

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«САМАРСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ ИМЕНИ АКАДЕМИКА С. П. КОРОЛЕВА»
(САМАРСКИЙ УНИВЕРСИТЕТ)

А. Ю. ТРУСОВА, А. И. ИЛЬИНА

ЭКОНОМЕТРИЧЕСКОЕ МОДЕЛИРОВАНИЕ

Рекомендовано редакционно-издательским советом федерального государственного автономного образовательного учреждения высшего образования «Самарский национальный исследовательский университет имени академика С. П. Королева» в качестве практикума для обучающихся по основной образовательной программе высшего образования по направлению подготовки 38.03.05 Бизнес-информатика

САМАРА

Издательство Самарского университета
2025

УДК 330.43(075)

ББК Ув631я7

Т789

Рецензенты: канд. техн. наук, доц. З. Ф. Камальдинова,
канд. экон. наук, доц. Ю. И. Ряжева

Трусова, Алла Юрьевна

Т789 **Эконометрическое моделирование:** практикум / *А. Ю. Трусова, А. И. Ильина.* – Самара: Издательство Самарского университета, 2025. – 96 с.

ISBN 978-5-7883-2265-0

Практикум рассчитан на обучающихся по программе бакалавриата. Призван помочь студентам лучше усвоить содержание учебного материала и разобраться в основах эконометрического моделирования. Практикум соответствует требованиям государственного образовательного стандарта высшего профессионального образования по указанному направлению.

Подготовлен на кафедре математики и бизнес-информатики.

УДК 330.43(075)

ББК Ув631я7

ISBN 978-5-7883-2265-0

© Самарский университет, 2025

ОГЛАВЛЕНИЕ

Введение	4
1. Расчет коэффициентов корреляции и ковариации	8
2. Парная линейная регрессия по исходным данным лабораторной работы №1	11
3. Проверка условий Гаусса-Маркова	16
4. Выявление эндогенных и экзогенных переменных в исходном массиве данных	24
5. Графический анализ и расчёт параметров парной линейной модели	34
6. Отбор экзогенных переменных в модель множественной линейной регрессии	39
7. Проверка на гетероскедастичность	47
8. Регрессионные модели с фиктивными переменными	62
9. Проверка условия Гаусса Маркова на автокорреляцию	76
10. Множественная линейная регрессия	85
Библиографический список	95

Введение

Эконометрическое моделирование в настоящий период позволяет решать разноплановые и разномасштабные проблемы. Инструмент эконометрического подхода в экономике находит широкое применение в анализе данных в социально-экономической сфере. Цифровая составляющая эконометрического моделирования постоянно обновляется: появляются новые функции, пакеты. Обучающимся, как оказывается, не так просто ориентироваться в функциональных особенностях программных продуктов. Для этого важно иметь хорошую теоретическую базу, которая отрабатывается в процессе лабораторных занятий. Данное пособие позволяет сформировать высокий уровень освоения дисциплин «Эконометрическое моделирование» и «Эконометрика».

Математическое описание инструментария эконометрического моделирования

Термин «эконометрика» включает в себя два понятия – «экономика» и «метрика», которые характеризуют данное понятие и позволяют дать его определение. Эконометрика – это область науки, занимающейся статистическими, экономическими и математическими данными, также – это предмет, помогающий определить количественную меру взаимосвязям экономических процессов.

Описание всевозможных конкретных видов зависимостей опирается на использование регрессионного анализа. Для этого необходимо осуществить подбор класса функций, который бы связывал результативный показатель Y и аргументы $X_1, X_2, X_3, \dots, X_k$, однако следует отбирать те аргументы, которые являются наиболее информативными.

Функция, выраженная формулой, называется уравнением регрессии:

$$f(x) = M_Y(X), \quad (1)$$

где X – это вектор заданных наиболее информативных значений аргументов.

Для построения моделей и нахождения неизвестного уравнения регрессии используется Метод наименьших квадратов. Данный метод минимизирует квадрат отклонения наблюдаемых значений эндогенного показателя y_i от модельных значений, которые вычисляются с помощью формулы $\tilde{y}_i = (x_i \beta)$, где β – это вектор коэффициентов уравнения регрессии, x_i – это значение аргумента в i наблюдении. Минимизируется выражение (2)

$$\sum_{i=1}^n (y_i - f(x_i, \beta))^2 \rightarrow \min. \quad (2)$$

В регрессионном анализе модель представляется формулой (3):

$$y = \sum_{i=1}^k \beta_i * \varphi(x_i) + \varepsilon \quad (3)$$

где $\varphi(x_j)$ – некоторая функция от $X_1, X_2, \dots, X_k$, ε – случайная величина. Самым простым видом регрессии является парная линейная регрессия. В ходе регрессионного анализа необходимо найти оценку параметров уравнения, представленного формулой (4).

$$\tilde{y} = \beta_0 + \beta_1 x, \quad (4)$$

где β_0 и β_1 – неизвестные параметры генеральной совокупности, подлежащих оценке по выборочным исследователем данным.

С помощью полученных значений оценок коэффициентов строится оценка уравнения регрессии, в виде формулы (23).

$$\tilde{y} = b_0 + b_1 x_i. \quad (5)$$

Оценка b_0 является случайно величиной и подчиняется нормальному закону распределения с параметрами $(\beta_0; \frac{\sigma^2}{n})$ [34]. Для определения значимости простейшего линейного уравнения регрессии $\tilde{y} = \beta_0 + \beta_1 x$ на уровне значимости α определяется значимостью коэффициента регрессии β_1 , то есть выдвигаются две гипотезы: $H_0: \beta_1 = 0$, $H_1: \beta_1 \neq 0$.

Для проверки гипотезы необходимо сравнить $t_{\text{набл}}$ и $t_{\text{кр}}$, которые находят по формулам $t_{\text{набл}} = \frac{b_1}{s_{b_1}}$, (6)

$$\text{где } \widehat{S_{b_1}} = \frac{\sqrt{\frac{n}{n-1}} S}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}}, t_{кр} (\alpha; n - 2). \quad (7)$$

Значение данной величины определяется по таблице распределения Стьюдента. Сравнивая полученные значения делается вывод. Если $|t_{набл}| > t_{кр}$ – верна гипотеза $H_1: \beta_1 \neq 0$, то есть уравнение регрессии значимо, а если $|t_{набл}| < t_{кр}$ – верна гипотеза $H_0: \beta_1 = 0$, т.е. уравнение регрессии не значимо.

Множественное линейное уравнение регрессии имеет вид, выраженный формулой (34):

$$\tilde{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_k x_k. \quad (8)$$

Для оценки неизвестных параметров β_i уравнения регрессии берется случайная выборка объемом n из $(k + 1)$ – мерной случайной величины $(y, x_1, x_2, x_3, \dots, x_k)$. Результат i -го наблюдения $y_i, x_{i1}, x_{i2}, x_{i3}, \dots, x_{ik}$.

Методом наименьших квадратов получим оценки неизвестных параметров β_i . Они определяются из условия минимизации величины Q по формуле (36):

$$Q = (Y - X\beta)^T (Y - X\beta) = \sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^k x_{ij} \beta_j)^2.$$

Матрицу b , имеет вид: $b = (X^T X)^{-1} X^T Y$.

Для проверки значимости уравнения регрессии используется F-критерий Фишера дисперсионного анализа. Предполагается, что случайная величина ε подчиняется нормальному закону распределения с параметрами $(\mu; \sigma^2 E_n)$ [37].

Таким образом при выполнении условия $\beta = 0$, т.е. когда $\beta_0 = \beta_1 = \beta_2 =$

$$= \beta_3 = \dots = \beta_k = 0, \text{ выражения } \frac{Q_{осм}}{n - m - 1} \text{ и } \frac{Q_R}{m + 1} \text{ являются не-}$$

зависимыми оценками одной и той же дисперсии σ^2 , следовательно для проверки гипотез H_0 и H_1 используется F-статистика Фишера-

Снедекора. Так выдвигаются две гипотезы H_0 и H_1 . Затем находятся значения $F_{\text{набл}}$ и $F_{\text{кр}}$. Последнее значение $F_{\text{кр}} = (α; m + 1; n - m - 1)$ находится по таблице, в зависимости от степеней свободы и количества наблюдений.

Сравнивая значения, делается вывод. Если $F_{\text{набл}} > F_{\text{кр}}$ – гипотеза H_0 отвергается, принимается гипотеза H_1 , т.е. $β \neq 0$. В противном случае принимается противоположная гипотеза [37].

Аналогично простой модели проводится проверка на значимость коэффициентов уравнения. Выдвигается две гипотезы, касаемо коэффициентов, H_0 и H_1 . С помощью формулы (39) составляется t-статистика Стьюдента.

$$D[b_j] = \left[(X^T X)^{-1} \right]_{jj}^{1/2}. \quad (9)$$

Также как и в простой модели вычисляются t-значения. Значение $t_{\text{кр}} (α; n - k - 1)$ вычисляется по таблице распределения Стьюдента. $t_{\text{набл}}$ по формуле $b_j / S(b_j)$. Сравнивая величины, делается вывод значимости коэффициента. Так, если $t_{\text{набл}} > t_{\text{кр}}$ – гипотеза H_0 отвергается, принимается гипотеза H_1 , $β_j \neq 0$, т.е. коэффициент $β_j$ значим в уравнении регрессии.

Множественный регрессионный анализ позволяет разделять влияние независимых переменных, оставляя возможность коррелированности [38]. Коэффициент регрессии $β_j$, соответствующий переменной X_j дает оценку ее влияния на величину y в случае неизменности влияния не нее всех остальных переменных (в единицах измерения X_j).

1. Расчет коэффициентов корреляции и ковариации

Лабораторная работа №1.

Цель работы: Изучить практическую реализацию математического инструментария расчета коэффициентов корреляции и ковариации.

Выполнение работы.

1. Сформировать исходный массив данных (табл.1).

Таблица 1. Исходные данные для расчета коэффициентов корреляции и ковариации

X	1	2	3	4	5
Y	19	16	19	14	23
X*Y	19	32		56	115

X	6	7	8	9	10
Y	22	27	33	28	29
X*Y	132	189	264	252	290

Вычислить среднее значение для X, Y и X*Y. Расчетные значения представлены в таблице 2.

Таблица 2. Средние значения

среднее значение	X	Y	X*Y
	5,5	23	140,6

Вычислить ковариацию несколькими способами: по формуле $(X*Y)_{\text{ср}} - X_{\text{ср}}*Y_{\text{ср}}$. Ответ $K_{xy}=14,1$

С помощью встроенных функций для смещенной ковариации $=\text{КОВАРИАЦИЯ.В}(\text{массив X}; \text{массив Y})= 15,666$

С помощью коррекции $(n-1)/n$, где $n=10$ для смещенной ковариации $K_{xy}=9/10*15,666=14,1$

С помощью встроенных функций для несмещенной ковариации =КОВАРИАЦИЯ.Г(массив X; массив Y)=14,1

2. Расчет коэффициента корреляции. По формуле $r = K_{xy} / (S_x * S_y)$

С использованием встроенных функций Excel. Предварительно необходимо рассчитать среднее квадратическое отклонение для X Sx и Sy для Y. Функции и расчетные значения представлены в таблице 3.

Таблица 3. Значения среднего квадратического отклонения

=СТАНДОТК ЛОН.В(массив X)	=СТАНДОТК ЛОН.В(массив Y)	=СТАНДОТК ЛОН.Г(массив X)	=СТАНДОТ КЛОН.Г(массив Y)
Sx	Sy	Sx	Sy
3,027	6,146	2,872	5,830

Значение коэффициента корреляции $r = 0,841$.

3. Проверка на значимость коэффициента корреляции.

H_0 статистически не значим в генеральной совокупности (связь отсутствует).

H_1 статистически значим в генеральной совокупности (связь присутствует)

Расчетное значение наблюдаемой статистики $t_{\text{набл}} = \frac{|r| \sqrt{n-2}}{\sqrt{1-r^2}}$,

$t_{\text{набл}} = 4,412$. Критическое значение статистики равно $t_{\text{кр}} = 2,306$ при использовании функции =СТЮДЕНТ.ОБР.2X(альфа; к), $\alpha = 0,05$, $k = 10 - 2 = 8$. Вывод: верна гипотеза H_1 , коэффициент корреляции статистически значим.

4. Интервальная оценка.

Необходимо рассчитать величину Z- преобразования Фишера для коэффициента корреляции =ФИШЕР(0,841) = 1,227. Далее значение $t = 1,959$ с помощью функции =НОРМ.СТ.ОБР(0,5+(1- α)/2). Далее рассчитать Корень(n-3)= 2,645. Далее рассчитать

$\Delta = t / \sqrt{n-2} = 0,740$. Следующим действием рассчитывается левая граница Z-преобразования Фишера. Z- Δ левая и правая граница Z+ Δ преобразования Фишера. Z- $\Delta = 0,486$, Z+ $\Delta = 1,968$. И последнее действие. Необходимо используя обратное преобразование Фишера, вычислить левую и правую границы генерального коэффициента корреляции. $=\text{ФИШЕРОБР}(0,486)=0,451$; $=\text{ФИШПРОБР}(1,968)=0,961$. Получаем доверительный интервал для коэффициента корреляции $[0,451; 0,961]$.

2. Парная линейная регрессия по исходным данным лабораторной работы №1

Лабораторная работа №2.

Цель работы: Изучить практическую реализацию математического инструментария модели парной линейной регрессии.

Выполнение работы.

Используем результаты лабораторной работы №1. Исходные данные из лабораторной работы №1.

Для переменных X и Y необходимо вычислить дисперсии.

В Excel используются две функции для оценки дисперсии. Для смещенной и несмещенной дисперсий.

=ДИСП.В(массив данных) – смещенная дисперсия

=ДИСП.Г(массив данных) – несмещенная дисперсия

В таблице 4 представлены результаты расчетов дисперсии для переменных X и Y .

Таблица 4. Расчетные значения дисперсии для X и Y

Переменная	X	Y
ДИСП.Г	9,166	37,777
ДИСП.В	8,25	34

1. Проверка гипотез о статистической значимости параметров модели регрессии.

$H_0: b_j = 0$

$H_1: b_j \neq 0$

Рассчитанные значения параметров модели и их статистика представлены в таблице 5.

Таблица 5. Рассчитанные значения параметров модели и их статистика

Обозначение	Название	b_1	b_0
b_j	Параметры модели	1,709	13,6
S_{bj}	стандартная ошибка	0,387	2,403
b_j/S_{bj}	t набл	4,412	5,658
	t крит	2,306	
α	Уровень значимости альфа	0,05	
$k=n-2$	Степени свободы	8	

Вывод $t \text{ набл} > t \text{ крит}$, оба параметра модели статистически значимы.

2. Интервальная оценка параметров модели регрессии b_0 и b_1 .

Таблица 6. Рассчитанные значения параметров модели

доверительная вероятность $\gamma = 0,95$		t крит	2,306
		b_1	b_0
S_{bj}	стандартная ошибка	0,387	2,403
$\Delta = t_{кр} * S_{bj}$	предельная ошибка	0,893	5,542
интервальная оценка для b_j :		b_1	b_0
		1,709	13,6
$b_j - \Delta$	левая граница	0,815	8,057
$b_j + \Delta$	правая граница	2,602	19,142

Аналогичные расчеты позволяют провести интервальную оценку параметров модели. В таблице 7 представлены для доверительной вероятности 0,9 и 0,99. Рекомендуется получить их самостоятельно.

Пусть $\gamma = 0,9$, $t_{\text{крит}} = 1,859$. Доверительный интервал для $b_1 = 1,709$ имеет вид $0,815 < b_1 < 2,602$, для $b_0 = 13,6$ имеет вид $8,057 < b_0 < 19,142$

Для $\gamma = 0,99$ $t_{\text{крит}} = 3,355$ получить самостоятельно.

3. Интервальная оценка для среднего значения Y при $x=6,5$.

Таблица 7. Результаты расчетов интервальной оценки для среднего значения

Прогнозное значение X	X_0	6,5
Прогнозное значение Y	Y_0	24,709
Стандартная ошибка	S	3,518

$$S_{\hat{y}_0}^2 = S^2 \left(1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right)$$

Провести расчеты самостоятельно.

Рассмотрим следующую задачу.

Имеются следующие данные об уровне механизации работ $X(\%)$ и производительности труда Y (т/ч) для 14 однотипных предприятий

Таблица 8. Исходные данные

№	x_i	y_i	№	x_i	y_i
1	32	20	8	54	37
2	30	24	9	60	38
3	36	28	10	55	40
4	40	30	11	61	41
5	41	31	12	67	43
6	47	33	13	69	45
7	56	34	14	76	48

Вычислить коэффициенты корреляции и детерминации, используя встроенные функции MS Excel. Результаты в таблице.

Таблица 9. Коэффициент парной корреляции

0,938	КВПИРСОН (Функция MS Excel)
0,968	КОРРЕЛ
0,938	Коэффициент детерминации = КОРРЕЛ^2

Расчет оценок параметров парной линейной модели $Y = v_0 + v_1 \cdot X$

Таблица 10. Расчет оценки параметра модели v_1

$Y = v_0 + v_1 \cdot X$			
v_1	равно ковариация/дисперсия(x)		
Дисперсия(x)	207,142	ДИСП.В	несмещенная
Дисперсия(x)	192,346	ДИСП.Г	смещенная
Ковариация(xу)	112,582	КОВАРИАЦИЯ.Г	
Ковариация(xу)	104,540	КОВАРИАЦИЯ.В	
Параметр модели $v_1 =$	0,543	0,543	0,543
Функция MS Excel	НАКЛОН	ковариация Г/ ДИСП.В	ковариация В/ ДИСП.Г

Таблица 11. Расчет v_1 через коэффициент корреляции

расчет	v_1 через коэффициент корреляции	
	0,968	КОРРЕЛ
Стандартное отклонение		
	Sx(кореньД(x))	14,392
	Sy (стандотклонВ)	8,075
	v_1	0,543

Таблица 12. Расчет оценки параметра модели b_0

$b_0 = \bar{Y} - b_1 * \bar{X}$		
$\bar{X}$	51,714	
$\bar{Y}$	35,142	
b_0	7,036	расчет по формуле
b_0	7,036	ОТРЕЗОК (Функция MS Excel)

Вид парной линейной модели $Y = 7,036 + 0,5435 * X$

Самостоятельно построить регрессию X по Y и провести все интервальные оценки.

3. Проверка условий Гаусса-Маркова

Лабораторная работа № 3.

Цель работы: Расчет остатков. Проверка математического ожидания и дисперсии остатков. Расчет дисперсий оценок параметров линейной регрессии.

Выполнение работы.

Имеются данные о количестве внесенных удобрений (X , кг/га) и урожайности пшеницы Y (ц/га) по 10 фермерским хозяйствам.

Таблица 13. Исходные данные

x_i	y_i	x_i	y_i	x_i	y_i
1	19	5	23	8	33
2	16	6	22	9	28
3	19	7	27	10	29
4	14				

Корреляционное поле изучаемых показателей представлено на рисунке 1

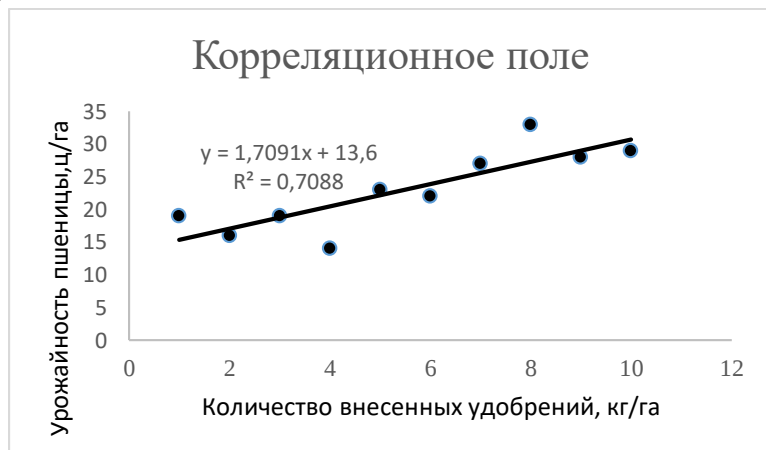


Рис. 1. Корреляционное поле изучаемых показателей

Из графика видно, что показатели могут быть описаны моделью парной линейной регрессии.

В таблице 15 представлена информация об оценках параметров модели, остатках и квадратах остатков. В качестве оценок параметров модели берутся значения $b_1=1,709$ $b_0=13,6$.

Таблица 14. Остатки и квадраты остатков

x_i	y_i^3	y_i^T	$e_i=y_i^3-y_i^T$	$e_i=(y_i^3-y_i^T)^2$
1	19	15,309	3,690	13,622
2	16	17,018	-1,018	1,036
3	19	18,727	0,272	0,074
4	14	20,436	-6,436	41,426
5	23	22,145	0,854	0,730
6	22	23,854	-1,854	3,439
7	27	25,563	1,436	2,063
8	33	27,272	5,727	32,801
9	28	28,981	-0,981	0,963
10	29	30,690	-1,690	2,859
		$\sum(e_i)$	-0,001	99,018
		$M(e_i)$	-0,001	12,377

$\sum(e_i)^2$ = Сумма квадратов остатков 99,018. $M(e_i)$ – математическое ожидание остатков.

На рисунке 2 представлен график остатков, которые позволяют оценить случайную компоненту

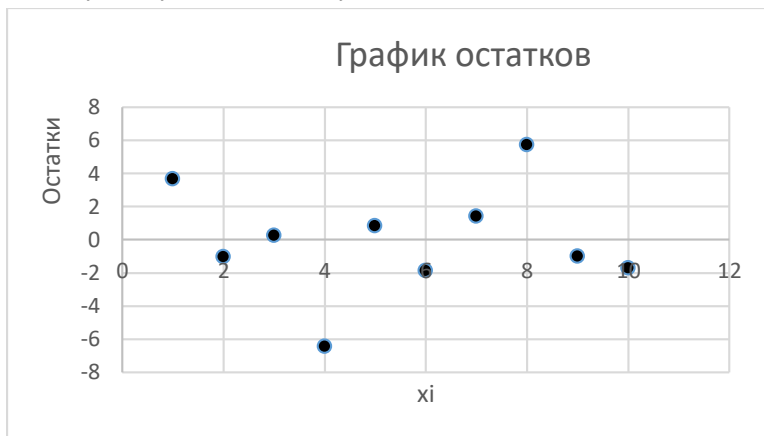


Рис. 2. График остатков

Видно, что случайная компонента не имеет тенденции с x_i .

На рисунке 3 представлен график квадратов остатков, которые позволяют оценить дисперсию случайной компоненты и обнаружить гетероскедастичность.

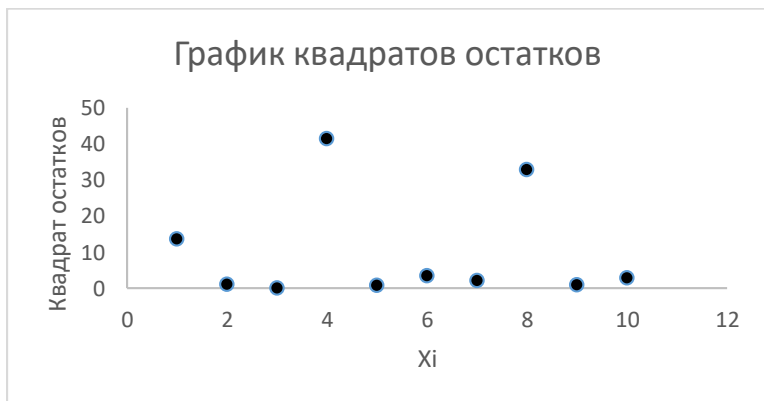


Рис. 3. График квадратов остатков

Из рисунка 3 видно, что 7 значений лежат в полосе от 0 до 5, а три значения можно охарактеризовать как выбросы. Однако, считаем из графического анализа, что данные гомоскедастичные.

Корреляция и ковариация остатков и остатков, сдвинутых на лаг 1 позволяют проверить условие независимости остатков (3 условие Гаусса-Маркова).

Таблица 15. Корреляция и ковариация остатков и остатков, сдвинутых на лаг1

ϵ_i	ϵ_{i-1}	ϵ_i	ϵ_{i-1}
3,690	-1,018	-1,854	1,436
-1,018	0,272	1,436	5,727
0,272	-6,436	5,727	-0,981
-6,436	0,854	-0,981	-1,690
0,854	-1,854	-1,690	

Коэффициент корреляции равен -0,118. Коэффициент ковариации равен -1,322.

Вывод: малая величина коэффициентов позволяет принять условие Гаусса-Маркова о независимости остатков.

Проверка наличия степени тесноты взаимосвязи между X_i и остатками.

Таблица 16. Ковариация и корреляция между X_i и остатками

X_i	ϵ_i	X_i	ϵ_i
1	3,690	6	-1,854
2	-1,018	7	1,436
3	0,272	8	5,727
4	-6,436	9	-0,981
5	0,854	10	-1,690

Коэффициент корреляции равен 0,001.

Вывод. Установлено, что X_i и остатки независимы.

Самостоятельно проверить условия Гаусса-Маркова для данной в случае для модели X от Y.

Изучим дисперсии оценок параметров модели регрессии (парной). В таблице 18 представлены результаты расчетов теоретических значений эндогенной переменной. В качестве параметров модели использовались значения $v_1=1,709$ и $v_0=13,6$.

Таблица 17. Результаты расчетов теоретических значений эндогенной переменной

x_i	y_i^3	y_i^T	x_i	y_i^3	y_i^T
1	19	15,309	6	22	23,854
2	16	17,018	7	27	25,563
3	19	18,727	8	33	27,272
4	14	20,436	9	28	28,981
5	23	22,145	10	29	30,690

Вычислим несмещенную оценку остаточной дисперсии и исправленное выборочное стандартное отклонение остатков регрессии Se .

Таблица 18. Значения квадратов остатков и сумма

$e_i^2=(y_i^3-y_i^T)^2$	$e_i^2=(y_i^3-y_i^T)^2$
13,622	3,439
1,036	2,063
0,074	32,801
41,426	0,963
0,730	2,859
$\sum e_i^2=99,018$	

В результате получим при $n=10$ значение $Se^2=12,377$ и $Se=3,518$.

Вычислим стандартные ошибки эмпирических коэффициентов регрессии $S(b_1)$

$$\sigma_{b_1}^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2 \sigma^2}{(\sum_{i=1}^n (x_i - \bar{x})^2)^2} = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}.$$

Таблица 19. Вспомогательная таблица для расчетов

xi	yi эмп	yi теор	(xi-хср)^2
1	19	15,309	20,25
2	16	17,018	12,25
3	19	18,727	6,25
4	14	20,436	2,25
5	23	22,145	0,25
6	22	23,854	0,25
7	27	25,563	2,25
8	33	27,272	6,25
9	28	28,981	12,25
10	29	30,690	20,25

В результате среднее значение показателя $X=5,5$ Дисперсия для оценки параметра модели $b_1=0,150$. Значение суммы $\sum((x_i - x_{ср})^2)=82,5$ Стандартная ошибка $b_1=0,387$.

Вычислим стандартные ошибки эмпирических коэффициентов регрессии $S(b_0)$.

$$Sb_0^2 = Sb_1^2 \cdot (X^2)_{ср}$$

Дисперсия $b_0=5,776$ Стандартная ошибка $b_0=2,403$.

Результаты проверки на статистическую значимость параметров модели на уровне значимости $\alpha=0,05$ и степени свободы $k=n-2=8$ следующие. Для параметров b_1 и b_0 табл составили соответственно 4,412 и 5,658. Критическое значение статистики Стьюдента равно 2,306. Оба параметра статистически значимы.

Интервальная оценка параметров модели Y по X .

Будем использовать доверительные вероятности γ равными 0,95; 0,9 и 0,99. Соответствующие значения t крит 2,306; 1,859 и 3,355. Стандартные ошибки $Sb_1=0,387$ и $Sb_0=2,403$.

Таблица 20. Вспомогательные расчеты для интервальной оценки параметров модели регрессии b_0 и b_1

γ	0,95	0,9	0,99
Предельная ошибка $\Delta=t_{кр}*Sb_1$	0,893	0,720	1,299
Предельная ошибка $\Delta=t_{кр}*Sb_0$	5,542	4,469	8,06

Интервальная оценка для $b_1=1,709$ при $\gamma=0,95$: (0,815; 2,602).

Интервальная оценка для $b_1=1,709$ при $\gamma=0,9$: (0,988; 2,429).

Интервальную оценку для $b_1=1,709$ при $\gamma=0,99$ рассчитать самостоятельно.

Интервальная оценка для $b_0=13,6$ при $\gamma=0,95$: (8,057; 19,142).

Интервальная оценка для $b_0=13,6$ при $\gamma=0,9$: (9,13; 18,069).

Интервальную оценку для $b_0=13,6$ при $\gamma=0,99$ рассчитать самостоятельно.

Проведем интервальную оценку для значения $X_0=6,5$ (внутреннее интерполирование). Значение $Y_0=24,709$. Стандартная ошибка $Se=3,518$. Среднее значение $X=5,5$. Необходимо вычислить $\sum(x_i - x_{cp})^2$, Вычислить $(x_0 - x_{cp})^2$. В результате $S^2(Y_0)= 11,732$ и $S(Y_0)= 3,710$. Предельная ошибка $\Delta=t_{кр}* S(Y_0)= 6,899$, $t_{крит}=3,334$ для $\gamma=0,99$. Тогда доверительный интервал имеет вид (17,809; 31,608). Для доверительной вероятности 0,95 вычислить самостоятельно. При расчете использовать формулу.

$$S_{y_0}^2 = S^2 \left(1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right),$$

Проведем интервальную оценку для $X_p=10,5$ (внешнее интерполирование). Используя формулу, получим доверительный интервал (19,97;43,12).

$$S_y^2 = S^2 \left(\frac{1}{n} + \frac{(x - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right),$$

Рассмотрим показатели качества уравнения регрессии. Важным является дисперсия остатков Q_e (случайностей), не объясненная часть дисперсии. В качестве исходных данных используем $Se=3,518$. $Se^2=12,377$. $b_1= 1,709$; $b_0=13,6$.

Таблица 21. Расчеты значения дисперсия, обусловленной регрессией Q_R

$(y^3 - y_{\text{мод}})^2$	y_i^3	$y_{\text{мод}}$	$(y_{\text{мод}} - y_{\text{ср}})^2$	$(y_i^3 - y_{\text{ср}})^2$
13,622	19	15,309	59,150	16
1,036	16	17,018	35,782	49
0,074	19	18,727	18,256	16
41,426	14	20,436	6,572	81
0,730	23	22,145	0,730	0
3,439	22	23,854	0,730	1
2,063	27	25,563	6,572	16
32,801	33	27,272	18,256	100
0,963	28	28,981	35,782	25
2,859	29	30,690	59,150	36

$Q_e = 99,018$, $Q_R = 240,981$ и $Q_{\text{общ}} = 340$.

4. Выявление эндогенных и экзогенных переменных в исходном массиве данных

Лабораторная работа № 4.

Цель работы: Используя корреляционный анализ, выявить степень тесноты взаимосвязи.

Для достижения цели воспользуемся инструментом корреляционного анализа. В корреляционном анализе предлагается коэффициент множественной корреляции, который вычисляется по формуле.

$$R(x_j) = \sqrt{1 - \frac{\det(R)}{R_{jj}}},$$

где $R(x_j)$ – коэффициент множественной корреляции x_j показателя со всеми остальными показателями, участвующими в исследовании;

$\det(R)$ - определитель матрицы корреляции, рассчитывается с помощью функции Excel МОПРЕД;

R_{jj} - алгебраическое дополнение x_j в матрице корреляции. Для этого используется функция МОПРЕД. Предварительно из матрицы корреляции необходимо удалить j -строчку и j -столбец (удаление элементов матрицы проводить со сдвигом).

Выполнение работы.

Таблица 22. Исходный массив данных

X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈
515,8	97,3	85,2	86,9	455,4	106,7	65,5	93,1
469,6	118,9	29,4	43,0	323,1	190,7	24,3	109,7
652,9	95,2	73,5	59,2	583,9	111,7	73,2	100,5
1037,4	102,9	187,7	87,3	724,8	121,6	106,7	126,5

273,8	83,0	44,6	98,4	219,4	110,1	34,2	127, 2
615,0	92,7	73,7	103,7	500,7	94,5	73,7	143, 7
212,3	106,1	32,4	123,6	168,1	117,1	21,3	88,4
398,3	121,2	87,1	246,2	261,1	112,5	42,3	125, 6
327,9	80,0	43,5	103,7	291,5	87,9	39,0	100, 9
6 906,5	102,1	1 103,1	140,6	5 907,0	124,7	906,6	131, 0
183,2	95,5	21,5	59,5	132,6	125,9	21,4	143, 7
309,6	69,0	43,3	62,1	219,2	90,6	43,3	140, 6
309,5	102,2	39,7	85,8	234,5	118,8	33,1	109, 4
242,1	96,0	24,9	75,5	210,7	120,4	23,2	118, 4
588,4	140,8	119,0	254,3	431,5	142,1	72,4	171, 6
568,9	100,8	68,3	74,7	431,8	121,0	68,2	91,7
554,6	94,0	63,9	123,0	349,4	115,6	53,0	167, 4
2 243,6	63,7	507,9	228,6	442,4	86,9	70,6	134, 5
190,4	101,4	26,4	122,9	151,5	122,6	18,0	126, 0
122,6	101,2	12,9	128,9	82,5	95,7	12,9	129, 8

Продолжение таблицы 22

296,3	114,4	36,5	113,1	145,8	127,5	20,9	158,6
19,6	141,0	4,6	171,4	11,4	81,9	1,9	69,7
276,7	112,9	32,0	107,9	134,4	133,8	19,0	181,2
431,3	107,2	65,8	130,5	299,9	115,6	40,8	114,3
624,1	100,3	61,7	53,7	387,0	109,6	51,5	88,4
2 585,1	107,8	405,5	142,8	2 008,0	126,0	262,7	122,0
50,9	153,3	7,5	265,5	45,0	135,3	6,5	231,5
199,8	83,5	29,2	62,5	157,8	92,8	24,4	103,9
219,8	120,6	26,0	131,6	171,0	133,7	22,6	132,3
1 613,9	72,5	215,9	48,1	99,3	75,6	13,6	86,8
345,4	89,3	40,2	104,9	255,6	101,9	38,6	100,8
61,2	82,0	10,9	88,1	56,7	99,4	10,3	118,5
837,3	113,3	101,7	104,5	689,7	126,0	96,4	141,1
3 739,4	91,9	640,5	107,3	2 574,3	104,8	359,3	101,0
373,9	102,1	56,0	120,6	337,0	104,7	56,0	120,9
644,6	134,5	97,1	139,7	378,6	125,2	65,1	129,1

1 480,9	95,9	174,4	76,7	1 232,1	114,0	149,8	102, 8
160,9	65,1	29,4	141,4	143,2	62,4	29,4	165, 1
2 260,5	435,4	167,3	167,5	1 949,9	742,4	78,0	162, 4
127,0	156,4	32,3	153,3	84,0	132,3	12,5	119, 5
349,5	133,8	64,9	122,5	271,6	142,5	38,0	126, 8
223,4	84,0	31,9	71,3	200,8	107,3	27,6	115, 0
379,9	116,3	31,1	98,0	183,0	139,3	20,1	100, 0
1 319,5	134,2	131,4	64,2	1 243,1	135,5	131,4	64,9
1 273,9	139,3	191,7	153,6	925,3	131,9	124,2	135, 8
1 622,0	92,1	241,2	127,0	1 360,5	115,7	228,1	152, 1
274,9	94,3	43,8	110,0	189,2	120,8	24,8	119, 4
170,9	91,5	16,9	58,9	142,4	127,9	16,1	79,1
2 587,5	103,6	167,4	95,8	1 880,6	138,6	49,8	54,1
863,9	91,5	130,2	87,7	533,5	97,1	89,0	110, 7
450,2	95,7	37,7	111,0	220,5	106,1	29,0	111, 3

Продолжение таблицы 22

945,7	89,2	173,5	129,6	683,4	85,9	108,0	106, 3
306,2	105,3	45,2	143,3	240,3	104,1	33,6	106, 4
1 388,2	109,4	161,7	96,3	970,2	112,9	138,5	125, 1
758,1	144,2	118,3	141,0	616,6	150,0	81,9	126, 1
544,3	116,4	62,1	93,4	298,6	114,2	38,3	85,6
1 097,5	120,3	131,3	126,4	837,4	124,6	100,6	99,5
426,2	94,7	69,7	103,7	363,1	98,0	52,8	131, 6
328,4	84,8	46,0	69,4	234,4	96,9	28,0	114, 0
234,9	118,2	40,9	100,9	171,6	125,2	20,9	104, 9
1 969,6	110,1	255,9	85,3	1 302,4	122,2	129,2	84,5
2 088,1	114,3	199,6	62,9	1 083,3	127,1	145,2	99,1
381,3	91,9	36,4	61,9	200,0	122,2	28,1	130, 1
190,5	105,4	9,6	85,2	79,7	161,7	8,3	146, 6
1 516,3	123,2	153,7	62,1	803,6	125,7	108,7	91,2
1 180,2	92,0	148,4	108,2	856,2	121,0	121,0	130, 6

Продолжение таблицы 22

160,1	130,5	21,6	132,4	146,5	130,0	21,5	157, 3
195,1	152,2	38,1	305,2	174,6	184,0	38,1	305, 2
252,1	115,2	29,4	118,5	157,4	112,3	22,6	134, 9
625,3	113,0	65,2	107,8	433,6	118,0	64,8	107, 9
757,7	96,0	114,3	105,7	507,8	113,6	62,1	91,1
978,9	110,9	124,6	102,4	839,4	130,4	101,3	114, 3
442,5	122,8	57,7	116,1	332,0	126,5	44,7	117, 5
1 546,9	86,7	223,3	155,2	705,2	131,2	114,0	154, 6
523,9	118,6	82,6	110,4	433,1	138,3	57,7	108, 3
261,9	88,5	38,5	125,3	150,7	102,1	28,3	173, 9
415,2	136,1	45,4	135,2	278,6	163,4	45,4	148, 4
452,8	141,9	84,3	130,6	291,6	166,3	49,2	129, 9
299,3	137,4	53,8	218,7	247,7	147,8	30,2	168, 4
45,7	94,9	5,4	61,6	44,8	124,1	4,8	128, 3
803,1	117,4	94,1	159,0	442,5	121,1	53,1	125, 8
266,7	102,0	35,1	96,2	149,4	100,3	14,9	89,5

275,5	109,6	61,6	217,9	146,5	105,9	18,2	94,6
4,9	142,8	2,7	318,6	3,0	87,4	0,8	96,1
169,4	98,6	24,9	114,7	137,3	123,4	19,5	109,6
23,1	151,8	2,2	101,9	23,0	150,6	2,0	94,0
3,6	120,6	0,3	46,6	0,4	121,3	0,0	0,0

Открываем пакет анализа данных.

Рисунок 4 – исходный пакет анализа данных

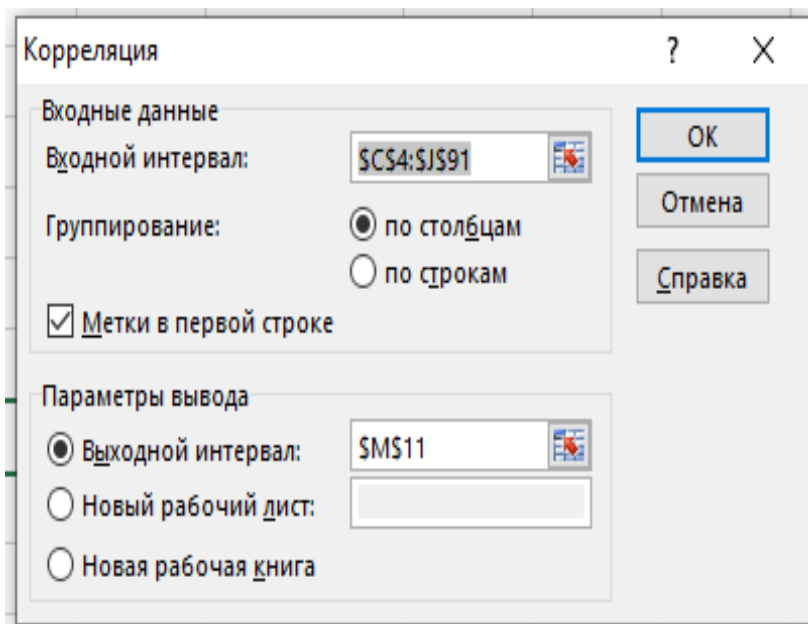


Рис. 4. Пакет анализа данных

В результате получаем матрицу корреляции (Таблица 34).

Таблица 23. Матрица корреляции

	x1	x2	x3	x4	x5	x6	x7	x8
x1	1							
x2	0,076	1						
x3	0,957	-0,04	1					
x4	-0,01	0,316	0,075	1				
x5	0,957	0,144	0,903	-0,01	1			
x6	0,161	0,918	0,027	0,12 1	0,211	1		
x7	0,907	- 0,029	0,923	0,00 6	0,95	0,01 5	1	
x8	- 0,075	0,18	- 0,021	0,52 5	- 0,052	0,20 3	0,00 6	1

Далее необходимо транспонировать значения матрицы корреляции, чтобы матрица корреляций была симметрична. Учитывая округления до двух знаков после запятой, необходимо получить матрицу корреляции, представленную в таблице 35.

Таблица 24. Полная Матрица корреляции

	x1	x2	x3	x4	x5	x6	x7	x8
1	1	0,08	0,96	-0,01	0,96	0,16	0,91	-0,08
2	0,08	1	-0,04	0,32	0,14	0,92	-0,03	0,19
3	0,96	-0,04	1	0,08	0,90	0,03	0,92	-0,02

4	-0,01	0,32	0,08	1	-0,01	0,12	0,01	0,53
5	0,96	0,14	0,90	-0,01	1	0,21	0,95	-0,05
6	0,16	0,92	0,03	0,12	0,21	1	0,02	0,20
7	0,91	-0,03	0,92	0,01	0,95	0,02	1	0,01
8	-0,08	0,19	-0,02	0,53	-0,05	0,20	0,01	1

Матрица корреляции размерности $m \times m$ (8×8).

Используя формулу для расчёта коэффициента множественной корреляции, рассчитываются численные значения коэффициентов множественной корреляции для всех x_j , $j=1; m$. таблица 36.

Таблица 25. Алгебраические дополнения и коэффициент множественной корреляции

Величина алгебраического дополнения R_{jj}	Величина коэффициента множественной корреляции $R(X_j)$
$R_{11}=0,001$	$R(X_1)=0,991$
$R_{22}=0,001$	$R(X_2)=0,962$
$R_{33}=0,001$	$R(X_3)=0,987$
$R_{44}=0,001$	$R(X_4)=0,832$
$R_{55}=0,001$	$R(X_5)=0,992$
$R_{66}=0,001$	$R(X_6)=0,964$
$R_{77}=0,001$	$R(X_7)=0,988$
$R_{88}=0,001$	$R(X_8)=0,744$

Все коэффициенты множественной корреляции высокие (больше чем 0,7)

Вывод: таким образом, любой показатель X может быть эндогенной переменной. Необходимо провести проверку на значимость полученных значений коэффициентов множественной корреляции. Гипотезы для проверки.

H_0 : коэффициент множественной корреляции статистически не значим

H_1 : коэффициент множественной корреляции статистически значим

Используется F – статистика Фишера-Снедекора, представленная соотношением.

$$F_{\text{набл}} = \frac{R_{jj}^2(n-m)}{(1-R_{j0}^2)(m-1)}$$

Таблица 26. Значение $F_{\text{набл}}$

	$F_{\text{набл}}$		$F_{\text{набл}}$
$R(X_1)$	0,001	$R(X_5)$	0,001
$R(X_2)$	0,001	$R(X_6)$	0,001
$R(X_3)$	0,001	$R(X_7)$	0,001
$R(X_4)$	0,001	$R(X_8)$	0,001

$$F_{\text{крит}}(\alpha; k_1; k_2), k_1 = m-1, k_2 = n - m$$

$F_{\text{крит}}$ рассчитывается с помощью функции Excel F.ОБР находим обратную функцию распределения Фишера, учитываем, что $m=8$, $n=87$. В результате получаем значение 0,305.

Вывод: по результатам лабораторной работы №1 можно утверждать, так как все $F_{\text{набл}}$ больше, чем $F_{\text{крит}}$, то все переменные могут быть как экзогенными, так и эндогенными.

5. Графический анализ и расчёт параметров парной линейной модели

Лабораторная работа № 5.

Массив содержит $m=8$, $n=47$ (где m — число независимых переменных, n — объем выборки).

Цель работы: выявить оптимальные модели.

Выполнение работы.

Задание 1. Сформировать массив данных

Построить графическую зависимость X_3 от X_6 и X_6 от X_3 . График называется корреляционным полем. (рисунки 5 и 6).



Рис. 5. Графическая зависимость X_3 от X_6

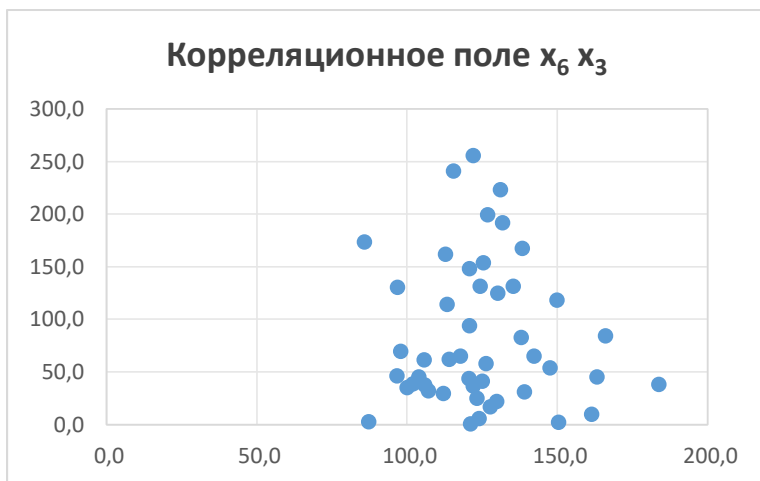


Рис. 6. Графическая зависимость X_6 от X_3

Задание 2. Расчёт параметров модели.

Необходимо оценить среднее значение экзогенной переменной $\bar{x}$, среднее значение эндогенной переменной $\bar{y}$, дисперсию $D(x)$ $D(y)$, стандартное отклонение S_x S_y , коэффициент корреляции r , коэффициент ковариации b_{yx} b_{xy} (таблица 36 и 37).

$$K_{xy} = \overline{xy} - \bar{x}\bar{y}; \quad r = \frac{K_{xy}}{S_x S_y}; \quad b_{yx} = \frac{K_{xy}}{S_x^2}; \quad b_{xy} = \frac{K_{xy}}{S_y^2}$$

$$x - \bar{x} = b_{xy}(y - \bar{y}); \quad y - \bar{y} = b_{yx}(x - \bar{x}); \quad y = b_{yx}x + \bar{y} - b_{yx}\bar{x}$$

$$y = b_0 + b_{yx}x$$

Таблица 27. Результаты расчетов для парной линейной модели 1 (X_3 от X_6)

$\bar{x}$	83,953	S_x	68,192
$\bar{y}$	124,526	S_y	20,617
$\overline{x \cdot y}$	10380,896	r	-0,053
K_{xy}	-73,469	b_{yx}	-0,016
$D(x)$	4650,114	b_{xy}	-0,173
$D(y)$	425,058	b_0	125,881

Модель1 имеет вид:

$$X_6 = b_0 + b_{yx}X_3 \quad X_6=125,881+(-0,016 \cdot X_3)$$

Таблица 28. Результаты расчетов для парной линейной модели 2 (X_6 от X_3)

$\bar{x}$	124,526	S_x	20,617
$\bar{y}$	83,953	S_y	68,192
$x \cdot y$	10380,896	r	-0,053
K_{xy}	-73,469	b_{yx}	-0,177
$D(x)$	425,058	b_{xy}	-0,016
$D(y)$	4650,114	b_0	105,945

Модель 2 имеет вид:

$$X_3 = b_0 + b_{yx}X_6 \quad X_3=105,945+(-0,177 \cdot X_6)$$

С помощью функции ЛИНЕЙН пакета Excel находим вывод дополнительной регрессионной статистики для модели 1 и модели 2 (таблица 29 и 30).

Таблица 29. Вывод регрессионной статистики для модели 1, используя функцию ЛИНЕЙН.

b_{yx}	-0,016	125,881	b_0
$S_{b_{yx}}$	0,045	4,847	S_{b_0}
R^2	0,003	20,815	S_e
$F_{\text{набл}}$	0,129	45	K
Q_R	55,741	19496,944	$Q_{\text{ост}}$

Таблица 30. Вывод регрессионной статистики для модели 2, используя функцию ЛИНЕЙН.

b_{yx}	-0,176	105,945	b_0
$S_{b_{yx}}$	0,492	62,128	S_{b_0}
R^2	0,003	68,847	S_e
$F_{\text{набл}}$	0,128	45	K
Q_R	609,807	213295,424	$Q_{\text{ост}}$

$F_{\text{крит}}$ рассчитывается с помощью функций Excel ФРАСПРОБОР или F.ОБР. Находим значение обратной функции распределения Фишера с двумя степенями свободы.

$$F_{\text{крит}}(\alpha; k_1; k_2), k_1 = m-1, k_2 = n - m,$$

В нашем случае $k_1=7, k_2=39$.

Например, $F_{\text{крит}}$ рассчитывается с помощью функции Excel F.ОБР. В результате получаем значения критической статистики для разных уровней значимости (таблица 31).

Таблица 31. Обратная функция распределения Фишера для первой модели

α	$F_{\text{крит}}$	α	$F_{\text{крит}}$
0,05	0,299	0,15	0,470
0,1	0,394	0,2	0,538

Вывод: $F_{\text{набл}}$ меньше $F_{\text{кр}}$, следовательно, нет достаточных оснований утверждать, что экзогенная переменная статистически значима.

Задание 3. Оценка прогнозных значений.

Расчёт прогнозных значений для линейной модели $y = f(x)$, где Y — показатель X_6 (эндогенная переменная), а X_3 — экзогенная переменная.

Проведём оценку прогнозного уровня экзогенной переменной (модель экстраполирования). Для этого рассчитываем величину $X_{\text{макс}}$ и $X_{\text{мин}}$ для экзогенной переменной используя функцию Excel МАКС и МИН в результате получаем значения, представленные в таблице 32.

Используя функцию Excel ТЕНДЕНЦИЯ рассчитываем прогнозные значения для величины экзогенной переменной при увеличении $X_{\text{макс}}$ экзогенной переменной на 20, 15, 10, 5 и 2%. В результате получаем прогнозные значения, представленные в таблице 32.

Таблица 32. Максимальные, минимальные и прогнозные значения.

		20%	15%	10%	5%	2%
$X_{\text{макс}}$	255,9	307,0	294,24 3	281,45 0	268,65 7	260,981
$X_{\text{мин}}$	0,34	0,410	0,3933	0,3762	0,3591	0,34884
Тенден ция		120,9	121,13 1	121,33 7	121,54 4	121,668

Проведём расчёт прогнозного значения эндогенной переменной в случае внутреннего интерполирования. Рассчитаем $X_{\text{прогн}}$ из условия

$$(X_{\text{макс}} - X_{\text{мин}})/2; (X_{\text{макс}} - X_{\text{мин}})/4; (X_{\text{макс}} - X_{\text{мин}})/8; (X_{\text{макс}} - X_{\text{мин}})/3;$$

$$(X_{\text{макс}} - X_{\text{мин}})/5.$$

В результате получаем прогнозные значения, представленные в таблице 33.

Таблица 33. Прогнозные значения.

$X_{\text{макс}} - X_{\text{мин}}$	255,5	Значение прогноза
$X_{\text{макс}} - X_{\text{мин}}$	127,761	123,818
$X_{\text{макс}} - X_{\text{мин}}$	63,881	124,849
$X_{\text{макс}} - X_{\text{мин}}$	31,940	125,365
$X_{\text{макс}} - X_{\text{мин}}$	85,174	124,506
$X_{\text{макс}} - X_{\text{мин}}$	51,104	125,056

Таким образом, для линейной модели можно осуществлять расчёт прогнозных значений для любого промежутка экзогенной переменной.

6. Отбор экзогенных переменных в модель множественной линейной регрессии

Лабораторная работа № 6.

Цель работы: Построение оптимальной модели множественной линейной регрессии.

Выполнение работы.

Исходными данными служит матрица корреляций, полученная в ходе первой лабораторной работы (таблица 45). В качестве эндогенной переменной выбрана X_1 (таким образом, X_1 переобозначается как Y_1).

Таблица 34. Матрица корреляций

R	Y_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
X Y_1	1	0,08	0,96	-0,01	0,96	0,16	0,91	-0,08
XX ₂	0,08	1	-0,04	0,32	0,14	0,92	-0,03	0,19
XX ₃	0,96	-0,04	1	0,08	0,90	0,03	0,92	-0,02
XX ₄	-0,01	0,32	0,08	1	-0,01	0,12	0,01	0,53
XX ₅	0,96	0,14	0,90	-0,01	1	0,21	0,95	-0,05
XX ₆	0,16	0,92	0,03	0,12	0,21	1	0,02	0,20
XX ₇	0,91	-0,03	0,92	0,01	0,95	0,02	1	0,01
XX ₈	-0,08	0,19	-0,02	0,53	-0,05	0,20	0,01	1

Шаг 1: В матрице корреляций находим переменную X_i , имеющую максимальный по модулю коэффициент парной корреляции. Из матрицы корреляций определяем наибольший коэффициент корреляции $r(Y_1; X_3)=0,958$.

Используя функцию Excel ЛИНЕЙН получим регрессионную статистику для модели $y=f(x_3)$ (таблица 34)

Таблица 34. Регрессионная статистика

b_{yx}	6,156	113,268	b_0
S_{byx}	0,199	36,080	S_{b0}
R^2	0,917	277,711	S_e
$F_{набл}$	947,994	85	K
Q_R	73112500	6555485	$Q_{ост}$

Используя полученные данные, рассчитываем скорректированный коэффициент детерминации при $n = 87$ и $m = 1$ (где n — объем выборки, m — число независимых переменных).

$$R_{ск}^2 = 1 - \frac{Q_{ост}/(n - m - 1)}{Q/(n - 1)}$$

$$Q = Q_{рег} + Q_{ост}$$

Тогда значение скорректированного коэффициента детерминации для модели $y = f(X_3)$ равно 0,917.

Шаг 2: Отбор переменной. По матрице корреляций находим переменную, имеющую максимальный коэффициент корреляции с Y_1 (кроме X_3). Такой переменной является X_5 , так как $r(Y_1; X_5) = 0,957$.

Выясним степень тесноты взаимосвязи X_3 и X_5 . Коэффициент корреляции $r(X_3, X_5)$ равен 0,903, что указывает на сильную корреляцию. Однако использовать этот коэффициент корреляции в модели нельзя, так как это может привести к мультиколлинеарности.

Продолжаем подбор второй переменной.

Отбор второй переменной. По матрице корреляций выбираем переменную (кроме X_3), имеющую максимальный коэффициент корреляции с Y_1 . Этой переменной является X_7 , так как $r(Y_1; X_7) = 0,91$. Далее выясняем степень тесноты взаимосвязи между X_3 и X_7 .

Коэффициент корреляции $r(X_3, X_7) = 0,92$ указывает на высокую мультиколлинеарность. Поэтому в качестве второй переменной выбираем X_6 , где $r(Y_1, X_6) = 0,16$ и $r(X_3, X_6) = 0,03$.

Используя функцию Excel ЛИНЕЙН получим регрессионную статистику для модели $Y_1=f(X_3;X_6)$ (таблица 35).

Таблица 35. Регрессионная статистика для модели $y=f(x_3; x_6)$.

$b_2=1,849$	$b_1=6,132$	$b_0=-118,504$
$S_{b2}=0,379$	$S_{b1}=0,178$	$S_{b0}=57,247$
$R^2=0,936$	$S_e=246,531$	
$F_{\text{набл}}=613,408$	$k=84$	
$Q_R=74562684$	$Q_{\text{ост}}=5105300,567$	

Используя полученные данные, рассчитаем скорректированный коэффициент детерминации при $n = 87$ и $m = 2$ (где n — объем выборки, m — число независимых переменных).

$$R_{\text{ск}}^2 = 1 - \frac{Q_{\text{ост}}/(n - m - 1)}{Q/(n - 1)}$$

Тогда значение скорректированного коэффициента детерминации для модели $Y_1=f(X_3; X_6)$ равно 0,934.

Шаг 3. При отборе третьей переменной рассматриваем переменные X_2 , X_4 и X_8 . Коэффициент корреляции между Y_1 и X_2 равен $r(Y_1; X_2) = 0,08$. Проверка на мультиколлинеарность показывает: $r(X_2; X_3) = -0,04$, $r(X_2; X_6) = 0,92$. Высокая корреляция $r(X_2; X_6)$ указывает на мультиколлинеарность; поэтому X_2 не включается в модель.

Рассмотрим X_8 . Проверка на мультиколлинеарность: $r(X_8, X_3) = -0,02$, $r(X_8, X_6) = 0,2$. Мультиколлинеарность отсутствует.

Используя функцию Excel ЛИНЕЙН получим регрессионную статистику для модели $Y_1=f(X_3; X_6; X_8)$ (таблица 36)

Таблица 36. Регрессионная статистика для модели $Y_1=f(X_3; X_6; X_8)$.

$b_8=-0,261$	$b_6=2,093$	$b_3=6,117$	$b_0=125,961$
$S_{b8}=0,697$	$S_{b6}=0,366$	$S_{b3}=0,168$	$S_{b0}=92,837$
$R^2=0,943$	$S_e=233,633$		
$F_{\text{набл}}=458,844$	$k=83$		
$Q_R=75137466,24$	$Q_{\text{ост}}=4530518,529$		

Используя полученные данные, рассчитаем скорректированный коэффициент детерминации при $n = 87$ и $m = 3$ (где n — объем выборки, m — число независимых переменных).

$$R_{\text{ск}}^2 = 1 - \frac{Q_{\text{ост}}/(n - m - 1)}{Q/(n - 1)}$$

Тогда значение скорректированного коэффициента детерминации для модели $Y_1=f(X_3;X_6;X_8)$ равно 0,941.

Шаг 4. Добавляем переменную x_4

Используя функцию Excel ЛИНЕЙН получим регрессионную статистику для модели $Y_1=f(X_3;X_6;X_8;X_4)$ (таблица 37)

Таблица 37. Регрессионная статистика для модели $Y_1=f(X_3;X_6;X_8;X_4)$.

$b_4=-1,418$	$b_8=-1,166$	$b_6=2,106$	$b_3=6,162$	$b_0=154,947$
$S_{b4}=0,528$	$S_{b8}=0,786$	$S_{b6}=0,354$	$S_{b3}=0,163$	$S_{b0}=90,195$
$R^2=0,948$	$S_e=225,353$			
$F_{\text{набл}}=371,689$	$k=82$			
Q_R $=75503683,00$	$Q_{\text{Рост}}$ $=4164302,1$			

Используя полученные данные, рассчитаем скорректированный коэффициент детерминации при $n = 87$ и $m = 4$ (где n — объем выборки, m — число независимых переменных).

$$R_{\text{ск}}^2 = 1 - \frac{Q_{\text{ост}}/(n - m - 1)}{Q/(n - 1)}$$

Тогда значение скорректированного коэффициента детерминации для модели $Y_1=f(X_3;X_6;X_8;X_4)$ равно 0,945.

Вывод: Пошаговый отбор переменных, с учетом скорректированного коэффициента детерминации и исключением мультиколлинеарности, приводит к множественной линейной регрессии вида $Y_1=f(X_3;X_6;X_8;X_4)$.

Основные параметры модели будут рассчитаны с помощью пакета «Анализ данных» (таблица 38).

Таблица 38. Регрессионная статистика.

Регрессионная статистика	
Множественный R	0,973
R-квадрат	0,947
Нормированный R-квадрат	0,945
Стандартная ошибка	225,353
Наблюдения	87

Множественный R — это множественный коэффициент корреляции, который показывает степень тесноты взаимосвязи показателя Y_1 (эндогенной переменной X_1) со всеми оставшимися в исследовании показателями (X_2 - X_8). Данная связь может быть охарактеризована как очень сильная.

R-квадрат-коэффициент детерминации. Определяет степень обоснования эндогенной переменной X_1 всеми экзогенными переменными, т.е. является мерой, позволяющей определить, в какой степени уравнение регрессии дает лучший результат при использовании всех экзогенных переменных.

Нормированный R-квадрат – скорректированный коэффициент детерминации.

Стандартная ошибка – несмещенная оценка остаточной дисперсии.

$$S_e = \frac{\sum_{i=1}^n e_i^2}{n - m - 1}$$

Наблюдения – n объём выборки (87)

Таблица 39. Дисперсионный анализ в рамках регрессионного анализа.

	df	SS	MS	F	Значи- мость F
Регрессия (Q_r)	4	755036 83	188759 21	371,6 89	0,001
Остаток ($Q_{\text{остат}}$)	82	416430 2	50784, 17		
Итого ($\sum Q_r + Q_{\text{остат}}$)	86	796679 85			

df – степени свободы

SS – вариативность обусловленная регрессией

Q_r

MS – вариативность деления на степень свободы

$$MS = SS/df$$

F – наблюдаемое значение статистики F Фишера -Снедекора

Значимость F – уровень значимости, соответствующий численному значению $F_{\text{наблюдаемого}}$.

Если значимость F меньше либо равна заданного уровня значимости альфа, то уравнение регрессии статистически значимо.

Значимость F много меньше, чем 0,05, следовательно, уравнение значимо.

m количество экзогенных переменных равно четырём.

Вариативность обусловленная регрессией SS

Остаток ($Q_{\text{остат}}$) Значимость F.

Таблица 40. Регрессионная статистика параметров модели.

Обозначение экзогенных переменных	Коэффициенты	Стандартная ошибка	t-статистика	P-Значение
Y-пересечение	154,946	90,195	1,717	0,089
X ₃	6,161	0,163	37,753	0,001
X ₆	2,106	0,353	5,955	0,001
X ₈	-1,166	0,786	-1,483	0,141
X ₄	-1,418	0,528	-2,685	0,008

Y-пересечение – свободный член;

Коэффициенты – параметры модели;

Параметры модели, которая имеет вид:

$$X_1 = 154,947 + 6,162X_3 + 2,106 X_6 + (-1,166) X_8 + (-1,418) X_4$$

Множественная линейная модель регрессии

Стандартная ошибка – это стандартная ошибка данного параметра модели.

t-статистика используется для расчёта наблюдаемого значения t-статистики Стьюдента (коэффициент делить на стандартную ошибку).

P-Значение – уровень значимости, соответствующий параметру модели

Для b₀ уровень значимости 0,08

Для X₃ уровень значимости 0,001

Для X₆ уровень значимости 0,001

Для X₈ уровень значимости 0,141

Для X₄ уровень значимости 0,008

Если альфа 0,05, то значимы все значения кроме X₈ – 0,008

Таблица 41. Интервальная оценка параметров модели.

Интер-вал	Нижние 95%	Верхние 95%	Нижние 90,0%	Верхние 90,0%
У-пере-сечение	-24,480	334,374	4,893	305,000
X ₃	5,836	6,486	5,890	6,433
X ₆	1,402	2,809	1,517	2,694
X ₈	-2,730	0,398	-2,474	0,141
X ₄	-2,469	-0,367	-2,297	-0,539

В таблице 41 представлен вывод интервальной оценки параметров модели с доверительной вероятностью 0,95 и 0,9.

Интервальная оценка параметров модели рассчитывается по формулам:

$$b_j - \Delta < \beta_j < b_j + \Delta$$

$$\Delta = t_{\gamma} \hat{S}[(X^T X)^{-1}]_{jj}^{\frac{1}{2}}$$

7. Проверка на гетероскедастичность

Лабораторная работа № 7.

Цель работы: Выявить с помощью графического метода и коэффициента ранговой корреляции по Спирмену и теста Голдфелда-Квандта гетероскедастичность.

Выполнение работы.

Задание 1. Графический метод.

Необходимо построить точечный график, по оси абсцисс откладывается показатель x . По оси ординат квадраты остатков.

Исходный массив данных представлен в таблице 1.

Используя пакет Анализа данных – Регрессия, вывести остатки, вычислить квадраты остатков. В таблице 42 представлены результаты вычислений.

Таблица 42. Расчетные значения остатков и квадратов остатков для модели $Y=f(X_3 X_6 X_8 X_4)$. Примеры данных.

Остатки	Квадрат остатков	Остатки	Квадрат остатков
-157,195	24710,267	-6,782	45,991
-78,971	6236,348	-11,575	133,969
10,940	119,686	7,401	54,780
-258,792	66973,202	-53,990	2914,937
-99,685	9937,039	37,720	1422,809
...	...	...	...
-183,629	33719,772	-15,436	238,268
-118,567	14058,109	-71,951	5176,953
640,208	409866,205	213,085	45405,087

Используя пакет Анализа данных – Регрессия, вывести графики остатков и квадратов остатков. В качестве примера приведены следующие графики.

На рисунках 7-11 представлены графики квадратов остатков.

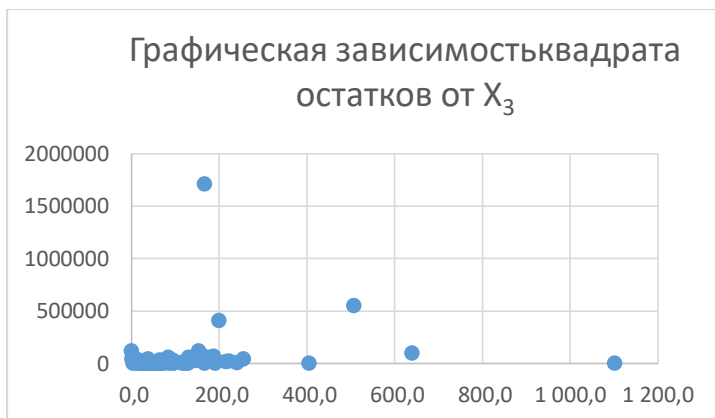


Рис. 7. Графическая зависимость квадрата остатков от X_3 .

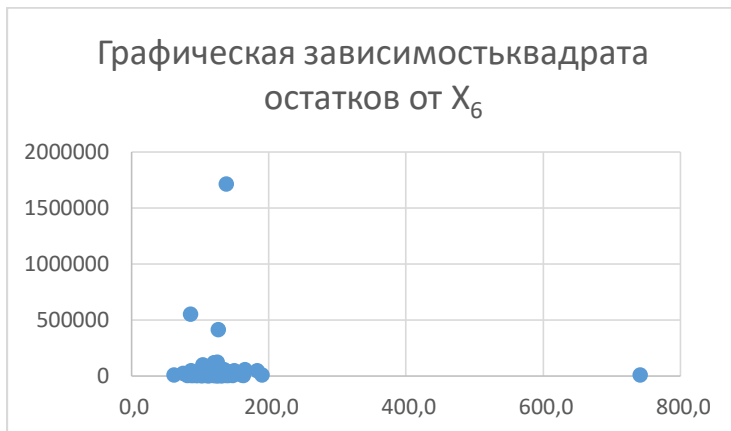


Рис. 8. Графическая зависимость квадрата остатков от X_6 .

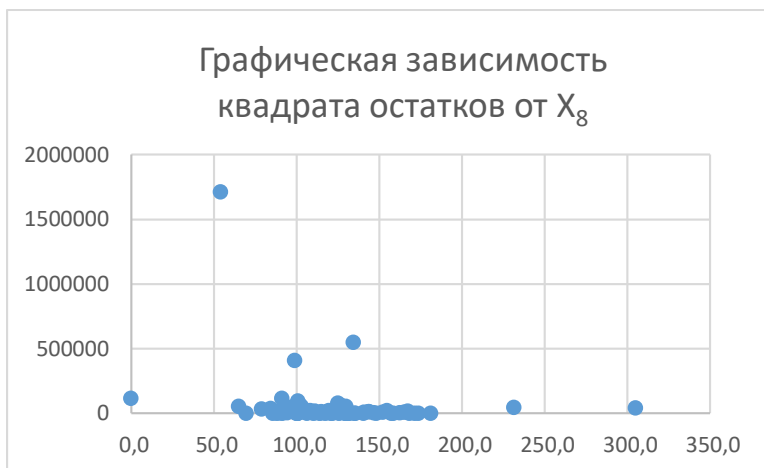


Рис. 9. Графическая зависимость квадрата остатков от X_8 .

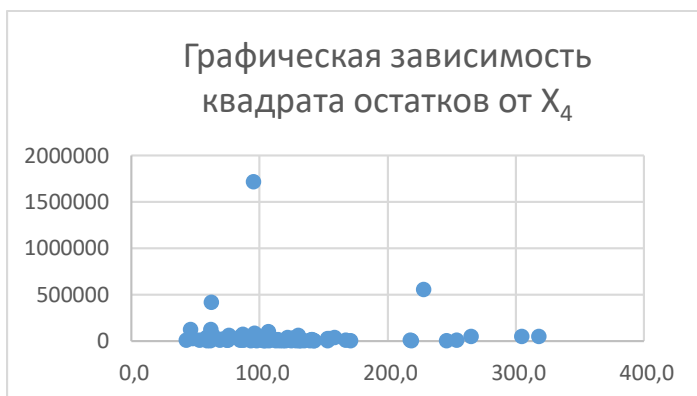


Рис. 10. Графическая зависимость квадрата остатков от X_4 .

Из анализа графиков можно сделать вывод, что исходный массив данных не предполагает наличие гетероскедастичности, так как не выявляется функциональная зависимость между представленными переменными.

Задание 2. Выявление гетероскедастичности с помощью коэффициента ранговой корреляции по Спирмену.

Используя функцию Excel РАНГ, были рассчитаны ранги для изучаемых показателей и модуля остатков рисунок 11.

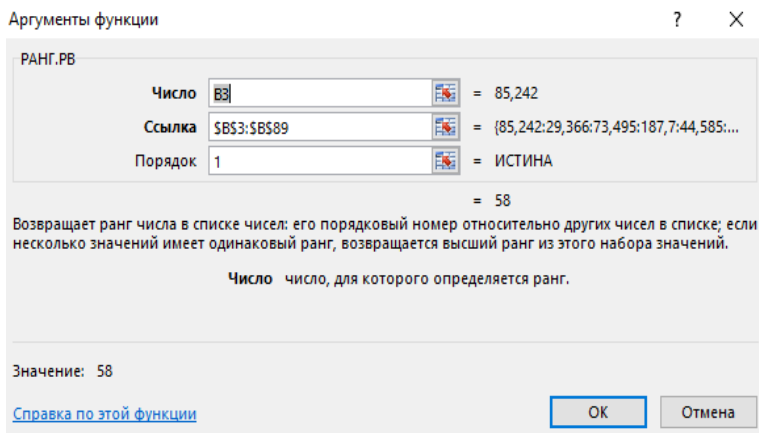


Рис. 11. Расчёт ранга с помощью функции Excel.

Далее в таблице представлены расчетные значения рангов

Таблица 43. Ранги изучаемых показателей и модулей остатков.

Ранг (X_3)	Ранг (X_6)	Ранг (X_8)	Ранг (X_4)	Ранг (остатки в модуле)
58	24	15	23	64
18	86	35	1	39
54	28	22	6	9
77	47	54	24	80
...	...	...	...	...
7	82	73	20	36
71	57	13	11	84
2	81	16	34	74
1	46	1	2	83

После расчёта рангов для каждого показателя X рассчитывается квадрат разности между рангами показателя X и абсолютной

величиной остатков. Полученные значения суммируются для каждого столбика. В таблице представлены результаты расчётов.

Таблица 44. Пример результатов расчётов.

(ранг X_3 - ранг ABS e_i) ²	(ранг X_6 - ранг ABS e_i) ²	(ранг X_8 - ранг ABS e_i) ²	(ранг X_4 - ранг ABS e_i) ²
36	1600	2401	1681
441	2209	16	1444
2025	361	169	9
9	1089	676	3136
...	...	...	...
6724	1369	6724	6561
Сумма			
86700	102382	134864	128918

Коэффициент ранговой корреляции по Спирмену рассчитывается по формуле:

$$r_s = 1 - \frac{6 \sum (rang(x) - rang(|e_i|))^2}{n(n^2 - 1)}$$

Таблица 45. Численные значения коэффициентов ранговой корреляции

Коэффициент ранговой корреляции по Спирмену			
$e_i X_3$	$e_i X_6$	$e_i X_8$	$e_i X_4$
0,210	0,067	0,229	-0,175

Проверим гипотезу о статистической значимости коэффициента ранговой корреляции по Спирмену.

H_0 : коэффициент ранговой корреляции по Спирмену статистически не значим (гетероскедастичность отсутствует).

H_1 : коэффициент ранговой корреляции по Спирмену статистически значим (гетероскедастичность имеет место).

Воспользуемся статистикой

$$t_{\text{набл}} = \frac{|r_s| \sqrt{n-2}}{\sqrt{1-r_s^2}}$$

Используется функция Excel СТЬЮДРАСПОБР (0,05;n-2).

Вывод по тесту с использованием коэффициента ранговой корреляции Спирмана:

Если $t_{\text{набл}}$ меньше $t_{\text{кр}}$, то гетероскедастичность отсутствует.

Таблица 46. Расчеты $t_{\text{кр}}$ и $t_{\text{набл}}$.

	$e_i X_3$	$e_i X_6$	$e_i X_8$	$e_i X_4$
r_s	0,210	-0,175	0,067	-0,229
$t_{\text{набл}}$	1,979	0,619	2,169	1,637
$t_{\text{кр}}$	1,988			

Показатели X_3 , X_4 , X_6 не вносят гетероскедастичность.

Показатель X_8 – вносит гетероскедастичность.

Задание 3. Гетероскедастичность модели $Y=f(X_3; X_4; X_6; X_8)$.

Тест Голдфелда-Квандта.

Для изучения зависимости дисперсии случайной компоненты от величины x в случае линейной модели имеем зависимость:

$$\sigma^2(Ei) \approx \lambda x_i$$

Необходимо выявить данную зависимость и проверить её на значимость.

1. Для этого необходимо сформировать массив ($Y; X_3$).
2. Отсортировать столбик X_3 сортировать по возрастанию.
3. Общий объем выборки $n=87$.
4. Объем выборки n делим на 3 примерно равные части — это 29. Берём первые и последние 29. Между ними выборка исключается. Используем функцию Excel ЛИНЕЙН.

Таблица 47. Верхний массив (Y; X₃)

Y	X ₃	Y	X ₃
3,6	0,3	190,4	26,4
23,1	2,2	199,8	29,2
4,9	2,7	469,6	29,4
19,6	4,6	252,1	29,4
45,7	5,4	160,9	29,4
50,9	7,5	379,9	31,1
190,5	9,6	223,4	31,9
61,2	10,9	276,7	32,0
122,6	12,9	127,0	32,3
170,9	16,9	212,3	32,4
183,2	21,5	266,7	35,1
160,1	21,6	381,3	36,4
242,1	24,9	296,3	36,5
169,4	24,9	450,2	37,7
219,8	26,0		

Таблица 48. ЛИНЕЙН для верхнего массива (Y; X₃).

b_{yx}	8,708	-1,003	b_0
$S_{b_{yx}}$	1,145	28,680	S_{b_0}
R^2	0,682	72,520	S_e
$F_{\text{набл}}$	57,812	27,000	K
Q_R	304038,900	141996,100	$Q_{\text{ост}}$

Вывод по ЛИНЕЙН позволяет вычислить дисперсию по верхней выборке. Дисперсия по верхней части $\sigma^2 = 5259,115$.

Таблица 49. Нижняя часть массива (Y; X₃).

Y	X ₃	Y	X ₃
398,3	87,1	2 587,5	167,4
803,1	94,1	945,7	173,5
644,6	97,1	1 480,9	174,4
837,3	101,7	1 037,4	187,7
757,7	114,3	1 273,9	191,7
758,1	118,3	2 088,1	199,6
588,4	119,0	1 613,9	215,9
978,9	124,6	1 546,9	223,3
863,9	130,2	1 622,0	241,2
1 097,5	131,3	1 969,6	255,9
1 319,5	131,4	2 585,1	405,5
1 180,2	148,4	2 243,6	507,9
1 516,3	153,7	3 739,4	640,5
1 388,2	161,7	6 906,5	1 103,1
2 260,5	167,3		

Таблица 50. ЛИНЕЙН для нижнего массива (Y; X₃).

b_{yx}	5,639	344,792	b_0
$S_{b_{yx}}$	0,399	121,939	S_{b_0}
R^2	0,881	441,454	S_e
$F_{\text{набл}}$	200,129	27,000	K
Q_R	39001455,000	5261797,00 0	$Q_{\text{ост}}$

Вывод по ЛИНЕЙН позволяет вычислить дисперсию по верхней выборке. Дисперсия по верхней части $\sigma^2 = 194881,363$.

Проверим гипотезу о равенстве дисперсий:

H₀: Дисперсия верхней части массива равна дисперсии нижней массива части (гетероскедастичность отсутствует).

H_1 : Дисперсия верхней части массива не равна дисперсии нижней части (гетероскедастичность присутствует).

Рассчитаем $F_{набл}$ и $F_{кр}$ по формулам:

$$F_{набл} = \frac{\sigma_{больш}^2}{\sigma_{меньш}^2} F_{кр}(\alpha; K_1; K_2) \text{ ФРАСПРОБР}(0.05; 27; 27)$$

Если $F_{набл} > F_{кр}$, то результаты противоречат гипотезе H_0 и она отвергается, принимается гипотеза H_1 .

Если $F_{набл} < F_{кр}$, то результаты не противоречат гипотезе H_0 и она принимается.

$F_{набл}=37,056$, $F_{кр}=1,905$, следовательно, показатель X_3 дает гетероскедастичность.

Аналогично необходимо провести тест для всех показателей модели. Для показателя X_4 имеем следующие результаты.

Таблица 51. Верхняя часть массива (y; x_4)

Y	X_4	Y	X_4
469,6	43,0	223,4	71,3
3,6	46,6	568,9	74,7
1 613,9	48,1	242,1	75,5
624,1	53,7	1 480,9	76,7
170,9	58,9	190,5	85,2
652,9	59,2	1 969,6	85,3
183,2	59,5	309,5	85,8
45,7	61,6	515,8	86,9
381,3	61,9	1 037,4	87,3
309,6	62,1	863,9	87,7
1 516,3	62,1	61,2	88,1
199,8	62,5	544,3	93,4
2 088,1	62,9	2 587,5	95,8
1 319,5	64,2	266,7	96,2
328,4	69,4		

Таблица 52 Вывод по функции ЛИНЕЙН для верхней части массива (Y; X₄).

b _{yx}	6,736	236,364	b ₀
S _{b_{yx}}	8,535	621,559	S _{b₀}
R ²	0,023	696,838	S _e
F _{набл}	0,623	27	K
Q _R	302462,3	13110760	Q _{ост}

Вывод по ЛИНЕЙН позволяет вычислить дисперсию по верхней выборке. Дисперсия по верхней части $\sigma^2 = 485583,705$.

Таблица 53. Нижняя часть массив (Y; X₄)

Y	X ₄	Y	X ₄
1 097,5	126,4	127,0	153,3
1 622,0	127,0	1 273,9	153,6
122,6	128,9	1 546,9	155,2
945,7	129,6	803,1	159,0
431,3	130,5	2 260,5	167,5
452,8	130,6	19,6	171,4
219,8	131,6	275,5	217,9
160,1	132,4	299,3	218,7
415,2	135,2	2 243,6	228,6
644,6	139,7	398,3	246,2
6 906,5	140,6	588,4	254,3
758,1	141,0	50,9	265,5
160,9	141,4	195,1	305,2
2 585,1	142,8	4,9	318,6
306,2	143,3		

Таблица 54. Вывод по функции ЛИНЕЙН для нижней части массива (Y; X₄).

b_{yx}	-4,875	1774,723	b_0
$S_{b_{yx}}$	4,499	820,523	S_{b_0}
R^2	0,042	1351,021	S_e
$F_{\text{набл}}$	1,174	27	K
Q_R	2143538	49281971	$Q_{\text{ост}}$

Вывод по ЛИНЕЙН позволяет вычислить дисперсию по нижней выборке. Дисперсия по нижней части $\sigma^2 = 1825258,185$.

$F_{\text{набл}}=0,001$, $F_{\text{кр}}= 1,905$, следовательно, по показателю x_4 отсутствует гетероскедастичность.

Таблица 55. Верхняя часть массива (Y; X₆)

Y	X ₆	Y	X ₆
160,9	62,4	266,7	100,3
1 613,9	75,6	345,4	101,9
19,6	81,9	261,9	102,1
945,7	85,9	306,2	104,1
2 243,6	86,9	373,9	104,7
4,9	87,4	3 739,4	104,8
327,9	87,9	275,5	105,9
309,6	90,6	450,2	106,1
199,8	92,8	515,8	106,7
615,0	94,5	223,4	107,3
122,6	95,7	624,1	109,6
328,4	96,9	273,8	110,1
863,9	97,1	652,9	111,7
426,2	98,0	252,1	112,3
61,2	99,4		

Таблица 56. Вывод по функции ЛИНЕЙН для верхней части массива (Y; X₆).

-2,140	787,637
12,803	1253,776
0,001	781,852
0,028	27
17080,1	16504884

Вывод по ЛИНЕЙН позволяет вычислить дисперсию по верхней выборке. Дисперсия по верхней части $\sigma^2 = 611292,015$.

Таблица 57. Нижний массив (Y; X₆)

Y	X ₆	Y	X ₆
837,3	126,0	523,9	138,3
2 585,1	126,0	2 587,5	138,6
442,5	126,5	379,9	139,3
2 088,1	127,1	588,4	142,1
296,3	127,5	349,5	142,5
170,9	127,9	299,3	147,8
160,1	130,0	758,1	150,0
978,9	130,4	23,1	150,6
1 546,9	131,2	190,5	161,7
1 273,9	131,9	415,2	163,4
127,0	132,3	452,8	166,3
219,8	133,7	195,1	184,0
276,7	133,8	469,6	190,7
50,9	135,3	2 260,5	742,4
1 319,5	135,5		

Таблица 58. Вывод по функции ЛИНЕЙН для нижней части массива (Y; X₆).

2,255	387,646
1,235	242,930
0,109	737,119
3,333	27
1810968	14670321

Вывод по применению функции ЛИНЕЙН: позволяет вычислить дисперсию по нижней выборке. Дисперсия по нижней части $\sigma^2 = 543345,211$. $F_{\text{набл}} = 1,125$, $F_{\text{кр}} = 1,905$, следовательно, показатель X_6 не дает гетероскедастичность.

Таблица 59. Верхний массив (Y; X₈)

Y	X ₈	Y	X ₈
3,6	0,0	23,1	94,0
2 587,5	54,1	275,5	94,6
1 319,5	64,9	4,9	96,1
19,6	69,7	2 088,1	99,1
170,9	79,1	1 097,5	99,5
1 969,6	84,5	379,9	100,0
544,3	85,6	652,9	100,5
1 613,9	86,8	345,4	100,8
624,1	88,4	327,9	100,9
212,3	88,4	3 739,4	101,0
266,7	89,5	1 480,9	102,8
757,7	91,1	199,8	103,9
1 516,3	91,2	234,9	104,9
568,9	91,7	945,7	106,3
515,8	93,1		

Таблица 60. ЛИНЕЙН для верхнего массива (Y; X₈).

2,253	645,321
8,186	742,457
0,003	904,334
0,076	27
61951,79	22081156

Вывод по ЛИНЕЙН позволяет вычислить дисперсию по верхней выборке. Дисперсия по верхней части $\sigma^2 = 817820,589$.

Таблица 61. Нижний массив (Y; X₈).

Y	X ₈	Y	X ₈
122,6	129,8	415,2	148,4
452,8	129,9	1 622,0	152,1
381,3	130,1	1 546,9	154,6
1 180,2	130,6	160,1	157,3
6 906,5	131,0	296,3	158,6
426,2	131,6	2 260,5	162,4
219,8	132,3	160,9	165,1
2 243,6	134,5	554,6	167,4
252,1	134,9	299,3	168,4
1 273,9	135,8	588,4	171,6
309,6	140,6	261,9	173,9
837,3	141,1	276,7	181,2
183,2	143,7	50,9	231,5
615,0	143,7	195,1	305,2
190,5	146,6		

Таблица 62. ЛИНЕЙН для (Y; X8).

-8,015	2090,554
6,878	1102,641
0,047	1310,386
1,357	27
2331563	46362026

Вывод по ЛИНЕЙН позволяет вычислить дисперсию по нижней выборке. Дисперсия по нижней части $\sigma^2 = 1717112,079$.

$F_{\text{набл}} = 2,099$, $F_{\text{кр}} = 1,905$, следовательно, по показателю x_8 имеет место гетероскедастичность.

Общий вывод по лабораторной работе 7: в результате использования графического метода, теста ранговой корреляции по Спирмену и теста Голфилда-Квандта можно установить, что графический метод является слабо чувствительным, так как по графикам зависимости квадрата остатков от x не очевидна связь между ними. Обнаружить связь позволяет коэффициент ранговой корреляции Спирмена, который позволил выявить гетероскедастичность по показателю X_8 .

Более чувствительным является тест Голфилда-Квандта, который выявил гетероскедастичность для показателей X_3 X_4 и X_8 .

8. Регрессионные модели с фиктивными переменными

Лабораторная работа № 8.

Цель работы: Изучить регрессионные модели с фиктивными переменными.

Выполнение работы.

Задание 1. Используя модель ЛР №2, ввести фиктивную переменную C , имеющую два уровня 0 и 1. Объем выборки $n=47$. Правило введения фиктивной переменной: для введения фиктивной переменной для каждого показателя, который является экзогенной переменной, вычислить максимальное значение и разделить его на 3. Используя логическую функцию ЕСЛИ, в качестве условия задать x меньше, чем $x_{\max}/3$ ($x_{\max}/3$ взять в доллары). Если истина-0, если ложь-1. Протаскиваем по всему массиву. Используя пакет анализа данных, выведем всю регрессионную статистику.

Таблица 63. Исходный массив данных и фиктивные переменные $x_3(y)$ и x_6 .

$x_3(y)$	x_6	Фиктивные переменные
64,9	142,5	1
31,9	107,3	0
31,1	139,3	1
131,4	135,5	1
191,7	131,9	1
241,2	115,7	0
43,8	120,8	0
16,9	127,9	1
167,4	138,6	1
130,2	97,1	0
37,7	106,1	0
173,5	85,9	0
45,2	104,1	0

161,7	112,9	0
118,3	150,0	1
62,1	114,2	0
131,3	124,6	1
69,7	98,0	0
46,0	96,9	0
40,9	125,2	1
255,9	122,2	0
199,6	127,1	1
36,4	122,2	0
9,6	161,7	1
153,7	125,7	1
148,4	121,0	0
21,6	130,0	1
38,1	184,0	1
29,4	112,3	0
65,2	118,0	0
114,3	113,6	0
124,6	130,4	1
57,7	126,5	1
223,3	131,2	1
82,6	138,3	1
38,5	102,1	0
45,4	163,4	1
84,3	166,3	1
53,8	147,8	1
5,4	124,1	0
94,1	121,1	0
35,1	100,3	0

61,6	105,9	0
2,7	87,4	0
24,9	123,4	0
2,2	150,6	1
0,3	121,3	0

Используем пакет Анализа данных рисунок 12. Для кодирования используем переменную x_3 и $x_{cp} = 124,5$ Фиктивную переменную формировали по правилу $x < x_{max} / 3$.

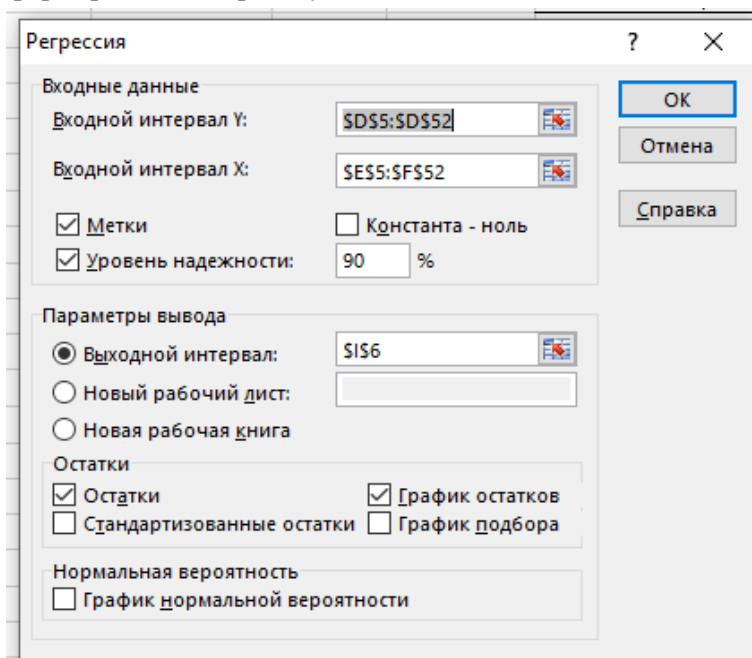


Рис. 12. Окно регрессионной функции.

Используя пакет анализа данных «Регрессия» получили регрессионную статистику таблица 2.

Таблица 64. Регрессионной статистики.

Множественный R	0,205
R-квадрат	0,042
Нормированный R-квадрат	-0,002
Стандартная ошибка	68,243
Наблюдения	47

Множественный $R=0,205$, это меньше $0,75$, следовательно, слабая связь между показателями.

Таблица 65. Дисперсионный анализ.

	df	SS	MS	F	Значимость F
Регрессия	2	8991,296	4495,648	0,965	0,389
Остаток	4	204913,9	4657,135		
Итого	6	213905,2			

Значимость $F=0,388$, следовательно модель статистически не значима. Роль фиктивной переменной нивелируется.

Таблица 66. Вывод регрессионной статистики

	Коэффициенты	Стандартная ошибка	t-статистика	P-Значение
Y-пересечение	179,614	82,511	2,177	0,035
X_6	-0,920	0,739	-1,246	0,219
Фиктивная переменная	40,506	30,194	1,342	0,187

P-значение свидетельствует о том, что фиктивная переменная значима на уровне 0,187.

Таблица 67. Интервальная оценка параметров модели

	Нижние 95%	Верхние 95%	Нижние 90,0%	Верхние 90,0%
У- пе- ресе- че- ние	13,324	345,905	40,976	318,252
X ₆	-2,409	0,568	-2,161	0,321
Фик тивн ая пере менн ая	-20,346	101,359	-10,226	91,239

Далее приводятся графики остатков за переменной X₆ и фиктивной переменной

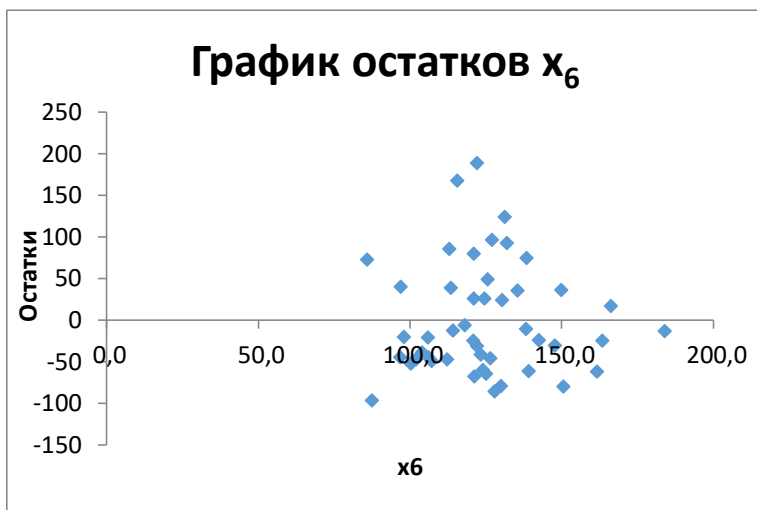


Рис. 13. График остатков для модели $x_3(y)$ и x_6 .

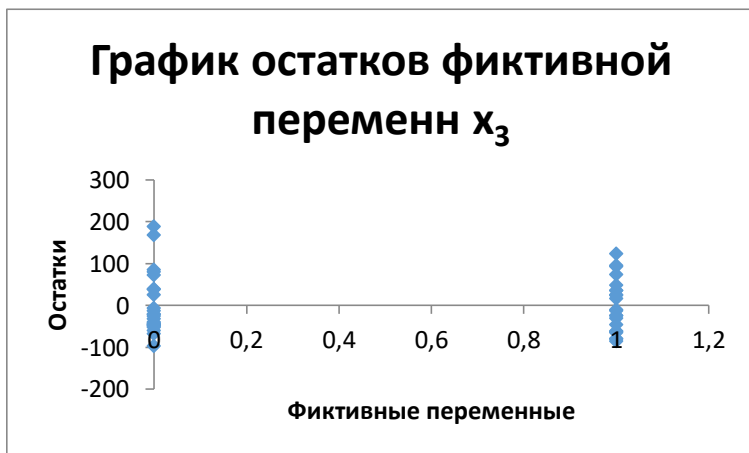


Рис. 14. График остатков для модели $x_3(y)$ и x_6

Множественный R равен 0,20 и меньше 0,75, значит слабая связь между переменными, что свидетельствует о наличии нелинейной связи.

Нахождение фиктивных переменных с массивом данных $x_6(y)$ и x_3 .

Таблица 68. Исходные данные $x_6(y)$ и x_3

$x_6(y)$	x_3	Фик- тивные пере- менные	$x_6(y)$	x_3	Фик- тивные пере- менные
142,5	64,9	0	125,7	153,7	1
107,3	31,9	0	121,0	148,4	1
139,3	31,1	0	130,0	21,6	0
135,5	131,4	1	184,0	38,1	0
131,9	191,7	1	112,3	29,4	0
115,7	241,2	1	118,0	65,2	0
120,8	43,8	0	113,6	114,3	1
127,9	16,9	0	130,4	124,6	1
138,6	167,4	1	126,5	57,7	0
97,1	130,2	1	131,2	223,3	1
106,1	37,7	0	138,3	82,6	0
85,9	173,5	1	102,1	38,5	0
104,1	45,2	0	163,4	45,4	0
112,9	161,7	1	166,3	84,3	1
150,0	118,3	1	147,8	53,8	0
114,2	62,1	0	124,1	5,4	0
124,6	131,3	1	121,1	94,1	1
98,0	69,7	0	100,3	35,1	0
96,9	46,0	0	105,9	61,6	0
125,2	40,9	0	87,4	2,7	0
122,2	255,9	1	123,4	24,9	0
127,1	199,6	1	150,6	2,2	0
122,2	36,4	0	121,3	0,3	0
161,7	9,6	0			

Таблица 69. Регрессионная статистика

	df	SS	MS	F	Значи- мость F
Ре- грес- сия	2	399,735	199,867	0,459	0,635
Оста- ток	44	19152,950	435,294		
Итого	46	19552,685			

Используя пакет Анализа данных получили регрессионную статистику. Для кодирования используем переменную x_6 и $x_{cp} = 84,0$

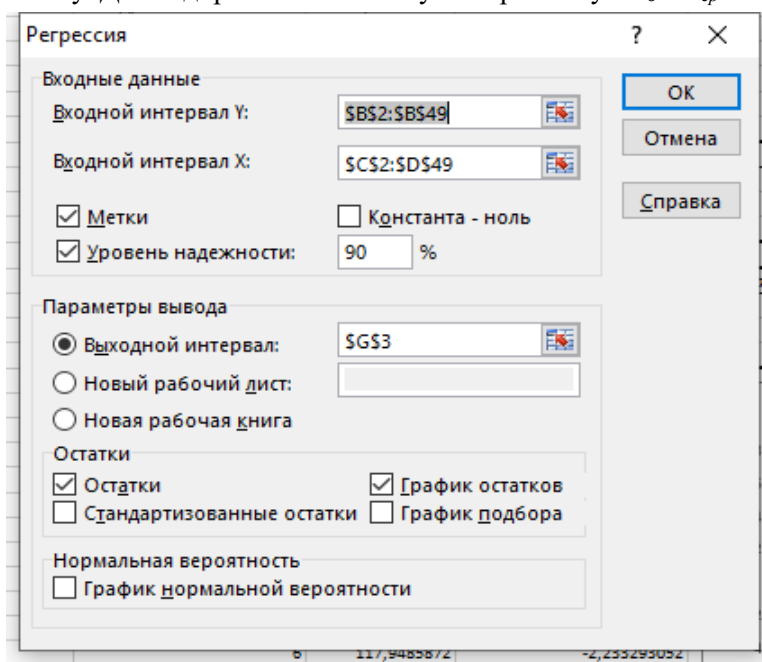


Рис. 15. Окно регрессионной функции для $x_6(y)$ и x_3

Используя пакет Анализа данных получили регрессионную статистику таблица 69.

Таблица 70. Дисперсионный анализ

Регрессионная статистика	
Множественный R	0,143
R-квадрат	0,020
Нормированный R-квадрат	-0,024
Стандартная ошибка	20,864
Наблюдения	47

Таблица 71. Дисперсионный анализ для модели с фиктивной переменной $X_3=f(X_6)$

	Коэффициенты	Стандартная ошибка	t-статистика	P-Значение
Y-пересечение	127,442	5,166	24,669	0,001
	-0,085	0,090	-0,948	0,347
Фиктивная переменная	11,107	12,494	0,888	0,378

Таблица 72. Интервальная оценка для модели с фиктивной переменной $X_3=f(X_6)$

Нижнее 95%	Верхние 95%	Нижние 90,0%	Верхние 90,0%
117,031	137,853	118,761	136,122
-0,266	0,096	-0,236	0,066
-14,073	36,286	-9,886	32,099

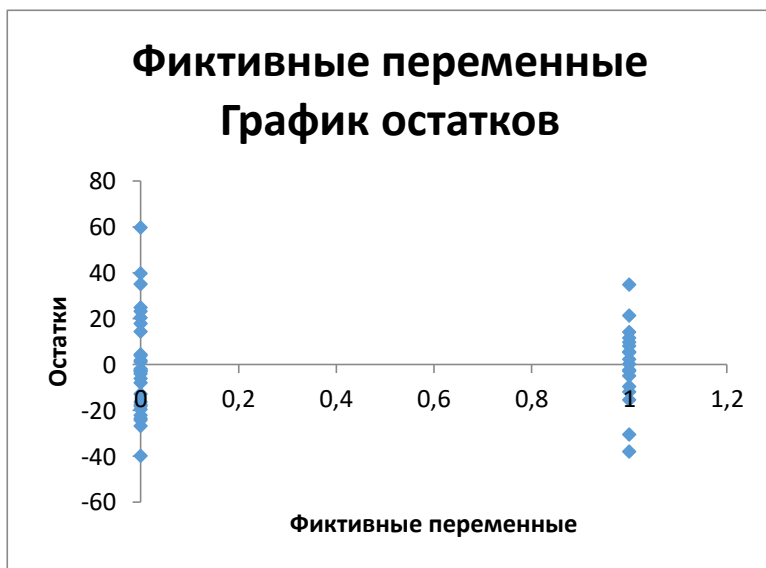


Рис.16. График остатков для модели график остатков для модели с фиктивной переменной $X_3=f(X_6)$

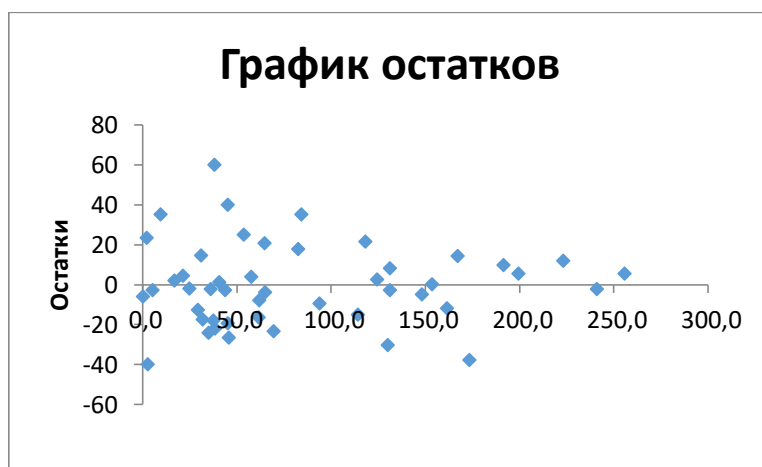


Рис. 17. График остатков для модели с фиктивной переменной $X_3=f(X_6)$

Множественный R равен 0,14 и меньше 0,75, значит слабая связь между переменными, что свидетельствует о наличии не линейной связи.

Она значима на уровне 0,37.

Нахождение фиктивных переменных с массивом данных график остатков для модели с фиктивной переменной $X_6=f(X_3)$

При введении фиктивной переменной в модель, в которой x_6 эндогенная x_3 экзогенная $y=f(x_6)$, фиктивная переменная :

$$\begin{cases} 1, & x_3 > x_{cp} \\ 0, & x_3 < x_{cp} \end{cases}$$

Модель статистически не значима т.к. ϕ значение много больше 0,05. Параметры также статистически не значимы т.к. значения больше 0,05.

Задание 2: Эндогенная переменная в модели является фиктивной. Формируем модель фиктивная переменная (x_3 x_6), фп 1 фп 2

Таблица 73. Регрессионная статистика

Множественный R	0,762
R-квадрат	0,581
Нормированный R-квадрат	0,562
Стандартная ошибка	0,334
Наблюдения	47

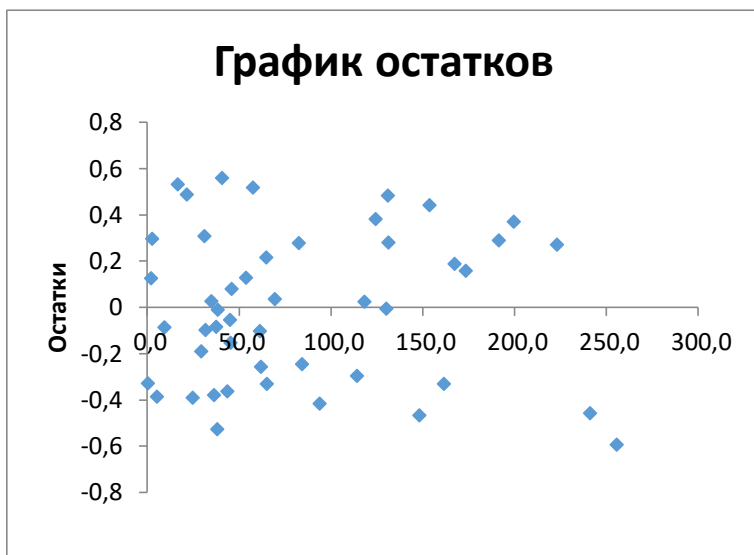


Рис. 18. График остатков для модели с фиктивной переменной $X_6=f(X_3)$

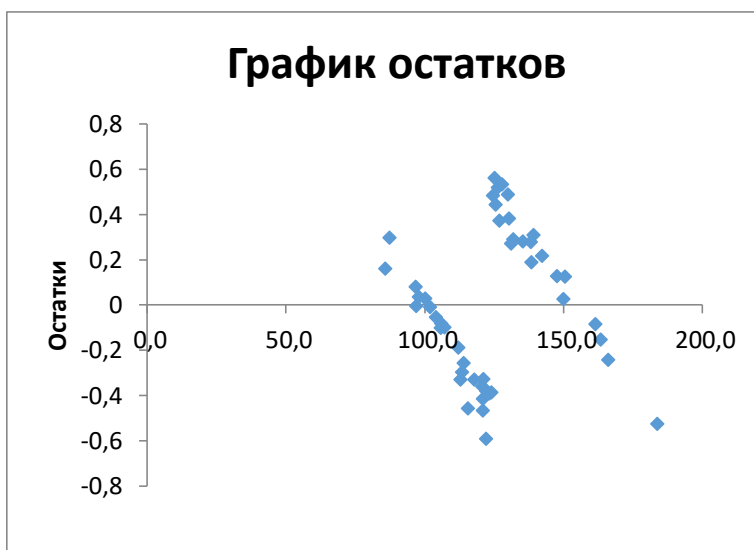


Рис. 19. График остатков для модели с фиктивной переменной $X_6=f(X_3)$

Таблица 74. Регрессионная статистика для модели с фиктивной переменной $X_6=f(X_3)$

Множественный R	0,868
R-квадрат	0,753
Нормированный R-квадрат	0,742
Стандартная ошибка	0,249
Наблюдения	47

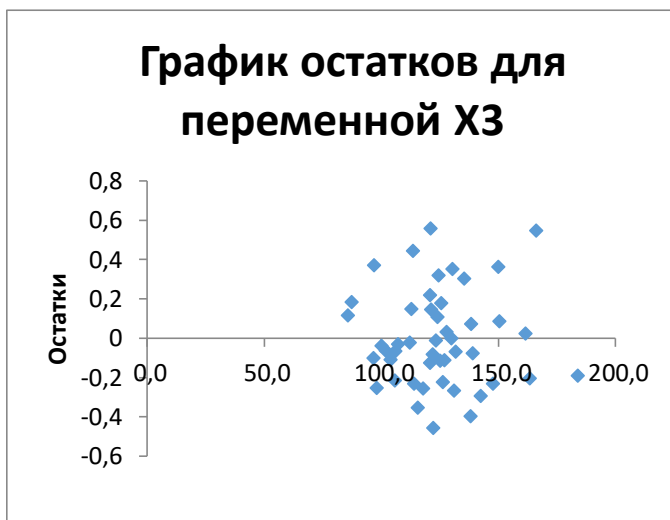


Рис. 20. График остатков для модели с фиктивной переменной $X_6=f(X_3)$.

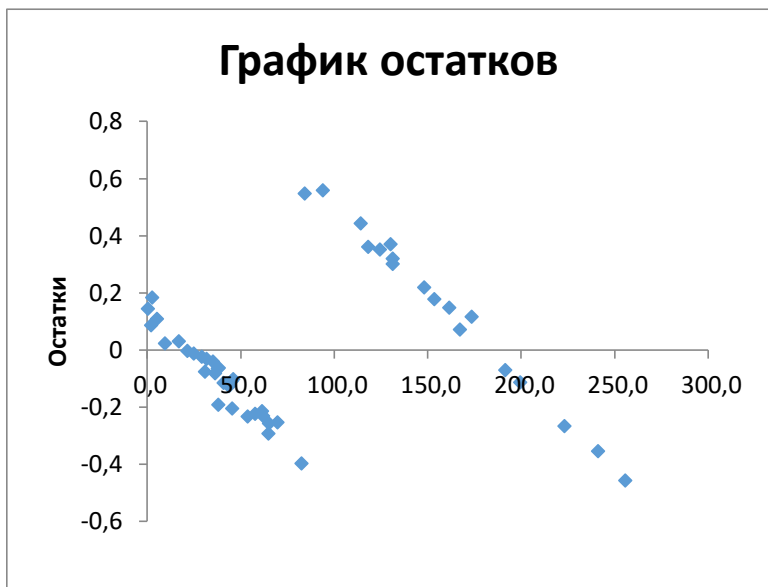


Рис. 21. График остатков для модели с фиктивной переменной $X_6=f(X_3)$

Вывод по варианту, в котором фиктивная переменная является эндогенной. Так как уровень значимости F-статистики больше, чем 0,05, значит модель не значима.

Эндогенная переменная это фиктивная переменная $ФП\ 1 = f(x_3\ x_6)$. Значит она статистически значима, если Р- значение 0,05 или меньше то статистически значима. Если больше, чем 0,02, то значима.

Для уравнения ФП 2 от тех же переменных и тоже самое описываем.

Задание 3: Формирование модели с двумя экзогенными фиктивными переменными две модели. В качестве У выбираем показатель X_1 . Экзогенные переменные $x_3\ x_4\ x_6\ x_8$

Для переменной x_4 найти среднее значение и формируем столбик ФП 1

9. Проверка условия Гаусса Маркова на автокорреляцию

Лабораторная работа № 9.

Цель работы: Изучить автокорреляцию остатков с помощью коэффициента автокорреляции.

Выполнение работы.

Автокорреляция остатков для модели $x_1 = f(x_3 \ x_4 \ x_6 \ x_8)$ изучается с помощью коэффициента автокорреляции:

$$r_{e_i e_{i-1}} = \frac{\sum_{i=2}^n e_i e_{i-1}}{\sqrt{\sum_{i=2}^n e_i^2 \sum_{i=2}^n e_{i-1}^2}}$$

В формуле e_i – значения остатков и e_{i-1} – сдвинутые на один лаг остатки

Таблица 75. Остатки исходные данные

Наб лю- де- ние	Остатки	Набл юде- ние	Остатк и	Набл юде- ние	Остатки
1	-157,195	30	138,874	59	-82,422
2	-78,971	31	-5,415	60	-170,213
3	10,940	32	- 107,059	61	200,235
4	-258,792	33	103,169	62	640,208
5	-99,685	34	- 312,660	63	-15,436
6	121,610	35	-34,653	64	-71,951
7	-110,756	36	-23,670	65	344,140
8	-34,665	37	239,908	66	161,792
9	-15,269	38	86,342	67	-30,798
10	44,248	39	-61,876	68	206,650

11	-117,313	40	- 148,911	69	5,198
12	-51,141	41	- 183,629	70	98,642
13	-90,772	42	- 118,567	71	-84,606
14	-74,364	43	-4,482	72	60,002
15	-38,202	44	236,004	73	-33,059
16	-48,870	45	36,224	74	140,167
17	132,359	46	94,614	75	-148,106
18	-742,793	47	- 108,816	76	35,466
19	-64,371	48	- 181,717	77	1,058
20	20,713	49	1308,35 0	78	-235,166
21	-6,782	50	-44,379	79	8,128
22	-11,575	51	126,550	80	-167,203
23	7,401	52	- 151,764	81	185,599
24	-53,990	53	-19,266	82	-75,090
25	37,720	54	281,409	83	-62,621
26	11,119	55	-94,343	84	213,085
27	211,272	56	-1,684	85	-108,399
28	-120,518	57	166,257	86	-208,276
29	-36,049	58	-64,261	87	-342,881

Таблица 76. Остатки $e_i; e_{i-1}$

e_i	e_{i-1}	e_i	e_{i-1}
-157,195	-78,971	236,004	36,224
-78,971	10,940	36,224	94,614
10,940	-258,792	94,614	-108,816
-258,792	-99,685	-108,816	-181,717
-99,685	121,610	-181,717	1308,350
121,610	-110,756	1308,350	-44,379
-110,756	-34,665	-44,379	126,550
-34,665	-15,269	126,550	-151,764
-15,269	44,248	-151,764	-19,266
44,248	-117,313	-19,266	281,409
-117,313	-51,141	281,409	-94,343
-51,141	-90,772	-94,343	-1,684
-90,772	-74,364	-1,684	166,257
-74,364	-38,202	166,257	-64,261
-38,202	-48,870	-64,261	-82,422
-48,870	132,359	-82,422	-170,213
132,359	-742,793	-170,213	200,235
-742,793	-64,371	200,235	640,208
-64,371	20,713	640,208	-15,436
20,713	-6,782	-15,436	-71,951
-6,782	-11,575	-71,951	344,140
-11,575	7,401	344,140	161,792
7,401	-53,990	161,792	-30,798
-53,990	37,720	-30,798	206,650
37,720	11,119	206,650	5,198
11,119	211,272	5,198	98,642
211,272	-120,518	98,642	-84,606
-120,518	-36,049	-84,606	60,002
-36,049	138,874	60,002	-33,059
138,874	-5,415	-33,059	140,167

-5,415	-107,059	140,167	-148,106
-107,059	103,169	-148,106	35,466
103,169	-312,660	35,466	1,058
-312,660	-34,653	1,058	-235,166
-34,653	-23,670	-235,166	8,128
-23,670	239,908	8,128	-167,203
239,908	86,342	-167,203	185,599
86,342	-61,876	185,599	-75,090
-61,876	-148,911	-75,090	-62,621
-148,911	-183,629	-62,621	213,085
-183,629	-118,567	213,085	-108,399
-118,567	-4,482	-108,399	-208,276
-4,482	236,004	-208,276	-342,881

Расчёты коэффициента автокорреляции Лаг 1

Используя функцию Excel СУММПРОИЗВ, вычисляем сумму произведений остатков и сдвинутых на лаг остатков. В результате получили -264498,9

Используя функцию Excel СУММКВ вычисляем сумму квадратов остатков и сумму квадратов остатков, сдвинутых на лаг. В результате получили соответственно значения 4046734,89 и 4139591,875.

Далее рассчитываем произведение суммы квадратов остатков и суммы квадратов остатков сдвинутых на лаг. В результате получили 10^{13} .

В результате расчетное значение коэффициента автокорреляции $r_{e_i e_{i-1}}$ равно -0,065. Таким образом, связь практически отсутствует.

Аналогично формируется массив данных по остаткам и сдвинутым остаткам на лаг 2.

Таблица 77. Остатки $r_{e_i e_{i-2}}$

e_i	e_{i-2}	e_i	e_{i-2}
-157,195	10,940	236,004	94,614
-78,971	-258,792	36,224	-108,816
10,940	-99,685	94,614	-181,717
-258,792	121,609	-108,816	1308,350
-99,685	-110,756	-181,717	-44,379
121,609	-34,665	1308,350	126,550
-110,756	-15,269	-44,379	-151,764
-34,665	44,248	126,550	-19,266
-15,269	-117,313	-151,764	281,409
44,248	-51,141	-19,266	-94,343
-117,313	-90,772	281,409	-1,684
-51,145	-74,364	-94,343	166,257
-90,772	-38,202	-1,684	-64,261
-74,364	-48,870	166,257	-82,422
-38,202	132,359	-64,261	-170,213
-48,871	-742,793	-82,422	200,235
132,359	-64,371	-170,213	640,208
-742,793	20,713	200,235	-15,436
-64,371	-6,782	640,208	-71,951
20,713	-11,575	-15,436	344,140
-6,782	7,401	-71,951	161,792
-11,575	-53,990	344,140	-30,798
7,401	37,720	161,792	206,650
-53,990	11,119	-30,798	5,198
37,720	211,272	206,650	98,642
11,119	-120,518	5,198	-84,606
211,272	-36,049	98,642	60,002

-120,518	138,874	-84,606	-33,059
-36,049	-5,415	60,002	140,167
138,874	-107,059	-33,059	-148,106
-5,415	103,169	140,167	35,466
-107,059	-312,660	-148,106	1,058
103,169	-34,653	35,466	-235,166
-312,660	-23,670	1,058	8,128
-34,653	239,908	-235,166	-167,203
-23,670	86,342	8,128	185,599
239,908	-61,876	-167,203	-75,090
86,342	-148,911	185,599	-62,621
-61,876	-183,629	-75,090	213,085
-148,911	-118,567	-62,621	-108,399
-183,629	-4,482	213,085	-208,276
-118,567	236,004	-108,399	-342,881
-4,482	36,224		

Расчёты коэффициента автокорреляции Лаг 2

Используя функцию Excel СУММПРОИЗВ вычислили сумму произведений остатков и сдвинутых на лаг остатков. В результате получили -138079,5

Используя функцию Excel СУММКВ вычислили сумму квадратов остатков и сумму квадратов остатков сдвинутых на два лага. В результате получили 4003355,827 и 4133355,527.

Произведение суммы квадратов остатков и суммы квадратов остатков сдвинутых на лаг, В результате получили 10^{13} . Итоговый коэффициент автокорреляции r_{eie_i-2} равен -0,034. Таким образом, связь практически отсутствует.

Изучаем форму взаимосвязи остатков и остатков, сдвинутых на лаг 1 с помощью парной линейной регрессии.

$$e_i = f(e_{i-1})$$

Используя пакет Анализа данных, найдем параметры модели рисунок 1.

Рис. 22. Регрессия для коэффициента корреляции e_{i-1}

Таблица 78. Регрессионная статистика для e_{i-1}

Регрессионная статистика	
Множественный R	0,065
R-квадрат	0,004
Нормированный R-квадрат	-0,008
Стандартная ошибка	218,991
Наблюдения	86

Таблица 79. Дисперсионный анализ для e_{i-1}

	df	SS	MS	F	Значимость F
Регрессия	1	16981,503	16981,503	0,354	0,553
Остаток	84	4028386,326	47956,980		
Итого	85	4045367,829			

Таблица 80. Продолжение вывода дисперсионного анализа для e_{i-1}

	Коэффициенты	Стандартная ошибка	t-статистика	P-Значение
У-пересечение	4,104	23,615	0,174	0,862
	-0,064	0,108	-0,595	0,553

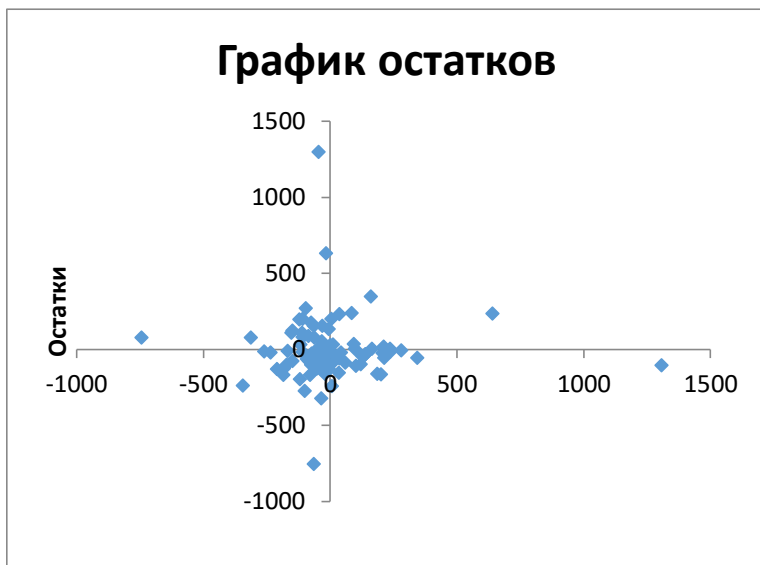


Рис. 23. График остатков для модели $e_i = f(e_{i-1})$

Таким образом, модель не значима. параметры не значимы, так как Значимость-F много больше чем 0,05. Модель статистически не значима.

Критерии Дарбина Уотсона вычисляется с помощью данной формулы:

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

Числитель равен 8715324,599. А знаменатель можно вычислить с помощью функции СУММКВ. Знаменатель равен 4046734,89.

Таким образом, коэффициент Дарбина Уотсона равен 2,154.

8. Множественная линейная регрессия

Лабораторная работа № 10.

Цель работы: Построить модель множественной линейной регрессии МЛР

Выполнение работы.

С помощью МЛР исследовать зависимость фондоотдачи в процентах на единицу ОПФ (Y) от среднечасовой производительности детали (X_1) и удельного веса активной части ОПФ (X_2 , %).

Таблица 81. Исходные данные

№	Y	X_1	X_2
1	26	37	39
2	33	33	40
3	24	15	35
4	29	36	48
5	42	26	53
6	24	24	42
7	52	15	54
8	56	33	54
9	26	44	50
10	45	34	53
11	27	63	46
12	54	8	50
13	34	44	43
14	48	43	55
15	45	31	51

Таблица 82. Матрица X

1	37	39
1	33	40
1	15	35
1	36	48
1	26	53
1	24	42
1	15	54
1	33	54
1	44	50
1	34	53
1	63	46
1	8	50
1	44	43
1	43	55
1	31	51

Таблица 83. Матрица Хтрансп

1	1	1	1	1	1	1	1	1
37	33	15	36	26	24	15	33	44
39	40	35	48	53	42	54	54	50

Таблица 83 продолжение

1	1	1	1	1	1
34	63	8	44	43	31
53	46	50	43	55	51

Таблица 84. Матрица Хтрансп*Х

15	486	713
486	18416	23132
713	23132	34455

Таблица 85. Матрица $(X_{тр} * X)^{(-1)}$

4,408	-0,011	-0,084
-0,011	0,001	-0,001
-0,084	-0,001	0,002

Таблица 86. Матрица $(X_{тр} * X)^{(-1)} * X_{тр}$

0,729	0,691	1,310	-	-	0,624	-	-0,481
0,002	0,001	0,006	0,001	0,003	0,003	0,007	0,001
-	-	-	0,001	0,009	0,009	0,012	0,011
0,015	0,013	0,022					

Таблица Продолжение 86

-		-				
0,269	-0,409	0,1468	0,133	0,317	-0,67	-0,208
0,004	0,001	0,012	-0,009	0,004	0,004	-0,001
0,004	0,009	-0,003	0,005	-0,008	0,013	0,006

Таблица 87. Расчетные значения параметров модели

$b = (X_{тр} * X)^{(-1)} * X_{тр} * Y$	
b0	-20,414
b1	-0,314
b2	1,436

Модель множественной линейной регрессии имеет вид:

$$Y = -20,4137 - 0,31361 * X_1 + 1,435654 * X_2$$

Таблица 88. Проверка на значимость уравнения регрессии

У		X ₁	X ₂	У _{мод}	(У- У _{мод})^2	(У _{мод} - У _{ср})^2	(У- У _{ср})^2
26	1	37	39	23,973	4,108	187,513	136,111
33	1	33	40	26,663	40,155	121,075	21,778
24	1	15	35	25,129	1,277	157,168	186,778
29	1	36	48	37,208	67,365	0,211	75,111
42	1	26	53	47,522	30,493	97,128	18,778
24	1	24	42	32,357	69,840	28,192	186,778
52	1	15	54	52,407	0,166	217,289	205,444
56	1	33	54	46,762	85,333	82,732	336,111
26	1	44	50	37,570	133,866	0,009	136,111
45	1	34	53	45,013	0,001	53,971	53,778
27	1	63	46	25,869	1,279	139,189	113,778
54	1	8	50	48,860	26,419	125,292	266,778

3							
4	1	44	43	27,520	41,984	102,945	13,444
4							
8	1	43	55	45,060	8,632	54,689	106,778
4							
5	1	31	51	43,083	3,677	29,333	53,778

$Y_{cp}=37,667$ $Q_e=514,595$ $Q_R=1396,739$ $Q=1911,333$

$Q_R/m=698,369$ $Q_e/k_2=42,883$ $F_{набл}=16,286$ $F_{кр}=3,491$

Вывод. $F_{набл} > F_{кр}$ Уравнение регрессии статистически значимо.

Расчет коэффициента детерминации

Используем полученные выше значения

$Q_e=514,595$ $Q_R=1396,739$ $Q=1911,333$

$R^2=Q_R/Q$, $Q=Q_R+Q_e$

$R^2=0,731$

Для проверки на значимость параметров модели необходимо вычислить диагональные элементы матрицы $Se^2 \cdot (X_{тр} \cdot X)^{(-1)}$. В таблице 106 представлена Матрица $Se^2 \cdot (X_{тр} \cdot X)^{(-1)}$

Таблица 89. Матрица $Se^2 \cdot (X_{тр} \cdot X)^{(-1)}$

189,008	-0,479	-3,589
-0,479	0,016	-0,001
-3,589	-0,001	0,076

Далее используем регрессионную статистику, которая представлена в таблице 90.

Далее рассчитываем наблюдаемое и критическое

Таблица 90. Регрессионная статистика

Параметры	b_0	b_1	b_2
Значения	-20,414	-0,314	435653816
Дисперсия	189,008	0,016	0,076
Стандартная ошибка	13,748	0,127	0,276
$t_{\text{набл}}$	-1,485	-2,474	5,204

значения статистики для параметров модели при $\alpha=0,05$, степени свободы $k_2=12$

$$/t_{\text{набл}}() = 1,485 \quad /t_{\text{набл}}(b_1) = 2,474$$

$/t_{\text{набл}}(b_2) = 5,204$. Вывод параметр b_0 статистически не значим, параметры b_1 и b_2 статистически значимы. $t_{\text{кр}}=2,179$. Рассчитывается с помощью функции =СТЮДЕНТ.ОБР.2X(α ; k_2)

Проведем проверку с использованием функции ЛИНЕЙН(У;Х;1;1). В таблице 110 представлен вывод данной функции.

Таблица 91. Вывод с применением функции ЛИНЕЙН(У;Х;1;1)

b_2	b_1	b_0
1,436	-0,314	-20,414
0,276	0,127	13,748
0,731	6,549	#Н/Д
16,286	12	#Н/Д
1396,739	514,594	#Н/Д
1911,333		

Далее рассмотрим описание полученных числовых значений.

Таблица 92. Описание вывода значений функции линейн. (y;x;1;1)

Значения параметров модели		b_0	b_1	b_2
		1,436	-0,314	-20,414
Стандартные ошибки	S(bj)	0,276	0,127	13,748
Коэффициент детерминации	R^2	0,731	6,549	Se несмещенная оценка стандартной ошибки
Проверка на значимость у регрессии	Fнабл	16,286	12	k^2 – число степеней свободы
Q_R вариативность, обусловленная регрессией		1396,739	514,595	Q_e Остаточная вариативность

Рассмотрим решение задачи эконометрического моделирования через встроенный пакет Анализа данных «Регрессия».

В таблице 93 представлен выводной листинг пакета Анализ данных «Регрессия»

Таблица 93. Вывод пакета «Анализ данных» Регрессия

Регрессионная статистика		
Множественный R	0,855	Множественный коэффициент корреляции
R-квадрат	0,731	Коэффициент детерминации
Нормированный R-квадрат	0,686	Скорректированный коэффициент детерминации
Стандартная ошибка	6,549	Se
Наблюдения	15	Объем выборки

Таблица 94.Продолжение выводного листинга пакета «Анализ данных»
Регрессия

Дисперсионный анализ	Степени свободы	Значения Q	MS=SS /df (Q/k)	Fнабл=MS (Q _R)/MS(Q _e)	Уровень значимости
Виды вариативности	df	SS	MS	F	Значимость F
Регрессия (Q _R)	2	1396,739	698,369	16,286	0,001
Остаток (Q _e)	12	514,595	42,883	Вывод	
Итого Q ₀ =Q _R +Q _e	14	1911,333	Уравнение регрессии статистически значимо		

Таблица 95. Продолжение выводного листинга пакета «Анализ данных»

Регрессия

	Параметры модели b_j	$S(b_j)$	$t=b_j/S(b_j)$	Уровень значимости
	Коэффициенты	Стандартная ошибка	t-статистика	P-Значение
Y-пересечение	-20,414	13,748	-1,485	0,163
X_1	-0,314	0,127	-2,474	0,029
X_2	1,436	0,276	5,204	0,001

Таблица 96. Интервальная оценка параметров модели b_j (доверительные интервалы)

Нижние 95%	Верхние 95%	Нижние 90,0%	Верхние 90,0%
-50,368	9,541	-44,917	4,089
-0,589	-0,037	-0,539	-0,088
0,835	2,037	0,944	1,927

Таблица 97. Продолжение выводного листинга пакета «Анализ данных»

Регрессия. Вывод остатка

	Модельное значение	$Y_i - Y_{\text{мод}}$	
Наблюдение	Предсказанное Y	Остатки	Стандартные остатки
1	23,973	2,027	0,334
2	26,663	6,338	1,045
3	25,129	-1,129	-0,186
4	37,208	-8,208	-1,354
5	47,522	-5,522	-0,911
6	32,357	-8,357	-1,378

Продолжение табл. 97

7	52,407	-0,407	-0,067
8	46,762	9,238	1,524
9	37,570	-11,570	-1,908
10	45,013	-0,013	-0,002
11	25,869	1,131	0,187
12	48,860	5,139	0,848
13	27,520	6,479	1,069
14	45,062	2,938	0,485
15	43,083	1,917	0,316

Таблица 98. Продолжение выводного листинга пакета «Анализ данных»
Регрессия. Вывод вероятности

Персентиль	У	Персентиль	У
3,333	24	56,667	42
10	24	63,333	45
16,667	26	70	45
23,333	26	76,667	48
30	27	83,333	52
36,667	29	90	54
43,333	33	96,667	56
50	34		

Используя функцию Процентиль.вкл ($y; 0,5$), можно рассчитать Персентили.

Библиографический список

1. Кулинич, Е.И. «Эконометрия». М.: Финансы и статистика, 2001. – 304 с.
2. Магнус, Я.Р. «Эконометрика». М.: Дело, 2001. – 399 с.
3. Мардас, А.Н. «Эконометрика». СПб.: Питер, 2001. – 304 с.
4. Кремер, Н.Ш. «Эконометрика». М.: ЮНИТИ-ДАНА, 2008. – 311 с.
5. Кузнецова, О.А. «Эконометрическое моделирование». Самара: Изд-во Самарского гос. аэрокосмического ун-та, 2012. – 203 с.
6. Зеткина, О. В. «Эконометрика: основы математического моделирования социально-экономических процессов». Ярославль: ЯрГУ, 2013. – 260 с.
7. Валентинов В.А.: Эконометрика. – М.: Дашков и К, 2010. – 416 с.
8. Валентинов В.А.: Эконометрика: Практикум. – М.: Дашков и К, 2010. – 288 с.
9. Гладилин А.В.: Эконометрика. – М.: КНОРУС, 2009. – 272 с.
10. Бывшев В.А.: Эконометрика. – М.: Финансы и статистика, 2008. – 480 с.
11. Чураков Е.П.: Прогнозирование эконометрических временных рядов. – М.: Финансы и статистика, 2008. – 272 с.
12. Яковлева А.В.: Эконометрика. – М.: ЭКСМО, 2008. – 320 с.
13. Елисеева И.И., Курышева С.В., Костеева Т.В. и др.; под ред. чл.-корр. РАН Елисеевой И.И.: Эконометрика. – М.: Финансы и статистика, 2007. – 576 с.
14. Кочетыгов А.А.: Основы эконометрики. – Ростов на/Д: МарТ, 2007. – 192 с.
15. Шелобаев С.И.: Математические методы и модели в экономике, финансах, бизнесе. – М.: ЮНИТИ-ДАНА, 2000. – 367 с.
16. Ежеманская С.Н.: Эконометрика. – Ростов н/Д: Феникс, 2003. – 192 с.
17. Магнус Я.Р.: Эконометрика. Начальный курс. – М.: Дело, 2007. – 504 с.

Учебное издание

Трусова Алла Юрьевна
Ильина Алла Ивановна

ЭКОНОМЕТРИЧЕСКОЕ МОДЕЛИРОВАНИЕ

Практикум

Редакционно-издательская обработка
издательства Самарского университета

Подписано в печать 11.12.2025. Формат 60х84 1/16.

Бумага офсетная. Печ. л. 6,0.

Тираж 27 экз. Заказ № . Арт. – 5(P2ПР)/2025.

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«САМАРСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ ИМЕНИ АКАДЕМИКА С.П. КОРОЛЕВА»
(САМАРСКИЙ УНИВЕРСИТЕТ)
443086, САМАРА, МОСКОВСКОЕ ШОССЕ, 34.

Издательство Самарского университета.
443086, Самара, Московское шоссе, 34.