

МИНИСТЕРСТВО СЕЛЬСКОГО ХОЗЯЙСТВА РФ

ФГБОУ ВО «Кубанский государственный
аграрный университет им. И.Т. Трубилина»

Учетно-финансовый факультет
Кафедра статистики и прикладной математики

ЭКОНОМЕТРИКА

(продвинутый уровень)

Методические рекомендации
к выполнению рубежной контрольной работы
для студентов магистратуры заочной формы обучения
направления «Экономика»

Краснодар
КубГАУ
2020

Составители:

Н. Х. Ворокова, А. Е. Жминько, А. Е. Сенникова

Эконометрика (продвинутый уровень): метод. рекомендации к выполнению контрольной работы / сост. А. Е. Сенникова, Н. Х. Ворокова, А.Е. Жминько – Издательство: Краснодарский ЦНТИ – филиал ФГБУ «РЭА» Минэнерго России, 2021 –63 с.

Методические рекомендации содержат задания к выполнению рубежной контрольной работы, отражающей необходимый уровень знаний, а также примеры решения задач, позволяющие освоить учебный материал самостоятельно.

Предназначены для студентов-магистров заочной формы обучения направления 38.04.01 «Экономика».

© Н. Х. Ворокова,
А. Е. Жминько,
А. Е. Сенникова, 2021

© ФГБОУ ВО «Кубанский государственный аграрный университет имени И. Т. Трубилина»,
2021

ВВЕДЕНИЕ

Успешная практическая деятельность специалиста экономического профиля все в большей мере зависит от рациональной организации сбора и обработки информации, от умения строить и анализировать эконометрические модели взаимосвязей экономических явлений и процессов. Совершенствование технологий получения, хранения и обработки данных – создание баз, хранилищ (банков) данных и знаний во всех сферах деятельности человека предъявляет новые требования к уровню подготовки специалистов. Большой объем информации, которой сопровождается деятельность практически любой организации, обычно содержит полезные сведения, благодаря которым можно значительно повысить эффективность работы, совершенствуя технологию, организацию, управление и т.д.

Методика получения, обработки и последующего анализа данных включает в себя описательную и аналитическую статистику, визуализацию данных и др.

Эконометрика – это наука, изучающая конкретные количественные закономерности и взаимосвязи экономических явлений, объектов и процессов с помощью статистических и математических методов и моделей. Она базируется на экономической теории, экономической статистике, статистико-математическом инструментарии. Эконометрика придает количественное выражение качественным закономерностям экономических явлений и процессов.

Задачи дисциплины:

а) освоение методов и приемов эконометрического анализа статистических данных;

б) изучение аппарата и техники разработки математических моделей связей и зависимостей между экономическими явлениями и процессами;

в) формирование навыков количественной оценки состояния, развития и прогнозирования социально-экономических явлений;

г) подготовка специалистов, обладающих навыками исследовательской деятельности.

Выделяется семь этапов построения эконометрической модели:

1. Постановка проблемы, исходя из целей и задач конкретного исследования. Выбор включаемых в модель переменных, взаимосвязи между которыми подлежат изучению.

2. Предварительный теоретический анализ сущности изучаемых явлений, по результатам которого формируется априорная информация и выдвигаются различные статистические гипотезы.

3. Информационный этап включает выбор объекта исследования, сбор необходимой информации и предварительный анализ ее качества. Определяются значения переменных по объектам наблюдения, моментам и периодам времени.

4. Спецификация модели заключается в обосновании общего вида модели, связывающей зависимые и независимые переменные (входные и выходные, экзогенные и эндогенные).

5. Идентификация модели включает ее статистический анализ и оценку параметров модели по имеющимся статистическим данным.

6. На этапе верификации модели осуществляется анализ ее точности и адекватности моделируемому явлению или процессу, строятся точечные и интервальные оценки (прогнозы) параметров эконометрической модели.

7. Проводится интерпретация полученных результатов и дается оценка возможности использования формализованных выводов в практических целях.

Основные разделы эконометрики: регрессионный анализ, анализ временных рядов, системы одновременных уравнений, статистические методы классификации и снижения размерности. Сферы применения: макроуровень (модели национальной

экономики); мезоуровень (модели региональной экономики, отраслей, секторов); микроуровень (модели поведения потребителей, домашних хозяйств, фирм, предприятий). Эконометрику можно рассматривать как приложение методов математической статистики к изучению социально-экономических явлений, систем и процессов. Современные коммерческие организации интенсивно внедряют информационные базы, банки (хранилища) данных и знаний, многие из которых содержат средства интеллектуального анализа данных или предполагают возможность их применения. Изучение эконометрики является первым шагом к освоению современных компьютерных технологий интеллектуального анализа данных.

Выделяется три основных класса эконометрических моделей:

- регрессионные модели в виде одного уравнения, когда резуль­тативный признак зависит от одного или множества факторов;

- модели временных рядов;

- системы одновременных уравнений.

Настоящие методические рекомендации предназначены для студентов экономических специальностей заочной формы обучения. Студент, на основании изучения рекомендуемой литературы, выполняет одну контрольную работу, которая представляется на кафедру до начала лабораторно-экзаменационной сессии в соответствии с указанным в таблице 1 вариантом. Контрольная работа, выполненная по другому варианту, не рецензируется.

Для самостоятельного изучения эконометрики рекомендуется следующая литература:

Основная учебная литература

1. Ниворожкина, Л. И. Эконометрика : теория и практика : учеб. пособие / Л.И. Ниворожкина, С.В. Арженовский, Е.П. Кокина. — Москва : РИОР : ИНФРА-М, 2018. — 207 с. — (Высшее образование). — DOI: <https://doi.org/10.12737/1698-5>. - ISBN 978-5-369-01698-5. - Текст : электронный. - URL:

<https://znanium.com/catalog/product/907587>

2. Кремер, Н. Ш. Эконометрика : учебник для студентов вузов / Н. Ш. Кремер, Б. А. Путко ; под редакцией Н. Ш. Кремер. — 3-е изд. — Москва : ЮНИТИ-ДАНА, 2017. — 328 с. — ISBN 978-5-238-01720-4. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. — URL:

<http://www.iprbookshop.ru/71071.html>

3. Крянев, А. В. Эконометрика (продвинутый уровень): Конспект лекций / Крянев А.В. - Москва :КУРС, НИЦ ИНФРА-М, 2017. - 62 с.: ISBN 978-5-906818-62-1. - Текст : электронный. - URL:

<https://znanium.com/catalog/product/767248>

Дополнительная учебная литература

1. Айвазян, С. А. Методы эконометрики : учебник / С. А. Айвазян ; Московская школа экономики МГУ им. М.В. Ломоносова (МШЭ). — Москва: Магистр : ИНФРА-М, 2020. — 512 с. - ISBN 978-5-9776-0153-5. - Текст : электронный. - URL:

<https://znanium.com/catalog/product/1043084>

2. Бабешко, Л. О. Эконометрика и эконометрическое моделирование : учебник / Л.О. Бабешко, М.Г. Бич, И.В. Орлова. - Москва: Вузовский учебник : ИНФРА-М, 2019. - 385 с. : ил. — (Высшее образование: Бакалавриат). - ISBN 978-5-9558-0576-4. - Текст : электронный. - URL:

<https://znanium.com/catalog/product/1029152>

3. Ершова, Н. А. Современная эконометрика : учебное пособие / Н. А. Ершова, С. Н. Павлов. — Москва : Российский государственный университет правосудия, 2018. — 52 с. — ISBN 978-5-93916-650-8. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. — URL: <http://www.iprbookshop.ru/78311.html>

4. Соколов, Г. А. Эконометрика: теоретические основы : учеб. пособие / Г.А. Соколов. — Москва : ИНФРА-М, 2018. — 216 с. + Доп. материалы [Электронный ресурс; Режим доступа: <https://new.znanium.com>]. — (Высшее образование: Бакалавриат). - ISBN 978-5-16-010851-3. - Текст : электронный. - URL: <https://znanium.com/catalog/product/944383>

5. Эконометрика в среде GRETЛ : учебное пособие для студентов, обучающихся по направлениям 38.03.01 Экономика, 38.03.02 Менеджмент, 38.03.05 Бизнес-информатика, 38.04.01 Экономика / В. А. Балаш, О. С. Балаш, Т. И. Солодкая, Е. В. Чистопольская. — Саратов : Издательство Саратовского университета, 2019. — 96 с. — ISBN 978-5-292-04617-2. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. — URL: <http://www.iprbookshop.ru/99048.html>

Перечень электронно-библиотечных систем:

№	Наименование	Тематика	Ссылка
1	<i>Znanium.com</i>	Универсальная	https://znanium.com/
2	<i>IPRbook</i>	Универсальная	http://www.iprbookshop.ru/
3	Образовательный портал КубГАУ	Универсальная	https://edu.kubsau.ru/

Перечень Интернет-сайтов:

– Московская Межбанковская валютная биржа: [Электронный ресурс]. – Режим доступа: www.micex.ru, свободный.

- Загл. с экрана;
 - Федеральная служба государственной статистики[Электронный ресурс]. – Режим доступа: www.gks.ru, свободный. – Загл. с экрана.
 - Информационный портал Всемирного банка[Электронный ресурс]. – Режим доступа: <http://data.worldbank.org>, свободный. – Загл. с экрана;
 - Универсальная база данных зарубежных полнотекстовых научных журналов по всем областям знаний: [Электронный ресурс]. – Режим доступа: <http://search.epnet.com/>, свободный. – Загл. с экрана;
 - Сайт эконометрического журнала «Квантиль»:[Электронный ресурс]. – Режим доступа:<http://quantile.ru/>, свободный. – Загл. с экрана;
 - Банк России (ЦБ):[Электронный ресурс]. – Режим доступа: www.cbr.ru, свободный. – Загл. с экрана;
 - Минфин России: Документы МСФО: [Электронный ресурс]. – Режим доступа: http://www.minfin.ru/ru/accounting/mej_standart_fo/docs, свободный. – Загл. с экрана.

Методические указания для обучающихся по освоению дисциплины

1.Эконометрика (продвинутый уровень): метод. указания для контактной и самостоятельной работы обучающихся по направлению подготовки 38.04.01 Экономика / И. А. Кацко, А. Е. Сенникова, Н. Н. Яроменко – Краснодар: Краснодарский Ц+НТИ – филиал ФГБУ «РЭА» Минэнерго России, 2020. – 22 с. – Режим доступа:
https://edu.kubsau.ru/file.php/120/ENkonometrika_prodvinutyi_uroven_Magistry_ENkonomika_2020_g._527192_v1_.PDF

2.Эконометрика. Практикум: учебно-практическое пособие / Коллектив авторов под ред. Кацко И.А. – М.: КНОРУС,

2019. – 216 с. – 80 экземпляров.

3. Эконометрика. Практикум [Электронный ресурс]: учебно-практическое пособие / И.А. Кацко, под ред. и др. – Электрон. текстовые данные. – Москва: КноРус, 2019. – 216 с. 978-5-406-06368-2. – Режим доступа:

<https://www.book.ru/book/931003>

Контрольная работа содержит четыре задания по основным темам. Первое и четвертое задания призваны позволить студенту вспомнить основные идеи эконометрики на уровне бакалавриата. Второе и третье задание соответствуют уровню магистратуры. Задания выполняются в отдельной тетради. По каждому заданию записывается условие, проводятся подробные расчеты с необходимыми пояснениями, формулируются выводы по полученным результатам. В конце контрольной работы приводится список использованной литературы, ставится дата и подпись студента. Для облегчения выполнения контрольной работы по каждому заданию изложены необходимые краткие методические указания, приводится решение типовых задач, формулируются краткие выводы.

Все расчеты рекомендуется выполнять на компьютере с использованием табличного процессора *Microsoft Excel*. Можно использовать также эконометрические пакеты программ: *Statistika*, *Stata*, *IBM SPSS*, *Gretl* и другие профессиональные пакеты.

Студентом контрольная работа выполняется по одному варианту заданий в соответствии с первой и второй буквой фамилии. Варианты и номера решаемых задач указаны в таблице 1.

Таблица 1 – Варианты и номера задач для выполнения контрольной работы (задания 1 и 4)

Первая буква фамилии студента	Вторая буква фамилии студента			
	А, Б, В, Г, Д, Е, Ё,	Ж, З, И, Й, К, Л, М	Н, О, П, Р, С, Т, У	Остальные буквы
А, Б	1, 86	12, 53	23, 64	34, 75
В, Г	2, 43	13, 54	24, 65	35, 76
Д, Е, Ё	3, 44	14, 55	25, 66	36, 77
Ж, З	4, 45	15, 56	26, 67	37, 78
И, Й, К	5, 46	16, 57	27, 68	38, 79
Л, М	6, 47	17, 58	28, 69	39, 80
Н, О	7, 48	18, 59	29, 70	40, 81
П, Р	8, 49	19, 60	30, 71	41, 82
С, Т	9, 50	20, 61	31, 72	41, 83
У, Ф, Х, Ц	10, 51	21, 62	32, 73	42, 84
Остальные буквы	11, 52	22, 63	33, 74	42, 85

1 Методические рекомендации к выполнению контрольной работы

1 Множественный корреляционно-регрессионный анализ

В экономических исследованиях результативный признак Y формируется, как правило, под влиянием не одного, а нескольких факторных признаков X_1, X_2, \dots, X_p . Уравнение множественной регрессии в таком случае имеет вид:

$$y = f(x_1, x_2, \dots, x_p, e).$$

В зависимости от вида функции используются как линейные, так и нелинейные модели. Линейная модель множественной регрессии с несколькими переменными имеет вид:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + e, \quad (2.1)$$

где $\beta_0, \beta_1, \beta_2, \dots, \beta_p$ – параметры модели,
 e – случайные отклонения или остаток,
 p – количество переменных.

Наиболее распространенным методом оценивания параметров линейных эконометрических моделей является метод наименьших квадратов. Его идея сводится к выбору таких значений оценок b_0, b_1, \dots, b_p структурных параметров $\beta_0, \beta_1, \beta_2, \dots, \beta_p$, при которых сумма квадратов отклонений наблюдаемых значений зависимой переменной (y_i) от ее теоретических значений (\hat{y}_i), рассчитанных по уравнению регрессии в натуральном масштабе $\hat{y}_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + \dots + b_p x_{ip}$, оказывается наименьшей. Это условие записывается в виде:

$$S = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n \varepsilon_i^2 \rightarrow \min. \quad (2.2)$$

Чтобы найти экстремум функции, необходимо определить частные производные по параметрам и приравнять их к нулю:

$$\frac{\partial S}{\partial b_0} = 0, \quad \frac{\partial S}{\partial b_1} = 0, \quad \dots, \quad \frac{\partial S}{\partial b_p} = 0. \quad (2.3)$$

Параметры линейного уравнения множественной регрессии находятся путем составления и решения следующей системы нормальных уравнений:

$$\begin{cases} \Sigma y = nb_o + b_1 \Sigma x_1 + b_2 \Sigma x_2 + \dots + b_p \Sigma x_p, \\ \Sigma yx_1 = b_o \Sigma x_1 + b_1 \Sigma x_1^2 + b_2 \Sigma x_1 x_2 + \dots + b_p \Sigma x_1 x_p, \\ \dots\dots\dots \dots\dots\dots \dots\dots\dots \dots\dots\dots \dots\dots\dots \dots\dots\dots \dots\dots\dots \dots\dots\dots \dots\dots\dots \\ \Sigma yx_p = b_o \Sigma x_p + b_1 \Sigma x_1 x_p + b_2 \Sigma x_2 x_p + \dots + b_p \Sigma x_p^2. \end{cases} \quad (2.4)$$

В матричной форме уравнение множественной регрессии имеет вид:

$$Y = XB + \varepsilon, \quad (2.5)$$

где Y – вектор-столбец размерности $(n \times 1)$;

X – матрица наблюдаемых значений факторных признаков размерности $(n \times (p+1))$;

B – вектор-столбец коэффициентов регрессии $((p+1) \times 1)$;

ε – вектор-столбец остаточных значений результирующего признака размерности $(n \times 1)$.

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}; B = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_p \end{bmatrix}; \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}; X = \begin{bmatrix} 1 & x_{11} & x_{21} & \dots & x_{p1} \\ 1 & x_{12} & x_{22} & \dots & x_{p2} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_{1n} & x_{2n} & \dots & x_{pn} \end{bmatrix}. \quad (2.6)$$

Тогда оценки параметров линейной модели b находятся из выражения:

$$b = (X^T X)^{-1} X^T Y, \quad (2.7)$$

где X^T – транспонированная матрица X ;

$(X^T X)^{-1}$ – обратная матрица.

Множественный коэффициент регрессии (b_j) показывает, на сколько единиц изменяется в среднем результирующий признак Y , если j -й факторный признак X_j увеличить на единицу

при условии, что все другие факторы в линейной модели закреплены на постоянном, обычно среднем, уровне.

Уравнение множественной регрессии может быть построено в стандартизованном масштабе, когда единицей измерения признаков принимается их среднее квадратическое отклонение:

$$t_y = \beta_1 t_{x_1} + \beta_2 t_{x_2} + \dots + \beta_p t_{x_p}. \quad (2.8)$$

$$t_y = \frac{y - \bar{y}}{\sigma_y}, \quad t_{x_j} = \frac{x_j - \bar{x}_j}{\sigma_{x_j}},$$

где β_j – стандартизованные коэффициенты регрессии.

Параметры уравнения регрессии определяются методом наименьших квадратов путем составления и решения следующей системы уравнений:

$$\begin{cases} r_{yx_1} = \beta_1 + \beta_2 r_{x_1 x_2} + \dots + \beta_p r_{x_1 x_p} \\ r_{yx_2} = \beta_1 r_{x_1 x_2} + \beta_2 + \dots + \beta_p r_{x_2 x_p} \\ \dots \\ \dots \\ r_{yx_p} = \beta_1 r_{x_1 x_p} + \beta_2 r_{x_2 x_p} + \dots + \beta_p \end{cases} \quad (2.9)$$

Зная стандартизованные коэффициенты можно получить множественные коэффициенты регрессии:

$$b_j = \beta_j \cdot \frac{\sigma_y}{\sigma_{x_j}}, \quad b_0 = \bar{y} - b_1 \bar{x}_1 - b_2 \bar{x}_2 - \dots - b_p \bar{x}_p. \quad (2.10)$$

По абсолютной величине -коэффициентов судят об относительной силе влияния факторов на изменение результативного признака. Для характеристики силы влияния факторов на результативный признак используется также коэффициент

эластичности, который представляет отношение прироста результативного признака Y к приросту факторного признака X_j :

$$\varepsilon_{x_j} = \frac{dy}{\hat{y}} : \frac{dx_j}{x_j} = \frac{dy}{dx_j} \cdot \frac{x_j}{\hat{y}}. \quad (2.11)$$

По линейной модели множественной регрессии коэффициент эластичности определяется по формуле:

$$\varepsilon_{x_j} = b_j \frac{x_j}{b_0 + b_1 x_{i1} + b_2 x_{i2} + \dots + b_p x_{ip}}. \quad (2.12)$$

Если в этой формуле значения факторов взять на среднем уровне, то будет получен средний коэффициент эластичности, который показывает, на сколько процентов в среднем изменится результативный признак, если j -й фактор увеличить на один процент, при условии что все другие факторы закреплены на среднем уровне.

$$\bar{\varepsilon}_{x_j} = b_j \frac{\bar{x}_j}{\bar{y}}. \quad (2.13)$$

Для оценки тесноты связи между признаками применяются парные, частные и множественные коэффициенты (индексы) корреляции и детерминации.

Множественный коэффициент (индекс) корреляции ($R_{yx_1x_2\dots x_p}$) характеризует совместное влияние всех факторов, включенных в уравнение регрессии. Он рассчитывается по следующим формулам:

$$R_{yx_1x_2\dots x_p} = \sqrt{1 - \frac{\sigma_{ост.}^2}{\sigma_y^2}} = \sqrt{\frac{\sigma_{рег.}^2}{\sigma_y^2}} = \sqrt{1 - \frac{\sum_i^n (y - \hat{y}_{x_1x_2\dots x_p})^2}{\sum_i^n (y - \bar{y})^2}} = \sqrt{1 - \frac{SS_{ост.}}{SS_{общ.}}}, \quad (2.14)$$

где σ_y^2 – общая дисперсия результативного признака,
 $\sigma_{рег.}^2$ – дисперсия, объяснимая регрессией,
 $\sigma_{ост.}^2$ – остаточная дисперсия, причем

$$\sigma_y^2 = \sigma_{рег.}^2 + \sigma_{ост.}^2; \quad \sigma_y^2 = \frac{\sum (y - \bar{y})^2}{n}; \quad \sigma_{рег.}^2 = \frac{\sum (\hat{y} - \bar{y})^2}{n}; \quad \sigma_{ост.}^2 = \frac{\sum (y - \hat{y})^2}{n}.$$

$$SS_{общ.} = SS_{факт.} + SS_{ост.};$$

$$\Sigma(y - \bar{y})^2 = \Sigma(\hat{y} - \bar{y})^2 + \Sigma(y - \hat{y})^2, \quad (2.15)$$

где $SS_{общ.}$ – общая сумма квадратов отклонений
результативного признака;

$SS_{факт.}$ – факторная сумма квадратов отклонений;

$SS_{ост.}$ – остаточная сумма квадратов отклонений.

Квадрат множественного коэффициента (индекса) корреляции называется множественным коэффициентом (индексом) детерминации. Он показывает, какая часть вариации результативного признака объясняется влиянием факторов, включенных в уравнение регрессии. Если используется линейное уравнение множественной регрессии в стандартизованном масштабе, то множественный коэффициент детерминации рассчитывается по формуле:

$$R_{yx_1x_2...x_p}^2 = \beta_1 r_{yx_1} + \beta_2 r_{yx_2} + \dots + \beta_p r_{yx_p} = \Sigma \beta_j r_{x_j}. \quad (2.16)$$

Частные коэффициенты корреляции, характеризующие тесноту связи между фактором x_j и результативным признаком, при исключении влияния других факторов, включенных в модель, определяется по формулам:

$$r_{yx_j \cdot x_1 x_2 \dots x_{j-1} x_{j+1} \dots x_p} = \sqrt{1 - \frac{1 - R_{yx_1 x_2 \dots x_j \dots x_p}^2}{1 - R_{y x_1 x_2 \dots x_{j-1} x_{j+1} \dots x_p}^2}}, \quad (2.17)$$

$$r_{yx_j \cdot x_1 x_2 \dots x_{j-1} x_{j+1} \dots x_p} =$$

$$= \frac{r_{yx_j \cdot x_1 x_2 \dots x_{j-1} x_{j+1} \dots x_{p-1}} - r_{yx_p \cdot x_1 x_2 \dots x_{j-1} x_{j+1} \dots x_{p-1}} \cdot r_{x_j x_p \cdot x_1 \dots x_{j-1} x_{j+1} \dots x_{p-1}}}{\sqrt{(1 - r_{y x_p \cdot x_1 x_2 \dots x_{j-1} x_{j+1} \dots x_{p-1}}^2)(1 - r_{x_j x_p \cdot x_1 \dots x_{j-1} x_{j+1} \dots x_{p-1}}^2)}} \quad (2.18)$$

В формуле частные коэффициенты корреляции j -го порядка рассчитываются через частные коэффициенты корреля-

ции $(j-1)$ -го порядка. Значения частных коэффициентов корреляции изменяются от -1 до 1 . Они могут быть использованы при отсеве несущественно влияющих факторов.

С учетом поправки на число степеней свободы рассчитывается скорректированный коэффициент (индекс) множественной корреляции:

$$R_{ск}^2 = 1 - \frac{\sum(y - \hat{y})^2 : (n-m-1)}{\sum Y - \bar{Y})^2 : (n-1)}, \quad (2.19)$$

$$R_{ск}^2 = 1 - (1 - R^2) \cdot \frac{(n-1)}{(n-m-1)}, \quad (2.20)$$

где m – число параметров уравнения регрессии без учета свободного члена. В линейном уравнении $m = p$.

Оценка значимости множественного уравнения регрессии производится с помощью F -критерия Фишера-Снедекора.

Определяется наблюдаемое значение критерия по следующей формуле:

$$F_n = \frac{SS_{факт.}}{m} : \frac{SS_{ост.}}{n-m-1} = \frac{R^2}{1-R^2} \cdot \frac{n-m-1}{m}. \quad (2.21)$$

При заданном уровне значимости α и числе степеней свободы факторной (k_1) и остаточной дисперсий (k_2) по таблицам находится критическое значение критерия Фишера-Снедекора. Сравнивается наблюдаемое и критическое значения критерия. Если $F_n < F_{кр}$, то нулевая гипотеза о незначимости уравнения регрессии принимается. Если $F_n > F_{кр}$, то нулевая гипотеза отвергается и принимается альтернативная гипотеза о статистической значимости всего уравнения регрессии.

Оценка значимости параметров множественного линейного уравнения регрессии производится с помощью t -критерия Стьюдента. Выдвигается основная гипотеза о равенстве нулю параметров уравнения регрессии ($H_0: \beta_j = 0$), при конкурирующей гипотезе, что параметры уравнения

отличны от нуля ($H_0: \beta_j \neq 0$). Наблюдаемое значение t -критерия для параметра уравнения b_j определяется по формуле:

$$t_{b_j} = \frac{b_j}{s_{b_j}}, \quad s_{b_j} = \sqrt{\frac{SS_{ocm.}}{n-p-1} [(X^T X)^{-1}]_{jj}}, \quad (2.22)$$

где s_{b_j} – стандартная ошибка параметра уравнения регрессии b_j ,

$[(X^T X)^{-1}]_{jj}$ – диагональный элемент матрицы $(X^T X)^{-1}$.

Стандартная ошибка множественного коэффициента регрессии b_j может быть найдена также по формуле:

$$s_{b_j} = \frac{\sigma_y}{\sigma_{x_j}} \sqrt{\frac{1 - R_{y x_1 x_2 \dots x_p}^2}{(1 - R_{x_j x_1 x_2 \dots x_{j-1} x_{j+1} \dots x_p}^2)(n-m-1)}}, \quad (2.23)$$

где σ_y – среднее квадратическое отклонение результативного признака;

σ_{x_j} – среднее квадратическое отклонение факторного признака x_j .

Критическое значение t находится по таблице значений t -критерия Стьюдента при уровне значимости α и числе степеней свободы $k = n - m - 1$. Если $|t_{b_j}| > |t_{kp}|$, то параметр уравнения статистически значим. Если $|t_{b_j}| < |t_{kp}|$, то параметр уравнения статистически не значим и j -ая переменная исключается из уравнения регрессии. Доверительные интервалы для коэффициентов линейного уравнения регрессии находятся по формуле:

$$b_j \pm t_{kp} s_{b_j}. \quad (2.24)$$

Условия, необходимые для получения несмещенных, состоятельных эффективных оценок, представляют собой предпосылки метода наименьших квадратов, соблюдение которых желательно для получения достоверных результатов регрессии:

- 1) остаток является случайной величиной;
- 2) математическое ожидание случайного остатка равно 0:

$$M(\varepsilon_i) = 0, i = 1, 2, \dots, n;$$

3) дисперсия случайных остатков постоянна для любого наблюдения $D(\varepsilon_i) = \sigma^2$, это предположение называется условием гомоскедастичности;

- 4) случайные остатки ε_i и ε_j не коррелированы,

$$\text{cov}(\varepsilon_i, \varepsilon_j) = 0, i \neq j;$$

- 5) остатки распределены по нормальному закону.

Первые 4 условия известны как условия Гаусса-Маркова.

При построении уравнения множественной регрессии обычно используются следующие нелинейные функции:

степенная $y = b_0 \cdot x_1^{b_1} \cdot x_2^{b_2} \dots x_p^{b_p} \cdot \varepsilon$; (2.25)

экспонента $y = e^{b_0 + b_1 x_1 + b_2 x_2 + \dots + b_p x_p + \varepsilon}$; (2.26)

гипербола $y = b_0 + \frac{b_1}{x_1} + \frac{b_2}{x_2} + \dots + \frac{b_p}{x_p} + \varepsilon$; (2.27)

логлинейная $\ln y = b_0 + b_1 \ln x_1 + b_2 \ln x_2 + \dots + b_p \ln x_p + \varepsilon$. (2.28)

Довольно часто применяются и другие виды функций, например:

$$y = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_1^2 + b_4 x_2^2 + b_5 x_1 x_2 + \varepsilon . \quad (2.29)$$

Если уравнение регрессии нелинейное, то оно вначале приводится путем соответствующего преобразования к линейному виду.

Методы подгонки зависимости

Модели параметрической регрессии, требуют выполнения ряда априорных положений, достоверность которых не всегда достаточно убедительна: вид зависимости заранее известен как условное математическое ожидание результативного признака при условии, что факторный признак принял конкретное значение – $M(y/x)$, дисперсия погрешности постоянна (гомоскедастичность) и т. д. Для получения более точных моделей в затрудненных условиях (когда предпосылки параметрического регрессионного анализа не выполняются, прежде всего в технических науках) были предложены:

- *локально-параметрические модели* регрессии, использующие простые модели (линейные, параболические) для под областей изменения регрессора;
- *сплайны* – совокупности гладко склеенных полиномов;
- *методы сглаживания регрессии (ядерное, методом ближайших соседей, сплайнами)*.

Известно, что различные процедуры сглаживания регрессии асимптотически стремятся к ядерным оценкам.

В середине XX века Кенуй (1949) предложил метод, а Тьюки (1958) принимавший активное участие в его (метода) совершенствовании дал ему название «метод складного ножа» или *бутстреп метод*, согласно которому вся совокупность данных разбивается на части и проводится статистический анализ ее частей для достижения несмещённости оценок. В основу положена идея рандомизации (случайной выборки) Р. Фишера, который направлял ее на планирование эксперимента, а Б. Эфрон (1977) [239], А. Г. Ивахненко (1969) [159,

160] на управление выборкой для получения несмещенных оценок. Осуществление рандомизации с использованием компьютера сегодня позволило перейти от теоретических предпосылок к более точным прогнозам для реальных объектов и процессов.

Для получения устойчивых к выбросам оценок параметров регрессионных моделей (и далее моделей машинного обучения) используют идею *бутстреп метода*, позволяющего получить дополнительную информацию из выборки за счет многократной обработки ее различных частей.

Другим направлением, позволяющим учесть недостатки параметрической регрессии (1970-1980) является *квантильная регрессия*.

Итак, в зависимости от целей моделирования рассматриваются различные способы оценки параметров регрессии. Без потери общности рассмотрим эти методы на модели парной функции регрессии:

$$Y = \beta_0 + \beta_1 x + \varepsilon,$$

где β_0, β_1 – параметры модели,
 ε – свободный член или остаток.

Если целью построения эконометрической модели является поиск среднего значению y в зависимости от значения переменной x или иначе, $M(y_i/x_i) = \beta_0 + \beta_1 \cdot x_i$, то параметры модели определяются методом наименьших квадратов.

Если целью исследования является нахождение условной медианы

$$M_e(y_i/x_i) = \beta_0 + \beta_1 x_i, \quad (2.30)$$

то

$$M_e(\hat{\beta}_0, \hat{\beta}_1) = \sum |y_i - \hat{y}_i| \rightarrow \min \quad (2.31)$$

К минусам медианной регрессии относится то, что нет формул для нахождения параметров и нахождения стандартных ошибок, так как нахождение медианы решается как задача численной оптимизации на компьютере. Плюсы медианной

регрессии: другой взгляд на данные (с точки зрения порядковой шкалы), устойчивость к выбросам.

Медианная регрессия – частный случай квантильной регрессии. Вместо поиска *среднего значения* прогнозируемой переменной, квантильная регрессия направлена на поиск *медианы* и любых других квантилей (которые иногда называют *процентилями*).

Одним из важных свойств квантильной регрессии является то, что данная модель устойчива к «выбросам», которые часто встречаются на практике. Поэтому при прогнозе использование модели квантильной регрессии более предпочтительно, чем использование классических моделей.

Непараметрическая регрессия: ядерное сглаживание.

Непараметрическое восстановление регрессии основано на непараметрическом восстановлении плотности распределения вероятностей и использует ту же идею, что и непараметрическое восстановление функции плотности распределения вероятностей.

Обычно рассматриваются следующие функции расстояния $\rho(x_i, x_j)$: евклидово, взвешенное евклидово, хемингово (расстояние городских кварталов), Махаланобиса и др. Например, евклидово расстояние можно определить по формуле

$$\rho_E(x_i, x_j) = \sqrt{\sum_{k=1}^p (x_i^{(k)} - x_j^{(k)})^2}. \quad (2.32)$$

1. Рассмотрим неотрицательную, невозрастающую, гладкую ограниченную функцию K (интеграл от которой равен единице), определенную для неотрицательных значений переменной – *ядро*:

$$w(x_i) = K\left(\frac{\rho(x_i, x_j)}{h}\right), \quad (2.33)$$

где h - ширина (полоса пропускания) окна сглаживания (ширина ядра);

$Q = \sum_{i=1}^n w(x_i) f(\varepsilon_i)$ – функционал качества аппроксимации, например,

$$Q(b) = \sum_{i=1}^n w(x_i) (b - y_i)^2, \quad (2.34)$$

где $y = b$ – простейшее уравнение регрессии.

Приравняв к нулю производную $\frac{dQ}{db} = 0$ получим формулу ядерного сглаживания Надарая-Ватсона (Nadaraya, 1964; Watson, 1964):

$$b_h(x_j) = \frac{\sum_{i=1}^n y_i w(x_j)}{\sum_{i=1}^n w(x_j)} = \frac{\sum_{i=1}^n y_i K\left(\frac{\rho(x_i, x_j)}{h}\right)}{\sum_{i=1}^n K\left(\frac{\rho(x_i, x_j)}{h}\right)}, \quad (2.35)$$

Loess или Lowess (locally weighted regression; W. Cleveland, 1979) – локально взвешенная полиномиальная регрессия (0, 1 или 2 порядка) в каждой точке i по r ближайшим соседям подбирается полином (0-2 порядка) с помощью которого в точке x_i предсказывается значение y_i . Для каждой точки x_i вычисляются расстояния $|x_k - x_i|$, где $k = \overline{1, r}$

($r = qn$, q – параметр полосы пропускания может подбираться итеративно). Веса подбираются по формуле, предложенной Кливлендом:

$$\omega_k(x_i) = W\left(\frac{x_k - x_i}{h_i}\right), \quad (2.36)$$

где h_i – расстояние между x_i и x_r ; $W(x) = (1 - |x|^3)^3$ при $|x| < 1$, иначе 0.

Модель *Loess* для каждого не пропущенного значения аргумента строит прогноз.

Часто для ядерного сглаживания используется гауссовское или квадратическое ядро, соответственно:

$$K_G(r) = \exp\left(-\frac{1}{2h} r^2\right), \quad (2.37)$$

$$K_Q(r) = (1 - r^2)^2, \quad [|r| < 1]. \quad (2.38)$$

На качество построенной модели главным образом влияет выбор ширины окна h . При ($h \rightarrow 0$) функция $b_h(x)$ проходит через все точки выборки, а при слишком большом окне

функция $b_h(x)$ сглаживается и в пределе $h \rightarrow \infty$ вырождается в *const.*

Пример 1. Исследовать влияние среднегодовой стоимости основных фондов на 1 га сельскохозяйственных угодий и численности работников на 100 га сельскохозяйственных угодий на стоимость реализованной продукции с 1 га сельскохозяйственных угодий по сельскохозяйственным организациям Краснодарского края за 2013 год. Исходные данные для выполнения задания представлены в приложениях 1 и 3.

Результативным признаком (Y) является стоимость реализованной продукции на 1 га сельхозугодий, тыс. руб.

Факторные признаки: X_1 – среднегодовая стоимость основных фондов на 1 га сельскохозяйственных угодий, характеризующая оснащенность сельскохозяйственных организаций основными фондами, тыс. руб.; X_2 – среднегодовая численность работников на 100 га сельхозугодий, чел., выражающая обеспеченность сельскохозяйственных организаций рабочей силой.

Требуется определить:

- 1) параметры множественного уравнения регрессии в натуральной и стандартизованной форме;
- 2) средние коэффициенты эластичности для каждого фактора;
- 3) коэффициенты частной и множественной корреляции;
- 4) общий и частные критерии F -Фишера.

Решение

Связь между результативным признаком Y и факторами X_1 и X_2 выражается множественным линейным уравнением регрессии, которое имеет вид:

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2. \quad (2.39)$$

Рассмотрим применение пакета анализа данных в *Excel MS Office* для решения задачи. Исходные данные введем на листе *MS Excel* в виде, представленном таблице 2.

Таблица 2 – Исходные данные для регрессионного анализа в *MS Excel*

№ п/п	Выручка на 1 га сельскохозяйственных угодий, тыс. руб. (Y)	Среднегодовая стоимость основных фондов на 1 га сельскохозяйственных угодий, тыс. руб. (X_1)	Среднегодовая численность работников на 100 га сельскохозяйственных угодий, чел. (X_2)
1	37,8	44,9	3,02
2	60,7	69,3	6,87
3	36,8	22,7	2,55
4	41,4	21,6	4,82
5	40,0	26,9	4,55
6	42,3	29,5	1,92
7	42,6	61,2	3,29
8	45,2	59,9	6,81
9	49,6	60,1	6,77
10	57,2	71,7	5,84
11	41,7	48,5	3,16
12	47,5	50,3	5,93
13	30,4	21,4	1,54
14	38,0	40,1	2,83
15	54,5	55,0	5,97
16	44,4	59,2	4,99
17	38,2	40,1	1,58
18	38,5	45,3	5,02
19	35,9	29,1	4,04
20	35,7	21,3	2,93

Для проведения анализа предварительно установим пакет анализа, выполнив последовательно действия: кнопка *Office – Параметры Excel – Надстройки – Пакет анализа – Перейти* (выделим в окне доступных надстроек *Пакет анализа*), после этого во вкладке *Данные* ленты появится инструмент *Пакет анализа*.

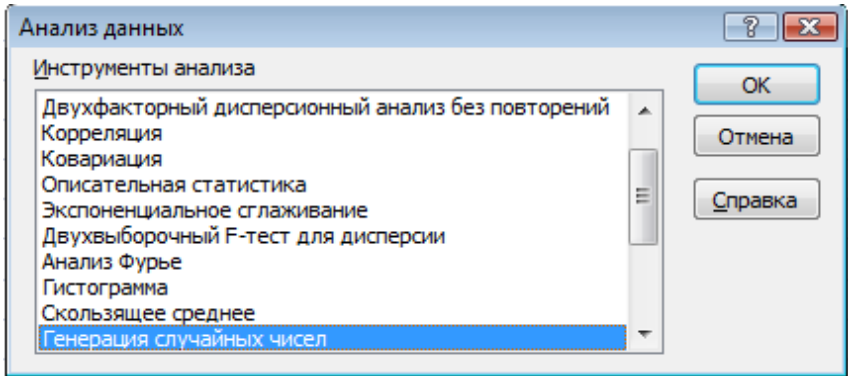


Рисунок 1 – Анализ данных

Выберем в *Пакете анализа* инструмент *Описательная статистика* и заполним параметры диалогового окна (рисунок 1). В результате будут рассчитаны обобщающие характеристики по каждому признаку (таблица 3).

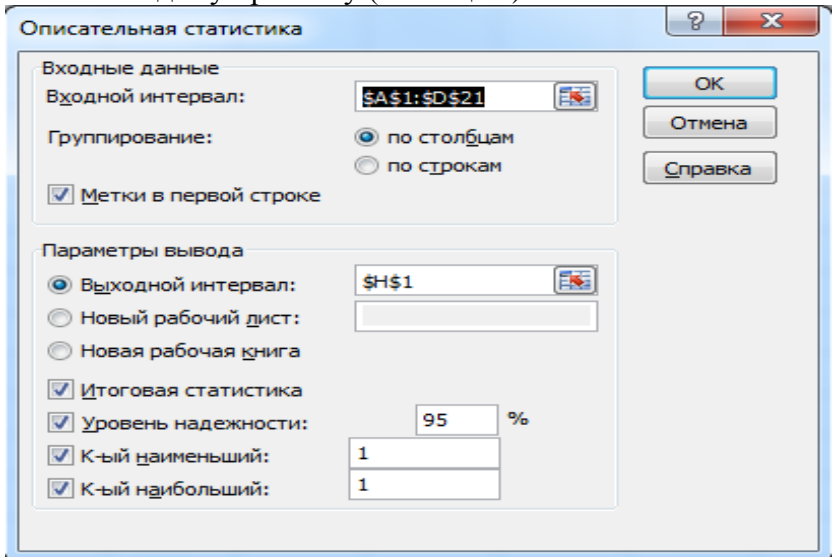


Рисунок 2 – Диалоговое окно описательной статистики

Таблица 3 – Обобщающие характеристики исследуемых признаков по совокупности сельскохозяйственных организаций

Показатель	У	X ₁	X ₂	Принятые обозначения
Среднее значение	42,92 1	43,908	4,222	$\bar{X} = \sum x_i n_i / n$
Стандартная ошибка	1,722	3,760	0,397	$s_{\bar{x}} = s / \sqrt{n}$
Медиана	41,54 8	45,108	4,295	M _e
Мода	н/д	н/д	н/д	M _o
Стандартное отклонение	7,700	16,816	1,774	s
Дисперсия выборки	59,28 9	282,76 8	3,148	$s^2 = \sum (x_i - \bar{X})^2 n_i / (n - 1)$
Экссесс	0,471	- 1,321	1,289	$Ex = \sum ((x_i - \bar{X}) / S)^4 n_i / n - 3$
Асимметричность	0,909	0,022	0,053	$Sk = \sum ((x_i - \bar{X}) / S)^3 n_i / n$
Интервал	30,3	50,426	5,33	W=X _{max} - X _{min}
Минимум	30,4	21,3	1,54	X _{min}
Максимум	60,7	71,7	6,87	X _{max}
Сумма	858,4 2	878,2	84,43	ΣX _i
Счет	20	20	20	n=Σn _i
Наибольший (1)	60,7	71,7	6,87	-
Наименьший (1)	30,4	21,3	1,54	-
Уровень надежности (95,0%)	3,61	7,87	0,83	$\Delta = t_{\alpha; n-1} s_{\bar{x}}$

Данные таблицы 3 показывают, что по совокупности предприятий средняя выручка от реализации продукции на 1 га сельскохозяйственных угодий составила 42,9 тыс. руб. и в среднем между предприятиями выручка на 1 га сельскохозяйственных угодий колеблется в границах 42,9 ± 7,7 тыс. руб., т. е. от 35,2 до 50,6 тыс. руб. Коэффициент вариации составил 17,9 %, что свидетельствует о больших различиях в выручке от реализации продукции на 1га сельскохозяйственных угодий

между предприятиями. По значению медианы видно, что половина предприятий имеет размер выручки до 41,5 тыс. руб./га, а половина более. Распределение предприятий по данному признаку является несимметричными ($Ka = 0,909$) и островершинным ($\Theta = -0,471$). Наименьшее значение выручки на 1 га сельскохозяйственных угодий составило 30,4, а наибольшее – 60,7 тыс. руб.

Аналогичные выводы можно сделать и по факторным признакам X_1 и X_2 .

Для нахождения парных коэффициентов корреляции применим инструмент пакета анализа *Корреляция*, для этого заполним параметры диалогового окна как на рисунке 3.

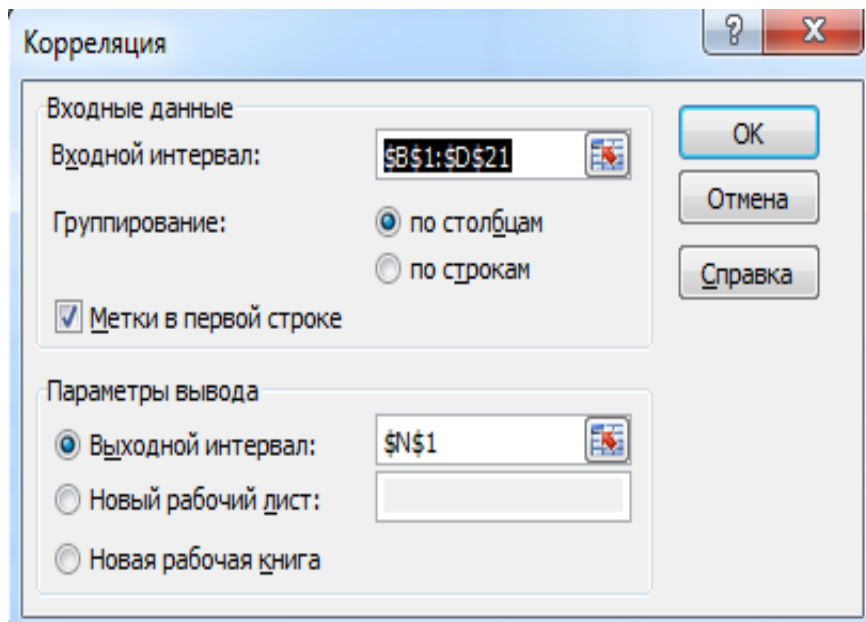


Рисунок 3 – Диалоговое окно «Корреляция»

В результате будет получена матрица парных коэффициентов корреляции между всеми изучаемыми переменными (таблица 4).

Таблица 4 – Парные коэффициенты корреляции между признаками

	Y	X ₁	X ₂
Y	1	0,7963	0,7692
X ₁	0,7963	1	0,6586
X ₂	0,7692	0,6586	1

Значит: $r_{yx_1} = 0,7963$; $r_{yx_2} = 0,7692$; $r_{x_1x_2} = 0,6586$. Парные коэффициенты корреляции показывают, что связь между выручкой от реализации, фондооснащенностью и трудообеспеченностью довольно тесная, а между факторными признаками X₁ и X₂ – средняя.

Линейное уравнение множественной регрессии в натуральной форме имеет вид:

$$y = b_0 + b_1 x_1 + b_2 x_2.$$

Найдем параметры этого уравнения, используя инструмент Пакета анализа – Регрессия. Заполним параметры диалогового окна (рисунок 4).

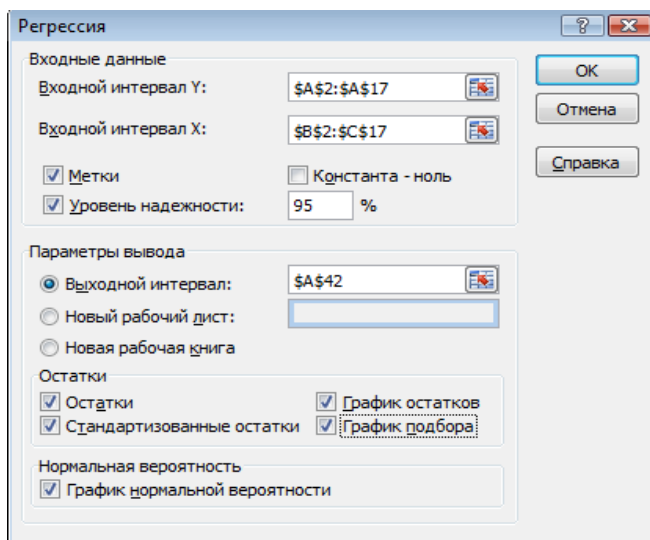


Рисунок 4 – Диалоговое окно «Регрессия»

ВЫВОД ИТОГОВ

Регрессионная статистика						
Множественный R		0,8602				
R-квадрат		0,7399				
Нормированный R-квадрат		0,7093				
Стандартная ошибка		4,1536				
Наблюдения		20				
Дисперсионный анализ						
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Значимость F</i>	
Регрессия	2	834,4994099	417,249705	24,18487	1,0669E-05	
Остаток	17	293,2925901	17,2525053			
Итого	19	1127,792				
<i>Коэффициенты</i>		<i>Стандартная ошибка</i>	<i>t-статистика</i>	<i>P-Значение</i>	<i>Нижние 95%</i>	<i>Верхние 95%</i>
Y-пересечение	24,699	2,78480	8,8693	8,718E-08	18,8239	30,5747
X1	0,2345	0,0753	3,1125	0,0063	0,0756	0,3934
X2	1,8770	0,71356	2,6304	0,0175	0,3715	3,3825

Рисунок 5 – Вывод итогов регрессионного анализа

Получим линейное уравнение множественной регрессии:

$$y = 24,699 + 0,234x_1 + 1,877x_2.$$

Коэффициенты множественной регрессии показывают, что при увеличении среднегодовой стоимости основных фондов на 1 га сельскохозяйственных угодий на 1 тыс. руб. выручка от реализации продукции на 1 га сельскохозяйственных угодий в среднем увеличивается на 234 руб. (при исключении

влияния второго фактора X_2), а при росте численности работников на 100 га сельскохозяйственных угодий на одного работника она в среднем возрастает на 1877 руб./ га.

В стандартизированной форме уравнение регрессии имеет вид:

$$t_y = \beta_1 \cdot t_{x_1} + \beta_2 t_{x_2}, \quad t_y = \frac{y - \bar{y}}{\sigma_y}; \quad t_{x_1} = \frac{x_1 - \bar{x}_1}{\sigma_{x_1}}; \quad t_{x_2} = \frac{x_2 - \bar{x}_1}{\sigma_{x_1}}.$$

Найдем β -коэффициенты, используя их связь с коэффициентами b_j уравнения регрессии в нормальной форме:

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

$$\beta_1 = 0,234 \cdot \frac{16,816}{7,7} = 0,512,$$

$$\beta_2 = 1,877 \cdot \frac{1,774}{7,7} = 0,432.$$

B -коэффициенты, можно также найти с помощью парных коэффициентов корреляции по формулам:

$$\beta_1 = \frac{r_{yx_1} - r_{yx_2} \cdot r_{x_1x_2}}{1 - r_{x_1x_2}^2} = \frac{0,7963 - 0,7692 \cdot 0,6568}{1 - 0,6568^2} = 0,512,$$

$$\beta_2 = \frac{r_{yx_2} - r_{yx_1} \cdot r_{x_1x_2}}{1 - r_{x_1x_2}^2} = \frac{0,7692 - 0,7963 \cdot 0,6568}{1 - 0,6568^2} = 0,432.$$

Линейное уравнение множественной регрессии в стандартизованном масштабе имеет вид:

$$t_y = 0,512 t_{x_1} + 0,432 t_{x_2}.$$

По абсолютной величине β -коэффициентов можно сделать вывод об относительной силе влияния факторов на изменение резульативного признака. Видно, что на выручку от реализации продукции большее влияние оказывает обеспеченность основными фондами и меньшее – обеспеченность рабочей силой.

2. Средние коэффициенты эластичности находятся по формуле:

$$\text{Э}x_1 = b_j \cdot \frac{\bar{x}_j}{\bar{y}},$$

$$\text{Э}x_1 = b_1 \cdot \frac{\bar{x}_1}{\bar{y}} = 0,234 \cdot \frac{43,908}{42,921} = 0,239,$$

$$\text{Э}x_2 = b_2 \cdot \frac{\bar{x}_2}{\bar{y}} = 1,877 \cdot \frac{4,222}{42,921} = 0,185.$$

Значит, при увеличении обеспеченность основными фондами на 1 % выручка от реализации продукции на 1 га сельскохозяйственных угодий увеличивается в среднем на 0,239 %, исключив влияние второго фактора. Если увеличить численность работников на 100 га сельхозугодий на 1 %, то выручка от реализации в среднем возрастет на 0,185 %, исключив влияние фондооснащенности.

3. Коэффициенты частной корреляции определяются через парные коэффициенты корреляции по формулам:

$$r_{yx_1-x_2} = \frac{r_{yx_1} - r_{yx_2}r_{x_1x_2}}{\sqrt{(1-r_{yx_2}^2)(1-r_{x_1x_2}^2)}} = \frac{0,7963 - 0,7692 \cdot 0,6568}{\sqrt{(1-0,7692^2)(1-0,6568^2)}} = 0,604;$$

$$r_{yx_2-x_1} = \frac{r_{yx_2} - r_{yx_1}r_{x_1x_2}}{\sqrt{(1-r_{yx_1}^2)(1-r_{x_1x_2}^2)}} = \frac{0,7692 - 0,7963 \cdot 0,6568}{\sqrt{(1-0,7963^2)(1-0,6568^2)}} = 0,540;$$

$$r_{x_1x_2-y} = \frac{r_{x_1x_2} - r_{yx_1}r_{yx_2}}{\sqrt{(1-r_{yx_1}^2)(1-r_{yx_2}^2)}} = \frac{0,6568 - 0,7963 \cdot 0,7692}{\sqrt{(1-0,7963^2)(1-0,7692^2)}} = 0,115.$$

Коэффициенты частной корреляции характеризуют тесноту связи между двумя переменными, исключив влияние третьей переменной. Значит, связь между фондооснащенностью и выручкой от реализации прямая и довольно тесная, между трудообеспеченностью и выручкой от реализации прямая и средняя. А связь между факторами x_1 и x_2 слабая и также прямая.

Коэффициент множественной корреляции находится по формуле:

$$R_{yx_1x_2} = \sqrt{\beta_1 r_{yx_1} + \beta_2 r_{yx_2}} = \sqrt{0,512 \cdot 0,7963 + 0,432 \cdot 0,7692} = \sqrt{0,408 + 0,332} = \sqrt{0,74} = 0,86.$$

Величина коэффициента множественной корреляции показывает, что связь между выручкой и обоими факторами очень тесная, причем 86 % вариации выручки от реализации продукции объясняется влиянием фондо- и трудообеспеченности, из которой на долю фондообеспеченности приходится 40,8 % вариации, а трудообеспеченности – 33,2 %.

4. Оценим значимость уравнения регрессии и множественного коэффициента детерминации R^2 с помощью критерия F -Фишера. Фактически рассматривается нулевая гипотеза $H_0: R^2=0, (b_1=b_2=0)$ и альтернативная гипотеза $H_1: R^2 \neq 0, (b_1 \neq 0, b_2 \neq 0)$.

Наблюдаемое значение критерия находится по формуле:

$$F_n = \frac{R_{yx_1x_2}^2}{1 - R_{yx_1x_2}^2} : \frac{m}{n - m - 1},$$

где m – число факторов в линейном уравнении регрессии;
 n – число единиц наблюдения.

$$F_n = \frac{0,737}{1-0,737} : \frac{2}{20-2-1} = 24,19.$$

При уровне значимости $\alpha = 0,05$ и числе степеней свободы $k_1=m=2$, $k_2=n-m-l=20-2-1=17$, по таблице значений критерия F -Фишера критическое значения составляет 3,59, т. е. $F_{кр}=3,59$. Сравниваем F_n с $F_{кр}$. Так как $F_n > F_{кр}$, то нулевую гипотезу о незначимости величины R^2 отклоняем, т.е. уравнение множественной регрессии и R^2 статистически значимы.

В уравнении множественной регрессии не все факторы могут оказывать статистически существенное влияние на изменение результативного признака. Оценка значимости факторов в уравнении регрессии может быть дана с помощью частного F -критерия или критерия t -Стьюдента.

$$F_{nx_1} = \frac{R_{yx_1x_2}^2 - r_{yx_2}^2}{1 - R_{yx_1x_2}^2} \cdot \frac{n-m-1}{1} = \frac{0,74-0,5917}{1-0,74} \cdot \frac{20-2-1}{1} = 9,70.$$

При $\alpha = 0,05$, $k_1 = 1$, $k_2 = 17$, $F_{кр} = 4,45$. Так как $F_{nx_1} > F_{кр}$, то в уравнение регрессии целесообразно включение фактора X_1 после X_2 . Фактор X_1 оказывает статистически значимое влияние на Y .

$$F_{nx_2} = \frac{R_{yx_1x_2}^2 - r_{yx_1}^2}{1 - R_{yx_1x_2}^2} \cdot \frac{n-m-1}{1} = \frac{0,737-0,6341}{1-0,74} \cdot \frac{20-2-1}{1} = 6,92.$$

В этом случае также наблюдаемое значение критерия Фишера больше критического, это свидетельствует о статистической значимости влияния фактора X_2 и целесообразности включения его в уравнение множественной регрессии. В данной задаче на выручку от реализации продукции статистиче-

ски значимое влияние оказывают оба фактора. Небольшие расхождения в результатах расчетов в компьютерном и ручном вариантах расчетов обусловлено округлением расчетных значений.

Пример 2. Имеется выборка из базы данных по стоимости жилья г. Краснодара (2016 г.) – 93 наблюдения общей площади и цены на вторичном рынке жилья. Переменные X – общая площадь (m^2), Y – цена (тыс. руб.). С использованием пакета *Gretl* построить модели медианной, квантильной регрессии и непараметрических моделей ядерного сглаживания: *loess*, Надарая-Ватсона.

N п/п	X	Y	N п/п	X	Y	N п/п	X	Y
1	39	1350	32	65	4000	63	37	1650
2	41	1350	33	75	4500	64	31	1650
3	36	2000	34	83	4500	65	80	4000
4	45	1800	35	106	4660	66	36	1650
5	40	2150	36	88	4700	67	41	1660
6	48	2900	37	70	4700	68	38,4	1690
7	64	2940	38	85	6500	69	31	1700
8	45	3000	39	85	6900	70	40	1700
9	64	3100	40	36	1820	71	33	1730
10	56	2400	41	38	1850	72	32	1250
11	44	2450	42	37	1850	73	68	2156
12	45	2500	43	110	4950	74	58	1899
13	45	2600	44	85	4000	75	70	3180
14	68	2675	45	83	6000	76	38	1470
15	46,6	1900	46	37	1900	77	37	1309
16	44	2050	47	40	1900	78	62	2700
17	32	1850	48	38	1950	79	88	2870
18	87	7400	49	38	2000	80	42	1830

N п/п	X	Y	N п/п	X	Y	N п/п	X	Y
19	90	5990	50	43	2000	81	42,6	1850
20	97	6800	51	41	2100	82	43	1850
21	42	1680	52	39	1600	83	40	2050
22	41	1700	53	44	1650	84	46	2200
23	42	1730	54	43,4	1650	85	56	3400
24	46	1770	55	44	1850	86	142	8520
25	44	2800	56	73	3950	87	62	2237
26	56	2800	57	64	4350	88	54	2270
27	60	2800	58	74	4850	89	64	2350
28	60	3000	59	52	3100	90	54	2400
29	67	3100	60	53	3200	91	54	1298
30	60	3100	61	70	3450	92	50	1774
31	110	4000	62	67	3450	93	70	2100

Решение. Модели квантильной регрессии представлены на рисунках 6, 7.

Модель 1: Квантильная оценка, использованы наблюдения 1-93

Зависимая переменная: Y

tau = 0,5

Асимптотические стандартные ошибки считаются независимыми и одинаково распределенными

	Коэффициент	Ст. ошибка	t-статистика	P-значение
const	-508,333	217,543	-2,337	0,0216 **
X	58,3333	3,60894	16,16	1,73e-028 ***

Медиана зав. перемен	2237,000	Ст. откл. зав. перемен	1507,376
Сумма модулей ошибок	50078,67	Сумма кв. остатков	55745292
Лог. правдоподобие	-742,3165	Крит. Акаике	1488,633
Крит. Шварца	1493,698	Крит. Хеннана-Куинна	1490,678

Тест на нормальное распределение ошибок -

Нулевая гипотеза: ошибки распределены по нормальному закону

Тестовая статистика: Хи-квадрат(2) = 13,2635

p-значение = 0,00131786

Рисунок 6 – Медианная ($\tau = 0,5$) регрессия

Первая из них – медианная статистически значима и отражает значение (медианное) стоимости одного квадратного метра жилья как 58, 33 тыс. руб.

Модель 2: Квантильная оценка, использованы наблюдения 1-93

Зависимая переменная: Y

Асимптотические стандартные ошибки считаются независимыми и одинаково распределенными

	tau	Коэффициент	Ст. ошибка	t-статистика
const	0,100	-40,2083	167,396	-0,240199
	0,200	-126,000	135,506	-0,929845
	0,300	-231,159	146,754	-1,57515
	0,400	-319,813	156,466	-2,04398
	0,500	-508,333	217,543	-2,33670
	0,600	-518,605	231,932	-2,23602
	0,700	-653,846	239,060	-2,73507
	0,800	-1026,47	106,521	-9,63631
	0,900	-1063,83	331,872	-3,20554
X	0,100	36,7292	2,77702	13,2261
	0,200	43,0000	2,24799	19,1282
	0,300	47,1014	2,43458	19,3469
	0,400	50,9346	2,59569	19,6227
	0,500	58,3333	3,60894	16,1636
	0,600	60,4651	3,84764	15,7149
	0,700	68,7179	3,96590	17,3272
	0,800	79,4118	1,76714	44,9381
	0,900	85,1064	5,50560	15,4581

Медиана зав. перемен 2237,000 Ст. откл. зав. перемен 1507,376

Рисунок 7 – Квантильная регрессия

Коэффициенты второй регрессии также статистически значимы и показывают значение оценки стоимости одного квадратного метра жилья по децилям (рис. 8).

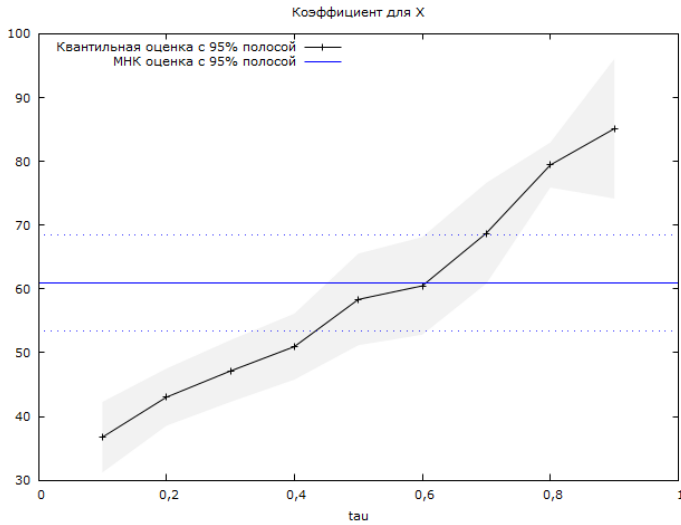


Рисунок 8 – Оценки коэффициентов регрессии по децилям

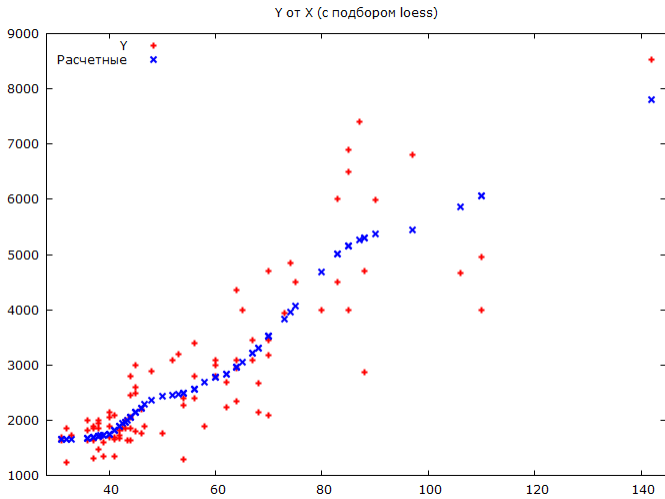


Рисунок 9 – Визуализация прогноза согласно модели *Loess*-регрессии

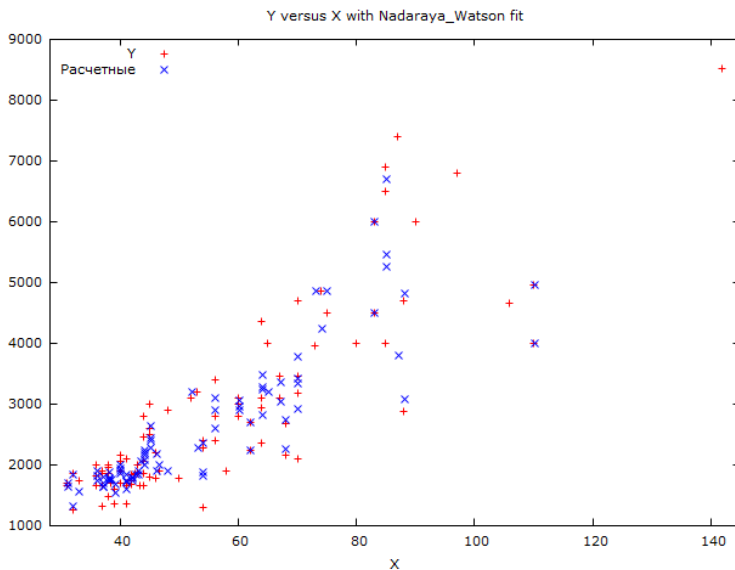


Рисунок 10 – Визуализация прогноза согласно модели ядерного сглаживания Надарая-Ватсона

На рисунке 10 графически представлено сравнение коэффициента простой регрессии (60,9), полученной с использованием МНК и коэффициентов квантильной регрессии. Уравнение обычной регрессии имеет вид:

$$\hat{Y} = -602 + 60,9X, \quad (231) \quad (3,83)$$

в скобках указаны стандартные ошибки.

Наглядность рисунка 18.8 позволяет утверждать, что использование обычной регрессии для оценки стоимости жилья не является корректным.

На рисунках 18.9-18.10 представлены прогнозные значения стоимости квадратного метра жилья, полученные согласно указанным моделям ядерного сглаживания (точные значения можно получить с использованием *Gretl*).

2 Временные ряды

Экономические явления, их связи и зависимости могут рассматриваться как в пространстве, так и во времени, путем построения и анализа одного или нескольких временных рядов.

Временной ряд – это совокупность числовых значений изучаемого показателя в последовательные моменты или периоды времени. Он состоит из значений или уровней временного ряда (Y) и периодов или моментов времени (t). Первый член временного ряда называют начальным (y_1), а последний конечным (y_n). Тогда временной ряд имеет вид:

$$\begin{array}{cccccc} t: & 1 & 2 & 3 & \dots & n \\ Y_t: & Y_1 & Y_2 & Y_3 & \dots & Y_n \end{array} \quad (2.1)$$

Модели, построенные по данным, характеризующим один объект за ряд последовательных моментов или периодов времени, называются моделями временных рядов.

Уровни временного ряда формируются под воздействием большого числа факторов, которые условно можно подразделить на три группы:

- факторы, формирующие тенденцию изменения уровней временного ряда – трендовая компонента (T);
- факторы, формирующие циклические или сезонные колебания уровней ряда – циклическая компонента (S);
- случайные факторы – случайная компонента (ε).

Модель, в которой временной ряд представлен как сумма перечисленных выше компонент, называется аддитивной моделью временного ряда ($Y=T+S+\varepsilon$). Если временной ряд представлен как произведение компонент, то она называется мультипликативной моделью временного ряда ($Y=T \cdot S \cdot \varepsilon$).

Основная задача эконометрического исследования отдельного временного ряда – выявление и количественная оценка

каждой из компонент с целью использования полученной информации для анализа и прогнозирования будущих значений ряда.

При наличии во временном ряду тенденции и циклических колебаний значения каждого последующего уровня ряда зависят от предыдущих. Корреляционную зависимость между последовательными уровнями временного ряда называют автокорреляцией уровней ряда. Количественно ее можно измерить с помощью линейного коэффициента корреляции между уровнями исходного временного ряда и уровнями этого ряда, сдвинутыми на один или несколько периодов или моментов времени, называемого коэффициентом автокорреляции.

Коэффициент автокорреляции уровней ряда первого порядка, смещенных на одну единицу времени, определяется по формуле:

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

$$\text{где } \bar{y}_1 = \frac{\sum_{t=2}^n y_t}{n-1}; \quad \bar{y}_2 = \frac{\sum_{t=2}^n y_{t-1}}{n-1}; \quad (2.2)$$

Коэффициент автокорреляции уровней ряда второго порядка:

$$r_2 = \frac{\sum_{t=3}^n (y_t - \bar{y}_3) \cdot (y_{t-2} - \bar{y}_4)}{\sqrt{\sum_{t=3}^n (y_t - \bar{y}_3)^2 \cdot \sum_{t=3}^n (y_{t-2} - \bar{y}_4)^2}}, \quad (2.3)$$

$$\text{где } \bar{y}_3 = \frac{\sum_{t=3}^n y_t}{n-2}; \quad \bar{y}_4 = \frac{\sum_{t=3}^n y_{t-2}}{n-2}.$$

Аналогично можно определить коэффициенты автокорреляции более высоких порядков.

Так как коэффициент автокорреляции строится по аналогии с линейным коэффициентом корреляции, то по нему можно судить о наличии линейной или близкой к линейной тенденции. Чем ближе коэффициент автокорреляции первого порядка к единице, тем более выражена линейная тенденция. Для некоторых временных рядов, имеющих сильную нелинейную тенденцию, коэффициент автокорреляции уровней исходного ряда может приближаться к нулю.

Последовательность значений коэффициентов автокорреляции уровней первого, второго и т. д. порядков называют автокорреляционной функцией временного ряда. Если наиболее высоким оказался коэффициент автокорреляции первого порядка, исследуемый ряд содержит только тенденцию. Если наиболее высоким оказался коэффициент автокорреляции порядка τ , то ряд содержит циклические или сезонные колебания с периодичностью в τ моментов времени. Если ни один коэффициент не является значимым, можно сделать вывод о том, что либо ряд не содержит тенденции и циклических колебаний, либо содержит сильную нелинейную тенденцию.

Число периодов или моментов времени, по которым рассчитывается коэффициент автокорреляции, называют лагом.

Построение аналитической функции для моделирования тенденции (тренда) временного ряда называют аналитическим выравниванием временного ряда. Тенденция во времени может принимать разные формы, для ее формализации используют следующие функции:

- линейная: $y_t = a + bt$;
- степенная: $y_t = at^b$;
- гиперболола: $y_t = a + b/t$;
- экспонента: $y_t = e^{a-bt}$;
- показательная: $y_t = ab^t$;
- полином k -ого порядка: $y_t = a + b_1t + b_2t^2 + \dots + b_k t^k$;
- логическая: $y_t = \frac{1}{1+be^{-ct}}$;
- Гомперца: $\log_c f(x) = a - bc^t$, где $0 < c < 1$.

Параметры каждой из перечисленных выше функций определяются методом наименьших квадратов, используя в качестве независимой переменной время (t), а в качестве зависимой переменной – фактические уровни временного ряда (y_t). Для нелинейных трендов предварительно проводят стандартную процедуру линеаризации (таблица 8).

При выборе конкретной функции предпочтение отдается той, которая имеет меньшую сумму квадратов отклонений фактических уровней временного ряда от теоретических, найденных по уравнениям тренда.

Таблица 1 – Линеаризующие преобразования

Функция	Преобразования переменных	
	y	t
$y_t = a + b/t$	y	$1/t$
$y_t = e^{a+bt}$	$\ln y$	t
$y_t = a t^b$	$\lg y$	$\lg t$
$y_t = a + b_1t + b_2t^2 + \dots + b_kt^k$	y	$t_1=t, t_2=t^2, \dots, t_k=t^k$
$y_t = ab^t$	$\lg y$	t

Наиболее простую экономическую интерпретацию имеют параметры линейного и экспоненциального трендов. Для линейного тренда: a – начальный уровень временного ряда в момент времени $t = 0$; b – средний за единицу времени абсолютный прирост уровней ряда. Для показательного тренда: a – начальный уровень временного ряда в момент времени $t = 0$; b – средний за единицу времени коэффициент роста уровней ряда. Критерием отбора наилучшей формы тренда является значение скорректированного коэффициента детерминации: чем выше его значение, тем лучше форма тренда отражает тенденцию изменения уровней ряда.

Пример 1. Имеются данные о средней цене реализации пшеницы сельскохозяйственными производителями Краснодарского края.

Таблица 2 – Средняя цена реализации 1 тонны пшеницы с.-х. производителями Краснодарского края

Год	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Средняя цена 1 т, руб.	2635	2331	3105	5186	5101	4212	3962	5255	7413	7033

Требуется:

1. Построить график динамики цены реализации 1 т пшеницы.
2. Рассчитать коэффициент автокорреляции первого порядка.
3. Обосновать выбор типа уравнения тренда и рассчитать его параметры.
4. Дать интерпретацию параметров тренда и сделать выводы по результатам решения.

Решение

1. Рассмотрим систему координат $Y0t$, где Y_t – средняя цена реализации 1 т пшеницы, т; t – порядковый номер года и нанесем в ней данные примера на график.

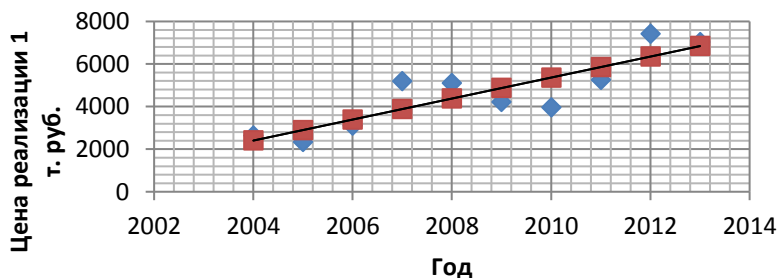


Рисунок 1 – Динамика средней цены реализации 1 т пшеницы сельскохозяйственными производителями Краснодарского края

2. Определим коэффициент автокорреляции первого порядка, характеризующего степень тесноты связи между последовательными уровнями временного ряда цены реализации, сдвинутыми на один год. Заполним вспомогательную таблицу 3.

Таблица 3 – Вспомогательная таблица для расчета коэффициента автокорреляции

t	y_t	y_{t-1}	$y_t - \bar{y}_1$	$y_{t-1} - \bar{y}_2$	$(y_t - \bar{y}_1)(y_{t-1} - \bar{y}_2)$	$(y_t - \bar{y}_1)^2$	$(y_{t-1} - \bar{y}_2)^2$
1	2635	–	–	–	–	–	–
2	2331	2635	-2513,22	-1720,56	4324145,8	6316274,8	2960326,7
3	3105	2331	-1739,22	-2024,56	3521155,2	3024886,2	4098843,2
4	5186	3105	341,78	-1250,56	-427416,4	116813,6	1563900,3
5	5101	5186	256,78	830,44	213240,4	65936,0	689630,6
6	4212	5101	-632,22	745,44	-471282,1	399702,1	555680,8
7	3962	4212	-882,22	-143,56	126651,5	778312,1	20609,5
8	5255	3962	410,78	-393,56	-161666,6	168740,2	154889,5
9	7413	5255	2568,78	899,44	2310463,5	6598630,7	808992,3
10	7033	7413	2188,78	3057,44	6692063,5	4790757,9	9347939,4
Сумма	46233	39200	0,02	-0,04	16127354,8	22260053,6	20200812,3

$$\bar{y}_1 = \frac{\sum_{t=2}^n y_t}{n-1} = \frac{46233 - 2635}{9} = 4844,22;$$

$$\bar{y}_2 = \frac{\sum_{t=2}^n y_{t-1}}{n-1} = \frac{39200}{9} = 4355,56;$$

$$r_1 = \frac{\sum_{t=2}^n (y_t - \bar{y}_1) \cdot (y_{t-1} - \bar{y}_2)}{\sqrt{\sum_{t=2}^n (y_t - \bar{y}_1)^2 \cdot \sum_{t=2}^n (y_{t-1} - \bar{y}_2)^2}} = \frac{16127354,8}{\sqrt{22260053,6 \cdot 20200812,3}} = 0,7605.$$

3. Полученное значение коэффициента автокорреляции и графическое изображение временного ряда позволяют сделать

вывод о том, что ряд средней цены реализации пшеницы содержит тенденцию, близкую к линейной. Поэтому для моделирования его тенденции используем линейную функцию

$$y = a + bt.$$

Для расчета параметров линейного тренда a и b используем метод наименьших квадратов, для чего составим и решим следующую систему:

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

Для составления и решения системы уравнений заполним таблицу 4.

Таблица 4 – Вспомогательная таблица для расчета параметров тренда

№ п/п	y	t	Yt	t^2	\hat{y}_t	$(y-\hat{y}_t)^2$
1	2635	1	2635	1	2404,8	52992,04
2	2331	2	4662	4	2897,8	321262,24
3	3105	3	9315	9	3390,8	81681,64
4	5186	4	20744	16	3883,8	1695724,84
5	5101	5	25505	25	4376,8	524465,64
6	4212	6	25272	36	4869,8	432700,84
7	3962	7	27734	49	5362,8	1962240,64
8	5255	8	42040	64	5855,8	360960,64
9	7413	9	66717	81	6348,8	1132521,64
10	7033	10	70330	100	6841,8	36557,44
Сумма	46233	55	294954	385	46233	6601107,6
Среднее значение	4623,3	5,5	29495,4	38,5		

Воспользуемся формулами, вытекающими из системы:

$$b = \frac{\overline{yt} - \bar{y} \cdot \bar{t}}{\overline{t^2} - \bar{t}^2} = \frac{29495,4 - 4623,3 \cdot 5,5}{38,5 - 5,5^2} = 493;$$

$$a = \bar{y} - b\bar{t} = 4623,3 - 493 \cdot 5,5 = 1911,8 \Rightarrow Y = 1911,8 + 463 t.$$

Таким образом, в среднем ежегодно за 2004–2013 гг. средняя цена реализации пшеницы сельскохозяйственными товаропроизводителями Краснодарского края увеличивалась на 493 руб. за 1 т. Парный коэффициент корреляции между Y и t составил 0,867, что свидетельствует о достаточно хорошем отражении тенденции роста цены линейным уравнением.

Рассчитаем прогнозное значение средней цены реализации пшеницы сельскохозяйственными товаропроизводителями Краснодарского края на 2016 г. путем подстановки в уравнение линейного тренда значения $t_{np}=13$:

$$\hat{Y}_{np} = 1911,8 + 493 \cdot 13 \rightarrow \hat{Y}_{np} = 8320,8 \text{ руб. за 1 т.}$$

Однако точечный прогноз нереален и он дополняется расчетом интервальной оценки \hat{y}_{np}^* с учетом 95 %-й доверительной вероятности:

$$\hat{Y}_{np} - t_{\alpha} \cdot m_{\hat{y}_t} \leq y_{np}^* \leq \hat{Y}_{np} + t_{\alpha} \cdot m_{\hat{y}_t},$$

где: t_{α} – критическое значение t -критерия Стьюдента при уровне значимости α и числе степеней свободы

$$df = n-2;$$

$$m_{\hat{y}_t} = \sqrt{S_{ocm}^2 \left(\frac{1}{n} + \frac{(t_{np} - \bar{t})^2}{\sum(t - \bar{t})^2} \right)} \quad \text{— стандартная ошибка прогноза.}$$

$$S^2 = \frac{\sum(y - \hat{y}_t)^2}{n-2} = \frac{6601107,6}{8} = 825138,45,$$

$$m_{\hat{y}_t} = \sqrt{825138,45 \cdot \left(\frac{1}{10} + \frac{(13 - 5.5)^2}{82.5} \right)} = 803,2$$

при $t_{\alpha=0.05; df=8} = 2,3$.

$$8320,8 - 2,3 \cdot 803,2 \leq y_{np}^* \leq 8320,8 + 2,3 \cdot 803,2,$$

$$6473,4 \leq y_{np}^* \leq 10168,2.$$

Следовательно, с доверительной вероятностью 0,95 можно утверждать, что средняя цена реализации пшеницы сельскохозяйственными товаропроизводителями Краснодарского края в 2016 г. будет находиться в интервале от 6473,2 до 10168,2 руб. за 1 т.

2. Задания по контрольной работе

Задание 1. Имеются данные по совокупности сельскохозяйственным предприятиям центральной зоны Краснодарского края за 2013 г.

По одному варианту задания требуется:

1. Определить параметры множественного уравнения регрессии в натуральной и стандартизованной форме.
2. Найти средние коэффициенты эластичности для каждого фактора.
3. Рассчитать коэффициенты частной и множественной корреляции.
4. Определить общий и частные критерии F -Фишера.
Написать выводы по результатам расчетов.

Варианты задач по данным приложений А и В

1. Производственная себестоимость молока, надой молока на среднегодовую корову, затраты на корма на 1 ц молока.
2. Производственная себестоимость молока, надой молока на среднегодовую корову, затраты труда на 1 ц молока.

3. Производственная себестоимость молока, надой молока на среднегодовую корову, среднегодовое поголовье коров.
4. Производственная себестоимость молока, надой молока на среднегодовую корову, доля молока в выручке от реализации продукции животноводства.
5. Производственная себестоимость молока, надой молока на среднегодовую корову, оплата труда на 1 чел.-ч.
6. Коммерческая себестоимость молока, надой молока на среднегодовую корову, затраты на корма на 1 ц молока.
7. Коммерческая себестоимость молока, надой молока на среднегодовую корову, затраты труда на 1 ц молока.
8. Коммерческая себестоимость молока, надой молока на среднегодовую корову, среднегодовое поголовье коров.
9. Коммерческая себестоимость молока, надой молока на среднегодовую корову, доля молока в выручке от реализации продукции животноводства.
10. Коммерческая себестоимость молока, надой молока на среднегодовую корову, оплата труда на 1 чел.-ч.
11. Надой молока на среднегодовую корову, среднегодовое поголовье коров, доля молока в выручке от реализации продукции животноводства.
12. Надой молока на среднегодовую корову, среднегодовое поголовье коров, оплата труда на 1 чел.-ч.
13. Надой молока на среднегодовую корову, доля молока в выручке от реализации продукции животноводства, оплата труда на 1 чел.-ч.
14. Надой молока на среднегодовую корову, среднегодовое поголовье коров, затраты на корма на среднегодовую корову.
15. Надой молока на среднегодовую корову, оплата труда на 1 чел.-ч, затраты на корма на среднегодовую корову.
16. Надой молока на среднегодовую корову, доля молока в выручке от реализации продукции животноводства, затраты на корма на среднегодовую корову.
17. Затраты труда на 1 ц молока, надой молока на среднегодовую корову, среднегодовое поголовье коров.

18. Затраты труда на 1 ц молока, надой молока на среднегодовую корову, доля молока в выручке от реализации продукции животноводства.
19. Затраты труда на 1 ц молока, надой молока на среднегодовую корову, оплата труда на 1 чел.-ч.
20. Затраты труда на 1 ц молока, среднегодовое поголовье коров, доля молока в выручке от реализации продукции животноводства.
21. Затраты труда на 1 ц молока, среднегодовое поголовье коров, оплата труда на 1 чел.-ч.
22. Затраты труда на 1 ц молока, доля молока в выручке от реализации продукции животноводства, оплата труда на 1 чел.-ч.

Варианты задач по данным приложений Б и Г

23. Выручка от реализации продукции на 1 га сельскохозяйственных угодий, среднегодовая стоимость основных фондов на 1 га сельскохозяйственных угодий, среднегодовая стоимость оборотных средств на 1 га сельскохозяйственных угодий.
24. Выручка от реализации продукции на 1 га сельскохозяйственных угодий, среднегодовая стоимость основных фондов на 1 га сельскохозяйственных угодий, среднегодовая численность работников на 100 га сельскохозяйственных угодий.
25. Выручка от реализации продукции на 1 га сельскохозяйственных угодий, среднегодовая стоимость основных фондов на 1 га сельскохозяйственных угодий, материальные затраты на 1 га сельскохозяйственных угодий.
26. Выручка от реализации продукции на 1 га сельскохозяйственных угодий, среднегодовая стоимость основных фондов на 1 га сельскохозяйственных угодий начислено заработной платы на 1 га сельскохозяйственных угодий.
27. Выручка от реализации продукции на 1 га сельскохозяйственных угодий, среднегодовая численность работников на

100 га сельскохозяйственных угодий, материальные затраты на 1 га сельскохозяйственных угодий.

28. Выручка от реализации продукции на 1 га сельскохозяйственных угодий, среднегодовая численность работников на 100 га сельскохозяйственных угодий, среднегодовая стоимость оборотных средств на 1 га сельскохозяйственных угодий.

29. Выручка от реализации продукции на 1 га сельскохозяйственных угодий, материальные затраты на 1 га сельскохозяйственных угодий, начислено заработной платы на 1 га сельскохозяйственных угодий.

30. Выручка от реализации продукции на 1 га сельскохозяйственных угодий, материальные затраты на 1 га сельскохозяйственных угодий, энергетические мощности на 1 га сельскохозяйственных угодий.

31. Выручка от реализации продукции на 1 га сельскохозяйственных угодий, энергетические мощности на 1 га сельскохозяйственных угодий, среднегодовая стоимость оборотных средств на 1 га сельскохозяйственных угодий.

32. Выручка от реализации продукции на 1 га сельскохозяйственных угодий, энергетические мощности на 1 га сельскохозяйственных угодий, начислено заработной платы на 1 га сельскохозяйственных угодий.

33. Выручка от реализации продукции на 1 га пашни, среднегодовая стоимость основных фондов на 1 га пашни, среднегодовая стоимость оборотных средств на 1 га пашни.

34. Выручка от реализации продукции на 1 га пашни, среднегодовая стоимость основных фондов на 1 га с пашни, среднегодовая численность работников на 100 га пашни.

35. Выручка от реализации продукции на 1 га пашни, среднегодовая стоимость основных фондов на 1 га пашни, материальные затраты на 1 га пашни.

36. Выручка от реализации продукции на 1 га пашни, среднегодовая стоимость основных фондов на 1 га пашни, начислено заработной платы на 1 га пашни.

37. Выручка от реализации продукции на 1 га пашни, среднегодовая численность работников на 100 га пашни, материальные затраты на 1 га пашни.
38. Выручка от реализации продукции на 1 га, пашни среднегодовая численность работников на 100 га пашни, среднегодовая стоимость оборотных средств на 1 га пашни.
39. Выручка от реализации продукции на 1 га пашни, материальные затраты на 1 га пашни, начислено заработной платы на 1 га пашни.
40. Выручка от реализации продукции на 1 га пашни, материальные затраты на 1 га пашни, энергетические мощности на 1 га пашни.
41. Выручка от реализации продукции на 1 га пашни, энергетические мощности на 1 га пашни а, среднегодовая стоимость оборотных средств на 1 га пашни.
42. Выручка от реализации продукции на 1 га пашни, энергетические мощности на 1 га пашни, начислено заработной платы на 1 га пашни.

Исходные данные для задач 1 – 22 представлены в приложениях А и В, а для задач 23 –42 в приложениях Б и Г.

Задание 2. *Провести корреляционно-регрессионный анализ по варианту, имеющихся данных о деятельности сельскохозяйственных организаций 2018 г., , проверив данные на: мультиколлинеарность, гетероскедастичность, нормальность остатков, однородность (тест Чоу). Описать данные и сделать выводы (данные скачать на сайте кафедр).*

Задание 3. *Ввести фиктивные переменные, выбрав заданную зону в качестве базовой, получить оценки медианной и квантильной регрессии. Сделать выводы, сравнить результаты с задачей 2.*

ЗАМЕЧАНИЕ. Переменная Y и зона заданы по варианту, переменные Xj выбираются самостоятельно – содержательно с учетом статистического смысла.

Вариант	Результативная переменная	Факторные (X1-X8)	Природно-климатическая зона
1	Y1	-	Центральная
2	Y2	-	Центральная
3	Y3	-	Центральная
4	Y4	-	Центральная
5	Y5	-	Центральная
6	Y6	-	Центральная
7	Y7	-	Центральная
8	Y8	-	Центральная
9	Y9	-	Центральная
10	Y1	-	Северная
11	Y2	-	Северная
12	Y3	-	Северная
13	Y4	-	Северная
14	Y5	-	Северная
15	Y6	-	Северная
16	Y7	-	Северная
17	Y8	-	Северная
18	Y9	-	Северная

Характеристика исходных данных о деятельности сельскохозяйственных организаций в 2018 г

Y1 – Выручка, тыс. руб. на 1 га с/х угодий;

Y2 – Выручка, тыс. руб. на среднегодового работника;

Y3 – Выручка, тыс. руб. на 1000руб. основных средств;

Y4 – Выручка, тыс. руб. на 1000 руб. оборотных средств;

- У5 – Выручка, тыс. руб. на 1000 руб. затрат;
- У6 – Прибыль от реализации на 1 га, тыс. руб.;
- У7 – Чистая прибыль на 1000 руб. капитала, тыс. руб.;
- У8 – Чистая прибыль на 1000 руб. основных средств, тыс. руб.;
- У9 – Чистая прибыль на 1000 руб. оборотных активов, тыс. руб.;
- Х1 – основные средства, на 1 работника, тыс. руб.;
- Х2 – основные средства, на 1 га с/х угодий, тыс. руб.;
- Х3 – оборотных средств на 1000 руб. основных;
- Х4 – оборотных средств на 1 га;
- Х5 – доля активной части основных средств, %;
- Х6 – производственные затраты на 1 га с/х угодий, тыс. руб.;
- Х7 – энергетические мощности на 100 га с/х угодий, л. с.;
- Х8 – энергетические мощности на 1 работника, л. с.

Задание 4. По статистическим данным Краснодарского края, указанным в приложениях Д и Е, в соответствии с заданным вариантом контрольной работы:

- а) построить график временного ряда;
- б) рассчитать коэффициент автокорреляции первого порядка;
- в) обосновать выбор типа уравнения тренда и рассчитать его параметры;
- г) определить точечный и интервальный прогноз уровня временного ряда на 2016 год;
- д) сделать выводы по задаче.

Приложение А

Таблица А 1 – Показатели производства молока в с/х предприятиях
центральной зоны Краснодарского края, 2013 г.

№ п/п	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9
1	1575	1575	60,6	738	1,69	760	35,6	162,4	44,7
2	1308	1394	71,9	582	1,23	700	91,0	314,0	41,9
3	1695	1728	35,3	860	3,70	460	41,5	58,8	30,3
4	1645	1802	52,2	660	3,86	536	82,5	106,7	34,5
5	1474	1474	49,0	521	1,71	1650	72,3	194,8	25,5
6	1421	1429	60,8	744	1,41	2600	82,2	196,4	45,2
7	1532	1543	26,6	642	3,20	742	71,9	121,8	17,0
8	1611	1773	34,4	607	3,52	331	58,3	89,1	20,8
9	1296	1367	54,6	663	1,45	1100	76,0	363,4	36,2
10	1536	1536	53,7	703	1,88	712	90,2	128,0	37,8
11	1728	1731	53,8	689	2,57	500	89,4	131,2	37,1
12	1762	1764	42,1	1123	4,33	686	71,8	87,8	47,3
13	1177	1177	52,5	695	2,56	2021	79,6	119,8	36,5
14	1532	1532	62,4	523	3,21	800	71,4	164,0	32,6
15	1458	1540	43,7	516	1,63	337	61,5	210,1	22,6
16	1204	1259	73,3	534	1,49	1450	82,0	298,6	39,1
17	1110	1203	62,3	538	1,38	1637	67,3	113,5	33,5
18	2030	2169	39,3	1272	2,92	500	41,6	278,9	50,0
19	2348	2388	30,5	1106	3,69	632	51,2	92,0	33,7

Обозначения:

X_1 – производственная себестоимость 1 ц молока, руб.;

X_2 – коммерческая себестоимость 1 ц молока, руб.;

X_3 – надой молока на среднегодовую корову, ц;

X_4 – затраты на корма на 1 ц молока, руб.;

X_5 – затраты труда на 1 ц молока, чел.-ч.;

X_6 – среднегодовое поголовье коров, гол.;

X_7 – доля молока в выручке от реализации продукции животноводства, %;

X_8 – оплата труда на 1 чел.-ч, руб.;

X_9 – затраты на корма на среднегодовую корову, тыс. руб.

Приложение Б

Таблица Б 1 – Показатели производства продукции и факторов на 1 га сельхозугодий и пашни в сельскохозяйственных организациях Краснодарского края 2013 г.

№ п/п	Y_1	Y_2	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}
1	55,0	55,5	43,5	65,0	65,6	6,52	37,0	37,3	6,43	6,48	12,85	12,95	43,9	80,8	81,4	6,58
2	38,9	39,2	31,6	35,2	35,5	4,65	25,7	25,9	3,37	3,40	7,59	7,66	31,9	20,3	20,5	4,69
3	35,9	36,6	22,9	36,8	37,5	2,64	14,1	14,4	1,73	1,76	5,13	5,22	23,3	24,2	24,7	2,69
4	47,7	52,7	38,8	116,2	128,5	4,16	23,4	25,8	4,73	5,23	9,47	10,47	42,9	39,6	43,8	4,60
5	37,3	37,3	31,2	12,9	12,9	2,98	14,7	14,7	1,45	1,45	6,44	6,44	31,2	49,4	49,4	2,98
6	71,0	71,4	52,5	90,3	90,8	6,36	34,1	34,3	5,25	5,28	11,77	11,84	52,8	74,0	74,4	6,40
7	46,4	46,7	38,5	105,7	106,6	2,52	24,2	24,5	2,27	2,29	6,05	6,10	38,8	18,1	18,3	2,54
8	42,5	43,0	35,9	32,7	33,0	4,92	25,3	25,6	1,65	1,66	9,03	9,13	36,3	51,1	51,6	4,97
9	47,6	48,7	39,6	93,4	95,4	4,48	27,4	28,0	3,34	3,41	9,01	9,20	40,4	31,8	32,5	4,58
10	35,4	35,6	29,9	27,2	27,7	2,36	20,9	21,0	4,19	4,21	4,33	4,35	30,0	24,3	24,4	2,37
11	42,5	42,5	42,1	30,3	30,3	4,39	21,6	21,6	1,27	1,27	8,35	8,35	42,1	47,9	47,9	4,39
12	36,9	37,1	25,9	25,8	26,0	1,90	21,7	21,8	1,89	1,90	4,40	4,43	26,1	28,4	28,6	1,92
13	34,1	34,1	29,7	38,1	38,1	2,86	18,1	18,1	1,80	1,80	4,80	4,80	29,7	18,3	18,3	2,86
14	21,3	21,3	11,0	16,2	16,2	0,83	12,3	12,3	0,76	0,76	1,46	1,46	11,0	20,4	20,4	0,83
15	23,7	23,7	15,8	30,0	30,0	1,24	9,3	9,3	2,10	2,10	2,88	2,88	15,8	34,9	34,9	1,24
16	54,8	54,8	32,6	37,4	37,4	3,00	35,6	35,6	2,38	2,38	6,07	6,07	32,6	33,9	33,9	3,00
17	23,3	23,3	15,3	23,6	23,6	2,30	12,5	12,5	2,51	2,51	2,33	2,33	15,3	11,0	11,0	2,30
18	25,6	25,6	23,5	20,2	20,2	5,42	12,4	12,4	1,55	1,55	9,53	9,53	23,5	20,8	20,8	5,42
19	37,8	40,4	37,6	44,9	48,0	3,02	23,4	25,0	1,81	1,93	6,85	7,32	40,2	92,1	98,4	3,23
20	50,7	52,6	43,6	34,3	35,6	8,97	29,3	30,4	5,33	5,53	15,22	15,79	45,3	30,7	31,8	9,31

Условные обозначения:

U_1 – выручка от реализации на 1 га сельскохозяйственных угодий, тыс. руб.;

U_2 – выручка от реализации продукции на 1 га пашни, тыс. руб.;

X_1 – затраты на реализованную продукцию на 1 га сельскохозяйственных угодий, тыс. руб.;

X_2 – среднегодовая стоимость основных фондов на 1 га сельскохозяйственных угодий, тыс. руб.;

X_3 – среднегодовая стоимость основных фондов на 1 га пашни, тыс. руб.;

X_5 – материальные затраты на 1 га сельскохозяйственных угодий, тыс. руб.;

X_6 – материальные затраты на 1 га пашни, тыс. руб.;

X_7 – энергетические мощности на 1 га сельскохозяйственных угодий, л. с.;

X_8 – энергетические мощности на 1 га пашни, л. с.;

X_9 – начислено заработной платы на 1 га сельскохозяйственных угодий, тыс. руб.;

X_{11} – затраты на реализованную продукцию на 1 га пашни, тыс. руб.;

X_{12} – среднегодовая стоимость оборотных средств на 1 га сельскохозяйственных угодий, тыс. руб.

X_{13} – среднегодовая стоимость оборотных средств на 1 га пашни, тыс. руб.

X_{14} – среднегодовая численность работников на 100 га пашни, чел.

Приложение В

Таблица В 1 – Парные коэффициенты корреляции между показателями производства молока в сельскохозяйственных предприятиях

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9
X_1	1								
X_2	0,983	1							
X_3	-0,641	-0,640	1						
X_4	0,765	0,748	-0,488	1					
X_5	0,650	0,651	-0,674	0,526	1				
X_6	-0,541	-0,589	0,439	-0,235	-0,493	1			
X_7	-0,483	-0,498	0,487	-0,465	-0,290	0,333	1		
X_8	-0,331	-0,287	0,509	-0,159	-0,653	0,172	0,208	1	
X_9	0,112	0,091	0,464	0,528	-0,135	0,210	0,023	0,307	1

Приложение Г

Таблица Г 1– Коэффициенты корреляции между выручкой от реализации продукции на 1 га сельхозугодий (пашни) и факторами

№ п/п	Y_1	Y_2	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}
Y_1	1															
Y_2	0,996	1														
X_1	0,901	0,909	1													
X_2	0,624	0,661	0,601	1												
X_3	0,607	0,649	0,590	0,998	1											
X_4	0,628	0,634	0,722	0,250	0,245	1										
X_5	0,904	0,897	0,813	0,495	0,479	0,610	1									
X_6	0,906	0,906	0,825	0,527	0,514	0,620	0,997	1								
X_7	0,657	0,675	0,603	0,515	0,515	0,659	0,668	0,684	1							
X_8	0,653	0,676	0,605	0,537	0,540	0,658	0,659	0,679	0,998	1						
X_9	0,720	0,734	0,817	0,392	0,390	0,967	0,677	0,693	0,675	0,680	1					
X_{10}	0,715	0,735	0,816	0,415	0,416	0,960	0,671	0,691	0,683	0,691	0,998	1				
X_{11}	0,890	0,906	0,996	0,629	0,623	0,718	0,801	0,820	0,613	0,621	0,820	0,825	1			
X_{12}	0,522	0,534	0,597	0,224	0,225	0,350	0,494	0,510	0,324	0,324	0,474	0,476	0,605	1		
X_{13}	0,511	0,528	0,593	0,239	0,243	0,342	0,485	0,505	0,322	0,325	0,470	0,475	0,606	0,998	1	
X_{14}	0,628	0,639	0,725	0,271	0,269	0,999	0,609	0,622	0,669	0,671	0,971	0,967	0,726	0,353	0,348	1

Приложение Д

Таблица Д 1 – Урожайность сельскохозяйственных культур в
сельскохозяйственных организациях
Краснодарского края, ц с 1 га

№ варианта	Культура	Год									
		2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
43	Зерновые и зернобобовые	42,7	44,5	42,7	41,4	54,3	45,9	48,7	54,5	41,9	52,8
44	Пшеница озимая	44,2	48,2	43,7	46,5	57,4	47,0	51,1	55,9	39,9	51,3
45	Рожь озимая	23,3	28,0	29,5	22,0	27,6	47,9	30,6	47,5	30,9	55,6
46	Кукуруза на зерно	48,1	44,1	43,3	23,8	52,8	38,0	36,4	51,1	43,8	59,1
47	Ячмень озимый	46,5	42,0	44,9	49,2	53,5	49,1	51,7	55,4	38,0	55,4
48	Ячмень яровой	23,4	24,2	28,0	21,5	40,1	30,2	28,9	36,6	28,1	33,7
49	Овес	26,3	26,6	26,4	21,6	36,8	25,7	25,1	31,4	26,4	28,2
50	Просо	9,9	13,4	11,6	12,2	25,8	11,9	13,3	23,8	19,9	21,5
51	Рис	39,7	44,4	46,9	48,4	50,4	60,1	61,8	61,1	63,5	57,6
52	Зернобобовые	23,3	19,7	23,3	14,8	33,8	23,6	23,9	28,6	42,3	21,1
53	Горох	23,6	19,8	23,6	14,8	34,5	24,0	24,1	28,7	22,1	21,9
54	Сахарная свекла	396	328	369	268	448	394	369	448	432	524
55	Масличные культуры	18,0	19,3	19,6	17,1	23,6	21,3	20,3	22,5	21,9	24,9
56	Подсолнечник	18,2	20,9	22,5	20,7	25,3	22,4	22,1	24,1	24,2	27,0
57	Соя	18,0	15,1	12,3	9,4	16,4	18,2	15,8	18,9	18,6	21,2
58	Картофель	86	88	130	105	195	132	126	163	153	170
59	Овощи	91	100	114	85	127	117	98	125	94	102
60	Кукуруза на силос	188	156	170	126	191	166	141	194	155	206
61	Сено многолетних трав	28,5	24,8	27,8	16,2	33,0	23,3	29,5	31,7	25,7	28,3
62	Сено однолетних трав	17,9	29,9	34,6	22,7	32,1	25,7	24,5	33,0	27,3	24,0
63	Плоды и ягоды	58,1	77,5	63,3	72,7	89,7	93,8	73,7	96,0	109	141
64	Виноград	57,1	69,6	49,1	81,8	74,8	86,1	81,7	116	76,3	104

Приложение Е

Таблица Е 1 – Динамика показателей по Краснодарскому краю

№ варианта	Показатель	Год									
		2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
1	2	3	4	5	6	7	8	9	10	11	12
Потребление продуктов питания на члена семьи за год, кг:											
65	Хлебные продукты	113	110	106	116	113	116	118	102	92	94
66	Картофель	65	58	64	71	66	64	63	52	49	54
67	Овощи и бахчевые	98	118	122	134	122	131	127	129	122	119
68	Фрукты, ягоды	38	50	42	53	71	65	92	85	77	81
69	Мясо и мясопродукты	68	70	74	81	85	81	87	88	77	86
70	Молоко и молочные продукты	222	230	251	275	314	283	292	278	263	272
71	Яйца, штук	240	211	231	241	235	239	235	208	193	221
72	Рыба и рыбопродукты	16	17	18	19	23	23	28	24	20	20
73	Площадь жилых помещений на 1 жителя, кв. м.	19,0	19,4	19,8	20,5	21,1	21,6	22,4	22,6	23,1	23,3
Добыча отдельных видов полезных ископаемых											
74	Нефть, тыс. т	1789	1689	1838	1752	1424	1227	1144	1117	1121	1077
75	Газовый конденсат	40	67	102	152	157	115	102	111	114	147
76	Газ природный, млн. куб. м	1970	2656	2999	3281	3372	3126	2818	2900	2970	3144

Продолжение таблицы Е 1

1	2	3	4	5	6	7	8	9	10	11	12
77	Материалы строительные нерудные, млн. куб. м	7,6	7,7	7,9	14,0	15,5	14	19	21	21	21
Производство отдельных видов пищевых продуктов, тыс. т:											
78	Мясо, включая субпродукты 1 категории	93,3	86,4	101	144	121	111	99	116	122	133
79	Изделия колбасные	90,1	93,5	98,7	103	105	96	102	93	87	77
80	Масла растительные нерафинированные	478	484	653	616	538	719	657	639	917	815
81	Масла растительные рафинированные	164	116	195	183	192	270	237	339	433	508
82	Хлеб и хлебобулочные изделия	426	405	377	359	340	326	318	318	321	326
83	Сахар песок	1849	1890	1950	2122	1795	1374	1698	2114	1229	1076
84	Майонез	8,3	10,9	13,6	13,6	14,4	12,9	19,4	19,0	19,6	23,7
85	Воды минеральные, млн. полулитров	118	118	139	210	250	262	276	244	284	275
86	Топливо дизельное	1933	2253	2800	3256	3267	3318	3276	3534	3896	4349
87	Мазут топочный	2924	3355	4002	4707	4903	4819	5427	5804	5947	6843
88	Топливо печное бытовое	124	187	101	76	118	201	353	553	545	621
89	Электрoэнергия, млн. кВт-часов	6165	6697	6905	7079	6689	6194	6620	6589	7980	9944

Оглавление

ВВЕДЕНИЕ.....	3
1 Методические рекомендации к выполнению контрольной работы.....	8
1 Множественный корреляционно-регрессионный анализ.....	27
2 Временные ряды.....	48
2 Задания по контрольной работе.....	57
ПРИЛОЖЕНИЯ.....	65

Эконометрика (продвинутый уровень)

**Методические рекомендации
к выполнению рубежной контрольной работы**

Составитель:

**Ворокова Нодира Хасановна,
Сенникова Алина Евгеньевна,
Жминько Альбина Евгеньевна**

В авторской редакции

Подписано в печать 30.11.2021. Формат 60 × 84 1/16.
Усл. печ. л. – 3,29. Уч.- изд. л.–2,08. Тираж 500 экз. Заказ №
....