МИНИСТЕРСТВО СЕЛЬСКОГО ХОЗЯЙСТВА РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ОБРАЗОВАНИЯ
БАШКИРСКИЙ ГОСУДАРСТВЕННЫЙ АГРАРНЫЙ УНИВЕРСИТЕТ

Е.В. Кабашова

ОСНОВЫ ЭКОНОМЕТРИКИ

ЛАБОРАТОРНЫЙ ПРАКТИКУМ

Учебное пособие

УФА
Башкирский ГАУ
2017

УДК 330.43(07)
ББК 65в6(я7)
К 12

РЕЦЕНЗЕНТЫ:

А.Р. Кузнецова
доктор экономических наук, профессор кафедры экономики и менеджмента
ФГБОУ ВО Башкирский ГАУ

Р.Р. Бакирова
кандидат экономических наук, доцент кафедры бухгалтерского учета, аудита, статистики Уфимского филиала Финансового университета
при Правительстве РФ

Кабашова Е.В.

К 12 Основы эконометрики : лабораторный практикум : учебное пособие / Е. В. Кабашова. – Уфа : Башкирский ГАУ, 2017. – 108 с.

В учебном пособии рассмотрены основные вопросы построения эконометрических моделей в пространстве и во времени, а именно: понятие о корреляционной связи, расчет и интерпретация параметров парной (линейной и нелинейной) и множественной регрессии и корреляции, оценка их надежности, а также построение мультипликативной и аддитивной моделей временных рядов, оценка автокорреляции в остатках.

Учебное пособие содержит цель занятия, основной теоретический материал, примеры решения типовых задач, практические задания для самостоятельной работы обучающихся, контрольные вопросы, тестовые задания, библиографический список, глоссарий.

При разработке пособия, как правило, использовались официальные статистические данные Федеральной службы государственной статистики, что дает обучающимся представление о фактически складывающейся социально-экономической ситуации в России и ее регионах.

Учебное пособие предназначено для студентов очной и заочной формы обучения направлений подготовки *38.03.01 Экономика*, *38.03.02 Менеджмент* и других экономических направлений.

УДК 330.43(07)
ББК 65в6(я7)
К 12

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ

Развитие экономики, усложнение социально-экономических процессов и явлений, а также взаимосвязей между ними, повышение требований к качеству принимаемых управленческих решений требует комплексного анализа протекающих процессов на основе привлечения современных математических и статистических методов.

Эконометрика – это самостоятельная научная дисциплина, объединяющая совокупность теоретических результатов, приемов, методов и моделей, предназначенных для того, чтобы на базе экономической теории, экономической статистики, математико-статистического инструментария придавать конкретное количественное выражение социально-экономическим процессам и явлениям и их взаимосвязям.

Практическая значимость эконометрики определяется тем, что применение ее методов позволяет выявить реально существующие связи между явлениями, дать обоснованный прогноз развития социально-экономических явлений, проверить и численно оценить экономические последствия принимаемых управленческих решений.

Успех эконометрического исследования наряду с правильным применением математического и статистического инструментария также зависит от экономически грамотной постановки задачи и последующей интерпретации полученных результатов.

Развитию эконометрических методов и их широкому использованию в практической деятельности способствовало внедрение средств вычислительной техники. Разработка программных пакетов (SPSS, Statistica, STADIA), реализующих методы построения и исследования эконометрических моделей, привело к тому, что выполнение эконометрических процедур стало доступным широкому кругу аналитиков, экономистов и менеджеров.

Цель дисциплины «Эконометрика» – овладение теоретическими знаниями и практическими навыками построения эконометрических моделей и использования их для описания, анализа и прогнозирования экономических процессов и явлений на микро- и макроуровне.

Задачи изучения дисциплины:

- усвоение студентами знаний в области спецификации моделей;
- развитие навыков построения эконометрических моделей и проверки их значимости;
- выработка умения оценивать неизвестные параметры выбранных моделей, делать прогнозы и оценивать их точность;
- развитие навыков исследовательской деятельности.

В результате усвоения материала дисциплины студент должен:

а) знать:

- основные категории и понятия эконометрики;
- принципы и методы построения пространственных и временных эконометрических моделей;

- методику оценки существенности и значимости моделей и их параметров;

- возможности реализации типовых задач на компьютере с помощью ППП Excel, Statistica;

б) уметь и обладать навыками:

- формировать концепцию эконометрической модели на основе качественного анализа объекта исследования;

- строить эконометрические модели с использованием процедур регрессионного анализа и анализа временных рядов, проводить оценку взаимосвязей экономических показателей с помощью статистических методов;

- оценивать качество построенных эконометрических моделей, интерпретировать полученные результаты и делать прогнозы социально-экономических явлений и оценивать их точность.

Настоящее учебное пособие состоит из четырех глав. В первой главе рассмотрены вопросы построения парных линейных регрессионных моделей: постановка задачи, спецификация и оценка параметров моделей, оценка качества полученных моделей, получение точечного и интервального прогнозных значений, экономическая интерпретация модели. Во второй главе изучены основы нелинейной регрессии и корреляции. Третья глава посвящена построению множественных регрессионных моделей. Подробно рассмотрены вопросы спецификации и оценки параметров модели, оценки качества полученной модели и ее статистической значимости. В четвертой главе рассмотрены вопросы моделирования одномерных временных рядов: структура временного ряда, явление автокорреляции, моделирование тенденции, циклических и сезонных колебаний. Отдельное внимание уделено построению аддитивных и мультипликативных моделей временного ряда.

Все главы пособия содержат идентичные разделы:

Теоретические положения – раздел, содержащий основные понятия и формулы для определения параметров регрессии и корреляции.

Решение типовых задач – в данном разделе подробно рассматриваются примеры решения задач по формулам, а также с помощью табличного процессора MS Excel.

Задания для самостоятельной работы обучающихся – практические задачи, составленные на основе официальных данных Росстата, служащие для закрепления теоретического материала и практических навыков по эконометрическому моделированию.

Контрольные вопросы для самопроверки и тестовые задания – помогают проверить уровень освоения пройденного материала.

Учебное пособие предназначено для студентов направлений подготовки «Экономика», «Менеджмент» очного и заочного обучения. Пособие будет также полезно всем желающим ознакомиться с основными задачами, моделями и методами эконометрики.

1 ПАРНАЯ ЛИНЕЙНАЯ РЕГРЕССИЯ И КОРРЕЛЯЦИЯ В ЭКОНОМЕТРИЧЕСКИХ ИССЛЕДОВАНИЯХ

Цель занятия – ознакомиться с основными понятиями парной линейной регрессии и корреляции, овладеть навыками определения параметров линейной регрессии и корреляции и оценки из значимости с использованием формул и табличного процессора MS Excel.

1.1 Теоретические положения

1.1.1 Понятие о корреляционной связи. Этапы эконометрического моделирования

1.1.2 Оценка и интерпретация параметров парной линейной регрессии и корреляции

1.1.3 Оценка существенности уравнения регрессии и его параметров

1.1.1 Понятие о корреляционной связи. Этапы эконометрического моделирования

Корреляционная зависимость – это взаимосвязь между признаками, состоящая в том, что средняя величина значений одного признака меняется в зависимости от изменения другого признака. Корреляционную зависимость можно выявить только в виде общей тенденции при массовом изучении факторов. Корреляционная связь – это частный случай статистической связи.

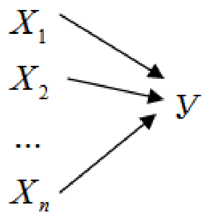
По направлению различают ***прямую*** и ***обратную*** связь. Если увеличение факторного признака x приводит к увеличению средних значений результативного признака y, то связь прямая (например, с увеличением доходов населения возрастает объем потребления); если увеличение факторного признака приводит к уменьшению результативного признака, то связь обратная (с увеличением урожайности сельскохозяйственных культур снижается себестоимость единицы продукции) [3].

Различают следующие варианты зависимостей (таблица 1.1).

Таблица 1.1 Зависимости между результативным и факторным признаками

Название	Сущность	Графическое представление
Парная корреляция	Отражает связь между двумя признаками: результативным и факторным	$X \rightarrow У$
Частная корреляция	Характеризует зависимость между результативным и одним факторным признаком при фиксированном значении других факторных признаков	X_1, $\overline{X}_2$, ..., $\overline{X}_n$ $\rightarrow У$

Продолжение таблицы 1.1

Множественная корреляция	Исследует зависимость результативного признака от нескольких факторных признаков	

Этапы эконометрического моделирования представлены на рисунке 1.1.

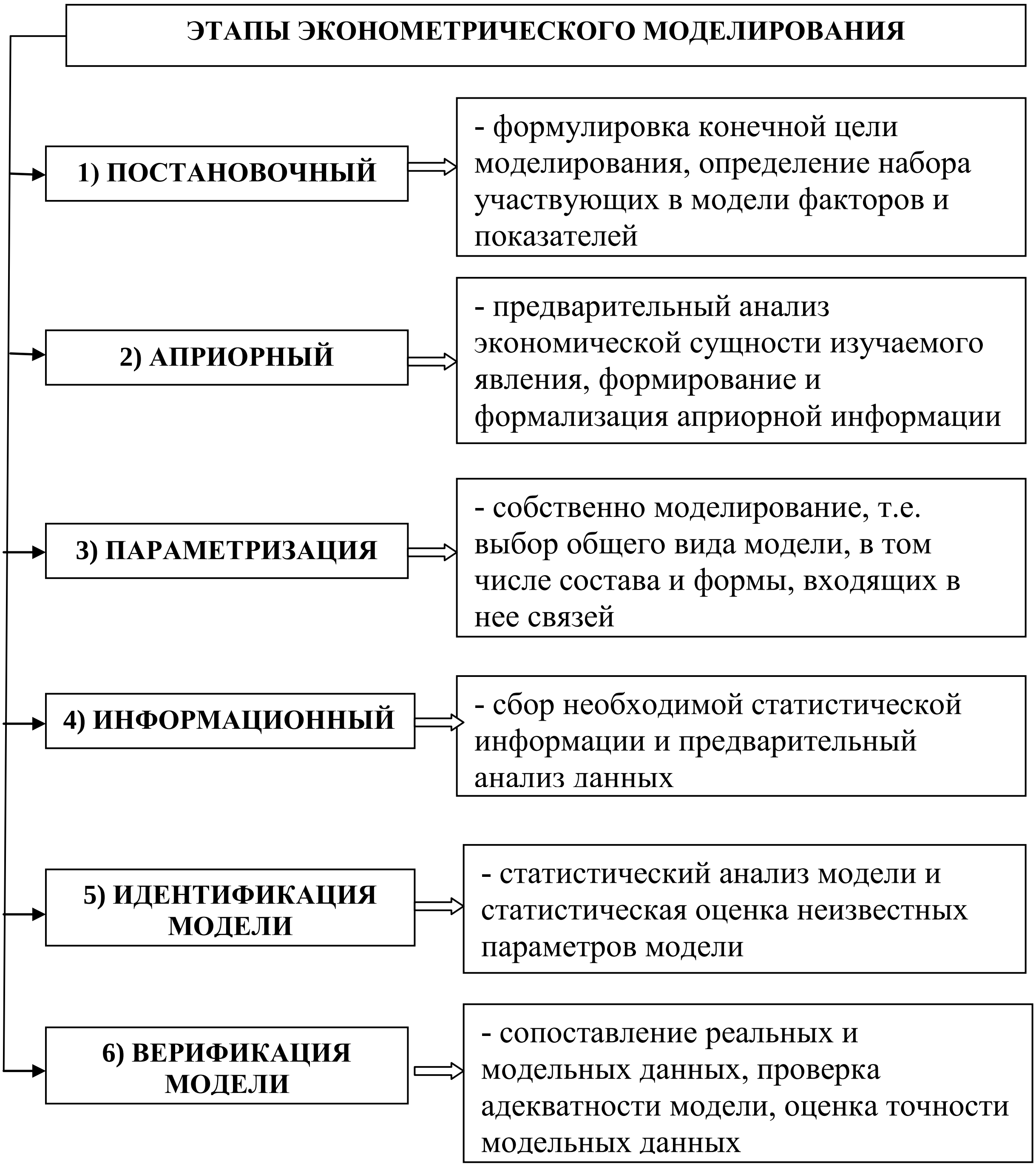


Рисунок 1.1 Этапы эконометрического моделирования

1.1.2 Оценка и интерпретация параметров парной линейной регрессии и корреляции

Уравнение парной линейной регрессии имеет вид:

$$\tilde{y} = a + b \cdot x + \varepsilon, \qquad (1.1)$$

где $\tilde{y}$ – теоретическое значение результативного признака, найденное из уравнения регрессии;

x – независимая переменная (факторный признак);

a, b – параметры уравнения регрессии (a – экономического содержания не имеет; b – коэффициент регрессии);

ε – случайная величина, характеризующая отклонения реального значения результативного признака от теоретического.

Параметры линейной регрессии оценивают с помощью метода наименьших квадратов (МНК).

МНК позволяет получить такие оценки параметров a и b, при которых сумма квадратов отклонений фактических значений результативного признака от расчетных (теоретических) минимальна:

$$\sum (y - \tilde{y})^2 \to \min.$$

Система нормальных уравнений МНК имеет вид:

$$\begin{cases} na + b\sum x = \sum y; \\ a\sum x + b\sum x^2 = \sum xy, \end{cases} \qquad (1.2)$$

где n – количество наблюдений.

Для решения системы можно воспользоваться готовыми формулами:

$$b = \frac{\operatorname{cov}(x, y)}{\sigma_x^2} = \frac{\overline{yx} - \bar{y} \cdot \bar{x}}{\overline{x^2} - \bar{x}^2}, \qquad (1.3)$$

$$a = \bar{y} - b \cdot \bar{x} \qquad (1.4)$$

где $\operatorname{cov}(x, y)$ – ковариация признаков;

σ_x^2 – дисперсия признака x.

Параметр b называется ***коэффициентом регрессии.*** Его величина показывает среднее изменение результата при изменении фактора на одну единицу [19].

Тесноту связи изучаемых явлений характеризует ***коэффициент корреляции (r),*** который определяется по формуле:

$$r = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x \cdot \sigma_y} = b\frac{\sigma_x}{\sigma_y}. \qquad (1.5)$$

где средние квадратические отклонения:

$$\sigma_x = \sqrt{\frac{\sum (x-\bar{x})^2}{n}} = \sqrt{\bar{x^2} - \bar{x}^2}\,; \tag{1.6}$$

$$\sigma_y = \sqrt{\frac{\sum (y-\bar{y})^2}{n}} = \sqrt{\bar{y^2} - \bar{y}^2}\,. \tag{1.7}$$

Коэффициент корреляции может принимать значения: $-1 \le r \le 1$. Если $r > 0$, то связь между признаками прямая, если $r < 0$ – связь обратная.

Для оценки тесноты связи используют шкалу Чеддока:

до $\pm 0{,}3$ – связь отсутствует или очень слабая;

от $\pm 0{,}3$ до $\pm 0{,}5$ – связь слабая;

от $\pm 0{,}5$ до $\pm 0{,}7$ – связь умеренная;

от $\pm 0{,}7$ до $\pm 1{,}0$ – связь сильная.

Для оценки качества подбора линейной функции рассчитывается квадрат линейного коэффициента корреляции – ***коэффициент детерминации*** (r^2), который показывает, на сколько процентов вариация результативного признака определяется вариацией факторов, включенных в модель [20].

Качество построенной модели оценивает также ***средняя ошибка аппроксимации*** – это среднее отклонение расчетных значений от фактических:

$$\bar{A} = \frac{1}{n} \sum \left| \frac{y - \tilde{y}}{y} \right| \cdot 100\%. \tag{1.8}$$

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

Средний коэффициент эластичности рассчитывается по формуле:

$$\bar{Э} = f'(x) \frac{\bar{x}}{\bar{y}}. \tag{1.9}$$

Для линейной функции $y = a + bx$ первая производная $f'(x) = b$, откуда $\bar{Э} = b \cdot \frac{\bar{x}}{\bar{y}}$.

Средний коэффициент эластичности показывает, на сколько процентов в среднем изменится результат при изменении фактора на 1%.

1.1.3 Оценка существенности уравнения регрессии и его параметров

Значимость уравнения регрессии в целом оценивается с помощью *F*-критерия Фишера. При этом выдвигается нулевая гипотеза (H_0), что коэффициент регрессии равен нулю, т.е. b=0, и, следовательно, фактор x не оказывает влияния на результат y. Расчету F – критерия Фишера предшествует ***анализ дисперсии.*** Центральное место в нем занимает разложение общей суммы квадратов отклонений переменной y от среднего значения $\bar{y}$ на две части – факторную (или объясненную регрессией) и остаточную [4].

Дисперсионный анализ оформляется в таблице 1.2.

Таблица 1.2 Дисперсионный анализ

Вариация результата	Число степеней свободы, *df*	Сумма квадратов отклонений, *S*	Дисперсия на одну степень свободы, *D*	$F_{факт.}$	$F_{табл.}$
Общая	*n-1*	$\sum(y-\bar{y})^2$	$D_{общ.}=\dfrac{\sum(y-\bar{y})^2}{n-1}$		
Факторная	*m =1*	$\sum(\tilde{y}-\bar{y})^2$	$D_{ф.}=\dfrac{\sum(\tilde{y}-\bar{y})^2}{m}$	$\dfrac{\sum(\tilde{y}-\bar{y})^2/m}{\sum(y-\tilde{y})^2/(n-m-1)}$	$F_{табл.}$
Остаточная	*n-m-1*	$\sum(y-\tilde{y})^2$	$D_{ост.}=\dfrac{\sum(y-\tilde{y})^2}{n-m-1}$		

где n – число единиц совокупности;

m – число параметров при переменных х.

Если сумма квадратов отклонений, обусловленная регрессией, будет больше остаточной суммы квадратов, то уравнение регрессии статистически значимо и фактор х оказывает существенное воздействие на результат у.

Любая сумма квадратов отклонений связана с числом степеней свободы (df – degrees of freedom), т.е. с числом свободы независимого варьирования признака. Число степеней свободы связано с числом единиц совокупности n и с числом определяемых по ней констант.

Разделив каждую сумму квадратов на соответствующее ей число степеней свободы, получим дисперсию на одну степень свободы (D).

$F_{факт}$ определяется из соотношения значений факторной и остаточной дисперсии, рассчитанных на одну степень свободы:

$$F_{факт.}=\frac{\sum(\tilde{y}-\bar{y})^2/m}{\sum(y-\tilde{y})^2/(n-m-1)}. \qquad (1.10)$$

Для оценки значимости уравнения регрессии $F_{факт.}$ сравнивается с $F_{табл.}$ при $\alpha=0{,}05$, $k_1=m=1$, $k_2=n-m-1=n-2$. Если $F_{факт.} > F_{табл.}$, то уравнение регрессии значимо, статистически надежно и может быть использовано для прогнозирования.

Для оценки статистической значимости коэффициентов регрессии и корреляции рассчитываются t-критерии Стьюдента и доверительные интервалы каждого из показателей.

t-критерии Стьюдента рассчитываются по формулам:

$$t_a=\frac{a}{m_a}; \qquad t_b=\frac{b}{m_b}; \qquad t_r=\frac{r}{m_r}, \qquad (1.11)$$

где m_a, m_b, m_r - случайные ошибки параметров линейной регрессии, которые определяются по формулам:

$$m_a = \sqrt{D_{ост.}} \frac{\sqrt{\sum x^2}}{n\sigma_x}; \tag{1.12}$$

$$m_b = \frac{\sqrt{D_{ост.}}}{\sigma_x \sqrt{n}}; \tag{1.13}$$

$$m_r = \sqrt{\frac{1-r^2}{n-2}}, \tag{1.14}$$

где $D_{ост.}$ – остаточная дисперсия на одну степень свободы.

Фактическое значение t-критерия Стьюдента сравнивают с табличным, которое устанавливается по специальной таблице (стр.188 практикума) на основе заданной вероятности (90, 95, 99%) и числа степеней свободы (n-2).

Если $t_{факт.} > t_{табл.}$, то параметры значимы и статистически надежны; если $t_{факт.} < t_{табл.}$, то параметры уравнения незначимы и ненадежны.

Для установления доверительного интервала определяют предельную ошибку (Δ) для каждого показателя:

$$\Delta_a = t_{табл.} \cdot m_a; \qquad \Delta_b = t_{табл.} \cdot m_b; \qquad \Delta_r = t_{табл.} \cdot m_r \tag{1.15}$$

Формулы для расчета доверительного интервала имеют вид:

$$\gamma_a = a \pm \Delta_a; \qquad \gamma_b = b \pm \Delta_b; \qquad \gamma_r = r \pm \Delta_r \tag{1.16}$$

Если в границы доверительного интервала попадает ноль, т.е. нижняя граница отрицательна, а верхняя положительна, то оцениваемый параметр принимается нулевым, так как не может одновременно принимать и положительное, и отрицательное значения.

1.2 Решение типовых задач

Пример 1.1

По данным регионов Центрального федерального округа Российской Федерации за 2015 год изучается зависимость среднедушевых денежных доходов населения от уровня безработицы (таблица 1.3).

Задание:

1) Для характеристики зависимости среднедушевых денежных доходов населения от уровня безработицы рассчитать параметры уравнения линейной регрессии.
2) Определить средний коэффициент эластичности.
3) Рассчитать коэффициент корреляции.
4) Оценить значимость модели с помощью показателя детерминации, средней ошибки аппроксимации и F-критерия Фишера.
5) С вероятностью 0,95 указать доверительный интервал ожидаемого значения величины среднедушевых денежных доходов населения в предположении сокращения уровня безработицы на 5,0% от своего среднего уровня и найти доверительный интервал прогноза.
6) Проанализировать полученные результаты.

Таблица 1.4 Исходные данные для корреляционно-регрессионного анализа

Регионы	Уровень безработицы, %	Среднедушевые денежные доходы, тыс. руб.
1.Белгородская область	4,0	25,4
2.Брянская область	5,0	22,0
3.Владимирская область	4,3	20,6
4.Воронежская область	4,5	25,5
5.Ивановская область	4,3	20,4
6.Калужская область	4,2	25,0
7.Костромская область	4,3	19,3
8.Курская область	3,9	23,2
9.Липецкая область	3,7	25,3
10.Московская область	2,7	34,9
11.Орловская область	5,1	20,0
12.Рязанская область	4,4	22,0
13.Смоленская область	5,1	21,8
14.Тамбовская область	4,3	22,4
15.Тверская область	5,3	20,6
16.Тульская область	4,1	23,0
17.Ярославская область	3,8	23,9

РЕШЕНИЕ:

1) Параметры a и b линейной регрессии $\tilde{y} = a + b \cdot x$ рассчитываются с помощью метода наименьших квадратов. Для этого составим систему нормальных уравнений (1.2).

По исходным данным определим $\sum y$, $\sum x$, $\sum y \cdot x$, $\sum x^2$, $\sum y^2$ в расчетной таблице 1.4.

Система нормальных уравнений составит:

$$\begin{cases} 17a + 73{,}0b = 395{,}2; \\ 73{,}0a + 319{,}8b = 1669{,}3. \end{cases}$$

Решив систему, получим: a = 42,064; b = -4,382.

Уравнение линейной регрессии имеет вид:

$$\tilde{y} = 42{,}064 - 4{,}382 \cdot x.$$

Параметры уравнения можно определить и по формулам (1.3), (1.4):

$$b = \frac{98{,}2 - 4{,}3 \cdot 23{,}2}{18{,}8 - 4{,}3^2} = 5{,}032; \qquad a = 23{,}2 - (-5{,}032) \cdot 4{,}3 = 44{,}838.$$

Таблица 1.4 Расчет показателей парной линейной регрессии и корреляции

№	x	y	$x \cdot y$	x^2	y^2	$\tilde{y}$	$\left\lvert \frac{y-\tilde{y}}{y} \right\rvert$	$(y-\tilde{y})^2$	$(x-\bar{x})^2$
1	4,0	25,4	101,5	16,0	643,7	24,6	0,032	0,67	0,09
2	5,0	22,0	110,2	25,0	485,7	20,1	0,087	3,71	0,49
3	4,3	20,6	88,4	18,5	423,1	23,2	0,129	7,05	0,00
4	4,5	25,5	114,8	20,3	650,5	22,3	0,124	10,05	0,04
5	4,3	20,4	87,8	18,5	416,5	23,2	0,138	7,92	0,00
6	4,2	25,0	104,9	17,6	624,2	23,7	0,053	1,73	0,01
7	4,3	19,3	83,1	18,5	373,3	23,2	0,202	15,24	0,00
8	3,9	23,2	90,4	15,2	537,7	25,0	0,078	3,28	0,16
9	3,7	25,3	93,5	13,7	638,2	25,9	0,025	0,39	0,36
10	2,7	34,9	94,4	7,3	1221,4	30,3	0,132	21,32	2,56
11	5,1	20,0	101,9	26,0	399,2	19,7	0,016	0,10	0,64
12	4,4	22,0	96,7	19,4	483,5	22,8	0,036	0,63	0,01
13	5,1	21,8	111,1	26,0	474,7	19,7	0,097	4,49	0,64
14	4,3	22,4	96,2	18,5	500,7	23,2	0,038	0,72	0,00
15	5,3	20,6	109,2	28,1	424,4	18,8	0,088	3,31	1,00
16	4,1	23,0	94,5	16,8	530,8	24,1	0,047	1,15	0,04
17	3,8	23,9	90,7	14,4	570,1	25,4	0,066	2,46	0,25
Итого	**73,0**	**395,2**	**1669,3**	**319,8**	**9397,8**	**395,2**	**1,389**	**84,20**	**6,29**
Среднее	**4,3**	**23,2**	**98,2**	**18,8**	**552,8**	**23,2**	**0,082**	**х**	**х**

Однако, оперируя средними величинами, мы можем столкнуться с ошибками округления. Действительно, как видно, параметры a и b, рассчитанные двумя способами не совпадают. В дальнейшем при решении мы будем использовать значения параметров, полученные при решении системы нормальных уравнений.

Величина коэффициента регрессии b = -4,382 означает, что с ростом уровня безработицы на 1% величина среднедушевых денежных доходов населения сократится в среднем на 4,382 тыс. руб.

2) Средний коэффициент эластичности для линейной регрессии находится по формуле (1.9):

$$\bar{Э} = -4{,}382 \cdot \frac{4{,}3}{23{,}2} = 0{,}812.$$

При увеличении уровня безработицы на 1% величина среднедушевых денежных доходов населения сократится в среднем на 0,812%.

3) Линейный коэффициент парной корреляции *(r)* определяется по формуле (1.5), где $\sigma_x = \sqrt{18{,}8 - 4{,}3^2}$ = 0,557, $\sigma_y = \sqrt{552{,}8 - 23{,}2^2}$ = 3,816.

Тогда $r_{xy} = \dfrac{98{,}2 - 4{,}3 \cdot 23{,}2}{0{,}557 \cdot 3{,}816}$ = 0,734, значит связь между уровнем безработицы и уровнем денежных доходов населения в Центральном федеральном округе РФ обратная сильная.

4) Определим коэффициент детерминации: $r_{xy}^2 = 0{,}734^2 = 0{,}539$.

Таким образом, вариация среднедушевых денежных доходов населения на 53,9% зависит от вариации уровня безработицы, а на остальные (100%-53,9%) 46,1% − от вариации факторов, не включенных в модель.

Подставляя в уравнение регрессии фактические значения x, определим теоретические (расчетные) значения $\tilde{y}$ (таблица 1.5) и найдем величину средней ошибки аппроксимации ($\overline{A}$) по формуле (1.8):

$$\overline{A} = \frac{1}{17} \cdot 1{,}389 \cdot 100\% = 8{,}17\%.$$

Так как допустимый предел значений $\overline{A}$ не более 8-10%, качество модели по данному показателю удовлетворительное. Однако средняя ошибка аппроксимации не является главным критерием оценки значимости модели.

С помощью F−критерия Фишера оценим статистическую надежность уравнения регрессии:

$$F_{факт} = \frac{r_{xy}^{\ 2}}{1 - r_{xy}^{\ 2}}(n-2) = \frac{0{,}539}{1 - 0{,}539}(17-2) = 17{,}538.$$

$F_{табл}$ = 4,54 при $\alpha = 0{,}05$; $k_1 = m = 1$; $k_2 = n - m - 1 = 15$.

Так как $F_{факт} > F_{табл}$, уравнение регрессии значимо, статистически надежно.

5) Если существенность уравнения регрессии $\tilde{y} = 42{,}064 - 4{,}382 \cdot x$ доказана, то оно используется для составления прогнозов.

Рассчитаем прогнозное значение уровня безработицы *(x)*:

$$x_p = \bar{x} - 0{,}05 \cdot \bar{x} = 4{,}3 - 0{,}05 \cdot 4{,}3 = 4{,}1\%.$$

Точечный прогноз $\tilde{y}_p$ определим путем подстановки прогнозного значения x_p в уравнение регрессии:

$$\tilde{y}_p = a + bx_p = 42{,}064 - 4{,}382 \cdot 4{,}1 = 24{,}1 \text{ тыс. руб.}$$

Среднюю ошибку прогнозируемого значения находим по формуле:

$$m_{\tilde{y}} = \sigma_{ост} \cdot \sqrt{1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{\sum (x - \bar{x})^2}} = 2{,}37\sqrt{1 + \frac{1}{17} + \frac{(4{,}1 - 4{,}3)^2}{6{,}29}} = 2{,}44,$$

где $\sigma_{ост.} = \sqrt{\dfrac{\sum (y - \tilde{y}_x)^2}{n - m - 1}} = \sqrt{\dfrac{84{,}20}{17 - 1 - 1}} = 2{,}37$.

Определим предельную ошибку прогноза:

$$\Delta_{\tilde{y}_p} = t_{стьюд.табл.} \cdot m_{\tilde{y}} = 2{,}131 \cdot 2{,}37 = 5{,}05,$$

где $t_{стьюд.табл.} = 2{,}131$ (при $df = n\text{-}m\text{-}1 = 15$ и $\alpha = 0{,}05$).

Доверительный интервал прогноза составит:

$$\gamma_{\tilde{y}_p} = \tilde{y}_p \pm \Delta_{\tilde{y}_p} = 24{,}1 \pm 5{,}05; \qquad 19{,}05 < \gamma_{\tilde{y}_p} < 29{,}15.$$

Таким образом, с вероятностью 0,95 доверительный интервал ожидаемого значения среднедушевых денежных доходов населения в Центральном федеральном округе Российской Федерации в предположении сокращения уровня безработицы на 5% от своего среднего уровня составит от 19,05 до 29,15 тыс. руб.

Решение задачи с помощью MS Excel

I. Параметры линейной регрессии $\tilde{y} = a + b \cdot x$ можно определить с помощью встроенной статистической функции **ЛИНЕЙН** MS Excel. Порядок вычисления следующий:

1) введите исходные данные, как показано на рисунке 1.2.

	A	B	C
1	Регионы	x	y
2	Белгородская область	4,0	25,4
3	Брянская область	5,0	22,0
4	Владимирская область	4,3	20,6
5	Воронежская область	4,5	25,5
6	Ивановская область	4,3	20,4
7	Калужская область	4,2	25,0
8	Костромская область	4,3	19,3
9	Курская область	3,9	23,2
10	Липецкая область	3,7	25,3
11	Московская область	2,7	34,9
12	Орловская область	5,1	20,0
13	Рязанская область	4,4	22,0
14	Смоленская область	5,1	21,8
15	Тамбовская область	4,3	22,4
16	Тверская область	5,3	20,6
17	Тульская область	4,1	23,0
18	Ярославская область	3,8	23,9

Рисунок 1.2 Ввод данных для корреляционно-регрессионного анализа

2) выделите область пустых ячеек 5×2 (5 строк, 2 столбца) с целью вывода результатов регрессионной статистики или область 1×2 – для получения только оценок коэффициентов регрессии;

3) активизируйте ***Мастер функций;***

4) в окне «Категория» выберите ***Статистические,*** в окне «Функция» – **ЛИНЕЙН**. Щелкните по кнопке **ОК** (рисунок 1.3);

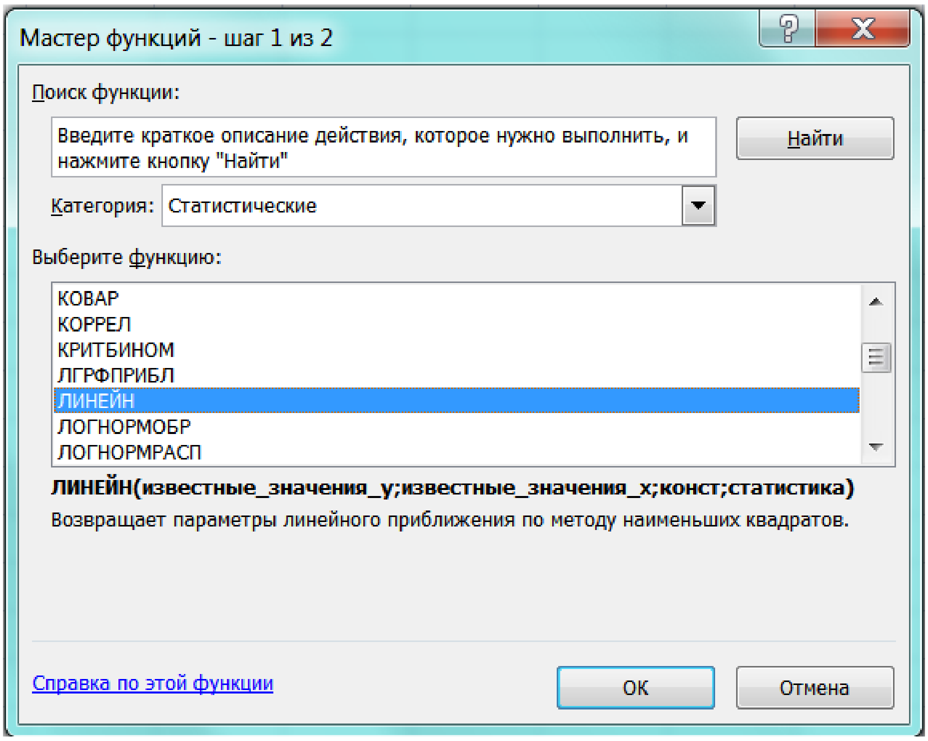


Рисунок 1.3 Диалоговое окно ***Мастер функций***

5) заполните аргументы функции (рисунок 1.4):

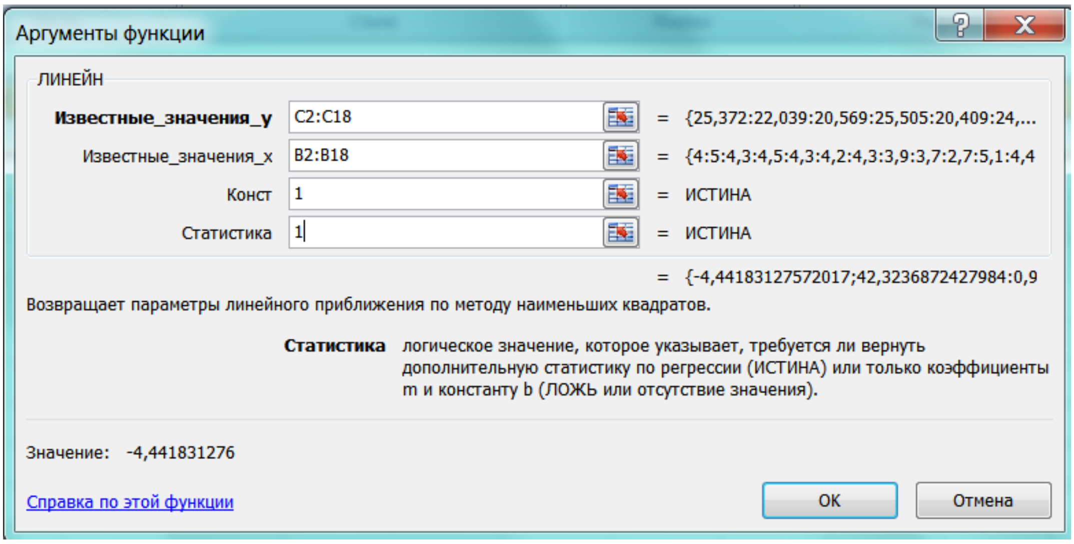


Рисунок 1.4 Диалоговое окно ***Аргументы функции***

Известные_значения_ у – диапазон, содержащий данные результативного признака.

Известные_значения_х – диапазон, содержащий данные факторного признака.

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

Статистика – логическое значение, которое указывает, выводить дополнительную информацию по регрессионному анализу или нет. При *Статистике = 1* дополнительная информация выводится, а при *Статистике = 0* – выводятся только оценки параметров уравнения.

Щелкните по кнопке **ОК**;

6) в левой верхней ячейке выделенной области появится первый элемент итоговой таблицы. Чтобы раскрыть всю таблицу, нажмите на клавишу <**F2**>, а затем – на комбинацию клавиш <**CTRL**> + <**SHIFT**> + <**ENTER**>.

Дополнительная регрессионная статистика будет выводиться в порядке, указанном в следующей схеме:

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

Результаты вычислений функции **ЛИНЕЙН** представлены на рисунке 1.5.

=ЛИНЕЙН(C2:C18;B2:B18;1;1)		
0,945	4,097	
0,596	2,369	
22,106	15,000	
124,089	84,202	

Рисунок 1.5 Результаты вычислений функции **ЛИНЕЙН**

II. С помощью инструмента анализа данных ***Регрессия***, помимо результатов регрессионной статистики, дисперсионного анализа и доверительных интервалов, можно получить остатки и графики подбора линии регрессии, остатков и нормальной вероятности. Порядок действий следующий:

1) в главном меню выберите ***Данные / Анализ данных / Регрессия***. Щелкните по кнопке **ОК**;

2) после вызова режима ***Регрессия*** на экране появляется диалоговое окно (рисунок 1.6), в котором задаются следующие параметры:

Входной интервал Y – диапазон адресов ячеек, содержащих значения y.

Входной интервал X – диапазон адресов ячеек, содержащих значения *x*.

Метки – флажок, который указывает, содержит ли первая строка название столбца или нет.

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

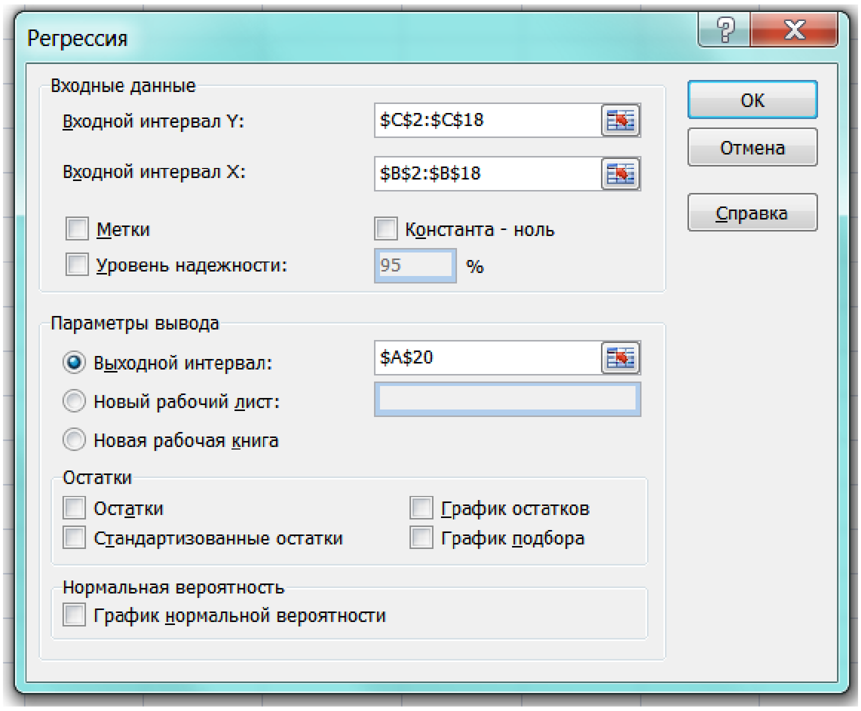


Рисунок 1.6 Диалоговое окно режима ***Регрессия***

Константа-ноль – флажок, указывающий на наличие или отсутствие свободного члена в уравнении (при включении этого параметра коэффициент $a = 0$).

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

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

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

Если необходимо получить информацию и графики остатков, установите соответствующие флажки в диалоговом окне. Щелкните по кнопке **ОК**.

Результаты регрессионного анализа для исходных данных представлены на рисунке 1.7.

ВЫВОД ИТОГОВ

Регрессионная статистика

Множественный R	0,772
R-квадрат	0,596
Нормированный R-квадрат	0,569
Стандартная ошибка	2,369
Наблюдения	17

Дисперсионный анализ

	df	*SS*	*MS*	*F*	*Значимость F*
Регрессия	1	124,089	124,089	22,106	0,0003
Остаток	15	84,202	5,613		
Итого	16	208,291			

	Коэф-фициенты	*Стандартная ошибка*	*t-статистика*	*P-Значение*	*Нижние 95%*	*Верхние 95%*
Y-пересечение	42,324	4,097	10,330	3E-08	33,590	51,057
Переменная X 1	-4,442	0,945	-4,702	0,0003	-6,455	-2,428

Рисунок 1.7 Результат применения инструмента ***Регрессия***

Ниже объясняются используемые в MS Excel терминология и определения.

Регрессионная статистика:

Множественный R – коэффициент корреляции.

$R-$*квадрат* – коэффициент детерминации R^2.

Нормированный $R-$*квадрат* – скорректированный коэффициент детерминации.

Стандартная ошибка – оценка среднеквадратического отклонения.

Наблюдения – число наблюдений n.

Дисперсионный анализ:

df – число степеней свободы. Для строки *Регрессия* показатель равен m, то есть числу параметров при переменных x; для строки *Остаток* – (n-m-*1*); для строки *Итого* – *(*n*-1)*.

SS – сумма квадратов отклонений.

$MS-$дисперсии на одну степень свободы, вычисленные по формуле: $MS = \dfrac{SS}{df}$.

F – значение $F-$критерия Фишера.

Значимость F – значение уровня значимости. Чем меньше уровень значимости, тем больше вероятность того, что вычисленная регрессия значима.

Перейдем к следующей группе показателей.

Коэффициенты – вычисленные значения параметров уравнения регрессии.

Стандартная ошибка – стандартные отклонения значений коэффициентов a и b.

$t-$*статистика* – значения t – критериев Стьюдента.

P – значение – содержит вероятности случайных событий. Если эта вероятность меньше принятого уровня значимости α, то принимается гипотеза о значимости соответствующего коэффициента регрессии.

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

Перейдем к следующей группе показателей:

Наблюдение – содержит номера наблюдений.

Предсказанное У – значения $\tilde{y}$, вычисленные по построенному уравнению регрессии.

Остатки – значения остатков $(y-\tilde{y})$.

Пример 1.2

Зависимость объема продаж y (тыс. долл.) от расходов на рекламу *x* (тыс. долл.) характеризуется по 12 предприятиям концерна следующим образом:

Уравнение регрессии $\tilde{y} = 10{,}6 + 0{,}6x$

Среднее квадратическое отклонение *x* $\sigma_x = 4{,}7$

Среднее квадратическое отклонение *y* $\sigma_y = 3{,}4$

Задание:

1) Определить коэффициент корреляции.
2) Построить таблицу дисперсионного анализа для оценки значимости уравнения регрессии в целом.
3) Найти стандартную ошибку оценки коэффициента регрессии.
4) Оценить значимость коэффициента регрессии через *t*-критерий Стьюдента.
5) Определить доверительный интервал для коэффициента регрессии с вероятностью 95,0% и сделать экономический вывод.

РЕШЕНИЕ:

1) Определим коэффициент корреляции по формуле:

$r = b\dfrac{\sigma_x}{\sigma_y} = 0{,}6\dfrac{4{,}7}{3{,}4} = 0{,}829$, значит связь между признаками прямая сильная.

2) Таблица дисперсионного анализа:

Вариация результата y	Число степеней свободы	Сумма квадратов отклонений, S	Дисперсия на одну степень свободы, D	$F_{факт.}$	$F_{табл.}$
Общая	df = n-1=11	138,72	12,61		
Факторная	df =m=1	95,33	95,33	21,97	4,96
Остаточная	df =n-m-1=10	43,39	4,34		

Среднее квадратическое отклонение находится по формуле:

$$\sigma_y = \sqrt{\frac{\sum(y-\bar{y})^2}{n}};$$

$$\sum(y-\bar{y})^2 = \sigma_y{}^2 \cdot n = 3{,}4^2 \cdot 12 = 138{,}72.$$

Долю дисперсии, объясняемую регрессией, в общей дисперсии результативного признака *y* характеризует коэффициент детерминации:

$$r^2 = \frac{\sum(\tilde{y}-\bar{y})^2}{\sum(y-\bar{y})^2}\text{, откуда } \sum(\tilde{y}-\bar{y})^2 = r^2 \cdot \sum(y-\bar{y})^2 = 0{,}829^2 \cdot 138{,}72 = 95{,}33.$$

$$F_{факт} = \frac{\sum(\tilde{y}-\bar{y})^2 / m}{\sum(y-\tilde{y})^2 / (n-m-1)} = \frac{95{,}33}{4{,}34} = 21{,}97.$$

Так как $F_{факт} > F_{табл}$, то уравнение регрессии значимо и статистически надежно.

3) Стандартная ошибка коэффициента регрессии определяется по формуле:

$$m_b = \sqrt{\frac{\sum(y-\tilde{y})^2 / (n-2)}{\sum(x-\bar{x})^2}} = \sqrt{\frac{D_{ост.}}{\sum(x-\bar{x})^2}} = \frac{D_{ост.}}{\sigma_x \cdot \sqrt{n}} = \frac{4{,}34}{4{,}7 \cdot \sqrt{12}} = 0{,}27.$$

4) Для оценки существенности коэффициента регрессии его величина сравнивается с его стандартной ошибкой, т.е. определяется фактическое значение *t*-критерия Стьюдента: $t_b = \frac{b}{m_b}$, которое затем сравнивается с табличным значением при определенном уровне значимости и числе степеней свободы (n-2).

$$t_b = \frac{b}{m_b} = \frac{0{,}6}{0{,}27} = 2{,}2222, \quad t_{табл.} = 2{,}2281 \text{ (при } df = 10;\ \alpha = 0{,}05).$$

Так как $t_b < t_{табл.}$, значит, признается случайная природа формирования коэффициента регрессии *b*.

5) Доверительный интервал для коэффициента регрессии определяется по формуле: $b \pm t \cdot m_b$.

$$b - t \cdot m_b \le \gamma_b \le b + t \cdot m_b;$$

$$0{,}6 - 2{,}2281 \cdot 0{,}27 \le \gamma_b \le 0{,}6 + 2{,}2281 \cdot 0{,}27;$$

$$0 \le \gamma_b \le 1{,}2.$$

1.3 Задания для самостоятельной работы

Задача 1.1

Имеются следующие социально-экономические и демографические категории:

- валовой региональный продукт на душу населения;
- величина прожиточного минимума:

- среднедушевые денежные доходы населения;
- среднемесячная номинальная начисленная заработная плата;
- доля населения трудоспособного возраста в общей численности населения;
- доля пенсионеров в общей численности населения;
- доля городского населения в общей численности населения;
- уровень занятости;
- уровень безработицы;
- общий коэффициент брачности;
- общий коэффициент рождаемости;
- индекс потребительских цен;
- оборот розничной торговли на душу населения;
- инвестиции в основной капитал.

Задание:

1) Установить взаимосвязь между перечисленными категориями. Результаты представить в виде таблицы 1.5

Таблица 1.5 Исходные данные для эконометрического моделирования

№ модели	Результативный признак у, ед. изм.	Факторный признак х, ед.изм.	Направление связи (прямая / обратная)
1			
2			
3			
4			
5			

2) По одному из федеральных округов Российской Федерации собрать статистические данные с сайта Росстата (Статистический сборник «Регионы России. Социально-экономические показатели») по каждой модели из таблицы 1.6. Статистические данные скопировать в MS Excel (рисунок 1.8).

3) По исходным данным каждой модели:

а) построить поле корреляции и сформулировать гипотезу о форме связи;

б) определить параметры уравнения парной линейной регрессии и дать интерпретацию коэффициента регрессии.

в) рассчитать линейный коэффициент корреляции и коэффициент детерминации, пояснить их смысл.

г) с вероятностью 95,0% оценить статистическую значимость коэффициента регрессии b и уравнения регрессии в целом. Сделать выводы.

Все необходимые расчеты проводить с использованием формул и с помощью MS Excel.

	A	B	C
1	**Модель 1**		
2	Регионы	Денежные доходы, руб.	ВРП на душу населения, руб.
3	Республика Башкортостан	27744	306771,3
4	Республика Марий Эл	18533	209488,1
5	Республика Мордовия	17878	210858,7
6	Республика Татарстан	32163	434509,1
7	Удмуртская Республика	24465	291287,5
8	Чувашская Республика	18492	189736,4
9	Пермский край	32053	367086,6
10	Кировская область	22170	191444,5
11	Нижегородская область	30837	310866,4
12	Оренбургская область	22948	364761,5
13	Пензенская область	21829	219181,9
14	Самарская область	27732	358648,8
15	Саратовская область	20070	225374,5
16	Ульяновская область	22782	220575,7

Рисунок 1.8 Исходные данные для моделирования

Задача 1.2

По данным Приволжского федерального округа Российской Федерации изучается зависимость среднедушевых денежных доходов от факторов (таблица 1.6).

Таблица 1.6 Доходы населения и определяющие их факторы

Регионы	Средне-душевые денежные доходы, тыс.руб.	Доля городского населения, %	Доля населения в трудо-способном возрасте, %	ВРП на душу, тыс.руб.	Уровень занятос-ти, %	Средне-месячная заработная плата, тыс.руб.
Республика Башкортостан	27,7	61,8	57,3	306,8	62,6	25,9
Республика Марий Эл	18,5	65,5	56,8	209,5	65,1	21,9
Республика Мордовия	17,9	61,9	58,2	210,9	67,9	22,0
Республика Татарстан	32,2	76,4	57,6	434,5	68,5	29,1
Удмуртская Республика	24,5	65,6	56,3	291,3	68,5	24,7
Чувашская Республика	18,5	61,3	57,5	189,7	68,3	21,4
Пермский край	32,1	75,6	56,3	367,1	61,8	28,5
Кировская область	22,2	75,9	54,6	191,4	65	22,1
Нижегородская область	30,8	79,5	56,6	310,9	67,5	26,5
Оренбургская область	22,9	59,9	56,4	364,8	64,3	24,6
Пензенская область	21,8	68,3	55,9	219,2	64,2	23,2
Самарская область	27,7	80,2	57,1	358,6	68,6	26,8
Саратовская область	20,1	75,3	56,9	225,4	62,9	22,5
Ульяновская область	22,8	74,7	56,6	220,6	63,3	22,8

Задание:

1) Построить поле корреляции и сформулировать гипотезу о форме связи среднедушевых доходов населения с каждым из факторов.

2) Определить параметры уравнения парной линейной регрессии и дать интерпретацию коэффициента регрессии.

3) Рассчитать линейный коэффициент корреляции и коэффициент детерминации, пояснить их смысл.

4) С вероятностью 95,0% оценить статистическую значимость коэффициента регрессии b и уравнения регрессии в целом. Сделать выводы.

Задача 1.3

Имеются данные о зависимости объема продаж (y) фирмы от затрат на рекламу (x) (усл. ден. ед.):

x	5	8	6	5	4	9	12
y	72	76	78	70	68	80	82

Уравнение регрессии имеет вид: $\tilde{y} = 64{,}25 + 1{,}59 \cdot x$; $\sum (y - \tilde{y})^2 = 34{,}12$.

Задание:

1) Определить объем продаж фирмы, если по прогнозу затраты на рекламу составят 13,5 усл. ед.

2) Найти стандартную ошибку прогноза и 99%-ный доверительный интервал прогноза.

Задача 1.4

По данным обследования 150 безработных получены следующие результаты:

Показатель	Средние значения	Коэффициент вариации, %
Возраст, лет	37	14
Продолжительность поиска работы, мес.	7	21

Известно также, что коэффициент корреляции между признаками составил 0,76.

Задание:

1) Определить, на сколько дней в среднем увеличивается продолжительность поиска работы с ростом возраста безработных на 1 год.

2) Оценить существенность рассматриваемой связи с помощью F-критерия Фишера.

3) Определить, в какой мере вариация продолжительности поиска работы обусловлена вариацией возраста.

Задача 1.5

По данным регионов Центрального федерального округа за 2015 год изучается зависимость уровня безработицы (%) от выпуска бакалавров, специалистов, магистров (тыс. чел.):

Уравнение регрессии: $\tilde{y} = 5{,}620 - 0{,}069x$

F-критерий Фишера равен 7,15.

Задание:

1) Оценить, на сколько процентов вариация уровня безработицы зависит от вариации выпуска бакалавров, специалистов, магистров.

2) Дать интерпретацию коэффициента регрессии.

3) Дать точечный прогноз уровня безработицы по совокупности регионов, если выпуск бакалавров, специалистов, магистров составит 13 тыс. чел.

Задача 1.6

Пусть имеется следующая модель регрессии, характеризующая зависимость y от x: $y = 10{,}80 - 7{,}25x$. Известно также, что $r_{xy} = -0{,}65$; $n = 27$.

Задание:

1) Построить доверительный интервал для коэффициента регрессии в этой модели:

а) с вероятностью 90%;

б) с вероятностью 99%.

2) Проанализировать результаты, пояснить причины их различий.

Задача 1.7

Имеются данные об урожайности зерновых и себестоимости зерновых культур в одном из регионов Российской Федерации:

№ хозяйств	Урожайность зерновых, ц с 1 га	Себестоимость 1 ц зерновых культур, руб.
1	18	900
2	19	850
3	20	840
4	22	770
5	24	720
6	26	700

Задание:

1) Обосновать выбор уравнения регрессии.

2) Рассчитать параметры уравнения регрессии.

3) Сделать выводы.

Задача 1.8

Модель регрессии, характеризующая зависимость среднедушевых денежных доходов населения от ВРП на душу населения в Приволжском федеральном округе Российской Федерации, имеет вид: $\tilde{y} = 9{,}50 + 0{,}05x$. Известно также, что коэффициент корреляции равен 0,84, число регионов – 14.

Задание:

1) Построить доверительный интервал для коэффициента регрессии в этой модели:

а) с вероятностью 90%;

б) с вероятностью 99%.

2) Проанализировать результаты, пояснить причины их различий.

Задача 1.9

Имеются данные об урожайности зерновых и дозе внесения удобрений по регионам Приволжского федерального округа Российской Федерации:

Регион	Урожайность зерновых с 1 га, ц	Внесено минеральных удобрений на 1 га посева, кг
Республика Башкортостан	13,9	14,9
Республика Марий Эл	15,7	13,7
Удмуртская Республика	17,0	15,6
Пермский край	16,0	15,3
Кировская область	21,1	25,1
Самарская область	17,7	20,0
Ульяновская область	20,2	23,7

Задание:

1) Выбрать модель уравнения регрессии.

2) Рассчитать параметры уравнения регрессии.

3) Определить среднюю ошибку аппроксимации.

4) Сделать выводы.

1.4 Тестовые задания

1 Линейная регрессия имеет вид:

а) $y = a + \frac{b}{x} + \varepsilon$; б) $y = a + bx + \varepsilon$; в) $y = a \cdot b^x \cdot \varepsilon$.

2 Связь между урожайностью и дозой внесения удобрений характеризуется уравнением $\tilde{y} = 10{,}655 + 0{,}4076x$. Выберите правильный ответ.

а) при увеличении дозы внесения удобрений на 1кг, урожайность в среднем повысится на 10,655 ц с 1 га;

б) при увеличении дозы внесения удобрений на 1кг, урожайность в среднем повысится на 0,4076 ц с 1 га;

в) при увеличении дозы внесения удобрений на 0,407 кг, урожайность в среднем повысится на 1 ц с 1 га.

3 Оценку качества модели дает:

а) коэффициент регрессии;
б) коэффициент эластичности;
в) средняя ошибка аппроксимации.

4 Коэффициент корреляции может принимать значения:
а) $-1 \le r \le 1$; б) $r \le 1$; в) $r \ge 1$.

5 Тесноту связи определяют с помощью:
а) коэффициента корреляции;
б) коэффициента эластичности;
в) коэффициента детерминации;
г) коэффициента регрессии.

6 Коэффициент корреляции равен r =0,432. Охарактеризуйте связь между признаками.
а) связь прямая средняя;
б) связь прямая слабая;
в) связь обратная средняя.

7 Значимость уравнения регрессии в целом оценивается с помощью:
а) F – критерия Фишера;
б) коэффициента корреляции.

8 Задача дисперсионного анализа состоит:
а) в анализе дисперсии зависимой переменной;
б) в анализе дисперсии независимой переменной.

9 Какой коэффициент не зависит от принятых единиц измерения?
а) коэффициент регрессии;
б) коэффициент корреляции.

10 F-критерий Фишера определяется по формуле:
а) $F = \frac{D_{факт.}}{D_{ост.}}$; б) $F = \frac{D_{ост.}}{D_{общ.}}$; в) $F = \frac{D_{факт.}}{D_{общ.}}$.

11 В уравнении $y = a + bx + \varepsilon$ коэффициентом регрессии является:
а) b; б) a.

12 Система нормальных уравнений метода наименьших квадратов для прямой линии имеет вид:
а) $\begin{cases} a + b\sum x = \sum y; \\ a\sum x + b\sum x^2 = \sum xy \end{cases}$; б) $\begin{cases} na + b\sum x = \sum y; \\ a\sum x + \sum x^2 = \sum xy \end{cases}$; в) $\begin{cases} na + b\sum x = \sum y; \\ a\sum x + b\sum x^2 = \sum xy \end{cases}$.

13 Коэффициент регрессии показывает:

а) среднее изменение зависимой переменной $\tilde{y}_x$ при изменении независимой переменной x на единицу;

б) на сколько процентов изменится в среднем результат, если фактор изменится на 1%.

14 Коэффициент корреляции вычисляется по формуле:

а) $r = \dfrac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x \cdot \sigma_y}$; б) $r = \dfrac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sum (x - \bar{x})^2}$;

в) $r = \dfrac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x}$; г) $r = b \cdot \dfrac{\sigma_x}{\sigma_y}$.

15 Коэффициент детерминации показывает:

а) на сколько процентов вариация результативного признака определяется изменением факторов, включенных в модель;

б) на сколько процентов изменится в среднем результативный признак, если факторный изменится на 1%.

16 Выберите правильный ответ:

а) по значению коэффициента регрессии нельзя судить о тесноте связи;

б) по значению коэффициента регрессии можно судить о тесноте связи.

17 Коэффициент корреляции ввел:

а) И. Фишер; б) К. Пирсон; в) Я. Тинберген.

18 Если сумма квадратов отклонений, обусловленная регрессией ($\sum (\tilde{y}_x - \bar{y})^2$), будет больше остаточной суммы квадратов отклонений ($\sum (y - \tilde{y}_x)^2$), то:

а) уравнение регрессии статистически значимо;

б) уравнение регрессии статистически незначимо;

в) фактор x оказывает существенное воздействие на результат y.

19 Если параметр a в уравнении регрессии больше нуля ($a>0$), то:

а) вариация результата больше вариации фактора ($V_x < V_y$);

б) вариация результата меньше вариации фактора ($V_x > V_y$).

20 При оценке значимости уравнения регрессии выдвигается нулевая гипотеза, что:

а) фактор x существенно влияет на результат y;

б) коэффициент регрессии равен нулю, т.е. $b=0$;

в) фактор x не оказывает влияния на результат y.

1.5 Контрольные вопросы для самопроверки

1) Как оцениваются параметры уравнения линейной регрессии, что обозначают?
2) В чем заключается сущность метода наименьших квадратов?
3) Что такое число степеней свободы и как оно определяется для факторной и остаточной сумм квадратов отклонений?
4) Как оценивается значимость уравнения регрессии?
5) В чем смысл средней ошибки аппроксимации и как она определяется?
6) Что показывает коэффициент эластичности?
7) Как определить показатель тесноты связи для линейной регрессии?
8) Для чего проводится дисперсионный анализ результатов регрессии?
9) Как определяется коэффициент детерминации и что он показывает?
10) Как применяется уравнение парной линейной регрессии для прогнозирования?

2 НЕЛИНЕЙНАЯ РЕГРЕССИЯ И КОРРЕЛЯЦИЯ

Цель занятия – ознакомиться с основными понятиями парной нелинейной регрессии и корреляции, овладеть навыками определения параметров нелинейной регрессии и корреляции с использованием формул и табличного процессора MS Excel.

2.1Теоретические положения

2.1.1 Классы нелинейных регрессий. Приведение нелинейных регрессий к линейному виду.

2.1.2 Расчет средних коэффициентов эластичности.

2.1.3 Корреляция в нелинейной регрессии. Оценка существенности уравнения нелинейной регрессии.

2.1.1 Классы нелинейных регрессий. Приведение нелинейных регрессий к линейному виду

Если между экономическими явлениями существуют нелинейные соотношения, то они выражаются с помощью соответствующих нелинейных функций. Различают два класса нелинейных регрессий (рисунок 2.1).

ДВА КЛАССА НЕЛИНЕЙНЫХ РЕГРЕССИЙ

1) Регрессии, нелинейные относительно включенных в анализ объясняющих переменных, но линейные по оцениваемым параметрам

2) Регрессии, нелинейные по оцениваемым параметрам

П Р И М Е Р Ы

Полиномы различных степеней

$$y_x = a + b \cdot x + c \cdot x^2$$

$$y_x = a + b \cdot x + c \cdot x^2 + d \cdot x^3$$

Равносторонняя гипербола

$$y_x = a + b/x$$

Полулогарифмическая функция

$$y_x = a + b \cdot \ln x$$

Степенная

$$y_x = a \cdot x^b$$

Показательная

$$y_x = a \cdot b^x$$

Экспоненциальная

$$y_x = \mathrm{e}^{a + b \cdot x}$$

Рисунок 2.1 Классы нелинейных регрессий

Регрессии нелинейные по включенным переменным приводятся к линейному виду простой заменой переменных, а дальнейшая оценка параметров производится с помощью метода наименьших квадратов.

Например, парабола второй степени $\tilde{y} = a + b \cdot x + c \cdot x^2$ приводится к линейному виду с помощью замены: $x = x_1$, $x^2 = x_2$. В результате получают двухфакторное уравнение регрессии $\tilde{y} = a + b \cdot x_1 + c \cdot x_2$, параметры которого определяют с помощью МНК.

Парабола второй степени обычно применяется в тех случаях, когда меняется характер связи рассматриваемых признаков: прямая связь меняется на обратную или обратная на прямую.

Равносторонняя гипербола $\tilde{y} = a + \frac{b}{x}$ приводится к линейному виду простой заменой: $z = \frac{1}{x}$.

Уравнение гиперболы используется для характеристики связи:

- удельных расходов сырья, материалов, топлива от объема выпускаемой продукции;
- времени обращения товаров от величины товарооборота;
- процента прироста заработной платы от уровня безработицы (например, кривая А.В. Филлипса);
- расходов на непродовольственные товары от доходов или общей суммы расходов (например, кривые Э. Энгеля) и в других случаях.

Аналогичным образом приводятся к линейному виду зависимости $y_x = a + b \cdot \ln x$, $y_x = a + b \cdot \sqrt{x}$ и другие.

Иначе обстоит дело с регрессиями нелинейными по оцениваемым параметрам, которые делятся на два типа: нелинейные модели внутренне линейные (приводятся к линейному виду с помощью соответствующих преобразований, например, логарифмированием) и нелинейные модели внутренне нелинейные (к линейному виду не приводятся).

К внутренне линейным моделям относятся, например, степенная функция $y_x = a \cdot x^b$, показательная $y_x = a \cdot b^x$, экспоненциальная $y_x = \mathrm{e}^{a+b\cdot x}$, логистическая $y_x = \frac{a}{1 + b \cdot \mathrm{e}^{-c\cdot x}}$, обратная $y_x = \frac{1}{a + b \cdot x}$.

К внутренне нелинейным моделям можно, например, отнести следующие модели: $y_x = a + b \cdot x^c$, $y_x = a \cdot \left(1 - \frac{1}{1 - x^b}\right)$.

2.1.2 Расчет средних коэффициентов эластичности

Широкое использование степенной функции связано с тем, что параметр b в ней имеет четкое экономическое истолкование, так как он является коэффициентом эластичности:

$$Э = f'(x) \cdot \frac{x}{y}. \tag{2.1}$$

В таблице 2.1 приведены формулы для расчета средних коэффициентов эластичности для наиболее часто используемых типов уравнений регрессии.

Таблица 2.1 Расчет средних коэффициентов эластичности

Вид функции, y	Первая производная, y'	Средний коэффициент эластичности, $\overline{Э}$
$y = a + b \cdot x + \varepsilon$	b	$\frac{b \cdot \overline{x}}{a + b \cdot \overline{x}}$
$y = a + b \cdot x + c \cdot x^2 + \varepsilon$	$b + 2c \cdot x$	$\frac{(b + 2c \cdot \overline{x}) \cdot \overline{x}}{a + b \cdot \overline{x} + c \cdot \overline{x}^2}$
$y = a + \frac{b}{x} + \varepsilon$	$-\frac{b}{x^2}$	$-\frac{b}{a \cdot \overline{x} + b}$
$y = a \cdot x^b \cdot \varepsilon$	$a \cdot b \cdot x^{b-1}$	b
$y = a \cdot b^x \cdot \varepsilon$	$a \cdot \ln b \cdot b^x$	$\overline{x} \cdot \ln b$
$y = a + b \cdot \ln x + \varepsilon$	$\frac{b}{x}$	$\frac{b}{a + b \cdot \ln \overline{x}}$
$y = \frac{a}{1 + b \cdot \mathrm{e}^{-c \cdot x + \varepsilon}}$	$\frac{a \cdot b \cdot c \cdot \mathrm{e}^{-c \cdot x}}{(1 + b \cdot \mathrm{e}^{-c \cdot x})^2}$	$\frac{b \cdot c \cdot \overline{x}}{b + \mathrm{e}^{c \cdot \overline{x}}}$
$y = \frac{1}{a + b \cdot x + \varepsilon}$	$-\frac{b}{(a + b \cdot x)^2}$	$-\frac{b \cdot \overline{x}}{a + b \cdot \overline{x}}$

2.1.3 Корреляция в нелинейной регрессии. Оценка существенности уравнения нелинейной регрессии

Уравнение нелинейной регрессии дополняется показателем тесноты связи – ***индексом корреляции (ρ),*** который находится по формуле:

$$\rho_{xy} = \sqrt{1 - \frac{\sigma^2_{\text{ост}}}{\sigma^2_y}}, \tag{2.2}$$

где $\sigma^2_y = \frac{1}{n}\sum (y - \overline{y})^2$ – общая дисперсия результативного признака y;

$$\sigma^2_{\text{ост}} = \frac{1}{n}\sum \left(y - y_x\right)^2$$ – остаточная дисперсия.

Величина данного показателя находится в пределах: $0 \le \rho_{xy} \le 1$. 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:

$$\rho^2_{xy} = 1 - \frac{\sigma^2_{\text{ост}}}{\sigma^2_y} = \frac{\sigma^2_{\text{объясн}}}{\sigma^2_y}. \tag{2.3}$$

Индекс детерминации используется для проверки существенности уравнения регрессии по F-критерию Фишера:

$$F = \frac{\rho^2_{xy}}{1-\rho^2_{xy}} \cdot \frac{n-m-1}{m}, \tag{2.4}$$

где ρ^2_{xy} – индекс детерминации;

n – число наблюдений;

m – число параметров при переменной x.

Фактическое значение F-критерия сравнивается с табличным 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) [4].

2.2 Решение типовых задач

Пример 2.1

Для характеристики зависимости среднедушевых денежных доходов населения от уровня безработицы в Центральном федеральном округе Российской Федерации (таблица 1.4) с помощью MS Excel рассчитать параметры степенной, экспоненциальной, гиперболической, логарифмической и полиномиальной функций. Найти показатели тесноты связи по каждой модели. Оценить значимость каждой модели с помощью *F*-критерия Фишера и выбрать наилучшую из них (включая линейную).

РЕШЕНИЕ:

1) Регрессия в виде степенной функции имеет вид:

$$y = a \cdot x^b \cdot \varepsilon. \tag{2.5}$$

Для оценки параметров модели линеаризуем (приводим к линейному виду) модель путем логарифмирования: $\ln y = \ln a + b \ln x$.

Обозначим *lny =Y, lna =A, lnx =X.*

Тогда получим*: Y=A+bX.*

Для расчетов составим с помощью MS Excel вспомогательную таблицу, в которой рассчитаем натуральные логарифмы с помощью математической функции **LN** (рисунок 2.2).

	A	B	C	D	E
1	Регионы	*x*	*y*	*X=ln(x)*	*Y=ln(y)*
2	Белгородская область	4,0	25,4	=LN(B2)	3,234
3	Брянская область	5,0	22,0	1,609	3,093
4	Владимирская область	4,3	20,6	1,459	3,024
5	Воронежская область	4,5	25,5	1,504	3,239
6	Ивановская область	4,3	20,4	1,459	3,016
7	Калужская область	4,2	25,0	1,435	3,218
8	Костромская область	4,3	19,3	1,459	2,961
9	Курская область	3,9	23,2	1,361	3,144
10	Липецкая область	3,7	25,3	1,308	3,229
11	Московская область	2,7	34,9	0,993	3,554
12	Орловская область	5,1	20,0	1,629	2,995
13	Рязанская область	4,4	22,0	1,482	3,090
14	Смоленская область	5,1	21,8	1,629	3,081
15	Тамбовская область	4,3	22,4	1,459	3,108
16	Тверская область	5,3	20,6	1,668	3,025
17	Тульская область	4,1	23,0	1,411	3,137
18	Ярославская область	3,8	23,9	1,335	3,173

Рисунок 2.2 Расчет натуральных логарифмов

Далее с помощью команды ***Данные*** / ***Анализ данных*** / ***Регрессия*** рассчитаем параметры уравнения (рисунок 2.3).

Таким образом, уравнение регрессии имеет вид:

$$\ln y = 4{,}165 - 0{,}711 \ln x .$$

Выполнив потенцирование, получим:

$$y = e^{4{,}165} x^{-0{,}711} = 64{,}393 x^{-0{,}711},$$

где $e^{4{,}165}$ можно рассчитать с помощью математической функции **EXP.**

Параметр b=-0,711 означает коэффициент эластичности, который показывает, что с увеличением уровня безработицы на 1% среднедушевые доходы в Центральном федеральном округе сократятся на 0,711%;.

ВЫВОД ИТОГОВ

Регрессионная статистика	
Множественный R	0,806
R-квадрат	0,649
Нормированный R-квадрат	0,626
Стандартная ошибка	0,085
Наблюдения	17

Дисперсионный анализ

	df	*SS*	*MS*	*F*	*Значимость F*
Регрессия	1	0,201	0,201	27,776	9,42E-05
Остаток	15	0,108	0,007		
Итого	16	0,309			

	Коэффициенты	*Стандартная ошибка*	*t-статистика*	*P-Значение*	*Нижние 95%*	*Верхние 95%*
Y-пересечение	4,165217298	0,196266088	21,22229749	1,33356E-12	3,746886037	4,58354856
Переменная X 1	-0,711276176	0,134959188	-5,270305684	9,42235E-05	-0,998934874	-0,423617477

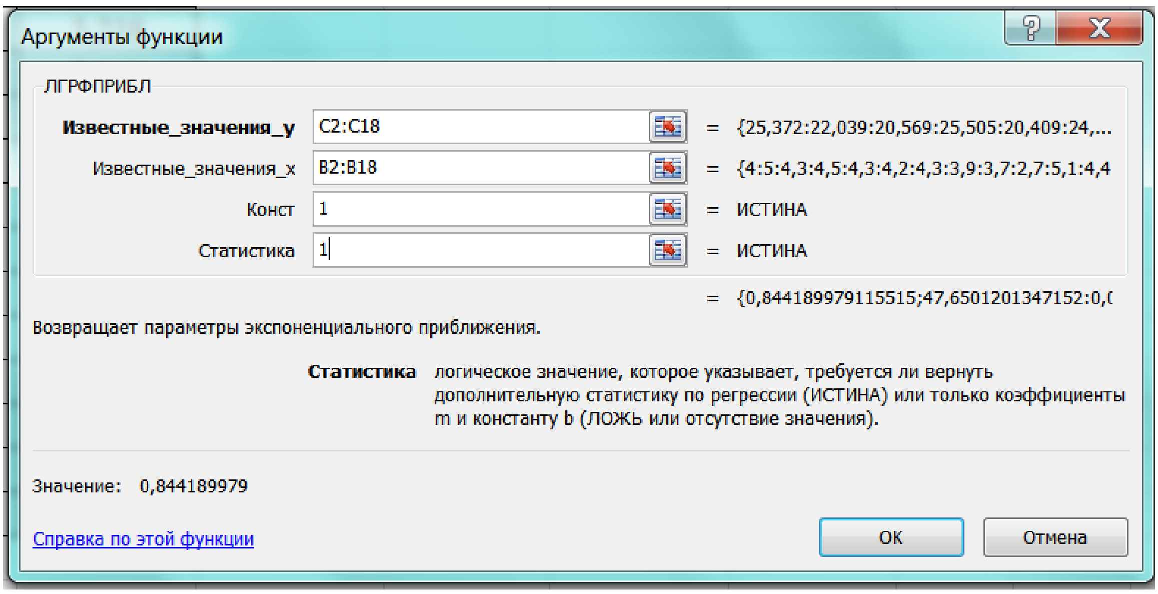
Рисунок 2.3 Результаты расчета параметров степенной функции

2) Регрессия в виде экспоненты имеет вид:

$$y = e^{a+bx}. \qquad (2.6)$$

Для оценки ее параметров необходимо привести уравнение к линейному виду: $\ln y = \ln a + bx$.

Для расчета параметров экспоненциальной прямой можно воспользоваться статистической функцией **ЛГРФПРИБЛ** MS Excel. Порядок вычислений аналогичен применению функции **ЛИНЕЙН** (см. стр. 15). Результаты вычислений представлены на рисунке 2.4.



0,844	47,650
0,037	0,160
0,584	0,093
21,041	15
0,180	0,129

Рисунок 2.4 Результаты вычислений параметров экспоненциальной функции

Таким образом, уравнение регрессии в виде экспоненты имеет вид:

$$\tilde{y} = 47{,}650e^{0{,}844x}.$$

3) Регрессия в виде равносторонней гиперболы имеет вид: $y = a + \frac{b}{x}$, чтобы оценить параметры *a* и *b*, приведем модель к линейному виду, заменив $\frac{1}{x} = z$. Тогда $y = a + bz$. Результаты замены представлены на рисунке 2.5.

	A	B	C	D	E	F
1	Регионы	*x*	*y*	*X=ln(x)*	*Y=ln(y)*	*z=1/x*
2	Белгородская область	4,0	25,4	1,386	3,234	=1/B2
3	Брянская область	5,0	22,0	1,609	3,093	0,20
4	Владимирская область	4,3	20,6	1,459	3,024	0,23
5	Воронежская область	4,5	25,5	1,504	3,239	0,22
6	Ивановская область	4,3	20,4	1,459	3,016	0,23
7	Калужская область	4,2	25,0	1,435	3,218	0,24
8	Костромская область	4,3	19,3	1,459	2,961	0,23
9	Курская область	3,9	23,2	1,361	3,144	0,26
10	Липецкая область	3,7	25,3	1,308	3,229	0,27
11	Московская область	2,7	34,9	0,993	3,554	0,37
12	Орловская область	5,1	20,0	1,629	2,995	0,20
13	Рязанская область	4,4	22,0	1,482	3,090	0,23
14	Смоленская область	5,1	21,8	1,629	3,081	0,20
15	Тамбовская область	4,3	22,4	1,459	3,108	0,23
16	Тверская область	5,3	20,6	1,668	3,025	0,19
17	Тульская область	4,1	23,0	1,411	3,137	0,24
18	Ярославская область	3,8	23,9	1,335	3,173	0,26

Рисунок 2.5 Вспомогательная таблица для расчета параметров гиперболы

Далее с помощью инструмента ***Регрессия*** рассчитаем параметры уравнения. Результаты расчета представлены на рисунке 2.6.

ВЫВОД ИТОГОВ

Регрессионная статистика	
Множественный R	0,860
R-квадрат	0,740
Нормированный R-квадрат	0,722
Стандартная ошибка	1,901
Наблюдения	17

Дисперсионный анализ

	df	*SS*	*MS*	*F*	*Значимость F*
Регрессия	1	154,064	154,064	42,617	0,000
Остаток	15	54,227	3,615		
Итого	16	208,291			

	Коэффициенты	*Стандартная ошибка*	*t-статистика*	*P-Значение*	*Нижние 95%*	*Верхние 95%*
Y-пересечение	5,434	2,768	1,963	0,068	-0,466	11,333
Переменная X 1	74,733	11,448	6,528	0,000	50,333	99,134

Рисунок 2.6 Результаты вычислений параметров гиперболической функции

Таким образом, уравнение гиперболы имеет вид: $\tilde{y} = 5{,}434 + \dfrac{74{,}733}{x}$.

4) Для построения полиномиальной (параболической) функции построим поле корреляции, выполнив команду ***Вставка / Диаграммы / Точечная / Точечная с маркерами*** (рисунок 2.7).

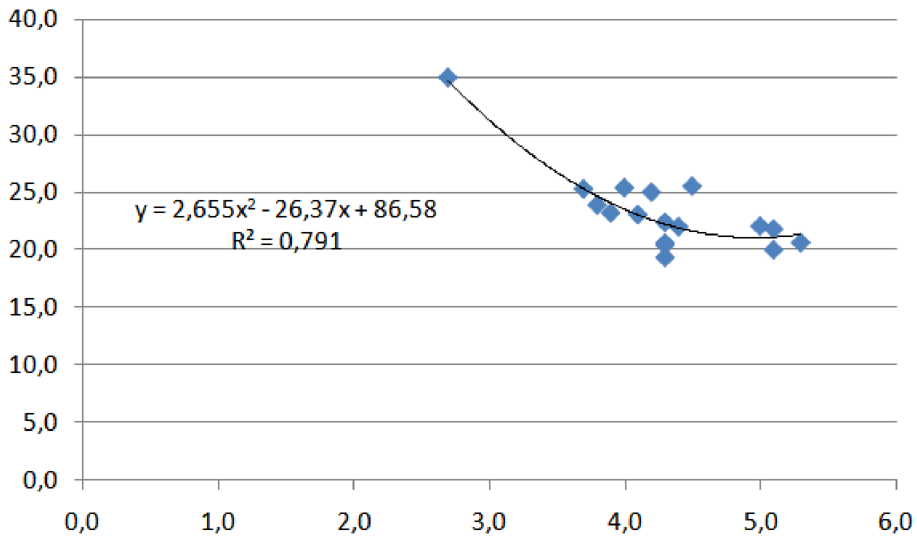


Рисунок 2.7 Полиномиальная зависимость среднедушевых доходов населения от уровня безработицы

Добавим на диаграмму линию тренда. Для этого выполним команду ***Макет / Линия тренда / Дополнительные параметры линии тренда*** (рисунок 2.8).

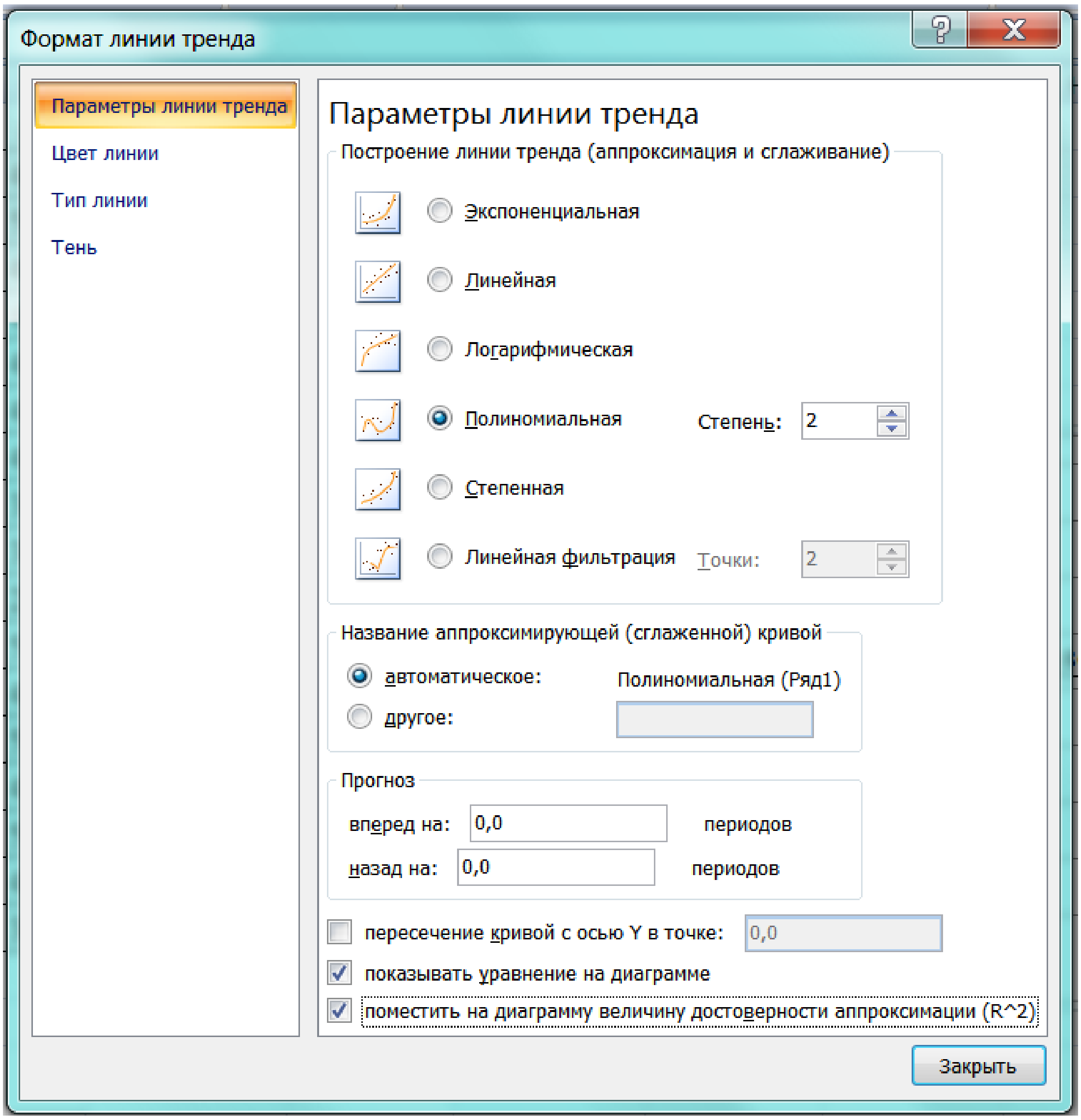


Рисунок 2.8 Построение полиномиального тренда

В результате на диаграмме получили уравнение полиномиальной функции (параболы второго порядка) и значение коэффициента детерминации (рисунок 2.7).

Аналогично строится уравнение логарифмической функции.

Выберем наилучшую модель, для чего объединим результаты построения парных регрессий в сводной таблице 2.2.

Таблица 2.2 Сравнительная оценка результатов регрессии и корреляции

Регрессия	Уравнение регрессии	Коэффициент корреляции	Коэффициент детерминации	F-критерий Фишера
Линейная	$\tilde{y} = 42{,}324 - 4{,}442 \cdot x$	0,772	0,596	22,106
Степенная	$\tilde{y} = 64{,}393x^{-0{,}711}$	0,806	0,649	27,776
Экспоненциальная	$\tilde{y} = 47{,}650e^{0{,}844x}$	0,764	0,584	21,041
Гиперболическая	$\tilde{y} = 5{,}434 + \dfrac{74{,}733}{x}$	0,860	0,740	42,617
Параболическая	$\tilde{y} = 86{,}58 - 26{,}37x + 2{,}655x^2$	0,889	0,791	26,493
Логарифмическая	$\tilde{y} = 50{,}49 - 18{,}81\ln(x)$	0,821	0,675	31,154

Предпочтение можно отдать параболической (полиномиальной) функции, для которой значения коэффициентов корреляции и детерминации наибольшие.

Пример 2.2

Для двух видов продукции А и Б зависимость расходов предприятия y (тыс. руб.) от объема производства x (шт.) характеризуется следующими данными:

Уравнение регрессии	Показатели корреляции	Число наблюдений
$y_a = 160 + 0{,}8x$	0,85	30
$y_b = 50x^{0{,}6}$	0,72	25

Задание:

1) Пояснить смысл величин 0,8 и 0,6 в уравнениях регрессии.

2) Сравнить эластичность расходов от объема производства для продукции А и Б при выпуске продукции А в 500 единиц.

3) Определить, каким должен быть выпуск продукции А, чтобы эластичность ее расходов совпадала с эластичностью расходов на продукцию Б.

4) Оценить значимость каждого уравнения регрессии с помощью F-критерия Фишера.

РЕШЕНИЕ:

1) В линейной функции коэффициент регрессии, равный 0,8, показывает, что при увеличении объема производства на 1 шт., расходы предприятия в среднем увеличатся на 0,8 тыс. руб. или на 800 руб.

В степенной функции коэффициент регрессии является коэффициентом эластичности. Он показывает, что при увеличении объема производства на 1 %, расходы предприятия увеличатся в среднем на 0,6%.

2) Средний коэффициент эластичности определяется по формуле:

$$\overline{Э} = f'(x)\frac{\overline{x}}{\overline{y}}$$

Для линейной функции: $\bar{Э} = b\dfrac{\bar{x}}{\bar{y}} = b\dfrac{\bar{x}}{a + b\bar{x}} = 0{,}8\dfrac{500}{160 + 0{,}8 \cdot 500} = 0{,}71$.

Для степенной функции: $\bar{Э} = 0{,}6$.

3) $Э_a = Э_b$;

$b\dfrac{\bar{x}}{a + b\bar{x}} = 0{,}6$, $\quad 0{,}8\dfrac{x}{160 + 0{,}8 \cdot x} = 0{,}6$, откуда x = 300.

4) $F_{факт.}$ определяется по формуле:

$$F_{факт.} = \frac{r^2}{1 - r^2} \cdot \frac{n - m - 1}{m}.$$

Для линейной функции: $F_{факт.} = \dfrac{0{,}85^2}{1 - 0{,}85^2} \cdot \dfrac{30 - 1 - 1}{1} = 72{,}9$.

$F_{табл.}$ = 4,20 ($\alpha = 0{,}05; \quad k_1 = m = 1; \quad k_2 = n - m - 1 = 30 - 1 - 1 = 28$).

Для степенной функции: $F_{факт} = \dfrac{0{,}72^2}{1 - 0{,}72^2} \cdot \dfrac{25 - 1 - 1}{1} = 24{,}8$.

$F_{табл}$= 4,28 ($\alpha = 0{,}05; \quad k_1 = m = 1; \quad k_2 = n - m - 1 = 25 - 1 - 1 = 23$).

Так как $F_{факт}$ > $F_{табл}$, то уравнения регрессии значимы, статистически надежны.

2.3 Задания для самостоятельной работы

Задача 2.1

При изучении зависимости вида $y = a \cdot x^b$ для преобразованных в логарифмах переменных получены следующие данные: $\sum xy = 9{,}23$; $\sum x = 10{,}56$, $\sum x^2 = 16{,}9$, $\sum y = 4{,}85$, $\sum (y - \tilde{y})^2 = 0{,}204$.

Задание:

1) Определить коэффициент регрессии, если $n = 9$.
2) Рассчитать показатель корреляции, предполагая, что $\sigma_y = 0{,}38$.
3) Оценить значимость уравнения регрессии.

Задача 2.2

Зависимость урожайности зерновых культур от дозы внесения удобрений в Центральном федеральном округе Российской Федерации характеризуется уравнением: $\tilde{y} = 3{,}81 \cdot 0{,}42^x$. Индекс корреляции равен 0,765; остаточная дисперсия $\sigma^2_{ост.} = 1{,}045$. Для оценки значимости уравнения регрессии требуется провести дисперсионный анализ полученных результатов.

Задача 2.3

По данным Северо-Западного федерального округа Российской Федерации за 2014 год изучается параболическая зависимость среднедушевых денежных доходов населения от валового регионального продукта в расчете на душу населения. Результаты представлены в таблице 2.3.

Таблица 2.3 Фактические и расчетные среднедушевые денежные доходы

Регион	Среднедушевые денежные доходы, тыс. руб.	
	фактические	расчетные
Республика Карелия	22,9	23,1
Республика Коми	30,8	32,2
Архангельская область	29,4	28,0
Вологодская область	22,8	23,4
Калининградская область	23,0	23,5
Ленинградская область	20,9	26,9
Мурманская область	34,1	27,0
Новгородская область	23,7	23,4
Псковская область	19,5	19,8

Задание:

1) Оценить качество модели, определив среднюю ошибку аппроксимации, показатель корреляции и *F*-критерий Фишера.

2) Сделать выводы.

Задача 2.4

Зависимость расходов 20 семей на покупку кондитерских изделий от среднемесячных доходов: $\tilde{y} = 1{,}2x^{0{,}3}$. Индекс корреляции равен 0,65; остаточная дисперсия $\sigma^2_{ост.} = 0{,}7$. Для оценки значимости уравнения регрессии, провести дисперсионный анализ полученных результатов.

Задача 2.5

Зависимость валового регионального продукта на душу населения y (млн. руб.) от уровня занятости x (%) по 14 регионам Приволжского федерального округа РФ характеризуется следующим образом:

Уравнение регрессии	$\tilde{y} = 42{,}60 - 1{,}32x + 0{,}09x^2$
Доля остаточной дисперсии в общей	26,4%

Задание:

1) Определить показатель корреляции.

2) Рассчитать коэффициент эластичности, предполагая, что численность занятых составляет 30 человек.

3) Оценить значимость уравнения регрессии.

Задача 2.6

По группе предприятий, производящих однородную продукцию, известна зависимость себестоимости единицы продукции от факторов, приведенных в таблице.

Факторный признак	Уравнение парной регрессии	Среднее значение фактора
Объем производства, млн. руб., x_1	$\tilde{y} = 0{,}9 + \frac{80{,}2}{x}$	$\bar{x}_1 = 4{,}2$
Трудоемкость единицы продукции, чел.-час., x_2	$\tilde{y} = 12{,}4 + 4{,}3 \cdot x$	$\bar{x}_2 = 0{,}7$
Оптовая цена за 1 т энергоносителя, млн. руб., x_3	$\tilde{y} = 13{,}6 \cdot x_3^{1,2}$	$\bar{x}_3 = 1{,}9$

Задание:

1) Определить с помощью коэффициента эластичности влияние каждого фактора на себестоимость.
2) Ранжировать факторы по силе влияния на результат.
3) Сделать выводы.

Задача 2.7

Для трех видов продукции А, Б и С модели зависимости удельных постоянных расходов от объема выпускаемой продукции выглядит следующим образом:

$$\tilde{y}_A = 700; \qquad \tilde{y}_Б = 65 + 0{,}95 \cdot x; \qquad \tilde{y}_C = 70 \cdot x^{0,6}.$$

Задание:

1) Определить коэффициенты эластичности по каждому виду продукции, если объем продукции Б составляет 600 ед.
2) Определить, каким должен быть объем выпускаемой продукции Б, чтобы коэффициенты эластичности для продукции Б и С были равны.
3) Сделать выводы.

Задача 2.8

При изучении спроса на телевизоры марки N по 19 торговым точкам аналитики компании АВС выявили следующую зависимость:

$$\ln y = 10{,}5 - 0{,}8 \ln x + \varepsilon,$$
$$(2{,}5) \quad (-4{,}0)$$

где y – объем продаж телевизоров марки N в отдельной торговой точке;

x – средняя цена телевизора в данной торговой точке.

В скобках приведены фактические значения t-критерия Стьюдента для параметров уравнения регрессии.

Задание:

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?

Задача 2.9

По группе 12 заводов, производящих однородную продукцию, получено уравнение регрессии себестоимости единицы продукции (тыс. руб.) y от уровня технической оснащенности (тыс. руб.) x: $y = 20{,}8 + \frac{785{,}2}{x}$. Доля остаточной дисперсии в общей составила 0,287.

Задание:

1) Определить коэффициент эластичности, предполагая, что стоимость активных производственных фондов составляет 530 тыс. руб.

2) Рассчитать индекс корреляции и F-критерий Фишера.

Задача 2.10

Имеются следующие данные по 17 регионам Центрального федерального округа РФ.

Показатель	Средние значения	Коэффициент вариации, %
Урожайность зерновых, ц с 1 га	24,3	26,8
Внесено удобрений на 1 га посевов, кг	7,7	25,3

Фактическое значение F-критерия Фишера равно 16,8.

Задание:

1) Определить линейный коэффициент детерминации.
2) Построить уравнение линейной регрессии.
3) Найти обобщающий коэффициент эластичности.
4) С вероятностью 95,0% указать доверительный интервал ожидаемого значения урожайности в предположении роста количества внесенных удобрений на 10,0% от своего среднего уровня.

Задача 2.11

По регионам Центрального федерального округа РФ изучается зависимость уровня безработицы (%) y от индекса потребительских цен (% к предыдущему году) x. Логарифмы исходных показателей представлены в таблице.

Показатель	$\ln x$	$\ln y$
Среднее значение	0,54	1,16
Среднее квадратическое отклонение	0,48	0,35

Показатель корреляции между логарифмами составил 0,758.

Задание:

1) Построить уравнение регрессии зависимости уровня безработицы от индекса потребительских цен в степенной форме.

2) Дать интерпретацию коэффициента эластичности данной модели регрессии.

3) Определить значение коэффициента детерминации и пояснить его смысл.

2.4 Тестовые задания

1 Какие функции относятся к нелинейным регрессиям по оцениваемым параметрам?

а) равносторонняя гипербола;

б) степенная;

в) показательная;

г) полиномы разных степеней.

2 Методом наименьших квадратов определяются параметры:

а) нелинейной регрессии по оцениваемым параметрам;

б) нелинейной регрессии по включенным в нее объясняющим переменным;

в) линейной регрессии.

3 Зависимость спроса от цен характеризуется уравнением вида $\tilde{y}_x = 106 \cdot x^{-1,5}$. Что это означает?

а) с увеличением цен на 1,5 руб., спрос снижается на 1%;

б) с увеличением цен на 106 руб., спрос снижается на 1,5 руб.;

в) со снижением цен на 1,5%, спрос увеличивается на 1%;

г) с увеличением цен на 1%, спрос снижается в среднем на 1,5%.

4 В степенной функции $y = a \cdot x^b \cdot \varepsilon$ параметр b является:

а) коэффициентом регрессии;

б) коэффициентом корреляции;

в) коэффициентом эластичности.

5 Индекс корреляции для нелинейной регрессии определяется по формуле:

а) $\rho = \left(1 - \frac{\sigma^2_{ост.}}{\sigma^2_{общ.}}\right)^2$; б) $\rho = \sqrt{1 - \frac{\sigma^2_{ост.}}{\sigma^2_{общ.}}}$; в) $\rho = \sqrt{1 - \frac{\sum (y - \tilde{y}_x)^2}{\sum (y - \bar{y})^2}}$.

6 Для оценки параметров уравнения $y = a + \frac{b}{x} + \varepsilon$ используется:

а) метод наименьших квадратов; б) итеративный метод.

7 Классическим примером гиперболы в экономике является:
а) кривая Филлипса;
б) кривая Лизера;
в) кривая Энгеля;
г) кривая Смита.

8 Какой класс нелинейных моделей делится на внутренне линейные и внутренне нелинейные?
а) регрессии, нелинейные по оцениваемым параметрам.
б) регрессии, нелинейные по объясняющим переменным.

9 Модель вида $y = a \cdot x^b + \varepsilon$ относится:
а) к внутренне нелинейным;
б) к внутренне линейным.

10 Если для определенного интервала значений фактора меняется характер связи рассматриваемых признаков: прямая связь меняется на обратную или наоборот, то используется:
а) степенная функция;
б) парабола второй степени.
в) полином третьего порядка;

11 Индекс корреляции принимает значения:
а) $-1 \le \rho \le 1$; б) $\rho \le 1$; в) $0 \le \rho \le 1$.

12 F-критерий Фишера для нелинейной регрессии рассчитывается по формуле:
а) $F = \frac{\rho^2}{1+\rho^2} \cdot \frac{n-m-1}{m}$; б) $F = \frac{\rho^2}{1-\rho^2} \cdot \frac{n-m-1}{m}$; в) $F = \frac{\rho^2}{1-\rho^2} \cdot \frac{m}{n-m-1}$.

13 Для оценки параметров уравнения $y = a + b \cdot x + c \cdot x^2$ используется:
а) метод наименьших квадратов;
б) итеративный метод.

14 Кривая Филлипса характеризует:
а) взаимосвязь доли расходов на товары длительного пользования и общих сумм расходов (или доходов);
б) соотношение между нормой безработицы и процентом прироста заработной платы.

15 Кривая Филлипса выражается уравнением:
а) прямой;
б) параболы второй степени;
в) гиперболы.

16 Параметры каких уравнений могут быть оценены с помощью метода наименьших квадратов?

а) $y = a + b \cdot x + c \cdot x^2$; б) $y = a + \frac{b}{x}$; в) $y = a \cdot x^b$; г) $y = a + b \cdot \ln x$.

17 Для оценки параметров внутренне нелинейных моделей используют:

а) метод наименьших квадратов;

б) итеративный метод.

18 Коэффициент эластичности для гиперболы равен:

а) $Э = b$; б) $Э = \frac{b \cdot x}{a + b \cdot x}$; в) $Э = -\frac{b}{a \cdot x + b}$.

19 Какие функции относятся к нелинейным регрессиям по включенным в нее объясняющим переменным?

а) полиномы разных степеней;

б) равносторонняя гипербола;

в) степенная;

г) показательная.

20 В уравнении регрессии $y = a + b \cdot x + c \cdot x^2 + \varepsilon$ m равно:

а) 2; б) 1; в) 3

2.5 Контрольные вопросы

1) Назовите два класса нелинейных регрессий.
2) Приведите примеры моделей, нелинейных относительно включаемых переменных.
3) Приведите примеры моделей, нелинейных относительно оцениваемых параметров.
4) Какой нелинейной функцией может быть заменена парабола второй степени, если не наблюдается смена направленности связи признаков?
5) Как уравнения гиперболы и параболы второй степени приводятся к линейному виду?
6) Показатели тесноты связи для нелинейной регрессии.
7) Оценка существенности уравнения нелинейной регрессии.
8) Что показывает коэффициент регрессии в степенной функции?
9) Как определяется коэффициент эластичности для гиперболы?
10) Значима ли модели, если $F_{факт.}$ = 6,35, n = 15, а связь между x и y описывается параболой второго порядка?

3 ЛИНЕЙНЫЕ МОДЕЛИ МНОЖЕСТВЕННОЙ РЕГРЕССИИ

Цель занятия – изучить основные понятия множественной регрессии и корреляции, овладеть навыками оценки параметров множественной регрессии и корреляции и их статистической значимости.

3.1 Теоретические положения

3.1.1 Спецификация модели

3.1.2 Оценка параметров уравнения множественной регрессии

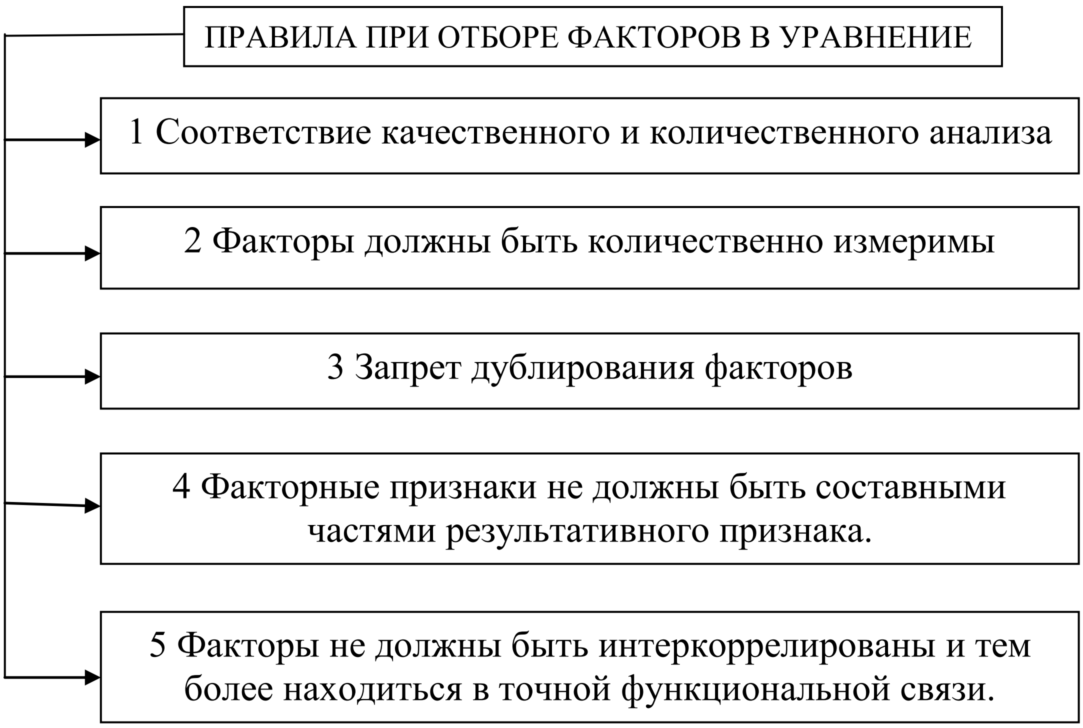
3.1.3 Множественная корреляция

3.1.4 Оценка надежности результатов множественной регрессии и корреляции

3.1.1 Спецификация модели

Построение уравнения множественной регрессии начинается с решения вопроса о спецификации модели. Спецификация модели состоит из двух этапов: отбора факторов и выбора вида уравнения регрессии.

При ***отборе факторов*** при построении множественной регрессии необходимо соблюдать следующие правила**,** представленные на рисунке 3.1.



3.1 Правила при отборе факторов в уравнение множественной регрессии

Коллинеарность означает достаточно тесную неслучайную линейную корреляцию одних факторов с другими. Считается, что два фактора явно коллинеарны, т.е. находятся между собой в линейной зависимости, если $r_{xixj} \geq 0{,}7$.

Если факторы явно коллинеарны, то они дублируют друг друга и один из них следует исключить из модели. Предпочтение при этом отдается не фактору, более тесно связанному с результатом, а тому фактору, который при достаточно тесной связи с результатом имеет наименьшую тесноту связи с другими факторами.

При построении моделей множественной регрессии применяются разные виды уравнений: линейные и нелинейные. Однако из-за четкой интерпретации параметров наиболее широко используются линейная и степенная функции.

3.1.2 Оценка параметров уравнения множественной регрессии

Уравнение линейной множественной регрессии имеет вид:

$$\tilde{y}_x = a + b_1 x_1 + b_2 x_2 + ... + b_n x_n, \tag{3.1}$$

где $b_1, b_2, ..., b_n$ – коэффициенты условно чистой регрессии. Они характеризуют среднее изменение результата при изменении соответствующего фактора на единицу при неизменном значении других факторов, закрепленных на среднем уровне.

Параметры уравнения множественной регрессии оцениваются, как и в парной регрессии, методом наименьших квадратов (МНК). При его применении строится система нормальных уравнений, решение которой позволяет получить оценки параметров регрессии.

Так, для уравнения $y = a + b_1 x_1 + b_2 x_2 + ... + b_n x_n + \varepsilon$ система нормальных уравнений составит:

$$\begin{cases} \sum y = n \cdot a + b_1 \sum x_1 + b_2 \sum x_2 + ... + b_n \sum x_n, \\ \sum y \cdot x_1 = a \sum x_1 + b_1 \sum x_1^2 + b_2 \sum x_1 \cdot x_2 + ... + b_n \sum x_n x_1, \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ \sum y x_n = a \sum x_n + b_1 \sum x_1 \cdot x_n + b_2 \sum x_2 \cdot x_n + ... + b_n \sum x_n^2. \end{cases} \tag{3.2}$$

Ее решение может быть осуществлено методом определителей:

$$a = \frac{\Delta a}{\Delta}; \qquad b_1 = \frac{\Delta b_1}{\Delta}; \quad \dots \; ; \; b_n = \frac{\Delta b_n}{\Delta}, \tag{3.3}$$

где Δ – определитель системы;

$\Delta a, \Delta b_1, ..., \Delta b_n$ – частные определители.

Определитель системы имеет вид:

$$\Delta = \begin{vmatrix} n & \sum x_1 & \sum x_2 & \ldots & \sum x_n \\ \sum x_1 & \sum x_1^2 & \sum x_2 x_1 & \ldots & \sum x_n x_1 \\ \sum x_2 & \sum x_1 x_2 & \sum x_2^2 & \ldots & \sum x_n x_2 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \sum x_n & \sum x_1 x_n & \sum x_2 x_n & \ldots & \sum x_n^2 \end{vmatrix} \qquad (3.4)$$

Частные определители $\Delta a,\ \Delta b_1, \ldots, \Delta b_n$ получаем путем замены соответствующего столбца матрицы определителя системы данными левой части системы.

Параметры множественной регрессии можно оценить также с помощью построения уравнение регрессии в ***стандартизованном масштабе:***

$$t_y = \beta_1 \cdot t_{x1} + \beta_2 \cdot t_{x2} + \ldots + b_n \cdot t_{xn} + \varepsilon, \qquad (3.5)$$

где t – стандартизованные переменные, для которых среднее значение равно 0, а среднее квадратическое отклонение равно 1;

β – стандартизованные коэффициенты регрессии.

Применяя МНК к уравнению множественной регрессии в стандартизованном масштабе, после соответствующих преобразований получим систему нормальных уравнений вида:

$$\begin{cases} r_{yx1} = \beta_1 + \beta_2 \cdot r_{x2x1} + \beta_3 \cdot r_{x3x1} + \ldots + \beta_n \cdot r_{xnx1}, \\ r_{yx2} = \beta_1 \cdot r_{x2x1} + \beta_2 + \beta_3 \cdot r_{x3x2} + \ldots + \beta_n \cdot r_{xnx2}, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ r_{yxn} = \beta_1 \cdot r_{xnx1} + \beta_2 \cdot r_{xnx2} + \beta_3 \cdot r_{x3xn} + \ldots + \beta_n \end{cases}, \qquad (3.6)$$

где r_{yx1}, r_{yx2} – парные коэффициенты корреляции.

Решая ее методом определителей, находят параметры – стандартизованные коэффициенты регрессии ($\beta -$коэффициенты) [19].

Для нахождения $\beta -$коэффициентов в уравнении с двумя факторами (x1, x2) можно воспользоваться готовыми рабочими формулами:

$$\beta_1 = \frac{r_{yx1} - r_{yx2} \cdot r_{x1x2}}{1 - r_{x1x2}^2}; \qquad \beta_2 = \frac{r_{yx2} - r_{yx1} \cdot r_{x1x2}}{1 - r_{x1x2}^2}. \qquad (3.7)$$

Стандартизированные коэффициенты регрессии ($\beta -$коэффициенты) показывают, на сколько сигм изменится в среднем результат, если соответствующий фактор x_i изменится на одну сигму при неизменном среднем уровне других факторов.

Основное достоинство стандартизованных коэффициентов регрессии в отличие от коэффициентов «условно чистой» регрессии в том, что их можно сравнивать друг с другом и ранжировать факторы по силе их влияния на результат.

Во множественной регрессии коэффициенты условно чистой регрессии b_i связаны со стандартизованными коэффициентами регрессии β_i, а именно:

$$b_i = \beta_i \frac{\sigma_y}{\sigma_{xi}}; \tag{3.8}$$

Это позволяет от уравнения в стандартном масштабе переходить к уравнению регрессии в натуральном масштабе. При этом параметр a определяется по формуле:

$$a = \bar{y} - b_1\bar{x}_1 - b_2\bar{x}_2 - ... - b_n\bar{x}_n \tag{3.9}$$

3.1.3 Множественная корреляция

Показатель множественной корреляции характеризует тесноту связи рассматриваемого набора факторов с исследуемым признаком, то есть оценивает тесноту совместного влияния факторов на результат.

В зависимости от исходных данных показатель множественной корреляции может быть определен следующими способами.

1. Независимо от формы связи показатель множественной корреляции может быть найден как ***индекс множественной корреляции:***

$$R_{yx1x2...xn} = \sqrt{1 - \frac{\sigma^2_{ост.}}{\sigma^2_y}}, \tag{3.10}$$

где σ^2_y – общая дисперсия результативного признака;

$\sigma^2_{ост}$ – остаточная дисперсия.

$R_{yx1x2...xn}$ принимает значения от 0 до 1. 2. При линейной зависимости признаков формула индекса корреляции может быть представлена следующим выражением:

$$R_{yx1x2...xn} = \sqrt{\sum \beta_{xi} \cdot r_{yxi}} = \sqrt{\beta_{x1} \cdot r_{yx1} + \beta_{x2} \cdot r_{yx2} + ... + \beta_n \cdot r_{yn}}, \tag{3.11}$$

где β_{xi} – стандартизованные коэффициенты регрессии;

r_{yxi} – парные коэффициенты корреляции результата с каждым фактором.

3. При линейной зависимости совокупный коэффициент корреляции можно также определить через матрицу парных коэффициентов корреляции:

$$R_{yx1x2...чт} = \sqrt{1 - \frac{\Delta r}{\Delta r_{11}}}, \tag{3.12}$$

где Δr – определитель матрицы парных коэффициентов корреляции;

Δr_{11} – определитель матрицы межфакторной корреляции.

Для уравнения $y = a + b_1 \cdot x_1 + b_2 \cdot x_2 + ... + b_n \cdot x_n + \varepsilon$ определитель матрицы коэффициентов парной корреляции принимает вид:

$$\Delta r = \begin{vmatrix} 1 & r_{yx1} & r_{yx2} & \ldots & r_{yxn} \\ r_{yx1} & 1 & r_{x1x2} & \ldots & r_{x1xn} \\ r_{yx2} & r_{x2x1} & 1 & \ldots & r_{x2xn} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ r_{yxn} & r_{xnx1} & r_{xnx2} & \ldots & 1 \end{vmatrix} \qquad (3.13)$$

Определитель более низкого порядка Δr_{11} остается, когда вычеркиваются из матрицы коэффициентов парной корреляции первый столбец и первая строка, что соответствует матрице коэффициентов парной корреляции между факторами:

$$\Delta r_{11} = \begin{vmatrix} 1 & r_{x1x2} & \ldots & r_{x1xn} \\ r_{x2x1} & 1 & \ldots & r_{x2xn} \\ \ldots & \ldots & \ldots & \ldots \\ r_{xnx1} & r_{xnx2} & \ldots & 1 \end{vmatrix} \qquad (3.14)$$

Чем ближе его значение к 1, тем теснее связь результативного признака со всем набором исследуемых факторов. Величина индекса множественной корреляции должна быть больше или равна максимальному парному индексу корреляции.

Коэффициент множественной детерминации рассчитывается как квадрат индекса множественной корреляции: R^2_{yx1x2}.

Скорректированный индекс множественной детерминации содержит поправку на число степеней свободы и рассчитывается по формуле:

$$\overline{R}^2 = 1 - (1 - R^2) \cdot \frac{(n-1)}{(n-m-1)}, \qquad (3.15)$$

где n – число наблюдений;

m – число параметров при переменных x (число факторов, включенных в модель).

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

3.1.4 Оценка надежности результатов множественной регрессии и корреляции

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

$$F = \frac{R^2}{1-R^2} \cdot \frac{n-m-1}{m}, \qquad (3.16)$$

где R^2 – коэффициент множественной детерминации.

Если $F_{факт.} > F_{табл}$, то уравнение множественной регрессии в целом статистически значимо и надежно.

Оценивается значимость не только уравнения в целом, но и фактора, дополнительно включенного в регрессионную модель.

Частные F-критерии оценивают статистическую значимость присутствия факторов x_1 и x_2 в уравнении множественной регрессии, оценивают целесообразность включения в уравнение одного фактора после другого фактора.

Рассмотрим расчет частных F-критериев на примере двухфакторной модели.

F_{x1} оценивает целесообразность включения в уравнение фактора x_1 после того, как в него был включен фактор x_2. Соответственно, F_{x2} указывает на целесообразность включения в модель фактора x_2 после фактора x_1. Определим частные F-критерии для факторов $x1$ и $x2$ по формулам:

$$F_{x1} = \frac{R^2_{yx1x2} - r^2_{yx2}}{1 - R^2_{yx1x2}}(n - m - 1); \qquad (3.17)$$

$$F_{x2} = \frac{R^2_{yx1x2} - r^2_{yx1}}{1 - R^2_{yx1x2}}(n - m - 1). \qquad (3.18)$$

Зная величину F_{x_i}, можно определить t-критерий Стьюдента для коэффициента регрессии по формуле:

$$t_{bi} = \sqrt{F_{xi}}. \qquad (3.19)$$

3.2. Решение типовых задач

Пример 3.1

Имеются данные по регионам Центрального федерального округа Российской Федерации: (таблица 3.1). y – среднедушевые денежные доходы населения, тыс. руб.; x_1 – доля сельского населения в общей численности населения региона, %; x_2 – среднемесячная номинальная начисленная заработная плата работников, тыс. руб.

Задание:

1) Оценить параметры уравнения множественной регрессии с помощью метода определителей.
2) Построить уравнение регрессии в стандартизованном масштабе.
3) Построить частные уравнения регрессии.
4) Определить средние и частные коэффициенты эластичности.
5) Рассчитать коэффициент множественной корреляции.
6) Определить коэффициент детерминации (скорректированный, нескорректированный).
7) Найти частные коэффициенты корреляции.

Таблица 3.1 Исходные данные для корреляционно-регрессионного анализа

Регионы	x_1	x_2	y
1 Белгородская область	34,22	2,51	2,13
2 Брянская область	31,18	1,82	1,69
3 Владимирская область	19,57	2,29	1,63
4 Воронежская область	37,92	1,91	2,02
5 Ивановская область	17,59	1,76	1,26
6 Калужская область	25,72	2,49	1,69
7 Костромская область	33,85	2,19	1,84
8 Курская область	38,41	2,01	1,86
9 Липецкая область	35,55	2,64	2,25
10 Московская область	19,84	3,45	2,68
11 Орловская область	37,15	2,25	1,92
12 Рязанская область	31,47	2,19	1,84
13 Смоленская область	29,44	2,38	2,26
14 Тамбовская область	42,00	1,76	2,00
15 Тверская область	26,41	2,34	1,62
16 Тульская область	18,53	2,41	2,01
17 Ярославская область	19,63	2,83	2,50

8) Оценить значимость уравнения множественной регрессии с помощью F-критерия Фишера.

9) Рассчитать частные F-критерии Фишера.

10) Оценить значимость коэффициентов чистой регрессии по t-критерию Стьюдента.

РЕШЕНИЕ:

1) Параметры уравнения множественной регрессии оцениваются, как и в парной регрессии, методом наименьших квадратов (МНК).

Для оценки параметров уравнения множественной регрессии построим с помощью MS Excel вспомогательную таблицу 3.2.

На основе расчетов, представленных в таблице 3.2, получили следующую систему:

$$\begin{cases} 45{,}34 = 17a + 498{,}48b_1 + 44{,}15b_2, \\ 975{,}4675 = 498{,}48a + 15641b_1 + 1127{,}023b_2 \\ 138{,}08 = 44{,}15a + 1127{,}023b_1 + 117{,}6303b_2. \end{cases}$$

Решив систему с помощью метода определителей, получили следующие значения:

Δ=144107,059;

Δa=-609468,266, $\quad a = -\dfrac{609468{,}266}{144107{,}059} = -4{,}2293$,

Таблица 3.2 Вспомогательная таблица для расчета параметров уравнения множественной регрессии

№	x_1	x_2	y	x_1^2	$x_1 \cdot x_2$	x_2^2	$y \cdot x_1$	$y \cdot x_2$	y^2
1	34,22	2,51	2,13	1171,008	85,892	6,300	72,889	5,346	4,537
2	31,18	1,82	1,69	972,192	56,748	3,312	52,694	3,076	2,856
3	19,57	2,29	1,63	382,985	44,815	5,244	31,899	3,733	2,657
4	37,92	1,91	2,02	1437,926	72,427	3,648	76,598	3,858	4,080
5	17,59	1,76	1,26	309,408	30,958	3,098	22,163	2,218	1,588
6	25,72	2,49	1,69	661,518	64,043	6,200	43,467	4,208	2,856
7	33,85	2,19	1,84	1145,823	74,132	4,796	62,284	4,030	3,386
8	38,41	2,01	1,86	1475,328	77,204	4,040	71,443	3,739	3,460
9	35,55	2,64	2,25	1263,803	93,852	6,970	79,988	5,940	5,063
10	19,84	3,45	2,68	393,626	68,448	11,903	53,171	9,246	7,182
11	37,15	2,25	1,92	1380,123	83,588	5,063	71,328	4,320	3,686
12	31,47	2,19	1,84	990,361	68,919	4,796	57,905	4,030	3,386
13	29,44	2,38	2,26	866,714	70,067	5,664	66,534	5,379	5,108
14	42,00	1,76	2,00	1764,000	73,920	3,098	84,000	3,520	4,000
15	26,41	2,34	1,62	697,488	61,799	5,476	42,784	3,791	2,624
16	18,53	2,41	2,01	343,361	44,657	5,808	37,245	4,844	4,040
17	19,63	2,83	2,50	385,337	55,553	8,009	49,075	7,075	6,250
Итого	498,48	44,15	45,34	15641,000	1127,023	117,630	975,468	138,080	214,138
Среднее	27,693	2,453	2,519	868,944	62,612	6,549	54,193	7,671	11,897

Δb_1=-841,435, $\quad b_1 = -\dfrac{841{,}435}{144107{,}059} = -0{,}0058$,

Δb_2=405972,391, $\quad b_2 = \dfrac{405972{,}391}{144107{,}059} = 2{,}8172$.

Таким образом, уравнение множественной регрессии имеет вид:

$$\tilde{y} = -4{,}2293 - 0{,}0058x_1 + 2{,}8172x_2.$$

При увеличении доли сельского населения на 1%, среднедушевой денежный доход в Центральном федеральном округе снизится на 5,8 руб., а при увеличении средней начисленной заработной платы на 1000 руб., среднедушевой денежный доход увеличится на 2817,2 руб.

2) Применяя МНК к уравнению множественной регрессии в стандартизованном масштабе, после соответствующих преобразований получим систему нормальных уравнений вида:

$$\begin{cases} r_{yx1} = \beta_1 + \beta_2 r_{x2x1}, \\ r_{yx2} = \beta_1 r_{x2x1} + \beta_2, \end{cases} \tag{3.20}$$

где r_{yx1}, r_{yx2} – парные коэффициенты корреляции.

Парные коэффициенты корреляции найдем по формулам:

$$r_{x1x2} = \frac{\overline{x_1 \cdot x_2} - \bar{x}_1 \cdot \bar{x}_2}{\sigma_{x1}\sigma_{x2}},$$

$$r_{yx1} = \frac{\overline{y \cdot x_1} - \bar{y} \cdot \bar{x}_2}{\sigma_y \sigma_{x1}},$$

$$r_{yx2} = \frac{\overline{y \cdot x_2} - \bar{y} \cdot \bar{x}_2}{\sigma_y \sigma_{x2}},$$

где $\sigma_{x1} = \sqrt{\frac{\sum (x_1 - \bar{x}_1)^2}{n}} = \sqrt{\overline{x_1}^2 - \bar{x}_1^2} = \sqrt{868{,}9445 - 766{,}9189} = 10{,}1008,$

$$\sigma_{x2} = \sqrt{\frac{\sum (x_2 - \bar{x}_2)^2}{n}} = \sqrt{\overline{x_2}^2 - \bar{x}_2^2} = \sqrt{6{,}5488 - 6{,}0162} = 0{,}7298,$$

$$\sigma_y = \sqrt{\frac{\sum (y - \bar{y})^2}{n}} = \sqrt{\overline{y^2} - \bar{y}^2} = \sqrt{11{,}9878 - 6{,}3449} = 2{,}3565,$$

$$r_{x1x2} = \frac{62{,}6124 - 27{,}6933 \cdot 2{,}4528}{10{,}1008 \cdot 0{,}7298} = -0{,}6208,$$

$$r_{yx1} = \frac{54{,}1926 - 2{,}5189 \cdot 27{,}6933}{2{,}3565 \cdot 10{,}1008} = -0{,}654,$$

$$r_{yx2} = \frac{7{,}6711 - 2{,}5189 \cdot 2{,}4528}{2{,}3565 \cdot 0{,}7298} = 0{,}8680.$$

Система уравнений имеет вид: $\begin{cases} -0{,}654 = \beta_1 - 0{,}6208\beta_2, \\ 0{,}8680 = -0{,}6208\beta_1 + \beta_2. \end{cases}$

Решив систему методом определителей, получили формулы:

$$\beta_1 = \frac{r_{yx1} - r_{yx2} \cdot r_{x1x2}}{1 - r_{x1x2}^2} = \frac{-0{,}654 - 0{,}868 \cdot (-0{,}6208)}{1 - (-0{,}6208)^2} = -0{,}059,$$

$$\beta_2 = \frac{r_{yx2} - r_{yx1} \cdot r_{x1x2}}{1 - r_{x1x2}^2} = \frac{0{,}868 - (-0{,}654) \cdot (-0{,}6208)}{1 - (-0{,}6208)^2} = 0{,}825.$$

Уравнение в стандартизированном масштабе имеет вид:

$$t_y = -0{,}059 t_{x1} + 0{,}825 t_{x2} + \varepsilon$$

Таким образом, с ростом доли сельского населения на 1 сигму при неизменном среднем уровне начисленной заработной платы, среднедушевые денежные доходы в среднем уменьшатся на 0,059 сигмы; а с увеличением средней месячной заработной платы на 1 сигму при неизменной доле сельского населения среднедушевые денежные доходы возрастут на 0,825 сигмы.

Зная значения стандартизованных коэффициентов регрессии $\beta_{i,}$,можно определить коэффициенты условно «чистой» регрессии b_i :

$$b_1 = -0{,}0589 \cdot \frac{2{,}3565}{10{,}1008} = -0{,}0137,$$

$$b_2 = -0{,}8256 \cdot \frac{2{,}3565}{0{,}7298} = 2{,}6658.$$

3) Частные уравнения регрессии связывают результативный признак с соответствующими факторами x при закреплении других учитываемых во множественной регрессии факторов на среднем уровне. Частные уравнения имеют вид:

$$y_{x1 \cdot x2} = a + b_1 x_1 + b_2 \bar{x}_2 + \varepsilon, \qquad (3.21)$$

$$y_{x2 \cdot x1} = a + b_1 \bar{x}_1 + b_2 x_2 + \varepsilon\,. \qquad (3.22)$$

В данной задаче частные уравнения имеют вид:

$$y_{x1 \cdot x2} = -4{,}2293 - 0{,}0058 x_1 + 2{,}8172 \cdot 2{,}4528 = 2{,}6807 - 0{,}0058 x_1,$$

$$y_{x2 \cdot x1} = -4{,}2293 - 0{,}0058 \cdot 27{,}6933 + 2{,}8172 x_2 = -4{,}3899 + 2{,}8172 x_2.$$

4) На основе частных уравнений регрессии можно определить частные коэффициенты эластичности для каждого региона по формуле:

$$Э_{yxi} = b_i \cdot \frac{x_i}{\tilde{y}_{x1 \cdot x2}}, \qquad (3.23)$$

где b_i – коэффициенты регрессии для фактора x_i в уравнении множественной регрессии;

$\tilde{y}_{x1 \cdot x2}$ – частное уравнение регрессии.

Рассчитаем частные коэффициенты эластичности для Белгородской и Брянской областей.

Для Белгородской области x_1=34,22, x_2=2,51, тогда:

$$Э_{yx1} = -0{,}0058 \cdot \frac{34{,}22}{2{,}6807 - 0{,}0058 \cdot 34{,}22} = -0{,}08,$$

$$Э_{yx2} = 2{,}8172 \cdot \frac{2{,}51}{-4{,}3899 + 2{,}8172 \cdot 2{,}51} = 2{,}64.$$

Для Брянской области x_1 =31,18, x_2=1,82:

$$Э_{yx1} = -0{,}0058 \cdot \frac{31{,}18}{2{,}6807 - 0{,}0058 \cdot 31{,}18} = -0{,}07,$$

$$Э_{yx2} = 2{,}8172 \cdot \frac{1{,}82}{-4{,}3899 + 2{,}8172 \cdot 1{,}82} = 6{,}95.$$

Таким образом, в Белгородской области при увеличении доли сельского населения в общей численности населения на 1%, среднедушевые денежные доходы населения сократятся на 0,08%, а при увеличении заработной платы на 1%, денежные доходы возрастут на 2,64%. В Брянской области при увеличении доли сельского населения на 1%, среднедушевые денежные доходы населения сократятся на 0,07, а при увеличении заработной платы на 1%, денежные доходы возрастут на 6,95%.

Средние по совокупности показатели эластичности находим по формуле:

$$\overline{Э}_{yxi} = b_i \cdot \frac{\vec{x}_i}{\overline{y}_{xi}}, \tag{3.24}$$

Для данной задачи они окажутся равными:

$$\overline{Э}_{yx1} = -0{,}0058 \cdot \frac{27{,}6933}{2{,}5189} = -0{,}064\%,$$

$$\overline{Э}_{yx2} = 2{,}8172 \cdot \frac{27{,}6933}{2{,}5189} = 30{,}97\%.$$

Таким образом, с ростом доли сельского населения на 1%, размер среднедушевого денежного дохода населения в среднем по совокупности сократится на 0,064% при неизменном размере заработной платы. При увеличении среднемесячной заработной платы на 1%, величина денежного дохода в среднем по изучаемой совокупности возрастет на 30,97% при неизменной доле сельского населения.

5) Индекс множественной корреляции определим по формуле (3.11):

$$R_{yx1x2} = \sqrt{(-0{,}0589) \cdot (-0{,}654) + 0{,}8256 \cdot 0{,}8680} = 0{,}869.$$

При линейной зависимости совокупный коэффициент корреляции можно также определить через матрицу парных коэффициентов корреляции по формуле (3.12): $R_{yx1x2} = \sqrt{1 - \frac{0{,}1182}{0{,}4804}} = 0{,}868.$

Таким образом, связь средних денежных доходов населения с долей сельского населения и средней месячной заработной платой сильная.

6) Качество построенной модели в целом оценивает коэффициент детерминации. Коэффициент множественной детерминации рассчитывается как квадрат индекса множественной корреляции: $R^2_{yx1x2} = 0{,}869^2 = 0{,}755$.

Скорректированный индекс множественной детерминации содержит поправку на число степеней свободы и рассчитывается по формуле (3.15):

$$\overline{R}^2 = 1-(1-0{,}755)\cdot\frac{17-1}{17-2-1} = 0{,}7026.$$

Таким образом, вариация среднедушевых денежных доходов населения на 75,5% (70,26% - при скорректированном индексе детерминации) зависит от вариации доли сельского населения в общей численности населения и среднемесячной начисленной заработной платы, а на остальные 24,5% (29,74%) от других факторов, не включенных в модель.

7) Частные коэффициенты корреляции характеризуют тесноту связи между результатом и соответствующим фактором при устранении влияния других факторов, включенных в модель. Формула коэффициента частной корреляции, выраженная через показатель детерминации, для x_1 принимает вид:

$$r_{yx1\cdot x2} = \frac{r_{yx1} - r_{yx2}\cdot r_{x1x2}}{\sqrt{(1-r_{yx2}^2)\cdot(1-r_{x1x2}^2)}} = \frac{-0{,}654-0{,}868\cdot(-0{,}6208)}{\sqrt{(1-0{,}868^2)\cdot(1-(-0{,}6208)^2)}} = -0{,}3159, \qquad (3.25)$$

$$r_{yx2\cdot x1} = \frac{r_{yx2} - r_{yx1}\cdot r_{x1x2}}{\sqrt{(1-r_{yx1}^2)\cdot(1-r_{x1x2}^2)}} = \frac{0{,}868-(-0{,}654)\cdot(-0{,}6208)}{\sqrt{(1-(-0{,}654^2)\cdot(1-(-0{,}6208)^2)}} = 0{,}7563. \qquad (3.26)$$

Таким образом, при закреплении фактора x_2 на постоянном уровне (элиминировании) корреляция y и x_1 равна -0,3159, то есть связь обратная слабая. При закреплении фактора x_1 на постоянном уровне корреляция y и x_2 равна 0,7563, то есть связь прямая сильная.

8) Значимость уравнения множественной регрессии в целом, оценивается с помощью F-критерия Фишера по формуле (3.16):

$$F_{факт.} = \frac{0{,}755}{1-0{,}755}\cdot\frac{17-2-1}{2} = 21{,}571.$$

$F_{табл.}$ =3,74 (при $k_1=m=2$ и $k_2=n-m-1=17-2-1=14$).

Так как $F_{факт.} > F_{табл}$, с вероятностью 95% делаем вывод о статистической значимости уравнения в целом и показателя тесноты связи, которые сформировались под неслучайным воздействием факторов x_1, x_2.

9) Определим частные F-критерии для факторов x_1 и x_2 по формулам (3.17) и (3.18):

$$F_{x1} = \frac{0{,}755-0{,}868^2}{1-0{,}755}(17-1-1) = 0{,}096,$$

$$F_{x2} = \frac{0{,}755-(-0{,}654)^2}{1-0{,}755}(17-1-1) = 20{,}038.$$

$F_{табл.}$ = 4,54.

Таким образом, низкое значение $F_{x1факт.}$ свидетельствует о нецелесообразности включения в модель фактора x_1 (доли сельского населения). Включение же фактора x_2 в модель статистически целесообразно.

Это означает, что парная регрессионная модель зависимости среднего дохода от средней заработной платы является достаточно статистически значимой, надежной и нет необходимости улучшать ее, включая дополнительный фактор x_1.

10) Частный F-критерий оценивает значимость коэффициентов чистой регрессии по формуле (3.19):

$t_{b1} = \sqrt{0{,}096} = 0{,}310$,

$t_{b2} = \sqrt{20{,}038} = 4{,}476$,

$t_{табл.}$=2,1448.

Так как $t_{b1} < t_{табл.}$, то фактор x_1 статистически незначим, а так как $t_{b2} > t_{табл.}$, то фактор x_2 статистически значим.

Пример 3.2

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 представлены в таблице 3.3.

Таблица 3.3 Исходные данные для анализа

Признак	Среднее значение	Среднее квадратическое отклонение	Парный коэффициент корреляции
y	250	38	$r_{yx1} = 0{,}68$
x_1	47	12	$r_{yx2} = 0{,}63$
x_2	112	21	$r_{x1x2} = 0{,}42$

Задание:

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

2) Найти уравнение множественной регрессии в стандартизованном и натуральном масштабе.

3) Рассчитать множественный коэффициент корреляции, общий и частные *F*-критерии Фишера и сделать выводы.

РЕШЕНИЕ:

1) Линейные уравнения парной регрессии имеет вид:

$$\tilde{y} = a + bx,$$

$r = b\dfrac{\sigma_x}{\sigma_y}$, откуда $b = r\dfrac{\sigma_y}{\sigma_x}$,

$$b_1 = 0{,}68 \cdot \frac{38}{12} = 2{,}15$$

$a = \bar{y} - b \cdot \bar{x}$ = 250 – 2,15·47 = 148,95;

$$\tilde{y} = 148{,}95 + 2{,}15 \cdot x_1 \tag{3.27}$$

$$b_2 = 0{,}63 \cdot \frac{38}{21} = 1{,}14,$$

$a = \bar{y} - b \cdot \bar{x}$ = 250 – 1,14·112 = 122,32;

$$\tilde{y} = 122{,}32 + 1{,}14 \cdot x_2 \tag{3.28}$$

Оценим значимость уравнения (3.27):

$$F_{факт} == \frac{r^2}{1-r^2}(n-2) = \frac{0{,}68^2}{1-0{,}68^2}(30-2) = 24{,}1,$$

$F_{табл.}$=4,20

Оценим значимость уравнения (3.28):

$$F_{факт} = \frac{r^2}{1-r^2}(n-2) = \frac{0{,}63^2}{1-0{,}63^2}(30-2) = 18{,}43,$$

Так как $F_{факт}$ > $F_{табл}$, то уравнения регрессии значимы, статистически надежны.

2) Уравнение регрессии в стандартизованном масштабе имеет вид:

$$t_y = \beta_1 t_{x1} + \beta_2 t_{x2} + \varepsilon,$$

где t – стандартизованные переменные, для которых среднее значение равно 0, а среднее квадратическое отклонение равно 1;

β – стандартизованные коэффициенты регрессии.

$$\beta_1 = \frac{r_{yx1} - r_{yx2} \cdot r_{x1x2}}{1 - r_{x1x2}^2} = \frac{0{,}68 - 0{,}63 \cdot 0{,}42}{1 - 0{,}42^2} = 0{,}504,$$

$$\beta_2 = \frac{r_{yx2} - r_{yx1} \cdot r_{x1x2}}{1 - r_{x1x2}^2} = \frac{0{,}63 - 0{,}68 \cdot 0{,}42}{1 - 0{,}42^2} = 0{,}418.$$

Получим следующее уравнение в стандартизированном масштабе:

$$t_y = 0{,}504 t_{x1} + 0{,}418 t_{x2}.$$

Таким образом, с увеличением выработки продукции на одного работника на 1 сигму при неизменном среднем уровне индекса цен на продукцию прибыль в среднем увеличится на 0,504 сигмы, а с увеличением индекса цен на продукцию на 1 сигму прибыль возрастет на 0,418 сигмы.

Так как $\beta_1 > \beta_2$, то можно сделать вывод, что большее влияние на прибыль оказывает фактор x_1.

Уравнение в естественной форме имеет вид:

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

где $b_i = \beta_i \dfrac{\sigma_y}{\sigma_{xi}}$

Тогда $b_1 = 0{,}504 \cdot \dfrac{38}{12} = 1{,}60, \qquad b_2 = 0{,}418 \cdot \dfrac{38}{21} = 0{,}76.$

$a = \bar{y} - b_1\bar{x}_1 - b_2\bar{x}_2 = 250 - 1{,}6 \cdot 47 - 0{,}76 \cdot 112 = 89{,}68$.

Следовательно, уравнение регрессии имеет вид:

$$\tilde{y} = 89{,}68 + 1{,}6x_1 + 0{,}76x_2,$$

Таким образом, с увеличением выработки продукции на одного работника на 1 ед., прибыль возрастет на 1,6 тыс. руб. при неизменном среднем уровне индекса цен, а с увеличением индекса цен на 1%, прибыль возрастет в среднем на 0,76 тыс. руб.

3) Множественный коэффициент корреляции определяется по формуле:

$$R_{yx1x2} = \sqrt{\sum \beta_{xi} \cdot r_{yxi}} = \sqrt{\beta_{x1} \cdot r_{yx1} + \beta_{x2} \cdot r_{yx2}},$$

$R_{yx1x2} = \sqrt{0{,}504 \cdot 0{,}68 + 0{,}418 \cdot 0{,}63} = 0{,}778$, то есть связь между признаками сильная.

F-критерий Фишера определим по формуле: $F = \dfrac{R^2}{1-R^2} \cdot \dfrac{n-m-1}{m}$,

$$F_{факт.} = \frac{0{,}778^2}{1-0{,}778^2} \cdot \frac{30-2-1}{2} = 20{,}7.$$

$F_{табл.}$ =3,35 (при k_1=m=2 и k_2=n-m-1=30-2-1=27).

Так как $F_{факт.}$>$F_{табл}$, делаем вывод о статистической значимости уравнения регрессии в целом.

F_{x1} оценивает целесообразность включения в уравнение фактора x_1 после того, как в него был включен фактор x_2. Соответственно, F_{x2} указывает на целесообразность включения в модель фактора x_2 после фактора x_1. Определим частные *F*-критерии для факторов x_1 и x_2 по формулам:

$$F_{x1} = \frac{R^2_{yx1x2} - r^2_{yx2}}{1 - R^2_{yx1x2}}(n-m-1) = \frac{0{,}778^2 - 0{,}68^2}{1-0{,}778^2}(30-1-1) = 10{,}137,$$

$$F_{x2} = \frac{R^2_{yx1x2} - r^2_{yx1}}{1 - R^2_{yx1x2}}(n-m-1) = \frac{0{,}778^2 - 0{,}63^2}{1-0{,}778^2}(30-1-1) = 14{,}751,$$

$F_{табл.}$ =4,21.

Таким образом, включение факторов x_1 и x_2 в модель статистически целесообразно.

3.3 Задания для самостоятельной работы

Задача 3.1

По регионам Российской Федерации за последний год собрать данные с сайта Росстата по следующим социально-экономические показателям:

- удельный вес сельского населения в общей численности населения, %;
- удельный вес населения в трудоспособном возрасте в общей численности населения, %;
- общие коэффициенты рождаемости (число родившихся на 1000 человек населения);
- уровень безработицы, %;
- среднедушевые денежные доходы населения, тыс. руб. в месяц;
- среднемесячная номинальная начисленная заработная плата работников организаций, тыс. руб. в месяц;
- величина прожиточного минимума (в среднем на душу трудоспособного населения), тыс. руб. в месяц;
- потребительские расходы в среднем на душу населения, тыс. руб. в месяц;
- валовой региональный продукт (ВРП) на душу населения, тыс. руб.;
- инвестиции в основной капитал на душу населения, тыс. руб.

В соответствии с заданием, выданным преподавателем, охарактеризовать зависимость результативного признака y от факторных x_1 и x_2.

Задание:

1) Оценить параметры линейной модели множественной регрессии с помощью метода определителей.

2) Построить уравнение регрессии в стандартизованном масштабе.

3) Построить частные уравнения регрессии.

4) Определить частные и средние коэффициенты эластичности.

5) Рассчитать индекс множественной корреляции.

6) Определить коэффициент детерминации (скорректированный и нескорректированный).

7) Рассчитать частные коэффициенты корреляции.

8) Оценить значимость уравнения множественной регрессии с помощью F-критерия Фишера.

9) Рассчитать частные F-критерии Фишера.

10) Оценить значимость коэффициентов чистой регрессии по t-критерию Стьюдента.

Корреляционно-регрессионный анализ проводить с использованием формул и с помощью табличного процессора MS Excel (см. Главу 1).

Задача 3.2

По регионам Центрального федерального округа Российской Федерации изучается зависимость среднемесячной начисленной заработной платы работников (руб.) от факторов: X1 - величины прожиточного минимума в среднем на душу населения, руб. в месяц; x2 – уровень занятости населения, %. Уравнение регрессии имеет вид:

$$\begin{array}{lccc} \tilde{y} = & ? & + \ 1{,}86 \cdot x_1 & + \ 574{,}42 \cdot x_2 \\ m & 22121{,}67 & ? & 398{,}11 \\ t & -1{,}23 & 1{,}15 & ? \end{array}$$

Задание:

1) Восполнить недостающие показатели.

2) Построить доверительный интервал для параметра b_2 с вероятностью 99,0%.

3) Сделать выводы.

Задача 3.3

По 14 регионам Приволжского федерального округа Российской Федерации изучается зависимость среднедушевого денежного дохода населения (тыс. руб.) от следующих факторов: x_1 – валовой региональный продукт на душу населения, тыс. руб.; x_2 – уровень участия в рабочей силе, %; x_3 – доля городского населения в общей численности населения, %. В результате корреляционно-регрессионного анализа получены следующие данные:

$$\tilde{y} = a + 0{,}045x_1 - 0{,}098x_2 + 0{,}254x_3$$

$R^2 = 0{,}844$; $\bar{y} = 21{,}5$; $\bar{x}_1 = 204{,}8$; $\bar{x}_2 = 65{,}8$; $\bar{x}_3 = 74{,}5$.

Задание:

1) Определить значение скорректированного коэффициента детерминации.

2) Рассчитать средние коэффициенты эластичности и параметр a равны.

3) Сделать выводы.

Задача 3.4

По данным регионов Центрального федерального округа Российской Федерации исследуется зависимость производства молока от факторов: x_1 - расход кормов в расчете на одну голову, ц к.ед.; x_2 – надой молока на одну корову, кг. Матрица коэффициентов парной корреляции имеет вид:

	y	x_1	x_2
y	1		
x_1	0,048	1	
x_2	0,549	-0,422	1

Задание:

1) Проверить гипотезу H_0.

2) Сделать выводы.

Задача 3.5

По 30 предприятиям отрасли были получены следующие результаты регрессионного анализа зависимости объема выпуска продукции *y* (млн. руб.) от численности занятых на предприятии x_1 (чел.) и среднегодовой стоимости основных фондов x_2 (млн. руб.):

Коэффициент детерминации	0,72
Множественный коэффициент корреляции	0,85
Уравнение регрессии	$y = ??? + 0{,}48x_1 + 20x_2$
Стандартные ошибки параметров	2 0,06 ???
t- критерий для параметров	1,5 ??? 4

Задание:

1) Восстановить пропущенные характеристики.

2) С вероятностью 0,95 построить доверительные интервалы для коэффициентов регрессии.

Задача 3.6

По 25 предприятиям легкой промышленности получена следующая информация, характеризующая зависимость объема выпуска продукции *y* (млн. руб.) от количества отработанных за год человеко-часов x_1 (тыс.чел.-час.) и среднегодовой стоимости производственного оборудования x_2 (млн. руб.):

Уравнение регрессии	$\tilde{y} = 35 + 0{,}06x_1 + 2{,}5x_2$
Множественный коэффициент корреляции	0,85
Сумма квадратов отклонений расчетных значений результата от фактических	3200

Задание:

1) Определить коэффициент детерминации.

2) Составить таблицу результатов дисперсионного анализа.

3) Проанализировать полученные результаты регрессионного анализа.

Задача 3.7

Изучается зависимость уровня доходов на душу населения (тыс. руб.) в регионе от факторов: x_1 – доля населения с высшим образованием, %; x_2 – доля пенсионеров, %; x_3 – инвестиции на душу населения, тыс. руб. Уравнение регрессии, построенное по 20 наблюдениям, имеет вид:

$y =$	? +	$0{,}36 \cdot x_1 -$	$0{,}28 \cdot x_2 +$	$? \cdot x_3$
m	0,35	?	0,12	0,07
t	3,20	2,94	?	3,43

Задание:

1) Определить недостающие показатели.
2) Дать интерпретацию коэффициентов регрессии.
3) Построить доверительный интервал для параметра b_2 с вероятностью 95,0%. Сделать выводы.

Задача 3.8

Изучается зависимость урожайности сахарной свеклы (ц с 1 га) от факторов: x_1 – количество осадков в период вегетации, г; x_2 – затраты труда на 1 га, чел.-час.; x_3 – доля затрат на удобрения в общих затратах, %. По 18 наблюдениям получены следующие результаты:

$$\tilde{y} = a + 0{,}46 \cdot x_1 - 0{,}25 \cdot x_2 + 2{,}30 \cdot x_3$$

$$R^2 = 0{,}75; \quad \bar{y} = 70; \quad \bar{x}_1 = 110; \quad \bar{x}_2 = 130; \quad \bar{x}_3 = 85.$$

Задание:

1) Определить значения скорректированного коэффициента детерминации, средних коэффициентов эластичности и параметра a.
2) Сделать выводы.

Задача 3.9

Изучается зависимость потребления хлебных продуктов (кг) от факторов: x_1 – доход, тыс. руб.; x_2 – доля пенсионеров в общей численности населения, %. Получены следующие результаты:

$$\bar{y} = 35; \quad \sigma_y = 8; \quad r_{yx1} = 0{,}67;$$

$$\bar{x}_1 = 18; \quad \sigma_{x1} = 6; \quad r_{yx2} = 0{,}82;$$

$$\bar{x}_2 = 12; \quad \sigma_{x2} = 4; \quad r_{x1x2} = 0{,}40$$

Задание:

1) Построить уравнение регрессии в стандартизированном и натуральном масштабе.
2) Сделать выводы.

Задача 3.10

Зависимость потребления электроэнергии (тыс. кВт·ч) от объемов производства продукции А (тыс. ед.) (x_1) и продукции Б (тыс. ед.) (x_2) характеризуется следующим образом:

Уравнение в стандартизованном виде:	$\tilde{t}_y = 0{,}76 \cdot t_{x1} + 0{,}63 \cdot t_{x2}$
Коэффициент детерминации:	0,95
Коэффициент вариации y:	17%
Коэффициент вариации x_1:	27%
Коэффициент вариации x_2:	30%

Задание:

1) Сделать вывод о силе влияния факторов на результат.

2) Определить средние коэффициенты эластичности, сделать по ним выводы.

3) Оценить значимость уравнения регрессии, учитывая, что оно построено по 27 наблюдениям.

Задача 3.11

При анализе деятельности 30 фермерских хозяйств изучается зависимость объема выпуска продукции растениеводства (млн. руб.) от факторов: x_1 – численность работников, чел.; x_2 – количество минеральных удобрений на 1 га посева, кг; x_3 – количество осадков в период вегетации, г.

Уравнение регрессии имеет вид: $\tilde{y} = 3{,}0 + 0{,}6 \cdot x_1 + 1{,}8 \cdot x_2 + 2{,}1 \cdot x_3$; $R^2 = 0{,}75$. С вероятностью 95% получены следующие доверительные интервалы для коэффициентов регрессии:

Граница	Доверительные интервалы для коэффициентов регрессии при факторе		
	x_1	x_2	x_3
Нижняя	0,1	?	1,2
Верхняя	?	2,3	?

Задание:

1) Восстановить пропущенные границы доверительных интервалов.

2) Дать интерпретацию параметров уравнения регрессии и их доверительных интервалов.

3) Оценить значимость уравнения регрессии.

Задача 3.12

Изучается зависимость численности безработных (тыс. чел.) в регионах РФ от следующих факторов: x_1 – доля населения с высшим образованием, %; x_2 – число малых предприятий, тыс. ед. По 30 регионам получены следующие данные:

Уравнение регрессии	$\tilde{y} = a - 0{,}17 \cdot x_1 - 0{,}24 \cdot x_2$
Коэффициент детерминации	0,65
$\bar{y}$	94
$\bar{x}_1$	60
$\bar{x}_2$	40

Задание:

1) Найти параметр a.

2) Определить средние коэффициенты эластичности.

3) Оценить значимость уравнения регрессии в целом. Сделать выводы.

3.4 Тестовые задания

1 Чем ближе к нулю определитель матрицы межфакторной корреляции, тем:

1) сильнее мультиколлинеарность факторов;
2) слабее мультиколлинеарность факторов;
3) ненадежнее результаты множественной регрессии;
4) надежнее результаты множественной регрессии.

2 Если в уравнение множественной регрессии включено три фактора, то число наблюдений должно быть:

1) 6-10; 2) не менее 18-30; 3) не более 18-30.

3 В линейной множественной регрессии параметры при x называются:

1) коэффициентами корреляции;
2) стандартизованными коэффициентами регрессии;
3) коэффициентами условно чистой регрессии.

4 Коэффициенты β называются:

1) коэффициентами корреляции;
2) стандартизованными коэффициентами регрессии;
3) коэффициентами условно чистой регрессии.

5 Стандартизованные коэффициенты регрессии показывают:

1) на сколько процентов изменится в среднем результат, если соответствующий фактор изменится на один процент при неизменном среднем уровне других факторов;
2) на сколько единиц изменится в среднем результат, если соответствующий фактор изменится на одну единицу при неизменном среднем уровне других факторов;
3) на сколько сигм изменится в среднем результат, если соответствующий фактор изменится на одну сигму при неизменном среднем уровне других факторов;
4) на сколько сигм изменится в среднем фактор, если результативный признак изменится на одну сигму.

6 Уравнение множественной регрессии в стандартизованной форме имеет вид:

1) $y = a + b_1x_1 + b_2x_2 + ... + b_nx_n + \varepsilon$;
2) $t_y = \beta_1 \cdot t_{x1} + \beta_2 \cdot t_{x2} + ... + b_n \cdot t_{xn} + \varepsilon$
3) $t_y = a + \beta_1 \cdot t_{x1} + \beta_2 \cdot t_{x2} + ... + b_n \cdot t_{xn} + \varepsilon$.

7 Уравнение регрессии в стандартизованном масштабе имеет вид: $t_y = -1{,}256 \cdot t_{x1} + 0{,}825 \cdot t_{x2} + \varepsilon$. Наибольшее влияние на результат оказывает фактор:

1) x_1; 2) x_2; 3) коэффициенты сравнивать нельзя.

8 При линейной зависимости признаков формула индекса корреляции имеет вид:

1) $R_{yx1x2...xn} = \sqrt{\sum \beta_{xi} + r_{yxi}}$; 2) $R_{yx1x2...xn} = \sqrt{\sum b_{xi} \cdot r_{yxi}}$;

3) $R_{yx1x2...xn} = \sqrt{\sum \beta_{xi} \cdot r_{yxi}}$.

9 Тесноту связи совместного влияния факторов на результат в модели множественной регрессии оценивает:

1) частный коэффициент корреляции;

2) индекс множественной корреляции;

3) множественный коэффициент детерминации.

10 На стадии формирования модели, в частности в процедуре отсева факторов, используют:

1) частные коэффициенты корреляции;

2) коэффициенты условно чистой регрессии;

3) стандартизованные коэффициенты регрессии.

11 Частный F-критерий Фишера оценивает:

1) статистическую значимость присутствия соответствующего фактора в уравнении множественной регрессии;

2) целесообразность включения в уравнение одного фактора после другого фактора;

3) значимость коэффициентов чистой регрессии;

4) значимость уравнения регрессии в целом.

12 Отличия скорректированного коэффициента детерминации от обычного:

1) содержит поправку на число степеней свободы;

2) меньше обычного коэффициента детерминации;

3) больше обычного коэффициент детерминации;

4) характеризует вариацию результативного признака.

13 Дана матрица парных коэффициентов корреляции:

	y	x_1	x_2	x_3
y	1,0	-	-	-
x_1	0,7	1,0	-	-
x_2	-0,5	0,4	1,0	-
x_3	0,4	0,8	-0,1	1,0

Какие факторы являются коллинеарными?

1) x_1 и x_2; 2) x_1 и x_3; 3) x_2 и x_3.

14 Множественный коэффициент корреляции $R_{yx_1x_2} = 0,9$. Определите, какой процент дисперсии зависимой переменной y объясняется влиянием факторов x_1 и x_2:

1) 90% 2) 81%; 3) 19%.

15 При расчете модели множественной регрессии и корреляции в MS Excel для вывода матрицы парных коэффициентов корреляции используется:

1) инструмент анализа данных Корреляция;
2) инструмент анализа данных Регрессия;
3) инструмент анализа данных Описательная статистика.

16 Для расчета множественного коэффициента корреляции в MS Excel используется:

1) инструмент анализа данных Корреляция;
2) инструмент анализа данных Регрессия;
3) инструмент анализа данных Описательная статистика.

17 Для проведения дисперсионного анализа в Excel используется:

1) инструмент анализа данных Корреляция;
2) инструмент анализа данных Регрессия;
3) инструмент анализа данных Описательная статистика.

18 Для расчета множественного коэффициента детерминации в Excel используется:

1) инструмент анализа данных Корреляция;
2) инструмент анализа данных Регрессия;
3) инструмент анализа данных Описательная статистика.

19 Для расчета коэффициентов регрессии в Excel используется:

1) инструмент анализа данных Корреляция;
2) инструмент анализа данных Регрессия;
3) инструмент анализа данных Описательная статистика.

20 Для определения F-критерия в Excel используется:

1) инструмент анализа данных Корреляция;
2) инструмент анализа данных Регрессия;
3) инструмент анализа данных Описательная статистика.

3.5 Контрольные вопросы

1) Что такое спецификация модели множественной регрессии?

2) Каким требованиям должны отвечать факторы, включаемые во множественную регрессию?

3) С помощью какого метода оцениваются параметры уравнения множественной регрессии?

4) Что характеризуют коэффициенты «чистой» регрессии?

5) Что показывают стандартизованные коэффициенты регрессии?

6) Какая связь существует между стандартизованными коэффициентами регрессии и «чистыми» коэффициентами регрессии?

7) Приведите формулу для расчета скорректированного индекса множественной детерминации, от чего он зависит?

8) Как определить совокупный коэффициент корреляции через матрицу парных коэффициентов корреляции?

9) Какие уравнения называют частными уравнениями регрессии?

10) Для чего определяются частные F-критерии?

4 МОДЕЛИРОВАНИЕ ОДНОМЕРНЫХ ВРЕМЕННЫХ РЯДОВ

Цель занятия – изучить основные понятия временных рядов, овладеть навыками определения трендовой, сезонной и случайной компонент аддитивных и мультипликативных моделей временных рядов, а также оценки взаимосвязей по временным рядам.

4.1 Теоретические положения

4.1.1 Модели временных рядов. Автокорреляция уровней временного ряда.

4.1.2 Моделирование сезонных и циклических колебаний.

4.1.3 Методы исключения тенденции

4.1.4 Автокорреляция в остатках. Критерий Дарбина-Уотсона.

4.1.1 Модели временных рядов. Автокорреляция уровней временного ряда.

Модели временных рядов – это модели, построенные по данным, характеризующим один объект за ряд последовательных моментов (периодов) времени.

В большинстве случаев фактический уровень временного ряда можно представить как сумму или произведение трендовой (T), циклической (S) и случайной компонент (E).

Модель, в которой временной ряд представлен как сумма перечисленных компонент, называется ***аддитивной моделью*** временного ряда: Y = T + S + E.

Модель, в которой временной ряд представлен как произведение перечисленных компонент, называется ***мультипликативной моделью*** временного ряда: Y = T · S · E.

Автокорреляцией уровней ряда называют корреляционную зависимость между последовательными уровнями временного ряда. Корреляционную зависимость можно измерить с помощью линейного коэффициента корреляции между уровнями исходного временного ряда и уровнями этого ряда, сдвинутыми на несколько шагов во времени.

Коэффициент автокорреляции уровней ряда первого порядка измеряет зависимость между соседними уровнями ряда t и $t-1$, то есть при лаге 1 и определяется по формуле:

$$r_1 = \frac{\sum_{t=2}^{n}(y_t - \bar{y}_1)\cdot(y_{t-1} - \bar{y}_2)}{\sqrt{\sum_{t=2}^{n}(y_t - \bar{y}_1)^2 \cdot \sum_{t=2}^{n}(y_{t-1} - \bar{y}_2)^2}}, \qquad (4.1)$$

где $\bar{y}_1 = \frac{\sum_{t=2}^{n} y_t}{n-1}$, $\qquad$ (4.2)

$$\bar{y}_2 = \frac{\sum_{t=2}^{n} y_{t-1}}{n-1} \quad . \tag{4.3}$$

4.1.2 Моделирование сезонных и циклических колебаний

Моделирование циклических колебаний аналогично моделированию сезонных колебаний.

Если амплитуда колебаний приблизительно постоянна, строят аддитивную модель временного ряда.

Если амплитуда сезонных колебаний вырастает или уменьшается, строят мультипликативную модель, которая ставит уровни ряда в зависимости от значений сезонной компоненты.

Процесс построения модели состоит из следующих шагов:

1. Выравнивание исходного ряда методом скользящей средней.
2. Расчет значений сезонной компоненты S.
3. Устранение сезонной компоненты из исходных уравнений ряда и получение выравненных данных (T+E) в аддитивной или (T×E) в мультипликативной модели.
4. Аналитическое выравнивание уровней (T+E) или (T×E) и расчет значений с использованием полученного уравнения тренда.
5. Расчет полученных по модели значений (T+S) или (T×S).
6. Расчет абсолютных и относительных ошибок.

4.1.3 Методы исключения тенденции

Сущность всех методов исключения тенденции заключается в том, чтобы устранить или зафиксировать воздействие фактора времени на формирование уровней ряда.

1) Метод отклонений от тренда. Пусть имеются два временных ряда x_t и y_t, каждый из которых содержит трендовую компоненту T и случайную компоненту ε. Проведение аналитического выравнивания по каждому из этих рядов позволяет найти параметры соответствующих уравнений трендов и определить по тренду расчетные (теоретические) уровни $\tilde{x}_t$ и $\tilde{y}_t$. Эти расчетные значения можно принять за оценку трендовой компоненты T каждого ряда. Поэтому влияние тенденции можно устранить путем вычитания из фактических значений уровней расчетных значений, то есть:

$$U_x = x_t - \tilde{x}_t, \tag{4.4}$$

$$U_y = y_t - \tilde{y}_t, \tag{4.5}$$

где U_x, U_y – отклонения от трендов.

Учитывая, что средние величины отклонений от линейных и параболических трендов всегда равны нулю, а от других форм тренда близки к нулю, то $\overline{U}_x = \overline{U}_y = 0$, следовательно, формула расчета коэффициента корреляции приобретает следующий вид:

$$r_{UxUy} = \frac{\sum_{i=1}^{n} U_{Xi} U_{Yi}}{\sqrt{\sum_{i=1}^{n} U_{Xi}^2 \cdot \sum_{i=1}^{n} U_{Yi}^2}}. \qquad (4.6)$$

Метод корреляции отклонений от трендов является основным, и используется независимо от того, одинаковы типы трендов коррелируемых показателей или нет.

2) ***Метод последовательных разностей.*** Если временной ряд содержит ярко выраженную линейную тенденцию, ее можно устранить путем замены исходных уровней ряда цепными абсолютными приростами (первыми разностями):

$$\Delta_t = y_t - y_{t-1} \qquad (4.7)$$

Если временной ряд содержит тенденцию в форме параболы второго порядка то для ее устранения можно заменить исходные уровни ряда на вторые разности: $\Delta_t = y_t - y_{t-1}$ – первые разности; $\Delta_t = \Delta_t - \Delta_{t-1}$ – вторые разности.

Если тенденции временного ряда соответствует экспоненциальный или степенной тренд, метод последних разностей не к исходным уровням ряда, а к их логарифмам.

3) Включение в модель регрессии фактора времени. В данном методе тенденция фиксируется через включение фактора времени в модель в качестве независимой переменной.

Модель, включающая фактор времени имеет вид:

$$y_t = a + b_1 \cdot x_1 + b_2 \cdot t + \varepsilon. \qquad (4.8)$$

Для расчета параметров такого уравнения используется МНК (метод наименьших квадратов). Система нормальных уравнений имеет вид:

$$\begin{cases} n \cdot a + b_1 \cdot \sum x + b_2 \cdot \sum t = \sum y \\ a \cdot \sum x + b_1 \cdot \sum x^2 + b_2 \cdot \sum t \cdot x = \sum x \cdot y \\ a \cdot \sum t + b_1 \sum t \cdot x + b_2 \cdot \sum x^2 = \sum t \cdot y \end{cases} \qquad (4.9)$$

К преимуществам модели с включением фактора времени относятся: учет всей информации, содержащейся в исходных данных; построение по всей совокупности данных за рассматриваемый период; определение параметров уравнения обычным МНК.

4.1.4 Автокорреляция в остатках. Критерий Дарбина-Уотсона.

Существует два наиболее распространенных метода определения автокорреляции остатков:

1) Построение графика зависимости остатков от времени и визуальное определение наличия или отсутствия автокорреляции.

2) Использование критерия Дарбина – Уотсона и расчет величины:

$$d = \frac{\sum_{t=2}^{n}(\varepsilon_t - \varepsilon_{t-1})^2}{\sum_{t=1}^{n}\varepsilon_t^2} \ . \qquad (4.10)$$

d есть отношение суммы квадратов разностей последовательных значений остатков к остаточной сумме квадратов по модели регрессии.

Между критерием Дарбина – Уотсона и коэффициентом автокорреляции остатков первого порядка существует следующая связь:

$$d \approx 2 \cdot (1 - r_1). \qquad (4.11)$$

Таким образом, если в остатках существует полная положительная автокорреляция и $r_1 = 1$, то $d = 0$. Если в остатках полная отрицательная автокорреляция, то $r_1 = -1$ и, следовательно, $d = 4$. Если автокорреляция остатков отсутствует, то $r_1 = 0$ и $d = 2$.

Следовательно, $0 \le d \le 4$.

Алгоритм выявления автокорреляции остатков на основе критерия Дарбина – Уотсона следующий.

Выдвигаются следующие гипотезы:

H_0 - об отсутствии автокорреляции остатков (в остатках нет автокорреляции);

H_1 – в остатках есть положительная автокорреляция;

Далее по специальным таблицам (Приложение В) определяются критические значения критерия Дарбина- Уотсона d_L и d_U для заданного числа наблюдений n, числа независимых переменных модели k и уровня значимости α. По этим значениям числовой промежуток [redacted]; 4 разбивают на пять отрезков. Принятие или отклонение каждой из гипотез с вероятностью (1-α) рассматривается на рисунке 4.1.

Есть положительная автокорреляция остатков. H_0 отклоняется. С вероятностью P=(1-α) принимается H_1	Зона неопределен-ности	Нет оснований отклонять H_0 (автокорреляция остатков отсутствует)	Зона неопределен-ности	Есть отрицательная автокорреляция остатков. H_0 отклоняется. С вероятностью P=(1-α) принимается H_1^*

0 d_L d_U 2 4-d_U 4-d_L 4

Рисунок 4.1 Механизм проверки гипотезы о наличии автокорреляции остатков

Если фактическое значение критерия Дарбина – Уотсона попадает в зону неопределенности, то на практике предполагают существование автокорреляции остатков и отклоняют гипотезу H_0.

4.2 Решение типовых задач

Пример 4.1

По данным о выручке от реализации продукции в ООО «Рассвет» (таблица 4.1) построить аддитивную модель временного ряда.

Таблица 4.1 Исходные данные для построения модели временного ряда

Годы	2013 г.				2014 г.				2015 г.			
Квартал, t	1	2	3	4	1	2	3	4	1	2	3	4
Выручка, млн. руб., y_t	50,2	45,2	39,0	32,5	46,3	40,5	35,4	31,2	43,5	35,0	32,1	27,5

РЕШЕНИЕ:

Временной ряд выручки от реализации продукции изобразим графически (рисунок 4.2).

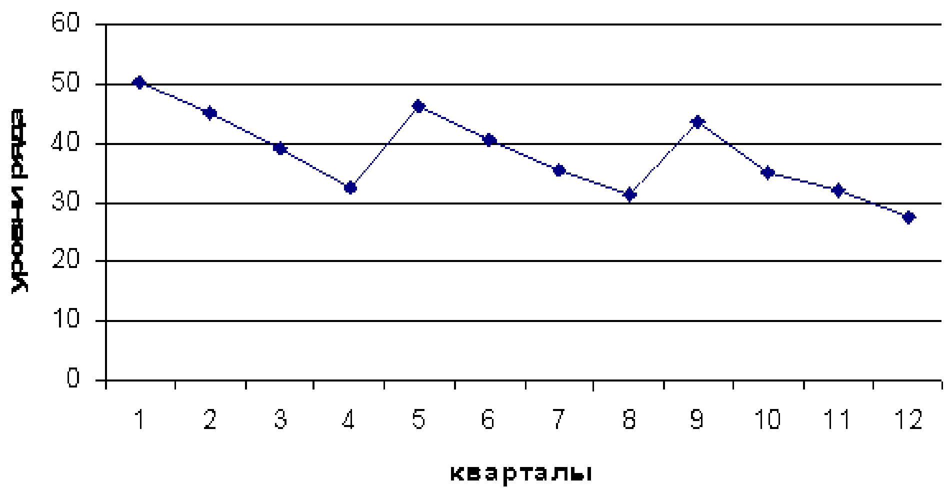


Рисунок 4.2 Временной ряд выручки от реализации продукции

Если амплитуда колебаний приблизительно постоянна, строят аддитивную модель временного ряда, в которой значения сезонной компоненты предполагаются постоянными для различных циклов.

График показывает, что ряд содержит сезонные колебания периодичностью 4. По графику данного ряда можно установить наличие приблизительно равной амплитуды колебаний, что свидетельствует о возможности построения аддитивной модели временного ряда.

Шаг 1. Проведем выравнивание исходных уровней ряда методом скользящей средней. Для этого составим вспомогательную таблицу 4.2:

1) просуммируем уровни ряда последовательно за каждые четыре квартала со сдвигом на один момент времени (графа 3);

Таблица 4.2 Расчет оценок сезонной компоненты в аддитивной модели

Номер квартала, t	Уровни ряда, y_t	Итого за четыре квартала	Скользящая средняя за 4 квартала	Центрированная скользящая средняя	Оценка сезонной компоненты
1	2	3	4	5	6
1	50,2	-	-	-	-
2	45,2	166,9	41,725	-	-
3	39,0	163,0	40,750	41,238	-2,238
4	32,5	158,3	39,575	40,163	-7,663
5	46,3	154,7	38,675	39,125	7,175
6	40,5	153,4	38,350	38,513	1,988
7	35,4	150,6	37,650	38,000	-2,600
8	31,2	145,1	36,275	36,963	-5,763
9	43,5	141,8	35,450	35,863	7,638
10	35,0	138,1	34,525	34,988	0,013
11	32,1	-	-	-	-
12	27,5	-	-	-	-

2) разделим полученные суммы на 4 и найдем скользящие средние (графа 4). Полученные таким образом выравненные значения уже не содержат сезонной компоненты (их количество будет меньше количества уровней исходного временного ряда на 3 единицы);

3) приведем эти значения в соответствие с фактическими моментами времени, для этого найдем средние значения их двух последовательных скользящих средних – центрированные скользящие средние (графа 5) (их количество будет меньше количества уровней исходного временного ряда на 4 единицы).

Шаг 2. Найдем оценки сезонной компоненты как разность между фактическими уровнями ряда и центрированными скользящими средними (графа 6 таблицы 2). Используем эти оценки для расчета значений сезонной компоненты S (таблица 4.3).

Таблица 4.3 Расчет значений сезонной компоненты в аддитивной модели

Показатель	Год	Номер квартала			
		1	2	3	4
	1	-	-	-2,238	-7,663
	2	7,175	1,988	-2,600	-5,763
	3	7,638	0,013	-	-
Итого за i-й квартал (за все годы)	×	14,813	2,001	-4,838	-13,426
Средняя оценка сезонной компоненты для i-ого квартала, $\overline{S}_i$	×	7,407	1,001	-2,419	-6,713
Скорректированная сезонная компонента, S_i	×	7,588	1,182	-2,238	-6,532

Найдем средние за каждый квартал (по всем годам) оценки сезонной компоненты S_i. В моделях с сезонной компонентой обычно предполагается, что сезонные воздействия за период взаимопогашаются. В аддитивной модели это выражается в том, что сумма значений сезонной компоненты по всем кварталам должна быть равна нулю.

Имеем для данной модели: 7,407+1,001-2,419-6,713 = -0,724.

Определим корректирующий коэффициент:

$$k = \frac{-0{,}724}{4} = -0{,}181.$$

Рассчитаем скорректированные значения сезонной компоненты как разность между ее средней оценкой и корректирующим коэффициентом k:

$S_i = \bar{S}_i - k$, где $i = 1:4$.

Проверим условие равенства нулю суммы значений сезонной компоненты: 7,588+1,182-2,238-6,532 = 0.

Таким образом, получены следующие значения сезонной компоненты:

1 квартал: $S_1 = 7{,}588$; 2 квартал: $S_2 = 1{,}182$;

3 квартал: $S_3 = -2{,}238$; 4 квартал: $S_4 = -6{,}532$.

Занесем полученные значения в таблицу 4.4 для соответствующих кварталов каждого года (графа 3).

Шаг 3. Вычтем значение сезонной компоненты из каждого уровня исходного временного ряда, чтобы устранить ее влияние. Получим: T + E = Y – S (графа 4 таблицы 4.4). Эти значения рассчитываются для каждого момента времени и содержат только тенденцию и случайную компоненту.

Таблица 4.4 Расчет выравненных значений T и ошибок E в аддитивной модели

t	y_t	S_i	T + E = $y_t - S_i$	T	T + S	E = y_t – (T + S)	E^2	$(y_t - \bar{y})^2$
1	2	3	4	5	6	7	8	9
1	50,2	7,588	42,612	43,078	50,666	-0,466	0,217	144,0
2	45,2	1,182	44,018	42,191	43,373	1,827	3,337	49,0
3	39,0	-2,238	41,238	41,304	39,066	-0,066	0,004	0,6
4	32,5	-6,532	39,032	40,417	33,885	-1,385	1,919	32,5
5	46,3	7,588	38,712	39,530	47,118	-0,818	0,670	65,6
6	40,5	1,182	39,318	38,643	39,825	0,675	0,455	5,3
7	35,4	-2,238	37,638	37,757	35,519	-0,119	0,014	7,8
8	31,2	-6,532	37,732	36,870	30,338	0,862	0,744	49,0
9	43,5	7,588	35,912	35,983	43,571	-0,071	0,005	28,1
10	35,0	1,182	33,818	35,096	36,278	-1,278	1,633	10,2
11	32,1	-2,238	34,338	34,209	31,971	0,129	0,017	37,2
12	27,5	-6,532	34,032	33,322	26,790	0,710	0,504	114,5
Итого	458,4	0,000	458,4	458,400	×	×	9,519	543,9

Шаг 4. Определим компоненту Т данной модели. Для этого проведем выравнивание ряда (Т + Е) с помощью линейного тренда (таблица 4.5).

Таблица 4.5 Расчет линейного тренда уровней временного ряда

№ п/п	t	$(T+E)$	$(T+E)\cdot t$	t^2
1	1	42,612	42,612	1
2	2	44,018	88,036	4
3	3	41,238	123,714	9
4	4	39,032	156,128	16
5	5	38,712	193,560	25
6	6	39,318	235,908	36
7	7	37,638	263,466	49
8	8	37,732	301,856	64
9	9	35,912	323,208	81
10	10	33,818	338,180	100
11	11	34,338	377,718	121
12	12	34,032	408,384	144
Итого	78	458,400	2852,770	650

Для оценки параметров a и b необходимо составить систему нормальных уравнений:

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

Система нормальных уравнений имеет вид:

$$\begin{cases} 12\cdot a + 78\cdot b = 458{,}4, \\ 78\cdot a + 650\cdot b = 2852{,}77. \end{cases}$$

Решив ее, получили параметры:

$a = 43{,}965, \quad b = -0{,}887$

Итак, линейный тренд имеет вид: $T = 43{,}965 - 0{,}887\cdot t$.

Найдем уровни Т для каждого момента времени (графа 5 таблицы 4.4). Для этого в уравнение тренда подставим порядковый номер квартала.

Шаг 5. Найдем значения уровней ряда, полученные по аддитивной модели. Для этого прибавим к уровням Т значения сезонной компоненты для соответствующих кварталов (графа 6 таблицы 4.4).

Шаг 6. Рассчитаем ошибку (случайную компоненту Е) модели. Численные значения абсолютных ошибок приведены в таблице 4.4 (графа 7).

Таким образом, мы рассчитали количественные значения трендовой, сезонной и случайной компонент уровней временного ряда за каждый квартал за три года по аддитивной модели.

Так, например, расчеты за четвертый квартал 2015 г. (12-й уровень ряда) показывают, что если бы ряд содержал только трендовую составляющую (тенденцию уровней – ежеквартальное уменьшение выручки от реализации на 0,887 млн. руб.), то выручка составила бы 33,322 млн. руб.

Отнимая (прибавляя) сезонную компоненту, равную за четвертый квартал -6,532 млн. руб., мы получаем уровень ряда 33,322 – 6,532 = 26,79 млн. руб. Однако из-за воздействия случайной составляющей (о причинах которой мы можем предполагать), равной 0,710, фактическая выручка в четвертом квартале 2015 г. составила 26,79 +0,710 = 27,5 млн. руб.

Для оценки качества построения модели можно использовать сумму квадратов абсолютных ошибок. Для данной построенной аддитивной модели сумма квадратов абсолютных ошибок равна $\sum E^2 =$ 9,519. По отношению к общей сумме квадратов отклонений уровней ряда от его среднего уровня, равной $(y_t - \bar{y})^2 = 543{,}9$, эта величина (коэффициент детерминации) составляет: $R^2_{yt} = 1 - \frac{9{,}519}{543{,}9} = 0{,}982$ или 98,2%, то есть аддитивная модель объясняет 98,2% общей вариации уровней временного ряда выручки от реализации за 2013 – 2015 гг.

На основе построенной модели сделаем точечный прогноз ожидаемой выручки в течение первого квартала 2016 года. Прогнозное значение уровня временного ряда F_t в аддитивной модели – это сумма трендового значения T_t и соответствующего значения сезонной компоненты S_t.

Для расчета трендовых значений воспользуемся уравнением тренда: $T = 43{,}965 - 0{,}887 \cdot t$ Первый квартал четвертого в ряду года будет стоять под номером 13, следовательно:

$T_{13} = 43{,}965 - 0{,}887 \cdot 13 = 32{,}434$.

Значение сезонной компоненты за первый квартал равно $S_1 = 7{,}588$.

Прогнозное значение составит:

$F_{13} = T_{13} + S_1 = 32{,}434 + 7{,}588 = 40{,}022$.

Таким образом, величина выручки от реализации в первом квартале 2016 года составит 40,022 млн. руб.

Пример 4.2

На основе данных таблицы 4.6 о выплате дивидендов (усл. ден.ед) в ООО «Весна» построить мультипликативную модель временного ряда.

Таблица 4.6 Исходные данные для построения моделей временного ряда

Годы	2013 г.				2014 г.				2015 г.			
Квартал, t	1	2	3	4	1	2	3	4	1	2	3	4
y_t	40	60	50	30	50	80	70	50	60	100	80	60

РЕШЕНИЕ:

Изобразим временной ряд графически (рисунок 4.2).

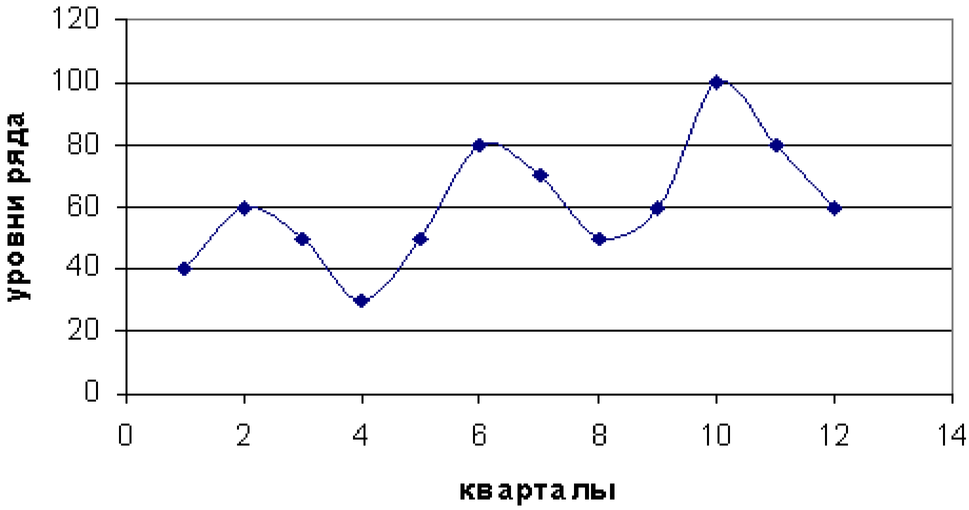


Рисунок 4.2 Временной ряд выплаты дивидендов

По графику временного ряда можно установить наличие приблизительно линейного тренда и сезонных колебаний (период равен 4) возрастающей амплитуды, поэтому строим мультипликативную модель.

Шаг 1. Проведем выравнивание исходных уровней ряда методом скользящей средней по той же методике расчета, что и для аддитивной модели (таблица 4.7).

Таблица 4.7 Расчет оценок сезонной компоненты в мультипликативной модели

Номер квартала, t	Уровни ряда, y_t	Итого за четыре квартала	Скользящая средняя за 4 квартала	Центрированная скользящая средняя	Оценка сезонной компоненты
1	2	3	4	5	6
1	40	-	-	-	-
2	60	180	45,0	-	-
3	50	190	47,5	46,25	1,081
4	30	210	52,5	50,00	0,600
5	50	230	57,5	55,00	0,909
6	80	250	62,5	60,00	1,333
7	70	260	65,0	63,75	1,098
8	50	280	70,0	67,50	0,741
9	60	290	72,5	71,25	0,842
10	100	300	75,0	73,75	1,356
11	80	-	-	-	-
12	60			-	-

Шаг 2. Найдем оценки сезонной компоненты как частное от деления фактических уровней ряда (графа 6 таблица 4.7). Используем эти оценки для расчета значений сезонной компоненты S (таблица 4.8). Для этого найдем средние за каждый квартал оценки сезонной компоненты S_i. Взаимопогашаемость сезонных воздействий в мультипликативной модели выражается в том, что сумма значений сезонной компоненты по всем кварталам должна быть равна числу периодов в цикле, т.е. четырем (4 квартала в году).

Таблица 4.8 Расчет сезонной компоненты в мультипликативной модели

Показатель	Год	Номер квартала			
		1	2	3	4
	1 2 3	- 0,909 0,842	- 1,333 1,356	1,081 1,098 -	0,600 0,741 -
Итого за i-й квартал (за все годы)	×	1,751	2,689	2,179	1,341
Средняя оценка сезонной компоненты для i-ого квартала, $\bar{S}_i$	×	0,876	1,345	1,090	0,671
Скорректированная сезонная компонента, S_i	×	0,880	1,351	1,095	0,674

Имеем: 0,876 + 1,345 + 1,090 + 0,671 = 3,980.

Рассчитаем корректирующий коэффициент:

$$k = \frac{4}{3{,}98} = 1{,}005.$$

Определим скорректированные значения сезонной компоненты, умножив ее средние оценки на корректирующий коэффициент k: $S_i = \bar{S}_i \cdot k$, где $i = 1:4$.

Проверим условие равенства суммы значений сезонной компоненты четырем: 0,880 + 1,351 + 1,095 + 0,674 = 4.

Получили следующие значения сезонной компоненты:

1 квартал: $S_1 = 0{,}880$; 2 квартал: $S_2 = 1{,}351$;

3 квартал: $S_3 = 1{,}095$; 4 квартал: $S_4 = 0{,}674$.

Занесем полученные значения в таблицу 4.8 для соответствующих кварталов каждого года (графа 3).

Шаг 3. Разделим каждый уровень исходного ряда на соответствующие значения сезонной компоненты. Получим T · E = Y / S (графа 4 таблицы 4.9). Эти значения рассчитываются для каждого момента времени и содержат только тенденцию и случайную компоненту.

Таблица 4.9 Расчет выравненных значений Т и ошибок Е
в мультипликативной модели

t	y_t	S_i	T · E = y_t / S_i	T	T · S	E = y_t : (T·S)	E = y_t – (T·S)	E^2	$(y_t - \bar{y})^2$
1	2	3	4	5	6	7	8	9	10
1	40	0,880	45,45	40,178	35,357	1,131	4,643	21,561	433,889
2	60	1,351	44,41	44,062	59,528	1,008	0,472	0,223	0,689
3	50	1,095	45,66	47,946	52,501	0,952	-2,501	6,254	117,289
4	30	0,674	44,51	51,830	34,933	0,859	-4,933	24,339	950,489
5	50	0,880	56,82	55,714	49,028	1,020	0,972	0,944	117,289
6	80	1,351	59,22	59,598	80,517	0,994	-0,517	0,267	367,489
7	70	1,095	63,93	63,482	69,513	1,007	0,487	0,237	84,089
8	50	0,674	74,18	67,366	45,405	1,101	4,595	21,117	117,289
9	60	0,880	68,18	71,250	62,700	0,957	-2,700	7,290	0,689
10	100	1,351	74,02	75,134	101,506	0,985	-1,506	2,268	1534,289
11	80	1,095	73,06	79,018	86,525	0,925	-6,525	42,572	367,489
12	60	0,674	89,02	82,902	55,876	1,074	4,124	17,008	0,689
Итого	730	12,000	738,46	738,480	733,388	12,013	-3,388	144,080	4091,667

Шаг 4. Определим компоненту Т в мультипликативной модели. Для этого проведем выравнивание ряда (Т•Е) с помощью линейного тренда (таблица 4.10).

Для оценки параметров a и b необходимо составить систему нормальных уравнений:

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

Система нормальных уравнений составит: $\begin{cases} 12 \cdot a + 78 \cdot b = 738{,}46, \\ 78 \cdot a + 650 \cdot b = 5355{,}380. \end{cases}$

Решив ее, получаем: a = 36,294; b = 3,884.

Таким образом, линейный тренд имеет вид: $T = 36{,}294 + 3{,}884 \cdot t$.

Найдем уровни Т для каждого момента времени (графа 5 таблицы 4.9).

Шаг 5. Найдем уровни ряда по мультипликативной модели, умножив уровни Т на значения сезонной компоненты для соответствующих кварталов (графа 6 таблицы 4.9).

Таблица 4.10 Расчет линейного тренда уровней временного ряда

№ п/п	t	$(T \cdot E)$	$(T \cdot E) \cdot t$	t^2
1	1	45,45	45,455	1
2	2	44,41	88,823	4
3	3	45,66	136,986	9
4	4	44,51	178,042	16
5	5	56,82	284,091	25
6	6	59,22	355,292	36
7	7	63,93	447,489	49
8	8	74,18	593,472	64
9	9	68,18	613,636	81
10	10	74,02	740,192	100
11	11	73,06	803,653	121
12	12	89,02	1068,249	144
Итого	78	738,46	5355,380	650

Шаг 6. Расчет ошибки в мультипликативной модели проводится по формуле Е = y_t: (Т · S). Численные значения ошибки приведем в графе 7 таблицы 4.9.

Чтобы сравнить мультипликативную модель ряда с другими моделями, используем сумму квадратов абсолютных ошибок. Абсолютные ошибки в мультипликативной модели определяются по формуле: Е = y_t - (Т·S). В данной модели сумма квадратов абсолютных ошибок составляет 144,080. Общая сумма квадратов отклонений фактических уровней ряда от среднего значения $(y_t - \bar{y})^2 = 4091{,}667$. Доля объясненной дисперсии уровней ряда динамики равна: $R^2_{y_t} = 1 - \dfrac{144{,}080}{4091{,}667} = 0{,}965$ или 96,5% − мультипликативная модель объясняет 96,5% общей вариации уровней временного ряда.

Прогнозное значение уровня временного ряда F_t в мультипликативной модели есть произведение трендового значения T_t и соответствующего значения сезонной компоненты S_t.

Для расчета трендовых значений воспользуемся уравнением тренда: $T = 36{,}294 + 3{,}884 \cdot t$. Первый квартал 2016 г. будет стоять под номером 13 во временном ряду, следовательно: $T_{13} = 36{,}294 + 3{,}884 \cdot 13 = 86{,}786.$ Значение сезонной компоненты за первый квартал равно $S_1 = 0{,}880$.

Прогнозное значение составит: $F_{13} = T_{13} \cdot S_1 = 86{,}786 \cdot 0{,}88 = 76{,}372.$

Выплата дивидендов в первом квартале 2016 года составит 76,372 усл. ден. ед.

Пример 4.3

На основе помесячных данных о потреблении электроэнергии в регионе (млн. кВт·ч) за последние 3 года построена аддитивная модель временного ряда. Скорректированные значения сезонной компоненты за соответствующие месяцы приводятся ниже:

январь	+25	май	−32	сентябрь	+2
февраль	+10	июнь	−38	октябрь	+15
март	+6	июль	−25	ноябрь	+27
апрель	−4	август	−18	декабрь	?

Уравнение тренда выглядит следующим образом: T = 300 + 1,5t. При расчете параметров тренда для моделирования переменной времени использовались натуральные числа t = 1:36.

Задание:

1) Определить значение сезонной компоненты за декабрь.

2) На основе построенной модели дать точечный прогноз ожидаемого потребления электроэнергии в течение первого квартала следующего года.

РЕШЕНИЕ:

1) В аддитивной модели сумма значений сезонной компоненты по всем кварталам должна быть равна нулю, значит:

$$S = 0 – (25+10+6-4-32-38-25-18+2+15+27) = 32.$$

2) Общий вид аддитивной модели следующий:

$$Y = T + S + E.$$

Прогнозное значение F_t уровня временного ряда в аддитивной модели есть сумма трендовой и сезонной компонент. Объем электроэнергии, потребленной в течение первого квартала следующего года, рассчитывается как сумма объемов потребления электроэнергии в январе, феврале, марте, соответственно F_{37}, F_{38} и F_{39}.

Для определения трендовой компоненты воспользуемся уравнением тренда:

$$T = 300 + 1,5t.$$

Получим:

$T_{37} = 300 + 1,5·37=355,5;$

$T_{38} = 300 + 1,5·38=357;$

$T_{39} = 300 + 1,5·39 =358,5.$

Значения сезонной компоненты равны:

$S_{янв.(1)} = +25$; $S_{февр.(2)} = +10$; $S_{март\ (3)} = +6$.

Таким образом,

$F_{37} = T_{37}+S_1=355,5 + 25 = 380,5$;

$F_{38} = T_{38}+S_2= 357 + 10 = 367$;

$F_{39} = T_{39}+S_3= 358,5 + 6=364,5$.

Прогноз объема потребления электроэнергии на первый квартал следующего года составит:

380,5 + 367 + 364,5 = 1112 млн. кВт·ч.

3.3 Задания для самостоятельной работы

Задача 4.1

Имеются данные об урожайности зерновых культур в сельскохозяйственных организациях по Российской Федерации:

Годы	Урожайность зерновых, ц с 1 га убранной площади
2007	20,5
2008	24,6
2009	23,6
2010	19,0
2011	23,3
2012	19,3
2013	23,1
2014	25,4
2015	25,0

Задание:

1) Определить среднегодовое изменение урожайности зерновых культур.

2) Сделать прогноз урожайности зерновых на 2017, 2018 и 2019 гг., если тенденция устойчива.

Задача 4.2

Администрация Сбербанка изучает динамику вкладов (депозитов) на рублевых счетах за ряд лет (млрд. руб.). Исходные данные представлены ниже.

Годы	2009	2010	2011	2012	2013	2014	2015
Депозиты физических лиц, y	8,5	9,9	12,4	13,7	14,6	16,4	14,9

Задание:

1) Построить уравнение линейного тренда и дать интерпретацию его параметров.

2) Определить коэффициент детерминации для линейного тренда.

3) Администрация банка предполагает, что среднегодовой абсолютный прирост депозитов физических лиц составляет не менее 1,1 млрд. руб. Подтверждается ли это предположение результатами, которые вы получили?

Задача 4.3

Имеются данные об урожайности сахарной свеклы в хозяйствах всех категорий в регионах Приволжского федерального округа Российской Федерации (таблица 4.11).

Таблица 4.11 Урожайность сахарной свеклы, ц с 1 га убранной площади

Регион ПФО	2009г.	2010г.	2011г.	2012г.	2013г.	2014г.	2015г.
Республика Башкортостан	254	97	249	209	338	250	259
Республика Мордовия	394	287	393	453	411	372	368
Республика Татарстан	274	167	244	371	399	292	356
Чувашская Республика	223	211	274	263	338	264	262
Нижегородская область	175	123	235	290	315	187	166
Пензенская область	254	141	349	355	401	304	325
Саратовская область	192	131	289	331	432	300	312
Ульяновская область	150	161	307	309	368	262	302

Задание:

1) Обоснуйте выбор типа уравнения тренда.
2) Рассчитайте параметры уравнения тренда.
3) Дайте прогноз урожайности зерновых на следующий год.

Задача 4.4

Имеются данные об урожайности зерновых и зернобобовых культур по странам мира:

Годы	Урожайность зерновых и зернобобовых, ц с 1 га						
	Россия	Австрия	Беларусь	Индия	Австралия	Канада	Египет
2006	18,9	55,9	24,9	21,0	10,2	29,0	73,5
2007	19,8	57,0	28,5	21,9	12,1	28,1	72,5
2008	23,8	66,8	35,2	22,6	16,5	32,1	73,1
2009	22,7	60,6	33,3	22,4	16,6	31,0	70,1
2010	18,3	60,8	27,7	22,5	16,7	31,8	63,8
2011	22,4	69,5	32,2	23,7	19,9	33,1	71,1
2012	18,3	59,0	34,4	24,7	21,5	32,9	71,7

Задание:

1) По каждой стране выбрать уравнение тренда.
2) Рассчитать параметры уравнения.

Задача 4.5

Имеются следующие данные об уровне безработицы y_t (%) в Российской Федерации:

Годы	2008	2009	2010	2011	2012	2013	2014	2015
y_t	6,2	8,3	7,3	6,5	5,5	5,5	5,2	5,6

Задание:

1) Определить коэффициент автокорреляции уровней ряда I порядка.

2) Обосновать выбор уравнения тренда и определить его параметры.

3) Интерпретировать полученные результаты.

Задача 4.6

По данным Российской Федерации изучается зависимость оборота розничной торговли пищевыми продуктами (y_t) от динамики потребительских цен (x_t). Полученные за последние 6 лет данные представлены в таблице 4.12.

Таблица 4.12 Динамика потребительских цен

Годы	2009	2010	2011	2012	2013	2014
ИПЦ, раз	1,09	1,09	1,06	1,07	1,07	1,11
Оборот розничной торговли, млрд. руб.	7,1	8,0	9,1	10,0	11,1	12,4

Известно также, что $\sum x_t$=6,49; $\sum y_t$=57,7; $\sum x_t\, y_t$ = 62,42; $\sum x_t^2$ =7,02.

Задание:

1) Построить модель зависимости оборота розничной торговли пищевыми от ИПЦ с включением фактора времени.

2) Дать интерпретацию параметров полученной модели.

Задача 4.7

Годовое потребление мяса и мясопродуктов и располагаемые ресурсы домашних хозяйств в Российской Федерации приведены в таблице 4.13.

Таблица 4.13 Динамика показателей домашних хозяйств

Показатели	Годы								
	2007	2008	2009	2010	2011	2012	2013	2014	2015
Потребление мяса и мясопродуктов, кг в год	71	75	73	79	81	83	85	85	85
Располагаемые ресурсы, тыс.руб.на человека	9,1	12,0	12,4	14,6	16,6	18,6	21,2	22,9	23,1

Задание:

1) Определите уравнение регрессии, включив в него фактор времени, если известно, что $\sum y=717$, $\sum x=150{,}5$, $\sum xy=12208{,}9$, $\sum x^2=2724{,}2$, $\sum y^2=57361$.

2) Интерпретируйте полученные результаты.

Задача 4.8

По данным таблицы 4.14 на основе предварительного анализа амплитуды колебаний построить аддитивную или мультипликативную модели временного ряда.

Таблица 4.14 Исходные данные для построения модели временного ряда

Годы	2014 г.				2015 г.				2016 г.			
Квартал, *t*	1	2	3	4	1	2	3	4	1	2	3	4
Выручка, млн. руб.,	60,5	55,2	50,0	43,5	56,3	51,5	44,4	40,2	53,5	45,0	41,1	37,5
Объем продукции, млн. руб.	50	70	60	40	60	70	80	60	70	110	90	70
Выплата дивидендов, млн. руб.	30	50	40	20	40	70	60	40	50	90	70	50

Задача 4.9

Для прогнозирования объема продаж ООО «Заря» (млн. руб.) на основе поквартальных данных за 2012-2016 гг. была построена аддитивная модель временного ряда объема продаж. Уравнение, моделирующее динамику трендовой компоненты этой модели, имеет вид: T = 100 + 2t (при построении тренда для моделирования переменной времени использовались натуральные числа, начиная с 1). Показатели за 2015 г., полученные в ходе построения аддитивной модели, представлены ниже.

Время года	Фактический объем продаж в 2015 г.	Компонента, полученная по аддитивной модели		
		трендовая	сезонная	случайная
Зима	?	103	−6	?
Весна	119	104	?	?
Лето	?	106	?	19
Осень	121	108	−29	42

Задание:

Определить недостающие в таблице данные, учитывая, что объем продаж ООО «Заря» за 2015 г. в целом составил 510 млн. руб.

Задача 4.10

На основе поквартальных данных об уровне безработицы за последние 5 лет была построена мультипликативная модель временного ряда. Скорректированные значения сезонной компоненты за каждый квартал приводятся ниже:

I квартал	1,4	III квартал	0,7
II квартал	0,8	IV квартал	?

Уравнение тренда выглядит следующим образом: T = 9,2-0,3t (при расчете параметров тренда для нумерации кварталов использовались натуральные числа t = 1:20).

Задание:

1) Определить значения сезонной компоненты за IV квартал.

2) На основе построенной модели дать точечные прогнозы уровня безработицы на I и II квартал следующего года.

Задача 4.11

В результате регрессионного моделирования уровня рождаемости по данным Российской Федерации получены остатки регрессии:

Год	Остатки
2004	0,2
2005	0,1
2006	-0,5
2007	-0,6
2008	0,0
2009	0,4
2010	0,4
2011	0,2
2012	-0,1
2013	0,3
2014	-0,1
2015	-0,4

Задание:

1) Рассчитать Критерий Дарбина-Уотсона.

2) Сделать выводы относительно рассматриваемой регрессии.

Задача 4.10

На основе квартальных данных объемов продаж 2013-2016 гг. была построена аддитивная модель временного ряда. Трендовая компонента имеет вид: $T = 260 + 3 \cdot t, \quad (t = 1,2,3,\ldots)$. Показатели за 2015 г. приведены ниже.

Квартал	Фактический объем продаж, тыс.т	Компонента аддитивной модели		
		трендовая	сезонная	случайная
1	270	?	?	-9
2	?	?	-15	+4
3	324	?	8	?
4	?	?	?	?
Итого	1342	х	х	×

Задание:

1) Определите недостающие показатели в таблице.
2) Сделать выводы.

4.4 Тестовые задания

1 Моделями временных рядов называются модели, построенные по данным:

1) характеризующим совокупность различных объектов в определенный момент (период) времени;

2) характеризующим один объект за ряд последовательных моментов (периодов) времени;

3) любым имеющимся данным об изучаемой совокупности.

2 Обязательными элементами временного ряда являются:

1) тенденция; 2) циклические колебания;

3) время; 4) уровень ряда.

3 Модель, в которой временной ряд представлен как сумма трендовой, сезонной и случайной компонент – это:

1) аддитивная модель; 2) мультипликативная модель;

3) регрессионная модель.

4 Модель, в которой временной ряд представлен как произведение трендовой, сезонной и случайной компонент – это:

1) аддитивная модель;

2) мультипликативная модель;

3) регрессионная модель.

5 Аддитивную модель строят, если:

1) амплитуда сезонных колебаний приблизительно постоянна;

2) амплитуда сезонных колебаний возрастает или уменьшается.

6 Мультипликативную модель строят, если:

1) амплитуда сезонных колебаний приблизительно постоянна;

2) амплитуда сезонных колебаний возрастает или уменьшается.

7 Сумма значений сезонной компоненты по всем кварталам в аддитивной модели должна быть равна:

1) нулю;

2) четырем;

3) числу периодов в цикле.

8 Сумма значений сезонной компоненты по всем периодам в мультипликативной модели в цикле должна быть равна:

1) нулю;

2) числу периодов в цикле;

3) четырем.

9 Прогнозное значение уровня временного ряда в аддитивной модели – это:

1) произведение трендовой и сезонной компонент;

2) сумма трендовой и сезонной компонент;

3) разница между трендовой и сезонной компонентами;

4) сумма трендовой и случайной компонент.

10 Прогнозное значение уровня временного ряда в мультипликативной модели – это:

1) произведение трендовой и сезонной компонент;

2) сумма трендовой и сезонной компонент;

3) разница между трендовой и сезонной компонентами;

4) произведение трендовой и случайной компонент.

11 В моделях с сезонной компонентой сезонные воздействия за период:

1) взаимопогашаются;

2) взаимо не погашаются;

3) среди ответов нет правильного.

12 Значение уровней ряда в аддитивной модели находится как:

1) $T - S$;

2) $T \cdot S$;

3) $T + S$.

13 Для определения трендовой компоненты в мультипликативной модели временного ряда необходимо:

1) выровнять ряд ($T \cdot E$);

2) выровнять ряд ($T + E$);

3) трендовая компонента обычно уже дается в условии задачи.

14 Расчет абсолютных ошибок в мультипликативной модели проводится по формуле:

1) $У_t-(T+S)$;   2) $У_t:(T\cdot S)$;   3) $У_t-(T\cdot S)$.

15 Число периодов, по которым рассчитывается коэффициент автокорреляции, – это:

1) лаг;

2) моменты ряда;

3) среди ответов нет правильного.

16 По знаку коэффициента автокорреляции:

1) можно делать вывод о возрастающей или убывающей тенденции в уровнях ряда;

2) нельзя делать вывод о возрастающей или убывающей тенденции в уровнях ряда;

3) среди ответов нет правильного.

17 Если наиболее высоким оказался коэффициент автокорреляции первого порядка, то исследуемый ряд содержит:

1) только тенденцию;

2) циклические колебания;

3) нельзя сделать никаких определенных выводов о ряде.

18 Метод отклонений от тренда предполагает:

1) вычисление трендовых значений для каждого временного ряда модели и расчет отклонений от трендов с целью дальнейшего анализа по этим данным;

2) замену исходных данных разностями;

3) включение фактора времени.

19 Критерий Дарбина – Уотсона и коэффициент автокорреляции остатков первого порядка связаны соотношением:

1) $d=2\cdot(1-r_1)$;

2) $r_1=2\cdot(1-d)$;

3) $d=4\cdot(2-r_1)$.

20 Фактическое значение критерия Дарбина – Уотсона, попадающее в интервал от d_U до ($4-d_U$) означает, что:

1) автокорреляция остатков существует, гипотеза H_0 отклоняется;

2) автокорреляция остатков отсутствует, гипотеза H_0 не отклоняется;

3) есть отрицательная автокорреляция остатков, гипотеза H_0 отклоняется, принимается другая гипотеза H_1.

4.5 Контрольные вопросы

1) Что такое временной ряд?
2) Назовите основные элементы временного ряда.
3) Из каких компонент состоит уровень ряда?
4) В каком случае строят аддитивную модель, а в каком – мультипликативную?
5) Назовите этапы построения модели временного ряда.
6) В чем сущность методов исключения тенденции?
7) Что такое автокорреляция остатков? Чем она вызвана?
8) Для чего используется критерий Дарбин-Уотсона?
9) Можно ли по знаку коэффициента автокорреляции делать вывод о возрастающей или убывающей тенденции в уровнях ряда?
10) Как находится прогнозное значение уровня временного ряда в аддитивной модели ?

ВОПРОСЫ К ЗАЧЕТУ

1 Определение, предмет и задачи дисциплины «Эконометрика».
2 Эконометрика и ее место в ряду математико-статистических и экономических дисциплин.
3 Задачи экономического анализа, решаемые на основе эконометрики.
4 Некоторые сведения об истории возникновения эконометрики.
5 Особенности эконометрического метода.
6 Эконометрические модели и их типы.
7 Основные понятия эконометрического моделирования.
8 Типы эконометрических моделей.
9 Основные этапы и проблемы эконометрического моделирования.
10 Виды переменных: экзогенные (объясняющие, факторные), эндогенные (результирующие, прогнозируемые).
11 Понятие о функциональной, статистической и корреляционной связях.
12 Спецификация модели.
13 Метод наименьших квадратов (МНК) и условия его применения для определения параметров уравнения парной регрессии.
14 Уравнение регрессии, его смысл и назначение.
15 Показатели качества регрессии (F-критерий Фишера, t-критерий Стьюдента, доверительные интервалы для коэффициентов регрессии).
16 Внутренне линейные и внутренне нелинейные модели. Модель Филлипса. Модель Энгеля.
17 Расчет коэффициентов эластичности для наиболее распространенных типов уравнений регрессии.
18 Корреляция для нелинейной регрессии.
19 Отбор факторов при построении множественной регрессии.
20 Выбор формы уравнения регрессии.
21 Оценка и интерпретация параметров уравнения множественной регрессии
22 Уравнение множественной регрессии в стандартизованном масштабе.
23 Частные уравнения регрессии и частные коэффициенты корреляции.
24 Множественный коэффициент корреляции и детерминации.
25 Оценка надежности результатов множественной регрессии и корреляции.
26 Понятие, примеры гетероскедастичности.
27 Проверка линейной регрессии на гетероскедастичность.
28 Обобщенный метод наименьших квадратов (ОМНК).
29 Регрессионные модели с переменной структурой (фиктивные переменные).

30 Предпосылки метода наименьших квадратов.

31 Основные этапы изучения, моделирования и прогнозирования временного ряда.

32 Основные элементы временного ряда.

33 Автокорреляция уровней временного ряда и выявление его структуры.

34 Моделирование тенденции временного ряда.

35 Моделирование сезонных и циклических колебаний (аддитивная и мультипликативная модели временного ряда).

36 Специфика статистической оценки взаимосвязи двух временных рядов.

37 Методы исключения тенденции (метод отклонений от тренда, метод последовательных разностей, включение в модель регрессии фактора времени).

38 Автокорреляция в остатках. Критерий Дарбина-Уотсона.

39 Оценивание параметров уравнения регрессии при наличии автокорреляции в остатках.

40 Коинтеграция временных рядов.

ГЛОССАРИЙ

Автокорреляция – корреляция между временной переменной и лаговой переменной, составленной от той же переменной.

Авторегрессионная модель – разновидность динамической эконометрической модели, которая содержит в качестве факторных переменных лаговые значения эндогенных переменных.

Авторегрессия – регрессия, учитывающая влияние предыдущих уровней ряда на последующие.

Аддитивная модель временного ряда — модель, в которой все компоненты ряда динамики представлены как сумма этих составляющих.

Адекватность модели – соответствие построенной модели моделируемому реальному экономическому объекту или процессу.

Бета-коэффициент – коэффициент, который показывает, на какую часть своего среднего квадратического отклонения изменится в среднем значение результативного признака при изменении факторного признака на величину своего среднеквадратического отклонения.

Вариация – изменение значений признака внутри изучаемой совокупности.

Временной ряд – ряд последовательно расположенных во времени числовых показателей, которые характеризуют уровень, состояния и изменения явления или процесса.

Гетероскедастичность – нарушение равенства дисперсий ошибок регрессии

Статистическая гипотеза – различного рода предположения относительно характера или параметров распределения случайной переменной, которые можно проверить, опираясь на результаты наблюдений в случайной выборке.

Гомоскедастичность – свойство постоянства дисперсий ошибок регрессии.

Графический метод — способ распознавания типа тренда, при котором временные интервалы откладывают на оси абсцисс, величины уровней – по оси ординат.

Дисперсионный анализ – статистический метод для оценки влияния различных факторов на результат эксперимента, а также для последующего планирования аналогичных экспериментов.

Дисперсия случайной величины – математическое ожидание квадрата ее отклонения от математического ожидания.

Доверительный интервал – числовой интервал, который с заданной доверительной вероятностью накрывает неизвестное значение параметра.

Зависимая переменная – в регрессионной модели некоторая переменная Y, являющаяся функцией регрессии с точностью до случайного возмущения.

Задачи эконометрики – построение экономических моделей, оценка их параметров, проверка гипотез о свойствах экономических показателей, установление видов их взаимосвязей.

Задачи регрессионного анализа – установление формы зависимости между переменными, оценка функции регрессии, оценка неизвестных значений (прогноз значений) зависимой переменной.

Интервальный прогноз – интервал, в котором с определенной вероятностью находится фактическое значение прогнозируемой переменной.

Ковариация двух случайных величин – математическое ожидание произведения отклонений этих величин от своих математических ожиданий.

Коинтеграция – причинно-следственная связь в уровнях двух или более временных рядов, которая выражается в совпадении или противоположной направленности их тенденций и случайной колеблемости.

Коррелированность – наличие линейной зависимости между двумя случайными величинами.

Коррелограмма – график выборочной автокорреляционной функции.

Корреляционная связь – зависимость, при которой изменение среднего значения результативного признака обусловлено изменением факторных признаков.

Корреляционный анализ заключается в количественном определении тесноты связи между двумя признаками (при парной связи) и между результативным и множеством факторных признаков (при многофакторной связи).

Корреляция – это статистическая зависимость между случайными величинами, при которой изменение одной из случайных величин приводит к изменению математического ожидания другой.

Коэффициент вариации – показатель относительной колеблемости признака, отношение среднего квадратического отклонения случайной величины к ее математическому ожиданию.

Коэффициент детерминации – квадрат коэффициента корреляции, показывает, какая часть дисперсии результативного признака объяснена уравнением регрессии; мера качества уравнения регрессии.

Коэффициент корреляции двух случайных величин – величина, рассчитываемая по наблюдениям над двумя случайными величинами и характеризующая степень их связи; отношение ковариации двух случайных величин

Коэффициент эластичности оказывает, на сколько процентов изменяется результативный признак у при изменении факторного признака х на один процент.

Криволинейная зависимость – это связь, при которой с возрастанием величины факторного признака возрастание (или убывание) результативного признака происходит неравномерно (выражаются уравнениями кривых линий).

Критерий Стьюдента (t-критерий) – применяется для оценки статистической значимости коэффициентов полученного уравнения регрессии.

Критерий Фишера (F-критерий) – применяется для оценки статистической значимости полученного уравнения регрессии в целом.

Критерий Энгеля-Грэнджера – применяется для тестирования временных рядов на коинтеграцию.

Лаг – смещение во времени изменения одного показателя по сравнению с изменением другого.

Математическая модель – представляет реальный объект определенными математическими структурами: функциями, уравнениями, неравенствами и т.п.

Метод наименьших квадратов (МНК) – метод оценки параметров уравнения регрессии, основанный на минимизации суммы квадратов отклонений расчетных значений (по уравнению регрессии) зависимой переменной от ее наблюдаемых значений.

Множественная корреляция – зависимость между результативным признаком и двумя и более факторными признаками, включенными в исследование.

Модель временного ряда – разновидность эконометрической модели, в которой результативный признак является функцией переменной времени или переменных, относящихся к другим моментам времени.

Мультиколлинеарность – тесная зависимость между факторными признаками, включенными в модель.

Мультипликативная модель временного ряда – модель, в которой факторы влияния представлены в виде произведения составляющих

Обобщенный метод наименьших квадратов – метод оценки параметров уравнения регрессии, в котором МНК применяется к уравнению регрессии, преобразованному таким образом, чтобы исключить гетероскедастичность остатков.

Парная корреляция – связь между двумя признаками (результативным и факторным или двумя факторными).

Коэффициент регрессии показывает, на какую величину в среднем изменится результативный признак, если переменную х увеличить на единицу измерения.

Поле корреляции – совокупность точек на координатной плоскости, изображающих наблюдения.

Предмет эконометрики – количественная оценка взаимосвязи между случайными событиями, признаками, показателями, факторами переменных экономических объектов.

Пространственные данные – набор сведений по разным объектам, взятым за один и тот же период времени.

Прямолинейная зависимость – связь, при которой с возрастанием величины факторного признака происходит равномерное возрастание (или убывание) величин результативного признака.

Уравнение регрессии – уравнение линии, вокруг которой группируются точки корреляционного поля; указывает основное направление и тенденцию связи.

Регрессионный анализ – раздел математической статистики, изучающий характер связи между случайными переменными.

Регрессия – зависимость среднего значения какой-либо величины (у) от некоторой другой величины или от нескольких величин (xi).

Результативный признак – признак, изменяющийся под действием факторных признаков.

Статистическая зависимость – связь, при которой каждому значению независимой переменной x соответствует множество значений зависимой переменной y, причем неизвестно заранее, какое именно значение примет y.

Тест Глейзера – применяется для проверки остатков регрессии на гетероскедастичность.

Тест Гольдфельда–Квандта – применяется для проверки остатков регрессии на гетероскедастичность.

Точечный прогноз – среднее прогнозное значение изучаемой переменной экономического объекта.

Тренд – изменение, определяющее общее направление развития, основную тенденцию временного ряда.

Фиктивные переменные – переменные, полученные путем перевода качественных признаков переменных в количественных, то есть при присвоении цифровых меток.

Функциональная связь – связь, при которой определенному значению факторного признака соответствует одно и только одно значение результативного признака.

Частная корреляция – зависимость между результативным и одним факторным признаками или двумя факторными признаками при фиксированном значении других факторных признаков.

Частные уравнения регрессии – характеризующие изолированное влияние одного из факторов xi на результативную переменную у при исключении влияния остальных факторов, входящих в общее уравнение регрессии.

Эконометрика – наука, изучающая конкретные количественные и качественные взаимосвязи экономических объектов и процессов с помощью математических и статистических методов и моделей.

***Этапы эконометрического исследования*: 1)постановка проблемы; 2) получение данных, анализ их качества; 3)спецификация модели; 4)оценка параметров; 5)интерпретация результатов.

БИБЛИОГРАФИЧЕСКИЙ СПИКОК

1. Айвазян С. А. Прикладная статистика и основы эконометрики : учебник / С. А. Айвазян, В. С. Мхитарян. – Москва : ЮНИТИ, 1998. – 1022 с.
2. Айвазян, С. А. Эконометрика-2: продвинутый курс с приложениями в финансах : учебник / С. А. Айвазян, Д. Фантаццини. – Москва : Магистр : Инфра-М, 2015. – 942 с.
3. Афанасьев В. Н. Эконометрика : учебник для студентов вузов / В. Н. Афанасьев, М. М. Юзбашев, Т. И. Гуляева ; под ред. В. Н. Афанасьева. – Москва : Финансы и статистика, 2006. – 255 с.
4. Грубер Й. Эконометрия : учебное пособие / Й. Грубер. – Киев : Нічлава. – 1999. – Т. 2 : Эконометрические прогнозные и оптимизационные модели. – 1999. – 308 с.
5. Гусаров В. М. Общая теория статистики : учебное пособие / В. М. Гусаров, С. М. Проява. – 2-е изд., перераб. и доп. – Москва : ЮНИТИ, 2008. – 207 с.
6. Дубина И. Н. Математико-статистические методы в эмпирических социально-экономических исследованиях : учебное пособие по дисциплине "Эконометрика" для студентов вузов / И. Н. Дубина. – Москва : Финансы и статистика : ИНФРА-М, 2010. – 414 с.
7. Елисеева И. И. Общая теория статистики : учебник / И. И. Елисеева, М. М. Юзбашев ; под ред. И. И. Елисеевой. – 5-е изд., перераб. и доп. – Москва : Финансы и статистика, 2008. – 655 с.
8. Ефимова М. Р. Общая теория статистики : учебник / М. Р. Ефимова, Е. В. Петрова, В. Н. Румянцев. – 2-е изд., испр. и доп. – Москва : ИНФРА-М, 2009. – 413 с.
9. Зинченко А. П. Статистика : учебник / А. П. Зинченко. – Москва : КолосС, 2007. – 567 с.
10. Колемаев, В. А. Эконометрика : учебник для студентов вузов / В. А. Колемаев. – Москва : ИНФРА-М, 2007. – 160 с.
11. Кремер, Н. Ш. Эконометрика : учебник для студентов вузов / Н. Ш. Кремер, Б. А. Путко ; под ред. Н. Ш. Кремера. – Москва : ЮНИТИ, 2007. – 311 с.
12. Лезина М. Л. Статистика : учебное пособие / М. Л. Лезина. – Москва : Наука и образование, 2008. – 367 с.
13. Мазуркин, П. М. Статистическая эконометрика : учебное пособие / П. М. Мазуркин. – Йошкар-Ола : МарГТУ, 2006. – 374 с.
14. Новиков А. И. Эконометрика : учебное пособие / А. И. Новиков. – Москва : Дашков и К, 2015. – 223 с.
15. Новиков, А. И. Эконометрика : учебное пособие / А. И. Новиков. – 2-е изд., испр. и доп. – Москва : ИНФРА-М, 2008. – 144 с.
16. Новиков, А. И. Эконометрика : учебное пособие / А. И. Новиков. – Москва : Дашков и К, 2015. – 223 с.

17. Плохотников, К. Э. Основы эконометрики в пакете STATISTICA : учебное пособие для студентов вузов / К. Э. Плохотников. – Москва : Вузовский учебник, 2011. – 297 с.

18. Практикум по эконометрике : учебное пособие / И. И. Елисеева [и др.] ; под ред. И. И. Елисеевой. – 2-е изд., испр. и доп. – Москва : Финансы и статистика, 2007. – 1 эл. опт. диск (CD-ROM).

19. Практикум по эконометрике : учебное пособие / И. И. Елисеева [и др.] ; под ред. И. И. Елисеевой. – 2-е изд., перераб. и доп. – Мосуква : Финансы и статистика, 2008. – 344 с.

20. Практикум по эконометрике : учебное пособие / И. И. Елисеева [и др.] ; под ред. И. И. Елисеевой. – Москва : Финансы и статистика, 2003. – 192 с.

21. Регионы России. Социально-экономические показатели. 2015: Стат. сб. / Росстат. − Москва, 2015. − 1266 с.

22. Статистика : учебник / В. С. Мхитарян [и др.] ; под ред. В. С. Мхитаряна. – 14-е изд., стер. – Москва : Академия, 2015. – 304 с.

23. Статистика : учебник / И. И. Елисеева [и др.] ; под ред. И. И. Елисеевой. – Москва : Проспект, 2013. – 448 с.

24. Статистика : учебник / Л. И. Ниворожкина [и др.] ; под общ. ред. Л. И. Ниворожкиной. – 2-е изд., доп. и перераб. – Москва : Дашков и К, 2015. – 415 с.

25. Статистика : учебник / Л. П. Харченко [и др.]. – 3-е изд., перераб. и доп. – Москва : ИНФРА-М, 2010. – 445 с.

26. Статистика и бухгалтерский учет : учебное пособие / А. П. Зинченко [и др.]. – Москва : КолосС, 2008. – 435 с.

27. Статистика. Базовый курс : учебник для бакалавров : рек. М-вом образования / [М. В. Боченина и др.] ; под ред. И. И. Елисеевой ; Санкт-Петербургский университет экономики и финансов. - М. : Юрайт, 2011. - 483 с.

28. Тимофеев В. С. Эконометрика : учебник для студентов вузов / В. С. Тимофеев, А. В. Фаддеенков, В. Ю. Щеколдин. – 2-е изд., перераб. и доп. – Москва : Юрайт, 2013. – 328 с.

29. Чураков, Е. П. Прогнозирование экономических временных рядов : учебное пособие для студентов вузов / Е. П. Чураков. – Москва : Финансы и статистика, 2008. – 205 с.

30. Шундалов Б. М. Статистика агропромышленного комплекса : учебное пособие / Б. М. Шундалов. – Минск : ИВЦ Минфина, 2008. – 298 с.

31. Эконометрика : учебник / под ред. В. Б. Уткина. – 2-е изд. – Москва : Дашков и К, 2012. – 561 с.

32. Эконометрика : учебник для магистров : для студентов вузов / И. И. Елисеева [и др.] ; под ред. И. И. Елисеевой. – Москва : Юрайт, 2012. – 449 с.

33. Эконометрика : учебник для студентов вузов / И. И. Елисеева [и др.] ; под ред. И. И. Елисеевой. – 2-е изд., перераб. и доп. – Москва : Финансы и статистика, 2008. – 575 с.

34. Эконометрика : учебник для студентов вузов / И. И. Елисеева [и др.] ; под ред. И. И. Елисеевой. – Москва : Проспект, 2010. – 288 с.

35. Эконометрика [Электронный ресурс] : учебник для магистров / И. И. Елисеева [и др.] ; под ред. И. И. Елисеевой. – Москва : Юрайт, 2012. – 453 с.
36. Экономическая статистика : учебник / А. Р. Алексеев [и др.] ; под ред. Ю. Н. Иванова. – 4-е изд., перераб. и доп. – Москва : ИНФРА-М, 2011. – 667 с.

ПРИЛОЖЕНИЕ А

Таблица А1 Значения F-критерия Фишера для уровня значимости $\alpha = 0,05$

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

Таблица А2 Критические значения t-критерия Стьюдента для уровней значимости 0,10, 0,05, 0,01 (двухсторонний)

Число степеней свободы $d.f.$	α			Число степеней свободы $d.f.$	α		
	0,10	0,05	0,01		0,10	0,05	0,01
1	6,3138	12,706	63,657	18	1,7341	2,1009	2,8784
2	2,9200	4,3027	9,9248	19	1,7291	2,0930	2,8609
3	2,3534	3,1825	5,8409	20	1,7247	2,0860	2,8453
4	2,1318	2,7764	4,6041	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	∞	1,6449	1,9600	2,5758

Таблица А3 Значения статистик Дарбина-Уотсона $d_L d_U$ при 5%-ном уровне значимости

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

ПРИЛОЖЕНИЕ Б

Таблица Б1 Формулы расчета показателей парной регрессии и корреляции

Показатель	Формула расчета
1 Система нормальных уравнений МНК	$\begin{cases} na + b\sum x = \sum y; \\ a\sum x + b\sum x^2 = \sum xy \end{cases}$
2 Коэффициент регрессии	$b = \frac{\operatorname{cov}(x,y)}{\sigma_x^2} = \frac{\overline{yx} - \bar{y}\cdot\bar{x}}{\overline{x^2} - \bar{x}^2}$
3 Параметр *a*	$a = \bar{y} - b\cdot\bar{x}$
4 Линейный коэффициент парной корреляции	$r = \frac{\overline{xy} - \bar{x}\cdot\bar{y}}{\sigma_x\cdot\sigma_y} = b\frac{\sigma_x}{\sigma_y}$
5 Индекс корреляции	$\rho_{xy} = \sqrt{1 - \frac{\sum(y - \tilde{y}_x)^2}{\sum(y - \bar{y})^2}}$
6 Средняя ошибка аппроксимации	$\bar{A} = \frac{1}{n}\sum\left\lvert\frac{y - \tilde{y}}{y}\right\rvert\cdot 100\%$
7 Средний коэффициент эластичности	$\bar{Э} = f'(x)\frac{\bar{x}}{\bar{y}}$
8 *F*-критерий Фишера	$F_{факт.} = \frac{\sum(\tilde{y} - \bar{y})^2 / m}{\sum(y - \tilde{y})^2 / (n - m - 1)} = \frac{r^2}{1 - r^2}\cdot\frac{(n - m - 1)}{m}$
9 *t*-критерий Стьюдента	$t_a = \frac{a}{m_a}$; $t_b = \frac{b}{m_b}$; $t_r = \frac{r}{m_r}$
10 Случайные ошибки параметров линейной регрессии	$m_a = \sqrt{D_{ост.}}\frac{\sqrt{\sum x^2}}{n\sigma_x}$; $m_b = \frac{\sqrt{D_{ост.}}}{\sigma_x\sqrt{n}}$; $m_r = \sqrt{\frac{1 - r^2}{n - 2}}$
11 Предельная ошибка параметров регрессии	$\Delta_a = t_{табл.}\cdot m_a$; $\Delta_b = t_{табл.}\cdot m_b$; $\Delta_r = t_{табл.}\cdot m_r$
12 Доверительные интервалы параметров регрессии	$\gamma_a = a \pm \Delta_a$; $\gamma_b = b \pm \Delta_b$; $\gamma_r = r \pm \Delta_r$

Таблица Б2 Основные показатели множественной линейной регрессии и корреляции

Показатель	Формула расчета
1	2
1 Уравнение множественной регрессии в естественной форме	$\tilde{y} = a + b_1 x_1 + b_2 x_2 + \ldots + b_n x_n + \varepsilon$
2 Уравнение регрессии в стандартизованном масштабе	$t_y = \beta_1 \cdot t_{x1} + \beta_2 \cdot t_{x2} + \ldots + b_n \cdot t_{xn} + \varepsilon$
3 Стандартизованные коэффициенты регрессии ($\beta -$коэффициенты)	$\beta_1 = \frac{r_{yx1} - r_{yx2} \cdot r_{x1x2}}{1 - r_{x1x2}^2}$; $\beta_2 = \frac{r_{yx2} - r_{yx1} \cdot r_{x1x2}}{1 - r_{x1x2}^2}$
4 Связь коэффициентов условно чистой регрессии b_i со стандартизованными коэффициентами регрессии β_i	$b_i = \beta_i \frac{\sigma_y}{\sigma_{xi}}$
5 Связь стандартизованных коэффициенты регрессии с коэффициентами эластичности	$\bar{Э}_i = \beta_i \cdot \frac{V_y}{V_{xi}}$
6 Частные уравнения регрессии	$\tilde{y}_{x1 \cdot x2,\ldots xn} = a + b_1 \cdot x_1 + b_2 \cdot \bar{x}_2 + \ldots + b_n \cdot \bar{x}_n + \varepsilon,$ $\tilde{y}_{x2 \cdot x1,\ldots,xn} = a + b_1 \cdot \bar{x}_1 + b_2 \cdot x_2 + \ldots + b_n \cdot \bar{x}_n + \varepsilon$
7 Частные коэффициенты эластичности	$Э_{y_{xi}} = b_i \frac{x_i}{\tilde{y}_{xi \cdot x1x2\ldots xn}}$
8 Средний коэффициент эластичности	$\bar{Э}_{yxi} = b_i \cdot \frac{\bar{x}_i}{\bar{y}_{xi}},$
9 Индекс множественной корреляции	$R_{yx1x2\ldots xn} = \sqrt{1 - \frac{\sigma_{ост.}^2}{\sigma_y^2}}$; $R_{yx1x2\ldots xn} = \sqrt{\beta_{x1} \cdot r_{yx1} + \beta_{x2} \cdot r_{yx2} + \ldots + \beta_n \cdot r_{yn}}$; $R_{yx1x2\ldots чт} = \sqrt{1 - \frac{\Delta r}{\Delta r_{11}}}$
10 Скорректированный индекс множественной детерминации	$\bar{R}^2 = 1 - (1 - R^2) \cdot \frac{(n-1)}{(n-m-1)}$
1	2
11 Коэффициенты частной корреляции	$r_{yx_1 \cdot x_2} = \sqrt{1 - \frac{1 - R_{yx_1x_2}^2}{1 - r_{yx_2}^2}}$; $r_{yx_2 \cdot x_1} = \sqrt{1 - \frac{1 - R_{yx_1x_2}^2}{1 - r_{yx_1}^2}}$
12 F-критерий Фишера	$F = \frac{R^2}{1 - R^2} \cdot \frac{n-m-1}{m}$

1	2
13 Частные F-критерии Фишера	$F_{x1} = \dfrac{R^2_{yx1x2} - r^2_{yx2}}{1 - R^2_{yx1x2}}(n - m - 1)$; $F_{x2} = \dfrac{R^2_{yx1x2} - r^2_{yx1}}{1 - R^2_{yx1x2}}(n - m - 1)$
14 t-критерий Стьюдента коэффициентов регрессии	$t_{b_i} = \dfrac{b_i}{m_{b_i}}$
15 Связь частных F-критериев с t-критерием Стьюдента	$t_{bi} = \sqrt{F_{xi}}$