

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ УКРАИНЫ

*ХАРЬКОВСКИЙ НАЦИОНАЛЬНЫЙ ЭКОНОМИЧЕСКИЙ
УНИВЕРСИТЕТ ИМЕНИ СЕМЕНА КУЗНЕЦА*

*Малярец Л. М.
Железнякова Э. Ю.
Норик Л. А.*

ЭКОНОМЕТРИКА

в примерах и задачах

**Учебное пособие
для иностранных студентов**

Харьков. Изд. ХНЭУ им. С. Кузнецца, 2014

УДК 330.43(075.8)

ББК 65в6я73

M21

Рецензенты: докт. экон. наук, профессор кафедры математики и математических методов экономики Донецкого национального университета *Христиановский В. В.*; докт. экон. наук, доцент, проректор по подготовке научных кадров Восточноевропейского университета экономики и менеджмента *Ус Г. А.*

Рекомендовано к изданию решением ученого совета Харьковского национального экономического университета имени Семена Кузнеця.

Протокол № 9 от 22.04.2014 г.

Малярец Л. М.

M21 Эконометрика в примерах и задачах : учебное пособие для иностранных студентов / Л. М. Малярец, Э. Ю. Железнякова, Л. А. Норик. – Х. : Изд. ХНЭУ им. С. Кузнеця, 2014. – 268 с. (Рус. яз.)

Изложен материал учебной дисциплины, которая преподается бакалаврам отрасли знаний 0305 "Экономика и предпринимательство". Изложен материал, который сопровождается большим количеством примеров, имеющих как учебный, так и исследовательский характер. Приведены реальные примеры экономики субъектов хозяйствования различных уровней управления.

Рекомендовано для студентов, магистров, аспирантов, которые изучают эконометрические модели.

ISBN 978-966-676-550-8

УДК 330.43(075.8)

ББК 65в6я73

© Харьковский национальный
экономический университет
имени Семена Кузнеця, 2014

© Малярец Л. М., Железнякова Э. Ю.,
Норик Л. А., 2014

Введение

Широкое распространение моделирования и количественного анализа обусловило формирование такого направления экономических исследований, как эконометрика. Эконометрика – это раздел экономической науки, который изучает методы количественного измерения взаимосвязей между экономическими показателями, характеризующими реальные экономические явления и процессы. Эконометрика является одним из базовых курсов фундаментальной подготовки бакалавров экономики.

Основным *заданием* учебной дисциплины "Эконометрика" является изучение методов оценивания параметров эконометрической модели и особенностей экономической информации с целью количественного измерения взаимосвязи между исследуемыми процессами и явлениями. *Целью* изучения данной учебной дисциплины является формирование у студентов системы знаний математического инструментария построения и использования разных типов эконометрических моделей, приобретения необходимых теоретических и практических знаний для решения конкретных заданий по построению экономико-математических моделей реальных экономических объектов.

При изучении данной учебной дисциплины студент получает знания и приобретает навыки, которые имеет возможность применять при последующем изучении дисциплин экономического направления и в научно-исследовательской работе. Приобретенные студентами в результате изучения дисциплины знания и навыки по применению математических методов для построения эконометрических моделей, обоснования методики их анализа составляют основу *компетентностей* современного специалиста в контексте Национальной рамки квалификаций:

знания основных математических методов и принципов эконометрики для построения эконометрических моделей, которые применяются для решения задач количественного анализа и прогноза процессов функционирования и развития социально-экономических систем;

знания способов определения точности и надежности методов, которые применяются для обработки эмпирических данных;

умение выполнять постановку и формализацию практических задач согласно общей технологии моделирования в экономике;

умение осуществлять отбор факторов для построения уравнения множественной регрессии, определять статистические оценки параметров многофакторной эконометрической модели и проверять их статистическую значимость;

умение проводить оценивание параметров уравнения регрессии для временных рядов и определять пригодность этого уравнения для прогнозирования;

умение применять статистические критерии для определения соответствия между эконометрической моделью, которая построена по выборочным данным и основным гипотезам, которые лежат в ее основе;

умение строить экономические модели на основе системы структурных уравнений;

умение строить динамические эконометрические модели;

умение моделировать и анализировать временные ряды;

навыки использования математических методов, которые являются инструментом экономико-математического моделирования при решении практических задач экономики;

навыки применения программной среды MS Excel к решению задач количественного (эконометрического) анализа;

способности автономно и ответственно выполнять задания;

коммуникации по взаимодействию студентов при совместном решении задач и принятии согласованных решений.

В процессе изложения материала учебного пособия "Эконометрика в примерах и задачах" авторами систематически обсуждаются причины, по которым приходится пересматривать модели, уточнять систему представлений, опираясь на которые с помощью статистических данных можно конструировать новые варианты соответствующей модели. Такой подход практически обусловил структуру данного учебного пособия, где сначала рассматривается применение классического регрессионного анализа для построения эконометрической модели, а далее исследуются эконометрические задачи, соответствующие реальным экономическим условиям, когда нарушаются исходные гипотезы регрессионного анализа.

1. Особенности эконометрических моделей и принципы их построения

1.1. Основные теоретические сведения

Эконометрическое знание выделилось и сформировалось как результат развития и взаимодействия экономической теории, экономической статистики, математической статистики и теории вероятностей (рис. 1.1).

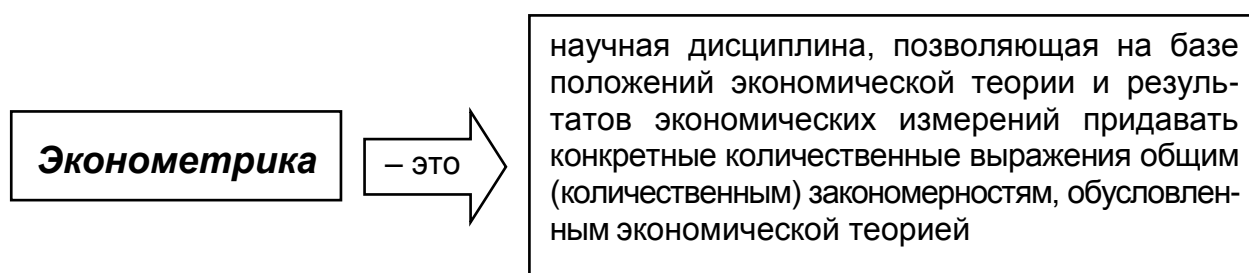


Рис. 1.1. Понятие "эконометрика"

Цель эконометрики – это разработка способов моделирования, количественного анализа и прогнозирования показателей, характеризующих реальные экономические объекты.

Задачи эконометрики:

спецификация модели – построение эконометрических моделей для эмпирического анализа;

параметризация – оценка параметров моделей;

верификация модели – проверка качества параметров модели и модели в целом;

прогнозирование – составление прогнозов и рекомендаций по результатам эконометрического моделирования.

Типы данных в эконометрическом моделировании:

пространственные данные – набор сведений по разным объектам, взятым за один и тот же период времени;

временные – набор сведений, характеризующих один и тот же объект за разные периоды времени.

Виды переменных эконометрического моделирования:

экзогенные переменные – признаки или показатели, значения которых задаются извне модели и они являются управляемыми или планируемыми (независимые);

эндогенные – переменные, которые формируются в процессе функционирования социально-экономической системы и зависят от экзогенных переменных или от их взаимодействия (зависимые);

лаговые переменные – эндогенные или экзогенные переменные, значения которых измерены в предыдущие моменты времени и являются уже заданными.

Процесс построения эконометрической модели в экономике имеет свои особенности (рис.1.2).



Рис. 1.2. **Этапы построения эконометрической модели в экономике**

Описание количественных взаимосвязей между признаками осуществляют эконометрические модели разных типов (рис. 1.3).

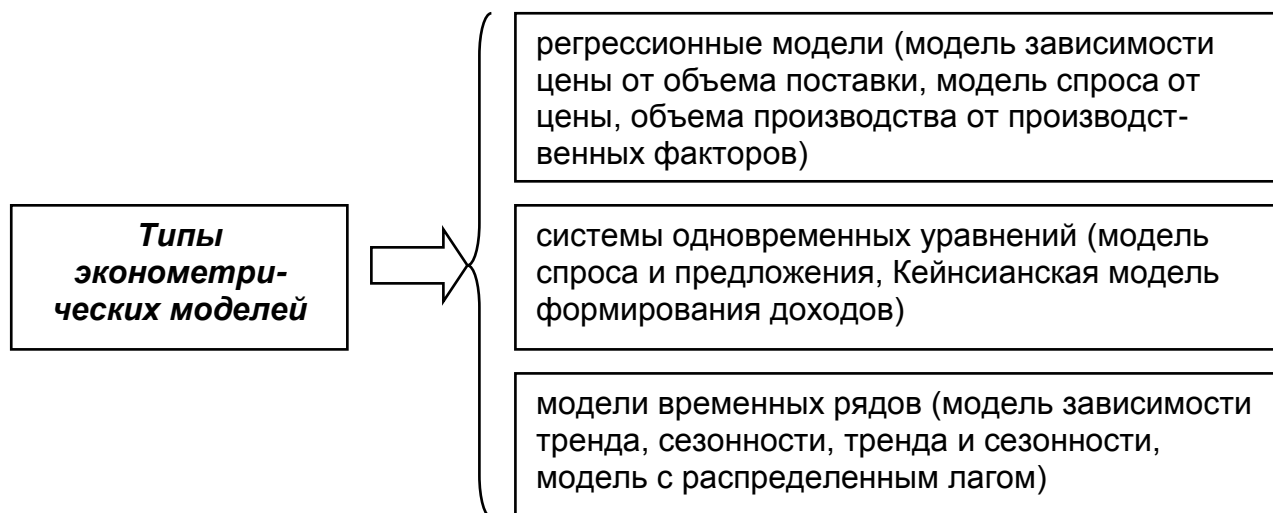


Рис. 1.3. Типы эконометрических моделей

1.2. Примеры

Пример 1.1. Модель "Спрос – предложение" содержит функции спроса, предложения и условие равновесия:

$$\begin{cases} q_t^D = \alpha_0 + \alpha_1 p_t + \varepsilon_{t1}, & \alpha_1 < 0, \\ q_t^S = \beta_0 + \beta_1 p_t + \varepsilon_{t2}, & \beta_1 < 0, \\ q_t^D = q_t^S. \end{cases}$$

В представленной системе первое уравнение – функция спроса, второе уравнение – функция предложения, третье уравнение – условие равновесия, p_t – цена товара в момент времени t , ε_{t1} и ε_{t2} – погрешности моделей спроса и предложения соответственно.

Охарактеризовать типы данных и виды переменных данной модели.

Решение.

В модели исследуются временные данные, так как они характеризуют определенные объекты за разные периоды времени.

Переменные в системах одновременных уравнений бывают двух видов: эндогенные (их значения определяются внутри модели) и экзогенные (внешние по отношению к модели, их значения считаются фиксированными).

В представленной модели переменные q_t^D , q_t^S оцениваются внутри модели, поэтому они являются эндогенными переменными.

p_t – цена товара в момент времени является внешней переменной по отношению к модели – экзогенная переменная.

Пример 1.2. Указать, какие переменные в модели являются эндогенными, какие экзогенными, какие лаговыми в упрощенной модели рынка, которая представлена уравнением спроса, уравнением предложения и условием равновесия:

$$Q_t^s = a_2 p_t + e_t;$$

$$Q_t^d = b_1 p_t + b_2 y_{t-1} + e_2;$$

$$Q_t^s = Q_t^d,$$

где Q^S – предложение товара;

Q^d – спрос на товар;

p – цена товара;

y – величина совокупного дохода;

t – текущий период времени;

$t-1$ – предыдущий период времени.

Решение.

В модели исследуются временные данные, так как они характеризуют определенные объекты за разные периоды времени.

Переменные Q_t^d , Q_t^s оцениваются внутри модели, поэтому они являются эндогенными переменными.

p_t – цена товара в момент времени является внешней переменной по отношению к модели – экзогенная переменная.

y_{t-1} – величина совокупного дохода, значения которой измерены в предыдущий момент времени и являются уже заданными – лаговая переменная.

Пример 1.3. Модель динамики цены и заработной платы имеет вид:

$$\begin{cases} y_1 = b_{12}y_2 + a_{11}x_1 + \varepsilon_1; \\ y_2 = b_{21}y_1 + a_{22}x_2 + a_{23}x_3 + \varepsilon_2, \end{cases}$$

где y_1 – темп изменения месячной заработной платы;

y_2 – темп изменения цен;

x_1 – процент безработных;

x_2 – темп изменения постоянного капитала;

x_3 – темп изменения цен на импорт сырья.

Идентифицировать тип эконометрической модели.

Решение.

В системе переменные y_1, y_2 одновременно рассматриваются как зависимые и как независимые. Такая система взаимозависимых уравнений является эконометрической моделью системы совместных, одновременных уравнений.

Пример 1.4. Привести пример построения системы одновременных уравнений эконометрической модели зависимости между заработной платой и производительностью работы на предприятии.

Решение

Построим эконометрическую модель, которая описывается системой одновременных уравнений:

$$\begin{aligned} Y_1 &= f(Y_2, X_1, X_2, X_3, u_1) \\ Y_2 &= f(Y_1, X_1, X_2, X_4, X_5, u_2), \end{aligned}$$

где Y_1 – заработная плата; Y_2 – производительность труда; X_1 – уровень квалификации работающих; X_2 – стаж работающих; X_3 – форма оплаты труда; X_4 – фондоотдача; X_5 – текучесть рабочей силы; u_1, u_2 – остатки в первом и втором уравнениях модели соответственно.

1.3. Задачи

Задача 1.1. Упрощенная модель закрытой экономики состоит из уравнений функции потребления, инвестиционной функции и тождества для национального дохода:

$$C_t = a_1 + a_2 Y_t + e_t;$$

$$I_t = b_1 + b_2 i + e_2;$$

$$Y_t = C_t + I_t + G_{t-1},$$

где C – совокупный объем личных потребительских расходов;

I – объем инвестиций;

G – совокупные государственные расходы;

Y – валовой выпуск;

i – ставка процента;

t – текущий период времени;

$t-1$ – предыдущий период времени.

Произвести деление переменных модели на эндогенные, экзогенные, лаговые. Обосновать ответ.

Задача 1.2. Макроэкономическая модель экономики представлена следующей системой уравнений:

$$C_t = a_{01} + a_{11} Y_t + a_{12} C_{t-1} + \varepsilon_1;$$

$$I_t = a_{02} + a_{21} Y_t + a_{22} i_t + \varepsilon_2;$$

$$i_t = a_{03} + a_{31} Y_t + a_{32} M_t + a_{33} i_{t-1} + \varepsilon_3;$$

$$Y_t = C_t + I_t + G_t,$$

где Y – валовой внутренний продукт;

C – расходы на личное потребление;

I – валовые инвестиции;

G – государственные расходы;

M – денежная масса;

i – процентная ставка на капитал;

t – текущий период времени;
 $t-1$ – предыдущий период времени.

Произвести деление переменных модели на эндогенные, экзогенные, лаговые. Обосновать ответ.

Задача 1.3. Модифицированная модель Кейнса выглядит следующим образом:

$$C_t = a_{01} + a_{11}Y_t + a_{21}Y_{t-1} + \varepsilon_1;$$

$$I_t = a_{02} + a_{12}Y_t + a_{22}Y_{t-1} + \varepsilon_2;$$

$$Y_t = C_t + I_t + G_t,$$

где Y – национальный доход;
 C – расходы на личное потребление;
 I – чистые инвестиции;
 G – государственные расходы, не связанные с заработной платой;
 t – текущий период времени;
 $t-1$ – предыдущий период времени.

Охарактеризовать типы данных и виды переменных данной модели. Идентифицировать тип эконометрической модели.

Задача 1.4. Известна модель денежного рынка:

$$R_t = a_{01} + a_{11}M_t + a_{21}Y_t + \varepsilon_1;$$

$$Y_t = a_{02} + a_{12}R_t + a_{22}I_t + \varepsilon_2,$$

где R – процентная ставка;
 Y – валовой внутренний продукт;
 M – денежная масса;
 I – внутренние инвестиции;
 t – время.

Охарактеризовать типы данных и виды переменных данной модели. Идентифицировать тип эконометрической модели.

2. Парная регрессия и корреляция в эконометрических исследованиях

2.1. Основные теоретические сведения

Для случайных величин X и Y не всегда можно установить функциональную зависимость. Между такими величинами существует зависимость, при которой с изменением одной величины изменяется распределение другой величины. Такая зависимость называется *стохастической (статистической)*. Частным случаем статистической зависимости является *корреляционная зависимость* между двумя случайными величинами, при которой с изменением одной из них изменяется среднее значение другой.

Цель корреляционно-регрессионного анализа заключается в установлении факта наличия или отсутствия зависимостей между несколькими показателями и описании этих связей достаточно простыми выражениями. Среди всех заданных показателей (признаков, переменных) один показатель считается результативным признаком (откликом) и на этот показатель оказывают влияние остальные объясняющие переменные факторы.

Различают такие *виды корреляции и регрессии*:

1) *относительно характера* корреляции и регрессии имеем положительную или отрицательную корреляцию и регрессию;

2) *относительно числа переменных*: парная или множественная корреляция и регрессия;

3) *относительно формы связи*: линейная или нелинейная корреляция и регрессия.

Линейная парная регрессия сводится к нахождению уравнения вида: $\hat{y}_x = a + b \cdot x$ или $y = a + b \cdot x + \varepsilon$. Реальные значения зависимой переменной не всегда совпадают с ее условными средними и могут быть различными при одном и том же значении объясняющей переменной. Поэтому зависимость должна быть дополнена слагаемым ε , что есть случайной величиной и указывает на стохастическую суть зависимости. Присутствие в модели случайной величины ε обусловлено тремя причинами: спецификацией модели, выборочным характером исходных данных, особенностями измерения переменных.

Параметр b называется *коэффициентом линейной регрессии*. Он характеризует наклон прямой к оси абсцисс или $b = \operatorname{tg}\alpha$, где α – угол, который прямая регрессии образует с осью абсцисс.

Коэффициент регрессии является мерой зависимости переменной y от переменной x или мерой влияния, оказываемого изменением переменной x на переменную y . Если $b > 0$ имеем положительную линейную регрессию, демонстрирующую рост зависимой переменной при увеличении значений объясняющей переменной x . Если $b < 0$, то имеем отрицательную линейную регрессию, при которой с увеличением значений x значения переменной y убывают.

Построение линейной регрессии сводится к оценке ее параметров a и b . Классический подход к оцениванию параметров линейной регрессии основан на *методе наименьших квадратов* (МНК), позволяющем получить такие оценки параметров, при которых сумма квадратов отклонений фактических значений результативного признака y от теоретических \hat{y}_x минимальна:

$$\sum_{i=1}^n (y_i - \hat{y}_{x_i})^2 \rightarrow \min, \quad \varepsilon_i = y_i - \hat{y}_{x_i}; \quad \sum_{i=1}^n \varepsilon_i^2 \rightarrow \min.$$

Формулы для определения значения параметров регрессии имеют вид:

$$b = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2}; \quad a = \frac{\sum_{i=1}^n y_i - b \sum_{i=1}^n x_i}{n};$$

или

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

$$\text{где } \bar{y} = \frac{\sum_{i=1}^n y_i}{n}; \quad \bar{x} = \frac{\sum_{i=1}^n x_i}{n}; \quad \overline{yx} = \frac{\sum_{i=1}^n y_i x_i}{n}.$$

Тесноту связи изучаемых явлений оценивает линейный коэффициент парной корреляции для линейной регрессии:

$$r_{xy} = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sqrt{\left(n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2 \right) \left(n \sum_{i=1}^n y_i^2 - \left(\sum_{i=1}^n y_i \right)^2 \right)}};$$

или

$$r_{xy} = b \frac{\sigma_x}{\sigma_y} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x \sigma_y} \quad (-1 \leq r_{xy} \leq 1).$$

Коэффициент корреляции r_{xy} показывает на сколько величин σ_y изменится в среднем Y , если X увеличится на одну величину σ_x .

Величина $\overline{xy} - \bar{x} \cdot \bar{y}$ называется ковариацией или корреляционным моментом и обозначается μ_{xy} или $\text{cov}(x, y)$. Ковариация характеризует сопряженность вариаций двух признаков и представляет собой статистическую меру взаимодействия двух случайных величин.

Для оценки качества подбора линейной функции используют квадрат линейного коэффициента корреляции r_{xy}^2 , который называется коэффициентом детерминации $R^2 = r_{xy}^2$.

Нелинейные регрессии делятся на два класса:

регрессии, нелинейные относительно включенных в анализ объясняющих переменных, но линейные по оцениваемым параметрам (полиномы разных степеней, например, $y = a + b_1 x + b_2 x^2 + b_3 x^3 + \varepsilon$;

равносторонняя гипербола $y = a + \frac{b}{x} + \varepsilon$);

регрессии, нелинейные по оцениваемым параметрам (показательная $y = a \cdot b^x \cdot \varepsilon$; степенная $y = a \cdot x^b \cdot \varepsilon$; экспоненциальная $y = e^{a+b \cdot x} \cdot \varepsilon$).

Для оценки качества нелинейных регрессионных моделей целесообразно использовать корреляционное отношение (η):

$$\eta = \sqrt{1 - \frac{\sum (y - \hat{y}_x)^2}{\sum (y - \bar{y})^2}}, \quad 0 \leq \eta \leq 1,$$

или индекс детерминации η^2 .

Чем ближе индекс детерминации к единице, тем ближе корреляционная зависимость к функциональной.

Приведенные виды нелинейной зависимости между функциональным признаком (y) и фактором (x) можно с помощью некоторых преобразований привести к линейному виду. Способы линеаризации нелинейных функций приведены в табл. 2.1.

Таблица 2.1

Способы линеаризации нелинейных функций

№ п/п	Вид функции	Преобразования исходных данных	Оцениваемые параметры
1	$y = ax + b$	$X = x, Y = y$	$A = a, B = b$
2	$y = ax^b$	$X = \ln x, Y = \ln y$	$A = b, B = \ln a$
3	$y = ab^x$	$X = x, Y = \ln y$	$A = \ln b, B = \ln a$
4	$y = \frac{a}{x} + b$	$X = \frac{1}{x}, Y = y$	$A = a, B = b$
5	$y = \frac{1}{ax + b}$	$X = x, Y = \frac{1}{y}$	$A = a, B = b$
6	$y = \frac{ax}{1 + bx}$	$X = \frac{1}{x}, Y = \frac{1}{y}$	$A = \frac{b}{a}, B = \frac{1}{a}$
7	$y = ae^{bx}$	$X = x, Y = \ln y$	$A = b, B = \ln a$
8	$y = ae^{b/x}$	$X = \frac{1}{x}, Y = \ln y$	$A = b, B = \ln a$
9	$y = a \ln x + b$	$X = \ln x, Y = y$	$A = a, B = b$

2.2. Примеры

Пример 2.1. Изучается зависимость себестоимости единицы изделия (y , тыс. грн) от величины выпуска продукции (x , тыс. шт.) по группам предприятий за отчетный период.

Экономист провел исследование $n = 5$ групп предприятий и получил результаты, представленные в табл. 2.2.

Таблица 2.2

Исходные данные

x_i	1	1,5	2	2,5	3
y_i	2,15	2,3	2,6	2,8	2,5

В предположении линейной зависимости определить выборочное уравнение линейной регрессии $\hat{y}_x = b + a \cdot x$. Выполнить прогноз на основе построенной линейной модели для $x = 3,5$.

Решение.

Исходные данные и результаты вспомогательных вычислений представлены в табл. 2.3.

Таблица 2.3

Вспомогательные вычисления

n	x_i	y_i	x_i^2	$x_i y_i$
1	1	2,15	1	2,15
2	1,5	2,3	2,25	3,45
3	2	2,6	4	5,2
4	2,5	2,8	6,25	7
5	3	2,5	9	7,5
Σ	10	12,35	22,5	25,3

Выполняя вычисления основных числовых характеристик, получим:

$$\bar{x} = \frac{10}{5} = 2; \quad \bar{y} = \frac{12,35}{5} = 2,47; \quad \overline{x^2} = \frac{22,5}{5} = 4,5; \quad \overline{xy} = \frac{25,3}{5} = 5,06;$$

$$a = \frac{\overline{yx} - \bar{y}\bar{x}}{\overline{x^2} - \bar{x}^2} = \frac{5,06 - 2,47 \cdot 2}{4,5 - 2^2} = 0,24;$$

$$b = \bar{y} - a \cdot \bar{x} = 2,47 - 0,24 \cdot 2 = 1,99.$$

Искомое уравнение имеет вид: $\bar{y}_x = 0,24x + 1,99$.

Проверим уравнение:

$$\bar{y}_{x=1,5} = 0,24 \cdot 1,5 + 1,99 = 2,35; \quad \bar{y}_{x=2,5} = 0,24 \cdot 2,5 + 1,99 = 2,6.$$

Отклонения: $\varepsilon_1 = |2,35 - 2,3| = 0,05$; $\varepsilon_2 = |2,8 - 2,6| = 0,2$.

На рис. 2.1 изображены эмпирические точки и найденная теоретическая линия регрессии \hat{y}_x .

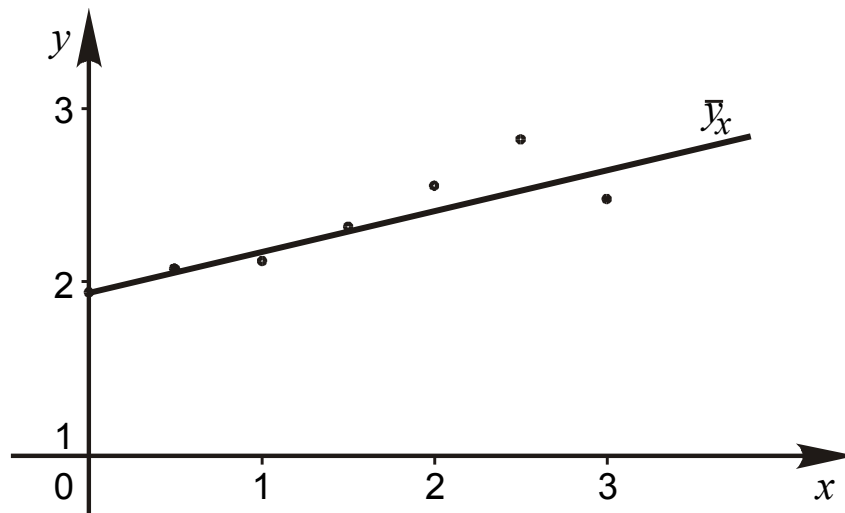


Рис. 2.1. Теоретическая линия регрессии \hat{y}_x и эмпирические точки

Найдем прогнозное значение себестоимости y при выпуске продукции $x = 3,5$ тыс. шт.: $y(3,5) = 1,99 + 0,24 \cdot 3,5 = 2,83$ тыс. грн.

Пример 2.2. Для признаков x_i и y_k при объеме выборки $n = 100$ получено: $\sum x_i h_i = 2\,296$, $\sum y_k l_k = 269$, $\sum x_i y_k m_{ik} = 6\,272$, $s_x^2 = 7,45$, $s_y^2 = 10,21$, где h_i, l_k, m_{ik} – соответствующие частоты.

Оценить тесноту связи.

Решение.

Из данных условий получим:

$$\bar{x} = \frac{2\,296}{100} = 22,96; \quad \bar{y} = 2,69; \quad \overline{xy} = 62,72;$$

$$\mu_{xy} = \overline{xy} - \bar{x} \cdot \bar{y} = 62,72 - 22,96 \cdot 2,69 = 0,9576.$$

Коэффициент корреляции: $r = \frac{\mu_{xy}}{s_x \cdot s_y} = \frac{0,9576}{\sqrt{7,45 \cdot 10,21}} = 0,110.$

Так как число r близко к нулю, значит линейная корреляционная связь очень слабая.

Пример 2.3. На основании данных 2012 г. о валовом региональном продукте и величине основных средств по регионам Украины (табл. 2.4) построить уравнение парной линейной регрессионной модели, которая описывает зависимость величины валового регионального продукта от величины основных средств.

Дать экономическую интерпретацию полученных результатов.

Таблица 2.4

Исходные данные

№ п/п	Область	Валовой региональный продукт, млн грн	Основные средства, млн грн
		y_i	x_i
1	2	3	4
1	Автономная Республика Крым	27 365	136 268
2	Винницкая	20 094	98 021
3	Волынская	12 784	34 569
4	Днепропетровская	104 687	341 076
5	Донецкая	117 646	430 793
6	Житомирская	15 008	44 715
7	Закарпатская	13 208	50 544
8	Запорожская	42 445	154 835
9	Ивано-Франковская	17 883	50 434
10	Киевская	35 687	111 959
11	Кировоградская	13 961	82 085
12	Луганская	42 985	113 452
13	Львовская	35 534	152 120

1	2	3	4
14	Николаевская	19 410	93 411
15	Одесская	46 994	151 666
16	Полтавская	34 118	144 432
17	Ровенская	14 074	49 826
18	Сумская	16 210	49 337
19	Тернопольская	10 618	26 809
20	Харьковская	59 389	399 665
21	Херсонская	13 174	49 369
22	Хмельницкая	16 061	56 896
23	Черкасская	19 101	62 838
24	Черновицкая	8 833	29 696
25	Черниговская	14 918	53 025
26	г. Киев	169 564	921 296
27	г. Севастополь	6 305	14 579

Решение.

Расчеты для построения уравнения регрессии удобно оформить в виде табл. 2.5.

По приведенным расчетным данным вычислим параметры уравнения регрессии $\hat{y}_x = a_0 + a_1 \cdot x$.

Система нормальных уравнений имеет вид:

$$\begin{cases} 27 \cdot a_0 + a_1 \cdot 3\,903\,716 = 948\,056; \\ a_0 \cdot 3\,903\,716 + a_1 \cdot 14\,979\,155\,897\,74 = 316\,281\,967\,655; \end{cases}$$

$$\begin{cases} a_0 = 7\,357,0691; \\ a_1 = 0,1920. \end{cases}$$

Уравнение парной линейной корреляционно-регрессионной модели, которая описывает зависимость величины валового регионального продукта (млн грн) от величины основных средств (млн грн) по регионам Украины имеет вид:

$$\hat{y} = 7\,357,0691 + 0,192x.$$

Таблица 2.5

Расчетная таблица

№ п/п	y_i	x_i	y_i^2	x_i^2	$y_i \cdot x_i$
1	27 365	136 268	748 843 225	18 568 967 824	3 728 973 820
2	20 094	98 021	403 768 836	9 608 116 441	1 969 633 974
3	12 784	34 569	163 430 656	1 195 015 761	441 930 096
4	104 687	341 076	10 959 367 969	116 332 837 776	35 706 223 212
5	117 646	430 793	13 840 581 316	185 582 608 849	50 681 073 278
6	15 008	44 715	225 240 064	1 999 431 225	671 082 720
7	13 208	50 544	174 451 264	2 554 695 936	667 585 152
8	42 445	154 835	1 801 578 025	23 973 877 225	6 571 971 575
9	17 883	50 434	319 801 689	2 543 588 356	901 911 222
10	35 687	111 959	1 273 561 969	12 534 817 681	3 995 480 833
11	13 961	82 085	194 909 521	6 737 947 225	1 145 988 685
12	42 985	113 452	1 847 710 225	12 871 356 304	4 876 734 220
13	35 534	152 120	1 262 665 156	23 140 494 400	5 405 432 080
14	19 410	93 411	376 748 100	8 725 614 921	1 813 107 510
15	46 994	151 666	2 208 436 036	23 002 575 556	7 127 392 004
16	34 118	144 432	1 164 037 924	20 860 602 624	4 927 730 976
17	14 074	49 826	198 077 476	2 482 630 276	701 251 124
18	16 210	49 337	262 764 100	2 434 139 569	799 752 770
19	10 618	26 809	112 741 924	718 722 481	284 657 962
20	59 389	399 665	3 527 053 321	159 732 112 225	23 735 704 685
21	13 174	49 369	173 554 276	2 437 298 161	650 387 206
22	16 061	56 896	257 955 721	3 237 154 816	913 806 656
23	19 101	62 838	364 848 201	3 948 614 244	1 200 268 638
24	8 833	29 696	78 021 889	881 852 416	262 304 768
25	14 918	53 025	222 546 724	2 811 650 625	791 026 950
26	169 564	921 296	28 751 950 096	848 786 319 616	156 218 634 944
27	6 305	14 579	39 753 025	212 547 241	91 920 595
Σ	948 056	3 903 716	70 954 398 728	1 497 915 589 774	316 281 967 655

В предлагаемой модели $a_1 = 0,1920$. Поскольку значение коэффициента регрессии отличается от нуля, то корреляционная связь между величиной валового регионального продукта и величиной основных

средств существует. Так как $a_1 = 0,1920 > 0$, то при увеличении величины основных средств в регионах на один млн грн среднее значение валового регионального продукта по регионам Украины должно расти на 0,192 млн грн.

Свободный член представленной модели $a_0 = 7\,357,0691$ показывает, что при нулевом значении величины основных средств валовой региональный продукт в среднем должен иметь значение, которое равно 7 357,0691 млн грн.

Пример 2.4. В табл. 2.6 приведены эмпирические данные о расходах на оплату труда x_i (экзогенный фактор) и цене продукции y_i (эндогенный фактор), которые получены из семи наблюдений.

Таблица 2.6

Исходные данные

№ наблюдения	x_i	y_i
1	45,1	68,8
2	59,0	61,2
3	57,2	59,9
4	61,8	56,7
5	58,8	55,0
6	47,2	54,3
7	55,2	49,3

Построить линейную, степенную и показательную модели, вычислить коэффициент корреляции, сделать выводы.

Решение.

Рассмотрим линейную аппроксимацию, поскольку другие модели с помощью преобразований и использования вспомогательных переменных сводятся к построению линейной модели. В предположении линейной аппроксимации уравнение регрессии имеет вид:

$$\tilde{y}_x = b_0 + b_1x.$$

Выполним вспомогательные вычисления (табл. 2.7).

Вспомогательные вычисления для оценки параметров линейной модели "Цена товара – расходы на оплату труда"

№ п/п	x_i	y_i	$x_i y_i$	x_i^2	y_i^2
1	45,1	68,8	3 102,88	2 034,01	4 733,44
2	59,0	61,2	3 610,80	3 481,00	3 745,44
3	57,2	59,9	3 426,28	3 271,84	3 588,01
4	61,8	56,7	3 504,06	3 819,24	3 214,89
5	58,8	55,0	3 234,00	3 457,44	3 025,00
6	47,2	54,3	2 562,96	2 227,84	2 948,49
7	55,2	49,3	2 721,36	3 047,04	2 430,49
Σ	384,3	405,2	22 162,34	21 338,41	23 685,76

Последняя строка табл. 2.7 содержит суммы по каждому из столбиков и используется для вычисления средних значений.

По данным полученной расчетной таблицы вычисляем выборочные средние, ковариацию, дисперсию и параметры уравнения линейной модели:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{384,3}{7} = 54,9; \quad \bar{y} = \frac{\sum_{i=1}^n y_i}{n} = 57,89;$$

$$\mu_{xy} = \overline{xy} - \bar{x} \bar{y} = \frac{\sum_{i=1}^n x_i y_i}{n} - \bar{x} \cdot \bar{y} = \frac{22\,162,34}{7} - 54,9 \cdot 57,89 = -11,88;$$

$$\sigma_x^2 = \overline{x^2} - \bar{x}^2 = \frac{\sum_{i=1}^n x_i^2}{n} - \bar{x}^2 = \frac{21\,338,41}{7} - (54,9)^2 = 34,33;$$

$$b_1 = \frac{\mu_{xy}}{\sigma_x^2} = \frac{\overline{xy} - \bar{x} \bar{y}}{\overline{x^2} - \bar{x}^2} = -0,35; \quad b_0 = \bar{y} - b_1 \bar{x} = 76,88.$$

Таким образом, теоретическое уравнение линейной регрессии имеет вид:

$$\tilde{y}_x = 76,88 - 0,35x.$$

Объясним, какую информацию можно получить по значению коэффициента регрессии. Поскольку $b_1 < 0$, то увеличение значения x ведет к уменьшению значения y . То есть корреляционная связь между факторами является обратной. По абсолютной величине коэффициента регрессии можно сделать вывод, что увеличение расходов на оплату труда на 1 единицу приводит к уменьшению цены товара на 0,35 единиц.

Для оценки тесноты корреляционной связи вычислим коэффициент корреляции:

$$r_{xy} = \frac{\mu_{xy}}{\sigma_x \sigma_y} = b_1 \frac{\sigma_x}{\sigma_y} = -0,357.$$

Поскольку $|r| > 0,35$, то связь можно считать средней.

Вычислим коэффициент детерминации. В случае линейной регрессии имеем:

$$R^2 = r_{xy}^2 = 0,127.$$

Это значит, что изменчивость эндогенного фактора Y лишь на 12,7 % можно объяснить изменчивостью экзогенного фактора X .

Построим эконометрическую модель с помощью показательной функции. В этом случае теоретическое уравнение регрессии имеет вид:

$$\tilde{y}_x = b_0 \cdot b_1^x.$$

Для оценки параметров уравнения регрессии проведем линеаризацию исходного уравнения путем его логарифмирования:

$$\ln \tilde{y} = \ln(b_0 \cdot b_1^x); \ln \tilde{y} = \ln b_0 + x \cdot \ln b_1.$$

Если ввести вспомогательную переменную $\tilde{Y} = \ln \tilde{y}$ и обозначить $B_0 = \ln b_0$, $B_1 = \ln b_1$, то получим линейное уравнение:

$$\tilde{Y} = B_0 + B_1 \cdot x.$$

Следовательно, осуществлено преобразование показательной модели в линейную, параметры которой вычисляются с помощью МНК, то есть

$$B_1 = \frac{\mu_{xY}}{\sigma_x^2} = \frac{\overline{xY} - \bar{x} \cdot \bar{Y}}{\overline{x^2} - \bar{x}^2}; \quad B_0 = \bar{Y} - B_1 \cdot \bar{x}.$$

Выполненные вычисления приведены в табл. 2.8.

Таблица 2.8

Вспомогательные вычисления оценки параметров показательной модели "Цена товара – расходы на оплату труда"

№ п/п	x	y	$Y = \ln y$	xY	x^2	Y^2
1	45,1	68,8	4,23	190,83	2 034,01	17,90
2	59,0	61,2	4,11	242,73	3 481,00	16,93
3	57,2	59,9	4,09	234,10	3 271,84	16,75
4	61,8	56,7	4,04	249,53	3 819,24	16,30
5	58,8	55,0	4,01	235,63	3 457,44	16,06
6	47,2	54,3	3,99	188,54	2 227,84	15,96
7	55,2	49,3	3,90	215,17	3 047,04	15,19
Σ	384,3	405,2	28,38	1 556,54	21 338,41	115,09

Следовательно:

$$B_1 = \frac{\mu_{xY}}{\sigma_x^2} = \frac{\overline{xY} - \bar{x} \cdot \bar{Y}}{\overline{x^2} - \bar{x}^2} = -0,005; \quad B_0 = \bar{Y} - B_1 \cdot \bar{x} = 4,35.$$

Теоретическое уравнение линейной регрессии имеет вид:

$$\tilde{Y} = 4,35 - 0,005 \cdot x.$$

Вернемся к исходным переменным. Для этого выполним потенцирование полученного теоретического уравнения. Тогда с использованием показательной функции эконометрическая модель примет вид:

$$\tilde{y} = e^{4,35} e^{-0,005x} = 77,5 \cdot 0,995^x.$$

Пусть для аппроксимации используется степенная функция. Тогда теоретическое уравнение регрессии будет иметь вид:

$$\tilde{y} = b_0 \cdot x^{b_1}.$$

Преобразуем его к линейному виду путем логарифмирования:

$$\ln \tilde{y} = \ln(b_0 \cdot x^{b_1}).$$

Введем новые обозначения $\tilde{Y} = \ln \tilde{y}$, $B_0 = \ln b_0$, $B_1 = b_1$, $X = \ln x$, тогда получим:

$$\hat{Y} = B_0 + B_1 \cdot X.$$

При вычислении параметров модели, которую получили в результате линеаризации функции степени, применяем соотношение:

$$B_1 = \frac{\mu_{XY}}{\sigma_X^2} = \frac{\overline{XY} - \bar{X} \cdot \bar{Y}}{\overline{X^2} - \bar{X}^2}, \quad B_0 = \bar{Y} - B_1 \cdot \bar{X}.$$

Результаты вычислений, которые необходимо сделать для определения параметров модели, приведены в табл. 2.9.

Таблица 2.9

Вспомогательные вычисления оценки параметров модели степенной модели "Цена товара – расходы на оплату труда"

№ п/п	x	y	$Y = \ln y$	$X = \ln x$	XY	X^2
1	45,1	68,8	4,23	3,81	16,12	14,52
2	59,0	61,2	4,11	4,08	16,77	16,65
3	57,2	59,9	4,09	4,05	16,56	16,40
4	61,8	56,7	4,04	4,12	16,64	16,97
5	58,8	55,0	4,01	4,07	16,32	16,56
6	47,2	54,3	3,99	3,85	15,36	14,82
7	55,2	49,3	3,90	4,01	15,64	16,08
Σ	384,3	405,2	28,37	27,99	113,42	112,01

Следовательно:

$$\bar{Y} = \frac{\sum_{i=1}^n Y_i}{n} = 4,05; \quad \bar{X} = \frac{\sum_{i=1}^n X_i}{n} = 4,0; \quad \mu_{XY} = \frac{\sum_{i=1}^n X_i Y_i}{n} - \bar{X}\bar{Y} = -0,0034;$$

$$\sigma_X^2 = \frac{\sum_{i=1}^n X_i^2}{n} - \bar{X}^2 = 0,0124; \quad B_1 = -0,27; \quad B_0 = 5,13.$$

Тогда теоретическое уравнение регрессии во вспомогательных переменных имеет вид:

$$\tilde{Y} = 5,13 - 0,27 \cdot X.$$

Выполним потенцирование полученного уравнения и запишем его в виде степенной функции:

$$\tilde{y} = e^{5,13} \cdot x^{-0,27} = 169,44 \cdot x^{-0,27}.$$

Пример 2.5. Найти эмпирическую формулу для аналитической записи зависимости между величинами X и Y по данным табл. 2.10.

Таблица 2.10

Исходные данные

x_i	0,10	0,15	0,20	0,25	0,30	0,35	0,40
y_i	102,2	87,2	77,8	70,8	66,1	62,0	58,8

Решение.

Построим график зависимости $y = f(x)$ по исходным данным (рис. 2.2).

По полученной эмпирической кривой предполагаем, что имеет место степенная зависимость $y = ax^b$ ($b < 0$).

Линеаризуем зависимость:

$$Y = \lg y, \quad A = b, \quad B = \lg a, \quad X = \lg x: \quad Y = AX + B.$$

Проведем дополнительные вычисления (табл. 2.11).

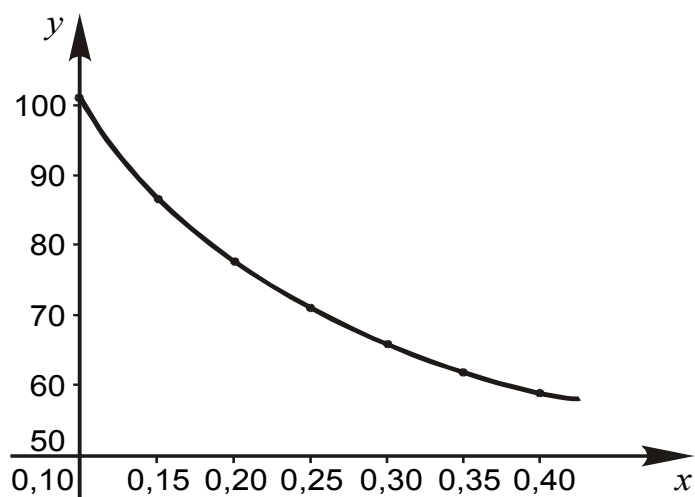


Рис. 2.2. График зависимости $y = f(x)$

Таблица 2.11

Расчетная таблица

								Σ
X	-1,00	-0,82	-0,70	-0,60	-0,523	-0,46	-0,40	-4,50
Y	2,01	1,94	1,89	1,85	1,82	1,79	1,76	13,06
X^2	1,00	0,67	0,49	0,36	0,27	0,21	0,16	3,17
XY	-2,01	-1,59	-1,32	-1,11	-0,95	-0,82	-0,70	-8,51

Согласно МНК найдем для линеаризованной функции $Y = AX + B$ параметры модели:

$$A = \frac{\overline{XY} - \bar{X}\bar{Y}}{\overline{X^2} - \bar{X}^2} = \frac{\frac{-8,51}{7} - \left(\frac{-4,5}{7}\right) \cdot \frac{13,06}{7}}{\frac{3,17}{7} - \left(\frac{-4,5}{7}\right)^2} = -0,41;$$

$$B = \bar{Y} - A \cdot \bar{X} = \frac{13,06}{7} - (-0,41) \cdot \left(\frac{-4,5}{7}\right) = 1,6.$$

Однако $A = b$, $B = \lg a$, тогда $b = -0,41$, $a = 10^B = 10^{1,6} = 39,81$.

Следовательно, нелинейное уравнение регрессии будет иметь вид:

$$y = 39,81x^{-0,4}.$$

2.3. Задачи

Задача 2.1. Из нижеперечисленных функций, определить, какие являются линейными по переменным, линейными по параметрам, нелинейными ни по переменным, ни по параметрам.

- 1) $Y_t = \alpha_0 + \alpha_1 x^3 + \varepsilon_t$; 2) $Y_t = \alpha_0 + \alpha_1 \ln(x) + \varepsilon_t$;
3) $\ln(Y_t) = \alpha_0 + \alpha_1 \ln(x) + \varepsilon_t$; 4) $Y_t = \alpha_0 + \alpha_1 x^c + \varepsilon_t$;
5) $Y^a = \alpha_0 + \alpha_1 x^2 + \varepsilon_t$; 6) $Y_t = 1 + a_0(1 - x^b) + \varepsilon_t$;
7) $Y_t = \alpha_0 + \alpha_1 \frac{x}{10} + \varepsilon_t$.

Задача 2.2. Заданы две регрессии, параметры которых вычислены по данным 25-ти квартальных наблюдений:

- а) $y_t = -30 + 0,18x_t$ (y_t – затраты на оплату жилья, x_t – прибыль);
б) $y_t = 50 + 4,5t$ (y_t – затраты на оплату жилья, t – время).

Дать экономическую интерпретацию параметров построенных уравнений регрессии.

Задача 2.3. В табл. 2.12 представлены значения цены товара (X , грн) и объемов предложения (Y , тыс. грн). Построить линейную модель регрессии и ее графическое изображение, вычислить прогнозное значение, если цена будет составлять 7,2.

Таблица 2.12

Исходные данные

X	5,4	7,6	2,3	5,9	11,0	12,6	10,4	4,9	2,4	1,6
Y	13,7	18,0	6,2	15,5	24,1	24,8	25,0	13,0	18,1	6,7

Задача 2.4. Данные, приведенные в табл. 2.13, характеризуют расходы на рекламу продукции (млн грн) – X , и объем ее реализации (млн грн) – Y . Определить коэффициенты уравнения парной линейной регрессии, начертить график линии регрессии; определить прогнозный объем реализации продукции, если расходы на рекламу составят 12 млн грн.

Таблица 2.13

Исходные данные

X	2	3	5	7	10
Y	14	16	20	25	32

Задача 2.5. В табл. 2.14 представлены значения цены на товар (грн) и объем спроса (тыс. грн). Построить линейную, степенную и показательную модели, вычислить коэффициент корреляции, сделать выводы.

Таблица 2.14

Исходные данные

Цена (X)	Спрос (Y)
1,25	11,18
1,38	10,68
1,85	8,36
2,01	7,25
2,16	7,46
1,72	9,26
2,34	6,35
1,92	7,73
2,26	6,34
2,04	6,58
2,03	7,58
1,72	9,56
2,02	7,48
2,27	6,93
1,94	8,83

Задача 2.6. В табл. 2.15 Y – месячный объем спроса на товары первой необходимости семьи из трех человек (усл. ден. ед.), X – месячный уровень дохода семьи (усл. ден. ед.).

Проверить, какая из моделей лучше приближает эмпирические данные: 1) линейная, 2) степенная; 3) гиперболическая.

Таблица 2.15

Исходные данные

X	2,5	1,4	0,9	2,7	1,8	2,2	2,4	1,9	1,6	1,2
Y	0,8	1,1	0,7	0,9	1,0	1,2	0,9	0,6	0,7	0,5

Задача 2.7. Известны данные о среднемесечном уровне занятости (X , %) и уровне инфляции (Y , %) (табл. 2.16).

Таблица 2.16

Исходные данные

X	32	35	36	34	38	36	37	40
Y	5,4	6,1	6,2	5,8	6,3	6,0	5,9	6,3

Построить гиперболическую модель, определить коэффициент детерминации, сделать выводы.

Задача 2.8. Изучается динамика потребления мяса в регионе. Для этого были собраны данные об объемах среднедушевого потребления мяса (кг) за год. Предварительная обработка данных путем логарифмирования привела к получению результатов, представленных в табл. 2.17.

Таблица 2.17

Исходные данные

Месяц	1	2	3	4	5	6	7	8	9	10	11	12
$\ln(y_t)$	1,99	2,10	2,12	2,17	2,22	2,20	2,25	2,28	2,30	2,31	2,29	2,31

Построить уравнение экспоненциальной модели вида $Y = a_0 \cdot e^{a_1 t}$.
Дать интерпретацию его параметров.

Задача 2.9. На основе данных, приведенных в табл. 2.18, определить коэффициенты корреляции между независимыми переменными X_1 , X_2 , X_3 и зависимой переменной Y .

Сделать выводы.

Таблица 2.18

Исходные данные

Y	X_1	X_2	X_3
30	4	10	16
35	2	13	18
40	3	14	14
44	8	15	10
45	7	18	12
42	10	20	9
48	15	23	12
50	14	25	8
55	10	27	5

Задача 2.10. Изучается зависимость объема продаж товара (y) от индекса потребительских цен (x). Полученные за 12 кварталов данные представлены в табл. 2.19.

Таблица 2.19

Исходные данные

Квартал	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
Индекс потребительских цен, в % к I кварталу	100	102	104	110	112	112	114	117	120	124	124	130
Средний за день объем продаж товара в течение квартала, кг	20	24	22	30	35	37	45	47	46	50	54	60

Известно также, что

$$\sum_{t=1}^{12} x_t = 1\,360; \quad \sum_{t=1}^{12} y_t = 470; \quad \sum_{t=1}^{12} x_t y_t = 54\,945; \quad \sum_{t=1}^{12} x_t^2 = 20\,320.$$

Построить модель зависимости объема продаж товара от индекса потребительских цен и дать интерпретацию параметров полученной модели.

Задача 2.11. Проведено исследование 15 объектов и получены такие результаты:

$$\sum x_i = 23,19; \quad \sum y_i = 149,95; \quad \sum x_i y_i = 227,238;$$
$$\sum x_i^2 = 36,829; \quad \sum y_i^2 = 1523,986.$$

Оценить коэффициенты уравнений линейной регрессии \hat{y}_x и коэффициент корреляции, сделать выводы.

Задача 2.12. Проведено исследование зависимости себестоимости продукции (y) от затрат на автоматизацию труда (x). Составить уравнение прямой регрессии \hat{y}_x , если:

$$\sum x_i = 54,40; \quad \sum x_i^2 = 165,64; \quad \sum y_i = 789,0; \quad n = 20;$$
$$\sum y_i^2 = 33\,803,36; \quad \sum x_i y_i = 2\,322,26.$$

Задача 2.13. Составить уравнение зависимости y (стоимость основных фондов предприятия) и x (объем валовой продукции), если $r_{yx} = 0,68$, $\bar{y} = 250$, $\sigma_y = 38$, $\bar{x} = 47$, $\sigma_x = 12$; $n = 16$.

Пояснить результаты.

3. Проверка качества уравнения парной регрессии

3.1. Основные теоретические сведения

Для проверки качества построенной регрессионной модели используется *F-критерий Фишера*, которому предшествует дисперсионный анализ.

Из курса изучения математической статистики известно разложение общей суммы квадратов отклонений переменной y от среднего значения \bar{y} на две части – "объясненную" и "необъясненную":

$$\begin{array}{l} \sum(y - \bar{y})^2 = \sum(\hat{y}_x - \bar{y})^2 + \sum(y - \hat{y}_x)^2 \\ \text{общая сумма} \quad \text{сумма квадратов} \quad \text{остаточная сумма} \\ \text{квадратов} = \text{отклонений,} \quad + \quad \text{квадратов} \\ \text{отклонений} \quad \text{объясненная} \quad \text{отклонений} \\ \quad \quad \text{регрессией} \quad \quad \text{необъясненная} \\ \quad \quad \quad \quad \quad \quad \text{регрессией} \end{array}$$

или

$$SS_y = SS_{\hat{y}_x} + SS_{\varepsilon}.$$

Все выкладки дисперсионного анализа оформляются в виде стандартной таблицы (m – число оцениваемых параметров уравнения регрессии при независимых переменных, n – число наблюдений) (табл. 3.1).

Таблица 3.1

Таблица дисперсионного анализа

Компоненты дисперсии	Число степеней свободы	Дисперсия на одну степень свободы	Дисперсионное отношение	Уровень значимости
Факторная	m	$S^2_{\hat{y}_x} = \frac{\sum(\hat{y}_x - \bar{y})^2}{m}$	$F = \frac{S^2_{\hat{y}_x}}{S^2_{\varepsilon}}$	α
Остаточная	$n - m - 1$	$S^2_{\varepsilon} = \frac{\sum(y - \hat{y}_x)^2}{n - m - 1}$		
Общая	$n - 1$	$S^2_y = \frac{\sum(y - \bar{y})^2}{n - 1}$		

При $F > F_{\alpha=0,01}(n-1, n-m-1)$ результативная переменная Y и показатель X не являются независимыми, они связаны между собой, между ними имеется статистическая (корреляционная) зависимость.

Для оценки значимости уравнения регрессии в целом с помощью F -критерия Фишера выдвигается нулевая гипотеза (H_0) о том, что коэффициент регрессии $\hat{y}_x = a + b \cdot x$ равен нулю, то есть $H_0 : b = 0$.

Значение статистики F -критерия Фишера для парной линейной регрессии рассчитывается по формуле:

$$F = \frac{r^2}{1 - r^2} \cdot (n - 2),$$

где $r = \sqrt{1 - \frac{S_\varepsilon^2}{S_y^2}} = \sqrt{\frac{S_{\hat{y}}^2}{S_y^2}}$, и сравнивается с табличным $F_\alpha(1, n - 2)$

(приложения А, Б), где α – уровень значимости.

Если $F > F_{0,01}(1, n - 2)$, то H_0 отвергается и делается вывод о существенности (значимости) изучаемой связи. Если $F < F_{0,05}(1, n - 2)$, то H_0 не может быть отклонена без серьезного риска сделать неправильный вывод о наличии связи и уравнение регрессии считается статистически незначимым.

В линейной регрессии $\hat{y}_x = a + b \cdot x$ необходимо оценить значимость не только уравнения в целом, но и отдельных его параметров, поэтому определяются стандартные ошибки m_b и m_a :

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

$$m_a = \sqrt{\frac{\sum (y - \hat{y}_x)^2}{n - 2} \cdot \frac{\sum x^2}{n \sum (x - \bar{x})^2}} = \sqrt{S_\varepsilon^2 \cdot \frac{\sum x^2}{n \sum (x - \bar{x})^2}}.$$

Вычисляются значения статистик t -критерия Стьюдента:

$$t_b = \frac{b}{m_b}; \quad t_a = \frac{a}{m_a},$$

которые сравниваются с табличными значениями $t_{\alpha, n-2}$ (приложение В) при числе степеней свободы $n - 2$.

Если $t_b > t_{\alpha=0,05; n-2}$ ($t_a > t_{\alpha=0,05; n-2}$), то гипотезу о несущественности коэффициента регрессии (параметра a) можно отклонить.

Значимость линейного коэффициента корреляции проверяется с помощью величины стандартной ошибки коэффициента корреляции m_r :

$$m_r = \sqrt{\frac{1-r^2}{n-2}}, \quad t_r = \frac{r}{\sqrt{1-r^2}} \cdot \sqrt{n-2}.$$

Если $t_r > t_{\alpha=0,05; n-2}$, то гипотезу о несущественности коэффициента корреляции можно отклонить.

Заметим, что $t_r^2 = F$ и $t_b^2 = F$, а значит $t_r^2 = t_b^2$. То есть проверка гипотез о значимости коэффициентов регрессии и корреляции равносильна проверке гипотезы о существенности линейного уравнения регрессии.

Зная значения стандартных ошибок m_b и m_a , можно вычислить доверительные интервалы для параметров уравнения регрессии:

$$b \pm t_{\alpha, n-2} m_b; \quad a \pm t_{\alpha, n-2} m_a.$$

Прогноз результативного показателя будет реальным, если вычислить доверительный интервал для предсказываемого среднего значения переменной y при заданном значении переменной $x = x_0$:

$$\hat{y}_{x=x_0} \pm t_{\alpha, n-2} \cdot m_{\hat{y}_x},$$

где $\hat{y}_{x=x_0} = a + b \cdot x_0$;

$m_{\hat{y}_x}$ – стандартная ошибка предсказываемого среднего значения y

при заданном значении x_0 , $m_{\hat{y}_x} = S \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum (x_i - \bar{x})^2}}$,

S – стандартная ошибка регрессии, $S = \sqrt{\frac{\sum_{i=1}^n \varepsilon_i^2}{n-2}}$;

$t_{\alpha, n-2}$ – значение теоретической статистики критерия из таблицы критических точек распределения Стьюдента (приложение В).

Вычисление доверительного интервала для прогнозируемого индивидуального значения \hat{y}_x при $x = x_k$ с надежностью 95 % выполняется по формуле:

$$\hat{y}_{x=x_k} \pm t_{\alpha, n-2} \cdot m_{y(x_k)},$$

где $\hat{y}_{x=x_k} = a + b \cdot x_k$;

$$m_{y(x_k)} = S \sqrt{1 + \frac{1}{n} + \frac{(x_k - \bar{x})^2}{\sum (x_i - \bar{x})^2}} - \text{стандартная ошибка прогнози-}$$

руемого индивидуального значения $y(x_k)$.

Чтобы иметь общее суждение о качестве модели определяют среднюю ошибку аппроксимации:

$$A = \frac{1}{n} \sum \left| \frac{(y - \hat{y}_x)}{y} \right| \cdot 100 \%$$

Допустимая граница средней ошибки аппроксимации – не более чем 8 – 10 %.

В анализе моделей влияния фактора на результат полезным является расчет коэффициента эластичности:

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

где $f'(x)$ – первая производная, характеризующая соотношение приростов результата и фактора для соответствующей формы связи.

Поскольку коэффициент эластичности для линейной функции не является величиной постоянной, а зависит от соответствующего

значения x , то рассчитывается средний показатель эластичности по формуле:

$$\bar{\varepsilon} = b \frac{\bar{x}}{\bar{y}}.$$

3.2. Примеры

Пример 3.1. По данным табл. 3.2 оценены параметры разных типов модели ценообразования, которая характеризует зависимость между затратами труда и ценой товара при условии относительного соответствия спроса и предложения.

Таблица 3.2

Исходные данные

Затраты труда x , тыс. грн	65	70	74	72	68	73	75	76
Цена y , тыс. грн	310,5	314,2	317,4	315,1	311,5	316,2	319,7	321,6

Получены следующие уравнения зависимости между затратами труда и ценой товара:

$$\tilde{y}_1 = 245,58 + 0,9801 \cdot x; \quad \tilde{y}_2 = 252,77 \cdot e^{0,0031x}; \quad \tilde{y}_3 = 124,67 \cdot x^{0,2176}.$$

Оценить качество всех типов уравнений зависимости с помощью средней ошибки аппроксимации.

Решение.

Средняя ошибка аппроксимации вычисляется по формуле:

$$\bar{A} = \frac{\sum_{i=1}^n \left| \frac{y_i - \tilde{y}_i}{y_i} \right| \cdot 100\%}{n},$$

где y_i – фактическое значение результативного признака;

\tilde{y}_i – теоретическое значение результативного признака;

n – количество наблюдений.

Выполним вспомогательные расчеты (табл. 3.3 – 3.5).

Таблица 3.3

Вычисления для оценки показателя качества линейной зависимости между затратами труда и ценой товара

№ п/п	x_i	y_i	\tilde{y}_{1i}	$\left \frac{y_i - \tilde{y}_{1i}}{y_i} \right \cdot 100 \%$
1	65	310,5	309,3	0,391
2	70	314,2	314,2	0,004
3	74	317,4	318,1	0,223
4	72	315,1	316,1	0,332
5	68	311,5	312,2	0,233
6	73	316,2	317,1	0,293
7	75	319,7	319,1	0,192
8	76	321,6	320,1	0,476
Σ	–	–	–	2,145

Таблица 3.4

Вычисления для оценки показателя качества экспоненциальной зависимости между затратами труда и ценой товара

№ п/п	x_i	y_i	\tilde{y}_{2i}	$\left \frac{y_i - \tilde{y}_{2i}}{y_i} \right \cdot 100 \%$
1	65	310,5	309,2	0,420
2	70	314,2	314,0	0,055
3	74	317,4	317,9	0,172
4	72	315,1	316,0	0,279
5	68	311,5	312,1	0,188
6	73	316,2	317,0	0,241
7	75	319,7	318,9	0,240
8	76	321,6	319,9	0,521
Σ	–	–	–	2,116

Вычисления для оценки показателя качества степенной зависимости между затратами труда и ценой товара

№ п/п	x_i	y_i	\tilde{y}_{3i}	$\left \frac{y_i - \tilde{y}_{3i}}{y_i} \right \cdot 100 \%$
1	65	310,5	309,2	0,416
2	70	314,2	314,2	0,011
3	74	317,4	318,1	0,207
4	72	315,1	316,2	0,339
5	68	311,5	312,3	0,244
6	73	316,2	317,1	0,290
7	75	319,7	319,0	0,222
8	76	321,6	319,9	0,526
Σ	–	–	–	2,256

Получено: $\overline{A_1} = \frac{2,145}{8} = 0,27 \%$, то есть в среднем значения, кото-

рые вычислены по уравнению линейной зависимости, отклоняются от фактических на 0,27 % ,

$\overline{A_2} = \frac{2,116}{8} = 0,26 \%$ (по экспоненте отклонения от фактических значений – 0,26 %),

$\overline{A_3} = \frac{2,256}{8} = 0,28 \%$ (по степенной функции отклонения от фактических значений – 0,28 %).

Таким образом, по величине средней ошибки аппроксимации среди представленных уравнений моделей более качественной является регрессия экспоненциального типа.

Осуществление окончательного выбора типа зависимости позволит в последующем использовать ее для краткосрочного прогноза цены товара.

Пример 3.2. Зависимость потребления молочных продуктов (y) от дохода по данным 20 семей (x) характеризуется уравнением регрессии $y = 2x^{0,3}$, где $r = 0,9$, остаточная дисперсия $S_{\varepsilon}^2 = 0,06$.

Провести дисперсионный анализ полученных результатов.

Решение.

Проверим значимо ли статистически заданное уравнение регрессии и надежен ли показатель тесноты связи.

Используем формулу $r = \sqrt{1 - \frac{S_{\varepsilon}^2}{S_y^2}} = \sqrt{\frac{S_{\hat{y}}^2}{S_y^2}}$.

Результаты дисперсионного анализа представлены в табл. 3.6.

Сравниваем полученное значение F с табличным $F_{\alpha}(1, n-2) = F_{0,05}(1, 18) = 4,41$ (приложение А).

Так как $F = 4,33 < F_{0,05}(1, 18) = 4,41$, то заданное уравнение регрессии считаем статистически незначимым и показатель тесноты связи не надежным.

Таблица 3.6

Результаты дисперсионного анализа

Компоненты дисперсии	Число степеней свободы	Дисперсия на одну степень свободы	Дисперсионное отношение	Уровень значимости
Факторная	$m = 1$	$S_{\hat{y}_x}^2 = R^2 \cdot S_y^2 =$ $= 0,81 \cdot 0,32 = 0,26$	$F = \frac{S_{\hat{y}_x}^2}{S_{\varepsilon}^2} =$ $= 4,33$	$\alpha = 0,05$
Остаточная	$n - m - 1 =$ $= 18$	$S_{\varepsilon}^2 = 0,06$		
Общая	$n - 1 = 19$	$S_y^2 = \frac{S_{\varepsilon}^2}{1 - R^2} =$ $= \frac{0,06}{1 - 0,81} = 0,32$		

Пример 3.3. Проверить гипотезу о наличии линейной связи между переменными: себестоимостью единицы изделия (y , тыс. грн) и величиной выпуска продукции (x , тыс. шт.) в генеральной совокупности, если $r = -0,904$, $n = 5$. Доверительная вероятность $p = 95\%$.

Решение.

Выдвигаем гипотезу $H_0: r = 0$, то есть между переменными отсутствует линейная связь. Тогда альтернативная гипотеза $H_1: r \neq 0$, то есть между переменными есть линейная связь.

Вычисляем статистику критерия:

$$t_r = \sqrt{\frac{r^2(n-2)}{1-r^2}} = \sqrt{\frac{(-0,904)^2(5-2)}{1-(-0,904)^2}} = 3,66.$$

$\alpha = 0,05$, поэтому $t_{\alpha=0,05;3} = 3,1825$; $t > t_{\alpha=0,05;3}$; $3,66 > 3,1825$.

Следовательно отклоняем гипотезу H_0 и принимаем гипотезу H_1 на уровне значимости 5%. То есть между себестоимостью единицы изделия (y , тыс. грн) и величиной выпуска продукции (x , тыс. шт.) есть линейная связь.

Пример 3.4. Найти интервальную оценку коэффициента регрессии $y = 2,12 - 0,11 \cdot x$, которая описывает зависимость себестоимости единицы изделия (y , тыс. грн) от величины выпуска продукции (x , тыс. шт.) по данным табл. 3.7 при $p = 95\%$.

Таблица 3.7

Исходные данные

Номер наблюдения	x	y	$(x - \bar{x})^2$	$(y - \hat{y})^2$
1	2	1,9	4	0,00
2	3	1,7	1	0,0081
3	4	1,8	0	0,0144
4	5	1,6	1	0,0009
5	6	1,4	4	0,0036
Сумма	20	8,4	10	0,027

Решение.

Доверительный интервал для коэффициента регрессии:

$$b \pm t_{\alpha, n-2} m_b,$$

где $m_b = \sqrt{\frac{\sum (y - \hat{y}_x)^2 / (n - 2)}{\sum (x - \bar{x})^2}} = \sqrt{\frac{S_\varepsilon^2}{\sum (x - \bar{x})^2}}.$

По данным табл. 3.7 имеем:

$$m_b = \sqrt{\frac{\sum (y - \hat{y}_x)^2 / (n - 2)}{\sum (x - \bar{x})^2}} = \sqrt{\frac{S_\varepsilon^2}{\sum (x - \bar{x})^2}} = \sqrt{\frac{0,027 / 3}{10}} = 0,03.$$

$\alpha = 0,05$, поэтому $t_{\alpha=0,05;3} = 3,182$.

Тогда $b \pm t_{\alpha, n-2} m_b = -0,11 \pm 3,182 \cdot 0,03 = -0,11 \pm 0,1$, то есть

$$-0,21 < b < -0,01.$$

Пример 3.5. По данным примера 3.4 найти доверительный интервал для предсказываемого среднего значения переменной y и индивидуального значения переменной y при $x_0 = 5,5$ тыс. шт. Доверительная вероятность $p = 95\%$.

Решение.

Для вычисления границ доверительного интервала выполним вспомогательные расчеты (табл. 3.8).

Таблица 3.8

Вспомогательные расчеты

№ п/п	x	x^2	$(y - \hat{y})^2$
1	2	4	0,00
2	3	9	0,0081
3	4	16	0,0144
4	5	25	0,0009
5	6	36	0,0036
Сумма	20	90	0,027

Доверительный интервал для предсказываемого среднего значения переменной y при заданном значении переменной $x = x_0$:

$$\hat{y}_{x=x_0} \pm t_{\alpha, n-2} \cdot m_{\hat{y}_x},$$

где $m_{\hat{y}_x} = S \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum (x_i - \bar{x})^2}}$; $S = \sqrt{\frac{\sum_{i=1}^n \varepsilon_i^2}{n-2}}$.

Используем формулу доверительного интервала:

$$\hat{y}_{x=x_0} \pm t_{\alpha, n-2} \cdot m_{\hat{y}_x} = a + bx_0 \pm t_{\alpha, n-2} \cdot S \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^n x_i^2 - \frac{\left(\sum_{i=1}^n x_i\right)^2}{n}}}.$$

По данным табл. 3.8 получаем:

$$\sum_{i=1}^n x_i^2 - \frac{\left(\sum_{i=1}^n x_i\right)^2}{n} = 90 - \frac{400}{5} = 10; S = \sqrt{\frac{0,027}{3}} = 0,0949;$$

$$(x_0 - \bar{x})^2 = (5,5 - 4)^2 = 2,25.$$

$$\alpha = 0,05, \text{ поэтому } t_{\alpha=0,05;3} = 3,182.$$

Тогда:

$$\begin{aligned} \hat{y}_{x=x_0} \pm t_{\alpha, n-2} \cdot m_{\hat{y}_x} &= 2,12 + (-0,11) \cdot 5,5 \pm 3,182 \cdot 0,0949 \sqrt{\frac{1}{5} + \frac{2,25}{10}} = \\ &= 1,515 \pm 0,197. \end{aligned}$$

То есть доверительный интервал для предсказываемого среднего значения переменной y при заданном значении $x_0 = 5,5$ тыс. шт.:

$$1,318 < y < 1,712, \text{ то есть } (1,318; 1,712).$$

Найдем доверительный интервал для предсказываемого индивидуального значения переменной y при $x_0 = 5,5$ тыс. шт.:

$$a + bx_0 \pm t_{\alpha, n-2} S \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^n x_i^2 - \frac{\left(\sum_{i=1}^n x_i\right)^2}{n}}},$$

$$2,12 + (-0,11)5,5 \pm 3,182 \cdot 0,0949 \sqrt{1 + \frac{1}{5} + \frac{2,25}{10}} = 1,515 \pm 0,3605.$$

То есть доверительный интервал для предсказываемого индивидуального значения переменной y при $x_0 = 5,5$ тыс. шт. :

$$1,1545 < y < 1,8755, \text{ то есть } (1,1545; 1,8755).$$

Доверительный интервал для предсказываемого индивидуального значения переменной y шире доверительного интервала для предсказываемого среднего значения переменной y при заданном значении x_0 .

3.3. Задачи

Задача 3.1. По данным табл. 3.9 проанализировать модели зависимости себестоимости продукции от затрат труда на единицу продукции.

Таблица 3.9

Исходные данные

Уравнение зависимости	Среднее значение независимой переменной
$y = 9,3 + 9,83 \cdot x$	$\bar{x} = 1,38$
$y = 14,87 \cdot 1,016^x$	$\bar{x} = 26,3$
$y = 189,7 \cdot x^{-0,298}$	$\bar{x} = 1,74$

Сделать вывод о коэффициенте регрессии каждого уравнения. С помощью среднего коэффициента эластичности определить силу влияния фактора x .

Задача 3.2. Для трех видов продукции A, B, C уравнение модели зависимости затрат (y) от объема производства продукции (x) имеют вид:

$$y_A = 30 + 0,8x; \quad y_B = 200; \quad y_C = 60x^{0,3}.$$

Определить коэффициенты эластичности для каждого вида продукции, объяснить их значение;

сравнить при $x = 400$ эластичность затрат для продукции A и C ;

определить объем производства продукции, при котором коэффициенты эластичности для продукции A и C будут равны.

Задача 3.3. Зависимость производительности труда (y) от продолжительности рабочего дня (x), часов по 20 заводам характеризуется следующим образом: $y = 200 + 0,2x - 0,06x^2$, $r^2 = 0,75$.

Проверить гипотезу о значимости уравнения регрессии при уровне значимости 0,1, используя критерий Фишера.

Определить средний коэффициент эластичности, предполагая, что средняя продолжительность рабочего дня равна 7 час.

Задача 3.4. Модель $y = 62,9 - 1,4x$ характеризует зависимость объема предоставленных кредитов (млн грн) от уровня процентной ставки. Уравнение построено на основе эмпирических данных, приведенных в табл. 3.10.

Таблица 3.10

Исходные данные

x_i	18	20	24	25	27
y_i	40	34	32	28	27

Провести эконометрический анализ модели:

проверить статистическую значимость коэффициента корреляции r и параметров модели при уровне значимости $\alpha = 0,05$;

построить доверительные интервалы для теоретических параметров модели при уровне надежности $\gamma = 0,95$;

вычислить коэффициент детерминации, объяснить, что он значит.

Построить точечный и интервальный прогноз для зависимой переменной при уровне надежности $\gamma = 0,95$ (точку прогноза задать самостоятельно).

Задача 3.5. По данным за 9 лет построена такая эмпирическая регрессия:

$$y_t = -70,85 + 0,888x_t \quad R^2 = 0,685,$$
$$t = (-5,89) \quad (5,9)$$

где y – индекс цен оптовой торговли;

x – процент использования производственных мощностей,

t – фактические значения статистик t -критерия Стьюдента.

Как трактуется угловой коэффициент данного уравнения регрессии? Оценить значимость параметров регрессии и качество модели в целом.

Задача 3.6. Получены данные прибыли X и Y (%) двух компаний за 10 лет (табл. 3.11).

Таблица 3.11

Исходные данные

X	19,2	15,8	12,5	10,3	5,7	-5,8	-3,5	5,2	7,3	6,7
Y	20,1	18,0	10,3	12,5	6,0	-6,8	-2,8	3,0	8,5	8,0

Построить регрессионную модель $Y = b_0 + b_1X + e$, оценить статистическую значимость коэффициентов регрессии и коэффициента корреляции.

Возможно ли на основе построенного уравнения регрессии утверждать, что прибыль одной компании является следствием прибыли другой?

Задача 3.7. На основании 32 данных о зависимости ожидаемой продолжительности жизни (y) от суточной калорийности питания населения в различных странах (x) построена следующая регрессионная модель: $y = 44,4 + 0,009 \cdot x$.

Известно также, что коэффициент $r = 0,5562$.

Построить доверительный интервал для коэффициента регрессии с вероятностью 90 и 95 %. Проанализировать результаты для различных вероятностей, пояснить причины различий результатов.

Задача 3.8. По случайной выборке из 25 предприятий торговли изучается зависимость между признаками: x – средняя величина покупки в торговом центре, тыс грн; y – выручка торгового предприятия, млн грн. При оценке регрессионной модели были получены следующие промежуточные результаты:

$$\sum_{i=1}^{25} (y_j - \bar{y}_j)^2 = 1300; \quad \sum_{i=1}^{25} (y_j - \bar{y}_j)^2 = 2000.$$

Какой показатель можно определить по этим данным? Построить таблицу дисперсионного анализа для расчета значений статистики F -критерия Фишера. Является ли модель существенной? Если да, то с какой вероятностью?

Задача 3.9. По 20 фермерским хозяйствам области получена информация, представленная в табл. 3.12.

Фактическое значение статистики F -критерия Фишера составило 45. Определить линейный коэффициент детерминации. Сделать вывод. Восстановить уравнение линейной регрессии $\hat{y}_x = a + b \cdot x$.

С вероятностью 90 % определить доверительный интервал ожидаемого значения урожайности в предположении роста количества внесенных удобрений на 15 % от своего среднего уровня.

Таблица 3.12

Исходные данные

Показатель	Среднее значение	Дисперсия
Урожайность, ц/га (y)	27	20
Внесено удобрений на 1 га посева, кг (x)	5	15

Задача 3.10. Для двух видов продукции A и B зависимость расходов предприятия Y (тыс. грн) от объема производства x (шт.) характеризуется данными, представленными в табл. 3.13.

Таблица 3.13

Исходные данные

Уравнение регрессии	Показатель корреляции	Число наблюдений
$Y_A = 130 + 0,85x$	0,76	40
$Y_B = 30x^{0,8}$	0,82	50

Пояснить смысл величин 0,85 и 0,8 в уравнениях регрессии.

Оценить значимость каждого уравнения регрессии с помощью критерия Фишера, приняв уровень значимости, равным 0,01.

Задача 3.11. По данным табл. 3.14 провести точечную и интервальную оценку коэффициентов линейной регрессии с вероятностью 90 %, если x – независимая переменная, Y – зависимая.

Таблица 3.14

Исходные данные

x	1,2	1,7	2,1	3,4	3,2	4,1	5,5	4,9	4,3	4,0
Y	31	30,8	29,2	28,3	27,1	26,3	26,1	28,2	28	27

Является ли коэффициент регрессии значимым? Если да, то с какой вероятностью?

Задача 3.12. Зависимость объема продаж товара – Y (тыс. \$) от расходов на рекламу x (тыс. \$) характеризуется по 12 филиалам торгового предприятия данными табл. 3.15.

Исходные данные

Уравнение регрессии	$Y = 12,42 + 0,729x$
Среднее квадратическое отклонение x	0,5629
Среднее квадратическое отклонение Y	0,0399

Определить коэффициент корреляции между Y и x .

Вычислить стандартную ошибку оценки коэффициента регрессии; оценить значимость коэффициента регрессии.

Построить таблицу дисперсионного анализа для оценки существенности уравнения регрессии (принять уровень значимости равным 0,05).

4. Линейные модели множественной регрессии**4.1. Основные теоретические сведения**

Основная цель множественной регрессии – построить модель с большим числом факторов, определив при этом влияние каждого из них в отдельности, а также совокупное их воздействие на результивный показатель. Построение модели начинается с определения ее спецификации: отбора факторов и выбора вида уравнения регрессии.

Уравнение линейной множественной регрессии ("чистой" регрессии) имеет вид:

$$\hat{y} = a + b_1x_1 + b_2x_2 + \dots + b_mx_m + \varepsilon.$$

Коэффициенты "чистой" регрессии b_i характеризуют среднее изменение результата с изменением соответствующего фактора на единицу при неизменной значении других факторов, закрепленных на среднем уровне. Постоянная a выполняет в уравнении регрессии функцию выравнивания и определяет точку пересечения гиперповерхности регрессии с осью ординат. Эта величина позволяет получить при средних значениях объясняющих переменных среднее значение результивного признака.

3. Отсутствие автокорреляции. Случайные отклонения ε_i и ε_j являются независимыми друг от друга для всех $i \neq j$:

$$\sigma_{\varepsilon_i \varepsilon_j} = \text{cov}(\varepsilon_i, \varepsilon_j) = \begin{cases} 0, & \text{если } i \neq j \\ \sigma^2, & \text{если } i = j \end{cases}$$

С учетом выполнимости предпосылки 1 данное соотношение можно записать в виде $M(\varepsilon_i \varepsilon_j) = 0 (i \neq j)$.

4. Случайное отклонение должно быть независимо от объясняющих переменных: $\sigma_{\varepsilon_i x_j} = 0$.

5. Модель является линейной относительно параметров.

Для случая множественной линейной регрессии существенными являются еще две предпосылки.

6. Отсутствие мультиколлинеарности, то есть между объясняющими переменными отсутствует строгая (сильная) линейная зависимость.

7. Ошибки $\varepsilon_i, i = \overline{1, n}$ имеют нормальное распределение ($\varepsilon_i \sim N(0, \sigma)$).
Выполнимость данной предпосылки важна для проверки статистических гипотез и построение интервальных оценок.

При выполнении предпосылок МНК относительно ошибок ε_i оценки $a = b_0, b_1, \dots, b_m$ параметров множественной линейной регрессии по МНК являются несмещенными, эффективными и состоятельными.

В матричной форме уравнение линейной множественной регрессии можно записать в виде:

$$Y = XB + E,$$

где $Y^T = (y_1, y_2, \dots, y_n)$;

$$X = (X_0 \ X_1 \ X_2 \ \dots \ X_m) = \begin{pmatrix} 1 & x_{11} & x_{21} & \dots & x_{m1} \\ 1 & x_{12} & x_{22} & \dots & x_{m2} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_{1n} & x_{2n} & \dots & x_{mn} \end{pmatrix};$$

$$X_0 = (1, 1, \dots, 1);$$

$$B = (b_0, b_1, b_2, \dots, b_m)^T;$$

$$E = (e_1, e_2, \dots, e_n)^T.$$

Откуда находим вектор коэффициентов регрессии в виде:

$$B = (X^T X)^{-1} X^T Y = CX^T Y;$$

$$B = CX^T Y,$$

где X^T – матрица, транспонированная к матрице X ;

$C = (X^T X)^{-1}$ – матрица, обратная к матрице системы нормальных уравнений $X^T X$, то есть $C(X^T X) = I$ (единичная матрица).

Матрица X кроме векторов независимых переменных содержит вектор единиц. Он дописывается в этой матрице слева тогда, когда эконометрическая модель имеет свободный член.

Для характеристики параметров b_1, b_2, \dots, b_m используются их дисперсия и ковариация. Значения этих величин в классической эконометрической модели формируют дисперсионно-ковариационную матрицу:

$$\text{var}(B) = S_\varepsilon^2 (X^T X)^{-1} = \begin{pmatrix} \sigma_{b_1}^2 & \mu_{b_1 b_2} & \dots & \mu_{b_1 b_j} & \dots & \mu_{b_1 b_m} \\ \mu_{b_2 b_1} & \sigma_{b_2}^2 & \dots & \mu_{b_2 b_j} & \dots & \mu_{b_2 b_m} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \mu_{b_j b_1} & \mu_{b_j b_2} & \dots & \sigma_{b_j}^2 & \dots & \mu_{b_j b_m} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \mu_{b_m b_1} & \mu_{b_m b_2} & \dots & \mu_{b_m b_j} & \dots & \sigma_{b_m}^2 \end{pmatrix},$$

где $S_\varepsilon^2 = \frac{E^T E}{n - m} = \frac{Y^T Y - B^T X^T Y}{n - m}$ – несмещенная оценка дисперсии остатков модели.

Оценки дисперсионно-ковариационной матрицы используются для вычисления стандартных ошибок:

$$m_{b_i} = \sqrt{\sigma_{b_i}^2} = \sqrt{\frac{\sum e_i^2}{n-m-1} \cdot c_{ii}},$$

где c_{ii} – диагональный элемент матрицы $C = (X^T X)^{-1}$ и доверительных интервалов оценок параметров b_j .

Они используются и при проверке их статистической значимости.

По аналогии с парной регрессией после определения точечных оценок b_i коэффициентов теоретического уравнения регрессии могут быть рассчитаны интервальные оценки указанных коэффициентов.

Доверительный интервал, накрывающий с надежностью $1-\alpha$ неизвестное значение параметра модели, определяется неравенством:

$$b_i - t_{\alpha, n-m-1} \cdot \sigma_{b_i} < \hat{b}_i < b_i + t_{\alpha, n-m-1} \cdot \sigma_{b_i}.$$

Интервальная оценка для среднего предсказываемого значения результативного показателя модели с надежностью $1-\alpha$ определяется неравенством:

$$\hat{Y}_{X=X_0} - t_{\alpha, n-m-1} \cdot S_p < \bar{Y} < \hat{Y}_{X=X_0} + t_{\alpha, n-m-1} \cdot S_p,$$

или

$$\hat{Y}_0 - t_{\alpha} S_{\varepsilon} \sqrt{X_0^T (X^T X)^{-1} X_0} \leq \bar{Y} \leq \hat{Y}_0 + t_{\alpha} S_{\varepsilon} \sqrt{X_0^T (X^T X)^{-1} X_0},$$

где $S_p = S_{\varepsilon} \sqrt{\frac{1}{n} + \frac{(x_{0j} - \bar{x}_j)^2}{\sum_{i=1}^n x_{ij}^2}} = S_{\varepsilon} \sqrt{X_0^T (X^T X)^{-1} X_0}$ – стандартная ошибка

прогноза среднего значения зависимой переменной.

Интервальная оценка для индивидуального предсказываемого значения зависимой переменной модели с надежностью $1 - \alpha$ определяется неравенством:

$$\hat{Y}_{X=X_0} - t_{\alpha, n-m-1} \cdot S_{p_1} < \bar{Y} < \hat{Y}_{X=X_0} + t_{\alpha, n-m-1} \cdot S_{p_1},$$

или

$$\hat{Y}_0 - t_{\alpha} S_{\varepsilon} \sqrt{1 + X_0^T (X^T X)^{-1} X_0} \leq \bar{Y} \leq \hat{Y}_0 + t_{\alpha} S_{\varepsilon} \sqrt{1 + X_0^T (X^T X)^{-1} X_0},$$

где $S_{p_1} = \sqrt{S_p^2 + S_{\varepsilon}^2} = S_{\varepsilon} \sqrt{1 + \frac{1}{n} + \frac{(x_{0j} - \bar{x}_j)^2}{\sum_{i=1}^n x_{ij}^2}} = S_{\varepsilon} \sqrt{1 + X_0^T (X^T X)^{-1} X_0}$ -

стандартная ошибка прогноза индивидуального значения зависимой переменной.

Для определения рейтинга влияния выбранных в модели факторов формируют линейную регрессионную модель в стандартизованных переменных:

$$t_y = \beta_1 t_1 + \beta_2 t_2 + \dots + \beta_m t_m + \varepsilon,$$

где $t_y = \frac{y - \bar{y}}{s_y}$, $t_{x_i} = \frac{x_i - \bar{x}_i}{s_{x_i}}$, $\varepsilon = \frac{e}{s_y}$.

Коэффициенты "чистой" регрессии связаны с коэффициентами регрессии в стандартизованных переменных:

$$b_i = \beta_i \frac{S_y}{S_{x_i}}, \quad a = \bar{y} - \sum_i b_i \bar{x}_i.$$

Применяя МНК к вычислению параметров уравнения множественной регрессии в стандартизованном масштабе, после преобразований получаем систему нормальных уравнений:

$$\begin{cases} r_{yx_1} = \beta_1 & + \beta_2 r_{x_1 x_2} & + \dots + \beta_m r_{x_1 x_m} \\ r_{yx_2} = \beta_1 r_{x_2 x_1} + \beta_2 & & + \dots + \beta_m r_{x_2 x_m} \\ \dots & \dots & \dots & \dots & \dots \\ r_{yx_m} = \beta_1 r_{x_m x_1} + \beta_2 r_{x_m x_2} & + \dots + \beta_m & & & \end{cases}.$$

Стандартизованные коэффициенты регрессии показывают, на сколько сигм (среднеквадратических отклонений) изменится в среднем результат, если соответствующий фактор x_i изменится на одну сигму при неизменном среднем уровне других факторов.

Так как переменные центрированные и нормированные, то стандартизованные коэффициенты регрессии β_i сравнимы между собой и поэтому можно ранжировать факторы по силе их воздействия на результат.

Вычисляя регрессионные модели для экономического анализа следует знать, что может получиться уравнение с β -коэффициентами по модулю большими единицы. Причем статистические критерии значимости отдельных членов модели подтвердят ее качество; они не предотвратят включение в модель посторонних помех, в результате чего может быть получено уравнение регрессии, лишённое какого-либо смысла.

Тесноту связи рассматриваемого набора факторов с исследуемым признаком или тесноту совместного влияния факторов на результат независимо от формы связи характеризует или оценивает индекс множественной корреляции:

$$R_{yx_1 \dots x_m} = \sqrt{1 - \frac{S_\varepsilon^2}{S_y^2}} = \sqrt{\frac{S_{\hat{y}}^2}{S_y^2}},$$

где $S_y^2 = \frac{\sum (y - \bar{y})^2}{n}$ – общая дисперсия результативного признака (по эмпирическим данным);

$S_{\hat{y}}^2 = \frac{\sum (\hat{y} - \bar{\hat{y}})^2}{n}$ – дисперсия результативного признака (по модели регрессии);

$S_\varepsilon^2 = \frac{\sum (y - \hat{y})^2}{n}$ – остаточная дисперсия.

При линейной зависимости индекс корреляции можно представить в виде:

$$R = \sqrt{\beta_{x_1} \cdot r_{yx_1} + \beta_{x_2} \cdot r_{yx_2} + \dots + \beta_{x_m} \cdot r_{yx_m}} = \sqrt{\sum \beta_{x_i} \cdot r_{yx_i}},$$

или

$$R_{yx_1 \dots x_m} = \sqrt{1 - \frac{\Delta r}{\Delta r_{11}}},$$

где $\Delta r = \begin{vmatrix} 1 & r_{yx_1} & r_{yx_2} & \dots & r_{yx_m} \\ r_{yx_1} & 1 & r_{x_1x_2} & \dots & r_{x_1x_m} \\ r_{yx_2} & r_{x_2x_1} & 1 & \dots & r_{x_2x_m} \\ \dots & \dots & \dots & \dots & \dots \\ r_{yx_m} & r_{x_mx_1} & r_{x_mx_2} & \dots & 1 \end{vmatrix}$; $\Delta r_{11} = \begin{vmatrix} 1 & r_{x_1x_2} & \dots & r_{x_1x_m} \\ r_{x_2x_1} & 1 & \dots & r_{x_2x_m} \\ \dots & \dots & \dots & \dots \\ r_{x_mx_1} & r_{x_mx_2} & \dots & 1 \end{vmatrix}$.

Чтобы не допустить возможного преувеличения тесноты связи, при небольшом числе наблюдений используется скорректированный индекс (коэффициент) линейной множественной корреляции:

$$\bar{R} = \sqrt{R^2 \cdot \frac{n-1}{n-m-1}}.$$

Ранжирование факторов по силе влияния на результат можно провести и с помощью частных коэффициентов корреляции (для линейных связей) и частных индексов детерминации (при нелинейных связях).

Частные коэффициенты (индексы) корреляции характеризуют тесноту связи между результатом и соответствующим фактором при устранении влияния других факторов, включенных в уравнение регрессии.

Частные коэффициенты корреляции высоких порядков можно определить через частные коэффициенты корреляции низких порядков по рекуррентной формуле:

$$r_{yx_i \cdot x_1 x_2 \dots x_m} = \frac{r_{yx_i \cdot x_1 x_2 \dots x_{m-1}} - r_{yx_m \cdot x_1 x_2 \dots x_{m-1}} \cdot r_{x_i x_m \cdot x_1 x_2 \dots x_{m-1}}}{\sqrt{(1 - r_{yx_m \cdot x_1 x_2 \dots x_{m-1}}^2) \cdot (1 - r_{x_i x_m \cdot x_1 x_2 \dots x_{m-1}}^2)}}.$$

При двух факторах и $i = 1$ данная формула примет вид:

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

Соответственно при $i = 2$ и двух факторах частный коэффициент корреляции y с фактором x_2 можно определить по формуле:

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

Для модели с двумя независимыми факторами существует взаимосвязь между β -коэффициентами и парными коэффициентами корреляции:

$$\begin{cases} \beta_{x_1} = \frac{r_{yx_1} - r_{yx_2} \cdot r_{x_1 x_2}}{1 - r_{x_1 x_2}^2}, \\ \beta_{x_2} = \frac{r_{yx_2} - r_{yx_1} \cdot r_{x_1 x_2}}{1 - r_{x_1 x_2}^2}. \end{cases}$$

С помощью коэффициента множественной корреляции R и частных коэффициентов корреляции $R_{x_i x_1 \dots x_m}^2$ можно вычислить стандартную ошибку коэффициента регрессии:

$$m_{b_i} = \frac{\sigma_y \sqrt{1 - R^2}}{\sigma_{x_i} \sqrt{1 - R_{x_i x_1 \dots x_m}^2}} \cdot \frac{1}{\sqrt{n - m - 1}}.$$

4.2. Примеры

Пример 4.1. Оценить параметры эконометрической модели, которая характеризует зависимость между расходами на питание, общими расходами и размером дохода на семью.

Исходные данные приведены в табл. 4.1.

Таблица 4.1

Исходные данные

№ п/п	Расходы на питание (y)	Общие расходы (x_1)	Размер семьи (x_2)
1	22	45	1,5
2	34	75	1,6
3	50	125	1,9
4	67	223	1,8
5	47	92	3,4
6	66	146	3,6
7	81	227	3,4
8	106	358	3,5
9	70	135	5,5
10	95	218	5,4
11	119	331	5,4
12	147	490	5,3
13	93	175	8,5
14	133	305	8,3
15	169	468	8,1
16	197	749	7,3

Решение.

Запишем эконометрическую модель:

$$y = b_0 + b_1x_1 + b_2x_2 + \varepsilon,$$

где y – значения расходов на питание;

x_1 – общие расходы;

x_2 – размер дохода на семью;

ε – остатки модели.

Находим вектор оценок коэффициентов регрессии в виде:

$$B = (X^T X)^{-1} X^T Y,$$

где $B = \begin{pmatrix} b_0 \\ b_1 \\ b_2 \end{pmatrix};$

$$X = \begin{pmatrix} 1 & 45 & 1,5 \\ 1 & 75 & 1,6 \\ 1 & 125 & 1,9 \\ 1 & 223 & 1,8 \\ 1 & 92 & 3,4 \\ 1 & 146 & 3,6 \\ 1 & 227 & 3,4 \\ 1 & 358 & 3,5 \\ 1 & 135 & 5,5 \\ 1 & 218 & 5,4 \\ 1 & 331 & 5,4 \\ 1 & 490 & 5,3 \\ 1 & 175 & 8,5 \\ 1 & 305 & 8,3 \\ 1 & 468 & 8,1 \\ 1 & 749 & 7,3 \end{pmatrix}; \quad Y = \begin{pmatrix} 22 \\ 34 \\ 50 \\ 67 \\ 47 \\ 66 \\ 81 \\ 106 \\ 70 \\ 95 \\ 119 \\ 147 \\ 93 \\ 133 \\ 169 \\ 197 \end{pmatrix};$$

X^T – матрица, транспонированная к матрице X .

Согласно формуле оценивания найдем:

$$(X^T X) = \begin{pmatrix} 16 & 416,2 & 74,5 \\ 416,2 & 1601562 & 23271,2 \\ 74,5 & 23271,2 & 436,69 \end{pmatrix};$$

$$(X^T X)^{-1} = \begin{pmatrix} 0,314 & -0,00017 & -0,0446 \\ -0,00017 & 0,00003 & -0,00012 \\ -0,0446 & -0,00012 & 0,0165 \end{pmatrix}; \quad X^T Y = \begin{pmatrix} 1496 \\ 520090 \\ 8367,9 \end{pmatrix};$$

$$B = (X^T X)^{-1} X^T Y = \begin{pmatrix} 8,8 \\ 0,2 \\ 6,97 \end{pmatrix}.$$

Следовательно, эконометрическая модель имеет вид:

$$y = 8,8 + 0,2x_1 + 6,97x_2.$$

Когда при всех одинаковых условиях независимая переменная x_1 (общие расходы) увеличивается (уменьшается) на единицу, то зависимая переменная y (оценка расходов на питание) также увеличивается (уменьшается) на 0,2 единиц.

Если при неизменных условиях независимая переменная x_2 (размер дохода на семью) увеличивается (уменьшается) на единицу, то зависимая переменная y (оценка расходов на питание) также увеличивается (уменьшается) на 6,97 единиц.

Если значение двух независимых переменных равно нулю, то значение y равно $b_0 = 8,8$.

Пример 4.2. Для эконометрической модели (пример 4.1) вычислить дисперсионно-ковариационную матрицу.

Решение.

Из примера 4.1 имеем:

$$B = \begin{pmatrix} 8,8 \\ 0,2 \\ 6,97 \end{pmatrix}, \quad C = (X^T X)^{-1} = \begin{pmatrix} 0,314 & -0,00017 & -0,0446 \\ -0,00017 & 0,00003 & -0,00012 \\ -0,0446 & -0,00012 & 0,0165 \end{pmatrix};$$

$$X^T Y = \begin{pmatrix} 1496 \\ 520090 \\ 8367,9 \end{pmatrix}, \quad n = 16; \quad m = 3.$$

Вычислим оценку несмещенной дисперсии остатков по формуле:

$$S_{\varepsilon}^2 = \frac{Y^T Y - B^T X^T Y}{n - m}.$$

$$Y^T Y = \sum_{i=1}^n y_i^2 = 22^2 + 34^2 + \dots + 197^2 = 176\,394;$$

$$B^T X^T Y = (8,8 \quad 0,2 \quad 6,97) \cdot \begin{pmatrix} 1\,496 \\ 520\,090 \\ 8\,367,9 \end{pmatrix} = 13\,156 + 104\,018 + 58\,324,0 = \\ = 175\,498,0;$$

$$S_{\varepsilon}^2 = \frac{176\,394 - 175\,498}{16 - 3} = \frac{896}{13} = 68,92.$$

Определим дисперсии оценок каждого параметра:

$$\text{var}(b_1) = \hat{\sigma}_{b_1}^2 = S_{\varepsilon}^2 \cdot c_{11} = 68,92 \cdot 0,314 = 21,64;$$

$$\text{var}(b_2) = \hat{\sigma}_{b_2}^2 = S_{\varepsilon}^2 \cdot c_{22} = 68,92 \cdot 0,00003 = 0,00207;$$

$$\text{var}(b_3) = \hat{\sigma}_{b_3}^2 = S_{\varepsilon}^2 \cdot c_{33} = 68,92 \cdot 0,0165 = 1,137.$$

Вычислим ковариации соответствующих оценок параметров:

$$\mu_{b_1 b_2} = S_{\varepsilon}^2 \cdot c_{12} = 68,92 \cdot (-0,00017) = -0,0118;$$

$$\mu_{b_1 b_3} = S_{\varepsilon}^2 \cdot c_{13} = 68,92 \cdot (-0,0446) = -3,0738;$$

$$\mu_{b_2 b_3} = S_{\varepsilon}^2 \cdot c_{23} = 68,92 \cdot (-0,00012) = -0,00827.$$

Знак "минус" перед оценками ковариаций $\mu_{b_j b_k}$ указывает на то, что увеличение одной оценки параметров приводит к уменьшению другой и наоборот.

Следовательно, дисперсионно-ковариационная матрица имеет вид:

$$\text{var}(B) = \begin{pmatrix} 21,64 & -0,0118 & -3,0738 \\ -0,0118 & 0,00207 & -0,00827 \\ -3,0738 & -0,00827 & 1,137 \end{pmatrix}.$$

Пример 4.3. Необходимо рассчитать для эконометрической модели (пример 4.1) точечный и интервальный прогнозы среднего и индивидуального значения зависимой переменной, когда для прогнозного периода

известен вектор $X_0 = \begin{pmatrix} 1 \\ 500 \\ 6 \end{pmatrix}$ с надежностью 95 %.

Решение.

Определяем точечный прогноз среднего значения зависимой переменной, когда $X_0 = \begin{pmatrix} 1 \\ 500 \\ 6 \end{pmatrix}$:

$$y_0 = 8,8 + 0,2x_1 + 6,97x_2 = 8,8 + 0,2 \cdot 500 + 6,97 \cdot 6 = 150,62.$$

Среднее y_0 можно интерпретировать как оценку прогнозного значения среднего и индивидуального значения расходов на питание, когда известны общие расходы x_1 и размер семьи составляет x_2 .

Определяем прогнозный интервал среднего значения y_0 .

Найдем дисперсию прогноза среднего значения:

$$\begin{aligned} S_p^2 &= S_\varepsilon^2 X_0^T (X^T X)^{-1} X_0 = X_0^T \text{var}(B) X_0 = \\ &= (1 \quad 500 \quad 6) \begin{pmatrix} 21,64 & -0,0118 & -3,0738 \\ -0,0118 & 0,00207 & -0,00827 \\ -3,0738 & -0,00827 & 1,137 \end{pmatrix} \begin{pmatrix} 1 \\ 500 \\ 6 \end{pmatrix} = \end{aligned}$$

$$= (-2,703 \quad 0,9736 \quad -0,3868) \cdot \begin{pmatrix} 1 \\ 500 \\ 6 \end{pmatrix} = 481,777.$$

Тогда стандартная ошибка прогноза среднего y_0 :

$$S_p = \sqrt{S_p^2} = \sqrt{481,777} = 21,95.$$

Определяем интервальный прогноз для среднего значения зависимой переменной.

При этом $\alpha = 0,05$, $n - m = 13$; тогда $t_{0,05;13} = 2,160$.

Следовательно:

$$\hat{y}_0 - t_\alpha S_p \leq \bar{y}_0 \leq \hat{y}_0 + t_\alpha S_p;$$

$$150,62 - 2,160 \cdot 21,95 \leq \bar{y}_0 \leq 150,62 + 2,160 \cdot 21,95;$$

$$150,62 - 47,412 \leq \bar{y}_0 \leq 150,62 + 47,412;$$

$$103,208 \leq \bar{y}_0 \leq 198,032.$$

Вычисляем дисперсию и стандартную ошибку прогнозного индивидуального значения зависимой переменной:

$$S_{p_1}^2 = S_p^2 + S_\varepsilon^2 = 481,777 + 68,92 = 550,697.$$

Стандартная ошибка прогноза индивидуального значения такая:

$$S_{p_1} = \sqrt{S_{p_1}^2} = \sqrt{550,697} = 23,467.$$

Определяем интервальный прогноз индивидуального значения зависимой переменной:

$$\hat{y}_0 - t_\alpha S_{p_1} \leq y_0 \leq \hat{y}_0 + t_\alpha S_{p_1};$$

$$150,62 - 2,160 \cdot 23,467 \leq y_0 \leq 150,62 + 2,160 \cdot 23,467;$$

$$150,62 - 50,689 \leq y_0 \leq 150,62 + 50,689;$$

$$99,931 \leq y_0 \leq 201,309.$$

Экономическая интерпретация: если в прогнозном периоде общие расходы имеют уровень 500 единиц, а семья состоит из шести лиц, то средние расходы на питание попадают в интервал:

$$103,208 \leq \overline{y_0} \leq 198,032.$$

В то же время отдельное (индивидуальное) значение этих расходов будет содержаться в более широком интервале, а именно:

$$99,931 \leq y_0 \leq 201,309.$$

Пример 4.4. В процессе изучения зависимости прибыли (тыс. грн) y от выработки продукции на одного работника (ед.) x_1 и индекса цен на продукцию (%) x_2 получены данные по 30 предприятиям (табл. 4.2).

Таблица 4.2

Исходные данные

Признак	Среднее значение	Среднее квадратическое отклонение	Парный коэффициент корреляции
y	250	38	$r_{yx_1} = 0,68$
x_1	47	12	$r_{yx_2} = 0,63$
x_2	112	21	$r_{x_1x_2} = 0,42$

Построить уравнение линейной множественной регрессии в стандартизованных переменных и уравнение "чистой" регрессии.

Рассчитать коэффициент множественной корреляции и скорректированный коэффициент множественной корреляции;

вычислить линейные частные коэффициенты корреляции, сравнить их с линейными парными коэффициентами корреляции, пояснить различия между ними.

Решение.

Применяя формулы для вычисления β -коэффициентов, получаем:

$$\beta_{x_1} = \frac{r_{yx_1} - r_{yx_2} \cdot r_{x_1x_2}}{1 - r_{x_1x_2}^2} = \frac{0,68 - 0,63 \cdot 0,42}{1 - 0,42^2} = 0,5044 ;$$

$$\beta_{x_2} = \frac{r_{yx_2} - r_{yx_1} \cdot r_{x_1x_2}}{1 - r_{x_1x_2}^2} = \frac{0,63 - 0,68 \cdot 0,42}{1 - 0,42^2} = 0,4182 .$$

Тогда получаем уравнение:

$$t_y = 0,5044t_{x_1} + 0,4182t_{x_2} .$$

Таким образом, фактор – выработка продукции на одного работника (ед.) x_1 влияет больше на изменение прибыли (тыс. грн) y , по сравнению с индексом цен на продукцию (%) x_2 .

Для построения уравнения "чистой" регрессии рассчитаем b_1 и b_2 , используя формулы для перехода от β_i к b_i :

$$b_i = \beta_i \frac{s_y}{s_{x_i}} .$$

$$\text{Тогда } b_1 = 0,5044 \frac{38}{12} = 1,5973 ; b_2 = 0,4182 \frac{38}{21} = 0,7567 .$$

Значение a определяем из соотношения:

$$a = \bar{y} - b_1\bar{x}_1 - b_2\bar{x}_2 = 250 - 1,5973 \cdot 47 - 0,7567 \cdot 21 = 159,0362 .$$

Уравнение "чистой" регрессии имеет вид:

$$\hat{y}_x = 159,0362 + 1,5973x_1 + 0,7567x_2 .$$

Таким образом, с изменением выработки продукции на одного работника (ед.) x_1 на единицу при неизменном значении второго фактора, закрепленного на среднем уровне, среднее изменение прибыли составит 1,5973 тыс. грн, а с изменением индекса цен на продукцию на 1 % (x_2) среднее изменение прибыли составит 0,7567 тыс. грн.

Частные коэффициенты корреляции:

$$r_{yx_1 \cdot x_2} = \frac{r_{yx_1} - r_{yx_2} \cdot r_{x_1 x_2}}{\sqrt{(1 - r_{yx_2}^2) \cdot (1 - r_{x_1 x_2}^2)}} = \frac{0,68 - 0,63 \cdot 0,42}{\sqrt{(1 - 0,63^2) \cdot (1 - 0,42^2)}} = 0,5894;$$

$$r_{yx_2 \cdot x_1} = \frac{r_{yx_2} - r_{yx_1} \cdot r_{x_1 x_2}}{\sqrt{(1 - r_{yx_1}^2) \cdot (1 - r_{x_1 x_2}^2)}} = \frac{0,63 - 0,68 \cdot 0,42}{\sqrt{(1 - 0,68^2) \cdot (1 - 0,42^2)}} = 0,5176.$$

Так как межфакторной связь средняя ($r_{x_1 x_2} = 0,42$), то наблюдаем небольшое расхождение между значениями парными и частными коэффициентами корреляции:

$$\begin{aligned} r_{yx_1} &= 0,68; & r_{yx_2} &= 0,63; \\ r_{yx_1 \cdot x_2} &= 0,5894; & r_{yx_2 \cdot x_1} &= 0,5176. \end{aligned}$$

Выводы о тесноте и направлении связи на основе парных и частных коэффициентов корреляции совпадают.

Коэффициент множественной корреляции:

$$\begin{aligned} R &= \sqrt{\beta_{x_1} \cdot r_{yx_1} + \beta_{x_2} \cdot r_{yx_2}} = \sqrt{0,5044 \cdot 0,68 + 0,4182 \cdot 0,63} = \\ &= \sqrt{0,6065} = 0,78, \end{aligned}$$

коэффициент детерминации: $R^2 \approx 0,61$.

Скорректированный коэффициент линейной множественной корреляции:

$$\bar{R} = \sqrt{0,61 \cdot \frac{30 - 1}{30 - 2 - 1}} = \sqrt{0,6552} = 0,81, \quad \bar{R}^2 \approx 0,66.$$

Таким образом, зависимость y от x_1 и x_2 характеризуется как средняя, в которой 66 % вариации средней прибыли (тыс. грн) определяются вариацией выработкой продукции на одного работника (ед.) и индексом цен на продукцию (%). Прочие факторы, не включенные в модель, составляют соответственно 34 % от общей вариации y .

4.3. Задачи

Задача 4.1. По результатам исследования зависимости прибыльности (y , млн грн) от затрат на оформление интерьера (x_1 , тис. грн) и стажа работы персонала (x_2 , год) на 20-ти торговых точках получены такие данные:

$$\sum x_{i1} = 4,88; \quad \sum x_{i1}^2 = 2,518; \quad \sum x_{i2} = 26,7;$$

$$\sum x_{i2}^2 = 75,15; \quad \sum y_i = 44,7; \quad \sum x_{i1}x_{i2} = 13,75;$$

$$\sum x_{i1}y_i = 22,1; \quad \sum x_{i2}y_i = 125,75; \quad \sum y_i^2 = 210,4; \quad \sum e_i^2 = 0,015.$$

Необходимо:

оценить коэффициенты линейной регрессии

$$Y = b_0 + b_1x_1 + b_2x_2 + \varepsilon;$$

определить стандартные ошибки коэффициентов модели;

вычислить R^2 и сделать вывод о модели.

Задача 4.2. По результатам исследования зависимости объема продаж (y , тис. грн) от скидки на товар (x_1 , %) и возраста персонала (x_2 , год.) на 20-ти торговых площадках получены такие данные:

$$\bar{x}_1 = 7,3; \quad \bar{x}_2 = 420,7; \quad \bar{y} = 350,3;$$

$$\sum (y_i - \bar{y})^2 = 62\,050,35; \quad \sum (x_{i1} - \bar{x}_1)^2 = 265,52;$$

$$\sum (x_{i2} - \bar{x}_2)^2 = 92\,845,072; \quad \sum (x_{i1} - \bar{x}_1)(x_{i2} - \bar{x}_2) = 4\,803,5;$$

$$\sum (x_{i1} - \bar{x}_1)(y_i - \bar{y}) = 3\,950,9; \quad \sum (x_{i2} - \bar{x}_2)(y_i - \bar{y}) = 75\,380,645.$$

Необходимо:

оценить коэффициенты линейной регрессии

$$Y = b_0 + b_1X_1 + b_2X_2 + \varepsilon;$$

вычислить коэффициент детерминации R^2 и сделать вывод.

Задача 4.3. По 10-ти актам купли-продажи однокомнатной квартиры получены данные: стоимость квартиры (тыс усл. ед.) y , общая площадь (м^2) x_1 и расстояние от нее до метро (мин.) x_2 :

$$\sum x_1 = 739; \quad \sum x_2 = 180; \quad \sum y = 734; \quad \sum x_1^2 = 27551; \quad \sum x_1 x_2 = 6615;$$

$$\sum x_2^2 = 1806; \quad \sum y^2 = 28020; \quad \sum y x_1 = 27513; \quad \sum y x_2 = 6357.$$

Построить линейное уравнение множественной регрессии.

Найти коэффициент множественной детерминации; сделать выводы.

Задача 4.4. По приведенным в табл. 4.3 статистическим данным построить регрессионные модели:

$$Y = \alpha_0 + \alpha_1 X_1 + \varepsilon, \quad Y = \gamma_0 + \gamma_2 X_2 + \varepsilon, \quad Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon.$$

Таблица 4.3

Исходные данные

Y	X_1	X_2
1	0	3
3	1	1
5	3	0
11	4	-2

Какие выводы можно сделать по построенным моделям?

Задача 4.5. Для исследования изменений валового национального продукта (*ВНП*) за 10 лет требуется построить регрессионную модель с объясняющими переменными – потребление (C) и инвестиции (I). Получены статистические данные, которые представлены в табл. 4.4.

Необходимо:

оценить коэффициенты уравнения линейной регрессионной модели

$$\text{ВНП} = b_0 + b_1 I + b_2 C + \varepsilon;$$

оценить стандартную ошибку регрессии и стандартные ошибки коэффициентов;

построить доверительные интервалы для параметров теоретической модели при уровне надежности $\gamma = 0,95$;

вычислить коэффициент детерминации R^2 и сделать вывод;
 через три года предполагается: $C = 22$, $I = 3,8$. Какой уровень *ВНП* можно ожидать при этом?

Рассчитать точечный и интервальный прогнозы среднего и индивидуального значения зависимой переменной для заданного прогнозного периода.

Таблица 4.4

Статистические данные

C , млрд	8	9,5	11	12	13	14	15	16,5	17	18
I , млрд	1,65	1,8	2,0	2,1	2,2	2,4	2,65	2,85	3,2	3,55
<i>ВНП</i> , млрд	14	16	18	20	23	23,5	25	26,5	28,5	30,5

Задача 4.6. Проведено исследование влияния климатических условий: x_1 – количество осадков, x_2 – средняя температура воздуха на y – урожай зерновых в 25-ти областях Украины.

Получена матрица парных коэффициентов корреляции (табл. 4.5).

Таблица 4.5

Матрица парных коэффициентов корреляции

Показатели	y	x_1	x_2
y	1	0,6	0,5
x_1	0,6	1	0,9
x_2	0,5	0,9	1

Построить эконометрическую модель в стандартизованных переменных. Сделать вывод о влиянии факторов x_1 , x_2 на результативный признак y . Вычислить частные коэффициенты корреляции с каждым из факторов.

Сравнить их с парными коэффициентами корреляции, сделать вывод.

Задача 4.7. На 30 производственных участках изучается зависимость потребления электроэнергии (y , грн) от объемов производства на 1 рабочего (x_1 , шт.) и уровня автоматизации работы (x_2 , грн). Данные приведены в табл. 4.6.

Таблица 4.6

Основные характеристики

Показатели	Среднее значение	Среднее квадратичное отклонение	Парный коэффициент корреляции
y	100	25	$r_{yx_1} = 0,67$
x_1	300	45	$r_{yx_2} = 0,54$
x_2	36,7	20	$r_{x_1x_2} = 0,23$

Построить уравнение линейной множественной регрессии в стандартизованных переменных и уравнение "чистой" регрессии.

Рассчитать коэффициент множественной корреляции, частные коэффициенты корреляции и сравнить их с парными коэффициентами корреляции, пояснить различия между ними.

Задача 4.8. По 20-ти наблюдениям получена матрица парных коэффициентов корреляции (табл. 4.7).

Таблица 4.7

Матрица парных коэффициентов корреляции

Показатели	y	x_1	x_2
y	1,0000	0,9699	0,9408
x_1	0,9699	1,0000	0,9428
x_2	0,9408	0,9428	1,0000

Определить показатель множественной корреляции.

Построить уравнение регрессии в стандартизованных переменных и сделать выводы.

Задача 4.9. Эконометрическая модель, которая характеризует зависимость между среднемесячной зарплатой \hat{y} производительностью труда x_1 и коэффициентом текучести кадров x_2 , имеет вид:

$$\hat{y} = 227,48 + 2,95x_1 - 0,22x_2.$$

Оценить стандартизованные параметры этой модели, если известны $\sigma_{x_1} = 21,58$; $\sigma_y = 4,42$; $\sigma_{x_2} = 4,49$.

Сделать выводы.

Задача 4.10. Известны статистические данные о цене на товар данного предприятия – X_1 , грн., и цене на товар-аналог конкурирующего предприятия – X_2 , грн (табл. 4.8).

Таблица 4.8

Исходные данные

X_1 , грн	10	8	12	15	14	16	18
X_2 , грн	9	7,2	10,8	13,5	12,6	14,4	16,2

Найти $\det(X^T X)$ и сделать вывод – можно ли эти переменные использовать в качестве регрессоров модели множественной линейной регрессии.

Задача 4.11. Корреляционная матрица для переменных задачи 4.10 имеет вид:

$$r = \begin{pmatrix} 1 & 0,63 & -0,39 \\ 0,63 & 1 & -0,57 \\ -0,39 & -0,57 & 1 \end{pmatrix}.$$

Сделать анализ этой матрицы.

Вычислить коэффициент множественной корреляции, дать интерпретацию его значения.

Задача 4.12. Построена модель зависимости среднемесячной рыночной цены акций предприятия (Y , грн/акция) от объема оплаченных дивидендов на акцию (X_1 , грн/месяц) и объема средств, направленных предприятием на расширение производства (X_2 , тис. грн/месяц):

$$Y = 2,4 + 1,6X_1 + 0,9X_2 + E.$$

Известно, что $S_\varepsilon^2 = 10,5$, $(X^T X)^{-1} = \begin{pmatrix} 1,22 & -0,22 & -0,11 \\ -0,22 & 0,14 & -0,07 \\ -0,11 & -0,07 & 0,101 \end{pmatrix}$,

$n = 20$.

Построить точечный и интервальный прогноз для среднего и индивидуального значения зависимой переменной, если в прогнозном

периоде известен вектор $X_0 = \begin{pmatrix} 1 \\ 10 \\ 2 \end{pmatrix}$, а уровень надежности $1 - \alpha = 0,99$.

Задача 4.13. В процессе изучения зависимости прибыли (тыс. грн) y от производительности труда одного рабочего (ед.) x_1 и индекса цен на продукцию (%) x_2 получены данные по 30-ти производственным участкам корпорации (табл. 4.9).

Таблица 4.9

Основные характеристики

Признак	Среднее значение	Среднее квадратичное отклонение	Парный коэффициент корреляции
y	250	38	$r_{yx_1} = 0,68$
x_1	47	12	$r_{yx_2} = 0,63$
x_2	112	21	$r_{x_1x_2} = 0,42$

Построить уравнение множественной регрессии в стандартизованных и натуральных переменных.

5. Оценка надежности многофакторной линейной модели

5.1. Основные теоретические сведения

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

Проверяется нулевая гипотеза H_0 о статистической незначимости коэффициента детерминации ($H_0: R^2 = 0$).

Для проверки гипотезы H_0 вычисляется F -статистика:

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

где $D_{\text{факт}} = \frac{\sum(\hat{y}_x - \bar{y})^2}{m}$ – факторная сумма квадратов на одну степень свободы;

$$D_{\text{ост}} = \frac{\sum(y - \hat{y}_x)^2}{n - m - 1}$$

свободы;

R^2 – коэффициент детерминации;

m – число параметров при переменных (в линейной регрессии совпадает с числом включенных в модель факторов);

n – число наблюдений.

При заданном уровне значимости α по таблицам критических точек распределения Фишера (приложения А, Б) находится критическое значение $F_{кр} = F_{\alpha}(m, n - m - 1)$.

H_0 гипотеза отклоняется, если $F_{\alpha}(m, n - m - 1) > F_{кр}$. Это равносильно тому, что $R^2 > 0$, то есть R^2 статистически значим.

Мерой для оценки включения соответствующего фактора в модель служит частный F -критерий Фишера.

В общем виде для фактора x_i статистика частного F -критерия Фишера определяется как:

$$F_{x_i} = \frac{R_{yx_i x_1 x_2 \dots x_i \dots x_m}^2 - R_{yx_1, \dots, x_{i-1}, x_{i+1}, \dots, x_m}^2}{1 - R_{yx_1 x_2 \dots x_i \dots x_m}^2} \cdot \frac{n - m - 1}{1}.$$

В числителе $R_{yx_i x_1 x_2 \dots x_i \dots x_m}^2 - R_{yx_1, \dots, x_{i-1}, x_{i+1}, \dots, x_m}^2$ – прирост доли объясненной вариации y за счет дополнительного включения в модель фактора x_i .

Когда в модели два фактора имеем:

$$F_{x_1} = \frac{R_{yx_1 x_2}^2 - r_{yx_2}^2}{1 - R_{yx_1 x_2}^2} \cdot \frac{n - m - 1}{1}; \quad F_{x_2} = \frac{R_{yx_1 x_2}^2 - r_{yx_1}^2}{1 - R_{yx_1 x_2}^2} \cdot \frac{n - m - 1}{1}.$$

Частный F -критерий Фишера оценивает значимость коэффициентов "чистой" регрессии или целесообразность включения в модель соответствующего фактора.

Зная величину F_{x_i} можно определить и статистику t -критерия для коэффициента регрессии при i -м факторе:

$$t_{b_i} = \sqrt{F_{x_i}}.$$

Если расчетное значение t_{b_i} с $n - m - 1$ степенями свободы больше табличного (приложения А, Б), то подтверждается значимость включенного в модель фактора.

Оценка значимости коэффициентов чистой регрессии (существенности факторного признака) по t -критерию Стьюдента может быть проведена без расчетов частных F -критериев Фишера:

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

где $b_i = \frac{S_y \sqrt{R_{yx_1 \dots x_m}^2 - R_{yx_1 \dots x_{i-1} x_{i+1} \dots x_m}^2}}{S_{x_i} \sqrt{1 - R_{x_i x_1 \dots x_m}^2}};$

стандартная ошибка коэффициента регрессии –

$$m_{b_i} = \frac{S_y \sqrt{1 - R_{yx_1 \dots x_m}^2}}{S_{x_i} \sqrt{1 - R_{x_i x_1 \dots x_m}^2}} \cdot \frac{1}{\sqrt{n - m - 1}},$$

или

$$m_{b_i} = S_\varepsilon \sqrt{c_{ii}},$$

c_{ii} – диагональный элемент матрицы $C = (X^T X)^{-1}$;

стандартная ошибка регрессии – $S_\varepsilon = \sqrt{\frac{\sum (y - \hat{y})^2}{n - m - 1}}$.

Поскольку коэффициент корреляции является также выборочной характеристикой, которая может отклоняться от своего "истинного" значения, значимость коэффициента корреляции также нуждается в проверке. Базируется она на t -критерии Стьюдента:

$$t = \frac{R \sqrt{n - m}}{\sqrt{1 - R^2}},$$

где R^2 – коэффициент детерминации модели;

R – коэффициент корреляции;

$n - m$ – число степеней свободы.

Если $|t| > t_\alpha$, где t_α – соответствующее табличное значение (приложение В) с $n - m$ степенями свободы, то можно сделать вывод о значимости коэффициента корреляции между зависимой и независимыми переменными модели.

5.2. Примеры

Пример 5.1. По данным табл. 5.1 построено уравнение множественной регрессии:

$$\hat{y} = -3,54 + 0,854x_1 + 0,367x_2,$$

где y – производительность труда рабочего;

x_1 – уровень квалификации рабочего;

x_2 – уровень автоматизации производства.

Определить множественный коэффициент детерминации и проверить значимость полученного уравнения на уровне $\alpha = 0,05$.

Таблица 5.1

Исходные данные эконометрической модели

y	5	10	10	7	5	6	6	5	6	8
x_1	8	11	12	9	8	8	9	9	8	12
x_2	5	8	8	5	7	8	6	4	5	7

Решение.

Для вычисления коэффициента детерминации воспользуемся формулой:

$$R^2 = 1 - \frac{S_{\varepsilon}^2}{S_y^2},$$

где $S_y^2 = \frac{\sum y^2}{n} - \left(\frac{\sum y}{n}\right)^2 = \overline{y^2} - (\bar{y})^2$ – общая дисперсия результативного признака (по эмпирическим данным);

$$S_{\varepsilon}^2 = \frac{\sum (y - \hat{y})^2}{n} \text{ – остаточная дисперсия.}$$

Для удобства вычислений составим вспомогательную таблицу (табл. 5.2).

Вспомогательные расчеты

№ п/п	y_i	x_{1i}	x_{2i}	y_i^2	\hat{y}_i	$\varepsilon_i^2 = (y_i - \hat{y}_i)^2$
1	5	8	5	25	5,13	0,016
2	10	11	8	100	8,79	1,464
3	10	12	8	100	9,64	1,127
4	7	9	5	49	5,98	1,038
5	5	8	7	25	5,86	0,741
6	6	8	8	36	6,23	0,052
7	6	9	6	36	6,35	0,121
8	5	9	4	25	5,61	0,377
9	6	8	5	36	5,13	0,762
10	8	12	7	64	9,28	1,631
Σ	68	94	63	496	–	6,329

Тогда:

$$S_y^2 = \frac{\sum y^2}{n} - \left(\frac{\sum y}{n} \right)^2 = \frac{496}{10} - \left(\frac{68}{10} \right)^2 = 3,36,$$

$$S_\varepsilon^2 = \frac{\sum (y - \hat{y})^2}{n} = \frac{6,329}{10} = 0,6329,$$

$$R^2 = 1 - \frac{S_\varepsilon^2}{S_y^2} = 1 - \frac{0,6329}{3,36} = 0,811.$$

Полученный коэффициент детерминации свидетельствует о том, что вариация исследуемой переменной y на 81,1 % объясняется изменчивостью включенных в модель объясняющих переменных.

Проверим значимость уравнения регрессии.

Имеем:

$$F_p = \frac{R^2}{1 - R^2} \cdot \frac{n - m - 1}{m} = \frac{0,811}{1 - 0,811} \cdot \frac{10 - 2 - 1}{2} = 15,0.$$

Табличное значение равно $F_{\alpha=0,05}(2, 7) = 4,74$.

Фактическое значение статистики F -критерия больше табличного, то есть уравнение регрессии значимо, следовательно, с надежностью 95 % исследуемая зависимая переменная достаточно хорошо описывается включенными в регрессионную модель объясняющими переменными.

Пример 5.2. Для эконометрической модели зависимости прибыли (тыс. грн) y от выработки продукции на одного работника (ед.) x_1 и индекса цен на продукцию (%) x_2 при $n = 30$ рассчитан коэффициент детерминации $R^2 = 0,6065$. Известно, что $r_{yx_2} = 0,63$, $r_{yx_1} = 0,68$.

Рассчитать общий и частные F -критерий Фишера с надежностью 95 %.

Решение

Определяем фактическое и табличное значения F -статистики:

$$F_p = \frac{R^2}{1 - R^2} \cdot \frac{n - m - 1}{m} = \frac{0,6065}{1 - 0,6065} \cdot \frac{30 - 2 - 1}{2} = 20,8075 ;$$

$$F_{\alpha=0,05}(2, 27) = 3,35.$$

Видим, что $F_p > F_{табл.}$. Следовательно, уравнение зависимости статистически значимо с надежностью 95 %.

Проверим целесообразность включения в модель соответствующего фактора с помощью частных критериев.

Табличное значение F -статистики равно:

$$F_{\alpha=0,05}(1, 27) = 4,21.$$

Частный F -критерий для фактора x_1 определим по формуле:

$$F_{x_1} = \frac{R_{yx_1x_2}^2 - r_{yx_2}^2}{1 - R_{yx_1x_2}^2} \cdot \frac{n - m - 1}{1} = \frac{0,6065 - 0,63^2}{1 - 0,6065} \cdot \frac{30 - 2 - 1}{1} = 14,3817 ,$$

что больше табличного. Поэтому включение в модель фактора x_1 после фактора x_2 статистически оправдано – доля объясненной вариации возрастает с включением фактора на

$$(0,6065 - 0,63^2) \cdot 100 = 20,96 \%$$

Частный F -критерий для фактора x_2 определим по формуле:

$$F_{x_2} = \frac{R_{yx_1x_2}^2 - r_{yx_1}^2}{1 - R_{yx_1x_2}^2} \cdot \frac{n - m - 1}{1} = \frac{0,6065 - 0,68^2}{1 - 0,6065} \cdot \frac{30 - 2 - 1}{1} = 9,8824,$$

что больше табличного. Поэтому включение в модель фактора x_2 после фактора x_1 статистически оправдано – доля объясненной вариации возрастает с включением фактора на

$$(0,6065 - 0,68^2) \cdot 100 = 14,41 \%$$

Пример 5.3. Вычислить F -критерий ($1 - \alpha = 95 \%$) для эконометрических моделей, представленных в табл. 5.3.

Сравнить полученные результаты с расчетами значений статистик t -критерия.

Таблица 5.3

Эконометрические модели

Эконометрическая модель	Число степеней свободы	F -критерий
1) $\hat{y} = -9,77 + 0,23x_1$, $R^2 = 0,9$	$\begin{cases} m - 1 = 1 \\ n - m = 8 \end{cases}$	34,24
2) $\hat{y} = -20,63 + 0,23x_1 + 0,09x_3$, $R^2 = 0,91$	$\begin{cases} m - 1 = 2 \\ n - m = 7 \end{cases}$	19,45
3) $\hat{y} = -23,83 + 0,23x_1 + 0,10x_2 + 9,02x_3$, $R^2 = 0,904$	$\begin{cases} m - 1 = 3 \\ n - m = 6 \end{cases}$	12,09

Решение

$F_{1табл}(0,95)$ для первой модели равно 5,32.

$F_{2табл}(0,95)$ для второй модели равно 4,74.

$F_{3табл}(0,95)$ для третьей модели равно 4,76.

Следовательно, при уровне значимости $\alpha = 0,05$:

$$F_{1факт} > F_{1табл}(0,95);$$

$$F_{2факт} > F_{2табл}(0,95);$$

$$F_{3факт} > F_{3табл}(0,95).$$

Это значит, что соответствующие эконометрические модели являются достоверными. То есть подтверждается гипотеза о том, что количественная оценка связи между зависимой и независимыми переменными в модели является существенной.

Для множественных коэффициентов корреляции, которые приведены в табл. 5.3, вычислим значения статистик t -критерия:

$$t_1 = \frac{0,9\sqrt{8}}{\sqrt{1-0,81}} = \frac{0,9\sqrt{8}}{\sqrt{0,19}} = 5,84;$$

$$t_2 = \frac{0,91\sqrt{7}}{\sqrt{1-0,828}} = \frac{0,91\sqrt{7}}{\sqrt{0,172}} = 5,82;$$

$$t_3 = \frac{0,904\sqrt{6}}{\sqrt{1-0,817}} = \frac{0,904\sqrt{6}}{\sqrt{0,183}} = 5,18.$$

Табличные значения этого критерия при уровне значимости $\alpha = 0,05$ и соответствующих степенях свободы такие:

$$t_{1табл} = 1,860; \quad t_{2табл} = 1,895; \quad t_{3табл} = 1,943.$$

Сравнивая их с фактическими, приходим к выводу, что коэффициенты корреляции, которые характеризуют тесноту связи между зависимой и независимыми переменными в моделях, являются достоверными.

Пример 5.4. Определить с надежностью 95 % статистическую значимость коэффициентов линейной множественной регрессионной модели зависимости объема предложения товара y от цены товара x_1 и зарплаты сотрудников x_2 :

$$\hat{y} = 90,74 + 0,88x_1 - 7,32x_2,$$

если $n = 10$; $m_a = 24,87$; $m_{b_1} = 0,36$; $m_{b_2} = 1,91$.

Решение.

Обозначим $H_0 : b_i = 0$, то есть объясняющая переменная x_i не влияет на результирующий признак y . Тогда $H_1 : b_i \neq 0$, то есть объясняющая переменная x_i влияет на результирующий признак y .

Вычислим значения t -статистики:

$$t_a = \frac{a}{m_a} = \frac{90,74}{24,87} = 3,649;$$

$$t_{b_1} = \frac{b_1}{m_{b_1}} = \frac{0,88}{0,36} = 2,44;$$

$$t_{b_2} = \frac{b_2}{m_{b_2}} = -\frac{7,32}{1,91} = -3,832.$$

Поскольку граничные точки $\pm t_{\alpha, n-m-1} = \pm 2,365$, то гипотезу H_0 отвергаем и принимаем гипотезу H_1 с надежностью 95 %. Все коэффициенты статистически значимы, то есть x_1 и x_2 влияют на результирующий признак y .

5.3. Задачи

Задача 5.1. В результате исследования факторов, которые определяют экономический рост, по 73 странам получено уравнение регрессии:

$$\hat{G} = 1,4 - 0,52P + 0,17S + 11,16I - 0,38D - 4,75In, R^2 = 0,60,$$

(−5,9) (4,34) (3,91) (−0,79) (−2,7)

где \hat{G} – темп экономического роста (темп роста среднего ВВП в % к базисному периоду);

P – реальный среднедушевой ВВП, %;

S – бюджетный дефицит, % к ВВП;

I – объем инвестиций, % к ВВП;

D – внешний долг, % к ВВП;

In – уровень инфляции, %.

Проверить гипотезу о достоверности полученной модели в целом.

В скобках указаны наблюдаемые значения t -критерия Стьюдента.

Проверить общее качество данной модели.

Проверить значимость параметров модели при уровне значимости $\alpha = 0,05$ и записать, учитывая сделанные выводы, теоретическую модель.

Указывают ли эмпирические результаты данной модели на наличие обратной связи между темпами экономического роста и объемом внешнего долга страны (% к ВВП)?

Построить доверительные интервалы для параметров теоретической модели при уровне надежности $\gamma = 0,95$.

Задача 5.2. Проведен анализ зависимости уровня рентабельности предприятий y от уровня инвестиций – x_1 и объемов производства – x_2 на 50 предприятиях.

Получены такие данные: $R^2 = 0,86$, $r_{yx_1} = 0,65$, $r_{yx_2} = 0,52$.

Сделать вывод об индексе детерминации. Оценить с надежностью 95 % значимость уравнения множественной регрессии по критерию Фишера; исследовать целесообразность включения соответствующих факторов в уравнение регрессии.

Задача 5.3. По результатам работы 20-ти производственных участков предприятия легкой промышленности получена информация, которая

характеризует зависимость объемов производства продукции (млн грн) y от количества обработанных человеко-часов (тыс. чел.-часов) x_1 и среднегодовой стоимости производственного оборудования (млн грн) x_2 (табл. 5.4).

Таблица 5.4

Исходные данные

Уравнение регрессии	$y = 35 + 0,06x_1 + 2,5x_2$
Множественный коэффициент корреляции	0,9
Сумма квадратов отклонений вычисленных значений результатов от фактических	3 000

Найти коэффициент детерминации в этой модели и оценить его статистическую значимость с надежностью 95 %.

Задача 5.4. Производственная функция, которая получена по данным 7 лет, характеризуется уравнением:

$$\lg \hat{P} = 0,552 + 0,276 \cdot \lg Z + 0,521 \cdot \lg K,$$

(0,584) (0,065)

$$R^2 = 0,9843, \quad r_{PZ}^2 = 0,7826, \quad r_{PK}^2 = 0,9836,$$

где P – индекс промышленного производства;

Z – численность рабочих;

K – капитал.

В скобках указаны значения стандартных ошибок для коэффициентов регрессии.

Дать интерпретацию параметров уравнения регрессии.

Оценить значимость параметров регрессии с помощью t -критерия Стьюдента и сделать соответствующие выводы о целесообразности включения факторов в модель.

Найти частные значения F -критерия и сделать выводы.

Оценить значимость уравнения регрессии в целом с помощью F -критерия Фишера. Какая роль факторов, которые не учтены в модели, в вариации индекса промышленного производства?

Задача 5.5. Проведено исследование влияния x_1 – численности работников, x_2 – среднегодовой стоимости основных средств на y – объем производства продукции на 20 производственных участках предприятия. Получена матрица парных коэффициентов корреляции (табл. 5.5).

Таблица 5.5

Матрица парных коэффициентов

Показатели	y	x_1	x_2
y	1	0,3	0,6
x_1	0,3	1	0,8
x_2	0,6	0,8	1

Оценить значимость уравнения регрессии с помощью F -критерия Фишера.

Задача 5.6. Получена информация по 25-ти объектам наблюдения (табл. 5.6). Построить уравнение регрессии в стандартизованных переменных и сделать выводы.

Таблица 5.6

Исходные данные

Признак	Среднее значение	Коэффициент вариации, %	Уравнение регрессии
y	35	20	$\hat{y} = 20 + x_1 - 2x_2$
x_1	16	30	$\hat{y} = 9 + 1,1x_1$
x_2	8	10	$\hat{y} = 4 - 4,1x_2$

Оценить значимость каждого уравнения регрессии, если известно, что $r_{x_1x_2} = -0,35$.

Оценить значимость коэффициентов регрессии уравнения с двумя объясняющими переменными.

Определить частные коэффициенты корреляции. Сделать вывод.

Задача 5.7. При изучении уровня потребления мяса (кг на душу населения) y в зависимости от дохода (грн на одного члена семьи) x_1 и уровня потребления рыбы (кг на душу населения) x_2 получены такие результаты (по 50-ти семьям) (табл. 5.7).

Таблица 5.7

Основные результаты

Уравнение регрессии	$\hat{y} = -180 + 0,2 x_1 - 0,4 x_2$
Стандартные ошибки параметров	20 0,01 0,25
Множественный коэффициент корреляции	0,85

Используя t -критерий, оценить значимость параметров уравнения с надежностью 95 %.

Вычислить F -критерий Фишера, сделать выводы.

Оценить с помощью частных F -критериев Фишера целесообразность включения в модель фактора x_1 после фактора x_2 и фактора x_2 после фактора x_1 с надежностью 99 %.

Задача 5.8. На 40 производственных участках предприятия исследовалась зависимость производительности труда y от уровня квалификации рабочих x_1 и энерговооруженности их труда x_2 (табл. 5.8).

Таблица 5.8

Исходные данные

Уравнение регрессии	$\hat{y} = a + 10 x_1 + 2 x_2$
Стандартные ошибки параметров	0,5 2 ?
t -критерий для параметров	3 ? 5
Множественный коэффициент корреляции	0,85

Определить параметр a и заполнить пропущенные значения. Какой из факторов имеет более сильное влияние на результат?

Оценить значимость уравнения в целом, используя значение множественного коэффициента корреляции.

Задача 5.9. По 30-ти предприятиям были получены результаты регрессионного анализа зависимости объемов производства продукции (млн грн) y от численности занятых на предприятии (лиц) x_1 и среднегодовой стоимости основных фондов x_2 (млн грн) (табл. 5.9).

Исходные данные

Коэффициент детерминации	???
Множественный коэффициент корреляции	0,85
Уравнение регрессии	$y = ??? + 0,48x_1 + 20x_2$
Стандартные ошибки параметров	2 0,06 ?
t -критерий для параметров	1,5 ? 4

Восстановить пропущенные характеристики.

С вероятностью 0,95 построить доверительные интервалы для коэффициентов регрессии.

Проанализировать результаты регрессионного анализа.

Задача 5.10. По 20-ти актам купли-продажи однокомнатной квартиры получены данные:

$$\sum x_1 = 739; \quad \sum x_2 = 180; \quad \sum y = 734; \quad \sum x_1^2 = 27551; \quad \sum x_1 x_2 = 6615;$$

$$\sum x_2^2 = 1806; \quad \sum y^2 = 28020; \quad \sum y x_1 = 27513; \quad \sum y x_2 = 6357.$$

Здесь y – стоимость квартиры (тыс. усл. ед.); x_1 – общая площадь (м^2) квартиры; x_2 – расстояние от нее до метро (мин.).

Построить линейное уравнение множественной регрессии.

Найти коэффициент множественной детерминации, в том числе скорректированный; сделать выводы.

Оценить значимость уравнения регрессии по F -критерию Фишера.

Оценить значимость коэффициентов регрессии по t -критерию Стьюдента.

Задача 5.11. Эконометрическая модель, которая характеризует зависимость между среднемесячной зарплатой \hat{y} , производительностью труда x_1 и коэффициентом текучести рабочей силы x_2 , имеет вид:

$$\hat{y} = 227,48 + 2,95x_1 - 0,22x_2.$$

Известно, что $\sigma_{x_1} = 21,58$; $\sigma_y = 4,42$; $\sigma_{x_2} = 4,49$.

Корреляционная матрица для переменных имеет вид:

$$r = \begin{pmatrix} 1 & 0,63 & -0,39 \\ 0,63 & 1 & -0,57 \\ -0,39 & -0,57 & 1 \end{pmatrix}.$$

Вычислить множественный коэффициент корреляции и детерминации.

Найти дисперсию остатков для модели, если известен вектор остатков:

$$E = \begin{pmatrix} 2,8 \\ -16,2 \\ 1,4 \\ 13,5 \\ -12,7 \\ 10,3 \\ 16,1 \\ 24,3 \\ -10,8 \\ -28,8 \end{pmatrix}.$$

Вычислить стандартные ошибки параметров модели и статистику t -критерия для оценки достоверности параметров модели.

Задача 5.12. Построить доверительные интервалы для параметров модели примера 5.11.

Вычислить F -критерий на основе R^2 .

Можно ли согласно этому критерию отбросить или принять нулевую гипотезу о существенности связи в эконометрической модели?

Задача 5.13. Имеются данные о годовых ставках доходов за месяц по трем акциям за шестимесячный период (табл. 5.10). Есть основания предполагать, что доходы акции Y зависят от доходов по акциям X_1 и X_2 .

Исходные данные

Акция	Доходы по месяцам, %					
Y	5,4	5,3	4,9	4,9	5,4	6,0
X_1	6,3	6,2	6,1	5,8	5,7	5,7
X_2	9,2	9,2	9,1	9,0	8,7	8,6

Необходимо:

составить уравнение линейной регрессии $Y = f(X_1, X_2)$;

найти множественный коэффициент детерминации и пояснить его смысл;

проверить значимость полученного уравнения регрессии на уровне $\alpha = 0,05$;

оценить значимость коэффициентов регрессии по t -критерию Стьюдента.

6. Мультиколлинеарность, ее последствия и методы устранения

6.1. Основные теоретические сведения

Серьезной проблемой при построении моделей множественной линейной регрессии по МНК является мультиколлинеарность – линейная взаимосвязь двух или нескольких объясняющих переменных.

Причиной возникновения мультиколлинеарности в экономических исследованиях является существование соотношений между объясняющими переменными.

Существует несколько признаков, по которым может быть установлено наличие мультиколлинеарности:

1) коэффициент детерминации R^2 достаточно высок, но некоторые из коэффициентов регрессии статистически незначимы, то есть они имеют низкие значения t -статистики;

2) близость определителя матрицы $X^T X$ к нулю;

3) парная корреляция между малозначимыми объясняющими переменными достаточно сильная ($|r_{ij}| > 0,8$);

4) достаточно большие частные коэффициенты корреляции;

5) сильная вспомогательная (дополнительная регрессия), то есть какая-либо из объясняющих переменных является линейной комбинацией других объясняющих переменных. В этом случае для анализа строятся уравнения регрессии каждой из объясняющих переменных от оставшихся. Вычисляются соответствующие коэффициенты детерминации R_j^2 и рассчитывается их статистическая значимость по критерию Фишера.

Если коэффициент R_j^2 статистически незначим, то соответствующая переменная не является линейной комбинацией других переменных и ее можно оставить в уравнении регрессии. В противном случае есть основания считать, что эта переменная существенно зависит от других объясняющих переменных и имеет место мультиколлинеарность.

Для обнаружения наличия мультиколлинеарности используют *алгоритм Феррара – Глобера*.

Алгоритм Феррара – Глобера

1. Стандартизация (нормализация переменных):

$$x_i^* = \frac{x_{ik} - \bar{x}_i}{\sqrt{n\sigma_{x_i}^2}},$$

где n – число наблюдений ($k = \overline{1, n}$);

m – число объясняющих переменных ($i = \overline{1, m}$);

\bar{x}_i – среднее значение объясняющей переменной;

$\sigma_{x_i}^2$ – дисперсия объясняющей переменной.

2. Вычисление корреляционной матрицы $X^{*T} X^*$.

3. Определение статистики критерия χ^2 ("хи-квадрат"):

$$\chi^2 = - \left[n - 1 - \frac{1}{6}(2m + 5) \right] \ln|r|,$$

где $|r|$ – определитель корреляционной матрицы $X^{*T} X^*$.

Значение этого критерия сравнивается с табличным значением (приложение А) при $\frac{1}{2}m(m-1)$ степенях свободы и уровне значимости α . Если $\chi_{\text{факт}}^2 > \chi_{\text{табл}}^2$, то в массиве объясняющих переменных существует мультиколлинеарность.

4. Определение обратной матрицы $C = (X^{*T} X^*)^{-1}$.

5. Вычисление значений статистики F -критерия:

$$F = |c_{ii} - 1| \frac{n - m}{m - 1},$$

где c_{ii} – диагональные элементы обратной матрицы $C = (X^{*T} X^*)^{-1}$;

n – количество наблюдений;

m – количество аргументов в модели.

Величина F имеет распределение Фишера с $m-1$, $n-m$ степенями свободы.

По таблицам критерия Фишера (приложения А, Б) находим $F_{\alpha}(m-1; n-m)$.

Если $F < F_{0,05}(m-1; n-m)$, то гипотеза о наличии мультиколлинеарности отклоняется;

если $F > F_{0,01}(m-1; n-m)$, то соответствующая переменная мультиколлинеарна с другими.

Коэффициент детерминации для каждой переменной:

$$R_{x_i}^2 = 1 - \frac{1}{c_{ii}}.$$

6. Вычисление частных коэффициентов корреляции:

$$r_{ij} = \frac{-c_{ij}}{\sqrt{c_{ii} \cdot c_{jj}}},$$

где c_{ij} – элемент матрицы $C = (X^{*T} X^*)^{-1}$, который находится в i -ой строке и j -ом столбце;

c_{ii} и c_{jj} – диагональные элементы матрицы $C = (X^{*T} X^*)^{-1}$.

7. Вычисление статистик t -критерия (определение статистической значимости частных коэффициентов корреляции):

$$t_{ij} = \frac{|r_{ij}| \sqrt{n-m}}{\sqrt{1-r_{ij}^2}}.$$

Фактические значения t_{ij} сравниваются с табличными (приложение В) при $n - m$ степенях свободы и уровне значимости α .

Если $t_{ij} > t_{\alpha}(n-m)$, то между независимыми переменными x_i и x_j существует мультиколлинеарность.

Основными методами устранения мультиколлинеарности являются:

1. Метод сравнения значений линейных коэффициентов корреляции. Этот метод заключается в том, что высоко коррелирующие объясняющие переменные ($|r_{ij}| > 0,8$) исключаются из регрессии, и она оценивается еще раз. Какую именно переменную необходимо исключить определяют на основании экономического анализа зависимой переменной, предпочтение обычно отдается той объясняющей переменной, которая наиболее тесно связана с результативным признаком.

2. Исключение тренда (при построении регрессии по данным, полученным из временных рядов).

3. Получение дополнительных данных или новой выборки. Использование предварительной информации о некоторых параметрах.

4. Изменение спецификации модели: либо изменяется форма модели, либо добавляются объясняющие переменные, не учтенные в первоначальной модели, но существенно влияющие на зависимую переменную.

5. Метод пошаговой регрессии – метод включения переменных. В анализ последовательно включают по одной объясняющей переменной. На каждом шаге проверяется значимость коэффициентов регрессии и оценивается мультиколлинеарность переменных.

Если оценка коэффициента получается незначимой, то переменная исключается и рассматривается другая объясняющая переменная. Таким образом, пошагово определяются все составляющие регрессии с выполнением условия отсутствия мультиколлинеарности.

7. Преобразование переменных (например, переход к относительным величинам).

8. Метод гребневой (ридж-) регрессии. Перед обращением матрицы $X^T X$ к ее диагональным элементам прибавляются малые числа k – гребневые константы. Это предупреждает вырождение матрицы, увеличивает ее определитель и уменьшает дисперсию параметров.

Новые оценки (ридж-оценки) параметров вычисляются по формуле:

$$B = (X^T X + kE)^{-1} X^T Y, \text{ где } E \text{ – единичная матрица.}$$

Устойчивость оценок параметров характеризуется суммой их дисперсий $\sum_{i=1}^m s_{b_i}^2$.

Если имеется мультиколлинеарность, то при $k \rightarrow 0$ начинается резкое возрастание некоторых (если не всех) коэффициентов регрессии. Необходимо выбрать такое наименьшее значение параметра k , при котором поведение всех коэффициентов регрессии уже стабилизировалось, но еще не начался интенсивный рост остаточной дисперсии.

9. Метод главных компонент. При этом исходную стандартизованную систему объясняющих переменных X_i преобразовывают некоррелируемой системой компонент F_i , имеющей свойства: первая компонента F_1 объясняет максимум общей изменчивости данных, вторая компонента F_2 – максимум остатка изменчивости и так далее. Несколько первых (главных) компонент объясняют почти всю изменчивость данных, а остальные компоненты отображают случайные ошибки и могут быть отброшены. В модели с главными компонентами выполняют обратный переход к исходным переменным X_i , тем самым обходят проблему мультиколлинеарности. Компоненты составляются как линейные комбинации: $F_i = u_1X_1 + u_2X_2 + \dots + u_mX_m$, коэффициенты которой являются элементами собственных векторов $U_i = (u_1, u_2, \dots, u_m)^T$ корреляционной матрицы $R = (X^T X) / n$; дисперсии компонент равны соответствующим собственным числам λ_i корреляционной матрицы.

6.2. Примеры

Пример 6.1. Проведено исследование зависимости трудоемкости обработки детали (y) от количества операций (x_1), количества обработанных поверхностей (x_2), количества деталей (x_3).

Промежуточные расчетные данные приведены в табл. 6.1 и табл. 6.2.

Таблица 6.1

Промежуточные расчетные данные

№ п/п	x_1	x_2	x_3	y	\hat{y}	ε
Средние значения	295,46	524,33	157,38	181,68	181,68	0
Дисперсия	872,92	1 803,4	4 644,4	7 163,32	7 110,01	53,31

Корреляционная таблица

Показатели	x_1	x_2	x_3	y
x_1	1	0,9967	0,9939	0,9813
x_2	0,9967	1	0,9955	0,9901
x_3	0,9939	0,9955	1	0,9925
y	0,9813	0,9901	0,9925	1

На одном из этапов метода пошаговой регрессии получено следующее уравнение:

$$\hat{y} = 254,03 - 1,174x_1 + 1,738x_3.$$

С помощью МНК получено уравнение регрессии:

$$\hat{y} = -253,61 - 3,027x_1 + 2,181x_2 + 1,184x_3,$$

(t_b) $(5,7)$ $(5,0)$ $(5,9)$

где в скобках указаны наблюдаемые значения t -критерия Стьюдента.

Провести исследование наличия мультиколлинеарности и эффективности метода пошаговой регрессии.

Решение.

Полученная модель значима в целом и имеет только статистически значимые факторы (значения статистик t -критерия Стьюдента $t_{b_i} > t_{0,05} = 2,0$).

Коэффициент детерминации вычислим двумя способами:

$$R^2 = 1 - \frac{S_\varepsilon^2}{S_y^2} = 1 - \frac{53,31}{7\ 163,32} = 0,9926$$

или

$$R^2 = \frac{S_{\hat{y}}^2}{S_y^2} = \frac{7\ 110,01}{7\ 163,32} = 0,9926.$$

Коэффициент множественной корреляции очень высокий $R = \sqrt{R^2} = 0,996$. На первый взгляд полученная модель качественная. Однако имеем отрицательный коэффициент регрессии при x_1 . Это интерпретируется так: чем больше операций при изготовлении детали, тем меньше трудоемкость, что противоречит экономическому смыслу задачи.

Построим модель зависимости в стандартизованных переменных:

$$\hat{Y} = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3, \text{ где } \beta_i = b_i \frac{S_{x_i}}{S_y}.$$

Вычислим β -коэффициенты:

$$\beta_1 = -3,027 \cdot \sqrt{\frac{872,92}{7\,163,32}} = -1,057;$$

$$\beta_2 = 2,181 \cdot \sqrt{\frac{1\,803,4}{7\,163,32}} = 1,094;$$

$$\beta_3 = 1,184 \cdot \sqrt{\frac{4\,644,4}{7\,163,32}} = 0,953.$$

Тогда модель зависимости в стандартизованных переменных имеет вид:

$$\hat{Y} = -1,057 X_1 + 1,094 X_2 + 0,953 X_3,$$

где β -коэффициенты оказались большими единицы (по модулю).

Одной из причин этого может быть мультиколлинеарность.

Действительно, из табл. 6.2 следует, что между объясняющими переменными существуют очень большие корреляции. Фактически имеем не три объясняющие переменные, а три разные формы одной переменной. Метод пошаговой регрессии в данном примере не устранил проблему подключения в модель тесно связанных групп показателей.

Рассмотрим пошагово процесс составления регрессионной модели. Результативный признак y теснее связан с количеством деталей x_3 . Учет одной переменной x_3 объясняет $(0,9925)^2 = 0,9851$ (98,51 %) общей изменчивости результативного признака y .

Трехфакторная модель объясняет 99,26 % общей изменчивости y , то есть дополнительный вклад остальных факторов x_1, x_2 составляет только 0,75 % изменчивости y . Учет x_3 фактически исчерпывает изменчивость всех переменных, от которых остаются только малые случайные остатки. Однако подключение в модель x_1 дает абсурдную с экономической точки зрения модель: $\hat{y} = 254,03 - 1,174x_1 + 1,738x_3$.

Далее подключается x_2 . Следовательно, метод пошаговой регрессии не устраняет проблему мультиколлинеарности и в модель может войти группа тесно связанных показателей.

Пример 6.2. На среднемесячную плату (Y) влияет ряд независимых факторов: производительность труда (x_1), фондоемкость (x_2) и коэффициент текучести рабочей силы (x_3).

Исходные данные приведены в табл. 6.3.

Таблица 6.3

Исходные данные

Номер цеха	Производительность труда, человеко-дней	Фондоемкость, млн усл. ед.	Коэффициент текучести рабочей силы, %
1	32	0,89	19,5
2	29	0,43	15,6
3	30	0,70	13,5
4	31	0,61	9,5
5	25	0,51	23,5
6	34	0,51	12,5
7	29	0,65	17,5
8	24	0,43	14,5
9	20	0,51	14,5
10	33	0,92	7,5

Исследовать наличие или отсутствие мультиколлинеарности между приведенными независимыми переменными для модели, которую необходимо построить с помощью МНК.

Решение.

Используем алгоритм Феррара – Глобера.

1. Стандартизация переменных:

$$x_{ik}^* = \frac{x_{ik} - \bar{x}_k}{\sqrt{n\sigma_{xk}^2}},$$

где n – число наблюдений ($n = 10$);

\bar{x}_k – среднее значение переменной x_k ;

σ_{xk}^2 – дисперсия переменной x_k .

Дисперсии каждой независимой переменной имеют значение:

$$\sigma_{x_1}^2 = \frac{1}{n} \sum_{i=1}^{10} (x_{i1} - \bar{x}_1)^2 = \frac{176,1}{10} = 17,614;$$

$$\sigma_{x_2}^2 = \frac{1}{n} \sum_{i=1}^{10} (x_{i2} - \bar{x}_2)^2 = \frac{0,19524}{10} = 0,0195;$$

$$\sigma_{x_3}^2 = \frac{1}{n} \sum_{i=1}^{10} (x_{i3} - \bar{x}_3)^2 = \frac{182,4}{10} = 18,24.$$

Расчеты для стандартизации с переменной x_1, x_2, x_3 приведены в табл. 6.4.

Таблица 6.4

Основные расчеты стандартизации переменных

$(x_{i1} - \bar{x}_1)^2$	$(x_{i2} - \bar{x}_2)^2$	$(x_{i3} - \bar{x}_3)^2$	x_{i1}^*	x_{i2}^*	x_{i3}^*
1	2	3	4	5	6
10,89	0,000016	11,56	0,2487	0,0091	-0,2518
0,09	0,0247336	2,56	0,0226	-0,3531	0,1185
1,89	0,012995	0,16	0,0980	0,2580	-0,0296

1	2	3	4	5	6
5,29	0,000576	19,36	0,1733	0,0543	-0,3258
13,89	0,005776	92,16	-0,2788	-0,1720	0,7108
28,09	0,005776	1,96	0,3994	-0,1720	-0,1037
0,09	0,004096	12,96	0,0226	0,1448	0,2666
22,09	0,024336	0,35	-0,3541	-0,3531	0,0444
75,89	0,005778	0,35	-0,6556	-0,1720	0,0444
14,49	0,111555	40,95	0,3240	0,7559	-0,4739

Матрица стандартизированных переменных имеет вид:

$$X^* = \begin{pmatrix} 0,2487 & 0,0091 & -0,2518 \\ 0,0226 & -0,3531 & 0,1185 \\ 0,0980 & 0,2580 & -0,0296 \\ 0,1733 & 0,0543 & -0,3258 \\ -0,2788 & -0,1720 & 0,7108 \\ 0,3994 & -0,1720 & -0,1037 \\ 0,0226 & 0,1448 & 0,2666 \\ -0,3542 & -0,3531 & 0,0444 \\ -0,6556 & -0,1720 & 0,0444 \\ 0,3240 & 0,7559 & -0,4739 \end{pmatrix}.$$

2. Определение корреляционной матрицы:

$$r = X^{*T} X^*,$$

где X^{*T} – транспонированная матрица X^* .

Матрица $r = X^{*T} X^*$ симметричная и имеет размер 3×3 :

$$r = \begin{pmatrix} 1 & 0,494 & -0,551 \\ 0,494 & 1 & -0,5168 \\ -0,551 & -0,5168 & 1 \end{pmatrix}.$$

Элементами матрицы являются парные коэффициенты корреляции, числовые значения которых свидетельствуют о тесноте связи между независимыми переменными. Но можно ли утверждать, что эта связь является проявлением мультиколлинеарности? Для ответа на этот вопрос применяются статистические критерии оценки мультиколлинеарности, которые рассматриваются далее.

3. Вычисление детерминанта матрицы $r = X^{*T} X^*$ и критерия χ^2 :

$$D = |r| = 0,466;$$

$$\chi^2 = -\left[n - 1 - \frac{1}{6}(2m + 5) \right] \ln|r| = -\left[9 - \frac{1}{6}(6 + 5) \right] \ln 0,466 = 5,48.$$

При $\frac{1}{2}m(m-1) = \frac{1}{2} \cdot 3 \cdot 2 = 3$ степенях свободы и уровне значимости $(1-\alpha) = 0,95$, $\chi_{\text{табл}}^2 = 7,91$. Так как $\chi^2 < \chi_{\text{табл}}^2$, то для всего массива независимых переменных X мультиколлинеарность не наблюдается.

4. Определение матрицы, обратной к матрице r :

$$C = (X^{*T} X^*)^{-1} = \begin{pmatrix} 1,57 & -0,45 & 0,63 \\ -0,45 & 1,49 & 0,52 \\ 0,63 & 0,52 & 1,62 \end{pmatrix}.$$

5. Вычисление значений статистики F -критерия Фишера:

$$F_k = (c_{kk} - 1) \frac{n - m_1}{m_1 - 1},$$

где c_{kk} – диагональные элементы матрицы $C = (X^{*T} X^*)^{-1}$;

m_1 – число параметров модели, $m_1 = 4$.

$$F_1 = (1,57 - 1) \cdot \frac{6}{3} = 1,14; \quad F_2 = (1,49 - 1) \cdot \frac{6}{3} = 0,98; \quad F_3 = (1,62 - 1) \cdot \frac{6}{3} = 1,24.$$

Для уровня значимости $\alpha = 0,05$ и степеней свободы числителя $(n - m_1) = 6$ и знаменателя $(m_1 - 1) = 3$ по статистическим таблицам (приложение А) находим критическое значение $F_{\text{табл}} = 8,94$.

Поскольку $F_k < F_{\text{табл}} (k = 1, 2, 3)$, то каждая из независимых переменных не мультиколлинеарна с двумя другими.

6. Вычисление частных коэффициентов корреляции.

Частные коэффициенты корреляции вычисляются по элементам матрицы $C = (X^{*T} X^*)^{-1}$:

$$r_{12.3} = \frac{-c_{12}}{\sqrt{c_{11} \cdot c_{22}}} = \frac{0,45}{\sqrt{1,57 \cdot 1,49}} = 0,293;$$

$$r_{13.2} = \frac{-c_{13}}{\sqrt{c_{11} \cdot c_{33}}} = \frac{-0,63}{\sqrt{1,57 \cdot 1,62}} = -0,397;$$

$$r_{23.1} = \frac{-c_{23}}{\sqrt{c_{22} \cdot c_{33}}} = \frac{-0,52}{\sqrt{1,49 \cdot 1,62}} = -0,337.$$

Частные коэффициенты корреляции значительно меньше чем парные. Это еще раз показывает, что частные и парные коэффициенты корреляции не свидетельствуют о наличии или отсутствии мульти-коллинеарности.

7. Определение значений статистики t -критерия Стьюдента:

$$t_{12} = \frac{r_{12.3} \sqrt{n - m_1}}{\sqrt{1 - r_{12.3}^2}} = \frac{0,293 \sqrt{6}}{\sqrt{1 - 0,0858}} = 0,750;$$

$$t_{13} = \frac{r_{13.2} \sqrt{n - m_1}}{\sqrt{1 - r_{13.2}^2}} = \frac{0,39 \sqrt{6}}{\sqrt{1 - 0,1521}} = 1,060;$$

$$t_{23} = \frac{r_{23,1} \sqrt{n - m_1}}{\sqrt{1 - r_{23,1}^2}} = \frac{0,34 \sqrt{6}}{\sqrt{1 - 0,1156}} = 0,896.$$

Табличное значение t -критерия при степенях свободы $(n - m_1) = 6$ и уровне значимости $\alpha = 0,05$ равно $t_{\text{табл}} = 1,943$.

Все числовые значения статистик t -критерия, найденные для каждой пары переменных, меньше чем табличное значение.

Отсюда делаем вывод, что все пары независимых переменных не является мультиколлинеарными.

Пример 6.3. Исследовать наличие мультиколлинеарности, если $n = 72$; $m = 4$; $c_{11} = 2,86$; $c_{22} = 1,07$; $c_{33} = 1,18$; $c_{44} = 2,58$; $c_{12} = -0,45$; $c_{13} = 0,63$; $c_{14} = 0,78$; $c_{23} = 0,52$; $c_{24} = -0,42$; $c_{34} = 0,46$; $\Delta r = 0,3126$.

Решение.

Выполняем алгоритм Феррара – Глобера с пункта 5.

$$F_{1\text{стат}} = |2,86 - 1| \frac{72 - 4}{4 - 1} = 42,16, \quad F_{2\text{стат}} = |1,07 - 1| \frac{68}{3} = 1,58,$$

$$F_{3\text{стат}} = |1,18 - 1| \frac{68}{3} = 4,08, \quad F_{4\text{стат}} = |2,58 - 1| \frac{68}{3} = 35,81.$$

По таблицам приложений А, Б получаем:

$$F_{0,05}(3;68) = 2,76, \quad F_{0,01}(3;68) = 4,13;$$

$F_{1\text{стат}} > F_{0,01}(3;68)$, $F_{4\text{стат}} > F_{0,01}(3;68)$, то есть, переменные x_1 и x_4 мультиколлинеарны с другими.

Вычисляем частные коэффициенты корреляции:

$$r_{12} = \frac{0,45}{\sqrt{2,86 \cdot 1,07}} = 0,26; \quad r_{13} = \frac{-0,63}{\sqrt{2,86 \cdot 1,18}} = -0,34;$$

$$r_{14} = \frac{-0,78}{\sqrt{2,86 \cdot 2,58}} = -0,29; \quad r_{23} = \frac{-0,52}{\sqrt{1,07 \cdot 1,18}} = -0,46;$$

$$r_{24} = \frac{-0,52}{\sqrt{1,07 \cdot 2,58}} = -0,31; \quad r_{34} = \frac{-0,46}{\sqrt{1,18 \cdot 2,58}} = -0,26.$$

Вычисляем значения статистик t -критерия:

$$t_{12} = \frac{0,26\sqrt{72-4}}{\sqrt{1-0,26^2}} = 2,22; \quad t_{13} = \frac{0,34\sqrt{72-4}}{\sqrt{1-(-0,34)^2}} = 2,98;$$

$$t_{14} = \frac{0,29\sqrt{72-4}}{\sqrt{1-(-0,29)^2}} = 2,50; \quad t_{23} = \frac{0,46\sqrt{72-4}}{\sqrt{1-(-0,46)^2}} = 4,27;$$

$$t_{24} = \frac{0,34\sqrt{72-4}}{\sqrt{1-(-0,31)^2}} = 2,98; \quad t_{34} = \frac{0,26\sqrt{72-4}}{\sqrt{1-(-0,26)^2}} = 2,22.$$

Фактические значения статистик t -критерия t_{ij} сравниваем с табличными при $n-m=72-4$ степенях свободы и уровне значимости $\alpha=0,05$: $t_{0,05}(68) = 1,99$.

В представленной задаче все расчетные $t_{ij} > t_{\alpha}(n-m)$, то есть между независимыми переменными существует мультиколлинеарность.

Один из факторов следует исключить. Это нужно сделать с учетом экономического смысла фактора и коэффициентов корреляции между признаком и фактором. Далее необходимо построить новую обратную матрицу и проверить наличие мультиколлинеарности.

6.3. Задачи

Задача 6.1. Вычислить значения статистик F -критерия и t -критерия для определения наличия мультиколлинеарности трех объясняющих переменных, если задана матрица:

$$C = \begin{pmatrix} 1,60 & -0,41 & 0,61 \\ -0,41 & 1,50 & 0,83 \\ 0,61 & 0,81 & 1,47 \end{pmatrix} = r^{-1}.$$

Количество наблюдений $n = 20$.

Задача 6.2. По выборке объема $n = 50$ для X_1, X_2, X_3 построена следующая корреляционная матрица:

$$r = \begin{pmatrix} 1,0 & 0,45 & -0,35 \\ 0,45 & 1,0 & 0,52 \\ -0,35 & 0,52 & 1,0 \end{pmatrix}.$$

Найти и оценить на 5 %-ом уровне статистическую значимость частных коэффициентов корреляции.

При рассмотрении какой регрессии будет иметь место мультиколлинеарность?

Задача 6.3. По данным социологического опроса семи случайно выбранных семей изучалась зависимость накоплений (y) от дохода (x_1), расходов на питание (x_2) и стоимости коммунальных услуг (x_3).

Получена корреляционная матрица:

$$r = \begin{pmatrix} & y & x_1 & x_2 & x_3 \\ y & 1 & & & \\ x_1 & 0,85 & 1 & & \\ x_2 & 0,81 & 0,93 & 1 & \\ x_3 & -0,65 & -0,38 & -0,28 & 1 \end{pmatrix}.$$

Исследовать наличие мультиколлинеарности и проанализировать целесообразность включения каждого фактора в модель.

Задача 6.4. Проведено исследования влияния объемов производства (x_1) и затрат на сырье (x_2) на прибыль 30 участков предприятия (y). Получены парные коэффициенты корреляции (табл. 6.5).

С помощью t -критерия Стьюдента исследовать наличие мультиколлинеарности.

Таблица 6.5

Парные коэффициенты корреляции

Показатели	y	x_1	x_2
y	1	0,75	-0,55
x_1	0,75	1	0,8
x_2	-0,55	0,8	1

Задача 6.5. Имеется матрица парных коэффициентов корреляции (табл. 6.6).

Таблица 6.6

Парные коэффициенты корреляции

Показатели	y	x_1	x_2	x_3
y	1			
x_1	-0,782	1		
x_2	0,451	0,564	1	
x_3	0,842	-0,873	0,303	1

Исследовать наличие мультиколлинеарности и проанализировать целесообразность включения каждого фактора в модель при $n = 75$.

Задача 6.6. С целью исследования влияния факторов X_1 – среднемесячного количества профилактических наладок автоматической линии и X_2 – среднемесячного числа обрывов нити на показатель Y – средне-месячную характеристику качества ткани по данным 37 участков предприятия легкой промышленности были вычислены парные коэффициенты корреляции: $r_{yx_1} = 0,105$, $r_{yx_2} = 0,024$, $r_{x_1x_2} = 0,996$.

Определить частные коэффициенты корреляции, оценить их значимость на 5 %-ом уровне и сделать вывод о наличии мультиколлинеарности.

Задача 6.7. На расходы оборота влияют объем грузооборота, запасы по грузообороту и трудоемкость его единицы. Чтобы построить

эконометрическую модель этой зависимости на основе метода МНК, необходимо быть уверенным, что между факторами грузооборота, запасов и трудоемкости не существует мультиколлинеарности.

Исследовать наличие мультиколлинеарности между этими факторами на основе данных, которые приведены в табл. 6.7.

Таблица 6.7

Исходные данные задачи

№ п/п	Грузооборот	Запасы	Трудоемкость
1	15,6	40,4	2,11
2	13,5	38,9	2,78
3	15,3	36,6	2,17
4	14,9	41,4	2,15
5	15,1	32,2	2,11
6	16,1	31,4	1,97
7	16,7	32,6	1,96
8	15,4	38,7	2,12
9	17,1	44,3	2,02
10	16,8	39,3	2,13

Задача 6.8. Известны данные (усл. ед.) затрат (X) на непродовольственные товары для 30-ти домохозяйств (табл. 6.8).

Определить с помощью МНК оценки параметров уравнения парной регрессии $y_i = b_0 + b_1x_i + e_i$;

исследовать наличие мультиколлинеарности по алгоритму Феррара – Глобера.

Таблица 6.8

Исходные данные

X	26,2	33,1	42,5	47,0	48,5	49,0	49,1	50,9	52,4	53,2
Y	10,0	11,2	15,0	20,5	21,2	19,5	23,0	19,0	19,5	18,0
X	54,0	54,8	59,0	61,3	62,5	63,1	64,0	66,2	70,0	71,5
Y	24,5	21,5	35,4	25,0	17,3	21,6	15,3	32,6	34,0	23,8
X	73,2	75,4	76,0	80,6	81,2	83,3	92,0	95,5	103,2	110,4
Y	22,5	27,4	40,0	23,5	20,0	40,1	15,5	39,0	47,4	21,3

Задача 6.9. Для девяти географических районов имеются данные о потреблении электроэнергии Y в зависимости от уровня доли городского населения X_1 , относительного образовательного уровня X_2 и относительного заработка X_3 (табл. 6.9).

Таблица 6.9

Исходные данные задачи

Номер района	X_1	X_2	X_3	Y
1	42,2	11,2	31,9	167,1
2	48,6	10,6	13,2	174,4
3	42,6	10,6	28,7	160,8
4	39,0	10,4	26,1	162,0
5	34,7	9,3	30,1	140,8
6	44,5	10,7	8,5	174,6
7	39,1	10,7	24,3	163,7
8	40,1	10,0	18,6	174,5
9	45,9	12,0	20,4	185,7

Используя пошаговую процедуру отбора наиболее информативных объясняющих переменных, определить подходящую регрессионную модель, исключив при этом мультиколлинеарность.

Задача 6.10. Исследуется зависимость прибыли (X) предприятия от расходов на сырье (Y).

Выборочные данные 27-ми производственных участков предприятия приведены в табл. 6.10.

Таблица 6.10

Исходные данные

X	25,5	26,5	25,2	29,6	28,7	38,6	39,0	39,3	39,9
Y	14,5	11,3	14,7	10,2	13,5	9,9	12,4	8,6	10,3
X	41,9	42,5	44,2	44,8	45,5	45,5	48,3	49,5	52,3
Y	13,9	14,9	11,6	12,5	10,8	13,8	10,0	11,2	9,1
X	55,7	59,0	61,0	61,7	62,5	64,7	69,7	71,2	73,8
Y	6,3	7,5	5,9	6,1	4,5	4,6	3,0	3,2	2,7

Построить уравнение регрессии $X(Y)$;
 исследовать наличие мультиколлинеарности;
 вычислить отклонения e_i ;
 проанализировать графически отклонения и зависимость дисперсии отклонений от значений Y .

Задача 6.11. По данным табл. 6.11 исследовать независимые переменные на мультиколлинеарность с помощью алгоритма Феррара – Глобера.

Таблица 6.11

Исходные данные

	X_1	X_2	X_3	X_4	Y
X_1	1	0,716	-0,226	0,379	0,853
X_2		1	-0,229	-0,01	0,747
X_3			1	0,181	-0,239
X_4				1	0,157
\bar{X}_i	603,7	45,7	1079,3	188,3	166,6
σ	282,1	14,3	680,3	34,5	78,2

Построить эконометрическую линейную модель методом пошаговой регрессии.

Определить значимость оценок параметров уравнения с помощью t -статистики Стьюдента.

7. Гетероскедастичность и методы ее определения.

Обобщенный метод наименьших квадратов

7.1. Основные теоретические сведения

Одной из предпосылок МНК является условие постоянства дисперсий случайных отклонений. Выполнимость этой предпосылки называется *гомоскедастичностью*, невыполнимость – *гетероскедастичностью*.

Наиболее популярными и наглядными методами определения гетероскедастичности являются:

графический анализ отклонений, при котором по оси абсцисс откладываются значения объясняющей переменной либо линейной комбинации объясняющих переменных, а по оси ординат отклонения, либо их квадраты;

тест ранговой корреляции Спирмена:

- ранжирование (упорядочивание по величинам) значений x_i и ε_i ,
- определение коэффициента ранговой корреляции:

$$r_{x,\varepsilon} = 1 - 6 \cdot \frac{\sum d_i^2}{n(n^2 - 1)},$$

где d_i – разность между рангами x_i и ε_i ;

- вычисление статистики $t = \frac{r_{x,\varepsilon} \sqrt{n-2}}{\sqrt{1-r_{x,\varepsilon}^2}}$ (если наблюдаемое значение

статистики t -критерия превышает $t_{кр} = t_{\frac{\alpha}{2}, n-2}$, то необходимо отклонить

гипотезу о об отсутствии гетероскедастичности, в противном случае гипотеза об отсутствии гетероскедастичности принимается);

тест Парка:

- построение уравнения регрессии $y_i = a + b_1 x_i + \varepsilon_i$,
- определение для каждого наблюдения $\ln \varepsilon_i^2 = \ln (y_i - \hat{y}_i)^2$;
- построение регрессии $\ln \varepsilon_i^2 = \alpha + \beta \ln x_i + v_i$, где $\alpha = \ln \sigma^2$ (для

множественной регрессии эта зависимость строится для каждой объясняющей переменной);

- проверка статистической значимости коэффициента β на основе

статистики t -критерия $t = \frac{\beta}{\sigma_\beta}$ (если коэффициент β статистически значим,

то это означает наличие связи между $\ln \varepsilon_i^2$ и $\ln x_i$, то есть гетероскедастичности в статистических данных);

тест Глейзера:

- рассматривается регрессия абсолютных значений остатков $|\varepsilon_i|$ как некоторая функция от x_i , где x_i является той независимой переменной, которая соответствует изменению дисперсии σ_ε^2 :

$$|\varepsilon_i| = a_0 + a_1 x_i,$$

или

$$|\varepsilon_i| = a_0 + a_1 x_i^{-1},$$

или

$$|\varepsilon_i| = a_0 + a_1 \sqrt{x_i}$$

и т. п.,

- решение об отсутствии гетероскедастичности остатков принимается на основе статистической значимости коэффициентов a_0 и a_1 ;

тест Голдфелда-Квандта:

- упорядоченная выборка объема n делится на три примерно равные части $k, n - 2k, k$ соответственно, оптимальным соотношением между параметрами $n - 2k$ и n считается:

$$\frac{n - 2k}{n} = \frac{4}{15};$$

- для каждой из двух выборок объема k оценивается свое уравнение регрессии и находятся суммы квадратов отклонений $S_1 = \sum_{i=1}^k \varepsilon_i^2$ и

$$S_3 = \sum_{i=n-k+1}^k \varepsilon_i^2 \text{ соответственно;}$$

- при заданной доверительной вероятности $p, \alpha = 1 - p$ по таблицам F -критерия находится граничная точка $F_{\alpha, k-m-1, k-m-1}$, где m – число факторов модели;

- рассчитывается значение $F = \frac{S_3}{S_1}$ (если $F < F_{\alpha, k-m-1, k-m-1}$, то на

уровне значимости α принимается гипотеза об отсутствии гетероскедастичности, в противном случае гипотеза отклоняется).

Для устранения гетероскедастичности при известных значениях $\sigma_{\varepsilon_i} = \sqrt{\sigma_{\varepsilon_i}^2}$ отклонений ε_i рекомендуется применять метод взвешенных наименьших квадратов (МВНК).

При этом следует разделить каждое наблюдаемое значение $y_i = a + b_1 x_i + \varepsilon_i$ на соответствующее ему значение $\sigma_{\varepsilon_i} = \sqrt{\sigma_{\varepsilon_i}^2}$:

$$\frac{y_i}{\sigma_{\varepsilon_i}} = \frac{a}{\sigma_{\varepsilon_i}} + b_1 \frac{x_i}{\sigma_{\varepsilon_i}} + \frac{\varepsilon_i}{\sigma_{\varepsilon_i}}.$$

Положив $\frac{y_i}{\sigma_{\varepsilon_i}} = y_i^*$, $\frac{x_i}{\sigma_{\varepsilon_i}} = x_i^*$, $\frac{\varepsilon_i}{\sigma_{\varepsilon_i}} = v_i$, $\frac{1}{\sigma_{\varepsilon_i}} = z_i$, получим уравнение

регрессии без свободного члена, но с дополнительной объясняющей переменной Z и с преобразованным отклонением v_i :

$$y_i^* = a z_i + b_1 x_i^* + v_i.$$

При этом для всех преобразованных остатков v_i выполняется условие гомоскедастичности.

Эконометрическая модель, которой присуща гетероскедастичность, является *обобщенной линейной моделью*, и для оценивания ее параметров следует воспользоваться *методом Эйткена – обобщенным методом наименьших квадратов (ОМНК)*.

Оценка параметров регрессии по ОМНК осуществляется по формуле:

$$\hat{B} = (X^T S^{-1} X)^{-1} X^T S^{-1} Y,$$

где $S^{-1} = \begin{pmatrix} \lambda_1 & 0 & 0 & 0 & \dots & 0 \\ 0 & \lambda_2 & 0 & 0 & \dots & 0 \\ 0 & 0 & \lambda_3 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & \lambda_n \end{pmatrix};$

$$\lambda_i = \frac{1}{x_{ij}}, \text{ если дисперсия остатков пропорциональна изменениям}$$

объясняющей переменной,

или

$$\lambda_i = \frac{1}{x_{ij}^2}, \text{ если дисперсия остатков пропорциональна изменениям}$$

квадрата объясняющей переменной,

или

$$\lambda_i = \{|\varepsilon_i|\}, \text{ если дисперсия остатков пропорциональна изменениям}$$

квадрата остатков по модулю.

7.2. Примеры

Пример 7.1. По данным табл. 7.1 проверить с надежностью 95 % гипотезу об отсутствии гетероскедастичности с помощью теста ранговой корреляции Спирмена.

Таблица 7.1

Исходные данные задачи

x_i	ε_i
2	0
3	-0,09
4	0,12
5	0,03
6	-0,06

Решение.

Выполним вспомогательные расчеты (табл. 7.2).

Имеем:

$$n = 5, r_{x,\varepsilon} = 1 - 6 \cdot \frac{\sum d_i^2}{n(n^2 - 1)} = 1 - 6 \cdot \frac{16}{5(5^2 - 1)} = 0,2.$$

$$\alpha = \frac{1 - p}{2} = \frac{1 - 0,95}{2} = 0,25.$$

Расчеты величин теста ранговой корреляции Спирмена

x_i	ε_i	$ \varepsilon_i $	d_1	d_2	$d = d_1 - d_2$	d^2
2	0	0	5	5	0	0
3	-0,09	0,09	4	2	2	4
4	0,12	0,12	3	1	2	4
5	0,03	0,03	2	4	-2	4
6	-0,06	0,06	1	3	-2	4
Сумма						16

По таблице приложения В находим $t_{кр} = t_{\frac{\alpha}{2}, n-2} = t_{0,025; 5-2} = 3,182$.

Значение статистики $t = \frac{r_{x,\varepsilon} \sqrt{n-2}}{\sqrt{1-r_{x,\varepsilon}^2}} = \frac{0,2\sqrt{5-2}}{\sqrt{1-0,2^2}} = 0,352 < 3,182$.

Следовательно, принимаем гипотезу об отсутствии гетероскедастичности на уровне значимости 5 %.

Пример 7.2. Имеется регрессионная линейная модель с числом факторов $m=2$, и числом наблюдений $n=30$. Для первых и последних $k=11$ наблюдений суммы квадратов отклонений $S_1=20$ и $S_3=45$ соответственно. С помощью теста Голдфелда – Квандта проверить гипотезу об отсутствии гетероскедастичности с надежностью 95 %.

Решение.

Так как $\alpha=0,05$, то по F -таблицам находим граничную точку $F_{0,05, 8, 8} = 3,44$.

Вычислим фактическое значение статистики критерия:

$$F = \frac{S_3}{S_1} = \frac{45}{20} = 2,25 < 3,44.$$

Следовательно, на уровне значимости 5 % принимается гипотеза об отсутствии гетероскедастичности.

Пример 7.3. По данным табл. 7.3 построена модель множественной линейной регрессии:

$$Y = 204,66 + 0,62 \cdot X_1 - 5,05 \cdot X_2,$$

где Y – цена товара (тыс. грн);

X_1 – расходы на оплату труда (тыс. грн);

X_2 – расходы на усовершенствование процесса производства (тыс. грн).

Провести исследование гомоскедастичности по каждому фактору.

Таблица 7.3

Исходные данные

X_1	161	96	74	165	152	102	110	146	140	168	144	150	108	100	99
X_2	21	28	30	14	25	32	27	25	23	17	21	20	22	29	28
Y	164	122	99	232	174	122	152	174	160	274	178	183	157	115	99

Решение.

Вычислим остатки модели (табл. 7.4).

Таблица 7.4

Расчет остатков модели

X_1	X_2	Y	\hat{Y}	e_i
161	21	164	198,43	-34,43
96	28	122	122,78	-0,78
74	30	99	99,04	-0,04
165	14	232	236,26	-4,26
152	25	174	172,65	1,35
102	32	122	106,3	15,7
110	27	152	136,51	15,49
146	25	174	168,93	5,07
140	23	160	175,31	-15,31
168	17	274	222,97	51,03
144	21	178	187,89	-9,89
150	20	183	196,66	-13,66
108	22	157	160,52	-3,52
100	29	115	120,21	-5,21
99	28	99	124,64	-25,64

Исследуем гомоскедастичность по каждому фактору.

В соответствии с условиями теста Голдфелда – Квандта ранжируем остатки модели множественной регрессии отдельно по фактору X_1 и X_2 . В табл. 7.5 приведены результаты ранжирования по каждому фактору и разделения выборки на три равные части.

Таблица 7.5

Результаты ранжирования

Ранжирование по фактору X_1			Ранжирование по фактору X_2		
X_1	X_2	Y	X_1	X_2	Y
74	30	99	165	14	232
96	28	122	168	17	274
99	28	99	150	20	183
100	29	115	144	21	178
102	32	122	161	21	164
108	22	157	108	22	157
110	27	152	140	23	160
140	23	160	146	25	174
144	21	178	152	25	174
146	25	174	110	27	152
150	20	183	96	28	122
152	25	174	99	28	99
161	21	164	100	29	115
165	14	232	74	30	99
168	17	274	102	32	122

По данным табл. 7.5 для первой и третьей частей каждой ранжированной выборки построены уравнения регрессии и вычислены суммы квадратов их отклонений (предлагается проверить это самостоятельно):

$$Y1_{X_1} = -6,78 + 0,61 \cdot X_1 + 2,08 \cdot X_2, \quad S1_{X_1} = 311,95;$$

$$Y3_{X_1} = -163,97 + 2,84 \cdot X_1 - 4,24 \cdot X_2, \quad S3_{X_1} = 3062,5;$$

$$Y1_{X_2} = 124,42 + 1,48 \cdot X_1 - 8,12 \cdot X_2, \quad S1_{X_2} = 2961,4;$$

$$Y3_{X_2} = -6,78 + 0,61 \cdot X_1 + 2,08 \cdot X_2, \quad S3_{X_2} = 311,95.$$

Вычисляем значения статистики критерия Фишера:

$$F_{X_1} = \frac{S3_{X_1}}{S1_{X_1}} = \frac{3062,5}{311,95} \approx 9,82, \quad F_{X_2} = \frac{S1_{X_2}}{S3_{X_2}} = \frac{2961,4}{311,95} \approx 9,49.$$

Находим табличное значение критерия Фишера: $F_{0,05}(2,2) = 19$.

Имеем, что $F_{X_1} < F_{0,05}$ и $F_{X_2} < F_{0,05}$, то есть условие о гомоскедастичности остатков по переменной X_1 и переменной X_2 не нарушается. Следовательно, нулевая гипотеза об отсутствии гетероскедастичности принимается.

Пример 7.4. В предположении нарушения гипотезы о постоянстве дисперсии необходимо оценить параметры эконометрической модели, которая характеризует зависимость расходов на питание от общих расходов на основе данных, которые приведены в табл. 7.6.

Таблица 7.6

Исходные данные

Номер наблюдения	Расходы на питание, Y	Общие расходы, X
1	2	3
1	2,30	15
2	2,20	15
3	2,08	16
4	2,20	17
5	2,10	17
6	2,32	18
7	2,45	19
8	2,50	20
9	2,20	20
10	2,50	22
11	3,10	64
12	2,40	68
13	2,82	72
14	3,04	80

1	2	3
15	2,70	85
16	3,91	90
17	3,10	95
18	3,99	100

Решение.

Проверим наличие гетероскедастичности для приведенных исходных данных на основе теста Гольдфелда – Квандта.

Упорядочим значение независимой переменной X по возрастанию и отбросим $n - 2k$ значений, которые находятся внутри упорядоченного

$$\text{ряда: } \frac{n-2k}{n} = \frac{18-2k}{18} = \frac{4}{15}; \quad k \approx 7.$$

На основе полученных двух совокупностей наблюдений (от первого к седьмому включительно и от двенадцатого к восемнадцатому значению) с помощью МНК построены (предлагается проверить это самостоятельно) две эконометрические модели:

$$1\text{-я модель: } \hat{Y}_1 = 1,475 + 0,046X; \quad 2\text{-я модель: } \hat{Y}_2 = -0,093 + 0,039X.$$

Остатки и квадраты остатков по этим двум моделям приведены в табл. 7.7.

Таблица 7.7

Расчет остатков моделей

Номер наблюдения	Расходы на питание, Y	Общие расходы, X	\hat{Y}	ε	ε^2
1	2	3	4	5	6
1	2,30	15	2,16	0,14	0,020
2	2,20	15	2,16	0,04	0,002
3	2,08	16	2,20	-0,12	0,015
4	2,20	17	2,25	-0,05	0,002
5	2,10	17	2,25	-0,15	0,022
6	2,32	18	2,29	0,26	0,0007
7	2,45	19	2,34	0,11	0,012
				Σ	0,074

1	2	3	4	5	6
12	2,40	68	2,37	0,13	0,016
13	2,82	72	2,52	0,29	0,085
14	3,04	80	2,68	0,36	0,128
15	2,70	85	2,99	-0,29	0,084
16	3,91	90	3,18	0,76	0,573
17	3,10	95	3,38	-0,28	0,076
18	3,99	100	3,57	0,42	0,178
				Σ	1,14

Найдем значение статистики F -критерия как соотношение остаточных дисперсий:

$$F = \frac{S_{\varepsilon_2}^2}{S_{\varepsilon_1}^2} = \frac{\frac{1,14}{7}}{\frac{0,074}{7}} = 15,41.$$

Сравним значение F с критическим значением F -критерия при $\gamma_1 = 5$ и $\gamma_2 = 5$ степенях свободы и уровне значимости $\alpha = 0,01$ ($F_{0,01}(5;5) = 11$). Поскольку $F > F_{0,01}(5;5)$, то исходные данные имеют гетероскедастичность.

При наличии гетероскедастичности оценку параметров модели выполним методом Эйткена:

$$\hat{B} = (X^T S^{-1} X)^{-1} X^T S^{-1} Y.$$

Запишем матрицу переменных, которые входят в формулу метода Эйткена:

$$X^T = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 15 & 15 & 16 & 17 & 17 & 18 & 19 & 68 & 72 & 80 & 85 & 90 & 95 & 100 \end{pmatrix}.$$

Определим матрицу S^{-1} , пользуясь гипотезой: $\lambda_i = \frac{1}{x_i}$, то есть:

$$S^{-1} = \begin{pmatrix} 0,0667 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0,0667 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0,0625 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0,0588 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0,0588 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0,0555 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0,0526 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,05 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,05 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,0154 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,0156 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,0148 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,0139 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,0125 \end{pmatrix}$$

Определим произведения матриц:

$$X^T S^{-1} = \begin{pmatrix} 0,067 & 0,067 & 0,063 & 0,059 & 0,059 & 0,056 & 0,053 & 0,05 & 0,05 & 0,015 & 0,016 & 0,015 & 0,014 & 0,013 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

$$X^T S^{-1} X = \begin{pmatrix} 0,6672 & 18 \\ 18 & 833 \end{pmatrix}.$$

Найдем обратную матрицу:

$$(X^T S^{-1} X)^{-1} = \begin{pmatrix} 3,5934 & -0,0776 \\ -0,0776 & 0,0029 \end{pmatrix}$$

и вектор:

$$X^T S^{-1} Y = \begin{pmatrix} 1,5998 \\ 48,04 \end{pmatrix}.$$

Вычислим вектор оценок параметров модели:

$$\hat{B} = \begin{pmatrix} 3,5934 & -0,0776 \\ -0,0776 & 0,0029 \end{pmatrix} \cdot \begin{pmatrix} 1,5998 \\ 48,04 \end{pmatrix} = \begin{pmatrix} 2,0187 \\ 0,0141 \end{pmatrix}.$$

Отсюда

$$\hat{b}_0 = 2,0187; \hat{b}_1 = 0,0141.$$

Эконометрическая модель расходов на питание имеет вид:

$$\hat{Y} = 2,0187 + 0,0141X.$$

Остаточная дисперсия $\sigma_{\varepsilon}^2 = 0,083$ (предлагается вычислить самостоятельно) показывает, что расчетные значения расходов на питания очень близкие к фактическим.

7.3. Задачи

Задача 7.1. По 11 наблюдениям построена модель зависимости расходов на питание от доходов $\hat{y} = 2,872 + 0,1789 \cdot x$. В табл. 7.8 приведены величины остатков при каждом значении x_i .

Таблица 7.8

Исходные данные

x_i	57	46	29	35	38	30	45	60	39	50	49
e_i	-2,52	1,2	-0,23	1,97	-2,01	1,02	0,58	0,9	-3,52	1,12	1,49

Проверить остатки на наличие гетероскедастичности с помощью теста ранговой корреляции Спирмена.

Задача 7.2. В табл. 7.9 приведены данные о доходах по акциям (X) и балансовой прибыли (Y) 11 предприятий легкой промышленности.

Таблица 7.9

Исходные данные

X	3	4	5	7	8	10	11	12	15	20	30
Y	12	13	20	19	31	24	41	28	52	55	103

Проверить гипотезу о наличии гетероскедастичности в линейной парной регрессии с помощью теста ранговой корреляции Спирмена при уровне значимости 0,05.

Задача 7.3. В табл. 7.10 приведены данные о производительности труда (X) и выработке продукции (Y) 10 угледобывающих предприятий.

Таблица 7.10

Исходные данные

X	7	8	14	16	19	21	20	32	42	22
Y	7	2	8	17	9	10	12	26	35	15

Проверить гипотезу об отсутствии гетероскедастичности в линейной парной регрессии с помощью теста ранговой корреляции Спирмена при уровне значимости 0,05.

Задача 7.4. В табл. 7.11 приведены данные о прибыли 15 предприятий отрасли – X (млн грн) и объеме дивидендов, оплаченных этими предприятиями за месяц – Y (млн грн).

Таблица 7.11

Исходные данные

X	3	5	8	10	12	14	7	6	9	10	5	7	4	12	15	18
Y	0,2	1,2	4,0	1,5	2,0	3,5	0,8	2,2	1,4	5,0	2,1	1,8	2,3	8	1,6	10

Построить модель парной линейной регрессии.

Проверить с помощью теста Голдфелда–Квандта, выполняется ли условие гомоскедастичности.

Задача 7.5. Проводится анализ средней заработной платы рабочих (З/п) от стажа работы. Статистические данные приведены в табл. 7.12.

Таблица 7.12

Исходные данные

Стаж	5	10	15	20	25	30	35	40
З/п	10 000	12 500	14 300	18 700	25 400	29 000	32 000	34 300

Построить уравнение регрессии, в которой заработная плата является зависимой переменной, а стаж работы – объясняющей переменной; оценить качество построенной регрессии. Есть ли основания считать, что для заданной регрессионной модели вероятна гетероскедастичность? Если да, то почему?

Задача 7.6. Известны данные (усл. ед.) затрат (X) на непродовольственные товары для 30-ти домохозяйств (табл. 7.13).

Таблица 7.13

Исходные данные

X	26,2	33,1	42,5	47,0	48,5	49,0	49,1	50,9	52,4	53,2
Y	10,0	11,2	15,0	20,5	21,2	19,5	23,0	19,0	19,5	18,0
X	54,0	54,8	59,0	61,3	62,5	63,1	64,0	66,2	70,0	71,5
Y	24,5	21,5	35,4	25,0	17,3	21,6	15,3	32,6	34,0	23,8
X	73,2	75,4	76,0	80,6	81,2	83,3	92,0	95,5	103,2	110,4
Y	22,5	27,4	40,0	23,5	20,0	40,1	15,5	39,0	47,4	21,3

Определить с помощью МНК оценки параметров уравнения парной регрессии $y_i = b_0 + b_1x_i + e_i$;

применить тест ранговой корреляции Спирмена;

определить, значительно ли влияет гетероскедастичность на качество оценок в уравнении, которое построено по МНК.

Задача 7.7. Исследуется зависимость прибыли (X) предприятия от расходов на сырье (Y). Выборочные данные 27-ми участков предприятия приведены в табл. 7.14.

Таблица 7.14

Исходные данные

X	25,5	26,5	27,2	29,6	35,7	38,6	39,0	39,3	40,0
Y	14,5	11,3	14,7	10,2	13,5	9,9	12,4	8,6	10,3
X	41,9	42,5	44,2	44,8	45,5	45,5	48,3	49,5	52,3
Y	13,9	14,9	11,6	21,5	10,8	13,8	16,0	18,2	19,1
X	55,7	59,0	61,0	61,7	62,5	64,7	69,7	71,2	73,8
Y	16,3	17,5	10,9	16,1	10,5	10,6	29,0	8,2	14,3

Построить уравнение регрессии $x_i = b_0 + b_1 y_i + e_i$;

проанализировать модель на гетероскедастичность по тесту ранговой корреляции Спирмена, проанализировать графически отклонения и зависимость дисперсии отклонений от значений Y .

Задача 7.8. В табл. 7.15 приведены данные модели по 14 наблюдениям.

Таблица 7.15

Исходные данные

y_i	1	2	3	4	5	6	7	8	9	10	11	12	13	14
x_i	21,3	22,6	32,7	41,9	43,8	49,7	56,9	59,7	67,8	71,5	75,7	76,0	78,9	79,8
e_i^2	2,3	5,6	12,8	10,1	14,6	13,9	24,0	21,9	19,7	23,8	45,7	34,7	56,9	56,8

Проверить гипотезу о гомоскедастичности, используя тест Голдфелда – Квандта.

Задача 7.9. Построить эконометрическую модель, которая характеризует зависимость между затратами на реализацию продукции, объемом товарооборота и средним уровнем товарных запасов. Исходные данные представлены в табл. 7.16.

Необходимо оценить параметры модели по МНК и ОМНК, провести сравнительный анализ оценок параметров, полученных разными методами, сделать выводы.

Таблица 7.16

Исходные данные

№ склада	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Затраты на реализацию продукции, млн грн	300	280	350	340	330	320	310	300	320	280	340	360	320	300
Объем товарооборота, млн т	25	20	30	30	28	28	25	24	27	22	35	30	29	28
Средний уровень товарных запасов, млн т	5	4	6	7	7	5	6	4	5	4	6	7	7	5

Задача 7.10. В табл. 7.17 приведены статистические данные изменения заработной платы (Y), роста производительности труда (x_1) и уровня инфляции (x_2) за 20 месяцев.

Таблица 7.17

Исходные данные

Y	6,0	8,9	9,0	7,1	3,2	6,5	9,1	14,6	11,9	9,4
x_1	2,8	6,3	4,5	3,1	1,5	7,6	6,7	4,2	2,7	3,5
x_2	3,0	3,1	3,8	3,8	1,1	2,3	3,6	7,5	8,0	6,3
Y	12,0	12,5	8,5	5,9	6,8	5,6	4,8	6,7	5,5	4,0
x_1	5,0	2,3	1,5	6,0	2,9	2,8	2,6	0,9	0,6	0,7
x_2	6,1	6,9	7,1	3,1	3,7	3,9	3,9	4,8	4,3	4,8

По методу МНК построить уравнение регрессии

$$y_t = b_0 + b_1x_{1t} + b_2x_{2t} + e_t.$$

Оценить качество построенного уравнения, учитывая проверку наличия гетероскедастичности.

Задача 7.11. В табл. 7.18 приведены статистические данные за 25 лет по темпам прироста заработной платы Y (%), производительности труда X_1 (%), а также уровню инфляции X_2 (%).

Таблица 7.18

Исходные данные

Год	1	2	3	4	5	6	7	8	9	10	11	12	13
X_1	3,5	2,8	6,3	4,5	3,1	1,5	7,6	6,7	4,2	2,7	4,5	3,5	5,0
X_2	4,5	3,0	3,1	3,8	3,8	1,1	2,3	3,6	7,5	8,0	3,9	4,7	6,1
Y	9,0	6,0	8,9	9,0	7,1	3,2	6,5	9,1	14,6	11,9	9,2	8,8	12,0
Год	14	15	16	17	18	19	20	21	22	23	24	25	
X_1	2,3	2,8	1,5	6,0	2,9	2,8	2,6	1,5	0,9	0,6	0,7	3,1	
X_2	6,9	3,5	7,1	3,1	3,7	3,9	4,0	4,8	4,8	4,2	4,9	3,2	
Y	12,5	6,7	8,5	5,9	6,8	5,6	4,8	4,5	6,7	5,5	4,0	3,3	

Оценить параметры уравнения множественной регрессии $Y = b_0 + b_1X_1 + b_2X_2 + \varepsilon$.

Оценить качество построенного уравнения, проводя при этом проверку наличия гетероскедастичности.

8. Автокорреляция остатков модели и методы ее устранения

8.1. Основные теоретические сведения

Важной предпосылкой МНК является независимость значений случайных отклонений ε_i от значений отклонений во всех других наблюдениях.

Отсутствие зависимости гарантирует отсутствие коррелированности между любыми отклонениями $\sigma(\varepsilon_i, \varepsilon_j) = 0$ при $i \neq j$, в противном случае наблюдается явление *автокорреляции*.

Основные методы определения автокорреляции:

графический метод, отображающий связь отклонений ε_i с моментами t их получения (их порядковыми номерами i);

критерий Дарбина – Уотсона:

определяются значения отклонений $\varepsilon_t = y_t - \hat{y}_t$;

рассчитывается статистика:

$$DW = \frac{\sum_{t=2}^T (\varepsilon_t - \varepsilon_{t-1})^2}{\sum_{t=1}^T \varepsilon_t^2};$$

по таблице критических точек Дарбина – Уотсона (приложение Д) при заданном уровне значимости α , числе наблюдений n и количестве объясняющих переменных m определяются нижняя и верхняя границы критерия: d_l и d_u , затем делаются выводы по правилу:

$0 \leq DW \leq d_l$ – существует положительная корреляция;

$d_u \leq DW \leq 4 - d_u$ – автокорреляция отсутствует;

$d_l \leq DW \leq d_u$ или $4 - d_u \leq DW \leq 4 - d_l$ – вывод о наличии автокорреляции сделать нельзя (зона неопределенности).

Статистика Дарбина – Уотсона тесно связана с выборочным коэффициентом корреляции $r_{\varepsilon_t \varepsilon_{t-1}}$:

$$DW \approx 2(1 - r_{\varepsilon_t \varepsilon_{t-1}}).$$

Таким образом, $0 \leq DW \leq 4$ и значения DW могут указать на наличие либо отсутствие автокорреляции.

Если $r_{\varepsilon_t \varepsilon_{t-1}} = 0$ (автокорреляция отсутствует), то $DW \approx 2$.

Если $r_{\varepsilon_t \varepsilon_{t-1}} = 1$ (положительная автокорреляция), то $DW \approx 0$.

Если $r_{\varepsilon_t \varepsilon_{t-1}} = -1$ (отрицательная автокорреляция), то $DW \approx 4$.

Так как автокорреляция чаще всего вызывается неправильной спецификацией модели, то необходимо, прежде всего, скорректировать саму модель.

Для оценивания параметров эконометрической модели, которая имеет автокорреляцию остатков, можно применить ОМНК (метод Эйткена), который базируется на откорректированной исходной информации с учетом ковариации остатков:

$$\hat{B} = (X^T S^{-1} X)^{-1} X^T S^{-1} Y,$$

где $S = \begin{pmatrix} 1 & \rho & \rho^2 & \rho^3 & \rho^4 & \dots & \rho^{n-1} \\ \rho & 1 & \rho & \rho^2 & \rho^3 & \dots & \rho^{n-2} \\ \rho^2 & \rho & 1 & \rho & \rho^2 & \dots & \rho^{n-3} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \rho^{n-4} & \rho^{n-5} & \dots & 1 \end{pmatrix},$

ρ^s – коэффициент автокорреляции s -го порядка для остатков,

$$S^{-1} = \frac{1}{1-\rho^2} \begin{pmatrix} 1 & -\rho & 0 & 0 & 0 & \dots & 0 \\ -\rho & 1+\rho^2 & -\rho & 0 & 0 & \dots & 0 \\ 0 & -\rho & 1+\rho^2 & -\rho & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & 0 & \dots & 1 \end{pmatrix} \quad (\text{при } s > 2 \quad \rho^s$$

приближается к нулю);

$$\rho = \frac{n}{n-1} \cdot \frac{\sum_{t=2}^n \varepsilon_t \varepsilon_{t-1}}{\sum_{t=1}^n \varepsilon_t^2} + \frac{m+1}{n} \quad (\text{при этом осуществляется подбор}$$

соответствующей модели для остатков, начиная с самых простых, которые постепенно усложняются, пока не будет достигнута идентификация);

$\frac{m+1}{n}$ – величина смещения (m – количество независимых переменных).

Существует альтернативный подход к поиску оценок параметров модели с помощью двухшаговой процедуры:

преобразование исходной информации с помощью параметра ρ ;

использование МНК для оценки параметров на основе преобразованных данных.

В качестве матрицы преобразования строится матрица размером $n \times n$:

$$T_1 = \begin{pmatrix} \sqrt{1-\rho^2} & 0 & 0 & 0 & \dots & 0 \\ -\rho & 1 & 0 & 0 & \dots & 0 \\ 0 & -\rho & 1 & 0 & \dots & 0 \\ 0 & 0 & -\rho & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & 1 \end{pmatrix}.$$

Тогда

$$T_1 Y = \begin{pmatrix} \sqrt{1-\rho^2} y_1 \\ y_2 - \rho y_1 \\ y_3 - \rho y_2 \\ y_4 - \rho y_3 \\ \dots \\ y_n - \rho y_{n-1} \end{pmatrix}; \quad T_1 X = \begin{pmatrix} \sqrt{1-\rho^2} & \sqrt{1-\rho^2} x_1^1 & \dots & \sqrt{1-\rho^2} x_1^m \\ 1-\rho & x_2^1 - \rho x_1^1 & \dots & x_2^m - \rho x_1^m \\ \dots & \dots & \dots & \dots \\ 1-\rho & x_n^1 - \rho x_1^1 & \dots & x_n^m - \rho x_1^m \end{pmatrix}.$$

Иногда для преобразования исходной информации используется матрица T_2 , которая получается из матрицы T_1 вычеркиванием первой строки:

$$T_2 = \begin{pmatrix} -\rho & 1 & 0 & 0 & \dots & 0 \\ 0 & -\rho & 1 & 0 & \dots & 0 \\ 0 & 0 & -\rho & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & 1 \end{pmatrix}.$$

В случае, когда невозможно применить ни метод Эйткена, ни метод преобразования исходной информации, в эконометрической литературе предлагаются приближенные методы Кохрана – Оркатта и Дарбина.

8.2. Примеры

Пример 8.1. Пусть объем предложения товара y зависит от цены товара x_1 и зарплаты сотрудников x_2 (табл. 8.1):

$$\hat{y} = 90,74 + 0,88x_1 - 7,32x_2.$$

Исследовать наличие автокорреляции с помощью критерия Дарбина – Уотсона.

Расчетные данные зависимости

y_i	x_1	x_2	\hat{y}_i	ε_i	ε_i^2	$\varepsilon_i - \varepsilon_{t-1}$	$(\varepsilon_i - \varepsilon_{t-1})^2$
20	10	12	11,7	8,3	68,89		
35	15	10	30,74	4,26	18,15	-4,04	16,32
30	20	9	42,46	-12,46	155,25	-16,72	279,56
45	25	9	46,86	-1,86	3,46	10,6	112,36
60	40	8	67,38	-7,38	54,46	-5,52	30,47
70	37	8	64,74	5,26	27,67	12,64	159,77
75	43	6	84,66	-9,66	93,32	-14,92	22,61
90	35	4	92,26	-2,26	5,11	7,4	54,76
105	40	4	96,66	8,34	69,56	10,6	112,36
110	55	5	102,54	7,46	55,65	-0,88	0,77
640	320	75			551,52		988,98

Решение.

Рассчитываем статистику критерия Дарбина – Уотсона:

$$DW = \frac{\sum_{t=2}^T (\varepsilon_t - \varepsilon_{t-1})^2}{\sum_{t=1}^T \varepsilon_t^2} = \frac{988,98}{551,52} = 1,793.$$

По таблице распределения Дарбина – Уотсона (приложение Д) находим нижнюю и верхнюю границы критерия:

$$d_l = 0,697 \text{ и } d_u = 1,641.$$

Тогда

$$4 - d_u = 4 - 1,641 = 2,359.$$

Поскольку $d_u \leq DW \leq 4 - d_u$ ($1,641 < 1,793 < 2,359$), то гипотеза об отсутствии автокорреляции остатков не отклоняется на уровне значимости 0,05. Это является одним из подтверждений высокого качества модели.

Пример 8.2. По данным изменений объемов продажи промышленной продукции Y (табл. 8.2) получено уравнение линейной регрессии $Y = 1,4286 \cdot t + 45,286$, где Y – объем продажи; t – период времени; $R^2 = 0,7519$ – коэффициент детерминации.

Исследовать наличие автокорреляции остатков по критерию Дарбина – Уотсона.

Таблица 8.2

Исходные данные

t	1	2	3	4	5	6	7
Y	46	50	48	53	51	52	57

Решение.

Вычислим теоретические значения результативного признака (предсказанное Y) и остатки модели (рис. 8.1), которые позволят определить фактическое значение статистики критерия Дарбина – Уотсона.

Наблюдение	Предсказанное Y	Остатки $e(t)$	$e(t-1)$	$(e(t-1) - e(t))^2$	$e^2(t)$
1	46,714	-0,714	-	-	0,510
2	48,143	1,857	-0,714	6,612	3,449
3	49,571	-1,571	1,857	11,755	2,469
4	51,000	2,000	-1,571	12,755	4,000
5	52,429	-1,429	2,000	11,755	2,041
6	53,857	-1,857	-1,429	0,184	3,449
7	55,286	1,714	-1,857	12,755	2,939
Сумма				55,816	18,857

Рис. 8.1. Вычисление элементов статистики Дарбина – Уотсона

Имеем:
$$\sum_{t=2}^T (e_t - e_{t-1})^2 = 55,816, \quad \sum_{t=1}^T e_t^2 = 18,857.$$

Следовательно, фактическое значение статистики критерия Дарбина – Уотсона:

$$DW = \frac{\sum_{t=2}^T (e_t - e_{t-1})^2}{\sum_{t=1}^T e_t^2} = \frac{55,816}{18,857} = 2,96.$$

При уровне значимости $\alpha = 0,05$, числе наблюдений $n = 7$ и количестве объясняющих переменных $m = 1$ определяем: $d_l = 0,70$ – нижняя граница и $d_u = 1,37$ – верхняя граница (приложение Д).

Далее необходимо проверить выполнение соответствующих неравенств, а именно:

если $0 \leq DW \leq d_l$, то существует положительная автокорреляция;

если $d_l \leq DW \leq d_u$ или $4 - d_u \leq DW \leq 4 - d_l$ – вывод о наличии автокорреляции нельзя сделать;

если $d_u \leq DW \leq 4 - d_u$ – автокорреляция отсутствует;

если $4 - d_l \leq DW \leq 4$ – существует отрицательная автокорреляция.

В приведенном примере $4 - d_u \leq DW \leq 4 - d_l$. Поэтому вывод о наличии автокорреляции нельзя сделать.

Пример 8.3. По данным табл. 8.3 построить эконометрическую модель зависимости розничного товарооборота от дохода.

Таблица 8.3

Исходные данные

Год	1	2	3	4	5	6	7	8	9	10
Розничный товарооборот, ден. ед.	24,0	25,0	25,7	27,0	28,8	30,8	33,8	38,1	43,4	45,5
Доход, ден. ед.	27,1	28,2	29,3	31,3	34,0	36,0	38,7	43,2	50,0	52,1

Необходимо:

исследовать наличие автокорреляции;

оценить параметры модели методом преобразования исходной информации.

Решение.

Идентифицируем переменные модели:

y_t – розничный товарооборот за период t , зависимая переменная;

x_t – доход за период t , независимая переменная.

Специфицируем эконометрическую модель:

$$y_t = b_0 + b_1 x_t + \varepsilon_t; \quad \hat{y}_t = \hat{b}_0 + \hat{b}_1 x_t + \varepsilon_t; \quad \varepsilon_t = y_t - \hat{y}_t.$$

Определим оценки параметров модели с помощью МНК, считая, что остатки ε_t некоррелированные:

$$\hat{B} = (X^T X)^{-1} X^T Y.$$

$$X = \begin{pmatrix} 1 & 27,1 \\ 1 & 28,2 \\ 1 & 29,3 \\ 1 & 31,3 \\ 1 & 34,0 \\ 1 & 36,0 \\ 1 & 38,7 \\ 1 & 43,7 \\ 1 & 50,0 \\ 1 & 52,1 \end{pmatrix}; \quad (X^T X) = \begin{pmatrix} 10 & 370,4 \\ 370,4 & 14\,441,62 \end{pmatrix};$$

$$(X^T X)^{-1} = \begin{pmatrix} 2,0002 & -0,513 \\ -0,513 & 0,0014 \end{pmatrix}; \quad (X^T Y) = \begin{pmatrix} 322,1 \\ 12\,555,9 \end{pmatrix};$$

$$\hat{B} = \begin{pmatrix} 2,0002 & -0,513 \\ -0,513 & 0,0014 \end{pmatrix} \cdot \begin{pmatrix} 322,1 \\ 12\,555,9 \end{pmatrix} = \begin{pmatrix} 0,172 \\ 0,865 \end{pmatrix};$$

$$\hat{b}_0 = 0,172; \quad \hat{b}_1 = 0,865.$$

Эконометрическая модель имеет вид:

$$\hat{Y}_t = 0,172 + 0,865 X_t.$$

Вычислим остатки модели ε_t (табл. 8.4).

Расчет остатков модели по МНК

Год	y_t	\hat{y}_t	ε_t	ε_t^2	$\varepsilon_t - \varepsilon_{t-1}$	$(\varepsilon_t - \varepsilon_{t-1})^2$	$\varepsilon_t \varepsilon_{t-1}$
1	24,0	23,613	0,388	0,150	-	-	-
2	25,0	24,564	0,436	0,190	0,049	0,0024	0,1691
3	25,7	25,515	0,485	0,034	-0,252	0,0632	0,0806
4	27,0	27,245	-0,245	0,060	-0,430	0,1848	-0,045
5	28,8	29,581	-0,779	0,609	-0,535	0,2866	0,1913
6	30,8	31,310	-0,510	0,261	0,270	0,0729	0,3984
7	33,8	33,646	0,154	0,023	0,665	0,4417	-0,0787
8	38,1	37,971	0,129	0,017	-0,025	0,0006	0,0199
9	43,4	43,420	-0,020	0,0002	-0,149	0,0222	-0,003
10	45,5	42,236	0,264	0,070	0,284	0,0804	-0,005
Σ	322,1	-	-	1,4152	-	1,1550	0,7276

Оценим значение статистики критерия Дарбина – Уотсона:

$$DW = \frac{\sum_{t=2}^n (\varepsilon_t - \varepsilon_{t-1})^2}{\sum_{t=1}^n \varepsilon_t^2} = \frac{1,155}{1,4152} = 0,816.$$

Сравним расчетное значение статистики критерия DW с данными статистических таблиц для уровня значимости $\alpha = 0,05$, числа наблюдений $n = 10$ и числа независимых переменных $m = 1$ (приложение Д).

Нижняя граница критерия $DW_1 = 0,88$, верхняя граница – $DW_2 = 1,32$.

Так как $DW < DW_1$, то можно утверждать, что остатки ε_t имеют положительную автокорреляцию.

Применим метод Эйткена для оценивания параметров эконометрической модели с автокоррелированными остатками:

$$\hat{B} = (X^T S^{-1} X)^{-1} X^T S^{-1} Y,$$

где $S = \begin{pmatrix} 1 & \rho & \rho^2 & \rho^3 & \rho^4 & \rho^5 & \rho^6 & \rho^7 & \rho^8 & \rho^9 \\ \rho & 1 & \rho & \rho^2 & \rho^3 & \rho^4 & \rho^5 & \rho^6 & \rho^7 & \rho^8 \\ \rho^2 & \rho & 1 & \rho & \rho^2 & \rho^3 & \rho^4 & \rho^5 & \rho^6 & \rho^7 \\ \rho^3 & \rho^2 & \rho & 1 & \rho & \rho^2 & \rho^3 & \rho^4 & \rho^5 & \rho^6 \\ \rho^4 & \rho^3 & \rho^2 & \rho & 1 & \rho & \rho^2 & \rho^3 & \rho^4 & \rho^5 \\ \rho^5 & \rho^4 & \rho^3 & \rho^2 & \rho & 1 & \rho & \rho^2 & \rho^3 & \rho^4 \\ \rho^6 & \rho^5 & \rho^4 & \rho^3 & \rho^2 & \rho & 1 & \rho & \rho^2 & \rho^3 \\ \rho^7 & \rho^6 & \rho^5 & \rho^4 & \rho^3 & \rho^2 & \rho & 1 & \rho & \rho^2 \\ \rho^8 & \rho^7 & \rho^6 & \rho^5 & \rho^4 & \rho^3 & \rho^2 & \rho & 1 & \rho \\ \rho^9 & \rho^8 & \rho^7 & \rho^6 & \rho^5 & \rho^4 & \rho^3 & \rho^2 & \rho & 1 \end{pmatrix}$ – матрица

ковариаций остатков.

Чтобы сформировать матрицу S , необходимо определить величину ρ , которая характеризует взаимосвязь между последовательными членами ряда остатков.

Пусть остатки описываются моделью $\varepsilon_t = \rho\varepsilon_{t-1} + e_t$.

Тогда на основании данных табл. 8.4 получаем:

$$\rho \approx r = \frac{n}{n-1} \frac{\sum_{t=2}^n \varepsilon_t \varepsilon_{t-1}}{\sum_{t=1}^n \varepsilon_t^2} + \frac{m+1}{n} = \frac{10}{9} \cdot \frac{0,7276}{1,4156} + \frac{2}{10} = 0,77.$$

Следовательно, матрица S имеет вид:

$$S = \begin{pmatrix} 1 & 0,77 & 0,595 & 0,459 & 0,354 & 0,273 & 0,211 & 0,162 & 0,125 & 0,097 \\ 0,77 & 1 & 0,77 & 0,595 & 0,459 & 0,354 & 0,273 & 0,211 & 0,162 & 0,125 \\ 0,595 & 0,77 & 1 & 0,77 & 0,595 & 0,459 & 0,354 & 0,273 & 0,211 & 0,162 \\ 0,459 & 0,595 & 0,77 & 1 & 0,77 & 0,595 & 0,459 & 0,354 & 0,273 & 0,211 \\ 0,354 & 0,459 & 0,595 & 0,77 & 1 & 0,77 & 0,595 & 0,459 & 0,354 & 0,273 \\ 0,273 & 0,354 & 0,459 & 0,595 & 0,77 & 1 & 0,77 & 0,595 & 0,459 & 0,354 \\ 0,211 & 0,273 & 0,354 & 0,459 & 0,595 & 0,77 & 1 & 0,77 & 0,595 & 0,459 \\ 0,162 & 0,211 & 0,273 & 0,354 & 0,459 & 0,595 & 0,77 & 1 & 0,77 & 0,595 \\ 0,125 & 0,162 & 0,211 & 0,273 & 0,354 & 0,459 & 0,595 & 0,77 & 1 & 0,77 \\ 0,097 & 0,125 & 0,162 & 0,211 & 0,273 & 0,354 & 0,459 & 0,595 & 0,77 & 1 \end{pmatrix}.$$

Тогда:

$$(X^T S^{-1}) = \begin{pmatrix} 0,565 & 0,129 & 0,129 & 0,129 & 0,129 & 0,129 & 0,129 & 0,129 & 0,129 & 0,565 \\ 13,204 & 3,640 & 2,069 & 2,708 & -5,722 & 3,314 & 0,616 & 3,165 & 14,453 & 33,412 \end{pmatrix};$$

$$(X^T S^{-1} X) = \begin{pmatrix} 28,56 & 47,26 \\ 52,53 & 51,13 \end{pmatrix};$$

$$(X^T S^{-1} X)^{-1} = \begin{pmatrix} -0,08 & 0,07 \\ 0,07 & -0,07 \end{pmatrix}; \quad (X^T S^{-1} Y) = \begin{pmatrix} 71,8148 \\ 3\,089,991 \end{pmatrix};$$

$$\hat{A} = \begin{pmatrix} 3,9815 & -0,0924 \\ -0,0924 & 0,0024 \end{pmatrix} \cdot \begin{pmatrix} 71,8148 \\ 3\,089,991 \end{pmatrix} = \begin{pmatrix} 0,442 \\ 0,861 \end{pmatrix};$$

$$\hat{b}_0 = 0,442; \hat{b}_1 = 0,861.$$

Эконометрическая модель имеет вид:

$$\hat{Y}_t = 0,442 + 0,861 X_t.$$

Определим оценку параметров модели по методу Эйткена, воспользовавшись обратной матрицей S^{-1} :

$$S^{-1} = \frac{1}{1-\rho^2} \begin{pmatrix} 1 & -\rho & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -\rho & 1+\rho^2 & -\rho & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -\rho & 1+\rho^2 & -\rho & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -\rho & 1+\rho^2 & -\rho & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -\rho & 1+\rho^2 & -\rho & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -\rho & 1+\rho^2 & -\rho & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -\rho & 1+\rho^2 & -\rho & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -\rho & 1+\rho^2 & -\rho & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\rho & 1+\rho^2 & -\rho \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\rho & 1 \end{pmatrix}.$$

Подставив $\rho = 0,77$, получим:

$$S^{-1} = \begin{pmatrix} 2,469 & -1,904 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1,904 & 3,938 & -1,904 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1,904 & 3,938 & -1,904 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1,904 & 3,938 & -1,904 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1,904 & 3,938 & -1,904 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1,904 & 3,938 & -1,904 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1,904 & 3,938 & -1,904 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1,904 & 3,938 & -1,904 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1,904 & 3,938 & -1,904 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1,904 & 3,938 \end{pmatrix}.$$

Тогда оценки параметров модели:

$$\hat{B} = \begin{pmatrix} 3,9815 & -0,0924 \\ -0,0924 & 0,0024 \end{pmatrix} \cdot \begin{pmatrix} 71,8148 \\ 3089,991 \end{pmatrix} = \begin{pmatrix} 0,442 \\ 0,861 \end{pmatrix},$$

$$\hat{b}_0 = 0,442; \hat{b}_1 = 0,861.$$

Эконометрическая модель имеет вид:

$$\hat{Y}_t = 0,442 + 0,861 X_t.$$

Сравнив полученные эконометрические модели, видим что, при оценивании параметров методом Эйткена проще пользоваться матрицей S^{-1} , когда ковариация остатков отсутствует.

Вычислим расчетные значения \hat{Y}_t на основе построенной эконометрической модели и определим остатки (табл. 8.5).

Вычислим значение статистики критерия Дарбина – Уотсона:

$$DW = \frac{1,1514}{1,6069} = 0,716.$$

Сравним значение критерия DW с данными статистических таблиц для уровня значимости $\alpha = 0,05$, числа наблюдений $n = 10$ и числа независимых переменных $m = 1$ (приложение Д).

Расчет остатков модели по методу Эйткена

Год	Y_t	\hat{Y}_t^*	e_t	e_t^2	$e_t - e_{t-1}$	$(e_t - e_{t-1})^2$	$e_t e_{t-1}$
1	24,0	23,784	0,216	0,0468	-	-	-
2	25,0	24,731	0,269	0,0724	0,0526	0,0028	0,0528
3	25,7	25,678	0,022	0,0005	-0,2774	0,0612	0,0058
4	27,0	27,401	-0,401	0,1608	-0,4226	0,1786	-0,0086
5	28,8	29,727	-0,927	0,8586	-0,5255	0,2762	0,3716
6	30,8	31,449	-0,649	0,4215	0,2774	0,0769	0,6016
7	33,8	33,775	0,025	0,0006	0,6745	0,4549	-0,0164
8	38,1	38,081	0,019	0,0004	-0,0066	0,00004	0,0005
9	43,4	43,508	-0,108	0,0116	-0,1262	0,0159	0,0020
10	45,5	45,316	0,184	0,0937	0,2912	0,0848	0,9908
				1,6069		1,1514	0,9908

Нижняя граница критерия $DW_1 = 0,88$, верхняя граница – $DW_2 = 1,32$. Так как $DW < DW_1$, то можно утверждать, что остатки e_t все еще имеют положительную автокорреляцию.

Исходная гипотеза, о том, что остатки описываются моделью $\varepsilon_t = \rho\varepsilon_t + e_t$, не выполняется.

В этом случае целесообразно выполнить оценку параметров модели другим методом.

Пример 8.4. По данным примера 8.3 оценить параметры эконометрической модели с автокоррелированными остатками методом преобразования исходной информации.

Решение.

Сформируем матрицу T_1 для преобразования исходных данных и преобразуем переменные Y_t и X_t на основе матрицы T_1 :

$$T_1 = \begin{pmatrix} 0,64 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -0,77 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -0,77 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -0,77 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -0,77 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -0,77 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -0,77 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -0,77 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0,77 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0,77 & 1 \end{pmatrix},$$

$$Y^* = T_1 Y_t = \begin{pmatrix} 15,27 \\ 6,49 \\ 6,42 \\ 7,18 \\ 7,97 \\ 8,58 \\ 10,04 \\ 12,03 \\ 14,01 \\ 12,02 \end{pmatrix}, \quad X^* = T_1 X_t = \begin{pmatrix} 0,64 & 17,25 \\ 0,23 & 7,30 \\ 0,23 & 7,55 \\ 0,23 & 8,70 \\ 0,23 & 9,86 \\ 0,23 & 9,77 \\ 0,23 & 10,93 \\ 0,23 & 13,85 \\ 0,23 & 16,29 \\ 0,23 & 13,53 \end{pmatrix}.$$

Для преобразованных данных воспользуемся формулой:

$$\hat{B} = (X^T X)^{-1} X^T Y \rightarrow \hat{B} = (X^{*T} X^*)^{-1} X^{*T} Y^*.$$

Тогда:

$$X^{*T} X^* = \begin{pmatrix} 0,8756 & 33,3368 \\ 33,3368 & 1436,007 \end{pmatrix}, \quad X^{*T} Y^* = \begin{pmatrix} 29,1003 \\ 1251,584 \end{pmatrix},$$

$$(X^{*T} X^*)^{-1} = \begin{pmatrix} 9,8299 & -0,2282 \\ -0,2282 & 0,006 \end{pmatrix},$$

$$\hat{B} = \begin{pmatrix} 3,9815 & -0,0924 \\ -0,0924 & 0,0024 \end{pmatrix} \cdot \begin{pmatrix} 71,8148 \\ 3089,991 \end{pmatrix} = \begin{pmatrix} 0,442 \\ 0,861 \end{pmatrix}.$$

Отсюда $\hat{b}_0 = 0,442$, $\hat{b}_1 = 0,861$ и эконометрическая модель имеет вид:

$$\hat{Y}_t = 0,442 + 0,861X_t.$$

Оценки параметров модели, которые определены с помощью метода преобразования исходной информации, не отличаются от оценок, полученных методом Эйткена при разных матрицах ковариаций остатков. Это значит, что два метода являются альтернативными.

8.3. Задачи

Задача 8.1. Получены данные (табл. 8.6), которые отражают динамику изменений объемов продажи промышленной продукции за период 15-ти месяцев.

Таблица 8.6

Исходные данные

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
y_i	187	331	408	408	395	426	455	506	505	528	544	590	605	629	753

При условии линейности зависимости найти ее уравнение и оценить его значимость на уровне $\alpha = 0,05$.

Исследовать с помощью критерия Дарбина – Уотсона наличие или отсутствие автокорреляции остатков на 5 %-ом уровне значимости.

Задача 8.2. С целью прогнозирования объема экспорта страны на будущие периоды были собраны данные за 30 лет по следующим показателям: Y_t – объемы экспорта (млрд. \$, в сопоставимых ценах); X_t – индекс физического объема промышленного производства (в % к предыдущему году).

Далее представлены результаты предварительной обработки исходных данных.

1. Уравнения линейных трендов:

а) для ряда Y_t

$$Y_t = 3,1 + 1,35t + \varepsilon_t; \quad R^2 = 0,91; \quad DW = 2,31;$$

б) для ряда X_t

$$X_t = -8,4 + 4,8t + \varepsilon_t; \quad R^2 = 0,89; \quad DW = 2,08.$$

2. Уравнение регрессии по уровням временных рядов:

$$Y_t = -10,5 + 0,5x_t + \varepsilon_t; \quad R^2 = 0,95; \quad DW = 2,21.$$

3. Уравнение регрессии по первым разностям уровней временных рядов:

$$\Delta Y_t = 1,4 + 0,03\Delta x_t + \varepsilon_t; \quad R^2 = 0,86; \quad DW = 2,25.$$

4. Уравнение регрессии по вторым разностям уровней временных рядов:

$$\Delta^2 Y_t = 0,7 + 0,012 \Delta^2 x_t + \varepsilon_t; \quad R^2 = 0,47; \quad DW = 2,69.$$

5. Уравнение регрессии по уровням временных рядов с включением фактора времени:

$$Y_t = 4,23 + 0,24 x_t + 0,78t + \varepsilon_t; \quad R^2 = 0,97; \quad DW = 0,9.$$

Сформулировать свои предположения о величине коэффициента автокорреляции первого порядка в каждом из рядов. Обосновать ответ.

Выбрать наилучшее уравнение регрессии, которое можно использовать для прогнозирования объема экспорта, и дать интерпретацию его параметров.

Задача 8.3. Проводится анализ расходов за период 36-ти месяцев построено следующее уравнение регрессии:

$$y_t = 43,5 + 0,251x_t + 0,545y_{t-1}, \quad DW = 1,9,$$

где y_t – расходы на развлечения молодежи в момент времени t ;

x_t – прибыль в момент времени t .

Проверить гипотезу об отсутствии автокорреляции остатков.

Задача 8.4. Для изучения изменений *ВНП* за 10 месяцев необходимо построить регрессионную модель с объясняющими переменными – потреблением (*C*) и инвестициями (*I*) (табл. 8.7).

Таблица 8.7

Исходные данные

<i>C</i> , млрд	8	9,5	11	12	13	14	15	16,5	17	18
<i>I</i> , млрд	1,65	1,8	2,0	2,1	2,2	2,4	2,65	2,85	3,2	3,55
<i>ВНП</i> , млрд	14	16	18	20	23	23,5	25	26,5	28,5	30,5

Оценить коэффициенты уравнения линейной регрессионной модели $ВНП = b_0 + b_1I + b_2C + \varepsilon$.

Определить значение статистики *DW* Дарбина – Уотсона. Имеет ли место автокорреляция остатков?

Задача 8.5. Проведен анализ средних годовых расходов (*Y*) студентов на развлечения. По статистическим данным за 22 года построено уравнение регрессии:

$$y_t = 41,2 + 0,254x_t + 0,539y_{t-1}; \quad R^2 = 0,783, \quad DW = 1,86,$$

(*S*) (0,107) (0,133)

где *X* – прибыль студента после выплаты за учебу и общежитие.

Оценить качество построенной модели. Насколько увеличатся расходы на развлечения при увеличении прибыли на единицу?

Проверить гипотезу об отсутствии автокорреляции остатков при альтернативной гипотезе о положительной автокорреляции с уровнем значимости $\alpha = 0,01$.

Задача 8.6. В табл. 8.8 приведены статистические данные изменения заработной платы (*Y*), роста производительности труда (*x₁*) и уровня инфляции (*x₂*) за 20 месяцев. По методу МНК построить уравнение регрессии $y_t = b_0 + b_1x_{1t} + b_2x_{2t} + e_t$. Оценить качество построенного уравнения, учитывая проверку наличия автокорреляции.

Таблица 8.8

Исходные данные

Y	6,0	8,9	9,0	7,1	3,2	6,5	9,1	14,6	11,9	9,4
x_1	2,8	6,3	4,5	3,1	1,5	7,6	6,7	4,2	2,7	3,5
x_2	3,0	3,1	3,8	3,8	1,1	2,3	3,6	7,5	8,0	6,3
Y	12,0	12,5	8,5	5,9	6,8	5,6	4,8	6,7	5,5	4,0
x_1	5,0	2,3	1,5	6,0	2,9	2,8	2,6	0,9	0,6	0,7
x_2	6,1	6,9	7,1	3,1	3,7	3,9	3,9	4,8	4,3	4,8

Задача 8.7. В табл. 8.9 приведены статистические данные за 25 лет о темпах роста заработной платы Y (%), производительности труда X_1 (%), а также уровне инфляции X_2 (%).

Таблица 8.9

Исходные данные

Год	1	2	3	4	5	6	7	8	9	10	11	12	13
X_1	3,5	2,8	6,3	4,5	3,1	1,5	7,6	6,7	4,2	2,7	4,5	3,5	5,0
X_2	4,5	3,0	3,1	3,8	3,8	1,1	2,3	3,6	7,5	8,0	3,9	4,7	6,1
Y	9,0	6,0	8,9	9,0	7,1	3,2	6,5	9,1	14,6	11,9	9,2	8,8	12,0
Год	14	15	16	17	18	19	20	21	22	23	24	25	
X_1	2,3	2,8	1,5	6,0	2,9	2,8	2,6	1,5	0,9	0,6	0,7	3,1	
X_2	6,9	3,5	7,1	3,1	3,7	3,9	4,0	4,8	4,8	4,2	4,9	3,2	
Y	12,5	6,7	8,5	5,9	6,8	5,6	4,8	4,5	6,7	5,5	4,0	3,3	

Оценить параметры уравнения множественной регрессии $Y = b_0 + b_1X_1 + b_2X_2 + \varepsilon$.

Оценить качество построенного уравнения, проводя при этом проверку наличия автокорреляции.

Задача 8.8. Проводится анализ зависимости между инфляцией (INF) и безработицей (U). Используются статистические данные за 12 лет (табл. 8.10).

Построить линейную парную регрессию, вычислить статистику DW Дарбина – Уотсона и на ее основе проверить наличие автокорреляции.

Таблица 8.10

Исходные данные

<i>INF</i>	3,07	0,70	4,08	2,20	2,38	0,90	1,10	5,12	0,93	2,54	1,55	3,45
<i>U</i>	3,69	9,10	3,92	6,50	4,63	8,50	9,55	3,71	5,80	3,60	6,53	4,32

Задача 8.9. Вычисленные остатки модели (ε_t) зависимости прибыли предприятия (y) от стоимости на сырье (x_1) и производительности труда (x_2) $y = f(x_1, x_2)$ приведены в табл. 8.11.

Таблица 8.11

Исходные данные

<i>t</i>	1	2	3	4	5	6
ε_t	12,0	21,2	-6,7	10,1	1,5	12,6

Вычислить статистику критерия Дарбина – Уотсона; оценить полученный результат при 5 % уровне значимости.

Задача 8.10. Проведен анализ зависимости себестоимости продукции (y) от стоимости сырья и уровня автоматизации производства. Получена модель, результаты которой приведены в табл. 8.12.

Таблица 8.12

Исходные данные

№	1	2	3	4	5	6	7	8
$Y_{\text{эмп}}$	220	215	250	305	352	390	356	360
$Y_{\text{теор}}$	209	210	258	309	360	380	350	335

Проверить гипотезу о наличии автокорреляции, используя критерий Дарбина – Уотсона.

Задача 8.11. По результатам 50 наблюдений 3-х объясняющих переменных статистика DW принимает следующие значения:

- а) 0,91; б) 1,37; в) 2,34; г) 3,01; д) 3,72.

Не обращаясь к таблице критических точек Дарбина – Уотсона, сделать выводы о наличии автокорреляции.

Проверить свои выводы с помощью таблицы критических точек Дарбина – Уотсона для $\alpha = 0,05$ и $\alpha = 0,01$:

определить значения статистики DW , которые дают возможность отклонить гипотезу о наличии автокорреляции в зависимости от объема выборки n и количества объясняющих переменных m :

- а) $n = 20, m = 1$; б) $n = 25, m = 2$; в) $n = 50, m = 1$;
 г) $n = 50, m = 4$; д) $n = 100, m = 2$.

Задача 8.12. На основе данных табл. 8.13 построить эконометрическую модель занятости населения. Исследовать остатки модели на наличие автокорреляции.

Таблица 8.13

Исходные данные

Год	Производство	Занятость	Экспорт
1-й	28	11	25
2-й	30	14	26
3-й	37	16	26
4-й	35	16	24
5-й	36	17	25
6-й	38	15	28
7-й	40	18	30
8-й	45	20	32
9-й	46	21	35
10-й	50	24	38

Задача 8.13. С уровнем значимости 0,05 исследовать наличие автокорреляции остатков по данным табл. 8.14.

Таблица 8.14

Исходные данные

t	1	2	3	4	5	6	7	8
y	213	171	291	309	317	362	351	361
\hat{y}	207	232,7	258,4	284	309,7	335,4	361,1	386,6