



МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Магнитогорский государственный технический университет им. Г.И. Носова»



УТВЕРЖДАЮ
Директор ИЭиАС
В.Р. Храмин

03.03.2021 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

ЭКОНОМЕТРИКА

Направление подготовки (специальность)
09.03.03 Прикладная информатика

Направленность (профиль/специализация) программы
Разработка компьютерных игр и приложений виртуальной/дополненной реальности

Уровень высшего образования - бакалавриат

Форма обучения
очная

Институт/ факультет	Институт энергетики и автоматизированных систем
Кафедра	Бизнес-информатики и информационных технологий
Курс	2
Семестр	4

Магнитогорск
2021 год

Рабочая программа составлена на основе ФГОС ВО - бакалавриат по направлению подготовки 09.03.03 Прикладная информатика (приказ Минобрнауки России от 19.09.2017 г. № 922)

Рабочая программа рассмотрена и одобрена на заседании кафедры Бизнес-информатики и информационных технологий

18.02.2021 г. протокол № 6

Зав. кафедрой  Г.Н. Чусавитина

Рабочая программа одобрена методической комиссией ИЭиАС
03.03.2021 г. протокол № 5

Председатель  В.Р. Храмшин

Рабочая программа составлена:

доцент кафедры БИиИТ, канд. пед. наук  М.В. Романова

Рецензент:

Генеральный директор ООО «Корпоративные системы Плюс»,

 Ю.А. Чудинова

Лист актуализации рабочей программы

Рабочая программа пересмотрена, обсуждена и одобрена для реализации в 2022 - 2023 учебном году на заседании кафедры Бизнес-информатики и информационных

Протокол от _____ 20__ г. № ____
Зав. кафедрой _____ Г.Н. Чусавитина

Рабочая программа пересмотрена, обсуждена и одобрена для реализации в 2023 - 2024 учебном году на заседании кафедры Бизнес-информатики и информационных

Протокол от _____ 20__ г. № ____
Зав. кафедрой _____ Г.Н. Чусавитина

Рабочая программа пересмотрена, обсуждена и одобрена для реализации в 2024 - 2025 учебном году на заседании кафедры Бизнес-информатики и информационных

Протокол от _____ 20__ г. № ____
Зав. кафедрой _____ Г.Н. Чусавитина

Рабочая программа пересмотрена, обсуждена и одобрена для реализации в 2025 - 2026 учебном году на заседании кафедры Бизнес-информатики и информационных

Протокол от _____ 20__ г. № ____
Зав. кафедрой _____ Г.Н. Чусавитина

1 Цели освоения дисциплины (модуля)

обучение студентов основным математическим понятиям и методам в области эконометрического моделирования, применительно к решению задач анализа, прогнозирования и управления экономикой на макро - и микро - уровнях. В процессе изучения этой дисциплины у студентов должны быть сформированы теоретические и практические навыки решения эконометрических задач и анализа полученных результатов.

2 Место дисциплины (модуля) в структуре образовательной программы

Дисциплина Эконометрика входит в часть учебного плана формируемую участниками образовательных отношений образовательной программы.

Для изучения дисциплины необходимы знания (умения, владения), сформированные в результате изучения дисциплин/ практик:

Прикладная математика

Методологии и инструментальные средства моделирования бизнес-процессов и данных

Знания (умения, владения), полученные при изучении данной дисциплины будут необходимы для изучения дисциплин/практик:

Методы научных исследований в сфере ИКТ

Подготовка к сдаче и сдача государственного экзамена

3 Компетенции обучающегося, формируемые в результате освоения дисциплины (модуля) и планируемые результаты обучения

В результате освоения дисциплины (модуля) «Эконометрика» обучающийся должен обладать следующими компетенциями:

Код индикатора	Индикатор достижения компетенции
ОПК-1	Способен применять естественнонаучные и общеинженерные знания, методы математического анализа и моделирования, теоретического и экспериментального исследования в профессиональной деятельности;
ОПК-1.1	Решает стандартные профессиональные задачи с применением естественнонаучных и общеинженерных знаний, методов математического анализа и моделирования
ОПК-1.2	Решает профессиональные задачи с применением методов теоретического и экспериментального исследования

4. Структура, объём и содержание дисциплины (модуля)

Общая трудоемкость дисциплины составляет 4 зачетных единиц 144 акад. часов, в том числе:

- контактная работа – 51,95 акад. часов;
- аудиторная – 51 акад. часов;
- внеаудиторная – 0,95 акад. часов
- самостоятельная работа – 92,05 акад. часов;

Форма аттестации - зачет

Раздел/ тема дисциплины	Семестр	Аудиторная контактная работа (в акад. часах)			Самостоятельная работа студента	Вид самостоятельной работы	Форма текущего контроля успеваемости и промежуточной аттестации	Код компетенции
		Лек.	лаб. зан.	практ. зан.				
1. Введение в предмет								
1.1 Предмет эконометрики, её цель, задачи. Базовые понятия	4	1/0,5И			1	Изучение литературы, подготовка к выполнению лабораторных работ	устный опрос Отчет по лабораторной работе	ОПК-1.1
1.2 Повторение теории вероятности и математической статистики		0,5	2/0,5И		4	Изучение литературы, подготовка к семинару	устный опрос доклад к семинару	ОПК-1.1, ОПК-1.2
1.3 Математическое моделирование в экономике. Классы моделей. Виды переменных. Этапы эконометрического моделирования		1,5/0,5И			2	Изучение литературы	устный опрос	ОПК-1.1, ОПК-1.2
Итого по разделу		3/1И	2/0,5И		7			
2. Регрессионные модели								
2.1 Линейная регрессия.	4	2/0,5И	6/1И		8	Изучение литературы, подготовка к выполнению лабораторных работ Прохождения МООК	отчет по лабораторным работам отчет по прохождению модуля МООК	ОПК-1.1, ОПК-1.2
2.2 Нелинейные модели регрессии		4/1И	8/3И		12	Изучение литературы, подготовка к выполнению лабораторных работ Прохождение МООК	Отчет по лабораторной работе отчет о прохождении модуля МООК	ОПК-1.1, ОПК-1.2
Итого по разделу		6/1,5И	14/4И		20			

3. Временные ряды								
3.1 Основные элементы временного ряда	4	2/0,5И	4/1И		16	Изучение литературы, подготовка к выполнению лабораторных работ Прохождение модуля МООК	Отчет по лабораторной работе Отчет о прохождении модуля МООК	ОПК-1.1, ОПК-1.2
3.2 Модели стационарных и нестационарных рядов		2/0,5И	6/3И		16	Изучение литературы, подготовка к выполнению лабораторных работ Прохождение модуля МООК	Отчет по лабораторной работе Отчет о прохождении модуля МООК	ОПК-1.1, ОПК-1.2
Итого по разделу		4/1И	10/4И		32			
4. Системы эконометрических уравнений								
4.1 Системы уравнений в эконометрике	4	2/0,5И	4/1И		16	Изучение литературы, подготовка к выполнению лабораторных работ прохождение модуля МООК	Отчет по лабораторной работе отчет о прохождении модуля МООК	ОПК-1.1, ОПК-1.2
4.2 Системы линейных одновременных уравнений		2/0,3И	4/0,4И		16	Изучение литературы, подготовка к выполнению лабораторных работ прохождение модуля МООК	Отчет по лабораторной работе отчет о прохождении модуля МООК	ОПК-1.1, ОПК-1.2
Итого по разделу		4/0,8И	8/1,4И		32			
5. Зачет								
5.1 Зачет	4				1,05	Изучение литературы, подготовка к зачету	тест	ОПК-1.1, ОПК-1.2
Итого по разделу					1,05			
Итого за семестр		17/4,3И	34/9,9И		92,05		зачёт	
Итого по дисциплине		17/4, 3И	34/9,9 И		92,05		зачет	

5 Образовательные технологии

При проведении занятий и организации самостоятельной работы используются:

Традиционные технологии обучения, предполагающие передачу информации в готовом виде, формирование учебных умений по образцу: лекция-изложение, лекция-объяснение, лабораторные работы, контрольная работа и др.

Использование традиционных технологий обеспечивает ориентирование студента в потоке информации, связанной с различными подходами к определению сущности, содержания, методов, форм развития и саморазвития личности; самоопределение в выборе оптимального пути и способов личностно-профессионального развития; систематизацию знаний, полученных студентами в процессе аудиторной и самостоятельной работы. Лабораторные занятия обеспечивают развитие и закрепление умений и навыков определения целей и задач саморазвития, а также принятия наиболее эффективных решений по их реализации.

Интерактивные формы обучения, предполагающие организацию обучения как продуктивной творческой деятельности в режиме взаимодействия студентов друг с другом и с преподавателем.

Использование интерактивных образовательных технологий способствует повышению интереса и мотивации учащихся, активизации мыслительной деятельности и творческого потенциала студентов, делает более эффективным усвоение материала, позволяет индивидуализировать обучение и ввести экстренную коррекцию знаний.

При проведении лабораторных занятий используются групповая работа, технология коллективной творческой деятельности, технология сотрудничества, обсуждение проблемы в форме дискуссии, Case-study. Данные технологии обеспечивают высокий уровень усвоения студентами знаний, эффективное и успешное овладение умениями и навыками в предметной области, формируют познавательную потребность и необходимость дальнейшего самообразования, позволяют активизировать исследовательскую деятельность, обеспечивают эффективный контроль усвоения знаний.

6 Учебно-методическое обеспечение самостоятельной работы обучающихся

Представлено в приложении 1.

7 Оценочные средства для проведения промежуточной аттестации

Представлены в приложении 2.

8 Учебно-методическое и информационное обеспечение дисциплины (модуля)

а) Основная литература:

1. Эконометрика : учебник для вузов / И. И. Елисеева [и др.]. — Москва : Издательство Юрайт, 2020. — 449 с. — (Высшее образование). — ISBN 978-5-534-00313-0. — Текст : электронный // ЭБС Юрайт [сайт]. — URL: <https://urait.ru/bcode/449677>

б) Дополнительная литература:

1. Костюнин, В. И. Эконометрика : учебник и практикум для вузов / В. И. Костюнин. — Москва : Издательство Юрайт, 2020. — 285 с. — (Высшее образование). — ISBN 978-5-534-02660-3. — Текст : электронный // ЭБС Юрайт [сайт]. — URL: <https://urait.ru/bcode/450113>

2. Демидова, О. А. Эконометрика : учебник и практикум для среднего профессионального образования / О. А. Демидова, Д. И. Малахов. — Москва : Издательство Юрайт, 2020. — 334 с. — (Профессиональное образование). — ISBN 978-5-534-13226-7. — Текст : электронный // ЭБС Юрайт [сайт]. — URL:

<https://urait.ru/bcode/449565>

в) Методические указания:

1. Шанченко, Н.И. Эконометрика [Электронный ресурс]: лабораторный практикум: учебное пособие / Н. И. Шанченко. – Ульяновск: УлГТУ, 2011. – 117 с. – Режим доступа: <http://window.edu.ru/resource/537/74537/files/ulstu2011-90.pdf>

г) Программное обеспечение и Интернет-ресурсы:

Программное обеспечение

Наименование ПО	№ договора	Срок действия лицензии
MS Windows 7 Professional(для классов)	Д-1227-18 от 08.10.2018	11.10.2021
MS Office 2007 Professional	№ 135 от 17.09.2007	бессрочно
7Zip	свободно	бессрочно

Профессиональные базы данных и информационные справочные системы

Название курса	Ссылка
Национальная информационно-аналитическая система – Российский индекс научного цитирования (РИНЦ)	URL: https://elibrary.ru/project_risc.asp
Поисковая система Академия Google (Google Scholar)	URL: https://scholar.google.ru/
Информационная система - Единое окно доступа к информационным ресурсам	URL: http://window.edu.ru/
Электронные ресурсы библиотеки МГТУ им. Г.И. Носова	http://magtu.ru:8085/marcweb2/Default.asp

9 Материально-техническое обеспечение дисциплины (модуля)

Материально-техническое обеспечение дисциплины включает:

Материально-техническое обеспечение дисциплины включает:

Учебные аудитории для проведения занятий лекционного типа - Специализированная (учебная) мебель (столы, стулья, доска аудиторная), мультимедийное оборудование (проектор, компьютер, экран) для презентации учебного материала по дисциплине;

Учебные аудитории для проведения лабораторных занятий, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации - Специализированная (учебная) мебель (столы, стулья, доска аудиторная), персональные компьютеры объединенные в локальные сети с выходом в Internet и с доступом в электронную информационно-образовательную среду университета, оснащенные современными программно-методическими комплексами

Аудитории для самостоятельной работы (компьютерные классы; читальные залы библиотеки) - Специализированная (учебная) мебель (столы, стулья, доска аудиторная), персональные компьютеры объединенные в локальные сети с выходом в Internet и с доступом в электронную информационно-образовательную среду университета, оснащенные современными программно-методическими комплексами

Помещение для хранения и профилактического обслуживания учебного оборудования - Мебель (столы, стулья, стеллажи для хранения учебно-наглядных пособий и учебно-методической документации), персональные компьютеры.

Учебно-методическое обеспечение самостоятельной работы обучающихся

Раздел 1. Построение модели линейной парной регрессии

Лабораторная работа №1.

Вычисление коэффициентов уравнения линейной регрессии

Парная регрессия – это уравнение связи двух переменных y и x :

$$y = f(x),$$

где y - зависимая переменная (результативный признак); x - независимая, объясняющая переменная (признак - фактор).

Линейная парная регрессия: $y = a + bx + \varepsilon$.

Построение уравнения регрессии сводится к оценке ее параметров (в приведенных выше уравнениях регрессии – параметры a и b). Для оценки параметров регрессий, линейных по параметрам, используют **метод наименьших квадратов** (МНК). МНК позволяет получить такие оценки параметров, при которых сумма квадратов отклонений фактических значений результативного признака y от теоретических $\hat{y}(x)$ минимальна, т.е.

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

Для линейных уравнений решается следующая система относительно a и b :

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

Для решения данной системы и нахождения a и b можно использовать два подхода:

1. Используются следующие соотношения:

$$\Delta = \begin{vmatrix} n & \sum x \\ \sum x & \sum x^2 \end{vmatrix}, \quad \Delta a = \begin{vmatrix} \sum y & \sum x \\ \sum yx & \sum x^2 \end{vmatrix}, \quad \Delta b = \begin{vmatrix} n & \sum y \\ \sum x & \sum yx \end{vmatrix},$$

$$a = \frac{\Delta a}{\Delta}; \quad b = \frac{\Delta b}{\Delta}$$

2. Можно воспользоваться готовыми формулами, которые вытекают из этой системы:

$$a = \bar{y} - b\bar{x}, \tag{1}$$

$$b = \frac{Cov(x, y)}{\sigma_x^2} = \frac{\overline{y \cdot x} - \bar{y} \cdot \bar{x}}{\overline{x^2} - \bar{x}^2}; \quad (2)$$

где $Cov(x, y) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$ - коэффициент выборочной ковариации;

$\sigma_x^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$ - дисперсия факторного признака или выборочная вариация, обозначаемая как $Var(x)$.

Задача

Имеются статистические данные поступления договоров в консолидированный бюджет Санкт-Петербурга (y , млрд руб.) в зависимости от численности работающих на крупных и средних предприятиях (x , тыс. чел.) экономики районов за 2012 г. представлены в таблице 1.

Таблица 1

№ района города	Поступление договоров в бюджет у (млрд. руб.)	Численность работающих x (тыс. чел.)
1	4,4	3,3
2	8,1	6,2
3	12,9	8,1
4	20,8	18,3
5	15,5	20,2
6	28,8	23,1
7	37,5	39
8	48,7	49,1
9	68,6	60,1
10	104,6	74,2
11	90,5	79
12	88,3	95
13	132,4	106
14	122	112,2
15	99,1	115
16	144,2	125,1
17	150,6	132
18	156,1	149
19	209,5	157
20	342,9	282

Необходимо построить линейное уравнение парной регрессии y от x .

Решение.

1. Для расчета параметров уравнения линейной регрессии постройте расчетную таблицу 2.

Таблица 2

№ района	y	x	xy	x^2	y^2
1	4,4	3,3			

2	8,1	6,2			
3	12,9	8,1			
4	20,8	18,3			
5	15,5	20,2			
6	28,8	23,1			
7	37,5	39			
8	48,7	49,1			
9	68,6	60,1			
10	104,6	74,2			
11	90,5	79			
12	88,3	95			
13	132,4	106			
14	122	112,2			
15	99,1	115			
16	144,2	125,1			
17	150,6	132			
18	156,1	149			
19	209,5	157			
20	342,9	282			
Итого					
Средние значения	\bar{y}	\bar{x}	\overline{xy}	$\overline{x^2}$	$\overline{y^2}$

2. Используя формулы 1, 2 рассчитайте параметры a и b .

3. Запишите полученной уравнение парной регрессии.

4. Для проверки правильности полученного уравнения воспользуйтесь статистической функцией ЛИНЕЙН табличного процессора Microsoft Excel, для этого:

- выделите интервал ячеек, содержащий 2 столбца и 5 строк;
- вставьте функцию ЛИНЕЙН, задайте значение константы и статистики равными единице.
- введите формулу как формулу массива, для чего нажмите клавишу $F2$ а затем комбинацию клавиш $Ctrl+Shift+Enter$.

В таблице 3. показано, в каком порядке возвращается дополнительная регрессионная статистика.

Таблица 3

b	a
se_b	se_a
r^2	se_y
F	df
$SS_{рег.}$	$SS_{ост.}$

Лабораторная работа №2.

Вычисление коэффициента корреляции и средней ошибки аппроксимации

Тесноту связи изучаемых явлений оценивает **линейный коэффициент парной корреляции** r_{xy} :

$$r_{xy} = b \frac{\sigma_x}{\sigma_y} = \frac{Cov(x, y)}{\sigma_x \sigma_y} = \frac{\overline{yx} - \bar{y} \cdot \bar{x}}{\sigma_x \sigma_y} ,$$

где $\sigma_x = \sqrt{\overline{x^2} - \bar{x}^2}$, $\sigma_y = \sqrt{\overline{y^2} - \bar{y}^2}$.

Линейный коэффициент корреляции изменяется в пределах $-1 \leq r_{xy} \leq 1$.

Равенство коэффициента нулю свидетельствует об отсутствии **линейной связи**. Равенство коэффициента -1 или +1 показывает наличие функциональной связи. Знак «+» указывает на **связь прямую** (увеличение или уменьшение одного признака сопровождается аналогичным изменением другого признака), знак «-» указывает на **связь обратную** (увеличение или уменьшение одного признака сопровождается противоположным по направлению изменением другого признака). Показателям тесноты связи можно дать качественную оценку:

Количественная мера тесноты связи	Качественная характеристика силы связи
до 0,25	Очень слабая
0,25 – 0,44	Слабая
0,45 – 0,64	Умеренная
0,65 – 0,80	Сильная
0,81 – 0,90	Весьма сильная
0,91 – 1,00	Очень сильная

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

Коэффициент детерминации r_{xy}^2 :

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

Коэффициент детерминации принимает значения от 0, когда факторы x не оказывают никакого влияния на зависимую переменную, до 1, когда изменения зависимой переменной y полностью объяснимы влиянием факторов модели.

Функциональная связь возникает при значении равном 1, а отсутствие связи - 0. При значениях показателей тесноты связи меньше 0,7 величина коэффициента детерминации

σ^2								
r_{xy}								
r_{xy}^2								
\bar{A}								

2. Сравните полученное значение коэффициента детерминации со значением, вычисленным функцией ЛИНЕЙН.

3. Сделайте выводы о тесноте связи и об оценке качества построенной модели.

Лабораторная работа №3.

Оценка значимости уравнения регрессии с помощью F-критерия Фишера

F-критерий Фишера заключается в проверке гипотезы H_0 о статистической незначимости уравнения регрессии. Для этого выполняется сравнение фактического $F_{факт}$ и критического (табличного) $F_{табл}$ значений F-критерия Фишера.

$F_{факт}$ определяется из соотношения значений факторной и остаточной дисперсий, рассчитанных на одну степень свободы:

$$F_{факт.} = \frac{\sum \frac{(\hat{y} - \bar{y})^2}{m}}{\sum \frac{(y - \hat{y})^2}{n - m - 1}} = \frac{r_{xy}^2}{1 - r_{xy}^2} \cdot \frac{n - m - 1}{m}, \quad (3)$$

где n – число единиц совокупности; m – число параметров при переменных (для линейной регрессии $m = 1$).

$F_{табл}$ – максимально возможное значение критерия под влиянием случайных факторов при степенях свободы $k_1 = m, k_2 = n - m - 1$ (для линейной регрессии $m = 1$) и уровне значимости α .

Уровень значимости α – вероятность отвергнуть правильную гипотезу при условии, что она верна. Обычно величина α принимается равной 0,05 или 0,01.

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

Задача

Определим F-критерий Фишера для уравнения парной регрессии, полученного в Лабораторной работе №1.

Решение.

1. Рассчитайте $F_{факт}$ по формуле 3 и сравните полученное значение со значением, рассчитанным с помощью функции ЛИНЕЙН в Лабораторной работе №1.
2. Определите $F_{табл}$ при заданном уровне значимости $\alpha = 0,05$ (Приложение 1). Критическое значение также можно найти с помощью функции Excel ФРАСПОБР.
3. Сравните полученные значения $F_{табл}$ и $F_{факт}$, сделайте вывод о существенности связи.

Лабораторная работа №4.

Оценка статистической значимости параметров регрессии и корреляции

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

$$t_b = \frac{b}{se_b}; \quad t_a = \frac{a}{se_a}; \quad t_r = \frac{r_{xy}}{se_{r_{xy}}}. \quad (4)$$

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

$$se_b = \sqrt{\frac{\sum (y - \hat{y}_x)^2 / (n-2)}{\sum (x - \bar{x})^2}} = \sqrt{\frac{S^2_{ocm.}}{\sum (x - \bar{x})^2}} = \frac{S_{ocm.}}{\sigma_x \sqrt{n}}; \quad (5)$$

$$se_a = \sqrt{\frac{\sum (y - \hat{y}_x)^2}{n-2} \cdot \frac{\sum x^2}{n \sum (x - \bar{x})^2}} = \sqrt{S^2_{ocm.} \cdot \frac{\sum x^2}{n^2 \sigma_x^2}} = S_{ocm.} \cdot \frac{\sqrt{\sum x^2}}{n \sigma_x}; \quad (6)$$

$$se_{r_{xy}} = \sqrt{\frac{1 - r_{xy}^2}{n - 2}}. \quad (7)$$

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

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

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

Если $t_{табл} < t_{факт}$ или $t_{факт} < -t_{табл}$, то H_0 отклоняется, т.е. a , b и r_{xy} не случайно отличаются от нуля и сформировались под влиянием систематически действующего фактора x .

Если $t_{табл} > t_{факт}$ или $t_{факт} > -t_{табл}$, то гипотеза H_0 не отклоняется и признается случайная природа формирования a , b или r_{xy} .

Задача

Оценим статистическую значимость коэффициентов регрессии, рассчитанных в Лабораторной работе №1.

Решение.

1. Примите гипотезу H_0 о статистически незначимом отличии показателей от нуля $a = b = r_{xy} = 0$.
2. Рассчитайте величины случайных ошибок, используя формулы 4, 5, 6, 7. Сравните полученные значения со значениями, рассчитанными с помощью функции ЛИНЕЙН в Лабораторной работе №1.
3. Определите $t_{\text{табл}}$ при заданном уровне значимости $\alpha = 0,05$ и степени свободы $k_2 = n - m - 1$ (Приложение 2). Критическое значение также можно найти с помощью функции Excel СТЬЮДРАСПОБР.
4. Сравните полученные значения $t_{\text{табл}}$ и $t_{\text{факт}}$, сделайте вывод о статистической значимости параметров регрессии.

Лабораторная работа №5.

Расчет доверительных интервалов

Заметим, что все оценки, приведенные ранее, определяются одним числом, т.е. являются точечными. При малых объемах выборки точечная оценка может приводить к большим ошибкам и значительно отличаться от оцениваемого параметра. Более широкое применение получил метод доверительных интервалов, разработанный американским статистиком Ю. Нейманом.

Таким образом, рассчитанные значения показателей (коэффициенты a , b , r_{xy}) являются приближенными, полученными на основе имеющихся выборочных данных. Для оценки того, насколько точные значения показателей могут отличаться от рассчитанных, осуществляется построение доверительных интервалов. Доверительные интервалы определяют пределы, в которых лежат точные значения определяемых показателей с заданной степенью уверенности, соответствующей заданному уровню значимости α . При построении эконометрических регрессионных моделей доверительные интервалы обычно определяют для двух уровней значимости - $\alpha = 0,05$ и $\alpha = 0,01$. Соответственно говорят о 5-процентном уровне значимости или об 1-процентном уровне значимости.

Доверительные интервалы (уровни доверия) при этом будут равны $1 - \alpha = 0,95$ и $1 - \alpha = 0,99$. Соответственно говорят о 95-процентном или о 99-процентном уровне доверия (надежности). Подчеркнем, что чем меньше уровень значимости (больше уровень доверия), тем шире соответствующий доверительный интервал (при прочих равных условиях).

Для расчета доверительных интервалов для параметров a и b уравнения линейной регрессии определяем предельную ошибку Δ для каждого показателя:

$$\Delta a = t_{\text{табл.}} \cdot se_a; \quad \Delta b = t_{\text{табл.}} \cdot se_b. \quad (8)$$

Величина $t_{\text{табл}}$ представляет собой табличное значение t -критерия Стьюдента под влиянием случайных факторов при степени свободы $k = n - 2$ и заданном уровне значимости α .

Формулы для расчета доверительных интервалов имеют следующий вид:

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

$$\gamma_b = b \pm \Delta b; \quad \gamma_{b_{\min}} = b - \Delta b; \quad \gamma_{b_{\max}} = b + \Delta b. \quad (9)$$

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

Задача

Рассчитаем доверительный интервал для коэффициентов a и b , значения которых найдены в Лабораторной работе №1.

Решение.

1. Определите предельную ошибку для каждого показателя по формуле 8.
2. Определите доверительные интервалы для коэффициентов a , b по формуле 9 и запишите полученные границы доверительных интервалов.

Лабораторная работа №6.

Точечный и интервальный прогноз по уравнению линейной регрессии

Точечный прогноз заключается в получении прогнозного значения y_p , которое определяется путем подстановки в уравнение регрессии: $\hat{y}_x = a + bx$

соответствующего (прогнозного) значения x_p : $y_p = a + b \cdot x_p$.

Интервальный прогноз заключается в построении доверительного интервала прогноза, т. е. нижней и верхней границ $y_{p\min}$, $y_{p\max}$ интервала, содержащего точную величину для прогнозного значения \hat{y}_p ($y_{p\min} < \hat{y}_p < y_{p\max}$). Доверительный интервал всегда определяется с заданной вероятностью (степенью уверенности), соответствующей принятому значению уровня значимости α .

Предварительно вычисляется стандартная ошибка прогноза $m_{\hat{y}_p}$:

$$m_{\hat{y}_p} = \sigma_{\text{ост.}} \cdot \sqrt{1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{\sum (x - \bar{x})^2}}, \quad (10)$$

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

Затем строится доверительный интервал прогноза, т. е. определяются нижняя $\gamma_{\hat{y}_{p\min}}$ и верхняя $\gamma_{\hat{y}_{p\max}}$ границы интервала прогноза:

$$\gamma_{\hat{y}_{p\min}} = \hat{y}_p - \Delta_{\hat{y}_p}, \quad \gamma_{\hat{y}_{p\max}} = \hat{y}_p + \Delta_{\hat{y}_p}.$$

где $\Delta_{\hat{y}_p} = t_{табл.} \cdot m_{\hat{y}_p}$.

Задача

Выполним прогноз ожидаемых объемов продаж y при прогнозном значении затрат на рекламу x , составляющем 107% от среднего уровня. Оценим точность прогноза, рассчитав ошибку прогноза и его доверительный интервал.

Решение.

1. Рассчитайте прогнозное значение x_p , а затем прогнозное значение y_p .
2. Рассчитайте ошибку прогноза по формуле 10.
3. Постройте доверительный интервал прогноза, сделайте выводы о надежности прогноза.

Лабораторная работа №7.

Расчет коэффициента эластичности

В экономических исследованиях широкое применение находит такой показатель, как коэффициент эластичности, вычисляемый по формуле:

$$\mathcal{E} = f'(x) \frac{x}{y}.$$

Коэффициент эластичности показывает на сколько процентов изменится результат y при изменении фактора x на 1 % от своего номинального значения.

Для линейной регрессии коэффициент эластичности равен:

$$\mathcal{E} = b \frac{x}{y},$$

и зависит от x , поэтому рассчитывают средний коэффициент эластичности:

$$\bar{\varepsilon} = f'(\bar{x}) \frac{\bar{x}}{\bar{y}} = b \frac{\bar{x}}{\bar{y}}.$$

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

Задача

Рассчитаем коэффициент эластичности и средний коэффициент эластичности.

Решение.

1. Рассчитайте коэффициент эластичности и средний коэффициент эластичности.
2. Сделайте выводы об изменении y при изменении фактора x .

Контрольные вопросы:

1. Что понимается под парной регрессией?
2. Какой вид имеет система нормальных уравнений метода наименьших квадратов в случае линейной регрессии?
3. По какой формуле вычисляется линейный коэффициент парной корреляции?
4. Как вычисляется и что показывает индекс детерминации?
5. Как проверяется значимость уравнения регрессии и отдельных коэффициентов?
6. Как строится доверительный интервал прогноза в случае линейной регрессии?
7. Как вычисляются и что показывают коэффициент эластичности ε и средний коэффициент эластичности $\bar{\varepsilon}$?

Раздел 2. Построение модели множественной линейной регрессии

Лабораторная работа №1.

Отбор факторов для построения множественной линейной зависимости

и оценка степени коллинеарности и мультиколлинеарности регрессоров

Множественная линейная регрессия – уравнение связи с несколькими независимыми переменными:

$$y = b_0 + b_1x_1 + b_2x_2 + \dots + b_mx_m + \varepsilon, \quad (1)$$

где y – зависимая (объясняемая, эндогенная) – регрессанд; x_1, x_2, \dots, x_m – независимые (объясняющие, экзогенные) переменные – регрессоры, ε – случайная составляющая модели.

Зависимость (1) во многих случаях довольно хорошо отражает сложившиеся экономические взаимосвязи. Исходная информация для построения данной зависимости обычно задается в виде некоторой таблицы.

№	Факторы, для которых получены данные					
	x_1	x_2	...	x_j	x_m	y
1	x_{11}	x_{21}	...	x_{j1}	x_{m1}	y_1
2	x_{12}	x_{22}	...	x_{j2}	x_{m2}	y_2
...
n	x_{1n}	x_{2n}	...	x_{jn}	x_{mn}	y_n

Следует определить все ли переменные следует включать в уравнение или есть переменные, которые существенно не влияют на величину y и их нецелесообразно включать в (1). В первом случае $p = m$, во втором $p < m$.

Для решения этого вопроса часто используется таблица, составленная из коэффициентов парной корреляции. Элементами такой таблицы являются коэффициенты парной корреляции для всех m факторов. Таблица имеет вид:

	y	x_1	x_2	...	x_m
y	1	r_{yx_1}	r_{yx_2}	...	r_{yx_m}
x_1	r_{x_1y}	1	$r_{x_1x_2}$...	$r_{x_1x_m}$
x_2	r_{x_2y}	$r_{x_2x_1}$	1	...	$r_{x_2x_m}$
...
x_m	r_{x_my}	$r_{x_mx_1}$	$r_{x_mx_2}$...	1

В клетках таблицы записаны парные коэффициенты корреляции, например, $r_{x_2x_1}$ – парный коэффициент корреляции между переменными x_2 и x_1 .

Коэффициенты r_{ij} и r_{ji} , а также r_{x_jy} и r_{yx_j} совпадают, так как теснота связи между переменными y и x_j такая же, как между переменными x_j и y , аналогично для x_j и x_i .

Поэтому таблицу записывают в упрощенной симметричной форме (треугольная форма).

	y	x_1	x_2	...	x_m
y	1	r_{yx_1}	r_{yx_2}	...	r_{yx_m}
x_1	—	1	$r_{x_1x_2}$...	$r_{x_1x_m}$
x_2	—	—	1	...	$r_{x_2x_m}$
...
x_m	—	—	—	...	1

По данным такой таблицы можно примерно оценить, какие факторы существенно влияют на переменную y , а какие – несущественно, а также выявить зависимость между факторами.

Если величина коэффициента парной корреляции между y и x_j мала, то нецелесообразно включать фактор x_j в уравнение (1). Высокое значение коэффициента парной корреляции между x_i и x_j , показывает их тесную корреляционную связь. Считается, что две переменные x_i и x_j явно коллинеарны, т.е. находятся между собой в линейной зависимости, если $r_{x_i x_j} \geq 0,7$. Если факторы явно коллинеарны, то они дублируют друг друга и один из них рекомендуется исключить из регрессии. Предпочтение при этом отдается не фактору, более тесно связанному с результатом, а тому фактору, который при достаточно тесной связи с результатом имеет наименьшую тесноту связи с другими факторами или выбирают один из них в зависимости от их смысла и мнения исследователя. Нецелесообразно включать в уравнение одновременно показатели, представляющие сумму некоторых факторов или их составных частей, а также характеризующие один и тот же фактор, выраженный в различных единицах измерения, например, абсолютных и относительных.

По величине парных коэффициентов корреляции обнаруживается лишь явная корреляция факторов. Наибольшие трудности возникают при *мультиколлинеарности* факторов. Для оценки мультиколлинеарности факторов используется *определитель матрицы парных коэффициентов межфакторной корреляции* $DET|R|$.

$$DET|R| = \begin{vmatrix} 1 & r_{x_1 x_2} & \dots & r_{x_1 x_m} \\ r_{x_2 x_1} & 1 & \dots & r_{x_2 x_m} \\ \dots & \dots & \dots & \dots \\ r_{x_m x_1} & r_{x_m x_2} & \dots & 1 \end{vmatrix}.$$

Чем ближе к 0 определитель матрицы межфакторной корреляции, тем сильнее мультиколлинеарность факторов и ненадежнее результаты множественной регрессии. И наоборот, чем ближе к 1 определитель матрицы межфакторной корреляции, тем меньше мультиколлинеарность факторов.

Задача

Требуется построить статистическую зависимость стоимости квартиры от трех факторов в виде множественной регрессии (табл. 1). Для построения модели, в первую очередь, необходимо провести отбор факторов и оценить степень коллинеарности регрессоров.

Решение.

1. Рассчитайте коэффициенты парной корреляции. Проверьте правильность расчетов с помощью статистической функции КОРРЕЛ электронной таблицы Microsoft Excel.
2. По полученным значениям парных коэффициентов корреляции отберите наиболее существенные факторы для уравнения (1).
3. Проверьте факторы на коллинеарность и мультиколлинеарность.
4. Сделайте выводы.

Таблица 1.

Исходные данные задачи

№ изм.	стоимость квартиры, тыс. дол., y	общая площадь, m^2 , x_1	жилая площадь, m^2 , x_2	расстояние до метро, минут пешком' x_3
1	16	80	84	3
2	22	62	37	8
3	23	69,7	42	18
4	19,8	79	80,3	28
5	34	96,4	88	8
6	24,8	90	64	8
7	23,7	102	66	7
8	41	87	86,8	10
9	31	114,8	74	10
10	38,6	114,3	74,7	8
11	46	90	62	8
12	38	116	81	10
13	42,7	107	78,8	10
14	27	93	66	18
15	78	176	129	10
16	38	96	69,4	8
17	23,8	92	72,8	10
18	68	176	110	20
19	23	74	49	18
20	48,8	106	73,7	10
21	34	88	61,7	3
22	23	74	48,8	10
23	26,8	74,7	80,8	10
24	37	118	76	8
25	30	92	62	18
26	43	110	79,8	8

Лабораторная работа №2.*Оценка параметров уравнения множественной регрессии*

Для оценки параметров уравнения множественной линейной регрессии применяют метод наименьших квадратов (МНК).

Для линейных уравнений и нелинейных уравнений, приводимых к линейным, строится следующая система нормальных уравнений, решение которой позволяет получить оценки параметров регрессии, в частности для уравнения (1):

$$\begin{cases} \sum y = nb_0 + b_1 \sum x_1 + b_2 \sum x_2 + \dots + b_m \sum x_p, \\ \sum yx_1 = b_0 \sum x_1 + b_1 \sum x_1^2 + b_2 \sum x_1x_2 + \dots + b_m \sum x_mx_1, \\ \sum yx_m = b_0 \sum x_m + b_1 \sum x_1x_p + b_2 \sum x_2x_m + \dots + b_m \sum x_m^2. \end{cases} \quad (2)$$

Одним из методов для ее решения является метод определителей:

$$b_0 = \frac{\Delta b_0}{\Delta}; \quad b = \frac{\Delta b_1}{\Delta}; \quad b_m = \frac{\Delta b_m}{\Delta};$$

$$\Delta \begin{vmatrix} n & \sum x_1 & \sum x_2 & \dots & \sum x_m \\ \sum x_1 & \sum x_1^2 & \sum x_2x_1 & \dots & \sum x_mx_1 \\ \dots & \dots & \dots & \dots & \dots \\ \sum x_m & \sum x_1x_m & \sum x_2x_m & \dots & \sum x_m^2 \end{vmatrix} \text{ — определитель системы;}$$

где

$\Delta b_0, \Delta b_1, \dots, \Delta b_m$ – частные определители; которые получаются путем замены соответствующего столбца матрицы определителя системы данными левой части системы.

Задача

1. Рассчитайте параметры уравнения множественной регрессии с помощью метода определителей. Для вычисления определителей используйте статистическую функцию МОПРЕД. Проверьте полученные результаты с помощью функции ЛИНЕЙН.
2. Запишите полученное уравнение множественной регрессии.

Лабораторная работа №3.

Построение уравнения множественной регрессии в стандартизованной форме

Если коэффициенты множественной регрессии рассматривать в качестве показателей влияния факторов, то следует иметь в виду, что коэффициенты регрессии между собой прямо не сравнимы. Их численные значения зависят от выбранных единиц измерения каждого фактора. Чтобы коэффициенты регрессии стали сравнимы, их приводят к стандартизованному масштабу. Для этого все переменные выражают в безразмерных, так называемых стандартизованных единицах измерения при помощи следующих соотношений:

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

где σ_y , σ_{x_j} – соответственно средние квадратичные отклонения.

Свободный член b_0 в стандартизованном уравнении множественной регрессии отсутствует, т.е. уравнение (1) можно записать в виде:

$$t_y = \beta_1 t_{x_1} + \beta_2 t_{x_2} + \dots + \beta_m t_{x_m} + \varepsilon. \quad (2)$$

Все переменные уравнения выражены в сравнимых единицах измерения. Коэффициенты $\beta_1, \beta_2, \dots, \beta_m$ называются коэффициентами регрессии в стандартизованном масштабе.

Для их определения можно перейти от коэффициентов b_j к β_j ($j = 1, m$) и обратно осуществляется по формуле:

$$\beta_j = \frac{b_j \sigma_{x_j}}{\sigma_y}.$$

Коэффициенты регрессии (2) показывают влияния изменения каждой переменной на изменение фактора y . Все коэффициенты выражены в сравнимых единицах измерения. Чем больше $|\beta_j|$, тем сильнее влияет соответствующий факторный показатель на результативный.

Таким образом, сравнивая между собой стандартизованные коэффициенты регрессии можно ранжировать факторы по силе их воздействия на результат. В этом основное достоинство стандартизованных коэффициентов регрессии в отличие от коэффициентов «чистой» регрессии, которые несравнимы между собой.

На практике часто используют уравнение регрессии в стандартизованном масштабе в виде:

$$y = \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_m x_m + \varepsilon.$$

Как было показано выше, ранжирование факторов, участвующих во множественной линейной регрессии, может быть проведено через стандартизованные коэффициенты регрессии (β -коэффициенты). Эта же цель может быть достигнута с помощью частных коэффициентов корреляции – для линейных связей. Кроме того, частные показатели корреляции широко используются при решении проблемы отбора факторов: целесообразность включения того или иного фактора в модель доказывается величиной показателя частной корреляции.

Частные коэффициенты (или индексы) корреляции характеризуют тесноту связи между результатом и соответствующим фактором при устранении влияния других факторов, включенных в уравнение регрессии. Показатели частной корреляции представляют собой отношение сокращения остаточной дисперсии, произошедшего в результате

дополнительного включения в анализ нового фактора, к остаточной дисперсии, имевшей место до введения в модель нового фактора.

Коэффициент частной корреляции, измеряющий влияние на y фактора x_i при неизменном уровне других факторов, можно определить по формуле:

$$r_{yx_j * x_1 x_2 \dots x_{j-1} x_{j+1} \dots x_m} = \sqrt{1 - \frac{1 - R_{yx_1 \dots x_j \dots x_m}^2}{1 - R_{yx_1 x_2 \dots x_{j-1} x_{j+1} \dots x_m}^2}},$$

где $R_{yx_1 \dots x_j \dots x_m}^2$ – множественный коэффициент детерминации всего комплекса m факторов с результатом; $R_{yx_1 x_2 \dots x_{j-1} x_{j+1} \dots x_m}^2$ – показатель детерминации, но без введения в модель фактора x_j .

Средние коэффициенты эластичности для линейной множественной регрессии рассчитываются по формуле:

$$\bar{\varepsilon}_{yx_j} = b_j \frac{\bar{x}_j}{\bar{y}}$$

и показывают, на сколько процентов в среднем по совокупности изменится результат y от своей величины при изменении фактора x на 1 % от своего значения при неизменных значениях других факторов.

Задача

1. Рассчитайте стандартизованные коэффициенты регрессии. Запишите полученное уравнение в стандартизованном масштабе вида (2).
2. Проранжируйте стандартизованные коэффициенты регрессии по силе их воздействия на результат.
3. Рассчитайте средние коэффициенты эластичности.

Лабораторная работа №4.

Составление частных уравнений регрессии

На основе линейного уравнения множественной регрессии (1) могут быть найдены частные уравнения регрессии, т.е. уравнения регрессии, которые связывают результативный признак с соответствующими факторами x_j при закреплении других, учитываемых во множественной регрессии, факторов на среднем уровне. Частные уравнения регрессии имеют следующий вид:

$$y_{x_1, x_2, x_3, \dots, x_m} = b_0 + b_1 x_1 + b_2 \bar{x}_2 + b_3 \bar{x}_3 + \dots + b_m \bar{x}_m + \varepsilon;$$

$$y_{x_2, x_1, x_3, \dots, x_m} = b_0 + b_1 \bar{x}_1 + b_2 x_2 + b_3 \bar{x}_3 + \dots + b_m \bar{x}_m + \varepsilon;$$

.....;

$$y_{x_m, x_1, x_2, \dots, x_{m-1}} = b_0 + b_1 \bar{x}_1 + b_2 \bar{x}_2 + \dots + b_{m-1} \bar{x}_{m-1} + \dots + b_m x_m + \varepsilon;$$

При подстановке в эти уравнения средних значений соответствующих факторов они принимают вид парных уравнений линейной регрессии, т. е. имеем:

$$\hat{y}_{x_1, x_2, x_3, \dots, x_m} = B_1 + b_1 x_1;$$

$$\hat{y}_{x_2, x_1, x_3, \dots, x_m} = B_2 + b_2 x_2;$$

.....;

$$\hat{y}_{x_m, x_1, x_2, \dots, x_{m-1}} = B_m + b_m x_m.$$

где

$$B_1 = b_0 + b_2 \bar{x}_2 + b_3 \bar{x}_3 + \dots + b_m \bar{x}_m;$$

$$B_2 = b_0 + b_1 \bar{x}_1 + b_3 \bar{x}_3 + \dots + b_m \bar{x}_m;$$

.....;

$$B_m = b_0 + b_1 \bar{x}_1 + b_2 \bar{x}_2 + \dots + b_{m-1} \bar{x}_{m-1};$$

В отличие от парной регрессии, частные уравнения регрессии характеризуют изолированное влияние фактора на результат, ибо другие факторы закреплены на неизменном уровне. Эффекты влияния других факторов присоединены в них к свободному члену уравнения множественной регрессии. Это позволяет на основе частных уравнений регрессии определять **частные коэффициенты эластичности**:

$$\mathcal{E}_{yx_i} = b_j \frac{x_i}{\hat{y}_{x_j, x_1, x_2, \dots, x_{j-1}, x_{j+1}, \dots, x_m}},$$

где b_j – коэффициенты регрессии для фактора x_i в уравнении множественной регрессии;

$\hat{y}_{x_j, x_1, x_2, \dots, x_{j-1}, x_{j+1}, \dots, x_m}$ – значение результативного фактора, полученное из частного уравнения регрессии при данном значении фактора x_j .

Задача

1. Составьте частные уравнения регрессии.
2. Определите частные коэффициенты эластичности и проранжируйте факторы по силе их воздействия на результат.

Лабораторная работа №5.

Оценка надежности результатов множественной регрессии и корреляции

Практическая значимость уравнения множественной регрессии оценивается с помощью показателя множественной корреляции и его квадрата – коэффициента детерминации.

Показатель множественной корреляции характеризует тесноту связи рассматриваемого набора факторов с исследуемым признаком, или, иначе, оценивает тесноту совместного влияния факторов на результат.

Независимо от формы связи показатель множественной корреляции может быть найден как индекс множественной корреляции:

$$R_{yx_1, x_2, \dots, x_m} = \sqrt{1 - \frac{\sigma_{ост}^2}{\sigma_y^2}},$$

где σ_y^2 – общая дисперсия результативного признака; $\sigma_{ост}^2$ – остаточная дисперсия для

уравнения $\hat{y} = f(x_1, x_2, \dots, x_m)$, $\sigma_{ост}^2 = \frac{\sum (y - \hat{y}_{x_1, x_2, \dots, x_m})^2}{n}$.

При линейной зависимости признаков формула индекса корреляции может быть также представлена следующим выражением:

$$R_{yx_1, x_2, \dots, x_m} = \sqrt{\sum_{j=1}^m \beta_{x_j} r_{yx_j}},$$

где β_{x_j} – стандартизованные коэффициенты регрессии; r_{yx_j} – парные коэффициенты корреляции результата с каждым фактором.

Формула индекса множественной корреляции для линейной регрессии получила название **линейного коэффициента множественной корреляции**, или, что то же самое, **совокупного коэффициента корреляции**.

Индекс множественной корреляции изменяется от 0 до 1. Чем ближе его значение к 1, тем теснее связь результативного признака со всем набором исследуемых факторов. Величина индекса множественной корреляции больше или равна максимального парного индекса корреляции:

$$R_{yx_1x_2\dots x_m} \geq R_{yx_j(\max)}, \quad j = \overline{1, m}.$$

При правильном включении факторов в регрессионный анализ величина индекса множественной корреляции будет существенно отличаться от индекса корреляции парной зависимости.

Коэффициент (индекс) множественной детерминации рассчитывается как квадрат индекса множественной корреляции:

$$R^2_{yx_1x_2\dots x_m}.$$

Скорректированный коэффициент (индекс) множественной детерминации содержит поправку на число степеней свободы и рассчитывается по формуле:

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

где n – число наблюдений; m – число факторов.

Значимость уравнения множественной регрессии в целом оценивается с помощью F -критерия Фишера:

$$F_{\text{факт.}} = \frac{R^2}{1 - R^2} \cdot \frac{n - m - 1}{m}.$$

$F_{\text{табл}}$ определяется при заданном α (0,05; 0,01) и степенях свободы $k_1 = m, k_2 = n - m - 1$, (m – число параметров при факторных переменных в уравнении множественной регрессии).

Частный F -критерий оценивает статистическую значимость присутствия каждого из факторов в уравнении. В общем виде, для факторов частный F -критерий определится как:

$$F_{\text{част } x_j} = \frac{R^2_{yx_1, \dots, x_j, x_m} - R^2_{yx_j, \dots, x_{j-1}, x_{j+1}, \dots, x_m}}{1 - R^2_{yx_1, \dots, x_j, x_m}} \cdot \frac{n - m - 1}{1}.$$

$F_{\text{табл}}$ для частного F -критерия определяется при заданном α и степенях свободы $k_1 = 1, k_2 = n - m - 1$.

Оценка значимости коэффициентов чистой регрессии с помощью t -критерия Стьюдента сводится к вычислению значения:

$$t_{\text{факт.}} = \frac{b_j}{m_{b_j}} = \sqrt{F_{x_j}},$$

где m_{b_i} – средняя квадратичная ошибка коэффициента регрессии b_j , она может быть определена по следующей формуле:

$$m_{b_i} = \frac{\sigma_y \cdot \sqrt{1 - R_{yx_1x_2\dots x_m}^2}}{\sigma_{x_j} \cdot \sqrt{1 - R_{x_jx_1\dots x_m}^2}} \cdot \frac{1}{\sqrt{n - m - 1}}$$

$R_{yx_1x_2\dots x_m}^2$ – множественный коэффициент детерминации всего комплекса m факторов с результатом;
 $R_{x_jx_1\dots x_m}^2$ – коэффициент детерминации для зависимости фактора x_j со всеми другими факторами уравнения множественной регрессии.

$t_{табл}$ определяется при заданном α и степени свободы $k = n - m - 1$.

Задача

1. Рассчитайте линейного коэффициента множественной корреляции и сделайте вывод о тесноте связи результативного признака со всем набором исследуемых факторов.
2. Рассчитайте коэффициент множественной детерминации и сделайте вывод об оценке качества построенной модели.
3. Рассчитайте $F_{факт}$ и сравните полученное значение со значением, рассчитанным с помощью функции ЛИНЕЙН. Определите $F_{табл}$ при заданном уровне значимости $\alpha = 0,05$. Сравните полученные значения $F_{табл}$ и $F_{факт}$, сделайте вывод о существенности связи.
4. Определите $t_{факт}$ и $t_{табл}$ при заданном уровне значимости $\alpha = 0,05$ Сравните полученные значения и сделайте вывод о статистической значимости параметров регрессии.

Контрольные вопросы:

1. Что понимается под линейной множественной регрессией?
2. Какие задачи решаются при построении уравнения регрессии?
3. Какие требования предъявляются к факторам, включаемым в уравнение регрессии?
4. Что понимается под коллинеарностью и мультиколлинеарностью факторов?
5. Как проверяется наличие коллинеарности и мультиколлинеарности?
6. Какой вид имеет система нормальных уравнений метода наименьших квадратов в случае линейной регрессии?

7. По какой формуле вычисляется индекс множественной корреляции?
8. Как вычисляются индекс множественной детерминации и скорректированный индекс множественной детерминации?
9. Что означает низкое значение коэффициента (индекса) множественной корреляции?
10. Как проверяется значимость уравнения регрессии и отдельных коэффициентов?
11. Как строятся частные уравнения регрессии?
12. Как вычисляются средние частные коэффициенты эластичности?
13. Что такое стандартизированные переменные?
14. Какой вид имеет уравнение линейной регрессии в стандартизированном масштабе?
15. Что понимается под гомоскедастичностью?
16. Как проверяется гипотеза о гомоскедастичности ряда остатков?

Раздел 3. Построение регрессионных моделей с фиктивными переменными

В регрессионных моделях в качестве объясняющих переменных часто приходится использовать не только количественные (определяемые численно), но и качественные переменные. Например, спрос на какое-либо благо может определяться как количественными переменными (цена данного блага), так и качественными (вкусы потребителей). Качественные показатели в численном виде представить нельзя. Возникает проблема отражения в модели влияния таких переменных на исследуемую величину.

Обычно в моделях влияние качественного фактора выражается в виде фиктивной (искусственной) переменной, которая отражает два противоположных состояния качественного фактора. В этом случае фиктивная переменная может выражаться в двоичной форме:

$$D = \begin{cases} 0, & \text{фактор не действует;} \\ 1, & \text{фактор действует.} \end{cases}$$

Переменная D называется *фиктивной (искусственной, двоичной) переменной (индикатором)*.

Регрессионные модели, содержащие лишь качественные объясняющие переменные, называются *моделями дисперсионного анализа (ANOVA-моделями)*.

Например, пусть Y – начальная заработная плата.

$$D = \begin{cases} 0, & \text{если претендент не имеет высшего образования;} \\ 1, & \text{если претендент имеет высшее образование.} \end{cases}$$

Тогда зависимость можно выразить моделью парной регрессии:

$$Y = \beta_0 + \gamma D + \varepsilon$$

Очевидно,

$$M(Y | D = 0) = \beta_0 + \gamma \cdot 0 = \beta_0;$$

$$M(Y | D = 1) = \beta_0 + \gamma \cdot 1 = \beta_0 + \gamma.$$

Коэффициент β_0 определяет среднюю начальную заработную плату при отсутствии высшего образования. Коэффициент γ указывает, на какую величину отличаются средние начальные заработные платы при наличии и при отсутствии высшего образования у претендента. Проверая статическую значимость коэффициента γ с помощью t -статистики, либо значимость коэффициента детерминации R^2 или F -статистики, можно определить, влияет или нет наличие высшего образования на начальную заработную плату.

Модели, в которых объясняющие переменные носят как количественный, так и качественный характер, называются *моделями ковариационного анализа* (ANCOVA-моделями).

Существует несколько разновидностей моделей ковариационного анализа.

1. Модели ковариационного анализа при наличии у фиктивной переменной двух альтернатив.

Рассмотрим простейшую модель с одной количественной и одной качественной переменными, имеющую два альтернативных состояния:

$$Y = \beta_0 + \beta_1 x + \gamma D + \varepsilon$$

Пусть, например, Y – заработная плата сотрудника фирмы, x – стаж сотрудника, D – пол сотрудника, т.е.

$$D = \begin{cases} 0, & \text{если сотрудник-женщина;} \\ 1, & \text{если сотрудник-мужчина.} \end{cases}$$

Тогда ожидаемое значение заработной платы сотрудников при x годах трудового стажа будет:

$$M(Y | x, D = 0) = \beta_0 + \beta_1 x;$$

$$M(Y | x, D = 1) = \beta_0 + \beta_1 x + \gamma = (\beta_0 + \gamma) + \beta_1 x.$$

Заработная плата в данном случае является линейной функцией от стажа работы.

При составлении моделей с фиктивными переменными необходимо руководствоваться следующим *правилом моделирования*: если качественная переменная имеет k альтернативных значений, то при моделировании используется $(k - 1)$ фиктивных переменных. Таким образом, если переменная имеет два альтернативных значения (например, пол), то в модель можно ввести только одну фиктивную переменную.

Если не следовать данному правилу, то при моделировании исследователь попадает в ситуацию совершенной мультиколлинеарности или так называемую *ловушку фиктивной переменной*.

Значение качественной переменной, для которого принимается $D = 0$, называется *базовым* или *сравнительным*. Выбор базового значения обычно диктуется целями исследования, но может быть и произвольным.

Коэффициент γ в модели иногда называется *дифференциальным коэффициентом свободного члена*, так как он показывает, на какую величину отличается свободный член модели при значении фиктивной переменной, равном единице, от свободного члена модели при базовом значении фиктивной переменной.

2. Модели ковариационного анализа при наличии у качественных переменных более двух альтернатив.

Рассмотрим модель с двумя объясняющими переменными, одна из которых количественная, а другая – качественная. Причем качественная переменная имеет три альтернативы. Например, расходы на содержание ребенка могут быть связаны с доходами домохозяйства и возрастом ребенка: дошкольный, младший школьный и старший школьный. Так как качественная переменная имеет три альтернативы, то по общему правилу моделирования необходимо использовать две фиктивные переменные. Таким образом, модель может быть представлена в виде:

$$Y = \beta_0 + \beta_1 x + \gamma_1 D_1 + \gamma_2 D_2 + \varepsilon,$$

где Y – расходы, x – доходы домохозяйств.

$$D_1 = \begin{cases} 0, & \text{если дошкольник;} \\ 1, & \text{в противоположном случае.} \end{cases}$$

$$D_2 = \begin{cases} 0, & \text{если дошкольник или младший школьник;} \\ 1, & \text{в противоположном случае.} \end{cases}$$

Образуются следующие зависимости:

1. Средний расход на дошкольника:

$$M(Y | D_1 = 0, D_2 = 0) = \beta_0 + \beta_1 x. \quad (1)$$

2. Средний расход на младшего школьника:

$$M(Y | D_1 = 1, D_2 = 0) = (\beta_0 + \gamma_1) + \beta_1 x. \quad (2)$$

3. Средний расход на старшего школьника:

$$M(Y | D_1 = 1, D_2 = 1) = (\beta_0 + \gamma_1 + \gamma_2) + \beta_1 x. \quad (3)$$

Здесь γ_1, γ_2 – дифференциальные свободные члены. Базовым значением качественной переменной является значение «дошкольник». После вычисления коэффициентов уравнений регрессии (1) – (3) определяется статистическая значимость коэффициентов γ_1 и γ_2 на основе обычной t -статистики.

Если коэффициенты γ_1 и γ_2 оказываются статистически незначимыми, то можно сделать вывод, что возраст ребенка не оказывает влияния на расходы по его содержанию.

3. Регрессия с одной количественной и двумя качественными переменными.

Техника фиктивных переменных может быть распространена на произвольное число качественных факторов. Рассмотрим ситуацию с двумя качественными переменными.

Пусть Y – заработная плата сотрудников фирмы, x – стаж работы, D_1 – наличие высшего образования, D_2 – пол сотрудника:

$$D_1 = \begin{cases} 0, & \text{если сотрудник-женщина;} \\ 1, & \text{если сотрудник-мужчина.} \end{cases}$$

$$D_2 = \begin{cases} 0, & \text{если нет высшего образования;} \\ 1, & \text{в противоположном случае.} \end{cases}$$

Таким образом, получим следующую модель:

$$Y = \beta_0 + \beta_1 x + \gamma_1 D_1 + \gamma_2 D_2 + \varepsilon.$$

Из этой модели выводятся следующие регрессионные модели:

1. Средняя зарплата женщины без высшего образования:

$$M(Y | D_1 = 0, D_2 = 0) = \beta_0 + \beta_1 x.$$

2. Средняя зарплата женщины с высшим образованием:

$$M(Y | D_1 = 0, D_2 = 1) = (\beta_0 + \gamma_2) + \beta_1 x.$$

3. Средняя зарплата у мужчины без высшего образования:

$$M(Y | D_1 = 1, D_2 = 0) = (\beta_0 + \gamma_1) + \beta_1 x.$$

4. Средняя зарплата мужчины с высшим образованием:

$$M(Y | D_1 = 1, D_2 = 1) = (\beta_0 + \gamma_1 + \gamma_2) + \beta_1 x.$$

Очевидно, что все регрессии отличаются только свободными членами. Дальнейшее определение статистической значимости коэффициентов γ_1 и γ_2 позволяет убедиться, влияют ли образование и пол сотрудника на его заработную плату.

Задача 1

Исследуется зависимость между заработной платой рабочего за месяц y (\$), x - возрастом рабочего (лет) и фиктивной переменной D – пол рабочего.

№ п/п	Заработная плата за месяц y (\$)	Возраст рабочего x (лет)	Пол, м/ж
1	300	29	Ж
2	400	40	М
3	300	36	Ж
4	320	32	Ж
5	200	23	М
6	350	45	М
7	350	38	Ж
8	400	40	М
9	380	50	М
10	400	47	М
11	400	47	Ж
12	250	28	М
13	350	30	М
14	200	25	М
15	400	48	Ж
16	220	30	М
17	320	40	М
18	390	40	М
19	360	38	Ж
20	260	29	М

1. Необходимо построить модель $Y = b_0 + b_1x + \gamma D$ с фиктивной переменной D , которая принимает два значения: 1 - если пол рабочего мужской; 0 - если пол женский.
2. Проверить статистическую значимость коэффициентов. Сделать выводы.

Задача 2

На предприятии используются станки трех фирм (А, В, С). Исследуется надежность станков. При этом учитывается возраст станка (x , мес.) и время безаварийной работы до последней поломки (y , час). Выборка из 40 станков дала результаты, представленные в таблице.

№ измерения	Возраст станка в месяцах	Время работы станка без ремонта в часах	Фирма - производитель
1	23	280	А

2	30	230	B
3	65	112	C
4	69	176	A
5	75	90	C
6	63	176	A
7	25	216	B
8	75	110	C
9	75	45	B
10	52	200	A
11	20	265	B
12	70	148	C
13	62	150	C
14	40	176	B
15	66	123	A
16	20	245	A
17	39	176	C
18	25	260	B
19	48	236	A
20	59	205	A
21	25	240	A
22	69	65	B
23	71	115	A
24	26	200	C
25	45	126	B
26	40	225	A
27	30	210	C
28	69	45	B

29	30	260	A
30	22	220	B
31	33	194	B
32	48	156	C
33	75	100	A
34	21	240	B
35	56	170	A
36	58	116	C
37	50	120	B
38	37	140	A
39	56	88	B
40	67	120	A

1. Оцените уравнение регрессии без учета различия станков разных фирм.
2. Оценить уравнение регрессии, учитывающее различие качества разных фирм (F-критерий Фишера, t-критерий Стьюдента, скорректированный коэффициент детерминации).
3. Сравните качество построенных моделей.
4. Построить корреляционное поле и нанести на него графики функций.
5. Сделать выводы о необходимости использования фиктивных переменных в этом случае.

ПОСТРОЕНИЕ МОДЕЛИ НЕЛИНЕЙНОЙ ПАРНОЙ РЕГРЕССИИ

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

Примеры регрессий, нелинейных по объясняющим переменным, но линейных по оцениваемым параметрам:

– полиномы разных степеней
$$y_x = a + b_1x + b_2x^2 + b_3x^3 + \varepsilon;$$

$$y_x = a + \frac{b}{x} + \varepsilon$$

– равносторонняя гипербола

Примеры регрессий, нелинейных по оцениваемым параметрам:

– степенная $y_x = a \cdot x^b \cdot \varepsilon$;

– показательная $y_x = a \cdot b^x \cdot \varepsilon$;

– экспоненциальная $y_x = e^{a+bx} \cdot \varepsilon$.

Наиболее часто применяются следующие нелинейные модели регрессий:

гиперболы – $y_x = a + \frac{b}{x}$;

параболы – $y_x = a + bx + cx^2$;

показательной функции $y_x = a \cdot b^x$;

степенной функции – $y_x = a \cdot x^b$ и др.

Лабораторная работа №1.

Вычисление коэффициентов уравнения нелинейной регрессии

Построению нелинейных регрессионных моделей предшествует процедура линеаризации переменных.

Линеаризация – (от лат. *linearis* — линейный), один из методов приближённого представления замкнутых нелинейных систем, при котором исследование нелинейной системы заменяется анализом линейной системы, в некотором смысле эквивалентной исходной. Методы линеаризации имеют ограниченный характер, т. е. эквивалентность исходной нелинейной системы и её линейного приближения сохраняется лишь для ограниченных пространственных или временных масштабов системы, либо для определенных процессов, причем, если система переходит с одного режима работы на другой, то следует изменить и её линеаризованную модель. Применяя линеаризацию, можно выяснить многие качественные и особенно количественные свойства нелинейной системы.

Для нелинейных уравнений регрессии, приводимых к линейным с помощью преобразования $(x, y) \rightarrow (x', y')$, система нормальных уравнений имеет вид:

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

в преобразованных переменных x', y' .

Коэффициент b при факторной переменной x имеет следующую интерпретацию: он показывает, на сколько изменится в среднем величина y при изменении фактора x на 1 единицу измерения.

Гиперболическая регрессия:
$$y_x = a + \frac{b}{x}.$$

Линеаризующее преобразование:
$$x' = \frac{1}{x}; y' = y.$$

Тогда

$$\begin{cases} na + b \sum \frac{1}{x} = \sum y, \\ a \sum \frac{1}{x} + b \sum \frac{1}{x^2} = \sum y \frac{1}{x}. \end{cases}$$

$$a = \frac{1}{n} \sum y - \frac{1}{n} b \sum \frac{1}{x}, \quad b = \frac{n \sum \frac{y}{x} - \sum \frac{1}{x} \sum y}{n \sum \frac{1}{x^2} - (\sum \frac{1}{x})^2}.$$

Экспоненциальная регрессия:
$$y_x = e^{a+bx}.$$

Линеаризующее преобразование:
$$x' = x; y' = \ln y.$$

$$a = \frac{1}{n} \sum \ln y - \frac{1}{n} b \sum x, \quad b = \frac{n \sum x \ln y - \sum x \sum \ln y}{n \sum x^2 - (\sum x)^2}$$

Степенная функция:
$$y_x = a \cdot x^b.$$

Линеаризующее преобразование:
$$x' = \ln x; y' = \ln y.$$

$$\ln a = \frac{1}{n} \sum \ln y - \frac{1}{n} b \sum \ln x, \quad b = \frac{n \sum (\ln x \cdot \ln y) - \sum \ln x \sum \ln y}{n \sum (\ln x)^2 - (\sum \ln x)^2}$$

Переход к параметру a уравнения степенной зависимости производят с помощью потенцирования по формуле:

$$a = e^{\ln a}$$

Показательная функция: $y_x = a \cdot b^x$

Линеаризующее преобразование: $x' = x; y' = \ln y$

$$\ln a = \frac{1}{n} \sum \ln y - \frac{1}{n} \ln b \sum x, \quad \ln b = \frac{n \sum x \cdot \ln y - \sum x \sum \ln y}{n \sum x^2 - (\sum x)^2}$$

Переход к параметрам уравнения показательной зависимости (a, b) производят по формулам:

$$a = e^{\ln a}, \quad b = e^{\ln b}$$

Логарифмическая функция: $y_x = a + b \ln x$

Линеаризующее преобразование: $x' = \ln x; y' = y$

$$a = \frac{1}{n} \sum y - \frac{1}{n} b \sum \ln x, \quad b = \frac{n \sum \ln x \cdot y - \sum \ln x \sum y}{n \sum (\ln x)^2 - (\sum \ln x)^2}$$

Задача

Для определения зависимости между рождаемостью в России (переменная y измеряемая в тыс. чел.) и средней заработной платой (переменная x , измеряемая в руб.) были проведены исследования, результаты которых представлены в таблице 1.

Таблица 1.

Годы	Рождаемость в тыс. чел. y	Средняя заработная плата, руб. в мес. x
2014	886,9	2290,1
2015	928,6	3078,4
2016	998,1	3947,2
2017	1050,6	5170,4
2018	1074,3	6410,4
2019	1036,9	8023,2

Для характеристики зависимости y от x рассчитать параметры следующих функций:

- равносторонней гиперболы;
- степенной функции;
- показательной функции.

Решение.

$$y_x = a + \frac{b}{x}$$

1. Уравнение равносторонней гиперболы $y_x = a + \frac{b}{x}$ линеаризуется

$$x' = \frac{1}{x}; y' = y, \text{ тогда } y_x = a + bx'$$

2. Для расчета используем данные таблицы 2.

Годы	y	x'	$y \cdot x'$	$(x')^2$	y^2
2014	886,9				
2015	928,6				
2016	998,1				
2017	1050,6				
2018	1074,3				
2019	1036,9				
Итого					
Среднее значение					
σ					
σ^2					

Значения параметров регрессии a и b рассчитываем по формулам:

$$b = \frac{\overline{yx'} - \bar{y} \cdot \bar{x'}}{\sigma_{x'}^2}; a = \bar{y} - b \cdot \bar{x'}$$

3. Запишите полученное уравнение регрессии.

4. С помощью линеаризующих преобразований постройте уравнение регрессии для степенной и показательной функций.

Лабораторная работа №2.

Вычисление индекса корреляции и средней ошибки аппроксимации

Тесноту связи изучаемых явлений в нелинейных регрессиях оценивает **индекс корреляции**, который вычисляется по формуле:

$$\rho_{xy} = \sqrt{1 - \frac{\sigma_{\text{ост.}}^2}{\sigma_y^2}} = \sqrt{1 - \frac{\sum (y - \hat{y}_x)^2}{\sum (y - \bar{y})^2}}; \quad 0 \leq \rho_{xy} \leq 1.$$

Средняя ошибка аппроксимации - среднее отклонение расчетных значений от фактических:

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

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

Задача

1. Рассчитать индекс корреляции и среднюю ошибку аппроксимации для нелинейных регрессий Лабораторной №1.
2. Определить по какому уравнению регрессии получена наибольшая оценка тесноты связи.
3. Сделать выводы.

Лабораторная работа №3.

Оценка значимости уравнения регрессии с помощью F-критерия Фишера

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

Задача

Определим F-критерий Фишера для уравнений парной нелинейной регрессии, полученных в Лабораторной работе №1.

Решение.

1. Рассчитайте $F_{\text{факт}}$ и сравните полученное значение со значением, $F_{\text{табл}}$ при заданном уровне значимости $\alpha = 0,05$.
2. Сделайте выводы о существенности связи.

Лабораторная работа №4.

Расчет средних коэффициентов эластичности

В экономических исследованиях широкое применение находит такой показатель, как средний коэффициент эластичности, вычисляемый по формуле:

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

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

Задача

По группе предприятий, производящих однородную продукцию, известно, как зависит себестоимость единицы продукции y от факторов, приведенных в табл. 3.

Таблица 3.

Признак-фактор	Уравнение парной регрессии	Среднее значение фактора
Объем производства, млн руб., x_1	$\hat{y}_{x_1} = 0,62 + \frac{58,74}{x}$	$\bar{x}_1 = 2,64$
Оптовая цена за 1 т энергоносителя, млн. руб., x_2	$\hat{y}_{x_2} = 11,75 \cdot x_2^{1,6281}$	$\bar{x}_2 = 1,503$
Доля прибыли, изымаемой государством, %, x_3	$\hat{y}_{x_3} = 14,87 \cdot 1,016^{x_3}$	$\bar{x}_3 = 26,3$

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

Решение.

1. Для уравнения равнобочной гиперболы $\hat{y}_{x_1} = 0,62 + \frac{58,74}{x}$:

$$\bar{\varepsilon}_{yx_1} = f'(x_1) \frac{\bar{x}_1}{y} = -\frac{b}{x_1^2} \cdot \frac{\bar{x}_1}{a + b/x_1} = \frac{-b}{ax_1 + b}$$

2. Рассчитайте средние коэффициенты эластичности для степенной и показательной функций (Приложение 4).

ВРЕМЕННЫЕ РЯДЫ

Совокупность упорядоченных во времени наблюдений некоторого показателя называют *временным рядом*.

Ряд наблюдений за значениями определенного показателя, упорядоченный в зависимости от возрастающих или убывающих значений другого показателя, называют *динамическим рядом*, *временным рядом*, *рядом динамики*.

Отдельные наблюдения временного ряда называются *уровнями* этого ряда.

Лабораторная работа №1

Выявление аномальных уровней временных рядов

Уровни временных рядов могут иметь аномальные значения. Появление таких значений может быть вызвано ошибками при сборе, записи или передаче информации – это ошибки технического порядка, или ошибки первого рода. Однако аномальные значения могут отражать реальные процессы, например скачок доллара или падение курса ценных бумаг и т.д.; такие аномальные значения относят к ошибкам второго рода, они не подлежат устранению. Для выявления аномальных уровней временных рядов можно использовать *метод Ирвина*.

Пусть имеется временной ряд:

$$y_t = y_1, y_2, \dots, y_t, \dots, y_n, \quad t = \overline{1, n}. \quad (1)$$

Метод Ирвина предполагает использование следующей формулы:

$$\lambda_t = \frac{|y_t - y_{t-1}|}{\sigma_y}, \quad t = \overline{1, n}.$$

где $\sigma_y = \sqrt{y^2 - \bar{y}^2}$.

Расчетные значения $\lambda_2, \lambda_3, \dots$ сравниваются с табличными значениями критерия Ирвина λ_a ; если какое-либо из них оказывается больше табличного, то соответствующее значение y_t уровня ряда считается аномальным. Значения критерия Ирвина для уровня значимости $\alpha = 0,05$ приведены в таблице.

n	2	3	10	20	30	50	100
λ_a	2,8	2,3	1,5	1,3	1,2	1,1	1,0

После выявления аномальных уровней необходимо определить причины их возникновения. Если они вызваны ошибками технического порядка, то они устраняются либо заменой аномальных уровней соответствующими значениями по кривой, аппроксимирующей временной ряд, либо заменой уровней средней арифметической двух соседних уровней ряда.

Ошибки, возникающие из-за воздействия факторов, имеющих объективный характер, устранению не подлежат.

Задача

Временной ряд задан в табличной форме.

t	1	2	3	4	5	6	7	8	9	10
y	1,6	1,9	2,1	2,4	4,5	2,8	3,1	3,3	3,6	3,8

1. Проверить наличие аномальных значений точки $t = 2$, $t = 5$, $t = 10$.
2. Устранить аномальные уровни, считая причинами их возникновения ошибки первого уровня.

Лабораторная работа №2

Проверка гипотезы существования тенденции

Если во временном ряду проявляется длительная тенденция изменения экономического показателя, то в этом случае говорят, что имеет место *тренд*.

Под трендом понимают изменение, определяющее общее направление развития или основную тенденцию временного ряда.

Другими словами, *тренд (тенденция)* – это устойчивое, систематическое изменение значений временного ряда в течение достаточно долгого периода.

При прогнозировании временных рядов необходимо выяснить, существует ли тенденция во временном ряду или она отсутствует.

Для временного ряда (1) рассмотрим критерий «восходящих и нисходящих» серий, согласно которому тенденция определяется по следующему алгоритму.

1. Для исследуемого временного ряда определяется последовательность знаков, исходя из условий:

$$\delta_i = \begin{cases} +, & \text{если } y_{t+1} - y_t > 0, \\ -, & \text{если } y_{t+1} - y_t < 0. \end{cases} \quad (2).$$

При этом если последующее наблюдение равно предыдущему, то учитывается только одно наблюдение.

2. Подсчитывается число серий $\nu(n)$. Под серией понимается последовательность подряд расположенных плюсов и минусов, причем один плюс или минус считается серией.

3. Определяется протяженность самой длинной серии $l_{\max}(n)$.

4. По таблице, приведенной ниже, находится значение $l(n)$.

Длина ряда, n	$n \leq 26$	$26 < n < 153$	$153 < n < 170$
Значение $l(n)$	5	6	7

5. Если нарушается хотя бы одно из следующих неравенств, то гипотеза об отсутствии тренда отвергается с доверительной вероятностью 0,95:

$$\begin{cases} \nu(n) > \left[\frac{1}{3} \cdot (2n-1) - 1,96 \cdot \sqrt{\frac{16n-29}{90}} \right], \\ l_{\max}(n) \leq l(n). \end{cases}$$

Квадратные скобки неравенства означают целую часть числа.

Задача

Дана динамика ежеквартального выпуска продукции фирмы в ден. ед. с помощью критерия «восходящих и нисходящих» серий сделать вывод о присутствии или отсутствии тренда. Доверительную вероятность принять равной 0,95.

<i>t</i>	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
<i>y</i>	10	14	7	16	15	17	16	20	17	7	15	16	20	14	19	21

Решение.

1. Добавьте в таблицу дополнительную строку δ_i для записи последовательности знаков, для определения которой используйте формулу (2).
2. Подсчитайте число серий $\nu(n)$, протяженность самой длинной серии $l_{\max}(n)$ и определите $l(n)$ по таблице.
3. Запишите систему неравенств, проверьте их выполнение.
4. Сделайте вывод.

Лабораторная работа №3

Вычисление основных показателей динамики

Для количественной оценки динамики экономических процессов применяют такие статистические показатели, как *абсолютные приросты, темпы роста и прироста*. Они подразделяются на *цепные, базисные и средние*.

Если сравнение уровней временного ряда осуществляется с одним и тем же уровнем, принятым за базу, то показатели называются *базисными*. Если сравнение осуществляется с переменной базой, причем каждый последующий уровень сравнивается с предыдущим, то вычисленные показатели называются *цепными*.

Формулы для вычисления цепных, базисных и средних абсолютных приростов, темпов роста и темпов прироста даны в таблице (1). В формулах приняты обозначения:

$Y_t = Y_1, Y_2, \dots, Y_t, \dots, Y_n$ – уровни временного ряда, n – длина ряда, Y_δ – уровень временного ряда, принятый за базу для сравнения.

Обозначение	Абсолютный прирост	Темп роста	Темп прироста
--------------------	---------------------------	-------------------	----------------------

Цепной	$\Delta y_t = y_t - y_{t-1}$	$T_t = \frac{y_t}{y_{t-1}} \cdot 100\%$	$K_t = T_t - 100\%$
Базисный	$\Delta y_{\delta t} = y_t - y_{\delta}$	$T_{\delta t} = \frac{y_t}{y_{\delta}} \cdot 100\%$	$K_{\delta t} = T_{\delta t} - 100\%$
Средний	$\Delta \bar{y}_t = \frac{y_n - y_1}{n - 1}$	$\bar{T} = \sqrt[n-1]{\frac{y_n}{y_1}} \cdot 100\%$	$K_{\delta t} = \bar{T} - 100\%$

Описание динамики ряда средним приростом соответствует его представлению в виде прямой, проходящей через две крайние точки. Для получения прогнозного значения на один шаг вперед достаточно к последнему наблюдению добавить значение среднего абсолютного прироста:

$$\hat{y}_{n+1} = y_n + \Delta \bar{y}, \quad (3)$$

где y_n – значение показателя в n точке временного ряда, \hat{y}_{n+1} – прогнозируемое значение показателя в точке $n + 1$, $\Delta \bar{y}$ – значение среднего прироста временного ряда.

Получение прогнозного значения по формуле (3) корректно, если динамика ряда близка к линейной. На такой равномерный характер динамики указывают примерно одинаковые цепные абсолютные приросты.

Использование среднего темпа роста (среднего темпа прироста) для описания динамики развития ряда соответствует его представлению в виде показательной или экспоненциальной кривой, проведенной через две крайние точки, и характерно для процессов, изменение динамики которых происходит с постоянным темпом роста. Прогнозируемое значение на один шаг вперед определяется по формуле:

$$\hat{y}_{n+1} = y_n \cdot \bar{T}$$

Недостатком прогнозирования с использованием среднего прироста и среднего темпа роста является то, что они учитывают начальный и конечный уровни ряда, исключая влияние промежуточных уровней. Тем не менее, они используются как простейшие, приближенные способы прогнозирования.

Задача.

Ежеквартальная динамика фонда заработной платы работников фирмы в ден. ед. представлена в таблице.

t	1	2	3	4	5
y	8946,00	8981,50	9024,10	9063,15	9105,75

Обосновать правомерность использования среднего прироста для определения прогнозного значения фонда заработной платы в 6-м квартале.

Задача.

Изменение ежеквартальной динамики фонда заработной платы работников фирмы происходило примерно с постоянным темпом роста в течение 5 кварталов. Фонд заработной платы в 1-квартале составлял 9523,08 ден. ед., а в 5-м квартале – 9693,13 ден. ед.

Определить прогноз фонда заработной платы работников фирмы в 6-м квартале, используя средний темп роста.

Лабораторная №4

Сглаживание временных рядов

Предварительный анализ временных рядов экономических показателей заключается в выявлении аномальных значений уровней ряда, которые не соответствуют реальным возможностям рассматриваемой экономической системы, а также в определении наличия тренда.

Наиболее распространенным приемом для устранения аномальных значений показателей и отсутствия тенденции временного ряда является *сглаживание временного ряда*. При этом производится замена фактических уровней временного ряда расчетными, что способствует более четкому проявлению тенденции ряда. Скользящие средние позволяют сгладить случайные и периодические колебания временного ряда.

Наиболее распространенной процедурой сглаживания является метод *простой скользящей средней*. Сначала для временного ряда определяется интервал сглаживания (g). Если необходимо сгладить мелкие колебания, то интервал сглаживания берут по возможности большим, если нужно сохранить более мелкие колебания, то интервал сглаживания уменьшают. Для первых (g) уровней временного ряда вычисляется их среднее арифметическое значение. Это будет сглаженное значение уровня ряда, находящегося в интервале сглаживания. Затем интервал сглаживания сдвигается на один уровень вправо, повторяется вычисление средней арифметической и т.д. В результате такой процедуры получим ряд сглаженных значений, при этом в зависимости от (g) первые и последние уровни теряются.

Длину интервала сглаживания (g) удобно брать в виде нечетного числа, в этом случае расчетное значение скользящей средней будет приходиться на средний интервал ряда.

Например, для интервала $g = 3$ сглаженные уровни рассчитываются по формуле:

$$\hat{y}_t = (y_{t-1} + y_t + y_{t+1}) / 3$$

Чтобы не потерять первый и последний уровни ряда, их можно вычислить по формулам параболического интерполирования:

$$\hat{y}_1 = (5y_1 + 2y_2 - y_3) / 6,$$

$$\hat{y}_n = (-y_{n-2} + 2y_{n-1} + 5y_n) / 6.$$

Метод простой скользящей средней дает хорошие результаты в динамических рядах с линейной тенденцией развития.

Для рядов с нелинейной тенденцией развития необходимо применять метод *взвешенной скользящей средней*. Этот метод отличается от метода простой скользящей средней тем, что уровни, входящие в интервал сглаживания, суммируются с разными весами. Для полиномов 2-го и 3-го порядков по 5-членной взвешенной скользящей средней центральное значение определяется по формуле:

$$\hat{y}_t = 1/35 \cdot (-3y_{t-2} + 12y_{t-1} + 17y_t + 12y_{t+1} - 3y_{t+2}).$$

Весовые коэффициенты при сглаживании по полиномам 2-го и 3-го порядков в зависимости от длины интервала сглаживания представлены в таблице.

Длина интервала сглаживания	Весовой коэффициент
5	1/35 (-3, +12, +17)
7	1/21 (-2, +3, +6, +7)
9	1/231 (-21, +14, +39, +54, +59)

Выравнивание временных рядов может быть произведено методом *экспоненциального сглаживания*. Суть метода заключается в том, что в процедуре нахождения сглаженного уровня используются значения только предшествующих уровней рядов, взятые с определенным весом, причем вес наблюдения уменьшается по мере удаления его от момента времени, для которого определяется сглаженное значение уровня ряда. Если для исходного временного ряда $y_t = y_1, y_2, \dots, y_t, \dots, y_n$ соответствующие сглаженные значения уровней обозначить S_t , где $t = \overline{1, n}$, то экспоненциальное сглаживание производится по рекуррентному соотношению:

$$S_t = \alpha \cdot y_t + (1 - \alpha) \cdot S_{t-1},$$

где α – параметр сглаживания, причем $0 < \alpha < 1$, величина $(1 - \alpha)$ называется коэффициентом дисконтирования.

Используя рекуррентное соотношение для всех уровней ряда, начиная с первого и кончая моментом уровня t , можно получить, что экспоненциальная средняя, т.е. сглаженное данным методом значение уровня ряда, является взвешенной средней всех предшествующих уровней:

$$S_t = \alpha \sum (1 - \alpha)^i \cdot y_{t-i} + (1 - \alpha)^t S_0.$$

Обычно для временных рядов в экономических величинах величину параметра сглаживания выбирают в интервале от 0,1 до 0,3.

Начальный параметр S_0 принимают равным значению первого уровня ряда или равным средней арифметической нескольких первых членов ряда, например:

$$S_0 = \frac{y_1 + y_2 + y_3}{3}.$$

Задача

По данным динамики урожайности за 10 лет, приведенным в таблице, рассчитать:

- 3-, 5-летние скользящие средние простые;
- 5-летние скользящие средние взвешенные;
- сравнить результаты расчетов.

t	1	2	3	4	5	6	7	8	9	10
y	16,3	21,2	18,1	8,7	16,3	17,3	20,9	15,4	19,7	21,7

- постройте графики ряда динамики и трендов линейной фильтрации.

Задача

В таблице приведены данные численности преподавателей высших учебных заведений РФ (тыс. человек) по годам. Произвести сглаживание временного ряда с использованием экспоненциальной средней, приняв параметр сглаживания $\alpha = 0,1$ и $\alpha = 0,3$. По результатам расчетов определить, какой из сглаженных временных рядов носит более гладкий характер.

Год	1992	1993	1994	1995	1996	1997	1998	1999
y_t	233,5	239,9	239,8	261,9	261,8	268,7	260,7	298,6

Лабораторная №5

Выбор наилучшего вида тренда с помощью Excel

Для определения параметров линейного тренда по методу наименьших квадратов используется статистическая функция *ЛИНЕЙН*, для определения экспоненциального тренда *ЛГРФПРИБЛ*. В качестве независимой переменной выступает время $t = \overline{1, n}$.

В Excel линия тренда может быть добавлена в диаграмму с областями гистограммы или в график. Для этого необходимо:

- выделить область диаграммы; в главном меню выбрать *Диаграмма* → *Добавить линию тренда*;
- в появившемся диалоговом окне *Линия тренда* выбрать вид линии тренда и задать соответствующие параметры. Для полиномиального тренда необходимо задать степень аппроксимирующего полинома.

В качестве дополнительной информации на диаграмме можно отобразить уравнение регрессии и значение среднеквадратического отклонения, установив соответствующие флажки на закладке *Параметры* диалогового окна *Линия тренда*.

Задача

Динамика выпуска продукции Финляндии характеризуется данными (млн. долл.), представленными в таблице.

Год	Выпуск продукции	Год	Выпуск продукции
1961	1054	1979	11172
1962	1104	1980	14150
1963	1149	1981	14004
1964	1291	1982	13088
1965	1427	1983	12518
1966	1505	1984	13471
1967	1513	1985	13617
1968	1635	1986	16356
1969	1987	1987	20037
1970	2306	1988	21748
1971	2367	1989	23298
1972	2913	1990	26570
1973	3837	1991	23080
1974	5490	1992	23981
1975	5502	1993	23446
1976	6342	1994	29658
1977	7665	1995	39573
1978	8570	1996	38435

1. Провести расчет параметров линейного и экспоненциального трендов.
2. Построить графики ряда динамики и трендов.
3. Выбрать наилучший вид тренда на основании графического изображения и значения коэффициента детерминации.

Оценочные средства для проведения промежуточной аттестации

а) Планируемые результаты обучения и оценочные средства для проведения промежуточной аттестации:

Код индикатора	Индикатор достижения компетенции	Оценочные средства
ОПК-1 Способен применять естественнонаучные и общеинженерные знания, методы математического анализа и моделирования, теоретического и экспериментального исследования в профессиональной деятельности		
ОПК-1.1	Решает стандартные профессиональные задачи с применением естественнонаучных и общеинженерных знаний, методов математического анализа и моделирования	<p>Что такое регрессионный анализ?</p> <ol style="list-style-type: none"> 1. Что понимается под парной регрессией? 2. Опишите суть традиционного МНК для линейной парной регрессии. 3. Что такое «оценка параметра»? Чем отличаются «истинные» значения параметров регрессии от их оценок? 4. Что такое остатки в регрессионном анализе? 5. Опишите процедуру проверки гипотезы. 6. Что такое уровень доверия? 7. Что такое интервальный прогноз? Почему возникает необходимость построения точечных прогнозов? 8. Какой вид имеет система нормальных уравнений метода наименьших квадратов в случае линейной регрессии? 9. По какой формуле вычисляется линейный коэффициент парной корреляции? 10. Как вычисляется и что показывает индекс детерминации? Как вычисляется и что показывает индекс детерминации? 11. Как проверяется значимость уравнения регрессии и отдельных коэффициентов? 12. Как строится доверительный интервал прогноза в случае линейной регрессии? 13. Как вычисляются и что показывают коэффициент эластичности ε и средний коэффициент эластичности? <p><i>Пример практического задания</i></p> <ol style="list-style-type: none"> 1. Постройте линейную парную модель регрессии, описывающую зависимость заработной платы рабочего от его возраста по экспериментальным данным. 2. Рассчитать следующие показатели качества модели регрессии: <ul style="list-style-type: none"> - коэффициент детерминации; - коэффициент множественной корреляции; - средняя квадратическая ошибка уравнения регрессии; - ошибка аппроксимации. <p>Сделать выводы по каждому показателю.</p>

Код индикатора	Индикатор достижения компетенции	Оценочные средства
		<p><i>Пример комплексного задания</i></p> <ol style="list-style-type: none"> 1. Определите вид функции, наилучшим образом описывающей зависимость расходов на покупку продовольственных товаров в общих расходах (%) - y от среднедневной заработной платы одного работающего (руб) – x по данным семи территорий Уральского региона за 199X г. 2. Для характеристики зависимости y от x рассчитайте параметры следующих функций: <ul style="list-style-type: none"> • линейной; • степенной; • показательной; • равносторонней гиперболы. <p>Оцените каждую модель через среднюю ошибку аппроксимации A и F-критерий Фишера.</p>
ОПК-1.2	Решает профессиональные задачи с применением методов теоретического и экспериментального исследования	<p><i>Перечень вопросов для подготовки к зачету</i></p> <ol style="list-style-type: none"> 1. Что означает понятие «качественная» модель регрессии? 2. В каком случае верно соотношение ? 3. Что означает высказывание «тесная линейная связь»? 4. Что показывает коэффициент детерминации равный 0,75? 5. Что такое ошибка аппроксимации? 6. Приведите пример нелинейной регрессии по включаемым в нее объясняющим переменным, но линейной по оцениваемым параметрам. 7. Что значит внутренне линейная модель? 8. Опишите, что означает высказывание «функция, наилучшим образом описывающая зависимость y от x»? 9. Перечислите все виды моделей, нелинейных относительно: а) включаемых переменных; б) оцениваемых параметров. 10. Как проводится подбор линеаризующего преобразования для внутренне нелинейных моделей? 11. Назовите показатели корреляции, используемые при нелинейных соотношениях рассматриваемых признаков. 12. Какие задачи решаются при построении уравнения регрессии? 13. Какие требования предъявляются к факторам, включаемым в уравнение регрессии? 14. Что понимается под линейной множественной регрессией?

Код индикатора	Индикатор достижения компетенции	Оценочные средства
		<p>15. Чем отличаются стандартизованные коэффициенты регрессии от коэффициентов в естественном виде?</p> <p>16. Что показывает отрицательное значение коэффициента эластичности?</p> <p>17. Чем отличаются частный и общий критерии Фишера?</p> <p>18. Что понимается под коллинеарностью и мультиколлинеарностью факторов?</p> <p>19. Как проверяется наличие коллинеарности и мультиколлинеарности?</p> <p>20. Какой вид имеет система нормальных уравнений метода наименьших квадратов в случае линейной регрессии?</p> <p>21. По какой формуле вычисляется индекс множественной корреляции?</p> <p>22. Как вычисляются индекс множественной детерминации и скорректированный индекс множественной детерминации?</p> <p>23. Как проверяется значимость уравнения регрессии и отдельных коэффициентов?</p> <p>24. Как строятся частные уравнения регрессии?</p> <p>25. Как вычисляются средние частные коэффициенты эластичности?</p> <p>26. Что такое стандартизованные переменные?</p> <p>27. Какой вид имеет уравнение линейной регрессии в стандартизованном масштабе?</p> <p>28. Что понимается под гомоскедастичностью?</p> <p>29. Как проверяется гипотеза о гомоскедастичности ряда остатков?</p> <p>30. При каких условиях строится уравнение множественной регрессии с фиктивными переменными?</p> <p>31. Как трактуются коэффициенты модели при фиктивных переменных?</p> <p>32. Что такое ловушка фиктивных переменных и как избежать такой ситуации при моделировании?</p> <p><i>Пример практического задания</i></p> <p>Определить факторы, формировавшие цену квартир в строящихся домах в Санкт-Петербурге в 1996 г. по данным о рынке строящегося жилья в Санкт-Петербурге (по состоянию на декабрь 1996 г.)</p>

Код индикатора	Индикатор достижения компетенции	Оценочные средства
		<p>Сгенерируйте фиктивную переменную, отражающую местоположение квартиры и позволяющую разделить всю совокупность квартир на две группы: квартиры на севере города (Приморский район, Шувалово-Озерки, Гражданка) и на юге города (Юго-Запад, Красносельский район). Постройте уравнение регрессии, характеризующее зависимость цены от всех факторов, в линейной и степенной форме.</p> <p>Существует ли разница в ценах квартир, расположенных в северной и южной частях Санкт-Петербурга? Является ли наличие балкона и лоджии преимуществом квартиры на рынке? Как вы объясните этот факт?</p> <p><i>Пример комплексного задания</i></p> <p>По 20 предприятиям региона определить зависимость выработки продукции на одного работника y (тыс.руб.) от ввода в действие новых основных фондов x_1 (% от стоимости фондов на конец года) и от удельного веса рабочих высокой квалификации в общей численности рабочих x_2 (%).</p>

б) Порядок проведения промежуточной аттестации, показатели и критерии оценивания:

Промежуточная аттестация по дисциплине проводится в форме зачета.

Показатели и критерии:

– на оценку «**зачтено**» – обучающийся показывает высокий уровень сформированности компетенций, т.е. принимает активное участие в обсуждении, владеет терминологическим аппаратом, демонстрирует знания в области эконометрического моделирования; осуществляет выбор эффективной модели, на основе проведения необходимых расчетов и учета всех представленных в условии показателей, грамотно обосновывает свое решение и формулирует необходимые выводы.

– на оценку «**не зачтено**» – результат обучения не достигнут, обучающийся не может показать знания на уровне воспроизведения и объяснения информации, не может показать интеллектуальные навыки решения простых задач

