



Министерство образования Республики Беларусь

Учреждение образования
«Гомельский государственный технический
университет имени П. О. Сухого»

Кафедра «Экономика и управление в отраслях»

О. Г. Винник

**ЭКОНОМЕТРИКА
И ЭКОНОМИКО-МАТЕМАТИЧЕСКИЕ
МЕТОДЫ И МОДЕЛИ**

ЗАДАЧНИК

**по одноименной дисциплине для студентов
экономических специальностей
дневной формы обучения**

Гомель 2011

УДК 330.43(075.8)
ББК 65в631я73
В48

*Рекомендовано научно-методическим советом
гуманитарно-экономического факультета ГГТУ им. П. О. Сухого
(протокол № 9 от 23.05.2011 г.)*

Рецензент: канд. экон. наук, доц. каф. алгебры и геометрии ГГУ им. Ф. Скорины
В. В. Подгорная

Винник, О. Г.
В48 Эконометрика и экономико-математические методы и модели : задачник по одной дисциплине для студентов экон. специальностей днев. формы обучения / О. Г. Винник. – Гомель : ГГТУ им. П. О. Сухого, 2011. – 62 с. – Систем. требования: PC не ниже Intel Celeron 300 МГц ; 32 Mb RAM ; свободное место на HDD 16 Mb ; Windows 98 и выше ; Adobe Acrobat Reader. – Режим доступа: <http://lib.gstu.local>. – Загл. с титул. экрана.

Содержит основные понятия и формулы эконометрики из разделов «Парная регрессия» и «Множественная регрессия». Рассмотрены решения типовых задач по обоим разделам. Представлены упражнения для самостоятельного решения.

Для студентов экономических специальностей дневной формы обучения.

**УДК 330.43(075.8)
ББК 65в631я73**

© Учреждение образования «Гомельский государственный технический университет имени П. О. Сухого», 2011

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	4
РАЗДЕЛ 1. ПАРНАЯ РЕГРЕССИЯ И КОРРЕЛЯЦИЯ	5
ПРИМЕРЫ ТИПОВЫХ ЗАДАЧ.....	14
УПРАЖНЕНИЯ.....	23
РАЗДЕЛ II. МНОЖЕСТВЕННАЯ РЕГРЕССИЯ И КОРРЕЛЯЦИЯ	28
РЕШЕНИЕ ТИПОВЫХ ЗАДАЧ	35
УПРАЖНЕНИЯ.....	49
СПИСОК РЕКОМЕНДУЕМОЙ ЛИТЕРАТУРЫ	58
ПРИЛОЖЕНИЯ	59

ВВЕДЕНИЕ

В настоящее время деятельность любой сферы требует от специалиста применения современных методов работы, понимания научного языка. Анализ информации в условиях ее недостатка и неполноты исходных данных требует знания специальных методов, которые составляют один из аспектов эконометрики. Большинство новых методов основано на эконометрических моделях, которые невозможно освоить без знаний эконометрики и навыков ее практического применения.

Методическое пособие предназначено для студентов, изучающих начальный курс эконометрики. Изучение этой дисциплины предполагает приобретение студентами опыта построения экономических моделей, принятия решения о спецификации и идентификации модели, выбора метода оценки параметров модели, интерпретации результатов, получения прогнозных оценок.

В разделах методического пособия представлены краткие методические положения: понятия, определения, формулы для расчетов. Кроме того, в первом и втором разделах рассмотрены решения наиболее типичных задач по парной и множественной регрессии с использованием приведенных формул. Для самостоятельной проработки изучаемых разделов в каждом из них приведены условия задач разной степени сложности, которые студентам предлагается рассмотреть с преподавателем на практических занятиях по данной дисциплине или решить самостоятельно. Формулировки приведенных заданий направлены на применение результатов эконометрического анализа.

В конце пособия приведены основные статистические таблицы, необходимые при решении приведенных задач.

Цель предлагаемого пособия – помощь в изучении основ дисциплины «Эконометрика» по разделам парная и множественная регрессия.

РАЗДЕЛ 1. ПАРНАЯ РЕГРЕССИЯ И КОРРЕЛЯЦИЯ

Парная регрессия представляет собой регрессию между двумя переменными – y и x , т. е. модель вида: $y = \hat{f}(x)$, где y – зависимая переменная (результативный признак); x – независимая, или объясняющая, переменная (признак-фактор). Знак « $\hat{}$ » означает, что между переменными x и y нет строгой функциональной зависимости, поэтому практически в каждом отдельном случае величина y складывается из двух слагаемых: $y = \hat{y}_x + \varepsilon$, где y – фактическое значение результативного признака; \hat{y}_x – теоретическое значение результативного признака, найденное исходя из уравнения регрессии; ε – случайная величина, характеризующая отклонения реального значения результативного признака от теоретического, найденного по уравнению регрессии.

Случайная величина ε называется также *возмущением*. Она включает влияние не учтенных в модели факторов, случайных ошибок и особенностей измерения. Ее присутствие в модели порождено тремя источниками: спецификацией модели, выборочным характером исходных данных, особенностями измерения переменных.

В парной регрессии выбор вида математической функции $\hat{y} = f(x)$ может быть осуществлен тремя методами:

- 1) графическим;
- 2) аналитическим, т.е. исходя из теории изучаемой взаимосвязи;
- 3) экспериментальным.

Различают линейные и нелинейные регрессии.

Нелинейные регрессии делятся на два класса: регрессии, нелинейные относительно включенных в анализ объясняющих переменных, но линейные по оцениваемым параметрам, и регрессии, нелинейные по оцениваемым параметрам.

Линейная регрессия: $y = a + b \cdot x + \varepsilon$.

Регрессии, нелинейные по объясняющим переменным:

- полиномы разных степеней: $y = a + b_1 \cdot x + b_2 \cdot x^2 + b_3 \cdot x^3 + \varepsilon$;

- равносторонняя гиперболола: $y = a + \frac{b}{x} + \varepsilon$;

Регрессии, нелинейные по оцениваемым параметрам:

- степенная: $y = a \cdot x^b \cdot \varepsilon$;

- показательная: $y = a \cdot b^x \cdot \varepsilon$;
- экспоненциальная: $y = e^{a+b \cdot x} \cdot \varepsilon$.

При изучении зависимости между двумя признаками графический метод подбора вида уравнения регрессии достаточно нагляден. Он основан на поле корреляции. Основные типы кривых, используемые при количественной оценке связей, представлены на рис. 1.1:

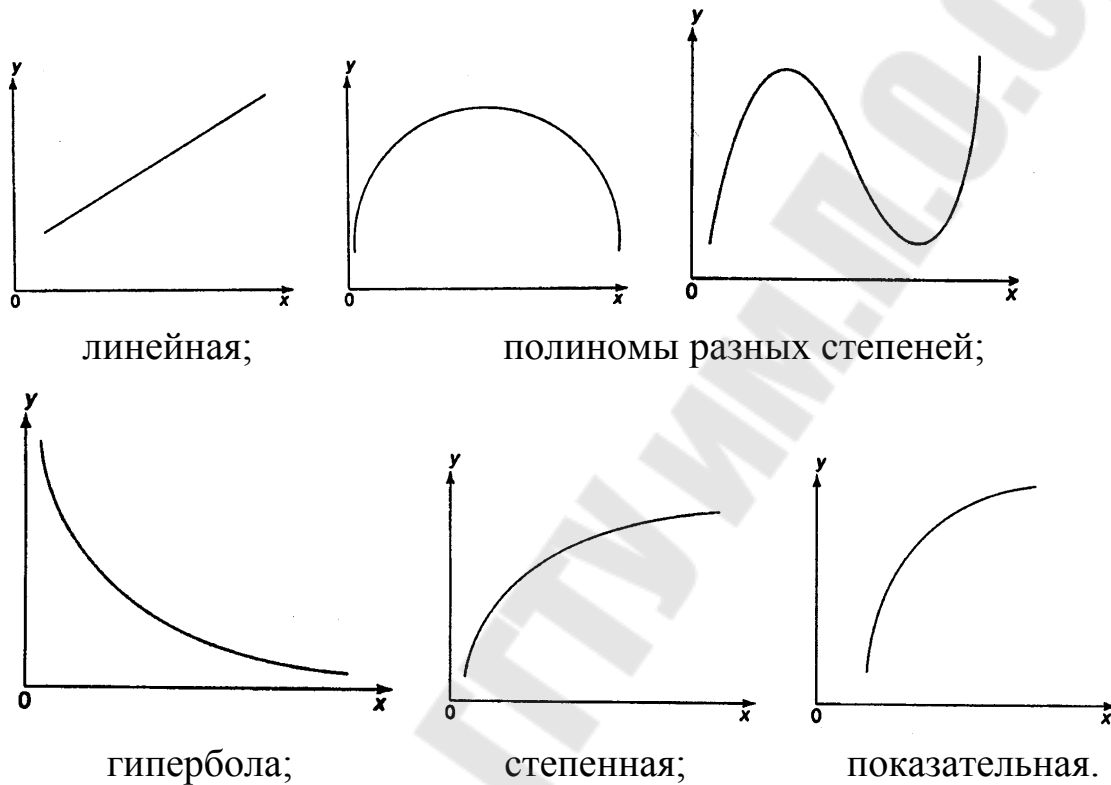


Рис. 1.1 – Основные типы кривых, используемые при количественной оценке связей между двумя переменными.

Класс математических функций для описания связи двух переменных достаточно широк. Кроме уже указанных используются и другие типы кривых:

$$y = \frac{1}{a + b \cdot x}; \quad y = a + b_1 \cdot x + b_2 \cdot \frac{1}{x}; \quad y = a + b \cdot \lg x;$$

$$y = \frac{1}{a + b_1 \cdot x + b_2 \cdot x^2}; \quad y = \frac{a}{1 + b_1 \cdot e^{-b_2 \cdot x}};$$

$$\lg y = a + b_1 \cdot x + b_2 \cdot x^2$$

Построение уравнения регрессии сводится к оценке ее параметров. Для оценки параметров исподльзуют метод наименьших квадратов (МНК). МНК позволяет получить такие оценки параметров, при которых сумма квадратов отклонений фактических значений резуль- тативного признака y от теоретических \hat{y}_x минимальна, то есть

$$\sum (y - \hat{y}_x)^2 \rightarrow \min \quad (1.1)$$

В практических исследованиях, величина отклонений фактиче- ских данных от теоретических лежит в основе расчета *остаточной дисперсии*:

$$\sigma_{ост}^2 = \frac{1}{n} \sum (y - \hat{y}_x)^2 \quad (1.2)$$

Чем меньше величина остаточной дисперсии, тем меньше влия- ние не учитываемых в уравнении регрессии факторов и тем лучше уравнение регрессии подходит к исходным данным.

Считается, что число наблюдений должно в 7-8 раз превышать число рассчитываемых параметров при переменной x . Это означает, что искать линейную регрессию, имея менее 7 наблюдений, вообще не имеет смысла. Если вид функции усложняется, то требуется увели- чение объема наблюдений, ибо каждый параметр при x должен рас- считываться хотя бы по 7 наблюдениям. Значит, если мы выбираем параболу второй степени $y = a + b_1 \cdot x + b_2 \cdot x^2$, то требуется объем ин- формации уже не менее 14 наблюдений.

Для линейных и нелинейных уравнений, приводимых к линей- ным, решается следующая система относительно a и b :

$$\begin{cases} na + b \sum x = \sum y, \\ a \sum x + b \sum x^2 = \sum yx. \end{cases} \quad (1.3)$$

Можно воспользоваться формулами, которые вытекают из этой системы:

$$a = \bar{y} - b \cdot \bar{x}; \quad b = \frac{\overline{y \cdot x} - \bar{y} \cdot \bar{x}}{\overline{x^2} - \bar{x}^2}. \quad (1.4)$$

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

Параметр a – формально показывает значение y при $x=0$. Если признак-фактор x не может иметь нулевого значения, то вышеуказанная трактовка свободного члена a не имеет смысла, т.е. параметр a может не иметь экономического содержания.

Уравнение регрессии всегда дополняется показателем тесноты связи. При использовании линейной регрессии в качестве такого показателя выступает линейный *коэффициент парной корреляции* r_{xy} , который можно рассчитать по следующим формулам:

$$r_{xy} = b \frac{\sigma_x}{\sigma_y} = \frac{\overline{y \cdot x} - \bar{y} \cdot \bar{x}}{\sigma_x \cdot \sigma_y} \quad (1.5)$$

Для линейной регрессии его значение находится в пределах $-1 \leq r_{xy} \leq 1$. Чем ближе абсолютное значение r_{xy} к единице, тем сильнее линейная связь между факторами (при $r = \pm 1$ имеем строгую функциональную зависимость). Теснота линейной связи между переменными может быть оценена на основании шкалы Чеддока:

Теснота связи	Значение коэффициента корреляции при наличии:	
	Прямой связи	Обратной связи
Слабая	0,1–0,3	(–0,3)–(–0,1)
Умеренная	0,3–0,5	(–0,5)–(–0,3)
Заметная	0,5–0,7	(–0,7)–(–0,5)
Высокая	0,7–0,9	(–0,9)–(–0,7)
Весьма высокая	0,9–1	(–1)–(–0,9)

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

Для оценки качества подбора линейной функции рассчитывается квадрат линейного коэффициента корреляции r_{xy}^2 , называемый *коэффициентом детерминации*. Коэффициент детерминации характеризует долю дисперсии результативного признака y , объясняемую регрессией, в общей дисперсии результативного признака:

$$r_{xy}^2 = 1 - \frac{\sigma_{ост}^2}{\sigma_y^2} \quad (1.6)$$

где

$$\sigma_{ост}^2 = \frac{1}{n} \sum (y - \hat{y}_x)^2; \quad \sigma_y^2 = \frac{1}{n} \sum (y - \bar{y})^2 = \overline{y^2} - \bar{y}^2. \quad (1.7)$$

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

Для нелинейной регрессии показателем тесноты связи выступает *индекс корреляции*:

$$\rho_{xy} = \sqrt{1 - \frac{\sigma_{ост}^2}{\sigma_y^2}} = \sqrt{1 - \frac{\sum (y - \hat{y}_x)^2}{\sum (y - \bar{y})^2}} \quad (1.8)$$

Величина данного показателя находится в пределах $0 \leq \rho_{xy} \leq 1$. Чем ближе значение индекса корреляции к единице, тем теснее связь рассматриваемых признаков, тем более надежно уравнение регрессии.

Квадрат индекса корреляции носит название *индекса детерминации* и характеризует долю дисперсии результативного признака y , объясняемую регрессией, в общей дисперсии результативного признака:

$$\rho_{xy}^2 = 1 - \frac{\sigma_{ост}^2}{\sigma_y^2} = \frac{\sigma_{объясн}^2}{\sigma_y^2} \quad (1.9)$$

Индекс детерминации ρ_{xy}^2 можно сравнивать с коэффициентом детерминации r_{xy}^2 для обоснования возможности применения линейной функции. Чем больше кривизна линии регрессии, тем величина r_{xy}^2 меньше ρ_{xy}^2 . А близость этих показателей указывает на то, что нет необходимости усложнять форму уравнения регрессии и можно использовать линейную функцию.

Оценка значимости уравнения регрессии в целом производится на основе F -критерия Фишера, которому предшествует дисперсионный анализ.

Согласно основной идее дисперсионного анализа, общая сумма квадратов отклонений переменной y от среднего значения \bar{y} раскладывается на две части – «объясненную» и «необъясненную»:

$$\sum(y - \bar{y})^2 = \sum(\hat{y}_x - \bar{y})^2 + \sum(y - \hat{y}_x)^2 \quad (1.10)$$

где $\sum(y - \bar{y})^2$ – общая сумма квадратов отклонений; $\sum(\hat{y}_x - \bar{y})^2$ – сумма квадратов отклонений, объясненная регрессией (или факторная сумма квадратов отклонений); $\sum(y - \hat{y}_x)^2$ – остаточная сумма квадратов отклонений, характеризующая влияние неучтенных в модели факторов.

Таблица 1.1

Компоненты дисперсии	Сумма квадратов	Число степеней свободы	Дисперсия на одну степень свободы
Общая	$\sum(y - \bar{y})^2$	$n-1$	$S_{общ}^2 = \frac{\sum(y - \bar{y})^2}{n-1}$
Факторная	$\sum(\hat{y}_x - \bar{y})^2$	m	$S_{факт}^2 = \frac{\sum(\hat{y}_x - \bar{y})^2}{m}$
Остаточная	$\sum(y - \hat{y}_x)^2$	$n-m-1$	$S_{ост}^2 = \frac{\sum(y - \hat{y}_x)^2}{n-m-1}$

Схема дисперсионного анализа имеет вид, представленный в табл. 1.1 (n – число наблюдений, m – число параметров при переменной x).

Определение дисперсии на одну степень свободы приводит дисперсии к сравнимому виду. Сопоставляя факторную и остаточную дисперсии в расчете на одну степень свободы, получим величину F -критерия Фишера:

$$F = \frac{S_{\text{факт}}^2}{S_{\text{ост}}^2} \quad (1.11)$$

Величина F -критерия связана с коэффициентом детерминации r_{xy}^2 и индексом детерминации ρ_{xy}^2 , и ее можно рассчитать по следующим формулам:

$$F = \frac{r_{xy}^2}{1-r_{xy}^2} \cdot (n-2); \quad F = \frac{\rho_{xy}^2}{1-\rho_{xy}^2} \cdot \frac{n-m-1}{m} \quad (1.12)$$

где r_{xy}^2 - коэффициент детерминации, ρ_{xy}^2 - индекс детерминации, n - число наблюдений, m - число параметров при переменной x .

Фактическое значение F -критерия Фишера, полученное по приведенным формулам, сравнивается с табличным значением $F_{\text{табл}}(\alpha; k_1; k_2)$ (приложение А) при уровне значимости α и числе степеней свободы $k_2=n-m-1$ (для остаточной суммы квадратов) и $k_1=m$ (для факторной суммы квадратов). При этом, если фактическое значение F -критерия больше табличного, то признается статистическая значимость и надежность уравнения в целом.

О качестве как линейного, так и нелинейного уравнения регрессии можно также судить и по средней ошибке аппроксимации, которая отражает среднее отклонение расчетных значений от фактических:

$$\bar{A} = \frac{1}{n} \sum \left| \frac{y - \hat{y}_x}{y} \right| \cdot 100\% \quad (1.13)$$

Допустимый предел значений \bar{A} - не более 8-10%.

Средний коэффициент эластичности показывает, на сколько процентов в среднем по совокупности изменится результат y от своей средней величины, если фактор x изменится на 1% от своего среднего

значения. Формула для расчета среднего коэффициента эластичности имеет вид:

$$\bar{\varepsilon} = f'(x) \cdot \frac{\bar{x}}{y} \quad (1.14)$$

Формулы для расчета средних коэффициентов эластичности для наиболее часто используемых типов уравнений регрессии приведены в приложении Б.

Для оценки статистической значимости коэффициентов регрессии и корреляции рассчитываются t -критерий Стьюдента и доверительные интервалы каждого из показателей. Оценка значимости коэффициентов регрессии и корреляции с помощью t -критерия Стьюдента проводится путем сопоставления их значений с величиной случайной ошибки:

$$t_b = \frac{b}{m_b}, \quad t_a = \frac{a}{m_a}, \quad t_r = \frac{r}{m_r}, \quad (1.15)$$

которое затем сравнивается с табличным значением (приложение В) при определенном уровне значимости α и числе степеней свободы $(n-2)$. Доверительный интервал для коэффициента регрессии определяется как $b \pm t_{табл} \cdot m_b$. Границы доверительного интервала для коэффициента регрессии не должны содержать противоречивых результатов, например, $-1,5 \leq b \leq 0,8$. Такого рода запись указывает, что истинное значение коэффициента регрессии одновременно содержит положительные и отрицательные величины и даже ноль, чего не может быть, и коэффициент признается незначимым.

Случайные ошибки параметров линейной регрессии и коэффициента корреляции определяются по формулам:

$$m_b = \sqrt{\frac{\sum (y - \hat{y}_x)^2 / (n-2)}{\sum (x - \bar{x})^2}} = \sqrt{\frac{S_{оцм}^2}{\sum (x - \bar{x})^2}} = \frac{S_{оцм}}{\sigma_x \cdot \sqrt{n}}; \quad (1.16)$$

$$m_a = \sqrt{\frac{\sum (y - \hat{y}_x)^2}{(n-2)} \cdot \frac{\sum x^2}{n \sum (x - \bar{x})^2}} = \sqrt{S_{оцм}^2 \frac{\sum x^2}{n^2 \cdot \sigma_x^2}} = S_{оцм} \frac{\sqrt{\sum x^2}}{n \cdot \sigma_x}; \quad (1.17)$$

$$m_{r_{xy}} = \sqrt{\frac{1-r_{xy}^2}{n-2}}. \quad (1.18)$$

Существует связь между t -критерием Стьюдента и F -критерием Фишера:

$$t_b = t_r = \sqrt{F} \quad (1.19)$$

Если $t_{\text{табл}} < t_{\text{факт}}$, то гипотеза о случайной природе показателей отклоняется, т.е. a , b и r_{xy} не случайно отличаются от нуля и сформировались под влиянием систематически действующего фактора x .

Прогнозное значение y_p определяется путем подстановки в уравнение регрессии $\hat{y}_x = a + b \cdot x$ соответствующего значения x . Однако точечный прогноз явно не реален. Поэтому он дополняется расчетом стандартной ошибки прогноза:

$$m_{\hat{y}_p} = \sigma_{\text{ост}} \cdot \sqrt{1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{\sum(x - \bar{x})^2}} \quad (1.20)$$

где

$$\sigma_{\text{ост}} = \sqrt{\frac{\sum(y - \hat{y})^2}{n - m - 1}} \quad (1.21)$$

и соответственно интервальной оценкой прогнозного значения:

$$\hat{y}_p - \Delta \hat{y}_p \leq \hat{y}_p \leq \hat{y}_p + \Delta \hat{y}_p \quad (1.22)$$

где

$$\Delta \hat{y}_p = m_{\hat{y}_p} \cdot t_{\text{табл}}. \quad (1.23)$$

$$m_{\hat{y}_p} = S_{\text{ост}} \cdot \sqrt{1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{n \cdot \sigma_x^2}}. \quad (1.24)$$

ПРИМЕРЫ ТИПОВЫХ ЗАДАЧ

Задача 1.

По семи территориям Уральского района за 20XX г. известны значения двух признаков (табл. 1.1).

Таблица 1.1

Район	Расходы на покупку продовольственных товаров в общих расходах, %, y	Среднедневная заработная плата одного работающего, руб., x
Удмуртская респ.	68,8	45,1
Свердловская обл.	61,2	59,0
Башкортостан	59,9	57,2
Челябинская обл.	56,7	61,8
Пермская обл.	55,0	58,8
Курганская обл.	54,3	47,2
Оренбургская обл.	49,3	55,2

Требуется:

- Для характеристики зависимости y от x рассчитать параметры следующих функций:
 - линейной;
 - степенной;
 - равносторонней гиперболы.
- Оценить каждую модель через среднюю ошибку аппроксимации \bar{A} и F -критерий Фишера.

Решение:

1а. Для расчета параметров a и b линейной регрессии $y = a + bx$ решаем систему нормальных уравнений относительно a и b :

$$\begin{cases} na + b\sum x = \sum y, \\ a\sum x + b\sum x^2 = \sum yx. \end{cases}$$

По исходным данным рассчитываем $\sum y$, $\sum x$, $\sum yx$, $\sum x^2$, $\sum y^2$.

Таблица 1.1

	y	x	yx	x^2	y^2	\hat{y}_x	$y - \hat{y}_x$	A_i
1	68,8	45,1	3102,88	2034,01	4733,44	61,3	7,5	10,9
2	61,2	59,0	3610,80	3481,00	3745,44	56,5	4,7	7,7
3	59,9	57,2	3426,28	3271,84	3588,01	57,1	2,8	4,7
4	56,7	61,8	3504,06	3819,24	3214,89	55,5	1,2	2,1
5	55,0	58,8	3234,00	3457,44	3025,00	56,5	-1,5	2,7
6	54,3	47,2	2562,96	2227,84	2948,49	60,5	-6,2	11,4
7	49,3	55,2	2721,36	3047,04	2430,49	57,8	-8,5	17,2
Итого:	405,2	384,3	22162,34	21338,41	23685,76	405,2	0,0	56,7
Среднее значение:	57,89	54,90	3166,05	3048,34	3383,68	X	X	8,1
σ	5,74	5,86	X	X	X	X	X	X
σ^2	32,92	34,34	X	X	X	X	X	X

$$b = \frac{\overline{y \cdot x} - \bar{y} \cdot \bar{x}}{\overline{x^2} - \bar{x}^2} = \frac{3166,05 - 57,89 \cdot 54,9}{5,86^2} \approx -0,35;$$

$$a = \bar{y} - b \cdot \bar{x} = 57,89 - 0,35 \cdot 54,9 \approx 76,88.$$

Уравнение регрессии: $\hat{y} = 76,88 - 0,35 \cdot x$. С увеличением среднедневной заработной платы на 1 руб. доля расходов на покупку продовольственных товаров снижается в среднем на 0,35 %-ных пункта.

Рассчитаем линейный коэффициент парной корреляции:

$$r_{xy} = b \frac{\sigma_x}{\sigma_y} = -0,35 \frac{5,86}{5,74} = -0,357.$$

Связь умеренная, обратная.

Определим коэффициент детерминации:

$$r_{xy}^2 = (-0,35)^2 = 0,127.$$

Вариация результата на 12,7% объясняется вариацией фактора x . Подставляя в уравнение регрессии фактические значения x , определим теоретические (расчетные) значения \hat{y}_x . Найдем величину средней ошибки аппроксимации A :

$$\bar{A} = \frac{1}{n} \sum A_i = \frac{1}{n} \sum \left| \frac{y - \hat{y}_x}{y} \right| \cdot 100\% = \frac{56,7 \cdot 100\%}{7} = 8,1\%.$$

В среднем расчетные значения отклоняются от фактических на 8,1%.

Рассчитаем F -критерий:

$$F_{\text{факт}} = \frac{0,127}{0,873} \cdot 5 = 0,7.$$

поскольку $1 \leq F \leq \infty$, следует рассмотреть F^{-1} .

Полученное значение указывает на необходимость принять гипотезу H_0 о случайной природе выявленной зависимости и статистической незначимости параметров уравнения и показателя тесноты связи.

16. Построению степенной модели $y = a \cdot x$ предшествует процедура линеаризации переменных. В примере линеаризация производится путем логарифмирования обеих частей уравнения:

$$\lg y = \lg a + b \cdot \lg x;$$

$$Y = C + b \cdot X,$$

где $Y = \lg y$, $X = \lg x$, $C = \lg a$.

Для расчетов используем данные табл. 1.2.

Таблица 1.2

	Y	X	YX	Y^2	X^2	\hat{y}_x	$y - \hat{y}_x$	$(y - \hat{y}_x)^2$	A_i
1	1,8376	1,6542	3,0398	3,3768	2,7364	61,0	7,8	60,8	11,3
2	1,7868	1,7709	3,1642	3,1927	3,1361	56,3	4,9	24,0	8,0
3	1,7774	1,7574	3,1236	3,1592	3,0885	56,8	3,1	9,6	5,2
4	1,7536	1,7910	3,1407	3,0751	3,2077	55,5	1,2	1,4	2,1
5	1,7404	1,7694	3,0795	3,0290	3,1308	56,3	-1,3	1,7	2,4
6	1,7348	1,6739	2,9039	3,0095	2,8019	60,2	-5,9	34,8	10,9
7	1,6928	1,7419	2,9487	2,8656	3,0342	57,4	-8,1	65,6	16,4
Итого:	12,3234	12,1587	21,4003	21,7078	21,1355	403,5	1,7	197,9	56,3
Среднее значение	1,7605	1,7370	3,0572	3,1011	3,0194	X	X	28,27	8,0
σ	0,0425	0,0484	X	X	X	X	X	X	X
σ^2	0,0018	0,0023	X	X	X	X	X	X	X

Рассчитаем C и b :

$$b = \frac{\bar{Y} \cdot \bar{X} - \bar{Y} \cdot \bar{X}}{\sigma_x^2} = \frac{3,0572 - 1,7605 \cdot 1,7370}{0,0484^2} \approx -0,298;$$

$$C = \bar{Y} - b \cdot \bar{X} = 1,7605 - 0,298 \cdot 1,7370 = 2,278.$$

Получим линейное уравнение: $Y = 2,278 - 0,298 \cdot X$.

Выполнив его потенцирование, получим:

$$\hat{y} = 10^{2,278} \cdot x^{-0,298} = 189,7 \cdot x^{-0,298}$$

Подставляя в данное уравнение фактические значения x , получаем теоретические значения результата \hat{y}_x . По ним рассчитаем показатели: тесноты связи - индекс корреляции ρ_{xy} и среднюю ошибку аппроксимации \bar{A}_i :

$$\rho_{xy} = \sqrt{1 - \frac{\sum(y - \hat{y}_x)^2}{\sum(y - \bar{y})^2}} = \sqrt{1 - \frac{28,27}{32,92}} = 0,3758, \quad \bar{A} = 8,0\%$$

Характеристики степенной модели указывают, что она несколько лучше линейной функции описывает взаимосвязь.

1в. Уравнение равносторонней гиперболы $y = a + \frac{b}{x}$ линеаризуется при замене: $z = \frac{1}{x}$. Тогда $y = a + b \cdot z$.

Для расчетов используем данные табл. 1.3.

Таблица 1.3.

	y	z	yz	z^2	y^2	\hat{y}_x	$y - \hat{y}_x$	$(y - \hat{y}_x)^2$	A_i
1	68,8	0,0222	1,5255	0,000492	4733,44	61,8	7,0	49,00	10,2
2	61,2	0,0169	1,0373	0,000287	3745,44	56,3	4,9	24,01	8,0
3	59,9	0,0175	1,0472	0,000306	3588,01	56,9	3,0	9,00	5,0
4	56,7	0,0162	0,9175	0,000262	3214,89	55,5	1,2	1,44	2,1
5	55	0,0170	0,9354	0,000289	3025,00	56,4	-1,4	1,96	2,5
6	54,3	0,0212	1,1504	0,000449	2948,49	60,8	-6,5	42,25	12,0
7	49,3	0,0181	0,8931	0,000328	2430,49	57,5	-8,2	67,24	16,6
Итого:	405,2	0,1291	7,5064	0,002413	23685,76	405,2	0,0	194,90	56,5
Среднее значение	57,9	0,0184	1,0723	0,000345	3383,68	X	X	27,84	8,1
σ	5,74	0,002145	X	X	X	X	X	X	X
σ^2	32,9476	0,000005	X	X	X	X	X	X	X

Значения параметров регрессии a и b составили:

$$a = \bar{y} - b \cdot \bar{z} = 57,89 - 1051,4 \cdot 0,0184 = 38,5;$$

$$b = \frac{\bar{y} \cdot \bar{z} - \bar{y} \cdot \bar{z}}{\sigma_x^2} = \frac{1,0723 - 57,9 \cdot 0,0184}{0,002145^2} \approx -1051,4,$$

Получено уравнение: $\hat{y} = 38,5 + 1051,4 \cdot \frac{1}{x}$.

Индекс корреляции:

$$\rho_{xy} = \sqrt{1 - \frac{27,84}{32,92}} = 0,3944$$

По уравнению гиперболы получена наибольшая оценка тесноты связи (0,3944) по сравнению с линейной и степенной регрессиями. \bar{A} остается на допустимом уровне.

2.

$$F = \frac{\rho_{xy}^2}{1 - \rho_{xy}^2} \cdot \frac{n - m - 1}{m} = \frac{0,1555}{0,8445} = 0,92,$$

где $F_{\text{табл}} = 6,6 > F_{\text{факт}}, \alpha = 0,05$.

Следовательно, принимается гипотеза H_0 о статистически незначимых параметрах этого уравнения. Этот результат можно объяснить сравнительно невысокой теснотой выявленной зависимости и небольшим числом наблюдений.

Задача 2.

По территориям региона приводятся данные за 20XX г. (табл. 1.4).

Таблица 1.4.

Номер региона	Среднедушевой прожиточный минимум в день одного трудоспособного, руб., x	Среднедневная заработная плата, руб., y
1	78	133
2	82	148
3	87	134
4	79	154
5	89	162
6	106	195
7	67	139
8	88	158
9	73	152
10	87	162
11	76	159
12	115	173

Требуется:

1. Построить линейное уравнение парной регрессии y от x .
2. Рассчитать линейный коэффициент парной корреляции и среднюю ошибку аппроксимации.

3. Оценить статистическую значимость параметров регрессии и корреляции.
4. Выполнить прогноз заработной платы y при прогнозном значении среднедушевого прожиточного минимума, составляющем 107% от среднего уровня.
5. Оценить точность прогноза, рассчитав ошибку прогноза и его доверительный интервал.

Решение:

1. Для расчета параметров уравнения линейной регрессии строим расчетную таблицу (табл. 1.5).

Таблица 1.5.

	y	x	yx	x^2	y^2	\hat{y}_x	$y - \hat{y}_x$	A_i
1	78	133	10374	6084	17689	149	-16	12,0
2	82	148	12136	6724	21904	152	-4	2,7
3	87	134	11658	7569	17956	157	-23	17,2
4	79	154	12166	6241	23716	150	4	2,6
5	89	162	14418	7921	26244	159	3	1,9
6	106	195	20670	11236	38025	174	21	10,8
7	67	139	9313	4489	19321	139	0	0,0
8	88	158	13904	7744	24964	158	0	0,0
9	73	152	11096	5329	23104	144	8	5,3
10	87	162	14094	7569	26244	157	5	3,1
11	76	159	12084	5776	25281	147	12	7,5
12	115	173	19895	13225	29929	183	-10	5,8
Итого:	1027	1869	161808	89907	294377	1869	0	68,8
Среднее значение	85,6	155,8	13484,0	7492,3	24531,4	X	X	5,7
σ	12,95	16,53	X	X	X	X	X	X
σ^2	167,7	273,4	X	X	X	X	X	X

$$b = \frac{\overline{y \cdot x} - \bar{y} \cdot \bar{x}}{\overline{x^2} - \bar{x}^2} = \frac{13484 - 85,6 \cdot 155,8}{7492,3 - 85,6^2} = \frac{151,8}{164,94} = 0,92,$$

$$a = \bar{y} - b \cdot \bar{x} = 155,8 - 0,92 \cdot 85,6 = 77,0.$$

Получено уравнение регрессии: $y = 77,0 + 0,92 \cdot x$. С увеличением среднедушевого прожиточного минимума на 1 руб. среднедневная заработная плата возрастает в среднем на 0,92 руб.

2. Тесноту линейной связи оценит коэффициент корреляции:

$$r_{xy} = b \frac{\sigma_x}{\sigma_y} = \frac{\overline{y \cdot x} - \bar{y} \cdot \bar{x}}{\sigma_x \cdot \sigma_y} = 0,92 \cdot \frac{12,95}{16,53} = 0,721; \quad r_{xy}^2 = 0,52.$$

Это означает, что 52% вариации заработной платы (y) объясняется вариацией фактора x - среднедушевого прожиточного минимума.

Качество модели определяет средняя ошибка аппроксимации:

$$\bar{A} = \frac{1}{n} \sum A_i = \frac{68,9}{12} = 5,7\%.$$

Качество построенной модели оценивается как хорошее, так как \bar{A} не превышает 8-10%.

3. Оценку статистической значимости параметров регрессии проведем с помощью t -статистики Стьюдента и путем расчета доверительного интервала каждого из показателей.

Выдвигаем гипотезу H_0 о статистически незначимом отличии показателей от нуля: $a = b = r_{xy} = 0$.

$t_{\text{табл}}$ для числа степеней свободы $df = n - 2 = 12 - 2 = 10$ и $\alpha = 0,05$ составит 2,23.

Определим случайные ошибки $m_a, m_b, m_{r_{xy}}$:

$$m_a = 12,6 \frac{\sqrt{89907}}{12 \cdot 12,95} = 24,3;$$

$$m_b = \frac{12,6}{12,95 \cdot \sqrt{12}} = 0,281;$$

$$m_{r_{xy}} = \sqrt{\frac{1 - 0,520}{12 - 2}} = 0,219$$

Тогда

$$t_a = \frac{77}{24,3} = 3,2; \quad t_b = \frac{0,92}{0,281} = 3,3; \quad t_{r_{xy}} = \frac{0,721}{0,219} = 3,3.$$

Фактические значения t -статистики превосходят табличные значения:

$$t_a = 3,2 > t_{\text{табл}} = 2,3; \quad t_b = 3,3 > t_{\text{табл}} = 2,3; \quad t_{r_{xy}} = 3,3 > t_{\text{табл}} = 2,3,$$

поэтому гипотеза H_0 отклоняется, т.е. a, b и r_{xy} не случайно отличаются от нуля, а статистически значимы.

Рассчитаем доверительный интервал для a и b . Для этого определим предельную ошибку для каждого показателя:

$$\Delta_a = 2,23 \cdot 24,3 = 54; \quad \Delta_b = 2,23 \cdot 0,281 = 0,62.$$

Доверительные интервалы:

$$\begin{aligned}\gamma_a &= a \pm \Delta_a = 77 \pm 54; \\ \gamma_{amin} &= 77 - 54 = 23; \\ \gamma_{amax} &= 77 + 54 = 131; \\ \gamma_b &= b \pm \Delta_b = 0,92 \pm 0,62; \\ \gamma_{bmin} &= 0,92 - 0,62 = 0,30; \\ \gamma_{bmax} &= 0,92 + 0,62 = 1,54;\end{aligned}$$

Анализ верхней и нижней границ доверительных интервалов приводит к выводу о том, что с вероятностью $p = 1 - \alpha = 0,95$ параметры a и b , находясь в указанных границах, не принимают нулевых значений, т.е. не являются статистически незначимыми и существенно отличны от нуля.

4. Полученные оценки уравнения регрессии позволяют использовать его для прогноза. Если прогнозное значение прожиточного минимума составит: $x_p = \bar{x} \cdot 1,07 = 85,6 \cdot 1,07 = 91,6$ тыс. руб., тогда прогнозное значение прожиточного минимума составит: $\hat{y}_p = 77 + 0,92 \cdot 91,6 = 161$ тыс. руб.

5. Ошибка прогноза составит:

$$m_{\hat{y}_p} = 12,6 \cdot \sqrt{1 + \frac{1}{12} + \frac{(91,6 - 85,6)^2}{12 \cdot 12,95^2}} = 13,2 \text{ (тыс. руб.)}.$$

Предельная ошибка прогноза, которая в 95% случаев не будет превышена, составит:

$$\Delta \hat{y}_p = t_{табл} \cdot m_{\hat{y}_p} = 2,23 \cdot 13,2 = 29,4.$$

Доверительный интервал прогноза:

$$\begin{aligned}\gamma_{\hat{y}_p} &= 91,6 \pm 29,4; \\ \gamma_{\hat{y}_{min}} &= 91,6 - 29,4 = 62,2; \\ \gamma_{\hat{y}_{max}} &= 91,6 + 29,4 = 121 \text{ (руб.)}.\end{aligned}$$

Выполненный прогноз среднемесячной заработной платы оказался надежным ($p = 1 - \alpha = 1 - 0,05 = 0,95$), но неточным, так как диапазон верхней и нижней границ доверительного интервала D_r составляет 1,95 раза:

Задача 3.

По группе предприятий, производящих однородную продукцию, известно, как зависит себестоимость единицы продукции y от факторов, приведенных в табл. 1.6.

Таблица 1.6.

Признак-фактор	Уравнение парной регрессии	Среднее значение фактора
Объем производства, млн руб., x_1	$\hat{y}_{x_1} = 0,62 + 58,74 \cdot \frac{1}{x_1}$	$\bar{x}_1 = 2,64$
Трудоемкость единицы продукции, чел.-час, x_2	$\hat{y}_{x_2} = 9,30 + 9,83 \cdot x_2$	$\bar{x}_2 = 1,38$
Оптовая цена за 1 т энергоносителя, МЛН руб., x_3	$\hat{y}_{x_3} = 11,75 + x_3^{1,6281}$	$\bar{x}_3 = 1,503$
Доля прибыли, изымаемой государством, %, x_4	$\hat{y}_{x_4} = 14,87 \cdot 1,016^{x_4}$	$\bar{x}_4 = 26,3$

Требуется:

1. Определить с помощью коэффициентов эластичности силу влияния каждого фактора на результат.
2. Ранжировать факторы по силе влияния.

Решение:

1. Для уравнения равнобочной гиперболы $\hat{y}_{x_1} = 0,62 + 58,74 \cdot \frac{1}{x_1}$:

$$\bar{\varepsilon}_{yx_1} = f'(x_1) \frac{\bar{x}_1}{y} = -\frac{b}{\bar{x}_1^2} \cdot \frac{\bar{x}_1}{a + b/\bar{x}_1} = -\frac{b}{a \cdot \bar{x}_1 + b} = -\frac{58,74}{0,62 \cdot 2,64 + 58,74} = -0,973\%.$$

Для уравнения прямой $\hat{y}_{x_2} = 9,30 + 9,83 \cdot x_2$:

$$\bar{\varepsilon}_{yx_2} = f'(x_2) \frac{\bar{x}_2}{y} = \frac{b \cdot \bar{x}_2}{a + b/\bar{x}_2} = \frac{9,83 \cdot 1,38}{9,30 + 9,83 \cdot 1,38 + 58,74} = 0,59\%.$$

Для уравнения степенной зависимости $\hat{y}_{x_3} = 11,75 + x_3^{1,6281}$:

$$\bar{\varepsilon}_{yx_3} = f'(x_3) \frac{\bar{x}_3}{y} = a \cdot b \cdot \bar{x}_3^{b-1} \cdot \frac{\bar{x}_3}{a \cdot \bar{x}_3} = b = 1,63\%.$$

Для уравнения показательной зависимости $\hat{y}_{x_4} = 14,87 \cdot 1,016^{x_4}$:

$$\bar{\varepsilon}_{yx_4} = f'(x_4) \frac{\bar{x}_4}{y} = a \cdot b \bar{x}_4 \cdot \ln b \cdot \frac{\bar{x}_4}{a \cdot b \bar{x}_4} = \ln b \cdot \bar{x}_4 = \ln 1,016 \cdot 26,3 = 0,42\%.$$

2. Сравнивая значения $\bar{\mathcal{E}}_{yx_j}$, ранжируем x_j по силе их влияния на себестоимость единицы продукции:

а) $\bar{\mathcal{E}}_{yx_3} = 1,63\%$;

б) $\bar{\mathcal{E}}_{yx_1} = -0,973\%$;

в) $\bar{\mathcal{E}}_{yx_2} = 0,59\%$;

г) $\bar{\mathcal{E}}_{yx_4} = 0,42\%$.

Для формирования уровня себестоимости продукции группы предприятия первоочередное значение имеют цены на энергоносители; в гораздо меньшей степени влияют трудоемкость продукции и отчисляемая часть прибыли. Фактором снижения себестоимости выступает размер производства: с ростом его на 1% себестоимость единицы продукции снижается на -0,97%.

УПРАЖНЕНИЯ

1.1. Получены функции:

1. $y = a + b \cdot x^3 + \varepsilon$,

2. $y = a + b \cdot \lg x + \varepsilon$,

3. $\ln y = a + b \cdot \ln x + \varepsilon$,

4. $y = a + b_1 \cdot x^{b_2} + \varepsilon$,

5. $y^a = b_1 + b_2 x^2 + \varepsilon$,

6. $y = 1 + a \cdot (1 - x^b) + \varepsilon$,

7. $y = a + b \frac{x}{10} + \varepsilon$.

Определите, какие из представленных выше функций линейны по переменным, линейны по параметрам, нелинейны ни по переменным, ни по параметрам.

1.2. Имеются следующие данные об уровне механизации работ $x(\%)$ и производительности труда y (т/ч) для 14 однотипных предприятий:

x_i	32	30	36	40	41	47	56	54	60	55	61	67	69	76
y_i	20	24	28	30	31	33	34	37	38	40	41	43	45	48

Требуется:

- а) оценить тесноту и направление связи между переменными с помощью коэффициента корреляции;
- б) построить уравнение регрессии y по x .

1.3. При исследовании корреляционной зависимости между ценой на нефть x и индексом нефтяных компаний y получены следующие данные:

$$\bar{x}=16,2 \text{ (ден. ед.)}, \bar{y} = 4000 \text{ (ден. ед.)}, s_x^2 = 4, s_y^2 = 500, \text{Cov}(X,Y)=40.$$

Необходимо:

- а) составить уравнение регрессии y по x ;
- б) используя уравнение регрессии, найти среднее значение индекса при цене на нефть 16,5 ден. ед.

1.4. По данным примера 1.1:

- а) найти уравнение регрессии y по x ;
- б) найти коэффициент детерминации и пояснить его смысл;
- в) проверить значимость уравнения регрессии на 5%-ном уровне по F -критерию;
- г) оценить среднюю производительность труда на предприятиях с уровнем механизации работ 60% и построить для нее 95%-ный доверительный интервал; аналогичный доверительный интервал найти для индивидуальных значений производительности труда на тех же предприятиях.

1.5. По данным 30 нефтяных компаний получено следующее уравнение регрессии между оценкой y (ден. ед.) и фактической стоимостью x (ден. ед.) этих компаний: $y_x = 0,8750x + 295$. Найти: 95%-ные доверительные интервалы для среднего и индивидуального значений оценки предприятий, фактическая стоимость которых составила 1300 ден. ед., если коэффициент корреляции между переменными равен 0,76, а среднее квадратическое отклонение переменной x равно 270 ден. ед.

1.6. Исследуя спрос на телевизоры марки N , аналитический отдел компании ABC по данным, собранным по 19 торговым точкам компании, выявил следующую зависимость:

$$\ln y = 10,5 - 0,8 \cdot \ln x + \varepsilon$$

где y - объем продаж телевизоров марки N в отдельной торговой точке; x - средняя цена телевизора в данной торговой точке;

Фактические значения t -критерия Стьюдента составляют: для параметра a уравнения регрессии – 2,5, для параметра b уравнения регрессии – -4.0 .

До проведения этого исследования администрация компании предполагала, что эластичность спроса по цене для телевизоров марки N составляет -0,9. Подтвердилось ли предположение администрации результатами исследования?

1.7. Для трех видов продукции A , B и C модели зависимости удельных постоянных расходов от объема выпускаемой продукции выглядят следующим образом:

$$y_A = 600,$$

$$y_B = 80 + 0,7x,$$

$$y_C = 40x^{0,5}$$

Требуется:

1. Определить коэффициенты эластичности по каждому виду продукции и пояснить их смысл.
2. Сравнить при $x = 1000$ эластичность затрат для продукции B и C .
3. Определить, каким должен быть объем выпускаемой продукции, чтобы коэффициенты эластичности для продукции B и C были равны.

1.8. Пусть имеется следующая модель регрессии, характеризующая зависимость y от x :

$$y = 8 - 7x + \varepsilon$$

Известно также, что $r_{xy} = -0,5$; $n = 20$.

Требуется:

1. Построить доверительный интервал для коэффициента регрессии в этой модели:
 - а) с вероятностью 90%;
 - б) с вероятностью 99%.
2. Проанализировать результаты, полученные в п.1, и пояснить причины их различий.

1.9. Изучается зависимость потребления материалов y от объема производства продукции x . По 20 наблюдениям были получены следующие варианты уравнения регрессии:

1. $y = 3 + 2x + \varepsilon$.
(6,48)

2. $\ln y = 2,5 + 0,2 \cdot \ln x + \varepsilon$, $r^2 = 0,68$.
(6,19)

3. $\ln Y = 1,1 + 0,8 \ln X + \varepsilon$, $r^2 = 0,69$.

$$4. Y = 3 + 1,5 \cdot X + 0,1 \cdot X^2, \quad r^2 = 0,701. \quad (6.2)$$

(3,0) (2.65)

В скобках указаны фактические значения t -критерия.

Требуется:

1. Определить коэффициент детерминации для 1-го уравнения;
2. Записать функции, характеризующие зависимость y от x во 2-м и 3-м уравнениях;
3. Определить коэффициенты эластичности для каждого из уравнений.
4. Выбрать наилучший вариант уравнения регрессии.

1.10. По совокупности 30 предприятий торговли изучается зависимость между признаками: x - цена на товар A , тыс. руб.; y - прибыль торгового предприятия, млн руб.

При оценке регрессионной модели были получены следующие промежуточные результаты:

$$\sum (y_j - \hat{y}_x)^2 = 39\,000;$$

$$\sum (y_j - \bar{y})^2 = 120\,000.$$

Требуется:

1. Пояснить, какой показатель корреляции можно определить по этим данным.
2. Построить таблицу дисперсионного анализа для расчета значения F -критерия Фишера.
3. Сравнить фактическое значение F -критерия с табличным. Сделать выводы.

1.11. Зависимость среднемесячной производительности труда от возраста рабочих характеризуется моделью: $y = a + b_1x + b_2x^2$. Ее использование привело к результатам, представленным в табл. 1.7.

Таблица 1.7

№ п/п	Производительность труда рабочих, тыс. руб., y		№ п/п	Производительность труда рабочих, тыс. руб., y	
	фактическая	расчетная		фактическая	расчетная
1	12	10	6	11	12
2	8	10	7	12	13
3	13	13	8	9	10
4	15	14	9	11	10
5	16	15	10	9	9

Требуется оценить качество модели, определив ошибку аппроксимации, индекс корреляции и F -критерий Фишера.

1.12. Моделирование прибыли фирмы по уравнению $y = ab^x$ привело к результатам, представленным в табл. 1.8.

Таблица 1.8.

№ п/п	Прибыль фирмы, тыс. руб., y		№ п/п	Прибыль фирмы, тыс. руб., y	
	фактическая	расчетная		фактическая	расчетная
1	10	11	5	18	20
2	12	11	6	11	И
3	15	17	7	13	14
4	17	15	8	19	16

Требуется оценить качество модели. Для этого необходимо:

- определить ошибку аппроксимации;
- рассчитать показатель тесноты связи прибыли с исследуемым в модели фактором;
- рассчитать F -критерий Фишера. Сделать выводы.

1.13. Изучалась зависимость вида $y = ax^b$. Для преобразованных в логарифмах переменных получены следующие данные:

$$\sum x = 8,2370$$

$$\sum x^2 = 9,2334$$

$$\sum y = 3,9310$$

$$\sum xy = 4,2087$$

$$\sum (Y - \hat{Y}_x)^2 = 0,0014.$$

Требуется:

- Найти параметр b .
- Найти показатель корреляции, предполагая $\sigma_y = 0,08$. Оценить его значимость.
- Оценить его значимость, если известно, что $n = 9$.

1.14. Зависимость объема производства y (тыс. ед.) от численности занятых x (чел.) по 15 заводам концерна характеризуется следующим образом:

$$\text{Уравнение регрессии} - y = 30 - 0,4x + 0,04x^2;$$

Доля остаточной дисперсии в общей – 20%.

Требуется определить::

- индекс корреляции;

- б) значимость уравнения регрессии;
 в) коэффициент эластичности, предполагая, что численность занятых составляет 30 человек.

РАЗДЕЛ II. МНОЖЕСТВЕННАЯ РЕГРЕССИЯ И КОРРЕЛЯЦИЯ

Множественная регрессия - это уравнение связи с несколькими независимыми переменными:

$$y = f(x_1, x_2, \dots, x_m) + \varepsilon, \quad (2.1)$$

где y - зависимая переменная (результативный признак); x_1, x_2, \dots, x_m - независимые переменные (признаки-факторы).

Для построения уравнения множественной регрессии чаще используются следующие функции:

• линейная – $y = a + b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_m \cdot x_m + \varepsilon$;

• степенная – $y = a \cdot x_1^{b_1} \cdot x_2^{b_2} \cdot \dots \cdot x_m^{b_m} \cdot \varepsilon$;

• экспонента – $y = e^{a + b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_m \cdot x_m + \varepsilon}$

• гипербола – $y = \frac{1}{a + b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_m \cdot x_m + \varepsilon}$

Можно использовать и другие функции, приводимые к линейному виду.

Для оценки параметров уравнения множественной регрессии применяют *метод наименьших квадратов* (МНК). Для линейных уравнений

$$y = a + b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_m \cdot x_m + \varepsilon \quad (2.2)$$

строится следующая система нормальных уравнений, решение которой позволяет получить оценки параметров регрессии:

сии β_i можно сравнивать между собой. Сравнивая их друг с другом, можно ранжировать факторы по силе их воздействия на результат. В этом основное достоинство стандартизованных коэффициентов регрессии в отличие от коэффициентов «чистой» регрессии, которые несравнимы между собой.

Применяя МНК к уравнению множественной регрессии в стандартизованном масштабе, получим систему нормальных уравнений вида

$$\begin{cases} r_{yx_1} = \beta_1 + \beta_2 r_{x_1 x_2} + \beta_3 r_{x_1 x_3} + \dots + \beta_m r_{x_1 x_m} \\ r_{yx_2} = \beta_1 r_{x_1 x_2} + \beta_2 + \beta_3 r_{x_1 x_3} + \dots + \beta_m r_{x_1 x_m} \\ \dots \\ r_{yx_m} = \beta_1 r_{x_1 x_m} + \beta_2 r_{x_2 x_m} + \beta_3 r_{x_3 x_m} + \dots + \beta_m \end{cases} \quad (2.8)$$

где r_{yx_i} и $r_{x_i x_j}$ - коэффициенты парной и межфакторной корреляции.

Коэффициенты «чистой» регрессии b_i связаны со стандартизованными коэффициентами регрессии β_i следующим образом:

$$b_i = \beta_i \frac{\sigma_y}{\sigma_{x_i}} \quad (\beta_i = b_i \frac{\sigma_{x_i}}{\sigma_y}). \quad (2.9)$$

Поэтому можно переходить от уравнения регрессии в стандартизованном масштабе (2.7) к уравнению регрессии в натуральном масштабе переменных (2.2), при этом параметр a определяется как

$$a = \bar{y} - b_1 \bar{x}_1 - b_2 \bar{x}_2 - \dots - b_m \bar{x}_m. \quad (2.10)$$

Рассмотренный смысл стандартизованных коэффициентов регрессии позволяет их использовать при отсеивании факторов - из модели исключаются факторы с наименьшим значением β_i .

Средние коэффициенты эластичности для линейной регрессии рассчитываются по формуле

$$\bar{\varepsilon}_{yx_j} = b_j \frac{\bar{x}_j}{\bar{y}}, \quad (2.11)$$

которые показывают на сколько процентов в среднем изменится результат, при изменении соответствующего фактора на 1%. Средние показатели эластичности можно сравнивать друг с другом и соответственно ранжировать факторы по силе их воздействия на результат.

Тесноту совместного влияния факторов на результат оценивает *индекс множественной корреляции*:

$$R_{yx_1x_2\dots x_m} = \sqrt{1 - \frac{\sigma_{y_{ост}}^2}{\sigma_y^2}}. \quad (2.12)$$

Значение индекса множественной корреляции лежит в пределах от 0 до 1 и должно быть больше или равно максимальному парному индексу корреляции:

$$R_{yx_1x_2\dots x_m} \geq r_{yx_i} \quad (i = \overline{1, m}). \quad (2.13)$$

При линейной зависимости *коэффициент множественной корреляции* можно определить через матрицы парных коэффициентов корреляции:

$$R_{yx_1x_2\dots x_m} = \sqrt{1 - \frac{\Delta r}{\Delta r_{11}}} \quad (2.14)$$

где

$$\Delta r = \begin{vmatrix} 1 & r_{yx_1} & r_{yx_2} & \dots & r_{yx_m} \\ r_{yx_1} & 1 & r_{x_1x_2} & \dots & r_{x_1x_m} \\ r_{yx_2} & r_{x_2x_1} & 1 & \dots & r_{x_2x_m} \\ \dots & \dots & \dots & \dots & \dots \\ r_{yx_m} & r_{x_mx_1} & r_{x_mx_2} & \dots & 1 \end{vmatrix} \quad (2.15)$$

- определитель матрицы парных коэффициентов корреляции;

$$\Delta r_{11} = \begin{vmatrix} 1 & r_{x_1x_2} & \dots & r_{x_1x_m} \\ r_{x_2x_1} & 1 & \dots & r_{x_2x_m} \\ \dots & \dots & \dots & \dots \\ r_{x_mx_1} & r_{x_mx_2} & \dots & 1 \end{vmatrix} \quad (2.16)$$

- определитель матрицы межфакторной корреляции.

Так же при линейной зависимости признаков формула коэффициента множественной корреляции может быть также представлена следующим выражением:

$$R_{yx_1x_2\dots x_m} = \sqrt{\sum \beta_i \cdot r_{yx_i}} \quad (2.17)$$

где β_i - стандартизованные коэффициенты регрессии; r_{yx_i} - парные коэффициенты корреляции результата с каждым фактором.

Качество построенной модели в целом оценивает коэффициент (индекс) детерминации. Коэффициент множественной детерминации рассчитывается как квадрат индекса множественной корреляции $R^2_{yx_1x_2\dots x_m}$.

Для того чтобы не допустить преувеличения тесноты связи, применяется скорректированный индекс множественной детерминации, который содержит поправку на число степеней свободы и рассчитывается по формуле

$$\hat{R}^2 = 1 - (1 - R^2) \frac{(n-1)}{(n-m-1)}, \quad (2.18)$$

где n - число наблюдений, m - число факторов.

При небольшом числе наблюдений нескорректированная величина коэффициента множественной детерминации R^2 имеет тенденцию переоценивать долю вариации результативного признака, связанную с влиянием факторов, включенных в регрессионную модель.

Частные коэффициенты (или индексы) корреляции, измеряющие влияние на y фактора x_i , при элиминировании (исключении влияния) других факторов, можно определить по формуле

$$r_{yx_1 \cdot x_2 \dots x_{i-1} x_{i+1} \dots x_m} = \sqrt{1 - \frac{1 - R_{yx_1 x_2 \dots x_i \dots x_m}^2}{1 - R_{yx_1 x_2 \dots x_{i-1} x_{i+1} \dots x_m}^2}}, \quad (2.19)$$

или по рекуррентной формуле:

$$\begin{aligned} & r_{yx_1 \cdot x_2 \dots x_{i-1} x_{i+1} \dots x_m} = \\ & = \frac{r_{yx_1 \cdot x_2 \dots x_{i-1} x_{i+1} \dots x_{m-1}} \cdot r_{yx_m \cdot x_1 x_2 \dots x_{m-1}} \cdot r_{x_1 x_m \cdot x_1 x_2 \dots x_{i-1} x_{i+1} \dots x_{m-1}}}{\sqrt{(1 - r_{yx_m \cdot x_1 x_2 \dots x_{m-1}}^2)(1 - r_{x_1 x_m \cdot x_1 x_2 \dots x_{i-1} x_{i+1} \dots x_{m-1}}^2)}}. \end{aligned} \quad (2.20)$$

Рассчитанные по рекуррентной формуле частные коэффициенты корреляции изменяются в пределах от -1 до +1, а по формулам через множественные коэффициенты детерминации - от 0 до 1. Сравнение их друг с другом позволяет ранжировать факторы по тесноте их связи с результатом. Частные коэффициенты корреляции дают меру тесноты связи каждого фактора с результатом в чистом виде.

При двух факторах формулы (2.19) и (2.20) примут вид:

$$\begin{aligned} r_{yx_1 \cdot x_2} &= \sqrt{1 - \frac{1 - R_{yx_1 x_2}^2}{1 - r_{yx_2}^2}}; & r_{yx_2 \cdot x_1} &= \sqrt{1 - \frac{1 - R_{yx_1 x_2}^2}{1 - r_{yx_1}^2}} \\ r_{yx_1 \cdot x_2} &= \frac{r_{yx_1} \cdot r_{yx_2} \cdot r_{x_1 x_2}}{\sqrt{(1 - r_{yx_2}^2) \cdot (1 - r_{x_1 x_2}^2)}}; & r_{yx_2 \cdot x_1} &= \frac{r_{yx_2} \cdot r_{yx_1} \cdot r_{x_1 x_2}}{\sqrt{(1 - r_{yx_1}^2) \cdot (1 - r_{x_1 x_2}^2)}} \end{aligned}$$

Значимость уравнения множественной регрессии в целом оценивается с помощью F -критерия Фишера:

$$F = \frac{R^2}{1 - R^2} \cdot \frac{n - m - 1}{m} \quad (2.21)$$

Частный F -критерий оценивает статистическую значимость присутствия каждого из факторов в уравнении. В общем виде для фактора x частный F -критерий определится как

$$F_{x_i} = \frac{R_{yx_1 \dots x_i \dots x_m}^2 - R_{yx_1 \dots x_{i-1} x_{i+1} \dots x_m}^2}{1 - R_{yx_1 \dots x_i \dots x_m}^2} \cdot \frac{n - m - 1}{1} \quad (2.22)$$

Фактическое значение частного F -критерия сравнивается с табличным при уровне значимости α и числе степеней свободы: $k_1 = 1$ и $k_2 = n - m - 1$. Если фактическое значение F_{xi} превышает $F_{табл}(\alpha, k_1, k_2)$, то дополнительное включение фактора x_i в модель статистически оправданно и коэффициент чистой регрессии b_i при факторе x_i статистически значим.

Если же фактическое значение F_{xi} меньше табличного, то дополнительное включение в модель фактора x_i не увеличивает существенно долю объясненной вариации признака y , следовательно, нецелесообразно его включение в модель; коэффициент регрессии при данном факторе в этом случае статистически незначим.

Оценка значимости коэффициентов чистой регрессии проводится по t -критерию Стьюдента. В этом случае, как и в парной регрессии, для каждого фактора используется формула

$$t_{b_i} = \frac{b_i}{m_{b_i}} \quad (2.23)$$

Для уравнения множественной регрессии (3.1) средняя квадратическая ошибка коэффициента регрессии может быть определена по формуле:

$$m_{b_i} = \frac{\sigma_y \cdot \sqrt{1 - R_{yx_1 \dots x_m}^2}}{\sigma_{x_i} \cdot \sqrt{1 - R_{x_i x_1 \dots x_m}^2}} \cdot \frac{1}{\sqrt{n - m - 1}}, \quad (2.24)$$

где $R_{x_i x_1 \dots x_m}^2$ - коэффициент детерминации для зависимости фактора x_i со всеми другими факторами уравнения множественной регрессии. Для двухфакторной модели ($m = 2$) имеем:

$$m b_1 = \frac{\sigma_y \cdot \sqrt{1 - R_{yx_1 x_2}^2}}{\sigma_{x_1} \cdot \sqrt{1 - r_{x_1 x_2}^2}} \cdot \frac{1}{\sqrt{n-3}} \quad m b_2 = \frac{\sigma_y \cdot \sqrt{1 - R_{yx_1 x_2}^2}}{\sigma_{x_2} \cdot \sqrt{1 - r_{x_1 x_2}^2}} \cdot \frac{1}{\sqrt{n-3}}$$

Существует связь между t -критерием Стьюдента и частным F -критерием Фишера:

$$\left| t_{b_i} \right| = \sqrt{F_{x_i}} \quad (2.25)$$

Уравнения множественной регрессии могут включать в качестве независимых переменных качественные признаки (например, профессия, пол, образование, климатические условия, отдельные регионы и т.д.). Чтобы ввести такие переменные в регрессионную модель, их необходимо упорядочить и присвоить им те или иные значения, т.е. качественные переменные преобразовать в количественные.

Такого вида сконструированные переменные принято в эконометрике называть *фиктивными переменными*. Например, включать в модель фактор «пол» в виде фиктивной переменной можно в следующем виде:

$$z = \begin{cases} 1 - \text{мужской пол} \\ 0 - \text{женский пол} \end{cases}$$

Коэффициент регрессии при фиктивной переменной интерпретируется как среднее изменение зависимой переменной при переходе от одной категории (женский пол) к другой (мужской пол) при неизменных значениях остальных параметров.

РЕШЕНИЕ ТИПОВЫХ ЗАДАЧ

Задача 1.

По 20 предприятиям региона изучается зависимость выработки продукции на одного работника y (тыс. руб.) от ввода в действие но-

вых основных фондов x_1 (% от стоимости фондов на конец года) и от удельного веса рабочих высокой квалификации в общей численности рабочих x_2 (%).

Номер предприятия	y	x_1	x_2	Номер предприятия	y	x_1	x_2
1	7,0	3,9	10,0	11	9,0	6,0	21,0
2	7,0	3,9	14,0	12	11,0	6,4	22,0
3	7,0	3,7	15,0	13	9,0	6,8	22,0
4	7,0	4,0	16,0	14	11,0	7,2	25,0
5	7,0	3,8	17,0	15	12,0	8,0	28,0
6	7,0	4,8	19,0	16	12,0	8,2	29,0
7	8,0	5,4	19,0	17	12,0	8,1	30,0
8	8,0	4,4	20,0	18	12,0	8,5	31,0
9	8,0	5,3	20,0	19	14,0	9,6	32,0
10	10,0	6,8	20,0	20	14,0	9,0	36,0

Требуется:

1. Построить линейную модель множественной регрессии. Записать стандартизованное уравнение множественной регрессии. На основе стандартизованных коэффициентов регрессии и средних коэффициентов эластичности ранжировать факторы по степени их влияния на результат.

2. Найти коэффициенты парной, частной и множественной корреляции. Проанализировать их.

3. Найти скорректированный коэффициент множественной детерминации. Сравнить его с нескорректированным (общим) коэффициентом детерминации.

4. С помощью F -критерия Фишера оценить статистическую надежность уравнения регрессии и коэффициента детерминации R^2_y .

5. С помощью t -критерия оценить статистическую значимость коэффициентов чистой регрессии.

6. С помощью частных F -критериев Фишера оценить целесообразность включения в уравнение множественной регрессии фактора x_1 после x_2 и фактора x_2 после x_1 .

7. Составить уравнение линейной парной регрессии, оставив лишь один значащий фактор.

Решение:

Для удобства проведения расчетов поместим результаты промежуточных расчетов в таблицу:

Таблица 2.1.

№	y	x ₁	x ₂	yx ₁	yx ₂	x ₁ x ₂	x ₁ ²	x ₂ ²	y ²
1	7,0	3,9	10,0	27,3	70,0	39,0	15,21	100,0	49,0
2	7,0	3,9	14,0	27,3	98,0	54,6	15,21	196,0	49,0
3	7,0	3,7	15,0	25,9	105,0	55,5	13,69	225,0	49,0
4	7,0	4,0	16,0	28,0	112,0	64,0	16,0	256,0	49,0
5	7,0	3,8	17,0	26,6	119,0	64,6	14,44	289,0	49,0
6	7,0	4,8	19,0	33,6	133,0	91,2	23,04	361,0	49,0
7	8,0	5,4	19,0	43,2	152,0	102,6	29,16	361,0	64,0
8	8,0	4,4	20,0	35,2	160,0	88,0	19,36	400,0	64,0
9	8,0	5,3	20,0	42,4	160,0	106,0	28,09	400,0	64,0
10	10,0	6,8	20,0	68,0	200,0	136,0	46,24	400,0	100,0
11	9,0	6,0	21,0	54,0	189,0	126,0	36,0	441,0	81,0
12	11,0	6,4	22,0	70,4	242,0	140,8	40,96	484,0	121,0
13	9,0	6,8	22,0	61,2	198,0	149,6	46,24	484,0	81,0
14	11,0	7,2	25,0	79,2	275,0	180,0	51,84	625,0	121,0
15	12,0	8,0	28,0	96,0	336,0	224,0	64,0	784,0	144,0
16	12,0	8,2	29,0	98,4	348,0	237,8	67,24	841,0	144,0
17	12,0	8Д	30,0	97,2	360,0	243,0	65,61	900,0	144,0
18	12,0	8,5	31,0	102,0	372,0	263,5	72,25	961,0	144,0
19	14,0	9,6	32,0	134,4	448,0	307,2	92,16	1024,0	196,0
20	14,0	9,0	36,0	126,0	504,0	324,0	81,0	1296,0	196,0
Сумма	192	123,8	446	1276,3	4581	2997,4	837,74	10828,0	1958,0
Ср.знач.	9,6	6,19	22,3	63,815	229,05	149,87	41,887	541,4	97,9

Найдем средние квадратические отклонения признаков:

$$\sigma_y = \sqrt{y^2 - \bar{y}^2} = \sqrt{97,9 - 9,6^2} = \sqrt{5,74} = 2,396;$$

$$\sigma_{x_1} = \sqrt{x_1^2 - \bar{x}_1^2} = \sqrt{41,887 - 6,19^2} = \sqrt{3,571} = 1,890;$$

$$\sigma_{x_2} = \sqrt{x_2^2 - \bar{x}_2^2} = \sqrt{541,4 - 22,3^2} = \sqrt{44,11} = 6,642.$$

1. Для нахождения параметров линейного уравнения множественной регрессии необходимо решить систему линейных уравнений относительно неизвестных параметров a , b_1 , b_2 либо воспользоваться готовыми формулами (2.4-2.6).

Рассчитаем сначала парные коэффициенты корреляции:

$$r_{yx_1} = \frac{\text{cov}(y, x_1)}{\sigma_y \cdot \sigma_{x_1}} = \frac{63,815 - 6,19 \cdot 9,6}{1,890 \cdot 2,396} = 0,970;$$

$$r_{yx_2} = \frac{\text{cov}(y, x_2)}{\sigma_y \cdot \sigma_{x_2}} = \frac{229,05 - 22,3 \cdot 9,6}{6,642 \cdot 2,396} = 0,941;$$

$$r_{x_1 x_2} = \frac{\text{cov}(x_1, x_2)}{\sigma_{x_1} \cdot \sigma_{x_2}} = \frac{149,87 - 6,19 \cdot 22,3}{1,890 \cdot 6,642} = 0,943.$$

Находим по формулам (3.4) коэффициенты чистой регрессии и параметр a :

$$b_1 = \frac{\sigma_y}{\sigma_{x_1}} \cdot \frac{r_{yx_1} - r_{yx_2} \cdot r_{x_1 x_2}}{1 - r_{x_1 x_2}^2} = \frac{2,396}{1,890} \cdot \frac{0,970 - 0,941 \cdot 0,943}{1 - 0,943^2} = 0,946;$$

$$b_2 = \frac{\sigma_y}{\sigma_{x_2}} \cdot \frac{r_{yx_2} - r_{yx_1} \cdot r_{x_1 x_2}}{1 - r_{x_1 x_2}^2} = \frac{2,396}{6,642} \cdot \frac{0,941 - 0,970 \cdot 0,943}{1 - 0,943^2},$$

$$a = \bar{y} - b_1 \bar{x}_1 - b_2 \bar{x}_2 = 9,6 - 0,946 \cdot 6,19 - 0,0856 \cdot 22,3 = 1,835.$$

Таким образом, получили следующее уравнение множественной регрессии:

$$\hat{y} = 1,835 + 0,946 \cdot x_1 + 0,0856 \cdot x_2$$

Уравнение регрессии показывает, что при увеличении ввода в действие основных фондов на 1% (при неизменном уровне удельного веса рабочих высокой квалификации) выработка продукции на одного рабочего увеличивается в среднем на 0,946 тыс. руб., а при увеличении удельного веса рабочих высокой квалификации в общей численности рабочих на 1% (при неизменном уровне ввода в действие новых основных фондов) выработка продукции на одного рабочего увеличивается в среднем на 0,086 тыс. руб.

После нахождения уравнения регрессии составим новую расчетную таблицу для определения теоретических значений результативного признака, остаточной дисперсии и средней ошибки аппроксимации.

Таблица 2.2

№	y	x_1	x_2	\hat{y}	$y - \hat{y}$	$(y - \hat{y})^2$	$A_i, \%$
1	7,0	3,9	10,0	6,380	0,620	0,384	8,851
2	7,0	3,9	14,0	6,723	0,277	0,077	3,960
3	7,0	3,7	15,0	6,619	0,381	0,145	5,440
4	7,0	4,0	16,0	6,989	0,011	0,000	0,163
5	7,0	3,8	17,0	6,885	0,115	0,013	1,643
6	7,0	4,8	19,0	8,002	-1,002	1,004	14,317
7	8,0	5,4	19,0	8,570	-0,570	0,325	7,123
8	8,0	4,4	20,0	7,709	0,291	0,084	3,633
9	8,0	5,3	20,0	8,561	-0,561	0,315	7,010
10	10,0	6,8	20,0	9,980	0,020	0,000	0,202
11	9,0	6,0	21,0	9,309	-0,309	0,095	3,429
12	11,0	6,4	22,0	9,773	1,227	1,507	11,158
13	9,0	6,8	22,0	10,151	-1,151	1,325	12,789
14	11,0	7,2	25,0	10,786	0,214	0,046	1,944
15	12,0	8,0	28,0	11,800	0,200	0,040	1,668
16	12,0	8,2	29,0	12,075	-0,075	0,006	0,622
17	12,0	8,5	30,0	12,066	-0,066	0,004	0,547
18	12,0	8,5	31,0	12,530	-0,530	0,280	4,413
19	14,0	9,6	32,0	13,656	0,344	0,118	2,459
20	14,0	9,0	36,0	13,431	0,569	0,324	4,067
Сумма	192	123,8	446	191,992	0,008	6,093	95,437
Ср. знач.	9,6	6,19	22,3	9,6	—	0,305	4,77

Остаточная дисперсия:

$$\sigma_{ост}^2 = \frac{\sum (y - \hat{y})^2}{n} = \frac{6,093}{20} = 0,305.$$

Средняя ошибка аппроксимации:

$$\bar{A} = \frac{1}{n} \sum \left| \frac{y - \hat{y}}{y} \right| \cdot 100\% = \frac{95,437\%}{20} = 4,77\%.$$

Качество модели, исходя из относительных отклонений по каждому наблюдению, признается хорошим, т.к. средняя ошибка аппроксимации не превышает 10%.

Коэффициенты β_1 и β_2 стандартизованного уравнения регрессии

$\hat{t}_y = \beta_1 t_{x_1} + \beta_2 t_{x_2} + \varepsilon$ находятся по формуле (2.9):

$$\beta_1 = b_1 \frac{\sigma_{x_1}}{\sigma_y} = 0,946 \cdot \frac{1,890}{2,396} = 0,746;$$

$$\beta_2 = b_2 \frac{\sigma_{x_2}}{\sigma_y} = 0,0856 \cdot \frac{6,642}{2,396} = 0,237.$$

Т.е. уравнение будет выглядеть следующим образом:
 $\hat{t}_y = 0,746t_{x_1} + 0,237t_{x_2}$

Так как стандартизованные коэффициенты регрессии можно сравнивать между собой, то можно сказать, что ввод в действие новых основных фондов оказывает большее влияние на выработку продукции, чем удельный вес рабочих высокой квалификации.

Сравнивать влияние факторов на результат можно также при помощи средних коэффициентов эластичности (3.8). Вычисляем:

$$\bar{\varepsilon}_1 = 0,946 \cdot \frac{6,19}{9,6} = 0,61; \quad \bar{\varepsilon}_2 = 0,0856 \cdot \frac{22,3}{9,6} = 0,20$$

Т.е. увеличение только основных фондов (от своего среднего значения) или только удельного веса рабочих высокой квалификации на 1% увеличивает в среднем выработку продукции на 0,61% или 0,20% соответственно. Таким образом, подтверждается большее влияние на результат у фактора x_1 , чем фактора x_2 .

2. Коэффициенты парной корреляции найдены ранее:

$$r_{yx_1} = 0,970; \quad r_{yx_2} = 0,941; \quad r_{x_1x_2} = 0,943.$$

Они указывают на весьма сильную связь каждого фактора с результатом, а также высокую межфакторную зависимость (факторы x_1 и x_2 явно коллинеарны, т.к. $r_{x_1x_2} = 0,943 > 0,7$). При такой

сильной межфакторной зависимости рекомендуется один из факторов исключить из рассмотрения.

Частные коэффициенты корреляции характеризуют тесноту связи между результатом и соответствующим фактором при элиминировании (устранении влияния) других факторов, включенных в уравнение регрессии.

При двух факторах частные коэффициенты корреляции рассчитываются следующим образом:

$$r_{yx_1 \cdot x_2} = \frac{r_{yx_1} - r_{yx_2} \cdot r_{x_1 x_2}}{\sqrt{(1 - r_{yx_2}^2) \cdot (1 - r_{x_1 x_2}^2)}} = \frac{0,970 - 0,941 \cdot 0,943}{\sqrt{(1 - 0,941^2) \cdot (1 - 0,943^2)}} = 0,734$$

$$r_{yx_2 \cdot x_1} = \frac{r_{yx_2} - r_{yx_1} \cdot r_{x_1 x_2}}{\sqrt{(1 - r_{yx_1}^2) \cdot (1 - r_{x_1 x_2}^2)}} = \frac{0,941 - 0,970 \cdot 0,943}{\sqrt{(1 - 0,970^2) \cdot (1 - 0,943^2)}} = 0,325.$$

Если сравнить коэффициенты парной и частной корреляции, то можно увидеть, что из-за высокой межфакторной зависимости коэффициенты парной корреляции дают завышенные оценки тесноты связи. Именно по этой причине рекомендуется при наличии сильной коллинеарности (взаимосвязи) факторов исключать из исследования тот фактор, у которого теснота парной зависимости меньше, чем теснота межфакторной связи.

Коэффициент множественной корреляции определить через матрицы парных коэффициентов корреляции (3.9):

$$R_{yx_1 x_2} = \sqrt{1 - \frac{\Delta r}{\Delta r_{11}}} \quad (2.26)$$

где $\Delta r = \begin{vmatrix} 1 & r_{yx_1} & r_{yx_2} \\ r_{yx_1} & 1 & r_{x_1 x_2} \\ r_{yx_2} & r_{x_2 x_1} & 1 \end{vmatrix}$

- определитель матрицы парных коэффициентов корреляции;

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

- определитель матрицы межфакторной корреляции.

Находим:

$$\Delta r = \begin{vmatrix} 1 & 0,970 & 0,941 \\ 0,970 & 1 & 0,943 \\ 0,941 & 0,943 & 1 \end{vmatrix} = 1 + 0,8607 + 0,8607 - 0,8855 - 0,8892 - 0,9409 = 0,0058$$

$$\Delta r_{11} = \begin{vmatrix} 1 & 0,943 \\ 0,943 & 1 \end{vmatrix} = 1 - 0,8892 = 0,1108.$$

Коэффициент множественной корреляции:

$$R_{yx_1x_2} = \sqrt{1 - \frac{0,0058}{0,1108}} = 0,973.$$

Аналогичный результат получим при использовании формул (3.8) и (3.10):

$$R_{yx_1x_2} = \sqrt{1 - \frac{\sigma_{ост}^2}{\sigma_y^2}} = \sqrt{1 - \frac{0,305}{5,74}} = 0,973;$$

$$R_{yx_1x_2} = \sqrt{\sum \beta_i \cdot r_{yx_i}} = \sqrt{0,746 \cdot 0,970 + 0,237 \cdot 0,941} = 0,973;$$

Коэффициент множественной корреляции указывает на весьма сильную связь всего набора факторов с результатом.

3. Нескорректированный коэффициент множественной детерминации $R^2_{yx_1x_2} = 0,947$ оценивает долю дисперсии результата за счет представленных в уравнении факторов в общей вариации результата. Здесь эта доля составляет 94,7% и указывает на весьма высокую степень обусловленности вариации результата вариацией факторов, иными словами - на весьма тесную связь факторов с результатом.

Скорректированный коэффициент множественной детерминации

$$\hat{R}^2 = 1 - (1 - R^2) \cdot \frac{(n-1)}{(n-m-1)} = 1 - (1 - 0,947) \cdot \frac{20-1}{20-2-1} = 0,941$$

определяет тесноту связи с учетом степеней свободы общей и остаточной дисперсий. Он дает такую оценку тесноты связи, которая не зависит от числа факторов и поэтому может сравниваться по разным моделям с разным числом факторов. Оба коэффициента указывают на весьма высокую (более 94%) детерминированность результата y в модели факторами x_1 и x_2 .

4. Оценку надежности уравнения регрессии в целом и показателя тесноты связи $R_{yx_1x_2}$ дает F -критерий Фишера. В нашем случае фактическое значение F -критерия Фишера:

$$F = \frac{R^2}{1-R^2} \cdot \frac{n-m-1}{m} = \frac{0,973^2}{1-0,973^2} \cdot \frac{20-2-1}{2} = 151,88.$$

Получили, что $F_{факт} = 151,88 > F_{табл} = 3,59$ (при $n = 20$), т.е. вероятность случайно получить такое значение F -критерия не превышает допустимый уровень значимости 5%. Следовательно, полученное значение не случайно, оно сформировалось под влиянием существен-

ных факторов, т.е. подтверждается статистическая значимость всего уравнения и показателя тесноты связи $R_{yx_1x_2}$.

5. Оценим статистическую значимость параметров чистой регрессии с помощью t -критерия Стьюдента. Рассчитаем стандартные ошибки коэффициентов регрессии по формулам (3.19) и (3.20):

$$m_{b_1} = \frac{\sigma_y \cdot \sqrt{1 - R_{yx_1x_2}^2}}{\sigma_{x_1} \cdot \sqrt{1 - r_{x_1x_2}^2}} \cdot \frac{1}{\sqrt{n-3}} = \frac{2,396 \cdot \sqrt{1 - 0,973^2}}{1,890 \cdot \sqrt{1 - 0,943^2}} \cdot \frac{1}{\sqrt{20-3}} = 0,2132;$$

$$m_{b_2} = \frac{\sigma_y \cdot \sqrt{1 - R_{yx_1x_2}^2}}{\sigma_{x_2} \cdot \sqrt{1 - r_{x_1x_2}^2}} \cdot \frac{1}{\sqrt{n-3}} = \frac{2,396 \cdot \sqrt{1 - 0,973^2}}{6,642 \cdot \sqrt{1 - 0,943^2}} \cdot \frac{1}{\sqrt{20-3}} = 0,0607.$$

Фактические значения t -критерия Стьюдента:

$$t_{b_1} = \frac{b_1}{m_{b_1}} = \frac{0,946}{0,2132} = 4,44; \quad t_{b_2} = \frac{b_2}{m_{b_2}} = \frac{0,0856}{0,0607} = 1,41.$$

Табличное значение критерия при уровне значимости $\alpha = 0,05$ и числе степеней свободы $k = 17$ составит $t_{табл}(\alpha = 0,05; k = 17) = 2,11$. Таким образом, признается статистическая значимость параметра b_1 , т.к. $t_{b_1} > t_{табл}$, и случайная природа формирования параметра b_2 , т.к. $t_{b_2} < t_{табл}$.

Доверительные интервалы для параметров чистой регрессии:

$$b_1 - m_{b_1} \cdot t_{табл} \leq b_1^* \leq b_1 + m_{b_1} \cdot t_{табл}, \quad 0,496 \leq b_1^* \leq 1,396;$$

$$b_2 - m_{b_2} \cdot t_{табл} \leq b_2^* \leq b_2 + m_{b_2} \cdot t_{табл}, \quad -0,0425 \leq b_2^* \leq 0,2137$$

6. С помощью частных F -критериев Фишера оценим целесообразность включения в уравнение множественной регрессии фактора x_1 после x_2 и фактора x_2 после x_1 при помощи формул (3.16):

$$F_{x_1} = \frac{R_{yx_1x_2}^2 - R_{yx_2}^2}{1 - R_{yx_1x_2}^2} \cdot \frac{n-m-1}{1}; \quad F_{x_2} = \frac{R_{yx_1x_2}^2 - R_{yx_1}^2}{1 - R_{yx_1x_2}^2} \cdot \frac{n-m-1}{1}$$

Найдем $R_{yx_1}^2$ и $R_{yx_2}^2$:

$$R_{yx_1}^2 = r_{yx_1}^2 = 0,970^2 = 0,941;$$

$$R_{yx_2}^2 = r_{yx_2}^2 = 0,941^2 = 0,885.$$

Имеем:

$$F_{x_1} = \frac{0,947 - 0,885}{1 - 0,947} \cdot \frac{20 - 2 - 1}{1} = 19,89;$$

$$F_{x_2} = \frac{0,947 - 0,941}{1 - 0,947} \cdot \frac{20 - 2 - 1}{1} = 1,924.$$

Получили, что $F_{x_2} = 0,89 < F_{\text{табл}} (\alpha = 0,05; k_1=1; k_2=17) = 4,45$. Следовательно, включение в модель фактора x_2 после того, как в модель включен фактор x_1 статистически нецелесообразно: прирост факторной дисперсии за счет дополнительного признака x_2 оказывается незначительным, несущественным; фактор x_2 включать в уравнение после фактора x_1 не следует.

Если поменять первоначальный порядок включения факторов в модель и рассмотреть вариант включения x_1 после x_2 , то результат расчета частного F -критерия для x_1 будет иным. $F_{x_1} = 17,86 > F_{\text{табл}} = 4,45$, т.е. вероятность его случайного формирования меньше принятого стандарта $\alpha = 0,05$ (5%). Следовательно, значение частного F -критерия для дополнительно включенного фактора x_1 не случайно, является статистически значимым, надежным, достоверным: прирост факторной дисперсии за счет дополнительного фактора x_1 является существенным.

Фактор x_1 должен присутствовать в уравнении, в том числе в варианте, когда он дополнительно включается после фактора x_2 .

7. Общий вывод состоит в том, что множественная модель с факторами x_1 и x_2 с $R^2_{yx_1x_2} = 0,947$ содержит неинформативный фактор x_2 .

Если исключить фактор x_2 , то можно ограничиться уравнением парной регрессии:

$$\hat{y}_{x_1} = \alpha + \beta x_1.$$

Найдем его параметры:

$$\beta = \frac{\text{cov}(y, x_1)}{\sigma_{x_2}} = \frac{63,815 - 6,19 \cdot 9,6}{3,571} = 1,23 ;$$

$$\alpha = \bar{y} - \beta \cdot \bar{x} = 9,6 - 1,23 \cdot 6,19 = 1,99.$$

Таким образом [4],

$$\hat{y}_{x_1} = 1,99 + 1,23 \cdot x_1, \quad r_{yx_1}^2 = 0,941.$$

Задача 2.

Зависимость спроса на свинину x_1 от цены на нее x_2 и от цены на говядину x_3 представлена уравнением:

$$\lg x_1 = 0,1274 - 0,2143 * \lg x_2 + 2,8254 * \lg x_3.$$

Требуется:

1. Представить данное уравнение в естественной форме (не в логарифмах).
2. Оценить значимость параметров данного уравнения, если известно, что t -критерий для параметра b_2 при x_2 составил 0,827, а для параметра b_3 при x_3 - 1,015.

Решение:

1. Представленное степенное уравнение множественной регрессии приводим к естественной форме путём потенцирования обеих частей уравнения:

$$x_1 = 10^{0,1274} \cdot x_2^{-0,2143} \cdot x_3^{2,8254};$$

$$x_1 = 1,3409 \cdot \frac{1}{x_2^{0,2143}} \cdot x_3^{2,8254}.$$

Значения коэффициентов регрессии b_1 и b_2 в степенной функции равны коэффициентам эластичности результата x_1 от x_2 и x_3 .

$$\bar{\varepsilon}_{x_1 x_2} = -0,2143\%; \quad \bar{\varepsilon}_{x_1 x_3} = 2,8254\%.$$

Спрос на свинину x_1 сильнее связан с ценой на говядину - он увеличивается в среднем на 2,33% при росте цен на 1%. С ценой на свинину спрос на нее связан обратной зависимостью: с ростом цен на 1% потребление снижается в среднем на 0,21%.

2. Табличное значение t -критерия для $\alpha = 0,05$ обычно лежит в интервале 2-3 – в зависимости от степеней свободы. В данном примере $t_{b_2} = 0,827, t_{b_3} = 1,015$. Это весьма небольшие значения t -критерия, которые свидетельствуют о случайной природе взаимосвязи, о статистической ненадежности всего уравнения, поэтому применять полученное уравнение для прогноза не рекомендуется.

Задача 3.

По 30 территориям России имеются данные, представленные в табл. 2.1.

Таблица 2.1

Признак	Среднее значение	Среднее квадратическое отклонение	Линейный коэффициент парной корреляции
Среднедневной душевой доход, руб., y	86,8	11,44	-
Среднедневная заработная плата одного работающего, руб., x_1	54,9	5,86	$r_{yx_1} = 0,8405$
Средний возраст безработного, лет, x_2	33,5	0,58	$r_{yx_1} = -0,2101$ $r_{x_1x_2} = -0,1160$

Требуется:

1. Построить уравнение множественной регрессии в стандартизованной и естественной форме; рассчитать частные коэффициенты эластичности, сравнить их с β_1 и β_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 частные F -критерии Фишера.

Решение:

1. Линейное уравнение множественной регрессии y от x_1 и x_2 имеет вид: $y = a + b_1x_1 + b_2x_2$. Для расчета его параметров применим метод стандартизации переменных и построим искомое уравнение в стандартизованном масштабе: $t_y = \beta_1 \cdot t_{x_1} + \beta_2 \cdot t_{x_2}$.

Расчет β -коэффициентов выполним по формулам:

$$\beta_1 = \frac{r_{yx_1} - r_{yx_2} \cdot r_{x_1x_2}}{1 - r_{x_1x_2}^2} = \frac{0,8405 - 0,2101 \cdot 0,116}{1 - 0,116^2} = \frac{0,8161}{0,9865} = 0,8273;$$

$$\beta_2 = \frac{r_{yx_2} - r_{yx_1} \cdot r_{x_1x_2}}{1 - r_{x_1x_2}^2} = \frac{-0,2101 + 0,8405 \cdot 0,116}{1 - 0,116^2} = \frac{-0,1126}{0,9865} = -0,1141$$

Получим уравнение

$$t_y = 0,8273 \cdot t_{x_1} - 0,1141 \cdot t_{x_2}.$$

Для построения уравнения в естественной форме рассчитаем b_1 и b_2 , используя формулы для перехода от β_i и b_i :

$$b_i = \beta_i \frac{\sigma_y}{\sigma_{x_i}} \quad (\beta_i = b_i \frac{\sigma_{x_i}}{\sigma_y});$$

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

Значение a определим из соотношения

$$a = \bar{y} - b_1 \cdot \bar{x}_1 - b_2 \cdot \bar{x}_2 = 86,8 - 1,6151 \cdot 54,9 + 2,2505 \cdot 33,05 = -73,52276 \quad ;$$

$$y_{x_1 x_2} = -73,52 + 1,62 \cdot x_1 - 2,25 \cdot x_2.$$

Для характеристики относительной силы влияния x_1 и x_2 на y рассчитаем средние коэффициенты эластичности:

$$\bar{\varepsilon}_{yx_j} = b_j \frac{\bar{x}_j}{\bar{y}};$$

$$\bar{\varepsilon}_{yx_1} = \frac{1,62 \cdot 54,9}{86,8} = 1,0246\%; \quad \bar{\varepsilon}_{yx_2} = \frac{-2,25 \cdot 33,5}{86,8} = -0,8684\%.$$

С увеличением средней 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 y 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 β_1 и β_2 :

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

Различия в силе влияния фактора на результат, полученные при сравнении $\bar{\varepsilon}_{yx_j}$ и β_j , объясняются тем, что коэффициент эластичности исходит из соотношения средних: $\bar{\varepsilon}_{yx_j} = b_j \frac{\bar{x}_j}{\bar{y}}$, а β -коэффициент

– из соотношения средних квадратических отклонений: $\beta_i = b_i \frac{\sigma_{x_i}}{\sigma_y}$.

2. Линейные коэффициенты частной корреляции здесь рассчитываются по рекуррентной формуле:

$$r_{yx_1 \cdot x_2} = \frac{r_{yx_1} - r_{yx_2} \cdot r_{x_1 x_2}}{\sqrt{(1 - r_{yx_2}^2) \cdot (1 - r_{x_1 x_2}^2)}} = \frac{0,8405 - 0,2101 \cdot 0,116}{\sqrt{(1 - 0,2101^2) \cdot (1 - 0,116^2)}} = 0,8404;$$

$$r_{yx_2 \cdot x_1} = \frac{r_{yx_2} - r_{yx_1} \cdot r_{x_1 x_2}}{\sqrt{(1-r_{yx_1}^2) \cdot (1-r_{x_1 x_2}^2)}} = \frac{-0,2101 + 0,8405 \cdot 0,116}{\sqrt{(1-0,8405^2) \cdot (1-0,116^2)}} = -0,2092;$$

$$r_{x_1 x_2 \cdot y} = \frac{r_{x_1 x_2} - r_{yx_1} \cdot r_{yx_2}}{\sqrt{(1-r_{yx_1}^2) \cdot (1-r_{yx_2}^2)}} = \frac{-0,116 + 0,8405 \cdot 0,2101}{\sqrt{(1-0,8405^2) \cdot (1-0,2101^2)}} = 0,1144.$$

Если сравнить значения коэффициентов парной и частной корреляции, то приходим к выводу, что из-за слабой межфакторной связи ($r_{x_1 x_2} = -0,116$) коэффициенты парной и частной корреляции отличаются незначительно: выводы о тесноте и направлении связи на основе коэффициентов парной и частной корреляции совпадают:

$$r_{yx_1} = 0,8405; r_{yx_2} = -0,2101; r_{x_1 x_2} = 0,1160;$$

$$r_{yx_1 \cdot x_2} = 0,8404; r_{yx_2 \cdot x_1} = -0,2092; r_{x_1 x_2 \cdot y} = 0,1144.$$

Расчет линейного коэффициента множественной корреляции выполним с использованием коэффициентов r_{yx_j} и β_j .

$$R_{yx_1 x_2} = \sqrt{r_{yx_1} \cdot \beta_1 + r_{yx_2} \cdot \beta_2} = \sqrt{0,8405 \cdot 0,8273 + 0,2101 \cdot 0,1141} = \sqrt{0,7193} = 0,8481.$$

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 проверяет гипотезу H_0 о статистической значимости уравнения регрессии и показателя тесноты связи ($R^2 = 0$):

$$F_{\text{факт}} = \frac{R_{yx_1 x_2}^2}{1 - R_{yx_1 x_2}^2} \cdot \frac{n - m - 1}{m} = \frac{0,7193}{0,2807} = 34,6;$$

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

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 $R_{yx_1 x_2}$, которые сформировались под неслучайным воздействием факторов x_1 и x_2 .

Частные F -критерии – F_{x_1} и F_{x_2} оценивают статистическую значимость присутствия факторов x_1 и x_2 в уравнении множественной регрессии, оценивают целесообразность включения в уравнение од-

ного фактора после другого фактора, т.е. F_{x_1} оценивает целесообразность включения в уравнение фактора x_1 после того, как в него был включен фактор x_2 . Соответственно F_{x_2} указывает на целесообразность включения в модель фактора x_2 после фактора x_1 :

$$F_{x_1 \text{ факт}} = \frac{R_{yx_1x_2}^2 - r_{yx_2}^2}{1 - R_{yx_1x_2}^2} \cdot \frac{n - m - 1}{1} = \frac{0,8481^2 - 0,2101^2}{1 - 0,8481^2} \cdot \frac{30 - 2 - 1}{1} = 64,9.$$

$$F_{\text{табл}} = 4,21; \alpha = 0,05.$$

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

$$F_{x_2 \text{ факт}} = \frac{R_{yx_1x_2}^2 - r_{yx_1}^2}{1 - R_{yx_1x_2}^2} \cdot \frac{n - m - 1}{1} = \frac{0,8481^2 - 0,8405^2}{1 - 0,8481^2} \cdot \frac{30 - 2 - 1}{1} = 1,234.$$

Низкое значение $F_{x_2 \text{ факт}}$ (немногом больше 1) свидетельствует о статистической незначимости прироста $r_{yx_1}^2$ 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).

УПРАЖНЕНИЯ

1.1. По статистическим данным, описывающим зависимость производительности труда по плодоконсервным заводам области за год от удельного веса рабочих с технической подготовкой и удельного веса механизированных работ, построить модель множественной регрессии и определить ее значимость.

№ завода	Факторы		Производительность труда
	Удельный вес рабочих с технической подготовкой, %	Удельный вес механизированных работ, %	
1	64	84	4300
2	61	83	4150
3	47	67	3000
4	46	63	3420
5	49	69	3300
6	54	70	3400
7	53	73	3420
8	61	81	4100
9	57	77	3700
10	54	72	3500
11	60	80	4000
12	67	85	4450
13	63	83	4270
14	50	70	3300
15	67	87	4500

1.2. По статистическим данным, описывающим зависимость накопления пяти случайно выбранных семей от дохода и размера, построить модель множественной регрессии и определить ее значимость. Спрогнозируйте накопления семьи, имеющей доход 40 тыс. руб. и имущество стоимостью 25 тыс. руб.

Семья	Накопления, S	Доход, Y	Имущество, W
1	3,0	40	60
2	6,0	55	36
3	5,0	45	36
4	3,5	30	15
5	1,5	30	90

1.3. После финансового кризиса спрос на чебуреки упал, и менеджер был вынужден тратить часть средств на рекламу. Для изучения зависимости объема продаж от цены и расходов на рекламу менеджер использует модель вида: $q_t = a_0 + a_1 p_t + a_2 c_t + a_3 c_t^2 + \varepsilon_t$. В таблице приведены данные наблюдений за 18 недель. Построить модель множественной регрессии. Оцените ее значимость. Найдите оптимальную цену при расходах на рекламу, равных 280 руб. Найдите оптимальный уровень расходов на рекламу при цене чебурека 6 руб. Помогите менеджеру найти оптимальное решение (максимизирующее чистый доход).

t	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
q _t	525	567	396	726	265	615	370	789	513	661	407	608	399	631	545	512	845	571
p _t	5,9	6,5	6,5	6,1	6,6	5,2	5,1	5,1	6,7	5,5	6,6	6,9	6,9	6,5	6,5	6,8	5,1	6,1
c _t	4,8	3,6	5,5	2,7	5,7	1,3	5,8	3,4	3,7	3,6	5,1	5,5	4,7	3,8	4,3	2,7	2,2	3,1

1.4. В кейнсианской теории спрос на деньги зависит от доходов и и процентных ставок. Рассмотрим модель:

$$m_t = a_0 + a_1 y_t + a_2 i_t + \varepsilon_t;$$

где m_t – агрегат денежной массы М1 (млрд. долл.), y_t – валовой национальный продукт (млрд. долл.); i_t — процентные ставки по государственным облигациям (%).

Рассчитайте модель множественной регрессии и оцените спрос на деньги при

1) ВВП = 1000 и $i = 10$;

2) ВВП = 2500 и $i = 5$.

Оцените значимость модели.

t	1	2	3	4	5	6	7	8	9	10	11	12	13	14
m _t	141	146	149	154	161	173	199	216	251	277	310	363	414	478
y _t	506	524	565	597	637	756	873	992	1185	1434	1718	2164	2631	3073
i _t	3,3	2,6	2,9	3,2	3,7	5	5,5	6,6	4,5	7,9	5,3	7,6	11,4	11,1

1.5. В таблице приведены реальный доход на душу населения y (тыс. долл.), процент рабочей силы, занятой в сельском хозяйстве – x_1 , и средний уровень образования населения в возрасте после 25 лет x_2 (число лет, проведенных в учебных заведениях) для 15 развитых стран с 1983 г. Постройте модель множественной линейной регрессии и оцените ее значимость.

Страна	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
y	7	9	9	8	8	14	9	8	10	11	11	12	9	10	12
x ₁	8	9	7	6	10	4	5	5	6	7	6	4	8	5	8
x ₂	9	13	11	11	12	16	11	11	12	14	11	15	15	10	13

1.6. В таблице приведена зависимость выпуска Q от трудозатрат L и капиталовложений K . Оцените по этим данным производственную функцию Кобба-Дугласа $Q = \alpha L^{\beta_1} K^{\beta_2}$. Оцените значимость модели. Рассчитайте объем выпуска при $L = 2500$ и $K = 1800$.

Фирма №	Q	L	K
1	2350	2334	1570
2	2470	2425	1850
3	2110	2230	1150
4	2560	2463	1940
5	2650	2565	2450
6	2240	2278	1340
7	2430	2380	1700
8	2530	2437	1860
9	2550	2446	1880
10	2450	2403	1790
11	2290	2301	1480
12	2160	2253	1240
13	2400	2367	1660
14	2490	2430	1850
15	2590	2470	2000

1.7. Имеются следующие данные о выработке литья на одного работающего x_1 (т), браке литья x_2 (%) и себестоимости 1 т литья y (руб.) по 25 литейным цехам заводов:

i	x_{1i}	x_{2i}	y_i	i	x_{1i}	x_{2i}	y_i	i	x_{1i}	x_{2i}	y_i
1	14,6	4,2	239	10	25,3	0,9	198	19	17,0	9,3	282
2	13,5	6,7	254	11	56,0	1,3	170	20	33,1	3,3	196
3	21,5	5,5	262	12	40,2	1,8	173	21	30,1	3,5	186
4	17,4	7,7	251	13	40,6	3,3	197	22	65,2	1,0	176
5	44,8	1,2	158	14	75,8	3,4	172	23	22,6	5,2	238
6	111,9	2,2	101	15	27,6	1,1	201	24	33,4	2,3	204
7	20,1	8,4	259	16	88,4	0,1	130	25	19,7	2,7	205
8	28,1	1,4	186	17	16,6	4,1	251				
9	22,3	4,2	204	18	33,4	2,3	195				

Требуется:

- найти множественный коэффициент детерминации и пояснить его смысл;
- найти уравнение множественной регрессии y по x_1 и x_2 , оценить значимость этого уравнения и его коэффициентов на уровне $\alpha=0,05$;

- в) сравнить раздельное влияние на зависимую переменную каждой из объясняющих переменных, используя стандартизованные коэффициенты регрессии и коэффициенты эластичности;
- г) найти 95%-ные доверительные интервалы для коэффициентов регрессии, а также для среднего и индивидуальных значений себестоимости 1 т литья в цехах, в которых выработка литья на одного работающего составляет 40 т, а брак литья — 5%.

1.8. Имеются следующие данные о годовых ставках месячных доходов по трем акциям за шестимесячный период:

Акция	Доходы по месяцам, %					
	<i>A</i>	5,4	5,3	4,9	4,9	5,4
<i>B</i>	6,3	6,2	6,1	5,8	5,7	5,7
<i>C</i>	9,2	9,2	9,1	9,0	8,7	8,6

Есть основания предполагать, что доходы y по акции *C* зависят от доходов x_1 и x_2 по акциям *A* и *B*.

Требуется:

- а) составить уравнение регрессии y по x_1 и x_2 ;
- б) найти множественный коэффициент детерминации R^2 и пояснить его смысл;
- в) проверить значимость полученного уравнения регрессии на уровне $\alpha=0,05$;
- г) оценить средний доход по акции *C*, если доходы по акциям *A* и *B* составили соответственно 5,5 и 6,0%.

1.9. Имеются данные о деятельности крупнейших компаний США в 19XX г.

п/п	Чистый доход, млрд. долл.	Оборот капитала, млрд. долл. США	Использованный капитал, млрд. долл.	Численность служащих, тыс. чел.	Рыночная капитализация компании, млрд. долл.
	y	x_1	x_2	x_3	x_4
1	0,9	31,3	18,9	43,0	40,9
2	1,7	13,4	13,7	64,7	40,5
3	1,2	16,9	16,5	55,6	45,0
4	2,0	25,6	14,3	74,3	42,3
5	2,7	27,9	13,7	42,0	39,8
6	2,5	16,7	20,9	39,4	51,3
7	1,9	21,8	12,4	64,5	27,9

8	1,6	32,4	16,7	70,2	39,4
9	2,1	12,6	21,3	71,3	45,1
10	1,8	30,8	10,4	61,2	42,6
11	1,1	25,0	13,2	55,5	29,1
12	1,3	17,6	18,9	61,3	40,8
13	2,3	12,2	17,6	72,5	32,9
14	0,7	31,1	21,0	80,1	45,5
15	1,6	32,0	20,7	61,9	41,3
16	2,0	19,9	16,4	62,3	37,9
17	0,9	20,7	13,7	70,3	42,1
18	1,0	22,1	15,8	59,1	40,6
19	0,9	17,3	11,0	49,8	35,8
20	0,7	15,5	5,8	80,8	27,2

Требуется:

- а) Рассчитать матрицу парных коэффициентов корреляции; оценить статистическую значимость коэффициентов корреляции.
- б) Рассчитайте параметры линейного уравнения множественной регрессии с полным перечнем факторов.
- в) Оценить статистическую значимость параметров регрессионной модели с помощью t -критерия; нулевую гипотезу о значимости уравнения проверить с помощью F -критерия; оценить качество уравнения регрессии с помощью коэффициента детерминации R^2 .
- г) Дать сравнительную оценку силы связи факторов с результатом с помощью коэффициентов эластичности, β и Δ коэффициентов.
- д) Отобрать информативные факторы в модель по t -критерию для коэффициентов регрессии. Построить модель только с информативными факторами и оценить ее параметры [5].
- е) Рассчитать прогнозное значение результата, если прогнозные значения факторов составляют 80% от их максимальных значений.
- ж) Рассчитать ошибки и доверительный интервал прогноза для уровня значимости 5 или 10% ($\alpha = 0,05$; $\alpha = 0,10$).

1.10. В таблице указаны парные коэффициенты корреляции. Проведите анализ целесообразности включения заданных факторов в уравнение множественной линейной регрессии.

	y	x ₁	x ₂	x ₃	x ₄
y	1	0,71	0,58	0,08	0,62
x ₁	-	1	0,53	0,2	0,81
x ₂	-	-	1	0,13	0,3
x ₃	-	-	-	1	0,25
x ₄	-	-	-	-	1

1.11. По 19 предприятиям оптовой торговли изучается зависимость объема реализации y от размера торговой площади x_1 и товарных запасов x_2 . Были получены следующие варианты уравнений регрессии:

1. $y = 25 + 15x_1$, $r^2 = 0,90$
2. $y = 42 + 27x_2$, $r^2 = 0,84$
3. $y = 30 + 10x_1 + 8x_2$, $R^2 = 0,92$
(2,5) (4,0)
4. $y = 21 + 14x_1 + 20x_2 + 0,6x_2^2$, $R^2 = 0,95$
(5,0) (12,0) (0,2)

В скобках указаны значения стандартных ошибок для коэффициентов регрессии.

Требуется:

1. Проанализировать тесноту связи результата с каждым из факторов.
2. Выбрать наилучшее уравнение регрессии, обоснуйте принятое решение.

1.12. По 25 территориям страны изучается влияние климатических условий на урожайность зерновых y (ц/га). Для этого были отобраны две объясняющие переменные:

x_1 – количество осадков в период вегетации, мм;

x_2 – средняя температура воздуха, °С.

Матрица парных коэффициентов корреляции этих показателей имеет следующий вид:

	y	x ₁	x ₂
y	1,0		
x ₁	0,6	1,0	
x ₂	-0,5	-0,9	1,0

Требуется:

1. Определить частные коэффициенты корреляции результата с каждым из факторов. Объяснить различие полученных парных и частных коэффициентов корреляции результатов;

2. Исследователь, анализирующий данную зависимость, намерен определить на основе приведенной выше матрицы, какое уравнение регрессии лучше строить:

- а) парную линейную регрессию y на x_1 ;
- б) парную линейную регрессию y на x_2 ;
- в) множественную линейную регрессию.

Дать ответ на этот вопрос.

3. Построить уравнение регрессии в стандартизированном масштабе и сделать выводы.

1.13. Для изучения рынка жилья в городе по данным о 46 коттеджах было построено уравнение множественной регрессии:

$$y = 21,1 - 6,2x_1 + 0,95x_2 + 3,57x_3; \quad R^2 = 0,7,$$

(1.8) (0.54) (0.83)

где y - цена объекта, тыс. долл.;

x_1 - расстояние до центра города, км;

x_2 - полезная площадь объекта, кв. м;

x_3 - число этажей в доме, ед.;

R^2 - коэффициент множественной детерминации.

В скобках указаны значения стандартных ошибок для коэффициентов множественной регрессии.

Требуется:

1. Проверить гипотезу о том, что коэффициент регрессии b_1 в генеральной совокупности равен нулю.
2. Проверить гипотезу о том, что коэффициент регрессии b_2 в генеральной совокупности равен нулю.
3. Проверить гипотезу о том, что коэффициент регрессии b_3 в генеральной совокупности равен нулю.
4. Проверить гипотезу о том, что коэффициент регрессии b_1 , b_2 и b_3 в генеральной совокупности одновременно равны нулю (или коэффициент детерминации равен нулю).
5. Пояснить причины расхождения результатов, полученных в пп.1, 2 и 3, с результатами, полученными в п.4.

1.14. По 30 наблюдениям матрица парных коэффициентов корреляции оказалась следующей:

	y	x_1	x_2	x_3
y	1,00			
x_1	0,30	1,00		
x_2	0,60	0,10	1,00	
x_3	0,40	0,15	0,80	1,00

Требуется:

1. Построить уравнение регрессии в стандартизированном виде и сделать выводы;
2. Определить показатель множественной корреляции (нескорректированный и скорректированный);
3. Оценить целесообразность включения переменной x_1 в модель после введения в нее переменных x_2 и x_3 .

СПИСОК РЕКОМЕНДУЕМОЙ ЛИТЕРАТУРЫ

1. Доугерти К. Введение в эконометрику: Пер. с англ. - М.: ИНФРА-М, 1999.-402 с.
2. Катышев П.К., Магнус Я.Р., Пересецкий А.А. Сборник задач к начальному курсу эконометрики. - М.: Дело, 2002. - 208 с.
3. Кремер Н.Ш., Путко Б.А. Эконометрика: Учебник для вузов / Под ред. проф. Н.Ш. Кремера. - М.: ЮНИТИ-ДАНА, 2002. - 311 с.
4. Магнус Я.Р., Катышев П.К., Пересецкий А.А. Эконометрика. Начальный курс: Учебник. - М.: Дело, 2001. - 400 с.
5. Практикум по эконометрике: Учебн. пособие / Под ред. И.И. Елисеевой. - М.: Финансы и статистика, 2006. - 344 с.
6. Сборник задач по эконометрике: Учебное пособие для студентов экономических вузов / Сост. Е.Ю. Дорохина, Л.Ф. Преснякова, Н.П. Тихомиров. - М.: Издательство «Экзамен», 2003. - 224 с.
7. Учебно-методическое пособие по курсу «Основы эконометрики» / Составитель Леванова Л.Н. – Саратов, 2003. – 34 с.
8. Эконометрика: Учебник / Под ред. И.И. Елисеевой. - М.: Финансы и статистика, 2006. - 576 с.
9. Эконометрика: Учебник / Тихомиров Н.П., Дорохина Е.Ю. -М.: Издательство «Экзамен», 2003. - 512 с.
10. Эконометрика: Учебно-методическое пособие / Шалабанов АХ, Роганов Д.А. - Казань: ТИСБИ, 2004. - 198 с.

ПРИЛОЖЕНИЯ

ПРИЛОЖЕНИЕ А

ТАБЛИЦА ЗНАЧЕНИЙ F -КРИТЕРИЯ ФИШЕРА ПРИ УРОВНЕ
ЗНАЧИМОСТИ $A = 0,05$

	1	2	3	4	5	6	8	12	24	∞
1	2	3	4	5	6	7	8	9	10	11
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
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

Продолжение прилож. А

1	2	3	4	5	6	7	8	9	10	11
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
150	3,90	3,06	2,66	2,43	2,27	2,16	2,00	1,82	1,59	1,18
200	3,89	3,04	2,65	2,42	2,26	2,14	1,98	1,80	1,57	1,14
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
∞	3,84	2,99	2,60	2,37	2,21	2,09	1,94	1,75	1,52	1

ПРИЛОЖЕНИЕ Б

ФОРМУЛЫ ДЛЯ РАСЧЕТА СРЕДНИХ КОЭФФИЦИЕНТОВ ЭЛАСТИЧНОСТИ ДЛЯ НАИБОЛЕЕ ЧАСТО ИСПОЛЬЗУЕМЫХ ТИПОВ УРАВНЕНИЙ РЕГРЕССИИ

Вид функции, y	Первая производная, y'	Средний коэффициент эластичности, $\bar{\varepsilon}$	Линеаризация
$y = a + b \cdot x + \varepsilon$	b	$\frac{b \cdot \bar{x}}{a + b \cdot \bar{x}}$	-
$y = a + b_1 \cdot x + b_2 \cdot x^2$	$b_1 + 2 \cdot b_2 \cdot x$	$\frac{(b_1 + 2 \cdot b_2 \cdot \bar{x}) \cdot \bar{x}}{a + b_1 \cdot \bar{x} + b_2 \cdot \bar{x}^2}$	$X_1 = x, X_2 = x^2$
$y = a + \frac{b}{x} + \varepsilon$	$-\frac{b}{x^2}$	$-\frac{b}{a \cdot \bar{x} + b}$	$X = 1/x, Y = y$
$y = a \cdot x^b \cdot \varepsilon$	$a \cdot b \cdot x^{b-1}$	b	$X = \ln x, Y = \ln y$
$y = a \cdot b^x \cdot \varepsilon$	$a \cdot \ln b \cdot b^x$	$\bar{x} \cdot \ln b$	$X = x, Y = \ln y$
$y = a + b \cdot \ln x + \varepsilon$	$\frac{b}{x}$	$\frac{b}{a + b \cdot \ln \bar{x}}$	$X = \ln x, Y = y$
$y = e^{a+b \cdot x} \cdot \varepsilon$	$b \cdot e^{a+b \cdot x}$	$\bar{x} \cdot b$	$X = x, Y = \ln y$
$y = \frac{a}{1 + b_1 \cdot e^{-b_2 \cdot x + \varepsilon}}$	$\frac{a \cdot b_1 \cdot b_2 \cdot e^{-b_2 \cdot x}}{(1 + b_1 \cdot e^{-b_2 \cdot x})^2}$	$\frac{b_1 \cdot b_2 \cdot \bar{x}}{b_1 + e^{-b_2 \cdot \bar{x}}}$	-
$y = \frac{1}{a + b \cdot x + \varepsilon}$	$-\frac{b}{(a + b \cdot x)^2}$	$-\frac{b \cdot \bar{x}}{a + b \cdot \bar{x}}$	$X = x, Y = 1/y$

ПРИЛОЖЕНИЕ В

КРИТИЧЕСКИЕ ЗНАЧЕНИЯ *T*-КРИТЕРИЯ СТЬЮДЕНТА ПРИ УРОВНЕ ЗНАЧИМОСТИ 0,10, 0,05, 0,01 (ДВУХСТОРОННИЙ)

Число степеней свободы k_1, k_2	α			Число степеней свободы k_1, k_2	α		
	00,10	0,05	0,01		00,10	0,05	0,01
1	6,3138	12,706	63,657	18	1,7341	2,1009	2,8784
2	2,9200	4,3027	9,9248	19	1,7291	2,0930	2,8609
3	2,3534	3,1825	5,8409	20	1,7247	2,0860	2,8453
4	2,1318	2,7764	4,5041	21	1,7207	2,0796	2,8314
5	2,0150	2,5706	4,0321	22	1,7171	2,0739	2,8188
6	1,9432	2,4469	3,7074	23	1,7139	2,0687	2,8073
7	1,8946	2,3646	3,4995	24	1,7109	2,0639	2,7969
8	1,8595	2,3060	3,3554	25	1,7081	2,0595	2,7874
9	1,8331	2,2622	3,2498	26	1,7056	2,0555	2,7787
10	1,8125	2,2281	3,1693	27	1,7033	2,0518	2,7707
11	1,7959	2,2010	3,1058	28	1,7011	2,0484	2,7633
12	1,7823	2,1788	3,0545	29	1,6991	2,0452	2,7564
13	1,7709	2,1604	3,0123	30	1,6973	2,0423	2,7500
14	1,7613	2,1448	2,9768	40	1,6839	2,0211	2,7045
15	1,7530	2,1315	2,9467	60	1,6707	2,0003	2,6603
16	1,7459	2,1199	2,9208	120	1,6577	1,9799	2,6174
17	1,7396	2,1098	2,8982	∞	1,6449	1,9600	2,5758

Винник Ольга Григорьевна

**ЭКОНОМЕТРИКА
И ЭКОНОМИКО-МАТЕМАТИЧЕСКИЕ
МЕТОДЫ И МОДЕЛИ**

Задачник

**по одноименной дисциплине для студентов
экономических специальностей
дневной формы обучения**

Подписано к размещению в электронную библиотеку
ГГТУ им. П. О. Сухого в качестве электронного
учебно-методического документа 17.10.11.

Рег. № 41Е.

E-mail: ic@gstu.by

<http://www.gstu.by>