

Новокузнецкий институт (филиал)
Федерального государственного бюджетного
образовательного учреждения высшего образования
«Кемеровский государственный университет»

М. А. Кречетова

ЭКОНОМЕТРИКА

Практикум

по дисциплине «Эконометрика»
для студентов, обучающихся по направлению
38.03.01 – Экономика (все профили)

Новокузнецк
2020

УДК 330.4 (075.8)

ББК 65в6я73

Печатается по решению методического совета Новокузнецкого института (филиала) Федерального государственного бюджетного образовательного учреждения высшего образования «Кемеровский государственный университет»

Рецензенты:

- заместитель главы Мысковского городского округа по экономике и промышленности, кандидат экономических наук С. Н. Часовников;

- кандидат экономических наук, доцент кафедры экономики, учета и финансовых рынков. АБ «Кузнецкбизнесбанк» Сибирского государственного индустриального университета в г. Новокузнецке В.В. Шипунова

Эконометрика : практикум для студентов, обучающихся по направлению 38.03.01 Экономика / М.А. Кречетова; М-во образования и науки Рос. Федерации, Новокузнец. ин-т (фил.) Кемеров. гос. ун-та. – Новокузнецк: НФИ КемГУ, 2020. – 91 с.

Данный практикум является учебно-практическим изданием по основным темам дисциплины «Эконометрика» для студентов направления 38.03.01 Экономика (всех профилей). Оно обеспечивает получение практических навыков по построению эконометрических моделей на основе статистических данных.

В практикуме приводятся необходимые теоретические сведения и формулы по основным темам курса, разобраны решения типовых примеров, представлены задания для аудиторных практических занятий и самостоятельных лабораторных работ. Структура практикума способствует организации аудиторной и самостоятельной работы и предназначено для студентов всех форм обучения.

УДК 330.4 (075.8)

ББК 65в6я73

© Кречетова М.А., 2020

© Новокузнецкий институт (филиал) Федерального государственного бюджетного образовательного учреждения высшего образования «Кемеровский государственный университет», 2020

ПРЕДИСЛОВИЕ	5
Тема 1 ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ	7
1.1 Краткий теоретический материал	7
1.2 Типовые примеры и образцы их решения	9
1.3 Практические задания	11
1.4 Семестровое задание 1 «Проверка статистических гипотез»	11
Тема 2 ПАРНАЯ ЛИНЕЙНАЯ РЕГРЕССИЯ	12
2.1 Краткий теоретический материал	12
2.2 Типовые примеры и образцы их решения	15
2.3 Практические задания	18
2.4 Семестровая работа 2 «Парная линейная регрессия»	19
Тема 3 ПАРНАЯ НЕЛИНЕЙНАЯ РЕГРЕССИЯ	22
3.1 Краткий теоретический материал	22
3.2 Типовые примеры и образцы их решения	26
3.3 Практические задания	32
3.4 Семестровая работа 3 «Нелинейная регрессия»	33
Тема 4 МНОЖЕСТВЕННАЯ РЕГРЕССИЯ	35
4.1 Краткий теоретический материал	35
4.2 Типовые примеры и образцы их решения	40
4.3 Задания на практику	47
4.4 Семестровая работа 4 «Множественная регрессия»	49
Тема 5 МОДЕЛИРОВАНИЕ ВРЕМЕННЫХ РЯДОВ	52
5.1 Краткий теоретический материал	52
5.2 Типовые примеры и образцы их решения	57
5.3 Задания на практику	63
5.4.1 Семестровая работа 5.1 «Динамика и адаптивные методы»	65
5.4.2 Семестровая работа 5.2 «Моделирование сезонных колебаний»	65
Тема 6 СИСТЕМЫ ОДНОВРЕМЕННЫХ УРАВНЕНИЙ	67
6.1 Краткий теоретический материал	67
6.2. Типовые примеры и образцы их решения	71
6.3 Задания на практику	75
6.4 Семестровая работа 6 «Системы одновременных уравнений»	76
СПИСОК ЛИТЕРАТУРЫ	79
Приложения 1-4 СТАТИСТИКО-МАТЕМАТИЧЕСКИЕ ТАБЛИЦЫ	81
Приложение 1. Таблица значений F – критерия Фишера при уровне значимости $\alpha = 0,05$	81
Приложение 2. Критические значения t – критерия Стьюдента при уровне значимости 0,10, 0,05, 0,01 (двухсторонний)	82
Приложение 3. Значения статистик Дарбина-Уотсона $d_L d_U$ при 5%-ном уровне	83

значимости	83
Приложение 4. Критические значения RS критерия	84
Приложение 5. Решение практических заданий в Excel	85

ПРЕДИСЛОВИЕ

Практикум по дисциплине «Эконометрика» посвящен вопросам получения умений и навыков построения эконометрических моделей по реальным данным из различных сфер экономики. Он ориентирован на студентов – бакалавров, изучающих дисциплину «Эконометрика», направления 38.03.01 « Экономика», все профили. Практикум направлен на формирование у обучающихся общепрофессиональной компетенции ОПК-3, а также профессиональной компетенции ПК-4.

Эконометрика, наряду с микро и макроэкономикой является базовой дисциплиной для подготовки экономистов любого уровня – от бакалавров до магистров. Дисциплина входит в базовую часть образовательной программы (Б1.Б.16) подготовки студентов факультета информатики, математики и экономики профилей «Финансы и кредит» и «Бухгалтерский учет, анализ и аудит».

Изучение дисциплины «Эконометрика» направлено на формирование эконометрического мышления студентов, развитие их способности моделировать различные экономические явления с помощью эконометрических моделей и прогнозировать развитие этих явлений в будущем по построенным моделям. Студенты должны также научиться давать статистическую оценку значимости таких искажающих эффектов, как гетероскедастичность остатков зависимой переменной, мультиколлинеарность объясняющих переменных, автокорреляция.

В результате изучения дисциплины «Эконометрика» студенты должны:

Знать:

- основные этапы проведения эконометрического исследования,
- основные типы эконометрических данных и основные эконометрические модели,

Уметь:

- Формулировать задачу в пригодном для эконометрического исследования виде,
- Проверять статистические гипотезы,
- Оценивать параметры эконометрических моделей,
- Строить точечные и интервальные прогнозы.

Владеть:

- Навыками построения и диагностики построенных эконометрических моделей,
- Навыками интерпретации основных результатов построения эконометрических моделей.

Разработке современных учебников и практикумов посвятили свое внимание многие авторы, такие как Кремер Н.Ш., Галочкин В.Т., Валентинов В.А., Демидова О.А., , Бородич С.А., , Новиков А.И., Басовский Л.Е. и другие. В учебниках и учебных пособиях данных авторов значительное внимание уделяется теории эконометрических исследований, меньше внимания уделяется представлению решений практических задач, иллюстрирующих теоретические положения. Также в представленных заданиях для самостоятельного решения отсутствует вариативность, они представлены в одном варианте. Особенностью данного практикума является предложение для решения самостоятельных заданий на практике в одном варианте, при этом задания приведены различной степени сложности. Для самостоятельной проработки полученных на практике умений и закрепления навыков приведены задания для домашней семестровой работы в 10 вариантах с решением данных заданий в пакете Microsoft Excel. Выбор данного пакета для построения эконометрических моделей обоснован следующими его преимуществами:

1. Распространенность. В настоящее время на подавляющем большинстве компьютеров стоит это программное обеспечение.

2. Кросс-платформенность. Помимо Excel для операционной системы (ОС) Windows существуют его почти полные аналоги для Mac и Unix / Linux (для последних операционных систем самым популярным аналогом является пакет LibreOffice, в котором бесплатный аналог Excel называется Calc).

3. Интуитивность и простота. Не нужно обладать особыми знаниями в программировании, что бы уметь работать с этим пакетом.

4. Стабильность работы и малое количество ошибок в коде (по сравнению со свободно распространяемым программным обеспечением).

5. Возможность программирования и создания пользовательских функций.

Данный практикум преследует цель – помочь студентам получить и закрепить навыки построения эконометрических моделей в различных экономических направлениях и умения интерпретировать полученные результаты эконометрического анализа. Это даст возможность развить у студентов эконометрическое мышление, необходимое им для профессиональной практической деятельности.

Практикум охватывает основные темы курса и построен по следующей структуре:

- краткие теоретические сведения, включающие основные понятия, определения, формулы;
- типовые примеры и образцы их решения;
- Практические задания для аудиторных практических занятий;
- Семестровые работы для самостоятельной работы студентов.

Использование данного практикума позволит более рационально и эффективно организовать аудиторную и самостоятельную работу студентов по изучению дисциплины «Эконометрика».

При использовании практикума студентам необходимо:

- изучить теоретический материал по теме по базовым учебникам и представленным материалам,
- проработать решение типовых примеров,
- решить задания, выносимые на аудиторные практические занятия,
- выполнить дома семестровые задания по своему варианту.

Номер варианта выбирается для дневного отделения по последней цифре в списке группы, для заочного отделения по последней цифре номера зачетки. Все расчеты выполнять в Microsoft Excel и оформлять по требованиям к оформлению.

В приложениях 1-4 приведены статистические таблицы, необходимые для расчетов, а в приложении 5 примеры использования *Пакета анализа* Microsoft Excel для проведения расчетов.

Тема 1 ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ

1.1 Краткий теоретический материал

При статистическом анализе важно знать закон распределения генеральной совокупности. Если закон неизвестен, но имеются основания предположить, что он имеет определенный вид, то говорят: генеральная совокупность распределена по закону А. Если закон известен, но неизвестны его параметры и, например, есть основания предполагать, что $\theta = \theta_0$, то выдвигают гипотезу: $\theta = \theta_0$ (здесь речь идет о гипотезе: равенстве θ_0 параметру закона).

Возможны и другие гипотезы:

1. о равенстве параметров двух или нескольких распределений;
2. о независимости выборок и др.

Наряду с выдвинутой гипотезой рассматривают и противоположную ей. Если выдвинутая гипотеза будет отвергнута, то имеет место противоположная ей. По этой причине гипотезы делят на:

1. нулевую. Нулевой (основной) называют выдвинутую гипотезу H_0 ;
2. конкурирующую. Конкурирующей (альтернативной) называют гипотезу H_1 , которая противоречит основной.

Кроме того гипотез делят на простые и сложные.

Простой называют гипотезу, содержащую только одно предположение.

Сложной называют гипотезу, содержащую несколько простых гипотез. Например гипотеза $H: \lambda > 5$ содержит бесчисленное множество простых гипотез $H_i: \lambda = b_i$, где b_i - любое число большее 5.

Ошибки первого и второго рода

Выдвинутая гипотеза может быть правильной или неправильной. Поэтому необходима ее проверка. Проверку осуществляют статистическими методами. При проверке возможны ошибки двух родов.

Ошибка первого рода состоит в том, что будет отвергнута правильная гипотеза.

Ошибка второго рода состоит в том, что будет принята неправильная гипотеза.

Критерий используется для проверки гипотезы о равенстве средних значений двух выборок, взятых из нормально распределенных совокупностей.

Пусть заданы две генеральные совокупности x и y , имеющие нормальное распределение, из них взяты выборки $\{x_i\}_{i=1}^{n_1}$ и $\{y_i\}_{i=1}^{n_2}$, т.е. n_1 и n_2 - объемы первой и второй выборки соответственно. Выдвигается гипотеза H_0 , что средние значения выборок равны (альтернативная гипотеза H_1 - средние значения не равны).

Значение $t_{набл}$ вычисляют по формуле:

$$t_{набл} = \frac{\bar{x} - \bar{y}}{S}, \quad (1)$$

где $\bar{x}, \bar{y}$ — средние арифметические выборок $\{x_i\}_{i=1}^{n_1}$ и $\{y_i\}_{i=1}^{n_2}$;

S - стандартная ошибка разности средних значений.

Число степеней свободы вычисляют по формуле:

$$k = n_1 + n_2 - 2 \quad (2)$$

Из таблиц для заданного уровня значимости α и числа степеней свободы k определяют $t_{\text{крит}}$ (критическое значение).

Если $|t_{\text{набл}}| < t_{\text{крит}}$, то гипотеза H_0 принимается, в противном случае принимается альтернативная гипотеза.

Стандартная ошибка разности средних значений S вычисляется различными способами в зависимости от поставленной задачи:

- сравнение двух выборок;
- сравнение двух зависимых выборок;
- сравнение более двух независимых выборок.

СЛУЧАЙ ДВУХ НЕЗАВИСИМЫХ ВЫБОРОК

Требуется сравнить средние значения двух независимых выборок. Здесь возможны два варианта:

1. Дисперсии выборок равны.
2. Дисперсии выборок не равны.

Рассмотрим первый вариант (дисперсии выборок равны). В этом случае значение S вычисляется по формуле

$$S = \sqrt{\frac{\sum (x_i - \bar{x})^2 + \sum (y_i - \bar{y})^2}{n_1 + n_2 - 2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}, \quad (3)$$

где n_1 и n_2 - объемы первой и второй выборки; $\bar{x}$ и $\bar{y}$ — средние арифметические выборок.

СЛУЧАЙ ДВУХ ЗАВИСИМЫХ ВЫБОРОК

Требуется сравнить средние значения двух зависимых выборок, полученных из нормально распределенной совокупности. Объем выборок одинаков.

В этом случае значения $t_{\text{набл}}$ вычисляют по формуле (4.11), которая в данном случае примет вид:

$$t_{\text{набл}} = \frac{\bar{d}}{S_d}, \quad (4)$$

где $\bar{d} = x_i - y_i$ — разности между соответствующими значениями переменной x и переменной y ,

$\bar{d}$ - среднее значение этих разностей;

S_d - стандартная ошибка разности средних значений.

S_d вычисляется по формуле:

$$S_d = \sqrt{\frac{\sum d_i^2 - \frac{[\sum d_i]^2}{n}}{n \cdot (n - 1)}} \quad (5)$$

Число степеней свободы k определяется по формуле $k = n - 1$, где n - объем выборки. Рассмотрим пример использования t -критерия Стьюдента для связанных и, очевидно, равных по численности выборок.

1.2 Типовые примеры и образцы их решения

Пример 1. В двух группах учащихся — экспериментальной и контрольной — применялись две различные методики обучения: экспериментальная и традиционная. После завершения обучения был проведен тест и получены следующие результаты по учебному предмету (тестовые баллы; см. табл. 1.1).

Таблица 1.1 Результаты эксперимента

Первая группа (экспериментальная), $N_1=11$ человек										
12	14	13	16	11	9	13	15	15	18	14
Вторая группа (контрольная), $N_2=9$ человек										
13	9	11	10	7	6	8	10	11		

Имеет ли экспериментальный метод обучения преимущество по сравнению с традиционным?

Решение:

Для выявления преимущества экспериментального метода обучения по сравнению с традиционным можно проверить совпадения средних оценок в двух группах. Если средние отличаются незначительно, то преимущества нет, в противном случае преимущество есть. Гипотеза H_0 : средние значения выборок равны, альтернативная гипотеза H_1 : средние значения не равны.

Общее количество членов выборки: $n_1=11$, $n_2=9$; средние значения: $\bar{x}=13,636$; $\bar{y}=9,444$. По формуле (3) находим стандартную ошибку разности средних значений:

$$S = \sqrt{\frac{\sum (x_i - \bar{x})^2 + \sum (y_i - \bar{y})^2}{n_1 + n_2 - 2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)} = \sqrt{\frac{60.545 + 38.222}{11 + 9 - 2} \left(\frac{1}{11} + \frac{1}{9} \right)} = 1.053$$

Вычисляем значение $t_{набл}$

$$t_{набл} = \frac{\bar{x} - \bar{y}}{S} = \frac{13.636 - 9.444}{1.053} = 3.981$$

Вычислим табличное значение $t_{крит}$ с помощью таблицы распределения Стьюдента (приложение 2). Для этого определим число степеней свободы по формуле $k = n_1 + n_2 - 2 = 11 + 9 - 2 = 18$,

и с учетом уровня значимости $\alpha = 5\%$ (или $\alpha = 0,05$) получим $t_{крит} = 2.100$. Так как $|t_{набл}| > t_{крит}$, гипотеза H_0 отклоняется, принимается гипотеза H_1 . Из этого следует вывод о преимуществе экспериментального обучения.

Решение задачи с помощью Пакета анализа Excel приведено в приложении 5.

Пример 2. Исследовали влияния тренинга на частоту сердечных сокращений (ЧСС) у группы пациентов, страдающих тахикардией. В первом случае пациенты принимали традиционные лекарства, тренинг не проводился, величина ЧСС обозначена через X (рис.1). В другом случае эти же пациенты принимали традиционные лекарства после проведения сеанса

тренинга, величина ЧСС обозначена через Y . Требуется оценить эффективность проведения сеанса тренинга на ЧСС.

Решение

В столбце В (рис.1.1) содержатся значения $\{x_i\}_{i=1}^n$ ЧСС после приема лекарств у пациентов без тренинга, в столбце С $\{y_i\}_{i=1}^n$ ЧСС при приеме лекарств после сеанса тренинга ($n=10$).

Поскольку группа пациентов одна и та же, в данном примере применима методика для связанных и равных по численности выборок. Вначале произведем расчет $\bar{d}$ (ячейка D20, рис. 1):

$$\bar{d} = 2.1$$

Затем по формуле (5), получим:

$$S_d = \sqrt{\frac{\sum d_i^2 - \frac{[\sum d_i]^2}{n}}{n \cdot (n-1)}} = \sqrt{\frac{103 - \frac{[21]^2}{10}}{10 \cdot (10-1)}} = 0.809.$$

Далее следует применить формулу (5). Получим:

	A	B	C	D	E	F
1	Критерий Стьюдента (t-критерий)					
3	Сравнение двух зависимых выборок					
4	Гипотеза H_0 : средние значения выборок равны					
5	альтернативная гипотеза H_1: среднее Y меньше среднего X					
6						
7	Пациенты	x	y	d	(d)^2	
8	Иванов	80	76	4	16	
9	Новиков	92	92	0	0	
10	Сидоров	80	79	1	1	
11	Пирогов	89	85	4	16	
12	Агапов	87	80	7	49	
13	Суворов	86	86	0	0	
14	Рыжиков	85	85	0	0	
15	Серов	89	90	-1	1	
16	Топоров	90	86	4	16	
17	Быстров	92	90	2	4	
18	Сумма	870	849	21	103	
19	Объем выборки	10	10	10		
20	Ср.значение	87	84.9	2.1		
21						
22	S=	0.808977	k=	9		
23	tнабл=	2.59587	alfa=	0.05		
24						
25	 t_{набл} > t_{крит}		tкрит=	1.83		
26						
27	Гипотеза H_0 отклоняется, принимаем гипотезу H_1.					

Рис.1 . Проверка гипотезы о совпадении двух выборочных средних в случае двух зависимых выборок

$$t_{набл} = \frac{\bar{d}}{S_d} = \frac{2.1}{0.809} = 2.596.$$

Число степеней свободы: $k = n - 1 = 10 - 1 = 9$. таблицы распределения Стьюдента (приложение 2), $t_{крит} = 1.83$. Так как $t_{набл} = 2.596$, то возможно принять альтернативную гипотезу (H_1) о достоверном уменьшении ЧСС у пациентов группы Y . Отсюда можно сделать вывод об эффективности тренинга перед приемом лекарств.

В терминах проверки статистических гипотез полученный результат будет звучать так: на 5% уровне гипотеза H_0 отклоняется и принимается гипотеза H_1 .

1.3 Практические задания

1. Имеются данные об объеме продаж товара по фирме до и после проведения ею рекламной компании. С помощью критерия Стьюдента проверить гипотезу о равенстве средних в данных совокупностях (зависимые выборки) – повлияла ли рекламная компания на изменение объема продаж?

Таблица 1.2 Объем продаж до и после рекламной компании

До	150	200	180	175	168	190	149	165	173	144	130	207	178
после	165	190	189	170	177	159	144	163	198	171	182	164	153
До	150	160	185	176	138	140							
после	157	167	154	172	155	129							

2. Известны данные о самооценке стрессоустойчивости по учителям и менеджерам продаж. С помощью критерия Стьюдента проверить гипотезу о равенстве средних в данных совокупностях (независимые выборки) – влияет ли на профессию на уровень стрессоустойчивости?

Таблица 1.3 Оценка уровня стрессоустойчивости по профессиям

Учителя	23	17	19	18	22	18	19	17	20	21	24	19	21	20	22	23
менеджеры	25	14	17	23	24	22	24	20	21	23	19	23	21	20	24	26
Учителя	18	16	17	25	21	20	15	16	21	20	19	18	17	16	17	20
менеджеры	21	19	24	23	21	19	25	22	23	20	19	21	24	25	22	21

1.4 Семестровое работа 1 «Проверка статистических гипотез»

Имеется выборка прибыли коммерческой фирмы за 14 месяцев до (x_i) и после (y_i) проведения новой экономической политики. На уровне значимости $\alpha = 0,05$ проверить по критерию Стьюдента гипотезу о том, привело ли введение новой экономической политики в среднем к увеличению производительности.

Таблица 1.4 Задания по вариантам

Вариант	Выборка														
1.	x	21	32	26	34	25	33	31	32	28	33	28	34	27	26
	y	27	26	35	32	34	33	32	19	25	31	25	30	30	28
2.	x	28	28	29	27	28	27	29	29	30	30	29	28	29	29
	y	31	32	32	29	30	31	30	30	29	29	30	30	30	31
3.	x	26	34	28	33	33	26	21	23	31	23	27	24	24	29
	y	35	31	40	29	40	31	29	31	36	33	35	37	36	36
4.	x	42	32	46	39	39	37	35	38	35	42	39	40	38	47
	y	50	39	52	49	52	49	45	37	49	40	45	39	44	24
5.	x	42	32	46	39	39	37	35	38	35	42	39	40	38	47
	y	35	36	39	39	41	48	33	41	35	38	43	36	36	39
6.	x	59	63	54	61	57	52	54	61	63	61	56	55	55	55
	y	52	71	54	53	45	59	48	58	71	61	59	65	74	63
7.	x	46	51	48	45	53	51	46	53	48	53	49	58	56	49
	y	47	54	45	46	55	51	46	56	53	51	49	50	56	56
8.	x	52	51	48	52	54	50	51	51	52	52	53	56	51	50
	y	44	47	57	54	39	65	46	51	58	46	62	52	65	47

9.	x	73	76	77	76	76	75	74	72	75	79	76	78	71	75
	y	70	71	83	76	79	71	74	66	80	81	78	69	73	85
10.	x	21	20	20	17	21	22	23	19	25	21	20	17	21	22
	y	29	21	21	25	16	23	22	27	31	27	22	32	27	22

Тема 2 ПАРНАЯ ЛИНЕЙНАЯ РЕГРЕССИЯ

2.1 Краткий теоретический материал

Парная регрессия - уравнение связи двух переменных y и x :

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

где y - зависимая переменная (результативный признак);

x - независимая, объясняющая переменная (признак-фактор).

Различают *линейные* и *нелинейные* регрессии.

Линейная регрессия: $y = a + b \cdot x + \varepsilon$.

Нелинейные регрессии делятся на два класса: регрессии, нелинейные относительно включенных в анализ объясняющих переменных, но линейные по оцениваемым параметрам, и регрессии, нелинейные по оцениваемым параметрам.

Регрессии, *нелинейные по объясняющим переменным*:

- полиномы разных степеней $y = a + b_1 \cdot x + b_2 \cdot x^2 + b_3 \cdot x^3 + \varepsilon$;
- равнобочная гиперболa $y = a + \frac{b}{x} + \varepsilon$.

Регрессии, *нелинейные по оцениваемым параметрам*:

- степенная $y = a \cdot x^b \cdot \varepsilon$;
- показательная $y = a \cdot b^x \cdot \varepsilon$;
- экспоненциальная $y = e^{a+b \cdot x} \cdot \varepsilon$

Построение уравнения линейной регрессии сводится к оценке ее параметров. Для оценки параметров регрессий, линейных по параметрам, используют *метод наименьших квадратов (МНК)*. МНК позволяет получить такие оценки параметров, при которых сумма квадратов отклонений фактических значений результативного признака y от теоретических $\hat{y}_x$ минимальна, т.е.

$$\sum (y - \hat{y}_x)^2 \longrightarrow \min$$

Для линейных и нелинейных уравнений, приводимых к линейным, решается следующая система относительно a и b :

$$\begin{cases} na + b \sum x = \sum y, \\ a \sum x + b \sum x^2 = \sum xy. \end{cases}$$

Можно воспользоваться готовыми формулами, которые вытекают из этой системы:

$$a = \bar{y} - b \cdot \bar{x}, \quad b = \frac{\text{cov}(x, y)}{\sigma_x^2} = \frac{\overline{x \cdot y} - \bar{x} \cdot \bar{y}}{\overline{x^2} - \bar{x}^2}.$$

Тесноту связи изучаемых явлений оценивает *линейный коэффициент парной корреляции* r_{xy} для линейной регрессии ($-1 \leq r_{xy} \leq 1$):

$$r_{xy} = b \cdot \frac{\sigma_x}{\sigma_y} = \frac{\text{cov}(x, y)}{\sigma_x \cdot \sigma_y} = \frac{\overline{x \cdot y} - \bar{x} \cdot \bar{y}}{\sigma_x^2 - \bar{x}^2},$$

и индекс корреляции p_{xy} - для нелинейной регрессии ($0 \leq p_{xy} \leq 1$):

$$p_{xy} = \sqrt{1 - \frac{\sigma_{ост}^2}{\sigma_y^2}} = \sqrt{1 - \frac{\sum (y - \hat{y}_x)^2}{\sum (y - \bar{y})^2}}.$$

Оценку качества построенной модели даст коэффициент (индекс) детерминации, а также средняя ошибка аппроксимации.

Средняя ошибка аппроксимации - среднее отклонение расчетных значений от фактических:

$$\bar{A} = \frac{1}{n} \cdot \sum \left| \frac{y - \hat{y}}{y} \right| \cdot 100\%.$$

Допустимый предел значений $\bar{A}$ - не более 8 - 10%.

Средний коэффициент эластичности $\bar{\epsilon}$ показывает, на сколько процентов в среднем по совокупности изменится результат y от своей средней величины при изменении фактора x на 1% от своего среднего значения:

$$\bar{\epsilon} = f'(x) \cdot \frac{\bar{x}}{\bar{y}}.$$

Задача *дисперсионного анализа* состоит в анализе дисперсии зависимой переменной:

$$\sum (y - \bar{y})^2 = \sum (\hat{y}_x - \bar{y})^2 + \sum (y - \hat{y}_x)^2,$$

где $\sum (y - \bar{y})^2$ - общая сумма квадратов отклонений;

$\sum (\hat{y}_x - \bar{y})^2$ - сумма квадратов отклонений, обусловленная регрессией («объясненная» или «факторная»);

$\sum (y - \hat{y}_x)^2$, - остаточная сумма квадратов отклонений.

Долю дисперсии, объясняемую регрессией, в общей дисперсии результативного признака y характеризует *коэффициент (индекс) детерминации* R^2 :

$$R^2 = \frac{\sum (\hat{y}_x - \bar{y})^2}{\sum (y - \bar{y})^2}.$$

Коэффициент детерминации - квадрат коэффициента или индекса корреляции.

F-тест - оценивание качества уравнения регрессии - состоит в проверке гипотезы H_0 о *статистической незначимости уравнения регрессии и показателя тесноты связи*. Для этого выполняется сравнение фактического $F_{\text{факт}}$ и критического (табличного) $F_{\text{табл}}$ значений *F-критерия Фишера*. $F_{\text{факт}}$ определяется из соотношения значений факторной и остаточной дисперсий, рассчитанных на одну степень свободы:

$$F_{\text{факт}} = \frac{\sum (\hat{y} - \bar{y})^2 / m}{\sum (y - \hat{y})^2 / (n - m - 1)} = \frac{r_{xy}^2}{1 - r_{xy}^2} \cdot (n - 2),$$

n - число единиц совокупности;

m — число параметров при переменных x .

$F_{\text{табл}}$ - это максимально возможное значение критерия под влиянием случайных факторов при данных степенях свободы и уровне значимости α . Уровень значимости α - вероятность отвергнуть правильную гипотезу при условии, что она верна. Обычно α принимается равной 0,05 или 0,01.

Если $F_{\text{табл}} < F_{\text{факт}}$, то H_0 - гипотеза о случайной природе оцениваемых характеристик отклоняется и признается их статистическая значимость и надежность. Если $F_{\text{табл}} > F_{\text{факт}}$, то гипотеза H_0 не отклоняется и признается статистическая незначимость, ненадежность уравнения регрессии.

Для оценки *статистической значимости коэффициентов регрессии и корреляции* рассчитываются *t-критерий Стьюдента* и *доверительные интервалы* каждого из показателей. Выдвигается гипотеза H_0 о случайной природе показателей, т.е. о незначимом их отличии от нуля. Оценка значимости коэффициентов регрессии и корреляции с помощью *f-критерия Стьюдента* проводится путем сопоставления их значений с величиной случайной ошибки:

$$t_b = \frac{b}{m_b}; t_a = \frac{a}{m_a}; t_r = \frac{r}{m_r}.$$

Случайные ошибки параметров линейной регрессии и коэффициента корреляции определяются по формулам:

$$m_b = \sqrt{\frac{\sum(y - \hat{y}_x)^2 / (n-2)}{\sum(x - \bar{x})^2}} = \sqrt{\frac{S_{\text{ост}}^2}{\sum(x - \bar{x})^2}} = \frac{S_{\text{ост}}}{\sigma_x \cdot \sqrt{n}};$$

$$m_a = \sqrt{\frac{\sum(y - \hat{y}_x)^2}{(n-2)} \cdot \frac{\sum x^2}{n \cdot \sum(x - \bar{x})^2}} = \sqrt{S_{\text{ост}}^2 \cdot \frac{\sum x^2}{n^2 \cdot \sigma_x^2}} = S_{\text{ост}} \frac{\sqrt{\sum x^2}}{n \cdot \sigma_x};$$

$$m_{r_{xy}} = \sqrt{\frac{1 - r_{xy}^2}{n-2}}$$

Сравнивая фактическое и критическое (табличное) значения *t-статистики* – $t_{\text{табл}}$ и $t_{\text{факт}}$ – принимаем или отвергаем гипотезу H_0 .

Связь между *F-критерием Фишера* и *t-статистикой Стьюдента* выражается равенством

$$t_r^2 = t_b^2 = \sqrt{F}.$$

Если $t_{\text{табл}} < t_{\text{факт}}$, то H_0 отклоняется, т.е. a , b и r_{xy} не случайно отличаются от нуля и сформировались под влиянием систематически действующего фактора x . Если $t_{\text{табл}} > t_{\text{факт}}$, то гипотеза H_0 не отклоняется и признается случайная природа формирования a , b или r_{xy} .

Для расчета доверительного интервала определяем *предельную ошибку* Δ для каждого показателя:

$$\Delta_a = t_{\text{табл}} \cdot m_a, \quad \Delta_b = t_{\text{табл}} \cdot m_b.$$

Формулы для расчета *доверительных интервалов* имеют следующий вид:

$$\gamma_a = a \pm \Delta_a; \quad \gamma_{a_{\min}} = a - \Delta_a; \quad \gamma_{a_{\max}} = a + \Delta_a;$$

$$\gamma_b = b \pm \Delta_b; \quad \gamma_{b_{\min}} = b - \Delta_b; \quad \gamma_{b_{\max}} = b + \Delta_b;$$

Если в границы доверительного интервала попадает ноль, т.е. нижняя граница отрицательна, а верхняя положительна, то оцениваемый параметр принимается нулевым, так как он не может одновременно принимать и положительное, и отрицательное значения.

Прогнозное значение y_p определяется путем подстановки в уравнение регрессии

$\hat{y}_x = a + b \cdot x$ соответствующего (прогнозного) значения x_p . Вычисляется *средняя стан-*

дартная ошибка прогноза $m_{\hat{y}_p}$: $m_{\hat{y}_p} = \sigma_{\text{ост}} \cdot \sqrt{1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{\sum(x - \bar{x})^2}},$

$$\text{где } \sigma_{ост} = \sqrt{\frac{\sum (y - \hat{y})^2}{n - m - 1}};$$

и строится *доверительный интервал прогноза*:

$$\gamma_{\hat{y}_p} = \hat{y}_p \pm \Delta_{\hat{y}_p}; \quad \gamma_{\hat{y}_{p_{\min}}} = \hat{y}_p - \Delta_{\hat{y}_p}; \quad \gamma_{\hat{y}_{p_{\max}}} = \hat{y}_p + \Delta_{\hat{y}_p}, \quad \text{где } \Delta_{\hat{y}_p} = t_{табл} \cdot m_{\hat{y}_p}.$$

2.2 Типовые примеры и образцы их решения

По данным проведенного опроса восьми групп семей известны данные связи расходов населения на продукты питания с уровнем доходов семьи.

Таблица 2.1 Данные по доходам и расходам семей

Расходы на продукты питания, тыс. руб., Y	0,9	1,2	1,8	2,2	2,6	2,9	3,3	3,8
Доходы семьи, тыс. руб., X	1,2	3,1	5,3	7,4	9,6	11,8	14,5	18,7

- 1) Построить график – **поле корреляции** и сделать вывод о виде модели.
- 2) Рассчитать параметры парной регрессии Y от X.
- 3) Рассчитать линейный коэффициент парной корреляции, коэффициент детерминации и среднюю ошибку аппроксимации. Сделать выводы.
- 4) Оценить статистическую значимость параметров регрессии и корреляции, а также всего уравнения с помощью F-критерия Фишера и t-критерия Стьюдента. Сделать выводы.
- 5) На одном графике построить исходные данные и теоретическую прямую.
- 6) Выполнить прогноз Y* при прогнозном значении X*, равном 110% от максимального значения. Оценить точность прогноза, рассчитав ошибку прогноза и его доверительный интервал.

Решение: Предположим, что связь между доходами семьи и расходами на продукты питания линейная. Для подтверждения нашего предположения построим поле корреляции.

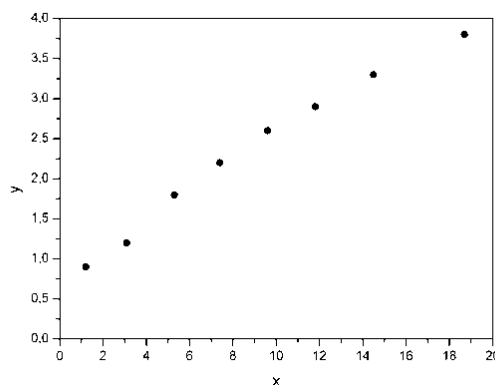


Рис. 2.1 - Поле корреляции

По графику видно, что точки выстраиваются в некоторую прямую линию. Для удобства дальнейших вычислений составим таблицу 2.2.

Таблица 2.2 Рабочая таблица

	x	y	$x \cdot y$	x^2	y^2	y_x	$y - y_x$	$(y - y_x)^2$	$A_i, \%$
1	2	3	4	5	6	7	8	9	10
1	1,2	0,9	1,08	1,44	0,81	1,038	-0,138	0,0190	15,33
2	3,1	1,2	3,72	9,61	1,44	1,357	-0,157	0,0246	13,08
3	5,3	1,8	9,54	28,09	3,24	1,726	0,074	0,0055	4,11
4	7,4	2,2	16,28	54,76	4,84	2,079	0,121	0,0146	5,50
5	9,6	2,6	24,96	92,16	6,76	2,449	0,151	0,0228	5,81
6	11,8	2,9	34,22	139,24	8,41	2,818	0,082	0,0067	2,83
7	14,5	3,3	47,85	210,25	10,89	3,272	0,028	0,0008	0,85
8	18,7	3,8	71,06	349,69	14,44	3,978	-0,178	0,0317	4,68
Итого	71,6	18,7	208,71	885,24	50,83	18,717	-0,017	0,1257	52,19
Среднее значение	8,95	2,34	26,09	110,66	6,35	2,34	–	0,0157	6,52
σ	5,53	0,935	–	–	–	–	–	–	–
σ^2	30,56	0,874	–	–	–	–	–	–	–

Рассчитаем параметры линейного уравнения парной регрессии $y_x = a + b \cdot x$. Для этого воспользуемся формулами (1.5):

$$b = \frac{\text{cov}(x, y)}{\sigma_x^2} = \frac{\overline{x \cdot y} - \bar{x} \cdot \bar{y}}{\bar{x^2} - \bar{x}^2} = \frac{26,09 - 8,95 \cdot 2,34}{30,56} = 0,168;$$

$$a = \bar{y} - b \cdot \bar{x} = 2,34 - 0,168 \cdot 8,95 = 0,836.$$

Получили уравнение: $y_x = 0,836 + 0,168 \cdot x$. Т.е. с увеличением дохода семьи на 1000 руб. расходы на питание увеличиваются на 168 руб.

Как было указано выше, уравнение линейной регрессии всегда дополняется показателем тесноты связи – линейным коэффициентом корреляции r_{xy} :

$$r_{xy} = b \cdot \frac{\sigma_x}{\sigma_y} = 0,168 \cdot \frac{5,53}{0,935} = 0,994.$$

Близость коэффициента корреляции к 1 указывает на тесную линейную связь между признаками.

Коэффициент детерминации $r_{xy}^2 = 0,987$ (примерно тот же результат получим, если воспользуемся формулой (1.7)) показывает, что уравнением регрессии объясняется 98,7% дисперсии результативного признака, а на долю прочих факторов приходится лишь 1,3%.

Оценим качество уравнения регрессии в целом с помощью F -критерия Фишера. Сосчитаем фактическое значение F -критерия:

$$F = \frac{r_{xy}^2}{1 - r_{xy}^2} \cdot (n - 2) = \frac{0,987}{1 - 0,987} \cdot 6 = 455,54.$$

Табличное значение ($k_1 = 1$, $k_2 = n - 2 = 6$, $\alpha = 0,05$): $F_{\text{табл}} = 5,99$. Так как $F_{\text{факт}} > F_{\text{табл}}$, то признается статистическая значимость уравнения в целом.

Для оценки статистической значимости коэффициентов регрессии и корреляции рассчитаем t -критерий Стьюдента и доверительные интервалы каждого из показателей. Рассчитаем случайные ошибки параметров линейной регрессии и коэффициента корреляции

$$\left(S_{\text{ост}}^2 = \frac{\sum (y - y_x)^2}{n - 2} = \frac{0,1257}{8 - 2} = 0,021 \right):$$

$$m_b = \frac{S_{\text{ост}}}{\sigma_x \cdot \sqrt{n}} = \frac{\sqrt{0,021}}{5,53 \cdot \sqrt{8}} = 0,0093, \quad m_a = S_{\text{ост}} \cdot \frac{\sqrt{\sum x^2}}{\sigma_x \cdot n} = \frac{\sqrt{0,021 \cdot 885,24}}{5,53 \cdot 8} = 0,0975,$$

$$m_r = \sqrt{\frac{1 - r^2}{n - 2}} = \sqrt{\frac{1 - 0,987}{6}} = 0,0465.$$

Фактические значения t -статистик: $t_b = \frac{0,168}{0,0093} = 18,065, \quad t_a = \frac{0,836}{0,0975} = 8,574,$

$$t_r = \frac{0,994}{0,0465} = 21,376. \text{ Табличное значение } t\text{-критерия Стьюдента при } \alpha = 0,05 \text{ и числе}$$

степеней свободы $\nu = n - 2 = 6$ есть $t_{\text{табл}} = 2,447$. Так как $t_b > t_{\text{табл}}, t_a > t_{\text{табл}}$ и $t_r > t_{\text{табл}}$, то признаем статистическую значимость параметров регрессии и показателя тесноты связи.

Средняя ошибка аппроксимации (находим с помощью столбца 10 таблицы 2.2:

$$A_i = \left| \frac{y_i - y_{x_i}}{y_i} \right| \cdot 100\%$$

$\bar{A} = 6,52\%$ говорит о хорошем качестве уравнения регрессии, т.е. свидетельствует о хорошем подборе модели к исходным данным.

И, наконец, найдем прогнозное значение результативного фактора y_p при значении признака-фактора, составляющем 110% от среднего уровня $x_p = 1,1 \cdot \bar{x} = 1,1 \cdot 8,95 = 9,845$, т.е. найдем расходы на питание, если доходы семьи составят 9,85 тыс. руб.

$$y_p = 0,836 + 0,168 \cdot 9,845 = 2,490 \text{ (тыс. руб.)}$$

Значит, если доходы семьи составят 9,845 тыс. руб., то расходы на питание будут 2,490 тыс. руб.

Найдем доверительный интервал прогноза. Ошибка прогноза

$$m_{y_p} = S_{\text{ост}} \cdot \sqrt{1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{n \cdot \sigma_x^2}} = \sqrt{0,021 \cdot \left(1 + \frac{1}{8} + \frac{(9,845 - 8,95)^2}{8 \cdot 30,56} \right)} = 0,154,$$

а доверительный интервал ($y_p - \Delta_{y_p} \leq y_p \leq y_p + \Delta_{y_p}$): $2,113 < y_p < 2,867$.

Т.е. прогноз является статистически надежным.

Теперь на одном графике изобразим исходные данные и линию регрессии:

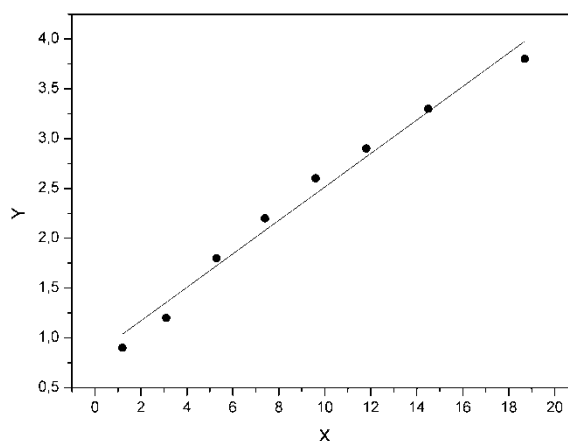


Рис. 2.2 - Линейная модель

2.3 Практические задания

1. $\sum_{\text{факт}}=1170$, $\sum_{\text{общ}}=1206$. Чему будет равна остаточная сумма квадратов отклонений?
2. Известны общая и остаточная суммы квадратов отклонений: $\sum_{\text{общ}}= 1500$, $\sum_{\text{ост}} = 260$. Чему будет равна факторная сумма квадратов отклонений?
3. Торговое предприятие имеет сеть, состоящую из 12 магазинов, информация о деятельности которых представлена в таблице 2.3

Таблица 2.3 Данные для решения

№ магазина	Годовой товарооборот, млн.руб.	Торговая площадь, тыс.м ²	Среднее число посетителей в день, тыс. чел.
1	19,8	0,2	8,3
2	38,1	0,3	10,2
3	41,0	0,6	9,3
4	41,1	0,5	11,0
5	56,3	0,8	8,5
6	68,5	1,0	7,5
7	75,0	0,9	12,4
8	89,1	1,2	10,8
9	91,1	1,3	10,0
10	91,3	1,1	13,7
11	100,0	1,3	12,3
12	108,6	1,5	13,9

Требуется: построить диаграммы рассеяния годового товарооборота (y) в зависимости от торговой площади (x_1) и среднего числа посетителей в день (x_2); определить форму связи и уравнение связи между результатом и каждым из факторов.

4. Исследуется зависимость затрат на рекламу y от годового оборота x в некоторой отрасли. Для этого собрана информация по 20 случайно выбранным предприятиям этой отрасли:

$$\sum x = 96 ; \sum y = 64 ; \sum xy = 492 ; \sum x^2 = 657 ; \sum y^2 = 526 .$$

Предполагается, что зависимость y от x описывается уравнением $y=a+bx$. Оценить параметры регрессии с помощью метода наименьших квадратов.

5. По совокупности 8 предприятий, выпускающих один и тот же вид продукции изучается зависимость между признаками x (тыс.ед.) – выпуск продукции; y (млн.руб.) – затраты на производство. При оценке регрессионной модели были получены следующие промежуточные результаты: $F_{\text{факт}}=276$. Определите коэффициент детерминации.

6. Пусть имеются данные по 10 квартирам по двум признакам (таблица 2.4):

Таблица 2.4 Данные для решения

№	Жилая площадь, m^2 , X	Цена квартиры, тыс. долларов, Y
1	34	37
2	20	17
3	40	36
4	42	45
5	36	32
6	37	40
7	12	15
8	26	24
9	45	44
10	18	20

1. Построить график – **поле корреляции** и сделать вывод о виде модели.
2. Построить линейное уравнение парной регрессии Y от X .
3. Рассчитать линейный коэффициент парной корреляции, коэффициент детерминации и среднюю ошибку аппроксимации. Сделать выводы.
4. Оценить статистическую значимость параметров регрессии и корреляции, а также всего уравнения с помощью F -критерия Фишера и t -критерия Стьюдента. Сделать выводы.
5. На одном графике построить исходные данные и теоретическую прямую.
6. Выполнить прогноз Y^* при прогнозном значении X^* , равном 120% от максимального значения.
7. Оценить точность прогноза, рассчитав ошибку прогноза и его доверительный интервал.

2.4 Семестровая работа 2 «Парная регрессия»

1. Построить график – **поле корреляции** и сделать вывод о виде модели.
2. Построить линейное уравнение парной регрессии Y от X .
3. Рассчитать линейный коэффициент парной корреляции, коэффициент детерминации и среднюю ошибку аппроксимации. Сделать выводы.
4. Оценить статистическую значимость параметров регрессии и корреляции с помощью F -критерия Фишера и t -критерия Стьюдента. Сделать выводы.
5. На одном графике построить исходные данные и теоретическую прямую.
6. Выполнить прогноз Y^* при прогнозном значении X^* , заданном в задании
7. Оценить точность прогноза, рассчитав ошибку прогноза и его доверительный интервал.

ЗАДАНИЯ по вариантам

1. Туристическая компания предлагает места в гостиницах приморского курорта. Менеджера компании интересует, насколько возрастает привлекательность гостиницы в зависимости от ее удаленности от пляжа. С этой целью по 10 гостиницам города была выяснена среднегодовая наполняемость номеров и расстояние от пляжа:

Расстояние, км X	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,7	0,8	0,9
Наполняемость, % Y	95	96	90	89	90	83	85	80	78	75

К пункту 6. Значение $x^* = 0,75$.

2. Компанию по прокату автомобилей интересует зависимость между пробегом автомобилей и стоимостью ежемесячного обслуживания. Для выяснения характера этой связи было отобрано 10 автомобилей. Результаты исследования представлены в таблице:

Пробег, тыс. км, X	6	7	8	9	10	11	12	13	14	15
Стоимость обслуживания, долл., Y	13	16	15	20	19	21	26	23	30	32

К пункту 6. Значение $x^* = 18$.

3. Торговцу нужно выяснить, как изменяется количество пучков салата, продаваемых ежедневно в розницу. Имеются следующие сведения о количестве и цене:

Цена, у.е. за единицу, X	30	31	25	26	22	24	16	12	10	8
Количество, тыс./день, Y	28	29	34	35	37	37	41	46	48	52

К пункту 6. Значение $x^* = 20$.

4. Компания, занимающаяся продажей радиоаппаратуры, установила на видеомангитофон определенной модели цену, дифференцированную по регионам. Следующие данные показывают цены на видеомангитофон в 10 различных регионах и соответствующее им число продаж:

Цена, у.е., X	5,5	6,0	6,5	6,0	5,0	5,6	4,5	5,0	4,4	4,0
Число продаж, шт., Y	420	380	350	400	440	380	450	420	460	500

К пункту 6. Значение $x^* = 5,7$.

5. Некоторая компания недавно провела рекламную кампанию в магазинах с демонстрацией антисептических качеств своего нового моющего средства. Через 10 недель компания решила проанализировать эффективность этого вида рекламы, сопоставив еженедельные объемы продаж с расходами на рекламу (тыс. руб.):

Расходы на рекламу, у.е. X	5	8	6	5	3	9	12	4	3	10
Объем продаж, у.е., Y	72	76	78	70	68	80	82	65	62	90

К пункту 6. Значение $x^* = 7$.

6. По 10 однородным предприятиям имеются данные о количестве рабочих с профессиональной подготовкой и количестве бракованной продукции:

Количество рабочих с профессиональной подготовкой, %, X	10	12	14	17	24	28	30	35	40	50
Количество бракованной продукции, %, Y	18	17	14	12	10	10	8	9	6	6

К пункту 6. Значение $x^* = 27$.

7. При исследовании годового дохода и сбережений населения в случайном порядке отобрано 10 человек. Получены следующие данные:

Доход, тыс. у.е., X	15	6	9	3	20	11	14	10	12	14
Сбережения, у.е., Y	2000	200	500	500	2500	1800	1500	1500	1600	1800

К пункту 6. Значение $x^* = 17$.

8. Проведен опрос случайно выбранных 10 студентов, проживающих в общежитии университета, для выявления зависимости между средним баллом по результатам предыдущей сессии и числом часов в неделю, затраченных студентом на самостоятельную подготовку:

Число часов, X	25	22	19	15	15	30	20	30	10	17
Средний балл, Y	4,6	4,3	3,8	3,8	4,2	4,3	3,8	4,0	3,1	3,9

К пункту 6. Значение $x^* = 27$.

9. Имеются данные по 10 предприятиям о производительности труда и коэффициенте механизации работ:

Коэффициент механизации работ, %, X	36	40	41	47	56	54	60	55	61	67
Производительность труда, шт., Y	28	30	31	33	34	37	38	40	41	43

К пункту 6. Значение $x^* = 65$.

10. Представлены данные, характеризующие зависимость между % лишнего веса и количеством больных диабетом по району:

Количество больных с сахарным диабетом, Y	30	33	45	48	53	25	29	34	65	25	28	35
Лишний вес, %, X	65	60	70	77	80	50	45	66	90	40	48	59

К пункту 6. Значение $x^* = 30$.

3. ПАРНАЯ НЕЛИНЕЙНАЯ РЕГРЕССИЯ

3.1 Краткий теоретический материал

Во многих практических случаях моделирование экономических зависимостей линейными уравнениями дает вполне удовлетворительный результат и может использоваться для анализа и прогнозирования. Однако многие экономические зависимости не являются линейными по своей сути, и поэтому их моделирование линейными уравнениями регрессии не даст положительного результата.

Если между экономическими явлениями существуют нелинейные соотношения, то они выражаются с помощью соответствующих нелинейных функций.

Различают два класса нелинейных регрессий:

1. Регрессии, нелинейные относительно включенных в анализ объясняющих переменных, но линейные по оцениваемым параметрам, например

- полиномы различных степеней – $y_x = a + b \cdot x + c \cdot x^2$, $y_x = a + b \cdot x + c \cdot x^2 + d \cdot x^3$;
- равносторонняя гипербола – $y_x = a + b/x$;
- полулогарифмическая функция – $y_x = a + b \cdot \ln x$.

2. Регрессии, нелинейные по оцениваемым параметрам, например

- степенная – $y_x = a \cdot x^b$;
- показательная – $y_x = a \cdot b^x$;
- экспоненциальная – $y_x = e^{a+b \cdot x}$.

Регрессии нелинейные по включенным переменным приводятся к линейному виду простой заменой переменных, а дальнейшая оценка параметров производится с помощью метода наименьших квадратов. Рассмотрим некоторые функции.

Парабола второй степени $y_x = a + b \cdot x + c \cdot x^2$ приводится к линейному виду с помощью замены: $x = x_1$, $x^2 = x_2$. В результате приходим к двухфакторному уравнению $y_x = a + b \cdot x_1 + c \cdot x_2$, оценка параметров которого при помощи МНК приводит к системе следующих нормальных уравнений:

$$\begin{cases} a \cdot n + b \cdot \sum x_1 + c \cdot \sum x_2 = \sum y; \\ a \cdot \sum x_1 + b \cdot \sum x_1^2 + c \cdot \sum x_1 \cdot x_2 = \sum x_1 \cdot y; \\ a \cdot \sum x_2 + b \cdot \sum x_1 \cdot x_2 + c \cdot \sum x_2^2 = \sum x_2 \cdot y. \end{cases}$$

А после обратной замены переменных получим

$$\begin{cases} a \cdot n + b \cdot \sum x + c \cdot \sum x^2 = \sum y; \\ a \cdot \sum x + b \cdot \sum x^2 + c \cdot \sum x^3 = \sum x \cdot y; \\ a \cdot \sum x^2 + b \cdot \sum x^3 + c \cdot \sum x^4 = \sum x^2 \cdot y. \end{cases}$$

Парабола второй степени обычно применяется в случаях, когда для определенного интервала значений фактора меняется характер связи рассматриваемых признаков: прямая связь меняется на обратную или обратная на прямую.

Кривая, для которой $b > 0$, $c < 0$ используется при изучении зависимости з/п работников физического труда от возраста. При $b < 0$, $c > 0$ – зависимость затрат на производство от объема выпуска. Часто можно использовать лишь сегмент параболы.

Равносторонняя гипербола $y_x = a + b/x$ может быть использована для характеристики связи удельных расходов сырья, материалов, топлива от объема выпускаемой продукции, времени обращения товаров от величины товарооборота, процента прироста заработной платы от уровня безработицы (например, кривая А.В. Филлипса), расходов на непродовольственные товары от доходов или общей суммы расходов (например, кривые Э. Энгеля) и в других случаях. Гипербола приводится к линейному уравнению простой заменой: $z = 1/x$.

Система линейных уравнений при применении МНК будет выглядеть следующим образом:

$$\begin{cases} a \cdot n + b \cdot \sum \frac{1}{x} = \sum y; \\ a \cdot \sum \frac{1}{x} + b \cdot \sum \frac{1}{x^2} = \sum \frac{1}{x} \cdot y. \end{cases}$$

Аналогичным образом приводятся к линейному виду зависимости $y_x = a + b \cdot \ln x$, $y_x = a + b \cdot \sqrt{x}$ и другие.

Класс функций, нелинейных по параметрам, в свою очередь, делится на два типа:

- *нелинейные модели внутренне линейные;*
- *нелинейные модели внутренне нелинейные.*

Внутренне линейные модели могут быть приведены к линейному виду.

К внутренне линейным моделям относятся, например,

– степенная функция $y_x = a \cdot x^b$,

– показательная $y_x = a \cdot b^x$,

– экспоненциальная $y_x = e^{a+b \cdot x}$,

– логистическая $y_x = \frac{a}{1 + b \cdot e^{-c \cdot x}}$,

– обратная $y_x = \frac{1}{a + b \cdot x}$.

1) $y = a \cdot x^b \cdot \varepsilon$ - внутренне линейна, так как $\ln y = \ln a + b \ln x + \ln \varepsilon$ - линейна по параметрам;

2) $y = e^{a+bx} \cdot \varepsilon$ - внутренне линейна, так как $\ln y = a + bx + \ln \varepsilon$;

3) $y = \frac{1}{a + bx + \varepsilon}$ - внутренне линейна, так как $z = \frac{1}{y} \Rightarrow z = a + bx + \varepsilon$;

4) $y = \frac{a}{1 + be^{-cx+\varepsilon}}$ - логистическая функция – внутренне линейна –

$$\frac{a}{y} = 1 + be^{-cx+\varepsilon}; \quad \frac{a}{y} - 1 = be^{-cx+\varepsilon}; \quad \ln\left(\frac{a}{y} - 1\right) = \ln b - cx + \varepsilon.$$

К **внутренне нелинейным** моделям можно, например, отнести следующие модели:

$y_x = a + b \cdot x^c$, $y_x = a \cdot \left(1 - \frac{1}{1 - x^b}\right)$, которые к линейному виду не приводятся.

Среди нелинейных моделей наиболее часто используется **степенная функция** $y = a \cdot x^b \cdot \varepsilon$, которая приводится к линейному виду логарифмированием:

$$\ln y = \ln(a \cdot x^b \cdot \varepsilon);$$

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

$$Y = A + b \cdot X + E,$$

где $Y = \ln y$, $X = \ln x$, $A = \ln a$, $E = \ln \varepsilon$. Т.е. МНК мы применяем для преобразованных данных:

$$\begin{cases} A \cdot n + b \cdot \sum X = \sum Y, \\ A \cdot \sum X + b \cdot \sum X^2 = \sum X \cdot Y, \end{cases}$$

а затем потенцированием находим искомое уравнение.

Широкое использование степенной функции связано с тем, что параметр b в ней имеет четкое экономическое истолкование – он является коэффициентом эластичности.

Коэффициент эластичности показывает, на сколько процентов измениться в среднем результат, если фактор изменится на 1%.

Формула для расчета коэффициента эластичности имеет вид:

$$\varepsilon = f'(x) \cdot \frac{x}{y}$$

Так как для остальных функций коэффициент эластичности не является постоянной величиной, а зависит от соответствующего значения фактора X , то обычно рассчитывается средний коэффициент эластичности: $\bar{\varepsilon} = f'(\bar{x}) \cdot \frac{\bar{x}}{\bar{y}}$.

Приведем формулы для расчета средних коэффициентов эластичности для наиболее часто используемых типов уравнений регрессии:

Таблица 3.1 Формулы для расчета среднего коэффициента эластичности

Вид функции, y	Первая производная, y'	Средний коэффициент эластичности, $\bar{\varepsilon}$
$y = a + b \cdot x + \varepsilon$	b	$\frac{b \cdot \bar{x}}{a + b \cdot \bar{x}}$
$y = a + b \cdot x + c \cdot x^2 + \varepsilon$	$b + 2c \cdot x$	$\frac{(b + 2c \cdot \bar{x}) \cdot \bar{x}}{a + b \cdot \bar{x} + c \cdot \bar{x}^2}$
$y = a + \frac{b}{x} + \varepsilon$	$-\frac{b}{x^2}$	$-\frac{b}{a \cdot \bar{x} + b}$
$y = a \cdot x^b \cdot \varepsilon$	$a \cdot b \cdot x^{b-1}$	b
$y = a \cdot b^x \cdot \varepsilon$	$a \cdot \ln b \cdot b^x$	$\bar{x} \cdot \ln b$
$y = a + b \cdot \ln x + \varepsilon$	$\frac{b}{x}$	$\frac{b}{a + b \cdot \ln \bar{x}}$
$y = \frac{1}{a + b \cdot x + \varepsilon}$	$-\frac{b}{(a + b \cdot x)^2}$	$-\frac{b \cdot \bar{x}}{a + b \cdot \bar{x}}$

Возможны случаи, когда расчет коэффициента эластичности не имеет смысла. Это происходит тогда, когда для рассматриваемых признаков бессмысленно определение изменения в процентах.

Уравнение нелинейной регрессии, так же, как и в случае линейной зависимости, дополняется показателем тесноты связи. В данном случае это **индекс корреляции**:

$$\rho_{xy} = \sqrt{1 - \frac{\sigma_{\text{ост}}^2}{\sigma_y^2}},$$

где $\sigma_y^2 = \frac{1}{n} \sum (y - \bar{y})^2$ – общая дисперсия результативного признака y ,

$\sigma_{\text{ост}}^2 = \frac{1}{n} \sum (y - y_x)^2$ – остаточная дисперсия.

Величина данного показателя находится в пределах: $0 \leq \rho_{xy} \leq 1$. Чем ближе значение индекса корреляции к единице, тем теснее связь рассматриваемых признаков, тем более надежно уравнение регрессии.

Квадрат индекса корреляции носит название **индекса детерминации** и характеризует долю дисперсии результативного признака y , объясняемую регрессией, в общей дисперсии результативного признака:

$$\rho_{xy}^2 = 1 - \frac{\sigma_{\text{ост}}^2}{\sigma_y^2} = \frac{\sigma_{\text{объясн}}^2}{\sigma_y^2},$$

т.е. имеет тот же смысл, что и в линейной регрессии; $\sigma_{\text{объясн}}^2 = \frac{1}{n} \sum (y_x - \bar{y})^2$.

Индекс детерминации ρ_{xy}^2 можно сравнивать с коэффициентом детерминации r_{xy}^2 для обоснования возможности применения линейной функции. Чем больше кривизна линии регрессии, тем величина r_{xy}^2 меньше ρ_{xy}^2 . А близость этих показателей указывает на то, что нет необходимости усложнять форму уравнения регрессии и можно использовать линейную функцию.

Индекс детерминации используется для проверки существенности в целом уравнения регрессии по F -критерию Фишера:

$$F = \frac{\rho_{xy}^2}{1 - \rho_{xy}^2} \cdot \frac{n - m - 1}{m},$$

где ρ_{xy}^2 – индекс детерминации,

n – число наблюдений,

m – число параметров при переменной X .

Фактическое значение F -критерия (1.23) сравнивается с табличным при уровне значимости α и числе степеней свободы $k_2 = n - m - 1$ (для остаточной суммы квадратов) и $k_1 = m$ (для факторной суммы квадратов).

О качестве нелинейного уравнения регрессии можно также судить и по средней ошибке аппроксимации, которая, так же как и в линейном случае.

3.2 Типовые примеры и образцы их решения

Пример . По семи предприятиям легкой промышленности региона получена информация, характеризующая зависимость объема выпуска продукции (Y , млн. руб.) от объема капиталовложений (X , млн. руб.).

Y	64	56	52	48	50	46	38
X	64	68	82	76	84	96	100

Требуется:

1 . Для характеристики Y от X построить следующие модели:

- линейную (для сравнения с нелинейными),
- степенную,
- показательную,
- гиперболическую.

2. Оценить каждую модель, определив:

- индекс корреляции,
- среднюю относительную ошибку,
- коэффициент детерминации,
- F -критерий Фишера.

3. Составить сводную таблицу вычислений, выбрать лучшую модель, дать интерпретацию рассчитанных характеристик.

4. Рассчитать прогнозные значения результативного признака по лучшей модели, если объем капиталовложений составит 89,573 млн. руб.

5. Результаты расчетов отобразить на графике.

Решение: 1. Построение линейной модели парной регрессии

Определим линейный коэффициент парной корреляции по следующей формуле:

$$r_{Y,X} = \frac{\sum (y - \bar{y}) \times (x - \bar{x})}{\sqrt{\sum (y - \bar{y})^2 \times \sum (x - \bar{x})^2}} = \frac{-593,714}{\sqrt{397,71 \times 1077,71}} = -0,907.$$

Можно сказать, что связь между объемом капиталовложений X и объемом выпуска продукции Y обратная, достаточно сильная.

Уравнение линейной регрессии имеет вид: $\hat{y} = a + b \times x$.

Значения параметров a и b линейной модели определим, используя данные таблицы 3.2.

$$b = \frac{\overline{y \times x} - \bar{y} \times \bar{x}}{\overline{x^2} - \bar{x}^2} = \frac{4033,14 - 50,57 \times 81,43}{6784,57 - 81,43^2} = -0,55,$$

$$a = \bar{y} - b \times \bar{x} = 50,57 - 0,55 \times 81,43 = 95,36$$

Уравнение линейной регрессии имеет вид: $\hat{y} = 95,36 - 0,55 \times x$.

Таблица 3.2 Расчет параметров линейной регрессии

t	y	x	$y \times x$	$x \times x$	$(y_i - \bar{y})$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	$\hat{y}$	$e_i = y_i - \hat{y}$	$\left \frac{e_i}{y_i} \right \times 100\%$
1	64	64	4096	4096	13,43	180,36	-17,4	303,8	60,2	3,84	6,000
2	56	68	3808	4624	5,43	29,485	-13,4	180,36	58,0	-1,96	-3,500
3	52	82	4264	6724	1,43	2,0449	0,57	0,3249	50,3	1,74	3,346
4	48	76	3648	5776	-2,57	6,6049	-5,43	29,485	53,6	-5,56	-11,583
5	50	84	4200	7056	-0,57	0,3249	2,57	6,6049	49,2	0,84	1,680
6	46	96	4416	9216	-4,57	20,885	14,57	212,28	42,6	3,44	7,478
7	38	100	3800	10000	-12,6	158,0	18,57	344,84	40,4	-2,36	-6,211
Итого	354,00	570,00	28232	47492	0,01	397,71		1077,7		-0,02	39,798
ср. знач.	50,57	81,43	4033,14	6784,57							5,685
диспер.	56,80	154,00									

С увеличением объема капиталовложений на 1 млн. руб. объем выпускаемой продукции уменьшится в среднем на 550 тыс. руб. Это свидетельствует о неэффективности работы предприятий, и необходимо принять меры для выяснения причин и устранения этого недостатка.

Рассчитаем коэффициент детерминации:

$$R^2 = r_{y-x}^2 = 0,822.$$

Вариация результата Y (объема выпуска продукции) на 82,2 % объясняется вариацией фактора X (объемом капиталовложений).

Оценку значимости уравнения регрессии проведем с помощью F -критерия Фишера:

$$F = \frac{r_{y,x}^2}{1 - r_{y,x}^2} \times (n - 2) = \frac{0,822}{1 - 0,822} \times (7 - 2) = 23,09;$$

$$F > F_{\text{табл}} = 6,61 \text{ для } \alpha = 0,05; k_1 = m = 1, k_2 = n - m - 1 = 5.$$

Уравнение регрессии с вероятностью 0,95 в целом статистически значимое, т. к. $F > F_{\text{табл}}$.

Определим среднюю относительную ошибку:

$$\bar{E}_{\text{ош}} = \frac{1}{n} \times \sum \frac{|E_i|}{y} \times 100\% = \frac{1}{n} \times \sum \left| \frac{y - \hat{y}}{y} \right| \times 100\% = \frac{39,798}{7} = 5,685\%.$$

В среднем расчетные значения $\hat{y}$ для линейной модели отличаются от фактических значений на 5,685%.

2. Построение степенной модели парной регрессии

Уравнение степенной модели имеет вид: $\hat{y} = a \times x^b$.

Для построения этой модели необходимо произвести линеаризацию переменных. Для этого произведем логарифмирование обеих частей уравнения: $\lg \hat{y} = \lg a + b \lg x$.

Обозначим $Y = \lg y$, $X = \lg x$, $A = \lg a$.

Тогда уравнение примет вид: $Y = A + b X$ — линейное уравнение регрессии.

Рассчитаем его параметры, используя данные таблицы 3.3.

Таблица 3.3 Расчет параметров степенной регрессии

№	y	Y=lgy	x	X =lgx	YX	X ²	$\hat{y}$	e _i	e _i /y ×100%	e _i ²
1	64	1,8062	64	1,8062	3,2623	3,2623	61,294	2,706	4,23	7,322
2	56	1,7482	68	1,8325	3,2036	3,3581	58,066	-2,066	3,69	4,270
3	52	1,7160	82	1,9138	3,2841	3,6627	49,133	2,867	5,51	8,220
4	48	1,6812	76	1,8808	3,1621	3,5375	52,580	-4,580	9,54	20,976
5	50	1,6990	84	1,9243	3,2693	3,7029	48,088	1,912	3,82	3,657
6	46	1,6628	96	1,9823	3,2960	3,9294	42,686	3,314	7,20	10,982
7	38	1,5798	100	2,0000	3,1596	4,0000	41,159	-3,159	8,31	9,980
Итого	354	11,8931		13,3399	22,6370	25,4528		0,51	42,32	65,407

$$b = \frac{\overline{Y \cdot X} - \bar{Y} \cdot \bar{X}}{\overline{X^2} - \bar{X}^2} = \frac{3,2339 - 1,699 \cdot 1,9057}{3,6361 - 1,9057 \cdot 1,9057} = -0,8921,$$

$$A = \bar{Y} - b \cdot \bar{X} = 1,699 + 0,8921 \cdot 1,9057 = 3,3991.$$

Уравнение регрессии будет иметь вид: $Y = 3,3991 - 0,8921 X$.

Перейдем к исходным переменным x и y , выполнив потенцирование данного уравнения.

$$\hat{y} = 10^{3,399} \times x^{-0,892}$$

Получим уравнение степенной модели регрессии:

$$\hat{y} = 2506,915 \times x^{-0,892}.$$

Определим индекс корреляции:

$$\rho_{xy} = \sqrt{1 - \frac{\sum (y - \hat{y})^2}{\sum (y - \bar{y})^2}} = 0,914.$$

Связь между показателем y и фактором x можно считать достаточно сильной.

Коэффициент детерминации равен 0,836:

$$R^2 = \rho_{xy}^2 = 0,914^2 = 0,836.$$

Вариация результата Y (объема выпуска продукции) на 83,6% объясняется вариацией фактора X (объемом капиталовложений).

Рассчитаем F -критерий Фишера:

$$F = \frac{R^2}{1 - R^2} \times (n - 2) = \frac{0,836}{1 - 0,836} \times 5 = 25,5.$$

$F > F_{\text{табл}} = 6,61$ для $\alpha = 0,05$; $k_1 = m = 1$, $k_2 = n - m - 1 = 5$.

Уравнение регрессии с вероятностью 0,95 в целом статистически значимое, т.к. $F > F_{\text{табл}}$.

Средняя относительная ошибка

$$\bar{E}_{\text{отн}} = \frac{1}{n} \times \sum \frac{|E_i|}{y} \times 100\% = \frac{1}{n} \times \sum \left| \frac{y - \hat{y}}{y} \right| \times 100\%;$$

$$\bar{E}_{\text{отн}} = \frac{42,32}{7} = 6,04\%.$$

В среднем расчетные значения $\hat{y}$ для степенной модели отличаются от фактических значений на 6,04%.

3. Построение показательной функции

Уравнение показательной кривой: $\hat{y} = ab^x$.

Для построения этой модели необходимо произвести линеаризацию переменных. Для этого осуществим логарифмирование обеих частей уравнения: $\lg \hat{y} = \lg a + x \lg b$.

Обозначим: $Y = \lg \hat{y}$, $B = \lg b$, $A = \lg a$.

Получим линейное уравнение регрессии: $Y = A + Bx$.

Рассчитаем его параметры, используя данные таблицы 3.4

Таблица 3.4 Расчет параметров показательной функции

t	y	Y	x	Yx	x^2	$Y - \bar{Y}$	$(Y - \bar{Y})^2$	$X - \bar{X}$	$(X - \bar{X})^2$	$\hat{y}$	$(y - \hat{y})^2$	e_i	$ e_i/y \times 100$
1	64	1,8062	64	115,60	4096	0,1072	0,0115	-17,43	303,76	60,6	11,464	3,3859	5,290
2	56	1,7482	68	118,88	4624	0,0492	0,0024	-13,43	180,33	58	3,9632	-1,991	3,555
3	52	1,7160	82	140,71	6724	0,0170	0,0003	0,57	0,33	49,7	5,4221	2,3285	4,478
4	48	1,6812	76	127,77	5776	-0,017	0,0003	-5,43	29,47	53,1	25,804	-5,08	10,583
5	50	1,6990	84	142,71	7056	0,0000	0,0000	2,57	6,61	48,6	2,0031	1,4153	2,831
6	46	1,6628	96	159,62	9216	-0,036	0,0013	14,57	212,33	42,5	11,933	3,4544	7,509
7	38	1,5798	100	157,98	10000	-0,119	0,0142	18,57	344,90	40,7	7,3132	-2,704	7,117
итого	354	11,8931	570	963,28	4749		0,0300		1077,7		67,903	0,8093	41,363
Сред. знач.	50,57	1,6990	81,4	137,61	6785								5,909

$$b = \frac{\overline{Y \times x} - \bar{Y} \times \bar{x}}{\overline{x^2} - \bar{x}^2} = \frac{137,61 - 1,699 \cdot 81,43}{6785 - 81,43 \cdot 81,43} = -0,0048;$$

$$A = \bar{Y} - b \times \bar{x} = 1,699 + 0,0048 \cdot 81,43 = 2,09.$$

Уравнение будет иметь вид: $Y = 2,09 + 0,0048 x$.

Перейдем к исходным переменным x и y , выполнив потенцирование данного уравнения:

$$\hat{y} = 10^{2,09} \times (10^{-0,0048})^x = 123,03 \times 0,989^x.$$

Определим индекс корреляции:

$$\rho_{xy} = \sqrt{1 - \frac{\Sigma(y - \hat{y})^2}{\Sigma(y - \bar{y})^2}} = \sqrt{1 - \frac{67,9}{397,71}} = 0,91.$$

Связь между показателем y и фактором x можно считать тесной.

Индекс детерминации: $R^2 = \rho_{xy}^2 = 0,91^2 = 0,828$.

Вариация результата Y (объема выпуска продукции) на 41,1 % объясняется вариацией фактора X (объем капиталовложений).

Рассчитаем F -критерий Фишера:

$$F = \frac{R^2}{1 - R^2} \times (n - 2) = 24,06.$$

$F > F_{\text{табл}} = 6,61$ для $\alpha = 0,05$; $k_1 = m = 1$, $k_2 = n - m - 1 = 5$.

Уравнение регрессии с вероятностью 0,95 в целом статистически значимое, т. е. $F > F_{\text{табл}}$.

Средняя относительная ошибка:

$$\bar{E}_{\text{отн}} = \frac{41,363}{7} = 5,909\%.$$

В среднем расчетные значения $\hat{y}$ для линейной модели отличаются от фактических значений на 5,909 %.

4. Построение гиперболической функции

Уравнение гиперболической функции:

Произведем линеаризацию модели путем замены $X = 1/x$. В результате получим линейное уравнение $\hat{y} = a + b/x$.

Рассчитаем его параметры по данным таблицы 3.5.

$$A = \bar{Y} - b \times \bar{X} = 50,57 - 3571,9 \cdot 0,0126 = 5,7.$$

$$b = \frac{\overline{y \times X} - \bar{y} \times \bar{X}}{\bar{X}^2 - \bar{X}^2} = \frac{0,6941 - 50,57 \cdot 0,0126}{0,0001618 - 0,0126 \cdot 0,0126} = 3571,95;$$

Получим следующее уравнение гиперболической модели: $\hat{y} = 5,7 - 3571,9/x$.

Определим индекс корреляции:

$$\rho_{yx} = \sqrt{1 - \frac{\Sigma(y - \hat{y})^2}{\Sigma(y - \bar{y})^2}} = \sqrt{1 - \frac{65,159}{397,71}} = 0,914.$$

Связь между показателем y и фактором x можно считать достаточно сильной.

Индекс детерминации: $R^2 = \rho_{yx}^2 = 0,914^2 = 0,835$.

Вариация результата Y (объема выпуска продукции) на 83,5% объясняется вариацией фактора X (объемом капиталовложений).

Таблица 3.5 Расчет параметров гиперболической функции

t	y	x	X	yX	X^2	$Y - \bar{Y}$	$(Y - \bar{Y})^2$	$(y - \bar{y})$	e_i	$\hat{y}$	$ e_i/y \times 100$
1	64	64	0,0156	1,0000	0,0002441	13,43	180,33	61,5	2,489	6,195	3,889
2	56	68	0,0147	0,8235	0,0002163	5,43	29,47	58,2	-2,228	4,963	3,978
3	52	82	0,0122	0,6341	0,0001487	1,43	2,04	49,3	2,740	7,508	5,270
4	48	76	0,0132	0,6316	0,0001731	-2,57	6,61	52,7	-4,699	22,07	9,789
5	50	84	0,0119	0,5952	0,0001417	-0,57	0,32653	48,2	1,777	3,159	3,555
6	46	96	0,0104	0,4792	0,0001085	-4,57	20,90	42,9	3,093	9,564	6,723
7	38	100	0,0100	0,3800	0,0001000	-12,57	158,04	41,4	-3,419	11,69	8,997
итого	354		0,0880	4,5437	0,0011325		397,71	354,2	-0,246	65,15	42,202
Сред. нач.	50,5		0,0126	0,6491	0,0001618						6,029

F -критерий Фишера:

$$F = \frac{R^2}{1 - R^2} \times (n - 2) = \frac{0,835}{1 - 0,835} \times 5 = 25,3.$$

$F > F_{\text{табл}} = 6,61$ для $\alpha = 0,05$; $k_1 = m = 1$, $k_2 = n - m - 1 = 5$.

Уравнение регрессии с вероятностью 0,95 в целом не является статистически значимым, т. к. $F > F_{\text{табл}}$.

Определим среднюю относительную ошибку:

$$\bar{E}_{\text{отн}} = \frac{1}{n} \sum \frac{|E_i|}{y} \times 100\% = \frac{42,202}{7} = 6,029\%.$$

В среднем расчетные значения для линейной модели отличаются от фактических значений на 6,029 %. Для выбора лучшей модели построим сводную таблицу результатов.

Таблица 3.6 Сравнение результатов

Параметры	Коэффициент детерминации R^2	F -критерий Фишера	Индекс корреляции $r_{yx}(r_{yx})$	Средняя относительная ошибка аппроксимации
Модель				
Линейная	0,822	23,09	0,907	5,685
Степенная	0,828	24,06	0,910	6,054
Показательная	0,828	24,06	0,910	5,909
Гиперболическая	0,835	25,30	0,914	6,029

Все модели имеют примерно одинаковые характеристики, но большее значение F -критерия Фишера и большее значение коэффициента детерминации R^2 имеет гиперболическая модель. Ее можно взять в качестве лучшей для построения прогноза.

Расчет прогнозного значения результативного показателя:

Прогнозное значение результативного признака (объема выпуска продукции) определим по уравнению гиперболической модели, подставив в него планируемую (заданную по условию) величину объема капиталовложений:

$$\hat{y}_{\text{тп}} = 5,7 + 3571,9 \times X_{\text{тп}} = 5,7 + 3571,9 / 89,573 = 45,542 \text{ (млн. руб.)}$$

3.3 Практические задания

1. Получены функции:

$$y = a + bx^3, y = a + b \ln x, \ln y = a + b \ln x, y = a + bx^c, y^a = b + cx^2, y = 1 + a(1 - x^b), y = a + bx/10.$$

Определите, какие из представленных выше функций линейны по переменным, линейны по параметрам, не линейны, не линейны ни по переменным, ни по параметрам.

2. Для трех видов продукции A , B и C модели зависимости удельных постоянных расходов от объема выпускаемой продукции выглядит следующим образом:

$$y_A = 60 + 1/x; y_B = 80 + 0,7x; y_C = 40 x^{0,5}.$$

Определите коэффициенты эластичности по каждому виду продукции и поясните их смысл при $\bar{x} = 50$.

3. По группе 10 заводов, производящих однородную продукцию, получено уравнение регрессии себестоимости единицы продукции y (тыс. руб.) от уровня технической оснащенности x (тыс. руб.): $y = 20 + 700/x$. Доля остаточной дисперсии в общей составила 0,20.

Определите:

- 1) индекс корреляции;
- 2) значимость уравнения регрессии;
- 3) коэффициент эластичности, предполагая, что стоимость активных производственных фондов составляет 200 тыс. руб.

3. Предложить аналитическую форму модели по следующим данным

Таблица 3.7

y_t	40	47	45	57	65	60	55	84	76	94	96
x_t	30	33	40	43	46	49	55	56	58	59	63

Таблица 3.8

y_t	43	61	39	25	27	26	35	52	63	59
x_t	11,2	11,3	12,4	13	14,3	15,7	16,7	17,6	17,6	18,1

4. Для анализа работы торгового предприятия произведено 10 наблюдений числа покупателей X за день и выручки Y (тыс. руб.) (табл. 3.9):

Таблица 3.9

№	1	2	3	4	5	6	7	8	9	10
X	31	75	89	26	35	73	91	21	56	21
Y	64	100	103	50	63	95	109	43	93	37

Для характеристики зависимости Y от X построить следующие модели (с помощью графического анализа):

— линейную, степенную, параболу, гиперболу.

2. Оценить каждую модель, определив:

— коэффициент детерминации,

— F -критерий Фишера

— ошибку аппроксимации,

— коэффициент эластичности.

Дать интерпретацию рассчитанных характеристик для каждой модели.

3. Составить сводную таблицу вычислений, обосновать выбор лучшей модели.

4. Рассчитать прогнозные значения результативного признака Y , если прогнозное значение фактора X увеличится на 20 % относительно максимального значения.

3.4 Лабораторная 2 «Нелинейная регрессия»

По предприятиям промышленности региона получена информация, характеризующая зависимость **объема выпуска продукции (Y , млн. руб.)** от **объема капиталовложений (X , млн. руб.)**. Требуется:

1. Для характеристики зависимости Y от X построить следующие модели (с помощью графического анализа):

— линейную, степенную, параболу, гиперболу.

2. Оценить каждую модель, определив:

— коэффициент детерминации,

— F -критерий Фишера

— ошибку аппроксимации,

— коэффициент эластичности.

Дать интерпретацию рассчитанных характеристик для каждой модели.

3. Составить сводную таблицу вычислений, обосновать выбор лучшей модели.

4. Рассчитать прогнозные значения результативного признака Y , если прогнозное значение фактора X увеличится на 20 % относительно максимального значения.

Вариант 1

X	66	58	73	82	81	84	55	67	81	59
Y	133	107	145	162	163	170	104	132	159	116

Вариант 2

X	72	52	73	74	76	79	54	68	73	64
Y	121	84	119	117	129	128	102	111	112	98

Вариант 3

X	38	28	27	37	46	27	41	39	28	44
Y	69	52	46	63	73	48	67	62	47	67

Вариант 4

X	36	28	43	52	51	54	25	37	51	29
Y	104	77	117	137	143	144	82	101	132	77

Вариант 5

X	31	23	38	47	46	49	20	32	46	24
Y	38	26	40	45	51	49	34	35	42	24

Вариант 6

X	33	17	23	17	36	25	39	20	13	12
Y	43	27	32	29	45	35	47	32	22	24

Вариант 7

X	36	28	43	52	51	54	25	37	51	29
Y	85	60	99	117	118	125	56	86	115	68

Вариант 8

X	17	22	10	7	12	21	14	7	20	3
Y	26	27	22	19	21	26	20	15	30	13

Вариант 9

X	12	4	18	27	26	29	1	13	26	5
Y	21	10	26	33	34	37	9	21	32	14

Вариант 10

X	26	18	33	42	41	44	15	27	41	19
Y	43	28	51	62	63	67	26	43	61	33

Тема 4 МНОЖЕСТВЕННАЯ РЕГРЕССИЯ

4.1 Краткий теоретический материал

Линейная множественная регрессия: $\tilde{y} = a + b_1x_1 + b_2x_2 + \dots + b_kx_k$

Степенная функция: $\tilde{y} = a \cdot x_1^{b_1} \cdot x_2^{b_2} \cdot \dots \cdot x_k^{b_k}$

Экспонента: $\tilde{y} = e^{\alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k}$

Гипербола: $\tilde{y} = \frac{1}{\alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k}$

Оценка параметров линейной множественной регрессии

1) в натуральном масштабе, т.е. для уравнения $\tilde{y} = a + b_1x_1 + b_2x_2 + \dots + b_kx_k$ система нормальных уравнений имеет вид:

[illegible]

Ее решение может быть найдено, например, методом определителей.

Вычисление параметров линейной множественной регрессии можно провести с помощью инструмента *Сервис/Анализ данных/Регрессия*.

2) в стандартизированном масштабе:

$$\overline{t_y} = \beta_1 \cdot t_{x_1} + \beta_2 \cdot t_{x_2} + \dots + \beta_k \cdot t_{x_k},$$

где $t_y, t_{x_1}, t_{x_2}, \dots, t_{x_k}$ – стандартизированные переменные

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

β_i – стандартизированные коэффициенты регрессии. Решают систему нормальных уравнений вида

$$\left\{ \begin{array}{l} r_{yx_1} = \beta_1 + \beta_2 \cdot r_{x_2 x_1} + \beta_3 \cdot r_{x_3 x_1} + ... + \beta_k \cdot r_{x_k x_1} \\ r_{yx_2} = \beta_1 \cdot r_{x_1 x_2} + \beta_2 + \beta_3 \cdot r_{x_3 x_2} + ... + \beta_k \cdot r_{x_k x_2} \\ \\ r_{yx_k} = \beta_1 \cdot r_{x_1 x_k} + \beta_2 \cdot r_{x_2 x_k} + \beta_3 \cdot r_{x_3 x_k} + ... + \beta_k \end{array} \right.$$

Решая ее методом определителей, найдем β -коэффициенты.

Определение β -коэффициентов:

1) Находим матрицу парных коэффициентов корреляции. Для двухфакторной линейной регрессии она имеет вид:

	y	x_1	x_2
y	1		
x_1	r_{yx_1}	1	
x_2	r_{yx_2}	$r_{x_1x_2}$	1

Удобнее всего найти эту матрицу Excel, используя инструмент анализа данных *Корреляция*. Для этого в главном меню нужно последовательно выбрать *Сервис/Анализ данных/Корреляция*.

2) для стандартизированного уравнения регрессии

$$\bar{t}_y = \beta_1 \cdot t_{x_1} + \beta_2 \cdot t_{x_2}$$

имеем

$$\beta_{x_1} = \frac{r_{yx_1} - r_{yx_2} \cdot r_{x_1x_2}}{1 - (r_{x_1x_2})^2}; \beta_{x_2} = \frac{r_{yx_2} - r_{yx_1} \cdot r_{x_1x_2}}{1 - (r_{x_1x_2})^2}.$$

Коэффициенты «чистой» регрессии связаны с β -коэффициентами следующим образом:

$$\beta_i = \frac{\sigma_{x_i}}{\sigma_y} \cdot b_i.$$

Методика построения уравнения регрессии при двухфакторном регрессионном анализе $\bar{y} = a + bx + cz$

приводит к следующим формулам для оценки параметров:

$$b = \frac{r_{yx} - r_{yz}r_{xz}}{1 - r_{xz}^2} \cdot \frac{\sigma_y}{\sigma_x}, c = \frac{r_{yz} - r_{yx}r_{xz}}{1 - r_{xz}^2} \cdot \frac{\sigma_y}{\sigma_z}, a = \bar{y} - b\bar{x} - c\bar{z}.$$

Методика построения уравнения регрессии в виде степенной функции

$$y = a \cdot x_1^{b_1} \cdot x_2^{b_2} \cdot \dots \cdot x_k^{b_k} \cdot \varepsilon$$

Преобразуем ее в линейный вид:

$$\lg y = \lg a + b_1 \cdot \lg x_1 + b_2 \cdot \lg x_2 + \dots + b_k \cdot \lg x_k + \lg \varepsilon,$$

где переменные выражены в логарифмах. Далее процедура МНК такая же, что и описана выше: строится система нормальных уравнений и определяются параметры, которые затем следует потенцировать.

Оценка тесноты связи и статистической значимости во множественной регрессии

1) коэффициент множественной детерминации R^2 ,

$$R^2 = 1 - \frac{\sigma_{ост}^2}{\sigma_y^2} = \frac{\sum(\tilde{y} - \bar{y})^2}{\sum(y - \bar{y})^2};$$

2) индекс множественной корреляции R ;

3) **линейный коэффициент множественной корреляции** (для

$$\tilde{y} = a + b_1x_1 + b_2x_2 + \dots + b_kx_k)$$

$$R = \sqrt{\sum(\beta_i \cdot r_{yx_i})};$$

4) в случае двухфакторной линейной модели индекс множественной корреляции R может быть найден по формуле:

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

5) **Скорректированный индекс** (коэффициент) корреляции:

$$R_{коррект}^2 = 1 - (1 - R^2) \cdot \frac{n-1}{n-k-1}; k - \text{число параметров при переменных.}$$

В статистических пакетах прикладных программ в процедуре множественной регрессии обычно приводится скорректированный коэффициент (индекс) множественной корреляции (детерминации).

б) **дельта-коэффициенты** Δ_i :

$$\Delta_i = r_{yx_i} \cdot \frac{\beta_i}{R^2},$$

где r_{yx_i} – коэффициент парной корреляции между y и x_i ;

R^2 – множественный коэффициент детерминации.

7) **частные коэффициенты эластичности**:

$$\mathcal{E}_i = b_i \cdot \frac{\bar{x}_i}{\bar{y}},$$

где b_i – коэффициент «чистой» регрессии при факторе x_i ;

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

$\bar{x}_i$ – среднее значение признака x_i .

Значимость уравнения множественной регрессии в целом

оценивается с помощью F -критерия Фишера:

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

где n – число наблюдений, m – число параметров при переменной x . Если расчетное значение критерия с $k_1 = m$ и $k_2 = n - m - 1$ степенями свободы больше табличного при заданном уровне значимости, то модель считается значимой.

Прогнозирование по уравнению линейной множественной регрессии

$$\tilde{y}_{точеч} - t_{табл} \cdot \mu_{\tilde{y}} \leq \tilde{y}_{прогн} \leq \tilde{y}_{точеч} + t_{табл} \cdot \mu_{\tilde{y}}$$

где $\mu_{\tilde{y}}$ – ошибка прогнозного значения, вычисляемая по формуле

$$\mu_{\tilde{y}} = S \sqrt{1 + (X_0)^T \cdot (X^T X)^{-1} \cdot X_0}$$

$$X_0 = \begin{pmatrix} 1 \\ x_{1\text{прогн}} \\ x_{2\text{прогн}} \end{pmatrix} \quad X = \begin{pmatrix} n & \sum x_1 & \sum x_2 \\ \sum x_1 & \sum x_1^2 & \sum x_1 x_2 \\ \sum x_2 & \sum x_1 x_2 & \sum x_2^2 \end{pmatrix}$$

для двухфакторной модели.

Мерой для оценки включения фактора в модель служит частный F -критерий, т.е. F_{x_i} . Так, если оцениваем значимость влияния фактора x_p после включения в модель факторов $x_1, x_2, \dots, x_{p-1}$, то формула частного F -критерия примет вид:

$$F_{x_p} = \frac{R_{yx_1x_2\dots x_p}^2 - R_{yx_1x_2\dots x_{p-1}}^2}{1 - R_{yx_1x_2\dots x_p}^2} \cdot \frac{n - m - 1}{1}.$$

Если фактическое значение критерия с $k_1 = 1$ и $k_2 = n - m - 1$ степенями свободы больше табличного при заданном уровне значимости, то дополнительное включение фактора x_p в модель статистически оправдано и коэффициент регрессии при данном факторе статистически значим.

Оценка	значимости	коэффициентов	«чистой»	регрессии
$\tilde{y} = a + b_1x_1 + b_2x_2 + \dots + b_kx_k$				

Для каждого фактора используется формула

$$t_{b_i} = \frac{b_i}{m_{b_i}},$$

где b_i – коэффициент «чистой» регрессии при факторе x_i ; m_{b_i} – средняя квадратическая ошибка коэффициента регрессии b_i ,

$$m_{b_i} = \frac{\sigma_y \cdot \sqrt{1 - R_{yx_1x_2\dots x_k}^2}}{\sigma_{x_i} \sqrt{1 - R_{x_ix_1x_2\dots x_k}^2}} \cdot \frac{1}{\sqrt{n - m - 1}},$$

где σ_y – среднее квадратическое отклонение для признака y ;

$R_{yx_1x_2\dots x_k}^2$ – коэффициент детерминации для уравнения множественной регрессии;

σ_{x_i} – среднее квадратическое отклонение для признака x_i ;

$R_{x_ix_1x_2\dots x_k}^2$ – коэффициент детерминации для зависимости фактора x_i со всеми другими факторами уравнения множественной регрессии.

Проверка наличия гетероскедастичности в остатках

При малом объеме выборки для оценки гетероскедастичности может использоваться метод Голдфельда — Квандта.

Данный тест используется для проверки такого типа гетероскедастичности, когда дисперсия остатков возрастает пропорционально квадрату фактора. При этом делается предположение, что, случайная составляющая ε распределена нормально.

Чтобы оценить нарушение гомоскедастичности по тесту Голдфельда - Квандта необходимо выполнить следующие шаги.

1. Упорядочение n наблюдений по мере возрастания переменной x .
2. Разделение совокупности на две группы (соответственно с малыми и большими значениями фактора x) и определение по каждой из групп уравнений регрессии.

3. Определение остаточной суммы квадратов для первой регрессии

$$S_{1\hat{y}} = \sum_{i=1}^{n1} (y_i - \hat{y}_{1i})^2 \quad \text{и второй регрессии} \quad S_{2\hat{y}} = \sum_{i=n-n1+1}^n (y_i - \hat{y}_{2i})^2.$$

4. Вычисление отношений $\frac{S_{2\hat{y}}}{S_{1\hat{y}}}$ (или $\frac{S_{1\hat{y}}}{S_{2\hat{y}}}$). В числителе должна быть большая

сумма квадратов.

Полученное отношение имеет F распределение со степенями свободы $k1=n1-m$ и $k2=n-n1-m$, (m – число оцениваемых параметров в уравнении регрессии).

Если $F_{набл} = \frac{S_{1\hat{y}}}{S_{2\hat{y}}} > F_{кр(\alpha; k1; k2)}$, то гетероскедастичность имеет место.

Чем больше величина F превышает табличное значение F -критерия, тем более нарушена предпосылка о равенстве дисперсий остаточных величин.

Применение фиктивных переменных в регрессионном анализе

При проведении регрессионного анализа часто случается, что отдельные факторы, включенные в модель, являются качественными, и, следовательно, не измеряются в числовой шкале. Один из возможных подходов к решению таких задач – учет влияния качественных факторов посредством введения фиктивных переменных. В качестве фиктивных переменных чаще всего используются булевы переменные, которые принимают только два значения: “0” или “1”. Пусть, например, требуется исследовать зависимость рыночной цены квартиры Y в городе Самара не только от ряда количественных факторов $X_1, X_2, \dots, X_m$, но и от качественного фактора – месторасположения продаваемой квартиры D_1 . Качественный признак D_1 указывает на то, расположена ли квартира достаточно близко к историческому центру города, или же она находится на окраине города. Регрессионная модель цены квартиры может быть записана в виде:

$$Y_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + \dots + b_m x_{im} + c_1 d_{i1} + \varepsilon_i, \quad i = \overline{1, n},$$

где $d_{i1} = \begin{cases} 1, & \text{если } i - \text{ая квартира расположена в центре города;} \\ 0, & \text{если } i - \text{ая квартира расположена на окраине.} \end{cases}$

Если $d_{i1} = 0$, то свободный член регрессии равен b_0 , если же $d_{i1} = 1$, то свободный член регрессии станет $b_0 + c_1$. Коэффициенты уравнения $b_0, b_1, b_2, \dots, b_m, c_1$ можно оценить по МНК. Если МНК-оценка $\hat{c}_1$ окажется статистически значимой, то можно считать, что цена за квартиру в центре города в среднем на $\hat{c}_1$ выше, чем на окраине при неизменных значениях других факторов, включенных в модель.

Если качественный признак имеет q уровней (т.е. количество его различных качественных состояний равно q), то число булевых переменных в модели следует взять равным $q - 1$. Например, месторасположение продаваемой квартиры можно описать $q = 3$ уровнями: первый уровень означает, что квартира находится в центре города; второй уровень – на окраине; третий уровень – в пригороде. В этом случае следует ввести $q - 1 = 3 - 1 = 2$ булевых переменных: D_1 и D_2 . Регрессионная модель цены квартиры примет вид:

$$Y_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + \dots + b_m x_{im} + c_1 d_{i1} + c_2 d_{i2} + \varepsilon_i, \quad i = \overline{1, n},$$

$$\text{где } d_{i1} = \begin{cases} 1, & \text{если } i - \text{ая квартира расположена в центре города;} \\ 0, & \text{во всех остальных случаях;} \end{cases}$$

$$d_{i2} = \begin{cases} 1, & \text{если } i - \text{ая квартира расположена на окраине города;} \\ 0, & \text{во всех остальных случаях.} \end{cases}$$

Если i -ая квартира расположена в центре города, то это будет отражено следующей парой значений фиктивных переменных: $d_{i1} = 1$ и $d_{i2} = 0$; если же i -ая квартира расположена на окраине города, то это будет отражено такой парой значений: $d_{i1} = 0$ и $d_{i2} = 1$; а если i -ая квартира расположена в пригороде, то $d_{i1} = 0$ и $d_{i2} = 0$.

4.2 Типовые примеры и образцы из решения

Пример 1. Пусть имеются следующие данные (условные) о сменной добыче угля на одного рабочего Y (т), мощности пласта X_1 (м) и уровне механизации работ X_2 (%), характеризующие процесс добычи угля в 10 шахтах.

Таблица 4.1 Данные для анализа

№	1	2	3	4	5	6	7	8	9	10
X_1	8	11	12	9	8	8	9	9	8	12
X_2	5	8	8	5	7	8	6	4	5	7
Y	5	10	10	7	5	6	6	5	6	8

Предполагая, что между переменными Y , X_1 , X_2 существует линейная корреляционная зависимость, найдем уравнение регрессии Y по X_1 и X_2 .

Для удобства дальнейших вычислений составляем таблицу ($\varepsilon = Y - Y_x$):

Таблица 4.2 Расчет параметров модели

№	X_1	X_2	Y	X_1^2	X_2^2	Y^2	$X_1 \cdot X_2$	$X_1 \cdot Y$	$X_2 \cdot Y$	Y_x	ε^2
1	2	3	4	5	6	7	8	9	10	11	12
1	8	5	5	64	25	25	40	40	25	5,13	0,016
2	11	8	10	121	64	100	88	110	80	8,79	1,464
3	12	8	10	144	64	100	96	120	80	9,64	0,127
4	9	5	7	81	25	49	45	63	35	5,98	1,038
5	8	7	5	64	49	25	56	40	35	5,86	0,741
6	8	8	6	64	64	36	64	48	48	6,23	0,052
7	9	6	6	81	36	36	54	54	36	6,35	0,121
8	9	4	5	81	16	25	36	45	20	5,61	0,377
9	8	5	6	64	25	36	40	48	30	5,13	0,762
10	12	7	8	144	49	64	84	96	56	9,28	1,631
Сумма	94	63	68	908	417	496	603	664	445	68	6,329
Среднее значение	9,4	6,3	6,8	90,8	41,7	49,6	60,3	66,4	44,5	—	—
σ^2	2,44	2,01	3,36	—	—	—	—	—	—	—	—
σ	1,56	1,42	1,83	—	—	—	—	—	—	—	—

Для нахождения параметров уравнения регрессии в данном случае необходимо решить следующую систему нормальных уравнений:

$$\begin{cases} 10a + 94b_1 + 63b_2 = 68, \\ 94a + 908b_1 + 603b_2 = 664, \\ 63a + 603b_1 + 417b_2 = 445. \end{cases}$$

Откуда получаем, что $a = -3,54$, $b_1 = 0,854$, $b_2 = 0,367$. Т.е. получили следующее уравнение множественной регрессии:

$$y_x = -3,54 + 0,854 \cdot x_1 + 0,367 \cdot x_2.$$

Оно показывает, что при увеличении только мощности пласта X_1 (при неизменном X_2) на 1 м добыча угля на одного рабочего y увеличится в среднем на 0,854 т, а при увеличении только уровня механизации работ X_2 (при неизменном X_1) на 1% – в среднем на 0,367 т.

Найдем уравнение множественной регрессии в стандартизованном масштабе:

$$t_y = \beta_1 t_{x_1} + \beta_2 t_{x_2} + \varepsilon,$$

при этом стандартизованные коэффициенты регрессии будут

$$\beta_1 = b_1 \frac{\sigma_{x_1}}{\sigma_y} = 0,854 \cdot \frac{1,56}{1,83} = 0,728,$$

$$\beta_2 = b_2 \frac{\sigma_{x_2}}{\sigma_y} = 0,367 \cdot \frac{1,42}{1,83} = 0,285.$$

Т.е. уравнение будет выглядеть следующим образом: $\hat{t}_y = 0,728 \cdot t_{x_1} + 0,285 \cdot t_{x_2}$.

Так как стандартизованные коэффициенты регрессии можно сравнивать между собой, то можно сказать, что мощность пласта оказывает большее влияние на сменную добычу угля, чем уровень механизации работ.

Сравнивать влияние факторов на результат можно также при помощи средних коэффициентов эластичности (2.11):

$$\bar{\varepsilon}_i = b_i \cdot \frac{\bar{x}_i}{\bar{y}}. \text{ Вычисляем: } \bar{\varepsilon}_1 = 0,854 \cdot \frac{9,4}{6,8} = 1,18, \quad \bar{\varepsilon}_2 = 0,367 \cdot \frac{6,3}{6,8} = 0,34.$$

Т.е. увеличение только мощности пласта (от своего среднего значения) или только уровня механизации работ на 1% увеличивает в среднем сменную добычу угля на 1,18% или 0,34% соответственно. Таким образом, подтверждается большее влияние на результат y фактора X_1 , чем фактора X_2 .

Оценим качество уравнения, полученного в предыдущем параграфе. Сначала найдем значения парных коэффициентов корреляции:

$$r_{yx_1} = \frac{\overline{y \cdot x_1} - \bar{y} \cdot \bar{x}_1}{\sigma_y \cdot \sigma_{x_1}} = \frac{66,4 - 6,8 \cdot 9,4}{1,83 \cdot 1,56} = 0,869;$$

$$r_{yx_2} = \frac{\overline{y \cdot x_2} - \bar{y} \cdot \bar{x}_2}{\sigma_y \cdot \sigma_{x_2}} = \frac{44,5 - 6,8 \cdot 6,3}{1,83 \cdot 1,42} = 0,639;$$

$$r_{x_1 x_2} = \frac{\overline{x_1 \cdot x_2} - \bar{x}_1 \cdot \bar{x}_2}{\sigma_{x_1} \cdot \sigma_{x_2}} = \frac{60,3 - 9,4 \cdot 6,3}{1,56 \cdot 1,42} = 0,488.$$

Значения парных коэффициентов корреляции указывают на достаточно тесную связь сменной добычи угля на одного рабочего Y с мощностью пласта X_1 и на умеренную связь с уровнем механизации работ X_2 . В то же время межфакторная связь $r_{x_1x_2}$ не очень сильная ($r_{x_1x_2} = 0,49 < 0,7$), что говорит о том, что оба фактора являются информативными, т.е. и X_1 , и X_2 необходимо включить в модель.

Теперь рассчитаем совокупный коэффициент корреляции $R_{yx_1x_2}$. Для этого сначала найдем определитель матрицы парных коэффициентов корреляции:

$$\Delta r = \begin{vmatrix} 1 & 0,87 & 0,64 \\ 0,87 & 1 & 0,49 \\ 0,64 & 0,49 & 1 \end{vmatrix} = 0,139064,$$

и определитель матрицы межфакторной корреляции:

$$\Delta r_{11} = \begin{vmatrix} 1 & 0,49 \\ 0,49 & 1 \end{vmatrix} = 0,7599.$$

Тогда коэффициент множественной корреляции по формуле (2.16):

$$R_{yx_1x_2} = \sqrt{1 - \frac{\Delta r}{\Delta r_{11}}} = \sqrt{1 - \frac{0,139064}{0,7599}} = 0,904.$$

Т.е. можно сказать, что 81,7% (коэффициент детерминации $R_{yx_1x_2}^2 = 0,817$) вариации результата объясняется вариацией представленных в уравнении признаков, что указывает на весьма тесную связь признаков с результатом.

Примерно тот же результат (различия связаны с ошибками округлений) для коэффициента множественной регрессии получим, если воспользуемся формулами (2.12) и (2.15):

$$R_{yx_1x_2} = \sqrt{1 - \frac{\sigma_{\text{ост}}^2}{\sigma_y^2}} = \sqrt{1 - \frac{0,6329}{3,36}} = 0,901;$$

$$R_{yx_1x_2} = \sqrt{\sum \beta_i \cdot r_{yx_i}} = \sqrt{0,728 \cdot 0,87 + 0,285 \cdot 0,64} = 0,903.$$

Скорректированный коэффициент множественной детерминации

$$R = 1 - (1 - R^2) \cdot \frac{n-1}{n-m-1} = 1 - (1 - 0,817) \cdot \frac{10-1}{10-2-1} = 0,765$$

указывает на умеренную связь между результатом и признаками. Это связано с малым количеством наблюдений.

Оценим надежность уравнения регрессии в целом и показателя связи с помощью F -критерия Фишера. Фактическое значение F -критерия (2.22)

$$F_{\text{факт}} = \frac{R^2}{1-R^2} \cdot \frac{n-m-1}{m} = \frac{0,817}{1-0,817} \cdot \frac{10-2-1}{2} = 15,63.$$

Табличное значение F -критерия при пятипроцентном уровне значимости ($\alpha = 0,05$, $k_1 = 2$, $k_2 = 10 - 2 - 1 = 7$): $F_{\text{табл}} = 4,74$. Так как $F_{\text{факт}} = 15,63 > F_{\text{табл}} = 4,10$, то уравнение признается статистически значимым. Уравнение регрессии, включающее только один значимый аргумент X_2 : $y = -2,754 + 1,016x_1$.

Пример 2. В базе данных магазина, торгующего поддержанными автомобилями, содержится информация об их потребительских свойствах и ценах.

Для анализа зависимости цены автомобиля Y от его возраста X_1 и мощности двигателя X_2 из базы данных выбраны сведения о 16 автомобилях. Эти сведения приведены в таблице 4.3

Таблица 4.3 Данные для анализа

Номер автомобиля №	Цена (тыс. у.е.) y_i	Возраст (лет) x_{i1}	Мощность двигателя (л.с.) x_{i2}
1	9,0	5,0	119
2	8,6	4,0	89
3	7,4	6,0	114
4	4,4	7,0	54
5	5,6	7,0	82
6	4,2	8,0	78
7	10,3	3,0	104
8	9,3	5,0	145
9	5,5	7,0	83
10	8,4	4,0	84
11	9,7	4,0	139
12	11,0	3,0	147
13	10,4	3,0	132
14	9,4	4,0	123
15	9,3	4,0	112
16	10,8	3,0	125

1. По методу наименьших квадратов найти оценки коэффициентов множественной линейной регрессионной модели $Y = a_0 + a_1X_1 + a_2X_2 + \varepsilon$.

2. Проверить статистическую значимость параметров и уравнения множественной регрессии с точностью 0,9.

3. Рассчитать точечный и интервальный прогноз среднего значения цены поступивших автомобилей возраста 3 года и мощностью двигателя 165 л.с. с доверительной вероятностью 0,95.

Решение: Нахождение оценок коэффициентов множественной линейной регрессионной модели по методу наименьших квадратов в матричном виде (таблица 4.4).

$$X^T X = \begin{pmatrix} 16 & 77 & 1730 \\ 77 & 413 & 7873 \\ 1730 & 7873 & 198240 \end{pmatrix}, X^T Y = \begin{pmatrix} 133,3 \\ 587,1 \\ 15172,1 \end{pmatrix}.$$

Найдем матрицу $(X^T X)^{-1}$, обратную к матрице $X^T X$.

Главный определитель:

$$|X^T X| = 16 \cdot 413 \cdot 198240 + 77 \cdot 7873 \cdot 1730 + 77 \cdot 7873 \cdot 1730 - 1730 \cdot 413 \cdot 1730 - 77 \cdot 77 \cdot 198240 - 7873 \cdot 7873 \cdot 16 = 4\,315\,856,00.$$

Таблица 4.4 Расчет параметров модели

N_2	y_i	x_{i1}	x_{i1}^2	$x_{i1} * y_i$	x_{i2}	x_{i2}^2	$x_{i2} * y_i$	$x_{i1} * x_{i2}$
1	9,0	5,0	25,0	45,0	119,0	14161,0	1071,0	595
2	8,6	4,0	16,0	34,4	89,0	7921,0	765,4	356
3	7,4	6,0	36,0	44,4	114,0	12996,0	843,6	684
4	4,4	7,0	49,0	30,8	54,0	2916,0	237,6	378
5	5,6	7,0	49,0	39,2	82,0	6724,0	459,2	574
6	4,2	8,0	64,0	33,6	78,0	6084,0	327,6	624
7	10,3	3,0	9,0	30,9	104,0	10816,0	1071,2	312
8	9,3	5,0	25,0	46,5	145,0	21025,0	1348,5	725
9	5,5	7,0	49,0	38,5	83,0	6889,0	456,5	581
10	8,4	4,0	16,0	33,6	84,0	7056,0	705,6	336
11	9,7	4,0	16,0	38,8	139,0	19321,0	1348,3	556
12	11,0	3,0	9,0	33,0	147,0	21609,0	1617,0	441
13	10,4	3,0	9,0	31,2	132,0	17424,0	1372,8	396
14	9,4	4,0	16,0	37,6	123,0	15129,0	1156,2	492
15	9,3	4,0	16,0	37,2	112,0	12544,0	1041,6	448
16	10,8	3,0	9,0	32,4	125,0	15625,0	1350,0	375
Сумма	133,3	77,0	413,0	587,1	1730,0	198240,0	15172,1	7873

Определим матрицу алгебраических дополнений:

$$D = \begin{pmatrix} D_{11} & D_{12} & D_{13} \\ D_{21} & D_{22} & D_{23} \\ D_{31} & D_{32} & D_{33} \end{pmatrix}, \text{ где } D_{ij} = (-1)^{i+j} M_{ij}. M_{ij} - \text{это минор элемента, стоящего на пересечении } i\text{-й строки и } j\text{-го столбца матрицы } X^T X.$$

ресечения i -й строки и j -го столбца матрицы $X^T X$.

$$D_{11} = (-1)^{1+1} \begin{vmatrix} 413 & 7873 \\ 7873 & 198240 \end{vmatrix} = 19\,888\,991,0.$$

$$D_{12} = (-1)^{1+2} \begin{vmatrix} 77 & 7873 \\ 1730 & 198240 \end{vmatrix} = -1\,644\,190,0.$$

$$D_{13} = (-1)^{1+3} \begin{vmatrix} 77 & 413 \\ 1730 & 7873 \end{vmatrix} = -108\,269,0.$$

$$D_{21} = (-1)^{2+1} \begin{vmatrix} 77 & 1730 \\ 7873 & 198240 \end{vmatrix} = -1\,644\,190,0.$$

$$D_{22} = (-1)^{2+2} \begin{vmatrix} 16 & 1730 \\ 1730 & 198240 \end{vmatrix} = 178\,940,0.$$

$$D_{23} = (-1)^{2+3} \begin{vmatrix} 16 & 77 \\ 1730 & 7873 \end{vmatrix} = 7\,242,0.$$

$$D_{31} = (-1)^{3+1} \begin{vmatrix} 77 & 1730 \\ 413 & 7873 \end{vmatrix} = -108\,269,0.$$

$$D_{32} = (-1)^{3+2} \begin{vmatrix} 16 & 1730 \\ 77 & 7873 \end{vmatrix} = 7242,0.$$

$$D_{33} = (-1)^{3+3} \begin{vmatrix} 16 & 77 \\ 77 & 413 \end{vmatrix} = 679,0.$$

$$D = \begin{pmatrix} 19888991 & -1644190 & -108269 \\ -1644190 & 178940 & 7242 \\ -108269 & 7242 & 679 \end{pmatrix}.$$

Составляем присоединенную матрицу.

$$(X^T X)^* = D^T = \begin{pmatrix} 19888991 & -1644190 & -108269 \\ -1644190 & 178940 & 7242 \\ -108269 & 7242 & 679 \end{pmatrix}.$$

$$(X^T X)^{-1} = \frac{(X^T X)^*}{|X^T X|} = \frac{1}{4315856} * \begin{pmatrix} 19888991 & -1644190 & -108269 \\ -1644190 & 178940 & 7242 \\ -108269 & 7242 & 679 \end{pmatrix} =$$

$$\begin{pmatrix} 4,60835 & -0,38096 & -0,02509 \\ -0,38096 & 0,041461 & 0,001678 \\ -0,02509 & 0,001678 & 0,000157 \end{pmatrix}.$$

Находим вектор оценок А:

$$A = (X^T X)^{-1} * X^T Y = \begin{pmatrix} 4,60835 & -0,38096 & -0,02509 \\ -0,38096 & 0,041461 & 0,001678 \\ -0,02509 & 0,001678 & 0,000157 \end{pmatrix} * \begin{pmatrix} 133,3 \\ 587,1 \\ 15172,1 \end{pmatrix} =$$

$$\begin{pmatrix} 4,60835 * 133,3 + (-0,38096) * 587,1 + (-0,02509) * 15172,1 \\ (-0,38096) * 133,3 + 0,041461 * 587,1 + 0,001678 * 15172,1 \\ (-0,02509) * 133,3 + 0,001678 * 587,1 + 0,000157 * 15172,1 \end{pmatrix} \approx \begin{pmatrix} 9,9640 \\ -0,9814 \\ 0,0227 \end{pmatrix}.$$

Таким образом, оценки коэффициентов множественной линейной регрессионной модели: $\hat{y} = 9,9640 - 0,9814x_1 + 0,0227x_2$.

2.2. Проверка статистической значимости параметров и уравнения множественной линейной регрессии с надежностью 0,9 (таблица 4.5).

$$\hat{y} = 9,9640 - 0,9814x_1 + 0,0227x_2. \quad \bar{y} = 8,3313$$

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}. \quad \sum(y_i - \bar{y})^2 = \sum y_i^2 - n\bar{y}^2 = 1186,21 - 16 * (8,3313)^2 = 75,6410$$

$$\text{Коэффициент детерминации: } R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2} = 1 - \frac{7,6767}{75,6410} = 0,8985.$$

Коэффициент детерминации = 0,8985. Следовательно, регрессия у на x_1 и x_2 объясняет 89,85% колебаний значений у. Это свидетельствует о значительном суммарном влиянии независимых переменных x_1 и x_2 на зависимую переменную у.

Таблица 4.5 Оценка качества модели

№	y_i	x_{i1}	x_{i2}	y_i^2	$\hat{y}_i$	e_i	e_i^2
1	9,0	5,0	119	81	7,7583	1,2417	1,5418
2	8,6	4,0	89	73,96	8,0587	0,5413	0,2930
3	7,4	6,0	114	54,76	6,6634	0,7366	0,5426
4	4,4	7,0	54	19,36	4,3200	0,0800	0,0064
5	5,6	7,0	82	31,36	4,9556	0,6444	0,4153
6	4,2	8,0	78	17,64	3,8834	0,3166	0,1002
7	10,3	3,0	104	106,09	9,3806	0,9194	0,8453
8	9,3	5,0	145	86,49	8,3485	0,9515	0,9054
9	5,5	7,0	83	30,25	4,9783	0,5217	0,2722
10	8,4	4,0	84	70,56	7,9452	0,4548	0,2068
11	9,7	4,0	139	94,09	9,1937	0,5063	0,2563
12	11,0	3,0	147	121	10,3567	0,6433	0,4138
13	10,4	3,0	132	108,16	10,0162	0,3838	0,1473
14	9,4	4,0	123	88,36	8,8305	0,5695	0,3243
15	9,3	4,0	112	86,49	8,5808	0,7192	0,5172
16	10,8	3,0	125	116,64	9,8573	0,9427	0,8887
Сумма	133,3	77	1730	1186,21	123,127	10,1728	7,6767

Выдвигаем гипотезу H_0 о статистической незначимости уравнения регрессии и показателя тесноты связи. Фактическое значение F-статистики Фишера:

$$F_{\Phi} = \frac{R^2/m}{R^2/(n-m-1)}$$

$$df_1 = m = 2, df_2 = n-m-1 = 16-2-1=13.$$

$$F_{\Phi} = \frac{0,8985/2}{0,8985/(16-2-1)}=6,5.$$

Статистической надежности 0,9 соответствует уровень значимости $(\alpha) = 0,1$. При уровне значимости 0,1, $df_1 = 2, df_2 = 13$ табличное значение $F_m = 2,7632$.

$$F_m < F_{\Phi} [2,7632 < 6,5].$$

Поскольку неравенство $F_m < F_{\Phi}$ выполняется, то гипотеза $H_0: \alpha_1 = 0, \alpha_2 = 0$ отклоняется и признается статистическая значимость уравнения регрессии.

2.3. Точечный и интервальный прогноз среднего значения цены поступивших автомобилей возраста 3 года и мощностью двигателя 165 л.с. с доверительной вероятностью 0,95.

а) Под *точечным прогнозом* среднего значения цены новой партии автомобилей понимается значение $\bar{y}(x^p) = \alpha_0 + \alpha_1 x_1^p + \alpha_2 x_2^p$, где $x^p = (1, x_1^p, x_2^p)$ – вектор независимых переменных, для которого определяется прогноз. $x_1^p = 3$ года – это возраст автомобиля. $x_2^p = 165$ л.с. – это мощность двигателя.

Определяем вектор независимых переменных $x^p = (1; 3; 165)$.

Находим *точечный прогноз*:

$$\bar{y}(x^p) = 9,9640 - 0,9814x_1^p + 0,0227x_2^p = 9,9640 - 0,9814*3 + 0,0227*165 = 10,7653 \text{ тыс. у.е.}$$

б) Под *интервальным прогнозом* среднего значения цены автомобилей понимается доверительный интервал цены, который находится по формуле:

$$\hat{y}_{B,H} = \hat{y}(x^p) \pm t_{1-\alpha/2, n-3} * S_{\hat{y}},$$

где $\hat{y}_B$ - верхняя граница доверительного интервала, $\hat{y}_H$ – нижняя граница доверительного интервала; $x^p = (1, x_1^p, x_2^p)$ - вектор независимых переменных, для которого определяется интервал.

$$S_{\hat{y}} = S\sqrt{x^p(X^T X)^{-1}(x^p)^T}, S = \sqrt{S^2}, S^2 = \frac{\sum(y_i - \hat{y}_i)^2}{n-3}.$$

Находим *интервальный прогноз*:

$$x^p(X^T X)^{-1}(x^p)^T = (1; 3; 165) \begin{pmatrix} 4,608354 & -0,38096 & -0,02509 \\ -0,38096 & 0,041461 & 0,001678 \\ -0,02509 & 0,001678 & 0,000157 \end{pmatrix} \begin{pmatrix} 1 \\ 3 \\ 165 \end{pmatrix} =$$

$$(-0,6744; 0,0203; 0,0058) \begin{pmatrix} 1 \\ 3 \\ 165 \end{pmatrix} = 0,5085.$$

$$S^2 = \frac{\sum(y_i - \hat{y}_i)^2}{n-3} = \frac{7,6767}{16-3} = 0,5905.$$

$$S = \sqrt{S^2} = \sqrt{0,5905} = 0,7684.$$

$$S_{\hat{y}} = S\sqrt{x^p(X^T X)^{-1}(x^p)^T} = 0,7684 * \sqrt{0,5085} = 0,5480.$$

Доверительная вероятность = 0,95. Тогда уровень значимости = 0,05.

$$t_{1-\alpha/2, n-3} = t_{1-0,05/2, 16-3} = t_{0,975, 13} = 2,1604.$$

$$\hat{y}_B = \hat{y}(x^p) \pm t_{1-\alpha/2, n-3} * S_{\hat{y}} = 10,7653 + (2,1604 * 0,5480) = 11,9492 \text{ тыс. у.е.}$$

$$\hat{y}_H = \hat{y}(x^p) \pm t_{1-\alpha/2, n-3} * S_{\hat{y}} = 10,7653 - (2,1604 * 0,5480) = 9,5814 \text{ тыс. у.е.}$$

Экономическая интерпретация

В результате исследования зависимости объема цены от возраста и мощности двигателя получено уравнение множественной линейной регрессии: $\hat{y} = 9,9640 - 0,9814x_1 + 0,0227x_2$.

Величина $\alpha_1 = -0,9814$ показывает, что при увеличении возраста на 1 год и неизменной мощности двигателя следует ожидать снижения цены автомобиля на 0,9814 тыс. у.е.

Коэффициент $\alpha_2 = 0,0227$ показывает, что при увеличении мощности двигателя на 1 л.с. и фиксированном возрасте автомобиля следует ожидать увеличения цены на 0,0227 тыс. у.е.

1.1. Выдвинуты гипотезы о том, что зависимость цены y от возраста автомобиля x_1 описывается линейной моделью вида: $y = a_0 + a_1x_1 + \varepsilon$, где a_0, a_1 – неизвестные постоянные коэффициенты, ε – случайная переменная, отражающая влияние неучтенных факторов от погрешностей измерений; зависимость цены y от мощности автомобиля x_2 описывается линейной моделью вида: $y = \beta_0 + \beta_1x_2 + \delta$, где β_0, β_1 – неизвестные постоянные коэффициенты, δ – случайная переменная, отражающая влияние неучтенных факторов от погрешностей измерений.

1.2. Методом наименьших квадратов найдены оценки линейных уравнений регрессии: $\hat{y} = 14,5014 - 1,2821x_1$; $\hat{y} = 0,9896 + 0,0679x_2$.

1.3. Коэффициент корреляции r_{yx_1} существенно отличается от нуля, и существует сильная линейная отрицательная связь между ценой и возрастом автомобиля.

Коэффициент корреляции r_{yx_2} существенно отличается от нуля, и существует сильная линейная положительная связь между ценой автомобиля и мощностью двигателя.

2.1. Мо методом наименьших квадратов найдены оценки коэффициентов множественной линейной регрессии: $\hat{y} = 9,9640 - 0,9814x_1 + 0,0227x_2$.

2.2. Коэффициент детерминации = 0,8985. Следовательно, регрессия y на x_1 и x_2 объясняет 89,85% колебаний значений y . Это свидетельствует о значительном суммарном влиянии независимых переменных x_1 и x_2 на зависимую переменную y . С использованием F – статистики признается статистическая значимость уравнения регрессии.

2.3. Точечный прогноз: $\bar{y}(x^p) = 10,7653$ тыс. у.е. Интервальный прогноз:

Верхняя граница доверительного интервала: $\hat{y}_B = 11,9492$ тыс. у.е.

Нижняя граница доверительного интервала: $\hat{y}_H = 9,5814$ тыс. у.е..

4.3 Задания на практику

1. На основе приведенных ниже данных (табл.4.6) для объясняемой переменной y и потенциальных объясняющих переменных x_1, x_2 рассчитать коэффициенты парной корреляции переменной y с переменными x_1, x_2 , а также переменной x_1 с переменной x_2 . Сделайте выводы.

Таблица 4.6 данные для анализа

номер	y	x_1	x_2
1	1,9	11	1,1
2	2,1	13	1,2
3	2,5	21	1,4
4	2,7	31	2,1
5	3,2	29	2,2
6	2,0	12	0,9

2. Для некоторой модели получена следующая последовательность остатков (табл. 4.7):

Таблица 4.7

i	1	2	3	4	5	6
e_i	-5	10	-1	-5	2	-1

Известно, что дисперсия объясняемой переменной $\sigma_y = 500$. Чему равен коэффициент детерминации?

3. По 20 предприятиям региона изучается зависимость выработки продукции на одного работника y (тыс.руб.) от ввода в действие новых основных фондов x_1 (% от стоимости фондов на конец года) и от удельного веса рабочих высокой квалификации в общей численности рабочих x_2 (%).

$$y = -1,8353 + 0,9459x_1 + 0,0856x_2, \text{ где } y = 9,6; x_1 = 6,19; x_2 = 22,3$$

Задание:

- 1.) Определить с помощью коэффициентов эластичности силу влияния каждого фактора на результат.
- 2.) Ранжировать факторы по силе влияния.

4. Для маркетингового исследования мелкой бытовой техники в компании **Redmond**, мирового лидера среди производителей бытовой техники, по 20 магазинам (предприятиям) изучается зависимость объема продаж мультиварок Y (млн. руб.) от цены мультиварки X_1 (тыс. руб.) и бюджета рекламной компании X_2 (тыс. руб.), проводимой в магазинах. (таблица 4.8)

1. Рассчитайте матрицу линейных коэффициентов корреляции для всех переменных X_1 , X_2 и Y . Сделайте выводы о наличии мультиколлинеарности между факторами X_1 , X_2 и целесообразности включения их в модель.
2. Рассчитайте параметры линейного уравнения множественной регрессии с полным перечнем факторов.
3. Записать стандартизированное уравнение регрессии и рассчитать средние коэффициенты эластичности для каждого фактора. На основе стандартизированных коэффициентов и коэффициентов эластичности сделать выводы о степени влияния факторов на результативный признак Y .
4. Найти скорректированный коэффициент множественной детерминации и сделать по нему выводы.
5. С помощью F- критерия Фишера и t-критерия Сьюдента оценить значимость коэффициентов модели и всего уравнения регрессии, сделать выводы.
6. Оцените качество уравнения через среднюю ошибку аппроксимации.
7. Проверить вычисления в Excel с помощью пакета анализа.

Таблица 4.8 Данные для анализа

№	y	x_1	x_2
1	7	4,2	10
2	7	4,3	14
3	7	4,4	15
4	7	4,7	16
5	8	5	17
6	9	5,8	19
7	10	5,9	19
8	11	6,1	20
9	11	6,2	20

10	11	6,8	20
11	11	6,9	21
12	12	7,6	22
13	13	8,3	22
14	13	8,4	25
15	13	8,4	28
16	13	8,6	29
17	13	8,7	30
18	14	8,8	31
19	14	9	32
20	14	9	36

4.4 Семестровая работа 4 «Множественная регрессия»

Имеются данные по парку машин (таблица 4.9)

1. Рассчитайте матрицу линейных коэффициентов корреляции для всех переменных X_1 , X_2 и Y . Сделайте выводы о наличии мультиколлинеарности между факторами X_1 , X_2 и целесообразности включения их в модель.
2. Рассчитайте параметры линейного уравнения множественной регрессии с полным перечнем факторов.
3. Записать стандартизированное уравнение регрессии и рассчитать средние коэффициенты эластичности для каждого фактора. На основе стандартизированных коэффициентов и коэффициентов эластичности сделать выводы о степени влияния факторов на результирующий признак Y .
4. Найти скорректированный коэффициент множественной детерминации и сделать по нему выводы.
5. С помощью F- критерия Фишера и t-критерия Сьюдента оценить значимость коэффициентов модели и всего уравнения регрессии, сделать выводы.
6. Оцените качество уравнения через среднюю ошибку аппроксимации.
7. Проверить вычисления в Excel с помощью пакета анализа.
8. Построить модель с одним наиболее значимым фактором и сравнить ее с множественной моделью. Сделайте выводы.
9. Рассчитайте прогнозное значение результата по наиболее лучшей модели и, если прогнозные значения факторов составляют 80 % от их максимальных значений. (Максимум найти с помощью пакета анализа (описательные статистики))
10. Проведите выполнение всех предпосылок и оценку гетероскедастичности остатков с помощью критерия Годфрида-Кванта и оценки его значимости по модели с наиболее значимым фактором сделайте выводы.
11. Оцените полученные результаты и сделайте выводы в отчете.

Таблица 4.9 Данные для анализа

№ компа нии	Стоимость грузовой автомобильной перевозки у (тыс. руб.)									
	№ варианта									
	1	2	3	4	5	6	7	8	9	10
1	51,	56,1	20,4	84,2	38,8	127,5	71,4	182,2	83,8	89,8
2	16,	17,6	6,4	26,4	12,2	40,0	22,4	57,2	26,3	28,2
3	74,	81,4	29,6	122,1	56,2	185,0	103,6	264,3	121,6	130,2
4	7,5	8,3	3,0	12,4	5,7	18,8	10,5	26,8	12,3	13,2
5	33,	36,3	13,2	54,5	25,1	82,5	46,2	117,9	54,2	58,1
6	26,	28,6	10,4	42,9	19,8	65,0	36,4	92,9	42,7	45,8
7	11,	12,7	4,6	19,0	8,7	28,8	16,1	41,1	18,9	20,2
8	52,	57,2	20,8	85,8	39,5	130,0	72,8	185,7	85,4	91,5
9	15,	17,4	6,3	26,1	12,0	39,5	22,1	56,4	26,0	27,8
10	8,0	8,8	3,2	13,2	6,1	20,0	11,2	28,6	13,1	14,1
11	26,	28,6	10,4	42,9	19,8	65,0	36,4	92,9	42,7	45,8
12	6,0	6,6	2,4	9,9	4,6	15,0	8,4	21,4	9,9	10,6
13	5,8	6,4	2,3	9,6	4,4	14,5	8,1	20,7	9,5	10,2
14	13,	15,2	5,5	22,8	10,5	34,5	19,3	49,3	22,7	24,3
15	6,2	6,8	2,5	10,2	4,7	15,5	8,7	22,1	10,2	10,9
16	7,9	8,7	3,2	13,0	6,0	19,8	11,1	28,2	13,0	13,9
17	5,4	5,9	2,2	8,9	4,1	13,5	7,6	19,3	8,9	9,5
18	56,	61,6	22,4	92,4	42,6	140,0	78,4	200,0	92,0	98,6
19	25,	28,1	10,2	42,1	19,4	63,8	35,7	91,1	41,9	44,9
20	7,1	7,8	2,8	11,7	5,4	17,8	9,9	25,4	11,7	12,5
	Вес груза x1 (тонн)									
1	35,	45,5	14,0	54,6	26,6	91,0	54,6	125,0	61,3	72,8
2	16,	20,8	6,4	25,0	12,2	41,6	25,0	57,2	28,0	33,3
3	18,	23,4	7,2	28,1	13,7	46,8	28,1	64,3	31,5	37,4
4	2,0	2,6	0,8	3,1	1,5	5,2	3,1	7,1	3,5	4,2
5	14,	18,2	5,6	21,8	10,6	36,4	21,8	50,0	24,5	29,1
6	33,	42,9	13,2	51,5	25,1	85,8	51,5	117,9	57,8	68,6
7	20,	26,0	8,0	31,2	15,2	52,0	31,2	71,4	35,0	41,6
8	25,	32,5	10,0	39,0	19,0	65,0	39,0	89,3	43,8	52,0
9	13,	16,9	5,2	20,3	9,9	33,8	20,3	46,4	22,8	27,0
10	2,0	2,6	0,8	3,1	1,5	5,2	3,1	7,1	3,5	4,2
11	21,	27,3	8,4	32,8	16,0	54,6	32,8	75,0	36,8	43,7
12	11,	14,3	4,4	17,2	8,4	28,6	17,2	39,3	19,3	22,9
13	3,0	3,9	1,2	4,7	2,3	7,8	4,7	10,7	5,3	6,2
14	3,5	4,6	1,4	5,5	2,7	9,1	5,5	12,5	6,1	7,3
15	2,8	3,6	1,1	4,4	2,1	7,3	4,4	10,0	4,9	5,8
16	17,	22,1	6,8	26,5	12,9	44,2	26,5	60,7	29,8	35,4
17	3,4	4,4	1,4	5,3	2,6	8,8	5,3	12,1	6,0	7,1
18	24,	31,2	9,6	37,4	18,2	62,4	37,4	85,7	42,0	49,9
19	9,0	11,7	3,6	14,0	6,8	23,4	14,0	32,1	15,8	18,7
20	4,5	5,9	1,8	7,0	3,4	11,7	7,0	16,1	7,9	9,4

	Расстояние χ_2 (тыс. км)									
	№ варианта									
	1	2	3	4	5	6	7	8	9	10
1	2,0	2,4	0,6	3,8	1,1	5,4	2,2	5,1	2,5	3,8
2	1,1	1,3	0,3	2,1	0,6	3,0	1,2	2,8	1,4	2,1
3	2,6	3,1	0,8	4,9	1,5	6,9	2,8	6,5	3,1	4,9
4	1,7	2,0	0,5	3,3	1,0	4,6	1,8	4,4	2,1	3,3
5	2,4	2,9	0,7	4,6	1,4	6,5	2,6	6,2	3,0	4,6
6	1,6	1,9	0,5	3,0	0,9	4,2	1,7	4,0	1,9	3,0
7	0,6	0,7	0,2	1,2	0,3	1,6	0,6	1,5	0,7	1,2
8	2,3	2,8	0,7	4,4	1,3	6,2	2,5	5,9	2,8	4,4
9	1,4	1,7	0,4	2,7	0,8	3,8	1,5	3,6	1,7	2,7
10	2,1	2,5	0,6	4,0	1,2	5,7	2,3	5,4	2,6	4,0
11	1,3	1,6	0,4	2,5	0,7	3,5	1,4	3,3	1,6	2,5
12	0,4	0,4	0,1	0,7	0,2	0,9	0,4	0,9	0,4	0,7
13	1,7	2,0	0,5	3,2	0,9	4,5	1,8	4,2	2,0	3,2
14	2,9	3,5	0,9	5,6	1,7	7,8	3,1	7,4	3,6	5,6
15	0,8	0,9	0,2	1,4	0,4	2,0	0,8	1,9	0,9	1,4
16	0,6	0,7	0,2	1,2	0,3	1,6	0,6	1,5	0,7	1,2
17	0,9	1,1	0,3	1,7	0,5	2,4	1,0	2,3	1,1	1,7
18	2,5	3,0	0,8	4,8	1,4	6,8	2,7	6,4	3,1	4,8
19	2,2	2,6	0,7	4,2	1,3	5,9	2,4	5,6	2,7	4,2
20	1,0	1,1	0,3	1,8	0,5	2,6	1,0	2,4	1,2	1,8

Тема 5 МОДЕЛИРОВАНИЕ ВРЕМЕННЫХ РЯДОВ

5.1 Краткий теоретический материал

Временной ряд (ряд динамики) – это совокупность значений какого-либо показателя за несколько последовательных моментов или периодов времени. Каждый уровень временного ряда формируется под воздействием большого числа факторов, которые условно можно подразделить на три группы:

- 1) факторы, формирующие тенденцию ряда;
- 2) факторы, формирующие циклические колебания ряда;
- 3) случайные факторы.

Для выявления специфики развития изучаемых явлений за отдельные периоды времени определяют:

- ✓ абсолютные приросты уровней ряда;
- ✓ относительные приросты уровней ряда, т.е. *темпы роста*;
- ✓ темпы прироста.

1. Для выражения абсолютной скорости роста или снижения уровней ряда вычисляют *абсолютный прирост*. Его величина определяется как разность двух сравниваемых уровней. Так, для цепных приростов, используется формула:

$$\Delta_t = y_{t+1} - y_t.$$

2. Интенсивность изменения уровней ряда оценивается отношением текущего уровня к предыдущему или базисному. Этот показатель принято называть *темпом роста*:

$$T_t = \frac{y_{t+1}}{y_t}.$$

3. Для выражения изменения величины абсолютного прироста уровней ряда в относительных величинах используется *темп прироста*, который рассчитывается как отношение абсолютного прироста к предыдущему или к базисному уровню:

$$T_{np} = \frac{\Delta_{t+1}}{y_t} \text{ или } T_{np} = \frac{\Delta_{t+1}}{y_1}.$$

Также справедлива формула $T_{np} = T_y - 1$.

Особое внимание уделяют расчетам средних показателей рядов динамики, среди них различают:

- ✓ средний уровень $\bar{y} = \frac{1}{n} \sum y_t$;
- ✓ средний абсолютный прирост $\bar{\Delta} = \frac{1}{n-1} \sum \Delta_t$;
- ✓ средний темп роста $\bar{T}_p = \sqrt[m]{T_1 \cdot T_2 \cdot \dots \cdot T_m}$, где m – число периодов, по которым вычисляется среднее.
- ✓ средний темп прироста $\bar{T}_{np} = \bar{T}_p - 1$.

При наличии во временном ряде тенденции и циклических колебаний значения каждого последующего уровня ряда зависят от предыдущих. Корреляционную зависимость между последовательными уровнями временного ряда называют автокорреляцией уровней ряда.

Количественно ее можно измерить с помощью линейного коэффициента корреляции между уровнями исходного временного ряда и уровнями этого ряда, сдвинутыми на несколько шагов во времени.

Формула для расчета коэффициента автокорреляции имеет вид:

$$r_1 = \frac{\sum_{t=2}^n (y_t - \bar{y}_1)(y_{t-1} - \bar{y}_2)}{\sqrt{\sum_{t=2}^n (y_t - \bar{y}_1)^2 \sum_{t=2}^n (y_{t-1} - \bar{y}_2)^2}},$$

где

$$\bar{y}_1 = \frac{1}{n-1} \sum_{t=2}^n y_t, \quad \bar{y}_2 = \frac{1}{n-1} \sum_{t=2}^n y_{t-1}.$$

Эту величину называют коэффициентом автокорреляции уровней ряда первого порядка, так как он измеряет зависимость между соседними уровнями ряда t и y_{t-1} .

Аналогично можно определить коэффициенты автокорреляции второго и более высоких порядков. Так, коэффициент автокорреляции второго порядка характеризует тесноту связи между уровнями y_t и y_{t-2} и определяется по формуле:

$$r_2 = \frac{\sum_{t=3}^n (y_t - \bar{y}_3)(y_{t-2} - \bar{y}_4)}{\sqrt{\sum_{t=3}^n (y_t - \bar{y}_3)^2 \sum_{t=3}^n (y_{t-2} - \bar{y}_4)^2}},$$

где

$$\bar{y}_3 = \frac{1}{n-2} \sum_{t=3}^n y_t, \quad \bar{y}_4 = \frac{1}{n-2} \sum_{t=3}^n y_{t-2}.$$

Число периодов, по которым рассчитывается коэффициент автокорреляции, называют *лагом*. С увеличением лага число пар значений, по которым рассчитывается коэффициент автокорреляции, уменьшается. Считается целесообразным для обеспечения статистической достоверности коэффициентов автокорреляции использовать правило – максимальный лаг должен быть не больше $n/4$.

Свойства коэффициента автокорреляции.

1. Он строится по аналогии с линейным коэффициентом корреляции и таким образом характеризует тесноту только линейной связи текущего и предыдущего уровней ряда. Поэтому по коэффициенту автокорреляции можно судить о наличии линейной (или близкой к линейной) тенденции. Для некоторых временных рядов, имеющих сильную нелинейную тенденцию (например, параболу второго порядка или экспоненту), коэффициент автокорреляции уровней исходного ряда может приближаться к нулю.

2. По знаку коэффициента автокорреляции нельзя делать вывод о возрастающей или убывающей тенденции в уровнях ряда. Большинство временных рядов экономических данных содержат положительную автокорреляцию уровней, однако при этом могут иметь убывающую тенденцию.

Последовательность коэффициентов автокорреляции уровней первого, второго и т.д. порядков называют *автокорреляционной функцией* временного ряда. График зависимости ее значений от величины лага (порядка коэффициента автокорреляции) называется *коррелограммой*.

Анализ автокорреляционной функции и коррелограммы позволяет определить лаг, при котором связь между текущим и предыдущим уровнями наиболее тесная. Причем,

- ✓ если наиболее высоким оказался коэффициент автокорреляции первого порядка, то исследуемый ряд содержит только тренд и не содержит сезонных колебаний;
- ✓ если временной ряд имеет линейную тенденцию, то его соседние уровни y_t и y_{t-1} тесно коррелируют, т.е. $r_1 > 0,9$;
- ✓ если временной ряд содержит сильную нелинейную тенденцию, например, в форме экспоненты, то коэффициент автокорреляции первого порядка по логарифмам исходного уровня будет выше, чем соответствующий коэффициент, подсчитанный по непреобразованным уровням ряда;
- ✓ если наиболее высоким оказался коэффициент автокорреляции порядка τ , ряд содержит циклические колебания с периодичностью в τ моментов времени;
- ✓ если ни один из коэффициентов автокорреляции не является значимым, можно сделать предположение относительно структуры этого ряда: либо ряд содержит только случайную компоненту, либо содержит сильную нелинейную тенденцию, для выявления которой нужно провести дополнительный анализ.

Выбор уравнения тренда

При выборе уравнения тренда необходимо руководствоваться принципом простоты. Чем сложнее уравнение линии тренда и чем большее число параметров содержит, тем при равной степени приближения труднее дать надежную оценку этим параметрам.

На практике чаще всего используют следующие основные виды трендов временных рядов:

- ✓ линейный тренд $\tilde{y}_t = a + bt$;
- ✓ гипербола $\tilde{y}_t = a + \frac{b}{t}$;
- ✓ параболический тренд $\tilde{y}_t = a + b_1t + b_2t^2$;
- ✓ экспоненциальный тренд $\tilde{y}_t = a \cdot b^t$ или $\tilde{y}_t = e^{a+bt}$;
- ✓ тренд в форме степенной функции $\tilde{y}_t = a \cdot t^b$;
- ✓ логарифмический тренд $\tilde{y}_t = a + b \cdot \ln t$;

Для правильного выбора типа тренда, который наилучшим способом отражает тенденцию фактического ряда уровней, следует руководствоваться следующим:

- ✓ Построить график ряда в правильно выбранном масштабе.
- ✓ Провести сглаживание ряда методом скользящих средних.
- ✓ Линейный тип тренда подходит для отображения тенденции примерно равномерного изменения уровней: равных в среднем величин абсолютного пророста (или абсолютного сокращения) за равные промежутки времени.
- ✓ Параболический тренд используют, если цепные темпы изменений либо уменьшаются, либо некоторое время возрастают, но при достаточно большом периоде рано или поздно темпы роста обязательно начинают уменьшаться (темпы сокращения уровней начинают возрастать).
- ✓ Степенной или показательный тренд используют при значительном ускорении уровней ряда.
- ✓ Полином степени 3 используют, если направление изменения показателя происходит два раза.
- ✓ Логистическая форма тренда используется для описания процессов, при которых изучаемый показатель проходит полный цикл развития,

- начиная от нулевого уровня, сначала медленно, но с ускорением возрастая;
- затем ускорение становится нулевым в середине цикла, т.е. рост происходит по линейному тренду;
- далее, в завершающей части цикла рост замедляется по гиперболе по мере приближения к предельному значению показателя.

Моделирование сезонных колебаний

Простейший подход к моделированию сезонных колебаний – это расчет значений сезонной компоненты методом скользящей средней и построение аддитивной или мультипликативной модели временного ряда.

Общий вид аддитивной модели следующий:

$$Y = T + S + E.$$

Эта модель предполагает, что каждый уровень временного ряда может быть представлен как сумма трендовой (T), сезонной (S) и случайной (E) компонент.

Общий вид мультипликативной модели выглядит так:

$$Y = T \cdot S \cdot E.$$

Эта модель предполагает, что каждый уровень временного ряда может быть представлен как произведение трендовой (T), сезонной (S) и случайной (E) компонент.

Выбор одной из двух моделей осуществляется на основе анализа структуры сезонных колебаний. Если амплитуда колебаний приблизительно постоянна, строят аддитивную модель временного ряда, в которой значения сезонной компоненты предполагаются постоянными для различных циклов. Если амплитуда сезонных колебаний возрастает или уменьшается, строят мультипликативную модель временного ряда, которая ставит уровни ряда в зависимость от значений сезонной компоненты.

Построение аддитивной и мультипликативной моделей сводится к расчету значений T , S и E для каждого уровня ряда.

Процесс построения модели включает в себя следующие шаги.

- 1) Выравнивание исходного ряда методом скользящей средней.
- 2) Расчет значений сезонной компоненты S .
- 3) Устранение сезонной компоненты из исходных уровней ряда и получение выровненных данных ($T + E$) в аддитивной или ($T \cdot E$) в мультипликативной модели.
- 4) Аналитическое выравнивание уровней ($T + E$) или ($T \cdot E$) и расчет значений T с использованием полученного уравнения тренда.
- 5) Расчет полученных по модели значений ($T + E$) или ($T \cdot E$).
- 6) Расчет абсолютных и/или относительных ошибок. Если полученные значения ошибок не содержат автокорреляции, ими можно заменить исходные уровни ряда и в дальнейшем использовать временной ряд ошибок E для анализа взаимосвязи исходного ряда и других временных рядов.

Оценка качества построенной модели

Коэффициент детерминации R^2 показывает *долю дисперсии объясненную регрессией*. Чем больше факторов включается в модель, тем большая часть дисперсии объясняется, поэтому при включении даже несущественной переменной можно заметить увеличение коэффициента детерминации R^2 . Для устранения этого эффекта при анализе переменных используют скорректированный коэффициент детерминации:

$$R_{adj}^2 = 1 - (1 - R^2) \frac{n-1}{n-k-1}, \text{ где } n - \text{ количество наблюдений, } k - \text{ количество}$$

объясняющих переменных.

Значимость построенного уравнения тренда оценивается по критерию Фишера. Фактическое значение F – критерия для уравнения регрессии определяется по формуле:

$$F_{\text{факт.}} = \frac{SSR}{SSE} \cdot \frac{n-k}{k-1} = \frac{R^2}{1-R^2} \cdot (n-k),$$

где k - общее число параметров в уравнении регрессии

Критерий поворотных точек применяется для оценки случайности в остатках.

Критерий поворотных точек

1. Каждый элемент ряда e_i сравнивается с двумя рядом стоящими элементами e_{i-1} и e_{i+1} . Если e_i больше (или меньше) как e_{i-1} , так и e_{i+1} , то она считается поворотной.

2. Подсчитывается сумма поворотных точек P .

Если выполняется условие:

$$P > \left[\frac{2(n-2)}{3} - 1,96 \sqrt{\frac{16n-29}{90}} \right], \text{ то ряд } e_i \text{ является случайным с}$$

вероятностью 95 %. Здесь n – число наблюдений, а квадратные скобки означают, что от результата берется целая часть.

Нормальность распределения в остатках проверяется с помощью RS критерия. Крите-

рий размахов RS. Предпосылку нельзя отклонить, если $a \leq RS = \frac{\max_i \varepsilon_i - \min_i \varepsilon_i}{S_E} \leq b$

Где S_E - среднеквадратическое отклонение остатков. Критические точки критерия RS представлены в приложении 4.

Для анализа коррелированности отклонений используют статистику Дарбина-Уотсона:

$$DW = \frac{\sum (e_i - e_{i-1})^2}{\sum e_i^2}.$$

Критические значения d_1 и d_2 определяются на основе специальных таблиц (приложение 3) для требуемого уровня значимости α , числа наблюдений n и количества объясняющих переменных m . Значение критерия DW может находиться в промежутке от 0 до 4.

Выводы об отсутствии автокорреляции остатков осуществляются по следующей схеме (рис. 5.1).



Рис. 5.1 Проверка гипотезы Дарбина–Уотсона

Не обращаясь к таблицам, можно пользоваться «грубым» правилом и считать, что автокорреляция остатков отсутствует, если $1,5 < DW < 2,5$. Для более надежного вывода целесообразно обращаться к табличным значениям.

Доверительный интервал для прогноза

Границы доверительного интервала, в котором с заданной надежностью γ будут найдены прогнозные значения. Общепринятый в экономике уровень надежности $\gamma = 1 - \alpha = 1 - 0,05 = 0,95$.

$$\hat{y}_t - \Delta \leq y_t \leq \hat{y}_t + \Delta,$$

где y_t - прогноз значения переменной y на момент (период) времени t ;

$\hat{y}_t$ - точечная оценка значения переменной y на момент (период) времени t ;

Δ - предельная ошибка прогноза.

Для того, чтобы получить интервальную оценку, определим величину предельной ошибки прогноза. Она рассчитывается по формуле:

$$\Delta = t_{\alpha/2} \cdot se(\hat{y}_t),$$

где $t_{\alpha/2}$ - табличное значение t - критерия Стьюдента для уровня значимости α и числа степеней свободы ($k = n - 2$);

$se(\hat{y}_t)$ - стандартная ошибка точечного прогноза, которая, в свою очередь, рассчитывается по формуле:

$$se(\hat{y}_t) = \sqrt{\frac{\sum (y_t - \hat{y}_t)^2}{n - 2} \left(1 + \frac{1}{n} + \frac{3(n + 2l - 1)^2}{n^3 - n} \right)},$$

где l - длина периода упреждения (срок прогноза).

5.2 Типовые примеры и образцы их решения

Пример 1. Пусть имеются некоторые условные данные об общем количестве правонарушений на таможне одного из субъектов РФ (таблица 5.1).

Таблица 5.1 Данные для анализа

Год	Квартал	t	Количество возбужденных дел, y_t
2012	I	1	375
	II	2	371
	III	3	869
	IV	4	1015
2013	I	5	357
	II	6	471
	III	7	992
	IV	8	1020
2014	I	9	390
	II	10	355
	III	11	992
	IV	12	905
2015	I	13	461
	II	14	454
	III	15	920
	IV	16	927

Построим поле корреляции (рисунок 5.2):

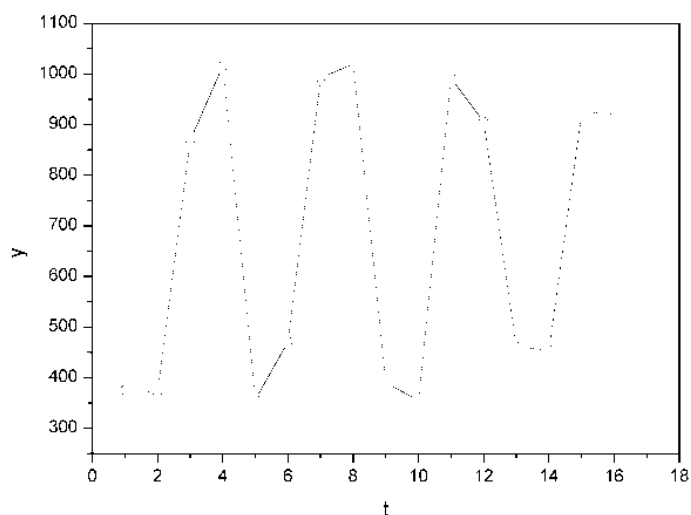


Рис. 5.2 – Корреляционное поле

Исходя из графика видно, что значения y образуют пилообразную фигуру. Рассчитаем несколько последовательных коэффициентов автокорреляции. Для этого составляем первую вспомогательную таблицу 5.2.

Таблица 5.2 Расчет коэффициента автокорреляции 1-го порядка

t	y_t	y_{t-1}	$y_t - \bar{y}_1$	$y_{t-1} - \bar{y}_2$	$(y_t - \bar{y}_1) \times (y_{t-1} - \bar{y}_2)$	$(y_t - \bar{y}_1)^2$	$(y_{t-1} - \bar{y}_2)^2$
1	2	3	4	5	6	7	8
1	375	—	—	—	—	—	—
2	371	375	-328,33	-288,13	94601,72	107800,59	83018,90
3	869	371	169,67	-292,13	-49565,70	28787,91	85339,94
4	1015	869	315,67	205,87	64986,98	99647,55	42382,46
5	357	1015	-342,33	351,87	-120455,66	117189,83	123812,50
6	471	357	-228,33	-306,13	69898,66	52134,59	93715,58
7	992	471	292,67	-192,13	-56230,69	85655,73	36913,94
8	1020	992	320,67	328,87	105458,74	102829,25	108155,48
9	390	1020	-309,33	356,87	-110390,60	95685,05	127356,20
10	355	390	-344,33	-273,13	94046,85	118563,15	74600,00
11	992	355	292,67	-308,13	-90180,41	85655,73	94944,10
12	905	992	205,67	328,87	67638,69	42300,15	108155,48
13	461	905	-238,33	241,87	-57644,88	56801,19	58501,10
14	454	461	-245,33	-202,13	49588,55	60186,81	40856,54
15	920	454	220,67	-209,13	-46148,72	48695,25	43735,36
16	927	920	227,67	256,87	58481,59	51833,63	65982,20
Сумма	10499	9947	9,05	0,05	74085,16	1153766,39	1187469,73
Среднее значение	699,33	663,13	—	—	—	—	—

Следует заметить, что среднее значение получается путем деления не на 16, а на 15, т.к. у нас теперь на одно наблюдение меньше.

Теперь вычисляем коэффициент автокорреляции первого порядка по формуле:

$$r_1 = \frac{74085,16}{\sqrt{1153756,39 \cdot 1187469,73}} = 0,063294 \cdot$$

Аналогично находим коэффициенты автокорреляции более высоких порядков, а все полученные значения заносим в сводную таблицу 5.3.

Таблица 5.3 коэффициенты автокорреляции

Лаг	Коэффициент автокорреляции уровней
1	0,063294
2	-0,961183
3	-0,036290
4	0,964735
5	0,050594
6	-0,976516
7	-0,069444
8	0,964629
9	0,162064
10	-0,972918
11	-0,065323
12	0,985761

Коррелограмма примет вид (рис.5.3):

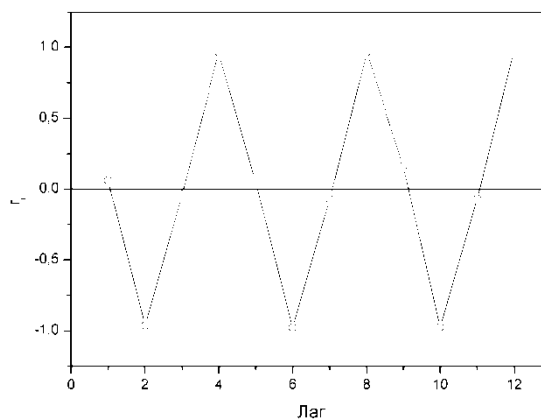


Рисунок 5.3 – Коррелограмма

Анализ коррелограммы и графика исходных уровней временного ряда позволяет сделать вывод о наличии в изучаемом временном ряде сезонных колебаний периодичностью в четыре квартала. Построим аддитивную модель временного ряда. Для этого проведем ряд шагов.

Шаг 1. Проведем выравнивание исходных уровней ряда методом скользящей средней. Для этого:

1.1. Просуммируем уровни ряда последовательно за каждые четыре квартала со сдвигом на один момент времени и определим условные годовые объемы потребления электроэнергии (гр. 3 табл. 5.4).

1.2. Разделив полученные суммы на 4, найдем скользящие средние (гр. 4 табл. 5.4). Полученные таким образом выровненные значения уже не содержат сезонной компоненты.

1.3. Приведем эти значения в соответствие с фактическими моментами времени, для чего найдем средние значения из двух последовательных скользящих средних – центрированные скользящие средние (гр. 5 табл. 5.4).

Шаг 2. Найдем оценки сезонной компоненты как разность между фактическими уровнями ряда и центрированными скользящими средними (гр. 6 табл. 5.4). Используем эти оценки для расчета значений сезонной компоненты S (табл. 5.5). Для этого найдем средние за каждый квартал (по всем годам) оценки сезонной компоненты S_i . В моделях с сезонной компонентой обычно предполагается, что сезонные воздействия за период взаимопогашаются.

Таблица 5.4 Расчет скользящих средних и оценка сезонной компоненты

№ квартала, t	Количество правонарушений, y_t	Итого за четыре квартала	Скользящая средняя за четыре квартала	Центрированная скользящая средняя	Оценка сезонной компоненты
1	2	3	4	5	6
1	375	–	–	–	–
2	371	2630	657,5	–	–
3	869	2612	653	655,25	213,75
4	1015	2712	678	665,5	349,5
5	357	2835	708,75	693,75	-336,75
6	471	2840	710	709,375	-238,375
7	992	2873	718,25	714,125	277,875
8	1020	2757	689,25	703,75	316,25
9	390	2757	689,25	689,25	-299,25
10	355	2642	660,5	674,875	-319,875
11	992	2713	678,25	669,375	322,625
12	905	2812	703	690,625	214,375
13	461	2740	685	694	-233
14	454	2762	690,5	687,75	-233,75
15	920	–	–	–	–
16	927	–	–	–	–

В аддитивной модели это выражается в том, что сумма значений сезонной компоненты по всем кварталам должна быть равна нулю (таблица 5.5).

Таблица 5.5 Оценка скорректированной сезонной компоненты

Показатели	Год	№ квартала, i			
		I	II	III	IV
	2012	–	–	213,75	349,5
	2013	-336,75	-238,375	277,875	316,25
	2014	-299,25	-319,875	322,625	214,375
	2015	-233	-233,75	–	–
Всего за i -й квартал		-869	-792	814,25	880,125
Средняя оценка сезонной компоненты для i -го квартала, $\bar{S}_i$		-289,667	-264	271,417	293,375
Скорректированная сезонная компонента, S_i		-292,448	-266,781	268,636	290,593

Для данной модели имеем:

$$-289,667 - 264 + 271,417 + 293,375 = 11,125.$$

Корректирующий коэффициент: $k = 11,125/4 = 2,781$.

Рассчитываем скорректированные значения сезонной компоненты ($S_i = \bar{S}_i - k$) и заносим полученные данные в таблицу 5.6.

Проверим равенство нулю суммы значений сезонной компоненты:

$$-292,448 - 266,781 + 268,636 + 290,593 = 0.$$

Шаг 3. Исключим влияние сезонной компоненты, вычитая ее значение из каждого уровня исходного временного ряда. Получим величины $T + E = Y - S$ (гр. 4 табл. 5.6). Эти значения рассчитываются за каждый момент времени и содержат только тенденцию и случайную компоненту (табл.5.6).

Таблица 5.6 Расчет параметров аддитивной модели

t	y_t	S_i	$y_t - S_i$	T	$T + S$	$E = y_t - (T + S)$	E^2
1	2	3	4	5	6	7	8
1	375	-292,448	667,448	672,700	380,252	-5,252	27,584
2	371	-266,781	637,781	673,624	406,843	-35,843	1284,721
3	869	268,636	600,364	674,547	943,183	-74,183	5503,117
4	1015	290,593	724,407	675,470	966,063	48,937	2394,830
5	357	-292,448	649,448	676,394	383,946	-26,946	726,087
6	471	-266,781	737,781	677,317	410,536	60,464	3655,895
7	992	268,636	723,364	678,240	946,876	45,124	2036,175
8	1020	290,593	729,407	679,163	969,756	50,244	2524,460
9	390	-292,448	682,448	680,087	387,639	2,361	5,574
10	355	-266,781	621,781	681,010	414,229	-59,229	3508,074
11	992	268,636	723,364	681,933	950,569	41,431	1716,528
12	905	290,593	614,407	682,857	973,450	-68,450	4685,403
13	461	-292,448	753,448	683,780	391,332	69,668	4853,630
14	454	-266,781	720,781	684,703	417,922	36,078	1301,622
15	920	268,636	651,364	685,627	954,263	-34,263	1173,953
16	927	290,593	636,407	686,550	977,143	-50,143	2514,320

Шаг 4. Определим компоненту T данной модели. Для этого проведем аналитическое выравнивание ряда ($T + E$) с помощью линейного тренда. Результаты аналитического выравнивания следующие:

$$T = 671,777 + 0,9233 \cdot t.$$

Подставляя в это уравнение значения $t = 1, 2, \dots, 16$, найдем уровни T для каждого момента времени (гр. 5 табл. 5.6).

Шаг 5. Найдем значения уровней ряда, полученные по аддитивной модели. Для этого прибавим к уровням T значения сезонной компоненты для соответствующих кварталов (гр. 6 табл. 5.6).

На одном графике (рис. 5.4) отложим фактические значения уровней временного ряда и теоретические, полученные по аддитивной модели.

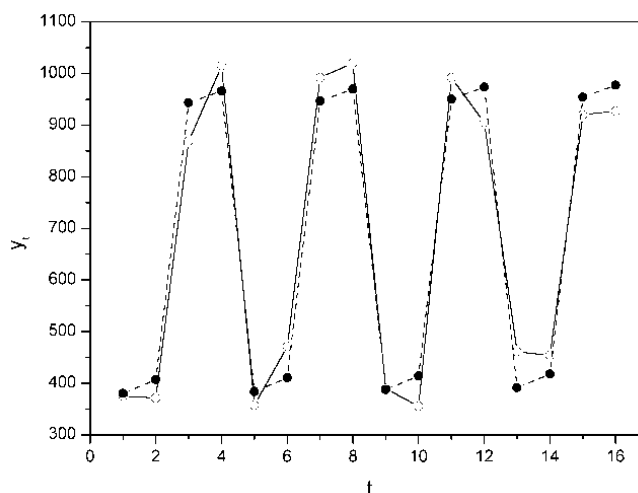


Рис. 5.4 – Исходные данные и аддитивная модель.

Для оценки качества построенной модели применим сумму квадратов полученных абсолютных ошибок.

$$R^2 = 1 - \frac{E^2}{(y_t - \bar{y})^2} = 1 - \frac{37911,973}{1252743,75} = 0,970.$$

Следовательно, можно сказать, что аддитивная модель объясняет 97% общей вариации уровней временного ряда количества правонарушений по кварталам за 4 года.

Шаг 6. Прогнозирование по аддитивной модели. Предположим, что по нашему примеру необходимо дать прогноз об общем объеме правонарушений на I и II кварталы 2016 года. Прогнозное значение F_t уровня временного ряда в аддитивной модели есть сумма трендовой и сезонной компонент. Для определения трендовой компоненты воспользуемся уравнением тренда

$$T = 671,777 + 0,9233 \cdot t.$$

Получим

$$T_{17} = 671,777 + 0,9233 \cdot 17 = 687,473;$$

$$T_{18} = 671,777 + 0,9233 \cdot 18 = 688,396.$$

Значения сезонных компонент за соответствующие кварталы равны: $S_1 = -292,448$ и $S_2 = -266,781$. Таким образом,

$$F_{17} = T_{17} + S_1 = 687,473 - 292,448 \approx 395;$$

$$F_{18} = T_{18} + S_2 = 688,396 - 266,781 \approx 422.$$

Т.е. в первые два квартала 2018 г. следовало ожидать порядка 395 и 422 правонарушений соответственно.

5.3 Задания на практику

1. Администрация банка изучает динамику депозитов физических лиц за ряд лет (млн. долл. в сопоставимых ценах). Исходные данные представлены в табл. 5.7:

Таблица 5.7

								Сумма
Время, лет	1	2	3	4	5	6	7	28
Депозиты физических лиц, x	2	6	7	3	10	12	13	53

Известно также следующее: $\sum x^2 = 511$. Задание:

- 1) Постройте уравнение линейного тренда и дайте интерпретацию его параметров.
- 2) Определите коэффициент детерминации для линейного тренда.
- 3) Администрация банка предполагает, что среднегодовой абсолютный прирост депозитов физических лиц составляет не менее 2,5 млн. долл. Подтверждается ли это предположение результатами, которые вы получили?

2. Имеются следующие данные об уровне безработицы месяцев y_t (%) за 8 месяцев (таблица 5.8):

Таблица 5.8 Данные для анализа

Месяц	1	2	3	4	5	6	7	8
y_t	8,8	8,6	8,4	8,1	7,9	7,6	7,4	7,0

- 1) Определите коэффициенты автокорреляции уровней этого ряда первого и второго порядка.
- 2) Обоснуйте выбор уравнения тренда и определите его параметры.

3. Имеются данные динамики производства электроэнергии по региону (тыс. квт. ч.) за 12 лет (таблица 5.9).

- 1) Провести сглаживание временного ряда с помощью трехчленной скользящей средней. Сделать выводы о типе модели.
- 2) Построить модель тенденции временного ряда (с помощью подбора в Excel). Выбрать лучшую по скорректированному коэффициенту детерминации, рассчитать значения по модели.
- 3) Оценить качество модели (по коэффициенту детерминации, критерию поворотных точек, ошибке аппроксимации, критерию Дарбина-Уотсона, критерию RS). Сделать выводы о качестве и возможности прогнозирования по модели.
- 4) Построить прогноз на 2020 г. по лучшей модели ($t^* = 14$).
- 5) Рассчитать адаптивную модель Брауна для параметров $\alpha = 0,5; 0,8$ для каждой модели найти стандартную ошибку и выбрать лучшую. Построить прогноз по модели на 2020 г.

Таблица 5.9 Данные для анализа

год	Производство электроэнергии, млн. кВт. ч.
2007	165
2008	170
2009	182
2010	198
2011	190
2012	225
2013	250
2014	232
2015	264
2016	297
2017	356
2018	381

6. Дана выборка курса акции за 12 месяцев, тыс. руб. (таблица 5.10).

- 1) Найти коэффициенты автокорреляции со смещением на 1,2,3,4 и их значимость с вероятностью 0,95. Сделать выводы.
- 2) Построить коррелограмму.
- 3) Провести сглаживание ряда трехчленными скользящими средними.
- 4) Определить скорректированную сезонную компоненту. И исключить ее из ряда.
- 5) Построить аддитивную (или мультипликативную) модель временного ряда.
- 6) Оценить качество модели по коэффициенту детерминации, ошибке аппроксимации и RS. Сделать выводы.
- 7) Построить прогноз по модели на февраль следующего года.

Таблица 5.10

№ месяца	Курс Акции, долл., Y_t
1	13,1
2	11,9
3	11,8
4	17,3
5	15,9
6	16,1
7	20,5
8	19,2
9	19,9
10	23,9
11	22,8
12	23,8

5.4.1 Семестровая 5.1 «Динамика и адаптивные методы»

Дана выборка динамики производства продукции (тыс. ед.) за 12 лет по предприятию (таблица 5.11).

- 1) Провести сглаживание временного ряда с помощью трехчленной скользящей средней. Сделать выводы о типе модели.
- 2) Построить модель тенденции временного ряда (с помощью подбора в Excel). Выбрать лучшую по скорректированному коэффициенту детерминации, рассчитать значения по модели.
- 3) Оценить качество модели (по коэффициенту детерминации, критерию поворотных точек, ошибке аппроксимации, критерию Дарбина-Уотсона, критерию RS). Сделать выводы о качестве и возможности прогнозирования по модели.
- 4) Построить прогноз на 2020 г. по лучшей модели ($t^* = 14$).
- 5) Рассчитать адаптивную модель Брауна для параметров $\alpha = 0,5; 0,2; 0,8$ для каждой модели найти стандартную ошибку и выбрать лучшую. Построить прогноз по модели на 2020 г.

Таблица 5.11 Данные для анализа по вариантам

год	Вариант									
	1	2	3	4	5	6	7	8	9	10
2007	20	25	200	115	105	38	240	102	15	200
2008	25	29	183	105	130	41	232	97	18	143
2009	27	27	153	95	150	42	214	89	15	140
2010	30	36	140	92	250	49	200	77	30	170
2011	29	48	107	84	242	64	230	81	35	250
2012	30	73	87	72	254	53	280	87	26	280
2013	35	68	68	57	275	47	300	94	34	340
2014	33	77	83	49	272	52	310	92	37	354
2015	40	78	85	46	262	51	330	100	44	370
2016	43	84	79	40	259	70	290	104	59	320
2017	42	87	80	35	330	92	237	87	64	250
2018	48	80	73	32	362	100	252	80	68	210

5.4.2 Семестровая работа 5.2 «Моделирование сезонных колебаний»

Дана выборка объемов продаж разных товаров по одному из магазинов за 12 месяцев (таблица 5.12).

- 1) Найти коэффициенты автокорреляции со смещением на 1,2,3,4 и их значимость с вероятностью 0,95. Сделать выводы.
- 2) Построить коррелограмму.
- 3) Провести сглаживание ряда трехчленными скользящими средними.
- 4) Определить скорректированную сезонную компоненту. И исключить ее из ряда.
- 5) Построить аддитивную (или мультипликативную) модель временного ряда.
- 6) Оценить качество модели по коэффициенту детерминации, ошибке аппроксимации и RS. Сделать выводы.
- 7) Построить прогноз по модели на февраль следующего года.

Таблица 5.12 Данные для анализа по вариантам

месяц	<i>№ варианта</i>									
	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>	<i>7</i>	<i>8</i>	<i>9</i>	<i>10</i>
1	52	79	74,4	107	84,1	112	32,8	46,7	13,3	35
2	52,1	78,2	73,2	108	82,6	111	32	46,1	12,5	33
3	57,3	78,6	74,3	111	83,8	112	32,8	45,7	12,7	34
4	56,5	82	79,9	110	87,5	117	34,7	49,7	17,2	38
5	56,1	81	78,7	109	87,3	116	34,1	47,4	15,9	36
6	56,2	82,3	79,7	112	88,1	116	34,2	47,8	16,1	38
7	61,3	87,1	84,1	117	93	122	37,5	52	20,5	42
8	60,9	86,3	84,3	116	92,3	121	35,8	50,1	19,2	40
9	60,5	85,5	85,4	117	93,6	121	35,7	49,8	19,9	41
10	65,4	92,4	89,3	122	99,4	126	40,1	54,6	23,9	45
11	65,6	90,6	89,6	121	97,2	125	38,8	51,9	22,8	44
12	65	90,7	91	122	98	126	39	52,3	23,5	45

Структурная модель (3.3) в полном виде содержит $m \cdot (m + n - 1)$ параметров, а приведенная форма модели в полном виде содержит $m \cdot n$ параметров. Т.е. в полном виде структурная модель содержит большее число параметров, чем приведенная форма модели. Соответственно $m \cdot (m + n - 1)$ параметров структурной модели не могут быть однозначно определены из $m \cdot n$ параметров приведенной формы модели.

Чтобы получить единственно возможное решение для структурной модели, необходимо предположить, что некоторые из структурных коэффициентов модели ввиду слабой взаимосвязи признаков с эндогенной переменной из левой части системы равны нулю. Тем самым уменьшится число структурных коэффициентов модели. Уменьшение числа структурных коэффициентов модели возможно и другим путем: например, путем приравнивания некоторых коэффициентов друг к другу, т.е. путем предположений, что их воздействие на формируемую эндогенную переменную одинаково. На структурные коэффициенты могут накладываться, например, ограничения вида $b_{ik} + a_{ij} = 0$.

С позиции идентифицируемости структурные модели можно подразделить на три вида:

- 1) идентифицируемые;
- 2) неидентифицируемые;
- 3) сверхидентифицируемые.

Модель *идентифицируема*, если все структурные ее коэффициенты определяются однозначно, единственным образом по коэффициентам приведенной формы модели, т. е. если число параметров структурной модели равно числу параметров приведенной формы модели. В этом случае структурные коэффициенты модели оцениваются через параметры приведенной формы модели и модель идентифицируема.

Модель *неидентифицируема*, если число приведенных коэффициентов меньше числа структурных коэффициентов, и в результате структурные коэффициенты не могут быть оценены через коэффициенты приведенной формы модели.

Модель *сверхидентифицируема*, если число приведенных коэффициентов больше числа структурных коэффициентов. В этом случае на основе коэффициентов приведенной формы можно получить два или более значений одного структурного коэффициента. В этой модели число структурных коэффициентов меньше числа коэффициентов приведенной формы. Сверхидентифицируемая модель в отличие от неидентифицируемой модели практически решается, но требует для этого специальных методов исчисления параметров.

Выполнение условия идентифицируемости модели проверяется для каждого уравнения системы. Чтобы уравнение было идентифицируемо, необходимо, чтобы число предопределенных переменных, отсутствующих в данном уравнении, но присутствующих в системе, было равно числу эндогенных переменных в данном уравнении без одного.

Если обозначить число эндогенных переменных в i -м уравнении системы через H , а число экзогенных (предопределенных) переменных, которые содержатся в системе, но не входят в данное уравнение, — через D , то условие идентифицируемости модели может быть записано в виде следующего счетного правила:

$D + 1 = H$	уравнение идентифицируемо
$D + 1 < H$	уравнение неидентифицируемо
$D + 1 > H$	уравнение сверхидентифицируемо

Для оценки параметров структурной модели система должна быть идентифицируема или сверхидентифицируема.

Рассмотренное счетное правило отражает необходимое, но недостаточное условие идентификации. Более точно условия идентификации определяются, если накладывать ограничения на коэффициенты матриц параметров структурной модели. Уравнение идентифицируемо, если по отсутствующим в нем переменным (эндогенным и экзогенным) можно из коэффициентов при них в других уравнениях системы получить матрицу, определитель которой не равен нулю, а ранг матрицы не меньше, чем число эндогенных переменных в системе без одного.

Целесообразность проверки условия идентификации модели через определитель матрицы коэффициентов, отсутствующих в данном уравнении, но присутствующих в других, объясняется тем, что возможна ситуация, когда для каждого уравнения системы выполнено счетное правило, а определитель матрицы названных коэффициентов равен нулю. В этом случае соблюдается лишь необходимое, но недостаточное условие идентификации.

Коэффициенты структурной модели могут быть оценены разными способами в зависимости от вида системы одновременных уравнений. Наибольшее распространение в литературе получили следующие методы оценивания коэффициентов структурной модели:

- 1) косвенный метод наименьших квадратов;
- 2) двухшаговый метод наименьших квадратов;
- 3) трехшаговый метод наименьших квадратов;

Рассмотрим вкратце сущность каждого из этих методов.

Косвенный метод наименьших квадратов (КМНК) применяется в случае точно идентифицируемой структурной модели. Процедура применения КМНК предполагает выполнение следующих этапов работы.

1. Структурная модель преобразовывается в приведенную форму модели.
2. Для каждого уравнения приведенной формы модели обычным МНК оцениваются

приведенные коэффициенты δ_{ij} .

3. Коэффициенты приведенной формы модели трансформируются в параметры структурной модели.

Если система сверхидентифицируема, то КМНК не используется, ибо он не дает однозначных оценок для параметров структурной модели. В этом случае могут использоваться разные методы оценивания, среди которых наиболее распространенным и простым является двухшаговый метод наименьших квадратов (ДМНК).

Основная идея ДМНК – на основе приведенной формы модели получить для сверхидентифицируемого уравнения теоретические значения эндогенных переменных, содержащихся в правой части уравнения.

Далее, подставив их вместо фактических значений, можно применить обычный МНК к структурной форме сверхидентифицируемого уравнения. Метод получил название двухшагового МНК, ибо дважды используется МНК: на первом шаге при определении приведенной формы модели и нахождении на ее основе оценок теоретических значений эндогенной переменной $y_i = \delta_{i1}x_1 + \delta_{i2}x_2 + \dots + \delta_{in}x_n$ и на втором шаге применительно к структурному сверхидентифицируемому уравнению при определении структурных коэффициентов модели по данным теоретических (расчетных) значений эндогенных переменных.

Сверхидентифицируемая структурная модель может быть двух типов:

- 1) все уравнения системы сверхидентифицируемы;

2) система содержит наряду со сверхидентифицируемыми точно идентифицируемые уравнения.

Если все уравнения системы сверхидентифицируемые, то для оценки структурных коэффициентов каждого уравнения используется ДМНК. Если в системе есть точно идентифицируемые уравнения, то структурные коэффициенты по ним находятся из системы приведенных уравнений.

6.2. Типовые примеры и образцы их решения

Пример 1. Задана структурная система эконометрических уравнений:

$$\begin{cases} y_{t1} = b_{10} + b_{12}y_{t2} + a_{11}x_{t1} + a_{12}x_{t2} + \varepsilon_{t1} \\ y_{t2} = b_{20} + b_{21}y_{t1} + b_{23}y_{t3} + a_{23}x_{t3} + \varepsilon_{t2} \\ y_{t3} = b_{30} + b_{32}y_{t1} + \varepsilon_{t3} \end{cases}$$

В табл.6.1 приведены данные наблюдений эндогенных y_{it} и экзогенных x_{it} переменных модели:

Таблица 6.1						
t	y_{t1}	y_{t2}	y_{t3}	x_{t1}	x_{t2}	x_{t3}
1	19	23	13	5	5	4
2	15	31	15	7	8	3
3	22	13	17	4	8	2
4	6	32	9	6	2	5
5	8	66	16	12	2	9

Необходимо, опираясь на данные наблюдений, оценить параметры структурной системы.

Решение: Определим, являются ли уравнения системы идентифицируемыми. Для этого применим необходимое условие идентификации или счетное правило:

$D+1=N$ – уравнение идентифицируемо;

$D+1<N$ – уравнение неидентифицируемо;

$D+1>N$ – уравнение сверхидентифицируемо.

Здесь - N – число эндогенных переменных в уравнении; D – число отсутствующих экзогенных и лаговых переменных в уравнении. В приведенной системе лаговые или запаздывающие переменные отсутствуют.

В первом уравнении эндогенных переменных две: y_{t1} , y_{t2} , следовательно, $N=2$. Всего у нас три экзогенных переменных, из них в первом уравнении отсутствует одна x_{t3} . Поэтому $D=1$. Для первого уравнения выполняется $D+1=N=2$. Первое уравнение идентифицируемо. Аналогично определяем, что для второго уравнения выполняется $D+1=N=3$. Для третьего уравнения находим, что $D+1=4>N=2$. Третье уравнение сверхидентифицируемо.

Прямая оценка параметров структурной системы уравнений недопустима. Из-за корреляции регрессоров с остатками получаются смещенные оценки. Для оценки параметров используют косвенный МНК (в случае идентификации) или двухшаговый МНК (в случае идентификации и сверхидентификации). Оба метода требуют оценки параметров приведенной формы модели:

$$\begin{cases} \hat{y}_{t1} = \hat{c}_{10} + \hat{c}_{11}x_{t1} + \hat{c}_{12}x_{t2} + \hat{c}_{13}x_{t3} \\ \hat{y}_{t2} = \hat{c}_{20} + \hat{c}_{21}x_{t1} + \hat{c}_{22}x_{t2} + \hat{c}_{23}x_{t3} \\ \hat{y}_{t3} = \hat{c}_{30} + \hat{c}_{31}x_{t1} + \hat{c}_{32}x_{t2} + \hat{c}_{33}x_{t3}. \end{cases}$$

Опираясь на данные наблюдений, выполняем регрессию y_{t1} на x_{t1} , x_{t2} , x_{t3} . Получаем уравнение прогноза

$$\hat{y}_{t1} = -14,982 - 5,421 x_{t1} + 5,228 x_{t2} + 8,632 x_{t3} \quad R^2 = 0,998$$

(2,726) (2,726) (0,367) (0,809)

В скобках приводятся стандартные ошибки оценок параметров.

Затем выполняем регрессию y_{t2} на x_{t1} , x_{t2} , x_{t3}

$$\hat{y}_{t2} = 1,129 + 5,912 x_{t1} - 1,327 x_{t2} - 0,368 x_{t3} \quad R^2 = 0,9997$$

(2,726) (0,469) (0,367) (0,809)

и y_{t3} на x_{t1} , x_{t2} , x_{t3}

$$\hat{y}_{t3} = -7,298 - 1,342 x_{t1} + 2,623 x_{t2} + 3,763 x_{t3} \quad R^2 = 0,920$$

(7,496) (1,291) (1,008) (2,223)

Приведенная система эконометрических уравнений имеет вид

$$\begin{cases} \hat{y}_{t1} = -14,982 - 5,421x_{t1} + 5,228x_{t2} + 8,632x_{t3} \\ \hat{y}_{t2} = 1,129 + 5,912x_{t1} - 1,327x_{t2} - 0,368x_{t3} \\ \hat{y}_{t3} = -7,298 - 1,342x_{t1} + 2,623x_{t2} + 3,763x_{t3}. \end{cases}$$

Параметры первого уравнения структурной системы можно идентифицировать методом, который называется косвенный МНК. Для этого второе уравнение приведенной системы решается относительно x_{t3} :

$$x_{t3} = (\hat{y}_{t2} - 1,129 - 5,912x_{t1} + 1,327x_{t2}) / (-0,368) = 3,068 - 2,717\hat{y}_{t2} + 16,065x_{t1} - 3,606x_{t2}$$

Затем найденное значение подставляется в первое уравнение:

$$\begin{aligned} \hat{y}_{t1} &= -14,982 - 5,421x_{t1} + 5,228x_{t2} + 8,632 \cdot (3,068 - 2,717\hat{y}_{t2} + 16,065x_{t1} - 3,606x_{t2}) \\ &= 11,501 - 23,453\hat{y}_{t2} + 133,252x_{t1} - 25,899x_{t2}. \end{aligned}$$

Для идентификации второго уравнения структурной системы решаем первое и третье уравнения приведенной системы относительно x_{t1} и x_{t2} и подставляем найденные значения во второе уравнение:

$$\begin{cases} -5,421x_{t1} + 5,228x_{t2} = \hat{y}_{t1} + 14,982 - 8,632x_{t3} \\ -1,342x_{t1} + 2,623x_{t2} = \hat{y}_{t3} + 7,298 - 3,763x_{t3} \end{cases}$$

$$x_{t1} = \frac{\Delta_1}{\Delta} = \frac{\begin{vmatrix} \hat{y}_{t1} + 14,982 - 8,632x_{t3} & 5,228 \\ \hat{y}_{t3} + 7,298 - 3,763x_{t3} & 2,623 \end{vmatrix}}{\begin{vmatrix} -5,421 & 5,228 \\ -1,342 & 2,623 \end{vmatrix}} = -0,159 - 0,364\hat{y}_{t1} + 0,726\hat{y}_{t3} + 0,412x_{t3}$$

$$x_{t2} = \frac{\Delta_2}{\Delta} = \frac{\begin{vmatrix} -5,421 & \hat{y}_{t1} + 14,982 - 8,632x_{t3} \\ -1,342 & \hat{y}_{t3} + 7,298 - 3,763x_{t3} \end{vmatrix}}{\begin{vmatrix} -5,421 & 5,228 \\ -1,342 & 2,623 \end{vmatrix}} = 2,701 - 0,186\hat{y}_{t1} + 0,753\hat{y}_{t3} - 1,224x_{t3}.$$

$$\begin{aligned} \hat{y}_{t2} &= 1,129 + 5,912 \cdot (-0,159 - 0,364\hat{y}_{t1} + 0,726\hat{y}_{t3} + 0,412x_{t3}) - \\ &\quad - 1,327 \cdot (2,701 - 0,186\hat{y}_{t1} + 0,753\hat{y}_{t3} - 1,224x_{t3}) - 0,368x_{t3} = \\ &= -3,395 - 1,905\hat{y}_{t1} + 3,293\hat{y}_{t3} + 3,692x_{t3}. \end{aligned}$$

Третье уравнение структурной системы сверхидентифицируемо. Для идентификации его параметров необходимо применить двухшаговый МНК. На первом шаге определяем значения инструментальной переменной

$$\hat{y}_{t1} = -14,982 - 5,421x_{t1} + 5,228x_{t2} + 8,632x_{t3}.$$

Значения $\hat{y}_{t1}$ рассчитываются в табл.6.2

Таблица 6.2					
t	y _{t3}	x _{t1}	x _{t2}	x _{t3}	$\hat{y}_{t1}$
1	13	5	5	4	18,579
2	15	7	8	3	14,789
3	17	4	8	2	22,421
4	9	6	2	5	6,105
5	16	12	2	9	8,105

На втором шаге осуществляем регрессию y_{t3} на $\hat{y}_{t1}$. Получаем уравнение

$$\hat{y}_{t3} = 10,370 + 0,259 \hat{y}_{t1} \quad R^2 = 0,319$$

(3,348) (0,219)

Идентификация параметров произведена. Структурная система эконометрических уравнений имеет вид:

$$\begin{cases} \hat{y}_{t1} = 11,501 - 23,453\hat{y}_{t2} + 133,252x_{t1} - 25,899x_{t2} \\ \hat{y}_{t2} = -3,395 - 1,905\hat{y}_{t1} + 3,293\hat{y}_{t3} + 3,692x_{t3} \\ \hat{y}_{t3} = 10,370 + 0,259\hat{y}_{t1} + \varepsilon_{t3}. \end{cases}$$

Пример 2. Имеется следующая структурная модель:

$$\begin{cases} y_1 = b_{12}y_2 + a_{11}x_1 + a_{12}x_2, \\ y_2 = b_{21}y_1 + b_{23}y_3 + a_{22}x_2, \\ y_3 = b_{32}y_2 + a_{31}x_1 + a_{33}x_3. \end{cases}$$

Соответствующая ей приведенная форма модели имеет вид:

$$\begin{cases} y_1 = 3x_1 - 4x_2 + 2x_3, \\ y_2 = 2x_1 + 4x_2 + 5x_3, \\ y_3 = -5x_1 + 6x_2 + 5x_3. \end{cases}$$

Определить, если это возможно, неизвестные параметры структурной модели.

Решение. Сначала определим идентифицируемость структурной модели. Ограничимся для простоты применением счетного правила. Приведем кратко информацию об этом правиле.

Обозначим H – число эндогенных переменных в i -ом уравнении системы, D – число экзогенных переменных, которые содержатся в системе, но не входят в данное уравнение. Тогда условие идентифицируемости уравнения может быть записано в виде следующего счетного правила:

$D+1 = H$ – уравнение идентифицируемо;

$D+1 < H$ – уравнение неидентифицируемо;

$D+1 > H$ – уравнение сверхидентифицируемо.

Первое и третье уравнения структурной модели имеют $H = 2$, $D = 1$. В первом уравнении две эндогенные переменные – y_1 , y_2 , в третьем тоже две – y_2 , y_3 ; в обоих уравнениях не хватает по одной экзогенной переменной: в первом отсутствует x_3 , в третьем – x_2 . В этих уравнениях выполняется равенство $D + 1 = H$, и они идентифицируемы. Во втором уравнении присут-

ствуют все три эндогенные переменные ($H=3$), а отсутствуют две экзогенные – x_1 и x_3 ($D=2$). Здесь также выполняется равенство $D + 1 = H$, и второе уравнение также идентифицируемо. Поскольку все три уравнения структурной модели идентифицируемы, система также идентифицируема.

Для идентифицируемых систем методом оценки структурных параметров является косвенный МНК. Он заключается в том, что уравнения приведенной формы модели (ПФМ), полученные обычным МНК как уравнения множественной регрессии, с помощью алгебраических преобразований превращаются в уравнения структурной формы модели (СФМ). Здесь, как видим, МНК применяется только один раз – для оценки коэффициентов приведенной формы.

Начнем с построения первого уравнения СФМ. Из всех уравнений ПФМ к нему ближе всех по структуре первое уравнение: в обоих уравнениях слева стоит y_1 , а справа стоят x_1 и x_2 . Однако они отличаются тем, что в первом уравнении ПФМ стоит x_3 , а в первом уравнении СФМ стоит y_2 . Поэтому, чтобы получить первое уравнение СФМ из первого уравнения ПФМ, надо в последнем заменить x_3 на выражение, в котором появилась бы y_2 . Эту замену делаем с помощью второго уравнения ПФМ:

$$x_3 = \frac{1}{5}(y_2 - 2x_1 - 4x_2).$$

Подставим в первое уравнение ПФМ, получаем после элементарных преобразований:

$$y_1 = 3x_1 - 4x_2 + 2\left[\frac{1}{5}(y_2 - 2x_1 - 4x_2)\right],$$

или

$$y_1 = 0,4y_2 + 2,2x_1 - 5,64x_2.$$

Это и есть первое уравнение СФМ.

Для получения третьего уравнения СФМ действуем аналогично: в третьем уравнении ПФМ заменяем x_2 так, чтобы в результате замены появилась y_2 . такую замену снова делаем через второе уравнение ПФМ:

$$x_2 = \frac{1}{4}(y_2 - 2x_1 - 5x_3).$$

Подставим в третье уравнение ПФМ, получаем:

$$y_3 = -5x_1 + 6\left[\frac{1}{4}(y_2 - 2x_1 - 5x_3)\right] + 5x_3,$$

или

$$y_3 = 1,5y_2 - 8x_1 - 2,5x_3.$$

Это и есть третье уравнение СФМ.

Для получения второго уравнения СФМ требуются более сложные преобразования. Это связано с тем, что из второго уравнения ПФМ, как наиболее похожего на второе уравнение СФМ, надо исключить сразу две переменные – x_1 и x_3 , чтобы при этом появились y_1 и y_3 . Последовательное исключение здесь не годится, их надо исключать одновременно. Для этого запишем первое и третье уравнения ПФМ как систему относительно исключаемых переменных:

$$\begin{cases} 3x_1 + 2x_3 = y_1 + 4x_2; \\ -5x_1 + 5x_3 = y_3 - 6x_2. \end{cases}$$

Решаем эту систему любым способом, например, методом определителей:

$$\Delta = \begin{vmatrix} 3 & 2 \\ -5 & 5 \end{vmatrix} = 25;$$

$$\Delta_1 = \begin{vmatrix} y_1 + 4x_2 & 2 \\ y_3 - 6x_2 & 5 \end{vmatrix} = 5(y_1 + 4x_2) - 2(y_3 - 6x_2) = 5y_1 - 2y_3 + 32x_2;$$

$$\Delta_3 = \begin{vmatrix} 3 & y_1 + 4x_2 \\ -5 & y_3 - 6x_2 \end{vmatrix} = 3(y_3 - 6x_2) - (-5)(y_1 + 4x_2) = 5y_1 + 3y_3 + 2x_2;$$

$$x_1 = \frac{\Delta_1}{\Delta} = 0,2y_1 - 0,08y_3 + 1,28x_2;$$

$$x_3 = \frac{\Delta_3}{\Delta} = 0,2y_1 + 0,12y_3 + 0,08x_2.$$

Подставим полученные решения во второе уравнение ПФМ, получаем второе уравнение СФМ:

$$y_2 = 2(0,2y_1 - 0,08y_3 + 1,28x_2) + 4x_2 + 5(0,2y_1 + 0,12y_3 + 0,08x_2),$$

или

$$y_2 = 1,4y_1 + 0,44y_3 + 6,96x_2.$$

Теперь можем полностью записать структурную модель:

$$\begin{cases} y_1 = 0,4y_2 + 2,2x_1 - 5,6x_2, \\ y_2 = 1,4y_1 + 0,44y_3 + 6,96x_2, \\ y_3 = 1,5y_2 - 8x_1 - 2,5x_3. \end{cases}$$

6.3 Задания на практику

1. Применив необходимое и достаточное условие идентификации, определите, идентифицировано ли каждое из уравнений модели и запишите приведенную форму модели.

Модель денежного рынка:

$$R_t = a_1 + b_{11}M_t + b_{12}Y_t + e_{1t},$$

$$Y_t = a_2 + b_{21}R_t + b_{22}I_t + e_{2t},$$

$$I_t = a_3 + b_{33}R_t + e_{3t},$$

где R – процентные ставки;

Y – ВВП;

M – денежная масса;

I – внутренние инвестиции.

2. Имеется модель, построенная по шести наблюдениям:

$$y_1 = a_1 + b_{12}y_2,$$

$$y_2 = a_2 + b_{21}y_1 + c_{21}x_1,$$

$$y_3 = y_2 + x_2.$$

Ей соответствует следующая приведенная форма:

$$y_1 = -1,25 + 22x_1 + 0,67x_2,$$

$$y_2 = 2 - 4x_1 + 10x_2,$$

$$y_3 = -30 + 12x_1 + 8x_2.$$

Известны также следующие исходные данные:

n	1	2	3	4	5	6
y_1	3	2	4	1	5	3
x_1	2	3	5	6	10	8
x_2	4	7	3	6	5	5

Определите структурные параметры первого и второго уравнений, если это возможно.

3. Имеется следующая модель:

$$y_1 = b_{11}x_1 + b_{12}x_2 + c_{12}y_2,$$

$$y_2 = b_{22}x_2 + c_{23}y_3 + c_{21}y_1,$$

$$y_3 = b_{31}x_1 + b_{33}x_3 + c_{32}y_2.$$

Приведенная форма этой модели имеет вид

$$y_1 = 3x_1 - 6x_2 + 2x_3,$$

$$y_2 = 2x_1 + 4x_2 + 10x_3,$$

$$y_3 = -5x_1 + 6x_2 + 5x_3.$$

Определите структурные коэффициенты на основе приведенной формы модели проверьте структурную форму модели на идентификацию.

6.4 Семестровая работа 6 «Системы одновременных уравнений»

Задание 6.1

1. Проверить систему на идентифицируемость по необходимому и достаточному условию. Сделать выводы.
2. Построить приведенную систему одновременных уравнений и рассчитать коэффициенты приведенной системы с помощью КМНК (на основе данных).
3. Выразить коэффициенты структурной системы из приведенной

$$\begin{cases} y_1 = f(y_2, x_1) \\ y_2 = f(y_1, x_2) \end{cases} \Rightarrow \begin{cases} y_1 = b_{12}y_2 + a_{12}x_1 \\ y_2 = b_{21}y_1 + a_{21}x_2 \end{cases}$$

по исходным данным. Исходные данные к вариантам 1 – 2 (Таблица 6.3).

Таблица 6.3 Данные по вариантам

№	y1		y2		x1		x2	
	1	2	1	2	1	2	1	2
1	10	15	3	5	6	20	3	4
2	20	20	4	6	8	16	3	5
3	35	15	5	12	5	22	5	3
4	40	25	4	15	10	24	7	7
5	44	35	7	14	12	25	6	9
6	50	40	8	18	14	30	8	10
7	60	50	10	20	15	35	9	14
8	65	55	15	21	18	40	11	12
9	70	65	13	22	21	44	13	15
10	80	70	15	25	24	40	14	18

Задание 6.2 По данным таблицы 6.4:

- 1) Определите к какому типу относится каждое из уравнений структурной системы эконометрических уравнений (идентифицируемо, неидентифицируемо или сверхидентифицируемо) с помощью необходимого и достаточного условия.

2) Опираясь на данные наблюдений и построенную на их основе приведенную систему эконометрических уравнений, выразите параметры структурной модели через приведенную. параметров структурной системы.

Таблица 6.4 Данные по вариантам

Вариант 1. Структурная система: $y_{t1} = b_{10} + b_{12}y_{t2} + \varepsilon_{t1}$ $y_{t2} = b_{20} + b_{21}y_{t1} + b_{23}y_{t3} + a_{23}x_{t3} + \varepsilon_{t2}$ $y_{t3} = b_{30} + b_{31}y_{t1} + a_{31}x_{t1} + a_{32}x_{t2} + \varepsilon_{t3}.$ Приведенная система: $\hat{y}_{t1} = 2 - x_{t1} + 3x_{t2} - x_{t3}$ $\hat{y}_{t2} = -1 + 2x_{t1} - 2x_{t2} + x_{t3}$ $\hat{y}_{t3} = 5 + 2x_{t1} - 4x_{t2} + 3x_{t3}$ Данные наблюдений: <table><tr><th>t</th><th>y_{t1}</th><th>y_{t2}</th><th>y_{t3}</th><th>x_{t1}</th><th>x_{t2}</th><th>x_{t3}</th></tr><tr><td>1</td><td>14</td><td>-8</td><td>-5</td><td>1</td><td>5</td><td>3</td></tr><tr><td>2</td><td>6</td><td>0</td><td>3</td><td>2</td><td>3</td><td>2</td></tr><tr><td>3</td><td>20</td><td>-7</td><td>-8</td><td>4</td><td>8</td><td>3</td></tr><tr><td>4</td><td>-5</td><td>14</td><td>23</td><td>6</td><td>2</td><td>5</td></tr><tr><td>5</td><td>6</td><td>2</td><td>14</td><td>2</td><td>4</td><td>7</td></tr></table>	t	y _{t1}	y _{t2}	y _{t3}	x _{t1}	x _{t2}	x _{t3}	1	14	-8	-5	1	5	3	2	6	0	3	2	3	2	3	20	-7	-8	4	8	3	4	-5	14	23	6	2	5	5	6	2	14	2	4	7	Вариант 2. Структурная система: $y_{t1} = b_{10} + b_{12}y_{t2} + b_{13}y_{t3} + a_{13}x_{t3} + \varepsilon_{t1}$ $y_{t2} = b_{20} + b_{21}y_{t1} + \varepsilon_{t2}$ $y_{t3} = b_{30} + b_{31}y_{t1} + a_{31}x_{t1} + a_{32}x_{t2} + \varepsilon_{t3}.$ Приведенная система: $\hat{y}_{t1} = 5 + 3x_{t1} + 2x_{t2} + 5x_{t3}$ $\hat{y}_{t2} = 1 + 4x_{t1} - 2x_{t2} - 2x_{t3}$ $\hat{y}_{t3} = -1 - 3x_{t1} + x_{t2} + x_{t3}$ Данные наблюдений: <table><tr><th>t</th><th>y_{t1}</th><th>y_{t2}</th><th>y_{t3}</th><th>x_{t1}</th><th>x_{t2}</th><th>x_{t3}</th></tr><tr><td>1</td><td>26</td><td>-5</td><td>-1</td><td>1</td><td>1</td><td>3</td></tr><tr><td>2</td><td>19</td><td>4</td><td>-4</td><td>2</td><td>2</td><td>1</td></tr><tr><td>3</td><td>47</td><td>-4</td><td>-1</td><td>4</td><td>7</td><td>3</td></tr><tr><td>4</td><td>32</td><td>19</td><td>-16</td><td>6</td><td>3</td><td>1</td></tr><tr><td>5</td><td>48</td><td>-5</td><td>-1</td><td>3</td><td>4</td><td>5</td></tr></table>	t	y _{t1}	y _{t2}	y _{t3}	x _{t1}	x _{t2}	x _{t3}	1	26	-5	-1	1	1	3	2	19	4	-4	2	2	1	3	47	-4	-1	4	7	3	4	32	19	-16	6	3	1	5	48	-5	-1	3	4	5
t	y _{t1}	y _{t2}	y _{t3}	x _{t1}	x _{t2}	x _{t3}																																																																															
1	14	-8	-5	1	5	3																																																																															
2	6	0	3	2	3	2																																																																															
3	20	-7	-8	4	8	3																																																																															
4	-5	14	23	6	2	5																																																																															
5	6	2	14	2	4	7																																																																															
t	y _{t1}	y _{t2}	y _{t3}	x _{t1}	x _{t2}	x _{t3}																																																																															
1	26	-5	-1	1	1	3																																																																															
2	19	4	-4	2	2	1																																																																															
3	47	-4	-1	4	7	3																																																																															
4	32	19	-16	6	3	1																																																																															
5	48	-5	-1	3	4	5																																																																															
Вариант 3. Структурная система: $y_{t1} = b_{10} + b_{12}y_{t2} + a_{12}x_{t2} + a_{13}x_{t3} + \varepsilon_{t1}$ $y_{t2} = b_{20} + b_{21}y_{t1} + b_{23}y_{t3} + a_{21}x_{t1} + \varepsilon_{t2}$ $y_{t3} = b_{30} + b_{32}y_{t2} + \varepsilon_{t3}.$ Приведенная система: $\hat{y}_{t1} = 2 - 3x_{t1} + x_{t2} + 4x_{t3}$ $\hat{y}_{t2} = -4 + 2x_{t1} - x_{t2} + 2x_{t3}$ $\hat{y}_{t3} = -3 - 2x_{t1} + 2x_{t2} + 5x_{t3}$ Данные наблюдений: <table><tr><th>t</th><th>y_{t1}</th><th>y_{t2}</th><th>y_{t3}</th><th>x_{t1}</th><th>x_{t2}</th><th>x_{t3}</th></tr><tr><td>1</td><td>-6</td><td>3</td><td>-1</td><td>1</td><td>5</td><td>3</td></tr><tr><td>2</td><td>15</td><td>0</td><td>20</td><td>2</td><td>3</td><td>2</td></tr><tr><td>3</td><td>7</td><td>-1</td><td>17</td><td>4</td><td>8</td><td>3</td></tr><tr><td>4</td><td>4</td><td>18</td><td>13</td><td>6</td><td>2</td><td>5</td></tr><tr><td>5</td><td>29</td><td>20</td><td>38</td><td>2</td><td>4</td><td>7</td></tr></table>	t	y _{t1}	y _{t2}	y _{t3}	x _{t1}	x _{t2}	x _{t3}	1	-6	3	-1	1	5	3	2	15	0	20	2	3	2	3	7	-1	17	4	8	3	4	4	18	13	6	2	5	5	29	20	38	2	4	7	Вариант 4. Структурная система: $y_{t1} = b_{10} + b_{12}y_{t2} + \varepsilon_{t1}$ $y_{t2} = b_{20} + b_{21}y_{t1} + a_{21}x_{t1} + a_{23}x_{t3} + \varepsilon_{t2}$ $y_{t3} = b_{30} + b_{31}y_{t1} + b_{32}y_{t2} + a_{32}x_{t2} + \varepsilon_{t3}.$ Приведенная система: $\hat{y}_{t1} = -3 + 4x_{t1} + x_{t2} - 5x_{t3}$ $\hat{y}_{t2} = 1 - 2x_{t1} - 2x_{t2} + 3x_{t3}$ $\hat{y}_{t3} = 5 - 1x_{t1} - x_{t2} - x_{t3}$ Данные наблюдений: <table><tr><th>t</th><th>y_{t1}</th><th>y_{t2}</th><th>y_{t3}</th><th>x_{t1}</th><th>x_{t2}</th><th>x_{t3}</th></tr><tr><td>1</td><td>11</td><td>-12</td><td>-9</td><td>1</td><td>1</td><td>3</td></tr><tr><td>2</td><td>-1</td><td>-6</td><td>-4</td><td>2</td><td>2</td><td>1</td></tr><tr><td>3</td><td>3</td><td>-13</td><td>-7</td><td>4</td><td>7</td><td>3</td></tr><tr><td>4</td><td>-5</td><td>4</td><td>-8</td><td>6</td><td>3</td><td>1</td></tr><tr><td>5</td><td>-26</td><td>12</td><td>-7</td><td>3</td><td>4</td><td>5</td></tr></table>	t	y _{t1}	y _{t2}	y _{t3}	x _{t1}	x _{t2}	x _{t3}	1	11	-12	-9	1	1	3	2	-1	-6	-4	2	2	1	3	3	-13	-7	4	7	3	4	-5	4	-8	6	3	1	5	-26	12	-7	3	4	5
t	y _{t1}	y _{t2}	y _{t3}	x _{t1}	x _{t2}	x _{t3}																																																																															
1	-6	3	-1	1	5	3																																																																															
2	15	0	20	2	3	2																																																																															
3	7	-1	17	4	8	3																																																																															
4	4	18	13	6	2	5																																																																															
5	29	20	38	2	4	7																																																																															
t	y _{t1}	y _{t2}	y _{t3}	x _{t1}	x _{t2}	x _{t3}																																																																															
1	11	-12	-9	1	1	3																																																																															
2	-1	-6	-4	2	2	1																																																																															
3	3	-13	-7	4	7	3																																																																															
4	-5	4	-8	6	3	1																																																																															
5	-26	12	-7	3	4	5																																																																															

Вариант 5.

Структурная система:

$$y_{t1} = b_{10} + b_{12}y_{t2} + \varepsilon_{t1}$$

$$y_{t2} = b_{20} + b_{21}y_{t1} + b_{23}y_{t3} + a_{23}x_{t3} + \varepsilon_{t2}$$

$$y_{t3} = b_{30} + b_{31}y_{t1} + a_{31}x_{t1} + a_{32}x_{t2} + \varepsilon_{t3}.$$

Приведенная система:

$$\hat{y}_{t1} = 1 + x_{t1} + 2x_{t2} + 2x_{t3}$$

$$\hat{y}_{t2} = -1 + 3x_{t1} + x_{t2} + 3x_{t3}$$

$$\hat{y}_{t3} = 2 + 3x_{t1} + 2x_{t2} + 3x_{t3}$$

Данные наблюдений:

t	y _{t1}	y _{t2}	y _{t3}	x _{t1}	x _{t2}	x _{t3}
1	26	35	44	1	5	3
2	13	17	21	2	3	2
3	14	16	25	4	8	3
4	13	23	24	6	2	5
5	20	17	26	2	4	7

Вариант 6.

Структурная система:

$$y_{t1} = b_{10} + b_{12}y_{t2} + b_{13}y_{t3} + a_{13}x_{t3} + \varepsilon_{t1}$$

$$y_{t2} = b_{20} + b_{21}y_{t1} + \varepsilon_{t2}$$

$$y_{t3} = b_{30} + b_{31}y_{t1} + a_{31}x_{t1} + a_{32}x_{t2} + \varepsilon_{t3}.$$

Приведенная система:

$$\hat{y}_{t1} = 1 + 3x_{t1} + x_{t2} - 2x_{t3}$$

$$\hat{y}_{t2} = -2 + 2x_{t1} + 3x_{t2} + 2x_{t3}$$

$$\hat{y}_{t3} = 1 + 3x_{t1} + 4x_{t2} + 5x_{t3}$$

Данные наблюдений:

t	y _{t1}	y _{t2}	y _{t3}	x _{t1}	x _{t2}	x _{t3}
1	13	25	47	1	1	3
2	15	18	33	2	2	1
3	16	35	62	4	7	3
4	11	27	50	6	3	1
5	10	16	28	3	4	5

Вариант 7.

Структурная система:

$$y_{t1} = b_{10} + b_{12}y_{t2} + a_{12}x_{t2} + a_{13}x_{t3} + \varepsilon_{t1}$$

$$y_{t2} = b_{20} + b_{21}y_{t1} + b_{23}y_{t3} + a_{21}x_{t1} + \varepsilon_{t2}$$

$$y_{t3} = b_{30} + b_{32}y_{t2} + \varepsilon_{t3}.$$

Приведенная система:

$$\hat{y}_{t1} = 1 + 2x_{t1} + 2x_{t2} + 2x_{t3}$$

$$\hat{y}_{t2} = 2 - 1x_{t1} - 2x_{t2} + x_{t3}$$

$$\hat{y}_{t3} = -1 - 4x_{t1} + 4x_{t2} + 3x_{t3}$$

Данные наблюдений:

t	y _{t1}	y _{t2}	y _{t3}	x _{t1}	x _{t2}	x _{t3}
1	26	-11	11	1	5	3
2	30	-3	-18	2	3	2
3	14	-4	15	4	8	3
4	29	-9	37	6	2	5
5	16	-4	9	2	4	7

Вариант 8.

Структурная система:

$$y_{t1} = b_{10} + b_{12}y_{t2} + \varepsilon_{t1}$$

$$y_{t2} = b_{20} + b_{21}y_{t1} + a_{21}x_{t1} + a_{23}x_{t3} + \varepsilon_{t2}$$

$$y_{t3} = b_{30} + b_{31}y_{t1} + b_{32}y_{t2} + a_{32}x_{t2} + \varepsilon_{t3}.$$

Приведенная система:

$$\hat{y}_{t1} = 1 + 1x_{t1} + 4x_{t2} + 2x_{t3}$$

$$\hat{y}_{t2} = -2 - 1x_{t1} + 1x_{t2} - 3x_{t3}$$

$$\hat{y}_{t3} = 2 + 2x_{t1} + x_{t2} + 2x_{t3}$$

Данные наблюдений:

t	y _{t1}	y _{t2}	y _{t3}	x _{t1}	x _{t2}	x _{t3}
1	30	-10	16	1	1	3
2	24	-12	25	2	2	1
3	44	-15	29	4	7	3
4	23	-19	25	6	3	1
5	16	-15	15	3	4	5

Вариант 9.

Структурная система:

$$y_{t1} = b_{10} + b_{12}y_{t2} + \varepsilon_{t1}$$

$$y_{t2} = b_{20} + b_{21}y_{t1} + b_{23}y_{t3} + a_{23}x_{t3} + \varepsilon_{t2}$$

$$y_{t3} = b_{30} + b_{31}y_{t1} + a_{31}x_{t1} + a_{32}x_{t2} + \varepsilon_{t3}.$$

Приведенная система:

$$\hat{y}_{t1} = 2 + 2x_{t1} + x_{t2} + 2x_{t3}$$

$$\hat{y}_{t2} = 3 - 3x_{t1} - 2x_{t2} + x_{t3}$$

$$\hat{y}_{t3} = 1 + x_{t1} + x_{t2} + 3x_{t3}$$

Данные наблюдений:

t	y _{t1}	y _{t2}	y _{t3}	x _{t1}	x _{t2}	x _{t3}
1	16	-9	15	1	5	3
2	12	-6	12	2	3	2
3	25	-23	24	4	8	3
4	24	-12	23	6	2	5
5	25	-4	28	2	4	7

Вариант 10.

Структурная система:

$$y_{t1} = b_{10} + b_{12}y_{t2} + b_{13}y_{t3} + a_{13}x_{t3} + \varepsilon_{t1}$$

$$y_{t2} = b_{20} + b_{21}y_{t1} + \varepsilon_{t2}$$

$$y_{t3} = b_{30} + b_{31}y_{t1} + a_{31}x_{t1} + a_{32}x_{t2} + \varepsilon_{t3}.$$

Приведенная система:

$$\hat{y}_{t1} = 2 + x_{t1} + 2x_{t2} + 2x_{t3}$$

$$\hat{y}_{t2} = 6 - 4x_{t1} - 5x_{t2} + 2x_{t3}$$

$$\hat{y}_{t3} = -2 + 3x_{t1} + 4x_{t2} + x_{t3}$$

Данные наблюдений:

t	y _{t1}	y _{t2}	y _{t3}	x _{t1}	x _{t2}	x _{t3}
1	12	1	7	1	1	3
2	9	-9	13	2	2	1
3	27	-40	43	4	7	3
4	14	-29	28	6	3	1
5	24	-16	28	3	4	5

СПИСОК ЛИТЕРАТУРЫ

Основная учебная литература

1. Эконометрика : учебник для бакалавриата и магистратуры / И. И. Елисеева [и др.] ; под редакцией И. И. Елисеевой. — Москва : Издательство Юрайт, 2019. — 449 с. — (Бакалавр и магистр. Академический курс). — ISBN 978-5-534-00313-0. — С. 251 — 282 — Текст : электронный // ЭБС Юрайт [сайт]. — URL: <https://biblio-online.ru/bcode/431129/p.251-282> (дата обращения: 12.01.2019).
2. Галочкин, В. Т. Эконометрика : учебник и практикум для бакалавриата и специалитета / В. Т. Галочкин. — Москва : Издательство Юрайт, 2019. — 288 с. — (Бакалавр и специалист). — ISBN 978-5-534-10751-7. — Текст : электронный // ЭБС Юрайт [сайт]. — URL: <https://biblio-online.ru/bcode/431440> (дата обращения: 12.01.2019).

Дополнительная литература

1. Ниворожкина, Л. И. Эконометрика : теория и практика : учеб. пособие / Л.И. Ниворожкина, С.В. Арженовский, Е.П. Кокина. — Москва : РИОР : ИНФРА-М, 2018. — 207 с. — (Высшее образование). — DOI: <https://doi.org/10.12737/1698-5>. - ISBN 978-5-16-103231-2. - Текст : электронный. - URL: <https://new.znaniy.com/catalog/product/90758> (дата обращения: 31.01.2020)
2. Бородич, С. А. Эконометрика. Практикум : учеб. пособие / С.А. Бородич. — Минск : Новое знание ; Москва : ИНФРА-М, 2018. — 329 с. : ил. — (Высшее образование: Бакалавриат). - ISBN 978-5-16-100513-2. - Текст : электронный. - URL: <https://new.znaniy.com/catalog/product/988809> (дата обращения: 31.01.2019)
3. Яковлев, В. П. Эконометрика: Учебник для бакалавров / Яковлев В.П. - Москва : Дашков и К, 2016. - 384 с. (Учебные издания для бакалавров) ISBN 978-5-394-02532-7. - Текст : электронный. - URL: <https://new.znaniy.com/catalog/product/519496> (дата обращения: 31.01.2019)
4. Мардас, А. Н. Эконометрика : учебник и практикум для академического бакалавриата / А. Н. Мардас. — 2-е изд., испр. и доп. — Москва : Издательство Юрайт, 2019. — 180 с. — (Бакалавр. Академический курс). — ISBN 978-5-9916-8164-3. — Текст : электронный // ЭБС Юрайт [сайт]. — URL: <https://biblio-online.ru/bcode/434110> (дата обращения: 12.01.2019).
5. Кремер, Н. Ш. Эконометрика : учебник и практикум для вузов / Н. Ш. Кремер, Б. А. Путко ; под редакцией Н. Ш. Кремера. — 4-е изд., испр. и доп. — Москва : Издательство Юрайт, 2020. — 308 с. — (Высшее образование). — ISBN 978-5-534-08710-9. — Текст : электронный // ЭБС Юрайт [сайт]. — URL: <https://biblio-online.ru/bcode/449750> (дата обращения: 01.03.2020).
6. Костюнин, В. И. Эконометрика : учебник и практикум для вузов / В. И. Костюнин. — Москва : Издательство Юрайт, 2020. — 285 с. — (Высшее образование). — ISBN 978-5-534-02660-3. — Текст : электронный // ЭБС Юрайт [сайт]. — URL: <https://biblio-online.ru/bcode/450113> (дата обращения: 01.03.2020).
7. Эконометрика : учебник и практикум для вузов / О. А. Демидова, Д. И. Малахов. — Москва : Издательство Юрайт, 2020. — 334 с. — (Высшее образование). — ISBN 978-5-534-00625-4. — Текст : электронный // ЭБС Юрайт [сайт]. с. 15 — URL: <https://biblio-online.ru/bcode/450357/p.15> (дата обращения: 01.03.2020).

8. *Евсеев, Е. А.* Эконометрика : учебное пособие для бакалавриата и специалитета / Е. А. Евсеев, В. М. Буре. — 2-е изд., испр. и доп. — Москва : Издательство Юрайт, 2019. — 186 с. — (Бакалавр и специалист). — ISBN 978-5-534-10752-4. — Текст : электронный // ЭБС Юрайт [сайт]. — URL: <https://biblio-online.ru/bcode/431441> (дата обращения: 12.01.2019).

Приложения 1-4 СТАТИСТИКО-МАТЕМАТИЧЕСКИЕ ТАБЛИЦЫ

Приложение 1. Таблица значений F – критерия Фишера при уровне значимости $\alpha = 0,05$

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

Приложение 2. Критические значения t – критерия Стьюдента при уровне значимости 0,10, 0,05, 0,01 (двухсторонний)

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

Приложение 3. Значения статистик Дарбина-Уотсона $d_L d_U$ при 5%-ном уровне

значимости

n	$k = 1$		$k = 2$		$k = 3$		$k = 4$		$k = 5$	
	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
6	0,61	1,40								
7	0,70	1,36	0,47	1,90						
8	0,76	1,33	0,56	1,78	0,37	2,29				
9	0,82	1,32	0,63	1,70	0,46	2,13				
10	0,88	1,32	0,70	1,64	0,53	2,02				
11	0,93	1,32	0,66	1,60	0,60	1,93				
12	0,97	1,33	0,81	1,58	0,66	1,86				
13	1,01	1,34	0,86	1,56	0,72	1,82				
14	1,05	1,35	0,91	1,55	0,77	1,78				
15	1,08	1,36	0,95	1,54	0,82	1,75	0,69	1,97	0,56	2,21
16	1,10	1,37	0,98	1,54	0,86	1,73	0,74	1,93	0,62	2,15
17	1,13	1,38	1,02	1,54	0,90	1,71	0,78	1,90	0,67	2,10
18	1,16	1,39	1,05	1,53	0,93	1,69	0,82	1,87	0,71	2,06
19	1,18	1,40	1,08	1,53	0,97	1,68	0,85	1,85	0,75	2,02
20	1,20	1,41	1,10	1,54	1,00	1,68	0,90	1,83	0,79	1,99
21	1,22	1,42	1,13	1,54	1,03	1,67	0,93	1,81	0,83	1,96
22	1,24	1,43	1,15	1,54	1,05	1,66	0,96	1,80	0,86	1,94
23	1,26	1,44	1,17	1,54	1,08	1,66	0,99	1,79	0,90	1,92
24	1,27	1,45	1,19	1,55	1,10	1,66	1,01	1,78	0,93	1,99
25	1,29	1,45	1,21	1,55	1,12	1,66	1,04	1,77	0,95	1,89
26	1,30	1,46	1,22	1,55	1,14	1,65	1,06	1,76	0,98	1,88
27	1,32	1,47	1,24	1,56	1,16	1,65	1,08	1,76	1,01	1,86
28	1,33	1,48	1,26	1,56	1,18	1,65	1,10	1,75	1,03	1,85
29	1,34	1,48	1,27	1,56	1,20	1,65	1,12	1,74	1,05	1,84
30	1,35	1,49	1,28	1,57	1,21	1,65	1,14	1,74	1,07	1,83

Приложение 4. Критические значения RS критерия

Значения a и b при уровне значимости $\alpha=0.05$.		
n	a	b
12	2,800	3,910
13	2,860	4,000
14	2,920	4,090
15	2,965	4,165
16	3,010	4,240
17	3,055	4,275
18	3,100	4,310
19	3,140	4,400
20	3,180	4,490
21	3,212	4,534
22	3,244	4,578
23	3,276	4,622

Приложение 5. Решение практических заданий в Excel

Применение инструмента регрессия для расчета параметров парной линейной регрессии

С помощью инструмента анализа данных **Регрессия**, помимо результатов регрессионной статистики, дисперсионного анализа и доверительных интервалов, можно получить остатки и графики подбора линии регрессии, остатков и нормальной вероятности. Порядок действий следующий:

1) проверьте доступ к пакету анализа. В главном меню последовательно **выберите Параметры Excel /Надстройки**. Установите флажок **Пакет анализа** (рис 1)^

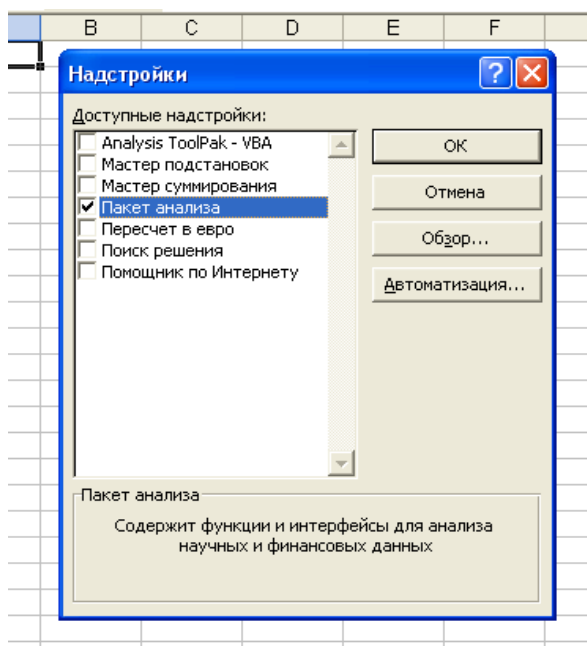


Рис. 1. Подключение надстройки **Пакет анализа**

2) в главном меню выберите **Вкладка Данные/ Анализ данных/Регрессия**. Щелкните по кнопке ОК;

3) заполните диалоговое окно ввода данных и параметров вывода (рис. 2):

Входной интервал Y - диапазон, содержащий данные результативного признака;

Входной интервал X - диапазон, содержащий данные факторов независимого признака;

Метки - флажок, который указывает, содержит ли первая строка названия столбцов или нет;

Константа - ноль - флажок, указывающий на наличие или отсутствие свободного члена в уравнении;

Выходной интервал - достаточно указать левую верхнюю ячейку будущего диапазона;

Новый рабочий лист - можно задать произвольное имя нового **листа**.

Если необходимо получить информацию и графики остатков, установите соответствующие флажки в диалоговом окне. Щелкните по кнопке ОК.

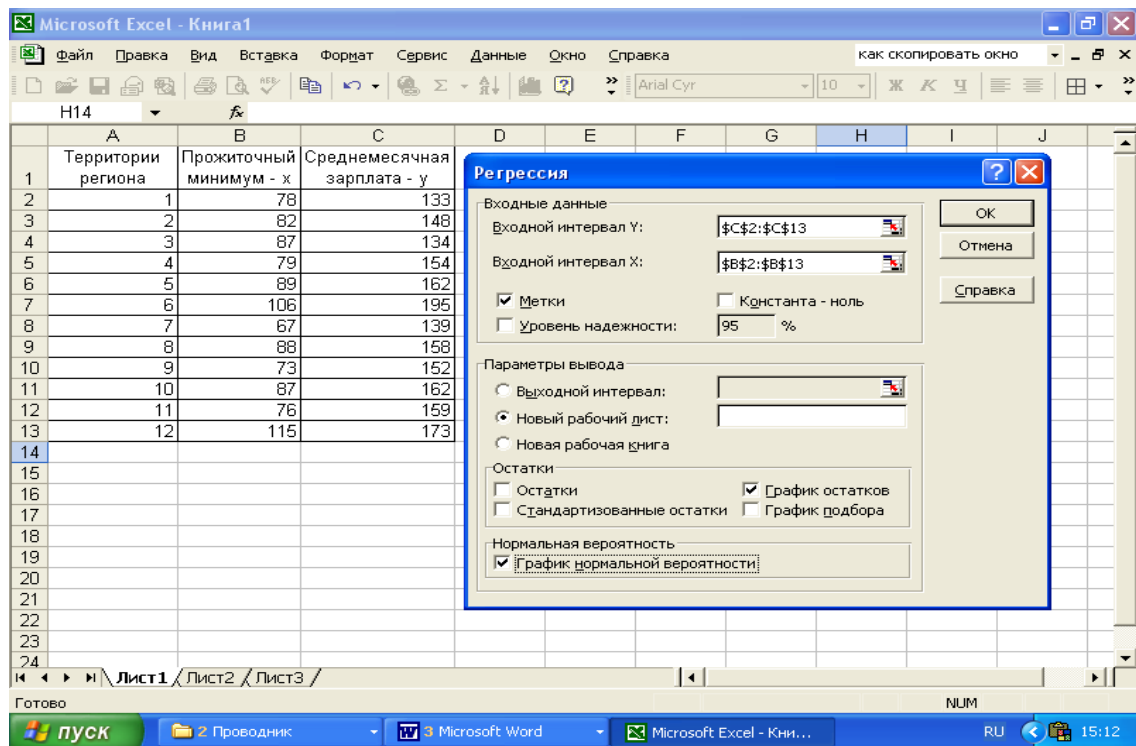


Рис. 2. Диалоговое окно ввода параметров инструмента Регрессия

Результаты регрессионного анализа для данных из примера 2 представлены на рис. 3.

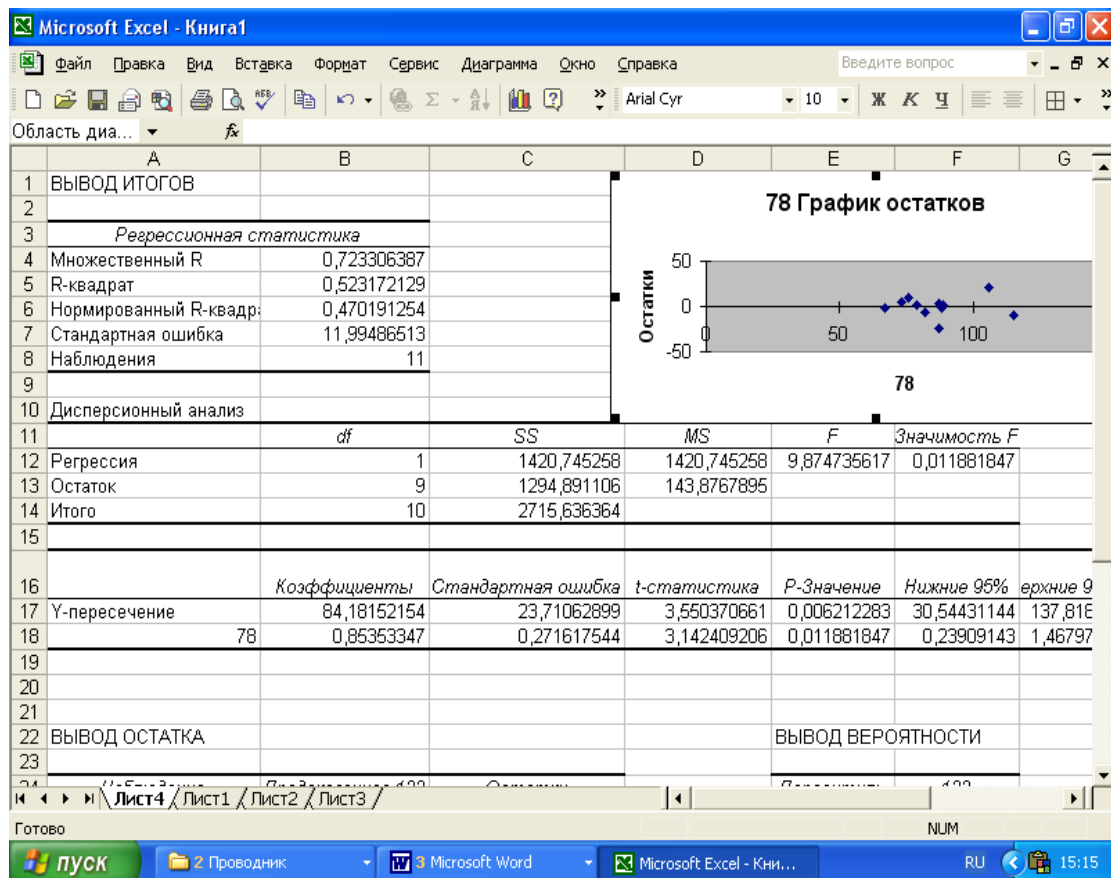


Рис. 3. Результат применения инструмента Регрессия

Нахождение модели нелинейной регрессии в Excel

Пример: В таблице 31 приведены значения независимой переменной X (доход американской семьи в тысяч долларов) и значения зависимой переменной Y (доля расходов на товары длительного пользования в процентах от общей суммы расходов).

Таблица 1

x_i	1	2	3	4	5	6
y_i	10	13.4	15.4	16.5	18.6	19.1

Решение:

Команда «Добавить линию тренда». Используется для выделения тренда (медленных изменений) при анализе временных рядов. Однако эту команду можно использовать и для построения уравнения нелинейной регрессии, рассматривая в качестве времени t независимую переменную x .

Эта команда позволяет построить следующие уравнения регрессии:

- линейную $\hat{y} = b_0 + b_1x$
- полиномиальную $\hat{y} = b_0 + b_1x + \dots + b_kx^k$ ($k \leq 6$);
- логарифмическую $\hat{y} = b_0 + b_1 \cdot \ln x$
- степенную $\hat{y} = b_0 \cdot x^{b_1}$;
- экспоненциальную $\hat{y} = b_0 e^{b_1x}$.

Для построения одной из перечисленных регрессий необходимо выполнить следующие шаги:

Шаг 1. В выбранном листе Excel ввести по столбцам исходные данные $\{x_i, y_i\}, i = 1, 2, \dots, n$ (см. рис. 2.1).

Шаг 2. По этим данным построить график в декартовой системе координат (см. рис 4).

Шаг 3. Установить курсор на построенном графике, сделать щелчок правой кнопкой и в появившемся контекстном меню выполнить команду *Добавить линию тренда* (см. рис. 4).

Шаг 4. В появившемся диалоговом окне (см. рис. 5) активизировать закладку «Тип» и выбрать нужное уравнение регрессии.

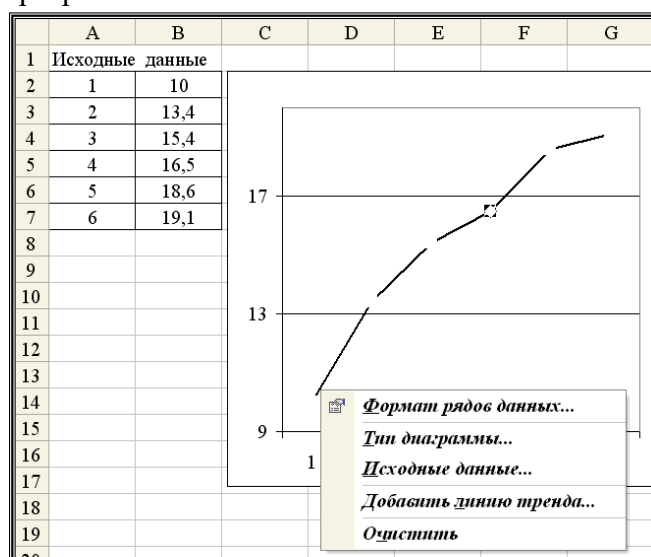


Рис. 4 – Построение графика по исходным данным

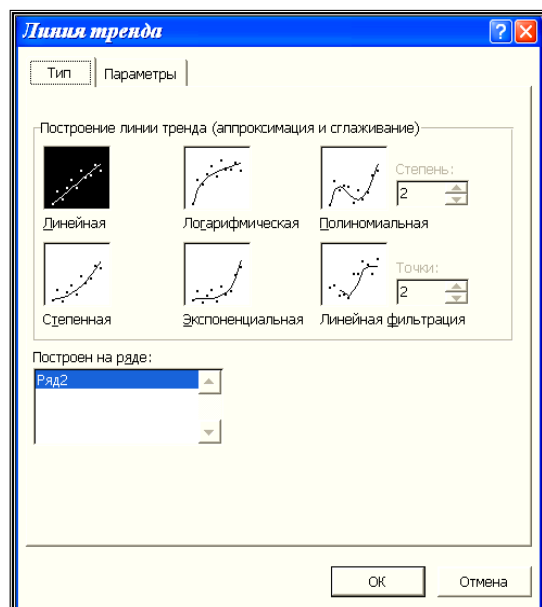


Рис. 5 – Выбор вида уравнения регрессии

Шаг 5. Активизировать закладку «Параметры» (см. рис. 6) и «включить» необходимые для нас опции:

- «Показать уравнение на диаграмме» - на диаграмме будет показано выбранное уравнение регрессии с вычисленными коэффициентами;

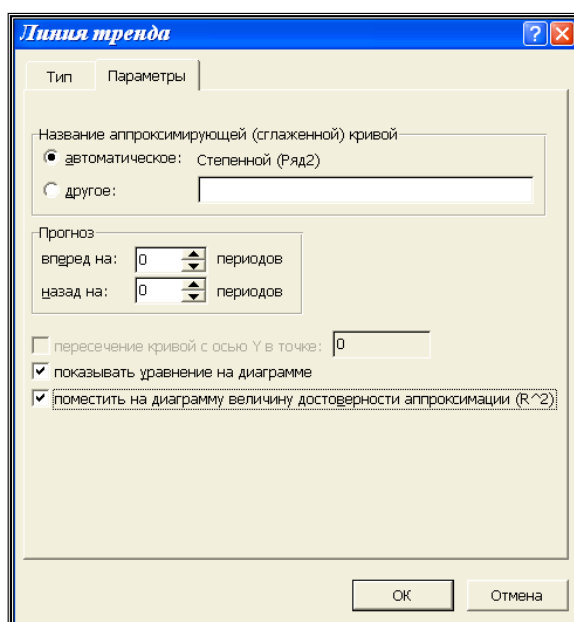


Рис. 6– Задание опций вывода информации

- «Поместить на диаграмму величину достоверности аппроксимации (R^2)» - на диаграмме будет показана значение коэффициент детерминации R^2 (для нелинейной регрессии - индекс детерминации). Если по построенному уравнению регрессии необходимо выполнить прогноз, то нужно указать число периодов прогноза (см. рис. 2.3).

Назначение других опций понятны из своих названий.

Шаг 6. После задания всех перечисленных опций щелкнуть на кнопке «ОК» и на диаграмме появиться формула построенного уравнения регрессии и значение индекса детерминации R^2 (выделено на рис. 7 затемнением).

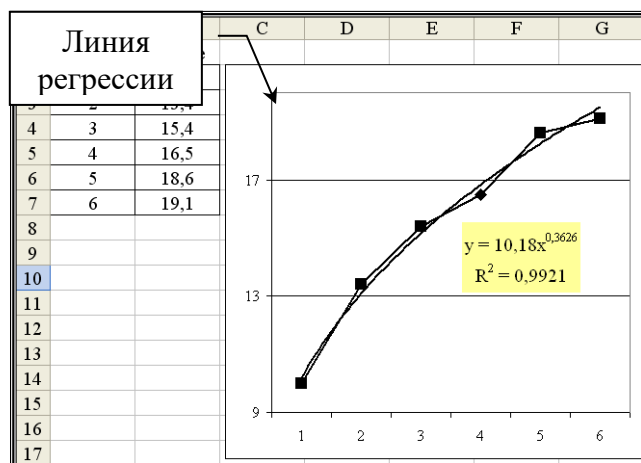


Рис. 7– График и уравнение построенной регрессии

Решение. Построение уравнения $\hat{y} = b_0 \cdot x^{b_1}$ осуществляем по описанным выше шагам. Получаем уравнение

$$\hat{y}(x) = 10.18 \cdot x^{0.3626},$$

для которого коэффициент детерминации равен $R^2 = 0.9921$ (см. рис. 2.4). Такая величина говорит о хорошем соответствии построенного уравнения исходным данным.

Выбор наилучшей нелинейной регрессии по приведенному коэффициенту детерминации

Приведенный коэффициент детерминации. Коэффициент детерминации R^2 характеризует близость построенной регрессии к исходным данным, которые содержат «нежелательную» случайную составляющую ε . Очевидно, что, построив по данным таб. 2.1 полином 5-ого порядка, получаем «идеальное» значение $R^2 = 1$, но такое уравнение содержит в себе не только независимую переменную X , но составляющую ε и это снижает точность использования построенного уравнения для прогноза. Поэтому при выборе уравнения регрессии надо учитывать не только величину R^2 , но и «сложность» регрессионного уравнения, определяемое количеством коэффициентов уравнения. Такой учет удачно реализован в так называемом *приведенном коэффициенте детерминации* $\hat{R}^2$.

Для построения каждого уравнения выполняем шаги 2 – 6 (для первого уравнения еще и шаг 1) и размещаем в одном документе шесть окон, в которых выводятся найденные уравнения регрессии уравнения и величина R^2 . Затем формулу уравнения и R^2 заносим в таблицу 2.2. Далее по формуле (2.1) вычисляем приведенный коэффициент детерминации $\hat{R}^2$ и заносим эти значения также в таблицу (см. таб. 2).

Таблица 2

№	Уравнение	R^2	$\hat{R}^2$
1	$\hat{y} = 9.28 + 1.777x$	0.949	0.938
2	$\hat{y} = 9.8759 + 5.1289 \cdot \ln x$	0.9916	0.9895
3	$\hat{y} = 6.93 + 3.5396x - 0.2518x^2$ (полиномиальная, $m = 2$)	0.9896	0.9827
4	$\hat{y} = 5.8333 + 4.9192 \cdot x - 0.7087 \cdot x^2 -$ $- 0.0435 \cdot x^3$ (полиномиальная, $m = 3$)	0.9917	0.9792
5	$\hat{y} = 10.18x^{0.3626}$	0.9921	0.9901
6	$\hat{y} = 9.8675 \cdot e^{0.1225x}$	0.9029	0,8786

В качестве «наилучшего» уравнения регрессии выбираем уравнение, имеющее наибольшую величину приведенный коэффициент детерминации $\hat{R}^2$. Из таб. 2.2 видно, что таким уравнением является степенная функции (в таблице строка с этой функцией выделена серым цветом)

$$\hat{y} = 10.18x^{0.3626},$$

имеющая величину $\hat{R}^2 = 0.9901$.

Учебное издание

Кречетова Марина Александровна

ЭКОНОМЕТРИКА

Практикум по дисциплине «Эконометрика»
для студентов, обучающихся по направлению:
38.03.01 – Экономика
(все профили)

Редактор **Т. И. Головки**
Технический редактор **М. А. Гомонок**

Подписано в печать **14.11.2019 г.**
Бумага писчая. Ризография. Формат 60 × 84 ¹/₁₆.
Уч.-изд. л. 4,38. **Тираж 40 экз. Заказ 474.**

Новокузнецкий институт (филиал)
Кемеровского государственного университета.
654000, г. Новокузнецк, просп. Metallургов, 19, тел. 74-14-95.
root@nbikemsu.ru
Центр издательской деятельности
Цена договорная