Sons, Ltd., 2005. 302p.

25. Greene W.H. Econometric Analysis. 4th ed. Prentice Hall, 1999.

26. Pindyck R. S., Rubinfeld D. L. Econometric models. Economic forecasts. 4th ed. McGraw-Hill, 1998.

КОРРЕЛЯЦИОННЫЙ И РЕГРЕССИОННЫЙ АНАЛИЗ С ИСПОЛЬЗОВАНИЕМ ППП MICROSOFT EXCEL

Для работы со статистическими данными можно воспользоваться в Microsoft Excel встроенным *Пакетом анализа*. Если этот пункт в данном меню *Сервис* отсутствует, необходимо его включить, используя пункт меню *Надстройки* (*рис*. 1):

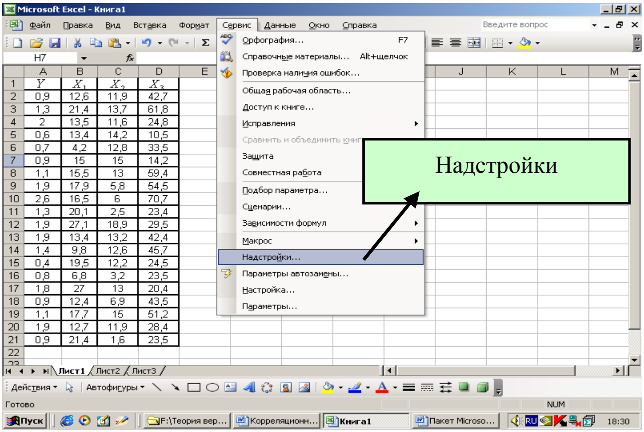


Рис. 1

Поставьте флажок в *Надстройках* в строке *Пакет анализа* (*рис*. 2), нажмите клавишу *ОК*.

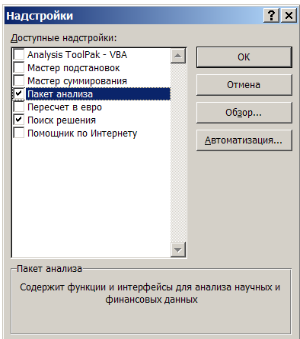


Рис. 2

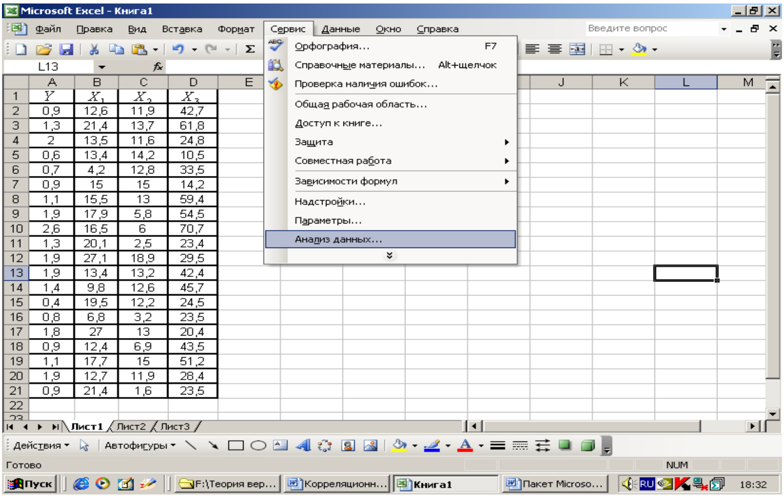


Рис. 3

После активизации *Анализ данных* выберите модуль *Описательная статистика* (*рис*. 4), нажмите кнопку *ОК*.

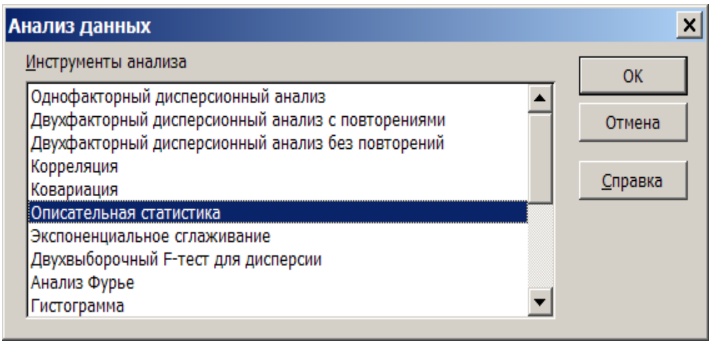


Рис. 4

После этого появится диалоговое окно *Описательная статистика*, в котором нужно определить *Входной интервал*, поставить флажок *Метки в первой строке*, если вместе с входными данными были выделены заголовки, задать *параметры вывода* (*рис*. 5).

В параметрах вывода по умолчанию указано, что результаты будут размещаться на *Новом рабочем листе*. Поставьте флажок в строке *Итоговая статистика.* Затем нажмите кнопку *ОК*.

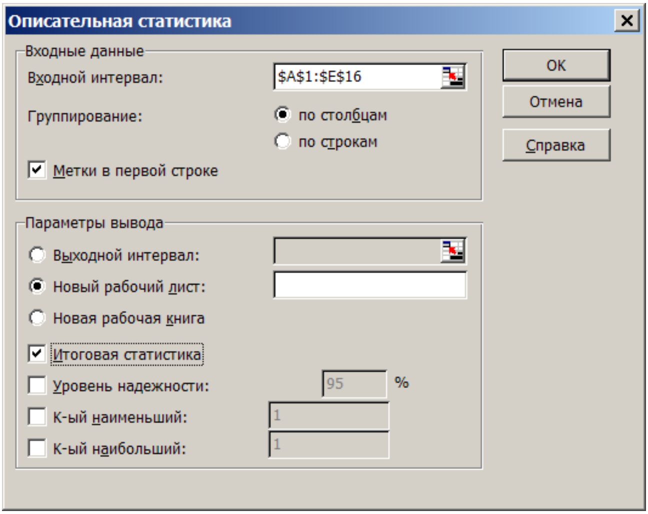


Рис. 5

Рассмотрим более подробно элементы Диалогового окна *Описательная статистика* и операции в этом окне. Окно содержит (*рис*. 5):

1) *Входной диапазон*. Ссылка на диапазон, содержащий анализируемые данные. Ссылка должна состоять не менее чем из двух смежных диапазонов данных, данные в которых расположены по строкам или столбцам.

2) *Группирование*. Установите переключатель в положение *По столбцам* или *По строкам* в зависимости от расположения данных во входном диапазоне.

3) *Метки в первой строке/Метки в первом столбце*. Если первая строка исходного диапазона содержит названия столбцов, установите переключатель в положение *Метки в первой строке*. Если названия строк находятся в первом столбце входного диапазона, установите переключатель в положение *Метки в первом столб-*

це. Если входной диапазон не содержит меток, то необходимые заголовки в выходном диапазоне будут созданы автоматически.

4) *Уровень надежности*. Установите флажок, если в выходную таблицу необходимо включить строку для уровня надежности. В поле введите требуемое значение. Например, значение 95% вычисляет уровень надежности среднего со значимостью 0.05.

5) *К-й наибольший*. Установите флажок, если в выходную таблицу необходимо включить строку для k-го наибольшего значения для каждого диапазона данных. В соответствующем окне введите число k. Если k равно 1, эта строка будет содержать максимум из набора данных.

6) *К-й наименьший*. Установите флажок, если в выходную таблицу необходимо включить строку для k-го наименьшего значения для каждого диапазона данных. В соответствующем окне введите число k. Если k равно 1, эта строка будет содержать минимум из набора данных.

7) *Выходной диапазон*. Введите ссылку на левую верхнюю ячейку выходного диапазона. Этот инструмент анализа выводит два столбца сведений для каждого набора данных. Левый столбец содержит метки статистических данных; правый столбец содержит статистические данные. Состоящий из двух столбцов диапазон статистических данных будет выведен для каждого столбца или для каждой строки входного диапазона в зависимости от положения переключателя Группирование.

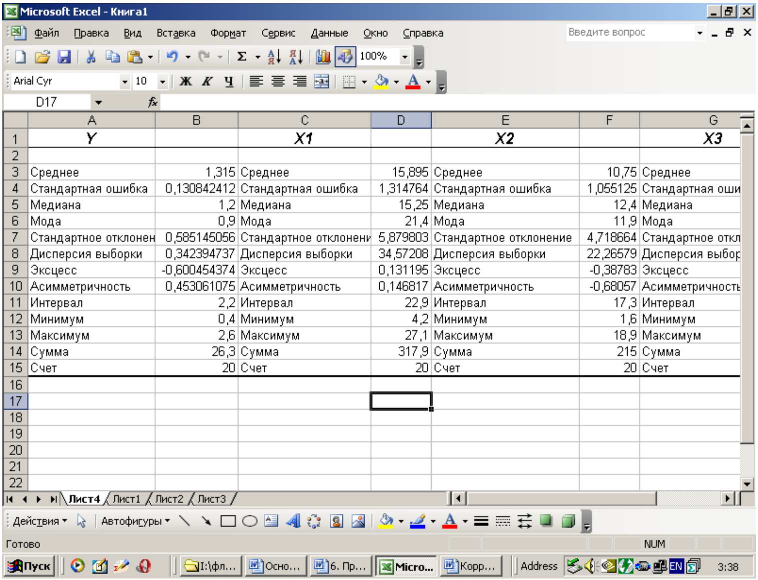
8) *Новый лист*. Установите переключатель, чтобы открыть новый лист в книге и вставить результаты анализа, начиная с ячейки *A*1. Если в этом есть необходимость, введите имя нового листа в поле, расположенном напротив соответствующего положения переключателя.

9) *Новая книга*. Установите переключатель, чтобы открыть новую книгу и вставить результаты анализа в ячейку *A*1 на первом листе в этой книге.

10) *Итоговая статистика*. Установите флажок, если в выходном диапазоне необходимо получить по одному полю для каждого из следующих видов статистических данных: Среднее, Стан-

дартная ошибка (среднего), Медиана, Мода, Стандартное отклонение, Дисперсия выборки, Эксцесс, Асимметричность, Интервал, Минимум, Максимум, Сумма, Счет, Наибольшее (#), Наименьшее (#), Уровень надежности.

В результате Вы получите статистические показатели, такие как среднее, дисперсия, среднее квадратичное отклонение и т.д. (*рис*. 6).



	A	B	C	D	E	F	G
1	*Y*		*X1*		*X2*		*X3*
2							
3	Среднее	1,315	Среднее	15,895	Среднее	10,75	Среднее
4	Стандартная ошибка	0,130842412	Стандартная ошибка	1,314764	Стандартная ошибка	1,055125	Стандартная оши
5	Медиана	1,2	Медиана	15,25	Медиана	12,4	Медиана
6	Мода	0,9	Мода	21,4	Мода	11,9	Мода
7	Стандартное отклонен	0,585145056	Стандартное отклонени	5,879803	Стандартное отклонение	4,718664	Стандартное откл
8	Дисперсия выборки	0,342394737	Дисперсия выборки	34,57208	Дисперсия выборки	22,26579	Дисперсия выбор
9	Эксцесс	-0,600454374	Эксцесс	0,131195	Эксцесс	-0,38783	Эксцесс
10	Асимметричность	0,453061075	Асимметричность	0,146817	Асимметричность	-0,68057	Асимметричность
11	Интервал	2,2	Интервал	22,9	Интервал	17,3	Интервал
12	Минимум	0,4	Минимум	4,2	Минимум	1,6	Минимум
13	Максимум	2,6	Максимум	27,1	Максимум	18,9	Максимум
14	Сумма	26,3	Сумма	317,9	Сумма	215	Сумма
15	Счет	20	Счет	20	Счет	20	Счет

Рис. 6

ОПРЕДЕЛЕНИЕ ОПИСАТЕЛЬНЫХ СТАТИСТИК

Рассмотрим определения описательных статистик для их дальнейшего использования. Наряду со средними величинами в качестве описательных характеристик выборочных данных применяют моду и медиану.

Медианой называют значение признака, приходящееся на середину ранжированного ряда наблюдений.

Модой называют такое значение признака, которое наблюдалось наибольшее число раз.

Средние величины, характеризующие ряд числом, не отражают изменчивости наблюдавшихся значений признака, т.е. вариацию. Простейший показатель вариации – *вариационный размах*, равный разности между наибольшим и наименьшим значениями в ряду. Вариационный размах – приближенный показатель вариации, так как не зависит от изменения вариантов, а крайние варианты, которые используются для его вычисления, как правило, ненадежны. Правильнее использовать меру рассеяния наблюдений вокруг средних величин: *это выборочные дисперсия и среднеквадратическое отклонение.*

Выборочная дисперсия находится по формуле

$$S^2 = \sum_{i=1}^{n}(x_i - \overline{x})^2 / n,$$

а *выборочное среднее квадратическое отклонение* равно арифметическому значению корня квадратного из дисперсии и имеет ту же размерность, что и значение признака.

Выборочным *коэффициентом асимметрии* называют отношение центрального момента третьего порядка к кубу среднеквадратического отклонения. Если в выборочном ряду преобладают варианты меньше $\overline{x}$, то выборочный коэффициент асимметрии отрицателен, в этом случае имеет место *левосторонняя асимметрия*.

Если преобладают варианты, большие $\overline{x}$, то выборочный коэффициент асимметрии положителен, в этом случае имеет место *правосторонняя асимметрия.* Выборочный коэффициент асимметрии не имеет верхней или нижней границы, что снижает его ценность как меры асимметрии. Практически коэффициент асимметрии редко бывает особенно велик, а для умеренно асимметричных рядов он обычно меньше 1.

Выборочный эксцесс – это уменьшенное на 3 единицы отношение центрального момента четвертого порядка к четвертой степени среднеквадратического отклонения. За стандартное значение эксцесса принимают нуль-эксцесс нормальной кривой. Кривые, у которых эксцесс отрицателен по сравнению с нормальной кривой, менее крутые, имеют более плоскую вершину и называются плосковершинными.

Рассмотрим отличие функции распределения данного ряда наблюдений от нормального распределения.

1. Коэффициенты асимметрии и эксцесса для нормального распределения равны 0. Коэффициент асимметрии характеризует скошенность распределения по отношению к математическому ожиданию.

2. Наряду со средним значением в качестве показателя центра группирования используется медиана. Для нормального распределения медиана и среднее совпадают.

3. Для нормального распределения мода и среднее должны совпадать.

ДЛЯ ПОСТРОЕНИЯ ПОЛЯ КОРРЕЛЯЦИИ можно воспользоваться мастером ДИАГРАММ.

Сначала выделите курсором исходные данные, для которых будете строить *Поле корреляции* (*рис.*7).

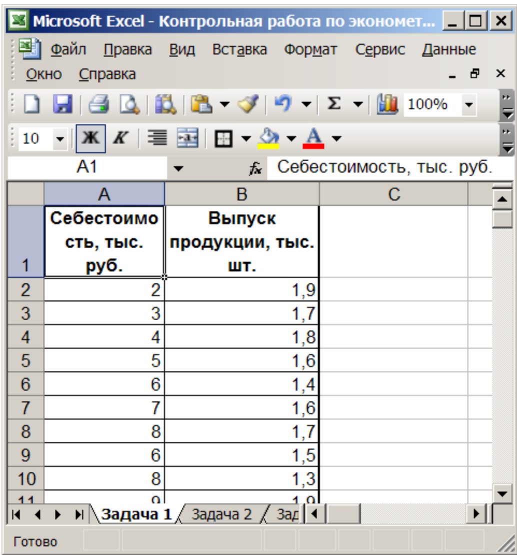


Рис. 7

Затем выберите *ВСТАВКА Диаграмма* или на панели инструментов воспользуйтесь значком *Мастер диаграмм* После этого появится диалоговое окно *Мастер диаграмм (Шаг 1 из 4)* (*рис*.8). Выберите *Точечную Диаграмму* и нажмите *Далее*.

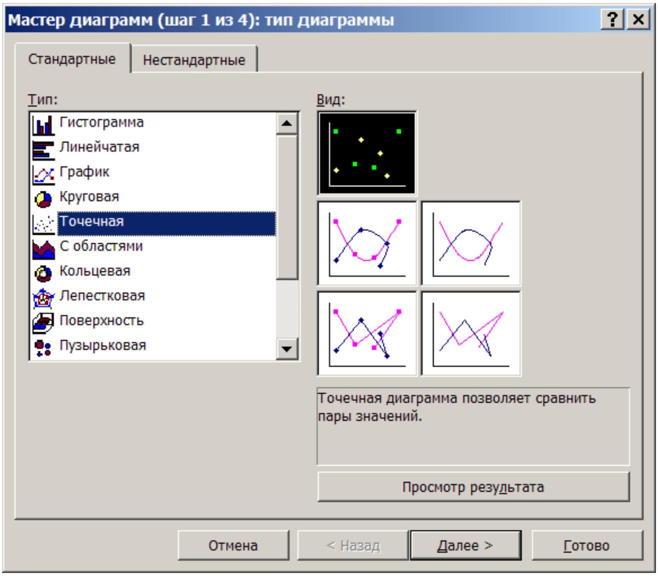


Рис. 8

Если перед вызовом *Мастера диаграмм* Вы выделили *Диапазон исходных данных*, для которого хотите построить диаграмму, то в следующем диалоговом окне *Мастер диаграмм (Шаг 2 из 4)* Вы увидите *Поле корреляции для выделенных исходных данных* (*рис.* 9). На *рис*.7 выделенные ряды исходных данных *Себестоимость* и *Выпуск продукции*.

Если исходные данные не были выделены перед началом работы *Мастера диаграмм*, то в следующем диалоговом окне *Мастер диаграмм (Шаг 2 из 4)* в пунктах меню *Диапазон данных* или *Ряды*, можно задать исходные данные. Если исходные данные располагаются в смежных столбцах, то следует воспользоваться пунктом меню *Диапазон данных*.

Если данные для построения графика находятся не в смежных столбцах, выберите меню *Ряды* и, установив курсор сначала в окно *Значения X,* выделите курсором мыши *на рабочем листе* соответствующие значения *X*, а затем*,* установив курсор в окно *Значения Y*, выделите на *рабочем листе* соответствующие значения *Y* (*рис*.9).

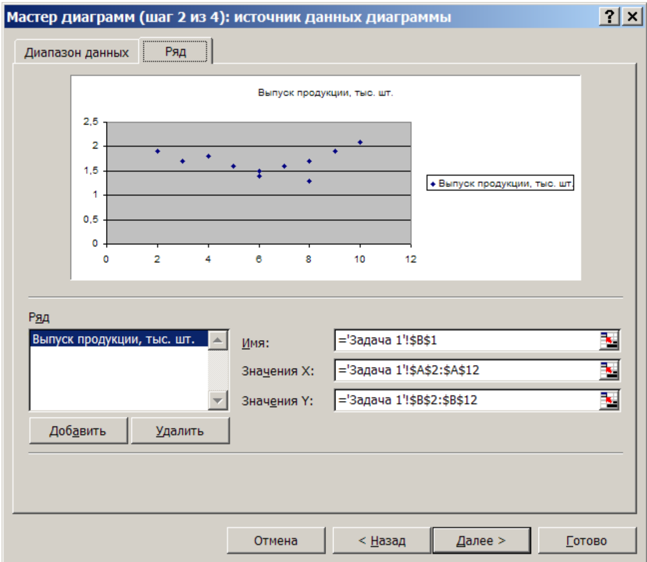


Рис. 9

После чего нажмите кнопку *Далее.* Появится диалоговое окно *Мастер диаграмм (шаг 3 из 4*). В этом окне можно задать *Параметры диаграммы,* например подписи данных, названия для осей координат и т.д.

После шага *2* можно сразу перейти на *шаг 4,* выполнив команду *Готово*. В этом случае Вы сразу перейдете в диалоговое окно *Мастер диаграмм (шаг 4 из 4)*. В этом окне выберите месторасположение *Поля корреляции*: на листе, где находятся исходные данные (текущем), или на новом. После выполнения команды *Готово* на заданном листе будет нарисована диаграмма рассеяния переменных (*Поле корреляции*) (*рис*.10).

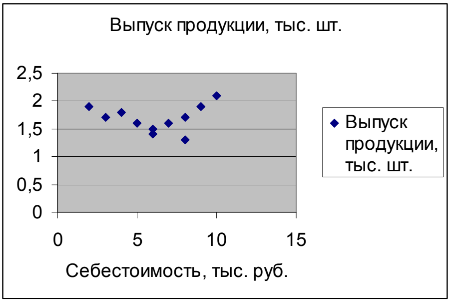


Рис. 10

В EXCEL ПАРНЫЕ КОЭФФИЦИЕНТЫ КОРРЕЛЯЦИИ МОЖНО ВЫЧИСЛИТЬ с помощью встроенных статистических функций КОРРЕЛ и ПИРСОН.

Для этого надо выбрать:

ВСТАВКА

$f(x)$ Функции

Статистические

КОРРЕЛ

или *ПИРСОН,*

отметив в качестве массивов 1 и 2 столбцы интересующих переменных.

Расчет матрицы выборочных парных коэффициентов корреляции осуществляется с помощью *Пакета анализа*. Выбираем в меню *Сервис – Анализ данных – Корреляция.*

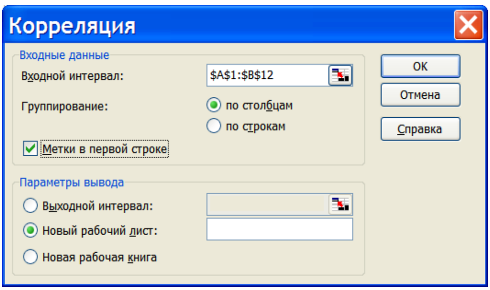


Рис. 11

Диалоговое окно «Корреляция» (рис. 11) содержит:

Входной диапазон. Необходимо указать массив исходных показателей (выделив мышкой все значения исследуемых переменных).

Группирование. В зависимости от расположения данных (в нашем случае *по столбцам*) необходимо установить переключатель в положение *По столбцам или По строкам.*

Метки в первой строке/Метки в первом столбце. Если первая строка исходного диапазона данных содержит названия столбцов, установите переключатель. Если входной диапазон не содержит меток, то необходимые заголовки в выходном диапазоне будут созданы автоматически.

Выходной интервал. Можно указать ссылку на левую верхнюю ячейку выходного диапазона.

Новый лист. Если в этом есть необходимость, то можно установить переключатель, чтобы открыть новый лист в книге и вставить результаты анализа, начиная с ячейка A1.

Рассмотрим пример, в котором исследуется линейная зависимость между переменными Y, X_1, X_2, X_3 (*рис.*12).

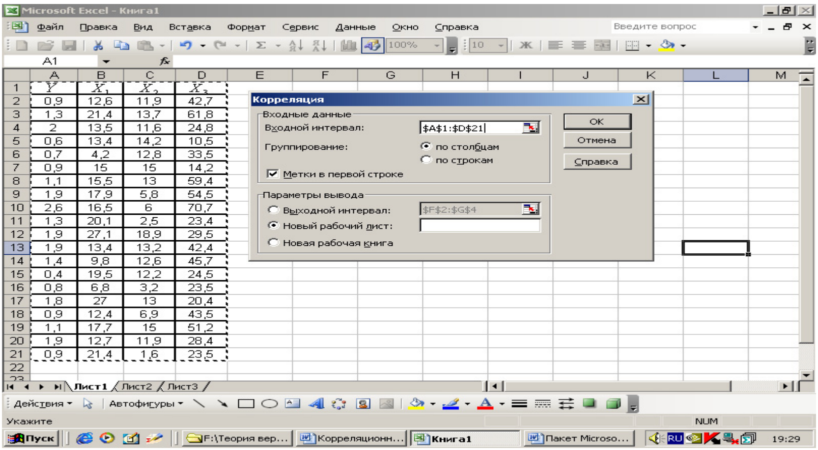


Рис. 12

После нажатия на кнопку «ОК» на новом рабочем листе появится корреляционная матрица (рис. 13).

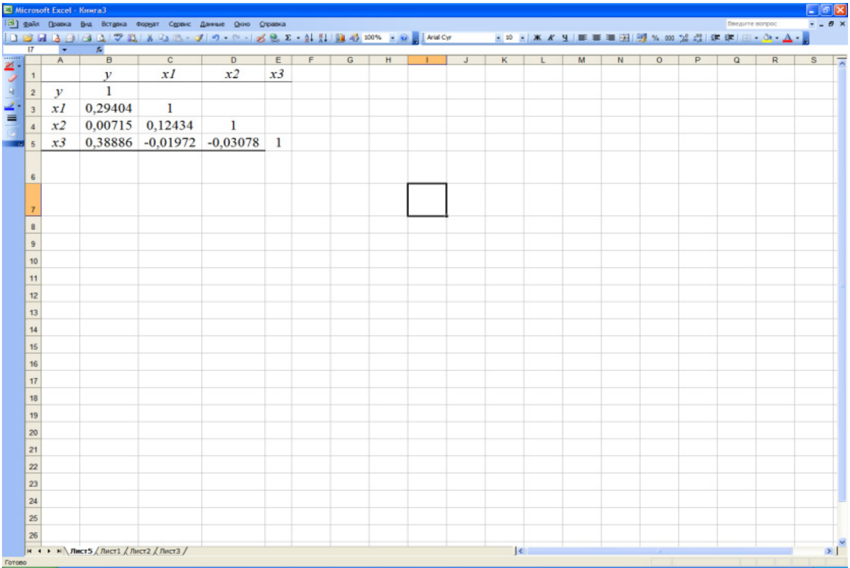


Рис. 13

Эту матрицу можно записать в виде

r_*	Y	X1	X2	X3
Y	1			
X1	0,46250	1		
X2	0,00715	0,12434	1	
X3	0,56890	-0,01972	-0,03078	1

Если в настройке корреляционного анализа не был поставлен флажок в *Метки в первой строке*, то матрица будет выведена без названий переменных и их необходимо ввести вместо номеров столбцов:

r_*	*Столбец 1*	*Столбец 2*	*Столбец 3*	*Столбец 4*
Столбец 1	1			
Столбец 2	0,46250	1		
Столбец 3	0,00715	0,12434	1	
Столбец 4	0,56890	-0,01972	-0,03078	1

Для построения регрессионной зависимости между переменными воспользуйтесь модулем «Регрессия».

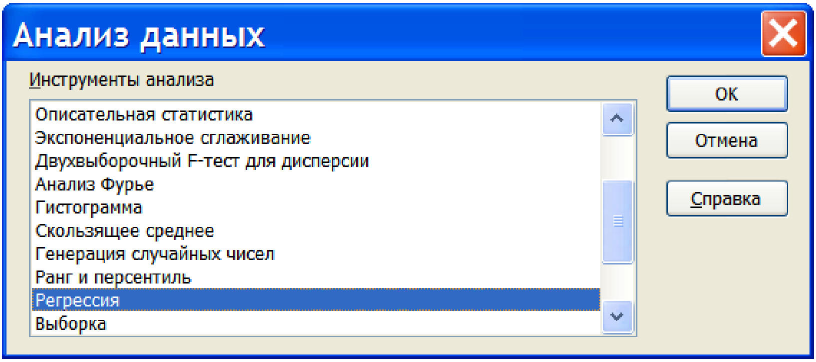


Рис. 14

После активизации *Регрессии* в окне *Анализ данных* появится соответствующее диалоговое окно *Регрессия*, в котором нужно задать *Входные данные*, *Метки*, *Параметры вывода* и *Остатки* (*рис*.15).

Элементы диалогового окна *Регрессия* содержат (*рис*. 16):

Входной интервал Y. Введите ссылку на диапазон анализируемых зависимых данных. Диапазон должен состоять из одного столбца.

Входной интервал X. Введите ссылку на диапазон независимых данных, подлежащих анализу. Microsoft Excel располагает независимые переменные этого диапазона слева направо в порядке возрастания. Максимальное число входных диапазонов равно 16.

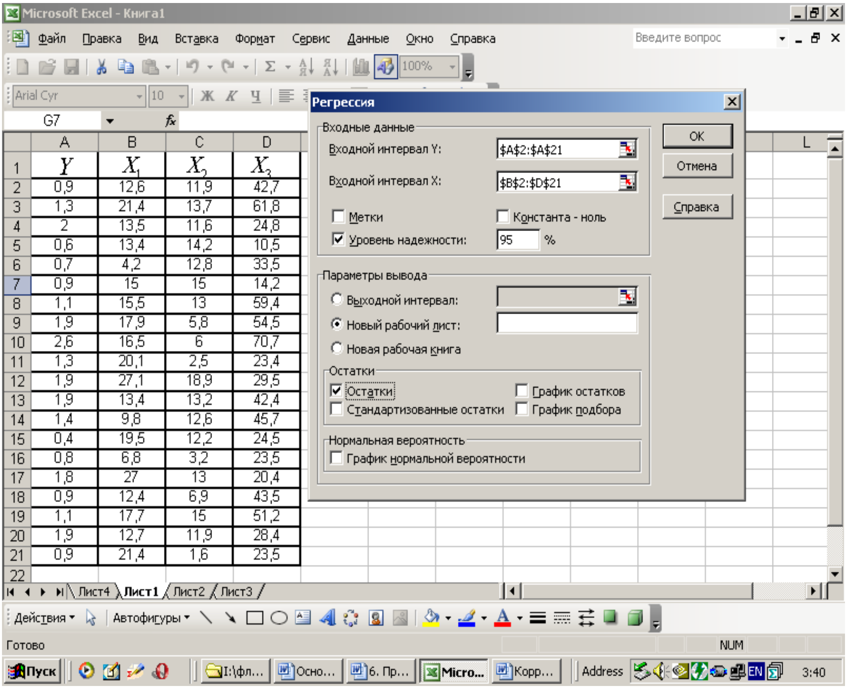


Рис. 15

Заголовки. Установите флажок, если первая строка или первый столбец входного интервала содержит заголовки. Снимите флажок, если заголовки отсутствуют; в этом случае подходящие названия для данных выходного диапазона будут созданы автоматически.

Уровень надежности. Установите флажок, чтобы включить в выходной диапазон дополнительный уровень. В соответствующее поле введите уровень надежности, который будет использован дополнительно к уровню 95%, применяемому по умолчанию.

Константа – ноль. Установите флажок, чтобы линия регрессии прошла через начало координат.

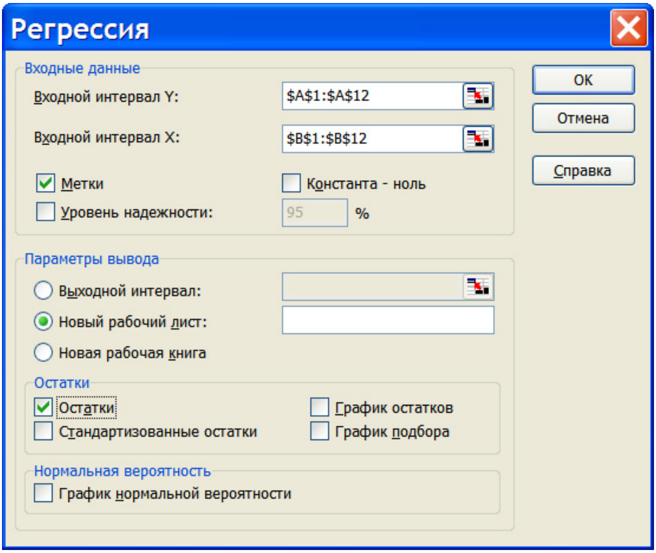


Рис. 16

Выходной диапазон. Введите ссылку на левую верхнюю ячейку выходного диапазона. Отведите, по крайней мере, семь столбцов для итогового диапазона, который будет включать в себя: результаты дисперсионного анализа, коэффициенты регрессии, стандартную погрешность вычисления Y, среднеквадратичные отклонения, число наблюдений, стандартные погрешности для коэффициентов.

Новый лист. Установите переключатель, чтобы открыть новый лист в книге и вставить результаты анализа, начиная с ячейки A1. Если в этом есть необходимость, введите имя нового листа в поле, расположенном напротив соответствующего положения переключателя.

Новая книга. Установите переключатель, чтобы открыть новую книгу и вставить результаты анализа в ячейку A1 на первом листе в этой книге.

Остатки. Установите флажок, чтобы включить остатки в выходной диапазон.

Стандартизированные остатки. Установите флажок, чтобы включить стандартизированные остатки в выходной диапазон.

График остатков. Установите флажок, чтобы построить диаграмму остатков для каждой независимой переменной.

График подбора. Установите флажок, чтобы построить диаграммы наблюдаемых и предсказанных значений для каждой независимой переменной.

График нормальной вероятности. Установите флажок, чтобы построить диаграмму нормальной вероятности.

В результате Вы получите следующее решение:

ВЫВОД ИТОГОВ

Регрессионная статистика		
Множественный R	0,49	
R-квадрат	0,24	Коэффициент детерминации
Нормированный R-квадрат	0,10	
Стандартная ошибка	0,55	Среднеквадратичное отклонение = корень из оценки дисперсии модели
Наблюдения	20	Объем выборки

Дисперсионный анализ

	df	*SS*	*MS*	*F*	*Значимость F*
Регрессия	3,00	1,58	0,53	1,71	0,21
Остаток	16,00	4,93	0,31	Проверка гипотезы о значимости модели	
Итого	19,00	6,51			

	Коэффициенты	*Стандартная ошибка*	*t-статистика*	*P-Значение*
Y-пересечение	0,36	0,53	0,68	0,50
Переменная X 1	0,03	0,02	1,39	0,18
Переменная X 2	0,00	0,03	-0,08	0,93
Переменная X 3	0,01	0,01	1,81	0,09

Нижние 95,0%	*Верхние 95,0%*
-0,76	1,48
-0,02	0,08
-0,06	0,06
0,00	0,03

ВЫВОД
ОСТАТКА

Наблюдение	*Предсказанное Y*	*Остатки*
1	1,30	-0,40
2	1,82	-0,52
3	1,08	0,92
4	0,88	-0,28
5	0,92	-0,22
6	0,98	-0,08
7	1,61	-0,51
8	1,63	0,27
9	1,81	0,79
10	1,28	0,02
11	1,54	0,36
12	1,32	0,58
13	1,25	0,15
14	1,26	-0,86
15	0,88	-0,08
16	1,43	0,37
17	1,31	-0,41
18	1,56	-0,46
19	1,11	0,79
20	1,33	-0,43

ДЛЯ ПОСТРОЕНИЯ ПРОГНОЗА ЗАВИСИМОЙ ПЕРЕМЕННОЙ Y необходимо в массиве исходных данных добавить строку для значений независимых переменных X_1, X_2, X_3, для которых Вы хотите найти прогнозное значение *Y*. В нашем примере – это 22-я *строка* (*рис*.17). И затем активируйте функцию *ТЕНДЕНЦИЯ*.

Использование функции *ТЕНДЕНЦИЯ* для определения точечного прогноза по линейной функции. Для этого надо выбрать:

ВСТАВКА $f(x)$ Функции

Статистические

ТЕНДЕНЦИЯ,

отметив в качестве массивов столбцы интересующих переменных.

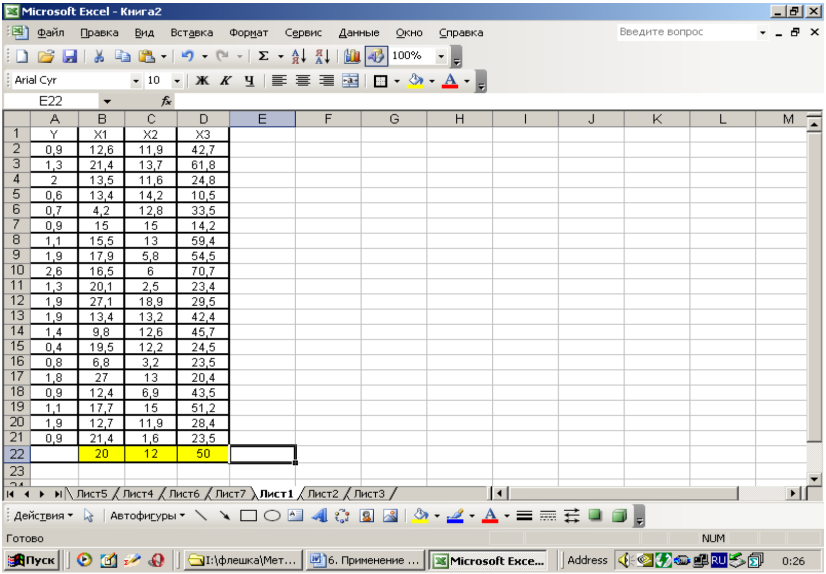

Microsoft Excel - Книга2
Файл Правка Вид Вставка Формат Сервис Данные Окно Справка
E22
Y X1 X2 X3
0,9 12,6 11,9 42,7
1,3 21,4 13,7 61,8
2 13,5 11,6 24,8
0,6 13,4 14,2 10,5
0,7 4,2 12,8 33,5
0,9 15 15 14,2
1,1 15,5 13 59,4
1,9 17,9 5,8 54,5
2,6 16,5 6 70,7
1,3 20,1 2,5 23,4
1,9 27,1 18,9 29,5
1,9 13,4 13,2 42,4
1,4 9,8 12,6 45,7
0,4 19,5 12,2 24,5
0,8 6,8 3,2 23,5
1,8 27 13 20,4
0,9 12,4 6,9 43,5
1,1 17,7 15 51,2
1,9 12,7 11,9 28,4
0,9 21,4 1,6 23,5
20 12 50
Лист5 Лист4 Лист6 Лист7 Лист1 Лист2 Лист3
Готово

Рис. 17

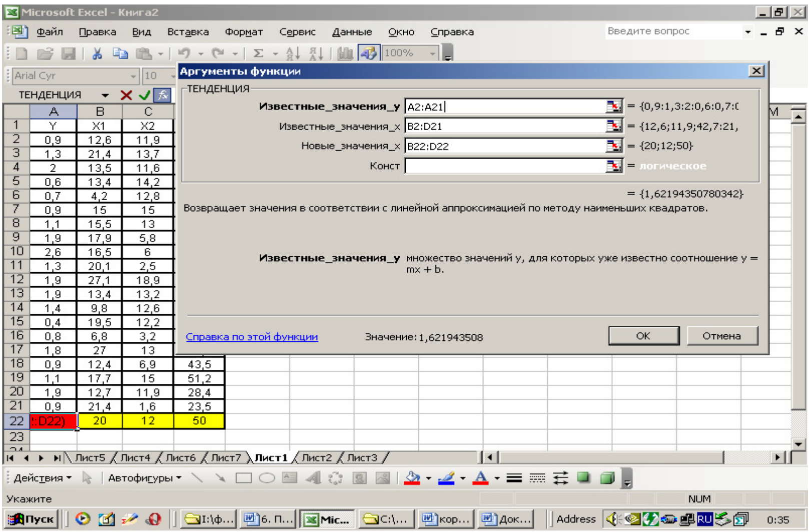
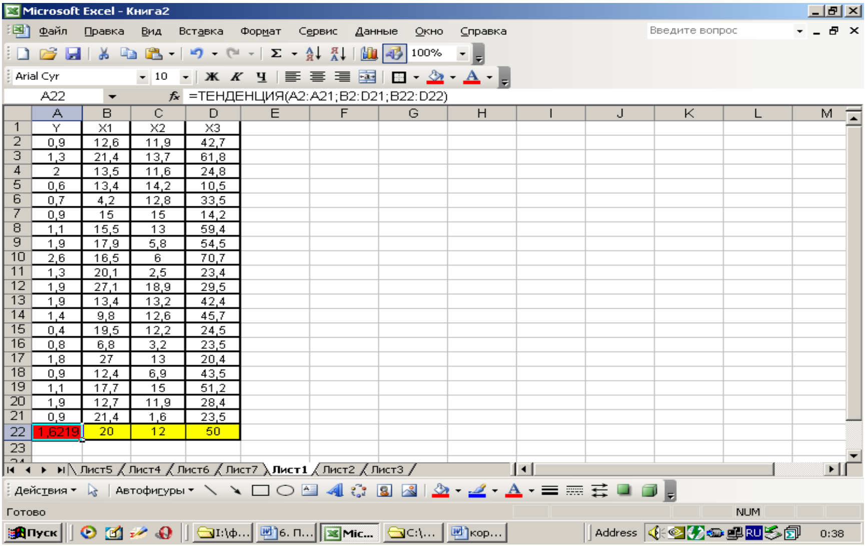

Microsoft Excel - Книга2
Аргументы функции
ТЕНДЕНЦИЯ
Известные_значения_y A2:A21 = {0,9:1,3:2:0,6:0,7:(
Известные_значения_x B2:D21 = {12,6;11,9;42,7:21,
Новые_значения_x B22:D22 = {20;12;50}
Конст = логическое
= {1,62194350780342}
Возвращает значения в соответствии с линейной аппроксимацией по методу наименьших квадратов.
Известные_значения_y множество значений y, для которых уже известно соотношение y = mx + b.
Справка по этой функции
Значение: 1,621943508
ОК
Отмена

Рис. 18

В результате получим прогнозное значение переменной Y. В примере результат помещен в ячейку 22A, и при прогнозных значений $X_1 = 20,\ X_2 = 12,\ X_3 = 50$, прогнозное значение Y будет равно 1,6219.



	A	B	C	D
1	Y	X1	X2	X3
2	0,9	12,6	11,9	42,7
3	1,3	21,4	13,7	61,8
4	2	13,5	11,6	24,8
5	0,6	13,4	14,2	10,5
6	0,7	4,2	12,8	33,5
7	0,9	15	15	14,2
8	1,1	15,5	13	59,4
9	1,9	17,9	5,8	54,5
10	2,6	16,5	6	70,7
11	1,3	20,1	2,5	23,4
12	1,9	27,1	18,9	29,5
13	1,9	13,4	13,2	42,4
14	1,4	9,8	12,6	45,7
15	0,4	19,5	12,2	24,5
16	0,8	6,8	3,2	23,5
17	1,8	27	13	20,4
18	0,9	12,4	6,9	43,5
19	1,1	17,7	15	51,2
20	1,9	12,7	11,9	28,4
21	0,9	21,4	1,6	23,5
22	1,6219	20	12	50

Рис. 19

ПРИЛОЖЕНИЕ А

СТАТИСТИЧЕСКИЕ ТАБЛИЦЫ

Таблица А1

Функция плотности стандартного нормального распределения $\varphi(x)=\dfrac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$

x	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
0,0	0,3989	0,3989	0,3989	0,3988	0,3986	0,3984	0,3982	0,3980	0,3977	0,3973
0,1	0,3970	0,3965	0,3961	0,3956	0,3951	0,3945	0,3939	0,3932	0,3925	0,3918
0,2	0,3910	0,3902	0,3894	0,3885	0,3876	0,3867	0,3857	0,3847	0,3836	0,3825
0,3	0,3814	0,3802	0,3790	0,3778	0,3765	0,3752	0,3739	0,3725	0,3712	0,3697
0,4	0,3683	0,3668	0,3653	0,3637	0,3621	0,3605	0,3589	0,3572	0,3555	0,3538
0,5	0,3521	0,3503	0,3485	0,3467	0,3448	0,3429	0,3410	0,3391	0,3372	0,3352
0,6	0,3332	0,3312	0,3292	0,3271	0,3251	0,3230	0,3209	0,3187	0,3166	0,3144
0,7	0,3123	0,3101	0,3079	0,3056	0,3034	0,3011	0,2989	0,2966	0,2943	0,2920
0,8	0,2897	0,2874	0,2850	0,2827	0,2803	0,2780	0,2756	0,2732	0,2709	0,2685
0,9	0,2661	0,2637	0,2613	0,2589	0,2565	0,2541	0,2516	0,2492	0,2468	0,2444
1,0	0,2420	0,2396	0,2371	0,2347	0,2323	0,2299	0,2275	0,2251	0,2227	0,2203
1,1	0,2179	0,2155	0,2131	0,2107	0,2083	0,2059	0,2036	0,2012	0,1989	0,1965
1,2	0,1942	0,1919	0,1895	0,1872	0,1849	0,1826	0,1804	0,1781	0,1758	0,1736
1,3	0,1714	0,1691	0,1669	0,1647	0,1626	0,1604	0,1582	0,1561	0,1539	0,1518
1,4	0,1497	0,1476	0,1456	0,1435	0,1415	0,1394	0,1374	0,1354	0,1334	0,1315
1,5	0,1295	0,1276	0,1257	0,1238	0,1219	0,1200	0,1182	0,1163	0,1145	0,1127
1,6	0,1109	0,1092	0,1074	0,1057	0,1040	0,1023	0,1006	0,0989	0,0973	0,0957
1,7	0,0940	0,0925	0,0909	0,0893	0,0878	0,0863	0,0848	0,0833	0,0818	0,0804
1,8	0,0790	0,0775	0,0761	0,0748	0,0734	0,0721	0,0707	0,0694	0,0681	0,0669
1,9	0,0656	0,0644	0,0632	0,0620	0,0608	0,0596	0,0584	0,0573	0,0562	0,0551
2,0	0,0540	0,0529	0,0519	0,0508	0,0498	0,0488	0,0478	0,0468	0,0459	0,0449
2,1	0,0440	0,0431	0,0422	0,0413	0,0404	0,0396	0,0387	0,0379	0,0371	0,0363
2,2	0,0355	0,0347	0,0339	0,0332	0,0325	0,0317	0,0310	0,0303	0,0297	0,0290
2,3	0,0283	0,0277	0,0270	0,0264	0,0258	0,0252	0,0246	0,0241	0,0235	0,0229
2,4	0,0224	0,0219	0,0213	0,0208	0,0203	0,0198	0,0194	0,0189	0,0184	0,0180
2,5	0,0175	0,0171	0,0167	0,0163	0,0158	0,0154	0,0151	0,0147	0,0143	0,0139
2,6	0,0136	0,0132	0,0129	0,0126	0,0122	0,0119	0,0116	0,0113	0,0110	0,0107
2,7	0,0104	0,0101	0,0099	0,0096	0,0093	0,0091	0,0088	0,0086	0,0084	0,0081
2,8	0,0079	0,0077	0,0075	0,0073	0,0071	0,0069	0,0067	0,0065	0,0063	0,0061
2,9	0,0060	0,0058	0,0056	0,0055	0,0053	0,0051	0,0050	0,0048	0,0047	0,0046

Окончание табл. А1

x	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
3,0	0,0044	0,0043	0,0042	0,0040	0,0039	0,0038	0,0037	0,0036	0,0035	0,0034
3,1	0,0033	0,0032	0,0031	0,0030	0,0029	0,0028	0,0027	0,0026	0,0025	0,0025
3,2	0,0024	0,0023	0,0022	0,0022	0,0021	0,0020	0,0020	0,0019	0,0018	0,0018
3,3	0,0017	0,0017	0,0016	0,0016	0,0015	0,0015	0,0014	0,0014	0,0013	0,0013
3,4	0,0012	0,0012	0,0012	0,0011	0,0011	0,0010	0,0010	0,0010	0,0009	0,0009
3,5	0,0009	0,0008	0,0008	0,0008	0,0008	0,0007	0,0007	0,0007	0,0007	0,0006
3,6	0,0006	0,0006	0,0006	0,0005	0,0005	0,0005	0,0005	0,0005	0,0005	0,0004
3,7	0,0004	0,0004	0,0004	0,0004	0,0004	0,0004	0,0003	0,0003	0,0003	0,0003
3,8	0,0003	0,0003	0,0003	0,0003	0,0003	0,0002	0,0002	0,0002	0,0002	0,0002
3,9	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0001	0,0001
4,0	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001

Таблица А2

Функция стандартного нормального распределения

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}}\, dt$$

x	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3,0	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990
3,1	0,9990	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993
3,2	0,9993	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995
3,3	0,9995	0,9995	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9997
3,4	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9998
3,5	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998
3,6	0,9998	0,9998	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,7	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,8	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,9	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
4,0	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000

Таблица А3

Квантили уровней 0,99, 0,98, 0,975, 0,95, 0,9, 0,8 распределения Стьюдента с v степенями свободы

v	**0,99**	**0,98**	**0,975**	**0,95**	**0,9**	**0,8**
1	31,82	15,89	12,71	6,31	3,08	1,38
2	6,96	4,85	4,30	2,92	1,89	1,06
3	4,54	3,48	3,18	2,35	1,64	0,98
4	3,75	3,00	2,78	2,13	1,53	0,94
5	3,36	2,76	2,57	2,02	1,48	0,92
6	3,14	2,61	2,45	1,94	1,44	0,91
7	3,00	2,52	2,36	1,89	1,41	0,90
8	2,90	2,45	2,31	1,86	1,40	0,89
9	2,82	2,40	2,26	1,83	1,38	0,88
10	2,76	2,36	2,23	1,81	1,37	0,88
11	2,72	2,33	2,20	1,80	1,36	0,88
12	2,68	2,30	2,18	1,78	1,36	0,87
13	2,65	2,28	2,16	1,77	1,35	0,87
14	2,62	2,26	2,14	1,76	1,35	0,87
15	2,60	2,25	2,13	1,75	1,34	0,87
16	2,58	2,24	2,12	1,75	1,34	0,86
17	2,57	2,22	2,11	1,74	1,33	0,86
18	2,55	2,21	2,10	1,73	1,33	0,86
19	2,54	2,20	2,09	1,73	1,33	0,86
20	2,53	2,20	2,09	1,72	1,33	0,86
21	2,52	2,19	2,08	1,72	1,32	0,86
22	2,51	2,18	2,07	1,72	1,32	0,86
23	2,50	2,18	2,07	1,71	1,32	0,86
24	2,49	2,17	2,06	1,71	1,32	0,86
25	2,49	2,17	2,06	1,71	1,32	0,86
26	2,48	2,16	2,06	1,71	1,31	0,86
27	2,47	2,16	2,05	1,70	1,31	0,86
28	2,47	2,15	2,05	1,70	1,31	0,85
29	2,46	2,15	2,05	1,70	1,31	0,85
30	2,46	2,15	2,04	1,70	1,31	0,85
35	2,44	2,13	2,03	1,69	1,31	0,85
40	2,42	2,12	2,02	1,68	1,30	0,85
45	2,41	2,12	2,01	1,68	1,30	0,85
50	2,40	2,11	2,01	1,68	1,30	0,85
55	2,40	2,10	2,00	1,67	1,30	0,85
60	2,39	2,10	2,00	1,67	1,30	0,85
65	2,39	2,10	2,00	1,67	1,29	0,85

Окончание табл. А3

v	**0,99**	**0,98**	**0,975**	**0,95**	**0,9**	**0,8**
70	2,38	2,09	1,99	1,67	1,29	0,85
75	2,38	2,09	1,99	1,67	1,29	0,85
80	2,37	2,09	1,99	1,66	1,29	0,85
85	2,37	2,09	1,99	1,66	1,29	0,85
90	2,37	2,08	1,99	1,66	1,29	0,85
95	2,37	2,08	1,99	1,66	1,29	0,85
100	2,36	2,08	1,98	1,66	1,29	0,85
110	2,36	2,08	1,98	1,66	1,29	0,84
120	2,36	2,08	1,98	1,66	1,29	0,84
130	2,36	2,07	1,98	1,66	1,29	0,84
140	2,35	2,07	1,98	1,66	1,29	0,84
150	2,35	2,07	1,98	1,66	1,29	0,84
160	2,35	2,07	1,97	1,65	1,29	0,84
170	2,35	2,07	1,97	1,65	1,29	0,84
180	2,35	2,07	1,97	1,65	1,29	0,84
190	2,35	2,07	1,97	1,65	1,29	0,84
200	2,35	2,07	1,97	1,65	1,29	0,84
$+\infty$	2,33	2,05	1,96	1,64	1,28	0,84

Таблица А4

Квантили уровней 0,01, 0,05, 0,1, 0,9, 0,95, 0,99
распределения хи-квадрат с v степенями свободы

v	**0,01**	**0,05**	**0,1**	**0,9**	**0,95**	**0,99**
1	0,00	0,00	0,02	2,71	3,84	6,63
2	0,02	0,10	0,21	4,61	5,99	9,21
3	0,11	0,35	0,58	6,25	7,81	11,34
4	0,30	0,71	1,06	7,78	9,49	13,28
5	0,55	1,15	1,61	9,24	11,07	15,09
6	0,87	1,64	2,20	10,64	12,59	16,81
7	1,24	2,17	2,83	12,02	14,07	18,48
8	1,65	2,73	3,49	13,36	15,51	20,09
9	2,09	3,33	4,17	14,68	16,92	21,67
10	2,56	3,94	4,87	15,99	18,31	23,21
11	3,05	4,57	5,58	17,28	19,68	24,73
12	3,57	5,23	6,30	18,55	21,03	26,22
13	4,11	5,89	7,04	19,81	22,36	27,69
14	4,66	6,57	7,79	21,06	23,68	29,14
15	5,23	7,26	8,55	22,31	25,00	30,58
16	5,81	7,96	9,31	23,54	26,30	32,00
17	6,41	8,67	10,09	24,77	27,59	33,41
18	7,01	9,39	10,86	25,99	28,87	34,81
19	7,63	10,12	11,65	27,20	30,14	36,19
20	8,26	10,85	12,44	28,41	31,41	37,57
21	8,90	11,59	13,24	29,62	32,67	38,93
22	9,54	12,34	14,04	30,81	33,92	40,29
23	10,20	13,09	14,85	32,01	35,17	41,64
24	10,86	13,85	15,66	33,20	36,42	42,98
25	11,52	14,61	16,47	34,38	37,65	44,31
26	12,20	15,38	17,29	35,56	38,89	45,64
27	12,88	16,15	18,11	36,74	40,11	46,96
28	13,56	16,93	18,94	37,92	41,34	48,28
29	14,26	17,71	19,77	39,09	42,56	49,59
30	14,95	18,49	20,60	40,26	43,77	50,89
35	18,51	22,47	24,80	46,06	49,80	57,34
40	22,16	26,51	29,05	51,81	55,76	63,69
45	25,90	30,61	33,35	57,51	61,66	69,96
50	29,71	34,76	37,69	63,17	67,50	76,15
55	33,57	38,96	42,06	68,80	73,31	82,29
60	37,48	43,19	46,46	74,40	79,08	88,38
65	41,44	47,45	50,88	79,97	84,82	94,42
70	45,44	51,74	55,33	85,53	90,53	100,43
75	49,48	56,05	59,79	91,06	96,22	106,39
80	53,54	60,39	64,28	96,58	101,88	112,33
85	57,63	64,75	68,78	102,08	107,52	118,24
90	61,75	69,13	73,29	107,57	113,15	124,12
95	65,90	73,52	77,82	113,04	118,75	129,97
100	70,06	77,93	82,36	118,50	124,34	135,81

Таблица А5

Квантили уровня 0,9 распределения Фишера
с v_1 и v_2 степенями свободы

$\frac{v_1}{v_2}$	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
1	39,86	49,50	53,59	55,83	57,24	58,20	58,91	59,44	59,86	60,19	60,47	60,71	60,90	61,07	61,22
2	8,53	9,00	9,16	9,24	9,29	9,33	9,35	9,37	9,38	9,39	9,40	9,41	9,41	9,42	9,42
3	5,54	5,46	5,39	5,34	5,31	5,28	5,27	5,25	5,24	5,23	5,22	5,22	5,21	5,20	5,20
4	4,54	4,32	4,19	4,11	4,05	4,01	3,98	3,95	3,94	3,92	3,91	3,90	3,89	3,88	3,87
5	4,06	3,78	3,62	3,52	3,45	3,40	3,37	3,34	3,32	3,30	3,28	3,27	3,26	3,25	3,24
6	3,78	3,46	3,29	3,18	3,11	3,05	3,01	2,98	2,96	2,94	2,92	2,90	2,89	2,88	2,87
7	3,59	3,26	3,07	2,96	2,88	2,83	2,78	2,75	2,72	2,70	2,68	2,67	2,65	2,64	2,63
8	3,46	3,11	2,92	2,81	2,73	2,67	2,62	2,59	2,56	2,54	2,52	2,50	2,49	2,48	2,46
9	3,36	3,01	2,81	2,69	2,61	2,55	2,51	2,47	2,44	2,42	2,40	2,38	2,36	2,35	2,34
10	3,29	2,92	2,73	2,61	2,52	2,46	2,41	2,38	2,35	2,32	2,30	2,28	2,27	2,26	2,24
11	3,23	2,86	2,66	2,54	2,45	2,39	2,34	2,30	2,27	2,25	2,23	2,21	2,19	2,18	2,17
12	3,18	2,81	2,61	2,48	2,39	2,33	2,28	2,24	2,21	2,19	2,17	2,15	2,13	2,12	2,10
13	3,14	2,76	2,56	2,43	2,35	2,28	2,23	2,20	2,16	2,14	2,12	2,10	2,08	2,07	2,05
14	3,10	2,73	2,52	2,39	2,31	2,24	2,19	2,15	2,12	2,10	2,07	2,05	2,04	2,02	2,01
15	3,07	2,70	2,49	2,36	2,27	2,21	2,16	2,12	2,09	2,06	2,04	2,02	2,00	1,99	1,97
16	3,05	2,67	2,46	2,33	2,24	2,18	2,13	2,09	2,06	2,03	2,01	1,99	1,97	1,95	1,94
17	3,03	2,64	2,44	2,31	2,22	2,15	2,10	2,06	2,03	2,00	1,98	1,96	1,94	1,93	1,91
18	3,01	2,62	2,42	2,29	2,20	2,13	2,08	2,04	2,00	1,98	1,95	1,93	1,92	1,90	1,89
19	2,99	2,61	2,40	2,27	2,18	2,11	2,06	2,02	1,98	1,96	1,93	1,91	1,89	1,88	1,86
20	2,97	2,59	2,38	2,25	2,16	2,09	2,04	2,00	1,96	1,94	1,91	1,89	1,87	1,86	1,84
21	2,96	2,57	2,36	2,23	2,14	2,08	2,02	1,98	1,95	1,92	1,90	1,87	1,86	1,84	1,83
22	2,95	2,56	2,35	2,22	2,13	2,06	2,01	1,97	1,93	1,90	1,88	1,86	1,84	1,83	1,81
23	2,94	2,55	2,34	2,21	2,11	2,05	1,99	1,95	1,92	1,89	1,87	1,84	1,83	1,81	1,80
24	2,93	2,54	2,33	2,19	2,10	2,04	1,98	1,94	1,91	1,88	1,85	1,83	1,81	1,80	1,78
25	2,92	2,53	2,32	2,18	2,09	2,02	1,97	1,93	1,89	1,87	1,84	1,82	1,80	1,79	1,77
26	2,91	2,52	2,31	2,17	2,08	2,01	1,96	1,92	1,88	1,86	1,83	1,81	1,79	1,77	1,76
27	2,90	2,51	2,30	2,17	2,07	2,00	1,95	1,91	1,87	1,85	1,82	1,80	1,78	1,76	1,75
28	2,89	2,50	2,29	2,16	2,06	2,00	1,94	1,90	1,87	1,84	1,81	1,79	1,77	1,75	1,74
29	2,89	2,50	2,28	2,15	2,06	1,99	1,93	1,89	1,86	1,83	1,80	1,78	1,76	1,75	1,73
30	2,88	2,49	2,28	2,14	2,05	1,98	1,93	1,88	1,85	1,82	1,79	1,77	1,75	1,74	1,72
31	2,87	2,48	2,27	2,14	2,04	1,97	1,92	1,88	1,84	1,81	1,79	1,77	1,75	1,73	1,71
32	2,87	2,48	2,26	2,13	2,04	1,97	1,91	1,87	1,83	1,81	1,78	1,76	1,74	1,72	1,71
33	2,86	2,47	2,26	2,12	2,03	1,96	1,91	1,86	1,83	1,80	1,77	1,75	1,73	1,72	1,70
34	2,86	2,47	2,25	2,12	2,02	1,96	1,90	1,86	1,82	1,79	1,77	1,75	1,73	1,71	1,69
35	2,85	2,46	2,25	2,11	2,02	1,95	1,90	1,85	1,82	1,79	1,76	1,74	1,72	1,70	1,69
36	2,85	2,46	2,24	2,11	2,01	1,94	1,89	1,85	1,81	1,78	1,76	1,73	1,71	1,70	1,68
37	2,85	2,45	2,24	2,10	2,01	1,94	1,89	1,84	1,81	1,78	1,75	1,73	1,71	1,69	1,68
38	2,84	2,45	2,23	2,10	2,01	1,94	1,88	1,84	1,80	1,77	1,75	1,72	1,70	1,69	1,67

Окончание табл. А5

$\frac{v_1}{v_2}$	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
39	2,84	2,44	2,23	2,09	2,00	1,93	1,88	1,83	1,80	1,77	1,74	1,72	1,70	1,68	1,67
40	2,84	2,44	2,23	2,09	2,00	1,93	1,87	1,83	1,79	1,76	1,74	1,71	1,70	1,68	1,66
41	2,83	2,44	2,22	2,09	1,99	1,92	1,87	1,82	1,79	1,76	1,73	1,71	1,69	1,67	1,66
42	2,83	2,43	2,22	2,08	1,99	1,92	1,86	1,82	1,78	1,75	1,73	1,71	1,69	1,67	1,65
43	2,83	2,43	2,22	2,08	1,99	1,92	1,86	1,82	1,78	1,75	1,72	1,70	1,68	1,67	1,65
44	2,82	2,43	2,21	2,08	1,98	1,91	1,86	1,81	1,78	1,75	1,72	1,70	1,68	1,66	1,65
45	2,82	2,42	2,21	2,07	1,98	1,91	1,85	1,81	1,77	1,74	1,72	1,70	1,68	1,66	1,64
46	2,82	2,42	2,21	2,07	1,98	1,91	1,85	1,81	1,77	1,74	1,71	1,69	1,67	1,65	1,64
47	2,82	2,42	2,20	2,07	1,97	1,90	1,85	1,80	1,77	1,74	1,71	1,69	1,67	1,65	1,64
48	2,81	2,42	2,20	2,07	1,97	1,90	1,85	1,80	1,77	1,73	1,71	1,69	1,67	1,65	1,63
49	2,81	2,41	2,20	2,06	1,97	1,90	1,84	1,80	1,76	1,73	1,71	1,68	1,66	1,65	1,63
50	2,81	2,41	2,20	2,06	1,97	1,90	1,84	1,80	1,76	1,73	1,70	1,68	1,66	1,64	1,63

Таблица А6

Квантили уровня 0,95 распределения Фишера
с ν_1 и ν_2 степенями свободы

ν_1 / ν_2	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	161,45	199,50	215,71	224,58	230,16	233,99	236,77	238,88	240,54	241,88	242,98	243,90	244,69	245,36	245,95
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38	19,40	19,40	19,41	19,42	19,42	19,43
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,76	8,74	8,73	8,71	8,70
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,94	5,91	5,89	5,87	5,86
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,70	4,68	4,66	4,64	4,62
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,03	4,00	3,98	3,96	3,94
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,60	3,57	3,55	3,53	3,51
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,31	3,28	3,26	3,24	3,22
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,10	3,07	3,05	3,03	3,01
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,94	2,91	2,89	2,86	2,85
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,82	2,79	2,76	2,74	2,72
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75	2,72	2,69	2,66	2,64	2,62
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,63	2,60	2,58	2,55	2,53
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,57	2,53	2,51	2,48	2,46
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,51	2,48	2,45	2,42	2,40
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,46	2,42	2,40	2,37	2,35
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45	2,41	2,38	2,35	2,33	2,31
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,37	2,34	2,31	2,29	2,27
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38	2,34	2,31	2,28	2,26	2,23
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,31	2,28	2,25	2,22	2,20
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32	2,28	2,25	2,22	2,20	2,18
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30	2,26	2,23	2,20	2,17	2,15
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32	2,27	2,24	2,20	2,18	2,15	2,13
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25	2,22	2,18	2,15	2,13	2,11
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28	2,24	2,20	2,16	2,14	2,11	2,09
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22	2,18	2,15	2,12	2,09	2,07
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31	2,25	2,20	2,17	2,13	2,10	2,08	2,06
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24	2,19	2,15	2,12	2,09	2,06	2,04
29	4,18	3,33	2,93	2,70	2,55	2,43	2,35	2,28	2,22	2,18	2,14	2,10	2,08	2,05	2,03
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	2,13	2,09	2,06	2,04	2,01
31	4,16	3,30	2,91	2,68	2,52	2,41	2,32	2,25	2,20	2,15	2,11	2,08	2,05	2,03	2,00
32	4,15	3,29	2,90	2,67	2,51	2,40	2,31	2,24	2,19	2,14	2,10	2,07	2,04	2,01	1,99
33	4,14	3,28	2,89	2,66	2,50	2,39	2,30	2,23	2,18	2,13	2,09	2,06	2,03	2,00	1,98
34	4,13	3,28	2,88	2,65	2,49	2,38	2,29	2,23	2,17	2,12	2,08	2,05	2,02	1,99	1,97
35	4,12	3,27	2,87	2,64	2,49	2,37	2,29	2,22	2,16	2,11	2,07	2,04	2,01	1,99	1,96
36	4,11	3,26	2,87	2,63	2,48	2,36	2,28	2,21	2,15	2,11	2,07	2,03	2,00	1,98	1,95
37	4,11	3,25	2,86	2,63	2,47	2,36	2,27	2,20	2,14	2,10	2,06	2,02	2,00	1,97	1,95

Окончание табл. А6

$\frac{v_1}{v_2}$	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
38	4,10	3,24	2,85	2,62	2,46	2,35	2,26	2,19	2,14	2,09	2,05	2,02	1,99	1,96	1,94
39	4,09	3,24	2,85	2,61	2,46	2,34	2,26	2,19	2,13	2,08	2,04	2,01	1,98	1,95	1,93
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08	2,04	2,00	1,97	1,95	1,92
41	4,08	3,23	2,83	2,60	2,44	2,33	2,24	2,17	2,12	2,07	2,03	2,00	1,97	1,94	1,92
42	4,07	3,22	2,83	2,59	2,44	2,32	2,24	2,17	2,11	2,06	2,03	1,99	1,96	1,94	1,91
43	4,07	3,21	2,82	2,59	2,43	2,32	2,23	2,16	2,11	2,06	2,02	1,99	1,96	1,93	1,91
44	4,06	3,21	2,82	2,58	2,43	2,31	2,23	2,16	2,10	2,05	2,01	1,98	1,95	1,92	1,90
45	4,06	3,20	2,81	2,58	2,42	2,31	2,22	2,15	2,10	2,05	2,01	1,97	1,94	1,92	1,89
46	4,05	3,20	2,81	2,57	2,42	2,30	2,22	2,15	2,09	2,04	2,00	1,97	1,94	1,91	1,89
47	4,05	3,20	2,80	2,57	2,41	2,30	2,21	2,14	2,09	2,04	2,00	1,96	1,93	1,91	1,88
48	4,04	3,19	2,80	2,57	2,41	2,29	2,21	2,14	2,08	2,03	1,99	1,96	1,93	1,90	1,88
49	4,04	3,19	2,79	2,56	2,40	2,29	2,20	2,13	2,08	2,03	1,99	1,96	1,93	1,90	1,88
50	4,03	3,18	2,79	2,56	2,40	2,29	2,20	2,13	2,07	2,03	1,99	1,95	1,92	1,89	1,87

Учебное издание

ЭКОНОМЕТРИКА
в схемах и таблицах

Учебно-методическое пособие

Составители:
Радионова Марина Владимировна
Фролова Наталья Владимировна

Печатается в авторской редакции

Корректор *М.Н. Афанасьева*

Подписано в печать 16.12.2018. Формат 60×90/16.
Усл. печ. л. 11,7. Тираж 500 экз. Заказ 301/2018

Издательство
Пермского национального исследовательского
политехнического университета.
Адрес: 614990, г. Пермь, Комсомольский пр., 29, к. 113.
Тел. (342) 219-80-33.