

Федеральное государственное бюджетное образовательное учреждение
высшего образования
"КАЛИНИНГРАДСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ"

Институт отраслевой экономики и управления

Ю. Я. Настин

ЭКОНОМЕТРИКА

Учебно-методическое пособие по изучению дисциплины
для студентов бакалавриата по направлению подготовки 38.03.01 Экономика,
профиль "Прикладная экономика"

Калининград
Издательство ФГБОУ ВО «КГТУ»
2023

330.43(075)

Рецензент

кандидат экономических наук, доцент кафедры экономики и финансов Института отраслевой экономики и управления ФГБОУ ВО "КГТУ"

И. А. Крамаренко

Настин, Ю. Я.

Эконометрика: учеб.-метод. пособие по изучению дисциплины для бакалавриата по напр. подгот. 38.03.01 Экономика, профиль "Прикладная экономика" / Ю. Я. Настин. – Калининград: Изд-во ФГБОУ ВО «КГТУ», 2023. – 107 с.

Учебно-методическое пособие по дисциплине "Эконометрика" - части 1 и 2 - подготовлено в соответствии с учебным планом и рабочей программой модуля "Анализ в решении экономических задач" по направлению подготовки 38.03.01 Экономика, профиль "Прикладная экономика". В учебно-методическое пособие по изучению дисциплины вошёл традиционный материал, который изучается в высшей школе. Особенностью пособия является то, что оно учитывает: а) трёхкратное увеличение трудоёмкости изучения дисциплины (было 2 зет, стало 6 зет); б) изучение дисциплины в двух семестрах с двумя промежуточными аттестациями. Фактически в одном пособии содержатся два. Этим объясняется несколько увеличенный объём пособия. Большое внимание уделено методике решения исследовательских задач как ручным способом, так и в MS Excel.

Рис. 25, табл. 19, список лит. – 12 наименований

Учебно-методическое пособие рассмотрено и одобрено в качестве локального электронного методического материала кафедрой экономической теории и инструментальных методов, протокол № 12 от 30.06.2023 г.

Учебно-методическое пособие рассмотрено и рекомендовано к изданию в качестве локального электронного методического материала для использования в учебном процессе методической комиссией Института отраслевой экономики и управления ФГБОУ ВО "КГТУ", протокол № 9 от 05.07.2023 г.

330.43(075)

© Федеральное государственное бюджетное образовательное учреждение высшего образования "Калининградский государственный технический университет", 2023 г.

© Настин Ю. Я., 2023 г.

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ	5
ЧАСТЬ 1.....	8
1 Тематический план части 1 дисциплины и методические указания по её изучению	8
Тема 1. Основные понятия эконометрики.....	8
Тема 2. Парный регрессионный анализ.....	12
Тема 3. Множественный регрессионный анализ.....	18
Тема 4. Нелинейный регрессионный анализ.....	25
2 Методические указания для подготовки к практическим занятиям по 1-й части дисциплины	30
3 Методические указания по выполнению контрольной работы № 1	34
3.1 Общие сведения, выбор варианта и вопросов, исходных данных для задач.....	34
3.2 Пример ручного решения задачи-1: "Моделирование магазина на основе парной линейной регрессии"	36
3.3 Пример ручного решения задачи-2: "Моделирование магазина на основе троичной линейной регрессии"	40
3.4 Пример решения задачи-3: "Моделирование магазина на основе линейной троичной регрессии в MS Excel"	45
3.5 Пример решения задачи-4: "Моделирование магазина на основе нелинейной регрессии в MS Excel"	48
4 Методические указания по подготовке к промежуточной аттестации - зачёту	51
5 Методические указания по выполнению самостоятельной работы по дисциплине – части 1	53
ЧАСТЬ 2.....	54
6 Тематический план части 2 дисциплины и методические указания по её изучению	54
Тема 5. Инструментальные методы регрессионного анализа	54
Тема 6. Регрессионные модели временных рядов и прогнозирование	57
Тема 7. Обобщённая линейная модель и гетероскедастичность пространственной выборки	63
Тема 8. Обобщённая линейная модель и автокорреляция остатков.....	70
7 Методические указания по подготовке к практическим занятиям по 2-й части дисциплины.....	74
8 Методические указания по выполнению контрольной работы № 2	78

8.1 Общие требования, выбор варианта, вопросов и исходных данных для задач.....	78
8.2 Пример ручного решения задачи-2.1: "Проверка возможности объединения выборок для линейной парной модели магазина"	79
8.3 Пример ручного решения задачи-2.2: "Проверка линейной парной модели магазина на гомоскедастичность по двум критериям"	82
8.4 Пример решения задачи 2.3: "Выделение тренда временного ряда методами скользящей средней в MS Excel"	84
8.5 Пример решения задачи 2.4: "Выделение тренда временного ряда методом экспоненциального сглаживания в MS Excel"	87
8.6 Пример решения задачи 2.5: "Прогнозирование на основе авторегрессии 1-го порядка в MS Excel".....	89
9 Методические указания по подготовке к промежуточной аттестации – дифференцированному зачёту	92
10 Методические указания по выполнению самостоятельной работы по дисциплине – части 2.....	95
СПИСОК ИСТОЧНИКОВ	96
ПРИЛОЖЕНИЕ А. Исходные данные для контрольных работ № 1 и № 2.....	97
ПРИЛОЖЕНИЕ Б. Образец страницы СОДЕРЖАНИЕ для контрольных работ № 1.....	100
ПРИЛОЖЕНИЕ В. Образец страницы СОДЕРЖАНИЕ для контрольных работ № 2.....	101
ПРИЛОЖЕНИЕ Г. Построение графика линейной парной регрессии в Word через Excel.....	102
ПРИЛОЖЕНИЕ Д. Некоторые формулы из математической статистики и эконометрики.....	103
ПРИЛОЖЕНИЕ Е. Статистические таблицы (учебные).....	105

ВВЕДЕНИЕ

Учебно-методическое пособие по дисциплине "Эконометрика" - части 1 и 2 - подготовлено в соответствии с учебным планом и рабочей программой модуля "Анализ в решении экономических задач" по направлению подготовки 38.03.01 Экономика, профиль "Прикладная экономика".

Дисциплина изучается студентами очной и очно-заочной формы обучения в двух семестрах: 4-м и 5-м. Трудоёмкость изучения дисциплины составляет 6 зет ($6 \cdot 36 = 216$ академических часов) и разделена пополам по семестрам на две части. После 4-го семестра студенты сдают зачёт по 1-й части, после 5-го семестра – зачёт с оценкой по всей дисциплине. В каждом из семестров для всех форм обучения предусмотрено выполнение контрольной работы, всего их две. В таблице 1 показано распределение аудиторных часов по форме обучения, по частям дисциплины и темам.

Таблица 1 - Распределение часов аудиторных занятий по формам обучения и по темам, час.

Тема	Форма обучения			
	очная		очно-заочная	
	лек	ПЗ	лек	ПЗ
Часть 1				
1 Основные понятия эконометрики	4	2	2	2
2 Парный регрессионный анализ	4	4		2
3 Множественный регрессионный анализ	4	4	2	2
4 Нелинейный регрессионный анализ	4	4		2
Итого часть 1	16	14	4	8
Часть 2				
5 Инструментальные методы регрессионного анализа	2	4	2	2
6 Регрессионные модели временных рядов и прогнозирование	4	4		2
7 Обобщённая линейная модель и гетероскедастичность пространственной выборки	4	4	2	2
8 Обобщённая линейная модель и автокорреляция остатков временного ряда	4	4		2
Итого часть 2	14	16	4	8
Итого части 1 и 2	30	30	8	16

Контрольные работы № 1 и № 2 представляют собой исследования, по форме близкие для всех вариантов, но с индивидуальными исходными данными.

По очной форме обучения на начальном этапе студенты занимаются выполнением контрольной работы в аудитории на практических занятиях – темы

совпадают. Завершение контрольных работ происходит в режиме самостоятельной работы дома.

По очно-заочной форме обучения в каждом из двух семестров всего по две лекции, из которых первая - установочная. Её главное назначение – "зарядить" студентов минимальным объёмом знаний для начала выполнения ими контрольной работы (выбор варианта, выбор исходных данных, рекомендованные источники, требования к оформлению и пр.). Далее выполнение контрольной работы будет только поддерживаться на практических занятиях.

По результатам аудиторных занятий и самостоятельной работы студенты обеих форм обучения представляют свои контрольные работы, успешная защита которых – обязательное условие для промежуточных аттестаций.

Результаты обучения (знания, умения и владения) включают следующие компетенции студента:

знать:

базовые понятия эконометрического моделирования, относящиеся к сфере профессиональной деятельности в области экономических процессов;

уметь:

разрабатывать эконометрические модели исследуемых процессов, относящихся к области экономических процессов;

владеть:

навыками применения эконометрических моделей в области экономических процессов.

Оценочные средства поэтапного формирования результатов освоения дисциплины включают тестовые задания, типовые задания для проведения практических занятий и типовые задания для контрольных работ студентов всех форм обучения.

Тестовые задания используются для оценки освоения тем дисциплины. Тестирование обучающихся по дневной форме проводится на практических занятиях после рассмотрения на лекциях соответствующих тем; по очно-заочной форме – в основном в рамках самостоятельной работы.

Выполнение тестового задания по темам состоит, как правило, в выборе одного правильного ответа из предлагаемых трёх вариантов. Оценка по результатам тестирования соответствует следующей шкале (% правильных ответов):

- от 0 до 55 – "неудовлетворительно";
- от 56 до 70 – "удовлетворительно";
- от 71 до 85 – "хорошо";
- от 86 до 100 – "отлично".

Типовые задания для проведения практических занятий дают примеры формирования эконометрических моделей и с их помощью исследований соответствующих экономических ситуаций.

Структура учебно-методического пособия. Оно состоит из двух частей: в 4-м семестре изучается 1-я часть дисциплины, в 5-м семестре - 2-я часть дисциплины. Каждая часть состоит из 4-х тем: 1-4 и 5-8. В каждой теме выделено по 4-6 вопросов. Структура каждой - из двух частей:

- разделы 1-й и 6-й: методические указания по изучению дисциплины – теория;
- разделы 2-й и 7-й: методические указания по практическим занятиям;
- разделы 3-й и 8-й: методические указания по контрольным работам № 1 и № 2 соответственно;
- разделы 4-й и 9-й: методические указания по сдаче зачёта и зачёта с оценкой соответственно;
- разделы 5-й и 10-й: методические указания по выполнению самостоятельной работы.

В работе [3, с. 7] приведены четыре особенности эконометрики, которые затрудняют её изучение. С ними полезно познакомиться.

ЧАСТЬ 1

1 Тематический план части 1 дисциплины и методические указания по её изучению

Тема 1. Основные понятия эконометрики

Форма занятий – лекции.

Вопросы для изучения:

Вопрос 1. История возникновения эконометрики.

Вопрос 2. Предмет и метод эконометрики.

Вопрос 3. Две модели магазина (предприятия) в форме линейной регрессии (сквозная задача-1 и сквозная задача-2).

Вопрос 4. Модель прогнозирования банкротства предприятия в форме линейной регрессии.

Вопрос 5. Методология и этапы эконометрического моделирования.

Методические указания по изучению темы 1

Цель изучения темы – освоить элементы методологии эконометрического моделирования, познакомиться с разными формами таких моделей, понять особенности их корректного построения и использования. Примерно об этом в середине XIX века К. Маркс написал своему соратнику Ф. Энгельсу; ниже - цитата из его письма.

"... ты знаешь таблицы, в которых цены, учетный процент и т. д. представлены в их движении ... в виде зигзагообразных линий. Я неоднократно пытался - для анализа кризисов - вычислить эти up and downs и думал ... математически вывести из этого главные законы кризисов".

По сути, Маркс здесь сформулировал главную задачу эконометрики.

Вопрос 1. История возникновения эконометрики.

Изучение этого вопроса поможет составить общее представление о формировании науки эконометрики, её границах, роли отечественных учёных и способе её появления – на основе синтеза нескольких смежных наук. При изучении этого вопроса рекомендуем историю разделить на этапы. Подробно - в работах [1, с. 6-8] и [3, с. 10-11].

Вопрос 2. Предмет и метод эконометрики.

В этом вопросе важно чётко уяснить предмет эконометрики, главное в нём – массовость в экономических явлениях и взаимосвязях, в том числе в финансах, менеджменте. Экономическая теория эти взаимосвязи выделяет, качественно

описывает, и это – её результат. Эконометрика же с этого этапа познания только начинает свою работу и доводит словесное описание взаимосвязей до математико-статистических (эконометрических) моделей, базируясь на фактических данных. Рекомендуем вдуматься в определение эконометрики.

Эконометрика – это наука, призванная наполнить эмпирическим содержанием *априорные* экономические рассуждения. - Лоуренс Клейн (1920-2013).

Традиционно основным разделом для эконометрики является *регрессионный анализ*. Регрессией в строгом смысле называется условное математическое ожидание случайной величины Y , которое является функцией, зависящей от переменной X : $M(Y|X) = \mu(X)$. На практике неизвестную функцию $\mu(X)$ с помощью метода наименьших квадратов заменяют приближённой функцией

$$\hat{y} = \hat{y}(X, b_0, b_1, \dots, b_p),$$

где b_0, b_1, \dots, b_p - оценочные значения параметров функции.

Подробно этот вопрос раскрывается ниже и в работе [1, Кремер-Путко, с. 6-11].

Вопрос 3. Две модели магазина (предприятия) в форме линейной регрессии (сквозная задача-1 и сквозная задача-2).

Сквозная задача-1. Компания ООО "Морепродукты" имеет семь магазинов и планирует открыть восьмой с проектными показателями: торговая площадь $X_1=1100$ кв. м, площадь паркинга $X_2 = 80$ а/м. Разрабатывается бизнес-план и эконометрическая модель - уравнение парной линейной регрессии:

$$\hat{y} = b_0 + b_1x, \quad (1.1)$$

где \hat{y} – теоретическая – регрессионная – суточная выручка магазина, руб./сут.; x – торговая площадь, кв. м; b_0 и b_1 – выборочные оценки параметров β_0 и β_1 уравнения регрессии (1.1).

Исходными данными для нас (вариант А-Ъ, таблицы А.1 и А.6) служат две выборки по семь магазинов, таблица 1.1, показатели X_1 и Y .

Таблица 1.1 – Исходные данные для сквозных задач 1 и 2

Показатель	Номер магазина								Ед. изм.
	1	2	3	4	5	6	7	8*	
X_1 – торговая площадь	1	1	2	3	4	5	8	11	×100 кв. м
X_2 - площадь паркинга	1	2	2	3	3	4	6	8	×10 шт.
Y – среднесуточная выручка	2	3	4	5	5	7	14	?	×10 ⁵ руб./сут.

*Графа 8 - проектные показатели магазина-8.

Построение и анализ модели. Фактические – наблюдаемые - "точки" из таблицы 1.1 (корреляционное поле) можно видеть на рисунке 1.1. Они хорошо формируют тренд в виде прямой линии.

Для определения значений параметров b_0 и b_1 используем метод наименьших квадратов и получаем эконометрическую модель:

$$\hat{y} = 0,43 + 1,54x_1. \quad (1.2)$$

Анализ модели позволяет получить важную информацию, например, какие магазины "передовые" (расположены выше линии регрессии), а какие – "отстающие".

Сквозная задача-2. Модель магазина в форме тройной линейной регрессии. Постановка задачи. Специалисты ООО "Морепродукты" решают ввести в модель вторую объясняющую переменную x_2 – площадь паркинга:

$$\hat{y} = b_0 + b_1x_1 + b_2x_2. \quad (1.3)$$

Построение и анализ модели. Для отыскания значений параметров b_0 , b_1 , b_2 применяем МНК-метод, получаем модель:

$$\hat{y} = - 0,88 + 0,50x_1 + 1,63x_2. \quad (1.4)$$

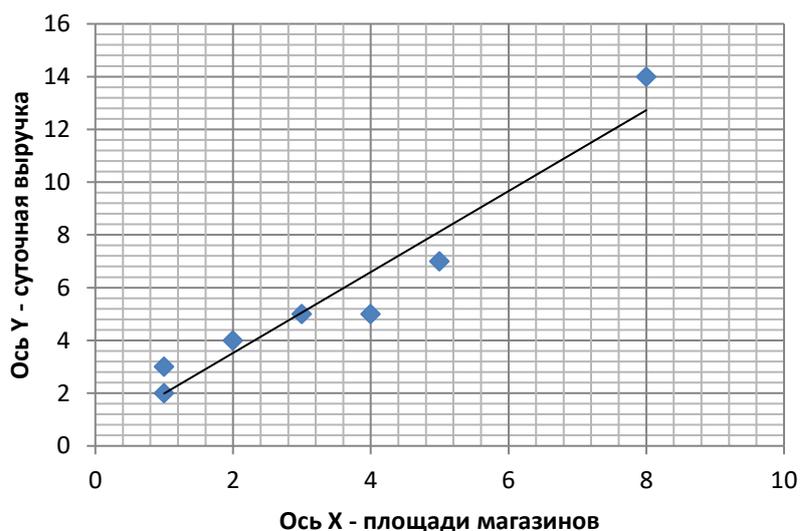


Рисунок 1.1 - Зависимость выручки магазина от торговой площади

Вопрос 4. Модель прогнозирования банкротства предприятия в форме линейной регрессии.

Одним из первых подобные модели начал строить Э. Альтман (г. р. 1941) – американский ученый. Модель строилась для предприятий, чьи акции *не торгуются* на фондовом рынке:

$$Z = 0,72 \cdot X_1 + 0,85 \cdot X_2 + 3,11 \cdot X_3 + 0,42 \cdot X_4 + 1,00 \cdot X_5, \quad (1.5)$$

где Z - искусственная результирующая переменная;

X_1 - X_5 - относительные финансовые показатели предприятия за прошлый год, таблица 1.2.

Результат прогнозирования определяется по шкале на рисунке 1.2. Точность модели – 91 % при прогнозировании банкротства предприятия на предстоящий год.

Красная зона высокого финансового риска	Серая зона неопределенности, риск средний	Зелёная зона финансовой устойчивости, риск банкротства в предстоящий год малый
	1,23	2,9 → Z

Рисунок 1.2 - Шкала прогноза финансовой устойчивости по значению Z

Таблица 1.2 – Факторы модели Альтмана – традиционные относительные финансовые показатели предприятия за прошедший год

Фактор	Формула для расчёта	
	содержательная	бухгалтерская*
X1	(Оборотные активы – Краткосрочные обязательства) / Активы	(1200-1500) / 1600
X2	Чистая прибыль / Активы	2400 / 1600
X3	Прибыль до налогообложения / Активы	2300 / 1600
X4	Собственный капитал / Обязательства	1300 / (1400+1500)
X5	Выручка / Активы	2110 / 1600

* Числа в 3-й графе – это номера строк, 1xxx – в бухгалтерском балансе, 2xxx – в отчёте о финансовых результатах.

Студентам важно уяснить, что модели Альтмана построены на основе дискриминантного анализа. Дискриминация состоит в разделении объектов по зонам на три класса, зоны: красная, серая, зелёная. Особенность таких моделей – отсутствие естественной результирующей переменной, например, такой, как прибыль. Возникает необходимость в создании такой искусственной переменной, как Z.

Вопрос 5. Методология и этапы эконометрического моделирования.

Это последний вопрос темы 1 и, наверное, самый важный (если следовать афоризму "Нет ничего практичнее, чем хорошая теория" - автор точно неизвестен, возможно, это И. Кант).

Студентам следует внимательно отнестись к этому вопросу. Мы рекомендуем учебник [8], а также [3, с. 15-16].

Методические материалы по теме 1

Учебник [1, Кремер-Путко, с. 3-24], учебник [2, Настин], ФОС в ЭИОС. В ходе работы по теме студенту следует использовать лекционный материал; материалы, полученные в ходе практического занятия, материалы в разделе дисциплины в ЭИОС КГТУ, а главное – рекомендованные здесь источники.

Ссылки на рекомендуемые источники по теме 1:

[1, Кремер-Путко, с. 3-24] – это базовый учебник, которого мы будем придерживаться, в том числе в терминологии, обозначениях.

[2, Настин, с. 5-25, 122-123] – это основной учебник.

[3, Настин, с.1-16] – это учебно-методическое пособие 2023 г.

Тема 2. Парный регрессионный анализ

Форма занятий – лекции.

Вопросы для изучения:

Вопрос 1. Метод наименьших квадратов.

Вопрос 2. Связь коэффициентов регрессии и корреляции.

Вопрос 3. Пять предпосылок для парного регрессионного анализа.

Вопрос 4. Качество оценок для параметров σ^2 , β_0 и β_1 .

Вопрос 5. Интервальные оценки функции регрессии $M_x(Y)$.

Вопрос 6. Интервальные оценки параметров регрессии.

Вопрос 7. Оценка значимости уравнения парной регрессии.

Методические указания по изучению темы 2

Цель изучения темы – освоить методы, применяемые в парном регрессионном анализе, и предпосылки, при которых применение этих методов правомерно. В процессе формулирования постановок задач и приёмов их моделирования рекомендуем придерживаться напутствия А. Эйнштейна:

"Всё должно быть изложено так просто, как только возможно, но не проще".

Вопрос 1. Метод наименьших квадратов.

Регрессионный анализ занимает центральное место в эконометрике. Напомним: регрессией называют функцию – зависимость математического ожидания СВ Y от значений X :

регрессия есть условное математическое ожидание СВ Y :

$$M_x(Y) = \varphi(x) \text{ или } M(Y|X) = \varphi(x). \quad (1.6)$$

На практике $\varphi(x)$ точно неизвестна, можно говорить лишь о её оценке. Исходными данными для этого служат результаты системного анализа проблемы и n пар значений x_i и y_i , $i \in 1:n$. На практике очень часто используется самая простая – линейная парная регрессия: $\hat{y}_x = b_0 + b_1 x$. Неизвестные параметры b_0 и b_1 определяются МНК-методом.

Суть метода наименьших квадратов состоит в отыскании таких значений параметров регрессии b_0 , b_1 , которые доставляют минимум функции $S(b_0, b_1)$ – сумме квадратов отклонений наблюдаемых значений y_i от теоретических – регрессионных – значений \hat{y}_i :

$$S(b_0, b_1) = \sum (\hat{y}_i - y_i)^2 = \sum (b_0 + b_1 x_i - y_i)^2 = \sum e_i^2 \rightarrow \min. \quad (1.7)$$

На основе МНК-метода получаем формулы:

$$b_1 = \frac{\overline{x \cdot y} - \bar{x} \cdot \bar{y}}{\overline{x^2} - \bar{x}^2} = \frac{\text{Cov}(X, Y)}{s_x^2}, \quad b_0 = \bar{y} - b_1 \bar{x}, \quad (1.8)$$

где $\overline{x \cdot y}$ - среднее произведения x на y ;

$\bar{x} \cdot \bar{y}$ - произведение средних x и y ;

$\overline{x^2}$ - среднее квадрата переменной x ;

\bar{x}^2 - квадрат среднего x ;

$\text{Cov}(X, Y)$ - *ковариация переменных x и y ;

s_x^2 - *дисперсия переменной x .

*см. приложение Д.

Вопрос 2. Связь коэффициентов регрессии и корреляции.

Рекомендуем студентам внимательно отнестись к содержанию вопроса, а также к коэффициенту корреляции, который является излюбленным инструментом анализа взаимосвязей переменных. В то же время, он имеет свои ограничения на применение. Подробнее см. [2, Настин, с. 33-35].

Нетрудно вывести формулу для r_{xy} – коэффициента парной корреляции и одновременно коэффициент регрессии:

$$r_{xy} = \frac{b_1 \cdot s_x}{s_y}. \quad (1.9)$$

Другие формулы для расчета r_{xy} (см. Приложение Д):

$$r_{xy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y}) / n}{s_x s_y} = \frac{\text{Cov}(x, y)}{s_x s_y}. \quad (1.10)$$

На рисунке 1.3 представлены графические образы различных статистических зависимостей и соответствующие им примерные значения коэффициентов корреляции. Следует иметь в виду, что:

коэффициент парной корреляции r_{XY} является оценкой генерального коэффициента корреляции ρ_{XY} лишь в случае двумерного *нормального закона распределения* векторной случайной величины (X, Y) . В других случаях он не является строгой мерой взаимосвязи переменных.

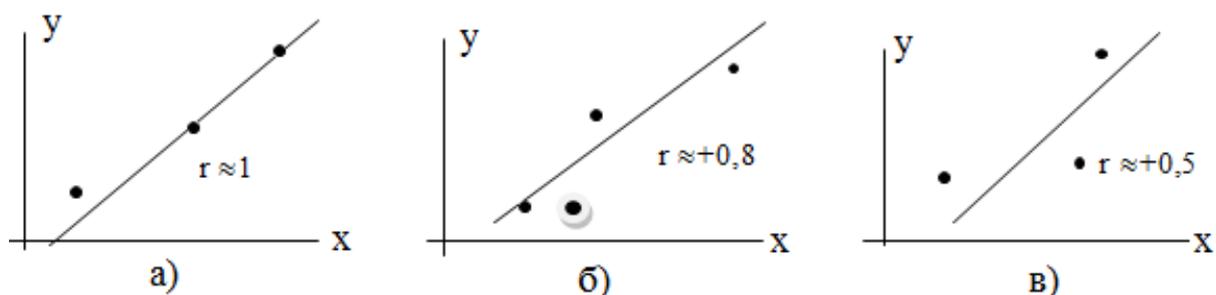


Рисунок 1.3 - Поля корреляции, линии аппроксимации и значения коэффициента корреляции r_{xy}

Вопрос 3. Пять предпосылок для парного регрессионного анализа.

Студентам следует обратить внимание на два термина: уравнение регрессии – мы его уже широко использовали - и стохастическая регрессионная модель (см. ниже), их следует различать. Подробнее см. в работах [1, с. 61] и [2, с. 35-36].

Линейная парная стохастическая регрессионная модель имеет вид:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \text{ для } \forall i \in 1:n. \quad (1.11)$$

Предпосылка 1. В модели (1.11) отклонение ε_i (и результирующая переменная y_i) есть величина случайная, а фактор x_i – величина неслучайная.

Предпосылка 2. Математическое ожидание отклонения ε_i равно нулю - $M(\varepsilon_i) = 0$; ε_i – это центрированная СВ, уравнение регрессии примет вид:

$$M(y_i) = \beta_0 + \beta_1 x_i \text{ для } \forall i \in 1:n. \quad (1.12)$$

Предпосылка 3. Отклонения ε_i гомоскедастичны (синоним - равноизменчивы) – их дисперсии одинаковы для разных значений переменной X :

$$D(\varepsilon_i) = \sigma^2 = D(y_i) = \text{const для } \forall i \in 1:n. \quad (1.13)$$

При невыполнении этого условия мы имеем дело с гетероскедастичностью (разноизменчивостью).

Предпосылка 4. Отклонения ε_i и ε_j (и переменные y_i и y_j) некоррелированы – коэффициенты корреляции между ними должны равняться нулю:

$$\rho_{\varepsilon_i, \varepsilon_j} = M(\varepsilon_i \cdot \varepsilon_j) = 0 \text{ для } \forall i, j \in 1:n \text{ и } i \neq j. \quad (1.14)$$

Предпосылка 5. Отклонение ε_i (и переменная y_i) есть нормально распределенная (распределена по нормальному закону) СВ для $\forall i \in 1:n$.

Модель (1.11), для которой выполняются все пять предпосылок, называется *нормальной классической линейной регрессионной моделью* (НКЛРМ).

Для получения уравнения регрессии достаточно предпосылок 1-4, при их наличии будем иметь *классическую линейную регрессионную модель* (КЛРМ). Предпосылка 5 необходима для оценки точности уравнения и его параметров.

Вопрос 4. Качество оценок для параметров σ^2 , β_0 и β_1 .

Оценку s^2 для остаточной дисперсии (дисперсии отклонений) σ^2 вычисляют по формуле:

$$s^2 = \frac{\sum (\hat{y}_i - y_i)^2}{n-2} = \frac{\sum e_i^2}{n-2}, \quad (1.15)$$

где n - длина выборки (в нашей задаче-1 $n=7$);

$e_i = (\hat{y}_i - y_i)$ - отклонение (остаток, ошибка);

$n - 2$ - число степеней свободы.

Оценки b_0 , b_1 для β_0 , β_1 получают методом наименьших квадратов. Вопрос: являются ли они наилучшими? Ответ "да" дает теорема.

Теорема Гаусса-Маркова. Если регрессионная модель (1.11) удовлетворяет предпосылкам 1-4, то оценки уравнения b_0 , b_1 среди несмещённых *линейных* (относительно y_i) оценок имеют наименьшую дисперсию, то есть, они эффективные.

Вопрос 5. Доверительный интервал для регрессии $M_x(Y)$ и индивидуального прогноза.

1) Переменная \hat{y}_x - случайная величина, поскольку параметры регрессии являются случайными величинами, ведь они находятся по выборке. Для построения доверительного интервала нужно найти её дисперсию – оценку $s_{\hat{y}}^2(x)$ по формуле:

$$s_{\hat{y}}^2(x) = s^2 \left(\frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum (x_i - \bar{x})^2} \right). \quad (1.16)$$

Обоснование этого выражения см. в работе [1, с. 64-66].

Теперь можно записать систему неравенств для доверительного интервала регрессии $M_x(Y)$:

$$\hat{y}_x - t_{1-\alpha, k} s_{\hat{y}} \leq M_x(Y) \leq \hat{y}_x + t_{1-\alpha, k} s_{\hat{y}}, \quad (1.17)$$

где $t_{1-\alpha, k}$ – критическое значение статистики Стьюдента для доверительной вероятности $\gamma=1-\alpha$ (α – уровень значимости, обычно $\alpha=0,05$).

2) **Доверительный интервал** для индивидуального значения прогноза. При определении такого интервала для значений зависимой переменной $y(x_0) = y_0$ нужно учесть ее рассеяние вокруг линии регрессии, т. е. к суммарной дисперсии $s_{\hat{y}}^2$ нужно добавить еще и остаточную дисперсию s^2 . Тогда и получим оценку дисперсии СВ y_0 :

$$s_{y_0}^2 = s^2 \left(1 + \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum (x_i - \bar{x})^2} \right). \quad (1.18)$$

Искомый доверительный интервал для СВ – индивидуального значения y_0 :

$$\hat{y}_0 - t_{1-\alpha, n-2} \cdot s_{\hat{y}_0} \leq y_0 \leq \hat{y}_0 + t_{1-\alpha, n-2} \cdot s_{\hat{y}_0}. \quad (1.19)$$

Вопрос 6. Доверительные интервалы для параметров регрессии.

1) *Доверительный интервал для коэффициента регрессии.* Иногда анализ требует интервального оценивания коэффициента β_1 . Доказано, что при выполнении предпосылки-5 статистика $t = (b_1 - \beta_1) / \sigma_{b_1}$ имеет стандартный НЗР, а статистика:

I_1 и I_2 – доверительные интервалы для $M(Y|\bar{x})$ и $M(Y| x_0)$,

I_3 – доверительный интервал для индивидуального значения $y_0(x_0)$,

$$t = \frac{b_1 - \beta_1}{s} \sqrt{\sum (x_i - \bar{x})^2} \quad (1.20)$$

имеет t-распределение Стьюдента с $k = n-2$ степенями свободы. Поэтому доверительный интервал для β_1 с доверительной вероятностью $\gamma=1-\alpha$ имеет вид:

$$b_1 - \frac{t_{1-\alpha, n-2} \cdot s}{\sqrt{\sum (x_i - \bar{x})^2}} \leq \beta_1 \leq b_1 + \frac{t_{1-\alpha, n-2} \cdot s}{\sqrt{\sum (x_i - \bar{x})^2}}. \quad (1.21)$$

Примечание: в формулах (1.18) и (1.19) s – это СКО для b_1 .

2) *Доверительный интервал для дисперсии отклонения (ошибки) σ^2* (напомним: s^2 и σ^2 - остаточные дисперсии – выборочная и генеральная). Статистика ns^2/σ^2 имеет распределение χ^2 с $k = n-2$ степенями свободы. Поэтому интервальная оценка отклонения σ^2 с уровнем значимости α имеет вид:

$$\frac{ns^2}{\chi_{\frac{\alpha}{2}, n-2}^2} \leq \sigma^2 \leq \frac{ns^2}{\chi_{1-\frac{\alpha}{2}, n-2}^2}. \quad (1.22)$$

Вопрос 7. Оценка значимости уравнения парной регрессии.

Подробно этот вопрос изложен в работах [1, с. 102-104], [2, с. 42-44].

1) *Оценка парной регрессии по критерию Стьюдента.* Здесь значимость оценивается по одному коэффициенту b_1 . Нулевая гипотеза H_0 : b_1 незначимо отличается от 0. Если она верна, то $b_1=0$, и линия регрессии параллельна оси ОХ, а значит, Y не зависит от X . Обратная гипотеза H_1 : b_1 значимо отличается от 0; если это так, то регрессия значима. Критерий на основе статистики Стьюдента и гипотезы H_1 :

$$H_1: \beta_1=0 \text{ отвергается, если } t = \frac{b_1 - 0}{s} \sqrt{\sum (x_i - \bar{x})^2} > t_{1-\alpha, n-2}.$$

2) *Оценка регрессии по критерию Фишера-Снедекора.* Здесь используется выражение из дисперсионного анализа (ниже добавлены англоязычные обозначения).

$$Q = Q_R + Q_e, \quad (1.23)$$

где $Q = \sum (\bar{y} - y_i)^2 = TSS$ - общая сумма квадратов отклонений зависимой переменной от её средней; Total Sum of Squares;

$Q_R = \sum (\hat{y}_i - \bar{y})^2 = RSS$ - сумма квадратов отклонений, объяснённая регрессией; Regression Sum of Squares;

$Q_e = \sum (\hat{y}_i - y_i)^2 = \sum e_i^2 = ESS$ - остаточная сумма квадратов отклонений, обусловленная влиянием неучтенных факторов; Error Sum of Squares.

Таблица 1.3 – Выражения из дисперсионного анализа

Вид дисперсии	Сумма квадратов	Число степеней свободы	Средние квадраты* (дисперсии)
Регрессионная	$Q_R = \sum (\hat{y}_i - \bar{y})^2$	m-1	$s_R^2 = Q_R / (m-1)$
Остаточная	$Q_e = \sum (y_i - \hat{y}_i)^2$	n-m	$s^2 = Q_e / (n-m)$
Общая	$Q = \sum (y_i - \bar{y})^2$	n-1	$s_y^2 = Q / (n-1)$

*Графа 4: средние квадраты – это несмещенные оценки соответствующих дисперсий.

Гипотеза H_0 о равенстве нулю всех коэффициентов модели ($\beta_1 = \beta_2 = \dots = \beta_p = 0$) отвергается (регрессия значима), если фактическое значение статистики Фишера-Снедекора больше ее табличного значения:

$$F = \frac{Q_R / (m-1)}{Q_e / (n-m)} = \frac{s_R^2}{s^2} > F_{\alpha, k_1, k_2}, \quad (1.24)$$

где F_{α, k_1, k_2} - табличное (критическое) значение F-критерия Фишера-Снедекора для уровня значимости α , степеней свободы $k_1 = m-1$, $k_2 = n-m$.

Смысл статистики F прозрачен: её значения показывают, во сколько раз объясненная регрессией дисперсия s_R^2 СВ \hat{y} больше остаточной (необъясненной) дисперсии s^2 СВ e . Чем больше это отношение, тем более значимо уравнение регрессии.

Для парной регрессии $m=2$, поэтому критерий (1.24) формулируется проще: уравнение значимо, если выполняется неравенство:

$$F = \frac{Q_R}{Q_e / (n-2)} = \frac{s_R^2}{s^2} > F_{\alpha, 1, n-2}. \quad (1.25)$$

3) Оценка регрессии по коэффициенту детерминации R^2 . Нетрудно вывести формулу:

$$R^2 = Q_R / Q = 1 - Q_e / Q. \quad (1.26)$$

Очевидно, что $0 \leq Q_R \leq Q$. Отсюда важное свойство коэффициента – наличие диапазона значений: $0 \leq R^2 \leq 1$.

Для парной регрессии коэффициент детерминации равен квадрату коэффициента корреляции: $R^2 = r_{xy}^2$.

4) Оценка регрессии по скорректированному коэффициенту детерминации:

$$\hat{R}^2 = 1 - \frac{n-1}{n-p-1} (1 - R^2). \quad (1.27)$$

Диапазон изменения \hat{R}^2 тот же, что и у R^2 (от 0 до 1). Различие в том, что \hat{R}^2 обратно зависит от числа факторных переменных: чем их больше, тем меньше \hat{R}^2 при прочих равных условиях вводится штраф за использование

каждой переменной. При отборе значимых факторов коэффициент \hat{R}^2 сначала растёт, а потом уменьшается при включении в уравнение малозначимых переменных.

Методические материалы по теме 2

Конспект лекций, учебник [1, Кремер-Путко], учебник [2, Настин], ФОС в ЭИОС.

Ссылки на рекомендуемые источники по теме 2:

[1, Кремер-Путко, с. 50-77] – это базовый учебник, которого мы будем придерживаться, в том числе в терминологии, обозначениях.

[2, Настин, с. 29-42] – это основной учебник.

Тема 3. Множественный регрессионный анализ

Форма занятий – лекция.

Вопросы для изучения:

Вопрос 1. Классическая нормальная линейная модель множественной регрессии.

Вопрос 2. Оценка параметров классической нормальной линейной модели множественной регрессии методом наименьших квадратов.

Вопрос 3. Оценки ковариационной матрицы для случайного вектора b и дисперсии отклонений.

Вопрос 4. Доверительные интервалы для параметров регрессии, условного математического ожидания $M_x(Y)$ и для индивидуального прогноза.

Вопрос 5. Оценка значимости множественной регрессии по трём критериям.

Вопрос 6. Мультиколлинеарность и отбор значимых факторов

Методические указания по изучению темы 3

Цель изучения темы – освоить методы, применяемые в множественном регрессионном анализе, а также предпосылки, при которых применение этих методов правомерно. Особенностью темы 3 является применяемый здесь математический аппарат – матричная алгебра, см. [2, с. 119-121]. Погружаясь в теорию эконометрики, полезно помнить о предостережении, высказанном А. Эйнштейном:

"Существует поразительная возможность овладеть предметом математически, не поняв существа дела".

Вопрос 1. Классическая нормальная линейная модель множественной регрессии.

Сложное явление описывается многими объясняющими переменными, примерно 2-10. Отсюда и название – множественная регрессия. Стохастическая модель множественной линейной регрессии (ММЛР) имеет вид:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \varepsilon_i, \forall i \in 1:n; \quad (1.28)$$

где ε_i – случайная величина, отклонение, ошибка.

Соответствующее уравнение множественной регрессии примет вид:

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + \dots + b_p x_p. \quad (1.29)$$

Пусть $\beta = (\beta_0 \beta_1 \dots \beta_p)'$ и $\varepsilon = (\varepsilon_1 \dots \varepsilon_n)'$ - матрицы-столбцы параметров регрессии и отклонений, размерность $(p+1) \times 1$ и $n \times 1$ соответственно. Теперь выражение (1.28) можно записать в матричной форме:

$$Y = X\beta + \varepsilon. \quad (1.30)$$

Оценкой этой модели будет выборочная стохастическая модель:

$$Y = Xb + e, \quad (1.31)$$

где Y – результирующая переменная;

$$b = (b_0 \ b_1 \dots \ b_p)';$$

$$e = (e_1 \dots \ e_n)';$$

К сформулированным ранее предпосылкам 1-5 для множественного анализа нужно добавить *предпосылку-б*: векторы значений объясняющих переменных (столбцы матрицы-плана X) должны быть линейно независимыми, другими словами, ранг r матрицы X должен быть равен $p+1$, сокращённая запись $r(X)=p+1$.

Сформулируем кратко в матричной форме все 6 предпосылок.

- 1) В модели (1.40) ε - случайный вектор, X - неслучайная матрица плана.
- 2) $M(\varepsilon) = 0_n$, где 0_n , -вектор-столбец, состоящий из n нулей.
- 3) $\Sigma_\varepsilon = M(\varepsilon \cdot \varepsilon') = \sigma^2 E_n$, где E_n - единичная матрица $n \times n$: на главной диагонали находятся одинаковые дисперсии σ^2 - гомоскедастичность.
- 4) $\Sigma_\varepsilon = M(\varepsilon \cdot \varepsilon') = \sigma^2 E_n$: над главной диагональю находятся нулевые ковариации – отклонения ε_i и ε_j , некоррелированные для всех $i, j \in 1:n, i \neq j$.
- 5) ε - случайный вектор с n -мерным НЗР $N_n(0_n, \sigma^2 E_n)$.
- 6) Ранг матрицы X : $\text{Rang}(X) = p+1$, причем $p+1 \ll n$ (знак \ll означает много меньше): число объясняющих переменных должно быть много меньше, чем число наблюдений n .

Итак, если модель (1.40) удовлетворяет предпосылкам 1-6 регрессионного анализа, то такая модель называется *классической нормальной линейной моделью множественной регрессии (КНЛММР)*.

Если из шести предпосылок не выполняется одна - пятая - о НЗР распределения ε , то модель называется классической линейной моделью множественной регрессии (КЛММР).

В пособии [3, с. 27] дано определение каждого из пяти терминов аббревиатуры КНЛММР.

Вопрос 2. Оценка параметров классической нормальной линейной модели множественной регрессии методом наименьших квадратов.

Условие минимизации суммы квадратов отклонений для множественной регрессии в 4-х формах записи:

$$S = \sum (\hat{y}_{xi} - y_i)^2 = \sum e_i^2 = e'e = (Y-Xb)'(Y-Xb) \rightarrow \min. \quad (1.32)$$

Составим систему линейных уравнений, приравняв нулю частные производные от S по b_i :

$$\nabla_b S = 0_n, \quad (1.33)$$

где ∇_b – набла, векторный дифференциальный оператор;

S – уже знакомая нам функция – сумма квадратов отклонения наблюдаемых значений от регрессионных;

0_n - 0-вектор.

Опустим все промежуточные преобразования (подробности см. [1, Кремер-Путко, с. 84]) и получим результат: систему нормальных линейных уравнений в матричной форме относительно искомого вектора b :

$$X'Xb = X'Y. \quad (1.34)$$

Чтобы разрешить систему (1.43) уравнений, нужно обе части уравнения умножить *слева* (умножение матриц некоммукативно) на матрицу, обратную к матрице $X'X$, т. е. на $(X'X)^{-1}$. В результате получим систему уравнений, разрешенную относительно искомого вектора b :

$$b = (X'X)^{-1}X'Y. \quad (1.35)$$

Теперь можно *предпосылку-б* переформулировать так:

Матрица $A=X'X$ должна быть неособенной (ее определитель не должен равняться нулю: $|X'X| \neq 0$).

И третья эквивалентная формулировка: матрица $X'X$ должна иметь обратную матрицу.

Приведём важную для нас формулировку теоремы Гаусса-Маркова для множественной регрессии.

Теорема Гаусса-Маркова. При выполнении предпосылок 1-4 и 6 множественного регрессионного анализа оценка вектора $b = (X'X)^{-1}X'Y$, полученного по МНК-методу, является эффективной, т. е. она несмещённая и обладает наименьшей дисперсией в классе *линейных* оценок.

Вопрос 3. Оценки ковариационной матрицы для случайного вектора b и дисперсии отклонений e .

Здесь мы используем понятия: "случайный вектор", многомерный закон распределения, ковариационная матрица как аналог дисперсии, см. [2, с. 115], подробнее – в работе [4, с. 187-190].

1) *Оценка ковариационной матрицы для случайного вектора b .* В многомерном статистическом анализе вместо дисперсии для одной случайной величины рассматривается ковариационная матрица для случайного вектора как набора случайных величин. Для случайного вектора b ковариационная матрица Σ_b (не путать с оператором суммирования) имеет размерность $(p+1) \times (p+1)$ и в общем случае она состоит из элементов:

$\sigma_{ij} = M[(b_i - M(b_i)) \cdot (b_j - M(b_j))]$, $i \neq j$ - ковариация двух СВ - коэффициентов b_i и b_j ;

$\sigma_{ii} = M[(b_i - M(b_i)) \cdot (b_i - M(b_i))] = \sigma_{b_i}^2$ - дисперсия СВ – коэффициента b_i ; находится на главной диагонали.

Путем преобразований [1, Кремер-Путко, с. 92] получаем формулу для ковариационной матрицы случайного вектора b :

$$\Sigma_b = \sigma^2 (X'X)^{-1}, \quad (1.36)$$

где $\sigma^2 = D(\varepsilon_i)$ – дисперсия - константа.

2) *Оценки дисперсии отклонений.* Для практического применения (1.46) нужно вычислить точечную оценку s^2 – остаточной дисперсии [1, Кремер-Путко, с. 95]:

$$s^2 = \frac{e'e}{n-p-1} = \frac{\sum e_i^2}{n-p-1}. \quad (1.37)$$

Интервальная оценка - доверительный интервал - для остаточной дисперсии множественной регрессии с соответствующим числом степеней свободы для χ^2 :

$$\frac{ns^2}{\chi_{\alpha/2, n-p-1}^2} \leq \sigma^2 \leq \frac{ns^2}{\chi_{1-\alpha/2, n-p-1}^2}. \quad (1.38)$$

Вопрос 4. Доверительные интервалы для параметров регрессии, условного математического ожидания $M_x(Y)$ и для индивидуального прогноза.

1) *Доверительный интервал для параметров β_j .* Начнём со значимости параметров b_j . Она проверяется с помощью нулевой гипотезы $\{H_0: \beta_j = 0\}$. Гипотеза отвергается с уровнем значимости α , если выполняется неравенство:

$$t = |b_j - \beta_j| / S_{b_j} > t_{1-\alpha, n-p-1} \cdot S_{b_j}.$$

Доверительный интервал для β_j вытекает из неравенства:

$$b_j - t_{1-\alpha, n-p-1} \cdot S_{b_j} \leq \beta_j \leq b_j + t_{1-\alpha, n-p-1} \cdot S_{b_j}. \quad (1.39)$$

2) *Доверительный интервал для регрессии $M_x(Y)$.* Искомый интервал определяется аналогично доверительному интервалу для парной регрессии и имеет вид:

$$\hat{y} - t_{1-\alpha, k} s_{\hat{y}} \leq M_x(Y) \leq \hat{y} + t_{1-\alpha, k} s_{\hat{y}}, \quad (1.40)$$

где \hat{y} - значение функции регрессии;

$s_{\hat{y}}$ - стандартная ошибка регрессии \hat{y} :

$$s_{\hat{y}} = s \sqrt{X_0' (X'X)^{-1} X_0}; \quad (1.41)$$

где $X_0' = (1 \ x_{10} \ x_{20} \ \dots \ x_{p0})$ - вектор заданных – проектных - значений объясняющих переменных, например, в задаче-1 $x_{10}=11$ (площадь магазина-8).

3) *Доверительный интервал для индивидуального значения прогноза y_0* . Используем понятие индивидуального значения прогноза. Для индивидуальных значений результирующей переменной множественной регрессии:

$$\hat{y}_0 - t_{1-\alpha, n-p-1} s_{\hat{y}_0} \leq y_0 \leq \hat{y}_0 + t_{1-\alpha, n-p-1} s_{\hat{y}_0}, \quad (1.42)$$

$$\text{где } s_{\hat{y}_0} = s \sqrt{1 + X_0' (X'X)^{-1} X_0}. \quad (1.43)$$

Вопрос 5. Оценка значимости множественной регрессии по трём критериям. Критерий Стьюдента здесь неприменим.

1) *Критерий Фишера-Снедекора*. Гипотеза H_0 о равенстве нулю всех коэффициентов ($\beta_1 = \beta_2 = \dots = \beta_p = 0$) *отвергается*, если фактическое значение статистики Фишера-Снедекора F больше ее табличного значения:

$$F = \frac{Q_R (n - p - 1)}{Q_e p} > F_{\alpha, p, n-p-1}. \quad (1.44)$$

2) *Критериальный показатель – коэффициент детерминации R^2* , рассчитывается по формулам:

$$R^2 = Q_R / Q = 1 - Q_e / Q.$$

В матричной форме коэффициент R^2 может быть рассчитан также по формуле:

$$R^2 = \frac{\mathbf{b}' \mathbf{X}' \mathbf{Y} - n \bar{y}^2}{\mathbf{Y}' \mathbf{Y} - n \bar{y}^2}, \quad (1.45)$$

где $\mathbf{e} = \mathbf{Y} - \mathbf{Xb}$,

$\bar{\mathbf{Y}} = (\bar{y}, \bar{y}, \dots, \bar{y})'$ – n -мерные векторы;

$\dot{\mathbf{y}} = (\mathbf{Y} - \bar{\mathbf{Y}})$ - центрированный n -мерный вектор;

$\mathbf{e}'\mathbf{e} = \sum e_i^2 = \sum (y_i - \hat{y}_i)^2$ – сумма квадратов отклонений (ошибок);

$\dot{\mathbf{y}}' \cdot \dot{\mathbf{y}} = \sum (y_i - \bar{y})^2$ – скалярное произведение центрированных векторов.

Коэффициент детерминации принимает значения в диапазоне от 0 до +1, а значимость уравнения регрессии тем больше, чем R^2 больше.

3) *Критериальный показатель - скорректированный коэффициент детерминации \hat{R}^2* :

$$\hat{R}^2 = 1 - \frac{n-1}{n-p-1}(1-R^2). \quad (1.46)$$

Из этого выражения можно видеть: чем больше объясняющих переменных p , тем меньше \hat{R}^2 в сравнении с R^2 при прочих равных условиях.

Вопрос 6. Мультиколлинеарность и отбор значимых факторов.

Вопрос является, наверное, самым важным в регрессионном анализе. Снижение мультиколлинеарности и отбор факторов – залог успеха моделирования. Понятие мультиколлинеарности в регрессионном анализе относят к векторам-столбцам матрицы факторных переменных X_1, X_2, \dots, X_p . Существует функциональная (полная) коллинеарность и статистическая (неполная).

Функциональная коллинеарность. Векторы коллинеарны, если они лежат на параллельных прямых, рисунок 1.4а. Здесь вектора линейно зависят между собой, например, $X_2 = k \cdot X_1$. Коэффициент корреляции между X_1 и X_2 равен 1. Если это имеет место в матрице X , то оно будет иметь место и в матрице $A = (X'X)$ ("'" - это оператор транспонирования). В таком случае получится $|A| = 0$. Такую матрицу называют особенной, она не имеет обратной матрицы. Пример:

$$A = \begin{bmatrix} 1 & 2 & 4 \\ 2 & 4 & 8 \end{bmatrix} \times \begin{bmatrix} 1 & 2 \\ 2 & 4 \\ 4 & 8 \end{bmatrix} = \begin{bmatrix} 21 & 42 \\ 42 & 84 \end{bmatrix}.$$

В матрице X - слева и в матрице $A = (X'X)$ - справа - столбцы линейно зависимы, определитель $|A| = 0$.

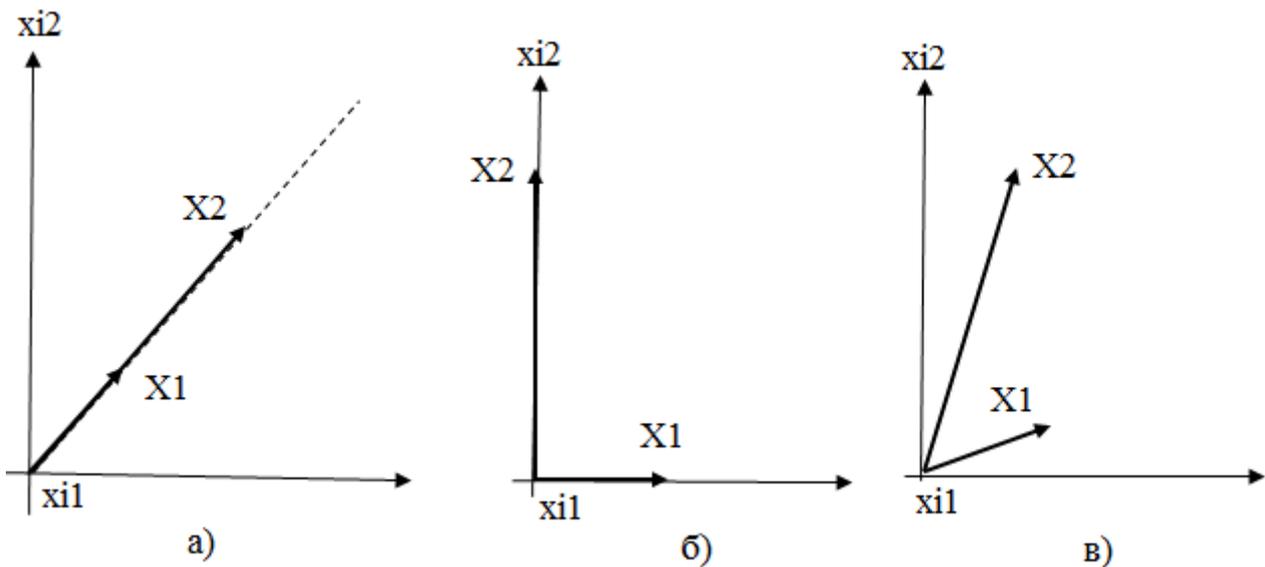


Рисунок 1.4 - Три отношения между парой векторов: а) полная коллинеарность; б) нет коллинеарности (вектора ортогональны); в) неполная коллинеарность (статистическая)

Статистическая мультиколлинеарность, рисунок 1.4в. Здесь между двумя или большим числом факторов имеет место существенная корреляция. Определитель матрицы A близок к нулю. Как следствие, параметры регрессии

крайне чувствительны к изменениям исходных данных – неустойчивы: а) увеличивается их дисперсия, б) возникают противоречия: модель значима, а её коэффициенты незначимы, в) затруднена интерпретация модели.

Два метода выявления мультиколлинеарности и отбора значимых факторов.

1) *Метод анализа матрицы $R_{p,p}$ парных коэффициентов корреляции между результирующей и факторными переменными.* Отбор значимых факторов проводится путём *исключения* незначимых факторов из первоначального большого набора факторов.

1-й шаг. Экспертно формируем большой набор факторов.

2-й шаг. Вычисляем парные коэффициенты корреляции - элементы матрицы R_{pp} . Например, для $p = 3$ имеем матрицу:

$$R_{p=3} = \begin{pmatrix} 1 & r_{12} & r_{13} \\ r_{21} & 1 & r_{23} \\ r_{31} & r_{32} & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0,5 & 0,6 \\ 0,5 & 1 & 0,8 \\ 0,6 & 0,8 & 1 \end{pmatrix}. \quad (1.47)$$

Дополнительно вычисляются коэффициенты корреляции между факторами и результирующей переменной y .

3-й шаг. Ищем наибольший по модулю коэффициент в матрице R_{pp} . Пусть это будет $r_{23} = 0,8$. Фактор x_2 либо x_3 можно исключить из списка претендентов. Пусть корреляция между результатом Y и факторами равна: $r_{y2}=0,4$; $r_{y3}=0,7$. Фактор x_3 включаем в итоговый список, а фактор x_2 исключаем.

4-й шаг. Проверяем: либо конец решения, набор факторов сформирован; либо продолжаем решение – переход ко 2-му шагу. В качестве критерия останова можно взять такой: в новой матрице коэффициентов корреляции, полученной после исключения очередного фактора, больше нет коэффициентов r_{ij} со значением по модулю больше 0,4; конец решения.

2) *Метод "вращения факторов".* Здесь в качестве критерия уместно использовать скорректированный коэффициент детерминации \hat{R}^2 . Рассмотрим алгоритм этого метода по шагам.

1-й шаг. Экспертно формируем первоначальный набор из шести факторов: $X_a, X_b, X_c, X_d, X_e, X_f$.

2-й шаг. Строим шесть однофакторных линейных уравнений регрессий с каждым из факторов и вычисляем для каждой регрессии коэффициент детерминации \hat{R}^2 . Отложим эти шесть значений на вертикали для $k=1$. Фактор, у которого коэффициент \hat{R}^2 максимальный, мы включаем в регрессию. В нашем примере это фактор X_c .

3-й шаг. Строим пять двухфакторных уравнений регрессий с фактором X_c и с каждым из оставшихся пяти факторов и вычисляем для каждой регрессии

коэффициент детерминации \hat{R}^2 . Отложим эти пять значений на вертикали для $k=2$, X_e - это фактор, с которым в нашем примере коэффициент максимальный, его мы и включаем в регрессию.

4-й шаг. При работе с пятифакторными регрессиями (с X_a и затем с X_d) мы сталкиваемся с уменьшением коэффициента детерминации \hat{R} .

Вывод. Мы отобрали четыре наиболее значимых фактора, которые одновременно имеют малую степень мультиколлинеарности: X_c , X_e , X_f , X_a . Факторы X_b , X_d незначимы и существенно повышают мультиколлинеарность, мы не включаем их в целевую регрессию.

Конец алгоритма.

На рисунке 1.4б показан случай, когда два вектора независимы – ортогональны. Этот то, к чему мы должны стремиться при отборе факторов.

Вопрос 7. Функции и коэффициенты эластичности.

Функции эластичности представляют большой интерес при анализе экономического явления.

Функция эластичности $E_x(\hat{y})$ показывает, на сколько процентов изменяется \hat{y} при изменении переменной x на 1 %.

По определению частной функцией эластичности $E_{x_i}(\hat{y})$ регрессии $\hat{y}=f(x_1, x_2, \dots, x_p)$ называется функция:

$$E_{x_i}(\hat{y}) = \frac{\partial \ln \hat{y}}{\partial \ln x_i} = \frac{\partial \hat{y}}{\hat{y}} \cdot \frac{\partial x_i}{x_i} = \frac{\partial \hat{y}}{\partial x_i} \cdot \frac{x_i}{\hat{y}}. \quad (1.48)$$

Например, для модели $y = \beta_0 + \beta_1 x + \varepsilon$ функция эластичности имеет вид $E(x) = \beta_1 x / (\beta_0 + \beta_1 x)$.

Коэффициентом эластичности называется значение функции эластичности для одного значения x . Обычно используют среднее значение: $x = \bar{x}$ и получают "средний" коэффициент эластичности $E(\hat{y}|\bar{x})$.

Методические материалы по теме 3

Конспект лекций, учебник [1, Кремер-Путко], учебник [2, Настин], ФОС в ЭИОС.

Ссылки на рекомендуемые источники по теме 3:

[1, Кремер-Путко, с. 82-107] – базовый учебник, которого мы будем придерживаться, в том числе в терминологии, обозначениях.

[2, Настин, с. 48-55] – основной учебник.

Тема 4. Нелинейный регрессионный анализ

Форма занятий – лекция.

Вопросы для изучения:

Вопрос 1. Нелинейные модели регрессии, их классификация и линейаризация.

Вопрос 2. Простейшая нелинейная модель магазина.

Вопрос 3. Фиктивные переменные в линейных моделях.

Вопрос 4. Фиктивные переменные в нелинейных моделях.

Вопрос 5. Нелинейные производственные функции Кобба-Дугласа.

Вопрос 6. Пример функции Кобба-Дугласа в рыбной отрасли.

Методические указания по изучению темы 4

Цель изучения темы – освоение новых разделов классического регрессионного анализа. Другими словами, мы продолжим говорить об эконометрическом моделировании. При этом рекомендуем не поддаваться пессимизму, который звучит в высказывании А. Эйнштейна (ведь это было сказано в первой половине XX века, ещё не было информационных технологий):

"Законы математики, имеющие какое-либо отношение к реальному миру, ненадёжны; а надёжные математические законы не имеют отношения к реальному миру".

Вопрос 1. Нелинейные модели регрессии, их классификация и линейаризация.

В природе и экономике вряд ли найдутся чисто линейные зависимости – они все нелинейные. Нелинейные регрессионные модели разнообразны, поэтому нужна их классификация, таблица 1.4.

Таблица 1.4 - Классификация нелинейных регрессионных моделей

Н - Нелинейные		
Н1 - Нелинейные по переменным, но линейные по параметрам	Н2 - Нелинейные по параметрам, но внутренне они:	
	Н.2.1 - линейные	Н.2.2 - нелинейные

Опишем классы моделей.

Класс Н1: модели нелинейные относительно переменных, но линейные по параметрам, например:

- полином второй степени: $y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon$;
- гипербола: $y = \beta_0 + \beta_1 / x + \varepsilon$;
- логарифмическая функция: $y = \beta_0 + \beta_1 \ln x + \varepsilon$.

Модели этого класса легко приводятся к линейной форме с помощью замены переменных. Например, в полиноме $y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon$ заменим переменные: $x = x_1$, $x^2 = x_2$ и получим двухфакторную линейную регрессионную модель: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$. Далее применяем МНК-метод.

Вопрос 2. Простейшая нелинейная модель магазина.

Выше мы рассматривали линейную парную модель магазина: $y=b_0+b_1x+\varepsilon$, в которой параметр b_0 не имеет смысла. Более того, параметр b_1 из-за этого также теряет смысл (удельная выручка). Может возникнуть и третья проблема: реальная зависимость y (выручки) от x (площади) существенно нелинейная. Все это устраняется, если использовать нелинейную квадратичную модель: $y=b_1x+b_2x^2+\varepsilon$.

Что важно, здесь выполняется равенство: при $x = 0$ и $y = 0$. Модель относится к классу $H1$ – она нелинейная по переменным, но линейная по параметрам, поэтому легко линеаризуется.

Вопрос 3. Фиктивные переменные в линейных моделях.

С помощью фиктивных переменных путём их ввода в уравнение регрессии можно решить две проблемы: а) использовать качественные – атрибутивные – факторные переменные; б) наполнить уравнение сложной логикой. Это открывает новые возможности моделирования.

Фиктивной (булевой, бинарной) называется переменная, которая принимает два значения: 0 либо 1.

Пусть изучается зависимость зарплаты сотрудников от $X1$ – рабочего стажа и $X2$ – образования. Переменная $X2$ атрибутивная, она может иметь три нечисловых значения: среднее общее, среднее профессиональное, высшее. Для использования этой переменной в регрессии предлагается вместо неё ввести две фиктивных переменных (вместо k значений атрибутивной переменной вводится $(k-1)$ фиктивных переменных). Подробно этот вопрос раскрыт в [3].

Вопрос 4. Фиктивные переменные в нелинейных моделях.

В качестве примера рассмотрим упрощённую модель оценки стоимости квартиры [12, Хачатрян и др., с. 205].

Первый шаг: составляем нелинейную регрессионную модель в исходной форме:

$$Y=(e^{b_0} \cdot X1^{b_1} X2^{b_2} \cdot e^{K1 \cdot Z1} \cdot e^{K3 \cdot Z3} / e^{K2 \cdot Z2}) \mu, \quad (1.49)$$

где Y - рыночная цена квартиры на дату сделки, руб.

b_0 - коэффициент конъюнктуры на рынке жилья, определяется экспертно, например, в диапазоне от -1 до +1;

b_1 - b_2 и $K1$ - $K3$ – параметры уравнения регрессии;

$X1$ - площадь жилых помещений в квартире, кв. м;

$X2$ - площадь нежилых помещений в квартире, кв. м;

$Z1$ - $Z3$ - фиктивные переменные;

$Z1 = 1$, если тип жилого дома с улучшенной планировкой, иначе 0;

$Z2 = 1$, если этаж первый или последний, иначе 0;

$Z3 = 1$, если есть балкон или лоджия, иначе 0;

М - отклонение – мультипликативная случайная величина.

Второй шаг: собираем статистику о сделках на рынке жилья в городе;

Третий шаг: анализируем регрессионную модель (1.49):

- она мультипликативная, состоит из произведения степенных ($y=x^n$) и показательных ($y=a^x$) функций;
- она нелинейная по параметрам, но "внутренне линейная" - подкласс Н.2.1 - её можно путём логарифмирования перевести в линейную форму.

Четвёртый шаг: логарифмируем выражение (1.49) и получаем модель регрессии в аддитивной форме:

$$\ln Y = b_0 + b_1 \cdot \ln X_1 + b_2 \cdot \ln X_2 + K_1 \cdot Z_1 + K_3 \cdot Z_3 - K_2 \cdot Z_2 + \ln \mu. \quad (1.50)$$

Пятый шаг: заменяем функции от переменных на обычные переменные, у нас их 3: $\ln Y = \hat{u}$, $\ln X_i = V_i$ для $\forall i \in 1:2$, $\ln \mu = \varepsilon$. Получаем *линейное* пятифакторное регрессионное уравнение (слагаемое ε уже можно не писать):

$$\hat{u} = b_0 + b_1 \cdot V_1 + b_2 \cdot V_2 + K_1 \cdot Z_1 + K_3 \cdot Z_3 - K_2 \cdot Z_2. \quad (1.51)$$

Шестой шаг: готовим исходные данные, соответствующие преобразованиям на 5-м шаге, и решаем задачу МНК-методом.

Вопрос 5. Нелинейные производственные функции Кобба-Дугласа.

Эту функцию можно применять к отдельному предприятию, к отрасли, к национальной экономике. Можно увеличить число факторных переменных с двух до трёх и более. Рассмотрим один вариант такой функции.

Первый вариант. Функция имеет вид:

$$\hat{Y} = AK^\alpha L^\beta, \quad (1.52)$$

где Y - объем производства;

K - фактор-1, затраты капитала (прямые инвестиции), руб.;

L - фактор-2, затраты труда: среднесписочный состав за год; чел.*мес./год; фонд оплаты труда, руб./год;

A , α , β - параметры модели; α , β - коэффициенты эластичности Y по K и по L : $E_K(Y) = \alpha$, $E_L(Y) = \beta$.

Достоинство функции Кобба-Дугласа состоит в том, что она способна отражать технологическое соотношение между трудом и капиталом. С учетом случайных отклонений функция Кобба-Дугласа выглядит как мультипликативная степенная модель:

$$Y = AK^\alpha L^\beta \varepsilon. \quad (1.53)$$

Эта модель относится к подклассу Н.2.1 – нелинейная по параметрам, но внутренне линейная. Поэтому её можно свести к линейной относительно параметров путем логарифмирования:

$$\ln Y = \ln A + \alpha \ln K + \beta \ln L + \ln \varepsilon. \quad (1.54)$$

Вводим новые переменные: $Z = \ln Y$, $B = \ln A$, $X_1 = \ln K$, $X_2 = \ln L$, $\mu = \ln \varepsilon$, получаем линейную двухфакторную регрессионную модель:

$$Z = B + \alpha X_1 + \beta X_2 + \mu. \quad (1.55)$$

Вопрос 6. Пример функции Кобба-Дугласа в рыбной отрасли [11, Сергеев].

Для студентов пример интересен тем, что рассматривается модель, полученная на реальном статистическом материале. На основании статистики за 2006-2016 гг. МНК-методом была построена функция Кобба-Дугласа (рисунок 1.9), у которой: Y - объем добычи водных биологических ресурсов, I - объем инвестиций, P - размер текущих расходов:

$$Y = 1208,5 \cdot I^{0,044} \cdot P^{0,246}, \text{ тыс. т.}$$

Параметры эконометрической модели: $A = 1208,5$ - коэффициент нейтрального технологического прогресса; $\alpha = 0,044 \%$ - рост биоресурсов при росте инвестиций на 1 %; $\beta = 0,246 \%$ - рост биоресурсов при росте текущих расходов на 1 %.

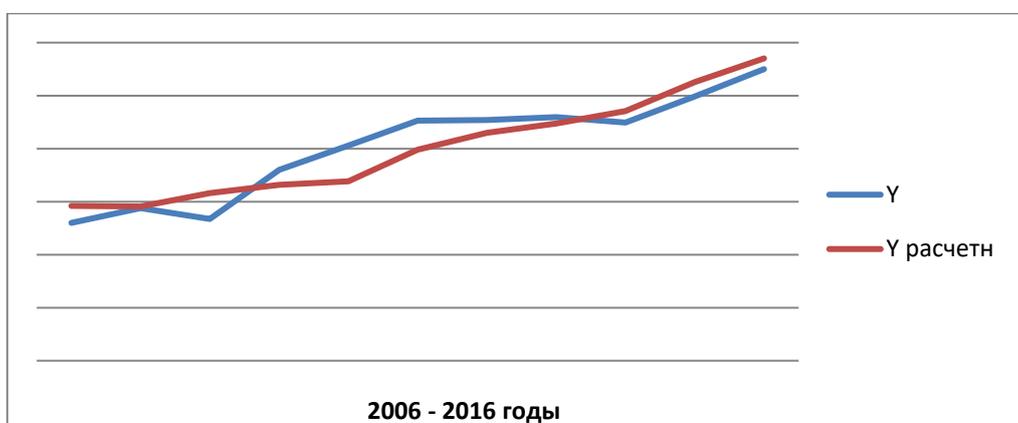


Рисунок 1.5 - Динамика фактического и расчетного объема добычи, тыс. т

Коэффициент корреляции фактических и рассчитанных по производственной функции результатов составил 0,925, что характеризует высокую степень связи фактических и модельных результатов. Полученную модель можно использовать для прогнозирования объема добычи.

Методические материалы по теме 4

Конспект лекций, Учебник [1, Кремер-Путко], Учебник [2, Настин], ФОС в ЭИОС.

Ссылки на рекомендуемые источники по теме 4:

[1, Кремер-Путко, с. 108-114, 124-127] – базовый учебник, которого мы придерживаемся, в том числе в терминологии, обозначениях.

[2, Настин, с. 59-63, 66-73] – основной учебник.

2 Методические указания для подготовки к практическим занятиям по 1-й части дисциплины

Ниже по первым четырём темам приведены вопросы для подготовки к практическим занятиям. Цель проведения практических занятий – закрепление теоретических знаний, полученных на лекциях и при самостоятельном изучении дисциплины "Эконометрика" часть 1, а также для выработки профессиональных умений и навыков, сформулированных в рабочей программе модуля "Анализ в решении экономических задач" по направлению подготовки 38.03.01 Экономика, профиль "Прикладная экономика".

Во введении настоящего пособия дана таблица 1, в которой показано распределение аудиторных часов для практических занятий по формам обучения и темам.

Практическими занятиями предусматриваются сочетание индивидуальных и групповых методов работы, решение задач, работа с тестами, ответы на вопросы. **Перед практическим занятием** студент должен постараться ответить на вопросы по очередной теме (вопросы приведены ниже), а для этого изучить конспект лекции, просмотреть рекомендованную преподавателем литературу. Следует обращать внимание на **практическое применение теории и на методику решения типовых задач. Наиболее целесообразная форма проведения занятия - интерактивное** практическое занятие: в процессе выполнения заданий приветствуются вопросы со стороны студентов и коллективный поиск ответов на них, обсуждение сложных моментов.

Для того, чтобы каждое практическое занятие прошло с максимальной пользой, настоятельно рекомендуем готовиться к ним. В самом кратком варианте такая подготовка может состоять в следующем. Ищется ответ на каждый вопрос по теме, который представлен здесь. При этом не следует стремиться охватить слишком большой объём материала. Нужно постараться ответить кратко и точно на поставленный вопрос. Ответы желательно законспектировать, вдумчиво несколько раз повторить их для себя. Для поиска ответов, выявления смысла рекомендуем использовать раздел 1 настоящего учебно-методического пособия, другие рекомендованные источники. При разборе материала крайне желательно использовать графические образы, стараться представить такой образ для любого объекта дисциплины. Например, для моделей магазина создать графический образ остаточной дисперсии, образ регрессии с доверительным интервалом прогноза.

С полученными в процессе подготовки знаниями студенту будет интересно участвовать в опросе по теме, работать с тестами, решать задачи.

Ниже даются рекомендации по формам проведения занятий, вопросы по темам. Заголовки даны для очной формы (во множественном числе - на каждую тему, как правило, 2 практических занятия по 2 часа). По очно-заочной форме на каждую тему выделен 1 час практических занятий.

Занятия по теме 1 "Основные понятия эконометрики".

Форма занятия: интерактивное практическое занятие.

План занятия:

- а) Опрос по материалам лекций.
- б) Работа с тестом.
- в) Решение задач, в том числе с использованием табличного редактора Excel.

Вопросы по теме 1:

- 1) Расшифровать аббревиатуры: ЗР, МНК, МО, НЗР, СВ, СКО.
- 2) Что является предметом эконометрики?
- 3) Как можно охарактеризовать метод эконометрики?
- 4) Какое место занимает эконометрика в системе экономических наук?
- 5) Сформулировать одно- и двухфакторные эконометрические модели магазина.
- 6) Сформулировать 5-факторную модель прогнозирования банкротства предприятия.
- 7) На какие две группы линия регрессии разделяет магазины?
- 8) Назвать несколько приёмов исследования функции.
- 9) Что такое дифференциальный и интегральный законы распределения случайной величины?
- 10) Дать определение регрессии, записать её формульное выражение, создать графический образ.

Занятия по теме 2: "Парный регрессионный анализ".

Форма занятия: интерактивное практическое занятие.

План занятия:

- а) Опрос по материалам лекций.
- б) Работа с тестом.
- в) Решение задач, в том числе с использованием табличного редактора Excel.

Вопросы по теме 2:

- 1) Раскрыть смысл переменных с именами: n , m , p , ε , s и σ .
- 2) В чём суть метода наименьших квадратов?
- 3) В чём различие между параметрами уравнения регрессии β_0 , β_1 и b_0 , b_1 ?

- 4) Назвать свойства парного коэффициента корреляции.
- 5) Сформулировать пять предпосылок регрессионного анализа.
- 6) Что утверждает теорема Гаусса-Маркова?
- 7) Как оценить значимость уравнения парной регрессии по критерию Стьюдента?
- 8) Что общего и в чем отличие коэффициента корреляции r_{xy} и ковариации $Cov(X, Y)$?
- 9) Определить понятия: доверительная вероятность, доверительный интервал.
- 10) Раскрыть смысл понятия доверительного интервала для регрессии $M_x(Y)$ как модели магазина при площади $x=11$.

Занятия по теме 3: "Множественный регрессионный анализ".

Форма занятия: интерактивное практическое занятие.

План занятия:

- а) Опрос по материалам лекций.
- б) Работа с тестом.
- в) Решение задач, в том числе с использованием табличного редактора Excel.

Вопросы по теме 3:

- 1) Пусть A и B - матрицы; верно ли утверждение: всегда выполняется равенство $A \cdot B = B \cdot A$?
- 2) Записать формулу обращения матрицы.
- 3) Дать толкование каждому из 6-ти терминов названия "Классическая нормальная линейная модель множественной регрессии".
- 4) Записать критерий метода наименьших квадратов в скалярной и матричной формах.
- 5) Записать алгебраическое выражение для выборочной оценки ковариационной матрицы вектора параметров регрессии b .
- 6) Записать в скалярной и матричной форме выражения для оценки дисперсии отклонений (ошибок).
- 7) Записать в общем виде доверительный интервал для коэффициента регрессии, создать графический образ интервала.
- 8) Записать в общем виде доверительный интервал для регрессии $M_x(Y)$, создать графический образ этого интервала.
- 9) В чём суть мультиколлинеарности, как отбирать значимые факторы на основе корреляционной матрицы?
- 10) В чём суть мультиколлинеарности и как отбирать значимые факторы методом их "вращения"?

11) Какие алгебраические и содержательные неприятности влечет высокая мультиколлинеарность?

12) В чём смысл, что показывает функция эластичности?

13) Как соотносятся между собой коэффициент эластичности и функция эластичности?

Занятия по теме 4: "Нелинейный регрессионный анализ.

Форма занятия: интерактивное практическое занятие.

План занятия:

а) Опрос по материалам лекций.

б) Работа с тестом.

в) Решение задач.

Вопросы по теме 4:

1) Нарисуйте бинарное дерево классификации нелинейных моделей регрессии.

2) К какому классу нелинейности относится модель $y = \beta_0 + \beta_1 x^2 + \varepsilon$?

3) К какому классу нелинейности относится модель $y = \beta_0 + \beta_1/x + \varepsilon$?

4) К какому классу нелинейностей относится степенная модель $y = \beta_0 x^{\beta_1} \varepsilon$?

5) Привести степенную модель $y = \beta_0 x^{\beta_1} \varepsilon$ к линейному виду.

6) . Какая проблема возникает при применении МНК-метода для моделей, нелинейных по параметрам и внутренне нелинейных?

7) Если атрибутивная переменная имеет 3 значения, то сколько фиктивных (булевских) переменных нужно включить в модель?

8) Сколько факторных переменных в функции Кобба-Дугласа $\hat{Y} = AK^\alpha L^\beta$, в каких единицах они измеряются?

9) Сколько факторных переменных в функции Кобба-Дугласа $\hat{Y} = AK^\alpha L^\beta$?

10) В каких единицах можно измерять факторные переменные в функции Кобба-Дугласа $\hat{Y} = AK^\alpha L^\beta$?

3 Методические указания по выполнению контрольной работы № 1

3.1 Общие сведения, выбор варианта и вопросов, исходных данных для задач

Структура работы. Контрольная работа №1 состоит из 6 разделов.

- 1) Вопрос №__ : <Название вопроса>.
- 2) Вопрос №__ : <Название вопроса>.
- 3) Задача-1 (линейная парная регрессия, ручное решение).
- 4) Задача-2 (линейная троичная регрессия, ручное решение).
- 5) Задача-3 (линейная троичная регрессия, решение в MS Excel).
- 6) Задача-4 (нелинейная регрессия, решение в MS Excel).

В приложении Б дана структура работы в форме листа "СОДЕРЖАНИЕ".

Общие требования к оформлению и объёму

1) В основу правил оформления положен ГОСТ 7.32-2017 "Структура и правила оформления отчёта по НИР", другие стандарты. Подробно правила оформления изложены в источнике [5, Мнацаканян и др.].

2) Из текста контрольной работы следует исключить пропуски строк между абзацами; должен быть включён автоперенос.

3) Объём работы не должен превышать 25 страниц (от титульного листа до последней страницы, включая приложения). Контрольные работы с объёмом больше 25 страниц возвращаются на доработку.

Правила заимствования. При заимствовании материала из источников нужно соблюдать следующие правила:

- не следует выхватывать из сети первые попавшиеся тексты, они обычно адресованы студентам по высшей математике;
- следует выбирать источники, которые по терминологии и обозначениям близки к тексту настоящего пособия; обычно можно ограничиться нашими источниками [1, 2];
- терминология и обозначения должны быть едиными во всех частях работы;
- следует делать ссылки на источники, например, так: [3], или так: [5, с. 116], или так: [6, Берндт].

Особенность работы. Трудоемкость выполнения контрольной работы снижается на порядок за счёт использования ворд-файла настоящего раздела пособия. Ниже для решения *каждой* задачи даются подробнейшие инструкции и наглядные примеры.

Большинство дальнейших решений (примеров) выполнено для условного варианта А-Б.

Выбор варианта и номеров вопросов. Вариант выбирается по первым подходящим буквам полного имени и фамилии. Например, если Вы – тАтьяна Соколова, то ваш вариант А-С, ваши вопросы 14 и 31, см. таблицу 3.1.

Таблица 3.1 - Варианты и номера теоретических вопросов

Имя*	Первая подходящая буква фамилии													
	А	Б	В	Г	Д	Е	К	Л	М	Н	О	П	Р	С
<i>А</i>	1, 18	2, 19	3, 20	4, 21	5, 22	6, 23	7, 24	8, 25	9, 26	10, 27	11, 28	12, 29	13, 30	14, 31
<i>В</i>	15, 32	16, 33	17, 34	18, 35	19, 36	20, 37	21, 1	22, 2	23, 3	24, 4	25, 5	24, 6	25, 7	26, 8
<i>Е</i>	27, 9	28, 10	29, 11	30, 12	31, 13	32, 14	33, 15	34, 16	35, 17	36, 18	37, 19	1, 20	2, 21	3, 22
<i>О</i>	4, 23	5, 24	6, 25	7, 26	8, 27	9, 28	10, 29	11, 30	12, 31	13, 32	14, 33	15, 34	16, 35	17, 36
<i>И</i>	18, 37	19, 1	20, 2	21, 3	22, 4	23, 5	24, 6	25, 7	26, 8	27, 9	28, 10	29, 11	30, 12	31, 13

*Первая подходящая буква полного имени.

Список вопросов для контрольной работы № 1 см. в разделе 4. Ответ на каждый вопрос не должен занимать больше двух страниц.

Исходные данные. Во всех вариантах задачи схожи по экономическому смыслу, но у каждого варианта имеются свои исходные данные. Для 1-й контрольной работы это столбцы X1, X2, Y, см Приложение А1, параметры, которые задаются преподавателем через таблицу А.7.

Принятые в наших примерах значения параметров: $\Gamma_1 = \Gamma_2 = 0$, $X_{10} = 11$, $X_{20} = 8$, $\alpha = 0,05$ (отсюда доверительная вероятность $\gamma = 1 - 0,05 = 0,95$).

Порядок действий на начальном этапе.

1) После того, как выбран вариант (например, А-С), по первой его букве выбирается подходящая таблица из набора таблиц А.1-А.5, кроме того, используются таблицы А.6 и А.7 (последняя задаётся преподавателем).

2) Из выбранной таблицы, для нашего варианта это таблица А.1, считывается столбец X1 (он самый левый, выделен жирно).

3) Из этой же таблицы А.1 по второй букве варианта выбирается и считывается столбец Y с учётом значения слагаемого параметра Γ_1 .

4) Из таблицы А.6 выбирается столбец X2 по букве имени с учётом значения слагаемого параметра Γ_2 .

5) Далее выбираются значения параметров из таблицы А.7.

3.2 Пример ручного решения задачи-1: "Моделирование магазина на основе парной линейной регрессии"

Постановка задачи-1. Торговая компания имеет семь магазинов типа "Морепродукты" и планирует построить еще один с торговой площадью 1100 м². Для этого она разрабатывает бизнес-план и в том числе - эконометрическую модель магазина. На этой модели специалисты хотят исследовать зависимость среднегодовой суточной выручки (у - в сотнях тыс. руб./день) от площади магазина (х - в сотнях м²) и спрогнозировать среднесуточную выручку восьмого магазина.

Решение задачи-1.

1) Выберем исходные данные из таблицы А.1 и перепишем их в ещё не заполненную таблицу 3.2, графы 2 и 3. Построим линейный график зависимости Y от X с помощью MS Excel, инструкцию см. в Приложении Г, см. также рисунок 3.1. Как видно, корреляционное поле хорошо аппроксимируется прямой линией, поэтому принимаем гипотезу о правомерности применения линейной парной регрессии.

Заполним графы 4-8 таблицы 3.2 на основе формул, указанных в головке таблицы и в последней строке. Выражения с чертой – это средние арифметические, например, $\overline{x \cdot y}$ - это среднее арифметическое произведения выборок X на Y.

2) Найдём на основе таблицы 3.2 МНК-методом параметры b_0 и b_1 уравнения регрессии:

$$\hat{y} = b_0 + b_1 x. \quad (3.1)$$

Для этого используем формулы:

$$b_1 = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{x^2 - \bar{x}^2}; \quad b_0 = \bar{y} - b_1 \bar{x}. \quad (3.2)$$

Таблица 3.2 – Расчётная таблица для задачи-1

№ п/п	x_i	y_i	x^2	y^2	$x_i y_i$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$	\hat{y}_{x_i}	$e_i =$ $= \hat{y}_{x_i} - y_i$	$e_i^2 =$ $= (\hat{y}_{x_i} - y_i)^2$
1	1	2	1	4	2	5,90	13,76	1,97	-0,03	0,00
2	1	3	1	9	3	5,90	7,34	1,97	-1,03	1,06
3	2	4	4	16	8	2,04	2,92	3,51	-0,49	0,22
4	3	5	9	25	15	0,18	0,50	5,05	0,05	0,00
5	4	5	16	25	20	0,32	0,50	6,59	1,59	2,53
6	5	7	25	49	35	2,46	1,66	8,13	1,13	1,28
7	8	14	64	196	112	20,88	68,72	12,75	-1,25	1,56
Σ	24	40	120	324	195	37,68	95,40	-	-0,03	6,65

Найдём средние значения показателей по графам (обращаем внимание: в графах 7, 8, 10 числители делятся не на $n=7$; это делается, чтобы получить несмещённые оценки параметров, подробно см. Приложение Г):

Графа 2. $\bar{x}=24/7=3,43.$	Графа 3. $\bar{y}=40/7=5,71.$	Графа 4. $\bar{x}^2=120/7=17,14.$
Графа 5. $\bar{y}^2=324/7=46,29.$	Графа 6. $\overline{xy}=195/7=27,86.$	Графа 7. Дисперсия $s_x^2=37,68/(7-1) = 6,28.$
Графа 8. Дисперсия $s_y^2=95,40/(7-1)=15,90.$	Графа 9. -	Графа 10: сумма стремится к 0.
Графа 11. Остаточная дисперсия. $s^2=6,65/(7-2)=1,33.$	-	-

Рассчитаем искомые оценки параметров регрессии:

$$b_1 = (27,86 - 3,43 \cdot 5,71) / (17,14 - 3,43^2) = 8,27 / 5,38 = 1,54; \quad b_0 = 5,71 - 1,54 \cdot 3,43 = 0,43.$$

Наше уравнение регрессии будет иметь вид:

$$\hat{y} = 0,43 + 1,54x. \quad (3.3)$$

3) Построим по исходным данным линию регрессии на координатной плоскости XOY с помощью MS Excel, рисунок 1.1 (инструкцию см. в Приложении Г). Определяем по рисунку приблизительные значения b_1 и b_0 , они совпадают с расчётными, следовательно, последние определены верно.

4) Покажем графически и аналитически, что линия регрессии проходит через точку средних (\bar{x}, \bar{y}) . Из рисунка 1.1 видно, что линия регрессии проходит через точку “средних” $(\bar{x}=3,43; \bar{y}=5,71)$. Проверим это аналитически: $\hat{y} = 0,43 + 1,54 \cdot 3,43 = 5,71$, что и требовалось показать.

5) Сформулируем экономический смысл параметров регрессии:

а) коэффициент $b_1 = 1,54 = 1,54 \cdot 10^5 \text{руб}/100\text{м}^2 = 1540 \text{руб./м}^2$ – удельная суточная выручка – с одного квадратного метра – если без учёта b_0 ;

б) свободный член $b_0 = 0,43$ смысла не имеет - при нулевой площади выручка может быть только нулевой.

6) Вычислим коэффициент корреляции между переменными X и Y по формуле:

$$r_{xy} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{s_x s_y} = \frac{C\hat{o}v(X, Y)}{s_x s_y}. \quad (3.4)$$

Здесь известно все, кроме СКО x и СКО y в знаменателе формулы, значения *несмещённых* дисперсий см. выше. Вычислим *несмещённые* СКО:

$$\text{СКО } s_x = \sqrt{s_x^2} = \sqrt{6,28} = 2,51; \text{ СКО } s_y = \sqrt{s_y^2} = \sqrt{15,90} = 3,99.$$

Переведём числитель – ковариацию - с помощью коэффициента $k=n/(n-1)$ в разряд несмещённых и получим корректное значение коэффициента корреляции:

$$r_{xy} = \frac{(27,86 - 3,43 \cdot 5,71) \cdot 7 / (7-1)}{2,51 \cdot 3,99} = \frac{9,65}{10,01} = 0,96.$$

Значение коэффициента корреляции – очень большое - говорит о высокой зависимости выручки от размера торговой площади магазина, именно она в основном и определяет размер выручки.

7) Определим графически и аналитически точечный прогноз - регрессионное среднесуточное значение выручки для магазина-8 с площадью $x = 11$. Прогнозное значение из рисунка 3.1 и из формулы совпадают:

$$\hat{y} = 0,43 + 1,54 \cdot 11 = 17,37 \text{ (1737000 руб./сут.)}$$

8) Определим интервальный регрессионный прогноз выручки для магазина-8 (площадь $x = 11$), используя доверительную вероятность $\gamma = 0,95 = 95\%$ (уровень значимости $\alpha = 0,05 = 5\%$).

Чтобы построить доверительный интервал для СВ $\hat{y}_{x=11}$, определим её дисперсию $s_{\hat{y}_{x=11}}^2$ и СКО, используя результаты вычислений из таблицы 3.2 (остаточное СКО $s = \sqrt{1,33} = 1,15$):

$$s_{\hat{y}_{x=11}}^2 = s^2 \left(\frac{1}{n} + \frac{(x - \bar{x})^2}{\sum (x_i - \bar{x})^2} \right) = 1,15^2 \left(\frac{1}{7} + \frac{(11 - 3,43)^2}{37,68} \right) = 2,20, \quad s_{\hat{y}_{x=11}} = \sqrt{2,20} = 1,48.$$

Искомая система неравенств, определяющая доверительный интервал для регрессионного значения выручки магазина-8, имеет вид:

$$\hat{y} - t_{1-\alpha, k} s_{\hat{y}} \leq M_x(Y) \leq \hat{y} + t_{1-\alpha, k} s_{\hat{y}}.$$

Из таблицы Е.1 значений критерия Стьюдента находим значение критерия $t_{0,95; 7-2} = 2,57$. Вычисляем границы интервала:

- нижнее значение интервала: $17,37 - 2,57 \cdot 1,48 = 13,57$;
- верхнее значение интервала: $17,37 + 2,57 \cdot 1,48 = 21,37$.

Окончательно доверительный интервал – прогноз выручки - имеет вид:

$$\mathbf{13,57 \leq M_{x=11}(Y) \leq 21,37.}$$

9) Найдём с надёжностью 0,95 интервальную оценку для коэффициента регрессии β_1 . Общая формула для расчета интервала:

$$b_1 - \Delta \leq \beta_1 \leq b_1 + \Delta,$$

где $\Delta = t_{1-\alpha, n-2} s / \sqrt{\sum(x_i - \bar{x})^2} = 2,57 \cdot 1,15 / \sqrt{37,68} = 0,48$.

Вычисляем:

- нижнее значение интервала: $1,54 - 0,48 = 1,06$;
- верхнее значение интервала: $1,54 + 0,48 = 2,02$.

Окончательно доверительный интервал имеет вид: $1,06 \leq \beta_1 \leq 2,02$. Здесь числа – значения тангенса углов. Применяя к ним функцию \arctg , получаем интервал в градусах: $47^\circ \leq \alpha \leq 64^\circ$. Такой интервал легко изобразить на рисунке.

10) Найдём с надёжностью 0,95 интервальную оценку для остаточной дисперсии σ^2 . Из статистической таблицы Е.2 для значений критерия Пирсона находим два значения статистики $\chi_{\alpha, k}^2$ (хи-квадрат):

$$\chi_{\frac{\alpha}{2}; n-2}^2 = \chi_{0,025; 5}^2 = 12,8; \quad \chi_{1-\alpha/2; n-2}^2 = \chi_{0,975; 5}^2 = 0,831;$$

Формула для расчёта доверительного интервала:

$$\frac{ns^2}{\chi_{\alpha/2; n-2}^2} \leq \sigma^2 \leq \frac{ns^2}{\chi_{1-\alpha/2; n-2}^2}$$

Границы интервала:

$$s_{\text{нижн}}^2 = 7 \cdot 1,15^2 / 12,8 = 0,72; \quad s_{\text{верх}}^2 = 7 \cdot 1,15^2 / 0,83 = 11,14.$$

Ответ: доверительный интервал:

$$0,72 \leq \sigma^2 \leq 11,14.$$

11) Оценим на уровне $\alpha = 0,05$ значимость уравнения регрессии Y по X по критерию Стьюдента.

Уравнение парной регрессии значимо, если $t > t_{\text{крит}}$. Значение статистики Стьюдента:

$$t = \frac{b_1}{s} \sqrt{\sum(x_i - \bar{x})^2} = \frac{1,54}{1,15} \sqrt{37,68} = 8,22.$$

По статистической таблице находим $t_{\text{крит}} = t_{0,95; 7-2=5} = 2,57$. Так как $8,22 > 2,57$, то нулевую гипотезу H_0 ($H_0: \beta_1 = 0$) отвергаем и принимаем противоположную гипотезу $H_1: \beta_1$ значимо отличается от 0, следовательно, уравнение значимо.

12) Оценим на уровне $\alpha = 0,05$ значимость уравнения регрессии Y по X по критерию Фишера. Для этого используем таблицу 3.2:

- общая сумма – восьмая графа: $Q = \sum(y_i - \bar{y})^2 = 95,40$;
- регрессионная сумма: из 9-й графы вычтем среднее 5,71, разности возведём в квадрат и суммируем:

$$Q_R = \sum (\hat{y}_i - \bar{y})^2 = 13,99 + 13,99 + 4,84 + 0,44 + 0,78 + 5,86 + 49,56 = 89,46;$$

- остаточная сумма – одиннадцатая графа: $Q_e = \sum (\hat{y}_i - y_i)^2 = 6,65$.

Значение статистики Фишера рассчитаем по формуле:

$$F = \frac{Q_R / (m-1)}{Q_e / (n-m)} = \frac{s_R^2}{s^2} = \frac{89,46 / (2-1)}{1,15^2} = 67,64.$$

Уравнение регрессии значимо, если $F > F_{\alpha, k_1, k_2}$, здесь степени свободы $k_1 = m-1 = 2-1 = 1$, $k_2 = n-m = 7-2 = 5$. По статистической таблице Е.3 для критерия Фишера находим: $F_{0,05;1;5} = 6,61$. Так как $67,64 > 6,61$, то уравнение значимо: коэффициент регрессии $b_1 = 1,54$ значимо отличается от нуля.

13) Оценим значимость уравнения регрессии Y по X по степени близости коэффициента детерминации R^2 к 1:

$$R^2 = Q_R / Q = 89,46 / 95,40 = 0,94.$$

Как видно, R^2 очень близок к 1. Следовательно, уравнение регрессии весьма значимо. Смысл R^2 : коэффициент показывает, какая доля вариации результирующей переменной Y обусловлена вариацией фактора X , эта доля составляет 94 %.

3.3 Пример ручного решения задачи-2: "Моделирование магазина на основе троичной линейной регрессии"

Постановка задачи-2. Специалисты компании решили усложнить модель магазина путём введения в неё второй факторной переменной. Из различных факторов была отобрана переменная X_2 – размер паркинга магазина. Проектное значение паркинга у магазина-8 равно 80 автомобилям. Требуется построить новую модель, исследовать её, выполнить с её помощью прогноз выручки магазина-8.

Решение задачи-2.

1) Выберем данные нашего варианта "А-Ъ" для X_1 и Y из таблицы А.1, значения X_2 из таблицы А.6 и перенесём их в таблицу 3.3.

Таблица 3.3 – Исходные данные для задачи-2, вариант "А-Ъ"

X_1	1	1	2	3	4	5	8	<i>11*</i>	$m^2 * 100$
X_2	1	2	2	3	3	4	6	<i>8*</i>	ед.*10
Y	2	3	4	5	5	7	14	-	руб./сут.* 10^5

*Курсив – наши проектные данные для магазина-8.

Как видно из таблицы 3.3 зависимость $\hat{y}(x_1, x_2)$ прямая. Проверим, можно ли аппроксимировать её плоскостью – линейной функцией:

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2. \quad (3.5)$$

Для этого положим $x_1 = 0$ и получим функцию-линию $\hat{y} = b_0 + b_2 x_2$ – след от пересечения плоскости (3.5) с плоскостью $x_2 = 0$. Чтобы увидеть этот след, нужно построить график зависимости y от x_2 , используя табличный редактор Excel (инструкцию см. в Приложении Г). Из рисунка 3.1 - точка (3, 5) сдвояна - видно, что прямая линия хорошо аппроксимирует связь между y и x_2 .

2) Запишем матрицу X значений объясняющих переменных (матрицу плана), см. ниже среднюю матрицу. Запишем слева от неё транспонированную матрицу плана X' (здесь штрих – оператор транспонирования). Найдём матрицу A как произведение матриц $X'X$:

$$A = X'X = \begin{pmatrix} 11 & 11 & 11 & 11 & 11 & 11 \\ 1 & 1 & 2 & 3 & 4 & 5 & 8 \\ 1 & 2 & 2 & 3 & 3 & 4 & 6 \end{pmatrix} \cdot \begin{pmatrix} 11 & 1 \\ 11 & 2 \\ 12 & 2 \\ 13 & 3 \\ 14 & 3 \\ 15 & 4 \\ 18 & 6 \end{pmatrix} = \begin{pmatrix} 7 & 24 & 21 \\ 24 & 120 & 96 \\ 21 & 96 & 79 \end{pmatrix}.$$

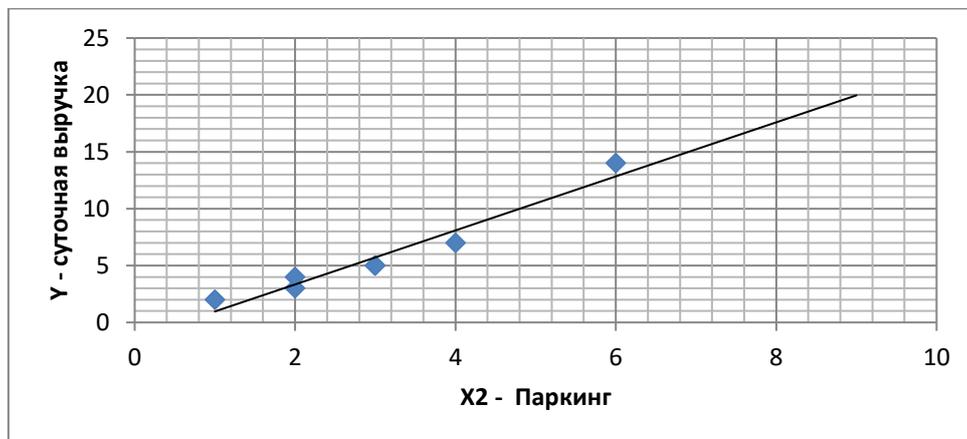


Рисунок 3.1 - Корреляционное поле и линия регрессии $\hat{y} = b_0 + b_2 x_2$

3) Найдём обратную матрицу $A^{-1} = (X'X)^{-1}$ по формуле:

$$A^{-1} = \frac{1}{|A|} (A_{ij})',$$

где $|A|$ - определитель матрицы A ;

$(A_{ij})'$ – транспонированная матрица, составленная из алгебраических дополнений матрицы A .

Находим определитель:

$$|A| = 7 \cdot 120 \cdot 79 + 24 \cdot 96 \cdot 21 + 21 \cdot 96 \cdot 24 - 21 \cdot 120 \cdot 21 - 96 \cdot 96 \cdot 7 - 79 \cdot 24 \cdot 24 = 192.$$

Находим алгебраические дополнения (слева сумма индексов чётная, справа - нечётная):

$$\begin{aligned} A_{11} &= 120 \cdot 79 - 96 \cdot 96 = 264; & A_{12} &= -(24 \cdot 79 - 96 \cdot 21) = 120; \\ A_{13} &= 24 \cdot 96 - 120 \cdot 21 = -216; & A_{21} &= -(24 \cdot 79 - 21 \cdot 96) = 120; \\ A_{22} &= 7 \cdot 79 - 21 \cdot 21 = 112; & A_{23} &= -(7 \cdot 96 - 24 \cdot 21) = -168; \\ A_{31} &= 24 \cdot 96 - 21 \cdot 120 = -216; & A_{32} &= -(7 \cdot 96 - 21 \cdot 24) = -168; \\ A_{33} &= 7 \cdot 120 - 24 \cdot 24 = 264. \end{aligned}$$

Искомая обратная матрица:

$$A^{-1} = \frac{1}{192} \begin{pmatrix} 264 & 120 & -216 \\ 120 & 112 & -168 \\ -216 & -168 & 264 \end{pmatrix} = \begin{pmatrix} 1,375 & 0,625 & -1,125 \\ 0,625 & 0,584 & -0,875 \\ -1,125 & -0,875 & 1,375 \end{pmatrix}.$$

Проверим, выполняется ли равенство: $A^{-1}A = E$.

$$\begin{aligned} AA^{-1} &= \frac{1}{192} \begin{pmatrix} 7 & 24 & 21 \\ 24 & 120 & 96 \\ 21 & 96 & 79 \end{pmatrix} \begin{pmatrix} 264 & 120 & -216 \\ 120 & 112 & -168 \\ -216 & -168 & 264 \end{pmatrix} = \\ &= \frac{1}{192} \begin{pmatrix} 192 & 0 & 0 \\ 0 & 192 & 0 \\ 0 & 0 & 192 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \end{aligned}$$

Вывод: равенство выполняется, расчет обратной матрицы выполнен верно.

4) Найдём произведение матриц $X'Y$:

$$X'Y = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 2 & 3 & 4 & 5 & 8 \\ 1 & 2 & 2 & 3 & 3 & 4 & 6 \end{pmatrix} \cdot \begin{pmatrix} 2 \\ 3 \\ 4 \\ 5 \\ 5 \\ 7 \\ 14 \end{pmatrix} = \begin{pmatrix} 40 \\ 195 \\ 158 \end{pmatrix}.$$

5) Найдём МНК-методом параметры уравнения троичной регрессии (3.5) (ниже b – это вектор параметров регрессии):

$$b = (X'X)^{-1} X'Y = (1/192) \begin{pmatrix} 264 & 120 & -216 \\ 120 & 112 & -168 \\ -216 & -168 & 264 \end{pmatrix} \cdot \begin{pmatrix} 40 \\ 195 \\ 158 \end{pmatrix} = (1/192) \cdot \begin{pmatrix} -168 \\ 96 \\ 312 \end{pmatrix} = \begin{pmatrix} -0,88 \\ 0,50 \\ 1,63 \end{pmatrix}.$$

Ответ: $b_0 = -0,88$; $b_1 = 0,50$; $b_2 = 1,63$. Наше уравнение троичной регрессии имеет вид:

$$\hat{y} = -0,88 + 0,50x_1 + 1,63x_2. \quad (3.6)$$

6) **Найдём точечный прогноз** выручки – регрессионное значение - для проектируемого магазина-8 с площадью $x_1 = 11$ (1100 м²) и площадью паркинга $x_2 = 8$ (80 автомашин): $\hat{y} = -0,88 + 0,50 \cdot 11 + 1,63 \cdot 8 = 17,66$.

7) **Найдём 95 %-ный доверительный интервальный прогноз** выручки для регрессионного значения магазина-8 при проектных значениях факторов $x_1 = 11$, $x_2 = 8$. Для этого нужно найти остаточную дисперсию s^2 отклонения ε и дисперсию $s_{\hat{y}}^2$ СВ \hat{y} по формулам:

$$s^2 = \frac{e'e}{n - (p + 1)} = \frac{\sum e_i^2}{n - p - 1}, \quad s_{\hat{y}} = s \sqrt{X_0'(X'X)^{-1}X_0};$$

где $X_0 = (1 \ 11 \ 8)'$ – вектор проектных значений переменных $x_0 = 1$, $x_1 = 11$, $x_2 = 8$. Составим расчётную таблицу 3.4.

Таблица 3.4 – Расчётная таблица для задачи-2

№ п/п	x_{i1}	x_{i2}	y_i	$x_{i1} \cdot x_{i2}$	\hat{y}_i	e_i	e_i^2 *	\dot{x}_1^2 *	\dot{x}_2^2 *
1	2	3	4	5	6	7	8	9	10
1	1	1	2	1	1,25	0,75	0,56	5,90	4,00
2	1	2	3	2	2,88	0,12	0,02	5,90	1,00
3	2	2	4	4	3,38	0,62	0,39	2,04	1,00
4	3	3	5	9	5,51	-0,51	0,26	0,18	0,00
5	4	3	5	12	6,01	-1,01	1,02	0,32	0,00
6	5	4	7	20	8,14	-1,14	1,30	2,46	1,00
7	8	6	14	48	12,90	1,10	1,21	20,88	9,00
Итого	24	21	40	96	40,07	-0,07	4,76	37,71	16,00
Средн. значен.	3,43	3,0	5,71	13,71	-	-	4,76/4 = 1,19	37,71/6 = 6,29	16,0/6 = 2,67

*В графах 9 и 10 переменные x_1 и x_2 центрированы и возведены в квадрат. В графах 8-10 дисперсии s^2 , $s_{x_1}^2$ и $s_{x_2}^2$ исправленные (они несмещённые – "хорошие").

На основе табличных данных рассчитаем стандартные отклонения:

$$\text{остаточное СКО СВ } \varepsilon: s = \sqrt{1,19} = 1,09;$$

$$\text{СКО } x_1: s_{x_1} = \sqrt{6,29} = 2,51; \text{ СКО } x_2: s_{x_2} = \sqrt{2,67} = 1,63.$$

Продолжим вычисления.

$$X_0'(X'X)^{-1} = (1 \ 11 \ 8) \begin{pmatrix} 1,375 & 0,625 & -1,125 \\ 0,625 & 0,584 & -0,875 \\ -1,125 & -0,875 & 1,375 \end{pmatrix} = (-0,750 \ 0,049 \ 0,250),$$

$$(X_0'(X'X)^{-1})X_0 = (-0,750 \ 0,049 \ 0,250) \begin{pmatrix} 1 \\ 11 \\ 8 \end{pmatrix} = 1,789 \approx 1,79,$$

$$s_{\hat{y}} = 1,09\sqrt{1,79} = 1,46.$$

По таблице Е.1 находим критическое значение статистики Стьюдента $t_{0,95; 7-2-1=4} = 2,78$. Вычисляем полуинтервал:

$$\Delta = t_{0,95; 4} \cdot s_{\hat{y}} = 2,78 \cdot 1,46 = 4,05.$$

Нижняя граница доверительного интервала:

$$\hat{y}_{min} = \hat{y}_{X_0} - \Delta = 17,66 - 4,05 = 13,61.$$

Верхняя граница доверительного интервала:

$$\hat{y}_{max} = \hat{y}_{X_0} + \Delta = 17,66 + 4,05 = 21,71.$$

Ответ: искомый доверительный интервал для прогнозного регрессионного значения \hat{y}_{X_0} задаётся системой неравенств:

$$13,61 \leq M_{X_0}(Y) \leq 21,71. \quad (3.7)$$

8) Рассчитаем коэффициент корреляции между факторными переменными – проверка на мультиколлинеарность:

$$r_{x_1x_2} = \frac{\overline{X_1 \cdot X_2} - \overline{X_1} \cdot \overline{X_2}}{s_{x_1} \cdot s_{x_2}}.$$

Здесь все выражения и переменные вычислены в таблице 3.4 и под ней. Искомый коэффициент корреляции факторных переменных (в числителе поправочный коэффициент $k = 7/6 = 1,17$, см. Приложение Д):

$$r_{x_1x_2} = \frac{(13,71 - 3,43 \cdot 3,00) \cdot 7/6}{2,51 \cdot 1,63} = 0,97.$$

Вывод: мультиколлинеарность факторов X_1 и X_2 очень большая, поэтому в дальнейшем можно ожидать противоречивых результатов.

9) Проверим значимость коэффициентов регрессии. СКО коэффициента b_j рассчитывается по формуле:

$$s_{b_j} = s \sqrt{[(X'X)^{-1}]_{jj}},$$

где $j = 0, 1, 2$, а выражение под корнем есть диагональный элемент обратной матрицы $(X'X)^{-1}$, см. п. 3. Отсюда:

$$s_{b_1} = 1,09 \sqrt{0,584} = 0,83; \quad s_{b_2} = 1,09 \sqrt{1,375} = 1,28.$$

Вычислим: $t = |b_1| / s_{b_1} = 0,50 / 0,83 = 0,60$. Так как $0,60 < t_{0,95; 4} = 2,78$, то **коэффициент b_1 незначим** (незначимо отличается от нуля).

Вычислим: $t = |b_2| / s_{b_2} = 1,63 / 1,28 = 1,27$. Так как $1,27 < t_{0,95; 4} = 2,78$, то **коэффициент b_2 незначим**.

Выводы: а) на такой результат могли повлиять: мультиколлинеарность факторов, см. п. 8, и слишком малые объёмы выборок ($n=7$); б) полученный результат влечёт незначимость уравнения регрессии, см. п. 10.

10) Оценим значимость уравнения регрессии на уровне $\alpha = 0,05$ по коэффициенту детерминации:

$$R^2 = Q_R / Q = (b'X'Y - n\bar{y}^2) / (Y'Y - n\bar{y}^2)$$

$$b'(X'Y) = (-0,88 \quad 0,50 \quad 1,63) \begin{pmatrix} 40 \\ 195 \\ 158 \end{pmatrix} = 319,84.$$

$\bar{y}=40/7=5,72$ (см. таблицу 4.1); $Y'Y=\sum y_i^2=324$. Окончательно:
 $R^2 = (319,84-7 \cdot 5,72^2)/(324-7 \cdot 5,72^2)=90,81/94,97=0,96$.

Вывод: значимость регрессии здесь очень высокая, хотя в п. 9 была установлена незначимость обоих коэффициентов регрессии и самой регрессии. Противоречие.

11) Оценим значимость уравнения регрессии на уровне $\alpha = 0,05$ по критерию Фишера. Регрессия значима, если справедливо неравенство:

$$F = R^2 (n-p-1) / ((1-R^2) p) > F_{\alpha; k_1; k_2}.$$

Вычислим: $F = 0,96(7-2-1) / ((1-0,96)2) = 48,0$. Отсюда неравенство:
 $48,0 > (F_{0,05; 2; 4} = 6,94)$.

Вывод: уравнение значимо, снова противоречие с п. 9.

Вывод по решению задачи-2: полученная двухфакторная модель магазина незначима. Её нельзя использовать. План её доработки:

- 1) рассчитать коэффициенты корреляции между x_1 и y , между x_2 и y ;
- 2) если первый коэффициент больше второго, то фактор x_2 исключить из регрессии, а вместо него включить, например, фактор x_3 – средняя ширина ассортимента магазина (ширина ассортимента – это общее количество категорий товаров);
- 3) если второй коэффициент больше первого, то фактор x_1 исключить из регрессии, а вместо него включить какой-нибудь фактор x_3 .

3.4 Пример решения задачи-3: "Моделирование магазина на основе линейной троичной регрессии в MS Excel"

Постановка задачи-3 и исходные данные те же, что и для задачи-2, см. таблицу 3.3. Меняется метод решения.

Для решения используется функция "Регрессия" из пакета анализа данных.

- 1) **Внесём** в лист MS Excel исходные данные своего варианта, рисунок 3.3.

	A	B	C	D	E
1	Y_i	X_{1i}	X_{2i}		
2	2	1	1		
3	3	1	2		
4	4	2	2		
5	5	3	3		
6	5	4	3		
7	7	5	4		
8	14	8	6		

Рисунок 3.3 – Исходные данные для построения тройной регрессии

2) Выберем пункт "Анализ данных" во вкладке "Данные". В появившемся окне найдём инструмент "Регрессия".

3) Зададим входные интервалы переменных X, Y и укажем интервал вывода итогов регрессионного анализа (верхнюю левую ячейку диапазона вывода данных). Выбор пункта "Остатки" позволит получить данные об остатках (ошибках) регрессии, рисунок 3.4.

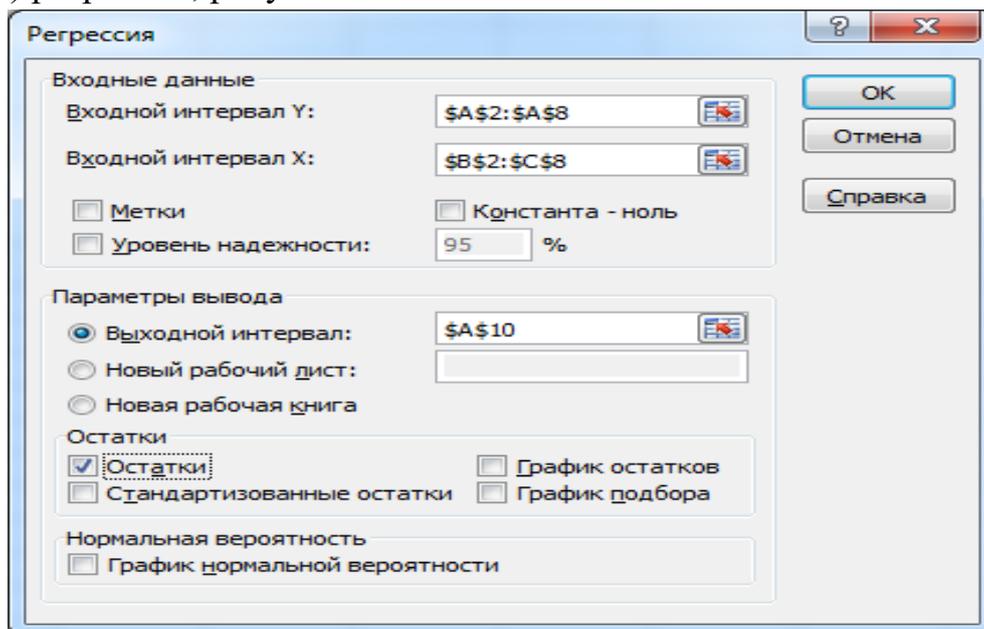


Рисунок 3.4 – Параметры построения тройной регрессии

4) Проведём расчёты и представим результаты в виде нескольких таблиц, см. рисунок 3.5. Расшифруем полученные данные.

а) Таблица "Регрессионная статистика" содержит:

- "Множественный R" – коэффициент множественной корреляции;
- "R-квадрат" – коэффициент детерминации;
- "Нормированный R-квадрат" – коэффициент детерминации, скорректированный на количество степеней свободы;
- "Стандартная ошибка" – квадратный корень из остаточной дисперсии;

- "Наблюдения" – количество наблюдений n .

б) Таблица "Дисперсионный анализ" содержит данные:

- столбец " df " - три числа степеней свободы: для регрессии, остатка и общее;
- столбец " SS " - три суммы квадратов отклонений: для регрессии, остатка и общее;
- столбец " MS " - два числа дисперсии на одну степень свободы: для регрессии и остатка;
- столбец " F " - значение СВ F , распределённой по ЗР Фишера;
- столбец "*Значимость F* " - p -вероятность для СВ F , т. е. вероятность того, что выполняется нулевая гипотеза о случайном отклонении коэффициента Фишера от нуля; в нашем примере p -значение ничтожно мало, поэтому гипотеза H_0 отклоняется.

в) Следующая таблица содержит информацию о параметрах уравнения регрессии:

- столбец "Коэффициенты" - значения параметров уравнения регрессии; строка "Y-пересечение" содержит значения параметра b_0 , строка "Переменная X1" - параметра b_1 , строка "Переменная X2" - параметра b_2 ;

- столбец "Стандартная ошибка" – стандартные ошибки параметров уравнения регрессии. Строка "Y-пересечение" - стандартная ошибка для параметра b_0 , строка "Переменная X1" – стандартная ошибка для параметра b_1 , строка "Переменная X2" – стандартная ошибка для параметра b_2 ;

- столбец "t-статистика" - значения t-критериев Стьюдента для параметров уравнения регрессии. Строка "Y-пересечение" – значение t-критерия Стьюдента для параметра b_0 , строка "Переменная X1" – значение t-критерия Стьюдента для параметра b_1 , строка "Переменная X2" – значение t-критерия Стьюдента для параметра b_2 ;

- столбец "p-значение" – p -значения t-критерия Стьюдента для параметров регрессии. Строка "Y-пересечение" – p -значение t-критерия Стьюдента для параметра b_0 , строка "Переменная X1" – p -значение t-критерия Стьюдента для параметра b_1 , строка "Переменная X2" – p -значение t-критерия Стьюдента для параметра b_2 ;

- столбцы "Верхние 95 %" и "Нижние 95 %" – интервалы для параметров уравнения регрессии, построенные с доверительной вероятностью 95 %.

г) Таблица "Вывод остатка" содержит данные:

- "Предсказанное Y" - значения, рассчитанные по уравнению регрессии (\hat{y}).

- "Остатки" - разница между предсказанными значениями \hat{y} и наблюдаемыми значениями y - (e).

5) Конец решения задачи-3.

Ниже в таблице 3.5 приведён количественный анализ совпадения ручного и компьютерного решений.

Таблица 3.5 – Сравнение ручного и компьютерного решений

Обозначение и решение ручное – задача-2		Обозначение и решение в MS Excel – задача-3		Относит. отклонение, %
1	2	3	4	5*
а) Коэффициент детерминации R^2 , п. 10	0,96	R-квадрат	0,951	0,9
б) СКО отклонений s , п. 7	1,09	Стандартная ошибка	1,090	0,0
в) Свободный член b_0 , п. 5	-0,88	Y-пересечение Коэффициенты	-0,875	0,6
г) Коэффициент b_1 , п.5	0,50	Переменная X1 Коэффициенты	0,500	0,0
9 Коэффициент регрессии b_2 , п. 5	1,63	Переменная X1 Коэффициенты	1,625	0,3

* (графа 5: $\frac{(2)-(4)}{(4)}100\%$).

Регрессионная статистика		Дисперсионный анализ							
		df	SS	MS	F	Значимость F			
Множественный R	0,97479462								
R-квадрат	0,950224551								
Нормированный R-квадрат	0,925336826								
Стандартная ошибка	1,089724736								
Наблюдения	7								
		df	SS	MS	F	Значимость F			
Регрессия		2	90,67857143	45,33928571	38,18045113	0,002477595			
Остаток		4	4,75	1,1875					
Итого		6	95,42857143						
		Коэффициенты	Стандартная ошибка	t-статистика	P-Значение	Нижние 95%	Верхние 95%	Нижние 95,0%	Верхние 95,0%
Y-пересечение		-0,875	1,277815519	-0,684762383	0,531111679	-4,422784642	2,672784642	-4,422784642	2,672784642
Переменная X 1		0,5	0,832291015	0,600751409	0,580386945	-1,810810314	2,810810314	-1,810810314	2,810810314
Переменная X 2		1,625	1,277815519	1,271701569	0,27238898	-1,922784642	5,172784642	-1,922784642	5,172784642
Вывод остатка									
Наблюдение	Предсказанное Y	Остатки							
1	1,25	0,75							
2	2,875	0,125							
3	3,375	0,625							
4	5,5	-0,5							
5	6	-1							
6	8,125	-1,125							
7	12,875	1,125							

Рисунок 3.5 – Вывод итогов решения задачи-3

Вывод: относительное отклонение ручного и компьютерного решений не превысило 0,9 %, следовательно, ручное решение было выполнено верно.

3.5 Пример решения задачи-4: "Моделирование магазина на основе нелинейной регрессии в MS Excel"

Постановка задачи.

Ранее в задаче-1 мы отмечали два (возможных) недостатка линейной регрессионной модели: а) наличие b_0 , которое не имеет экономического смысла; б) реальная нелинейная зависимость выручки от площади заменялась линейной. Эти недостатки устраняются, если использовать квадратичное уравнение регрессии $\hat{y}=b_1x+b_2x^2$. Его особенности: а) при $x=0$ и $\hat{y}=0$; б) есть возможность создавать плавный изгиб линии регрессии.

Решение.

1) В ячейки табличного пространства внесём исходные данные, рисунок 3.6.

	A	B
1	X	Y
2	1	2
3	1	3
4	2	4
5	3	5
6	4	5
7	5	7
8	8	14

Рисунок 3.6 – Исходные данные для построения нелинейной регрессии

2) Жмём "Вставка".

3) На вставке "Диаграммы" жмём "Точечная", жмём "Точечная с маркерами". Появляется система координат и корреляционное поле – 7 точек-магазинов, рисунок 3.7.

4) Жмём левой клавишей мыши на любую из 7-ми точек, все точки становятся "крестиками".

5) Наводим курсор на любой из "крестиков" и жмём правую клавишу. Раскрывается меню, жмём "Добавить линию тренда". Возникает меню "Параметры линии тренда".

б) Жмём на параметры меню:

- Полиномиальная, Степень 2;
- другое, Магазины;
- Прогноз вперёд на 3 (для нашего варианта: $X_8 = 11, 11-8 = 3$);
- Прогноз назад на 1 (для нашего варианта: в начало координат);
- Пересечение линии с осью OY;
- Показывать уравнение на диаграмме;
- Поместить на диаграмму коэффициент детерминации R^2 .

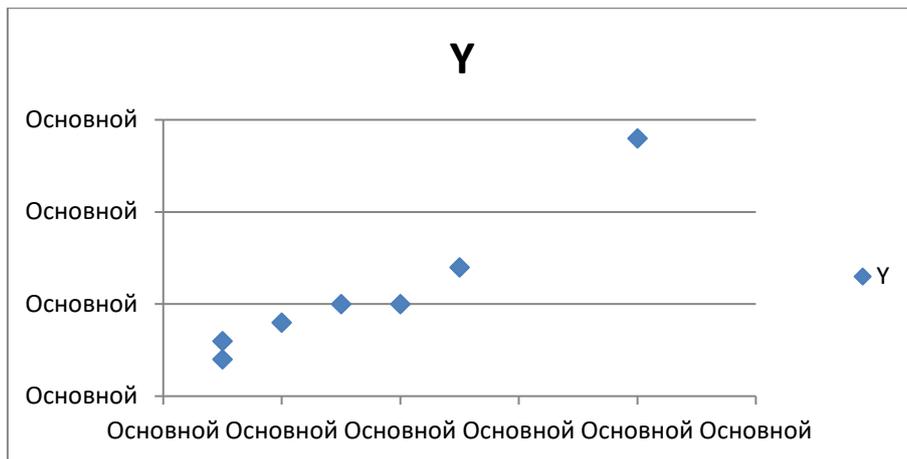


Рисунок 3.7 – Корреляционное поле из семи точек-магазинов

7) Закрываем меню. Появляется диаграмма, рисунок 3.8.

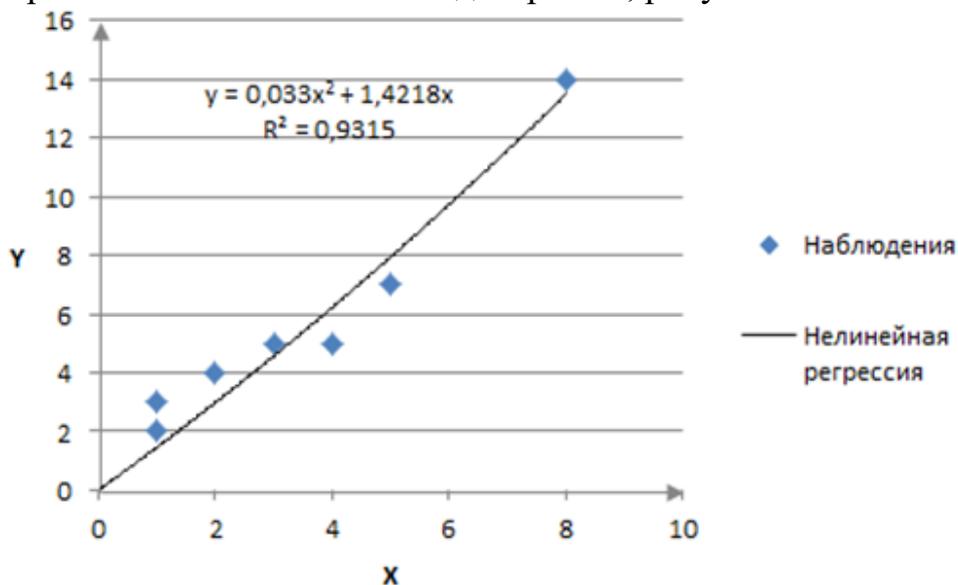


Рисунок 3.8 – Диаграмма с квадратичной линией регрессии

8) Анализ полученной диаграммы.

- Значительной разницы между прямой линией на рисунке 1.1 и квадратичной линией на рисунке 3.8 нет.
- Прогноз для магазина-8 $\hat{y}(x = 11)$: на линейной модели 17,37; на нелинейной 19,64.
- Коэффициент детерминации R^2 : на линейной модели 0,94; на нелинейной 0,93.
- Отношение коэффициентов в квадратичном уравнении: $1,422/0,033 = 43$ – коэффициент при X в 43 раза больше, чем коэффициент при X^2 .

Вывод: предпочтительнее нелинейная модель, но можно использовать и линейную, у неё есть свои достоинства.

4 Методические указания по подготовке к промежуточной аттестации - зачёту

Промежуточная аттестация в форме зачета проходит по результатам прохождения всех видов текущего контроля успеваемости. В качестве оценочных средств для текущего контроля успеваемости используются различные задания, выполняемые студентами на учебных занятиях и во время самостоятельной учебной работы. Сюда относятся:

- а) тестовые задания;
- б) задания по темам практических занятий;
- в) задания по контрольным работам по всем формам обучения.

Последней группе заданий при подготовке к зачёту следует уделить особое внимание, поскольку, во-первых, выполнение и защита контрольной работы предусмотрены учебными планами по всем формам обучения, во-вторых, контрольная работа представляет собой целостное исследование, которое проходит через основные темы дисциплины, и включает в себя решение многих эконометрических задач, в том числе на компьютере.

Вопросы для зачёта и для контрольных работ № 1.

Тема 1. Основные понятия эконометрики

- 1) История возникновения эконометрики.
- 2) Предмет и метод эконометрики.
- 3) Две модели предприятия (магазина) в форме линейной регрессии (сквозные задача-1 и задача-2).
- 4) Модель прогнозирования банкротства предприятия в форме линейной регрессии.
- 5) Дискриминантный анализ для модели прогнозирования банкротства предприятия.
- 6) Методология и этапы эконометрического моделирования.
- 7) Случайные величины и их числовые характеристики.
- 8) Функция распределения случайной величины. Непрерывные случайные величины, законы распределения: равномерный и нормальный.
- 9) Случайные величины и законы распределения (ЗР) непрерывных случайных величин: ЗР χ^2 (хи-квадрат), ЗР Стьюдента, ЗР Фишера-Снедекора.
- 10) Точечные и интервальные оценки параметров.

Тема 2. Парный регрессионный анализ

- 11) Метод наименьших квадратов.
- 12) Связь коэффициентов регрессии и корреляции.
- 13) Пять предпосылок для парного регрессионного анализа.
- 14) Точечные оценки параметров и теорема Гаусса-Маркова.

- 15) Доверительный интервал для регрессии $M_x(Y)$ и линии границ.
- 16) Доверительный интервал для индивидуального прогноза и линии границ.
- 17) Доверительные интервалы для коэффициентов регрессии и остаточной дисперсии.
- 18) Оценка значимости уравнения парной регрессии по критериям Стьюдента и Фишера-Снедекора.
- 19) Оценка значимости уравнения парной регрессии по коэффициентам детерминации: простому и скорректированному.

Тема 3. Множественный регрессионный анализ

- 20) Применение фиктивных переменных в регрессионном анализе.
- 21) Классическая нормальная линейная модель множественной регрессии.
- 22) Шесть предпосылок для классического линейного регрессионного анализа.
- 23) Оценка параметров классической нормальной линейной модели множественной регрессии методом наименьших квадратов.
- 24) Оценки ковариационной матрицы для случайного вектора b и дисперсии отклонений.
- 25) Доверительные интервалы для параметров регрессии.
- 26) Доверительные интервалы для условного математического ожидания $M_x(Y)$.
- 27) Доверительные интервалы для индивидуального прогноза.
- 28) Оценка значимости множественной регрессии по трём критериям.
- 29) Мультиколлинеарность и отбор значимых факторов по корреляционной матрице.
- 30) Мультиколлинеарность и отбор значимых факторов "методом вращения".

Тема 4. Нелинейный регрессионный анализ

- 31) Нелинейные модели регрессии, их классификация и линеаризация.
- 32) Особенности применения МНК-метода для моделей, нелинейных по параметрам и внутренне нелинейных.
- 33) Линейные модели с фиктивными переменными (на примерах).
- 34) Нелинейные модели с фиктивными переменными (на примере оценки рыночной стоимости квартир).
- 35) Нелинейные производственные функции Кобба-Дугласа.
- 36) Пример функции Кобба-Дугласа в рыбной отрасли.
- 37) Пример функции Кобба-Дугласа как модели магазина (методика применения).

5 Методические указания по выполнению самостоятельной работы по дисциплине – части 1

Самостоятельная работа студента – главная форма овладения знаниями. Для дисциплины "Эконометрика" эта работа выстраивается, главным образом, вокруг выполнения контрольной работы № 1 для всех форм обучения. Контрольная работа охватывает 70 % теоретического материала, выполняется по индивидуальным исходным данным.

Примерно половину контрольной работы студенты-очники самостоятельно (с консультационной поддержкой) выполняют на практических занятиях в рамках соответствующих тем. Оставшуюся половину они готовят вне аудитории полностью самостоятельно. Сюда включается не только решение задач, но и подготовка ответов на два теоретических вопроса.

Помимо этого, самостоятельная работа выполняется в рамках подготовки к аудиторным практическим занятиям.

Студенты очно-заочной формы обучения (и для них учебным планом установлено выполнение контрольных работ) также начинают выполнять контрольную работу № 1 на практических занятиях, а затем окончательно выполняют её уже вне аудитории самостоятельно. Помимо контрольной работы, студенты очно-заочной формы обучения имеют возможность самостоятельно работать с тестами, контрольными вопросами при подготовке к сдаче зачёта.

ЧАСТЬ 2

6 Тематический план части 2 дисциплины и методические указания по её изучению

Тема 5. Инструментальные методы регрессионного анализа

Форма занятий – лекция.

Вопросы для изучения:

Вопрос 1. Критерий Чоу объединения двух выборок.

Вопрос 2. Стандартизированные коэффициенты регрессии.

Вопрос 3. Частные коэффициенты корреляции между факторными переменными.

Вопрос 4. Коэффициент ранговой корреляции Спирмена.

Методические указания по изучению темы 5

Цель изучения темы: овладение отдельными методами, которые способствуют построению эффективных моделей и усилению их аналитических возможностей. Для облегчения усвоения этого материала рекомендуем принять на вооружение такую парадоксальную точку зрения Ж. Адамара:

"Задача математики состоит в том, чтобы узаконить интуицию, и никакой другой задачи у неё нет".

Парадоксальность этого высказывания мы видим в том, что Адамар соединяет два разных начала: строгую логику математики и интуицию, которая присуща скорее гуманитарному знанию и искусству.

Вопрос 1. Критерий Чоу объединения двух выборок.

Рассмотрим пример. В организации изучается зависимость размера среднемесячной зарплаты сотрудников Y от их образования X (лет). Всего в организации работают n_1 мужчин и n_2 женщин. Вопрос: можно ли выборки по мужчинам Y_1 и X_1 объединить с выборками по женщинам - Y_2 и X_2 ?

Критерий (метод) Чоу предусматривает построение трёх регрессионных моделей на основе трёх пар выборок: смешанных, отдельно для мужчин, отдельно для женщин. Результатами работы с этими моделями должны служить три набора значений отклонений e .

Критерий Чоу: выборки с уровнем значимости α нельзя объединять, если выполняется неравенство:

$$F = \frac{\sum_{i=1}^n e_i^2 - \sum_{i=1}^{n_1} \gamma_i^2 - \sum_{i=n_1+1}^n \mu_i^2}{\sum_{i=1}^{n_1} \gamma_i^2 + \sum_{i=n_1+1}^n \mu_i^2} \cdot \frac{(n-2p-2)}{(p+1)} > F_{\alpha, p+1, n-2p-2}, \quad (6.1)$$

где n_1 - число работников-мужчин;

n_2 - число работников-женщин;

n - $n = n_1 + n_2$ общее число работников;

p - число факторных переменных в регрессиях;

e_i - остатки, полученные от регрессии на объединённых выборках;

γ_i - остатки, полученные от регрессии на выборках по мужчинам;

μ_i - остатки, полученные от регрессии на выборках по женщинам.

Вопрос 2. Стандартизированные коэффициенты регрессии.

При исследовании экономического явления с помощью модели множественной регрессии часто бывает необходимым сравнить влияние объясняющих переменных на зависимую переменную. Иногда это можно сделать прямо по значениям коэффициентов регрессии b_i . Однако они, как правило, имеют разную размерность, и их влияние трудно сопоставить, поэтому используют стандартизированные коэффициенты регрессии b_j' , которые являются одновременно парными коэффициентами корреляции:

$$r_{x_j y} = b_j' = b_j s_{x_j} / s_y, \quad (6.2)$$

где s_{x_j} и s_y – СКО переменных x_i и y соответственно.

Стандартизированный коэффициент регрессии b_j' показывает, на сколько величин s_y изменится в среднем зависимая переменная Y при увеличении только j -й объясняющей переменной на s_{x_j} . Эти коэффициенты безразмерны.

С помощью стандартизированных коэффициентов можно, например, трёх-факторную стохастическую модель с переменными в отклонениях и относительных величинах записать так:

$$Y = r_{x_1 y} \cdot X_1 + r_{x_2 y} \cdot X_2 + r_{x_3 y} \cdot X_3 + \varepsilon,$$

где $r_{x_i y}$ - коэффициент корреляции между фактором x_i и результирующей переменной y ;

$Y = (y - \bar{y}) / s_y$ – результирующая переменная в отклонениях и в относительной форме;

$X_i = (x_i - \bar{x}_i) / s_{x_i}$ - факторная переменная в отклонениях и в относительной форме.

Вопрос 3. Частные коэффициенты корреляции между факторными переменными.

Обращаем внимание студентов на то, что в этом вопросе рассматривается "глубинный" анализ взаимосвязи переменных. При этом техника анализа не очень сложна. Более подробно этот вопрос изложен в работе [2, с. 64-65].

Часто возникает необходимость исследовать корреляцию между переменными при исключении (элиминировании) влияния остальных переменных. Такая корреляция и коэффициент называются частными. Выборочный частный

коэффициент корреляции между переменными X_i и X_j при фиксированных остальных $p-2$ переменных определяется выражением:

$$r_{i-j,1,2,\dots,p} = \frac{q_{ij}}{\sqrt{(q_{ii}q_{jj})}}, \quad (6.3)$$

где q_{ij} - алгебраическое дополнение элемента r_{ij} матрицы коэффициентов корреляции.

Для частного случая ($p = 3$) матрица коэффициентов корреляции q может иметь такой вид:

$$R_{p=3} = \begin{bmatrix} 1 & r_{12} & r_{13} \\ r_{21} & 1 & r_{23} \\ r_{31} & r_{32} & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0,7 & 0,6 \\ 0,7 & 1 & 0,5 \\ 0,6 & 0,5 & 1 \end{bmatrix} \quad (6.4)$$

Как и обычный частный коэффициент, принимает значения от -1 до +1.

Вопрос 4. Коэффициент ранговой корреляции Спирмена.

Коэффициент ранговой корреляции применяется при наличии атрибутивных (качественных) переменных [2, с. 43-44].. Например, переменная "высшее образование" может принимать значения: бакалавр, специалист, магистр. Для оценки связи между переменными объекты упорядочивают (ранжируют) по степени выраженности признаков переменных. Формула для расчёта коэффициента ранговой корреляции Спирмена:

$$\rho_{x,y} = 1 - \frac{6\sum d_i^2}{n(n^2-1)}, \quad (6.5)$$

где d_i - разность между рангами r_i и s_i значений по переменным X и Y i -го объекта;

n - число объектов, пар наблюдений.

Если невозможно найти существенные различия между объектами (отношение тождества " \equiv "), то им присваивают одинаковые *средние* ранги. Например, два "равных" объекта претендуют на третье место - ранг "3", в таком случае каждому из них присваивают средний ранг $(3+4)/2 = 3,5$.

Методические материалы по теме 5

Учебник [1, Кремер-Путко], Учебник [2, Настин], [3, Настин-Тылик] – пособие по практическим занятиям.

ФОС в ЭИОС. В ходе работы по теме студенту следует использовать лекционный материал; материалы, полученные в ходе практических занятий, материалы в разделе дисциплины в ЭИОС КГТУ, а главное – рекомендованные здесь источники.

Ссылки на рекомендуемые источники по теме 5:

[1, Кремер-Путко, с. 57, 78-79, 122-123, 128-129] – базовый учебник, которого мы будем придерживаться, в том числе в терминологии, обозначениях.

[2, Настин, с. 43, 52, 64-65] – основной учебник.

[4, Карлов, с. 156-193] – учебное пособие, здесь основы математической статистики.

Тема 6. Регрессионные модели временных рядов и прогнозирование

Форма занятий – лекция.

Вопросы для изучения:

Вопрос 1. Структура и классификация временных рядов.

Вопрос 2. Автокорреляционная функция временного ряда.

Вопрос 3. Моделирование временного ряда.

Вопрос 4. Сглаживание временного ряда.

Вопрос 5. Методология и методы прогнозирования.

Методические указания по изучению темы 6

Цель изучения темы – получить представление о важнейшем разделе эконометрики, в теоретическом плане и с точки зрения его практики применения. Действительно, всё течёт и всё меняется, образуя, таким образом, поток временных рядов. И пусть читателей вдохновляет позиция монаха, профессора богословия Роджера Бэкона, высказанная им ещё в **XIII веке**, 1220-1292.

"Устройство нашего мира непостижимо без знания математики".

Вопрос 1. Структура и классификация временных рядов.

Временным (динамическим) рядом называется набор значений (уровней) переменной, измеренной в последовательные моменты времени: y_t ($t=1, 2, \dots, n$), где n - длина ряда. В структуре экономических временных рядов в общем случае выделяют четыре составляющих (слагаемых) рядов (в разных случаях смысл составляющих может быть разным):

$$y_t = u_t + v_t + c_t + \varepsilon_t, \quad (6.6)$$

где u_t - тренд, отражает изменение явления (процесса), происходящее в течение значительного промежутка времени, периодичность отсутствует;

v_t - компонента отражает цикличность с большими периодами T_1 (до десятков лет).;

c_t - компонента отражает цикличность с периодами T_2 до нескольких лет;

ε_t - компонента случайная, отражает влияние факторов с небольшими периодичностями T ; иногда это называют помехой, шумом.

Первые три составляющие ряда рассматриваются как *неслучайные*, поэтому полной задачей анализа временного ряда является выявление всех этих

трех закономерностей. Частными задачами может быть выявление какой-либо пары закономерностей или даже одной.

Ряды не всегда содержат все три неслучайных составляющих, либо какие-то из них пренебрежимо малы, либо их оценка невозможна ввиду ограниченности статистического материала. Классической задачей анализа временных рядов является выявление одной главной закономерности - тренда.

Временные ряды (процессы) делятся на стационарные и нестационарные. Временной ряд y_t , $t \in 1:n$, называется *строго стационарным* (стационарным в узком смысле), если совместное распределение вероятностей n наблюдений $y_{1+\tau}$, $y_{2+\tau}$, ..., $y_{n+\tau}$ не зависит от n , t и временного сдвига τ . Поэтому такие характеристики распределения, как математическое ожидание $a_y(t) = a$, среднеквадратичное отклонение $\sigma_y(t) = \sigma$ можно оценить на основе наблюдений y_t , $t \in 1:n$, по формулам:

$$\bar{y}_t = \sum \frac{y_t}{n}, \quad s_t^2 = \frac{\sum (y_t - \bar{y}_t)^2}{n}. \quad (6.7)$$

Классическим стационарным рядом в узком смысле является случайный процесс типа "белый шум", у которого математическое ожидание $a_y(t) = 0$, а отклонения ε_t не коррелированы между собой. Отклонения ε_t в классической линейной регрессионной модели образуют белый шум, а в случае их нормального распределения - нормальный белый шум.

Временной ряд y_t , $t \in 1:n$, называется *стационарным в широком смысле*, если лишь числовые характеристики совместного распределения вероятностей n наблюдений $y_{1+\tau}$, $y_{2+\tau}$, ..., $y_{n+\tau}$ не зависят от n , t и временного сдвига τ .

При нарушении условия постоянства числовых характеристик временной ряд называется *нестационарным*. Он отражает переход из одного устойчивого состояния в другое - развитие (или деградация), у него есть начало и конец.

Вопрос 2. Автокорреляционная функция временного ряда.

Слово "авто" здесь означает "с самим собой" (авто-биография - сам о себе, авто-мобиль - везу сам себя без лошадей). Идея функции очень проста: делается копия временного ряда, и эти два ряда сдвигаются относительно друг друга на временные отрезки $\tau = 0, 1, 2, \dots$; делается вертикальное сечение и рассчитываются коэффициенты корреляции для $\tau = 0, 1, 2, \dots$. Получается корреляционная функция $r = r(\tau)$.

Автокорреляционная функция r является важной характеристикой временных рядов. Для стационарных рядов она зависит только от одной переменной - временного сдвига (лага) τ , для нестационарных - от двух переменных: сдвига

(лага) τ и момента начала отсчета t . Функция определяется по аналогии с коэффициентом корреляции:

$$\rho(\tau) = \frac{M[(y_t - a)(y_{t+\tau} - a)]}{\sigma_x(t) \cdot \sigma_x(t + \tau)} = \frac{M[(y_t - a)(y_{t+\tau} - a)]}{\sigma^2}. \quad (6.8)$$

Статистической оценкой для $\rho(\tau)$ является выборочная автокорреляционная функция $r(\tau)$. График этой функции называют также коррелограммой:

$$r(\tau) = \frac{(n - \tau) \sum_{t=1}^{n-\tau} y_t y_{t+\tau} - \sum_{t=1}^{n-\tau} y_t \sum_{t=1}^{n-\tau} y_{t+\tau}}{\sqrt{(n - \tau) \sum_{t=1}^{n-\tau} y_t^2 - (\sum_{t=1}^{n-\tau} y_t)^2} \sqrt{(n - \tau) \sum_{t=1+\tau}^n y_{t+\tau}^2 - (\sum_{t=1+\tau}^{n-\tau} y_{t+\tau})^2}}. \quad (6.9)$$

Рассмотрим свойства автокорреляционной функции (рисунок 6.1):

- $\rho(\tau) = \rho(-\tau)$ - функция чётная;
- $\rho(\tau=0) = 1$ - корреляция случайной величины y_t ($t=1, 2, \dots, n$) сама с собой равна 1;
- при $\tau \rightarrow \infty$ $|\rho| \rightarrow 0$, при большом сдвиге τ корреляция между значениями временного ряда падает до нуля, рисунок 6.3.

Значения $r(\tau = 1)$, $r(\tau = 2)$, ... называются коэффициентами корреляции 1-го порядка, 2-го порядка и т. д.

Автокорреляционная функция:

- дает представление о характере временного ряда; например, для белого шума она будет иметь вид остроконечного импульса; для временного ряда типа синусоиды она будет иметь вид синусоиды;
- помогает подобрать наиболее подходящую функцию регрессии, например, включить в неё сезонную синусоидальную составляющую.

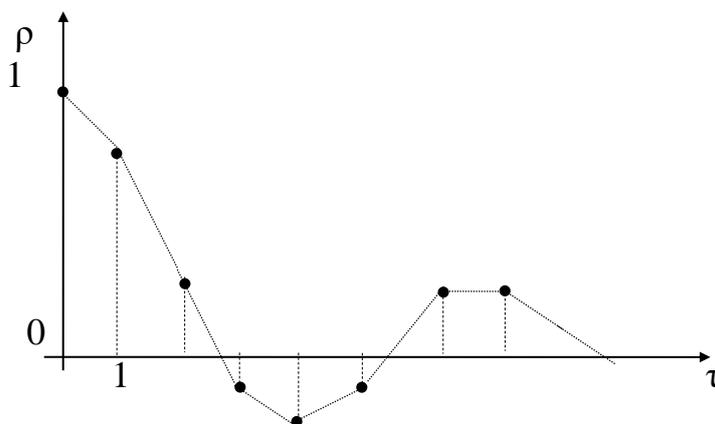


Рисунок 6.1 - Типичная автокорреляционная функция (коррелограмма) временного ряда с циклической составляющей

Вопрос 3. Моделирование временного ряда.

Важнейшая задача моделирования временного ряда - выявление в нем основной тенденции - сглаживающей функции $f(t)$. Для её выбора используются, по меньшей мере, *три типа информации*.

Первый тип - содержательная информация. Здесь строится качественная теория изучаемого процесса. Выдвигается гипотеза о причинах именно такого поведения процесса в прошлом и настоящем и о поведении в будущем.

Второй тип - графическая информация: временной ряд представляется в виде графика. Соответственно ему подбирается функция $f(t)$ – линейная, квадратичная и т. п.

Третий тип информации – аналитическая. На основе содержательной и графической информации выбираются несколько функций, для каждой из которых строится регрессионное уравнение и рассчитываются показатели качества сглаживания. Далее решается задача выбора наилучшей сглаживающей функции $f(t)$, например:

линейная $f(t) = b_0 + b_1 t$; квадратичная $f(t) = b_0 + b_1 t + b_2 t^2$; логистическая $f(t) = a / (1 + b e^{-ct})$.	(6.10)
-------------------------------------------------------------------------------------------------------------------------------	--------

Метод наименьших квадратов. Для сглаживания и прогнозирования этот метод используется чаще всего. Здесь сначала выбирают вид функции на основе компромисса между несколькими критериями. После этого применяется МНК-метод; роль факторной переменной играет время $t \in 1:n$. Получаем стохастическую регрессионную модель:

$$y_t = f(t) + e_t, \quad (6.11)$$

где e_t – отклонения.

Методы скользящих средних (МСС). Обычно они используются только для сглаживания временного ряда, реже – для прогнозирования. Их достоинства по сравнению с МНК:

- 1) простое регулирование степени сглаживания;
- 2) гарантированное сохранение "характера" первичного временного ряда;
- 3) возможность на заключительном этапе присоединить ещё и МНК.

Вопрос 4. Сглаживание временного ряда методами скользящих средних (МСС).

В основе МСС-методов лежит переход от исходных значений членов ряда к их средним значениям на очередном рабочем интервале времени – окне с шириной m . Методы делятся по способу сглаживания (усреднения).

1) *МСС-метод на основе простого среднего арифметического, с нечётной шириной окна:* $m = 2k + 1$. Здесь вычисленное среднее определяет значение тренда в средней $(k+1)$ -й точке этого окна.

Пример. Для моментов времени $t = \{1, 2, 3, 4, 5, 6\}$ дан временной ряд $y = \{5, 7, 11, 12, 13, 16\}$. Выполнить сглаживание методом скользящих средних, используя простую среднюю арифметическую с интервалом сглаживания $m = 3$ (здесь $k = 1$). Вычислим все значения сглаженного ряда z (значения z_1 и z_6 не существуют):

$$z_2 = (5+7+11)/3 = 7,7; z_3 = (7+11+12)/3 = 10,0; z_4 = (11+12+13)/3 = 12,0 \text{ и т. д.}$$

МСС-метод на основе простой средней арифметической, с чётной шириной окна: $m = 2k$. Здесь среднее определяет значение тренда для промежуточной точки - середины отрезка времени между k -м и $(k+1)$ -м, например, между 2-м и 3-м. Но таких моментов нет во временном ряде, поэтому после нахождения значений тренда во всех промежуточных точках определяют значения тренда для моментов ряда как среднее в соседних промежуточных точках.

Пример. Дан ряд: $t \in 1:6$, $y = \{5, 7, 11, 12, 13, 16\}$, ширина окна $m = 4$ (здесь $k = 2$). Вычислить значения тренда z .

1) Вычислим для промежуточных значений времени: 2,5; 3,5; 4,5 (других значений нет):

$$z_{2,5} = (5+7+11+12)/4 = 8,75; z_{3,5} = (7+11+12+13)/4 = 10,75; z_{4,5} = (11+12+13+16)/4 = 13,0.$$

2) Вычислим значения тренда в заданных моментах времени:

$$t = 3: z_3 = (z_{2,5} + z_{3,5})/2 = (8,75 + 10,75)/2 = 9,75; t = 4: z_4 = (z_{3,5} + z_{4,5})/2 = (10,75 + 13,0)/2 = 11,875.$$

3) *МСС-метод на основе средневзвешенной арифметической при нечётной ширине окна: $2k+1$.* В подборе весов обычно реализуется идея актуальности и старения информации.

Пример. Дан ряд: $t \in 1:6$, $y = \{5, 7, 11, 12, 13, 16\}$, ширина окна $m = 3$. Рассчитать средневзвешенные значения тренда.

1) Брокер на основе своего опыта задаёт веса по росту курсов акций $\{0,2; 0,3; 0,5\}$, их сумма равна 1,0.

2) Рассчитываем средневзвешенные значения тренда:

$$\text{- для } t = 2: z_2 = (0,2 \cdot 5 + 0,3 \cdot 7 + 0,5 \cdot 11) = 8,6;$$

$$\text{- для } t = 3: z_3 = (0,2 \cdot 7 + 0,3 \cdot 11 + 0,5 \cdot 12) = 10,7; \text{ и т.д. для } t = 4 \text{ и } t = 5.$$

4) *МСС-метод на основе экспоненциального сглаживания (метод Брауна).*

Метод состоит в расчёте экспоненциальных средних z_t по рекуррентной формуле (значение "на сегодня" рассчитывается через значение "на вчера"):

$$z_t = \alpha \cdot y_t + (1 - \alpha) \cdot z_{t-1}, \quad (6.12)$$

где α – весовой коэффициент (параметр сглаживания), $0 < \alpha \leq 1$.

Если последовательно выразить z_t через z_{t-1} , а потом z_{t-1} через z_{t-2} и так далее, то значение z_t будет выражено через все прошлые значения временного ряда y с соответствующими весами:

$$z_t = \alpha \cdot \sum_{j=0}^{t-1} (1 - \alpha)^j + (1 - \alpha)^t y_0, \quad (6.13)$$

где y_0 – значение, которое характеризует начальное условие; рекомендуется его определять по формуле простого скользящего среднего.

Рекомендуемый диапазон значений α :

$$0,1 \leq \alpha \leq 0,4. \quad (6.14)$$

Напомним: для нахождения значения y_0 рекомендуется использовать формулу простого скользящего среднего с выбранным интервалом сглаживания m .

Вопрос 5. Методология и методы прогнозирования.

Прогноз – это научно обоснованное вероятное суждение о возможных будущих состояниях объекта прогнозирования. Прогнозирование – это процесс разработки механизма прогноза и/или самого прогноза. Объект прогнозирования – совокупность процессов, на которую направлена познавательная деятельность субъекта прогнозирования.

Прогноз характеризуют 7 параметрами.

1) Период упреждения: отрезок времени от настоящего в будущее, на который разрабатывается прогноз.

2) Прогнозный горизонт: максимально возможный период упреждения прогноза – ограничен точностью прогноза.

3) Период основания прогноза: отрезок времени от прошлого к настоящему, на основании которого строится прогноз.

4) Точность прогноза: доверительный интервал прогноза для заданной вероятности его осуществления.

5) Достоверность прогноза: вероятность осуществления прогноза для заданного доверительного интервала.

6) Обоснованность прогноза: степень соответствия методов и исходной информации целям и задачам прогнозирования.

7) Ошибка прогноза: апостериорная величина отклонения прогноза от действительного состояния объекта прогнозирования или путей и сроков его осуществления.

Прогностика выделяет 6 этапов разработки прогноза.

Этап 1 - предпрогнозная ориентация: определяются объект, цели и задачи прогнозирования, периоды основания и упреждения, исполнители, сроки, финансирование. Основной метод этапа – системный анализ. Результат – задание на прогноз.

Этап 2 - прогнозная ретроспекция: собираются исходные данные, изучается история развития объекта прогнозирования и прогнозного фона. Методы - системный анализ, статистика.

Этап 3 - прогнозный диагноз и синтез: диагностируется сущность объекта прогнозирования и прогнозного фона, выбираются и/или разрабатываются

модели и методы прогнозирования, например, авторегрессионное уравнение прогнозирования курса акций: $y_t = \beta_0 + \beta_1 y_{t-1}$ (руб.).

Этап 4 - прогнозная проспекция: разрабатывается собственно прогноз на основании выбранных и/или построенных моделей. Обычно для этого используется компьютер и пакет прикладных программ.

Этап 5 - верификация прогноза: априори оценивается достоверность и точность прогноза и/или его обоснованность.

Этап 6 - корректировка прогноза: прогноз уточняется на основании его верификации и/или дополнительных данных.

Классификация методов прогнозирования. По степени формализации методы делят на фактографические и экспертные. Фактографические методы основаны, в основном, на фактической информации, а экспертные – на опыте и интуиции экспертов.

Фактографические методы делят на методы экстраполяции и методы моделирования (граница между ними условная).

Методы экстраполяции основаны на гипотезе: закономерности прошлого без особых изменений будут действовать и в будущем.

Для прогнозирования в качестве объясняющих переменных естественно использовать лаговые переменные (лаг естественно временной сдвиг); сдвиг в прошлое от момента t : $y_{t-1}, y_{t-2}, \dots, y_{t-p}$. Например, для прогнозирования курса акций можно использовать авторегрессионную модель p -го порядка:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \dots + \beta_p y_{t-p} + \varepsilon_t; t \in 1: n. \quad (6.15)$$

Методические материалы по теме 6

Учебник [1, Кремер-Путко], Учебник [2, Настин], ФОС в ЭИОС. В ходе работы по теме студенту следует использовать лекционный материал; материалы, полученные в ходе практических занятий, материалы в разделе дисциплины в ЭИОС КГТУ, а главное – рекомендованные здесь источники.

Ссылки на рекомендуемые источники по теме 6:

[1, Кремер-Путко, с. 133-148] – это базовый учебник, которого мы будем придерживаться, в том числе в терминологии, обозначениях.

[2, Настин, с. 86-96] – это основной учебник.

[4, Карлов, с.156-193] – учебное пособие, здесь основы математической статистики.

Тема 7. Обобщённая линейная модель и гетероскедастичность пространственной выборки

Форма занятий – лекция.

Вопросы для изучения:

Вопрос 1. Признаки обобщенной линейной модели и четыре предпосылки.

Вопрос 2. Обобщённый метод наименьших квадратов.

Вопрос 3. Сущность и последствия гетероскедастичности.

Вопрос 4. Тест Спирмена на наличие гетероскедастичности.

Вопрос 5. Тест Голдфелда-Квандта на наличие гетероскедастичности.

Вопрос 6. Тест Уайта на наличие гетероскедастичности.

Вопрос 7. Устранение гетероскедастичности взвешенным МНК-методом.

Методические указания по изучению темы 7

Цель изучения темы – получить представление о важнейшем разделе эконометрики, который выводит нас за рамки классического линейного регрессионного анализа. Происходит это в тех случаях, когда одна или несколько предпосылок из шести не выполняются. Всё усложняется, конечно, и тут мы, следуя авторитету и словам А. Эйнштейна, рекомендуем дать волю такому инструменту познания, как воображение:

"Воображение важнее, чем знание".

Вопрос 1. Признаки обобщенной линейной модели и четыре предпосылки.

Коренное отличие *обобщенной* модели от классической состоит только в виде ковариационной (напоминаем – квадратной) матрицы вектора отклонений ε . У классической модели матрица имеет специальный вид $\sum_{\varepsilon} = \sigma^2 E_n$ – см (6.16), у обобщенной модели она имеет произвольные значения ковариаций и дисперсий, другими словами – матрица имеет самый общий вид: $\sum_{\varepsilon} = \Omega$. Пример для сравнения: ковариационные матрицы классической (6.16) и обобщенной (6.17) моделей для двух наблюдений ($n = 2$) будут иметь вид:

$$\sum_{\varepsilon} = \sigma^2 E_n = \sigma^2 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix}, \quad (6.16)$$

$$\sum_{\varepsilon} = \Omega = \begin{pmatrix} M(\varepsilon_1^2) & M(\varepsilon_1 \varepsilon_2) \\ M(\varepsilon_2 \varepsilon_1) & M(\varepsilon_2^2) \end{pmatrix}. \quad (6.17)$$

Обобщенная стохастическая линейная модель множественной регрессии (ОЛММР) в матричной форме внешне имеет такой же вид, как и классическая:

$$Y = X\beta + \varepsilon. \quad (6.18)$$

Четыре предпосылки для обобщённой ЛММР:

1) ε - случайный n -мерный вектор отклонений; X - неслучайная матрица значений объясняющих переменных (матрица плана), её размерность $n \times (p+1)$; напомним, что 1-й столбец этой матрицы состоит из n единиц;

2) $M(\varepsilon) = 0_n$ - математическое ожидание вектора отклонений равно ноль-вектору 0_n ;

3) Матрица $\Sigma_{\varepsilon} = M(\varepsilon\varepsilon') = \Omega$, где Ω - положительно определённая квадратная матрица $n \times n$; (заметим, что произведение векторов $\varepsilon'\varepsilon$ даёт скаляр, а произведение векторов $\varepsilon\varepsilon'$ даёт матрицу размером $n \times n$).

4) Ранг матрицы X равен $p+1$ – это число факторных переменных в модели (вместе с "единичной" переменной X_0), n - число наблюдений (в наших задаче-1 и задаче-2 это 7 - число магазинов).

Следствие 1. Оценка параметров обобщённой модели (6.18) обычным МНК

$$b = (X'X)^{-1}X'Y \quad (6.19)$$

является несмещенной и состоятельной, что хорошо, но неэффективной (неоптимальной в смысле теоремы Гаусса-Маркова – не обладает минимальной дисперсией). Для получения эффективной оценки нужно использовать обобщенный метод наименьших квадратов (ОМНК).

Следствие 2. Для классической модели ковариационная матрица вектора оценок параметров определяется по формуле:

$$\Sigma_b = \sigma^2(X'X)^{-1}. \quad (6.20)$$

Эта же оценка для обобщенной модели является смещенной, следовательно, и неэффективной. Её называют неработоспособной и неприменимой для обобщенной ЛММР.

Следствие 3. Для обобщенной модели ковариационная матрица вектора оценок параметров определяется другой формулой:

$$\Sigma_{b^*} = (X'X)^{-1}X'\Omega X(X'X)^{-1}. \quad (6.21)$$

Вопрос 2. Обобщенный метод наименьших квадратов.

Проблема эффективности линейной несмещенной оценки вектора β для обобщенной ЛММР решается с помощью теоремы Айткена.

Теорема Айткена: в классе линейных несмещенных оценок вектора β для обобщенной ЛММР оценка

$$b^* = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}XY \quad (6.22)$$

имеет наименьшую ковариационную матрицу. При этом математическое ожидание оценки b^* равно β : $M(b^*) = \beta$, т. к. $M(\varepsilon) = 0$ (нет смещения).

Вопрос 3. Сущность и последствия гетероскедастичности.

Формально условие гомоскедастичности записывается в виде матрицы $\Sigma_\varepsilon = \sigma^2 E_n$, где E_n – единичная матрица, а на главной диагонали матрицы Σ_ε всюду *одинаковое* число σ^2 – остаточная дисперсия.

Однако на практике это условие часто не выполняется, и тогда мы имеем дело с гетероскедастичностью. В парной регрессии она может проявляться так: с ростом факторной переменной X растет (либо уменьшается, если обратная зависимость) в среднем значение результирующей переменной Y и одновременно увеличивается разброс точек относительно линии регрессии.

Последствия гетероскедастичности. Пусть для оценки регрессии Y по X_1, \dots, X_p мы применили обычный МНК и получили оценочный вектор b для вектора параметров β : $b = (X'X)^{-1}X'Y$. Если вместо Y подставить его модель $Y = X\beta + \varepsilon$, то после несложных преобразований получим (заметим, что вектор b зависит от случайного вектора ε):

$$b = (X'X)^{-1}X'Y = \beta + (X'X)^{-1}X'\varepsilon. \quad (6.23)$$

Эта оценка – несмещенная и состоятельная для обобщенной линейной модели множественной регрессии, в том числе для случая гетероскедастичности (это очевидно, если учесть $M(\varepsilon) = 0$). Следовательно:

для построения регрессионной модели и использования ее в качестве прогностического инструмента обычный МНК-метод *применим* и в случае гетероскедастичности модели.

Неприятности начинаются, когда мы хотим оценить точность модели, ее значимость, получить интервальные оценки ее коэффициентов. Результаты оказались бы непригодными.

Дело в том, что при расчете критических t - и F -статистик для тестирования гипотез важное значение имеют оценки дисперсий и ковариаций оценок b_i , т. е. ковариационная матрица Σ_b . Если модель не является классической, то ковариационная матрица вектора отклонений $\Sigma_\varepsilon \neq \sigma^2 E_n$, и вместо $\Sigma_b = \sigma^2 (X'X)^{-1}$ мы имеем существенно иную ковариационную матрицу:

$$\Sigma_b = (X'X)^{-1}X'\Omega X(X'X)^{-1}.$$

Добавим, что несмещенная и состоятельная оценка в случае гетероскедастичности не будет оптимальной в смысле теоремы Гаусса-Маркова, т. е. эффективной. Это может привести к тому, что оценка b будет значительно отличаться от истинного значения β .

Вопрос 4. Тест Спирмена на наличие гетероскедастичности.

Для проверки наличия или отсутствия гетероскедастичности используются тесты. Мы рассмотрим три таких теста.

1) Тест ранговой корреляции Спирмена использует следующую идею. Модули отклонений $|e_i|$ являются оценками СКО σ_i , поэтому при гетероскедастичности модули отклонений $|e_i|$ и значения факторов x_i будут коррелированы. Здесь применяется коэффициент ранговой корреляции $\rho_{x,e}$, который рассчитывается по формуле:

$$\rho_{x,e} = 1 - \frac{6\sum d_i^2}{n(n^2-1)}, \quad (6.24)$$

где d_{ii} - разность между рангами значений x_i и e_i i -го объекта;
 n - число объектов, пар наблюдений.

Итак, критерий Спирмена: коэффициент ранговой корреляции значим на уровне значимости α (справедлива гипотеза H_1 , имеет место гетероскедастичность), если выполняется условие:

$$|t| = \frac{|\rho_{x,e}| \sqrt{n-2}}{\sqrt{1-\rho_{x,e}^2}} > t_{1-\alpha, n-2}, \quad (6.25)$$

где $t_{1-\alpha, n-2}$ - табличное значение t -критерия Стьюдента с уровнем значимости α и числом степеней свободы $(n-2)$.

Вопрос 5. Тест Голдфелда-Квандта на наличие гетероскедастичности.

Этот тест можно применять тогда, когда ошибки регрессии имеют нормальный закон распределения. Пусть средние квадратичные отклонения σ_i ошибок пропорциональны значениям переменной X . Упорядочим наблюдения в порядке возрастания X и выберем m первых и m последних наблюдений (две m -выборки).

Гипотеза H_0 о наличии гомоскедастичности будет равносильна утверждению о том, что две нормально распределенные m -выборки имеют одинаковые дисперсии (разность дисперсий между ними равна нулю). Гипотеза о равенстве дисперсий отвергается (гетероскедастичность), если выполняется неравенство (в числитель помещается бóльшая сумма):

$$F = \frac{\sum_{i=1}^m e_i^2}{\sum_{i=n-m+1}^n e_i^2} > F_{\alpha, m-p, m-p}. \quad (6.26)$$

Мощность теста оказывается максимальной, если выбирать длину m -выборки $m \approx n/3$. Например, для примера с семью магазинами можно взять две m -выборки с индексами 1-3 и 5-7. Здесь $m_{\text{факт}} = 3$, хотя рекомендованное значение $m = 7/3 = 2,3 \dots \approx 2$. Следует взять выборки с индексами 1-2 и 6-7.

Вопрос 6. Тест Уайта на наличие гетероскедастичности.

Этот тест, в отличие от двух рассмотренных выше, позволяет проследить количественный характер зависимости дисперсий отклонений регрессии от значений факторов. Для этого предполагается, что существует функция, которая свяжет дисперсии отклонений с факторами. Обычно в качестве таковой берется квадратичная функция:

$$\sigma_i^2 = f(X_i) = \sum_{j=1}^p g_j x_{ij} + \sum_{k=1}^{p-1} \sum_{j=1}^{p-k} h_{jk} x_{ij} x_{ij+k}, \quad i = 1, \dots, n. \quad (6.27)$$

Здесь первая сумма – это линейная составляющая $f(X_i)$, а вторая сумма – квадратичная составляющая, она включает в себя квадраты факторов и их произведения. Простой пример квадратичной функции:

$$\sigma_i^2 = ax_1^2 + bx_2^2 + cx_1x_2 + dx_1 + g.$$

После того как выбран вид функции $f(X_i)$, она оценивается с помощью регрессионной модели для квадратов отклонений:

$$e_i^2 = f(X_i) + u_i, \quad i \in 1: n, \quad (6.28)$$

где u_i – отклонение в i -м наблюдении.

Гипотеза об отсутствии гетероскедастичности - условие $f(X_i) = \text{const}$ - принимается, если регрессия (6.28) незначима (нет зависимости).

Вопрос 7. Устранение гетероскедастичности *взвешенным МНК-методом*.

Пусть критерий показал, что модель

$$Y = X\beta + \varepsilon \quad \text{или} \quad y_i = \beta_0 + \sum \beta_i x_{ij} + \varepsilon_i \quad \forall i \in 1: n \quad (6.29)$$

гетероскедастична: дисперсии отклонений σ_i^2 не равны между собой, но отклонения ε_i и ε_k некоррелированы. Отсюда следует, что ковариационная матрица $\sum_{\varepsilon} = \Omega$ вектора отклонений диагональная:

$$\Omega = \begin{pmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sigma_n^2 \end{pmatrix}. \quad (6.30)$$

Если дисперсии отклонений σ_i^2 известны, то гетероскедастичность можно устранить. На практике, однако, эти значения почти никогда не бывают известны. Если исходить из предположения (6.32), то в качестве состоятельных оценок σ_i^2 берутся прогнозные значения e_i^2 регрессии (6.33).

Рассмотрим алгоритм оценки параметров взвешенным МНК-методом.

1) Проверить с помощью одного или нескольких тестов факт наличия в исходной модели (6.34) гетероскедастичности.

2) Применить к модели (6.34) обычный МНК-метод, найти оценочный вектор коэффициентов \mathbf{b} , вычислить квадраты ошибок e_i^2 .

3) Найти регрессию квадратов ошибок на квадратичные функции регрессоров, т.е. найти регрессионную модель (6.33).

4) Вычислить по уравнению (6.33) прогнозные значения \hat{e}_i^2 и получить набор весов $\hat{\sigma}_i = \sqrt{\hat{e}_i^2}$.

5) Ввести новые переменные $X_j^* = \frac{X_j}{\hat{\sigma}_i}$ ($j=1, \dots, p$), $Y_j^* = \frac{Y_j}{\hat{\sigma}_i}$ и записать уравнение регрессии $\hat{Y}^* = X^* \mathbf{b}$, которое гомоскедастично.

Полученная при этом оценка $\mathbf{b}^* = (X^{*'} X^*)^{-1} X^{*'} Y^*$ и есть оценка **взвешенным методом** наименьших квадратов исходной регрессионной модели.

Подведём итоги. Неизвестные параметры уравнения регрессии находим следующим образом:

- если применяем *обычный МНК-метод*, то минимизируем остаточную сумму квадратов:

$$S = \mathbf{e}' \mathbf{e} = \sum (\hat{y}_i - y_i)^2; \quad (6.31)$$

- если применяем *обобщенный МНК-метод*, то минимизируем выражение:

$$S = \mathbf{e}' \Omega^{-1} \mathbf{e}; \quad (6.32)$$

- если применяем *взвешенный МНК-метод*, то минимизируем взвешенную остаточную сумму квадратов:

$$S = \sum (\hat{y}_i - y_i)^2 / \hat{\sigma}_i. \quad (6.33)$$

В последнем случае мы добиваемся равномерного вклада остатков в общую сумму, что и приводит к получению наиболее эффективных оценок параметров модели.

Методические материалы по теме 7

Учебник [1, Кремер-Путко], Учебник [2, Настин], [3, Настин-Тылик] – пособие по практическим занятиям.

ФОС в ЭИОС. В ходе работы по теме студенту следует использовать лекционный материал; материалы, полученные в ходе практических занятий, материалы в разделе дисциплины в ЭИОС КГТУ, а главное – рекомендованные здесь источники.

Ссылки на рекомендуемые источники по теме 7:

[1, Кремер-Путко, с. 150-166] – базовый учебник, которого мы будем придерживаться, в том числе в терминологии и обозначениях.

[2, Настин, с. 77-83] – основной учебник.

[4, Карлов, с. 156-193] – учебное пособие, здесь основы математической статистики.

Тема 8. Обобщённая линейная модель и автокорреляция остатков

Форма занятий – лекция.

Вопросы для изучения:

Вопрос 1. Автокорреляция остатков временного ряда.

Вопрос 2. Тест Дарбина-Уотсона на наличие автокорреляции между соседними наблюдениями.

Вопрос 3. Тест Бреуша-Годфри на наличие автокорреляции между соседними и отдалёнными наблюдениями.

Вопрос 4. Устранение автокорреляции.

Методические указания по изучению темы 8

Цель изучения темы – освоение новых разделов неклассического линейного регрессионного анализа. Другими словами, мы продолжим углублять наши знания об эконометрическом - математическом - моделировании. Да, при этом неизбежны упрощения, и пусть поддержкой служит высказывание знаменитого философа и математика Бертрана Рассела:

"Всякая точная наука основывается на приближительности".

Вопрос 1. Автокорреляция остатков временного ряда.

При рассмотрении временного ряда отчётливо видно влияние предыдущих результатов наблюдений на последующие. Примером может служить курс акций ПАО "Газпром" вчера и сегодня. Математически эта связь выражается в том, что случайные величины – отклонения ε_i и ε_j регрессионной модели - зависимые, условие $r(\varepsilon_i, \varepsilon_j) = 0$ не выполняется. В качестве причин такого "не белого шума" можно назвать экономическую устойчивость и инерционность. Такие модели называются моделями с автокорреляцией.

Соответствующая регрессионная модель является обобщённой, и применяя к ней обычный МНК-метод, мы получим несмещённые и состоятельные оценки - это хорошо, но они неэффективные – а это плохо.

При наличии автокорреляции наибольшее влияние оказывает предыдущий результат измерения. Поэтому отсутствие корреляции между соседними членами позволяет сделать вывод об отсутствии корреляции в целом.

Вопрос 2. Тест Дарбина-Уотсона на наличие автокорреляции между соседними наблюдениями.

Этот тест основан на гипотезе о том, что при наличии автокорреляции наибольшее влияние на последующее значение оказывает предыдущее значение. Поэтому отсутствие корреляции между соседними членами ряда может служить хорошим критерием того, что корреляция отсутствует вообще.

Критическая статистика Дарбина-Уотсона рассчитывается по формуле:

$$d = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2}. \quad (6.34)$$

Статистика d связана с выборочным коэффициентом корреляции и тем точнее, чем больше длина выборки n :

$$d \approx 2(1-r). \quad (6.35)$$

Анализ связи r и d :

- при $r = -1$ имеем $d = 4$ (отрицательная автокорреляция);
- при $r = 0$ имеем $d = 2$ (отсутствие автокорреляции);
- при $r = +1$ имеем $d = 0$ (положительная автокорреляция).

На рисунке 6.6 изображена связь значения d -статистики с результатами тестирования для $n > 14$ (d_n и d_b - нижняя и верхняя границы - находятся по таблице, Приложение Ж):

- 1) $d_n < d < 4 - d_b$ - гипотеза H_0 не отвергается (принимается);
- 2) $d_n < d < d_b$ - гипотеза H_0 не отвергается и не принимается - область неопределенности критерия;
- 3) $0 < d < d_n$ - принимается альтернативная гипотеза о положительной автокорреляции;
- 4) $4 - d_n < d < 4 - d_b$ - принимается альтернативная гипотеза об отрицательной автокорреляции.

0	d_n	d_b	2	$4 - d_b$	$4 - d_n$	4
H_0 отвергается, H_1 - положительная ав- токорреляция		Зона-1 неопреде- лённости	H_0 принимается - отсутствие автокорреляции	Зона-2 неопреде- лённости	H_0 отвергается, H_1 - отрицательная ав- токорреляция	

Рисунок 6.2 - Схема теста (шкалы) Дарбина-Уотсона

Два недостатка теста Дарбина-Уотсона: 1) наличие двух зон неопределенности критерия; 2) корреляция учитывается только между двумя соседними членами ряда отклонений.

Вопрос 3. Тест Бреуша-Годфри на наличие автокорреляции между соседними и отдалёнными наблюдениями.

Этот тест основан на учёте корреляции между соседними наблюдениями авторегрессионным уравнением первого порядка:

$$e_t = \rho e_{t-1}, t \in 1:n, \quad (6.36)$$

где e_t - отклонения регрессии, полученные обычным МНК;

ρ - коэффициент регрессии.

Суть теста состоит в следующем: если коэффициент регрессии ρ значимо отличается от нуля, то уравнение значимо, автокорреляция имеет место.

Два преимущества этого теста перед тестом Дарбина-Уотсона:

- он не содержит зон неопределенностей;
- он может измерять наличие автокорреляции не только между смежной парой отклонений, но и с учетом других отклонений: e_{t-2}, e_{t-3} . Для этого в уравнение (6.40) нужно ввести эти факторы, например, так:

$$e_t = \rho_1 e_{t-1} + \rho_2 e_{t-2}.$$

Существуют и другие тесты.

Вопрос 4. Устранение автокорреляции.

Автокорреляция устраняется, если мы качественно выполним идентификацию временного ряда, т. е. построим модель такую, в которой остатки представляют собой "белый шум", отсутствует мультиколлинеарность и все факторы значимы.

Такая модель не единственная, поэтому выбирают одну из трех типов. Рассмотрим их (расшифровку аббревиатур AR, MA, ARMA см. в предисловии).

1) Авторегрессионная модель p -го порядка AR(p), она имеет вид:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \dots + \beta_p y_{t-p} + \varepsilon_t; t \in 1:n. \quad (6.37)$$

Модель 2-го порядка AR(2) запишется так:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \varepsilon_t; t \in 1:n.$$

2) Модель скользящей средней q -го порядка MA(q), она имеет вид:

$$y_t = \varepsilon_t + \gamma_1 \varepsilon_{t-1} + \gamma_2 \varepsilon_{t-2} + \dots + \gamma_q \varepsilon_{t-q}; t \in 1:n. \quad (6.38)$$

Модель 2-го порядка MA(2) запишется так:

$$y_t = \varepsilon_t + \gamma_1 \varepsilon_{t-1} + \gamma_2 \varepsilon_{t-2}; t \in 1:n.$$

3) Авторегрессионная модель скользящей средней ARMA(p, q):

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \dots + \beta_p y_{t-p} + \varepsilon_t + \gamma_1 \varepsilon_{t-1} + \dots + \gamma_q \varepsilon_{t-q}; t \in 1:n. \quad (6.39)$$

Модель ARMA($p=2, q=2$) запишется так:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \varepsilon_t + \gamma_1 \varepsilon_{t-1} + \gamma_2 \varepsilon_{t-2}; t \in 1:n.$$

Чаще всего на практике используется метод проб и ошибок: проверяются на пригодность различные модели, начиная с самых простых.

В качестве примера рассмотрим авторегрессионную модель 1-го порядка. Пусть мы имеем регрессионную модель:

$$Y=X\beta+\varepsilon \text{ или } y_t=\beta_0 + \sum\beta_jx_{tj} + \varepsilon_t \text{ } t\in 1:n \quad (6.40)$$

и пусть отклонения связаны наиболее просто – авторегрессионным процессом 1-го порядка:

$$\varepsilon_t = \rho\varepsilon_{t-1} + v_t, \quad (6.41)$$

где v_t - "белый шум" ($M(v_t)=0$ и $D(v_t)=\sigma_0^2$);

ρ - коэффициент авторегрессии.

Теперь из выражения (6.41) нетрудно получить формулу:

$$\sigma^2 = \sigma_0^2 / (1-\rho^2) \quad (6.42)$$

и далее для коэффициентов автокорреляции 1-го и m -го порядков:

$$\rho(\varepsilon_t, \varepsilon_{t-1}) = \rho \text{ и } \rho(\varepsilon_t, \varepsilon_{t-m}) = \rho^m. \quad (6.43)$$

На основе формул (6.42) и (6.43) можно записать ковариационную матрицу вектора отклонений ε :

$$\sum_{\varepsilon} = \Omega = \frac{\sigma_0^2}{1-\rho^2} \begin{pmatrix} 1 & \rho & \dots & \rho^{n-1} \\ \rho & 1 & \dots & \rho^{n-2} \\ \dots & \dots & \dots & \dots \\ \rho^{n-1} & \rho^{n-2} & \dots & 1 \end{pmatrix}. \quad (6.44)$$

Для получения наиболее эффективных значений оценок параметра β применяется обобщённый МНК. Неизвестное значение ρ оценивается применением к регрессии (6.40) обычного МНК.

Можно пойти другим путем. Исключая ε_t из уравнений (6.40) и (6.41), получим классическую ЛММР: отклонения v_t независимы и имеют постоянную дисперсию σ_0^2 :

$$y_t - \rho y_{t-1} = \beta_0(1-\rho) + \sum\beta_j(x_{tj} - \rho x_{t-1j}) + v_t \text{ } t=1, \dots, n. \quad (6.45)$$

Таким образом, автокорреляция довольно легко устраняется.

Методические материалы по теме 8

Конспект лекций, Учебник [1, Кремер-Путко], Учебник [2, Настин], ФОС в ЭИОС. В ходе работы по теме студенту следует использовать лекционный материал; материалы, полученные в ходе практического занятия, материалы в разделе дисциплины в ЭИОС КГТУ, а главное – рекомендованные здесь источники.

Ссылки на рекомендуемые источники по теме 8:

[1, Кремер-Путко, с. 167-175] – это базовый учебник, которого мы придерживаемся, в том числе в терминологии и обозначениях.

[2, Настин, с. 97-101] – это основной учебник.

[4, Карлов, с.155-193] – учебное пособие, здесь основы математической статистики.

7 Методические указания по подготовке к практическим занятиям по 2-й части дисциплины

Вторая часть дисциплины "Эконометрика" так же, как и первая часть, состоит из 4-х разделов-тем, их номера 5-8. Ниже по темам приведены вопросы для подготовки к практическим занятиям. В темах 5, 6, 8 - по 10 вопросов, в теме 7 – 12 вопросов.

Рекомендуем студентам освежить в памяти методические указания по подготовке к практическим занятиям, которые приведены выше в самом начале раздела 2 для 1-й части дисциплины.

Здесь мы приведём только главное: 1) для того, чтобы каждое практическое занятие прошло с максимальной пользой, настоятельно рекомендуем готовиться к ним; 2) при разборе материала крайне желательно использовать графические образы, стараться составить такой образ для любого объекта дисциплины. Например, для условной регрессионной модели магазина создать графический образ гетероскедастичности с постоянным относительным стандартным отклонением 15 %.

Ниже даются рекомендации по формам проведения занятий, вопросы по темам. Заголовки даны по очной форме (во множественном числе - на каждую тему, как правило, 2 практических занятия по 2 часа). По очно-заочной форме на каждую тему выделен 1 час практических занятий.

Занятия по теме 5. Инструментальные методы регрессионного анализа.

Форма занятия: интерактивное практическое занятие.

План занятия:

- а) Опрос по материалам лекций.
- б) Работа с тестом.
- в) Решение задач, в том числе с использованием табличного редактора Excel.

Вопросы по теме 5:

- 1) Назовите три и более причины, заставляющие объединять выборки.
- 2) Раскройте реализацию метода Чоу по шагам, на примере.
- 3) Какую размерность имеет стандартизированный коэффициент регрессии $b_j' = b_j s_{xj}/s_y$?
- 4) Верно ли выражение: $r_{xjy} = b_j s_{xj}/s_y$?
- 5) В каком диапазоне изменяются факторные переменные в уравнении регрессии со стандартизированными коэффициентами?

- б) Какие возможности для анализа открывает запись уравнения регрессии со стандартизированными коэффициентами регрессии?
- 7) В чём смысл элиминирования переменных?
- 8) Частные коэффициенты корреляции. В чём смысл выражения $r_{i-j,1,2,\dots,p}$?
- 9) Когда оправдано применение коэффициента ранговой корреляции Спирмена?
- 10) Как следует ранжировать объекты при отсутствии существенных различий между ними?

Занятия по теме 6. Регрессионные модели временных рядов и прогнозирование

Форма занятия: интерактивное практическое занятие.

План занятия:

- а) Опрос по материалам лекций.
- б) Работа с тестом.
- в) Решение задач, в том числе с использованием табличного редактора Excel.

Вопросы по теме 6:

- 1) Что такое временной ряд?
- 2) Сколько составляющих учитывается в структуре временного ряда – по максимуму?
- 3) Сколько случайных составляющих в структуре временного ряда?
- 4) Какой ряд называется стационарным?
- 5) Что собой представляет объект, для которого формируется автокорреляционная функция?
- 6) Что такое временной лаг?
- 7) Как выглядит типичная корреляционная функция?
- 8) Приведите пример временного ряда y_t и его сглаживания на основе простого среднего арифметического с шириной окна $m = 3$.
- 9) Приведите пример временного ряда y_t и его сглаживания на основе средневзвешенной средней арифметической с шириной окна $m = 3$.
- 10) Что такое точность прогноза (априорная)?

Занятия по теме 7. Обобщённая линейная модель и гетероскедстичность пространственной выборки.

Форма занятия: интерактивное практическое занятие.

План занятия:

- а) Опрос по материалам лекций.

- б) Работа с тестом.
- в) Решение задач, в том числе с использованием табличного редактора Excel.

Вопросы по теме 7:

- 1) Каковы признаки обобщенной линейной модели?
- 2) Можно ли утверждать по виду, что стохастическая модель $Y = X\beta + \varepsilon$ является обобщённой?
- 3) Является ли эффективной оценка параметров обобщённой модели МНК-методом ($b = (X'X)^{-1}X'Y$)?
- 4) Верно ли утверждение: для обобщенной модели ковариационная матрица вектора оценок параметров определяется формулой: $\Sigma_{b^*} = (X'X)^{-1}X'\Omega X(X'X)^{-1}$?
- 5) Что утверждает теорема Айткена?
- 6) Назовите два недостатка коэффициента детерминации для обобщённой линейной регрессионной модели.
- 7) Как в матричной форме записать условие гетероскедастичности для линейной регрессионной модели?
- 8) К чему может привести неэффективность оценки b для вектора β (в смысле теоремы Гаусса-Маркова)?
- 9) Какая идея используется в тесте ранговой корреляции Спирмена для проверки наличия – отсутствия гетероскедастичности?
- 10) Какая идея используется в тесте Голдфелда-Квандта для проверки наличия – отсутствия гетероскедастичности?
- 11) Какая идея используется в тесте Уайта для проверки наличия – отсутствия гетероскедастичности?
- 12) Какая идея используется во взвешенном МНК-методе, как можно с помощью него устранить гетероскедастичность?

Занятия по теме 8. Обобщённая линейная модель и автокорреляция остатков.

Форма занятия: интерактивное практическое занятие.

План занятия:

- а) Опрос по материалам лекций.
- б) Работа с тестом.
- в) Решение задач, в том числе с использованием табличного редактора Excel.

Вопросы по теме 8:

- 1) Какой смысл вкладывается в термин "автокорреляция временного ряда"?
- 2) Как выглядит ковариационная матрица Σ_b для обобщённой регрессионной модели с автокорреляцией остатков?
- 3) В чём состоит идея теста Дарбина-Уотсона на наличие автокорреляции между соседними наблюдениями?
- 4) Как графически выглядит шкала теста Дарбина-Уотсона?
- 5) Назовите два недостатка теста Дарбина-Уотсона.
- 6) В чём состоит идея теста Бреуша-Годфри на наличие автокорреляции между соседними и отдалёнными наблюдениями?
- 7) Что отражает уравнение регрессии $e_t = \rho_1 e_{t-1} + \rho_2 e_{t-2}$ в случае его значимости и незначимости коэффициента ρ_2 ?
- 8) В чём основная причина автокорреляции отклонений?
- 9) Как классифицируется линейная стохастическая модель $y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \varepsilon_t$; $t \in 1:n$?
- 10) Как классифицируется линейная стохастическая модель $y_t = \varepsilon_t + \gamma_1 \varepsilon_{t-1} + \gamma_2 \varepsilon_{t-2}$; $t \in 1:n$?

8 Методические указания по выполнению контрольной работы № 2

8.1 Общие требования, выбор варианта, вопросов и исходных данных для задач

Структура работы. Контрольная работа состоит из 7 разделов:

- 1) Вопрос №__ : <Название вопроса>.
- 2) Вопрос №__ : <Название вопроса>.
- 3) Задача-2.1 (объединение выборок - критерий Чоу, ручное решение).
- 4) Задача-2.2 (проверка на гомоскедастичность, ручное решение).
- 5) Задача-2.3 (выделение тренда скользящей средней в MS Excel).
- 6) Задача-2.4 (выделение тренда экспоненциальным сглаживанием в MS Excel).
- 7) Задача 2.5 (прогнозирование на основе авторегрессии в MS Excel).

В приложении В дана структура работы в форме листа "СОДЕРЖАНИЕ".

Примечание. Каждая задача здесь кодируется двумя числами: номер части дисциплины и её номер по порядку; кодирование задач в 1-й части дисциплины оставлено прежним: номер по порядку.

Общие требования к оформлению и объёму

1) В основу правил оформления положен ГОСТ 7.32-2017 "Структура и правила оформления отчёта по НИР", другие стандарты.

2) Подробно правила оформления изложены в источнике [5, Мнацаканян и др.]. Здесь имеется образец титульного листа, правила оформления текста с его формулами, рисунками, таблицами, списком использованных источников и приложениями.

3) Из текста следует исключить пропуски строк между абзацами; должен быть включён автоперенос.

4) Объём работы не должен превышать 25 страниц (от титульного листа до последней страницы, включая приложения). Работы с объёмом больше 25 страниц к рассмотрению не принимаются.

Правила заимствования. При заимствовании материалов из источников нужно соблюдать следующие правила:

1) не следует выхватывать из сети первые попавшиеся тексты, они обычно адресованы студентам, изучающим высшую математику;

2) следует выбирать источники, которые по терминологии и обозначениям близки к тексту настоящего пособия и к источникам [1, 2];

3) терминология и обозначения должны быть едиными во всех частях работы.

4) следует делать ссылки на использованные источники, например, так: [3], или так [5, с. 116], или так: [6, Берндт].

Особенность работы. Трудоёмкость выполнения контрольной работы снижается на порядок за счёт использования ворд-файла настоящего раздела пособия. Ниже для решения каждой задачи даются подробнейшие инструкции и наглядные примеры.

Выбор варианта и номеров вопросов. Вариант выбирается по первым подходящим буквам полного имени и фамилии. Например, если Вы – тАтьяна Соколова, то Ваш вариант А-С, Ваши вопросы 14 и 17, см. таблицу 8.1. В разделе 9 приведено 40 вопросов, из них 20 по первой части дисциплины и 20 по 2-й части.

Таблица 8.1 - Варианты и номера вопросов для контрольной работы № 2 (1-е число – из 1-й части эконометрики, 2-е число – из 2-й части).

Имя*	Первая подходящая буква фамилии													
	А	Б	В	Г	Д	Е	К	Л	М	Н	О	П	Р	С
А	1, 10	2, 11	3, 12	4, 13	5, 14	6, 15	7, 16	8, 17	9, 18	10, 19	11, 20	12, 19	13, 18	14, 17
В	15, 16	16, 15	17, 14	18, 13	19, 12	20, 11	19, 10	18, 9	19, 8	20, 7	19, 6	18, 5	17, 4	16, 3
Е	15, 2	14, 1	13, 2	12, 1	11, 3	10, 4	9, 5	8, 6	7, 7	6, 8	5, 9	4, 10	3, 11	2, 12
О	1, 13	2, 14	3, 15	4, 16	5, 17	6, 18	7, 19	8, 20	9, 19	10, 18	11, 17	12, 16	13, 15	14, 14
И	15, 13	16, 12	17, 11	18, 10	19, 9	20, 8	19, 7	18, 6	17, 5	16, 4	15, 3	14, 2	13, 1	12, 2

*Первая подходящая буква полного имени.

8.2 Пример ручного решения задачи-2.1: "Проверка возможности объединения выборок для линейной парной модели магазина"

Постановка задачи-2.1. В организации имеется семь магазинов, $n = 7$. Три из них торгуют рыбными продуктами (тип Р, $n_1 = 3$), а четыре – мясными (тип М, $n_2 = 4$). Изучается зависимость суточной выручки Y от размера торговой площади X . Возникает вопрос: можно ли объединить выборки X_1 и Y_1 рыбных и X_2 и Y_2 мясных магазинов и получить смешанные выборки X и Y ? Задача решается с помощью критерия Чоу. Он предусматривает построение трёх регрессионных моделей на основе трёх пар выборок: смешанных, для рыбных и для мясных магазинов.

Исходные данные для задачи-2.1:

- выборки X и Y для варианта А-Б из таблицы 2, графы 2-3 и 10-11 из той же таблицы (см. также таблицу А.1);
- подраздел А.2 с отдельными исходными данными, здесь примем: $\alpha = 0,05$; $\Gamma_1 = \Gamma_2 = 0$.

Решение задачи-2.1.

1) Критерий Чоу: если выполняется неравенство:

$$F = \frac{\sum_{i=1}^n e_i^2 - \sum_{i=1}^{n_1} \gamma_i^2 - \sum_{i=n_1+1}^n \mu_i^2 \cdot (n-2p-2)}{\sum_{i=1}^{n_1} \gamma_i^2 + \sum_{i=n_1+1}^n \mu_i^2} < F_{\alpha, p+1, n-2p-2}, \quad (8.1)$$

то выборки объединять можно (они "однородные");

- здесь p - число факторных переменных в каждом уравнении регрессии;
 e_i - остатки, полученные от регрессии на выборках типа О;
 γ_i - остатки, полученные от регрессии на выборках типа Р;
 μ_i - остатки, полученные от регрессии на выборках типа М.

2) Решение для обобщённых выборок. Эта задача уже была решена в подразделе 3.2, было получено уравнение регрессии (3.3):

$$\hat{y} = 0,43 + 1,54x.$$

Отклонения и их квадраты возьмём из таблицы 3.2, графы 10 и 11. Заполним расчётную таблицу 8.2.

Таблица 8.2 - Типы магазинов, исходные данные, остатки (отклонения) от регрессии на обобщённых выборках

№ п/п	Тип магазина	Площадь X	Выручка Y	e_i	e_i^2
1	2	3	4	5	6
1	Р	1	2	-0,03	0
2	М	1	3	-1,03	1,06
3	Р	2	4	-0,49	0,24
4	М	3	5	0,05	0
5	М	4	5	1,59	2,53
6	М	5	7	1,13	1,28
7	Р	8	14	-1,25	1,56
Итого	-	24	40	-	6,65

Наш результат-1: сумма квадратов отклонений равна 6,65.

3) Решение для выборок рыбных магазинов. Заполним расчётную таблицу 8.3, графы 1-6.

Таблица 8.3 – Расчётная таблица для рыбных магазинов

№ п/п	X_i	Y_i	$X_i \cdot Y_i$	X_i^2	Y_i^2	\hat{y}_i	$\gamma_i^2 = (\hat{y} - y_i)^2$
1	2	3	4	5	6	7	8
1	1	2	2	1	4	2,13	0,02
3	2	4	8	4	16	3,83	0,03
7	8	14	112	64	196	14,03	0
Σ	11	20	122	69	216		0,05
	11/3=3,67	20/3=6,67	122/3=40,67	69/3=23	216/3=72		

Рассчитаем параметры регрессии:

$$b_1 = (\overline{xy} - \bar{x} \cdot \bar{y}) / (\overline{x^2} - \bar{x}^2) = (40,67 - 3,67 \cdot 6,67) / (23 - 3,67^2) = 16,19 / 9,53 = 1,70.$$

$$b_0 = \bar{y} - b_1 \bar{x} = 6,67 - 1,7 \cdot 3,67 = 0,43.$$

Запишем второе уравнение регрессии – для рыбных магазинов:

$$\hat{y} = 0,43 + 1,7x.$$

Рассчитаем и заполним графу 7:

$$\hat{y} = 0,43 + 1,7x = 0,43 + 1,7 \cdot 1 = 2,13.$$

$$\hat{y} = 0,43 + 1,7 \cdot 2 = 3,83.$$

$$\hat{y} = 0,43 + 1,7 \cdot 8 = 14,03.$$

Рассчитаем и заполним графу 8.

Наш результат-2: сумма квадратов отклонений равна 0,05.

4) Решение для выборок мясных магазинов. Заполним расчетную таблицу 8.4, графы 1-6.

Таблица 8.4 - Расчётная таблица для мясных магазинов

№ п/п	X _i	Y _i	X _i ·Y _i	X _i ²	Y _i ²	\widehat{y}_i	$\gamma_i^2 = (\widehat{y}_i - y_i)^2$
1	2	3	4	5	6	7	8
2	1	3	3	1	9	2,95	0,00
4	3	5	15	9	25	4,77	0,05
5	4	5	20	16	25	5,68	0,46
6	5	7	35	25	49	6,59	0,17
Σ	13	20	73	51	108		0,68
	13/4=3,25	20/4=5	73/4=18,25	51/4=12,75	108/4=27		

Рассчитаем параметры регрессии:

$$b_1 = (\overline{xy} - \bar{x} \cdot \bar{y}) / (\overline{x^2} - \bar{x}^2) = (18,25 - 3,25 \cdot 5) / (12,75 - 3,25^2) = 2 / 2,19 = 0,91.$$

$$b_0 = \bar{y} - b_1 \bar{x} = 5 - 0,91 \cdot 3,25 = 2,04.$$

Запишем третье уравнение регрессии – для рыбных магазинов:

$$\hat{y} = 2,04 + 0,91x.$$

Рассчитаем и заполним графу 7:

$$\hat{y} = 2,04 + 0,91x = 2,04 + 0,91 \cdot 1 = 2,95;$$

$$\hat{y} = 2,04 + 0,91 \cdot 3 = 4,77;$$

$$\hat{y} = 2,04 + 0,91 \cdot 4 = 5,68;$$

$$\hat{y} = 2,04 + 0,91 \cdot 5 = 6,59.$$

Рассчитаем и заполним графу 8.

Наш результат-3: сумма квадратов отклонений равна 0,68.

5) Рассчитаем значение статистики Фишера-Снедекора:

$$F = \frac{6,65 - 0,05 - 0,68}{0,05 + 0,68} \cdot \frac{(7 - 2 \cdot 1 - 2)}{(1 + 1)} = \frac{8,88}{0,73} = 12,16.$$

Определим по таблице критическое значение статистики Фишера-Снедекора:

$$F_{\alpha, p+1, n-2p-2} = F_{0,05; 1+1; 7-2 \cdot 1-2} = F_{0,05; 2; 3} = 19,16.$$

Мы получили неравенство: $12,16 < 19,16$.

Вывод: Выборки по рыбным и мясным магазинам "однородные", и их можно объединять (с вероятностью ошибки 0,05).

8.3 Пример ручного решения задачи-2.2: "Проверка линейной парной модели магазина на гомоскедастичность по двум критериям"

Постановка задачи-2.2. В регрессионном анализе сформулированы шесть предпосылок, выполнение которых облегчает решение задач. Одна из предпосылок формулируется так: дисперсия отклонений должна быть постоянной и не зависеть от факторных переменных – условие гомоскедастичности. Задача-6 посвящена методам проверки выполнения этого условия на примере парной регрессионной модели магазина.

Исходные данные для задачи-2.2:

- выборки X и отклонений e_i для варианта А-Ъ, см. таблицу 3.2, графы 2-10; таблицу А.1;
- отдельные исходные данные из подраздела А.2, примем: $\alpha = 0,05$; $\Gamma_1 = \Gamma_2 = 0$.

Решение задачи-2.2.

1) **Тест ранговой корреляции Спирмена** использует такую идею: модули отклонений $|e_i|$ являются оценками СКО σ_i , поэтому при гетероскедастичности модули отклонений $|e_i|$ и значения факторов x_i будут коррелированы. Здесь применяется коэффициент ранговой корреляции $\rho_{x,e}$, который рассчитывается по формуле:

$$\rho_{x,e} = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}, \quad (8.2)$$

где d_{ii} - разность между рангами значений x_i и e_i i -го объекта;

n - число объектов, пар наблюдений.

Критерий Спирмена: коэффициент ранговой корреляции значим на уровне значимости α (справедлива гипотеза H_1 , имеет место гетероскедастичность), если выполняется условие:

$$|t| = \frac{|\rho_{x,e}| \sqrt{n-2}}{\sqrt{1-\rho_{x,e}^2}} > t_{1-\alpha, n-2}, \quad (8.3)$$

где $t_{1-\alpha, n-2}$ - табличное значение t -критерия Стьюдента с уровнем значимости α и числом степеней свободы $(n-2)$.

Построим расчётную таблицу 8.5, в которой графы 1 и 2 возьмём из граф 2 и 3 таблицы 3.2 (Задача-1), остальные графы рассчитаем.

Таблица 8.5 – Расчётная таблица

X_i	e_i	Ранг r_{x_i}	Ранг r_{e_i}	$d_i=r_{x_i}-r_{e_i}$	d_i^2
1	-0,03	$(1+2)/2=1,5$	1	0,5	0,25
1	-1,03	$(1+2)/2=1,5$	4	-2,5	6,25
2	-0,49	3	3	0	0
3	0,05	4	2	2	4
4	1,59	5	7	-2	4
5	1,13	6	5	1	1
8	-1,25	7	6	1	1
Итого					16,5

Теперь можно вычислить коэффициент ранговой корреляции Спирмена:

$$\rho_{x,e} = 1 - \frac{6 \cdot 16,5}{7(7^2 - 1)} = 0,71.$$

Рассчитаем значение статистики Стьюдента t :

$$t = \frac{0,71\sqrt{7-2}}{\sqrt{1-0,71^2}} = 2,26.$$

Из таблицы возьмём критическое значение статистики Стьюдента:

$$t_{1-\alpha, n-2} = t_{1-0,05; 7-2} = t_{0,95; 5} = 2,57.$$

Вывод: поскольку $2,26 < 2,57$, имеет место гомоскедастичность (хотя и на пределе). Подобный же вывод можно сделать и "на глаз" по рисунку 3.1 – нет увеличения разброса точек с увеличением фактора X .

2) Тест Голдфелда-Квандта на наличие гетероскедастичности. Пусть ошибки регрессии имеют нормальный закон распределения и пусть средние квадратичные отклонения для e_i пропорциональны значениям переменной X . Упорядочим наблюдения в порядке возрастания X и выберем m первых и m последних наблюдений (две m -выборки).

Гипотеза H_0 о наличии гомоскедастичности будет равносильна утверждению о том, что две нормально распределенные m -выборки имеют одинаковые дисперсии (разность дисперсий между ними равна нулю).

Гипотеза H_0 о равенстве дисперсий по критерию Фишера-Снедекора отвергается (имеет место гетероскедастичность), если выполняется неравенство (в числитель помещается бóльшая сумма):

$$F = \frac{\sum_{i=n-m+1}^n e_i^2}{\sum_{i=1}^m e_i^2} > F_{\alpha, n-m, n-m}. \quad (8.4)$$

Мощность теста максимальна, если выбирать длину m -выборки по формуле $m \approx n/3$. Для нашего примера с семью магазинами (очень маленькая выборка) примем $m = 3$.

Из таблицы 3.2 (графы 2, 10, 11) выбираем нужные нам данные для таблицы 8.6.

Таблица 8.6 – Расчётная таблица для критерия Голдфелда-Квандта.

№ п/п	X_i	e_i	e_i^2
1	1	-0,03	0,001
2	1	-1,03	1,061
3	2	-0,49	0,240
Итого	-	-	1,302
5	4	1,59	2,528
6	5	1,13	1,277
7	8	-1,25	1,563
Итого	-	-	2,840

Из таблицы критических точек распределения F Фишера-Снедекора находим значение:

$$F_{кр} = F_{\alpha, m-p, m-p} = F_{0,05, 2-1, 2-1} = F_{0,05; 1; 1} = 161,$$

где $m = 2$ – число параметров в уравнении регрессии;

$p = 1$ – число факторных переменных.

Вычисляем F по формуле (8.4) и получим неравенство:

$$F = \frac{2,840}{1,302} = 2,18; \quad \mathbf{2,18 < 161.}$$

Вывод: значение F во много раз меньше значения $F_{кр}$ (различия в дисперсиях невелики), поэтому принимаем гипотезу H_0 : разность дисперсий равна нулю, имеет место гомоскедастичность.

8.4 Пример решения задачи 2.3: "Выделение тренда временного ряда методами скользящей средней в MS Excel"

Постановка задачи-2.3. Дан временной ряд – 20 значений поквартального объёма выпуска морепродуктов рыбокомбинатом в сопоставимых ценах. Необходимо выделить из него неслучайную составляющую – тренд – двумя методами скользящей средней.

Исходные данные для задачи-2.3. В части, общей для всех вариантов, временной ряд приведён в таблице А.7. В нём всего 20 уровней - клеток. Числа в 13-ти клетках (с 5-й по 17-ю) уже проставлены - для всех вариантов одни и те же, а 7 клеток (1-4 и 18-20) пустые. Они заполняются каждым студентом самостоятельно на основе столбца Y задачи-1 для своего варианта из 1-й контрольной работы, см. таблицы А1-А5.

Алгоритм заполнения таблицы:

- 1) очередное число из столбца Y умножается на 100;
- 2) если произведение меньше 300, то в соответствующую клетку таблицы А.1 записывается 300;

3) если произведение больше 1200, то в соответствующую клетку таблицы А.1 записывается 1200;

4) иначе произведение без изменения записывается в соответствующую клетку таблицы А.1.

В таблице 8.7 приведён пример сформированного таким образом временного ряда.

Таблица 8.7 – Пример формирования поквартального временного ряда для варианта АО

Квартал	Год наблюдения				
	1	2	3	4	5* ?
I	$2*100 \div 300,0$	660,0	830,0	750,0	750,0
II	$3*100=300,0$	670,0	810,0	710,0	$8*100=800,0$
III	$6*100=600,0$	680,0	750,0	660,0	$9*100=900,0$
IV	$6*100=600,0$	690,0	780,0	690,0	$20*100 \div 1200,0$

Решение задачи-2.3.

1) Ввод временного ряда, построение его графика, анализ структуры ряда по его графику. Расположим имя временного ряда в ячейке А1, например, Y, а значения - в ячейках А2-А21, $n = 20$. Для построения графика временного ряда выделим курсором ячейки А1-А21. Во вкладке "Вставка" выберем вид графика "Точечная с маркерами", получим график временного ряда, рисунок 8.1. Из рисунка видно, что уровни ряда в начальном периоде от $t = 1$ до $t = 10$ имеют тенденцию возрастания, а затем – убывания. Уровни ряда содержат колебания, но их характер не говорит о наличии сезонной компоненты, связанной с кварталами года.

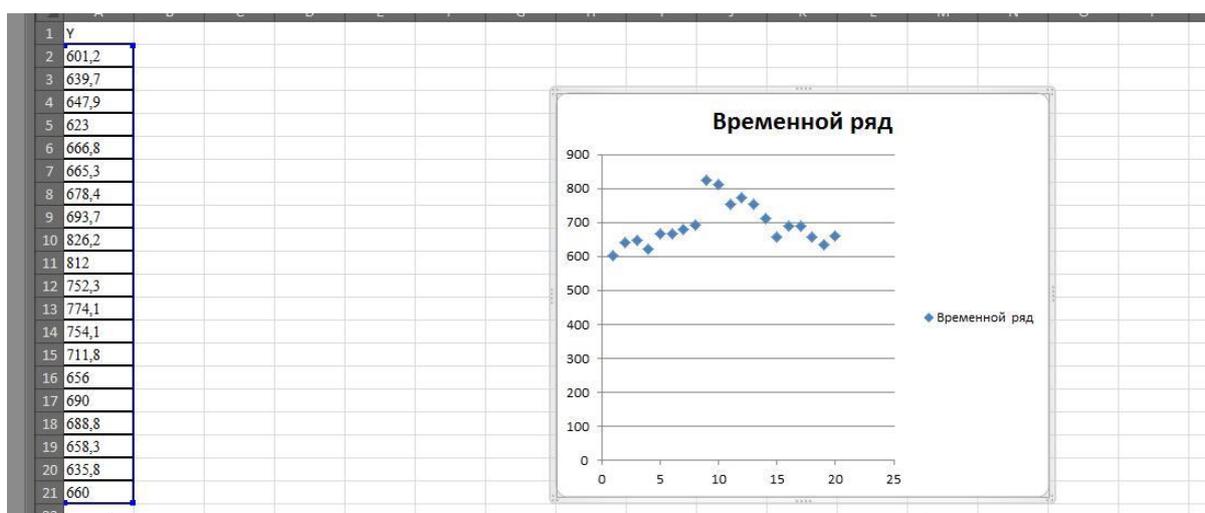


Рисунок 8.1 – Временной ряд и его график

2) Выделение тренда временного ряда методом простой скользящей средней с **чётной** длиной окна скольжения, равной 4. По вкладке "Данные" выберем "Анализ данных", здесь в окне выберем "Скользящее среднее", далее заполним поля его окна:

- в поле *входной интервал* укажем ячейки **A1-A21**, содержащие временной ряд; это легко сделать выделением ячеек A1-A21 курсором;
- выберем "метки" в первой строке, так как в ячейке **A1** у нас имя временного ряда;
- в поле *интервал* укажем длину окна скольжения, равную 4;
- в поле *выходной интервал* для удобства построения графиков ряда и скользящих средних укажем ячейку **I2**;
- выберем "Вывод графика" и "Стандартные погрешности" - в этом случае. наряду со сглаженным рядом, выводятся и вычисленные по окну скольжения средние квадратичные отклонения уровней ряда от сглаженных значений:

$$e_t = \sqrt{\frac{\sum_{i=0}^{T-1} (y_{t-i} - \bar{y}_{t-i})^2}{T}}, \quad (8.5)$$

где T – длина окна скольжения; \bar{y}_{t-i} – скользящее среднее для момента времени $t-i$.

По "ОК" в ячейках **I2-I21** получим сглаженные значения временного ряда, а в ячейках **J2-J21** - значения средних квадратичных отклонений, рисунок 8.2.

Внимание: процедура сглаживания относит сглаженные значения уровней, попавших в окно, к последнему моменту времени в окне скольжения. В данном случае сглаженные значения располагаются, начиная с ячейки **I5**. Необходимо рассчитать сглаженные значения для моментов времени, находящихся в середине окна скольжения (длина окна 4). Значения скользящей средней для 3-го момента времени в окне скольжения находятся как среднее арифметическое скользящего среднего для данного и следующего положения окна скольжения на временной оси. Для этого выделим ячейку **B4** и в строке формул введём =СРЗНАЧ(**I5:I6**), по "ОК" получим значение скользящей средней для момента времени $t = 3$. Аналогично в ячейках B5-B19 получаем значения скользящей средней для последующих моментов времени, до момента $n-2$ включительно, рисунок 8.2.

3) Построение скользящей средней с **нечётной** длиной окна скольжения, равной пяти, производится так же, результат см. в ячейках S2-S21, рисунок 8.2. Среднее значение уровней, попавших в интервал сглаживания, приписывается последнему моменту времени, вошедшему в интервал сглаживания, поэтому их нужно переместить в середины интервалов. Для этого выделим ячейку C4 и в строке формул введём =S6, тем самым среднее первых пяти уровней относим к

моменту времени $t = 3$. Аналогично в ячейках C5-C19 получаем значения скользящей средней для последующих моментов времени.

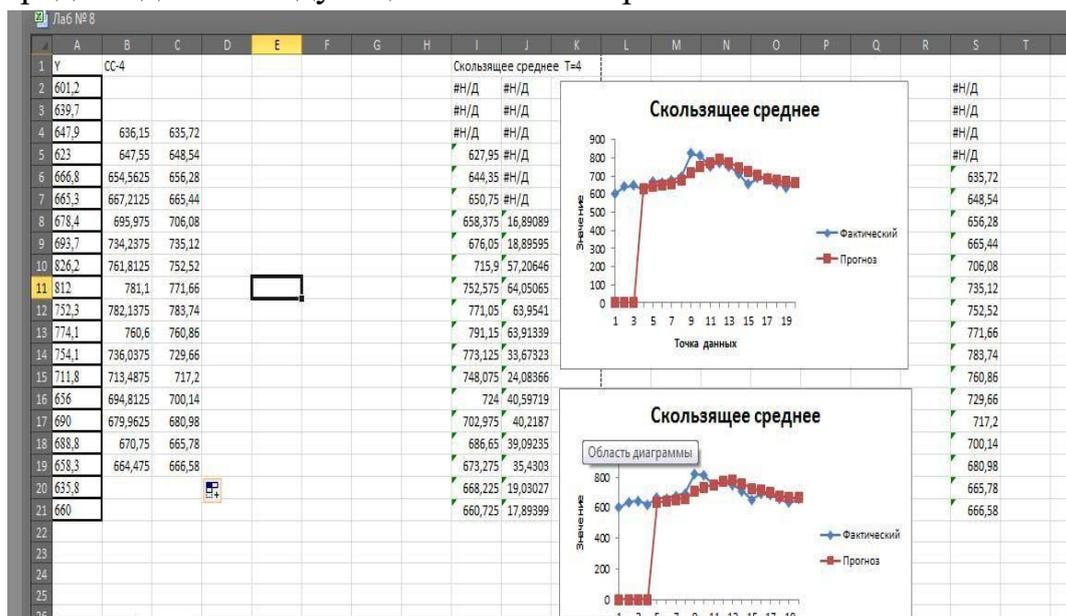


Рисунок 8.2 - Построение скользящей средней с длиной окна, равной 4 и 5

8.5 Пример решения задачи 2.4: "Выделение тренда временного ряда методом экспоненциального сглаживания в MS Excel"

Постановка задачи-2.4 и исходные данные для неё точно такие, как в задаче-2.3. Исходный временной ряд уже полностью сформирован и представлен в таблице 8.7. Задачи 2.3 и 2.4 отличаются только методами решения.

Решение задачи-2.4. Экспоненциальное сглаживание проведем для двух значений параметра сглаживания α : 0,1 и 0,3. В окне "Анализ данных" выберем инструмент анализа "Экспоненциальное сглаживание". Далее работаем с полями окон этого инструмента:

- в поле *входной интервал* укажем ячейки **A1-A21**, содержащие временной ряд;
- выберем метки в первой строке, так как ячейка **A1** содержит имя временного ряда;
- в поле *фактор затухания* зададим число, равное $1-\alpha$ (в примере 0,9 и 0,7);
- в поле *выходной интервал* укажем ячейку **F2** для $\alpha=0,1$ и **G2** для $\alpha=0,3$;
- выберем *вывод графика*.

Колебания уровней относительно тенденции в начальный период, от $t = 1$ до $t = 8$, имеют сезонный характер с возрастающей амплитудой колебаний. При $t > 8$ характер колебаний изменяется, поэтому нельзя говорить о наличии сезонных колебаний во всем временном ряде.

Результаты экспоненциального сглаживания приведены на рисунке 8.3. Прогноз уровней временного ряда по формуле экспоненциального сглаживания $S_{t+1} = S_t + \alpha(y_t - S_t)$ для последующих моментов времени $t = 21$ и $t = 22$ рассчитаем в ячейках F22-F23 и G22-G23, рисунок 8.3.

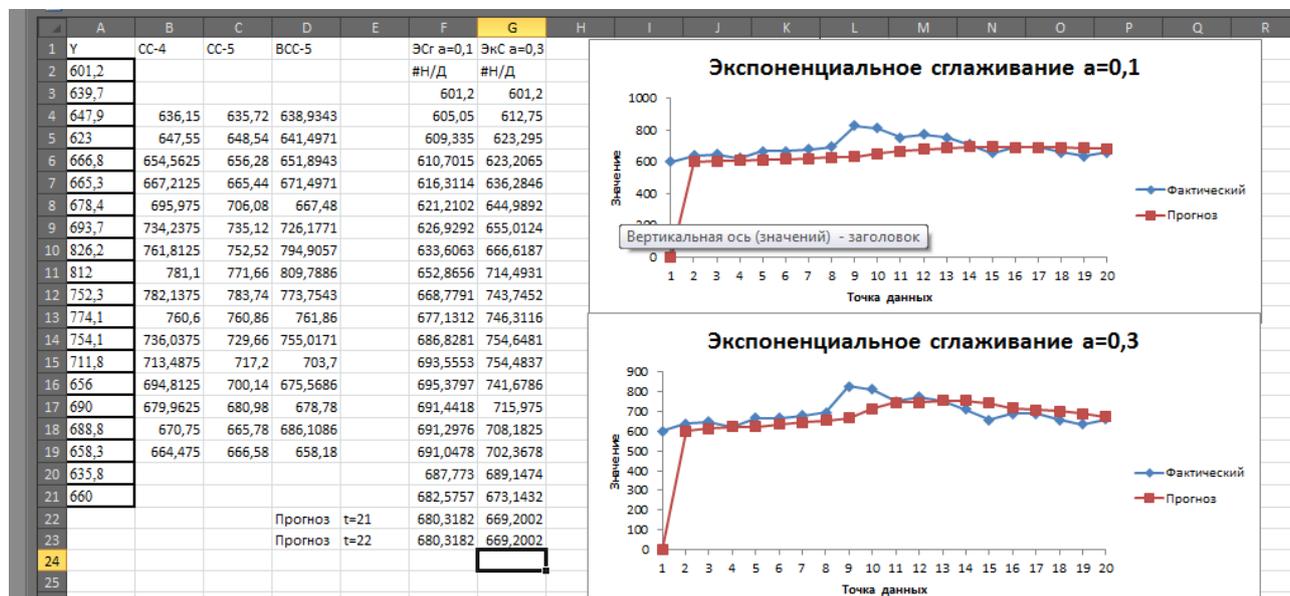


Рисунок 8.3 - Графики экспоненциального сглаживания

Общее заключение. Проведенное сглаживание уровней временного ряда квартальных объемов выпуска продукции говорит о тенденции возрастания в начальном периоде и тенденции убывания в конечном периоде. Изменение тенденции происходит в момент времени $t = 10$. Графики сглаженных уровней показывают, что простые скользящие средние с окном скольжения 4 и 5 хорошо отражают тенденцию возрастания квартальных объемов выпуска продукции в начальном периоде и тенденцию их убывания в конечном периоде, но сдвигают на шаг вправо точку изменения тенденции.

Экспоненциальное сглаживание с параметром $\alpha = 0,1$ занижает тенденцию возрастания в начальном периоде и тенденцию убывания в конечном периоде, а также значительно сдвигает вправо точку изменения характера тенденции. Экспоненциальное сглаживания с параметром $\alpha = 0,3$ также занижает тенденцию возрастания в начальном периоде и значительно сдвигает вправо точку изменения характера тенденции. В обоих случаях этот метод хорошо сглаживает колебания уровней временного ряда.

Из рассмотренных методов сглаживания и фильтрации тенденцию временного ряда квартальных объемов выпуска продукции лучше представляют простые скользящие средние с окном скольжения 4 или 5.

8.6 Пример решения задачи 2.5: "Прогнозирование на основе авторегрессии 1-го порядка в MS Excel"

Постановка задачи-2.5. Один из важнейших экономических показателей региона – годовой объём валового регионального продукта (ВРП). В задаче он представлен временным рядом за 10 лет, таблица А.8. Требуется произвести сглаживание ряда и прогнозирование на один год вперёд.

Исходные данные для задачи 2.5 – это временной ряд из 10 чисел. В таблице А.8 он представлен так: первых 4 числа отсутствуют, они вычисляются по алгоритму, приведенному выше в подразделе 8.4. Другими словами, в задаче 2.5 используются первые 10 чисел временного ряда из таблицы 8.7 задачи 2.3. В таблице 8.8 дан пример этого ряда.

Таблица 8.8 – Валовой региональный продукт (ВРП) по годам за 10 лет, млрд. руб.

Год	ВРП	Год	ВРП	Год	ВРП
1	300,0	5	660,0	9	830,0
2	300,0	6	670,0	10	810,0
3	600,0	7	680,0	-	-
4	600,0	8	690,0	-	-

Метод решения основан на авторегрессии первого порядка, сокращённое обозначение AR I. Стохастическая регрессионная модель на её основе имеет вид:

$$Y_i = a_0 + a_1 * Y_{i-1} + \varepsilon_i.$$

Как видно из этой формулы, линейная модель авторегрессии 1-го порядка состоит только из одного влияющего фактора Y_{i-1} , т. е. изучается зависимость только от того, каким был итоговый показатель на один шаг в прошлое, например, вчера.

Рассмотрим построение модели с помощью "Пакета анализа" в Excel на примере некоего ряда динамики валового регионального продукта (ВРП) с 2004 по 2012 г., рисунок 8.4.

Исходные данные предварительно нужно подготовить, а именно прописать в четвёртой графе Y_{i-1} – это и будет фактором. Таким образом, наша совокупность, т. е. динамический ряд, который будет использоваться для регрессии, сократился на одну позицию, обратите внимание на графы 1 и 2, с 2005 по 2012 г.

Далее в Excel активируем вкладку "Данные" и нажимаем "Анализ данных", указываем диапазон исходных данных (см. пример на скриншоте) и жмем кнопку ОК. Если по указанному пути нет кнопки "Анализ данных", то "Пакет анализа" нужно активировать, это несложно (инструкцию см. в Интернете).

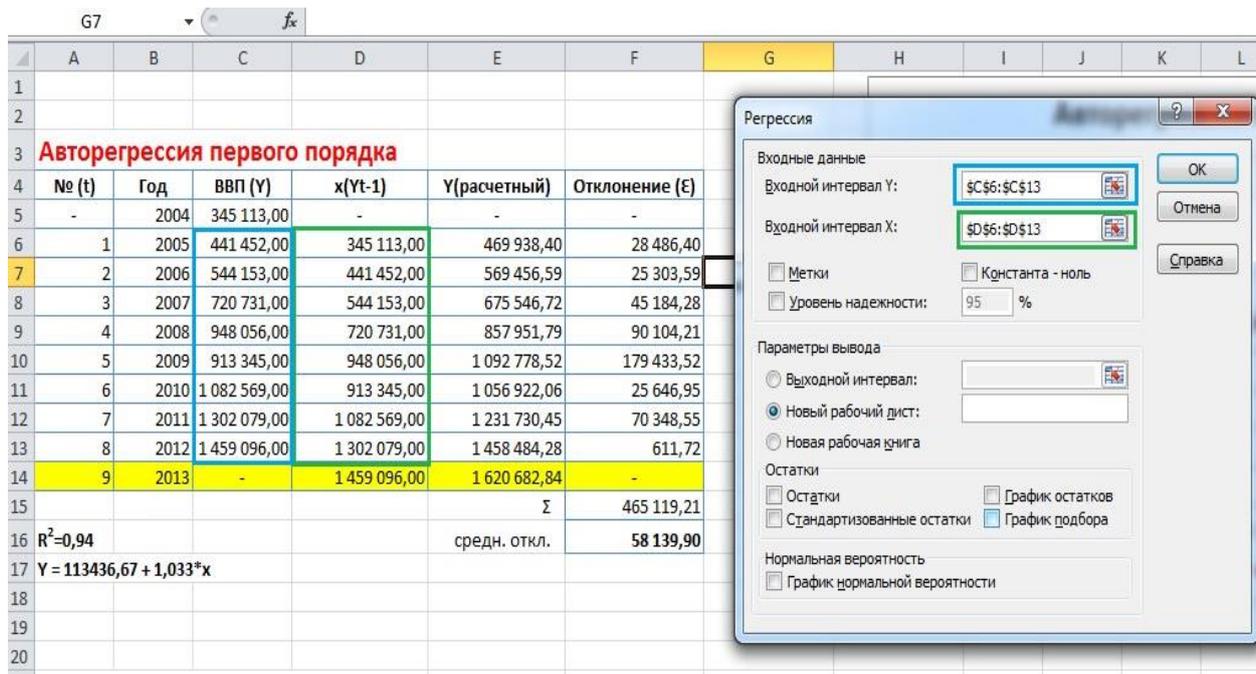


Рисунок 8.4 – Построение авторегрессии 1-го порядка

Вывод итогов								
Регрессионная статистика								
Множественный R	0,971302904							
R-квадрат	0,943429332							
Нормированный R-квадрат	0,934000887							
Стандартная ошибка	90751,79091							
Наблюдения	8							
Дисперсионный анализ								
	df	SS	MS	F	Значимость F			
Регрессия	1	8,241E+11	8,241E+11	100,0620322	5,78175E-05			
Остаток	6	49415325316	8235887553					
Итого	7	8,73515E+11						
Коэффициенты								
	Коэффициенты	Стандартная ошибка	t-статистика	P-Значение	Нижние 95%	Верхние 95%	Нижние 95%	Верхние 95%
Y-пересечение	113436,6764	87378,79114	1,298217507	0,241878233	-100371,5232	327244,9	-100372	327244,9
Переменная X 1	1,032789147	0,103246897	10,00310113	5,78175E-05	0,780153093	1,285425	0,780153	1,285425
Y = 113436,67 + 1,033*x								

Рисунок 8.5 – Результаты расчётов для авторегрессии 1-го порядка

Результаты расчетов – см. рисунок 8.5 – "Пакет анализа" выдает нам на новом листе (если в настройках не было указано иначе), первоочередные по важности ячейки выделены тёмным. Из этих данных собираем авторегрессию с числовыми параметрами:

$$Y = 113436,67 + 1,033 * X$$

ИЛИ

$$Y_i = 113436,67 + 1,033 * Y_{i-1}$$

Возвращаемся в нашу табличку с исходными данными и подставляем полученное уравнение в столбец в качестве формулы, таким образом, получаем расчетные значения по модели, что мы и сделали в столбце Y (расчетный), протягивая формулу на период ниже, получаем прогноз (в табличке строка выделена тёмным).

Для справки. Авторегрессия второго порядка, сокращённое обозначение AR II:

$$Y_i = a_0 + a_1 * Y_{i-1} + a_2 * Y_{i-2} + \varepsilon_i.$$

Модель авторегрессии 2-го порядка отличается от первой тем, что она включает в себя второй фактор Y_{i-2} . Это может повысить значимость регрессии и привести к более точному прогнозу.

9 Методические указания по подготовке к промежуточной аттестации – дифференцированному зачёту

Промежуточная аттестация в форме дифференцированного зачета проходит по результатам прохождения всех видов текущего контроля успеваемости. В качестве оценочных средств для текущего контроля успеваемости используются различные задания, выполняемые студентами на учебных занятиях и во время самостоятельной учебной работы. Сюда относятся:

- тестовые задания;
- задания по темам практических занятий;
- задания по контрольным работам № 2 по всем формам обучения.

Последней группе заданий при подготовке к зачёту следует уделить особое внимание, поскольку:

- выполнение и защита контрольной работы предусмотрены учебными планами по всем формам обучения;
- контрольная работа № 2 представляет собой целостное исследование (продолжение контрольной работы № 1), которое проходит через основные темы дисциплины и включает в себя решение многих эконометрических задач, в том числе на компьютере.

Вопросы для контрольной работы № 2 (часть 2) и для дифференцированного зачёта (части 1 и 2).

Предложенные вопросы охватывают обе части дисциплины "Эконометрика". В основном они взяты из учебника [2, Настин], там же можно найти ответы на них. Но, конечно, рекомендуем посмотреть и другие источники, в первую очередь, это учебник [1, Кремер-Путко], другие источники, приведённые в библиографическом списке.

Всего ниже приведено 40 вопросов, по 20 из каждой части дисциплины.

Часть 1

Тема 1. Основные понятия эконометрики

- 1) Модель прогнозирования банкротства предприятия в форме линейной регрессии.
- 2) Методология и этапы эконометрического моделирования.

Тема 2. Парный регрессионный анализ

- 3) Метод наименьших квадратов.
- 4) Связь коэффициентов регрессии и корреляции.
- 5) Пять предпосылок для парного регрессионного анализа.
- 6) Точечные оценки параметров и теорема Гаусса-Маркова.

- 7) Интервальные оценки функции регрессии $M_x(Y)$.
- 8) Интервальные оценки параметров регрессии.
- 9) Оценка значимости уравнения парной регрессии.

Тема 3. Множественный регрессионный анализ

- 10) Оценка параметров классической нормальной линейной модели множественной регрессии методом наименьших квадратов.
- 11) Оценки ковариационной матрицы для случайного вектора b и дисперсии отклонений.
- 12) Доверительные интервалы для условного математического ожидания $M_x(Y)$ и для индивидуального прогноза.
- 13) Оценка значимости множественной регрессии по трём критериям.
- 14) Мультиколлинеарность и отбор значимых факторов
- 15) Функции и коэффициенты эластичности.

Тема 4. Нелинейный регрессионный анализ

- 16) Нелинейные модели регрессии, их классификация и линеаризация.
- 17) Нелинейная квадратичная модель магазина.
- 18) Фиктивные переменные в линейных моделях.
- 19) Фиктивные переменные в нелинейных моделях.
- 20) Нелинейные производственные функции Кобба-Дугласа.

Часть 2

Тема 5. Инструментальные методы регрессионного анализа

- 1) Критерий Чоу объединения двух выборок.
- 2) Стандартизированные коэффициенты регрессии.
- 3) Частные коэффициенты корреляции между факторными переменными.
- 4) Коэффициент ранговой корреляции Спирмена.

Тема 6. Регрессионные модели временных рядов и прогнозирование

- 5) Структура и классификация временных рядов.
- 6) Автокорреляционная функция временного ряда.
- 7) Моделирование временного ряда.
- 8) Сглаживание временного ряда методами скользящих средних.
- 9) Методология и методы прогнозирования.

Тема 7. Обобщённая линейная модель и гетероскедастичность пространственной выборки

- 10) Признаки обобщенной линейной модели и четыре предпосылки.

- 11) Обобщённый метод наименьших квадратов.
- 12) Сущность и последствия гетероскедастичности.
- 13) Тест Спирмена на наличие гетероскедастичности.
- 14) Тест Голдфелда-Квандта на наличие гетероскедастичности.
- 15) Тест Уайта на наличие гетероскедастичности.
- 16) Устранение гетероскедастичности взвешенным МНК-методом.

Тема 8. Обобщённая линейная модель и автокорреляция остатков

- 17) Автокорреляция остатков временного ряда.
- 18) Тест Дарбина-Уотсона на наличие автокорреляции между соседними наблюдениями.
- 19) Тест Бреуша-Годфри на наличие автокорреляции между соседними и отдалёнными наблюдениями.
- 20) Устранение автокорреляции.

10 Методические указания по выполнению самостоятельной работы по дисциплине – части 2

Самоподготовка - самостоятельная работа студента – главная форма овладения знаниями. В нашей дисциплине "Эконометрика" эта работа выстраивается, в основном, вокруг выполнении контрольной работы № 2, и это справедливо для всех форм обучения. Контрольная работа № 2 охватывает 70 % теоретического материала, она выполняется по индивидуальным исходным данным.

Примерно половину контрольной работы студенты-очники самостоятельно (с консультационной поддержкой) выполняют на практических занятиях в рамках соответствующих тем. Оставшуюся половину они готовят вне аудитории полностью самостоятельно. Сюда включается не только решение задач, но и подготовка ответов на два теоретических вопроса.

Помимо этого, самостоятельная работа выполняется в рамках подготовки к аудиторным практическим занятиям, к сдаче зачёта.

Студенты очно-заочной формы обучения (и для них учебным планом установлено выполнение контрольных работ № 2) также начинают выполнять контрольную работу на практических занятиях, а затем окончательно выполняют её уже вне аудитории самостоятельно. Помимо контрольной работы, студенты очно-заочной формы обучения имеют возможность самостоятельно работать с тестами, контрольными вопросами при подготовке к сдаче зачёта.

СПИСОК ИСТОЧНИКОВ

Основные

1. Кремер, Н. Ш. Эконометрика: учеб. / Н. Ш. Кремер, Б. А. Путко. – Москва: ЮНИТИ-ДАНА, 2016. – 328 с. (ЭБС "Электронная библиотека онлайн").

2. Настин, Ю. Я. Эконометрика: учеб. для бакалавриата по направлению подготовки 380301 "Экономика" / Ю. Я. Настин. – Калининград: Изд-во ФГБОУ ВО "КГТУ", 2021. – 142 с.

3. Настин, Ю. Я. Эконометрика: учеб.-метод. пособие по изучению дисциплины для бакалавриата по напр. подгот. 38.03.01 Экономика, профиль "Прикладная экономика" / Ю. Я. Настин. - Калининград: Изд-во ФГБОУ ВО "КГТУ", 2023. – 120 с.

Дополнительные

4. Карлов, А. М. Теория вероятностей и математическая статистика для экономистов: учеб. пособие / А. М. Карлов. – Москва: КноРус, 2015. – 264 с.

5. Мнацаканян, А. Г. Методические указания по оформлению учебных текстовых работ (рефератов, контрольных, курсовых, выпускных квалификационных) / А. Г. Мнацаканян, Ю. Я. Настин, Э. С. Круглова. – Калининград: Изд-во КГТУ, 2017. – 22 с.

6. Балдин, К. В. Эконометрика: учеб пособие / К. В. Балдин, О. Ф. Быстров, М. Соколов. – Москва: ЮНИТИ-ДАНА, 2015. – 254 с. (ЭБС "Электронная библиотека онлайн").

7. Дубова, Т. А. Статистические методы прогнозирования в экономике: учеб. пособие, практикум, тесты, программа курса, руководство по изучению дисциплины / Т. А. Дубова, М. Ю. Архипова. – Москва: Изд-во МЭСИ, 2004. – 136 с.

8. Орлов, А. И. Эконометрика: учеб. / А. И. Орлов. – Москва: "Экзамен", 2002. – 413 с.

9. Абдуллин, Р. З. Эконометрика в MS EXCEL: практикум / Р. З. Абдуллин, В. Р. Абдуллин. – Иркутск: Изд-во БГУ, 2016. – 135 с.

10. Мхитарян, В. С. Эконометрика: учеб.-практ. пособие / В. С. Мхитарян, М. Ю. Архипова, В. П. Сиротин. – Москва: АНО "Евразийский открытый университет", 2012. – 221 с. (ЭБС "Электронная библиотека онлайн").

11. Сергеев, Л. И. Регрессионный анализ макроэкономических параметров развития рыбной отрасли / Л. И. Сергеев // Балтийский экономический журнал. – 2018. - № 1(21). - С. 86-99.

12. Хачатрян, С. Р. Методы и модели решения экономических задач: учеб. пособие / С. Р. Хачатрян, М. В. Пинегина, В. П. Буянов. – Москва: "Экзамен", 2005. – 384 с.

Исходные данные для контрольных работ № 1 и № 2

А.1 Исходные табличные данные

Таблица А.1 - Исходные данные при первой подходящей букве имени А

X _i	Y _i – первая подходящая буква фамилии														
	Ъ	А	Б	В	Г	Д	Е	К	Л	М	Н	О	П	Р	С
1	2	2	2	2	2	3	3	3	3	2	2	2	2	2	4
1	3	3	3	3	3	3	2	2	2	2	3	3	3	3	5
2	4	4	4	5	5	6	6	4	4	5	5	6	6	4	3
3	5	5	6	6	6	6	7	7	7	4	5	6	6	5	6
4	5	6+ Г1	7+ Г1	8+ Г1	5+ Г1	6+ Г1	7+ Г1	8	5	6	7	8	5	6	8
5	7	8	9	7	8	9	7	8+ Г1	9+ Г1	7+ Г1	8+ Г1	9+ Г1	6+ Г1	5+ Г1	9+Г1
8	14	10	11	12	13	14	15	16	17	18	19	7	8	9	10

Таблица А.2 - Исходные данные при первой подходящей букве имени В

X _i	Y _i – первая подходящая буква фамилии														
	А	Б	В	Г	Д	Е	К	Л	М	Н	О	П	Р	С	
1	2	2	2	2	2	3	3	3	3	2	2	2	2	2	
2	3	3	3	3	3	3	2	2	2	2	3	3	3	3	
4	4	4	4	5	5	6	6	4	4	5	5	6	6	4	
4	5	5	6	6	6	6	7	7	7	6	5	6	6	5	
5	5	6	7	8	5	6	7	8+ Г1	5+ Г1	6+ Г1	7+ Г1	8+ Г1	5+ Г1	6+ Г1	
6	7+ Г1	8+ Г1	9+ Г1	7+ Г1	8+ Г1	9+ Г1	7+ Г1	8	9	7	8	9	11	12	
8	8	9	10	11	12	13	14	15	16	15	14	13	12	11	

Таблица А.3 - Исходные данные при первой подходящей букве имени Е

X _i	Y _i – первая подходящая буква фамилии														
	А	Б	В	Г	Д	Е	К	Л	М	Н	О	П	Р	С	
2	2	2	2	2	2	3	3	3	3	2	2	2	2	2	
3	3	3	3	3	3	3	2	2	2	2	3	3	3	3	
4	5	5	5	5	5	6	6	5	5	5	5	6	6	5	
5	5	5	4	6	6	6	7	7	7	4	5	6	4	5	
7	5+ Г1	6+ Г1	7+ Г1	8+ Г1	5	6	7	8	5	6	7+ Г1	8+ Г1	5+ Г1	6+ Г1	
7	7	8	9	7	8+ Г1	9+ Г1	7+ Г1	8+ Г1	9+ Г1	7+ Г1	8	9	11	12	
9	8	9	10	11	12	13	14	15	16	15	14	13	12	11	

Таблица А.4 - Исходные данные при первой подходящей букве имени **И**

X_i	Y_i – первая подходящая буква фамилии													
	А	Б	В	Г	Д	Е	К	Л	М	Н	О	П	Р	С
И	2	2	2	2	2	3	3	3	3	2	2	2	2	2
2	2	2	2	2	2	3	3	3	3	2	2	2	2	2
4	3	3	3	3	3	3	2	2	2	2	3	3	3	3
4	6	6	6	6	6	6	6	6	6	6	6	6	6	6
6	5	5	4	6	6	6	7	7	7	4	5	6	4	5
6	5	6+ Г1	7	8+ Г1	5+ Г1	6+ Г1	7+ Г1	8+ Г1	5+ Г1	6+ Г1	7	8	5	6
7	7+ Г1	8	9+ Г1	7	8	9	7	8	9	7	8+ Г1	9+ Г1	6+ Г1	5+ Г1
8	14	16	10	10	14	16	11	12	13	14	15	8	12	10

Таблица А.5 - Исходные данные при первой подходящей букве имени **О**

X_i	Y_i – первая подходящая буква фамилии													
	А	Б	В	Г	Д	Е	К	Л	М	Н	О	П	Р	С
О	2	2	2	2	2	3	3	3	3	2	2	2	2	2
1	2	2	2	2	2	3	3	3	3	2	2	2	2	2
3	3	3	3	3	3	3	2	2	2	2	3	3	3	3
5	5	5	5	5	5	6	6	5	5	5	5	6	6	5
5	5+ Г1	5+ Г1	4+ Г1	6+ Г1	6	6	7	7	7	4+ Г1	5+ Г1	6+ Г1	4+ Г1	5+ Г1
6	5	6	7	8	5+ Г1	6+ Г1	7+ Г1	8+ Г1	5+ Г1	6	7	8	5	6
8	7	8	9	7	8	9	7	8	9	7	8	9	11	12
8	16	15	14	13	12	11	8	9	10	11	12	13	14	15

Таблица А.6 - Исходные данные для задачи-2

	X₂ – площадь паркинга, выбор по букве имени					
	Ъ	А	В	Е	И	О
1		2	1	3	1	1
2		2	4	1	3	2
2		4	3	4	4	4
3		5	6	4	4	6
3		5	5	5	5	5
4		3+Г2	4+Г2	6+Г2	5+Г2	6+Г2
6		6+Г2	11	6+Г2	9	8

А.2 Дополнительные исходные данные

Таблицу А.7 преподаватель в начале семестра выдаёт на учебную группу. Здесь она приведена как пример.

Таблица А.7 – Дополнительные исходные данные, пример. *

Название параметра, обозначение	Гр. 23-ОК-1, Иван Петров**		
	1-8	9-17	18-∞
1) дополнительное слагаемое в таблицах А.1-А.5: Г1	0	2	4
2) дополнительное слагаемое в таблице А.6: Г2	4	2	0
3) проектные параметры магазина-8, X10 и X20	11 и 8	12 и 9	13 и 10
4) пороговое значение для формирования временного ряда: П	250	300	350
5) уровень значимости для критических точек распределений Стьюдента, Пирсона, Фишера: α – (альфа)	0,10	0,05	0,10

*Таблица помещается на 2-й странице контрольной работы под таблицей СОДЕРЖАНИЕ.

**Студент из 3-х последних граф удаляет две графы, например, 2-ю и 3-ю, оставляя ту, которая относится к нему.

А.3 Исходные данные для задач 2.3 и 2.4

Таблица А.8 – Поквартальный объём выпуска морепродуктов в сопоставимых ценах, млн. руб.

Квартал	Год				
	1*	2	3	4	5*
I	?	660,0	830,0	750,0	750,0
II	?	670,0	810,0	710,0	?
III	?	680,0	750,0	660,0	?
IV	?	690,0	780,0	690,0	?

*Алгоритм заполнения клеток со знаками ? индивидуальными данными см. в подразделе 8.4.

А.4 Исходные данные для задачи 2.5

Таблица А.9 – Валовой региональный продукт по годам за 10 лет, млрд. руб.

Год	ВРП	Год	ВРП	Год	ВРП
1	?	5	660,0	9	830,0
2	?	6	670,0	10	810,0
3	?	7	680,0	-	-
4	?	8	690,0	-	-

*Правило заполнения клеток со знаками ? индивидуальными данными см. в подразделе 8.6.

Образец страницы СОДЕРЖАНИЕ для контрольных работ № 1

СОДЕРЖАНИЕ

1	Вопрос 11. <Название вопроса>	3
2	Вопрос 31. <Название вопроса>.....	5
3	Задача-1: "Моделирование магазина на основе парной линейной регрессии" - ручное решение	7
4	Задача-2: "Моделирование магазина на основе троичной линейной регрессии" - ручное решение	11
5	Задача-3: "Моделирование магазина на основе троичной линейной регрессии в MS Excel"	14
6	Задача-4. "Моделирование магазина на основе нелинейной регрессии в MS Excel".	17
	СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ.....	20

Методические указания

- 1) Страницы проставлены ориентировочно.
- 2) Границы после заполнения таблицы убрать.
- 3) Можно эту страницу сформировать иначе – средствами редактора.
- 4) Здесь межстрочный интервал 1,5, так предписано ГОСТом и нашими правилами.
- 5) Использование прописных букв здесь и в тексте контрольных работ предписано ГОСТом и правилами.

Образец страницы СОДЕРЖАНИЕ для контрольных работ № 2

СОДЕРЖАНИЕ

1*	Вопрос 1.12. <Название вопроса>.....	3
2*	Вопрос 2.19. <Название вопроса>.....	5
3*	Задача-2.1: "Проверка возможности объединения выборок для линейной парной модели магазина" - ручное решение.....	
4	Задача-2.2: "Проверка линейной парной модели магазина на гомоскедастичность по двум критериям" – ручное решение	7
5	Задача-2.3: "Выделение тренда временного ряда методами скользящей средней в MS Excel"	11
6	Задача-2.4: "Выделение тренда временного ряда методом экспоненциального сглаживания в MS Excel”	14
7	Задача-2.5: "Прогнозирование на основе авторегрессии 1-го порядка в MS Excel"	17
	СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ.....	20

*Для определённости к номерам вопросов, к слову Задача впереди добавлять 1 или 2 – номер части дисциплины (всего 20 вопросов из 1-й части и 20 вопросов из 2-й части).

Методические указания

- 1) Страницы проставлены ориентировочно.
- 2) Границы после заполнения таблицы убрать.
- 3) Можно эту страницу сформировать иначе – средствами редактора.
- 4) Здесь межстрочный интервал 1,5, так предписано ГОСТом и нашими правилами.
- 5) Использование прописных букв здесь и в тексте контрольных работ предписано ГОСТом и правилами.

Построение графика линейной парной регрессии в Word через Excel

Работа в Excel

1) Создать таблицу с исходными данными в левом верхнем углу от A1 до B8 (слева X1, справа Y).

1	2
2	3
2	4
3	5
4	5
5	7
8	14

2) Выделить таблицу.

3) Нажать "Вставка".

4) Нажать "Точечная", затем 1-ю иконку (точечная с маркерами), получим рисунок с полем корреляции.

5) Нажать сверху иконку 3-ю слева: сетка с прямой линией, получим искомый график с регрессией – прямой линией и точками.

6) Курсор в белое поле рисунка, правой клавишей вызвать меню.

7) Вырезать рисунок.

Работа в Word

8) Вставить рисунок в нужное место ворд-файла.

9) Вписать названия координат, легенду лучше удалить.

10) Можно поэкспериментировать: раздвинуть границы, применить заливку и цвета и т. п.

Некоторые формулы из математической статистики и эконометрики

В приведённых ниже формулах акцент делается на анализе смещённости оценок параметров ("плохие") и несмещённости "хорошие". Это особенно важно в статистике коротких выборок, где n не более 30, а в наших примерах $n=7$! Вопрос: почему учебники используют, в основном, только смещённые оценки? Ответ: а) при $n>30$ разницы между оценками, практически, нет; б) формулы для смещённых оценок проще и "красивее".

Обозначения:

- $D1_x = s1_x^2$ – оценка дисперсии СВ X смещённая ("плохая");
- $D2_x = s2_x^2$ – оценка дисперсии СВ X несмещённая (исправленная, "хорошая");
- $Cov1(x,y)$ – оценка ковариации двух СВ смещённая ("плохая");
- $Cov2(x,y)$ – оценка ковариация двух СВ несмещённая (исправленная, "хорошая");

Пример. Дано: выборки двух СВ: $X=(1, 2, 3)$ и $Y=(6, 0, 3)$. Рассчитаем для них некоторые статистики.

- 1) Средние арифметические: $\bar{x} = (1+2+3)/3 = 2$; $\bar{y} = (6+0+3)/3 = 3$.
- 2) Центрированные выборки: $\dot{x} = (1-2, 2-2, 3-2) = (-1, 0, 1)$; $\dot{y} = (3, -3, 0)$.
- 3) Среднее произведения: $\overline{XY} = (1 \cdot 6 + 2 \cdot 0 + 3 \cdot 3)/3 = 5$.
- 4) Среднее квадрата: $\overline{x^2} = (1^2 + 2^2 + 3^2)/3 = 4,67$.

Таблица Д.1 – Некоторые полезные формулы и преобразования.

Формула	Комментарий и пример
1	2
1) $D1_x = s1_x^2 = \Sigma(x_i - \bar{x})^2 / n = \Sigma \dot{x}_i^2 / n$; $s1_x^2 = (\overline{x^2} - \bar{x}^2)$	Дисперсия, смещённая из-за деления на n , а не на $n-1$: $s1_x^2 = (1+0+1)/3 = 0,667$; $s1_y^2 = (9+9+0)/3 = 6$
2) $D2_x = s2_x^2 = \Sigma(x_i - \bar{x})^2 / (n-1) = \Sigma \dot{x}_i^2 / (n-1)$; $s2_x^2 = s1_x^2 \cdot n / (n-1) = (\overline{x^2} - \bar{x}^2) n / (n-1)$.	Дисперсия, несмещённая либо а) из-за деления на $n-1$, либо б) умножения $s1_x^2$ на поправочный коэффициент $n/n-1$: а) $s2_x^2 = (1+0+1)/(3-1) = 1$; б) $s2_y^2 = 6 \cdot 3/(3-1) = 9$. Сравни значения: 0,667 и 1, 6 и 9!
3) $Cov1(x,y) = \Sigma(x_i - \bar{x})(y_i - \bar{y}) / n = \Sigma \dot{x}_i \cdot \dot{y}_i / n$; $Cov1(x,y) = \overline{xy} - \bar{x}\bar{y}$.	Ковариация, смещённая из-за деления на n , а не на $n-1$: $Cov1(x,y) = (-1 \cdot 3 + 0 \cdot (-3) + 1 \cdot 0) / 3 = -1$.
4) $Cov2(x,y) = \Sigma(x_i - \bar{x})(y_i - \bar{y}) / (n-1) = \Sigma \dot{x}_i \cdot \dot{y}_i / (n-1)$; $Cov2(x,y) = Cov1(x,y) \cdot n / (n-1) = (\overline{xy} - \bar{x}\bar{y}) \cdot n / (n-1)$.	Ковариация, несмещённая из-за: а) деления на $(n-1)$ либо б) умножения $Cov1(x,y)$ на поправочный коэффициент $n/n-1$: $Cov2(x,y) = (-1 \cdot 3 + 0 \cdot (-3) + 1 \cdot 0) / (3-1) = -1,5$

1	2
5) $b_1 = (\overline{xy} - \bar{x}\bar{y}) / (\overline{x^2} - \bar{x}^2) = [(\overline{xy} - \bar{x}\bar{y})n / (n-1)] / [(\overline{x^2} - \bar{x}^2)n / (n-1)]$; $b_1 = \text{Cov}1(x,y) / s1_x^2 = \text{Cov}2(x,y) / s2_x^2$	Важно: МНК-оценка (1.14) коэффициента парной регрессии <i>инвариантна</i> - она несмещённая, если числитель и знаменатель однотипные: оба смещённые либо оба несмещённые. Покажем это: а) оба смещённые: $(\overline{xy} - \bar{x}\bar{y}) / (\overline{x^2} - \bar{x}^2) = (5 - 2 \cdot 3) / (4,67 - 2^2) \approx -1,5$; б) оба несмещённые: $\text{Cov}2(x,y) / s2_x^2 = -1,5 / 1 = -1,5$
6) $r_{xy} = (\sum \dot{x}_i \cdot \dot{y}_i / n) / (s_x s_y) = \text{Cov}(x,y) / (s_x s_y) = \sum \dot{x}_i \cdot \dot{y}_i / (\sqrt{\sum \dot{x}_i^2} \cdot \sqrt{\sum \dot{y}_i^2})$	Оценка коэффициента r_{xy} парной корреляции инвариантна – она несмещённая, если числитель и знаменатель однотипные: оба смещённые либо оба несмещённые

Статистические таблицы (учебные)

Таблица Е.1 - Критические значения $t_{1-\alpha, k}$ -критерия Стьюдента

Число степе- ней свободы k	Уровень значимости α		Число степе- ней свободы k	Уровень значимости α	
	0,10	0,05		0,10	0,05
1	6,31	12,71	6	1,94	2,45
2	2,92	4,40	7	1,89	2,36
3	2,35	3,18	8	1,86	2,31
4	2,13	2,78	9	1,83	2,26
5	2,02	2,57	10	1,81	2,23
			15	1,75	2,13

Таблица Е.2 - Критические значения $\chi^2_{\alpha, k}$ -критерия Пирсона

Число степе- ней свободы k	Уровень значимости α		Число степе- ней свободы k	Уровень значимости α	
	0,10	0,05		0,10	0,05
1	2	3	1	2	3
1	2,71	12,71	6	10,6	12,6
2	4,60	4,40	7	12,0	14,1
3	6,25	3,18	8	13,4	15,5
4	7,78	2,78	9	14,7	16,9
5	9,24	2,57	10	16,0	18,3
-	-	-	15	22,3	25,0

Таблица Е.3 - Критические значения F-критерия Фишера при $\alpha=0,05$

k2 \ k1	1	2	3	4	5	6	7	8
1	161,5	199,5	215,7	224,6	230,2	233,9	237	238,9
2	18,5	19,0	19,2	19,3	19,3	19,3	19,4	19,4
3	10,1	9,6	9,3	9,1	9,0	8,9	8,9	8,8
4	7,7	6,9	6,6	6,4	6,3	6,2	6,1	6,0
5	6,6	5,8	5,4	5,2	5,1	5,0	4,9	4,8
6	6,0	5,1	4,8	4,5	4,4	4,3	4,2	4,2
7	5,6	4,7	4,4	4,1	4,0	3,9	3,8	3,7
8	5,3	4,5	4,1	3,8	3,7	3,6	3,5	3,4

k1 и k2 – числа степеней свободы большей и меньшей дисперсии.

Таблица Е.4 - Критические значения F-критерия Фишера при $\alpha=0,10$

k2 \ k1	1	2	3	4	5	6	7	8
1	2	3	4	5	6	7	8	9
1	39,9	49,5	54,0	55,8	57,2	58,2	58,9	59,4
2	8,5	9,0	9,2	9,2	9,3	9,3	9,4	9,4
3	5,5	5,5	5,4	5,3	5,3	5,3	5,3	5,3
4	4,5	4,3	4,2	4,1	4,1	4,0	4,0	4,0
5	4,1	3,8	3,6	3,5	3,5	3,4	3,4	3,3
6	3,8	3,5	3,3	3,2	3,1	3,1	3,0	3,0
7	3,6	3,3	3,1	3,0	2,9	2,8	2,8	2,8
8	3,5	3,1	2,9	2,8	2,7	2,6	2,6	2,6

k1 и k2 – числа степеней свободы большей и меньшей дисперсии.

Таблица Е.5 - Критические значения Дарбина-Уотсона при $\alpha=0,05$

n	p=1		p=2		p=3		p=4	
	d_n	d_b	d_n	d_b	d_n	d_b	d_n	d_b
6	0,61	1,40	-	-	-	-	-	-
7	0,70	1,36	0,47	1,90	-	-	-	-
8	0,76	1,33	0,36	1,78	0,37	2,29	-	-
9	0,82	1,32	0,63	1,70	0,44	2,13	0,30	2,39

n - число наблюдений, p – число факторных переменных, d_n и d_b – нижнее и верхнее значения.

Локальный электронный методический материал

Юрий Яковлевич Настин

ЭКОНОМЕТРИКА

Редактор Э. С. Круглова

Уч.-изд. л. 7,5 Печ. л. 6,7

Федеральное государственное бюджетное
образовательное учреждение высшего образования
"Калининградский государственный технический университет".
236022, Калининград, Советский проспект, 1