

ОБРАБОТКА ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ В MS EXCEL

Хабаровск

2012

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное бюджетное учреждение

высшего профессионального образования

«Тихоокеанский государственный университет»

ОБРАБОТКА ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ В MS EXCEL

Методические указания к выполнению лабораторных работ
для студентов дневной формы обучения

Хабаровск
Издательство ТОГУ
2012

Обработка экспериментальных данных в MS Excel : методические указания к выполнению лабораторных работ для студентов дневной формы обучения / сост. Е. Г. Агапова, Е. А. Битехтина. – Хабаровск : Изд-во Тихоокеан. гос. ун-та, 2012. – 32 с.

Методические указания содержат описание наиболее распространенных методов обработки результатов наблюдений, решений основных задач парного и множественного регрессионного анализа. Все методы снабжены подробными комментариями и рассматриваются на простых и конкретных примерах, что дает возможность использовать работу в качестве справочника по приведенным методам.

Методические указания предназначены для студентов, обучающихся по направлению «Прикладная математика», и могут быть использованы студентами других направлений при изучении дисциплин «Обработка экспериментальных данных на ЭВМ», «Математическая статистика», «Эконометрика».

Печатается в соответствии с решениями кафедры «Прикладная математика» и методического совета факультета компьютерных и фундаментальных наук.

ОБРАБОТКА ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ В MS EXCEL

Методические указания к выполнению лабораторных работ
для студентов дневной формы обучения

Агапова Елена Григорьевна

Битехтина Екатерина Андреевна

Главный редактор *Л. А. Суевалова*

Редактор *Л. С. Бакаева*

Подписано в печать Формат 60x84 1/16.

Бумага писчая. Гарнитура «Таймс». Печать цифровая.

Усл. печ. л. Тираж 100 экз. Заказ

Издательство Тихоокеанского государственного университета.

680035, Хабаровск, ул. Тихоокеанская, 136.

Отдел оперативной полиграфии издательства Тихоокеанского государственного университета.

680035, Хабаровск, ул. Тихоокеанская, 136.

ОБЩИЕ СВЕДЕНИЯ

Методы обработки экспериментальных данных (ОЭД) начали разрабатываться более двух веков тому назад в связи с необходимостью решения практических задач по агробиологии, медицине, экономике, социологии. Полученные при этом результаты составили фундамент такой научной дисциплины, как математическая статистика. Первыми работами, положившими начало статистике как самостоятельной науке, были книги Дж. Граунта «Естественные и политические наблюдения, перечисленные в прилагаемом оглавлении и сделанные над бюллетенями смертности ...» (1662 г.) и У. Пети «Два очерка по политической арифметике, относящиеся к людям, зданиям, больницам Лондона и Парижа» (1682 г.).

Современный уровень – естественнонаучного эксперимента характеризуется большими потоками информации. При этом визуальный просмотр данных, не говоря уже об анализе, невозможен без применения ЭВМ. Обработка результатов экспериментов предполагает знание основных понятий и методов теории вероятностей и математической статистики. Выявление характерных классов задач в ОЭД и стандартных методов их решения позволяет выделить обработку результатов экспериментов из многообразия задач прикладной статистики.

Как правило, основным подходом в решении многих задач является метод наименьших квадратов (МНК) в его различных модификациях. Однако МНК эффективно работает только для линейных моделей, а на практике встречаются ситуации, когда связь искомого параметра с измеряемой величиной сугубо нелинейная. В этом случае применяют нелинейный МНК или другие методы обработки. Знакомство со всеми этими методами расширяет арсенал средств, находящихся в распоряжении обработчика, что особенно важно в сложных случаях, например, когда измерения производятся при воздействии большого числа факторов, мешающих их проведению.

Появление электронных таблиц (табличных процессоров) привело к тому, что статистические методы, ранее доступные лишь узкому кругу математиков, стали использоваться широким кругом специалистов разных областей. Дальнейшее развитие программного обеспечения привело к созданию большого количества прикладных пакетов по статистике. Удобной универсальной вычислительной средой для решения задач ОЭД является табличный процессор *MS Excel*. Поэтому основной целью данных методических указаний является изложение (в форме лабораторных работ) численных методик решения основных задач парного и множественного регрессионного анализа в вычислительной среде табличного процессора *MS Excel*. Жирным шрифтом в работе выделены команды меню, названия панелей инструментов и диалоговых окон и их элементов.

Тема 1. ОЦЕНКА ПОКАЗАТЕЛЕЙ КАЧЕСТВА ОБЪЕКТА ПО РЕЗУЛЬТАТАМ ЭКСПЕРИМЕНТОВ

В этой теме рассмотрены основные задачи математической статистики о первичной статистической обработке данных. В статистике, как правило, статистические данные являются результатами наблюдений над некоторой случайной величиной X .

Пример. В результате эксперимента были получены значения переменной X (табл. 1.1).

Таблица 1.1

i	1	2	3	4	5	6	7	8	9	10
X	14,85	14,80	14,84	14,81	14,63	14,81	14,80	14,85	14,84	14,80

Лабораторная работа 1.1

СТАТИСТИЧЕСКИЕ ХАРАКТЕРИСТИКИ СЛУЧАЙНЫХ ВЕЛИЧИН

Цель работы. Научиться вычислять оценки основных числовых характеристик по экспериментальным данным.

Статистическая функция **СЧЕТ** используется для определения числа значений (рис. 1.1).

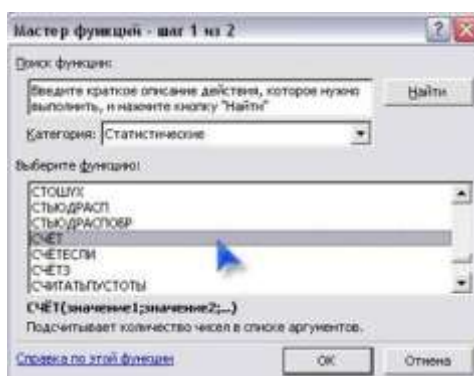


Рис. 1.1.

Диапазон ячеек указывается адресами первой и последней ячейки данных, записанными через двоеточие, например **B6:B15** (рис. 1.2).

Функция **СРЗНАЧ** рассчитывает среднее значение выборки.

Функция **СТАНДОТКЛОН** – стандартное отклонение выборки. Аргументами этих функций служит все тот же диапазон ячеек.

Статистическая функция **СТЮДРАСПОБР** используется для нахождения коэффициента Стьюдента. Этот коэффициент зависит от вероятности ошибки (при обычно задаваемой надежности 95 % вероятность ошибки составляет 5 %) и от числа степеней свободы $n-1$. Также это значение можно найти по таблице

критических значений t -критерия (прил. 1).

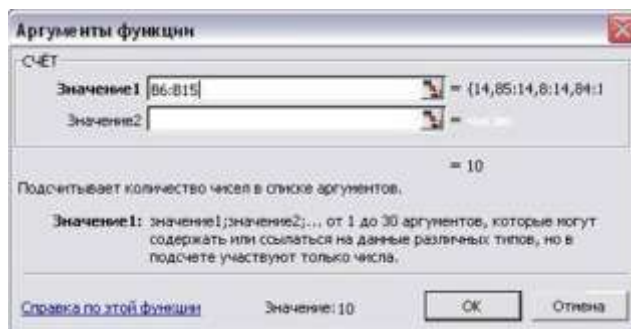


Рис. 1.2.

Для нахождения доверительного интервала используется обычная формула умножения «=СТЮДРАСПОБР* X_{cp} ».

Решение. Данные из примера введем в столбец **В** (табл. 1.2). В столбцах **Д** и **Е** – подсказки характеристик, которые мы будем рассчитывать. Используя приведенные выше функции, в столбец **Ф** поместим результаты. По найденному значению ячейки **F11** окончательный результат доверительного интервала можно записать так: с 95 %-ной надежностью $X = 14,81 \pm 0,046$. В заключение вычислим относительную ошибку определения доверительного интервала: $\delta = ДИ/X_{cp}$ (формула: «=F11/F7»). Значение относительной ошибки обычно выражают в процентах, в нашем случае 0,3 %.

Таблица 1.2

	А	В	С	Д	Е	Ф
1						
2						
3	Данные эксперимента			Обработка		
4						
5	№	X				
6	1	14,85	Число значений n	СЧЕТ		10
7	2	14,80	Среднее значение X_{cp}	СРЗНАЧ		14,803
8	3	14,84	Станд. отклонение S	СТАНДОТКЛОН		0,0643
9	4	14,81	Ст. откл. среднего S_{cp}	=S / КОРЕНЬ(n)		0,0203
10	5	14,63	К.Стюд (5%, n-1) t	СТЮДРАСПОБР		2,2622
11	6	14,81	Доверит. интервал ДИ	= t * S_{cp}		0,046
12	7	14,80	Относит. ошибка δ	= ДИ / X_{cp}		0,0031
13	8	14,85				
14	9	14,84				
15	10	14,80				

Лабораторная работа 1.2

ОЦЕНКА ПАРАМЕТРОВ И ОПРЕДЕЛЕНИЕ ЗАКОНА РАСПРЕДЕЛЕНИЯ

Пример. Исследуется случайная величина X – число правонарушений в течение одних суток в некотором городе N . Получены данные за первые 150

суток года.

3	5	4	4	5	8	2	3	1	6	6	1	2	5	5
4	4	4	3	4	5	5	2	2	3	4	3	2	4	4
8	10	1	4	3	3	2	5	7	5	3	6	7	5	6
1	4	6	4	5	4	5	7	6	5	3	5	5	8	7
7	5	5	4	5	3	3	6	3	5	2	2	2	6	2
5	6	8	4	4	8	3	6	4	4	5	5	7	5	5
3	5	4	5	5	4	7	6	9	3	3	5	6	6	3
4	5	2	6	7	5	5	4	2	5	4	2	6	2	7
5	5	8	5	3	5	2	5	3	7	4	6	3	6	0
4	4	4	5	2	7	7	3	1	1	3	6	5	7	6

Цель работы. Требуется провести первичную статистическую обработку данных, проверить гипотезу о виде распределения случайной величины с помощью критерия согласия Пирсона.

Расчетные соотношения. Данная задача решается с помощью статистических процедур **Анализа данных** и статистических функций библиотеки встроенных функций *MS Excel*. Приведем алгоритм решения задачи.

1. **Ввод данных.** В диапазон ячеек **A1:AN** ввести выборочные значения x_1, x_2, \dots, x_N .

2. **Построение вариационного ряда.** Скопировать содержимое ячеек **A1:AN** в ячейки **B1:BN**. Упорядочить выборочные значения, используя кнопку сортировки по возрастанию.

3. **Построение статистического ряда выборки.** В ячейки **C1:CK** ввести k различных выборочных значений. В меню **Данные** выделить строку **Анализ данных**, выделить процедуру **Гистограмма**. В поле **Входной интервал** диалогового окна **Гистограмма** ввести ссылку на диапазон **A1:AN**. В поле **Интервал карманов** ввести ссылку на диапазон **C1:CK**. Активизировать поле **Выходной интервал** и ввести в это поле ссылку – левая верхняя ячейка, в которую будет введена таблица результатов решений. Установить флажок **Вывод графика**. Составить табл. 1.3 статистического ряда по следующему образцу:

Таблица 1.3

x_i различные выборочные значения	n_i частота выборочного значения x_i	$\frac{n_i}{N}$ относительная частота выборочного значения x_i	$\frac{n_i^*}{N}$ накопленная относительная частота

Первые столбцы заполнить копированием. Относительные и накопленные частоты вычислить с использованием формул.

4. **Построение полигонов относительных и накопленных относительных частот.**

Скопировать первый и третий столбцы табл. 1.3. Выделить их. Используя меню **Вставка**, применить к выделенным числам средство диаграммы **Точечная**. Полученный график есть полигон относительных частот. Если эти же действия проделать с первым и четвертым столбцами табл. 1.3, то получим полигон накопленных частот – сглаженный график эмпирической функции распределения.

5. *Определение выборочных характеристик.* В меню **Данные** выделить подменю **Анализ данных**, выделить процедуру **Описательная статистика**, в поле ввода **Входной интервал** ввести ссылку на диапазон ячеек, содержащий статистические данные **A1:AN**. Установить флажок **Итоговая статистика**. Активизировать поле **Выходной интервал**, ввести в это поле ссылку – левая верхняя ячейка, в которую будет введена таблица результатов решений.

6. *Проверка гипотезы о виде распределения случайной величины с помощью критерия согласия Пирсона.* Заполнить табл. 1.4.

Таблица 1.4

x_i различные выборочные значения	n_i частота выборочного значения x_i	p_i теоретическая вероятность выборочного значения x_i	$n'_i = N \cdot p_i$ теоретическая частота выборочного значения x_i	$\frac{(n'_i - n_i)^2}{n'_i}$
--	---	--	---	-------------------------------

Первые столбцы заполнить копированием, а оставшиеся – вычисленными по формулам значениями.

Если проверяется гипотеза *о распределении Пуассона*, то теоретические вероятности p_i вычислить с помощью функции **ПУАССОН**($x_i, \bar{x}, 0$). Здесь \bar{x} – выборочное среднее, оно определяется в пункте 5, **0** – параметр, показывающий, что вычисляется вероятность того, что случайная величина, распределенная по закону Пуассона, принимает значение x_i .

Если проверяется гипотеза *о биномиальном распределении* случайной величины, то теоретические вероятности p_i вычислить с помощью функции **БИНОМРАСП**($x_i, N, P, 0$), при этом вероятность успеха **P** в одном испытании определить по формуле $P = \frac{\bar{x}}{N}$, где \bar{x} – выборочное среднее.

В случае других распределений воспользоваться справкой о статистических функциях библиотеки встроенных функций *MS Excel*.

Значение

$$\chi^2 = \sum_{i=1}^k \frac{(n'_i - n_i)^2}{n'_i}$$

является наблюдаемым значением случайной величины χ^2 . Число степеней свободы этой случайной величины равно $r = k - 2$ при проверке гипотезы о распределении Пуассона и $r = k - 3$, если проверяется гипотеза о биномиальном распределении. Критическое значение случайной величины χ^2 определить с помощью функции $\chi^2_{кр} = \text{ХИ2ОБР}(\alpha, r)$, где α – уровень

значимости. Полученное наблюдаемое значение χ^2 сравнить с $\chi^2_{кр}$. Если $\chi^2 < \chi^2_{кр}$, то гипотеза о виде распределения принимается при уровне значимости α . Если $\chi^2 > \chi^2_{кр}$, то гипотеза отвергается с уровнем значимости α .

Решение. По предложенному алгоритму проведем первичную статистическую обработку данных. Согласно пункту 3 алгоритма находим $k = 11$ различных выборочных значений

0 1 2 3 4 5 6 7 8 9 10

С помощью пакета **Анализ данных** получаем статистический ряд выборки и его графическое представление (рис. 1.3).



Рис. 1.3.

Построенная гистограмма позволяет сделать предположение о виде распределения случайной величины X . В результате заполнения табл. 1.3 получим табл. 1.5, в третьем столбце которой, представлены относительные, а в четвертом – накопленные относительные частоты выборочных значений.

Таблица 1.5

x_i	n_i	n_i/N	n_i^*/N
0	1	0,006667	0,006667
1	6	0,04	0,046667
2	16	0,106667	0,153333
3	22	0,146667	0,3
4	27	0,18	0,48
5	38	0,253333	0,733333
6	19	0,126667	0,86
7	13	0,086667	0,946667
8	6	0,04	0,986667
9	1	0,006667	0,993333
10	1	0,006667	1

Согласно пункту 4 алгоритма получим полигоны относительных и накопленных частот (рис. 1.4, 1.5).

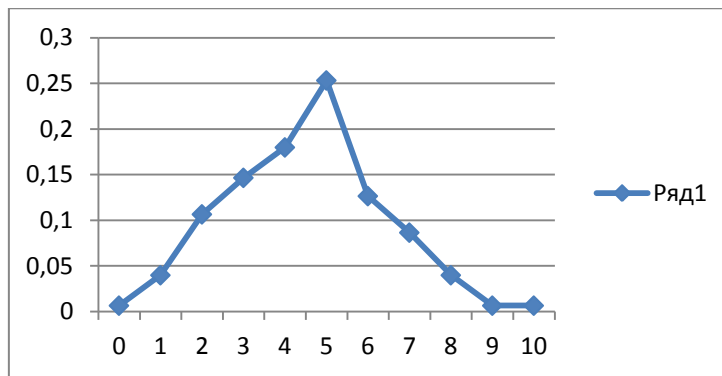


Рис. 1.4.

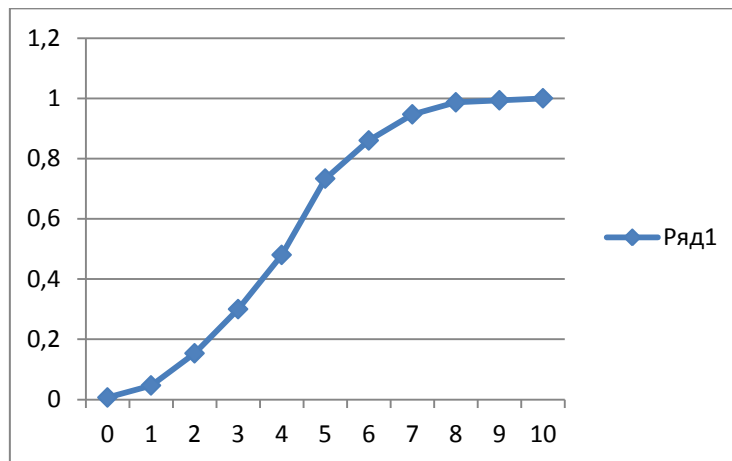


Рис.1.5.

Согласно пункту 5 алгоритма получаем выборочные характеристики (табл. 1.6).

Таблица 1.6

Среднее	4,493333
Стандартная ошибка	0,150687
Медиана	5
Мода	5
Стандартное отклонение	1,845534
Дисперсия выборки	3,405996
Экцесс	-0,12004
Асимметричность	0,130829
Интервал	10
Минимум	0
Максимум	10
Сумма	674
Счет	150

Проверим гипотезу о распределении случайной величины по закону Пуассона. В качестве точечной оценки параметра распределения выбираем выборочное

среднее $\lambda = \bar{x} = 4,49$. В результате заполнения табл. 1.4 получим табл. 1.7. Наблюдаемое значение случайной величины $\chi^2 = 9,261528$. Оно получено суммированием чисел последнего столбца табл. 1.6. Критическое значение $\chi^2_{кр} = \text{ХИ2ОБР}(\alpha, r) = 16,91898$, где $\alpha = 0,05$, $r = k - 2 = 11 - 2 = 9$. Так как $\chi^2 < \chi^2_{кр}$, то гипотеза о распределении по закону Пуассона при уровне значимости $\alpha = 0,05$ не противоречит опытными данным.

Таблица 1.7

xi	ni	pi	ni'	хи2
0	1	0,011183	1,677501	0,273626
1	6	0,05025	7,537566	0,313644
2	16	0,112896	16,93439	0,051557
3	22	0,169093	25,36393	0,446146
4	27	0,189948	28,49213	0,078142
5	38	0,170699	25,6049	6,00035
6	19	0,127835	19,17521	0,001601
7	13	0,082058	12,30865	0,038831
8	6	0,046089	6,913355	0,120667
9	1	0,02301	3,451554	1,741278
10	1	0,010339	1,550897	0,195685
				9,261528

Тема 2. ЛИНЕЙНАЯ ПАРНАЯ РЕГРЕССИЯ

Эта тема включает выполнение лабораторных работ, посвященных построению и исследованию уравнения линейной регрессии вида

$$\hat{y} = f(x) = a + bx. \quad (2.1)$$

Построение линейной регрессии сводится к оценке ее параметров a , b . Классический подход к оцениванию параметров линейной регрессии основан на **методе наименьших квадратов** (МНК).

Пример. Для определения зависимости между сменной добычей угля на одного рабочего (переменная Y , измеряемая в тоннах) и мощностью угольного пласта (переменная X , измеряемая в метрах) на 10 шахтах были проведены исследования, результаты которых представлены в табл. 2.1.

Таблица 2.1

i	1	2	3	4	5	6	7	8	9	10
X	8	11	12	9	8	8	9	9	8	12
Y	5	10	10	7	5	6	6	5	6	8

Лабораторная работа 2.1

ВЫЧИСЛЕНИЕ КОЭФФИЦИЕНТОВ УРАВНЕНИЯ ЛИНЕЙНОЙ РЕГРЕССИИ

Цель работы. Вычисление коэффициентов уравнения линейной регрессии по пространственной выборке табл. 2.1.

Расчетные соотношения. Коэффициенты, определяемые на основе МНК, являются решением системы уравнений

$$\begin{cases} na + b \sum x = \sum y, \\ a \sum x + b \sum x^2 = \sum xy. \end{cases} \quad (2.2)$$

Решая эту систему уравнений, получим

$$a = \bar{y} - b\bar{x}, \quad (2.3)$$

$$b = \frac{\overline{xy} - \bar{x}\bar{y}}{x^2 - (\bar{x})^2} = \frac{cov(x, y)}{s_x^2}, \quad (2.4)$$

где $cov(x, y) = \overline{xy} - \bar{x}\bar{y}$ – выборочная ковариация; s_x^2 – выборочное значение дисперсии величины x , определяемой по формуле:

$$s_x^2 = \overline{x^2} - (\bar{x})^2. \quad (2.5)$$

Решение. Вычислим значения, используя данные табл. 2.2:

$$\bar{x} = 9,4; \quad \bar{y} = 6,8; \quad \overline{x^2} = 90,8; \quad \overline{xy} = 66,4.$$

Таблица 2.2

А	В	С	Д	Е	Ф	Г	Н	І	Ј	К
№ п/п	x	y	xy	x^2	y^2	$(y - \bar{y})$	$(y - \bar{y})^2$	\hat{y}_i	$(y - \hat{y}_i)$	$(y - \hat{y}_i)^2$
1	8	5	40	64	25	-1,8	3,24	5,38	0,38	0,1429
2	8	5	40	64	25	-1,8	3,24	5,38	0,38	0,1429
3	8	6	48	64	36	-0,8	0,64	5,38	-0,62	0,3869
4	8	6	48	64	36	-0,8	0,64	5,38	-0,62	0,3869
5	9	7	63	81	49	0,2	0,04	6,39	-0,61	0,3672
6	9	6	54	81	36	-0,8	0,64	6,39	0,39	0,1552
7	9	5	45	81	25	-1,8	3,24	6,39	1,39	1,9432
8	11	10	110	121	100	3,2	10,24	8,43	-1,57	2,4775
9	12	10	120	144	100	3,2	10,24	9,44	-0,56	0,3114
10	12	8	96	144	64	1,2	1,44	9,44	1,44	2,0794
Сумма	94	68	664	908	496	0	33,6	68	0	8,3934
Среднее	9,4	6,8	66,4	90,8	49,6					

Коэффициенты a , b вычислим по формулам (2.3), (2.4) соответственно:

$$a = -2,75; \quad b = 1,02,$$

а само уравнение регрессии (2.1) примет вид

$$\hat{y} = f(x) = -2,75 + 1,02x. \quad (2.6)$$

Лабораторная работа 2.2

ВЫЧИСЛЕНИЕ ВЫБОРОЧНОГО КОЭФФИЦИЕНТА КОРРЕЛЯЦИИ

Цель работы. Вычисление выборочного коэффициента корреляции по пространственной выборке табл. 2.1.

Расчетные соотношения. Выборочный коэффициент корреляции определяется соотношением

$$r_{XY} = \frac{\overline{x \cdot y} - \bar{x} \cdot \bar{y}}{s_X \cdot s_Y}, \quad (2.7)$$

где

$$s_X = \sqrt{\overline{x^2} - (\bar{x})^2}; \quad s_Y = \sqrt{\overline{y^2} - (\bar{y})^2}; \quad \bar{y}^2 = \frac{1}{n} \sum_{i=1}^n y_i^2. \quad (2.8)$$

Решение. Используя вычисления лабораторной работы 2.1, данные табл. 2.2 и формулы (2.7), (2.8), получим:

$$s_x = 1,56; \quad \bar{y}^2 = 49,6; \quad s_y = 1,83; \quad r_{xy} = \frac{66,4 - 9,4 \cdot 6,8}{1,56 \cdot 1,83} = \frac{2,48}{2,86} = 0,87.$$

Величина $r_{xy} = 0,87$ говорит о сильной положительной линейной связи.

Лабораторная работа 2.3

ПРОВЕРКА ЗНАЧИМОСТИ УРАВНЕНИЯ ЛИНЕЙНОЙ РЕГРЕССИИ ПО КРИТЕРИЮ ФИШЕРА

Цель работы. По данным табл. 2.1 оценить при уровне $\alpha = 0,05$ значимость уравнения регрессии (2.6), построенного в лабораторной работе 2.1.

Критерий Фишера (F-критерий). Уравнение парной регрессии значимо при уровне значимости α , если выполняется следующее неравенство:

$$F = \frac{Q_r(n-2)}{Q_e} > F_{1-\alpha; 1; n-2}. \quad (2.9)$$

Величины Q_r , Q_e являются факторной и остаточной суммами квадратов соответственно:

$$Q_r = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2, \quad Q_e = \sum_{i=1}^n (\hat{y}_i - y_i)^2. \quad (2.10)$$

Величина $F_{1-\alpha; 1; n-2}$ – табличное значение F -распределения с числами степеней свободы $k_1 = 1$ и $k_2 = n - 2$ квантиля уровня $\gamma = 1 - \alpha$ (прил. 2). Эту вероятность можно также определить с помощью функции **ФРАСПОБР**:

$$F_{1-\alpha; 1; n-2} = \text{ФРАСПОБР}(\alpha; 1; n-2). \quad (2.11)$$

Значение F для *линейной* парной регрессии можно вычислить через коэффициент корреляции:

$$F = \frac{r_{xy}^2}{1 - r_{xy}^2} (n - 2). \quad (2.12)$$

Величина $r_{xy}^2 = R^2$ называется коэффициентом детерминации и характеризует долю дисперсии результативного признака y , объясняемую регрессией, в общей дисперсии результативного признака:

$$R^2 = 1 - \frac{Q_e}{Q}, \quad \text{где } Q = \sum_{i=1}^n (y_i - \bar{y})^2. \quad (2.13)$$

Соответственно величина $1 - r_{xy}^2$ характеризует долю дисперсии y , вызванную влиянием остальных, не учтенных в модели факторов.

Решение. Вычислим значения Q_e , $Q_r = Q - Q_e$ и критерий F по данным табл. 2.2. В столбце **I** значения вычисляются по формуле (2.6). Итак, получены следующие значения:

$$Q_e = 8,39, \quad Q_r = Q - Q_e = 33,6 - 8,39 = 25,21,$$

$$F = \frac{Q_r(n - 2)}{Q_e} = \frac{25,21 \cdot 8}{8,39} = 24,03.$$

По формуле (2.11) при $k_1 = 1$, $k_2 = 8$ или по таблице (прил. 2) вычисляем квантиль $F_{0,95; 1; 8} = 5,32$. Неравенство (2.9) выполняется, т. е. $24,03 > 5,32$, и поэтому уравнение регрессии (2.6) значимо.

Лабораторная работа 2.4

ВЫЧИСЛЕНИЕ КОЭФФИЦИЕНТОВ УРАВНЕНИЯ ЛИНЕЙНОЙ РЕГРЕССИИ

Цель работы. Вычисление коэффициентов уравнения линейной регрессии.

В вычислительной среде табличного процессора *MS Excel* эта задача решается при помощи статистических функций **НАКЛОН** (наклон прямой относительно оси X , коэффициент b) и **ОТРЕЗОК** (отрезок, отсекаемый прямой на оси Y , коэффициент a).

Для знакомства с этими возможностями введем необходимые исходные данные (табл. 2.3). В столбцах **B** и **C** вводятся данные табл. 2.1, записи в столбце **E** играют роль подсказок, столбец **F** заполняется по мере обработки. В ячейку **F3** вводится функция **НАКЛОН**, в ячейку **F4** – **ОТРЕЗОК**. Обе эти функции имеют два аргумента: диапазон ячеек со значениями Y (**C3:C12**) и диапазон ячеек со значениями X (**B3:B12**).

Статистическая функция **КВПИРСОН** вычисляет значение коэффициента детерминации.

Таблица 2.3

	A	B	C	D	E	F
1	Линейная зависимость			Обработка		
2	№	X	Y			
3	1	8	5	НАКЛОН	1,0164	
4	2	11	10	ОТРЕЗОК	-2,754	
5	3	12	10	КВПИРСОН	0,7502	
6	4	9	7			
7	5	8	5			
8	6	8	6			
9	7	9	6			
10	8	9	5			
11	9	8	6			
12	10	12	8			
13						

Функция **ЛИНЕЙН**(изв_знач_y; изв_знач_x; константа; стат) вычисляет коэффициенты линейной регрессии, коэффициент детерминации R^2 , F -статистику.

В поле «изв_знач_y» вводится диапазон значений Y (C3:C12); «изв_знач_x» – диапазон значений X (B3:B12); константа устанавливается на 0, если заранее известно, что свободный член равен 0, и на 1 в противном случае; стат устанавливается на 0, если не нужен вывод дополнительных сведений регрессионного анализа, и на 1 в противном случае.

Порядок использования функции **ЛИНЕЙН**:

1. Выделить область пустых ячеек 5×2 (5 строк, 2 столбца) для вывода результатов регрессионной статистики и 1×2 для вывода только коэффициентов a , b .

2. Ввести функцию **ЛИНЕЙН** вручную или через **Мастер функций**.

3. После корректного ввода функции в левой верхней ячейке выделенной таблицы появится первый итоговый элемент таблицы. Чтобы раскрыть всю таблицу, следует сначала нажать клавишу **F2**, а затем одновременно нажать клавиши **[Ctrl]**, **[Shift]**, **[Enter]**. Далее появляется следующая регрессионная статистика, представленная в табл. 2.4.

Таблица 2.4

Значение коэффициента b	Значение коэффициента a
Среднеквадратическое отклонение b	Среднеквадратическое отклонение a
Коэффициент детерминации R^2	Среднеквадратическое отклонение y
F -статистика	Число степеней свободы
Регрессионная сумма квадратов	Остаточная сумма квадратов

Решение. В результате выполнения вышеуказанных действий получим табл. 2.5. Значения в табл. 2.5 совпадают со значениями, полученными в лабораторных работах 2.1, 2.3.

Таблица 2.5

Значение коэффициента b	1,0164	-2,754	Значение коэффициента a
Среднеквадратическое отклонение b	0,2074	1,9759	Среднеквадратическое отклонение a
Коэффициент детерминации R^2	0,7502	1,0243	Среднеквадратическое отклонение y
F -статистика	24,025	8	Число степеней свободы
Регрессионная сумма квадратов	25,207	8,3934	Остаточная сумма квадратов

Лабораторная работа 2.5

АВТОКОРРЕЛЯЦИЯ ОСТАТКОВ. СТАТИСТИКА ДАРБИНА–УОТСОНА

Цель работы. Научиться пользоваться статистикой Дарбина–Уотсона.

Одна из предпосылок МНК – это независимость значений случайных отклонений от значений отклонений во всех других наблюдениях. Автокорреляция (последовательная корреляция) – это корреляция между наблюдаемыми показателями, упорядоченными во времени (временные ряды) или в пространстве (перекрестные ряды). Нам неизвестны истинные значения отклонений $e_i = y_i - f(x_i)$, $i = 1, \dots, n$. Поэтому выводы об их независимости делаются на основе оценок e_i , $i = 1, \dots, n$. При этом обычно проверяется некоррелированность только соседних величин. При наличии автокорреляции остатков полученное уравнение регрессии считается неудовлетворительным.

Критерий Дарбина–Уотсона – наиболее известный способ обнаружения автокорреляции первого порядка. Пусть n – число наблюдений, k – число факторов модели, уровень значимости α . Для n , k , α по таблицам распределения Дарбина–Уотсона (прил. 3) находим числа d_l и d_u . Вычисляем статистику Дарбина–Уотсона

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

Здесь $e_i = y_i - (ax_i + b)$ – остатки, вычисленные по уравнению линейной регрессии $y = a + bx$ для статистических данных (x_i, y_i) .

Если $DW < d_l$, то это свидетельствует о положительной автокорреляции остатков. Если $DW > 4 - d_l$, то это свидетельствует об отрицательной автокорреляции остатков. При $d_u < DW < 4 - d_u$ гипотеза об отсутствии автокорреляции остатков принимается. Если

$$d_l < DW < d_u \text{ или } 4 - d_u < DW < 4 - d_l,$$

то гипотеза об отсутствии автокорреляции остатков не может быть ни принята, ни отвергнута.

Решение. Вернемся к примеру. Определим наличие автокорреляции с помощью критерия Дарбина–Уотсона. В столбце **K** табл. 2.2 получили

$$\sum_{i=1}^n e_i^2 = 8,39.$$

Во второй столбец e_i табл. 2.6 введем данные столбца **J** табл. 2.2. Из каждого числа 2-го столбца вычитаем предыдущее число этого же столбца и результат введем в третий столбец $e_i - e_{i-1}$. Статистика Дарбина–Уотсона равна

$$DW = \frac{\sum_{i=2}^n (e_i - e_{i-1})^2}{\sum_{i=1}^n e_i^2} = \frac{16,78}{8,39} = 2.$$

Таблица 2.6

№ п/п	e_i	$e_i - e_{i-1}$	$(e_i - e_{i-1})^2$
1	0,38	–	–
2	0,38	0	0
3	–0,62	–1,00	1
4	–0,62	0	0
5	–0,61	0,01	0
6	0,39	1	1
7	1,39	1	1
8	–1,57	–2,96	8,76
9	–0,56	1,01	1,02
10	1,44	2	4
Сумма			16,78

По таблице распределения Дарбина–Уотсона (прил. 3) находим числа $d_l = 0,88$ и $d_u = 1,32$ при $n = 10$, $k = 1$. Так как $d_u < DW < 4 - d_u$ ($1,32 < 2 < 2,68$), то гипотеза об отсутствии автокорреляции остатков не отклоняется на уровне значимости $0,05$. Это является одним из подтверждений высокого качества модели.

Тема 3. НЕЛИНЕЙНАЯ ПАРНАЯ РЕГРЕССИЯ

Эта тема включает выполнение лабораторных работ, посвященных построению уравнения нелинейной парной регрессии.

Пример. В табл. 3.1 приведены значения независимой переменной X (товaroоборот, тыс. р.) и значения зависимой переменной Y (уровень издержек обращения по отношению к товарообороту, %).

Таблица 3.1

X	7	10	15	20	30	45	60	120
Y	10	9	7,2	6	6,3	5,8	5,4	5

Лабораторная работа 3.1

ПОСТРОЕНИЕ НЕЛИНЕЙНОЙ РЕГРЕССИИ

Цель работы. Построить уравнение нелинейной регрессии и вычислить индекс детерминации R^2 .

Расчетные соотношения. В случае нелинейных зависимостей с помощью некоторых преобразований можно перейти к линейным. Приведем некоторые из нелинейных функций и их модификацию к линейной функции – табл. 3.2.

Таблица 3.2

№ п/п	Исходная спецификация	Преобразование $x \rightarrow x^*$	Преобразование $y \rightarrow y^*$	Вычисление b по b^*	Вычисление a по a^*
1	$y = a + \frac{b}{x} + \varepsilon$	$x^* = \frac{1}{x}$	$y^* = y$	$b = b^*$	$a = a^*$
2	$y = \frac{1}{a + bx + \varepsilon}$	$x^* = x$	$y^* = \frac{1}{y}$	$b = b^*$	$a = a^*$
3	$y = \frac{x}{a + bx + x\varepsilon}$	$x^* = \frac{1}{x}$	$y^* = \frac{1}{y}$	$b = a^*$	$a = b^*$
4	$y = ae^{bx + \varepsilon}$	$x^* = x$	$y^* = \ln y$	$b = b^*$	$a = e^{a^*}$
5	$y = ae^{\frac{b}{x} + \varepsilon}$	$x^* = \frac{1}{x}$	$y^* = \ln y$	$b = b^*$	$a = e^{a^*}$
6	$y = \frac{1}{a + be^{-x} + \varepsilon}$	$x^* = e^{-x}$	$y^* = \frac{1}{y}$	$b = b^*$	$a = a^*$
7	$y = ax^b e^\varepsilon$	$x^* = \ln x$	$y^* = \ln y$	$b = b^*$	$a = e^{a^*}$

Решение. На рис. 3.1 изображены данные в виде диаграммы. Из этого рисунка видно, что зависимость между показателями нелинейная. Будем искать регрессионную зависимость в виде степенной: $y = ax^b$; переходя к логарифмам

(модификация табл. 3.2), получим линейную зависимость $y^* = a^* + bx^*$, где $y^* = \ln y$, $x^* = \ln x$.

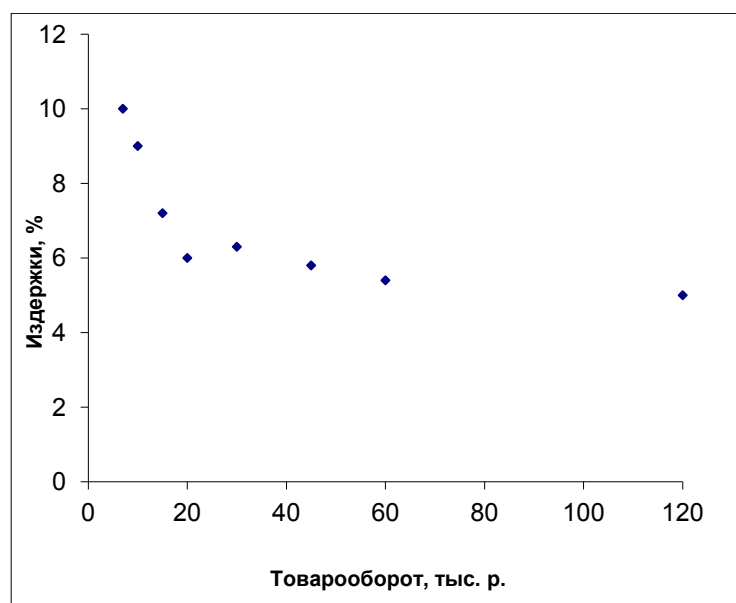


Рис. 3.1.

Определяем коэффициенты a^* , b для модифицированных данных. Исходные данные и основные вычисления приведены в табл. 3.3.

Таблица 3.3

№ п/п	x	$x^* = \ln x$	$y^* = \ln y$	x^2	y^2	$(y - \bar{y}_x)^2$	\hat{y}_x	$(y - \hat{y}_x)^2$
1	7	1,95	2,30	3,79	5,30	10,00	9,12	0,77
2	10	2,30	2,20	5,30	4,83	4,68	8,37	0,40
3	15	2,71	1,97	7,33	3,90	0,13	7,59	0,15
4	20	3,00	1,79	8,97	3,21	0,70	7,08	1,17
5	30	3,40	1,84	11,57	3,39	0,29	6,42	0,02
6	45	3,81	1,76	14,49	3,09	1,08	5,82	0,00
7	60	4,09	1,69	16,76	2,84	2,07	5,43	0,00
8	120	4,79	1,61	22,92	2,59	3,38	4,60	0,16
Итого		26,04	15,16	91,14	29,15	22,32	-	2,67

Аналогично лабораторной работе 2.1 определим коэффициенты a^* и b из соотношений (2.3), (2.4): $a^* \approx 2,68$; $b \approx -0,24$. По величине $a^* = 2,68$ потенцированием находим исходный коэффициент a : $a = e^{2,68} \approx 14,58$.

Таким образом, уравнение регрессии имеет вид: $y = 14,58 x^{-0,24}$. Изобразим эту кривую на диаграмме с исходными данными (рис. 3.2).

Аналогично лабораторной работе 2.3 вычислим

$$Q_e = 2,67; \quad Q_r = Q - Q_e = 22,32 - 2,67 = 19,65.$$

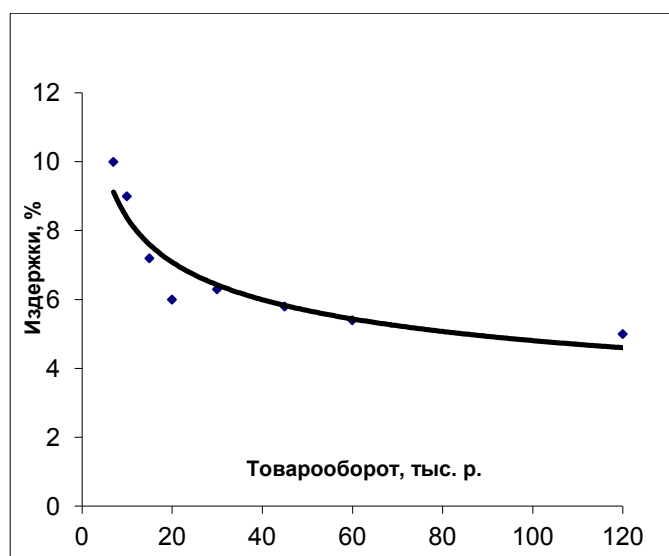


Рис. 3.2.

Согласно формуле (2.13) индекс детерминации равен

$$R^2 = 1 - \frac{Q_e}{Q} = 1 - \frac{2,67}{22,32} = 1 - 0,12 = 0,88.$$

Вычисленное значение говорит о том, что уравнением регрессии объясняется 88 % дисперсии результативного признака, а на долю прочих факторов приходится лишь 12 %.

Лабораторная работа 3.2

ПОСТРОЕНИЕ НЕЛИНЕЙНОЙ РЕГРЕССИИ С ИСПОЛЬЗОВАНИЕМ КОМАНДЫ «ДОБАВИТЬ ЛИНИЮ ТРЕНДА»

Цель работы. Построить уравнение нелинейной регрессии с использованием команды «Добавить линию тренда».

Команда «Добавить линию тренда» используется для выделения тренда (медленных изменений) при анализе временных рядов. Однако эту команду можно использовать и для построения уравнения нелинейной регрессии, рассматривая в качестве времени t независимую переменную x .

Эта команда позволяет построить следующие уравнения регрессии:

- линейная $\hat{y} = a + bx$;
- полиномиальная $\hat{y} = a + b_1x + b_2x^2 + b_3x^3 + b_4x^4 + b_5x^5 + b_6x^6$;
- логарифмическая $\hat{y} = a + b$;
- степенная $\hat{y} = ax^b$;
- экспоненциальная $\hat{y} = ae^{bx}$.

Для построения одной из перечисленных регрессий необходимо выполнить следующие шаги:

Шаг 1. Ввести по столбцам исходные данные $\{x_i, y_i\}$, $i = 1, 2, \dots, n$ (рис. 3.3).

Шаг 2. По этим данным построить график в декартовой системе координат (рис. 3.3).

Шаг 3. Установить курсор на любую точку построенного графика, сделать щелчок правой кнопкой и в появившемся контекстном меню выполнить команду **Добавить линию тренда** (рис. 3.3).

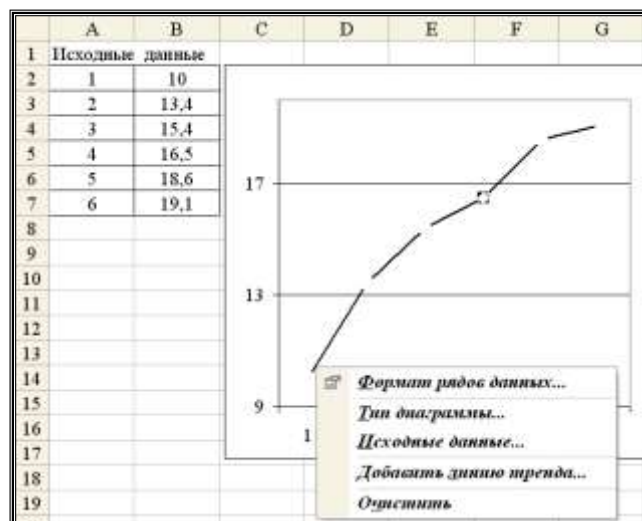


Рис. 3.3.

Шаг 4. В появившемся диалоговом окне выбрать нужное уравнение регрессии.

Шаг 5. «Включить» необходимые опции:

- **«Показать уравнение на диаграмме»** – на диаграмме будет показано выбранное уравнение регрессии с вычисленными коэффициентами;

- **«Поместить на диаграмму величину достоверности аппроксимации (R^2)»** – на диаграмме будет показано значение коэффициента детерминации R^2 (для нелинейной регрессии – индекс детерминации).

Если по построенному уравнению регрессии необходимо выполнить прогноз, то нужно указать число периодов прогноза.

Шаг 6. После задания всех перечисленных опций на диаграмме появится формула построенного уравнения регрессии и значение индекса детерминации R^2 .

Решение. По данным табл. 3.1 построить степенное уравнение регрессии $\hat{y} = ax^b$. После выполнения перечисленных шагов получим уравнение $\hat{y} = 14,58 x^{-0,24}$, для которого индекс детерминации равен $R^2 = 0,877$ (рис. 3.4). Такая величина говорит о хорошем соответствии построенного уравнения исходным данным.

	x	y
1	7	10
2	10	9
3	15	7,2
4	20	6
5	30	6,3
6	45	5,8
7	60	5,4
8	120	5

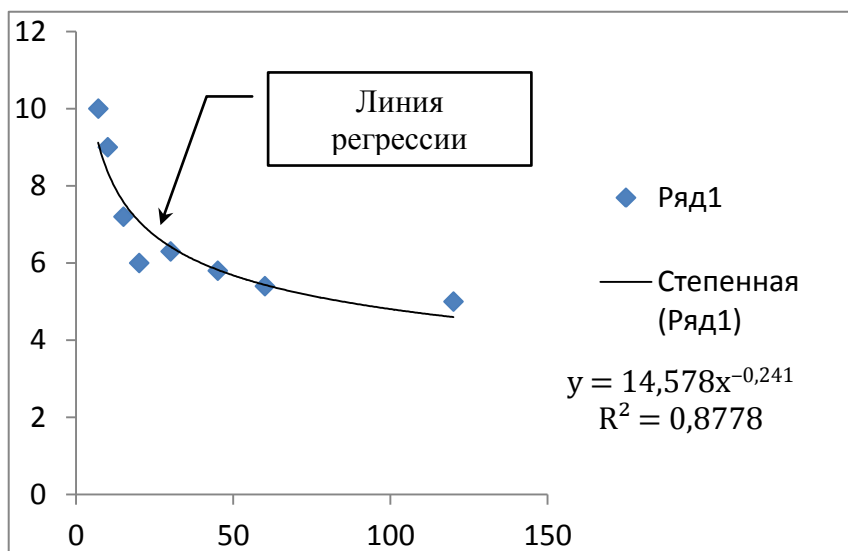


Рис. 3.4.

Лабораторная работа 3.3

ВЗВЕШЕННЫЙ МЕТОД НАИМЕНЬШИХ КВАДРАТОВ

Цель работы. Освоение применения взвешенного метода наименьших квадратов для коррекции гетероскедастичности остатков.

Расчетные соотношения. После применения обычного МНК выясняется гетероскедастичность остатков: стандартное отклонение остатков линейно увеличивается при увеличении независимой переменной. Необходимо применить модификацию взвешенного МНК для коррекции такой гетероскедастичности.

Решение. Построить точечную диаграмму исходных данных, поместить на нее линию тренда, его уравнение и индекс детерминации.

1. Сформировать массив остатков.

2. Модифицировать массив независимой переменной следующим образом:
 $x_i^* = 1/x_i$.

3. Модифицировать массив зависимой переменной следующим образом:
 $y_i^* = y_i/x_i$.

4. По полученным данным модифицированной регрессии x_i^* , y_i^* построить диаграмму облака рассеяния, поместить на нее линию тренда с уравнением линии регрессии и индексом детерминации. Сравнить это уравнение с уравнением исходной регрессии.

Тема 4. ЛИНЕЙНАЯ МНОЖЕСТВЕННАЯ РЕГРЕССИЯ

Эта тема включает выполнение лабораторных работ, посвященных построению и исследованию уравнения линейной множественной регрессии вида

$$\hat{y}(x_1, x_2) = b_0 + b_1 x_1 + b_2 x_2. \quad (4.1)$$

Пример. Данные о сменной добыче угля на одного рабочего (переменная Y , измеряемая в тоннах), мощности пласта (переменная X_1 , измеряемая в метрах) и об уровне механизации работ в шахте (переменная X_2 , измеряемая в процентах), характеризующие процесс добычи угля в 10 шахтах, приведены в табл. 4.1.

Таблица 4.1

Номер шахты i	X_1	X_2	Y
1	8	5	5
2	11	8	10
3	12	8	10
4	9	5	7
5	8	7	5
6	8	8	6
7	9	6	6
8	9	4	5
9	8	5	6
10	12	7	8

Предполагая, что между переменными Y , X_1 , X_2 существует линейная зависимость, необходимо найти аналитическое выражение для этой зависимости, т. е. построить уравнение линейной регрессии.

Лабораторная работа 4.1

ВЫЧИСЛЕНИЕ КОЭФФИЦИЕНТОВ ЛИНЕЙНОЙ МНОЖЕСТВЕННОЙ РЕГРЕССИИ

Цель работы. Для пространственной выборки табл. 4.1 необходимо вычислить вектор коэффициентов $b = \begin{pmatrix} b_0 \\ b_1 \\ b_2 \end{pmatrix}$ уравнения регрессии (4.1).

Расчетные соотношения. Вектор коэффициентов, найденный методом наименьших квадратов, является решением следующей системы уравнений:

$$X^T X b = X^T Y,$$

где X – матрица размерности 10×3 , первый столбец которой составлен из 1, а другие два столбца составлены из значений x_{i1} , x_{i2} , т. е. матрица X имеет структуру

$$X = \begin{pmatrix} 1 & 8 & 5 \\ 1 & 11 & 8 \\ \dots & \dots & \dots \\ 1 & 12 & 7 \end{pmatrix},$$

а Y – вектор, составленный из 10 значений y_i , т. е.

$$Y = \begin{pmatrix} 5 \\ 10 \\ \dots \\ 8 \end{pmatrix}.$$

Матрица $X^T X$ имеет обратную матрицу $(X^T X)^{-1}$, и тогда вектор коэффициентов равен

$$b = (X^T X)^{-1} (X^T Y). \quad (4.2)$$

Для реализации этой матричной формулы необходимо выполнить следующие операции: транспонирование; умножение матриц (частный случай – умножение матрицы на вектор); вычисление обратной матрицы. Все эти операции можно реализовать с помощью матричных функций *MS Excel* категории функций **Ссылки и массивы**. Для работы с этими функциями можно либо обратиться к **Мастеру функций** и выбрать нужную категорию функций, затем указать имя функции и задать соответствующие диапазоны ячеек, либо непосредственно ввести с клавиатуры имя функции и задать соответствующие диапазоны ячеек.

Функция **ТРАНСП(диапазон ячеек)** осуществляет транспонирование матрицы, где параметр **диапазон ячеек** задает все элементы транспонируемой матрицы (или вектора).

Функция **МУМНОЖ(диапазон_1; диапазон_2)** осуществляет умножение матриц в категории функций **Математические**. Параметр **диапазон_1** задает элементы первой из перемножаемых матриц, а параметр **диапазон_2** – элементы второй матрицы. При этом перемножаемые матрицы должны иметь соответствующие размерности (если первая матрица размерности $n \times k$, вторая – $k \times m$, то результатом будет матрица размерности $n \times m$).

Функция **МОБР(диапазон ячеек)** осуществляет обращение матрицы (вычисление обратной матрицы) в категории функций **Математические**. Параметр **диапазон ячеек** задает все элементы обращаемой матрицы, которая должна быть квадратной и невырожденной.

При использовании этих функций необходимо соблюдать следующий порядок действий:

1. Выделить фрагмент ячеек, в которые будет занесен результат выполнения матричных функций (при этом надо учитывать размеры исходных матриц);
2. Ввести арифметическое выражение, содержащее обращение к матричным функциям *MS Excel*;

3. Одновременно нажать клавиши [Ctrl], [Shift], [Enter]. Если этого не сделать, то вычислится только один элемент результирующей матрицы или вектора.

Решение. Сформируем матрицу X и вектор Y . Затем выполним формирование матрицы $X^T X$, вектора $X^T Y$ и вычисление вектора $b = (b_0, b_1, b_2)^T$ по формуле (4.2). В результате получим вектор коэффициентов $b = \begin{pmatrix} -3,5393 \\ 0,8539 \\ 0,3670 \end{pmatrix}$ и тогда уравнение регрессии (4.1) примет вид

$$\hat{y}(x_1, x_2) = -3,5393 + 0,8539 x_1 + 0,367 x_2. \quad (4.3)$$

Лабораторная работа 4.2

ВЫЧИСЛЕНИЕ КОЭФФИЦИЕНТОВ ЛИНЕЙНОЙ МНОЖЕСТВЕННОЙ РЕГРЕССИИ И ПРОВЕРКА ЗНАЧИМОСТИ В РЕЖИМЕ РЕГРЕССИЯ

Цель работы. Используя режим **Регрессия**, вычислить вектор коэффициентов уравнения регрессии (4.1).

Табличный процессор *MS Excel* содержит модуль **Анализ данных**. Этот модуль позволяет выполнить статистический анализ выборочных данных (построение гистограмм, вычисление числовых характеристик и т. д.). Режим работы **Регрессия** этого модуля осуществляет вычисление коэффициентов линейной множественной регрессии с k переменными, построение доверительных интервалов и проверку значимости уравнения регрессии.

После вызова режима **Регрессия** на экране появляется диалоговое окно (рис. 4.1), в котором задаются следующие параметры:

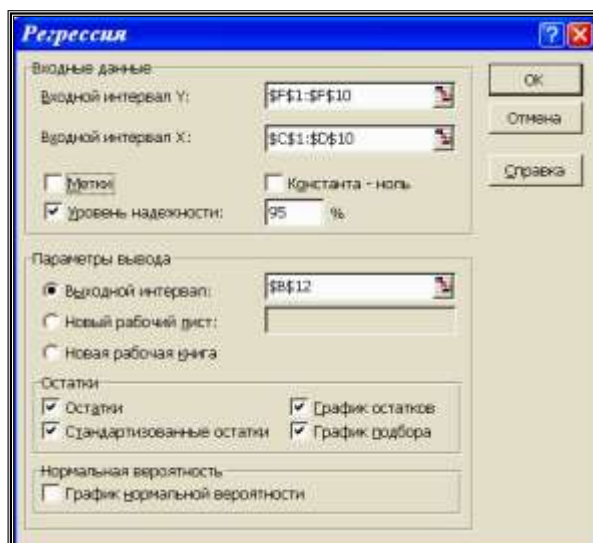


Рис. 4.1.

1. *Входной интервал Y* – вводится диапазон адресов ячеек, содержащих значения y_i (ячейки должны составлять один столбец).

2. *Входной интервал X* – вводится диапазон адресов ячеек, содержащих значения независимых переменных. Значения каждой переменной представляются одним столбцом. Количество переменных – не более 16.

3. *Метки* – включается, если первая строка во входном диапазоне содержит заголовок. В этом случае автоматически будут созданы стандартные названия.

4. *Уровень надежности* – при включении этого параметра задается надежность γ при построении доверительных интервалов.

5. *Константа-ноль* – при включении этого параметра коэффициент $b_0 = 0$.

6. *Выходной интервал* – при включении активизируется поле, в которое необходимо ввести адрес левой верхней ячейки выходного диапазона, который содержит ячейки с результатами вычислений режима **Регрессия**.

7. *Новый рабочий лист* – при включении этого параметра открывается новый лист, в который, начиная с ячейки **A1**, вставляются результаты работы режима **Регрессия**.

8. *Новая рабочая книга* – при включении этого параметра открывается новая книга, на первом листе которой, начиная с ячейки **A1**, вставляются результаты работы режима **Регрессия**.

9. *Остатки* – при включении вычисляется столбец, содержащий невязки $y_i - \hat{y}_i, i = 1, \dots, n$.

10. *Стандартизованные остатки* – при включении вычисляется столбец, содержащий стандартизованные остатки.

11. *График остатков* – при включении выводятся точечные графики невязки $y_i - \hat{y}_i, i = 1, \dots, n$ в зависимости от значений переменных $x_j, j = 1, \dots, k$. Количество графиков равно числу k переменных x_j .

12. *График подбора* – при включении выводятся точечные графики предсказанных по построенной регрессии значений \hat{y}_i от значений переменных $x_j, j = 1, \dots, k$. Количество графиков равно числу k переменных x_j .

Решение. Первоначально введем в столбец **C** десять значений первой переменной X_1 , в столбец **D** – десять значений второй переменной X_2 , а в столбец **F** – десять значений зависимой переменной Y .

После этого вызовем режим **Регрессия** и в диалоговом окне зададим необходимые параметры (рис. 4.1). Результаты работы приведены на рис. 4.2 – 4.3. Заметим, что из-за большой «ширины» таблиц, в которых выводятся результаты работы режима **Регрессия**, часть результатов помещена в другие ячейки.

Дадим краткую интерпретацию показателям, значения которых вычисляются в режиме **Регрессия**. Первоначально рассмотрим показатели, объединенные названием *Регрессионная статистика* (рис. 4.2).

ВЫВОДИТОГОВ			
<i>Регрессионная статистика</i>			
Множественный R	0,9009		
R-квадрат	0,8116		
Нормированный R-квадрат	0,7578		
Стандартная ошибка	0,9509		
Наблюдения	10		
<i>Дисперсионный анализ</i>			
	<i>df</i>	<i>SS</i>	<i>MS</i>
Регрессия	2	27,2704	13,635
Остаток	7	6,3296	0,904
Итого	9	33,6000	
		<i>F</i>	<i>Значимость F</i>
		15,0794	0,0029

Рис. 4.2.

Множественный R – корень квадратный из коэффициента детерминации.
R-квадрат – коэффициент детерминации R^2 .

Нормированный R-квадрат – приведенный коэффициент детерминации

$$\hat{R}^2 = 1 - \frac{(n-1)Q_e}{(n-k)Q} = 1 - \frac{n-1}{n-k}(1-R^2).$$

Стандартная ошибка – оценка s для среднеквадратического отклонения σ .

Наблюдения – число наблюдений n .

Перейдем к показателям, объединенным названием *Дисперсионный анализ* (рис. 4.2).

Столбец df – число степеней свободы. Для строки *Регрессия* показатель равен числу независимых переменных $k_1 = k - 1$; для строки *Остаток* он равен $k_2 = n - (k_1 + 1) = n - k$; для строки *Итого* $k_1 + k_2 = n - 1$.

Столбец SS – сумма квадратов отклонений. Для строки *Регрессия* показатель равен величине Q_r (см. формулы (2.10)); для строки *Остаток* – равен величине Q_e (см. формулы (2.10)); для строки *Итого* $Q = Q_r + Q_e$.

Столбец MS – дисперсии, вычисленные по формуле $MS = \frac{SS}{df}$, т. е. дисперсия на одну степень свободы.

Столбец F – значение F_c , равное F -критерию Фишера, вычисленному по формуле

$$F_c = \frac{Q_r/k_1}{Q_e/k_2}.$$

Столбец *Значимость F* – значение уровня значимости, соответствующее вычисленной величине *F*-критерия и равное вероятности $P(F(k_1, k_2) \geq F_c)$, где $F(k_1, k_2)$ – случайная величина, подчиняющаяся распределению Фишера с k_1, k_2 степенями свободы. Если вероятность меньше уровня значимости α (обычно $\alpha = 0,05$), то построенная регрессия является значимой.

Перейдем к следующей группе показателей, объединенных в табл. 4.2.

Таблица 4.2

	<i>Коэффициенты</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>
<i>Y</i> -пересечение	-3,539	1,907	-1,8564
Переменная X_1	0,854	0,221	3,8726
Переменная X_2	0,367	0,243	1,5108
	<i>P-значение</i>	<i>Нижние 95 %</i>	<i>Верхние 95 %</i>
	0,1058	-8,0477	0,9690
	0,0061	0,3325	1,3753
	-0,1746	-0,2074	0,9415

Столбец *Коэффициенты* – вычисленные значения коэффициентов b_0, b_1, \dots, b_k , расположенных сверху вниз.

Столбец *Стандартная ошибка* – значения s_{b_j} , $j = 0, \dots, k$, вычисленные по формуле

$$s_{b_j} = \sqrt{s^2 \{(X^T X)^{-1}\}_{jj}}$$

Столбец *t-статистика* – значения статистик T_{b_j} .

Столбец *P-значение* – содержит вероятности случайных событий $P(t(n-k) \geq T_{b_j})$, где $t(n-k)$ – случайная величина, подчиняющаяся распределению Стьюдента с $n-k$ степенями свободы.

Если эта вероятность меньше уровня значимости α , то принимается гипотеза о значимости соответствующего коэффициента регрессии.

Из табл. 4.2 видно, что значимым коэффициентом является только коэффициент при x_1 : $b_1 = 0,854$, так как *P-значение* при x_1 равно 0,0061, что меньше α .

Столбцы *Нижние 95 %* и *Верхние 95 %* – соответственно нижние и верхние интервалы для оцениваемых коэффициентов β_j .

Перейдем к следующей группе показателей, объединенных в табл. 4.3.

Столбец *Наблюдение* – содержит номера наблюдений.

Столбец *Предсказанное Y* – значения \hat{y}_i , вычисленные по построенному уравнению регрессии.

Столбец *Остатки* – значения невязок $y_i - \hat{y}_i$.

Таблица 4.3

ВЫВОД ОСТАТКА			
<i>Наблюдение</i>	<i>Предсказанное Y</i>	<i>Остатки</i>	<i>Стандартные остатки</i>
1	5,127	-0,127	-0,152
2	8,790	1,210	1,443
3	9,644	0,356	0,424
4	5,981	1,019	1,215
5	5,861	-0,861	-1,027
6	6,228	-0,228	-0,272
7	6,348	-0,348	-0,415
8	5,614	-0,614	-0,732
9	5,127	0,873	1,041
10	9,277	-1,277	-1,523

В заключение рассмотрения результатов работы режима *Регрессия* приведем график невязок – остатков $y_i - \hat{y}_i$ при заданных значениях только второй переменной (рис. 4.3). Наличие чередующихся положительных и отрицательных значений невязок является косвенным признаком отсутствия систематической ошибки (неучтенной независимой переменной) в построенном уравнении регрессии.

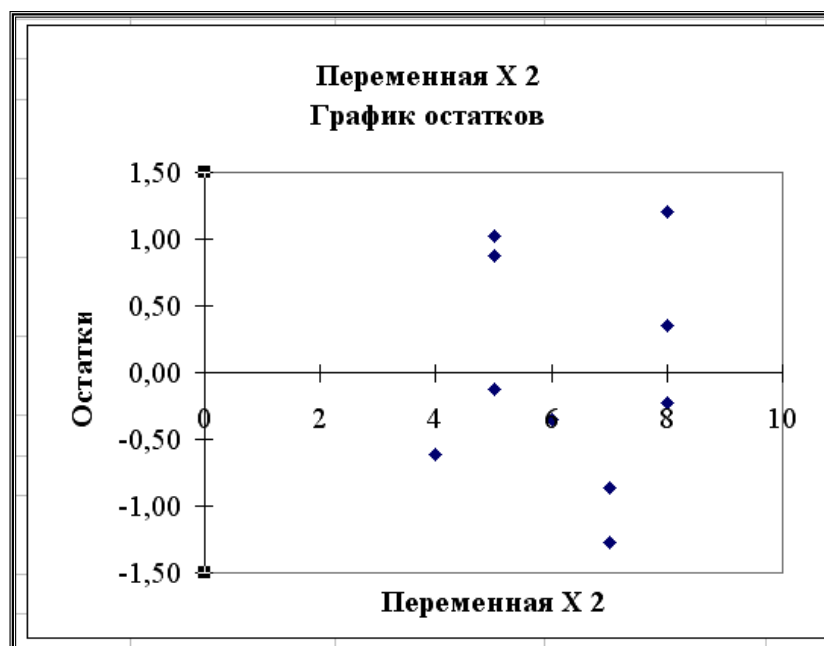


Рис. 4.3.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Брандт З.* Анализ данных. Статистические и вычислительные методы для научных работников и инженеров : учеб. пособие / З. Брандт ; пер. с англ. О. И. Волкова. – М. : Мир, 2003. – 686 с.
2. *Вадзинский Р.* Статистические вычисления в среде Excel. Библиотека пользователя Р. Вадзинский. – СПб. : Питер, 2008. – 608 с.
3. *Горелова Г. В.* Теория вероятностей и математическая статистика в примерах и задачах с применением Excel : учеб. пособие / Г. В. Горелова, И. А. Кацко. – 3-е изд., доп. и перераб. – Ростов н/Д : Феникс, 2005. – 480 с.
4. *Колемаев В. А.* Теория вероятностей и математическая статистика : учебник / В. А. Колемаев, В. Н. Калинина. – М. : ИНФРА-М, 2000. – 336 с.
5. *Лемешко Б. Ю.* Некоторые приложения, рекомендуемые к использованию в задачах прикладной математической статистики [Электронный ресурс] : учеб. пособие / Б. Ю. Лемешко. – Режим доступа: <http://ami.nstu.ru/~headrd/applied/index.html>.
6. *Минько А. А.* Статистический анализ в MS Excel / А. А. Минько. – М. : Издательский дом «Вильямс», 2004. – 448 с.
7. *Розанов Ю. А.* Теория вероятностей, случайные процессы и математическая статистика : учебник / Ю. А. Розанов. – 2-е изд., доп. – М. : Наука, 1989. – 312 с.
8. *Себер Дж.* Линейный регрессионный анализ / Дж. Себер ; пер. с англ. В. П. Носко. – М. : Мир, 1980. – 456 с.
9. *Тюрин Ю. Н.* Анализ данных на компьютере / Ю. Н. Тюрин, А. А. Макаров ; под ред. В. Э. Фигурнова. – 3-е изд., перераб. и доп. – М. : ИНФРА-М, 2003. – 544 с.
10. *Ходасевич Г. Б.* Обработка экспериментальных данных на ЭВМ. Ч. 1. Обработка одномерных данных [Электронный ресурс] : учеб. пособие / Г. Б. Ходасевич. – Режим доступа: http://www.dvo.sut.ru/libr/opds/i130/hodo_part1/4.htm.

Критические значения t -критерия Стьюдента

при уровне значимости 0,10; 0,05; 0,01 (двухсторонний)

Число степеней свободы d.f.	α			Число степеней свободы d.f.	α		
	0,10	0,05	0,01		0,10	0,05	0,01
1	6,3138	12,706	63,657	18	1,7341	2,1009	2,8784
2	2,9200	4,3027	9,9248	19	1,7291	2,0930	2,8609
3	2,3534	3,1825	5,8409	20	1,7247	2,0860	2,8453
4	2,1318	2,7764	4,5041	21	1,7207	2,0796	2,8314
5	2,0150	2,5706	4,0321	22	1,7171	2,0739	2,8188
6	1,9432	2,4469	3,7074	23	1,7139	2,0687	2,8073
7	1,8946	2,3646	3,4995	24	1,7109	2,0639	2,7969
8	1,8595	2,3060	3,3554	25	1,7081	2,0595	2,7874
9	1,8331	2,2622	3,2498	26	1,7056	2,0555	2,7787
10	1,8125	2,2281	3,1693	27	1,7033	2,0518	2,7707
11	1,7959	2,2010	3,1058	28	1,7011	2,0484	2,7633
12	1,7823	2,1788	3,0545	29	1,6991	2,0452	2,7564
13	1,7709	2,1604	3,0123	30	1,6973	2,0423	2,7500
14	1,7613	2,1448	2,9768	40	1,6839	2,0211	2,7045
15	1,7530	2,1315	2,9467	60	1,6707	2,0003	2,6603
16	1,7459	2,1199	2,9208	120	1,6577	1,9799	2,6174
17	1,7396	2,1098	2,8982	∞	1,6449	1,9600	2,5758

Таблица значений F -критерия Фишера при уровне значимости $\alpha = 0,05$

k_2	k_1									
	1	2	3	4	5	6	8	12	24	∞
1	161,5	199,5	215,7	224,6	230,2	233,9	238,9	243,9	249,0	254,3
2	18,51	19,00	19,16	19,25	19,30	19,33	19,37	19,41	19,45	19,50
3	10,13	9,55	9,28	9,12	9,01	8,94	8,84	8,74	8,64	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,04	5,91	5,77	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,82	4,68	4,53	4,36
6	5,99	5,14	4,76	4,53	4,39	4,28	4,15	4,00	3,84	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,73	3,57	3,41	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,44	3,28	3,12	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,23	3,07	2,90	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,07	2,91	2,74	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	2,95	2,79	2,61	2,40
12	4,75	3,88	3,49	3,26	3,11	3,00	2,85	2,69	2,50	2,30
13	4,67	3,80	3,41	3,18	3,02	2,92	2,77	2,60	2,42	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,70	2,53	2,35	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,64	2,48	2,29	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,59	2,42	2,24	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,55	2,38	2,19	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,51	2,34	2,15	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,48	2,31	2,11	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,45	2,28	2,08	1,84
21	4,32	3,47	3,07	2,84	2,68	2,57	2,42	2,25	2,05	1,81
22	4,30	3,44	3,05	2,82	2,66	2,55	2,40	2,23	2,03	1,78
23	4,28	3,42	3,03	2,80	2,64	2,53	2,38	2,20	2,00	1,76
24	4,26	3,40	3,01	2,78	2,62	2,51	2,36	2,18	1,98	1,73
25	4,24	3,38	2,99	2,76	2,60	2,49	2,34	2,16	1,96	1,71
26	4,22	3,37	2,98	2,74	2,59	2,47	2,32	2,15	1,95	1,69
27	4,21	3,35	2,96	2,73	2,57	2,46	2,30	2,13	1,93	1,67
28	4,20	3,34	2,95	2,71	2,56	2,44	2,29	2,12	1,91	1,65
29	4,18	3,33	2,93	2,70	2,54	2,43	2,28	2,10	1,90	1,64
30	4,17	3,32	2,92	2,69	2,53	2,42	2,27	2,09	1,89	1,62
35	4,12	3,26	2,87	2,64	2,48	2,37	2,22	2,04	1,83	1,57
40	4,08	3,23	2,84	2,61	2,45	2,34	2,18	2,00	1,79	1,51
45	4,06	3,21	2,81	2,58	2,42	2,31	2,15	1,97	1,76	1,48
50	4,03	3,18	2,79	2,56	2,40	2,29	2,13	1,95	1,74	1,44
60	4,00	3,15	2,76	2,52	2,37	2,25	2,10	1,92	1,70	1,39
70	3,98	3,13	2,74	2,50	2,35	2,23	2,07	1,89	1,67	1,35
80	3,96	3,11	2,72	2,49	2,33	2,21	2,06	1,88	1,65	1,31
90	3,95	3,10	2,71	2,47	2,32	2,20	2,04	1,86	1,64	1,28
100	3,94	3,09	2,70	2,46	2,30	2,19	2,03	1,85	1,63	1,26
125	3,92	3,07	2,68	2,44	2,29	2,17	2,01	1,83	1,60	1,21
150	3,90	3,06	2,66	2,43	2,27	2,16	2,00	1,82	1,59	1,18
200	3,89	3,04	2,65	2,42	2,26	2,14	1,98	1,80	1,57	1,14

Значения статистик Дарбина–Уотсона d_l, d_u при 5%-ном уровне значимости

n	$k = 1$		$k = 2$		$k = 3$		$k = 4$		$k = 5$	
	d_l	d_u	d_l	d_u	d_l	d_u	d_l	d_u	d_l	d_u
6	0,61	1,40								
7	0,70	1,36	0,47	1,90						
8	0,76	1,33	0,56	1,78	0,37	2,29				
9	0,82	1,32	0,63	1,70	0,46	2,13				
10	0,88	1,32	0,70	1,64	0,53	2,02				
11	0,93	1,32	0,66	1,60	0,60	1,93				
12	0,97	1,33	0,81	1,58	0,66	1,86				
13	1,01	1,34	0,86	1,56	0,72	1,82				
14	1,05	1,35	0,91	1,55	0,77	1,78				
15	1,08	1,36	0,95	1,54	0,82	1,75	0,69	1,97	0,56	2,21
16	1,10	1,37	0,98	1,54	0,86	1,73	0,74	1,93	0,62	2,15
17	1,13	1,38	1,02	1,54	0,90	1,71	0,78	1,90	0,67	2,10
18	1,16	1,39	1,05	1,53	0,93	1,69	0,82	1,87	0,71	2,06
19	1,18	1,40	1,08	1,53	0,97	1,68	0,85	1,85	0,75	2,02
20	1,20	1,41	1,10	1,54	1,00	1,68	0,90	1,83	0,79	1,99
21	1,22	1,42	1,13	1,54	1,03	1,67	0,93	1,81	0,83	1,96
22	1,24	1,43	1,15	1,54	1,05	1,66	0,96	1,80	0,86	1,94
23	1,26	1,44	1,17	1,54	1,08	1,66	0,99	1,79	0,90	1,92
24	1,27	1,45	1,19	1,55	1,10	1,66	1,01	1,78	0,93	1,99
25	1,29	1,45	1,21	1,55	1,12	1,66	1,04	1,77	0,95	1,89
26	1,30	1,46	1,22	1,55	1,14	1,65	1,06	1,76	0,98	1,88
27	1,32	1,47	1,24	1,56	1,16	1,65	1,08	1,76	1,01	1,86
28	1,33	1,48	1,26	1,56	1,18	1,65	1,10	1,75	1,03	1,85
29	1,34	1,48	1,27	1,56	1,20	1,65	1,12	1,74	1,05	1,84
30	1,35	1,49	1,28	1,57	1,21	1,65	1,14	1,74	1,07	1,83

ОГЛАВЛЕНИЕ

ОБЩИЕ СВЕДЕНИЯ	3
Тема 1. ОЦЕНКА ПОКАЗАТЕЛЕЙ КАЧЕСТВА ОБЪЕКТА ПО РЕЗУЛЬТАТАМ ЭКСПЕРИМЕНТОВ.....	4
Тема 2. ЛИНЕЙНАЯ ПАРНАЯ РЕГРЕССИЯ.....	10
Тема 3. НЕЛИНЕЙНАЯ ПАРНАЯ РЕГРЕССИЯ.....	16
Тема 4. ЛИНЕЙНАЯ МНОЖЕСТВЕННАЯ РЕГРЕССИЯ.....	22
Библиографический список	29
Приложение 1. Критические значения t -критерия Стьюдента при уровне значимости 0,10; 0,05; 0,01 (двухсторонний).....	30
Приложение 2. Таблица значений F -критерия Фишера при уровне значимости $\alpha = 0,05$	31
Приложение 3. Значения статистик Дарбина–Уотсона d_l, d_u при 5%-ном уровне значимости.....	32