

**Л.Ф. Филатова**

# **МНОЖЕСТВЕННАЯ РЕГРЕССИЯ**

**Учебное пособие**



МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ  
РОССИЙСКОЙ ФЕДЕРАЦИИ  
Федеральное государственное автономное образовательное учреждение  
высшего профессионального образования  
**«Национальный исследовательский ядерный университет «МИФИ»**  
**Северский технологический институт – филиал НИЯУ МИФИ**  
**(СТИ НИЯУ МИФИ)**

Л.Ф. Филатова

**МНОЖЕСТВЕННАЯ РЕГРЕССИЯ**

Учебное пособие

Северск 2013

УДК 330  
ББК У.в 611 я 73.  
Ф 517  
**Филатова Л.Ф.**

Ф 517 Множественная регрессия: учебное пособие/ Л.Ф. Филатова. –  
Северск: Изд. СТИ НИЯУ МИФИ, 2013. – 50 с.

В данном учебном пособии изложены основные теоретические вопросы и этапы выполнения лабораторных работ по множественной регрессии. Для проверки полученных теоретических знаний составлен список контрольных вопросов, приведенный в конце первого раздела. Лабораторные работы оснащены заданиями для самостоятельной работы.

В приложениях приведены статистические таблицы критериев Фишера, Стьюдента и Пирсона, которые используются для проведения необходимых оценок.

Учебное пособие подготовлено на кафедре «ВМиИТ» СТИ НИЯУ МИФИ и предназначено для студентов экономических специальностей очной, очно-заочной и заочной форм обучения.

Пособие одобрено на заседании кафедры высшей математики и информационных технологий (протокол № 3 от " 13 " марта 2013 г.).

Печатается в соответствии с планом выпуска учебно – методической литературы на 2013 г., утвержденным Советом СТИ НИЯУ МИФИ.

Рег.№ 5/13 от « 8 » апреля 2013г.

Рецензенты

Канд. физ.-мат. наук, зав.кафедрой ИОИД ТГУ  
С.Л. Миньков

Канд. техн. наук, доцент ТПУ  
Е.И. Сметанина

Редактор  
Р.В. Фирсова

ISBN

© Северский технологический институт НИЯУ МИФИ, 2013

## Содержание

|                                                                                                                         |    |
|-------------------------------------------------------------------------------------------------------------------------|----|
| Введение.....                                                                                                           | 4  |
| 1 Теоретическая часть.....                                                                                              | 5  |
| 1.1 Построение уравнения множественной регрессии.....                                                                   | 5  |
| 1.1.1 Отбор факторов при построении множественной регрессии.....                                                        | 5  |
| 1.1.2 Выбор формы уравнения множественной регрессии.....                                                                | 10 |
| 1.1.3 Расчет параметров уравнений множественной регрессии.....                                                          | 12 |
| 1.2 Оценка множественной регрессии.....                                                                                 | 15 |
| 1.2.1 Общие оценки уравнений множественной регрессии.....                                                               | 16 |
| 1.2.2 Частные оценки уравнений множественной регрессии.....                                                             | 18 |
| 1.2.3 Оценка значимости коэффициентов чистой регрессии по критерию Стьюдента.....                                       | 21 |
| 1.3 Исследование остаточных величин регрессии (предпосылки МНК).....                                                    | 22 |
| 1.3.1 Свойства случайных остатков.....                                                                                  | 22 |
| 1.3.2 Первая предпосылка - случайный характер остатков.....                                                             | 24 |
| 1.3.3 Вторая предпосылка - нулевая средняя величина остатков, независимая от $X_i$ .....                                | 25 |
| 1.3.4 Третья предпосылка - гомоскедастичность остатков.....                                                             | 26 |
| 1.3.5 Четвертая предпосылка - отсутствие автокорреляции в остатках.....                                                 | 28 |
| 1.3.6 Пятая предпосылка - остатки подчиняются нормальному распределению.....                                            | 28 |
| 1.4 Контрольные вопросы к теоретическому разделу .....                                                                  | 30 |
| 2 Практическая часть .....                                                                                              | 31 |
| 2.1 Лабораторная работа 1. Расчет параметров и оценка уравнения множественной регрессии .....                           | 31 |
| 2.2 Лабораторная работа 2. Расчет парных коэффициентов корреляции и проверка наличия мультиколлинеарности факторов..... | 36 |
| Литература.....                                                                                                         | 46 |
| Приложение А (справочное) Значения F-критерия Фишера.....                                                               | 47 |
| Приложение Б (справочное) Критические значения t-критерия Стьюдента.....                                                | 48 |
| Приложение В (справочное) Значения критерия Пирсона.....                                                                | 49 |

## Введение

Специфической особенностью деятельности экономиста является работа в условиях недостатка информации и неполноты исходных данных. Анализ такой информации требует специальных методов, которые составляют один из аспектов эконометрики. Центральной проблемой эконометрики являются построение эконометрической модели и определение возможностей ее использования для описания, анализа и прогнозирования реальных экономических процессов.

В зависимости от эмпирических данных, которые берутся во внимание при определении спецификации модели, различают следующие виды эконометрических моделей: парная регрессия, множественная регрессия, системы эконометрических (одновременных) уравнений и модели временных рядов.

Парная регрессия может дать хороший результат при моделировании, если влиянием других факторов, воздействующих на объект исследования, можно пренебречь.

Однако на практике часто имеет место совокупное влияние факторов на результат, что приводит к необходимости построения множественной регрессии. Множественный регрессионный анализ имеет свою специфику, отличную от парной регрессии. В данном пособии изложены основные теоретические вопросы и этапы выполнения лабораторных работ по множественной регрессии. Для проверки полученных теоретических знаний составлен список контрольных вопросов, приведенный в конце первого раздела. Лабораторные работы оснащены заданиями для самостоятельной работы.

## 1 Теоретическая часть

Множественная регрессия используется в тех случаях, когда результативный признак зависит от двух и более факторов. Общий вид спецификации модели:

$$y_i = f(x_{1i}, x_{2i}, \dots, x_{ji}, \dots, x_{pn}),$$

где  $i = (1; n)$  – число наблюдений;  $j = (1; p)$  – число объясняющих переменных или факторов.

Множественная регрессия широко используется в решении проблем спроса, доходности акций, при изучении функции издержек производства, в макроэкономических расчетах и т.д.

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

### 1.1 Построение уравнения множественной регрессии

Построение уравнения множественной регрессии начинается с решения вопроса о спецификации модели, которая включает в себя два этапа:

- отбор факторов;
- выбор вида уравнения регрессии.

#### 1.1.1 Отбор факторов при построении множественной регрессии

Факторы, включенные во множественную регрессию должны отвечать следующим требованиям:

**а) факторы должны быть количественно измеримы.** Если необходимо включить в модель качественный фактор, то ему надо придать количественную определенность. Например, качество почвы может быть задано в виде баллов, а место нахождения недвижимости (районы) могут быть проранжированы. Такие факторы носят название фиктивных переменных [1].

*Фиктивные переменные* – это факторы, которые имеют качественные характеристики. Например, профессия, пол, образование, климатические условия, принадлежность к определённому региону.

Чтобы ввести такие переменные в регрессионную модель, им должны быть присвоены те или иные цифровые метки, то есть качественные переменные должны быть преобразованы в количественные. Такие преобразованные или сконструированные переменные называются *фиктивными* или *структурными* переменными.

Рассмотрим применение фиктивных переменных для функции спроса.

Предположим, что по группе лиц мужского и женского пола изучается линейная зависимость потребления кофе от цены.

В общем виде уравнение регрессии имеет вид:

$$y_i = a + bx_i + \varepsilon_i,$$

где  $y_i$  – количество потребляемого кофе;  $x_i$  – цена.

Для учета потребления кофе мужчинами и женщинами введем в модель фиктивную переменную  $z$ :

$$y_i = A + A_1 z + bx_i + \varepsilon_i.$$

При этом переменная  $z$  может принимать два значения: 1 – для мужчин, 2 – для женщин. Тогда уравнение зависимости потребления кофе от цены для мужчин будет иметь вид:

$$y_i = A + A_1 + bx_i,$$

а для женщин:

$$y_i = A + bx_i.$$

Таким образом, различие в уровне потребления кофе мужчинами и женщинами состоит в различии свободных членов данных уравнений:  $A$  – для женщин,  $(A + A_1)$  – для мужчин;

**б) при дополнительном включении в регрессию фактора с номером  $(p+1)$ , коэффициент детерминации  $R^2_{xy}$  должен возрастать, а остаточная дисперсия уменьшаться.** Насыщение модели лишними факторами не увеличивает  $R^2_{xy}$  и приводит к статистической незначимости параметров регрессии по критерию Стьюдента;

в) факторы не должны быть *интеркоррелированы* и тем более находиться в точной функциональной связи. Для проверки этого условия надо рассчитать коэффициент корреляции по формуле:

$$r_{x_1 x_2} = \frac{\overline{x_1 x_2} - \overline{x_1} \cdot \overline{x_2}}{\sigma_{x_1} \cdot \sigma_{x_2}},$$

где  $\sigma_{x_1} = \sqrt{\overline{x_1^2} - (\overline{x_1})^2}$  -- среднеквадратическое отклонение  $x_1$ ,

$\sigma_{x_2} = \sqrt{\overline{x_2^2} - (\overline{x_2})^2}$  -- среднеквадратическое отклонение  $x_2$ .

Если между факторами есть связь, то коэффициент корреляции для этих переменных стремится к единице. Это значит, что нельзя определить изолированное влияние каждого фактора на результативный признак, и параметры уравнения регрессии оказываются не интерпретируемыми.

Рассмотрим это на примере. Пусть имеется трехфакторная модель следующего вида:

$$y_i = a + b_1 x_{1i} + b_2 x_{2i} + b_3 x_{3i},$$

где  $y_i$  – цена автомобиля,  $x_{1i}$  – год выпуска,  $x_{2i}$  – мощность двигателя,  $x_{3i}$  – пробег.

Для проверки корреляции между факторами необходимо рассчитать коэффициенты корреляции по следующим формулам:

$$r_{x_1 x_2} = \frac{\overline{x_1 x_2} - \overline{x_1} \cdot \overline{x_2}}{\sigma_{x_1} \cdot \sigma_{x_2}},$$

$$r_{x_1 x_3} = \frac{\overline{x_1 x_3} - \overline{x_1} \cdot \overline{x_3}}{\sigma_{x_1} \cdot \sigma_{x_3}},$$

$$r_{x_2 x_3} = \frac{\overline{x_2 x_3} - \overline{x_2} \cdot \overline{x_3}}{\sigma_{x_2} \cdot \sigma_{x_3}}.$$

Если полученные значения коэффициентов корреляции стремятся к нулю, то можно говорить, что параметр  $b_1$  измеряет силу влияния фактора  $x_1$  на результат при неизменном значении факторов  $x_2, x_3$ . И так же для всех остальных параметров.

Если один из коэффициентов, например,  $r_{x_1 x_2}$  стремится к единице, то с изменением фактора  $x_1$  фактор  $x_2$  не может оставаться неизменным. Следовательно, параметры  $b_1$  и  $b_2$  нельзя интерпретировать как показатели раздельного влияния  $x_1$  и  $x_2$  на результативный признак  $y_i$ .

Коэффициенты корреляции позволяют исключать из модели дублирующие факторы. Считается, что две переменные явно коррелированы, то есть находятся между собой в линейной зависимости, если коэффициент корреляции больше или равен 0,7. При этом предпочтение отдается тому фактору, который имеет наименьшую тесноту связи с другими факторами, даже если этот фактор имеет хорошую связь с фактическим значением результативного признака  $y_i$ .

Для этого дополнительно рассчитываются линейные коэффициенты корреляции между каждым фактором и результативным признаком. Полу-

ченные данные сводятся в таблицу. Например, получены следующие значения коэффициентов корреляции, занесенные в таблицу 1.

Таблица 1 - Коэффициенты корреляции между переменными

|          | $y_i$ | $x_{1i}$   | $x_{2i}$ | $x_{3i}$ |
|----------|-------|------------|----------|----------|
| $y_1$    | 1     |            |          |          |
| $x_{1i}$ | 0,8   | 1          |          |          |
| $x_{2i}$ | 0,7   | <b>0,8</b> | 1        |          |
| $x_{3i}$ | 0,6   | 0,5        | 0,2      | 1        |

Очевидно, что факторы  $x_{1i}$  и  $x_{2i}$  дублируют друг друга ( $r_{x_1x_2} = 0,8$ ). Какой из них исключить? Корреляция фактора  $x_{1i}$  с  $y_i$  сильнее, чем  $x_{2i}$  с  $y_i$  ( $r_{yx_1} > r_{yx_2}$ ), но зато межфакторная корреляция с третьим фактором для  $x_{2i}$  меньше ( $r_{x_2x_3} < r_{x_1x_3}$ ).

Значит, предпочтение следует отдать второму фактору, то есть оставить его в модели, а первый фактор исключить.

Наибольшие трудности возникают при наличии *мультиколлинеарности* факторов, когда более чем два фактора связаны между собой линейной зависимостью, то есть имеет место совокупное воздействие факторов друг на друга.

Для оценки мультиколлинеарности факторов используют определитель матрицы парных коэффициентов корреляции между факторами:

$$\Delta R = \begin{vmatrix} r_{x_1x_1} & r_{x_2x_1} & r_{x_3x_1} \\ r_{x_1x_2} & r_{x_2x_2} & r_{x_3x_2} \\ r_{x_1x_3} & r_{x_2x_3} & r_{x_3x_3} \end{vmatrix}.$$

Если факторы не коррелируют между собой, то определитель матрицы равен единице:

$$\Delta R = \begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix} = 1.$$

Если между факторами существует полная линейная зависимость, то определитель матрицы равен нулю:

$$\Delta R = \begin{vmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{vmatrix} = 0.$$

Чем ближе  $\Delta R$  к нулю, тем сильнее мультиколлинеарность факторов и тем ненадежнее результаты множественной регрессии. И наоборот, чем ближе  $\Delta R$  к единице, тем слабее мультиколлинеарность факторов и тем ненадежнее результаты множественной регрессии.

Оценка значимости мультиколлинеарности факторов может быть проведена методом испытания гипотезы  $H_0$  о независимости переменных, то есть когда  $\Delta R=1$ . Эта оценка проводится по критерию Пирсона  $\chi^2$  по следующей формуле:

$$\chi^2 = \left[ n - 1 - \frac{1}{6}(2m + 5) \lg \Delta R \right].$$

Далее по таблице критериев Пирсона определяется критическое значение  $\chi^2_{\text{табл}}$  при числе степенной свободы  $df$ :

$$df = \frac{1}{2}n(n-1).$$

Если фактическое значение  $\chi^2$  больше табличного, то гипотеза  $H_0$  отклоняется, следовательно,  $\Delta R \neq 1$  и недиагональные (ненулевые) коэффициенты корреляции указывают на коллинеарность факторов. Мультиколлинеарность считается доказанной.

Включение в модель мультиколлинеарных факторов нежелательно в силу следующих последствий:

- затрудняется интерпретация параметров множественной регрессии, они теряют эконометрический смысл;
- оценки параметров ненадежны, обнаруживают больше стандартных ошибок и меняются с изменением объема наблюдений, что делает модель непригодной для анализа и прогнозирования.

Отбор факторов, включаемых в регрессию, является одним из важнейших этапов построения уравнения множественной регрессии. Подходы к отбору факторов на основе показателей корреляции могут быть разные. Они определяют методику построения уравнения регрессии.

Наиболее широкое распространение получили следующие методы построения уравнения множественной регрессии:

- метод исключения (отсев факторов из полного его набора);
- метод включения (введение дополнительного фактора);

- пошаговый регрессионный анализ (исключение ранее введенного фактора).

Матрица парных коэффициентов корреляции играет важную роль в отборе факторов. Но решение вопроса о целесообразности включения того или иного фактора окончательно принимается после расчета показателей частной корреляции, которые оценивают тесноту связи фактора с результатом в чистом виде.

При отборе факторов рекомендуется пользоваться следующим правилом: число включаемых факторов должно быть в 6-7 раз меньше объема совокупности  $n$ , по которой строится регрессия. Например, для трех факторов  $n=18-21$ . Если это соотношение нарушено, то параметры уравнения регрессии оказываются статистически незначимыми, а  $F$ -критерий получается меньше табличного значения.

### 1.1.2 Выбор формы уравнения множественной регрессии

Уравнения множественной регрессии могут быть линейными и нелинейными.

Линейная регрессия:

$$\hat{y}_i = a + b_1 x_{1i} + b_2 x_{2i} + K + b_p x_{pi}.$$

Нелинейные регрессии:

- степенная:

$$\hat{y}_i = a \cdot x_{1i}^{b_1} \cdot x_{2i}^{b_2} \cdot K \cdot x_{pi}^{b_p};$$

- экспонента:

$$\hat{y}_i = e^{a + b_1 x_{1i} + b_2 x_{2i} + K + b_p x_{pi}};$$

- гипербола:

$$\hat{y}_i = \frac{1}{a + b_1 x_{1i} + b_2 x_{2i} + K + b_p x_{pi}}.$$

Можно использовать и другие функции, приводимые к линейному виду. Наиболее широко используются линейная и степенная функции ввиду четкой интерпретации параметров.

В *линейной* множественной регрессии параметры при объясняющих переменных называются коэффициентами «чистой» регрессии. Они характеризуют среднее изменение результата при изменении соответствующего фактора на единицу при неизменном значении других факторов, закрепленных на среднем уровне. Параметр  $a$ , как и в парной регрессии, не подлежит экономической интерпретации.

Например, предположим, что зависимость расходов на продукты питания характеризуется следующим уравнением [3]:

$$y_i = 0,5 + 0,35x_{1i} + 0,73x_{2i},$$

где  $y_i$  – расходы семьи за один месяц на продукты питания, тыс. руб.;

$x_{1i}$  – месячный доход на одного члена семьи, тыс. руб.;

$x_{2i}$  – размер семьи, человек.

По этому уравнению можно сделать следующие выводы:

а) с ростом дохода ( $x_1$ ) на одного члена семьи на 1 тыс.руб. расходы на питание возрастут в среднем на 350 руб. при том же среднем размере семьи;

б) увеличение размера семьи ( $x_2$ ) на одного человека, при тех же доходах ( $x_1$ ) предполагает дополнительный рост расходов на питание на 730 руб.

В *степенной* функции коэффициент  $b_j$  является коэффициентом эластичности. Он показывает, *на сколько процентов изменится результат с изменением соответствующего фактора на 1 % при неизменном действии других факторов.*

Этот вид уравнения получил наибольшее распространение в производственных функциях, в исследованиях спроса и потребления.

Например, предположим, что при исследовании спроса на мясо получено уравнение:

$$y_i = 0,82 \cdot x_{1i}^{-2,63} + x_{2i}^{1,11},$$

где  $y_i$  – количество спрашиваемого мяса;  $x_{1i}$  – цена;  $x_{2i}$  – доход.

Из уравнения следует, что рост цен на 1 %, при неизменном доходе вызывает снижение спроса в среднем на 2,63 %. А увеличение дохода на 1 % ведет к росту спроса на 1,11 % при неизменных ценах.

В производственных функциях вида:

$$P = a \cdot F_1^{b_1} \cdot F_2^{b_2} \cdot K \cdot F_m^{b_m} \cdot \varepsilon,$$

где  $P$  – количество продукта, которое изготовлено с помощью  $m$  производственных факторов,  $b_j$  – эластичность количества продукции по отношению к количеству соответствующих производственных факторов, экономический смысл имеют не только коэффициенты каждого фактора, но и их сумма (сумма эластичностей):

$$B = b_1 + b_2 + \dots + b_m.$$

Эта величина фиксирует обобщенную характеристику эластичности производства.

Например, построена следующая производственная функция:

$$P = 2 \cdot F_1^{0,3} \cdot F_2^{0,2} \cdot F_3^{0,5} \cdot \varepsilon, \quad (1)$$

где  $P$  – выпуск продукции;

$F_1$  – стоимость основных производственных фондов;

$F_2$  – отработано человеко-дней;

$F_3$  – затраты на производство.

Согласно уравнению (1) эластичность выпуска по отдельным факторам производства составляет в среднем 0,3 % с ростом  $F_1$  на 1 % при неизменном уровне других факторов; 0,2 % – с ростом  $F_2$  на 1 % также при неизменности других факторов производства и на 0,5 % с ростом  $F_3$  на 1 % при неизменном уровне факторов  $F_1$  и  $F_2$ .

Для данного уравнения  $B=b_1+b_2+b_3=1$ . Следовательно, в целом с ростом каждого фактора производства на 1% коэффициент эластичности выпуска продукции составляет 1 %, что в микроэкономике соответствует постоянной отдаче на масштаб.

### 1.1.3 Расчет параметров уравнений множественной регрессии

Для расчета параметров линейного уравнения множественной регрессии, применяется метод наименьших квадратов (МНК):

- непосредственно к уравнению множественной регрессии;
- к уравнению регрессии в стандартизованном масштабе.

По *первому способу* для расчета параметров линейного уравнения множественной регрессии, как и в парной регрессии, определяется минимальное значение суммы квадратов отклонений:

$$S = \sum (y_i - \hat{y}_i)^2 \rightarrow \min. \quad (2)$$

Рассмотрим вывод формул для расчета параметров трехфакторного уравнения линейной регрессии:

$$\hat{y}_i = a + b_1 x_{1i} + b_2 x_{2i} + b_3 x_{3i}. \quad (3)$$

Приравниваем к нулю частные производные от выражения (2) по коэффициентам уравнения множественной регрессии:

$$\frac{\partial S}{\partial a} = 0, \quad \frac{\partial S}{\partial b_1} = 0, \quad \frac{\partial S}{\partial b_2} = 0, \quad \frac{\partial S}{\partial b_3} = 0$$

и получаем систему нормальных уравнений:

$$\begin{cases} \sum y_i = na + b_1 \sum x_{1i} + b_2 \sum x_{2i} + b_3 \sum x_{3i}, \\ \sum y_i x_{1i} = a \sum x_{1i} + b_1 \sum x_{1i}^2 + b_2 \sum x_{1i} x_{2i} + b_3 \sum x_{1i} x_{3i}, \\ \sum y_i x_{2i} = a \sum x_{2i} + b_1 \sum x_{2i} x_{1i} + b_2 \sum x_{2i}^2 + b_3 \sum x_{2i} x_{3i}, \\ \sum y_i x_{3i} = a \sum x_{3i} + b_1 \sum x_{3i} x_{1i} + b_2 \sum x_{3i} x_{2i} + b_3 \sum x_{3i}^2. \end{cases}$$

Для её решения может быть применён метод определителей (метод Крамера), и коэффициенты уравнения будут вычисляться по формулам:

$$a = \frac{\Delta a}{\Delta}, \quad b_1 = \frac{\Delta b_1}{\Delta}, \quad b_2 = \frac{\Delta b_2}{\Delta}, \quad b_3 = \frac{\Delta b_3}{\Delta},$$

где  $\Delta$  - определитель системы:

$$\Delta = \begin{vmatrix} n & \sum x_{1i} & \sum x_{2i} & \sum x_{3i} \\ \sum x_{1i} & \sum x_{1i}^2 & \sum x_{1i} x_{2i} & \sum x_{1i} x_{3i} \\ \sum x_{2i} & \sum x_{2i} x_{1i} & \sum x_{2i}^2 & \sum x_{2i} x_{3i} \\ \sum x_{3i} & \sum x_{3i} x_{1i} & \sum x_{3i} x_{2i} & \sum x_{3i}^2 \end{vmatrix},$$

а  $\Delta a, \Delta b_1, \Delta b_2, \Delta b_3$  – частные определители, которые получаются путем замены соответствующего столбца матрицы определителя  $\Delta$  данными левой части системы уравнений [7].

**Второй способ** основан на построении уравнения регрессии в стандартизованном масштабе:

$$t_{yi} = \beta_1 t_{x_{1i}} + \beta_2 t_{x_{2i}} + \beta_3 t_{x_{3i}} + \varepsilon_i, \quad (4)$$

где  $t_{yi}, t_{x_{1i}}, t_{x_{2i}}, t_{x_{3i}}$  – стандартизованные переменные, которые рассчитываются по следующим формулам:

$$t_{yi} = \frac{y_i - \bar{y}}{\sigma_y},$$

$$t_{x_{ji}} = \frac{x_{ji} - \bar{x}_j}{\sigma_{x_{ji}}},$$

где  $j=1, 2, 3$  (для трехфакторной регрессии).

Коэффициенты  $\beta_j$  – это стандартизованные коэффициенты уравнения регрессии, которые находятся по методу МНК. Для этого минимизируется квадрат отклонений расчетных и теоретических значений стандартизованных переменных:

$$S = \sum (t_{y_i} - \hat{t}_{y_i})^2 \rightarrow \min$$

и после соответствующих преобразований строится система нормальных уравнений вида:

$$\begin{cases} r_{yx_i} = \beta_1 + \beta_2 r_{x_1 x_2} + \beta_3 r_{x_1 x_3}, \\ r_{yx_2} = \beta_1 r_{x_2 x_1} + \beta_2 + \beta_3 r_{x_2 x_3}, \\ r_{yx_3} = \beta_1 r_{x_3 x_1} + \beta_2 r_{x_3 x_2} + \beta_3, \end{cases}$$

где  $r$  – это парные коэффициенты корреляции.

Решая систему методом определителей, находят параметры – стандартизованные коэффициенты регрессии  $\beta_j$ . Для нахождения «чистых» параметров уравнения множественной регрессии пользуются соотношением:

$$b_j = \beta_j \frac{\sigma_y}{\sigma_{x_j}}.$$

Параметр  $a$  определяется как:

$$a = \bar{y} - b_1 \bar{x}_1 - b_2 \bar{x}_2 - b_3 \bar{x}_3.$$

Это позволяет от уравнения регрессии в стандартизованном масштабе (4) переходить к уравнению регрессии в натуральном масштабе переменных (3).

Стандартизованные коэффициенты  $\beta_j$  сравнивают между собой, чтобы ранжировать факторы по силе их воздействия на результат. В этом основное достоинство стандартизованных коэффициентов регрессии в отличие от коэффициентов «чистой» регрессии, которые несравнимы между собой.

Рассмотрим это на примере.

Пусть функция издержек производства ( $y_i$ , тыс. руб.) характеризуется следующим уравнением:

$$y_i = 200 + 1,2 \cdot x_{1i} + 1,1 \cdot x_{2i} + \varepsilon_i, \quad (5)$$

где  $x_{1i}$  – основные производственные фонды, тыс. руб.;

$x_{2i}$  – численность занятых в производстве, чел.

Анализ уравнения (5) показывает, что при той же занятости дополнительный рост стоимости основных производственных фондов на одну тыс. рублей влечет за собой увеличение затрат в среднем на 1,2 тыс. руб., а увеличение численности занятых на одного человека способствует при той же технической оснащенности предприятий росту затрат в среднем на 1,1 тыс. руб.

Однако, это не означает, что фактор  $x_{1i}$  оказывает более сильное влияние на издержки производства по сравнению с фактором  $x_{2i}$ . Такое сравнение возможно, если обратиться к уравнению регрессии в стандартизованном масштабе (4). Предположим, оно выглядит так:

$$t'_{yi} = 0,5 \cdot t_{x_{1i}} + 0,8 \cdot t_{x_{2i}}.$$

Так как  $\beta_1 < \beta_2$  ( $0,5 < 0,8$ ), то можно заключить, что большее влияние оказывает фактор  $x_{2i}$ .

Для *нелинейной* зависимости признаков, приводимой к линейному виду, параметры множественной регрессии также определяются методом наименьших квадратов.

Например, степенная функция:

$$y_i = a \cdot x_{1i}^{b_1} \cdot x_{2i}^{b_2} \cdot K \cdot x_{pi}^{b_p} \cdot \varepsilon_i$$

после преобразования в линейную будет выглядеть так:

$$\ln y_i = \ln a + b_1 \cdot \ln x_{1i} + b_2 \cdot \ln x_{2i} + K + b_p \cdot \ln x_{pi} + \ln \varepsilon_i.$$

Полученное уравнение обрабатывается МНК, то есть строится система нормальных уравнений, определяются параметры  $b_j$  и т.д.

## 1.2 Оценка множественной регрессии

Особенность оценки множественной регрессии состоит в том, что, помимо расчета общих характеристик, рассчитывают частные характеристики. Это необходимо для того, чтобы определить или оценить влияния каждого фактора на результат и сделать вывод о целесообразности включения этого фактора в регрессионную модель.

### 1.2.1 Общие оценки уравнений множественной регрессии

Как и в парной регрессии, общие оценки оценивают качество построенной модели. К ним относятся:

**а) индекс множественной корреляции** оценивает тесноту связи результата со всей совокупностью факторов и вычисляется по формуле (для трехфакторной регрессии):

$$R_{yx_1x_2x_3} = \sqrt{1 - \frac{\sigma_{ост}^2}{\sigma_{общ}^2}} = \sqrt{1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2}}.$$

Индекс множественной корреляции находится в границах:  
 $0 \leq R_{yx_1x_2x_3} \leq 1$ .

Связь результативного признака со всем набором исследуемых факторов должна быть больше или равна максимальному значению парного коэффициента корреляции:

$$R_{yx_1x_2x_3} \geq r_{yx_j} (\max).$$

Если нет, то уравнение имеет лишние факторы, которые не способствуют увеличению доли объясненной вариации результативного признака.

Для линейной множественной регрессии формула индекса множественной корреляции получила название *линейного коэффициента множественной корреляции* или, что тоже самое, *совокупного коэффициента корреляции*.

При линейной зависимости определение коэффициента множественной корреляции возможно через матрицу парных коэффициентов корреляции по формуле:

$$R_{yx_1x_2x_3} = \sqrt{1 - \frac{\Delta_R}{\Delta_{R_{11}}}}, \quad (6)$$

где  $\Delta_R$  – определитель матрицы парных коэффициентов корреляции:

$$\Delta R = \begin{vmatrix} 1 & r_{yx_1} & r_{yx_2} & r_{yx_3} \\ r_{yx_1} & 1 & r_{x_1x_2} & r_{x_1x_3} \\ r_{yx_2} & r_{x_2x_1} & 1 & r_{x_2x_3} \\ r_{yx_3} & r_{x_3x_1} & r_{x_3x_2} & 1 \end{vmatrix},$$

а  $\Delta_{R_{11}}$  – определитель матрицы межфакторной корреляции. Определитель  $\Delta_{R_{11}}$  получается путем вычеркивания из определителя  $\Delta_R$  первого столбца и первой строки. Получается матрица коэффициентов парной корреляции между факторами:

$$\Delta R_{11} = \begin{vmatrix} 1 & r_{x_1 x_2} & r_{x_1 x_3} \\ r_{x_2 x_1} & 1 & r_{x_2 x_3} \\ r_{x_3 x_1} & r_{x_3 x_2} & 1 \end{vmatrix}.$$

Таким образом, формула (6) позволяет определять совокупный коэффициент корреляции, не обращаясь к уравнению множественной регрессии, а используя лишь парные коэффициенты корреляции;

**б) индекс множественной детерминации** оценивает качество подбора уравнения регрессии и для линейной множественной регрессии вычисляется по формуле:

$$R^2_{yx_1 x_2 x_3} = (R_{yx_1 x_2 x_3})^2 = 1 - \frac{\sigma^2_{ост}}{\sigma^2_{общ}}.$$

Индекс множественной детерминации находится в границах:  $0 \leq R^2_{yx_1 x_2 x_3} \leq 1$ .

Чем ближе  $R^2_{yx_1 x_2 x_3}$  к единице, тем больше доля объясненной вариации, тем меньше влияние остальных, не учтенных в модели факторов.

Для регрессий, нелинейных по оцениваемым параметрам, рассчитывают скорректированный индекс множественной детерминации по формуле:

$$\overline{R^2} = 1 - (1 - R^2) \cdot \frac{(n-1)}{(n-m-1)}, \quad (7)$$

где  $m$  – число параметров при переменных  $x_{ji}$ ;

$n$  – число наблюдений.

Дело в том, что при расчете параметров регрессий, нелинейных по параметрам, МНК применяется не к исходным данным, а к их логарифмам. Это приводит к систематической ошибке при расчете остаточной дисперсии. Если число параметров при  $x_{ji}$  равно  $m$  и приближается к объему наблюдений, то остаточная дисперсия будет близка к нулю даже при слабой связи факторов с результатом. Поэтому в формуле (7) скорректированный индекс множественной детерминации содержит поправку на число степеней свободы. Чем больше величина  $m$ , тем сильнее различия между  $\overline{R^2}$  и  $R^2$ ;

**в) критерий Фишера** оценивает надежность результатов множественной регрессии и вычисляется по формуле:

$$F = \frac{D_{\text{регр}}}{D_{\text{ост}}} = \frac{R^2_{y x_1 x_2 x_3}}{1 - R^2_{y x_1 x_2 x_3}} \cdot \frac{n - m - 1}{m},$$

где  $D_{\text{регр}}$  – факторная сумма квадратов отклонений, рассчитанная на одну степень свободы;

$D_{\text{ост}}$  – остаточная сумма квадратов отклонений, рассчитанная на одну степень свободы;

$R^2_{y x_1 x_2 x_3}$  – множественный коэффициент детерминации;

$m$  – число параметров при объясняющих переменных;

$n$  – число наблюдений.

Уравнение множественной регрессии считается надежным, если  $F > F_{\text{табл.}}$ .

### 1.2.2 Частные оценки уравнений множественной регрессии

Для расчета частных оценок множественной регрессии строят частные уравнения, которые связывают результативный признак с одной объясняющей переменной при закреплении других факторов на среднем уровне [1].

Для трехфакторного уравнения регрессии  $\hat{y}_i = a + b_1 x_{1i} + b_2 x_{2i} + b_3 x_{3i}$  используются следующие частные уравнения:

$$\begin{aligned} \hat{y}_{x_1 \bullet x_2 x_3} &= a + b_1 x_{1i} + b_2 \bar{x}_2 + b_3 \bar{x}_3, \\ \hat{y}_{x_2 \bullet x_1 x_3} &= a + b_1 \bar{x}_1 + b_2 x_{2i} + b_3 \bar{x}_3, \\ \hat{y}_{x_3 \bullet x_1 x_2} &= a + b_1 \bar{x}_1 + b_2 \bar{x}_2 + b_3 x_{3i}, \end{aligned}$$

в которых  $\hat{y}_{x_1 \bullet x_2 x_3}$ ,  $\hat{y}_{x_2 \bullet x_1 x_3}$ ,  $\hat{y}_{x_3 \bullet x_1 x_2}$  означают зависимость результативного признака от фактора  $x_j$  при закреплении других факторов на среднем уровне. Например, первое уравнение показывает зависимость результативного признака от фактора  $x_1$  при закреплении других факторов на среднем уровне.

Если  $A_1 = a + b_2 \bar{x}_2 + b_3 \bar{x}_3$ , то уравнение приобретает вид парного линейного уравнения регрессии:

$$\hat{y}_{x_1 \bullet x_2 x_3} = A_1 + b_1 x_{1i}.$$

Аналогично для остальных уравнений:

$$\hat{y}_{x_2 \bullet x_1 x_3} = A_2 + b_2 x_{2i} ,$$

$$\hat{y}_{x_3 \bullet x_1 x_2} = A_3 + b_3 x_{3i} .$$

В отличие от парной регрессии частные уравнения множественной регрессии характеризуют изолированное влияние фактора на результат, так как другие факторы закреплены на среднем уровне. Это позволяет определять частные оценки, к которым относятся:

**а) частный коэффициент эластичности** показывает, на сколько процентов возрастет результативный признак с ростом фактора на 1%. Вычисляется по формуле:

$$\bar{\varepsilon}_{yx_1} = b_1 \cdot \frac{\bar{x}_1}{\bar{y}_{x_1 \bullet x_2 x_3}} ;$$

**б) частный коэффициент корреляции** позволяет измерить тесноту связи между результативным признаком и одним фактором при устранении влияния всех других факторов. Например, чтобы измерить тесноту связи результативного признака с первым фактором, необходимо рассчитать коэффициент корреляции по следующей формуле:

$$r_{y_{x_1 \bullet x_2 x_3}} = \sqrt{1 - \frac{1 - R^2_{yx_1 x_2 x_3}}{1 - R^2_{yx_2 x_3}}} ,$$

где  $R^2_{yx_1 x_2 x_3}$  – множественный коэффициент детерминации всего комплекса факторов с результатом;

$R^2_{yx_2 x_3}$  – множественный коэффициент детерминации, но без введения в модель фактора  $x_1$ , то есть для модели вида:  
 $\hat{y}_i = a + b_2 x_{2i} + b_3 x_{3i} .$

Частный коэффициент корреляции можно рассчитать по *рекуррентной* формуле через коэффициенты частной корреляции более низких порядков. Рассмотрим формулы для расчета частных коэффициентов корреляции для трехфакторного уравнения множественной регрессии:

$$\hat{y}_i = a + b_1 x_{1i} + b_2 x_{2i} + b_3 x_{3i} .$$

Порядок коэффициентов корреляции зависит от количества взятых во внимание факторов:

- $r_{yx_1}$  – коэффициент парной корреляции – коэффициент *нулевого* порядка;
- $r_{yx_1 \bullet x_2}$  – коэффициент частной корреляции *первого* порядка;
- $r_{yx_1 \bullet x_2 x_3}$  – коэффициент частной корреляции *второго* порядка и т.д.

Общая формула для расчета частных коэффициентов корреляции выглядит следующим образом:

$$r_{yx_j \bullet x_1 x_2 \dots x_p} = \frac{r_{yx_j \bullet x_1 x_2 \dots x_{p-1}} - r_{yx_p \bullet x_1 x_2 \dots x_{p-1}} \cdot r_{x_j x_p \bullet x_1 x_2 \dots x_{p-1}}}{\sqrt{(1 - r_{yx_p \bullet x_1 x_2 \dots x_{p-1}}^2) \cdot (1 - r_{x_j x_p \bullet x_1 x_2 \dots x_{p-1}}^2)}} \quad (8)$$

Согласно выражению (8) формула для расчета коэффициента частной корреляции для трехфакторной множественной регрессии между  $y$  и  $x_1$  при закреплении факторов  $x_2$  и  $x_3$  на среднем уровне будет следующая:

$$r_{yx_1 \bullet x_2 x_3} = \frac{r_{yx_1 \bullet x_2} - r_{yx_3 \bullet x_2} \cdot r_{x_1 x_3 \bullet x_2}}{\sqrt{(1 - r_{yx_3 \bullet x_2}^2) \cdot (1 - r_{x_1 x_3 \bullet x_2}^2)}} \quad (9)$$

В формуле (9) использованы коэффициенты корреляции *первого* порядка:

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

$$r_{yx_3 \bullet x_2} = \frac{r_{yx_3} - r_{yx_2} \cdot r_{x_3 x_2}}{\sqrt{(1 - r_{yx_2}^2) \cdot (1 - r_{x_3 x_2}^2)}},$$

$$r_{x_1 x_3 \bullet x_2} = \frac{r_{x_1 x_3} - r_{x_1 x_2} \cdot r_{x_3 x_2}}{\sqrt{(1 - r_{x_1 x_2}^2) \cdot (1 - r_{x_3 x_2}^2)}}.$$

Сравнение частных коэффициентов корреляции позволяет ранжировать факторы по тесноте их связи с результатом. Они дают, в отличие от стандартизованных коэффициентов  $\beta_j$ , конкретную меру тесноты связи каждого фактора с результатом в чистом виде. Причем соблюдается следующее соответствие: если  $\beta_1 > \beta_2 > \beta_3$ , то  $r_{yx_1 \bullet x_2 x_3} > r_{yx_2 \bullet x_1 x_3} > r_{yx_3 \bullet x_1 x_2}$ .

**в) частный критерий Фишера** рассчитывают для оценки фактора, дополнительно включенного в модель, так как не каждый фактор может существенно увеличить долю объяснённой вариации результативного признака. Например, расчет частного критерия Фишера для оценки первого фактора в трехфакторной регрессии производится по формуле:

$$F_{x_1} = \frac{R^2_{yx_1 x_2 x_3} - R^2_{yx_2 x_3}}{1 - R^2_{yx_1 x_2 x_3}} \cdot \frac{n - m - 1}{1},$$

где  $R^2_{yx_1 x_2 x_3}$  – коэффициент множественной детерминации для модели с полным набором факторов;

$R^2_{yx_2 x_3}$  – тот же показатель, но без включения в модель фактора  $x_1$ ;

$n$  – число наблюдений;

$m$  – число параметров в модели (без свободного члена).

Величина  $(R^2_{yx_1 x_2 x_3} - R^2_{yx_2 x_3})$  является приростом доли объяснённой вариации за счет дополнительного включения в модель соответствующего фактора (в данном случае  $x_1$ ).

Фактическое значение частного F-критерия сравнивается с табличным при 5%-ном или 1%-ном уровне значимости  $\alpha$  и числе степеней свободы:  $df_1 = 1$  и  $df_2 = n - m - 1$ . Число степеней свободы  $df_1 = 1$ , так как прирост факторной суммы квадратов отклонений обусловлен дополнительным включением в модель одного исследуемого фактора (например,  $x_1$ ). Для остаточной суммы квадратов отклонений по регрессионной модели число степеней свободы равно:  $df_2 = n - m - 1$ .

Если фактическое значение частного критерия Фишера больше табличного значения, то дополнительное включение фактора  $x_1$  в модель статистически оправдано, и коэффициент чистой регрессии  $b_1$  при факторе  $x_1$  статистически значим.

### 1.2.3 Оценка значимости коэффициентов чистой регрессии по критерию Стьюдента

Для оценки каждого коэффициента используется, как и в парной регрессии, формула для расчета фактических критериев Стьюдента [1,6].

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

где  $b_j$  – коэффициент чистой регрессии при факторе  $X_{ji}$ ;

$m_{b_j}$  – среднеквадратическая ошибка коэффициента регрессии  $b_j$ , которая рассчитывается по формуле:

$$m_{b_j} = \frac{\sigma_y \sqrt{1 - R^2_{yX_1 X_2 \dots X_p}}}{\sigma_{x_j} \sqrt{1 - R^2_{x_j X_1 \dots X_p}}} \cdot \frac{1}{\sqrt{n - m - 1}},$$

где  $\sigma_y$  – среднеквадратическое отклонение  $y$ ;

$\sigma_{x_j}$  – среднеквадратическое отклонение  $x_j$ ;

$R^2_{yX_1 X_2 \dots X_p}$  – коэффициент детерминации для уравнения множественной регрессии;

$R^2_{x_j X_1 \dots X_p}$  – коэффициент детерминации для зависимости фактора  $x_j$  со всеми другими факторами уравнения множественной регрессии;

$(n-m-1)$  – число степеней свободы для остаточной суммы квадратов отклонений.

Для расчета  $R^2_{x_j X_1 \dots X_p}$  необходима матрица межфакторной корреляции и расчет по ней соответствующих коэффициентов детерминации.

Например, для уравнения  $y_i = a + b_1 x_{1i} + b_2 x_{2i} + b_3 x_{3i} + \varepsilon_i$  оценка значимости коэффициентов регрессии  $b_1, b_2, b_3$ , предполагает расчет трёх межфакторных коэффициентов детерминации:  $R^2_{x_1 \bullet x_2 x_3}$ ,  $R^2_{x_2 \bullet x_1 x_3}$ ,  $R^2_{x_3 \bullet x_1 x_2}$ .

Фактическое значение  $t$ -критерия можно также рассчитать через частные  $F$ -критерии по формуле:

$$t_{b_j} = \sqrt{F_{x_j}}.$$

Если величина  $t_{b_j} > t_{табл}$ , то значение коэффициента регрессии статистически достоверно.

### 1.3 Исследование остаточных величин регрессии $\varepsilon_i$ (предпосылки МНК)

#### 1.3.1 Свойства случайных остатков

Метод наименьших квадратов (МНК) строит оценки регрессии на основе минимизации суммы квадратов остатков [1]:

$$\sum \varepsilon_i^2 = \sum (y_i - \hat{y}_i)^2 \rightarrow \min .$$

Случайная составляющая  $\varepsilon_i$  представляет собой ненаблюдаемую величину, т.е. она появляется только в результате расчёта  $\hat{y}_i$ . При изменении спецификации модели, в том числе при добавлении в неё новых наблюдений, остатки  $\varepsilon_i$  могут меняться и влиять при этом на свойства регрессии. Поэтому после построения уравнения регрессии проводится проверка наличия у случайных остатков  $\varepsilon_i$  следующих свойств:

- остатки представляют собой независимые случайные величины и их среднее значение или математическое ожидание равно нулю;
- остатки имеют одинаковую постоянную дисперсию и подчиняются нормальному закону распределения.

Наличие этих свойств говорит о том, что оценки параметров регрессии являются *несмещёнными, состоятельными и эффективными*.

*Несмещённость* оценки означает, что математическое ожидание остатков равно нулю:

$$\frac{\sum \varepsilon_i}{n} = 0 .$$

*Эффективность* оценки означает, что остаточная дисперсия стремится к минимуму:

$$D_{ост} \rightarrow \min .$$

*Состоятельность* оценок – это увеличение их точности с увеличением объёма выборки. То есть доверительный интервал стремится к минимуму:

$$\gamma \rightarrow \min .$$

Условия, необходимые для получения несмещённых, состоятельных и эффективных оценок, представляют собой *предпосылки МНК*, соблюдение которых желательно для получения достоверных результатов регрессии.

Исследование остатков  $\varepsilon_i$  предполагает проверку наличия следующих пяти предпосылок МНК:

- случайный характер остатков;
- нулевая средняя величина остатков, независящая от  $x_i$ ;

- гомоскедастичность остатков означает, что дисперсия каждого отклонения  $\varepsilon_i$  одинакова для всех значений  $X_i$ ;

- отсутствие автокорреляции, т.е. значения  $\varepsilon_i$  распределены независимо друг от друга;

- остатки подчиняются нормальному распределению.

Если распределение случайных остатков не соответствует некоторым предпосылкам МНК, то следует корректировать модель.

Рассмотрим предпосылки МНК более подробно.

### 1.3.2 Первая предпосылка – случайный характер остатков

Для проверки случайного характера остатков строится график зависимости остатков от значений результативного признака  $y_i$ , как на рисунке 1.

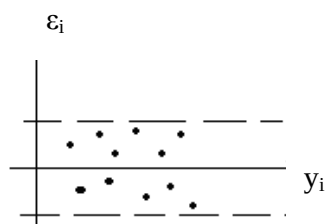


Рисунок 1 – График зависимости остатков от значений результативного признака (остатки случайны)

Если на графике получена горизонтальная полоса, то остатки представляют собой случайные величины, т.е. не зависят от  $y_i$ ; МНК оправдан, теоретические значения  $y_i$  хорошо описывают (аппроксимируют) фактические значения  $y_i$ . Если  $\varepsilon_i$  зависит от  $y_i$ , то возможны следующие случаи:

а) остатки  $\varepsilon_i$  неслучайны, как на рисунке 2;

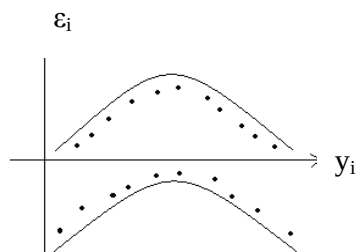


Рисунок 2 – График зависимости остатков от значений результативного признака  $y_i$  (остатки не случайны)

б) остатки  $\varepsilon_i$  не имеют постоянной дисперсии, как показано на рисунке 3;

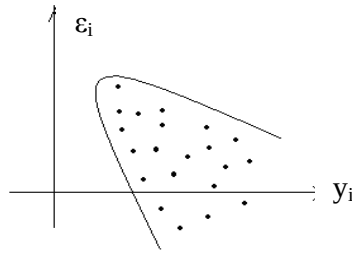


Рисунок 3 – График зависимости остатков от значений  
результативного признака  $\hat{y}_i$   
(остатки не имеют постоянной дисперсии)

в) остатки  $\varepsilon_i$  носят систематический характер, показанный на рисунке 4. В данном случае отрицательные значения остатков соответствуют низким значениям  $\hat{y}_i$ , а положительные – высоким.

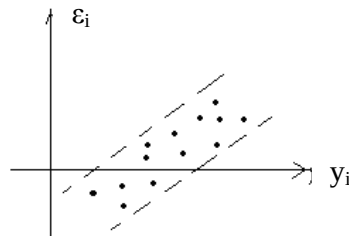


Рисунок 4 – График зависимости остатков от значений  
результативного признака  $\hat{y}_i$   
(остатки носят систематический характер)

В случае обнаружения зависимости  $\varepsilon_i$  от  $\hat{y}_i$  надо либо применить другую функцию, либо вводить дополнительную информацию и заново строить уравнение регрессии до тех пор, пока остатки не станут случайными величинами.

### 1.3.3 Вторая предпосылка – нулевая средняя величина остатков, независимая от $x_i$

Нулевая средняя величина остатков, независимая от  $x_i$ , означает несмещенность оценки. Для линейных моделей и моделей, нелинейных по переменным, это означает соблюдение условия:

$$\sum (y_i - \hat{y}_i) = 0. \quad (10)$$

Для моделей, нелинейных по параметрам:

$$\sum (\ln y_i - \ln \hat{y}_i) = 0. \quad (11)$$

Условия (10) и (11) говорят о том, что остатки не зависят от  $x$ . Они равномерно распределены на графике. Дополнительно строятся графики зависимости  $\varepsilon_i$  от факторов, включённых в регрессию в соответствии с рисунком 5.

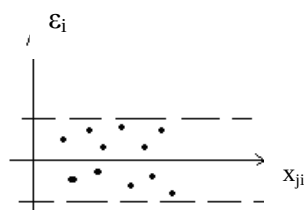


Рисунок 5 – График зависимости  $\varepsilon_i$  от фактора, включённого в регрессию

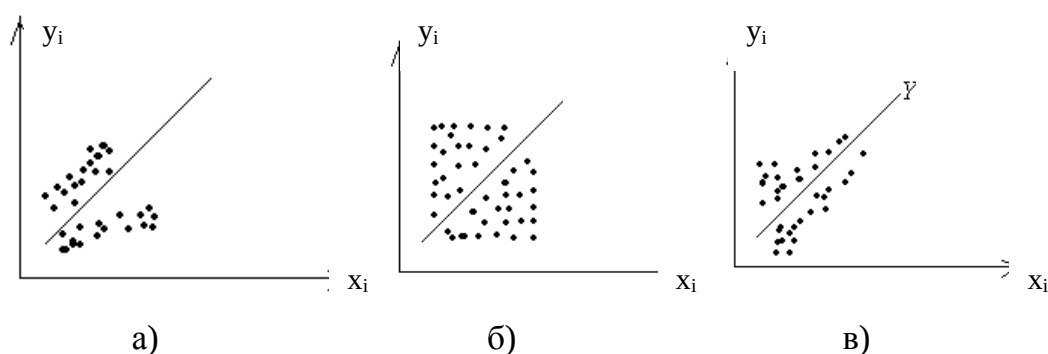
Если остатки  $\varepsilon_i$  на графике расположены в виде горизонтальной полосы, то они независимы от значений  $x_{ji}$ .

Если график показывает наличие зависимости  $\varepsilon_i$  от  $x_{ji}$ , то модель неадекватна. Скопление точек в определённых участках говорит о наличии систематической погрешности модели.

### 1.3.4 Третья предпосылка – гомоскедастичность остатков

*Гомоскедастичность* остатков означает, что дисперсия каждого отклонения  $\varepsilon_i$  одинакова для всех значений  $x_i$ . Если это условие не соблюдается, то имеет место *гетероскедастичность*.

Гетероскедастичность имеет место в следующих случаях, представленных на рисунке 6.



а – дисперсия растёт по мере увеличения  $x_i$ ; б – дисперсия остатков растёт при средних значениях  $x_i$  и уменьшается при min и max значениях  $x_i$ ; в – максимальная дисперсия остатков при малых значениях  $x_i$ , а по мере увеличения  $x_i$  дисперсия однородна

Рисунок 6 – Примеры гетероскедастичности остатков

Проверку на гомоскедастичность можно осуществить и с помощью графиков зависимости  $\varepsilon_i$  от  $y_i$ . Для случаев, рассмотренных на рисунке 6, графики зависимости будут выглядеть в соответствии с рисунком 7.

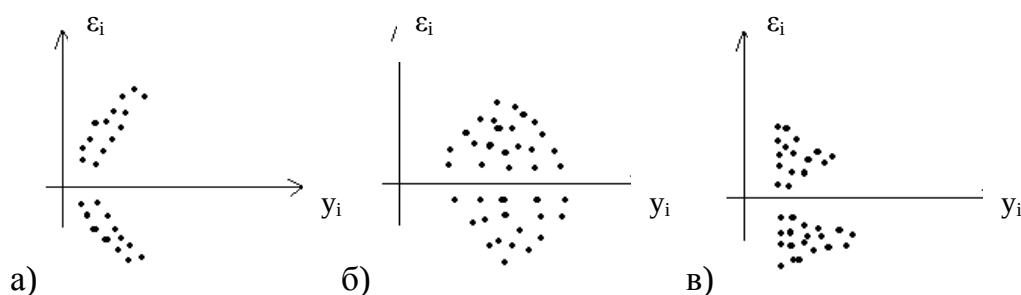


Рисунок 7 – Примеры гетероскедастичности остатков для зависимости  $\varepsilon_i$  от  $y_i$

Для оценки гетероскедастичности может использоваться метод Гольдфельда-Квандта. В 1965 г. Гольдфельд и Квандт рассмотрели однофакторную линейную модель, для которой дисперсия остатков возрастает пропорционально квадрату фактора. И они предложили параметрический тест, который включает следующие шаги:

- упорядочивание  $n$  наблюдений по мере возрастания переменной  $x_i$ , то есть  $x_{i+1} > x_i$ ;
- исключение из рассмотрения  $C$  центральных наблюдений. Число  $C$  находится из выражения:

$$\frac{n - C}{2} > p,$$

где  $p$  – число оцениваемых параметров.

Например, если  $n = 30$ , то  $C = 8$ ; если  $n = 60$ , то  $C = 16$ ; если  $n=20$ , то  $C=4$  и т.д.;

- разделение совокупности из оставшихся  $(n - C)$  наблюдений на две группы (соответственно с малыми и большими значениями фактора  $x_i$ ).

Определение по каждой из групп уравнения регрессии;

- определение остаточной суммы квадратов отклонений для первой ( $S_1$ ) и второй ( $S_2$ ) групп и нахождение их отношения  $R$ :

$$R = \frac{S_1}{S_2}.$$

Если  $R > F_{\text{табл.}}$  при  $df = \frac{n - c - 2p}{2}$ , то предпосылка о равенстве дисперсий остаточных величин нарушена.

### 1.3.5 Четвертая предпосылка – отсутствие автокорреляции в остатках

Автокорреляция в остатках означает наличие корреляции между остатками текущих и предыдущих (или последующих) наблюдений.

Коэффициент автокорреляции рассчитывается по формуле линейного коэффициента корреляции:

$$r_{\varepsilon_i \varepsilon_{i\pm 1}} = \frac{\overline{\varepsilon_i \varepsilon_{i\pm 1}} - \bar{\varepsilon}_i \bar{\varepsilon}_{i\pm 1}}{\sigma_{\varepsilon_i} \sigma_{\varepsilon_{i\pm 1}}},$$

где  $\bar{\varepsilon}_i$  – значение остатка текущего наблюдения;

$\bar{\varepsilon}_{i\pm 1}$  – значение остатка предыдущего или последующего наблюдения;

$\sigma_{\varepsilon_i}$  – среднее квадратическое отклонение текущего наблюдения, которое определяется по формуле:

$$\sigma_{\varepsilon_i} = \sqrt{\varepsilon_i^2 - (\bar{\varepsilon}_i)^2};$$

$\sigma_{\varepsilon_{i\pm 1}}$  – среднее квадратическое отклонение предыдущего или последующего наблюдения, которое определяется по формуле:

$$\sigma_{\varepsilon_{i\pm 1}} = \sqrt{\varepsilon_{i\pm 1}^2 - (\bar{\varepsilon}_{i\pm 1})^2}.$$

Если коэффициент автокорреляции окажется существенно отличным от нуля, то остатки автокоррелированы, т.е. связаны между собой.

Отсутствие автокорреляции остаточных величин обеспечивает состоятельность и эффективность оценок коэффициентов регрессии.

### 1.3.6 Пятая предпосылка - остатки подчиняются нормальному распределению

Предпосылка о нормальном распределении остатков позволяет проводить проверку параметров регрессии и корреляции с помощью критериев Стьюдента и Фишера. Однако бывают случаи, когда оценки регрессии, найденные с применением МНК, обладают хорошими свойствами даже при отсутствии нормального распределения остатков, то есть при нарушении пятой предпосылки.

Случайная величина  $\varepsilon_i$  имеет нормальный закон распределения (закон Гаусса), если её плотность вероятности имеет вид, представленный на рисунке 8.

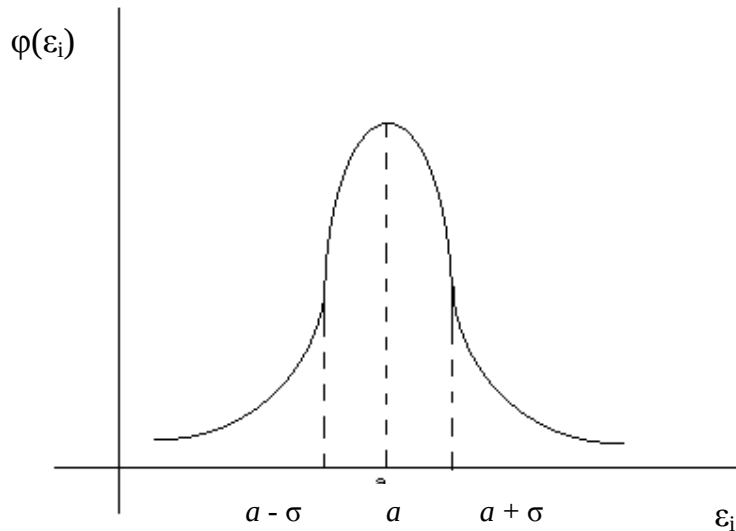


Рисунок 8 - График нормального распределения случайной величины  $\varepsilon_i$

Нормальный закон распределения случайной величины описывается выражением [6]:

$$\varphi(\varepsilon_i) = \frac{1}{\sigma_\varepsilon \sqrt{2\pi}} e^{-\frac{(\varepsilon_i - a)^2}{2\sigma_\varepsilon^2}}, \quad (12)$$

где  $a = M(\varepsilon_i)$  – математическое ожидание;  $\sigma_\varepsilon^2$  – дисперсия остатков.

Согласно формуле (12) при  $\varepsilon \rightarrow a$ :

$$\varphi(\varepsilon_i) \rightarrow \frac{1}{\sigma \sqrt{2\pi}}.$$

При нормальном распределении действует *правило трех сигм*: если случайная величина имеет нормальный закон распределения, то практически достоверно, что её значения заключены в интервале  $(a - 3\sigma; a + 3\sigma)$ .

При несоблюдении основных предпосылок МНК приходится корректировать модель, изменяя ее спецификацию, добавлять (исключать) некоторые факторы, преобразовывать исходные данные для того, чтобы получить оценки коэффициентов регрессии, которые обладают свойством несмещенности, имеют меньшее значение дисперсии остатков и обеспечивают, в связи с этим, более эффективную статистическую проверку значимости параметров регрессии.

## 1.4 Контрольные вопросы к теоретическому разделу

- 1 Назовите, в чем состоит спецификация модели множественной регрессии.
- 2 Сформулируйте требования, предъявляемые к факторам для включения их в модель множественной регрессии.
- 3 К каким трудностям приводит мультиколлинеанность факторов, включенных в модель, и как они могут быть разрешены?
- 4 Назовите методы устранения мультиколлинеарности факторов.
- 5 Что означает взаимодействие факторов и как оно может быть представлено графически?
- 6 Как интерпретируются коэффициенты регрессии линейной модели потребления?
- 7 Какой смысл приобретает  $\sum b_j$  в производственных функциях и что означает  $\sum b_j > 1$ ?
- 8 Какие коэффициенты используются для оценки сравнительной силы воздействия факторов на результат?
- 9 В каких случаях рассчитывается  $\overline{R^2}$ ?
- 10 От чего зависит величина скорректированного коэффициента множественной корреляции?
- 11 Каково назначение частной корреляции при построении модели множественной регрессии?
- 12 Составьте матрицу частных коэффициентов корреляции разного порядка для регрессионной модели с четырьмя факторами.
- 13 Что такое частный  $F$ -критерий и чем он отличается от общего  $F$ -критерия?
- 14 Как связаны между собой  $t$ -критерий Стьюдента для оценки значимости  $b_j$  и частные  $F$ -критерии?
- 15 При каких условиях строится уравнение множественной регрессии с фиктивными переменными?
- 16 Как трактуются коэффициенты модели, построенной только на фиктивных переменных?
- 17 Сформулируйте основные предпосылки применения МНК для построения регрессионной модели.
- 18 В чем сущность анализа остатков при наличии регрессионной модели?
- 19 Как можно проверить наличие гомо- и гетероскедастичности остатков?
- 20 Как оценивается отсутствие автокорреляции остатков при построении статистической регрессионной модели?

## 2 Практическая часть

### 1.1 Лабораторная работа 1. Расчет параметров и оценка уравнения множественной регрессии

**Цель работы:** освоить методику расчета коэффициентов и оценки уравнения множественной регрессии.

**Дано:** по 30 территориям России имеются данные, представленные в таблице 2.

Таблица 2 – Данные для лабораторной работы

| Признак                                                        | Среднее значение | Среднее квадратическое отклонение | Линейный коэффициент парной корреляции        |
|----------------------------------------------------------------|------------------|-----------------------------------|-----------------------------------------------|
| Среднедушевой доход, руб., $y$                                 | 86,8             | 11,44                             | -                                             |
| Среднедневная заработная плата одного работающего, руб., $x_1$ | 54,9             | 5,86                              | $r_{yx1} = 0,8405$                            |
| Средний возраст безработного, лет, $x_2$                       | 33,5             | 0,58                              | $r_{yx2} = - 0,2101$<br>$r_{x1x2} = - 0,1160$ |

#### Задание

1 Построить уравнение множественной регрессии в стандартизованной и естественной формах; рассчитать частные коэффициенты эластичности, сравнить их с  $\beta_1$  и  $\beta_2$ , пояснить различия между ними.

2 Рассчитать линейные коэффициенты частной корреляции и коэффициент множественной корреляции, сравнить их с линейными коэффициентами парной корреляции, пояснить различия между ними.

3 Рассчитать общий и частный  $F$ -критерии Фишера.

#### Решение

1 Линейное уравнение множественной регрессии имеет вид:  $y_i = a + b_1 \cdot x_{1i} + b_2 \cdot x_{2i}$ . Для расчета его параметров надо применить метод стандартизованных переменных и построить искомое уравнение в стандартизованном масштабе:  $t_y = \beta_1 \cdot t_{x1} + \beta_2 \cdot t_{x2}$ .

Расчет  $\beta$ -коэффициентов выполнить по формулам:

$$\beta_1 = \frac{r_{yx1} - r_{yx2} r_{x1x2}}{1 - r_{x1x2}^2} = \frac{0,8405 - 0,2101 \cdot 0,116}{1 - 0,116^2} = \frac{0,8161}{0,9865} = 0,8273;$$

$$\beta_2 = \frac{r_{yx2} - r_{yx1} r_{x1x2}}{1 - r_{x1x2}^2} = \frac{-0,2101 + 0,8405 \cdot 0,116}{1 - 0,116^2} = \frac{-0,1126}{0,9865} = -0,1141.$$

Получится уравнение:

$$t_y = 0,8273t_{x12} - 0,1141t_{x2}.$$

Для построения уравнения в естественной форме рассчитать  $b_1$  и  $b_2$ , используя формулы для перехода от  $\beta_j$  к  $b_j$ :

$$\beta_j = b_j \frac{\sigma_{xj}}{\sigma_y}; \quad b_j = \beta_j \frac{\sigma_y}{\sigma_{xj}};$$

$$b_1 = \frac{0,8273 \cdot 11,44}{5,86} = 1,6151; \quad b_2 = \frac{-0,1141 \cdot 11,44}{0,58} = -2,2505.$$

Значение  $a$  определить из соотношения:

$$a = \bar{y} - b_1 \cdot \bar{x}_1 - b_2 \cdot \bar{x}_2 = 86,8 - 1,6151 \cdot 54,9 + 2,2505 \cdot 33,5 = -73,52276.$$

Уравнение множественной регрессии в естественной форме будет следующим:

$$y_i = -73,52 + 1,62 \cdot x_{1i} - 2,25 \cdot x_{2i}.$$

Для характеристики относительной силы влияния  $x_1$  и  $x_2$  на  $y$  рассчитать средние коэффициенты эластичности по формуле:

$$\bar{\mathcal{E}}_{yxj} = b_j \frac{\bar{x}_j}{\bar{y}}.$$

$$\bar{\mathcal{E}}_{yx1} = \frac{1,62 \cdot 54,9}{86,8} = 1,0246\%; \quad \bar{\mathcal{E}}_{yx2} = \frac{-2,25 \cdot 33,5}{86,8} = -0,8684\%.$$

По полученным значениям частных коэффициентов эластичности можно сделать вывод, что с увеличением заработной платы  $x_1$  на 1 % от ее среднего уровня средний душевой доход  $y$  возрастает на 1,02 % от своего среднего уровня; при повышении среднего возраста безработного  $x_2$  на 1 % среднедушевой доход снижается на 0,87 % от своего среднего уровня.

Очевидно, что сила влияния средней заработной платы  $x_1$  на средний душевой доход  $y$  оказалась большей, чем сила влияния среднего возраста безработного  $x_2$ . К аналогичным выводам о силе связи приходим при сравнении модулей значений  $\beta_1$  и  $\beta_2$ :

$$|\beta_1| = |0,8273| > |\beta_2| = |-0,1141|.$$

Различия в силе влияния фактора на результат, полученные при сравнении  $\bar{\mathcal{E}}_{yxj}$  и  $\beta_j$ , объясняются тем, что коэффициент эластичности ис-

ходит из соотношения средних:  $\bar{\mathcal{E}}_{yxj} = b_j \frac{\bar{x}_j}{\bar{y}}$ , а  $\beta$ -коэффициент – из соотно-

шения средних квадратических отклонений:  $\beta_j = b_j \frac{\sigma_{xj}}{\sigma_y}$ .

2 Линейные коэффициенты частной корреляции для данных исходных данных рассчитываются по рекуррентным формулам:

$$r_{yx1 \cdot x2} = \frac{r_{yx1} - r_{yx2} \cdot r_{x1x2}}{\sqrt{(1 - r_{yx2}^2)(1 - r_{x1x2}^2)}} = \frac{0,8405 - 0,2101 \cdot 0,116}{\sqrt{(1 - 0,2101^2)(1 - 0,116^2)}} = 0,8404;$$

$$r_{yx2 \cdot x1} = \frac{r_{yx2} - r_{yx1} \cdot r_{x1x2}}{\sqrt{(1 - r_{yx1}^2)(1 - r_{x1x2}^2)}} = \frac{-0,2101 + 0,8405 \cdot 0,116}{\sqrt{(1 - 0,8405^2)(1 - 0,116^2)}} = -0,2092;$$

$$r_{x1x2 \cdot y} = \frac{r_{x1x2} - r_{yx1} \cdot r_{yx2}}{\sqrt{(1 - r_{yx1}^2)(1 - r_{yx2}^2)}} = \frac{-0,116 + 0,8405 \cdot 0,2101}{\sqrt{(1 - 0,8405^2)(1 - 0,2101^2)}} = 0,1144.$$

Если сравнить значения коэффициентов парной и частной корреляции, то приходим к выводу, что из-за слабой межфакторной связи ( $r_{x1x2} = -0,116$ ) коэффициенты парной и частной корреляции отличаются незначительно: выводы о тесноте и направлении связи на основе коэффициентов парной и частной корреляции совпадают:

$$r_{yx1} = 0,8405; \quad r_{yx2} = -0,2101; \quad r_{x1x2} = -0,1160;$$

$$r_{yx1 \cdot x2} = 0,8404; \quad r_{yx2 \cdot x1} = -0,2092; \quad r_{x1x2 \cdot y} = 0,1144.$$

Расчет линейного коэффициента множественной корреляции выполнить с использованием коэффициентов  $r_{yxj}$  и  $\beta_j$ :

$$R_{yx1x2} = \sqrt{r_{yx1} \cdot \beta_1 + r_{yx2} \cdot \beta_2} = \sqrt{0,8405 \cdot 0,8273 + 0,2101 \cdot 0,1141} = \sqrt{0,7193} = 0,8481.$$

Зависимость  $y$  от  $x_1$  и  $x_2$  характеризуется как тесная, в которой 72% вариации среднего душевого дохода определяются вариацией учтенных в модели факторов: средней заработной платы и среднего возраста безработного. Прочие факторы, не включенные в модель, составляют соответственно 28 % от общей вариации  $y$ .

3 Общий  $F$ -критерий Фишера проверяет гипотезу  $H_0$  о статистической значимости уравнения регрессии и показателя тесноты связи ( $R^2 = 0$ ):

$$F_{\text{факт}} = \frac{R_{yx1x2}^2}{1 - R_{yx1x2}^2} \cdot \frac{n - m - 1}{m} = \frac{0,7193}{0,2807} \cdot \frac{27}{2} = 34,6;$$

$$F_{\text{табл}} = 3,4; \quad \alpha = 0,05.$$

Сравнивая  $F_{\text{табл}}$  и  $F_{\text{факт}}$ , приходим к выводу о необходимости отклонить гипотезу  $H_0$ , так как  $F_{\text{табл}} = 3,4 < F_{\text{факт}} = 34,6$ . С вероятностью  $1 - \alpha = 0,95$  делаем заключение о статистической значимости уравнения в целом и показателя тесноты связи  $R_{yx1x2}$ , которые сформировались под неслучайным воздействием факторов  $x_1$  и  $x_2$ .

Частные  $F$ -критерии –  $F_{x1}$  и  $F_{x2}$  оценивают статистическую значимость присутствия факторов  $x_1$  и  $x_2$  в уравнении множественной регрессии, оценивают целесообразность включения в уравнение одного фактора после другого фактора, то есть  $F_{x1}$  оценивает целесообразность включения в уравнение фактора  $x_1$  после того, как в него был включен фактор  $x_2$ . Соот-

ветственно  $F_{x_2}$  указывает на целесообразность включения в модель фактора  $x_2$  после фактора  $x_1$ :

$$F_{x_1} = \frac{R_{yx1x_2}^2 - r_{yx2}^2}{1 - R_{yx1x_2}^2} \cdot \frac{n - m - 1}{1} = \frac{0,8481^2 - 0,2101^2}{1 - 0,8481^2} \cdot \frac{30 - 2 - 1}{1} = 64,9.$$

$$F_{табл} = 4,21; \alpha = 0,05.$$

Сравнивая  $F_{табл}$  и  $F_{факт}$ , приходим к выводу о целесообразности включения в модель фактора  $x_1$  после фактора  $x_2$ , так как  $F_{x_1} = 64,9 > F_{табл}$ . Гипотезу  $H_0$  о несущественности прироста  $R_y^2$  за счет включения дополнительного фактора  $x_1$  отклоняем и приходим к выводу о статистически подтвержденной целесообразности включения фактора  $x_1$  после фактора  $x_2$ .

Целесообразность включения в модель фактора  $x_2$  после фактора  $x_1$  проверяет  $F_{x_2}$ :

$$F_{x_2} = \frac{R_{yx1x_2}^2 - r_{yx1}^2}{1 - R_{yx1x_2}^2} \cdot \frac{n - m - 1}{1} = \frac{0,8481^2 - 0,8405^2}{1 - 0,8481^2} \cdot \frac{30 - 2 - 1}{1} = 1,234.$$

Низкое значение  $F_{x_2}$  (немного больше 1) свидетельствует о статистической незначимости прироста  $r_{yx1}^2$  за счет включения в модель фактора  $x_2$  после фактора  $x_1$ . Следовательно, подтверждается нулевая гипотеза  $H_0$  о нецелесообразности включения в модель фактора  $x_2$  (средний возраст безработного). Это означает, что парная регрессионная модель зависимости среднего дохода от средней заработной платы является достаточно статистически значимой, надежной и что нет необходимости улучшать ее, включая дополнительный фактор  $x_2$  (средний возраст безработного).

## Задачи для самостоятельной работы

### Задача 1

По совокупности 30 предприятий концерна изучается зависимость прибыли  $y$  (тыс. руб.) от выработки продукции на одного работника  $x_1$  (ед.) и индекса цен на продукцию  $x_2$  (%). Данные приведены в таблице 3.

Таблица 3 - Данные для задачи 1

| Признак | Среднее значение | Среднее квадратическое отклонение | Линейный коэффициент парной корреляции |
|---------|------------------|-----------------------------------|----------------------------------------|
| $y$     | 250              | 38                                | -                                      |
| $x_1$   | 47               | 12                                | $r_{yx1} = 0,68$                       |
| $x_2$   | 112              | 21                                | $r_{yx2} = 0,63$<br>$r_{x1x2} = 0,42$  |

### Задание

1 Построить линейные уравнения парной регрессии, оценить их значимость с помощью  $F$ -критерия Фишера.

2 Найти уравнение множественной регрессии в стандартизованном и натуральном масштабах.

3 Рассчитать множественный коэффициент корреляции, общий и частные критерии Фишера и сделать выводы.

### Задача 2

По 30 заводам, выпускающим продукцию А, изучается зависимость потребления электроэнергии  $y$  (тыс. кВт·ч) от производства продукции –  $x_1$  (тыс.ед.) и уровня механизации труда –  $x_2$  (%). Данные приведены в таблице 4.

Таблица 4 - Данные для задачи 2

| Признак | Среднее значение | Среднее квадратическое отклонение | Линейный коэффициент парной корреляции |
|---------|------------------|-----------------------------------|----------------------------------------|
| $y$     | 1000             | 27                                | -                                      |
| $x_1$   | 420              | 45                                | $r_{yx1} = 0,77$                       |
| $x_2$   | 41,5             | 18                                | $r_{yx2} = 0,43$<br>$r_{x1x2} = 0,38$  |

### Задание

1 Построить уравнение множественной регрессии в стандартизованном и натуральном масштабах.

2 Определить показатели частной и множественной корреляции.

3 Найти частные коэффициенты эластичности и сравнить их с  $\beta$ -коэффициентами.

4 Рассчитать общий и частные  $F$ -критерии Фишера.

### Задача 3

Изучается зависимость по 25 предприятиям концерна потребления материалов  $y$  (т) от энерговооруженности труда  $x_1$  (кВт·ч на одного рабочего) и объема производственной продукции  $x_2$  (тыс.ед.). Данные приведены в таблице 5.

Таблица 5 – Данные для задачи 3

| Признак | Среднее значение | Среднее квадратическое отклонение | Линейный коэффициент парной корреляции |
|---------|------------------|-----------------------------------|----------------------------------------|
| $y$     | 12               | 2                                 | -                                      |
| $x_1$   | 4,3              | 0,5                               | $r_{yx1} = 0,52$                       |
| $x_2$   | 10               | 1,8                               | $r_{yx2} = 0,84$<br>$r_{x1x2} = 0,43$  |

**Задание**

- 1 Построить уравнение множественной регрессии и пояснить экономический смысл его параметров
- 2 Определить частные коэффициенты эластичности и стандартизованные коэффициенты регрессии.
- 3 Найти частные и множественный коэффициенты корреляции.
- 4 оценить значимость уравнения регрессии с помощью критерия Фишера.

## 2.2 Лабораторная работа 2. Расчет парных коэффициентов корреляции и проверка наличия мультиколлинеарности факторов

**Цель работы:** освоить методику расчета парных коэффициентов корреляции и проверки наличия мультиколлинеарности факторов, а также исследования остаточных величин уравнения множественной регрессии.

**Дано:** по 20 предприятиям региона (см. таблицу 5) изучается зависимость выработки продукции на одного 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  $x_2$  (%). Данные приведены в таблице 6.

Таблица 6 – Данные для лабораторной работы №2

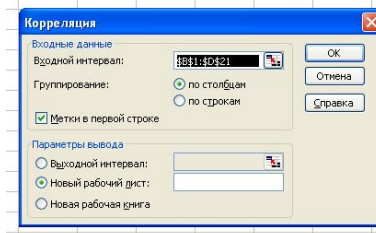
| Номер предприятия | $y$ | $x_1$ | $x_2$ | Номер предприятия | $y$ | $x_1$ | $x_2$ |
|-------------------|-----|-------|-------|-------------------|-----|-------|-------|
| 1                 | 7   | 3,9   | 10    | 11                | 9   | 6     | 21    |
| 2                 | 7   | 3,9   | 14    | 12                | 11  | 6,4   | 22    |
| 3                 | 7   | 3,7   | 15    | 13                | 9   | 6,8   | 22    |
| 4                 | 7   | 4     | 16    | 14                | 11  | 7,2   | 25    |
| 5                 | 7   | 3,8   | 17    | 15                | 12  | 8     | 28    |
| 6                 | 7   | 4,8   | 19    | 16                | 12  | 8,2   | 29    |
| 7                 | 8   | 5,4   | 19    | 17                | 12  | 8,1   | 30    |
| 8                 | 8   | 4,4   | 20    | 18                | 12  | 8,5   | 31    |
| 9                 | 8   | 5,3   | 20    | 19                | 14  | 9,6   | 32    |
| 10                | 10  | 6,8   | 20    | 20                | 14  | 9     | 36    |

## Задание

1 С помощью инструмента анализа данных «Корреляция» в MS Excel рассчитать парные коэффициенты корреляции. Для этого:

1) в главном меню последовательно выбрать пункты: **Сервис/Анализ данных/Корреляция**, кнопка **ОК**.

2) заполнить окно ввода данных и параметров вывода:



Результат будет следующим:

|                | y        | x <sub>1</sub> | x <sub>2</sub> |
|----------------|----------|----------------|----------------|
| y              | 1        |                |                |
| x <sub>1</sub> | 0,969881 | 1              |                |
| x <sub>2</sub> | 0,9408   | 0,942839       | 1              |

2 Осуществить проверку факторов на мультиколлинеарность с помощью определителя матрицы парных коэффициентов корреляции:

|                | x <sub>1</sub> | x <sub>2</sub> |
|----------------|----------------|----------------|
| x <sub>1</sub> | 1              | 0,9428         |
| x <sub>2</sub> | 0,9428         | 1              |

Определитель матрицы ( $\text{Det}|R| = 0,111$ ) указывает на наличие мультиколлинеарности факторов.

Проверка мультиколлинеарности факторов методом испытания гипотезы о независимости переменных  $H_0: \text{Det}|R| = 1$  осуществляется расчетом приближенного распределения  $\chi^2 = [n-1-1/6(2m+5)]\lg\text{Det}R$  с  $(1/2 n(N-1))$  степенями свободы. Если фактическое значение  $\chi^2_{\text{факт}} > \chi^2_{\text{табл}}$ , то гипотеза  $H_0$  отклоняется. Это означает, что  $\text{Det}|R| \neq 1$ , недиагональные ненулевые коэффициенты корреляции указывают на коллинеарность факторов.

3 С помощью инструмента анализа данных «Регрессия» (**Сервис/Анализ данных/Регрессия**) рассчитать параметры уравнения множественной регрессии:

**Регрессия**

Входные данные  
Входной интервал Y:   
Входной интервал X:   
☒ Метки  
☐ Константа - ноль  
☐ Уровень надежности:  %

Параметры вывода  
☐ Выходной интервал:  
☒ Новый рабочий лист:  
☐ Новая рабочая книга

Остатки  
☒ Остатки  
☐ Стандартизованные остатки  
☒ График остатков  
☐ График подбора

Нормальная вероятность  
☐ График нормальной вероятности

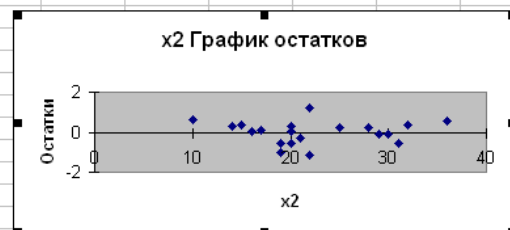
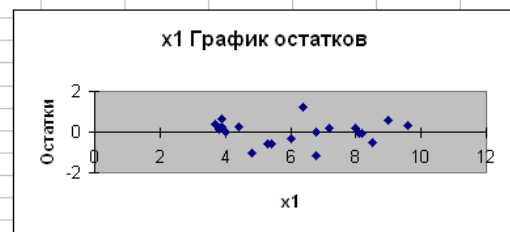
OK Отмена Справка

Результат будет следующим:

| A                        | B               | C        | D        | E        | F            | G        | H        | I        | J |
|--------------------------|-----------------|----------|----------|----------|--------------|----------|----------|----------|---|
| ВЫВОД ИТОГОВ             |                 |          |          |          |              |          |          |          |   |
| Регрессионная статистика |                 |          |          |          |              |          |          |          |   |
| Множественный R          | 0,973101182     |          |          |          |              |          |          |          |   |
| R-квадрат                | 0,94692591      |          |          |          |              |          |          |          |   |
| Нормированный R-квадрат  | 0,9406819       |          |          |          |              |          |          |          |   |
| Стандартная ошибка       | 0,598670364     |          |          |          |              |          |          |          |   |
| Наблюдения               | 20              |          |          |          |              |          |          |          |   |
| Дисперсионный анализ     |                 |          |          |          |              |          |          |          |   |
|                          | df              | SS       | MS       | F        | значимость F |          |          |          |   |
| Регрессия                | 2               | 108,7071 | 54,35355 | 151,6535 | 1,45E-11     |          |          |          |   |
| Остаток                  | 17              | 6,092905 | 0,358406 |          |              |          |          |          |   |
| Итого                    | 19              | 114,8    |          |          |              |          |          |          |   |
| Коэффициенты             |                 |          |          |          |              |          |          |          |   |
| Y-пересечение            | 1,83530694      | 0,471065 | 3,89608  | 0,001162 | 0,841447     | 2,829167 | 0,841447 | 2,829167 |   |
| x1                       | 0,945947723     | 0,212576 | 4,449917 | 0,000351 | 0,497451     | 1,394445 | 0,497451 | 1,394445 |   |
| x2                       | 0,085617787     | 0,060483 | 1,415561 | 0,174964 | -0,04199     | 0,213226 | -0,04199 | 0,213226 |   |
|                          | a=              | 1,84     |          |          |              |          |          |          |   |
|                          | b1=             | 0,95     |          |          |              |          |          |          |   |
|                          | b2=             | 0,09     |          |          |              |          |          |          |   |
| ВЫВОД ОСТАТКА            |                 |          |          |          |              |          |          |          |   |
| Наблюдение               | Предсказанное y | Остатки  |          |          |              |          |          |          |   |
| 1                        | 6,380680931     | 0,619319 |          |          |              |          |          |          |   |
| 2                        | 6,723152081     | 0,276848 |          |          |              |          |          |          |   |
| 3                        | 6,619580323     | 0,38042  |          |          |              |          |          |          |   |
| 4                        | 6,988982427     | 0,011018 |          |          |              |          |          |          |   |
| 5                        | 6,88541067      | 0,114589 |          |          |              |          |          |          |   |
| 6                        | 8,002593967     | -1,00259 |          |          |              |          |          |          |   |
| 7                        | 8,570162601     | -0,57016 |          |          |              |          |          |          |   |
| 8                        | 7,709832666     | 0,290167 |          |          |              |          |          |          |   |
| 9                        | 8,561185616     | -0,56119 |          |          |              |          |          |          |   |
| 10                       | 9,9801072       | 0,019893 |          |          |              |          |          |          |   |
| 11                       | 9,308966809     | -0,30897 |          |          |              |          |          |          |   |
| 12                       | 9,772963686     | 1,227036 |          |          |              |          |          |          |   |
| 13                       | 10,15134277     | -1,15134 |          |          |              |          |          |          |   |
| 14                       | 10,78657523     | 0,213425 |          |          |              |          |          |          |   |
| 15                       | 11,80018677     | 0,199813 |          |          |              |          |          |          |   |
| 16                       | 12,0749941      | -0,07499 |          |          |              |          |          |          |   |
| 17                       | 12,06601711     | -0,06602 |          |          |              |          |          |          |   |
| 18                       | 12,53001399     | -0,53001 |          |          |              |          |          |          |   |
| 19                       | 13,65617427     | 0,343826 |          |          |              |          |          |          |   |
| 20                       | 13,43107679     | 0,568923 |          |          |              |          |          |          |   |

x1 График остатков

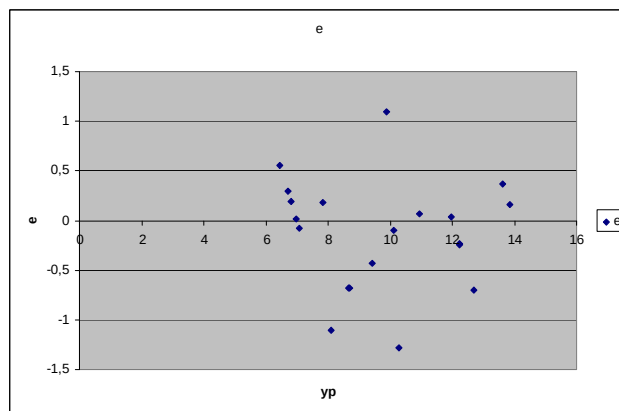
x2 График остатков



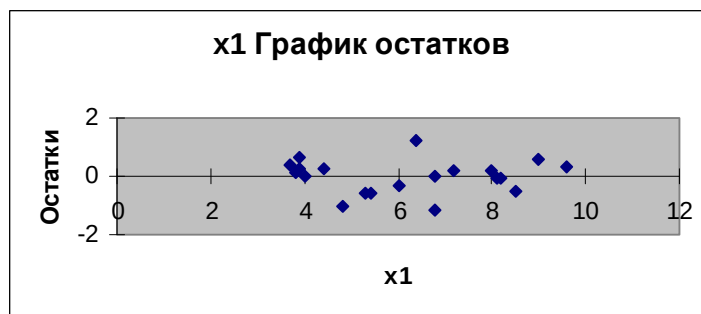
4 По полученным значениям параметров регрессии рассчитать теоретические значения результативного признака и значения остатков:

| Номер предприятия | y  | x <sub>1</sub> | x <sub>2</sub> | y <sub>p</sub> | e      | Номер предприятия | y  | x <sub>1</sub> | x <sub>2</sub> | y <sub>p</sub> | e     |
|-------------------|----|----------------|----------------|----------------|--------|-------------------|----|----------------|----------------|----------------|-------|
| 1                 | 7  | 3,9            | 10             | 6,445          | 0,555  | 11                | 9  | 6              | 21             | 9,43           | -0,43 |
| 2                 | 7  | 3,9            | 14             | 6,805          | 0,195  | 12                | 11 | 6,4            | 22             | 9,9            | 1,1   |
| 3                 | 7  | 3,7            | 15             | 6,705          | 0,295  | 13                | 9  | 6,8            | 22             | 10,28          | -1,28 |
| 4                 | 7  | 4              | 16             | 7,08           | -0,08  | 14                | 11 | 7,2            | 25             | 10,93          | 0,07  |
| 5                 | 7  | 3,8            | 17             | 6,98           | 0,02   | 15                | 12 | 8              | 28             | 11,96          | 0,04  |
| 6                 | 7  | 4,8            | 19             | 8,11           | -1,11  | 16                | 12 | 8,2            | 29             | 12,24          | -0,24 |
| 7                 | 8  | 5,4            | 19             | 8,68           | -0,68  | 17                | 12 | 8,1            | 30             | 12,235         | 0,235 |
| 8                 | 8  | 4,4            | 20             | 7,82           | 0,18   | 18                | 12 | 8,5            | 31             | 12,705         | 0,705 |
| 9                 | 8  | 5,3            | 20             | 8,675          | -0,675 | 19                | 14 | 9,6            | 32             | 13,84          | 0,16  |
| 10                | 10 | 6,8            | 20             | 10,1           | -0,1   | 20                | 14 | 9              | 36             | 13,63          | 0,37  |

5 Построить графики зависимости для проверки предпосылок МНК:  
1) проверка случайного характера остатков (e от y<sub>p</sub>):

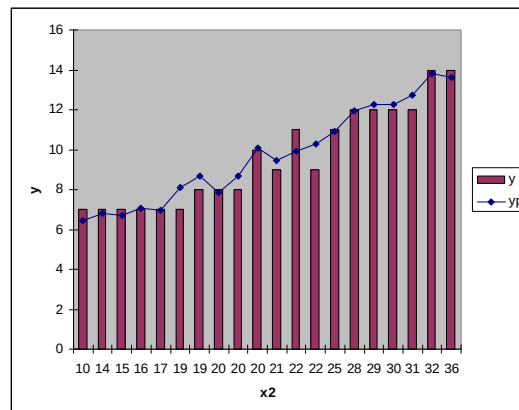
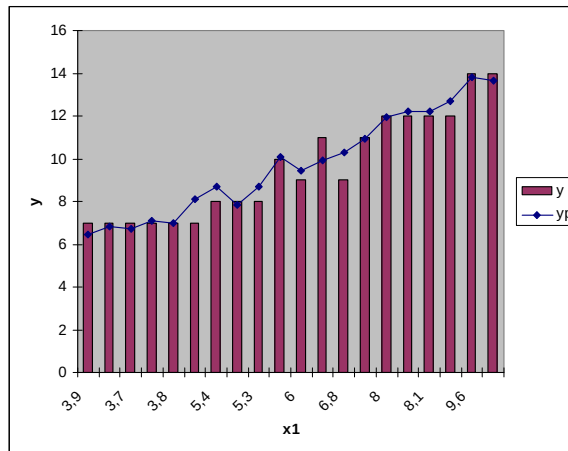


2) нулевая средняя величина остатков (e от x<sub>1</sub>):





3) проверка на гомоскедастичность (у от  $x_j$ ):



6 Осуществить проверку остатков на гомоскедастичность с помощью метода Гольдфелда-Квандта в следующем порядке:

- 1) упорядочение  $n$  наблюдений по мере возрастания переменной  $x$ ;
- 2) исключение из рассмотрения  $C$  центральных наблюдений;
- 3) разделение совокупности из  $(n - C)$  наблюдений на две группы и определение по каждой из групп уравнения регрессии;

| Номер предприятия | $y$ | $x_1$ | $x_2$ | $y_p$ | $e$ | $(y_p - y)^2$ |
|-------------------|-----|-------|-------|-------|-----|---------------|
| 1                 | 7   | 3,9   | 10    | 6,509 |     |               |
| 2                 | 7   | 3,9   | 14    | 6,669 |     |               |
| 3                 | 7   | 3,7   | 15    | 6,647 |     |               |
| 4                 | 7   | 4     | 16    | 6,78  |     |               |
| 5                 | 7   | 3,8   | 17    | 6,758 |     |               |
| 6                 | 7   | 4,8   | 19    | 7,148 |     |               |
| 7                 | 8   | 5,4   | 19    | 7,334 |     |               |
| 8                 | 8   | 4,4   | 20    | 7,064 |     |               |

| Номер предприятия | $y$ | $x_1$ | $x_2$ | $y_p$    | $e$ | $(y_p - y)^2$ |
|-------------------|-----|-------|-------|----------|-----|---------------|
| 13                | 9   | 6,8   | 22    | 9,486136 |     |               |
| 14                | 11  | 7,2   | 25    | 10,36709 |     |               |
| 15                | 12  | 8     | 28    | 11,64898 |     |               |
| 16                | 12  | 8,2   | 29    | 12,00946 |     |               |
| 17                | 12  | 8,1   | 30    | 12,06922 |     |               |
| 18                | 12  | 8,5   | 31    | 12,63017 |     |               |
| 19                | 14  | 9,6   | 32    | 13,89278 |     |               |
| 20                | 14  | 9     | 36    | 13,93136 |     |               |

4) определение остаточной суммы квадратов для первой ( $S_1$ ) и второй ( $S_2$ ) групп и нахождение их отношения:  $R = S_1:S_2$ ;

5) сравнение полученного значения  $R$  с табличным значением критерия Фишера ( $df = (n - C - 2p) : 2$ ). Чем больше величина  $R$  превышает  $F_{\text{табл.}}$ , тем более нарушена предпосылка о равенстве дисперсий остаточных величин.

### Задачи для самостоятельной работы

#### Задача 1

Имеются данные о деятельности крупнейших компаний США в 1996г., представленные в таблице 7.

Таблица 7 – Данные для задачи 1

| №  | Чистый доход,<br>у,<br>млрд долл.<br>США | Оборот капита-<br>ла, $x_1$ ,<br>млрд долл.<br>США | Использованный<br>капитал, $x_2$ ,<br>млрд долл. США | Численность<br>служащих, $x_3$ ,<br>тыс.чел. |
|----|------------------------------------------|----------------------------------------------------|------------------------------------------------------|----------------------------------------------|
| 1  | 6,6                                      | 6,9                                                | 83,6                                                 | 222                                          |
| 2  | 3                                        | 18                                                 | 6,5                                                  | 32                                           |
| 3  | 6,5                                      | 107,9                                              | 50,4                                                 | 82                                           |
| 4  | 3,3                                      | 16,7                                               | 15,4                                                 | 45,2                                         |
| 5  | 0,1                                      | 79,6                                               | 29,6                                                 | 299,3                                        |
| 6  | 3,6                                      | 16,2                                               | 13,3                                                 | 41,6                                         |
| 7  | 1,5                                      | 5,9                                                | 5,9                                                  | 17,8                                         |
| 8  | 5,5                                      | 53,1                                               | 27,1                                                 | 151                                          |
| 9  | 2,4                                      | 18,8                                               | 11,2                                                 | 82,3                                         |
| 10 | 3                                        | 35,3                                               | 16,4                                                 | 103                                          |
| 11 | 4,2                                      | 71,9                                               | 32,5                                                 | 225,4                                        |
| 12 | 2,7                                      | 93,6                                               | 25,4                                                 | 675                                          |
| 13 | 1,6                                      | 10                                                 | 6,4                                                  | 43,8                                         |
| 14 | 2,4                                      | 31,5                                               | 12,5                                                 | 102,3                                        |
| 15 | 3,3                                      | 36,7                                               | 14,3                                                 | 105                                          |
| 16 | 1,8                                      | 13,8                                               | 6,5                                                  | 49,1                                         |
| 17 | 2,4                                      | 64,8                                               | 22,7                                                 | 50,4                                         |
| 18 | 1,6                                      | 30,4                                               | 15,8                                                 | 480                                          |
| 19 | 1,4                                      | 12,1                                               | 9,3                                                  | 71                                           |
| 20 | 0,9                                      | 31,383,6                                           | 18,9                                                 | 43                                           |

### Задание

По имеющимся данным рассчитать парные коэффициенты корреляции и проверить наличие мультиколлинеарности факторов, построить уравнение множественной регрессии и исследовать остаточные величины уравнения множественной регрессии.

### Задача 2

Имеются данные по 36 странам за 1997г., представленные в таблице 8.

Таблица 8 – Данные для задачи 2

| Страна           | Индекс человеческого развития,<br>у | Ожидаемая продолжительность жизни при рождении в 1997г.,<br>$x_1$ , лет | Суточная калорийность питания населения,<br>$x_2$ , ккал на душу |
|------------------|-------------------------------------|-------------------------------------------------------------------------|------------------------------------------------------------------|
| Австрия          | 0,904                               | 77,0                                                                    | 3343                                                             |
| Австралия        | 0,922                               | 78,2                                                                    | 3001                                                             |
| Аргентина        | 0,827                               | 72,9                                                                    | 3136                                                             |
| Белоруссия       | 0,763                               | 68,0                                                                    | 3101                                                             |
| Бельгия          | 0,923                               | 77,2                                                                    | 3543                                                             |
| Бразилия         | 0,739                               | 66,8                                                                    | 2938                                                             |
| Великобритания   | 0,918                               | 77,2                                                                    | 3237                                                             |
| Венгрия          | 0,795                               | 70,9                                                                    | 3402                                                             |
| Германия         | 0,906                               | 77,2                                                                    | 3330                                                             |
| Греция           | 0,867                               | 78,1                                                                    | 3575                                                             |
| Дания            | 0,905                               | 75,7                                                                    | 3808                                                             |
| Египет           | 0,616                               | 66,3                                                                    | 3289                                                             |
| Израиль          | 0,883                               | 77,8                                                                    | 3272                                                             |
| Индия            | 0,545                               | 62,6                                                                    | 2415                                                             |
| Испания          | 0,894                               | 78,0                                                                    | 3295                                                             |
| Италия           | 0,900                               | 78,2                                                                    | 3504                                                             |
| Канада           | 0,932                               | 79,0                                                                    | 3056                                                             |
| Казахстан        | 0,740                               | 67,7                                                                    | 3007                                                             |
| Китай            | 0,701                               | 69,8                                                                    | 2844                                                             |
| Латвия           | 0,744                               | 68,4                                                                    | 2861                                                             |
| Нидерланды       | 0,921                               | 77,9                                                                    | 3259                                                             |
| Норвегия         | 0,927                               | 78,1                                                                    | 3350                                                             |
| Польша           | 0,802                               | 72,5                                                                    | 3344                                                             |
| Республика Корея | 0,852                               | 72,4                                                                    | 3336                                                             |
| Россия           | 0,747                               | 66,6                                                                    | 2704                                                             |
| Румыния          | 0,752                               | 69,9                                                                    | 2043                                                             |
| США              | 0,927                               | 76,6                                                                    | 3642                                                             |
| Турция           | 0,728                               | 69,0                                                                    | 3568                                                             |
| Украина          | 0,721                               | 68,8                                                                    | 2753                                                             |
| Финляндия        | 0,913                               | 76,8                                                                    | 2916                                                             |
| Франция          | 0,918                               | 78,1                                                                    | 3551                                                             |
| Чехия            | 0,833                               | 73,9                                                                    | 3177                                                             |
| Швейцария        | 0,914                               | 78,6                                                                    | 3280                                                             |
| Швеция           | 0,923                               | 78,5                                                                    | 3160                                                             |
| ЮАР              | 0,695                               | 64,1                                                                    | 2933                                                             |
| Япония           | 0,924                               | 80                                                                      | 2905                                                             |

### Задание

- 1 Построить матрицу парных коэффициентов корреляции.
- 2 Построить парные уравнения регрессии.
- 3 Оценить статистическую значимость уравнения и их параметров с помощью критериев Фишера и Стьюдента.
- 4 Построить уравнение множественной регрессии.
- 5 Построить графики остатков. Сделать выводы.
- 6 Провести тестирование ошибок уравнения множественной регрессии на гетероскедастичность, применив тест Гольдфельда-Квандта.
- 7 Оценить статистическую значимость уравнения множественной регрессии. Определить, какое уравнение лучше использовать для прогноза:
  - а) парную регрессию  $y$  на  $x_1$ ;
  - б) парную регрессию  $y$  на  $x_2$ ;
  - с) множественную регрессию.

### Задача 3

Имеются данные по 25 странам, представленные в таблице 9:

- $y$  – индекс человеческого развития;  
 $x_1$  – ВВП 1997 г., % к 1990 г.;  
 $x_2$  – расходы на конечное потребление в текущих ценах, % к ВВП;  
 $x_3$  – расходы домашних хозяйств, % к ВВП;  
 $x_4$  – валовое накопление, % к ВВП;  
 $x_5$  – суточная калорийность питания населения, ккал на душу населения;  
 $x_6$  – ожидаемая продолжительность жизни при рождении 1997 г., число лет.

Таблица 9 – Данные для задачи 3

| Страна         | $y$   | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ |
|----------------|-------|-------|-------|-------|-------|-------|-------|
| Австрия        | 0,904 | 115   | 75,5  | 56,1  | 25,2  | 3343  | 77,0  |
| Австралия      | 0,922 | 123   | 78,5  | 61,8  | 21,8  | 3001  | 78,2  |
| Белоруссия     | 0,763 | 74    | 78,4  | 59,1  | 25,7  | 3101  | 68,0  |
| Бельгия        | 0,923 | 111   | 77,7  | 63,3  | 17,8  | 3543  | 77,2  |
| Великобритания | 0,918 | 113   | 84,4  | 64,1  | 15,9  | 3237  | 77,2  |
| Германия       | 0,906 | 110   | 75,9  | 57,0  | 22,4  | 3330  | 77,2  |
| Дания          | 0,905 | 119   | 76,0  | 50,7  | 20,6  | 3808  | 75,7  |
| Индия          | 0,545 | 146   | 67,5  | 57,1  | 25,2  | 2415  | 62,6  |
| Испания        | 0,894 | 113   | 78,2  | 62,0  | 20,7  | 3295  | 78,0  |
| Италия         | 0,900 | 108   | 78,1  | 61,8  | 17,5  | 3504  | 78,2  |
| Канада         | 0,932 | 113   | 78,6  | 58,6  | 19,7  | 3056  | 79,0  |

Окончание таблицы 9

| Страна     | y     | x <sub>1</sub> | x <sub>2</sub> | x <sub>3</sub> | x <sub>4</sub> | x <sub>5</sub> | x <sub>6</sub> |
|------------|-------|----------------|----------------|----------------|----------------|----------------|----------------|
| Казахстан  | 0,740 | 71             | 84,0           | 71,7           | 18,5           | 3007           | 67,6           |
| Китай      | 0,701 | 210            | 59,2           | 48,0           | 42,4           | 2844           | 69,8           |
| Латвия     | 0,744 | 94             | 90,2           | 63,9           | 23,0           | 2861           | 68,4           |
| Нидерланды | 0,921 | 118            | 72,8           | 59,1           | 20,2           | 3259           | 77,9           |
| Норвегия   | 0,927 | 130            | 67,7           | 47,5           | 25,2           | 3350           | 78,1           |
| Польша     | 0,802 | 127            | 82,6           | 65,3           | 22,4           | 3344           | 72,5           |
| Россия     | 0,747 | 61             | 74,4           | 53,2           | 22,7           | 2704           | 66,6           |
| США        | 0,927 | 117            | 83,3           | 67,9           | 18,1           | 3642           | 76,7           |
| Украина    | 0,721 | 46             | 83,7           | 61,7           | 20,1           | 2753           | 68,8           |
| Финляндия  | 0,913 | 107            | 73,8           | 52,9           | 17,3           | 2916           | 76,8           |
| Франция    | 0,918 | 110            | 79,2           | 59,9           | 16,8           | 3551           | 78,1           |
| Чехия      | 0,833 | 99,2           | 71,5           | 51,5           | 29,9           | 3177           | 73,9           |
| Швейцария  | 0,914 | 101            | 75,3           | 61,2           | 20,3           | 3280           | 78,6           |
| Швеция     | 0,923 | 105            | 79,0           | 53,1           | 14,1           | 3160           | 78,5           |

### Задание

1 Построить матрицу парных коэффициентов корреляции. Рассчитать коэффициенты множественной детерминации, используя в качестве зависимой переменной каждый фактор. Установить, какие факторы мультиколлинеарны.

2 Построить уравнение множественной регрессии в линейной форме с полным набором факторов.

3 Оценить статистическую значимость уравнения регрессии и его параметров с помощью критериев Фишера и Стьюдента.

4 Отобрать информативные факторы по п.1 и 3. Построить уравнение регрессии со статистически значимыми факторами.

## Литература

- 1 Эконометрика: учебник / под ред. И.И. Елисеевой. – М.: Финансы и статистика, 2004. – 344 с.: ил.
- 2 Практикум по эконометрике: учебное пособие / И.И. Елисеева [и др.]; под ред. И.И. Елисеевой. – М.: Финансы и статистика, 2004. – 192 с.: ил.
- 3 Новак Эдвард. Введение в методы эконометрики. Сборник задач: пер. с польск./ под ред. И.И. Елисеевой. – М.: Финансы и статистика, 2004. – 248 с.: ил.
- 4 Бабешко Л.О. Основы эконометрического моделирования: учебное пособие / Л.О. Бабешко. – М.: КомКнига, 2006. – 432 с.
- 5 Доугерти К. Введение в эконометрику: пер. с англ./к. – М.: ИНФРА-М, 1997. – XIV, 402 с.
- 6 Кремер Н.Ш. Эконометрика: учебник для вузов / под ред. проф. Н.Ш. Кремера. – М.: ЮНИТИ-ДАНА, 2003. – 311 с.
- 7 Магнус Я.Р. Эконометрика: начальный курс / Я.Р. Магнус, П.К. Катышев, А.А. Пересецкий. – М.: Дело, 1997. – 248 с.

# **Приложение А** **(справочное)**

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

| $k1 \backslash k2$ | 1      | 2      | 3      | 4      | 5      | 6      | 8      | 12     | 24     | $\infty$ |
|--------------------|--------|--------|--------|--------|--------|--------|--------|--------|--------|----------|
| 1                  | 161,45 | 199,50 | 215,72 | 224,57 | 230,17 | 233,97 | 238,89 | 243,91 | 249,04 | 254,32   |
| 2                  | 18,51  | 19,00  | 19,16  | 19,25  | 19,30  | 19,33  | 19,37  | 19,41  | 19,45  | 19,50    |
| 3                  | 10,13  | 9,55   | 9,28   | 9,12   | 9,01   | 8,94   | 8,84   | 8,74   | 8,64   | 8,53     |
| 4                  | 7,71   | 6,94   | 6,59   | 6,39   | 6,26   | 6,16   | 6,04   | 5,91   | 5,77   | 5,63     |
| 5                  | 6,61   | 5,79   | 5,41   | 5,19   | 5,05   | 4,95   | 4,82   | 4,68   | 4,53   | 4,36     |
| 6                  | 5,99   | 5,14   | 4,76   | 4,53   | 4,39   | 4,28   | 4,15   | 4,00   | 3,84   | 3,67     |
| 7                  | 5,59   | 4,74   | 4,35   | 4,12   | 3,97   | 3,87   | 3,73   | 3,57   | 3,41   | 3,23     |
| 8                  | 5,32   | 4,46   | 4,07   | 3,84   | 3,69   | 3,58   | 3,44   | 3,28   | 3,12   | 2,93     |
| 9                  | 5,12   | 4,26   | 3,86   | 3,63   | 3,48   | 3,37   | 3,23   | 3,07   | 2,90   | 2,71     |
| 10                 | 4,96   | 4,10   | 3,71   | 3,48   | 3,33   | 3,22   | 3,07   | 2,91   | 2,74   | 2,54     |
| 11                 | 4,84   | 3,98   | 3,59   | 3,36   | 3,20   | 3,09   | 2,95   | 2,79   | 2,61   | 2,40     |
| 12                 | 4,75   | 3,88   | 3,49   | 3,26   | 3,11   | 3,00   | 2,85   | 2,69   | 2,50   | 2,30     |
| 13                 | 4,67   | 3,80   | 3,41   | 3,18   | 3,02   | 2,92   | 2,77   | 2,60   | 2,42   | 2,21     |
| 14                 | 4,60   | 3,74   | 3,34   | 3,11   | 2,96   | 2,85   | 2,70   | 2,53   | 2,35   | 2,13     |
| 15                 | 4,54   | 3,68   | 3,29   | 3,06   | 2,90   | 2,79   | 2,64   | 2,48   | 2,29   | 2,07     |
| 16                 | 4,49   | 3,63   | 3,24   | 3,01   | 2,85   | 2,74   | 2,59   | 2,42   | 2,24   | 2,01     |
| 17                 | 4,45   | 3,59   | 3,20   | 2,96   | 2,81   | 2,70   | 2,55   | 2,38   | 2,19   | 1,96     |
| 18                 | 4,41   | 3,55   | 3,16   | 2,93   | 2,77   | 2,66   | 2,51   | 2,34   | 2,15   | 1,92     |
| 19                 | 4,38   | 3,52   | 3,13   | 2,90   | 2,74   | 2,63   | 2,48   | 2,31   | 2,11   | 1,88     |
| 20                 | 4,35   | 3,49   | 3,10   | 2,87   | 2,71   | 2,60   | 2,45   | 2,28   | 2,08   | 1,84     |
| 21                 | 4,32   | 3,47   | 3,07   | 2,84   | 2,68   | 2,57   | 2,42   | 2,25   | 2,05   | 1,81     |
| 22                 | 4,30   | 3,44   | 3,05   | 2,82   | 2,66   | 2,55   | 2,40   | 2,23   | 2,03   | 1,78     |
| 23                 | 4,28   | 3,42   | 3,03   | 2,80   | 2,64   | 2,53   | 2,38   | 2,20   | 2,00   | 1,76     |
| 24                 | 4,26   | 3,40   | 3,01   | 2,78   | 2,62   | 2,51   | 2,36   | 2,18   | 1,98   | 1,73     |
| 25                 | 4,24   | 3,38   | 2,99   | 2,76   | 2,60   | 2,49   | 2,34   | 2,16   | 1,96   | 1,71     |
| 26                 | 4,22   | 3,37   | 2,98   | 2,74   | 2,59   | 2,47   | 2,32   | 2,15   | 1,95   | 1,69     |
| 27                 | 4,21   | 3,35   | 2,96   | 2,73   | 2,57   | 2,46   | 2,30   | 2,13   | 1,93   | 1,67     |
| 28                 | 4,20   | 3,34   | 2,95   | 2,71   | 2,56   | 2,44   | 2,29   | 2,12   | 1,91   | 1,65     |
| 29                 | 4,18   | 3,33   | 2,93   | 2,70   | 2,54   | 2,43   | 2,28   | 2,10   | 1,90   | 1,64     |
| 30                 | 4,17   | 3,32   | 2,92   | 2,69   | 2,53   | 2,42   | 2,27   | 2,09   | 1,89   | 1,62     |
| 35                 | 4,12   | 3,26   | 2,87   | 2,64   | 2,48   | 2,37   | 2,22   | 2,04   | 1,83   | 1,57     |
| 40                 | 4,08   | 3,23   | 2,84   | 2,61   | 2,45   | 2,34   | 2,18   | 2,00   | 1,79   | 1,51     |
| 45                 | 4,06   | 3,21   | 2,81   | 2,58   | 2,42   | 2,31   | 2,15   | 1,97   | 1,76   | 1,48     |
| 50                 | 4,03   | 3,18   | 2,79   | 2,56   | 2,40   | 2,29   | 2,13   | 1,95   | 1,74   | 1,44     |
| 60                 | 4,00   | 3,15   | 2,76   | 2,52   | 2,37   | 2,25   | 2,10   | 1,92   | 1,70   | 1,39     |
| 70                 | 3,98   | 3,13   | 2,74   | 2,50   | 2,35   | 2,23   | 2,07   | 1,89   | 1,67   | 1,35     |
| 80                 | 3,96   | 3,11   | 2,72   | 2,49   | 2,33   | 2,21   | 2,06   | 1,88   | 1,65   | 1,31     |
| 90                 | 3,95   | 3,10   | 2,71   | 2,47   | 2,32   | 2,20   | 2,04   | 1,86   | 1,64   | 1,28     |
| 100                | 3,94   | 3,09   | 2,70   | 2,46   | 2,30   | 2,19   | 2,03   | 1,85   | 1,63   | 1,26     |
| 125                | 3,92   | 3,07   | 2,68   | 2,44   | 2,29   | 2,17   | 2,01   | 1,83   | 1,60   | 1,21     |

### Продолжение приложения А

| $k1 \backslash k2$ | 1    | 2    | 3    | 4    | 5    | 6    | 8    | 12   | 24   | $\infty$ |
|--------------------|------|------|------|------|------|------|------|------|------|----------|
|                    |      |      |      |      |      |      |      |      |      |          |
| 200                | 3,89 | 3,04 | 2,65 | 2,42 | 2,26 | 2,14 | 1,98 | 1,80 | 1,57 | 1,14     |
| 300                | 3,87 | 3,03 | 2,64 | 2,41 | 2,25 | 2,13 | 1,97 | 1,79 | 1,55 | 1,10     |
| 400                | 3,86 | 3,02 | 2,63 | 2,40 | 2,24 | 2,12 | 1,96 | 1,78 | 1,54 | 1,07     |
| 500                | 3,86 | 3,01 | 2,62 | 2,39 | 2,23 | 2,11 | 1,96 | 1,77 | 1,54 | 1,06     |
| 1000               | 3,85 | 3,00 | 2,61 | 2,38 | 2,22 | 2,10 | 1,95 | 1,76 | 1,53 | 1,03     |
| $\infty$           | 3,84 | 2,99 | 2,60 | 2,37 | 2,21 | 2,09 | 1,94 | 1,75 | 1,52 | 1,00     |

### Приложение Б (справочное)

**Критические значения t-критерия Стьюдента  
при уровне значимости 0,10; 0,05; 0,01(двухсторонний)**

| Число<br>степеней<br>свободы,<br>df | $\alpha$ |        |        | Число<br>степеней<br>свободы,<br>df | $\alpha$ |        |        |
|-------------------------------------|----------|--------|--------|-------------------------------------|----------|--------|--------|
|                                     | 0,10     | 0,05   | 0,01   |                                     | 0,10     | 0,05   | 0,01   |
| 1                                   | 6,3138   | 12,706 | 63,657 | 18                                  | 1,7341   | 2,1009 | 2,8784 |
| 2                                   | 2,9200   | 4,3027 | 9,9248 | 19                                  | 1,7291   | 2,0930 | 2,8609 |
| 3                                   | 2,3534   | 3,1825 | 5,8409 | 20                                  | 1,7247   | 2,0860 | 2,8453 |
| 4                                   | 2,13 18  | 2,7764 | 4,6041 | 21                                  | 1,7207   | 2,0796 | 2,8314 |
| 5                                   | 2,0150   | 2,5706 | 4,0321 | 22                                  | 1,7171   | 2,0739 | 2,8188 |
| 6                                   | 1,9432   | 2,4469 | 3,7074 | 23                                  | 1,7139   | 2,0687 | 2,8073 |
| 7                                   | 1,8946   | 2,3646 | 3,4995 | 24                                  | 1,7109   | 2,0639 | 2,7969 |
| 8                                   | 1,8595   | 2,3060 | 3,3554 | 25                                  | 1,7081   | 2,0595 | 2,7874 |
| 9                                   | 1,8331   | 2,2622 | 3,2498 | 26                                  | 1,7056   | 2,0555 | 2,7787 |
| 10                                  | 1,8125   | 2,2281 | 3,1693 | 27                                  | 1,7033   | 2,0518 | 2,7707 |
| 11                                  | 1,7959   | 2,2010 | 3,1058 | 28                                  | 1,7011   | 2,0484 | 2,7633 |
| 12                                  | 1,7823   | 2,1788 | 3,0545 | 29                                  | 1,6991   | 2,0452 | 2,7564 |
| 13                                  | 1,7709   | 2,1604 | 3,0123 | 30                                  | 1,6973   | 2,0423 | 2,7500 |
| 14                                  | 1,7613   | 2,1448 | 2,9768 | 40                                  | 1,6839   | 2,0211 | 2,7045 |
| 15                                  | 1,7530   | 2,1315 | 2,9467 | 60                                  | 1,6707   | 2,0003 | 2,6603 |
| 16                                  | 1,7459   | 2,1199 | 2,9208 | 120                                 | 1,6577   | 1,9799 | 2,6174 |
| 17                                  | 1,7396   | 2,1098 | 2,8982 | $\infty$                            | 1,6449   | 1,9600 | 2,5758 |

**Приложение В**  
**(справочное)**  
**Значения критерия Пирсона  $\chi^2$**

| Число<br>степен-<br>ней<br>свобо-<br>ды, <i>k</i> | Вероятность $\alpha$ |      |      |       |      |      |       |      |      |      |      |      |      |
|---------------------------------------------------|----------------------|------|------|-------|------|------|-------|------|------|------|------|------|------|
|                                                   | 0,99                 | 0,98 | 0,95 | 0,90  | 0,80 | 0,70 | 0,50  | 0,30 | 0,20 | 0,10 | 0,05 | 0,02 | 0,01 |
| 1                                                 | 0,00                 | 0,00 | 0,00 | 0,02  | 0,06 | 0,15 | 0,45  | 1,07 | 1,64 | 2,71 | 3,84 | 5,41 | 6,64 |
| 2                                                 | 0,02                 | 0,04 | 0,10 | 0,21  | 0,45 | 0,71 | 1,39  | 2,41 | 3,22 | 4,60 | 5,99 | 7,82 | 9,21 |
| 3                                                 | 0,11                 | 0,18 | 0,35 | 0,58  | 1,00 | 1,42 | 2,37  | 3,66 | 4,64 | 6,25 | 7,82 | 9,84 | 11,3 |
| 4                                                 | 0,30                 | 0,43 | 0,71 | 1,06  | 1,65 | 2,20 | 3,36  | 4,88 | 5,99 | 7,78 | 9,49 | 11,7 | 13,3 |
| 5                                                 | 0,55                 | 0,75 | 1,14 | 1,61  | 2,34 | 3,00 | 4,35  | 6,06 | 7,29 | 9,24 | 11,1 | 13,4 | 15,1 |
| 6                                                 | 0,87                 | 1,13 | 1,63 | 2,20  | 3,07 | 3,83 | 5,35  | 7,23 | 8,56 | 10,6 | 12,6 | 15,0 | 16,8 |
| 7                                                 | 1,24                 | 1,56 | 2,17 | 2,83  | 3,82 | 4,67 | 6,35  | 8,38 | 9,80 | 12,0 | 14,1 | 16,6 | 18,5 |
| 8                                                 | 1,65                 | 2,03 | 2,73 | 3,49  | 4,59 | 5,53 | 7,34  | 9,52 | 11,0 | 13,4 | 15,5 | 18,2 | 20,1 |
| 9                                                 | 2,09                 | 2,53 | 3,32 | 4,17  | 5,38 | 6,39 | 8,34  | 10,7 | 12,2 | 14,7 | 16,9 | 19,7 | 21,7 |
| 10                                                | 2,56                 | 3,06 | 3,94 | 4,86  | 6,18 | 7,27 | 9,34  | 11,8 | 13,4 | 16,0 | 18,3 | 21,2 | 23,2 |
| 11                                                | 3,05                 | 3,61 | 4,58 | 5,58  | 6,99 | 8,15 | 10,3  | 12,9 | 14,6 | 17,3 | 19,7 | 22,6 | 24,7 |
| 12                                                | 3,57                 | 4,18 | 5,23 | 6,30  | 7,81 | 9,03 | 11,3  | 14,0 | 15,8 | 18,5 | 21,0 | 24,1 | 26,2 |
| 13                                                | 4,11                 | 4,76 | 5,89 | 7,04  | 8,63 | 9,93 | 12,3  | 15,1 | 17,0 | 19,8 | 22,4 | 25,5 | 27,7 |
| 14                                                | 4,66                 | 5,37 | 6,57 | 7,79  | 9,47 | 10,8 | 13,3  | 16,2 | 18,1 | 21,1 | 23,7 | 26,9 | 29,1 |
| 15                                                | 5,23                 | 5,98 | 7,26 | 8,55' | 10,3 | 11,7 | 14,3  | 17,3 | 19,3 | 22,3 | 25,0 | 28,3 | 30,6 |
| 16                                                | 5,81                 | 6,61 | 7,96 | 9,31  | 11,1 | 12,6 | 15,3  | 18,4 | 20,5 | 23,5 | 26,3 | 29,6 | 32,0 |
| 17                                                | 6,41                 | 7,26 | 8,67 | 10,1  | 12,0 | 13,5 | 16,3  | 19,5 | 21,6 | 24,8 | 27,6 | 31,0 | 33,4 |
| 18                                                | 7,02                 | 7,91 | 9,39 | 10,9  | 12,9 | 14,4 | 17,3  | 20,6 | 22,8 | 26,0 | 28,9 | 32,3 | 34,8 |
| 19                                                | 7,63                 | 8,57 | 10,1 | 11,6  | 13,7 | 15,3 | 18,3  | 21,7 | 23,9 | 27,2 | 30,1 | 33,7 | 36,2 |
| 20                                                | 8,26                 | 9,24 | 10,8 | 12,4  | 14,6 | 16,3 | 19,3  | 22,8 | 25,0 | 28,4 | 31,4 | 35,0 | 37,6 |
| 21                                                | 8,90                 | 9,92 | 11,6 | 13,2  | 15,4 | 17,2 | 20,3  | 23,9 | 26,2 | 29,6 | 32,7 | 36,3 | 38,9 |
| 22                                                | 9,54                 | 10,6 | 12,3 | 14,0  | 16,3 | 18,1 | 21,3  | 24,9 | 27,3 | 30,8 | 33,9 | 37,7 | 40,3 |
| 23                                                | 10,2                 | 11,3 | 13,1 | 14,8  | 17,2 | 19,0 | 22,3- | 26,0 | 28,4 | 32,0 | 35,2 | 39,0 | 41,6 |
| 24                                                | 10,9                 | 12,0 | 13,8 | 15,7  | 18,1 | 19,9 | 23,3  | 27,1 | 29,6 | 33,2 | 36,4 | 40,3 | 43,0 |
| 25                                                | 11,5                 | 12,7 | 14,6 | 16,5  | 18,9 | 20,9 | 24,3  | 28,2 | 30,7 | 34,4 | 37,7 | 41,7 | 44,3 |
| 26                                                | 12,2                 | 13,4 | 15,4 | 17,3  | 19,8 | 21,8 | 25,3  | 29,2 | 31,8 | 35,6 | 38,9 | 42,9 | 45,6 |
| 27                                                | 12,9                 | 14,1 | 16,1 | 18,1  | 20,7 | 22,7 | 26,3  | 30,3 | 32,9 | 36,7 | 40,1 | 44,1 | 47,0 |
| 28                                                | 13,6                 | 14,8 | 16,9 | 18,9  | 21,6 | 23,6 | 27,3  | 31,4 | 34,0 | 37,9 | 41,3 | 45,4 | 48,3 |
| 29                                                | 14,3                 | 15,6 | 17,7 | 19,8  | 22,5 | 24,6 | 28,3  | 32,5 | 35,1 | 39,1 | 42,6 | 46,7 | 49,6 |
| 30                                                | 14,9                 | 16,3 | 18,5 | 20,6  | 23,4 | 25,5 | 29,3  | 33,5 | 36,2 | 40,3 | 43,8 | 48,0 | 50,9 |

Лариса Федоровна Филатова

## МНОЖЕСТВЕННАЯ РЕГРЕССИЯ

Учебное пособие

Научный редактор канд. физ.-мат. наук, доцент В.Н. Брендаков

Редактор Р.В. Фирсова

Технический редактор Л.Ф. Филатова

Подписано к печати \_\_\_\_\_ Формат 60х84/32.  
Гарнитура Times New Roman. Бумага писчая №2  
Плоская печать. Усл. печ. л. 1,45 Уч. изд. л. 2,63  
Тираж 50 экз. Заказ \_\_\_\_\_

Отпечатано в ИИО СТИ НИЯУ МИФИ  
636036. Томская обл., г. Северск,  
пр. Коммунистический. 65