

Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
"САМАРСКИЙ ГОСУДАРСТВЕННЫЙ ЭКОНОМИЧЕСКИЙ УНИВЕРСИТЕТ"

**ЭКОНОМЕТРИКА
В ТАБЛИЧНОМ РЕДАКТОРЕ
MS EXCEL**

Практикум

Самара
Издательство
Самарского государственного экономического университета
2019

УДК 330.43(075.8)
ББК У.в6я7
Э40

Рецензенты: кафедра статистики Саратовского социально-экономического института (филиала) РЭУ им. Г.В. Плеханова (зав. кафедрой д-р экон. наук, доцент М.Н. Толмачев);
С.И. Харитонов, д-р физ.-мат. наук, профессор кафедры нанотехнологий Самарского национального исследовательского университета им. академика С.П. Королева

Издается по решению
редакционно-издательского совета университета

Авторы: канд. экон. наук, доцент **Е.Г. Репина**; канд. экон. наук, доцент **А.П. Цыпин**;
канд. физ.-мат. наук, доцент **Н.А. Зайчикова**; ст. преподаватель **С.Ю. Ширнаева**

Э40 **Эконометрика в табличном редакторе MS Excel** [Электронный ресурс] : практикум / **Е.Г. Репина, А.П. Цыпин, Н.А. Зайчикова, С.Ю. Ширнаева.** - Самара : Изд-во Самар. гос. экон. ун-та, 2019. - 1 электрон. опт. диск. - Систем. требования: процессор Intel с тактовой частотой 1,3 ГГц и выше ; 256 Мб ОЗУ и более ; MS Windows XP/Vista/7/10 ; Adobe Reader ; разрешение экрана 1024×768 ; привод CD-ROM. - Загл. с титул. экрана. - № гос.регистрации: 0321903942.

ISBN 978-5-94622-970-8

Рассматриваются функциональные возможности табличного редактора MS Excel для проведения эконометрического моделирования на персональном компьютере. Подробно описан алгоритм работы с инструментом анализа данных «Пакет анализа» и со встроенными математическими, статистическими, логическими функциями. Содержатся краткие сведения из теории эконометрического моделирования, а также комплекс задач и тестов для закрепления практических навыков эконометрического моделирования в MS Excel.

Практикум разработан в соответствии с рабочей программой дисциплины «Эконометрика» и с требованиями Федерального государственного образовательного стандарта высшего образования по подготовке бакалавров направлений «Экономика», «Социология», «Менеджмент».

УДК 330.43(075.8)
ББК У.в6я7

ISBN 978-5-94622-970-8

© ФГБОУ ВО «Самарский государственный
экономический университет», 2019

СОДЕРЖАНИЕ

Введение. Краткое руководство пользователя табличного редактора MS Excel	5
1. Некоторые графические возможности MS Excel	8
1.1. Решение типовых задач	8
1.1.1. Построение линейной диаграммы	8
1.1.2. Построение диаграммы рассеяния	10
1.1.3. Построение лепестковой диаграммы	13
1.1.4. Построение диаграммы структуры	14
1.2. Задачи для самостоятельного выполнения	16
1.3. Тесты	26
1.4. Контрольные вопросы	32
2. Выявление корреляционной взаимосвязи. Оценка параметров классической модели парной линейной регрессии	33
2.1. Решение типовых задач	33
2.1.1. Расчет парного линейного коэффициента корреляции	33
2.1.2. Оценка параметров классической модели парной линейной регрессии в MS Excel	36
2.2. Задачи для самостоятельного выполнения	40
2.3. Тесты	42
2.4. Контрольные вопросы	45
3. Инструмент анализа данных «РЕГРЕССИЯ» в табличном редакторе MS Excel	46
3.1. Решение типовых задач	46
3.2. Задачи для самостоятельного выполнения	51
3.3. Тесты	51
3.4. Контрольные вопросы	54
4. Оценивание модели множественной линейной регрессии с использованием инструмента анализа данных «РЕГРЕССИЯ»	56
4.1. Решение типовых задач	56
4.2. Задача для самостоятельного выполнения	63
4.3. Тесты	67
4.4. Контрольные вопросы	69
5. Фиктивная переменная в регрессионной модели	70
5.1. Решение типовых задач	70
5.2. Задача для самостоятельного выполнения	76
5.3. Тесты	78
5.4. Контрольные вопросы	79
6. Регрессионная модель с двумя фиктивными переменными	81
6.1. Решение типовых задач	81
6.2. Задача для самостоятельного выполнения	88
6.3. Тесты	93
6.4. Контрольные вопросы	96
7. Мультиколлинеарность в модели множественной линейной регрессии	98
7.1. Решение типовых задач	98
7.1.1. Выявление мультиколлинеарности на основе матрицы парных коэффициентов корреляции (МПКК)	98
7.1.2. Показатель VIF (<i>Variance Inflation Factor</i>)	100

7.2. Задача для самостоятельного выполнения.....	102
7.3. Тесты.....	102
7.4. Контрольные вопросы.....	105
8. Гетероскедастичность случайной компоненты эконометрической модели	106
8.1. Решение типовых задач.....	106
8.1.1. Графический способ обнаружения гетероскедастичности.....	106
8.1.2. Аналитический способ выявления гетероскедастичности	109
8.2. Задача для самостоятельного выполнения.....	113
8.3. Тесты.....	114
8.4. Контрольные вопросы.....	116
9. Временные ряды: линейный тренд	117
9.1. Решение типовых задач.....	117
9.1.1. Расчет абсолютных, относительных и средних показателей, характеризующих динамику исследуемого показателя (уровней временного ряда).....	118
9.1.2. Построение линейного тренда с использованием графических инструментов табличного редактора MS Excel.....	119
9.1.3. Аналитическое выравнивание на основе линейной функции	121
9.2. Задача для самостоятельного выполнения.....	126
9.3. Тесты.....	127
9.4. Контрольные вопросы.....	128
10. Временные ряды: параболический тренд второго порядка.....	130
10.1. Решение типовых задач.....	130
10.1.1. Построение параболического тренда второго порядка с использованием графических инструментов табличного редактора MS Excel	131
10.1.2. Аналитическое выравнивание на основе полиномиальной функции	133
10.2. Задача для самостоятельного выполнения.....	138
10.3. Тесты.....	140
10.4. Контрольные вопросы.....	141
11. Временные ряды: гиперболический тренд	142
11.1. Решение типовых задач.....	142
11.2. Задача для самостоятельного выполнения.....	147
11.3. Тесты.....	148
11.4. Контрольные вопросы.....	150
12. Временные ряды: фиктивные переменные в анализе сезонности.....	151
12.1. Решение типовых задач	151
12.2. Задача для самостоятельного выполнения.....	155
12.3. Тесты	157
12.4. Контрольные вопросы.....	159
13. Автокорреляция.....	161
13.1. Решение типовых задач.....	161
13.2. Задача для самостоятельного выполнения.....	166
13.3. Тесты.....	167
13.4. Контрольные вопросы.....	168
Рекомендуемая литература.....	170
Приложения.....	172

ВВЕДЕНИЕ. КРАТКОЕ РУКОВОДСТВО ПОЛЬЗОВАТЕЛЯ ТАБЛИЧНОГО РЕДАКТОРА MS EXCEL

Функции в MS Excel оперируют аргументами, которые могут иметь вид постоянных чисел, ссылок на ячейки или массивы.

Выражения (аргументы) можно вводить вручную в определенную ячейку или в строку формул, если хорошо знать синтаксис конкретной функции. Намного удобнее воспользоваться специальным окном аргументов, которое содержит подсказки и уже готовые поля для ввода данных. Перейти в окно аргумента статистических выражений можно через **«Мастер функций»** или с помощью кнопок **«Библиотеки функций»** на ленте.

Запустить **«Мастер функций»** можно:

- 1) кликнуть по пиктограмме ;
- 2) набрать на клавиатуре сочетание клавиш Shift+F3.

Справочная информация по некоторым функциям MS Excel:

1. СУММ

Функция **СУММ** относится к категории **«Математические»**. Она возвращает сумму указанных аргументов. Синтаксис функции: =СУММ(*число1*; *число2*;...).

2. СРЗНАЧ

Функция **СРЗНАЧ** относится к категории **«Статистические»**. Она ищет среднее арифметическое значение указанных чисел. Результат этого расчета выводится в отдельную ячейку, в которой и содержится формула. Синтаксис функции: =СРЗНАЧ(*число1*; *число2*;...).

3. ДИСПР

Данная функция относится к категории **«Статистические»**. Функция **ДИСПР** (или **ДИСПР.Г** в последних версиях MS Excel) вычисляет значение неисправленной дисперсии для заданной совокупности или дисперсии генеральной совокупности. Синтаксис функции: =ДИСПР(*число1*; *число2*;...) (в MS Excel 2007 и более ранних версиях) или =ДИСПР.Г(*число1*; *число2*;...) (в последних версиях MS Excel).

4. КОРЕНЬ

Функция **КОРЕНЬ** относится к категории **«Математические»**. Она возвращает квадратный корень указанного аргумента. Синтаксис функции: =КОРЕНЬ(*число*).

	1	2	3	4	5	6	7	8	9	10	11	12	13
1			<i>Эконометрика в табличном редакторе MS Excel</i>										
2													
3													

5. СТАНДОТКЛОНП

Данная функция относится к категории **«Статистические»**. Формула **СТАНДОТКЛОНП** так же, как и **ДИСПР**, является пережитком старых версий программы. Сейчас современный ее подвид - **СТАНДОТКЛОН.Г** - используется для вычисления среднеквадратического отклонения генеральной совокупности или неисправленного среднеквадратического отклонения. Синтаксис функций: =СТАНДОТКЛОНП(*число1*; *число2*;...) или =СТАНДОТКЛОН.Г(*число1*; *число2*;...) (в последних версиях MS Excel).

6. КОВАР

Функция **КОВАР** относится к категории **«Статистические»**. Функция **КОВАР** (или **КОВАР.Г** в последних версиях MS Excel). Вычисляется неисправленная ковариация выборки, среднее попарных произведений отклонений. Синтаксис функции: =КОВАР(*массив1*; *массив2*).

7. КОРРЕЛ

Функция **КОРРЕЛ** относится к категории **«Статистические»**. Она возвращает значение коэффициента корреляции между двумя массивами данных. Синтаксис функции: =КОРРЕЛ(*массив1*; *массив2*).

8. СТЬЮДРАСПОБР

Функция **СТЬЮДРАСПОБР** относится к категории **«Статистические»**. Она возвращает обратное значение функции распределения Стьюдента, определяя критическую точку двустороннего *t*-распределения. Синтаксис функции: =СТЬЮДРАСПОБР(*вероятность*; *степени_свободы*) или =СТЬЮДЕНТ.ОБР.2Х(*вероятность*; *степени_свободы*) (в последних версиях MS Excel).

9. ФРАСПОБР

Функция **ФРАСПОБР** (или **Ф.ОБР.ПХ** в последних версиях MS Excel) относится к категории **«Статистические»**. Она возвращает обратное значение функции распределения Фишера-Снедекора, определяя критическую точку F-распределения. Синтаксис функции следующий: =ФРАСПОБР(*вероятность*; *степени_свободы1*; *степени_свободы_2*) или =Ф.ОБР.ПХ(*вероятность*; *степени_свободы1*; *степени_свободы_2*).

10. ЛИНЕЙН

Функция **ЛИНЕЙН** относится к категории **«Статистические»**. Она возвращает параметры линейного приближения, полученные методом наименьших квадратов, т.е. массив со значениями оценок коэффициентов линейной регрессии и, по выбору, регрессионную статистику, другие параметры регрессии. Синтаксис функции: =ЛИНЕЙН(*известные_значения_у*; *известные_значения_х*; *константа*; *статистика*).

ЗАМЕЧАНИЕ: **«известные_значения_у»** - множество наблюдаемых значений *y*; **«известные_значения_х»** - множество наблюдаемых значений *x*; **«константа»** - логическое значение (1 (ИСТИНА) - оцениваемая модель содержит свободный член,

	1	2	3	4	5	6	7	8	9	10	11	12	13
1			<i>Введение. Краткое руководство пользователя</i>										
2			<i>табличного редактора MS Excel</i>										
3													

0 (ЛОЖЬ) - иначе); **«статистика»** - логическое значение (1(ИСТИНА) - вычислять дополнительные регрессионные статистики, 0 (ЛОЖЬ) – иначе).

11. ЕСЛИ

Функция **ЕСЛИ** относится к категории **«Логические»**. Она проверяет, выполняется ли заданное условие, и возвращает значение 1 в случае положительного ответа или значение 0 в случае отрицательного. Синтаксис функции: =ЕСЛИ(*логическое_выражение*; *значение_если_истина*; *значение_если_ложь*).

12. РАНГ

Функция **РАНГ** (или **РАНГ.РВ** в последних версиях MS Excel) относится к категории **«Статистические»**. Она определяет ранг числа в списке чисел. Синтаксис функции: =РАНГ(*число*; *ссылка*; *порядок*) или =РАНГ.РВ(*число*; *ссылка*; *порядок*).

ЗАМЕЧАНИЕ: **«число»** - обязательный аргумент, число, для которого определяется ранг; **«ссылка»** - обязательный аргумент, массив чисел или ссылка на список чисел (нечисловые значения в ссылке игнорируются); **«порядок»** - число, задающее направление сортировки (0 или значение опущено - убывание; число не равное 0 - возрастание).

Функция **РАНГ.РВ** в случае, если несколько чисел имеют одинаковый ранг, возвращает высший ранг из этого набора значений.

13. СЧЕТ

Функция **СЧЕТ** относится к категории **«Статистические»**. Она подсчитывает количество ячеек в диапазоне, который содержит числа. В качестве аргумента выделяются значения, которые могут содержать различные типы данных, при этом будут учитываться только числовые значения. Синтаксис функции: =СЧЕТ(*значение1*; *значение2*; *значение3*...).

1. НЕКОТОРЫЕ ГРАФИЧЕСКИЕ ВОЗМОЖНОСТИ MS EXCEL



Показаны способы графического представления статистических данных в табличном редакторе Excel.

1.1. Решение типовых задач

1.1.1. Построение линейной диаграммы

Динамическое развитие экономического показателя можно представить в графическом виде. Возможности редактора Excel проиллюстрируем на примере данных в табл. 1.1.

Таблица 1.1

Поквартальная динамика ВВП РФ, млрд руб.

Период	Объем ВВП в текущих ценах согласно ОКВЭД	Период	Объем ВВП в текущих ценах согласно ОКВЭД
2005		2008	
I	4458,6	I	8877,7
II	5077,9	II	10238,3
III	5845,2	III	11542
IV	6228,1	IV	10618,9
2006		2009	
I	5792,9	I	8334,6
II	6368,1	II	9244,8
III	7275,8	III	10411,3
IV	7480,3	IV	10816,4
2007		2010	
I	6780,2	I	9995,8
II	7767,5	II	10977
III	8902,7	III	12086,5
IV	9797	IV	13249,3

1. Некоторые графические возможности MS Excel

Окончание табл. 1.1

Период	Объем ВВП в текущих ценах согласно ОКВЭД	Период	Объем ВВП в текущих ценах согласно ОКВЭД
2011		2015	
I	12844,3	I	18209,7
II	14313,7	II	19284,1
III	15663,6	III	21294,4
IV	16876,6	IV	22016,1
2012		2016	
I	14925	I	19110,1
II	16149	II	20624,9
III	17442,1	III	22515,8
IV	18410,7	IV	23897,8
2013		2017	
I	15891,7	I	20549,8
II	17015,1	II	22035,1
III	18543,5	III	23948,8
IV	19566,5	IV	25503,4
2014		2018	
I	17138,9	I	22239,4
II	18884,6	-	-
III	20406,9	-	-
IV	21514,7	-	-

Источник: Динамические ряды макроэкономической статистики РФ / Единый архив экономических и социологических данных. URL: <http://sophist.hse.ru/hse/nindex.shtml>.

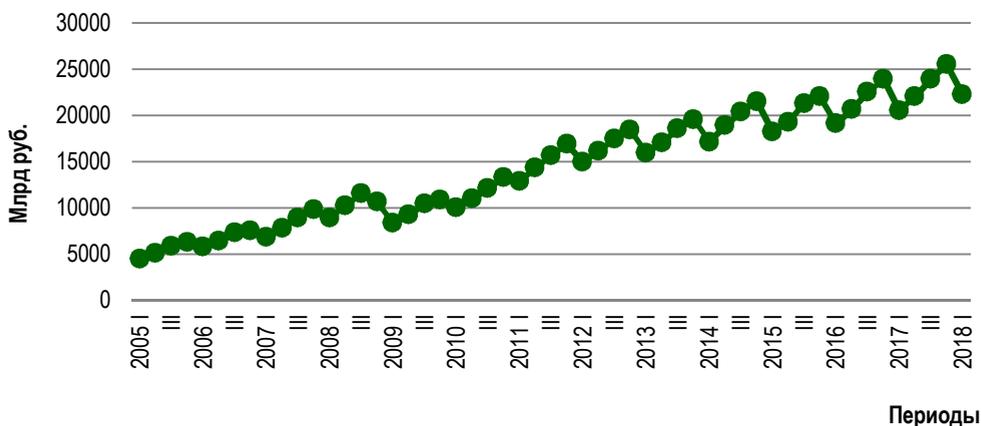


Рис. 1.1. Динамика ВВП РФ в текущих ценах согласно ОКВЭД

Для построения линейного графика необходимо следующее.

Шаг 1. Выделить курсором данные вместе с обозначениями периодов.

Шаг 2. В главном меню необходимо выбрать **Вставка/График/График с маркером**.

Результатом данных манипуляций будет линейный график (рис. 1.1).

Опираясь на представленный график, можно сделать вывод, что ВВП России имеет тенденцию к росту с ежегодными спадами в I квартале.

1.1.2. Построение диаграммы рассеяния

В качестве исходных данных для иллюстрации построения диаграммы рассеяния воспользуемся информацией по 15 промышленным предприятиям (табл. 1.2).

Таблица 1.2

Вариация экономических показателей промышленных предприятий

№ п/п	Произведено продукции в месяц, млн руб.	Среднедневная заработная плата одного работающего, руб.
1	14,6	245,1
2	15,8	259,0
3	15,0	257,2
4	15,8	261,8
5	15,1	258,8
6	14,5	247,2
7	14,6	255,2
8	12,5	234,0
9	14,7	256,0
10	12,2	231,0
11	16,8	276,0
12	13,7	234,0
13	13,2	245,0
14	16,2	265,0
15	12,4	212,0

Для того чтобы построить диаграмму рассеяния, необходимо следующее.

Шаг 1. В главном меню выберем **Вставка / Точечный / Точечный с маркером**. В результате получаем рис. 1.2.

Данный график малоинформативен, так как точки на нем сосредоточены в правом верхнем углу. В подобных случаях рекомендуется сместить начала координат. В нашем случае по оси ОУ изменим минимальное значение (200), а по оси ОХ минимальное значение выберем равным 10.

1. Некоторые графические возможности MS Excel

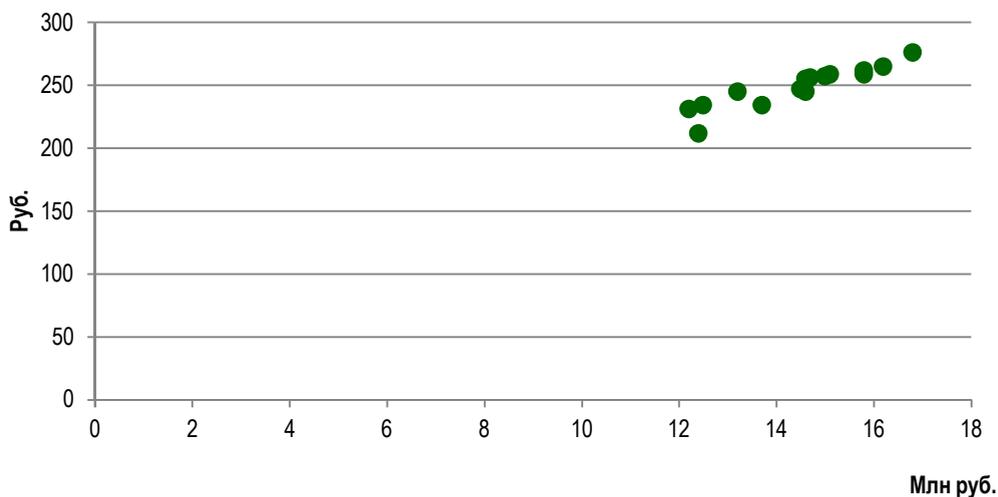


Рис. 1.2. Диаграмма рассеяния зависимости объемов произведенной продукции от средней заработной платы

Шаг 2. Направим курсор на шкалу ОУ и нажатием правой кнопки мыши вызовем дополнительное меню (рис. 1.3), в котором следует выбрать **Формат оси**.

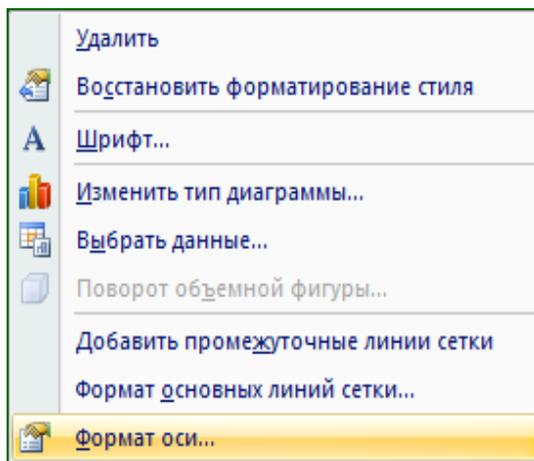


Рис. 1.3. Вспомогательное меню

Шаг 3. В окне **Формат оси** активируем минимальное значение и устанавливаем значение 200 (рис. 1.4).

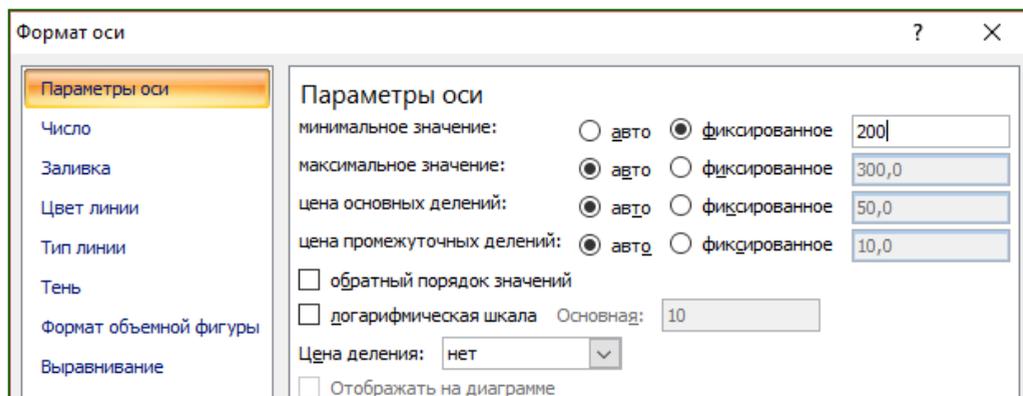


Рис. 1.4. Установка начальной точки отсчета

Аналогичным образом проведем установки по оси ОХ.
В результате получаем скорректированную диаграмму рассеяния (рис. 1.5).

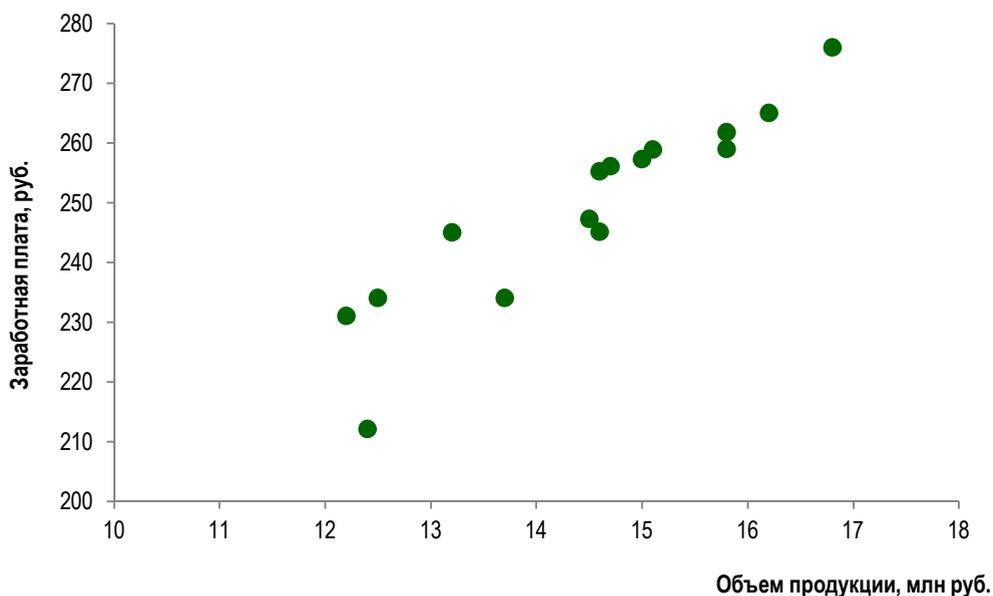


Рис. 1.5. Диаграмма рассеяния зависимости объемов произведенной продукции от средней заработной платы (с коррекцией)

Анализируя полученный график, можно предположить наличие прямой корреляционной взаимосвязи между исследуемыми показателями.

1. Некоторые графические возможности MS Excel

1.1.3. Построение лепестковой диаграммы

Для иллюстрации процесса построения лепестковой диаграммы используем данные, характеризующие инвестиции в основной капитал в России за 2008 и 2010 гг. (табл. 1.3).

Таблица 1.3

Динамика инвестиций в основной капитал в России, млрд руб.

Месяц	2008 г.	2010 г.
Январь	364,3	331,7
Февраль	447,3	421,2
Март	503	489,7
Апрель	544,1	527,2
Май	672,2	642,2
Июнь	775,2	793,1
Июль	725,5	683,5
Август	782	794,7
Сентябрь	861,5	882,9
Октябрь	879	931,2
Ноябрь	877,4	977,7
Декабрь	1350,1	1676,3

Источник: Динамические ряды макроэкономической статистики РФ / Единый архив экономических и социологических данных. URL: <http://sophist.hse.ru/hse/nindex.shtml>.

Для построения графика на основе представленных данных необходимо в главном меню выбрать **Вставка / Другие диаграммы / Лепестковая с маркером** (рис. 1.6).

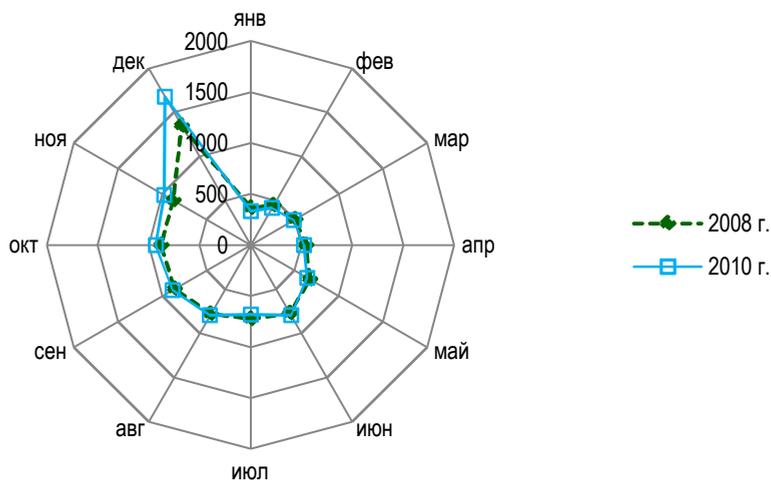


Рис. 1.6. Результат построения лепестковой диаграммы

На рис. 1.6 видно, что наблюдается рост значения показателя в декабре каждого года и отмечается спад в первом квартале, что указывает на наличие сезонной составляющей.

1.1.4. Построение диаграммы структуры

В социально-экономических исследованиях часто возникают ситуации, в которых необходимо анализировать структуру явлений или процессов за ряд лет. Одним из инструментов анализа структуры в данном случае являются структурные диаграммы.

Для построения структурных диаграмм воспользуемся данными, представленными в табл. 1.4.

Таблица 1.4

Состав и структура безработных в России по уровню образования, %

Уровень системы образования	1992 г.	1998 г.	2002 г.	2003 г.
Высшее профессиональное	10,0	10,2	10,4	11,2
Неполное высшее профессиональное	3,5	2,4	2,7	2,6
Среднее профессиональное	27,3	28,2	24,6	20,7
Среднее (полное) общее	38,1	41,8	47,2	48,9
Основное общее	18,6	14,7	13,9	14,3
Начальное общее, не имеют начального общего образования	2,5	2,7	1,2	2,3

Источник: Российский статистический ежегодник, 2004 : стат. сб. / Росстат. Москва, 2004. 725 с.

Самой распространенной структурной диаграммой является секторная или круговая. Для построения подобной диаграммы в главном меню выберем **Вставка / Круговая / Разрезанная круговая** (рис. 1.7).

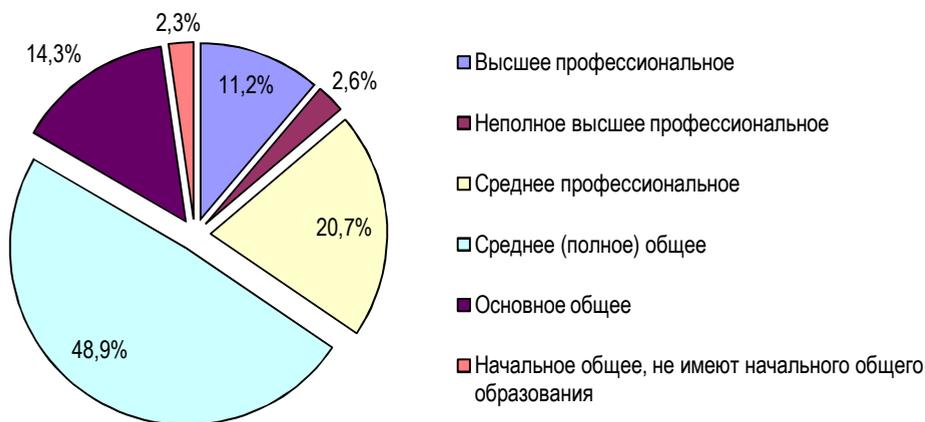


Рис. 1.7. Состав и структура безработных по уровню образования в 2003 г.

1. Некоторые графические возможности MS Excel

Данный вид диаграмм удобнее всего использовать при иллюстрации структуры явления за один год (см. рис. 1.7).

На практике может возникнуть ситуация, когда необходимо сравнивать структуру за ряд лет. В данном случае необходимо использовать кольцевую диаграмму. Следует выбрать **Вставка / Другие диаграммы / Кольцевая** (рис. 1.8).

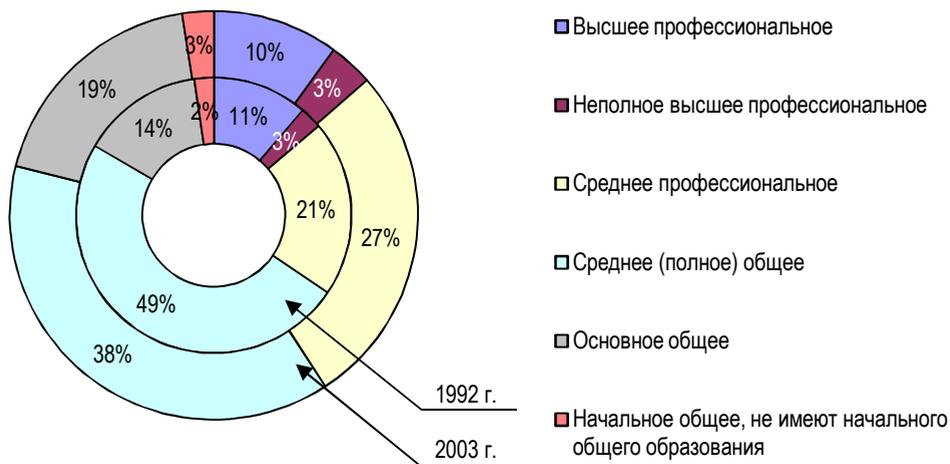


Рис. 1.8. Состав и структура безработных по уровню образования

Данный вид структурной диаграммы рекомендуется использовать, если необходимо проиллюстрировать показатели за 2-3 года. Если рассматривается период продолжительностью более 3 лет, необходимо использовать особый вид столбиковых диаграмм - нормированную гистограмму. Для этого в главном меню выбираем **Вставка / Гистограммы / Гистограммы с накоплением** (рис. 1.9).

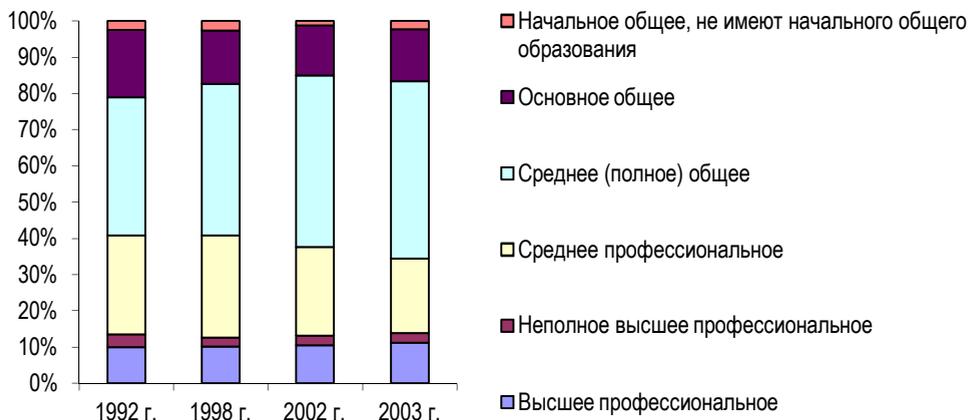


Рис. 1.9. Состав и структура контингента безработных по уровню образования

Анализируя рис. 1.9, можно сделать вывод, что наибольшее число безработных имеет среднее (полное) общее образование. При этом год от года доля таких безработных увеличивается.

1.2. Задачи для самостоятельного выполнения

Задача 1. С использованием табличного редактора Excel постройте диаграмму динамики по приведенным ниже показателям.

Период	Варианты				
	1	2	3	4	5
	Расходы консолидированного бюджета, млрд (трлн) руб.	Объем наличных денег в обращении вне банков (M ₀), млрд (трлн) руб. на конец квартала	Объем наличных денег в обращении (вне банков) и остатков средств нефинансовых предприятий, организаций, индивидуальных предпринимателей, населения на расчетных, текущих, депозитных и иных счетах до востребования (M ₂), млрд (трлн) руб. на конец квартала	Международные резервы ЦБР, млрд долл.	Валютные резервы ЦБР, млрд долл.
1998					
I	144,8	119,1	360,4	16,9	11,9
II	186,9	129,8	368,6	16,2	11,2
III	156,6	154,2	365,8	12,7	8,8
IV	264,7	187,8	448,3	12,2	7,8
1999					
I	189	174,1	473,8	10,8	6,7
II	297,8	216,4	567,7	12,2	8,2
III	309	212,8	597,4	11,2	6,6
IV	455,7	266,6	704,7	12,5	8,5
2000					
I	330,2	251,5	751,4	15,5	11,5
II	430,7	321,8	892,2	21,0	17,7
III	443,9	351,0	992,4	25,0	21,5
IV	666,8	419,3	1144,3	28,0	24,3
2001					
I	445,7	399,4	1149,5	29,709	25,942
II	580,8	474,7	1294,3	35,053	31,251

1. Некоторые графические возможности MS Excel

Продолжение таблицы

Период	Варианты				
	1	2	3	4	5
	Расходы консолидированного бюджета, млрд (трлн) руб.	Объем наличных денег в обращении вне банков (M ₀), млрд (трлн) руб. на конец квартала	Объем наличных денег в обращении (вне банков) и остатков средств нефинансовых предприятий, организаций, индивидуальных предпринимателей, населения на расчетных, текущих, депозитных и иных счетах до востребования (M ₂), млрд (трлн) руб. на конец квартала	Международные резервы ЦБР, млрд долл.	Валютные резервы ЦБР, млрд долл.
III	629,2	531	1414,4	37,957	34,04
IV	751,8	584,3	1602,6	36,622	32,542
2002					
I	586,8	552,9	1562,4	37,295	33,179
II	805,7	645,9	1751,1	43,579	39,848
III	847	672,6	1846,6	45,619	41,887
IV	1164,2	763,3	2119,6	47,793	44,054
2003					
I	770	749,5	2208,5	55,525	51,79
II	962,1	917,1	2604,5	64,43	60,691
III	1005,7	957,1	2732,1	62,073	58,33
IV	1227,1	1147	3212,7	76,938	73,175
2004					
I	889,6	1165,5	3430,4	83,398	79,639
II	1122,8	1276,1	3694	88,226	84,463
III	1132,9	1293,7	3727,5	95,082	91,328
IV	1524,4	1534,8	4363,3	124,541	120,809
2005					
I	1060,9	1481,7	4474,6	137,381	133,651
II	1382,2	1650,7	4927,4	151,578	147,776
III	1655,1	1740,7	5292,8	159,56	155,83
IV	1843,2	2009,2	6045,6	182,24	175,891
2006					
I	1274,2	1928,8	6169,4	205,881	198,676
II	2707,6	2233,4	7092,3	250,561	243,175
III	1499,4	2400,8	7757,5	266,197	258,705

Продолжение таблицы

Период	Варианты				
	1	2	3	4	5
	Расходы консолидированного бюджета, млрд (трлн) руб.	Объем наличных денег в обращении вне банков (M ₀), млрд (трлн) руб. на конец квартала	Объем наличных денег в обращении (вне банков) и остатков средств нефинансовых предприятий, организаций, индивидуальных предпринимателей, населения на расчетных, текущих, депозитных и иных счетах до востребования (M ₂), млрд (трлн) руб. на конец квартала	Международные резервы ЦБР, млрд долл.	Валютные резервы ЦБР, млрд долл.
IV	2902,8	2785,2	8995,8	303,732	295,568
2007					
I	1754,7	2741,2	9412,6	338,83	330,334
II	2475,1	3027,5	10857,7	405,841	397,399
III	2491	3220,9	11494	425,378	415,264
IV	4525	3702,2	13272,1	476,391	464,379
2008					
I	2335,9	3475,5	13382,9	512,584	498,89
II	3179,6	3724,9	14244,7	569,966	555,181
III	3226	3904,2	14374,6	556,813	542,839
IV	5250,3	3794,8	13493,2	427,08	412,547
2009					
I	2807,1	3278,3	12111,7	383,889	368,146
II	3812,6	3522,5	13161	412,591	395,97
III	3884,2	3485,6	13649,5	413,448	394,602
IV	5343,4	4038,1	15697,7	435,83	413,514
2010					
I	3220,7	3986,1	15996,5	447,004	423,321
II	4110,5	4367,7	17190,3	461,2	433
III	3978,6	4524,5	17909,2	490,1	458,3
IV	5991,2	5062,7	20173,5	479,4	443,6
2011					
I	3575,4	4916,4	19786,8	502,5	465,5
II	4563,3	5189,6	20742,6	524,5	484
III	4606,7	5420,4	21497,4	516,8	472,5
IV	7249,2	5938,6	24543,4	498,6	454

1. Некоторые графические возможности MS Excel

Продолжение таблицы

Период	Варианты				
	1	2	3	4	5
	Расходы консолидированного бюджета, млрд (трлн) руб.	Объем наличных денег в обращении вне банков (M ₀), млрд (трлн) руб. на конец квартала	Объем наличных денег в обращении (вне банков) и остатков средств нефинансовых предприятий, организаций, индивидуальных предпринимателей, населения на расчетных, текущих, депозитных и иных счетах до востребования (M ₂), млрд (трлн) руб. на конец квартала	Международные резервы ЦБР, млрд долл.	Валютные резервы ЦБР, млрд долл.
2012					
I	4583,6	5704,3	24041,3	513,5	465,7
II	5377,6	6003,9	24764	514,3	468
III	5138,6	5969,2	24657,5	529,9	476,4
IV	7726	6430,1	27405,4	537,6	486,6
2013					
I	5110,7	6181,4	27465,9	527,7	477,3
II	5724,7	6470,3	28506,1	513,8	475,2
III	5668,8	6414,4	28629,3	522,6	479,5
IV	8426,9	6985,6	31404,7	509,6	469,6
2014					
I	5432	6608,2	29800,1	486,1	442,8
II	6151,6	6763,5	30426,2	478,3	432
III	6095,5	6959,3	30644,8	454,2	409,2
IV	9536,9	7171,5	32110,5	385,5	339,4
2015					
I	6491,8	6540,8	31636,7	356,4	309,1
II	7139,7	6659,5	32492,8	361,6	313,3
III	6616,5	6744,9	32950,8	371,3	322,4
IV	9059,8	7239,1	35809,2	368,4	319,8
2016					
I	6339,1	7142,9	34698,7	387	328,9
II	7243,8	7372,7	35867,9	392,8	329,3
III	6910,7	7412,1	36148,7	397,7	332,2
IV	10395,2	7714,7	38417,8	377,7	317,5

Окончание таблицы

Период	Варианты				
	1	2	3	4	5
	Расходы консолидированного бюджета, млрд (трлн) руб.	Объем наличных денег в обращении вне банков (M ₀), млрд (трлн) руб. на конец квартала	Объем наличных денег в обращении (вне банков) и остатков средств нефинансовых предприятий, организаций, индивидуальных предпринимателей, населения на расчетных, текущих, депозитных и иных счетах до востребования (M ₂), млрд (трлн) руб. на конец квартала	Международные резервы ЦБР, млрд долл.	Валютные резервы ЦБР, млрд долл.
2017					
I	6892,4	7610,3	38555,2	397,9	330,3
II	7550,6	7946,9	39623,1	412,2	343,5
III	7236,7	8089,5	39571	424,8	351,2
IV	10309,4	8446	42442,1	432,7	356,1
2018					
I	7090,7	8442,6	42375,3	458	377,5
II	8165,2	8945,5	44125,3	456,7	378,6
III	7677,9	9138,7	44253,4	459,2	381,7
IV	10946,9	9339	47108,1	468,5	381,6

Источник: URL: <http://sophist.hse.ru/hse/nindex.shtml>.

Задача 2. С использованием табличного редактора Excel постройте диаграмму рассеяния по приведенным ниже показателям.

Регион	Для всех вариантов ВРП, млн руб.	Варианты				
		1	2	3	4	5
		Выпуск по виду деятельности "Добыча полезных ископаемых", млн руб.	Выпуск по виду деятельности "Обрабатывающие производства", млн руб.	Выпуск по виду деятельности "Обеспечение электрической энергией, газом и паром; кондиционирование воздуха", млн руб.	Продукция сельского хозяйства, млн руб.	Численность населения, тыс. чел.
Республика Башкортостан	1344360,1	233703	1082923	122987	157270,1	4063
Республика Марий Эл	160464,0	656	152086	11546	38912,7	682
Республика Мордовия	198132,8	77,3	167180	11745	59724,0	805

1. Некоторые графические возможности MS Excel

Окончание таблицы

Регион	Для всех вариантов	Варианты				
	ВРП, млн руб.	1	2	3	4	5
		Выпуск по виду деятельности "Добыча полезных ископаемых", млн руб.	Выпуск по виду деятельности "Обрабатывающие производства", млн руб.	Выпуск по виду деятельности "Обеспечение электрической энергией, газом и паром; кондиционирование воздуха", млн руб.	Продукция сельского хозяйства, млн руб.	Численность населения, тыс. чел.
Республика Татарстан	1937637,1	547515	1596330	126863	235291,8	3895
Удмуртская Республика	540115,0	191064	321066	37608	65189,8	1513
Чувашская Республика	261574,3	706	172565	24578	38736,9	1231
Пермский край	1091268,7	294130	933960	118342	41397,6	2623
Кировская область	290990,3	1107	199822	39382	39271,8	1283
Нижегородская область	1182265,0	1168	1237833	99013	68976,8	3235
Оренбургская область	772107,3	389692	304238	57928	113038,3	1978
Пензенская область	338589,0	890	195823	22153	73919,4	1332
Самарская область	1275063,6	273744	969528	115780	87488,7	3193
Саратовская область	655053,7	30492	352459	96236	134973,4	2463
Ульяновская область	328249,3	7096	256480	27120	38461,6	1247

Источник: URL: http://www.gks.ru/bgd/regl/b18_14p/Main.htm.

Задача 3. С использованием табличного редактора Excel постройте лепестковую диаграмму по приведенным ниже показателям.

Период	Инвестиции в основной капитал, млрд (трлн) руб.					
	Для всех вариантов	Варианты				
		1	2	3	4	5
	2008 г.	2011 г.	2012 г.	2013 г.	2014 г.	2015 г.
Январь	364,3	367,8	457,3	511,4	492,2	516,9
Февраль	447,3	486,3	602,3	652,5	643,2	680,7
Март	503	567,9	670,5	741,5	728,4	772,1
Апрель	544,1	611,1	709,9	771,6	770,4	812,8
Май	672,2	766,9	931,6	987	991,1	1004,2

Окончание таблицы

Период	Инвестиции в основной капитал, млрд (грлн) руб.					
	Для всех вариантов	Варианты				
		1	2	3	4	5
	2008 г.	2011 г.	2012 г.	2013 г.	2014 г.	2015 г.
Июнь	775,2	928	1089	1160,1	1180,5	1203,8
Июль	725,5	824	976,6	1053,2	1075,1	1078,4
Август	782	969,2	1099,1	1147,2	1168,5	1209,1
Сентябрь	861,5	1060,9	1149,3	1202,5	1204	1272,7
Октябрь	879	1175,4	1373	1449,9	1468,5	1547,6
Ноябрь	877,4	1192,3	1303,3	1431,9	1372,5	1446,8
Декабрь	1350,1	2085,9	2224,2	2341,5	2433,3	2460,3

Источник: URL: <http://sophist.hse.ru/hse/nindex.shtml>.

Задача 4. С использованием табличного редактора Excel постройте диаграмму структуры по приведенным ниже показателям.

Вариант 1

Возраст	Структура занятых по возрастным группам, % к итогу								
	2000 г.	2010 г.	2011 г.	2012 г.	2013 г.	2014 г.	2015 г.	2016 г.	2017 г.
До 20	2,6	1	0,9	0,7	0,7	0,6	0,6	0,6	0,5
20-24	9,6	9,4	9,2	8,3	8,5	7,8	7	6,4	5,7
25-29	12,1	13,6	13,8	14,3	14,2	14,5	14,5	14,5	14,4
30-34	11,6	12,9	13,1	13,3	13,2	13,5	13,9	14,3	14,5
35-39	15	12,5	12,5	12,7	12,7	12,8	13	13,2	13,6
40-44	15,8	11,5	11,5	11,6	11,6	12	12,2	12,5	12,7
45-49	14,1	13,7	12,9	12,3	12,3	11,8	11,4	11,2	11,4
50-54	9,8	13	13,1	13,4	13,3	13,3	13	12,7	12,1
55-59	4,4	8,3	8,5	8,8	8,8	9	9,3	9,4	9,6
60-72	5	4,2	4,5	4,6	4,7	4,8	5,1	5,2	5,5

Источник: URL: http://www.gks.ru/wps/wcm/connect/rosstat_main/rosstat/ru/statistics/publications/catalog/doc_1135075100641.

1. Некоторые графические возможности MS Excel

Вариант 2

Вид экономической деятельности	Среднегодовая численность занятых по видам экономической деятельности, % к итогу								
	2000 г.	2010 г.	2011 г.	2012 г.	2013 г.	2014 г.	2015 г.	2016 г.	2017 г.
Сельское хозяйство, охота и лесное хозяйство	13,9	9,8	9,7	9,5	9,4	9,2	9,2	9,2	6,9
Рыболовство, рыбоводство	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2
Добыча полезных ископаемых	1,7	1,6	1,6	1,6	1,6	1,6	1,6	1,6	1,6
Обрабатывающие производства	19,1	15,2	15,2	15	14,8	14,6	14,4	14,3	14,3
Производство и распределение электроэнергии, газа и воды	2,9	2,9	2,9	2,9	2,9	2,8	2,8	2,8	2,2
Строительство	6,7	8	8,1	8,3	8,4	8,4	8,3	8,1	8,9
Оптовая и розничная торговля; ремонт автотранспортных средств, мотоциклов, бытовых изделий и предметов личного пользования	13,7	17,9	18	18,1	18,3	18,7	18,8	19,1	19,3
Гостиницы и рестораны	1,5	1,8	1,8	1,8	1,9	1,9	2,0	2,0	2,4
Транспорт и связь	7,8	7,9	7,9	8,0	8,0	8,0	8,0	5,0	5,0
Финансовая деятельность	1	1,7	1,7	1,8	1,9	1,9	1,9	1,8	1,9
Операции с недвижимым имуществом, аренда и предоставление услуг	7	8	8,1	8,4	8,5	8,7	8,8	9,0	8,7
Государственное управление и обеспечение военной безопасности; социальное страхование	4,8	5,8	5,6	5,5	5,5	5,5	5,5	5,4	5,2
Образование	9,3	8,7	8,6	8,4	8,2	8,1	8,1	8,0	7,6
Здравоохранение и предоставление социальных услуг	6,8	6,8	6,8	6,7	6,7	6,6	6,6	6,6	6,2
Предоставление прочих коммунальных, социальных и персональных услуг	3,6	3,7	3,7	3,7	3,7	3,7	3,7	3,8	3,7

Источник: URL: http://www.gks.ru/bgd/regl/b18_11/IssWWW.exe/Stg/d01/06-04.doc.

Вариант 3

Период	Структура безработных по продолжительности поиска работы, % к итогу						
	2000 г.	2010 г.	2011 г.	2012 г.	2013 г.	2014 г.	2015 г.
До 3 месяцев	22,2	30	30	32,8	33,5	34,2	34,3
От 3 до 6 месяцев	12,6	22,6	17,1	18,7	16,6	17	17,8
От 6 до 12 месяцев	19	17,4	20	17,5	18,8	20,7	20,6
12 месяцев и более	46,2	30	32,9	30,9	31	28,1	27,3

Источник: URL: http://www.gks.ru/bgd/regl/b17_11/Main.htm.

Вариант 4

Способ поиска работы	Структура безработных по способам поиска работы, % к итогу						
	2000 г.	2010 г.	2011 г.	2012 г.	2013 г.	2014 г.	2015 г.
Обращение в государственные учреждения службы занятости населения	27,9	38,3	34,7	30,1	29,5	28	28,3
Обращение в коммерческую службу занятости	2,4	2,7	3	3	3,4	4,2	4,4
Подача объявлений в СМИ, Интернет, отклик на объявления	21,9	21,9	23,9	29,2	33,1	39,1	45,3
Обращение к друзьям, родственникам, знакомым	59	53,9	57,5	60,8	59,4	62,9	65,5
Непосредственное обращение к администрации/работодателю	31,2	23,7	25,5	28,5	28	29,1	31,3
Поиск земли, зданий, машин и оборудования и др. для открытия собственного дела	1	0,7	0,9	1,2	0,9	0,9	1,8
Другие способы	11,6	8,6	7,8	7,1	6,4	8	9

Источник: URL: http://www.gks.ru/bgd/regl/b17_11/Main.htm.

Вариант 5

Возраст	Структура безработных по возрастным группам, % к итогу								
	2000 г.	2010 г.	2011 г.	2012 г.	2013 г.	2014 г.	2015 г.	2016 г.	2017 г.
До 20	9,6	5,9	5,4	4,8	4,3	4,2	4,7	4,2	4,2
20-24	17,2	20,8	20,5	22,3	21,3	20,3	19,8	19,1	19,1
25-29	12,5	15	15,2	16	15,6	15,8	16,1	16,5	16,5
30-34	11,6	11,7	12	11,7	11,9	12,4	12,7	13,1	13,1
35-39	13,6	9,6	9,7	10,2	10,2	10,2	10,3	10,8	10,8
40-44	12,9	8,5	8,3	7,9	8,5	8,6	8,8	9,0	9,0
45-49	10,4	10,5	9,7	9,1	9,1	8,7	8,1	7,9	7,9
50-54	6,4	10,1	10,6	10,1	10,5	10,5	10,4	9,8	9,8
55-59	3	5,7	6	5,6	6	6,3	6,4	6,4	6,4
60-72	2,8	2,2	2,5	2,3	2,7	2,9	2,8	3,1	3,1

Источник: URL: http://www.gks.ru/bgd/regl/b17_11/lssWWW.exe/Stg/d01/06-12.doc.

Задача 5. С использованием табличного редактора Excel постройте диаграмму, соответствующую типу статистических данных Банка РФ, Минфина РФ об объеме внешнего корпоративного долга в иностранной валюте, млрд долл. США.

1. Некоторые графические возможности MS Excel

Квартал	Годы			
	2013	2014	2015	2016
I	583,1	642,6	565,2	560,3
II	607,5	631,7	560,1	533,7
III	615,8	621,4	567,3	533,7
IV	632,9	597,8	559,3	515,5

Источник: URL: <https://www.minfin.ru/ru>.

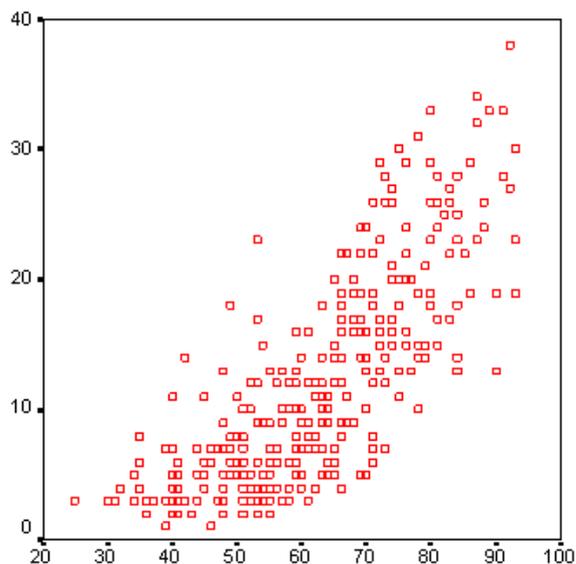
Задача 6. С использованием табличного редактора Excel постройте диаграмму, соответствующую типу статистических данных по исполнению расходов федерального бюджета за 2017 г. в разрезе государственных программ РФ по направлению «Новое качество жизни».

Государственная программа	Исполнение расходов федерального бюджета, млн руб.
Развитие образования	450132,6
Социальная поддержка граждан	1272170,2
Доступная среда	51992,7
Обеспечение доступным и комфортным жильем и коммунальными услугами граждан РФ	105198,4
Содействие занятости населения	40944,6
Обеспечение общественного порядка и противодействие преступности	661705,1
Защита населения и территории от чрезвычайных ситуаций, обеспечение пожарной безопасности и безопасности людей на водных объектах	179286,4
Развитие культуры и туризма	90155,4
Охрана окружающей среды	31537
Развитие физической культуры и спорта	94411,8
Реализация государственной национальной политики	1961,8

Источник: URL: <https://www.minfin.ru/ru>.

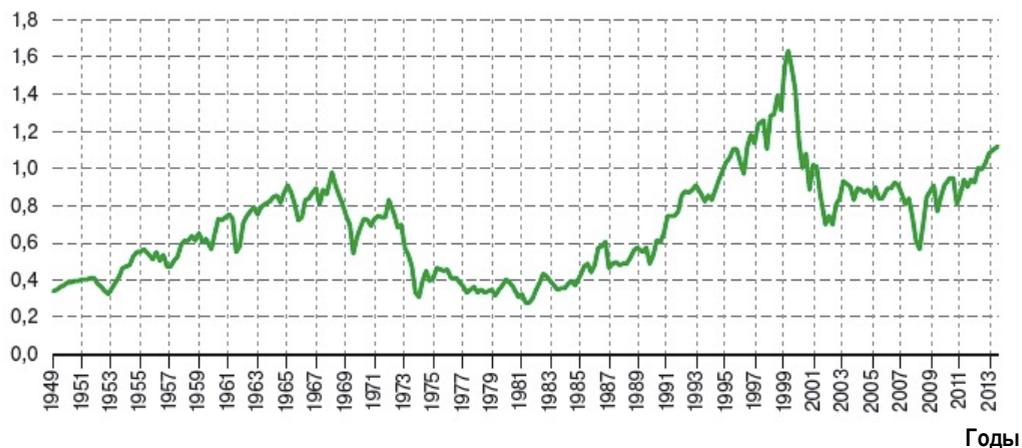
1.3. Тесты

1. Представленный ниже график является:



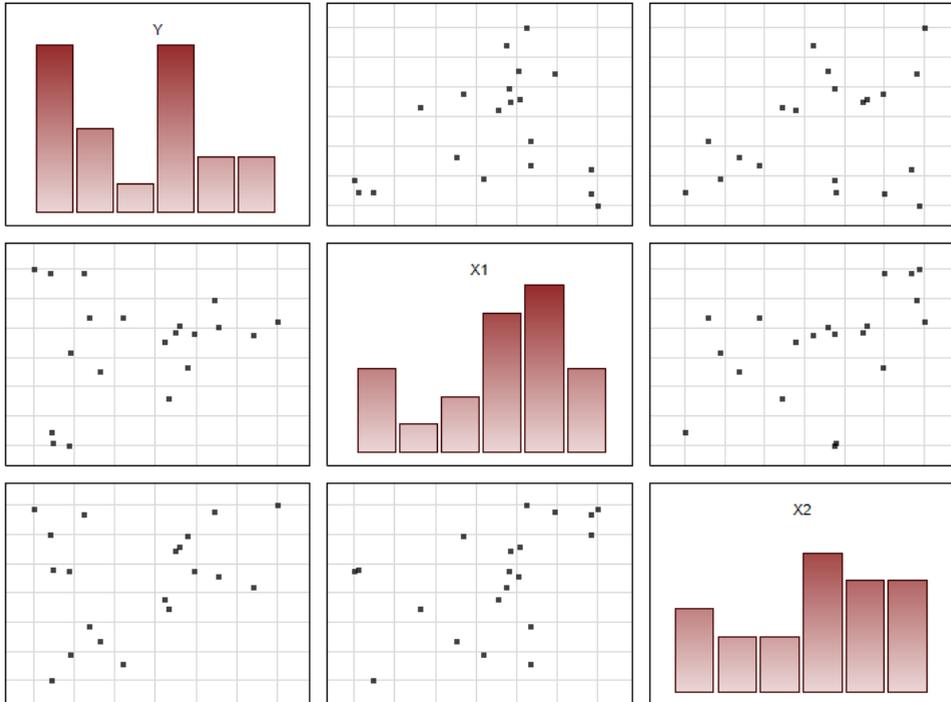
- А) картодиаграммой;
- Б) диаграммой вариации;
- В) структурной диаграммой;
- Г) диаграммой рассеяния.

2. График, представленный ниже, относится к типу:



- А) линейных диаграмм;
- Б) диаграмм вариации;
- В) структурных диаграмм;
- Г) диаграмм рассеяния.

3. Представленный ниже график является:



- А) матричным графиком;
- Б) полем корреляции;
- В) диаграммой рассеяния.

4. Для отображения совместной вариации взаимосвязанных показателей следует использовать следующий график:

- А) столбиковую диаграмму;
- Б) кривую распределения;
- В) поле корреляции;
- Г) секторную диаграмму.

5. Для отображения взаимосвязи трех показателей используется график следующего вида:

- А) столбиковая диаграмма;
- Б) кривая распределения;
- В) секторная диаграмма;
- Г) поле корреляции;
- Д) трехмерная диаграмма рассеяния.

6. Установите тип диаграмм на рисунках, представленных ниже.

Добыча полезных ископаемых и обрабатывающая промышленность в общем объеме промышленного производства Волжского р-на Самарской обл.



А)

Распределение основных предприятий промышленности Волжского р-на по объему отгруженных товаров собственного производства, выполненных работ и услуг собственными силами



Б)

Сравнение индекса промышленного производства в Волжском р-не с областными значениями



В)

Степень износа основных фондов коммерческих организаций Волжского р-на



Г)

? - разрезанная круговая

? - линейчатая с накоплением

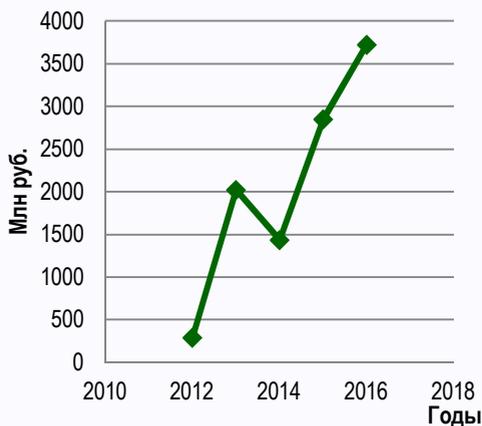
? - лепестковая

? - линейная

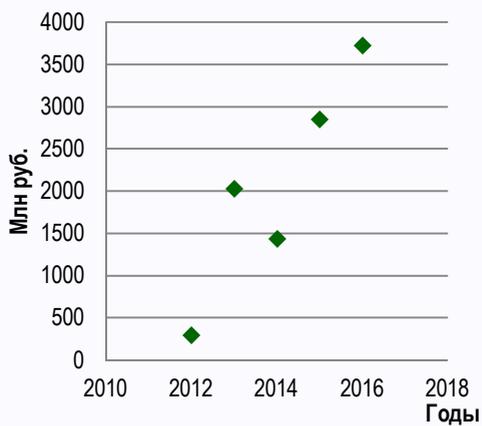
1. Некоторые графические возможности MS Excel

7. Установите тип диаграмм на рисунках, представленных ниже.

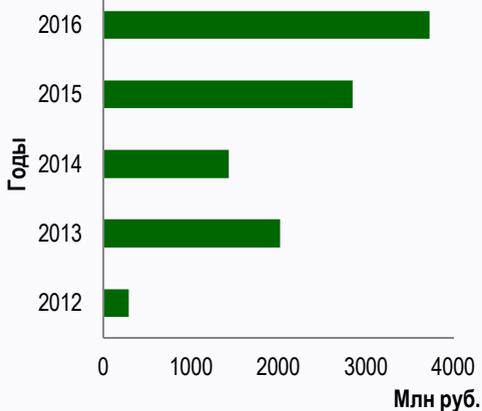
Сальдированный финансовый результат
по всем видам деятельности организаций Волжского р-на



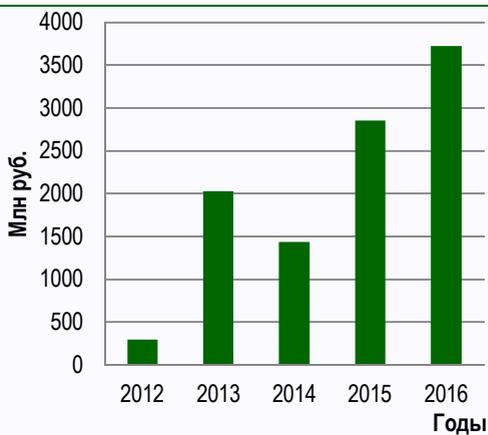
А)



Б)



В)



Г)

? - линейная диаграмма

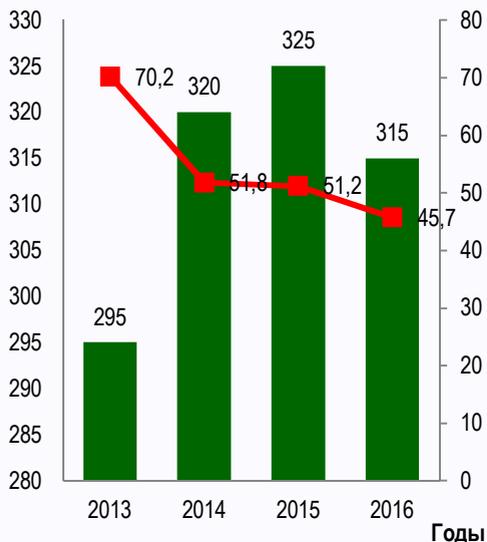
? - точечная диаграмма

? - линейный график

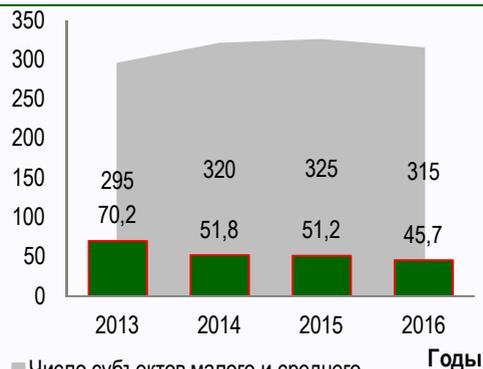
? - гистограмма

8. Установите тип диаграмм на рисунках, представленных ниже.

Число субъектов малого и среднего предпринимательства в расчете на 10 тыс. чел. населения и доля среднесписочной численности работников малых и средних предприятий в среднесписочной численности работников всех предприятий и организаций в Волжском р-не, %

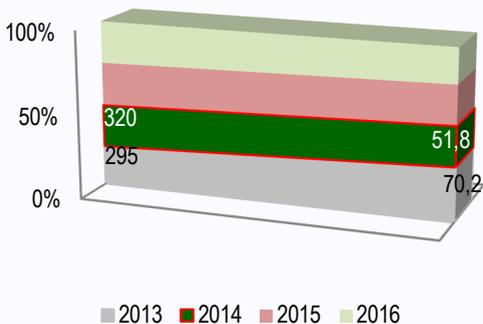


A)



■ Число субъектов малого и среднего предпринимательства в расчете на 10 тыс. человек населения
 ■ Доля среднесписочной численности работников малых и средних предприятий в среднесписочной численности работников всех предприятий и организаций (%)

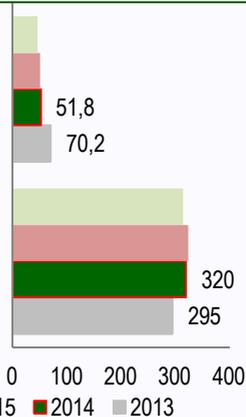
Б)



В)

Доля среднесписочной численности работников малых и средних предприятий в среднесписочной численности...

Число субъектов малого и среднего предпринимательства в расчете на 10 тыс. человек населения



Г)

? - объемная нормированная с областями и накоплением диаграмма

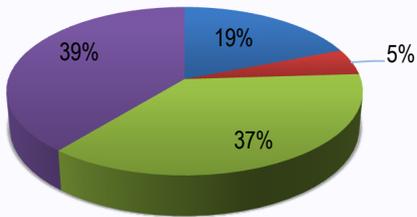
? - гистограмма с группировкой и с областями накопления

? - гистограмма с группировкой и с областями накопления

? - линейчатая с группировкой диаграмма

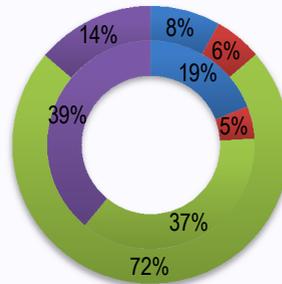
1. Некоторые графические возможности MS Excel

9. Установите тип диаграмм на рисунках, представленных ниже.



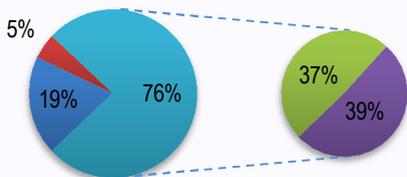
- ООО "Самарское Объединение Керамики"
- ООО "Авиакор-Железобетон"
- ООО "Роберт Бош Самара"
- ООО "Самарский Стройфарфор"

А)



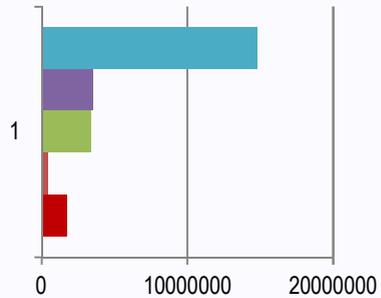
- ООО "Самарское Объединение Керамики"
- ООО "Авиакор-Железобетон"
- ООО "Роберт Бош Самара"
- ООО "Самарский Стройфарфор"

Б)



- ООО "Самарское Объединение Керамики"
- ООО "Авиакор-Железобетон"
- ООО "Роберт Бош Самара"
- ООО "Самарский Стройфарфор"

В)



- Другие предприятия
- ООО "Самарское Объединение Керамики"
- ООО "Самарский Стройфарфор"
- ООО "Роберт Бош Самара"
- ООО "Авиакор-Железобетон"

Г)

? - кольцевая диаграмма

? - линейчатая с группировкой диаграмма

? - объемная круговая диаграмма

? - вторичная круговая диаграмма

	1	2	3	4	5	6	7	8	9	10	11	12	13
1			<i>Эконометрика в табличном редакторе MS Excel</i>										
2													
3													

1.4. Контрольные вопросы

1. Дайте определение понятия «эконометрика».
2. Назовите этапы эконометрического моделирования.
3. Что понимается под спецификацией эконометрической модели?
4. Что понимается под параметризацией эконометрической модели?
5. Как реализуется этап верификации эконометрической модели?
6. Какие типы данных используются при эконометрическом моделировании?
7. Какие виды переменных используются при эконометрическом моделировании?
8. Назовите классы эконометрических моделей.
9. Что является инструментом эконометрики?
10. К какому типу эконометрических данных относятся следующие: объем платных услуг по регионам РФ за 2018 г.; ежемесячные расходы компании на рекламу; цена однокомнатных квартир в Промышленном районе г. Самары на 1 января 2019 г.?

2. ВЫЯВЛЕНИЕ КОРРЕЛЯЦИОННОЙ ВЗАИМОСВЯЗИ. ОЦЕНКА ПАРАМЕТРОВ КЛАССИЧЕСКОЙ МОДЕЛИ ПАРНОЙ ЛИНЕЙНОЙ РЕГРЕССИИ



На конкретных практических примерах вычисляется выборочный коэффициент корреляции, по методу наименьших квадратов оцениваются параметры линейной парной регрессионной модели.

2.1. Решение типовых задач

2.1.1. Расчет парного линейного коэффициента корреляции

Для иллюстрации вычисления парного линейного коэффициента корреляции в табличном редакторе Excel используем данные, представленные в табл. 2.1.

Таблица 2.1

**Инвестиции в основной капитал и среднегодовая численность занятых
в разрезе муниципальных образований Самарской области, 2017 г.**

Муниципальное образование	Y	X	Муниципальное образование	Y	X
Алексеевский м.р.	12878	1546	Приволжский м.р.	390401	3439
Безенчукский м.р.	6540234	8082	Сергеевский м.р.	6589916	9667
Богатовский м.р.	905057	2767	Ставропольский м.р.	2487599	9676
Большеглушицкий м.р.	1587301	2791	Сызранский м.р.	1079027	3327
Большечерниговский м.р.	631314	2795	Хворостянский м.р.	613808	2240
Борский м.р.	300426	3006	Челно-Вершинский м.р.	562301	2521
Волжский м.р.	18850964	18980	Шенталинский м.р.	143150	2297
Елховский м.р.	212229	1210	Шигонский м.р.	70668	2977
Иса克林ский м.р.	48713	2094	г.о. Самара	84044571	352170
Камышлинский м.р.	1064219	1510	г.о. Жигулевск	3497466	11790
Кинельский м.р.	3477251	4264	г.о. Кинель	644826	11300

Муниципальное образование	Y	X	Муниципальное образование	Y	X
Кинель-Черкасский м.р.	1777476	7183	г.о. Новокуйбышевск	13165517	30178
Клявлинский м.р.	1732117	2275	г.о. Октябрьск	175667	5300
Кошкинский м.р.	1825941	3933	г.о. Отрадный	8277553	12904
Красноармейский м.р.	457717	2831	г.о. Похвистнево	3201227	6472
Красноярский м.р.	1551759	8728	г.о. Сызрань	6119688	42744
Нефтегорский м.р.	12464092	7409	г.о. Тольятти	23217548	160124
Пестравский м.р.	1001766	2908	г.о. Чапаевск	1067078	12992
Похвистневский м.р.	957906	2230	-	-	-

Примечание: Y - инвестиции в основной капитал, тыс. руб.; X - среднегодовая численность работников организаций, чел.; м.р. - муниципальный район; г.о. - городской округ.
Источник: URL: http://www.gks.ru/scripts/db_inet2/passport/table.aspx?opt=366020002017.

Шаг 1. Необходимо отметить, что городские округа Самара и Тольятти имеют несопоставимо большие по сравнению с другими муниципальными образованиями Самарской области значения инвестиций в основной капитал. В силу этого данные два наблюдения следует исключить из дальнейшего анализа!

Для вычисления парного выборочного коэффициента корреляции необходимо в главном меню выбрать **Данные / Анализ данных**.

Шаг 2. В появившемся окне **Анализ данных** (рис. 2.1) выберем инструмент анализа **Корреляция**.

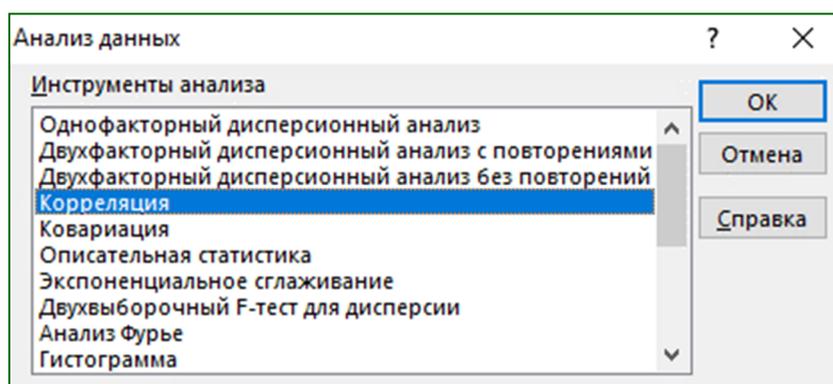


Рис. 2.1. Окно выбора инструментов анализа данных

Шаг 3. В окне **Корреляция** необходимо выделить диапазон (**Входной интервал**), в котором содержатся значения изучаемых показателей (рис. 2.2). Для этого следует нажать кнопку .

2. Выявление корреляционной взаимосвязи. Оценка параметров классической модели парной линейной регрессии

	A	B	C
1	Мун. образование	X	Y
2	Алексеевский м.р.	1546	12878
3	Безенчукский м.р.	8082	6540234
4	Богатовский м.р.	2767	905057
5	Большеглушицкий м.р.	2791	1587301
6	Большечерниговский м.р.	2795	631314
7	Борский м.р.	3006	300426
8	Волжский м.р.	18980	1,9E+07
9	Елховский м.р.	1210	212229
10	Иса克林ский м.р.	2094	48713
11	Камышлинский м.р.	1510	1064219
12	Кинельский м.р.	4264	3477251
13	Кинель-Черкасский м.р.	7183	1777476
14	Клявлинский м.р.	2275	1732117

Рис. 2.2. Выделение диапазона значений исследуемых показателей

Результатом работы процедуры вычисления выборочных коэффициентов корреляции является матрица парных коэффициентов корреляции (рис. 2.3).

	G	H	I
		X	Y
X		1	
Y		0,607188	1

Рис. 2.3. Результаты построения матрицы парных коэффициентов корреляции (МПКК)

Данная матрица симметрична относительно главной диагонали. В первом столбце матрицы расположено значение парного коэффициента корреляции, отражающее зависимость между зависимой переменной Y и независимой переменной X.

Полученное значение равно $r_{xy} = 0,607$, что указывает на заметную прямую взаимосвязь между анализируемыми показателями. Градации корреляционной связи между экономическими показателями здесь и далее соответствуют шкале Чеддока (табл. 2.2).

Таблица 2.2

Шкала Чеддока

Теснота связи	Коэффициент корреляции по абсолютной величине
Отсутствует	0
Слабая	0,1 - 0,3
Умеренная	0,3 - 0,5
Заметная	0,5 - 0,7
Высокая	0,7 - 0,9
Весьма высокая	0,9 - 0,99
Функциональная	1

2.1.2. Оценка параметров классической модели парной линейной регрессии в MS Excel

В табличном редакторе Excel содержится огромное количество встроенных функций - математических, статистических, финансовых и др. Все они значительно упрощают проводимые исследователем расчеты, так как автоматически применяют формулы, требующие трудоемких вычислений.

Рассмотрим некоторые важные статистические функции, необходимые для оценивания по МНК регрессионной модели вида

$$Y = b_0 + b_1X + \varepsilon,$$

где b_0 и b_1 - теоретические коэффициенты уравнения регрессии;

Y - объясняемая (зависимая) переменная;

X - объясняющая (независимая) переменная;

ε - случайная компонента (ошибка).

Шаг 1. Для вставки необходимой функции в главном меню выберем **Формулы / Встроить функцию** (рис. 2.4).

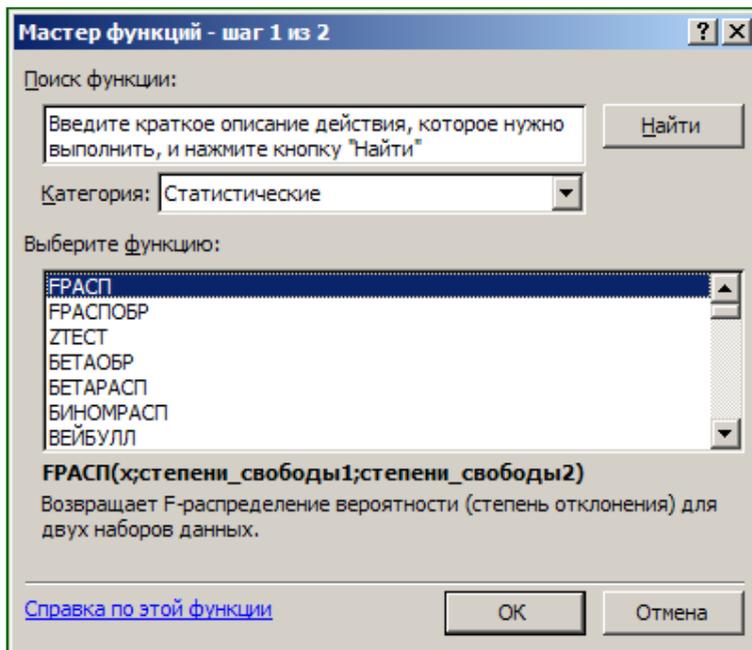


Рис. 2.4. Окно выбора необходимой функции

	1	2	3	4	5	6	7	8	9	10	11	12	13
1		2. Выявление корреляционной взаимосвязи. Оценка параметров											
2		классической модели парной линейной регрессии											
3													

Шаг 2. Для оценивания параметров линейной эконометрической модели служит функция **ЛИНЕЙН**.

Синтаксис данной функции следующий: **ЛИНЕЙН** (известные_значения_y; известные_значения_x; конст; статистика).

Поле **«Конст»** - логическое, которое устанавливает, требуется ли приравнять параметр b_0 к нулю. Если поле **«Конст»** имеет значение ИСТИНА или опущено, то оценка параметра b_0 вычисляется соответствующим образом. Если поле **«Конст»** имеет значение ЛОЖЬ, то b_0 полагается равным 0 и значения \hat{b}_0 подбираются так, чтобы выполнялось соотношение $\hat{y} = \hat{b}_1 x$.

Поле **«Статистика»** - логическое, которое устанавливает, требуется ли вывести дополнительные регрессионные статистики. Если поле **«Статистика»** имеет значение ИСТИНА, то функция **ЛИНЕЙН** рассчитывает дополнительные регрессионные статистики. Итоги работы функции **ЛИНЕЙН** имеют следующий вид:

\hat{b}_1	\hat{b}_0
$S_{\hat{b}_1}$	$S_{\hat{b}_0}$
R^2	S_e
$F_{набл}$	k_2
RSS	ESS

Если поле **«Статистика»** имеет значение ЛОЖЬ или опущено, то функция **ЛИНЕЙН** рассчитывает только оценки теоретических коэффициентов регрессии. В этом случае итоги работы функции **ЛИНЕЙН** имеют иной вид:

\hat{b}_1	\hat{b}_0
-------------	-------------

Более подробную информацию по данной функции можно получить в самом табличном редакторе Excel, для чего следует воспользоваться встроенной справкой.

Шаг 3. Практическое применение функции **ЛИНЕЙН** в табличном редакторе Excel показано на рис. 2.5.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Мун. образование	X	Y							X	Y		
2	Алексеевский м.р.	1546	12878						X	1			
3	Безенгукский м.р.	8082	6540234						Y	0,607188	1		
4	Богатовский м.р.	2767	905057										
5	Большеглушицкий м.р.	2791	1587301										
6	Большечерногоский м.р.	2795	631314										
7	Борский м.р.	3006	300426										
8	Волжский м.р.	18980	1,9E+07										
9	Елховский м.р.	1210	212229										
10	Исаклинский м.р.	2094	48713										
11	Камышлинский м.р.	1510	1064219										
12	Кинельский м.р.	4264	3477251										
13	Кинель-Черкасский м.р.	7183	1777476										
14	Клявлинский м.р.	2275	1732117										
15	Кожинский м.р.	3933	1825941										
16	Красноармейский м.р.	2831	457717										
17	Красноярский м.р.	8728	1551759										
18	Нефтегорский м.р.	7409	1,2E+07										
19	Пестравский м.р.	2908	1001766										
20	Похвистневский м.р.	2230	957906										
21	Пивновский м.р.	3439	390401										

Аргументы функции

ЛИНЕЙН

Изнестные_значения_у С2:С36 = {12878;6540234;905057;1587301;631314;...}

Изнестные_значения_x B2:В36 = {1546;8082;2767;2791;2795;3006;18980;12...}

Конст 1 = ИСТИНА

Статистика |1| = ИСТИНА

Возвращает параметры линейного приближения по методу наименьших квадратов.

Статистика логическое значение, которое указывает, требуется ли вернуть дополнительную статистику по регрессии (ИСТИНА) или только коэффициенты m и константу b (ЛОЖЬ или отсутствие значения).

Значение: 306,9986989

[Справка по этой функции](#)

OK Отмена

Рис. 2.5. Расчет оценок параметров парной линейной регрессии с применением функции ЛИНЕЙН

2. Выявление корреляционной взаимосвязи. Оценка параметров классической модели парной линейной регрессии

ЗАМЕЧАНИЕ: необходимо задать значения, равные 1, в полях «Конст» и «Статистика». В противном случае дополнительные регрессионные статистики не будут рассчитаны.

Шаг 4. Выделяем диапазон ячеек E5:F9. После активации работы функции, следует нажать клавишу F2, а затем сочетание клавиш Ctrl+Shift+Enter (ввод массива данных). В итоге получаем результаты, представленные на рис. 2.6.

	A	B	C	D	E	F	G
1	Мун. образование	X	Y				
2	Алексеевский м.р.	1546	12878				
3	Безенчукский м.р.	8082	6540234				
4	Богатовский м.р.	2767	905057				
5	Большеглушицкий м.р.	2791	1587301		306,9986989	708034,9585	
6	Большечерниговский м.р.	2795	631314		69,93305349	780223,0876	
7	Борский м.р.	3006	300426		0,368676847	3481735,217	
8	Волжский м.р.	18980	1,9E+07		19,27117021	33	
9	Елховский м.р.	1210	212229		2,33614E+14	4,00042E+14	
10	Исаклинский м.р.	2094	48713				
11	Камышлинский м.р.	1510	1064219				
12	Кинельский м.р.	4264	3477251				
13	Кинель-Черкасский м.р.	7183	1777476				
14	Клявлинский м.р.	2275	1732117				
15	Кошкинский м.р.	3933	1825941				
16	Красноармейский м.р.	2831	457717				

Рис. 2.6. Результаты применения функции ЛИНЕЙН

Исходя из полученных данных имеем следующее выборочное уравнение парной линейной регрессии: $\hat{y} = 708034,96 + 307x$. Таким образом, при росте среднегодовой численности работников организации на 1 чел. инвестиции в среднем увеличатся на 307 тыс. руб.

Шаг 5. Вычислим расчетные значения зависимой переменной \hat{y} , которые необходимы для расчета остатков регрессии. Для этого следует использовать оцененное уравнение регрессии и подставить в него значения переменной X (рис. 2.7).

ЛИНЕЙН		X		Y		Y^	
	A	B	C	D	E	F	
1	Мун. образование	X	Y	Y^			
2	Алексеевский м.р.	1546	12878	=F\$5+\$E\$5*B2			
3	Безенчукский м.р.	8082	6540234	3189198			
4	Богатовский м.р.	2767	905057	1557500			
5	Большеглушицкий м.р.	2791	1587301	1564868	306,9986989	708034,9585	
6	Большечерниговский м.р.	2795	631314	1566096	69,93305349	780223,0876	
7	Борский м.р.	3006	300426	1630873	0,368676847	3481735,217	
8	Волжский м.р.	18980	1,9E+07	6534870	19,27117021	33	
9	Елховский м.р.	1210	212229	1079503	2,33614E+14	4,00042E+14	
10	Исаклинский м.р.	2094	48713	1250800			

Рис. 2.7. Вычисление расчетных значений зависимой переменной

2.2. Задачи для самостоятельного выполнения

Задача 7. Необходимо оценить значение парного линейного коэффициента корреляции в табличном редакторе Excel. Дайте интерпретацию полученному значению коэффициента по шкале Чеддока.

Муниципальное образование	Для всех вариантов	Варианты			
		1	2	3	4
	Y	X ₁	X ₂	X ₃	X ₄
Алексеевский м.р.	12878	8099873,9	298908,6	21	21985,1
Безенчукский м.р.	6540234	37477617,4	2408364,2	78	27595,5
Богатовский м.р.	905057	3416586,7	564248	53	28063,1
Большеглушицкий м.р.	1587301	10632998,1	616777,8	49	25442
Большечерниговский м.р.	631314	5701175,5	635747,2	60	24704,2
Борский м.р.	300426	749634,6	619640,5	84	23187,1
Волжский м.р.	18850964	35563871,6	4596263,2	156	32224,2
Елховский м.р.	212229	3409398,2	205460,4	30	25006,1
Иса克林ский м.р.	48713	2359568,4	376598	60	21411,6
Камышлинский м.р.	1064219	1147359,5	472508	35	22649,1
Кинельский м.р.	3477251	18437543,5	469237	65	36582,7
Кинель-Черкасский м.р.	1777476	10558790,2	1259475,6	129	26102,1
Клявлинский м.р.	1732117	1258939,2	341277,6	33	23607
Кошкинский м.р.	1825941	20045991,3	1196246,9	98	26105,6
Красноармейский м.р.	457717	1944247,8	466704,9	40	24057,8
Красноярский м.р.	1551759	22358736	3766093,1	133	25940,9
Нефтегорский м.р.	12464092	77881929,2	1400748	48	30737
Пестравский м.р.	1001766	4475010,1	417830	37	24654
Похвистневский м.р.	957906	1192076,6	121142	106	21074,6
Приволжский м.р.	390401	644687,7	536390	54	21601,5
Сергиевский м.р.	6589916	49566446,5	2213281,1	112	31429,5
Ставропольский м.р.	2487599	20669895,6	4856283,2	95	31490,4
Сызранский м.р.	1079027	1422000,1	949733,2	49	24545,9
Хворостянский м.р.	613808	524071	3245	40	23150,3

	1	2	3	4	5	6	7	8	9	10	11	12	13
1	2. Выявление корреляционной взаимосвязи. Оценка параметров												
2	классической модели парной линейной регрессии												
3													

Окончание таблицы

Муниципальное образование	Для всех вариантов	Варианты			
		1	2	3	4
	Y	X ₁	X ₂	X ₃	X ₄
Челно-Вершинский м.р.	562301	7919030,5	265127,7	52	22900,5
Шенталинский м.р.	143150	3664546,7	411472,7	54	22797,6
Шигонский м.р.	70668	514786	674102,3	59	21106,2
г.о. Самара	84044571	504565563,4	148152794,5	1234	39394,5
г.о. Жигулевск	3497466	39553075,4	2098785,7	66	29245,1
г.о. Кинель	644826	6058949,7	3467209	71	27954,4
г.о. Новокуйбышевск	13165517	75556669,5	6412060,9	106	37408,2
г.о. Октябрьск	175667	855377,7	517818,9	49	28884,7
г.о. Отрадный	8277553	60809496	2578661,2	93	37136,3
г.о. Похвистнево	3201227	15260800,2	1344823,5	35	26719,6
г.о. Сызрань	6119688	54227968,5	11086240,6	184	27859
г.о. Тольятти	23217548	490997145,1	67103925,5	804	33455,9
г.о. Чапаевск	1067078	14035131,5	3707269,1	100	26328

Примечание: Y - инвестиции в основной капитал, тыс. руб.; X₁ - объем отгруженных товаров собственного производства, выполненных работ и услуг собственными силами, млн руб.; X₂ - оборот розничной торговли, млн руб.; X₃ - число спортивных сооружений, ед.; X₄ - среднемесячная заработная плата работников организаций, руб.
Источник: URL: http://www.gks.ru/scripts/db_inet2/passport/table.aspx?opt=367500002017.

Задача 8. Используя данные задачи 7, оцените по МНК параметры модели парной линейной регрессии с использованием табличного редактора Excel. Экономически интерпретируйте оценки теоретических коэффициентов регрессии.

Задача 9. Используя данные, полученные в результате решения задачи 8, вычислите расчетные значения зависимой переменной и нанесите их на корреляционное поле.

2.3. Тесты

1. Предпосылками метода наименьших квадратов (условиями теоремы Гаусса - Маркова) для модели парной линейной регрессии $Y = b_0 + b_1X + \varepsilon$ являются следующие утверждения (выберите несколько правильных ответов):

А) математическое ожидание случайной компоненты для любого наблюдения должно быть постоянно:

$$M(\varepsilon_i) = \text{const}, i = \overline{1, n};$$

Б) математическое ожидание случайной компоненты для любого наблюдения должно быть равно нулю:

$$M(\varepsilon_i) = 0, i = \overline{1, n};$$

В) теоретическая дисперсия случайной компоненты должна быть постоянна для всех наблюдений:

$$D(\varepsilon_i) = \text{const}, i = \overline{1, n};$$

Г) теоретическая дисперсия случайной компоненты должна стремиться к нулю при неограниченном возрастании числа наблюдений:

$$\lim_{n \rightarrow \infty} D(\varepsilon_i) = 0, i = \overline{1, n};$$

Д) случайные компоненты должны быть статистически независимы между собой:

$$\text{cov}(\varepsilon_i, \varepsilon_k) = 0, i, k = \overline{1, n}, i \neq k;$$

Е) распределение случайной компоненты должно быть независимым от значений объясняющей переменной X , т.е.

$$\text{cov}(\varepsilon_i, x_i) = 0, i, k = \overline{1, n}.$$

2. Если значение коэффициента корреляции, рассчитанное для модели парной линейной регрессии $y = b_0 + b_1x + \varepsilon$, равно единице, то (выберите несколько правильных ответов):

А) величина ε оказывает существенное влияние на переменную y ;

Б) связь между переменными y и x функциональная;

В) величина ε не оказывает влияния на переменную y ;

Г) связь между параметрами b_0 и b_1 функциональная.

3. Выборочный коэффициент корреляции r_{xy} между признаками X и Y (выберите несколько правильных ответов):

А) прямо пропорционален произведению их дисперсий: $\hat{D}(X) \cdot \hat{D}(Y)$;

Б) обратно пропорционален выборочной ковариации: $\hat{\text{cov}}(x, y)$;

	1	2	3	4	5	6	7	8	9	10	11	12	13
1		2. Выявление корреляционной взаимосвязи. Оценка параметров											
2		классической модели парной линейной регрессии											
3													

В) обратно пропорционален произведению выборочных средних квадратических отклонений: $\hat{\sigma}_x \cdot \hat{\sigma}_y$;

Г) прямо пропорционален выборочной ковариации $\hat{cov}(x, y)$.

4. При оценке статистической значимости линейного выборочного коэффициента корреляции r_{xy} используется:

- А) критерий Дарбина - Уотсона;
- Б) критерий Фишера - Снедекора;
- В) критерий Стьюдента;
- Г) тест Голдфелда - Квандта.

5. Коэффициент b_1 модели парной линейной регрессии $Y = b_0 + b_1X + \varepsilon$ показывает:

- А) тесноту связи между переменными X и Y ;
- Б) на сколько процентов изменится зависимая переменная Y , если независимая переменная X увеличится на единицу своего измерения;
- В) на сколько процентов изменится зависимая переменная Y , если независимая переменная X увеличится на 1%;
- Г) на сколько единиц изменится зависимая переменная Y , если независимая переменная X увеличится на единицу своего измерения.

6. По 12 фирмам проводился анализ взаимосвязи между ценой товара (X , ден. ед.) и объемом продаж данного товара (Y , тыс. ед.). Получено следующее уравнение парной линейной регрессии: $\hat{y} = 6,07 - 1,38x$. Выберите верное утверждение:

- А) если цена товара составит 2 ден. ед., то объем продаж товара составит в среднем 3,31 тыс. ед.;
- Б) при увеличении цены товара на 1 ден. ед. объем продаж снизится в среднем в 1,38 раза;
- В) при увеличении объема продаж товара на 1 тыс. ед. цена снизится в среднем на 1,38 тыс. руб.;
- Г) если цена товара составит 3 ден. ед., то объем продаж товара составит в среднем 5,23 тыс. ед.

7. Какая из приведенных ниже формул отражает суть метода наименьших квадратов:

А) $\sum_{i=1}^n (y_i - \hat{y}_i)^2 \rightarrow \min ;$

Б) $\sum_{i=1}^n (y_i - \hat{y}_i)^2 \rightarrow \max ;$

$$B) \sum_{i=1}^n |y_i - \hat{y}_i|^2 = 0 ;$$

$$Г) \sum_{i=1}^n (y_i - \bar{y})^2 = 0 .$$

8. Исследуется зависимость между ценой (Y , млн руб.) и общей площадью (X , м²) по 29 коттеджам в населенных пунктах Самарской области. Оценка параметров модели парной линейной регрессии проводилась с использованием функции ЛИНЕЙН:

0,12	22,31
0,04	7,02
0,24	20,83
8,60	27
3731,75	11719,49

Какое утверждение не является верным:

- А) модель парной линейной регрессии имеет вид $\hat{y} = 22,31 + 0,12x$;
 Б) общая сумма квадратов $TSS=15451,24$;
 В) линейный выборочный коэффициент корреляции между ценой и общей площадью коттеджа $r_{xy} = 0,49$;
 Г) остаточная сумма квадратов $ESS = 11719,49$.

9. По данным теста 7 коэффициент детерминации показывает:

- А) вариация цены коттеджа на 24% объясняется вариацией общей площади и на 76% вариацией неучтенных в модели факторов;
 Б) вариация цены коттеджа на 8,6% объясняется вариацией общей площади и на 91,4% вариацией неучтенных в модели факторов;
 В) вариация общей площади коттеджа на 24% объясняется вариацией цены и на 76% вариацией неучтенных в модели факторов;
 Г) вариация общей площади коттеджа на 8,6% объясняется вариацией цены и на 91,4% вариацией неучтенных в модели факторов.

10. Уравнение парной линейной регрессии имеет вид $\hat{y} = 4,3 - 1,8x$, $\hat{\sigma}_x = 1,2$, $\hat{\sigma}_y = 3,6$. Тогда выборочный коэффициент корреляции равен:

- А) -0,6;
 Б) 0,6;
 В) -0,33;
 Г) 0,33.

	1	2	3	4	5	6	7	8	9	10	11	12	13
1	2. Выявление корреляционной взаимосвязи. Оценка параметров												
2	классической модели парной линейной регрессии												
3													

2.4. Контрольные вопросы

1. В чем отличие между функциональной и корреляционной связью?
2. Каковы причины появления в модели парной линейной зависимости $y = b_0 + b_1x + \varepsilon$ случайной составляющей ε ?
3. Перечислите свойства линейного коэффициента корреляции.
4. Назовите градации шкалы Чеддока.
5. С какой целью необходимо проверить статистическую значимость линейного выборочного коэффициента корреляции?
6. Что показывают коэффициенты b_0 и b_1 в модели парной линейной зависимости $y = b_0 + b_1x + \varepsilon$?
7. Перечислите предпосылки МНК.
8. Сформулируйте теорему Гаусса - Маркова.
9. В чем суть метода наименьших квадратов (МНК)?
10. Докажите, что в модели парной линейной зависимости без константы $Y = bx + \varepsilon$

оценка коэффициента регрессии $\hat{b} = \frac{\overline{xy}}{\overline{x^2}}$.

3. ИНСТРУМЕНТ АНАЛИЗА ДАННЫХ «РЕГРЕССИЯ» В ТАБЛИЧНОМ РЕДАКТОРЕ MS EXCEL



На конкретных практических примерах реализуется алгоритм применения инструмента анализа данных «РЕГРЕССИЯ», осуществляется верификация парной линейной регрессионной модели с использованием результатов данного инструмента.

3.1. Решение типовых задач

Для иллюстрации возможностей инструмента анализа данных «РЕГРЕССИЯ» табличного редактора Excel на примере данных табл. 2.1 проведем оценку параметров парной линейной регрессионной модели.

Шаг 1. В главном меню выбираем **Данные / Анализ данных**.

Шаг 2. В появившемся окне **Анализ данных** выберем инструмент анализа данных «РЕГРЕССИЯ» (рис. 3.1).

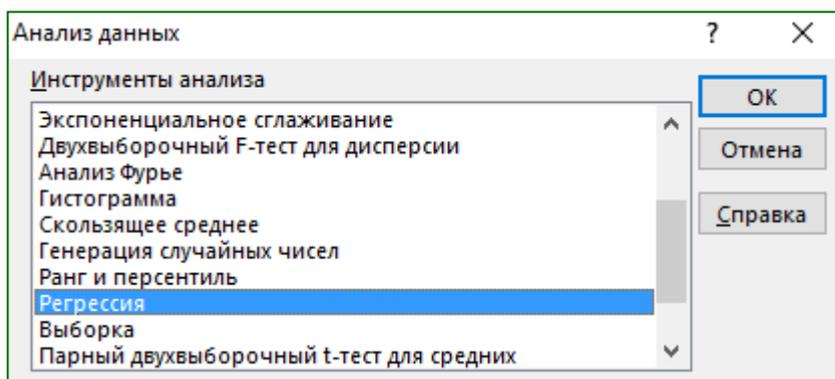


Рис. 3.1. Выбор инструмента анализа данных

Шаг 3. В окне установок данного инструмента необходимо указать значения зависимой переменной «**Входной интервал Y**» и значения независимой переменной «**Входной интервал X**», как показано на рис. 3.2. Значения анализируемых переменных выделяются в соответствующих диалоговых полях с наименованиями для получения резуль-

3. Инструмент анализа данных «РЕГРЕССИЯ» в табличном редакторе MS Excel

татов в более информативном виде. Соответственно, в этом случае необходимо активировать опцию «**Метки**». Опция «**Уровень надежности**» активируется при необходимости построения доверительных интервалов для теоретических коэффициентов регрессии при требуемой доверительной вероятности (надежности). По умолчанию инструмент анализа данных «**РЕГРЕССИЯ**» выдает результаты построения интервальных оценок с надежностью 95%.

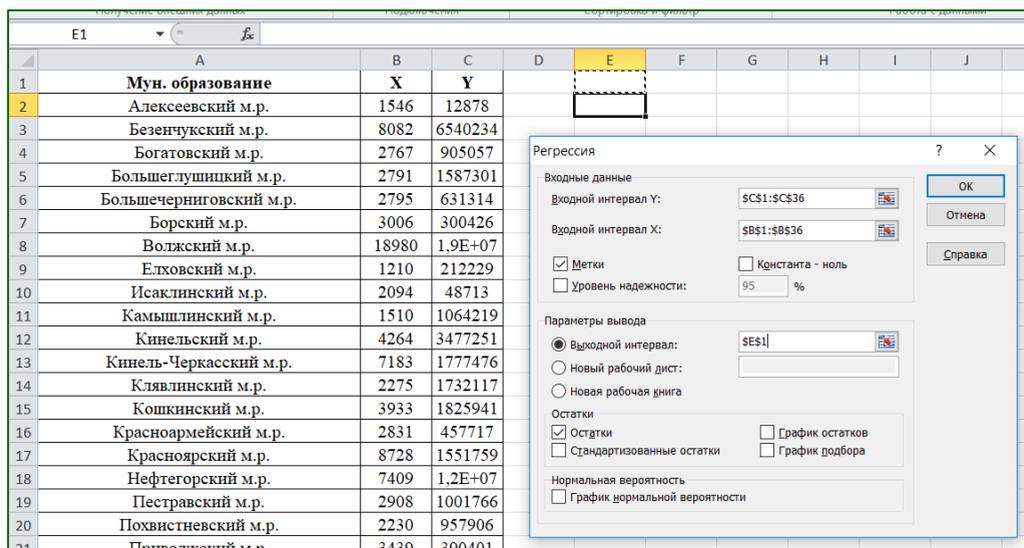


Рис. 3.2. Установки инструмента анализа данных «РЕГРЕССИЯ»

После нажатия кнопки ОК получаем результаты регрессионного анализа (рис. 3.3).

В приведенных ниже результатах обратимся к таблице на рис. 3.3, в которой рассмотрим столбец «**Коэффициенты**». В этом столбце представлены оценки параметров линейного регрессионного уравнения вида $Y = b_0 + b_1X + \varepsilon$. Выборочное уравнение регрессии имеет вид $\hat{y} = 708034,96 + 307x$.

Шаг 4. Для оценки статистической значимости параметров уравнения регрессии потребуется проанализировать стандартные ошибки коэффициентов, вычисленные в соответствующем столбце (рис. 3.3).

Необходимо проверить статистическую гипотезу вида $H_0 : b_j = 0; j = \overline{0,1}$. Справедливость гипотезы проверяется с помощью t -статистики Стьюдента. Наблюдаемое значение данной случайной величины вычисляется как $t_{набл} = \frac{\hat{b}_j}{S_{\hat{b}_j}}$. Эти значения рассчитаны в соответствующем столбце предпоследней таблицы в результатах инструмента «**РЕГРЕССИЯ**» - «**t-статистика**» (см. рис. 3.3).

L23	A	B	C	D	E	F	G	H	I	J	K
		X	Y		ВЫВОД ИТОГОВ						
1	Мун. образование	1546	12878								
2	Алексеевский м.р.	8082	6540234								
3	Белегучевский м.р.	2767	905057		Регрессионная статистика						
4	Богатовский м.р.	2791	1587301		Множественный R	0,607187654					
5	Большеглушицкий м.р.	2795	631314		R-квадрат	0,368676847					
6	Большечерниговский м.р.	3006	300426		Нормированный R-квадрат	0,349545842					
7	Борский м.р.	18980	1,9E+07		Стандартная ошибка	3481735,217					
8	Волжский м.р.	1210	212229		Наблюдения	35					
9	Елховский м.р.	2094	48713								
10	Исаковский м.р.	1510	1064219		Дисперсионный анализ						
11	Камышлинский м.р.	4264	3477251			df	SS	MS	F	Значимость F	
12	Кинельский м.р.	7183	1777476		Регрессия	1	2,33614E+14	2,33614E+14	19,27117021	0,000109901	
13	Кинель-Черкасский м.р.	2275	1732117		Остаток	33	4,00042E+14	1,21225E+13			
14	Клявлинский м.р.	3933	1825941		Итого	34	6,33656E+14				
15	Кошкинский м.р.	2831	457717								
16	Красноармейский м.р.	8728	1551759								
17	Красноярский м.р.	7409	1,2E+07		У-пересечение	708034,9585	780223,0876	0,907477579	0,370733745	-879340,8486	2295410,766
18	Нефтегорский м.р.	2908	1001766		X	306,9986989	69,93305349	4,3898941	0,000109901	164,7188318	449,2785661
19	Петровский м.р.	2230	957906								
20	Похвостневский м.р.	3439	390401								
21	Привольский м.р.	9667	6589916		ВЫВОД ОСТАТКА						
22	Сергеевский м.р.	9676	2487599								
23	Старопольский м.р.	3327	1079027		Наблюдение	Предсказанное Y	Остатки				
24	Сызранский м.р.	2240	613808			1	1182654,947			-1169776,947	
25	Хворостянский м.р.	2521	562301			2	3189198,443			3351035,557	
26	Челно-Вершинский м.р.										

Рис. 3.3. Результаты применения инструмента анализа данных «РЕГРЕССИЯ»

3. Инструмент анализа данных «РЕГРЕССИЯ» в табличном редакторе MS Excel

Критическое значение распределения Стьюдента вычисляют с помощью функции **СТЮДРАСПОБР**, исходя из заданного уровня значимости α и параметра распределения (числа степеней свободы) $k = n - 2$ (рис. 3.4).

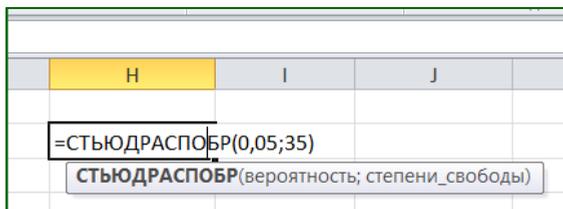


Рис. 3.4. Расчет критического значения t -статистики Стьюдента

При выполнении условия $|t_{набл}| > t_{кр.дв}(\alpha; k)$ оценку \hat{b}_j можно считать статистически значимой.

Таким образом, исходя из данных рис. 3.4 можно утверждать, что оценка коэффициента \hat{b}_1 является статистически значимой, а \hat{b}_0 - не значимой.

15					
16		Коэффициенты	Стандартная ошибка	t -статистика	P -Значение
17	Y-пересечение	708034,9585	780223,0876	0,907477579	0,370733745
18	X	306,9986989	69,93305349	4,3898941	0,000109901
19					
20				t кр. дв.	
21				2,030107928	

Рис. 3.5. Проверка статистической значимости оценок коэффициентов регрессии

На практике нередко используют **p -значение** (рис. 3.5), которое показывает вероятность принять или отвергнуть гипотезу о равенстве нулю (незначимости) соответствующего коэффициента. При этом предполагается, что ошибки имеют нормальное или асимптотически нормальное распределение. Значения вероятности, указанные в таблице на рис. 3.5, известны в статистике как уровни значимости α . Если **p -значение** ниже уровня значимости α , то гипотеза H_0 отвергается и соответствующая оценка коэффициента признается статистически значимой.

Общепринятыми уровнями значимости являются $\alpha = 0,01$; $\alpha = 0,05$; $\alpha = 0,1$.

Шаг 5. Проверка общего качества уравнения регрессии проводится с помощью таблицы **«Дисперсионный анализ»** инструмента анализа данных **«РЕГРЕССИЯ»** (рис. 3.3).

В качестве основной используется гипотеза о незначимости регрессионного уравнения в целом: $H_0 : R^2 = 0$. Данная гипотеза проверяется с помощью случайной величины F , имеющей распределение Фишера - Снедекора с числами степеней свободы

$k_1 = m$; $k_2 = n - m - 1$, где m - число независимых переменных в модели. Наблюдаемое значение случайной величины рассчитывается по формуле $F_{набл} = \frac{RSS(n - m - 1)}{ESS \cdot m}$.

Данное значение вычислено в столбце «F» таблицы «Дисперсионный анализ».

Критическое значение F-статистики вычисляется с помощью функции **ФРАСПОБР** исходя из заданного уровня значимости α и параметров распределения (числа степеней свободы) $k_1 = m$; $k_2 = n - m - 1$ (рис. 3.6).

✓ fx	=ФРАСПОБР(0,05;1;35)							
	B	C	D	E	F	G	H	I
	Y	X						
	12878	1546			=ФРАСПОБР(0,05;1;35)			
	6540234	8082						

Рис. 3.6. Расчет критического значения F-статистики Фишера

При выполнении условия $F_{набл} > F_{кр}(\alpha; k_1; k_2)$ построенная регрессионная модель считается статистически значимой. По данным результатов инструмента «РЕГРЕССИЯ», $F_{набл} = 19,27$, что много меньше $F_{кр}(\alpha = 0,05; k_1 = 1; k_2 = 35)$. Соответственно, построенная эконометрическая модель может быть признана статистически значимой.

ЗАМЕЧАНИЕ: число степеней свободы указано в результатах инструмента «РЕГРЕССИЯ» в таблице «Дисперсионный анализ» в столбце «df» (degrees of freedom).

линейн	=5G\$17+5G\$18*B2						
	A	B	C	D	E	F	G
1	Мун. образование	X	Y	Y^2		ВЫВОД ИТОГОВ	
2	Алексеевский м.р.	1546		=5G\$17+5G\$18*B2			
3	Безенчукский м.р.	8082	6540234	3189198		Регрессионная статистика	
4	Богатовский м.р.	2767	905057	1557500		Множественный R	0,607187654
5	Большеглушицкий м.р.	2791	1587301	1564868		R-квадрат	0,368676847
6	Большечерниговский м.р.	2795	631314	1566096		Нормированный R-квадрат	0,349545842
7	Борский м.р.	3006	300426	1630873		Стандартная ошибка	3481735,217
8	Волжский м.р.	18980	18850964	6534870		Наблюдения	35
9	Елховский м.р.	1210	212229	1079503			
10	Исаклинский м.р.	2094	48713	1350890		Дисперсионный анализ	
11	Камышлинский м.р.	1510	1064219	1171603			df
12	Кинельский м.р.	4264	3477251	2017077		Регрессия	1
13	Кинель-Черкасский м.р.	7183	1777476	2913207		Остаток	33
14	Клявлинский м.р.	2275	1732117	1406457		Итого	34
15	Кошкинский м.р.	3933	1825941	1915461			
16	Красноармейский м.р.	2831	457717	1577148			Коэффициенты
17	Красноярский м.р.	8728	1551759	3387520		Y-пересечение	708034,9585
18	Нефтегорский м.р.	7409	12464092	2982588		X	306,9986989

Рис. 3.7. Вычисление расчетных значений зависимой переменной

	1	2	3	4	5	6	7	8	9	10	11	12	13
1			3. Инструмент анализа данных «РЕГРЕССИЯ»										
2			в табличном редакторе MS Excel										
3													

Шаг 6. Вычисление расчетных значений зависимой переменной \hat{y}_i будем осуществлять в дополнительном столбце (рис. 3.7), используя полученное выборочное уравнение регрессии $\hat{y} = 708034,96 + 307x$.

При вычислении следует придать ячейкам, содержащим значения оценок коэффициентов регрессии, универсальный адрес путем нажатия клавиши F4.

3.2. Задачи для самостоятельного выполнения

Задача 10. Используя данные задачи 7, оцените параметры модели парной линейной регрессии с помощью инструмента анализа данных «РЕГРЕССИЯ».

Задача 11. Используя данные задачи 10, проверьте статистическую значимость оценок коэффициентов регрессии (уровень значимости принять равным 0,05).

Задача 12. Используя данные задачи 10, проверьте статистическую значимость модели в целом (уровень значимости принять равным 0,05).

Задача 13. Используя данные задачи 10, вычислите остатки регрессии.

3.3. Тесты

1. Проверить значимость параметров уравнения регрессии можно, используя:

- А) t -статистику;
- Б) коэффициент детерминации;
- В) коэффициент корреляции;
- Г) коэффициент ковариации.

2. Для оценки значимости коэффициента детерминации используется:

- А) критерий Смирнова - Граббса;
- Б) критерий Стьюдента;
- В) критерий Спирмена;
- Г) критерий Фишера - Снедекора;
- Д) метод Барта.

	1	2	3	4	5	6	7	8	9	10	11	12	13
1			<i>Эконометрика в табличном редакторе MS Excel</i>										
2													
3													

3. В качестве основной гипотезы H_0 для проверки по F -тесту при оценке модели $y = b_0 + b_1x$ в целом можно указать условие вида:

- А) $b_0 = b_1 = 0$;
- Б) $b_0 \neq b_1 \neq 0$;
- В) $b_1 = 0$;
- Г) $b_0 \neq 0$.

4. В качестве основной гипотезы H_0 для проверки по t -тесту статистической значимости оценок коэффициентов регрессии в модели вида $y = b_0 + b_1x + \varepsilon$ берется условие:

- А) $b_1 = b_2 = 0$;
- Б) $b_1 \neq b_2 \neq 0$;
- В) $b_1 \neq 0, b_2 \neq 0$;
- Г) нет правильного ответа.

5. Случайная величина $\frac{ESS / m}{RSS / (n - m - 1)}$, где m - число объясняющих переменных,

n - число наблюдений, имеет распределение:

- А) Фишера - Снедекора;
- Б) Хи-квадрат;
- В) Стьюдента;
- Г) Пирсона.

6. При проверке значимости оценки коэффициента регрессии \hat{b}_j t -статистика имеет:

- А) нормальное распределение;
- Б) распределение Стьюдента с $(n - m - 1)$ степенями свободы;
- В) распределение Фишера с m и $(n - m - 1)$ степенями свободы.

7. При проверке гипотезы $H_0: b_1 = 0$ оказалось, что $|t_{набл}| > t_{крит}$. Какое из приведенных ниже утверждений справедливо:

- А) $b_1 = 0$;
- Б) параметр b_1 не равен 0;
- В) параметр b_1 не равен нулю с вероятностью ошибки α ;
- Г) параметр b_1 равен нулю с вероятностью ошибки α .

8. Значение оценки коэффициента регрессии \hat{b}_1 было получено по МНК равным 12,4, а его стандартная ошибка $S_{\hat{b}_1}$ равна 2,34. Табличное значение t -критерия Стьюдента для данной выборки равно 2,20, тогда:

- А) оценка \hat{b}_1 будет статистически не значима;

	1	2	3	4	5	6	7	8	9	10	11	12	13
1			3. Инструмент анализа данных «РЕГРЕССИЯ»										
2			в табличном редакторе MS Excel										
3													

- Б) оценка \hat{b}_0 будет статистически значима;
 В) не представляется возможным вычислить.

9. В результате оценки параметров получены следующие результаты: $\hat{b}_0 = 1,845$, стандартная ошибка $S_{\hat{b}_0}$ параметра равна 0,471. Тогда наблюдаемое значение t -критерия равно:

- А) 3,126;
 Б) 0,471;
 В) 3,917.

10. В ходе оценки уравнения регрессии в целом было получено наблюдаемое значение F -критерия, равное 3,245, при этом табличное значение равно 3,021. Тогда можно сделать вывод, что:

- А) построенное регрессионное уравнение статистически не значимо;
 Б) построенное регрессионное уравнение статистически значимо;
 В) оценки коэффициентов регрессии не значимы.

11. В результате применения процедуры «РЕГРЕССИЯ» получены p -значения оценок коэффициентов регрессии, приведенные в таблице:

Коэффициенты	p -значения
Y-пересечение	0,05634362
X	0,02983635

По таблице регрессионной статистики можно сделать следующие выводы:

- А) все оценки коэффициентов регрессии значимы при уровне 5%;
 Б) оценка \hat{b}_0 статистически не значима, а оценка \hat{b}_1 статистически значима при уровне 5%;
 В) оценка \hat{b}_0 статистически значима, а оценка \hat{b}_1 статистически не значима при уровне 3%;
 Г) все оценки коэффициентов регрессии не значимы при уровне 5%.

12. В результате применения процедуры «РЕГРЕССИЯ» получено F -значение, равное 3,6E-15. Тогда в результате проверки на значимость уравнения в целом получим ответ:

- А) построенное уравнение регрессии значимо в целом при уровне 5%;
 Б) построенное уравнение регрессии в целом не значимо при уровне 5%;
 В) построенное уравнение регрессии значимо в целом при уровне 3%;
 Г) построенное уравнение регрессии в целом не значимо ни при каком уровне значимости.
 Д) построенное уравнение регрессии значимо в целом при уровне 1%.

Выберите несколько вариантов ответа.

13. В результате применения процедуры «РЕГРЕССИЯ» получены значения нижних и верхних границ 95% и 98% доверительных интервалов, приведенных в таблице:

	Нижние 95%	Верхние 95%	Нижние 98,0%	Верхние 98,0%
Y-пересечение	0,220435	26,85229	-3,07171	30,14443
Переменная X ₁	9,091219	13,01787	8,605819	13,50327

Тогда можно сделать следующие выводы по статистической значимости оценок коэффициентов регрессии:

- А) оценки коэффициентов регрессии значимы при уровне значимости 5%;
- Б) оценки коэффициентов регрессии значимы при уровне 2%;
- В) оценка \hat{b}_0 статистически не значима, а оценка \hat{b}_1 статистически значима при уровне 2%;
- Г) оценка \hat{b}_0 статистически значима, а оценка \hat{b}_1 статистически не значима при уровне 5%;
- Д) оценка \hat{b}_0 статистически не значима, а оценка \hat{b}_1 статистически значима при уровне 5%.

Выберите несколько вариантов ответа.

14. В процедуре «РЕГРЕССИЯ» не выводится по умолчанию:

- А) таблица дисперсионного анализа;
- Б) ВЫВОД ОСТАТКА;
- В) ВЫВОД ИТОГОВ;
- Г) регрессионная статистика.

3.4. Контрольные вопросы

- Какой статистический критерий используется для проверки значимости оценок параметров модели парной линейной регрессии?
- Перечислите показатели качества модели парной линейной регрессии.
- Перечислите свойства коэффициента детерминации, сформулируйте его экономический смысл.
- Приведите различные формулы расчета коэффициента детерминации для модели парной линейной регрессии.
- Какой статистический критерий используется для проверки значимости модели парной линейной регрессии в целом?

	1	2	3	4	5	6	7	8	9	10	11	12	13
1			3. Инструмент анализа данных «РЕГРЕССИЯ»										
2			в табличном редакторе MS Excel										
3													

6. Приведите схему построения доверительного интервала для параметров парной линейной регрессии.
7. Как изменится доверительный интервал оценки параметра линейной регрессии, если достоверность а) увеличится; б) уменьшится?
8. Приведите экономический смысл интервальных оценок коэффициентов регрессии.
9. Как рассчитать точечный прогноз в модели парной линейной регрессии?
10. Приведите схему построения интервального прогноза для случая парной линейной регрессии.

4. ОЦЕНИВАНИЕ МОДЕЛИ МНОЖЕСТВЕННОЙ ЛИНЕЙНОЙ РЕГРЕССИИ С ИСПОЛЬЗОВАНИЕМ ИНСТРУМЕНТА АНАЛИЗА ДАННЫХ «РЕГРЕССИЯ»



Рассмотрены особенности оценивания модели множественной линейной регрессии с помощью инструмента анализа данных «РЕГРЕССИЯ». Реализован один из способов отбора факторных признаков в эконометрическую модель.

4.1. Решение типовых задач

В качестве иллюстрационного примера особенностей построения и анализа множественного линейного уравнения регрессии использован статистический массив (табл. 4.1) по субъектам Центрального федерального округа РФ за 2016 г. в составе показателей, образующих следующую систему:

Y - валовой региональный продукт, руб.;

X_1 - стоимость основных фондов (по полной учетной стоимости), млн руб.;

X_2 - объем производства по виду деятельности «Добыча полезных ископаемых» (в фактически действовавших ценах), млн руб.;

X_3 - объем производства по виду деятельности «Обрабатывающие производства» (в фактически действовавших ценах), млн руб.;

X_4 - объем производства по виду деятельности «Производство и распределение электроэнергии, газа и воды» (в фактически действовавших ценах), млн руб.;

X_5 - продукция сельского хозяйства (в фактически действовавших ценах), млн руб.

Таблица 4.1

Система показателей по субъектам Центрального федерального округа РФ, 2016 г.

Субъект ЦФО РФ	Y	X_1	X_2	X_3	X_4	X_5
Белгородская область	730562	1400837	89048	597751	26504	226056,3
Брянская область	285847,9	733055	281	176003	15439	78437,4
Владимирская область	392051,8	787525	3068	406936	33727	30834,9
Воронежская область	841375,7	1534644	5576	408214	59609	199325,4
Ивановская область	179633,4	553289	892	120938	29674	14987

1 2 3 4 5 6 7 8 9 10 11 12 13
 1 4. **Оценивание модели множественной линейной регрессии с использованием**
 2 **инструмент анализа данных «РЕГРЕССИЯ»**
 3

Окончание табл. 4.1

Субъект ЦФО РФ	Y	X ₁	X ₂	X ₃	X ₄	X ₅
Калужская область	373403,5	912722	2983	536624	20076	33458
Костромская область	160705,4	419515	373	107792	36246	17090
Курская область	364601,7	805024	49246	160486	58522	124357,1
Липецкая область	470239,3	1178985	4941	572356	25824	106324,7
Московская область	3565258	7237874	12685	2126307	254767	98372,1
Орловская область	213924	472195	134	103982	13089	68568
Рязанская область	336973,5	927211	1503	252967	37227	51048
Смоленская область	262318	829388	1423	178891	54349	21243,4
Тамбовская область	311433,4	798260	127	137330	15520	108896,2
Тверская область	359345,1	1212653	1088	216368	85524	32882,2
Тульская область	517740,8	1033088	4336	564749	47813	56294,1
Ярославская область	469804,9	1226233	996	316472	33450	31852,3

Источник: URL: www.gks.ru.

Стоит отметить, что из рассмотрения исключен город Москва, так как значения показателей по этому субъекту РФ существенно отличаются от значений по остальным субъектам ЦФО и, как следствие, будут вносить существенное искажение в оценки параметров регрессионного уравнения.

Шаг 1. Для оценивания множественного линейного уравнения регрессии, включающего в себя все 5 независимых переменных X_1, X_2, X_3, X_4, X_5 , необходимо выбрать **Данные/Анализ данных**.

Теоретическое уравнение регрессии в этом случае имеет вид

$$Y = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4 + b_5X_5 + \varepsilon.$$

Шаг 2. В появившемся окне **Анализ данных** выбираем инструмент анализа данных **«РЕГРЕССИЯ»** (рис. 4.1).

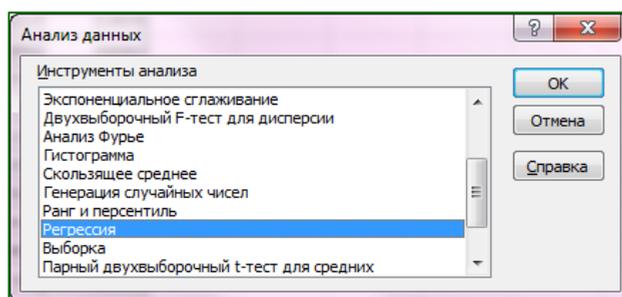


Рис. 4.1. Выбор инструмента анализа данных «РЕГРЕССИЯ»

	A	B	C	D	E	F	G
	Y	X1	X2	X3	X4	X5	
1	Субъект ЦФО РФ						
2	Белгородская область	730562	1400837	89048	597751	26504	226056,3
3	Брянская область	285847,9	733055	281	176003	15439	78437,4
4	Владимирская область	392051,8	787525	3068	406936	33727	30834,9
5	Воронежская область	841375,7	1534644	5576	408214	59609	199325,4
6	Ивановская область	179633,4	553289	892	120938	29674	14987
7	Калужская область	373403,5	912722	2983	536624	20076	33458
8	Костромская область	160705,4	419515	373	107792	36246	17090
9	Курская область	364601,7	805024	49246	160486	58522	124357,1
10	Липецкая область	470239,3	1178985	4941	572356	25824	106324,7
11	Московская область	3565258	7237874	12685	2126307	254767	98372,1
12	Орловская область	213924	472195	134	103982	13089	68568
13	Рязанская область	336973,5	927211	1503	252967	37227	51048
14	Смоленская область	262318	829388	1423	178891	54349	21243,4
15	Тамбовская область	311433,4	798260	127	137330	15520	108896,2
16	Тверская область	359345,1	1212653	1088	216368	85524	32882,2
17	Тульская область	517740,8	1033088	4336	564749	47813	56294,1
18	Ярославская область	469804,9	1226233	996	316472	33450	31852,3

Регрессия

Входные данные: \$B\$1:\$B\$18
 Входной интервал Y: \$C\$1:\$G\$18
 Входной интервал X: Метки Константа - ноль
 Уровень надежности: 95 %

Параметры вывода:
 Выходной интервал:
 Новый рабочий лист:
 Новая рабочая книга

Остатки:
 Остатки График остатков
 Стандартизованные остатки График подбора

Нормальная вероятность:
 Нормальная вероятность График нормальной вероятности

OK Отмена Справка

Рис. 4.2. Установки инструмента анализа данных «РЕГРЕССИЯ» при оценивании множественного уравнения регрессии

1 2 3 4 5 6 7 8 9 10 11 12 13
 1 4. Оценивание модели множественной линейной регрессии с использованием
 2 инструмента анализа данных «РЕГРЕССИЯ»
 3

Шаг 3. Указываем входной интервал для зависимой переменной Y и независимых переменных X_1, X_2, X_3, X_4, X_5 (рис. 4.2).

ЗАМЕЧАНИЕ: если в окне «РЕГРЕССИЯ» не установить флажок в поле «**Метки**», то при выводе результатов не будут указаны обозначения переменных. В поле «**Входной интервал X**», необходимо выделять все 5 независимых переменных X_1, X_2, X_3, X_4, X_5 .

Результаты оценки параметров представлены на рис. 4.3.

	A	B	C	D	E	F	G	H	I
1	Вывод итогов								
2									
3	Регрессионная статистика								
4	Множественный R	0,997482029							
5	R-квадрат	0,994970398							
6	Нормированный R-квадрат	0,992684215							
7	Стандартная ошибка	67594,32749							
8	Наблюдения	17							
9									
10	Дисперсионный анализ								
11		df	SS	MS	F	Значимость F			
12	Регрессия	5	9,94236E+12	1,98847E+12	435,2103261	2,98698E-12			
13	Остаток	11	50258924197	4568993109					
14	Итого	16	9,99262E+12						
15									
16		Коэффициенты	Стандартная ошибка	t-статистика	P-Значение	Нижние 95%	Верхние 95%	Нижние 95,0%	Верхние 95,0%
17	Y-пересечение	-138424,7282	35573,60286	-3,891220374	0,002514294	-216721,7002	-60127,75625	-216721,7002	-60127,75625
18	X1	0,418728192	0,075032377	5,580633472	0,000165016	0,253583045	0,583873339	0,253583045	0,583873339
19	X2	-0,552742902	1,06764987	-0,517719261	0,614906448	-2,902624423	1,797138619	-2,902624423	1,797138619
20	X3	0,217626958	0,149019473	1,460392751	0,1721473	-0,11036269	0,545616606	-0,11036269	0,545616606
21	X4	0,426345513	1,197326368	0,356081286	0,728515012	-2,208952056	3,061643082	-2,208952056	3,061643082
22	X5	0,898299049	0,433336286	2,072983679	0,062450423	-0,055467686	1,852065784	-0,055467686	1,852065784
23									
24									
25									
26	Вывод остатка								
27									
28		Наблюдение	Предсказанное Y	Остатки					
29		1	743377,3184	-12815,31839					
30		2	283716,3333	2131,566666					
31		3	320276,936	71774,86397					

Рис. 4.3. Результаты оценки параметров множественного линейного уравнения регрессии

Согласно результатам инструмента «РЕГРЕССИЯ», между рассматриваемыми переменными существует сильная корреляционная связь, так как множественный коэффициент корреляции равен 0,997.

Выборочное уравнение регрессии имеет вид

$$\hat{y} = -138424,72 + 0,42X_1 - 0,55X_2 + 0,21X_3 + 0,42X_4 + 0,89X_5.$$

Статистическую значимость оценок коэффициентов регрессии $\hat{b}_0, \hat{b}_1, \hat{b}_2, \hat{b}_3, \hat{b}_4, \hat{b}_5$ проверим с использованием **p-значения**, указанного в третьей таблице итогов инстру-

	A	B	C	D	E	F	G	H	I
1	ВЫВОД ИТОГОВ								
2									
3	Регрессионная статистика								
4	Множественный R	0,996942821							
5	R-квадрат	0,993894989							
6	Нормированный R-квадрат	0,993022844							
7	Стандартная ошибка	66011,40672							
8	Наблюдения	17							
9									
10	Дисперсионный анализ								
11		df	SS	MS	F	Значимость F			
12	Регрессия	2	9,93162E+12	4,96581E+12	1139,599023	3,16086E-16			
13	Остаток	14	61005081436	4357505817					
14	Итого	16	9,99262E+12						
15									
16		Коэффициенты	Стандартная ошибка	t-статистика	P-Значение	Верхние 95%	Нижние 95%	Верхние 95,0%	Верхние 95,0%
17	У-пересечение	-124605,0673	27653,75655	-4,505900204	0,000493587	-183916,4762	-65293,65836	-183916,4762	-65293,65836
18	X1	0,497356997	0,010801892	46,04350561	1,09877E-16	0,474189242	0,520524752	0,474189242	0,520524752
19	X5	0,754266429	0,272161935	2,771388393	0,015001222	0,170537133	1,337995724	0,170537133	1,337995724
20									
21									
22									
23	ВЫВОД ОСТАТКА								
24	Наблюдение	Предказанное Y	Остатки						
26		1	742617,694	-12055,69403					
27		2	299147,6635	-13299,76353					
28		3	290333,7315	101718,0685					

Рис. 4.4. Результаты оценивания множественного линейного уравнения регрессии с двумя зависимыми переменными X₁, X₅

	1	2	3	4	5	6	7	8	9	10	11	12	13
1	4. <i>Оценивание модели множественной линейной регрессии с использованием</i>												
2	<i>инструмента анализа данных «РЕГРЕССИЯ»</i>												
3													

мента «РЕГРЕССИЯ» (см. рис. 4.3). Примем уровень значимости α равным 0,08. В этом случае ряд оценок коэффициентов регрессии следует признать статистически не значимыми, так как соответствующее им ***p-значение*** превышает указанный уровень значимости. Незначимые оценки: $\hat{b}_2, \hat{b}_3, \hat{b}_4$. Оценки $\hat{b}_0, \hat{b}_1, \hat{b}_5$ являются статистически значимыми.

Таким образом, необходимо изменить спецификацию модели, исключив из правой части уравнения независимые переменные X_2, X_3, X_4 .

Шаг 4. Исключим указанные в **шаге 3** независимые переменные из анализа и проведем процедуру оценки следующего регрессионного уравнения (рис. 4.4):

$$Y = b_0 + b_1X_1 + b_5X_5 + \varepsilon.$$

ЗАМЕЧАНИЕ: так как независимые переменные в исходной таблице статистических показателей по субъектам ПФО стоят по порядку, при заполнении поля «**Входной интервал X**» инструмента анализа данных «РЕГРЕССИЯ» необходимо удалить лишние переменные (не участвующие в анализе).

Итоги инструмента анализа данных «РЕГРЕССИЯ» позволяют записать выборочное уравнение множественной регрессии вида

$$\hat{y} = -124605,07 + 0,497X_1 + 0,754X_5.$$

Оценки коэффициентов $\hat{b}_0, \hat{b}_1, \hat{b}_5$ признаны статистически значимыми, так как соответствующее им ***p-значение*** не превышает заданного уровня значимости $\alpha = 0,08$.

Коэффициент детерминации $R^2 = 0,994$. Гипотеза о статистической незначимости модели в целом отвергается, так как значение в поле «Значимость F» много меньше заданного уровня значимости $\alpha = 0,08$. Таким образом, валовой региональный продукт субъектов Центрального федерального округа на 99,4% зависит от стоимости основных фондов и от объема сельскохозяйственного производства, а на 0,6% - от прочих факторных признаков, не включенных в модель.

Шаг 5. Используем полученное уравнение регрессии вида $\hat{y} = -124605,07 + 0,497X_1 + 0,754X_5$ для имитационных расчетов. При этом оценим два варианта:

- 1) максимальные значения независимых переменных X_1, X_5 увеличатся на 10%;
- 2) минимальные значения независимых переменных X_1, X_5 снизятся на 10%.

Для определения максимального значения в ряду значений переменной следует использовать функцию **МАКС**, указывая в диалоговом поле все значения ряда. Для определения минимального значения - функцию **МИН** (рис. 4.5).

Увеличение максимальных показателей приведет к росту ВРП выше, чем в Московской области, а минимизация значений независимых переменных - к значению ВРП, не превышающему ВРП Костромской области.

Так как оценки коэффициентов регрессионного уравнения являются числами именованными (имеющими размерность), невозможно оценить, какой из факторов оказывает

наибольшее влияние на резульativenную переменную. Для этих целей следует рассчитать частный коэффициент эластичности: $\Theta = \hat{b}_j \cdot \frac{\bar{X}}{\bar{Y}}$.

J	K	L	M
Y	X1	X5	
730 562,00	1400837	226056	
285 847,90	733055	78437,4	
392 051,80	787525	30834,9	
841 375,70	1534644	199325	
179 633,40	553289	14987	
373 403,50	912722	33458	
160 705,40	419515	17090	
364 601,70	805024	124357	
470 239,30	1178985	106325	
3 565 258,00	7237874	98372,1	
213 924,00	472195	68568	
336 973,50	927211	51048	
262 318,00	829388	21243,4	
311 433,40	798260	108896	
359 345,10	1212653	32882,2	
517 740,80	1033088	56294,1	
469 804,90	1226233	31852,3	
Максимум	7237874	226056	
Минимум	419515	14987	
Максимум +10%	7961661,4	248662	
Минимум -10%	377563,5	13488,3	
Y^ (предсказанное) max+10%	4 018 837,06		
Y^ (предсказанное) мин-10%	=-124605,1+0,497*K24+0,75*L24		

Рис. 4.5. Имитационный расчет возможных значений зависимой переменной при заданных значениях независимых переменных

1 2 3 4 5 6 7 8 9 10 11 12 13
 1 4. **Оценивание модели множественной линейной регрессии с использованием**
 2 **инструмент анализа данных «РЕГРЕССИЯ»**
 3

ЛИНЕЙН $=D21*(D20/B20)$					
	A	B	C	D	E
1	Субъект ЦФО РФ	Y	X1	X5	
2	Белгородская область	730 562,00	1400837	226056,3	
3	Брянская область	285 847,90	733055	78437,4	
4	Владимирская область	392 051,80	787525	30834,9	
5	Воронежская область	841 375,70	1534644	199325,4	
6	Ивановская область	179 633,40	553289	14987	
7	Калужская область	373 403,50	912722	33458	
8	Костромская область	160 705,40	419515	17090	
9	Курская область	364 601,70	805024	124357,1	
10	Липецкая область	470 239,30	1178985	106324,7	
11	Московская область	3 565 258,00	7237874	98372,1	
12	Орловская область	213 924,00	472195	68568	
13	Рязанская область	336 973,50	927211	51048	
14	Смоленская область	262 318,00	829388	21243,4	
15	Тамбовская область	311 433,40	798260	108896,2	
16	Тверская область	359 345,10	1212653	32882,2	
17	Тульская область	517 740,80	1033088	56294,1	
18	Ярославская область	469 804,90	1226233	31852,3	
19					
20	Среднее	578 542,26	1 297 794,00	76 472,18	
21	b ^{4j}		0,497	0,75	
22	Коэффициент эластичности		1,11487728	=D21*(D20/B20)	
23					

Рис. 4.6. Оценка значений частных коэффициентов эластичности

Для решения поставленной задачи предварительно необходимо оценить средние значения переменных X_1 , X_5 (рис. 4.6).

Произведенные расчеты показали, что коэффициент эластичности при переменной X_1 получен больше, чем при переменной X_5 (1,11 против 0,099). Таким образом, можно утверждать, что стоимость основных фондов оказывает большее влияние на ВРП регионов Центрального федерального округа, чем объем сельскохозяйственного производства.

4.2. Задача для самостоятельного выполнения

Задача 14. По субъектам Приволжского федерального округа РФ за ряд лет собран статистический массив по показателям, образующим следующую систему (www.gks.ru):

Y - валовой региональный продукт, руб.;

X_1 - стоимость основных фондов (по полной учетной стоимости), млн руб.;

X_2 - объем производства по виду деятельности «Добыча полезных ископаемых» (в фактически действовавших ценах), млн руб.;

1	2	3	4	5	6	7	8	9	10	11	12	13
2	<i>Эконометрика в табличном редакторе MS Excel</i>											
3												

X_3 - объем производства по виду деятельности «Обрабатывающие производства» (в фактически действовавших ценах), млн руб.;

X_4 - объем производства по виду деятельности «Производство и распределение электроэнергии, газа и воды» (в фактически действовавших ценах), млн руб.;

X_5 - продукция сельского хозяйства (в фактически действовавших ценах), млн руб.

Оцените модель множественной линейной регрессии с применением инструмента анализа данных «РЕГРЕССИЯ». Проверьте статистическую значимость оценок коэффициентов регрессии ($\alpha = 0,07$), а также значимость модели в целом ($\alpha = 0,07$). Проведите имитационные расчеты при условии увеличения максимальных значений независимых переменных на 5% и снижении минимальных значений независимых переменных на 5%. Оцените значение частных коэффициентов эластичности, сделайте соответствующие выводы.

Вариант 1, 2012 г.

Субъект ПФО РФ	Y	X ₁	X ₂	X ₃	X ₄	X ₅
Республика Башкортостан	1149385	1799031	133473	924132	88443	106804
Республика Марий Эл	117201,1	286499	360	85833	9694	25194
Республика Мордовия	134315,6	406047	166	101320	10212	39225
Республика Татарстан	1437001	3110418	363156	995068	109757	150106
Удмуртская Республика	372782,7	817042	124512	171789	27220	46154
Чувашская Республика	217821,1	589437	695	131448	20790	32625
Пермский край	860342,7	2199176	213668	760028	90085	36029
Кировская область	208505,4	639525	910	131145	28303	26212
Нижегородская область	842195,5	1947537	1876	898070	80661	47899
Оренбургская область	628563,6	1377934	313118	211629	83189	71097
Пензенская область	239962,5	669188	1878	115379	16713	43812
Самарская область	937434,5	2173528	171702	757031	98619	58193
Саратовская область	478275,8	1319466	20837	246788	80369	89211
Ульяновская область	240556,1	571931	8891	157823	22469	27705

Вариант 2, 2013 г.

Субъект ПФО РФ	Y	X ₁	X ₂	X ₃	X ₄	X ₅
Республика Башкортостан	1163219	2105770	146854	961320	97642	126450
Республика Марий Эл	125950,2	314647	384	77450	11008	30232
Республика Мордовия	148705,7	433918	185	106093	10654	40289
Республика Татарстан	1551472,1	3342559	367108	1060483	121566	160157

1 2 3 4 5 6 7 8 9 10 11 12 13
 1 4. *Оценивание модели множественной линейной регрессии с использованием*
 2 *инструмент анализа данных «РЕГРЕССИЯ»*
 3

Окончание варианта 2

Субъект ПФО РФ	Y	X ₁	X ₂	X ₃	X ₄	X ₅
Удмуртская Республика	405126,4	870234	133616	190641	29524	47757
Чувашская Республика	223147,9	654078	633	128289	21476	30862
Пермский край	880264,4	2410614	209238	783008	96926	37652
Кировская область	224152,3	682835	848	135852	30721	27395
Нижегородская область	925182	2137855	1741	943843	90996	55094
Оренбургская область	717014,8	1596988	420880	192004	62496	82123
Пензенская область	270436,8	719888	649	128127	14480	47596
Самарская область	1048545,8	2342741	186506	780879	115666	67739
Саратовская область	526178,9	1456217	20374	242266	90217	99773
Ульяновская область	265288,7	616325	10212	164837	24930	28140

Вариант 3, 2014 г.

Субъект ПФО РФ	Y	X ₁	X ₂	X ₃	X ₄	X ₅
Республика Башкортостан	1260010,4	2306755	169093	893964	109823	136920
Республика Марий Эл	143396,1	330932	469	104874	10863	38514
Республика Мордовия	173872,7	477197	224	118328	11593	46280
Республика Татарстан	1661413,8	3431206	368991	1183555	126046	185974
Удмуртская Республика	450548,9	974795	137644	220811	33063	60293
Чувашская Республика	237447,2	664956	1909	133122	21955	37054
Пермский край	974192,9	2651647	206134	846723	105398	41669
Кировская область	254089,4	659265	898	153929	34879	34198
Нижегородская область	1009460,1	2381528	1776	998185	91099	67100
Оренбургская область	731277,7	1652428	348081	230679	65921	90370
Пензенская область	295238,7	864334	596	139216	19667	58218
Самарская область	1149147,8	2522834	205669	833357	122985	75793
Саратовская область	566646,1	1402337	20222	275766	89658	109571
Ульяновская область	278808,2	608107	11475	185638	25522	29369

Вариант 4, 2015 г.

Субъект ПФО РФ	Y	X ₁	X ₂	X ₃	X ₄	X ₅
Республика Башкортостан	1316598,3	2519215	195725	978473	113188	159470
Республика Марий Эл	171689,5	358656	501	139431	10959	48039

Окончание варианта 4

Субъект ПФО РФ	Y	X ₁	X ₂	X ₃	X ₄	X ₅
Республика Мордовия	180352,3	528791	231	132268	11874	53555
Республика Татарстан	1867258,7	3921931	439139	1369000	126650	217060
Удмуртская Республика	517999,8	1040617	165784	273381	34677	67282
Чувашская Республика	251307	715778	638	153829	20911	43955
Пермский край	1063780,3	2900859	249038	900639	105588	46537
Кировская область	282191	712522	925	189956	37173	36724
Нижегородская область	1104643,2	2579755	2303	1138576	97914	73587
Оренбургская область	774962,1	1820589	383955	274498	78012	106100
Пензенская область	343328,6	970828	411	176491	21534	76123
Самарская область	1264910,3	2735586	244656	928775	115748	87658
Саратовская область	625176,8	1551306	30646	342515	93886	131907
Ульяновская область	304479,1	675549	12851	235075	25529	35398

Вариант 5, 2016 г.

Субъект ПФО РФ	Y	X ₁	X ₂	X ₃	X ₄	X ₅
Республика Башкортостан	1344360,1	2868186	205472	965379	115343	158709
Республика Марий Эл	160464	404552	644	129024	11331	39217,4
Республика Мордовия	198132,8	599727	120	144706	11225	58710,9
Республика Татарстан	1937637,1	4256272	447744	1442974	123305	229693
Удмуртская Республика	540115	1169614	168379	332826	35291	64156,5
Чувашская Республика	261574,3	760719	518	160646	20814	39338,3
Пермский край	1091268,7	3204554	254573	858536	100802	40425,8
Кировская область	290990,3	785413	1207	194228	38150	38536,2
Нижегородская область	1182265	2790966	1240	1138876	92716	67365,9
Оренбургская область	772107,3	2039846	340354	269658	54816	110754
Пензенская область	338589	930071	822	180838	20440	83618,9
Самарская область	1275063,6	3012202	237648	907338	112420	93004
Саратовская область	655053,7	1779849	28187	344275	91650	142010
Ульяновская область	328249,3	727958	9541	235639	24583	38762,2

4.3. Тесты

1. В модель множественной линейной регрессии необходимо включить факторы, оказывающие на зависимую переменную:

- А) несущественное влияние;
- Б) существенное влияние;
- В) детерминированное влияние;
- Г) случайное влияние.

2. Модель множественной линейной регрессии представлена следующим уравнением:

- А) $y = b_0 + b_1x + b_2x^2 + \varepsilon$;
- Б) $y = b_0 + b_1x_1 + b_2x_2 + \varepsilon$;
- В) $y = b_0 + b_1x + \varepsilon$;
- Г) $y = a_0x^{a_1}\varepsilon$.

3. Модель множественной линейной регрессии в матричной форме имеет следующий вид:

- А) $\bar{X} = Y\bar{b} + \bar{\varepsilon}$;
- Б) $\bar{Y} = X\bar{b}$;
- В) $y = xb + \varepsilon$;
- Г) $\bar{Y} = X\bar{b} + \bar{\varepsilon}$.

4. Значимость модели множественной линейной регрессии в целом можно оценить с помощью:

- А) критерия Дарбина - Уотсона;
- Б) критерия Фишера - Снедекора;
- В) критерия Стьюдента;
- Г) теста Голдфелда - Квандта.

5. Получены следующие результаты оценки параметров модели множественной линейной регрессии:

Параметр	Кoeffициент	Стандартная ошибка	t-статистика
У-пересечение	12,4	8,0	1,55
X ₁	-9,6	3,2	-3,0
X ₂	\hat{b}_2	0,12	4,0
X ₃	-6,3	2,0	-3,15

Оценка \hat{b}_2 равна:

- А) 33,33;
- Б) 0,03;
- В) 0,48;
- Г) -9,6.

6. После оценки параметров модели множественной линейной цены мобильного телефона (Y , руб.) в зависимости от размера диагонали (X_1 , дюйм), от разрешения камеры (X_2 , мпикс) и объема встроенной памяти (X_3 , МБ) получены следующие результаты (для объема выборки, равного 31):

Параметр	Коэффициент	Стандартная ошибка	t-статистика	p-значения
Y-пересечение	-3108,35	2246,42	-1,38	0,1778
X_1	2931,63	1217,40	2,41	0,0231
X_2	648,04	232,35	2,79	0,0096
X_3	6,30	2,06	3,06	0,0049

Результатами проверки значимости оценок параметров модели на 5%-ном уровне значимости ($t_{крит} = 2,05$) будет вывод о том, что:

- А) все оценки параметров значимы;
- Б) все оценки параметров не значимы;
- В) все оценки параметров, кроме \hat{b}_0 , значимы;
- Г) только оценка \hat{b}_3 значима.

7. По данным теста б оценка \hat{b}_3 показывает, что:

- А) при увеличении объема встроенной памяти телефона на 1 МБ цена телефона увеличивается в среднем на 6,3 руб.;
- Б) при увеличении объема встроенной памяти телефона на 1 МБ цена телефона снизится в среднем на 6,3 руб.;
- В) при увеличении объема встроенной памяти телефона на 1 % цена телефона увеличивается в среднем на 6,3 руб.;
- Г) при увеличении объема встроенной памяти телефона на 1 МБ цена телефона увеличивается в среднем на 6,3 %.

8. По данным теста б найдите цену мобильного телефона, размер диагонали которого 4 дюйма, разрешение камеры - 8 мпикс., объем встроенной памяти - 512 МБ (округлите полученное значение до целых):

- А) 3586 руб.;
- Б) 17029 руб.;
- В) 20137 руб.;
- Г) 2932 руб.

	1	2	3	4	5	6	7	8	9	10	11	12	13
1	4. Оценивание модели множественной линейной регрессии с использованием												
2	инструмент анализа данных «РЕГРЕССИЯ»												
3													

9. По данным теста 6 получен частный коэффициент эластичности $\varepsilon_1 = 1,13$ при переменной X_1 . Выберите верное утверждение:

А) при увеличении размера диагонали телефона на 1 дюйм цена телефона увеличивается в среднем на 1,13%;

Б) при увеличении размера диагонали телефона на 1% цена телефона увеличивается в среднем на 1,13%;

В) при увеличении размера диагонали телефона на 1% цена телефона увеличивается в среднем на 1,13 руб.;

Г) при увеличении размера диагонали телефона на 1% цена телефона уменьшится в среднем на 1,13 %.

4.4. Контрольные вопросы

1. Перечислите предпосылки МНК для случая множественной линейной регрессии.
2. Каковы последствия невыполнимости предпосылок МНК для случая множественной линейной регрессии?
3. Сформулируйте теорему Гаусса - Маркова для модели множественной линейной регрессии.
4. Как проверяется статистическая значимость оценок параметров модели множественной линейной регрессии?
5. Как используется F -статистика при построении модели множественной линейной регрессии?
6. Как интерпретируются параметры множественной линейной регрессии?
7. В чем отличие скорректированного коэффициента детерминации от обычного?
8. Как рассчитать частный коэффициент эластичности?
9. Как рассчитать стандартизированный коэффициент регрессии?
10. Что означает незначимость оценок параметров модели множественной линейной регрессии?

5. ФИКТИВНАЯ ПЕРЕМЕННАЯ В РЕГРЕССИОННОЙ МОДЕЛИ



Изложены особенности эконометрического моделирования зависимости между показателями, один из которых выражен в номинальной шкале. Рассматриваются этапы оценивания эконометрической модели, содержащей фиктивную переменную.

5.1. Решение типовых задач

В качестве иллюстрационного примера особенностей оценивания регрессионной модели с фиктивной переменной воспользуемся результатами выборочного обследования арендного рынка жилья г. Казани (3-комнатные квартиры), проведенного в 2018 г. Объем выборки - 69 квартир (табл. 5.1).

Система показателей проведенного исследования приведена ниже:

Y - стоимость аренды квартиры, тыс. руб./мес.;

X - площади квартиры, м²;

D - наличие кабельного телевидения (1 - наличие; 0 - отсутствие).

ЗАМЕЧАНИЕ: Переменная «Наличие кабельного телевидения» выражена в качественной шкале. Для отражения ее влияния на стоимость аренды квартиры будет использована фиктивная (искусственная переменная) D (*dummy variable*) бинарного типа.

Таблица 5.1

Статистический массив проведенного исследования стоимости аренды квартиры
в г. Казани, 2018 г.

№ п/п	Y	X	D	№ п/п	Y	X	D	№ п/п	Y	X	D
1	13	63	1	24	20	63	1	47	25	73	1
2	14	68	1	25	20	76	1	48	25	84	0
3	14	54	0	26	20	73	1	49	25	80	1
4	15	71	0	27	20	58	1	50	25	70	1
5	15	50	0	28	20	60	1	51	25	90	1
6	16	50	0	29	20	80	1	52	25	74	1
7	20	78	0	30	20	70	1	53	25	57	1
8	17	55	1	31	20	67	0	54	25	78	1

5. Фиктивная переменная в регрессионной модели

Окончание табл. 5.1

№ п/п	Y	X	D	№ п/п	Y	X	D	№ п/п	Y	X	D
9	17	58	0	32	20	68	1	55	25	60	0
10	17	55	1	33	20	58	1	56	25	65	1
11	18	55	0	34	20	88	1	57	25	90	1
12	18	66	0	35	20	60	1	58	25	75	0
13	18	60	1	36	20	60	1	59	25	83	1
14	18	70	1	37	20	63	1	60	25	66	1
15	18	63	1	38	21	65	1	61	25	68	0
16	18	60	1	39	21	75	1	62	29	65	1
17	19	74	0	40	22	70	1	63	30	75	1
18	19	54	1	41	23	57	1	64	31	78	1
19	20	70	0	42	23	67	1	65	28	75	1
20	20	74	1	43	23	56	1	66	34	72	1
21	18	75	1	44	23	51	1	67	35	75	1
22	20	60	1	45	24	60	1	68	35	80	1
23	20	60	1	46	24	69	1	69	35	87	1

Шаг 1. Необходимо построить поле корреляции (точечную диаграмму). По оси OY изменяется стоимость аренды квартиры (Y), по оси OX - площадь квартиры (X). Для построения диаграммы необходимо выполнить: **Вставка / Точечная / Точечная с маркерами** (рис. 5.1).

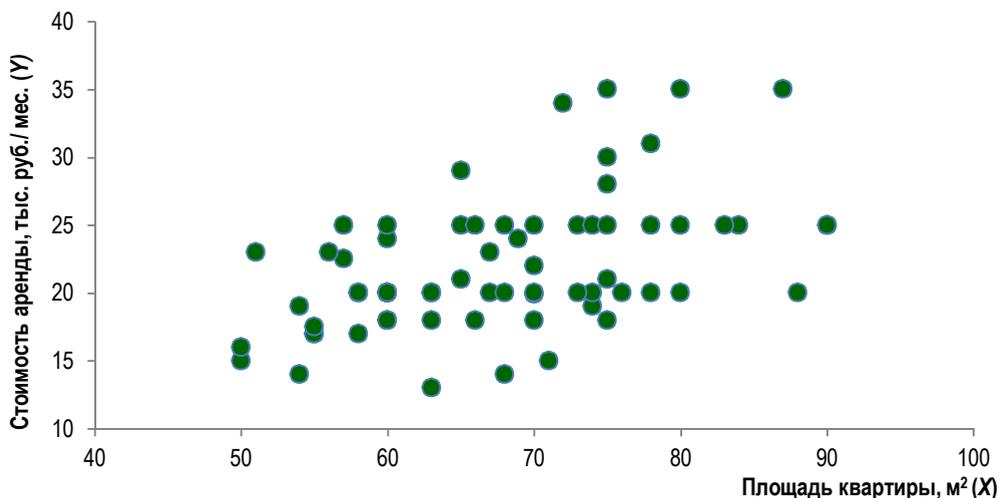


Рис. 5.1. Диаграмма зависимости стоимости аренды трехкомнатной квартиры и общей площади квартиры в г. Казани

По расположению точек на корреляционном поле можно предположить наличие прямой взаимосвязи между показателями. При этом выборочный парный коэффициент корреляции между рассматриваемыми переменными составляет 0,49 (рис. 5.2).

	A	B	C	D	E	F
1	№	Y	X	D		
63	62	29	65	1		
64	63	30	75	1		
65	64	31	78	1		
66	65	28	75	1		
67	66	34	72	1		
68	67	35	75	1		
69	68	35	80	1		
70	69	35	87	1		
71						
72	r=	0,49				

Рис. 5.2. Реализация функции КОРРЕЛ

Парный линейный выборочный коэффициент корреляции может быть вычислен в табличном редакторе Excel с помощью функции **КОРРЕЛ**. В аргументе функции необходимо выделить исходные массивы статистических показателей без заголовков столбцов. При этом очередность выделения массивов данных значения не имеет!

По шкале Чеддока связь между переменными можно охарактеризовать как умеренную.

Логично предположить, что на стоимость аренды жилья влияет и наличие/отсутствие кабельного телевидения. Следовательно, в процессе эконометрического моделирования разумно использовать две объясняющие переменные: площадь квартиры и наличие/отсутствие кабельного телевидения.

Теоретическое уравнение множественной линейной регрессии будет иметь вид

$$Y = b_0 + b_1X + c_1D + \varepsilon .$$

Шаг 2. Оценивание коэффициентов уравнения множественной линейной регрессии будем проводить с использованием инструмента анализа данных **«РЕГРЕССИЯ»**. Выбираем **Данные/Анализ данных/Регрессия** (рис. 5.3).

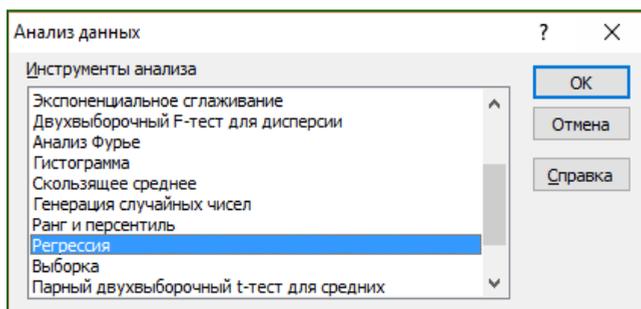


Рис. 5.3. Инструмент анализа данных «РЕГРЕССИЯ»

1 2 3 4 5 6 7 8 9 10 11 12 13

1
2
3

5. Фиктивная переменная в регрессионной модели

Шаг 3. В окне «РЕГРЕССИЯ» в диалоговом поле «Входной интервал Y» выбираем диапазон значений стоимости аренды жилья (Y), в диалоговом поле «Входной интервал X» - одновременно значения площади квартиры и наличие кабельного телевидения (X, D), как показано на рис. 5.4.

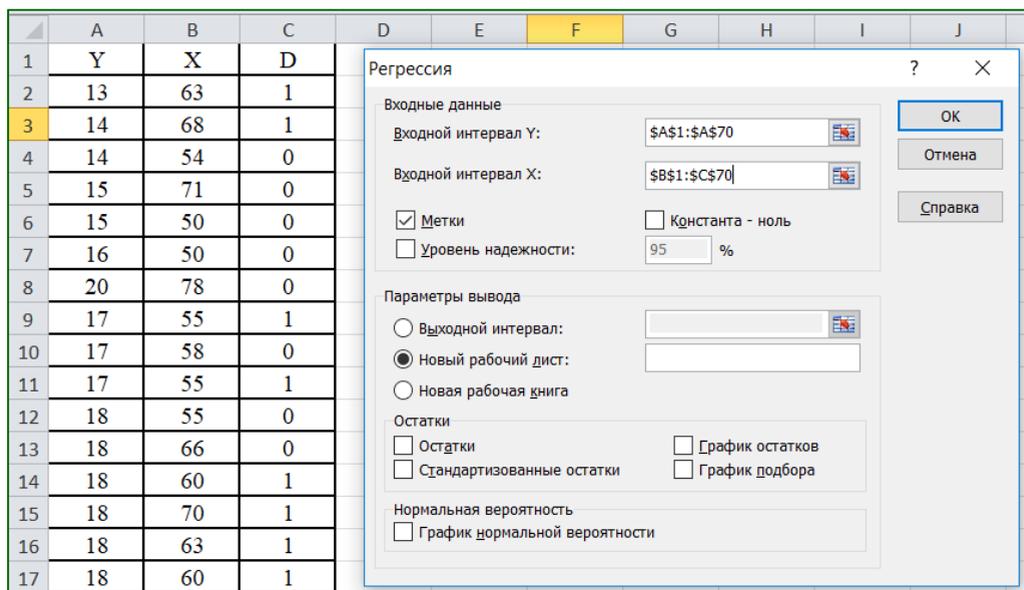


Рис. 5.4. Установки инструмента анализа данных «РЕГРЕССИЯ» при оценивании множественного уравнения регрессии

Шаг 4. Результаты работы инструмента анализа данных «РЕГРЕССИЯ» представлены на рис. 5.5.

Выборочное уравнение регрессии имеет вид

$$\hat{Y} = 4,48 + 0,23X + 2,55D.$$

Оценки коэффициентов регрессии \hat{b}_1 , \hat{c}_1 являются статистически значимыми в силу того, что **p-значение** оказалось меньше принятого уровня значимости α (0,05). Свободный член уравнения \hat{b}_0 статистически не значим.

Коэффициент \hat{b}_1 показывает, что при увеличении площади квартиры на 1 м² стоимость ее аренды возрастет на 0,23 тыс. руб. в месяц при условии неизменности факта наличия кабельного телевидения.

Коэффициент \hat{c}_1 показывает, что стоимость аренды квартиры с наличием кабельного телевидения дороже на 2,55 тыс. руб. в месяц по сравнению с его отсутствием (квартиры одинаковой площади).

Эконометрика в табличном редакторе MS Excel

	A	B	C	D	E	F	G	H	I
1	Вывод итогов								
2									
3	<i>Регрессионная статистика</i>								
4	Множественный R	0,533							
5	R-квадрат	0,285							
6	Нормированный R-квадрат	0,263							
7	Стандартная ошибка	4,239							
8	Наблюдения	69							
9									
10	<i>Дисперсионный анализ</i>								
11		<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Значимость F</i>			
12	Регрессия	2	471,74	235,87	13,12	1,59E-05			
13	Остаток	66	1186,15	17,97					
14	Итого	68	1657,89						
15									
16		<i>Коэффициенты</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>	<i>p-Значение</i>	<i>Нижние 95%</i>	<i>Верхние 95%</i>	<i>Нижние 95,0%</i>	<i>Верхние 95,0%</i>
17	Y-пересечение	4,48	3,58	1,25	0,21	-2,66	11,62	-2,66	11,62
18	X	0,23	0,05	4,39	0,00	0,12	0,33	0,12	0,33
19	D	2,55	1,25	2,05	0,04	0,06	5,04	0,06	5,04

Рис. 5.5. Результаты оценки множественного уравнения регрессии с фиктивной переменной (приведен фрагмент исходного окна)

Коэффициент детерминации R^2 равен 0,285. Модель в целом может быть признана статистически значимой, так как показатель **«значимость F»** меньше принятого уровня значимости (0,05). Следует отметить, что объясняющая способность модели невелика. Вариация стоимости аренды жилья в г. Казани лишь на 28,5 % зависит от вариации площади квартиры и от наличия кабельного телевидения. В дальнейшем, безусловно, необходимо попытаться подобрать дополнительные объясняющие переменные для увеличения объясняющей способности модели.

ЗАМЕЧАНИЕ: Оцененная регрессионная модель относится к классу ANCOVA (модели, содержащие как количественные, так и качественные переменные бинарного типа).

Шаг 5. Вычисление расчетных значений стоимости аренды трехкомнатной квартиры будем производить по уравнению $\hat{Y} = 4,48 + 0,23X + 2,55D$. Для этого следует в дополнительном столбце (рис. 5.6) провести необходимые вычисления с приданием ячейкам с оценками коэффициентов регрессии универсального адреса (F4).

5. Фиктивная переменная в регрессионной модели

ЛИНЕЙН					= \$G\$17+\$G\$18*B2+\$G\$19*C2		
A	B	C	D	E	F	G	H
1	Y	X	D	Y [^]	ВЫВОД ИТОГОВ		
2	13	63	1	= \$G\$17+\$G\$18*B2+\$G\$19*C2			
					Регрессионная статистика		
3	14	68	1	22,5985	Множественный R	0,530049926	
4	14	54	0	16,892	R-квадрат	0,280952924	
5	15	71	0	20,754	Нормированный R-квадрат	0,259163619	
6	15	50	0	15,9833	Стандартная ошибка	4,244968257	
7	16	50	0	15,9833	Наблюдения	69	
8	20	78	0	22,3442			
9	17	55	1	19,6452			
10	17	58	0	17,8007	Дисперсионный анализ		
11	17	55	1	19,6452		df	SS
12	18	55	0	17,1192	Регрессия	2	4
13	18	66	0	19,6181	Остаток	66	11
14	18	60	1	20,7811	Итого	68	
15	18	70	1	23,0529			
16	18	63	1	21,4627		Коэффициенты	Стандартна
17	18	60	1	20,7811	Y-пересечение	4,624536181	3,
18	19	74	0	21,4355	X	0,227175467	0,0
19	19	54	1	19,4181	D	2,526061966	1,

Рис. 5.6. Вычисление расчетных значений зависимой переменной

Для построения корреляционного поля с расчетными значениями результативного признака необходимо расположить столбцы в порядке X, Y, \hat{Y} , выделить данные столбцы мышью и построить точечную диаграмму (рис. 5.7). Фиктивная переменная в уравнение регрессии включена методом сдвига, так как в присутствии заданной градации ($D=1$) соответствующий коэффициент (\hat{c}_1) приводит к поднятию эмпирической линии регрессии на соответствующее значение вдоль оси OY .

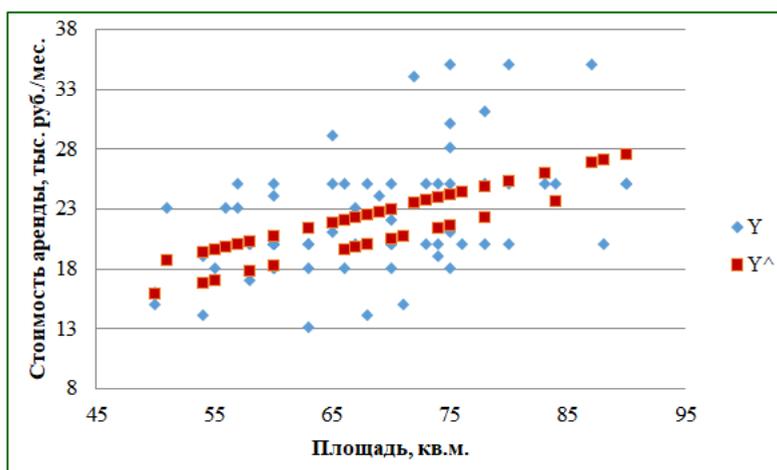


Рис. 5.7. Корреляционное поле с эмпирической линией регрессии при наличии фиктивной (бинарной) переменной в уравнении регрессии

5.2. Задача для самостоятельного выполнения

Задача 15. По результатам проведенного выборочного обследования стоимости аренды однокомнатных квартир в г. Казани (2018 г.) сформирован статистический массив (табл. 5.2), в основе которого лежит следующая система показателей:

- Y - стоимость аренды квартиры, тыс. руб./мес.;
- X - площадь квартиры, м²;
- D_1 - оплата коммунальных услуг арендатором (да/нет);
- D_2 - статус арендодателя (агентство / собственник);
- D_3 - наличие кабельного телевидения (да/нет);
- D_4 - оснащенность квартиры холодильником (да/нет);
- D_5 - оснащенность квартиры кондиционером (да/нет).

Таблица 5.2

Характеристики 1-комнатных квартир

№ п/п	Y	X	D_1	D_2	D_3	D_4	D_5
1	15	37	Да	Агентство	Да	Да	Да
2	14	40	Да	Агентство	Да	Да	Нет
3	12	40	Нет	Агентство	Да	Да	Нет
4	15	41	Да	Собственник	Да	Да	Нет
5	20	40	Да	Собственник	Да	Да	Нет
6	16	34	Да	Агентство	Да	Да	Да
7	12	40	Да	Агентство	Да	Да	Нет
8	18	39	Да	Собственник	Да	Да	Нет
9	14	42	Да	Агентство	Да	Да	Нет
10	14	38	Да	Собственник	Да	Да	Нет
11	17	45	Да	Собственник	Да	Да	Да
12	15	34	Нет	Агентство	Да	Да	Нет
13	12	33	Да	Собственник	Да	Да	Нет
14	16	34	Нет	Агентство	Да	Да	Нет
15	15	43	Да	Агентство	Да	Да	Нет
16	17	42	Да	Собственник	Да	Да	Да
17	13	37	Да	Агентство	Да	Да	Нет
18	13	35	Да	Собственник	Да	Да	Нет
19	17	40	Нет	Агентство	Нет	Да	Да

	1	2	3	4	5	6	7	8	9	10	11	12	13
1													
2													
3													

5. Фиктивная переменная в регрессионной модели

Окончание табл. 5.2

№ п/п	Y	X	D ₁	D ₂	D ₃	D ₄	D ₅
20	17	36	Да	Собственник	Нет	Да	Нет
21	15	38	Да	Собственник	Да	Да	Нет
22	18	40	Нет	Собственник	Да	Да	Нет
23	20	48	Да	Собственник	Да	Да	Нет
24	13	35	Да	Собственник	Да	Да	Нет
25	18	43	Нет	Агентство	Да	Да	Нет
26	15	46	Да	Собственник	Да	Да	Нет
27	11	40	Да	Собственник	Да	Да	Нет
28	15	38	Да	Собственник	Да	Да	Нет
29	15	44	Да	Собственник	Да	Да	Нет
30	10	36	Да	Собственник	Да	Да	Нет
31	18	40	Да	Собственник	Да	Да	Нет
32	15	33	Нет	Агентство	Да	Да	Нет
33	12	41	Да	Собственник	Нет	Да	Нет
34	19	37	Нет	Собственник	Да	Да	Нет
35	12	42	Да	Агентство	Да	Да	Нет
36	15	38	Нет	Агентство	Да	Да	Нет
37	18	41	Да	Агентство	Да	Да	Да
38	14	42	Да	Агентство	Да	Да	Нет
39	12	45	Да	Агентство	Да	Да	Да
40	12	44	Да	Агентство	Да	Да	Да
41	10	37	Да	Агентство	Нет	Да	Нет
42	11	32	Да	Агентство	Нет	Да	Нет
43	12	43	Да	Агентство	Да	Да	Нет
44	18	45	Нет	Агентство	Да	Да	Да
45	18	46	Да	Собственник	Да	Да	Нет
46	17	45	Да	Агентство	Да	Да	Нет
47	18	34	Нет	Агентство	Да	Да	Да
48	20	40	Нет	Агентство	Да	Да	Нет
49	14	44	Да	Собственник	Да	Да	Нет
50	15	48	Да	Собственник	Да	Да	Нет

	1	2	3	4	5	6	7	8	9	10	11	12	13
1			<i>Эконометрика в табличном редакторе MS Excel</i>										
2													
3													

Используя вышеприведенные данные и выбрав соответствующий вариант, следует:

- 1) оценить уравнение множественной линейной регрессии с фиктивной переменной, предварительно закодировав ее;
- 2) проверить статистическую значимость оценок коэффициентов регрессии и модели в целом ($\alpha = 0,08$);
- 3) сделать необходимые экономические выводы;
- 4) построить корреляционное поле с эмпирической линией регрессии;
- 5) если коэффициент при фиктивной переменной оказался статистически значимым, необходимо провести моделирование при максимальном и минимальном значениях переменной X .

Вариант 1: переменные Y, X, D_1 .

Вариант 2: переменные Y, X, D_2 .

Вариант 3: переменные Y, X, D_3 .

Вариант 4: переменные Y, X, D_4 .

Вариант 5: переменные Y, X, D_5 .

5.3. Тесты

1. В регрессионных моделях влияние качественного фактора выражается в виде:
 - А) фиктивной переменной;
 - Б) эндогенной переменной;
 - В) лаговой переменной.
2. Фиктивные переменные вводятся:
 - А) в линейные модели;
 - Б) в нелинейные модели;
 - В) как в линейные, так и в нелинейные модели, приводимые к линейному виду.
3. Фиктивные переменные в модели могут выступать в роли:
 - А) фактора;
 - Б) результата;
 - В) как фактора, так и результата.
4. Найти оценки параметра уравнения линейной регрессии с фиктивными переменными можно с помощью метода:
 - А) максимального правдоподобия;
 - Б) инструментальных переменных;
 - В) наименьших квадратов;
 - Г) моментов.

	1	2	3	4	5	6	7	8	9	10	11	12	13
1													
2													
3													

5. Фиктивная переменная в регрессионной модели

5. Оценка значимости параметров уравнения регрессии с фиктивными переменными осуществляется на основе:

- А) F -критерия (критерия Фишера - Снедекора);
- Б) t -критерия (критерия Стьюдента);
- В) стандартной ошибки оценки коэффициента регрессии;
- Г) средней ошибки аппроксимации.

6. Если качественный признак, который необходимо отразить в регрессионной модели, имеет две градации, то в уравнение включают:

- А) одну фиктивную переменную;
- Б) две фиктивные переменные;
- В) три фиктивные переменные;
- Г) нельзя включать качественные переменные.

7. Чтобы избежать «ловушки», число вводимых фиктивных переменных должно быть:

- А) на единицу меньше числа уровней качественного признака;
- Б) на единицу больше числа уровней качественного признака;
- В) больше числа уровней качественного признака;
- Г) нет правильного ответа.

8. Если качественная переменная имеет k альтернативных значений, то при моделировании используются:

- А) $(k - 1)$ фиктивная переменная;
- Б) k фиктивных переменных;
- В) $(k + 1)$ фиктивная переменная.

9. Фиктивные переменные могут принимать значения:

- А) только 1;
- Б) только 0;
- В) 1, 2, 3, 4, 5, ...;
- Г) 1 и 0;
- Д) нет правильного ответа.

5.4. Контрольные вопросы

1. Что такое фиктивная переменная?
2. Что такое «базовая категория»?
3. Каковы основные причины применения фиктивных переменных в эконометрических моделях?

	1	2	3	4	5	6	7	8	9	10	11	12	13
1			<i>Эконометрика в табличном редакторе MS Excel</i>										
2													
3													

4. Закончите фразу: «Если качественный признак имеет p уровней, то число фиктивных переменных в модели следует взять равным...»
5. Назовите виды регрессионных уравнений, в которых фиктивной является зависимая переменная.
6. Как проверяется значимость коэффициента при фиктивной переменной?
7. Что означает статистическая значимость коэффициента при фиктивной переменной?
8. В чем особенность проведения теста Чоу?
9. В чем преимущество использования фиктивных переменных?
10. Что такое «ловушка фиктивных переменных»?

6. РЕГРЕССИОННАЯ МОДЕЛЬ С ДВУМЯ ФИКТИВНЫМИ ПЕРЕМЕННЫМИ



Рассматриваются этапы оценивания эконометрической модели, содержащей более одной фиктивной переменной.

6.1. Решение типовых задач

Для иллюстрации оценивания регрессионной модели с одной количественной и двумя качественными (бинарными) переменными воспользуемся статистическими данными выборочного обследования факторов, определяющих цену автомобиля на вторичном рынке г. Оренбурга (2017 г.). Статистический массив представлен в табл. 6.1 (объем выборки 36 автомобилей).

Таблица 6.1

Данные выборочного обследования факторов, определяющих цену автомобиля на вторичном рынке (г. Оренбург, 2017 г.)

Марка авто	Y	X	D ₁	D ₂	Марка авто	Y	X	D ₁	D ₂
Mazda 6	315	170	0	1	Toyota Camry	500	150	1	1
Renault Megan	355	110	0	1	Chevrolet Cruze	505	57	1	0
Geely Emgrand Ec7	335	145	0	1	Mazda 3	520	75	1	0
Peugeot 308	342	120	0	1	Skoda Octavia	535	100	1	0
Citroen C3 Picasso	380	83	0	1	Kia Sorento	555	126	1	0
Ford Focus	385	95	0	1	Opel Astra	570	65	1	0
Chevrolet Cobalt	415	93	0	1	Opel Astra	570	65	1	0
Honda Civic	415	88	1	0	Skoda Octavia	580	62	1	0
Chevrolet Aveo	440	68	1	1	Audi A4	620	59	1	0
Mitsubishi Lancer	430	145	0	1	Audi A4	625	130	1	0
Renault Fluence	430	67	0	1	Opel Astra	625	64	1	0
Toyota Corolla	435	90	0	1	Skoda Octavia	635	79	1	1
Volkswagen Passat	450	170	0	1	Kia cee'd	650	50	1	0
Peugeot 301	459	87	0	1	Volkswagen Golf	650	81	1	0
Ford Focus	463	63	0	1	Volkswagen Jetta	680	75	1	0
Chevrolet Aveo	465	110	0	1	Toyota Corolla	685	89	1	0

Окончание табл. 6.1

Марка авто	Y	X	D ₁	D ₂	Марка авто	Y	X	D ₁	D ₂
Nissan Note	470	114	0	1	Toyota Auris	730	52	1	0
Opel Corsa	480	125	0	1	Opel Astra	750	35	1	0

Примечание: Y - цена автомобиля, тыс. руб.; X - пробег автомобиля, тыс. км; D₁ - участие автомобиля в дорожно-транспортных происшествиях (1 - автомобиль не участвовал в ДТП, 0 - в противном случае); D₂ - тип коробки передач (1 - коробка механическая, 0 - в противном случае).

Шаг 1. На начальном этапе исследования установим наличие корреляционной взаимосвязи между ценой автомобиля и пробегом. Воспользуемся функцией **КОРРЕЛ** (рис. 6.1).

	B	C	D	E	F	G
1	Y	X	D1	D2		
2	315	170	0	1		г выб.
3	355	110	0	1		-0,5547
4	335	145	0	1		
5	342	120	0	1		
6	380	83	0	1		
7	385	95	0	1		

Рис. 6.1. Вычисление парного линейного коэффициента корреляции

Между изучаемыми показателями наблюдается заметная корреляционная взаимосвязь ($r_{xy} = -0,55$). С ростом пробега автомобиля его стоимость на вторичном рынке уменьшается, что согласуется с положениями экономической теории.

Шаг 2. Оценим теоретическое уравнение парной регрессии вида $Y = b_0 + b_1X + \varepsilon$ с применением инструмента анализа данных **«РЕГРЕССИЯ»** (рис. 6.2).

Выборочное уравнение регрессии имеет вид $\hat{Y} = 687,83 - 1,88X$. Оценки коэффициентов регрессии \hat{b}_0, \hat{b}_1 являются статистически значимыми, так как соответствующее им **p-значение** меньше заданного уровня значимости $\alpha = 0,08$.

Коэффициент \hat{b}_1 показывает, что при увеличении пробега автомобиля на 1 тыс. км стоимость его на вторичном рынке падает на 1,88 тыс. руб.

Коэффициент \hat{b}_0 не имеет экономического содержания.

Коэффициент детерминации $R^2 = 0,31$. При этом эконометрическая модель может быть признана статистически значимой с использованием F-статистики. Показатель **«значимость F»** меньше заданного уровня значимости $\alpha = 0,08$.

Таким образом, стоимость автомобиля на вторичном рынке на 31% зависит от его пробега, а на 69% от прочих факторов, не включенных в процесс исследования.

1 2 3 4 5 6 7 8 9 10 11 12 13

1 **6. Регрессионная модель с двумя фиктивными переменными**

2

3

	A	B	C	D	E	F	G
1	Вывод ИТОВ						
2							
3	Регрессионная статистика						
4	Множественный R	0,554701426					
5	R-квадрат	0,307693672					
6	Нормированный R-квадрат	0,287331721					
7	Стандартная ошибка	99,00514421					
8	Наблюдения	36					
9							
10	Дисперсионный анализ						
11		<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Значимость F</i>	
12	Регрессия	1	148120,3405	148120,3405	15,1112079	0,000446731	
13	Остаток	34	333268,6317	9802,01858			
14	Итого	35	481388,9722				
15							
16		<i>Коэффициенты</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>	<i>P-Значение</i>	<i>Нижние 95%</i>	<i>Верхние 95%</i>
17	Y-пересечение	687,8271032	48,03277712	14,31995284	5,8188E-16	590,2127557	785,4414508
18	X	-1,880481298	0,483748274	-3,887313711	0,00044673	-2,863576071	-0,89738652
19							
20							
21							
22	Вывод ОСТАТКА						
23							
24		<i>Наблюдение</i>	<i>Предсказанное Y</i>	<i>Остатки</i>			
25		1	368,1452826	-53,14528262			
26		2	480,9741605	-125,9741605			
27		3	445,4573454	88,4573454			

Рис. 6.2. Итоги работы инструмента анализа данных «РЕГРЕССИЯ» (парная регрессионная модель с количественной переменной)

Для наглядной иллюстрации рассматриваемой выше корреляционной зависимости построим диаграмму рассеяния (корреляционное поле). В главном меню выберем **Вставка / Точечная / Точечная с маркером**.

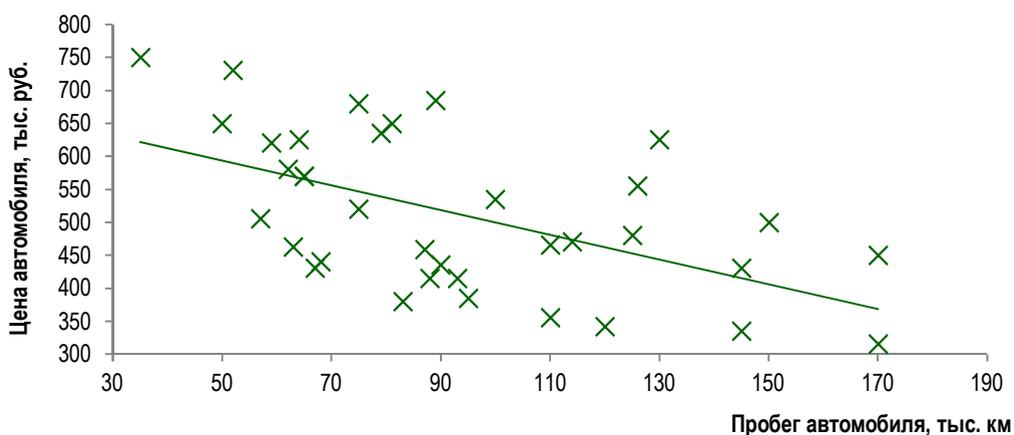


Рис. 6.3. Корреляционное поле «Пробег / Стоимость автомобиля»

На рис. 6.3 наглядно видно, как снижается цена при росте пробега автомобиля. Это логично, поскольку износ требует от будущих владельцев дополнительных затрат на восстановление определенных узлов автомобиля и на соответствующий ремонт.

Шаг 3. Здравый смысл позволяет предположить, что цена автомобиля может зависеть не только от его пробега, но и от факта участия в ДТП, а также от типа коробки передач. Данные признаки выражены в качественной шкале. Для обеспечения возможности включения данных переменных в эконометрическую модель следует воспользоваться бинарным кодом вида 0/1.

Таким образом, необходимо оценить уравнение множественной линейной регрессии (модель ANCOVA) вида $Y = b_0 + b_1X + c_1D_1 + c_2D_2 + \varepsilon$ с применением инструмента анализа данных «РЕГРЕССИЯ». Итоги вычислений данного инструмента представлены на рис. 6.4.

	A	B	C	D	E	F	G
1	Вывод итогов						
2							
3	<i>Регрессионная статистика</i>						
4	Множественный R	0,808977251					
5	R-квадрат	0,654444193					
6	Нормированный R-квадрат	0,622048336					
7	Стандартная ошибка	72,09948739					
8	Наблюдения	36					
9							
10	<i>Дисперсионный анализ</i>						
11		<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Значимость F</i>	
12	Регрессия	3	315042,2176	105014,0725	20,2014781	1,56788E-07	
13	Остаток	32	166346,7546	5198,336082			
14	Итого	35	481388,9722				
15							
16		<i>Коэффициенты</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>	<i>P-Значение</i>	<i>Нижние 95%</i>	<i>Верхние 95%</i>
17	Y-пересечение	552,3376356	60,54436907	9,122857238	2,03888E-10	429,0127915	675,6624797
18	X	-0,6903244	0,412076451	-1,675233801	0,103634975	-1,529696664	0,149047864
19	D1	103,1790092	45,6618986	2,25963029	0,030787246	10,1687654	196,189253
20	D2	-62,17452918	46,23130177	-1,344857852	0,18812613	-156,3446093	31,9955509
21							
22							
23							
24	Вывод остатка						
25							
26	<i>Наблюдение</i>	<i>Предсказанное Y</i>	<i>Остатки</i>				
27	1	372,8079584	-57,80795843				
28	2	414,2274224	-59,22742242				

Рис. 6.4. Итоги работы инструмента анализа данных «РЕГРЕССИЯ»
 $(Y = b_0 + b_1X + c_1D_1 + c_2D_2 + \varepsilon)$

По итогам вычислений (см. рис. 6.4) можно сделать вывод о незначимости оценок коэффициентов регрессии \hat{b}_1, \hat{c}_2 (соответствующее **p-значение** превышает заданный уровень значимости $\alpha = 0,08$). Следовательно, одновременное включение двух фиктивных переменных в модель ухудшило результаты моделирования цены автомобиля. Данная модель не может быть использована на практике.

6. Регрессионная модель с двумя фиктивными переменными

Шаг 4. Совершим попытку включения в процесс эконометрического моделирования фиктивных переменных поочередно.

1. Оценим модель множественной линейной регрессии вида $Y = b_0 + b_1X + c_1D_1 + \varepsilon$. Результаты оценивания представлены на рис. 6.5.

	A	B	C	D	E	F	G	H	I
1	Вывод итогов								
2									
3	Регрессионная статистика								
4	Множественный R	0,796814502							
5	R-квадрат	0,634913351							
6	Нормированный R-квадрат	0,612786887							
7	Стандартная ошибка	72,97751562							
8	Наблюдения	36							
9									
10	Дисперсионный анализ								
11		<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Значимость F</i>			
12	Регрессия	2	305640,2853	152820,1426	28,69475042	6,01912E-08			
13	Остаток	33	175748,6869	5325,717786					
14	Итого	35	481388,9722						
15									
16		<i>Коэффициенты</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>	<i>P-Значение</i>	<i>Нижние 95%</i>	<i>Верхние 95%</i>	<i>Нижние 95,0%</i>	<i>Верхние 95,0%</i>
17	Y-пересечение	503,4687339	49,01700981	10,27130655	8,24634E-12	403,7428776	603,1945902	403,7428776	603,1945902
18	X	-0,809457046	0,407343376	-1,98716143	0,055257188	-1,638203377	0,019289284	-1,638203377	0,019289284
19	D1	152,0736443	27,96245206	5,438494591	5,05592E-06	95,1836078	208,9636807	95,1836078	208,9636807

Рис. 6.5. Итоги работы инструмента анализа данных «РЕГРЕССИЯ»
 $(Y = b_0 + b_1X + c_1D_1 + \varepsilon)$

По итогам вычислений (см. рис. 6.5) можно сделать вывод о значимости оценок коэффициентов регрессии \hat{b}_0 , \hat{b}_1 , \hat{c}_1 (соответствующее **p-значение** не превышает заданного уровня значимости $\alpha = 0,08$). Коэффициент детерминации $R^2 = 0,63$. Выборочное уравнение регрессии вида $\hat{y} = 503,47 - 0,81X + 152,07D_1$ может быть признано статистически значимым (показатель **«значимость F»** много меньше принятого уровня значимости $\alpha = 0,08$).

Таким образом, цена автомобиля на вторичном рынке на 63% зависит от пробега и факта участия в ДТП, а на 37% - от прочих факторов. Экономическое содержание оценки \hat{c}_1 : автомобиль, не участвовавший в ДТП, стоит на 152,07 тыс. руб. дороже, чем автомобиль, имевший участие в происшествии, при условии равенства пробега.

2. Оценим модель множественной линейной регрессии вида $Y = b_0 + b_1X + c_2D_2 + \varepsilon$. Результаты оценивания представлены на рис. 6.6.

После применения инструмента анализа данных **«РЕГРЕССИЯ»** (см. рис. 6.6) можно сделать вывод о значимости оценок коэффициентов регрессии \hat{b}_0 , \hat{b}_1 , \hat{c}_1 (соответствующее **p-значение** не превышает заданного уровня значимости $\alpha = 0,08$). Коэффициент детерминации $R^2 = 0,60$. Выборочное уравнение регрессии вида $\hat{y} = 663,5 - 0,79X -$

$-145,35D_2$ может быть признано статистически значимым (показатель **«значимость F»** много меньше принятого уровня значимости $\alpha = 0,08$).

L11									
A	B	C	D	E	F	G	H	I	
Вывод итогов									
<i>Регрессионная статистика</i>									
4	Множественный R	0,774149347							
5	R-квадрат	0,599307211							
6	Нормированный R-квадрат	0,5750228							
7	Стандартная ошибка	76,45341022							
8	Наблюдения	36							
<i>Дисперсионный анализ</i>									
11		<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Значимость F</i>			
12	Регрессия	2	288499,8824	144249,9412	24,67867967	2,79505E-07			
13	Остаток	33	192889,0898	5845,123934					
14	Итого	35	481388,9722						
<i>Кoeffициенты Стандартная ошибка t-статистика P-Значение Нижние 95% Верхние 95% Нижние 95,0% Верхние 95,0%</i>									
17	Y-пересечение	663,5001673	37,42239906	17,73002758	1,96117E-18	587,3637239	739,6366106	587,3637239	739,6366106
18	X	-0,796938605	0,434087096	-1,835895636	0,075397641	-1,680095441	0,086218232	-1,680095441	0,086218232
19	D2	-145,3517435	29,65959367	-4,900665366	2,47117E-05	-205,6946405	-85,00884644	-205,6946405	-85,00884644

Рис. 6.6. Итоги работы инструмента анализа данных «РЕГРЕССИЯ»
 $(Y = b_0 + b_1X + c_2D_2 + \varepsilon)$

Таким образом, цена автомобиля на вторичном рынке на 60% зависит от пробега и типа коробки передач, а на 40% - от прочих факторов. Экономическое содержание оценки \hat{c}_2 : автомобиль с механической коробкой передач стоит на 145,35 тыс. руб. дешевле автомобиля с автоматической коробкой передач (вариатором) при условии равенства пробега.

Шаг 5. В результате эконометрического моделирования было получено три выборочных уравнения регрессии с разным набором объясняющих переменных. Для выбора оптимальной эконометрической модели следует сравнить полученные уравнения по величине скорректированного (нормированного) коэффициента детерминации:

$$\bar{R}^2 = 1 - \frac{ESS / (n - m - 1)}{TSS / (n - 1)}$$

Показатели дисперсионного анализа рассчитываются инструментом анализа данных **«РЕГРЕССИЯ»** в соответствующей таблице (например, см. рис. 6.6). Они расположены в столбце **«SS»**, англ. - *Sum of Squares* (рис. 6.7).

	1	2	3	4	5	6	7	8	9	10	11	12	13
1		6. Регрессионная модель с двумя фиктивными переменными											
2													
3		Регрессионная статистика											
4		Множественный R		0,774149347									
5		R-квадрат		0,599307211									
6		Нормированный R-квадрат		0,5750228									
7		Стандартная ошибка		76,45341022									
8		Наблюдения		36									
9													
10		Дисперсионный анализ											
11			<i>df</i>		<i>SS</i>	<i>RSS</i>	<i>MS</i>		<i>F</i>		<i>Значимость F</i>		
12		Регрессия	2		288499,8824	144249,9412		24,6786797		2,79505E-07			
13		Остаток	33	<i>ESS</i>	192889,0898								
14		Итого	35		481388,9722	<i>TSS</i>	23934						
15													
16				<i>Коэффициенты</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>	<i>P-Значение</i>	<i>Нижние 95%</i>					
17		Y-пересечение		663,5001673		37,42239906	17,73002758	1,9612E-18	587,3637239				
18		X		-0,796938605		0,434087096	-1,835895636	0,07539764	-1,680095441				

Рис. 6.7. Показатели дисперсионного анализа в итогах работы инструмента анализа данных «РЕГРЕССИЯ»

Нормированный коэффициент детерминации также уже рассчитан в первой таблице **«Регрессионная статистика»** инструмента **«РЕГРЕССИЯ»** на рис. 6.7.

Так как рассматриваемые модели имеют одинаковую линейную форму, их можно сравнить по величине \bar{R}^2 и выбрать оптимальную по максимизации данного коэффициента (табл. 6.2).

Таблица 6.2

Сравнение линейных эконометрических моделей

Выборочное уравнение регрессии	\bar{R}^2	Статистическая значимость оценок коэффициентов	Статистическая значимость модели в целом
$\hat{Y} = 687,83 - 1,88X$	0,29	++	+
$\hat{y} = 503,47 - 0,81X + 152,07D_1$	0,61	++	+
$\hat{y} = 663,5 - 0,79X - 145,35D_2$	0,58	++	+

Основываясь на данных табл. 6.2, можно сделать вывод, что наилучшей является модель с максимальным нормированным коэффициентом детерминации: $\hat{y} = 503,47 - 0,81X + 152,07D_1$.

6.2. Задача для самостоятельного выполнения

Задача 16. Используя предложенные в каждом варианте статистические данные выборочных обследований, постройте качественную эконометрическую модель с наличием одной или двух фиктивных переменных. Процесс моделирования следует фиксировать пошагово в аналитической записке в соответствии с алгоритмом, описанным в типовой задаче п. 6.1.

Вариант 1

Y	X	D ₁	D ₂	Y	X	D ₁	D ₂
1200	47	Балкон	Затемненная	1250	40	Лоджия	Солнечная
1150	49	Лоджия	Затемненная	3600	70	Лоджия	Солнечная
1250	43	Лоджия	Солнечная	1900	44	Лоджия	Затемненная
1370	36	Лоджия	Солнечная	1450	49	Лоджия	Солнечная
1420	38	Балкон	Затемненная	950	35	Лоджия	Затемненная
1360	35	Лоджия	Солнечная	1220	48	Лоджия	Солнечная
1550	33	Лоджия	Солнечная	1850	50	Лоджия	Солнечная
1250	30	Балкон	Солнечная	1630	60	Лоджия	Затемненная
1850	39	Балкон	Затемненная	1080	41	Лоджия	Затемненная
1100	30	Лоджия	Солнечная	1180	48	Лоджия	Затемненная
1050	31	Лоджия	Солнечная	1430	41	Лоджия	Затемненная
1380	33	Лоджия	Солнечная	1300	39	Лоджия	Затемненная
1650	37	Лоджия	Затемненная	1450	45	Лоджия	Солнечная
1430	38	Лоджия	Солнечная	1650	33	Лоджия	Солнечная
1100	37	Лоджия	Затемненная	1650	40	Лоджия	Солнечная
1170	36	Лоджия	Солнечная	1550	33	Балкон	Затемненная
1150	36	Лоджия	Затемненная	1850	42	Лоджия	Затемненная
1100	30	Лоджия	Солнечная	1650	40	Лоджия	Солнечная
1600	34	Лоджия	Солнечная	1550	33	Лоджия	Солнечная
950	29	Балкон	Затемненная	1100	37	Балкон	Затемненная

Здесь: Y - стоимость **однокомнатной** квартиры в городе N, тыс. руб.; X - общая площадь квартиры, м²;
D₁ - наличие балкона или лоджии; D₂ - сторона выхода окон квартиры.

	1	2	3	4	5	6	7	8	9	10	11	12	13
1	6. Регрессионная модель с двумя фиктивными переменными												
2													
3													

Вариант 2

Y	X	D ₁	D ₂	Y	X	D ₁	D ₂
2100	85	Балкон	Затемненная	1730	46	Балкон	Солнечная
1500	71	Балкон	Солнечная	1850	46	Балкон	Затемненная
1700	73	Балкон	Затемненная	1550	74	Лоджия	Солнечная
1850	80	Балкон	Затемненная	1830	77	Балкон	Затемненная
1100	57	Лоджия	Солнечная	1980	84	Балкон	Затемненная
1180	68	Балкон	Солнечная	1550	76	Балкон	Солнечная
2300	85	Балкон	Затемненная	1750	78	Балкон	Солнечная
1850	50	Лоджия	Солнечная	2250	81	Балкон	Затемненная
1150	46	Балкон	Затемненная	1820	80	Балкон	Солнечная
1700	54	Балкон	Солнечная	1570	65	Балкон	Затемненная
1850	52	Балкон	Солнечная	1841	86	Балкон	Солнечная
1750	53	Лоджия	Солнечная	2100	84	Балкон	Солнечная
1400	52	Лоджия	Затемненная	1600	79	Лоджия	Затемненная
1550	55	Балкон	Солнечная	1863	71	Балкон	Солнечная
1250	42	Лоджия	Солнечная	1900	79	Балкон	Солнечная
1500	47	Балкон	Затемненная	1680	72	Лоджия	Затемненная
1680	48	Балкон	Затемненная	1300	68	Лоджия	Затемненная
1150	76	Лоджия	Затемненная	1580	75	Лоджия	Затемненная
1700	57	Балкон	Солнечная	1180	58	Лоджия	Затемненная

Здесь: Y - стоимость **двухкомнатной** квартиры в городе N, тыс. руб.; X - общая площадь квартиры, м²;
D₁ - наличие балкона или лоджии; D₂ - сторона выхода окон квартиры.

Вариант 3

Y	X	D ₁	D ₂	Y	D ₁	D ₂
1980	84	Лоджия	Затемненная	2350	Лоджия	Затемненная
2250	81	Лоджия	Затемненная	1270	Балкон	Солнечная
1820	80	Балкон	Солнечная	1830	Балкон	Солнечная
1841	86	Балкон	Солнечная	2100	Балкон	Затемненная
2100	84	Балкон	Солнечная	2900	Балкон	Солнечная
1600	79	Лоджия	Затемненная	1800	Балкон	Солнечная
1900	79	Балкон	Солнечная	2100	Балкон	Солнечная
1950	96	Балкон	Затемненная	1100	Балкон	Затемненная

Окончание варианта 3

Y	X	D ₁	D ₂	Y	D ₁	D ₂
1650	88	Лоджия	Затемненная	2550	Балкон	Солнечная
1900	99	Балкон	Затемненная	1930	Балкон	Затемненная
2180	109	Балкон	Затемненная	2768	Балкон	Солнечная
2850	111	Балкон	Затемненная	1850	Лоджия	Солнечная
2430	111	Лоджия	Затемненная	2000	Балкон	Затемненная
3300	65	Балкон	Солнечная	2450	Балкон	Затемненная
2000	84	Лоджия	Затемненная	2100	Балкон	Солнечная
2500	70	Балкон	Солнечная	2800	Балкон	Солнечная
2350	54	Балкон	Затемненная	3200	Балкон	Затемненная
2380	64	Балкон	Солнечная	2400	Балкон	Солнечная
1650	45	Балкон	Солнечная	2650	Лоджия	Затемненная
2100	62	Балкон	Затемненная	2200	Балкон	Солнечная

Здесь: Y - стоимость **трехкомнатной** квартиры в городе N, тыс. руб.; X - общая площадь квартиры, м²;
D₁ - наличие балкона или лоджии; D₂ - сторона выхода окон квартиры.

Вариант 4

Марка	Стоимость, тыс. руб.	Пробег, тыс. км	Участие в ДТП	Топливо
	Y	X	D ₁	D ₂
Ваз-21099	110	40	Нет	Бензин
Ваз-2110	170	30	Нет	Газ
Ваз-2112	160	40	Нет	Бензин
Ваз-2109	80	50	Нет	Бензин
Ваз-2109	140	85	Нет	Бензин
Ваз-21099	160	90	Нет	Бензин
Ваз-21083	45	500	Да	Бензин
Ваз-2112	120	115	Нет	Газ
Ваз-2112	145	50	Да	Бензин
Ваз-2110	80	100	Да	Газ
Ваз-2115	40	50	Да	Газ
Ваз "ОКА"	75	18	Нет	Бензин
Ваз-21099	108	70	Нет	Газ
Ваз-2107 М	210	53	Да	Газ

	1	2	3	4	5	6	7	8	9	10	11	12	13
1	6. Регрессионная модель с двумя фиктивными переменными												
2													
3													

Окончание варианта 4

Марка	Стоимость, тыс. руб.	Пробег, тыс. км	Участие в ДТП	Топливо
	Y	X	D_1	D_2
Ваз-21102	105	99	Да	Бензин
Ваз-21099	110	47	Нет	Газ
Ваз-21099	90	70	Нет	Газ
Ваз-21102	155	50	Нет	Газ
Ваз-2114	230	14	Нет	Газ
Ваз-21140	185	27	Нет	Бензин
Ваз-2108	45	300	Да	Бензин
Ваз-21099	97	100	Да	Бензин
Ваз-2115	200	57	Нет	Газ
Ваз-21083	60	150	Нет	Газ
Ваз-2110	190	45	Да	Газ
Ваз-2112	175	50	Нет	Газ
Ваз-2110	160	84	Нет	Бензин
ОКА	70	30	Нет	Бензин
Ваз-21063	50	50	Нет	Бензин
Ваз-2107	110	35	Нет	Газ
Ваз-21074	125	32	Нет	Бензин
Ваз 21013	12	250	Да	Бензин
Ваз21013	15	200	Нет	Бензин
Уаз 3303	17	70	Нет	Бензин
Ваз21065	62	105	Нет	Газ
Ваз 2106	11	230	Да	Бензин
Ваз 21074	68	80	Нет	Газ
Ваз 21074	100	41	Нет	Бензин
Ваз 21053	40	58	Да	Бензин
Ваз 2107	80	25	Да	Бензин

Вариант 5

Марка	Стоимость, тыс. руб. (Y)	Год выпуска (X)	Участие в ДПТ (D ₁)	Топливо (D ₂)
Ваз-21099	110	2000	Нет	Бензин
Ваз-2110	170	2006	Нет	Газ
Ваз-2112	160	2005	Нет	Бензин
Ваз-2109	80	1995	Нет	Бензин
Ваз-2109	140	1995	Нет	Бензин
Ваз-21099	160	2004	Нет	Бензин
Ваз-21083	45	1988	Да	Бензин
Ваз-2112	120	2000	Нет	Газ
Ваз-2112	145	2001	Да	Бензин
Ваз-2110	80	2007	Да	Газ
Ваз-2115	40	2005	Да	Газ
Ваз "ОКА"	75	2006	Нет	Бензин
Ваз-21099	108	2000	Нет	Газ
Ваз-2107 М	210	2006	Да	Газ
Ваз-21102	105	2000	Да	Бензин
Ваз-21099	110	2001	Нет	Газ
Ваз-21099	90	2000	Нет	Газ
Ваз-21102	155	2003	Нет	Газ
Ваз-2114	230	2008	Нет	Газ
Ваз-21140	185	2007	Нет	Бензин
Ваз-2108	45	1993	Да	Бензин
Ваз-21099	97	1998	Да	Бензин
Ваз-2115	200	2004	Нет	Газ
Ваз-21083	60	1996	Нет	Газ
Ваз-2110	190	2007	Да	Газ
Ваз-2112	175	2005	Нет	Газ
Ваз-2110	160	2004	Нет	Бензин
ОКА	70	2006	Нет	Бензин
Ваз-21063	50	1995	Нет	Бензин
Ваз-2107	110	2003	Нет	Газ

	1	2	3	4	5	6	7	8	9	10	11	12	13
1	6. Регрессионная модель с двумя фиктивными переменными												
2													
3													

Окончание варианта 5

Марка	Стоимость, тыс. руб. (Y)	Год выпуска (X)	Участие в ДТП (D ₁)	Топливо (D ₂)
Ваз-21074	125	2007	Нет	Бензин
Ваз 21013	12	1983	Да	Бензин
Ваз21013	15	1987	Нет	Бензин
Уаз 3303	17	1988	Нет	Бензин
Ваз21065	62	1996	Нет	Газ
Ваз 2106	11	1985	Да	Бензин
Ваз 21074	68	2002	Нет	Газ
Ваз 21074	100	2006	Нет	Бензин
Ваз 21053	40	1997	Да	Бензин
Ваз 2107	80	2005	Да	Бензин

6.3. Тесты

1. Фиктивными переменными в модели множественной регрессии могут быть:

- А) количественные переменные;
- Б) переменные, исходные значения которых не имеют количественного значения;
- В) качественные переменные, преобразованные в количественные;
- Г) экономические переменные, преобразованные в количественные.

Выберите несколько вариантов ответа.

2. Если качественный признак, который необходимо отразить в регрессионной модели имеет две градации, то в уравнение включают:

- А) одну фиктивную переменную;
- Б) две фиктивные переменные;
- В) три фиктивные переменные.

3. Исследуется зависимость объема продаж компании от ряда факторов: X_1 - цены товара, X_2 - уровня образования персонала (высшее, среднее), X_3 - дохода потребителя, X_4 - расходов на рекламу. Какие из перечисленных переменных не являются фиктивными?

- А) X_4 ;
- Б) X_2 ;

	1	2	3	4	5	6	7	8	9	10	11	12	13
1			<i>Эконометрика в табличном редакторе MS Excel</i>										
2													
3													

В) X_1 ;

Г) X_3 .

Выберите несколько вариантов ответа.

4. Исследуется зависимость потребления кофе от следующих факторов: X_1 - страна - производитель кофе, X_2 - уровень крепости кофе (крепкий, средней крепости, слабой крепости), X_3 - доход потребителя, X_4 - цена кофе. Фиктивными переменными в модели являются:

А) X_4 ;

Б) X_2 ;

В) X_1 ;

Г) X_3 .

Выберите несколько вариантов ответа.

5. Для исследования заработной платы работника некоторого предприятия используется модель

$$Y_i = b_0 + b_1X_i + c_1D_{1i} + c_2D_{2i} + c_3D_{3i} + \varepsilon_i,$$

где Y_i - заработная плата i -го работника;

X_i - общий стаж его работы;

D_{1i} - уровень образования (переменная, принимающая значение 1, если работник с высшим образованием, и 0 в противном случае);

D_{2i} - наличие детей (переменная, принимающая значение 1, если у работника есть дети, и 0 в противном случае);

D_{3i} - пол.

Укажите, сколько факторов необходимо представить в модели фиктивными переменными:

А) 4;

Б) 3;

В) 1.

6. Исследуется влияние уровня образования на заработную плату работника некоторого предприятия, где трудятся люди со средним специальным, средним, высшим и незаконченным высшим образованием (всего 4 категории). Тогда максимальное количество фиктивных переменных, необходимых для проведения анализа, равно:

А) 4;

Б) 2;

В) 3;

Г) 1.

7. Исследуется зависимость цены туристической путевки на одного человека Y (руб.) от продолжительности отдыха X (число дней), типа питания D_1 (фиктивная переменная, принимающая значение 1 для типа питания «все включено» и значение 0 для типа питания «полупансион») и страны пребывания D_2 (фиктивная переменная, принимающая зна-

	1	2	3	4	5	6	7	8	9	10	11	12	13
1		6. Регрессионная модель с двумя фиктивными переменными											
2													
3													

чение 1 для Испании и значение 0 в остальных случаях). Получены следующие результаты оценки параметров эконометрических моделей.

Результаты оценки параметров модели (1) $Y = b_0 + b_1X + \varepsilon$:

$$R^2 = 0,508 ; \overline{R^2} = 0,485 ; F_{набл} = 22,69 \text{ («Значимость } F \text{» } 0,0000);$$

Показатель	Коэффициент	Стандартная ошибка	t-статистика	p-значение
Y-пересечение	-12712,7	11387,6	-1,11637	0,27631
X	3639,938	764,0704	4,763877	9,35E-05

Результаты оценки параметров модели (2) $Y = b_0 + b_1X + c_1D_1 + \varepsilon$:

$$R^2 = 0,587 ; \overline{R^2} = 0,548 ; F_{набл} = 14,94 \text{ («Значимость } F \text{» } 0,0000);$$

Показатель	Коэффициент	Стандартная ошибка	t-статистика	p-значение
Y-пересечение	-12727,6	10672,86	-1,19252	0,246357
X	3439,896	722,9876	4,757891	0,000106
D ₁	7097,213	3528,7	2,011283	0,057306

Результаты оценки параметров модели (3) $Y = b_0 + b_1X + c_2D_2 + \varepsilon$:

$$R^2 = 0,539 ; \overline{R^2} = 0,495 ; F_{набл} = 12,25 \text{ («Значимость } F \text{» } 0,0003);$$

Показатель	Коэффициент	Стандартная ошибка	t-статистика	p-значение
Y-пересечение	-3304,33	13807,49	-0,23931	0,813182
X	3120,066	875,564	3,563492	0,001835
D ₂	-5285,89	4469,211	-1,18274	0,250132

Результаты оценки параметров модели (4) $Y = b_0 + b_1X + c_1D_1 + c_2D_2 + \varepsilon$:

$$R^2 = 0,592 ; \overline{R^2} = 0,531 ; F_{набл} = 9,659 \text{ («Значимость } F \text{» } 0,0004).$$

Показатель	Коэффициент	Стандартная ошибка	t-статистика	p-значение
Y-пересечение	-8840,82	13744,47	-0,64323	0,527383
X	3246,155	847,5698	3,829956	0,001047
D ₁	6354,157	3938,912	1,613176	0,122374
D ₂	-2182,85	4717,833	-0,46268	0,648588

Моделью лучшего статистического качества из представленных моделей (при уровне значимости 0,1) является:

- А) модель (1);
- Б) модель (2);

	1	2	3	4	5	6	7	8	9	10	11	12	13
1			<i>Эконометрика в табличном редакторе MS Excel</i>										
2													
3													

- В) модель (3);
Г) модель (4).

8. Укажите, какая из моделей теста 5 содержит фиктивную переменную сдвига:

- А) модель (1);
Б) модель (2);
В) модель (3);
Г) модель (4).

Выберите несколько вариантов ответа.

9. Проанализировав результаты оценки параметров модели (2) из теста 5, выберите верное утверждение:

А) вариация цены путевки на 0,587% объясняется вариацией продолжительности отдыха и типа питания;

Б) при условии одинаковой продолжительности отдыха путевка с типом питания «все включено» в среднем на 7097,2 руб. дороже, чем путевка с типом питания «полупансион»;

В) при условии одинаковой продолжительности отдыха путевка с типом питания «полупансион» в среднем на 12712,7 руб. дешевле, чем путевка с типом питания «все включено»;

Г) при увеличении продолжительности отдыха на 1 день цена путевки увеличивается в среднем на 7097,2 тыс. руб.

10. Укажите, какая из нижеперечисленных моделей содержит фиктивную переменную наклона:

- А) $Y = b_0 + b_1X + \varepsilon$;
Б) $Y = b_0 + b_1X + c_1D_1 + \varepsilon$;
В) $Y = b_0 + b_1X + c_1D_1 + c_2XD_1 + \varepsilon$;
Г) $Y = b_0 + b_1X + c_1D_1 + c_2D_2 + \varepsilon$.

6.4. Контрольные вопросы

1. Приведите примеры фиктивных переменных.
2. Что представляет собой «фиктивная переменная сдвига»?
3. Как интерпретируется коэффициент при фиктивной переменной сдвига?
4. Что представляет собой «фиктивная переменная наклона»?
5. Как интерпретируется коэффициент при фиктивной переменной наклона?

	1	2	3	4	5	6	7	8	9	10	11	12	13
1		<i>6. Регрессионная модель с двумя фиктивными переменными</i>											
2													
3													

6. Как интерпретируется константа при использовании фиктивной переменной наклона?
7. Как определяется значимость одновременно нескольких фиктивных переменных?
8. Что представляет собой модель ANCOVA?
9. Что представляет собой модель ANOVA?
10. Предполагается, что ежемесячное потребление кофе студентами линейно связано с доходом, возрастом студента и формой обучения (заочная/очная). Сколько количественных и качественных объясняющих переменных должна содержать модель?

7. МУЛЬТИКОЛЛИНЕАРНОСТЬ В МОДЕЛИ МНОЖЕСТВЕННОЙ ЛИНЕЙНОЙ РЕГРЕССИИ



Изложен алгоритм обнаружения и устранения мультиколлинеарности в регрессионных моделях, продемонстрированный на конкретном примере с использованием некоторых инструментов анализа данных табличного редактора MS Excel.

7.1. Решение типовых задач

Для иллюстрации алгоритма построения множественной регрессионной модели в условиях возможной мультиколлинеарности воспользуемся статистическими данными из табл. 4.1. Зависимая переменная Y может быть объяснена пятью переменными X_1, X_2, X_3, X_4, X_5 . Расшифровка системы показателей:

Y - валовой региональный продукт, руб.;

X_1 - стоимость основных фондов (по полной учетной стоимости), млн руб.;

X_2 - объем производства по виду деятельности «Добыча полезных ископаемых» (в фактически действовавших ценах), млн руб.;

X_3 - объем производства по виду деятельности «Обрабатывающие производства» (в фактически действовавших ценах), млн руб.;

X_4 - объем производства по виду деятельности «Производство и распределение электроэнергии, газа и воды» (в фактически действовавших ценах), млн руб.;

X_5 - продукция сельского хозяйства (в фактически действовавших ценах), млн руб.

Статистический массив собран из официальных статистических источников по субъектам Центрального федерального округа РФ за 2016 г. (www.gks.ru).

7.1.1. Выявление мультиколлинеарности на основе матрицы парных коэффициентов корреляции (МПКК)

Для выявления возможных пар мультиколлинеарных объясняющих переменных следует воспользоваться МПКК. Для расчета матрицы парных коэффициентов в главном меню выберем **Данные / Анализ данных / Корреляция** (рис. 7.1).

7. Мультиколлинеарность в модели множественной линейной регрессии

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Субъект ЦФО РФ	Y	X1	X2	X3	X4	X5							
2	Белгородская область	730562	1400837	89048	597751	26504	226056							
3	Брянская область	285848	733055	281	176003	15439	78437,4							
4	Владимирская область	392052	787525	3068	406936	33727	30834,9							
5	Воронежская область	841376	1534644	5576	408214	59609	199325							
6	Ивановская область	179633	553289	892	120938	29674	14987							
7	Калужская область	373404	912722	2983	536624	20076	33458							
8	Костромская область	160705	419515	373	107792	36246	17090							
9	Курская область	364602	805024	49246	160486	58522	124357							
10	Липецкая область	470239	1178985	4941	572356	25824	106325							
11	Московская область	3565258	7237874	12685	2126307	254767	98372,1							
12	Орловская область	213924	472195	134	103982	13089	68568							
13	Рязанская область	336974	927211	1503	252967	37227	51048							
14	Смоленская область	262318	829388	1423	178891	54349	21243,4							
15	Тамбовская область	311433	798260	127	137330	15520	108896							
16	Тверская область	359345	1212653	1088	216368	85524	32882,2							
17	Тульская область	517741	1033088	4336	564749	47813	56294,1							
18	Ярославская область	469805	1226233	996	316472	33450	31852,3							

Корреляция

Входные данные: \$B\$1:\$G\$18

Входной интервал: \$B\$1:\$G\$18

Группирование: по столбцам по строкам

Метки в первой строке

Параметры вывода: Выходной интервал: \$I\$2:\$I\$18 Новый рабочий лист Новая рабочая книга

OK Отмена Справка

Рис. 7.1. Построение МПКК в условиях количества объясняющих переменных более одной

ЗАМЕЧАНИЕ: в поле «*Выходной интервал*» необходимо выделить весь статистический массив переменных. При выборе массива с заголовками столбцов требуется активировать флажок «*Метки в первой строке*».

Результат построения МПКК приведен на рис. 7.2.

	I	J	K	L	M	N	O
		Y	X1	X2	X3	X4	X5
Y		1					
X1		0,99526	1				
X2		0,13041	0,09319	1			
X3		0,96503	0,95927	0,1488	1		
X4		0,93367	0,94807	0,03357	0,86754	1	
X5		0,26348	0,20785	0,69682	0,22883	0,07823	1

Рис. 7.2. Итог построения МПКК

МПКК симметрична относительно главной диагонали, содержит значения парных коэффициентов корреляции между соответствующими переменными. Первый столбец (желтый) содержит коэффициенты корреляции: r_{YX_1} , r_{YX_2} , r_{YX_3} , r_{YX_4} , r_{YX_5} . Проанализировав значения этого столбца, можно сделать вывод, что на переменную Y наибольшее влияние оказывает переменная X_1 . В столбцах 2-5 приведены парные коэффициенты корреляции между объясняющими переменными. По шкале Чеддока мультиколлинеарными признаны пары переменных: X_1X_3 , X_1X_4 , X_3X_4 ($r_{X_iX_j} > 0,7$). Следовательно, совместное включение данных переменных в множественную линейную регрессионную модель непродуктивно, так как однозначно установить их влияние на объясняемую пере-

менную невозможно. Оценки, полученные по МНК, в этом случае будут неустойчивыми (неэффективными).

7.1.2. Показатель VIF (Variance Inflation Factor)

Одним из способов выявления мультиколлинеарности в модели множественной регрессии является коэффициент **VIF**, который вычисляется по формуле $VIF_j = \frac{1}{1-R_j^2}$, где R_j^2 - значение коэффициента детерминации, полученное при оценке регрессии X_j на остальные объясняющие переменные.

Считается, что в случае $VIF > 10$ объясняющие переменные, коррелирующие между собой, считаются мультиколлинеарными.

Таблица 7.1

Возможные значения коэффициента VIF

VIF	Статус объясняющих переменных
$VIF < 1$	Не коррелируют
$1 < VIF < 5$	Умеренно коррелируют
$5 < VIF < 10$	Сильно коррелированы
$VIF > 10$	Высоко коррелированы (мультиколлинеарность)

Шаг 1. Проиллюстрируем нахождение показателя VIF для зависимости вида $X_1 = s_0 + s_2X_2 + s_3X_3 + s_4X_4 + s_5X_5 + \varepsilon$. Оценим данную модель с помощью инструмента анализа данных «РЕГРЕССИЯ».

В окне «РЕГРЕССИЯ», в поле «**Входной интервал Y**» укажем значения переменной X_1 , а в поле «**Входной интервал X**» - оставшиеся объясняющие переменные X_2, X_3, X_4, X_5 . Итоги оценивания модели представлены на рис. 7.3.

Шаг 2. Значение VIF-коэффициента равно 24,29. Согласно шкале, представленной в табл. 7.1, существует сильная корреляция между переменной X_1 с остальными объясняющими переменными. Таким образом, наблюдается мультиколлинеарность объясняющих переменных.

7. Мультиколлинеарность в модели множественной линейной регрессии

линейн							
Н	И	Ж	К	Л	М	Н	О
2	ВЫВОД ИТОГОВ						
3							
4	Регрессионная статистика						
5	Множественный R	0,989549186		VIF			
6	R-квадрат	0,979207591		=1/((1-J6^2))			
7	Нормированный R-квадрат	0,972276788					
8	Стандартная ошибка	260058,424					
9	Наблюдения	17					
10	Дисперсионный анализ						
12		df	SS	MS	F	Значимость F	
13	Регрессия	4	3,82202E+13	9,55505E+12	141,2834257	5,55542E-10	
14	Остаток	12	8,11565E+11	67630383904			
15	Итого	16	3,90318E+13				
17		Коэффициенты	Стандартная ошибка	t-статистика	P-Значение	Нижние 95%	Верхние 95%
18	Y-пересечение	-225341,181	120416,5271	-1,871347616	0,085868834	-487706,2553	37023,89319
19	X2	-4,655595524	3,88152946	-1,199422952	0,253510675	-13,11272171	3,801530663
20	X3	1,712956734	0,290150691	5,903679656	7,21159E-05	1,080772685	2,345140783
21	X4	13,61164182	2,404229448	5,661540261	0,000105338	8,37327585	18,85000779
22	X5	2,482808249	1,505269375	1,649411255	0,124975269	-0,796891977	5,762508475

Н	И	Ж	К	Л
2	ВЫВОД ИТОГОВ			
3				
4	Регрессионная статистика			
5	Множественный R	0,989549186		VIF
6	R-квадрат	0,979207591		24,29986396
7	Нормированный R-квадрат	0,972276788		
8	Стандартная ошибка	260058,424		
9	Наблюдения	17		

Рис. 7.3. Результаты оценивания регрессии вида

$$X_1 = s_0 + s_2 X_2 + s_3 X_3 + s_4 X_4 + s_5 X_5 + \varepsilon$$

Аналогично проводят расчет для оставшихся вариантов зависимостей. Результаты данных расчетов представлены в табл. 7.2.

Таблица 7.2

Результаты расчета VIF-коэффициента для всех вариантов зависимостей между объясняющими переменными

Зависимость	R^2	VIF	Статус объясняющих переменных
$X_1 = s_0 + s_2 X_2 + s_3 X_3 + s_4 X_4 + s_5 X_5 + \varepsilon$	0,98	24,29	Мультиколлинеарность
$X_2 = s_0 + s_1 X_1 + s_3 X_3 + s_4 X_4 + s_5 X_5 + \varepsilon$	0,37	1,16	Умеренно коррелируют
$X_3 = s_0 + s_1 X_1 + s_2 X_2 + s_4 X_4 + s_5 X_5 + \varepsilon$	0,94	8,99	Сильно коррелированы
$X_4 = s_0 + s_1 X_1 + s_2 X_2 + s_3 X_3 + s_5 X_5 + \varepsilon$	0,95	9,97	Сильно коррелированы
$X_5 = s_0 + s_1 X_1 + s_2 X_2 + s_3 X_3 + s_4 X_4 + \varepsilon$	0,36	1,15	Умеренно коррелируют

Таким образом, если исходить из матрицы парных коэффициентов корреляции, оптимальной может быть признана модель $Y = b_0 + b_1X_1 + b_5X_5 + \varepsilon$. Результаты ее оценивания по МНК приведены на рис. 7.4 ($\alpha = 0,05$).

Вывод итогов					
Регрессионная статистика					
Множественный R		0,996942821			
R-квадрат		0,993894989			
Нормированный R-квадрат		0,993022844			
Стандартная ошибка		66011,40672			
Наблюдения		17			
Дисперсионный анализ					
	df	SS	MS	F	Значимость F
Регрессия	2	9,93162E+12	4,96581E+12	1139,599023	3,16086E-16
Остаток	14	61005081436	4357505817		
Итого	16	9,99262E+12			
	Коэффициенты	Стандартная ошибка	t-статистика	P-Значение	Нижние 95%
Y-пересечение	-124605,0673	27653,75655	-4,505900204	0,000493587	-183916,4762
Переменная X 1	0,497356997	0,010801892	46,04350561	1,09877E-16	0,474189242
Переменная X 2	0,754266429	0,272161935	2,771388393	0,015001222	0,170537133

Рис. 7.4. Результаты оценивания модели $Y = b_0 + b_1X_1 + b_5X_5 + \varepsilon$

ЗАМЕЧАНИЕ: переменная X_2 исключена из модели в силу незначимости оценки соответствующего коэффициента.

7.2. Задача для самостоятельного выполнения

Задача 17. Используя статистические данные задачи 14, необходимо определить вид спецификации модели множественной линейной регрессии с отсутствием мультиколлинеарности (если это возможно!). Использовать матрицу парных коэффициентов корреляции, показатель VIF. Сформировать аналитическую записку.

ЗАМЕЧАНИЕ: используйте соответствующий вариант задания!!!

7.3. Тесты

1. Полная мультиколлинеарность - это:
 - А) нестрогая линейная зависимость между объясняющими переменными;
 - Б) отсутствие корреляции переменных;
 - В) наличие функциональной связи между зависимой переменной и независимыми переменными;

7. Мультиколлинеарность в модели множественной линейной регрессии

- Г) оценка тесноты корреляционной взаимосвязи между переменными;
Д) нет правильного ответа.

2. Частичная мультиколлинеарность - это:

- А) отсутствие корреляции переменных;
Б) наличие корреляционной зависимости между объясняющими переменными;
В) оценка тесноты корреляционной взаимосвязи между переменными;
Г) нестрогая линейная зависимость между объясняющими переменными.

3. Совместно включать в множественное уравнение регрессии факторы X_1 и X_2 , если коэффициент корреляции между ними равен 0,87:

- А) можно;
Б) нельзя;
В) и да, и нет, корреляция между ними не имеет существенного значения.

4. В ходе проведения корреляционного анализа получена матрица парных коэффициентов корреляции, приведенная ниже.

	Y	X_1	X_2
Y	1		
X_1	0,607	1	
X_2	0,889	0,78	1

Ее анализ свидетельствует о том, что:

- А) между факторами X_1 и X_2 существует высокая корреляционная взаимосвязь;
Б) оценки коэффициентов при факторах X_1 и X_2 в уравнении множественной регрессии будут статистически значимы;
В) одна из оценок коэффициентов регрессии при факторах X_1 и X_2 в уравнении множественной регрессии ожидается статистически незначимой;
Г) между факторами X_1 и X_2 существует умеренная корреляционная взаимосвязь.

5. Стохастическая мультиколлинеарность, т.е. наличие сильной линейной корреляционной связи между независимыми переменными, классифицируется как:

- А) строгая;
Б) умеренная;
В) нестрогая;
Г) слабая.

6. Не способно исключить проблему мультиколлинеарности:

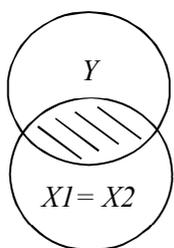
- А) исключение из регрессионной модели незначимых переменных;

- Б) исключение из регрессионной модели переменных, имеющих сильную корреляционную зависимость с другими факторами;
- В) получение дополнительных данных или новой выборки;
- Г) использование метода взвешенных наименьших квадратов.

7. Коэффициент детерминации R^2 показывает:

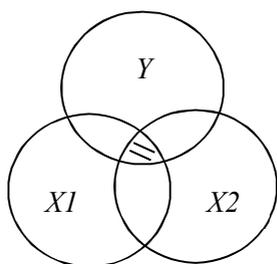
- А) наличие мультиколлинеарности в модели;
- Б) степень взаимосвязи между объясняющими переменными;
- В) долю вариации зависимой переменной, обусловленную вариацией объясняющих переменных.

8. На диаграмме Венна схематично изображена мультиколлинеарность:



- А) строгая;
- Б) умеренная;
- В) нестрогая;
- Г) слабая.

9. На диаграмме Венна схематично изображена мультиколлинеарность:



- А) строгая;
- Б) умеренная;
- В) нестрогая;
- Г) слабая.

10. Дана матрица парных коэффициентов корреляции ($n = 30$):

	Y	X_1	X_2	X_3	X_4
Y	1				
X_1	0,85	1			
X_2	0,74	0,13	1		
X_3	0,15	0,45	0,74	1	
X_4	0,12	0,12	0,71	0,82	1

После исключения мультиколлинеарности в модели останутся факторы:

- А) X_1, X_2 ;

	1	2	3	4	5	6	7	8	9	10	11	12	13
1	7. Мультиколлинеарность в модели множественной линейной регрессии												
2													
3													

- Б) X_1, X_2, X_3 ;
- В) X_1, X_2, X_6 ;
- Г) X_1, X_2, X_3, X_6 .

7.4. Контрольные вопросы

1. Что такое мультиколлинеарность?
2. Каковы основные причины возникновения мультиколлинеарности?
3. Какие виды мультиколлинеарности встречаются?
4. Как можно обнаружить мультиколлинеарность?
5. Для чего используется матрица парных коэффициентов корреляции?
6. Что такое показатель VIF?
7. Какие значения принимает показатель VIF?
8. Каковы последствия мультиколлинеарности в регрессионном анализе?
9. Как влияет мультиколлинеарность на значимость оценок коэффициентов регрессии?
10. Какие существуют методы устранения мультиколлинеарности?

8. ГЕТЕРОСКЕДАСТИЧНОСТЬ СЛУЧАЙНОЙ КОМПОНЕНТЫ ЭКОНОМЕТРИЧЕСКОЙ МОДЕЛИ



Представлена практическая реализация обнаружения нарушения одного из условий Гаусса - Маркова (гетероскедастичности) с использованием табличного редактора MS Excel. Рассмотрен графический способ обнаружения гетероскедастичности, а также выполнен тест Голдфелда - Квандта.

8.1. Решение типовых задач

Оценивание уравнения линейной регрессии по МНК позволяет получить эффективные оценки теоретических коэффициентов в случае выполнения предпосылки: $D(\varepsilon_i) = \sigma_\varepsilon^2, i = \overline{1, n}$. Выполнение данного условия (предпосылки) называется гомоскедастичностью остатков регрессии, нарушение - гетероскедастичностью остатков регрессии.

Практическую реализацию процесса обнаружения гетероскедастичности в табличном редакторе Excel продемонстрируем на статистических данных табл. 2.1, отражающей значения таких показателей, как «инвестиции в основной капитал» (Y) и «среднегодовая численность занятых» (X) в разрезе муниципальных образований Самарской области (2017 г.).

8.1.1 Графический способ обнаружения гетероскедастичности

В результате оценивания модели парной линейной регрессии вида $Y = b_0 + b_1X + \varepsilon$ получено следующее выборочное уравнение: $\hat{y} = 708034,96 + 307x$.

Проверка статистической значимости оценок коэффициентов регрессии показала значимость оценки \hat{b}_1 и незначимость оценки \hat{b}_0 (t -тест). Коэффициент детерминации $R^2 = 0,37$, при этом модель статистически значима (F -тест). Данные факты были доказательно изложены в рамках темы 2 настоящего учебного пособия.

Шаг 1. Необходимо проверить выполнимость предпосылки МНК вида $D(\varepsilon_i) = \sigma_\varepsilon^2, i = \overline{1, n}$. Для этого следует вычислить остатки регрессии: $e_i = y_i - \hat{y}_i$, где \hat{y}_i -

8. Гетероскедастичность случайной компоненты эконометрической модели

расчетные значения объясняемой переменной исходя из вида выборочного уравнения регрессии (рис. 8.1).

	A	B	C	D	E
1	Мун. образование	X	Y	Y[^]	ei
2	Алексеевский м.р.	1546	12878	1182655	=C2-D2
3	Безенчукский м.р.	8082	6540234	3189198	3351036
4	Богатовский м.р.	2767	905057	1557500	-652443
5	Большеглушицкий м.р.	2791	1587301	1564868	22432,7
6	Большечерниговский м.р.	2795	631314	1566096	-934782
7	Борский м.р.	3006	300426	1630873	-1330447

Рис. 8.1. Вычисление остатков регрессии ($\hat{y} = 708034,96 + 307x$)

Следует отметить, что остатки регрессии можно исчислить с помощью инструмента анализа данных «РЕГРЕССИЯ». Для этого необходимо установить флажок в поле «Остатки». Итоги работы данного инструмента будут содержать таблицу со значениями \hat{y}_i и e_i (рис. 8.2).

	D	E	F	G	H	I	J	K	L
1	Y[^]	ei		ВЫВОД ИТОГОВ					
2	1182655	-1169777							
3	3189198	3351036		<i>Регрессионная статистика</i>					
4	1557500	-652443		Множественный R	0,607187654				
5	1564868	22432,67		R-квадрат	0,368676847				
6	1566096	-934782		Нормированный R-квадрат	0,349545842				
7	1630873	-1330447		Стандартная ошибка	3481735,217				
8	6534870	12316094		Наблюдения	35				
9	1079503	-867274		<i>Дисперсионный анализ</i>					
10	1350890	-1302177							
11	1171603	-107384			<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Значимость F</i>
12	2017077	1460174		Регрессия	1	2,33614E+14	2,33614E+14	19,27117021	0,000109901
13	2913207	-1135731		Остаток	33	4,00042E+14	1,21225E+13		
14	1406457	325660		Итого	34	6,33656E+14			
15	1915461	-89519,8							
16	1577148	-1119431			<i>Коэффициенты</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>	<i>P-Значение</i>	<i>Нижние 95%</i>
17	3387520	-1835761		Y-пересечение	708034,9585	780223,0876	0,907477579	0,370733745	-879340,8486
18	2982588	9481504		X	306,9986989	69,93305349	4,3898941	0,000109901	164,7188318
19	1600787	-599021							
20	1392642	-434736					t кр. дв.		
21	1763803	-1373402					2,030107928		
22	3675791	2914125		ВЫВОД ОСТАТКА					
23	3678554	-1190955							
24	1729420	-650393		<i>Наблюдение</i>	<i>Предсказанное Y</i>	<i>Остатки</i>			
25	1395712	-781904		1	1182654,947	-1169776,947			
26	1481979	-919678		2	3189198,443	3351035,557			
27	1413211	-1270061		3	1557500,358	-652443,3585			
28	1621970	-1551302		4	1564868,327	22432,67273			

Рис. 8.2. Вывод остатков регрессии в итогах работы инструмента анализа данных «РЕГРЕССИЯ»

Шаг 2. Самым наглядным и технически простым методом распознавания гетероскедастичности является графический метод. Этот метод предполагает построение графика зависимости остатков регрессии от значений факторного признака (в случае парной зависимости).

ЗАМЕЧАНИЕ: если имеет место множественная регрессионная модель, необходимо графически исследовать зависимость остатков регрессии от расчетного значения зависимой переменной.

В табличном редакторе Excel построение данного графика возможно двумя способами.

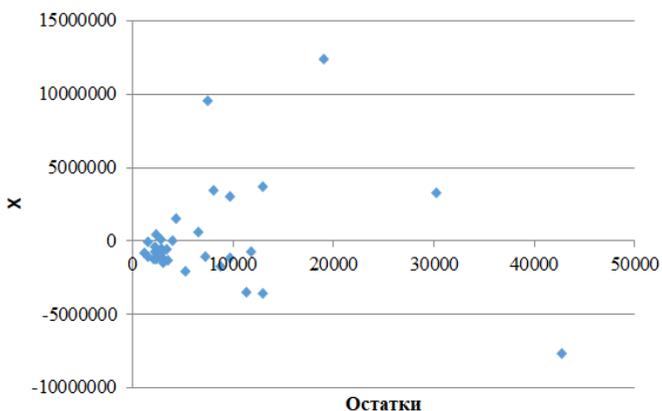


Рис. 8.3. Графический анализ остатков регрессии $\hat{y} = 708034,96 + 307x$ (стандартная диаграмма)

Первый способ - построение стандартной точечной диаграммы по технологии, описанной в теме 1. Результат построения отражен на рис. 8.3.

Второй способ - построение графика остатков с использованием инструмента анализа данных «РЕГРЕССИЯ». Необходимо включить флажок в поле «График остатков». Результат построения так же отражен на рис. 8.4.

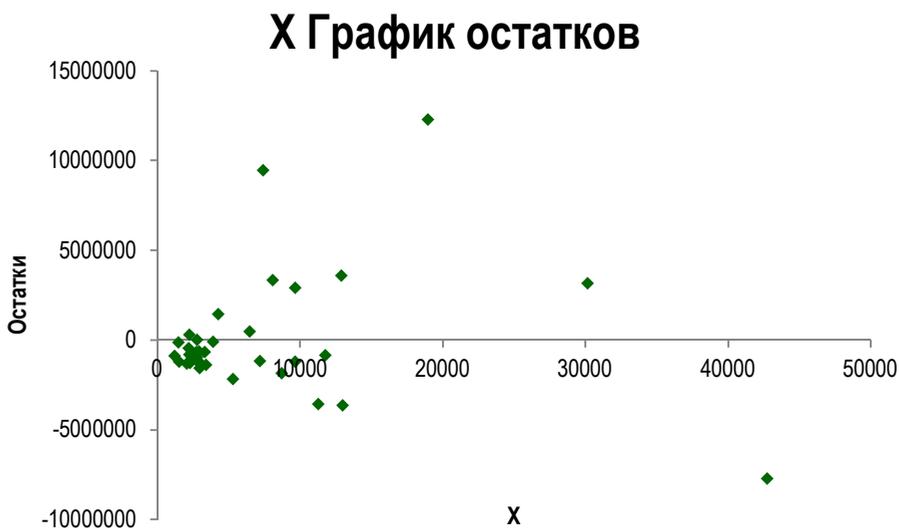


Рис. 8.4. Графический анализ остатков регрессии $\hat{y} = 708034,96 + 307x$ («РЕГРЕССИЯ»)

8. Гетероскедастичность случайной компоненты эконометрической модели

Если все остатки регрессии располагаются внутри полосы постоянной ширины, то очевидно выполнение предпосылки МНК $D(\varepsilon_i) = \sigma_\varepsilon^2, i = \overline{1, n}$. В данном случае этого не наблюдается. С большой долей вероятности можно говорить о гетероскедастичности остатков модели.

Вместе с тем, исходя из расположения точек на рис. 8.4 гетероскедастичность остатков можно описать выражением следующего вида: $\sigma_\varepsilon \approx \lambda x_i$ (прямопропорциональная гетероскедастичность).

8.1.2. Аналитический способ выявления гетероскедастичности

А) Тест ранговой корреляции Спирмена

Для нахождения значений коэффициента корреляции рангов Спирмена воспользуемся формулой

$$r_{xe} = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)},$$

где $d_i = R(x_i) - R(|e_i|)$;

$R(x_i), R(|e_i|)$ - ранги значений объясняющей переменной и остатков регрессии соответственно;

n - объем выборки.

Шаг 1. На первом этапе вычисления рангового коэффициента корреляции Спирмена дополним исходную таблицу тремя новыми столбцами. В этих столбцах вычислим ранги объясняющей переменной и остатков регрессии, а также найдем квадрат разности рангов d_i^2 (рис. 8.5).

В процессе вычислений использовались следующие функции табличного процессора Excel:

ABS (число) - находит абсолютное значение числа, указанного в диалоговом поле функции (рис. 8.5а);

РАНГ (число; ссылка; порядок) - определяет ранг числа (число) в массиве значений (ссылка) в определенном порядке (0 - убывание; 1 - возрастание) (рис. 8.5б).

Эконометрика в табличном редакторе MS Excel

ЛИНЕЙН =ABS(E2)						
	A	B	C	D	E	F
1	Мун. образование	X	Y	Y^	ei	ei
2	Алексеевский м.р.	1546	12878	1182655	-1169776,	=ABS(E2)
3	Безенчукский м.р.	8082	6540234	3189198	3351035,557	ABS(число)
4	Богатовский м.р.	2767	905057	1557500	-652443,3585	652443
5	Большеглушицкий м.р.	2791	1587301	1564868	22432,67273	22432,7
6	Большечерниговский м.р.	2795	631314	1566096	-934782,3221	934782

а)

ЛИНЕЙН =РАНГ(F2:\$F\$2:\$F\$36;1)										
	A	B	C	D	E	F	G	H	I	J
1	Мун. образование	X	Y	Y^	ei	ei	R(x)	R(ei)		
2	Алексеевский м.р.	1546	12878	1182655	-1169776,947	1169777		=РАНГ(F2:\$F\$2:\$F\$36;1)		
3	Безенчукский м.р.	8082	6540234	3189198	3351035,557	3351036		РАНГ(число; ссылка; [порядок])		
4	Богатовский м.р.	2767	905057	1557500	-652443,3585	652443	10	9		
5	Большеглушицкий м.р.	2791	1587301	1564868	22432,67273	22432,7	11	1		
6	Большечерниговский м.р.	2795	631314	1566096	-934782,3221	934782	12	14		
7	Борский м.р.	3006	300426	1630873	-1330447,048	1330447	16	21		

б)

ЛИНЕЙН =(G2-H2)^2									
	A	B	C	D	E	F	G	H	I
1	Мун. образование	X	Y	Y^	ei	ei	R(x)	R(ei)	d^2
2	Алексеевский м.р.	1546	12878	1182655	-1169776,947	1169777	3		=(G2-H2)^2
3	Безенчукский м.р.	8082	6540234	3189198	3351035,557	3351036	25	29	16
4	Богатовский м.р.	2767	905057	1557500	-652443,3585	652443	10	9	1
5	Большеглушицкий м.р.	2791	1587301	1564868	22432,67273	22432,7	11	1	100
6	Большечерниговский м.р.	2795	631314	1566096	-934782,3221	934782	12	14	4

в)

Рис. 8.5. Поэтапное вычисление квадрата разности рангов d_i^2

Шаг 2. Далее вычисляется непосредственно коэффициент ранговой корреляции Спирмена с использованием приведенной выше формулы (рис. 8.6).

ЛИНЕЙН =-1-6*СУММ(I2:I36)/(35*(35^2-1))										
	A	B	C	D	E	F	G	H	I	J
1	Мун. образование	X	Y	Y^	ei	ei	R(x)	R(ei)	d^2	r(xe)
2	Алексеевский м.р.	1546	12878	1182655	-1169776,947	1169777	3	17	196	=-1-6*СУММ(I2:I36)/(35*(35^2-1))
3	Безенчукский м.р.	8082	6540234	3189198	3351035,557	3351036	25	29	16	
4	Богатовский м.р.	2767	905057	1557500	-652443,3585	652443	10	9	1	
5	Большеглушицкий м.р.	2791	1587301	1564868	22432,67273	22432,7	11	1	100	
6	Большечерниговский м.р.	2795	631314	1566096	-934782,3221	934782	12	14	4	
7	Борский м.р.	3006	300426	1630873	-1330447,048	1330447	16	21	25	
8	Волжский м.р.	18980	18850964	6534870	12316093,74	1,2E+07	33	35	4	
9	Елховский м.р.	1210	212229	1079503	-867274,3843	867274	1	12	121	
10	Иса克林ский м.р.	2094	48713	1350890	-1302177,234	1302177	4	20	256	

Рис. 8.6. Вычисление рангового коэффициента корреляции Спирмена

8. Гетероскедастичность случайной компоненты эконометрической модели

В результате вычислений $r_{\text{хе}} = 0,67$, что свидетельствует о заметной корреляции между значениями объясняющей переменной и остатками регрессии по шкале Чеддока.

ЗАМЕЧАНИЕ: необходимо проверить статистическую значимость рангового коэффициента корреляции с помощью t -статистики (см. тему 2).

Б) Тест Голдфелда - Кванта

	A	B	C
1	Мун. образование	X	Y
2	Алексеевский м.р.	1546	12878
3	Безенчукский м.р.	8082	6540234
4	Богатовский м.р.	2767	905057
5	Большеглушицкий м.р.	2791	1587301
6	Большечерниговский м.р.	2795	631314
7	Борский м.р.	3006	300426
8	Волжский м.р.	18980	1.9E+07
9	Елховский м.р.	1210	212229
10	Исаклинский м.р.	2094	48713
11	Камышлинский м.р.	1510	1064219
12	Кинельский м.р.	4264	3477251
13	Кинель-Черкасский м.р.	7183	1777476
14	Клявлинский м.р.	2275	1732117

а)

	A	B	C
1	Мун. образование	X	Y
2	Елховский м.р.	1210	212229
3	Камышлинский м.р.	1510	1064219
4	Алексеевский м.р.	1546	12878
5	Исаклинский м.р.	2094	48713
6	Похвистневский м.р.	2230	957906
7	Хворостянский м.р.	2240	613808
8	Клявлинский м.р.	2275	1732117
9	Шенталинский м.р.	2297	143150
10	Челно-Вершинский м.р.	2521	562301
11	Богатовский м.р.	2767	905057
12	Большеглушицкий м.р.	2791	1587301
13	Большечерниговский м.р.	2795	631314
14	Красноармейский м.р.	2831	457717
15	Пестравский м.р.	2908	1001766

б)

Рис. 8.7. Этапы СОРТИРОВКИ данных в табличном редакторе Excel

Шаг 1. Для проведения теста необходимо исходный статистический массив (X, Y) перенести на новый лист табличного редактора Excel, а затем упорядочить всю совокупность значений по величине объясняющей переменной (X) . Для этого выбираем Главная / Сортировка и фильтры / Сортировка от А до Я (рис. 8.7).

Шаг 2. Упорядоченный массив разбиваем на три равные части и оцениваем регрессии Y на X по первой и последней подвыборкам с помощью функции ЛИНЕЙН (возможно также применение инструмента анализа данных «РЕГРЕССИЯ»). Результаты оценивания представлены на рис. 8.8.

	A	B	C	D	E	F	G
1	Мун. образование	X	Y				
2	Елховский м.р.	1210	212229		524,0654243	-405991,0208	
3	Камышлинский м.р.	1510	1064219		345,6274914	756965,7907	
4	Алексеевский м.р.	1546	12878		0,203475113	561525,0785	
5	Исаклинский м.р.	2094	48713		2,299081978	9	
6	Похвистневский м.р.	2230	957906		724 924 489 614,33	2 837 793 723 740,58	ESS
7	Хворостянский м.р.	2240	613808				
8	Клявлинский м.р.	2275	1732117		ЛИНЕЙН 1 подвыборка		
9	Шенталинский м.р.	2297	143150				
10	Челно-Вершинский м.р.	2521	562301				
11	Богатовский м.р.	2767	905057				
12	Большеглушицкий м.р.	2791	1587301				
13	Большечерниговский м.р.	2795	631314				
14	Красноармейский м.р.	2831	457717				
15	Пестравский м.р.	2908	1001766				
16	Шигонский м.р.	2977	70668				
17	Борский м.р.	3006	300426				
18	Сызранский м.р.	3327	1079027				
19	Приволжский м.р.	3439	390401				
20	Кошкинский м.р.	3933	1825941				
21	Кинельский м.р.	4264	3477251				
22	г.о. Октябрьск	5300	175667				
23	г.о. Похвистнево	6472	3201227				
24	Кинель-Черкасский м.р.	7183	1777476		ЛИНЕЙН 2 подвыборка		
25	Нефтегорский м.р.	7409	12464092				
26	Безенчукский м.р.	8082	6540234		197,4821537	3075469,275	
27	Красноярский м.р.	8728	1551759		158,6268714	3035680,534	
28	Сергеевский м.р.	9667	6589916		0,146910916	5447079,21	
29	Ставропольский м.р.	9676	2487599		1,549894698	9	
30	г.о. Кинель	11300	644826		45 986 417 090 475,00	267 036 047 313 291,00	ESS
31	г.о. Жигулевск	11790	3497466				

Рис. 8.8. Оценивание регрессий крайних подвыборок в рамках теста Голдфелда - Квандта

Шаг 3. Следует проверить статистическую гипотезу вида $H_0 : D(\varepsilon)_1 = D(\varepsilon)_2$ (гомоскедастичность) против $H_1 : D(\varepsilon)_2 > D(\varepsilon)_1$, так как в результате графического анализа остатков сделано предположение о прямо пропорциональной гетероскедастичности ($\sigma_\varepsilon \approx \lambda x_i$).

В качестве критерия согласия используется случайная величина F , имеющая распределение Фишера - Снедекора. Наблюдаемое значение вычисляется по формуле

$$F_{\text{набл}} = \frac{ESS_2}{ESS_1}. \text{ Критическая точка распределения вычисляется с использованием функции}$$

ФРАСПОБ исходя из выбранного уровня значимости и числа степеней свободы (рис. 8.9).

8. Гетероскедастичность случайной компоненты эконометрической модели

Мун. образование	X	Y	F набл.	F кр.
Елховский м.р.	1210	212229	524,0654243	-405991,0208
Камышлинский м.р.	1510	1064219	345,6274914	756965,7907
Алексеевский м.р.	1546	12878	0,203475113	561525,0785
Исаклинский м.р.	2094	48713	2,299081978	9
Похвистневский м.р.	2230	957906	724 924 489 614,33	2 837 793 723 740,58
Хворостянский м.р.	2240	613808		
Клявлинский м.р.	2275	1732117	ЛИНЕЙН 1 подвыборка	
Шенталинский м.р.	2297	143150		
Челно-Вершинский м.р.	2521	562301		
Богатовский м.р.	2767	905057		
Большеглушицкий м.р.	2791	1587301		
Большечерниговский м.р.	2795	631314	F набл.	94,09987945
Красноармейский м.р.	2831	457717	F кр.	=ФРАСПОБР(0,05;F5;F29)
Пестравский м.р.	2908	1001766		ФРАСПОБР(вероятность; степени_свободы1; степени_свободы2)
Шигонский м.р.	2977	70668		
Борский м.р.	3006	300426		
Сызранский м.р.	3327	1079027		
Приволжский м.р.	3439	390401		
Кошкинский м.р.	3933	1825941		
Кинельский м.р.	4264	3477251		
г.о. Октябрьск	5300	175667		
г.о. Похвистнево	6472	3201227		
Кинель-Черкасский м.р.	7183	1777476	ЛИНЕЙН 2 подвыборка	
Нефтегорский м.р.	7409	12464092		
Безенчукский м.р.	8082	6540234	197,4821537	3075469,275
Красноярский м.р.	8728	1551759	158,6268714	3035680,534
Сергеевский м.р.	9667	6589916	0,146910916	5447079,21
Ставропольский м.р.	9676	2487599	1,549894698	9
г.о. Кинель	11300	644826	45 986 417 090 475,00	267 036 047 313 291,00
г.о. Жигулевск	11790	3497466		

Рис. 8.9. Расчет наблюдаемого значения критерия согласия в тесте Голдфелда - Квандта

Так как $F_{набл} > F_{кр} (\alpha = 0,05, k_1 = k_2 = 9) = 3,18$, принимается гипотеза о наличии гетероскедастичности остатков регрессии.

Нарушение предпосылки МНК вида требует $D(\varepsilon_i) = \sigma_\varepsilon^2, i = \overline{1, n}$ коррекции с применением взвешенного МНК.

8.2. Задача для самостоятельного выполнения

Задача 18. Используя исходные статистические данные задачи 7, необходимо провести графический анализ остатков регрессии, выдвинуть гипотезу о наличии или отсутствии гетероскедастичности остатков и подтвердить свое предположение результатами тестов Спирмена и Голдфелда - Квандта.

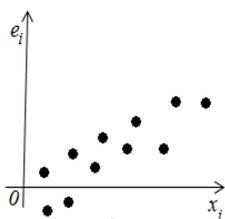
ЗАМЕЧАНИЕ: используйте соответствующий вариант задания!!!

8.3. Тесты

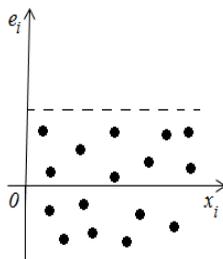
1. Свойство постоянства дисперсии остатков называется:

- А) мультиколлинеарностью;
- Б) гомоскедастичностью;
- В) гетероскедастичностью;
- Г) автокорреляцией.

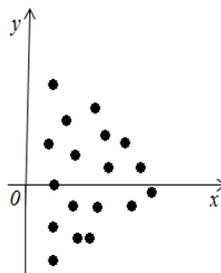
2. Расположение точек $(x_i; e_i)$, типичное для модели с гомоскедастичными остатками, находится на следующих рисунках (выберите несколько правильных ответов):



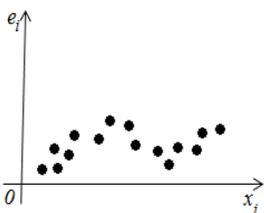
А)



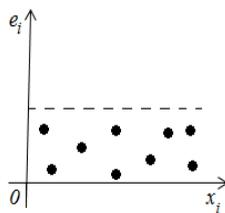
Б)



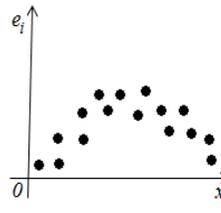
В)



Г)



Д)



Е)

3. Верным утверждением является следующее:

- А) статистические выводы на основе критерия Фишера при гетероскедастичности являются надежными;
- Б) наличие гетероскедастичности невозможно выявить, пользуясь критерием Дарбина - Уотсона;
- В) проблема гетероскедастичности не характерна для перекрестных данных;
- Г) остатки характеризуются постоянной дисперсией в случае гетероскедастичности.

4. Если предполагается, что дисперсия случайной компоненты может либо увеличиваться, либо уменьшаться с ростом X , то для обнаружения гетероскедастичности остатков рекомендуется провести тест:

- А) ранговой корреляции Спирмена;

	1	2	3	4	5	6	7	8	9	10	11	12	13
1	8. Гетероскедастичность случайной компоненты эконометрической модели												
2													
3													

- Б) Уайта;
- В) Чоу;
- Г) Брэуша - Пагана.

5. При проведении теста Голдфелда - Квандта используется критерий:

- А) Фишера - Снедекора;
- Б) Стьюдента;
- В) Дарбина - Уотсона;
- Г) Пирсона.

6. При проведении теста ранговой корреляции Спирмена используется критерий:

- А) Фишера - Снедекора;
- Б) Стьюдента;
- В) Дарбина - Уотсона;
- Г) Пирсона.

7. Обобщенный метод наименьших квадратов применяется, когда случайные отклонения такие, что:

- А) не имеют постоянной дисперсии и не коррелированы между собой;
- Б) не имеют постоянной дисперсии и коррелированы между собой;
- В) имеют постоянную дисперсию и коррелированы между собой;
- Г) имеют постоянную дисперсию и не коррелированы между собой.

8. Оценки параметров модели с гетероскедастичными остатками:

- А) оставаясь несмещенными, перестают быть состоятельными и эффективными;
- Б) остаются несмещенными, состоятельными и эффективными;
- В) оставаясь состоятельными и эффективными, перестают быть несмещенными;
- Г) оставаясь несмещенными и состоятельными, перестают быть эффективными.

9. Методом устранения гетероскедастичности в модели является (выберите несколько правильных ответов):

- А) применение теста Глейзера;
- Б) применение МНК;
- В) вычисление робастных стандартных ошибок;
- Г) изменение исходных статистических данных;
- Д) изменение спецификации уравнения регрессии;
- Е) применение обобщенного метода наименьших квадратов (ОМНК).

10. Коэффициент ранговой корреляции Спирмена рассчитывается по следующей формуле:

	1	2	3	4	5	6	7	8	9	10	11	12	13
1			<i>Эконометрика в табличном редакторе MS Excel</i>										
2													
3													

$$A) r_{xe} = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n-1)};$$

$$Б) r_{xe} = 1 - \frac{\sum_{i=1}^n d_i^2}{n(n^2-1)};$$

$$B) r_{xe} = 1 - \frac{6 \sum_{i=1}^n d_i}{n(n^2-1)};$$

$$Г) r_{xe} = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n^3-1}.$$

8.4. Контрольные вопросы

1. С нарушением какой из предпосылок Гаусса - Маркова связано наличие гетероскедастичности?
2. Назовите основные причины возникновения гетероскедастичности.
3. Назовите последствия гетероскедастичности.
4. Как проявляется гетероскедастичность на графике остатков?
5. Какие известны тесты для выявления гетероскедастичности?
6. Какой статистический критерий используется в тесте ранговой корреляции Спирмена? Приведите алгоритм теста ранговой корреляции Спирмена.
7. Какими свойствами обладает коэффициент ранговой корреляции Спирмена?
8. В каких случаях целесообразно использовать тест Голдфелда - Квандта? Приведите алгоритм теста Голдфелда - Квандта.
9. Какие методы устранения гетероскедастичности вы знаете?
10. В чем суть обобщенного метода наименьших квадратов (ОМНК)?

9. ВРЕМЕННЫЕ РЯДЫ: ЛИНЕЙНЫЙ ТРЕНД



Продемонстрирована технология оценивания линейного тренда с применением инструментов анализа данных табличного редактора MS Excel.

9.1. Решение типовых задач

В качестве иллюстрационного примера особенностей построения и анализа линейного тренда используем официальные статистические данные Российского статистического ежегодника (раздел «Культура, отдых и туризм»).

Таблица 9.1

Число общедоступных библиотек в РФ

Год	Число общедоступных библиотек, тыс. ед.	Год	Число общедоступных библиотек, тыс. ед.
1991	59,1	2005	49,5
1992	57,2	2006	48,3
1993	56,9	2007	47,5
1994	54,8	2008	47,0
1995	54,4	2009	46,7
1996	53,5	2010	46,1
1997	52,9	2011	43,2
1998	52,2	2012	40,8
1999	51,4	2013	39,8
2000	51,2	2014	40,1
2001	51,2	2015	38,9
2002	51,0	2016	38,2
2003	50,6	2017	37,4
2004	49,9	-	-

9.1.1. Расчет абсолютных, относительных и средних показателей, характеризующих динамику исследуемого показателя (уровней временного ряда)

Одними из показателей, характеризующих динамические изменения, являются абсолютный прирост, темп роста, темп прироста, абсолютное ускорение.

Шаг 1. Необходимо перенести данные табл. 9.1 на лист табличного редактора Excel в следующем порядке: первый столбец - год наблюдения (t), второй столбец - значения показателя «число доступных библиотек» (y_t).

Шаг 2. Для расчета цепных и базисных показателей применены формулы (табл. 9.2).

Таблица 9.2

Расчет абсолютных и относительных показателей ряда динамики (временного ряда)

Показатель	Формула	
	Базисный метод расчета	Цепной метод расчета
Абсолютный прирост	$\Delta_b = y_t - y_0$	$\Delta_c = y_t - y_{t-1}$
Темп роста	$T_p^b = \frac{y_t}{y_0}$	$T_p^c = \frac{y_t}{y_{t-1}}$
Темп прироста	$T_{пр}^b = T_p^b - 100$	$T_{пр}^c = T_p^c - 100$
Абсолютное ускорение	$A_{1\%} = \frac{\Delta_c}{T_{пр}^c}$ или $A_{1\%} = 0,01y_{t-1}$	

Непосредственный расчет выше указанных показателей приведен на рис. 9.1. Необходимо помнить, что при расчете базисных показателей ссылка на начальный уровень должна быть абсолютной (ячейке придается универсальный адрес путем нажатия F4).

	A	B	C		D		E		F		G		H		I	J	
1	Год	y	Δ, тыс. ед.		Тр. коэффициент		Тп, %		Абсолютное ускорение, тыс. ед.		А%						
2			цепной	базисный	цепной	базисный	цепной	базисный	цепной	базисный	цепной	базисный	цепной	базисный			
3	1991	59,1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
4	1992	57,2	-1,9	-1,9	0,97	0,97	-3,21	-3,21	-	-	-	-	-	-	-	0,591	
5	1993	56,9	-0,3	-2,2	0,99	0,96	-0,52	-3,72	1,6	0,572							
6	1994	54,8	-2,1	=B6-\$B\$3	0,96	0,93	-3,69	-7,28	-1,8	0,569							
7	1995	54,4	-0,4	-4,7	0,99	0,92	-0,73	-7,95	1,7	0,548							
8	1996	53,5	-0,9	-5,6	0,98	0,91	-1,65	-9,48	-0,5	0,544							
9	1997	52,9	-0,6	-6,2	0,99	0,90	-1,12	-10,49	0,3	0,535							

Рис. 9.1. Абсолютные и относительные показатели временного ряда показателя «число общедоступных библиотек в РФ»

9. Временные ряды: линейный тренд

Шаг 3. Вычислим средние значения рассчитанных показателей (рис. 9.2). При вычислениях исходим из того, что в имеющемся временном ряду равные интервалы времени между датами.

	A	B	C		D		E		F		G		H		I		J
1	Год	y	Δ, тыс. ед.		Тр, коэффициент		Тп, %				Абсолютное ускорение, тыс. ед.		А%				
2			цепной	базисный	цепной	базисный	цепной	базисный	цепной	базисный							
22	2010	46,1	-0,6	-13,0	0,99	0,78	-1,28	-22,00			-0,3		0,467				
23	2011	43,2	-2,9	-15,9	0,94	0,73	-6,29	-26,90			-2,3		0,461				
24	2012	40,8	-2,4	-18,3	0,94	0,69	-5,56	-30,96			0,5		0,432				
25	2013	39,8	-1,0	-19,3	0,98	0,67	-2,45	-32,66			1,4		0,408				
26	2014	40,1	0,3	-19,0	1,01	0,68	0,75	-32,15			1,3		0,398				
27	2015	38,9	-1,2	-20,2	0,97	0,66	-2,99	-34,18			-1,5		0,401				
28	2016	38,2	-0,7	-20,9	0,98	0,65	-1,80	-35,36			0,5		0,389				
29	2017	37,4	-0,8	-21,7	0,98	0,63	-2,09	-36,72			-0,1		0,382				
30	Средняя	48,51	-0,83	x	0,98	x	-1,74	x			x		x				

Рис. 9.2. Расчет средних значений для показателей временного ряда «число общедоступных библиотек в РФ»

По итогам проведенных вычислений можно сделать вывод, что количество общедоступных библиотек в РФ за период 1991-2017 гг. снижалось в среднем за год на 0,83 тыс. ед. при ежегодном замедлении в 1,74%.

9.1.2. Построение линейного тренда с использованием графических инструментов табличного редактора MS Excel

Динамику изменения показателя во времени можно представить графически.

Шаг 1. Представим исходные данные табл. 9.1 в графическом виде. В главном меню выберем **Вставка / График / График с маркерами**. Полученная линейная диаграмма представлена на рис. 9.3.

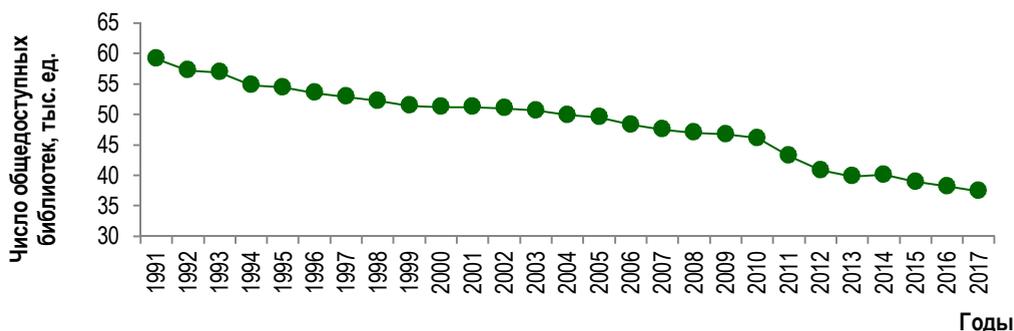


Рис. 9.3. Динамика числа общедоступных библиотек в РФ

На рис. 9.3 наблюдается непрерывное снижение значения показателя на всем протяжении рассматриваемого периода.

Шаг 2. Для отражения убывающего тренда на графике необходимо выделить уровни временного ряда. Нажатием правой кнопки мыши следует вызвать дополнительное меню настройки графика. В появившемся окне выбираем пункт **«Добавить линию тренда»**.

В окне **«Формат линии тренда»** (рис. 9.4) доступны к выбору большое количество функциональных видов трендов. Исходя из расположения точек на графике (см. рис. 9.3) следует подобрать оптимальный тренд, т.е. тот, который позволит наилучшим образом аппроксимировать тенденцию. В данном случае, это линейный тренд $y_t = b_0 + b_1t + \varepsilon$. Активируем соответствующий способ аппроксимации: **«Линейная»**. Также необходимо включить флажки в полях **«Показывать уравнение на диаграмме»** и **«Поместить на диаграмму величину достоверности аппроксимации (R²)»**. Тем самым на график будут добавлены линия тренда, уравнение регрессии и коэффициент детерминации (аппроксимации).

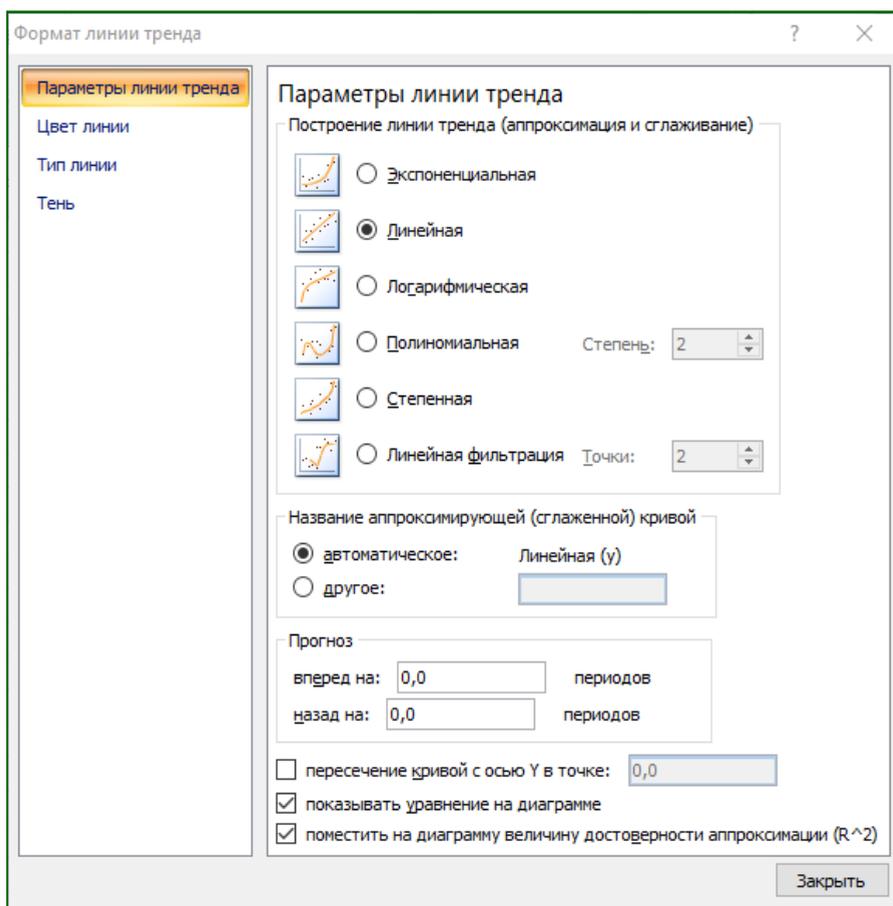


Рис. 9.4. Выбор вида тренда и прочие установки графика

9. Временные ряды: линейный тренд

После нажатия кнопки «Заккрыть» на экране появится график, представленный на рис. 9.5.

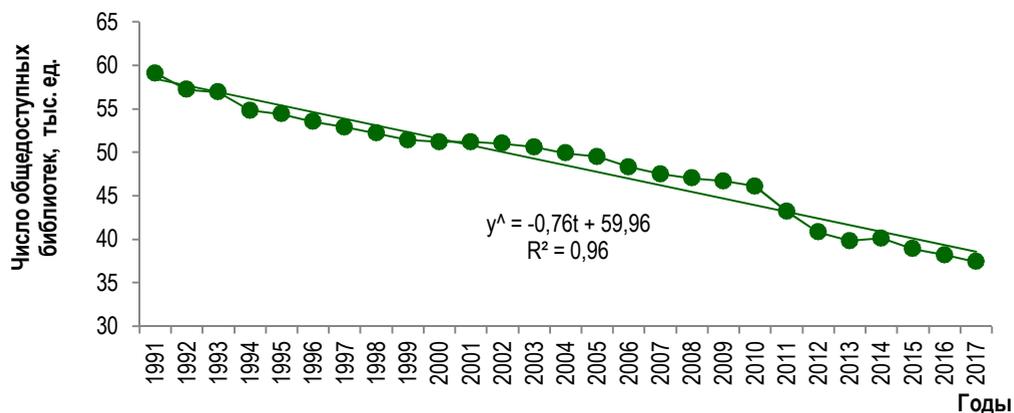


Рис. 9.5. Динамика числа общедоступных библиотек в РФ с линейной аппроксимацией уровней временного ряда

ЗАМЕЧАНИЕ: при построении модели временного ряда $y_t = b_0 + b_1t + \varepsilon$ в качестве объясняющей (независимой) переменной используется время (t).

Коэффициент детерминации R^2 равен 0,96. Это свидетельствует в пользу того, что линейный тренд на 96% описывает динамику изменения числа общедоступных библиотек в РФ за период 1991-2017 гг.

9.1.3. Аналитическое выравнивание на основе линейной функции

Наиболее простым в реализации способом построения тренда является использование пакета «АНАЛИЗ ДАННЫХ». Проиллюстрируем этапы построения линейного тренда с применением данного инструмента.

Шаг 1. Добавим к имеющимся данным дискретную переменную (момент или период времени) t (рис. 9.6). Переменная t принимает значения от 1 до 27 (период с 1991 по 2017 г. включает в себя 27 лет). Переменная t - независимая в уравнении регрессии вида $y_t = b_0 + b_1t + \varepsilon$.

	A	B	C
1	Год	y	t
2	1991	59,1	1
3	1992	57,2	2
4	1993	56,9	3
5	1994	54,8	4
6	1995	54,4	5
7	1996	53,5	6
8	1997	52,9	7
9	1998	52,2	8

Рис. 9.6. Начальный вид листа табличного редактора Excel при аналитическом выравнивании на основе уравнения прямой

Шаг 2. В главном меню следует выбрать **Данные / Анализ данных**, затем активировать инструмент анализа **РЕГРЕССИЯ** (рис. 9.7).

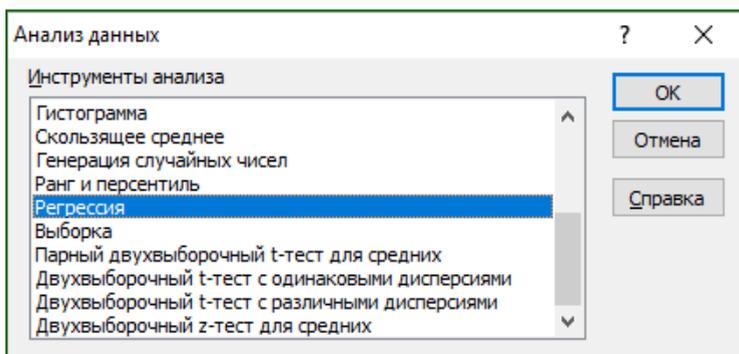


Рис. 9.7. Выбор инструмента анализа данных «РЕГРЕССИЯ»

Шаг 3. В поле **«Входной интервал Y»** следует указывать значения уровней временного ряда (число общедоступных библиотек, y), в поле **«Входной интервал X»** - объясняющую переменную t (рис. 9.8).

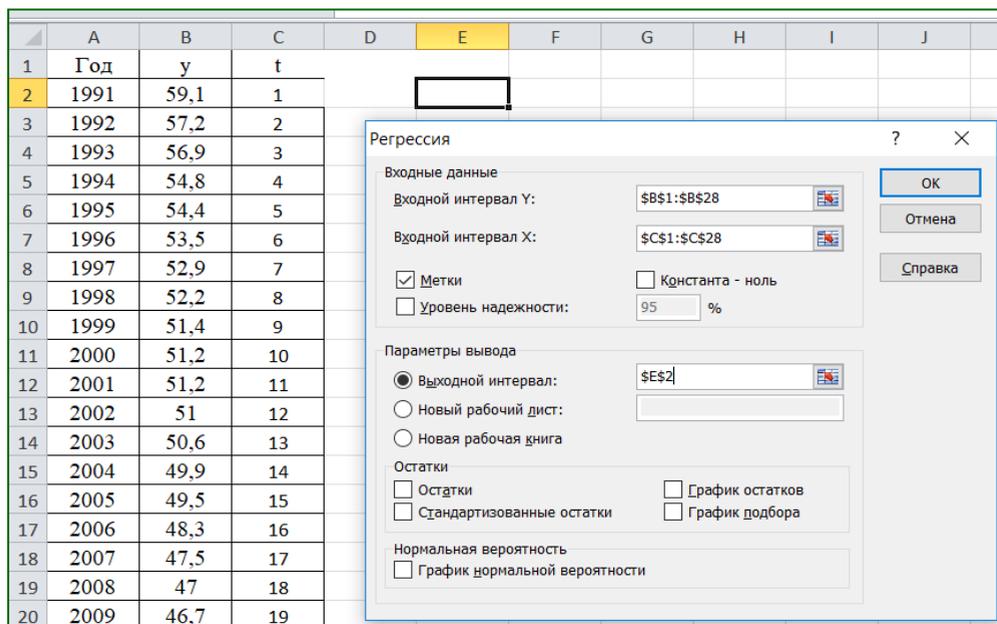


Рис. 9.8. Установки оценки параметров регрессионного уравнения

$$y_t = b_0 + b_1 t + \varepsilon$$

9. Временные ряды: линейный тренд

Результаты оценки параметров представлены на рис. 9.9.

	E	F	G	H	I	J
Вывод итогов						
<i>Регрессионная статистика</i>						
Множественный R		0,979452098				
R-квадрат		0,959326413				
Нормированный R-квадрат		0,95769947				
Стандартная ошибка		1,272417925				
Наблюдения		27				
<i>Дисперсионный анализ</i>						
		<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Значимость F</i>
Регрессия		1	954,6704823	954,6704823	589,6495045	6,65842E-19
Остаток		25	40,47618437	1,619047375		
Итого		26	995,1466667			
		<i>Коэффициенты</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>	<i>P-Значение</i>	<i>Нижние 95%</i>
Y-пересечение		59,1991453	0,503683315	117,5324723	8,11857E-36	58,16179009
t		-0,763431013	0,031439297	-24,28269969	6,65842E-19	-0,828181458

Рис. 9.9. Результаты оценки параметров тренда $y_t = b_0 + b_1t + \varepsilon$

Анализ итогов работы инструмента **РЕГРЕССИЯ** позволил сделать следующий вывод: между рассматриваемыми переменными существует сильная корреляционная связь, так как множественный коэффициент корреляции равен 0,979.

Выборочное уравнение линейной регрессии имеет вид $\hat{y} = 59,2 - 0,76t$.

Интерпретация параметра модели при дискретной переменной (\hat{b}_1) заключается в следующем: ежегодно число общественных библиотек в России снижается на 0,76 тыс. ед.

Оценки параметров модели линейного тренда \hat{b}_0 , \hat{b}_1 являются статистически значимыми, так как **p-значение** много меньше принятого в данном случае уровня значимости $\alpha = 0,05$.

Полученная регрессия также статистически значима по *F*-критерию Фишера ($F_{\text{набл}} > F_{\text{кр}}(\alpha = 0,05; k_1 = 1; k_2 = 25) = 4,24$).

Шаг 4. Для вычисления расчетных значений уровней временного ряда по уравнению вида $\hat{y} = 59,2 - 0,76t$ следует задать соответствующую формулу в столбце \hat{y} (рис. 9.10).

Эконометрика в табличном редакторе MS Excel

1	A	B	C	D	E	F	G
2	1991	59,1	1	= $\$G\$18+\$G\$19*C2$	Вывод ИТОГОВ		
3	1992	57,2	2	57,67228	<i>Регрессионная статистика</i>		
4	1993	56,9	3	56,90885	Множественный R	0,979452098	
5	1994	54,8	4	56,14542	R-квадрат	0,959326413	
6	1995	54,4	5	55,38199	Нормированный R-квадрат	0,95769947	
7	1996	53,5	6	54,61856	Стандартная ошибка	1,272417925	
8	1997	52,9	7	53,85513	Наблюдения	27	
9	1998	52,2	8	53,0917	<i>Дисперсионный анализ</i>		
10	1999	51,4	9	52,32827	<i>df</i>		
11	2000	51,2	10	51,56484	Регрессия	1	
12	2001	51,2	11	50,8014	Остаток	25	
13	2002	51	12	50,03797	Итого	26	
14	2003	50,6	13	49,27454	<i>Коэффициенты</i>		
15	2004	49,9	14	48,51111	Y-пересечение	59,1991453	
16	2005	49,5	15	47,74768	t	-0,763431013	
17	2006	48,3	16	46,98425			
18	2007	47,5	17	46,22082			
19	2008	47	18	45,45739			

Рис. 9.10. Вычисление расчетных значений числа библиотек в России по уравнению линейного тренда $\hat{y} = 59,2 - 0,76t$

Шаг 5. По полученной модели линейного тренда можно выполнить прогнозирование числа общедоступных библиотек в РФ в 2018-2020 гг. Для этого необходимо в столбец дискретной переменной t внести номер соответствующего периода: 28 - для 2018 г., 29 - для 2019 г., 30 - для 2020 г.

Автоматически выполнить прогнозирование позволяет дополнительное меню настройки графика. В окне «Добавить линию тренда» (см. рис. 9.4) активируем поле «Прогноз» и укажем число периодов прогнозирования (3). В результате получим график, представленный на рис. 9.11.

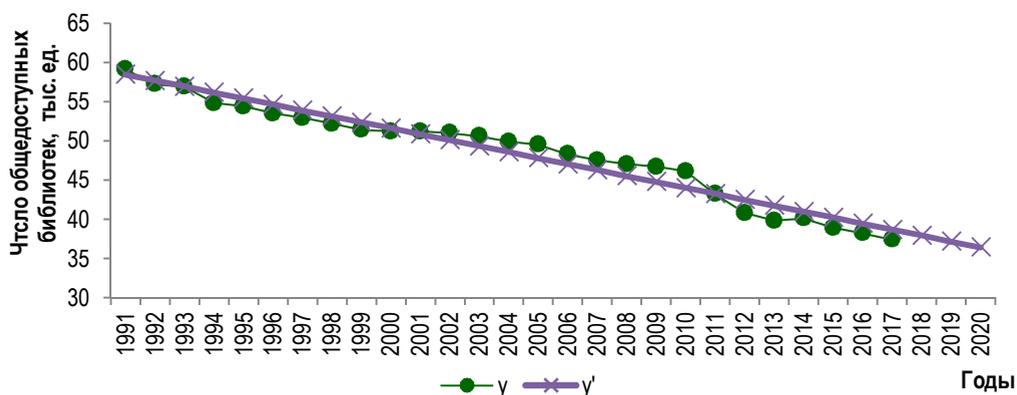


Рис. 9.11. Прогноз числа общедоступных библиотек в России

9. Временные ряды: линейный тренд

Согласно данным, представленным на рис. 9.11, прогнозные значения анализируемого показателя будут снижаться каждый год периода 2018-2020 гг.

Шаг 6. Для расчета средней ошибки аппроксимации используется формула

$$A = \frac{1}{n} \sum_{t=1}^n \left| \frac{y_t - \hat{y}_t}{y_t} \right| \cdot 100\% .$$

Для расчета показателя в исходной таблице добавляем еще один столбец. Для вычисления модуля числа необходимо использовать встроенную функцию **ABS(число)** (рис. 9.12).

ЛИНЕЙН						=ABS((B23-D23)/B23)	
	A	B	C	D	E	F	
1	Год	y	t	y^	$\left \frac{y_t - \hat{y}_t}{y_t} \right $		
21	2010	46,1	20	43,93053	0,04706		
22	2011	43,2	21	43,16709	0,000762		
23	2012	40,8	22	42,40366	=ABS((B23-D23)/B23)		
24	2013	39,8	23	41,64023	ABS(число)		
25	2014	40,1	24	40,8768	0,019372		
26	2015	38,9	25	40,11337	0,031192		
27	2016	38,2	26	39,34994	0,030103		
28	2017	37,4	27	38,58651	0,031725		
29					0,635268		
30					Сумма		
31							

Рис. 9.12. Расчет средней ошибки аппроксимации модели линейного тренда $\hat{y} = 59,2 - 0,76t$

Значение ошибки аппроксимации модели линейного тренда составило:

$A = \frac{1}{27} \cdot 0,6353 \cdot 100 = 2,35$. Следовательно, среднее отклонение расчетных значений уровней временного ряда от эмпирических составляет 2,35%. Объясняющая способность модели приемлема для такого рода исследований.

9.2. Задача для самостоятельного выполнения

Задача 19. Проведите графическую визуализацию уровней временного ряда показателя, представленного в таблице (по вариантам):

Год	Варианты				
	1	2	3	4	5
	Урожайность зерновых и зернобобовых культур в РФ, ц/га	Урожайность картофеля в РФ, ц/га	Перевозки пассажиров автобусным транспортом в РФ, млн чел.	Число легковых автомобилей (на конец года) в собственности граждан в РФ, тыс. шт.	Число троллейбусов (на конец года) в РФ, тыс. шт.
1991	15,1	109	27302,0	9417,3	14,0
1992	18,0	114	24874,2	10156,7	13,9
1993	17,1	109	24123,7	11099,2	13,8
1994	15,3	103	23438,0	12387,3	13,5
1995	13,1	118	22817,1	13688,5	13,2
1996	14,9	114	23185,0	15047,2	12,7
1997	17,8	111	23664,7	16591,2	12,5
1998	12,9	97	23102,8	17761,3	12,3
1999	14,4	97	22882,5	18543,4	12,2
2000	15,6	105	23001,1	19097,4	12,2
2001	19,4	108	22462,2	19983,9	12,1
2002	19,6	103	21920,6	21135,1	11,9
2003	17,8	117	20912,3	22082,1	11,8
2004	18,8	116	21018,3	22853,8	11,6
2005	18,5	124	16374,0	24124,8	11,4
2006	18,9	133	14733,6	25281,8	11,3
2007	19,8	132	14794,9	27754,5	11,2
2008	23,8	138	14717,8	30300,2	11,2
2009	22,7	143	13704,4	31340,6	11,0
2010	18,3	100	13433,7	32629,1	11,1
2011	22,4	148	13304,9	34624,3	11,0
2012	18,3	134	12766,2	36917,2	11,0
2013	22,0	145	11586,8	39237,1	10,7

Окончание таблицы

Год	Варианты				
	1	2	3	4	5
	Урожайность зерновых и зернобобовых культур в РФ, ц/га	Урожайность картофеля в РФ, ц/га	Перевозки пассажиров автобусным транспортом в РФ, млн чел.	Число легковых автомобилей (на конец года) в собственности граждан в РФ, тыс. шт.	Число троллейбусов (на конец года) в РФ, тыс. шт.
2014	24,1	150	11554,3	41433,4	10,7
2015	23,7	159	11522,9	42317,1	10,2
2016	26,2	158	11296,0	45163,0	10,0
2017	29,2	163	11184,0	46887,0	9,4

Источник: URL: www.gks.ru.

Рассчитайте основные показатели динамики. Сделайте соответствующие выводы. Оцените параметры модели линейного тренда по методу наименьших квадратов. Проверьте статистическую значимость оценок параметров модели, модели в целом ($\alpha = 0,05$). Выполните прогнозирование значений уровней временного ряда на период 2018-2020 гг. (при наличии такой возможности). Какова ошибка аппроксимации полученной модели линейного тренда?

9.3. Тесты

1. Выражение $\left(\frac{y_t - y_{t-1}}{y_{t-1}} - 1 \right) \cdot 100\%$ представляет собой:

- А) темп роста;
- Б) темп прироста (снижения);
- В) абсолютное значение одного процента прироста (снижения).

2. Формула $\frac{0,5x_1 + \sum_{i=2}^{n-1} x_i + 0,5x_n}{n-2}$ является представлением:

- А) средней квадратической;
- Б) средней хронологической;
- В) медианой.

	1	2	3	4	5	6	7	8	9	10	11	12	13
1			<i>Эконометрика в табличном редакторе MS Excel</i>										
2													
3													

3. Коэффициент b_1 в уравнении линейной регрессии показывает:

- А) наклон прямой;
- Б) сдвиг по оси ординат;
- В) среднее значение y .

4. Коэффициент b_0 в уравнении линейной регрессии показывает:

- А) наклон прямой;
- Б) сдвиг по оси ординат;
- В) среднее значение y .

5. Модели временных рядов - это:

- А) модели, построенные по данным, характеризующим один объект за ряд последовательных моментов времени;
- Б) модели, построенные по данным, характеризующим совокупность различных объектов в определенный момент времени.

6. Ряд динамики состоит:

- А) из частот;
- Б) частостей;
- В) уровней;
- Г) вариантов.

7. Модель $Y = T + S + V + \varepsilon$ - это:

- А) аддитивная форма представления исследуемой зависимости;
- Б) мультипликативная форма представления исследуемой зависимости;
- В) кратная форма представления исследуемой зависимости.

8. Если уровни временного ряда изменяются в арифметической прогрессии, т.е. когда первые разности уровней (абсолютные приросты) относительно постоянны, то для описания лучшим образом подойдет:

- А) линейная функция;
- Б) парабола второго порядка;
- В) гипербола.

9.4. Контрольные вопросы

1. Что такое временной ряд?
2. Какие классы эконометрических моделей используются для исследования временного ряда?
3. Назовите составляющие временного ряда.
4. Перечислите основные этапы анализа временного ряда.

	1	2	3	4	5	6	7	8	9	10	11	12	13	
1														
2				9. Временные ряды: линейный тренд										
3														

5. Перечислите основные типы трендов.
6. Какие инструменты Excel используются для подбора тренда временного ряда?
7. Можно ли использовать метод наименьших квадратов (МНК) для оценки параметров линейного тренда?
8. Сформулируйте экономический смысл параметра b_0 в модели $y_t = b_0 + b_1t + \varepsilon_t$.
9. Для чего используется средняя ошибка аппроксимации?
10. Как осуществляется проверка статистической значимости оценок параметров модели $y_t = b_0 + b_1t + \varepsilon_t$? Как проверить статистическую значимость данной модели?

10. ВРЕМЕННЫЕ РЯДЫ: ПАРАБОЛИЧЕСКИЙ ТРЕНД ВТОРОГО ПОРЯДКА



Продемонстрирована технология оценивания параболического тренда второго порядка с применением инструментов анализа данных табличного редактора MS Excel.

10.1. Решение типовых задач

В качестве иллюстрационного примера особенностей построения и анализа параболического тренда второго порядка воспользуемся официальными статистическими данными о численности населения России на начало года за период 1995-2019 гг. Данные табл. 10.1 получены из официального источника - портала «Единый архив экономических и социологических данных» (<http://sophist.hse.ru>).

Таблица 10.1

Численность населения РФ на начало года

Год	Численность населения РФ (на начало года), тыс. чел.	Год	Численность населения РФ (на начало года), тыс. чел.
1995	148306	2008	141980
1996	147976	2009	141900
1997	147502	2010	142962
1998	147105	2011	142914
1999	146693	2012	143103
2000	145925	2013	143395
2001	146304	2014	143700
2002	145649	2015	146267
2003	144964	2016	146545
2004	144168	2017	146804
2005	143474	2018	146880
2006	142754	2019	146781
2007	142220	-	-

10. Временные ряды: параболический тренд второго порядка

10.1.1. Построение параболического тренда второго порядка с использованием графических инструментов табличного редактора MS Excel

Визуально значения уровней временного ряда необходимо представить графически с использованием диаграммы.

Шаг 1. В главном меню выберем **Вставка / График / График с маркерами**. Результат построения графика представлен на рис. 10.1.

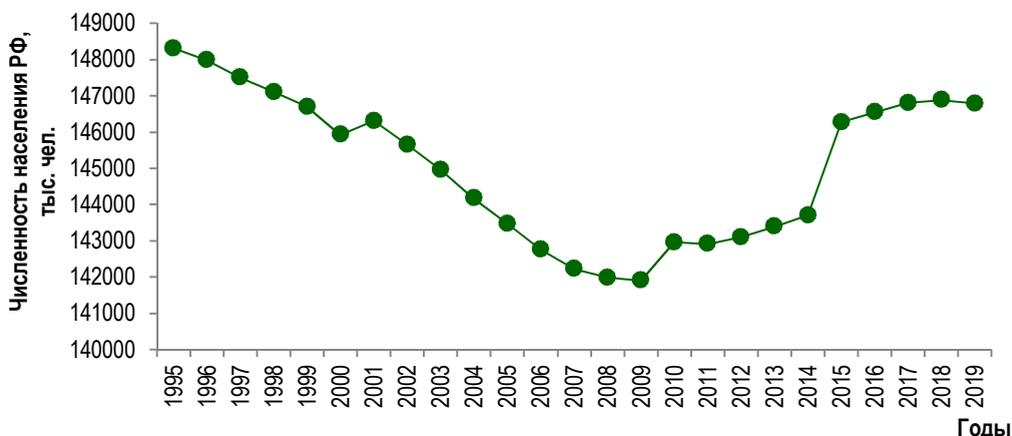


Рис. 10.1. Динамика численности населения России (на начало года)

Визуально на рисунке можно наблюдать близость уровней временного ряда параболической кривой второго порядка вида $y_t = b_0 + b_1t + b_2t^2 + \varepsilon$.

Шаг 2. Выделим правой кнопкой мыши любую точку, соответствующую уровню временного ряда, активируя дополнительное меню (рис. 10.2). В появившемся окне выберем «Добавить линию тренда».

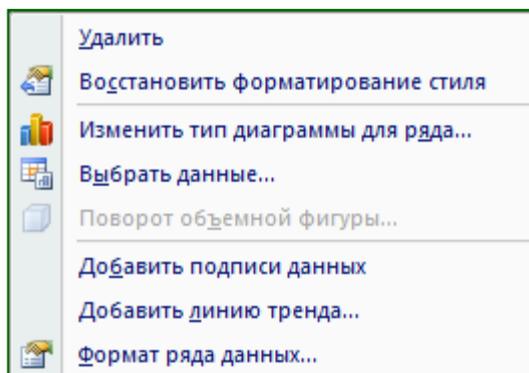


Рис. 10.2. Дополнительное меню настройки графиков

Шаг 3. Для нанесения кривой на график необходимо в окне формата линии тренда (рис. 10.3) выбрать тип аппроксимации **«Полиномиальная»**, указать степень, равную 2, установить флажок напротив полей **«показывать уравнение на диаграмме»** и **«поместить на диаграмму величину достоверности аппроксимации (R^2)»**.

ЗАМЕЧАНИЕ: возможности окна «Формат линии тренда» включают в себя осуществление перспективного и ретроспективного прогнозов. Установим в поле **«Прогноз вперед на:»** значение 3. На графике будет выполнено графическое отображение прогнозной численности населения РФ в 2020-2022 гг. (на начало соответствующего года).

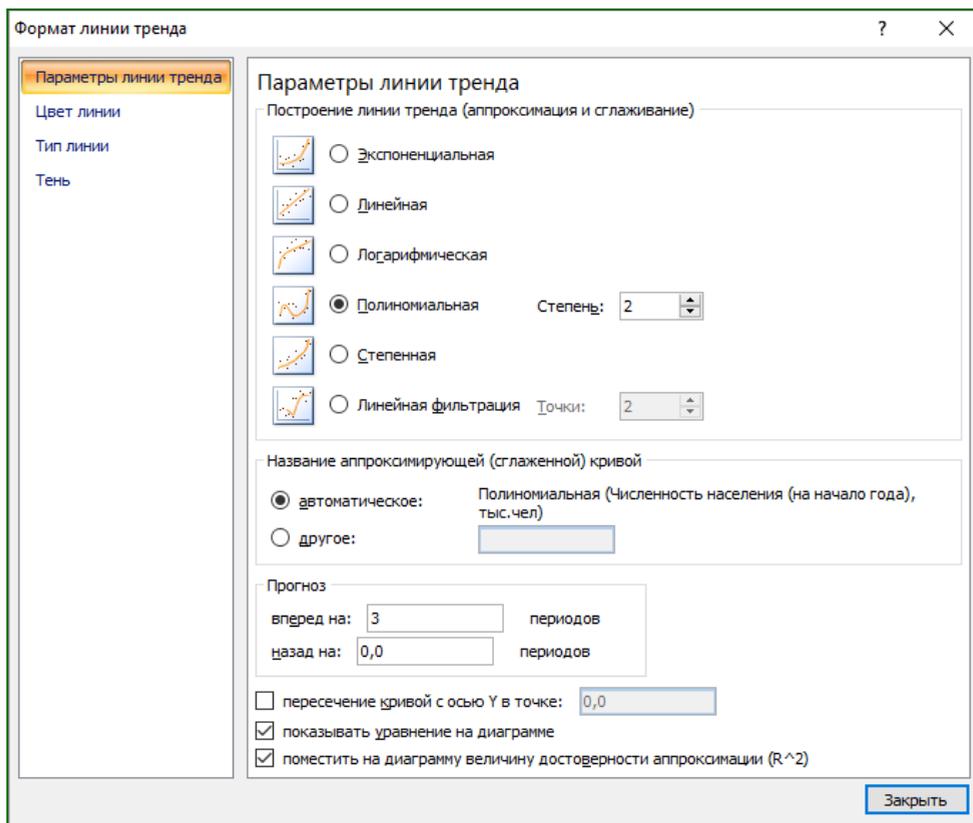


Рис. 10.3. Установки для вывода на график полинома второго порядка

В результате указанных манипуляций получаем диаграмму, приведенную на рис. 10.4.

10. Временные ряды: параболический тренд второго порядка

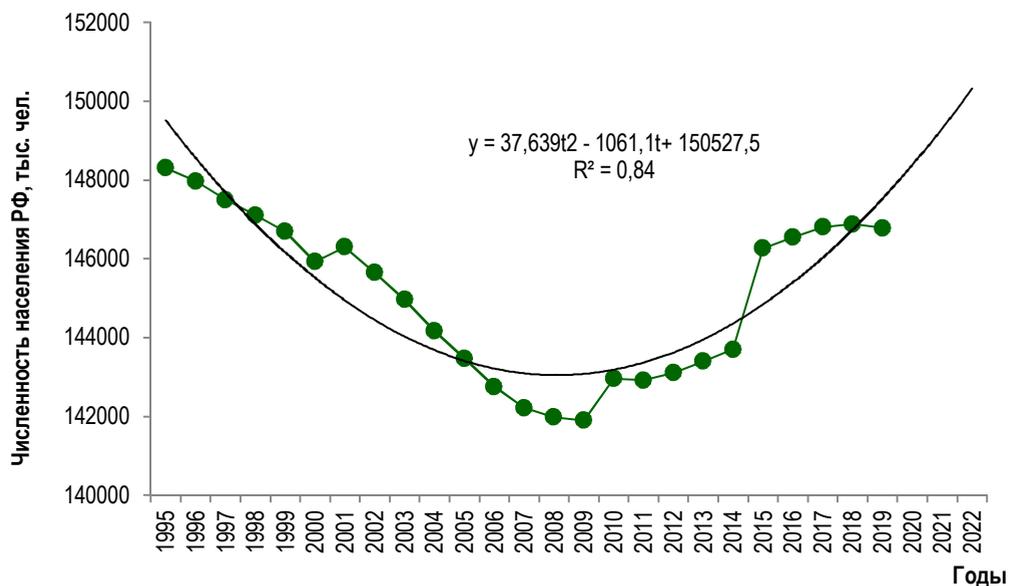


Рис. 10.4. Динамика численности населения РФ (на начало года) с выровненными уровнями временного ряда на основе параболы второго порядка

Стоит заметить, что данный способ можно использовать только в качестве инструмента разведочного анализа, так как получить числовые значения выровненных уровней и прогнозных значений не представляется возможным.

10.1.2. Аналитическое выравнивание на основе полиномиальной функции

Теоретическое уравнение множественной регрессии имеет вид $y_t = b_0 + b_1t + b_2t^2 + \varepsilon$. Данная модель линейна относительно параметров, но не линейна по переменным (полином второго порядка).

Шаг 1. Для оценки параметров модели воспользуемся макросом «Анализ данных», предварительно создав на листе табличного редактора Excel дополнительные столбцы t , t^2 для дискретных переменных (t - номер уровня временного ряда, t^2 - квадраты номеров уровней временного ряда). Указанные построения приведены на рис. 10.5.

	A	B	C	D
1	год	y	t	t^2
2	1995	148306	1	1
3	1996	147976	2	4
4	1997	147502	3	=C4^2
5	1998	147105	4	16
6	1999	146693	5	25
7	2000	145925	6	36
8	2001	146304	7	49
9	2002	145649	8	64
10	2003	144964	9	81
11	2004	144168	10	100
12	2005	143474	11	121
13	2006	142754	12	144

Рис. 10.5. Добавление дискретных переменных

Шаг 2. Для оценивания параметров модели $y_t = b_0 + b_1t + b_2t^2 + \varepsilon$ по методу наименьших квадратов воспользуемся инструментом анализа данных **РЕГРЕССИЯ**. В окне «Регрессия» в качестве «Входного интервала Y» укажем исходные значения уровней временного ряда, в качестве «Входного интервала X» - обе переменные t, t^2 (рис. 10.6).

	A	B	C	D	E	F	G	H	I
1	Вывод итогов								
2									
3	<i>Регрессионная статистика</i>								
4	Множественный R	0,92							
5	R-квадрат	0,84							
6	Нормированный R-квадрат	0,83							
7	Стандартная ошибка	852,90							
8	Наблюдения	25							
9									
10	<i>Дисперсионный анализ</i>								
11		<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>значимость F</i>			
12	Регрессия	2	85100663	42550331	58,49	1,56E-09			
13	Остаток	22	16003528	727433					
14	Итого	24	1,01E+08						
15									
16		<i>Коэффициенты</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>	<i>P-Значение</i>	<i>Нижние 95%</i>	<i>Верхние 95%</i>	<i>Нижние 95,0%</i>	<i>Верхние 95,0%</i>
17	Y-пересечение	150527,45	555,59	270,93	0,00	149375,22	151679,68	149375,22	151679,68
18	t	-1061,14	98,47	-10,78	0,00	-1265,36	-856,93	-1265,36	-856,93
19	t^2	37,64	3,68	10,24	0,00	30,01	45,26	30,01	45,26

Рис. 10.6. Результаты оценки параметров тренда $y_t = b_0 + b_1t + b_2t^2 + \varepsilon$

10. Временные ряды: параболический тренд второго порядка

Выборочное уравнение регрессии имеет вид $\hat{y}_t = 150527,45 - 1061,14t + 37,64t^2$.
 Оценки параметров модели линейного тренда $\hat{b}_0, \hat{b}_1, \hat{b}_2$ являются статистически значимыми, так как **p-значение** много меньше принятого в данном случае уровня значимости $\alpha = 0,05$.

Шаг 3. В случае нелинейной регрессии показателем тесноты связи выступает индекс детерминации:

$$\eta^2 = \sqrt{1 - \frac{\sum_{t=1}^n (y_t - \hat{y}_t)^2}{\sum_{t=1}^n (y_t - \bar{y})^2}}$$

Для расчета индекса детерминации необходимо создать два новых столбца на листе табличного редактора Excel (рис. 10.7).

	A	B	C	D	E	F	G
	Год	Y	t	t^2	Y^	$(y_t - \hat{y}_t)^2$	$(y_t - \bar{y})^2$
19	2012	143103	18	324	143622	269315,4	1128593,772
20	2013	143395	19	361	143953	311879	593443,654
21	2014	143700	20	400	144360	435923,3	216553,3599
22	2015	146267	21	441	144842	2029751	4416920,36
23	2016	146545	22	484	145400	1311834	5662720,125
24	2017	146804	23	529	146032	595574,4	6962458,301
25	2018	146880	24	576	146740	19554,65	7369308,654
26	2019	146781	25	625	147523	551064,4	6841609,536
27						10407995	20783648989
28						Сумма	Сумма
31					η	=КОРЕНЬ(1-F27/G27)	

Рис. 10.7. Вычисление индекса детерминации

Статистическая значимость нелинейной модели проверяется с помощью случайной величины F , имеющей распределение Фишера - Снедекора. Наблюдаемое значение критерия рассчитывается по формуле

$$F_{\text{набл}} = \frac{\eta^2}{1 - \eta^2} \cdot \frac{n - m}{m - 1},$$

где n - число наблюдений;

m - число параметров в регрессионном уравнении.

В данном случае $F_{набл} > F_{кр} (\alpha = 0,05; k_1 = 2; k_2 = 22) = 3,44$. Модель параболического тренда статистически значима.

Интерпретация параметров модели при дискретных переменных t, t^2 :

✓ оценка \hat{b}_1 равна -1061,14, следовательно, с каждым годом численность населения РФ сокращается на 1061,14 тыс. чел.;

✓ оценка \hat{b}_0 равна 37,64, следовательно, сокращение численности населения в РФ ускоряется, т.е. происходит рост депопуляции населения.

Шаг 4. Вычислим расчетные значения численности населения в РФ (на начало года) в период 1995-2019 гг. по полученному уравнению $\hat{y}_t = 150527,45 - 1061,14t + 37,64t^2$ (рис. 10.8).

ЛИНЕЙН					=SH\$17+SH\$18*C7+SH\$19*D7			
	A	B	C	D	E	F	G	H
	Год	Y	t	t^2	Y^			
1							ВЫВОД ИТОГОВ	
2	1995	148306	1	1	149504			
3	1996	147976	2	4	148556		<i>Регрессионная статистика</i>	
4	1997	147502	3	9	147683		Множественный R	0,917448915
5	1998	147105	4	16	146885		R-квадрат	0,841712512
6	1999	146693	5	25	146163		Нормированный R-квадрат	0,827322741
7	2000	145925			=SH\$17+SH\$18*C7+SH\$19*D7		аяя ошибка	852,8968934
8	2001	146304	7	49	144944		Наблюдения	25
9	2002	145649	8	64	144447			
10	2003	144964	9	81	144026		<i>Дисперсионный анализ</i>	
11	2004	144168	10	100	143680			<i>df</i>
12	2005	143474	11	121	143409		Регрессия	2
13	2006	142754	12	144	143214		Остаток	22
14	2007	142220	13	169	143094		Итого	24
15	2008	141980	14	196	143049			
16	2009	141900	15	225	143079			
17	2010	142962	16	256	143185		<i>Коэффициенты</i>	
18	2011	142914	17	289	143366		Y-пересечение	150527,4504
19	2012	143103	18	324	143622		t	-1061,143259
20	2012	143205	19	361	143952		t^2	37,63914902

Рис. 10.8. Вычисление расчетных значений численности населения РФ (на начало года) за период 1995-2019 гг. по уравнению параболического тренда

$$\hat{y}_t = 150527,45 - 1061,14t + 37,64t^2$$

Аналогичным образом можно получить прогнозные значения уровней временного ряда за период 2020-2022 гг., подставив в уравнение номера и квадраты номеров соответствующих лет (26 - 2020 г.; 27 - 2021 г., 28 - 2022 г.).

Шаг 5. Для расчета средней ошибки аппроксимации используется формула

10. Временные ряды: параболический тренд второго порядка

$$A = \frac{1}{n} \sum_{t=1}^n \left| \frac{y_t - \hat{y}_t}{y_t} \right| \cdot 100\% .$$

В исходной таблице следует добавить еще один столбец (рис. 10.9). Для вычисления модуля числа необходимо использовать встроенную функцию **ABS**.

ЛИНЕЙН							=ABS((B20-E20)/B20)
	A	B	C	D	E	F	G
	Год	Y	t	t^2	Y^	$\left \frac{y_t - \hat{y}_t}{y_t} \right $	
17	2010	142962	16	256	143185	0,0015583	
18	2011	142914	17	289	143366	0,0031608	
19	2012	143103	18	324	143622	0,0036265	
20	2013	143395	19	361	143953	=ABS((B20-E20)/B20)	
21	2014	143700	20	400	144360	ABS(число)	
22	2015	146267	21	441	144842	0,0097404	
23	2016	146545	22	484	145400	0,0078157	
24	2017	146804	23	529	146032	0,0052569	
25	2018	146880	24	576	146740	0,0009521	
26	2019	146781	25	625	147523	0,0050574	
27						0,11985	
28						Сумма	

Рис. 10.9. Расчет средней ошибки аппроксимации для тренда $\hat{y}_t = 150527,45 - 1061,14t + 37,64t^2$

Таким образом, средняя ошибка аппроксимации составила $A = \frac{1}{25} \cdot 0,11985 \cdot 100 = 0,479$. Среднее отклонение эмпирических и расчетных значений уровней временного ряда составляет 0,479%, что свидетельствует о хорошем качестве модели параболического тренда.

10.2. Задача для самостоятельного выполнения

Задача 20. Проведите графическую визуализацию уровней временного ряда показателя, представленного в таблице:

Год	Варианты				
	1	2	3	4	5
	Состояние парка основных строительных машин в строительных организациях (удельный вес машин с истекшим сроком службы в общем количестве машин, процентов): экскаваторы одноковшовые, %	Состояние парка основных строительных машин в строительных организациях (удельный вес машин с истекшим сроком службы в общем количестве машин, процентов): краны на пневмоколесном ходу, %	Перевозки пассажиров воздушным транспортом, млн чел.	Индекс физического объема инвестиций в основной капитал, % к предыдущему году	Производство мяса, включая субпродукты I категории, тыс. т
1991	14,5	24,2	86,0	85,1	5700
1992	14,7	26,0	62,6	60,3	4686
1993	15,7	27,6	41,5	88,3	3970
1994	16,6	28,7	33,8	75,7	3224
1995	21,5	33,0	33,0	89,9	2370
1996	23,6	34,7	28,1	81,9	1900
1997	25,6	36,3	26,2	95,0	1510
1998	31,2	43,2	23,3	88,0	1315
1999	37,6	48,6	22,4	105,3	1113
2000	42,5	55,0	23,0	117,4	1194
2001	44,5	59,6	26,4	111,7	1284
2002	47,4	62,6	28,0	102,9	1456
2003	48,4	66,0	30,9	112,7	1677
2004	47,5	65,5	35,2	116,8	1776
2005	46,8	64,6	36,5	110,2	1857
2006	44,7	63,4	39,5	117,8	2185
2007	42,3	64,1	46,8	123,8	2561

1	2	3	4	5	6	7	8	9	10	11	12	13
1	10. Временные ряды: параболический тренд второго порядка											
2												
3												

Окончание таблицы

Год	Варианты				
	1	2	3	4	5
	Состояние парка основных строительных машин в строительных организациях (удельный вес машин с истекшим сроком службы в общем количестве машин, процентов): экскаваторы одноковшовые, %	Состояние парка основных строительных машин в строительных организациях (удельный вес машин с истекшим сроком службы в общем количестве машин, процентов): краны на пневмоколёсном ходу, %	Перевозки пассажиров воздушным транспортом, млн чел.	Индекс физического объема инвестиций в основной капитал, % к предыдущему году	Производство мяса, включая субпродукты I категории, тыс. т
2008	37,7	59,2	51,5	109,5	2899
2009	37,5	58,5	46,6	86,5	3380
2010	37,3	54,9	58,6	106,3	3957
2011	34,7	51,5	66,0	110,8	4250
2012	32,1	49,0	76,1	106,8	4747
2013	31,2	45,7	86,3	100,8	5321
2014	30,9	45,4	95,0	98,5	5957
2015	31,2	42,8	93,6	91,6	6631
2016	33,4	43,2	91,0	99,8	6827
2017	32,4	37,1	108,0	104,4	7479

Примечание - Данные приведены по РФ из официальных статистических источников (www.gks.ru).

Рассчитайте основные показатели динамики. Сделайте соответствующие выводы. Оцените параметры модели параболического тренда второго порядка по методу наименьших квадратов. Проверьте статистическую значимость оценок параметров модели, модели в целом ($\alpha = 0,05$). Выполните прогнозирование значений уровней временного ряда на период 2018-2020 гг. (при условии наличия такой возможности). Какова ошибка аппроксимации полученной модели параболического тренда?

1	2	3	4	5	6	7	8	9	10	11	12	13
2	<i>Эконометрика в табличном редакторе MS Excel</i>											
3												

10.3. Тесты

1. Замена $x_1 = t$; $x_2 = t^2$ подходит для модели следующего вида:

А) $y_t = b_0 + b_1t + b_2t^2 + \varepsilon_t$;

Б) $y_t = b_0 + b_1t + b_2t^2 + b_3t^3 + \varepsilon_t$;

В) $y_t = \frac{1}{b_0 + b_1t + b_2t^2} + \varepsilon_t$;

Г) $y_t = b_0 + \frac{b_1}{t} + \frac{b_2}{t^2} + \varepsilon_t$.

2. Укажите, какие из представленных моделей являются моделями нелинейного тренда (задание с выбором нескольких правильных ответов из предложенных):

А) $y_t = b_0 + b_1t + b_2t^2 + \varepsilon_t$;

Б) $y = \alpha_0 + \alpha_1x_1 + \alpha_2x_2 + \varepsilon$;

В) $y_t = \alpha_0 + \alpha_1t + \varepsilon_t$;

Г) $y_t = \alpha_0 \cdot t^{\alpha_1} \cdot \varepsilon_t$.

3. Тренд модели временных данных характеризует:

А) влияние долговременных факторов, т.е. длительную “вековую” тенденцию изменения признака;

Б) повторяемость экономических процессов в течение не очень длительного периода;

В) повторяемость экономических процессов в течение длительных периодов;

Г) влияние не поддающихся учету и регистрации случайных факторов.

4. Модель вида $y_t = b_0 + b_1t + b_2t^2 + b_3t^3 + \dots + b_kt^k + \varepsilon_t$ называется:

А) полулогарифмической;

Б) моделью множественной линейной зависимости;

В) экспоненциальной;

Г) полиномиальной.

5. Нелинейной относительно переменных параметров является модель:

А) $y = a + bx + \varepsilon$;

Б) $y = \frac{a + bx}{c} + \varepsilon$;

В) $y = a + bx + cx^2 + \varepsilon$;

Г) $y = e^{a+bx} \cdot \varepsilon$.

1	2	3	4	5	6	7	8	9	10	11	12	13	
1	10. Временные ряды: параболический тренд второго порядка												
2													
3													

6. Укажите, какое из уравнений описывает степенную зависимость между переменными:

А) $y = e^{a+bx} \cdot \varepsilon$;

Б) $y = ax^b \cdot \varepsilon$;

В) $y = a + b \cdot \ln x + \varepsilon$;

Г) $y = a + \frac{b}{x} + \varepsilon$.

7. Пусть X и Y - случайные величины, η - корреляционное отношение. Свойствами корреляционного отношения являются следующие:

А) при вычислении корреляционного отношения не имеет значения, какую переменную считать независимой, а какую - зависимой, $\eta_{XY} = \eta_{YX}$;

Б) корреляционное отношение есть неотрицательная величина, не превосходящая единицы, $0 \leq \eta \leq 1$;

В) если $\eta = 0$, то между переменными корреляционная связь отсутствует;

Г) если $\eta = 1$, то между переменными корреляционная связь отсутствует.

Выберите несколько вариантов ответа.

10.4. Контрольные вопросы

1. С какой целью используется корреляционное отношение в нелинейной зависимости?
2. Перечислите свойства корреляционного отношения.
3. В каком случае целесообразно использовать параболический тренд второго порядка $y_t = b_0 + b_1 t + b_2 t^2 + \varepsilon$?
4. Как построить параболический тренд второго порядка с использованием графических инструментов табличного редактора Excel?
5. Какие преобразования следует выполнить для оценки параметров параболического тренда второго порядка методом наименьших квадратов?
6. Какие преобразования следует выполнить для оценки параметров параболического тренда более высоких порядков методом наименьших квадратов?
7. Как осуществляется проверка статистической значимости оценок параметров модели $y_t = b_0 + b_1 t + b_2 t^2 + \varepsilon$? Как проверить статистическую значимость данной модели?
8. Как интерпретируются параметры параболического тренда второго порядка $y_t = b_0 + b_1 t + b_2 t^2 + \varepsilon$?
9. Какие социально-экономические явления (процессы) можно моделировать с помощью параболического тренда второго порядка?
10. С чем связано ограничение в использовании параболического тренда более высоких порядков?

11. ВРЕМЕННЫЕ РЯДЫ: ГИПЕРБОЛИЧЕСКИЙ ТРЕНД



Продемонстрирована технология оценивания гиперболического тренда с применением инструментов анализа данных табличного редактора MS Excel.

11.1. Решение типовых задач

Для иллюстрации особенностей построения и анализа гиперболического тренда воспользуемся данными из Российского статистического ежегодника (раздел «Транспорт» (www.gks.ru)). Статистический массив величины перевозок грузов всеми видами транспорта в РФ за период 1991-2017 гг. представлен в табл. 11.1.

Таблица 11.1

Перевозки грузов всеми видами транспорта в РФ

Год	Перевозки грузов всеми видами транспорта, млн т	Год	Перевозки грузов всеми видами транспорта, млн т
1991	18307,8	2005	9167,1
1992	15736,3	2006	9300,7
1993	12752,0	2007	9450,6
1994	9940,0	2008	9451,1
1995	8813,7	2009	7469,5
1996	8166,2	2010	7750,4
1997	7482,5	2011	8337,2
1998	7078,2	2012	8519,2
1999	7381,5	2013	8264,5
2000	7907,1	2014	8006,4
2001	8200,3	2015	7582,0
2002	8487,5	2016	7954,0
2003	8767,8	2017	8071,0
2004	8978,3	-	-

11. Временные ряды: гиперболический тренд

Осуществим аналитическое выравнивание на основе гиперболической функции.

Визуально значения уровней временного ряда необходимо представить графически с использованием диаграммы.

Шаг 1. Для построения диаграммы в главном меню выберем **Вставка / График / График с маркерами**. Результат построения диаграммы представлен на рис. 11.1.

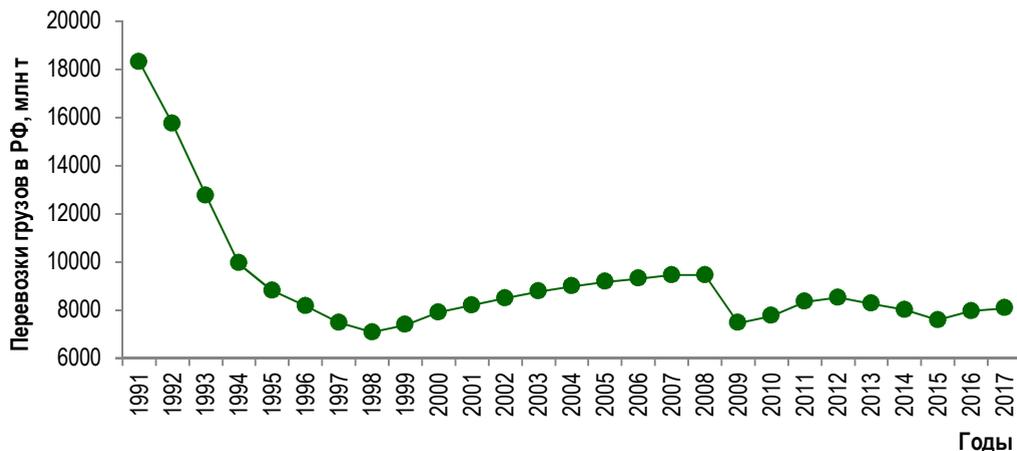


Рис. 11.1. Динамика величины перевозки грузов всеми видами транспорта в РФ

Визуализация расположения точек на диаграмме позволяет сделать предположение о нелинейной зависимости величины перевозки грузов от фактора времени. Функция, позволяющая наилучшим образом описать рассматриваемый тренд, имеет вид

$$y_t = b_0 + b_1 \cdot \frac{1}{t} + \varepsilon.$$

Шаг 2. Приведенная выше модель гиперболического тренда линейна относительно параметров, но не линейна по переменной (гипербола).

Для оценки параметров модели воспользуемся пакетом **«Анализ данных»**, предварительно создав на листе табличного редактора Excel дополнительные столбцы t , $\frac{1}{t}$ (t - номер уровня временного ряда, $\frac{1}{t}$ - соответствующий расчетный столбец). Указанные построения приведены на рис. 11.2.

	A	B	C	D
1	Год	y	t	1/t
2	1991	18307,8	1	1,000
3	1992	15736,3	2	0,500
4	1993	12752,0	3	0,333
5	1994	9940,0	4	=1/C5
6	1995	8813,7	5	0,200
7	1996	8166,2	6	0,167
8	1997	7482,5	7	0,143
9	1998	7078,2	8	0,125
10	1999	7381,5	9	0,111

Рис. 11.2. Добавление переменных t и $\frac{1}{t}$

Шаг 3. Для оценивания параметров модели $y_t = b_0 + b_1 \cdot \frac{1}{t} + \varepsilon$ по методу наименьших квадратов воспользуемся инструментом анализа данных «РЕГРЕССИЯ». В окне «Регрессия» в качестве «Входного интервала Y» укажем исходные значения уровней временного ряда, в качестве «Входного интервала X» - переменную $\frac{1}{t}$. Результаты работы инструмента РЕГРЕССИЯ приведены на рис. 11.3.

	A	B	C	D	E	F	G	H	I
1	ВЫВОД ИТОГОВ								
2									
3	<i>Регрессионная статистика</i>								
4	Множественный R	0,92							
5	R-квадрат	0,84							
6	Нормированный R-квадрат	0,83							
7	Стандартная ошибка	1043,10							
8	Наблюдения	27							
9									
10	<i>Дисперсионный анализ</i>								
11		<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>ачимость F</i>			
12	Регрессия	1	1,41E+08	1,41E+08	129,75	2,18E-11			
13	Остаток	25	27201670	1088067					
14	Итого	26	1,68E+08						
15									
16		<i>Коэффициенты</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>	<i>P-Значение</i>	<i>Нижние 95%</i>	<i>Верхние 95%</i>	<i>Нижние 95,0%</i>	<i>Верхние 95,0%</i>
17	Y-пересечение	7487,08	248,74	30,10	0,00	6974,79	7999,37	6974,79	7999,37
18	1/t	11607,93	1019,08	11,39	0,00	9509,09	13706,76	9509,09	13706,76

Рис. 11.3. Результаты оценки параметров тренда $y_t = b_0 + b_1 \cdot \frac{1}{t} + \varepsilon$

11. Временные ряды: гиперболический тренд

Выборочное уравнение регрессии имеет вид $\hat{y}_t = 7487,08 + 11607,93 \frac{1}{t}$. Оценки параметров модели линейного тренда \hat{b}_0, \hat{b}_1 являются статистически значимыми, так как **p-значение** много меньше принятого в данном случае уровня значимости $\alpha = 0,05$.

Полученная регрессия также статистически значима по F-критерию Фишера ($F_{набл} > F_{кр}(\alpha = 0,05; k_1 = 1; k_2 = 25) = 4,24$), при этом индекс детерминации равен 0,916.

ЗАМЕЧАНИЕ: гиперболическая функция используется при моделировании зависимости, в которой при неограниченном увеличении X эндогенная переменная Y асимптотически приближается к некоторому предельному значению. Параметр b_0 показывает значение (уровень) Y , которое устанавливается при больших значениях X ; b_1 характеризует скорость приближения к данному уровню. В этом случае при сохранении сложившихся социально-экономических условий в России объем перевезенных грузов не снизится менее 7487,08 млн т при скорости приближения к данному уровню 11607,93 млн т ежегодно.

Шаг 4. Вычислим расчетные значения величины перевозки грузов всеми видами транспорта в РФ в период 1995-2019 г. по полученному уравнению $\hat{y}_t = 7487,08 + 11607,93 \frac{1}{t}$ (рис. 11.4).

ЛИНЕЙН		=SHS17+SHS18*D2						
	A	B	C	D	E	F	G	H
1	Год	Y	t	1/t	Y^		ВЫВОД ИТОГОВ	
2	1991	18307,8	1		=SHS17+SHS18*D2			
3	1992	15736,3	2	0,5	13291,0434		Регрессионная статистика	
4	1993	12752	3	0,333333	11356,3888		Множественный R	0,915666293
5	1994	9940	4	0,25	10389,0616		R-квадрат	0,83844476
6	1995	8813,7	5	0,2	9808,6652		Нормированный R-квадрат	0,83198255
7	1996	8166,2	6	0,16667	9421,73429		Стандартная ошибка	1043,104407
8	1997	7482,5	7	0,14286	9145,35507		Наблюдения	27
9	1998	7078,2	8	0,125	8938,07065			
10	1999	7381,5	9	0,11111	8776,84944		Дисперсионный анализ	
11	2000	7907,1	10	0,1	8647,87247			df
12	2001	8200,3	11	0,09091	8542,34586		Регрессия	1
13	2002	8487,5	12	0,083333	8454,40701		Остаток	25
14	2003	8767,8	13	0,07692	8379,99722		Итого	26
15	2004	8978,3	14	0,07143	8316,2174			
16	2005	9167,1	15	0,06667	8260,94156			Коэффициенты
17	2006	9300,7	16	0,0625	8212,5752		Y-пересечение	7487,079742
18	2007	9450,6	17	0,05882	8169,89899		1/t	11607,92727

Рис. 11.4. Вычисление расчетных значений величины перевозки грузов всеми видами транспорта в РФ по уравнению гиперболического тренда $\hat{y}_t = 7487,08 + 11607,93 \frac{1}{t}$

Аналогичным образом можно получить прогнозные значения уровней временного ряда за период 2018-2020 гг., подставив в уравнение номера и обратные значения номеров уровней временного ряда (28 - 2019 г.; 29 - 2020 г., 20 - 2021 г.).

Шаг 5. Для нанесения гиперболического тренда на график динамики рассматриваемого показателя необходимо скопировать столбцы «Год», «Y», «Y^» в свободные столбцы листа табличного редактора Excel. Затем с помощью мастера диаграмм построить соответствующий график (рис. 11.5).

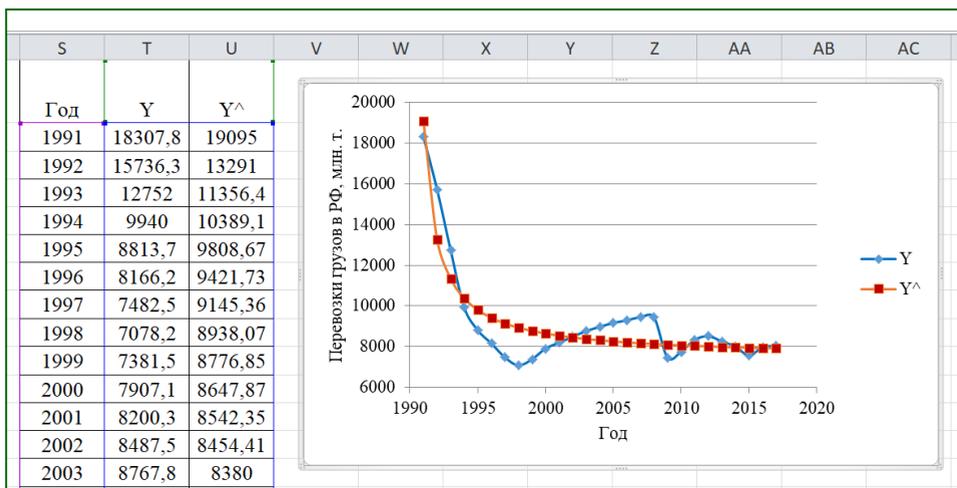


Рис. 11.5. Нанесение линии гиперболического тренда на график динамики перевозки грузов в РФ

Год	Y	t	1/t	Y^	$\frac{ y_t - \hat{y}_t }{y_t}$
2009	7469,5	19	0,05263	8098,02328	0,0841453
2010	7750,4	20	0,05	8067,47611	0,0409109
2011	8337,2	21	0,04762	8039,83818	0,0356669
2012	8519,2	22	0,04545	=ABS((B23-E23)/B23)	
2013	8264,5	23	0,04348	7991,772	0,029999
2014	8006,4	24	0,04167	7970,74338	0,0044535
2015	7582	25	0,04	7951,39683	0,0487202
2016	7954	26	0,03846	7933,53848	0,0025725
2017	8071	27	0,03704	7917,00297	0,0190803
					2,3696381
					Сумма

Рис. 11.6. Расчет средней ошибки аппроксимации

$$\text{для тренда } \hat{y}_t = 7487,08 + 11607,93 \frac{1}{t}$$

11. Временные ряды: гиперболический тренд

Шаг 6. Для расчета средней ошибки аппроксимации используется формула

$$A = \frac{1}{n} \sum_{t=1}^n \left| \frac{y_t - \hat{y}_t}{y_t} \right| \cdot 100\%.$$

В исходной таблице следует добавить еще один столбец (рис. 11.6). Для вычисления модуля числа необходимо использовать встроенную функцию **ABS**.

Таким образом, средняя ошибка аппроксимации составила $A = \frac{1}{27} \cdot 2,3696 \cdot 100 = 8,78$. Среднее отклонение эмпирических и расчетных значений уровней временного ряда составляет 8,78 %, что свидетельствует о приемлемом качестве модели гиперболического тренда.

11.2. Задача для самостоятельного выполнения

Задача 21. Проведите графическую визуализацию уровней временного ряда показателя, представленного в таблице:

Год	Варианты				
	1	2	3	4	5
	Посевная площадь сахарной свеклы, тыс. га	Валовой сбор льноволокна, тыс. т	Поголовье свиней, млн голов	Поголовье крупного рогатого скота, млн голов	Среднегодовая численность занятых в экономике, тыс. чел.
1991	1399	102	35,4	54,7	73848
1992	1439	78	31,5	52,2	71905
1993	1333	58	28,6	48,9	70720
1994	1104	54	24,9	43,3	68070
1995	1085	69	22,6	39,7	66330
1996	1060	59	19,1	35,1	65748
1997	933	23	17,3	31,5	64574
1998	810	34	17,2	28,5	63683
1999	900	24	18,3	28,1	64114
2000	805	51	15,8	27,5	64517
2001	773	58	16,2	27,4	64980
2002	808	38	17,6	26,8	65574
2003	923	55	16,3	25,1	65905
2004	849	58	13,7	23,2	66331

Окончание таблицы

Год	Варианты				
	1	2	3	4	5
	Посевная площадь сахарной свеклы, тыс. га	Валовой сбор льноволокна, тыс. т	Поголовье свиней, млн голов	Поголовье крупного рогатого скота, млн голов	Среднегодовая численность занятых в экономике, тыс. чел.
2005	799	56	13,8	21,6	66683
2006	996	36	16,2	21,6	67047
2007	1060	47	16,3	21,5	67922
2008	819	52	16,2	21,0	68397
2009	819	52	17,2	20,7	67418
2010	1160	35	17,2	20,0	67493
2011	1292	43	17,3	20,1	67644
2012	1143	46	18,8	19,9	67968
2013	904	39	19,1	19,6	67901
2014	919	37	19,5	19,3	67813
2015	1022	45	21,4	18,6	68389
2016	1107	41	21,9	18,3	72393
2017	1198	39	23,1	18,3	72142

Примечание - Данные приведены по РФ из официальных статистических источников (www.gks.ru).

Рассчитайте основные показатели динамики. Сделайте соответствующие выводы. Оцените параметры модели гиперболического тренда по методу наименьших квадратов. Проверьте статистическую значимость оценок параметров модели, модели в целом ($\alpha = 0,05$). Выполните прогнозирование значений уровней временного ряда на период 2018-2020 гг. (при условии наличия такой возможности). Какова ошибка аппроксимации полученной модели гиперболического тренда? Нанесите на график линию гиперболического тренда.

11.3. Тесты

1. Степенная модель нелинейной функции разделяется на два класса:
 - А) нелинейные модели внутренне линейны, нелинейные модели внутренне нелинейны;
 - Б) линейные модели внутренне линейны, нелинейные модели внутренне нелинейны;

	1	2	3	4	5	6	7	8	9	10	11	12	13
1				11. Временные ряды: гиперболический тренд									
2													
3													

В) нелинейные модели внутренне линейны, линейные модели внутренне нелинейны.

2. Обратная зависимость, характеризуемая нижней асимптотой, т.е. минимальным предельным значением Y , оценкой которого служит параметр b_0 в гиперболической модели, имеется:

- А) при $b_1 < 0$;
- Б) при $b_1 > 0$;
- В) при $b_1 = 0$.

3. К классу нелинейной модели по оцениваемым параметрам относятся функции:

- А) степенные, показательные, полиномы разных порядков;
- Б) степенные, полиномы разных порядков;
- В) степенные, показательные, экспоненциальные;
- Г) нет верного варианта ответа.

4. Модель $\hat{y}_i = b_0 + b_1 \frac{1}{t_i}$ относится к моделям:

- А) нелинейным относительно объясняющих переменных, но линейным по оцениваемым параметрам;
- Б) линейным относительно объясняющих переменных, но нелинейным по оцениваемым параметрам;
- В) нелинейным по оцениваемым параметрам;
- Г) нелинейным по зависимой переменной.

5. Модель $\hat{y}_i = b_0 + b_1 \frac{1}{t_i}$ является по функциональной форме:

- А) параболой;
- Б) степенной функцией;
- В) экспонентой;
- Г) гиперболой.

6. Для проверки существенности нелинейной регрессии с помощью F-критерия используют:

- А) параметры уравнения;
- Б) индекс детерминации;
- В) коэффициент корреляции;
- Г) коэффициент эластичности.

7. Оценка параметров нелинейной регрессии осуществляется методом:

- А) логарифмирования;
- Б) потенцирования;

	1	2	3	4	5	6	7	8	9	10	11	12	13
1			<i>Эконометрика в табличном редакторе MS Excel</i>										
2													
3													

- В) наименьших квадратов;
 Г) интегральным методом.

11.4. Контрольные вопросы

1. Перечислите основные показатели динамики.
2. Можно ли использовать метод наименьших квадратов для расчета оценок параметров гиперболического тренда $y_t = b_0 + b_1 \cdot \frac{1}{t} + \varepsilon_t$?
3. Как интерпретируются параметры гиперболического тренда $y_t = b_0 + b_1 \cdot \frac{1}{t} + \varepsilon_t$?
4. Как осуществляется проверка статистической значимости оценок параметров модели $y_t = b_0 + b_1 \cdot \frac{1}{t} + \varepsilon_t$? Как проверить статистическую значимость данной модели?
5. Как оценить параметры модели $y_t = b_0 + b_1 \cdot \frac{1}{t} + \varepsilon$ с использованием инструмента анализа данных «РЕГРЕССИЯ»?
6. Какие социально-экономические явления (процессы) можно моделировать с помощью гиперболического тренда?
7. Для чего используется средняя ошибка аппроксимации?
8. Какие виды прогнозирования вы знаете?
9. Как осуществляется прогнозирование по модели $y_t = b_0 + b_1 \cdot \frac{1}{t} + \varepsilon_t$?
10. Для чего используется тест П. Зарембки?

12. ВРЕМЕННЫЕ РЯДЫ: ФИКТИВНЫЕ ПЕРЕМЕННЫЕ В АНАЛИЗЕ СЕЗОННОСТИ



Рассмотрены особенности оценивания сезонной компоненты в структуре временного ряда с применением фиктивных переменных.

12.1. Решение типовых задач

Для иллюстрации особенностей анализа временного ряда с сезонной компонентой используем официальные ежеквартальные статистические данные по величине среднемесячной номинальной заработной платы в РФ за период 2010-2016 гг. (www.gks.ru). Статистический массив представлен в табл. 12.1.

Таблица 12.1

Среднемесячная заработная плата в РФ по годам

Период	У, руб.	Период	У, руб.
2010		2013	
I	19485	I	27339
II	20809	II	30245
III	21031	III	29578
IV	23491	IV	33269
2011		2014	
I	21354	I	30057
II	23154	II	32963
III	23352	III	31730
IV	26905	IV	35685
2012		2015	
I	24407	I	31566
II	26547	II	34703
III	26127	III	32983
IV	30233	IV	35692

Окончание табл. 12.1

Период	Y, руб.	Период	Y, руб.
2016		-	-
I	34000	-	-
II	37404	-	-
III	35744	-	-
IV	39649	-	-

Шаг 1. Проведем визуализацию имеющегося временного ряда. В главном меню выбираем **Вставка / График / График с маркерами**. Результат построения графика динамики исследуемого показателя представлен на рис. 12.1.

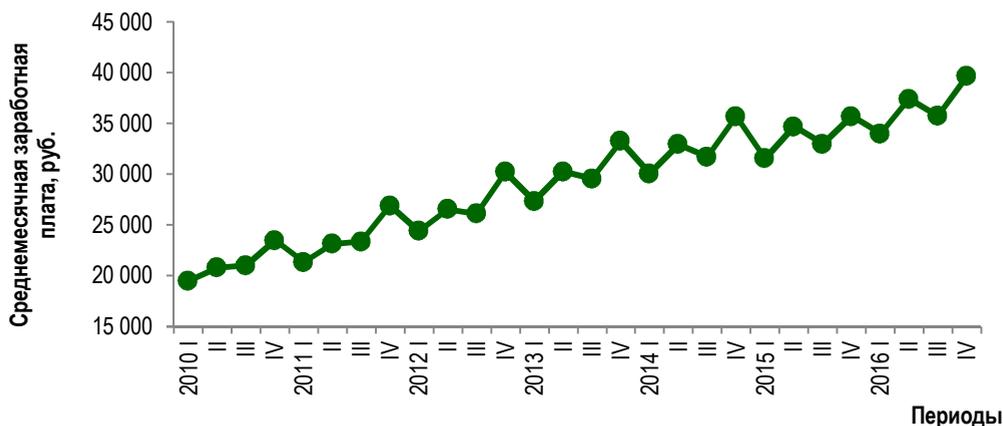


Рис. 12.1. Динамика среднемесячной заработной платы в РФ

Визуально на графике наблюдается рост среднемесячной заработной платы в течение всего периода 2010-2016 гг. Однако на протяжении года прослеживаются определенные колебания (сезонность). Динамика роста показателя может быть описана аддитивной моделью временного ряда, содержащей линейный тренд и сезонную компоненту. Теоретический вид модели следующий: $y_t = b_0 + b_1t + S_t + \varepsilon_t$.

Шаг 2. Так как сезонные изменения заработной платы носят ежеквартальный характер (4 квартала) в модель требуется включить три фиктивные переменные: $y_t = b_0 + b_1t + \alpha_1Z_{1t} + \alpha_2Z_{2t} + \alpha_3Z_{3t} + \varepsilon_t$.

Градации фиктивных переменных:

$$Z_{1t} = \begin{cases} 1, & \text{если рассматривается II квартал,} \\ 0 & \text{в противном случае;} \end{cases}$$

12. Временные ряды: фиктивные переменные в анализе сезонности

$$Z_{2t} = \begin{cases} 1, & \text{если рассматривается III квартал,} \\ 0 & \text{в противном случае;} \end{cases}$$

$$Z_{3t} = \begin{cases} 1, & \text{если рассматривается IV квартал,} \\ 0 & \text{в противном случае.} \end{cases}$$

Результат кодирования представлен на рис. 12.2.

	A	B	C	D	E	F
		Средняя номинальная заработная плата, рублей в месяц	t	Z1	Z2	Z3
1						
2	2010 I	19485	1	0	0	0
3	II	20809	2	1	0	0
4	III	21031	3	0	1	0
5	IV	23491	4	0	0	1
6	2011 I	21354	5	0	0	0
7	II	23154	6	1	0	0
8	III	23352	7	0	1	0
9	IV	26905	8	0	0	1

Рис. 12.2. Кодирование фиктивных переменных при оценивании модели вида

$$y_t = b_0 + b_1 t + \alpha_1 Z_{1t} + \alpha_2 Z_{2t} + \alpha_3 Z_{3t} + \varepsilon$$

Коэффициент $\alpha_j, j = 1, 2, 3$ в приведенной модели называется дифференциальным коэффициентом свободного члена, показывающим, на какую величину свободный член модели при значении фиктивной переменной, равном единице, отличается от свободного члена модели при базовом значении фиктивной переменной.

Шаг 3. Оценивание приведенной выше модели будем осуществлять с использованием инструмента анализа данных **РЕГРЕССИЯ**. В окне «**Регрессия**» в качестве «**Входного интервала Y**» укажем исходные значения уровней временного ряда, в качестве «**Входного интервала X**» - переменные Y, t, Z_1, Z_2, Z_3 . Результаты работы инструмента **РЕГРЕССИЯ** приведены на рис. 12.3.

Выборочное уравнение регрессии имеет вид

$$\hat{y}_t = 18488,45 + 646,03t + 1870,68Z_1 + 470,37Z_2 + 3307,05Z_3.$$

Исходя из оцененного уравнения, запишем частные выборочные уравнения для каждого квартала:

$$\text{I кв. - } \hat{Y} = 18488,45 + 646,03t;$$

$$\text{II кв. - } \hat{Y} = (18488,45 + 1870,68) + 646,03t;$$

$$\text{III кв. - } \hat{Y} = (18488,45 + 470,37) + 646,03t;$$

$$\text{IV кв. - } \hat{Y} = (18488,45 + 3307,05) + 646,03t.$$

Эконометрика в табличном редакторе MS Excel

	A	B	C	D	E	F	G	H	I
1	ВЫВОД ИТОГОВ								
2									
3	<i>Регрессионная статистика</i>								
4	Множественный R	0,992							
5	R-квадрат	0,984							
6	Нормированный R-квадрат	0,982							
7	Стандартная ошибка	767,987							
8	Наблюдения	28							
9									
10	<i>Дисперсионный анализ</i>								
11		<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>ачимость F</i>			
12	Регрессия	4	8,48E+08	2,12E+08	359,36394	2,28E-20			
13	Остаток	23	13565494	589804,1					
14	Итого	27	8,61E+08						
15									
16		<i>Кoeffициенты</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>	<i>P-Значение</i>	<i>Нижние 95%</i>	<i>Верхние 95%</i>	<i>Нижние 95,0%</i>	<i>Верхние 95,0%</i>
17	Y-пересечение	18488,45	374,01	49,43	0,00	17714,76	19262,14	17714,76	19262,14
18	t	646,03	18,14	35,61	0,00	608,50	683,56	608,50	683,56
19	Z1	1870,68	410,91	4,55	0,00	1020,66	2720,71	1020,66	2720,71
20	Z2	470,37	412,11	1,14	0,27	-382,14	1322,87	-382,14	1322,87
21	Z3	3307,05	414,10	7,99	0,00	2450,42	4163,68	2450,42	4163,68

Рис. 12.3. Результаты оценки параметров модели $y_t = b_0 + b_1t + \alpha_1Z_{1t} + \alpha_2Z_{2t} + \alpha_3Z_{3t} + \varepsilon$

Представляется возможным спрогнозировать величину среднемесячной номинальной заработной платы в каждом квартале 2017 г. Для этого следует использовать соответствующее частное уравнение регрессии с указанием номера уровня временного ряда: 29 - I квартал 2017 г. и т.д. (рис. 12.4).

	A	B	C	D	E	F
1		Средняя номинальная заработная плата, рублей в месяц	t	Z1	Z2	Z3
23	II	34703	22	1	0	0
24	III	32983	23	0	1	0
25	IV	35692	24	0	0	1
26	2016 I	34000	25	0	0	0
27	II	37404	26	1	0	0
28	III	35744	27	0	1	0
29	IV	39649	28	0	0	1
30	2017 I		29	0	0	0
31	II		30	1	0	0
32	III		31	0	1	0
33	IV		32	0	0	1

Рис. 12.4. Кодирование переменных для построения прогноза среднемесячной номинальной начисленной заработной платы в РФ в каждом квартале 2017 г.

12. Временные ряды: фиктивные переменные в анализе сезонности

Визуализация полученных прогнозных значений представлена на рис. 12.5.

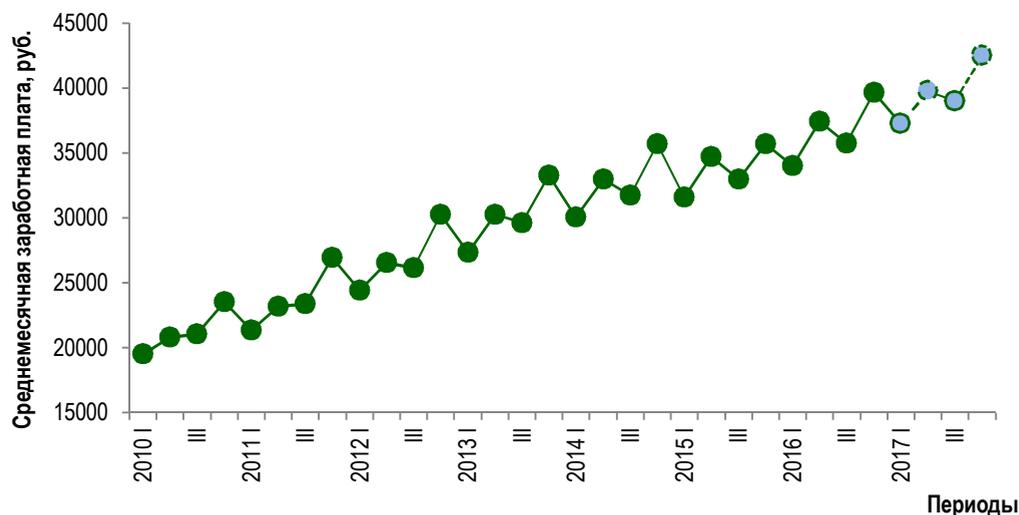


Рис. 12.5. Прогноз среднемесячной номинальной заработной платы россиян в 2017 г. с учетом сезонности и долговременной тенденции

Согласно полученным результатам моделирования тенденция роста сохраняется в 2017 г., но при этом будут наблюдаться периоды спада (I и III кварталы) и роста (II и IV кварталы) исследуемого показателя.

12.2. Задача для самостоятельного выполнения

Задача 22. Проведите графическую визуализацию уровней временного ряда показателя, представленного в таблице:

Период	Варианты				
	1	2	3	4	5
	Объем работ, выполненных по виду деятельности «Строительство» в текущих ценах, млрд руб.	Ввод в действие жилых домов, млн м ²	Коммерческий грузооборот транспорта, млрд т-км	Оборот розничной торговли в текущих ценах, млрд руб.	Доходы консолидированного бюджета, млрд руб.
2010					
I	646,9	9,6	1145	3625,7	3463,1
II	1055,8	12	1146	3934,5	4140,4

Окончание таблицы

Период	Варианты				
	1 Объем работ, выполненных по виду дея- тельности «Строитель- ство» в текущих ценах, млрд руб.	2 Ввод в действие жилых домов, млн м ²	3 Коммерческий грузооборот транспорта, млрд т-км	4 Оборот рознич- ной торговли в текущих ценах, млрд руб.	5 Доходы консо- лидированного бюджета, млрд руб.
III	1288,4	11,4	1139	4203,8	3815,9
IV	1463	25,5	1217	4704,6	4296,5
2011					
I	733,6	9,3	1189	4180,1	4392,5
II	1159,1	11,4	1205	4568,3	5492,6
III	1496,2	13	1164	4895	5285,7
IV	1751,4	28,6	1243	5439,2	5684,6
2012					
I	843,7	9,8	1245,3	4689,7	5103,5
II	1330,7	11,2	1213	5112,2	5998
III	1660,2	13,6	1211,1	5492,4	5583,7
IV	1879,5	31,1	1264,7	6100,2	6403,5
2013					
I	939,6	10,4	1225	5241,3	5401,6
II	1395,3	12,2	1211	5692,8	5969,1
III	1719,2	16,1	1217	6052	6046,8
IV	1965,4	30,7	1305	6699,8	6664,9
2014					
I	934,2	14	1244	5792,9	5960,4
II	1413	15,7	1227,1	6256,7	6710,8
III	1756,6	18,9	1206	6697,3	6550,2
IV	2021,4	35,6	1278	7609,3	7149,7
2015					
I	981,4	18,6	1225	6271,5	6044,6
II	1385,8	16,1	1197	6595,9	6704
III	1676,1	17,7	1223	7000,4	6747,6
IV	2105,1	32,9	1333	7670,6	6997,9
2016					
I	967,8	15,6	1245	6445,1	5876,1
II	1320,7	15,9	1210	6692,8	6645,4
III	1704,1	18	1260	7213,8	6853
IV	2191,8	30,3	1357,7	7785,4	8372,2
Примечание - Данные приведены по РФ из официальных статистических источников (www.gks.ru).					

12. Временные ряды: фиктивные переменные в анализе сезонности

Рассчитайте основные показатели динамики. Сделайте соответствующие выводы. Оцените параметры модели линейного тренда с наличием сезонной компоненты. Сделайте вывод о влиянии сезонности на динамику анализируемого показателя. Спрогнозируйте тенденцию динамического изменения показателя в каждом квартале 2017 г.

12.3. Тесты

1. Если качественный признак, который необходимо включить в эконометрическую модель, имеет четыре градации, то модель содержит:

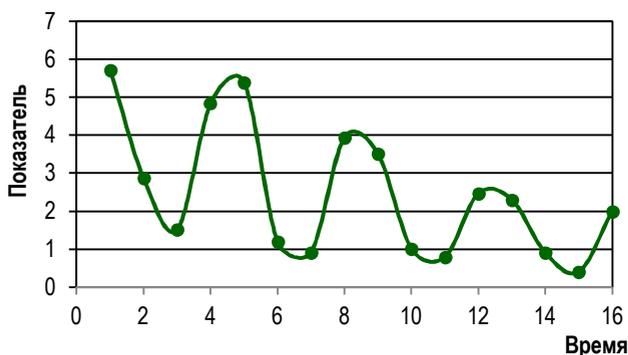
- А) четыре фиктивные переменные;
- Б) **три фиктивные переменные;**
- В) одну фиктивную переменную;
- Г) нельзя включать фиктивные переменные в модель.

2. Временной ряд, график которого приведен ниже, содержит следующие компоненты:



- А) только тренд и случайную компоненту;
- Б) тренд, аддитивную сезонную компоненту и случайную компоненту;
- В) тренд, мультипликативную сезонную компоненту и случайную компоненту;
- Г) только случайную компоненту.

3. Временной ряд, график которого приведен ниже, содержит следующие компоненты:



- А) только тренд и случайную компоненту;
- Б) тренд, аддитивную сезонную компоненту и случайную компоненту;
- В) тренд, мультипликативную сезонную компоненту и случайную компоненту;
- Г) только случайную компоненту.

4. Временной ряд, модель которого имеет вид

$$y_t = b_0 + b_1 t + \alpha_1 Z_{1t} + \alpha_2 Z_{2t} + \alpha_3 Z_{3t} + \varepsilon_t$$
, содержит следующие компоненты:

- А) тренд, квартальную сезонную компоненту и случайную компоненту;
- Б) только тренд и случайную компоненту;
- В) только случайную компоненту;
- Г) квартальную сезонную компоненту и случайную компоненту.

5. Пусть Y_t - временной ряд в поквартальной динамике, S_t - аддитивная сезонная компонента, $t = \overline{1, 4}$. Тогда сумма всех сезонных компонент равна:

- А) 4;
- Б) 1;
- В) 0;
- Г) 2.

6. Пусть Y_t - временной ряд в поквартальной динамике, S_t - мультипликативная сезонная компонента, $t = \overline{1, 4}$. Тогда сумма всех сезонных компонент равна:

- А) 4;
- Б) 1;
- В) 0;
- Г) 2.

7. Пусть Y_t - временной ряд в поквартальной динамике, S_t - аддитивная сезонная компонента, причем для второго квартала года $S_t = S_2 = 1$, для третьего квартала года $S_t = S_3 = 5$, для четвертого квартала года $S_t = S_4 = -8$. Тогда оценка сезонной компоненты для первого квартала года $S_t = S_1$ равна:

- А) 2;
- Б) 1;
- В) 4;
- Г) 8.

8. Пусть Y_t - временной ряд в поквартальной динамике, S_t - мультипликативная сезонная компонента, причем для первого квартала года $S_t = S_1 = 8$, для второго квартала года $S_t = S_2 = \frac{1}{2}$, для четвертого квартала года $S_t = S_4 = \frac{1}{4}$. Тогда оценка сезонной компоненты для третьего квартала года $S_t = S_3$ равна:

- А) 1;
- Б) $\frac{1}{2}$;

1	2	3	4	5	6	7	8	9	10	11	12	13
2	12. Временные ряды: фиктивные переменные в анализе сезонности											
3												

В) $\frac{1}{4}$;

Г) 0.

9. Временной ряд, модель которого имеет вид

$$y_t = b_0 + b_1t + \alpha_1Z_{1t} + \alpha_2Z_{2t} + \alpha_3Z_{3t} + \alpha_4Z_{4t} + \dots + \alpha_{11}Z_{11t} + \varepsilon_t,$$

содержит следующие компоненты:

- А) только линейный тренд и случайную компоненту;
- Б) линейный тренд, сезонные колебания с периодом, равным 4, и случайную компоненту;
- В) линейный тренд, сезонные колебания с периодом, равным 12, и случайную компоненту;
- Г) только случайную компоненту.

10. Выборочное уравнение регрессии для временного ряда, содержащего линейный тренд и представленного в поквартальной динамике, имеет вид

$$\hat{y}_t = 18,45 + 6,03t + 1,68Z_1 + 47,37Z_2 + 33,05Z_3,$$

где Z_1, Z_2, Z_3 - фиктивные переменные, соответственно, для I, II и III кварталов.

Тогда частное выборочное уравнение для четвертого квартала имеет вид:

А) $\hat{Y} = (18,45 + 33,05) + 6,03t$;

Б) $\hat{Y} = 18,45 + 6,03t$;

В) $\hat{Y} = 33,05 + 6,03t$.

12.4. Контрольные вопросы

1. Назовите составляющие временного ряда.
2. Какие методы моделирования сезонной составляющей временного ряда существуют?
3. Приведите примеры использования фиктивных переменных в сезонном анализе.
4. Закончите фразу: «Если качественный признак имеет p уровней, то число фиктивных переменных в модели следует взять равным...»
5. Сколько фиктивных переменных необходимо включить в модель для отражения сезонности, если данные представлены в поквартальной динамике?
6. Сколько фиктивных переменных необходимо включить в модель для отражения сезонности, если данные представлены в помесечной динамике?

	1	2	3	4	5	6	7	8	9	10	11	12	13
1			<i>Эконометрика в табличном редакторе MS Excel</i>										
2													
3													

7. Опишите отличие в поведении временного ряда в случае аддитивной и мультипликативной сезонной волны.
8. Применяются ли сезонные фиктивные переменные в случае мультипликативной сезонной волны?
9. Применяются ли сезонные фиктивные переменные в случае аддитивной сезонной волны?
10. Какие особенности имеет прогнозирование значений временного ряда, содержащего сезонную составляющую?

13. АВТОКОРРЕЛЯЦИЯ



Реализован алгоритм выявления нарушения условия Гаусса - Маркова (автокорреляции) в табличном редакторе MS Excel на примере моделирования линейного тренда во временных данных.

13.1. Решение типовых задач

Автокорреляция (последовательная корреляция) определяется как корреляция между наблюдаемыми показателями, упорядоченными во времени или в пространстве. Это нарушение одной из предпосылок МНК, а именно $\text{cov}(\varepsilon_i, \varepsilon_j) = 0, i \neq j$.

Иллюстрация обнаружения автокорреляции в процессе эконометрического моделирования будет выполнена на примере временного ряда. В табл. 13.1 приведены значения показателя «Добыча угля в РФ» за период 1995-2015 гг.

Таблица 13.1

Динамика добычи угля в РФ

Год	Добыча угля, млн т	Год	Добыча угля, млн т
1995	263	2006	310
1996	257	2007	314
1997	245	2008	329
1998	232	2009	301
1999	250	2010	322
2000	258	2011	336
2001	270	2012	357
2002	256	2013	353
2003	277	2014	357
2004	282	2015	372
2005	299	-	-

Примечание - Данные приведены по РФ из официальных статистических источников (www.gks.ru)

Шаг 1. Графическую визуализацию уровней временного ряда выполним с применением графических возможностей табличного редактора Excel. Выберем **Вставка / График / График с маркером**. Результаты визуализации приведены на рис. 13.1.

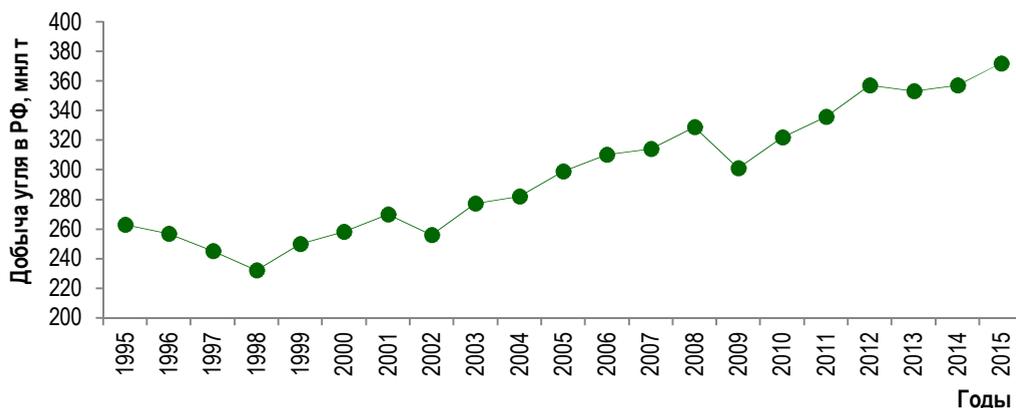


Рис. 13.1. Динамика добычи угля в РФ

Динамика добычи угля характеризуется возрастающим трендом с некоторыми сокращениями добычи в период экономических кризисов (1998 г., 2009 г.).

Шаг 2. Возрастающий тренд добычи угля можно описать с помощью эконометрической модели вида $y_t = b_0 + b_1t + \varepsilon$ (линейный тренд без сезонной компоненты).

M13			
	A	B	C
1	Год	Y	t
2	1995	263	1
3	1996	257	2
4	1997	245	3
5	1998	232	4
6	1999	250	5
7	2000	258	6
8	2001	270	7
9	2002	256	8
10	2003	277	9
11	2004	282	10
12	2005	299	11
13	2006	310	12
14	2007	314	13
15	2008	329	14
16	2009	301	15
17	2010	322	16
18	2011	336	17
19	2012	357	18
20	2013	353	19
21	2014	357	20
22	2015	372	21

Рис. 13.2. Добавление дискретной переменной t

Для оценки параметров модели воспользуемся макросом **«Анализ данных»**, предварительно создав на листе табличного редактора Excel дополнительный столбец t (t - номер уровня временного ряда). Указанные построения приведены на рис. 13.2.

Шаг 3. Для оценивания параметров модели $y_t = b_0 + b_1t + \varepsilon$ по методу наименьших квадратов воспользуемся инструментом анализа данных **РЕГРЕССИЯ**. В окне **«Регрессия»** в качестве **«Входного интервала Y»** укажем исходные значения уровней временного ряда, в качестве **«Входного интервала X»** - переменную t . Результаты работы инструмента **РЕГРЕССИЯ** приведены на рис. 13.3.

Выборочное уравнение регрессии имеет вид $\hat{y}_t = 225,57 + 6,51t$. Оценки параметров модели линейного тренда \hat{b}_0, \hat{b}_1 являются статистически значимыми, так как **p-значение** много меньше принятого в данном случае уровня значимости $\alpha = 0,05$. Полученная регрессия также статистически значима по F-критерию Фишера ($F_{набл} > F_{кр}(\alpha = 0,05; k_1 = 1; k_2 = 19) = 4,38$).

13. Автокорреляция

Год	Y	t	ВЫВОД ИТОГОВ					
1995	263	1	<i>Регрессионная статистика</i>					
1996	257	2						
1997	245	3	Множественный R	0,950797661				
1998	232	4	R-квадрат	0,904016193				
1999	250	5	Нормированный R-квадрат	0,898964414				
2000	258	6	Стандартная ошибка	13,49666423				
2001	270	7	Наблюдения	21				
2002	256	8	<i>Дисперсионный анализ</i>					
2003	277	9						
2004	282	10		<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Значимость F</i>
2005	299	11	Регрессия	1	32597,53247	32597,53247	178,9500563	4,05016E-11
2006	310	12	Остаток	19	3461,038961	182,1599453		
2007	314	13	Итого	20	36058,57143			
2008	329	14	<i>Коэффициенты</i>					
2009	301	15						
2010	322	16	Y-пересечение	225,5714286	6,107324895	36,93457159	3,71049E-19	212,7886507
2011	336	17	t	6,506493506	0,486386017	13,37722155	4,05016E-11	5,488475873

Рис. 13.3. Результаты оценки параметров тренда $y_t = b_0 + b_1t + \varepsilon$

Шаг 3. Визуально автокорреляция проявляется в систематическом отклонении точек корреляционного поля от линии регрессии. Произведем наложение линии регрессии на график динамики изменения добычи угля в РФ (рис. 13.4).

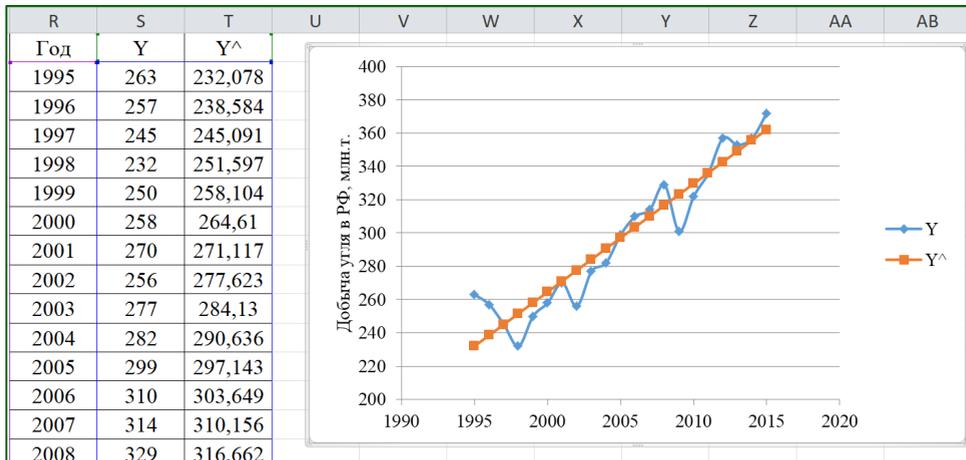


Рис. 13.4. Нанесение линейного тренда на график динамики добычи угля в РФ

На графике визуально можно наблюдать некоторую системность отклонений эмпирических значений уровней временного ряда от линии регрессии. Необходимо аналитическим методом подтвердить или опровергнуть гипотезу о наличии автокорреляции.

Шаг 4. Выявление нарушения предпосылки МНК вида $\text{cov}(\varepsilon_i, \varepsilon_j) = 0, i \neq j$ производится на основе анализа остатков регрессии $e_t = y_t - \hat{y}_t$. Результаты вычисления остатков регрессии приведены на рис. 13.5.

Год	Y	t	Y^	et	Вывод итогов																																						
1995	263	1	232,078	=B2-D2	Регрессионная статистика Множественный R 0,950797661 R-квадрат 0,904016193 Нормированный R-квадрат 0,898964414 Стандартная ошибка 13,49666423 Наблюдения 21 Дисперсионный анализ <table border="1"> <thead> <tr> <th></th> <th>df</th> <th>SS</th> <th>MS</th> <th>F</th> </tr> </thead> <tbody> <tr> <td>Регрессия</td> <td>1</td> <td>32597,53247</td> <td>32597,53247</td> <td>178,9500563</td> </tr> <tr> <td>Остаток</td> <td>19</td> <td>3461,038961</td> <td>182,1599453</td> <td></td> </tr> <tr> <td>Итого</td> <td>20</td> <td>36058,57143</td> <td></td> <td></td> </tr> </tbody> </table> <table border="1"> <thead> <tr> <th></th> <th>Коэффициенты</th> <th>Стандартная ошибка</th> <th>t-статистика</th> <th>P-Значение</th> </tr> </thead> <tbody> <tr> <td>Y-пересечение</td> <td>225,5714286</td> <td>6,107324895</td> <td>36,93457159</td> <td>3,71049E-19</td> </tr> <tr> <td>t</td> <td>6,506493506</td> <td>0,486386017</td> <td>13,37722155</td> <td>4,05016E-11</td> </tr> </tbody> </table>					df	SS	MS	F	Регрессия	1	32597,53247	32597,53247	178,9500563	Остаток	19	3461,038961	182,1599453		Итого	20	36058,57143				Коэффициенты	Стандартная ошибка	t-статистика	P-Значение	Y-пересечение	225,5714286	6,107324895	36,93457159	3,71049E-19	t	6,506493506	0,486386017	13,37722155	4,05016E-11
	df	SS	MS	F																																							
Регрессия	1	32597,53247	32597,53247	178,9500563																																							
Остаток	19	3461,038961	182,1599453																																								
Итого	20	36058,57143																																									
	Коэффициенты	Стандартная ошибка	t-статистика	P-Значение																																							
Y-пересечение	225,5714286	6,107324895	36,93457159	3,71049E-19																																							
t	6,506493506	0,486386017	13,37722155	4,05016E-11																																							
1996	257	2	238,584	18,4156																																							
1997	245	3	245,091	-0,0909																																							
1998	232	4	251,597	-19,597																																							
1999	250	5	258,104	-8,1039																																							
2000	258	6	264,61	-6,6104																																							
2001	270	7	271,117	-1,1169																																							
2002	256	8	277,623	-21,623																																							
2003	277	9	284,13	-7,1299																																							
2004	282	10	290,636	-8,6364																																							
2005	299	11	297,143	1,85714																																							
2006	310	12	303,649	6,35065																																							
2007	314	13	310,156	3,84416																																							
2008	329	14	316,662	12,3377																																							
2009	301	15	323,169	-22,169																																							
2010	322	16	329,675	-7,6753																																							
2011	336	17	336,182	-0,1818																																							

Рис. 13.5. Вычисление остатков регрессии $\hat{y}_t = 225,57 + 6,51t$

Шаг 5. Для оценивания тесноты связи между остатками в соседних наблюдениях используется выборочный коэффициент автокорреляции первого порядка:

$$r_{e_t, e_{t-1}} = \frac{\sum_{t=2}^n e_t e_{t-1}}{\sqrt{\sum_{t=1}^{n-1} e_t^2 \sum_{t=2}^n e_{t-1}^2}}$$

где e_t, e_{t-1} - остатки регрессии в соседних наблюдениях.

В табличном редакторе Excel данный коэффициент возможно вычислить применив функцию **КОРРЕЛ**, указав в поле «**Массив 1**» значения остатков e_{2-t} , в поле «**Массив 2**» - $e_{1-(t-1)}$ (рис. 13.6).

Год	Y	t	Y^	et	
1995	263	1	232,078	30,9221	r(et;et-1)
1996	257	2	238,584	18,4156	=КОРРЕЛ(Е3:Е22;Е2:Е21)
1997	245	3	245,091	-0,0909	КОРРЕЛ(массив1, массив2)
1998	232	4	251,597	-19,597	

Рис. 13.6. Вычисление коэффициента автокорреляции первого порядка

Таким образом, коэффициент автокорреляции первого порядка $r_{e_t e_{t-1}} = 0,37$, что говорит об умеренной корреляционной связи между остатками в соседних наблюдениях (по шкале Чеддока).

Шаг 6. Необходимо проверить статистическую значимость полученного коэффициента автокорреляции. Для этого предположим, что автокорреляция остатков отсутствует: $H_0: \rho = 0$. Справедливость данной гипотезы будем устанавливать с помощью статистики Дарбина - Уотсона, которая имеет вид

$$DW = \frac{\sum_{i=2}^n (e_i - e_{i-1})^2}{\sum_{i=1}^n e_i^2}.$$

Наблюдаемое значение критерия можно определить по альтернативной формуле: $DW \approx 2 \cdot (1 - r_{e_t e_{t-1}})$. Таким образом, получим $DW \approx 1,27$.

Критические точки распределения Дарбина - Уотсона табулированы, определяются по соответствующей таблице (прил. 1 или 2) по параметрам α, m, n . В данном случае $\alpha = 0,05; m = 1; n = 21; d_1 = 1,22; d_2 = 1,42$.

На рис. 13.7 приведен вид критической области соответствующей статистики.

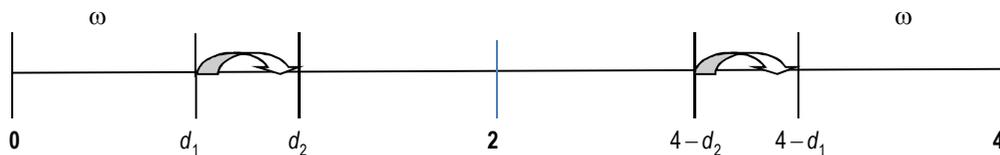


Рис. 13.7. Критическая область статистики Дарбина - Уотсона

В рассматриваемом примере наблюдаемое значение статистики Дарбина - Уотсона попало в зону неопределенности. В этом случае нельзя однозначно говорить об отсутствии автокорреляции остатков. Рационально предпринять попытку уменьшить корреляционную связь уровней временного ряда. Для этого применяют авторегрессионное преобразование Бокса - Дженкинса с поправкой Прайса - Уинстона.

13.2. Задача для самостоятельного выполнения

Задача 23. Проведите графическую визуализацию уровней временного ряда показателя, представленного в таблице (по вариантам):

Год	Варианты				
	1	2	3	4	5
	Численность рабочей силы в РФ, тыс. чел.	Численность безработных в РФ, тыс. чел.	Число общеобразовательных организаций в РФ, ед.	Число образовательных организаций высшего образования в РФ, ед.	Добыча нефти, включая газовый конденсат, в РФ, млн т
1998	67402	1929	69613	914	303
1999	72380	1263	69292	939	305
2000	72770	1037	68804	965	324
2001	71547	1123	68594	1008	348
2002	72357	1500	67431	1039	380
2003	72273	1639	66207	1044	421
2004	72985	1920	64908	1071	459
2005	73581	1830	63174	1068	470
2006	74419	1742	61028	1090	481
2007	75289	1553	57992	1108	491
2008	75700	1522	55792	1134	488
2009	75694	2147	53102	1114	494
2010	75478	1589	50793	1115	506
2011	75779	1286	48342	1080	512
2012	75676	1065	46881	1046	519
2013	75529	918	45419	969	522
2014	75428	883	44846	950	526
2015	76588	1001	43374	896	534

Примечание - Данные приведены по РФ из официальных статистических источников (www.gks.ru).

Рассчитайте основные показатели динамики. Сделайте соответствующие выводы. Оцените параметры модели линейного тренда. Сделайте вывод о возможной автокорреляции остатков. Подтвердите свое предположение аналитически при помощи нахождения коэффициента автокорреляции первого порядка и проверки его значимости ($\alpha = 0,05$).

	1	2	3	4	5	6	7	8	9	10	11	12	13
1													
2													
3													

13. Автокорреляция

13.3. Тесты

1. В эконометрической модели имеет место автокорреляция остатков, если выполняется следующее условие:

А) математическое ожидание случайной компоненты не равно нулю:

$$M(\varepsilon_i) \neq 0, \quad i = \overline{1, n};$$

Б) дисперсия случайной компоненты различна для разных наблюдений:

$$D(\varepsilon_i) \neq const, \quad i = \overline{1, n};$$

В) существует взаимосвязь между значениями случайной компоненты в различных наблюдениях:

$$\text{cov}(\varepsilon_i, \varepsilon_j) \neq 0, \quad i, j = \overline{1, n}, \quad i \neq j;$$

Г) остатки имеют нормальный закон распределения.

2. Причинами автокорреляции остатков в эконометрической модели являются:

А) отсутствие линейной зависимости между переменными;

Б) присутствие в модели фактора, несущественно влияющего на Y ;

В) инерция в изменении экономических показателей;

Г) статистический массив, представляющий собой временной ряд.

Выберите несколько вариантов ответа.

3. Последствием автокорреляции является:

А) потеря устойчивости оценок параметров регрессии \hat{b}_j ;

Б) смещенность дисперсии $D(\hat{b}_j)$ оценок коэффициентов регрессии;

В) неверный вывод о значимости модели в целом;

Г) несмещенность дисперсии оценок параметров регрессии $D(\hat{b}_j)$.

4. Вставьте пропущенное слово: «Отсутствие автокорреляции в остатках предполагает, что значения ... модели не зависят друг от друга»:

А) результативной переменной;

Б) параметров;

В) факторной переменной;

Г) остатков.

5. Отрицательная автокорреляция остатков имеет место, если:

А) сумма всех отклонений отрицательна;

Б) знаки соседних отклонений, как правило, противоположны;

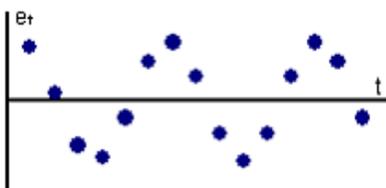
В) знаки соседних отклонений, как правило, одинаковы;

Г) подавляющее большинство отклонений отрицательно.

6. Для обнаружения автокорреляции первого порядка используется критерий:

- А) Чоу;
- Б) Стьюдента;
- В) Дарбина - Уотсона;
- Г) Фишера.

7. График зависимости остатков e_t от времени t свидетельствует:



- А) мультиколлинеарности данных;
- Б) положительной автокорреляции остатков;
- В) отрицательной автокорреляции остатков;
- Г) отсутствию корреляции в остатках.

8. Для оценки тесноты связи между соседними случайными ошибками используется:

- А) коэффициент ранговой корреляции Спирмена r_{xe} ;
- Б) выборочный коэффициент автокорреляции первого порядка $r(e_i; e_{i-1})$, $i = \overline{1, n}$;
- В) статистика Дарбина - Уотсона;
- Г) ковариация $\text{cov}(\varepsilon_i, \varepsilon_k) \neq 0$, $i \neq k$.

9. Если наблюдаемое значение статистики DW принимает значения из интервала $(d_2, 4 - d_2)$, то:

- А) принимается гипотеза H_0 об отсутствии автокорреляции в эконометрической модели;
- Б) отклоняется гипотеза H_0 о наличии автокорреляции остатков в эконометрической модели;
- В) ничего определенного о наличии автокорреляции остатков в эконометрической модели сказать нельзя.

13.4. Контрольные вопросы

1. Что понимают под автокорреляцией?
2. Какое из условий Гаусса - Маркова нарушает наличие автокорреляции?
3. Как автокорреляция проявляется на графике остатков? Какими признаками обладает график остатков при отсутствии автокорреляции?
4. Какими признаками обладает график остатков в случаях положительной и отрицательной автокорреляции?
5. Что означают термины «автокорреляция первого порядка», «автокорреляция второго порядка» и т.д.?

	1	2	3	4	5	6	7	8	9	10	11	12	13
1													
2													
3													

13. Автокорреляция

6. Как рассчитывается коэффициент автокорреляции? Какими свойствами обладает коэффициент автокорреляции?
7. Каковы основные последствия автокорреляции?
8. Каковы основные предпосылки и ограничения использования статистики Дарбина - Уотсона для обнаружения автокорреляции?
9. Как связаны статистика Дарбина - Уотсона и коэффициент автокорреляции первого порядка?
10. Какие способы устранения автокорреляции Вы знаете?

РЕКОМЕНДУЕМАЯ ЛИТЕРАТУРА

1. *Анатольев, С.* Оформление эконометрических отчетов / С. Анатольев // Квантиль. - 2008. - № 4. - URL: <http://quantile.ru/04/04-SA.pdf>.
2. *Афанасьев, В.Н.* Моделирование и прогнозирование временных рядов / В.Н. Афанасьев, Т.В. Лебедева. - Москва : Финансы и статистика, 2009.
3. *Бабешко, Л.О.* Основы эконометрического моделирования / Л.О. Бабешко. - Москва : КомКнига, 2007.
4. *Бородич, С.А.* Эконометрика / С.А. Бородич. - Минск : Новое знание, 2001.
5. *Галочкин, В.Т.* Эконометрика : учебник и практикум для прикладного бакалавриата / В.Т. Галочкин. - Москва : Юрайт, 2017. (Серия "Бакалавр. Прикладной курс").
6. *Демидова, О.А.* Эконометрика : учебник и практикум для прикладного бакалавриата / О.А. Демидова, Д.И. Малахов. - Москва : Юрайт, 2017. (Серия "Бакалавр. Прикладной курс").
7. *Доугерти, К.* Введение в эконометрику / К. Доугерти. - Москва : Инфра-М, 2009.
8. *Магнус, Я.Р.* Эконометрика / Я.Р. Магнус, П.К. Катышев, А.А. Пересецкий. - Москва : Дело, 2004.
9. Рабочая тетрадь по дисциплине "Эконометрика" : в 3 ч. / Н.А. Зайчикова, Л.А. Игнаткина, Н.П. Перстенева, Е.Г. Репина, Т.Ю. Субеева. - 4-е изд., испр. и доп. - Самара : Изд-во Самар. гос. экон. ун-та, 2018.
10. *Репина, Е.Г.* Вероятностно-статистические методы в научных студенческих исследованиях в области цифровой экономики: практическая значимость - образовательный эффект / Е.Г. Репина // Университетский комплекс как региональный центр образования, науки и культуры : материалы Всерос. науч.-метод. конф. - Оренбург, 2018.
11. *Репина, Е.Г.* Практикум по эконометрике: парная регрессия / Е.Г. Репина, Е.И. Суханова. - 2-е изд., перераб. и доп. - Самара : Изд-во Самар. гос. экон. ун-та, 2018.
12. *Репина, Е.Г.* Сравнительный анализ возможностей использования прикладных эконометрических пакетов в процессе преподавания эконометрики на примере Самарского государственного экономического университета / Е.Г. Репина, Н.А. Зайчикова // Информационно-телекоммуникационные системы и технологии : материалы Всерос. науч.-практ. конф. - Кемерово, 2017.
13. *Сток, Дж.* Введение в эконометрику : пер. с англ. / Дж. Сток, М. Уотсон ; под науч. ред. М.Ю. Турунцевай. - Москва : Дело : РАНХиГС, 2015. (Академический учебник).
14. *Суханова, Е.И.* Начальный курс эконометрики: руководство к решению задач / Е.И. Суханова, Л.К. Ширяева. - Самара : Изд-во Самар. гос. экон. ун-та, 2017.
15. *Цыпин, А.П.* Информационные технологии при проведении экономико-статистических исследований на основе исторических временных рядов / А.П. Цыпин, А.С. Сорокин // Азимут научных исследований: экономика и управление. - 2017. - Т. 6, № 2 (19).
16. *Цыпин, А.П.* Информационное обеспечение процесса построения исторических временных рядов социально-экономических показателей России / А.П. Цыпин, А.Г. Ковалева // Наукоедение : интернет-журн. - 2014. - № 6 (25).

	1	2	3	4	5	6	7	8	9	10	11	12	13
1													
2													
3													

13. Автокорреляция

17. Цыпин, А.П. Статистика в табличном редакторе Microsoft Excel : лабораторный практикум / А.П. Цыпин, Л.Р. Фаизова. - Оренбург : Изд-во Оренбур. гос. ун-та, 2016.

18. Цыпин, А.П. Статистические пакеты программ в социально-экономических исследованиях / А.П. Цыпин, А.С. Сорокин // Азимут научных исследований: экономика и управление. - 2016. - Т. 5, № 4 (17).

19. Эконометрика / под ред. И.И. Елисеевой. - Москва : Финансы и статистика, 2011.

20. Эконометрика : учебник / В.Н. Афанасьев, Т.В. Лебедева, Т.В. Леушина, А.П. Цыпин. - Оренбург : ООО ИПК «Университет», 2012.

ПРИЛОЖЕНИЯ

Приложение 1

DW-СТАТИСТИКА ДАРБИНА - УОТСОНА: $\alpha = 0,05$

(n - объем выборки, m - число объясняющих переменных в уравнении регрессии)

n	$m = 1$		$m = 2$		$m = 3$		$m = 4$	
	d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2
6	0,610	1,400						
7	0,700	1,356	0,467	1,896				
8	0,763	1,332	0,359	1,777	0,368	2,287		
9	0,824	1,320	0,629	1,699	0,435	2,128	0,296	2,388
10	0,879	1,320	0,697	1,641	0,525	2,016	0,376	2,414
11	0,927	1,324	0,658	1,604	0,595	1,928	0,444	2,283
12	0,971	1,331	0,812	1,579	0,658	1,864	0,512	2,177
13	1,010	1,340	0,861	1,562	0,715	1,816	0,574	2,094
14	1,045	1,330	0,905	1,551	0,767	1,779	0,632	2,030
15	1,077	1,361	0,946	1,543	0,814	1,750	0,685	1,977
16	1,106	1,371	0,982	1,539	0,857	1,728	0,734	1,935
17	1,133	1,381	1,015	1,536	0,89	1,710	0,779	1,900
18	1,158	1,391	1,046	1,535	0,933	1,696	0,820	1,872
19	1,180	1,401	1,074	1,536	0,967	1,685	0,859	1,848
20	1,201	1,411	1,100	1,537	0,998	1,676	0,894	1,828
21	1,221	1,420	1,125	1,538	1,026	1,669	0,927	1,812
22	1,239	1,429	1,147	1,541	1,053	1,664	0,958	1,797
23	1,257	1,437	1,168	1,543	1,078	1,660	0,986	1,785
24	1,273	1,446	1,188	1,546	1,101	1,656	1,013	1,775
25	1,288	1,454	1,206	1,550	1,123	1,654	1,038	1,767
26	1,302	1,461	1,224	1,553	1,143	1,652	1,062	1,759
27	1,316	1,469	1,240	1,556	1,162	1,651	1,084	1,753

Окончание прил. 1

n	$m = 1$		$m = 2$		$m = 3$		$m = 4$	
	d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2
29	1,341	1,483	1,270	1,563	1,198	1,650	1,124	1,743
30	1,352	1,489	1,284	1,567	1,214	1,650	1,143	1,739
32	1,373	1,502	1,309	1,574	1,244	1,650	1,177	1,732
33	1,383	1,508	1,321	1,577	1,258	1,651	1,193	1,730
35	1,402	1,519	1,343	1,584	1,283	1,653	1,222	1,726
37	1,419	1,530	1,364	1,590	1,307	1,655	1,249	1,723
39	1,435	1,540	1,382	1,597	1,328	1,658	1,273	1,722
40	1,442	1,544	1,391	1,600	1,338	1,659	1,285	1,721
45	1,475	1,566	1,430	1,615	1,83	1,666	1,336	1,720
50	1,503	1,585	1,462	1,628	1,421	1,674	1,378	1,721
55	1,528	1,601	1,490	1,641	1,452	1,681	1,414	1,724
60	1,549	1,616	1,514	1,652	1,480	1,689	1,444	1,727
65	1,567	1,629	1,536	1,662	1,503	1,696	1,471	1,731
70	1,538	1,641	1,554	1,672	1,525	1,703	1,494	1,735
75	1,598	1,650	1,571	1,680	1,543	1,709	1,515	1,739
80	1,611	1,662	1,586	1,688	1,560	1,715	1,534	1,743
85	1,624	1,671	1,600	1,696	1,575	1,721	1,550	1,747
90	1,635	1,679	1,612	1,703	1,589	1,726	1,566	1,751
95	1,645	1,687	1,623	1,709	1,602	1,732	1,579	1,755
100	1,654	1,694	1,634	1,715	1,613	1,736	1,592	1,758
100	1,654	1,694	1,634	1,715	1,613	1,736	1,592	1,758
150	1,720	1,746	1,706	1,760	1,693	1,774	1,679	1,788
200	1,758	1,778	1,748	1,789	1,738	1,799	1,728	1,810

DW-СТАТИСТИКА ДАРБИНА - УОТСОНА: $\alpha = 0,01$

(n - объем выборки, m - число объясняющих переменных в уравнении регрессии)

n	$m = 1$		$m = 2$		$m = 3$		$m = 4$	
	d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2
6	0,390	1,142						
7	0,433	1,036	0,294	1,676				
8	0,497	1,003	0,343	1,489	0,229	2,102		
9	0,554	0,998	0,408	1,389	0,279	1,873	0,183	2,433
10	0,604	1,001	0,466	1,333	0,340	1,733	0,230	2,193
11	0,633	1,010	0,319	1,297	0,396	1,640	0,286	2,030
12	0,697	1,023	0,369	1,274	0,449	1,373	0,339	1,913
13	0,738	1,038	0,616	1,261	0,499	1,326	0,391	1,826
14	0,776	1,034	0,660	1,234	0,347	1,490	0,441	1,737
15	0,811	1,070	0,700	1,232	0,391	1,464	0,488	1,704
16	0,844	1,086	0,737	1,232	0,633	1,446	0,332	1,663
17	0,874	1,102	0,772	1,233	0,672	1,432	0,374	1,630
18	0,902	1,118	0,803	1,239	0,708	1,422	0,613	1,604
19	0,928	1,132	0,833	1,263	0,742	1,413	0,630	1,384
20	0,932	1,147	0,863	1,271	0,773	1,411	0,683	1,367
21	0,973	1,161	0,890	1,277	0,803	1,408	0,718	1,334
22	0,997	1,174	0,914	1,284	0,831	1,407	0,748	1,343
23	1,018	1,187	0,938	1,291	0,838	1,407	0,777	1,334
24	1,037	1,199	0,960	1,298	0,882	1,407	0,803	1,328
25	1,033	1,211	0,981	1,303	0,906	1,409	0,831	1,323
26	1,072	1,222	1,001	1,312	0,928	1,411	0,833	1,318
27	1,089	1,233	1,019	1,319	0,949	1,413	0,878	1,313

1	2	3	4	5	6	7	8	9	10	11	12	13
2	<i>Приложения</i>											
3												

Окончание прил. 2

n	$m = 1$		$m = 2$		$m = 3$		$m = 4$	
	d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2
29	1,119	1,234	1,034	1,332	0,988	1,418	0,921	1,312
30	1,133	1,263	1,070	1,339	1,006	1,421	0,941	1,311
32	1,160	1,282	1,10	1,332	1,040	1,428	0,979	1,310
33	1,172	1,291	1,110	1,338	1,033	1,432	0,996	1,310
34	1,184	1,299	1,128	1,364	1,070	1,433	1,012	1,311
35	1,193	1,307	1,140	1,370	1,08	1,439	1,028	1,312
37	1,217	1,323	1,163	1,382	1,112	1,446	1,038	1,314
39	1,237	1,337	1,187	1,393	1,137	1,433	1,083	1,317
40	1,246	1,344	1,198	1,398	1,148	1,437	1,098	1,318
45	1,288	1,376	1,243	1,423	1,201	1,474	1,136	1,328
50	1,324	1,403	1,283	1,446	1,243	1,491	1,203	1,338
55	1,336	1,427	1,320	1,466	1,284	1,306	1,247	1,348
60	1,383	1,449	1,330	1,484	1,317	1,320	1,283	1,338
65	1,407	1,468	1,377	1,300	1,346	1,334	1,313	1,368
70	1,429	1,483	1,400	1,313	1,372	1,346	1,343	1,378
75	1,448	1,301	1,422	1,329	1,393	1,337	1,368	1,387
80	1,466	1,313	1,441	1,341	1,416	1,368	1,390	1,393
85	1,482	1,328	1,438	1,333	1,433	1,378	1,411	1,603
90	1,496	1,340	1,474	1,363	1,432	1,387	1,429	1,611
95	1,310	1,332	1,489	1,373	1,468	1,396	1,446	1,618
100	1,322	1,362	1,303	1,383	1,482	1,604	1,462	1,623
150	1,611	1,637	1,398	1,631	1,384	1,663	1,371	1,679
200	1,664	1,684	1,633	1,693	1,643	1,704	1,633	1,713

Учебное издание

**Репина Евгения Геннадьевна
Цыпин Александр Павлович
Зайчикова Надежда Анатольевна
Ширнаева Светлана Юрьевна**

**ЭКОНОМЕТРИКА
В ТАБЛИЧНОМ РЕДАКТОРЕ
MS EXCEL**

Практикум

Издательская группа
О.В. Егорова, Г.И. Конева, Л.И. Трофимова

Подписано к изданию 26.11.2019. Печ. л. 11,12.
ФГБОУ ВО "Самарский государственный экономический университет".
443090, Самара, ул. Советской Армии, 141.