МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное
учреждения высшего образования
«Юго-Западный государственный университет»
(ЮЗГУ)

Кафедра программной инженерии

УТВЕРЖДАЮ
Проректор по учебной работе
_____ О.Г. Локтионова
«__9__»__10__ 2024 г.

# Анализ данных в цифровой экономике

Методические указания к практическим занятиям по дисциплине «Анализ данных в цифровой экономике» для студентов направления подготовки 02.03.03 «Математическое обеспечение и администрирование информационных систем»

Курск 2024

УДК 004

Составитель: Халин Ю.А.

Рецензент

Доктор физико-математических наук, профессор В.П. Добрица

**Анализ данных в цифровой экономике**: методические указания к практическим занятиям / Юго-Зап. гос. ун-т; сост.: Ю.А. Халин. – Курск, 2024. – 42 с.: Библиогр.: с. 42.

Содержат сведения по вопросам анализа данных в цифровой экономике. Указывается порядок проведения практических занятий, правила оформления, содержание отчета.

Методические указания по проведению практических занятий по дисциплине «Анализ данных в цифровой экономике» предназначены для студентов направления подготовки 02.03.03 «Математическое обеспечение и администрирование информационных систем».

Текст печатается в авторской редакции
Подписано в печать 9.10.2024. Формат 60x84 1/16.
Усл. печ.л. 2,7. Уч. –изд.л. 2,47. Тираж 50 экз. Заказ 1125.
Бесплатно.
Юго-Западный государственный университет.
305040, г. Курск, ул. 50 лет Октября, 94.

# 1. Линейная регрессия. Коэффициент детерминации. Коэффициент корреляции. Его значимость

## 2. Цель и задачи

**Цель работы**: изучить возможности MS Excel для построения парной линейной регрессии и корреляционного анализа.

**3. Задачи**:

– приобрести навыки расчета коэффициента детерминации;

– приобрести навыки расчета коэффициента корреляции и определения его значимости;

– научиться находить коэффициенты регрессии и строить уравнение;

– научиться строить диаграмму рассеяния средствами MS Excel;

– приобрести навык использования статистических функций MS Excel для проведения корреляционного и регрессионного анализа.

## 4. Теоретическая часть

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

$$y=f(x)$$

где *y* – зависимая (эндогенная) переменная;

*x* – независимая (экзогенная), объясняющая переменная.

Различают *линейные* и *нелинейные* регрессии.

Линейная регрессия: $\hat{y}_x = \mathrm{a} + \mathrm{b} \cdot \mathrm{x}$.

Построение уравнения регрессии сводится к оценке ее параметров. Для оценки параметров регрессий, линейных по параметрам, используют метод наименьших квадратов (МНК). МНК позволяет получить такие оценки параметров, при которых сумма квадратов отклонений фактических значений результативного признака *y* от теоретических $y_x$ минимальна.

### *2.1. Определение параметров линейного уравнения регрессии*

Для линейных и нелинейных уравнений, приводимых к линейным, решается следующая система относительно $a$ и $b$:

$$\begin{cases} a \cdot n + b \cdot \sum x = \sum y; \\ a \cdot \sum x + b \cdot \sum x^2 = \sum x \cdot y. \end{cases}$$

Можно воспользоваться готовыми формулами, которые вытекают из этой системы:

$$a = \bar{y} - b \cdot \bar{x},\ \mathrm{b} = \frac{\mathrm{cov(x,y)}}{\sigma_{\mathrm{x}}^2} = \frac{\mathrm{cov(x,y)}}{\mathrm{var(x)}},$$

где $\mathrm{cov}(x,y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) = \overline{xy} - \overline{x} \cdot \overline{y}$ – ковариация признаков x и y,

$\sigma_x^2 = \overline{x^2} - \overline{x^2}$ - дисперсия признака x и

$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{x_1 + x_2 + \ldots + x_n}{n}$, $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{y_1 + y_2 + \ldots + y}{n}$,

$\overline{y \cdot x} = \frac{1}{n}\sum y \cdot x$, $\overline{x^2} = \frac{1}{n}\sum x^2$

$var(x) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \overline{x^2} - \bar{x}^2 = \sigma_x^2$,

$\mathrm{var(y)} = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \overline{\{y\}}\right)^2 = \overline{y^2} - \bar{y}^2 = \sigma_y^2$,

$\sigma_x = \sqrt{\frac{\sum_{k=1}^{n}(x_k - \bar{x})}{n}} = \sqrt{var(x)}$, $\sigma_y = \sqrt{\frac{\sum_{k=1}^{n}(y_k - \bar{y})}{n}} = \sqrt{var(y)}$

Параметр *b* называется коэффициентом регрессии. Его величина показывает среднее изменение результата с изменением фактора на одну единицу.

### *2.2. Расчет коэффициента корреляции*

Тесноту связи изучаемых явлений оценивает линейный коэффициент парной корреляции для линейной регрессии ($-1$ $r_{xy}$ 1):

$$r_{xy} = b \cdot \frac{\sigma_x}{\sigma_y} = \frac{cov(x,y)}{\sigma_x \cdot \sigma_y} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sqrt{(\overline{x^2} - \bar{x}^2)\,(\overline{y^2} - \bar{y}^2)}} = \frac{cov(x,y)}{\sqrt{var(x) \cdot var(xy)}}$$

Теснота линейной связи между переменными может быть оценена на основании шкалы Чеддока:

| Теснота связи | Значение коэффициента корреляции при наличии | |
|---|---|---|
| | прямой связи | обратной связи |
| Слабая | 0,1–0,3 | (–0,3) –(–0,1) |
| Умеренная | 0,3–0,5 | (–0,5) –(–0,3) |
| Заметная | 0,5–0,7 | (–0,7) –(–0,5) |
| Высокая | 0,7–0,9 | (–0,9) –(–0,7) |
| Весьма высокая (сильная) | 0,9–1 | (–1) –(–0,9) |

Положительное значение коэффициента корреляции говорит о положительной связи между *x* и *y*, когда с ростом одной из переменных другая тоже растет. Отрицательное значение коэффициента корреляции означает, с

ростом одной из переменных другая убывает, с убыванием одной из переменной другая растет.

### *2.3. Оценка значимости коэффициента корреляции*

Оценку статистической значимости коэффициента корреляции проводят с помощью *t*-критерия Стьюдента. Выдвигают гипотезу $H_0$ о статистически незначимом отличии коэффициента от нуля. Оценка значимости коэффициента корреляции с помощью t-критерия Стьюдента проводится путем сопоставления его значения с величиной случайной ошибки:

$$t_r = \frac{r}{m_r}$$

Стандартная (случайная) ошибка коэффициента корреляции определяется по формуле:

$$m_r = \sqrt{\frac{1 - r^2}{n - 2}}$$

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

Если $t_{табл} < t_{факт}$, то гипотеза $H_0$ отклоняется. Если $t_{табл} > t_{факт}$, то гипотеза $H_0$ не отклоняется и признается случайная природа формирования коэффициента корреляции.

### *2.4. Расчет коэффициента детерминации*

Коэффициент детерминации характеризует долю дисперсии, объясняемую регрессией, в общей дисперсии результативного признака *y*.

$$R^2 = 1 - \frac{\sum(y_x - y)^2}{\sum(y - \bar{y})^2}$$

Чем ближе коэффициент детерминации к 1, тем выше качество уравнения регрессии, тем в большей мере оно объясняет поведение эндогенной переменной.

## 5. Задание

По предприятиям легкой промышленности региона получена информация, характеризующая зависимость объема выпуска продукции (*y*, млн. руб.) от объема капиталовложений (*x*, млн. руб.)

Таблица 1.1. Варианты для заданий

| № | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | *x* | 66 | 58 | 73 | 82 | 81 | 84 | 55 | 67 | 81 | 59 |
| | *y* | 133 | 107 | 145 | 162 | 163 | 170 | 104 | 132 | 159 | 116 |
| 2 | *x* | 72 | 52 | 73 | 74 | 76 | 79 | 54 | 68 | 73 | 64 |
| | *y* | 121 | 84 | 119 | 117 | 129 | 128 | 102 | 111 | 112 | 98 |
| 3 | *x* | 38 | 28 | 27 | 37 | 46 | 27 | 41 | 39 | 28 | 44 |
| | *y* | 69 | 52 | 46 | 63 | 73 | 48 | 67 | 62 | 47 | 67 |
| 4 | *x* | 36 | 28 | 43 | 52 | 51 | 54 | 25 | 37 | 51 | 29 |
| | *y* | 104 | 77 | 117 | 137 | 143 | 144 | 82 | 101 | 132 | 77 |
| 5 | *x* | 31 | 23 | 38 | 47 | 46 | 49 | 20 | 32 | 46 | 24 |
| | *y* | 38 | 26 | 40 | 45 | 51 | 49 | 34 | 35 | 42 | 24 |
| 6 | *x* | 33 | 17 | 23 | 17 | 36 | 25 | 39 | 20 | 13 | 12 |
| | *y* | 43 | 27 | 32 | 29 | 45 | 35 | 47 | 32 | 22 | 24 |
| 7 | *x* | 36 | 28 | 43 | 52 | 51 | 54 | 25 | 37 | 51 | 29 |
| | *y* | 85 | 60 | 99 | 117 | 118 | 125 | 56 | 86 | 115 | 68 |
| 8 | *x* | 17 | 22 | 10 | 7 | 12 | 21 | 14 | 7 | 20 | 3 |
| | *y* | 26 | 27 | 22 | 19 | 21 | 26 | 20 | 15 | 30 | 13 |
| 9 | *x* | 12 | 4 | 18 | 27 | 26 | 29 | 1 | 13 | 26 | 5 |
| | *y* | 21 | 10 | 26 | 33 | 34 | 37 | 9 | 21 | 32 | 14 |
| 10 | *x* | 26 | 18 | 33 | 42 | 41 | 44 | 15 | 27 | 41 | 19 |
| | *y* | 43 | 28 | 51 | 62 | 63 | 67 | 26 | 43 | 61 | 33 |

По заданной выборке исследовать зависимость результата *y* от фактора *x*:

1. Создать таблицу данных.

2. Найти средние значения $\bar{x}, \bar{y}$ выборочные дисперсии $s_x^2, s_y^2$ исправленные средние квадратические отклонения $\overline{s_x}, \overline{s_y}$.
3. Найти коэффициент корреляции и проверить его значимость.
4. Найти коэффициент детерминации.
5. Найти коэффициенты линейного уравнения регрессии.
6. Дать экономическую интерпретацию значений коэффицента корреляции и параметров уравнения регрессии.
7. Построить диаграмму рассеяния и график уравнения регрессии.

## 6. 4. Методика выполнения заданий

В табл. 1.2 приведены данные об объеме производства *y* (тыс.ед.) в зависимости от численности занятых *x* (тыс.чел.) некоторой фирмы.

Таблица 1.2. Исходные данные

| *x* | 11 | 13 | 15 | 18 | 20 | 22 | 24 | 25 | 27 |
|---|---|---|---|---|---|---|---|---|---|
| *y* | 15 | 17 | 21 | 20 | 28 | 33 | 34 | 32 | 29 |

1. В диапазоне B3:C11 подготовим исходные данные (рис. 1.1).
2. В ячейках D3:D11 рассчитаем произведение *x* и *y*, в ячейках E3:E11 и F3:F11 квадраты значений *x* и *y*, в ячейках B12:F12 с помощью функции СРЗНАЧ рассчитаем средние значения рассмотренных величин.
3. В ячейках A17 и B17 рассчитаем выборочные дисперсии $s_x^2 = \overline{x^2} - \bar{x}^2$, $s_y^2 = \overline{y^2} - \bar{y}^2$
4. В ячейках A21 и B21 рассчитаем исправленные средние квадратические отклонения $\overline{s_x}, \overline{s_y}$. Для этого воспользуемся функцией СТАНДОТКЛОН. Она оценивает стандартное отклонение по выборке (мера того, насколько широко разбросаны точки данных относительно их среднего).
5. В ячейке E16 рассчитаем коэффициент корреляции. Для этого воспользуемся формулой $r_{xy} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sqrt{(\overline{x^2} - \bar{x}^2)(\overline{y^2} - \bar{y}^2)}}$

Из расчетов (рис. 1.2) следует, что коэффициент корреляции r = 0,902. Это свидетельствует о том, что связь между объемом выпуска продукции и численностью занятых весьма высокая и положительная.

| | A | B | C | D | E | F | G |
|---|---|---|---|---|---|---|---|
| 1 | **Простейшая обработка данных** | | | | | | |
| 2 | | *x* | *y* | *xy* | $x^2$ | $y^2$ | |
| 3 | 1 | 11 | 15 | 165 | 121 | 225 | |
| 4 | 2 | 13 | 17 | 221 | 169 | 289 | |
| 5 | 3 | 15 | 21 | 315 | 225 | 441 | |
| 6 | 4 | 18 | 20 | 360 | 324 | 400 | |
| 7 | 5 | 20 | 28 | 560 | 400 | 784 | |
| 8 | 6 | 22 | 33 | 726 | 484 | 1089 | |
| 9 | 7 | 24 | 34 | 816 | 576 | 1156 | |
| 10 | 8 | 25 | 32 | 800 | 625 | 1024 | |
| 11 | 9 | 27 | 29 | 783 | 729 | 841 | |
| 12 | среднее значение | 19,44 | 25,44 | 527,33 | 405,89 | 694,33 | |
| 13 | | | | | | | |
| 14 | | | | | | | |
| 15 | **Выборочные средние** | | | | | | |
| 16 | $S_x^2$ | $S_y^2$ | | | | | |
| 17 | 27,80 | 46,91 | | | | | |
| 18 | | | | | | | |
| 19 | **Исправленные средние квадратичные** | | | | | | |
| 20 | $S_{x\ испр}$ | $S_{y\ испр}$ | | | | | |
| 21 | 5,59 | 7,26 | | | | | |

Рис. 1.1 – Результаты простейшей обработки данных

6. Для проверки значимости коэффициента корреляции введем вспомогательные данные. Ячейка L16 – число предприятий ($n$): 9; ячейка L17 – уровень значимости: 0,05.
7. В ячейке H20 определим стандартную ошибку по следующей формуле:

$$m_r = \sqrt{\frac{1 - r^2}{n - 2}}$$

8. В ячейке H21 рассчитаем значение $t$-статистики по формуле:

$$t_r = \frac{r}{m_r}$$

9. Критическое значение $t$-статистики определим с помощью функции Excel СТЬЮДРАСПОБР. Она возвращает двустороннее обратное $t$-

распределения Стьюдента.

*Синтаксис:* ***СТЬЮДРАСПОБР (вероятность,степени_свободы).***

*Аргументы:*

*Вероятность – вероятность, соответствующая двустороннему распределению Стьюдента;*

*Степени_свободы – число степеней свободы, характеризующее распределение.*

В качестве вероятности укажем уровень значимости. Число степеней равно $n$-$m$-1, где $m$ – число независимых переменных в модели (в нашем случае она всего одна – $x$)

10. Для наглядного отображения вывода воспользуемся функцией ЕСЛИ и условным форматированием: если рассчетное значение t-статистики больше критического, то коэффициент корреляции значим (выделяем зеленым), в противном случае незначим (выделяем красным).
11. В ячейке H16 рассчитаем коэффициент детерминации как квадрат коэффициента корреляции.

| | C | D | E | F | G | H | I | J | K | L |
|---|---|---|---|---|---|---|---|---|---|---|
| 14 | | | | | | | | | | |
| 15 | | **Коэффициент корреляции** | | | **Коэффициент детерминации** | | | **Вспомогательные данные** | | |
| 16 | | $r_{xy}$ | 0,902 | | $R^2_{xy}$ | 0,814 | | n | | 9 |
| 17 | | | | | | | | уровень значимости | | 0,05 |
| 18 | | | | | | | | | | |
| 19 | | **Проверка значимости коэффициента корреляции** | | | | | | | | |
| 20 | | стандартная ошибка | | | | 0,1631 | | | | |
| 21 | | t-статистика | | | | 5,5315 | | | | |
| 22 | | Критическое значение t-статистики | | | | 2,3646 | | | | |
| 23 | | Вывод | | | | Значим | | | | |
| 24 | | | | | | | | | | |

Рис. 1.2 – Расчет коэффициента корреляции, коэффициента корреляции и анализ его значимости

12. Для определения коэффициентов уравнения линейной регрессии на основе формул (рис. 1.3):

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

В нашем примере уравнение имеет вид: $y = 2, + 1{,}1\ 2 \cdot x$

Значение коэффициента $b$=1,172 говорит о том, что при увеличении численности занятых на 1 тыс.чел. объем продукции увеличится на 1,172 тыс.ед.

| | G | H | I | J |
|---|---|---|---|---|
| 1 | | | | |
| 2 | | Коэффициенты регрессии | | |
| 3 | | *b* | 1,172 | |
| 4 | | *a* | 2,659 | |
| 5 | | | | |

Рис. 1.3 – Результаты расчета параметров уравнения регрессии

13. Для построения диаграммы рассеяния выделим диапазон B3:C11. Во вкладке «Вставка» выберем тип диаграммы – «Точечная». На построенной диаграмме выделим нанесенные значения, щелкнув по ним левой кнопкой мыши. Нажав правую кнопку мыши, выведем контекстно-зависимое меню, в котором выберем опцию Добавить линию тренда. В окне Линия тренда по вкладке «Параметры линии тренда» выберем тип функции «Линейная», установим флажок «показывать уравнение на диаграмме» и «поместить на диаграмму величину достоверности аппроксимации (R^2)». В результате на диаграмме появиться вид теоретической кривой – тренда, ее уравнение и коэффициент детерминации (рис.1.4). Добавим подписи осей, заголовок диаграммы и легенду.

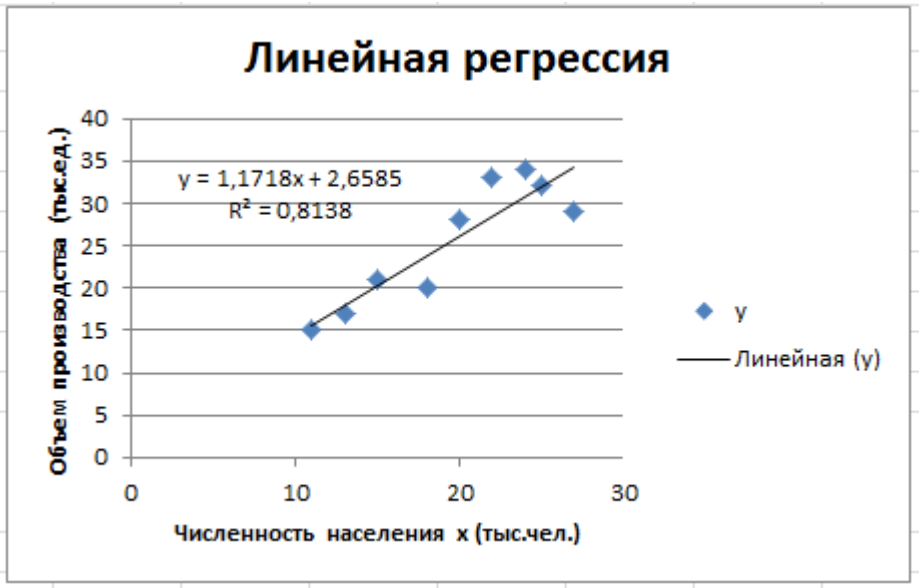



Рис. 1.4 – Графики фактических данных и построенной регрессии

14.Вычисление параметров регрессии с помощью статистических функций Excel:

КОРРЕЛ(массив1;массив2) вычисляет коэффициент корреляции между двумя переменными; значения первой из них приведены в диапазоне массив1, значения второй – в диапазоне массив2;

НАКЛОН(известные_значения_y;известные_значения_x) служит для определения коэффициента b;

ОТРЕЗОК(известные_значения_y;известные_значения_x) служит для определения коэффициента a.

Рассчитаем с их помощью коэффициент корреляции в ячейке E27, параметры *a* и *b* соответственно в ячейках E28 и E29 (рис. 1.5).

| | A | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|---|
| 25 | **Расчет параметров регрессии с помощью статистических функций Excel** | | | | | | | |
| 26 | **Линейн** | | | | | | | |
| 27 | 1,171847 | 2,658526 | | $r_{xy}$ | 0,902118 | | | |
| 28 | 0,21185 | 4,268075 | | *b* | 1,171847 | | | |
| 29 | 0,813817 | 3,351131 | | *a* | 2,658526 | | | |
| 30 | 30,59743 | 7 | | | | | | |
| 31 | 343,6117 | 78,61057 | | | | | | |

Рис. 1.5 – Расчет параметров регрессии с помощью функций Excel

15.Встроенная статистическая функция ЛИНЕЙН определяет параметры линейной регрессии. Порядок вычислений следующий:

1) выделите ячейку A27, нажмите на кнопку «Вставить функцию» ( $f_x$ );

2) в строке Категория (рис.1.6) выберите Статистические, в окне Функция – ЛИНЕЙН. Щелкните ОК.

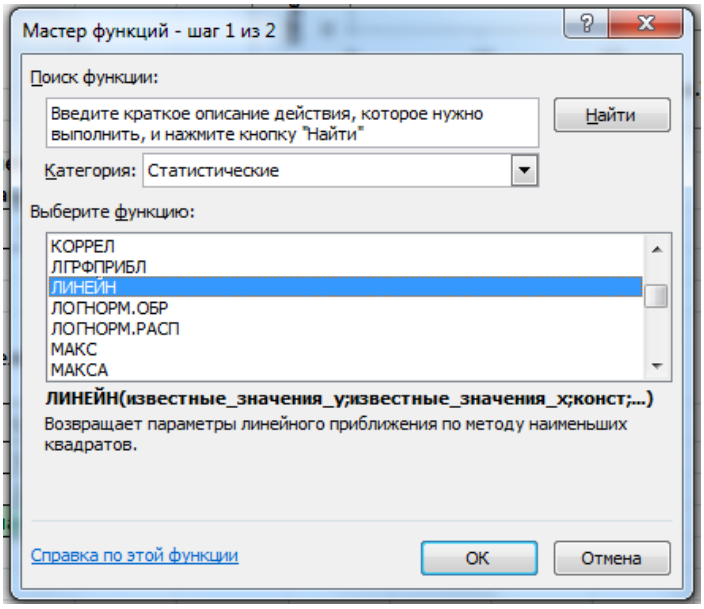



Рис. 1.6 – Диалоговое окно «Мастер функций»

4) Заполните аргументы функции (рис.1.7.):

Известные_значения_y – диапазон, содержащий данные результативного признака;

Известные_значения_x – диапазон, содержащий данные факторов независимого признака;

Константа – логическое значение, которое указывает на наличие или на отсутствие свободного члена в уравнении; если Константа = 1, то свободный член рассчитывается обычным образом, если Константа = 0, то свободный член равен 0.

Статистика – логическое значение, которое указывает выводить дополнительную информацию по регрессионному анализу или нет. Если

Статистика = 1, то дополнительная информация выводится, если Статистика = 0, то выводится только оценки параметров уравнения.

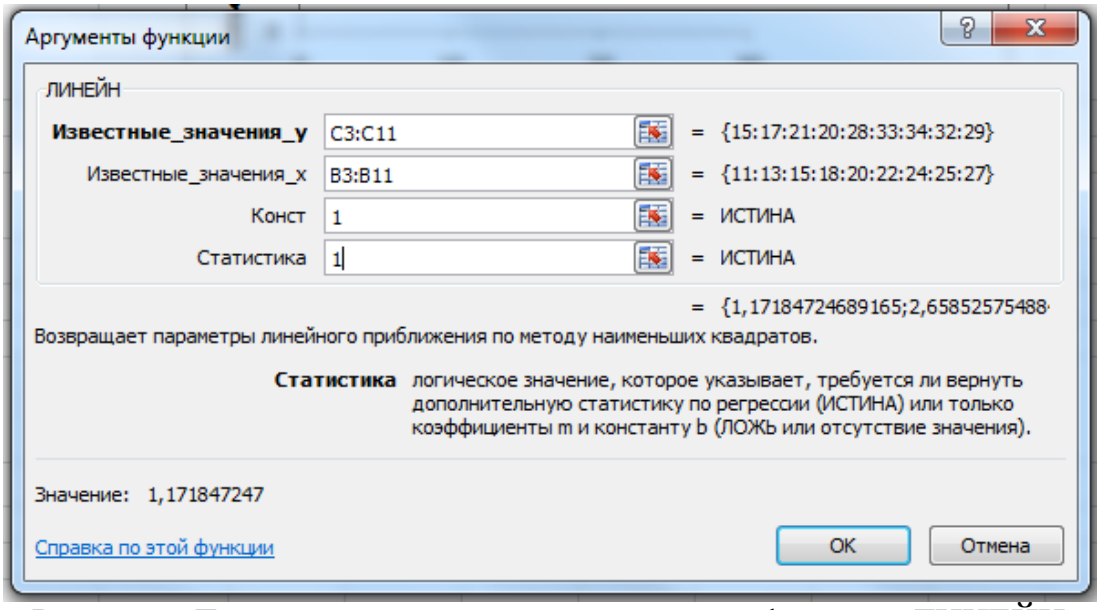



Рис. 1.7 – Диалоговое окно ввода аргументов функции ЛИНЕЙН

5) появится первый элемент итоговой таблицы. Чтобы вывести результаты регрессионной статистики, выделите область пустых ячеек 5x2 (A27:B31). Нажмите на клавишу F2, а затем на комбинацию клавиш CTRL+SHIFT+ENTER. Дополнительная регрессионная статистика будет выводиться в порядке, указанном в следующей схеме:

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

Результаты регрессионного анализа представлены на рис.1.5.

## 7. Контрольные вопросы

1. Какова экономическая интерпретация параметров уравнения регрессии?
2. Что означает отрицательное значение коэффициента корреляции?
3. Назовите диапазон изменения значений коэффициента корреляции и коэффициента детерминации
4. Что является показателем тесноты связи в парной линейной регрессии?
5. Каково значение коэффициента корреляции?
6. Каково значение коэффициента детерминации и что он характеризует?
7. Как оценивается значимость коэффициента корреляции?
8. Какие функции Excel можно использовать для определения параметров линейного уравнения регрессии?
9. Какие функции Excel можно использовать для определения коэффициен корреляции?
10. Для чего используется функция СТЬЮДРАСПОБР?

## 8. Требования к содержанию отчета

Отчет к лабораторной работе предоставляется в электронном виде и должен содержать:

- название и цель работы;
- номер и исходные данные своего варианта;
- описание хода выполнения заданий, в том числе:
  - скриншоты из MS Excel, отображающие заданные в ячейках формулы и функции, использованные для вычислений;
  - скриншоты из MS Excel с полученными при расчетах результатами;
  - анализ и интерпретация полученных результатов;
- выводы по лабораторной работе.

## 9. Требования к оформлению отчета

- Все рисунки в отчете должны быть подписаны.
- Все скриншоты должны быть читаемыми в масштабе документа 100%. При необходимости используйте обрезку и пропорциональное изменение размера рисунка.
- Все скриншоты таблиц MS Excel должны содержать системное наименование строк (1, 2, 3….) и столбцов (A, B, C, …)

**Построение многофакторной линейной регрессии с помощьюпакета Анализ данных MS Excel. Анализ остатков.**

## 1. Цель и задачи лабораторной работы

**Цель работы**: изучить возможности MS Excel для построения многофакторной линейной регрессии, оценки ее качества и анализа остатков.

16. **Задачи**:

– научиться проверять факторы на мультиколлинеарность;

– приобрести навыки нахождения параметров уравнения многофакторной линейной регрессии в естественной и стандартизованной форме;

– научиться ранжировать факторы по силе их воздействия на результат;

– научиться проверять качество уравнения многофакторной линейной регрессии;

– научиться проводить анализ остатков на выполнение пяти предпосылок метода наименьших квадратов.

## 1. Теоретическая часть

*Множественная регрессия* – уравнения связи с несколькими независимыми переменными:

$$y = f(x_1, x_2, \dots, x_m),$$

где $y$ – зависимая переменная (результативный признак);

$x_1, x_2, \dots, x_m$ – независимые переменные (факторы).

Множественная регрессия применяется в ситуациях, когда из множества факторов, влияющих на результативный признак, нельзя выделить один доминирующий фактор и необходимо учитывать влияние нескольких факторов.

Основная цель множественной регрессии – построить модель с большим числом факторов, определив при этом влияние каждого из них в отдельности, а также совокупное их воздействие на моделируемый показатель.

### *2.1. Отбор факторов при построении множественной регрессии*

Включение в уравнение множественной регрессии того или иного набора факторов связано, прежде всего, с представлением исследователя о природе взаимосвязи моделируемого показателя с другими экономическими явлениями.

Факторы, включаемые во множественную регрессию, должны отвечать следующим требованиям:

1. Факторы должны быть количественно измеримы.

2. Факторы не должны быть взаимно коррелированы. Если между факторами существует высокая корреляция, то нельзя определить их

изолированное влияние на результативный показатель, и параметры уравнения регрессии оказываются неинтерпретируемыми.

Коэффициенты интеркорреляции (т. е. корреляции между объясняющими переменными) позволяют исключать из модели дублирующие факторы. Считается, что две переменные явно коллинеарные, т. е. находятся между собой в линейной зависимости, если $\left|r_{x_i x_j}\right| \geq 0{,}7$.

Если факторы явно коллинеарны, то они дублируют друг друга и один из них рекомендуется исключить из регрессии. Предпочтение при этом отдается тому фактору, который при достаточно тесной связи с результатом имеет наименьшую тесноту связи с другими факторами.

3. Факторы должны иметь заметную связь с результирующей переменной, т.е. $\left|r_{yx_i}\right| \geq 0{,}4$.

### *2.2. Оценка параметров уравнения множественной регрессии*

Для оценки параметров уравнения множественной регрессии применяют метод наименьших квадратов (МНК). В результате мы получаем линейное уравнение регрессии в естественном виде:

$$y = b_0 + b_1 x_1 + \cdots + b_m x_m + u.$$

Для определения значимости факторов и повышения точности результата используется уравнение множественной регрессии в стандартизованном масштабе:

$$t_y = \beta_0 + \beta_1 t_{x1} + \cdots + \beta_m t_{xm} + u,$$

где $t_y,\ t_{x1}, \ldots, t_{xm}$ - стандартизованные

$$t_y = \frac{y - \bar{y}}{\sigma_y}, t_{x_i} = \frac{x_i - \overline{x_i}}{\sigma_{x_i}},$$

для которых среднее значение равно нулю $\bar{t}_y = \overline{t_{x_i}} = 0$, а среднее квадратическое отклонение равно единице$\sigma_y = \sigma_{x\mathrm{i}} = 1$.

Величины $\beta_i$ называются стандартизованными коэффициентами регрессии. Они показывают, на сколько сигм (средних квадратических отклонений) изменится в среднем результат, если соответствующий фактор $x_i$ изменится на одну сигму при неизменном среднем уровне других факторов. В силу того, что все переменные заданы как центрированные и нормированные, стандартизованные коэффициенты регрессии $\beta_i$ сравнимы между собой. Сравнивая их друг с другом, можно ранжировать факторы по силе их воздействия на результат. В этом основное достоинство стандартизованных коэффициентов регрессии в отличие от коэффициентов регрессии в естественном виде, которые несравнимы между собой.

Связь коэффициентов множественной регрессии $b_i$ со стандартизованными коэффициентами $\beta_i$ описывается соотношением

$$b_i = \beta_i \frac{\sigma_y}{\sigma_{x_i}}$$

Параметр $b_0$ определяется из соотношения: $b_0 = \overline{y} - b_1\overline{x}_1 - \cdots - b_m\overline{x}_m$.

Средние коэффициенты эластичности для линейной множественной регрессии рассчитываются по формуле

$$\overline{Э}_{yx_i} = b_J \cdot \frac{\overline{x_J}}{\overline{y}}$$

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

### *2.3. Множественная корреляция*

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

1. *Линейный коэффициент множественной корреляции* оценивает теснотуфакторов на результат и может быть рассчитан по формуле:

$$R_{yx_1x_2\ldots x_m} = \sqrt{\frac{1 - \sum\left(y - y_{yx_1x_2\ldots x_m}\right)^2}{\sum(y - \bar{y})^2}}$$

Низкое значение коэффициента множественной корреляции означает, что в регрессионную модель не включены существенные факторы – с одной стороны, а с другой стороны – рассматриваемая форма связи не отражает реальные соотношения между переменными, включенными в модель.

В этом случае требуются дальнейшие исследования по улучшению качества модели и увеличению ее практической значимости.

2. *Коэффициент детерминации* – ещё один показатель качества подгонки. $0 \leq R^2 \leq 1$, чем ближе $R^2$ к 1, тем лучше регрессионное уравнение (т.е. качество подгонки).

3. *Скорректированный коэффициент детерминации*. В многофакторном

регрессионном уравнении добавление дополнительных объясняющих увеличивает коэффициент детерминации. Следовательно, коэффициент детерминации должен быть скорректирован с учетом числа независимых переменных:

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

Чем больше величина $m$, тем сильнее различия $\bar{R}^2$ и $R^2$.

### *2.4. Оценка надежности результатов множественной регрессии и корреляции*

Значимость уравнения множественной регрессии в целом, так же как и в парной регрессии, оценивается с помощью *F-критерия Фишера*. Он состоит в проверке гипотезы $H_0$ о статистической незначимости уравнения регрессии и показателя тесноты связи. Для этого выполняется сравнение фактического $F_{факт}$ и критического (табличного) $F_{табл}$ значений F-критерия Фишера. $F_{факт}$ определяется из соотношения значений факторной иостаточной дисперсий, рассчитанных на одну степень свободы:

$$F = \frac{R^2}{1-R^2} \cdot \frac{n \cdot m - 1}{m}$$

Число степеней свободы критерия Фишера: $k_1=m$; $k_2=n-m-1$ и критическое значение этого критерия $F = F_{\alpha;k_1;k_2}$

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

Оценка значимости коэффициентов регрессии производится с помощью *t-критерия Стьюдента*. Вычисляются наблюдаемые значения *t*-статистики:

$$t_j = \frac{b_j}{m_{b_j}}$$

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

$$m_{b_j} = \frac{\sigma_y \cdot \sqrt{1 - R^2_{yx_1 \ldots x_m}}}{\sigma_{x_j} \cdot \sqrt{1 - R^2_{x_j x_1 \ldots x_m}}} \cdot \frac{1}{\sqrt{n - m - 1}}$$

Критическое значение t-статистики: $t_{кр} = t_{\alpha;n-m-1}$ где $k = n–m–1$ – число степеней свободы, $m$ – число факторов; Если $|t_{набл}| > t_{кр}$, то коэффициент регрессии статистически значим; в противном случае – статистически незначим.

***2.5. Анализ остатков***

Исследование остатков $u_i$ предполагает проверку наличия следующих пяти предпосылок МНК:

1) Случайный характер остатков;
2) Нулевая средняя величина остатков, не зависящая от $x_i$;
3) Гомоскедастичность;
4) Отсутствие автокорреляции остатков;
5) Остатки подчиняются нормальному закону распределения.

Для проверки *первой предпосылки* – случайного характера остатков – строится график зависимости остатков $u_i$ от теоретического значения результативного признака $\hat{y}_x$. Если на графике получена горизонтальная полоса (рис. 4.1а), внутри которой остатки расположены случайным образом, то первая предпосылка выполняется.

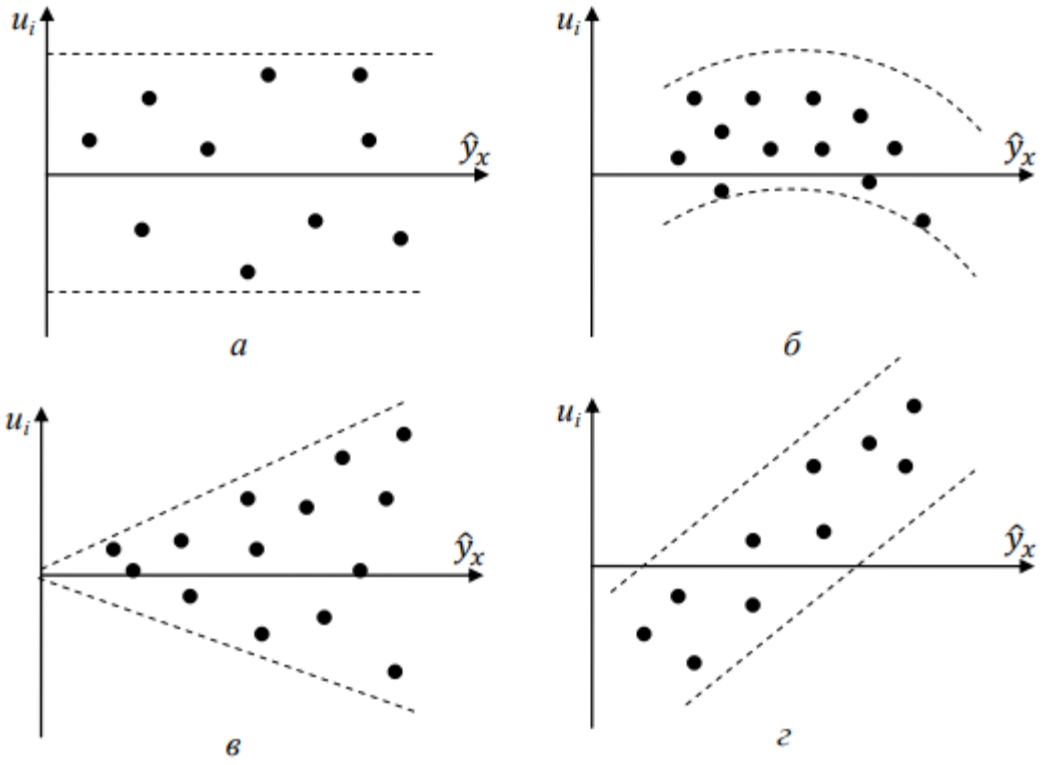



Рис. 4.1. Зависимость случайных остатков $u_i$ от теоретических значений $\hat{y}_x$

Возможны следующие случаи: если $u_i$ зависит от $\hat{y}_x$, то:

- Остатки $u_i$ не случайны (рис. 4.1б);
- Остатки $u_i$ не имеют постоянной дисперсии (рис. 4.1в);
- Остатки $u_i$ носят систематический характер (рис. 4.1г).

В этих случаях необходимо либо применить другую функцию, либо вводить дополнительную информацию и заново строить уравнение регрессии.

*Вторая предпосылка МНК* заключается в равенстве нулю средних значений остатков и независимости их от факторов. Она обеспечивает несмещенность оценок. Для ее проверки строится график зависимости случайных остатков $u_i$ от факторов, включенных в регрессию – $x_i$. Если на графике получена горизонтальная полоса, внутри которой остатки расположены случайным образом, то вторая предпосылка выполняется. Если же график показывает наличие зависимости $u_i$ от $x_i$, то модель неадекватна.

В соответствии с *третьей предпосылкой МНК* требуется, чтобы дисперсия остатков была гомоскедастичной. Это означает, что для каждого значения фактора остатки имеют одинаковую дисперсию. Если это условие не соблюдается, то имеет место гетероскедастичность. Наличие

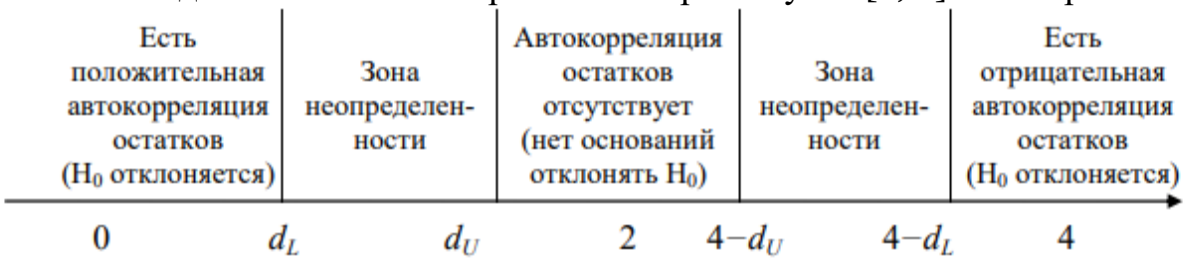
гомоскедастичности или гетероскедастичности можно видеть по графику зависимости остатков $u_i$ от теоретического значения результативного признака
$\hat{y}_x$. Так на рис. 4.1а остатки гомоскедастичны, а на рис. 4.1в – гетероскедастичны.

*Четвертая предпосылка МНК* – отсутствие автокорреляции остатков означает, что остатки $u_i$ распределены независимо друг от друга. Автокорреляция остатков означает наличие корреляции между остатками текущих и предыдущих наблюдений. Для проверки выполнения четвертой предпосылки можно воспользоваться критерием Дарбина-Уотсона:

- Выдвигается гипотеза $H_0$ об отсутствии автокорреляции в остатках.
- Рассчитывается статистика DW по формуле:

$$DW = \frac{\sum_{i=2}^{n}(u_i - u_{i-1})^2}{\sum_{1=1}^{n} u_i^2} \; ; \; 0 \le DW \le 4$$

- По специальным таблицам (см. Приложение 1) определяются критические значения Дарбина-Уотсона $d_L$ и $d_U$ для заданного числа наблюдений $n$, числа независимых переменных $m$ и уровня значимости $\alpha$.
- По найденным значениям разбиваем промежуток [0; 4] на 5 отрезков.




- Если расчетное значение попадает в зону неопределенности, то подтверждается существование автокорреляции и гипотезу $H_0$ отклоняют.

Для проверки *пятой предпосылки* о нормальном распределении остатков используют Q-Q график. Строятся квантили распределения остатков относительно нормального распределения (квантили – это значения, которые делят случаи на ряд групп одинакового размера). Если точки, соответствующие наблюдаемым данным, образуют прямую, проведенную из левого нижнего угла в правый верхний угол, значит, данные распределены приближенно нормально. С другой стороны, если эти точки отклоняются от прямой линии, распределение данных отличается от нормального.

Предпосылка о нормальном распределении остатков позволяет проводить

проверку параметров регрессии и корреляции с помощью критериев *t*, *F*. Вместе

с тем оценки регрессии, найденные с применением МНК, обладают хорошими свойствами даже при отсутствии нормального распределения остатков, т.е. при нарушении пятой предпосылки.

## 3. Задание

Для выданного варианта с помощью пакета Анализ данныхMS Excel построить модель множественной регрессии и проверить ее качество. Для этого необходимо:

1. Выбрать факторы для включения в модель:
    1) рассчитать коэффициенты парной корреляции (с помощью инструмента Корреляция):
        - между зависимой и независимыми переменными,
        - независимых переменных между собой;
    2) исключить незначимые факторы;
    3) исключить мультиколлинеарность;
2. Для выбранных факторов построить линейное уравнение множественной регрессии (с помощью инструмента Регрессия):
    1) определить коэффициенты уравнения множественной регрессии в естественной и стандартизованной форме;
    2) определить коэффициенты эластичности; на основе коэффициентов стандартизованного уравнения и коэффициентов эластичности сравнить степень влияниякторов на эндогенную переменную
3. Проверить качество построенного уравнения (с помощью инструмента Регрессия):
    1) исследовать качество модели на основе коэффициента множественной корреляции, коэффициента детерминации, скорректированного коэффициента детерминации;
    2) проверить значимость уравнения с помощью *F*-критерия Фишера (вместо расчета критического значения *F* использовать столбец Значимость *F*);
    3) проверить значимость параметров уравнения с помощью *t*-критерия Стьюдента(вместо расчета критического значения *t* использовать столбец *p*-значение)
4. Провести анализ остатков на выполнение пяти предпосылок МНК:
    1) проверить выполнение предпосылки о случайном характере остатковс помощью точечной диаграммы зависимости остатков от *y*;
    2) проверить выполнение предпосылки о нулевой средней величине остатков с помощью Графика остатков инструмента Регрессия;

3) проверить выполнение предпосылки о гомоскедастичности;
4) проверить выполнение предпосылки об отсутствии автокорреляции остатков с помощью теста Дарбина-Уотсона;
5) проверить выполнение предпосылки о нормальном распределении остатков с помощью Графика нормальной вероятности инструмента Регрессия.

***Варианты заданий***

Исследуется производительность труда на предприятиях одной из отраслей.

$y$ – производительность труда на предприятии, руб./чел.
$x_1$ – средняя заработная плата, руб.
$x_2$ – доля высококвалифицированных работников, %
$x_3$ – инвестиции в основные фонды в текущем квартале, тыс.руб.
$x_4$ – инвестиции в основные фонды в предыдущем квартале, тыс.руб.

Вариант 1

| $y$ | $x_1$ | $x_2$ | $x_3$ | $x_4$ |
|---|---|---|---|---|
| 14163,4 | 14343,9 | 37,9 | 3712,08 | 2896,18 |
| 15700,9 | 15736,6 | 42,2 | 2750,49 | 3299,60 |
| 10191,4 | 15851,6 | 43,7 | 1811,41 | 1520,13 |
| 15172,9 | 26047,8 | 25,3 | 2568,23 | 2017,30 |
| 8180,0 | 19781,5 | 36,8 | 1071,07 | 543,27 |
| 7871,5 | 14780,0 | 23,4 | 2297,24 | 1356,53 |
| 13750,9 | 18226,3 | 30,6 | 3811,98 | 2507,30 |
| 21531,9 | 23431,7 | 28,1 | 3609,77 | 2871,25 |
| 18951,8 | 18575,0 | 40,8 | 3796,96 | 2531,42 |
| 14812,4 | 14061,3 | 33,2 | 2801,13 | 2202,12 |
| 9863,4 | 19419,2 | 28,5 | 2927,71 | 2826,32 |
| 9461,9 | 17065,5 | 38,7 | 1768,73 | 2697,51 |
| 9446,1 | 13195,9 | 46,2 | 2630,41 | 1816,78 |
| 13895,7 | 14720,8 | 34,0 | 3592,92 | 2456,30 |
| 9111,3 | 18256,5 | 36,7 | 2991,88 | 2379,43 |
| 10593,1 | 12189,2 | 37,2 | 3346,85 | 2249,65 |
| 16296,4 | 20172,3 | 40,3 | 2918,41 | 2921,22 |
| 10426,8 | 12648,4 | 40,9 | 2755,49 | 2601,14 |
| 6068,3 | 14139,8 | 40,0 | 1094,32 | 778,60 |
| 11125,6 | 14866,3 | 42,0 | 2332,58 | 2225,41 |

Вариант 2

| $y$ | $x_1$ | $x_2$ | $x_3$ | $x_4$ |
|---|---|---|---|---|
| 15537,2 | 24985,3 | 39,2 | 2719,24 | 3154,72 |
| 20492,3 | 20866,0 | 60,0 | 1000,19 | 2391,42 |
| 25377,1 | 26352,3 | 62,0 | 1643,73 | 1789,64 |
| 16829,4 | 18454,5 | 62,2 | 1839,19 | 3330,46 |
| 23735,4 | 22073,7 | 65,8 | 2037,06 | 2077,28 |
| 31366,0 | 29346,9 | 48,5 | 1486,61 | 3327,56 |
| 16009,4 | 25767,9 | 50,7 | 2075,16 | 3348,70 |
| 21754,7 | 25919,2 | 47,5 | 1845,89 | 4004,76 |
| 20210,3 | 21953,3 | 59,2 | 1518,67 | 2652,36 |
| 50740,4 | 29218,5 | 52,2 | 2077,48 | 4432,78 |
| 5162,7 | 27258,4 | 26,4 | 2048,12 | 2949,15 |
| 13057,4 | 28558,2 | 30,1 | 1905,20 | 2717,73 |
| 11535,7 | 29302,3 | 44,6 | 1387,90 | 1297,64 |
| 49570,9 | 25787,5 | 71,1 | 2018,61 | 3961,47 |
| 16784,6 | 28079,8 | 38,6 | 2587,69 | 4165,44 |
| 12189,6 | 24380,5 | 56,4 | 1651,15 | 2271,13 |
| 8640,0 | 23793,8 | 31,2 | 1867,40 | 3891,67 |

Вариант 3

| $y$ | $x_1$ | $x_2$ | $x_3$ | $x_4$ |
|---|---|---|---|---|
| 7948,1 | 6773,9 | 34,8 | 1285,63 | 838,99 |
| 11193,7 | 17663,1 | 23,3 | 1816,52 | 1133,25 |
| 9923,6 | 15657,6 | 18,1 | 1523,58 | 1722,58 |
| 6441,2 | 14918,7 | 20,6 | 1539,46 | 1270,27 |
| 8024,9 | 14334,5 | 20,2 | 724,16 | 728,06 |
| 5094,2 | 9059,6 | 18,0 | 1036,19 | 1314,51 |
| 7755,2 | 16329,4 | 22,8 | 1773,48 | 396,21 |
| 6641,0 | 6668,6 | 27,6 | 1245,01 | 1176,72 |
| 9501,4 | 25650,8 | 16,5 | 1020,47 | 1715,50 |
| 11984,9 | 18391,0 | 18,9 | 1937,11 | 2354,55 |
| 6794,2 | 9919,6 | 12,8 | 253,16 | 1267,62 |
| 8462,2 | 12633,8 | 23,2 | 1060,90 | 1961,94 |
| 6572,9 | 15966,7 | 19,0 | 866,50 | -314,37 |
| 6952,0 | 11180,3 | 25,0 | 1298,73 | 2320,66 |
| 7288,0 | 12720,7 | 24,6 | 1626,90 | 1119,05 |
| 7364,4 | 19859,3 | 20,7 | 931,76 | 938,37 |

| $y$ | $x_1$ | $x_2$ | $x_3$ | $x_4$ |
|---|---|---|---|---|
| 5742,0 | 11110,2 | 32,1 | 1057,05 | 49,36 |
| 10202,2 | 19039,9 | 24,2 | 1683,96 | 1386,83 |
| 7895,2 | 14635,7 | 22,4 | 1202,91 | 711,93 |
| 6708,9 | 12762,5 | 18,8 | 154,19 | -200,88 |
| 10847,2 | 26646,4 | 21,0 | 739,32 | 1512,59 |
| 9407,9 | 17212,6 | 25,2 | 1646,26 | 2049,85 |
| 7807,5 | 12963,8 | 21,9 | 1814,23 | 1227,20 |
| 7775,7 | 14981,8 | 18,1 | 1046,08 | 1566,06 |
| 7921,4 | 14639,8 | 22,2 | 1784,37 | 804,19 |
| 6360,7 | 12833,3 | 15,7 | 689,68 | 830,51 |
| 9529,4 | 16760,3 | 21,3 | 1722,09 | 2505,53 |
| 8994,1 | 20765,1 | 25,1 | 756,30 | 1219,31 |
| 6309,3 | 11609,7 | 29,7 | 637,08 | 527,81 |
| 7422,4 | 13595,5 | 21,9 | 1035,73 | 1363,77 |

Вариант 4

| $y$ | $x_1$ | $x_2$ | $x_3$ | $x_4$ |
|---|---|---|---|---|
| 4602,2 | 11560,5 | 23,2 | 295,65 | 283,31 |
| 7745,3 | 10372,9 | 35,2 | 287,61 | 289,13 |
| 10413,8 | 12919,9 | 33,4 | 261,21 | 297,93 |
| 11773,7 | 11226,2 | 34,2 | 364,04 | 310,51 |
| 4365,0 | 11727,1 | 33,2 | 295,66 | 242,00 |
| 5691,7 | 11118,0 | 35,5 | 269,15 | 244,29 |
| 3795,0 | 11044,6 | 19,8 | 335,76 | 279,33 |
| 5414,8 | 11779,0 | 29,5 | 322,85 | 269,90 |
| 8040,2 | 12480,7 | 33,1 | 315,13 | 268,38 |
| 5730,2 | 10252,2 | 25,0 | 340,29 | 280,09 |
| 6712,6 | 12587,7 | 37,6 | 296,42 | 230,84 |
| 5911,2 | 9526,3 | 30,3 | 353,10 | 285,15 |
| 7329,2 | 10961,5 | 25,5 | 387,98 | 307,35 |
| 7044,2 | 11703,5 | 32,9 | 300,27 | 262,46 |
| 7107,5 | 10805,3 | 34,1 | 237,24 | 270,81 |
| 9348,3 | 12377,4 | 36,1 | 296,18 | 265,31 |
| 7216,9 | 10787,1 | 35,4 | 352,77 | 275,83 |
| 7049,4 | 10673,8 | 38,4 | 288,29 | 285,30 |
| 6661,1 | 9542,5 | 31,4 | 396,93 | 280,33 |
| 10387,3 | 11788,5 | 30,7 | 348,83 | 292,49 |
| 5013,5 | 9838,4 | 31,8 | 374,49 | 261,34 |
| 5433,4 | 12238,0 | 33,6 | 279,90 | 254,70 |

| $y$ | $x_1$ | $x_2$ | $x_3$ | $x_4$ |
|---|---|---|---|---|
| 4129,6 | 11888,8 | 36,1 | 284,75 | 237,54 |
| 6275,3 | 12030,8 | 22,4 | 340,93 | 283,29 |
| 5426,6 | 11847,1 | 22,9 | 308,16 | 279,93 |
| 5998,8 | 12298,2 | 29,9 | 303,35 | 265,64 |
| 5967,8 | 9156,8 | 28,0 | 352,87 | 283,18 |

Вариант 5

| $y$ | $x_1$ | $x_2$ | $x_3$ | $x_4$ |
|---|---|---|---|---|
| 7576,3 | 14239,9 | 38,1 | 959,32 | 711,85 |
| 6860,7 | 14024,0 | 31,6 | 636,33 | 431,74 |
| 13413,8 | 15202,1 | 33,0 | 1177,94 | 911,70 |
| 11588,1 | 18117,9 | 28,3 | 1311,09 | 834,23 |
| 12785,8 | 11918,9 | 34,0 | 2286,90 | 3308,22 |
| 19131,5 | 18692,5 | 34,1 | 2864,54 | 2559,67 |
| 15116,4 | 17185,1 | 62,8 | 1668,25 | 1761,92 |
| 13044,9 | 19210,9 | 66,5 | 545,31 | 476,47 |
| 16080,2 | 18898,9 | 35,0 | 760,93 | 1117,19 |
| 20117,6 | 20344,0 | 36,5 | 2266,12 | 1576,40 |
| 9238,6 | 18141,8 | 40,9 | 1218,70 | 701,55 |
| 10899,3 | 14667,1 | 38,6 | 1260,19 | 558,17 |
| 13373,5 | 15595,1 | 47,8 | 1184,23 | 1765,73 |
| 5661,4 | 13666,4 | 33,0 | 1098,98 | 1608,25 |
| 13339,8 | 20281,1 | 33,6 | 1182,13 | 1354,50 |
| 12377,8 | 18380,2 | 19,1 | 1462,09 | 1677,86 |
| 8922,9 | 17641,2 | 38,2 | 935,32 | 187,81 |
| 13192,2 | 15660,5 | 31,0 | 1719,43 | 2209,61 |
| 11817,4 | 16087,6 | 46,3 | 1599,66 | 2114,28 |
| 11612,6 | 15398,3 | 26,1 | 2815,79 | 3047,27 |
| 11377,3 | 17277,6 | 22,6 | 1917,00 | 1410,80 |
| 9771,1 | 15791,8 | 28,6 | 499,24 | 1176,23 |

Исследуется динамика цен на первичном рынке жилья.

$y$ – индекс цен на первичном рынке жилья, %

$x_1$ – среднедушевой доход населения, руб.

$x_2$ – индекс цен на строительные материалы в текущем году, %

$x_3$ – индекс цен на строительные материалы в предыдущем году, %

$x_4$ – индекс цен на строительно-монтажные работы, %

Вариант 6

| $y$ | $x_1$ | $x_2$ | $x_3$ | $x_4$ |
|---|---|---|---|---|
| 110,7 | 14780,6 | 104,2 | 107,2 | 117,3 |
| 118,0 | 16468,8 | 105,1 | 107,9 | 121,9 |
| 121,8 | 16608,1 | 108,0 | 112,1 | 110,8 |
| 109,8 | 28968,7 | 104,3 | 107,4 | 109,9 |
| 117,2 | 21371,5 | 109,3 | 107,9 | 104,2 |
| 74,1 | 15308,7 | 93,8 | 100,3 | 89,5 |
| 121,6 | 19486,5 | 107,8 | 112,3 | 118,9 |
| 133,1 | 25796,8 | 108,1 | 111,0 | 122,5 |
| 112,1 | 19908,9 | 102,3 | 108,5 | 111,5 |
| 121,4 | 14437,9 | 104,9 | 109,9 | 112,3 |
| 103,3 | 20932,7 | 103,6 | 109,3 | 115,8 |
| 85,5 | 18079,4 | 97,2 | 103,7 | 105,7 |
| 99,8 | 13389,8 | 102,3 | 107,2 | 105,3 |
| 126,6 | 15237,2 | 109,2 | 110,5 | 120,5 |
| 87,9 | 19523,0 | 99,5 | 106,5 | 106,2 |
| 105,5 | 12168,9 | 104,5 | 105,1 | 112,2 |
| 114,5 | 21844,9 | 102,5 | 113,3 | 115,1 |
| 127,7 | 12725,8 | 109,9 | 113,0 | 122,8 |
| 94,1 | 14533,3 | 100,8 | 105,2 | 94,4 |
| 105,5 | 12168,9 | 104,5 | 105,1 | 112,2 |
| 114,5 | 21844,9 | 102,5 | 113,3 | 115,1 |
| 127,7 | 12725,8 | 109,9 | 113,0 | 122,8 |
| 94,1 | 14533,3 | 100,8 | 105,2 | 94,4 |
| 101,6 | 15414,0 | 100,0 | 110,2 | 105,6 |

Вариант 7

| $y$ | $x_1$ | $x_2$ | $x_3$ | $x_4$ |
|---|---|---|---|---|
| 106,6 | 17634,7 | 105,7 | 102,1 | 107,2 |
| 110,4 | 8710,1 | 110,7 | 102,7 | 108,2 |
| 100,9 | 20596,6 | 101,2 | 100,8 | 101,2 |
| 97,7 | 3485,5 | 101,1 | 101,6 | 105,0 |
| 97,9 | 11327,0 | 98,8 | 102,0 | 100,5 |
| 119,8 | 27085,0 | 111,8 | 103,2 | 111,2 |
| 106,8 | 19330,5 | 107,4 | 103,7 | 108,7 |
| 104,7 | 19658,3 | 102,8 | 100,1 | 107,7 |
| 103,8 | 11065,8 | 104,6 | 101,8 | 105,3 |
| 130,3 | 26806,7 | 111,8 | 105,7 | 114,0 |
| $y$ | $x_1$ | $x_2$ | $x_3$ | $x_4$ |

| | | | | |
|---|---|---|---|---|
| 105,4 | 22559,8 | 106,5 | 99,6 | 107,2 |
| 108,9 | 25376,2 | 103,7 | 101,2 | 104,9 |
| 105,2 | 26988,3 | 105,4 | 101,9 | 102,4 |
| 113,3 | 19373,2 | 107,5 | 101,2 | 110,3 |
| 103,8 | 24339,6 | 102,6 | 100,9 | 107,9 |
| 101,9 | 16324,5 | 107,4 | 102,3 | 106,0 |
| 109,4 | 15052,9 | 105,3 | 102,4 | 108,8 |
| 107,6 | 11001,8 | 101,0 | 101,0 | 105,6 |
| 103,1 | 20705,9 | 102,7 | 99,1 | 105,1 |
| 103,6 | 15901,9 | 106,5 | 103,6 | 102,0 |

Вариант 8

| $y$ | $x_1$ | $x_2$ | $x_3$ | $x_4$ |
|---|---|---|---|---|
| 126,2 | 9352,6 | 122,6 | 108,3 | 112,0 |
| 123,2 | 19708,6 | 119,7 | 99,6 | 113,4 |
| 133,3 | 17801,2 | 127,1 | 107,6 | 121,3 |
| 101,8 | 17098,2 | 114,0 | 101,9 | 112,4 |
| 114,0 | 16542,6 | 117,6 | 99,9 | 109,1 |
| 109,4 | 11524,8 | 119,0 | 105,9 | 114,7 |
| 123,2 | 18440,1 | 130,4 | 108,6 | 111,2 |
| 117,4 | 9251,6 | 120,7 | 105,7 | 114,2 |
| 109,8 | 27307,5 | 109,9 | 103,2 | 114,6 |
| 129,1 | 20401,3 | 114,2 | 105,7 | 121,7 |
| 121,3 | 12342,5 | 123,2 | 102,9 | 115,9 |
| 109,9 | 14924,9 | 109,8 | 99,4 | 116,7 |
| 125,9 | 18095,2 | 132,5 | 111,2 | 105,9 |
| 116,0 | 13542,5 | 120,4 | 106,3 | 123,8 |
| 115,2 | 15007,6 | 126,5 | 102,1 | 115,9 |
| 104,9 | 21797,8 | 113,8 | 102,1 | 109,5 |
| 119,5 | 21018,2 | 112,1 | 106,9 | 112,6 |
| 110,0 | 16829,1 | 115,2 | 100,8 | 108,1 |
| 124,0 | 15047,2 | 127,3 | 106,4 | 104,9 |

Вариант 9

| $y$ | $x_1$ | $x_2$ | $x_3$ | $x_4$ |
|---|---|---|---|---|
| 105,3 | 17361,4 | 104,4 | 107,5 | 108,3 |
| 100,9 | 19261,8 | 104,1 | 108,4 | 106,8 |
| 102,8 | 16779,0 | 103,0 | 109,7 | 111,0 |
| 113,1 | 20812,9 | 108,5 | 111,6 | 118,7 |
| $y$ | $x_1$ | $x_2$ | $x_3$ | $x_4$ |

| | | | | |
|---|---|---|---|---|
| 99,3 | 17640,4 | 103,3 | 101,3 | 110,0 |
| 97,2 | 20607,1 | 102,1 | 101,6 | 104,3 |
| 105,2 | 10769,4 | 106,3 | 106,9 | 114,6 |
| 101,4 | 20662,5 | 105,4 | 105,5 | 104,5 |
| 101,2 | 18746,5 | 105,0 | 105,3 | 110,8 |
| 107,4 | 22141,8 | 106,5 | 107,0 | 117,2 |
| 97,5 | 16443,5 | 103,1 | 99,6 | 102,2 |
| 102,7 | 16948,8 | 107,3 | 107,8 | 110,4 |
| 108,9 | 13949,5 | 109,6 | 111,1 | 119,0 |
| 99,1 | 16359,4 | 104,1 | 104,4 | 111,8 |
| 102,4 | 20914,3 | 101,1 | 105,6 | 108,8 |
| 96,8 | 17175,2 | 103,9 | 104,8 | 100,9 |
| 113,8 | 19118,6 | 107,0 | 106,4 | 122,0 |
| 102,6 | 14026,2 | 104,0 | 107,8 | 115,7 |
| 108,5 | 17045,6 | 109,4 | 107,1 | 119,2 |
| 113,8 | 19118,6 | 107,0 | 106,4 | 122,0 |
| 102,6 | 14026,2 | 104,0 | 107,8 | 115,7 |
| 108,5 | 17045,6 | 109,4 | 107,1 | 119,2 |
| 108,0 | 18236,8 | 107,3 | 108,9 | 119,4 |

Вариант 10

| $y$ | $x_1$ | $x_2$ | $x_3$ | $x_4$ |
|---|---|---|---|---|
| 113,7 | 20666,8 | 104,9 | 107,2 | 102,2 |
| 110,5 | 23362,2 | 103,3 | 106,9 | 102,8 |
| 113,0 | 18346,0 | 105,0 | 106,4 | 104,4 |
| 104,7 | 18450,8 | 103,0 | 104,5 | 101,9 |
| 104,7 | 18942,3 | 102,1 | 107,1 | 100,5 |
| 112,6 | 17021,6 | 103,0 | 109,0 | 102,9 |
| 110,1 | 18301,7 | 104,0 | 103,3 | 104,0 |
| 101,7 | 21092,0 | 102,3 | 104,4 | 102,7 |
| 113,6 | 20504,0 | 104,2 | 104,6 | 104,2 |
| 97,4 | 16378,1 | 102,7 | 104,3 | 101,4 |
| 117,3 | 22126,6 | 105,7 | 105,5 | 103,8 |
| 110,9 | 18206,4 | 103,1 | 106,3 | 104,1 |
| 114,4 | 20063,5 | 104,3 | 102,5 | 105,1 |
| 100,2 | 24192,3 | 100,3 | 103,2 | 101,2 |
| 111,4 | 18782,1 | 103,8 | 106,3 | 102,0 |
| 110,1 | 18301,7 | 104,0 | 103,3 | 104,0 |
| 101,7 | 21092,0 | 102,3 | 104,4 | 102,7 |
| 113,6 | 20504,0 | 104,2 | 104,6 | 104,2 |

| $y$ | $x_1$ | $x_2$ | $x_3$ | $x_4$ |
|---|---|---|---|---|
| 97,4 | 16378,1 | 102,7 | 104,3 | 101,4 |
| 117,3 | 22126,6 | 105,7 | 105,5 | 103,8 |
| 110,9 | 18206,4 | 103,1 | 106,3 | 104,1 |
| 114,4 | 20063,5 | 104,3 | 102,5 | 105,1 |
| 100,2 | 24192,3 | 100,3 | 103,2 | 101,2 |
| 111,4 | 18782,1 | 103,8 | 106,3 | 102,0 |
| 106,2 | 20272,6 | 103,6 | 102,2 | 102,8 |

## 17. 4. Методика выполнения заданий

Исследуется взаимосвязь показателей качества жизни населения по выборке для 25 регионов (рис. 4.2).

$y$ – средняя ожидаемая продолжительность жизни при рождении, лет;

$x_1$ – уровень рождаемости, чел. на 1000 чел. населения;

$x_2$ – доля населения с денежными доходами ниже величины прожиточного минимума, % от всего населения;

$x_3$ – среднедушевые доходы населения, у.е.;

$x_4$ – объем социальных выплат, млрд. у.е.

| | A | B | C | D | E | F |
|---|---|---|---|---|---|---|
| 1 | i | Y | X1 | X2 | X3 | X4 |
| 2 | 1 | 68,1 | 10,2 | 11,2 | 14,04 | 6,09 |
| 3 | 2 | 68,2 | 10,5 | 14,0 | 16,27 | 6,79 |
| 4 | 3 | 69,0 | 11,7 | 11,9 | 23,41 | 4,50 |
| 5 | 4 | 68,2 | 11,3 | 12,0 | 16,41 | 4,71 |
| 6 | 5 | 66,6 | 8,8 | 14,3 | 11,25 | 5,72 |
| 7 | 6 | 68,6 | 11,9 | 11,0 | 21,22 | 4,69 |
| 8 | 7 | 68,3 | 11,4 | 11,3 | 14,72 | 6,11 |
| 9 | 8 | 67,3 | 9,0 | 14,3 | 11,31 | 6,65 |
| 10 | 9 | 68,6 | 11,4 | 12,6 | 23,04 | 5,18 |
| 11 | 10 | 68,4 | 12,0 | 12,5 | 21,67 | 5,41 |
| 12 | 11 | 69,1 | 11,1 | 10,5 | 20,80 | 5,83 |
| 13 | 12 | 69,1 | 12,3 | 11,2 | 21,55 | 4,85 |
| 14 | 13 | 68,8 | 12,0 | 12,5 | 18,08 | 5,57 |
| 15 | 14 | 68,7 | 12,5 | 13,0 | 19,81 | 5,58 |
| 16 | 15 | 68,6 | 11,2 | 15,1 | 16,16 | 6,52 |
| 17 | 16 | 68,6 | 12,5 | 12,8 | 18,87 | 5,70 |
| 18 | 17 | 69,0 | 12,2 | 12,2 | 22,43 | 5,72 |
| 19 | 18 | 68,5 | 10,5 | 13,9 | 17,06 | 6,84 |
| 20 | 19 | 67,9 | 10,9 | 12,9 | 20,53 | 5,43 |
| 21 | 20 | 69,7 | 13,1 | 11,8 | 23,49 | 6,02 |
| 22 | 21 | 68,5 | 10,4 | 11,6 | 21,98 | 5,11 |
| 23 | 22 | 68,6 | 11,9 | 13,1 | 19,48 | 5,34 |
| 24 | 23 | 68,3 | 12,5 | 12,1 | 21,30 | 4,95 |
| 25 | 24 | 67,0 | 8,1 | 15,2 | 11,22 | 7,43 |
| 26 | 25 | 68,0 | 10,1 | 12,3 | 20,33 | 6,06 |

Рис. 4.2. Исходные данные задачи

1. Для исследования целесообразности включения факторов в модель построим корреляционную матрицу. Воспользуемся инструментом Корреляция пакета Анализ данных (закладка Данные). В окне настройки параметров (рис. 4.3) в качестве входного интервала укажем столбцы значений $x$ и $y$ с заголовками, в качестве выходного интервала – область ячеек 6×6. Нажав OK, получим корреляционную матрицу (рис. 4.4).

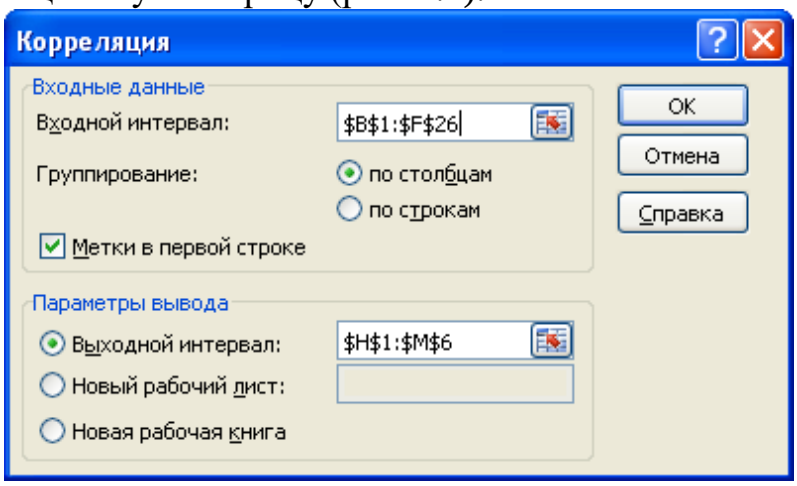



Рис. 4.3. Настройка параметров инструмента Корреляция

| | H | I | J | K | L | M |
|---|---|---|---|---|---|---|
| 1 | | *Y* | *X1* | *X2* | *X3* | *X4* |
| 2 | Y | 1 | | | | |
| 3 | X1 | 0,84358 | 1 | | | |
| 4 | X2 | -0,56285 | -0,52249 | 1 | | |
| 5 | X3 | 0,78743 | 0,75317 | -0,57922 | 1 | |
| 6 | X4 | -0,38186 | -0,56991 | 0,63499 | -0,63884 | 1 |

Рис. 4.4. Корреляционная матрица

Проанализируем полученную корреляционную матрицу:

1) Между зависимой переменной $y$ и независимыми переменными $x_1$, $x_2$, $x_3$наблюдается корреляционная заметная связь, в то время как между $y$ и $x_4$ связь не достаточно сильна ($|r_{yx}| < 0{,}4$). Следовательно, в модель не стоит включать фактор $x_4$.

2) Между независимыми переменными $x_1$и$x_3$наблюдается высокая нежелательная корреляционная связь ($|r_{x_1x_3}| > 0{,}7$ ). Для исключения мультиколлинеарности один из этих факторов нужно убрать из модели. По сравнению с $x_1$ фактор $x_3$ сильнее связан с $x_2$ и слабее с $y$, поэтому в модели следует оставить $x_1$.

Таким образом, включаем в модель факторы $x_1$ и $x_2$.

С помощью инструмента Регрессия построим и оценим уравнение линейной множественной регрессии. Для этого в окне параметров инструмента (рис. 4.5) зададим в качестве входного интервала – столбец со значениями *y*, в качестве выходного интервала – столбцы со значениями факторов *x*, которые мы ранее выбрали для включения в модель – $x_1$ и $x_2$. Для последующего анализа выполнения предпосылок нам также пригодится информация об остатках и график нормальной вероятности. Часть результатов Регрессии, необходимая для построения и оценки качества уравнения, представлена на рис. 4.6.

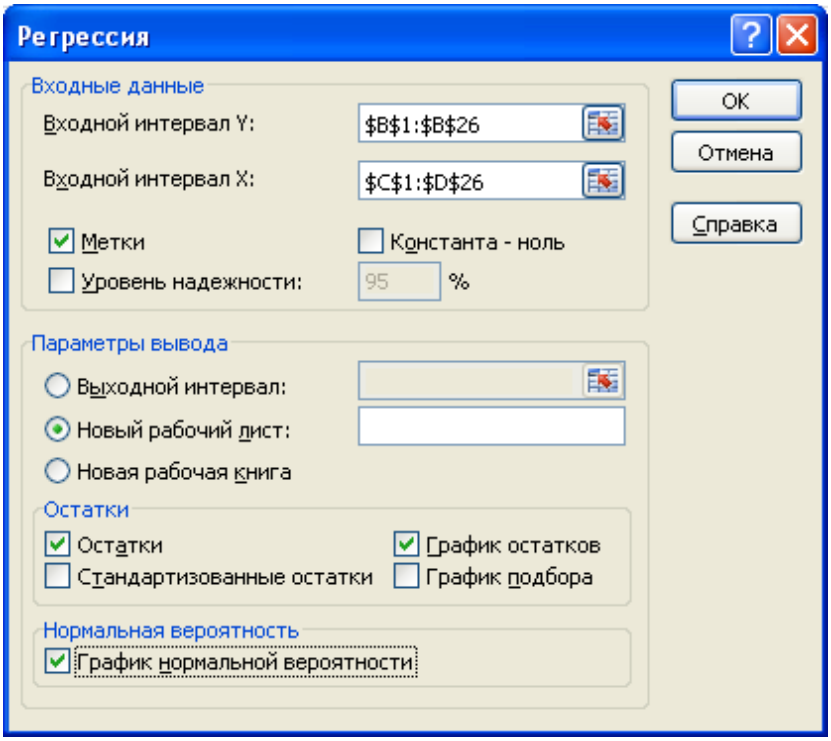



Рис. 4.5. Настройка параметров инструмента Регрессия

| | A | B | C | D | E | F | G | H | I |
|---|---|---|---|---|---|---|---|---|---|
| 1 | ВЫВОД ИТОГОВ | | | | | | | | |
| 2 | | | | | | | | | |
| 3 | *Регрессионная статистика* | | | | | | | | |
| 4 | Множественный R | 0,855645232 | | | | | | | |
| 5 | R-квадрат | 0,732128764 | | | | | | | |
| 6 | Нормированный R-квадрат | 0,707776833 | | | | | | | |
| 7 | Стандартная ошибка | 0,359476251 | | | | | | | |
| 8 | Наблюдения | 25 | | | | | | | |
| 9 | | | | | | | | | |
| 10 | Дисперсионный анализ | | | | | | | | |
| 11 | | *df* | *SS* | *MS* | *F* | *Значимость F* | | | |
| 12 | Регрессия | 2 | 7,770061865 | 3,885030932 | 30,06450605 | 0,000000510 | | | |
| 13 | Остаток | 22 | 2,842909854 | 0,129223175 | | | | | |
| 14 | Итого | 24 | 10,61297172 | | | | | | |
| 15 | | | | | | | | | |
| 16 | | *Коэффициенты* | *Стандартная ошибка* | *t-статистика* | *P-Значение* | *Нижние 95%* | *Верхние 95%* | *Нижние 95,0%* | *Верхние 95,0%* |
| 17 | Y-пересечение | 65,02301895 | 1,418055429 | 45,85365114 | 0,00000000 | 62,08215199 | 67,9639 | 62,0822 | 67,9639 |
| 18 | X1 | 0,400802485 | 0,068625655 | 5,840417654 | 0,00000708 | 0,258481588 | 0,54312 | 0,25848 | 0,54312 |
| 19 | X2 | -0,087990916 | 0,067805884 | -1,297688494 | 0,20783669 | -0,228611712 | 0,05263 | -0,22861 | 0,05263 |

Рис. 4.6. Итоги регрессии (регрессионная статистика и дисперсионный анализ)

2. Построим линейное уравнение множественной регрессии:

1) В ячейках B17:B19 (рис. 4.6) выведены параметры уравнения в естественной форме. Следовательно, наше уравнение имеет вид:

$$\hat{y}_x = 65{,}02 + 0{,}401 \cdot x_1 - 0{,}088 \cdot x_2$$

Параметры найденного уравнения показывают, что:

- при увеличении уровня рождаемости на 1 чел./1000 чел. средняя ожидаемая продолжительность жизни увеличится в среднем на 0,401 лет при неизменном значении других факторов;
- при увеличении доли населения с денежными доходами ниже величины прожиточного минимума на 1% средняя ожидаемая продолжительность жизни уменьшится в среднем на 0,088 лет;
- при нулевом уровне рождаемости и отсутствии населения с денежными доходами ниже величины прожиточного минимума средняя ожидаемая продолжительность жизни составит 65,02 года.

2) Определим параметры стандартизованного уравнения по формулам перехода $\beta_i = b_i \frac{\sigma_{x_i}}{\sigma_y}$. Стандартные отклонения рассчитаем с помощью функции Excel СТАНДОТКЛОН.Г. Результаты вычислений параметров уравнения в стандартизованной форме показаны на рис. 4.7. Уравнение в стандартизованной форме имеет вид:

$$t_y = 0{,}756 \cdot t_{x_1} - 0{,}168 \cdot t_{x_2}.$$

Исходя из коэффициентов уравнения регрессии в стандартизованном виде, $x_1$ оказывает б*о*льшее влияние на $y$, чем $x_2$.

3) Определим коэффициенты эластичности по формуле $\overline{Э}_{yx_i} = b_J \cdot \frac{\overline{x_J}}{\overline{y}}$.

$\overline{Э}_{yx_1} = 0{,}065$, $\overline{Э}_{yx_2} = -0{,}016$. Исходя из средних коэффициентов эластичности, $x_1$ оказывает б*о*льшее влияние на $y$, чем $x_2$. Это полностью подтверждает выводы, сделанные по стандартизованному уравнению.

B32 =СТАНДОТКЛОН.Г(B2:B26)

| | A | B | C | D | E | F |
|---|---|---|---|---|---|---|
| 28 | Средние | Y | X1 | X2 | | |
| 29 | значения | 68,4 | 11,2 | 12,6 | | |
| 30 | | | | | | |
| 31 | Стандартное | Y | X1 | X2 | | |
| 32 | отклонение | 0,6516 | 1,2287 | 1,2435 | | |
| 33 | | | | | | |
| 34 | Параметры | | $\beta_1$ | $\beta_2$ | | |
| 35 | стандартизованного уравнения | | 0,75583 | -0,1679 | | |
| 36 | | | | | | |
| 37 | Коэффициенты | | $Э_{yx1}$ | $Э_{yx2}$ | | |
| 38 | эластичности | | 0,06549 | -0,0162 | | |

Рис. 4.7. Сравнение степени влияния факторов

3. Проверим качество построенного уравнения. Для этого воспользуемся результатами Регрессии (рис. 4.6).

1) Коэффициент множественной корреляции R = 0,856 говорит о высокой степени тесноты линейной связи средней ожидаемой продолжительности жизни с совокупностью факторов $x_1$ и $x_2$. Коэффициент детерминации $R^2$ = 0,732 и скорректированный коэффициент детерминации $\bar{R}^2$ = 0,708 свидетельствуют о высоком качестве регрессионной модели.

2) проверим значимость уравнения с помощью $F$-критерия Фишера:на рис 4.6 расчетное значение $F$ = 30,06, при этом Значимость $F$ = 0,00000051, что существенно меньше заданного уровня значимости $\alpha$ = 0,05. Таким образом, уравнениестатистически значимо.

3) проверим значимость параметров уравнения с помощью $t$-критерия Стьюдента: на рис 4.6 расчетные значения $t$-статистики для свободного члена и параметров $b_1$, $b_2$ составляют соответственно 45,854, 5,840 и -1,298. При этом $p$-значения для свободного члена и $b_1$существенно меньше 0,05, следовательно, гипотезу о незначимости этих параметров следует отвергнуть. $p$-значение для $b_2$= 0,2 > 0,05, т.е. у нас нет оснований отвергать гипотезу о незначимости этого параметра.

4. Проведем анализ остатков на выполнение пяти предпосылок МНК:

1) Проверим выполнение предпосылки о случайном характере остатков: инструмент Регрессия имеет опцию вывода остатков, результат представлен на рис. 4.8. Построим на основе этих данных точечную диаграмму: по оси абсцисс отложим теоретические значения $\hat{y}_x$, по оси ординат – остатки. Из графика остатков на рис. 4.8 видно, что точки

расположены случайным
образом, то есть первая предпосылка выполняется.

2) Проверим выполнение предпосылки о нулевой средней величине остатков: инструмент Регрессия имеет опцию вывода Графиковостатков, результат представлен на рис. 4.9. На графиках видно, что остатки расположены случайным образом, то есть вторая предпосылка выполняется.

3) Проверим выполнение предпосылки о гомоскедастичности. Наличие гомоскедастичности или гетероскедастичности можно видеть по графику зависимости остатков от теоретического значения результативного признака $\hat{y}_x$ (рис. 4.8). На графике видно, что дисперсия остатков в целом не меняется при переходе от одного значения $\hat{y}_x$ к другому. Следовательно, третья предпосылка выполняется.

| | A | B | C | D | E | F |
|---|---|---|---|---|---|---|
| 23 | ВЫВОД ОСТАТКА | | | | ВЫВОД ВЕРОЯТНОСТИ | |
| 24 | | | | | | |
| 25 | *Наблюдение* | *Предсказанное Y* | *Остатки* | | *Персентиль* | *Y* |
| 26 | 1 | 68,1261753 | -0,050215406 | | 2 | 66,63889118 |
| 27 | 2 | 67,99777676 | 0,225167058 | | 6 | 67,02356858 |
| 28 | 3 | 68,65731585 | 0,38304324 | | 10 | 67,33789524 |
| 29 | 4 | 68,49660015 | -0,325331342 | | 14 | 67,88444355 |
| 30 | 5 | 67,28205069 | -0,643159505 | | 18 | 67,99174002 |
| 31 | 6 | 68,81157292 | -0,258767194 | | 22 | 68,0759599 |
| 32 | 7 | 68,5963647 | -0,290084532 | | 26 | 68,17126881 |
| 33 | 8 | 67,38268449 | -0,044789245 | | 30 | 68,22294382 |
| 34 | 9 | 68,46668241 | 0,123323421 | | 34 | 68,30628017 |
| 35 | 10 | 68,73400046 | -0,328925594 | | 38 | 68,32765768 |
| 36 | 11 | 68,5670487 | 0,542206845 | | 42 | 68,40507486 |
| 37 | 12 | 68,96921647 | 0,105573248 | | 46 | 68,541426 |
| 38 | 13 | 68,74142385 | 0,05938629 | | 50 | 68,54190711 |
| 39 | 14 | 68,87382312 | -0,194610569 | | 54 | 68,55280572 |
| 40 | 15 | 68,18341302 | 0,408605432 | | 58 | 68,5773495 |
| 41 | 16 | 68,90071264 | -0,323363141 | | 62 | 68,59000583 |
| 42 | 17 | 68,85620214 | 0,113680352 | | 66 | 68,59201845 |
| 43 | 18 | 67,99418822 | 0,547718884 | | 70 | 68,64390963 |
| 44 | 19 | 68,27297193 | -0,388528375 | | 74 | 68,67921255 |
| 45 | 20 | 69,23089102 | 0,471413842 | | 78 | 68,80081014 |
| 46 | 21 | 68,17785563 | 0,363570365 | | 82 | 68,96988249 |
| 47 | 22 | 68,62726096 | 0,016648677 | | 86 | 69,04035909 |
| 48 | 23 | 68,95986873 | -0,632211056 | | 90 | 69,07478972 |
| 49 | 24 | 66,92200027 | 0,101568308 | | 94 | 69,10925555 |
| 50 | 25 | 67,97366002 | 0,018079997 | | 98 | 69,70230486 |

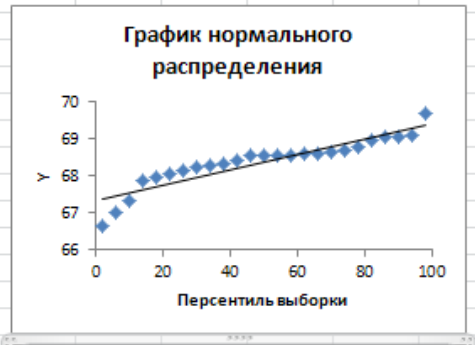
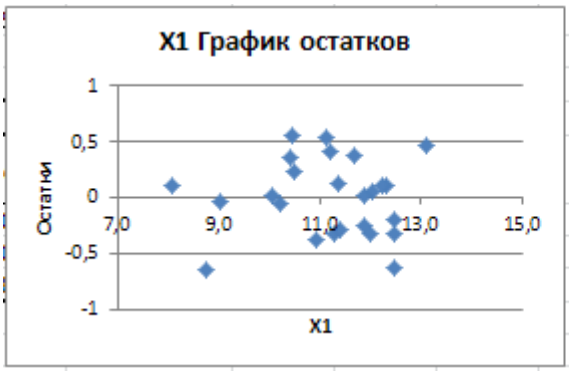




Рис. 4.8. Итоги регрессии (вывод остатка и вывод вероятности)



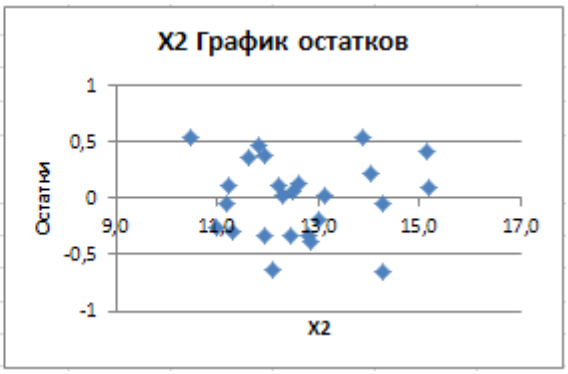



Рис. 4.9. Итоги регрессии (графики остатков)

4) Проверим выполнение предпосылки об отсутствии автокорреляции остатков с помощью теста Дарбина-Уотсона:

- Выдвигаем гипотезу $H_0$ об отсутствии автокорреляции;
- Для расчетов используем значения из столбца «Остатки» результатов Регрессии (рис. 4.8). Занесем их в столбец B нового листа (рис. 4.10);

B2 =Регрессия!C26

| | A | B | C | D | E | F | G |
|---|---|---|---|---|---|---|---|
| 1 | $i$ | $u_i$ | $u^2_i$ | $(u_i-u_{i-1})^2$ | | Проверка наличия автокорреляции в остатках | |
| 2 | 1 | -0,050 | 0,003 | --- | | Статистика Дарбина-Уотсона DW | 2,165 |
| 3 | 2 | 0,225 | 0,051 | 0,076 | | Число наблюдений $n$ | 25 |
| 4 | 3 | 0,383 | 0,147 | 0,025 | | Число независимых переменных $m$ | 2 |
| 5 | 4 | -0,325 | 0,106 | 0,502 | | Критическое значение $d_L$ | 1,21 |
| 6 | 5 | -0,643 | 0,414 | 0,101 | | Критическое значение $d_U$ | 1,55 |
| 7 | 6 | -0,259 | 0,067 | 0,148 | | Вывод | автокорреляция отсутствует |
| 8 | 7 | -0,290 | 0,084 | 0,001 | | | |
| 9 | 8 | -0,045 | 0,002 | 0,060 | | | |
| 10 | 9 | 0,123 | 0,015 | 0,028 | | | |
| 11 | 10 | -0,329 | 0,108 | 0,205 | | | |
| 12 | 11 | 0,542 | 0,294 | 0,759 | | | |
| 13 | 12 | 0,106 | 0,011 | 0,191 | | | |
| 14 | 13 | 0,059 | 0,004 | 0,002 | | | |
| 15 | 14 | -0,195 | 0,038 | 0,065 | | | |
| 16 | 15 | 0,409 | 0,167 | 0,364 | | | |
| 17 | 16 | -0,323 | 0,105 | 0,536 | | | |
| 18 | 17 | 0,114 | 0,013 | 0,191 | | | |
| 19 | 18 | 0,548 | 0,300 | 0,188 | | | |
| 20 | 19 | -0,389 | 0,151 | 0,877 | | | |
| 21 | 20 | 0,471 | 0,222 | 0,740 | | | |
| 22 | 21 | 0,364 | 0,132 | 0,012 | | | |
| 23 | 22 | 0,017 | 0,000 | 0,120 | | | |
| 24 | 23 | -0,632 | 0,400 | 0,421 | | | |
| 25 | 24 | 0,102 | 0,010 | 0,538 | | | |
| 26 | 25 | 0,018 | 0,000 | 0,007 | | | |
| 27 | сумма | | 2,843 | 6,155 | | | |

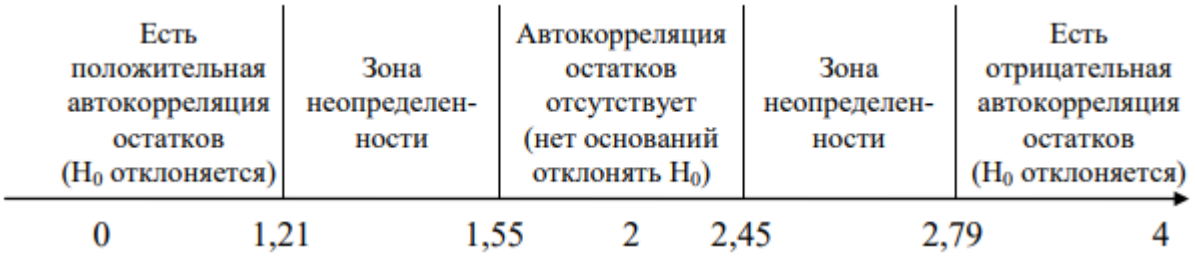
Рис. 4.10. Проверка гипотезы об отсутствии автокорреляции

- В столбце C возведем остатки в квадрат, в столбце D рассчитаем квадраты отклонений текущего значения остатка от предыдущего и просуммируем значения (рис. 4.10);
- Рассчитаем статистику Дарбина-Уотсона по формуле:

$$DW = \frac{\sum_{i=2}^{n}(u_i - u_{i-1})^2}{\sum_{1=1}^{n} u_i^2};$$

• Для числа наблюдений $n$=25 и числа независимых переменных $m$=2 и уровня значимости $\alpha$=0,05 по специальной таблице (см. приложение 1) найдем критические значения $d_L$ =1,21 и $d_U$ =1,55;

• Построим шкалу Дарбина-Уотсона и зададим автоматический вывод результата с помощью функции Excel ЕСЛИ. Используя условное форматирование, выделим ячейку красным цветом в случае наличия автокорреляции, зеленым –в случае отсутствия, желтым – при попадании значения в зонунеопределенности.




5) Проверим выполнение предпосылки о нормальном распределении остатков с помощью Графика нормальной вероятности инструмента Регрессия (рис. 4.8): поскольку все точки расположены близко к прямой, то можно считать, что распределение остатков близко к нормальному и пятая предпосылка выполняется.

Таким образом, все предпосылки МНК выполняются.

## 18. 5. Контрольные вопросы

1. В каких случаях строится уравнение множественной регрессии? Какова основная цель множественной регрессии?
2. Какой метод применяется для оценки параметров линейного уравнения множественной регрессии? В каких функциях MS Excel он реализован?
3. Каким требованиям должны отвечать факторы, включаемые во множественную регрессию?
4. Что показывает корреляционная матрица? Как построить корреляционную матрицу с помощью MS Excel?
5. Каким образом строится и для чего используется уравнение множественной регрессии в стандартизованном масштабе?
6. Что показывает линейный коэффициент множественной корреляции?
7. В чем особенности скорректированного коэффициента детерминации?
8. Как с помощью инструмента Регрессия можно оценить значимость уравнения регрессии и его параметров?
9. Назовите пять предпосылок МНК. Каким образом можно проверить

выполнение предпосылок средствами MS Excel?

10. Для чего используется тест Дарбина-Уотсона? Опишите его алгоритм.

Приложение 1

Значения статистик Дарбина-Уотсона при 5%-ном уровне значимости

| n | $m$=1 | | $m$=2 | | $m$=3 | | $m$=4 | | $m$=5 | |
|---|---|---|---|---|---|---|---|---|---|---|
| | $d_L$ | $d_U$ | $d_L$ | $d_U$ | $d_L$ | $d_L$ | $d_U$ | $d_L$ | $d_U$ | $d_L$ |
| 6 | 0,610 | 1,400 | - | - | - | - | - | - | - | - |
| 7 | 0,700 | 1,356 | 0,467 | 1,896 | - | - | - | - | - | - |
| 8 | 0,763 | 1,332 | 0,559 | 1,777 | 0,368 | 2,287 | - | - | - | - |
| 9 | 0,824 | 1,320 | 0,629 | 1,699 | 0,455 | 2,128 | - | - | - | - |
| 10 | 0,879 | 1,320 | 0,697 | 1,641 | 0,525 | 2,016 | - | - | - | - |
| 11 | 0,927 | 1,324 | 0,658 | 1,604 | 0,595 | 1,928 | - | - | - | - |
| 12 | 0,971 | 1,331 | 0,812 | 1,579 | 0,658 | 1,864 | - | - | - | - |
| 13 | 1,010 | 1,340 | 0,861 | 1,562 | 0,715 | 1,816 | - | - | - | - |
| 14 | 1,045 | 1,350 | 0,905 | 1,551 | 0,767 | 1,779 | - | - | - | - |
| 15 | 1,077 | 1,361 | 0,946 | 1,543 | 0,814 | 1,750 | 0,685 | 1,977 | 0,562 | 2,220 |
| 16 | 1,106 | 1,371 | 0,982 | 1,539 | 0,857 | 1,728 | 0,734 | 1,935 | 0,615 | 2,157 |
| 17 | 1,133 | 1,381 | 1,015 | 1,536 | 0,897 | 1,710 | 0,779 | 1,900 | 0,664 | 2,104 |
| 18 | 1,158 | 1,391 | 1,046 | 1,535 | 0,933 | 1,696 | 0,820 | 1,872 | 0,710 | 2,060 |
| 19 | 1,180 | 1,401 | 1,074 | 1,536 | 0,967 | 1,685 | 0,859 | 1,849 | 0,752 | 2,023 |
| 20 | 1,201 | 1,411 | 1,100 | 1,537 | 0,998 | 1,676 | 0,984 | 1,828 | 0,792 | 1,991 |
| 21 | 1,222 | 1,420 | 1,125 | 1,538 | 1,026 | 1,669 | 0,927 | 1,812 | 0,829 | 1,964 |
| 22 | 1,239 | 1,429 | 1,147 | 1,541 | 1,053 | 1,664 | 0,958 | 1,797 | 0,863 | 1,940 |
| 23 | 1,257 | 1,437 | 1,168 | 1,543 | 1,078 | 1,660 | 0,986 | 1,785 | 0,895 | 1,920 |
| 24 | 1,273 | 1,446 | 1,188 | 1,546 | 1,101 | 1,656 | 1,013 | 1,775 | 0,925 | 1,902 |
| 25 | 1,288 | 1,454 | 1,206 | 1,550 | 1,123 | 1,654 | 1,038 | 1,767 | 0,953 | 1,886 |
| 26 | 1,302 | 1,461 | 1,224 | 1,553 | 1,143 | 1,652 | 1,062 | 1,759 | 0,979 | 1,873 |
| 27 | 1,316 | 1,469 | 1,240 | 1,556 | 1,162 | 1,651 | 1,084 | 1,753 | 1,004 | 1,861 |
| 28 | 1,328 | 1,476 | 1,255 | 1,560 | 1,181 | 1,650 | 1,104 | 1,747 | 1,028 | 1,850 |
| 29 | 1,341 | 1,483 | 1,270 | 1,563 | 1,198 | 1,650 | 1,124 | 1,743 | 1,050 | 1,841 |
| 30 | 1,352 | 1,489 | 1,284 | 1,567 | 1,214 | 1,650 | 1,143 | 1,739 | 1,071 | 1,833 |
| 31 | 1,363 | 1,496 | 1,297 | 1,570 | 1,229 | 1,650 | 1,160 | 1,735 | 1,090 | 1,825 |
| 32 | 1,373 | 1,502 | 1,309 | 1,574 | 1,244 | 1,650 | 1,177 | 1,732 | 1,109 | 1,819 |
| 33 | 1,383 | 1,508 | 1,321 | 1,577 | 1,258 | 1,651 | 1,193 | 1,730 | 1,127 | 1,813 |
| 34 | 1,393 | 1,514 | 1,333 | 1,580 | 1,271 | 1,652 | 1,028 | 1,728 | 1,144 | 1,808 |
| 35 | 1,402 | 1,519 | 1,343 | 1,584 | 1,283 | 1,653 | 1,222 | 1,726 | 1,160 | 1,803 |
| 36 | 1,411 | 1,525 | 1,354 | 1,587 | 1,295 | 1,654 | 1,236 | 1,724 | 1,175 | 1,799 |
| 37 | 1,419 | 1,530 | 1,364 | 1,590 | 1,307 | 1,655 | 1,249 | 1,723 | 1,190 | 1,795 |
| 38 | 1,427 | 1,535 | 1,373 | 1,594 | 1,318 | 1,656 | 1,261 | 1,722 | 1,204 | 1,792 |
| 39 | 1,435 | 1,540 | 1,382 | 1,597 | 1,328 | 1,658 | 1,273 | 1,722 | 1,218 | 1,789 |

19. Приложение 1 (продолжение)

| n | *m*=1 | | *m*=2 | | *m*=3 | | *m*=4 | | *m*=5 | |
|---|---|---|---|---|---|---|---|---|---|---|
| | $d_L$ | $d_U$ | $d_L$ | $d_U$ | $d_L$ | $d_L$ | $d_U$ | $d_L$ | $d_U$ | $d_L$ |
| 40 | 1,442 | 1,544 | 1,391 | 1,600 | 1,338 | 1,659 | 1,285 | 1,721 | 1,230 | 1,786 |
| 45 | 1,475 | 1,566 | 1,430 | 1,615 | 1,383 | 1,666 | 1,336 | 1,720 | 1,287 | 1,776 |
| 50 | 1,503 | 1,585 | 1,462 | 1,628 | 1,421 | 1,674 | 1,378 | 1,721 | 1,335 | 1,771 |
| 55 | 1,528 | 1,601 | 1,490 | 1,641 | 1,452 | 1,681 | 1,414 | 1,724 | 1,374 | 1,768 |
| 60 | 1,549 | 1,616 | 1,514 | 1,652 | 1,480 | 1,689 | 1,444 | 1,727 | 1,408 | 1,767 |
| 65 | 1,567 | 1,629 | 1,536 | 1,662 | 1,503 | 1,696 | 1,471 | 1,731 | 1,438 | 1,767 |
| 70 | 1,583 | 1,641 | 1,554 | 1,672 | 1,525 | 1,703 | 1,494 | 1,735 | 1,464 | 1,768 |
| 75 | 1,598 | 1,652 | 1,571 | 1,680 | 1,543 | 1,709 | 1,515 | 1,739 | 1,487 | 1,770 |
| 80 | 1,611 | 1,662 | 1,586 | 1,688 | 1,560 | 1,715 | 1,534 | 1,743 | 1,507 | 1,772 |
| 85 | 1,624 | 1,671 | 1,600 | 1,696 | 1,575 | 1,721 | 1,550 | 1,747 | 1,525 | 1,774 |
| 90 | 1,635 | 1,679 | 1,612 | 1,703 | 1,589 | 1,726 | 1,566 | 1,751 | 1,542 | 1,776 |
| 95 | 1,645 | 1,687 | 1,623 | 1,709 | 1,602 | 1,732 | 1,579 | 1,755 | 1,557 | 1,778 |
| 100 | 1,654 | 1,694 | 1,634 | 1,715 | 1,613 | 1,736 | 1,592 | 1,758 | 1,571 | 1,780 |
| 150 | 1,720 | 1,746 | 1,706 | 1,760 | 1,693 | 1,774 | 1,679 | 1,788 | 1,665 | 1,802 |
| 200 | 1,758 | 1,778 | 1,748 | 1,789 | 1,738 | 1,799 | 1,728 | 1,810 | 1,718 | 1,820 |