МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования
РОССИЙСКИЙ ЭКОНОМИЧЕСКИЙ УНИВЕРСИТЕТ ИМ. Г.В. ПЛЕХАНОВА
(Оренбургский филиал РЭУ им. Г.В. Плеханова)

Лаптева Е. В.
Портнова Л.В.

Статистические методы исследований в экономике

Учебное пособие

2022

УДК 311
ББК 65в6
Л 24

РЕЦЕНЗЕНТЫ:

Д. Е. Бекбергенева, кандидат экономических наук, доцент кафедры общегуманитарных, социально-экономических, математических и естественнонаучных дисциплин Оренбургского института (филиала) университета имени О.Е.Кутафина (МГЮА);
Е. П. Огородникова, кандидат экономических наук, доцент кафедры экономической теории и управления ФГБОУ ВО «Оренбургский государственный аграрный университет»

Лаптева Е.В., Портнова Л.В.

Л 24 Статистические методы исследований в экономике: учебное пособие (второе издание, доработанное и дополненное) / Е.В.Лаптева, Л.В. Портнова. – Волгоград: Изд-во «Сфера», 2022. – 234 с.

В учебном пособии в систематизированном виде изложены и представлены статистические методы исследований в экономике, приведены конкретные примеры их применения в анализе социально-экономических явлений и процессов в экономике, представлены методики использования программных продуктов для исследований.

Учебное пособие предназначено для студентов магистратуры по направлениям подготовки «Экономика», «Менеджмент» высших учебных заведений, преподавателей, аспирантов, научных работников, работников органов государственного и муниципального управления, руководителей организаций.

Работа издана в авторской редакции

ISBN 978-5-00186-078-5

ОГЛАВЛЕНИЕ

Глава 1. ПОНЯТИЕ СТАТИСТИЧЕСКИХ МЕТОДОВ ИССЛЕДОВАНИЯ В ЭКОНОМИКЕ

Статистические методы исследования в экономике – понятие достаточно обширное, включающее в себя комплекс статистических методов с целю изучения конкретных социально-экономических явлений и процессов.

В большом медицинском словаре можно встретить следующее определение статистических методов: «Статистический метод – совокупность взаимосвязанных приемов исследования массовых объектов и явлений с целью получения количественных характеристик и выявления общих закономерностей путем устранения случайных особенностей отдельных единичных наблюдений»[1].

В научной литературе можно встретить более простое толкование данного понятия: «Статистические методы – это методы анализа статистических данных»[2].

Типовые примеры раннего этапа применения статистических методов описаны в Библии, в Ветхом Завете. Там, в частности, приводится число воинов в различных племенах. С математической точки зрения дело сводилось к подсчёту числа попаданий значений наблюдаемых признаков в определённые градации.

Сразу после возникновения теории вероятностей (Паскаль, Ферма, XVII век) вероятностные модели стали использоваться при обработке статистических данных. Например, изучалась частота рождения мальчиков и девочек, было установлено отличие вероятности рождения мальчика от 0,5, анализировались причины того, что в парижских приютах эта вероятность не та, что в самом Париже, и т. д.

В 1794 г. (по другим данным – в 1795 г.) немецкий математик Карл Гаусс формализовал один из методов современной математической статистики метод наименьших квадратов[3]. В XIX веке значительный вклад в развитие практической статистики внёс бельгиец Кетле, на основе анализа большого числа реальных данных показавший устойчивость относительных статистических показателей, таких, как доля самоубийств среди всех смертей[4].

Первая треть XX века прошла под знаком параметрической статистики. Изучались методы, основанные на анализе данных из параметрических семейств распределений, описываемых кривыми семейства Пирсона. Наиболее популярным было нормальное распределение. Для проверки гипотез использовались критерии Пирсона, Стьюдента, Фишера. Были предложены метод максимального правдоподобия, дисперсионный анализ, сформулированы основные идеи планирования эксперимента.

---

[1] Большой медицинский словарь. 2000.

[2] Свободная энциклопедия «Википедия» // http://ru.wikipedia.org/wiki/Статистика

[3] Клейн Ф. Лекции о развитии математики в XIX столетии. Часть I. — Москва, Ленинград: Объединенное научно-техническое издательство НКТП СССР, 1937.

[4] Плошко Б.Г., Елисеева И.И. История статистики: Учеб. пособие. — Москва, Ленинград: Финансы и статистика, 1990.

Разработанную в первой трети XX века теорию анализа данных называют параметрической статистикой, поскольку её основной объект изучения – это выборки из распределений, описываемых одним или небольшим числом параметров. Наиболее общим является семейство кривых Пирсона, задаваемых четырьмя параметрами. Как правило, нельзя указать каких-либо веских причин, по которым распределение результатов конкретных наблюдений должно входить в то или иное параметрическое семейство. Исключения хорошо известны: если вероятностная модель предусматривает суммирование независимых случайных величин, то сумму естественно описывать нормальным распределением; если же в модели рассматривается произведение таких величин, то итог, видимо, приближается логарифмически нормальным распределением и так далее.

Статистические методы анализа данных применяются практически во всех областях деятельности человека. Их используют всегда, когда необходимо получить и обосновать какие-либо суждения о группе (объектов или субъектов) с некоторой внутренней неоднородностью.

Целесообразно выделить три вида научной и прикладной деятельности в области статистических методов анализа данных (по степени специфичности методов, сопряженной с погруженностью в конкретные проблемы):

1) разработка и исследование методов общего назначения, без учета специфики области применения;

2) разработка и исследование статистических моделей реальных явлений и процессов в соответствии с потребностями той или иной области деятельности;

3) применение статистических методов и моделей для статистического анализа конкретных данных[1].

Применение статистических методов и моделей для статистического анализа конкретных данных тесно привязано к проблемам соответствующей области. Результаты третьего из выделенных видов научной и прикладной деятельности находятся на стыке дисциплин. Их можно рассматривать как примеры практического применения статистических методов. Но не меньше оснований относить их к соответствующей области деятельности человека.

Теория статистических методов нацелена на решение реальных задач. Поэтому в ней постоянно возникают новые постановки математических задач анализа статистических данных, развиваются и обосновываются новые методы. Обоснование часто проводится математическими средствами, то есть путем доказательства теорем. Большую роль играет методологическая составляющая — как именно ставить задачи, какие предположения принять с целью дальнейшего математического изучения. Велика роль современных информационных технологий, в частности, компьютерного эксперимента.

Актуальной является задача анализа истории статистических методов с целью выявления тенденций развития и применения их для прогнозирования.

[1] Свободная энциклопедия «Википедия» // http://ru.wikipedia.org/wiki/Статистика

Р.А.Шмойлова отмечает, что метод статистики – это целая совокупность приемов, пользуясь которыми статистика исследует свой предмет. Он включает в себя методы:

1. Статистическое наблюдение – заключается в сборе первичного статистического материала. Это первый этап всякого статистического исследования.

2. Метод группировок дает возможность все собранные в результате массового статистического наблюдения факты подвергать систематизации и классификации. Это второй этап статистического исследования.

3. Метод обобщающих показателей - позволяет характеризовать изучаемые явления и процессы при помощи статистических величин – абсолютных, относительных, средних. На этом этапе статистического исследования выявляются взаимосвязи и масштабы явлений, определяются закономерности развития, даются прогнозные оценки[1].

Классификация статистических методов с этапами исследования представлена в таблице 1.1.

Таблица 1.1 – Классификация статистических методов в соответствии с этапами исследования

Этап статистического исследования	Методы статического исследования
Статистическое наблюдение	По организационным формам: 1. Статическая отчетность 2. Специально организованное наблюдение 3. Регистры По видам статистического наблюдения: 1. Текущее (или непрерывное) 2. Прерывное - периодическое - единовременное По охвату единиц совокупности: 1. Сплошное 2. Несплошное - выборочное - основного массива - монографическое По способам: 1. Непосредственное 2. Документальное 3. Опрос - экспедиционный - корреспондентский

[1] Теория статистики / Под ред. Р. А. Шмойловой. – М.: Финансы и статистика, 2002. – 560 с.

Продолжение таблицы 1.1

	- анкетный - явочный
Первичная обработка, сводка и группировка результатов наблюдения	По задачам: 1. Типологическая группировка 2. Структурная группировка 3. Аналитическая группировка По признаку: 1. Простая 2. Сложная
Анализ полученных сводных материалов	1. Абсолютных, относительных и средних величин 2. Анализ вариации 3. Статистическое изучение взаимосвязи социально-экономических явлений 4. Анализ рядов динамики 5. Статистический анализ структуры и структурных изменений 6. Экономические индексы

В зависимости от цели исследования аналитик самостоятельно выбирает необходимые группы методов.

Следует отметить, что практика применения статистических методов в экономике тесно взаимосвязана с прикладной статистикой. Прикладная статистика – это наука о методах обработки статистических данных[1].

Методы прикладной статистики активно применяются в технических исследованиях, экономике менеджменте, социологии, медицине, геологии, истории и т. д. С результатами наблюдений, измерений, испытаний, опытов, с их анализом имеют дело специалисты во многих областях теоретической и практической деятельности.

В СССР термин «прикладная статистика» вошёл в широкое употребление в 1981 г. после выхода массовым тиражом сборника «Современные проблемы кибернетики (прикладная статистика)». В этом сборнике обосновывалась трёхкомпонентная структура прикладной статистики. Во-первых, в неё входят ориентированные на прикладную деятельность статистические методы анализа данных. Однако прикладную статистику нельзя целиком относить к математике. Она включает в себя две внематематические области: методологию организации статистического исследования и организацию компьютерной обработки данных, в том числе разработку и использование баз данных и электронных таблиц, статистических программных продуктов, например, диалоговых систем анализа данных. В СССР термин «прикладная статистика» использовался и

[1] Орлов А. И. О развитии прикладной статистики. – В сб.: Современные проблемы кибернетики (прикладная статистика). – М.: Знание, 1981, с.3-14.

ранее 1981 г., но лишь внутри сравнительно небольших и замкнутых групп специалистов[1].

Прикладная статистика – методическая дисциплина, являющаяся центром статистики. При применении методов прикладной статистики к конкретным областям знаний и социально-экономическим явлениям в экономике получаются научно-практические дисциплины типа «статистика в промышленности», «статистика в медицине», «статистика в психологии» и др. С этой точки зрения эконометрика – это «статистические методы в экономике»[2]. Математическая статистика играет роль математического фундамента для прикладной статистики.

К настоящему времени очевидно чётко выраженное размежевание этих двух научных направлений. Математическая статистика исходит из сформулированных в 1930 – 1950 гг. постановок математических задач, происхождение которых связано с анализом статистических данных. Начиная с 1970-х годов исследования по математической статистике посвящены обобщению и дальнейшему математическому изучению этих задач.

Прикладная статистика нацелена на решение реальных задач. Поэтому в ней возникают новые постановки математических задач анализа статистических данных, развиваются и обосновываются новые методы. Обоснование часто проводится математическими методами, то есть путём доказательства теорем. Большую роль играет методологическая составляющая – как именно ставить задачи, какие предположения принять с целью дальнейшего математического изучения. Велика роль современных информационных технологий, в частности, компьютерного эксперимента.

Хотя статистические данные собираются и анализируются с незапамятных времён, современная математическая статистика как наука была создана, по общему мнению специалистов, сравнительно недавно - в первой половине XX в. Именно тогда были разработаны основные идеи и получены результаты, излагаемые ныне в учебных курсах математической статистики. После чего специалисты по математической статистике занялись внутриматематическими проблемами, а для теоретического обслуживания проблем практического анализа статистических данных стала формироваться новая дисциплина – прикладная статистика. В настоящее время статистическая обработка данных проводится, как правило, с помощью соответствующих программных продуктов.

Статистический пакет – программный продукт, предназначенный для статистической обработки данных.

Являются надежным инструментом повышения качества принимаемых решений. В пакет, как правило, входит: деловая графика, дисперсионный анализ, регрессионный анализ, анализ временных рядов и пр.

Можно выделить два вида статистических пакетов.

---

[1] Орлов А. И. О развитии прикладной статистики. – В сб.: Современные проблемы кибернетики (прикладная статистика). – М.: Знание, 1981, с.3-14.

[2] Орлов А. И. Эконометрика. Учебник для вузов. Изд. 3-е, исправленное и дополненное. — М.: Изд-во «Экзамен», 2004. — 576 с.

Из зарубежных пакетов это STATGRAPHICS, SPSS, SYSTAT, BMDP,SAS, CSS, STATISTICA, S-plus, и др.

Из российских можно назвать такие пакеты, как STADIA, ЭВРИСТА, МЕЗОЗАВР, ОЛИМП: Стат-Эксперт, Статистик-Консультант, САНИ, КЛАСС-МАСТЕР и др.

Отечественные статистические пакеты, которые устойчиво представлены на рынке в течение последних лет, в значительной степени лишены таких недостатков, которые есть у западных продуктов. Они предполагают наличие широкого первоначального статистического образования, доступной литературы и консультационных служб. Поэтому они содержат мало экранных подсказок и требуют внимательного изучения документации на английском языке.

Основную часть имеющихся пакетов составляют специализированные пакеты и пакеты общего назначения.

Специализированные пакеты обычно содержат методы из одного – двух разделов статистики или методы, используемые в конкретной предметной области (контроль качества промышленной продукции, расчет страховых сумм и т.д.). Чаще всего встречаются пакеты для анализа временных рядов (например, ЭВРИСТА, МИЗОЗАВР, ОЛИМП: Стат-Эксперт), регрессионного и факторного анализа. Обычно эти пакеты содержат весьма полный набор традиционных методов в своей области, а иногда включают также и оригинальные методы и алгоритмы, созданные разработчиками пакета. Как правило, пакет и его документация ориентированы на специалистов, хорошо знакомых с соответствующими методами.

Особое место на рынке занимают так называемые статистические пакеты общего назначения. Широкий диапазон статистических методов, дружелюбный интерфейс пользователя привлекает в них не только начинающих пользователей, но и специалистов. Универсальность этих пакетов особенно полезна:

1) на начальных этапах обработки, когда речь идет о подборе статистической модели или метода анализа данных;

2) когда поведение статистических данных выходит за рамки использовавшейся ранее модели;

3) в процессе обучения основам статистики.

Именно пакеты общего назначения составляют большинство продаваемых на рынке статистических программ. К таким пакетам относятся системы STADIA и SPSS, а также пакеты STATGRAPHICS, STATISTICA, S-plus, и др.

Для того чтобы статистический пакет общего назначения был удобен и эффективен в работе, он должен удовлетворять многочисленным и весьма жестким требованиям. В частности, необходимо, чтобы он:

1) содержал достаточно полный набор стандартных статистических методов;

2) был достаточно прост для быстрого освоения и использования;

3) отвечал высоким требованиям к вводу, преобразованиям и организации хранения данных;

4) имел широкий набор средств графического представления данных и результатов обработки;

5) предоставлял удобные возможности для включения в отчеты таблиц исходных данных, графиков, промежуточных и окончательных результатов обработки;

6) имел подробную документацию, доступную для начинающих и информативную для специалистов-статистиков.

Пакеты, рассчитанные на массового пользователя, стоят дешевле, чем западные – обычно 500-1500 долларов. Эти пакеты отличаются от профессиональных, прежде всего ориентацией на индивидуального пользователя: преимущественно диалоговым режимом работы, наличием ограничений по объему обрабатываемых данных и т.д.

Отечественные статистические пакеты стоят существенно дешевле, как правило, их цена составляет от 50 до 300 долларов.

Глава 2. ГРУППИРОВКА СТАТИСТИЧЕСКИХ ДАННЫХ

Информация об отдельных единицах совокупности, получаемая в процессе статистического наблюдения, характеризует их с различных сторон. Важнейшим этапом исследования является систематизация первичных данных и получение на этой основе сводной характеристики объекта в целом при помощи обобщающих показателей, что достигается путем сводки и группировки первичного статистического материала.

Статистическая сводка – систематизация единичных фактов, позволяющая перейти к обобщающим показателям для выявления типичных черт и закономерностей, присущих изучаемому явлению в целом[1].

Статистическая сводка различается по ряду признаков:

1) по сложности построения;
2) месту проведения;
3) способу обработки материалов статистического наблюдения.

По первому признаку различают:

1) простую сводку – операции по подсчету общих итогов по совокупности единиц наблюдения;

2) сложную сводку – как комплекс операций, включающих группировку единиц наблюдения, подсчет итогов по каждой группе и по всему объекту и представлении результатов в виде статистических таблиц.

По второму признаку различают:

Централизованная, когда весь первичный материал поступает в одну организацию и подвергается в ней обработке от начала до конца; и децентрализованная, когда отчеты сводятся местными статистическими органами, а итоги поступают в Росстат и там определяются итоговые показатели в целом по стране.

По способу обработки сводка бывает механизированной (с использованием ЭВМ) и ручной.

Сводка статистической информации, как правило, не ограничивается получением общих итогов. Чаще всего исходная информация на этой стадии систематизируется, образуются отдельные статистические совокупности, т.е. осуществляется статистическая группировка.

Группировка – это процесс образования однородных групп на основе расчленения статистической совокупности на части или объединения единиц в частные совокупности по определенным, существенным для них признаками.

Признаки, по которым производится распределение единиц наблюдения совокупности на группы, называются группировочными признаками или основанием группировки.

---

[1] Балдин К.В., Рукосуев А.В. Общая теория статистики: Учебное пособие. – М.: Дашков и К, 2010. - 312 с.

С помощью метода группировок решаются следующие задачи: выделение социально-экономических типов явлений; изучение структуры явления и структурных сдвигов; выявление связи и зависимости между явлениями.

В соответствии с задачами группировки различают следующие их виды: типологическая, структурная, аналитическая.

Типологическая – это расчленение однородной совокупности на отдельные, качественно однородные группы и выявление на этой основе экономических типов явлений (группировка предприятий по формам собственности).

При построении такой группировки основное внимание уделяется выбору группировочного признака. Решение об основании группировки осуществляется на основе анализа сущности явлений.

Пример типологической группировки представлен в таблице 2.1.

Таблица 2.1 – Группировка населения в Оренбургской области по уровню образования (условные данные)

Показатели	2021 г.	
	тыс.чел.	в % к итогу
Численность обучающихся – всего	206,1	100
в том числе в: общеобразовательных учреждениях	124,67	60,49
учреждениях начального профессионального образования	11,60	5,63
учреждениях среднего профессионального образования	23,10	11,21
учреждениях высшего профессионального образования	46,25	22,44
аспирантуре и докторантуре	0,45	0,22

Структурной называется группировка, которая предназначена для изучения однородной совокупности по какому-либо варьирующему признаку.

Пример структурной группировки представлен в таблице 2.2.

Группировка, выявляющая взаимосвязи между явлениями и их признаками, называется аналитической.

Пример аналитической группировки представлен в таблице 2.3.

Все признаки в статистике подразделяются на факторные и результативные.

Факторными называются признаки, под воздействием которых изменяются другие признаки – они и образуют группу результативных признаков. Взаимосвязь проявляется в том, что с возрастанием значения факторного признака систематически возрастает или убывает значение результативного признака. Особенностями аналитической группировки является то, что единицы группируются по факторному признаку, каждая выделенная группа характеризуется средними значениями результативного признака.

Таблица 2.2 – Структура инвестиций в строительство по формам собственности в Оренбургской области (условные данные)

Показатели	2018 г.		2019 г.		2020г.		2021 г.		Изменение 2021 г.г. к 2018 г. , п.п.
	всего	% к итогу	всего	% к итогу	всего	% к итогу	всего	% к итогу	
Государственная	548,6	56,6	758,1	55,7	838,3	56,8	863,8	57,4	0,8
Муниципальная	176,2	18,2	188,8	13,6	203,8	13,3	202,1	15,1	-3,1
Частная	116,8	12,1	224,8	18,0	270	16,6	252,8	19,4	7,4
Смешанная российская	56,3	5,8	79,3	4,4	66,8	4,6	70,6	5,2	-0,6
Общественных и религиозных организаций	18,3	1,9	32	3,4	50,4	3,5	52,5	3,3	1,4
Потребительской кооперации	9,1	0,9	14,2	1,3	19,6	1,4	21,8	1,5	0,6
Совместная российская и иностранная	43,9	4,5	52	3,6	54,8	3,8	57,2	4,0	-0,5
Всего	969,2	100,0	1349,2	100,0	1503,7	100,0	1520,8	100,0	х

Таблица 2.3 – Группировка районов Оренбургской области по уровню производства молока в 2021 г. (условные данные)

Номер группы	Группы по объему производства молока, тыс.ц	Среднее значение в группе ($\bar{x}$), тыс.ц		Средне квадратическое отклонение (σ), тыс.ц		Коэффициент вариации (V_σ), %	
		В целом по группе	По подгруппам	В целом по группе	По подгруппам	В целом по группе	По подгруппам
1	До 47	45,6	25,01	26,27	13,27	57,57	53,06
	47,1- 95		94,36		24,17		25,61
2	95,1 – 143	141,2	109,15	27,31	9,69	19,34	8,88
	143,1 – 191		162,56		6,32		3,89
3	191,1 – 239	240,8	196,7	44,05	-	18,29	-
	свыше 239		284,8		-		-
В среднем по совокупности		63,53		47,73		75,13	

Все рассмотренные виды группировок могут быть построены по одному или нескольким существенным признакам.

Группировка, в которой, группы образованы по одному признаку, называются простой.

Сложной называется группировка, в которой расчленение совокупности на группы производится по двум и более признакам, взятыми в сочетании. Сначала группы формируются по одному признаку, затем группы делятся на подгруппы по другому признаку и т.д. Сложные группировки дают

возможность изучить единицы совокупности одновременно по нескольким признакам.

Сложная группировка строится в следующей последовательности: сначала производится группировка по атрибутивным признакам, затем по количественным.

Построение группировок начинается с определения состава группировочного признака. Выбор группировочного признака один из самых сложных вопросов теории группировки.

В основании группировки могут быть положены как количественные, так и атрибутивные признаки.

После этого определяют количество групп, на которые, надо разбить совокупность. Число групп зависит от задач исследования и вида показателя, положенного в основание. Если группировка строится по атрибутивному признаку то групп столько, сколько имеется градаций. Например, группировка по формам собственности.

Если группировка строится по количественному признаку, то следует обратить внимание на число единиц объекта. При небольшом объеме не следует образовывать слишком большое число групп.

Определение числа групп можно осуществить математическим путем с помощью формулы Стерджесса:

$$n=1+3,322*Lg(N) \qquad (2.1)$$

где n- число групп;

N-число единиц совокупности.

Когда определено число групп, следует определить интервал группировки. Интервал – это значение варьирующего признака, лежащее в определенных границах. Каждый интервал имеет свою величину, верхнюю и нижнюю границы или хотя бы одну из них.

Нижней границей называется наименьшее значение признака в интервале, верхней- наибольшее.

Если вариация признака проявляется в узких границах и распределение носит равномерный характер, то строят группировку с равными интервалами.

Величина равного интервала определяется по формуле (2.2).

$$h=R/n=\frac{X_{max}-X_{min}}{n} \qquad (2.2)$$

где R-размах совокупности

X_{max} и X_{min} - макс. и миним. значения признака в совокупности

n- число групп

Если h имеет один знак до запятой, то значения округляют до десятых: 0,66-0,7; 1,372-1,4; 5,8-5,8.

Если, имеется 2 значащие цифры до запятой сколько знаков после запятой, то величину округляют до целого числа: 12,785 – 13.

Если размах вариации велик и значения варьируют неравномерно, то используют группировку с неравными интервалами.

Интервалы группировки могут быть открытыми и закрытыми. Закрытые интервалы, у которых имеются верхняя и нижняя границы. У открытых указана только одна граница- верхняя или нижняя.

Разновидностью группировок является классификация. В основу классификации всегда кладется атрибутивный признак. Классификационные стандарты, устойчивы, разрабатываются органами государственной и международной статистики.

К построению группировок предъявляются следующие требования:

1) количество единиц в группе должно быть не менее 3;

2) количество групп не менее 3;

3) насыщенность центральных групп;

4) в одной группе не более 50% всей совокупности.

Результаты сводки и группировки оформляются в виде статистических рядов распределения.

Статистические ряды распределения – это упорядоченное распределение единиц совокупности по определенному признаку. В зависимости от признака, положенного в основу ряда, различают атрибутивные и вариационные ряды.

Атрибутивными называют ряды, построенные по качественным признакам. Они характеризуют состав совокупности по тем или иным существующим признакам. Пример: распределение студентов группы по полу.

Вариационными рядами называют ряды, построение по количественному признаку.

Любой вариационный ряд состоит из вариантов и частот. Вариантами считаются отдельные значения признака, которые он принимает в вариационном ряду, т.е. конкретные значения варьирующего признака. Частоты – это численность отдельных вариантов или каждой группы вариационного ряда, т.е. это числа, показывающие, как часто встречаются те или иные варианты в ряду распределения. Сумма всех частот определяет численность всей совокупности.

В зависимости от характера вариации признака различают дискретные и интервальные вариационные ряды.

Дискретный вариационный ряд характеризует распределение единиц совокупности по дискретному признаку, когда величина количественного признака принимает только целые значения.

Интервальный ряд характеризует распределение единиц совокупности в случае непрерывной вариации изучаемого признака, величина которого может принимать в определенных пределах любые значения.

Пример 2.1. Группировка данных

Произведем группировку магазинов (таблица 2.4) по признаку торговой площади, образовав 5 групп с равными интервалами. Каждую группу и всю совокупность магазинов охарактеризуем по количеству магазинов; размером торговой площади, товарооборота, издержек обращения; основных фондов; средним уровнем издержек обращения (в % к товарообороту); размером торговой площади, приходящейся на одного продавца. Построим групповую таблицу и сделаем выводы.

Таблица 2.4 - Показатели деятельности магазинов (условные данные)

№ магазина	Товарооборот, млн. руб.	Издержки обращения, млн. руб.	Среднегодовая стоимость основных фондов, млн. руб.	Численность продавцов, чел.	Торговая площадь, $м^2$
10	280	46,8	6,3	105	1353
11	156	30,4	5,7	57	1138
12	213	28,1	5,0	100	1216
13	298	38,5	6,7	112	1352
14	242	34,2	6,5	106	1445
15	130	20,1	4,8	62	1246
16	184	22,3	6,8	60	1332
17	96	9,8	3,0	34	680
18	304	38,7	6,9	109	1435
19	95	11,7	2,8	38	582
20	352	40,1	8,3	115	1677
21	101	13,6	3,0	40	990
22	148	21,6	4,1	50	1354
23	74	9,2	2,2	30	678
24	135	20,2	4,6	52	1380
25	320	40,0	7,1	140	1840
26	155	22,4	5,6	50	1442
27	262	29,1	6,0	102	1720
28	138	20,6	4,8	46	1520
29	216	28,4	8,1	96	1673

Решение:

Определяем шаг интервала: $i = \frac{x_{max} - x_{min}}{m}$,

Число групп: $m = 5, x_{max} = 1840, x_{min} = 582$

$$i = \frac{1840 - 582}{5} = 252\,м^2$$

Формируем интервалы:

Интервалы по величине торговой площади, м^2	Количество магазинов
(582; 834)	3
(834; 1085)	1
(1085; 1337)	4
(1337; 1588)	7
(1588; 1840)	4

Таблица 2.5 - Группировка магазинов по величине торговой площади

Количество магазинов		3	1	4	7	4
Размер торговой площади, м^2	В сумме	1940	990	4932	9928	6910
	В среднем на один магазин	646,7	990,0	1233,0	1418,3	1727,5
Товарооборот, млн.руб.	В сумме	265	101	683	1420	1150
	В среднем на один магазин	88,3	101,0	170,8	202,9	287,5
Издержки обращения, млн. руб.	В сумме	30,7	13,6	100,9	196,2	137,6
	В среднем на один магазин	10,2	13,6	25,2	28,0	34,4
Среднегодовая стоимость основных фондов, млн. руб.	В сумме	8	3	22,3	39,2	29,5
	В среднем на один магазин	2,7	3,0	5,6	5,6	7,4
Средний уровень издержек обращения (в % товарообороту)		11,6	13,5	14,8	13,8	12,0
Размер торговой площади, приходящейся на одного продавца, м^2		19,0	24,8	17,7	18,9	15,3

Вывод: Таким образом, на основе полученных расчетных данных можно сделать вывод, что первая группа размером торговой площади в интервале от 582 до 834 м2 характеризуется самыми низкими показателями в расчете на один магазин. Пятая группа, наоборот, самыми высокими показателями в расчете на один магазин. При этом наблюдается обратное соотношение в размере торговой площади, приходящейся на одного продавца – первая группа характеризуется наибольшей площадью (19 м2), пятая группа – наименьшей (15,2 м2).

Пример 2.2. Группировка данных

Для изучения зависимости между размером прибыли и среднегодовой стоимостью основных производственных фондов (таблица 2.6) произведем группировку предприятий по среднегодовой стоимости основных производственных фондов, образовав четыре группы с равными интервалами. По каждой группе и по совокупности предприятий в целом подсчитаем число предприятий; среднегодовую стоимость основных производственных фондов всего и в среднем на одно предприятие; прибыль всего и в среднем на одно предприятие. Результаты расчетов представим в виде групповой таблицы и сделаем выводы.

Таблица 2.6 – Распределение предприятий по среднегодовой стоимости основных производственных фондов и прибыли

Предприятие	Среднегодовая стоимость основных производственных фондов, млн.руб.	Прибыль, млн.руб.
А	45,0	55,0
Б	55,5	70,0
В	13,6	25,0
Г	65,7	86,0
Д	110,9	165,5
Ж	70,4	80,0
З	30,3	40,0
К	80,5	128,0
Л	115,5	160,0
М	15,9	32,0

Решение

Определяем шаг интервала: $i = \frac{x_{max} - x_{min}}{m}$,

Число групп- 4, x_{max}= 115,5 млн.руб.; x_{min}=13,6 млн.руб.

i= (115,5-13,6)/4 = 25,475 млн.руб.

Формируем интервалы:

Интервалы	Предприятия	Количество предприятий
(13,6; 39,075)	В, З, М	3
(39,075; 64,55)	А, Б	2
(64,55; 90,025)	Г, Ж, К	3
(90,025; 115,5)	Д, Л	2

Таблица 2.7 – Аналитическая группировка предприятий по среднегодовой стоимости основных производственных фондов и прибыли

Интервалы	Количество предприятий	Среднегодовая стоимость основных производственных фондов, млн.руб.		Прибыль, млн.руб.	
		В сумме	В среднем на одно предприятие	В сумме	В среднем на одно предприятие
(13,6; 39,075)	3	59,8	19,9	97	32,3
(39,075; 64,55)	2	100,5	50,25	125	62,5
(64,55; 90,025)	3	216,6	72,2	294	98
(90,025; 115,5)	2	226,4	113,2	325,5	162,75

Вывод: Таким образом, на основе полученных расчетных данных можно сделать вывод, что существует определенная взаимосвязь между величиной среднегодовой стоимостью основных производственных фондов и прибылью. С увеличением стоимости основных производственных фондов увеличивается и прибыль предприятий. Связь между признаками прямая.

Глава 3. СТАТИСТИЧЕСКОЕ ИССЛЕДОВАНИЕ РЯДОВ ДИНАМИКИ

Одним из наиболее точных и содержательных методов изучения социально-экономических явлений и процессов в настоящее время является метод анализа временных рядов. «Временной ряд – это последовательность упорядоченных во времени числовых показателей, характеризующих уровень состояния и изменения изучаемого явления»[1].

Для всестороннего и полного изучения временного ряда в первую очередь необходимо проанализировать изменения явления во времени, т.е. изучить динамику процесса. Далее необходимо проверить гипотезу о наличии тенденции развития, выбрать лучшее уравнение тренда, проверить адекватность выбранной модели. Затем сделать прогноз по выбранному уравнению тренда в пределах доверительных интервалов.

Для анализа динамики, и чтобы построить систему показателей динамики, необходимо рассчитать абсолютные и относительные показатели динамики. Анализ скорости и интенсивности развития явления во времени осуществляется с помощью статистических показателей, которые получаются в результате сравнения уровней между собой. К таким показателям относятся: абсолютный прирост, темп роста, темп прироста, абсолютное содержание 1% прироста, абсолютное ускорение, средний уровень ряда, средний темп роста, средний темп прироста, средний абсолютный прирост. Расчет показателей можно вести двумя способами: по базисной системе сравнения и по цепной системе сравнения[2].

Рассмотрим эти показатели подробнее:

1. Абсолютный прирост (Δ) характеризует размер увеличения (или уменьшения) уровня ряда за определенный промежуток времени. Он определяется как разница между любым уровнем ряда и уровнем, принятым за базу сравнения, т.е. с начальным уровнем – первым показателем ряда.

$$\Delta = y_i - y_{i-1}, \textit{или } \Delta' = y_i - y_k \qquad (3.1)$$

где i=1,2,3,…,n.

Δ – *цепной*

Δ' – *базисный*

2. Показатель интенсивности изменения уровня ряда в зависимости от того, выражается ли он в виде коэффициента или в процентах, принято называть коэффициентом роста (Кр) или темпом роста (Тр). Коэффициент роста показывает, во сколько раз данный уровень ряда больше базисного уровня (если он больше 1) или какую часть базисного уровня составляет уровень текущего периода за некоторый промежуток времени (если он меньше

[1] Дуброва Т.А.Статистические методы прогнозирования. – М. Юнити, 2003. – 206 с.

[2] Теория статистики / Под ред. Р. А. Шмойловой. – М.: Финансы и статистика, 2002. – 560 с.

1). В качестве базисного уровня может приниматься начальный уровень ряда при базисной системе либо для каждого последующего предшествующий ему при цепной системе.

$$K_p = \frac{y_i}{y_{i-1}}, или\ K'_p = \frac{y_i}{y_k};$$
$$T_p = K_p \cdot 100, или\ T'_p = K'_p \cdot 100. \quad (3.2)$$

3. Темп прироста (Тп) характеризует относительную скорость изменения уровня ряда в единицу времени и показывает, на какую долю (или процент) уровень данного периода или момента времени больше (или меньше) базисного уровня. Темп прироста рассчитывается как отношение абсолютного прироста к уровню ряда, принятого за базу:

$$T_n = T_p - 100, или\ T'_n = T'_p - 100. \quad (3.3)$$

4. Абсолютное значение 1% прироста (А) представляет собой сотую часть базисного уровня и в то же время отношение абсолютного прироста к соответствующему темпу прироста. Он показывает абсолютный рост явления при росте явления на 1%:

$$A = \frac{\Delta}{T_n}, илиA' = \frac{\Delta'}{T'_n}. \quad (3.4)$$

5. Абсолютное ускорение – это разность между последующим и предыдущим абсолютным приростами (Δ′=Δi-Δi-1) Ускорение показывает, насколько данная скорость больше (меньше) предыдущей.

6. Средний уровень ряда динамики ($\vec{o}$) – определяется по базисной системе как частное от деления суммы начального и конечного уровней на 2, а по цепной системе по средней арифметической простой.

$$\vec{o} = \frac{\sum y}{n}. \quad (3.5)$$

7. Средний абсолютный прирост ($\overline{\Delta}$) дает возможность установить, насколько в среднем за единицу времени должен увеличиться уровень ряда (в абсолютном выражении), чтобы, отправляясь от начального уровня за данное число периодов, достигнуть конечного уровня.

$$\overline{\Delta} = \frac{\sum \Delta}{n-1}. \quad (3.6)$$

8. Средний темп роста ($\overline{T}_p$) показывает, во сколько раз в среднем за единицу времени изменился уровень динамического ряда.

$$\overline{T}_p = \sqrt[n-1]{\frac{y_n}{y_1}} \cdot 100 . \qquad (3.7)$$

9. Средний темп прироста ($\overline{T}_n$) определяется как разница между средним темпом роста и единицей.

$$\overline{T}_n = \overline{T}_p - 100 . \qquad (3.8)$$

10. Средняя величина абсолютного значения 1% прироста ($\overline{A}$) определяется как:

$$\overline{A} = \frac{\overline{\Delta}}{\overline{T}_n} . \qquad (3.9)$$

Для выявления факта наличия или отсутствия неслучайной составляющей $f(x)$, то есть для проверки гипотезы о существовании тренда - $H_0 : E_y(t) = a = const$ можно использовать различные критерии.

Например:

1) критерии серии, который имеет две модификации
   - критерий серий, основанный на медиане выборки;
   - критерий «восходящих - нисходящих» серий
2) метод Фостера – Стюарта и др.

Рассмотрим более подробно каждый из них.

Критерий серий, основанный на медиане выборки[1].

Для этого необходимо:

1. из исходного ряда с уровнями: $y_1, y_2, ..., y_n$ образовать ранжированный ряд: $y'_1, y'_2, ..., y'_n$, где y'_1 - наименьшее значение уровней исходного ряда;

2. определить медиану Me этого вариационного ряда. В случае нечетного ряда $(n = 2m+1) - Me = y'_{m+1}$, в противном случае $(n = 2m)$, $Me = \frac{y'_m + y'_{m+1}}{2}$.

3. образовать последовательность δ, из «+» и «-» по следующему правилу: $\delta_i = \begin{cases} "+", y_i > Me \\ "-", y_i < Me \end{cases}$ для всех t $\forall \in (1, n)$

4. необходимо подсчитать $\nu(n)$ совокупности δ_1, где под серией понимается последовательность подряд идущих «+» и «-». Один «+» или один

[1] Дуброва Т.А.Статистические методы прогнозирования. – М. Юнити, 2003. – 206 с.

«-» тоже считается серией. Определяем $\tau_{\max}(n)$ - протяженность самой длинной серии.

5. проверка гипотезы основывается на том, что при условии случайности ряда (при отсутствии систематической составляющей) протяженность самой длинной серии не должно быть слишком большой, а общее число серии слишком маленьким. Поэтому, для того, чтобы не была отвергнута гипотеза о случайности исходного ряда, должны выполняться следующие неравенства:

$$\nu(n) > \left[\frac{1}{2}(n+1-1.96\sqrt{n-1})\right] \tag{3.10}$$

$$\tau_{\max}(n) < \left[1.43\ln(n+1)\right] \tag{3.11}$$

Если хотя бы одно из неравенств нарушается, то гипотеза отвергается с вероятностью ошибки α. Следовательно, подтверждается наличие зависящей от времени неслучайной составляющей.

Критерий «восходящих - нисходящих» серий[1].

1. образуется последовательность знаков – плюсов и минусов по следующему принципу:

$$\delta_i = \begin{cases} +, \text{ если } y_{t+1} - y_t > 0 \text{ если } t - 1,2,...,n-1 \\ -, \text{ если } y_{t+1} - y_t < 0 \text{ для } t = 1,2,...,n-1 \end{cases}$$

В случае если последующее наблюдение окажется равным предыдущему, учитывается только одно наблюдение. Таким образом, элементы этой последовательности принимают значение «+», если последующее значение уровня ряда больше предыдущего, и «-« - если меньше. Общее число знаков «+» и «-» заранее не известно. Индекс *i* может принимать значение *1, 2, … , к,* где $к \le n-1$.

2. подсчитывается общее число серий $\nu(n)$ и протяженность самой длинной серии $\tau_{\max}(n)$ аналогично тому, как это делалось в предыдущем варианте критерия. Очевидно, что при этом каждая серия, состоящая из плюсов, соответствует возрастанию уровней ряда («восходящая» серия), а последовательность минусов – их убыванию («нисходящая» серия).

3. для того, чтобы не была отвергнута гипотеза, должны выполняться следующие неравенства (при уровне значимости α, заключенном между 0,05 и 0,0975):

$$\nu(n) > \left[\frac{1}{3}(2n+1) - 1.96\sqrt{\frac{16n-29}{90}}\right] \tag{3.12}$$

[1] Дуброва Т.А.Статистические методы прогнозирования. – М. Юнити, 2003. – 206 с.

$$\tau_{max}(n) < \tau_0(n) \qquad (3.13)$$

где $\tau_0(n)$ - табличное значение, зависящее от длины временного ряда (таблица 3.1).

Таблица 3.1 – Значение $\tau_0(n)$ для критерия «восходящих – нисходящих» серий

n	n≤26	26 < n ≤ 153	153 < n ≤ 1170
$\tau_0(n)$	5	6	7

Если хотя бы одно неравенство нарушается, то нулевая гипотеза отвергается (следовательно, подтверждается наличие зависящей от времени неслучайной составляющей).

Метод Фостера – Стюарта[1]**.**

Алгоритм метода выглядит следующим образом:

1. сравнивается каждый уровень ряда со всеми предыдущими, при этом

если $y_i > y_{i-1}$, U=1, e=0 и наоборот

2. вычисляются значения величин S, d:

$$d = \sum d_i \qquad S = \sum S_i \qquad (3.14)$$

где $S_i = U_i + e_i$

$d_i = U_i - e_i$

Оба показателя асимптотически нормальны и имеют независимые распределения

3. проверка производится с использованием t-критерия Стьюдента гипотеза о том, можно ли считать случайным разности S-μ и d-0

$$t_S = \frac{S - \mu}{\sigma_1} \qquad t_d = \frac{d - 0}{\sigma_2} \qquad (3.15)$$

значения μ и σ табулированы и приведены в специальных таблицах.

4. сравниваются расчетные значения t с табличными при заданном уровне значимости. Если t_s , t_d < t табличного, то гипотеза об отсутствии тренда с средней и дисперсии подтверждается.

Кроме рассмотренных подходов определения тенденции во временном ряду используется также критерий квадратов последовательных разностей (метод Аббе), метод проверки разностей средних уровней и др.

При наличии тенденции в ряду динамики его уровни можно рассматривать как функцию времени.

Показатели силы и интенсивности колебаний аналогичны по построению, по форме показателям силы и интенсивности вариации признака в

[1] Теория статистики / Под ред. Р. А. Шмойловой. – М.: Финансы и статистика, 2002. – 560 с.

пространственной совокупности. По существу они отличаются тем, что показатели вариации вычисляются на основе отклонений от постоянной средней величины, а показатели, характеризующие колеблемость уровней временного ряда, - по отклонениям отдельных уровней от тренда, который можно считать «подвижной средней величиной»[1].

Абсолютные показатели колеблемости:

1. Амплитуда (размах) колебаний: $A = Y_{max} - Y_{min}$. (3.16)

Это разность между наибольшим и наименьшим по алгебраической величине отклонений от тренда.

2. Средняя по модулю отклонений от тренда: $\alpha(t) = \frac{\sum_{i=1}^{n} |y_i - \hat{y}_i|}{n}$. (3.17)

3. Среднее квадратическое отклонение уровней ряда от тренда:

$$S_y(t) = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{\hat{y}}_i)^2}{n - p}}, \quad (3.18)$$

где y_i - фактический уровень;

$\hat{\hat{y}}_i$ - выровненный уровень;

n -число уровней;

p - число параметров тренда;

t - номера лет (знак отклонения от тренда).

Относительные показатели:

$$V(t) = \frac{S(t)}{\bar{y}} * 100, \quad (3.19)$$

где $\bar{y}$ - средний уровень ряда.

Этот показатель отражает величину колеблемости в сравнении со средним уровнем ряда. Он необходим для сравнения двух различных явлений и чаще всего выражаются в %.

Устойчивость временного ряда – это наличие необходимой тенденции изучаемого статистического показателя с минимальным влиянием на него неблагоприятных условий.

Из этого вытекают основные требования устойчивости:

1) минимизация колебаний уровней временного ряда;

2) наличие определенной, необходимой для общества тенденции изменения [2].

Устойчивость временного ряда можно оценить на различных явлениях. При этом в зависимости от явления будет меняться показатели, которые

[1] Г.Л. Громыко. Статистика: Учебник. — Т11 2-е изд., перераб. и доп. - М.: ИНФРА-М,. - 476 с. — (Классический университетский учебник)., 2005

[2] Г.Л. Громыко. Статистика: Учебник. — Т11 2-е изд., перераб. и доп. - М.: ИНФРА-М,. - 476 с. — (Классический университетский учебник)., 2005

используются в качестве форм выражения существа исследуемого процесса, но содержание понятия устойчивости будет оставаться неизменным.

Наиболее простым, аналогичным процессу вариации при изменении устойчивости уровней временного ряда, является размах колеблемости средних уровней за благоприятный и неблагоприятные, в отношении к изучаемому явлению, периоды времени:

$$R_{\bar{y}} = \bar{Y}_{áëàã} - \bar{Y}_{íåáëàã}. \qquad (3.20)$$

К благоприятным периодам времени относятся все периоды с уровнями выше тренда, к неблагоприятным – ниже тренда (однако, например, при изучении динамики производительности труда если это трудоемкость, то все должно быть наоборот)[1].

Отношение средних уровней за благоприятные периоды времени к средним уровням за неблагоприятные также может служить показателем устойчивости уровней. Чем ближе отношение к единице, тем меньше колеблемости и соответственно выше устойчивости. Назовем это отношение индексом устойчивости динамических рядов:

$$i_{\bar{y}} = \frac{\bar{y}_{благ}}{\bar{y}_{неблаг}}. \qquad (3.21)$$

При изменении колеблемости уровней исчисляются обобщающие показатели отклонений уровней от тренда за исследуемый период.

Основным абсолютным показателем является среднее линейное отклонение:

$$a(t) = \frac{\sum_{i=1}^{n} |y_i - \tilde{y}_i|}{n - p}. \qquad (3.22)$$

Он выражается в единицах измерения анализируемых уровней и не может, служить для сравнения колебаний различных динамических рядов.

Величину (3.23) называют коэффициентом устойчивости:

$$K_y = (100 - V_y(t)). \qquad (3.23)$$

Такое определение коэффициента устойчивости интерпретируется как обеспечение устойчивости уровней ряда относительно лишь в $(100 - V_y(t))$ случаях.

Наиболее простым показателем устойчивости тенденции временного ряда является коэффициент (критерий) Спирмена:

[1] Г.Л. Громыко. Статистика: Учебник. — Т11 2-е изд., перераб. и доп. - М.: ИНФРА-М,. - 476 с. — (Классический университетский учебник)., 2005

$$Kp = 1 - \frac{6\sum_{i=1}^{n} d^2}{n^3 - n}, \qquad (3.24)$$

где d - разность рангов уровней изучаемого ряда (P_y) и рангов номеров периодов или момент времени в ряду (P_t);

n – число таких периодов или моментов.

Для определения коэффициента Спирмена величины уровней изучаемого явления y_i нумеруются в порядке возрастания, а при наличии одинаковых уровней им присваивается определенный ранг, равный частному от деления суммы рангов, приходящихся на эти значения, на число этих равных значений.

Коэффициент рангов периодов времени и уровней динамического ряда может применить значения в пределах от 0 до ± 1.

Интерпретация этого коэффициента такова: если каждый уровень ряда исследуемого периода выше, чем предыдущего, то ранги уровней ряда и номера лет совпадают, $Kp = +1$. Это означает полную устойчивость самого факта роста уровней ряда, непрерывность роста.

Чем ближе Kp к +1, тем ближе рост уровней к непрерывному, выше устойчивость роста. При $Kp = 0$ рост совершенно неустойчив. При отрицательных значениях чем ближе Kp к -1, тем устойчивее снижение изучаемого показателя.

В качестве характеристики устойчивости изменения можно применить индекс корреляции:

$$J_r = \sqrt{1 - \frac{\sum (y_i - \tilde{y}_i)^2}{\sum (y - \bar{y})^2}}, \qquad (3.25)$$

где y_i - уровни динамического ряда;

$\bar{y}$ - средний уровень ряда;

$\tilde{y}_i$ - теоретические уровни ряда.

Индекс корреляции показывает степень сопряженности колебаний исследуемых показателей с совокупностью факторов, изменяющих во времени. Приближение индекса корреляции к 1 означает большую устойчивость изменения уровней динамического ряда. Сравнение индексов корреляции по разным показателям возможно лишь при условии равенства числа уровней.

При наличии тенденции в ряду динамики его уровни можно рассматривать как функцию времени.

Существует несколько методов облегчающих процесс выбора формы кривой роста. Наиболее простой метод – визуальный, опирающийся на графическое изображении.

Проверка адекватности выбранных моделей реальному процессу строится при анализе случайных компонент. Ряд остатков как отклонение физических уровней от выровненных:

$$e_t = y_t - \hat{y}_t . \qquad (3.26)$$

Принято считать, что модель адекватно описываемому процессу, если значение остаточной компоненты подчиняется случайному закону распределения.

Если вид функции выбран неудачно, то последовательные значения ряда остатков могут не обладать свойствами независимости, так как они могут коррелировать между собой, т.е. будет иметь место автокорреляции ошибок. Существует несколько методов обнаружения автокорреляции. Наиболее распространен критерий Дарбина – Уотсона:

$$d = \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^{\,2}} \approx 2(1 - r_t), \qquad (3.27)$$

где r_t - коэффициент автокорреляции первого порядка.

Если в ряду остатков имеется сильная положительная автокорреляция, то d=0. Если сильная отрицательная, то d=4. При отсутствии автокорреляции d=2. Применение на практике данного критерия основано на сравнении величины d с теоретическими табулированными значениями d_1 и d_2.

Если $d \langle d_1$, то нулевая гипотеза о независимости случайных отклонений отвергается (отсутствие автокорреляции).

Если $d > d_2$, то нулевая гипотеза не отвергается.

Если $d_1 \le d \le d_2$, нет достаточных оснований для принятия решений.

Когда расчетное значение d превышает 2, то с d_1 и d_2 сравнивается не сам коэффициент d, а $(4 - d)$.

Если тенденция в динамике существует, то переходим к выбору кривой роста, наиболее точно описывающей процесс изменения в исследуемом явлении. После выявления тенденции приступаем к выбору лучшего уравнения тренда. Лучшим считается то уравнение тренда, в котором коэффициент аппроксимации очень близок к единице.

Самым простым типом тренда является прямая линия, описываемая линейным уравнением тренда:

$$\tilde{y}_i = a + b \cdot t_i , \qquad (24)$$

где $\tilde{y}_i$ - выровненные, т.е. лишенные колебаний, уровни тренда для лет с номером i;

a - свободный член уравнения, численно равный среднему выровненному уровню для момента или периода времени, принятого за начало отсчета;

b - средняя величина изменения уровней ряда за единицу изменения времени.

Помимо прямолинейного тренда в статистике при анализе временных рядов также могут использовать параболический, экспоненциальный, логарифмический, гиперболический и логистический тренды.

Важнейшими характеристическими качествами модели, выбранной для прогнозирования, являются показатели её точности. Одним из таких показателей является средняя относительная ошибка по модулю:

$$|\delta| = \frac{1}{n}\sum\left|\frac{\hat{y}_t - y_t}{y_t}\right| * 100\% . \qquad (3.28)$$

Если $|\delta|$ < 10%, то это свидетельствует о высокой точности модели;

Если 10% <$|\delta|$< 20%, это говорит о хорошей точности данной модели;

Если 20% <$|\delta|$< 50% - точность модели удовлетворительная.

После проведенного предварительного анализа переходим к прогнозированию простой трендовой модели. Простая трендовая модель динамики – это уравнение тренда с указанием начала отсчета единиц времени. Прогноз по этой модели заключается в подстановке в уравнение тренда номера периода, который прогнозируется. Такой прогноз называется точечным. Точечный прогноз – «это скорее абстракция, чем реальность», поэтому необходимо также делать интервальный прогноз. При таком прогнозе учитываются как вызванная колеблемостью ошибка репрезентативности выборочной оценки тренда, так и колебания уровней в отдельные периоды (моменты) относительно тренда.

Прогноз должен иметь вероятностный характер. Для этого вычисляется средняя ошибка прогноза положения тренда на год за номером t_k, обозначаемая m_y по формуле:

$$m_{\sigma_{\bar{y}_t}} = \sigma_{\bar{y}_t} \cdot \sqrt{\frac{1}{n} + \frac{t_k^{\,2}}{\sum t^2{}_i}} \ , \qquad (3.29)$$

где

$$\sigma_{\bar{y}_t} = \sqrt{\frac{\sum (y - \bar{y}_t)^2}{n - p}} . \qquad (3.30)$$

Источниками неопределенности, т.е. ошибки прогноза конкретного уровня, являются: во-первых, неопределенность положения тренда на прогнозируемое время и, во-вторых, колебания конкретных уровней относительно тренда.

Для вычисления доверительного интервала прогноза положения тренда среднюю ошибку необходимо умножить на величину t-критерия Стьюдента при имеющемся числе степеней свободы колебаний и при выбранной вероятности (надежности прогноза). Величина доверительных интервалов определяется:

$$\bar{y}_t +_- t_\alpha \cdot m_{\sigma \bar{y}_t}, \qquad (3.31)$$

где t_α - доверительная величина (надежностью 95%) и (n-1)- степенями свободы.

Прогнозирование по тренду имеет качественное ограничение: оно допустимо в условиях сохранения основной тенденции.

Под сезонностью понимается устойчивая закономерность внутригодовой динамики того или иного явления, которая проявляется во внутригодовых повышениях или понижениях уровней того или иного показателя на протяжении ряда лет.

Многие процессы имеет ярко выраженную зависимость от сезонных колебаний.

Изучение сезонности позволяет:

- ✓ определить степень влияния природно-климатических условий на формирование изучаемого потока;
- ✓ установить продолжительность сезона;
- ✓ раскрыть факторы, обусловливающие сезонность в; определить экономические последствия сезонности на уровне региона и фирмы;
- ✓ разработать комплекс мероприятий по снижению сезонной неравномерности.

Сезонность определяется целым рядом природно-климатических, экономических, социальных, демографических, психологических, материально-технических, технологических факторов.

В ходе исследования сезонности решаются следующие задачи:

1) определяется наличие сезонности, численное выражение проявления сезонных колебаний и выявляется их сила и характер в различных фазах годичного цикла;

2) дается характеристика факторов, вызывающих сезонные колебания;

3) оцениваются последствия, к которым приводит наличие сезонных колебаний;

4) осуществляется математическое моделирование сезонности.

Для измерения сезонных колебаний статистикой предложены различные методы, среди которых чаще всего используются:

а) метод абсолютных разностей;

б) метод относительных разностей;

в) построение индексов сезонности.

Первые два способа предполагают нахождение разностей фактических уровней и уровней, найденных при выявлении основной тенденции развития.

Применяя способ абсолютных разностей, оперируют непосредственно размерами этих разностей, а при использовании метода относительных разностей определяют отношение абсолютных размеров указанных разностей к выравненному уровню. При выявлении основной тенденции используют либо метод скользящей средней, либо аналитическое выравнивание.

Определение степени устойчивости (вариации) показателей во времени может быть осуществлено с помощью методов анализа рядов динамики и сезонности.

Характер общей тенденции ряда динамики определяется по графическому изображению ряда динамики и расчету показателей динамики:

1) базисного темпа роста:

$$T_{p_б} = \frac{y_i}{y_o} \times 100\% \qquad (3.32)$$

где y_i - уровень i-го периода (кроме первого); y_0 - уровень базисного периода.

2) среднегодового темпа роста

$$\overline{T}_p = \sqrt[n-1]{\frac{y_n}{y_0}} \cdot 100\,\% \qquad (3.33)$$

3) среднегодового темпа прироста:

$$\overline{T}_{np} = \overline{T}_p - 100\% \qquad (3.34)$$

Индекс сезонности определяется по формуле:

$$i_c = \frac{\bar{y}_n}{\bar{y}} \cdot 100\%, \qquad (3.35)$$

где $\bar{y}_n$ - средний за 3 года уровень n-ого месяца, $\bar{y}$ - средний уровень показателя по всей длине ряда.

При исследовании явлений периодического типа в качестве аналитической формы развития во времени принимается уравнение следующего типа (ряд Фурье):

$$\hat{y}_t = a_0 + \sum (a_k \cos kt + b_k \sin kt) \qquad (3.36)$$

В этом уравнении величина *k* определяет гармонику ряда Фурье и может быть взята с разной степенью точности (чаще всего от 1 до 4). Для отыскания параметров уравнения используется метод наименьших квадратов, т.е.

$$\sum_1^n (y_i - \hat{y}_t)^2 = \min \qquad (3.37)$$

Найдя частные производные этой функции и приравняв их к нулю, получим систему нормальных уравнений, решение которой дает следующие формулы для вычисления параметров:

$$a_0 = \frac{1}{n} \sum_1^n y_i \qquad (3.38)$$

$$a_k = \frac{2}{n} \sum_1^n y_i \cos kt \qquad (3.39)$$

$$b_k = \frac{2}{n}\sum_{1}^{n} y_i \sin kt \qquad (3.40)$$

Параметры уравнения зависят от значений y и связанных с ними последовательных значений cos *kt* и sin *kt.*

Для изучения сезонных колебаний на протяжении года необходимо взять n=12 (по числу месяцев в году). Тогда, представляя периоды как части длины окружности, ряд динамики можно записать в виде таблицы, в первой строке которой будут записаны периоды, а во второй – соответствующие им уровни.

Применяя к этим же данным вторую гармонику ряда Фурье для выражения модели сезонности, получим коэффициенты a_2 и b_2. Подставляя их в уравнение ряда Фурье, будем иметь следующую модель сезонности данного ряда динамики:

$$\hat{y}_t = a_0 + a_1 \cos t + b_1 \sin t + a_2 \cos 2t + b_2 \sin 2t . \qquad (3.41)$$

Простейший подход к моделированию сезонных колебаний это расчет значений сезонной компоненты методом скользящей средней и построение аддитивной или мультипликативной модели временного ряда.

Построение аддитивной и мультипликативной моделей сводится к расчету значений u_t, s_t, ε_t.

Алгоритм построения тренд – сезонной аддитивной модели

1. сглаживание (выравнивание) временного ряда с помощью простой скользящей средней. Период скольжения должен быть равен (периодичности сезонных колебаний) 1 году.

Для этого:

а) просуммируем уровни ряда последовательно за каждые четыре квартала (12 месяцев) со сдвигом на один момент времени и определим условные годовые объемы;

б) разделив полученные суммы на 4 (12), найдем скользящие средние. Отметим, что полученные таким образом выровненные значения не содержат сезонной компоненты;

в) если период четный, то проводится центрирование скользящей средней, для чего определяется среднее значение из двух последовательных скользящих средних;

2. рассчитывают абсолютные показатели сезонности:

$$S_i = y_i - \tilde{y} \qquad (3.42)$$

где $\tilde{y}$ - выровненные скользящие средние;

3. рассчитываются средние показатели сезонности для одноименных кварталов (месяцев):

$$\bar{S}_i = \frac{1}{n}\sum S_i \qquad (3.43)$$

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.

Если $\sum \bar{S}_i \neq 0$, проводится корректировка сезонной компоненты:

$$\hat{S}_i = \bar{S}_i - \frac{1}{n}\sum_{i=1}^{n} \bar{S}_i \qquad (3.44)$$

4. проводим десезоналирование временного ряда или элиминируем влияние сезонной компоненты: из исходных уровней вычитаем скорректированную сезонную компоненту:

$$y_i - \hat{S}_i \qquad (3.45)$$

5. по десезоналированному временному ряду проводим аналитическое выравнивание (определяем компоненту u_t);

6. рассчитываем тренд с учетом сезонности:

$$y_S = \hat{y}_t + \hat{S}_i . \qquad (3.46)$$

При мультипликативной модели временной ряд можно представить в виде сомножителей:

$$y_i = \hat{y}_t \cdot K_S \cdot E \qquad (3.47)$$

где K_S - коэффициент сезонности; E – коэффициент влияния случайности $\left(\frac{y_i}{y_S}\right)$.

Алгоритм построения тренд – сезонной мультипликативной модели

1. сглаживание временного ряда с помощью скользящей средней;
2. рассчитываем коэффициент сезонности

$$K_S = \frac{y_i}{\tilde{y}_i} \qquad (3.48)$$

3. определяем средние показатели сезонности для одноименных кварталов (месяцев):

$$\bar{K}_j = \frac{1}{n}\sum K_{Si} \qquad (3.49)$$

4. если при поквартальном наблюдении $\sum \bar{K} \neq 4$, а при помесячном $\sum \bar{K} \neq 12$, то выполняется корректировка коэффициента сезонности:

$$\hat{K}_j = \bar{K}_j \cdot \frac{4(12)}{\sum \bar{K}_j} \qquad (3.50)$$

5. исключаем сезонность из уровней ряда:

$$\frac{y_i}{\hat{K}_j}; \qquad (3.51)$$

6. проводится аналитическое выравнивание десезоналированного ряда;
7. рассчитываются уровни временного ряда, обусловленные влиянием тенденции и сезонности:

$$y_S = \hat{y}_t \cdot \hat{K}_j . \qquad (3.52)$$

Аддитивная модель целесообразна, если размах сезонных колебаний изменяется слабо.

<u>Пример 3.1. Расчет средних показателей ряда динамики, построение тренда, расчет критерия Дарбина-Уотсона.</u>

Для анализа численности населения Российской Федерации за 2009-2021гг. рассчитаем показатели интенсивности и скорости динамических рядов по базисной и цепной системе (таблица 3.2).

Таблица 3.2 - Расчет показателей динамики численности населения Российской Федерации (условные данные)

Годы	Численность населения, млн. чел.	Абсолютный прирост (убыль), млн.чел.		Темп изменения, %		Темп прироста (убыли), %		Абсолютное значение 1% прироста, млн. чел.
		базисный	цепной	базисный	цепной	базисный	цепной	
2009	148,3	-	-	-	-	-	-	-
2010	148,3	0,0	0,0	100	100	0	0	1,483
2011	146,3	-2,0	-2,0	98,65	98,65	-1,35	-1,35	1,483
2012	145,2	-3,1	-1,1	97,91	99,25	-2,09	-0,75	1,463
2013	145,0	-3,3	-0,2	97,77	99,86	-2,23	-0,14	1,452
2014	144,3	-4,0	-0,7	97,30	99,52	-2,70	-0,48	1,45
2015	143,8	-4,5	-0,5	96,97	99,65	-3,03	-0,35	1,443
2016	143,2	-5,1	-0,6	96,56	99,58	-3,44	-0,42	1,438
2017	142,8	-5,5	-0,4	96,29	99,72	-3,71	-0,28	1,432
2018	142,8	-5,5	0,0	96,29	100,00	-3,71	0,00	1,428
2019	142,7	-5,6	-0,1	96,22	99,93	-3,78	-0,07	1,428
2020	142,9	-5,4	0,2	96,36	100,14	-3,64	0,14	1,427
2021	142,9	-5,4	0,0	96,36	100,00	-3,64	0,00	1,429

Из таблицы 3.2 видно, что в течение всего анализируемого периода наблюдается снижение численности населения Российской Федерации по сравнении с 2009г. При расчет по цепной системе следует отметить не значительное увеличение в 2020г. и 2021г. по сравнению с предыдущими годами.

Для характеристики интенсивности численности населения Российской Федерации за период 2009-2021гг. рассчитываются средние показатели динамики:

1. Средний уровень интервального ряда ($\bar{y}$): $\bar{y}$=144,5 млн.чел.

Он показывает, что в среднем за период с 2009 по 2021гг. численность населения Российской Федерации составила 144,5 млн. чел.

2. Средний абсолютный прирост (убыль) ($\bar{\Delta}$): $\bar{\Delta}$ = -0,45 млн. чел.

Таким образом, в среднем за анализируемый период численность населения Российской Федерации за 2009-2021гг. уменьшался на 0,45 млн.чел.;

3. Средний темп изменения ($\overline{Tp}$): $\overline{Tp}$ =99,88%

Средний темп роста составляет 98,88% и является вспомогательным для определения следующего показателя, темпа прироста.

4. Средний темп прироста ($\overline{Tnp}$): $\overline{Tnp}$ =-0,12%

В среднем за период темп убыли численности населения Российской Федерации составляет 0,12%.

5. Средняя величина абсолютного значения 1% прироста($\overline{A}$): $\overline{A}$ =-3,75 млн. чел.

В среднем величина абсолютного значения 1% убыли за период с 2009 по 2021гг. по Российской Федерации составила 3,75 млн. чел.

Прежде чем переходить к определению тенденции и выделению тренда необходимо выяснить существует ли тенденция вообще в динамике численности населения Российской Федерации. Для этого можно воспользоваться наиболее часто используемым на практике методом проверки наличия тренда – критерием «восходящих - нисходящих» серий, основанного на проверке гипотезы «наличия – отсутствия» тренда (таблица 3.3).

Таблица 3.3 – Вспомогательная таблица для проверки гипотезы существенности ряда динамики численности населения Российской Федерации

Годы	y_t	y'_t	Серии
2009	148,3	5,6	+
2010	148,3	5,5	+
2011	146,3	3,5	+
2012	145,2	2,3	+
2013	145,0	2,1	+
2014	144,3	1,1	+
2015	143,8	0,0	+
2016	143,2	-1,1	-
2017	142,8	-2,2	-
2018	142,8	-2,4	-
2019	142,7	-3,6	-
2020	142,9	-5,4	-
2021	142,9	-5,4	-

Таким образом, пользуясь данными таблицы 3.3, определим число серий и их максимальное значение.

Число серий определяется путем подсчета: $\upsilon(13)=2$

Определяем протяженность самой длинной серии τ max (13)=7

Проверяем гипотезу о случайности исходного ряда, для этого должны выполняться неравенства. Для того чтобы не была отвергнута гипотеза о случайности исходного временного ряда должны выполняться данные

неравенств. Если хотя бы одно из неравенств нарушается, то тенденция имеет место.

Получаем: $\upsilon(13) > 6{,}6$ и $\tau \max(13) > 2$

Данные неравенства не выполняются, следовательно, гипотеза о случайности исходного ряда отклоняется, значит, тенденция в динамике численности населения Российской Федерации имеется.

Таким образом, используя данные о численности населения Российской Федерации за период с 2009 по 2021гг., построим график динамики численности населения и отобразим на нем тренды развития (рисунок 3.1).

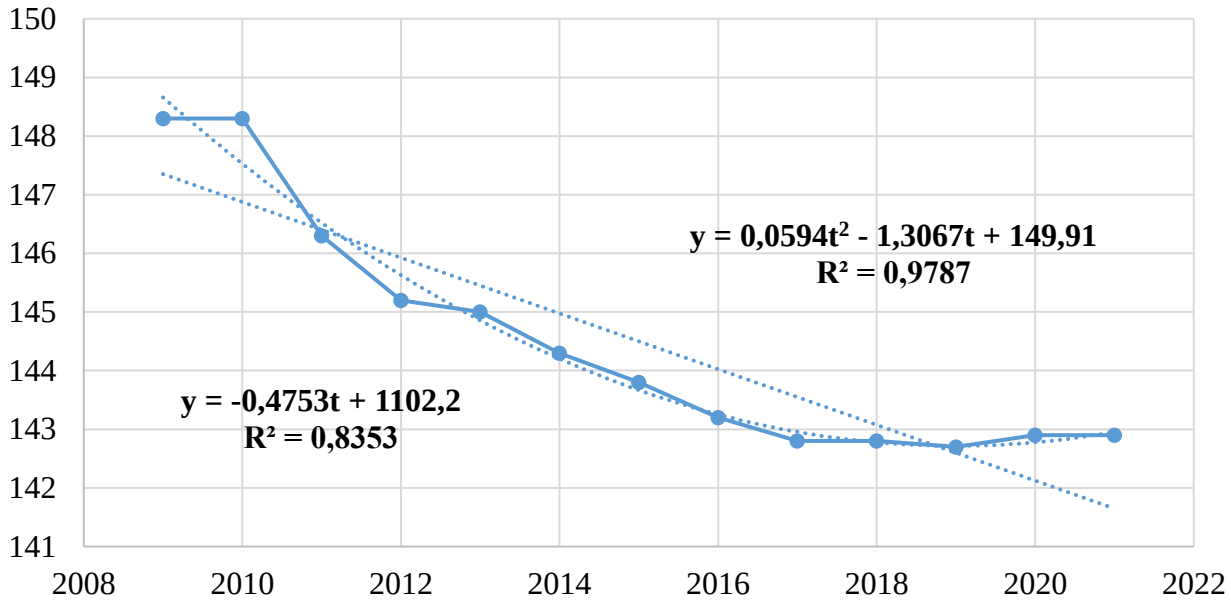


Рисунок 3.1 – Динамика численности населения Российской Федерации

На графике представлено два уравнения тренда - полином второй степени и прямая. Критерием при выборе лучшего уравнения тренда является коэффициент аппроксимации (R^2), чем ближе он к единице, тем точнее данная модель будет описывать исследуемое явление. В нашем случае для численности населения Российской Федерации R^2=0,9887 у полинома второй степени, следовательно, в дальнейших расчетах будем использовать именно его.

На графике изображен параболический тренд второго порядка вида:

$$y_t = 0{,}0594t^2 - 1{,}3067t + 149{,}91 \qquad (3.32)$$

Интерпретация параметров тренда такова: численность населения Российской Федерации в 2009-2021гг. уменьшалась в номинальной оценке ускоренно, со средним ускорением 0,0594 млн.чел. в год, средняя за весь период убыль населения за период 2009-2021 гг. в Российской Федерации составила 1,3067 млн. чел., средняя численность населения была равна 149,91 млн. чел.

Проверяем полученную модель развития на адекватность с помощью критерия Дарбина-Уотсона по формуле (3.27). Проверка адекватности выбранной модели реальному процессу строится на анализе случайной компоненты. Критерий Дарбина-Уотсона связан с гипотезой о существовании автокорреляции первого порядка, то есть автокорреляции между соседними остаточными членами ряда. Расчеты приводятся в таблице 3.4.

Таблица 3.4 - Вспомогательная таблица для расчета критерия Дарбина-Уотсона

Годы	y_i	t	$\hat{\hat{y}}$	e_t	$\lvert e_t \rvert$	e_t^2	$(e_t - e_{t-1})^2$	$\left\lvert \frac{\hat{y} - y_i}{y_i} \right\rvert$
2009	148,3	1	129,46	18,84	18,84	354,91	2555,94	-0,0379
2010	148,3	2	149,64	-1,34	1,34	1,78	407,03	0,0030
2011	146,3	3	107,44	38,86	38,86	1509,87	1615,52	-0,0845
2012	145,2	4	158,88	-13,68	13,68	187,17	2760,23	0,0347
2013	145,0	5	167,45	-22,45	22,45	504,00	76,90	0,0588
2014	144,3	6	170,25	-25,95	25,95	673,45	12,25	0,0674
2015	143,8	7	130,78	13,02	13,02	169,44	1518,50	-0,0295
2016	143,2	8	125,55	17,65	17,65	311,66	21,50	-0,0373
2017	142,8	9	113,44	29,36	29,36	861,95	137,01	-0,0563
2018	142,8	10	137,37	5,43	5,43	29,51	572,47	-0,0100
2019	142,7	11	159,03	-16,33	16,33	266,51	473,40	0,0281
2020	142,9	12	155,01	-12,11	12,11	146,75	17,73	0,0186
2021	142,9	13	129,88	13,02	13,02	169,44	1518,50	-0,0295
Итого	1878,5	-	1878,5	0,02	452,77	16967,51	37411,59	0,0476

Критерий Дарбина – Уотсона d=1,823.

Критерий Дарбина – Уотсона связан с гипотезой о существовании автокорреляции первого порядка, т.е. автокорреляции между соседними остаточными членами ряда. Если d=2, то автокорреляция отсутствует. Определив по специальным таблицам значения верхней и нижней доверительных границ критерия Дарбина – Уотсона, можно принять решение об отсутствии или наличии автокорреляции между соседними остаточными членами. $d_{1=}1,08$ и $d_{2=}1,36$. Так как $d>d_2$, следовательно, автокорреляция между остаточными членами отсутствует и модель адекватна численности населения Российской Федерации.

Следовательно, полученная модель развития численности населения Российской Федерации, можно использовать для дальнейшего прогнозирования и принятия управленческих решений.

Пример 3.2. Расчет средних показателей ряда динамики, построение тренда, расчет средней относительной ошибки прогноза

За последние десять лет наблюдается существенное изменение в динамическом ряду численности обучающихся в Оренбургской области. Проведем анализ динамики уровня образования.

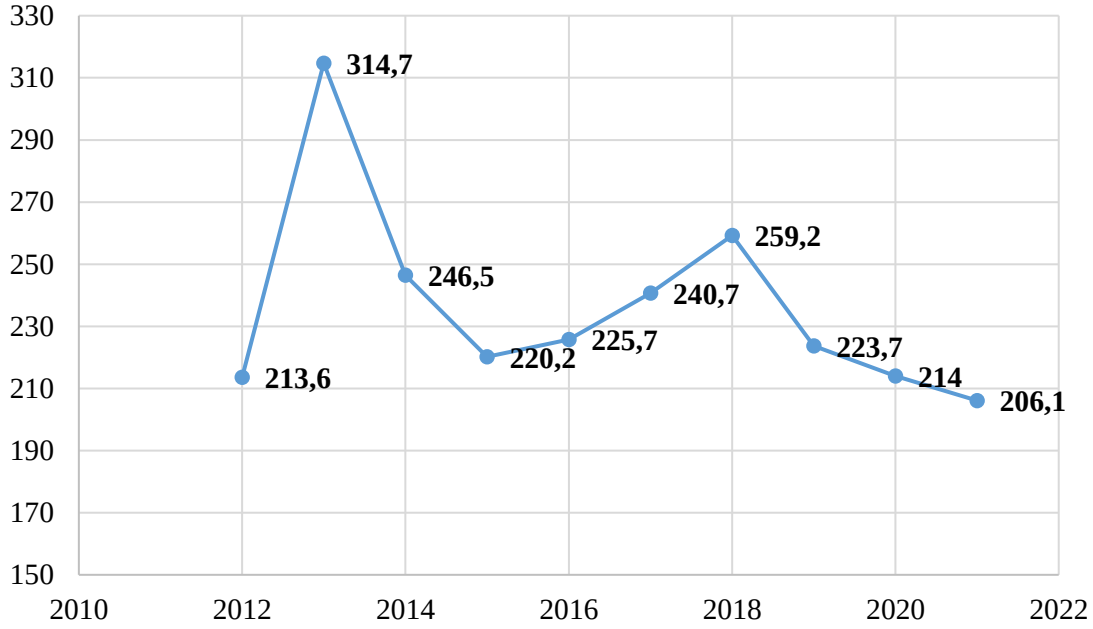


Рисунок 3.2 - Динамика численности обучающегося населения Оренбургской области, тыс. человек

Анализ скорости и интенсивности развития явления во времени осуществляется с помощью статистических показателей, которые получаются в результате сравнения уровней между собой (табл.3.5).

Таблица 3.5 - Динамика уровня образования населения по базисной системе

Годы	Абсолютный прирост (убыль)	Темп роста, %	Темп прироста, %
2012	-	-	-
2013	101,1	147,33	47,33
2014	32,9	115,38	15,38
2015	6,6	103,08	3,08
2016	12,1	105,68	5,68
2017	35,1	116,44	16,44
2018	45,6	121,33	21,33
2019	10,1	104,75	4,75
2020	1,2	100,57	0,57
2021	-7,5	96,48	-3,52

Из таблицы 3.5 видно, что в период с 2012 по 2020г. наблюдается прирост в численности обучающихся в образовательных учреждениях (наибольший в 2013г. – 101,1 тыс.чел. или 47,33%), а затем во все

последующие года прирост все меньше и меньше, а в 2021 по сравнению с 2012 г. идет убыль в размере 7,5 тыс. чел (или 3,52%).

Наряду с темпом роста можно рассчитать показатель темпа прироста, характеризующий относительную скорость изменения уровня ряда в единицу времени. Темп прироста показывает, на какую долю уровень данного периода или момента времени больше (или меньше) базисного уровня.

В статистической практике часто вместо расчета и анализа темпов роста и прироста рассматривают абсолютное значение одного процента прироста. Оно представляет собой одну часть базисного уровня и в то же время – отношение абсолютного прироста к соответствующему темпу роста. Этот показатель дает возможность установить, насколько в среднем за единицу времени должен увеличиваться уровень ряда, чтобы, отправляясь от начального уровня за данное число периодов, достигнуть конечного уровня. Расчет этого показателя имеет экономический смысл только по цепной системе.

Таблица 3.6 - Динамика уровня образования населения по цепной системе

Годы	Абсолютный прирост (убыль)	Темп роста, %	Темп прироста, %	Абсолютное значение 1% прироста
2012	-	-	-	-
2013	101,1	147,3	47,3	2,1
2014	-68,2	78,3	-21,7	3,1
2015	-26,3	89,3	-10,7	2,5
2016	5,6	102,5	2,5	2,2
2017	23,0	110,2	10,2	2,3
2018	10,4	104,2	4,2	2,5
2019	-35,4	86,3	-13,7	2,6
2020	-8,9	96,0	-4,0	2,2
2021	-8,7	95,9	-4,1	2,1

Из таблицы 3.6 видно, что после 2018г. наблюдается значительное уменьшение численности обучающихся: в 2020г. по сравнению с 2019г. сокращение составило 8,9 тыс. чел. (или 4,0 %); в 2021г. по сравнению с 2020г. – 95,9 (или 4,1%).

Особое внимание следует уделять методам расчета средних показателей рядов динамики, которые являются обобщающей характеристикой его абсолютных уровней, абсолютной скорости и интенсивности изменения уровней ряда динамики. Различают следующие показатели динамики: средний уровень ряда динамики, средний абсолютный прирост, средний темп роста и прироста.

В интервальном ряду динамики с равноотстоящими уровнями во времени расчет среднего уровня ряда ($\bar{y}$) производится по формуле средней арифметической простой:

$$\bar{y} = \frac{\sum y}{n} = 237{,}3 \text{ тыс. чел.}$$

Определение среднего абсолютного прироста производится по цепным абсолютным приростам по формуле:

$$\bar{\Delta} = \frac{\sum \Delta_ц}{n-1} = -0{,}83 \text{ тыс. чел.}$$

Среднегодовой темп роста вычисляется по формуле:

$$\bar{T}_p = \sqrt[n-1]{\frac{y_n}{y_0}} = 0{,}9822 \text{ или } 98{,}22\%$$

Среднегодовой темп прироста получим равный -1,78%

На основе рассчитанных показателей динамики можно сказать, что за период с 2012 по 2021 гг. в Оренбургской области наблюдалось сокращение численности обучающихся на 0,83 тыс.чел. или на 1,78%.

Средний уровень обучающихся за 10 лет составил 237,3 тыс.чел.

Перейдем к определению тенденции и выделению тренда. Можно предположить, что наиболее адекватно описывать тенденцию уровня образования населения Оренбургской области имеющихся данных может модель прямой или полинома второго порядка, приведем данные тренды на рисунке 3.3.

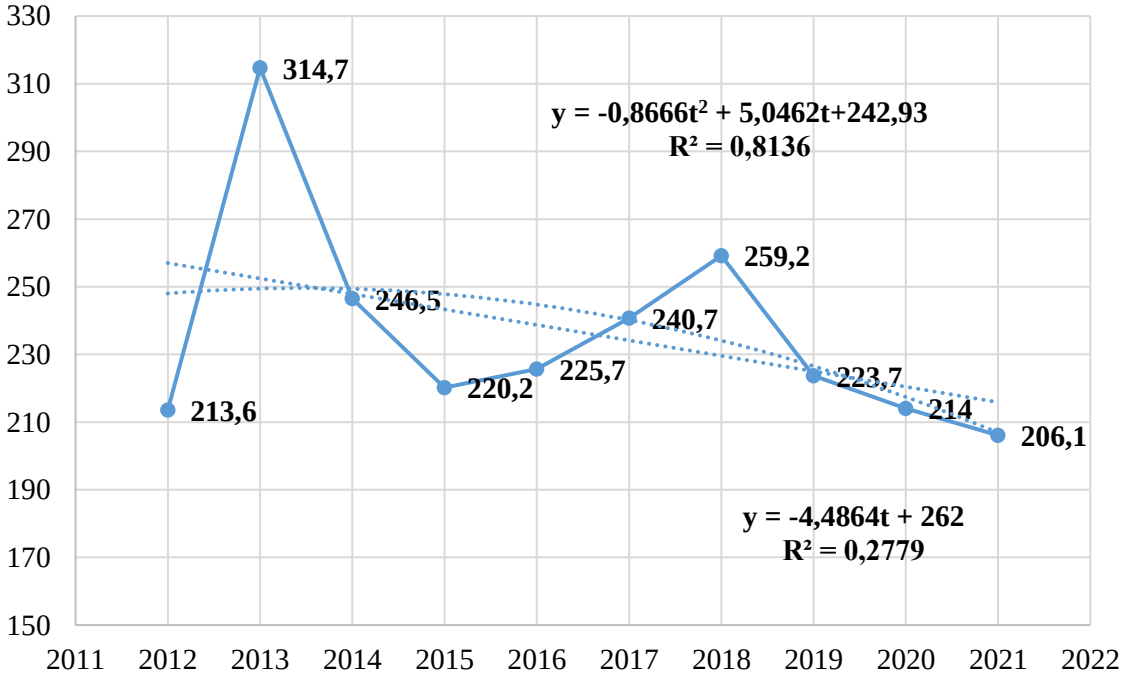


Рисунок 3.3 - Динамика уровня образования населения Оренбургской области, тренды развития

Для расчета параметров уравнения регрессии воспользуемся табличным редактором MS Excel XP, результаты расчетов представим в таблице 3.7.

Для определения наилучшего уравнения тренда следует обратить внимание на наибольший коэффициент аппроксимации и наименьшую среднеквадратическую ошибку.

Оценку надежности уравнения регрессии в целом дает R^2, в результате расчетов в случае параболы значение данного показателя выше, чем у прямой. Именно такой тренд будем использовать для приятия решений и прогнозирования.

Таблица 3.7 - Характеристики прямой и параболы второго порядка динамики численности обучающихся в Оренбургской области

Форма тренда	Модель	R^2	St
Прямая	$\tilde{y}_t$ = -4,4864t+262	0,2779	8,45
Парабола второго порядка	$\tilde{y}_t$ = -0,8666t^2+5,0462t+242,93	0,8136	3,44

Из таблицы 3.7 видно, что полиномиальный тренд наилучшим образом описывает динамику численности обучающихся в Оренбургской области.

Определяем среднюю относительную ошибку прогноза по модулю. Получаем: $\left|\bar{\delta}\right| = 1{,}23\%$

|δ|<10% это говорит о высокой точности модели.

Следовательно, полином второго порядка может быть использован нами в дальнейших расчетах для прогнозирования динамики числа обучающихся в Оренбургской области.

Пример 3.3. Построение уравнения тренда в Excel.

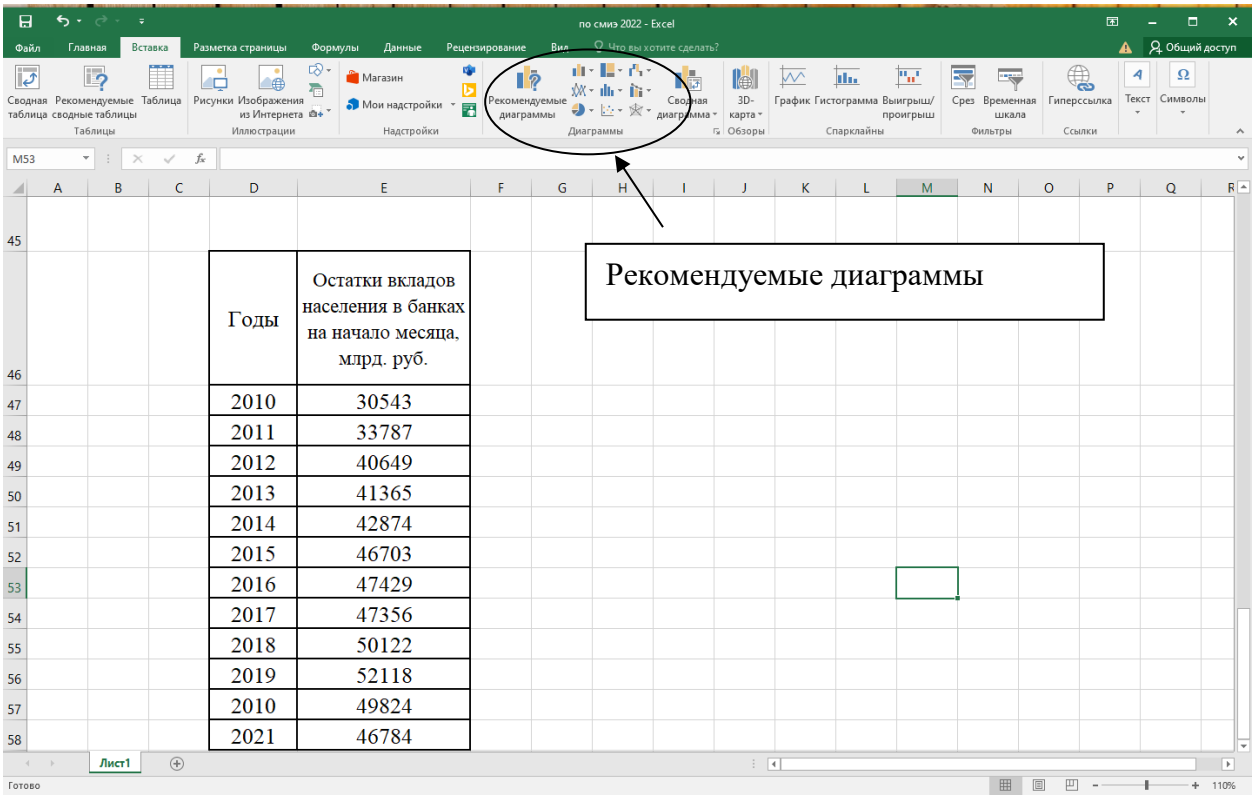
Имеются следующие данные (таблица 3.8). Определим наилучшее уравнение описывающее динамику изучаемого явления.

Таблица 3.8 – Динамика остатков вкладов населения в банках на начало месяца, млрд. руб. (условные данные)

Годы	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2010	2021
Остатки вкладов населения в банках на начало месяца, млрд. руб.	30543	33787	40649	41365	42874	46703	47429	47356	50122	52118	49824	46784

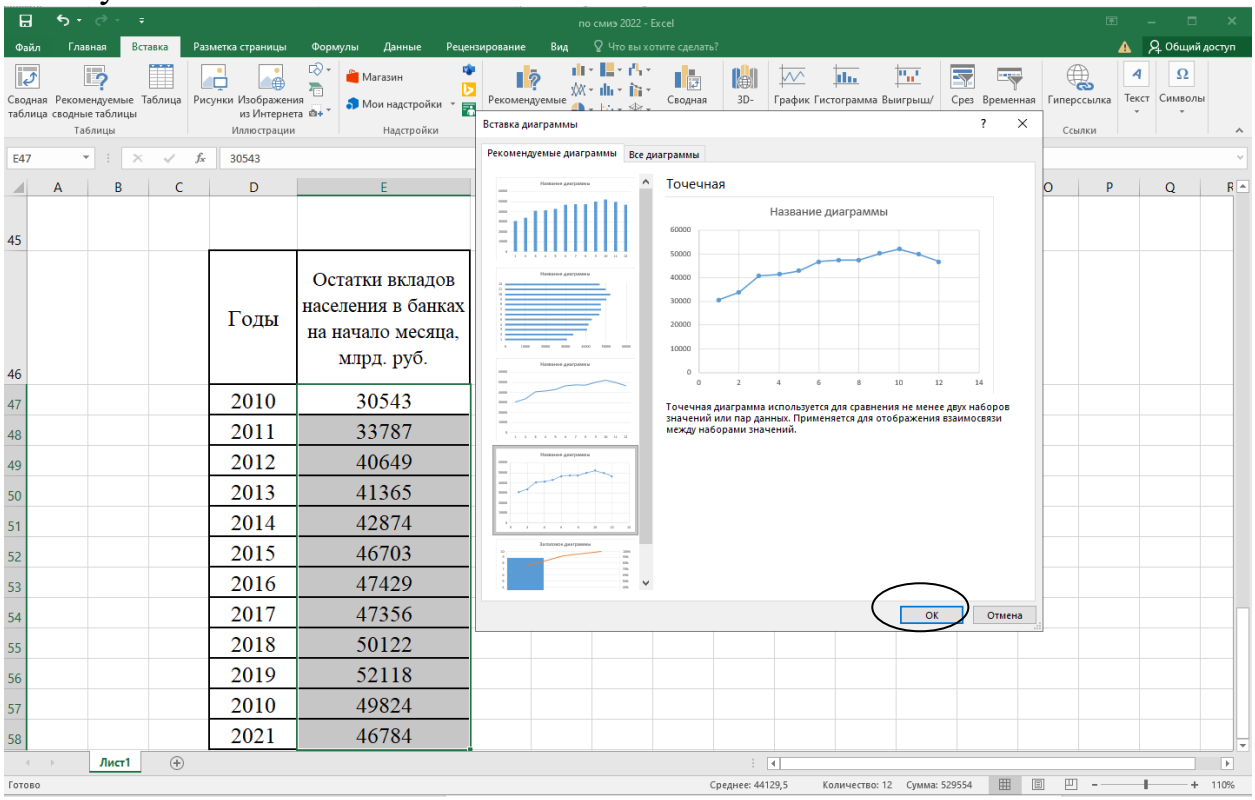
Последовательность построения тренда в Excel:

1. В окне Excel забиваем необходимые данные:

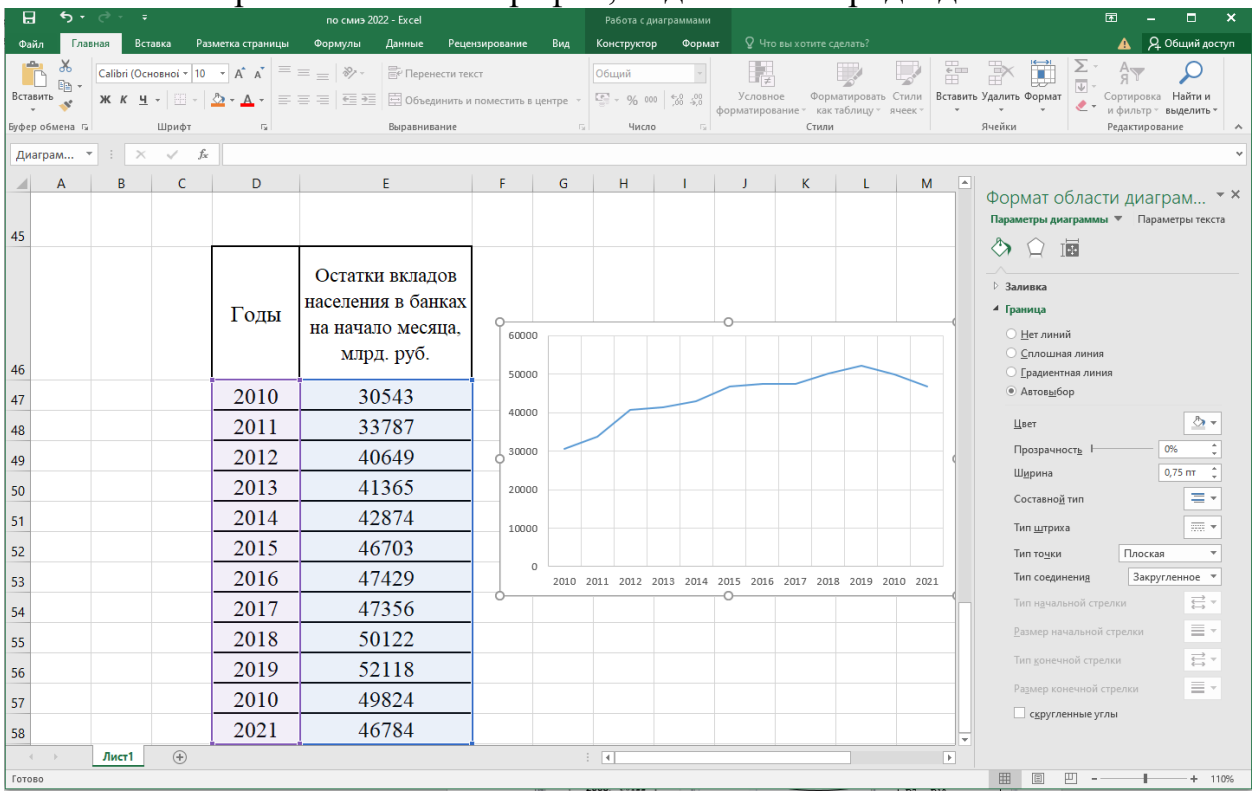


2. На панели инструментов выбираем вкладку Вставка, значок «Рекомендуемые диаграммы».

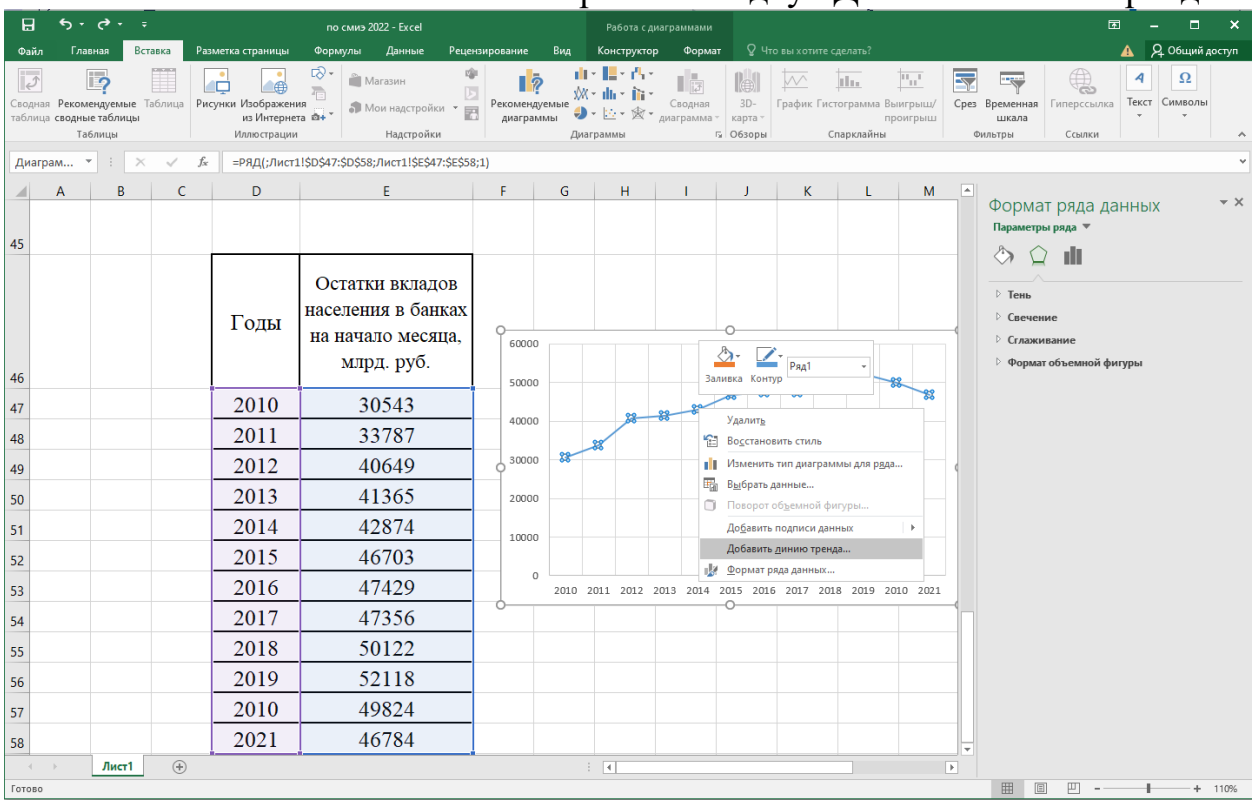
3. В появившемся окне выбираем линейный график, затем нажимаем кнопку «ОК».



4. Строим линейный график, подписываем ряды данных

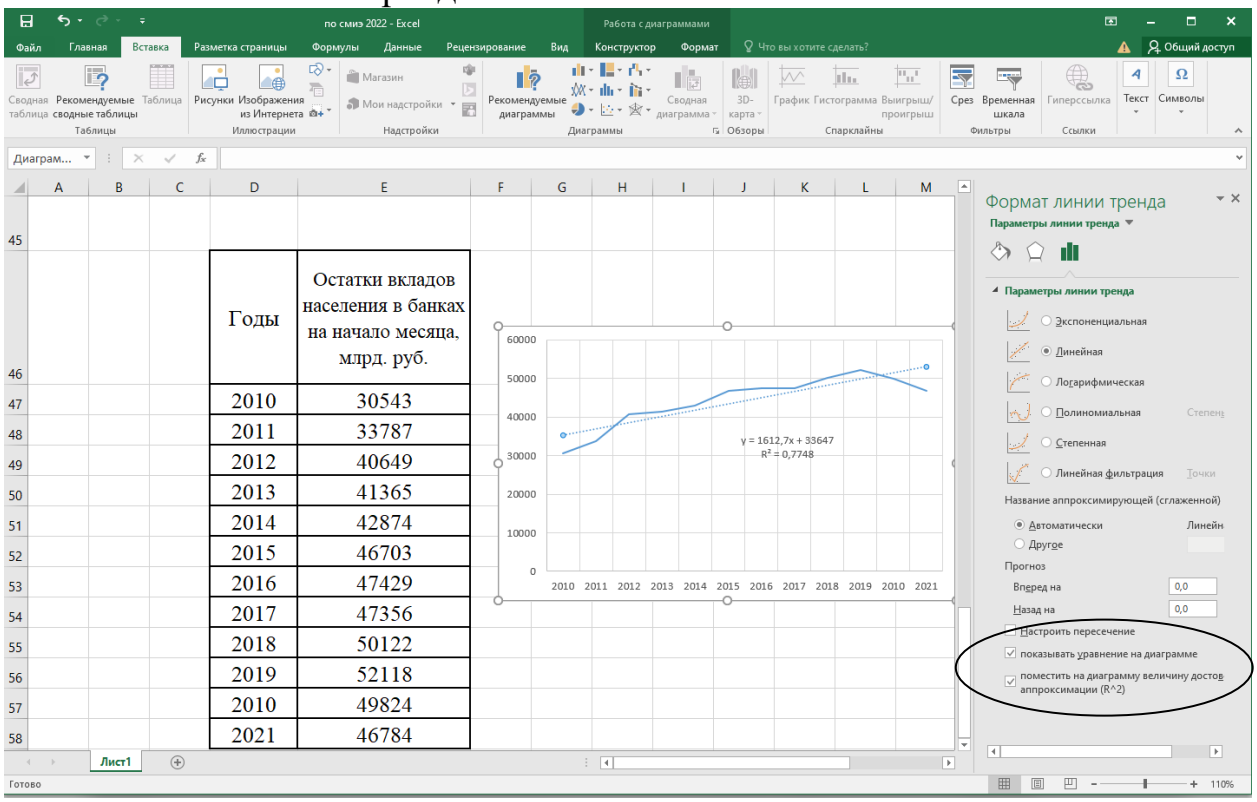


5. Для получения тренда щелкаем по полученному графику правой клавишей мыши и выбираем закладку «Добавить линию тренда».



6. В появившемся сбоку окне выбираем тип тренда (линейный, логарифмический, полиномиальный, степенной и т.д.), затем

переходим в нижнюю часть боковой панели. Мы выбрали линейный тренд.

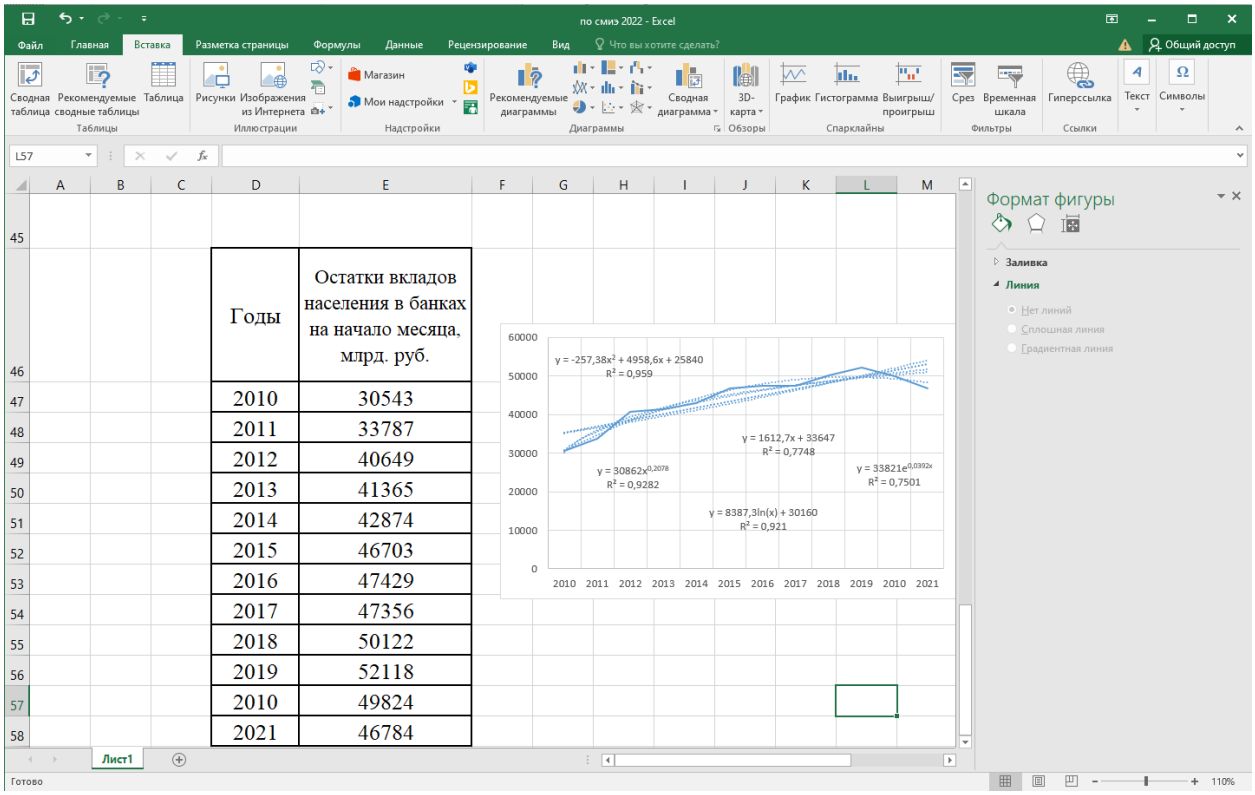


Годы	Остатки вкладов населения в банках на начало месяца, млрд. руб.
2010	30543
2011	33787
2012	40649
2013	41365
2014	42874
2015	46703
2016	47429
2017	47356
2018	50122
2019	52118
2010	49824
2021	46784

7. В появившемся окне ставим галочки напротив закладок «показать уравнение на диаграмме» и «поместить на диаграмму величину достоверности аппроксимации» и нажимаем кнопку «ОК»

8. Полученное уравнение показано на графике.

9. Аналогичная процедура проделывается для полиномиального тренд, степенного и т.д. и выбирается тренд, наилучшим образом описывающий изучаемое явление (по величине коэффициента аппроксимации R^2)
10. Результаты представляются на графике



11.По полученным данным наилучшим является полиномиальный тренд, т.к. у него самый высокий R^2.

Пример 3.4. Моделирование и прогнозирование по тренд-сезонной аддитивной модели

Имеются следующие условные данные о ценах картофеля на рынке города (табл.3.9), на основе которых:

1. определите характер общей тенденции динамики (тренда) цен картофеля (руб.);

2. используя соответствующую формулу индекса сезонности, измерьте сезонные колебания цен картофеля;

3. показатели сезонной волны изобразите графически;

4. на основе синтезированной модели сезонной волны сделайте прогноз цен на картофель по месяцам 2022 года, необходимый при составлении контракта.

Таблица 3.9 –Условные данные цен на картофель

Месяц	Годы		
	2019 г.	2020 г.	2021 г.
Январь	41	40	46
Февраль	37	38	43
Март	32	36	41
Апрель	30	34	40
Май	29	31	38
Июнь	30	31	36

Июль	25	28	33
Август	27	30	35
Сентябрь	42	40	51
Октябрь	46	50	56
Ноябрь	45	48	54
Декабрь	35	41	52
Итого	419	447	525

Решение:

1. Определим характер общей тенденции ряда динамики цен картофеля, для чего изобразим графически данные таблицы 3.1 и рассчитаем базисные темпы роста, среднегодовой темп роста и прироста [4]:

$$T_{p_б} = \frac{y_i}{y_o} \times 100\%$$

где y_i - уровень i-го периода (кроме первого); y_0 - уровень базисного периода.

$$\overline{T}_p = \sqrt[n-1]{\frac{y_n}{y_0}} \cdot 100\,\%$$

$$\overline{T}_{пр} = \overline{T}_p - 100\%$$

Графический анализ исходного временного ряда (рис. 3.4) свидетельствует о наличии трендовой компоненты, характер которой близок в равноускоренному развитию: имеется устойчивая, ярко выраженная тенденция снижения цен на картофель на рынке города.

Также отчетливо видны сезонные колебания. Наиболее существенные «всплески» в динамике показателя просматриваются в осенние месяцы каждого года. Снижение цен наблюдается в июле каждого года.

Таблица 3.10 - Показатели динамики цен картофеля на рынке города

Год	Среднемесячный уровень цен, руб.	$T_{p_б}$,%	$T_{пр_б}$,%
2019	419:12=35	100,0	-
2020	447:12=37	105,7	5,7
2021	525:12=44	125,7	25,7

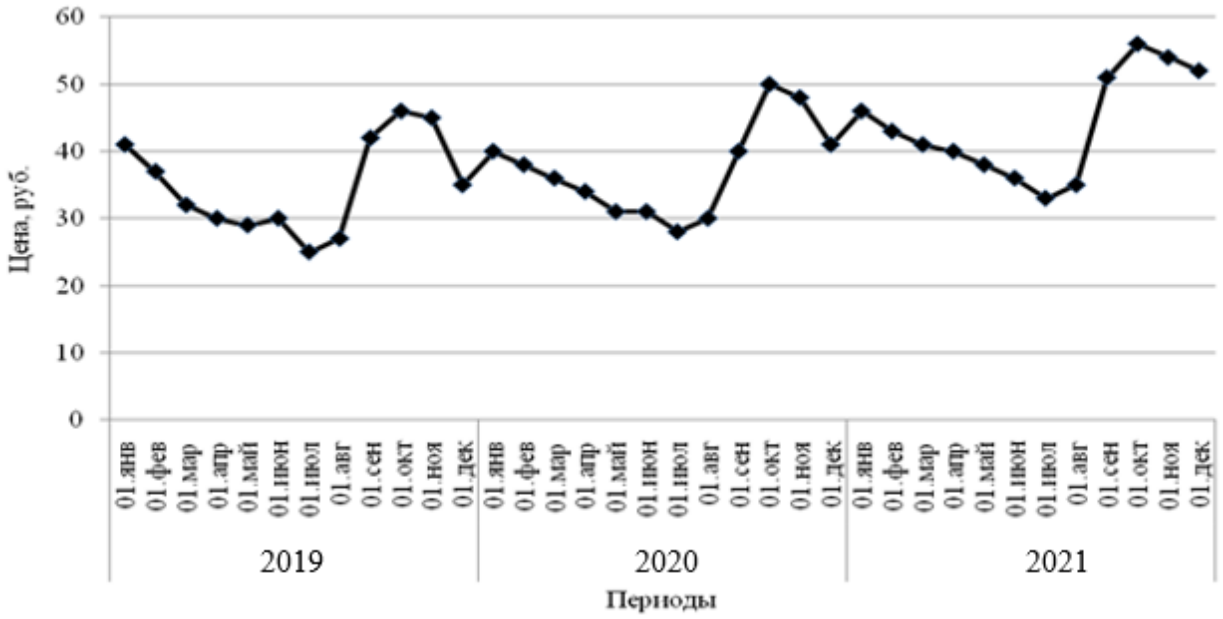


Рисунок 3.4 – Динамика цен картофеля на рынке города

Среднемесячный уровень цен в 2021 г. по сравнению с 2020 г. увеличился на 5,7%, в 2021 г. по сравнению с 2019 г. – увеличился на 25,7%.

Следовательно, выявилась закономерность к росту уровня цен.

Среднегодовой темп роста равен:

$$\overline{T}_p = \sqrt[3-1]{\frac{525}{419}} \cdot 100\,\% = 111{,}9\%$$

Среднегодовой темп прироста равен:

$$\overline{T}_{пр} = 111{,}9 - 100\% = 11{,}9\%.$$

О тенденциях роста свидетельствует и показатель среднегодового темпа роста. Так, в среднем за исследуемый период уровень цен на картофель увеличился на 11,9%.

2. Рассчитаем индексы сезонности по формуле, данные представим в таблице 3.11.

$$i_c = \frac{\overline{y}_n}{\overline{y}} \cdot 100\%,$$

[5].

Таблица 3.11 - Цены на картофель по сезонам

Месяц	Всего за 3 года, руб. $\sum y_n$	В среднем за 3 года, руб. $\overline{y}_n$	Индекс сезонности, % i_c
Январь	41+40+46=127	127:3=42,3	(42,3:38,6)*100%=109,6
Февраль	118	39,3	101,8
Март	109	36,3	94,0
Апрель	104	34,7	89,7
Май	98	32,7	84,5
Июнь	97	32,3	83,7
Июль	86	28,7	74,2

Август	92	30,7	79,4
Сентябрь	133	44,3	114,7
Октябрь	152	50,7	131,1
Ноябрь	147	49,0	126,8
Декабрь	128	42,7	110,4
Итого за 3 года	1391	463,7	-
В среднем	1391:12=115,9	$\bar{y}$ =38,6 (463,7:12)	-

3. Изобразим графически сезонную волну (рис. 3.5).

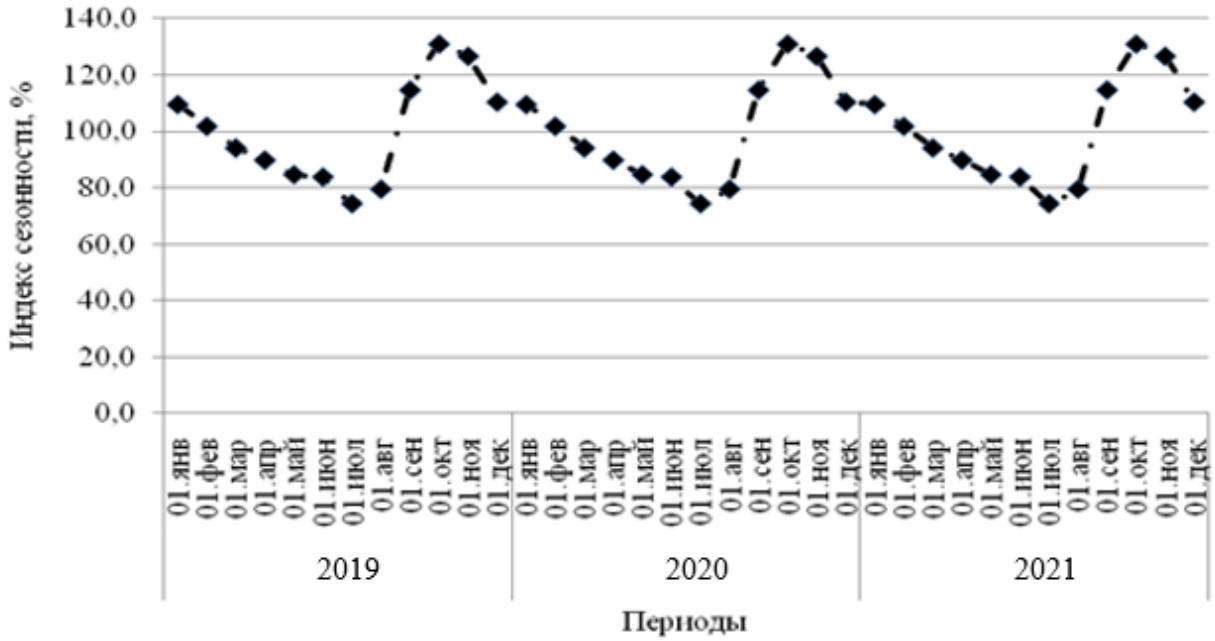


Рисунок 3.5 – Сезонная волна уровня цен на картофель на рынке города

Сезонные колебания цен на картофель характеризуются повышением в октябре (+31,1%), наибольшее снижение цен наблюдается в июле.

Прежде чем, переходить к определению тенденции и выделению тренда, необходимо выяснить существует ли тенденция в динамическом ряду уровня цен на картофель на рынке города. Для проверки наличия тренда воспользуемся методом критерия серий, основанном на медиане выборки (табл. 3.12).

Таблица 3.12 - Модификация критерия серий

Критерий серий	$\nu(36)$	$\nu(36)_{\text{расч}}$	$\tau_{\max}(36)$	$\tau_0(36)$	H_0
Критерий серий, основанный на медиане выборки	7,0	13,0	17,0	5,0	отвергается

Медиана исходного ряда (табл. 3.1) Me =29,5 руб.

Число серий определяется путем подсчета: υ(36) = 7,0.

Протяженность самой длинной серии τ max (36) = 17,0.

Проверяем гипотезу о случайности исходного ряда, для этого должны выполняться неравенства. Получаем: проверка гипотезы основывается на проверки гипотезы о случайности ряда, поэтому для того, чтобы не была отвергнута гипотеза о случайности исходного ряда должны выполнятся следующие неравенства [1]:

$$\nu(n) \rangle \left[\frac{1}{2}(n+1-1{,}96\sqrt{n-1}) \right]$$

$$\tau_{\max}(n) \langle [1{,}43 \ln(n+1)]$$

Данные неравенства не выполняются, следовательно, гипотеза о случайности исходного динамического ряда уровня цен на картофель на рынке города отклоняется – в исследуемом ряду динамики присутствует тенденция.

$$\nu(36) \prec \nu(36)_{расч} \quad \text{и} \quad \tau_{\max}(36) \succ \tau_0(36)$$

Поскольку амплитуда сезонных колебаний постоянна, не изменяется с течением времени, то для описания динамики временного ряда выбрана модель с аддитивной сезонностью.

Результаты расчётов представлены в приложении Л.

Последовательность построения тренд-сезонной аддитивной модели заключается в следующем:

1) показатели сглаживания исходного временного ряда с помощью скользящей средней определили по формуле:

$$y'_t = \frac{\frac{1}{2}y_{t-6} + y_{t-5} + \ldots + y_t + \ldots + y_{t+5} + \frac{1}{2}y_{t+6}}{12}$$

2) вычислим уровни x_t, полученные вычитанием значений исходного временного ряда и значений сглаженного ряда ($x_t = y_t - y'_t$) (уровни x_t отражают влияние случайных факторов и сезонности);

3) предварительную оценку сезонности получили усреднением уровней временного ряда x_t для одноименных месяцев (табл. 3.13).

4) Так как $\sum_{i=1}^{12} x_i$ = -0,69, то проведем корректировку ранее полученных оценок сезонности (-0,69:12=-0,06).

Окончательная оценка сезонности S_i приведена в табл. 3.13.

5) в расчетной таблице приложения Л представлен десезонализированный временной ряд $y_t^{(1)}$, полученный вычитанием данных исходного временного ряда и соответствующих коэффициентов сезонности (коэффициенты сезонности принимаются одинаковыми для каждого анализируемого периода).

6) для описания тенденции воспользовались моделью полиномиального тренда, так как это согласуется с результатами графического анализа динамики показателя (рис. 3.6). Модель для десезонализированного временного ряда имеет вид:

$$\hat{y}_t^{(1)} = 0{,}0134t^2 - 0{,}1237t + 34{,}876, \quad R^2 = 0{,}9033$$

Таблица 3.13 - Оценивание сезонной компоненты в аддитивной модели

Месяц	№	Предварительная оценка, x_t	Скорректированная оценка, S_t
Январь	1	(4,0+5,7):2=4,85	4,85-(-0,06)=4,91
Февраль	2	2,04	2,10
Март	3	-0,31	-0,26
Апрель	4	-2,21	-2,15
Май	5	-5,10	-5,05
Июнь	6	-6,65	-6,59
Июль	7	-9,69	-9,63
Август	8	-7,92	-7,86
Сентябрь	9	4,27	4,33
Октябрь	10	10,88	10,93
Ноябрь	11	8,98	9,03
Декабрь	12	0,17	0,20
Сумма		-0,69	0,00

*Рассчитано с помощью ППП «Microsoft Excel XP»

Рисунок 3.6 – Параболический тренд в динамике уровня цен на картофель на рынке города

7) на заключительном этапе расчетные уровни временного ряда «цена картофеля» были вычислены как сумма значений, полученных по полиномиальной модели для десезонализированного временного ряда, и соответствующих коэффициентов сезонности.

Прогнозная оценка цен на картофель на рынке города в 2022 г. представлена в таблице 3.14.

Таблица 3.14 - Результаты прогнозирования уровня цен на картофель на рынке города по аддитивной тренд-сезонной модели, руб.

Год	Месяц	t	Трендовая компонента T	Сезонная компонента S_i	Прогноз $y_s = T+S_i$
2022	Январь	37	48,64	4,93	53,57
	Февраль	38	49,53	2,10	51,62
	Март	39	50,43	-0,26	50,18
	Апрель	40	51,37	-2,15	49,22
	Май	41	52,33	-5,05	47,28
	Июнь	42	53,32	-6,59	46,73
	Июль	43	54,33	-9,63	44,70
	Август	44	55,38	-7,86	47,51
	Сентябрь	45	56,44	4,33	60,77
	Октябрь	46	57,54	10,93	68,47
	Ноябрь	47	58,66	9,03	67,70
	Декабрь	48	59,81	0,22	60,03

Таким образом, по модели десезонализированного временного ряда при условии сохранения тенденции ожидаемый уровень цен на картофель будет колебаться в 2022 г. в пределах от 53,57 руб. до 60,03 руб.

<u>Пример 3.5. Расчет индексов сезонности методом простой средней, построение сезонной волны</u>

Имеются условные данные о продаже яблок на рынках города по кварталам за 2019-2021гг.:

Таблица 3.15 - Условные данные о продаже яблок на рынках города

Квартал	Продано яблок (т)		
	2019 г.	2020 г.	2021 г.
1	320	312	325
2	275	290	283
3	344	369	357
4	330	325	341

Рассчитайте индексы сезонности методом простой средней, постройте сезонную волну. Результаты проанализируйте.

Решение:

Индекс сезонности рассчитывается по формуле:

$$i_{сез} = \frac{\bar{y}_i}{\bar{y}_0} \cdot 100 \text{ ,}$$

где числитель - среднее значение объема продаж по каждому кварталу, вычисляется как арифметическая простая; знаменатель – общая средняя объема продаж, вычисляется как арифметическая взвешенная (табл. 3.16).

Таблица 3.16 – Вспомогательные данные

Квартал	Продано яблок (т)			Средний объем продаж по кварталам ($\bar{y}_i$)	Индекс сезонности ($i_{сез}$)
	2019 г.	2020 г.	2021 г.		

1	320	312	325	(320+312+325):3=319	95,9
2	275	290	283	283	85,1
3	344	369	357	357	107,3
4	330	325	341	332,7	99,8

Общая средняя за весь период:

$$\bar{y}_0 = \frac{319 \times 90 + 283 \times 91 + 357 \times 92 + 332 \times 92}{365} = 332,7 \text{ т.}$$

Индекс сезонности:

1 квартал: $i_{сез} = \frac{\bar{y}_i}{\bar{y}_0} \cdot 100 = \frac{319}{332,7} \cdot 100 = 95,9\%$

2 квартал: $i_{сез} = \frac{\bar{y}_i}{\bar{y}_0} \cdot 100 = \frac{283}{332,7} \cdot 100 = 85,1\%$

3 квартал: $i_{сез} = \frac{\bar{y}_i}{\bar{y}_0} \cdot 100 = \frac{357}{332,7} \cdot 100 = 107,3\%$

4 квартал: $i_{сез} = \frac{\bar{y}_i}{\bar{y}_0} \cdot 100 = \frac{332}{332,7} \cdot 100 = 99,8\%$

Индексы сезонности показывают, что средний объем продаж яблок в 3 квартале больше среднего объема продаж за весь период на 7,3%, а во 2 квартале меньше среднего объема продаж яблок на 14,9%.

Построим сезонную волну по приведенным ниже данным (табл.3.17):

Таблица 3.17 – Расчет индекса сезонности

Годы	Квартал	Индекс сезонности ($i_{сез}$)
2019	1	95,9
	2	85,1
	3	107,3
	4	99,8
2020	1	95,9
	2	85,1
	3	107,3
	4	99,8
2021	1	95,9
	2	85,1
	3	107,3
	4	99,8

Рисунок 3.7. Динамика объема продажи яблок в 1-4 кварталах 2019-2021 гг.

По данным рисунка 3.7 видно, что пик продажи яблок приходится на третий квартал каждого года; снижение объема продаж приходятся на второй квартал каждого года.

Пример 3.6. Применение рядов Фурье при анализе сезонности

Постройте тренд процентных ставок по кредитам на основе статистических данных (табл. 3.18), опубликованных в «Бюллетене банковской статистики». (Используйте ряд Фурье по 1-ой и 2-ой гармонике).

Таблица 3.18 – Условные данные по процентным ставкам по кредитам (% годовых)

2019				2020				2021			
I кв.	II кв.	III кв.	IV кв.	I кв.	II кв.	III кв.	IV кв.	I кв.	II кв.	III кв.	IV кв.
37,7	34,4	30,4	29,6	25,6	23,1	20,4	14,1	11,1	10,5	9,0	7,8

Решение:

Ряд Фурье с одной гармоникой принимает вид:

$$y_t = a_0 + a_1 \cdot \cos t + b_1 \cdot \sin t$$

Ряд Фурье с двумя гармониками имеет вид:

$$y_t = a_0 + a_1 \cdot \cos t + b_1 \cdot \sin t + a_2 \cdot \cos 2t + b_2 \cdot \sin 2t$$

Параметры уравнений вычислим по формулам:

$$a_0 = \frac{\sum y_t}{N}$$

$$a_1 = \frac{2\sum y \cdot \cos t}{N}$$

$$b_1 = \frac{2\sum y \cdot \sin t}{N}$$

При определении второй гармоники рассчитываются a_2 и b_2 [9]:

$$a_2 = \frac{2}{N} \cdot \sum y \cdot \cos 2t$$

$$b_2 = \frac{2}{N} \cdot \sum y \cdot \sin 2t$$

Результаты промежуточных расчетов представим в таблице 3.19.

Таблица 3.19 - Расчет параметров по ряду Фурье

Годы	№ квар-тала	Y,%	t	2t	cos t	sin t	cos 2t	sin 2t	y*cos t	y*sin t	y*cos 2t	y*sin 2t
2019	I	37,7	0,000	0,000	1,000	0,000	1,000	0,000	37,700	0,000	37,700	0,000
	II	34,4	0,524	1,048	0,866	0,500	0,499	0,866	29,784	17,212	17,176	29,805
	III	30,4	1,048	2,097	0,499	0,867	-0,502	0,865	15,169	26,345	-15,262	26,291
	IV	29,6	1,573	3,146	-0,002	1,000	-1,000	-0,004	-0,058	29,600	-29,600	-0,116
2020	I	25,6	2,097	4,194	-0,502	0,865	-0,495	-0,869	-12,861	22,135	-12,678	-22,240
	II	23,1	2,622	5,243	-0,868	0,497	0,506	-0,862	-20,046	11,479	11,691	-19,923
	III	20,4	3,146	6,292	-1,000	-0,004	1,000	0,009	-20,400	-0,088	20,399	0,176
	IV	14,1	3,670	7,341	-0,863	-0,504	0,491	0,871	-12,175	-7,112	6,925	12,282
2021	I	11,1	4,195	8,389	-0,495	-0,869	-0,510	0,860	-5,493	-9,645	-5,662	9,547
	II	10,5	4,719	9,438	0,007	-1,000	-1,000	-0,013	0,070	-10,500	-10,499	-0,140
	III	9,0	5,243	10,487	0,506	-0,862	-0,487	-0,873	4,558	-7,761	-4,384	-7,860
	IV	7,8	5,768	11,536	0,870	-0,493	0,514	-0,858	6,787	-3,844	4,010	-6,690
Итого	-	253,7	34,605	69,210	0,017	-0,004	0,016	-0,008	23,035	67,822	19,817	21,131

*Рассчитано с помощью ППП «Microsoft Excel XP»

Подставим соответствующие значения в формулы:

$$a_0 = \frac{253,7}{12} = 21,142$$

$$a_1 = \frac{2 \cdot 23,035}{12} = 3,839$$

$$b_1 = \frac{2 \cdot 67,822}{12} = 11,304$$

$$a_2 = \frac{2}{12} \cdot 19,817 = 3,303$$

$$b_2 = \frac{2}{12} \cdot 21,131 = 3,522$$

Таким образом, ряд Фурье с одной гармоникой принимает вид:

$$y_t = 21,142 + 3,839 \cdot \cos t + 11,304 \cdot \sin t$$

Ряд Фурье с двумя гармониками принимает вид:

$$y_t = 21,142 + 3,839 \cdot \cos t + 11,304 \cdot \sin t + 3,303 \cdot \cos 2t + 3,522 \cdot \sin 2t$$

Теоретические значения по уравнениям представлены в таблице 3.20.

Таблица 3.20 - Теоретические значения по рядам Фурье, % годовых

Годы	№ квар-тала	Фактические данные	Теоретические значения ряду Фурье с одной гармоникой	Теоретические значения ряду Фурье с двумя гармониками
2019	I	37,7	25,0	28,3
	II	34,4	30,1	34,8
	III	30,4	32,9	34,2
	IV	29,6	32,4	29,1
2020	I	25,6	29,0	24,3
	II	23,1	23,4	22,1
	III	20,4	17,3	20,6
	IV	14,1	12,1	16,8
2021	I	11,1	9,4	10,8
	II	10,5	9,9	6,5
	III	9,0	13,3	8,7
	IV	7,8	18,9	17,6
Итого	-	253,7	253,7	253,7

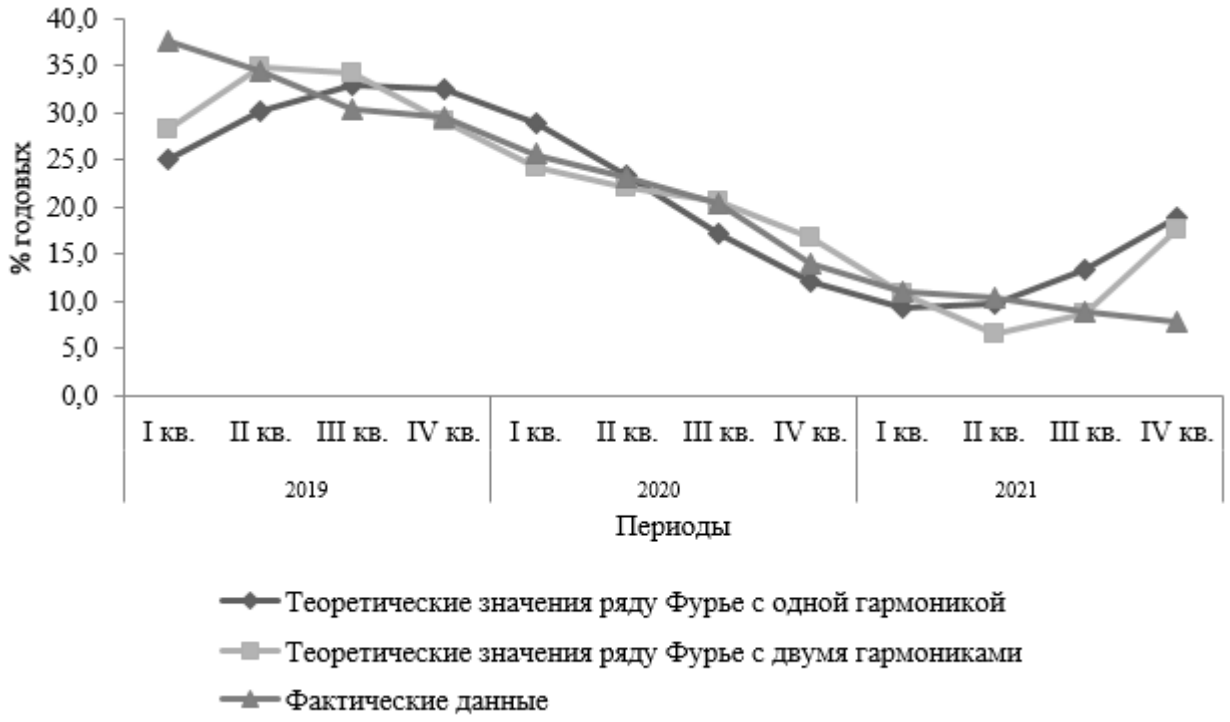


Рисунок 3.8 – Динамика процентных ставок по кредитам

<u>Пример 3.7. Определение показателей динамики при анализе сезонности, выявление сезонности, построение и анализ сезонной волны.</u>

Имеются условные данные об изменении денежной массы за год на начало месяца, млрд. руб. (табл.3.21).

Таблица 3.21 Условны данные о динамике денежной массы

Месяц	Денежная масса	в том числе	
		наличные деньги	безналичные средства
Январь	295,2	103,8	191,4
Февраль	297,4	96,3	201,1
Март	307,6	102,0	205,5
Апрель	315,0	105,2	209,8
Май	328,4	115,2	213,2
Июнь	339,4	120,4	219,0
Июль	363,8	136,8	227,0
Август	375,5	140,3	235,2
Сентябрь	377,7	141,6	236,1
Октябрь	376,2	134,8	241,4
Ноябрь	382,3	135,7	246,6
Декабрь	371,1	128,7	242,3

1) Определите долю наличности и безналичных средств, показатели динамики денежной массы и наличие денег по месяцам (абсолютное, относительное, среднее).

2) Выявите сезонность в динамике наличного и безналичного денежного оборота.

Решение:

1) Определим долю наличности и безналичных средств, результаты представим в таблице 3.22

Таблица 3.22 - Структура денежной массы за год на начало месяца, в % к итогу

Месяц	Денежная масса	в том числе	
		наличные деньги	безналичные средства
Январь	100,0	35,2	64,8
Февраль	100,0	32,4	67,6
Март	100,0	33,2	66,8
Апрель	100,0	33,4	66,6
Май	100,0	35,1	64,9
Июнь	100,0	35,5	64,5
Июль	100,0	37,6	62,4
Август	100,0	37,4	62,6
Сентябрь	100,0	37,5	62,5
Октябрь	100,0	35,8	64,2
Ноябрь	100,0	35,5	64,5
Декабрь	100,0	34,7	65,3

По данным таблицы 3.22 видно, что доля наличных денег в общем объеме денежной массы колеблется в разные месяцы от 32,4% до 37,6%. Наибольшая доля наличных денег приходится на июль, наименьшая – на февраль.

Доля безналичных средств от общей денежной массы составляет около 68%, и колеблется в разные месяцы от 62,4% до 67,6%. Наибольшая доля безналичных средств приходится на февраль, наименьшая – на июль.

Проведем анализ динамики денежной массы и наличия денег по месяцам, рассчитаем следующие показатели:

1) абсолютные приросты на цепной и базисной основе ($\Delta_{баз}, \Delta_{цеп}$);

2) цепные и базисные темпы роста ($T_{р\ баз}, T_{р\ цеп}$);

3) темпы прироста ($T_{пр\ баз}, T_{пр\ цеп}$).

Абсолютный прирост - это разность между сравниваемым уровнем и уровнем более раннего периода, принятым за базу сравнения. Если эта база - непосредственно предыдущий уровень, показатель называют цепным, если за базу взят, например, начальный уровень, показатель называют базисным.

$$\text{цепной: } \Delta_{баз} = y_n - y_{n-1},$$

$$\text{базисный: } \Delta_{цеп} = y_n - y_0$$

где y_n – сравниваемый уровень ряда динамики; y_0 – начальный (базисный) уровень; y_{n-1} – уровень ряда динамики, предшествующий сравниваемому.

Интенсивность изменения уровней временного ряда характеризуется темпами роста и прироста.

Темп роста - это отношение сравниваемого уровня (более позднего) к уровню, принятому за базу сравнения (более раннему).

Темп роста исчисляется в цепном варианте - к уровню предыдущего года, и в базисном - к одному и тому же, обычно начальному уровню, что иллюстрируется формулами ниже. Он свидетельствует о том, сколько процентов составляет сравниваемый уровень по отношению к уровню, принятому за базу, или во сколько раз сравниваемый уровень больше уровня, принятого за базу.

$$T_{р\ цеп} = \frac{y_n}{y_{n-1}} \cdot 100\%,$$

Базисный темп роста за весь период между базой (0) и текущим годом (n):

$$T_{р\ баз} = \frac{y_n}{y_o} \cdot 100\%$$

Темп прироста равен темпу роста минус единица (минус 100%).

$$\text{цепной: } T_{пр\ цеп} = T_{р\ цеп} - 100\%$$

$$\text{базисный: } T_{пр\ баз} = T_{р\ баз} - 100\%$$

Темп прироста характеризует абсолютный прирост в относительных величинах и показывает, на сколько процентов изменился сравниваемый уровень по отношению к базе сравнения.

Для обобщения данных по рядам динамики рассчитаем:

- средний абсолютный прирост;
- средние темпы роста и прироста.

Средний абсолютный прирост (абсолютное изменение) определяется как частное от деления базисного абсолютного изменения на число осредняемых отрезков времени от базисного до сравниваемого периода:

$$\overline{\Delta} = \frac{y_n - y_0}{n-1}$$

Средний темп роста определяется наиболее точно при аналитическом выравнивании динамического ряда по экспоненте. Если можно пренебречь колеблемостью, то средний темп определяют как геометрическую среднюю из цепных темпов роста за n лет или из общего (базисного) темпа роста за n лет:

$$\overline{T_p} = \sqrt[n-1]{\prod_{i=1}^{n} k_i} \cdot 100\% = \sqrt[n-1]{\frac{y_n}{y_0}} \cdot 100\%$$

Средний темп прироста определяется как разность между средним темпом роста и 1 (или 100%):

$$\overline{T}_{np} = \overline{T_p} - 1 \ (100\,\%)$$

Результаты расчета представим в таблице 3.23.

Анализ данных таблицы показал, что по сравнению с базисным месяцем – январем, объем денежной массы имел тенденцию к росту, в то время как в сравнении с каждым предыдущим месяцем наблюдается неустойчивая тенденция к увеличению наличных денег.

Таблица 3.23 - Показатели динамики денежной массы

Месяц	Наличные деньги, млрд. руб.	Абсолютный прирост, млрд. руб.		Темп роста, %		Темп прироста, %	
		цепной	базисный	цепной	базисный	цепной	базисный
Январь	295,2	-	-	-	-	-	-
Февраль	297,4	2,2	2,2	100,7	100,7	0,7	0,7
Март	307,6	10,2	12,4	103,4	104,2	3,4	4,2
Апрель	315,0	7,4	19,8	102,4	106,7	2,4	6,7
Май	328,4	13,4	33,2	104,3	111,2	4,3	11,2
Июнь	339,4	11,0	44,2	103,3	115,0	3,3	15,0
Июль	363,8	24,4	68,6	107,2	123,2	7,2	23,2
Август	375,5	11,7	80,3	103,2	127,2	3,2	27,2
Сентябрь	377,7	2,2	82,5	100,6	127,9	0,6	27,9
Октябрь	376,2	-1,5	81,0	99,6	127,4	-0,4	27,4
Ноябрь	382,3	6,1	87,1	101,6	129,5	1,6	29,5
Декабрь	371,1	-11,2	75,9	97,1	125,7	-2,9	25,7

Так, в октябре по сравнению с сентябрем и декабре по сравнению с ноябрем объем денежной массы снизился на 1,5 млрд. руб. и 11,2 млрд. руб. соответственно.

$$\overline{\Delta} = \frac{371,1 - 295,2}{12 - 1} = 6,9\ млрд.руб.$$

$$\overline{T_p} = \sqrt[12-1]{\frac{371,1}{295,2}} \cdot 100\% = 102,1\%$$

$$\overline{T}_{пр} = 102,1 - 100 = 2,1\%$$

Расчет средних показателей динамики показал, что в среднем за период объем денежной массы увеличивался на 6,9 млрд. руб. или на 2,1% ежемесячно.

Анализ данных таблицы 3.24 позволил сделать следующие выводы: по сравнению с базисным месяцем – январем, сумма наличных денег имела тенденцию к росту, в то время как в сравнении с каждым предыдущим месяцем наблюдается неустойчивая тенденция к увеличению наличных денег.

Так, в феврале по сравнению январем, октябре по сравнению с сентябрем, декабре по сравнению с ноябрем, сумма наличных денег снизилась на 7,5 млрд. руб., 6,8 млрд. руб. и 7,0 млрд. руб. соответственно.

$$\overline{\Delta} = \frac{128,7 - 103,8}{12 - 1} = 2,3\ млрд.руб.$$

$$\overline{T_p} = \sqrt[12-1]{\frac{128,7}{103,8}} \cdot 100\% = 102,0\%$$

$$\overline{T}_{пр} = 102,0 - 100 = 2,0\%$$

Расчет средних показателей динамики по формулам показал, что в среднем за период сумма наличных денег увеличивалась на 2,3 млрд. руб. или на 2% ежемесячно.

Таблица 3.24 - Показатели динамики наличных денег

Месяц	Наличные деньги, млрд. руб.	Абсолютный прирост, млрд. руб.		Темп роста, %		Темп прироста, %	
		цепной	базисный	цепной	базисный	цепной	базисный
Январь	103,8	-	-	-	-	-	-
Февраль	96,3	-7,5	-7,5	92,8	92,8	-7,2	-7,2
Март	102,0	5,7	-1,8	105,9	98,3	5,9	-1,7
Апрель	105,2	3,2	1,4	103,1	101,3	3,1	1,3
Май	115,2	10,0	11,4	109,5	111,0	9,5	11,0
Июнь	120,4	5,2	16,6	104,5	116,0	4,5	16,0
Июль	136,8	16,4	33,0	113,6	131,8	13,6	31,8
Август	140,3	3,5	36,5	102,6	135,2	2,6	35,2
Сентябрь	141,6	1,3	37,8	100,9	136,4	0,9	36,4
Октябрь	134,8	-6,8	31,0	95,2	129,9	-4,8	29,9
Ноябрь	135,7	0,9	31,9	100,7	130,7	0,7	30,7
Декабрь	128,7	-7,0	24,9	94,8	124,0	-5,2	24,0

2) Выявим сезонность в динамике наличного и безналичного денежного оборота, рассчитаем индекс сезонности по формуле, результаты представим в таблице 3.25.

$$i_c = \frac{y_n}{\bar{y}} \cdot 100\%,$$

где y_n - уровень п-ого месяца, $\bar{y}$ - средний уровень показателя по всей длине ряда.

Таблица 3.25 – Расчет индекса сезонности

Месяц	Наличные деньги, млрд. руб.	Индекс сезонности наличных денег, %	Безналичные средства, млрд. руб.	Индекс сезонности безналичного денежного оборота, %
Январь	103,8	85,3	191,4	86,1
Февраль	96,3	79,1	201,1	90,4
Март	102,0	83,8	205,5	92,4
Апрель	105,2	86,4	209,8	94,3
Май	115,2	94,6	213,2	95,9
Июнь	120,4	98,9	219,0	98,5
Июль	136,8	112,4	227,0	102,1
Август	140,3	115,3	235,2	105,8
Сентябрь	141,6	116,3	236,1	106,2
Октябрь	134,8	110,7	241,4	108,6
Ноябрь	135,7	111,5	246,6	110,9
Декабрь	128,7	105,7	242,3	109,0
Итого	1460,8	-	2668,6	-
В среднем	121,7	-	222,4	-

Рисунок 3.9 – Сезонная волна в динамике наличного и безналичного денежного оборота

При условии, что тенденции в динамике наличного и безналичного денежного оборота сохранятся, то:

-сезонные колебания в динамике наличного денежного оборота характеризуются повышением в сентябре (+16,3%), наибольшее снижение наличных денег наблюдается в феврале;

-сезонные колебания в динамике безналичного денежного оборота характеризуются увеличением в ноябре (+10,9%), наибольшее снижение безналичных средств наблюдается в январе.

Пример 3.8. Моделирование и прогнозирование по тренд-сезонной мультипликативной модели

В таблице 3.26 представлены квартальные данные с 2017 г. по 2021 г. о динамике оборота розничной торговли по РФ.

Требуется:

1) провести исследование компонентного состава анализируемого временного ряда;

2) рассчитать прогнозную оценку оборота розничной торговли в 2022 году по кварталам.

Таблица 3.26 _Динамика оборота розничной торговли по РФ

Годы	№ квартала	Оборот розничной торговли - всего, млн. руб.
2017	I кв.	4184798,0
	II кв.	4573224,4
	III кв.	4900496,2
	IV кв.	5445817,9
2018	I кв.	4689759,6
	II кв.	5112196,2
	III кв.	5492387,4
	IV кв.	6100183,0
2019	I кв.	5241300,6
	II кв.	5692793,9
	III кв.	6052021,7
	IV кв.	6699797,3
2020	I кв.	5792954,2
	II кв.	6256750,1
	III кв.	6697241,3
	IV кв.	7609291,7
2021	I кв.	6268188,6
	II кв.	6592601,0
	III кв.	6997485,7
	IV кв.	7668517,9

Источник данных: http://www.gks.ru/wps/wcm/connect/rosstat_main/rosstat/ru/statistics/enterprise/retail/

Решение:

1) С помощью функций MS Excel изобразим графически данные таблицы 3.26.

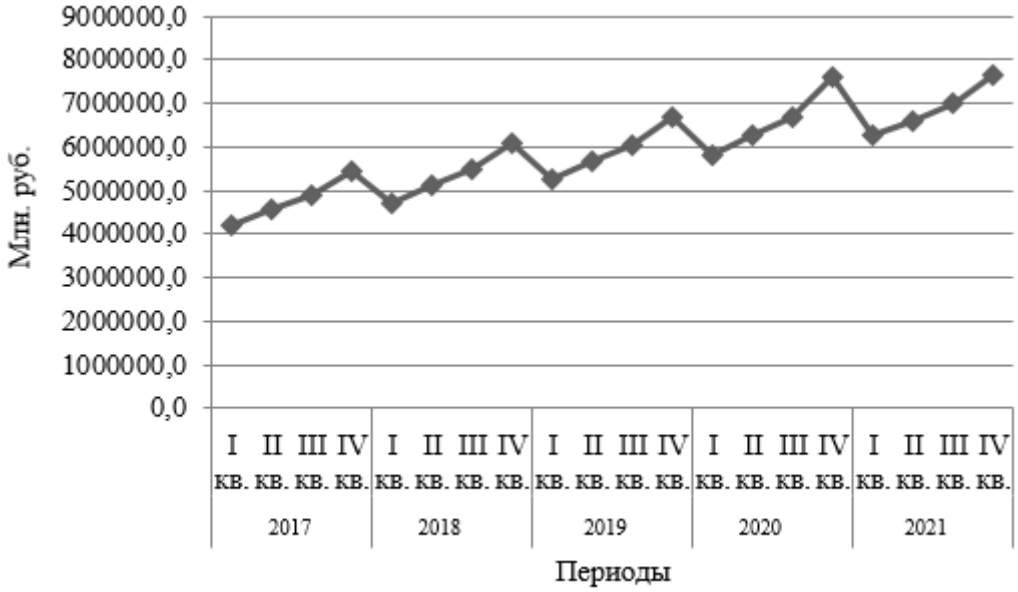


Рисунок 3.10 – Квартальная динамика оборота розничной торговли в РФ за 2017-2021 гг.

По данным графического анализа исходного временного ряда (рис.3.10) можно предположить о наличии трендовой компоненты в анализируемом периоде: наглядно выражена тенденция роста оборота розничной торговли по РФ, причем характер тенденции близок к линейному развитию.

На рис. 3.10 отчетливо видны сезонные колебания. Наблюдается устойчиво повторяющееся увеличение оборота розничной торговли в IV квартале по сравнению с I кварталом, для которого характерны повторяющиеся из года в год сезонные «спады».

Так как амплитуда сезонных колебаний изменяется с течением времени, увеличивается с ростом тренда, то для описания динамики временного ряда можно предложить мультипликативную тренд-сезонную модель.

Воспользуемся алгоритмом расчета по формулам 3.48-3.52.

Последовательность построения тренд-сезонной мультипликативной модели заключается в следующем:

1) показатели сглаживания исходного временного ряда с помощью скользящей средней определили по формуле:

$$\tilde{y}_i = \frac{\frac{1}{2}y_{t-2} + y_{t-1}. + y_t + y_{t+1} + \frac{1}{2}y_{t+2}}{4}$$

$$\tilde{y}_3 = \frac{\frac{1}{2}\cdot 4184798\ + 4573224\ ,4. + 4900496\ ,2 + 5445817\ ,9 + \frac{1}{2}\cdot 4689759\ ,6}{4} = 4839204\ ,3$$

2) вычислим коэффициенты сезонности:

$$K_S = \frac{y_i}{\tilde{y}_i} = \frac{4900496\ ,2}{4839204\ ,3} \approx 1{,}0127$$

Результаты расчетов скользящей средней и коэффициентов сезонности представлены в таблице 3.27.

Таблица 3.27 – Скользящие средние и коэффициенты сезонности

Годы	№ квартала	y_i	$\tilde{y}_i$	K_S
2017	I кв.	4184798,0	-	-
	II кв.	4573224,4	-	-
	III кв.	4900496,2	**4839204,3**	**1,0127**
	IV кв.	5445817,9	4969696,0	1,0958
2018	I кв.	4689759,6	5111053,9	0,9176
	II кв.	5112196,2	5266835,9	0,9706
	III кв.	5492387,4	5417574,2	1,0138
	IV кв.	6100183,0	5559091,5	1,0973
2019	I кв.	5241300,6	5701620,5	0,9193
	II кв.	5692793,9	5846526,6	0,9737
	III кв.	6052021,7	5990435,1	1,0103
	IV кв.	6699797,3	6129886,3	1,0930
2020	I кв.	5792954,2	6281033,3	0,9223
	II кв.	6256750,1	6475372,5	0,9662
	III кв.	6697241,3	6648463,6	1,0073
	IV кв.	7609291,7	6749849,3	1,1273
2021	I кв.	6268188,6	6829361,2	0,9178
	II кв.	6592601	6874295,0	0,9590
	III кв.	6997485,7	-	-
	IV кв.	7668517,9	-	-

3) Определяем средние показатели сезонности для одноимённых кварталов:

$$\overline{K}_j = \frac{1}{n}\sum K_{Si}$$

т.е. для I квартала средний коэффициент сезонности составит:

$$\overline{K}_1 = \frac{0{,}9176 + 0{,}9193 + 0{,}9223 + 0{,}9128}{4} \approx 0{,}9192$$

Аналогично рассчитываем средние показатели сезонности для других кварталов.

4) Так как сумма средних показателей сезонности не равна 4, проведем их корректировку по формуле:

$$\hat{K}_j = \overline{K}_j \cdot \frac{4(12)}{\sum \overline{K}_j}.$$

Так скорректированный коэффициент сезонности для I квартала составит:

Результаты расчетов средних и скорректированных коэффициентов сезонности представлены в таблице 3.28.

Таблица 3.28 – Расчет коэффициентов со скорретировнной оценкой

№ квартала	Предварительная оценка $\overline{K}_j$	Скорректированная оценка $\hat{K}_j$
I	0,9192	0,9190
II	0,9674	0,9672
III	1,0110	1,0108
IV	1,1034	1,1031
Итого	4,0010	4,0000

5) определяем уровни десезоналированного ряда оборота розничной торговли, результаты представим в таблице 3.29.

$$\frac{y_i}{\hat{K}_j} = \frac{4184798}{0,9190} = 4553620 \quad \text{и т.д.}$$

6) по десезоналированному ряду оборота розничной торговли проводим аналитическое выравнивание по линейному тренду (рис. 3.8), находим выравненные значения ($\hat{y}_i$) считая, что тенденция в исследуемом ряду динамики существует.

Таблица 3.29

Годы	№ квартала	y_i	$\tilde{y}_i$	K_S	$\hat{K}$	$\frac{y_i}{\hat{K}_j}$
2017	I кв.	4184798,0	-	-	0,9190	4553620
	II кв.	4573224,4	-	-	0,9672	4728541
	III кв.	4900496,2	4839204,3	1,0127	1,0108	4848307
	IV кв.	5445817,9	4969696,0	1,0958	1,1031	4936931
2018	I кв.	4689759,6	5111053,9	0,9176	0,9190	5103086
	II кв.	5112196,2	5266835,9	0,9706	0,9672	5285817
	III кв.	5492387,4	5417574,2	1,0138	1,0108	5433895
	IV кв.	6100183,0	5559091,5	1,0973	1,1031	5530149
2019	I кв.	5241300,6	5701620,5	0,9193	0,9190	5703236
	II кв.	5692793,9	5846526,6	0,9737	0,9672	5886133
	III кв.	6052021,7	5990435,1	1,0103	1,0108	5987569
	IV кв.	6699797,3	6129886,3	1,0930	1,1031	6073732
2020	I кв.	5792954,2	6281033,3	0,9223	0,9190	6303509
	II кв.	6256750,1	6475372,5	0,9662	0,9672	6469242
	III кв.	6697241,3	6648463,6	1,0073	1,0108	6625917
	IV кв.	7609291,7	6749849,3	1,1273	1,1031	6898238
2021	I кв.	6268188,6	6829361,2	0,9178	0,9190	6820628
	II кв.	6592601	6874295,0	0,9590	0,9672	6816499
	III кв.	6997485,7	-	-	1,0108	6922964
	IV кв.	7668517,9	-	-	1,1031	6951930

Рисунок 3.11 – Линейный тренд в динамике оборота розничной торговли в РФ за 2017-2021 гг.

Уравнение линейного тренда имеет вид:

$$\hat{y}_i = 136104 \cdot t + 4000000 \text{ , } R^2 = 0{,}9846$$

7) далее рассчитываются уровни временного ряда выравненные по десезоналированному ряду $\hat{y}_i$ и уровни временного ряда y_S, обусловленные влиянием тенденции и сезонности.

Результаты расчетов представлены в таблице 3.30:

-пояснение по столбцу 5:

$\hat{y}_i = 136104 \cdot t + 4000000 = 136104 \cdot 1 + 4000000 = 4136104$ *млн.руб.* и т.д.

-пояснение по столбцу 6:

$y_S = 4136104 \cdot 0{,}9190 = 3801099$ *млн.руб.* и т.д.

Таблица 3.30 – Результаты расчетов

Годы	№ квартала	Десезоналированный ряд $\frac{y_i}{\hat{K}_j}$	t	Выравненные значения по десезоналированному ряду $\hat{y}_i$	Выравненные значения по десезоналированному ряду с учетом сезонности y_S
1	2	3	4	5	6
2017	I кв.	4553620	1	**4136104**	**3801099**
	II кв.	4728541	2	4272208	4131881
	III кв.	4848307	3	4408312	4455765
	IV кв.	4936931	4	4544416	5012843
2018	I кв.	5103086	5	4680520	4301420
	II кв.	5285817	6	4816624	4658415
	III кв.	5433895	7	4952728	5006041

	IV кв.	5530149	8	5088832	5613376
2019	I кв.	5703236	9	5224936	4801740
	II кв.	5886133	10	5361040	5184948
	III кв.	5987569	11	5497144	5556317
	IV кв.	6073732	12	5633248	6213909
2020	I кв.	6303509	13	5769352	5302061
	II кв.	6469242	14	5905456	5711482
	III кв.	6625917	15	6041560	6106594
	IV кв.	6898238	16	6177664	6814442
2021	I кв.	6820628	17	6313768	5802382
	II кв.	6816499	18	6449872	6238016
	III кв.	6922964	19	6585976	6656870
	IV кв.	6951930	20	6722080	7414975
2022	I кв.	-	21	6858184	6302703
	II кв.	-	22	6994288	6764550
	III кв.	-	23	7130392	7207146
	IV кв.	-	24	7266496	8015508

Прогнозируемый объем оборота розничной торговли с учетом сезонности в 2022году составит: в I кв. - 6302703 млн. руб.; во II кв. - 6764550 млн. руб.; в III кв. – 7207146 млн. руб.; в IV кв. - 8015508 млн. руб.

Глава 4. ИССЛЕДОВАНИЕ СТРУКТУРЫ И СТРУКТУРНЫХ ВЗАИМОСВЯЗЕЙ[1]

Внутренним свойством систем любого типа, и социально-экономических - в частности, является движение. Оно может носить двойственный характер: во-первых, сохранять устойчивость объекта; во-вторых, переводить его в новое качественное состояние[2]. Кроме того, движение может быть следствием управленческого воздействия или же быть результатом объективного течения событий.

Движение системы во времени, носящее управляемый характер, мы считаем трансформацией. Для измерения силы и глубины трансформации, проявляющейся в структурных сдвигах, в статистике используются специальные методы, рассчитываются специфические показатели.

В то же время в исследованиях структурных различий и сдвигов остаются нерешённые или спорные вопросы.

1. Нет чёткого разграничения понятий «структурные сдвиги» и «структурные изменения», хотя сами эти категории активно используются в экономической литературе.

Под различиями в структуре совокупности в отдельные периоды времени мы будем понимать дифференциацию удельных весов (долей) частей этих совокупностей. Эти различия, рассматриваемые в динамике, мы можем назвать «структурными сдвигами».

Динамический анализ показателей структуры - одно из важнейших средств изучения закономерностей развития экономических явлений во времени. Структурные сдвиги, в частности, отражают различные темпы роста производства продукции видов экономической деятельности, изменение удельного веса занятого населения в регионе и т.д.

Применительно к сравнению двух структур в пространстве некоторые учёные (в частности, Л.С. Казинец[3]) предлагают термин «структурные различия».

2. Существует достаточное количество статистических показателей в сфере измерения структурных сдвигов (различий), но при этом не определена область применения каждого из них. В частности, нет чётких рекомендаций, какой из критериев применять в конкретной ситуации и для каких целей.

---

[1] Для написания данной главы был использован материал из научной статьи Перстенева Н.П. Критерии классификации показателей структурных различий и сдвигов // Фундаментальные исследования. – 2012. – № 3. – С. 478–482.

[2] Воложанина О.А. Развитие социально-экономических систем: теория и методология: автореф. дис. … д-ра экон. Наук. – СПб., 2011. – 42 с.

[3] Казинец Л.С. Темпы роста и структурные сдвиги в экономике (Показатели планировании и анализа). – М.: Экономика, 1981. – 184 с.

Такая ситуация затрудняет работу экономиста-аналитика, испытывающего сложности при попытке разобраться в этом многообразии показателей.

3. Из всего множества показателей структурных сдвигов (различий) только малая часть шкалирована и имеет какие-либо комментарии по их трактовке и особенностям. Причём эти комментарии зачастую сделаны не авторами-разработчиками, а исследователями, использовавшими эти показатели в своих работах.

4. Данные показатели в основном не отражают ни направленности структурных сдвигов, ни интенсивности изменений тех экономических характеристик, для которых они рассчитываются.

Указанные факторы стали причиной, побудившей нас обратить внимание на возможность систематизации, упорядочения соответствующих статистических показателей. Первым звеном в данном исследовании мы считаем разработку научной классификации индикаторов структурных сдвигов (различий).

Подобная оценка структурных сдвигов возможна с помощью системы обобщающих показателей. В её разработке большая роль принадлежит таким учёным-статистикам, как Л.С. Казинец, К. Гатев, А. Салаи и др.

Так, Л.С. Казинец предложил различать показатели двух видов структурных сдвигов: абсолютных и относительных, каждые из которых имеют самостоятельное значение[1].

Резкость и сила структурных сдвигов зависят от колеблемости (вариации) показателей абсолютных приростов и темпов роста удельных весов. Чем выше колеблемость абсолютных приростов, тем резче и сильнее абсолютные структурные сдвиги; чем выше колеблемость темпов роста, тем соответственно, резче и сильнее относительные структурные сдвиги. Следовательно, возникает вопрос об использовании показателей вариации, колеблемости абсолютных приростов и темпов роста удельных весов отдельных частей изучаемого целого для обобщающей оценки структурных сдвигов[2].

В условиях измерения абсолютных структурных сдвигов классическая формула среднего линейного отклонения трансформируется в следующую:

$$L_{àáñ} = \frac{\sum |d2 - d1|}{n}, \qquad (4.1)$$

где $|d2 - d1|$ - модуль абсолютного прироста долей (удельных весов) в текущем периоде по сравнению с базисным;

n - число градаций.

[1] Казинец Л.С. Темпы роста и структурные сдвиги в экономике (Показатели планировании и анализа). – М.: Экономика, 1981. – 184 с.

[2] Юзбашев М.М., Агапова Т.Н. О показателях вариации долей отдельных групп совокупности // Вестник статистики. -1988. - № 10. – С. 45-54.

Этот показатель Л.С. Казинец назвал линейным коэффициентом абсолютных структурных сдвигов. Статистически его смысл состоит в том, что он представляет собой среднюю арифметическую из модулей абсолютных приростов долей (удельных весов) всех частей сравниваемых целых.

Данный коэффициент характеризует среднюю величину отклонений от удельных весов, то есть показывает, на сколько процентных пунктов в среднем отклоняются друг от друга удельные веса частей в сравниваемых совокупностях.

Чем больше величина линейного коэффициента абсолютных структурных сдвигов, тем больше в среднем отклоняются друг от друга удельные веса отдельных частей за два сравниваемых периода, тем сильнее абсолютные структурные сдвиги. Если структуры за эти периоды совпадают (т.е. d2 - d1 = 0), то данный коэффициент будет равен нулю.

На основе формулы среднего квадратического отклонения Л.С. Казинец построил другой показатель абсолютных структурных сдвигов:

$$\sigma_{àáñ} = \sqrt{\frac{\sum (d2 - d1)^2}{n}} \qquad (4.2)$$

По мнению Л.С. Казинца, формула (2) есть частный случай формулы простого среднего квадратического отклонения в условиях измерения абсолютных структурных сдвигов. Этот показатель в литературе получил название «квадратический коэффициент абсолютных структурных сдвигов».

Он позволяет количественно оценить, на сколько процентных пунктов в среднем отклоняются друг от друга удельные веса частей в сравниваемых совокупностях.

Основу вычисления сводных показателей относительных структурных сдвигов составляют темпы роста удельных весов, рассматриваемых как часть целого, степень вариации которых и служит их сводной обобщающей характеристикой. Показатель относительных структурных сдвигов, основанный на среднем взвешенном линейном отклонении, вычисляется при использовании долей по формуле:

$$L_{îòí} = \sum \left| \frac{d2}{d1} - 1 \right| \cdot d1 \qquad (4.3)$$

Этот показатель называется линейным коэффициентом относительных структурных сдвигов.

Он показывает не среднюю скорость, а среднюю интенсивность изменения удельных весов отдельных частей совокупности. Иначе говоря, линейный коэффициент позволяет установить, на сколько процентов по сравнению с базисным периодом, удельные веса которого принимаются за единицу (100 %), изменился в среднем удельный вес частей целого, то есть каков средний относительный (а не «абсолютный» - по Л.С. Казинцу) прирост

удельного веса частей целого (взятых, естественно, по их абсолютному значению).

В случае тождественности структур сравниваемых совокупностей рассматриваемый коэффициент равен нулю, так как d2:d1 = const. И, наоборот, равенство нулю этого коэффициента означает тождественность структур сравниваемых совокупностей.

Чем больше количественное значение линейного коэффициента относительных структурных сдвигов, тем более резкими являются относительные структурные сдвиги, и, наоборот, менее резкие структурные сдвиги характеризуются меньшими значениями линейного коэффициента относительных структурных сдвигов.

Рассмотрим обобщающий показатель относительных структурных сдвигов, основанный на среднем взвешенном квадратическом отклонении. Он может быть вычислен по нижеследующей формуле:

$$\sigma_{îòí} = \sqrt{\sum\left(\frac{d2}{d1} - 1\right) \cdot d1} \qquad (4.4)$$

Этот показатель Л.С. Казинец предлагает называть квадратическим коэффициентом относительных структурных сдвигов.

Этот коэффициент показывает, на сколько в среднем отклоняются коэффициенты (темпы) роста отдельных частей совокупности от их среднего значения, равного единице (100 %), или, иначе говоря, какова средняя квадратическая величина относительного отклонения удельных весов.

Для дополнения системы показателей Казинца Перстеневой Н.П. [1] были предложены два коэффициента, которые являются модификациями линейного и квадратического коэффициентов структурных сдвигов. Они являются нормированными, то есть их значения варьируются в пределах от 0 (идентичность структур) до 1 (полное различие структур).

В знаменателе отношения долей, по нашему мнению, более целесообразно использовать не удельный вес базисного периода, а средний удельный вес (по двум периодам).

В этом случае формулы (4.5) и (4.6) примут следующий вид:

1) модификация линейного коэффициента:

$$L_{îòí\ (Ì)} = \sum \left| \frac{d2}{\frac{d1 + d2}{2}} - 1 \right| \cdot d1 \qquad (4.5)$$

2) модификация квадратического коэффициента:

[1] Перстенёва Н.П. Методология статистического исследования структурно-динамических изменений (на примере экономики Самарской области): дис. ... канд. экон. наук. – Самара, 2003. – 141 с.

$$\sigma_{\text{îòí (Ì)}} = \sqrt{\sum \left(\frac{d2}{\frac{d1+d2}{2}} - 1 \right)^2} \cdot d1 \qquad (4.6)$$

В теории и практике статистического анализа особое место принадлежит сводным показателям оценки структурных сдвигов. В отличие от рассмотренных ранее показателей Л.С. Казинца они, как правило, имеют более удобную и компактную шкалу значений - от 0 до 1, и в этом случае каждый отдельный коэффициент сам по себе имеет вполне определённый познавательный смысл и не требует обязательного сравнения с другим.

Болгарский статистик К. Гатев предложил нормировать линейный и квадратический коэффициенты абсолютных структурных сдвигов путём деления на их максимально возможную величину. Так как $|d2 - d1| \le 2$ и $\sum (d2 - d1) \cdot 2 \le 2$, то в результате получим следующие формулы[1]:

1) нормированного линейного коэффициента абсолютных структурных сдвигов:

$$L_{íîðì} = \frac{1}{2} \cdot \sum |d2 - d1| \qquad (4.7)$$

2) нормированного квадратического коэффициента абсолютных структурных сдвигов:

$$\sigma_{íîðì} = \sqrt{\frac{1}{2} \cdot \sum (d2 - d1)^2} \qquad (4.8)$$

С.В. Курышева в своих исследованиях интенсивности структурных сдвигов в составе рабочих промышленности[2] предлагает модифицировать показатель (8), добавив ему информативности. Она, в частности, предлагает учесть число доминантных групп, то есть тех, что в основном формируют структуру совокупности. Модифицированный показатель получил название «скорректированный нормированный квадратический коэффициент абсолютных структурных сдвигов»:

$$\sigma^{*}_{íîðì} = \sqrt{\frac{L}{2} \cdot \sum (d2 - d1)^2} \qquad (4.9)$$

где L - число доминантных групп.

---

[1] Агапова Т.Н. Статистические методы изучения структуры: дис. ... д-ра экон. наук. – СПб., 1996. – 215 с.

[2] Курашева С.В. Статистический анализ содержания труда рабочих. – Красноярск: Изд-во Красноярского ун-та, 1990. -184 с.

Ещё один показатель - интегральный коэффициент структурных сдвигов - был предложен К. Гатевым[1]:

$$\hat{E}_{\text{èíò}} = \sqrt{\frac{\sum (d2 - d1)^2}{\sum d2^2 + \sum d1^2}} \tag{4.10}$$

Как и показатели абсолютных структурных сдвигов Л.С. Казинца, данный коэффициент в принципе также основан на разностях удельных весов, однако при данном способе нормирования он учитывает значения самих удельных весов обоих периодов. В качестве недостатка можно отметить отсутствие реального смысла знаменателя.

Венгерский учёный А. Салаи предложил свой вариант обобщающего показателя структурных сдвигов:

$$I_c = \sqrt{\frac{\sum \frac{(d2 - d1)^2}{(d2 + d1)^2}}{n}} \tag{4.11}$$

Важный недостаток коэффициента: его значения зависят от числа градаций.

Позднее были сделаны попытки найти критерий, который был бы свободен от указанных недостатков. Так, в конце XX в. был разработан индекс В. Рябцева[2]:

$$I_R = \sqrt{\frac{\sum (d2 - d1)^2}{\sum (d2 + d1)^2}} \tag{4.12}$$

Рассмотренные показатели являются наиболее востребованными в практических исследованиях. Учитывая достаточно большое количество этих индикаторов, мы предложили ряд критериев, позволяющих их классифицировать - нормированность, универсальность, чувствительность, направленность.

1. Нормированность.

Этот критерий означает, что у показателя имеется интервал возможных значений - как правило, от 0 до 1. При этом 0 означает «идентичность структур», а 1 - «полное различие структур». Наличие крайних значений позволяет разработать шкалу, определяющую статус того или иного значения показателя. Естественно, что в этом вопросе присутствует некоторый элемент научного творчества, и каждый учёный теоретически мог бы составить

[1] Социальная статистика/ Под ред. И.И.Елисеевой. – М.: Финансы и статистика, 2002. – 480 с.

[2] Зарова Е.В., Чудилин Г.А. Региональная статистика: учебник. – М.: Финансы и статистика, 2006. – 624 с.

собственную шкалу. Этот вопрос не является принципиальным; важно, чтобы соблюдался общий подход к пониманию смысла крайних значений.

Отметим, что нормированность и шкалирование облегчают аналитику интерпретацию полученных результатов.

Критерий нормированности соблюдается в показателях (3, 5-8, 10-12), причём (3) варьируется от 0 до 200 %. В настоящее время шкала разработана только для показателя (12).

2. Универсальность.

Она может рассматриваться в двух аспектах: уровневом и пространственно-временном.

В первом случае речь идёт о возможности применения показателя при исследованиях на любом уровне экономики - макро-, мезо-, микро-. Например, интегральный коэффициент структурных сдвигов Гатева позиционируется как показатель социальной статистики (хотя это утверждение практически не обосновывается).

Во втором случае рассматривается применение того или иного индикатора для анализа как структурных различий (пространственных), так и структурных сдвигов (временных).

3. Чувствительность.

Она понимается как эластичность того или иного количественного показателя в зависимости от изменения удельных весов изучаемых совокупностей. Иными словами, чувствительность характеризует количественные изменения значений индикатора при определённых флуктуациях в структуре.

Наиболее чувствительными мы можем признать:

квадратический коэффициент относительных структурных сдвигов Казинца, который изменяется от 0 до 1000 %, однако имеет одну важную особенность: если в базисном периоде удельные веса каких-либо групп (двух и более) максимально близки к нулю (менее 0,3 %), то его значение может превышать 1000;

обобщающий показатель структурных сдвигов Салаи. Он не может быть рассчитан, если доли в каждом периоде у какой-либо группы равны 0. В этом случае будут нарушены правила элементарной математики: в числителе подкоренной дроби получится деление на 0. Кроме того, если хотя бы одна доля в одном из периодов равна 0 (даже при условии идентичности всех остальных), значение этого коэффициента резко возрастает и практически достигает 1.

Кроме того, чувствительность может означать различия в количественном изменении нескольких показателей при равном изменении структуры. С другой стороны, одинаковое количественное изменение нескольких индикаторов не обязательно может свидетельствовать об адекватном (одинаковом) изменении в структуре сравниваемых совокупностей.

4. Направленность.

Этот критерий определяет вектор развития, приближение или отдаление от «эталонной» структуры, положительные или отрицательные структурные сдвиги (различия). Например, увеличение доли бракованной продукции на предприятии свидетельствует об отрицательных структурных сдвигах (и наоборот). В какой-то мере ответ на эти вопросы дают индексы влияния структурных сдвигов, однако ни один из рассмотренных нами ранее статистических показателей не отвечает на подобные вопросы.

Предложенные нами критерии классификации являются начальным этапом исследования, в ходе которого мы ставим задачу не поиска оптимального (универсального) критерия измерения структурных сдвигов (различий), а систематизацию накопленных научных знаний по данной проблематике. Это позволит выявить особенности, достоинства и недостатки отдельных индикаторов, определить возможные сферы их применения.

Пример 4.1. Анализ структуры и расчет показателей структурных сдвигов

Одним из основных показателей, характеризующих уровень жизни населения, является уровень образования населения. Рассмотрим структуру охвата населения Оренбургской области образованием на начало года (табл. 4.1).

Таблица 4.1 - Структура уровня образования населения в Оренбургской области

Показатели	2012г.		2021г.		Изменения в структуре (+,-),%
	Тыс.чел.	В % к итогу	Тыс.чел.	В % к итогу	
Численность обучающихся - всего	327,9	100	206,1	100	327,9
в том числе в: общеобразовательных учреждениях	254,16	77,51	124,67	60,49	254,16
учреждениях начального профессионального образования	23,74	7,24	11,60	5,63	23,74
учреждениях среднего профессионального образования	30,82	9,4	23,10	11,21	30,82
учреждениях высшего профессионального образования	19,08	5,82	46,25	22,44	19,08
аспирантуре и докторантуре	0,07	0,02	0,45	0,22	0,07

Анализируя рассчитанные показатели, представленные в таблице 4.1 можно сделать следующие выводы: за рассматриваемый период с 2012г. по 2021г. в Оренбургской области наиболее значительные увеличения произошли в структуре уровня образования населения по следующим категориям: на 1,81 п.п. увеличилась доля обучающихся в учреждениях среднего профессионального образования, на 16,62 п.п. увеличилась доля населения, учащихся в учреждениях высшего профессионального образования и на 0,20

п.п. увеличилась доля обучающихся в аспирантуре и докторантуре. Численность обучающихся в общеобразовательных учреждениях и учреждениях начального образования сократилась на 17,02 п.п. и 1,61 п.п., соответственно.

Проанализируем структурные изменения, произошедшие в структуре уровня образования населения Оренбургской области за 2012 – 2021 гг. с помощью следующих относительных показателей:

1. Линейный коэффициент «абсолютных» структурных сдвигов:

$$L_{àáñ} = \frac{\sum |d2 - d1|}{n}$$

Для наших данных по образованию населения d = 7,45 п.п., это значит, что удельный вес отдельных статей изменился в среднем на 7,45 п.п.

2. Квадратический коэффициент «абсолютных» структурных сдвигов:

$$\sigma_{àáñ} = \sqrt{\frac{\sum (d2 - d1)^2}{n}}$$

Для нашего исследования по уровню образования населения σ =10,69 п.п., это свидетельствует о скорости изменения удельных весов отдельных частей совокупности, следовательно, скорость изменения удельных весов по уровню образования населения от года к году составляет 10,69 п.п.

Пример 4.2. Анализ структуры и расчет показателей структурных сдвигов

Структура производства продукции растениеводства в Оренбургской области в рассматриваемом периоде менялась в сторону сокращения долей картофеля и кукурузы (на 0,6 и 14,6 п.п. соответственно) и увеличения долей зерна, подсолнечника (на 5,9 и 9,4 п.п. соответственно) и овощей не изменилось (рис. 4.1).

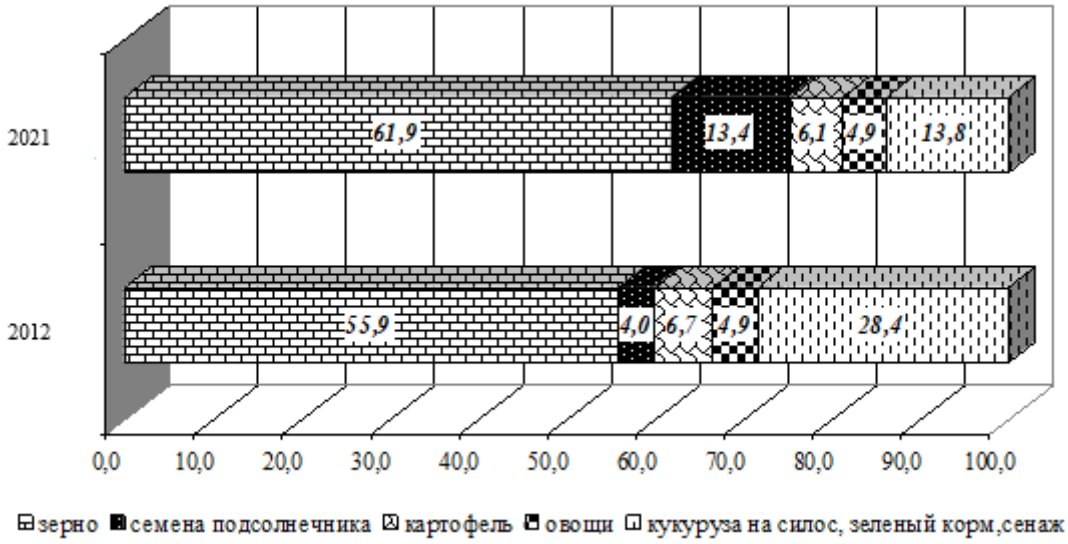


Рисунок 4.1 – Структура производства продукции растениеводства в Оренбургской области, %

Оценить существенность произошедших структурных изменений в производстве продукции растениеводства можно с помощью индексов структурных сдвигов: коэффициента абсолютных сдвигов, коэффициента К. Гатева, коэффициента А. Салаи, индекса Рябцева. В таблице 4.2 представлены исходные данные для расчета этих показателей.

Таблица 4.2 – Динамика структуры производства продукции растениеводства в Оренбургской области

Показатели	2012г.	2014г.	2016г.	2018г.	2020г.	2021г.	Изменения в структуре в 2021г. по сравнению с 2012г. (+/-)
Зерно	55,9	70,6	53,9	59,1	70,9	61,9	5,9
Семена подсолнечника	4,0	2,6	5,3	8,8	8,1	13,4	9,4
Картофель	6,7	6,1	11,6	8,2	6,0	6,1	-0,6
Овощи	4,9	5,0	8,5	5,1	4,4	4,9	0,0
Кукуруза на силос, зеленый корм, сенаж	28,4	15,7	20,5	18,8	10,5	13,8	-14,6
ИТОГО	100,0	100,0	100,0	100,0	100,0	100,0	x

Сводная таблица индексов структурных сдвигов представлена ниже (таблица 4.3).

Таблица 4.3 – Значения коэффициентов структурных сдвигов в производстве продукции растениеводства в Оренбургской области

Название показателя	Формула расчета	Значение, в коэффициентах
Коэффициент абсолютных структурных сдвигов	$L_{àáñ} = \frac{\sum \lvert d2 - d1 \rvert}{n}$	0,677
Коэффициент К.Гатева	$\hat{E}_{èíò} = \sqrt{\frac{\sum (d2 - d1)^2}{\sum d2^2 + \sum d1^2}}$	0,432
Коэффициент А.Салаи	$I_c = \sqrt{\frac{\sum \frac{(d2 - d1)^2}{(d2 + d1)^2}}{n}}$	0,567
Индекс Рябцева	$I_R = \sqrt{\frac{\sum (d2 - d1)^2}{\sum (d2 + d1)^2}}$	0,345

Вычисленные значения коэффициентов удовлетворяют условию: $J_R \langle K_\Gamma \langle J_c$, т.е. критерий Рябцева имеет среди других критериев наименьшее значение, интерпретируемое как существенный уровень различий.

Данные структурные изменения объясняются многими факторами, в том числе дефицитом финансовых средств у сельскохозяйственных товаропроизводителей, в связи, с чем они вынуждены были переключиться на более доходные виды растениеводческой продукции (подсолнечник, овощи, картофель).

<u>Пример 4.3 Определение показателей структурных сдвигов в динамике структуры доходов населения.</u>

Имеются следующие условные данные об источниках денежных доходов населения страны в текущих ценах:

Источники денежных доходов	Удельный вес, % к итогу	
	2018 г.	2019 г.
Денежные доходы – всего, в т.ч.	100,0	100,0
1. доходы от предпринимательской деятельности	11,9	10,0
2. оплата труда	65,8	67,5
3.социальные выплаты	15,2	11,6
4.доходы от собственности	5,2	8,9
5. другие доходы	1,9	2,0

Определите:

1) линейный коэффициент абсолютных структурных сдвигов

2) средний квадратический коэффициент абсолютных структурных сдвигов

3) средний квадратический коэффициент относительных структурных сдвигов

4) интегральный коэффициент структурных различий К. Гатева, коэффициент А. Салаи

Решение:

Линейный коэффициент абсолютных структурных сдвигов представляет собой сумму приростов удельных весов, взятых без учёта знака, деленную на число структурных частей. Определяется по формуле:

$$\overline{\Delta}_{d_1-d_0} = \frac{\sum_{i=1}^{n} |d_1 - d_0|}{n},$$

где d_1 – удельный вес (доля) части совокупности в отчетном периоде времени; d_0- удельный вес (доля) в базисном периоде времени.

Этот показатель отражает то среднее изменение удельных весов (в процентных пунктах), которое имело место за рассматриваемый период.

5) *квадратический коэффициент абсолютных структурных сдвигов* позволяет получить свободную оценку скорости изменения удельных весов отдельных частей совокупности:

$$\sigma_{d_1-d_0}=\sqrt{\frac{\sum_{i=1}^{n}(d_1-d_0)^2}{n}}$$

6) *средний квадратический коэффициент относительных структурных сдвигов* отражает тот средний относительный прирост удельного веса (в процентах), который наблюдается за рассматриваемый период:

$$\sigma_{\frac{d_1}{d_0}}=\sqrt{\sum_{i=1}^{n}\frac{(d_1-d_0)^2}{d_0}}\cdot 100$$

7) *коэффициент Гатева* определяется по формуле:

$$k_\gamma=\sqrt{\frac{\sum_{i=1}^{n}(d_1-d_0)^2}{\sum_{i=1}^{n}d_1^2+\sum_{i=1}^{n}d_0^2}}$$

8) *индекс Салаи* определяется по формуле:

$$I_c=\sqrt{\frac{\sum_{i=1}^{n}\left(\frac{d_1-d_0}{d_1+d_0}\right)^2}{n}}\ .$$

Коэффициенты структурных сдвигов изменяются от 0 до 1. Чем ближе значение коэффициентов к 1, тем существенней уровень различий в изучаемых структурах.

Чтобы вычислить выше перечисленные коэффициенты, построим расчетную таблицу (табл. 15.1).

Для расчета показателей структурных различий воспользуемся коэффициентами.

Таблица 4.4 – Расчетная таблица

Источники денежных доходов	Удельный вес, % к итогу		Расчетные графы							
	2018 г. d_0	2019 г. d_1	$\lvert d_1-d_0\rvert$	$(d_1-d_0)^2$	$\frac{(d_1-d_0)^2}{d_0}$	d_0^2	d_1^2	d_0+d_1	$\frac{d_1-d_0}{d_1+d_0}$	$(\frac{d_1-d_0}{d_1+d_0})^2$
1	11,9	10,0	1,9	3,61	0,30	141,6	100,0	21,9	0,087	0,0075
2	65,8	67,5	1,7	2,89	0,04	4329,6	4556,3	133,3	0,013	0,0002
3	15,2	11,6	3,6	12,96	0,85	231,0	134,6	26,8	0,134	0,0180
4	5,2	8,9	3,7	13,69	2,63	27,0	79,2	14,1	0,262	0,0688
5	1,9	2,0	0,1	0,01	0,01	3,6	4,0	3,9	0,025	0,0006
Итого	100,0	100,0	11,0	33,16	3,83	4732,8	4874,1	x	x	0,0951

1. Линейный коэффициент абсолютных структурных сдвигов равен:

$$\overline{\Delta}_{d_1-d_0}=\frac{\sum_{i=1}^{n}|d_1-d_0|}{n}=\frac{11{,}0}{5}=2{,}2\%$$

В отчетном периоде по сравнению с базисным периодом удельный вес отдельных источников денежных доходов населения изменился на 2,2%.

2. Средний квадратический коэффициент абсолютных структурных сдвигов равен:

$$\sigma_{d_1-d_0}=\sqrt{\frac{\sum_{i=1}^{n}(d_1-d_0)^2}{n}}=\sqrt{\frac{33{,}16}{5}}=2{,}58\%$$

3. Средний квадратический коэффициент относительных структурных сдвигов равен:

$$\sigma_{\frac{d_1}{d_0}} = \sqrt{\frac{\sum_{i=1}^{n} \frac{(d_1 - d_0)^2}{d_0}}{n}} \cdot 100 = \sqrt{\frac{3{,}83}{5} \cdot 100} = 8{,}76\%$$

За год удельный вес каждого источника доходов в среднем изменился на 8,76%.

4. Коэффициент К. Гатева:

$$k_\gamma = \sqrt{\frac{\sum_{i=1}^{n} (d_1 - d_0)^2}{\sum_{i=1}^{n} d_1^2 + \sum_{i=1}^{n} d_0^2}} = \sqrt{\frac{33{,}16}{4874{,}1 + 4732{,}8}} = 0{,}059$$

Коэффициент А. Салаи:

$$k_c = \sqrt{\frac{\sum_{i=1}^{n} \left(\frac{d_1 - d_0}{d_1 + d_0} \right)^2}{n}} = \sqrt{\frac{0{,}0951}{5}} = 0{,}138$$

Вычисленные значения критериев показывают, что наблюдается низкий уровень различий структур источников денежных доходов в отчетном году по сравнению с базисным.

Глава 5. ВЫБОРОЧНЫЙ МЕТОД ИССЛЕДОВАНИЯ ЭКОНОМИЧЕСКИХ ПРОЦЕССОВ

Под выборочным исследованием понимается такое несплошное наблюдение, при котором статистическому обследованию подвергаются единицы изучаемой совокупности, отобраны случайным способом. В нашем случае выборочному исследованию будет подвержена молочная продукция рынка Оренбургской области. Данное выборочное исследование ставит перед собой задачу – по обследуемой части (n) дать характеристику всей совокупности (N) при условии соблюдения всех правил и принципов проведения статистического наблюдения и научно организованной работы по отбору единиц.

Для аналитического исследования использование выборочного метода актуально по причинам, того что, сплошное исследование, состоящих из десятков и сотен тысяч единиц, потребовал бы огромных материальных и иных затрат. Использование же выборочного обследования позволяет значительно сэкономить силы и средства, что имеет немаловажное значение.

Наряду с экономией ресурсов, одной из причин исследования выборочного метода, является то, что это важнейший источник статистической информации, которая дает возможность значительно ускорить получение необходимых данных. Фактор времени важен для статистического исследования особенно в условиях изменяющейся социально-экономической ситуации.

Роль выборочного обследования в получении статистических данных возрастает в силу возможности – когда это необходимо – расширения программы наблюдения. Так как исследованию подвергается сравнительно небольшая часть всей совокупности, можно более широко и детально изучить отдельные единицы и их группы.

Следует также отметить, что на практике можно столкнуться со специфическими задачами изучения массовых процессов, которые решаются лишь с помощью методологии выборки. К таким задачам относится исследование, например предпочтений покупателей и удовлетворенность насыщенностью рынка.

Применение выборочного метода наблюдения включает следующие этапы:

1) определение генеральной совокупности и единиц наблюдения, обладающих первичной информацией, необходимой для решения задач обследования;

2) создание основы выборки;

3) формирование выборочной совокупности путем отбора элементов основы;

4) распространение собранных по выборке данных на генеральную совокупность.

Последний этап зависит от примененного способа отбора элементов в выборку и используемой формулы оценивания характеристик генеральной совокупности по данным выборки.

В статистической практике выборки извлекаются из конечных списочных основ. Однако единица основы, единица отбора и единица наблюдения могут отличаться. Например, это обычная ситуация при обследованиях населения и сельскохозяйственного сектора.

При рассмотрении любой схемы извлечения выборки должны быть учтены два фактора:

1) использовалась или нет вероятностная процедура;

2) наличие или отсутствие объективности в действиях специалиста, формирующего выборку.

Смысл объективности ясен и однозначен: любой специалист, производящий отбор, получил бы ту же самую выборку, т.е. выборку с теми же самыми свойствами. Субъективность означает, что специалисту, производящему отбор, позволено опираться на собственное суждение или интуицию относительно того, что является «хорошей» выборкой.

Рассматривая каждый из этих факторов на двух уровнях, можно выделить четыре типа выборок (таблица 5.1)

Таблица 5.1 - Типы выборок в статистической методологии

Роль, которую играет специалист, осуществляющий отбор	Процедура отбора	
	Вероятностная	Невероятностная
Объективная	Выборки, сформированные вероятностным (случайным) образом	Выборки, сформированные на основе направленного отбора
Субъективная	Выборки, сформированные квазислучайным образом	Выборки, сформированные на основе суждения эксперта

В статистической практике используются все четыре типа выборок. Однако обычно отдают предпочтение вероятностным (случайным) выборкам как наиболее объективным, поскольку имеется хорошо обоснованная теория, позволяющая понимать поведение таких выборок и оценивать их свойства (качество) отображения характеристик всей совокупности. Свойства и объективная ценность других выборок известны в меньшей мере.

Имеются два типа выборок, основывающихся на вероятностном способе отбора: выборки, отбираемые по объективным правилам вероятностного

(случайного) отбора, и выборки, отбираемые, строго говоря, не по этим правилам (квазислучайные).

В теории выборочных обследований рассматриваются выборки, извлеченные из совокупностей (основ выборки), содержащих некоторое конечное число единиц N. Эти единицы различимы между собой и число различных выборок объема n, которые могут быть извлечены из списка N единиц, равно числу сочетаний C_N^n.

В выборочных статистических обследованиях в целях расчета параметров совокупности основное внимание направлено на изучение определенных свойств единиц, которые измеряются и фиксируются в процессе наблюдения для каждой единицы, включенной в выборку. Эти свойства называют признаками.

Главным вопросом методологии выборочного наблюдения является обеспечение приемлемого уровня ошибок получаемых значений характеристик совокупности, в том числе по требуемым разрезам, например, отраслям экономики, формам собственности и регионам России.[17]

Полученные в результате выборочного наблюдения характеристики практически всегда несколько отличаются от характеристик генеральной совокупности. Эти отличия называются ошибками выборки (или репрезентативности), которые могут быть систематическими или случайными.

Систематические ошибки имеют место в том случае, когда нарушен принцип случайности отбора и в выборку попали единицы, обладающие какими-либо свойствами, не характерными для всех единиц генеральной совокупности. Случайные ошибки обусловлены тем обстоятельством, что даже при тщательной организации выборка не может в точности воспроизвести генеральную совокупность. В отличие от ошибок систематических, случайные ошибки являются вполне допустимыми, если они малы и могут быть оценены статистически.

Для измерения ошибки выборки, а также сравнения двух оценок, т.е. выявления более эффективной оценки, используют средний квадрат ошибки оценки (СКО), который измеряет ошибку относительно оцениваемого параметра совокупности (5.1).

$$\begin{aligned} &СКО(\hat{\theta}) = E(\hat{\theta} - \theta) = E\left[(\hat{\theta} - \bar{\theta}) + (\bar{\theta} - \theta)\right]^2 = \\ &= E(\hat{\theta} - \bar{\theta})^2 + 2(\hat{\theta} - \theta)E(\hat{\theta} - \bar{\theta}) + (\hat{\theta} - \theta)^2 = V(\hat{\theta}) + B^2, \\ &так\_как\_E(\hat{\theta} - \bar{\theta}) = 0, E(\hat{\theta}) = \bar{\theta} \end{aligned} \qquad (5.1)$$

где E - символ, заменяющий выражение "математическое ожидание величины";

$\hat{\theta}$ - оценка некоторой характеристики совокупности θ, получаемая согласно некоторой схеме отбора и примененной формуле оценивания;

$\bar{\theta}$ - математическое ожидание

$\hat{A} = \bar{\theta} - \hat{\theta}$ - смещение оценки;

$V(\hat{\theta})$ - дисперсия оценки.

Таким образом, СКО является критерием достоверности оценки, который характеризует величину отклонений от истинного значения характеристики совокупности $\bar{\theta}$.

Поскольку на практике трудно проследить, чтобы оценки не давали никаких смещений, для характеристики оценки используется понятие «точности», относящееся к величине отклонений от усредненного значения $\bar{\theta}$.

Степень точности оценки обычно характеризуется ее дисперсией, стандартной ошибкой, коэффициентом вариации (относительной стандартной ошибкой) и доверительным интервалом.

Точность какой-либо оценки, полученной по выборке, зависит от двух факторов: от способа, которым оценка вычисляется по данным выборки, и от способа формирования самой выборки.

В выборочных обследованиях способ оценивания называется состоятельным, если оценка становится в точности равной оцениваемому параметру для совокупности при n = N, т.е. когда выборку составляет вся совокупность. Очевидно, что при простом случайном отборе выборочное среднее $\bar{y}$ и произведение $N\bar{y}$ представляют собой состоятельные оценки соответственно среднего и суммарного значений для совокупности.

В данном контексте способ оценивания называется несмещенным, если среднее значение оценки, взятое по всем возможным выборкам данного объема n, в точности равно истинному значению для совокупности, и это утверждение справедливо для любой конечной совокупности значений y_i и для любого n. Например, при простом случайном отборе выборочное среднее $\bar{y}$ - несмещенная оценка среднего значения признака, $N\bar{y}$ - несмещенная оценка суммарного значения Y для совокупности, где $\bar{y}$ - среднее значение признака y_i по выборке.

В теории и практике выборочных обследований часто приходится рассматривать смещенные оценки. Это обусловлено следующими причинами. Во-первых, в некоторых случаях, особенно при оценивании отношений двух величин, смещенные оценки дают более достоверные результаты, чем несмещенные. Во-вторых, даже в случае использования теоретически несмещенных оценок ошибки наблюдения и неполучение ответов от респондентов могут привести к смещениям в распространенных результатах.

Кратко опишем некоторые, наиболее часто используемые в статистической практике способы формирования вероятностной выборки.

Простым случайным отбором называется способ, при котором извлечение единиц из совокупности для обследования осуществляется методом жеребьевки или с использованием таблиц или генератора случайных чисел без деления этой совокупности на какие-либо классы или группы.[9]

Простую случайную выборку получают, отбирая последовательно единицу за единицей. Единицы в совокупности нумеруются числами от 1 до N, после чего выбирается последовательность n случайных чисел,

заключенных между 1 и *N*. Единицы совокупности, имеющие эти номера, составляют выборку. На каждом этапе отбора такой процесс обеспечивает для всех еще не выбранных номеров равную вероятность быть отобранными. Легко показать, что равную вероятность быть отобранными имеют все C_N^n возможных выборок.

Уже отобранные номера исключаются из списка, иначе одна и та же единица могла бы попасть в выборку более одного раза. Поэтому такой отбор называется отбором без возвращения. Отбор с возвращением легко осуществим, но им, за исключением особых случаев, пользуются редко, поскольку нет особых оснований допускать, чтобы одна и та же единица встречалась в выборке дважды.

При простом случайном отборе для получения выводов о параметрах совокупности используют выборочное среднее в качестве оценки среднего значения признака совокупности, а дисперсию признака по выборке - для оценки дисперсии признака совокупности. Для простой случайной выборки усредненные выборочные средние и дисперсии точно равны среднему и дисперсии признака совокупности (приложение А).

Другие методы отбора часто оказываются предпочтительнее простого случайного отбора по соображениям удобства или повышения точности. Однако простая случайная выборка - наипростейший вид объективной вероятностной выборки, она служит основой для многих более сложных ее видов.

Расслоенный случайный отбор - это отбор, предусматривающий предварительное разделение совокупности, содержащей N единиц, на слои и проведение простого случайного отбора в каждом слое.

При расслоенном случайном отборе совокупность, содержащая N единиц, сначала подразделяется на подсовокупности, состоящие, соответственно, из N_1, N_2, ..., N_L единиц. Эти подсовокупности не содержат общих единиц и вместе исчерпывают всю совокупность:

$$N_1 + N_2 + ... + N_L = N \qquad (5.2)$$

Такие подсовокупности называются слоями. Для того чтобы можно было полностью воспользоваться преимуществами этого метода отбора, значения N_L должны быть известны. Когда слои определены, из каждого слоя извлекается простая случайная выборка, причем отбор в разных слоях производится независимо. Объемы выборок внутри слоев обозначаются соответственно через n_1, n_2, ..., n_L и, следовательно:

$$n_1 + n_2 + ... + n_L = n. \qquad (5.3)$$

Объем выборки из каждого слоя может быть пропорционален объему (размеру) этого слоя или определяется степенью дифференциации признака в данном слое, или устанавливается в соответствии с некоторым составным критерием, учитывающим оба названных фактора. Однако реализация второго и третьего вариантов на практике затруднена, так как они предполагают наличие информации о вариации признаков еще до проведения обследования, получаемой, например, по результатам предшествующих обследований.

Расслоение - довольно распространенный прием, что обусловлено многими причинами. Перечислим некоторые из них.

Расслоение можно рассматривать как процедуру извлечения выборок, в которой на простой случайный отбор наложены некоторые ограничения или условия. При выполнении определенных условий и наложении правильных ограничений можно получить значительный выигрыш в точности и, как правило, с малыми дополнительными затратами либо вовсе без них. В другом, но близком смысле, расслоение - это способ включения знаний об общей совокупности и ее совокупностях по признакам в процедуру отбора таким образом, чтобы повысить точность оценивания.

При расслоенном случайном отборе управление обследованием значительно упрощено. Однако сама процедура предполагает знание объемов слоев, общего числа единиц в выборке, а также определение долей отбора в каждом слое.

Расслоение может дать выигрыш в точности при оценивании характеристик всей совокупности. Часто неоднородную совокупность удается расслоить на подсовокупности (слои), каждый из которых внутренне однороден. Если каждый слой однороден в том смысле, что результаты измерений в нем мало изменяются от единицы к единице, то можно получить точную оценку среднего значения для любого слоя по небольшой выборке в этом слое. Затем эти оценки можно объединить в одну точную оценку для всей совокупности (приложение Б)

Гнездовой отбор - способ формирования выборки, при котором единица отбора состоит из группы или гнезда более мелких единиц, называемых элементами. Таким образом, гнездовая выборочная единица - группа элементов, которая в процессе извлечения выборки рассматривается как одна единица. В простейшем случае элементы, составляющие гнездо, либо входят в выборку как группа, либо не входят в нее вообще.

Например, если гнездом являются все квартиры в жилом квартале города, то квартиры из этого жилого квартала либо входят в выборку, либо нет - в зависимости от того, оказался ли отобранным этот квартал или нет. Ни в коем случае не может быть, чтобы одна часть квартир из жилого квартала попала в выборку, а другая из этого же квартала не попала. Это было бы возможно, если бы извлекалась, например, случайная выборка из списка квартир города.

Гнездовой метод отбора единиц наблюдения наиболее характерен для выборочных обследований в таких отраслях статистики, как статистика сельского хозяйства и статистика населения. Широкое применение гнездового отбора в статистической практике обусловлено двумя основными причинами.

Первая из них заключается в том, что для обследования может не существовать основы выборки (списка элементов совокупности), а ее составление или невозможно, или обошлось бы очень дорого. Например, эта ситуация имеет место, когда для обследований населения нет полных и неустаревших его списков. Однако по картам подлежащие обследованию районы могут быть разделены на территориальные участки с легко

идентифицируемыми границами. Относясь к таким участкам как к гнездам, возможно, решить задачу построения списка единиц отбора.

Вторая причина состоит в том, что, даже если имеется списочная основа элементов, экономические соображения могут диктовать выбор более крупных единиц отбора.

В статистической практике определение единицы отбора в значительной степени зависит от природы статистического исследования, от того, какого рода имеется основа, а также от ряда других факторов, которые не всегда поддаются количественной оценке. Этот выбор может быть сделан также на основании правила, согласно которому выбирается единица, обеспечивающая наибольшую точность оценок при заданной стоимости или наименьшую стоимость обследования при заданной точности.

Эффективность гнездового отбора можно повысить объединением в гнездах непохожих элементов. А затраты обычно уменьшаются, если в гнездах свести вместе территориально близкие элементы (приложение В).

Для осуществления систематического отбора все единицы совокупности нумеруются в некотором порядке числами от 1 до N. Для получения выборки объемом n единиц сначала извлекается, например, случайным образом какая-либо единица из первых k = N/n единиц совокупности. После этого в выборку включается каждая k-я единица, начиная с уже извлеченной. Извлечение первой единицы определяет всю выборку. Такая выборка называется систематической выборкой каждой k-й единицы. Отношение N/n называется интервалом или шагом отбора.

В отдельных случаях при наличии соответствующей информационной базы для повышения точности выборочных результатов единицы генеральной совокупности предварительно ранжируются по какому-либо существенному признаку. При таком подходе отбор единиц в выборочную совокупность начинается с единицы, находящейся в середине первого интервала.

В теории систематический отбор считается более эффективным, чем простая случайная выборка. Также его легче осуществлять при работе вручную, что потеряло актуальность с широким распространением персональных компьютеров. Следует отметить два недостатка этого способа отбора:

1. затруднено получение несмещенной оценки выборочной дисперсии;

2. существуют такие совокупности и значения n, при которых простая случайная выборка дает более точные оценки показателей (дисперсия оценки среднего систематической выборки для этих совокупностей может даже расти при увеличении объема выборки). Наличие периодического или циклического изменения в значениях признака, период которого равен интервалу отбора, - наихудшая из возможных ситуаций при систематическом отборе.

При систематическом отборе обычно применяются оценочные формулы простого случайного отбора, так как систематический отбор можно рассматривать как простой случайный, содержащий одну гнездовую единицу из совокупности k гнездовых единиц.

Дисперсия среднего значения при систематическом отборе определяется по формуле:

$$V(\bar{y}_{sy}) = \frac{N-1}{N} S^2 - \frac{k(n-1)}{N} S^2_{wsy} \qquad (5.4)$$

где N - объем генеральной совокупности;
S^2 - истинное значение дисперсии признака;
k - шаг отбора;
n - объем выборочной совокупности;
S^2_{wsy} - дисперсия единиц, принадлежащих одной и той же систематической выборке (*wsy* - от английского «within» - внутри и «systematic» - систематический):

$$S^2_{wsy} = \frac{1}{k(n-1)} \sum_{i=1}^{k} \sum_{j=1}^{n} (y_{ij} - \bar{y}_i)^2 \qquad (5.5)$$

где y_{ij} - значение признака j-го члена i-й систематической выборки, $j = 1, 2, ..., n$, $i = 1, 2, ..., k$;
$\bar{y}_i$ - среднее значение признака i-й выборки.

При организации статистических выборочных обследований широко применяется метод многоступенчатого отбора. Если исследуемая совокупность содержит некоторые группы и имеется информация о принадлежности элементов к той или иной группе, то в этом случае при выборочных обследованиях может быть удобным вначале осуществить случайную выборку из этих групп, а затем в целях экономии средств и времени не проводить обследование всех единиц отобранных групп, как при гнездовом отборе, а отобрать лишь часть элементов в каждой выбранной группе, т.е. осуществить двухступенчатый отбор. При многоступенчатом отборе извлечение единиц наблюдения осуществляется после нескольких последовательных случайных отборов групп.

В качестве примера многоступенчатого отбора рассмотрим его специальный случай - двухэтапный групповой отбор, в котором элементы второго этапа - единицы наблюдения, случайным образом отобранные из элементарных единиц выбранных групп (приложение Г).

В случае если на каждой ступени сохраняется одна и та же единица отбора, говорят о многофазном отборе. Многофазный отбор широко применяется в выборочных переписях населения, когда одна и та же совокупность обследуется на различных фазах отбора по разным, обычно расширяющимся от фазы к фазе, программам наблюдения.

В выборках квазислучайного типа предполагается наличие вероятностного отбора на том основании, что специалист, рассматривающий выборку, считает это допустимым (т.е. предполагается, обстоятельства таковы, что возможно рассматривать выборку как вероятностную).

Обоснованность этого решения всецело зависит от обстоятельств, поэтому делать обобщения сложно.

Примером использования квазислучайной выборки в статистической практике является «Выборочное обследование малых предприятий по изучению социальных процессов в малом предпринимательстве», проведенное в 1996 г. в некоторых регионах России. Единицы наблюдения (малые предприятия) отбирались экспертно с учетом представительства отраслей экономики из уже сформированной выборки обследования финансово-хозяйственной деятельности малых предприятий (форма № МП «Сведения об основных показателях финансово-хозяйственной деятельности малого предприятия»). При обобщении выборочных данных предполагалось, что выборочная совокупность сформирована методом простого случайного отбора.

В случае отбора невероятностным способом также имеются два типа выборок: сформированных на основе направленного отбора и сформированных на основе суждения эксперта. Отнесение выборки к тому или иному типу зависит от того, является ли процедура отбора объективной или нет.

Прямое использование суждения эксперта является наиболее общим методом намеренного включения единиц в выборку. Примером такого способа отбора является монографический метод, предполагающий получение информации только от одной единицы наблюдения, являющейся типичной, по мнению организатора обследования - эксперта.

По сравнению с вероятностными, выборки, построенные на основе суждения эксперта, наилучшим образом проявляют себя там, где:

1) выборка мала;

2) исследуемая совокупность весьма невелика и обозрима, или известна организатору наблюдения;

3) исследуемое свойство элементов общей совокупности существенно варьирует;

4) специалист, формирующий выборку, является большим и признанным мастером своего дела.

Выборки, сформированные на основе направленного отбора, извлекаются с помощью объективной процедуры, но без использования вероятностного механизма. Существует значительное число разнообразных способов направленного отбора.

Широко известен метод основного массива, при котором в выборку включаются наиболее крупные (существенные) единицы наблюдения, обеспечивающие основной вклад в показатель, например, суммарное значение признака, представляющего основной интерес обследования.

В заключение можно отметить, что выборки, сформированные на основе направленного отбора, несколько разочаровывают своими результатами по сравнению с вероятностными выборками, однако при работе с малыми выборками имеет смысл рассматривать виды отбора, приводящие к формированию таких выборок. Также в мировой статистической практике

существуют факты, свидетельствующие о том, что при очень небольших объемах выборки на основе суждения могут иметь преимущества перед, например, простыми случайными выборками.

«Квазислучайные» выборки нельзя сравнить с остальными тремя типами выборок каким-либо простым и удовлетворительным способом. Каждая квазислучайная выборка может оказаться уникальной. Можно сделать только единственное обобщение - такими выборками следует пользоваться крайне осторожно и обоснованно.

<u>Пример 5.1. Выборочное исследование общественного мнения (оформлено в виде статьи)</u>

УДК 336

Отношение студентов к высшему образованию в России.

***О.В. Николаева**, студент*
*Научный руководитель: **Е.В. Лаптева,** к.э.н., доцент*
Оренбургский филиал РЭУ им. Г.В. Плеханова, Россия

Аннотация. Образование - это один из главных каналов социальной мобильности, позволяющий добиться высокого социального статуса в обществе. Высшее образование перестало являться одной из самых важных ценностей в жизни для современных студентов. В статье проведен анализ отношения студентов к высшему образованию.

Ключевые слова: Россия, Оренбургская область, студенты, высшее образование, специальность, престиж образования.

Abstract. Education is one of the main channels of social mobility that allows you to achieve a high social status in society. Higher education has ceased to be one of the most important values in life for modern students. The article analyzes the attitude of students to higher education.

Keywords: Russia, Orenburg region, students, higher education, specialty, prestige of education.

Образование является одним из основных факторов, влияющих на жизнь каждого человека. А если человек имеет высокий уровень образования, то он может использовать свои знания, умения и навыки на деле, используя различные информационные ресурсы. Обладающий образованием человек общителен и имеет множество знакомых среди коллег и деловых партнеров. Уже в течение последних лет происходит постепенное выдвижение сферы высшего образования на первое место в общественной жизни, что находит

свое отражение в ее бурном развитии за прошедшие годы. В частности, это выразилось в том, что за три послевоенных десятилетия через систему обучения прошло больше учащихся, чем за всю предыдущую историю. К данному исследованию привело неадекватное отношение студентов вузов, которые не имеют возможности учиться на должном уровне. Сейчас, в наше время, ценность высшего образования в России как способа получения глубоких знаний и стать квалифицированным специалистом резко упала. Многие молодые люди стремятся к получению образования ради диплома.

На самом деле многие студенты просто-напросто не знают целей своего обучения в вузе. Они плохо готовятся к занятиям, прогуливают пары, каждый хочет иметь хорошие оценки, затрачивая при этом минимум усилий. У многих совершенно отсутствует желание учиться, изучать что-то новое и развиваться. Это приводит к снижению качества обучения, что является одним из главных критериев оценки качества подготовки специалистов любого образовательного учреждения.

Это связано с тем, что сейчас среди студентов наблюдается потеря интереса к обучению. По большей части, студенты посещают учебные заведения для того, чтобы провести время в компании друзей, а не для того, чтобы получить новые знания.

Большинство студентов обучаются на договорной основе и за них платят их родители, в некотором приходится себе отказывать, чтобы дать детям образование. Денег нет совсем, но дети это не ценят и воспринимают как должное. Не у всех родителей есть возможность дать детям образование, и они вынуждены работать, чтобы оплачивать свое обучение. Однако некоторые студенты не желают зависеть от родителей и самостоятельно зарабатывают деньги. Не все понимают, что без хорошего базиса знаний и опыта работы на рынке труда, никуда не устроиться, а тем более – быть востребованным и хорошим работником в своей области.

В современном обществе образованный человек всегда будет уважаемым, всегда сможет найти свое место в жизни и уверенно идти к намеченной цели. В современном мире образование является одним из основных каналов социальной мобильности, позволяющим достичь высокого социального статуса в обществе. Теперь высшее образование не является самой важной ценностью в жизни современного студента. Игнорирование этой проблемы может привести к серьезным последствиям в обществе.

Для исследования был проведен опрос студентов ОФ РЭУ им. Плеханова 1,2,3 курсов. Среди них приняли участие 155 студентов. Из них 71% (110 чел.) это студенты женского пола и 29% (45 чел.) студенты мужского пола. Возраст опрошенных варьируется от 18 до 23 лет.

Целью исследования является изучить отношение студентов к высшему образованию.

Главные задачи исследования это:

1.Выявить мотивы получения высшего образования студентами

2. Установить, что понимают студенты под высшим образованием как социальным ресурсом

3. Выявить место образования в иерархии ценностей студентов.

4.Определить влияние высшего образования на профессиональные планы студенчества.

5.Выявить факторы, влияющие на степень удовлетворенности получаемым образованием.

Анкетируемым был задан вопрос: какова Ваша общая оценка высшего образования в России на данный момент?

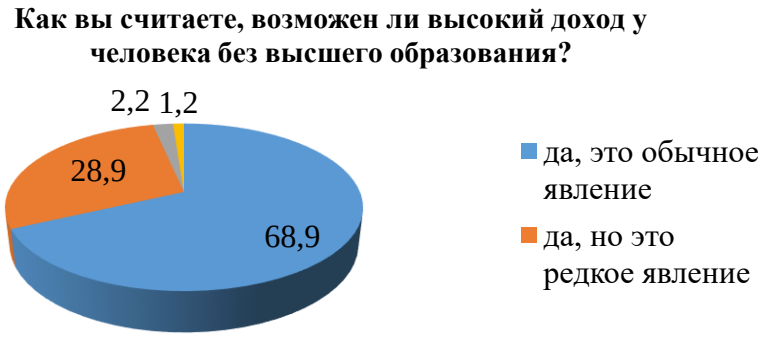
Больше половины опрошенных считают, что есть некоторые проблемы, но ожидаются изменения к лучшему. Однако 24,4% опрошенных считают, что есть некоторые проблемы и дальше будет только хуже. 11,1% опрошенных ответили, что это определенно упадок образования в России. И наименьший процент опрошенных отметил, что все хорошо. Из этого можно сделать вывод, о том, что люди видят какие-то недостатки в нашей системе образования и надеются, что в скором времени их исправят.

Следующий вопрос это: насколько, по вашему мнению, доступно высшее образование в России?

Более 80-ти % опрошенных отметили, что доступно, но требует больших материальных, эмоциональных, интеллектуальных затрат. А 8,9% ответили, что полностью доступно, не требует особых материальных, эмоциональных,

интеллектуальных затрат. Остальное меньшинство считает, что образование практически недоступно.

Третий вопрос: Как вы считаете, возможен ли доход у человека без высшего образования?



Большая часть опрашиваемых думает, что это вполне обычное явление. Почти 30% ответили, что это вполне возможно, но это редкое явление. Остальные считают, что это невозможно.

На следующий вопрос, который звучит так: насколько, по вашему мнению, высок престиж высшего образования в России?

Никто из студентов не считает, уровень высшего образования в России очень высокий. Но несмотря на это, больше 60-ти % опрошенных считают, что престиж выше среднего. Остальные затрудняются ответить.

Далее был представлен вопрос: насколько ответственно, по вашему мнению, нынешние студенты относятся к учебе?

Почти 85% опрошенных считают, что нынешние студенты относятся к учебе нормально. 13,3% - безответственно. Остальная малая часть считает, что студенты относятся к учебе ответственно.
Далее был задан вопрос, насколько сами опрошенные студенты относятся к учебе?

Наибольшее количество опрошенных, а точнее 61,4% дали ответ нормально. Несмотря на это 36,4% дали ответ очень ответственно. Таким образом, можно понять, что безответственных студентов среди 1-3 курсов в нашем университете почти нет.
Следующий вопрос прозвучал на тему финансовых затруднений у студентов.

Среди них 61,4% испытывают финансовые затруднения очень часто. 27,3% испытывают их всегда, и 11,4% испытывают финансовые затруднения иногда. Следовательно, исходя из данных ответов, студенты постоянно находятся в затруднительном финансовом положении.

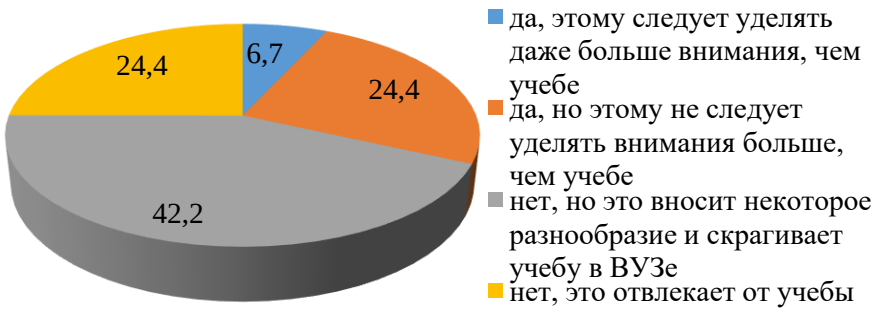
Далее был задан вопрос: способствует ли образование росту культурного уровня и кругозора студентов и выпускников ВУЗа?

Способствует ли высшее образование росту культурного уровня и кругозора студентов и выпускников ВУЗов?

Более 50% опрошенных считают, что многое зависит от самого человека, но при этом образование может оказывать существенное влияние. 26,7% думают, что образование не оказывает существенного влияние. Остальные два варианта собрали равное количество ответов.

Следующий заданный вопрос это: Считаете ли вы, что спорт и внеучебная деятельность является неотъемлемой частью высшего образования?

Считаете ли Вы, что спорт и внеучебная деятельность является неотъемлемой частью высшего образования?



42,2% опрошенных говорят о том, что спорт и внеучебная деятельность не являются частью высшего образования, но при этом они скрашивают будни и вносят какое-то разнообразие. Равное количество, а это 24,4% считают, что внеучебная деятельность и спорт это и есть часть высшего образования, но уделять большего внимания этому не стоит, а второй ответ, набравший также 24,4% показывает, что внеучебная деятельность лишь отвлекает и мешает учебе в университете.

Насколько высок, по Вашему мнению, уровень профессиональной подготовки в ВУЗах России?

Почти равное количество, а это 40% набрали два варианта: первый из них – все зависит от самого вуза, другой – удовлетворительно. 11,1% считают, что уровень подготовки удовлетворительный.

Как часто, по Вашему мнению, выпускники ВУЗов идут работать по специальности?

60% студентов дали ответ иногда. 24,4% ответили часто. И 13,3% выбрали вариант практически никогда.

Предпоследний вопрос: собираетесь ли Вы идти работать по специальности?

Половина опрошенных ответили да. Почти 30% затрудняются ответить на данный вопрос. И оставшиеся 20% дали ответ нет.

И последний вопрос, который прозвучал в опросе был Может ли, по Вашему мнению, молодой специалист получить высокооплачиваемую работу?

Почти 70% считают, что это редкое, но возможное явление для России. Оставшаяся часть ответила, что нет, но бывают единичные исключения.

Из полученных результатов опроса, можно сделать вывод о том, что по мнению опрошенных престиж образования в России для студентов находится на уровне выше среднего, что, несомненно, радует. Студенты видят некоторые проблемы в образовательной сфере, например, уровень подготовки в ВУЗах они оценивают как удовлетворительный. При этом образование в нашей стране, по их мнению, доступно, хоть и требует каких-либо затрат. Также студенты не видят смысла в организации для них внеучебной деятельности. Многие из опрошенных считают, что на спорт и различные мероприятия не стоит тратить много времени, а некоторые вообще отметили, что это мешает учебному процессу. Радует также, что среди опрошенных не наблюдается тенденция к тому, что образование в наше время совсем бесполезная трата времени.

Глава 6. СТАТИСТИЧЕСКИЕ МЕТОДЫ ИССЛЕДОВАНИЯ ВЗАИМОСВЯЗЕЙ

Изучение зависимости вариации признака от окружающих условий составляет содержание теории корреляции («корреляция» – соотношение, соответствие).

В действительности нетрудно заметить, что каждое явление находится в тесной связи и взаимодействует с другими явлениями.

При изучении конкретных зависимостей выделяют следующие виды признаков. Признаки, выступающие в роли факторов, обуславливающих изменение других признаков, называются факторными, а признаки, испытывающие воздействие – результативными.

Рассматривая зависимости между признаками, выделяют следующие категории связей:

1) функциональные связи (взаимооднозначные), где каждому значению фактора соответствует одно или несколько определенных значений результативного признака;

2) корреляционные связи, где между изменением факторного и результативного признака нет полного соответствия, влияние отдельных факторов проявляется лишь в среднем при массовом наблюдении факторов.

Сравнивая между собой функциональные и корреляционные связи отметим, что при наличии функциональной зависимости между признаками, зная величину факторного признака можно точно определить величину результативного признака, при наличии корреляционной: с изменением величины факторного признака меняется средняя величина результативного признака.

В обосновании связей решающая роль принадлежит экономической теории. Статистика же дает количественную оценку этой зависимости.

Условия применения корреляционно-регрессивного анализа:

- требование наблюдений (в массе единиц происходят взаимопогашения действия случайных факторов);
- требования однородности единиц (однотипные предприятия, по которым изучаются технико-экономические показатели).

Проведение корреляционно-регрессивного анализа имеет две основные цели:

1) измерение параметров уравнения, выражающего связь средних значений зависимой переменной со значениями независимой переменной;

2) измерение тесноты связи признаков между собой.

Существуют следующие методы выявления наличия корреляционной связи:

1) Сопоставление двух параллельных рядов – ряда значений факторного признака и соответствующих ему значений результативного признака.

Значения факторного признака располагают в возрастающем порядке и затем прослеживают направление изменения величины результативного признака.

Если с возрастанием величины факторного признака ведет к возрастанию результативного признака, то можно предположить прямую корреляционную связь;

Если с возрастанием факторного признака наблюдается убывание результативного признака, то можно предположить обратную связь между признаками.

2) Построение аналитической группировки, где все наблюдения разбиваются на группы по величине факторного признака, и по каждой группе вычисляется среднее значение результативного признака.

Предполагается, что все прочие причины, если они носят случайный характер, при определении средней по группам взаимопогашаются, т.е. дают в каждой группе один и тот же результат. Недостаток: неоднозначность результатов, которые зависят как от числа выделенных групп, так и от установления границ интервалов.

3) Графический метод

Для предварительного выявления наличия связи и раскрытия ее характера применяют графический метод. Используются данные о индивидуальных значения факторного признака и соответствующих ему значений результативного признака, можно построить точечный график, называемый полем корреляции. Точки корреляционного поля не лежат на одной линии, но вытянуты в определенном положении. Далее материал был сгруппирован, по каждому интервалу были определены значения средней. Соединив, получили эмпирическую линию связи, которая приближается к прямой, мы можем предположить наличие прямой корреляционной связи;

4) корреляционная таблица.

Задачи корреляционно-регрессионного анализа:

1. Выделение важных факторов, влияющих на результативный признак, на базе мер тесноты связи факторов с результативным признаком.

2. Оценка уравнения регрессии, где в качестве результативного признака выступает признак, являющийся следствием других признаков – причин.

3. Прогнозирование возможных значений результативного признака при задаваемых значениях факторных признаков. В уравнение подставляется планируемые значения факторных признаков и вычисляются ожидаемые значения результативного признака.

При рассмотрении количественных переменных основное внимание уделяется линейным связям.

При функциональной связи изменение результативного признака y всецело обусловлено действием факторного признака x: $y = f(x)$.

При корреляционной связи изменение результативного признака y обусловлено влиянием факторного признака x не всецело, а лишь частично, так как возможно влияние прочих факторов e: $y = \psi(x) + e$.

Характерной особенностью функциональной связи является то, что она проявляется с одинаковой силой у каждой единицы изучаемой совокупности. Иное дело при корреляционных связях. Здесь при одном и том же значении учтенного факторного признака возможны различные значения результативного признака. Это обусловлено наличием других факторов, которые могут быть различными по составу, направлению и силе действия на отдельные индивидуальные единицы статистической совокупности. Поэтому для изучаемой статистической совокупности в целом здесь устанавливается такое соотношение, в котором определенному изменению факторного признака соответствует среднее изменение признака результативного.

Следовательно, характерной особенностью корреляционных связей является то, что они проявляются не в единичных случаях, а в массе. Поэтому изучаются корреляционные связи по так называемым эмпирическим данным, полученным в статистическом наблюдении. В таких данных отображается совокупное действие всех причин и условий на изучаемый показатель.

Наиболее разработанной в теории статистики является методология так называемой парной корреляции, рассматривающая влияние вариации факторного признака x на результативный y.

Решение математических уравнений связи предполагает вычисление по исходным данным их параметров. Это осуществляется способом выравнивания эмпирических данных методом наименьших квадратов. В основу этого метода положено требование минимальности сумм квадратов отклонений эмпирических данных y_i от выровненных y_{xi}:

$$\sum (y_i - y_{x_i})^2 = \min . \qquad (6.1)$$

По проверенным на типичность параметрам уравнения регрессии производится построение математической модели связи. При этом параметры примененной в анализе математической функции получают соответствующие количественные значения.

Для статистической оценки тесноты связи между признаками x и y применяются следующие показатели вариации:

1) общая дисперсия результативного признака, отображающая совокупное влияние всех факторов:

$$\sigma_y^2 = \frac{\sum (y_i - \bar{y})^2}{n}; \qquad (6.2)$$

2) факторная дисперсия результативного признака отображающая вариацию y только от воздействия изучаемого фактора x:

$$\sigma_{y_x}^2 = \frac{\sum (y_{x_i} - \bar{y})^2}{n}; \qquad (6.3)$$

3) остаточная дисперсия отображающая вариацию результативного признака у от всех прочих, кроме х, факторов:

$$\sigma_{\varepsilon}^2 = \frac{\sum (y_i - y_{x_i})^2}{n}; \quad (6.4)$$

Соотношение между факторной и общей дисперсиями характеризует меру тесноты связи между признаками х и у:

$$\frac{\sigma_{y_x}^2}{\sigma_y^2} = R^2. \quad (6.5)$$

Показатель R^2 называется индексом детерминации. Он выражает долю факторной дисперсии, т. е. характеризует, какая часть общей вариации результативного признака у объясняется изучаемым фактором х.

Коэффициент детерминации используется для количественного определения характеристики, связывающей две переменные. Дает пропорцию общего изменения одной переменной (Y), которую можно объяснить изменением второй переменной (X). Иногда R^2 =1, но при этом связь отсутствует, поскольку X и Y связаны с третьей переменной (X_2).Индекс корреляции R определяется по формуле:

$$R = \sqrt{\frac{\sigma_{y_x}^2}{\sigma_y^2}} = \sqrt{\frac{\sigma_y^2 - \sigma_{\varepsilon}^2}{\sigma_y^2}} = \sqrt{1 - \frac{\sigma_{\varepsilon}^2}{\sigma_y^2}}. \quad (6.6)$$

При прямолинейной форме связи показатель тесноты связи определяется по формуле линейного коэффициента корреляции r:

$$r = \frac{\sum xy - \frac{\Sigma x \Sigma y}{n}}{\sqrt{\left[\sum x^2 - \frac{(\Sigma x)^2}{n}\right]\left[\sum y^2 - \frac{(\Sigma y)^2}{n}\right]}} = \frac{\sum (x_i - \bar{x}) \cdot (y_i - \bar{y})}{n \cdot \sigma_{x} \cdot \sigma_y} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x \cdot \sigma_y} \quad (6.7)$$

По абсолютной величине линейный коэффициент корреляции r равен индексу корреляции R только при прямолинейной связи.

Для получения выводов о практической значимости синтезированных в анализе моделей показаниям тесноты связи дается качественная оценка. Это осуществляется на основе шкалы Чеддока (таблица 6.1).

Таблица 6.1- Шкала Чеддока

Показания тесноты связи	0,1–0,3	0,3–0,5	0,5–0,7	0,7–0,9	0,9–0,99
Характеристика силы связи	слабая	умеренная	заметная	высокая	весьма высокая

При значениях показателей тесноты связи, превышающих 0,7 зависимость результативного признака у от факторного х является высокой, а при значениях более 0,9 – весьма высокой. Это означает, что более половины общей вариации результативного признака у объясняется влиянием изучаемого фактора х. Последнее позволяет считать оправданным применение

метода функционального анализа для изучения корреляционной связи, а синтезированные при этом математические модели признаются пригодными для их практического использования.

При показаниях тесноты связи ниже 0,7 величина индекса детерминации R всегда будет меньше 50%. Это означает, что на долю вариации факторного признака х приходится меньшая часть по сравнению с прочими признаками, влияющими на изменение общей дисперсии результативного признака. Синтезированные при таких условиях математические модели связи практического значения не имеют.

Значимые величины коэффициента корреляции (R) зависят от объёма выборки (N) и заданной вероятности получения результата. Для оценки значимости R можно использовать следующую шкалу (для вероятности 95%) (таблица 6.2).

Таблица 6.2 – Шкала оценки значимости R (для вероятности 95%)

Объём выборки, ед.	3	4	5	10	15	20	50	100
Значение R	0,997	0,95	0,878	0,632	0,51	0,44	0,35	0,19

Если R=0,3–0,5 - его трудно истолковать, и требуется проведение дополнительных исследований.

Показатели вариации результативного признака используются и при выборе наиболее соответствующего эмпирическим данным уравнения регрессии. В изучении корреляционной связи это наиболее важный и ответственный этап анализа. Именно от адекватности примененного уравнения регрессии зависит правильность выводов корреляционно-регрессионного анализа.

Прямолинейная форма зависимости между признаками х и у выражается уравнением

$$y = a_o + a_1 x.$$

Для определения параметров уравнения на основе требований метода наименьших квадратов составляется система нормальных уравнений:

$$\begin{cases} na_0 + a_1 \sum x = \sum y; \\ a_0 \sum x + a_1 \sum x^2 = \sum xy \end{cases}$$

Решение системы: $a_1 = \dfrac{\sum[(x-\bar{x})\times(y-\bar{y})]}{\sum(x-\bar{x})}$, $a_0 = \bar{y} - a_1 \bar{x}$

При использовании уравнения параболы $y = a_0 + a_1 x + a_2 x^2$ параметры уравнения вычисляются путем решения системы трех нормальных уравнений:

$$\begin{cases} na_0 + a_1 \Sigma x + a_2 \Sigma x^2 = \Sigma y \\ a_0 \Sigma x + a_1 \Sigma x^2 + a_2 \Sigma x^3 = \Sigma yx \\ a_0 \Sigma x^2 + a_1 \Sigma x^3 + a_2 \Sigma x^4 = \Sigma yx^2 \end{cases}$$

При использовании уравнения показательной (экспоненциальной) функции $y = a_0 \times a_1^t$ параметры уравнения вычисляются путем решения системы нормальных уравнений:

$$\begin{cases} n \times \lg a_0 + \lg a_1 \Sigma x = \Sigma \lg y \\ \lg a_0 \Sigma x + \lg a_1 \Sigma x^2 = \Sigma x \lg y \end{cases}$$

При статическом анализе криволинейной связи часто применяется полулогарифмическая функция: $y_x = a_o + a_1 \lg x$.

Параметры уравнения определяется из системы уравнений:

$$\begin{cases} n a_0 + a_1 \sum \lg x = \sum y; \\ a_0 \sum \lg x + a_1 \sum (\lg x)^2 = \sum y \cdot \lg x \end{cases}$$

На практике часто приходится исследовать зависимость результативного признака от нескольких факторных признаков. В этом случае статистическая модель может быть представлена уравнением регрессии с несколькими переменными величинами. Такая регрессия называется множественной.

F-критерий - это оценивание качества уравнения регрессии, которое состоит в проверке гипотезы H_0 о статистической незначимости уравнения регрессии и показателя тесноты связи. Для этого производится сравнение фактического $F_{факт}$ и $F_{табл}$ значений F критерия Фишера-Снедекора. $F_{факт}$ определяется из соотношения значений факторной и остаточной дисперсий, рассчитанных на одну степень свободы.

$F_{табл}$ - это максимальная величина отношения дисперсий, которая может иметь место при случайном их расхождении для данного уровня вероятности.

$$F_{факт} = \frac{\sum (y_x - \bar{y})^2 / m}{\sum (y - y_x)^2 / (n - m - 1)} = \frac{r_{xy}^2}{1 - r_{xy}^2}(n-2), \quad m = 1. \qquad (6.8)$$

$F_{табл}$ - это максимально возможное значение критерия под влиянием случайных факторов при данных степенях свободы и уровне значимости α. Уровень значимости α - это вероятность отвергнуть правильную гипотезу при условии, что она верна. Обычно $\alpha = 0{,}05 \ (0{,}01)$.

Если $F_{табл} < F_{факт}$, то H_0 - гипотеза о случайной природе оцениваемых характеристик отклоняется и признается их статистическая значимость и надежность.

Если $F_{табл} > F_{факт}$, то H_0 - гипотеза не отклоняется и признается статистическая незначимость, ненадежность уравнения регрессии.

Например, линейная регрессия с т независимыми переменными имеет вид:

$$\tilde{y}_i = a_0 \times x_0 + a_1 \times x_1 + a_2 \times x_2 + \ldots + a_m \times x_m.$$

При анализе экономических явлений множественная регрессия и корреляция применяются одновременно. С помощью регрессии определяется форма связи и оцениваются параметры регрессионной модели. Посредством корреляционного анализа определяется сила связи между факторами.

Для изучения основных задач и особенностей корреляционного анализа удобно рассматривать генеральную совокупность 3-х признаков.

Рассмотрим случай трех признаков $X=(x_1, x_2, x_3)$, (p=3). Будем предполагать, что поведение многомерного вектора X описывается нормальным законом распределения, т.е. плотность совместного распределения одномерных случайных величин x_1, x_2, x_3 задается в виде:

$$p(x_1, x_2, x_3) = \frac{1}{\sqrt{(2\pi)^2}} \cdot \frac{1}{\sqrt{\sigma_{x_1}^2 \sigma_{x_2}^2 \sigma_{x_3}^2 \det(R)}} e^{-0,5 z^T R^{-1} z}$$

где R — симметрическая положительно определенная матрица парных коэффициентов корреляции, а det(R) — определитель этой матрицы (обобщенная дисперсия случайной величины X), т.е.

$$R = \begin{bmatrix} 1 & r_{x_1x_2} & r_{x_1x_3} \\ r_{x_2x_1} & 1 & r_{x_2x_3} \\ r_{x_3x_1} & r_{x_3x_2} & 1 \end{bmatrix} \text{ и}$$

$$|R| = \det R = 1 + 2r_{12}r_{13}r_{23} - r_{13}^2 - r_{23}^2 - r_{12}^2 > 0$$

R^{-1} — матрица, обратная к R.

z - обозначен вектор значений нормированных случайных величин x_1, x_2, x_3.

$$z = \begin{bmatrix} \dfrac{x_1 - \mu_{x_1}}{\sigma_{x_1}} \\ \dfrac{x_2 - \mu_{x_2}}{\sigma_{x_2}} \\ \dfrac{x_3 - \mu_{x_3}}{\sigma_{x_3}} \end{bmatrix}$$

Таким образом, имеется трехмерная нормально распределенная случайная величина, которая определяется девятью параметрами: μ_{x1}, μ_{x2}, μ_{x3}, $\sigma_{x_1}^2$, $\sigma_{x_2}^2$, $\sigma_{x_3}^2$, $r_{x_1x_2}$, $r_{x_1x_3}$, $r_{x_2x_3}$.

Распределения одномерных x_1, x_2, x_3, двумерных (x_1, x_2), (x_1, x_3), (x_2, x_3), условные распределения при фиксировании одной из переменных $(x_1, x_2)|x_3$, $(x_1, x_3)|x_2$, $(x_2, x_3)|x_1$ и двух $x_1|x_2x_3$, $x_2|x_1x_3$, $x_3|x_1x_2$ являются нормальными.

Базовым инструментом измерения линейной связи между признаками является парный коэффициент корреляции.

Парный коэффициент корреляции характеризует тесноту линейной зависимости между двумя переменными на фоне действия всех остальных показателей, входящих в модель.

$$r_{jl} = \frac{1/n \sum_{i=1}^{n} (x_{ij} - \bar{x}_j)(x_{il} - \bar{x}_l)}{\sigma_j \cdot \sigma_l}, \quad -1 \le r_{jl} \le 1. \qquad (6.9)$$

Если r_{jl} близок к ±1, то связь между переменными сильная, если r_{jl}=0 – линейная связь отсутствует.

$r_{jl} > 0$ – связь между переменными прямая;

$r_{jl} < 0$ – связь обратная.

Для многомерной корреляционной модели важную роль играют частные и множественные коэффициенты корреляции, детерминации (квадраты соответствующих коэффициентов корреляции).

Частный коэффициент корреляции между x_1 и x_2 *k-2-го порядка* при фиксированном воздействии переменной x_3 может быть определен по следующей формуле:

$$r_{x_1x_2}(x_3) = \frac{-R_{12}}{\sqrt{R_{11}R_{22}}} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{(1-r_{13}^2)(1-r_{23}^2)}} \qquad (6.10)$$

R_{ij} — алгебраическое дополнение матрицы R к элементу r_{ij},

Остальные частные коэффициенты определяются аналогично, путем замены соответствующих индексов.

$$r_{yx_j}(x_1,...,x_{j-1},x_{j+1},...,x_k) = r_{yx_j/1,...,\,j-1,j+1,...,k} = \sqrt{1-\frac{1-R^2_{yx_1...x_k}}{R^2_{yx_1...x_{j-1},x_{j+1}...x_k}}}\,, \qquad (6.11)$$

$R^2_{yx_1...x_k}$ - множественный коэффициент детерминации всего комплекса *k* факторов с результатом;

$R^2_{yx_1...x_{j-1},x_{j+1,...}x_k}$ - тот же показатель детерминации, но без введения в модель фактора x_j.

Порядок частного коэффициента корреляции определяется количеством факторов, влияние которых исключается. В практических исследованиях предпочтение отдают показателям частной корреляции самого высокого порядка.

Частный коэффициент корреляции показывает тесноту линейной связи между двумя переменными независимо от влияния остальных случайных величин. Обладает всеми свойствами парного коэффициента корреляции.

Если частный коэффициент корреляции меньше парного, т.е. $r_{x_1x_2}(x_3) < r_{x_1x_2}$, то взаимодействие между x_1 и x_2 обусловлено частично (или полностью, если $r_{x_1x_2}(x_3) = 0$) воздействием фиксируемых прочих переменных, т.е. - x_3. Если частный коэффициент корреляции $r_{x_1x_2}(x_3) > r_{x_1x_2}$, то фиксируемые прочие переменные ослабляют линейную связь.

Корреляционная матрица не в полной мере отражает связи признака с множеством переменных.

Множественный коэффициент корреляции является мерой связи между одной переменной и другими (остальными), входящими в модель, и может быть рассчитан по формуле:

$$r_{x_3} = r_{x_3}(x_1 x_2) = \sqrt{1 - \frac{\det R}{R_{33}}} = \sqrt{\frac{r_{31}^2 + r_{32}^2 - 2r_{12}r_{31}r_{32}}{1 - r_{12}^2}}\ ; \qquad (6.12)$$

$$r_y = r_y(x_1, x_2, ..., x_k) = \sqrt{1 - \frac{\det R}{R_{yy}}}\ . \qquad (6.13)$$

Если $r_{X3} = 1$, то связь между x_3 и двумерной переменной (x_1, x_2) является функциональной, линейной. Если $r_{x_3} = 0$, то; линейной связи нет.

Из формулы r_{x_3} следует, что $r_3 > |r_{31}|$, $r_3 > |r_{32}|$, $r_3 > |r_{31|2}|$, $r_3 > |r_{32|1}|$.

Отсюда можно заметить, что коэффициент множественной корреляции может только увеличиваться, если в модель включать дополнительные признаки – случайные величины, и не увеличиться, если из имеющихся признаков производить исключение.

Если $r_3 = 0$, то $r_{31} = r_{32} = r_{31|2} = r_{32|1} = 0$.

Наибольшему множественному коэффициенту детерминации соответствуют большие частные коэффициенты детерминации (например, r_1^2 соответствуют $r_{12|3}^2$ и $r_{13|2}^2$)

Множественный коэффициент детерминации (квадрат соответствующего множественного коэффициента корреляции) показывает долю дисперсии, например, случайной величины x_3, обусловленную изменением случайных величин x_1 и x_2.

Уравнение нелинейной регрессии, так же как и в линейной зависимости, дополняется показателем корреляции, а именно индексом корреляции (R):

$$R = \sqrt{1 - \frac{\sum (y - \tilde{y}_x)^2}{\sum (y - \bar{y})^2}} \qquad (6.14)$$

Величина данного показателя находится в границах: $0 \le R \le 1$, чем ближе к единице, тем теснее связь рассматриваемых признаков, тем более надежно найденное уравнение регрессии.

Парабола второй степени, как и полином более высокого порядка, при линеаризации принимает вид уравнения множественной регрессии. Если же нелинейное относительно объясняемой переменной уравнение регрессии при линеаризации принимает форму линейного уравнения парной регрессии, то для оценки тесноты связи может быть использован линейный коэффициент корреляции, величина которого в этом случае совпадет с индексом корреляции $R_{yx} = r_{yz}$, где z - преобразованная величина признака-фактора, например $z = \frac{1}{x}$ или $z = \ln x$.

Иначе обстоит дело, когда преобразования уравнения в линейную форму связаны с зависимой переменной. В этом случае линейный коэффициент корреляции по преобразованным значениям признаков дает лишь приближенную оценку тесноты связи и численно не совпадает с индексом корреляции. Так, для степенной функции $\tilde{y}_x = a \cdot x^b \cdot \varepsilon$ после перехода к логарифмически линейному уравнению $\ln y = \ln a + b \ln x + \ln \varepsilon$ может быть найден линейный коэффициент корреляции не для фактических значений

переменных *x* и *y,* а для их логарифмов, т. е. $r_{\ln y \ln x}$. Соответственно квадрат его значения будет характеризовать отношение факторной суммы квадратов отклонений к общей, но не для *y,* а для его логарифмов:

$$r^2_{\ln y \ln x} = \frac{\sum(\ln \tilde{y} - \overline{\ln y})^2}{\sum\left(\ln y - \overline{\ln y}\right)^2} = 1 - \frac{\sum(\ln y - \ln \tilde{y})^2}{\sum\left(\ln y - \overline{\ln y}\right)^2} \qquad (6.15)$$

Поскольку в расчете индекса корреляции используется соотношение факторной и общей суммы квадратов отклонений, то R^2 имеет тот же смысл, что и коэффициент детерминации. В специальных исследованиях величину R^2 для нелинейных связей называют индексом детерминации.

Оценка существенности индекса корреляции проводится, так же как и оценка надежности коэффициента корреляции.

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

$$F = \frac{R^2}{1 - R^2} \cdot \frac{n - m}{m - 1} \qquad (6.16)$$

где R^2 - индекс детерминации;

n - число наблюдений;

m — число параметров.

Величина (*m*-1*)* характеризует число степеней свободы для факторной суммы квадратов, а *(n - m*) — число степеней свободы для остаточной суммы квадратов.

Оценку качества построенной модели можно определить через коэффициент (индекс) детерминации, а также с помощью средней ошибки аппроксимации.

Средняя ошибка аппроксимации - среднее отклонение расчетных значений от фактических в процентах:

$$\overline{A} = \frac{1}{n}\sum\left|\frac{y - y_x}{y}\right| \cdot 100\%. \qquad (6.17)$$

Предел значений $A \le 0.08 - 0.1$ $(8 - 10\%)$ считаем допустимым при построении модели.

Средний коэффициент эластичности $\overline{Э}$ показывает, на сколько % в среднем по совокупности изменится результат y от своей средней величины при изменении фактора x на 1% от своего среднего значения

$$\overline{Э} = f'(x)\frac{\bar{x}}{\bar{y}} \quad \Rightarrow \quad \overline{Э} \cdot \bar{y} = f'(x) \cdot \bar{x} \qquad (6.18)$$

f'- характеризует соотношение прироста результата и фактора для соответствующей формы связи.

Т.к., коэффициент Э не всегда const, то используем среднее значение - $\overline{Э}$.

В таблице 6.3 представлены формулы эластичности для наиболее употребительных функций.

Таблица 6.3 - Коэффициенты эластичности для ряда математических функций

Вид функции, $\tilde{y}_x$	Первая производная, $f'(x)$	Коэффициент эластичности, $Э = f'(x)\frac{x}{y}$
Линейная $y = a + bx$	b	$Э = \frac{bx}{a+bx}$
Параболическая $y = a + bx + cx^2$	$b + 2cx$	$Э = \frac{(b+2cx)x}{a+bx+cx^2}$
Гипербола $y = a + \frac{b}{x}$	$-\frac{b}{x^2}$	$Э = \frac{-b}{a+bx}$
Показательная $y = ab^x$	$\ln b \cdot a \cdot b^x$	$x \ln b$
Степенная $y = ax^b$	$a \cdot b \cdot x^{b-1}$	b
Логарифмическая $y = a + b \ln x$	$\frac{b}{x}$	$\frac{b}{a+b\ln x}$
Обратная $y = \frac{1}{a+bx}$	$\frac{-b}{(a+bx)^2}$	$\frac{-bx}{a+bx}$

Иногда коэффициент Э экономического смысла не имеет. Это происходит тогда, когда для рассматриваемых признаков бессмысленно определение изменения значений в процентах. Например, изменение роста заработной платы с ростом стажа работы на 1%.

Если линеаризация не затрагивает зависимую переменную, например $z = \ln x,\ z = \frac{1}{x}$, то требование МНК: $\sum (y - y_x)^2 \to \min$ выполнимо, то $r_{xy} = \rho_{xy}$ (коэффициент корреляции совпадает с индексом корреляции), в этом легко убедиться.

<u>Пример 6.1. Корреляционно - регрессионный анализ инвестиций в строительство в регионе</u>

На уровень инвестиций в строительство влияет большое количество факторов. Попробуем изучить взаимосвязь величины уровень инвестиций в строительство и других экономических явлений, происходящих в Оренбургской области. В корреляционно-регрессионном анализе можно устранить воздействие какого-либо фактора, если зафиксировать воздействие этого фактора на результат и другие, включенные в модель факторы. Данный прием широко применяется в анализе временных рядов, когда тенденция

фиксируется через включение фактора времени в модель в качестве независимой переменной.

Для проведения корреляционно-регрессионного анализа используем следующие факторные признаки:

X1 – количество строительных организаций;

X2 – среднесписочная численность работников строительных организаций, тыс.руб.;

X3 – среднемесячная номинальная начисленная заработная плата работников занятых на строительстве зданий и сооружений;

X4 – средний уровень использования производственных мощностей строительных организаций, %;

X5 – число убыточных строительных организаций, % от общего числа организаций;

X6 – материальные затраты на производство строительных работ, в % к итогу всех затрат;

X7 – задолженность заказчиков за выполненные объемы работ, в % от общей задолженности;

X8 – сальдированный финансовый результат.

Параметры модели с включением фактора времени оцениваются с помощью обычного метода наименьших квадратов (МНК).

С помощью ПК получаем матрицу парных коэффициентов, на основании которых необходимо сделать вывод о факторах, которые могут быть включены в модель множественной регрессии (таблица 6.4). Корреляционная матрица получена с помощью табличного редактора Excel XP в пакете анализа.

Таблица 6.4 – Корреляционная матрица влияния факторов на уровень инвестиций в строительство Оренбургской области

	y	*X1*	*X2*	*X3*	*X4*	*X5*	*X6*	*X7*	*X8*
У	1								
X1	**0,784196**	1							
X2	**0,740948**	0,472513	1						
X3	0,170823	0,20517	0,0347	1					
X4	**-0,57755**	-0,16364	-0,42623	-0,21985	1				
X5	-0,15388	-0,27971	-0,35876	0,479488	-0,20222	1			
X6	-0,11082	-0,74114	-0,51858	0,163678	-0,97561	0,446238	1		
X7	-0,3984	0,115731	-0,12415	-0,15768	0,78703	-0,42972	-0,9336	1	
X8	**-0,79397**	-0,7641	-0,84743	0,038866	0,410725	0,532786	0,816379	-0,01735	1

Из корреляционной матрицы видна достаточно сильная взаимосвязь между результативным (У) и факторными признаками (X1, X2, X4, X8). Связь очень сильная.

ВЫВОД ИТОГОВ

Регрессионная статистика	
Множественный R	0,910875
R-квадрат	0,829692
Нормированный R-квадрат	0,772923
Стандартная ошибка	1451,887
Наблюдения	17

Дисперсионный анализ

	df	*SS*	*MS*	*F*
Регрессия	4	1,23E+08	30808467	14,61519
Остаток	12	25295703	2107975	
Итого	16	1,49E+08		

	Коэффициенты	*Стандартная ошибка*	*t-статистика*	*P-Значение*
Y-пересечение	-17383,7	12756,71	-2,36271	0,00198001
X1	59,13016	16,01317	3,692595	0,00003077
X2	272,278	189,9122	3,433704	0,00177196
X4	-5,55431	1,960663	-2,83287	0,00015095
X8	171,312	241,8665	2,70829	0,00492294

Нижние 95%	*Верхние 95%*	*Нижние 95,0%*	*Верхние 95,0%*
-45178,2	10410,8	-45178,2	10410,8
24,24046	94,01987	24,24046	94,01987
-141,505	686,0612	-141,505	686,0612
-9,82623	-1,28239	-9,82623	-1,28239
-698,294	355,6697	-698,294	355,6697

Проведем регрессионный анализ. По результатам регрессионного анализа получено следующее уравнение регрессии:

$$y = -17383,73 + 59,13 \cdot x_1 + 272,3 \cdot x_2 + 5,55 \cdot x_4 + 171,3 \cdot x_8 \qquad (6.19)$$

(-2,36) (3,69) (3,43) (-2,83) (0,70)

В скобках указаны значения t-критерия Стьюдента.

В результате построения уравнения регрессии получили следующие результаты (таблица 6.5).

Таблица 6.5 – Результаты построения регрессии

Показатели	Значения
Коэффициент корреляции R	0,910
Коэффициент детерминации R^2	0,829
Скорректированный коэффициент детерминации R^2	0,773
Фактическое значении F-критерия Фишера	14,61
Табличное значении F-критерия Фишера	2,79
Стандартная ошибка	5,11

Множественный коэффициент регрессии равен 0,910. Это свидетельствует о высокой связи между признаками. Коэффициент детерминации – равен 0,829, следовательно, 82,9% вариации уровня инвестиций в строительство Оренбургской области обусловлено факторами, включенными в модель (6.19).

Анализ полученного уравнения позволяет сделать выводы о том, что с ростом количества строительных организаций – уровень инвестиций в строительство возрастает на 59,13 тыс.руб., с ростом среднесписочной численности работников строительных организаций Оренбургской области на 1 тыс.руб. - уровень инвестиций в строительство увеличивается на 272,3 тыс.руб., с ростом среднего уровня использования производственных мощностей строительных организаций на 1% - уровень инвестиций в строительство увеличивается на 5,55 тыс.тыс.руб., с ростом сальдированного финансового результата на 1 млн.руб. - уровень инвестиций в строительство увеличивается на 171,3 тыс.руб.

Проверка адекватности модели, построенной на основе уравнений регрессии, начинается с проверки значимости каждого коэффициента регрессии. Значимость коэффициента регрессии осуществляется с помощью t-критерия Стьюдента:

$$t_{расч} = \frac{|a_i|}{\sigma_{a_i}} \qquad (6.20)$$

Параметры уравнения все значимы, кроме параметра при факторе времени, так как их расчетные значения меньше табличных ($t_{табличное} = 2{,}02$, уровень значимости = 0,05, $t_{расч} \rangle t_{таблич}$)

Проверка адекватности всей модели осуществляется с помощью расчета F-критерия. Если $F_р > F_т$ при α=0,05, то модель в целом адекватна изучаемому явлению.

$$F_{расч} = 14{,}61 \quad F_{табл} = 2{,}79 \quad уровень\ значимости = 0{,}05 \quad F_{расч} \rangle\ F_{табл}$$

Следовательно, построенная модель на основе её проверки по F-критерию Фишера в целом адекватна, и все коэффициенты регрессии значимы. Такая модель может быть использована для принятия решений и осуществления прогнозов.

<u>Пример 6.2. Практика построения уравнения множественной регрессии в Excel</u>

На основании имеющихся данных проведем корреляционно-регрессионный анализ.

У - численность умерших на 1000 человек населения;

Х1 - уровень заболеваемости населения;

Х2 - число зарегистрированных преступлений по Оренбургской области на 100 000 человек;

X3 - уровень безработицы, %;

X4 - численность больных состоящих на учете в лечебно-профилактических учреждениях с диагнозом алкоголизм и алкогольный психоз на 100 000 человек;

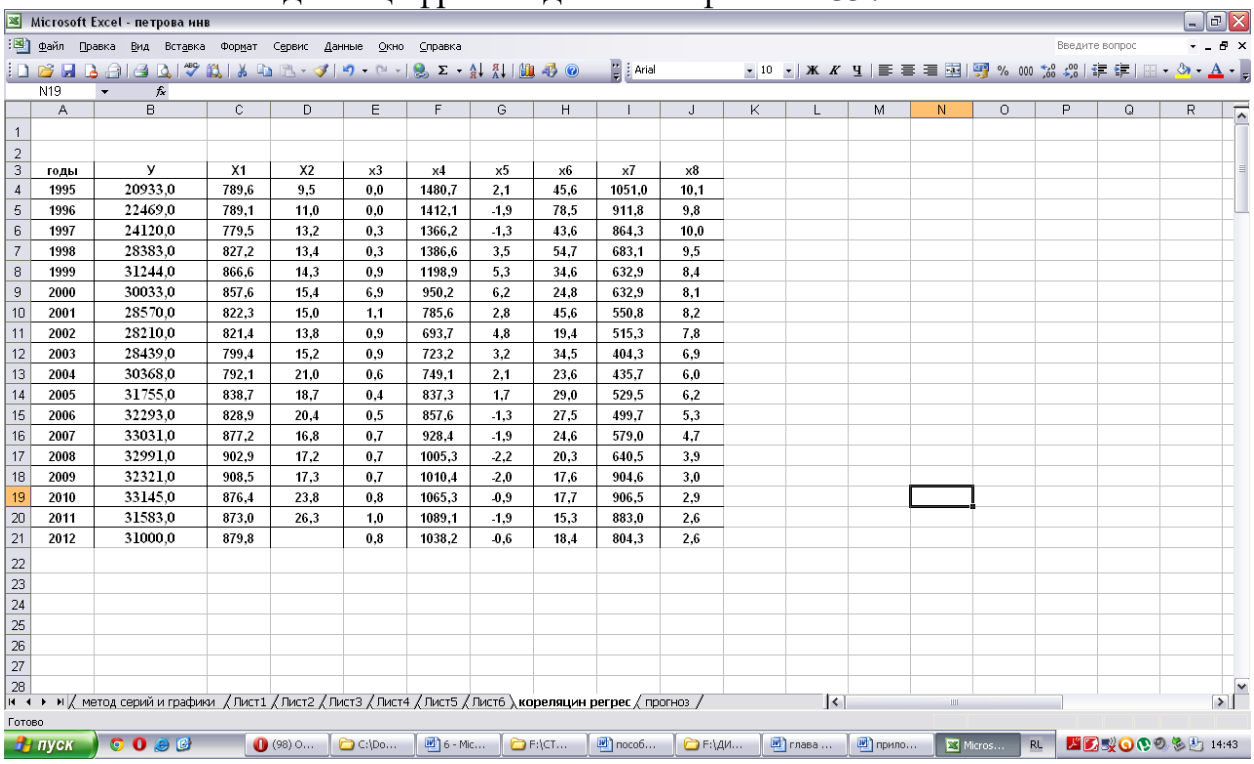
X5 - прирост миграции на 1000 человек;

X6 - численность населения с денежными доходами ниже величины прожиточного минимума, в % от общей численности населения;

X7- выбросы в окружающую среду;

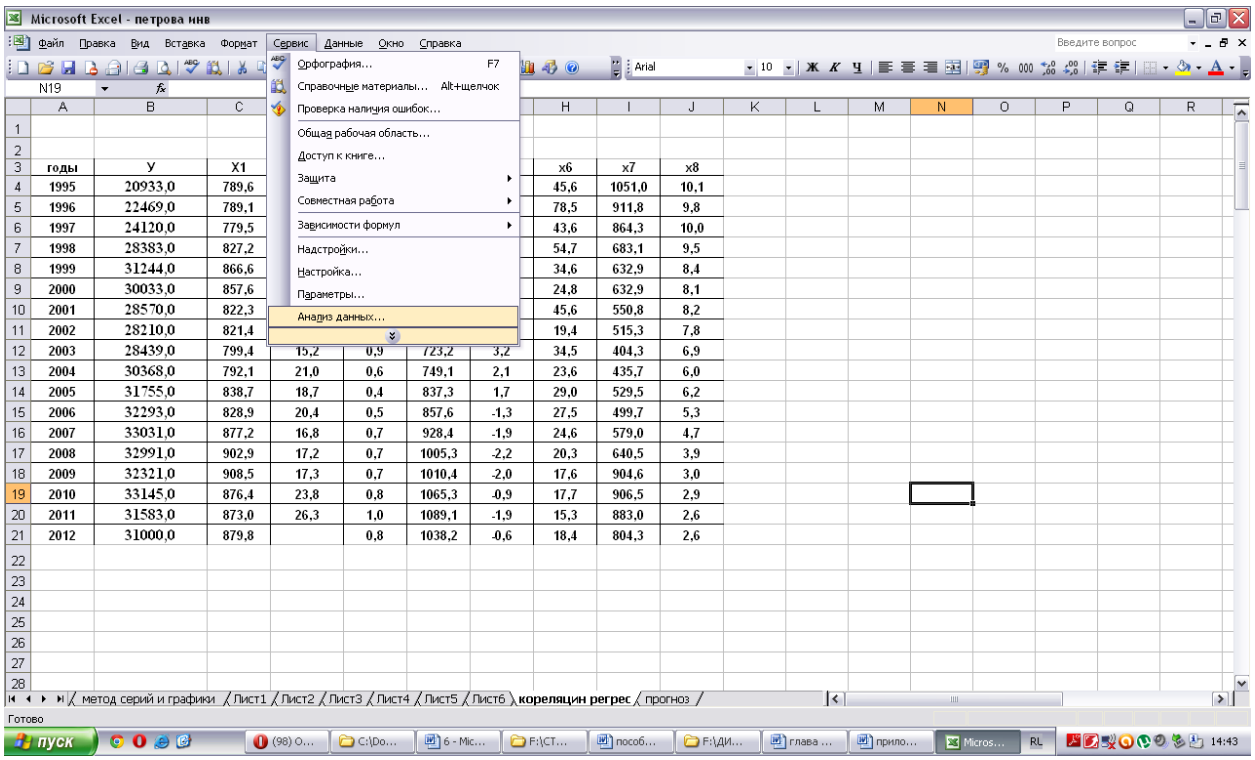
X8 - Численность пострадавших при несчастных случаях на производстве с утратой трудоспособности на один рабочий день и более и со смертельным исходом на 1000 работающих

1. Забиваем исходные цифровые данные в файл Excel.

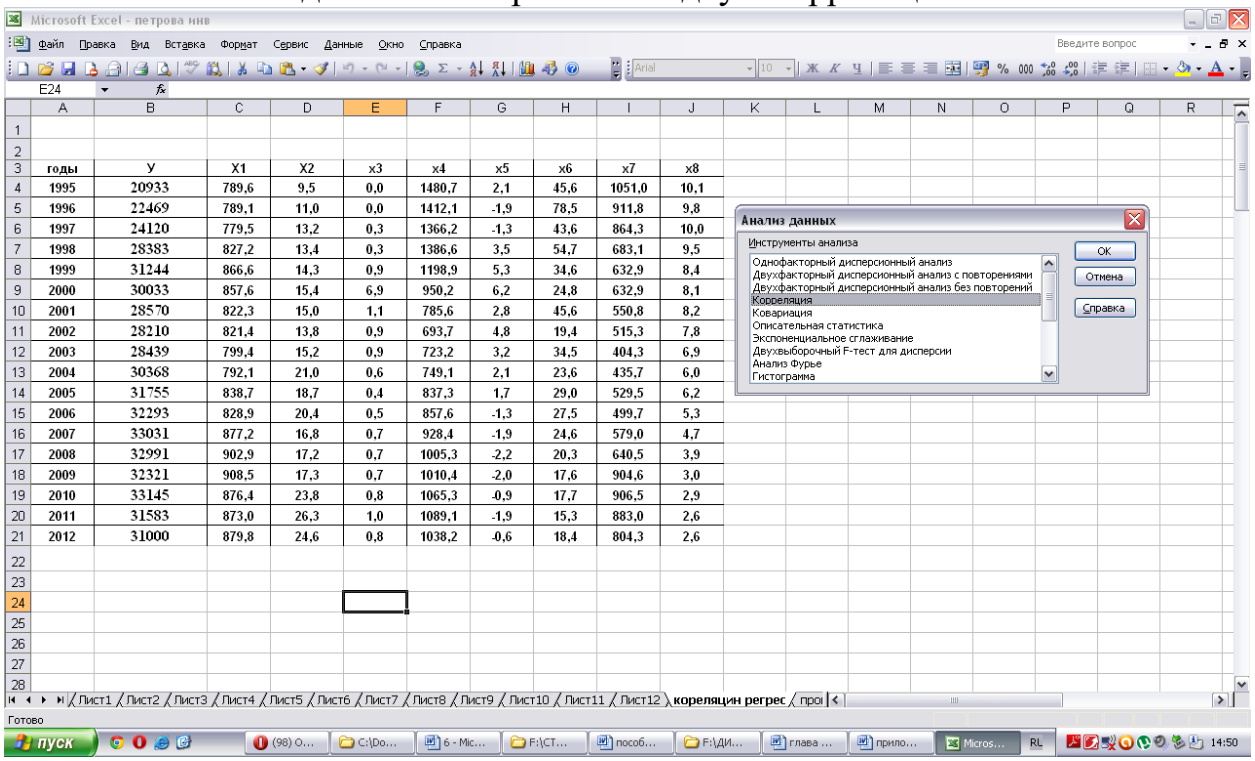


	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1																		
2																		
3	годы	У	X1	X2	x3	x4	x5	x6	x7	x8								
4	1995	20933,0	789,6	9,5	0,0	1480,7	2,1	45,6	1051,0	10,1								
5	1996	22469,0	789,1	11,0	0,0	1412,1	-1,9	78,5	911,8	9,8								
6	1997	24120,0	779,5	13,2	0,3	1366,2	-1,3	43,6	864,3	10,0								
7	1998	28383,0	827,2	13,4	0,3	1386,6	3,5	54,7	683,1	9,5								
8	1999	31244,0	866,6	14,3	0,9	1198,9	5,3	34,6	632,9	8,4								
9	2000	30033,0	857,6	15,4	6,9	950,2	6,2	24,8	632,9	8,1								
10	2001	28570,0	822,3	15,0	1,1	785,6	2,8	45,6	550,8	8,2								
11	2002	28210,0	821,4	13,8	0,9	693,7	4,8	19,4	515,3	7,8								
12	2003	28439,0	799,4	15,2	0,9	723,2	3,2	34,5	404,3	6,9								
13	2004	30368,0	792,1	21,0	0,6	749,1	2,1	23,6	435,7	6,0								
14	2005	31755,0	838,7	18,7	0,4	837,3	1,7	29,0	529,5	6,2								
15	2006	32293,0	828,9	20,4	0,5	857,6	-1,3	27,5	499,7	5,3								
16	2007	33031,0	877,2	16,8	0,7	928,4	-1,9	24,6	579,0	4,7								
17	2008	32991,0	902,9	17,2	0,7	1005,3	-2,2	20,3	640,5	3,9								
18	2009	32321,0	908,5	17,3	0,7	1010,4	-2,0	17,6	904,6	3,0								
19	2010	33145,0	876,4	23,8	0,8	1065,3	-0,9	17,7	906,5	2,9								
20	2011	31583,0	873,0	26,3	1,0	1089,1	-1,9	15,3	883,0	2,6								
21	2012	31000,0	879,8		0,8	1038,2	-0,6	18,4	804,3	2,6								
22																		
23																		
24																		
25																		
26																		
27																		
28																		

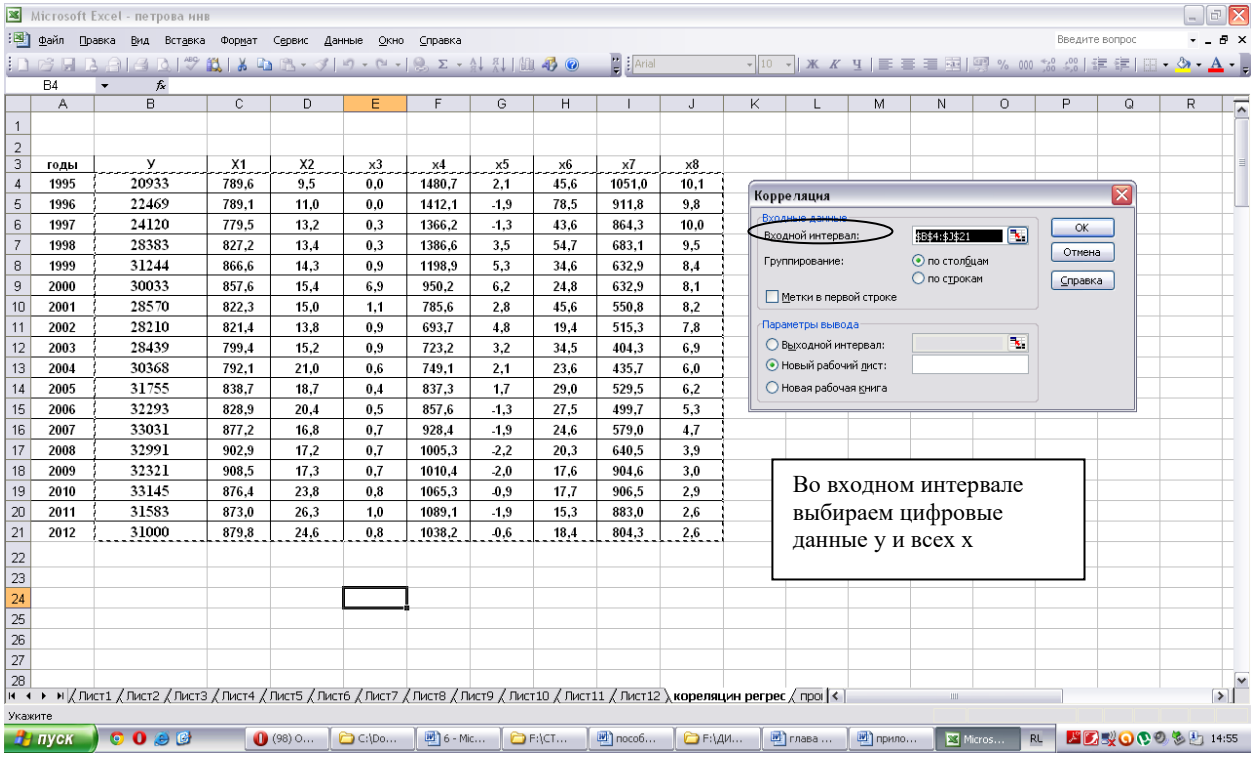
2. Выбираем на панели инструментов закладку Сервис → Анализ данных.



3. В окне Анализа данных выбираем закладку «Корреляция».

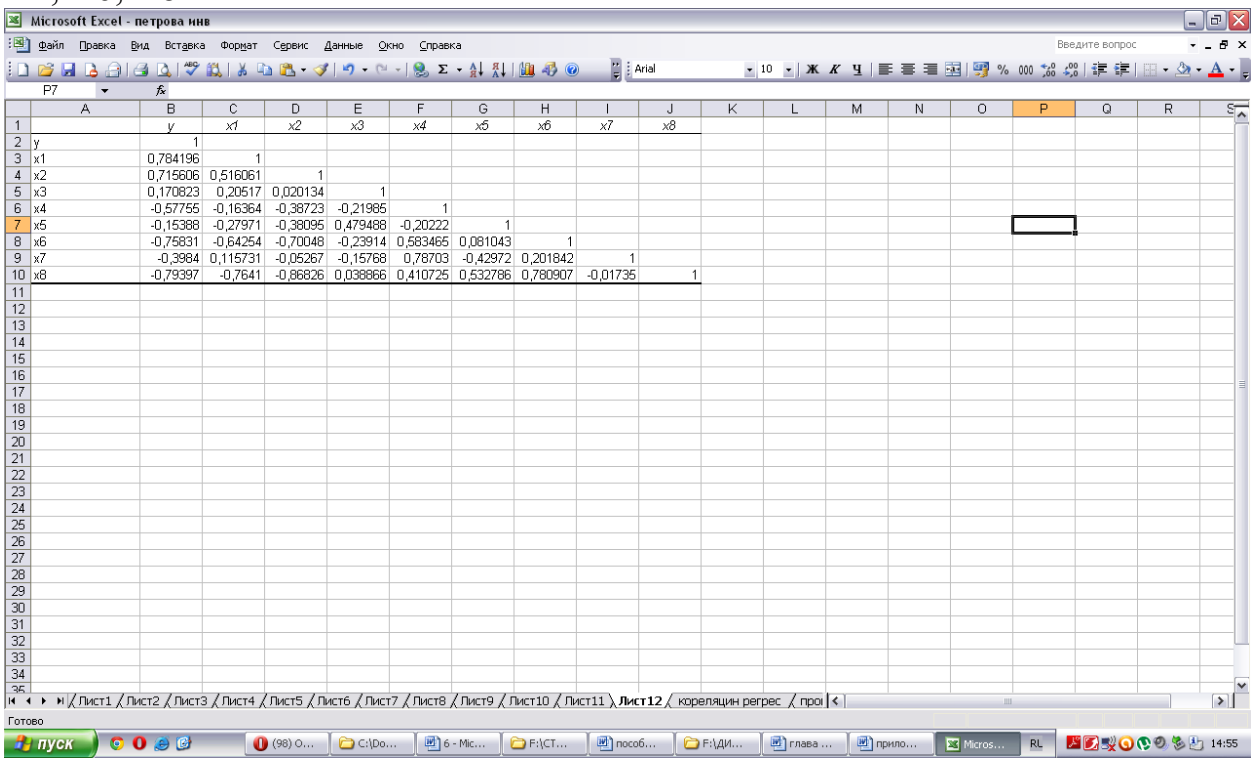


4. В появившемся окне «Корреляция» выбираем входной интервал, ставим галочку в способе группировки (по столбцам или по строкам) и выбираем входной интервал, т.е. то место, где будут отображаться полученные данные (на новом рабочем листе или в конкретно указанном месте).



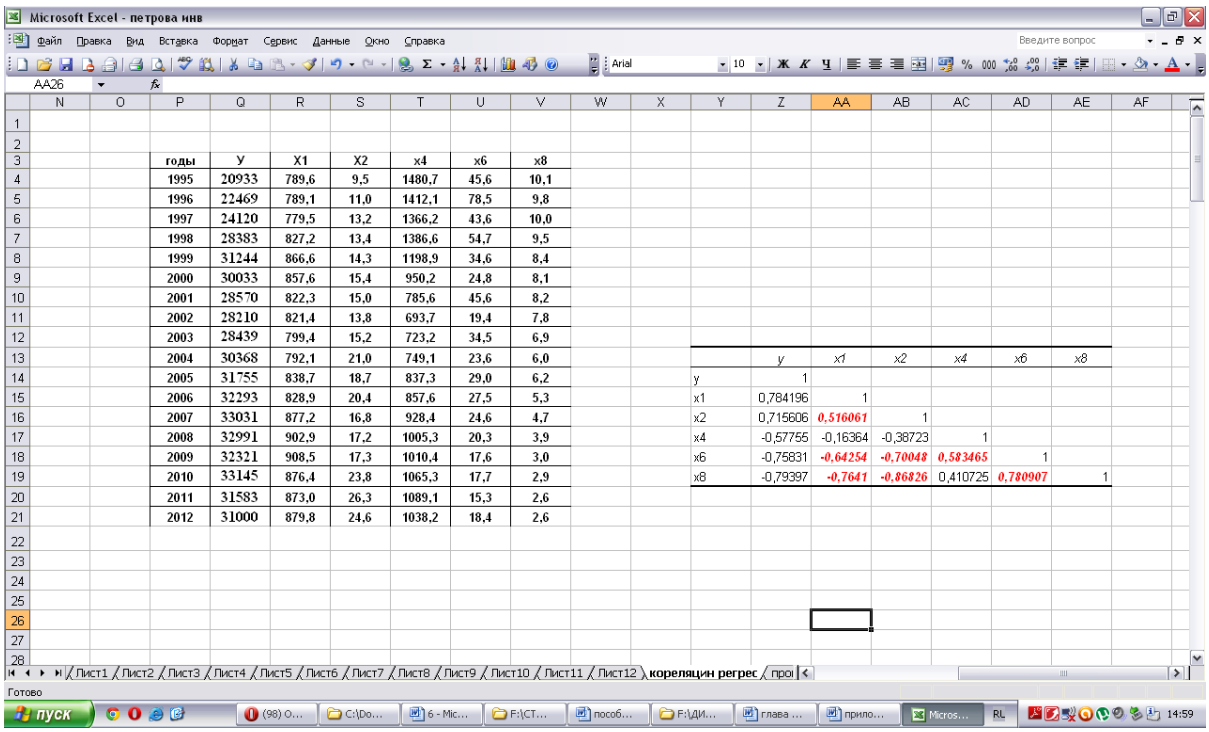
годы	y	X1	X2	x3	x4	x5	x6	x7	x8
1995	20933	789,6	9,5	0,0	1480,7	2,1	45,6	1051,0	10,1
1996	22469	789,1	11,0	0,0	1412,1	-1,9	78,5	911,8	9,8
1997	24120	779,5	13,2	0,3	1366,2	-1,3	43,6	864,3	10,0
1998	28383	827,2	13,4	0,3	1386,6	3,5	54,7	683,1	9,5
1999	31244	866,6	14,3	0,9	1198,9	5,3	34,6	632,9	8,4
2000	30033	857,6	15,4	6,9	950,2	6,2	24,8	632,9	8,1
2001	28570	822,3	15,0	1,1	785,6	2,8	45,6	550,8	8,2
2002	28210	821,4	13,8	0,9	693,7	4,8	19,4	515,3	7,8
2003	28439	799,4	15,2	0,9	723,2	3,2	34,5	404,3	6,9
2004	30368	792,1	21,0	0,6	749,1	2,1	23,6	435,7	6,0
2005	31755	838,7	18,7	0,4	837,3	1,7	29,0	529,5	6,2
2006	32293	828,9	20,4	0,5	857,6	-1,3	27,5	499,7	5,3
2007	33031	877,2	16,8	0,7	928,4	-1,9	24,6	579,0	4,7
2008	32991	902,9	17,2	0,7	1005,3	-2,2	20,3	640,5	3,9
2009	32321	908,5	17,3	0,7	1010,4	-2,0	17,6	904,6	3,0
2010	33145	876,4	23,8	0,8	1065,3	-0,9	17,7	906,5	2,9
2011	31583	873,0	26,3	1,0	1089,1	-1,9	15,3	883,0	2,6
2012	31000	879,8	24,6	0,8	1038,2	-0,6	18,4	804,3	2,6

5. Т.к. мы выбрали новый рабочий лист, корреляционная матрица появилась на новом листе. По значениям корреляционной матрицы определяем, какие факторные признаки следует исключить, а какие оставить. Рекомендательно, оставляют те факторные признаки, значения коэффициентов корреляции у которых больше 0,5. В нашем случае в модель могут быть включены X1, X2, X4, X6, X8.



	y	x1	x2	x3	x4	x5	x6	x7	x8
y	1								
x1	0,784196	1							
x2	0,715606	0,516061	1						
x3	0,170823	0,20517	0,020134	1					
x4	-0,57755	-0,16364	-0,38723	-0,21985	1				
x5	-0,15388	-0,27971	-0,38095	0,479488	-0,20222	1			
x6	-0,75831	-0,64254	-0,70048	-0,23914	0,583465	0,081043	1		
x7	-0,3984	0,115731	-0,05267	-0,15768	0,78703	-0,42972	0,201842	1	
x8	-0,79397	-0,7641	-0,86826	0,038866	0,410725	0,532786	0,780907	-0,01735	1

6. Аналогичную процедуру проделываем для оставшихся факторов. По корреляционной матрице проверяем мультиколлинеарность факторов (т.е. есть ли взаимосвязь между самими факторами), в случае, если существует тесная связь между факторами ($r \geq 0{,}5$), то включать их в одну модель нельзя. Из приведенного ниже примера видно, что существует тесная связь между X1 и X2, X6, X8; X2 и X6,X8; X4 и X6; X6 и X8. Принимаем решение исключить X1.



Microsoft Excel - петрова инв

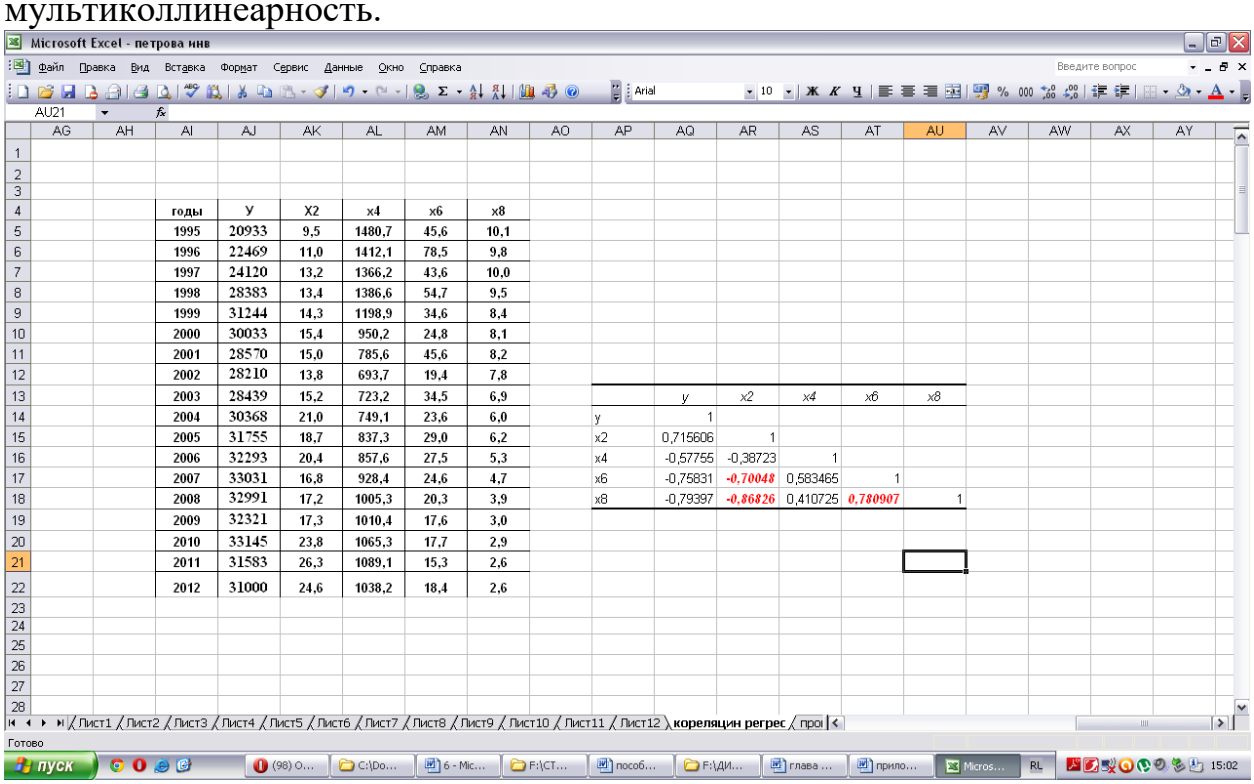
годы	У	X1	X2	x4	x6	x8
1995	20933	789,6	9,5	1480,7	45,6	10,1
1996	22469	789,1	11,0	1412,1	78,5	9,8
1997	24120	779,5	13,2	1366,2	43,6	10,0
1998	28383	827,2	13,4	1386,6	54,7	9,5
1999	31244	866,6	14,3	1198,9	34,6	8,4
2000	30033	857,6	15,4	950,2	24,8	8,1
2001	28570	822,3	15,0	785,6	45,6	8,2
2002	28210	821,4	13,8	693,7	19,4	7,8
2003	28439	799,4	15,2	723,2	34,5	6,9
2004	30368	792,1	21,0	749,1	23,6	6,0
2005	31755	838,7	18,7	837,3	29,0	6,2
2006	32293	828,9	20,4	857,6	27,5	5,3
2007	33031	877,2	16,8	928,4	24,6	4,7
2008	32991	902,9	17,2	1005,3	20,3	3,9
2009	32321	908,5	17,3	1010,4	17,6	3,0
2010	33145	876,4	23,8	1065,3	17,7	2,9
2011	31583	873,0	26,3	1089,1	15,3	2,6
2012	31000	879,8	24,6	1038,2	18,4	2,6

	y	x1	x2	x4	x6	x8
y	1					
x1	0,784196	1				
x2	0,715606	*0,516061*	1			
x4	-0,57755	-0,16364	-0,38723	1		
x6	-0,75831	*-0,64254*	*-0,70048*	*0,583465*	1	
x8	-0,79397	*-0,7641*	*-0,86826*	0,410725	*0,780907*	1

Листы: Лист1 … Лист12, кореляцин регрес, прои

Готово

7. Проводим корреляционный анализ. По результатам корреляционной матрицы видно, что существует связь между X2 и X6, X8; X6 и X8. Исключать и подбирать факторы необходимо до тех пор, пока не будет устранена мультиколлинеарность.



Microsoft Excel - петрова инв

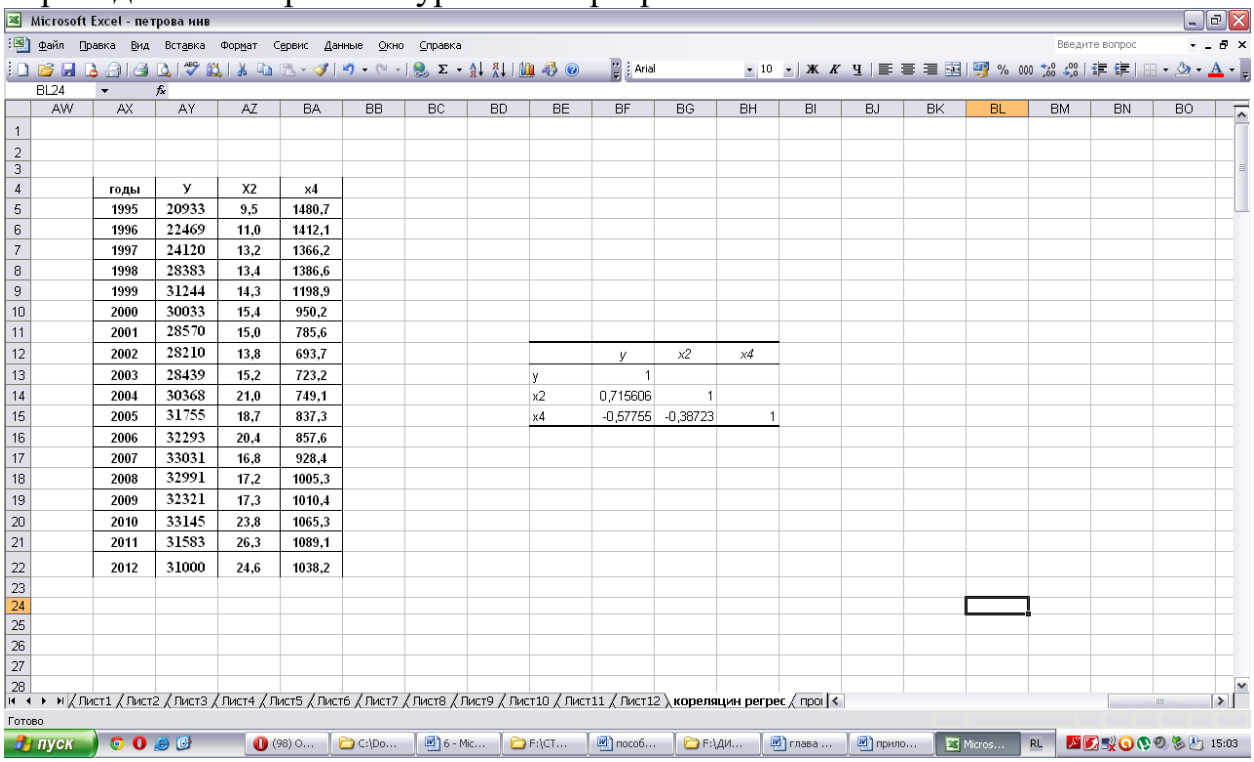
годы	У	X2	x4	x6	x8
1995	20933	9,5	1480,7	45,6	10,1
1996	22469	11,0	1412,1	78,5	9,8
1997	24120	13,2	1366,2	43,6	10,0
1998	28383	13,4	1386,6	54,7	9,5
1999	31244	14,3	1198,9	34,6	8,4
2000	30033	15,4	950,2	24,8	8,1
2001	28570	15,0	785,6	45,6	8,2
2002	28210	13,8	693,7	19,4	7,8
2003	28439	15,2	723,2	34,5	6,9
2004	30368	21,0	749,1	23,6	6,0
2005	31755	18,7	837,3	29,0	6,2
2006	32293	20,4	857,6	27,5	5,3
2007	33031	16,8	928,4	24,6	4,7
2008	32991	17,2	1005,3	20,3	3,9
2009	32321	17,3	1010,4	17,6	3,0
2010	33145	23,8	1065,3	17,7	2,9
2011	31583	26,3	1089,1	15,3	2,6
2012	31000	24,6	1038,2	18,4	2,6

	y	x2	x4	x6	x8
y	1				
x2	0,715606	1			
x4	-0,57755	-0,38723	1		
x6	-0,75831	*-0,70048*	0,583465	1	
x8	-0,79397	*-0,86826*	0,410725	*0,780907*	1

Листы: Лист1 … Лист12, кореляцин регрес, прои

Готово

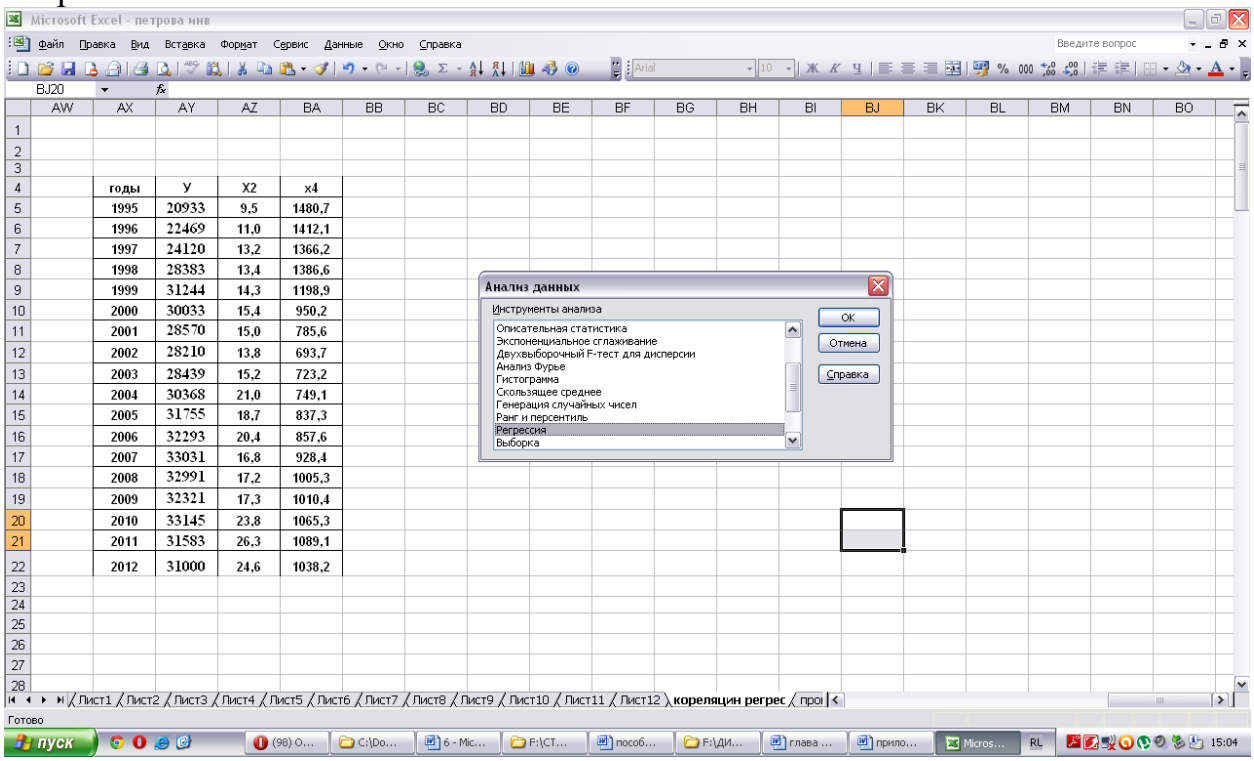
8. После устранения мультиколлинеарности между факторами и получения корреляционной матрицы со значимыми коэффициентами корреляции переходим к построению уравнения регрессии.



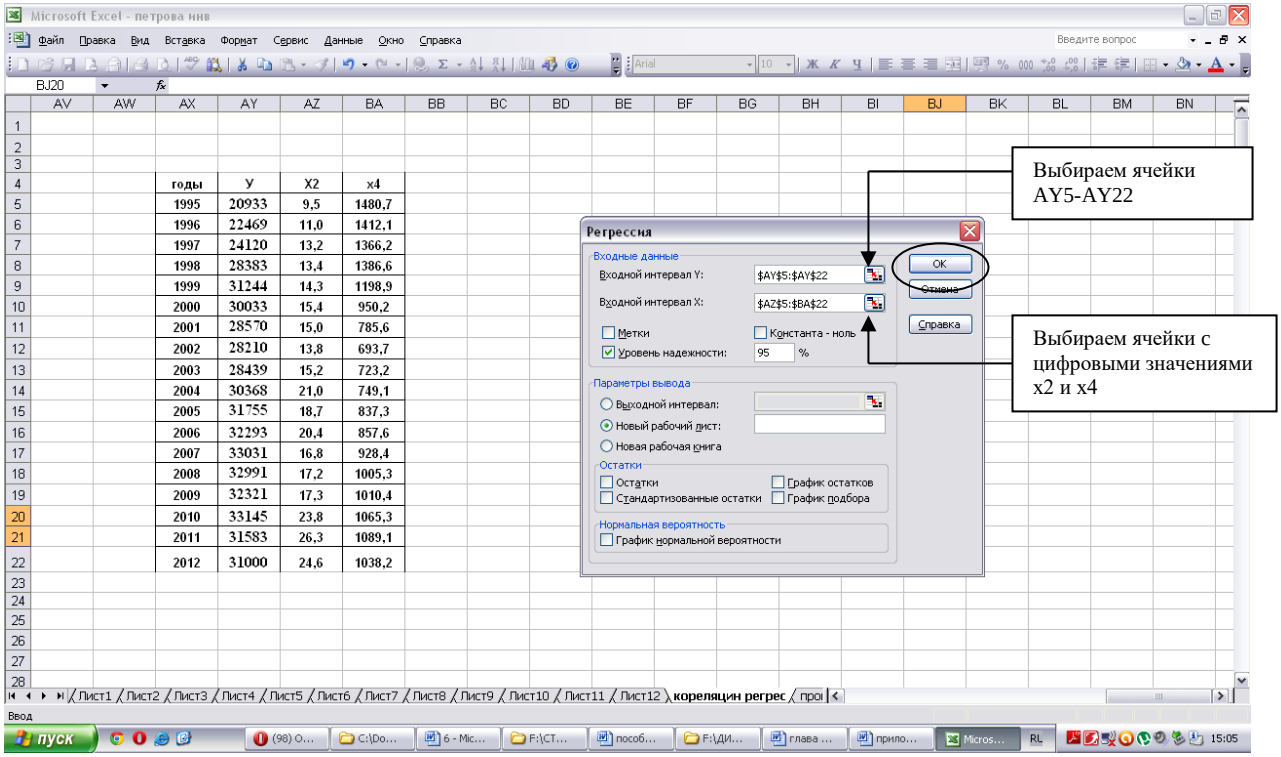
годы	У	X2	x4
1995	20933	9,5	1480,7
1996	22469	11,0	1412,1
1997	24120	13,2	1366,2
1998	28383	13,4	1386,6
1999	31244	14,3	1198,9
2000	30033	15,4	950,2
2001	28570	15,0	785,6
2002	28210	13,8	693,7
2003	28439	15,2	723,2
2004	30368	21,0	749,1
2005	31755	18,7	837,3
2006	32293	20,4	857,6
2007	33031	16,8	928,4
2008	32991	17,2	1005,3
2009	32321	17,3	1010,4
2010	33145	23,8	1065,3
2011	31583	26,3	1089,1
2012	31000	24,6	1038,2

	y	*x2*	*x4*
y	1		
x2	0,715606	1	
x4	-0,57755	-0,38723	1

9. Выбираем на панели инструментов закладку Сервис → Анализ данных → Регрессия.



10. В окне «Регрессия» выделяем входной интервал для У и входной интервал для X (выбираем только числовые данные); устанавливаем основные настройки и нажимаем ОК.

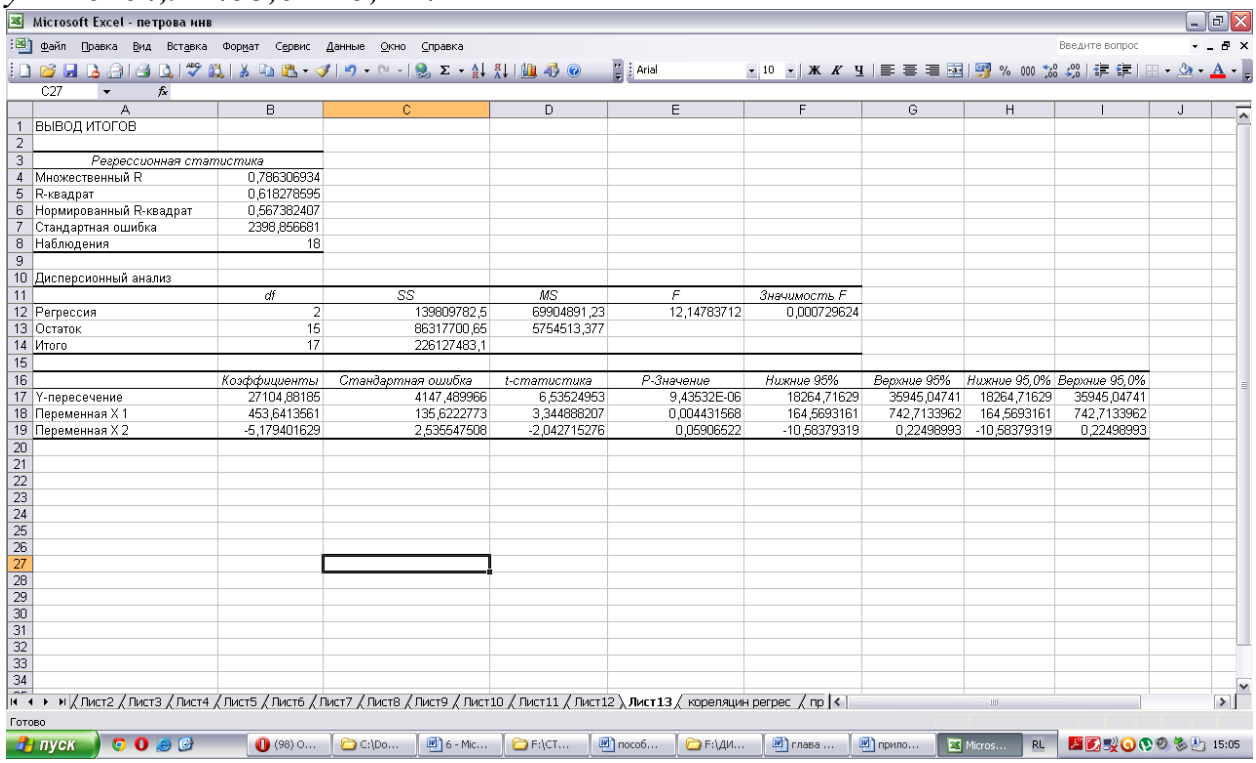


	AV	AW	AX	AY	AZ	BA	BB	BC	BD	BE	BF	BG	BH	BI	BJ	BK	BL	BM	BN
4			годы	У	X2	x4													
5			1995	20933	9,5	1480,7													
6			1996	22469	11,0	1412,1													
7			1997	24120	13,2	1366,2													
8			1998	28383	13,4	1386,6													
9			1999	31244	14,3	1198,9													
10			2000	30033	15,4	950,2													
11			2001	28570	15,0	785,6													
12			2002	28210	13,8	693,7													
13			2003	28439	15,2	723,2													
14			2004	30368	21,0	749,1													
15			2005	31755	18,7	837,3													
16			2006	32293	20,4	857,6													
17			2007	33031	16,8	928,4													
18			2008	32991	17,2	1005,3													
19			2009	32321	17,3	1010,4													
20			2010	33145	23,8	1065,3													
21			2011	31583	26,3	1089,1													
22			2012	31000	24,6	1038,2													

1. Получаем результаты регрессионного анализа.

Уравнение регрессии выглядит следующим образом:

$y = 26104,9+453,6X2-5,2X4$



	A	B	C	D	E	F	G	H	I	J
1	ВЫВОД ИТОГОВ									
2										
3	*Регрессионная статистика*									
4	Множественный R	0,786306934								
5	R-квадрат	0,618278595								
6	Нормированный R-квадрат	0,567382407								
7	Стандартная ошибка	2398,856681								
8	Наблюдения	18								
9										
10	Дисперсионный анализ									
11		*df*	*SS*	*MS*	*F*	*Значимость F*				
12	Регрессия	2	139809782,5	69904891,23	12,14783712	0,000729624				
13	Остаток	15	86317700,65	5754513,377						
14	Итого	17	226127483,1							
15										
16		*Коэффициенты*	*Стандартная ошибка*	*t-статистика*	*P-Значение*	*Нижние 95%*	*Верхние 95%*	*Нижние 95,0%*	*Верхние 95,0%*	
17	Y-пересечение	27104,88185	4147,489966	6,53524953	9,43532E-06	18264,71629	35945,04741	18264,71629	35945,04741	
18	Переменная X 1	453,6413561	135,6222773	3,344888207	0,004431568	164,5693161	742,7133962	164,5693161	742,7133962	
19	Переменная X 2	-5,179401629	2,535547508	-2,042715276	0,05906522	-10,58379319	0,22498993	-10,58379319	0,22498993	

Пример 6.3. Расчет параметров уравнения регрессии

Оценим тесноту между объемом вложений акционеров (x) и суммарным активом организации ООО «XXX» (y).

y	507,2	506,6	487,8	496	493,6	458,9	429,3	386,9	311,5	302,2	262	242,4	231,9	214,3	208,4
x	19,5	19,8	21,1	18,6	19,6	11,7	10,5	13,6	10,8	10,9	10,3	10,6	8,5	6,7	8,3

Решение:

1) Для определения связи между признаками построим поле корреляции

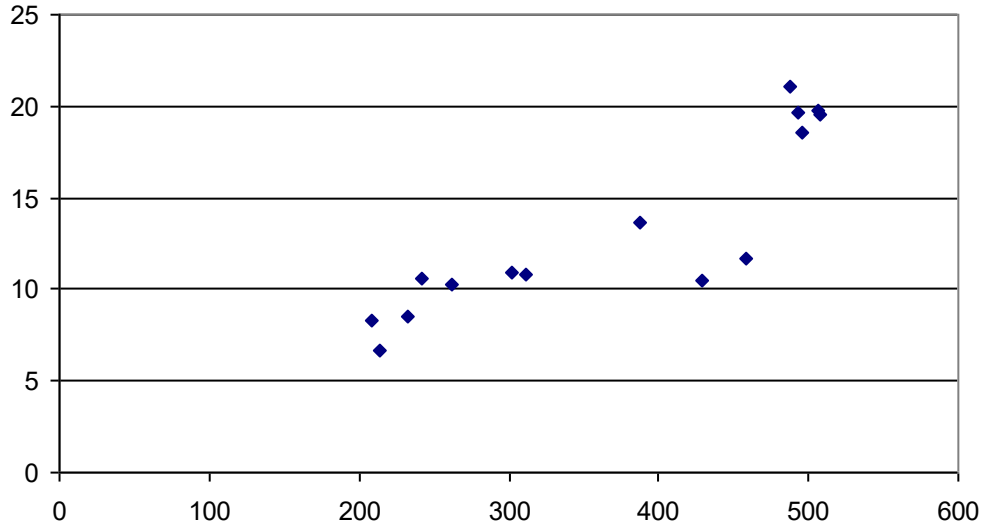


Рисунок 6.1 – Поле корреляции

Связь между признаками прямая. С увеличением объема вложений акционеров – суммарный актив увеличивается.

Таблица 6.6 – Вспомогательная таблица для расчета параметров уравнения регрессии

y	x	xy	y^2	x^2
507,2	19,5	9890,4	257251,8	380,25
506,6	19,8	10030,68	256643,6	392,04
487,8	21,1	10292,58	237948,8	445,21
496	18,6	9225,6	246016	345,96
493,6	19,6	9674,56	243641	384,16
458,9	11,7	5369,13	210589,2	136,89
429,3	10,5	4507,65	184298,5	110,25
386,9	13,6	5261,84	149691,6	184,96
311,5	10,8	3364,2	97032,25	116,64

302,2	10,9	3293,98	91324,84	118,81
262	10,3	2698,6	68644	106,09
242,4	10,6	2569,44	58757,76	112,36
231,9	8,5	1971,15	53777,61	72,25
214,3	6,7	1435,81	45924,49	44,89
208,4	8,30	1729,72	43430,56	68,89
5539	200,5	81315,34	2244972	3019,65
369,2667	13,36667	5421,023	149664,8	201,31

Рассчитаем оценки параметров парной линейной регрессии, где х – объем вложений, а у – суммарный актив.

$\tilde{y}_x = a + bx$ - уравнение прямой

$$b = \frac{\overline{yx} - \bar{y} \cdot \bar{x}}{\sigma_x^2} \qquad \bar{x} = 13{,}4 \qquad \bar{y} = 369{,}3 \qquad \overline{xy} = 5421$$

$$\sigma_x^2 = \frac{\sum x^2}{n} - \bar{x}^2 \qquad \sigma_x^2 = 201{,}3 - (13{,}4)^2 = 21{,}74$$

$$b = \frac{5421 - 13{,}4 \cdot 369{,}3}{21{,}74} = 21{,}7$$

$$a = \bar{y} - b\bar{x}$$

$$a = 369{,}3 - 21{,}7 \cdot 13{,}4 = 78{,}52$$

Получаем уравнение регрессии: $\tilde{y}_x = 78{,}52 + 21{,}7x$

Таким образом, с увеличением объема вложений на 1 % – суммарный актив увеличивается на 21,7 млрд .руб.

2) Коэффициент эластичности:

Средний коэффициент эластичности для линейной регрессии рассчитывается по формуле:

$$\bar{Э} = b \cdot \frac{\bar{x}}{\bar{y}} = 21{,}7 \cdot \frac{13{,}4}{369{,}3} = 0{,}78\%$$

Таким образом, получаем, что с увеличение объема вложений на 1% , суммарный актив увеличивается на 0,78 %.

3.) Линейный коэффициент корреляции: $r_{yx} = \frac{\overline{yx} - \bar{y} \cdot \bar{x}}{\sigma_x \cdot \sigma_y}$ или $r_{yx} = b\frac{\sigma_x}{\sigma_y}$.

σ_x – среднеквадратическое отклонение факторного признака х:

$$\sigma_x = \sqrt{\bar{x}^2 - (\bar{x})^2}$$

σ_y – среднеквадратическое отклонение результативного признака у:

$$\sigma_y = \sqrt{\bar{y}^2 - (\bar{y})^2}$$

$\bar{x} = 13{,}4$ $\bar{y} = 369{,}3$ $\bar{x}^2 = 201{,}31$ $\bar{y}^2 = 149664{,}8$

$$\sigma_x = \sqrt{201{,}31 - (13{,}4)^2} = 4{,}6$$
$$\sigma_y = \sqrt{149664{,}8 - (369{,}3)^2} = 115$$

$$r_{yx} = 21{,}7\frac{4{,}6}{115} = 0{,}868$$

Величина коэффициента регрессии свидетельствует о том, что связь между объемом вложений и суммарным активом прямая, тесная.

Расчет коэффициента детерминации:

Коэффициент детерминации составит: $r_{yx}^2 = (0{,}868)^2 = 0{,}753$, т.е. вариация суммарного актива на 75,3 % объясняется вариацией объема вложений. На долю прочих факторов, не учитываемых в регрессии, приходится 24,7%.

Пример 6.4. Практика построения уравнения множественной регрессии в Excel по пространственным данным

На основании имеющихся данных (табл.6.7) проведите КРА, постройте линейное уравнение множественной регрессии, дайте характеристику результатам, осуществите прогноз:

Таблица 6.7 – Данные по Оренбургской области за 2012 год

Города и районы	Y	X_1	X_2	X_3
Акбулакский район	581	11615	19,5	442
г.Абдулино	910	18932	13,7	267
г.Бугуруслан	519	19233	13,4	776
г.Бузулук	2610	24213	13,6	929
г.Новотроицк	703	21238	12,3	610
г.Оренбург	53383	24046	13,2	1308
г.Орск	6103	17939	13,6	1329

Y - численность студентов государственных и муниципальных образовательных учреждений ВПО в городах и районах Оренбургской области в 2012 г., человек;
X_1 – среднемесячная номинальная начисленная заработная плата работников организаций, рублей;
X_2 – общий коэффициент рождаемости в расчете на 1000 человек населения;

X_3 – численность учащихся образовательных учреждений начального профессионального образования, человек.

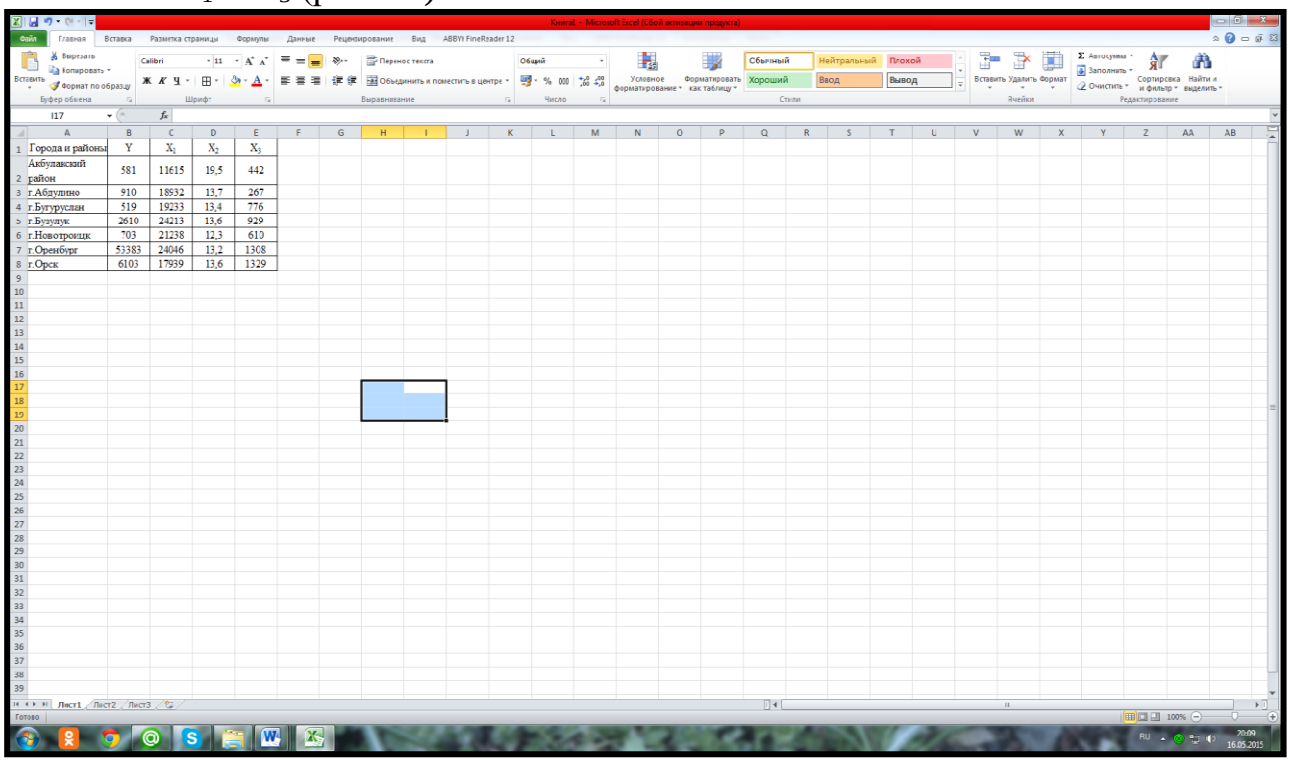
Ход процесса и решение:

1. Преобразуем исходные цифровые данные в файл Excel (рис.6.2).

2. Выбираем на панели инструментов закладку Данные → Анализ данных→ Корреляция (рис.6.3).

3. В появившемся окне «Корреляция» выбираем входной интервал, ставим галочку в способе группировки (по столбцам) и выбираем входной интервал, т.е. то место, где будут отображаться полученные данные (на новом рабочем листе или в конкретно указанном месте - ячейке) (рис.6.4).

4. Так как выбрана конкретная ячейка на рабочем листе, корреляционная матрица появилась на этом же листе. По значениям корреляционной матрицы определяем, какие факторные признаки следует исключить, а какие оставить. Рекомендовано оставлять те факторные признаки, значения коэффициентов корреляции которых больше 0,5. В нашем случае в модель могут быть включены X_1 и X_3 (рис.6.2).



Города и районы	Y	X_1	X_2	X_3
Акбулакский район	581	11615	19,5	442
г.Абдулино	910	18532	13,7	267
г.Бугуруслан	519	19233	13,4	776
г.Бузулук	2610	24213	13,6	929
г.Новотроицк	703	21238	12,3	610
г.Оренбург	53383	24046	13,2	1308
г.Орск	6103	17939	13,6	1329

Рисунок 6.2 – Результаты расчетов в программе Excel

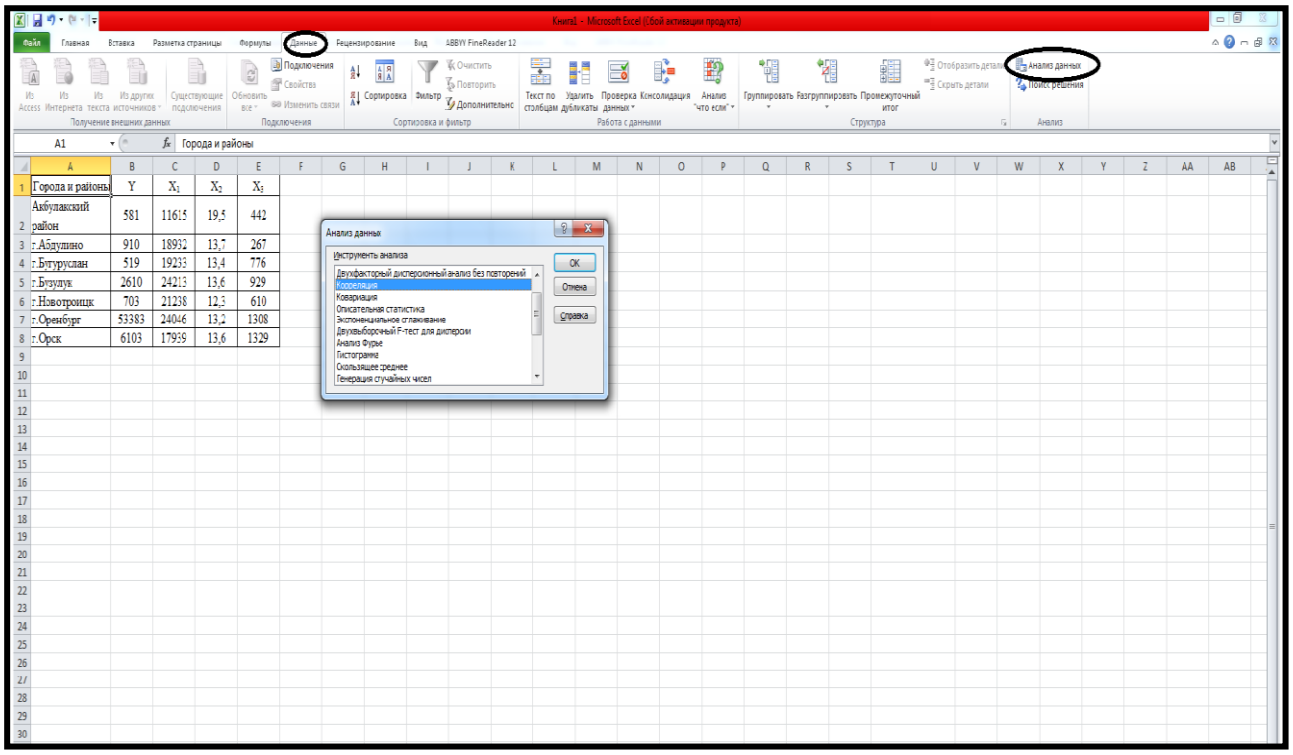


Рисунок 6.3 – Результаты расчетов в программе Excel

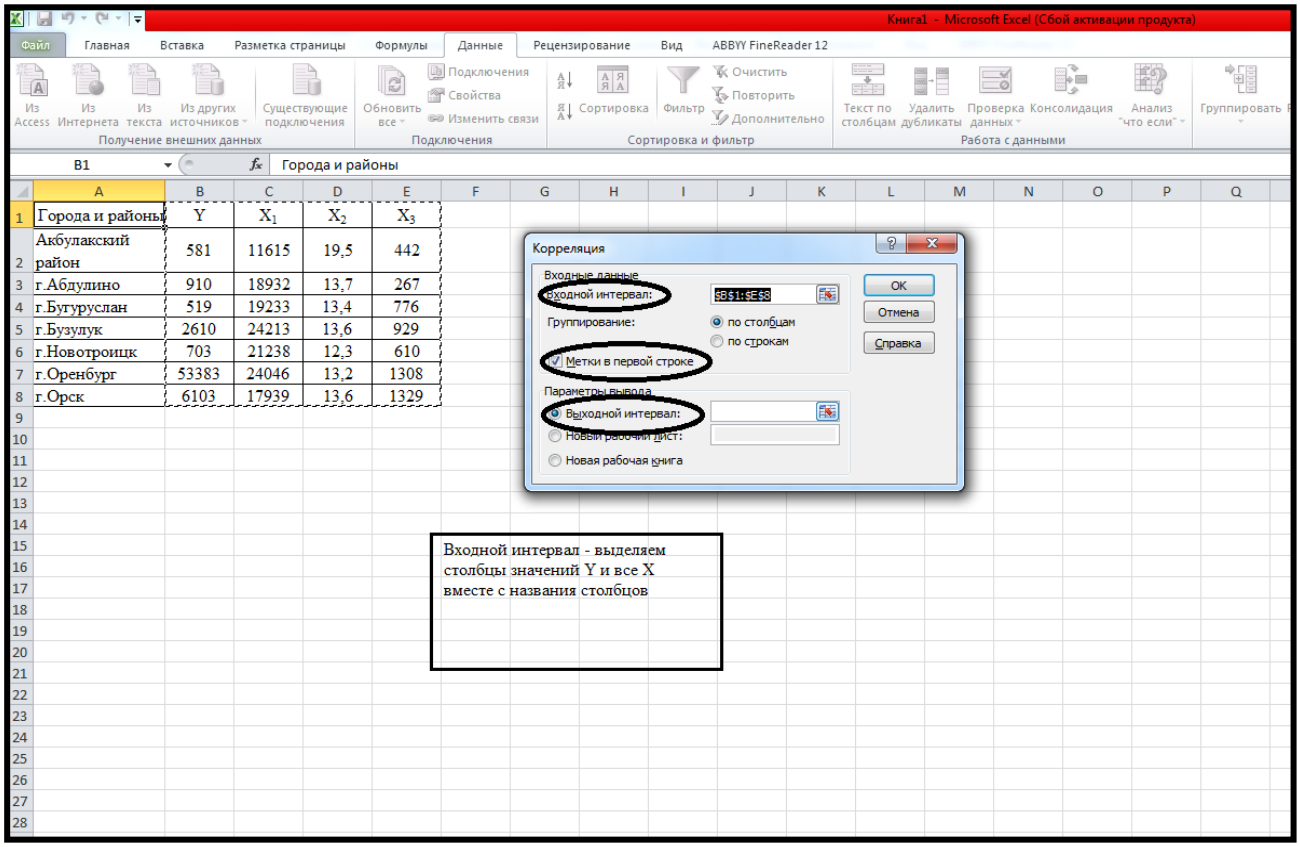
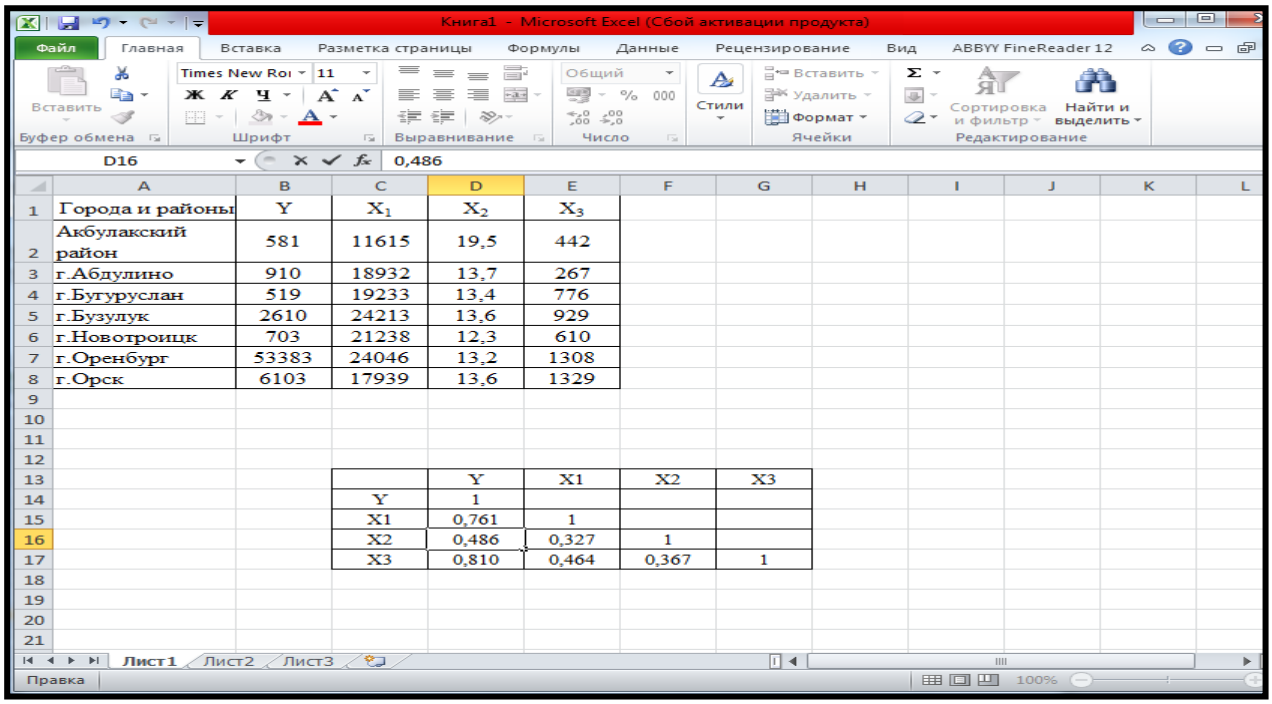


Рисунок 6.4 – Результаты расчетов в программе Excel



	A	B	C	D	E
1	Города и районы	Y	X_1	X_2	X_3
2	Акбулакский район	581	11615	19,5	442
3	г.Абдулино	910	18932	13,7	267
4	г.Бугуруслан	519	19233	13,4	776
5	г.Бузулук	2610	24213	13,6	929
6	г.Новотроицк	703	21238	12,3	610
7	г.Оренбург	53383	24046	13,2	1308
8	г.Орск	6103	17939	13,6	1329

	Y	X1	X2	X3
Y	1			
X1	0,761	1		
X2	0,486	0,327	1	
X3	0,810	0,464	0,367	1

Рисунок 6.5 – Результаты расчетов в программе Excel

5. Мультиколлинеарности между факторами нет (так как все коэффициенты корреляции между X_1 и X_2, X_1 и X_3, X_2 и X_3 меньше 0,75), далее переходим к построению уравнения регрессии.

6. Выбираем на панели инструментов закладку Данные → Анализ данных → Регрессия (рис.6.6).

7. В окне «Регрессия» выделяем входной интервал для Y и входной интервал для X (выбираем только числовые данные); устанавливаем основные настройки и нажимаем ОК (рис.6.7).

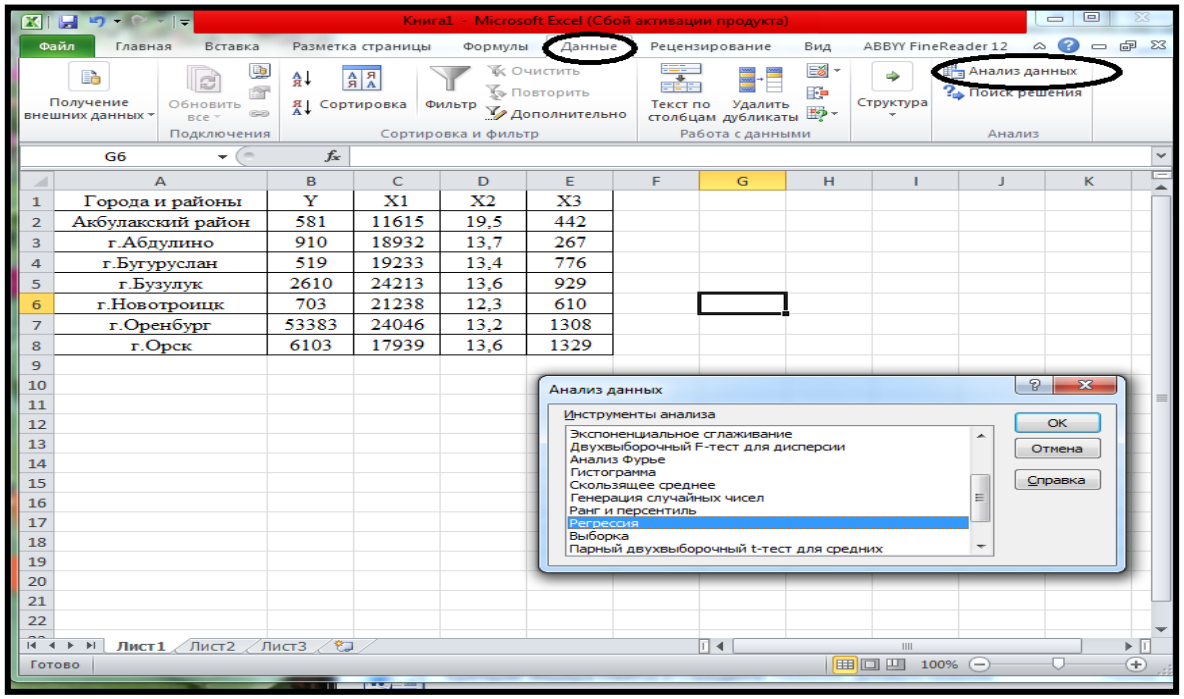
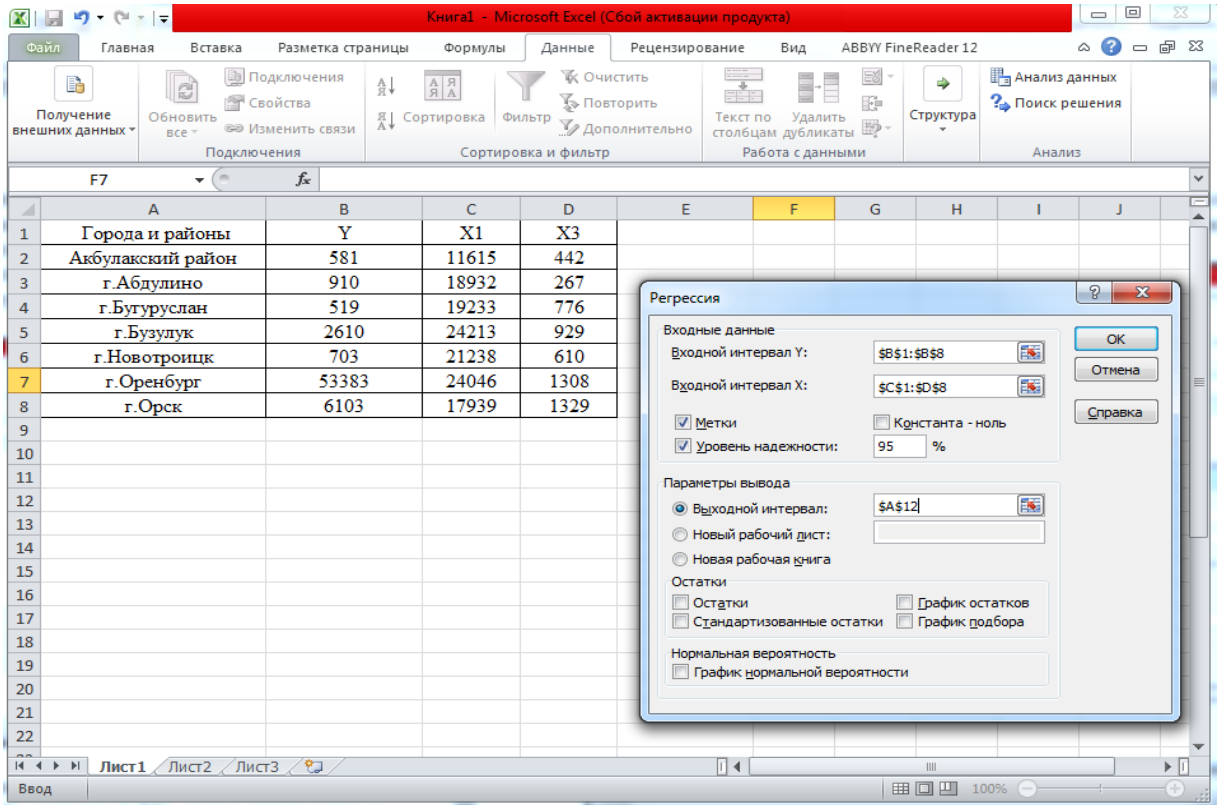


Рисунок 6.6 – Результаты расчетов в программе Excel

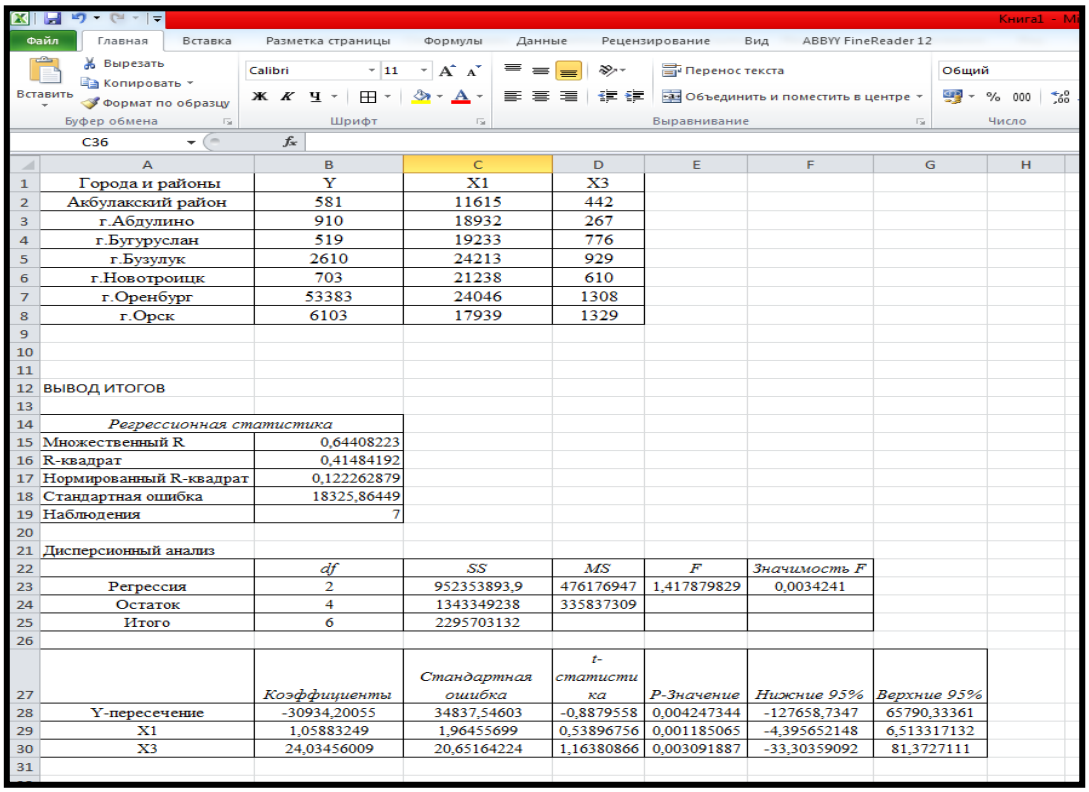


	A	B	C	D
1	Города и районы	Y	X1	X3
2	Акбулакский район	581	11615	442
3	г.Абдулино	910	18932	267
4	г.Бугуруслан	519	19233	776
5	г.Бузулук	2610	24213	929
6	г.Новотроицк	703	21238	610
7	г.Оренбург	53383	24046	1308
8	г.Орск	6103	17939	1329

Рисунок 6.7 – Результаты расчетов в программе Excel

8.Получаем результаты регрессионного анализа (рис.6.8). Уравнение регрессии выглядит следующим образом:

$$\hat{Y}_x = -30934{,}2 + 1{,}1x_1 + 24{,}1x_3$$



	A	B	C	D	E	F	G	H
1	Города и районы	Y	X1	X3				
2	Акбулакский район	581	11615	442				
3	г.Абдулино	910	18932	267				
4	г.Бугуруслан	519	19233	776				
5	г.Бузулук	2610	24213	929				
6	г.Новотроицк	703	21238	610				
7	г.Оренбург	53383	24046	1308				
8	г.Орск	6103	17939	1329				
9								
10								
11								
12	ВЫВОД ИТОГОВ							
13								
14	*Регрессионная статистика*							
15	Множественный R	0,64408223						
16	R-квадрат	0,41484192						
17	Нормированный R-квадрат	0,122262879						
18	Стандартная ошибка	18325,86449						
19	Наблюдения	7						
20								
21	Дисперсионный анализ							
22		*df*	*SS*	*MS*	*F*	*Значимость F*		
23	Регрессия	2	952353893,9	476176947	1,417879829	0,0034241		
24	Остаток	4	1343349238	335837309				
25	Итого	6	2295703132					
26								
27		*Коэффициенты*	*Стандартная ошибка*	*t-статистика*	*P-Значение*	*Нижние 95%*	*Верхние 95%*	
28	Y-пересечение	-30934,20055	34837,54603	-0,8879558	0,004247344	-127658,7347	65790,33361	
29	X1	1,05883249	1,96455699	0,53896756	0,001185065	-4,395652148	6,513317132	
30	X3	24,03456009	20,65164224	1,16380866	0,003091887	-33,30359092	81,3727111	
31								

Рисунок 6.8 – Результаты расчетов в программе Excel

9. В результате реализации процедуры корреляционного анализа (**пункт 4**) получена матрица парных коэффициентов корреляции (табл. 6.8).

Таблица 6.8 – Матрица парных коэффициентов корреляции

	Y	X_1	X_2	X_3
Y	1,000			
X_1	0,761	1,000		
X_2	0,502	0,327	1,000	
X_3	0,810	0,464	0,367	1,000

Проверку полученных значений парных коэффициентов корреляции проведем с помощью таблицы Фишера-Йейтса (приложение Ж).

В начале проверки задается *уровень значимости* (чаще всего обозначаемый буквой греческого алфавита «альфа» - α), который показывает *вероятность* принятия ошибочного решения. Возможность ошибки вытекает из того факта, что для определения взаимосвязи используются данные не всей совокупности, а лишь ее части. Обычно α принимает следующие значения: 0,05; 0,02; 0,01; 0,001. Например, если $\alpha = 0{,}05$, то это означает, что в среднем в пяти случаях из ста принятое решение о значимости (или незначимости) парных коэффициентов корреляции будет ошибочным; при $\alpha = 0{,}001$ - в одном случае из тысячи и т.д.

Вторым параметром при проверке значимости является *число степеней свободы* ν, которое в данном случае вычисляется как $\nu = n - 2$.

Коэффициенты, значения которых ***по модулю*** больше найденного критического значения, считаются значимыми.

При уровне значимости $\alpha = 0{,}05$ и числе степеней свободы $\nu = 5$ (7-2=5) критическое значение коэффициента корреляции $r_{кр} = 0{,}754$.

Значения полученных коэффициентов r_{yx1} и r_{yx3} больше критического $r_{кр}$, следовательно, они являются статистически значимыми.

Гипотеза о равенстве нулю коэффициента корреляции r_{yx2} принимается, поскольку его значение меньше найденного критического $r_{кр}$.

Для построения уравнения регрессии выбираем те факторы, у которых коэффициент корреляции с результативным признаком максимальный. В нашем случае – это факторы X_1 и X_3, значения коэффициентов связи:

$$r_{yx1} = 0{,}761 \quad r_{yx3} = 0{,}810$$

Указанные значения коэффициентов положительные, это говорит о тесной прямой связи между признаками, то есть с увеличением факторов X_1 и X_3 численность студентов государственных и муниципальных образовательных учреждений высшего профессионального образования в городах и районах области увеличивается.

10. Результаты регрессионного анализа (**пункт 8**) представим в таблице 6.9.

Коэффициент множественной корреляции R= 0,644, что говорит о прямой заметной взаимосвязи признаков в уравнении.

Таблица 6.9 – Регрессионная статистика

Показатели	Значения
Множественный R	**0,644**
R-квадрат	**0,415**
Нормированный R-квадрат	0,1223
Стандартная ошибка	18325,86
Наблюдения	7

Коэффициент детерминации R^2=0,415. Он показывает, что 41,5% вариации численности студентов государственных и муниципальных образовательных учреждений высшего профессионального образования в городах и районах области обусловлено вариацией включенных в модель факторов.

Таблица 6.10 – Регрессионная статистика

Показатели	df	SS	MS	F	Значимость F
Регрессия	2	9,52E-08	4,76E-08	14,1788	**0,0032**
Остаток	4	1,34E-09	3,36E-08		
Итого	6	2,3E-09			

Значение F-критерия Фишера равно 14,1788.

Значимость F-критерия (0,0032) показывает вероятность того, что множественный R будет равен нулю. Она крайне мала (меньше 0,05), следовательно, уравнение регрессии статистически значимо с вероятностью 95% (табл. 6.10).

Параметры уравнения регрессии представлены в таблице 6.11.

Таблица 6.11 – Параметры уравнения связи и показатели их значимости

	Коэффициенты	Стандартная ошибка	t-статистика	P-значение
Y-пересечение	-30934,2	34837,5	-0,887	**0,004**
X_1	1,1	1,9	0,538	**0,001**
X_3	24,1	20,6	1,163	**0,002**

Уравнение линейной множественной регрессии примет вид:

$$\tilde{Y}_x = -30934{,}2 + 1{,}1x_1 + 24{,}1x_3 \qquad (6.21)$$

Анализ параметров уравнения регрессии (6.21) дал следующие результаты.

При увеличении среднемесячной номинальной начисленной заработной платы работников организаций на 1000 рублей численность студентов государственных и муниципальных образовательных учреждений высшего профессионального образования в городах и районах области увеличится в среднем на 1100 чел.

С ростом численности учащихся образовательных учреждений начального профессионального образования на 1000 чел. численность студентов государственных и муниципальных образовательных учреждений высшего профессионального образования увеличится в среднем на 24100 чел.

Глава 7. СТАТИСТИЧЕСКИЕ МЕТОДЫ ИССЛЕДОВАНИЯ НЕЧИСЛОВЫХ ДАННЫХ

При исследовании степени тесноты связи между качественными признаками, каждый из которых представлен в виде альтернативных признаков, возможно использование так называемых «тетрахорических показателей». Тогда расчетная таблица состоит из четырех ячеек (обозначаемых буквами a, b, c, d). Каждая из клеток соответствует известной альтернативе того и другого признака.

	Да	Нет
Да	a	b
Нет	c	d

По этим данным рассчитываются коэффициенты ассоциации ($К_а$) и контингенции ($К_к$):

$$K_a = \frac{ad - bc}{ad + bc} \tag{7.1}$$

$$K_k = \frac{ad - bc}{\sqrt{(a+b)\cdot(b+d)\cdot(a+c)\cdot(c+d)}} \tag{7.2}$$

где a, b, c, d – числа (частоты) в четырехклеточной таблице.

Связь считается подтвержденной, если $К_а \geq 0{,}5$, $К_k \geq 0{,}3$.

Когда каждый из качественных признаков состоит более чем из двух групп, то для определения тесноты связи применяют коэффициенты взаимной сопряженности Пирсона и Чупрова.

Коэффициент Пирсона вычисляется по формуле:

$$P = \sqrt{\frac{\chi^2}{1+\chi^2}}\,, \tag{7.3}$$

где χ^2- критерий взаимной сопряженности, определяется суммой квадратов частот каждой клетки таблицы f_{ij}^2 к произведению частот итоговых соответствующего столбца m_j и строки n_i минус единица:

$$\chi^2 = \sum_i \sum_j \frac{f^2{}_{ij}}{m_j n_i} - 1\,. \tag{7.4}$$

Коэффициент Чупрова вычисляется по формуле:

$$P = \sqrt{\frac{\chi^2}{\sqrt{(K_1 - 1)(K_2 - 1)}}} \tag{7.5}$$

где $К_1$ - число значений (групп) первого признака;

$К_2$ - число значений (групп) второго признака

Коэффициент корреляции знаков, или коэффициент Фехнера, основан на оценке степени согласованности направлений отклонений индивидуальных значений факторного и результативного признаков от соответствующих средних. Вычисляется он следующим образом:

$$\hat{E}_{\hat{o}} = \frac{n_a - n_b}{n_a + n_b}, \tag{7.6}$$

где n_a - число совпадений знаков отклонений индивидуальных величин от средней;

n_b - число несовпадений.

Коэффициент Фехнера может принимать значения от -1 до +1. Кф = 1 свидетельствует о возможном наличии прямой связи, Кф =-1 свидетельствует о возможном наличии обратной связи.

Ранговый коэффициент корреляции Спирмена измеряет степень согласованности двух различных ранжировок $X^{(j)}$ и $X^{(k)}$ одного и того же множества из n объектов и рассчитывается по формуле:

$$r_c = 1 - \frac{6}{n^3 - n}\sum_{i=1}^{n}(x_i^{(j)} - x_i^{(k)})^2 \tag{7.7}$$

Причем $-1<r_c<1$. Соответственно -1 означает, что ранжировки противоположны, 1 - они совпадают, г =0 - между ранжировками связь отсутствует.

Связанные ранги возникают в случае дележки мест в ранжировке. Объектам, которые делят места, приписывается ранг, равный среднему арифметическому соответствующих мест.

Например: объекты делят места с 3 по 5, тогда (3+4+5)/3 = 4 и объектам на 3, 4 и 5 местах приписывается ранг 4.

Для связанных рангов формула коэффициента корреляции Спирмена усложняется:

$$r_c = \frac{\frac{1}{6}(n^3 - n) - \sum_{i=1}^{n}(x_i^{(j)} - x_i^{(k)})^2 - T^{(j)} - T^{(k)}}{\left[\frac{1}{6}(n^3 - n) - 2T^{(j)}\right]^{\frac{1}{2}}\left[\frac{1}{6}(n^3 - n) - 2T^{(k)}\right]^{\frac{1}{2}}} \tag{7.8}$$

где $T^{(l)} = \frac{1}{12}\sum_{t=1}^{m}(n_t^{(l)})^3 - n_t^{(l)}$ здесь m — число групп, для которых имеются связанные ранги, n_t — число рангов, входящих в t-ю группу.

Проверка гипотезы H_0: $r_c = 0$.

Если n >10, то статистика:

$$\gamma_n = \frac{r_c\sqrt{n-2}}{\sqrt{1 - r_c^2}} \tag{7.9}$$

распределена по закону Стьюдента с (n-2) степенями свободы.

При $\gamma_n > t_{\alpha/2}(n-2)$ гипотеза H_0 отклоняется, иначе - не отклоняется.

Если n < 10, то рассчитывают

$$r_c^{\max} = \frac{2S_c}{(n^3-n)/3}, \tag{7.10}$$

где S_c извлекается из таблицы для уровня значимости α/2 и если $|r_c| > r_c^{\max}$, то H_0 отвергается.

Ранговый коэффициент корреляции Кендалла.

Расчетная формула имеет вид:

$$r_k = \frac{S}{\frac{1}{2}n(n-1)} = \frac{P-Q}{P+Q}, \quad S = P-Q \tag{7.11}$$

Ранжируем все элементы по признаку $x^{(1)}$, по ряду другого признака $x^{(2)}$ подсчитываем для каждого ранга число последующих рангов, превышающих данный (их сумму обозначим P) и число последующих рангов ниже данного (их сумму обозначим Q)

Значения коэффициента $-1 < r_k < 1$.

На практике при $n \geq 10$, $r_c = \frac{3}{2} r_k$

Для связанных рангов формула коэффициента корреляции Кендалла имеет вид:

$$r_k^* = \frac{r_k - \frac{2(u^{(1)}+u^{(2)})}{n(n-1)}}{\sqrt{(1-\frac{2u^{(1)}}{n(n-1)})(1-\frac{2u^{(2)}}{n(n-1)})}}, \quad u^{(l)} = \frac{1}{2}\sum_{t=1}^{m^{(l)}} n_t^{(l)}(n_t^{(l)}-1), \; l=1,2. \tag{7.12}$$

Коэффициент конкордации (согласованности) Кендалла. Измеряет степень тесноты, статистической связи между m различными ранжировками:

$$W = \frac{12}{m^2(n^3-n)} \sum_{i=1}^{n} \left(\sum_{l=1}^{m} x_i^{(l)} - \frac{m(n+1)}{2}\right)^2. \tag{7.13}$$

Причем $0 < W < 1$, если $W = 1$ — ранжировки совпадают, $W = 0$ — связь между ранжировками отсутствует.

Для связанных рангов

$$W = \frac{\sum_{i=1}^{n}\left(\sum_{l=1}^{m} x_i^{(l)} - \frac{m(n+1)}{2}\right)}{\frac{1}{12}m^2(n^3-n) - m\sum_{l=1}^{m} T^{(l)}} \tag{7.14}$$

Проверка гипотезы H_0: W=0,

- при $n>7$ рассчитывают статистику $\gamma_n = m(n-1)W$ и при $\gamma_n > \chi_\alpha^2(n-1)$ гипотеза отклоняется, иначе не отклоняется.

Пример 7.1. Расчет коэффициентов ассоциации и контингенции

Оцените связь между образованием рабочих цеха и наличием прогулов:

	Полное среднее	Неполное среднее
Имеют прогулы	21	5
Не имеют прогулы	83	116

РЕШЕНИЕ:

Коэффициент ассоциации: $K_a = \frac{ad - bc}{ad + bc}$

a=21
b=5
c=83
d=116

$$K_a = \frac{ad - bc}{ad + bc} = \frac{21*116 - 5*83}{21*116 + 5*83} = \frac{2436 - 415}{2436 + 415} = \frac{2021}{2851} = 0{,}709$$

Коэффициент контингенции: $K_a = \frac{ad - bc}{\sqrt{(a+b)(b+d)(a+c)(c+d)}}$

a	b	a+b
c	d	c+d
a+c	b+d	a+b+c+d

21	5	26
83	116	199
104	121	225

$$K_a = \frac{ad - bc}{\sqrt{(a+b)(b+d)(a+c)(c+d)}} = \frac{2021}{\sqrt{26*121*104*199}} = \frac{2021}{8069} = 0{,}25$$

ОТВЕТ: связь между образованием рабочих цеха и наличием прогулов существенная.

Глава 8. МНОГОМЕРНЫЕ МЕТОДЫ СТАТИСТИЧЕСКОГО ИССЛЕДОВАНИЯ

Многомерные статистические методы – это раздел математической статистики, изучающий методы сбора многомерных статистических данных, их систематизации и обработки в целях выявления характера и структуры взаимосвязей между компонентами исследуемого многомерного признака, получения практических выводов[1].

Методы математической статистики, используемые для построения оптимальных планов сбора, систематизации и обработки многомерных статистических данных, направленные на выявление характера и структуры взаимосвязей между компонентами исследуемого многомерного признака и предназначенные для получения научных и практических выводов.

По содержанию методы многомерного статистического анализа могут быть условно разделены на три основные группы:

1) методы многомерного статистического анализа многомерных распределений и их основных характеристик;

2) методы многомерного статистического анализа характера и структуры взаимосвязей между компонентами исследуемого многомерного признака;

3) методы многомерного статистического анализа геометрической структуры исследуемой совокупности многомерных наблюдений.

Методы первой группы охватывают лишь те ситуации, в которых обрабатываемые наблюдения имеют вероятностную природу, т. е. интерпретируются как выборка из соответствующей генеральной совокупности. К основным задачам этого подраздела относятся: статистическое оценивание исследуемых многомерных распределений, их основных числовых характеристик и параметров, исследование распределения вероятностей для ряда статистик, с помощью которых строятся статистические критерии проверки различных гипотез о вероятностной природе анализируемых многомерных данных.

Вторая группа методов объединяет в себе понятия и результаты, обслуживающие такие методы и модели математического статистического анализа, как множественная регрессия, многомерный дисперсный анализ, факторный анализ и др. Результаты данных методов могут быть условно разделены на два основных типа:

а) построение наилучших статистических оценок для параметров этих моделей и анализ их свойств (точности, а в вероятностной постановке - законов их распределения, доверительных областей и т. д.);

б) построение статистических критериев для проверки различных гипотез о структуре исследуемых взаимосвязей.

---

[1] Орлов А. И. Эконометрика. Учебник для вузов. Изд. 3-е, исправленное и дополненное. - М.: Изд-во «Экзамен», 2004. - 576 с.

Третья группа - объединяет в себе понятия и результаты таких моделей и схем, как дискриминантный анализ, анализ многомерного шкалирования. Узловым во всех этих схемах является понятие расстояния (меры близости, меры сходства) между анализируемыми элементами. При этом анализируемыми могут быть как реальные объекты, так и сами показатели.

Прикладное назначение методов многомерного статистического анализа состоит в основном в обслуживании трех проблем:

1) проблема статистического исследования зависимостей между анализируемыми показателями;

2) проблема классификации элементов (объектов или показателей) в общей (нестрогой) постановке, чтобы всю анализируемую совокупность элементов, статистически представленную в виде матрицы, разбить на сравнительно небольшое число однородных групп;

3) проблема снижения размерности исследуемого факторного пространства и отбора наиболее информативных показателей[1].

К методам многомерного статистического анализа относятся: дисперсионный анализ, дискриминантный анализ, факторный анализ, канонический анализ, кластерный анализ, логлинейный анализ и др.

Дисперсионный анализ.

В процессе наблюдения за исследуемым объектом качественные факторы произвольно или заданным образом изменяются. Конкретная реализация фактора (например, определенный температурный режим, выбранное оборудование или материал) называется уровнем фактора или способом обработки. Модель дисперсионного анализа с фиксированными уровнями факторов называют моделью I, модель со случайными факторами - моделью II. Благодаря варьированию фактора можно исследовать его влияние на величину отклика. В настоящее время общая теория дисперсионного анализа разработана для моделей I.

В зависимости от количества факторов, определяющих вариацию результативного признака, дисперсионный анализ подразделяют на однофакторный и многофакторный.

Основными схемами организации исходных данных с двумя и более факторами являются:

- перекрестная классификация, характерная для моделей I, в которых каждый уровень одного фактора сочетается при планировании эксперимента с каждой градацией другого фактора;

- иерархическая (гнездовая) классификация, характерная для модели II, в которой каждому случайному, наудачу выбранному значению одного фактора соответствует свое подмножество значений второго фактора.

[1] Словарь социолингвистических терминов. — М.: Российская академия наук. Институт языкознания. Российская академия лингвистических наук. Ответственный редактор: доктор филологических наук В.Ю. Михальченко. 2006.

Если одновременно исследуется зависимость отклика от качественных и количественных факторов, т.е. факторов смешанной природы, то используется ковариационный анализ[1].

При обработке данных эксперимента наиболее разработанными и поэтому распространенными считаются две модели. Их различие обусловлено спецификой планирования самого эксперимента. В модели дисперсионного анализа с фиксированными эффектами исследователь намеренно устанавливает строго определенные уровни изучаемого фактора. Термин «фиксированный эффект» в данном контексте имеет тот смысл, что самим исследователем фиксируется количество уровней фактора и различия между ними. При повторении эксперимента он или другой исследователь выберет те же самые уровни фактора. В модели со случайными эффектами уровни значения фактора выбираются исследователем случайно из широкого диапазона значений фактора, и при повторных экспериментах, естественно, этот диапазон будет другим.

Таким образом, данные модели отличаются между собой способом выбора уровней фактора, что, очевидно, в первую очередь влияет на возможность обобщения полученных экспериментальных результатов. Для дисперсионного анализа однофакторных экспериментов различие этих двух моделей не столь существенно, однако в многофакторном дисперсионном анализе оно может оказаться весьма важным.

При проведении дисперсионного анализа должны выполняться следующие статистические допущения: независимо от уровня фактора величины отклика имеют нормальный (Гауссовский) закон распределения и одинаковую дисперсию. Такое равенство дисперсий называется гомогенностью. Таким образом, изменение способа обработки сказывается лишь на положении случайной величины отклика, которое характеризуется средним значением или медианой. Поэтому все наблюдения отклика принадлежат сдвиговому семейству нормальных распределений.

Говорят, что техника дисперсионного анализа является "робастной". Этот термин, используемый статистиками, означает, что данные допущения могут быть в некоторой степени нарушены, но несмотря на это, технику можно использовать.

При неизвестном законе распределения величин отклика используют непараметрические (чаще всего ранговые) методы анализа.

В основе дисперсионного анализа лежит разделение дисперсии на части или компоненты. Вариацию, обусловленную влиянием фактора, положенного в основу группировки, характеризует межгрупповая дисперсия σ^2. Она является мерой вариации частных средних по группам $\bar{x}_j$ вокруг общей средней $\bar{x}$ и определяется по формуле:

[1] www.sutd.ru

$$\overline{\sigma}^2 = \frac{\sum_{j=1}^{k} (\overline{x}_j - \overline{x})^2 \times n_j}{\sum_{j=1}^{k} n_j}, \qquad (8.1)$$

где k - число групп;

n_j - число единиц в j-ой группе;

$\overline{x}_j$ - частная средняя по j-ой группе;

$\overline{x}$ - общая средняя по совокупности единиц.

Вариацию, обусловленную влиянием прочих факторов, характеризует в каждой группе внутригрупповая дисперсия σ_j^2.

$$\sigma_j^2 = \frac{\sum_{i=1}^{n} (x_{ij} - \overline{x}_j)^2}{n}. \qquad (8.2)$$

Между общей дисперсией σ_0^2, внутригрупповой дисперсией σ^2 и межгрупповой дисперсией $\overline{\sigma}^2$ существует соотношение:

$$\sigma_0^2 = \overline{\sigma}^2 + \sigma^2. \qquad (8.3)$$

Внутригрупповая дисперсия объясняет влияние неучтенных при группировке факторов, а межгрупповая дисперсия объясняет влияние факторов группировки на среднее значение по группе[1].

В таблице 8.1 представлен общий вид вычисления значений, с помощью дисперсионного анализа.

Таблица 8.1 – Базовая таблица дисперсионного анализа

Компоненты дисперсии	Сумма квадратов	Число степеней свободы	Средний квадрат	Математическое ожидание среднего квадрата
Межгрупповая	$Q_1 = n\sum_{i=1}^{m} (\overline{x}_{i*} - \overline{x}_{**})^2$	m-1	$S_1^2 = Q_1/(m-1)$	$M(S_1^2 = \begin{cases} \frac{n}{m-1}\sum_{i=1}^{m}(F_i - F_*)^2 + \\ \sigma^2 (\text{модель I}) \\ n\sigma_F^2 + \sigma^2 (\text{модель II}) \end{cases}$
Внутригрупповая	$Q_2 = \sum_{i=1}^{m}\sum_{j=1}^{n} (x_{ij} - \overline{x}_{i*})^2$	mn-m	$S_1^2 = Q_2/(mn-m)$	$M(S^2) = \sigma^2$
Общая	$Q = \sum_{i=1}^{m}\sum_{j=1}^{n} (x_{ij} - \overline{x}_{**})^2$	mn-1		

Таким образом, процедура однофакторного дисперсионного анализа состоит в проверке гипотезы H_0 о том, что имеется одна группа однородных

[1] Гмурман В.Е. Теория вероятностей и математическая статистика. – М.: Высшая школа, 2003.-523с.

экспериментальных данных против альтернативы о том, что таких групп больше, чем одна. Под однородностью понимается одинаковость средних значений и дисперсий в любом подмножестве данных. При этом дисперсии могут быть как известны, так и неизвестны заранее. Если имеются основания полагать, что известная или неизвестная дисперсия измерений одинакова по всей совокупности данных, то задача однофакторного дисперсионного анализа сводится к исследованию значимости различия средних в группах данных[1].

Следует сразу же отметить, что принципиальной разницы между многофакторным и однофакторным дисперсионным анализом нет. Многофакторный анализ не меняет общую логику дисперсионного анализа, а лишь несколько усложняет ее, поскольку, кроме учета влияния на зависимую переменную каждого из факторов по отдельности, следует оценивать и их совместное действие. Таким образом, то новое, что вносит в анализ данных многофакторный дисперсионный анализ, касается в основном возможности оценить межфакторное взаимодействие. Тем не менее, по-прежнему остается возможность оценивать влияние каждого фактора в отдельности. В этом смысле процедура многофакторного дисперсионного анализа (в варианте ее компьютерного использования) несомненно более экономична, поскольку всего за один запуск решает сразу две задачи: оценивается влияние каждого из факторов и их взаимодействие[2].

Общая схема двухфакторного эксперимента, данные которого обрабатываются дисперсионным анализом имеет вид:

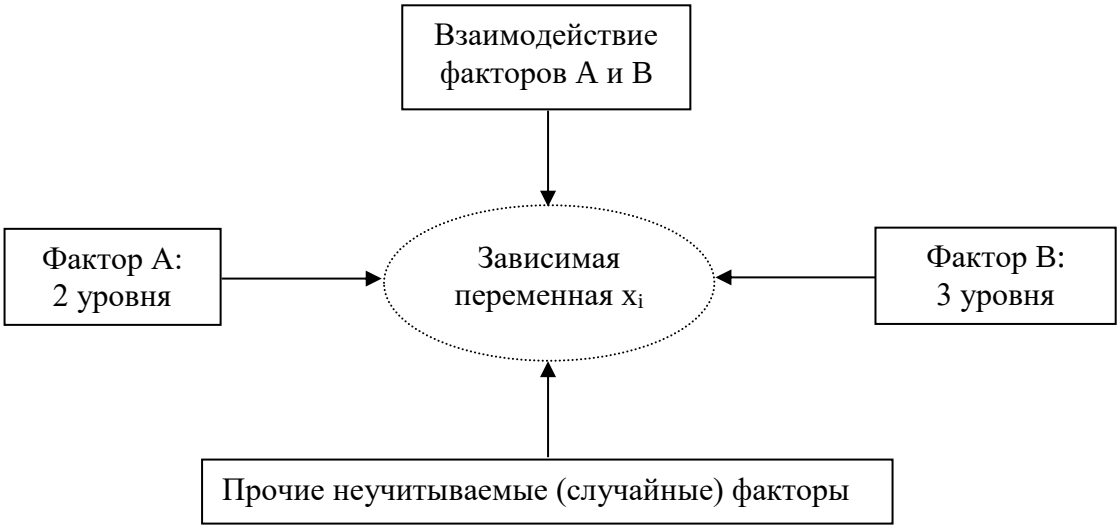


Рисунок 8.1 – Схема двухфакторного эксперимента

Данные, подвергаемые многофакторному дисперсионному анализу, часто обозначают в соответствии с количеством факторов и их уровней.

[1] Кремер Н.Ш. Теория вероятности и математическая статистика. М.: Юнити – Дана, 2002.-343с.
[2] Гусев А.Н. Дисперсионный анализ в экспериментальной психологии. – М.: Учебно-методический коллектор «Психология», 2000.-136с.

Дискриминантный анализ.

Дискриминантный анализ используется для принятия решения о том, какие переменные различают (дискриминируют) две или более возникающие совокупности (группы). Например, некий исследователь в области образования может захотеть исследовать, какие переменные относят выпускника средней школы к одной из трех категорий: (1) поступающий в колледж, (2) поступающий в профессиональную школу или (3) отказывающийся от дальнейшего образования или профессиональной подготовки. Для этой цели исследователь может собрать данные о различных переменных, связанных с учащимися школы. После выпуска большинство учащихся естественно должно попасть в одну из названных категорий. Затем можно использовать Дискриминантный анализ для определения того, какие переменные дают наилучшее предсказание выбора учащимися дальнейшего пути.

Медик может регистрировать различные переменные, относящиеся к состоянию больного, чтобы выяснить, какие переменные лучше предсказывают, что пациент, вероятно, выздоровел полностью (группа 1), частично (группа 2) или совсем не выздоровел (группа 3). Биолог может записать различные характеристики сходных типов (групп) цветов, чтобы затем провести анализ дискриминантной функции, наилучшим образом разделяющей типы или группы.

Основная идея дискриминантного анализа заключается в том, чтобы определить, отличаются ли совокупности по среднему какой-либо переменной (или линейной комбинации переменных), и затем использовать эту переменную, чтобы предсказать для новых членов их принадлежность к той или иной группе.

Вероятно, наиболее общим применением дискриминантного анализа является включение в исследование многих переменных с целью определения тех из них, которые наилучшим образом разделяют совокупности между собой. Например, исследователь в области образования, интересующийся предсказанием выбора, который сделают выпускники средней школы относительно своего дальнейшего образования, произведет с целью получения наиболее точных прогнозов регистрацию возможно большего количества параметров обучающихся, например, мотивацию, академическую успеваемость и т.д.

Модель. Другими словами, вы хотите построить «модель», позволяющую лучше всего предсказать, к какой совокупности будет принадлежать тот или иной образец. В следующем рассуждении термин «в модели» будет использоваться для того, чтобы обозначать переменные, используемые в предсказании принадлежности к совокупности; о неиспользуемых для этого переменных будем говорить, что они «вне модели».

Пошаговый анализ с включением. В пошаговом анализе дискриминантных функций модель дискриминации строится по шагам. Точнее, на каждом шаге просматриваются все переменные и находится та из них, которая вносит наибольший вклад в различие между совокупностями. Эта

переменная должна быть включена в модель на данном шаге, и происходит переход к следующему шагу.

Пошаговый анализ с исключением. Можно также двигаться в обратном направлении, в этом случае все переменные будут сначала включены в модель, а затем на каждом шаге будут устраняться переменные, вносящие малый вклад в предсказания. Тогда в качестве результата успешного анализа можно сохранить только «важные» переменные в модели, то есть те переменные, чей вклад в дискриминацию больше остальных.

F для включения, F для исключения. Эта пошаговая процедура «руководствуется» соответствующим значением F для включения и соответствующим значением F для исключения. Значение F статистики для переменной указывает на ее статистическую значимость при дискриминации между совокупностями, то есть, она является мерой вклада переменной в предсказание членства в совокупности. Если вы знакомы с пошаговой процедурой множественной регрессии, то вы можете интерпретировать значение F для включения (исключения) в том же самом смысле, что и в пошаговой регрессии.

Расчет на случай. Пошаговый дискриминантный анализ основан на использовании статистического уровня значимости. Поэтому по своей природе пошаговые процедуры рассчитывают на случай, так как они «тщательно перебирают» переменные, которые должны быть включены в модель для получения максимальной дискриминации. При использовании пошагового метода исследователь должен осознавать, что используемый при этом уровень значимости не отражает истинного значения альфа, то есть, вероятности ошибочного отклонения гипотезы H0 (нулевой гипотезы, заключающейся в том, что между совокупностями нет различия).

В общем, Дискриминантный анализ - это очень полезный инструмент:

1) для поиска переменных, позволяющих относить наблюдаемые объекты в одну или несколько реально наблюдаемых групп,

2) для классификации наблюдений в различные группы.

Факторный анализ.

Главными целями факторного анализа являются:

1) сокращение числа переменных (редукция данных);

2) определение структуры взаимосвязей между переменными, т.е. классификация переменных.

Поэтому факторный анализ используется либо как метод сокращения данных, либо как метод классификации (термин факторный анализ впервые ввел Thurstone, 1931).

Зависимость между переменными можно обнаружить с помощью диаграммы рассеяния. Полученная путем подгонки линия регрессии дает графическое представление зависимости. Если определить новую переменную на основе линии регрессии, изображенной на диаграмме рассевания, то такая переменная будет включать в себя наиболее существенные характеристики

обеих переменных. Итак, фактически, вы сократили число переменных и заменили две одной.

Факторный анализ - это метод разведочного анализа.

Канонический анализ.

Каноническая корреляция позволяет исследовать зависимость между двумя наборами переменных (и применяется для проверки гипотез или как метод разведочного анализа). Например, исследователь в сфере образования может оценить зависимость между навыками по трем учебным дисциплинам и оценками по пяти школьным предметам. Социолог может исследовать зависимость между прогнозами социальных изменений, печатаемыми в двух газетах, и реальными изменениями, оцененными с помощью четырех различных статистических признаков. Исследователь-медик может изучить зависимость между различными неблагоприятными факторами и появлением определенной группы симптомов. Во всех этих случаях нас интересуют зависимости между двумя группами переменных. Для анализа таких зависимостей и предназначен метод канонической корреляции.

Кластерный анализ.

Термин кластерный анализ (впервые ввел Tryon, 1939) в действительности включает в себя набор различных алгоритмов классификации. В отличие от комбинационных группировок кластерный анализ приводит к разбиению на группы с учетом всех группировочных признаков одновременно. При этом, как правило, не указаны четкие границы каждой группы, а также неизвестно заранее, сколько же групп целесообразно выделить в исследуемой совокупности.

Методы кластерного анализа используются для решения задачи уменьшения количества переменных, необходимых для описания исследуемого явления, объекта, системы и выделения в этом пространстве наиболее важных характеристик или скрытых факторов.

Метод кластерного анализа позволяет строить классификацию n объектов посредством объединения их в группы или кластеры на основе критерия минимума расстояния в пространстве т переменных, описывающих объекты. Метод позволяет находить множества объектов на заданное число кластеров.

Исходные данные для кластерного анализа представляются в виде матрицы размером Т х П, содержащей информацию одного из следующих трех типов: тип - измерения X_{ij} значений т переменных для п объектов; тип - квадратная (т = п) матрица расстояний между парами объектов; тип - квадратная (т = п) матрица близостей для всех пар п объектов.

В качестве ориентира для определения возможного числа кластеров используется графическое изображение процесса агломерации, представленное дендрограммой. В расчет также принимались величин расстояний между объединяемыми элементами.

Так как исследуемые показатели имеют разную размерность, то для приведения их в сопоставимый вид проводится процедура стандартизации.

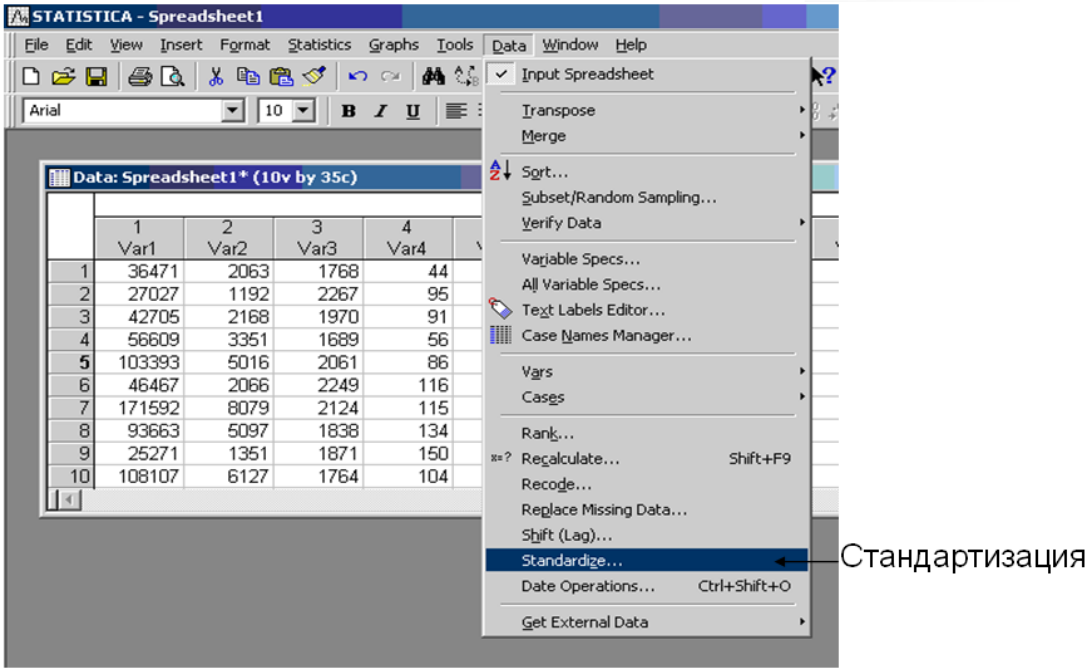
Логлинейный анализ.

Логлинейный анализ предоставляет «высоконаучный» подход к рассмотрению таблиц сопряженности (при проведении разведочного анализа данных и проверке гипотез), и иногда рассматривается в качестве эквивалента дисперсионного анализа для таблиц частот. Более точно, он позволяет пользователю проверить статистическую значимость влияния различных факторов (например, пол, место жительства и т.п.) и их взаимодействий при построении таблиц сопряженности.

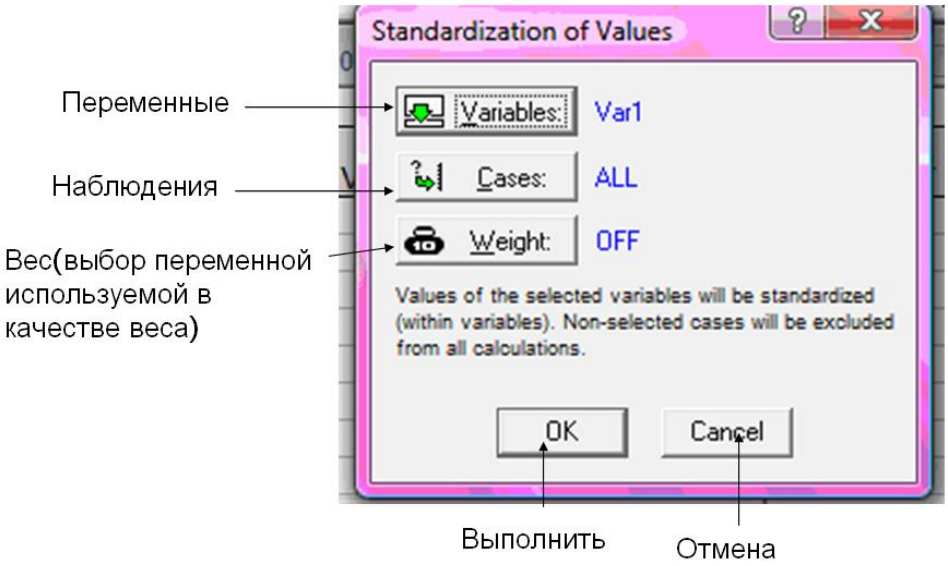
Помимо представленных методов исследования и с учетом ускоренного процесса развития статистического и эконометрического инструментария на практике применяется широкий спектр и других методов многомерного статистического исследования.

Пример 8.1. Реализация процедуры «Кластерный анализ» в пакете Statistica

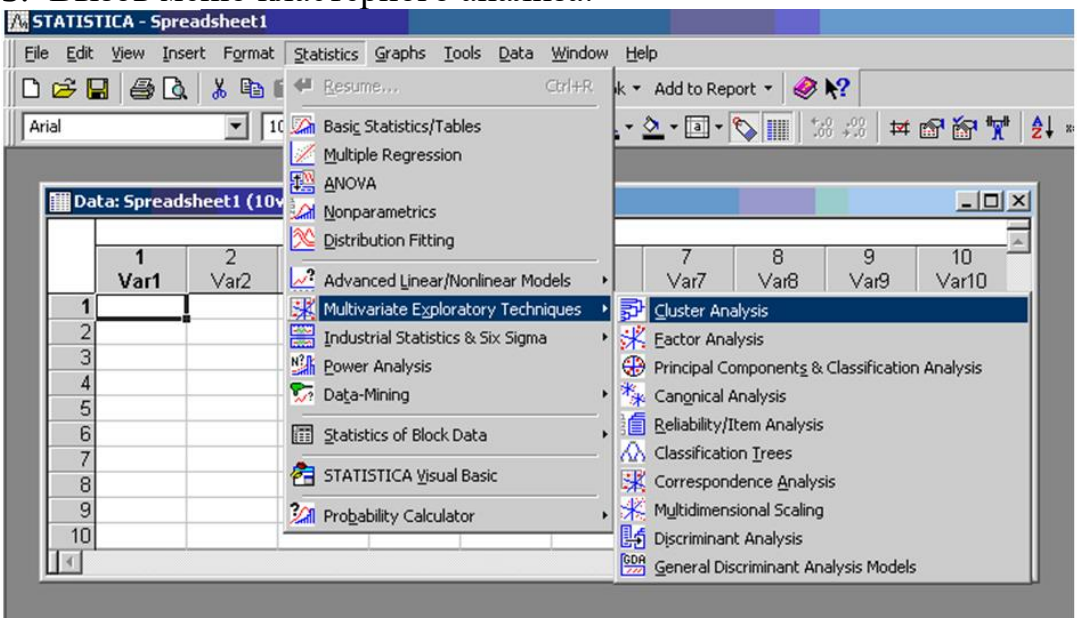
1. Нормирование (стандартизация) данных:



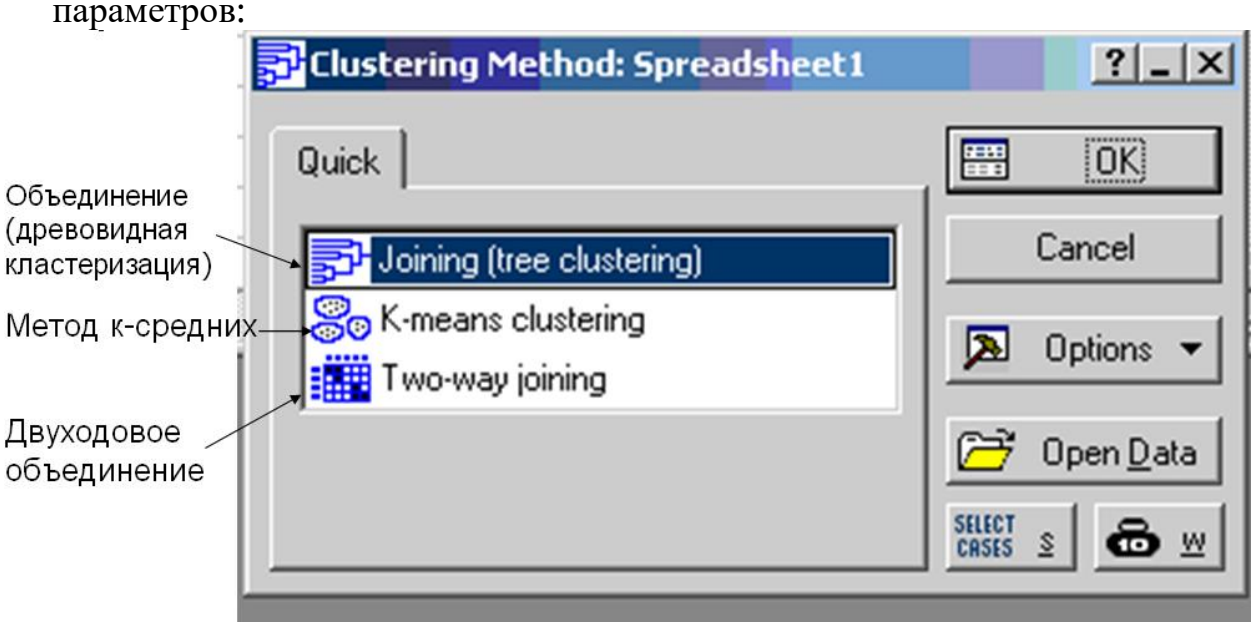
2. Меню стандартизации данных:



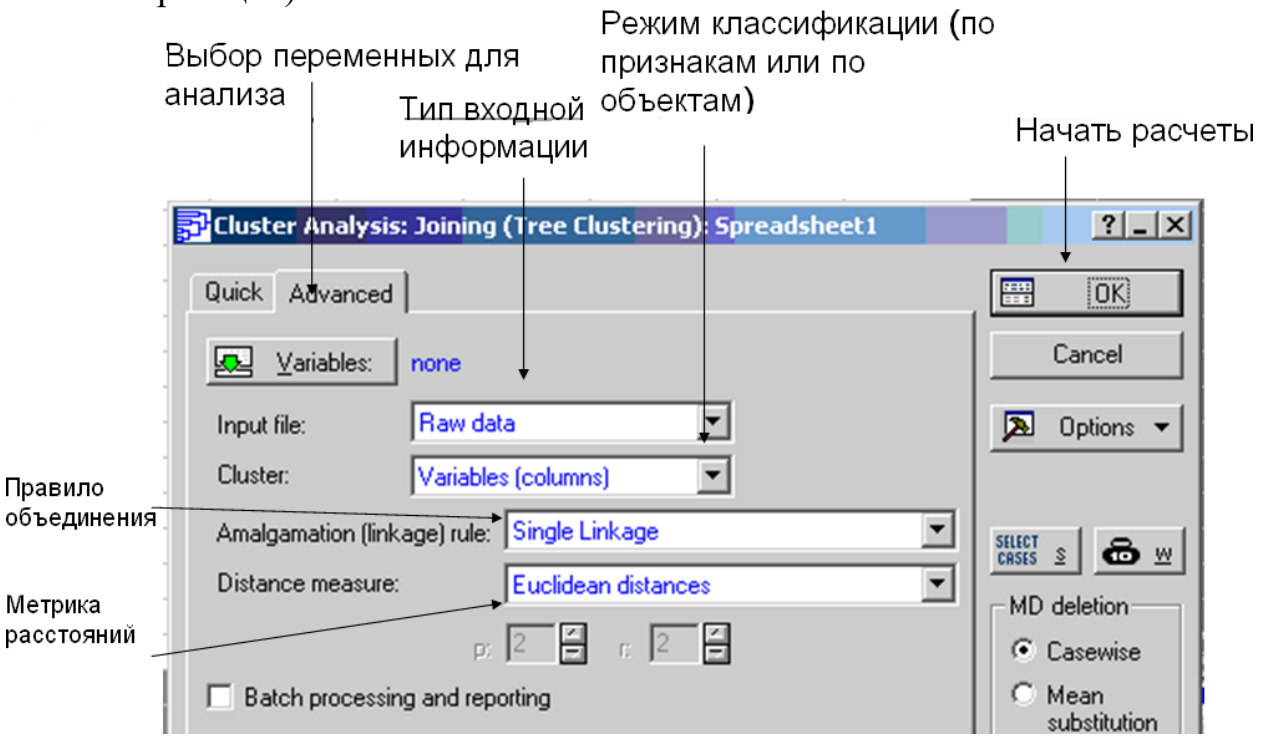
3. Вызов меню кластерного анализа:



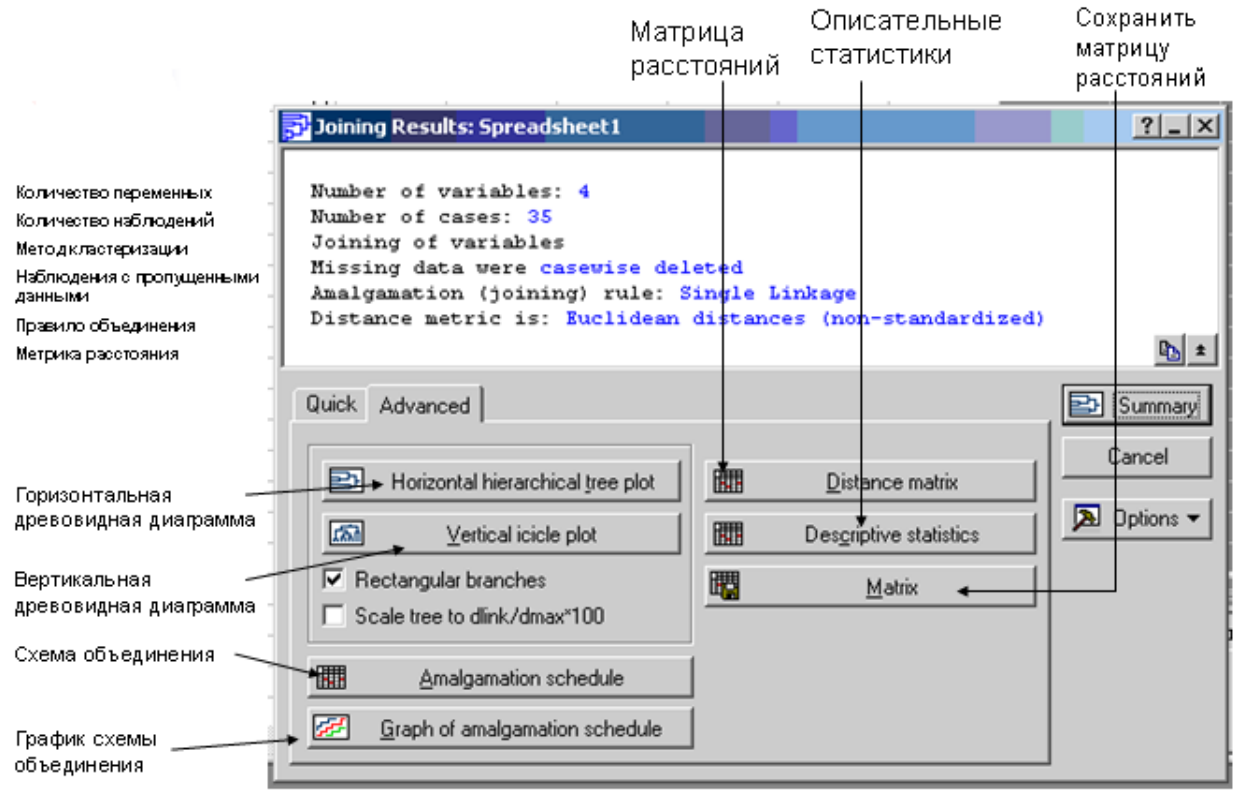
4. Иерархическая (древовидная) кластеризация. Задание выходных параметров:



5. Панель модуля кластерного анализа: объединение (древовидная кластеризация):

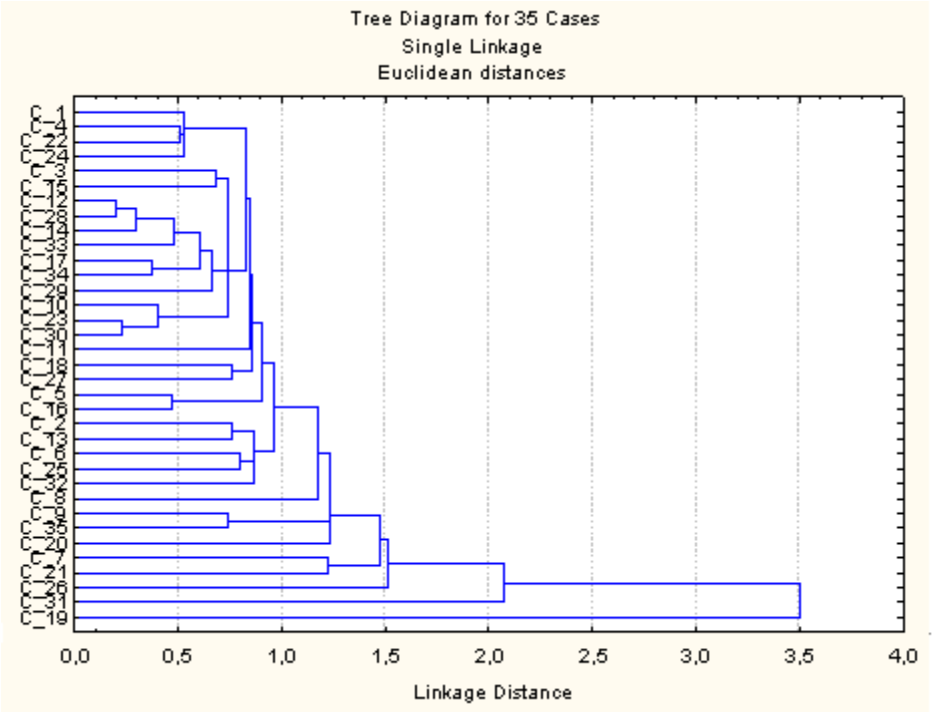


6. Результаты объединения:



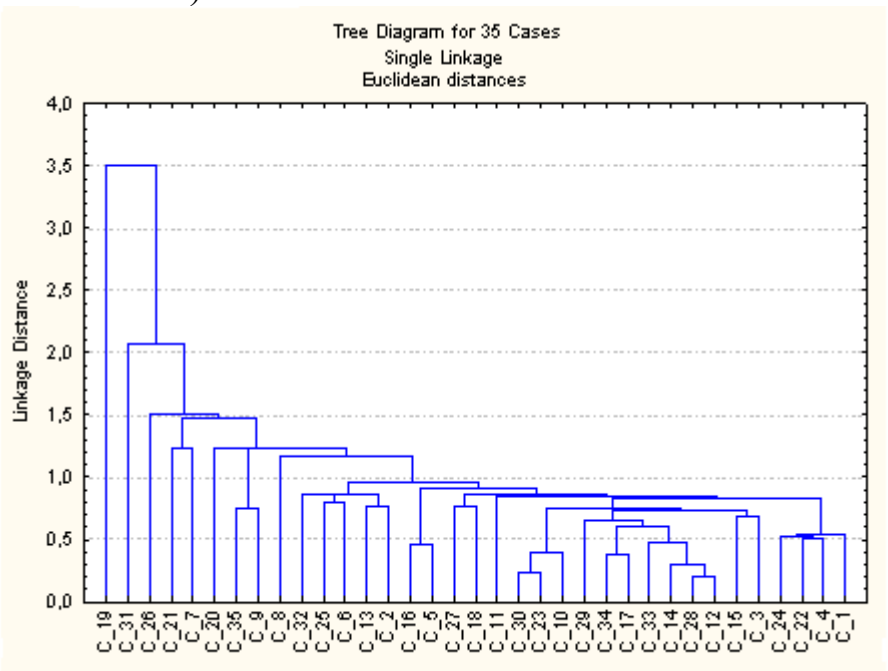
7. При нажатии клавиши 

появляется горизонтальная древовидная диаграмма (случай объединения по объектам):



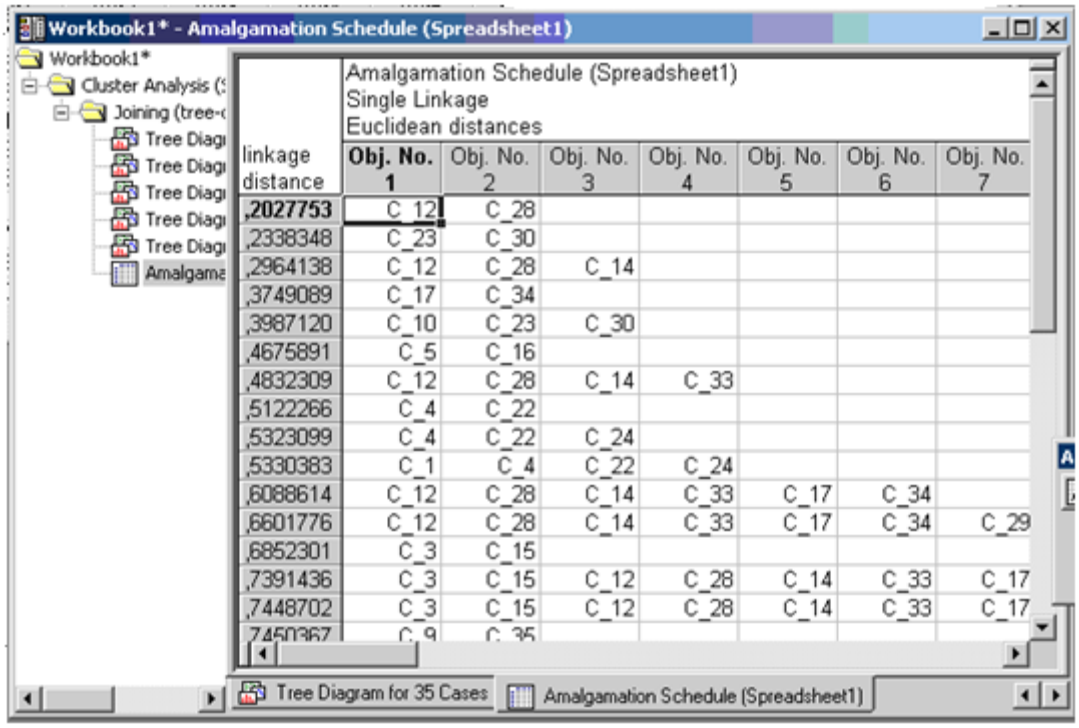
8. При нажатии клавиши 

появляется горизонтальная древовидная диаграмма (случай объединения по объектам):



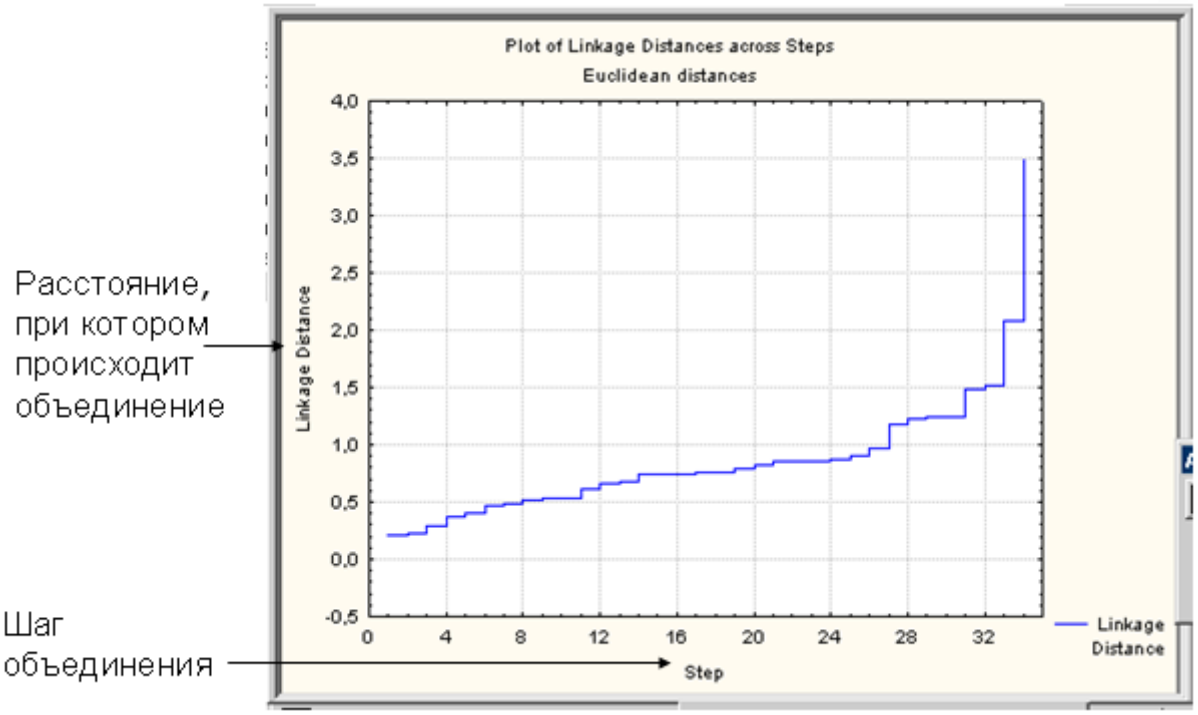
9. При нажатии клавиши 

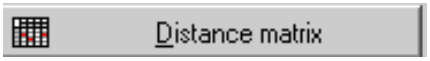
откроется окно, содержащее протокол объединения кластеров:



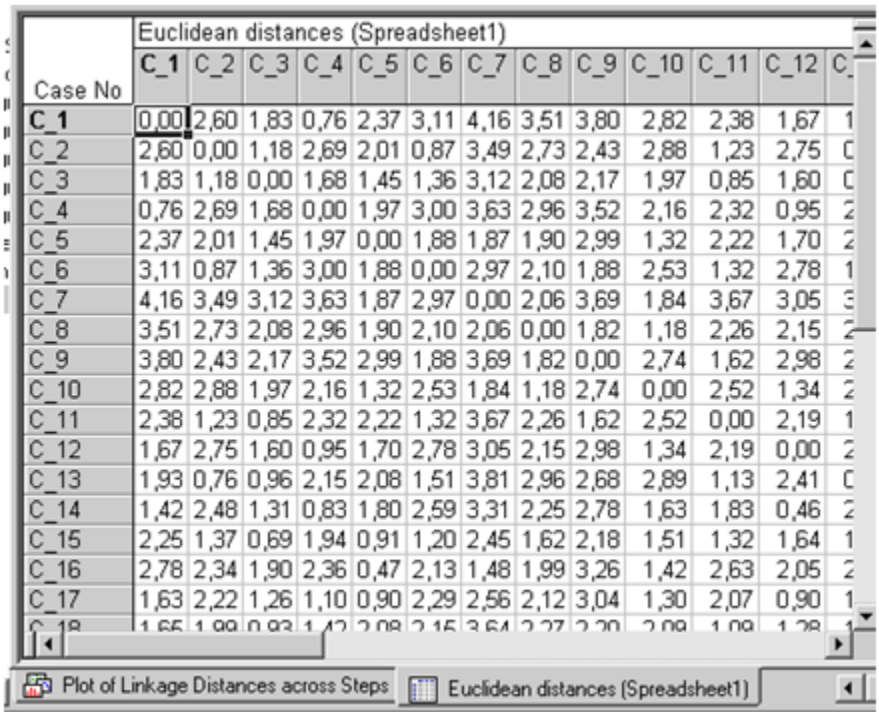
10.При нажатии клавиши 

выводится график изменений расстояний при объединении кластеров:



11. При нажатии клавиши 

выводится матрица расстояний:



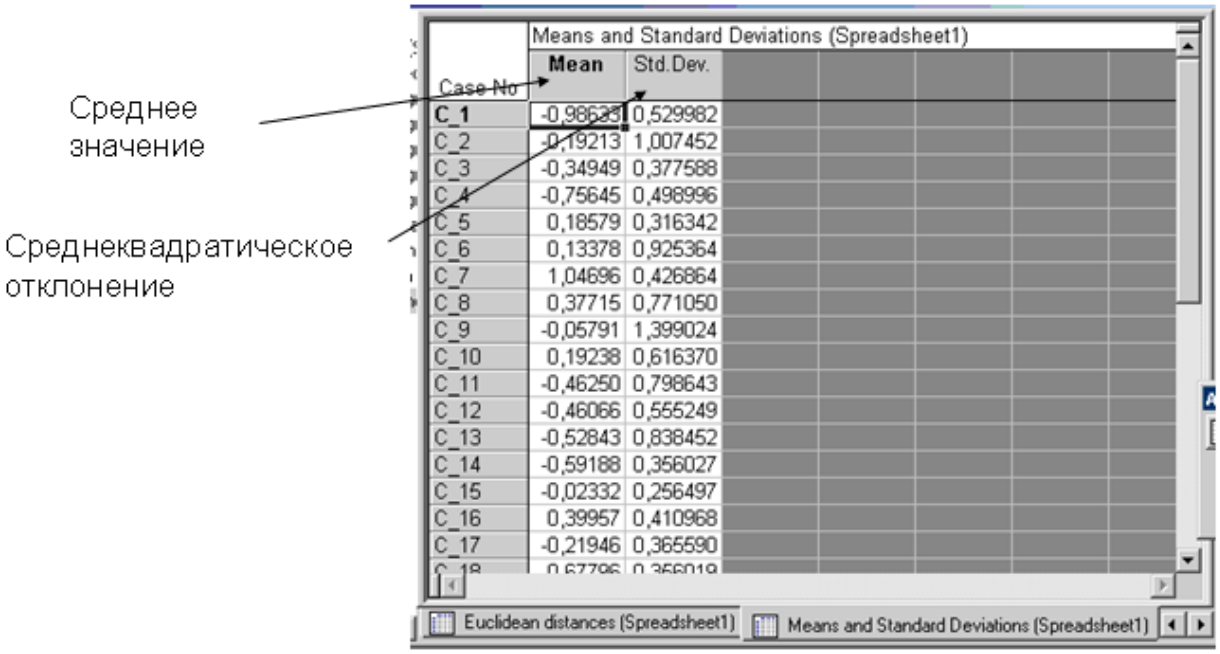
Case No	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_10	C_11	C_12	C_
C_1	0,00	2,60	1,83	0,76	2,37	3,11	4,16	3,51	3,80	2,82	2,38	1,67	1
C_2	2,60	0,00	1,18	2,69	2,01	0,87	3,49	2,73	2,43	2,88	1,23	2,75	0
C_3	1,83	1,18	0,00	1,68	1,45	1,36	3,12	2,08	2,17	1,97	0,85	1,60	0
C_4	0,76	2,69	1,68	0,00	1,97	3,00	3,63	2,96	3,52	2,16	2,32	0,95	2
C_5	2,37	2,01	1,45	1,97	0,00	1,88	1,87	1,90	2,99	1,32	2,22	1,70	2
C_6	3,11	0,87	1,36	3,00	1,88	0,00	2,97	2,10	1,88	2,53	1,32	2,78	1
C_7	4,16	3,49	3,12	3,63	1,87	2,97	0,00	2,06	3,69	1,84	3,67	3,05	3
C_8	3,51	2,73	2,08	2,96	1,90	2,10	2,06	0,00	1,82	1,18	2,26	2,15	2
C_9	3,80	2,43	2,17	3,52	2,99	1,88	3,69	1,82	0,00	2,74	1,62	2,98	2
C_10	2,82	2,88	1,97	2,16	1,32	2,53	1,84	1,18	2,74	0,00	2,52	1,34	2
C_11	2,38	1,23	0,85	2,32	2,22	1,32	3,67	2,26	1,62	2,52	0,00	2,19	1
C_12	1,67	2,75	1,60	0,95	1,70	2,78	3,05	2,15	2,98	1,34	2,19	0,00	2
C_13	1,93	0,76	0,96	2,15	2,08	1,51	3,81	2,96	2,68	2,89	1,13	2,41	0
C_14	1,42	2,48	1,31	0,83	1,80	2,59	3,31	2,25	2,78	1,63	1,83	0,46	2
C_15	2,25	1,37	0,69	1,94	0,91	1,20	2,45	1,62	2,18	1,51	1,32	1,64	1
C_16	2,78	2,34	1,90	2,36	0,47	2,13	1,48	1,99	3,26	1,42	2,63	2,05	2
C_17	1,63	2,22	1,26	1,10	0,90	2,29	2,56	2,12	3,04	1,30	2,07	0,90	1
C_18	1,65	1,99	0,93	1,42	2,08	2,15	3,64	2,27	2,20	2,09	1,09	1,28	1

Euclidean distances (Spreadsheet1)

Plot of Linkage Distances across Steps | Euclidean distances (Spreadsheet1)

12. При нажатии клавиши 

появляются описательные статистики:

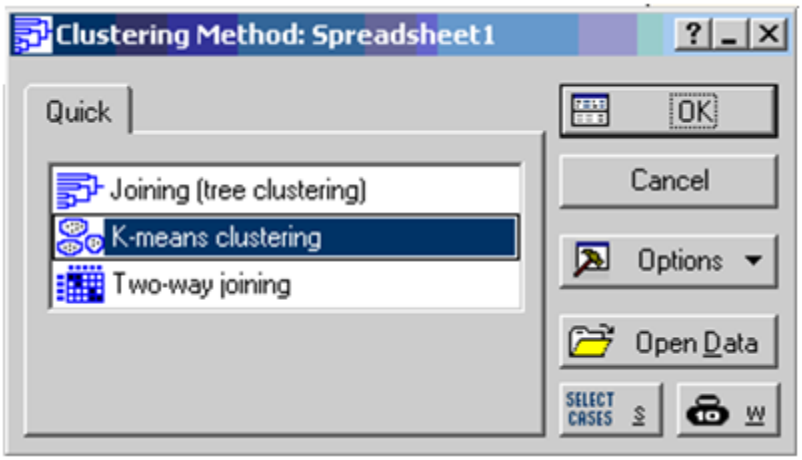


Means and Standard Deviations (Spreadsheet1)

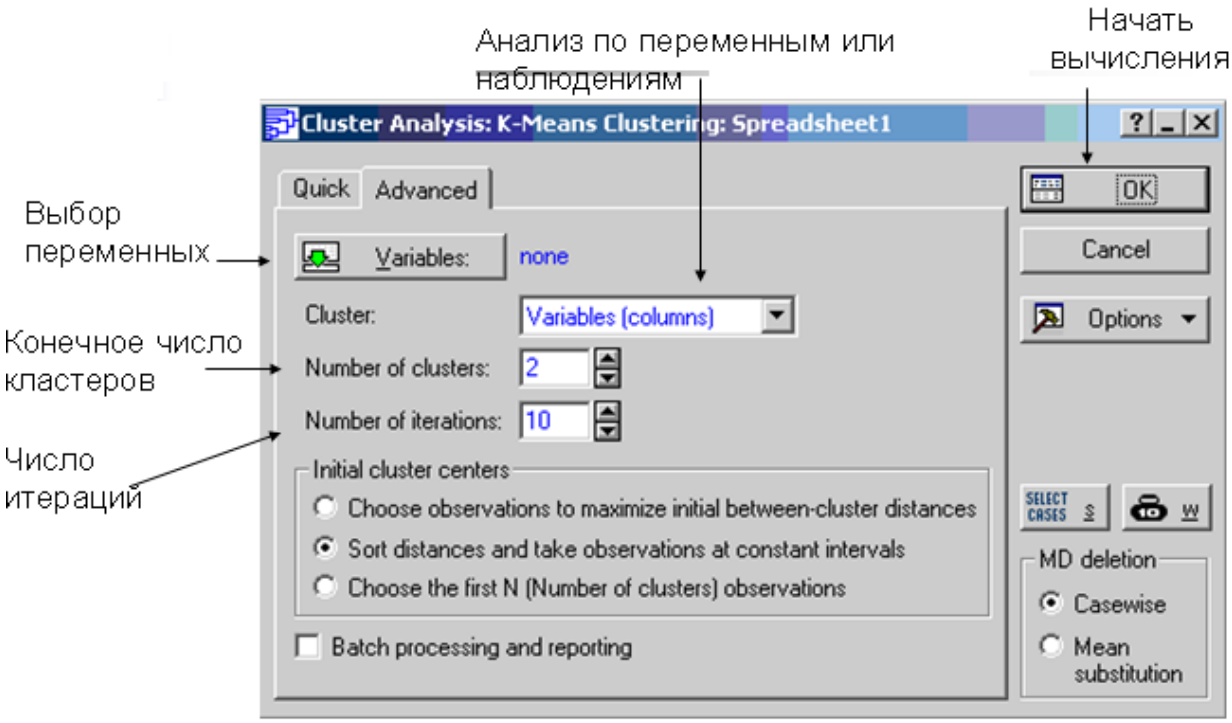
Case No	Mean	Std.Dev.
C_1	-0,98625	0,529982
C_2	-0,19213	1,007452
C_3	-0,34949	0,377588
C_4	-0,75645	0,498996
C_5	0,18579	0,316342
C_6	0,13378	0,925364
C_7	1,04696	0,426864
C_8	0,37715	0,771050
C_9	-0,05791	1,399024
C_10	0,19238	0,616370
C_11	-0,46250	0,798643
C_12	-0,46066	0,555249
C_13	-0,52843	0,838452
C_14	-0,59188	0,356027
C_15	-0,02332	0,256497
C_16	0,39957	0,410968
C_17	-0,21946	0,365590
C_18	0,67796	0,366019

Euclidean distances (Spreadsheet1) | Means and Standard Deviations (Spreadsheet1)

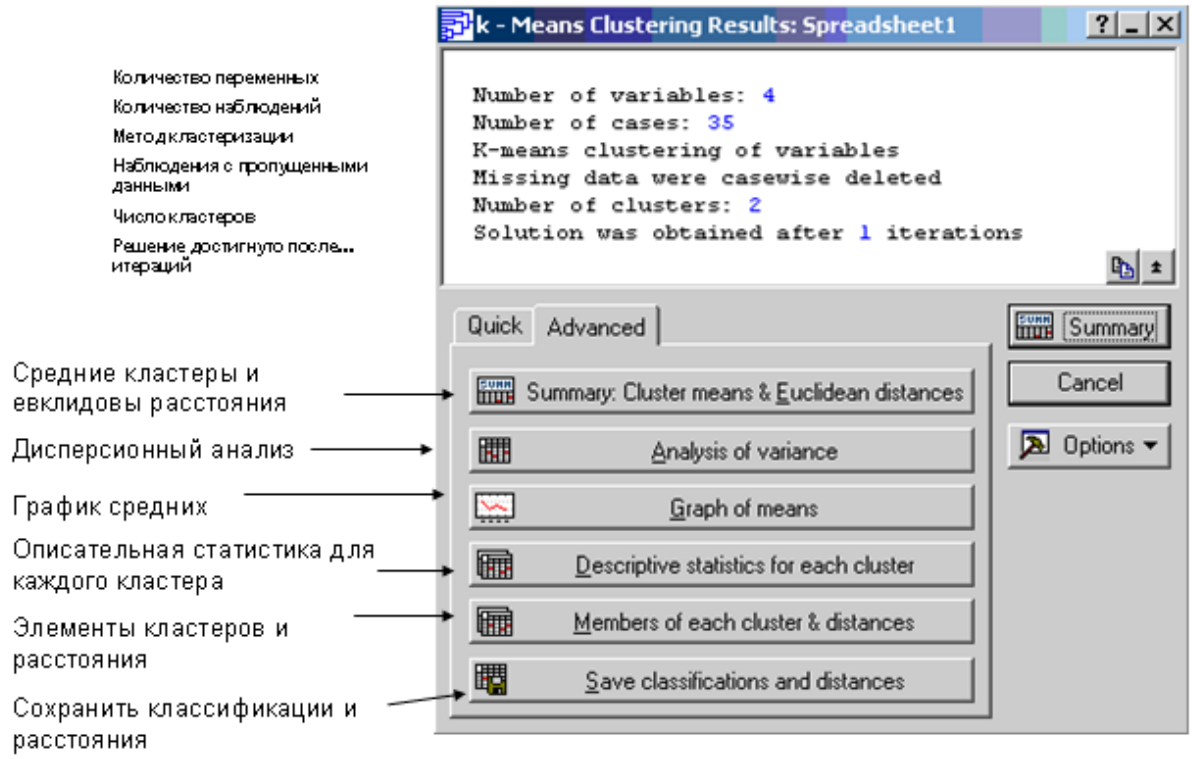
13. Метод к- средних. Панель модуля кластерный анализ: кластеризация методом к - средних:

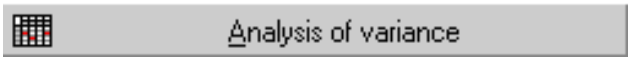


14.Панель модуля кластерного анализа: методом к – средних:



15. Панель результатов метода к – средних:

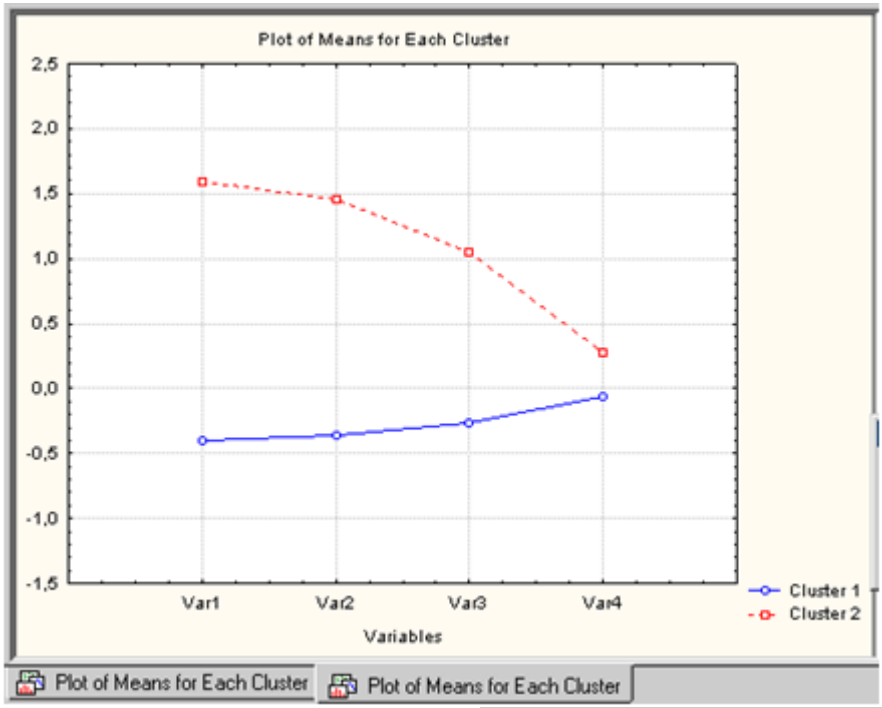


16. Нажатие клавиши 

вызывает появление результатов дисперсионного анализа:

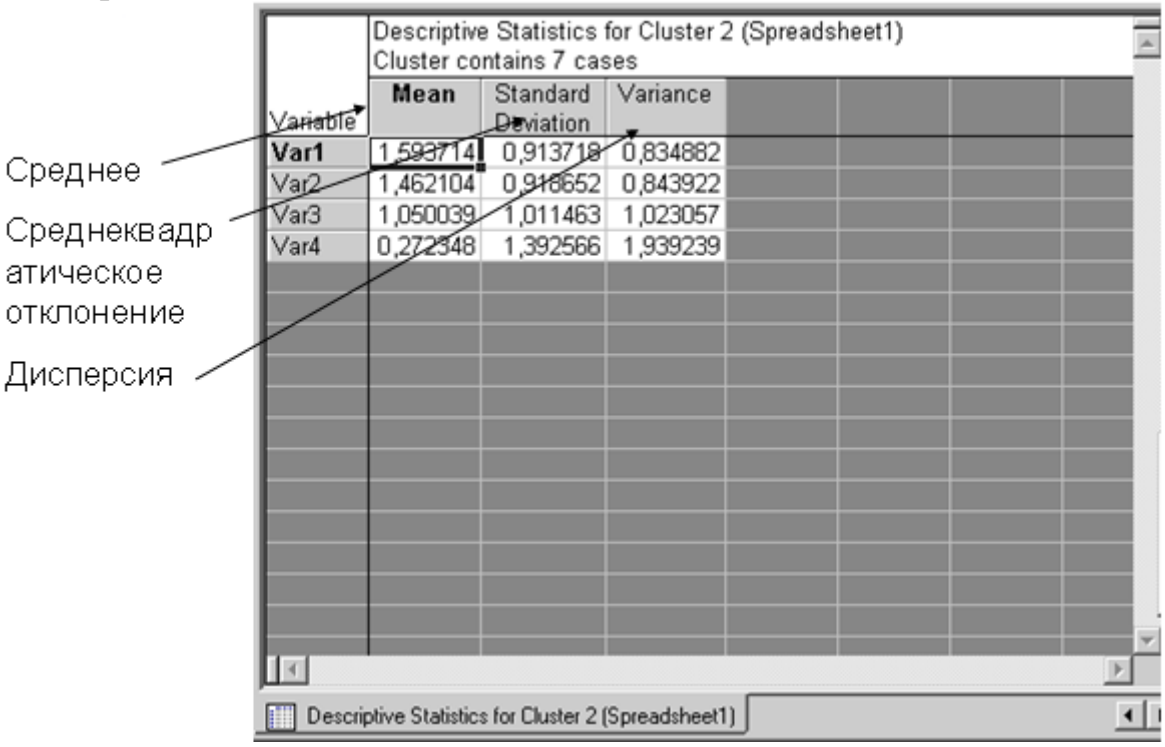
17. Нажатие клавиши 

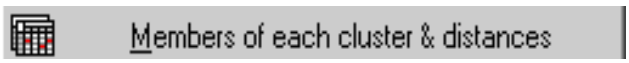
вызывает появление графика средних значений для каждого кластера по анализируемым переменным:



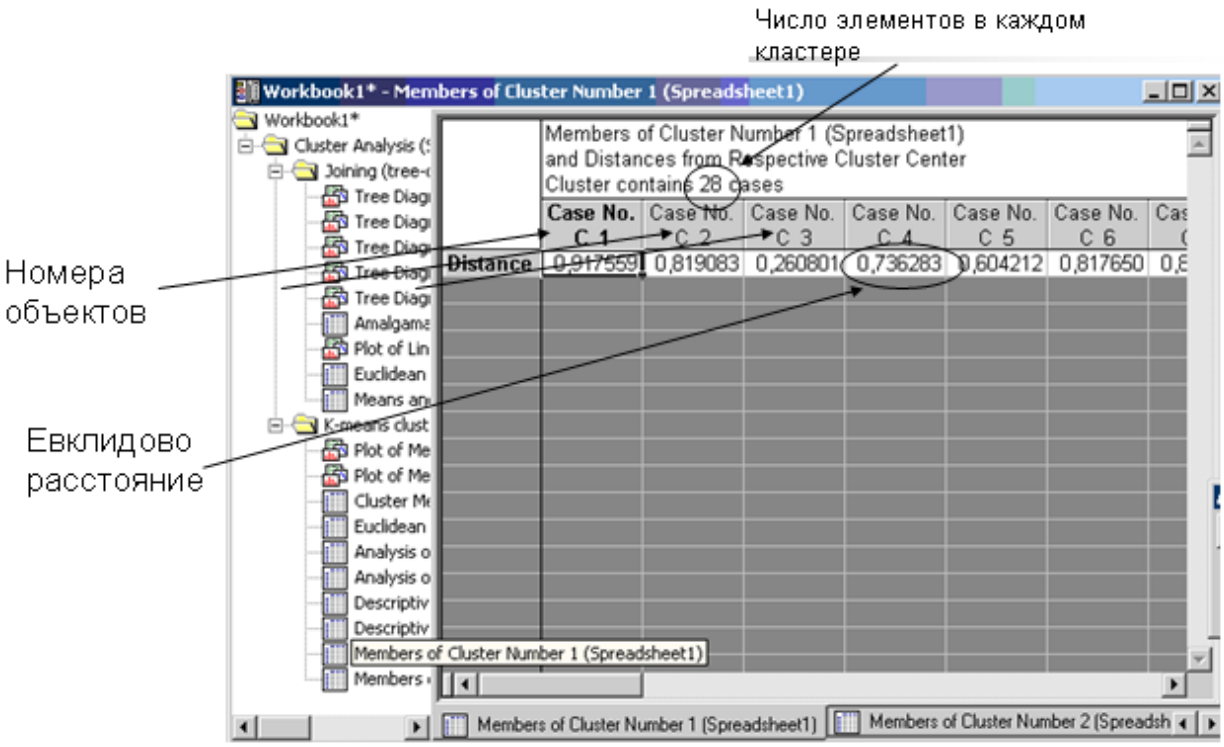
18. При нажатии клавиши 

появляются окна, количество которых, соответствует количеству кластеров. Вид окна:



19. При нажатии клавиши 

появляются окна, количество которых, соответствует количеству кластеров. Вид окна:



Пример 8.2. Кластер-процедура методом k-средних

На основе данных о развитии потребительского рынка Оренбургской области за 2005 г. и 2009 г. (приложения М, Н) осуществите кластерный анализ **методом k-средних**. Дайте характеристику результатам.

Ход процесса:

1. Проведем стандартизацию данных в ППП «STATISTICA», версия 6.0 (англоязычная) (рис.8.1).

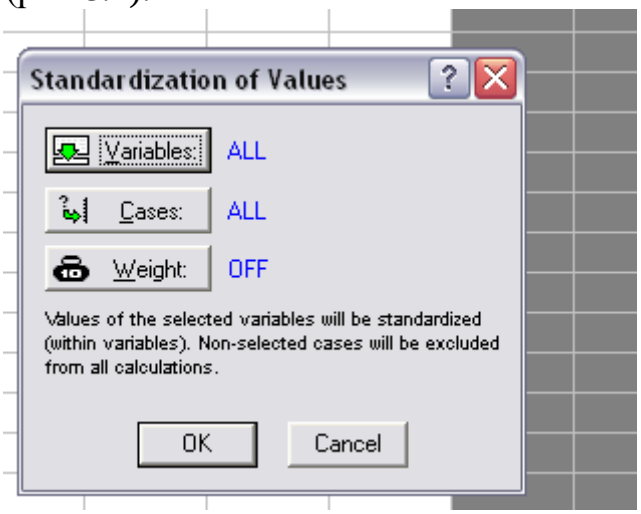


Рисунок 8.1 – Окно «Стандартизация данных»

2. Кластерный анализ проведем с помощью метода k-средних. Этот метод кластеризации существенно отличается от иерархических агломеративных методов. Он применяется, если пользователь уже имеет представление относительно числа кластеров, на которые необходимо разбить

наблюдения. Тогда метод k – средних строит ровно k различных кластеров, расположенных на возможно больших расстояниях друг от друга (рис.8.2).

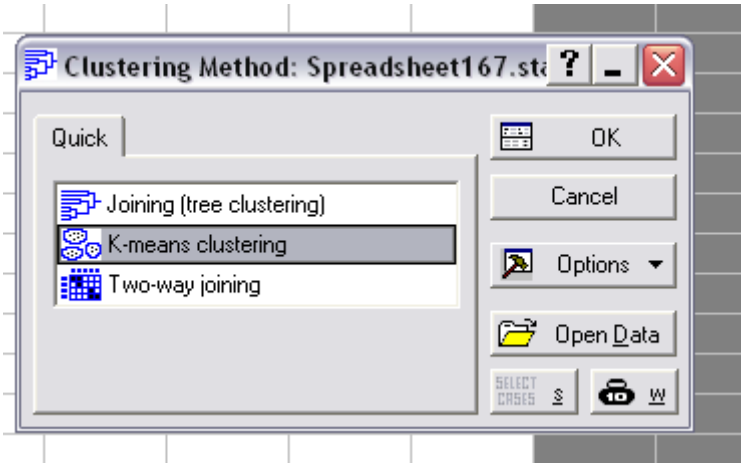


Рисунок 8.2 – Выбор метода кластерного анализа

3. С помощью кнопки Variables (Переменные) выберем показатели, по которым будет происходить кластеризация. В строке Cluster (Кластер) укажем объекты для классификации Cases [rows] (Наблюдения [строки]). Поле Number of clusters (Число кластеров) позволяет ввести желаемое число кластеров, которое должно быть больше 1 и меньше чем количество объектов. Выберем 3 (рис.8.3, 8.4).

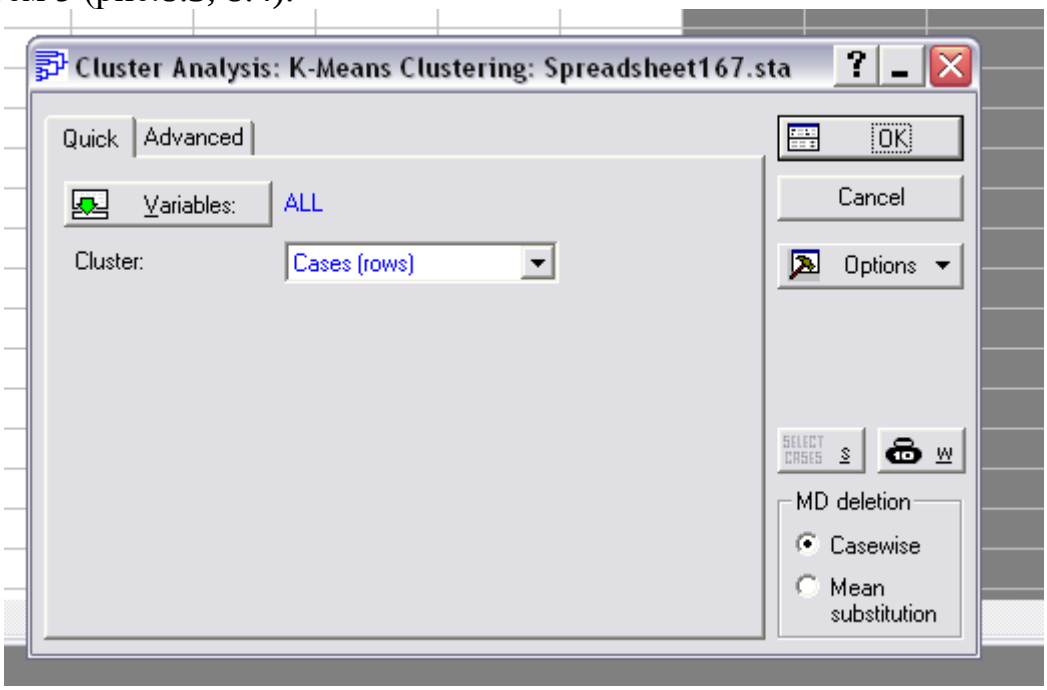


Рисунок 8.3 – Выбор параметров кластерного анализа

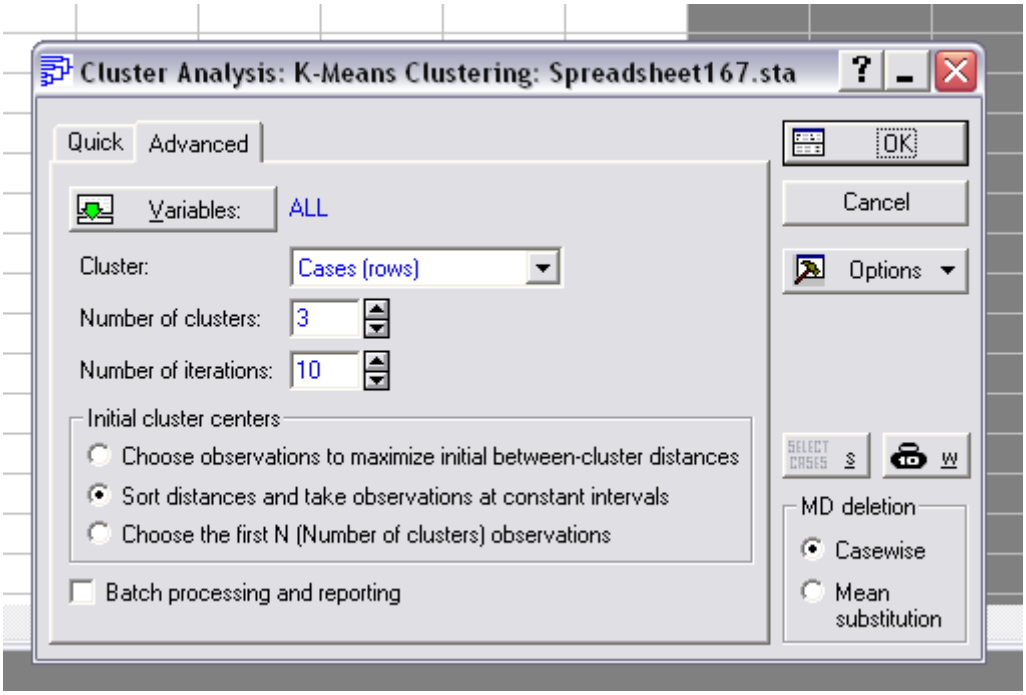


Рисунок 8.4 – Выбор параметров кластерного анализа

4. После соответствующего выбора нажмем кнопку OK. STATISTICA произведет вычисления и появится новое окно: "K –Means Clustering Results" (рис. 8.5).

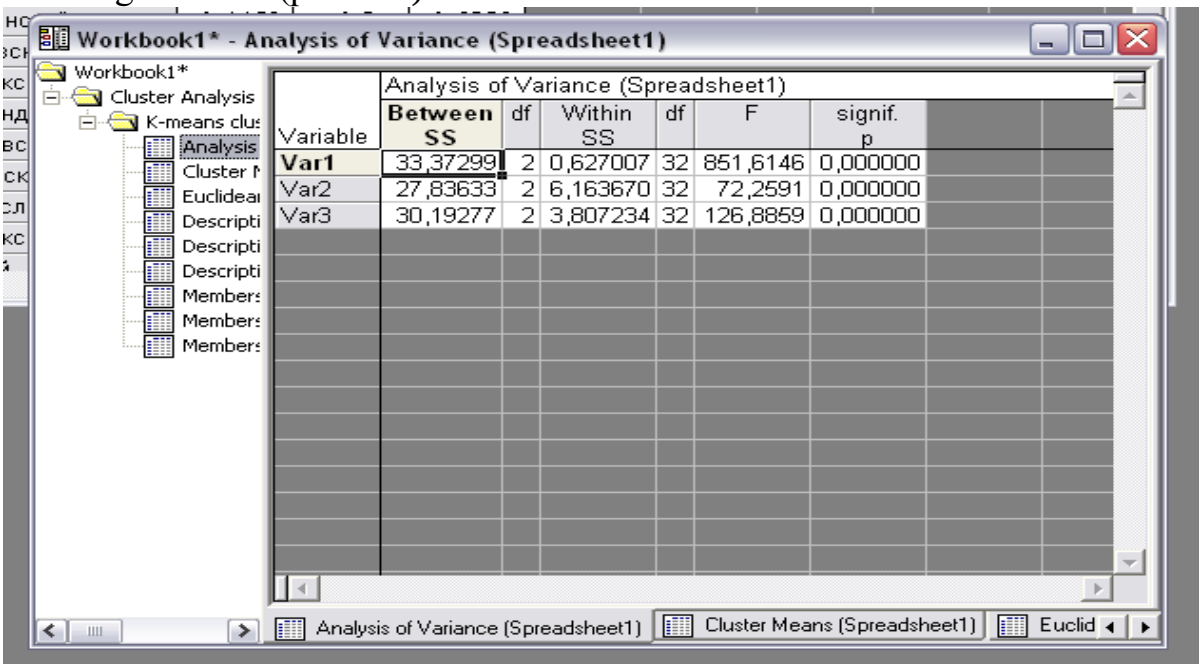


Рисунок 8.5 – Результат кластерного анализа

5. Проанализируем полученные данные.

Analysis of Variance (Дисперсионный анализ). Строки - переменные (наблюдения), столбцы - показатели для каждой переменной: дисперсия между кластерами, число степеней свободы для межклассовой дисперсии, дисперсия внутри кластеров, число степеней свободы для внутриклассовой

дисперсии, F - критерий, для проверки гипотезы о неравенстве дисперсий (рис.8.6).

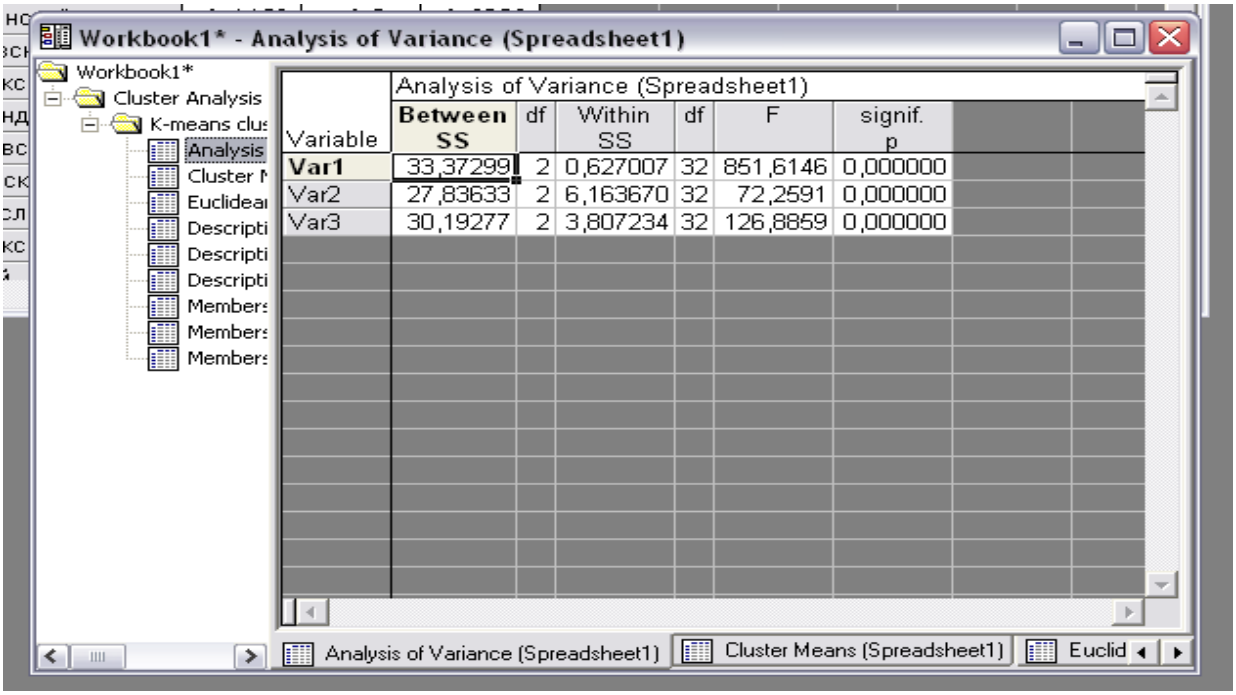


Рисунок 8.6 – Analysis of Variance (Дисперсионный анализ)

Cluster Means & Euclidean Distances (средние значения в кластерах и евклидово расстояние). Выводятся две таблицы. В первой (рис. 8.7) указаны средние величины 36 класса по всем переменным (наблюдениям). По вертикали указаны номера классов, а по горизонтали переменные (наблюдения).

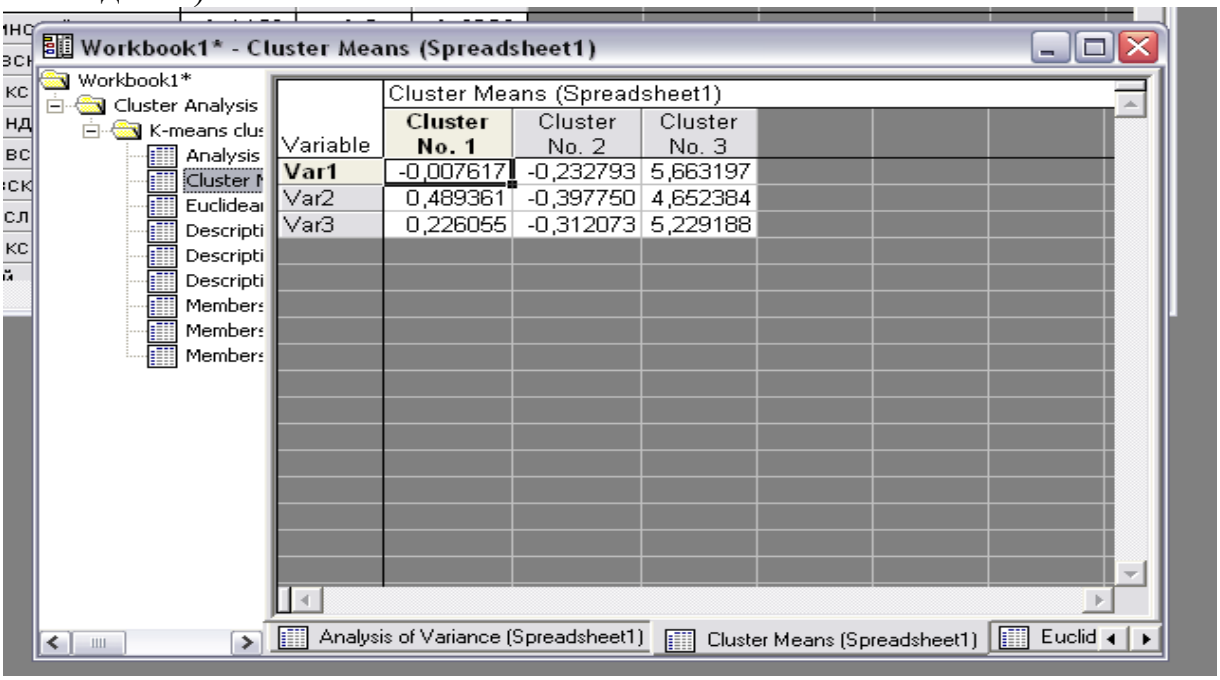


Рисунок 8.7 –Cluster Means (средние значения в кластерах)

Во второй таблице (рис. 8.8) приведены расстояния между классами. И по вертикали и по горизонтали указаны номера кластеров. Таким образом, при

пересечении строк и столбцов указаны расстояния между соответствующими классами. Причем выше диагонали (на которой стоят нули) указаны квадраты, а ниже просто евклидово расстояние.

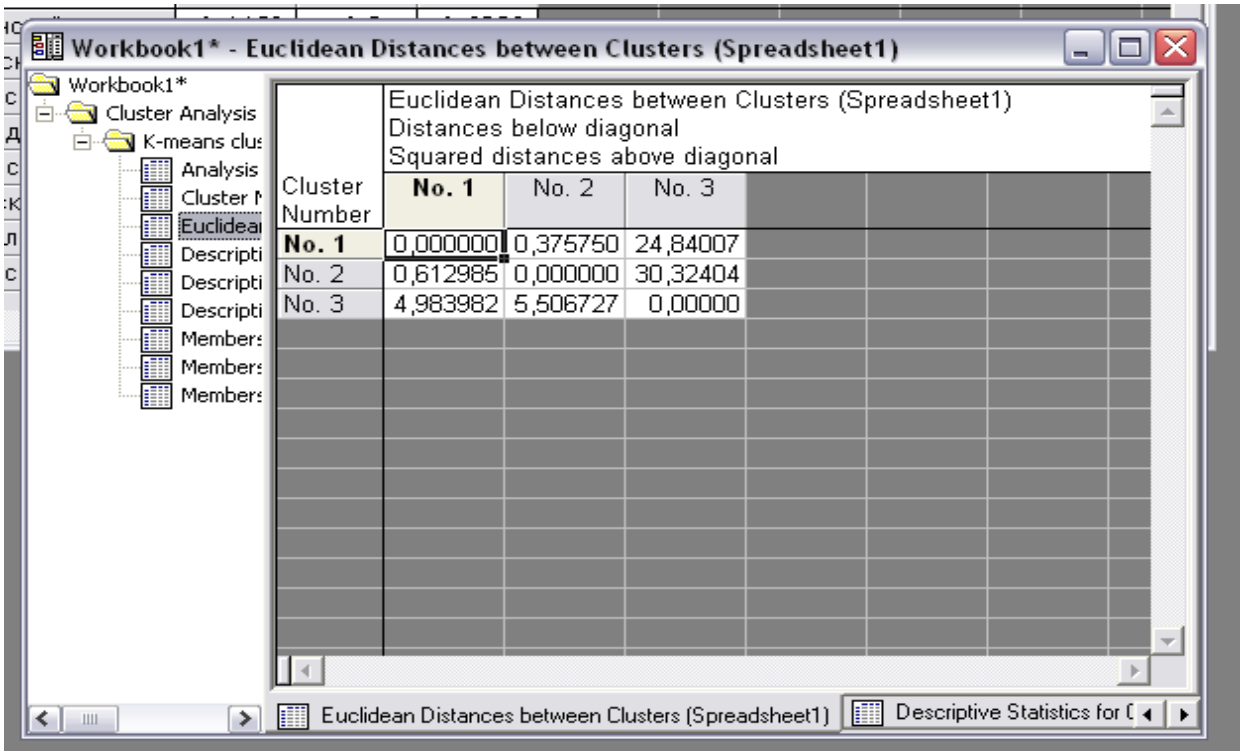


Рисунок 8.8 –Euclidean Distances (Евклидово расстояние)

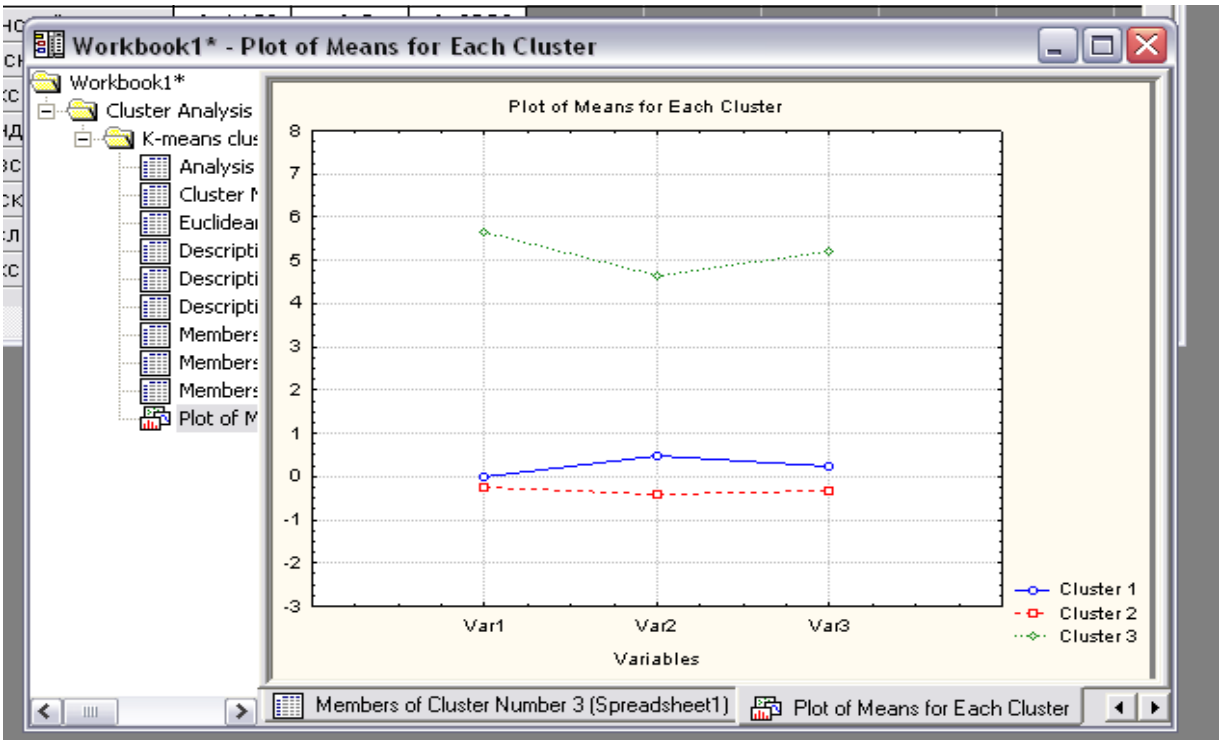


Рисунок 8.9 – Graph of means (График средних)

Щелкнув по кнопке Graph of means (График средних), можно получить графическое изображение информации, содержащейся в таблице, выводимой при нажатии на кнопку Analysis of Variance (Дисперсионный анализ). На графике показаны средние значения переменных для каждого кластера (рис. 8.9).

По горизонтали отложены участвующие в классификации переменные, а по вертикали - средние значения переменных в разрезе получаемых кластеров.

Descriptive Statistics for each cluster (Описательная статистика для каждого кластера). После нажатия этой кнопки выводятся окна, количество которых равно количеству кластеров (рис.8.10). В каждом таком окне в строках указаны переменные (наблюдения), а по горизонтали их характеристики, рассчитанные для данного класса: среднее, несмещенное среднеквадратическое отклонение, несмещенная дисперсия

Descriptive Statistics for Cluster 1 (Spreadsheet1)
Cluster contains 10 cases

Variable	Mean	Standard Deviation	Variance
Var1	-0,007617	0,204652	0,041883
Var2	0,489361	0,655057	0,429099
Var3	0,226055	0,554978	0,308001

Descriptive Statistics for Cluster 2 (Spreadsheet1)
Cluster contains 24 cases

Variable	Mean	Standard Deviation	Variance
Var1	-0,232793	0,104270	0,010872
Var2	-0,397750	0,316350	0,100077
Var3	-0,312073	0,212155	0,045010

Descriptive Statistics for Cluster 3 (Spreadsheet1)
Cluster contains 1 cases

Variable	Mean	Standard Deviation	Variance
Var1	5,663197	0,00	0,00
Var2	4,652384	0,00	0,00
Var3	5,229188	0,00	0,00

Рисунок 8.10 – Descriptive Statistics for each cluster (Описательная статистика для каждого кластера)

Members for each cluster & distances (Члены каждой группы и расстояния). Выводится столько окон, сколько задано классов (рис. 8.11). В каждом окне указывается общее число элементов, отнесенных к этому кластеру, в верхней строке указан номер наблюдения (переменной), отнесенной к данному классу и евклидово расстояние от центра класса до этого наблюдения (переменной). Центр класса - средние величины по всем переменным (наблюдениям) для этого класса.

Save classifications and distances. Позволяет сохранить в формате программы статистика таблицу, в которой содержатся значения всех переменных, их порядковые номера, номера кластеров к которым они

отнесены, и евклидовы расстояния от центра кластера до наблюдения. Записанная таблица может быть вызвана любым блоком или подвергнута дальнейшей обработке (рис.8.12).

Members of Cluster Number 1 (Spreadsheet1)
and Distances from Respective Cluster Center
Cluster contains 10 cases

	Адамовский	Новоорский	Новосергиевский	Переволоцкий	Сакмарский	Саракташский	Северный	Соль-Илецкий	Ташлинский	Тоцкий
Distance	0,289695	0,403673	1,094842	0,429919	0,244226	0,203316	0,398475	0,437761	0,315674	0,394350

Members of Cluster Number 2 (Spreadsheet1)
and Distances from Respective Cluster Center
Cluster contains 24 cases

	Абдулинский	Акбулакский	Александровский	Асекеевский	Беляевский	Бугурусланский	Бузулукский	Гайский	Грачевский	Домбаровский	Илекский	Кваркенский
Distance	0,314613	0,068096	0,040384	0,094111	0,063227	0,166431	0,179939	0,347546	0,092974	0,168886	0,218176	0,153685

Members of Cluster Number 3 (Spreadsheet1)
and Distances from Respective Cluster Center
Cluster contains 1 cases

	Оренбургский
Distance	0,00

Рисунок 8.11 - Members for each cluster & distances (Члены каждой группы и расстояния)

	Spreadsheet1					
	1 Var1	2 Var2	3 Var3	4 CASE_NO	5 CLUSTER	6 DISTANCE
Абдулинский	-0,446	-0,778	-0,639	1	2	0,31
Адамовский	-0,107	0,165	-0,143	2	1	0,29
Акбулакский	-0,119	-0,369	-0,298	3	2	0,07
Александровский	-0,270	-0,457	-0,313	4	2	0,04
Асекеевский	-0,136	-0,352	-0,189	5	2	0,09
Беляевский	-0,248	-0,403	-0,204	6	2	0,06
Бугурусланский	-0,238	-0,119	-0,384	7	2	0,17
Бузулукский	-0,078	-0,272	-0,551	8	2	0,18
Гайский	-0,357	-0,926	-0,572	9	2	0,35
Грачевский	-0,213	-0,557	-0,308	10	2	0,09
Домбаровский	-0,228	-0,681	-0,240	11	2	0,17
Илекский	-0,077	-0,106	-0,129	12	2	0,22
Кваркенский	-0,244	-0,561	-0,522	13	2	0,15
Красногвардейский	-0,207	-0,216	-0,056	14	2	0,18
Кувандыкский	-0,323	-0,599	-0,498	15	2	0,17
Курманаевский	-0,241	-0,355	0,048	16	2	0,21
Матвеевский	-0,318	-0,321	-0,346	17	2	0,07
Новоорский	0,275	-0,010	0,626	18	1	0,40
Новосергиевский	0,120	2,196	1,043	19	1	1,09
Октябрьский	-0,142	0,029	-0,268	20	2	0,25
Оренбургский	5,663	4,652	5,229	21	3	0,00
Первомайский	-0,135	-0,122	0,120	22	2	0,30
Переволоцкий	0,037	0,125	0,874	23	1	0,43
Пономаревский	-0,276	-0,107	0,011	24	2	0,25
Сакмарский	-0,079	0,116	0,041	25	1	0,24

Рисунок 8.12 - Save classifications and distances

Данные о кластерах и районах, вошедших в кластеры, представлены в таблице 8.1.

Таблица 8.1 – Результаты кластер-процедуры для данных 2005 г.

№ кластера	Количество районов	Районы
1	10	Адамовский, Новоорский, Новосергиевский, Переволоцкий, Сакмарский, Саракташский, Северный, Соль-Илецкий, Ташлинский, Тоцкий
2	24	Акбулакский, Октябрьский, Абдулинский, Александровский, Асекеевский, Беляевский, Бугурусланский, Бузулукский, Гайский, Грачевский, Домбаровский, Илекский, Кваркенский, Красногвардейский, Кувандыкский, Курманаевский, Матвеевский, Первомайский, Пономаревский, Светлинский, Сорочинский, Тюльганский, Шарлыкский, Ясненский
3	1	Оренбургский

Информация о средних величинах по кластерам представлена в таблице 8.2.

Так, в 2005 г. в первый кластер вошли 10 районов (29% от общего числа).

Во второй кластер – самый многочисленный – вошли 24 района (69% от общего числа).

Таблица 8.2 – Средние величины по кластерам в 2005 г.

Показатель	Среднее значение в 1 кластере	Среднее значение во 2 кластере	Среднее значение в 3 кластере
Объем платных услуг населению, тыс. руб.	82375,70	42278,25	1092188
Оборот розничной торговли, млн. руб.	278,04	136,15	943,9
Оборот общественного питания, тыс. руб.	18336,40	7001,96	123716

Из рисунков 8.9 и 8.11, данных таблицы 8.2 видно, что *наилучшим по показателям развития потребительского рынка является 3-ий кластер*. В данный кластер (аномальный) вошел один район - Оренбургский. Причиной этого является существенное превосходство данного района по показателям развития потребительского рынка, что обусловлено, прежде всего, высокой численностью населения района и, соответственно, более высокими по сравнению с другими районами значениями исследуемых показателей.

Оренбургский район расположен в центре области и располагает богатыми природными ресурсами и полезными ископаемыми, прежде всего, газоконденсатными и нефтяными месторождениями, для этого района характерна благоприятная экономическая ситуация, сопровождающаяся весьма высоким уровнем состояния и развития потребительского рынка, поскольку он является центром нефтедобывающей промышленности и сельскохозяйственного производства.

Районы, попавшие во второй кластер, занимают выгодное экономико-географическое положение, располагаясь на пересечении транспортных путей железнодорожного и автомобильного сообщения. Но, несмотря на это, средние показатели состояния и развития потребительского рынка данных районов достаточно низкие относительно аналогичных показателей районов первого и третьего кластера (табл.8.2). Это дает основание характеризовать районы, попавшие в данный кластер, как районы *с низким уровнем развития потребительского рынка*.

Районы, вошедшие в 1-ый кластер, имеют средние значения показателей состояния и развития потребительского рынка относительно регионов 2-ого и 3-его кластеров. Следовательно, районы, попавшие в данный кластер, характеризуются *средним уровнем развития потребительского рынка*.

На основе данных о развитии потребительского рынка Оренбургской области за 2009 г. осуществим процедуру кластеризации.

1. Проведем стандартизацию данных (рис.8.13).
2. Кластерный анализ проведем с помощью метода k-средних. Этот метод кластеризации существенно отличается от иерархических агломеративных методов. Он применяется, если пользователь уже имеет представление относительно числа кластеров, на которые необходимо разбить наблюдения. Тогда метод k – средних строит ровно k различных кластеров, расположенных на возможно больших расстояниях друг от друга (рис.8.14).

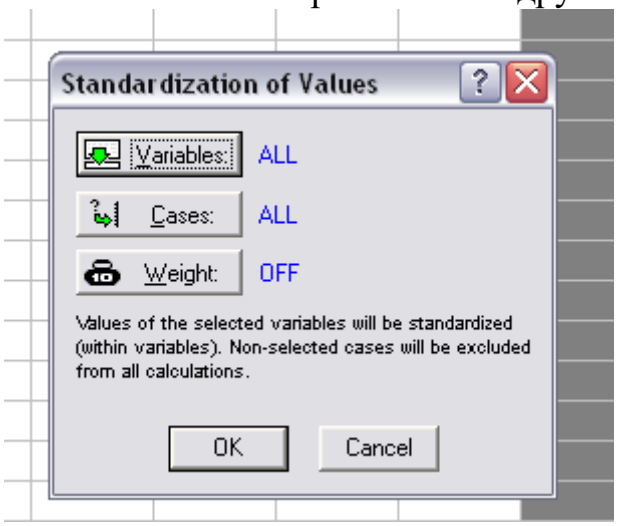


Рисунок 8.13 – Стандартизация данных

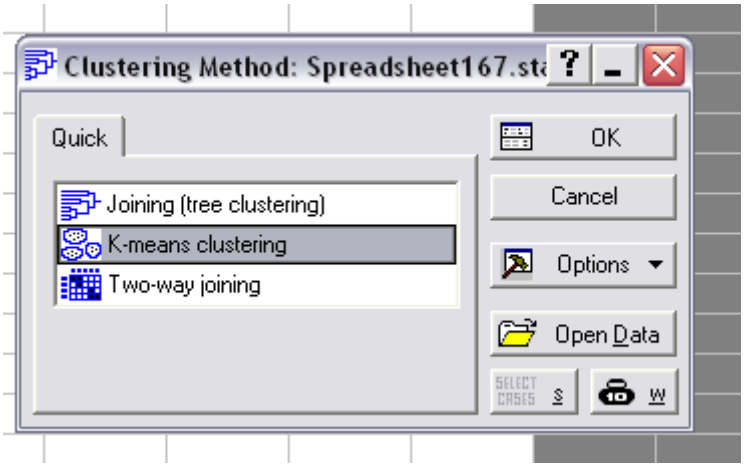


Рисунок 8.14 – Выбор метода кластерного анализа

3. С помощью кнопки Variables (Переменные) выберем показатели, по которым будет происходить кластеризация. В строке Cluster (Кластер) укажем объекты для классификации Cases [rows] (Наблюдения [строки]). Поле Number of clusters (Число кластеров) позволяет ввести желаемое число кластеров, которое должно быть больше 1 и меньше чем количество объектов. Выберем 3 (рис.8.15, 8.16).

4. После соответствующего выбора нажмем кнопку OK. STATISTICA произведет вычисления и появится новое окно: "K –Means Clustering Results" (рис. 8.17).

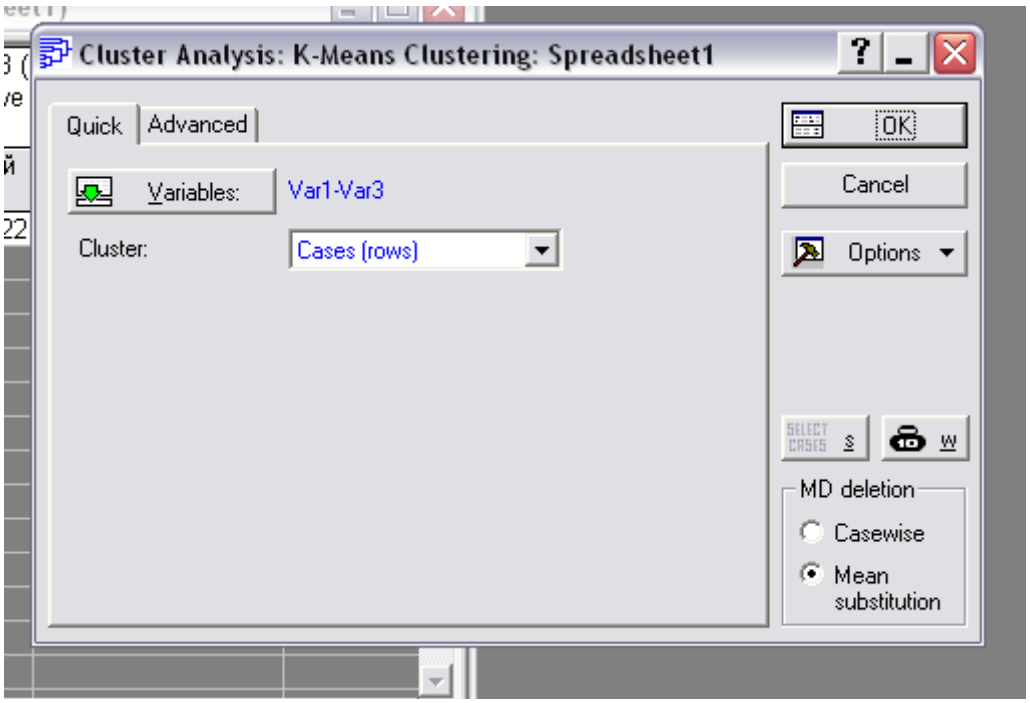


Рисунок 8.15 – Выбор параметров кластерного анализа

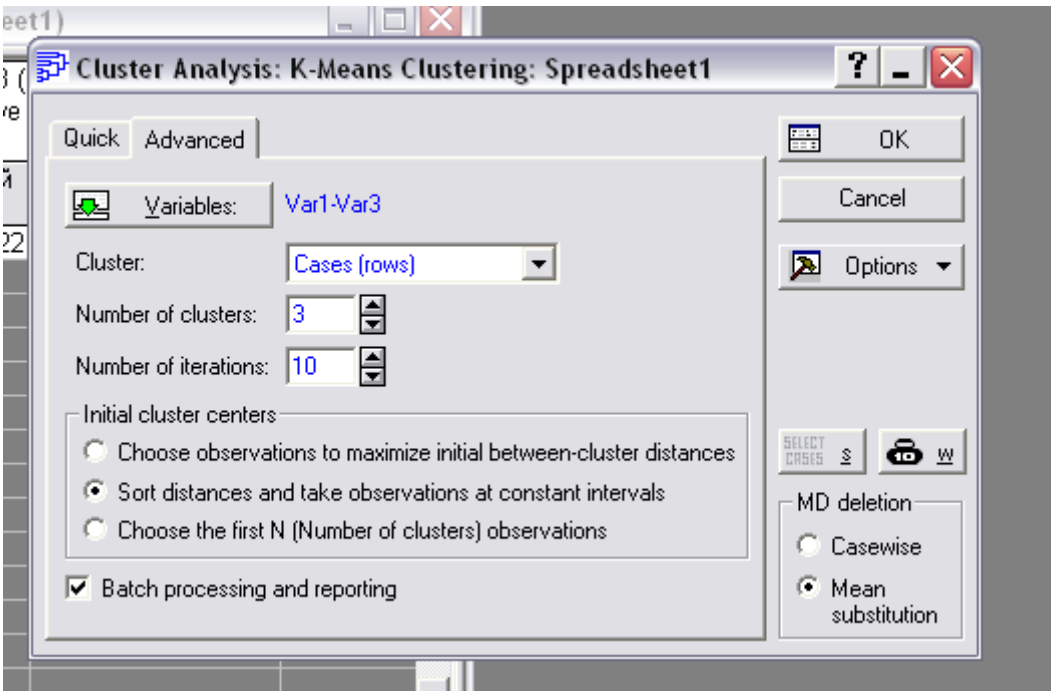


Рисунок 8.16 – Выбор параметров кластерного анализа

5. Проанализируем полученные данные.

Analysis of Variance (Дисперсионный анализ). Строки - переменные (наблюдения), столбцы - показатели для каждой переменной: дисперсия между кластерами, число степеней свободы для межклассовой дисперсии, дисперсия внутри кластеров, число степеней свободы для внутриклассовой дисперсии, F - критерий, для проверки гипотезы о неравенстве дисперсий (рис.8.18).

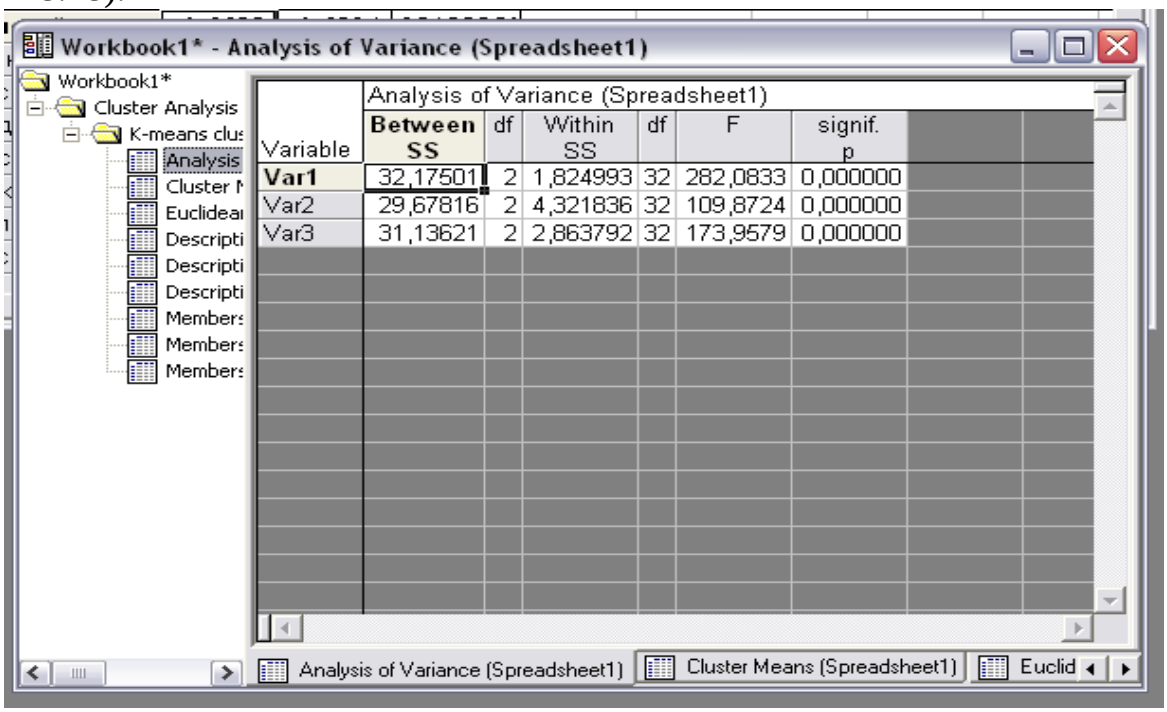


Рисунок 8.17 – Результат кластерного анализа

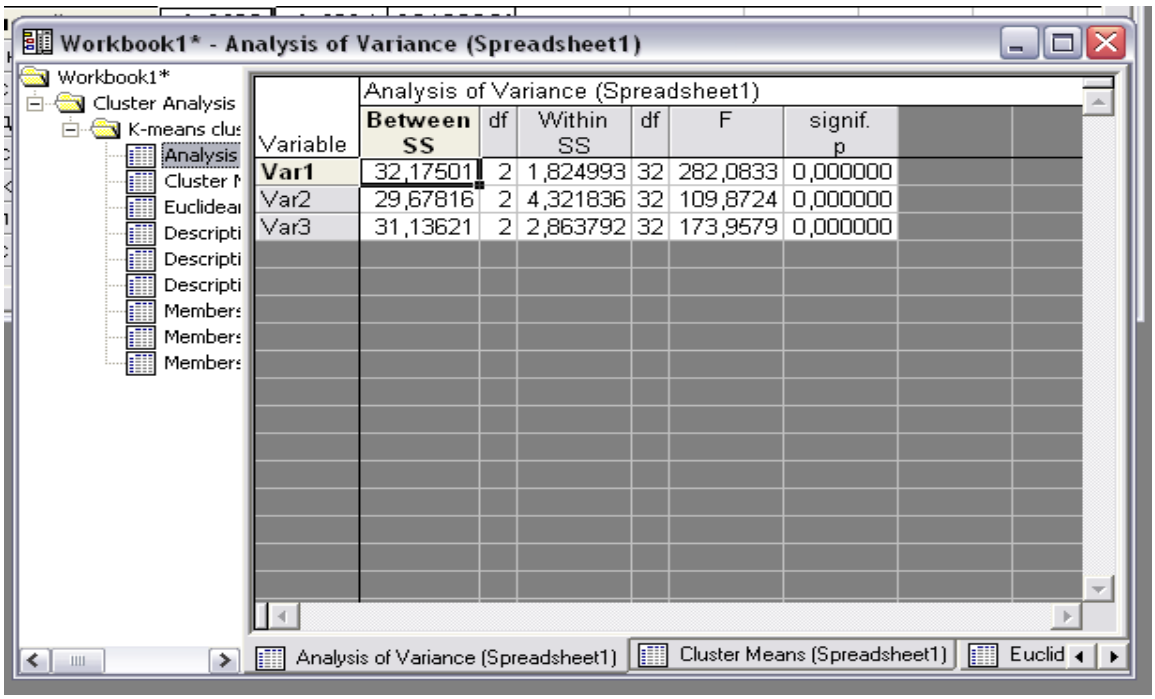


Рисунок 8.18 – Analysis of Variance (Дисперсионный анализ)

Cluster Means & Euclidean Distances (средние значения в кластерах и евклидово расстояние). Выводятся две таблицы. В первой (рис. 8.19) указаны средние величины 36 класса по всем переменным (наблюдениям). По вертикали указаны номера классов, а по горизонтали переменные (наблюдения).

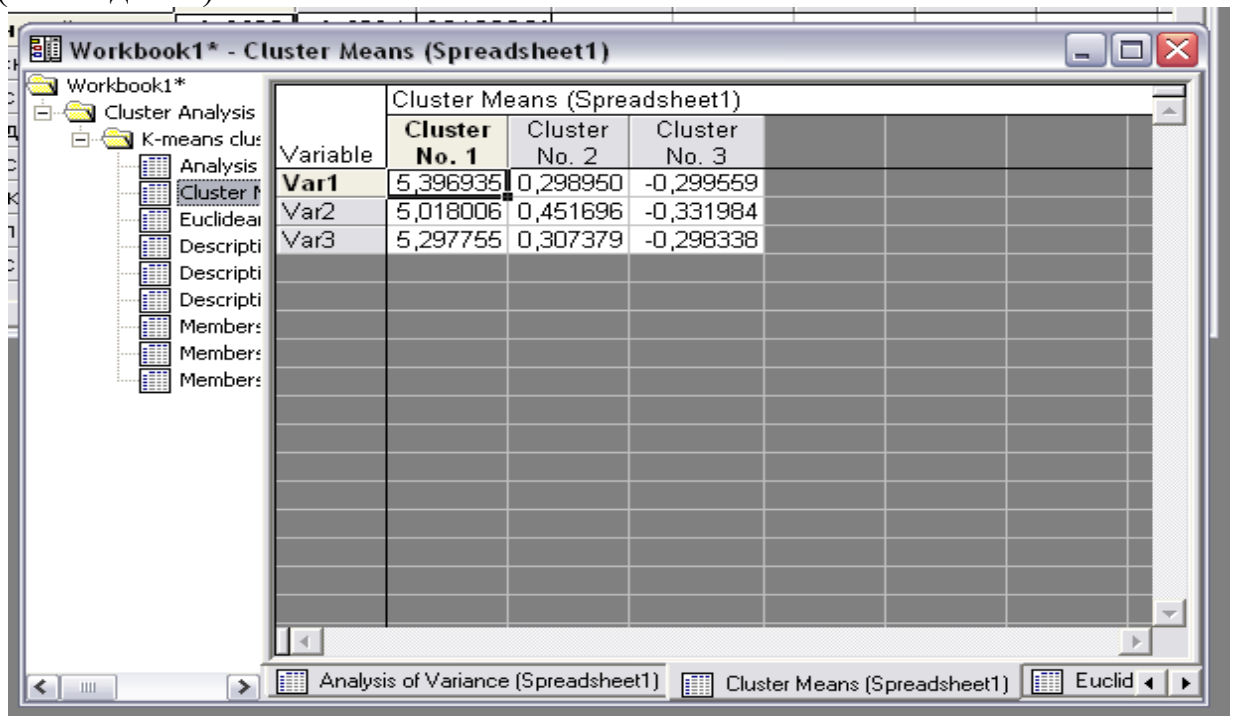


Рисунок 8.19 –Cluster Means (средние значения в кластерах)

Во второй таблице (рис. 8.20) приведены расстояния между классами. И по вертикали, и по горизонтали указаны номера кластеров. Таким образом, при

пересечении строк и столбцов указаны расстояния между соответствующими классами. Причем выше диагонали (на которой стоят нули) указаны квадраты, а ниже просто евклидово расстояние.

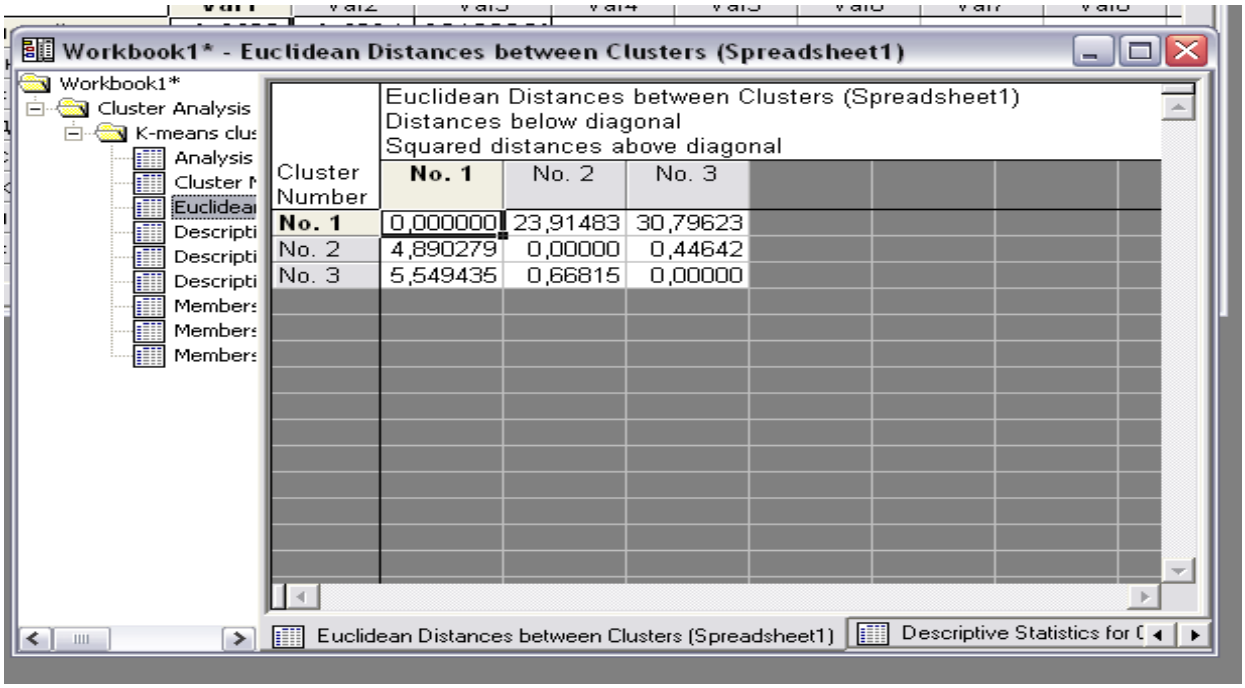


Рисунок 8.20 –Euclidean Distances (Евклидово расстояние)

Щелкнув по кнопке Graph of means (График средних), можно получить графическое изображение информации, содержащейся в таблице, выводимой при нажатии на кнопку Analysis of Variance (Дисперсионный анализ). На графике показаны средние значения переменных для каждого кластера (рис. 8.21).

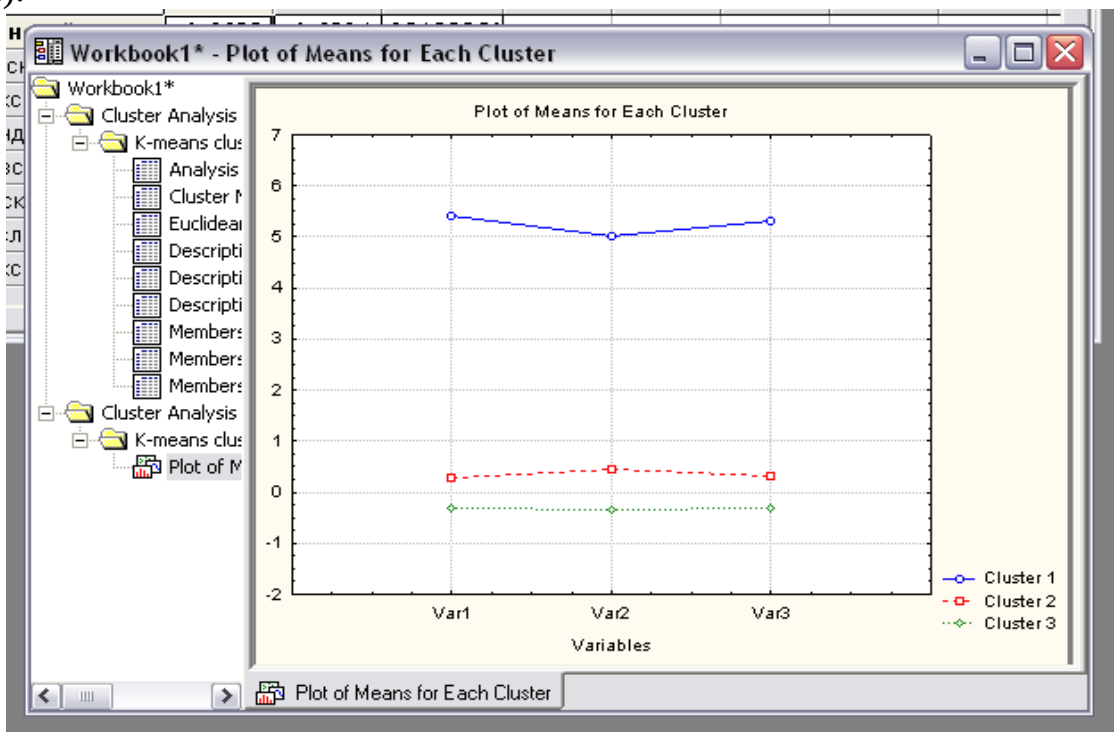


Рисунок 8.21 – Graph of means (График средних)

По горизонтали отложены участвующие в классификации переменные, а по вертикали - средние значения переменных в разрезе получаемых кластеров.

Descriptive Statistics for Cluster 1 (Spreadsheet1)
Cluster contains 1 cases

Variable	Mean	Standard Deviation	Variance
Var1	5,396935	0,00	0,00
Var2	5,018006	0,00	0,00
Var3	5,297755	0,00	0,00

Descriptive Statistics for Cluster 2 (Spreadsheet1)
Cluster contains 8 cases

Variable	Mean	Standard Deviation	Variance
Var1	0,298950	0,358351	0,128416
Var2	0,451696	0,557484	0,310789
Var3	0,307379	0,337362	0,113813

Descriptive Statistics for Cluster 3 (Spreadsheet1)
Cluster contains 26 cases

Variable	Mean	Standard Deviation	Variance
Var1	-0,299559	0,192467	0,037043
Var2	-0,331984	0,293006	0,085853
Var3	-0,298338	0,287548	0,082684

Рисунок 8.22 – Descriptive Statistics for each cluster (Описательная статистика для каждого кластера)

Descriptive Statistics for each cluster (Описательная статистика для каждого кластера). После нажатия этой кнопки выводятся окна, количество которых равно количеству кластеров (рис. 8.22). В каждом таком окне в строках указаны переменные (наблюдения), а по горизонтали их характеристики, рассчитанные для данного класса: среднее, несмещенное среднеквадратическое отклонение, несмещенная дисперсия.

Members for each cluster & distances (Члены каждой группы и расстояния). Выводится столько окон, сколько задано классов (рис. 8.23). В каждом окне указывается общее число элементов, отнесенных к этому кластеру, в верхней строке указан номер наблюдения (переменной), отнесенной к данному классу и евклидово расстояние от центра класса до этого наблюдения (переменной). Центр класса - средние величины по всем переменным (наблюдениям) для этого класса.

Members of Cluster Number 1 (Spreadsheet1)
and Distances from Respective Cluster Center
Cluster contains 1 cases

	Оренбургский
Distance	0,00

Members of Cluster Number 2 (Spreadsheet1)
and Distances from Respective Cluster Center
Cluster contains 8 cases

	Адамовский	Новоорский	Новосергиевский	Первомайский	Переволоцкий	Сакмарский	Саракташский	Тоцкий
Distance	0,330448	0,389673	0,745127	0,334573	0,362726	0,265295	0,368681	0,155889

Members of Cluster Number 3 (Spreadsheet1)
and Distances from Respective Cluster Center
Cluster contains 26 cases

	Абдулинский	Акбулакский	Александровский	Асекеевский	Беляевский	Бугурусланский	Бузулукский	Гайский	Грачевский	Домбаровский	Илекский
Distance	0,301456	0,175122	0,105373	0,102291	0,101139	0,195362	0,154291	0,211885	0,154006	0,130861	0,257708

Рисунок 8.23 - Members for each cluster & distances (Члены каждой группы и расстояния)

Spreadsheet1

	1 Var1	2 Var2	3 Var3	4 CASE_NO	5 CLUSTER	6 DISTANCE
Абдулинский	-0,668	-0,689	-0,204	1	3	0,30
Адамовский	0,072	0,250	-0,178	2	2	0,33
Акбулакский	-0,125	-0,097	-0,217	3	3	0,18
Александровский	-0,294	-0,173	-0,209	4	3	0,11
Асекеевский	-0,152	-0,380	-0,213	5	3	0,10
Беляевский	-0,321	-0,504	-0,325	6	3	0,10
Бугурусланский	-0,385	-0,062	-0,113	7	3	0,20
Бузулукский	-0,057	-0,411	-0,218	8	3	0,15
Гайский	-0,494	-0,457	-0,583	9	3	0,21
Грачевский	-0,245	-0,592	-0,276	10	3	0,15
Домбаровский	-0,324	-0,398	-0,514	11	3	0,13
Илекский	0,005	-0,006	-0,316	12	3	0,26
Кваркенский	-0,282	-0,487	-0,478	13	3	0,14
Красногвардейский	-0,153	-0,183	0,371	14	3	0,40
Кувандыкский	-0,543	-0,832	-0,624	15	3	0,37
Курманаевский	-0,260	-0,480	0,590	16	3	0,52
Матвеевский	-0,484	-0,346	-0,420	17	3	0,13
Новоорский	0,721	0,644	0,798	18	2	0,39
Новосергиевский	0,333	1,701	0,629	19	2	0,75
Октябрьский	-0,159	-0,067	-0,054	20	3	0,22
Оренбургский	5,397	5,018	5,298	21	1	0,00
Первомайский	-0,075	0,042	0,476	22	2	0,33
Переволоцкий	-0,055	-0,043	0,466	23	2	0,36
Пономаревский	-0,387	-0,254	-0,163	24	3	0,10

Рисунок 8.24 – Save classifications and distances

Save classifications and distances. Позволяет сохранить в формате программы статистика таблицу, в которой содержатся значения всех переменных, их порядковые номера, номера кластеров, к которым они отнесены, и евклидовы расстояния от центра кластера до наблюдения. Записанная таблица может быть вызвана любым блоком или подвергнута дальнейшей обработке (рис.8.24).

Данные о кластерах и районах, вошедших в кластеры, представлены в таблице 8.3.

Информация о средних величинах по кластерам представлена в таблице 8.4.

Таблица 8.3 – Результаты кластер-процедуры для данных 2009 г.

№ кластера	Количество районов	Районы
1	1	Оренбургский
2	8	Адамовский, Новоорский, Новосергиевский, Первомайский, Переволоцкий, Сакмарский, Саракташский, Тоцкий
3	26	Абдулинский, Акбулакский, Октябрьский, Абдулинский, Александровский, Алексеевский, Беляевский, Бугурусланский, Бузулукский, Гайский, Грачевский, Домбаровский, Илекский, Кваркенский, Красногвардейский, Кувандыкский, Курманаевский, Матвеевский, Первомайский, Пономаревский, Светлинский, Сорочинский, Тюльганский, Шарлыкский, Ясненский, Соль-Илецкий

Распределение районов по кластерам в 2009 году таково, что:

1) в первый кластер (аномальный) вошел Оренбургский район.

2) во второй кластер вошли 8 районов (23% от общего числа районов).

3) в третий кластер вошли 26 районов (75% от общего числа районов).

Таблица 8.4 – Средние величины по кластерам в 2009 г.

Показатель	Среднее значение в 1 кластере	Среднее значение во 2 кластере	Среднее значение в 3 кластере
Объем платных услуг населению, тыс. руб.	1192125	204756,5	88838,35
Оборот розничной торговли, млн. руб.	2520,8	606,15	277,55
Оборот общественного питания, тыс. руб.	277460	45470,375	17136,72

Районы, вошедшие во 2-ой кластер, имеют средние значения показателей состояния и развития потребительского рынка относительно регионов 1-ого и 3-его кластеров. Следовательно, районы, попавшие в данный кластер, характеризуются *средним уровнем развития потребительского рынка*.

Средние показатели состояния и развития потребительского рынка районов 3-его кластера достаточно низкие относительно аналогичных показателей районов первого и второго кластера (табл.8.4). Это дает основание характеризовать районы, попавшие в данный кластер, как районы *с низким уровнем развития потребительского рынка*.

В то же время, в 2009 г. по сравнению с 2005 г. произошли изменения в распределении районов по кластерам.

Так, из группы районов (кластер 2) со средним уровнем развития потребительского рынка перешли в группу (кластер 3) с низким уровнем развития потребительского рынка 3 района: Северный, Соль-Илецкий и Ташлинский. Первомайский район, напротив, перешел в кластер 2 со средним уровнем развития потребительского рынка.

Необходимо также сказать, что произошли изменения средних значений показателей, о чем свидетельствуют рис. 8.21 и табл. 8.4.

Сравнительный анализ данных таблиц 8.2 и 8.4 показал, что объем платных услуг населению в Оренбургском районе в 2009 г. по сравнению с 2005 г. увеличился на 99937 тыс. руб., оборот розничной торговли вырос на 1576,9 млн. руб., оборот общественного питания увеличился на 153744 тыс. руб.

Средний объем платных услуг населению для кластера со средним уровнем развития потребительского рынка вырос на 122380,8 тыс. руб., средний оборот розничной торговли увеличился на 328 млн. руб., средний оборот общественного питания вырос на 27133,9 тыс. руб.

В кластере с низким уровнем развития потребительского рынка средний объем платных услуг населению увеличился на 46560 тыс. руб., средний оборот розничной торговли увеличился на 141,3 млн. руб., средний оборот общественного питания вырос на 10134,8 тыс. руб.

Пример 8.3. Кластер-процедура методом Joining (tree clustering) - древовидная кластеризация и методом k-средних

Для оценки степени различия административных территорий Оренбургской области по социально-экономической ситуации в целом и на рынке труда в частности в 2021 году проведите кластерный анализ по 35 районам Оренбургской области. Постройте дендрограмму, дайте характеристику кластерам, полученным с помощью метода k-средних.

В основу классификации (стандартизированные данные приведены в приложении О) включены следующие факторы:

X_1 - коэффициент напряженности на рынке труда, чел. на одну вакансию;

X_2 - коэффициент миграционного прироста (на 1000 человек населения);

X_3 - доля убыточных организаций, % от общей численности организаций;

X_4 - индекс физического объема инвестиций в основной капитал, в % к предыдущему году;

X_5 - индекс физического объема оборота розничной торговли, в % к предыдущему году;

X_6 - ввод в действие жилых домов на 1000 человек населения (квадратных метров общей площади).

Ход процесса:

1. Города Оренбургской области не могут быть включены в кластеры, так как по некоторым признакам отсутствует необходимая информация, что затрудняет процедуру кластеризации.

2. Данные по района Оренбургской области предварительно стандартизированы с помощью вкладки Данные/Стандартизация… в ППП «STATISTICA» (русскоязычная версия).

3. Осуществляем запуск модуля Кластерный анализ (русскоязычная версия) (рис.8.2). Выбираем метод кластеризации *Joining (tree clustering)-древовидная кластеризация* (рис.8.25), щелкаем по кнопке *ОК*.

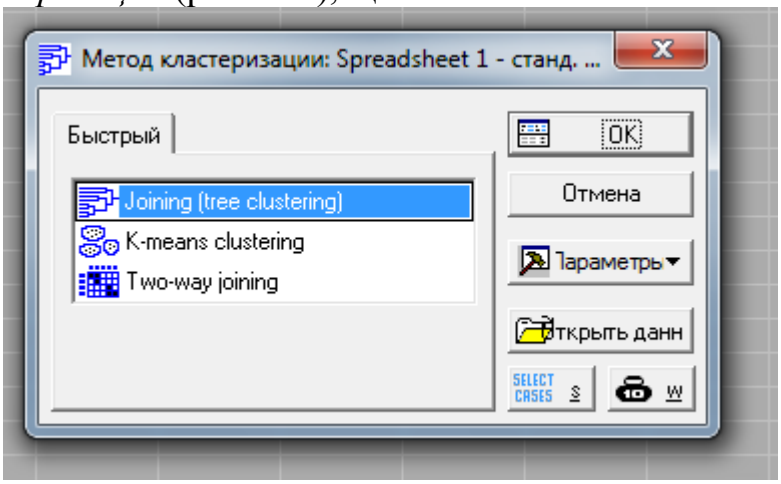


Рисунок 8.25 – Запуск модуля Joining (tree clustering)

4. В появившемся окне выбираем вкладку *Расширенный (Advanced)* (рис.8.26).

5. С помощью кнопки Variables (Переменные) выберем показатели $X_{1\text{-}6}$, по которым будет происходить кластеризация. В строке Cluster (Кластер) укажем объекты для классификации Cases [rows] (Наблюдения [строки]). Для объединения районов Оренбургской области в кластеры по признакам, указанным выше воспользовались правилом объединения - методом Уорда (*Ward's method*) и выбрали измерение Евклидово расстояние (*Euclidean distances*) (рис.8.26), щелкаем по кнопке *ОК*.

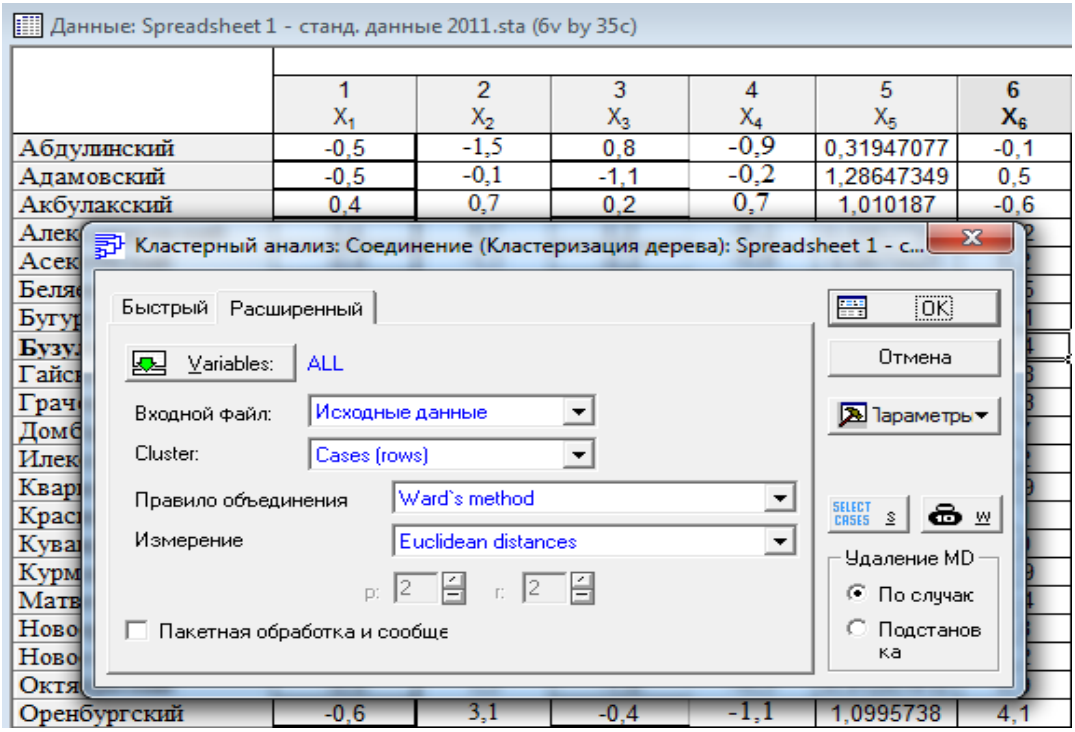


Рисунок 8.26 – Выбор параметров древовидной кластеризации

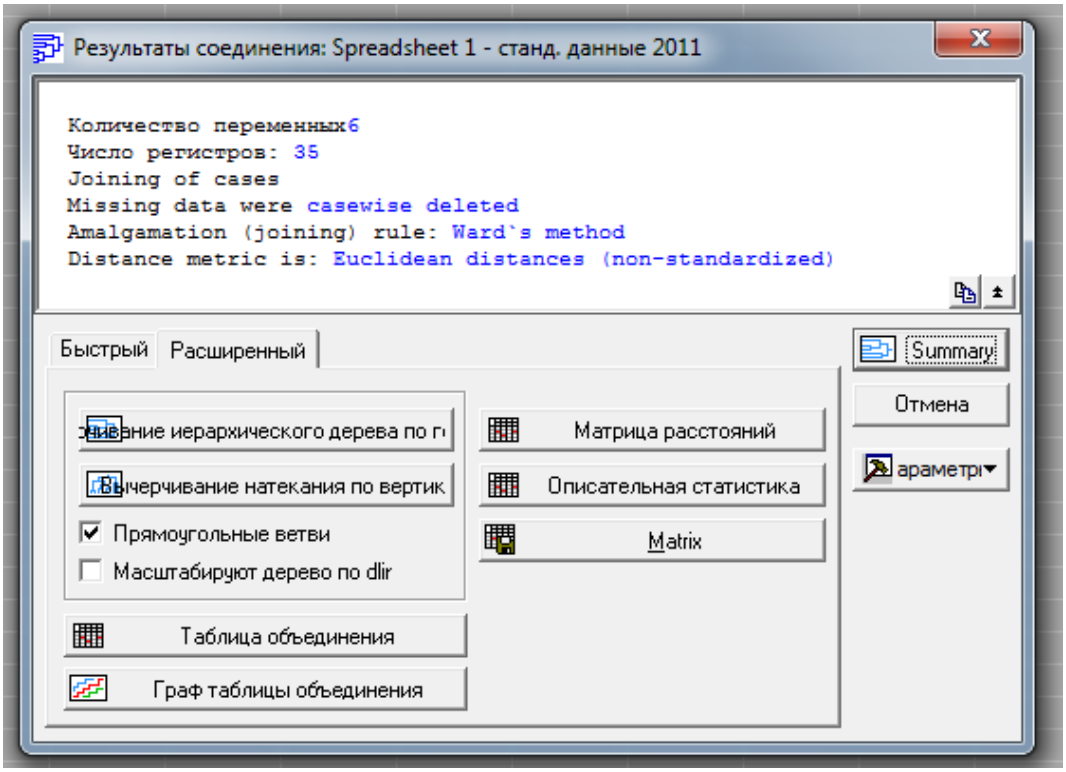


Рисунок 8.27 – Результаты соединения

6. В появившемся окне (рис.8.27) выбираем вкладки *Расширенный (Advanced) и Вычерчивание иерархического дерева по горизонтали (Horizontal hierarchical tree plot)*, нажимаем кнопку *Summary*. Результаты древовидной кластеризации представлены на рисунке 8.28.

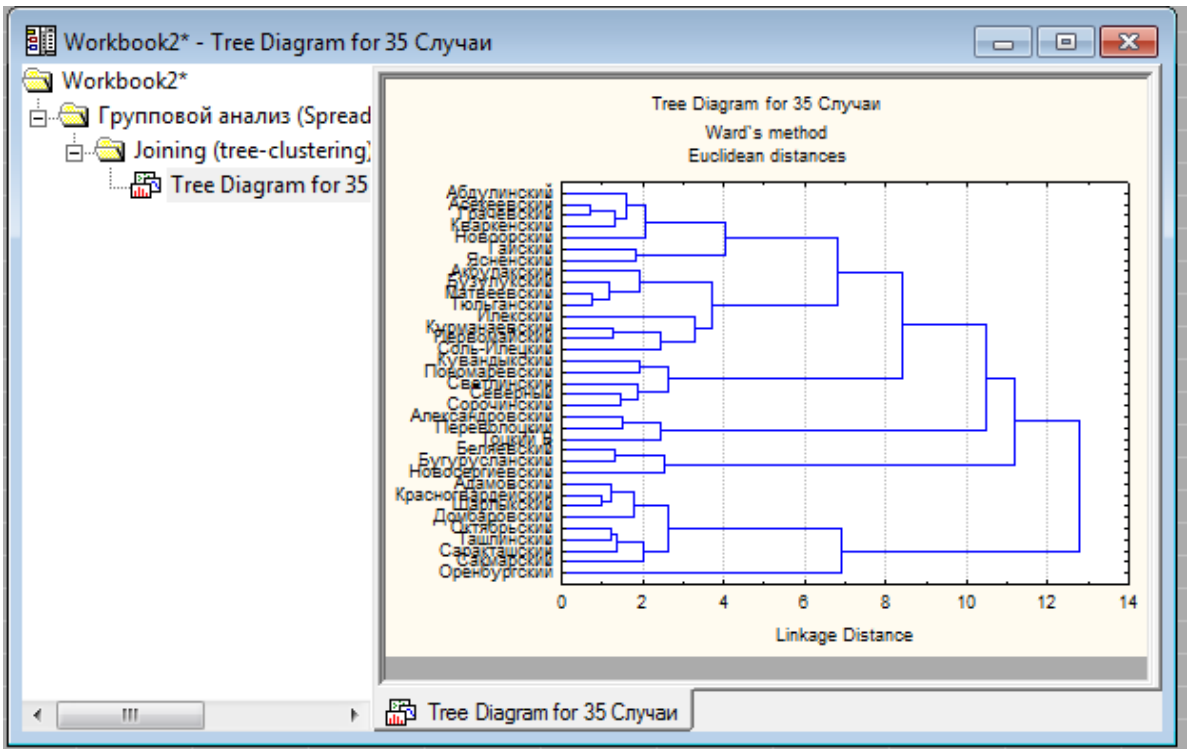


Рисунок 8.28 – Результаты древовидной кластеризации

Результат объединения – дендрограмма (рис.8.28), по оси ординат, которой отражены районы Оренбургской области, а по оси абсцисс показано значение интегрального показателя, представленного величиной, сформированной на основе исследуемых показателей. Данный показатель не имеет единицы измерения, а является многомерной статистической оценкой. В нашем случае - оценка социально-экономической ситуации в целом и на рынке труда Оренбургской области, в частности.

7. Осуществляем запуск модуля Кластерный анализ (русскоязычная версия) (рис.8.2). Выбираем метод кластеризации *K–means clustering* (рис.8.25), щелкаем по кнопке *OK*.

8. В появившемся окне (рис. 8.29) с помощью кнопки Variables (Переменные) выберем все показатели X , по которым будет происходить кластеризация. В строке Cluster (Кластер) укажем объекты для классификации Cases [rows] (Наблюдения [строки]). Поле Number of clusters (Число кластеров) позволяет ввести желаемое число кластеров, которое должно быть больше 1 и меньше чем количество объектов. Выберем разбиение на 3 кластера.

9. После соответствующего выбора нажимаем кнопку OK, ППП STATISTICA произведет вычисления и появится новое *окно k–Результаты кластеризации усреднений (k–means Clustering Results)* (рис. 8.30).

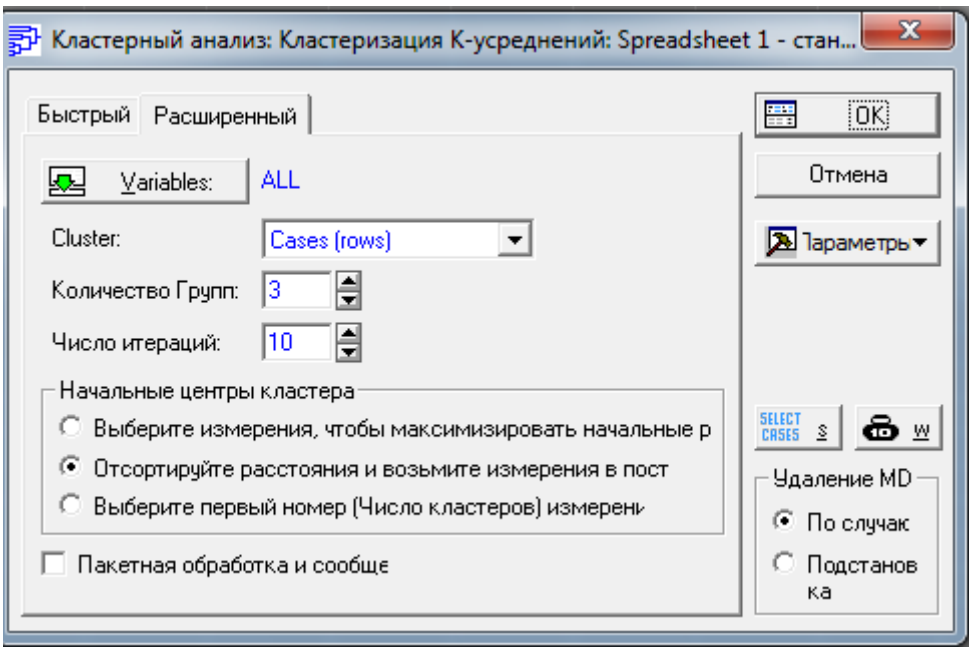


Рисунок 8.29 – Метод К-усреднений

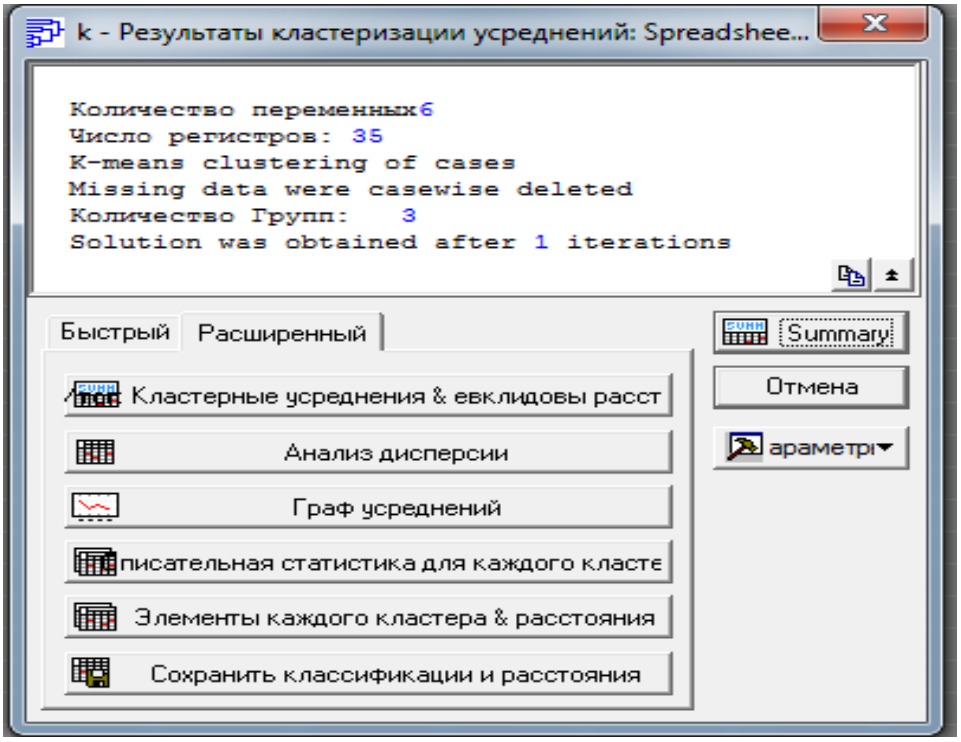


Рисунок 8.30 – Результаты применения метода К-усреднений

10. Для построения графика средних значений выбираем вкладку *«Граф усреднений» (Grapf of means)* (рис. 8.30). Результаты построения графика представлены на рис. 8.31.

11. Определяем наполняемость кластеров, выбрав вкладку *«Элементы каждого кластера & расстояния» (Members of each cluster & distances)* (рис. 8.30). Результаты представлены на рис. 8.32.

По результатам многомерной группировки получено 3 кластера (табл. 8.5), определяющие специфику социально-экономической ситуации в регионах.

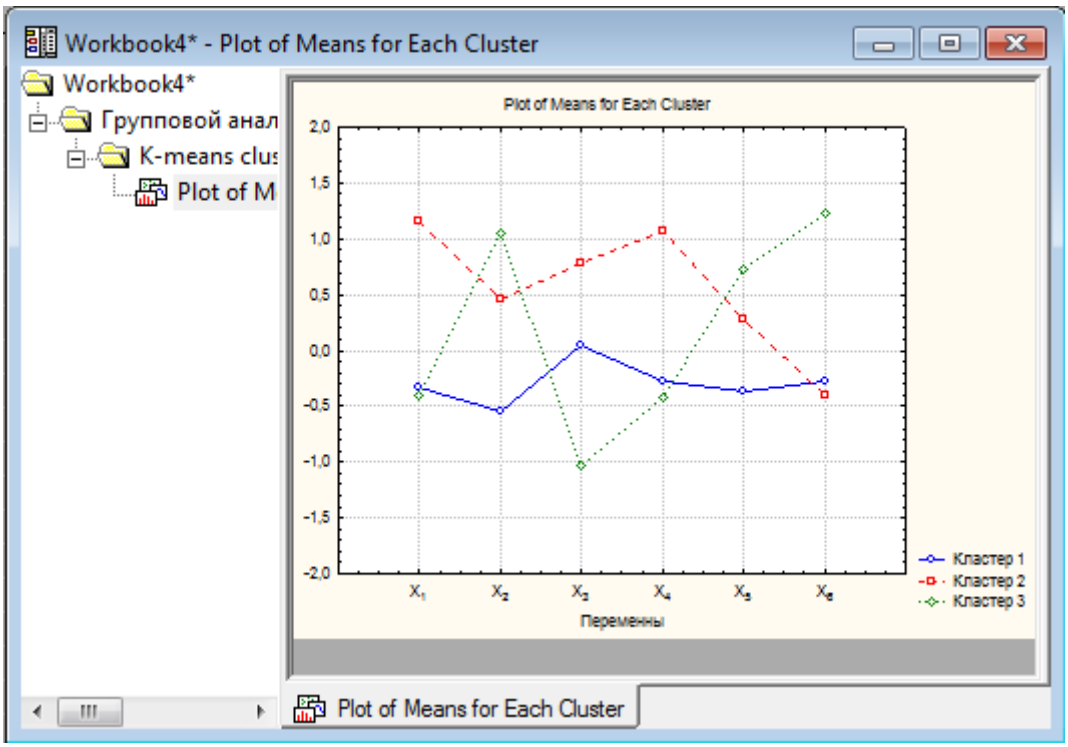


Рисунок 8.31 – График средних значений по кластерам

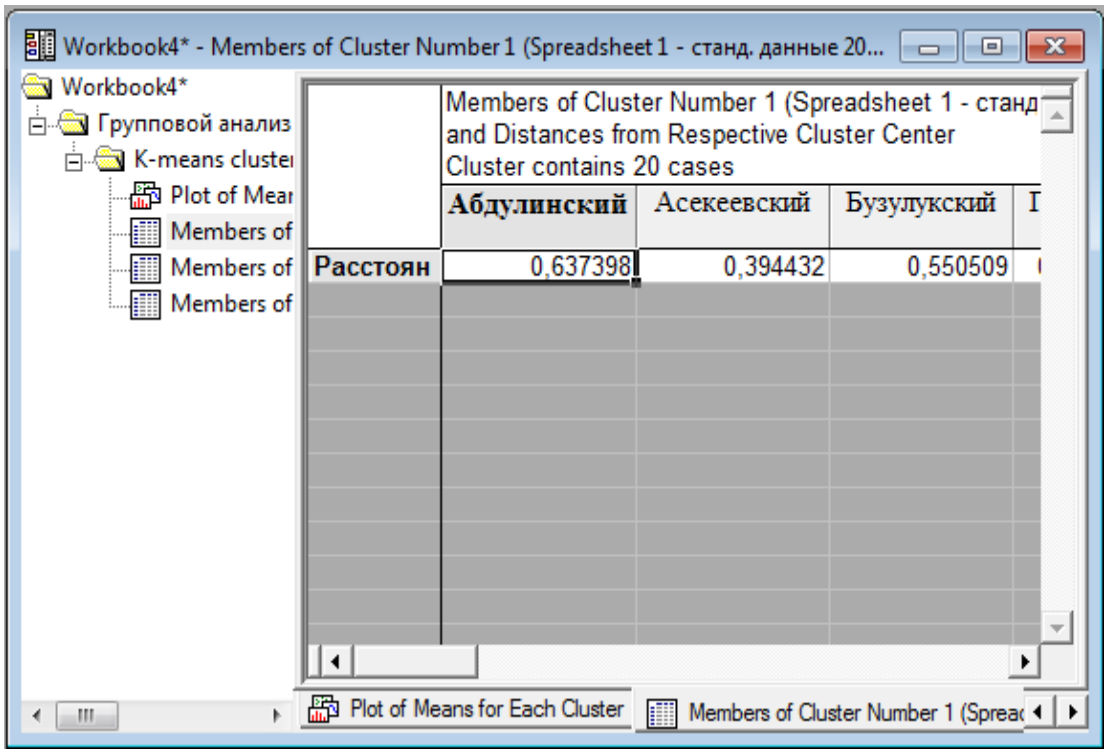


Рисунок 8.32 – Наполняемость кластеров

12. Для детальной характеристики кластеров рассчитаем средние значения показателей (по средней арифметической простой), составляющих основу классификации, по фактическим данным (приложение П). Результаты расчетов представлены в таблице 8.6.

Таблица 8.5 – Результаты кластеризации районов Оренбургской области

№ кластера	Количество районов	Наименование районов
1	20	Абдулинский, Асекеевский, Бузулукский, Гайский, Грачевский, Домбаровский, Кваркенский, Кувандыкский, Курманаевский, Матвеевский, Новоорский, Первомайский, Пономаревский, Светлинский, Северный, Соль-Илецкий, Сорочинский, Тюльганский, Шарлыкский, Ясненский
2	8	Акбулакский, Александровский, Беляевский, Бугурусланский, Илекский, Новосергиевский, Переволоцкий, Тоцкий
3	7	Адамовский, Красногвардейский, Октябрьский, Оренбургский, Сакмарский, Саракташский, Ташлинский

Таблица 8.6 – Средние значения по кластерам

№ кластера	X_1	X_2	X_3	X_4	X_5	X_6
1	9,77	-5,22	43,55	62,91	91,50	190,79
2	8,23	0,30	39,19	64,45	98,53	210,39
3	7,37	5,74	25,57	67,77	105,11	444,13

По данным рисунка 8.32 и таблицы 8.5 видно, что в первый кластер вошли 20 районов Оренбургской области (57% от общего числа районов).

Основу экономического потенциала районов первого кластера составляют предприятия промышленности, сельского хозяйства и субъекты малого и среднего предпринимательства. Районы занимают выгодное экономико-географическое положение, располагаясь на пересечении транспортных путей железнодорожного и автомобильного сообщения. Но, несмотря на это, темпы роста основных социально-экономических показателей данных районов достаточно низкие относительно аналогичных показателей районов второго и третьего кластера. В районах, попавших в первый кластер, отмечается определенное ухудшение показателей, характеризующих уровень и качество жизни населения.

Высокие средние значения показателей X_1, X_2 , X_3 и низкие значения показателей X_4, X_5 , X_6 для районов первого кластера (табл. 8.6), позволяют характеризовать социально-экономическую ситуацию, сложившуюся на рынке труда районов, попавших в первый кластер, как неблагоприятную. Так, доля убыточных организаций для районов первого кластера достаточно велика, что следует рассматривать как негативную тенденцию для рынка труда этих районов.

Во второй кластер вошли 8 районов (23%).

В данном кластере, в основном, присутствуют, районы, являющиеся центрами сельскохозяйственного производства. Промышленность этих районов представлена предприятиями обрабатывающих производств. Темпы роста основных социально-экономических показателей данных районов достаточно высокие относительно аналогичных показателей районов первого кластера. Данное обстоятельство позволяет отнести районы второго кластера

к районам со средними показателями социально-экономического развития и состояния рынка труда.

Значения показателей (табл.8.6), рассчитанные для районов второго кластера, по величине больше значений показателей первого кластера и меньше значений показателей третьего кластера.

Следовательно, социально-экономическую ситуацию в целом и на рынке труда в частности районов второго кластера следует считать хорошей относительно ситуации на рынке труда районов первого кластера.

В третий кластер вошли 7 районов Оренбургской области (20%).

Для этих районов характерна стабильная экономическая ситуация, сопровождающаяся стабильными темпами роста основных экономических показателей, здесь присутствуют в основном районы, являющиеся центрами нефтедобывающей, промышленности, крупными автотранспортными и железнодорожными узлами, центрами сельскохозяйственного, промышленного и обрабатывающего производства. В районах, попавших в третий кластер, отмечается определенное улучшение показателей, характеризующих развитие человеческого потенциала и качество жизни населения. Сохраняется положительная динамика роста заработной платы. Кроме того, наличие на территории районов третьего кластера значительного природно-ресурсного потенциала, развитой транспортно-энергетической инфраструктуры, производственных мощностей, характеризуют условия инвестиционной привлекательности.

На основе графического представления (рис. 8.31) средних значений признаков можно провести аналогичную экономическую интерпретацию для каждого кластера.

Анализируя полученные описательные характеристики (табл.8.6) можно отметить, что по средним значениям, районы, попавшие в третий кластер (20,0% от общего числа районов), можно отнести к районам, имеющим благополучную ситуацию на рынке труда. Об этом свидетельствуют низкие значения коэффициента напряженности и доли убыточных организаций. Для районов третьего кластера показатели X_4, X_5, X_6 достаточно высоки, поэтому экономическую ситуацию следует считать благополучной.

Пример 8.4. Метод главных компонент и кластеризация методом k-средних

Проблемы исследования текущего состояния финансов населения в регионах являются неотъемлемой частью комплексного анализа взаимосвязи финансов населения и экономического развития страны.

Так, по данным приложения Р за 2021г. проведите кластерный анализ регионов РФ методом k-средних по следующим показателям, предварительно выделив главные компоненты:

x_1 – среднемесячная номинальная начисленная заработная плата работников организаций, руб. в месяц;

x_2 – средний размер назначенных пенсий, руб.;

x_3 – удельный вес социальных выплат от общего объема денежных доходов населения субъекта, %;

x_4 – удельный вес доходов от предпринимательской деятельности от общего объема денежных доходов населения субъекта, %;

x_5 – величина прожиточного минимума (в среднем на душу населения), руб. в месяц;

x_6 – численность населения с денежными доходами ниже величины прожиточного минимума, в % от общей численности населения субъекта;

x_7 – коэффициент Джини;

x_8 – доля расходов на покупку и оплату товаров и услуг, в % от общего объема денежных доходов;

x_9 – доля обязательных платежей и разнообразных взносов, в % от общего объема денежных доходов;

x_{10} – доля расходов на прирост финансовых активов, в % от общего объема денежных доходов;

x_{11} – темпы роста (снижения) ВРП, в % к предыдущему году;

x_{12} – темпы роста (снижения) инвестиций в основной капитал, в % к предыдущему году;

x_{13} – темпы роста (снижения) объемов промышленного производства, в % к предыдущему году;

x_{14} – темпы роста (снижения) оборота розничной торговли, в % к предыдущему году;

x_{15}– индекс потребительских цен (декабрь к декабрю предыдущего года), %;

x_{16} – уровень зарегистрированной безработицы, %;

x_{17} – численность занятых в экономике, приходящаяся на одного пенсионера (в среднем за год), человек;

x_{18} – коэффициент напряженности на рынке труда, чел. на 1 вакансию;

x_{19} – удельный вес убыточных организаций, в % от общего числа организаций.

Ход процесса:

1. Показатели x_1– x_4 характеризуют формирование финансов населения; x_5– x_7 – дифференциацию доходов населения; x_8– x_{10} – использование финансов населения; x_{11}– x_{15} – экономический потенциал регионов; x_{16}– x_{19} – состояние рынка труда.

Так как исходные показатели ($X_1 – X_{19}$) несопоставимы, для вычисления компонент целесообразно использовать стандартизированные данные (приложение Р).

2. Для реализации процедуры метода главных компонент необходимо вызвать диалоговое окно **Factor Analysis** в англоязычной версии или через меню **Статистика \ Многомерные исследующие методы \ Анализ особенности** в русскоязычной версии (рис.8.5).

3. В появившемся окне (рис.8.33) выбираем вкладку Variables, во втором появившемся окне выделяем все переменные нажимаем ОК.

4. В новом окне (рис.8.34) указываем метод выборки – *Основные*

компоненты, указываем число главных компонент, в нашем случае их 3. Нажимаем кнопку ОК.

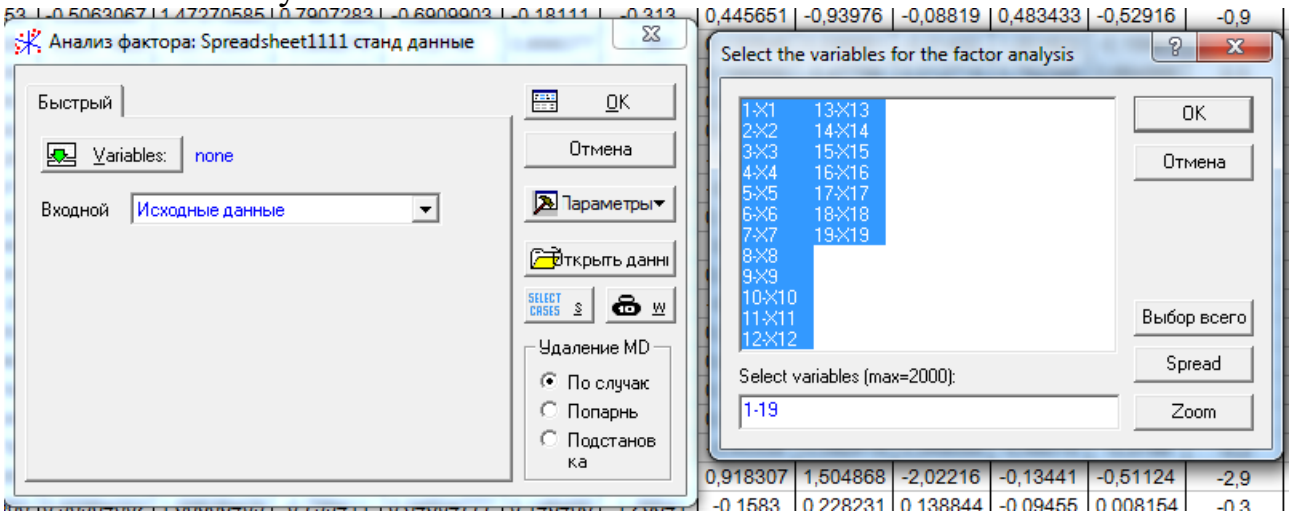


Рисунок 8.33 – Выбор переменных для проведения метода главных компонент

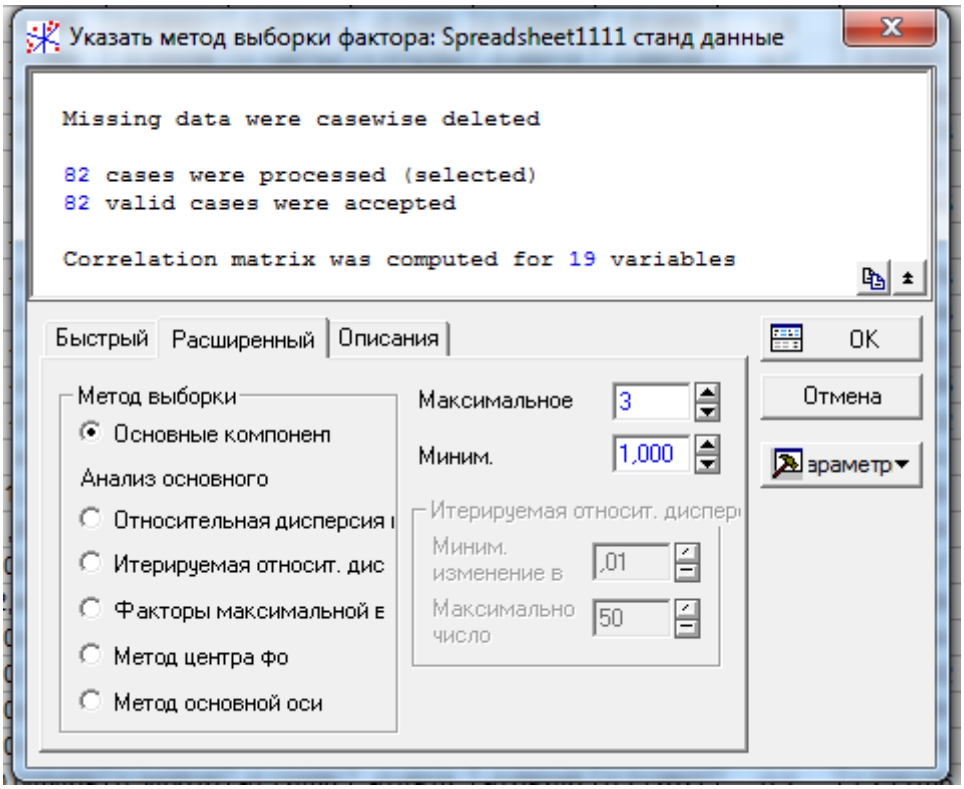


Рисунок 8.34 – Метод выборки фактора

5. Далее необходимо воспользоваться критерием «каменистой осыпи» и определить количество факторов (X_1 – X_{19}), то есть построить график собственных значений (рис. 8.37). Для этого переходим по вкладкам *Объяснимая дисперсия – Вычерчивание кусков – Summary* (рис.8.36).

Из рисунка 8.37 следует, что всю вариабельность показателей достаточно хорошо иллюстрируют 3 главные компоненты.

6. Возвращаемся к вкладке *Объяснимая дисперсия* (рис.8.36), выбираем *Eigenvalues*, результаты вклада первых трех компонент в общей дисперсии приведены на рис. 8.38.

Анализ данных рис. 8.38 показывает, что вклад первых трех компонент составляет 59,84% в общей дисперсии, следовательно, можно сделать вывод, что три полученные компоненты объясняют большую часть вариабельности переменных.

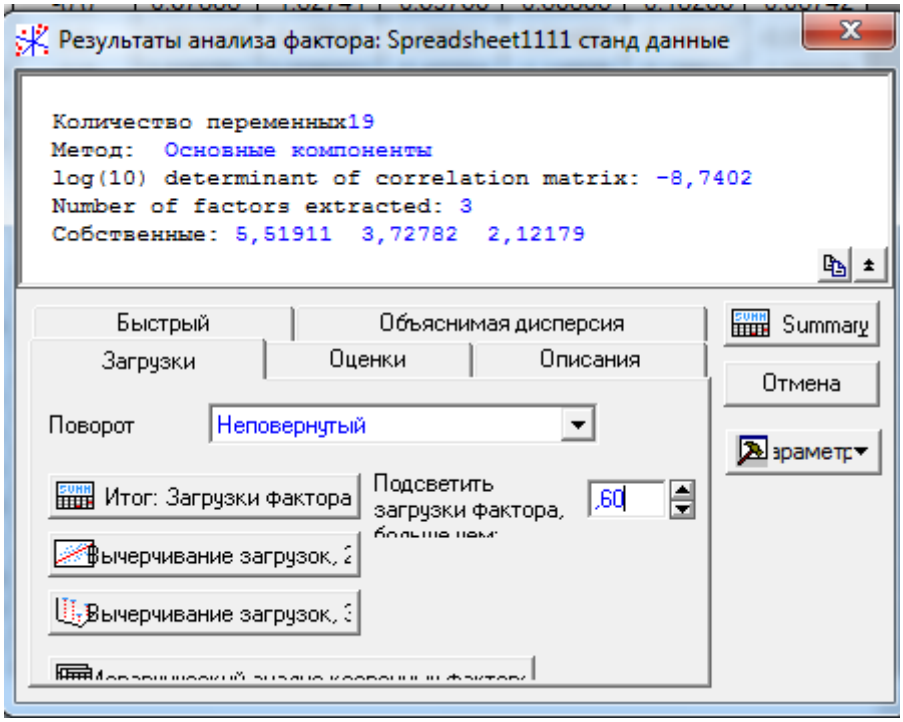


Рисунок 8.35 – Результаты анализа факторов

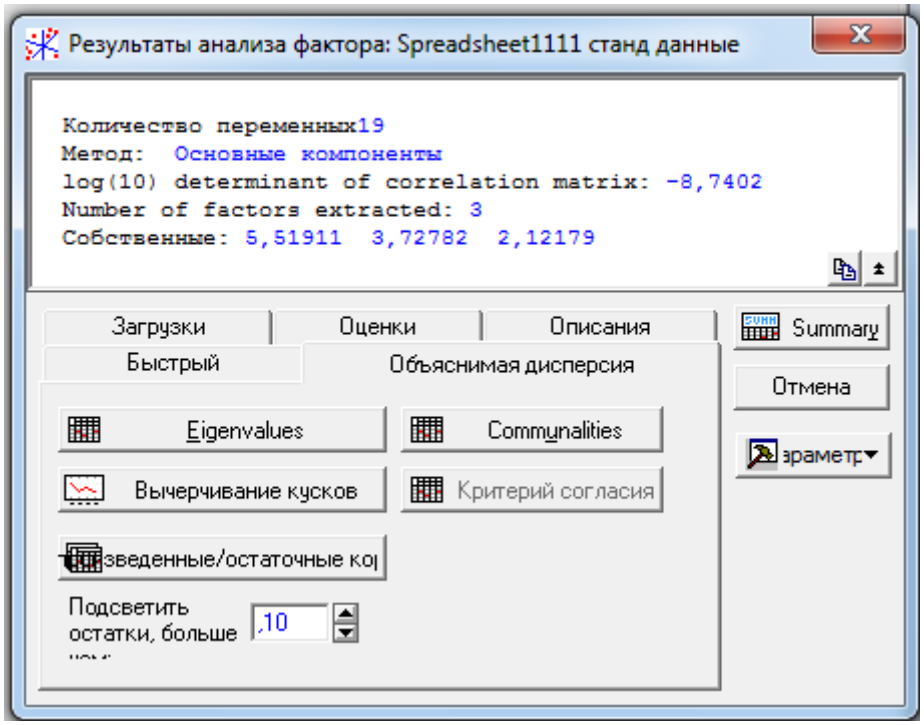


Рисунок 8.36 – Меню для построения графика собственных значений

7. Следующий шаг - построение матрицы факторных нагрузок на три выделенные главные компоненты и их интерпретация. При проведении

интерпретации оставленных для анализа главных компонент необходимо оставить только те признаки, с которыми главные компоненты имеют существенную связь (> 0,6).

Признаки, для которых величина факторных нагрузок превышает пороговое значение (> 0,6), в дальнейшем будут использованы при экономической интерпретации главных компонент.

Возвращаемся к вкладке *Загрузки* (рис.8.36), выбираем поле *Подсветить загрузки фактора больше чем:* в нашем случае 0,6, далее выбираем вкладку *Итог: загрузки фактора* (рис. 8.39).

Результаты представлены на рис. 8.40.

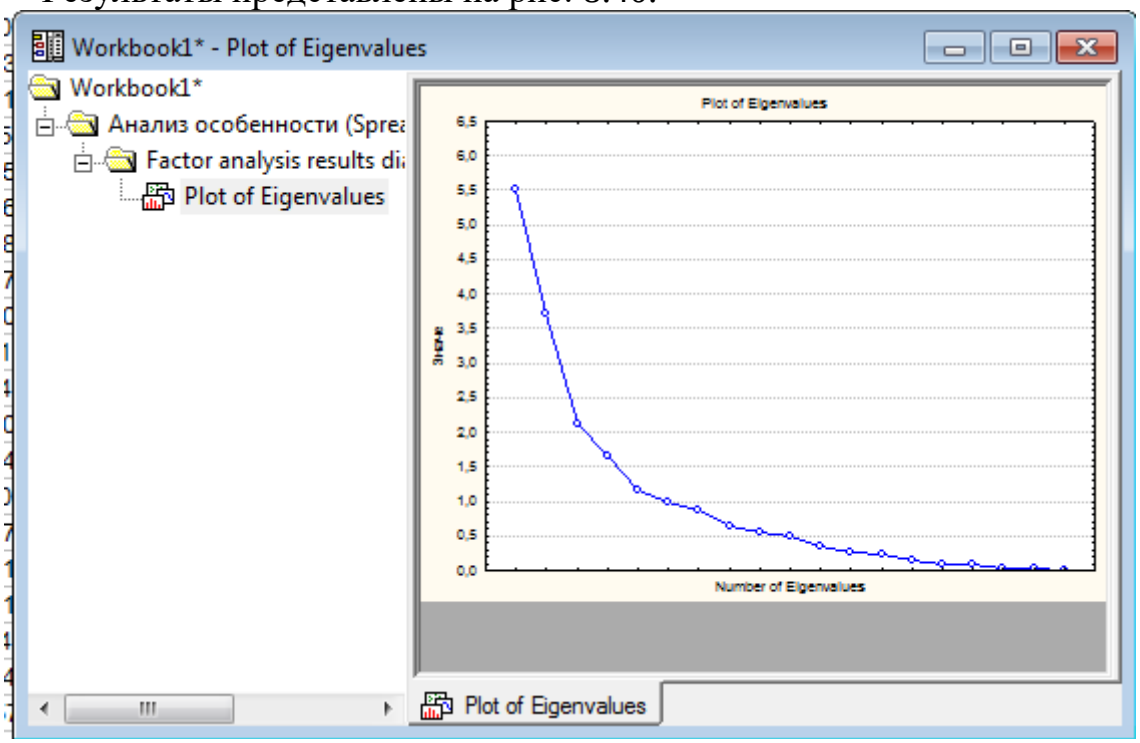


Рисунок 8.37 – Собственные значения факторов относительно номеров факторов для регионов РФ в 2021 году

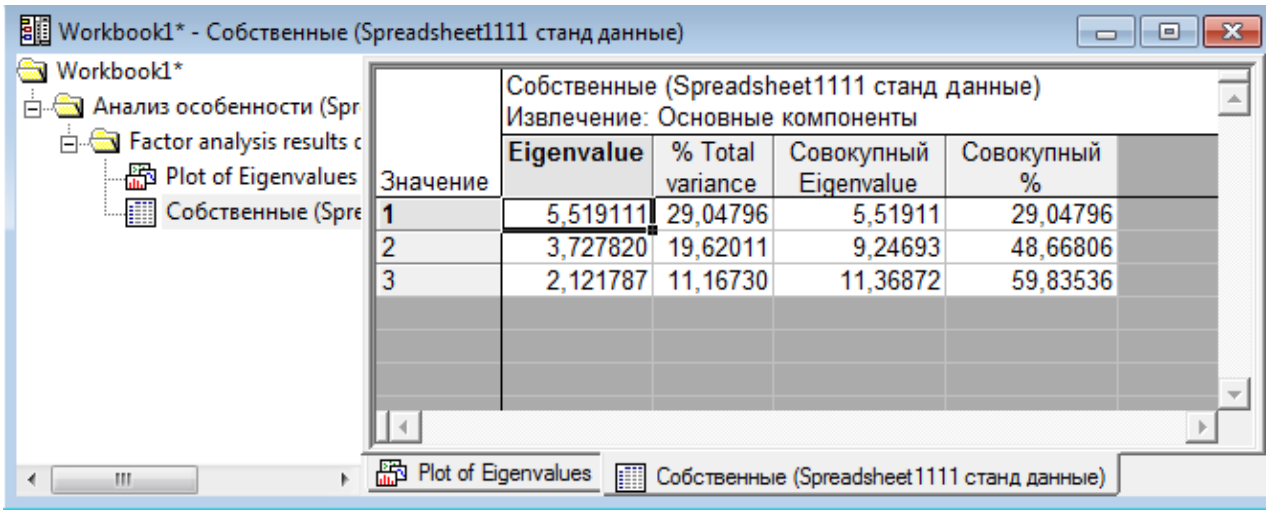


Рисунок 8.38 – Результаты вклада первых трех компонент в общей дисперсии

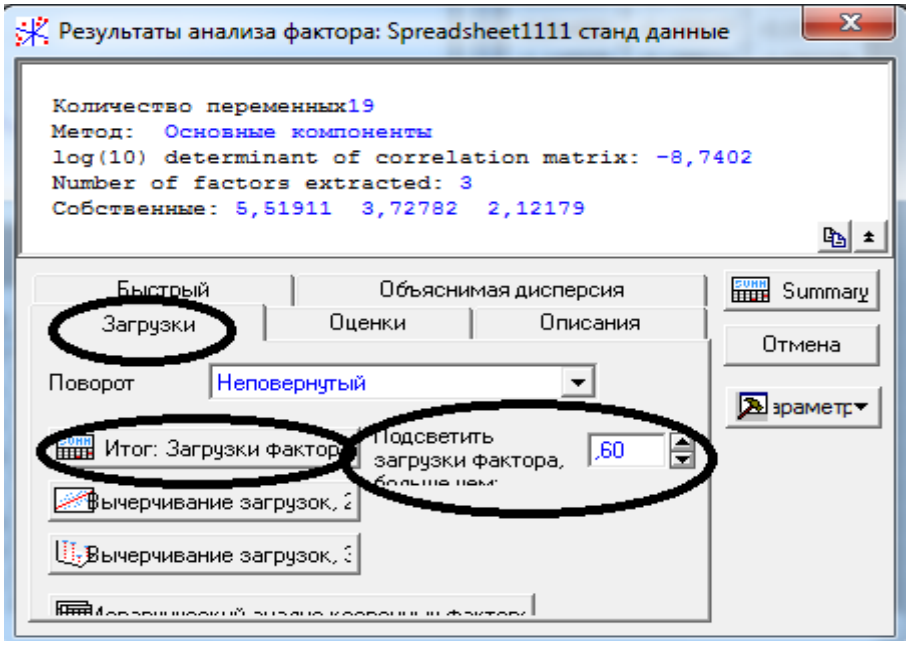


Рисунок 8.39 – Меню для построения матрицы факторных нагрузок

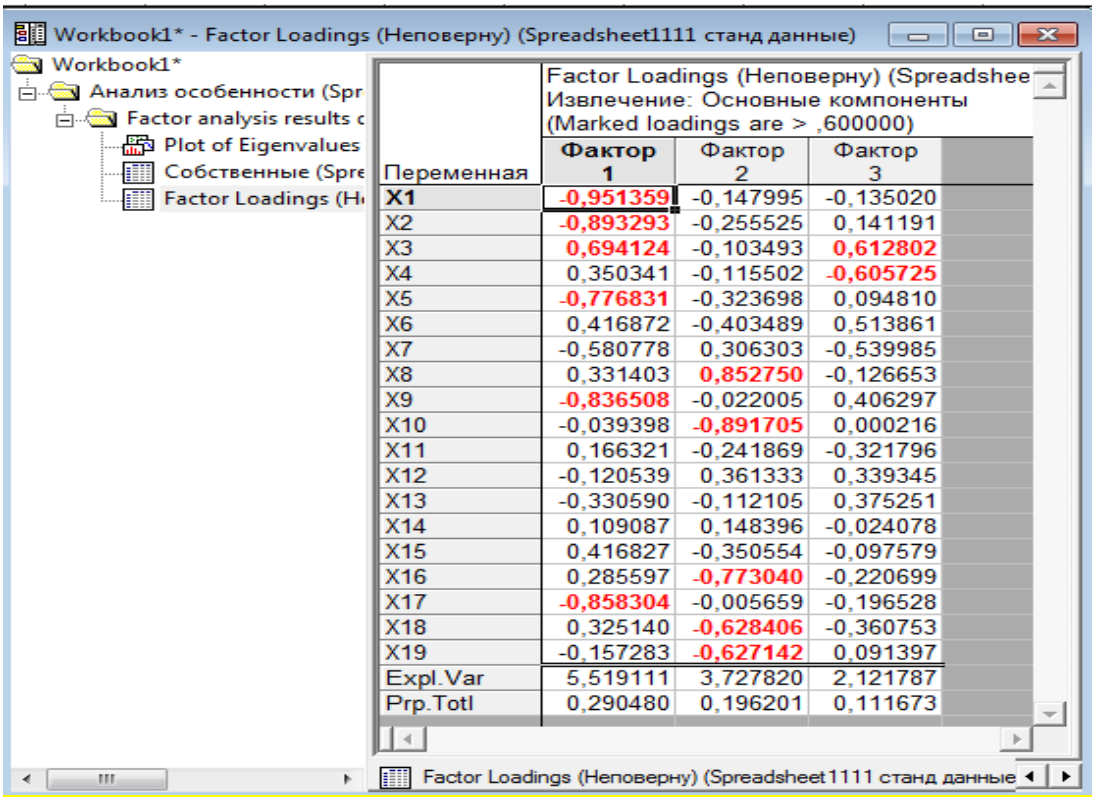


Рисунок 8.40 – Результаты построения матрицы факторных нагрузок

Согласно данным рисунка 8.40, *первая главная компонента* **(Фактор 1)** наиболее тесно связана со следующими показателями: X_1 (среднемесячная номинальная начисленная заработная плата работников организаций), X_2 (средний размер назначенных пенсий), X_5 (величина прожиточного минимума), X_9 (доля обязательных платежей и разнообразных взносов), X_{17} (численность

занятых в экономике, приходящаяся на одного пенсионера (в среднем за год)), поэтому названа «доходно-трудовой».

Вторая главная компонента **(Фактор 2)** связана с показателями, характеризующими использование финансов населения и состояние рынка труда, X_8 (доля расходов на покупку и оплату товаров и услуг), X_{10} (доля расходов на прирост финансовых активов), X_{16} (уровень зарегистрированной безработицы), X_{18} (коэффициент напряженности на рынке труда), X_{19} (удельный вес убыточных организаций), поэтому названа «расходно-трудовой».

Третья главная компонента **(Фактор 3)** наиболее тесно связана с показателями, характеризующими формирование финансов населения, X_3 (удельный вес социальных выплат от общего объема денежных доходов населения субъекта) и X_4 (удельный вес доходов от предпринимательской деятельности от общего объема денежных доходов населения субъекта), поэтому названа «доходный потенциал».

Таким образом, выделены три компоненты, достаточно адекватно описывающих исходное признаковое пространство. Необходимо отметить тот факт, что две из трех компонент связаны с показателями, характеризующими формирование, использование финансов населения и состояние рынка труда, что говорит о значимости данных показателей при анализе социально-экономического развития регионов России. Кроме того, ни одна из компонент не связана тесно с показателями экономического потенциала регионов. Объясняется это, на наш взгляд, тем, что в 2021 г. основу устойчивого развития регионов России, определяли, прежде всего, проблемы рынка труда и материальная составляющая уровня жизни населения – показатели формирования и использования финансов населения.

8. Для оценки степени различия субъектов РФ в 2021 году по ряду выбранных показателей был проведён *кластерный анализ* по 82 регионам РФ (n=82). Чеченская республика не включена в кластеры, так как по большинству признаков отсутствует необходимая информация, что затрудняет процедуру кластеризации.

Классификация n=82 регионов проводилась по трем первым главным компонентам с помощью метода k-средних (ход процесса и анализ в примерах 8.2-8.3). По содержательным и статистическим критериям наилучшим оказалось разбиение на 5 кластеров, результаты кластеризации представлены в таблице 8.7.

Средние значения, вычисленные как средние арифметические простые по фактическим данным для каждого кластера, показателей по кластерам 2021 г. представлены в табл. 8.8.

По данным таблицы 8.8 следует, что второй кластер характеризуется высокими значениями средних показателей формирования финансов населения, что характеризует уровень жизни населения этих регионов как относительно высокий. Большинство показателей, характеризующих состояние рынка труда, также имеют высокие значения, что свидетельствует о проблемах в области трудового потенциала регионов.

Первый кластер характеризуется средними значениями показателей формирования и использования финансов населения субъектов РФ, что характеризует уровень жизни населения этих регионов как стабильно высокий. Большинство показателей, характеризующих состояние рынка труда данных регионов, имеют низкие значения относительно аналогичных показателей 2–ого кластера, что свидетельствует о стабильности в сфере занятости.

Таблица 8.7 – Результаты кластеризации регионов РФ в 2021 году*

№ кластера	Количество субъектов РФ	Наименование субъектов РФ
1	16	Области: Московская, Архангельская без НАО, Калининградская, Мурманская, Челябинская, Томская, Амурская, Сахалинская, Республика Саха (Якутия); г. Москва; Республики: Карелия, Коми; Края: Красноярский, Приморский, Хабаровский; Еврейская авт. область
2	7	Автономные округа: Ненецкий, Ханты-Мансийский, Ямало-Ненецкий, Чукотский; Камчатский край; Области: Тюменская без авт. округа, Магаданская
3	1	Республика Ингушетия
4	41	Области: Белгородская, Брянская, Владимирская, Воронежская, Ивановская, Костромская, Курская, Липецкая, Орловская, Рязанская, Смоленская, Тамбовская, Тверская, Тульская, Ярославская, Вологодская, Ленинградская, Новгородская, Псковская, Кировская, Пензенская, Саратовская, Ульяновская, Курганская, Иркутская, Кемеровская, Омская Республики: Адыгея, Кабардино-Балкарская, Калмыкия, Карачаево-Черкесская, Северная Осетия – Алания, Марий Эл, Мордовия, Удмуртская, Алтай, Бурятия, Тыва, Хакасия, Края: Алтайский, Забайкальский
5	17	Области: Калужская, Астраханская, Волгоградская, Ростовская, Нижегородская, Оренбургская, Самарская, Свердловская, Новосибирская; г. Санкт-Петербург; Республики: Дагестан, Башкортостан, Татарстан, Чувашская; Края: Пермский, Краснодарский, Ставропольский

*Рассчитано авторами с помощью ППП «STATISTICA, версия 6.0»

Третий кластер, вследствие маленькой размерности (входит только Республика Ингушетия), следует считать аномальным.

Регионы четвертого (самого многочисленного) кластера характеризуются самыми низкими доходами, но в тоже время самой высокой долей расходов на товары и услуги от дохода, что, как и в первом кластере, характеризует сложное экономическое положение населения.

Таблица 8.8 – Средние уровни показателей, характеризующих социально-экономическое развитие регионов России по кластерам за 2021 г.

Макроэкономические показатели	Кластер 1	Кластер 2	Кластер 4	Кластер 5
X_1, рублей в месяц	19996,7	35748,6	12202,6	13746,2
X_2, рублей	5203,6	7138,8	4298,4	4338,2
X_3, в % от общего объема денежных доходов	13,7	9,9	17,5	13,7
X_4, в % от общего объема денежных доходов	10,1	8,3	11,3	12,5
X_5, руб. в месяц	6137,4	7454,3	4246,5	4387,8
X_8, в % от общего объема денежных доходов	67,6	48,6	67,0	79,1
X_9, в % от общего объема денежных доходов	12,7	18,4	10,2	10,4
X_{10}, в % от общего объема денежных доходов	16,2	31,9	20,4	7,9
X_{16}, %	2,1	2,3	2,1	1,5
X_{17}, человек	1,9	2,8	1,6	1,8
X_{18}, человек на 1 вакансию	2,2	1,8	4,5	9,2
X_{19}, в % от общего числа организаций	34,0	34,0	30,2	24,3

Пятый кластер включает регионы со средними значениями всех показателей, но в отличие от остальных кластеров более высокой долей расходов на товары и услуги от дохода. Рынок труда характеризуется как относительно благополучный, по сравнению с аналогичными показателями других кластеров, однако среднее значение коэффициента напряженности (чел. на 1 вакансию) значительно выше, чем по другим кластерам.

Глава 9. СТАТИСТИЧЕСКИЕ МЕТОДЫ ПРОГНОЗИРОВАНИЯ ЭКОНОМИЧЕСКИХ ПРОЦЕССОВ

Статистический прогноз (conditional prediction) - прогноз, применяемый тогда, когда предусматривается, что лицо, принимающее решение, может осуществлять различные меры, которые способны воздействовать на прогнозируемые показатели.

Различают активный и пассивный статистический прогноз. Например, если наблюдается неблагоприятная тенденция к понижению фондоотдачи, то пассивный прогноз предскажет дальнейшее снижение этого показателя. Активный же прогноз ответит на вопрос, что будет, если окажется принятой та или иная программа действий по повышению эффективности фондов.

Для реализации статистического прогноза используются определенные методы. Выделяют следующие методы статистического прогнозирования (рисунок 9.1).

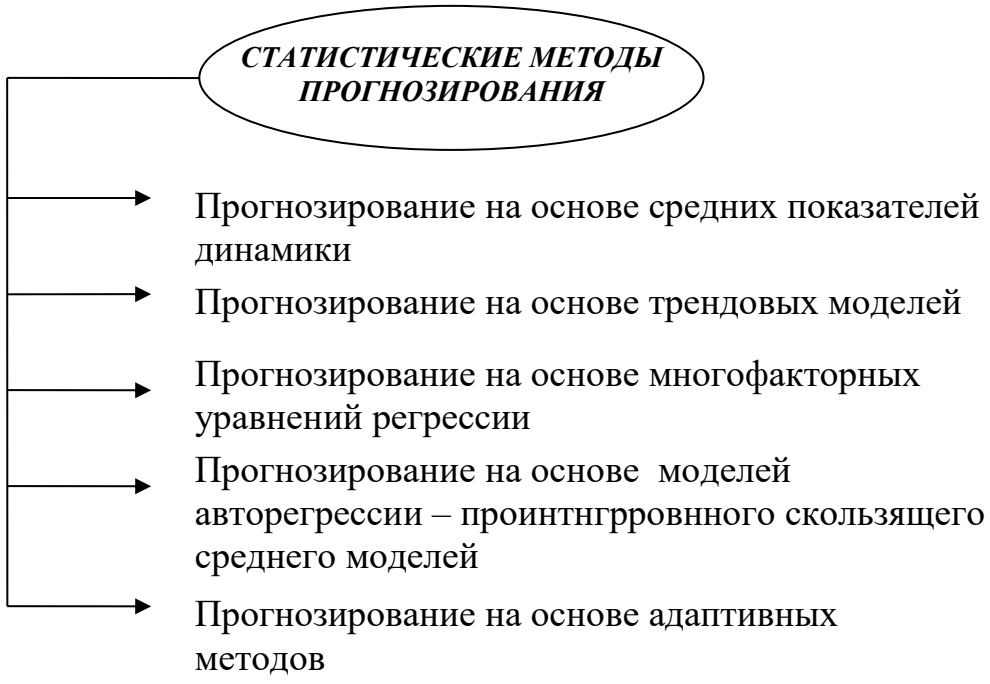


Рисунок 9.1 – Статистические методы прогнозирования, используемые в анализе экономических процессов

Статистические методы прогнозирования дополняются эконометрическими методами анализа, статистическими методами прогнозирования, приемами эконометрического моделирования и другими.

Рассмотрим более детально методы прогнозирования, представленные на рисунке 9.1.

Простейшими методами прогнозирования динамики являются методы **прогнозирования на основе среднего абсолютного прироста** и **среднего темпа роста.**

Прогноз на основе абсолютного прироста (9.1).

$$S_t = S_0 + \overline{\Delta}_s \cdot t, \qquad (9.1)$$

где S_t - прогнозируемый уровень;

S_0 - конечный уровень исходного ряда;

$\overline{\Delta}_s$ - средний абсолютный прирост за ряд лет, примыкающий к началу прогнозного периода: $\overline{\Delta}_s = \frac{S_n - S_0}{n-1}$

t – номер периода, на которое делается прогноз.

Прогноз на основе среднего темпа роста (9.2).

$$S_t = S_0 \cdot \overline{T}_s^t, \qquad (9.2)$$

где S_t - прогнозируемый уровень;

S_0 - - конечный уровень исходного ряда;

$\overline{T}_s^t$ - средний темп роста численности за ряд лет, примыкающих к началу прогнозируемого периода: $\overline{T}_s^t = \sqrt[n-1]{\frac{S_n}{S_0}}$

t – номер периода, на которое делается прогноз.

Прогнозирование на основе трендовых моделей. Для описания тенденции временного ряда на практике используют модели кривых роста, описывающиеся формулой вида:

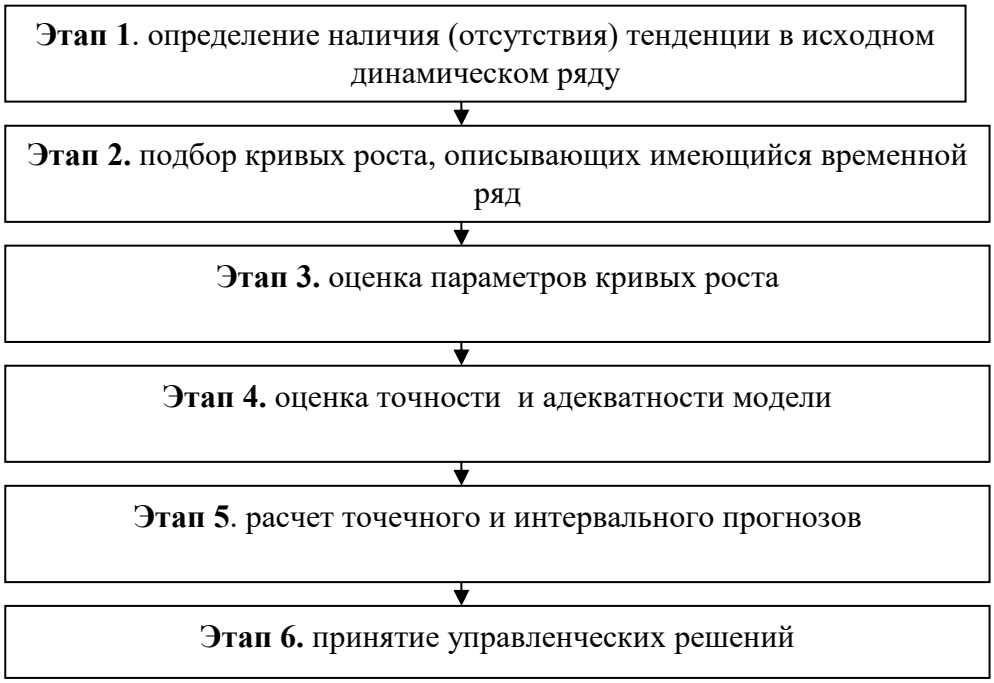
$$y = f(x) \qquad (9.3)$$

При таком научном подходе предполагается, что изучаемое явление связано с течением времени. Точно выбранная модель кривой роста должна соответствовать характеру изменений изучаемого явления. Прогнозирование на основе модели кривой роста основано на экстраполяции, т.е. на продлении в будущее тенденции, наблюдавшейся за предыдущие периоды времени.

Проблема выбора кривой роста – первоочередная при прогнозировании, т.к. от ее правильности зависит точность исследования в целом и, соответственно, точность прогноза.

В современной литературе описано несколько десятков кривых роста, к ним относятся: линейные (прямая) и нелинейные (гипербола, парабола, экспонента, логарифма, кривая Гомперца и т.д.)

Процедура разработки прогноза с использованием кривых роста включает несколько этапов:



Этап 1. Для того чтобы анализировать исходные данные и прогнозировать их дальнейшее развитие, в первую очередь необходимо выяснить существует ли тенденция вообще в изучаемом явлении. Это осуществляется путем проверки статистической гипотезы о случайности ряда. Для этого могут быль использованы критерий серий, основанный на медиане выборки, критерий «восходящих и нисходящих» серий или метод Фостара-Стюрта.

Этап 2. Если тенденция существует, то переходим к выбору кривой роста, наиболее точно описывающей процесс изменения в исследуемом явлении. После выявления тенденции приступаем к выбору лучшего уравнения тренда. Лучшим считается то уравнение тренда, в котором коэффициент апроксимациии очень близок к единице. Существует несколько методов облегчающих процесс выбора формы кривой роста. Наиболее простой метод – визуальный, опирающийся на графическое изображении.

Самым простым типом тренда является прямая линия, описываемая линейным уравнением тренда:

$$\tilde{y}_i = a + b \cdot t_i, \qquad (9.4)$$

где $\tilde{y}_i$ - выровненные, т.е. лишенные колебаний, уровни тренда для лет с номером i;

a - свободный член уравнения, численно равный среднему выровненному уровню для момента или периода времени, принятого за начало отсчета;

b - средняя величина изменения уровней ряда за единицу изменения времени.

Основные свойства тренда в форме прямой линии таковы:

- равные изменения за равные промежутки времени;

- если средний абсолютный прирост - положительная величина, то относительные приросты или темпы прироста постепенно уменьшаются;
- если тенденция к сокращению уровней, а изучаемая величина является по определению положительной, то среднее изменение b не может быть больше среднего уровня a;
- при линейном тренде ускорение, т.е. разность абсолютных приростов за последние периоды, равно 0.

Помимо прямолинейного тренда в статистике при анализе временных рядов также могут использовать параболический, экспоненциальный, логарифмический, гиперболический и логистический тренды.

Этап 3. Оценка параметров кривых роста осуществляется с помощью метода наименьших квадратов (МНК), суть которого сводится к минимизации квадрата отклонений фактических и теоретических значений временного ряда:

$$\sum (y_i - \tilde{y}_i)^2 \to \min \qquad (9.5)$$

В результате минимизации получается система нормальных уравнений, имеющая решение. Решение системы нормальных уравнений позволяет вычислить оценки искомых коэффициентов.

Этап 4. Проверка адекватности выбранных моделей реальному процессу строится при анализе случайных компонент. Ряд остатков как отклонение физических уровней от выровненных:

$$e_t = y_t - \hat{y}_t \qquad (9.6)$$

Принято считать, что модель адекватно описываемому процессу, если значение остаточной компоненты подчиняется случайному закону распределения.

Если вид функции выбран неудачно, то последовательные значения ряда остатков могут не обладать свойствами независимости, так как они могут коррелировать между собой, т.е. будет иметь место автокорреляции ошибок. Существует несколько методов обнаружения автокорреляции. Наиболее распространен критерий Дарбина – Уотсона:

$$d = \frac{\sum_{t=2}^{n} (e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^{\,2}} \approx 2(1 - r_t) \qquad (9.7)$$

где r_t - коэффициент автокорреляции первого порядка.

Если в ряду остатков имеется сильная положительная автокорреляция, то d=0. Если сильная отрицательная, то d=4. При отсутствии автокорреляции d=2. Применение на практике данного критерия основано на сравнении величины d с теоретическими табулированными значениями d_1 и d_2.

Если $d \langle d_1$, то нулевая гипотеза о независимости случайных отклонений отвергается (отсутствие автокорреляции).

Если $d > d_2$, то нулевая гипотеза не отвергается.

Если $d_1 \le d \le d_2$, нет достаточных оснований для принятия решений.

Когда расчетное значение d превышает 2, то с d_1 и d_2 сравнивается не сам коэффициент d, а $(4-d)$.

Важнейшими характеристиками качества моделей являются показатели ее точности:

Абсолютная ошибка прогноза: $$\Delta t = \hat{y}_t - y_t \quad (9.8)$$

Относительная ошибка прогноза: $$\delta_t = \frac{\hat{y}_t - y_t}{y_t} * 100 \quad (9.9)$$

Средняя абсолютная ошибка по модулю: $$S_t = \frac{\sum |\hat{y}_t - y_t|}{n} \quad (9.10)$$

Средняя относительная ошибка по модулю: $$|\bar{\delta}_t| = \frac{1}{n} \sum \left| \frac{\hat{y}_t - y_t}{y_t} \right| * 100 \quad (9.11)$$

Если $|\bar{\delta}_t| < 10\%$, это свидетельствует о высокой точности. Если $10\% \le |\bar{\delta}_t| \le 20\%$ - точность хорошей модели. Если $|\bar{\delta}_t| > 20\%$, но $|\bar{\delta}_t| < 50\%$ - удовлетворительная точность.

Этап 5. После проведенного предварительного анализа переходим к прогнозированию простой трендовой модели. Простая трендовая модель динамики – это уравнение тренда с указанием начала отсчета единиц времени. Прогноз по этой модели заключается в подстановке в уравнение тренда номера периода, который прогнозируется. Такой прогноз называется точечным. Точечный прогноз – «это скорее абстракция, чем реальность» [2], поэтому необходимо также делать интервальный прогноз. При таком прогнозе учитываются как вызванная колеблемостью ошибка репрезентативности выборочной оценки тренда, так и колебания уровней в отдельные периоды (моменты) относительно тренда.

Изучение закономерностей развития явлений, выявление и характеристика трендов создают базу для прогнозирования, т.е. для определения ориентированных размеров явлений в будущем.

Экстраполяция применяется в перспективном прогнозировании. Предполагает, что установленная тенденция в прошлом периоде будет сохраняться и в будущем. Для этих целей могут быть использованы выравнивание уровней динамического ряда по способу наименьших квадратов и подстановка в полученное уравнение соответствующих значений t.

Прежде всего, вычисляют точечный прогноз – значение уравнения тренда, получаемое при подстановке в уравнение тренда номера прогнозируемого года t_m, однако параметры тренда, вычисленные по ограниченному периоду – это лишь выборочные оценки генеральных параметров. Прогноз должен иметь вероятностный характер, как любое суждение о будущем. Для этого вычисляется средняя ошибка прогноза положения тренда на год за номером t_m, обозначающая m_y, по формуле:

$$m_y = S(t) \cdot \sqrt{\frac{1}{N} + \frac{(t_m - \bar{t})^2}{\sum (t_i - \bar{t})^2}} \quad (9.12)$$

где N- число уровней исходного ряда;

t_m - номер прогнозируемого года;

S(t) – среднее квадратическое отклонение уровней от тренда.

$$S_{\bar{y}} = \sqrt{\frac{\sum_{1}^{n}\left(y_t - \hat{y}_t\right)^2}{n-p}} \qquad (9.13)$$

где n – число уровней; p – число параметров тренда;

$y_t, \hat{y}_t$ - соответственно фактические и расчетные значения уровней динамического ряда.

Для вычисления доверительного интервала прогноза положения тренда среднюю ошибку необходимо умножить на величину t – критерия Стьюдента, при имеющимся числе степеней свободы колебаний и при выбранной вероятности (надежности прогноза). Следовательно, доверительный интервал прогноза положения тренда вычисляется по формуле:

$$y_{точ} \pm t_{ст} \cdot m_y \qquad (9.14)$$

где $y_{точ}$ - точечный прогноз.

где t_α - доверительная величина (надежностью 95%) и (n-1)- степенями свободы.

Прогнозирование по тренду имеет качественное ограничение: оно допустимо в условиях сохранения основной тенденции.

Этап 6. После получения прогноза можно принимать управленческие решения в зависимости от полученных значений.

Прогнозирование по уравнению регрессии. Уравнения регрессии используются для:

1. *Для оценки хозяйственной деятельности*

Сравнение фактических уровней результативного признака с расчетными позволяет установить эффективность использования фактора *x* (средств) в конкретном объекте (хозяйстве) по сравнению со средней эффективностью использования фактора *x* по совокупности объектов.

Линия регрессии отражает изменение среднего значения результативного признака. Объекты, чьи фактические значения результата превышают теоретические, наиболее эффективно используют ресурсы. А объекты, чьи фактические значения результата меньше теоретических, имеют неиспользованные ресурсы повышения результата.

Например, при изучении зависимости между надоем на одну корову (ц) и затратами на одну корову (тыс. руб.)

2. *Для прогнозирования возможных значений результативного признака*

- авторегрессионное прогнозирование по тренду и колеблемости;

- факторное прогнозирование, основанное на изучении и количественном измерении взаимосвязи между признаками.

Основным условием прогнозирования на основании регрессионного уравнения является стабильность или, по крайней мере, малая изменчивость других факторов и условий изучаемого процесса, не связанных с ними. Если

резко изменится «внешняя среда» протекающего процесса, прежнее уравнение регрессии потеряет свое значение.

Прогнозирование по уравнению регрессии проводится в два этапа:

1. Вычисляется «точечный прогноз».
2. Определяется доверительный интервал с достаточно большой вероятностью (интервальный прогноз).

Точечный прогноз – это значение результативного показателя, получаемого при подстановке в уравнение регрессии ожидаемой величины факторного признака. Однако, нельзя подставлять значения факторного признака, значительно отличающихся от входящих в базисную информацию, по которой вычислено уравнение регрессии. При качественно иных уровнях фактора, если они даже возможны в принципе, были бы другими параметры уравнения.

Вероятность точной реализации такого прогноза крайне мала. Необходимо сопроводить прогноз *доверительным интервалом*.

Статистический прогноз с учетом доверительного интервала:

«Точечный прогноз» ± α,

где α – доверительный интервал,

$\alpha = m \cdot t$,

m – стандартная ошибка прогноза

$$m = S(y)\sqrt{1+\frac{1}{n}+\frac{(x_k-\bar{x})^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}}, \qquad (9.15)$$

m – стандартная ошибка положения линии регрессии в генеральной совокупности при $x=x_k$;

n – объем выборки;

x_k - ожидаемое значение фактора;

$S(y)$ – среднее квадратическое отклонение результативного признака от линии регрессии в генеральной совокупности с учетом степеней свободы вариации.

$$S(y) = \sqrt{\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{n-2}}. \qquad (9.16)$$

t критерий Стьюдента определяется по таблице, зависит от вероятности и числа степеней свободы $df = n - p$.

По уравнению множественной регрессии можно прогнозировать несколькими способами:

1) при подстановке в уравнение регрессии желаемых показателей;
2) с учетом прогнозных значений факторного признака.

<u>Пример 9.1. Прогнозирование на основе среднего абсолютного прироста и среднего темпа роста.</u>

Осуществим процесс прогнозирования денежных потоков ООО «ХХХ» на основе среднегодового темпа роста и среднегодового абсолютного прироста. В качестве объекта возьмем положительный, отрицательный и чистый денежные потоки. Прогнозирование денежных потоков на основе среднего абсолютного прироста осуществляется по следующей формуле (9.1).

Для определения перспективной величины денежных потоков воспользуемся методом на основе среднего темпа роста (9.2).

Для наших исходных данных средний абсолютный прирост и среднегодовой темп роста денежных потоков представлен в таблице 3.9.

Таблица 9.1 - Средний абсолютный прирост и среднегодовой темп роста денежных потоков ООО «ХХХ»

Показатели	Средний абсолютный прирост, тыс.руб.	Среднегодовой темп роста, %
Положительный денежный поток	13665885	1,91
Отрицательный денежный поток	13555897	1,91
Чистый денежный поток	109988	2,42

При подстановке исходных данных из таблицы 9.1 в формулы (6) и (7) получатся следующие прогнозные данные (таблица 9.2).

Таблица 9.2 – Прогнозные значения величины денежных потоков ООО «ХХХ» по моделям

Показатели	Прогнозный год	Прогноз по модели среднегодового абсолютного прироста, тыс.руб.	Прогноз по модели среднегодового темпа роста, тыс.руб.
Положительный денежный поток	2022	92329819	500355040
	2023	105995704	955233759
	2024	119661588	1823648133
Отрицательный денежный поток	2022	91624702	496832615
	2023	105180599	948509056
	2024	118736495	1810809923
Чистый денежный поток	2022	705117	3522424,85
	2023	815105	6724703,18
	2024	925093	12838210,8

На основе полученных прогнозных значений по модели среднегодового абсолютного прироста, при условии, что тенденция не измениться, величина денежных потоков организации продолжит увеличиваться и к 2024г. положительный денежный поток может составить 119661588 тыс.руб.,

отрицательный денежный поток - 118736495 тыс.руб. и чистый денежный поток – 925093 тыс.руб.

На основе полученных прогнозных значений по модели среднегодового темпа роста, при условии, что тенденция не измениться, величина денежных потоков организации также продолжит увеличиваться и к 2024г. положительный денежный поток может составить 1823648133 тыс.руб., отрицательный денежный поток - 1810809923 тыс.руб. и чистый денежный поток – 12838210,8 тыс.руб.

Пример 9.2. Прогнозирование на основе трендовых моделей.

Для расчета параметров уравнения регрессии воспользуемся табличным редактором MS Excel XP, результаты расчетов представим в таблице 9.3.

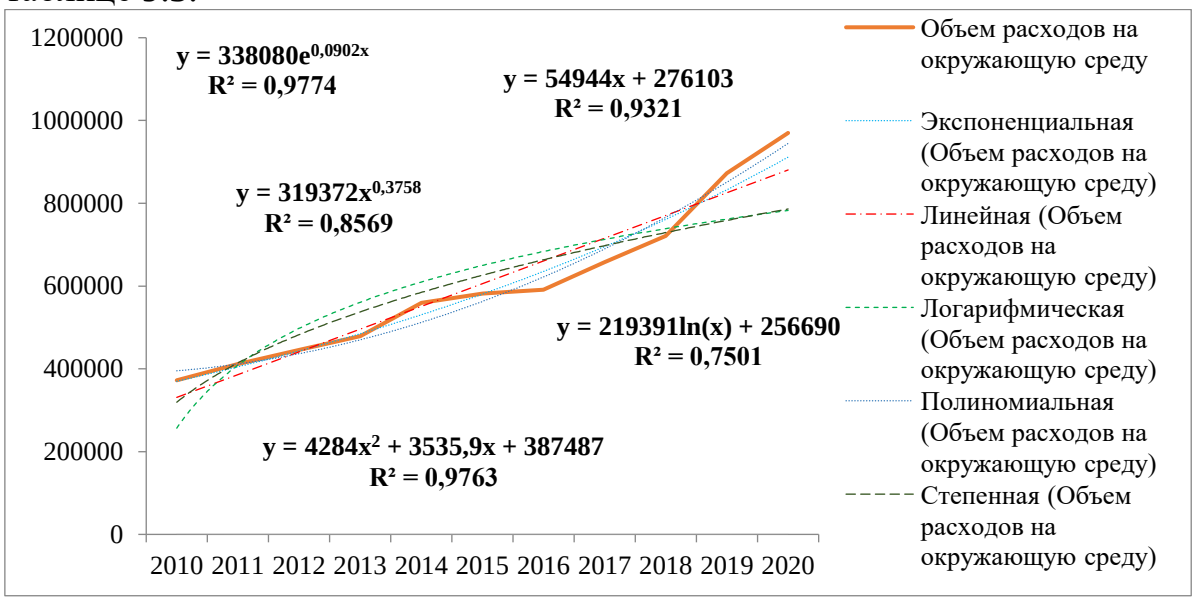


Рисунок 9.1 – Динамика объёмов расходов на охрану окружающей среды РФ.

То уравнение тренда, где коэффициент наибольший и наименьшая среднеквадратичная ошибка будет наилучшим вариантом.

Наиболее удачным является то уравнение, где R^2 наибольший, именно такой тренд и стоит выбрать.

Таблица 9.3 – Данные трендов величины объемов на охрану окружающей среды РФ

Форма тренда	Модель	R2
Линейный	y = 54944x + 27610	R² = 0,932
Логарифмический	y = 21939ln(x) + 25669	R² = 0,750
Полиномиальная	$y = 4284x^2 + 3535,x + 38748$	R² = 0,976
Степенная	$y=31937x^{0,375}$	R² = 0,856
Экспоненциальная	$y=33808e^{0,090x}$	R² = 0,977

С помощью уравнения тренда мы можем сделать точечный и интервальный прогноз.

Рисунок 9.2 – Доверительная граница прогнозных значений уровня объемов расходов на охрану окружающей среды РФ.

Осуществим прогноз по имеющимся данным и представим показатели в таблице. Согласно, прогнозу объем расходов сначала уменьшится, а потом возрастет.

Таблица 9.4 – Прогнозные значения величины объемов расходов на охрану окружающей среды Российской Федерации, тренды развития

Годы	Нижняя доверительная граница прогноза	Прогноз	Верхняя доверительная граница прогноза
2021	901897,3	935431	968964,7
2022	952397,3	990375	1028352,7
2023	1002786,7	104531	1087851,3

По исходным показателям мы можем сказать, что тренд в 2021 г. пройдет через точку с ординатой 935431руб., в 2022 г. – через точку 990375руб., а в 2023 г. – через точку 1045319руб.

Пример 9.3. Прогнозирование по уравнению регрессии.

С целью выявления влияния различных факторов на результативный показатель используется метод корреляционно- регрессионного анализа, суть которого заключается в выявлении зависимостей между показателями и моделировании на основе полученного уравнения регрессии прогнозных значений результата.

Выявим существенность влияния на уровень численности населения Российской Федерации таких социально-экономических факторов (приложение Д), как:

x_1 - уровень рождаемости, ‰;

x_2 - уровень смертности, ‰;

x_3 - уровень миграции населения, ‰;

x_4 - уровень безработицы, %;

x_5 – темп прироста уровня номинальной начисленной заработной платы к уровню предыдущего года, %.

Применение методов классической теории корреляции связано с определенными особенностями:

1) в рядах динамики зачастую наблюдается зависимость между последующими и предшествующими уровнями. Наличие такой связи в статистической литературе называют автокорреляцией. При изучении взаимосвязи между рядами динамики с применением методов корреляционно-регрессионного анализа автокорреляция должна быть исключена из каждого из сравниваемых рядов динамики;

2) в изменении уровней нескольких рядов динамики, как правило, существует лаг, т.е. смещение во времени по сравнению с изменением уровней другого ряда динамики. Для получения более правильной оценки степени тесноты корреляционной связи также необходимо исключить этот лаг, т.е. нужно сдвинуть уровни одного ряда относительно другого на некоторый промежуток времени;

3) условия формирования уровней рассматриваемых рядов, как правило, изменяются. Эти изменения могут быть и существенными. Соответственно может изменяться во времени и степень тесноты связи. В этих условиях речь идет о переменной корреляции.

Таким образом, при анализе корреляционной связи между рядами динамики необходимо:

1) измерить связь между предыдущими и последующими уровнями;

2) с учетом указанных выше особенностей изучить связь между рядами динамики.

Для устранения автокорреляции между факторными признаками при построении матрицы парных коэффициентов введем фактор времени t. Параметры модели, включающие фактор времени t определяются обычным МНК. Факторы включенные в модель не мультиколлинеарны, следовательно, могут быть совместно включены в модель.

Таблица 9.5 – Матрица парных коэффициентов корреляции

	y	x_1	x_2	x_3	x_4	x_5	t
y	1						
x_1	**0,856**	1					
x_2	**0,850**	0,123	1				
x_3	**0,713**	0,034	0,334	1			
x_4	0,343	0,118	0,007	0,223	1		
x_5	0,454	0,309	0,178	0,306	0,342	1	
t	-0,508	-0,313	-0,307	-0,192	-0,128	-0,112	1

На численность населения в большей степени оказывают влияние коэффициент рождаемости на 1000 человек (прямая, сильная связь), коэффициент смертности населения (прямая, сильная связь) и коэффициент миграции населения (прямая, умеренная связь). Показатель уровня номинальной начисленной заработной платы и уровень безработицы оказывают слабое влияние на численность населения, поэтому для построения уравнения множественной регрессии включим в модель факторы x_1, x_2, x_3. Также следует отметить, что численность населения в Российской Федерации зависит от фактора времени.

С использование Excel было получено уравнение регрессии. Результаты представлены в приложении Е.

По результатам регрессионного анализа получено следующее уравнение регрессии:

$$y = 594{,}45 + 20{,}40 \cdot x_1 - 14{,}43 \cdot x_2 + 2{,}07 \cdot x_3 - 1{,}12 \cdot t \qquad (9.18)$$

(8,7) (6,4) (-9,1) (5,8) (-11,1)

Анализ полученного уравнения позволяет сделать выводы о том, что при условии неизменной тенденции с ростом коэффициента рождаемости на 1000 человека населения - численность населения увеличивается на 20,40 человек, с ростом коэффициент смертности численность населения сокращается на 14,43 человека, а с ростом мигрантов, приходящихся на 1000 человек – численность населения Российской Федерации возрастает на 2,07 человека. Параметр при t равный -1,12 характеризует среднегодовую абсолютную убыль численности населения Российской Федерации под воздействием прочих факторов при условии неизменности факторов, включенных в модель.

Воспользуемся полученным множественным линейным регрессионным уравнением (9.18) проведем экстраполирование значений численности населения Российской Федерации при фиксированном значении факторов: максимальном, минимальном и среднем значениях факторных признаков. Получаем следующие результаты перспективного прогноза (таблица 9.6).

Таблица 9.6 - Прогнозные значения численности населения Российской Федерации при фиксированном значении факторных признаков множественной регрессионной модели

Вид прогноза	Нижняя доверительная граница α=0,05	Прогнозное значение	Верхняя доверительная граница α=0,05
Пессимистический	136,6	139,9	143,2
Реалистический	140,3	143,6	146,9
Оптимистический	146,5	149,8	153,1

Таким образом, на основе полученного уравнения множественной регрессии получаем следующие результаты: в Российской Федерации при

наименьших значениях факторных признаков, включенных в модель численность населения Российской Федерации может составить 139,9 млн.чел. и находится в интервале (136,6; 143,2) млн.чел.; при средних значениях факторных признаков численность населения Российской Федерации может составить 143,6 млн.чел. и находиться в интервале (140,3; 146,9) млн.чел.; при максимальных значениях факторных признаков численность населения Российской Федерации может составить 149,8 млн.чел. и находится в интервале (146,5; 153,1) млн.чел.

ЗАДАНИЕ ДЛЯ МАГИСТРАНТОВ ПО РЕЗУЛЬТАТАМ ИЗУЧЕНИЯ ДИСЦИПЛИНЫ

Результатом изучения и освоения курса «Статистические методы исследований в экономике» является написание научной статьи. Тематика статьи должна соответствовать теме магистерского исследования, в статье должны быть использованы статистические методы, изученные в рамках освоения дисциплины «Статистические методы исследований в экономике». Научная статья должна содержать не менее 4-5 статистических методов исследования, по результатам исследования должны быть даны полноценные выводы.

Требования к оформлению научной статьи:

Текст статьи должен быть объемом от 5 до 10 полных страниц, набранных в текстовом редакторе MS Word с расширением *.doc, *.rtf должен быть идентичным во всех вариантах. Шрифт Times New Roman Cyr, размер 14 пт. Межстрочный интервал – одинарный. Поля по 20мм. Размер бумаги – А4. Ориентация – книжная. Страницы не нумеровать. Выравнивание текста – по ширине страницы. Расстановка переносов – автоматическая. Список используемой литературы в конце работы. Обязательно ссылка на литературу по тексту. Сноски в конце работы. В тексте допускаются рисунки и таблицы. Цвет рисунков – черно-белый. Размер текста на рисунках не менее 11 пт., рисунки должны быть сгруппированы. Подрисуночные надписи и названия рисунков выполняется шрифтом Times New Roman Cyr 12 пт.

Пример 1.

Статистический анализ развития банковской системы Российской Федерации

Иванов И.И., Оренбургский филиал РЭУ им. Г.В.Плеханова, 1 курс, магистерская программа «Финансовая экономика»

Научный руководитель – Лаптева Е.В., к.э.н., доцент кафедры финансов и менеджмента Оренбургского филиала РЭУ им. Г.В.Плеханова

Одним из основных показателей, характеризующих банковскую систему, является величина банковских активов кредитных организации.

За последнее десятилетие наблюдается существенное изменение в динамическом ряду уровня банковских активов Российской Федерации. Проведем анализ динамики уровня банковских активов (рисунок 1).

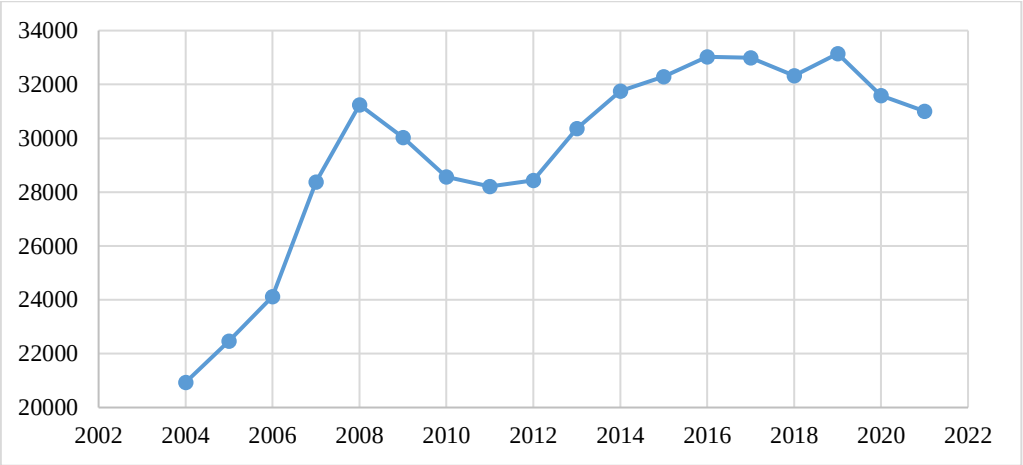


Рисунок 1 – Динамика величины банковских активов кредитных организаций Российской Федерации по состоянию на 1 января, млрд.руб.

Анализ скорости и интенсивности развития явления во времени осуществляется с помощью статистических показателей, которые получаются в результате сравнения уровней между собой (таблицы 1).

Таблица 1 – Динамика уровня величины банковских активов кредитных организаций по базисной системе

Годы	Абсолютный прирост (убыль), млрд.руб.	Темп роста, %	Темп прироста, %
2004	-	-	-
2005	1536,0	107,34	7,34
2006	3187,0	115,22	15,22
2007	7450,0	135,59	35,59
2008	10311,0	149,26	49,26
2009	9100,0	143,47	43,47
2010	7637,0	136,48	36,48
2011	7277,0	134,76	34,76
2012	7506,0	135,86	35,86
2013	9435,0	145,07	45,07
2014	10822,0	151,70	51,70
2015	11360,0	154,27	54,27
2016	12098,0	157,79	57,79
2017	12058,0	157,60	57,60
2018	11388,0	154,40	54,40
2019	12212,0	158,34	58,34
2020	10650,0	150,88	50,88
2021	10067,0	148,09	48,09

Из таблицы 1 видно, что в период с 2004 по 2021 г. наблюдается увеличение банковских активов кредитных организации РФ. По сравнению с 2004г. все последующие годы характеризуются приростом. Наибольший прирост в величине активов приходился на 2016 г. (57,79%).

Наряду с темпом роста можно рассчитать показатель темпа прироста, характеризующий относительную скорость изменения уровня ряда в единицу времени. Темп прироста показывает, на какую долю уровень данного периода или момента времени больше (или меньше) базисного уровня.

В статистической практике часто вместо расчета и анализа темпов роста и прироста рассматривают абсолютное значение одного процента прироста. Оно представляет собой одну часть базисного уровня и в то же время – отношение абсолютного прироста к соответствующему темпу роста. Этот показатель дает возможность установить, насколько в среднем за единицу времени должен увеличиваться уровень ряда, чтобы, отправляясь от начального уровня за данное число периодов, достигнуть конечного уровня. Расчет этого показателя имеет экономический смысл только по цепной системе.

Таблица 2 – Динамика уровня величины банковских активов кредитных организаций по цепной системе

Годы	Абсолютный прирост (убыль), млрд.руб.	Темп роста, %	Темп прироста, %	Абсолютное значение 1% прироста
2004	-	-	-	-
2005	1536,0	107,34	7,34	209,33
2006	1651,0	107,35	7,35	224,69
2007	4263,0	117,67	17,67	241,2
2008	2861,0	110,08	10,08	283,83
2009	-1211,0	96,12	-3,88	312,44
2010	-1463,0	95,13	-4,87	300,33
2011	-360,0	98,74	-1,26	285,7
2012	229,0	100,81	0,81	282,1
2013	1929,0	106,78	6,78	284,39
2014	1387,0	104,57	4,57	303,68
2015	538,0	101,69	1,69	317,55
2016	738,0	102,29	2,29	322,93
2017	-40,0	99,88	-0,12	330,31
2018	-670,0	97,97	-2,03	329,91
2019	824,0	102,55	2,55	323,21
2020	-1562,0	95,29	-4,71	331,45
2021	-583,0	98,15	-1,85	315,83

Из таблицы 2 видно, что 2004 г. и 2021 г., характеризующиеся финансовым кризисом, отмечен резким спадом величины банковских активов на 4,87% и 4,71%, соответственно.

Особое внимание следует уделять методам расчета средних показателей рядов динамики, которые являются обобщающей характеристикой его абсолютных уровней, абсолютной скорости и интенсивности изменения уровней ряда динамики. Различают следующие показатели динамики: средний уровень ряда динамики, средний абсолютный прирост, средний темп роста и прироста.

В интервальном ряду динамики с равноотстоящими уровнями во времени расчет среднего уровня ряда (у) производится по формуле средней арифметической простой:

$$y=\frac{\sum y}{n}=31228{,}7 \text{ млрд.руб.}$$

Определение среднего абсолютного прироста производится по цепным абсолютным приростам по формуле:

$$\Delta=\frac{\sum \Delta_{ц}}{n-1}=626{,}9 \text{ млрд.руб.}$$

Среднегодовой темп роста вычисляется по формуле:

$$Тр=\sqrt[n-1]{\frac{y_n}{y_0}}=1{,}103 \text{ или } 110{,}3\%$$

Среднегодовой темп прироста получим, вычтя из среднего темпа роста 100%.

$$Тпр=Тр-100=10{,}3\%$$

На основе рассчитанных показателей динамики можно сказать, что за период с 2004 по 2021 гг. в Российской Федерации наблюдалось увеличение величины банковских активов кредитных организаций на 626,9 млрд.руб. или на 10,3%.

Средний уровень величины банковских активов за 17 лет составил 31228,7 млрд.руб.

Прежде чем переходить к определению тенденции и выделению тренда необходимо выяснить существует ли тенденция вообще динамике величины банковских активов кредитных организаций Российской Федерации. Для этого можно воспользоваться наиболее часто используемым на практике методом проверки наличия тренда – критерием серий.

Считаем медиану исходного ряда Ме=30684 млрд.руб.

Число серий определяется путем подсчета: $\upsilon(17) = 4$.

Определяем протяженность самой длинной серии $\tau \max (17) = 8$.

Проверяем гипотезу о случайности исходного ряда, для этого должны выполняться неравенства (1) и (2). Получаем: проверка гипотезы основывается на проверки гипотезы о случайности ряда, поэтому для того чтобы не была отвергнута гипотеза о случайности исходного ряда должны выполнятся следующие неравенства:

$$\nu(n)\rangle\left[\frac{1}{2}(n+1-1{,}96\sqrt{n-1})\right] \quad (1)$$

$$\tau_{\max}(n)\langle[1{,}43\ln(n+1)] \quad (2)$$

Данные неравенства не выполняются, следовательно, гипотеза о случайности исходного ряда отклоняется, значит, тенденция в уровне величины банковских активов кредитных организаций в Российской Федерации имеется.

$\upsilon(17)> 5{,}4$ и $\tau \max (17)>4{,}2$

Можно предположить, что наиболее адекватно описывать тенденцию уровня величины банковских активов кредитных организаций Российской Федерации имеющихся данных может модель степенной, экспоненты или полинома второго порядка, приведем данные тренды на рисунке 2.

Для расчета параметров уравнения регрессии воспользуемся табличным редактором MS Excel XP, результаты расчетов представим в таблице 3.

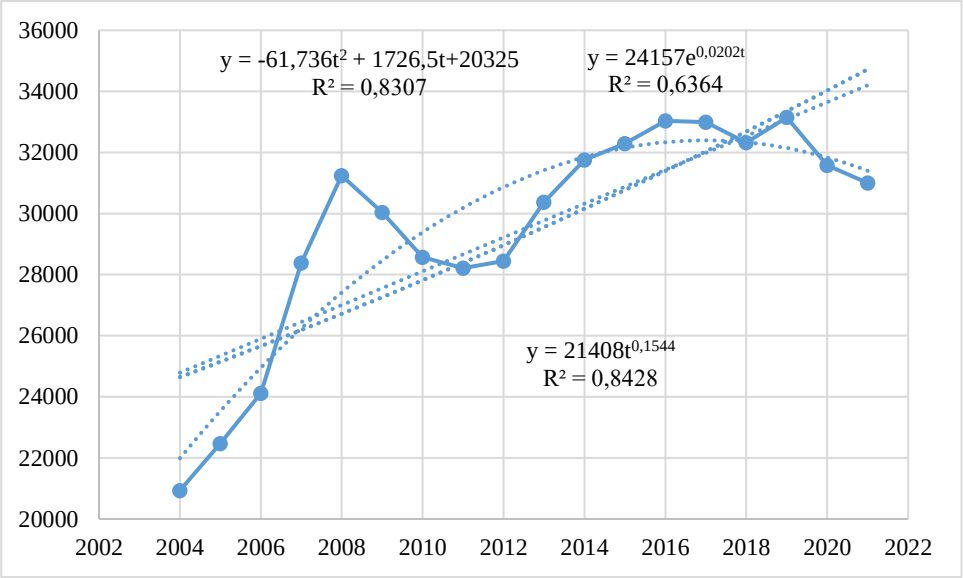


Рисунок 2 – Динамика величины банковских активов кредитных организаций Российской Федерации, тренды развития

Для определения наилучшего уравнения тренда следует обратить внимание на наибольший коэффициент аппроксимации и наименьшую среднеквадратическую ошибку.

Оценку надежности уравнения регрессии в целом дает R^2, в результате расчетов в случае параболы значение данного показателя выше, чем у прямой. Именно такой тренд будем использовать для приятия решений и прогнозирования.

Таблица 3 – Характеристики трендов развития уровня величины банковских активов кредитных организаций Российской Федерации

Форма тренда	Модель	R^2	$F_{факт}$	$F_{табл}$	Среднеквадратическая ошибка
Степенная	$\tilde{y}_t = 21408t^{0,1544}$ (8,4) (1,3)	0,8428	33,77	3,94	282,62
Парабола второго порядка	$\tilde{y}_t = -61,736t^2 - 1726,5t + 20325$ (-5,6) (4,7) (6,1)	0,8307	37,89	3,09	269,61
Экспонента	$\tilde{y}_t = 21408e^{0,0202t}$ (3,4) (1,2)	0,6364	42,56	3,94	236,88

Все полученные модели статистически значимы и пригодны для принятия решений. Степенной тренд значим по F-критерию Фишера, но параметр уравнения a_1 получен незначим, т.к. значение t-критерия Стьюдента получено

очень маленьким, поэтому данная модель может быть использована для дальнейших расчетов, но не пригодна для прогнозирования, также и экспоненциальный тренд имеет один не значимый параметр. Параболический тренд получен, значим по F-критерию Фишера, все параметры значимы по t-критерию Стьюдента, следовательно, в дальнейших исследованиях будем использовать именно его.

Проверяем полученную модель развития на адекватность с помощью критерия Дарбина-Уотсона. Для этого необходимо воспользоваться ППП Statistica 6.0.

Критерий Дарбина – Уотсона d=1,05. Определив по специальным таблицам значения верхней и нижней доверительных границ критерия Дарбина – Уотсона, можно принять решение об отсутствии или наличии автокорреляции между соседними остаточными членами. $d_{1=}0,08$ и $d_{2=}1,36$. Так как $d>d_1$, и $d<d_2$ следовательно, нет достаточных оснований для принятия решения.

Определяем среднюю относительную ошибку прогноза по модулю. Получаем:

$$\left|\bar{\delta}\right| = \frac{1}{n} \cdot \frac{\sum_{t=1}^{n} \left|\tilde{y}_t - y_t\right|}{y_t} \cdot 100\% = 1,45\% \qquad (3)$$

|δ|<10% это говорит о высокой точности модели.

Для характеристики модели рассмотрим нормальный вероятностный

Рассмотренные критерии значимости доказывают статистическую значимость полученной параболе и ее соответствие (адекватность) динамики уровня величины банковских активов кредитных организаций Российской Федерации, поэтому данная модель может быть использована для построения прогноза на период 2022-2024 гг.

Корреляционный анализ, разработанный К. Пирсоном и Дж. Юлом, является одним из методов статистического анализа взаимозависимости нескольких признаков. Основная задача корреляционного анализа состоит в оценке природы взаимозависимости между наблюдаемыми переменными, дополнительная задача (являющаяся основной в регрессионном анализе) состоит в оценке уравнений регрессии, где в качестве результативного признака выступает признак, являющийся следствием других признаков (факторов) – причин.

На уровень величины банковских активов кредитных организаций влияет большое количество факторов. Попробуем изучить взаимосвязь величины уровня величины банковских активов и других экономических явлений, происходящих в Российской Федерации. В корреляционно-регрессионном анализе можно устранить воздействие какого-либо фактора, если зафиксировать воздействие этого фактора на результат и другие, включенные в модель факторы. Данный прием широко применяется в анализе временных рядов, когда тенденция фиксируется через включение фактора времени в модель в качестве независимой переменной.

Для проведения корреляционно-регрессионного анализа используем следующие факторные признаки:

У - банковские активы кредитных организаций, млрд.руб.;

Х1- кредиты и прочие размещенные средства, предоставленные нефинансовым предприятиям и организациям, млрд.руб.;

Х2 – депозиты и прочие привлеченные средства физических лиц, млрд.руб.;

Х3 – число страховых организаций ед.;

Х4 – задолженность по кредитам, предоставленным кредитными организациями физическим лицам, млрд.руб.;

Х5 – задолженность по кредитам, предоставленным кредитными организациями юридическим лицам, млрд.руб.;

Х6 –доля кредитных организаций общего численности кредитных организаций;

Х7 – страховые выплаты, млн.руб.;

Х8 – инвестиции в основной капитал, млн.руб.

Параметры модели с включением фактора времени оцениваются с помощью обычного метода наименьших квадратов (МНК).

С помощью ПК получаем матрицу парных коэффициентов, на основании которых необходимо сделать вывод о факторах, которые могут быть включены в модель множественной регрессии (таблица 4). Корреляционная матрица получена с помощью табличного редактора Excel XP в пакете анализа.

Таблица 4 – Корреляционная матрица влияния факторов на уровень величины банковских активов кредитных организаций Российской Федерации

	y	X1	X2	X3	X4	X5	X6	X7	X8
У	1								
X1	0,784196	1							
X2	0,740948	0,02513	1						
X3	0,170823	0,20517	0,0347	1					
X4	-0,87755	-0,16364	-0,22623	-0,21985	1				
X5	-0,15388	-0,27971	-0,35876	0,479488	-0,20222	1			
X6	-0,11082	-0,74114	-0,51858	0,163678	-0,97561	0,446238	1		
X7	-0,3984	0,115731	-0,12415	-0,15768	0,78703	-0,42972	-0,9336	1	
X8	-0,79397	-0,7641	-0,84743	0,038866	0,410725	0,532786	0,816379	-0,01735	1

Из корреляционной матрицы видна достаточно сильная взаимосвязь между результативным (у) и факторными признаками (х1, х2, х4,t). Связь очень сильная.

Проведем регрессионный анализ. По результатам регрессионного анализа получено следующее уравнение регрессии:

$$y = -17383{,}73 + 59{,}13 \cdot x_1 + 272{,}3 \cdot x_2 - 5{,}55 \cdot x_4 + 171{,}3 \cdot t \qquad (4)$$

(-2,36) (3,69) (3,43) (-2,83) (2,70)

В скобках указаны значения t-критерия Стьюдента.

В результате построения уравнения регрессии получили следующие результаты (таблица 5).

Таблица 5 – Результаты построения регрессии

Показатели	Значения
Коэффициент корреляции R	0,910
Коэффициент детерминации R^2	0,829
Скорректированный коэффициент детерминации R^2	0,773
Фактическое значении F-критерия Фишера	14,61
Табличное значении F-критерия Фишера	2,79
Стандартная ошибка	5,11

Множественный коэффициент регрессии равен 0,910. Это свидетельствует о высокой связи между признаками. Коэффициент детерминации – равен 0,829, следовательно, 82,9% вариации уровня величины банковских активов кредитных организаций Российской Федерации обусловлено факторами, включенными в модель (6).

Анализ полученного уравнения позволяет сделать выводы о том, что с ростом кредитов и прочих размещений, предоставленных нефинансовым предприятиям и организациям на 1 млрд.руб. – банковские активы кредитных организаций увеличиваются на 59,13 млрд.руб.; с ростом депозитов и прочих привлеченных средств физических лиц – банковские активы кредитных организаций увеличиваются на 272,3 млрд.руб.; ростом задолженности по кредитам влечет уменьшение величины банковских активов кредитных организаций на 5, 55 млрд.руб. Параметр при t равный 171,3 характеризует среднегодовую абсолютный прирост величины банковских активов кредитных организаций Российской Федерации под воздействием прочих факторов при условии неизменности факторов, включенных в модель.

Проверка адекватности модели, построенной на основе уравнений регрессии, начинается с проверки значимости каждого коэффициента регрессии. Значимость коэффициента регрессии осуществляется с помощью t-критерия Стьюдента:

$$t_{расч} = \frac{|a_i|}{\sigma_{a_i}} \qquad (5)$$

Параметры уравнения все значимы, кроме параметра при факторе времени, так как их расчетные значения меньше табличных ($t_{табличное} = 2{,}02$, уровень значимости = 0,05, $t_{расч} \rangle t_{таблич}$)

Проверка адекватности всей модели осуществляется с помощью расчета F-критерия. Если $F_р > F_т$ при α=0,05, то модель в целом адекватна изучаемому явлению.

$$F_{расч} = 14{,}61 \quad F_{табл} = 2{,}79 \quad уровень\ значимости = 0{,}05 \quad F_{расч} \rangle\ F_{табл}$$

Следовательно, построенная модель на основе её проверки по F-критерию Фишера в целом адекватна, и все коэффициенты регрессии

значимы. Такая модель может быть использована для принятия решений и осуществления прогнозов.

Осуществим прогноз по имеющемуся уравнению тренда.

Рисунок 3 - Доверительная граница прогнозных значений уровня банковских активов кредитных организаций Российской Федерации

Представим прогнозные значения в таблице. Согласно, прогнозу уровень величины банковских активов будет снижаться.

Таблица 6 – Прогнозные значения величины банковских активов кредитных организаций Российской Федерации по уравнению тренда, млрд.руб.

Годы	Нижняя доверительная граница прогноза	Прогноз	Верхняя доверительная граница прогноза
2022	29034,5	31399,5	33764,5
2023	28476,8	30841,8	33206,8
2024	27795,6	30160,6	32525,6

Это означает, что тренд в 2022 г. пройдет через точку с ординатой 31399,5 млрд.руб., в 2023 г. – через точку 30841,8 млрд.руб., а в 2024 г. – через точку 30160,6 млрд.руб.. Однако параметры тренда, вычисленные по ограниченному периоду, - это лишь выборочные оценки генеральных параметров (рисунок 3). На рисунке представлена верхняя и нижняя доверительные границы прогноза.

Осуществим процесс прогнозирования по множественному уравнению регрессии (таблица 6).

Таблица 6 – Прогнозные значения величины банковских активов кредитных организаций в Российской Федерации по множественному уравнению регрессии

Прогнозы	Нижняя доверительная граница прогноза	Прогноз	Верхняя доверительная граница прогноза
Пессимистический	39845,4	38611,4	41079,4
Реалистический	31319,91	30085,91	32553,91
Оптимистический	23151,11	21917,11	24385,11

Таким образом, при среднем значении факторов, включенных в модель уровень величины банковских активов кредитных организаций при неизменности имеющейся тенденции может составить 31319,91млрд.руб. и находиться в интервале (30085,91; 32553,91) млрд.руб. При минимальных значениях факторов уровень величины банковских активов может составить 23151,11 млрд.руб. и принадлежать промежутку (21917,11; 24385,11) млрд.руб. При максимальных значениях – величина кредитных организаций может составить 39845,4 млрд.руб. и будет находиться в интервале (38611,4; 41079,4) млрд.руб.

Подводя итог проделанной работы, можно определить основные направления государственного регулирования банковской деятельности, которое осуществляется путем:

- установления обязательных требований к созданию кредитных организаций;
- лицензирования деятельности кредитных организаций;
- установления правил осуществления банковской деятельности (банковских правил);
- установления обязательных требований к деятельности кредитных организаций (банковское регулирование);
- надзора за соблюдением кредитными организациями требований законодательства о банковской деятельности и банковских правил (банковский надзор), в том числе принятия своевременных мер по приостановлению и (или) прекращению проведения небезопасной или неосновательной практики;
- создания системы защиты прав потребителей на рынке банковских услуг;
- запрещения и пресечения деятельности лиц, осуществляющих банковскую деятельность без соответствующей лицензии.

Литература:

1. Еремеева Н.С., Лебедева Т.В. Эконометрика: учебн. пособие для вузов. – Оренбург: ОАО «ИПК «Южный Урал», 2010. – 296 с.
2. Салин В.Н., Левит Б.Ю. Проверка значимости статистических показателей с помощью таблиц их критических значений. // Вопросы статистики. – 2009. - №9. – с. 69-77.

Пример 2.

Классификация городов и районов Оренбургской области по уровню инвестиций в строительство

Иванов И.И., Оренбургский филиал РЭУ им.Г.В.Плеханова, 1 курс, магистерская программа «Финансовая экономика»

Научный руководитель – Лаптева Е.В., к.э.н., доцент кафедры финансов и менеджмента Оренбургского филиала РЭУ им.Г.В.Плеханова

Привлечение инвестиций в реальный сектор экономики – вопрос ее выживания. Будут инвестиции – будет развитие реального сектора, а, следовательно, будет и экономический подъем. Не удастся привлечь их – неминуемо умирание производств, деградация экономики, обнищание страны, социальные взрывы и прочие сопутствующие явления.

Любое, даже самое незначительное, повышение инвестиционной привлекательности регионов – это дополнительные средства, позволяющие сделать шаг к выходу из кризиса. Но разовое привлечение инвестиций малоэффективно, после этого инвестиционная привлекательность остается статичной величиной, хотя и несколько более высокой. Спасти положение дел может лишь динамичное устойчивое движение, а не отдельные шаги. Только в этом случае отдельные порции инвестиций могут превратиться в постоянные. Осуществить это возможно, лишь управляя процессом повышения инвестиционной привлекательности регионов и правильно оценив региональный потенциал.

Значение статистического анализа для планирования и осуществления инвестиционной деятельности региона трудно переоценить. При этом особую важность имеет предварительный анализ, который проводится на стадии разработки инвестиционных проектов и способствует принятию разумных и обоснованных управленческих решений.

Прежде чем переходить к процедуре кластеризации необходимо исследовать вариацию городов и районов области по уровню инвестиций в строительство.

Для измерения вариации признака используют как абсолютные, так и относительные показатели.

К абсолютным показателям вариации относят: размах вариации, среднее линейное отклонение, среднее квадратическое отклонение, дисперсию. К относительным показателям вариации относят: коэффициент осцилляции, линейный коэффициент вариации, относительное линейное отклонение и др.

Рассчитаем показатели вариации по районам и городам Оренбургской области за 2021г.

Наибольший вклад в уровень инвестиций в строительство в Оренбургской области за 2021г. внес Абдулинский район 2270,3 тыс.руб. на.

Самый низкий уровень инвестиций в строительство в г.Кувандыке – 992,4 тыс.руб.

Размах вариации (размах колебаний) - важный показатель колеблемости признака, но он дает возможность увидеть только крайние отклонения, что ограничивает область его применения (1).

$$R=x_{max}- x_{min} \quad (1)$$

R =1277, 9 тыс.руб.

Различие между максимальной и минимальной величиной инвестиции в строительство по районам и городам Оренбургской области составляет 1278 тыс.руб.

Среднее значение уровня инвестиций в строительство по области составляет 1526,1 тыс.руб.

Для более точной характеристики вариации признака на основе учета его колеблемости используются другие показатели.

Среднее линейное отклонение d**,** которое вычисляют для того, чтобы учесть различия всех единиц исследуемой совокупности. Эта величина определяется как средняя арифметическая из абсолютных значений отклонений от средней. Так как сумма отклонений значений признака от средней величины равна нулю, то все отклонения берутся по модулю (2).

$$\bar{d} = \frac{\sum_{1}^{n} |x_i - \bar{x}|}{n} \quad (2)$$

$$\bar{d} = \frac{10394}{47} = 221{,}2 \text{ тыс.руб.}$$

В среднем отклонение значений уровня инвестиций в строительство по районам и городам от средней величины по Оренбургской области составляет 221,2 тыс.руб.

Обобщающие показатели, найденные с использованием вторых степеней отклонений, получили очень широкое распространение. К таким показателям относится среднее квадратическое отклонение σ (3).

$$\sigma = \sqrt{\frac{\sum_{1}^{n} (x - \bar{x})^2}{n}} \quad (3)$$

$$\sigma = \sqrt{\frac{3467914}{47}} = 271, \text{ тыс.руб.}$$

В среднем по Оренбургской области уровня инвестиций в строительство по районам и городам отклоняется от средней величины на 271,6 тыс.руб.

Рассчитаем относительные показатели вариации.

Коэффициент осцилляции: $V_R = \frac{R}{\bar{x}} * 100\%$ (4)

$$V_R = \frac{1277{,}9}{1526{,}1} * 100\% = 83{,}74\%$$

В среднем отклонение уровня инвестиций в строительство в Оренбургской области от средней величины составляет 16,26%.

Линейный коэффициент вариации: $V_{\bar{d}} = \frac{\bar{d}}{\bar{x}} * 100\%$ (5)

$$V_{\bar{d}} = \frac{221,2}{1526,1} * 100\% = 14,49\%$$

Значение данного показателя свидетельствует о том, что вариация признака достаточно большая и составляет 14,49% от среднего уровня инвестиций в строительство по районам.

Наиболее часто в практических расчетах применяется показатель относительной вариации – коэффициент вариации.

Коэффициент вариации: $V_{\sigma} = \frac{\sigma}{\bar{x}} * 100\%$ (6)

$$V_{\sigma} = \frac{271,6}{1526,1} * 100\% = 17,79\%$$

Следовательно, индивидуальные значения отличаются в среднем от средней арифметической на 1526,1 тыс.руб., или на 17,79%.

По рассчитанным показателям можем сделать вывод о том, что совокупность однородна по уровня инвестиций в строительство, так как рассчитанные коэффициенты не превышают 33% (совокупность считается однородной, если коэффициент вариации не превышает 33% (для распределений, близких к нормальному)).

Для более наглядного представления построим график уровня инвестиций в строительство по городам и районам Оренбургской области (рисунок 1).

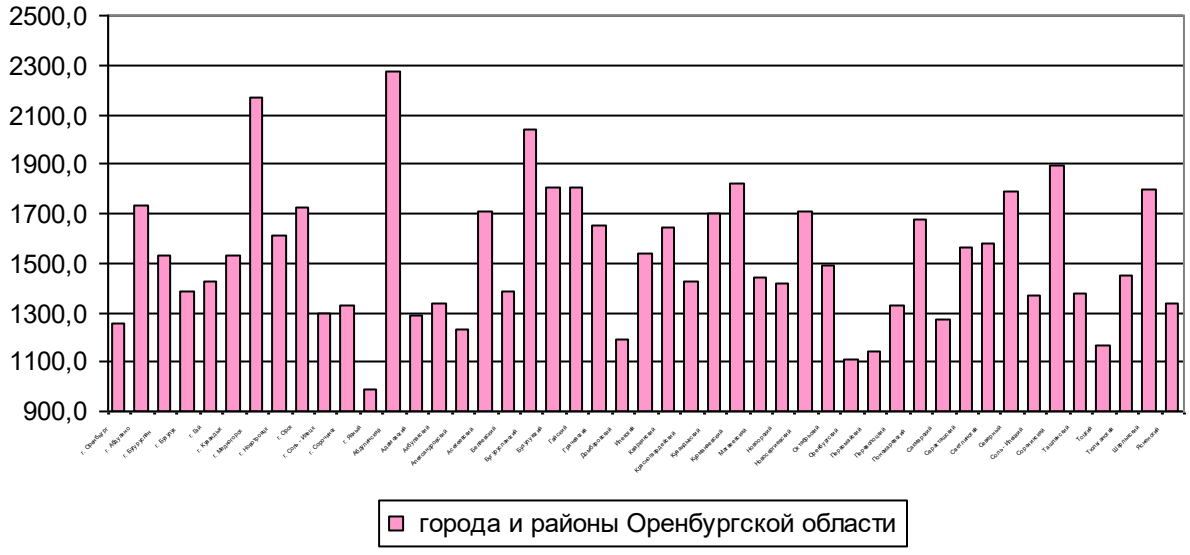


Рисунок 1 – Уровень инвестиций в строительство по районам и городам Оренбургской области в 2021г.

Графический анализ уровня инвестиций в строительство по районам Оренбургской области свидетельствует о том, что в 2021 г. наибольший уровень инвестиций в строительство наблюдалась в Абдулинском районе (2270,3тыс.руб.), чуть ниже в Бугурусланском районе (2041,7тыс.руб.); самый низкий уровень инвестиций в строительство в г.Кувандыке (992,4тыс.руб.).

При изучении вариации применяются также и такие характеристики вариационного ряда, которые описывают его структуру и строение. Такими характеристиками являются мода - Мо (наиболее часто встречающееся значение признака) и медиана – Ме (вариант середины ранжированного ряда).

Середина ранжированного ряда выпадает на Домбаровский и Илекский районы, следовательно, медиана (Ме) равна 1366 тыс.руб. Мода (Мо) равна 2270,3 тыс.руб., т.к. это самый высокий уровня инвестиций в строительство по Оренбургской области.

Как видно из рисунка 1 распределение уровня инвестиций в строительство по Оренбургской области несимметрично, поэтому определим показатель асимметрии (7).

$$A_s = \frac{\bar{x} - Mo}{\sigma} \tag{7}$$

$$A_s = \frac{1526{,}1 - 2270{,}3}{271{,}6} = -2{,}73$$

Перед полученным числом стоит знак «-», следовательно, асимметрия левосторонняя, значительная.

Чтобы оценить степень существенности асимметрии в нашем исследовании вариации уровня инвестиций в строительство по Оренбургской области, определим среднюю квадратическую ошибку по формуле:

$$\sigma_{A_s} = \sqrt{\frac{6(n-1)}{(n+1)(n+3)}} \tag{8}$$

$$\sigma_{A_s} = \sqrt{\frac{78}{255}} = 0{,}56$$

Средняя квадратическая ошибка составила 0,56.

$\frac{|A_s|}{\sigma_{A_s}} = \frac{2{,}73}{0{,}56} = 4{,}875 > 3$, значит, асимметрия существенна и распределение уровня инвестиций в строительство в Оренбургской области не является симметричным.

Таким образом, на основании проведенного анализа вариации уровня инвестиций в строительство в Оренбургской области, можно сделать вывод, что уровня инвестиций в строительство достаточно не симметричен (совокупность не однородна) и существует большой разброс уровня инвестиций в строительство, связанный как с социально-экономическими факторами, так и зависящий от региональной политики, проводимой в каждом районе в зависимости от специфики организационной структуры.

Выявим наличие групп по городам и районам Оренбургской области на основе многомерных статистических методов.

В качестве основного метода выявления групп районов по уровню инвестиций в строительство Оренбургской области используем один из многомерных методов статистического анализ – кластерный анализ.

Кластерный анализ - это совокупность методов, позволяющих классифицировать многомерные наблюдения, каждое из которых описывается набором исходных переменных $X_1, X_2, ..., Xm$.

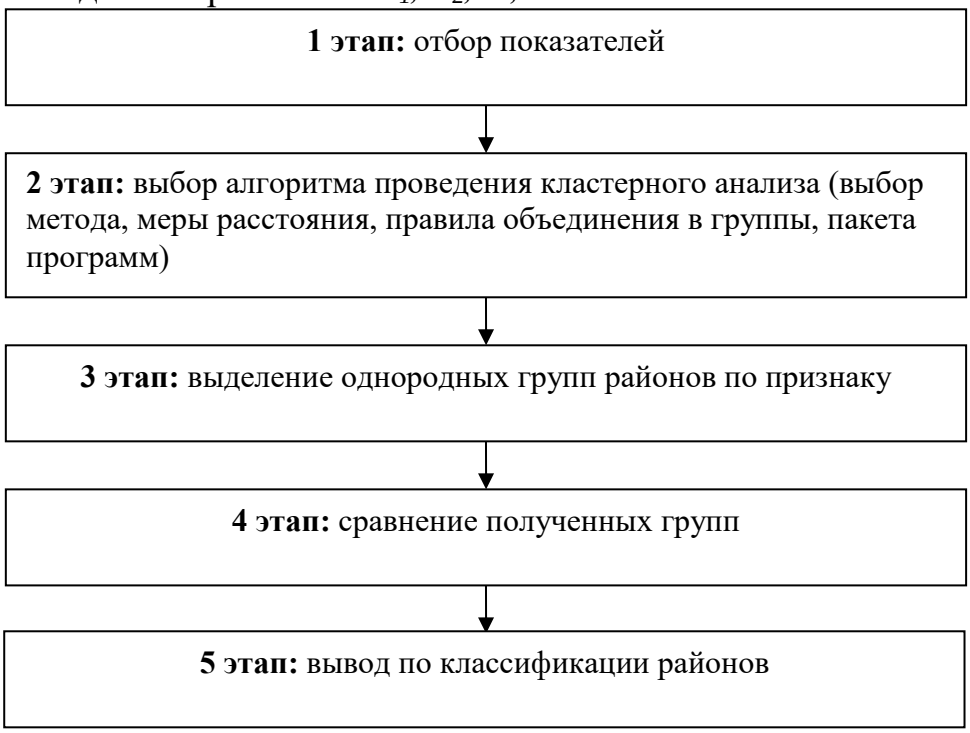


Рисунок 2 – Этапы анализа уровня инвестиций в строительство районов Оренбургской области

При проведении кластер-процедуры в качестве объектов анализа будут выступать 35 районов и 12 городов Оренбургской области. В качестве показателей уровня инвестиций в строительство можно использовать следующие показатели:

X1 – количество строительных организаций;

X2 – среднесписочная численность работников строительных организаций, тыс.тыс.руб.;

X3 – объем работ, выполненных по виду экономической деятельности «Строительство», млн.руб.;

X4 – уровень инвестиций в строительство, млн.руб.

В настоящее время для проведения кластер-процедур наибольшую популярность приобрели такие специализированные программные продукты как STATISTICA 6.0, SPSS 12.0, STATA 8 и другие. Обратимся к одному из перечисленных программных продуктов - статистическому пакету программ STATISTICA 6.0 и проведем разбиение имеющейся совокупности районов Оренбургской области на однородные группы по уровню инвестиций в строительство.

Нам представляется, что для наших целей классификации и построения типологии районов последующим статистическим анализом исследуемых показателей внутри каждого класса из перечисленных методов, представленных в пакете Statistica 6.0 в наибольшей степени отвечает *Ward's*

method (метод Варда), так как данный метод позволяет получать наиболее однородные в статистическом смысле кластеры.

В результате проведения кластерного анализа для 35 районов и 12 городов Оренбургской области методом древовидной кластеризации, в статистическом пакете Statistica 6.0 были получены следующие результаты, приведенные на рисунке 3.

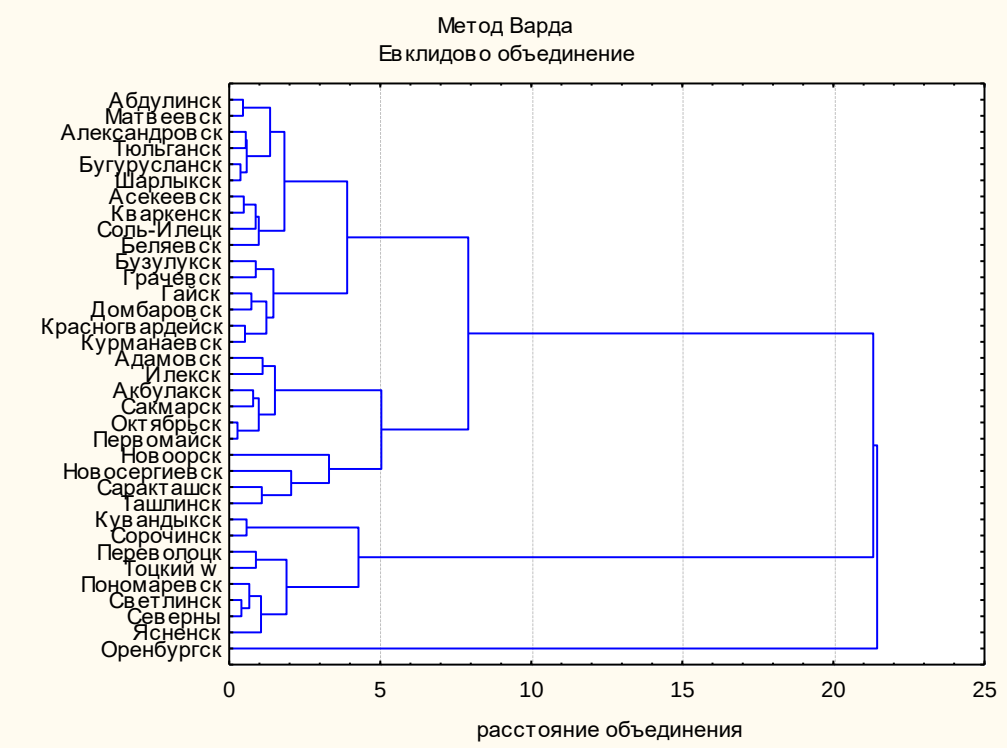


Рисунок 3 – Вертикальная древовидная диаграмма уровня инвестиций в строительство в городах и районах Оренбургской области в 2021г.

Согласно результатам, приведенным на рисунке 3, получаем три кластера которые характеризуются следующими показателями:

Таблица 1 – Характеристика кластеров районов области по уровню инвестиций в строительство

Показатели по группе	1 кластер	2 кластер	3 кластер
Число районов в группе	6	23	8
Среднее количество строительных организаций	48	51	23
Среднесписочная численность работников строительных организаций, тыс.тыс.руб.	14,9	33,9	14,6
Средний объем работ, выполненных по виду экономической деятельности «Строительство», млн.руб.	15,5	71,5	17,7
Уровень инвестиций в строительство, млн.руб.	98,7	76,4	34,9

Согласно данным, приведенным в таблице 1, первый кластер при малой численности характеризуются достаточно высокими показателями. Во второй кластер вошли районы со средними характеристиками, их можно отнести к районам со средним уровнем инвестиций в строительство. У районов третьего кластера самый низкий уровень инвестиций.

График средних значений по кластерам представлен на рисунке 4.

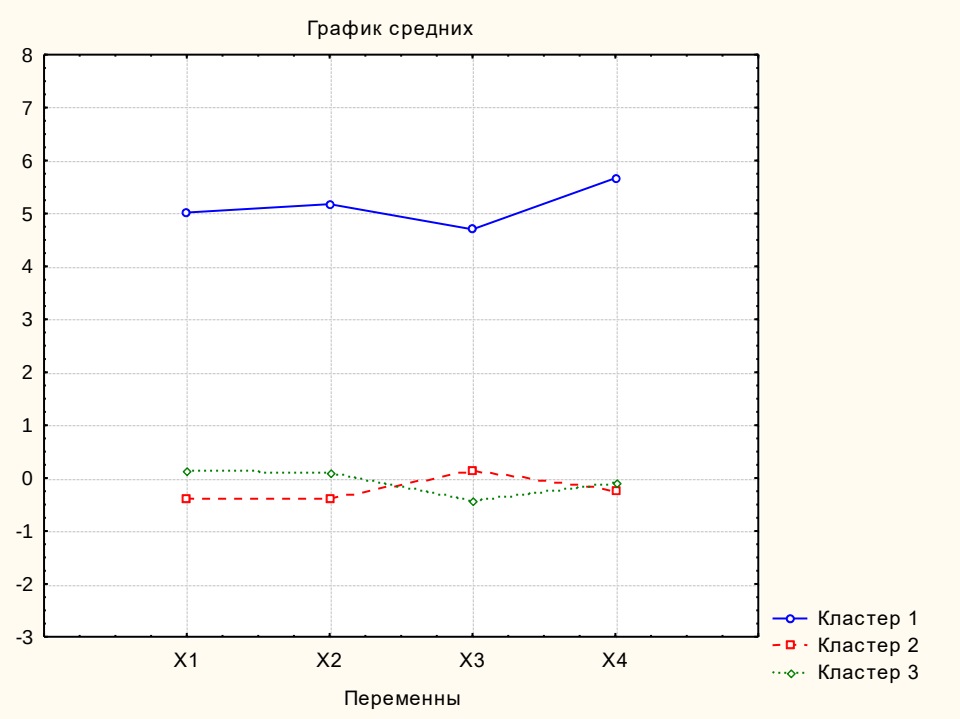


Рисунок 4 – График средних значений для кластеров

Состав кластеров и группы районов представлены в таблице 2.

Таблица 2 – Результаты кластеризации районов Оренбургской области по уровню инвестиций в строительство.

№ кластера	Количество субъектов	Состав кластера
1	6	Грачевский,Октябрьский, , Тюльганский, Пономаревский, Красногвардейский, Матвеевский
2	23	Абдулинский, Адамовский, Александровский, Алексеевский, Бугурусланский, Бузулукский, Илекский, Кувандыкский, Курманаеский, Новоорский, Новосергиевский, Оренбургский, Саракташский Первомайский, Переволоцкий, Светлинский, Северный, Соль-Илецкий, Сорочинский, Ташлинский, Шарлыкский
3	8	Акбулакский, Беляевский, Гайский, Домбаровский, Кваркенский, Сакмарский, Тоцкий, Ясненский
Итого	35	

Таким образом, в первую группу районов вошли 17,1% районов от общего числа, во вторую группу – 65,7% и в третью – 17,2%.

Формирование инвестиционной политики регионов происходило в весьма неоднородных стартовых условиях, в частности, при разной обеспеченности средствами производства. Неодинаковыми были и пути преодоления инвестиционных трудностей.

Основными элементами инвестиционной политики на мезоуровне национальной экономики являются:

1) принятие собственного законодательства, регулирующего инвестиционный процесс;

2) предоставление инвесторам в пределах своих полномочий различных льгот и стимулов финансового и нефинансового характера;

3) создание организационных структур по содействию инвестициям;

4) разработка и экспертиза инвестиционных проектов за счет государственных источников финансирования;

5) оказание содействия инвесторам в получении таможенных льгот;

6) предоставление гарантий и поручительств банкам под выделяемые ими средства для реализации отобранных на конкурсной основе инвестиционных проектов;

7) аккумулирование средств населения путем выпуска муниципальных займов.

На региональном уровне вопросы государственного стимулирования инвестиций прорабатываются лучше, нежели на федеральном, что свидетельствует о заинтересованном отношении властей к притоку капитала. С учетом сказанного о региональной инвестиционной политике это означает: разработку системы гарантий инвесторам в настоящее время следует интенсифицировать именно на федеральном уровне, а значит предстоит органически интегрировать весь тот опыт, который накоплен и на местах, и в «центре».

В рамках регионального механизма гарантами инвестиций должны быть органы государственной власти субъектов федерации и соответствующие региональные агентства по страхованию и гарантированию инвестиций. В рамках федерального - роль страхователя отводится федеральным органам государственной власти при посредничестве Российского государственного агентства по страхованию инвестиционных рисков (не исключено также участие иностранных государств и международных финансовых институтов).

СПИСОК ЛИТЕРАТУРЫ

1. Агапова Т.Н. Статистические методы изучения структуры: дис. ... д-ра экон. наук. – СПб., 1996. – 215 с.
2. Айвазян С. А., Мхитарян В. С. Прикладная статистика в задачах и упражнениях: Учебник для вузов. М.: ЮНИТИ- ДАНА, 2001. - 270 с.
3. Афанасьев и др. Эконометрика: учебник (В.Н.Афанасьев, М.М. Юзбашев, Т.И. Гуляева); под ред. В.Н. Афанасьева. – М.: Финансы и статистика, 2005. – 256 с.
4. Бабешко Л.О. Основы эконометрического моделирования: учебн. пособие для вузов / Л.О. Бабешко. – 2-е изд. перераб. и доп. – М.: Ком Книга, 2006. – 432 с.
5. Балдин К.В., Рукосуев А.В. Общая теория статистики: Учебное пособие. – М.: Дашков и К, 2010. – 312 с.
6. Богаткова Л.В., Пройдакова Е.В. Математические методы в исследовании экономического развития регионов Приволжского Федерального округа. // Вопросы статистики. – 2008. – №8. – С. 45–52.
7. Большой медицинский словарь. 2000.
8. Боровиков В. П. STATISTICA: искусство анализа данных на компьютере. Для профессионалов. – СПб., Питер, 2001. – 656 с.
9. Боровиков В. П., Ивченко Г. И. Прогнозирование в системе STATISTICA в среде WINDOWS. – М.: Финансы и статистика, 1999. – 382 с.
10. Бородич С.А. Эконометрика: Учеб. Пособие.- Мн.: Новое издание, 2001.
11. Бородич С.А. Эконометрика: учебное пособие. – Минск: ООО «Новое знание», 2005 – 408с.
12. Валентинов В.А. Эконометрика: Учебник. - 2-е изд.- Дашков и ко, 2009. – 488 с.
13. Воложанина О.А. Развитие социально-экономических систем: теория и методология: автореф. дис. ... д-ра экон. наук. – СПб., 2011. – 42 с.
14. Гмурман В.Е. Теория вероятностей и математическая статистика. – М.: Высшая школа, 2003. – 523с.
15. Громыко Г.Л., Боноев П.С.Использование кластерного анализа в классификации сельскохозяйственных предприятий по показателям эффективности их деятельности // Вопросы статистики. – 2008. – № 4. – С. 51–54.
16. Громыко Г.Л. Учебник. 2-е изд., перераб. и доп. – М.: ИНФРА-М, 2005. – 476 с. (Классический университетский учебник).
17. Гусев А.Н. Дисперсионный анализ в экспериментальной психологии. – М.: Учебно-методический коллектор «Психология», 2000. – 136 с.
18. Давнис В.В., Тинякова В.И., Мокшина С.И., Алексеева А.И. Компьютерные решения задач многомерной статистики: кластерный и дискриминантный анализ. – Воронеж: Издательство ВГУ, 2005. – 56 с.
19. Дуброва Т.А.Статистические методы прогнозирования. – М. Юнити, 2003. – 206 с.

20. Дюк В., "Data Mining - состояние, проблемы, новые решения", http://on.wplus.net/sparm/science/Data_mining.html, 1999.
21. Еремеева Н.С., Лебедева Т.В. Эконометрика: учебн. пособие для вузов. – Оренбург: ОАО «ИПК «Южный Урал», 2010. – 296 с.
22. Зарова Е.В., Чудилин Г.А. Региональная статистика: учебник. – М.: Финансы и статистика, 2006. – 624 с.
23. Золотова Е.А., Чернышева Н.А., Планирование финансовых показателей деятельности филиала коммерческого банка на основе линейных регрессионных моделей. // Финансовый менеджмент. – 2007. – № 27. – С. 7–15.
24. Илышев А.М., Шубат О.М. Многомерный статистический анализ предпринимательской активности в региональной сфере микробизнеса // Вопросы статистики. – 2008. – №4. – С. 42–51.
25. Казинец Л. С. Темпы роста и структурные сдвиги в экономике (Показатели планирования и анализа). – М.: Экономика, 1981. – 184 с.
26. Клейн Ф. Лекции о развитии математики в XIX столетии. Часть I. — Москва, Ленинград: Объединенное научно-техническое издательство НКТП СССР, 1937.
27. Кремер Н.Ш. Теория вероятности и математическая статистика. М.: Юнити – Дана, 2002. – 343 с.
28. Кремер Н.Ш. Эконометрика: учебник (Н.Ш. Кремер, Б.А. Путко). – М.: ЮНИТИ-ДАНА, 2006. – 311 с.
29. Кремер Н.Ш., Путко Б.А. Эконометрика: Учебник для вузов. – М.: ЮНИТИ – ДАНА, 2002.
30. Курс социально-экономической статистики: учебник для студентов вузов, обучающихся по специальности «Статистика». – М.: Омега-Л, 2010. – 1011 с.
31. Курышева С. В. Статистический анализ содержания труда рабочих. – Красноярск: Изд-во Красноярского ун-та, 1990. – 184 с.
32. Малюгин В.И., Демиденко М.В., Калечиц Д.Л., Миксюк А.Ю., Цукарев Т.В. Разработка и применение эконометрических моделей для прогнозирования и анализа вариантов денежно-кредитной политики // Прикладная эконометрика. – 2009. – №2. – С. 24–39.
33. Мицек С.А., Мицек Е.Б. Эконометрические и статистические оценки инвестиций в основной капитал в регионах России // Прикладная эконометрика. – 2009. – № 2. – С. 39–47.
34. Мхиторян В.С., Дуброва Т.А., Ткачев О.В. Кластерный анализ в системе «Statistica». – М.: Издательство МЭСИ, 2002. – 56 с.
35. Нейронные сети STATISTICA Neural Networks. – М., Горячая линия – Телеком, 2001. – 182 с.
36. Носко В.П. Эконометрика. Введение в анализ временных рядов: Курс лекций. М., 2002 // http://www.iet.ru/mipt/2/text/curs_econometrics_lectures.htm

37. Орлов А. И. О развитии прикладной статистики // Современные проблемы кибернетики (прикладная статистика). – М.: Знание, 1981. – С. 3–14.
38. Орлов А. И. Эконометрика. Учебник для вузов. Изд. 3-е, исправленное и дополненное. – М.: Изд-во «Экзамен», 2004. – 576 с.
39. Орлов Л. И. Эконометрика: Учебник для вузов. – 2-е изд., перераб. и доп. М.: Изд-во «Экзамен», 2003. – 576 с.
40. Первадчук В.П., Масенко И.Б. Математическая модель прогнозирования финансового состояния предприятия. // Вестник ОГУ. – 2007. – № 77. – С. 181–190.
41. Перстенва Н.П. Критерии классификации показателей структурных различий и сдвигов // Фундаментальные исследования. – 2012. – № 3. – С. 478–482.
42. Перстенёва Н.П. Методология статистического исследования структурно-динамических изменений (на примере экономики Самарской области): дис. ... канд. экон. наук. – Самара, 2003. – 141 с.
43. Петров А.Н. Эконометрические модели индекса валового регионального продукта // Экономический анализ: теория и практика. – 2010. – №31. – С. 43–52.
44. Плавинский С.Л. Биостатистика. Планирование, обработка и представление результатов биомедицинских исследований при помощи системы SAS. – СПб: Издательский дом СПб МАПО, 2005.
45. Плошко Б.Г., Елисеева И.И. История статистики: Учеб. пособие. – Москва, Ленинград: Финансы и статистика, 1990.
46. Попова И.Н. Долгосрочный прогноз производства зерна в России на основе гипертренда // Вопросы статистики. – 2009. – № 12. – С. 51–55.
47. Практикум по эконометрике: Учеб. пособие / И. И. Елисеева, С. В. Курышева, Н. М. Гордеенко и др.; Под ред. И. И. Елисеевой. М.: Финансы и статистика, 2002. 192 с: ил.
48. Практикум по эконометрике: Учеб. Пособие/ И.И.Елисеева (и др.); под ред. И.И. Елисеевой. – М.: Финансы и статистика, 2002.
49. Прикладная статистика. Основы эконометрики: учебник для вузов в 2 т. Т. 1. Теория вероятностей и прикладная статистика / С.А. Айвазян, В.С. Мхитарян. – 2-е изд., испр. – М.: ЮНИТИ-ДАНА, 2001. – 656 с.
50. Прикладная статистика. Основы эконометрики: учебник для вузов в 2 т. Т. 2. Основы эконометрики / С.А. Айвазян, В.С. Мхитарян. – 2-е изд., испр. – М.: ЮНИТИ-ДАНА, 2001. – 432 с.
51. Прикладная статистика. Основы эконометрики: Учебник для вузов: В 2 т. – 2-е изд., испр. Т. 1. Айвазян С. А., Мхитарян В. С. Теория вероятностей и прикладная статистика. – М.: ЮНИТИ-ДАНА, 2001. – 656 с.
52. Прикладная статистика. Основы эконометрики: Учебник для вузов: В 2 т. – 2-е изд., испр. Т. 2. – М.: ЮНИТИ-ДАНА, 2001. – 432 с.
53. Реннер А.Г., Седова Е.Н. Методы прогнозирования экономических показателей на основе временных рядов с учетом пространственной

неоднородности данных и нелинейной взаимосвязи между факторами. // Вестник ОГУ. – 2007. – № 4. – С. 105–111.
54. Садовникова Н.А., Шмойлова Р.А. Анализ временных рядов и прогнозирование. – М., 2001.
55. Салин В.Н., Левит Б.Ю. Проверка значимости статистических показателей с помощью таблиц их критических значений // Вопросы статистики. – 2009. – №9. – С. 69–77.
56. Свободная энциклопедия «Википедия» // http://ru.wikipedia.org/wiki/Статистика
57. Сивелькин В.А., Кузнецова В.Е. Многомерная классификация методом кластерного анализа с использованием пакета Statistica. – Оренбург: Издательство ОГАУ, 2003. – 40 с.
58. Словарь социолингвистических терминов. М., 2006.
59. Социальная статистика: учебник; под ред. И.И. Елисеевой. – М.: Финансы и статистика, 2002. – 480 с.
60. Статистический анализ продовольственной безопасности субъектов Российской Федерации (по материалам Оренбургстата)// Вопросы статистики. – 2008. – № 12. – С. 36–43.
61. Теория статистики / Под ред. Р. А. Шмойловой. – М.: Финансы и статистика, 2002. – 560 с.
62. Хендри Д. Эконометрика: алхимия или наука? // http://www.ipra.by/pdf/Hendry-36044.pdf
63. Шмойлова Р.А., Илюхина И.Е. Построение показателей прогноза связных рядов динамики основных показателей информационно-статистических услуг // Вопросы статистики. – 2008. – № 4. – С. 54–57.
64. Эконометрика: учебник / И.И. Елисеева, С.В. Курышева, Т.В. Костеева и др.; под ред. И.И. Елисеевой. – 2-е изд., перераб. и доп. – М.: Финансы и статистика, 2008. – 576 с.
65. Эконометрика: учебник / Под ред. И.И. Елисеевой. – М.: Финансы и статистика, 2005.
66. Эконометрическая страничка А. Цыплакова // http: // www.nsu.ru/ef/tsy/ecmr/
67. Юзбашев М. М., Агапова Т. Н. О показателях вариации долей отдельных групп в совокупности // Вестник статистики. – 1988. – № 10. – С. 45–54.
68. www.sutd.ru

ПРИЛОЖЕНИЯ

Приложение А

Формулы оценки при простом случайном отборе

Статистические показатели	Истинное значение	Оценка
Суммарное значение признака	$Y = \sum_{i=1}^{N} y_i$	$\hat{Y} = \frac{N}{n}\sum_{i=1}^{n} y_i$
Среднее значение признака	$\bar{Y} = \frac{1}{N}\sum_{i=1}^{N} y_i$	$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$
Дисперсия признака	$S^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \bar{Y})^2$	$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$
Дисперсия оценки суммарного значения признака	$V(\hat{Y}) = \frac{N^2 S^2}{n}\left(\frac{N-n}{N}\right)$	$v(\hat{Y}) = \frac{N^2 s^2}{n}\left(\frac{N-n}{N}\right)$
Дисперсия оценки среднего значения признака	$V(\bar{y}) = \frac{S^2}{n}\left(\frac{N-n}{N}\right)$	$v(\bar{y}) = \frac{s^2}{n}\left(\frac{N-n}{N}\right)$
Стандартная ошибка оценки суммарного значения признака	$S(\hat{Y}) = \frac{NS}{\sqrt{n}}\sqrt{\left(\frac{N-n}{N}\right)}$	$s(\hat{Y}) = \frac{Ns}{\sqrt{n}}\sqrt{\left(\frac{N-n}{N}\right)}$
Стандартная ошибка оценки среднего значения признака	$S(\bar{y}) = \frac{S}{\sqrt{n}}\sqrt{\left(\frac{N-n}{N}\right)}$	$s(\bar{y}) = \frac{s}{\sqrt{n}}\sqrt{\left(\frac{N-n}{N}\right)}$
Коэффициент вариации оценки	$CV = \frac{S(\hat{Y})}{Y}\times 100\% = \frac{S(\bar{y})}{\bar{Y}}\times 100\%$	$cv = \frac{s(\hat{Y})}{\hat{Y}}\times 100\% = \frac{s(\bar{y})}{\bar{y}}\times 100\%$

Приложение Б

Формулы оценивания при расслоенном отборе

Статистические показатели	Истинное значение	Оценка
Суммарное значение признака	$Y = \sum_{h=1}^{L} Y_h$	$\hat{Y}_{st} = \sum_{h=1}^{L} N_h \bar{y}_h$
Среднее значение признака	$\bar{Y} = \frac{1}{N} \sum_{h=1}^{L} N_h \bar{Y}_h$	$\bar{y}_{st} = \frac{1}{N} \sum_{h=1}^{L} N_h \bar{y}_h$
Дисперсия оценки суммарного значения признака	$V(\bar{Y}_{st}) = \sum_{h-1}^{L} N_h (N_h - n_h) \frac{S_h^2}{n_h}$	$v(\bar{Y}_{st}) = \sum_{h-1}^{L} N_h (N_h - n_h) \frac{S_h^2}{n_h}$
Дисперсия оценки среднего значения признака	$V(\bar{y}_{st}) = \frac{1}{N^2} \sum_{h-1}^{L} N_h (N_h - n_h) \frac{S_h^2}{n_h}$	$v(\bar{y}_{st}) = \frac{1}{N^2} \sum_{h-1}^{L} N_h (N_h - n_h) \frac{S_h^2}{n_h}$
Стандартная ошибка оценки суммарного значения признака	$S(\hat{Y}_{st}) = NS(\bar{y}_{st}) = \sqrt{\sum_{h=1}^{L} N_h (N_h - n_h) \frac{S_h^2}{n_h}}$	$s(\hat{Y}_{st}) = Ns(\bar{y}_{st}) = \sqrt{\sum_{h=1}^{L} N_h (N_h - n_h) \frac{s_h^2}{n_h}}$
Стандартная ошибка оценки среднего значения признака	$S(\bar{y}_{st}) = \sqrt{\frac{1}{N^2} \sum_{h=1}^{L} N_h (N_h - n_h) \frac{S_h^2}{n_h}}$	$s(\bar{y}_{st}) = \sqrt{\frac{1}{N^2} \sum_{h=1}^{L} N_h (N_h - n_h) \frac{s_h^2}{n_h}}$
Коэффициент вариации оценки	$CV = \frac{S(\hat{Y}_{st})}{Y} \times 100\% = \frac{S(\bar{y}_{st})}{\bar{Y}} \times 100\%$	$cv = \frac{s(\hat{Y}_{st})}{\hat{Y}_{st}} \times 100\% = \frac{s(\bar{y}_{st})}{\bar{y}_{st}} \times 100\%$

В таблице использованы следующие обозначения:

L **-** число слоев;

h **-** номер слоя;

Y_h – суммарное значение признака ***y*** в ***h***-м слое генеральной совокупности;

N_h – объем ***h***-го слоя генеральной совокупности;

$\bar{y}_h$ - среднее значение признака ***y*** в ***h***-м слое выборки;

N - объем генеральной совокупности;

$\bar{Y}_h$ - среднее значение признака ***y*** в ***h***-м слое генеральной совокупности;

n_h – объем ***h***-го слоя выборки;

S_h^2 - истинное значение дисперсии для ***h***-го слоя:

$$S_h^2 = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} (y_{hi} - \bar{Y}_h)^2;$$

i - номер элемента внутри слоя;

y_{hi} – значение признака ***y*** ***i***-го элемента слоя ***h***;

s_h^2 - несмещенная оценка дисперсии для ***h***-го слоя:

$$s_h^2 = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} (y_{hi} - \bar{y}_h)^2.$$

Приложение В

Формулы оценивания при гнездовом отборе

Статистические показатели	Истинное значение	Оценка
Суммарное значение признака	$Y=\sum_{i=1}^{M}T_i$	$\hat{Y}=\frac{M}{m}\sum_{i=1}^{m}T_i$
Среднее значение признака по гнездам	$\overline{T}=\frac{1}{M}\sum_{i=1}^{M}T_i$	$\hat{\overline{T}}=\frac{1}{m}\sum_{i=1}^{m}T_i$
Среднее значение признака	$\overline{Y}=\frac{Y}{N}$	$\hat{\overline{Y}}=\frac{\hat{Y}}{\hat{N}}$
Дисперсия оценки суммарного значения признака	$V(\hat{Y})=M^2(1-\frac{m}{M})\frac{1}{m}\frac{1}{M-1}\sum_{i=1}^{M}(T_i-\overline{T})^2$	$v(\hat{Y})=M^2(1-\frac{m}{M})\frac{1}{m}\frac{1}{m-1}\sum_{i=1}^{m}(T_i-\hat{\overline{T}})^2$
Дисперсия оценки среднего значения признака	$V(\hat{\overline{Y}})=\frac{1}{N^2}\left[M^2(1-\frac{m}{M})\frac{1}{m}\frac{1}{M-1}\sum_{i=1}^{M}(T_i-\overline{T})^2\right]$	$v(\hat{\overline{Y}})=\frac{1}{N^2}\left[M^2(1-\frac{m}{M})\frac{1}{m}\frac{1}{m-1}\sum_{i=1}^{m}(T_i-\hat{\overline{T}})^2\right]$
Стандартная ошибка оценки суммарного значения признака	$S(\hat{Y})=\sqrt{M^2(1-\frac{m}{M})\frac{1}{m}\frac{1}{M-1}\sum_{i=1}^{M}(T_i-\overline{T})^2}$	$s(\hat{Y})=\sqrt{M^2(1-\frac{m}{M})\frac{1}{m}\frac{1}{m-1}\sum_{i=1}^{m}(T_i-\hat{\overline{T}})^2}$
Стандартная ошибка оценки среднего значения признака	$S(\hat{\overline{Y}})=\frac{1}{N}\sqrt{M^2(1-\frac{m}{M})\frac{1}{m}\frac{1}{M-1}\sum_{i=1}^{M}(T_i-\overline{T})^2}$	$s(\hat{\overline{Y}})=\frac{1}{N}\sqrt{M^2(1-\frac{m}{M})\frac{1}{m}\frac{1}{m-1}\sum_{i=1}^{m}(T_i-\hat{\overline{T}})^2}$
Коэффициент вариации оценки	$CV=\frac{S(\hat{Y})}{Y}\times 100\%=\frac{S(\hat{\overline{Y}})}{\overline{Y}}\times 1$	$cv=\frac{s(\hat{Y})}{\hat{Y}}\times 100\%=\frac{s(\hat{\overline{Y}})}{\hat{\overline{Y}}}\times 100\%$

В таблице использованы следующие обозначения:

i **-** номер гнезда;

M **–** количество гнезд;

T_i – суммарное значение признака в ***i***-м гнезде;

m **–** количество выбранных гнезд;

N **–** объем генеральной совокупности;

$\hat{N}$ - оценка объема генеральной совокупности.

Приложение Г

Формулы оценивания при многоступенчатом (двухэтапном групповом) отборе

Статистические показатели	Истинное значение	Оценка
Суммарное значение признака	$Y = \sum_{i=1}^{N} Y_i$; $Y_i = \sum_{j=1}^{M_i} y_{ij}$	$\hat{Y} = \frac{N}{n} \sum_{i=1}^{n} \frac{M_i}{m_i} y_i$; $\hat{Y}_i = \frac{M_i}{m_i} y_i$
Количество элементов в генеральной совокупности	$M = \sum_{i=1}^{N} M_i$	$\hat{M} = \frac{N}{n} \sum_{i=1}^{n} M_i$
Среднее значение признака	$\bar{Y} = \frac{Y}{M}$; $\bar{Y}' = \frac{1}{N} \sum_{i=1}^{N} Y_i$; $\bar{Y}_i = \frac{1}{M_i} \sum_{j=1}^{M_i} Y_{ij}$	$\bar{y} = \frac{\hat{Y}}{\hat{M}}$; $\bar{y}' = \frac{1}{n} \sum_{i=1}^{n} \frac{M_i}{m_i} y_i$; $\bar{y}_i = \frac{1}{m_i} \sum_{j=1}^{m_i} y_{ij}$
Количество элементов в среднем на группу	$\bar{M} = \frac{M}{N} = \frac{1}{N} \sum_{i=1}^{N} M_i$	$\bar{m} = \frac{1}{n} \sum_{i=1}^{n} m_i$
Дисперсия оценки суммарного значения признака	$V(\hat{Y}) = K_1 S_b^2 + \sum_{i=1}^{N} K_2 S_{wi}^2$, где $K_1 = \frac{N^2 (N-n)}{Nn}$; $K_2 = \frac{N}{n} \frac{M_i^2 (M_i - m_i)}{M_i m_i}$	$v(\hat{Y}) = K_1 s_b^2 + \sum_{i=1}^{n} K_2 s_{wi}^2$
Дисперсия оценки среднего значения признака	$V(\bar{y}) = \frac{1}{M^2} \left[K_1 S_b^2 + \sum_{i=1}^{N} K_2 S_{wi}^2 \right]$	$v(\bar{y}) = \frac{1}{M^2} \left[K_1 s_b^2 + \sum_{i=1}^{n} K_2 s_{wi}^2 \right]$
Стандартная ошибка оценки суммарного значения признака	$S(\hat{Y}) = \sqrt{K_1 S_b^2 + \sum_{i=1}^{N} K_2 S_{wi}^2}$	$s(\hat{Y}) = \sqrt{K_1 s_b^2 + \sum_{i=1}^{n} K_2 s_{wi}^2}$
Стандартная ошибка оценки среднего значения признака	$S(\bar{y}) = \frac{1}{M} \sqrt{K_1 S_b^2 + \sum_{i=1}^{N} K_2 S_{wi}^2}$	$s(\bar{y}) = \frac{1}{M} \sqrt{K_1 s_b^2 + \sum_{i=1}^{n} K_2 s_{wi}^2}$
Коэффициент вариации оценки	$CV = \frac{S(\hat{Y})}{\hat{Y}} \times 100\% = \frac{S(\bar{y})}{\bar{Y}} \times 100\%$	$cv = \frac{s(\hat{Y})}{\hat{Y}} \times 100\% = \frac{s(\bar{y})}{\bar{y}} \times$

Приложение Д

Исходные данные для проведения корреляционно -регрессионного анализа

Годы	x_1	x_2	x_3	x_4	x_5	t
1999	9,3	15,0	19,4	9,89	12,41	1
2000	8,7	15,3	18,1	10,58	14,57	2
2001	9,0	15,6	16,0	8,98	15,83	3
2002	9,7	16,2	15,1	7,88	13,91	4
2003	10,2	16,4	15,0	8,23	13,08	5
2004	10,4	15,9	14,7	7,78	11,57	6
2005	10,2	16,1	14,5	7,17	10,77	7
2006	10,3	15,1	14,8	7,16	13,56	8
2007	11,3	14,6	16,0	6,11	12,97	9
2008	12,0	14,5	15,5	6,32	11,1	10
2009	12,3	14,1	13,9	8,42	13,03	11
2010	12,5	14,2	14,7	7,47	10,75	12
2011	12,6	13,5	23,9	6,60	11,76	13

Приложение Е

Результаты регрессионного анализа

ВЫВОД ИТОГОВ

Регрессионная статистика	
Множественный R	0,982509
R-квадрат	0,878822
Нормированный R-квадрат	0,668232
Стандартная ошибка	8,49583
Наблюдения	13

Дисперсионный анализ

	df	*SS*	*MS*	*F*	*Значимость F*
Регрессия	4	2033,278	508,3194	7,042471	0,009849
Остаток	8	577,4331	72,17913		
Итого	12	2610,711			

	Коэффициенты	*Стандартная ошибка*	*t-статистика*	*P-Значение*
Свободный член	594,455	842,7765	8,705353	0,000609
X 1	20,40115	44,49702	6,458484	0,0000 03
X 2	-14,4273	86,5637	-9,16667	0,000008
X 3	2,07235	2,457262	5,84336	0,000 527
t	-1,12389	0,940742	-11,19469	0,000005

Нижние 95%	*Верхние 95%*	*Нижние 95,0%*	*Верхние 95,0%*
-1348,99	2537,901	-1348,99	2537,901
-82,2092	123,0115	-82,2092	123,0115
-214,044	185,189	-214,044	185,189
-7,7388	3,59411	-7,7388	3,59411
-3,29325	1,045464	-3,29325	1,045464

Приложение Ж

Таблица значений F-критерия Фишера при уровне значимости 0,05

d.f.2	*d.f.1*									
	1	2	3	4	5	6	8	12	24	∞
1	161,45	199,5	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	307	2,91	2,74	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	2,95	2,79	2,61	2,40
12	4,75	3,88	3,49	3,26	3,11	3,00	2,85	2,69	2,50	2,30
13	4,67	3,80	3,41	3,18	3,02	2,92	2,77	2,60	2,42	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,70	2,53	2,35	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,64	2,48	2,29	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,59	2,42	2,24	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,55	2,38	2,19	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,51	2,34	2,15	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,48	2,31	2,11	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,45	2,28	2,08	1,84
21	4,32	3,47	3,07	2,84	2,68	2,57	2,42	2,25	2,05	1,81
22	4,30	3,44	3,05	2,82	2,66	2,55	2,40	2,23	2,03	1,78
23	4,28	3,42	3,03	2,80	2,64	2,53	2,38	2,20	2,00	1,76
24	4,26	3,40	3,01	2,78	2,62	2,51	2,36	2,18	1,98	1,73
25	4,24	3,38	2,99	2,76	2,60	2,49	2,34	2,16	1,96	1,71
26	4,22	3,37	2,98	2,74	2,59	2,47	2,32	2,15	1,95	1,69
27	4,21	3,35	2,96	2,73	2,57	2,46	2,30	2,13	1,93	1,67
28	4,20	3,34	2,95	2,71	2,56	2,44	2,29	2,12	1,91	1,65
29	4,18	3,33	2,93	2,70	2,54	2,43	2,28	2,10	1,90	1,64
30	4,17	3,32	2,92	2,69	2,53	2,42	2,27	2,09	1,89	1,62
35	4,12	3,26	2,87	2,64	2,48	2,37	2,22	2,04	1,83	1,57
40	4,08	3,23	2,84	2,61	2,45	2,34	2,18	2,00	1,79	1,51
45	4,06	3,21	2,81	2,58	2,42	2,31	2,15	1,97	1,76	1,48
50	4,03	3,18	2,79	2,56	2,40	2,29	2,13	1,95	1,74	1,44

Приложение И

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

Число степеней свободы	α			Число степеней свободы	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	3,495	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

Приложение К

Значения статистики Дарбина–Уотсона при 5-%-ном уровне значимости

n	$k^1 = 1$		$k^1 = 2$		$k^1 = 3$		$k^1 = 4$		$k^1 = 5$	
	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	0,44	2,13	0,30	2,39	-	-
10	0,88	1,32	0,70	1,64	0,53	2,02	0,38	2,41	0,24	2,82
11	0,93	1,32	0,66	1,60	0,60	1,93	0,44	2,28	0,32	2,65
12	0,97	1,33	0,81	1,58	0,66	1,86	0,51	2,18	0,38	2,51
13	1,01	1,34	0,86	1,56	0,72	1,82	0,57	2,09	0,45	2,40
14	1,05	1,35	0,91	1,55	0,77	1,78	0,63	2,03	0,51	2,30
15	1,08	1,36	0,95	1,54	0,92	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	0,86	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	1,90
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	,186
28	1,33	1,48	126	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 Л

Прогнозирование цен на картофель с помощью тренд-сезонной модели

Год	Месяц	t	Цена картофеля, руб. y_t	Скользящая средняя y'_t	$x_t = y_t - y'_t$	Десезонализированный ряд $y_t^{(1)}$	Расчетные уровни, полученные по тренд-сезонной модели $\hat{y}_t$
2022	Январь	1	41	-	-	36,07	34,77
	Февраль	2	37	-	-	34,90	34,68
	Март	3	32	-	-	32,26	34,63
	Апрель	4	30	-	-	32,15	34,60
	Май	5	29	-	-	34,05	34,59
	Июнь	6	30	-	-	36,59	34,62
	Июль	7	25	34,9	-9,9	34,63	34,67
	Август	8	27	34,9	-7,9	34,86	34,74
	Сентябрь	9	42	35,1	6,9	37,67	34,85
	Октябрь	10	46	35,4	10,6	35,07	34,98
	Ноябрь	11	45	35,7	9,3	35,97	35,14
	Декабрь	12	35	35,8	-0,8	34,78	35,32
2023	Январь	13	40	36,0	**4,0**	35,07	35,53
	Февраль	14	38	36,2	1,8	35,90	35,77
	Март	15	36	36,3	-0,3	36,26	36,04
	Апрель	16	34	36,3	-2,3	36,15	36,33
	Май	17	31	36,6	-5,6	36,05	36,65
	Июнь	18	31	37,0	-6,0	37,59	36,99
	Июль	19	28	37,5	-9,5	37,63	37,36
	Август	20	30	38,0	-8,0	37,86	37,76
	Сентябрь	21	40	38,4	1,6	35,67	38,19
	Октябрь	22	50	38,8	11,2	39,07	38,64
	Ноябрь	23	48	39,4	8,6	38,97	39,12
	Декабрь	24	41	39,9	1,1	40,78	39,63
2024	Январь	25	46	40,3	**5,7**	41,07	40,16
	Февраль	26	43	40,7	2,3	40,90	40,72
	Март	27	41	41,4	-0,4	41,26	41,30
	Апрель	28	40	42,1	-2,1	42,15	41,92
	Май	29	38	42,6	-4,6	43,05	42,56
	Июнь	30	36	43,3	-7,3	42,59	43,23
	Июль	31	33	-	-	42,63	43,92
	Август	32	35	-	-	42,86	44,64
	Сентябрь	33	51	-	-	46,67	45,39
	Октябрь	34	56	-	-	45,07	46,16
	Ноябрь	35	54	-	-	44,97	46,96
	Декабрь	36	52	-	-	51,78	47,79

*Рассчитано с помощью ППП «Microsoft Excel XP»

Приложение М

Данные по районам Оренбургской области за 2005 г.

Районы Оренбургской области	Объем платных услуг населению, тыс. руб.	Оборот розничной торговли, млн. руб.	Оборот общественного питания, тыс. руб.
Абдулинский	4327	75,3	119
Адамовский	64751	226,1	10554
Акбулакский	62512	140,7	7293
Александровский	35612	126,7	6987
Асекеевский	59475	143,5	9593
Беляевский	39541	135,3	9283
Бугурусланский	41294	180,8	5492
Бузулукский	69856	156,3	1961
Гайский	20242	51,6	1525
Грачевский	45831	110,6	7089
Домбаровский	43138	90,8	8514
Илекский	70001	182,8	10857
Кваркенский	40240	110,0	2584
Красногвардейский	46836	165,3	12387
Кувандыкский	26154	103,9	3079
Курманаевский	40811	143,0	14582
Матвеевский	27180	148,5	6279
Новоорский	132685	198,2	26758
Новосергиевский	105101	551,0	35543
Октябрьский	58393	204,4	7921
Оренбургский	1092188	943,9	123716
Первомайский	59747	180,3	16096
Переволоцкий	90349	219,8	31985
Пономаревский	34596	182,6	13815
Сакмарский	69603	218,3	14436
Саракташский	141970	295,4	18657
Светлинский	52563	138,9	3746
Северный	46054	259,5	4665
Соль-Илецкий	25458	320,2	4961
Сорочинский	19533	69,7	9089
Ташлинский	66355	289,4	7076
Тоцкий	81431	202,5	28729
Тюльганский	65880	167,9	6352
Шарлыкский	45683	237,8	2607
Ясненский	5233	20,9	797

Источник: статистический сборник «Города и районы Оренбургской области, 2006»

Приложение Н

Данные по районам Оренбургской области за 2009 г.

Районы Оренбургской области	Объем платных услуг населению, тыс. руб.	Оборот розничной торговли, млн. руб.	Оборот общественного питания, млн. руб.
Абдулинский	17414	127,7	21698
Адамовский	160766	521,7	22914
Акбулакский	122621	375,9	21079
Александровский	90002	344,3	21453
Асекеевский	117465	257,4	21266
Беляевский	84768	205,5	16069
Бугурусланский	72337	390,8	25921
Бузулукский	135768	244,6	21068
Гайский	51148	225	4073
Грачевский	99465	168,5	18350
Домбаровский	84199	249,9	7293
Илекский	147765	414,3	16493
Кваркенский	92330	212,4	8958
Красногвардейский	117149	340	48432
Кувандыкский	41698	68,1	2182
Курманаевский	96591	215,7	58592
Матвеевский	53038	271,6	11666
Новоорский	286447	686,7	68279
Новосергиевский	211334	1130	60435
Октябрьский	116054	388,7	28680
Оренбургский	1192125	2520,8	277460
Первомайский	132382	434,4	53304
Переволоцкий	136209	398,9	52841
Пономаревский	71823	310,2	23598
Сакмарский	162577	498	30978
Саракташский	317270	659,6	34663
Светлинский	116935	243,4	3094
Северный	98145	270,9	5024
Соль-Илецкий	41578	529	7900
Сорочинский	38288	82	22222
Ташлинский	132093	415,2	9701
Тоцкий	231067	519,9	40349
Тюльганский	150294	376,4	15024
Шарлыкский	90581	449,7	9421
Ясненский	30248	39,2	859

Источник: статистический сборник «Города и районы Оренбургской области, 2010»

Приложение О

Стандартизированные данные по районам Оренбургской области за 2021 г.

Районы	X_1	X_2	X_3	X_4	X_5	X_6
Абдулинский	-0,5	-1,5	0,8	-0,9	0,319470768	-0,1
Адамовский	-0,5	-0,1	-1,1	-0,2	1,28647349	0,5
Акбулакский	0,4	0,7	0,2	0,7	1,010187	-0,6
Александровский	3,4	0,7	0,2	-0,2	0,506370454	-1,2
Асекеевский	-0,4	-1,1	-0,2	-0,6	0,0675624952	0,2
Беляевский	-0,3	0,8	1,5	3,2	0,197579668	-0,5
Бугурусланский	-0,5	0,3	2,4	2,7	0,433235794	-0,1
Бузулукский	-0,6	0,7	-0,2	-0,4	-0,26560651	-0,4
Гайский	-0,6	-1,2	-1,4	0,1	0,286966474	-0,8
Грачевский	-0,6	-0,9	-0,3	-0,9	-0,0380764578	-0,3
Домбаровский	-0,2	-0,1	-0,4	0,7	-0,0380764578	0,7
Илекский	1,4	0,7	1,3	-0,6	1,69277716	0,2
Кваркенский	-0,5	-1,6	-0,4	-1,0	-0,257480437	-0,9
Красногвардейский	-0,6	0,2	-1,4	-0,3	0,400731501	0,1
Кувандыкский	0,2	-1,3	0,7	-0,5	-2,57341133	0,0
Курманаевский	-0,3	0,3	0,7	-0,7	0,49824438	-0,9
Матвеевский	-0,6	0,1	-0,4	0,2	0,335722914	-0,4
Новоорский	0,4	-0,7	0,1	-0,8	0,57137904	0,8
Новосергиевский	-0,5	0,3	0,0	2,4	0,368227207	0,2
Октябрьский	-0,6	1,2	-1,6	-0,3	0,912674119	0,9
Оренбургский	-0,6	3,1	-0,4	-1,1	1,0995738	4,1
Первомайский	-0,1	-0,5	1,3	-0,8	0,896421972	-0,3
Переволоцкий	2,8	0,7	0,3	-0,5	-0,785675202	-0,8
Пономаревский	-0,0	-0,3	1,5	-0,3	-1,23260923	0,5
Сакмарский	0,3	1,1	-0,8	-1,1	1,2214649	0,7
Саракташский	-0,3	1,3	-0,6	0,1	0,124445008	0,7
Светлинский	-0,6	-0,6	-1,2	0,3	-1,67954327	0,4
Северный	-0,6	-0,3	0,1	-0,2	-1,46826536	0,0
Соль-Илецкий	-0,4	-1,3	1,7	0,4	-0,0218243112	-1,3
Сорочинский	-0,2	0,2	0,3	0,5	-2,34588128	-0,6
Ташлинский	-0,5	0,6	-1,3	-0,1	0,100066788	1,5
Тоцкий	2,4	-0,6	0,3	0,7	-1,13509635	-0,4
Тюльганский	-0,3	0,3	0,2	0,1	0,400731501	-0,5
Шарлыкский	-0,1	0,3	-0,5	-0,5	0,335722914	0,1
Ясненский	-0,6	**-1,6**	-1,5	-0,3	-1,22448316	**-1,7**

Рассчитано на основе источника: статистический сборник «Города и районы Оренбургской области, 2012»

Приложение П

Фактические данные по районам Оренбургской области за 2021 г.

Районы	X1	X2	X3	X4	X5	X6
1 кластер						
Абдулинский	5,7	-12,1	62,5	48,5	100	242,7
Асекеевский	6,4	-9,0	44,8	56,1	96,9	298,1
Бузулукский	1,5	3,6	45,0	61,6	92,8	202,6
Гайский	2,4	-9,9	23,3	76,2	99,6	141,0
Грачевский	2,2	-7,4	42,4	46,0	95,6	219,9
Домбаровский	13,6	-2,2	40,7	91,6	95,6	365,7
Кваркенский	5,0	-12,2	41,4	43,6	92,9	130,9
Кувандыкский	24,7	-10,1	60,6	59,7	64,4	263,0
Курманаевский	11,1	0,7	60,0	53,5	102,2	123,5
Матвеевский	1,9	-0,6	42,1	78,5	100,2	196,6
Новоорский	32,0	-6,2	51,0	51,5	103,1	382,5
Первомайский	15,9	-4,8	70,9	50,7	107,1	214,2
Пономаревский	19,7	-3,2	74,9	64,2	80,9	335,5
Светлинский	1,4	-5,9	28,1	80,2	75,4	319,1
Северный	1,1	-3,6	50,6	66,4	78	267,9
Соль-Илецкий	6,8	-10,3	78,1	83,9	95,8	60,7
Сорочинский	13,8	0,1	53,2	86,5	67,2	172,2
Тюльганский	11,8	0,9	52,6	74,4	101	193,4
Шарлыкский	15,9	0,5	39,3	59,7	100,2	277,4
Ясненский	2,4	-12,7	22,3	65,4	81	8,9
2 кластер						
Абдулинский	5,7	-12,1	62,5	48,5	100	242,7
Александровский	12,6	3,2	52,3	68,1	102,3	88,4
Беляевский	11,7	4,1	74,9	164,7	98,5	181,6
Бугурусланский	5,9	0,4	91,0	148,4	101,4	243,4
Илекский	6,4	3,3	71,0	56,9	116,9	297,6
Новосергиевский	3,3	0,9	48,5	141,3	100,6	292,9
Переволоцкий	10,7	3,2	53,9	59,0	86,4	134,8
Тоцкий	9,5	-5,4	54,3	92,7	82,1	201,7
3 кластер						
Адамовский	3,9	-2,4	29,6	66,6	111,9	330,2
Красногвардейский	2,7	-0,4	24,0	65,0	101	278,5
Октябрьский	1,4	6,9	21,0	64,7	107,3	403,0
Оренбургский	1,3	20,0	40,9	41,9	109,6	881,1
Сакмарский	28,8	6,2	34,3	41,3	111,1	361,0
Саракташский	9,7	7,2	37,8	76,8	97,6	366,4
Ташлинский	3,8	2,7	26,4	69,1	97,3	488,7

Источник данных: статистический сборник «Города и районы Оренбургской области, 2012»

Приложение Р

Стандартизированные показатели для кластерного анализа (2021 г.)

Источник:
http://www.gks.ru/wps/wcm/connect/rosstat_main/rosstat/ru/statistics/publications/catalog/doc_1138631758656

Регионы России	x_1	x_2	x_3	x_4	x_6	x_7	x_8	x_9	x_{10}
Белгородская область	-0,330	-0,450	-0,033	0,221	-1,091	0,180	-0,360	-0,777	-0,330
Брянская область	-0,764	-0,506	1,473	0,791	-0,181	-0,313	0,446	-0,940	-0,764
Владимирская область	-0,513	-0,257	1,501	-0,989	0,656	-1,331	0,131	0,337	-0,513
Воронежская область	-0,596	-0,523	1,075	-0,087	0,510	0,475	0,166	-0,478	-0,596
Ивановская область	-0,765	-0,354	1,728	-0,799	0,692	-1,035	0,349	0,391	-0,765
Калужская область	-0,254	-0,247	0,166	-0,325	-0,709	-0,346	0,507	-0,098	-0,254
Костромская область	-0,601	-0,417	1,331	-0,776	0,510	-1,068	-0,333	0,147	-0,601
Курская область	-0,603	-0,677	0,336	0,981	-0,873	-0,477	-0,045	-0,777	-0,603
Липецкая область	-0,348	-0,493	0,251	-0,182	-1,127	-0,280	0,131	-0,587	-0,348
Московская область	0,724	0,059	-1,312	-1,203	-1,309	0,771	0,323	-0,071	0,724
Орловская область	-0,641	-0,358	1,473	-0,040	0,001	0,344	0,078	-0,587	-0,641
Рязанская область	-0,439	-0,457	0,762	-0,942	-0,454	-0,510	-0,088	-0,315	-0,439
Смоленская область	-0,523	-0,424	0,365	-0,087	-0,381	-0,444	0,314	-0,098	-0,523
Тамбовская область	-0,754	-0,658	0,393	0,981	-0,927	0,410	0,481	-1,239	-0,754
Тверская область	-0,389	-0,307	0,706	-0,301	-0,418	-1,397	0,761	-0,125	-0,389
Тульская область	-0,398	-0,293	1,132	-0,182	-0,618	-0,871	0,341	-0,261	-0,398
Ярославская область	-0,291	-0,172	0,194	-0,728	-0,400	-0,181	-0,403	0,092	-0,291
г. Москва	1,917	0,100	-1,596	-0,633	-0,818	4,614	0,918	1,505	1,917
Республика Карелия	0,116	0,940	1,586	-0,799	0,146	-1,265	-0,158	0,228	0,116
Республика Коми	0,635	1,221	-0,488	-0,681	-0,072	1,100	0,052	0,065	0,635
Архангельская область без НАО	0,286	1,043	0,507	-0,348	-0,218	-0,510	-0,325	0,255	0,286
Ненецкий авт. округ	3,319	2,440	-2,136	2,950	-1,965	2,446	-3,913	0,636	3,319
Вологодская область	0,013	-0,073	0,535	-0,254	-0,054	-0,477	-0,806	0,663	0,013
Калининградская область	-0,078	-0,337	-0,147	-1,179	-0,418	-0,707	0,341	0,147	-0,078
Ленинградская область	0,199	-0,074	0,166	-1,084	-0,454	-0,608	-0,202	0,174	0,199
Мурманская область	1,022	1,686	-0,431	-0,206	-0,272	-0,050	-0,220	0,717	1,022
Новгородская область	-0,307	-0,269	0,620	-0,230	0,310	0,048	0,341	-0,125	-0,307
Псковская область	-0,611	-0,422	0,819	-0,064	-0,036	-0,608	0,787	-0,559	-0,611
г. Санкт-Петербург	0,852	0,683	-0,147	-1,867	-0,964	1,297	1,776	1,939	0,852
Республика Адыгея	-0,780	-0,664	1,331	0,079	1,530	-0,970	0,621	-0,668	-0,780
Республика Дагестан	-1,110	-1,183	-1,568	2,761	-1,127	-0,181	1,093	-2,244	-1,110
Республика Ингушетия	-0,936	-1,228	0,421	0,435	2,094	-1,331	-2,697	-2,298	-0,936
Кабардино-Балкарская Республика	-0,920	-0,986	-0,772	3,995	-0,145	-0,576	-0,220	-1,863	-0,920
Республика Калмыкия	-0,914	-0,798	2,354	-0,894	4,023	-0,740	-1,244	0,038	-0,914
Карачаево-Черкесская Республика	-0,874	-0,757	0,563	1,242	0,037	-0,641	-0,176	-1,646	-0,874
Республика Северная Осетия - Алания	-0,905	-0,612	0,393	1,882	-0,873	-0,510	-0,570	-1,809	-0,905
Краснодарский край	-0,376	-0,523	-0,289	0,909	0,165	0,771	1,942	-0,397	-0,376
Ставропольский край	-0,647	-0,584	0,251	0,743	0,438	-0,181	1,461	-0,804	-0,647
Астраханская область	-0,484	-0,654	-0,715	0,957	0,037	-0,083	0,586	-0,614	-0,484
Волгоградская область	-0,529	-0,394	0,109	1,740	-0,509	-0,937	0,857	-0,424	-0,529
Ростовская область	-0,458	-0,517	-0,005	0,102	-0,254	0,180	1,531	-0,505	-0,458
Республика Башкортостан	-0,254	-0,519	-1,397	0,743	-0,891	1,362	1,338	-0,858	-0,254
Республика Марий Эл	-0,722	-0,718	1,359	-0,562	1,621	-0,247	0,612	0,337	-0,722
Республика Мордовия	-0,723	-0,671	1,331	-0,610	0,729	-0,806	-0,517	-0,315	-0,723
Республика Татарстан	-0,146	-0,429	-0,914	0,411	-1,437	0,574	1,015	-0,315	-0,146
Удмуртская Республика	-0,509	-0,338	0,450	-1,345	0,037	-1,167	0,306	0,527	-0,509
Чувашская Республика	-0,642	-0,706	0,507	0,269	0,674	-1,101	0,384	-0,179	-0,642
Пермский край	-0,163	-0,271	-1,198	0,126	-0,327	1,231	0,166	-0,179	-0,163
Кировская область	-0,665	-0,338	0,620	-0,467	0,219	-0,838	-0,211	-0,451	-0,665

Продолжение приложения Р

Стандартизированные показатели для кластерного анализа (2021 г.)

Регионы России	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}	x_{17}	x_{18}	x_{19}
Белгородская область	0,842	-0,087	0,261	0,693	-0,246	-0,363	-0,797	-0,159	-0,729
Брянская область	0,483	-0,529	-0,874	0,094	0,776	-0,324	-0,771	-0,139	-0,341
Владимирская область	1,002	-0,159	0,241	0,365	0,055	-0,012	-0,520	-0,143	-0,206
Воронежская область	0,782	0,665	0,256	-0,305	0,175	-0,129	-0,822	-0,146	-0,459
Ивановская область	0,663	0,904	-0,195	2,719	0,355	-0,012	-0,671	-0,125	1,362
Калужская область	0,723	1,519	1,550	-0,819	0,536	-0,598	-0,343	-0,184	-1,336
Костромская область	-0,493	-0,690	0,725	-0,006	-0,426	-0,441	-0,545	-0,146	1,413
Курская область	0,065	0,384	-0,802	-1,062	2,098	-0,363	-0,520	-0,163	-0,526
Липецкая область	-0,194	0,020	0,202	0,351	1,197	-0,480	-0,771	-0,150	-1,032
Московская область	0,304	-0,804	-0,005	0,679	-1,327	-0,598	-0,369	-0,183	-0,037
Орловская область	-0,593	-1,025	0,794	0,921	0,175	-0,285	-0,545	-0,155	-0,037
Рязанская область	-0,912	1,435	-0,848	0,836	0,836	-0,402	-1,149	-0,166	-1,032
Смоленская область	0,005	1,208	-0,571	0,080	1,678	-0,480	-0,419	-0,155	0,334
Тамбовская область	0,304	0,528	-0,285	0,465	0,716	-0,324	-0,923	-0,170	-0,105
Тверская область	-0,254	0,342	0,814	-0,206	-0,486	-0,480	-0,872	-0,168	0,334
Тульская область	0,623	0,283	-0,072	0,793	1,437	-0,480	-0,923	-0,170	0,519
Ярославская область	-0,493	-0,380	-0,187	0,194	0,476	-0,207	-0,217	-0,164	0,502
г. Москва	-0,134	-0,511	-2,913	-1,361	-1,027	-0,715	2,150	-0,188	-0,088
Республика Карелия	-0,095	0,008	-0,264	-0,776	-0,786	0,105	-0,570	-0,121	1,784
Республика Коми	-1,868	0,325	-0,072	-1,732	-0,125	-0,051	-0,268	-0,145	0,873
Архангельская область без НАО	0,643	-1,228	-0,062	-0,191	0,055	-0,051	-0,771	-0,150	1,430
Ненецкий авт. округ	1,938	-2,201	0,899	-0,676	2,639	0,066	3,006	-0,143	1,497
Вологодская область	-0,772	-1,640	0,317	-0,905	0,175	-0,129	-0,041	-0,125	-0,948
Калининградская область	2,178	1,226	0,054	0,037	0,716	-0,246	0,765	-0,148	2,205
Ленинградская область	-0,533	-0,177	0,102	-1,133	0,536	-0,519	-0,318	-0,173	0,401
Мурманская область	-1,350	2,539	-0,461	-0,320	0,055	0,066	0,085	-0,148	0,738
Новгородская область	-0,692	0,713	0,200	1,235	-0,005	-0,363	-0,545	-0,177	-0,560
Псковская область	-0,673	-0,266	0,269	-0,006	-0,606	0,027	-0,520	-0,119	0,249
г. Санкт-Петербург	0,822	-0,368	-0,582	-0,049	0,235	-0,676	0,362	-0,179	-1,875
Республика Адыгея	1,400	-1,258	0,397	3,404	0,656	0,339	-1,225	-0,137	-1,251
Республика Дагестан	1,600	1,363	-0,113	1,834	1,017	0,535	0,186	2,041	0,013
Республика Ингушетия	3,453	-2,738	-1,102	0,536	1,738	8,109	-1,854	8,654	1,295
Кабардино-Балкарская Республика	-0,832	-1,126	0,492	0,336	1,257	1,433	-0,092	-0,047	0,924
Республика Калмыкия	-1,111	0,354	-0,853	-1,732	1,257	0,535	-0,192	0,067	0,654
Карачаево-Черкесская Республика	-1,051	-0,141	-0,897	-0,933	0,295	0,144	-0,746	-0,016	0,468
Республика Северная Осетия - Алания	-0,573	-0,619	0,466	-0,049	0,175	0,339	-0,822	0,013	1,548
Краснодарский край	0,264	0,623	-0,502	0,251	-0,546	-0,519	-0,394	-0,179	-1,386
Ставропольский край	-0,852	0,217	-0,021	-0,477	0,656	-0,012	-0,041	-0,130	-1,656
Астраханская область	-0,095	-0,040	-0,092	0,921	-0,606	-0,324	0,437	-0,152	0,553
Волгоградская область	-0,394	-0,004	-0,477	-0,548	-0,546	-0,168	-0,041	-0,163	0,030
Ростовская область	0,683	0,122	0,036	0,194	0,235	-0,363	-0,520	-0,168	-1,235
Республика Башкортостан	0,105	0,259	-0,057	0,608	-0,967	-0,285	0,034	-0,137	-1,875
Республика Марий Эл	-0,314	-0,374	0,264	0,764	-0,065	-0,519	-0,041	-0,164	-0,290
Республика Мордовия	0,404	0,396	0,105	0,422	0,235	-0,441	-0,520	-0,148	-1,690
Республика Татарстан	0,344	-0,153	-0,062	0,722	-1,327	-0,363	0,085	-0,150	-1,049
Удмуртская Республика	-0,991	-0,941	-0,349	1,064	-0,185	-0,246	0,412	-0,159	-0,021
Чувашская Республика	0,483	-0,171	-0,016	1,135	0,115	-0,207	0,085	-0,141	-1,369
Пермский край	-0,174	-0,344	-0,177	-0,990	0,776	-0,090	0,060	-0,112	-0,729
Кировская область	-0,832	-0,881	-0,349	0,593	0,115	-0,324	-0,243	-0,155	-0,628

Продолжение приложения Р

Регионы России	x_1	x_2	x_3	x_4	x_6	x_7	x_8	x_9	x_{10}
Нижегородская область	-0,330	-0,450	-0,033	0,221	-1,091	0,180	-0,360	-0,777	-0,330
Оренбургская область	-0,764	-0,506	1,473	0,791	-0,181	-0,313	0,446	-0,940	-0,764
Пензенская область	-0,513	-0,257	1,501	-0,989	0,656	-1,331	0,131	0,337	-0,513
Самарская область	-0,596	-0,523	1,075	-0,087	0,510	0,475	0,166	-0,478	-0,596
Саратовская область	-0,765	-0,354	1,728	-0,799	0,692	-1,035	0,349	0,391	-0,765
Ульяновская область	-0,254	-0,247	0,166	-0,325	-0,709	-0,346	0,507	-0,098	-0,254
Курганская область	-0,601	-0,417	1,331	-0,776	0,510	-1,068	-0,333	0,147	-0,601
Свердловская область	-0,603	-0,677	0,336	0,981	-0,873	-0,477	-0,045	-0,777	-0,603
Тюменская область без АО	-0,348	-0,493	0,251	-0,182	-1,127	-0,280	0,131	-0,587	-0,348
Ханты-Мансийский АО	0,724	0,059	-1,312	-1,203	-1,309	0,771	0,323	-0,071	0,724
Ямало-Ненецкий авт. округ	-0,641	-0,358	1,473	-0,040	0,001	0,344	0,078	-0,587	-0,641
Челябинская область	-0,439	-0,457	0,762	-0,942	-0,454	-0,510	-0,088	-0,315	-0,439
Республика Алтай	-0,523	-0,424	0,365	-0,087	-0,381	-0,444	0,314	-0,098	-0,523
Республика Бурятия	-0,754	-0,658	0,393	0,981	-0,927	0,410	0,481	-1,239	-0,754
Республика Тыва	-0,389	-0,307	0,706	-0,301	-0,418	-1,397	0,761	-0,125	-0,389
Республика Хакасия	-0,398	-0,293	1,132	-0,182	-0,618	-0,871	0,341	-0,261	-0,398
Алтайский край	-0,291	-0,172	0,194	-0,728	-0,400	-0,181	-0,403	0,092	-0,291
Забайкальский край	1,917	0,100	-1,596	-0,633	-0,818	4,614	0,918	1,505	1,917
Красноярский край	0,116	0,940	1,586	-0,799	0,146	-1,265	-0,158	0,228	0,116
Иркутская область	0,635	1,221	-0,488	-0,681	-0,072	1,100	0,052	0,065	0,635
Кемеровская область	0,286	1,043	0,507	-0,348	-0,218	-0,510	-0,325	0,255	0,286
Новосибирская область	3,319	2,440	-2,136	2,950	-1,965	2,446	-3,913	0,636	3,319
Омская область	0,013	-0,073	0,535	-0,254	-0,054	-0,477	-0,806	0,663	0,013
Томская область	-0,078	-0,337	-0,147	-1,179	-0,418	-0,707	0,341	0,147	-0,078
Республика Саха (Якутия)	0,199	-0,074	0,166	-1,084	-0,454	-0,608	-0,202	0,174	0,199
Камчатский край	1,022	1,686	-0,431	-0,206	-0,272	-0,050	-0,220	0,717	1,022
Приморский край	-0,307	-0,269	0,620	-0,230	0,310	0,048	0,341	-0,125	-0,307
Хабаровский край	-0,611	-0,422	0,819	-0,064	-0,036	-0,608	0,787	-0,559	-0,611
Амурская область	0,852	0,683	-0,147	-1,867	-0,964	1,297	1,776	1,939	0,852
Магаданская область	-0,780	-0,664	1,331	0,079	1,530	-0,970	0,621	-0,668	-0,780
Сахалинская область	-1,110	-1,183	-1,568	2,761	-1,127	-0,181	1,093	-2,244	-1,110
Еврейская авт. область	-0,936	-1,228	0,421	0,435	2,094	-1,331	-2,697	-2,298	-0,936
Чукотский авт. округ	-0,920	-0,986	-0,772	3,995	-0,145	-0,576	-0,220	-1,863	-0,920

Продолжение приложения Р

Регионы России	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}	x_{17}	x_{18}	x_{19}
Нижегородская область	-0,330	-0,450	-0,033	0,221	-1,091	0,180	-0,360	-0,777	-0,330
Оренбургская область	-0,764	-0,506	1,473	0,791	-0,181	-0,313	0,446	-0,940	-0,764
Пензенская область	-0,513	-0,257	1,501	-0,989	0,656	-1,331	0,131	0,337	-0,513
Самарская область	-0,596	-0,523	1,075	-0,087	0,510	0,475	0,166	-0,478	-0,596
Саратовская область	-0,765	-0,354	1,728	-0,799	0,692	-1,035	0,349	0,391	-0,765
Ульяновская область	-0,254	-0,247	0,166	-0,325	-0,709	-0,346	0,507	-0,098	-0,254
Курганская область	-0,601	-0,417	1,331	-0,776	0,510	-1,068	-0,333	0,147	-0,601
Свердловская область	-0,603	-0,677	0,336	0,981	-0,873	-0,477	-0,045	-0,777	-0,603
Тюменская область без АО	-0,348	-0,493	0,251	-0,182	-1,127	-0,280	0,131	-0,587	-0,348
Ханты-Мансийский АО	0,724	0,059	-1,312	-1,203	-1,309	0,771	0,323	-0,071	0,724
Ямало-Ненецкий авт. округ	-0,641	-0,358	1,473	-0,040	0,001	0,344	0,078	-0,587	-0,641
Челябинская область	-0,439	-0,457	0,762	-0,942	-0,454	-0,510	-0,088	-0,315	-0,439
Республика Алтай	-0,523	-0,424	0,365	-0,087	-0,381	-0,444	0,314	-0,098	-0,523
Республика Бурятия	-0,754	-0,658	0,393	0,981	-0,927	0,410	0,481	-1,239	-0,754
Республика Тыва	-0,389	-0,307	0,706	-0,301	-0,418	-1,397	0,761	-0,125	-0,389
Республика Хакасия	-0,398	-0,293	1,132	-0,182	-0,618	-0,871	0,341	-0,261	-0,398
Алтайский край	-0,291	-0,172	0,194	-0,728	-0,400	-0,181	-0,403	0,092	-0,291
Забайкальский край	1,917	0,100	-1,596	-0,633	-0,818	4,614	0,918	1,505	1,917
Красноярский край	0,116	0,940	1,586	-0,799	0,146	-1,265	-0,158	0,228	0,116
Иркутская область	0,635	1,221	-0,488	-0,681	-0,072	1,100	0,052	0,065	0,635
Кемеровская область	0,286	1,043	0,507	-0,348	-0,218	-0,510	-0,325	0,255	0,286
Новосибирская область	3,319	2,440	-2,136	2,950	-1,965	2,446	-3,913	0,636	3,319
Омская область	0,013	-0,073	0,535	-0,254	-0,054	-0,477	-0,806	0,663	0,013
Томская область	-0,078	-0,337	-0,147	-1,179	-0,418	-0,707	0,341	0,147	-0,078
Республика Саха (Якутия)	0,199	-0,074	0,166	-1,084	-0,454	-0,608	-0,202	0,174	0,199
Камчатский край	1,022	1,686	-0,431	-0,206	-0,272	-0,050	-0,220	0,717	1,022
Приморский край	-0,307	-0,269	0,620	-0,230	0,310	0,048	0,341	-0,125	-0,307
Хабаровский край	-0,611	-0,422	0,819	-0,064	-0,036	-0,608	0,787	-0,559	-0,611
Амурская область	0,852	0,683	-0,147	-1,867	-0,964	1,297	1,776	1,939	0,852
Магаданская область	-0,780	-0,664	1,331	0,079	1,530	-0,970	0,621	-0,668	-0,780
Сахалинская область	-1,110	-1,183	-1,568	2,761	-1,127	-0,181	1,093	-2,244	-1,110
Еврейская авт. область	-0,936	-1,228	0,421	0,435	2,094	-1,331	-2,697	-2,298	-0,936
Чукотский авт. округ	-0,920	-0,986	-0,772	3,995	-0,145	-0,576	-0,220	-1,863	-0,920

Лаптева Е. В.
Портнова Л.В.

Статистические методы исследований в экономике

Учебное пособие

Электронное издание сетевого распространения
Доступ к сборнику – постоянный, свободный и бесплатный.
Сборник содержится в едином файле PDF.

http://sphere-publishing.ru/images/banners/lapteva2.pdf

Максимальный объем: 15 МБ.

Издательство ООО «Сфера»
400127, Волгоград, ул. Менделеева, 43

www.sphere-publishing,ru
sphere-vlg@mail.ru

Дата издания: 20.04.2022