МИНОБРНАУКИ РОССИИ
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Тверской государственный технический университет»
(ТвГТУ)

А.В. Ганичев, А.В. Ганичева

ПРАКТИКУМ ПО ЭКОНОМЕТРИКЕ

Учебное пособие

Тверь 2024

УДК 330.43(075.8)
ББК 65вбя

Рецензенты: профессор кафедры физико-математических дисциплин и информационных технологий Тверской государственной сельско-хозяйственной академии, доктор технических наук, профессор, академик РАЕН Попов П.Г.; доцент кафедры математической статистики и системного анализа Тверского государственного университета кандидат физико-математических наук Лесик А.И.

Ганичев А.В., Ганичева А.В. Практикум по эконометрике: учебное пособие. Тверь: Тверской государственный технический университет, 2024. 148 с.

Изложены основные разделы эконометрики: линейная парная регрессия, нелинейная регрессия, множественная регрессия и корреляция, системы эконометрических уравнений, модели временных рядов. Приведены типовые задачи с решениями, а также задачи для самостоятельной работы учащихся.

Предназначено для бакалавров, обучающихся на экономических специальностях.

ISBN 978-5-7995-1328-3

ВВЕДЕНИЕ

В современной системе высшего образования существенная роль отводится математическим и статистическим методам описания, анализа и прогнозирования развития различных социально-экономических явлений и процессов. Формирование профессиональных компетенций студентов экономических и управленческих специальностей невозможно без освоения основных методов и моделей эконометрики. Эконометрика является одной из самых сложных дисциплин в подготовке студентов-экономистов. От уровня подготовки по этой дисциплине зависят квалификация выпускника и его конкурентоспособность на рынке труда.

За последние три десятилетия дисциплина «Эконометрика» прочно вошла в учебный процесс и стала одной из базовых. По данному предмету выпущено большое количество учебно-методических материалов различной степени сложности. В первую очередь это фундаментальные учебники российских ученых, например [2, 3, 7, 28, 30, 40, 41, 49, 65], и зарубежных авторов, в частности [20].

Издано множество учебных пособий. Некоторые из них приведены в библиографическом списке. Большое количество и разнообразие пособий можно объяснить оригинальностью и своеобразием изложения учебных вопросов научными школами, сформировавшимися в различных вузах.

Учебные пособия для проведения практических и лабораторных работ (практикумы) можно условно разделить на две группы:

1) практикумы – сборники задач и решебники;

2) практикумы, ориентированные на эконометрические пакеты программ.

Имеются также учебные пособия, относящиеся одновременно к обеим группам.

Примерами практикумов первой группы (им уделено основное внимание при анализе источников учебной литературы) в библиографическом списке выступают [4, 9–11, 13, 15, 17, 19, 27, 32, 35, 38, 46, 54, 57, 60, 62].

Компьютерные практикумы второй группы, например [1, 6, 23, 42, 43, 61], ориентированы прежде всего на пакеты MS Excel и STATISTICA.

Образование двух групп практикумов вызвано большим объемом практического учебного материала по предмету «Эконометрика».

Предлагаемое учебное пособие относится к практикумам первой группы. Оно отличается тем, что в нем рассмотрено множество задач, описывающих различные вопросы социально-производственной сферы и служащих своеобразными шаблонами решения прикладных задач эконометрическими методами. Эти шаблоны представляют собой информационные структуры и могут использоваться как учащимися, так и сотрудниками различных организаций в любой сфере деятельности. При выполнении заданий по эконометрике с помощью прикладных программ на компьютере теряется наглядность алгоритмизации процесса решения задач. Предложенный метод шаблонов отличается иллюстративностью и дает полное представление о шагах алгоритма. Данный метод является обучающим, т.е. своеобразным репетитором по курсу «Эконометрика». Обилие полностью решенных прикладных задач способствует развитию интереса к предмету с учетом дальнейшего использования рассмотренных шаблонов в будущей производственной деятельности.

Пособие состоит из 5 глав и охватывает основные разделы курса эконометрики для бакалавров, а именно линейную парную регрессию и корреляцию, эластичность функций; нелинейную регрессию, эластичность нелинейных функций; множественную регрессию и корреляцию; системы эконометрических уравнений; модели временных рядов.

В каждой главе рассматриваются примеры решения типовых задач различной степени сложности. В конце главы предлагаются аналогичные задачи для самостоятельного решения.

Тестирование охватывает весь пройденный материал без выделения заданий по отдельным разделам курса (модулям).

При изучении пособия следует пользоваться учебниками или учебными пособиями, в которых изложены теоретический материал и формульные соотношения.

В главе 1 рассматриваются такие задачи, как оценивание параметров уравнения линейной регрессии, оценка качества уравнения регрессии, прогнозирование по регрессионной модели и точность прогнозов. Показаны задания по корреляционному анализу и вопросам исследования эластичности функций.

Глава 2 посвящена нелинейной регрессии. Приведены задачи на определение видов связей между случайными величинами. Рассмотрены задания, связанные с основными типами нелинейных зависимостей, применяемых в эконометрике. Показано, как рассчитывать эластичность нелинейных функций. Даны задания на линеаризацию нелинейных функций.

В главе 3 описываются задачи, посвященные модели множественной регрессии, системе нормальных уравнений, ковариационной матрице оценок, вычислению коэффициентов регрессии. Продемонстрирована методика проверки адекватности уравнения регрессии и значимости его коэффициентов для множественной линейной регрессионной модели. Рассмотрен метод отбора факторных переменных для включения их в модель множественной регрессии при наличии дублирующих и связанных признаков. Показано, как оценивать степень влияния факторных признаков на результативный признак. Имеются задания по расчету парных и частных коэффициентов корреляции, а также коэффициента множественной детерминации. Дан пример построения модели с фиктивными переменными.

Глава 4 посвящена системам эконометрических уравнений. Рассмотрены задания на исследование структурной и приведенной форм модели, экзогенные и эндогенные переменные, необходимое и достаточное условия идентификации систем уравнений. Показано оценивание системы одновременных уравнений.

В главе 5 имеются задания, связанные с моделями временных рядов, расчетом автокорреляционной функции, моделированием тенденции временного ряда. Рассмотрены мультипликаторы, средний и медианный лаги, а также особенности моделей с распределенными лагами. Показана методика проверки наличия автокорреляции в остатках по критерию Дарбина – Уотсона.

В приложении приведены вопросы для подготовки к экзамену, тестовые задания по курсу «Эконометрика», основные расчетные таблицы.

Библиографический список состоит из 65 наименований и содержит прежде всего практикумы первой группы.

Краткость, простота изложения, достаточное количество разобранных типовых иллюстрирующих примеров и множество задач для самостоятельного решения делают данное пособие незаменимым помощником при изучении курса эконометрики.

Авторы выражают большую признательность Попову Павлу Георгиевичу и Лесик Александре Ильиничне за рецензирование учебного пособия.

ГЛАВА 1
ЛИНЕЙНАЯ ПАРНАЯ РЕГРЕССИЯ

Примеры решения задач и выполнения заданий по теме «Линейная однофакторная регрессионная модель»

Задание 1.1.1. Построить линейное уравнение регрессии дохода x (ден. ед.) и прибыли y (ден. руб.) по данным для 7 предприятий:

x_i	10	20	40	30	50	30	40
y_i	3	7	15	10	17	10	15

По уравнению регрессии определить значение прибыли при доходе в 35 ден. ед.

Решение

Последовательно вычисляем:

$$\overline{X} = \frac{1}{7}\sum_{j=1}^{7} x_j = 31{,}43; \quad \overline{Y} = \frac{1}{7}\sum_{j=1}^{7} y_j = 11;$$

$$\overline{XY} = \frac{1}{7}\sum_{j=1}^{7} x_j y_j = 402{,}86;$$

$$\overline{X^2} = \frac{1}{7}\sum_{j=1}^{7} x_j^2 = 1143; \quad \overline{Y^2} = \frac{1}{7}\sum_{j=1}^{7} y_j^2 = 142{,}429;$$

$$S_x^2 = \overline{X^2} - (\overline{X})^2 = 155{,}155; \quad S_x = 12{,}456;$$

$$S_y^2 = \overline{Y^2} - (\overline{Y})^2 = 21{,}42; \quad S_y = 4{,}628;$$

$$\widehat{K}_{xy} = \overline{XY} - \overline{X}\cdot\overline{Y} = 57{,}13;$$

$$\widehat{b}_1 = \frac{\widehat{K}_{xy}}{s_x^2} = 0{,}368; \quad \widehat{b}_0 = \overline{Y} - \widehat{b}_1\overline{X} = -0{,}566;$$

$$\widehat{r}_{xy} = \frac{\widehat{K}_{xy}}{s_x \cdot s_y} = 0{,}991.$$

Уравнение регрессии:

$$\widehat{y} = -0{,}566 + 0{,}368x.$$

Если x = 35 ден. ед., то из полученного уравнения регрессии следует, что соответствующее значение прибыли

$$y = -0{,}566 + 0{,}368 \cdot 35 = 12{,}31 \text{ ден. ед.}$$

Задание 1.1.2. Построить линейное уравнение регрессии и оценить коэффициент корреляции между добычей угля на одного рабочего y (т) и мощностью пласта x (м) по данным для 10 шахт:

x_j	8	11	12	9	8	8	9	9	8	12
y_j	5	10	10	7	5	6	6	5	6	8

Оценить величину добычи угля на одного рабочего при мощности пласта в 7 м.

Решение

В результате расчетов получаем:

$$\overline{X} = 9,4;\ \overline{Y} = 6,8;\ \overline{X^2} = 90,8;\ \overline{Y^2} = 49,6;$$

$$\overline{XY} = 66,4;\ S_x^2 = 2,44;\ S_x = 1,56;$$

$$S_y^2 = 3,36;\ S_y = 1,83;\ \hat{K}_{xy} = 2,48;$$

$$\hat{r}_{xy} = 0,868;\ \hat{b}_1 = 1,016;\ \hat{b}_0 = -2,75;$$

$$\hat{y} = -2,75 + 1,016x.$$

При $x = 7$ $y = -2,75 + 1,016 \cdot 7 = 4,36$ м.

Задание 1.1.3. Известна выборка 15 значений случайных величин X и Y:

X	2	4	7	10	12	15	16	20	21	21	29	30	32	33	33
Y	5	8	8	12	16	16	16	25	30	31	30	40	45	47	50

Построить линейное уравнение регрессии зависимости Y от X. На сколько единиц изменится Y, если X увеличится на 1?

Решение

Общий вид линейного уравнения парной регрессии:

$$\hat{y}_i = a + b \cdot x_i,$$

где $\hat{y}_i$ – расчетные теоретические значения результативного признака для i-го наблюдения; a, b – оценки параметров линейного уравнения парной регрессии; x_i – значение факторного признака для i-го наблюдения.

Оценки параметров a, b найдем с помощью метода наименьших квадратов (МНК). Для определения оценок параметров необходимо решить систему линейных уравнений (систему нормальных уравнений):

$$
\begin{cases}
n \cdot a + \sum_{i=1}^{n} x_i \cdot b = \sum_{i=1}^{n} y_i; \\
\sum_{i=1}^{n} x_i \cdot a + \sum_{i=1}^{n} x_i^2 \cdot b = \sum_{i=1}^{n} \left(x_i \cdot y_i\right).
\end{cases}
$$

Вспомогательные результаты промежуточных расчетов приведены ниже:

№ п/п	x_i	y_i	x_i^2	$x_i y_i$	y_i^2	$\hat{y}_i$
1	2	5	4	10	25	1,982
2	4	8	16	32	64	4,721
3	7	8	49	56	64	8,830
4	10	12	100	120	144	12,939
5	12	16	144	192	256	15,679
6	15	16	225	240	256	19,788
7	16	16	256	256	256	21,158
8	20	25	400	500	625	26,636
9	21	30	441	630	900	28,006
10	21	31	441	651	961	28,006
11	29	30	841	870	900	38,964
12	30	40	900	1200	1600	40,333
13	32	45	1024	1440	2025	43,073
14	33	47	1089	1551	2209	44,442
15	33	50	1089	1650	2500	44,442
Итого	285	379	7019	9398	12785	379,000
Среднее	19,000	25,267	467,933	626,533	852,333	–

Для расчета оценок параметров можно использовать готовые формулы, которые вытекают из системы нормальных уравнений:

$$b = \frac{\operatorname{cov}(xy)}{\sigma_x^2} = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\overline{x^2} - \overline{x}^2} = \frac{626{,}533 - 19 \cdot 25{,}267}{467{,}933 - 19^2} = 1{,}370;$$

$$a = \overline{y} - b\overline{x} = 25{,}267 - 1{,}37 \cdot 19 = -0{,}758.$$

Получили линейное уравнение парной регрессии:

$$\hat{y}_i = -0{,}758 + 1{,}370 \cdot x_i.$$

Коэффициент регрессии b показывает, что при изменении фактора x на 1 ед. значение результативного признака y в среднем возрастает на 1,37 ед. Поскольку значение положительное, то связь между признаками прямая.

Задание 1.1.4. Указана зависимость между размером и доходностью дворянских имений в конце XIX века:

Размер имения (десятин)	240	255	265	270	285	295	310	320	352	330
Валовой доход (тыс. руб.)	1,5	1,25	1,55	1,4	1,46	1,6	1,8	1,8	1,85	1,9

Построить уравнение регрессии. Оценить, какой валовой доход соответствовал имению размером (десятины):

1) 275;

2) 248.

Как менялась в среднем доходность при увеличении размера имения на 1 десятину?

Решение

Вводим переменные: x – размер имения; y – валовой доход.

Строим таблицу промежуточных вычислений:

№	x	y	x^2	y^2	$x\ \ y$
1	240	1,5	57600	2,25	360
2	255	1,25	65025	1,5625	318,75
3	265	1,55	70225	2,4025	410,75
4	270	1,4	72900	1,96	378
5	285	1,46	81225	2,1316	416,1
6	295	1,6	87025	2,56	472
7	310	1,8	96100	3,24	558
8	320	1,8	102400	3,24	576
9	352	1,85	123904	3,4225	651,2
10	330	1,9	108900	3,61	627
Сумма	2922	16,11	865304	26,3791	4767,8
Среднее	292,2	1,611	86530,4	2,6379	476,78

Находим параметры уравнения регрессии по формулам:

$$b = \frac{\overline{x\ \ y} - \bar{x}\ \ \bar{y}}{\sigma_x^2} = \frac{476,78 - 292\ \ 1,611}{33,9052^2} = 0,0053;$$

$$a = \bar{y} - b\ \ \bar{x} = 1,611 - 0,0053\ \ 292,2 = 0,0743,$$

где

$$\sigma_x = \sqrt{\left(\overline{x^2}\right) - \left(\overline{x}\right)^2} = \sqrt{86530,4 - 292,2^2} = 33,9052;$$

$$\sigma_y = \sqrt{\left(\overline{y^2}\right) - \left(\overline{y}\right)^2} = \sqrt{2,6379 - 1,611^2} = 0,2064.$$

Уравнение регрессии имеет вид $\hat{y} = 0,0743 + 0,0053\ \ x$.

Валовой доход, который соответствовал имению размером 275 десятин:

$$\hat{y}_{275} = 0,0743 + 0,0053\ \ 275 = 1,521 \text{ тыс. руб.}$$

Валовой доход имения размером 248 десятин:

$$\hat{y}_{248} = 0,0743 + 0,0053\ \ 248 = 1,379 \text{ тыс. руб.}$$

При увеличении размера имения на 1 десятину доходность в среднем увеличивалась на 0,0053 тыс. руб.

Задание 1.1.5. Построить линейное уравнение регрессии зависимости прибыли предприятия *y* (ден. ед.) от объема выпускаемой продукции *x* (тыс. шт.) по данным, полученным для 7 предприятий:

x_i	10	20	40	30	50	30	40
y_i	3	7	15	10	17	10	15

По уравнению регрессии определить значение прибыли при выпуске продукции в 35 тыс. шт.

Решение

Вычисляем:

$$\overline{X} = \frac{1}{7}\sum_{j=1}^{7} x_j = 31,43;\ \overline{Y} = \frac{1}{7}\sum_{j=1}^{7} y_j = 11;$$

$$\overline{XY} = \frac{1}{7}\sum_{j=1}^{7} x_j y_j = 402,86;\ \overline{X^2} = \frac{1}{7}\sum_{j=1}^{7} x_j^2 = 1143;\ \overline{Y^2} = \frac{1}{7}\sum_{j=1}^{7} y_j^2 = 142,429;$$

$$S_x^2 = \overline{X^2} - (\overline{X})^2 = 155,155;\ S_x = 12,456;\ S_y^2 = \overline{Y^2} - (\overline{Y})^2 = 21,42;\ S_y = 4,628.$$

Ковариация составляет $\widehat{K}_{xy} = \overline{XY} - \overline{X}\ \ \overline{Y} = 57,13$; коэффициенты уравнения регрессии:

$$\widehat{b}_1 = \frac{\widehat{K}_{xy}}{s_x^2} = 0,368;\ \widehat{b}_0 = \overline{Y} - \widehat{b}_1\overline{X} = -0,566.$$

Уравнение регрессии имеет вид

$$\widehat{y} = -0,566 + 0,368x.$$

Оценка $\hat{b}_1$ интерпретируется как прибыль, полученная при увеличении выпуска продукции на 1 тыс. шт. Так как $\hat{b}_0 < 0$, то относительное изменение прибыли *y* происходит быстрее, чем изменение фактора *x* – объема выпускаемой продукции.

Задача прогноза заключается в определении значения $y_{прогн}$ при данном значении $x_{прогн}$. Для этого данное значение $x_{прогн}$ ставится в уравнение регрессии и находится соответствующее значение $y_{прогн}$.

Если $x_{прогн}$ = 35 тыс. шт., то из полученного уравнения регрессии следует, что соответствующее значение прибыли

$$y_{прогн} = -0{,}566 + 0{,}368\ \ 35 = 12{,}31 \text{ ден. ед.}$$

Задание 1.1.6. Представлены выборочные данные изменения по годам прибыли (ден. ед.) предприятия:

Годы x_i	2005	2008	2009	2011	2012	2013	2015	2016
Прибыль y_i (усл. ед.)	7	8	8	9	11	12	15	14

Определить, какую (предположительно) прибыль можно было ожидать в 2017 году.

Решение

Для упрощения вычислений поместим начало отсчета в 2005 год. Тогда 2008-й будет иметь номер 3, 2009-й – 4, 2011-й – 6, 2012-й – 7, 2013-й – 8, 2015-й – 10, 2016-й – 11. Вычислим:

$$\bar{x} = \frac{0+3+4+6+7+8+10+11}{8} = \frac{49}{8} = 6{,}125;$$

$$\bar{y} = \frac{7+8+8+9+11+12+15+14}{8} = \frac{84}{8} = 10{,}5;$$

$$\overline{xy} = \frac{0\ \ 7+3\ \ 8+4\ \ 8+6\ \ 9+7\ \ 11+8\ \ 12+10\ \ 15+11\ \ 14}{8} = \frac{587}{8} = 73{,}375;$$

$$S_x^2 = \frac{0^2+3^2+4^2+6^9+7^2+8^2+10^2+11^2}{8} - (6{,}125)^2 = \frac{395}{8} - 37{,}516 = 11{,}859;$$

$$b_1 = \frac{73{,}375 - 6{,}125\ \ 10{,}5}{11{,}859} = 0{,}76;\ b_0 = 10{,}5 - 0{,}76\ \ 6{,}125 = 5{,}845.$$

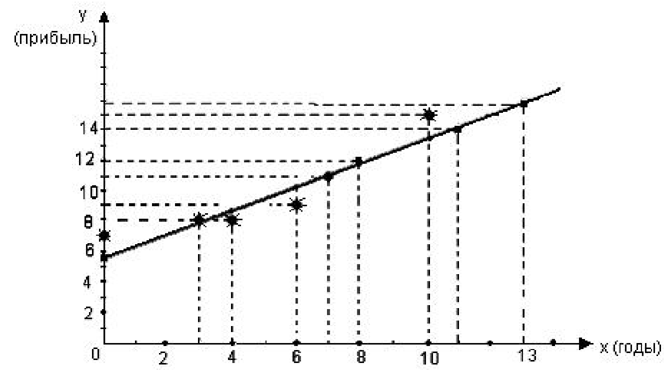
Уравнение регрессии: $y = 0{,}76x + 5{,}845$.

Номер 2017 года будет 13. Тогда ожидаемая прибыль составит

$$y(13) = 0{,}76\ \ 13 + 5{,}845 = 15{,}725 \text{ ден. ед.}$$

Уравнение регрессии: $y = 0{,}76x + 5{,}845$.

Проиллюстрируем решение графически:



Звездочками обозначены точки с координатами (x_i, y_i), соответствующие исходным данным, кружками – точки с координатами (x_i, $\hat{y}(x_i)$).

По построенной линии регрессии можно найти значения прибыли для любого года, не только будущего, но и предшествующего и промежуточного. Следует определить прибыль, полученную в 2000 году. Нужно обратить внимание на то, что 2000 год при введенной точке отсчета будет иметь номер –5, т.е. отрицательный. Коэффициент регрессии b_1 показывает, на сколько единиц в среднем изменяется переменная Y при увеличении переменной X на 1 ед. Для разобранного примера b_1 = 0,76, т.е. каждый год в течение рассмотренного периода прибыль возрастает в среднем на 0,76 ден. ед.

Задание 1.1.7. Зависимость результатов оценки Y (усл. ед.) от величины оцениваемого показателя X (усл. ед.) показана ниже:

x_i	78	82	87	79	89	106	67	88	73	87	76	115
y_i	133	148	134	154	162	195	139	158	152	162	159	173

Построить уравнение регрессии. Оценить среднее изменение оценки при изменении значения оцениваемого показателя на 1 %.

Решение

Последовательно находим:

$$\bar{x} = \frac{1}{12}\ (78+82+87+79+89+106+67+88+73+87+76+115) = 85{,}58 \text{ усл. ед.;}$$

$$\bar{y} = \frac{1}{12}\ (133+148+134+154+162+195+139+158+152+162+159+173) =$$

$$= 155{,}75 \text{ усл. ед.;}$$

$$\overline{xy} = \frac{1}{12}\ (78\ \ 133 + 82\ \ 148 + 87\ \ 134 + \ \ldots\ \ + 155\ \ 173) = 134\ 84;$$

$$S_x^2 = \frac{1}{12}\ \left(78^2 + 82^2 + 87^2 + \ \ldots\ + 76^2 + 115^2\right) - (85{,}58)^2 = 168{,}31;\ S_x = 12{,}97.$$

$$b_1 = \frac{13484 - 85{,}58\ \ 155{,}75}{168{,}31} = 0{,}92\,;\ b_0 = \bar{y} - b_1\ \ \bar{x} = 155{,}75 - 0{,}92\ \ 85{,}58 = 77{,}02\,.$$

Уравнение регрессии: $y = 77{,}02 + 0{,}92\ \ x$. Коэффициент эластичности

$$Э = 0{,}92\frac{85{,}58}{155{,}75} = 0{,}51.$$

Вывод: среднее увеличение оценки, приходящееся на 1 % изменения оцениваемого показателя, составляет 0,51.

Задание 1.1.8. Экспертами оценивались вкусовые качества разных вин для контроля цены в магазинах. Данные приведены ниже:

Марка вина	1	2	3	4	5	6	7	8	9	10
Оценка *X* (баллы)	11	14	17	15	13	13	18	10	19	25
Цена *Y* (усл. ед.)	1,57	1,6	2	2,1	1,7	1,85	1,8	1,15	2,3	2,4

Оценить изменение цен в зависимости от балла вина.

Вспомогательные расчеты приведены ниже:

Марка вина	Оценка *X* (баллы)	Цена *Y* (усл. ед.)	$x\ \ y$	x^2
1	11	1,57	17,27	121
2	14	1,60	22,4	196
3	17	2,00	34,0	289
4	15	2,10	31,5	225
5	13	1,70	22,1	169
6	13	1,85	24,05	169
7	18	1,80	32,4	324
8	10	1,15	11,5	100
9	19	2,30	43,7	361
10	25	2,40	60,0	625
	$\bar{x} = \frac{x}{10} = 15{,}5$	$\bar{y} = \frac{y}{10} = 1{,}85$	$\overline{xy} = \frac{xy}{10} =$ $= 29{,}89$	$\overline{x^2} = \frac{x^2}{10} =$ $= 257{,}9$

Решение

Поскольку в определении коэффициента эластичности участвуют значения $\overline{x\ \ y}$ и S_x^2, в последних двух столбцах выше были записаны значения произведений $x\ \ y$ и x^2. Итак, находим:

$$b_1 = \frac{29{,}89 - 15{,}5\ \ 1{,}85}{257{,}9 - 15{,}5^2} = 0{,}069;$$

$$Э = 0{,}069\ \ \frac{15{,}5}{1{,}85} = 0{,}58.$$

Вывод: при увеличении оценки на 1 % цена вин в среднем возрастает на 0,58 %.

Задание 1.1.9. Данные по урожайности пшеницы (ц/га) и количеству соответствующих осадков за год представлены ниже:

Количество осадков (мм)	240–250	250–260	260–270	270–280	280–285	285–290	290–295
Урожайность (ц/га)	30	31	29	34	37	37	36

Исследовать зависимость урожайности пшеницы от количества осадков за год.

Решение

Изучим зависимость урожайности y от количества осадков x. Представим функцию $y = f(x)$ как

$$y = ax + b.$$

В развернутой форме система МНК для определения параметров a и b будет иметь вид

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

$$a \sum_{i=1}^{n} x_i + n\ \ b = \sum_{i=1}^{n} y_i.$$

Заполним таблицу вычислений (в качестве x_i берем середину интервала):

Номер i	x_i	y_i	$x_i\,y_i$	x_i^2
1	245	30	7350	60025
2	255	31	7905	65025
3	265	29	7685	70225
4	275	34	9350	75625
5	282,5	37	10452,5	79806,25
6	287,5	37	10637,5	82656,35
7	292,5	36	10530	85556,25
Сумма	1902,5	234	63910	518918,75

Тогда система наименьших квадратов примет вид

$$\begin{cases} 518918{,}75\,a + 1902{,}5\,b = 63910; \\ 1902{,}5\,a + 7\,b. \end{cases}$$

Составим и вычислим определитель системы:

$$\Delta = \begin{vmatrix} 518918{,}75 & 1902{,}5 \\ 1902{,}5 & 7 \end{vmatrix} = 518918{,}75\cdot 7 - 1902{,}5\cdot 1902{,}5 = 12925.$$

Он отличен от нуля, система уравнений невырожденная. Найдем определители по элементам столбцов:

$$\Delta_a = \begin{vmatrix} 63910 & 1902{,}5 \\ 234 & 7 \end{vmatrix} = 63910\cdot 7 - 1902{,}5\cdot 234 = 2185;$$

$$\Delta_b = \begin{vmatrix} 518918{,}75 & 63910 \\ 1902{,}5 & 234 \end{vmatrix} = 518918{,}75\cdot 234 - 1902{,}5\cdot 63910 = -161787{,}5.$$

Решение системы имеет вид:

$$a = \frac{2185}{12925} = 0{,}169;\; b = \frac{-161787{,}5}{12925} = -12{,}517.$$

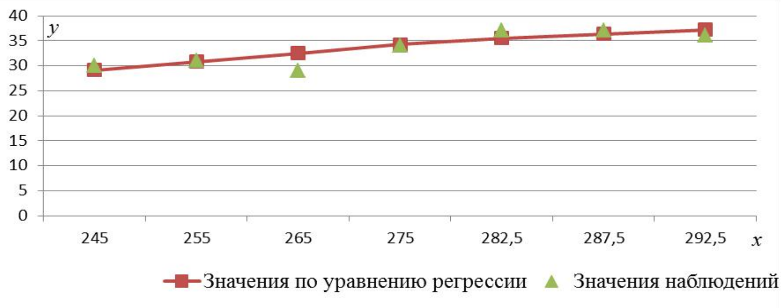
Тогда искомое уравнение регрессии будет

$$\hat{y} = 0{,}17x - 12{,}52.$$

Найдем значения, рассчитанные из уравнения регрессии. Получим

x_i	245,0	255,0	265,0	275,0	282,5	287,5	292,5
$\hat{y}_i$	29,130	30,830	32,530	34,230	35,505	36,355	37,205

Изобразим исходные наблюдения и значения, выведенные по уравнению регрессии, на графиках:



Задание 1.1.10. При выборочном осмотре 20 юношей из 10-х классов была выявлена следующая зависимость между ростом и размером обуви у молодых людей:

Рост (см)	175	178	180	178	182	172	172	180	175	176
Размер обуви (см)	42	43	44	44	45	42	43	45	45	44
Рост (см)	180	181	168	172	174	178	186	182	185	184
Размер обуви (см)	44	45	40	43	43	45	45	45	45	44

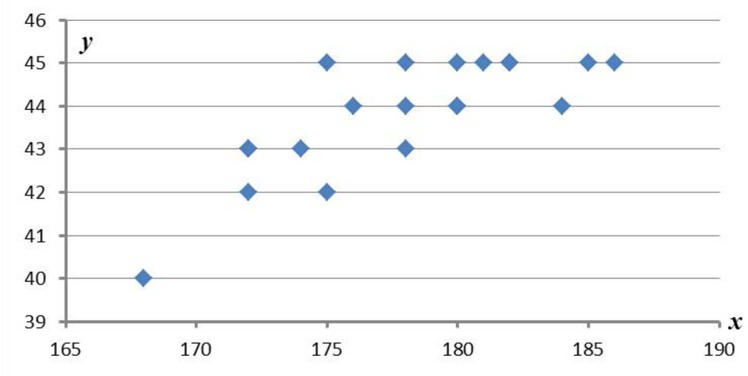
Построить уравнение регрессии и найти, какой размер обуви соответствует росту:

а) 179 см;

б) 177 см.

Решение

Пусть переменная y – размер обуви, а x – рост. Строим корреляционное поле:



Указанное поле не выявляет какого-то явно выраженного типа зависимости. Во всех таких случаях целесообразно использовать линейную регрессию, так как нет оснований применять более сложные виды регрессии. Представим функцию $y = f(x)$ в виде

$$y = ax + b.$$

В развернутой форме система МНК для определения параметров a и b будет иметь вид

$$a \sum_{i=1}^{n} x_i^2 + b \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} (x_i \; y_i);$$

$$a \sum_{i=1}^{n} x_i + n \; b = \sum_{i=1}^{n} y_i.$$

Заполним табличный массив, предварительно отсортировав значения по росту:

Номер точки i	x_i	y_i	x_iy_i	x_i^2
1	168	40	6720	28224
2	172	42	7224	29584
3	172	43	7396	29584
4	172	43	7396	29584
5	174	43	7482	30276
6	175	42	7350	30625
7	175	45	7875	30625
8	176	44	7744	30976
9	178	43	7654	31684
10	178	44	7832	31684
11	178	45	8010	31684
12	180	44	7920	32400
13	180	45	8100	32400
14	180	44	7920	32400
15	181	45	8145	32761
16	182	45	8190	33124
17	182	45	8190	33124
18	184	44	8096	33856
19	185	45	8325	34225
20	186	45	8370	34596
Сумма	3558	876	155939	633416

Тогда система наименьших квадратов примет вид

$$\begin{cases} a \cdot 633416 + b \cdot 3558 = 155939\ ; \\ a \cdot 3558 + 20 \cdot b = 876. \end{cases}$$

Рассмотрим систему уравнений

$$\begin{cases} 633416a + 3558b = 155939, \\ 3558a + 20b = 876. \end{cases}$$

Находим определитель системы

$$\Delta = \begin{vmatrix} 633416 & 3558 \\ 3558 & 20 \end{vmatrix} = 633416 \cdot 20 - 3558 \cdot 3558 = 8956.$$

Он не равен нулю, система уравнений невырожденная. Находим дополнительные определители:

$$\Delta_a = \begin{vmatrix} 155939 & 3558 \\ 876 & 20 \end{vmatrix} = 155939 \cdot 20 - 3558 \cdot 876 = 1972.$$

$$\Delta_b = \begin{vmatrix} 633416 & 155939 \\ 3558 & 876 \end{vmatrix} = 633416 \cdot 876 - 155939 \cdot 3558 = 41454.$$

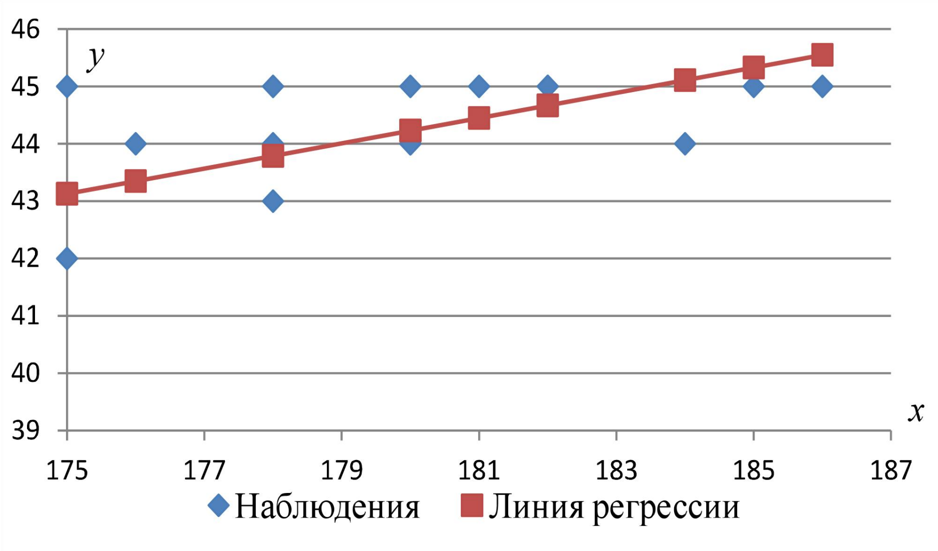
Теперь легко определить значения неизвестных:

$$a = \frac{1972}{8956} = 0,22; \; b = \frac{41454}{8956} = 4,63.$$

Тогда искомое уравнение будет

$$y = 0,22 \;\; x + 4,63.$$

Построим графики наблюдений и регрессионной зависимости y от x:



Прогноз для определения размера обуви:
а) $y(179) = 0,22 \cdot 179 + 4,63 \approx 44,01 \approx 44$;
б) $y(177) = 0,22 \cdot 176 + 4,63 \approx 43,35 \approx 43$.

Задание 1.1.11. Есть данные о динамике за год количества заболевших на тысячу жителей города:

Годы (t)	1	2	3	4	5	6	7	8	9	10	11	12
y (% хронических больных)	12	11	9	7	8	7	6	4	5	2	1	3

Построить линейное уравнение регрессии. Что можно сказать о прогнозе количества таких больных, когда t = 13 (t – номер месяца)?

Решение

$$\bar{y} = \frac{1}{12} \cdot \sum_{i=1}^{12} x_i = \frac{1}{12} \cdot (12 + 11 + 9 + 7 + 8 + 7 + 6 + 4 + 5 + 2 + 1 + 3) =$$
$$= 6,25;$$
$$\bar{t} = \frac{1}{12} \cdot \sum_{i=1}^{12} t_i = \frac{1}{12} \cdot (1 + 2 + 3 + 4 + 5 + 6 + 7 + 8 + 9 + 10 + 11 + 12) =$$
$$= 6,5;$$

$$\overline{yt} = \frac{1}{12} \cdot \sum_{i=1}^{12} y_i \cdot t_i = \frac{1}{12} \cdot (1 \cdot 12 + 2 \cdot 11 + 3 \cdot 9 + 4 \cdot 7 + 5 \cdot 8 + 6 \cdot 7 +$$
$$+7 \cdot 6 + 8 \cdot 4 + 9 \cdot 5 + 10 \cdot 2 + 11 \cdot 1 + 12 \cdot 3) = 29{,}75;$$
$$\overline{y^2} = \frac{1}{12} \cdot \sum_{i=1}^{6} y_i^2 = \frac{1}{12} \cdot (12^2 + 11^2 + 9^2 + 7^2 + 8^2 + 7^2 + 6^2 +$$
$$+4^2 + 5^2 + 2^2 + 1^2 + 3^2) \approx 49{,}916;$$
$$\overline{t^2} = \frac{1}{12} \cdot \sum_{i=1}^{6} t_i^2 = \frac{1}{12} \cdot (1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2 + 7^2 +$$
$$+8^2 + 9^2 + 10^2 + 11^2 + 12^2) \approx 54{,}167;$$
$$\sigma_y = \sqrt{\overline{y^2} - (\bar{y})^2} = \sqrt{49{,}917 - 6{,}25^2} \approx 3{,}295;$$
$$\sigma_t = \sqrt{\overline{t^2} - (\bar{t})^2} = \sqrt{54{,}167 - 6{,}5^2} \approx 3{,}452;$$
$$r = \frac{\overline{yt} - \bar{y} \cdot \bar{t}}{\sigma_y \sigma_t} = \frac{29{,}75 - 6{,}25 \cdot 6{,}5}{3{,}295 \cdot 3{,}452} \approx -0{,}956.$$

Составим линейное уравнение регрессии:

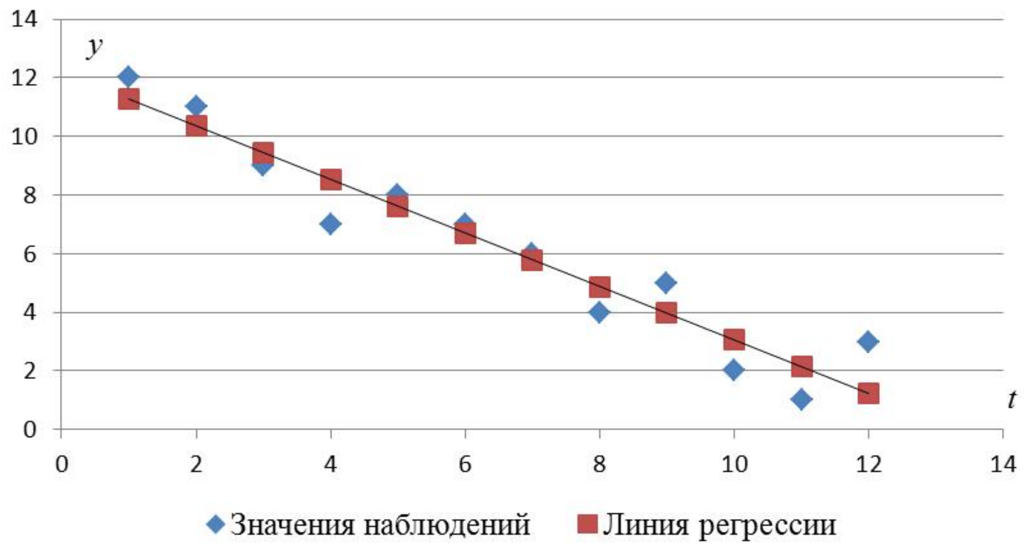
$$y - \bar{y} = r \cdot \frac{\sigma_y}{\sigma_t} \cdot (t - \bar{t}); \; y - 6{,}25 = -0{,}956 \cdot \frac{3{,}295}{3{,}452} \cdot (t - 6{,}5);$$

$$y = -0{,}913 \cdot t + 12{,}182.$$

Сделаем прогноз количества больных, когда $t = 13$ (на следующий год):

$$y(13) = -0{,}913 \cdot 13 + 12{,}182 = 0{,}318 \; (\%).$$

Построим поле корреляции и график линейной регрессии:



Задания для самостоятельного выполнения

Задание 1. Построить линию регрессии по данным, приведенным ниже:

x	1	2	3	4	5	6	7
y	2	1	3	5	4	6	4

Задание 2. Построить уравнение линейной регрессии по данным, представленным ниже:

x	3	4	5	6	8	10	12
y	–8	–10	–12	–11	–12	–14	–19

Задание 3. По данным, приведенным ниже, построить линию регрессии, сделать прогноз на 2024 год.

t_i	2016	2017	2018	2019	2020	2021	2022	2023
y_i	18	16	12	11	13	10	7	9

Задание 4. Построить линию регрессии по данным, показанным ниже:

x	2	4	6	10	8	9	11	13	12
y	5	3	7	9	11	12	14	13	15

Задание 5. Построить уравнение линейной регрессии по данным, представленным ниже, сделать прогноз на 2024 год.

t_i	2017	2018	2019	2020	2021	2022	2023
y_i	5	7	6	9	13	11	12

Задание 6. Построить уравнение регрессии по данным, приведенным ниже, сделать прогноз на 2010 год.

t_i	2000	2001	2002	2003	2004	2005	2006	2007	2008
y_i	8	6	10	12	14	15	13	17	16

Задание 7. Построить уравнение регрессии по данным, представленным ниже, сделать прогноз на 2022 год.

t_i	2015	2016	2017	2018	2019	2020	2021
y_i	8	6	10	12	11	13	12

Задание 8. Построить уравнение регрессии по данным, приведенным ниже, сделать прогноз на 2024 год.

t_i	2017	2018	2019	2020	2021	2022	2023
y_i	–12	–14	–13	–15	–17	–16	–19

Задание 9. Построить уравнение регрессии по данным, приведенным ниже, сделать прогноз на 2008 год.

t_i	2000	2001	2002	2003	2004	2005	2006
y_i	18	14	12	13	10	11	8

Задание 10. Построить линию регрессии на основе следующих данных:

x	12	14	16	20	18	21	23	22
y	–15	–13	–11	–12	–10	–8	–9	–7

Задание 11. Построить уравнение линейной регрессии, используя следующие данные:

x_i	2	4	6	8	10	12	14
y_i	11	13	12	14	20	15	17

Задание 12. Построить линию регрессии по данным, приведенным ниже:

x	12	14	16	20	18	19	21
y	15	13	11	9	11	5	7

Задание 13. Построить уравнение линейной регрессии по данным, приведенным ниже:

x_i	12	14	15	16	18	19	20
y_i	–11	–13	–12	–14	–20	–19	–18

Задание 14. Построить уравнение линейной регрессии на основе следующих данных:

t_i	2007	2008	2009	2010	2011	2012	2013	2014	2015
y_i	8	9	12	11	15	12	14	16	18

Для упрощения вычислений поместить начало отсчета в 2011 год.

Задание 15. Изменения средней заработной платы работников учреждения по годам показано ниже:

Год	2012	2013	2014	2015	2016	2017	2018	2019
Зарплата (усл. ед.)	18	27	27	28	29	31	30	33

Построить линейное уравнение регрессии. Чему была равна средняя заработная плата в 2021 году?

Задание 16. Тенденция домоуправления № 3 в исполнении обещаний по ремонту крыш, квартир, затопленных талыми водами, показана ниже:

Год	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024
Процентное исполнение обещаний	5	10	15	30	20	50	60	10	5	?

С целью выявления скрытых резервов и элемента саботирования в работе по обслуживанию населения и отработке полученных от населения денег предлагается построить линию регрессии и оценить вероятность того, что необходимый ремонт будет сделан в 2024 году.

Задание 17. Данные по урожайности пшеницы (ц/га) и количеству соответствующих осадков за год представлены ниже:

Количество осадков (мм)	240–250	250–260	260–270	270–280	280–285	285–290	290–295
Урожайность (ц/га)	30	31	29	34	37	37	36

Найти зависимость урожайности пшеницы от количества осадков за год. Оценить урожайность при минимальной норме осадков.

Задание 18. Статистические данные об уровне анализируемых преступлений приведены ниже:

Год	1991	1992	1993	1994	1995	1996	1997
Уровень преступности на 1000 осужденных	9,96	9,11	8,84	6,98	5,18	3,96	3,7

По уравнению регрессии оценить, какова была динамика изменений уровня преступности в 2000, 2002, 2004 годах.

Задание 19. Доходы населения Российской Федерации (руб.) по годам проиллюстрированы ниже:

Год	2013	2014	2015	2016	2017
Показатель	44230,6	47309,2	53153,2	54325,3	56205,1
Год	2018	2019	2020	2021	2022
Показатель	58781,5	62531,7	63692,0	70547,6	79113,9

Построить уравнение регрессии, дать прогноз на 2024 год.

Задание 20. Выборочное обследование 7 случайно выбранных работников показало следующую зависимость величины накопления денежных средств от дохода (тыс. руб.):

Наблюдение	Накопления Y	Доход X
1	4	45
2	7	59
3	6	46
4	3,7	35
5	1,6	32
6	4,7	55
7	3	36

Требуется:

1) построить однофакторную модель регрессии;

2) отобразить на графике исходные данные, результаты моделирования.

Примеры решения задач и выполнения заданий по теме «Линейный коэффициент корреляции и коэффициент эластичности»

Задание 1.2.1. Найти коэффициент корреляции между производительностью труда *Y* (тыс. руб.) и энерговооруженностью труда *X* (кВт) для 14 предприятий региона по данным, приведенным ниже:

x_i	2,8	2,2	3	3,5	3,2	3,7	4
y_i	6,7	6,9	7,2	7,3	8,4	8,8	9,1
x_i	4,8	6	5,4	5,2	5,4	6	9
y_i	9,8	10,6	10,7	11,1	11,8	12,1	12,4

Решение

Таблица промежуточных вычислений выглядит следующим образом:

№ п/п	x	y	x^2	y^2	$x \cdot y$
1	2,8	6,7	7,84	44,89	18,76
2	2,2	6,9	4,84	47,61	15,18
3	3	7,2	9	51,84	21,6
4	3,5	7,3	12,25	53,29	25,55
5	3,2	8,4	10,24	70,56	26,88
6	3,7	8,8	13,69	77,44	32,56
7	4	9,1	16	82,81	36,4
8	4,8	9,8	23,04	96,04	47,04
9	6	10,6	36	112,36	63,6
10	5,4	10,7	29,16	114,49	57,78
11	5,2	11,1	27,04	123,21	57,72
12	5,4	11,8	29,16	139,24	63,72
13	6	12,1	36	146,41	72,6
14	9	12,4	81	153,76	111,6
Сумма	64,2	132,9	335,26	1313,95	650,99
Среднее	4,5857	9,4929	23,9471	93,8536	46,4993

Находим коэффициент корреляции по формуле

$$r_{xy} = \frac{\overline{x \cdot y} - \overline{x} \cdot \overline{y}}{\sigma_x \cdot \sigma_y} = \frac{46{,}4993 - 4{,}5857 \cdot 9{,}4929}{1{,}7083 \cdot 1{,}9337} = 0{,}8984,$$

где

$$\sigma_x = \sqrt{\overline{(x^2)} - (\overline{x})^2} = \sqrt{23{,}9471 - 4{,}5857^2} = 1{,}7083;$$

$$\sigma_y = \sqrt{\overline{(y^2)} - (\overline{y})^2} = \sqrt{93{,}8536 - 9{,}4929^2} = 1{,}9337.$$

Связь между показателями высокая и прямая.

Задание 1.2.2. Имеются выборочные данные (усл. ед.) по показателям X и Y:

X	Y							
	20–40	40–60	60–80	80–100	100–120	120–140	140–160	160–180
30–60	2							
60–90	1	4	2		4		7	
90–120		6	1	5		2		
120–150			3	8	3	5	8	4
150–180				9		3	1	2

Числа, стоящие во внутренних клетках, означают частоту n_{ij} попадания X в i-й интервал, а Y – в j-й; n_i – частота попадания X в i-й интервал, n_j – частота попадания Y в j-й интервал. Найти коэффициент корреляции, используя формулу

$$\hat{r}_{xy} = \frac{n\sum_{i,j} n_{ij}\cdot x_i\cdot y_j - \left(\sum_i n_i x_i\right)\cdot\left(\sum_j n_j y_j\right)}{\sqrt{n\sum_i n_i (x_i)^2 - \left(\sum_i n_i x_i\right)^2}\sqrt{n\sum_j n_j \left(y_j\right)^2 - \left(\sum_j n_j y_j\right)^2}}.$$

Решение

Заменим в корреляционной таблице интервалы значениями их центров и добавим вспомогательные строки и столбцы, которые используем для расчета коэффициента корреляции.

X	Y								Сумма	$n_i \cdot x_i$	$n_i \cdot (x_i)^2$
	30	50	70	90	110	130	150	170			
45	2								2	90	4050
75	1	4	2		4		7		18	1350	101250
105		6	1	5		2			14	1470	154350
135			3	8	3	5	8	4	31	4185	564975
165				9		3	1	2	15	2475	408375
Сумма	3	10	6	22	7	10	16	6	$n = 80$	9570	1233000

Получим итоговый табличный массив

$n_j \cdot y_j$	90	500	420	1980	770	1300	2400	1020	8480
$n_j \cdot (y_j)^2$	2700	25000	29400	178200	84700	169000	360000	173400	1022400
$n_{ij} \cdot x_i \cdot y_j$	4950	46500	46200	278100	77550	179400	265500	147900	1046100

Рассчитаем коэффициент корреляции

$$\frac{80\ \ 1046100-9570\ \ 8480}{\sqrt{80\ \ 1233000-(9570)^2}\sqrt{80\ \ 1022400-(8480)^2}}=0{,}304.$$

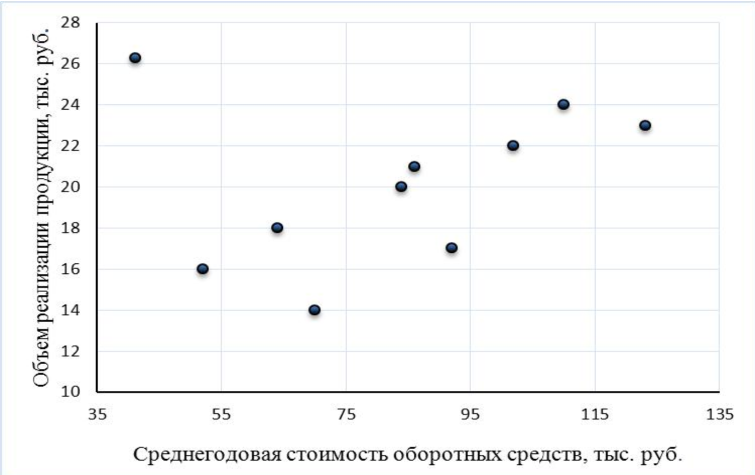
Задание 1.2.3. Методом корреляционного анализа исследовать зависимость между признаками. Построить график корреляционной зависимости (поле корреляции). Найти коэффициент корреляции для следующих данных:

Среднегодовая стоимость оборотных средств (тыс. руб.)	84	110	123,1	64	70	52	86	92	102	41,2
Объем реализации продукции (тыс. руб.)	20	24	23	18	14	16	21	17	22	26,3

Решение

Введем переменные: x – среднегодовая стоимость оборотных средств; y – объем реализации продукции.

Строим график корреляционной зависимости (поле корреляции):



Ориентируясь на расположение точек, можно предположить, что связь между показателями прямая и слабая.

Рассчитаем коэффициент корреляции. Для этого построим промежуточные вычисления в виде

№ п/п	x	y	x^2	y^2	$x \cdot y$
1	84	20	7056	400	1680
2	110	24	12100	576	2640
3	123,1	23	15153,61	529	2831,3
4	64	18	4096	324	1152
5	70	14	4900	196	980
6	52	16	2704	256	832
7	86	21	7396	441	1806
8	92	17	8464	289	1564
9	102	22	10404	484	2244
10	41,2	26,3	1697,44	691,69	1083,56
Сумма	824,3	201,3	73971,05	4186,69	16812,86
Среднее	82,4300	20,13	7397,105	418,669	1681,286

Находим коэффициент корреляции по формуле

$$r_{xy} = \frac{\overline{x \cdot y} - \overline{x} \cdot \overline{y}}{\sigma_x \cdot \sigma_y} = \frac{1681{,}286 - 82{,}43 \cdot 20{,}13}{24{,}5438 \cdot 3{,}6637} = 0{,}2441,$$

где

$$\sigma_x = \sqrt{\overline{(x^2)} - (\overline{x})^2} = \sqrt{7397{,}105 - 82{,}43^2} = 24{,}5438;$$

$$\sigma_y = \sqrt{\overline{(y^2)} - (\overline{y})^2} = \sqrt{418{,}669 - 20{,}13^2} = 3{,}6677.$$

Связь между показателями слабая и прямая.

Задание 1.2.4. Ниже приведены коэффициенты эластичности отраслей промышленности Тверской области (данные 1997–2000 годов), которые показывают, на сколько процентов изменились объем промышленного производства и налоговые поступления от отрасли при изменении доли инвестиций, вкладываемых в данную отрасль, на 1 %:

Отрасль	Коэффициент эластичности объема производства по объему инвестиций K_1	Коэффициент эластичности объема налоговых поступлений по объему инвестиций K_2
Электроэнергетика	1,137	3,08
Машиностроение	0,086	1
Пищевая	0,092	0,31
Химическая	0,016	1,48

Найти коэффициент корреляции между эластичностью объема производства и эластичностью объема налоговых поступлений. Выяснить, при каком уровне значимости можно говорить об отсутствии связи между указанными показателями.

Решение

В данной задаче $X = K_1$, $Y = K_2$. Находим:

$$\overline{K}_1 = \frac{1{,}137 + 0{,}086 + 0{,}092 + 0{,}016}{4} = 0{,}333; \; \overline{K}_2 = \frac{3{,}08 + 1 + 0{,}31 + 1{,}48}{4} = 1{,}468;$$

$$\overline{K_1 K_2} = \frac{1{,}137 \;\; 3{,}08 + 0{,}086 \;\; 1 + 0{,}092 \;\; 0{,}31 + 0{,}016 \;\; 1{,}48}{4} = 0{,}91;$$

$$S_{K_1}^2 = S_x^2 = \frac{1{,}137^2 + 0{,}086^2 + 0{,}092^2 + 0{,}016^2}{4} - 0{,}333^2 = 0{,}2163;$$

$$S_{K_1} = S_x = \sqrt{0{,}2163} = 0{,}4651;$$

$$S_{K_2}^2 = S_y^2 = \frac{3{,}08^2 + 1^2 + 0{,}31^2 + 1{,}48^2}{4} - 1{,}468^2 = 1{,}0397;$$

$$S_{K_2} = S_y = \sqrt{1{,}0397} = 1{,}0196; \; \hat{r}_{K_1 K_2} = \hat{r}_{xy} = \frac{0{,}91 - 0{,}333 \;\; 1{,}468}{0{,}4653 \;\; 1{,}0196} = 0{,}8889.$$

Вывод: связь между коэффициентами в данной выборке слабая.

Определим значимость коэффициента корреляции:

$$t_{набл} = r_{xy}\sqrt{\frac{n-2}{1-r_{xy}^2}} = 0,8889\sqrt{\frac{4-2}{1-0,12^2}} = 2,7443.$$

Об отсутствии связи между показателями можно говорить, если наблюдаемое значение t меньше табличного. Найдем табличные значения t при различных уровнях значимости. Количество степеней свободы 4 – 1 – 1 = 2.

При уровне значимости $\alpha = 0,01$ $t_{2;0,01} = 9,9248 > t_{набл} = 2,7443.$

При уровне значимости $\alpha = 0,05$ $t_{2;0,05} = 4,3027 > t_{набл} = 2,7443.$

При уровне значимости $\alpha = 0,1$ $t_{2;0,1} = 2,92 > t_{набл} = 2,7443.$

Отсюда видно, что табличное значение коэффициента корреляции превышает наблюдаемое при всех уровнях значимости. Следовательно, при любом уровне можно говорить об отсутствии связи между показателями.

Задание 1.2.5. Данные о потребительских расходах и денежных доходах на душу населения за ноябрь 1997 года по некоторым областям Центрального федерального округа приведены ниже:

Область	Потребительские расходы на душу населения (тыс. руб.)	Денежные доходы на душу населения (тыс. руб.)
Брянская	364	502
Владимирская	336	539
Ивановская	406	540
Калужская	452	682
Тверская	344	521
Тульская	401	638

Найти коэффициент корреляции между потребительскими расходами и денежными доходами на душу населения. Выяснить, при каком уровне значимости можно говорить об отсутствии связи между указанными показателями.

Решение

Введем переменные: x – денежные доходы на душу населения; y – потребительские расходы на душу населения. Составим таблицу промежуточных вычислений:

№ п/п	x	y	x^2	y^2	$x \cdot y$
1	502	364	252004	132496	182728
2	539	336	290521	112896	181104
3	540	406	291600	164836	219240
4	682	452	465124	204304	308264
5	521	344	271441	118336	179224
6	638	401	407044	160801	255838
Сумма	3422	2303	1977734	893669	1326398
Среднее	570,3333	383,8333	329622,3333	148944,8333	221066,3333

Находим коэффициент корреляции по формуле

$$r_{xy} = \frac{\overline{x \cdot y} - \overline{x} \cdot \overline{y}}{\sigma_x \cdot \sigma_y} = \frac{221066{,}3333 - 570{,}3333 \cdot 383{,}8333}{65{,}8955 \cdot 40{,}2095} = 0{,}8127,$$

где

$$\sigma_x = \sqrt{\overline{(x^2)} - (\overline{x})^2} = \sqrt{329622{,}3333 - 570{,}3333^2} = 65{,}8955,$$

$$\sigma_y = \sqrt{\overline{(y^2)} - (\overline{y})^2} = \sqrt{148944{,}8333 - 383{,}8333^2} = 40{,}2095.$$

Связь между показателями высокая и прямая.

Определим значимость коэффициента корреляции:

$$t_{\text{набл}} = r_{xy}\sqrt{\frac{n-2}{1-r_{xy}^2}} = 0{,}8127\sqrt{\frac{6-2}{1-0{,}8127^2}} = 2{,}7895.$$

Об отсутствии связи между показателями можно говорить, если наблюдаемое значение t меньше табличного. Найдем табличные значения tпри различных уровнях значимости. Количество степеней свободы 6 – 1 – 1 = 4. При уровне значимости $\alpha = 0{,}01$ $t_{4;0,01} = 4{,}6041 > t_{набл} = 2{,}7895.$ При уровне значимости $\alpha = 0{,}05$ $t_{4;0,05} = 2{,}7764 < t_{набл} = 2{,}7895.$ При уровне значимости $\alpha = 0{,}1$ $t_{4;0,1} = 2{,}138 < t_{набл} = 2{,}7895.$

Отсюда видно, что табличное значение коэффициента корреляции превышает наблюдаемое только при уровне значимости $\alpha = 0{,}01.$ При данном уровне можно говорить об отсутствии связи между показателями.

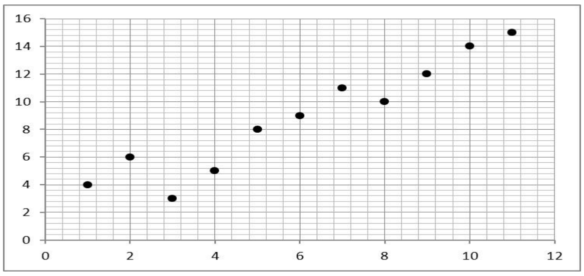
Задание 1.2.6. Рассчитать выборочный коэффициент корреляции для данных, приведенных ниже:

x_i	1	2	3	4	5	6	7	8	9	10	11
y_i	4	6	3	5	8	9	11	10	12	14	15

Построить поле корреляции. Определить значимость выборочного коэффициента корреляции при α = 0,005.

Решение

Строим поле корреляции:



Судя по расположению точек, можно предположить прямую и тесную связь между показателями.

Строим таблицу промежуточных вычислений:

№ п/п	x	y	x^2	y^2	$x\ \ y$
1	1	4	1	16	4
2	2	6	4	36	12
3	3	3	9	9	9
4	4	5	16	25	20
5	5	8	25	64	40
6	6	9	36	81	54
7	7	11	49	121	77
8	8	10	64	100	80
9	9	12	81	144	108
10	10	14	100	196	140
11	11	15	121	225	165
12	1	4	1	16	4
Сумма	67	101	507	1033	713
Среднее	5,5833	8,4167	42,2500	86,0833	59,4167

Находим коэффициент корреляции по формуле

$$r_{xy} = \frac{\overline{x\ \ y} - \bar{x}\ \ \bar{y}}{\sigma_x\ \ \sigma_y} = \frac{59,4167 - 5,5833\ \ 8,4167}{3,3281\ \ 3,9042} = 0,9561,$$

где

$$\sigma_x = \sqrt{\left(\overline{x^2}\right) - \left(\bar{x}\right)^2} = \sqrt{42,25 - 5,5833^2} = 3,3281;$$

$$\sigma_y = \sqrt{\left(\overline{y^2}\right) - \left(\bar{y}\right)^2} = \sqrt{86,0833 - 8,4167^2} = 3,9042.$$

Связь между показателями высокая и прямая.

Определим значимость коэффициента корреляции:

$$t_{набл} = r_{xy}\sqrt{\frac{n-2}{1-r_{xy}^2}} = 0{,}9561\sqrt{\frac{12-2}{1-0{,}9561^2}} = 9{,}377.$$

Найдем табличное значение t при уровне значимости α = 0,005. Количество степеней свободы 11 – 1 – 1 = 9.

При уровне значимости $\alpha = 0{,}005$ $t_{10;\ 0{,}005} = 3{,}6897 < t_{набл} = 9{,}377.$ Отсюда видно, что наблюдаемое значение коэффициента корреляции превышает табличное. Следовательно, коэффициент корреляции значим.

Задание 1.2.7. Рассчитать средний коэффициент эластичности линейной функции $y = 1{,}5 + 0{,}2x$; $\bar{x} = 3{,}5$; $\bar{y} = 2{,}5$. Определить, на сколько процентов в среднем изменится y, если x увеличится на 1 %.

Решение

Коэффициент эластичности $Э = b\,\dfrac{\bar{x}}{\bar{y}} = 0{,}2\,\dfrac{3{,}5}{2{,}5} = 0{,}28.$ Следовательно, при увеличении x на 1 % y возрастет на 0,28 %.

Задание 1.2.8. Проверить значимость выборочного коэффициента корреляции $\hat{r}_{xy} = 0{,}12$, n = 4 при $\alpha = 0{,}05$; $\alpha = 0{,}01$.

Решение

Найдем $t = 0{,}12\sqrt{\dfrac{4-2}{1-0{,}12}} = 0{,}17.$ Табличные значения равны: $t_{2;\ 0{,}95}$ = 4,3 и $t_{2;\ 0{,}99}$ = 9,92. Получаем $t < t_{2;\ 0{,}95}$ и $t < t_{2;\ 0{,}99}$.

Вывод: в обоих случаях гипотеза H_0 принимается, т.е. $r_{xy} = 0$, и с вероятностью 0,95 (0,99) можно утверждать, что и в генеральной совокупности отсутствует связь между признаками.

Задание 1.2.9. При условиях предыдущего примера найти, каков должен быть объем выборки n, чтобы гипотеза о значимости коэффициента $\hat{r}_{xy}$ могла быть принятой при $\alpha = 0{,}1$.

Решение

Имеем $t = 0{,}12\sqrt{\dfrac{n-2}{1-0{,}12^2}} > t_{n-2;\ 0{,}95}$, возведем обе части данного неравенства в квадрат. Получим

$$0{,}144\ \frac{(n-2)}{1-0{,}12^2} > t^2_{n-2;\ 0{,}95},$$

т.е. $0{,}0146\ \ n > 0{,}02922 + t^2_{n-2;\ 0{,}95}$.

Это неравенство не выполняется ни при каком n, т.е. гипотеза о значимости коэффициента не может быть принята.

Задания для самостоятельного выполнения

Задание 1. Рассчитать эмпирический коэффициент корреляции r_{xy} по вариантам.

Вариант 1

x	2	3	8	7
y	3	4	6	8

Вариант 2

x_i	1	2	3	4	5
y_i	–3	–4	–3	–5	–8

Вариант 3

x	–12	–13	–18	–17
y	16	14	12	10

Вариант 4

x_i	–2	–1	–3	–6	–5
y_i	–4	–6	–3	–5	–8

Вариант 5

x	12	10	8	7
y	6	4	2	1

Задание 2. Рассчитать средний коэффициент эластичности линейной функции:

$$y = 1{,}5 + 0{,}2x;\quad \bar{y} = 2{,}5;\quad \bar{x} = 3{,}5\,.$$

Определить, на сколько процентов в среднем изменится y, если x увеличится на 1 %.

Задание 3. В соответствии со своим вариантом методом корреляционного анализа исследовать зависимость между признаками. Построить график корреляционной зависимости.

Вариант 1

Прибыль предприятия (тыс. руб.)	50	55	60	50	65	60	62	61	63	51
Фонд материального поощрения работников (тыс. руб.)	4,0	4,2	4,1	4,2	4,5	4,3	4,0	4,2	4,1	4,0

Вариант 2

Основные производственные фонды в промышленности (тыс. руб.)	8	9	9	10	10,5	11	11,5	11,5	12	13
Производительность труда на одного рабочего в промышленности (%)	21,5	21	22,5	22,5	21,5	22	22,5	22,5	23,0	20

Вариант 3

Среднегодовая стоимость оборотных средств (тыс. руб.)	84	110	1231	64	70	52	86	92	102	41,2
Объем реализации продукции (тыс. руб.)	20	24	23	18	14	16	21	17	22	26,3

Вариант 4

Валовая продукция (тыс. руб.)	25,5	26,8	27	27,2	27,7	28	29	29,2	29,5	29,7
Производительность труда на одного рабочего (%)	20,5	23,5	23	23,5	22,7	23	25	24,8	20,6	20

Вариант 5

Обозначим: x – стоимость основных производственных сил (млн руб.); y – средняя месячная выработка продукции на одного рабочего (тыс. руб.).

x	33,7	35,5	34,5	37,0	38,0	38,3	39,0	39,3	40,0	40,0
y	22,1	24,4	24,0	26,0	25,8	23,7	22,4	25,8	26,0	23,9

Вариант 6

Производственная себестоимость продукции (тыс. руб.)	1,7	2,2	2,7	1,9	2,5	3,0	3,2	2,0	2,4	2,6
Выручка от реализации (тыс. руб.)	22	24	31	22	27	34	35	25	28	28

Вариант 7

Производственная себестоимость продукции (тыс. руб.)	2,9	2,1	2,3	2,2	1,5	2,0	2,2	1,7	1,7	1,8
Прибыль предприятия	6,3	5,1	1,7	4,5	4,3	4,3	4,3	4,2	4,2	4,2

Вариант 8

Обозначим: x – оборачиваемость оборотных средств (дн.); y – прибыль предприятия (тыс. руб.).

x	24	31	27	28	30	22	28	29	32	26
y	118	130	124	128	127	118	124	128	135	120

Вариант 9

Удельный вес продукции высшей категории качества (%)	43,1	35,9	36,9	32,3	34,9	30,2	31,2	48,1	30,1	42,9
Фонд отдачи (руб.)	1,47	1,25	1,82	1,45	1,75	1,37	1,61	1,93	1,68	1,66

Вариант 10

Объем произведенной продукции (тыс. шт.)	2	3	3	4	4	6	9	9	6	13
Себестоимость единицы продукции (руб.)	8	10	12	8	6	6	5	4	5	2

Задание 4. Имеется информация об объемах продаж и затратах на рекламу одной фирмы. Обозначим первые как Y (тыс. руб.), вторые – как X (тыс. руб.) Требуется построить диаграмму рассеяния (корреляционное поле) и определить степень влияния рекламы на объемы продаж (вычислить коэффициент парной корреляции).

Y	225	236	249	292	375	471	522	535	466	457	421	407	431	545
X	4	4,8	3,8	8,7	8,2	9,7	14,7	18,7	19,8	10,6	8,6	6,5	12,6	16,5

Примеры решения задач и выполнения заданий по теме «Оценка качества уравнения регрессии»

Задание 1.3.1. Для данных задачи произвести статистическую оценку точности уравнения регрессии. Дать прогноз значения Y при x_0 = 90 усл. ед. Найти 95%-е доверительные интервалы для индивидуального и среднего значения Y. Исходные данные по 12 регионам приведены ниже:

№ региона	1	2	3	4	5	6	7	8	9	10	11	12
x_i	78	82	87	79	89	106	67	88	73	87	76	115
y_i	133	148	134	154	162	195	139	158	152	162	159	173

Решение

В результате расчетов получается:

$$\overline{X} = 85,58;\ \overline{Y} = 155,75;\ \overline{XY} = 134\ 84;$$

$$S_x^2 = 168,31;\ S_x = 12,97;$$

$$\overline{X^2} = S_x^2 + \left(\overline{X}\right)^2 = 7492,25;$$

$$S_y^2 = 273,36;$$

$$S_y = 16,53;$$

$$\widehat{b}_1 = 0,92,$$

т.е. с увеличением значения фактора X на 1 усл. ед. значение Y возрастает в среднем на 0,92 усл. ед.; $\widehat{b}_0$ = 77,02.

Коэффициент корреляции рассчитывается по формуле

$$r_{xy} = \widehat{b}_1 \ \frac{S_x}{S_y} = 0,92 \ \frac{12,97}{16,53} = 0,722,$$

т.е. связь между X и Y прямая, достаточно сильная и описывается линейной регрессией.

Уравнение регрессии:

$$\widehat{y} = 77 + 0,92 \ \ x.$$

На основе уравнения регрессии строится таблица, в которой для исходных значений x_i указаны соответствующие $\widehat{y}_i$:

x_i	78	82	87	79	89	106	67	88	73	87	76	115
$\widehat{y}_i$	149	152	157	150	159	175	139	158	144	157	147	183

Найдем среднюю ошибку аппроксимации:

$$A=\frac{1}{n}\sum_{i=1}^{n}\left|\frac{\hat{y}_j-y_j}{y_j}\right|\cdot 100\,\%=$$

$$=\frac{1}{12}\left(\left|\frac{149-133}{133}\right|+\left|\frac{152-148}{148}\right|+\ldots+\left|\frac{183-173}{173}\right|\right)\cdot 100\,\%=5{,}7\,\%.$$

Полученное значение свидетельствует о том, что модель для исходных данных подобрана верная, т.е. точность описания экспериментальных данных уравнением регрессии можно считать хорошей.

По формуле $S_{ост}^2=\frac{1}{n-2}\sum_{j=1}^{n}(\hat{y}_j-y_j)^2$ находим:

$$S_{ост}^2=\frac{1}{10}\left[(149-133)^2+(152-148)^2+(157-134)^2+\ldots+(183-173)^2\right]=155{,}9;$$

$$S_{ост}=12{,}49.$$

Расчетное значение T – критерия для проверки значимости коэффициента b_1 – составляет

$$t_{расч}=|\hat{b}_1|\,S_x\sqrt{n}\,/\,S_{ост}=0{,}92\cdot 12{,}97\cdot\sqrt{12}/12{,}49=3{,}31.$$

Поскольку это больше, чем $t_{10;\,0,95}=2{,}23$, то коэффициент b_1 значимо отличается от нуля. Для b_0 $t=3{,}24$, поэтому b_0 тоже значимо отличается от нуля.

Оценки дисперсий коэффициентов регрессии и величины стандартных ошибок коэффициентов регрессии равны:

$$S_{\hat{b}_0}^2=\frac{S_{ост}^2\cdot\overline{X^2}}{nS_x^2}=\frac{155{,}9\cdot 7492{,}25}{12\cdot 168{,}3}=578{,}32;\qquad S_{\hat{b}_1}^2=\frac{S_{ост}^2}{nS_x^2}=\frac{155{,}9}{12\cdot 168{,}31}=0{,}077;$$

$$S_{\hat{b}_0}=\sqrt{578{,}32}=24{,}05;\; S_{\hat{b}_1}=\sqrt{0{,}077}=0{,}277.$$

Доверительные интервалы для коэффициентов уравнения регрессии согласно формулам $\hat{b}_0\pm t_{n-2;\,1-\alpha}\cdot S_{\hat{b}_0}$; $\hat{b}_1\pm t_{n-2;\,1-\alpha}\cdot S_{\hat{b}_1}$:

$$\hat{b}_0\pm t_{10;\,0,95}\cdot S_{\hat{b}_0}=77\pm 2{,}23\cdot 24{,}05=77\pm 53{,}63=(23\ldots 131);$$

$$\hat{b}_1\pm t_{10;\,0,95}\cdot S_{\hat{b}_1}=0{,}92\pm 2{,}23\cdot 0{,}277=0{,}92\pm 0{,}63=(0{,}3\ldots 1{,}54).$$

Применяя формулы $\hat{y}_0=\hat{b}_0+\hat{b}_1\cdot x_0$ и $\hat{y}_0\pm t_{n-2;\,1-\alpha}\cdot S_{\hat{y}_0}$, находим прогноз зависимой переменной для $x_0=90$ усл. ед. и доверительный интервал для прогноза:

$$\hat{y}_0=77+0{,}92\cdot 90=159{,}8\ \text{ден. ед.};$$

$$S_{\hat{y}_0}=12{,}49\sqrt{1+\frac{1}{12}+\frac{(90-85{,}58)^2}{12\cdot 168{,}31}}=13{,}06;$$

$$\hat{y}_0\pm t_{10;\,0,95}\cdot S_{\hat{y}_0}=159{,}8\pm 2{,}23\cdot 13{,}06=159{,}8\pm 29{,}12=(130{,}7\ \ 188{,}9).$$

Доверительный интервал для среднего значения фактора Y определяется согласно формулам:

$$\widehat{M}_x(Y) = \widehat{y}_0;$$

$$\widehat{y}_0 \pm t_{n-2,1-\alpha} S_{\widehat{M}_x(Y)};$$

$$S_{\widehat{M}_x(Y)} = S_{ocm}\sqrt{\frac{1}{n} + \frac{(x_0 - \overline{X})^2}{nS_x^2}}.$$

Тогда

$$S_{\widehat{M}_x(Y)} = 12{,}49\sqrt{\frac{1}{12} + \frac{(90-85{,}58)^2}{12\;\;168{,}31}} = 3{,}81;$$

$$\widehat{y}_0 \pm t_{10;0,95}\;\; S_{\widehat{M}_x(Y)} = 159{,}8 \pm 2{,}23\;\;3{,}81 = (151{,}3\quad 163{,}6).$$

Задание 1.3.2. Построить уравнение регрессии на основании приведенных ниже данных о динамике процента хронических больных на 1 тыс. больных:

Год t	0	1	2	3	4	5	6
Хронические больные (%)	12	10	6	4	4	5	2

Оценить точность и значимость уравнения регрессии и его коэффициентов (для α = 0,005). Что можно сказать о прогнозе количества таких больных, когда t = 7, 8, 9 (t – усл. ед. времени)?

Решение

Находим параметры уравнения регрессии по формулам:

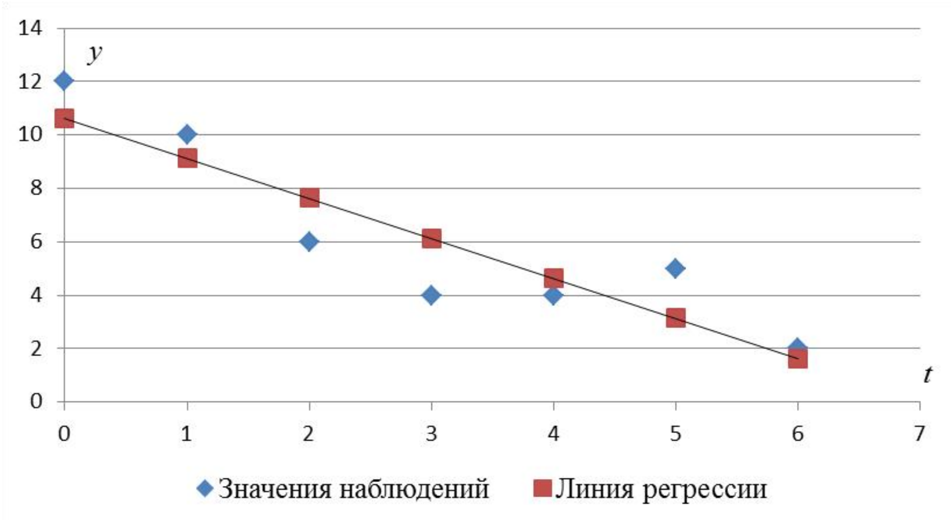
$$b = \frac{\overline{t\;\;y_t} - \bar{t}\;\;\bar{y}}{\sigma_t^2} = \frac{12{,}4286 - 3\;\;6{,}1429}{2^2} = -1{,}5;$$

$$a = \overline{y}_t - b\;\;\overline{t} = 6{,}1429 + 1{,}5\;\;3 = 10{,}6429,$$

где $\sigma_t = \sqrt{\overline{\left(t^2\right)} - \left(\bar{t}\right)^2} = \sqrt{13-3^2} = 2.$

Уравнение регрессии имеет вид $\hat{y} = 10{,}6429 - 1{,}5x.$

Построим графически линию регрессии:



Для оценки качества параметров уравнения построим расчетную таблицу:

t	y	y(t)	$(y_i - y_{cp})^2$	$(y_i - y(t))^2$	$(y_i - y(t)) : y_i$
0	12	10,643	34,306	1,842	0,113
1	10	9,143	14,878	0,735	0,0857
2	6	7,643	0,0204	2,699	0,274
3	4	6,143	4,592	4,592	0,536
4	4	4,643	4,592	0,413	0,161
5	5	3,143	1,306	3,449	0,371
6	2	1,643	17,163	0,128	0,179
		43	76,857	13,857	1,719

Средняя ошибка аппроксимации

$$\bar{A} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| 100\,\% = \frac{1,719}{7}\,100\,\% = 24,56\,\%.$$

Поскольку ошибка больше 7 %, то данное уравнение нежелательно использовать в качестве тренда.

Коэффициент детерминации

$$R^2_{xy} = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} = 1 - \frac{13,857}{76,857} = 0,8197.$$

Таким образом, в 81,97 % случаев t влияет на изменение y. Другими словами, точность подбора уравнения тренда высокая.

Коэффициент детерминации

$$F_{набл} = \frac{R_{xy}^2}{1 - R_{xy}^2} \ \frac{(n-m-1)}{m} = \frac{0{,}8197}{1-0{,}8197} \ \frac{(9-1-1)}{1} = 22{,}732.$$

Находим $F_{\text{кр}}(1;\,5;\,0{,}05) = 0$, где m – количество факторов в уравнении тренда ($m = 1$).

Поскольку $F > F_{\text{кр}}$, то коэффициент детерминации (и в целом уравнение тренда) статистически значим.

Дисперсия ошибки уравнения

$$S^2 = \frac{\sum(y_i - y_x)^2}{n-m-1} = \frac{13{,}857}{5} = 2{,}7714,$$

где $m = 1$ – количество влияющих факторов в модели тренда.

Проверим значимость параметров уравнения регрессии.

Находим стандартную ошибку уравнения:

$$\hat{\sigma} = \sqrt{\frac{(y_t - \hat{y}_t)^2}{7-1-1}} = \sqrt{\frac{13{,}8571}{5}} = 1{,}6648.$$

Определяем стандартные ошибки коэффициентов и значения их t-критериев:

$$\hat{\sigma}_b = \hat{\sigma} \ \sqrt{\frac{t^2}{n^2\sigma_t^2}} = 1{,}6648 \ \sqrt{\frac{91}{2^2 \ 7^2}} = 1{,}13437; \quad t_b = \frac{|b|}{\hat{\sigma}_b} = \frac{1{,}5}{1{,}13437} = 1{,}32232;$$

$$\hat{\sigma}_a = \sqrt{\frac{\hat{\sigma}^2}{n \ \hat{\sigma}_t^2}} = \sqrt{\frac{1{,}6648^2}{7 \ 2^2}} = 0{,}314618; \quad t_a = \frac{|a|}{\hat{\sigma}_a} = \frac{10{,}6429}{0{,}314618} = 33{,}82805.$$

Найдем табличное значение t при уровне значимости $\alpha = 0{,}005$. Количество степеней свободы 7 – 1 – 1 = 5, $t_{5;\ 0{,}005} = 4{,}7733$.

Для параметра a $t_{5;\ 0{,}005} = 4{,}7733 < t_{набл} = 33{,}82805.$ Следовательно, при уровне значимости $\alpha = 0{,}005$ параметр a значим.

Для параметра b $t_{5;\ 0{,}005} = 4{,}7733 > t_{набл} = 1{,}32232.$ Следовательно, при уровне значимости $\alpha = 0{,}005$ параметр b не значим.

Прогнозирование количества больных на последующие три года:

$$t = 7;\ \hat{y}_t = 10{,}6429 - 1{,}5 \ \ 7 = 0{,}14\ \%.$$

Если $t = 8$, то $y(8) = -1{,}5 \cdot 8 + 10{,}643 = -1{,}36$ %, а при $t = 9$ прогноз имеет следующий вид: $y(9) = -1{,}5 \cdot 9 + 10{,}643 = -2{,}86$ %.

Таким образом, при $t = 8$ и $t = 9$ будет происходить уменьшение количества хронических больных.

Задание 1.3.3. Рассчитать коэффициенты уравнения регрессии для данных, приведенных ниже:

t_i	2014	2015	2016	2017	2018	2019	2020	2021	2022
y_i	8	9	12	11	15	12	14	16	18

Построить графически линию регрессии. Дать прогноз на 2023 год. Определить качество уравнения регрессии (рассчитать среднюю ошибку аппроксимации, значимость уравнения регрессии и его коэффициентов для α = 0,005). Найти интервальные оценки коэффициентов уравнения регрессии и прогноза.

Решение

Для упрощения вычислений обозначим периоды как t = 1, 2, 3, , 9. Составим таблицу промежуточных вычислений:

№	t	y_t	t^2	y^2	$t\ y$	$\hat{y}_t$	$\left(y_t - \hat{y}_t\right)^2$	$\dfrac{\lvert y_t - \hat{y}_t\rvert}{y_t}$
1	1	8	1	64	8	8,378	0,143	0,047
2	2	9	4	81	18	9,478	0,229	0,053
3	3	12	9	144	36	10,578	2,023	0,119
4	4	11	16	121	44	11,678	0,459	0,062
5	5	15	25	225	75	12,778	4,938	0,148
6	6	12	36	144	72	13,878	3,526	0,157
7	7	14	49	196	98	14,978	0,956	0,07
8	8	16	64	256	128	16,078	0,006	0,005
9	9	18	81	324	162	17,178	0,676	0,046
Сумма	45	115	285	1555	641	115	12,956	0,706
Среднее	5	12,778	31,667	172,778	71,222	12,778	1,44	0,078

Находим параметры уравнения регрессии по формулам:

$$b = \frac{\overline{t\ y} - \bar{t}\ \bar{y}}{\sigma_t^2} = \frac{71,2222 - 5\ \ 12,7778}{2,582^2} = 1,1;$$

$$a = \bar{y}_t - b\ \ \bar{t} = 12,7778 + 1,1\ \ 5 = 7,2778.$$

Рассчитываем коэффициент корреляции по формуле

$$r_{xy} = \frac{\overline{t\ y} - \bar{t}\ \bar{y}}{\sigma_t\ \ \sigma_y} = \frac{71,2222 - 5\ \ 12,7778}{2,582\ \ 3,0832} = 0,9212,$$

где

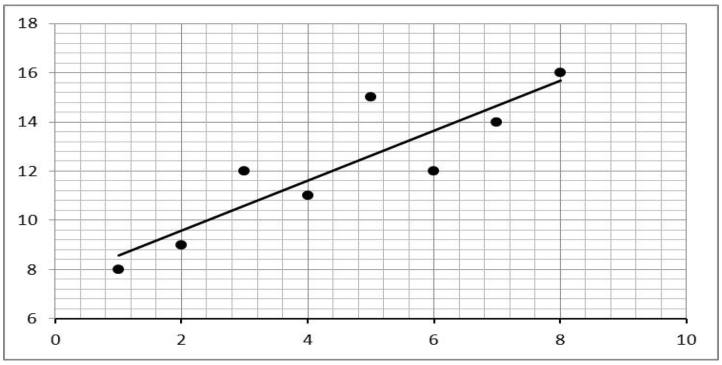
$$\sigma_t = \sqrt{\left(\overline{t^2}\right) - \left(\bar{t}\right)^2} = \sqrt{31,6667 - 5^2} = 2,582;$$

$$\sigma_y = \sqrt{\left(\overline{y^2}\right) - \left(\bar{y}\right)^2} = \sqrt{172,7778 - 12,7778^2} = 3,0832.$$

Уравнение регрессии имеет вид $\hat{y} = 7,2778 - 1,1x$.

Связь между показателями прямая и весьма высокая.

Построим графически линию регрессии:



Рассчитаем прогноз на 2016 год:

$$t_{2016} = 10;\ y_{2016} = 7,2778 + 1,1 \cdot 10 = 18,28.$$

Определим качество уравнения регрессии.

Средняя ошибка аппроксимации

$$A = \frac{1}{n}\sum\frac{\left|y_t - \hat{y}_t\right|}{y_t} \cdot 100\ \% = 0,0784 \cdot 100\ \% = 7,84.$$

Так как $A < 10$ %, данное уравнение подходит для прогнозирования.

Проверим значимость уравнения регрессии и его коэффициентов для α = 0,005. Найдем коэффициент детерминации: $R^2 = \left(0,9212\right)^2 = 0,8486.$ Определим значение F-критерия Фишера:

$$F_{набл} = \frac{R^2}{1 - R^2} \cdot \left(n - m - 1\right) = \frac{0,8486}{1 - 0,8486} \cdot \left(9 - 1 - 1\right) = 39,2264.$$

Найдем табличное значение F-критерия при уровне значимости α = 0,005. Количество степеней свободы $k_1 = 1$; $k_2 = 9 - 1 - 1 = 7$.

Так как $F_{набл} = 39,2264 > F_{табл} = 16,2356$, то при α = 0,005 уравнение признается значимым.

Проверим значимость параметров уравнения регрессии для α = 0,005.

Находим стандартную ошибку уравнения:

$$\hat{\sigma} = \sqrt{\frac{\sum\left(y_t - \hat{y}_t\right)^2}{n - 1 - 1}} = \sqrt{\frac{12,9556}{7}} = 1,3604.$$

Определим стандартные ошибки коэффициентов и значения их t-критериев:

$$\hat{\sigma}_b = \hat{\sigma}\ \sqrt{\frac{t^2}{n^2\sigma_t^2}} = 1{,}3604\ \sqrt{\frac{285}{9^2\ \ 2{,}582^2}} = 0{,}1756; \quad t_b = \frac{|b|}{\hat{\sigma}_b} = \frac{1{,}1}{0{,}1756} = 6{,}2631;$$

$$\hat{\sigma}_a = \sqrt{\frac{\hat{\sigma}^2}{n\ \ \hat{\sigma}_t^2}} = \sqrt{\frac{1{,}3604^2}{9\ \ 2{,}582^2}} = 0{,}9883; \quad t_a = \frac{|a|}{\hat{\sigma}_a} = \frac{7{,}2778}{0{,}9883} = 7{,}3637.$$

Найдем табличное значение t при уровне значимости α = 0,005. Количество степеней свободы равно 9 – 1 – 1 = 7; $t_{7;\ 0{,}005} = 4{,}0293$.

Для параметра a $t_{7;\ 0{,}005} = 4{,}0293 < t_{набл} = 6{,}2631$. Следовательно, при уровне значимости α = 0,005 параметр a значим.

Для параметра b $t_{7;\ 0{,}005} = 4{,}0293 < t_{набл} = 7{,}3637$. Следовательно, при уровне значимости α = 0,005 параметр b значим.

Рассчитаем интервальные оценки коэффициентов уравнения регрессии:

$$\left(b - t_{табл}\,\hat{\sigma}_b;\ b + t_{табл}\,\hat{\sigma}_b\right); \left(1{,}1 - 4{,}0293\ \ 0{,}1756;\ 1{,}1 + 4{,}0293\ \ 0{,}1756\right);$$

$$I = \left(0{,}3923;\ 1{,}8077\right).$$

С вероятностью 99,5 % можно утверждать, что значение данного параметра будет лежать в найденном интервале:

$$\left(a - t_{табл}\,\hat{\sigma}_a;\ a + t_{табл}\,\hat{\sigma}_a\right) =$$

$$= \left(7{,}2778 - 4{,}0293\ \ 0{,}9883;\ 7{,}2778 + 4{,}0293\ \ 0{,}9883\right);$$

$$I = \left(3{,}2954;\ 11{,}2601\right).$$

Рассчитаем доверительный интервал прогноза.

Предельная ошибка

$$e = t_{табл}\ \ \hat{\sigma}\ \sqrt{1 + \frac{1}{n} + \frac{\left(t - t_{пр}\right)^2}{\left(t - \overline{t}\right)^2}} = 4{,}0293\ \ 1{,}3604\ \sqrt{1 + \frac{1}{9} + \frac{\left(5 - 10\right)}{60}} = 6{,}78.$$

Доверительный интервал прогноза тогда будет

$$\left(y_{2016} - e;\ y_{2016} + e\right) = \left(18{,}28 - 6{,}78;\ 18{,}28 + 6{,}78\right) = \left(11{,}5;\ 25{,}06\right).$$

С вероятностью 99,5 % можно гарантировать, что прогнозное значение будет находиться в полученном интервале.

Задание 1.3.4. Задана корреляционная таблица:

X	Y					
	0	2	4	6	8	n_x
1	3					3
3	2	3	5			10
5		9	8			17
7			2	6		8
9				4	1	5
11					7	7
n_y	5	12	15	10	8	50

Требуется найти:
1) выборочные средние арифметические;
2) среднеквадратические отклонения;
3) коэффициент корреляции;
4) уравнение линейной регрессии Y на X;
5) значимость уравнения регрессии.

Решение

Выборочные средние арифметические:

$$\overline{X}=\frac{1\ \ 3+3\left(2+3+5\right)+5\left(9+8\right)+7\left(2+6\right)+9\left(4+1\right)+11\ \ 7}{50}=5,92;$$

$$\overline{Y}=\frac{0\left(3+2\right)+2\left(3+9\right)+4\left(5+8+2\right)+6\left(6+4\right)+8\left(1+7\right)}{50}=4,16.$$

Дисперсии:

$$\sigma_X^2=\frac{1^2\ \ 3+3^2\left(2+3+5\right)+5^2\left(9+8\right)+7^2\left(2+6\right)+9^2\left(4+1\right)+11^2\ \ 7}{50}\ \ 5,92^2=\ 8,19;$$

$$\sigma_Y^2=\frac{0^2\left(3+2\right)+2^2\left(3+9\right)+4^2\left(5+8+2\right)+6^2\left(6+4\right)+8^2\left(1+7\right)}{50}-4,16^2=\ 5,89.$$

Среднеквадратические отклонения:

$$\sigma_X=\sqrt{8,19}=2,862\ ;\ \sigma_Y=\sqrt{5,89}=2,428.$$

Для нахождения коэффициента корреляции строим в корреляционной таблице вспомогательные строки и столбцы:

X	Y					Сумма	$n_i\ \ x_i$	$n_i\ \left(x_i\right)^2$
	0	2	4	6	8			
1	3					3	3	3
3	2	3	5			10	30	90
5		9	8			17	85	425
7			2	6		8	56	392
9				4	1	5	45	405
11					7	7	77	847
Сумма	5	12	15	10	8	$n=50$	296	2162
$n_j\ \ y_j$	0	24	60	60	64	208		
$n_j\ \left(y_j\right)^2$	0	48	240	360	512	1160		
$n_{ij}\ \ x_i\ \ y_j$	0	108	276	468	688	1540		

Найдем коэффициент корреляции по формуле

$$\widehat{r}_{xy} = \frac{n \underset{i,j}{\;} n_{ij}\ x_i\ y_j - \underset{i}{\;} n_i x_i \quad \underset{j}{\;} n_j y_j}{\sqrt{n \underset{i}{\;} n_i \left(x_i\right)^2 - \quad \underset{i}{\;} n_i x_i \quad ^2} \sqrt{n \underset{j}{\;} n_j \left(y_j\right)^2 - \quad \underset{j}{\;} n_j y_j \quad ^2}}.$$

Отсюда

$$\widehat{r}_{xy} = \frac{50\ \ 1540 + 208\ \ 296}{\sqrt{50\ \ 2162 - 296^2}\sqrt{50\ \ 1160 - 208^2}} = 0{,}8882.$$

Следовательно, связь высокая и прямая.

Уравнение линейной регрессии Y на X имеет вид

$$y_x = r_{xy}\,\frac{x - \overline{x}}{\sigma_x}\sigma_y + \overline{y}.$$

Построим уравнение линии регрессии Y на X:

$$y_x = 0{,}8882\ \ \frac{x - 5{,}92}{2{,}862}\ \ 2{,}428 + 4{,}16\ , \text{ или } y_x = 0{,}75x - 0{,}3.$$

Оценим статистическую значимость уравнения с помощью F-критерия Фишера:

$$F_{набл} = \frac{r_{xy}^2}{1 - r_{xy}^2}\ \ \frac{n - m - 1}{m} = \frac{0{,}8882^2}{1 - 0{,}8882^2}\ \ \frac{50 - 1 - 1}{1} = 179{,}44.$$

Найдем табличное значение F-критерия при уровне значимости α = 0,005. Количество степеней свободы $k_1 = 1$; $k_2 = 50 - 1 - 1 = 48.$ Получаем $F_{табл} = 4{,}0427$.

Так как $F_{набл} = 179{,}44 > F_{табл} = 4{,}0427$, то при α = 0,005 уравнение признается значимым.

Задание 1.3.5. Рассчитать коэффициенты уравнения регрессии для данных, приведенных ниже:

t_i	2007	2008	2009	2010	2011	2012	2013	2014	2015
y_i	8	9	12	11	15	12	14	16	18

Построить графически линию регрессии. Дать прогноз на 2016 год. Определить качество уравнения регрессии (рассчитать среднюю ошибку аппроксимации, значимость уравнения регрессии и значимость его коэффициентов для α = 0,05). Найти интервальные оценки коэффициентов уравнения регрессии и прогноза.

Решение

Линейное уравнение регрессии имеет вид

$$у = b_0 + b_1 x,$$

где $b_1 = \dfrac{K_{xy}}{S_x^2}$; $b_0 = \overline{Y} - b_1 \overline{X}$; $K_{xy} = \overline{XY} - \overline{X} \cdot \overline{Y}$.

Для упрощения вычислений поместим начало отсчета в 2007 год и составим вспомогательную таблицу:

x_i	0	1	2	3	4	5	6	7	8
y_i	8	9	12	11	15	12	14	16	18

Выполняем вычисления:

$$\overline{X} = \frac{0+1+2+3+4+5+6+7+8}{9} = \frac{36}{9} = 4{,}00;$$

$$\overline{Y} = \frac{8+9+12+11+15+12+14+16+18}{9} = \frac{115}{9} = 12{,}78;$$

$$\overline{XY} = \frac{0\cdot 8 + 1\cdot 9 + 2\cdot 12 + 3\cdot 11 + 4\cdot 15}{9} +$$

$$+ \frac{5\cdot 12 + 6\cdot 14 + 7\cdot 16 + 8\cdot 18}{9} = \frac{526}{9} = 58{,}44;$$

$$\overline{X^2} = \frac{0^2+1^2+2^2+3^2+4^2+5^2+6^2+7^2+8^2}{9} = \frac{204}{9} = 22{,}67;$$

$$\overline{Y^2} = \frac{8^2+9^2+12^2+11^2+15^2+12^2+14^2+16^2+18^2}{9} = \frac{1555}{9} =$$

$$= 172{,}78;$$

$$S_x^2 = \overline{X^2} - (\overline{X})^2 = 22{,}67 - 4{,}00^2 = 6{,}67;$$

$$S_x = 2{,}58;$$

$$S_y^2 = \overline{Y^2} - (\overline{Y})^2 = 172{,}78 - 12{,}78^2 = 9{,}45;$$

$$S_y = 3{,}07.$$

Ковариация составляет

$$K_{xy} = \overline{XY} - \overline{X} \cdot \overline{Y} = 58{,}44 - 4{,}00 \cdot 12{,}78 = 7{,}32.$$

Коэффициент корреляции

$$r_{xy} = \frac{K_{xy}}{S_x \cdot S_y} = \frac{7{,}32}{2{,}58 \cdot 3{,}07} = 0{,}92.$$

Значение коэффициента корреляции, близкое к 1, свидетельствует о том, что полученное уравнение хорошо описывает экспериментальные данные.

Найдем коэффициенты уравнения регрессии:

$$b_1 = \frac{K_{xy}}{S_x^2} = \frac{7{,}32}{6{,}67} = 1{,}10;$$

$$b_0 = \overline{Y} - b_1\overline{X} = 12{,}78 - 1{,}10 \cdot 4{,}00 = 8{,}38.$$

Составим уравнение линейной регрессии:

$$y = 8{,}38 + 1{,}10x.$$

В принятых обозначениях 2016 год будет иметь номер 9. Если под y понимать, например, ожидаемую прибыль, то она должны быть

$$y = 8{,}38 + 1{,}10 \cdot 9 = 18{,}28 \text{ усл. ед.}$$

Коэффициент регрессии b_1 показывает, на сколько единиц в среднем изменяется переменная Y при увеличении переменной X на одну единицу. Следовательно, каждый год в течение рассматриваемого периода Y увеличивается в среднем на 1,1 усл. ед.

Составим расчетную таблицу. Значения $\hat{y}_x$ найдем, подставляя значения x в уравнение регрессии. Получаем

x_i	0	1	2	3	4	5	6	7	8
y_i	8	9	12	11	15	12	14	16	18
$\hat{y}_x$	8,38	9,48	10,58	11,68	12,78	13,88	14,98	16,08	17,18

Чтобы определить качество уравнения регрессии, сначала рассчитаем среднюю ошибку аппроксимации по формуле

$$A = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| 100\ \%;$$

$$\mathrm{A} = \frac{1}{9}\cdot\left(\left|\frac{8{,}38-8}{8}\right| + \left|\frac{9{,}48-9}{9}\right| + \left|\frac{10{,}58-12}{12}\right| + \left|\frac{11{,}68-11}{11}\right| + \right.$$

$$+ \left|\frac{12{,}78-15}{15}\right| + \left|\frac{13{,}88-12}{12}\right| + \left|\frac{14{,}98-14}{14}\right| + \left|\frac{16{,}08-16}{16}\right| +$$

$$\left. + \left|\frac{17{,}18-18}{18}\right|\right) 100\ \% = \frac{0{,}7062}{9}\cdot 100\,\% = 7{,}85\ \%.$$

Полученное значение средней ошибки аппроксимации свидетельствует о том, что модель для исходных данных была подобрана верная, т.е. точность описания экспериментальных данных уравнением регрессии можно считать удовлетворительной.

Оценку дисперсии ошибок (остатков) предсказания зависимой переменной находим по формуле

$$S^2_{\text{ост}} = \frac{1}{n-2}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right).$$

Получаем:

$$S_{\text{ост}}^2 = \frac{(8{,}38-8)^2+(9{,}48-9)^2+(10{,}58-12)^2}{7}+$$

$$+\frac{(11{,}68-11)^2+(12{,}78-15)^2+(13{,}88-12)^2}{7}+$$

$$+\frac{(14{,}98-14)^2+(16{,}08-16)^2+(17{,}18-18)^2}{7}=\frac{12{,}96}{7}=1{,}85;$$

$$S_{ост}=1{,}36.$$

Оценки дисперсий коэффициентов регрессии и величины стандартных ошибок коэффициентов регрессии определяются по формулам:

$$S_{\hat{b}_0}=\frac{S_{\text{ост}}\cdot\bar{X}}{\sqrt{n}\cdot S_x};\quad S_{\hat{b}_1}=\frac{S_{\text{ост}}}{\sqrt{n}\cdot S_x};$$

$$S_{\hat{b}_0}=\frac{S_{\text{ост}}\cdot\bar{\text{X}}}{\sqrt{n}\cdot S_{\text{x}}}=\frac{1{,}36\cdot 4}{\sqrt{9}\cdot 2{,}58}=0{,}70;\quad S_{\hat{b}_1}=\frac{1{,}36}{\sqrt{9}\cdot 2{,}58}=0{,}18.$$

Критерий Стьюдента для проверки значимости коэффициентов b_0 и b_1 выглядит следующим образом:

$$\frac{|\hat{b}_0|}{S_{\hat{b}_0}}=\frac{8{,}38}{0{,}70}=11{,}97;\quad \frac{|\hat{b}_1|}{S_{\hat{b}_1}}=\frac{1{,}10}{0{,}18}=6{,}11.$$

По табл. 2 прилож. определяем значение коэффициента Стьюдента:

$$t_{n-2;1-\alpha}=t_{7;0{,}95}=2{,}36.$$

Так как 11,97 > 2,36 и 6,11 > 2,36, то и коэффициент b_0, и коэффициент b_1 значимо отличаются от нуля.

Доверительные интервалы для коэффициентов b_0 и b_1:

$$\hat{b}_0\pm t_{7;0{,}95}\cdot S_{\bar{b}_0}=8{,}38\pm 2{,}36\cdot 0{,}70=8{,}38\pm 1{,}65=(6{,}73\div 10{,}03);$$

$$\hat{b}_1\pm t_{7;0{,}95}\cdot S_{\bar{b}_1}=1{,}10\pm 2{,}36\cdot 0{,}18=1{,}10\pm 0{,}42=(0{,}68\div 1{,}52).$$

Прогноз $\hat{y}_0$ зависимой переменной y для $x = x_0$ определяется по формуле

$$\hat{y}_0=\hat{b}_0+\hat{b}_1\cdot\text{x}_0.$$

Доверительный интервал для прогноза:

$$\hat{y}_0\pm t_{n-2;1-\alpha}\cdot S_{\hat{y}_0};$$

$$S_{\hat{y}_0}=S_{\text{ост}}\cdot\sqrt{1+\frac{1}{n}+\frac{(\text{x}_0-\bar{\text{x}})^2}{n\cdot S_{\text{x}}^2}}.$$

Поскольку требуется дать прогноз для 2016 года, то примем $x_0 = 9$. Тогда

$$y_0=8{,}38+1{,}10\cdot 9=18{,}28;$$

$$S_{\hat{y}_0} = 1{,}36 \cdot \sqrt{1 + \frac{1}{9} + \frac{(9-4)^2}{9 \cdot 2{,}58^2}} = 1{,}53;$$

$$\hat{y}_0 \pm t_{n-2;1-\alpha} \cdot S_{\hat{y}_0} = 18{,}28 \pm 2{,}36 \cdot 1{,}53 = 18{,}28 \pm 3{,}61 = (14{,}67 \div 21{,}89).$$

Точечная и интервальная оценка среднего значения Y (условного математического ожидания)

$$S_{\hat{M}_t(Y)} = S_{\text{ост}} \cdot \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{n \cdot S_x^2}}.$$

При $x = 9$

$$S_{\hat{M}_t(Y)} = 1{,}36 \cdot \sqrt{\frac{1}{9} + \frac{(9-4)^2}{9 \cdot 2{,}58^2}} = 0{,}99;$$

$$\hat{y}_0 \pm t_{7;0{,}95} \cdot S_{\hat{M}_t(Y)} = 18{,}28 \pm 2{,}36 \cdot 0{,}99 = 18{,}28 \pm 2{,}34 = (15{,}94 \div 20{,}62).$$

Задание 1.3.6. По 9 регионам имеются данные о среднедневной заработной плате за 2023 год:

Номер региона	Среднедушевой прожиточный минимум x в день одного трудоспособного, руб.	Среднедневная заработная плата y, руб.
1	449,20	1881,23
2	490,97	2015,14
3	464,83	1820,36
4	619,37	2934,91
5	605,87	2935,55
6	506,63	2097,77
7	459,63	2254,45
8	438,73	2212,77
9	492,67	1929,86

Требуется:

1) построить уравнение линейной парной регрессии y от x;

2) рассчитать линейный коэффициент корреляции и среднюю ошибку аппроксимации;

3) оценить статистическую значимость параметров регрессии и корреляции;

4) дать прогноз заработной платы y при $x = 1{,}07\ \bar{x}$;

5) оценить точность прогноза, рассчитав ошибку и доверительный интервал прогноза;

6) на одном графике изобразить исходные данные и построить прямую уравнения регрессии.

Решение

Составим таблицы промежуточных вычислений:

№ п/п	x	y	x^2	y^2	$x \cdot y$
1	449,20	1881,23	201780,64	3539026,31	845048,52
2	490,97	2015,14	241048,27	4060789,22	989366,57
3	464,83	1820,36	216070,03	3313710,53	846164,01
4	619,37	2934,91	383615,07	8613696,71	1817785,42
5	605,87	2935,55	367074,42	8617453,80	1778551,89
6	506,63	2097,77	256677,33	4400638,97	1062800,21
7	459,63	2254,45	211262,80	5082544,80	1036220,37
8	438,73	2212,77	192486,94	4896351,07	970815,96
9	492,67	1929,86	242720,44	3724359,62	950777,69
Сумма	4527,90	20082,04	2312735,94	46248571,04	10297530,64
Среднее	503,10	2231,34	256970,66	5138730,12	1144170,07

№ п/п	x	y	$\hat{y}$	$(x-\bar{x})^2$	$(y-\hat{y})^2$	$\frac{\lvert y-\hat{y}\rvert}{y}$
1	449,20	1881,23	1929,94	2905,21	122575,46	0,03
2	490,97	2015,14	2163,41	147,22	46741,48	0,07
3	464,83	1820,36	2017,33	1464,34	168902,73	0,11
4	619,37	2934,91	2881,17	13517,94	495013,87	0,02
5	605,87	2935,55	2805,70	10560,99	495914,85	0,04
6	506,63	2097,77	2250,99	12,48	17840,35	0,07
7	459,63	2254,45	1988,26	1889,35	534,17	0,12
8	438,73	2212,77	1871,43	4143,07	344,76	0,15
9	492,67	1929,86	2172,92	108,85	90888,85	0,13
Сумма	4527,90	20082,04	20081,15	34749,45	1438756,53	0,74
Среднее	503,10	2231,34	2231,24	3861,05	159861,84	0,08

Найдем параметры уравнения регрессии по формулам:

$$b=\frac{\overline{x\cdot y}-\bar{x}\cdot\bar{y}}{\sigma_x^2}=\frac{1144170,07-503,1\cdot 2231,34}{62,14^2}=5,59;$$

$$a=\bar{y}-b\cdot\bar{x}=2231,34-5,59\cdot 503,10=-581,09.$$

Рассчитаем коэффициент корреляции по формуле

$$r_{xy} = \frac{\overline{x\ y} - \bar{x}\ \bar{y}}{\sigma_x\ \sigma_y} = \frac{1144170,07 - 503,10\ \ 2231,34}{62,14\ \ 399,85} = 0,87,$$

где

$$\sigma_x = \sqrt{\left(\overline{x^2}\right) - \left(\bar{x}\right)^2} = \sqrt{256970,66 - 503,10^2} = 62,14;$$

$$\sigma_y = \sqrt{\left(\overline{y^2}\right) - \left(\bar{y}\right)^2} = \sqrt{5138730,12 - 2231,34^2} = 399,83.$$

Уравнение регрессии имеет вид $\hat{y} = 5,59x - 581,09.$

Так как $r_{xy} = 0,87,$ связь между показателями прямая и высокая.

Средняя ошибка аппроксимации

$$A = \frac{1}{n}\ \ \frac{\left|y_t - \hat{y}_t\right|}{y_t}\ 100\ \% = 0,08\ \ 100\ \% = 8\ \%.$$

Полученное значение свидетельствует о крайне высокой, но допустимой ошибке аппроксимации. Так как A < 10 %, то данное уравнение подходит для прогнозирования.

Проверим значимость уравнения регрессии и его коэффициентов для $\alpha = 0,05$.

Найдем коэффициент детерминации: $R^2 = \left(0,87\right)^2 = 0,75$.

Определим значение F-критерия Фишера:

$$F_{набл} = \frac{R^2}{1 - R^2}\ \left(n - m - 1\right) = \frac{0,75}{1 - 0,75}\ \left(9 - 1 - 1\right) = 21,54.$$

Найдем табличное значение F-критерия при уровне значимости α = 0,05. Количество степеней свободы $k_1 = 1$; $k_2 = 9 - 1 - 1 = 7$. Будем использовать табл. 1 прилож. Получаем $F_{табл} = 5,59$.

Так как $F_{набл} = 21,54 > F_{табл} = 5,59$, при $\alpha = 0,05$ уравнение признается значимым.

Проверим значимость параметров уравнения регрессии для $\alpha = 0,05$.

Рассчитаем стандартную ошибку уравнения:

$$\hat{S} = \sqrt{\frac{\left(y - \hat{y}\right)^2}{n - 1 - 1}} = \sqrt{\frac{1438756,53}{7}} = 453,36.$$

Найдем стандартные ошибки коэффициентов и значения их t-критериев:

$$\hat{S}_b = \hat{S}\ \ \sqrt{\frac{x^2}{n^2 \sigma_x^2}} = 453,36\ \ \sqrt{\frac{2312735,94}{9^2\ \ 62,14^2}} = 2,72;$$

$$t_b = \frac{|b|}{\hat{\sigma}_b} = \frac{5,59}{2,72} = 2,06;$$

$$\hat{\sigma}_a = \sqrt{\frac{\hat{\sigma}^2}{n \;\; \hat{\sigma}_x^2}} = \sqrt{\frac{453,36^2}{9 \;\; 62,14^2}} = 2,43;$$

$$t_a = \frac{|a|}{\hat{\sigma}_a} = \frac{581,09}{2,43} = 238,93.$$

Определим табличное значение t при уровне значимости α = 0,05. Количество степеней свободы 9 – 1 – 1 = 7. Будем использовать табл. 2 прилож. Получаем $t_{7;\ 0,05} = 2,36$.

Для параметра a

$$t_{7;\ 0,05} = 2,36 > t_{набл} = 2,06.$$

Следовательно, при уровне значимости α = 0,05 параметр a незначим.

Для параметра b

$$t_{7;\ 0,05} = 2,36 < t_{набл} = 238,93.$$

Следовательно, при уровне значимости α = 0,05 параметр b значим.

Определим значимость коэффициента корреляции:

$$t_{набл} = r_{xy}\sqrt{\frac{n-2}{1-r_{xy}^2}} = 0,87\sqrt{\frac{9-2}{1-0,87^2}} = 4,6.$$

Найдем табличное значение t при уровне значимости α = 0,05. Количество степеней свободы 9 – 1 – 1 = 7. Используем табл. 2 прилож.

При уровне значимости $\alpha = 0,05$ $t_{7;\ 0,05} = 2,36 < t_{набл} = 4,6$, следовательно, коэффициент корреляции значим.

Рассчитаем прогноз заработной платы y при $x = 1,07 \;\; \overline{x}$:

$$x_{пр} = 1,07 \;\; 503,1 = 538,32; \quad y_{пр} = 5,59 \;\; 538,32 - 581,09 = 2428,21.$$

Найдем доверительный интервал прогноза.

Предельная ошибка

$$e = t_{табл} \;\; \hat{\sigma} \sqrt{1 + \frac{1}{n} + \frac{\left(\overline{x} - x_{пр}\right)^2}{\left(x - \overline{x}\right)^2}} = 2,36 \;\; 453,36 \;\; \sqrt{1 + \frac{1}{9} + \frac{\left(503,10 - 538,32\right)^2}{34749,45}} = 1145,78.$$

Доверительный интервал прогноза тогда будет

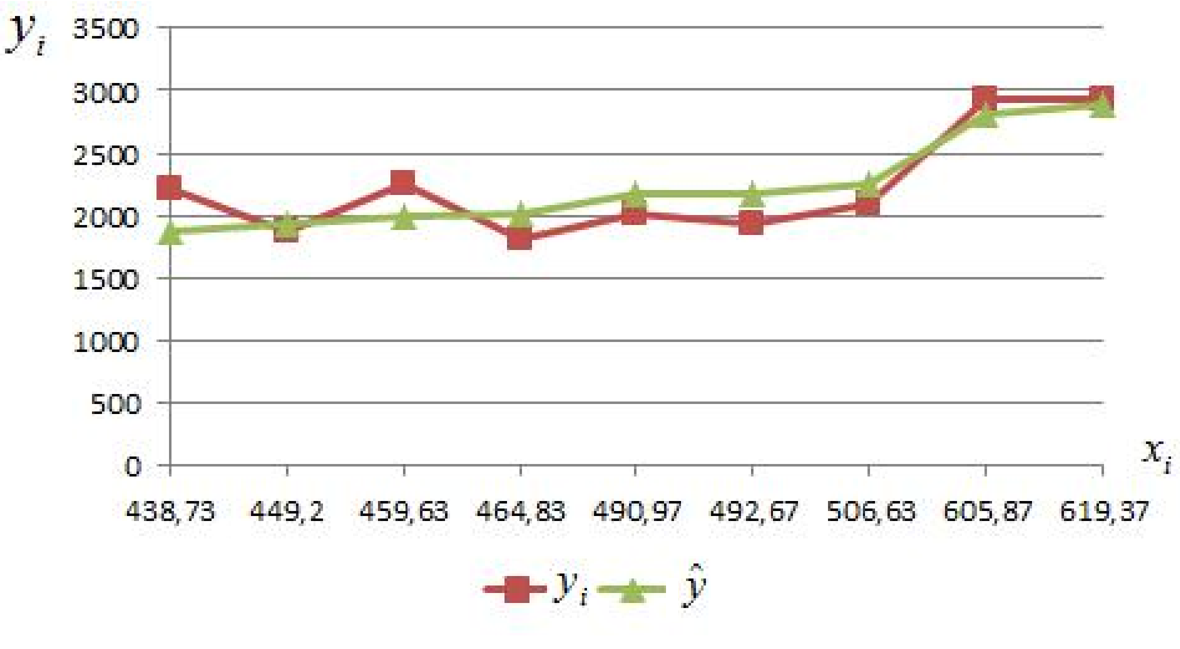
$$\left(y_{пр} - e;\ y_{пр} + e\right) = \left(2428,21 - 1145,78 \;\; 2428,21 - 1145,78\right) = \left(1282,43;\ 3573,99\right).$$

С вероятностью 95 % можно гарантировать, что прогнозное значение будет находиться в полученном интервале.

Упорядочим исходные данные по возрастанию переменной x:

x	y	$\hat{y}$
438,73	2212,77	1871,43
449,2	1881,23	1929,94
459,63	2254,45	1988,26
464,83	1820,36	2017,33
490,97	2015,14	2163,41
492,67	1929,86	2172,92
506,63	2097,77	2250,99
605,87	2935,55	2805,7
619,37	2934,91	2881,17

На одном графике изобразим исходные данные и построим прямую уравнения регрессии:



Отличие регрессионной линии от прямой объясняется неравномерностью значений по оси x.

Задания для самостоятельного выполнения

Задание 1. В таблице для каждого варианта приведены данные выборки об урожайности зерновых культур x (ц/га) и ценах их реализации y (усл. ден. ед/ц) по 10 районам. Требуется:

1) построить уравнение парной регрессии зависимости результативного признака y от фактора x, пояснить экономический смысл его параметров, изобразить графически полученную зависимость и поле корреляции;

2) оценить тесноту линейной связи цены реализации и урожайности зерновых культур с помощью коэффициентов корреляции и детерминации, значимость коэффициента корреляции через t-критерий Стьюдента на уровне $a = 0,05$;

3) определить статистическую надежность и качество полученного уравнения регрессии в целом с помощью F-критерия Фишера и средней ошибки аппроксимации;

4) рассчитать средний коэффициент эластичности и на его основе дать оценку силы связи результативного признака y (цены реализации) с фактором x_j (урожайностью зерновых культур);

5) рассчитать прогнозное значение результативного признака y (цены реализации), если прогнозное значение фактора x_j (урожайности зерновых культур) уменьшится на 10 % от его среднего уровня. Определить доверительный интервал прогноза на уровне $a = 0,05$.

Вариант / № региона	1		2		3		4		5	
	x	y	x	y	x	y	x	y	x	y
1	253	12,3	239	5,3	204	5,9	256	7,7	206	13,6
2	260	6,9	333	3,3	264	8,5	235	8,1	207	5,1
3	191	7,3	230	8,9	345	5,2	298	8,3	242	7,6
4	229	5,7	253	7,5	243	5,6	254	4,6	216	12,4
5	332	4,2	319	3,7	219	10,9	263	11,7	171	10,4
6	333	6,3	300	8,5	267	5,9	262	7,6	147	10,8
7	325	4,7	285	5,6	270	7,1	286	7,0	237	6,1
8	220	9,4	225	9,0	270	8,5	252	4,0	262	6,1
9	226	7,5	176	5,7	330	3,9	348	1,8	238	8,2
10	224	7,7	213	12,4	365	7,9	200	10,6	286	3,2
11	237	6,6	125	10,8	200	12,5	280	5,3	216	7,1
12	154	12,7	215	5,9	189	11,6	259	7,4	211	6,61

Вариант / № региона	6		7		8		9		10	
	x	y	x	y	x	y	x	y	x	y
1	268	6,0	291	7,1	160	8,5	205	7,9	149	10,2
2	333	5,3	266	7,3	303	4,3	231	6,2	337	5,0
3	229	5,9	245	7,9	359	3,7	247	6,4	201	6,4
4	278	4,5	199	3,6	298	8,4	233	10,8	185	6,8
5	353	2,5	224	7,9	266	1,8	303	7,7	221	7,7
6	295	7,3	236	5,3	280	6,9	164	3,9	244	7,1
7	109	6,9	289	2,7	264	8,4	264	2,8	296	5,9
8	128	8,8	263	11,7	171	11,3	195	7,4	248	65
9	202	13,3	234	9,9	253	3,8	213	8,4	196	4,8
10	198	12,5	308	4,0	235	6,1	191	11,8	241	8,1
11	164	10,0	285	76	221	8,5	227	8,9	293	4 8
12	233	9,6	204	12,0	223	7,7	250	2,4	304	5,8

Задание 2. На основе проведенного опроса 8 групп семей известны данные о связи расходов населения на продукты питания с уровнем доходов семьи:

Расходы на продукты питания y (тыс. руб.)	6	7	8	10	15	20	30
Доходы семьи x (тыс. руб.)	17	20	25	30	40	45	50

Следует:

1) построить линейную регрессионную модель связи переменных, где y – зависимая переменная, а x – независимая, используя МНК;

2) рассчитать линейный коэффициент парной корреляции, коэффициент детерминации и среднюю ошибку аппроксимации;

3) оценить статистическую значимость параметров регрессии и корреляции на уровне значимости $\alpha = 0{,}05$ с помощью F-критерия Фишера и t-критерия Стьюдента;

4) дать прогноз заработной платы y при прогнозном значении среднедушевого прожиточного минимума x, составляющем 110 % от среднего уровня;

5) оценить точность прогноза с уровнем значимости $\alpha = 0{,}05$;

6) найти с надежностью 0,95 интервальные оценки параметров уравнения регрессии α и β;

7) на одном графике (графике подбора) построить исходные данные и теоретическую прямую. Сделать выводы.

Задание 3. Изучается зависимость качества почв по плодородности *У* от факторного признака – запасов влаги в почве (бонитировочного балла) *X*.

Номер наблюдения	Запасы влаги в почве *X* (мм)	Плодородность *У* (%)
1	144	75
2	110	54
3	110	61
4	177	64
5	186	72
6	112	69
7	148	79
8	151	73
9	110	60
10	151	72
11	131	54
12	113	77
13	110	57
14	127	72
15	136	72
16	136	67
17	144	72
18	100	55
19	148	68
20	129	68

Требуется:

1. Рассчитать параметры парной линейной регрессии.
2. Оценить тесноту связи с помощью показателей корреляции и детерминации.
3. Установить с помощью средней ошибки аппроксимации качество уравнений.
4. Оценить статистическую значимость уравнения регрессии и его параметров с помощью критериев Фишера и Стьюдента.
5. Рассчитать прогнозное значение результата, если прогнозное значение фактора увеличится на 10 % от его среднего уровня ($x_{прогноз} = 1{,}1\ \bar{x}$). Определить доверительный интервал прогноза для уровня значимости *a* = 0,05.

ГЛАВА 2
НЕЛИНЕЙНАЯ РЕГРЕССИЯ

Примеры решения задач и выполнения заданий по теме «Исследование зависимостей между явлениями и их признаками. Нелинейная однофакторная регрессия и корреляция. Эластичность нелинейных функций»

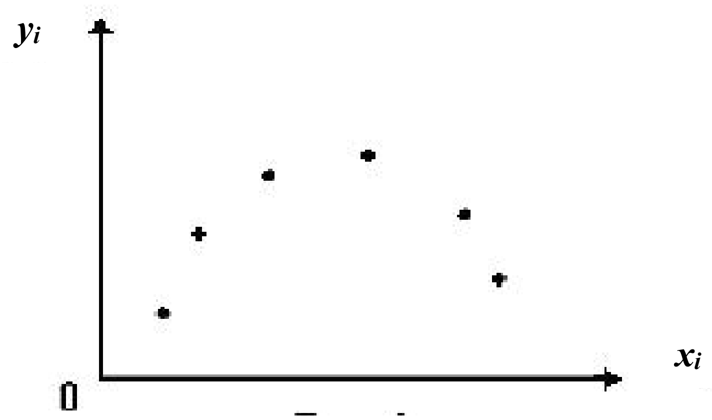
Задание 2.1.1. Данные об изменениях прибыли в зависимости от объема производства приведены ниже:

Объем производства (усл. ед.)	2	4	6	8	10	12
Прибыль (усл. ед.)	140	168	200	250	270	350
Объем производства (усл. ед.)	14	16	18	20	22	24
Прибыль (усл. ед.)	280	250	250	164	125	100

Найти регрессионную модель изменения прибыли. Определить величину прибыли при объеме производства в 13 усл. ед.

Решение

Изобразим схематично вид исходных данных:



На графике обозначены: y_i – прибыль, x_i – объем производства. По графику можно сделать предположение, что статистический материал аппроксимируется параболой типа $y = b_2x^2 + b_1x + b_0$. Параметры $b_0,\ b_1,\ b_2$ параболической функции находятся из решения системы линейных уравнений следующего вида:

$$b_2\sum_{i=1}^{n} x_i^4 + b_1\sum_{i=1}^{n} x_i^3 + b_0\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i^2;$$

$$b_2\sum_{i=1}^{n} x_i^3 + b_1\sum_{i=1}^{n} x_i^2 + b_0\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i x_i;$$

$$b_2\sum_{i=1}^{n} x_i^2 + b_1\sum_{i=1}^{n} x_i + nb_0 = \sum_{i=1}^{n} y_i.$$

Для расчета необходимых сумм в системе составим вспомогательную таблицу:

i	x_i	y_i	x_i^2	x_i^3	x_i^4	$y_i x_i$	$y_i x_i^2$
1	2	140	4	8	16	280	560
2	4	168	16	64	256	672	2688
3	6	200	36	216	1296	1200	7200
4	8	250	64	512	4096	2000	16000
5	10	270	100	1000	10000	2700	27000
6	12	350	144	1728	20736	4200	50400
7	14	280	196	2744	38416	3920	54880
8	16	250	256	4096	65536	4000	64000
9	18	250	324	5832	104976	4500	81000
10	20	164	400	8000	160000	3280	65600
11	22	125	484	10648	212960	2750	60500
12	24	100	576	13824	331776	2400	57600
	156	2547	2600	48672	950064	31902	487428

Система для рассматриваемого случая примет вид

$$950064\ b_2 + 48672\ b_1 + 2600\ b_0 = 487428;$$

$$48672\ b_2 + 2600\ b_1 + 156\ b_0 = 31902;$$

$$2600\ b_2 + 156\ b_1 + 12\ b_0 = 2547.$$

Решая ее, например, методом Гаусса, получим: $b_2 = -1{,}98$, $b_1 = 49{,}38$, $b_0 = -13{,}38$. Уравнение регрессии:

$$y = -19{,}8x^2 + 49{,}38x - 13{,}38.$$

Значение прибыли при объеме составляет 13 усл. ед., т.е. значение прогноза y при $x = 13$ $y(13) = -1{,}98\ 13^2 + 49{,}38\ 13 - 13{,}38 = 293{,}94$ усл. ед.

Задание 2.1.2. Используя условия предыдущего примера, найти среднюю абсолютную ошибку предсказания зависимой переменной (прибыли), а также оценить значимость уравнения регрессии.

Решение

Составим вспомогательную таблицу расчетов:

Объем производства	y_i	$y(x_i)$	$\left\lvert\frac{y(x_i)-y_i}{y_i}\right\rvert$	$(y_i - y(x))^2$	$\left(y(x_i)-\overline{y}\right)$
2	140	77,46	0,447	6000	18168
4	168	152,46	0,093	241,49	3574,84
6	200	211,62	0,058	135,02	0,4
8	250	254,94	0,02	24,4	1722,44
10	270	282,42	0,046	154,26	4923,83
12	350	294,06	0,16	3129,28	6692,88
14	280	239,86	0,035	97,22	6023,31
16	250	269,82	0,071	282,91	3314,3
18	250	233,94	0,064	257,92	470,46
20	164	182,22	0,111	331,97	901,8
22	125	114,66	0,088	106,92	4725
24	100	31,26	0,687	4725	32757,38
	2547	–	1,88	15486,39	88173,45

Найдем $\overline{y} = \frac{1}{12} \quad y_i = 212{,}25$.

Вычислим среднюю ошибку аппроксимации:

$$A = \frac{1}{12} \sum_{i=1}^{12} \left\lvert \frac{y_i - y(x_i)}{y_i} \right\rvert \ 100\ \% = \frac{1}{12}\ 1{,}88\ 100\ \% = 15{,}7\%.$$

Согласно классификации точности результатов регрессии можно сделать вывод о приближенной точности аппроксимации данного статистического материала.

Для определения значимости уравнения регрессии воспользуемся формулой

$$F_{факт} = \frac{(\text{€}_x - \overline{y})^2 / m}{(y - \overline{y})^2 / (n - m - 1)} = \frac{88173{,}45\ (12-3)}{15486{,}39\ (3-1)} = 25{,}62.$$

Из табл. 1 приложения для числа степеней свободы $k_1 = 3 - 1 = 2$, $k_2 = 12 - 3 = 9$ и $\alpha = 0{,}05$ находим $F_{0,05;\ 2;\ 9} = 19{,}4$. Поскольку в данном случае $F > F_{0,05;\ 2;\ 9}$, то согласно критерию Фишера уравнение регрессии значимо на уровне $a = 0{,}05$.

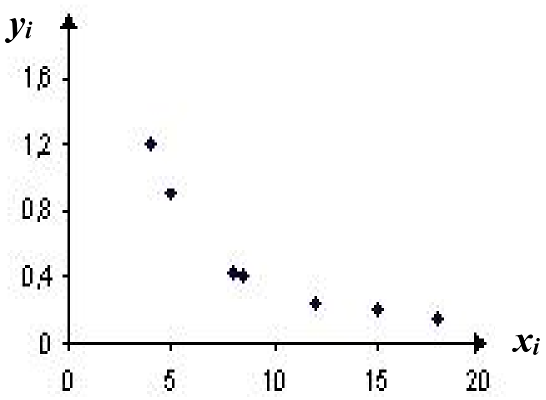
Задание 2.1.3. Имеется выборка данных зависимости спроса от цены следующего вида:

Цена x_i (ден. ед.)	4	5	8	8,5	12	15	18
Спрос y_i (ден. ед.)	1,2	0,9	0,42	0,4	0,24	0,2	0,14

Построить уравнение регрессии, оценить его точность и значимость на уровне $a = 0{,}05$.

Решение

Представим исходные данные задачи графически в виде поля корреляции:



Анализ графика позволяет сделать вывод об обратной зависимости вида $y = ax^k$, где $k < 0$. Путем замены переменных и коэффициентов в данной зависимости последнюю можно промоделировать линейной парной регрессией $z = b_0 + b_1 u$, где $z = \lg y,\quad b_0 = \lg a,\quad b_1 = k,\quad u = \lg x$. Составим таблицу изменения новых переменных *u* и *z*:

u_i	0,587	0,684	0,888	0,905	1,075	1,131	1,2
z_i	0,081	–0,043	–0,366	–0,387	–0,591	–0,684	–0,846

По ней построим линию регрессии. Имеем:

$$\bar{u} = \frac{1}{7}(0{,}587 + 0{,}684 + 0{,}888 + 0{,}905 + 1{,}075 + 1{,}131 + 1{,}2) = 0{,}923;$$

$$z = \frac{1}{7}(0{,}081 - 0{,}043 - 0{,}366 - 0{,}387 - 0{,}591 - 0{,}684 - 0{,}846) = -0{,}41;$$

$$\overline{uz} = \frac{1}{7}(0{,}587\ \ 0{,}081 - 0{,}684\ \ 0{,}043 - \ldots - 1{,}2\ \ 0{,}846) = -0{,}44;$$

$$S_u^2 = \frac{1}{7}(0{,}587^2 + \ldots + 1{,}2^2) - 0{,}92^2 = 0{,}899 - 0{,}92^2 = 0{,}053;$$

$$b_1 = \frac{\overline{uz} - \bar{u}\ \ \bar{z}}{S_u^2} = \frac{-0{,}44 - 0{,}92\ \ 0{,}41}{0{,}053} = -1{,}18;\ b_0 = \bar{z} - b_1\ \ \bar{u} = -0{,}41 + 1{,}18\ \ 0{,}92 = 0{,}68.$$

Уравнение регрессии:

$$z = 0{,}68 - 1{,}18\ u.$$

Отсюда

$$y = 10^{0{,}68}\ x^{-1{,}18},\ \text{или}\ y = 4{,}83\ x^{-1{,}18}.$$

Исходное и преобразованное уравнения при указанной выше замене эквивалентны. Поэтому если значимо преобразованное уравнение, то значимо и исходное уравнение, и если между переменными Z и U имеет место тесная корреляционная связь, то X и Y тоже будут сильно коррелированы. Точности исходной и преобразованной регрессионных моделей совпадают.

Задание 2.1.4. Зависимость расходов на питание от дохода выражается формулой $y = 6{,}13 \cdot x^{0{,}374}$. Определить эластичность спроса на продукты питания по доходу семьи. Что означает полученный результат?

Решение

В случае когда регрессионная зависимость выражается степенной функцией, коэффициент эластичности равен показателю степени b независимо от значений независимой переменной.

Определим коэффициент эластичности спроса для указанной регрессии:

$$E = b = 0{,}374.$$

Можно сделать вывод, что при росте доходов на 1 % расходы на питание возрастают на 0,374 %.

Задание 2.1.5. Задано уравнение двухфакторной линейной регрессии: $y = 20 + 14x_1 + 21x_2$. Средние значения признаков: $\bar{x}_1 = 3$; $\bar{x}_2 = 1$; $\bar{y} = 4$.

Требуется проанализировать влияние каждого из факторов на результат на основе частных коэффициентов эластичности.

Решение

Частные коэффициенты эластичности при линейной зависимости рассчитываются по формуле

$$\bar{Э}_j = b_j \frac{\bar{x}_j}{\bar{y}}.$$

Тогда

$$\bar{Э}_1 = 14\ \frac{3}{4} = 10{,}5;\ \bar{Э}_2 = 21\ \frac{1}{4} = 5{,}25.$$

Частный коэффициент эластичности показывает, на сколько процентов в среднем изменяется результат y с увеличением фактора x_j на 1 % от своего среднего уровня при фиксированном положении других факторов модели.

При увеличении фактора x_1 на 1 % от своего среднего уровня ($\overline{x_2} = 3$) результат y возрастет в среднем на 10,5 % при неизменном уровне других факторов модели.

При увеличении фактора x_2 на 1 % от своего среднего уровня ($\overline{x_2} = 1$) результат y возрастет в среднем на 5,25 % при неизменном уровне других факторов модели.

Задание 2.1.6. Какое из следующих уравнений нелинейно по оцениваемым параметрам:

1) $y_x = a + bx + \varepsilon$;

2) $y_x = a + b\ln x + \varepsilon$;

3) $y_x = ax^b\varepsilon$?

Решение

Алгоритм решения выглядит так:

1) $y_x = a + bx + \varepsilon.$ Данная модель является линейной моделью парной регрессии;

2) $y_x = a + b\ln x + \varepsilon.$ Эта модель является нелинейной относительно факторов, включенных в модель, но линейной относительно оцениваемых параметров;

3) $y_x = ax^b\varepsilon.$ Данная модель является степенной моделью, нелинейной по оцениваемым параметрам.

Вывод: только третье уравнение нелинейно по оцениваемым параметрам.

Задание 2.1.7. В группу входят 10 заводов, производящих однородную продукцию. Для этой группы получено уравнение регрессии себестоимости единицы продукции y (тыс. руб.) от уровня технической оснащенности x (тыс. руб.):

$$y = 20 + 700/x.$$

Определить коэффициент эластичности, предполагая, что стоимость активных производственных фондов составляет 200 тыс. руб.

Решение

Коэффициент эластичности определяется по формуле

$$Э = f(x)\ \frac{x}{y}.$$

Вычислим коэффициент эластичности для $x = 200$:

$$Э = f(x)\ \frac{x}{y} = -\frac{b}{x^2}\ \frac{x}{a + \frac{b}{x}} = -\frac{b}{ax + b} = -\frac{700}{20\ \ 200 + 700} = -0{,}149\ \%.$$

Таким образом, изменение технической оснащенности на 1 % приведет к снижению себестоимости единицы продукции на 0,149 %.

Спрос неэластичен, т.е. объем спроса на продукцию останется практически неизменным при любой цене.

Задание 2.1.8. Определить, какие из функций, представленных ниже, линейны по переменным; линейны по параметрам; нелинейны ни по переменным, ни по параметрам:

1) $y = a + bx^3 + \varepsilon$;
2) $y = a + b\ \ \ln x + \varepsilon$;
3) $\ln y = a + b\ \ \ln x + \varepsilon$;
4) $\ln y = a + b\ \ \ln x + \varepsilon$;
5) $y^a = b + cx^2 + \varepsilon$;
6) $y = 1 + a\left(1 - x^b\right) + \varepsilon$;
7) $y = a + b\dfrac{x}{10} + \varepsilon$.

Решение

Линейна по переменным (x, y) функция 7.

Линейны по оцениваемым параметрам (a, b, c) функции 1–4 и 7.

Нелинейны ни по переменным, ни по параметрам функции 5 и 6.

Задание 2.1.9. Определить коэффициент эластичности нелинейных функций:

1) $y = 15 \cdot 2^{x}$; $x = 2{,}6$; $\ln 2 = 0{,}69$;
2) $y = 0{,}6 + 0{,}5 \cdot \ln x$; $x = 1{,}5$; $\ln 1{,}5 = 0{,}405$;
3) $y = 12 + x^2$; $x = 10$;
4) $y = 200 + 700/x$; $x = 10$;
5) $y = 25 \cdot 2^x$; $x = 5$; $\ln 2 = 0{,}69$;
6) $y = 25 + 70/x$; $x = 5$;
7) $y = 1{,}6 + 2{,}2x$; $x = 23{,}5$.

Найти, на сколько процентов в среднем изменится y, если x в среднем увеличится на 1 %. Сделать выводы.

Решение

Последовательность расчетов и выводов:

1. $Э = \bar{x}\ \ \ln b = 2,6\ \ 0,69 = 1,794.$

При увеличении x на 1 % y возрастает на 1,794 %.

2. $Э = \dfrac{b}{a + b\ln x} = \dfrac{0,5}{0,6 + 0,5\ \ 0,405} = 0,623$ %.

При увеличении x на 1 % y возрастает на 0,623 %.

3. $Э = \dfrac{(b + 2c\ \ \bar{x})\ \ \bar{x}}{a + b\ \ \bar{x} + c\ \ \bar{x}^2} = \dfrac{(0 + 2\ \ 1\ \ 10)\ \ 10}{12 + \ 10 + 1\ \ 10^2} = 1,786$ %.

При увеличении x на 1 % y возрастает на 1,786 %.

4. $Э = -\dfrac{b}{a\bar{x} + b} = -\dfrac{1}{12\ \ 10 + 1} = -0,0083$ %.

При увеличении x на 1 % y снижается на 0,0083 %.

5. $Э = \bar{x}\ \ \ln b = 5\ \ 0,69 = 3,45$ %.

При увеличении x на 1 % y возрастает на 3,45 %.

6. $Э = -\dfrac{b}{a\bar{x} + b} = -\dfrac{70}{25\ \ 5 + 70} = -0,359$ %.

При увеличении x на 1 % y снижается на 0,359 %.

7. $Э = \dfrac{b\bar{x}}{a + b\ \ \bar{x}} = \dfrac{2,2\ \ 23,5}{1,6 + 2,2\ \ 23,5} = 0,97$ %.

При увеличении x на 1 % y возрастает на 0,97 %.

Задание 2.1.10. Доля остаточной дисперсии в общей составила 0,19. Определить индекс корреляции и детерминации.

Решение

Известно, что доля остаточной регрессии в общей составила 0,19. Значит, $\frac{\Sigma(y-\hat{y}_x)^2}{\Sigma(y-\bar{y}_x)^2} = 0,19$.

Найдем коэффициент детерминации: $r_{xy}^2 = 1 - \frac{\Sigma(y-\hat{y}_x)^2}{\Sigma(y-\bar{y}_x)^2} = 1 - 0,19 = 0,81$.

Вычислим коэффициент корреляции: $r_{xy} = \sqrt{0,81} = 0,9$.

Задание 2.1.11. Определить влияние численности работников x (чел.) на объем производства y (тыс. ед.) при x = 10, используя коэффициент эластичности, если $y = 30 - 0,4 \cdot x + 0,04 \cdot x^2$.

Решение

Рассчитаем эластичность:

$$Э = \frac{(b + 2c \cdot \bar{x}) \cdot \bar{x}}{a + b \cdot \bar{x} + c \cdot \bar{x}^2} = \frac{(-0,4 + 2 \cdot 0,04 \cdot 10) \cdot 10}{30 - 0,4 \cdot 10 + 0,04 \cdot 10^2} = 0,133\ \%.$$

Это означает, что при увеличении численности работников на 1 % объем производства возрастает на 0,133 %.

Коэффициент эластичности Э < 1. Следовательно, его влияние на результативный признак Y незначительно.

Задание 2.1.12. Для 3 видов продукции – А, В, С – модели зависимости постоянных расходов от объема продукции определяются выражениями $y_A = 600$; $y_B = 80 + 0,7 \cdot x$; $y_C = 40 \cdot x^{0,5}$. Определить коэффициенты эластичности по каждому виду продукции и пояснить их смысл. Сравнить при $x = 1000$ эластичность затрат для продукции В и С. При каком объеме продукции x $Э_B = Э_C$?

Решение

Определим коэффициенты эластичности по формуле $Э = f'(x) \cdot \frac{x}{y}$:

$$Э_A = 0;\ Э_B = \frac{bx}{a + b \cdot x} = \frac{0,7 \cdot x}{80 + 0,7 \cdot x};\ Э_C = 20 \cdot \frac{1}{\sqrt{x}} \cdot \frac{1}{40\sqrt{x}} = \frac{20 \cdot x}{40 \cdot x} = 0,5.$$

Коэффициент эластичности показывает, на сколько процентов уменьшится или увеличится y при изменении x на 1 %.

При $x = 1000$ $Э_B = \frac{0,7 \cdot 1000}{80 + 0,7 \cdot 1000} = 0,89,\ Э_C = 0,5.$

$Э_B = Э_C \Rightarrow \frac{0,7 \cdot x}{80 + 0,7 \cdot x} = 0,5 \Rightarrow 0,7x = 80.$

$x = 114,29$ – объем выпуска, при котором $Э_B = Э_C$.

Задания для самостоятельного выполнения

Задание 1. Определить коэффициент эластичности нелинейных функций:

1) $y = 12 \cdot 3^x$; $\bar{x} = 3$; $\ln 3 = 1,1$;
2) $y = 1,8 + 0,6 \cdot \ln x$; $\bar{x} = 2$; $\ln 2 = 0,69$;
3) $y = 10 + x^2$; $\bar{x} = 5$;
4) $y = 100 + 600/x$; $\bar{x} = 50$;
5) $y = 15 \cdot 5^x$; $\bar{x} = 5$; $\ln 5 = 1,61$;
6) $y = 15 + 60/x$; $\bar{x} = 15$;
7) $y = 1,2 + 1,2x$; $\bar{x} = 2,4$.

Рассчитать, на сколько процентов в среднем изменится y, если x в среднем увеличится на 1 %. Сделать выводы.

Задание 2. Доля остаточной дисперсии в общей составила 0,36. Определить индекс корреляции и детерминации.

Задание 3. В группу входят 20 заводов, производящих однородную продукцию. Для этой группы получено уравнение регрессии себестоимости единицы продукции Y (усл. ед.) от уровня технической оснащенности X (усл. ед.):

$$Y = 10 + 500/X.$$

Определить коэффициент эластичности, предполагая, что средняя стоимость активных производственных фондов составляет 100 усл. ед.

Задание 4. Определить влияние численности работников X (чел.) на объем производства Y (тыс. усл. ед.) при $\bar{x} = 5$, используя коэффициент эластичности, если $y = 20 - 0{,}5\ x + 0{,}01\ x^2$.

Задание 5. Зависимость расходов на питание от дохода $y = 5{,}26\ x^{0{,}256}$. Определить эластичность спроса на продукты питания по доходу семьи. Что означает полученный результат?

Задание 6. Определить, какие из функций линейны по переменным; линейны по параметрам; нелинейны ни по переменным, ни по параметрам:

1) $y = a + bx^4 + \varepsilon$;
2) $y = a + b \ln x^2 + \varepsilon$;
3) $\ln y = a + b\ln x_5 + \varepsilon$;
4) $\ln y^2 = a + b\ln x^3 + \varepsilon$;
5) $y^a = b + cx^3 + \varepsilon$;
6) $y = 1 + a(10 - x^b) + \varepsilon$;
7) $y = a + b\, x^2/20 + \varepsilon$.

Задание 7. Записать для функции $y_x = \dfrac{a}{1 + b\ e^{-cx}}$ средний коэффициент эластичности.

Задание 8. Есть данные по 9 однотипным сельскохозяйственным предприятиям, специализирующимся на выпуске молочной продукции в Тверской области. Годовой объем выпуска продукции y (усл. ед.) зависит от фонда оплаты труда x (усл. ед.) следующим образом:

x	2	2,2	2,5	3,1	3,3	3,5	3,8	3,9	4
y	1,5	1,84	2,425	3,865	4,425	5,025	6	6,345	6,7

Построить нелинейное уравнение парной регрессии y от x вида $y_x = a + b\ x + c\ x^2$. Оценить качество построенной модели: рассчитать среднюю ошибку аппроксимации; определить значимость уравнения регрессии и его параметров; дать прогноз для x = 5; рассчитать доверительный интервал прогноза (α = 0,05).

Задание 9. По 7 сельскохозяйственным предприятиям региона получены данные о зависимости объема выпуска продукции Y (млн руб.) от объема капиталовложений X (млн руб.):

Y	65	58	54	47	51	48	34
X	66	69	85	72	83	91	100

Требуется:

1) для определения вида зависимости Y от X построить следующие модели: степенную, показательную, гиперболическую. Изобразить регрессионные модели на графиках;

2) оценить каждую модель, определив индекс корреляции, среднюю ошибку аппроксимации, коэффициент детерминации, F-критерий Фишера;

3) выбрать лучшую модель;

4) рассчитать прогнозные значения результативного признака по лучшей модели, если объем капиталовложений составит 88 млн руб.

Задание 10. Известны динамика среднедушевого дохода населения в регионе по месяцам и средние расходы на парикмахерские услуги.

Номер месяца	1	2	3	4	5	6	7	8	9	10
Среднедушевой доход (тыс. руб.)	34,33	34,36	34,49	35,04	36,93	38,5	43,12	43,41	43,54	44,25
Расходы на бытовые услуги (тыс. руб.)	1,5	1,5	1,4	1,3	1,65	1,9	2,15	1,98	2,2	2,25

Для исследования расходов были построены модели:

1) $Y = -1{,}23 + 0{,}14X$;

2) $Y = 1{,}27 \cdot X^{0,27}$;

3) $Y = 0{,}47 \cdot \ln(X) + 1{,}22$.

Требуется:

1. Оценить точность построенных моделей с помощью коэффициента детерминации.

2. Построить на графике результаты наблюдений и результаты, полученные на построенных моделях.

ГЛАВА 3
МНОЖЕСТВЕННАЯ РЕГРЕССИЯ И КОРРЕЛЯЦИЯ

Примеры решения задач и выполнения заданий по теме «Отбор факторов для включения в модель множественной регрессии. Оценка точности, надежности, анализ влияния фактора на результат в множественной регрессии. Фиктивные переменные в множественной регрессии»

Задание 3.1.1. Матрица парных коэффициентов корреляции имеет следующий вид:

	y	x_1	x_2	x_3
y	1	0,8	0,7	0,6
x_1	0,8	1	0,8	0,5
x_2	0,7	0,8	1	0,2
x_3	0,6	0,5	0,2	1

Отобрать факторы для включения в модель множественной регрессии.

Решение

Матрица симметрична относительно главной диагонали, поэтому анализируется верхняя или нижняя часть.

Коэффициент корреляции $r_{x_1x_2} = 0,8 > 0,7$, т.е. факторы x_1 и x_2 дублируют друг друга. В уравнение регрессии нужно включить один из них.

При дублировании факторов один из них рекомендуется исключить из регрессии. Руководствуются двумя правилами:

1) предпочтение отдается фактору, имеющему наименьшую тесноту связи с другими факторами;

2) предпочтение отдается фактору, более тесно связанному с результатом.

Согласно первому правилу, так как $r_{x_1x_3} = 0,5 > r_{x_2x_3} = 0,2$, то в уравнение регрессии следует включить фактор x_2, а не x_1 (меньше связь с другим фактором, а именно x_3). По 2-му правилу, так как $r_{yx_2} = 0,7 < r_{yx_1} = 0,8$, то лучше признак x_1 (больше связь с результатом y). Таким образом, правило 2 противоречит правилу 1. Руководствуемся первым правилом, т.е. включаем в уравнение регрессии фактор x_2, а не x_1

(из дублирующих). Фактор x_3 не является дублирующим ($r_{x_1x_3} = 0,5 < 0,7$; $r_{x_2x_3} = 0,2 < 0,7$).

Задание 3.1.2. Данные (условные) о функции издержек производства y (ден. ед.), зависящей от основных производственных фондов x_1 (ден. ед.) и численности x_2 занятых в производстве человек, приведены ниже:

i	1	2	3	4	5	6	7	8	9	10
x_1	80	100	120	90	90	80	90	90	100	90
x_2	50	80	80	50	70	80	60	40	50	70
y	5	10	10	7	5	6	6	5	6	8

Предполагая, что между переменными Y, X_1 и X_2 существует линейная корреляционная зависимость, найти ее аналитическое выражение (вид уравнения регрессии $y = b_0 + b_1x_1 + b_2x_2$).

Решение

Обозначим: $Y = \begin{matrix} 5 \\ 10 \\ \dots \\ 8 \end{matrix}$, $\quad X = \begin{matrix} 1 & 80 & 50 \\ 1 & 100 & 80 \\ \dots & \dots & \dots \\ 1 & 90 & 70 \end{matrix}$.

Найдем $X^T \; X = \begin{matrix} 1 & 1 & \dots & 1 \\ 80 & 100 & \dots & 90 \\ 50 & 80 & \dots & 70 \end{matrix} \; \begin{matrix} 1 & 80 & 50 \\ 1 & 100 & 80 \\ \dots & \dots & \dots \\ 1 & 90 & 70 \end{matrix}$.

При умножении первой строки на первый столбец получается число 10. При умножении второй строки на первый столбец складываются все числа второй строки таблицы, а при умножении третьей строки на первый столбец – все числа третьей строки таблицы. При умножении первой строки на второй и третий столбцы складываются числа второй и третьей строки соответственно.

При умножении второй строки на второй столбец и третьей строки на третий столбец складываются соответственно квадраты чисел второй и третьей строк таблицы. Умножение второй строки на третий столбец и третьей строки на второй столбец сводится к сложению соответствующих произведений второй и третьей строк таблицы. В результате получим

$$X^T \; X = \begin{matrix} 10 & 9300 & 630 \\ 930 & 87700 & 59100 \\ 630 & 59100 & 41700 \end{matrix}.$$

Вычислим $X^T\ Y = \begin{matrix} 1 & 1...1 \\ 80 & 100...90 \\ 50 & 80...70 \end{matrix} \begin{matrix} 5 \\ 10 \\ ... \\ 8 \end{matrix}$. Здесь при умножении первой строки на Y получается сумма чисел третьей строки таблицы, а при умножении второй и третьей строки на Y соответственно произведение первой строки на третью и второй на третью. Итак, $X^T\ Y = \begin{matrix} 68 \\ 6470 \\ 4450 \end{matrix}$. Согласно приведенному выше алгоритму найдем $A^{-1} = \left(X^T\ X\right)^{-1}$. Для этого сначала вычислим определитель матрицы $A = X^T\ X$. Имеем

$$|A| = \left|X^T\ X\right| = \begin{vmatrix} 10 & 930 & 630 \\ 930 & 87700 & 59100 \\ 630 & 59100 & 41700 \end{vmatrix} = 10\ \begin{vmatrix} 87700 & 59100 \\ 59100 & 41700 \end{vmatrix} - 930\ \begin{vmatrix} 930 & 59100 \\ 630 & 41700 \end{vmatrix} +$$

$$+\ 630\ \begin{vmatrix} 930 & 87700 \\ 630 & 59100 \end{vmatrix} = 10\ \left(87700\ 41700 - 59100^2\right) - 930\ \left(930\ 41700 - 630\ 59100\right) +$$

$$+\ 630\ \left(930\ 59100 - 630\ 87700\right) = 21720000 = 2172\ 10^4.$$

Найдем $\left(X^T\ X\right)^T = \begin{matrix} 10 & 930 & 630 \\ 930 & 87700 & 59100 \\ 630 & 59100 & 41700 \end{matrix}$ и алгебраические дополнения с учетом того, что для данной матрицы $A_{12} = A_{21}, A_{13} = A_{31}, A_{23} = A_{32}$:

$$A_{11} = \begin{vmatrix} 87700 & 59100 \\ 59100 & 41700 \end{vmatrix}\ (-1)^{1+1} = 87700\ 41700 - 59100^2 = 16428\ 10^4;$$

$$A_{12} = \begin{vmatrix} 930 & 59100 \\ 630 & 41700 \end{vmatrix}\ (-1)^{1+2} = -(930\ 41700 - 630\ 59100) = -1548\ 10^3;$$

$$A_{13} = \begin{vmatrix} 930 & 87700 \\ 630 & 59100 \end{vmatrix}\ (-1)^{1+3} = 930\ 59100 - 87700\ 630 = -288\ 10^3;$$

$$A_{22} = \begin{vmatrix} 10 & 630 \\ 630 & 41700 \end{vmatrix}\ (-1)^{2+2} = 10\ 41700 - 630^2 = 20100;$$

$$A_{23} = \begin{vmatrix} 10 & 930 \\ 630 & 59100 \end{vmatrix}\ (-1)^{2+3} = -(10\ 59100 - 930\ 630) = -5100;$$

$$A_{33} = \begin{vmatrix} 10 & 930 \\ 930 & 87700 \end{vmatrix}\ (-1)^{3+3} = 10\ 87700 - 930^2 = 12100.$$

Тогда

$$A^{-1}=\left(X^T X\right)^{-1}=\frac{1}{|A|}\;\tilde{A}^T=\frac{1}{2172\cdot 10^4}\begin{pmatrix} 16428\cdot 10^4 & -1548\cdot 10^3 & -288\cdot 10^3 \\ -1548\cdot 10^3 & 20100 & -5100 \\ -288\cdot 10^3 & -5100 & 12100 \end{pmatrix}=$$

$$=\begin{pmatrix} 4107/543 & -38,7/543 & -7,2/543 \\ -38,7/543 & 0,5025/543 & -0,1275/543 \\ -7,2/543 & -0,1275/543 & 0,3025/543 \end{pmatrix}.$$

Здесь воспользовались следующим свойством: при умножении матрицы на число все элементы матрицы умножаются на это число. Исходя из этого свойства, поделили все элементы матрицы на $2172\cdot 10^4$. После этого воспользовались формулой $B=A^{-1}\cdot X^T\cdot Y$:

$$B=\begin{pmatrix} 4107/543 & -38,7/543 & -7,2/543 \\ -38,7/543 & 0,5025/543 & -0,1275/543 \\ -7,2/543 & -0,1275/543 & 0,3025/543 \end{pmatrix}\begin{pmatrix} 68 \\ 6470 \\ 4450 \end{pmatrix}=$$

$$\begin{pmatrix} 4107\cdot 68/543 & -38,7\cdot 6470/543 & -7,2\cdot 4450/543 \\ -38,7\cdot 68/543 & 0,5025\cdot 6470/543 & -0,127\cdot 4450/543 \\ -7,2\cdot 68/543 & -0,1275\cdot 6470/543 & 0,3025\cdot 4450/543 \end{pmatrix}=\begin{pmatrix} -5,8 \\ 0,096 \\ 0,058 \end{pmatrix}.$$

Таким образом, $b_0=-5,8;\quad b_1=0,096;\quad b_2=0,058$ и уравнение регрессии будет

$$y=-5,8+0,096x_1+0,058x_2.$$

Оно показывает, что при увеличении только производственных фондов X_1 (при неизменном X_2) на 1 усл. ед. издержки производства Y увеличиваются в среднем на 0,096 усл. ед., а при увеличении численности занятых в производстве человек X_2 – в среднем на 0,058 усл. ед.

Оценим тесноту связи между зависимой переменной Y и объясняющими переменными. Найдем R и R^2 согласно формуле $R=\sqrt{\dfrac{Q_R}{Q}}=\sqrt{\dfrac{B^T\cdot X^T\cdot Y-n\cdot\bar{y}^2}{Y^T\cdot Y-n\cdot\bar{y}^2}}$, имеем:

$B^T=\left(-5,8;\;\; 0,096;\;\; 0,058\right)$; $\bar{y}=\dfrac{1}{10}\cdot\left(5+10=10+7+5+6+6+5+6+8\right)=\dfrac{68}{10}=6,8;$

$$B^T X^T Y=\left(-5,8;\; 0,096;\; 0,058\right)\begin{pmatrix} 1 & 1 & \ldots 1 \\ 80 & 100 & \ldots 90 \\ 50 & 80 & \ldots 70 \end{pmatrix}\begin{pmatrix} 5 \\ 10 \\ \ldots \\ 8 \end{pmatrix}=(-5,8;\; 0,096;\; 0,058)\begin{pmatrix} 68 \\ 6470 \\ 4450 \end{pmatrix}=$$

$$=(-5,8\cdot 68+0,096\cdot 6470+0,058\cdot 4450)=484,82;$$

$$y^T y = (5,\ 10,\ 10,\ ...,\ 8)\begin{matrix} 5 \\ 10 \\ ... \\ 8 \end{matrix} = 5^2 + 10^2 + ... + 8^2 = 496.$$

Данное произведение равно сумме квадратов чисел последней строки таблицы; $R = \sqrt{\dfrac{484,82 - 10\ 6,82}{496 - 10\ 6,8^2}} = 0,82$ говорит о сильной связи между значениями переменных в рассматриваемых выборках; $R^2 = 0,82^2 = 0,67$ свидетельствует о хорошей точности аппроксимации экспериментальных данных полученным уравнением регрессии.

Задание 3.1.3. Оценить значимость уравнения регрессии $y = -5,8 + 0,096x_1 + 0,058x_2$ примера 3.1.2 при α = 0,05.

Решение

Имеем $n = 10;\ p = 3;\ k_1 = 3;\ k_2 = 10 - 3 - 1 = 6;\ R^2 = 0,67$ и $F_{0,05;3;6} = 4,76$ (см. табл. 1 прилож.). Тогда $F = \dfrac{0,67\ 6}{(1 - 0,67)\ 3} = 4,06$, поскольку $F < F_{0,05;3;6}$, то уравнение регрессии не значимо, т.е. полученную зависимость нельзя распространить на генеральную совокупность рассматриваемых величин.

Задание 3.1.4. Исходя из данных примера 3.1.2, сравнить раздельное влияние на издержки производства двух составляющих: основных производственных факторов и численности занятых в производстве человек, т.е. найти выражения $b_1 = b_1\ \dfrac{S_{x_1}}{S_y};\ b_2 = b_2\ \dfrac{S_{x_2}}{S_y};\ E_1 = b_1\ \dfrac{\overline{x}_1}{\overline{y}};\ E_2 = b_2\ \dfrac{\overline{x}_2}{\overline{y}}.$

Решение

Заметим, что при нахождении $b_1,\ b_2,\ E_1$ и E_2 сначала определяются $\overline{x}_1,\ S^2_{x_1},\ \overline{x}_2,\ S^2_{x_2},\ \overline{y},\ S^2_y,$ вычисление которых сводится к суммированию чисел и их квадратов во второй, третьей и четвертой строках таблицы, поэтому можно воспользоваться результатами примера 3.1.2:

$$\overline{x}_1 = \frac{1}{10}\left(80 + 100 + ... + 90\right) = \frac{930}{10} = 93;$$

$$S^2_{x_1} = \frac{1}{10}\left(80^2 + 100^2 + ... + 90^2\right) - \overline{x}_1^2 = \frac{87\,700}{10} - 93^2 = 121;$$

$$\overline{x}_2 = \frac{1}{10}\left(50+80+\ \dots +70\right) = \frac{630}{10} = 63;$$

$$S_{x_2}^2 = \frac{1}{10}\left(50^2+80^2+\ \dots +70^2\right) - \overline{x}_2^2 = \frac{41700}{10} - 63^2 = 201;$$

$$S_{x_1} = \sqrt{121} = 11; \quad S_{x_2} = \sqrt{201} = 14{,}18;$$

$$\overline{y} = \frac{1}{10}\left(5+10+\ \dots +8\right) = \frac{68}{10} = 6{,}8;$$

$$S_y^2 = \frac{1}{10}\left(5^2+10^2+\ \dots +8^2\right) - \overline{y}^2 = \frac{496}{10} - 6{,}8^2 = 3{,}36;$$

$$S_y = \sqrt{3{,}36} = 1{,}83; \quad b_1 = 0{,}096; \quad b_2 = 0{,}058;$$

$$b_1 = 0{,}096 \ \ \frac{11}{1{,}83} = 0{,}58; \quad b_2 = 0{,}058 \ \ \frac{14{,}18}{1{,}83} = 0{,}45;$$

$$E_1 = 0{,}096 \ \ \frac{93}{6{,}8} = 13{,}1; \quad E_2 = 0{,}058 \ \ \frac{63}{6{,}8} = 0{,}54.$$

Вывод: увеличение производственных фондов и численности занятых в производстве человек только на одно S_{x_1} или на одно S_{x_2} увеличивает в среднем издержки производства соответственно на 0,58 S_y или 0,45 S_y , а увеличение этих переменных на 1 % (от их средних значений) приводит в среднем к росту издержек производства на 1,31 и 0,54 % соответственно. Следовательно, на издержки производства большее влияние оказывает фактор основных производственных фондов по сравнению с фактором численности занятых в производстве человек.

Задание 3.1.5. Пусть зависимость спроса D на некоторый товар от цены p выражается формулой

$$D = D(p) = p^2 .$$

Определить, является ли спрос эластичным относительно цены, и найти, на сколько процентов изменится спрос при увеличении цены на 1 %.

Решение

По формуле $E_x(y) = x\frac{y}{y}$ находим эластичность спроса D относительно цены p:

$$E_p(D) = \frac{D_p}{D}\,p = \frac{2p^2}{p^2} > 1 .$$

Следовательно, спрос в данном случае эластичен, $E_p(D)$ показывает, на сколько процентов изменится спрос при увеличении цены на 1 %. Так, в рассмотренном примере получаем, что если цена $p = 6$ тыс.

руб., то спрос $E_{p=6}(D)=\frac{2\;6^2}{6^2}=2$, т.е. независимо от цены при увеличении ее на 1 % спрос возрастает на 2 %.

Рассмотрим, как применяется понятие эластичности спроса для анализа изменения суммарного дохода от реализации продукции.

Пусть $D=D(p)$ – спрос на некоторый товар при цене p. Если имеется достаточное количество спрашиваемого товара, то $D(p)$ – это также и количество проданного товара. В этом случае доход от продажи $R=p\;\; D(p)$. Найдем эластичность дохода по цене:

$$E_p(R)=\frac{R_p}{R}\,p=\frac{D+D_pP}{PD}\,p=1+\frac{D_p}{D}\,p=1+E_p(D)\cdot$$

Очевидно, можно считать, что $D(p)$ – убывающая функция, т.е. при увеличении цены на товар уменьшается спрос на него. Тогда $D_p<0$, а следовательно, при эластичном спросе $E_p(D)<-1$ и $E_p(R)<0$. При неэластичном спросе $E_p(D)>-1$ и $E_p(R)>0$. Но условие $E_p(R)<0$ означает, что $R_p<0$, т.е. в этом случае доход R – убывающая функция цены p. При $E_p(R)>0$ производная $R_p>0$, т.е. в этом случае доход R является возрастающей функцией цены p.

Вывод: если спрос эластичен, то изменение цены вызывает изменение дохода в противоположном направлении. При неэластичном спросе изменение цены и дохода происходит в одном и том же направлении.

Задание 3.1.6. Пусть спрос $D=D(p)=p^2+p$. Определить, как влияет увеличение цены p на доход $R=p\;\; D(p)$.

Решение

Найдем эластичность спроса относительно цены:

$$E_p(D)=\frac{D_p}{D}\,p=\frac{2p+1}{p^2+p}\,p=\frac{2p+1}{p+1}>1.$$

Следовательно, спрос эластичен, а поэтому увеличение (уменьшение) цены приводит к уменьшению (увеличению) дохода.

Относительную погрешность функции δy можно определить с помощью эластичности. Так как

$$\delta y=\left|\frac{\Delta y}{y}\right|\approx\left|\frac{dy}{y}\right|=\left|\frac{f\ (x)\Delta x}{f(x)}\right|=\left|\frac{xf\ (x)}{f(x)}\right|\ \left|\frac{\Delta x}{x}\right|,$$

то $\delta y = |E_x(y)|\delta x$, где $E_x(y)$ – эластичность функции $y = f(x)$; $\delta x = \left|\frac{\Delta x}{x}\right|$ – относительная погрешность измерения x.

Таким образом, относительная погрешность функции связана с относительной погрешностью измерения через эластичность функции.

Задание 3.1.7. Дана функция издержек $C(Q) = 2Q^3 - 12Q + 10$, где Q – объем выпускаемого товара. Оценить относительную погрешность вычисления $C(Q)$ при объеме товара $Q = 100$ ед., определенном с точностью до 3 %.

Решение

Найдем абсолютную величину эластичности функции $|E_Q(C)| = \left|\frac{Q\ \ C\ (Q)}{C(Q)}\right| = \left|\frac{6Q(Q^2 - 2)}{2(Q^3 - 6Q + 5)}\right|$. При $Q = 100$ $|E_Q(C)| = 3$. Тогда относительная погрешность вычисления $C(Q)$ $\delta_C = |E_Q(C)|\ \ \delta_Q = 3\ \ 5\ \% = 15\ \%$..

Задание 3.1.8. Найти предельные полезности функции полезности $U(x, y) = \ln x + \ln y - \ln(x + y)$ при количестве товара $x = 5$ ед. и $y = 6$ ед.

Решение

Найдем

$$U_x = \frac{1}{x} - \frac{1}{x + y} = \frac{y}{x(x + y)}; \quad U_y = \frac{1}{y} - \frac{1}{x + y} = \frac{x}{y(x + y)}.$$

Тогда $U_x(5{,}6) = \frac{6}{5(5 + 6)} = \frac{6}{55}$; $U_y(5{,}6) = \frac{5}{6(5 + 6)} = \frac{5}{66}$; $\frac{6}{55} > \frac{5}{66}$.

Следовательно, чувствительность набора (x, y) при количестве товара $x = 5$ ед. и $y = 6$ ед. больше для x, чем для y. Можно ли распространить этот вывод для любых значений x и y? Сравним U_x и U_y. Если $y > x$, то $U_x = \frac{y}{x(x + y)} > \frac{x}{y(x + y)} = U_y$, т.е. $U_x > U_y$ и чувствительность относительно товара x больше; если $y < x$, то, наоборот, $U_x < U_y$ и чувствительность относительно товара y больше.

Линии на плоскости с одинаковыми значениями функции называются линиями уровня.

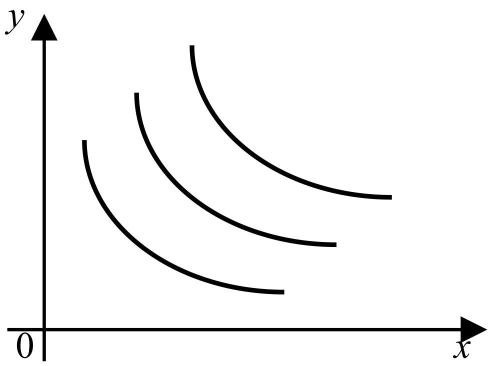
Линия уровня функции полезности называется линией безразличия. Эта линия соединяет потребительские наборы (x, y), имеющие один и тот же уровень удовлетворения потребностей индивидуума. Линии безразличия, соответствующие разным уровням удовлетворения потребностей, не касаются и не пересекаются.

Задание 3.1.9. Найти линии безразличия для функции полезности $U(x,y) = x^{1/3} \ y^{1/3}$.

Решение

Линии безразличия определяются уравнением $U(x,y) = C = \text{cons}$, поэтому $xy = C$, $x > 0$, $y > 0$. Графически такими кривыми являются ветви равнобочных гипербол, расположенных в первом квадранте (рисунок).

Дифференциал функции полезности записывается следующим образом: $dU(x,y) = \frac{\partial U}{\partial x}dx + \frac{\partial U}{\partial y}dy = \frac{\partial U}{\partial x}\Delta x + \frac{\partial U}{\partial y}\Delta y.$



Если переменные x и y изменяются незначительно, то результирующее (полное) изменение функции полезности можно получить по формуле

$$\Delta U = \frac{\partial U}{\partial x}\Delta x + \frac{\partial U}{\partial y}\Delta y.$$

Задание 3.1.10. Пусть функция полезности U (ютиль) задана следующим образом: $U(x,y) = 4x^{1/3} \ y^{2/3}$ (*x* – количество продукта первого вида (шт.), *y* – количество продукта второго вида (шт.)).

Требуется оценить (приближенно вычислить) полное изменение функции полезности, когда количество продукта x уменьшается от 15 до 13, а продукта y увеличивается от 120 до 130.

Решение

Вычислим предельные полезности продуктов:

$$\frac{\partial U}{\partial x} = \frac{4}{3}x^{-2/3}y^{2/3}, \quad \frac{\partial U}{\partial y} = \frac{8}{3}x^{1/3}y^{-1/3}.$$

Подставляя $x = 15$ и $y = 120$, находим:

$$\left.\frac{\partial U}{\partial x}\right|_{\substack{x=15\\y=120}} = \frac{4}{3}\ \left.\frac{y}{x}^{2/3}\right|_{\substack{x=15\\y=120}} = \frac{4}{3}\ 8^{2/3} = \frac{16}{3};$$

$$\left.\frac{\partial U}{\partial y}\right|_{\substack{x=15\\y=120}} = \frac{8}{3}\ \left.\frac{x}{y}^{1/3}\right|_{\substack{x=15\\y=120}} = \frac{8}{3}\ \frac{1}{8}^{1/3} = \frac{4}{3}.$$

Приращения независимых переменных $\Delta x = 13-15=-2$, Δy = $=130-120=10$.

Найдем приближенное значение функции полезности по формуле $\Delta U = \frac{\partial U}{\partial x}\Delta x + \frac{\partial U}{\partial y}\Delta y$:

$$\Delta U = \frac{16}{3}\ (-2) + \frac{4}{3}\ 10 = \frac{8}{3}.$$

Задание 3.1.11. Зависимость между себестоимостью единицы продукции y (ден. ед.) и выпуском продукции x (ден. ед.) выражается функцией $y=-0{,}6x+72$. Найти эластичность себестоимости при выпуске продукции, равном 40 ден. ед.

Решение

Эластичность себестоимости (для линейной функции) рассчитывается следующим образом:

$$E_x(y) = \frac{-0{,}6x}{-0{,}6x+72} = \frac{x}{x-120}.$$

При $x=40$ $E_{x=40}(y)=-0{,}5$, т.е. при выпуске продукции, равном 40 ден. ед., увеличение его на 1 % приводит к снижению себестоимости на 0,5 %.

Задание 3.1.12. Для двухфакторной линейной регрессии заданы парные коэффициенты корреляции:

$$r_{yx1}=0{,}8; \quad r_{yx2}=-0{,}2; \quad r_{x1x2}=-0{,}1.$$

Рассчитать частные коэффициенты корреляции и совокупный коэффициент множественной детерминации. Сделать выводы.

Решение

Частные коэффициенты корреляции характеризуют тесноту связи между результатом и соответствующим фактором при устранении влияния других факторов.

Рассчитаем частные коэффициенты корреляции:

$$r(yx_1/x_2)=\frac{r(yx_1)-r(yx_2)\ r(x_1x_2)}{\sqrt{(1-r(yx_2)^2)\ (1-r(x_1x_2)^2)}}=\frac{0,8-(-0,2)\ (-0,1)}{\sqrt{(1-(-0,2)^2)\ (1-(-0,1)^2)}}=$$
$$=0,8001.$$

Таким образом, если устранить влияние фактора x_2, то связь между y и x_1 будет тесной и прямой. Найдем

$$r(yx_2/x_1)=\frac{r(yx_2)-r(yx_1)\ r(x_1x_2)}{\sqrt{(1-r(yx_1)^2)\ (1-r(x_1x_2)^2)}}=\frac{-0,2-0,8\ (-0,1)}{\sqrt{(1-0,8^2)\ (1-(-0,1)^2)}}=$$
$$=-0,2010.$$

Получаем, что при устранении влияния фактора x_1 связь между y и x_2 будет слабой и обратной. Вычислим

$$r(x_1x_2/y)=\frac{r(x_1x_2)-r(yx_1)\ r(yx_2)}{\sqrt{(1-r(yx_1)^2)\ (1-r(yx_2)^2)}}=\frac{-0,1-0,8\ (-0,2)}{\sqrt{(1-0,8^2)\ (1-(-0,2)^2)}}=$$
$$=0,1021.$$

Если устранить влияние результативного признака y, то связь между x_1 и x_2 будет очень слабой и прямой.

Коэффициент множественной корреляции

$$R=\sqrt{\frac{r(yx_1)^2+r(yx_2)^2-2\ r(yx_1)\ r(yx_2)\ r(x_1x_2)}{1-r(x_1x_2)^2}}=$$
$$=\sqrt{\frac{0,8^2+(-0,2)^2-2\ 0,8\ (-0,2)\ (-0,1)}{1-(-0,1)^2}}=0,8481.$$

Вывод: связь результативного признака y с факторами x_1 и x_2 оценивается как тесная.

Задание 3.1.13. Заданы уравнение двухфакторной линейной регрессии (y = 80 – 0,4x_1 + 3x_2); средние значения признаков ($\overline{x}_1=11$; $\overline{x}_2=2$; $\overline{y}=90$); средние квадратические отклонения признаков ($S_{x1}=4$; $S_{x2}=1$; $S_y=6$); коэффициент корреляции факторных переменных ($r_{x1x2}=-0,7$).

Проанализировать влияние каждого из факторов на результат на основе частных коэффициентов эластичности, бета- (β-) и дельта-коэффициентов (Δ-коэффициентов).

Решение

Последовательность решения:

1. Определим частные коэффициенты эластичности.

Коэффициент эластичности показывает, на сколько процентов в среднем изменяется значение результативного признака при увеличении фактора x_j на 1 %, и рассчитывается по формуле

$$Э_j = b_j \frac{\overline{x_j}}{\overline{y}}.$$

Расчет коэффициентов эластичности:

$$Э_1 = b_1 \frac{\overline{x_1}}{y} = -0,4 \frac{11}{90} = -0,049;$$

$$Э_2 = b_2 \frac{\overline{x_2}}{y} = 3 \frac{2}{90} = 0,067.$$

Вывод: при увеличении фактора x_1 на 1 % значение результативного признака y в среднем снижается на 0,049 % при неизменном значении фактора x_2. При увеличении фактора x_2 на 1 % значение результативного признака y в среднем возрастает на 0,067 % при неизменном значении фактора x_1.

2. Определим β-коэффициенты.

β-коэффициент показывает, на какую часть своего среднего квадратического отклонения изменяется значение результативного признака при увеличении фактора x_j на величину своего среднего квадратического отклонения σ_{x_j}, и рассчитывается по формуле

$$\beta_j = b_j \frac{S_{x_j}}{S_y}.$$

Находим β-коэффициенты:

$$\beta_1 = b_1 \frac{S_{x_1}}{S_y} = -0,4 \frac{4}{6} = -0,267;$$

$$\beta_2 = b_2 \frac{S_{x_2}}{S_y} = 3 \frac{1}{6} = 0,500.$$

Вывод: при увеличении фактора x_1 на величину своего среднего квадратического отклонения (на 4) значение результативного признака y в среднем снижается на 0,267 от величины своего среднего квадратического отклонения при неизменном значении фактора x_2.

При увеличении фактора x_2 на величину своего среднего квадратического отклонения (на 1) значение результативного признака y в среднем увеличивается на 0,5 от величины своего среднего квадратического отклонения при неизменном значении фактора x_1.

3. Определим Δ-коэффициенты.

Δ-коэффициент показывает долю влияния фактора x_j в суммарном влиянии всех рассматриваемых факторов и рассчитывается по формуле

$$\Delta_j = \frac{r(yx_j)\cdot\beta_j}{\sum_{j=1}^{k} r(yx_j)\cdot\beta_j} = \frac{r(yx_j)\cdot\beta_j}{R^2},$$

где $R^2 = \sum_{j=1}^{k} r(yx_j)\cdot\beta_j$ – множественный коэффициент детерминации.

Находим парные коэффициенты корреляции:

$$r(yx_1) = \beta_1 + r(x_1x_2)\beta_2 = -0{,}267 + (-0{,}7)\cdot 0{,}5 = -0{,}6167;$$

$$r(yx_2) = r(x_1x_2)\beta_1 + \beta_2 = -0{,}7\cdot(-0{,}267) + 0{,}5 = 0{,}6867.$$

Определим значение множественного коэффициента детерминации:

$$R^2 = r(yx_1)\cdot\beta_1 + r(yx_2)\cdot\beta_2 = -0{,}267\cdot(-0{,}6167) + 0{,}5\cdot 0{,}6867 = 0{,}5078,$$

или

$$R^2 = \frac{r(yx_1)^2 + r(yx_2)^2 - 2\cdot r(yx_1)\cdot r(yx_2)\cdot r(x_1x_2)}{1 - r(x_1x_2)^2} =$$

$$= \frac{(-0{,}6167)^2 + 0{,}6867^2 - 2\cdot(-0{,}6167)\cdot 0{,}6867\cdot(-0{,}7)}{1-(-0{,}7)^2} = 0{,}5078.$$

Расчет Δ-коэффициентов:

$$\Delta_1 = \frac{r(yx_1)\cdot\beta_1}{R^2} = \frac{(-0{,}6167)\cdot(-0{,}267)}{0{,}5078} = 0{,}3239;$$

$$\Delta_2 = \frac{r(yx_2)\cdot\beta_2}{R^2} = \frac{0{,}6867\cdot 0{,}500}{0{,}5078} = 0{,}6761.$$

Вывод: 32,39 % влияния на результативный признак y обусловливает фактор x_1, 67,61 % изменения результативного признака y происходит за счет фактора x_2.

По модулю наибольшее значение всех показателей (коэффициента эластичности, β-коэффициента, Δ-коэффициента) наблюдается при факторе x_2, следовательно, он оказывает наиболее сильное влияние на результативный признак y.

Задание 3.1.14. Данные по западной природно-экономической зоне за 15 лет об урожайности многолетних трав, предназначенных для заготовки сена, y, внесении удобрений на 1 га пашни x_1 и осадках за май-июнь x_2 приведены ниже:

№ п/п	y	x_1	x_2
1	13,6	161	360
2	14,1	170	223
3	13,2	188	144
4	18,6	209	324
5	16,9	240	227
6	21	334	212
7	22,2	377	230
8	29,6	399	204
9	31,3	404	156
10	32,1	451	200
11	26,7	501	163
12	32,8	538	315
13	31,4	579	280
14	31	600	251
15	26,1	614	386

Требуется:

1) рассчитать параметры двухфакторного линейного уравнения регрессии и дать их интерпретацию;

2) охарактеризовать тесноту связи между признаками;

3) определить значимость уравнения регрессии и его коэффициентов;

4) определить степень влияния каждого из факторных признаков на результат, а также их совокупное влияние.

Решение

Последовательность решения:

1. Рассчитаем параметры двухфакторного линейного уравнения регрессии и дадим их интерпретацию.

Общий вид линейного двухфакторного уравнения регрессии:

$$\hat{y} = a + b_1 x_1 + b_2 x_2 ,$$

где $\hat{y}$ – расчетные (теоретические) значения результативного признака; a, b_1, b_2 – оценки параметров линейного уравнения регрессии; x_1, x_2 – значения факторных признаков.

Параметры линейного уравнения найдем с помощью МНК. Для определения параметров необходимо решить систему линейных уравнений:

$$na + \sum_{i=1}^{n} x_{1i} b_1 + \sum_{i=1}^{n} x_{2i} b_2 = \sum_{i=1}^{n} y_i;$$

$$\sum_{i=1}^{n} x_{1i} a + \sum_{i=1}^{n} x_{1i}^{2} b_1 + \sum_{i=1}^{n} x_{1i} x_{2i} b_2 = \sum_{i=1}^{n} y_i x_{1i};$$

$$\sum_{i=1}^{n} x_{2i} a + \sum_{i=1}^{n} x_{1i} x_{2i} b_1 + \sum_{i=1}^{n} x_{2i}^{2} b_2 = \sum_{i=1}^{n} y_i x_{2i}.$$

Вспомогательные промежуточные расчеты представлены ниже:

№ п/п	y	x_1	x_2	x_1^2	x_1x_2
1	13,6	161	360	25921	57960
2	14,1	170	223	28900	37910
3	13,2	188	144	35344	27072
4	18,6	209	324	43681	67716
5	16,9	240	227	57600	54480
6	21	334	212	111556	70808
7	22,2	377	230	142129	86710
8	29,6	399	204	159201	81396
9	31,3	404	156	163216	63024
10	32,1	451	200	203401	90200
11	26,7	501	163	251001	81663
12	32,8	538	315	289444	169470
13	31,4	579	280	335241	162120
14	31	600	251	360000	150600
15	26,1	614	386	376996	237004
Итого	360,6	5 765	3675	2583631	1438133
Среднее	24,04	384,33	245,00	172242,07	95875,53

№ п/п	yx_1	x_2^2	yx_2	y^2
1	2189,6	129600	4896	184,96
2	2397	49729	3144,3	198,81
3	2481,6	20736	1900,8	174,24
4	3887,4	104976	6026,4	345,96
5	4056	51529	3836,3	285,61
6	7014	44944	4452	441
7	8369,4	52900	5106	492,84
8	11810,4	41616	6038,4	876,16
9	12645,2	24336	4882,8	979,69
10	14477,1	40000	6420	1030,41
11	13376,7	26569	4352,1	712,89

12	17646,4	99225	10332	1075,84
13	18180,6	78400	8792	985,96
14	18600	63001	7781	961
15	16025,4	148996	10074,6	681,21
Итого	153156,8	976557	88034,7	9426,58
Среднее	10210,45	65103,80	5868,98	628,44

Получаем систему

$$15a + 5765b_1 + 365b_2 = 360,6;$$
$$5765a + 2583631b_1 + 1438133b_2 = 153156,8;$$
$$3675a + 1438133b_1 + 976557b_2 = 88034,7.$$

Решая систему относительно неизвестных параметров, получаем:

$$a = 12,726;\ b_1 = 0,041;\ b_2 = -0,018.$$

Вывели двухфакторную модель регрессии:

$$\hat{y} = 12,726 + 0,041x_1 - 0,018x_2.$$

Вывод: коэффициент регрессии b_1 показывает, что при увеличении факторного признака x_1 (внесении удобрений на 1 га пашни) на 1 ед. значение результативного признака (урожайности многолетних трав на сено) в среднем увеличится на 0,041 ед. при неизменном значении фактора x_2.

Коэффициент регрессии b_2 показывает, что при увеличении факторного признака x_2 (величины осадков) на 1 ед. значение результативного признака (урожайности многолетних трав) в среднем уменьшается на 0,018 ед. при неизменном значении фактора x_1.

Свободный член a = 12,726 оценивает влияние прочих факторов, оказывающих воздействие на результативный признак. Таким образом, воздействие прочих факторов увеличивает значение урожайности.

Теоретические (расчетные) значения результативного показателя (урожайности многолетних трав) получаем путем последовательной подстановки значений факторов x_1, x_2 в уравнение регрессии.

Рассчитываем остатки как разность фактических и расчетных значений:

$$\varepsilon_i = y_i - \hat{y}_i.$$

2. Оценим тесноту связи между признаками.

Заполним вспомогательную таблицу для оценки качества модели:

№ п/п	y_i	x_{i1}	x_{i2}	$y_i - \overline{y}_i$	$\hat{y}_i$	$\varepsilon_i = y_i - \hat{y}_i$	$\varepsilon_i^{\ 2}$
1	13,6	161	360	108,994	12,864	0,736	0,542
2	14,1	170	223	198,810	15,681	–1,581	2,499
3	13,2	188	144	174,240	17,828	–4,628	21,421
4	18,6	209	324	345,960	15,467	3,133	9,813
5	16,9	240	227	285,610	18,468	–1,568	2,458
6	21	334	212	441,000	22,575	–1,575	2,479
7	22,2	377	230	492,840	24,009	–1,809	3,272
8	29,6	399	204	876,160	25,372	4,228	17,876
9	31,3	404	156	979,690	26,434	4,866	23,674
10	32,1	451	200	1030,410	27,567	4,533	20,548
11	26,7	501	163	712,890	30,270	–3,570	12,748
12	32,8	538	315	1075,840	29,064	3,736	13,960
13	31,4	579	280	985,960	31,364	0,036	0,001
14	31	600	251	961,000	32,740	–1,740	3,027
15	26,1	614	386	681,210	30,898	–4,798	23,019
Итого	360,6	5765	3675	9 350,614	360,600	0,000	157,336

Найдем средние квадратические отклонения признаков:

$$\sigma_y = \sqrt{\overline{y^2} - \overline{y}^2} = \sqrt{628,44 - 24,04^2} = 7,108;$$

$$\sigma_{x_1} = \sqrt{\overline{x_1^{\ 2}} - \overline{x_1}^2} = \sqrt{172\,242,07 - 384,33^2} = 156,620;$$

$$\sigma_{x_2} = \sqrt{\overline{x_2^{\ 2}} - \overline{x_2}^2} = \sqrt{65103,80 - 245^2} = 71,266.$$

Расчет коэффициентов парной корреляции:

$r(yx_1) = \dfrac{\overline{yx_1} - \overline{y}\ \overline{x_1}}{\sigma_y\ \sigma_{x_1}} = 0,872$ (связь между количеством внесенных удобрений на 1 га пашни и урожайностью тесная и прямая);

$r(yx_2) = \dfrac{\overline{yx_2} - \overline{y}\ \overline{x_2}}{\sigma_y\ \sigma_{x_2}} = -0,041$ (связь между количеством осадков и урожайностью практически отсутствует);

$r(x_1x_2) = \dfrac{\overline{x_1x_2} - \overline{x_1}\ \overline{x_2}}{\sigma_{x_1}\ \sigma_{x_2}} = 0,154$ (связь между факторами очень слабая и прямая).

Вывод: фактор «количество осадков» x_2 следует исключить из модели. Он практически не влияет на урожайность.

Коэффициент множественной корреляции

$$R=\sqrt{\frac{r(yx_1)^2+r(yx_2)^2-2r(yx_1)r(yx_2)r(x_1x_2)}{1-r(x_1x_2)^2}}=$$

$$=\sqrt{\frac{0{,}872^2+(-0{,}041)^2-2\ \ 0{,}872\ \ (-0{,}041)\ \ 0{,}154}{1-0{,}154^2}}=0{,}890.$$

Вывод: связь урожайности y с факторами x_1, x_2, включенными в модель, тесная.

Коэффициент детерминации

$$R^2_{xy}=1-\frac{\sum\limits_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum\limits_{i=1}^{n}(y_i-\overline{y})^2}=1-\frac{157{,}336}{9\ 350{,}614}=0{,}983.$$

Вывод: коэффициент детерминации показывает, что 98,3 % вариации результативного признака y происходит под влиянием факторов x_1 и x_2, включенных в модель. Остальные 1,7 % вариации результативного признака y объясняются влиянием прочих случайных факторов.

3. Определим значимость уравнения регрессии.

Выдвигаем нулевую гипотезу о том, что найденные показатели тесноты связи случайны, т.е. равны нулю:

$$H_0: R^2_{xy}=0.$$

Для проверки нулевой гипотезы рассчитываем значение F-критерия Фишера:

$$F=\frac{R^2_{xy}}{1-R^2_{xy}}\ \frac{n-k-1}{k}=\frac{0{,}983}{1-0{,}983}\ \frac{15-2-1}{2}=350{,}585.$$

По таблице значений критерия Фишера находим табличное (критическое) значение критерия на уровне значимости $\alpha = 0{,}05$ и с числом степеней свободы $df_1 = k = 2$ (число факторов в модели равно двум) и $df_2 = n - k - 1 = = 12$:

$$F_{табл} = 3{,}885.$$

Вывод: поскольку $F > F_{\text{табл}}$, то нулевую гипотезу $H_0: R = R^2 = 0$ отклоняем с вероятностью допустить ошибку в 5 %. В качестве альтернативы принимаем гипотезу о том, что характеристики тесноты связи не случайны, т.е. построенная модель регрессии значима.

4. Определим значимость коэффициентов уравнения регрессии.

Выдвигаем нулевые гипотезы о том, что найденные параметры не являются статистически значимыми:

$$H_0: a=0;\ H_0: b_1=0,\ H_0: b_2=0.$$

Для проверки гипотез рассчитаем *t*-критерий Стьюдента по формулам:

$$t(a)=\frac{|a|}{S(a)};\quad t(b_j)=\frac{|b_j|}{S(b_j)},$$

где $S(a), S(b_j)$ – стандартные ошибки параметров модели.

Находим стандартную ошибку модели:

$$S_{\hat{y}}=\sqrt{\frac{\sum_{i=1}^{n}\varepsilon_i^2}{n-k-1}}=\sqrt{\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{n-k-1}}=\sqrt{\frac{157,336}{15-2-1}}=3,621.$$

Стандартные ошибки параметров определяют по формуле

$$S(b_j)=S_{\hat{y}}\sqrt{b_{jj}}\ ,$$

где b_{jj} – диагональные элементы матрицы $\left(X^TX\right)^{-1}$.

Находим обратную матрицу:

$$\left(X^TX\right)^{-1}=\begin{pmatrix} n & \sum x_1 & \sum x_2 \\ \sum x_1 & \sum x_1^2 & \sum x_1x_2 \\ \sum x_2 & \sum x_1x_2 & \sum x_2^2 \end{pmatrix}^{-1}=\begin{pmatrix} 15 & 5765 & 3675 \\ 5765 & 2583631 & 1438133 \\ 3675 & 1438133 & 976557 \end{pmatrix}^{-1}=$$

$$=\begin{pmatrix} 1,107862 & 0,000840 & -0,002933 \\ -0,000840 & 0,000003 & -0,000001 \\ -0,002933 & -0,000001 & 0,000013 \end{pmatrix}.$$

Стандартные ошибки параметров модели:

$$S(a)=S_{\hat{y}}\cdot\sqrt{b_{11}}=3,621\cdot\sqrt{1,107862}=3,811;$$

$$S(b_1)=S_{\hat{y}}\cdot\sqrt{b_{22}}=3,621\cdot\sqrt{0,000003}=0,006;$$

$$S(b_2)=S_{\hat{y}}\cdot\sqrt{b_{33}}=3,621\cdot\sqrt{0,000013}=0,013.$$

Расчет фактических значений критерия Стьюдента:

$$t(a)=\frac{12,726}{3,811}=3,339;\quad t(b_1)=\frac{0,041}{0,006}=6,760;\quad t(b_2)=\frac{0,018}{0,013}=1,347.$$

По таблице значений критерия Стьюдента находим табличное (критическое) значение критерия на уровне значимости $\alpha = 0,05$ и с числом степеней свободы $df = n - k - 1 = 15 - 3 = 12$:

$$t_{\text{табл}} = 2,179.$$

Вывод: поскольку $t(b_2) < t_{\text{табл}}$, то коэффициент регрессии b_2 не является статистически значимым, нулевую гипотезу принимаем, т.е.

$H_0 : b_2 = 0$. Поскольку коэффициент регрессии не значим, то и фактор x_2 при этом коэффициенте не значим, его следует исключить из модели.

Остальные параметры модели являются статистически значимыми, фактические значения критерия больше табличного, нулевые гипотезы отклоняем.

5. Определим степень влияния каждого из факторных признаков на результат.

Расчет коэффициентов эластичности:

$$Э_1 = b_1 \ \frac{\overline{x_1}}{y} = 0,041 \ \frac{384,33}{24,04} = 0,653;$$

$$Э_2 = b_2 \ \frac{\overline{x_2}}{y} = -0,018 \ \frac{245}{24,04} = -0,182.$$

Вывод: при увеличении фактора x_1 на 1 % значение результативного признака y в среднем возрастает на 0,653 %. При увеличении фактора x_2 на 1 % значение результативного признака y в среднем снижается на 0,182 %.

Находим β-коэффициенты:

$$\beta_1 = b_1 \ \frac{\sigma_{x_1}}{\sigma_y} = 0,041 \ \frac{156,620}{7,108} = 0,900;$$

$$\beta_2 = b_2 \frac{\sigma_{x_2}}{\sigma_y} = -0,018 \ \frac{71,266}{7,108} = -0,179.$$

Вывод: при увеличении фактора x_1 на величину своего среднего квадратического отклонения значение результативного признака y в среднем возрастает на 0,9 величины своего среднего квадратического отклонения.

При увеличении фактора x_2 на величину своего среднего квадратического отклонения значение результативного признака y в среднем снижается на 0,179 величины своего среднего квадратического отклонения.

Расчет Δ-коэффициентов:

$$\Delta_1 = \frac{r(yx_1) \ \beta_1}{R^2} = \frac{0,900 \ \ 0,872}{0,792} = 0,991;$$

$$\Delta_2 = \frac{r(yx_2) \ \beta_2}{R^2} = \frac{-0,179 \ \ (-0,041)}{0,792} = 0,009.$$

Вывод: 99,1 % влияния на результативный признак y обусловливает фактор x_1, 0,9 % изменения результативного признака y происходит за счет фактора x_2..

По модулю наибольшее значение коэффициента эластичности, β-коэффициента, Δ-коэффициента имеет место при факторе x_1, следовательно, он оказывает наиболее сильное воздействие на результативный признак y.

Задание 3.1.15. Построить модель двухфакторной линейной регрессии (рассчитать параметры уравнения регрессии матричным методом) по данным, приведенным ниже:

№ п/п	1	2	3	4	5	6	7	8	9	10	11	12	13	14
y	2	4	3	6	5	8	7	9	10	12	14	11	15	17
X_1	1	2	4	3	6	7	5	8	10	9	11	13	12	14
X_2	5	3	6	4	7	8	10	13	9	11	12	14	15	16

Решение

Уравнение двухфакторной регрессии имеет вид: $y = b_0 + b_1x_1 + b_2x_2$.

Обозначим $Y = \begin{matrix} 2 \\ 4 \\ \dots \\ 17 \end{matrix}$, $\quad X = \begin{matrix} 1 & 1 & 5 \\ 1 & 2 & 3 \\ \dots & \dots & \dots \\ 1 & 14 & 16 \end{matrix}$.

Найдем $Y = \begin{matrix} 2 \\ 7 \\ \dots \\ 17 \end{matrix}$, $\quad X^T\ X = \begin{matrix} 1 & 1\dots 1 \\ 1 & 2\dots 14 \\ 5 & 3\dots 16 \end{matrix} \begin{matrix} 1 & 1 & 5 \\ 1 & 2 & 3 \\ \dots & \dots & \dots \\ 1 & 14 & 16 \end{matrix} = \begin{matrix} 14 & 105 & 133 \\ 105 & 1015 & 1206 \\ 133 & 1206 & 1491 \end{matrix}$.

Вычислим $X^T\ Y = \begin{matrix} 1 & 1\dots 1 \\ 1 & 2\dots 14 \\ 5 & 3\dots 16 \end{matrix} \begin{matrix} 2 \\ 4 \\ \dots \\ 17 \end{matrix} = \begin{matrix} 123 \\ 1156 \\ 1391 \end{matrix}$.

Найдем $A^{-1} = \left(X^T\ X\right)^{-1} = \begin{matrix} 14 & 105 & 133 \\ 105 & 1015 & 1206 \\ 133 & 1206 & 1491 \end{matrix}^{-1} = \begin{matrix} 0,508 & 0,033 & -0,072 \\ 0,033 & 0,028 & -0,025 \\ -0,072 & -0,025 & 0,028 \end{matrix}$;

$$B = A^{-1}\ X^T\ Y = \begin{matrix} 0,508 & 0,033 & -0,072 \\ 0,033 & 0,028 & -0,025 \\ -0,072 & -0,025 & 0,028 \end{matrix} \begin{matrix} 123 \\ 1156 \\ 1391 \end{matrix} = \begin{matrix} 0,475 \\ 0,812 \\ 0,235 \end{matrix}.$$

Таким образом, $b_0 = 0,475$; $b_1 = 0,812$; $b_2 = 0,235$ и уравнение регрессии будет $y = 0,475 + 0,812\ x_1 + 0,235\ x_2$.

Задание 3.1.16. Предположим, что стоимость Y типа имущества зависит от двух основных факторов – X_1 и X_2. Построить уравнение двухфакторной регрессии по данным, приведенным ниже:

X_1 (усл. ед.)	8	11	12	9	8	8	9	10
X_2 (усл. ед.)	6	8	8	5	7	8	6	4
Y (усл. ед.)	5	10	10	7	5	6	6	5

Решение

Последовательность решения:

1. Расчет коэффициентов множественной линейной регрессии.

Система трех линейных уравнений с тремя неизвестными b_0, b_1, b_2:

$\sum y_i = nb_0 + b_1\sum x_{1i} + b_2\sum x_{2i}$;

$\sum x_{1i}y_i = b_0\sum x_{1i} + b_1\sum x_{1i}^2 + b_2\sum x_{1i}x_{2i}$;

$\sum x_{2i}y_i = b_0\sum x_{2i} + \mathrm{b}_1\sum x_{1i}x_{2i} + b_2\sum x_{2i}^2$.

Для удобства проведения расчетов оформим результаты промежуточных расчетов следующим образом:

	Y	X_1	X_2	X_1^2	X_2^2	X_1Y	X_2Y	X_1X_2	Y^2
	5	8	6	64	36	40	30	48	25
	10	11	8	121	64	110	80	88	100
	10	12	8	144	64	120	80	96	100
	7	9	5	81	25	63	35	45	49
	5	8	7	64	49	40	35	56	25
	6	8	8	64	64	48	48	64	36
	6	9	6	81	36	54	36	54	36
	5	10	4	100	16	50	20	40	25
Сумма	54	75	52	719	354	525	364	491	396
Среднее	6,75	9,375	6,5	89,875	44,25	65,625	45,5	61,375	49,5

Для исходных данных система уравнений имеет вид:

$$54 = 8\,b_0 + 75b_1 + 52b_2;$$

$$525 = 75b_0 + 719b_1 + 491b_2;$$

$$364 = 52b_0 + 491b_1 + 354b_2.$$

Решая систему методом Крамера, получим

$$b_0 = -6{,}904;\ b_1 = 1{,}053;\ b_2 = 0{,}582.$$

Уравнение двухфакторной регрессии будет:

$$Y = -6{,}904 + 1{,}053\ X_1 + 0{,}582\ X_2.$$

Найдем средние квадратические отклонения признаков:

$$S(y) = \sqrt{\overline{y^2} - \overline{y}^2} = \sqrt{49{,}5 - 6{,}75^2} = 1{,}984;$$

$$S(x_1) = \sqrt{\overline{x_1^2} - \overline{x_1}^2} = \sqrt{89{,}875 - 9{,}375^2} = 1{,}409;$$

$$S(x_2) = \sqrt{\overline{x_2^2} - \overline{x_2}^2} = \sqrt{44{,}25 - 6{,}5^2} = 1{,}414.$$

Рассчитаем парные коэффициенты корреляции.

Для y и x_1

$$r_{yx1} = \frac{\overline{y \cdot x_1} - \overline{y} \cdot \overline{x_1}}{S(Y) \cdot S(X_1)} = \frac{65{,}625 - 6{,}75 \cdot 9{,}375}{1{,}984 \cdot 1{,}409} = 0{,}838.$$

Для y и x_2

$$r_{yx2} = \frac{\overline{y \cdot x_2} - \overline{y} \cdot \overline{x_2}}{S(Y) \cdot S(X_2)} = \frac{45{,}5 - 6{,}75 \cdot 6{,}5}{1{,}984 \cdot 1{,}414} = 0{,}579.$$

Для x_1 и x_2

$$r_{x1x2} = \frac{\overline{x_1 \cdot x_2} - \overline{x_1} \cdot \overline{x_2}}{S(x_1) \cdot S(x_2)} = \frac{61{,}375 - 9{,}375 \cdot 6{,}5}{1{,}409 \cdot 1{,}414} = 0{,}22.$$

Для определения размеров погрешности или точности прогноза показателя Y найдем коэффициент несоответствия Тейла по формуле

$$K_T = \frac{\sqrt{\sum e^2}}{\sqrt{\sum Y_i^2}}.$$

Получаем

$$K_T = \frac{4.192}{396} = 0{,}0106.$$

Исследователь проблем экономического прогнозирования Г. Тейл предложил в качестве меры качества прогнозов коэффициент расхождения K_T (коэффициент несоответствия), числителем которого является среднеквадратическая ошибка прогноза, а знаменатель равен квадратному корню из среднего квадрата реализации. Если $K_T = 0$, то прогноз абсолютно точен (случай «идеального» прогнозирования). Этот показатель изменяется от 0 до 1. Чем ближе его значение к нулю, тем лучше результаты прогнозирования.

Задание 3.1.17. По экономическому уровню развития страны делятся на 3 группы: ведущие экономики; развивающиеся страны; страны с переходной экономикой.

Требуется определить, сколько фиктивных переменных следует включить в уравнение множественной регрессии динамики производства продукции.

Решение

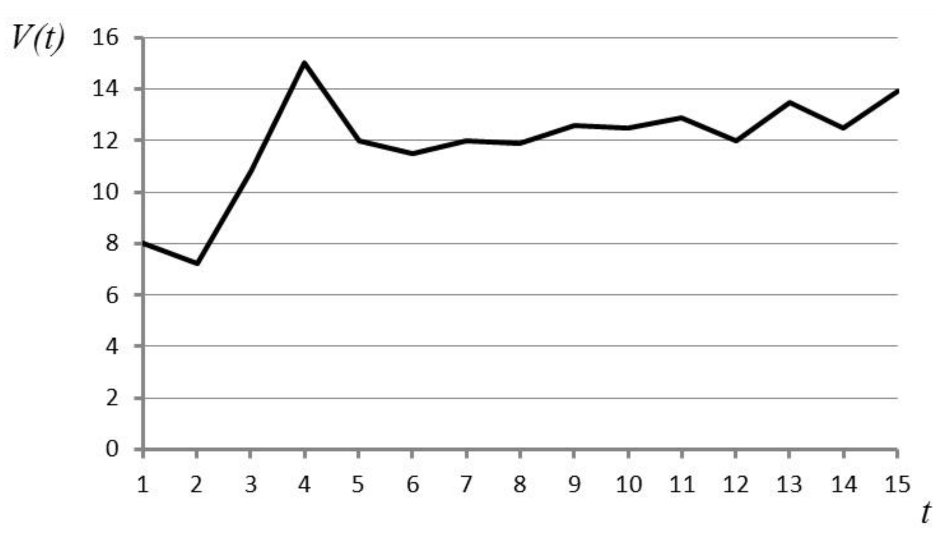
Если качественная переменная имеет 3 уровня, то для моделирования достаточно 2 фиктивных переменных – $d1$ и $d2$. Для обозначения третьего уровня достаточно принять, например, обе переменные равными нулю: $d1 = d2 = 0$. Тогда $d1 = d2 = 0$ означает страну с переходной экономикой. Нулевой уровень качественной переменной называется базовым, или сравнительным.

Задание 3.1.18. В одном из вузов объем математических знаний $V(t)$ студентов-экономистов по неделям распределяется в первом семестре согласно данным, приведенным ниже:

Недели	1	2	3	4	5	6	7	8
$V(t)$	8	7,2	10,8	15	12	11,5	12	11,9
Недели	9	10	11	12	13	14	15	
$V(t)$	12,6	12,5	12,9	12	13,5	12,5	13,9	

Общее количество учебных часов в первом семестре – 60.

График изменения объема часов по неделям имеет следующий вид:



Требуется на основе этих данных построить эконометрическую модель зависимости объема часов от времени.

Решение

Из графика зависимости видно, что в рассматриваемом случае имеется 2 этапа изучения учебного материала:

1) объем знаний резко возрастает в начальные 4 недели;

2) получаемый объем знаний остается практически стабильным в последующие 11 недель, немного повышается к концу семестра.

Единая линейная эконометрическая модель $V(t)$ за весь семестр имеет низкое качество (коэффициент детерминации $R^2 = 0,4$), поэтому в данном случае целесообразно построить эконометрическую модель, состоящую из двух уравнений для каждого из этапов. Тогда для первого этапа линейная регрессионная модель будет иметь вид

$$y = 2,46 \ \ x + 4,1 (R^2 = 0,81),$$

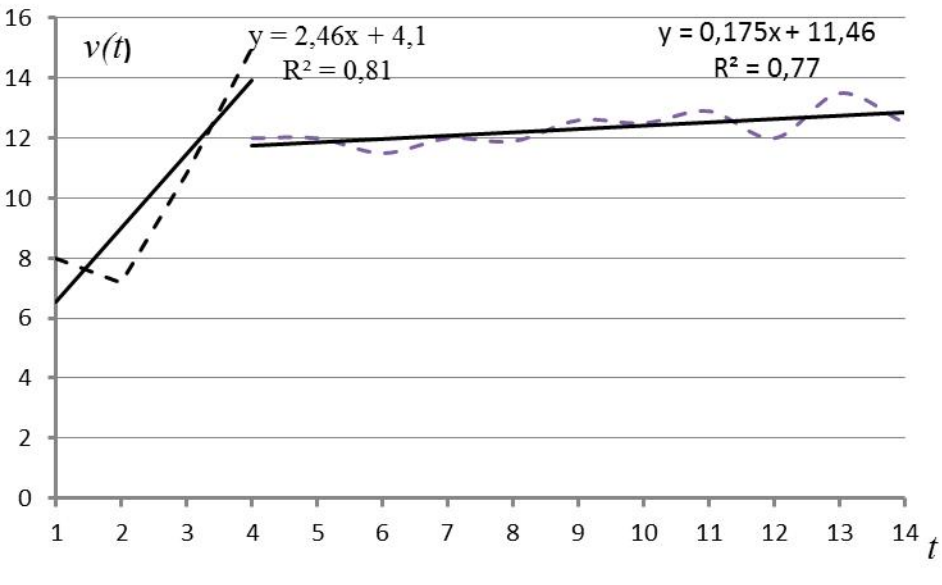
где $y = V(t)$, $x = t$.

Для второго этапа имеем

$$y = 0,175 \ \ x + 11,46 \ (R^2 = 0,77),$$

где $y = V(t)$, $x = t$.

Две линейных регрессионных модели изображены на графиках ниже:



Получить единую для двух этапов регрессионную модель можно с помощью метода фиктивных переменных. Введем фиктивную переменную *d*, которая определяется следующим образом: $d = \begin{cases} 0, \text{ если } d \quad 4; \\ 1 - \text{ иначе.} \end{cases}$

Тогда общее уравнение можно записать так:

$$y = 2,46 \ \ x + 4,1 + 0,175 \ \ x \ \ d + 11,46 \ \ d \, .$$

Отражение двух этапов получения знаний можно использовать для определения наличия структурных сдвигов в процессе обучения.

Объем получения знаний можно рассматривать как для одного обучаемого, так и для группы обучаемых. Для группы это может быть, например, усредненное по студентам этой группы значение показателя.

Планируемую производительность изучения дисциплины можно определить как первую производную от $V(t)$, вторая производная $V\ (t)$ будет соответствовать скорости изменения производительности. Еще одна характеристика – темп изменения производительности – определяется как $(\ln V\ (t))\ $.

Для рассматриваемого примера имеем $V\ (t) = 2{,}46$ для первой части; $V\ (t) = 0{,}175$ – для второй части; $V\ (t) = 0$ и $(\ln V\ (t))\ = 0$ – для обеих частей.

Таким образом, для первой части процесс ускорения более существенен.

Задания для самостоятельного выполнения

Задание 1. По матрице парных коэффициентов корреляции отобрать факторы для включения в модель множественной регрессии.

	y	x_1	x_2	x_3
y	1			
x_1	0,8	1		
x_2	0,7	0,1	1	
x_3	0,5	0,2	0,8	1

	y	x_1	x_2	x_3
y	1			
x_1	0,7	1		
x_2	0,6	0,7	1	
x_3	0,8	0,2	0,2	1

	y	x_1	x_2	x_3
y	1			
x_1	0,6	1		
x_2	0,7	0,1	1	
x_3	0,8	0,9	0,2	1

Задание 2. Построить модель двухфакторной линейной регрессии (рассчитать параметры уравнения регрессии матричным методом) по данным, приведенным ниже:

№	1	2	3	4	5	6	7	8	9	10	11	12	13	14
y	2	4	3	6	5	8	7	9	10	12	14	11	15	17
x_1	1	2	4	3	6	7	5	8	10	9	11	13	12	14
x_2	5	3	6	4	7	8	10	13	9	11	12	14	15	16

Задание 3. Построена эмпирическая модель двухфакторной линейной регрессии: $y = b_0 + b_1 x_1 + b_2 x_2$. Коэффициент корреляции $r_{x_1 x_2}$ переменных x_1 и x_2 равен точно 1. Чему равен определитель матрицы $X^T X$?

Задание 4. Предположим, что стоимость Y данного типа имущества зависит от двух основных факторов – X_1 и X_2. Данные выборочных наблюдений приведены ниже:

X_1 (усл. ед.)	8	11	12	9	8	8	9	10
X_2 (усл. ед.)	6	8	8	5	7	8	6	4
Y (усл. ед.)	5	10	10	7	5	6	6	5

По данным выборочных значений X_1, X_2 и Y построить уравнение регрессии. Оценить тесноту связи между переменными X_1, X_2 и Y, определить долю (часть) вариации Y, обусловленную регрессией.

Задание 5. Для двухфакторной линейной регрессии задана матрица парных коэффициентов корреляции:

	y	X_1	X_2
y	1		
X_1	0,6	1	
X_2	–0,5	–0,9	1

Определить частные коэффициенты корреляции результата с каждым из факторов. Объяснить различия значений парных и полученных частных коэффициентов корреляции.

Задание 6. Задано уравнение двухфакторной линейной регрессии: $y = 20 + 14x_1 + 21x_2$; средние значения признаков: $\overline{x}_1 = 3$; $\overline{x}_2 = 1$; $\overline{y} = 4$.

Проанализировать влияние каждого из факторов на результат на основе частных коэффициентов эластичности.

Задание 7. Дана матрица парных коэффициентов корреляции:

	y	x_1	x_2
y	1		
x_1	0,7	1	
x_2	0,6	0,3	1

Рассчитать по ней совокупный коэффициент множественной детерминации.

Задание 8. Определить наличие мультиколлинеарности факторов, если для двухфакторной линейной регрессии заданы значения парных коэффициентов корреляции между факторами: $r_{x_1x_2} = 0{,}1$; $r_{x_2x_3} = 0{,}05$; $r_{x_1x_2} = 0{,}1$. Объяснить полученный результат.

Задание 9. Определить коэффициенты эластичности ($Э_1$, $Э_2$, $Э_3$) для уравнения множественной регрессии $y = -66{,}028 + 0{,}135x_1 + 0{,}476x_2 + 0{,}343x_3$, если $\bar{x}_1 = 245{,}7$; $\bar{x}_2 = 3{,}7$; $\bar{x}_3 = 182{,}5$; $\bar{y} = 3$.

Задание 10. Имеются два качественных признака: тип потребительского поведения и сезон года (номер квартала). Согласно первому признаку все домашние хозяйства делятся на три социально-экономические страты: низко-, средне- и высокодоходные. Согласно второму признаку имеются четыре квартала (сезона). Сколько фиктивных переменных следует ввести в модель?

Задание 11. Предполагается, что ежемесячное потребление чая студентами определяется (линейно) доходом, возрастом, полом, а также периодом обучения (младшие курсы / старшие курсы). Сколько количественных и фиктивных переменных должна включать модель?

Задание 12. Пусть имеется уравнение регрессии с одним количественным и одним неколичественным фактором, выраженным тремя фиктивными переменными. Сколько возможных значений у неколичественного фактора? Как на основе заданного уравнения регрессии найти уравнения парной регрессии, содержащие только количественный фактор? Сколько будет таких уравнений и почему?

Задание 13. Имеются данные по 25 районам об уровне рентабельности производства и реализации молока (в процентах), об общих затратах на 1 корову (усл. ед.) и о среднегодовом поголовье коров. Получены средние значения признаков, стандартные отклонения и линейные коэффициенты корреляции:

Признак	Среднее значение	Стандартное отклонение	Линейные коэффициенты парной корреляции
Уровень рентабельности y	$\bar{y} = 6{,}4$	$S_y = 13{,}5$	$r_{yx_1} = -0{,}27$; $r_{yx_2} = 0{,}66$
Общие затраты на корову x_1 (тыс. руб.)	$\bar{x}_1 = 6{,}7$	$S_{x_1} = 1{,}3$	$r_{x_1x_2} = -0{,}11$
Поголовье коров x_2 (тыс. голов)	$\bar{x}_2 = 3{,}9$	$S_{x_2} = 2{,}2$	–

Требуется:

1) построить уравнение множественной регрессии зависимости результативного признака «уровень рентабельности» y от таких факторов, как общие затраты на корову x_1 и поголовье коров x_2. Пояснить экономический смысл параметров уравнения множественной регрессии;

2) оценить тесноту линейной связи с помощью коэффициента множественной корреляции и скорректированного коэффициента детерминации;

3) рассчитать средние коэффициенты эластичности и дать на их основе оценку силы влияния факторов (общих затрат на корову и поголовья коров) на результативный признак (уровень рентабельности);

4) оценить статистическую надежность и качество полученного уравнения регрессии в целом с помощью общего F-критерия Фишера, а также целесообразность включения в модель каждого из факторов с помощью частных F-критериев Фишера.

Задание 14. Исследуется динамика жилищного строительства в одном из регионов РФ: зависимость ввода площади жилого фонда (тыс. кв. м) от общего числа работников, занятых в жилищном строительстве (тыс. чел.), и затрат на строительство (тыс. усл. ед.). Данные на 2021–2022 годы приведены ниже:

Месяц и год	Строительство Y (тыс. кв. м)	Рабочие X_1 (тыс. чел.)	Затраты X_2 (тыс. усл. ед.)
Январь 2017 года	52,15	28,20	72,91
Февраль 2017 года	67,21	26,20	78,74
Март 2017 года	72,15	33,20	80,26
Апрель 2017 года	89,93	43,30	87,61
Май 2017 года	98,41	42,90	93,79
Июнь 2017 года	99,35	64,00	110,18
Июль 2017 года	96,49	40,10	128,38
Август 2017 года	88,83	34,40	109,76
Сентябрь 2017 года	80,88	24,00	99,00
Октябрь 2017 года	85,75	24,40	119,65
Ноябрь 2017 года	72,35	25,60	102,27
Декабрь 2017 года	61,20	24,80	81,09
Январь 2018 года	46,56	22,00	76,06
Февраль 2018 года	60,36	30,40	79,00

Требуется:

1) построить модель множественной регрессии. Оценить ее качество (значимость уравнения и коэффициентов, среднюю ошибку аппроксимации);

2) дать интерпретацию построенной модели и отдельных ее коэффициентов;

3) найти коэффициенты эластичности, бета-коэффициенты, дельта-коэффициенты, коэффициент множественной детерминации. Объяснить полученные значения.

Задание 15. Исследуется успеваемость в вузе (зависимость между результатами зимней X и летней Y сессий). Ниже приведены средние оценки, полученные по итогам сессий, а также указана принадлежность студента к группе А или Б:

№ п/п	X	Y	№ п/п	X	Y	Группа
1	3,7	4,8	11	4,7	3,7	Б
2	3,5	3,5	12	4,6	4,4	Б
3	4,3	5	13	4,6	3,8	Б
4	3	4	14	3,3	3,1	Б
5	4,6	4,2	15	4,3	3,6	Б
6	4,6	4,1	16	3,1	4,8	А
7	3,8	4,8	17	3,2	3	А
8	3,6	3,5	18	4,2	4,8	А
9	3,3	4,4	19	3,3	3,4	Б
10	3,9	3	20	3,5	4,2	А

Требуется:

1) построить линейную регрессионную модель зависимости Y по X;

2) проверить значимость коэффициентов уравнения и самого уравнения регрессии;

3) построить регрессионную модель зависимости Y по X с использованием фиктивной переменной «группа»;

4) проверить значимость коэффициентов уравнения и самого уравнения регрессии с фиктивной переменной;

5) вычислить коэффициенты детерминации для обычной модели и модели с фиктивной переменной.

Задание 16. Какое минимальное количество наблюдений необходимо для оценки параметров линейной множественной регрессии вида $\hat{y} = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 + b_5x_5$?

Задание 17. Расход бензина y (л), в зависимости от скорости x (км/ч), описывается для данного автомобиля функцией вида $y = 20 - 0{,}2x + 0{,}002x^2$.

Оценить относительную погрешность определения расхода бензина при скорости x = 80 км/ч, установленной с точностью до 2 %.

Задание 18. Найти коэффициенты эластичности для множественной регрессии ($Э_1$, $Э_2$, $Э_3$), если $y = -66{,}028 + 0{,}135x_1 + 0{,}476x_2 + 0{,}343x_3$; $\overline{y} = 31{,}5$; $\overline{x}_1 = 245{,}7$; $\overline{x}_2 = 3{,}7$; $\overline{x}_3 = 182{,}5$.

Задание 19. После обработки данных пространственной выборки по 25 районам об уровне рентабельности производства и реализации молока (в процентах), об общих затратах на 1 корову (усл. ден. ед.) и о среднегодовом поголовье коров получены их средние значения, стандартные отклонения и линейные коэффициенты парной корреляции, приведенные ниже:

Признак	Среднее значение	Стандартное отклонение	Линейные коэффициенты парной
Уровень рентабельности y (%)	$y = 6{,}4$	$\sigma_y = 13{,}5$	$r_{yx1} = -0{,}27$ $r_{yx2} = 0{,}66$
Общие затраты на 1 корову, x_1 (усл. ден. ед.)	$x_1 = 6{,}7$	$\sigma_{x1} = 1{,}3$	$r_{x1x2} = -0{,}11$
Поголовье коров x (тыс. голов)	$x_2 = 3{,}9$	$\sigma_{x2} = 2{,}2$	–

Требуется:

1. Построить уравнение множественной регрессии зависимости результативного признака y (уровня рентабельности) от таких факторов, как общие затраты на 1 корову x_1 и поголовье коров x_2. Пояснить экономический смысл его параметров.
2. Оценить тесноту линейной связи с помощью коэффициента множественной корреляции и скорректированного коэффициента детерминации.
3. Рассчитать средние коэффициенты эластичности и дать на их основе сравнительную оценку силы влияния факторов (общих затрат на 1 корову и поголовья коров) на результативный признак (уровень рентабельности).
4. Оценить статистическую надежность и качество полученного уравнения регрессии в целом с помощью общего F-критерия Фишера, а также целесообразность включения в модель каждого из факторов с помощью частных F-критериев Фишера.

Задание 20. Данные по природно-экономической зоне за 15 лет об урожайности многолетних трав, предназначенных для заготовки сена, y, внесении удобрений на 1 га пашни x_1 и осадках за май-июнь x_2 приведены ниже:

Номер года	y	x_1	x_2
1	13,6	161	360
2	14,1	170	223
3	13,2	188	144
4	18,6	209	324
5	16,9	240	227
6	21	334	212
7	22,2	377	230
8	29,6	399	204
9	31,3	404	156
10	32,1	451	200
11	26,7	501	163
12	32,8	538	315
13	31,4	579	280
14	31	600	251
15	26,1	614	386

Необходимо проанализировать степень зависимости урожайности y от факторов x_1 и x_2. Для этого определить для каждого ряда данных y, x_1, x_2 средние значения, дисперсии, первые разности (абсолютные приросты); рассчитать параметры двухфакторного линейного уравнения регрессии по первым разностям (по абсолютным приростам) и дать их интерпретацию. Охарактеризовать тесноту связи между рядами.

Задание 21. По 20 предприятиям региона изучается зависимость выработки продукции на 1 работника y (тыс. руб.) от ввода в действие новых основных фондов x_1 (% от стоимости фондов на конец года) и от удельного веса рабочих высокой квалификации в общей численности рабочих x_2 (%). Полученные данные по исследуемой зависимости представлены ниже:

Номер предприятия	y	x_1	x_2	Номер предприятия	y	x_1	x_2
1	70	3,6	12	110	10	7,3	23
2	70	4,1	14	120	11	7,6	25
3	80	4,3	16	130	12	7,8	26
4	70	4,4	17	140	11	7,9	28
5	70	4,5	18	150	12	8,2	30
6	80	4,8	19	160	12	8,4	31

7	80	5,3	20	170	12	8,6	32
8	80	5,6	20	180	13	8,8	32
9	90	6,7	21	190	13	9,2	33
10	100	6,9	22	200	14	9,6	34

Следует:

1) построить линейную модель множественной регрессии. Записать стандартизованное уравнение множественной регрессии. На основе стандартизованных коэффициентов регрессии и средних коэффициентов эластичности ранжировать факторы по степени их влияния на результат;

2) найти коэффициенты парной, частной и множественной корреляции. Проанализировать их;

3) найти скорректированный коэффициент множественной детерминации. Сравнить его с нескорректированным (общим) коэффициентом детерминации;

4) с помощью F-критерия Фишера оценить статистическую надежность уравнения регрессии и коэффициента детерминации $R^2_{yx_1x_2}$;

5) с помощью частных F-критериев Фишера оценить целесообразность включения в уравнение множественной регрессии фактора x_1 после x_2 и фактора x_2 после x_1;

6) составить уравнение линейной парной регрессии, оставив лишь один значащий фактор.

Задание 22. На предприятии имеются станки трех марок (*A*, *B*, *C*). Исследуется зависимость времени безаварийной работы до последней поломки (*y*, в часах) от возраста станка (*x*, в месяцах). За год (12 месяцев наблюдений) были получены следующие результаты:

Марка	*A*	*B*	*C*	*A*	*B*	*C*	*A*	*B*	*C*
x	57	53	62	20	40	50	49	47	65
y	225	284	263	100	176	187	113	244	263

Марка	*A*	*B*	*C*	*A*	*B*	*C*	*A*	*B*	*C*
x	28	30	63	52	69	59	34	55	52
y	206	282	140	119	148	126	168	296	95

Марка	*A*	*B*	*C*	*A*	*B*	*C*	*A*	*B*	*C*
x	54	61	52	26	61	49	37	38	71
y	139	212	144	91	285	172	227	259	253

Марка	*A*	*B*	*C*	*A*	*B*	*C*	*A*	*B*	*C*
x	60	63	52	34	63	52	40	52	73
y	210	266	232	78	293	284	227	280	212

Требуется:

1. Построить уравнение линейной регрессии вида $y = bx + a$ без учета различия станков разных марок.

2. Оценить значимость уравнения регрессии и его параметров.

3. Построить уравнение регрессии, учитывающее различие качества станков разных марок. Для этого в модель ввести две фиктивные (бинарные) переменные – d_1 и d_2, каждая из которых может принимать только значения 0 и 1. Первая из них равна 1 в случае, если станок фирмы A, и 0 в других случаях; другая равна 1 в случае, если станок фирмы B, и 0 в других случаях. Таким образом, для фирмы C обе фиктивные переменные должны быть равны 0. Модель имеет вид

$$y = \beta x + \alpha + d_1 D_1 + \gamma_2 d_2 .$$

4. Оценить значимость параметров полученного уравнения регрессии.

5. Сравнить качество построенных моделей. Сделать выводы о необходимости использования фиктивных переменных, а также о надежности станков разных марок.

Задание 23. Статистические данные о производстве молока с 2012 по 2018 год, поголовье коров и надое молока на 1 корову приведены ниже:

Год	Производство молока, тыс. т	Поголовье коров	Надой молока на 1 корову, кг
2012	31196,8	8657,2	3898
2013	29865,3	8430,9	3893
2014	29995,2	8263,2	4021
2015	29887,5	8115,2	4134
2016	29787,2	7966	4218
2017	30184,5	7950,6	4368
2018	30611,2	7942,6	4492

Требуется:

1. Рассчитать параметры уравнения двухфакторной регрессии вида $y = a + b_1x_1 + b_2x_2$, где y – результативный признак «производство молока»; x_1 – факторный признак «поголовье коров»; x_2 – факторный признак «надой молока».

2. Оценить значимость уравнения регрессии и его параметров. Произвести расчет ошибки аппроксимации.

3. Рассчитать парные и частные коэффициенты корреляции, а также коэффициент множественной корреляции.

4. Оценить влияние факторных признаков x_1 и x_2 на результат.

ГЛАВА 4
СИСТЕМЫ ЭКОНОМЕТРИЧЕСКИХ УРАВНЕНИЙ

Примеры решения задач и выполнения заданий по теме «Системы уравнений, используемых в эконометрике. Структурная и приведенная формы модели. Проблема идентификации. Необходимое и достаточное условия идентификации. Методы оценки параметров структурной формы модели»

Задание 4.1.1. Для структурной формы модели

$$\begin{aligned} y_1 &= b_{12}y_2 + a_{11}x_1 \\ y_2 &= b_{21}y_1 + a_{22}x_2 \end{aligned} \qquad (4.1)$$

записать коэффициенты приведенной формы.

Решение

Для структурной формы приведенная форма модели имеет вид

$$y_1 = \delta_{11}x_1 + \delta_{12}x_2;$$
$$y_2 = \delta_{21}x_1 + \delta_{22}x_2.$$

Из первого уравнения системы (4.1) $y_2 = \dfrac{y_1 - a_{11}x_1}{b_{12}}$. Подставим это значение во второе уравнение системы (4.1). Получим $\dfrac{y_1 - a_{11}x_1}{b_{12}} = b_{21}y_1 + a_{22}x_2$, т.е. $y_1 = \dfrac{a_{11}}{1 - b_{12}b_{21}}x_1 + \dfrac{a_{22}b_{12}}{1 - b_{12}b_{21}}x_2$.

Выразим y_1 из второго уравнения системы (4.1), подставим в первое, получим $y_2 = \dfrac{a_{11}b_{21}}{1 - b_{12}b_{21}}x_1 + \dfrac{a_{22}}{1 - b_{12}b_{21}}x_2$.

Приведенная форма модели имеет вид

$$y_1 = \frac{a_{11}}{1 - b_{12}b_{21}}x_1 + \frac{a_{22}b_{12}}{1 - b_{12}b_{21}}x_2;$$

$$y_2 = \frac{a_{11}b_{21}}{1 - b_{12}b_{21}}x_1 + \frac{a_{22}}{1 - b_{12}b_{21}}x_2.$$

Задание 4.1.2. Оценить структурную модель на идентификацию (проверить необходимое и достаточное условия):

$$\begin{aligned} y_1 &= b_{12}y_2 + b_{13}y_3 + a_{11}x_1 + a_{12}x_2; \\ y_2 &= b_{21}y_1 + a_{22}x_2 + a_{23}x_3 + a_{24}x_4; \\ y_3 &= b_{31}y_1 + b_{32}y_2 + a_{33}x_3 + a_{34}x_4. \end{aligned}$$

Решение

Проверим каждое уравнение системы на идентифицируемость. Для первого уравнения $H = 3$ (y_1, y_2, y_3), $D = 2$ (x_3, x_4), т.е. $D + 1 = H$, следовательно, необходимое условие выполнено. Для проверки достаточности составим следующую таблицу коэффициентов:

Уравнения	Переменные	
	x_3	x_4
2	a_{23}	a_{24}
3	0	a_{34}

Определитель, составленный из матрицы коэффициентов, не равен нулю ($\det A = a_{23} \quad a_{24} \quad 0$), а ранг матрицы равен 2. Следовательно, первое уравнение идентифицируемо.

Для второго уравнения $H = 2$ (y_1, y_2), $D = 1$ (x_1), т.е. $D + 1 = H$, необходимое условие выполнено. Для проверки достаточности составим следующую таблицу коэффициентов:

Уравнения	Переменные	
	y_3	x_1
1	b_{13}	a_{11}
3	–1	0

Определитель, составленный из матрицы коэффициентов, не равен нулю ($\det A = -a_{11} \quad 0$), а ранг матрицы равен 2. Следовательно, второе уравнение идентифицируемо.

Для третьего уравнения $H = 3$ (y_1, y_2, y_3), $D = 2$ (x_1, x_2), т.е. $D + 1 = H$, необходимое условие выполнено. Для проверки достаточности составим следующую таблицу коэффициентов:

Уравнения	Переменные	
	x_1	x_2
1	a_{11}	a_{12}
2	0	a_{22}

Определитель, составленный из матрицы коэффициентов, не равен нулю ($\det A = a_{11} \quad a_{22} \quad 0$), а ранг матрицы равен 2. Следовательно, третье уравнение идентифицируемо.

Таким образом, данная в задании структурная модель является идентифицируемой.

Задание 4.1.3. Оценить структурную модель на идентификацию (проверить необходимое и достаточное условия):

$$y_1 = b_{12}y_2 + a_{11}x_1 + a_{12}x_2;$$
$$y_2 = b_{21}y_1 + b_{23}y_3 + a_{22}x_2;$$
$$y_3 = b_{32}y_2 + a_{31}x_1 + a_{33}x_3.$$

Решение

Проверим каждое уравнение системы на идентифицируемость. Для первого уравнения $H = 2$ (y_1, y_2), $D = 1$ (x_3), т.е. $D + 1 = H$, необходимое условие выполнено. Для проверки достаточности составим следующую таблицу коэффициентов:

Уравнения	Переменные	
	y_3	x_3
2	b_{23}	*0*
3	–1	a_{33}

Определитель, составленный из матрицы коэффициентов, не равен нулю ($\det A = b_{23} \quad a_{33} \quad 0$), а ранг матрицы равен 2. Следовательно, первое уравнение идентифицируемо.

Для второго уравнения $H = 3$ (y_1, y_2, y_3), $D = 2$ (x_1, x_3), т.е. $D + 1 = H$, необходимое условие выполнено. Для проверки достаточности составим следующую таблицу коэффициентов:

Уравнения	Переменные	
	x_1	x_3
1	a_{11}	0
3	a_{31}	a_{33}

Определитель, составленный из матрицы коэффициентов, не равен нулю ($\det A = a_{11} \quad a_{11} \quad 0$), а ранг матрицы равен 2. Следовательно, второе уравнение идентифицируемо.

Для третьего уравнения $H = 2$ (y_2, y_3), $D = 1$ (x_2), т.е. $D + 1 = H$, необходимое условие выполнено. Для проверки достаточности составим следующую таблицу коэффициентов:

Уравнения	Переменные	
	y_1	x_2
1	–1	a_{12}
2	b_{21}	a_{22}

Определитель, составленный из матрицы коэффициентов, $\det A = -a_{22} - a_{12} \; b_{21}$, ранг матрицы равен 2. Следовательно, при $-a_{22} \quad a_{12} \; b_{21}$ третье уравнение идентифицируемо.

Таким образом, при $-a_{22} \quad a_{12} \; b_{21}$ данная в задании структурная модель является идентифицируемой.

Задание 4.1.4. Рассматривается модель денежного рынка [22, с. 122]:

$$\begin{aligned} R_t &= a_1 + b_{11}M_t + b_{12}Y_t + \varepsilon_1; \\ Y_t &= a_2 + b_{33}R_t + b_{22}I_t + \varepsilon_2, \end{aligned}$$

где R – процентные ставки; Y – ВВП; I – число удовлетворенных прошений; M – денежная масса; t – текущий период; a_1, a_2 – свободные члены уравнений.

Требуется проверить данную модель на идентификацию, определить, каким методом могут быть рассчитаны ее коэффициенты (в случае если модель сверх- или точно идентифицируема). Записать приведенную форму модели.

Решение

Модель содержит две эндогенные переменные (левые части уравнений) – R_t, Y_t, две экзогенные – M_t, I_t (в правой части первого уравнения). Проверим каждое уравнение на идентификацию.

Уравнение I. В этом уравнении присутствуют две эндогенные переменные (R_t, Y_t), отсутствует одна экзогенная переменная (I_t), поэтому $H = 2$, $D = 1$, необходимое условие идентификации выполняется, поскольку $D + 1 = H$. Первое уравнение идентифицируемо.

Уравнение II. В этом уравнении имеются две эндогенные переменные (R_t, Y_t), отсутствует одна экзогенная переменная (M_t), поэтому $H = 2$, $D = 1$, $D + 1 = H$, и второе уравнение по необходимому условию является идентифицируемым.

Проверим каждое уравнение на выполнение достаточного условия идентификации. Запишем расширенную матрицу системы:

Уравнение	R_t	Y_t	M_t	I_t
I	–1	b_{12}	b_{11}	0
II	b_{33}	–1	0	b_{12}

Достаточное условие идентификации для соответствующего уравнения будет выполнено, если определитель подматрицы, построенной только из коэффициентов при переменных, отсутствующих в этом уравнении, не равен нулю, а ранг равен количеству эндогенных переменных в системе минус единица.

Первому уравнению соответствует первая строка расширенной матрицы, поэтому первую строку не следует включать в подматрицу. Из остальной части расширенной матрицы оставим только столбцы, которые имеют нули в первой строке. Получаем подматрицу $\left[b_{12}\right]$. Определитель подматрицы не равен нулю ($\det A = b_{12} \quad 0$). Ранг подматрицы равен единице, т.е. числу эндогенных переменных в системе минус единица. Достаточное условие идентификации для первого уравнения выполнено.

Подматрица для второго уравнения имеет вид $\left[b_{11}\right]$. Ее определитель не равен нулю ($\det A = b_{11} \quad 0$). Ранг подматрицы равен единице (числу эндогенных переменных в системе минус единица).

Таким образом, достаточное условие идентификации выполнено для всех уравнений системы. Модель в целом идентифицирована. Для определения параметров уравнений может быть применен обычный МНК.

Приведенная форма модели имеет вид

$$R_t = A_1 + A_2 Y_t + A_3 M_t + A_4 I_t + V_1;$$
$$Y_t = B_1 + B_2 R_t + B_3 M_t + B_4 I_t + V_2,$$

где V_1, V_2 – случайные члены.

Как обычно, в правой части приведенной формы стоят только предопределенные переменные.

Задание 4.1.5. 1. Проверить на идентификацию структурную модель

$$\begin{cases} y_1 = b_{12} \quad y_2 + a_{11} \quad x_1; \\ y_2 = b_{21} \quad y_1 + a_{22} \quad x_2. \end{cases} \qquad (4.2)$$

2. Из известной приведенной формы модели уравнений

$$\begin{cases} y_1 = 3 \quad x_1 + 4 \quad x_2 \\ y_2 = 5 \quad x_1 - 6 \quad x_2 \end{cases} \qquad (4.3)$$

определить структурные коэффициенты модели.

Решение

1. В структурной модели (4.2) имеются две эндогенные (y_1, y_2) и две экзогенные (x_1, x_2) переменные.

Проверим уравнения системы (4.2) на необходимое и достаточное условия идентификации.

В первом уравнении системы (4.2) число эндогенных переменных H = 1 (y_1), число отсутствующих экзогенных D = 0. D + 1 = H, т.е. выполняется необходимое условие идентификации.

Проверим достаточное условие идентификации.

В первом уравнении отсутствует y_2. Построим матрицу из коэффициента при y_2 во втором уравнении системы:

Уравнение	Отсутствующие переменные
Первое	y_2
Второе	–1

$DetA = -1 \neq 0$.

Определитель матрицы не равен 0, ранг матрицы равен 1. Следовательно, выполняется достаточное условие идентификации (ранг этой матрицы не менее числа эндогенных переменных системы без единицы) и первое уравнение точно идентифицируемо.

Во втором уравнении системы (4.2) число эндогенных переменных $H = 1$ (y_1), число отсутствующих экзогенных $D = 0$. $D + 1 = H$, т.е. выполняется необходимое условие идентификации.

Проверим достаточное условие идентификации. Во втором уравнении отсутствует y_1. Построим матрицу из коэффициента при y_2 во втором уравнении системы:

Уравнение	Отсутствующие переменные
Первое	y_1
Второе	–1

$DetA = -1 \neq 0$.

Определитель матрицы не равен 0, ранг матрицы равен 1, следовательно, выполняется достаточное условие идентификации и второе уравнение точно идентифицируемо. Следовательно, исследуемая система точно идентифицируема и может быть решена косвенным МНК.

2. Определим структурные коэффициенты модели.

На основе решения задания 4.1.1 приведенная форма модели

$$\begin{cases} y_1 = b_{12} \ \ y_2 + a_{11} \ \ x_1 \\ y_2 = b_{21} \ \ y_1 + a_{22} \ \ x_2 \end{cases}$$

имеет вид

$$y_1 = \frac{a_{11}}{1 - b_{12}b_{21}} x_1 + \frac{a_{22}b_{12}}{1 - b_{12}b_{21}} x_2;$$

$$y_2 = \frac{a_{11}b_{21}}{1 - b_{12}b_{21}} x_1 + \frac{a_{22}}{1 - b_{12}b_{21}} x_2.$$

По условию задания 4.1.5 она равна

$$\begin{cases} y_1 = 3 \ \ x_1 + 4 \ \ x_2; \\ y_2 = 5 \ \ x_1 - 6 \ \ x_2, \end{cases}$$

т.е.

$$\frac{a_{11}}{1-b_{12}b_{21}}=3; \tag{4.4}$$

$$\frac{a_{22}b_{12}}{1-b_{12}b_{21}}=4; \tag{4.5}$$

$$\frac{a_{11}b_{21}}{1-b_{12}b_{21}}=5; \tag{4.6}$$

$$\frac{a_{22}}{1-b_{12}b_{21}}=6. \tag{4.7}$$

Тогда, приравняв выражение (4.4) к формуле (4.5), а (4.6) – к (4.7), имеем

$$\frac{1}{3}a_{11}=\frac{1}{4}a_{22}b_{12},\ \frac{1}{6}a_{22}=\frac{1}{5}a_{11}b_{21},$$

откуда $b_{12}b_{21}=\frac{10}{9}$. Получаем

$$a_{11}=-27,\ a_{22}=-54,\ b_{12}=\frac{2}{3},\ b_{21}=\frac{5}{3}.$$

Таким образом, структурная форма модели имеет вид

$$y_1=\frac{2}{3}\ y_2-27\ \ x_1;$$
$$y_2=\frac{5}{3}\ y_1-54\ \ x_2.$$

Задание 4.1.6. Определить, является ли следующая система системой независимых уравнений:

$$\begin{aligned}
y_1 &= a_{01} + a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + a_{14}x_4 + e_1;\\
y_2 &= a_{02} + a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + a_{24}x_4 + e_2;\\
y_3 &= a_{03} + a_{31}x_1 + a_{32}x_2 + a_{33}x_3 + a_{34}x_4 + e_3;\\
y_4 &= a_{04} + a_{41}x_1 + a_{42}x_2 + a_{43}x_3 + a_{44}x_4 + e_4.
\end{aligned}$$

Решение

Матрица параметров при зависимых переменных указанной системы имеет вид

$$\begin{vmatrix} 1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1 \end{vmatrix}.$$

Если матрица параметров диагональная, то рассматриваемая модель является системой независимых уравнений.

Задания для самостоятельного выполнения

Задание 1. Проверить на идентификацию статическую модель Кейнса народного хозяйства:

$$C = a + by + \varepsilon;$$
$$y = C + I,$$

где C – личное потребление; y – национальный доход в постоянных ценах; I – инвестиции в постоянных ценах.

Задание 2. Оценить следующие структурные модели на идентификацию:

$$\begin{aligned} y_1 &= a_1 + a_{11}x_1 + a_{12}x_2 + в_{12}y_2; \\ y_2 &= a_2 + a_{22}x_2 + a_{23}x_3 + в_{21}y_1; \\ y_3 &= a_3 + a_{31}x_1 + a_{32}x_3. \end{aligned} \qquad \begin{aligned} y_1 &= b_{11}y_2 + b_{13}y_3 + a_{12}x_2; \\ y_2 &= b_{21}y_1 + a_{22}x_2 + a_{23}x_3 + a_{24}x_4; \\ y_3 &= b_{31}y_1 + b_{32}y_2 + a_{31}x_1 + a_{32}x_2. \end{aligned}$$

$$\begin{aligned} y_1 &= b_{13}y_3 + a_{11}x_1 + a_{13}x_3; \\ y_2 &= b_{21}y_1 + b_{23}y_3 + a_{22}x_2; \\ y_3 &= b_{32}y_2 + a_{31}x_1 + a_{33}x_3. \end{aligned} \qquad \begin{aligned} y_1 &= b_{12}y_2 + a_{11}x_1 + a_{12}x_2; \\ y_2 &= b_{21}y_1 + b_{23}y_3 + a_{22}x_2; \\ y_3 &= b_{32}y_2 + a_{31}x_1 + a_{33}x_3. \end{aligned}$$

$$\begin{aligned} y_1 &= b_{12}y_2 + b_{13}y_3 + a_{11}x_1 + a_{12}x_2; \\ y_2 &= b_{21}y_1 + b_{23}y_3 + a_{22}x_2; \\ y_3 &= b_{31}y_1 + b_{32}y_2 + a_{31}x_1 + a_{33}x_3. \end{aligned} \qquad \begin{aligned} y_1 &= \mathrm{a}_{11}x_1 + b_{12}y_2 + \mathrm{b}_{13}y_3; \\ y_2 &= a_{22}x_2 + a_{24}x_4 + \mathrm{b}_{21}y_1; \\ y_3 &= \mathrm{a}_{31}x_2 + a_{32}x_3. \end{aligned}$$

$$\begin{aligned} y_1 &= b_{12}y_2 + a_{11}x_1; \\ y_2 &= b_{21}y_1 + a_{22}x_2. \end{aligned} \qquad \begin{aligned} y_1 &= b_{12}y_2 + b_{13}y_3 + a_{11}x_1 + a_{12}x_2; \\ y_2 &= b_{21}y_1 + a_{22}x_2 + a_{23}x_3; \\ y_3 &= b_{31}y_1 + b_{32}y_2 + a_{34}x_4. \end{aligned}$$

Задание 3. Проверить на идентификацию модель вида

$$\begin{aligned} y_1 &= b_{13}\ y_3 + a_{11}\ x_1 + a_{13}\ x_3 \\ y_2 &= b_{21}\ y_1 + b_{23}\ y_3 + a_{22}\ x_2 \\ y_3 &= b_{32}\ y_2 + a_{31}\ x_1 + a_{33}\ x_3 \end{aligned}$$

для приведенной формы модели

$$\begin{aligned} y_1 &= 2\ x_1 + 4\ x_2 + 10\ x_3 \\ y_2 &= 3\ x_1 - 6\ x_2 + 2\ x_3 \\ y_3 &= -5\ x_1 + 8\ x_2 + 5\ x_3, \end{aligned}$$

определить структурные коэффициенты модели.

ГЛАВА 5
МОДЕЛИ ВРЕМЕННЫХ РЯДОВ

Примеры решения задач и выполнения заданий по теме «Основные элементы временного ряда. Моделирование тенденции временного ряда. Модели с распределенными лагами»

Задание 5.1.1. Имеются данные об объемах потребления электроэнергии:

		$\tau = 1$	$\tau = 2$	$\tau = 3$
t	y_t	y_{t-1}	y_{t-2}	y_{t-3}
1	7	–	–	–
2	7	7	–	–
3	8	7	7	–
4	10	8	7	7
5	11	10	8	7
6	12	11	10	8
7	13	12	11	10
8	16	13	12	11

Рассчитать коэффициенты автокорреляции уровней для временного ряда.

Решение

По формулам $\bar{y}_1 = \dfrac{\sum_{t=2}^{n} y_t}{n-1}$ и $\bar{y}_2 = \dfrac{\sum_{t=2}^{n} y_{t-1}}{n-1}$ вычислим средние величины:

$$\bar{y}_1 = \frac{\sum_{t=2}^{n} y_t}{n-1} = \frac{7+8+10+11+12+13+16}{7} = 11;$$

$$\bar{y}_2 = \frac{\sum_{t=2}^{n} y_{t-1}}{n-1} = \frac{7+7+8+10+11+12+13}{7} = 9{,}7.$$

По формуле $r_1 = \dfrac{\sum_{t=2}^{n} (y_t - \bar{y}_1)(y_{t-1} - \bar{y}_2)}{\sqrt{\sum_{t=2}^{n} (y_t - \bar{y}_1)^2 \sum_{t=2}^{n} (y_{t-1} - \bar{y}_2)^2}}$ найдем коэффициенты автокорреляции для уровней ряда первого порядка:

$$r_1 = \frac{(7-11)\ (7-9{,}7) + \ ... \ + (16-11)\ (13-9{,}7)}{\sqrt{[(7-11)^2 + \ ... \ + (16-11)^2]\ [(7-9{,}7)^2 + \ ... \ + (13-9{,}7)^2]}} = \frac{43}{44{,}5} = 0{,}97.$$

По формулам $\bar{y}_3 = \dfrac{\sum\limits_{t=3}^{n} y_t}{n-2}$ и $\bar{y}_4 = \dfrac{\sum\limits_{t=3}^{n} y_{t-2}}{n-2}$ вычислим средние величины:

$$\bar{y}_3 = \frac{\sum\limits_{t=3}^{n} y_t}{n-2} = \frac{8+10+11+12+13+16}{6} = 11{,}7;$$

$$\bar{y}_4 = \frac{\sum\limits_{t=3}^{n} y_{t-2}}{n-2} = \frac{7+7+8+10+11+12}{6} = 9{,}2.$$

По формуле $r_2 = \dfrac{\sum\limits_{t=3}^{n} (y_t - \bar{y}_3)(y_{t-2} - \bar{y}_4)}{\sqrt{\sum\limits_{t=3}^{n} (y_t - \bar{y}_3)^2 \sum\limits_{t=2}^{n} (y_{t-2} - \bar{y}_4)^2}}$ рассчитаем коэффициенты автокорреляции для уровней ряда второго порядка:

$$r_2 = \frac{(8-11{,}7)\ (7-9{,}2) + ... + (16-11{,}7)\ (12-9{,}2)}{\sqrt{[(8-11{,}7)^2 + ... + (16-11{,}7)^2]\ [(7-9{,}2)^2 + ... + (12-9{,}2)^2]}} = \frac{27{,}33}{29{,}2} = 0{,}94.$$

Вычислим:

$$\bar{y}_5 = \frac{\sum\limits_{t=4}^{n} y_t}{n-3} = \frac{10+11+12+13+16}{5} = 12{,}4; \quad \bar{y}_6 = \frac{\sum\limits_{t=4}^{n} y_{t-3}}{n-3} = \frac{7+7+8+10+11}{5} = 8{,}6;$$

$$r_3 = \frac{15{,}8}{16{,}73} = 0{,}94.$$

Наибольшее значение коэффициентов автокорреляции первого порядка свидетельствует о наличии сильной линейной тенденции.

Задание 5.1.2. По результатам наблюдений деятельности магазина получена зависимость объема продаж за месяц от количества потенциальных покупателей. Зависимость имеет вид модели с распределенным лагом:

$$y_t = 4{,}25 + 2{,}87x_t + 0{,}88x_{t-1} + 9{,}52x_{t-2} + 2{,}6x_{t-3}.$$

Требуется определить краткосрочный и долгосрочный мультипликаторы.

Решение

Коэффициент регрессии b_0 = 2,87. Он показывает, что увеличение числа потенциальных покупателей на одного человека приведет к увеличению объема продаж на 2,87 ден. ед. Долгосрочный мультипликатор равен b = 2,87 + 0,88 + 9,52 + 2,6 = 15,87. Это означает, что увеличение в настоящее время числа потенциальных покупателей на одного человека приведет через 3 месяца к увеличению объема продаж на 15,87 ден. ед.

Задание 5.1.3. Статистические данные о доходе на одного члена семьи и расходах на покупку товара *T* показаны ниже:

Показатель	Год						
	2015	2016	2017	2018	2019	2020	2021
Доход на одного члена семьи (% к 2015 году)	100	103	105	107	110	112	115
Расходы на покупку товара *T* (тыс. руб.)	30	33	37	39	44	47	50

Требуется:

1. Найти ежегодные абсолютные приросты доходов и расходов.
2. Сделать выводы о наличии тенденции по доходам и расходам.
3. На основе первых разностей уравнений исходных динамических рядов составить линейную регрессионную модель спроса.
4. Построить линейную модель спроса на товар *T*, включив в нее фактор времени.

Решение

Последовательность решения:

1. Обозначим *T* через y, а доходы одного члена семьи – через x.

Ежегодные абсолютные приросты вычислим по формулам:

$$\Delta y_t = y_t - y_{t-1};\;\; \Delta x_t = x_t - x_{t-1}.$$

Получим

y_t	Δy_t	x_t	Δx_t
30	–	100	–
33	3	103	3
37	4	105	2
39	2	107	2
44	5	110	3
47	3	112	2
50	3	115	3

2. Значения абсолютных приростов Δy и Δx не имеют четко выраженной тенденции, они изменяются вокруг средних уровней. Это свидетельствует о равномерном развитии процессов в обоих рядах динамики, т.е. наличии линейных тенденций (линейных трендов).

3. Модель на основе первых разностей уравнений исходных динамических рядов имеет вид

$$\Delta\hat{y} = a + b\cdot\Delta x.$$

Для определения параметров a и b применим МНК. Система нормальных уравнений имеет вид

$$\Sigma\Delta y = n\ \ a + b\ \ \Sigma\Delta x;$$

$$\Sigma\Delta y\Delta x = a\ \ \Sigma\Delta x + b\ \ \Sigma(\Delta x)^2.$$

Для исходных данных

$$20 = 6\ \ a + 15\ \ b;$$

$$51 = 15\ \ a + 39\ \ b.$$

Решая эту систему по формулам Крамера, получим:

$$\Delta = 9;\ \Delta_1 = 15;\ \Delta_2 = 15.$$

Тогда $a = \frac{\Delta_1}{\Delta} = \frac{15}{9} = \frac{5}{3} \approx 1,67;\ b = \frac{\Delta_2}{\Delta} = \frac{6}{9} = \frac{2}{3} \approx 0,67;$ модель имеет вид

$$\Delta\hat{y} = 1,67 + 0,67\ \ \Delta x.$$

4. Построим модель по исходным данным, включив в нее в качестве самостоятельного фактора время, т.е. модель вида $\hat{y}_t = f(x, t)$.

Линейная двухфакторная модель будет

$$\hat{y}\ = a + b\cdot x + c\cdot t.$$

Для данной модели запишем систему нормальных уравнений:

$$\Sigma y = n\ \ a + b\ \ \Sigma x + c\ \ \Sigma t;$$

$$\Sigma yx = a\ \ \Sigma x + b\ \ \Sigma x^2 + c\ \ \Sigma x\ \ t;$$

$$\Sigma yt = a\ \ \Sigma t + b\ \ \Sigma x + c\ \ \Sigma t^2.$$

Вспомогательные расчеты приведены ниже:

	t	y	x	yx	y_t	x_t	x^2	t^2
	1	30	100	3000	30	100	10000	1
	2	33	103	3399	66	206	10609	4
	3	37	105	3885	111	315	11025	9
	4	39	107	4173	156	428	11449	16
	5	44	110	4840	220	550	12100	25
	6	47	112	5264	282	672	12544	36
	7	50	115	5750	350	805	13225	49
Сумма	28	280	752	30311	1215	3076	80952	140

Система нормальных уравнений будет

$$280 = 7\ a + 752\ b + 28\ c;$$

$$30311 = 752\ a + 80952\ b + 3076\ c;$$

$$1215 = 28\ a + 3076\ b + 140\ c.$$

Решая ее, получим параметры уравнения регрессии:

$$a = -22{,}43;\ b = 0{,}5;\ c = 2{,}18.$$

Уравнение регрессии имеет вид

$$\hat{y} = -22{,}43 + 0{,}5\ x + 2{,}18\ t.$$

Параметр b = 0,5 характеризует силу связи признаков y и x. Он показывает, что с ростом дохода на одного члена семьи на 1 % (при неизменной тенденции) расходы на товар T возрастают в среднем на 0,5 руб. Параметр c = 2,18 характеризует среднегодовой абсолютный прирост расходов на товар T под воздействием прочих факторов (при условии неизменного дохода).

Задание 5.1.4. Изучается зависимость между числом рабочих фирмы (x) и объемом выпуска продукции (y) в 8 периодах наблюдений. Данные о результатах исследования приведены ниже:

Период наблюдений (мес.)	1	2	3	4	5	6	7	8
Число рабочих x	10	12	11	12	14	15	17	20
Объем выпуска продукции y (усл. ед.)	7	8	8	10	11	12	14	16

Построить уравнение линейной регрессии. Проверить для него гипотезу о наличии автокорреляции в остатках по критерию Дарбина – Уотсона на уровне значимости α = 0,05.

Для построения графиков упорядочим наблюдения по переменной x. Вычисления приведены ниже:

	x	y	xy	x^2	$\hat{y}$
	10	7	70	100	6,842105
	11	9	99	121	7,957326
	12	8	96	144	9,072546
	14	11	154	196	11,30299
	15	12	180	225	12,41821
	17	15	255	289	14,64865
	18	16	288	324	15,76387
	20	18	360	400	17,99431
Среднее	14,625	12	187,75	224,875	–

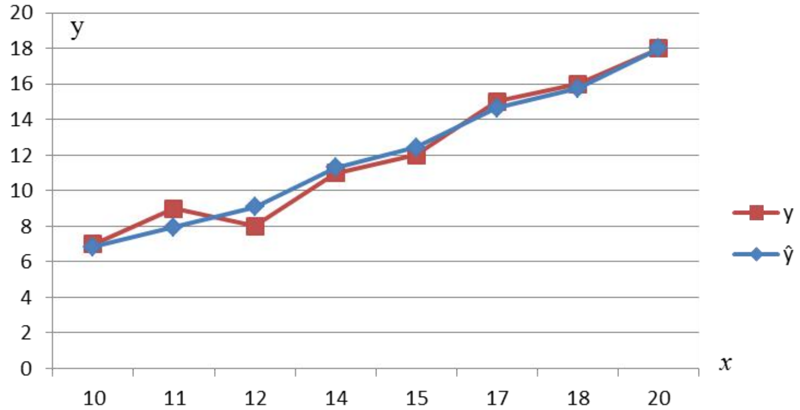
Параметры уравнения регрессии:

$$b = \frac{\overline{xy} - \overline{x}\ \overline{y}}{\overline{x^2} - \overline{x}^2} = \frac{187{,}75 - 14{,}625\ 12}{224{,}875\ 14{,}625^2} = 1{,}115;$$

$$a = \overline{y} - b\overline{x} = 12 - 1{,}115\ 14{,}625 = -4{,}31.$$

Искомое уравнение регрессии будет $\hat{y} = 1{,}115x - 4{,}31 + \varepsilon_t$, $\varepsilon_t = y - \hat{y}$.

Исходные наблюдения и линия регрессии показаны на графике, данном ниже:



Исходные данные и результаты промежуточных расчетов для проверки выполнения критерия Дарбина – Уотсона приведены ниже:

Период наблю-дений (мес.)	1	2	3	4	5	6	7	8
$\hat{y}$	6,842	7,957	9,073	11,303	12,418	14,649	15,764	17,994
ε_t	0,158	1,043	–1,073	–0,303	–0,418	0,351	0,236	0,006
$\varepsilon_t - \varepsilon_{t-1}$	–	0,885	–2,115	0,770	–0,115	0,770	–0,115	–0,230
ε_t^2	0,025	1,087	1,150	0,092	0,175	0,123	0,056	0,000
$(\varepsilon_t - \varepsilon_{t-1})^2$	–	0,783	4,474	0,592	0,013	0,592	0,013	0,053

$$\sum(\varepsilon_t)^2 = 2{,}708; \sum(\varepsilon_t - \varepsilon_{t-1})^2 = 6{,}521.$$

Имеем $d = \dfrac{\sum_{t=2}^{n}(\varepsilon_t - \varepsilon_{t-1})^2}{\sum_{t=1}^{n}\varepsilon_t^2} = \dfrac{(e_t - e_{t-1})^2}{(e_t)^2} = \dfrac{6{,}521}{2{,}708} = 2{,}408.$

По табл. 3 прилож. для $n = 8$ и $k = 1$ (однофакторная модель) находим критические значения $d_l = 0{,}76$, $d_u = 1{,}33$.

Так как в нашем случае $1{,}33 \le 2{,}408 \le 4 - 1{,}33 = 2{,}67$, автокорреляция остатков отсутствует.

Задания для самостоятельного выполнения

Задание 1. Модель с распределенным лагом имеет вид:

1) $y_t = 41{,}6 + 2{,}32x_t + 0{,}96x_{t-1} + 0{,}72x_{t-2} + 1{,}6x_{t-3}$;
2) $y_t = 65 + 2{,}78x_t + 0{,}345x_{t-1} + 7{,}2x_{t-2} + 2{,}55x_{t-3}$;
3) $y_t = 45 + 25{,}32x_t + 0{,}196x_{t-1} + 2{,}2x_{t-2} + 1{,}56x_{t-3}$;
4) $y_t = 3{,}6 + 2{,}35x_t + 0{,}94x_{t-1} + 0{,}52x_{t-2} + 2{,}6x_{t-3}$;
5) $y_t = 41{,}6 + 2{,}32x_t + 0{,}96x_{t-1} + 0{,}72x_{t-2} + 1{,}6x_{t-3}$;
6) $y_t = 45{,}6 + 2{,}37x_t + 0{,}66x_{t-1} + 0{,}82x_{t-2} + 1{,}69x_{t-3}$;
7) $y_t = 4{,}25 + 2{,}87x_t + 0{,}88x_{t-1} + 9{,}52x_{t-2} + 2{,}6x_{t-3}$.

Определить краткосрочный, промежуточные и долгосрочный мультипликаторы.

Задание 2. По результатам изучения зависимости удельных постоянных затрат от инвестиций в НИОКР для некоторого вида продукции администрация компании получила следующую модель (по данным за последние 38 лет):

$$y_t = 231 - 0{,}2x_{t-1} - 0{,}15x_{t-2} - 0{,}5x_{t-3}.$$

Какова структура лага в данной модели? Дать интерпретацию параметров этой модели.

Задание 3. Данные об уровне инфляции y_t (%) за 5 месяцев представлены ниже:

Месяц	1	2	3	4	5
y_t	3	4	5	6	7

Определить коэффициент автокорреляции уровней ряда первого порядка.

Задание 4. Данные об уровне y_t за 8 месяцев представлены ниже:

Номер месяца	1	2	3	4	5	6	7	8
y_t	87	9	8	7	8	6	6	5

Определить коэффициент автокорреляции уровней этого ряда первого порядка.

Задание 5. Имеется временной ряд

t	1	2	3	4	5	6	7	8
y_t	20	…	…	…	…	…	…	10

Известно также, что $x_t = 150$, $x_t^2 = 8100$, $\sum\limits_{t=2}^{n} x_t x_{t-1} = 7350$.

Определить коэффициент автокорреляции уровней ряда первого порядка. Установить, имеет ли исследуемый временной ряд тенденцию.

Задание 6. Данные об объеме производства продукции предприятия (усл. ден. ед.) приведены ниже:

Год	Квартал	Объем	Год	Квартал	Объем	Год	Квартал	Объем	Год	Квартал	Объем
2015	I	3,78	2016	I	4,78	2017	I	5,07	2018	I	5,12
	II	5,16		II	5,56		II	5,85		II	6,04
	III	4,94		III	5,15		III	5,9		III	6,11
	IV	5,95		IV	6,19		IV	6,25		IV	6,75

Требуется:

1. Определить наличие тенденции временного ряда, описывающего объем производства промышленного предприятия.
2. Проверить на значимость коэффициенты уравнения регрессии.
3. Проверить качество уравнения с помощью коэффициента детерминации.

Задание 7. Статистические данные (8 наблюдений) о зависимости капитальных затрат на ремонт оборудования Y от срока службы оборудования X приведены ниже:

Номер наблюдения	1	2	3	4	5	6	7	8
Y (усл. ед.)	90	125	180	110	120	102	250	82
X (месяц)	4	6	8	9	10	11	12	3

Требуется:

1. Определить абсолютные приросты капитальных затрат от срока службы оборудования, сделать выводы о наличии тенденции развития каждого ряда.
2. Построить линейную модель затрат, используя первые разности уравнений исходных динамических рядов.
3. Пояснить экономический смысл коэффициента регрессии.
4. Построить линейную модель затрат, включив в нее фактор времени.

Задание 8. Исследуется успеваемость в вузе (зависимость между результатами зимней X и летней Y сессий). Средние оценки по всем предметам зимней и летней сессий, полученные 20 студентами группы, приведены ниже:

Студент	X	Y	Студент	X	Y
1	3,6	4,9	11	4,3	3,6
2	3,4	3,7	12	4,6	4,5
3	4,5	5	13	4,6	4,8
4	3	4	14	3,3	3,3
5	4,7	4,2	15	4,3	4,6
6	4,8	5	16	3,1	3,8
7	3,9	4,9	17	3,2	3
8	3,5	3,6	18	4,2	4,3
9	3,2	4,2	19	3,3	3,4
10	3,9	3,1	20	3,5	3,2

Построить уравнение линейной регрессии зависимости результатов летней Y сессии от зимней X. Проверить для него по критерию Дарбина – Уотсона гипотезу о наличии автокорреляции в остатках на уровне значимости $\alpha = 0{,}05$.

Задание 9. По данным за 12 месяцев построено уравнение регрессии зависимости прибыли предприятия y (млн руб.) от цен на электроэнергию x_1 (тыс. руб. за 1 кВт) и производительности труда x_2 (ед. продукции на 1 работника):

$$\hat{y} = 300 - 1{,}8 \cdot x_1 + 5 \cdot x_2.$$

При анализе остаточных величин были использованы значения, приведенные ниже:

№ п/п	y	x_1	x_2
1	315	850	400
2	825	1100	600
3	350	1550	800

$$\Sigma\varepsilon_t^2 = 10500, \ \Sigma(\varepsilon_t - \varepsilon_{t-1})2 = 50000.$$

Требуется:

1. По трем позициям рассчитать $\hat{y}_t$, ε_t, ε_{t-1}, ε_t^2, $(\varepsilon_t - \varepsilon_{t-1})$.
2. Рассчитать критерий Дарбина – Уотсона.
3. Оценить полученный результат при 5%-м уровне значимости.
4. Указать, пригодно ли уравнение для прогноза.

ЗАКЛЮЧЕНИЕ

Эконометрические модели и методы находят широкое применение в промышленности и различных сферах деятельности человека. К таким сферам принадлежат:

1) банковские услуги (необходимы для оценки кредитоспособности клиентов и оптимизации вложения капиталов);

2) управление страхованием (нужны для оптимальной согласованности функций полезности страхователя и страховщика);

3) пенсионное обеспечение (используются с целью контроля отчисления части зарплаты, направляемой в пенсионный фонд);

4) управление транспортом и перевозками (требуются для определения кратчайших и наименее затратных маршрутов);

5) логистика, управление поставками и складское хозяйство (используются при планировании объемов поставок и сроков с целью минимизации складских площадей и цены аренды);

6) розничная торговля (задействуются для расстановки товаров таким образом, чтобы увеличивались продажи);

7) реклама (применяются в анализе методов подбора и распространения рекламной информации с целью охвата максимального количества потенциальных клиентов при минимальных затратах);

8) медицина (с помощью указанных моделей и методов сравнивается эффективность препаратов и методов лечения, устанавливаются возможные побочные эффекты);

9) сельское хозяйство (используются для прогнозирования урожайности, эффективности внесения удобрений, управления посевными площадями, севооборотом);

10) пищевая промышленность (применяются для прогнозирования объемов и качества выпускаемой продукции).

Рассмотренные модели и методы задействуются также при контроле и управлении производственными и технологическими процессами; разработке эффективных методов выборочного приемочного контроля; оптимизации производства изделий методами планирования эксперимента. Кроме того, они используются для повышения качества, надежности выпускаемой продукции и ее сертификации, диагностики дефектов материалов, изучения функции полезности потребителей в маркетинговых исследованиях.

Успешное освоение студентами дисциплины «Эконометрика» будет способствовать получению ими положительных результатов в профессиональной деятельности.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Абдуллин Р.З. Эконометрика в MS Excel: учебное пособие. Иркутск: БГУ, 2016. 135 с.
2. Агаларов З.С., Орлов А.И. Эконометрика: учебник. М.: Дашков и К°, 2023. 380 с.
3. Айвазян С.А. Методы эконометрики: учебник. М.: Магистр: ИНФРА-М, 2022. 512 с.
4. Анатольев С.А. Задачи и решения по эконометрике. М.: Российская экономическая школа, 2005. 164 с.
5. Артамонов Н.В. Введение в эконометрику: учебник. М.: МЦНМО, 2014. 222 с.
6. Афанасьев В.Н., Цыпкин А.П. Эконометрика в пакете STATISTICA: учебное пособие по выполнению лабораторных работ. Оренбург: ГОУ ОГУ, 2008. 204 с.
7. Бабешко Л.О., Бич М.Г., Орлова И.В. Эконометрика и эконометрическое моделирование: учебник. М.: ИНФРА-М, 2023. 387 с.
8. Балдин К.В., Быстров О.Ф., Соколов М.М. Эконометрика: учебное пособие для вузов. М.: ЮНИТИ-ДАНА, 2017. 254 с.
9. Борзых Д.А., Демешев Б.Б. Эконометрика в задачах и упражнениях. М.: URSS, 2017. 304 с.
10. Бородич С.А. Эконометрика. Практикум: учебное пособие. М.: ИНФРА-М, 2021. 329 с.
11. Булдык Г.М. Сборник задач и упражнений по эконометрике и экономико-математическим методам и моделям: учебно-методические рекомендации: в 3 ч. Минск: БИП, 2016. Ч. 1. 92 с.
12. Буравлев А.И. Эконометрика: учебное пособие. М.: Лаборатория знаний, 2021. 167 с.
13. Валентинов В.А. Эконометрика: практикум. Минск – М.: Новое знание: ИНФРА-М, 2016. 436 с.
14. Ганичев А.В., Ганичева А.В. Эконометрика: учебное пособие. Тверь: ТвГТУ, 2019. 144 с.
15. Гладилин А.В., Герасимов А.Н., Громов Е.И. Практикум по эконометрике: учебное пособие. Ростов н/Д.: Феникс, 2011. 326 с.
16. Грибанова Е.Б. Эконометрика: учебное пособие. Томск: ТУСУР, 2014. 156 с.
17. Григорьева С.В., Ильина Е.А. Сборник задач по эконометрике. Чебоксары: Волжский филиал МАДИ, 2012. 72 с.

18. Демидова О.А., Малахов Д.И. Эконометрика: учебник и практикум для прикладного бакалавриата. М.: Юрайт, 2019. 334 с.
19. Дорохина Е.Ю., Преснякова Л.Ф., Тихомиров Н.П. Сборник задач по эконометрике: учебное пособие для студентов экономических вузов. М.: Экзамен, 2003. 224 с.
20. Дэвидсон Р., Мак-Киннон Дж. Теория и методы эконометрики: учебник. М.: Дело, 2018. 936 с.
21. Евсеев Е.А., Буре В.М. Эконометрика: учебное пособие. М.: Юрайт, 2018. 186 с.
22. Заяц О.А. Эконометрика: учебное пособие. Волгоград: Волгоградский ГАУ, 2021. 140 с.
23. Ивин Е.А., Мироненков А.А., Шаклеина М.В. Практикум по эконометрике. Примеры реализации в пакетах Excel и R: учебное пособие для вузов. Вологда: ВолНЦ РАН, 2021. 104 с.
24. Игнашева Т.А. Методы прогнозирования социально-экономических процессов: учебное пособие. Йошкар-Ола: Волгатех, 2018. 104 с.
25. Каморников С.Ф., Каморников С.С. Эконометрика: учебное пособие. М.: Интеграция, 2012. 262 с.
26. Карпенко Н.В. Эконометрика. Анализ и прогнозирование временного ряда: учебное пособие. М.: РУТ (МИИТ), 2018. 132 с.
27. Козинова А.Т. Практикум по эконометрике: учебное пособие. Н. Новгород: Нижегородский госуниверситет, 2011. 96 с.
28. Колемаев В.А. Эконометрика: учебник. М.: ИНФРА-М, 2017. 160 с.
29. Комарова Е.С. Парный регрессионный анализ: учебное пособие. М. – Берлин: Директ-Медиа, 2019. 59 с.
30. Кундышева Е.С., Суслаков Б.А. Математические методы и модели в экономике: учебник. М.: Дашков и К°, 2023. 286 с.
31. Литвинова И.А. Эконометрика: учебное пособие. Кемерово: КемГУ, 2016. 112 с.
32. Лопатнюк Л.А., Марков А.С., Подашевская Е.И. Эконометрика и экономико-математические методы и модели. Практикум: учебно-методическое пособие. Минск: БГАТУ, 2019. 176 с.
33. Луговская Л.В. Эконометрика в вопросах и ответах: учебное пособие. М.: ТК «Велби»: Проспект, 2006. 208 с.
34. Марченко В.М., Можей Н.П., Шинкевич Е.А. Эконометрика и экономико-математические методы и модели: в 2 ч. Ч. 1. Эконометрика: учебное пособие. Минск: БГТУ, 2011. 157 с.
35. Молодых В.А., Рубежной А.А., Сосин А.И. Эконометрика. Практикум: учебное пособие. Ставрополь: СКФУ, 2017. 179 с.

36. Молотникова А.А. Основы эконометрики: учебное пособие. СПб.: Лань, 2018. 168 с.
37. Ниворожкина Л.И., Арженовский С.В., Кокина Е.П. Эконометрика: теория и практика: учебное пособие. М.: РИОР: ИНФРА-М, 2018. 207 с.
38. Новак Э. Введение в методы эконометрики. Сборник задач. М.: Финансы и статистика, 2004. 248 с.
39. Новиков А.И. Экономико-математические методы и модели: учебник. М.: Дашков и К°, 2022. 532 с.
40. Носко В.П. Эконометрика: учебник: в 2 кн.. М.: Дело, 2021. Кн. 2. 592 с.
41. Носко В.П. Эконометрика: учебник: в 2 кн. М.: Дело, 2021. Кн. 1. 704 с.
42. Орлова И.В., Филонова Е.С., Агеев А.В. Эконометрика: компьютерный практикум для студентов третьего курса, обучающихся по специальностям 080105.65 Финансы и кредит, 080109.65 Бухгалтерский учет, анализ и аудит. М.: ВЗФЭИ, 2011. 96 с.
43. Плохотников К.Э. Основы эконометрики в пакете STATISTICA: учебное пособие. М.: Вузовский учебник, 2020. 297 с.
44. Потахова И.В. Эконометрика: учебное пособие. Томск: ТУСУР, 2015. 110 с.
45. Практикум по эконометрике: учебное пособие / И.И. Елисеева [и др.]; под ред. И.И. Елисеевой. М.: Финансы и статистика, 2003. 192 с.
46. Просветов Г.И. Эконометрика. Задачи и решения: учебно-методическое пособие. М.: РДЛ, 2005. 104 с.
47. Репина О.М., Руденко С.А. Моделирование экономических процессов: учебное пособие. Йошкар-Ола: ПГТУ, 2021. 112 с.
48. Рожков И.М., Ларионова И.А., Исаева Н.А. Эконометрика. Продвинутый курс для начинающих исследователей: учебное пособие. М.: НИТУ «МИСиС», 2020. 268 с.
49. Сажин Ю.В., Иванова И.А. Эконометрика: учебник. Саранск: МГУ им. Н.П. Огарева, 2014. 316 с.
50. Саркисян Р.С. Эконометрика: учебное пособие. Новокузнецк: Знание-М, 2021. 328 с.
51. Сборник задач к начальному курсу эконометрики: учебное пособие / П.К. Катышев [и др.]. М.: Дело, 2021. 368 с.

52. Сборник задач по эконометрике временных рядов и панельных данных / Ф.С. Картаев [и др.]. М.: Экономический факультет МГУ им. М.В. Ломоносова, 2016. 64 с.
53. Середа В.А., Литаврин А.В., Собачкина Н.Л. Эконометрика: учебное пособие. Красноярск: СФУ, 2018. 148 с.
54. Снатенков А.А., Еремеева Н.С. Сборник задач по эконометрике. Оренбург: Издательский центр ОГАУ, 2004. 57 с.
55. Соколов Г.А. Эконометрика: теоретические основы: учебное пособие. М.: ИНФРА-М, 2022. 216 с.
56. Сорокина Е.А. Практикум по эконометрике (регрессионные модели с одним уравнением): лабораторный практикум по эконометрике для обучающихся по направлению «Экономика». Махачкала: ДГИНХ, 2011. 75 с.
57. Талызин В.А. Сборник задач по эконометрике: учебное пособие. Казань: РИЦ «Школа», 2009. 112 с.
58. Тимофеев В.С., Фаддеенков А.В., Щеколдин В.Ю. Эконометрика: учебник для вузов по экономическим направлениям и специальностям. Новосибирск: НГТУ, 2013. 340 с.
59. Тимофеева Н.Ю. Практикум по эконометрике: система экономических уравнений. Временные ряды: учебно-методическое пособие. Елец: ЕГУ им. И.А. Бунина, 2017. 90 с.
60. Хайруллина О.И., Баянова О.В. Эконометрика: практикум. Пермь: ИПЦ «Прокростъ», 2020. 127 с.
61. Шанченко Н.И. Эконометрика: лабораторный практикум: учебное пособие. Ульяновск: УлГТУ, 2011. 117 с.
62. Эконометрика. Практикум: учебно-практическое пособие / И.А. Кацко [и др]; под ред. И.А. Кацко. М.: КНОРУС, 2022. 216 с.
63. Эконометрика: учебник / И.И. Елисеева [и др.]; под ред. И.И. Елисеевой. М.: Финансы и статистика, 2003. 344 с.
64. Эконометрика: учебник / под ред. В.Б. Уткина. М.: Дашков и К°, 2017. 564 с.
65. Яковлев В.П. Эконометрика: учебник для бакалавров. М.: Дашков и К°, 2019. 384 с.

ПРИЛОЖЕНИЕ

Вопросы для подготовки к экзамену (зачету) по курсу «Эконометрика»

1. Предмет, задачи и принципы эконометрики.
2. Основные этапы эконометрических исследований.
3. Эконометрическая модель. Виды и классификация эконометрических моделей.
4. Этапы эконометрического моделирования. Проблемы, решаемые при эконометрическом исследовании.
5. Требования к факторам, включаемым в эконометрическую модель.
6. Система нормальных уравнений МНК и ее решение.
7. Свойства оценок параметров, полученных МНК. Условия Гаусса – Маркова.
8. Оценка параметров линейной регрессии по МНК. Система нормальных уравнений.
9. Экономический смысл параметров однофакторной регрессии.
10. Критерии качества оценки линейной регрессии.
11. Линейный коэффициент корреляции.
12. Коэффициент детерминации и его свойства.
13. Интервальный прогноз на основе линейного уравнения регрессии.
14. Методы выбора типа уравнения нелинейной регрессии.
15. Методы нелинейного оценивания коэффициентов модели регрессии.
16. Коэффициенты эластичности функций, применяемых в эконометрических расчетах.
17. Корреляция для нелинейной регрессии.
18. Средняя ошибка аппроксимации.
19. Нелинейные модели регрессии и их линеаризация.
20. Основные направления использования многофакторного регрессионного анализа в экономико-статистических исследованиях.
21. Построение уравнения двухфакторной линейной модели.
22. Методика отбора факторов, включаемых в множественную регрессию.
23. Мультиколлинеарность исходных данных и способы ее устранения.
24. Парные коэффициенты корреляции в множественной регрессии.
25. Совокупные коэффициенты множественной корреляции и детерминации.
26. Частные коэффициенты корреляции и детерминации.

27. Проверка значимости коэффициентов и адекватности регрессии для множественной линейной регрессионной модели.

28. Прогнозирование по регрессионной модели и его точность. Доверительные вероятности и интервалы прогноза.

29. Линеаризация регрессионных моделей путем логарифмических преобразований.

30. Экономическая интерпретация многофакторной регрессионной модели.

31. Фиктивные переменные в множественной регрессии.

32. Использование фиктивных переменных для учета структурных изменений в экономике.

33. Прогнозирование значения результирующей переменной по уравнению регрессии.

34. Виды систем уравнений. Формы эконометрических моделей. Зависимые и независимые переменные.

35. Проблема идентификации эконометрических моделей.

36. Необходимое и достаточные условия идентифицируемости системы уравнений.

37. Косвенный МНК.

38. Двухшаговый МНК.

39. Применение систем эконометрических уравнений в экономике.

40. Понятие о временных рядах и их моделях. Компоненты временного ряда. Тренд.

41. Аналитический вид тренда.

42. Автокорреляция уровней временного ряда и выявление его структуры.

43. Этапы построения модели временного ряда.

44. Временные ряды с лаговыми переменными.

45. Модели полиномиальных и геометрических лагов.

46. Проверка гипотезы о наличии автокорреляции в остатках по критерию Дарбина – Уотсона.

47. Выявление аномальных наблюдений. Метод Ирвина.

48. Применение временных рядов в экономике.

49. Состояние и перспективы развития эконометрики.

Тестовые задания по курсу «Эконометрика»

1. Суть МНК при построении уравнения и линии регрессии:

а) провести линию, соединяющую все точки фактических данных;

б) провести ее через наиболее важные точки фактических данных;

в) построить ее так, чтобы она прошла как можно ближе к фактическим данным.

2. Коэффициенты эластичности показывают, на какую величину в среднем изменится *y*, если *x* увеличить соответственно:

а) на 1 %;

б) на 100 %;

в) на единицу своего измерения.

3. Мультиколлинеарность – термин в эконометрике, обозначающий:

а) метод, позволяющий оценить параметры модели, опираясь на случайные выборки;

б) статистическую зависимость между последовательными элементами одного ряда, которые взяты со сдвигом;

в) наличие линейной зависимости между факторами (объясняющими переменными) регрессионной модели.

4. Если при построении уравнения регрессии получено значение коэффициента эластичности $Э = -15{,}3$, то:

а) при уменьшении фактора X на 1 % от своего среднего результативный признак Y увеличится в среднем на 15,3 % от своего среднего значения;

б) допущена ошибка в вычислениях;

в) при увеличении фактора X на 1 % от своего среднего результативный признак Y уменьшится в среднем на 1,53 % от своего среднего значения;

г) при увеличении фактора X на 1 % от своего среднего результативный признак Y уменьшится в среднем на 15,3 % от своего среднего значения.

5. Для проверки значимости отдельных коэффициентов регрессии, т.е. гипотез, используют:

а) нормальный закон распределения;

б) t-критерий;

в) распределение Фишера.

6. В каких пределах изменяется коэффициент корреляции:

а) от 0 до 1; б) от –1 до 0; в) от –1 до 1; г) от 0 до 10?

7. В каких пределах изменяется индекс множественный корреляции:

а) от 0 до 1; б) от –1 до 0; в) от –1 до 1; г) от 0 до 10?

8. Средняя ошибка аппроксимации A = 30 % свидетельствует:

а) о не очень хорошем подборе модели к наблюдаемым данным, прогнозы по этой модели следует строить с осторожностью;

б) о хорошем подборе модели к фактическим данным;

в) о неудовлетворительном подборе модели;

г) о том, что такую модель не следует применять для прогнозирования.

9. Сколько следует ввести фиктивных переменных для описания влияния образования (высшего, среднего, среднего специального, неполного среднего) на уровень заработной платы:

а) 1; б) 2; в) 3; г) 4?

10. Необходимое условие идентификации:

а) $D + 1 = H$; б) $D < H + 1$; в) $D - 1 > H$.

11. Как называются модели временных данных в эконометрике, объясняющие поведение результативного признака и зависимости от предыдущих значений результативных переменных:

а) модели ожиданий;

б) модели авторегрессий;

в) модели с распределенным лагом;

г) модели стационарных рядов;

д) модели нестационарных рядов?

12. Основная идея МНК для построения уравнения регрессии:

а) сумма квадратов остатков минимизируется;

б) сумма остатков минимизируется;

в) сумма квадратов остатков максимизируется;

г) сумма остатков максимизируется;

д) сумма квадратов фактора минимизируется.

13. Среди предложенных моделей выделить линейные парные регрессионные модели:

а) $Y = 7{,}1 - 0{,}5X + \varepsilon$;

б) $Y = 10{,}3X_1 + 6X_2$;

в) $Y = 1{,}2X_2 + \varepsilon$;

г) $Y = -54{,}2X + \varepsilon$;

д) $Y = 2{,}3X$.

14. Если в уравнении регрессии имеется несущественная переменная, то она обнаруживает себя по низкому значению:

а) t-статистики; б) F-статистики; в) коэффициента детерминации.

15. Отметьте основные виды ошибок спецификации:

а) отбрасывание значимой переменной;

б) отбрасывание незначимой переменной;

в) высокое значение коэффициента детерминации;

г) выбор неправильной формы модели.

16. Определить недостающее значение и выбрать нужный вариант ответа: «На практике о наличии мультиколлинеарности обычно судят по матрице парных коэффициентов корреляции (матрице R). Если один из элементов матрицы R больше ..., то считают, что имеет место

мультиколлинеарность и в уравнение регрессии следует включать только один из показателей x_j или x_i»:

а) 0,3; б) 0,5; в) 0,6; г) 0,7; д) 0,9; е) другое.

17. Оценки максимального правдоподобия и МНК:

а) могут не совпадать; б) совпадают; в) никогда не совпадают.

18. Квадрат какого коэффициента указывает долю дисперсии одной случайной величины, обусловленную вариацией другой:

а) коэффициент детерминации;

б) парный коэффициент корреляции;

в) частный коэффициент корреляции;

г) множественный коэффициент корреляции?

19. В каких пределах изменяется коэффициент детерминации:

а) от 0 до 1; б) от –1 до 0; в) от –1 до 1; г) от 0 до 10?

20. В правой части структурной формы взаимозависимой системы могут стоять:

а) только экзогенные лаговые переменные;

б) только экзогенные переменные (как лаговые, так и нелаговые);

в) только эндогенные лаговые переменные;

г) только эндогенные переменные (как лаговые, так и нелаговые);

д) любые экзогенные и эндогенные переменные.

21. Коэффициент регрессии изменяется в пределах:

а) от –1 до 1; б) от 0 до 1; в) любых.

22. С увеличением объема выборки:

а) возрастает точность оценок;

б) уменьшается ошибка регрессии;

в) расширяются интервальные оценки;

г) уменьшается коэффициент детерминации.

23. Коэффициент корреляции, равный нулю, означает, что между переменными:

а) линейная связь отсутствует;

б) существует линейная связь;

в) ситуация неопределенна.

24. Коэффициент корреляции, равный –1, означает, что между переменными:

а) линейная связь отсутствует;

б) существует слабая линейная связь;

в) функциональная зависимость;

г) ситуация неопределенна.

25. Если при построении уравнения регрессии получен коэффициент детерминации $R_2 = 0,98$:

а) то зависимость Y от X слабая, незначительная, изменения результативного признака Y по большей части обусловлены случайными (или не включенными в модель) факторами;

б) то изменения результативного признака Y на 0,98 % обусловлены изменениями фактора X;

в) то изменения результативного признака Y на 98 % обусловлены изменениями фактора X;

г) то допущена ошибка в вычислениях;

д) то изменения результативного признака Y на 98 % обусловлены случайными (или не включенными в модель) факторами.

26. Объяснить, что такое уровень значимости:

а) вероятность того, что статистика выйдет за пределы доверительного интервала, заданного данной критической границей, и тем самым будет отклонена верная нулевая гипотеза;

б) вероятность того, что статистика войдет в доверительный интервал, заданный данной критической границей, и тем самым будет отклонена верная нулевая гипотеза;

в) вероятность того, что статистика выйдет за пределы доверительного интервала, заданного данной критической границей, и тем самым будет принята верная нулевая гипотеза;

г) вероятность того, что статистика войдет в доверительный интервал, заданный данной критической границей, и тем самым будет принята верная нулевая гипотеза;

д) нет правильного ответа.

27. Аддитивная модель временного ряда имеет вид:

а) $Y = T \;\; S \;\; E$; б) $Y = T + S + E$; в) $Y = T \;\; S + E$.

28. Коэффициент автокорреляции характеризует:

а) тесноту линейной связи текущего и предыдущего уровней ряда;

б) тесноту нелинейной связи текущего и предыдущего уровней ряда;

в) наличие или отсутствие тенденции.

29. Критерий Дарбина – Уотсона применяется:

а) для определения автокорреляции в остатках;

б) определения наличия сезонных колебаний;

в) оценки существенности построенной модели.

30. Добавление в уравнение множественной регрессии новой объясняющей переменной:

а) уменьшает значение коэффициента детерминации;

б) увеличивает значение коэффициента детерминации;

в) не оказывает никакого влияния на коэффициент детерминации.

31. Что является предметом изучения эконометрики:

а) количественная сторона экономических процессов и явлений;

б) массовые экономические процессы и явления;

в) система внутренних связей между явлениями национальной экономики?

32. Теорема Гаусса – Маркова в эконометрике опирается:

а) на МНК;

б) метод наименьших модулей;

в) метод инструментальных переменных.

33. Коэффициент эластичности (формула в общем виде) в эконометрике имеет вид:

а) $Э = y_x \ \dfrac{x}{y}$; б) $Э = \dfrac{(2\beta_2\overline{x} + \beta_1) \ \overline{x}}{y(\overline{x})}$; в) $Э = \dfrac{\partial x}{\partial y} \ \dfrac{x}{y}$; г) $Э = \dfrac{\beta_1 \ x_1}{\beta_0 + \beta_1 \ x_1}$.

34. Модели временных рядов в эконометрике – это модели:

а) которые используются для того, чтобы определить, как себя будет вести тот или иной фактор в течение определенного промежутка времени;

б) которые позволяют максимально точно рассчитать период времени, требующийся для того, чтобы значение фактора изменилось на значимую величину;

в) для построения которых используются данные, характеризующие один объект за несколько последовательных периодов.

35. Модели в эконометрике – это:

а) средство прогнозирования значений определенных переменных;

б) экономические и статистические характеристики объектов, выраженные математическим языком;

в) данные одного типа, сгруппированные определенным образом.

36. Зависимая переменная в эконометрике – это:

а) параметр, состоящий из случайной и неслучайной величин;

б) некоторая переменная регрессионной модели, которая является функцией регрессии с точностью до случайного возмущения;

в) переменная, которая получается переводом качественных характеристик в количественные, т.е. путем присвоения цифровой метки.

37. Каковы цели эконометрики:

а) поиск, трактовка (с использованием математического инструментария) и систематизация факторов, которые влияют на поведение экономического объекта;

б) выявление качественных и количественных связей между характеристиками экономических объектов с целью построить экономическую модель их развития;

в) разработка инструментов для прогнозирования поведения экономического объекта в различных ситуациях и на их базе решение практических задач по управлению объектом, выбору поведения в сложившихся экономических условиях и т.д.?

38. Среди предложенных моделей выделить линейные парные регрессионные модели:

а) $Y = 7{,}1 - 0{,}5X + \varepsilon$; б) $Y = 10{,}3X_1 + 6X_2 + \varepsilon$; в) $Y = 1{,}2X_2 + \varepsilon$;

г) $Y = -54{,}2X + \varepsilon$; д) $Y = 2{,}3X$; е) $Y = 6X^2$.

39. Произвести отбор факторов для включения в уравнение множественной регрессии на основе матрицы парных коэффициентов корреляции:

	y	x_1	x_2	x_3
y	1			
x_1	0,6	1		
x_2	0,7	0,1	1	
x_3	0,9	0,5	0,7	1

Выбрать нужный вариант ответа:

а) x_1, x_2; б) x_1, x_3; в) x_2, x_3; г) x_1, x_2, x_3.

40. Определить доверительный интервал для математического ожидания нормально распределенной случайной величины X, если $\beta = 0{,}95$; $t_{5;0,95} = 2{,}57$, по данным, представленным ниже:

x_1	x_2	x_3	x_4	x_5	x_6
2	4	3	5	7	9

Выбрать нужный вариант ответа:

а) 2,1; б) 3,1; в) 4,1.

41. Рассчитать эмпирический коэффициент корреляции r_{xy} по данным, приведенным ниже:

x	2	3	8	7
y	6	4	6	8

Выбрать нужный вариант ответа:

а) –1; б) 0,56; в) 1.

42. Рассчитать средний коэффициент эластичности линейной функции

$$y = 1{,}5 + 0{,}2x;\ \bar{x} = 3{,}5,\ \bar{y} = 2{,}5.$$

Выбрать нужный вариант ответа:

а) –0,6; б) 0,28; в) 1.

43. Значения переменных X и Y указаны ниже:

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

Чему равен коэффициент корреляции этих переменных? Выберите нужный вариант ответа:

а) –1; б) 0,5; в) 1.

44. Построить линейное уравнение регрессии прибыли Y (ден. ед.) от объема выпускаемой продукции X (тыс. шт.) по данным, полученным для 7 предприятий:

x_i	10	20	40	30	50	30	40
y_i	3	7	15	10	17	10	15

Уравнение регрессии имеет вид:

а) $\hat{y} = 0,566 + 0,368x$;

б) $\hat{y} = -0,566 + 0,368x$;

в) $\hat{y} = -0,566 - 0,368x$.

45. Найти коэффициент эластичности нелинейной функции

$$y = 152^x, \quad \bar{x} = 2,6; \ \ln 2 = 0,69.$$

Определить, на сколько процентов в среднем изменится Y, если X в среднем увеличится на 1 %. Выбрать нужный вариант ответа:

а) 2,6; б) 90,1; в) 10,35.

46. Доля остаточной дисперсии в общей составила 0,19. Определить коэффициент детерминации:

а) 0–1; б) 0,81; в) 1.

47. Доля остаточной дисперсии в общей составила 0,36. Определить индекс корреляции:

а) 0,8; б) 0,9; в) 1.

48. Какое из следующих уравнений нелинейно по оцениваемым параметрам:

1) $y_x = a + bx + \varepsilon$;

2) $y_x = a + b\ln x + \varepsilon$;

3) $y_x = ax^b\varepsilon$?

Выбрать нужный вариант ответа:

а) первое; б) второе; в) третье.

49. Зависимость расходов на питание от дохода $y = 6,13 \cdot x^{0,374}$. Определить коэффициент эластичности спроса для указанной регрессии. Выбрать нужный вариант ответа:

а) 1; б) 0,374; в) 1.

50. Определить линейность функции $y = a + bx^3 + \varepsilon$ по факторам и параметрам:

а) нелинейная и линейная соответственно;

б) линейная и нелинейная соответственно;

в) линейная и линейная соответственно;

г) нелинейная и нелинейная соответственно.

51. Какое уравнение регрессии нельзя свести к линейному виду:

1) $y_x = a + b\ln x$;

2) $y_x = a\,x^b$;

3) $y_x = a + bx^c$?

Выбрать нужный вариант ответа:

а) первое; б) второе; в) третье.

52. Определить наличие мультиколлинеарности факторов, если заданы значения парных коэффициентов корреляции между факторами: $r_{x1x1} = 1$; $r_{x2x2} = 1$; $r_{x3x3} = 1$; $r_{x1x2} = 0{,}01$; $r_{x1x3} = 0$; $r_{x2x3} = 0{,}05$. Выбрать нужный вариант ответа:

а) она есть; б) ее нет;

53. Для двухфакторной линейной регрессии заданы парные коэффициенты корреляции: $r_{yx1} = 0{,}8$; $r_{yx2} = –0{,}2$; $r_{x1x2} = –0{,}1$. Рассчитать частные коэффициенты корреляции и совокупный коэффициент множественной детерминации. Выбрать нужный вариант ответа:

а) $r(yx_1/x_2) \approx 0{,}8$; $r(yx_2/x_1) \approx -0{,}2$; $r(x_1x_2/y) \approx 0{,}1$;

б) $r(yx_1/x_2) \approx 0{,}8$; $r(yx_2/x_1) \approx 0{,}2$; $r(x_1x_2/y) \approx 0{,}1$;

в) $r(yx_1/x_2) \approx 0{,}8$; $r(yx_2/x_1) \approx 0{,}2$; $r(x_1x_2/y) \approx -0{,}1$.

54. Задано уравнение двухфакторной линейной регрессии $y = 20 + 14x_1 + 21x_2$; средние значения признаков: $x_1 = 3$; $x_2 = 1$; $y = 4$. Проанализировать влияние каждого из факторов на результат на основе частных коэффициентов эластичности:

а) $\overline{Э_1} = 10{,}5\ \%$; $\overline{Э_2} = 50{,}25\ \%$;

б) $\overline{Э_1} = 10{,}5\ \%$; $\overline{Э_2} = -5{,}25\ \%$;

в) $\overline{Э_1} = 10{,}5\ \%$; $\overline{Э_2} = 5{,}25\ \%$.

55. Как называются модели временных данных в эконометрике, объясняющие поведение результативного признака в зависимости от предыдущих значений факторных переменных:

а) модели ожиданий;

б) модели авторегрессий;

в) модели с распределенным лагом;

г) модели стационарных рядов;

д) модели нестационарных рядов?

56. Какие Вы можете назвать возможные способы учета структурных сдвигов в экономических системах, описываемых эконометрическими моделями:

а) включение в модель фиктивных переменных;

б) включение в модель трендов;

в) включение в модель трендов и фиктивных переменных;

г) построение системы одновременных уравнений;

д) включение в модель трендов и фиктивных переменных, построение системы одновременных уравнений?

57. Для выявления мультиколлинеарности используются:

а) частные коэффициенты корреляции;

б) парные коэффициенты корреляции;

в) множественный коэффициент детерминации;

г) коэффициент регрессии.

58. Коэффициент множественной корреляции используется для оценки:

а) тесноты связи между Y и всеми факторами модели;

б) тесноты связи между Y и отдельным фактором;

в) тесноты связи между регрессорами;

г) влияния не учтенных в модели факторов.

59. Коэффициент уравнения регрессии b ($Y = bX + a$) показывает:

а) на сколько процентов изменится результат при изменении фактора на 1 %;

б) на сколько процентов изменится фактор при изменении результата на 1 %;

в) на сколько единиц изменится результат при изменении фактора на 1 ед.;

г) на сколько единиц изменится фактор при изменении результата на 1 ед.;

д) во сколько раз изменится результат при изменении фактора на 1 ед.

60. Критерий Стьюдента предназначен:

а) для нахождения экономической значимости каждого коэффициента уравнения;

б) определения статистической значимости каждого коэффициента уравнения;

в) проверки модели на автокорреляцию остатков;

г) определения экономической значимости модели в целом;

д) проверки на гомоскедастичность.

61. Табличное значение критерия Стьюдента зависит:

а) только от уровня доверительной вероятности;

б) только от числа факторов в модели;

в) только от длины исходного ряда;

г) только от уровня доверительной вероятности и длины исходного ряда;

д) и от доверительной вероятности, и от числа факторов, и от длины исходного ряда.

62. Для проверки значимости уравнения регрессии в целом используется критерий:

а) Стьюдента; б) Дарбина – Уотсона; в) Фишера; г) Вальда.

63. Какое значение может принимать коэффициент детерминации:

а) –0,5; б) –0,2; в) 0,4; г) 1,2?

64. Известно, что при фиксированном значении X_3 между величинами X_1 и X_2 существует положительная связь. Какое значение может принять частный коэффициент корреляции $r_{12/3}$:

а) –0,8; б) 0; в) 0,4; г) 1,3.

65. Метод максимального правдоподобия лучше работает на …, где он, как правило, дает оценки с минимальной дисперсией (выбрать необходимый пункт):

а) больших выборках;

б) малых выборках;

в) любых выборках.

66. Коэффициент детерминации – это:

а) квадрат парного коэффициента корреляции;

б) квадрат частного коэффициента корреляции;

в) квадрат множественного коэффициента корреляции.

67. Нормирование признаков производят с целью … (выбрать необходимый пункт):

а) устранения влияния различных единиц измерения;

б) уменьшить признаковое пространство;

в) упрощения расчетов.

68. В хорошо подобранной модели остатки должны … (выбрать необходимые пункты):

а) иметь нормальный закон распределения с нулевым математическим ожиданием и постоянной дисперсией;

б) не коррелировать друг с другом;

в) иметь экспоненциальный закон распределения;

г) быть хаотично разбросаны.

69. Неправильный выбор функциональной формы или объясняющих переменных называется:

а) ошибками спецификации;

б) ошибками прогноза;

в) гетероскедастичностью.

70. Коэффициент корреляции считается значимым с заданной вероятностью ошибки:

а) если $t_{набл}$ по модулю будет больше, чем $t_{кр}$;

б) не имеет значения;

в) $t_{набл}$ по модулю будет меньше, чем $t_{кр}$.

71. Матрица R парных коэффициентов корреляции является (выбрать необходимые пункты):

а) обратной;

б) транспонированной;

в) симметричной;

г) положительно определенной.

72. В регрессионном анализе x_j рассматриваются как:

а) неслучайные величины;

б) случайные величины;

в) любые величины.

73. На главной диагонали ковариационной матрицы находятся:

а) дисперсии коэффициентов регрессии;

б) средние значения коэффициентов регрессии;

в) коэффициенты корреляции;

г) квадраты коэффициентов корреляции.

74. Предельно допустимое значение средней ошибки аппроксимации составляет … процентов (выбрать нужный вариант ответа):

а) не более 10–12; б) не более 3–5; в) не более 8–10.

75. Какой показатель измеряет тесноту статистической связи между переменной и объясняющими переменными:

а) коэффициент детерминации;

б) коэффициент рекурсии;

в) коэффициент корреляции?

76. Коэффициентом детерминации R^2 характеризуют долю вариации … переменной … с помощью уравнения регрессии:

а) зависимой, объясненную;

б) зависимой, необъясненную;

в) независимой, объясненную;

г) независимой, необъясненную.

77. Экзогенные переменные – это:

а) внешние переменные, которые задаются вне моделей, являются автономными и управляемыми;

б) внутренние переменные;

в) переменные, формирующиеся в результате функционирования социально-экономической системы;

г) лаговые переменные.

78. Эндогенные переменные – это:

а) лаговые переменные;

б) внешние переменные;

в) автономные переменные;

г) внутренние переменные, которые формируются в результате функционирования социально-экономической системы.

79. Объем выборки должен превышать число рассчитываемых параметров при исследуемых факторах:

а) в 5–6 раз; б) в 2–3 раза; в) в 10–12 раз; г) в 20–25 раз.

80. К ошибкам спецификации относится:

а) неправильный выбор той или иной математической функции;

б) однородность выбранной совокупности;

в) учет в модели случайных факторов;
г) учет в модели существенных факторов.
81. Относительно формы зависимости различают … регрессии:
а) линейную и нелинейную;
б) простую и множественную;
в) непосредственную и косвенную;
г) положительную и отрицательную.
82. По количеству факторов, включенных в уравнение регрессии, различают … регрессии:
а) простую и множественную;
б) линейную и нелинейную;
в) непосредственную и косвенную;
г) множественную и многофакторную.
83. Простая линейная регрессия предполагает наличие:
а) одного фактора и линейность уравнения регрессии;
б) двух и более факторов и линейность уравнения регрессии;
в) одного фактора и нелинейность уравнения регрессии;
г) двух и более факторов и нелинейность уравнения регрессии.
84. Объем выборки определяется:
а) числом параметров при независимых переменных;
б) числом результативных переменных;
в) объемом генеральной совокупности;
г) числовыми значениями переменных, отбираемых в выборку.
85. Графическое изображение наблюдений на декартовой плоскости координат называется полем:
а) корреляции;
б) регрессии;
в) случайных воздействий;
г) автокорреляции.
86. В качестве показателя тесноты связи для линейного уравнения парной регрессии используется:
а) линейный коэффициент корреляции;
б) множественный коэффициент линейной корреляции;
в) линейный коэффициент детерминации;
г) линейный коэффициент регрессии.
87. Простая линейная регрессия предполагает наличие:
а) одного фактора и линейность уравнения регрессии;
б) двух и более факторов и линейность уравнения регрессии;
в) одного фактора и нелинейность уравнения регрессии;
г) двух и более факторов и нелинейность уравнения регрессии.
88. Объем выборки определяется:
а) числом параметров при независимых переменных;
б) числом результативных переменных;

в) объемом генеральной совокупности;

г) числовыми значениями переменных, отбираемых в выборку.

89. Мультипликативная модель временного ряда строится:

а) если значения сезонной компоненты предполагаются постоянными для различных циклов;

б) амплитуда сезонных колебаний возрастает или уменьшается;

в) отсутствует тенденция.

90. При моделировании линейного уравнения множественной регрессии вида $y = b_1x_1 + b_2x_2 + \varepsilon$ необходимо, чтобы выполнялось требование отсутствия взаимосвязи:

а) между x_1 и x_2; б) y и $\{x_1; x_2\}$; в) a и $\{b_1; b_2\}$; г) b_1 и b_2.

91. При идентификации модели множественной регрессии $y = a + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 + \varepsilon$ количество оцениваемых параметров составляет:

а) 6; б) 4; в) 5.

92. Косвенный МНК применим:

а) для идентифицируемой системы одновременных уравнений;

б) любой системы одновременных уравнений;

в) неидентифицируемой системы уравнений;

г) неидентифицируемой системы рекурсивных уравнений.

93. Оценить структурную модель на идентификацию (проверить необходимое и достаточное условия):

$$\begin{aligned}
y_1 &= b_{12}y_2 + b_{13}y_3 + a_{11}x_1 + a_{12}x_2;\\
y_2 &= b_{21}y_1 + a_{22}x_2 + a_{23}x_3 + a_{24}x_4;\\
y_3 &= b_{31}y_1 + b_{32}y_2 + a_{33}x_3 + a_{34}x_4.
\end{aligned}$$

Выбрать нужный вариант ответа:

а) неидентифицируема;

б) идентифицируема;

в) сверхидентифицируема.

94. Есть модель денежного рынка:

$$\begin{aligned}
R_t &= a_1 + b_{11}M_t + b_{12}Y_t + \varepsilon_1;\\
Y_t &= a_2 + b_{33}R_t + b_{22}I_t + \varepsilon_2,
\end{aligned}$$

где R – процентные ставки; Y – ВВП; I – число удовлетворенных прошений; M – денежная масса; t – текущий период; a_1, a_2 – свободные члены уравнений.

Проверить данную модель на идентификацию. Выбрать нужный вариант ответа:

а) идентифицируема;

б) неидентифицируема;

в) сверхидентифицируема.

95. Модель с распределенным лагом имеет вид $y_t = 4{,}25 + 2{,}87x_t + 0{,}88x_{t-1} + 9{,}52\,x_{t-2} + 2{,}6\,x_{t-3}$.

Определить краткосрочный и долгосрочный мультипликаторы. Выбрать нужный вариант ответа:

а) b_o= 2,87; b = 15,87;

б) b_o = 4,25; b = 15,87;

в) b_o= 4,25; b = 2,6.

96. Эконометрическая модель имеет вид:

а) $\hat{y} = a + bx + \varepsilon$;

б) $\hat{y} = a + b_1x + b_2x^2 + \varepsilon$;

в) $y = f(x)$;

г) $y = f(x + y)$.

97. Метод наименьших квадратов позволяет:

а) получить оценки параметров линейной регрессии, исходя из условия $\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \to min$;

б) получить оценки параметров регрессии, исходя из условия $\ln(\prod_{i=1}^{n} f(y_i, \Theta) \to max$;

в) проверить статистическую значимость параметров регрессии;

г) получить оценки параметров нелинейной регрессии, исходя из условия $\sum_{i=1}^{n}(\bar{y} - \hat{y}_i)^2 \to min$.

98. Уравнение линейной множественной регрессии:

а) $\hat{y} = a + bx$;

б) $\hat{y} = a + b_1x_1 + b_2x_2 + \cdots + b_px_p$;

в) $\hat{y} = ax_1^{b_1}x_2^{b_2} \dots x_p^{b_p}$.

99. Уравнение множественной регрессии в стандартизованном масштабе имеет вид $t_y = 20 + 0{,}9t_{x_1} + 0{,}5t_{x_2} + \varepsilon$. На результативный признак оказывает большее влияние:

а) x_1;

б) x_2;

в) нельзя сделать вывод.

100. Коэффициент множественной корреляции для линейной зависимости можно рассчитать по формуле:

а) $R_{yx_1\dots x_p} = \sqrt{\sum \beta_i r_{yx_i}}$; б) $R_{yx_1\dots x_p} = \sum \beta_i r_{yx_i}$; в) $r = \frac{\overline{xy} - \bar{x}\cdot\bar{y}}{\sigma_x\sigma_y}$;

г) $R_{yx_1\dots x_p} = \sqrt{1 - \frac{\sum(y-\hat{y})^2}{\sum(y-\bar{y})^2}}$.

101. Для отражения влияния качественной сопутствующей переменной, имеющей m состояний, обычно включают в модель … фиктивную переменную:

а) $m + 1$;

б) $(m + 1)^2$;

в) $m - 1$; г) $(m - 1)^2$.

102. Установите соответствие:

а) регрессионная модель; 1) $x\dot{-}1 = \begin{cases} 0, & x = 0, \\ x-1, & x > 0; \end{cases}$

б) система одновременных уравнений; 2) $\begin{cases} R = a_1 + b_{11}M + b_{12}Y + \varepsilon_1, \\ Y = a_2 + b_{21}R + \varepsilon_2; \end{cases}$

в) модель временного ряда. 3) $y = a + b_1x_1 + b_2x_2 + \varepsilon$;

4) $y_t = T_t + S_t + E_t$.

103. Для линейного уравнения множественной регрессии установите соответствие:

$$y = a + b_1x_1 + b_2x_2 + \varepsilon.$$

а) факторные переменные; 1) y;y

б) результативная переменная; 2) y, x_1;a

в) параметры; 3) a, ε;

г) случайная компонента. 4) x_1, x_2;

5) ε;

6) a, b_1, b_2.

104. Было замечено, что при увеличении количества вносимых удобрений урожайность также возрастает, однако по достижении определенного значения фактора моделируемый показатель начинает убывать. Для исследования данной зависимости можно использовать спецификацию уравнения регрессии:

а) $y = a + bx + cx^2 + \varepsilon$;

б) $y = a + b_1x_1 + b_2x_2 + \varepsilon$;

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

г) $y = a + x^b + \varepsilon$.

105. Если наиболее высоким оказался коэффициент автокорреляции первого порядка, то исследуемый ряд содержит:

а) тенденцию;

б) случайную компоненту;

в) циклические колебания;

г) сильную нелинейную тенденцию.

106. Значение коэффициента автокорреляции второго порядка характеризует связь:

а) между исходными уровнями и уровнями другого временного ряда;

б) двумя временными рядами;

в) исходными уровнями и уровнями другого ряда, сдвинутыми на 2 момента времени;

г) исходными уровнями и уровнями этого же ряда, сдвинутыми на 2 момента времени.

107. Какая функция используется при моделировании показателей с постоянным ростом:

а) линейная;

б) показательная;

в) степенная?

108. В каком случае рекомендуется применять для моделирования показателей с увеличивающимся ростом параболу:

а) если относительная величина прироста увеличивается неограниченно;

б) абсолютная величина прироста растет по линейному закону;

в) относительная величина прироста неизменна?

109. Для получения качественных оценок уравнений регрессии необходимо выполнение следующих предпосылок МНК (выбрать необходимые пункты):

а) отклонения ε_i должны быть нормально распределенными случайными величинами с нулевым математическим ожиданием и постоянной дисперсией;

б) отклонения ε_i не должны коррелировать друг с другом;

в) отклонения ε_i должны иметь показательный закон распределения.

110. Предпосылки регрессионного анализа исследуют поведение:

а) остаточных величин;

б) неслучайных величин;

в) переменных уравнения регрессии;

г) параметров уравнения регрессии.

111. Общая дисперсия служит для оценки влияния:

а) учтенных в модели факторов;

б) случайных воздействий;

в) не учтенных в модели факторов;

г) как учтенных факторов, так и случайных воздействий.

Таблица 1

Значения $F_{k_1,k_2,\alpha}$ – критерия Фишера – Снедекора –
при уровне значимости $\alpha = 0{,}05$

k_2	k_1									
	1	2	3	4	5	6	7	8	9	10
1	161	200	216	225	230	234	237	239	240	242
2	18,5	19,0	19,2	19,2	19,3	19,3	19,3	19,4	19,4	19,4
3	10,1	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32	2,27
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25
25	4,24	3,39	2,99	2,76	2,60	4,49	2,40	2,34	2,28	2,24
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31	2,25	2,20
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24	2,19
29	4,18	3,33	2,93	2,70	2,55	2,43	2,35	2,28	2,22	2,18
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99
120	3,92	3,07	2,68	2,45	2,29	2,17	2,09	2,02	1,96	1,91
∞	3,84	3,00	3,60	2,37	2,21	2,10	2,01	1,94	1,83	1,83

Примечание. k_1 – число степеней свободы большей дисперсии; k_2 – число степеней свободы меньшей дисперсии.

Окончание табл. 1

k_2	k_1								
	12	15	20	24	30	40	60	120	∞
1	244	246	248	249	250	251	252	253	254
2	19,4	19,4	19,4	19,4	19,5	19,5	19,5	19,5	19,5
3	8,74	8,70	8,66	8,64	8,62	8,59	8,57	8,55	8,53
4	5,91	5,86	5,80	5,77	5,75	5,72	5,69	5,66	5,63
5	4,68	5,62	4,56	4,53	4,50	4,46	4,43	4,40	4,36
6	4,00	3,94	3,87	3,84	3,81	3,77	3,74	3,70	3,67
7	3,57	3,51	3,44	3,41	3,38	3,34	3,30	3,27	3,23
8	3,28	3,22	3,15	3,12	3,08	3,04	3,01	2,97	2,93
9	3,07	3,01	2,94	2,90	2,86	2,83	2,79	2,75	2,71
10	2,91	2,85	2,77	2,74	2,70	2,66	2,62	2,58	2,54
11	2,79	2,72	2,65	2,61	2,57	2,53	2,49	2,45	2,40
12	2,69	2,62	2,54	2,51	2,47	2,43	2,38	2,34	2,30
13	2,60	2,53	2,46	2,42	2,38	2,34	2,30	2,25	2,21
14	2,53	2,46	2,39	2,35	2,31	2,27	2,22	2,18	2,13
15	2,48	2,40	2,33	2,29	2,25	2,20	2,16	2,11	2,07
16	2,42	2,35	2,28	2,24	2,19	2,15	2,11	2,06	2,01
17	2,38	2,31	2,23	2,19	2,15	2,10	2,06	2,01	1,96
18	2,34	2,27	2,19	2,15	2,11	2,06	2,02	1,97	1,92
19	2,31	2,23	2,16	2,11	2,07	2,03	1,98	1,93	1,88
20	2,28	2,20	2,12	2,08	2,04	1,99	1,95	1,90	1,84
21	2,25	2,18	2,10	2,05	2,01	1,96	1,92	1,87	1,81
22	2,23	2,15	2,07	2,03	1,98	1,94	1,89	1,84	1,78
23	2,20	2,13	2,05	2,01	1,96	1,91	1,86	1,81	1,76
24	2,18	2,11	2,03	1,98	1,94	1,89	1,84	1,79	1,73
25	2,16	2,09	2,01	1,96	1,92	1,87	1,82	1,77	1,71
26	2,15	2,07	1,99	1,95	1,90	1,85	1,80	1,75	1,69
27	2,13	2,06	1,97	1,93	1,88	1,84	1,79	1,73	1,67
28	2,12	2,04	1,96	1,91	1,87	1,82	1,77	1,71	1,65
29	2,10	2,03	1,94	1,90	1,85	1,81	1,75	1,70	1,64
30	2,09	2,01	1,93	1,89	1,84	1,79	1,74	1,68	1,62
40	2,00	1,92	1,84	1,79	1,74	1,69	1,64	1,58	1,51
60	1,92	1,84	1,75	1,70	1,65	1,59	1,53	1,47	1,39
120	1,83	1,75	0,66	1,61	1,55	1,50	1,43	1,35	1,25
∞	1,75	1,67	1,57	1,52	1,46	1,39	1,32	1,22	1,00

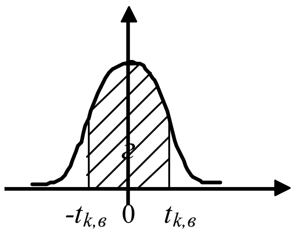


Таблица 2

Значения $t_{k,\beta}$ – критерия Стьюдента

Число степеней свободы k	Доверительная вероятность $\beta = 1 - \alpha$											
	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	0,95	0,98	0,99
1	0,16	0,32	0,51	0,73	1,00	1,38	1,96	3,08	6,31	12,71	31,82	63,66
2	14	29	44	62	0,82	06	34	1,89	2,92	4,30	6,96	9,92
3	14	28	42	58	76	0,98	25	64	35	3,18	4,54	5,84
4	13	27	41	57	74	94	19	53	13	2,78	3,75	4,60
5	13	27	41	56	73	92	16	48	01	57	36	03
6	0,13	0,26	0,40	0,55	1,72	1,91	1,13	1,44	1,94	2,45	3,14	3,71
7	13	26	40	55	71	90	12	41	89	36	00	50
8	13	26	40	55	70	89	11	40	86	31	2,90	35
9	13	26	40	54	70	88	10	38	83	26	82	25
10	13	26	40	54	70	88	09	37	81	23	76	17
11	0,13	0,26	0,40	0,54	0,70	0,88	1,09	1,36	1,80	2,20	2,72	3,11
12	13	26	39	54	69	87	08	36	78	18	68	05
13	13	26	39	54	69	87	08	35	77	16	65	01
14	13	26	39	54	69	87	08	34	76	14	62	2,98
15	13	26	39	54	69	87	07	34	75	13	60	95
16	0,13	0,26	0,39	0,53	0,69	0,86	1,07	1,34	1,75	2,12	2,58	2,92
17	13	26	39	53	69	86	07	33	74	11	57	90
18	13	26	39	53	69	86	07	33	73	10	55	88
19	13	26	39	53	69	86	07	33	73	09	54	86
20	13	26	39	53	69	86	06	32	72	09	53	84
21	0,13	0,26	0,39	0,53	0,69	0,86	1,06	1,32	1,72	2,08	2,52	2,83
22	13	26	39	53	69	86	06	32	72	07	51	82
23	13	26	39	53	68	86	06	32	71	07	50	81
24	13	26	39	53	68	86	06	32	71	06	49	80
25	13	26	39	53	68	86	06	32	71	06	48	79
26	0,13	0,26	0,39	0,53	0,68	0,86	1,06	1,31	1,71	2,06	2,48	2,78
27	13	26	39	53	68	85	06	31	70	05	47	77
28	13	26	39	53	68	85	06	31	70	05	47	76
29	13	26	39	53	68	85	05	31	70	04	46	76

Окончание табл. 2

Число степеней свободы k	Доверительная вероятность $\beta = 1 - \alpha$											
	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	0,95	0,98	0,99
30	13	26	39	53	68	85	05	31	70	04	46	75
40	0,13	0,25	0,39	0,53	0,68	0,85	1,05	1,30	1,68	2,02	2,42	2,70
60	13	25	39	53	68	85	05	30	67	00	39	66
120	0,13	0,25	0,39	0,53	0,68	0,84	1,04	1,29	1,66	1,98	2,36	2,62
∞	13	25	38	52	67	84	04	28	64	96	33	58

Примечание. α – уровень значимости.

Таблица 3

Значения статистик Дарбина – Уотсона d_L и d_u при 5%-м уровне значимости

n	$K^1 = 1$		$K^1 = 2$		$k^l = 3$		$K^1 = 4$		$K^1 = 5$	
	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	0,86	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	1,90
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

Алексей Валерианович Ганичев
Антонина Валериановна Ганичева

ПРАКТИКУМ ПО ЭКОНОМЕТРИКЕ

Учебное пособие

Редактор Ю.А. Якушева
Корректор С.В. Борисов

Подписано в печать 29.01.2024

Формат 60x84 1/16		Бумага писчая
Физ. п. л. 9,25	Усл. п. л. 8,60	Уч.-изд. л. 8,05
Тираж 50 экз.	Заказ № 7	С – 7

Редакционно-издательский центр
Тверского государственного технического университета
170026, г. Тверь, наб. А. Никитина, 22