**Владимирский государственный университет**

# СТАТИСТИЧЕСКИЕ МЕТОДЫ В УПРАВЛЕНИИ ФИРМОЙ

**Учебное пособие**

**Владимир 2024**

Министерство науки и высшего образования Российской Федерации

Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Владимирский государственный университет
имени Александра Григорьевича и Николая Григорьевича Столетовых»

# СТАТИСТИЧЕСКИЕ МЕТОДЫ В УПРАВЛЕНИИ ФИРМОЙ

Учебное пособие

*Электронное издание*

Владимир 2024


ISBN 978-5-9984-2102-0

УДК 311:33
ББК 65.051

**Авторы-составители:** Д. Ю. Фраймович, М. Л. Быкова

Рецензенты:
Доктор экономических наук, профессор
профессор кафедры бизнес-информатики и экономики
Владимирского государственного университета
имени Александра Григорьевича и Николая Григорьевича Столетовых
*А. М. Губернаторов*

Доктор экономических наук, доцент
профессор кафедры финансов
Российской академии народного хозяйства
и государственной службы при Президенте Российской Федерации
(Владимирский филиал)
*Т. В. Старикова*

**Статистические** методы в управлении фирмой [Электронный ресурс] : учеб. пособие / авт.-сост.: Д. Ю. Фраймович, М. Л. Быкова ; Владим. гос. ун-т им. А. Г. и Н. Г. Столетовых. – Владимир : Изд-во ВлГУ, 2024. – 404 с. – ISBN 978-5-9984-2102-0 – Электрон. дан. (8,12 Мб). – 1 электрон. опт. диск (CD-ROM). – Систем. требования: Intel от 1,3 ГГц ; Windows XP/7/8/10 ; Adobe Reader ; дисковод CD-ROM. – Загл. с титул. экрана.

Включает полный курс дисциплины «Статистические методы в управлении фирмой». Рассмотрены теоретические и практические основы дисциплины, проанализированы актуальные статистические методы в управлении фирмой.

Предназначено для студентов магистратуры, обучающихся по направлению подготовки «Экономика» всех форм обучения.

Рекомендовано для формирования профессиональных компетенций в соответствии с ФГОС ВО.

Табл. 15. Ил. 131. Библиогр.: 110 назв.

ISBN 978-5-9984-2102-0

## ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ .......... 6

Глава 1. ТЕОРЕТИЧЕСКИЕ АСПЕКТЫ ПРИМЕНЕНИЯ СТАТИСТИЧЕСКИХ МЕТОДОВ В УПРАВЛЕНИИ ФИРМОЙ .......... 8
1.1. Предмет, задачи и методы статистической науки .......... 8
1.2. Значение статистического наблюдения в управлении фирмой. Классификация статистических наблюдений .......... 14
1.3. Статистическая сводка и группировка .......... 22
1.4. Представление данных .......... 25
1.5. Роль прогнозирования в управлении фирмой .......... 32
1.6. Вероятностные характеристики, применяемые в управлении фирмой .......... 44
Контрольные вопросы .......... 47
ПРИМЕРЫ РЕШЕНИЯ ЗАДАЧ .......... 50

Глава 2. АНАЛИЗ ДИНАМИЧЕСКИХ ХАРАКТЕРИСТИК ДЕЯТЕЛЬНОСТИ ФИРМЫ .......... 56
2.1. Ряды динамики в управлении фирмой .......... 56
2.2. Проблема сопоставимости динамических рядов и пути ее решения .......... 61
2.3. Работа с аномальными статистическими значениями в процессе управления фирмой .......... 64
2.4. Модели временных рядов. Структурные элементы ряда .......... 74
Контрольные вопросы .......... 78
ПРИМЕРЫ РЕШЕНИЯ ЗАДАЧ .......... 80

Глава 3. ОСНОВНЫЕ ТЕНДЕНЦИИ В УПРАВЛЕНИИ ФИРМОЙ .......... 100
3.1. Основные критерии проверки гипотезы о существовании тренда .......... 100
3.2. Основные типы тенденций и уравнений тренда .......... 108
3.3. Определение типа тренда в развитии фирмы и оценка его параметров .......... 118
3.4. Метод наименьших квадратов как инструмент оценки параметров тренда .......... 122

3.5. Логарифмирование как инструмент оценки параметров экспоненциального, логарифмического и логистического уравнений тренда ........ 125
3.6. Сглаживание рядов, характеризующих динамику изменения показателей деятельности фирмы ........ 128
Контрольные вопросы ........ 131
ПРИМЕРЫ РЕШЕНИЯ ЗАДАЧ ........ 133

Глава 4. ОЦЕНКА ПЕРИОДИЧЕСКОЙ КОМПОНЕНТЫ В ДИНАМИКЕ ИЗМЕНЕНИЯ ПОКАЗАТЕЛЕЙ ДЕЯТЕЛЬНОСТИ ФИРМЫ ........ 159
4.1. Методы выявления периодической компоненты ........ 159
4.2. Сезонная составляющая в деятельности фирмы. Индексы сезонности ........ 160
4.3. Оценка параметров сезонности на основе абсолютных приростов ........ 164
Контрольные вопросы ........ 166
ПРИМЕРЫ РЕШЕНИЯ ЗАДАЧ ........ 167

Глава 5. КОРРЕЛЯЦИЯ И РЕГРЕССИЯ ........ 174
5.1. Регрессионный анализ связанных временных рядов ........ 174
5.2. Корреляция временных рядов ........ 179
Контрольные вопросы ........ 181
ПРИМЕРЫ РЕШЕНИЯ ЗАДАЧ ........ 183

Глава 6. ПРОГНОЗИРОВАНИЕ ДЕЯТЕЛЬНОСТИ ФИРМЫ С ИСПОЛЬЗОВАНИЕМ СТАТИСТИЧЕСКИХ МЕТОДОВ ........ 186
6.1. Сущность экстраполяции как метода прогнозирования ........ 186
6.2. Прогнозирование на основе аналитического выравнивания тренда ........ 190
6.3. Количественная оценка достоверности прогноза развития фирмы ........ 192
Контрольные вопросы ........ 193
ПРИМЕРЫ РЕШЕНИЯ ЗАДАЧ ........ 194

Глава 7. ОСНОВЫ КЛАСТЕРНОГО АНАЛИЗА ........ 201
7.1. Введение в кластерный анализ ........ 201

7.2. Понятие близости объектов и их характеристика........................ 208
7.3. Методы, используемые в кластерном анализе ............................ 213
7.4. Особенности решения задачи кластерного анализа
в программе Statistica ............................................................................ 221
Контрольные вопросы............................................................................. 235
ПРИМЕРЫ РЕШЕНИЯ ЗАДАЧ........................................................... 239

Глава 8. ДИСКРИМИНАНТНЫЙ АНАЛИЗ В СТАТИСТИКЕ
ФИРМЫ ................................................................................................. 276
8.1. Основы дискриминантного анализа ............................................. 276
8.2. Реализация процедуры дискриминантного анализа .................... 280
Контрольные вопросы............................................................................. 286

Глава 9. ФАКТОРНЫЙ АНАЛИЗ ....................................................... 288
9.1. Основные элементы факторного анализа ..................................... 288
9.2. Метод главных компонент в факторном анализе ........................ 294
9.3. Каноническая модель факторного анализа................................... 295
9.5. Значимость факторных признаков ................................................ 298
9.6. Особенности решения задачи факторного анализа
в программе Statitica............................................................................... 303
Контрольные вопросы............................................................................. 326

Глава 10. ДИСПЕРСИОННЫЙ АНАЛИЗ .......................................... 329
10.1. Сущность дисперсионного анализа в статистике фирмы ......... 329
10.2. Однофакторный дисперсионный анализ .................................... 330
10.3. Многофакторный дисперсионный анализ ................................... 334
10.4. Дисперсионный анализ для повторных наблюдений ................ 341
10.5. Апостериорные множественные сравнения средних ................ 349
Контрольные вопросы............................................................................. 353

МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ ПО ВЫПОЛНЕНИЮ
КУРСОВЫХ РАБОТ (ПРОЕКТОВ) ..................................................... 355

ПРИМЕРНАЯ ТЕМАТИКА КУРСОВЫХ РАБОТ
ПО ДИСЦИПЛИНЕ................................................................................ 376

ЗАКЛЮЧЕНИЕ....................................................................................... 378

ПРИЛОЖЕНИЯ ..................................................................................... 391

## ВВЕДЕНИЕ

В современном мире статистические методы активно используются в управлении фирмой. Их широкому распространению способствовало повсеместное внедрение персональных компьютеров. Применение пакетов общего назначения и специализированных программных продуктов позволило упростить задачу управления фирмой за счет быстрой и наглядной обработки больших массивов данных, выявления общих трендов развития, рассмотрения общих закономерностей изменения показателей деятельности фирмы.

Статистические методы активно применяются в фирмах, относящихся к различным сферам народно-хозяйственной деятельности. Рассмотренные в учебном пособии способы и приемы могут быть использованы при планировании и прогнозировании траектории развития фирмы на основе статистического анализа основных показателей, характеризующих динамику изменения основных параметров развития фирмы.

Цель изучения дисциплины состоит в формировании у студентов теоретических знаний в области статистических вычислений, а также развитии практических навыков решения конкретных профессиональных задач управления фирмой на основе знаний основных статистических методов и приемов.

Пособие содержит десять глав, в которых рассматриваются теоретические и практические основы курса.

Отдельное внимание уделено конкретным примерам, позволяющим оценить особенности применения различных методов и приемов при решении профессиональных задач. Дидактический материал пособия включает контрольные вопросы, примеры решения задач, тематику курсовых работ.

В случае успешного изучения курса предполагается формирование результатов, благодаря которым обучающийся:

– **знает** стратегические методы анализа и оценки текущего состояния фирмы, сущность планирования и прогнозирования в аспекте финансово-инвестиционных параметров фирмы, приемы планирования и прогнозирования деятельности фирмы, основывающиеся на статистических данных;

– **умеет** применять статистические методы анализа и оценки текущего состояния фирмы, использовать инструментарий планирования и прогнозирования в аспекте финансово-инвестиционных параметров фирм;

– **владеет** навыками планирования и прогнозирования деятельности фирмы на основе статистических данных, оценки текущего состояния фирмы.

# Глава 1. ТЕОРЕТИЧЕСКИЕ АСПЕКТЫ ПРИМЕНЕНИЯ СТАТИСТИЧЕСКИХ МЕТОДОВ В УПРАВЛЕНИИ ФИРМОЙ

## 1.1. Предмет, задачи и методы статистической науки

Слово «статистика» происходит от латинского «status», т.е. состояние. В Средние века данное слово использовалось для обозначения политического состояние государства.

В науку данный термин ввел немецкий ученый Готфрид Ахенвалль. Зарождение статистики как науки произошло только в XVII в., однако элементы статистического учета можно было наблюдать еще в глубокой древности. Из исторических источников известно, что переписи населения в Китае проводились еще за 5 тыс. лет до н. э., древние римляне регулярно проводили статистическое сравнение военного потенциала разных стран, вели учет имущества граждан. В Средние века активно велся учет домашнего имущества и земель [1].

У истоков статистической науки стояли две основные школы (рис. 1.1).

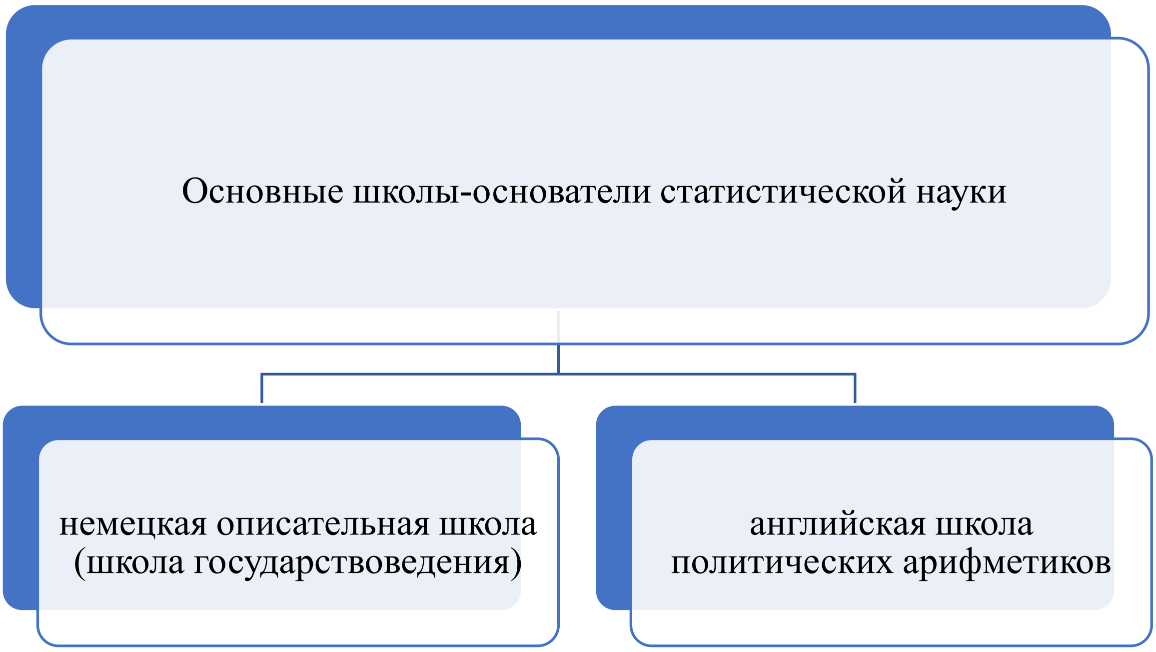



*Рис. 1.1. Основные школы-основатели статистической науки*

Представители описательной школы считали, что основная задача статистики состоит в описании территории государства, населения, климата, вероисповедания, способов ведения хозяйства и т. п..

Также они отмечали, что ведение данного учета важно исключительно в словесной форме.

По сути они продвигали идеи учета исключительно моментной статистики, без учета особенностей развития территорий в те или иные периоды, цифр и динамических характеристик

Основными представителями немецкой школы были Г. Конринг, Г. Ахенвалль, А. Бюшинг и др.

К представителям российской школы государствоведения, разделявшим аналогичные взгляды, можно отнести первооткрывателя табличного метода в статистике И. К. Кириллова, В. Н. Татищева, который занимался проблемой источниковедения, М.В. Ломоносова. Представившего экономико-географическое описание Российского государства и разработал подробную анкету для сбора статистических данных, руководителя первого в стране Статистического комитета К. Ф.

Несмотря на значительные достижения в данной области всех вышеперечисленных ученых, основателем русской государственной статистики принято считать П. П. Семенова-Тян-Шанского. Именно он стал инициатором проведения Всероссийской переписи населения в 1897 г. и отвечал за обработку ее материалов. Семенов-Тян-Шанский является автором множества сборников и справочников по фабрично-заводской статистике.

Основная цель политических арифметиков состояла в изучении общественных массовых явления с применением числовых характеристик.

Именно с открытием данной школы принято выделять новый этап в развитии статистической науки, поскольку от описания явлений и процессов статистика перешла к их измерению и исследованию, к выработке вероятных гипотез будущего развития.

Основной назначение статистической науки политические арифметики видели в изучении массовых общественных явлений. Они осознавали необходимость учета в статистическом исследовании требований закона больших чисел, поскольку закономерность может проявиться лишь при достаточно большом объеме анализируемой совокупности. История показала, что последнее слово в статистической науке осталось именно за школой политических арифметиков.

Основателем английской школы принято считать Уильяма Петти. Данный ученый активно интересовался хозяйственными процессами, закономерностями в экономической жизни страны. Именно Петти впервые сделал попытку оценки национального богатства и национального дохода.

Также яркими представителями английской школы политических арифметиков можно назвать Дж. Граунта, который исследовал закономерности воспроизводства населения и построил первую в историю таблицу смертности в своей работе «Естественные и политические наблюдения, перечисленные в прилагаемом оглавлении и сделанные над бюллетенями смертности, по отношению к управлению, религии, торговле, росту, болезням и пр.» , А. Кетле, который возглавлял национальную статистику Бельгии и был основоположником учения о средних величинах. Также Кетле изучал закономерности общественной жизни в области преступности и выявил действие постоянных и случайных причин и впервые ввел термин «средний человек».

Основное преемственное направление английской школы политических арифметиков в статистике – математическое. Оно возникло в XIX веке под влиянием идей Ф. Гальтона, К. Пирсона, У. Госсета (Стьюдента), Р. Фишера и других исследователей.

Существенное влияния на эволюцию статистической методологии оказали труды российских статистиков, представителей так называемой академической статистики, А. А. Чупрова, В. С. Немчинова, С. Г. Струмилина и др.

Развитие статистической науки и расширение области практической статистической работы привели к изменению сущности понятия «статистика». В настоящее время данный термин употребляется в трех значениях:

1) статистика как отрасль практической деятельности, основная цель которой состоит в сборе, обработке и анализе данных о разнообразных явлениях общественной жизни. Полученная в результате статистического исследования информация позволяет решать задачи выявления реально существующих закономерностей, свойственных описываемым процессам и явлениям;

2) статистика — это данные, которые служат для количественной характеристикой общественных явлений или территориального распределения показателя;

3) статистика как это наука. Как и любая другая наука, статистика имеет свой предмет и метод изучения. Предмет статистики заключается в изучении количественной стороны массовых социально-экономических явлений в связи с их качественной стороной, в исследовании количественно выраженных закономерностей общественного развития в конкретных условиях места и времени. Свой предмет статистика изучает при помощи специфического метода. Кратко статистический метод можно охарактеризовать следующим образом: это сбор, обобщение, представление, анализ и интерпретация данных [1].

Так как статистика изучает множество социально-экономических явлений и характерные для них закономерности, то и метод статистики представляет собой целую совокупность приемов, пользуясь которыми статистика исследует свой предмет.

К основным приемам статистической науки относят статистическое наблюдение, метод группировки и обобщения данных с последующим представлением результатов анализа и их интерпретацией

Задачи статистики как науки состоят в следующем:

- ✓ описание структуры экономики;
- ✓ описание тенденций развития экономики в будущем;
- ✓ анализ и прогнозирование различных экономических явлений;
- ✓ выявление факторов развития экономики для принятия управленческих решений.

В России экономико-статистические исследования проводятся научно-исследовательскими институтами, ведомственными статистическими органами и организациями, а также независимыми специалистами, однако преимущественная часть статистической информации формируется в системе официальной государственной статистики. В статистических управлениях первичная статистическая информация последовательно агрегируется с целью получения на уровне Росстата РФ макромодели функционирования экономики страны в виде системы национальных счетов.

Статистические органы преобразуют полученные от респондентов индивидуальные сведения и предоставляют потребителям информацию в полном соответствии с принципом конфиденциальности: только макроданные, относящиеся не менее чем к трем объектам наблюдения. Основные принципы организации работы органов официальной статистики в России (принцип легальности, принципы предметной централизации и региональной децентрализации) соответствуют требованиям Евростата и Департамента статистики ООН. В соответствии с международными стандартами ведения статистики и учета в России к официальной статистике относятся государственные статистические управления и ведомственная статистика (внутренняя и внешняя), т. е. определенные государственные организации, которые выполняют важные статистические работы, связанные с их собственной деятельностью (например, отделы ЗАГС) [1].

Права и обязанности официальной статистики детально урегулированы на федеральном уровне. Наряду с этим существует широкая и разнообразная сфера альтернативной статистики, т. е. частных, неофициальных статистических исследований, организаторы которых не имеют полномочий для проведения обследований с обязанностью предоставления информации широкому кругу лиц.

Основные принципы работы статистических управлений, в том числе в отношении сбора данных о населении представлены на рис. 1.2.

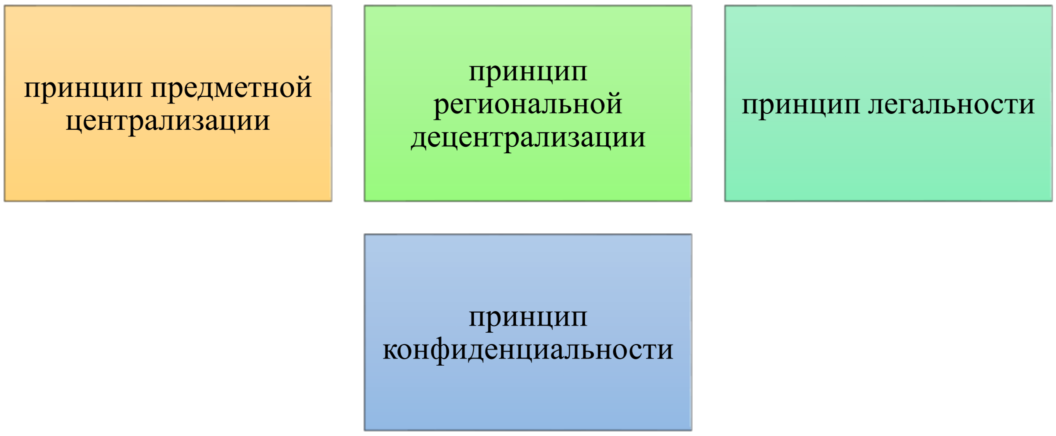



*Рис. 1.2. Основные принципы работы статистических управлений*

Реализация данных принципов позволяет добиться нейтральной и независимой позиции статистических управлений и тем самым укрепить доверие респондентов и пользователей, без которого статистика не может обойтись.

С мая 2012 г. деятельностью Росстата руководит Правительство РФ. Росстат РФ, его органы в республиках, краях, областях, автономных областях и округах, в городах Москве и Санкт-Петербурге, других городах и районах, а также подведомственные им организации, учреждения и учебные заведения составляют единую систему государственной статистики страны. Формы и методы сбора и обработки статистических данных, методология расчета статистических показателей, установленные Росстатом, являются статистическими стандартами РФ [1].

Основные задачи Росстата РФ представлены на рис. 1.3.

- предоставление официальной статистической информации Президенту РФ, Правительству РФ, Федеральному Собранию РФ, федеральным органам исполнительной власти, средствам массовой информации, другим организациям, в том числе международным;

- разработка и совершенствование научно обоснованной статистической методологии, соответствующей международным стандартам;

- координация статистической деятельности в государстве;

- разработка экономико-статистической информации, ее анализ, составление национальных счетов, проведение необходимых балансовых расчетов;

- развитие информационной системы государственной статистики, обеспечение ее совместимости и взаимодействия с другими государственными информационными системами.

*Рис. 1.3. Основные задачи Росстата*

Таким образом, в ходе эволюции статистической науки менялись представлении о ее предмете. В данном параграфе были рассмотрены задачи, предмет и методы статистической науки, а также основные принципы организации работы органов официальной статистики в России.

## 1.2. Значение статистического наблюдения в управлении фирмой. Классификация статистических наблюдений

Статистическое наблюдение – это научно обоснованная регистрация по единой разработанной программе фактов и их признаков, которые характеризуют явления общественной жизни, и сбор массовых данных.

Любое статистическое наблюдение начинается с планирования и организации исследования. Данный этап включает в себя разработку программы статистического наблюдения, определение критического момента наблюдения, времени и периода наблюдения, определение цели и задач исследования, объекта наблюдения.

Вторым этапом статистического исследования является непосредственно статистическое наблюдение. При его организации важная роль отводится планированию: от качества отобранных статистических данных зависит правильность и достоверность выводов, которые будут использоваться в рамках управляющего воздействия.

Для более четкой организации статистического наблюдения разрабатывают программу наблюдений, которая представляет собой перечень вопросов, по которым собираются сведения, либо перечень признаков и показателей, которые подлежат регистрации [1].

Программа наблюдения оформляется в виде бланка (анкеты, формуляра), в который заносятся первичные сведения. Необходимым дополнением к бланку является инструкция (или указания на самих формулярах), в которой разъясняется смысл представленных вопросов. Состав и содержание вопросов программы наблюдения напрямую зависит от конкретных исследовательских задач и особенностей рассматриваемого явления.

Важным элементом статистического наблюдения является понятие критического момента наблюдения. Данное понятие представляет собой момент или отрезок времени, по состоянию на который проводится регистрация значений признаков по каждой единице наблюдения.

В зависимости от целей и задач исследования, особенностей структуры совокупности, предмета исследования критическим моментом может быть дата (день, час), неделя, месяц и т. п.

Период наблюдения представляет собой интервал времени, в течение которого осуществляется сбор данных, заполнение бланков программы наблюдения.

Время наблюдения – это время, в течение которого проводится обследование по разработанной программе.

Формулировка цели исследования предполагает постановку научной проблемы, определение свойств и тенденций явления, которые подлежат анализу.

Задачи исследования — совокупность действий, необходимых для достижения цели исследования.

Объект наблюдения — совокупность социально-экономических явлений и процессов, которые подлежат исследованию, или точные границы, в пределах которых будут регистрироваться статистические сведения. Например, при переписи населения необходимо установить, какое именно население подлежит регистрации — наличное, т. е. фактически находящееся в данной местности в момент переписи, или постоянное, т. е. живущее в данной местности постоянно.

Совокупность (статистическая совокупность) представляет собой множество единиц изучаемого явления, объединенных единой качественной основой, но отличающихся друг от друга отдельными признаками. Таковы, например, совокупность домохозяйств, совокупность семей, совокупность предприятий, фирм, объединений и т. п.

Основными свойствами статистической совокупности является однородность, динамичность и независимость единиц. Совокупность называется однородной, если один или несколько изучаемых существенных признаков ее объектов являются общими для всех единиц. Совокупность, в которую входят явления разного типа, считается разнородной. Совокупность может быть однородна в одном отношении и разнородна в другом [1].

В каждом отдельном случае однородность совокупности устанавливается путем проведения качественного анализа, опираясь на сущность изучаемого процесса или явления.

Динамичность совокупности означает, что появление новых элементов совокупности и исчезновение существовавших ранее не отменяет существования совокупности как объекта наблюдения. Например, совокупность студентов высших учебных заведений не исчезает в результате отчисления одних студентов и восстановления других.

Независимость единиц показывает, что значения признаков одних единиц совокупности не могут быть получены как функция значений других ее единиц. Чтобы определить статистическую совокупность, необходимо ответить на два вопроса: какие единицы входят в совокупность и как эти единицы отличаются друг от друга [1].

В статистике выделяют три вида единиц (рис. 1.4).

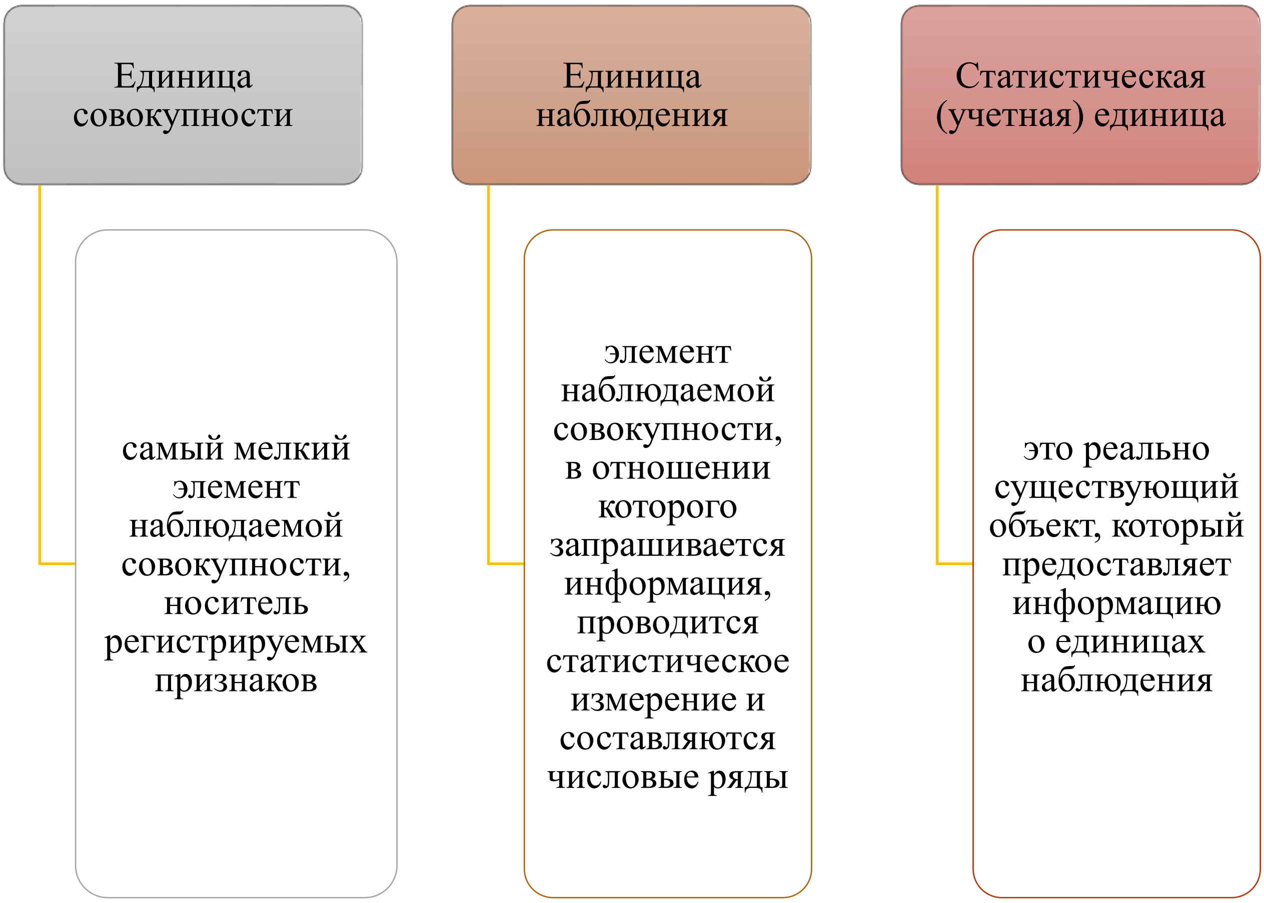



*Рис. 1.4. Основные виды единиц в статистике*

Стоит отметить, что единица совокупности и единица наблюдения могут совпадать (например, при анализе успеваемости студентов группы, каждый студент является единицей совокупности и единицей наблюдения).

Каждая единица наблюдения является набором значений различных признаков, которые определяют качественные особенности единицы совокупности.

Признаки можно разделить на три группы.

К первой группе относятся признаки, которые присущи всем единицам рассматриваемой статистической совокупности и помогают однозначно определить границы наблюдаемой совокупности. Значения признаков данной группы отвечают на вопросы:

- что изучается?
- когда изучается?
- где изучается?

Таким образом, значения данной группы признаков дают ответ на первый вопрос в определении статистической совокупности — какие единицы входят в данную совокупность.

Вторая группа включает в себя признаки, которые позволяют отличить единицы совокупности друг от друга. Данные признаки являются особенными, индивидуальными и неизменными для каждой единицы совокупности. Данная группа признаков отвечает на второй вопрос в определении статистической совокупности — как единицы совокупности отличаются друг от друга.

Третья группа представляет собой признаки как предмет статистического интереса, т. е. случайным образом варьирующие признаки единиц наблюдения (например, объем произведенных услуг какой-либо конкретной фирмой).

Значения таких признаков могут иногда совпадать у отдельных единиц совокупности, а могут быть различными. Именно эта группа признаков является непосредственным предметом изучения в статистическом исследовании [1].

По характеру отображения свойств единиц изучаемой совокупности признаки делятся на две основные группы (рис. 1.5).


*Рис. 1.5. Классификация признаков по характеру отображения свойств единиц изучаемой совокупности*

В статистическом исследовании принято различать пять основных шкал измерения признаков (в порядке повышения точности измерения) (рис. 1.6).

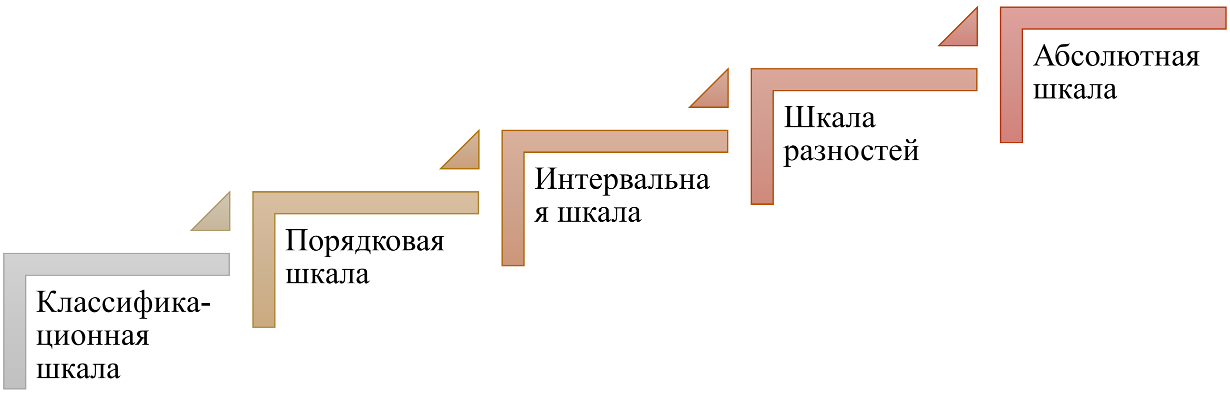



*Рис. 1.6. Основные шкалы измерения в статистике*

Классификационная шкала представляет собой перечень значений атрибутивного признака (например, телефонный справочник). Эта шкала не имеет ни нуля (начала отсчета), ни предпочтений, ни единицы измерения.

В порядковой (ранговой) шкале устанавливаются отношения предпочтений между вариантами значений признака. В данной шкале также нет нуля (начала отсчета) и единицы измерения.

Интервальная шкала устанавливает отношения следования между интервалами значений признака. Имеет произвольный нуль и произвольную единицу измерения.

Шкала разностей устанавливает отношения следования между разностями значений признака. Имеет фиксированную единицу измерения и произвольный нуль (например, шкала времени).

В отличие от шкалы разностей, шкала отношений имеет фиксированный нуль, а единица измерения в ней может быть произвольной.

Абсолютная шкала имеет фиксированный нуль и фиксированную единицу измерения показателя.

На этапе статистического наблюдения проводится сбор данных по разработанной программе. Однако не всякий сбор данных можно назвать статистическим наблюдением. О статистическом наблюдении можно говорить только в том случае, если обеспечивается регистрация устанавливаемых фактов в специальных учетных документах и изучаются статистические закономерности, которые проявляются в большом числе единиц некоторой совокупности. Именно поэтому статистическое наблюдение должно быть планомерным, массовым и систематическим.

К статистическому наблюдению предъявляется целый ряд требований, к которым можно отнести:

- ✓ Достоверность данных
- ✓ Полнота данных
- ✓ Точность данных
- ✓ Практическая ценность статистических данных
- ✓ Сопоставимость и единообразие данных [2].

Существуют различные признаки классификации статистических наблюдений, основные из которых приведены на рис. 1.7.

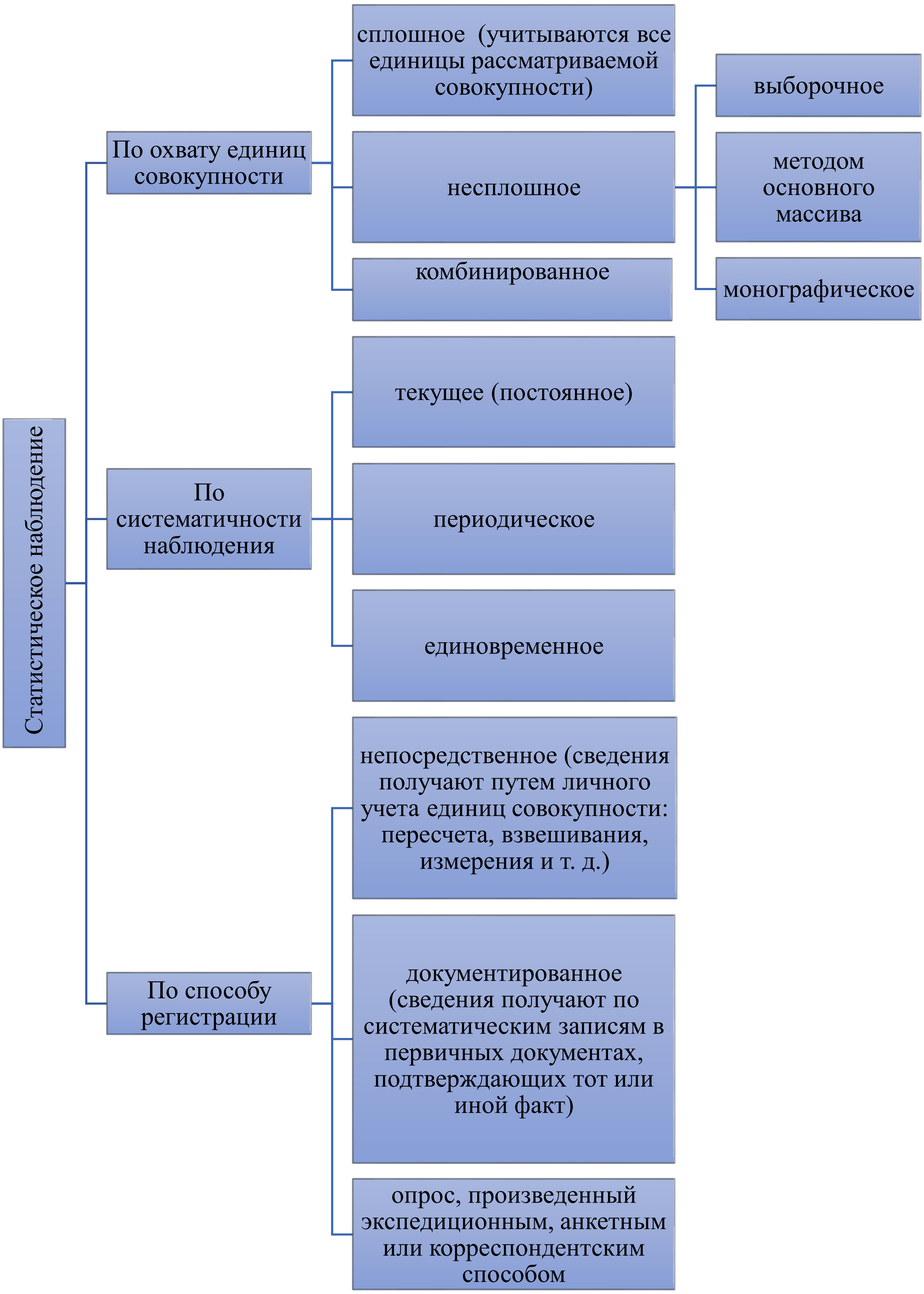



*Рис. 1.7. Классификация статистических наблюдений*

Рассмотрим подробнее разновидности несплошных наблюдений.

Выборочное наблюдение предполагает, что изучается отобранная в случайном порядке часть единиц совокупности с целью характеристики всей совокупности.

Несплошное наблюдение методом основного массива означает, что обследованию подвергается основная часть совокупности, и из генеральной совокупности исключается некоторая часть, о которой заведомо известно, что она не играет большой роли в характеристике всей совокупности [1,2].

Монографическое исследование предполагает изучение отдельных типичных единиц совокупности

В статистической практике используются три организационные формы наблюдения: отчетность, специальное организованное наблюдение и регистр.

Отчетность — это такая организационная форма, при которой единицы наблюдения предоставляют сведения о своей деятельности в виде формуляров регламентированного образца. Особенность отчетности состоит в том, что она обязательна, документально обоснована и юридически подтверждена подписью руководителя.

Специально организованное наблюдение проводится с целью получения сведений, отсутствующих в отчетности, или для проверки ее данных. Примером специально организованного наблюдения является перепись населения. Кроме этого, органы статистики проводят бюджетные обследования, которые характеризуют структуру потребительских расходов и доходов семей.

Регистр представляет собой систему, постоянно следящую за состоянием единицы наблюдения и оценивающую силу воздействия различных факторов на изучаемые показатели. В практике статистики различают регистры населения и регистры предприятий.

Выделяют следующие способы статистического наблюдения:

- ✓ экспедиционный (специально подготовленные регистраторы путем опроса заполняют формуляры, одновременно контролируя правильность получаемых сведений);
- ✓ саморегистрации (работники статистических органов раздают опросные бланки опрашиваемым лицам, инструктируют их, а

затем собирают заполненные формуляры, контролируя полноту и правильность полученных сведений);
- ✓ корреспондентский (статистическими органами организуется специальная сеть корреспондентов из лиц, проживающих на местах, которые проводят наблюдение согласно разработанному бланку и инструкции и сообщают сведения статистическим органам);
- ✓ анкетный (разработанная анкета рассылается кругу лиц и после заполнения возвращается органам, проводящим наблюдения);
- ✓ явочный (предусматривает предоставление сведений в органы, ведущие наблюдение в явочном порядке).

Перед началом третьего этапа статистического исследования необходимо провести арифметический и логический контроль собранных данных с целью устранения ошибок наблюдения. В статистике ошибкой наблюдения называют расхождение между расчетным и действительным значениями исследуемой величины. В зависимости от причин возникновения различают ошибки регистрации и ошибки репрезентативности. Контрольной проверкой собранных данных статистическое наблюдение завершается [1].

Таким образом, в рамках данного параграфа была рассмотрена сущность статистического наблюдения и определено его значение в процессе управления фирмой. На основании вышеизложенного можно сделать вывод о том, что грамотная организация статистического наблюдения является залогом успешного управления фирмой с использованием статистических методов и приемов.

## 1.3. Статистическая сводка и группировка

Одним их из этапов статистического исследования является сводка и группировка данных. Сводка представляет собой упорядочивание и обобщение первичного материала, сводку его в группы и выдачу на этой основе обобщающих характеристик совокупности.

Составными элементами сводки являются: программа сводки; подсчет групповых итогов; оформление конечных результатов сводки в виде таблиц и графиков [1,2].

Различают простую сводку (подсчет только общих итогов) и статистическую группировку, которая сводится к расчленению совокупности на группы по существенному для единиц совокупности признаку.

Группировка позволяет получить такие результаты, по которым можно выявить состав совокупности, характерные черты и свойства типичных явлений, обнаружить закономерности и взаимосвязи. Результаты сводки могут быть представлены в виде статистических рядов распределения.

Пусть из совокупности извлечена выборка, причем $x_1$ наблюдалось $n_1$ раз, $x_2$ – $n_2$ раз, $x_k$ – $n_k$ раз и $\sum n_i = n$ – объем выборки. Наблюдаемые значения $x_i$ называют вариантами, а последовательность вариант, записанных в возрастающем порядке, - вариационным рядом. Числа наблюдений называют частотами, а их отношения к объему выборки $n_i/n = w_i$ – относительными частотами.

Статистическое распределение выборки представляет собой перечень вариант и соответствующих им частот или относительных частот. Статистическое распределение можно задать также в виде последовательности интервалов и соответствующих им частот (в качестве частоты, соответствующей интервалу, принимают сумму частот, попавших в этот интервал) [2].

Статистическим рядом распределения называют упорядоченное распределение единиц совокупности на группы по изучаемому признаку. В зависимости от признака ряды могут быть вариационными (количественными) и атрибутивными (качественными). При построении вариационного ряда с равными интервалами определяют его число групп (n) и величину интервала (h). Число групп можно определить с помощью различных формул. Оптимальное число групп может быть определено по формуле Стерджесса (1.1):

$$n = 1 + 3{,}322 * \lg N, \qquad (1.1)$$

где n – число групп,
N – число единиц совокупности.

В зависимости от исследовательских целей можно использовать равные и неравные интервалы (в последнем случае – равномерно возрастающие или убывающие) открытые и закрытые. Величина равного интервала рассчитывается по формуле (1.2):

$$i = \frac{x_{max} - x_{min}}{n}, \qquad (1.2)$$

где i – длина интервала;

$x_{max}$ – максимальное значение признака в рассматриваемой совокупности;

$x_{min}$ – минимальное значение признака в рассматриваемой совокупности.

При проведении анализа вариационных рядов с неравными интервалами применяется показатель плотности распределения признака. Он может быть рассчитан как частное частоты или частости каждого интервала к его величине.

Для вариационного ряда возможен расчёт накопленных частот. По значению данного параметра можно сформулировать вывод о том, какое число единиц в совокупности имеет значение признака не выше того значения, которое соответствует выбранной величине накопленной частоты.

Группировка является процессом образования групп единиц совокупности однородных в каком-либо отношении, а также имеющих одинаковые или близкие значения группировочного признака.

Группировки бывают простые и комбинационные. Образование простых группировок осуществляется по какому-либо единому признаку, в том время как комбинационная группировка предполагает сочетание двух и более группировочных признаков. Основную задачу метода группировки можно определить, как выявление и выделение основных типов явлений, определение структуры совокупности, изучение взаимосвязи признаков [2].

Классификация видов группировок зависимости от цели и задач исследования приведена на рис. 1.8.


*Рис. 1.8. Виды группировок*

Если группировка используется для того, чтобы охарактеризовать качественные особенности и различия между типами явлений, принято говорить о типологической группировке.

Данный вид группировки активно используется в различных исследованиях социального и экономического характера.

Если группировка производится с целью выявления состава однородной в качественном отношении совокупности по какому-либо признаку, применяется структурная группировка.

В экономике фирмы примером применения данных видов группировок можно назвать группировку организаций по признаку численности рабочей силы, проценту реализации плана и прочее.

Аналитическая группировка применима в том случае, если основная исследовательская задача состоит в изучении взаимосвязи между процессами, явлениями или признаками. В результате проведения данного вида группировки определяются причинно-следственные связи в рассматриваемых объектах [1,2].

Таким образом, сводка и группировка данных является важнейшим этапом реализации статистического исследования. В управлении фирмой применение тех или иных методов и приемов опирается на результаты первичной обработки статистических данных. Именно поэтому грамотная реализация данного этапа является залогом успешности проводимых управляющих воздействий.

## 1.4. Представление данных

Грамотное представление статистических данных является важнейшим этапом исследования. При этом удобство пользования данными существенно упрощает задачу любого исследования с применением статистических методов или приемов. Статистические данные могут быть представлены в графическом, текстовом и числовом виде.

Наиболее рациональным способом представления итоговых результатов сводки и группировки является формирование статистических таблиц. Основное преимущество использования таблиц состоит в возможности компактного, удобного и рационального представления нужных данных. При этом важное значение уделяется грамотному выделению статистического подлежащего и сказуемого в данной форме. Как правило, строки в таблицах используются для отображения подлежащего, а столбцы – для сказуемого [2].

В зависимости от строения подлежащего принято выделять три вида статистических таблиц (рис. 1.9).

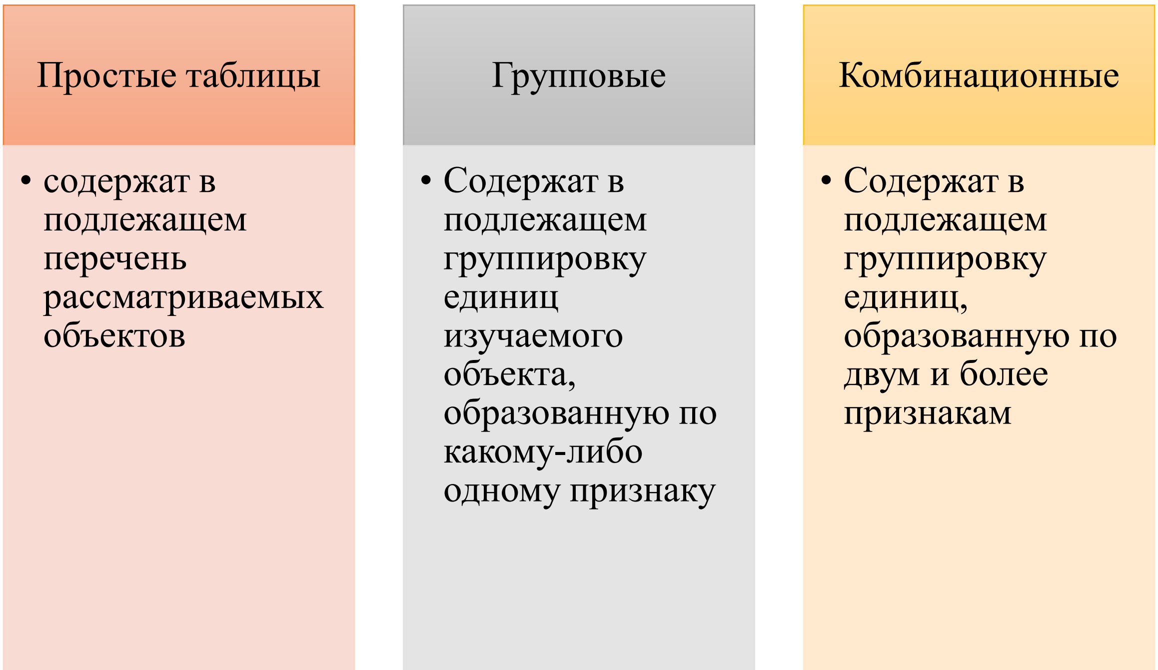



*Рис. 1.9. Виды статистических таблиц по типу строения подлежащего*

Обобщая все вышесказанное, можно сделать вывод о том, что статистическое подлежащее по своей сути представляет объект конкретного исследования, в сказуемое – систему показателей, которые позволяют охарактеризовать нужные стороны рассматриваемого процесса или явления.

Важно помнить, что наглядности и рациональности представляемых данных можно достичь только в том случае, если четко и грамотно сформулирован предмет и объект статистического исследования. Излишняя детализация представляемых данных в формируемой таблице только усложняет процесс интерпретации результатов сводки и группировки. Если представляется задача группировки большого объема данных, отражающих совокупность различных статистических показателей, на практике зачастую используют разбивку исходных параметров на несколько статистических таблиц, каждая из которых отображает блок рассматриваемых статистических данных по какому-либо обобщенному вопросу [1,2].

Существует целый ряд практических рекомендаций по составлению и оформлению таблиц, которые позволяют значительно рационализировать представляемые сведения (рис. 1.10).

Таблица по возможности должна отличаться краткостью и содержательностью

В каждой таблице должно содержаться подробное название, из которого можно сделать вывод о рассматриваемом в таблице круге вопросов, геограчиеских границах рассматриваемой совокупности, анализируемом интервале времени

Таблица может содержать примечания, содержащие источники данных, более подробное описание показателей и прочие появнения

При оформлении таблиц обычно применяются такие условные обозначения: знак тире (–) – когда явление отсутствует; x - если явление не имеет осмысленного содержания; многоточие (…) – когда отсутствуют сведения о размере явления (или делается запись «Нет сведений»)

Если числовое значение имеющихся сведений меньше принятой в таблице точности, оно выражается дробным числом (0,0)

Округленные числа приводятся в таблице с одинаковой степенью точности (до 0,1, до 0,01 и т. п.)

Если в таблице приводятся проценты роста, то во многих случаях целесообразно проценты от 300 и более заменять отношениями в разах, например, писать не «1 000 %», а «в 10,0 раз»

*Рис. 1.10. Основные правила составления и оформления статистических таблиц*

В таблицах важно уточнять, каковы используемые единицы измерения. Если единицы измерения для различных ячеек таблицы отличаются друг от друга, в верхних или боковых заголовках обязательно необходимо уточнять, в каких единицах измерения приводятся рассматриваемые статистические данные.

Применение графиков в процессе представления анализируемых статистических параметров дает возможность наглядного и выразительного отображения показателей, характеризующих рассматриваемый процесс или явление. Кроме того, грамотное графическое представление способно существенно облегчить восприятие и продемонстрировать наличие или отсутствие взаимосвязи между параметрами, а также характер данной взаимосвязи. В зависимости от того, что именно демонстрируется на графике (сравнение параметров, динамика их изменения, степень распространения процессов в конкретном подразделении фирмы и т.д.), могут быть использованы различные их виды [2].

Современные автоматизированные компьютерные программы обладают широким спектром возможностей для грамотного отображения на графике рассматриваемых процессов и явлений. Основная задача пользователя при этом сводится к грамотному определению типа представляемых данных и выбору инструментов, которые позволяют исследователю отобразить на графике именно те свойства и тенденции, которые нуждаются в отображении.

По способу построения графиков выделяют три основных вида данных объектов (рис. 1.11).

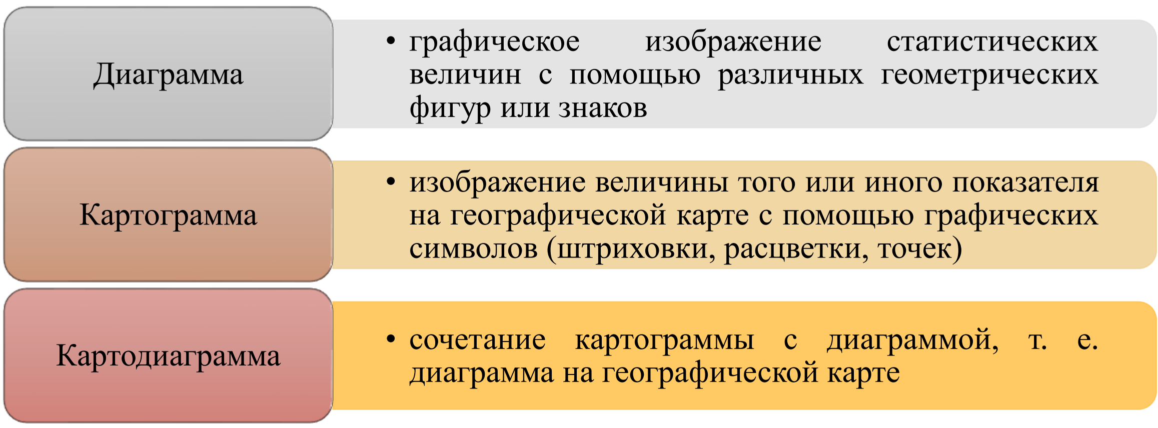



*Рис. 1.11. Классификация графиков по способу построения*

В зависимости от применяемых графических образов среди диаграмм различают столбиковые, плоскостные, объемные, линейные и др.

Для грамотного графического отображения вариационных рядов применяются линейные и плоскостные диаграммы, построенные в прямоугольной системе координат. Вариационный ряд можно изобразить, как и любой ряд значений аргумента и функции, используя прямоугольную систему координат и строя точки с координатой ($x_1$ , $f_1$ ); ($x_2$ , $f_2$ ); (…); ($x_n$ , $f_n$ ) в виде полигона, гистограммы, кумулятивной кривой (кумуляты), кривой Лоренца [2].

Полигон – графическое изображение дискретного вариационного ряда распределения. Полигон представляет собой замкнутый многоугольник, абсциссами вершин которого являются значения варьирующегося признака, а ординатами — соответствующие им частоты (рис. 1.12).

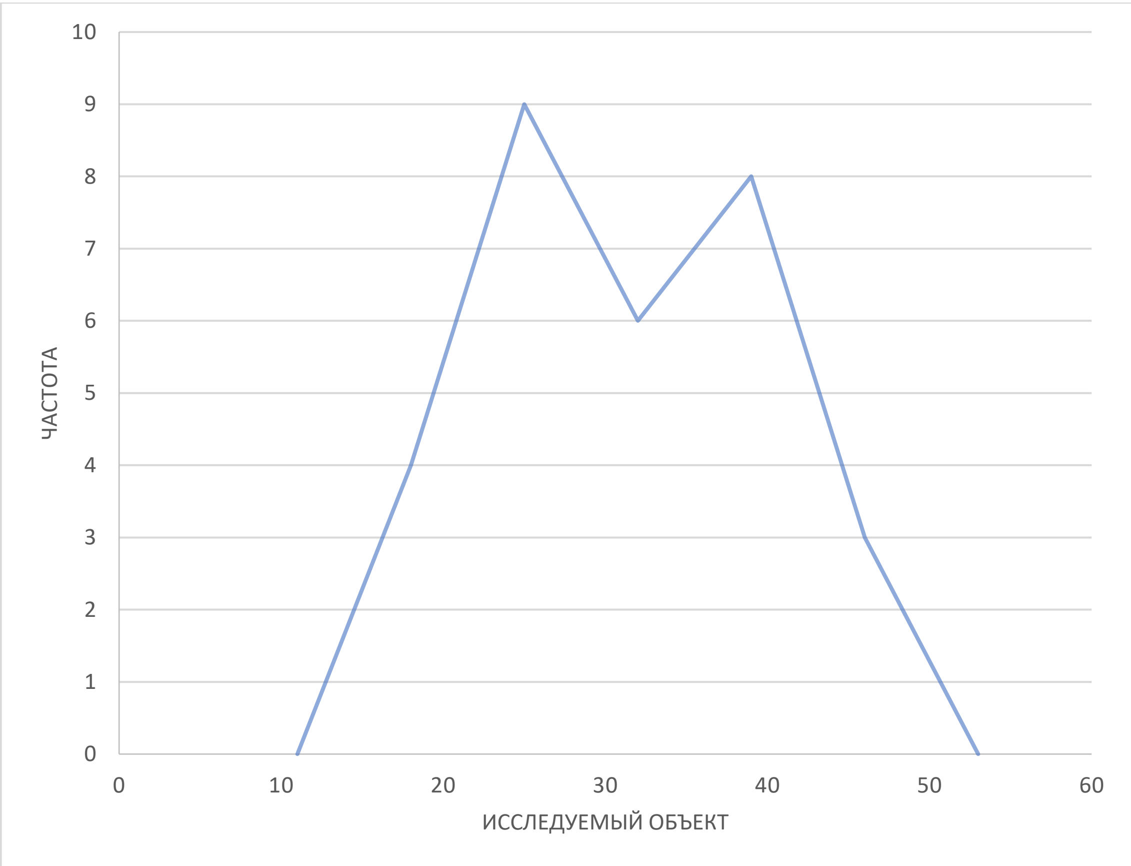



*Рис. 1.12. Пример полигона частот*

Гистограмма представляет собой графическое изображение интервального вариационного ряда. При ее построении на оси абсцисс откладывают не значения признака, а границы интервалов значений признака. По оси ординат откладывают частоты (частости) или плотности распределения (все зависит от вида интервального ряда). Если ряд интервальный с равными интервалами, то на оси ординат откладывают частоты (частости), т. е. строят прямоугольники с высотой, равной частоте (частости) заданного интервала. Если ряд интервальный с неравными интервалами, то строят гистограмму плотностей распределения, поскольку в ряду с неравными интервалами именно плотность дает точное представление о количестве единиц в каждом из интервалов. Площадь всей гистограммы, таким образом, численно равна сумме частот или численности единиц в совокупности. Пример гистограммы приведен на рис. 1.13.

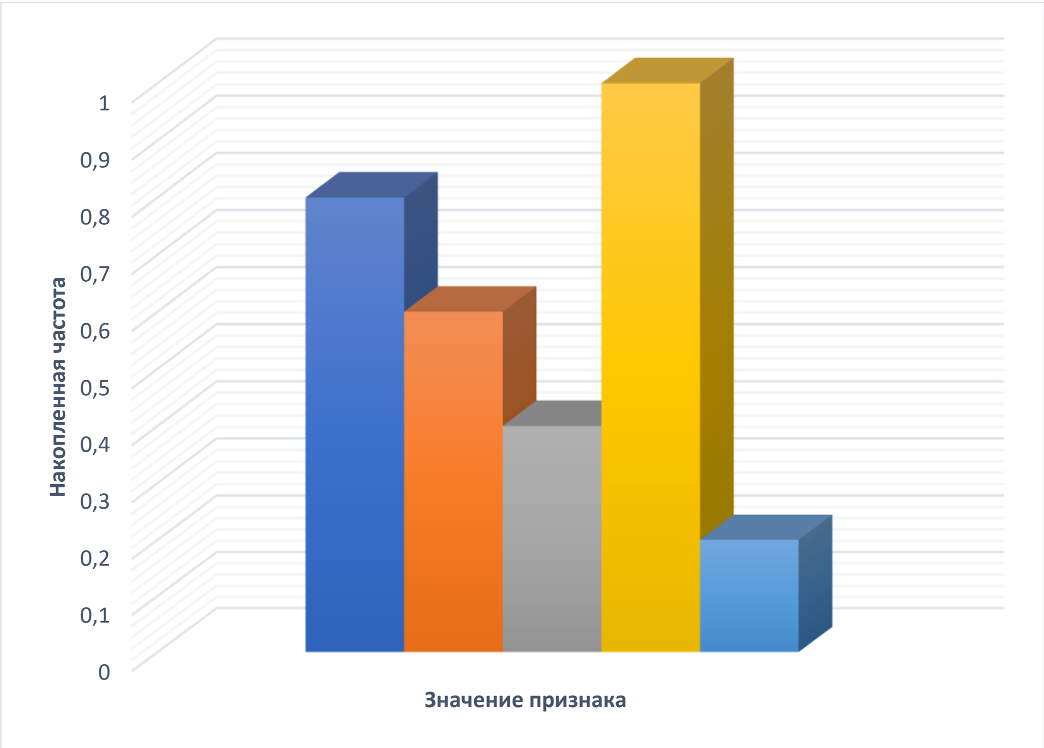



*Рис. 1.13. Пример гистограммы*

Кумулятивная кривая (кумулята) – кривая, характеризующая динамику накопленной частоты или частости. По оси абсцисс откладывают варианты значений признака (в интервальном ряду – верхние границы интервалов), а на оси ординат – соответствующие накопленные частоты или частости. Полученные точки соединяют отрезками и получают график, который называется кумулятой или кумулятивной кривой (рис. 1.14) [2].

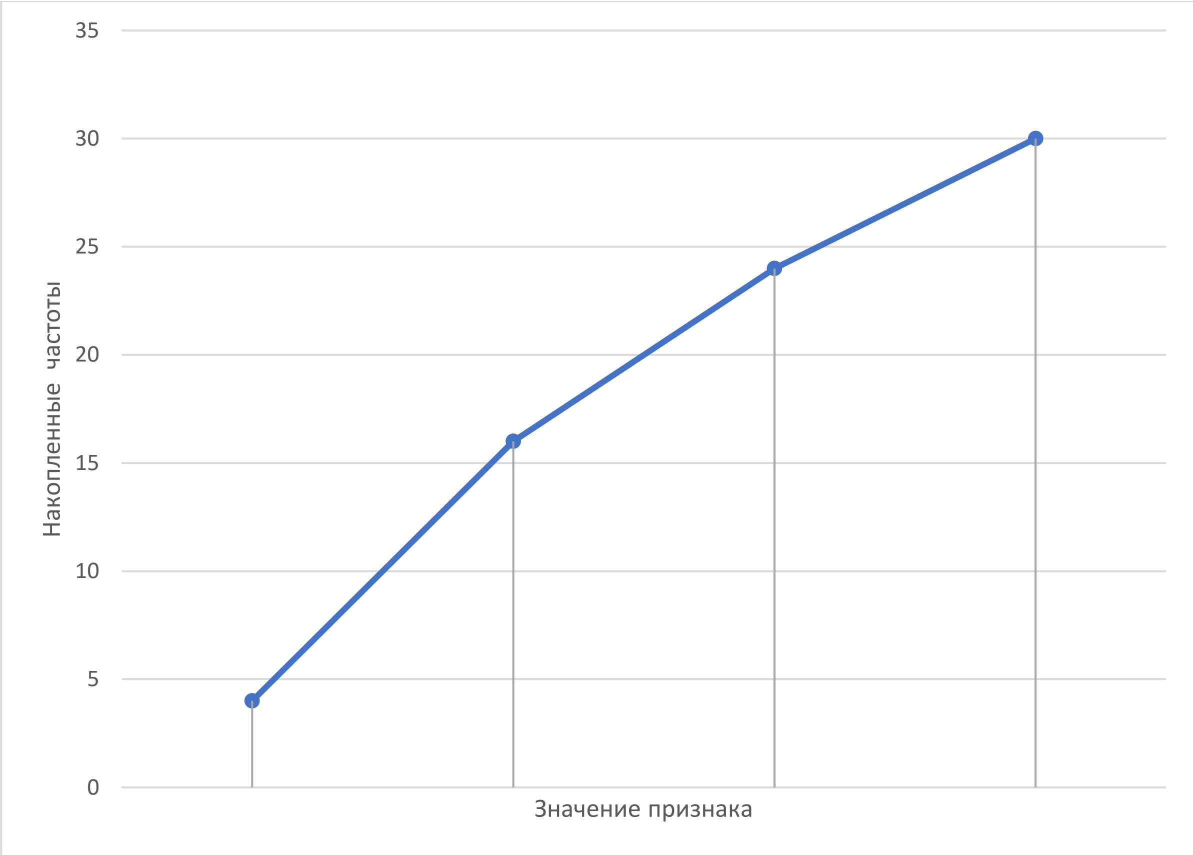



*Рис. 1.14. Пример кумуляты*

Кривая Лоренца — это график, который используется на практике для характеристики уровня относительной концентрации тех или иных явлений в совокупности. Построение такого графика предполагает указание значений накопленных частостей выделенных групп и значений накопленных долей признака в общем объеме совокупности (в процентах) (рис. 1.15).

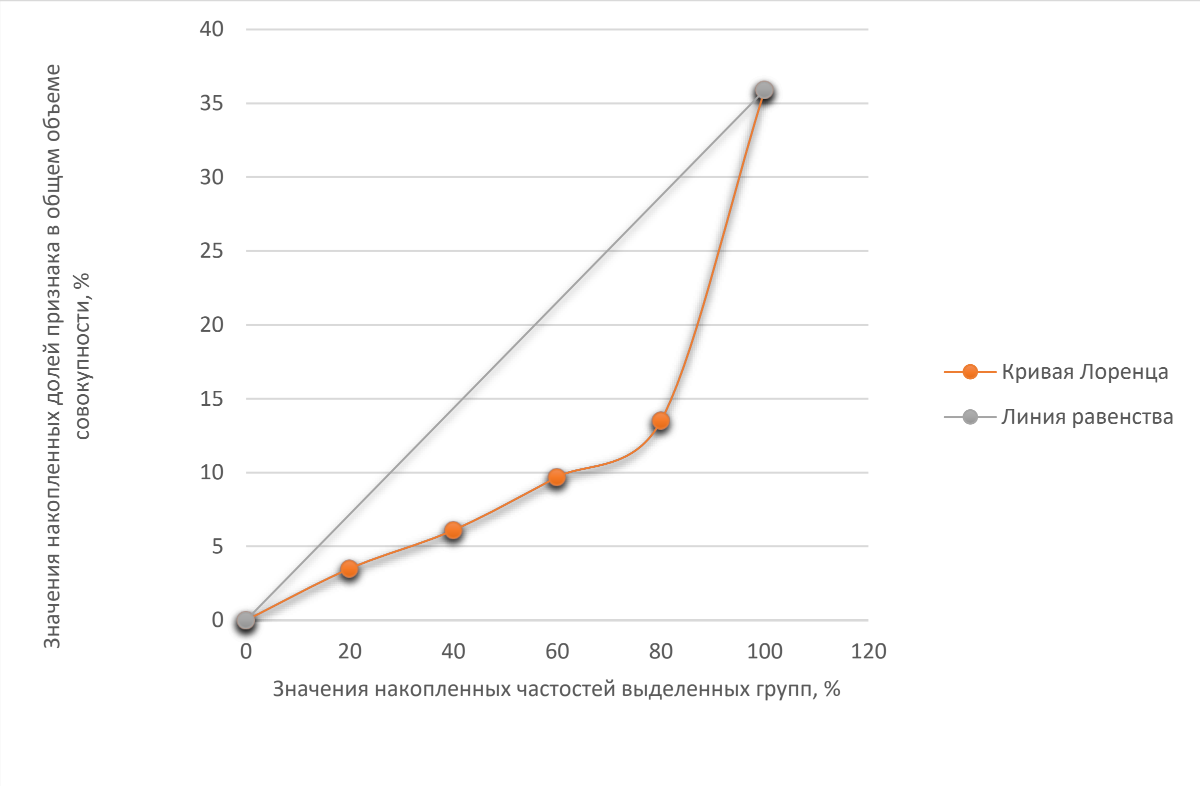



*Рис. 1.15. Пример кривой Лоренца*

Таким образом, были рассмотрены основные приемы грамотного представления данных при проведении статистического наблюдения. На основании вышеизложенного можно сделать вывод о том, что выбор инструмента представления данных оказывает существенное влияние на процедуру дальнейшего исследования показателей фирмы с использованием статистических методов.

## 1.5. Роль прогнозирования в управлении фирмой

Прогнозирование в процессах управления фирмой позволяет осуществлять контроль внешних факторов. Для того, чтобы процесс прогнозирования был максимально эффективным, необходимо четко понимать, где находится организация, где она должна находиться в будущем и какие для этого необходимы управленческие воздействия.

Процесс прогнозирования представляет собой выявление возможных альтернатив функционирования социально-экономических систем, с целью обоснованного выбора стратегии дальнейшего развития и принятия оптимального управленческого решения [3].

Термин "прогноз" происходит от греческого "prognosis" и обозначает предсказание, предвидение о развитии чего-либо, базирующееся на определенных данных [4].

Основные функции прогноза представлены на рис. 1.16

Анализ социально-экономических и научно-технических процессов и тенденций, объективных причинно-следственных связей этих явлений в конкретных условиях, в том числе оценка сложившейся ситуации и выявление проблем хозяйственного развития

Оценка этих тенденций в будущем; предвидение новых условий и проблем, требующих разрешения

Выявление альтернативы развития в перспективе; накопление информации и расчётов для обоснования выбора и принятия оптимального управленческого решения, в том числе в плане

*Рис. 1.16. Функции прогноза*

Следует четко понимать разницу между планом и прогнозом.

План представляет собой документ, содержащий в себе систему показателей и комплекс мероприятий, направленных на решение конкретных социально-экономических задач.

Прогноз является научно-обоснованным суждением, отражающим возможные состояния анализируемого объекта в будущем, а также характеристику альтернативных путей и конкретных сроков их достижения.

Основное отличие между планом и прогнозом состоит в том, что план носит обязательный характер, в то время как для прогноза характерно вероятностное содержание. Таким образом, план направлен на реализацию конкретного решения, выбранного в ходе процесса планирования в качестве оптимального. Прогноз нацелен на выявление всех возможных вариантов развития системы, то есть предполагает вероятностную оценку состояния исследуемой социально-экономической системы в будущем [5 – 7].

Цели планирования и прогнозирования также различны: если планирование нацелено на принятие и практическую реализацию управленческих решений с целью достижения конкретного ожидаемого результата, то прогнозирование ставит своей целью создание научных предпосылок для принятия управленческих решений.

По сути, прогнозирование является этапом, предшествующим процессу планирования. Прогнозы должны содержать в себе возможные варианты развития социально-экономических систем в случае реализации конкретных планов и стратегий. При этом прогнозирование должно учитывать влияние внешних факторов, управление которыми не представляется возможным [8].

Прогноз отличается от плана не только своей многовариантностью. Планирование возможно только для тех процессов, управление которыми представляется возможным для человека. Применение прогнозирования возможно и для неуправляемых или частично управляемых процессов.

Таким образом, прогноз – это вероятностное описание желаемого или возможного, а план – это директивное решение, содержащее в себе описание мероприятий, направленных на достижение желаемого или возможного [9].

Существуют различные основания для классификации прогнозов [10].

В частности, прогнозы можно подразделить в зависимости от целей, задач, объектов, времени упреждения, источников информации и т. д. (рис. 1.17).

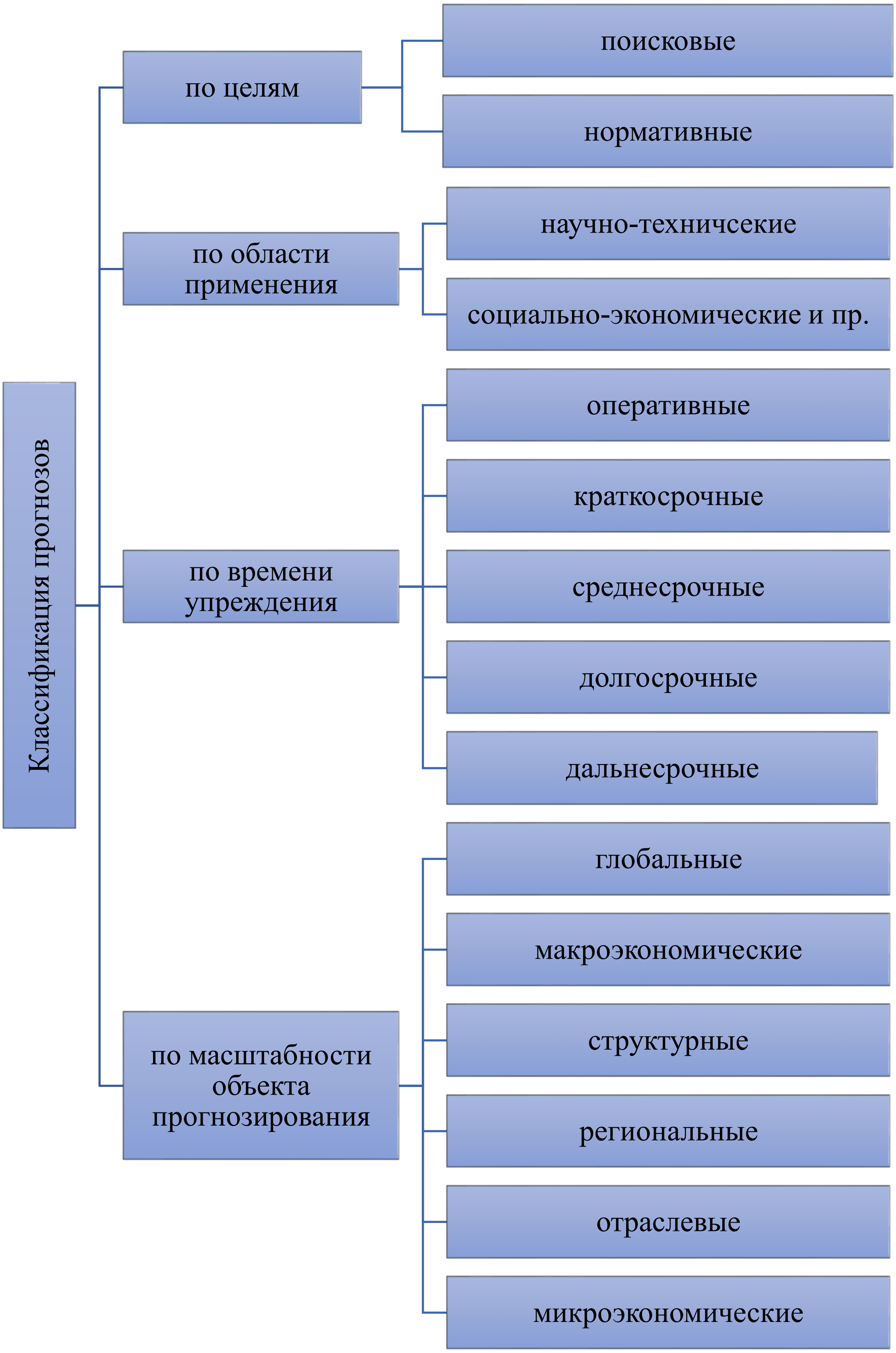



*Рис. 1.17. Классификация прогнозов*

Поисковый прогноз ориентирован, прежде всего, не на достижение конкретной цели, а на анализ возможных путей развития социально-экономического явления или процесса при условии сохранения существующих тенденций развития.

Нормативный прогноз базируется на достижении заранее намеченных целей и задач. Данный вид прогноза ориентирован на поиск возможных путей достижения заданной цели и на обоснование конкретных сроков достижения желаемого результата.

Научно-технические прогнозы рассматривают достижения научно-технического прогресса, развитие фундаментальных и прикладных исследований, новых видов техники и технологии, определяют последствия научно-технического прогресса.

Социально-экономические прогнозы исследуют вопросы динамики уровня жизни населения, доходов, потребления населением продуктов питания и непроизводственных товаров, развития отраслей социальной инфраструктуры, демографии, занятости населения и т. д. [11].

Стоит отметить, что по области применения существуют и другие виды прогнозов, и их классификация напрямую зависит от специфики рассматриваемой области.

Период упреждения прогноза - это отрезок времени от момента, для которого имеются последние статистические данные об изучаемом объекте, до момента, к которому относится прогноз.

Период упреждения прогноза зависит от специфики и особенностей изучаемого объекта исследования, от интенсивности изменения показателей, от продолжительности действия выявленных тенденций и закономерностей, от длины временного ряда и от многих других факторов.

По времени упреждения прогнозы подразделяются на оперативные, краткосрочные, среднесрочные, долгосрочные и дальнесрочные.

Оперативный прогноз имеет период упреждения до одного месяца, краткосрочный – от одного месяца до года, среднесрочный - от года до пяти лет, долгосрочный - от пяти до пятнадцати - двадцати лет, дальнесрочный - свыше этого периода.

Глобальные прогнозы предполагают рассмотрение наиболее общих тенденций и закономерностей в мировом масштабе.

В рамках макроэкономических прогнозов анализируются процессы и явления на уровне конкретного государства.

Структурные прогнозы направлены на исследование экономического развития в разрезе конкретной отраслевой структуры анализируемого объекта.

Соответственно объектом регионального прогнозирования являются отдельные регионы, а для отраслевых прогнозов характерно исследование развития отраслей.

Микроэкономические прогнозы строятся для отдельных предприятий, производств и других объектов прогностической деятельности на микроуровне.

Методы прогнозирования — это совокупность приемов и способов мышления, которые позволяют на основе анализа ретроспективных данных об исследуемом объекте вывести суждения определенной достоверности относительно будущего развития объекта.

По оценкам отечественных и зарубежных ученых, в настоящее время насчитывается более ста методов прогнозирования, однако на практике регулярно используются несколько десятков базовых методов [12].

Один из наиболее важных признаков методов прогнозирования — степень формализации, которая достаточно полно охватывает прогностические методы.

**По степени формализации** методы экономического прогнозирования можно разделить на интуитивные и формализованные (рис. 1.18).

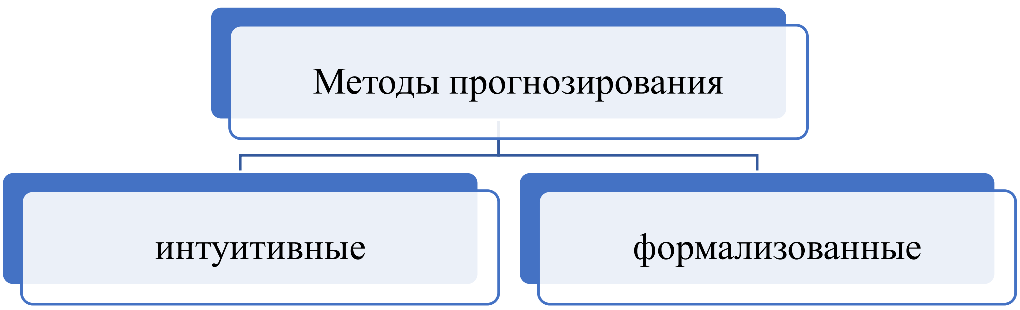



*Рис. 1.18. Классификация методов прогнозирования по степени их формализации*

Интуитивные методы прогнозирования имеют дело с суждениями и экспертными оценками. Их применено, как правило, обусловлено невозможностью учета влияния многих факторов из-за значительной сложности объекта прогнозирования. Также интуитивные методы применяют в том случае, если система предельно проста, и в математическом описании не нуждается.

Сущность метода экспертных оценок заключается в проведении экспертами интуитивно-логического анализа проблемы с количественной оценкой суждений и формальной обработкой результатов. При этом обобщенное мнение экспертов принимается как решение проблемы. Использование интуиции, логического мышления и количественных оценок с формальной обработкой позволяет получить эффективное решение проблемы. Особенностями метода экспертных оценок являются, во-первых, научно обоснованная организация проведения всех этапов экспертизы, обеспечивающая наибольшую эффективность работы на каждом из этапов; во-вторых, применение количественных методов как при организации экспертизы, так и при оценке суждений экспертов и формальной групповой обработке результатов. Наиболее часто эти методы используются при рассмотрении социально-экономических проблем, где невозможно выработать формализованную прогностическую модель.

Посредством метода экспертных оценок решаются следующие задачи:

- составляются перечни возможных событий за определенный промежуток времени по исследуемой проблеме;
- определяются наиболее вероятные интервалы времени совершения совокупности событий;
- определяются цели и задачи с упорядочением их по степени важности;
- разрабатываются альтернативные варианты решения проблем с оценкой их предпочтения;
- разрабатываются альтернативные варианты распределения ресурсов с ранжированием их очередности;
- разрабатываются альтернативные варианты принятия решений в определенной ситуации с оценкой их предпочтительности; и др.

Организация процедуры экспертной оценки включает несколько направлений:

- формирование экспертной группы; подготовку и проведение экспертизы;
- статистическую обработку полученных результатов опроса.

В зависимости от организации экспертной оценки и формы опроса различают методы индивидуальных и коллективных экспертных оценок.

Методы индивидуальных экспертных оценок включают в себя: метод анкетирования и интервьюирования, аналитический метод, метод написания сценария и др.

Метод анкетирования заключается в предъявлении экспертам опросных листов-анкет, на которые они должны дать ответы в письменной форме. Интервьюированием является устный вопрос эксперта членом группы управления интервьюером.

Все вопросы анкет можно классифицировать по содержанию и по форме. По содержанию вопросы делятся на три группы:

- объективные данные об эксперте;
- основные вопросы по сути анализируемой проблемы;
- дополнительные вопросы, позволяющие выявить источники информации и аргументации эксперта, самооценку компетентности эксперта.

По форме основные вопросы делятся на открытые, или свободные, закрытые и с «веером» ответов, а также на прямые и косвенные. Закрытый вопрос задается в форме, предполагающей лишь три возможных ответа – «да», «нет», «не знаю». Вопрос с «веером» ответов предоставляет эксперту возможность выбора одного из предлагаемых ответов, например, срока реализации определенной научно-технической идеи из ряда перечисленных сроков. К этой же форме относятся вопросы-задания на ранжирование заданных объектов, на оценку их весов, значимости в баллах на оценку вероятности некоторого события.

Кроме рассмотренных трех форм вопросов, можно ввести еще одну форму, промежуточную между открытыми вопросами и вопросами с «веером» ответов. Это вопрос-задание на проведение морфологического анализа, на построение дерева целей, альтернатив.

При задании вопроса в такой форме эксперту может быть предоставлено право дать две или три оценки одного объекта – минимальную, среднюю, максимальную (или оптимистическую, среднюю, пессимистическую).

При задании вопросов в любой форме эксперт должен быть поставлен в известность, что он вправе выдвинуть новые вопросы и дать на них ответы, а также назвать экспертов, не включенных в число опрашиваемых, которые способны дать ответы на вопросы анкеты или вопросы, выдвинутые им самим. Кроме того, эксперт должен изложить свои замечания и советы по форме и содержанию анкет.

Получение прогнозных оценок методом «интервью» осуществляется посредством беседы, в ходе которой интервьюер ставит вопросы эксперту по заранее разработанной программе. Одновременно может производиться опрос нескольких экспертов, однако в этом случае есть опасность потери самостоятельности экспертов.

От очного анкетирования этот метод отличается тем, что при интервью эксперт дает ответы в устной форме на устные вопросы, точное содержание которых до опроса ему, как правило, не было известно, хотя тематика интервью могла быть сообщена ему заранее.

Достоинством интервью является непрерывный живой контакт интервьюера и опрашиваемого, что позволяет быстро получить большое количество информации и всесторонне, хотя и поверхностно, осветить объект экспертизы.

Недостатками интервью являются возможность сильного влияния интервьюера на ответы эксперта, отсутствие времени для глубокого продумывания ответов, а также высокие требования к опрашивающему и большое время, расходуемое на опрос всего состава экспертов.

Получение прогнозных оценок аналитическим методом осуществляется посредством логического анализа какой-либо прогнозируемой ситуации. Он предполагает самостоятельную работу эксперта над анализом тенденции, оценкой состояния и путей развития прогнозируемого объекта.

Метод написания сценария основан на определении логики процесса или явления во времени при различных условиях. Основное назначение сценария — определение генеральной цели развития объекта прогнозирования, выявление основных факторов фона и формулирование критериев для оценки верхних уровней дерева целей. Ценность сценария тем выше, чем меньше степень неопределенности, т.е. чем больше степень согласованности мнений экспертов в осуществимости событий, в развитии процесса и т.д.

Основным преимуществом рассмотренных выше методов являются возможность максимального использования индивидуальных способностей экспертов и незначительность психологического давления, оказываемого на отдельных работников.

Методы коллективных экспертных оценок - группа методов коллективных экспертных оценок, основанных на том, что при коллективном мышлении, во-первых, выше точность результата и, во-вторых, при обработке индивидуальных независимых оценок, выносимых экспертами, могут возникнуть продуктивные идеи.

Существуют следующие разновидности методов коллективных экспертных оценок: метод «комиссий», метод Дельфи, метод «коллективной генерации идей» («мозговая атака»), метод морфологического анализа и др.

Метод «комиссий» предполагает создание рабочей группа, в функции которой входят: назначение экспертов, проведение опроса, обработка материалов, анализ результатов коллективной экспертной оценки. В ходе работы уточняются основные направления развития объекта, а также составляется матрица, отражающая генеральную цель, подцели и средства их достижения, т.е. направления научных исследований и разработок, результаты которых могут быть использованы для достижения цели.

Затем разрабатываются вопросы для экспертов. Это может быть перечень или таблица, но содержание вопросов должно определяться спецификой прогнозируемого объекта. Далее следуют проведение опроса экспертов и статистическая обработка материалов, которые характеризуют обобщенное мнение и степень согласованности индивидуальных оценок экспертов. Они служат исходной базой для синтеза прогнозных гипотез и вариантов развития исследуемого явления или процесса. Методика представляет собой совокупность оценок относительной важности, назначенных экспертами каждого из оцениваемых направлений исследований и разработок, выражающихся в баллах и принимающих значения от 0 до 1, от 0 до 10, от 0 до 100 и т. д.

Эти оценки по определенному вопросу сводятся в таблицу, строки которой соответствуют направлениям исследований, а столбцы — порядковым номерам экспертов.

Метод Дельфи – один из наиболее распространенных методов экспертных оценок. Его основными особенностями являются: анонимность экспертов, полный отказ от личных контактов экспертов и коллективных обсуждений; многотуровая процедура опроса экспертов посредством их анкетирования; обеспечение экспертов информацией, включая и обмен ею между экспертами, после каждого тура опроса при сохранении анонимности оценок, аргументации и критики; обоснование ответов экспертов по запросу организаторов.

Метод «коллективной генерации идей» включает два элемента: выявление вероятностных вариантов развития объекта прогнозирования и их оценка. При «мозговой атаке» сначала активизируется творческий потенциал специалистов, что находит отражение в генерации определенной идеи. Затем следует процесс деструирования (разрушения, критики) этой идеи и формулируется контридея. Это позволяет за короткое время путем вовлечения всех экспертов в активный творческий процесс получить продуктивные результаты.

Метод морфологического анализа – применяется при прогнозировании сложных процессов экспертный метод систематизированного обзора всех возможных комбинаций развития отдельных элементов исследуемой системы. Этот метод построен на полных и строгих классификациях объектов, явлений, свойств и параметров системы, позволяющих строить и оценивать возможные сценарии ее развития в целом.

Этой цели служит прием систематизированного охвата информации с последующим исследованием ее по методу «морфологического ящика». Последний строится в виде дерева или матрицы, в клетках которых помещены соответствующие характеристики объекта. Последовательное соединение одного из параметров первого уровня с одним из параметров последующего уровня представляет собой одно из возможных состояний объекта или решений проблемы. В результате создается новая информация об изучаемом объекте и вырабатывается оценка всех возможных альтернатив его состояния.

Результат реализации формализованного метода состоит в построении конкретной модели прогнозирования. То есть определяется конкретная математическая зависимость, позволяющая вычислить прогнозное значение [13].

Особое место в классификации методов прогнозирования занимают комбинированные методы, объединяющие в себе различные методы прогнозирования. Использование комбинированных методов особенно актуально для сложных социально-экономических систем, когда при разработке прогноза показателей каждого элемента системы могут быть использованы различные сочетания методов прогнозирования. Разновидностью комбинированных методов можно считать эконометрическое моделирование.

Модель прогнозирования есть функциональное представление, адекватно описывающее исследуемый процесс и являющееся основой для получения его будущих значений [14].

Классификация моделей прогнозирования представлена на рис. 1.19.

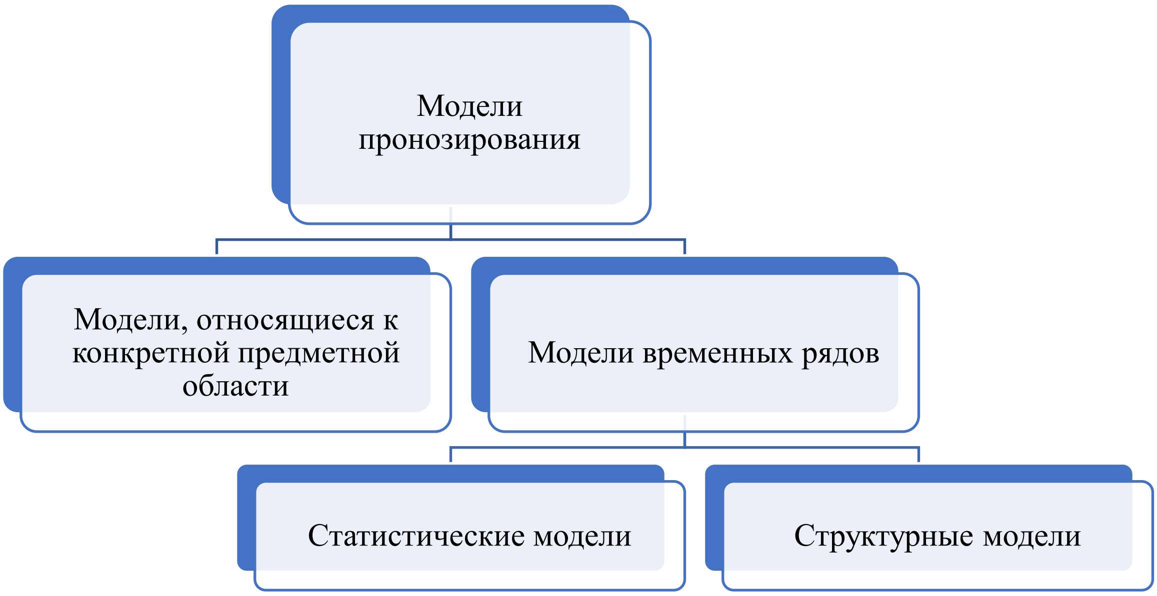



*Рис. 1.19. Классификация моделей прогнозирования*

Модели, относящиеся к конкретной предметной области, предполагают использование законов, свойственных определенным предметным областям. Для такого рода моделей требуется индивидуальный подход к построению и практическому применению, что делает невозможным их использование в рамках других областей прогнозирования [15].

Модели временных рядов направлены на поиск зависимостей внутри анализируемого процесса. Данный вид моделей базируется на знании прошлых значений объекта прогнозирования. Отличительной чертой моделей временных рядов является их универсальность, то есть возможность применения в рамках различных предметных областей. Более того, общий вид моделей не зависит от природы рассматриваемого в рамках прогностической деятельности явления или процесса [16].

В статистических моделях прогнозирования функциональная зависимость между будущими и фактическими значениями временного ряда, а также внешними факторами, если таковые учитываются, задана аналитически, то есть в формульном виде. Примерами статистических моделей являются регрессионные, авторегрессионные, модели экспоненциального сглаживания и др.

Структурные модели предполагают описание будущего значения в зависимости от прошлых параметров в виде некоторой структуры, а также конкретных правил перехода по ней. К структурным моделям можно отнести нейронные сети, цепи Маркова, различные классификационные деревья и др.

Таким образом, выбор адекватных методов и моделей прогнозирования зависит от целей исследования, специфики процесса, а также имеющихся данных о прошлом состоянии рассматриваемого объекта.

Прогнозирование является эффективным только в том случае, если субъект может всесторонне оценить ситуацию и выбрать эффективные инструменты прогнозирования в рамках решения конкретной исследовательской задачи. В экономике фирмы в процессе прогнозирования крайне важно учитывать специфику внешней среды, поскольку грамотное управленческое воздействие предполагает тщательный анализ факторов, оказывающих влияния на деятельность каждой конкретной фирмы.

## 1.6. Вероятностные характеристики, применяемые в управлении фирмой

Эксперимент или случайный опыт представляет собой процесс, для которого возможны различные исходы, т.е. невозможно однозначное предсказание результата. Результат случайного опыта X называется случайной величиной. Непостоянство результата такого опыта может быть связано с наличием случайных ошибок измерений или со статистической природой самой измеряемой величины. Отдельные значения, принимаемые случайной величиной, обозначаются $X_i$. Любая функция от $X_i$ также является случайной величиной [16].

Случайная величина может быть дискретной (значения которой могут быть либо конечными, либо счетными) или непрерывной (в результате испытания принимает все значения из некоторого числового промежутка). Также случайные величины могут быть одномерными (находящиеся в зависимости от единственной переменной) или многомерные (зависящие от двух и более переменных) [17].

Полной характеристикой случайной величины X с вероятностной точки зрения является ее закон распределения, т.е. заданная в той или иной форме связь между возможными значениями случайной величины и вероятностями их появления.

Естественная форма закона распределения дискретной случайной величины X, которая может принимать значения $X_1, X_2 .... X_n$, задается таблицей (1.3):

$$X = \begin{pmatrix} X_1 & X_2 & ... & X_n \\ P_1 & P_2 & ... & P_n \end{pmatrix}, \quad (1.3)$$

где $P_i$ - вероятность наступления события $X_i$, причем сумма вероятностей наступления всех возможных исходов равна единице [18, 19].

Универсальной формой закона распределения (непрерывных и дискретных величин) является функция распределения вероятностей – это такая функция F(x), значение которой в точке x равно вероятности P того, что при проведении опыта значение случайной величины X окажется меньше, чем x, т.е. F(x) = P(X < x). Числовые значения функции F(x) принадлежат промежутку от 0 до 1 [20].

Если случайная величина дискретна, то ее функция распределения представляет собой ступенчатую функцию, а у непрерывных случайных величин функция распределения также непрерывна. Функцию распределения вероятностей F(x) непрерывной случайной величины можно представить в виде интеграла от некоторой неотрицательной функции f (x) (1.4):

$$F(x) = \int_{-\infty}^{x} f(u)du \quad (1.4)$$

Функция f(x) характеризует плотность распределения вероятности, и может быть определена следующими свойствами:

1) $f(x) = \dfrac{dF(x)}{dx}; F(x) = \int\limits_{-\infty}^{x} f(u)du$;

2) $\int\limits_{-\infty}^{+\infty} f(u)du = 1$;

3) Пропорциональности плотности вероятности и вероятности события: $x \le X \le x + dx$

Помимо закона распределения, случайная величина может быть охарактеризована значениями параметров, определяющих наиболее значимые особенности ее распределения. К таким характеристикам можно отнести математическое ожидание и дисперсию случайной величины [21-26].

Математическое ожидание (среднее значение дискретной случайной величины) представляет собой сумму произведений всех возможных значений $x_i$ на соответствующие вероятности их возникновения $p_i$ (1.5):

$$M\{X\} = \bar{x} = \sum_{i=1}^{n} x_i P_i \quad (1.5)$$

Поскольку случайной величиной является также и функция от случайной величины, математическое ожидание функции Y=H(X) может быть определено следующим образом (1.6):

$$M\{H(X)\} = \sum_{i=1}^{n} H(x_i)P(X = x_i) = \sum_{i=1}^{n} H(x_i)P_i \quad (1.6)$$

Соответственно для непрерывных величин математическое ожидание примет следующий вид (1.7):

$$M\{X\} = \bar{x} = \int\limits_{-\infty}^{+\infty} xf(x)dx \quad (1.7)$$

А для функции непрерывной величины математическое ожидание может быть найдено по формуле (1.8):

$$M\{H(X)\} = \int\limits_{-\infty}^{+\infty} H(x)f(x)dx \quad (1.8)$$

Важным параметром, характеризующим разброс случайной величины относительно ее среднего значения, является дисперсия. Она может быть определена как математическое ожидание квадрата отклонения случайной величины X от своего среднего значения (1.9):

$$D\{X\} = \sigma^2(X) = M\{(X - M\{X\})^2\} = M\{(X - \bar{x})^2\} = M\{X^2\} - (M\{X\})^2 \quad (1.9)$$

Стандартное или среднеквадратическое отклонение показывает, насколько сильно значения случайной величины X разбросаны вокруг ее среднего значения, и может быть определено по формуле (1.10):

$$\sigma = +\sqrt{\sigma^2(X)} \quad (1.10)$$

Таким образом, владение базовыми статистическими знаниями является необходимым условием управления фирмой на основе статистических методов. Понимание сущности определяемых величин и их грамотная интерпретация с экономической точки зрения позволят не только выполнять расчеты, но и активно применять их в практической деятельности.

## Контрольные вопросы

1. Что обозначает термин «статистика»?
2. Кто ввел в науку термин «статистика»?
3. Какие основные школы стояли у истоков статистической науки?
4. Назовите представителей российской школы государствоведения.
5. В чем видели сущность статистики представители немецкой описательной школы?
6. Назовите основных представителей немецкой описательной школы
7. Какие идеи статистической мысли продвигали представители английской школы политических арифметиков?
8. Назовите основных представителей английской школы политических арифметиков
9. В чем видели основую цель в своих исследованиях представители школы политических арифметиков?
10. Что является назначением статистической науки согласно взглядам политических арифметиков?
11. Какой вклад в развитие статистической науки внес Уильям Петти?

12. Какие процессы исследовал представитель английской школы политических арифметиков Дж. Граунт?
13. Какое направление английской школы политических арифметиков в статистике является преемственным?
14. Назовите основных представителей академической статистики
15. В каких основных значениях употребляется термин «статистика»?
16. Перечислите основные приемы статистической науки, применяемые в современных исследованиях.
17. В чем состоят задачи статистики как науки?
18. Какими организациями осуществляют статситические исследования в Российской Федерации?
19. Какие статистические органы существуют в России?
20. В чем состоят основные принципы работы статистических управлений?
21. В чем состоит принцип региональной децентрализации?
22. Объясните сущность принципа предментной централизации в статистическом иследовании.
23. Какой орган руководит деятельностью Росстат?
24. Перечислите основные задачи Росстата.
25. Как в ходе эволюции статистической науки менялись представлении о ее предмете?
26. Что представляет собой статистическое наблюдение?
27. С чего начинается любое статистическое наблюдение?
28. Перечислите основные этапы статистического наблюдения
29. Что означает понятие «критический момент наблюдения»?
30. Чем отличаются понятия «период наблюдения» и «время наблюдения»?
31. Что предполагает формулировка цели исследования?
32. Что представляет собой объект наблюдения?
33. Дайте определение статистической совокупности.
34. Назовите основные свойста статистической совокупности.
35. Какое свойство статистической совокупности обозначает, что появление новых элементов совокупности и исчезновение существовавших ранее не отменяет существования совокупности как объекта наблюдения?

36. Что представляет собой каждая единица наблюдения?
37. На какие группы делятся признаки по характеру отображения свойств единиц изучаемой совокупности?
38. Какие призначи являются дискретными?
39. Какие признаки называются непрерывными? В чем состоит их сущность?
40. Какие существуют основные шкалы измерения в статистике?
41. Для каких данных используется порядковая шкала?
42. Какая шкала устанавливает отношения следования между разностями значений признака?
43. Всякий ли сбор данных можно назвать статистическим наблюдением?
44. Назовите основные требования к статистическому наблюдению
45. Какие бывают статистические налюдения по способу регистрации?
46. Что представляет собой регист в статистническом аблюдении?
47. Назовите основные признаки классификации статистических наблюдений
48. Что такое статистическая сводка?
49. Что является результатом статистической группировки?
50. Какие бывают основные виды группировок?
51. Перечислите основные правила составления и оформления статистических таблиц
52. Что показывает кривая Лоренца?
53. В чем состоит роль прогнозирования в управлении фирмой?
54. Что такое прогнозирование?
55. Какие признаки классификации прогнозов Вы знаете?
56. В чем состоит отличие плана и прогноза?
57. Как соотносятся понятия «метод прогнозирования» и «модель прогнозирования»?
58. Какие модели временных рядов Вам известны?

## ПРИМЕРЫ РЕШЕНИЯ ЗАДАЧ

**Пример 1.1.**

В таблице приведены сведения о группах фирм, сформированных по объему их прибыли, в некотором городе N:

| Группы организаций по объему выручки, тыс. руб. | Число фирм, входящих в данную группу |
|---|---|
| До 4 | 15 |
| 5 - 10 | 29 |
| 11 - 16 | 25 |
| 17 - 25 | 22 |
| 26 - 31 | 30 |
| 32 - 50 | 34 |
| 51 - 65 | 47 |
| 66 - 80 | 36 |
| 81 - 100 | 27 |
| 101-150 | 21 |

На основании представленных данных произведите вторичную группировку с образованием групп фирм по объему выручки до 25 тыс. руб., от 26 до 50 тыс. руб., от 51 до 100 тыс. руб., от 101 до 150 тыс. руб.

Также требуется определись частость, накопленную частость и абсолютную плотность распределения.

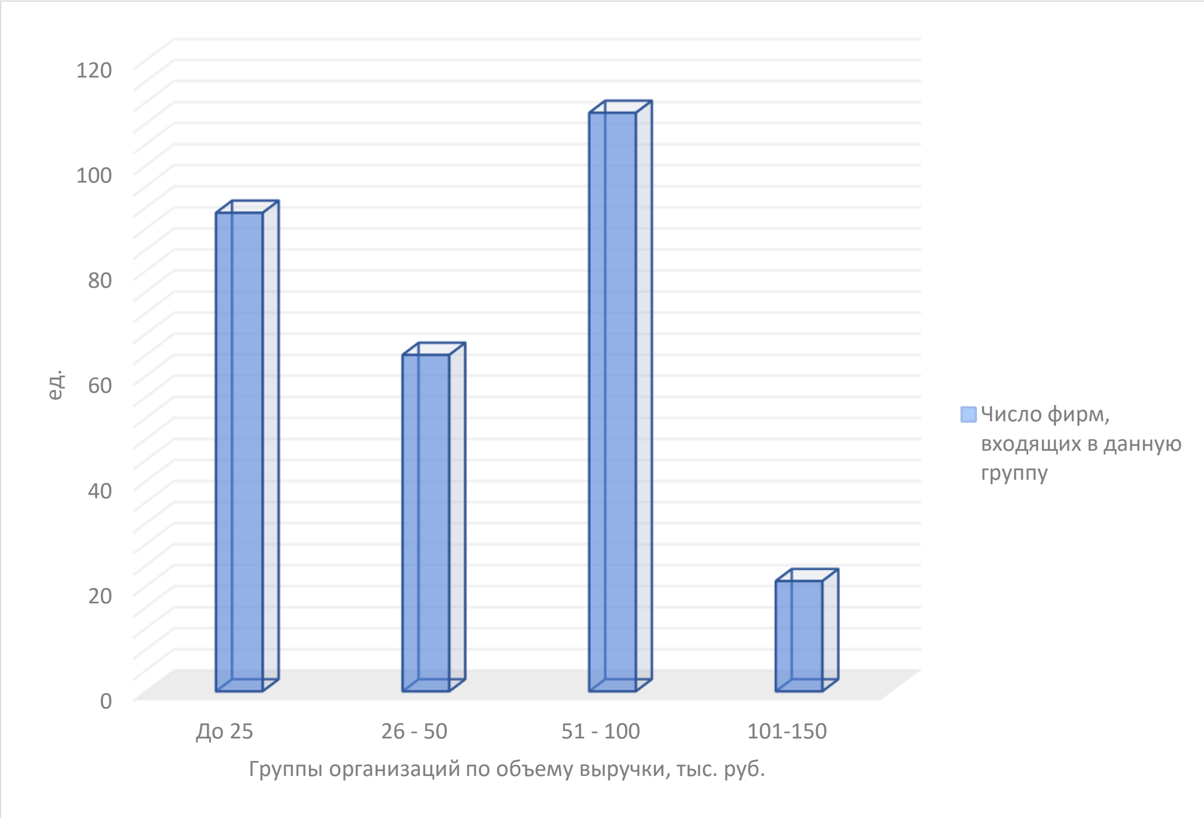
**Решение.**

Можно осуществить вторичную группировку, изменив интервалы представления исходных данных. В данном случае для формирования требуемых групп воспользуемся методом укрупнения интервалов, сложив необходимые значения по второму столбцу.

Тогда формирование таблицы в соответствии с требуемыми условиями будет выглядеть следующим образом:

| Группы организаций по объему выручки, тыс. руб. | Число фирм, входящих в данную группу |
|---|---|
| До 25 | 15 + 29 + 25 + 22 |
| 26 - 50 | 30 + 34 |
| 51 - 100 | 47 + 36 + 27 |
| 101 – 150 | 21 |

Результаты формирования новых групп в графическом виде могут быть представлены на гистограмме:




Второй способ вторичной группировки представляет собой процесс формирования новой группировки по удельному весу намечаемых к образованию групп.

Использование данного метода предполагает определение доли отдельных групп единиц совокупности с последующей перегруппировкой, согласно удельному весу групп, которые подлежат формированию.

Для определения частости найдем суммарное количество единиц рассматриваемой совокупности, а затем определим значение искомого параметра по группам путем деления числа единиц в каждой из рассматриваемых групп к найденному суммарному значению.

Накопленная частость определяется путем последовательного суммирования удельной доли значений предшествующих групп, в том числе и значений, входящих в определяемый интервал.

Для нахождения абсолютной плотности определим отношение частости к числу фирм, входящих в каждую группу.

Для упрощения расчетов результаты вычислений в формульном виде представим в сводной таблице:

| 28 | Группы организаций по объему выручки, тыс. руб. | Число фирм, входящих в данную группу (частота) | Частость, % | Накопленная частость, % | Плотность распределения |
|---|---|---|---|---|---|
| 29 | До 25 | =15+29+25+22 | =B29/СУММ($B$29:$B$32) | =C29 | =B29/25 |
| 30 | 26 - 50 | =30+34 | =B30/СУММ($B$29:$B$32) | =C29+C30 | =B30/(50-26) |
| 31 | 51 - 100 | =47+36+27 | =B31/СУММ($B$29:$B$32) | =C29+C30+C31 | =B31/(100-51) |
| 32 | 101-150 | 21 | =B32/СУММ($B$29:$B$32) | =C29+C30+C31+C32 | =B32/(150-101) |

В результате вычислений получим следующие итоги:

| Группы организаций по объему выручки, тыс. руб. | Число фирм, входящих в данную группу (частота) | Частость, % | Накопленная частость, % | Плотность распределения |
|---|---|---|---|---|
| До 25 | 91 | 31,82 | 31,82 | 3,64 |
| 26 - 50 | 64 | 22,38% | 54,20 | 2,67 |
| 51 - 100 | 110 | 38,46% | 92,66 | 2,24 |
| 101-150 | 21 | 7,34% | 100,00 | 0,43 |

Таким образом, можно сделать вывод о том, что самым многочисленным объединением из сформированных является группа фирм с объемом выручки от 51 до 100 тыс. руб., ее численность составляет 110 организаций, частость 38,46%.

**Пример 1.2.**

В таблице представлены данные о распределении фирм по выручке в двух различных регионах:

| Регион N | | Регион M | |
|---|---|---|---|
| Группы организаций по объему выручки, млн. руб. | Число фирм, входящих в данную группу | Группы организаций по объему выручки, млн. руб. | Число фирм, входящих в данную группу |
| 1 – 5 | 5 | Менее 5 | 2 |
| 6 – 10 | 7 | 5 – 10 | 8 |
| 11 – 15 | 11 | 11 – 24 | 20 |
| 16 – 20 | 13 | 25 – 35 | 22 |
| 21 – 40 | 20 | 36 – 40 | 14 |
| 41-50 | 17 | 41 -50 | 18 |

На основании представленных данных произведите вторичную группировку, пересчитав данные региона N в соответствии с группировкой региона M. Сделайте вывод о том, в каком из регионов средний объем выручки больше?

**Решение.**

Для того, чтобы осуществить вторичную перегруппировку в соответствии с категориями региона M, проанализируем, какая часть анализируемых объектов входит в состав вновь определенных групп.

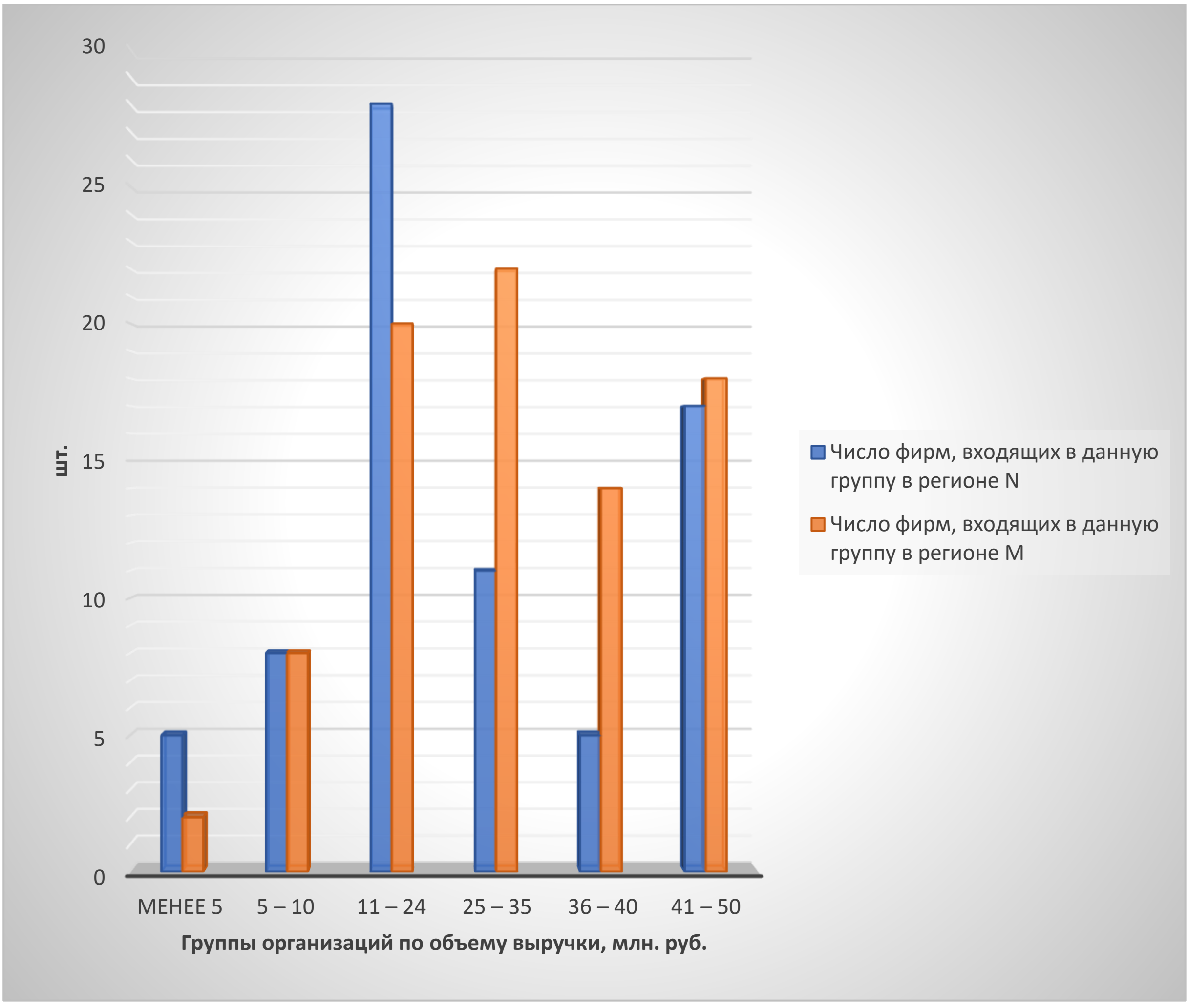
Для этого составим вспомогательную таблицу:

| Прежняя группировка | | Новая группировка | | |
|---|---|---|---|---|
| Группы организаций по объему выручки, млн. руб. | Число фирм, входящих в данную группу | Группы организаций по объему выручки, млн. руб. | Доля организаций, входящих в обновленную группировку из региона N | Число фирм, входящих обновленную группу |
| 1 – 5 | 5 | Менее 5 | 0,8 | = 0,8*5 = 4 |
| 6 – 10 | 7 | 5 – 10 | 0,2 из группы «1-5» и вся группа «6-10» | = 0,2*5 + 7 = 8 |
| 11 – 15 | 11 | 11 – 24 | Вся группа «11-15», вся группа «16-20», 0,2 из группы 21-40 | = 11 + 13 + 0,2*20 = 28 |
| 16 – 20 | 13 | 25 – 35 | 0,55 из группы 21-40 | = 0,5 * 20 = 11 |
| 21 – 40 | 20 | 36 – 40 | 0,25 из группы 21-40 | = 0,25 * 20 =5 |
| 41 – 50 | 17 | 41 – 50 | вся группа «41-50» | 17 |

Таким образом, вторичная группировка по признакам региона М для двух субъектов будет выглядеть следующим образом:

| Регион N | | Регион М | |
|---|---|---|---|
| Группы организаций по объему выручки, млн. руб. | Число фирм, входящих в данную группу | Группы организаций по объему выручки, млн. руб. | Число фирм, входящих в данную группу |
| Менее 5 | 5 | Менее 5 | 2 |
| 5 – 10 | 8 | 5 – 10 | 8 |
| 11 – 24 | 28 | 11 – 24 | 20 |
| 25 – 35 | 11 | 25 – 35 | 22 |
| 36 – 40 | 5 | 36 – 40 | 14 |
| 41 – 50 | 17 | 41 – 50 | 18 |

Графически сравнительные результаты после вторичной перегруппировки представлены ниже:



По результатам графического анализа сделать однозначный вывод о том, в каком из регионов средний объем выручки больше, достаточно трудно. В данном конкретном примере целесообразно применить среднюю арифметическую взвешенную, поскольку варианты исследуемой совокупности встречаются различное число раз.

Для расчета средней взвешенной для региона N воспользуемся вспомогательной таблицей:

| Группы организаций по объему выручки, млн. руб. | x | Число фирм, входящих в данную группу, f | x* f |
|---|---|---|---|
| Менее 5 | 3 | 5 | 15 |
| 5 – 10 | 7,5 | 8 | 60 |
| 11 – 24 | 17,5 | 28 | 490 |
| 25 – 35 | 30 | 11 | 330 |
| 36 – 40 | 38 | 5 | 190 |
| 41 – 50 | 45,5 | 17 | 773,5 |

Для региона N средний объем выручки составит:

$$x_{ср} = \frac{\sum x*f}{\sum f} = \frac{1858{,}5}{74} = 25{,}11 (\text{млн. руб.})$$

Аналогичные расчеты выполним для региона М:

| Группы организаций по объему выручки, млн руб. | x | Число фирм, входящих в данную группу, f | x* f |
|---|---|---|---|
| Менее 5 | 3 | 2 | 6 |
| 5 – 10 | 7,5 | 8 | 60 |
| 11 – 24 | 17,5 | 20 | 350 |
| 25 – 35 | 30 | 22 | 660 |
| 36 – 40 | 38 | 14 | 532 |
| 41 – 50 | 45,5 | 18 | 819 |

Для региона М средний объем выручки составит:

$$x_{ср} = \frac{\sum x*f}{\sum f} = \frac{2427}{84} = 28{,}89\ (\text{млн. руб.})$$

Таким образом, средний объем выручки фирм в регионе М больше.

# Глава 2. АНАЛИЗ ДИНАМИЧЕСКИХ ХАРАКТЕРИСТИК ДЕЯТЕЛЬНОСТИ ФИРМЫ

## 2.1. Ряды динамики в управлении фирмой

Временной ряд или ряд динамики представляет собой последовательность упорядоченных числовых параметров, характеризующий изменение исследуемого признака во времени [27]. Ряд динамики может быть представлен в графическом и в табличном виде. Графическое представление ряда динамики предполагает построение системы координат, в которой ось абсцисс представляет собой шкалу времени t, а ось ординат – шкалу уровней ряда Y.

В табличном виде числовые значения ряда динамики соответствуют моментам или периодам времени $t_i$, к которым относятся уровни. Причем $Y_1$ - начальное значение уровня ряда, а $Y_n$ - его конечное значение.

Существуют различные признаки классификации временных рядов [28, 29], основные из которых представлены на рис. 2.1.


*Рис. 2.1. Классификация рядов динамики*

Для **моментного временного ряда** характерно представление значения статистического показателя по состоянию на определенные последовательные моменты времени (например, на начало каждого месяца или года). При работе с моментными данными необходимо помнить, что сложение показателей такого ряда динамики невозможно, поскольку может появиться проблема двойного счета. Примером моментных рядов может служить численность населения на начало каждого календарного года, величина запасов на первое число каждого месяца.

**В интервальных рядах динамики** данные накапливаются за определенный промежуток времени. Например, годовое значение объема выпущенной продукции равно сумме данных за четыре квартала [30].

**Полный временной ряд** - ряд динамики, в котором одноименные моменты времени или периоды времени строго следуют один за другим в календарном порядке или равноотстоят друг от друга. В **неполном ряду** уровни зафиксированы в неравноотстоящие моменты или периоды времени.

**Изолированные временные ряды** строятся по одному конкретному показателю, а **комплексные ряды** являются многомерными и представляют собой систему взаимосвязанных параметров.

В зависимости от формы представления уровней ряды могут быть **абсолютными** (например, объем произведенной продукции), **относительными** (индекс инфляции) или **средними** (величина среднедушевых доходов) [31].

**Стационарный временной ряд** это ряд данных, статистические свойства которого не зависят от времени. Таким ряды не содержат тенденцию или циклическую компоненту. Значение их каждого последующего уровня может быть определено как сумма среднего уровня ряда и случайной компоненты.

Для нестационарного ряда характерно изменение статистических свойств во времени. Для нестационарных радов характерно наличие тренда, сезонной и/или циклической составляющей [32].

Таким образом, грамотная интерпретация исходных данных является залогом успешного моделирования и прогнозирования. Понимание исследователем природы и сущности рассматриваемых явлений позволяет грамотно использовать статистический инструментарий для решения конкретных прикладных задач.

К основным показателя временных рядов относятся: средний уровень временного ряды, абсолютные приросты, темпы роста и прироста, абсолютное значение одного процента прироста.

Средний уровень представляет собой среднюю величину изучаемого показателя за рассматриваемый период. Следует отметить, что для различных видов рядов динамики исчисление осуществляется по различным формулам [33].

Для равномерного интервального ряда вычисление производится по формуле простой средней арифметической (2.1):

$$\bar{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i \tag{2.1}$$

Средний уровень неравномерного интервального ряда определяется по формуле средней арифметической взвешенной (2.2):

$$\bar{Y} = \frac{1}{N} \sum_{i=1}^{n} Y_i t_i, N = \sum_{i=1}^{n} t_i \tag{2.2}$$

Для вычисления равномерного моментного ряда принято использовать формулу средней хорологической простой (2.3):

$$\bar{Y} = \frac{1}{n-1}\left(\frac{Y_1}{2} + Y_2 + Y_3 + \ldots Y_{n-1} + \frac{Y_n}{2}\right) = \frac{1}{2(n-1)} \sum_{i=1}^{n-1} (Y_i + Y_{i+1}) \tag{2.3}$$

Для неравномерного моментного ряда предполагается применение формулы средней хронологической взвешенной (2.4):

$$\bar{Y} = \frac{1}{\sum_{i=1}^{n} t_i} * \sum_{i=1}^{n} \bar{Y}_i t_i \tag{2.4}$$

Абсолютный прирост представляет собой разность между двумя последующими уровнями временного ряда (цепной прирост) или разность между данным уровнем ряда и уровнем, принятым за базу сравнения (бизисный прирост). Как правило, базой сравнения служит начальное значение уровня ряда.

Показатели абсолютного прироста измеряются в тех же единицах измерения, что и уровни ряда [34].

Соответственно базисный прирост отражает изменение величины показателя в анализируемом периоде по сравнению с базисным и может быть определен по формуле (2.5):

$$\Delta_{i\text{базис.}} = Y_i - Y_0 \tag{2.5}$$

где $Y_i$ - данный уровень ряда,

$Y_0$ - базисное значение уровня ряда.

Цепной прирост показывает, на сколько изменилась величина показателя в данном периоде по сравнению с предыдущим периодом (2.6):

$$\Delta_{i\,цепн.} = Y_i - Y_{i-1} \tag{2.6}$$

Где $Y_{i-1}$ - предыдущее значение уровня ряда относительно $Y_i$

Средний абсолютный прирост характеризует среднее изменение величины изучаемого показателя (2.7):

$$\bar{\Delta} = \frac{Y_n - Y_0}{n-1} \tag{2.7}$$

Где $Y_n$ и $Y_0$ конечный и начальный уровни ряда соответственно [35].

Либо средний абсолютный прирост может быть определен через значения цепных приростов (2.8):

$$\bar{\Delta} = \frac{\sum_{i=1}^{k} \Delta_{i\,цепн}}{K} \tag{2.8}$$

где К – количество анализируемых цепных приростов.

Базисный темп роста ($T_{рб}$) характеризует отношение каждого последующего уровня к базисному, выраженное в процентах [36] (2.9):

$$T_{рб} = \frac{Y_n}{Y_0} * 100\% \tag{2.9}$$

Цепной темп роста ($T_{рц}$) определяется следующим соотношением (2.10):

$$T_{рц} = \frac{Y_n}{Y_{n-1}} * 100\% \tag{2.10}$$

Темп роста показывает, какую долю составляет текущее значение уровня ряда относительно начального или предыдущего (в зависимости от базисного или цепного способа составления ряда).

Средний темп роста характеризует среднее процентное изменение изучаемого показателя [37]. Расчет данного показателя осуществляется по формуле средней геометрической простой (2.11):

$$\overline{T_p} = \sqrt[n-1]{T_{рц_1} * T_{рц_2} * \ldots * T_{рц_{n-1}}} = \sqrt[n-1]{\frac{Y_n}{Y_0}} *100\% \qquad (2.11)$$

Темп прироста представляет собой отношение абсолютного прироста (базисного или цепного) к начальному значению временного ряда, выраженное в процентах [38].

Соответственно базисный темп прироста может быть найден по формуле (2.12):

$$T_{пр_б} = \frac{\Delta_{базис}}{Y_0} *100\% = \frac{Y_n - Y_0}{Y_0} *100\% \qquad (2.12)$$

Также можно найти базисный темп прироста, зная значение базисного темпа роста (2.13):

$$T_{пр_б} = T_{р_б} - 100\% \qquad (2.13)$$

Цепной темп находится по следующим формулам (2.14 и 2.15):

$$T_{пр_ц} = \frac{\Delta_{цепн}}{Y_{n-1}} *100\% = \frac{Y_n - Y_{n-1}}{Y_{n-1}} *100\% \qquad (2.14)$$

$$T_{пр_ц} = T_{р_ц} - 100\% \qquad (2.15)$$

Темп прироста показывает, на сколько процентов изменилась величина показателя в данном периоде по сравнению с начальным или предыдущим (в зависимости от базисного или цепного способа составления ряда) [39].

Средний темп прироста позволяет оценить величину, на которую в среднем изменилась величина изучаемого показателя (2.16):

$$\overline{T_{пр}} = \overline{T_p} - 100\% \qquad (2.16)$$

Стоит отметить, что темп прироста может быть как положительной, так и отрицательной величиной.

Коэффициенты роста также могут быть базисными (2.17) или цепными (2.18):

$$K_{р_б} = \frac{Y_n}{Y_0} \qquad (2.17)$$

$$К_{р_ц} = \frac{Y_n}{Y_{n-1}} \quad (2.18)$$

В отличие от темповых характеристик они показывают относительную скорость изменения ряда, выраженную в долях, а не процентах [40].

Базисные или цепные коэффициенты прироста вычисляются по формулам (2.19 и 2.20):

$$К_{пр_б} = К_{р_б} - 1 \quad (2.19)$$

$$К_{пр_ц} = К_{р_ц} - 1 \quad (2.20)$$

Абсолютное значение одного процента прироста характеризует скорость изменения уровней ряда в единицу времени [41] и может быть найдено по следующей формуле (2.21):

$$|А| = \frac{\Delta_{цепн}}{Т_{пр_ц}} \quad (2.21)$$

Таким образом, в рамках данного параграфа были рассмотрены основные показатели временных рядов и формулы для их вычисления. Следует отметить, что понимание сущности рассчитываемых параметров является необходимым условием грамотного их применения в практической деятельности.

## 2.2. Проблема сопоставимости динамических рядов и пути ее решения

При построении динамического ряда исследователю важно помнить, что все уровни отвечать требованию сопоставимости: в частности, уровни должны относиться в одной территории, характеризовать один и тот же объект или явление, быть рассчитанными по единой методологии с использованием единых единиц измерения значений анализируемого показателя [40].

Сопоставимость по кругу охватываемых явлений означает, что процедура сравнения проводится для равных по числу элементов совокупностей, однородных по своему экономическому содержанию и обладающих одинаковыми границами. Несопоставимость может возникнуть в результате перехода ряда объектов из одного подчинения в другое. Например, если показатели из одной группы изучаемых социально-экономических явлений в течение анализируемого интервала времени были отнесены к другой группе.

Сопоставимость по времени регистрации для интервальных рядов динамики предполагает равенство временных промежутков, за которые собираются и представляются данные. Например, для обеспечения сопоставимости интервальных временных рядов, выделяют среднедневные показатели по декадам, кварталам, месяцам, которые затем подвергаются процедуре сравнения. Если речь идет о моментных рядах, наблюдения лучше проводить по состоянию на конкретно определенное число каждого месяца, полугодия, года [41].

Сопоставимость по территории предполагает исследование данных в рамках одних тех же территориальных границ в течение всего времени анализа. Например, изменение территориальных границ Москвы и Московской области привело к тому, что ранее параметры отдельных городов учитывались в показателях Московской области, в после расширения границ Москвы сведения по этим географическим единицам были включены в статистические данные по столице, то есть возникла проблема территориальной сопоставимости

Сопоставимость по методологии расчетов предполагает использование единого методологического аппарата для вычислений уровней динамического ряда в течение всего аналитического периода [42].

Существует также проблема сопоставимости по ценам. Существуют различные виды цен. Также стоимостные изменения могут не учитывать влияние инфляции. Для устранения данных проблем оценка одних и тех же процессов или явлений зачастую осуществляется в ценах одного и того же базисного периода. В статистических источниках такие приведенные данные принято называть сопоставимыми.

Следует отметить, что для экономических процессов изменение качественных характеристик является одним из наиболее часто встречающихся состояний. С течением времени меняется методология вычислений показателей, происходит корректировка границ их определения как в сторону сужения, так и расширения. Например, до октября 1995 года предприятие считалось малым, если число работников не превышало 200, а с октября ценз отнесения предприятий к числу малых по критерию численности сотрудников был снижен до 100 человек. Таким образом, возникла проблема сопоставимости динамического ряда, который бы характеризовал результаты деятельности малых предприятий до и после внесения изменений.

Приведение статистической информации к сопоставимому виду осуществляется при помощи особой вычислительной процедуры, которая носит название «смыкание временных рядов» [43].

Под смыканием принято понимать процесс объединения нескольких динамических рядов, относящихся к разным периодам времени и с первоначально несопоставимыми уровнями, в один динамический ряд с новыми, уже сопоставимыми уровнями, расположенными в хронологической последовательности.

Для того, чтобы проведение данной процедуры стало возможным, необходимо обладать информацией о значении показателя, рассчитанного по старой и новой методологии, но относящейся к одному аналитическому периоду времени [44].

Принято выделять два основных способа смыкания временных рядов.

Первый способ является абсолютным. Он предполагает корректировку данных до изменений на коэффициент перехода. Подробно схема реализации данного метода представлена на рис. 2.2.

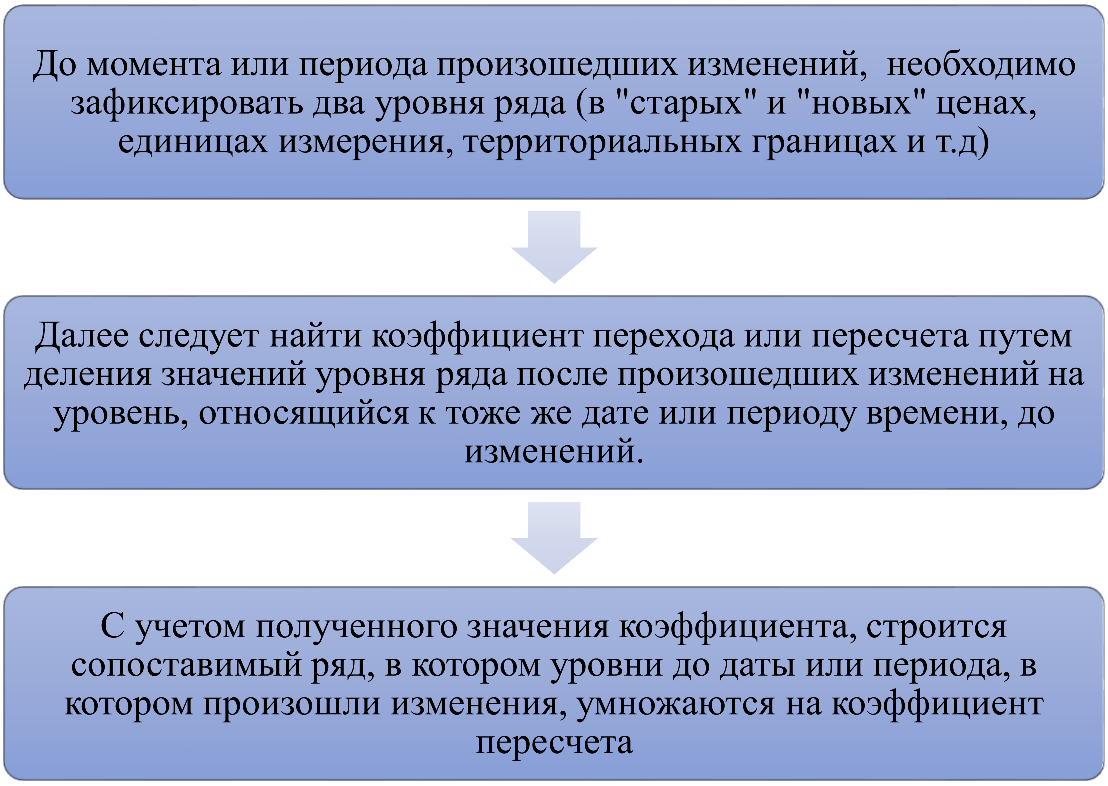



*Рис. 2.2. Схема реализации абсолютного способа смыкания данных*

Второй способ (относительный) предполагает получение временного ряда, выраженного в процентах или долях. При реализации данного метода уровень переходного периода принимается равным 100 % (как для прежнего значения ряда, так и для рассчитанного по новой методологии, в новых единица измерения и др.). Временной ряд пересчитывается, исходя из численного значения уровня, принятого за 100%. Применение данного способа предполагает пересчет значений как до, так и после «переломного момента».

Таким образом, анализ сопоставимости динамических рядов является важнейшим этапом научно-исследовательской работы. Для приведения данных в сопоставимый вид используется два основных способа смыкания: абсолютный и относительный [40]. Выбор способа смыкания обусловлен спецификой исходных временных рядов, а также и целями исследовательской работы в рамках решение конкретной практической задачи.

## 2.3. Работа с аномальными статистическими значениями в процессе управления фирмой

Под аномальным уровнем ряда понимается конкретное численное значение уровня, которое не отвечает потенциальным возможностям исследуемой экономической системы и, которое, оставаясь в качестве уровня ряда, оказывает существенное влияние на значения основных характеристик временного ряда, в том числе на соответствующую трендовую модель. Причинами аномальных наблюдений могут быть ошибки технического порядка, или **ошибки первого рода**. К ним относятся ошибки, возникающие при передаче информации, а также ошибки, возникающие при агрегировании и дезагрегировании показателей.

Кроме того, причины аномальных уровней во временных рядах могут быть обусловлены факторами, имеющими объективный характер, но проявляющимися эпизодически или очень редко (**ошибки второго рода**). Такого рода ошибки устранению не подлежат [45].

Одним из способов выявления аномальных уровней временных рядов является **метод Ирвина**.

Реализация данного метода основывается на вычислении параметра $\lambda_t$ (2.22):

$$\lambda_t = \frac{|y_t - y_{t-1}|}{\delta_y} \qquad (2.22)$$

где $\delta_y$ - среднее квадратическое отклонение, которое рассчитывается по формуле (2.23):

$$\delta_y = \sqrt{\frac{\sum_{t=1}^{n}(y_t - \overline{y})^2}{n-1}} \qquad (2.23),$$

Где $\overline{y}$ - среднее значение уровня ряда, которое рассчитывается по формуле (2.24):

$$\overline{y} = \frac{\sum_{t=1}^{n} y_t}{n} \qquad (2.24)$$

На следующем этапе расчетные значения $\lambda_t$ сравниваются с табличными значениями критерия Ирвина $\lambda_\alpha$. В случае, если найденное $\lambda_t > \lambda_\alpha$, можно сделать вывод о том, что соответствующее значение уровня $y_t$ является аномальным.

Значения критерия Ирвина для уровня значимости α = 0,05 и α = 0,01 приведены в табл. 2.1.

*Табл. 2.1. Табличные значения критерия Ирвина* $\lambda_\alpha$

| Число измерений *n* | Уровень значимости | |
|---|---|---|
| | *q* = 0,05 | *q* = 0,01 |
| 2 | 2,8 | 3,7 |
| 3 | 2,2 | 2,9 |
| 10 | 1,5 | 2,0 |
| 20 | 1,3 | 1,8 |
| 30 | 1,2 | 1,7 |
| 50 | 1,1 | 1,6 |
| 100 | 1,0 | 1,5 |
| 400 | 0,9 | 1,3 |
| 1000 | 0,8 | 1,2 |

После выявления аномальных уровней ряда, обязательным условием является определение причин их возникновения, а затем устранение этих значений.

**Критерий Романовского.** Данный критерий применяется для исследования грубых погрешностей (промахов), если число наблюдения n меньше 20 [46].

Конкурирующая гипотеза о наличии грубых погрешностей в подозрительных результатах подтверждается, если выполняется неравенство (2.25):

$$\beta_t = \frac{\left| y_t - \overline{y} \right|}{\delta_y} \tag{2.25}$$

При анализе грубых промахов с использованием критерия Романовского рассчитанное по формуле 2.25 значение $\beta_t$ сравнивается с критерием $\beta_q$, выбранным по табл. 2.2.

*Табл. 2.2. Табличные значения критерия Романовского* $\beta_q$

| n | q=0,01 | q=0,02 | q=0,05 | q=0,1 |
|---|---|---|---|---|
| 4 | 1,73 | 1,72 | 1,71 | 1,69 |
| 6 | 2,16 | 2,13 | 2,10 | 2,00 |
| 8 | 2,43 | 2,37 | 2,27 | 2,17 |
| 10 | 2,62 | 2,54 | 2,41 | 2,29 |
| 12 | 2,75 | 2,66 | 2,52 | 2,39 |
| 15 | 2,90 | 2,80 | 2,64 | 2,49 |
| 20 | 3,08 | 2,96 | 2,78 | 2,62 |

Если $\beta_t \geq \beta_q$, то значение уровня ряда считается промахом и отбрасывается.

**Критерий вариационного размаха** является одним из самых простых методов исключения грубой погрешности измерений (промаха) [47]. Для его использования определяют размах вариационного ряда упорядоченной совокупности наблюдений (2.26):

$$R = y_{max} - y_{min} \tag{2.26}$$

где $y_{max}$ и $y_{min}$ - максимальное и минимальное значение уровня ряда соответственно.

Если какой-либо член вариационного ряда, например $y_k$, резко отличается от всех других, то производят проверку, используя следующее неравенство (2.27):

$$\overline{Y} - z * R < y_k < \overline{Y} + z * R \tag{2.27}$$

где $\overline{Y}$ – выборочное среднее арифметическое значение, вычисленное после исключения предполагаемого промаха;

z – критериальное значение [48].

В табл. 2.3 приведены критериальные значения для метода вариационного размаха

*Табл. 2.3. Значения критерия вариационного размаха*

| *n* | 5 | 6 | 7 | 8-9 | 10-11 | 12-15 | 16-22 | 23-25 | 26-63 | 64-150 |
|---|---|---|---|---|---|---|---|---|---|---|
| *z* | 1,7 | 1,6 | 1,5 | 1,4 | 1,3 | 1,2 | 1,1 | 1,0 | 0,9 | 0,8 |

Нулевую гипотезу (об отсутствии грубой погрешности) принимают, если указанное неравенство выполняется. Если $y_k$ не удовлетворяет условию (2.27), то этот результат исключают из вариационного ряда [49].

**Критерий Диксона** основан на предположении, что погрешности измерений подчиняются нормальному закону (предварительно необходимо построение гистограммы результатов наблюдений) и проверка гипотезы о принадлежности нормальному закону распределения. При использовании критерия вычисляют коэффициент Диксона (наблюдаемое значение критерия) для проверки наибольшего или наименьшего экстремального значения в зависимости от числа измерений. В табл. 2.4 приведены формулы для вычисления коэффициентов. Коэффициенты $r_{10}, r_{11}$ применяют, когда имеется один выброс, а $r_{21}, r_{22}$ - когда два выброса. Требуется первоначальное упорядочение результатов измерений (объема выборки) по возрастанию. Критерий применяется, когда выборка может содержать более одной грубой погрешности [50].

*Табл. 2.4. Формулы коэффициентов Диксона*

| Число измерений $n$ (объем выборки) | Коэффициент Диксона | Для наименьшего экстремального значения параметра | Для наибольшего экспериментального параметра |
|---|---|---|---|
| 3-7 | $r_{10}$ | $\dfrac{y_2 - y_1}{y_n - y_1}$ | $\dfrac{y_n - y_{n-1}}{y_n - y_1}$ |
| 8-10 | $r_{11}$ | $\dfrac{y_2 - y_1}{y_{n-1} - y_1}$ | $\dfrac{y_n - y_{n-1}}{y_n - y_2}$ |
| 11-13 | $r_{21}$ | $\dfrac{y_3 - y_1}{y_{n-1} - y_1}$ | $\dfrac{y_n - y_{n-2}}{y_n - y_2}$ |
| 14-25 | $r_{22}$ | $\dfrac{y_3 - y_1}{y_{n-2} - y_1}$ | $\dfrac{y_n - y_{n-2}}{y_n - y_3}$ |

Вычисленные для выборки по формулам значения коэффициентов Диксона r сравнивают с табличным значением критерия Диксона $r_q$ (табл. 2.5).

*Табл. 2.5. Табличные значения критерия Диксона*

| Статистика | Число измерений | $r_q$ при уровне значимости q | | | |
|---|---|---|---|---|---|
| | | 0,1 | 0,05 | 0,02 | 0,01 |
| $r_{10}$ | 3 | 0,886 | 0,941 | 0,976 | 0,988 |
| | 4 | 0,679 | 0,765 | 0,846 | 0,899 |
| | 5 | 0,557 | 0,642 | 0,729 | 0,780 |
| | 6 | 0,482 | 0,560 | 0,644 | 0,698 |
| | 7 | 0,434 | 0,507 | 0,586 | 0,637 |
| $r_{11}$ | 8 | 0,479 | 0,554 | 0,631 | 0,683 |
| | 9 | 0,441 | 0,512 | 0,587 | 0,636 |
| | 10 | 0,409 | 0,477 | 0,551 | 0,597 |
| $r_{21}$ | 11 | 0,517 | 0,576 | 0,538 | 0,679 |
| | 12 | 0,490 | 0,546 | 0,605 | 0,642 |
| | 13 | 0,467 | 0,521 | 0,578 | 0,615 |
| $r_{22}$ | 14 | 0,462 | 0,546 | 0,602 | 0,641 |
| | 15 | 0,472 | 0,525 | 0,579 | 0,616 |
| | 16 | 0,452 | 0,507 | 0,559 | 0,595 |
| | 17 | 0,438 | 0,490 | 0,542 | 0,577 |
| | 18 | 0,424 | 0,475 | 0,527 | 0,561 |

| | | | | | |
|---|---|---|---|---|---|
| | 19 | 0,412 | 0,462 | 0,514 | 0,547 |
| | 20 | 0,401 | 0,450 | 0,502 | 0,535 |
| | 21 | 0,391 | 0,440 | 0,491 | 0,524 |
| | 22 | 0,382 | 0,430 | 0,481 | 0,514 |
| | 23 | 0,374 | 0,421 | 0,472 | 0,505 |
| | 24 | 0,367 | 0,413 | 0,464 | 0,497 |
| | 25 | 0,360 | 0,406 | 0,457 | 0,489 |

Нулевая гипотеза об отсутствии грубой погрешности выполняется, если выполняется неравенство $r < r_q$.

Если $r > r_q$, то результат признается грубой погрешностью и исключается из дальнейшей обработки [51,52].

**Критерии Райта и правило «трех сигм»** являются одними из простейших методов для проверки результатов, подчиняющихся нормальному закону распределения [53].

Сущность правила трех сигм заключается в следующем: если случайная величина распределена нормально, то абсолютная величина ее отклонения от математического ожидания не превосходит утроенного среднего квадратического отклонения [53].

На практике правило трех сигм применяют так: если распределение изучаемой случайной величины неизвестно, но условие, указанное в приведенном правиле, выполняется, то есть основания предполагать, что изучаемая величина распределена нормально; в противном случае она не распределена нормально. С этой целью для выборки вычисляется центр распределения и оценка среднего квадратического отклонения результата наблюдений. Результат, который удовлетворяет условию (2.28), считается имеющим грубую погрешность и удаляется, а ранее вычисленные характеристики распределения уточняются:

$$\left| y_k - \overline{y} \right| \geq 3\delta \qquad (2.28)$$

Правило «трех сигм» считается слишком жестким, поэтому рекомендуется назначать границу цензурирования в зависимости от объема выборки (критерий Райта) (табл. 2.6) [54].

*Табл. 2.6. Границы критерия Райта*

| Объём выборки | Условия критерий Райта |
|---|---|
| $6<n<100$ | $\left\| y_k - \overline{y} \right\| \geq 4\delta$ |
| $100<n<1000$ | $\left\| y_k - \overline{y} \right\| \geq 4{,}5\delta$ |
| $1000<n<10000$ | $\left\| y_k - \overline{y} \right\| \geq 5\delta$ |

Может оказаться, что при новых значениях $\overline{y}$ и $\delta$ другие результаты попадут в категорию аномальных, поэтому дважды использовать критерии грубой погрешности не рекомендуется [55].

**Критерий Смирнова** рекомендуется использовать при объемах выборки $n \geq 25$ или при известных значениях генеральных среднего и среднего квадратического отклонения. Он устанавливает менее жесткие границы грубой погрешности [56]. Для реализации этого критерия вычисляются действительные значения квантилей распределения (наблюдаемое значение критерия) по формуле (2.29):

$$\beta = \frac{\max\left| y_k - \overline{y} \right|}{\delta}, \tag{2.29}$$

где $y_k$ - сомнительное значение ряда (максимальное или минимальное) [57].

Найденное значение $\beta$ сравнивается с критериальным $\beta_q$, приведенным в табл. 2.7

*Табл. 2.7. Квантили распределения* $\beta_q$

| Объем выборки n | Предельное значение $\beta_q$ при уровне значимости *q* | | | | |
|---|---|---|---|---|---|
| | 0,100 | 0,050 | 0,0010 | 0,005 | 0,001 |
| 1 | 2 | 3 | 4 | 5 | 6 |
| 1 | 1,282 | 1,645 | 2,326 | 2,576 | 3,090 |
| 2 | 1,632 | 1,955 | 2,575 | 2,807 | 3,290 |
| 3 | 1,818 | 2,121 | 2,712 | 2,935 | 3,403 |

| | | | | | |
|---|---|---|---|---|---|
| 4 | 1,943 | 2,234 | 2,806 | 3,023 | 3,481 |
| 5 | 2,036 | 2,319 | 2,877 | 3,090 | 3,540 |
| 6 | 2,111 | 2,386 | 2,934 | 3,143 | 3,588 |
| 7 | 2,172 | 2,442 | 2,981 | 3,188 | 3,628 |
| 8 | 2,224 | 2,490 | 3,022 | 3,227 | 3,662 |
| 9 | 2,269 | 2,531 | 3,057 | 3,260 | 3,692 |
| 10 | 2,309 | 2,568 | 3,089 | 3,290 | 3,719 |
| 15 | 2,457 | 2,705 | 3,207 | 3,402 | 3,820 |
| 20 | 2,559 | 2,799 | 3,289 | 3,480 | 3,890 |
| 25 | 2,635 | 2,870 | 3,351 | 3,539 | 3,944 |
| 30 | 2,696 | 2,928 | 3,402 | 3,587 | 3,988 |
| 40 | 2,792 | 3,015 | 3,480 | 3,662 | 4,054 |
| 50 | 2,860 | 3,082 | 3,541 | 3,716 | 4,108 |
| 100 | 3,076 | 3,285 | 3,723 | 3,892 | 4,263 |
| 250 | 3,339 | 3,534 | 3,946 | 4,108 | 4,465 |
| 500 | 3,528 | 3,703 | 4,108 | 4,263 | 4,607 |

Если $\beta > \beta_q$, то значение $y_k$ считают грубой ошибкой и отбрасывают

Критерий Шовене применяется для законов, не противоречащих нормальному, и строится на определении числа ожидаемых результатов наблюдений $n_{ож}$, которые имеют столь же большие погрешности, как и подозрительный [58]. Гипотеза о наличии грубой погрешности принимается, если выполняется условие (2.30):

$$n_{ож} \le 0{,}5 \qquad (2.30)$$

Порядок проверки гипотезы следующий:

1) вычисляются среднее арифметическое $\overline{y}$ и $\delta$ результатов наблюдений для всей выборки;

2) из таблицы нормированного нормального распределения (Приложение 1 – интегральная функция нормированного нормального распределения) по величине $z = \dfrac{\left|y_k - \overline{y}\right|}{\delta}$ определяется вероятность появления подозрительного результата в генеральной совокупности чисел n [59]:

$$P\left( z \cdot \delta < \left| y_k - \overline{y} \right| \right); \qquad (2.31)$$

3) число ожидаемых результатов $n_{ож}$ определяется по формуле:

$$n_{ож} = n \cdot P \qquad (2.32)$$

Указанные выше критерии во многих случаях оказываются «жесткими». Тогда рекомендуется пользоваться критерием грубой погрешности «k», зависящим от объема выборки n и принятой доверительной вероятности P (табл.2.8) [60].

*Табл. 2.8. Зависимость критерия грубой погрешности k от объема выборки n и доверительной вероятности P*

| n | P = 95% | P = 99% | P = 99,73% |
|---|---|---|---|
| 9 | 4,42 | 7,10 | 11,49 |
| 10 | 4,31 | 6,99 | 10,26 |
| 12 | 4,16 | 6,38 | 8,80 |
| 15 | 4,03 | 5,88 | 7,66 |
| 20 | 3,90 | 5,41 | 6,73 |
| 25 | 3,84 | 5,14 | 6,25 |
| 30 | 3,80 | 5,00 | 5,95 |
| 40 | 3,75 | 4,82 | 5,56 |
| 50 | 3,73 | 4,70 | 5,34 |

Для распределений, отличных от нормального, таких классов, как двух модальных кругловершинных композиций нормального и дискретного распределения с эксцессом ε=1,5-3; островершинных двумодальных; композиций дискретного двузначного распределения и распределения Лапласа с эксцессом ε=1,5-6; композиций равномерного распределения с экспоненциальным распределением эксцесса ε=1,8-6 и классом экспоненциальных распределений в пределах изменения эксцесса ε=1,8-6 граница грубой погрешности определяется величиной $\pm\left(t_{гр}\cdot\sigma\right)$ [61], где:

$$t_{гр}=1{,}2+3{,}6\cdot\left(1-\gamma\right)\cdot\lg\frac{n}{10}, \tag{2.33}$$

где $\gamma$ – контрэксцесс;

$$t_{гр}=1{,}55+0{,}8\cdot\sqrt{\varepsilon-1}\cdot\lg\frac{n}{10}. \tag{2.34}$$

Погрешности в определении оценок $\delta$ и $t_{гр}$ являются отрицательно коррелированными, т. е. возрастание $\delta$ сопровождается уменьшением $t_{гр}$. Поэтому определение границ грубой погрешности для законов, отличных от нормального, с эксцессом $\varepsilon\leq 6$ с помощью критерия $t_{гр}$ является достаточно точным и может широко использоваться на практике [62].

Оценки $\overline{y}$, $\delta$ и $\varepsilon$ должны вычисляться после исключения подозрительных результатов из выборки. После расчета границ грубой погрешности результаты наблюдений, оказавшиеся внутри границ, возвращаются, а ранее найденные характеристики распределения уточняются [63,64].

Для равномерного распределения за границы грубой погрешности можно принять величину $\pm 1{,}8\cdot\delta$.

Таким образом, существуют различные способы выявления аномальных значений уровня ряда, выбор которых должен быть определен на основе требуемых целей исследования и специфики исходного временного ряда.

## 2.4. Модели временных рядов. Структурные элементы ряда

В практической деятельности принято полагать, что структурно-образующими элементами временного ряда являются:

-тренд T(t)

-сезонная компонента S(t)

-циклическая компонента Z(t)

-случайная составляющая R(t) [65].

Анализ временных рядов путём разложения их на перечисленные компоненты называется декомпозицией. В общем виде функция изменения временного ряда представляет собой следующую зависимость (2.35):

$$y = f(T, S, Z, R) \tag{2.35}$$

Тренд представляет собой изменение, которое определяет общий вектор развития, то есть тенденцию ряда динамики. Тренд является систематической составляющей долговременного действия. Тренд предполагает плавное изменение уровня ряда, которое выражается устойчивой тенденцией развития определенного экономического процесса или явления во времени [66].

Зачастую в рядах динамики, описывающих экономические процессы или явления, помимо долговременных тенденций имеют место относительно регулярные колебания, то есть периодические составляющие временного ряда [67].

Если период колебаний меньше или равен году, то такие колебания являются сезонными [68]. Часто причина их появления обусловлена природно-климатическими или иными условиями. Например, колебание цен на сельскохозяйственную продукцию. В период сбора урожая цена снижается, а затем фиксируется рост, обусловленный необходимостью хранения собранной продукции. Также с использованием сезонных колебаний описывается объём потребления газа в течение года. Логично, что чем ниже температура окружающей среды, тем больше будут объёмы потребления газа. Соответственно, с конца весны и до середины осени показатели потребления будут значительно ниже, чем зимой.

Таким образом, в рассматриваемых процессах прослеживается устойчивая годовая периодичность. Иногда причины сезонных колебаний имеют социальный характер, например, увеличение закупок в предпраздничный период, увеличение платежей в конце квартала и т.д. [69].

Если период колебания составляет больше года, имеет место циклическая составляющая. Примерами могут служить инвестиционные, промышленные, демографические и иные циклы [70].

Помимо тренда и периодических составляющих, временные ряды содержат в себе нерегулярную компоненту [71]. В экономической практике принято выделять два вида факторов, под влиянием которых формируется данная компонента (рис.2.3).

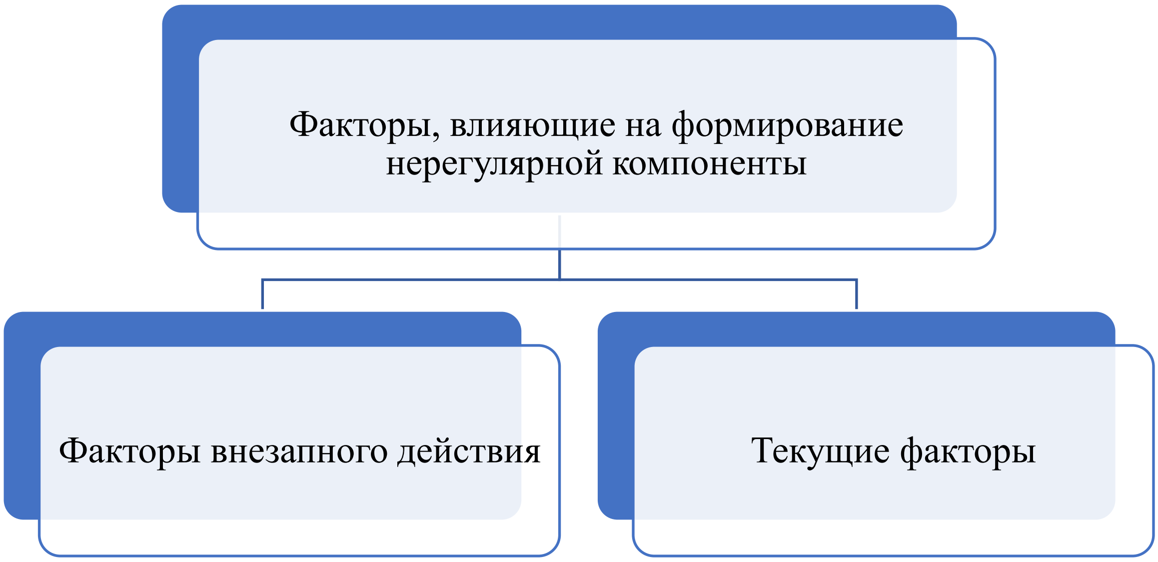



*Рис. 2.3. Факторы, оказывающие влияние на формирование нерегулярной компоненты временного ряда*

Факторы внезапного действия представляют собой эпидемии, стихийные бедствия и пр. Как правило, под влиянием данных факторов, возникают значительные отклонения, которые иногда носят название «катастрофические» [72].

Текущие факторы являются причиной случайных колебаний. Как правило, они обусловлены целым рядом побочных причин. Поскольку влияние каждого отдельно взятого текущего фактора является незначительным, в исследованиях воздействие данных факторов учитывается как суммарное [73, 74].

Рассмотренные в рамках данного параграфа положения позволяют сделать вывод о том, что выделяют четыре основных компонента, образующих структуру временного ряда, однако не все элементы имеют место в каждой конкретной модели.

Если ряд динамики представляется в виде суммы соответствующих компонент, то полученная модель является аддитивной [75] и записывается в следующем виде (2.36):

$$y_t = T(t) + S(t) + Z(t) + R(t) \tag{2.36}$$

Аддитивная модель применяется тогда, когда анализируемый ряд динамики имеет примерно одинаковую амплитуду колебаний в течение рассматриваемого промежутка времени (рис.2.4).

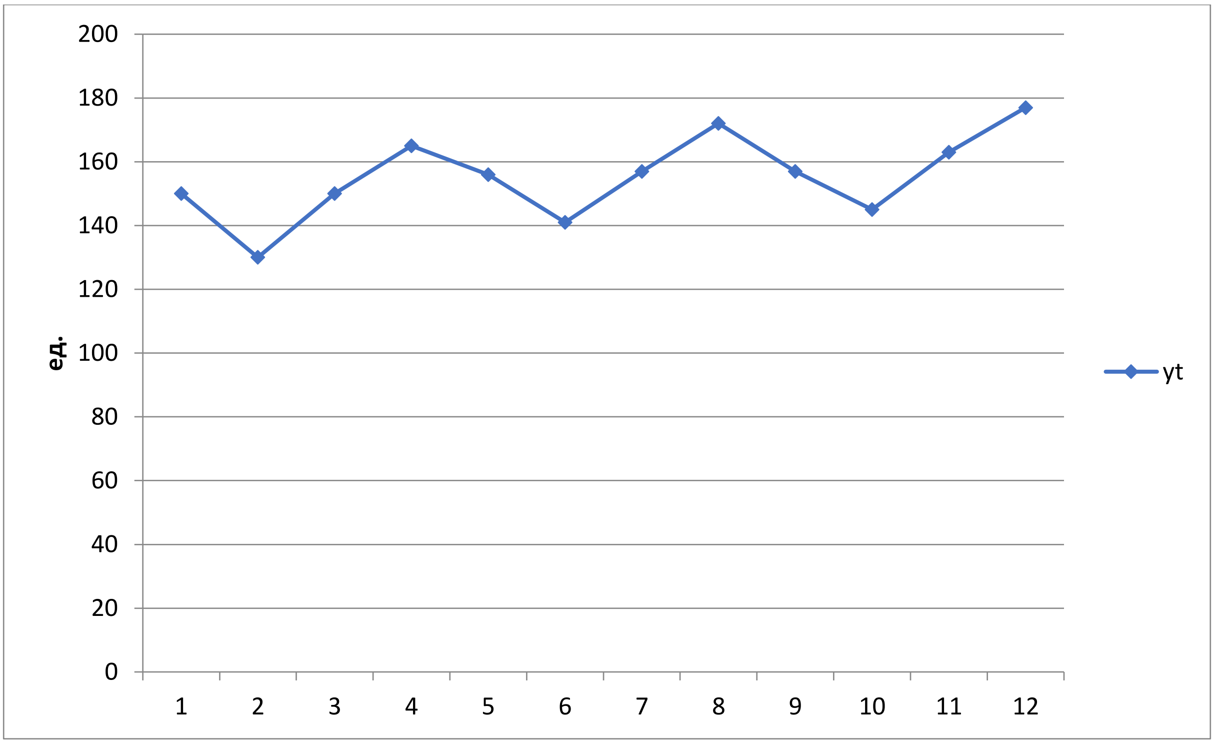



*Рис.2.4. Динамика изменения временного ряда с постоянной амплитудой*

В мультипликативной модели временного ряда компоненты представляют собой сомножители (2.37):

$$y_t = T(t) * S(t) * Z(t) * R(t) \tag{2.37}$$

Как правило, именно мультипликативная модель чаще всего применяется в экономических исследованиях [76].

Данный вид модели применяется в том случае, если на протяжении анализируемого промежутка времени амплитуда колебаний постоянно возрастает или уменьшается (рис.2.5 и рис.2.6).

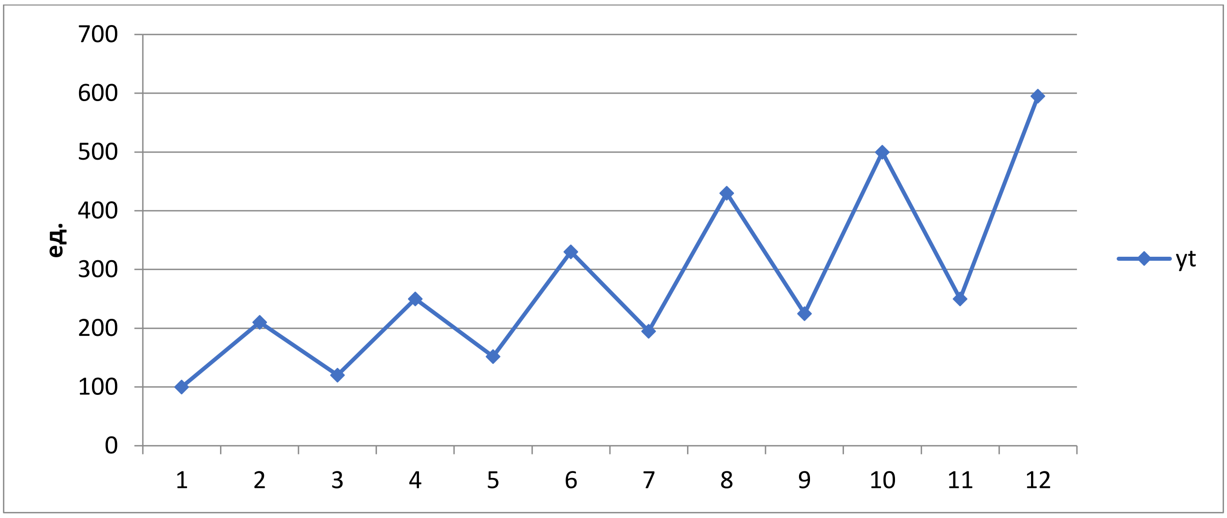



*Рис. 2.5. Динамика изменения временного ряда с возрастающей амплитудой*

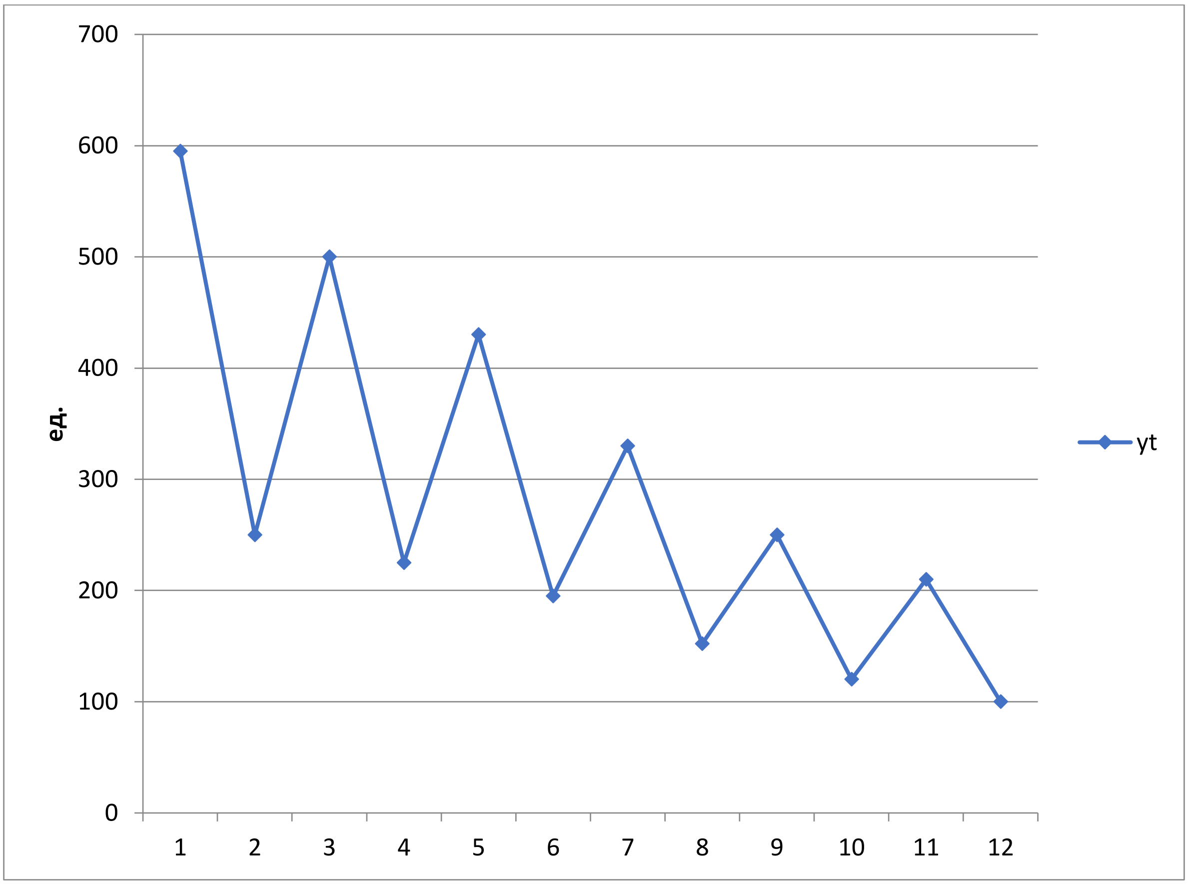



*Рис. 2.6. Динамика изменения временного ряда с амплитудой, имеющей тенденцию к снижению*

Смешанная модель представляет собой комбинированную модель временного ряда и записывается в следующем виде (2.38):

$$y_t = T(t) * S(t) * Z(t) + R(t) \tag{2.38}$$

При статистическом моделировании и прогнозировании следует четко понимать, какая из моделей наиболее точно описывает изменения временного ряда, с целью составления наиболее точного прогноза, учитывающего специфику рассматриваемых экономических процессов и явлений.

## Контрольные вопросы

1. Что такое временной ряд?
2. Какие существуют признаки для классификации временных рядов?
3. Какие основные показатели рядов динамики выделяют в статистическом моделировании и прогнозировании?
4. Отличаются ли способы вычисления основных показателей для различных видов рядов динамики?
5. Чем отличаются понятия «темп рост» и «темп прироста»? Каковы способы вычисления данных параметров?
6. В чем состоит проблема сопоставимости динамических рядов?
7. Какие предъявляются требования к сопоставимости временных рядов?
8. Каковы признаки сопоставимости для различных динамических рядов?
9. В чем состоит сущность абсолютного способа смыкания временного ряда?
10. Что представляет собой относительный способ смыкания динамического ряда?
11. Что такое «аномальный уровень ряда»?
12. Что представляют собой ошибки первого и второго рода?
13. Какие методы выявления аномальных значений уровней временного ряда Вам известны?
14. В чем состоит суть метода Ирвина?
15. В каких случаях применяется критерий Романовского?
16. В чем состоит сущность критерия вариационного размаха?

17. Что представляет собой основу критерия Диксона?
18. Каким образом определяется специфика применения конкретных методов выявления аномальных значений временного ряда?
19. Какие особенности графика свидетельствуют о том, что представленная модель временного ряда является аддитивной?
20. В чем состоят особенности мультипликативной модели?
21. Какие элементы ряда динамики Вам известны?

## ПРИМЕРЫ РЕШЕНИЯ ЗАДАЧ

**Пример 2.1**

В таблице приведены данные о внутренних текущих затратах на научные исследования и разработки в Российской Федерации за период с 2000 по 2020 год:

| Год | Величина внутренних текущих затрат на научные исследования и разработки, млн руб. |
|---|---|
| 2000 | 73873,3 |
| 2001 | 100507,4 |
| 2002 | 128243,3 |
| 2003 | 161202,7 |
| 2004 | 187210,5 |
| 2005 | 221119,5 |
| 2006 | 277784,8 |
| 2007 | 352917,7 |
| 2008 | 410865 |
| 2009 | 461006,2 |
| 2010 | 489450,8 |
| 2011 | 568386,7 |
| 2012 | 655061,7 |
| 2013 | 699948,9 |
| 2014 | 795407,9 |
| 2015 | 854288,0 |
| 2016 | 873778,7 |
| 2017 | 950257,0 |
| 2018 | 960689,4 |
| 2019 | 1060589,7 |
| 2020 | 1091333,5 |

На основании данных таблицы, определить основные показатели временного ряда.

**Решение.**

Первым этапом решения представленной задачи является определение вида ременного ряда. Поскольку приведены ежегодные данные, можно сделать вывод о том, что представленный временной ряд является равномерным.

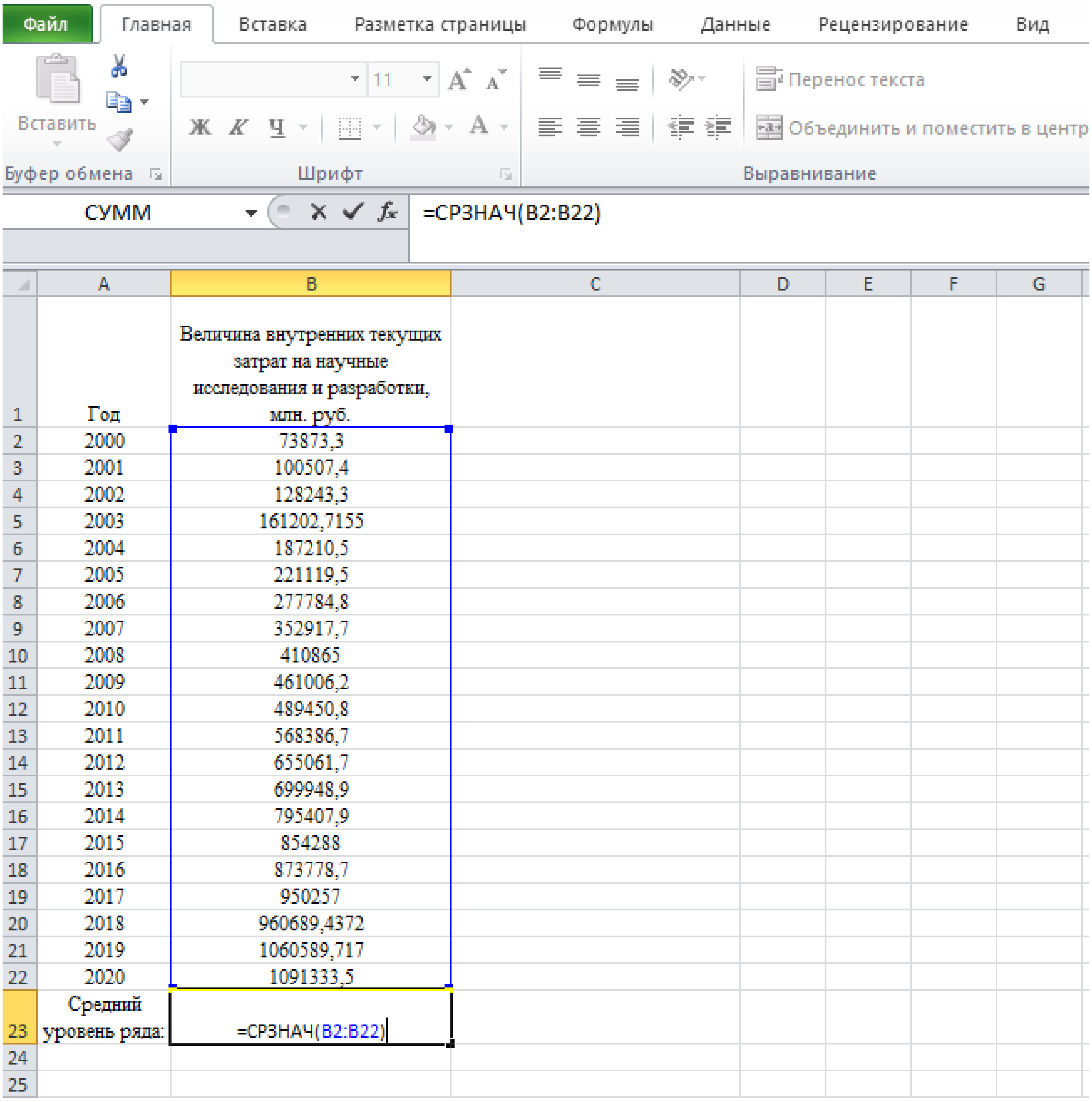
В данном ряду динамики  данные накапливаются за определенный промежуток времени, следовательно, он является интервальным.

Опираясь на сформулированные выводы, рассчитаем основные показатели представленного равномерного интервального ряда

**Средний ряд динамики**

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i = \frac{1}{21}\sum_{i=1}^{21} Y_i = \frac{1}{21} * 11373922{,}8 = 541615{,}37 \quad \text{млн.руб.}$$

Вычисления данного параметра в Ms Excel может быть произведено с помощью функции «срзнач»:




| | A | B |
|---|---|---|
| 1 | Год | Величина внутренних текущих затрат на научные исследования и разработки, млн. руб. |
| 2 | 2000 | 73873,3 |
| 3 | 2001 | 100507,4 |
| 4 | 2002 | 128243,3 |
| 5 | 2003 | 161202,7155 |
| 6 | 2004 | 187210,5 |
| 7 | 2005 | 221119,5 |
| 8 | 2006 | 277784,8 |
| 9 | 2007 | 352917,7 |
| 10 | 2008 | 410865 |
| 11 | 2009 | 461006,2 |
| 12 | 2010 | 489450,8 |
| 13 | 2011 | 568386,7 |
| 14 | 2012 | 655061,7 |
| 15 | 2013 | 699948,9 |
| 16 | 2014 | 795407,9 |
| 17 | 2015 | 854288 |
| 18 | 2016 | 873778,7 |
| 19 | 2017 | 950257 |
| 20 | 2018 | 960689,4372 |
| 21 | 2019 | 1060589,717 |
| 22 | 2020 | 1091333,5 |
| 23 | Средний уровень ряда: | =СРЗНАЧ(B2:B22) |
| 24 | | |
| 25 | | |

**Абсолютные приросты**

Вычисление базисного и цепного абсолютного прироста производилось по формулам:

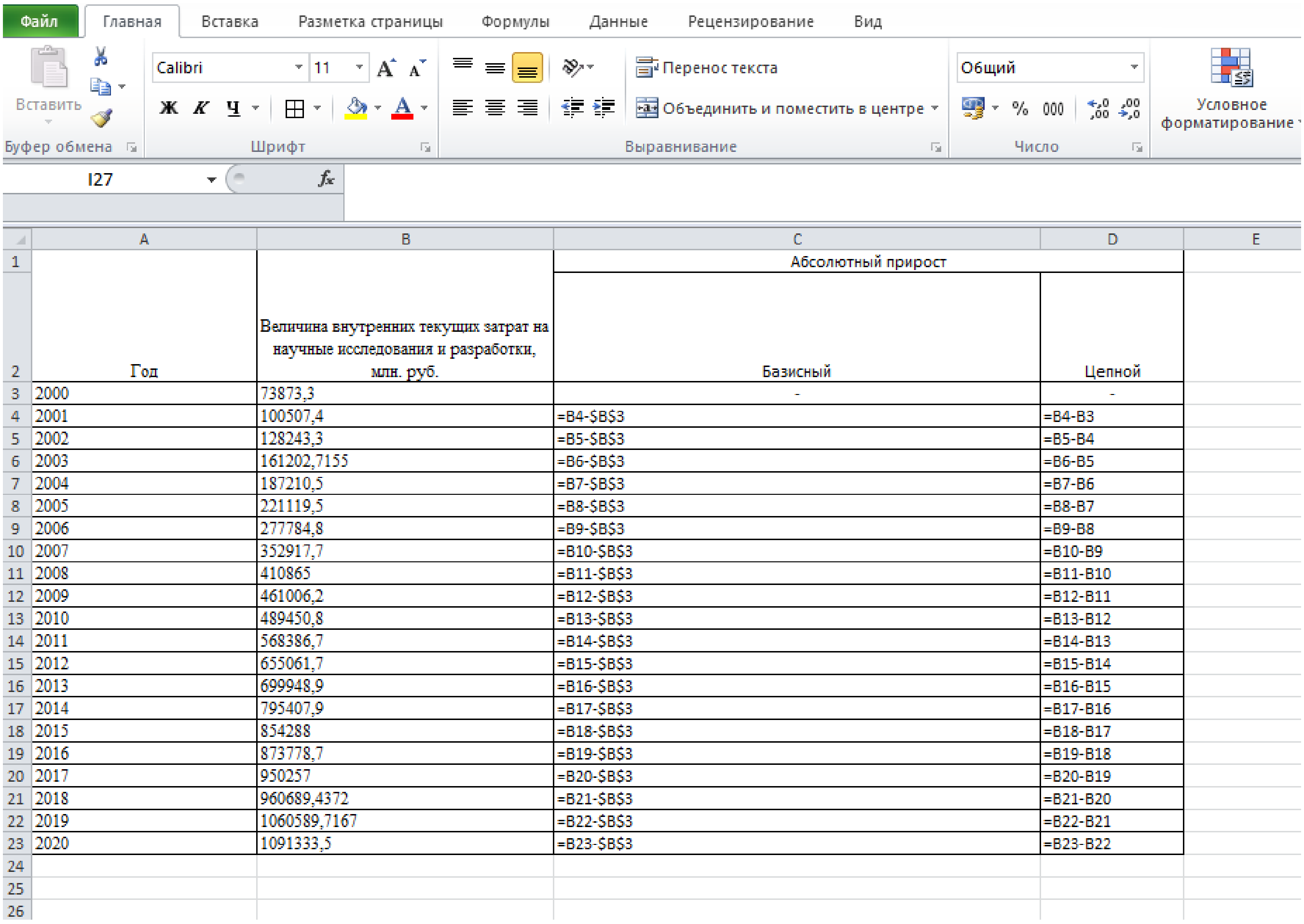
$$\Delta_{i\,базис.} = Y_i - Y_0,$$

$$\Delta_{i\,цепн.} = Y_i - Y_{i-1}.$$

Результаты вычислений для заданных исходных данных приведены в таблице:

| Год | Величина внутренних текущих затрат на научные исследования и разработки, млн. руб. | Абсолютный прирост | |
|---|---|---|---|
| | | Базисный | Цепной |
| 2000 | 73873,3 | - | - |
| 2001 | 100507,4 | 100507,4 - 73873,3 = 26634,1 | 100507,4 - 73873,3 = 26634,1 |
| 2002 | 128243,3 | 128243,3-73873,3 = 54370 | 128243,3-100507,4 = 27735,9 |
| 2003 | 161202,7 | 161202,7-73873,3 = 87329,4155 | 161202,7-128243,3=32959,42 |
| 2004 | 187210,5 | 113337,2 | 26007,78 |
| 2005 | 221119,5 | 147246,2 | 33909 |
| 2006 | 277784,8 | 203911,5 | 56665,3 |
| 2007 | 352917,7 | 279044,4 | 75132,9 |
| 2008 | 410865 | 336991,7 | 57947,3 |
| 2009 | 461006,2 | 387132,9 | 50141,2 |
| 2010 | 489450,8 | 415577,5 | 28444,6 |
| 2011 | 568386,7 | 494513,4 | 78935,9 |
| 2012 | 655061,7 | 581188,4 | 86675 |
| 2013 | 699948,9 | 626075,6 | 44887,2 |
| 2014 | 795407,9 | 721534,6 | 95459 |
| 2015 | 854288 | 780414,7 | 58880,1 |
| 2016 | 873778,7 | 799905,4 | 19490,7 |
| 2017 | 950257 | 876383,7 | 76478,3 |
| 2018 | 960689,4 | 886816,1 | 10432,4 |
| 2019 | 1060589,7 | 986716,4 | 99900,3 |
| 2020 | 1091333,5 | 1017460,2 | 30743,3 |

Представленное в формульном виде решение выглядит следующим образом:




| | A | B | C | D |
|---|---|---|---|---|
| 1 | | | Абсолютный прирост | |
| 2 | Год | Величина внутренних текущих затрат на научные исследования и разработки, млн. руб. | Базисный | Цепной |
| 3 | 2000 | 73873,3 | - | - |
| 4 | 2001 | 100507,4 | =B4-$B$3 | =B4-B3 |
| 5 | 2002 | 128243,3 | =B5-$B$3 | =B5-B4 |
| 6 | 2003 | 161202,7155 | =B6-$B$3 | =B6-B5 |
| 7 | 2004 | 187210,5 | =B7-$B$3 | =B7-B6 |
| 8 | 2005 | 221119,5 | =B8-$B$3 | =B8-B7 |
| 9 | 2006 | 277784,8 | =B9-$B$3 | =B9-B8 |
| 10 | 2007 | 352917,7 | =B10-$B$3 | =B10-B9 |
| 11 | 2008 | 410865 | =B11-$B$3 | =B11-B10 |
| 12 | 2009 | 461006,2 | =B12-$B$3 | =B12-B11 |
| 13 | 2010 | 489450,8 | =B13-$B$3 | =B13-B12 |
| 14 | 2011 | 568386,7 | =B14-$B$3 | =B14-B13 |
| 15 | 2012 | 655061,7 | =B15-$B$3 | =B15-B14 |
| 16 | 2013 | 699948,9 | =B16-$B$3 | =B16-B15 |
| 17 | 2014 | 795407,9 | =B17-$B$3 | =B17-B16 |
| 18 | 2015 | 854288 | =B18-$B$3 | =B18-B17 |
| 19 | 2016 | 873778,7 | =B19-$B$3 | =B19-B18 |
| 20 | 2017 | 950257 | =B20-$B$3 | =B20-B19 |
| 21 | 2018 | 960689,4372 | =B21-$B$3 | =B21-B20 |
| 22 | 2019 | 1060589,7167 | =B22-$B$3 | =B22-B21 |
| 23 | 2020 | 1091333,5 | =B23-$B$3 | =B23-B22 |

**Средний абсолютный прирост** можно вычислить по формуле:

$$\bar{\Delta} = \frac{Y_n - Y_0}{n-1} = \frac{1091333{,}5 - 73873{,}3}{21-1} = 50873{,}01 \text{ млн.руб.}$$

Также нахождение данного показателя временного ряда возможно по формуле:

$$\bar{\Delta} = \frac{\sum_{i=1}^{k} \Delta_{i\,\text{цепн}}}{K}$$

Учитывая, что сумма цепных абсолютных приростов равна базисному абсолютному приросту за весь анализируемый период, результат, вычисленный по каждой из предложенных формул, будет одинаковым.

| Год | Величина внутренних текущих затрат на научные исследования и разработки, млн. руб. | Коэффициенты роста | | Темпы роста | |
|---|---|---|---|---|---|
| | | Базисный | Цепной | Базисный | Цепной |
| 2000 | 73873,3 | =73873,3/ 73873,3=1 | - | =73873,3/ 73873,3*100% = 100,00% | - |
| 2001 | 100507,4 | =100507,4/ 73873,3 =1,36 | =100507,4/ 73873,3 =1,36 | =100507,4/ 73873,3* 100% =136,05% | =100507,4/ 73873,3 *100% = 136,05% |
| 2002 | 128243,3 | =128243,3/ 73873,3 =1,74 | =128243,3/ 100507,4= 1,28 | =128243,3/ 73873,3 *100% = 173,60% | =128243,3/ 100507,4*100% = 127,60% |
| 2003 | 161202,7 | =161202,7/ 73873,3 =2,18 | =161202,7/ 128243,3= 1,26 | =161202,7/ 73873,3 *100% = 218,22% | =1612025,7/ 128243,3*100% = 125,70% |
| 2004 | 187210,5 | 2,53 | 1,16 | 253,42% | 116,13% |
| 2005 | 221119,5 | 2,99 | 1,18 | 299,32% | 118,11% |
| 2006 | 277784,8 | 3,76 | 1,26 | 376,03% | 125,63% |
| 2007 | 352917,7 | 4,78 | 1,27 | 477,73% | 127,05% |
| 2008 | 410865,0 | 5,56 | 1,16 | 556,18% | 116,42% |
| 2009 | 461006,2 | 6,24 | 1,12 | 624,05% | 112,20% |
| 2010 | 489450,8 | 6,63 | 1,06 | 662,55% | 106,17% |
| 2011 | 568386,7 | 7,69 | 1,16 | 769,41% | 116,13% |
| 2012 | 655061,7 | 8,87 | 1,15 | 886,74% | 115,25% |
| 2013 | 699948,9 | 9,47 | 1,07 | 947,50% | 106,85% |
| 2014 | 795407,9 | 10,77 | 1,14 | 1076,72% | 113,64% |
| 2015 | 854288,0 | 11,56 | 1,07 | 1156,42% | 107,40% |
| 2016 | 873778,7 | 11,83 | 1,02 | 1182,81% | 102,28% |
| 2017 | 950257,0 | 12,86 | 1,09 | 1286,33% | 108,75% |
| 2018 | 960689,4 | 13,00 | 1,01 | 1300,46% | 101,10% |
| 2019 | 1060589,7 | 14,36 | 1,10 | 1435,69% | 110,40% |
| 2020 | 1091333,5 | 14,77 | 1,03 | 1477,30% | 102,90% |

Таким образом, величина внутренних текущих затрат на научные исследования и разработки увеличивалась в течение всего анализируемого промежутка времени. Наибольший абсолютный прирост по сравнению с базисным периодом наблюдается с 2020 году.

Анализируя цепные абсолютные приросты, можно сделать вывод, что наибольший рост (на 99900,3 млн. руб.) показателя по сравнению с уровнем предыдущего периода имеет место в 2019 году.

В среднем абсолютный прирост величины внутренних текущих затрат на научные исследования и разработки за весь анализируемый период с 2000 по 2020 годы составил 50843,01 млн. руб.

**Коэффициенты и темпы роста**

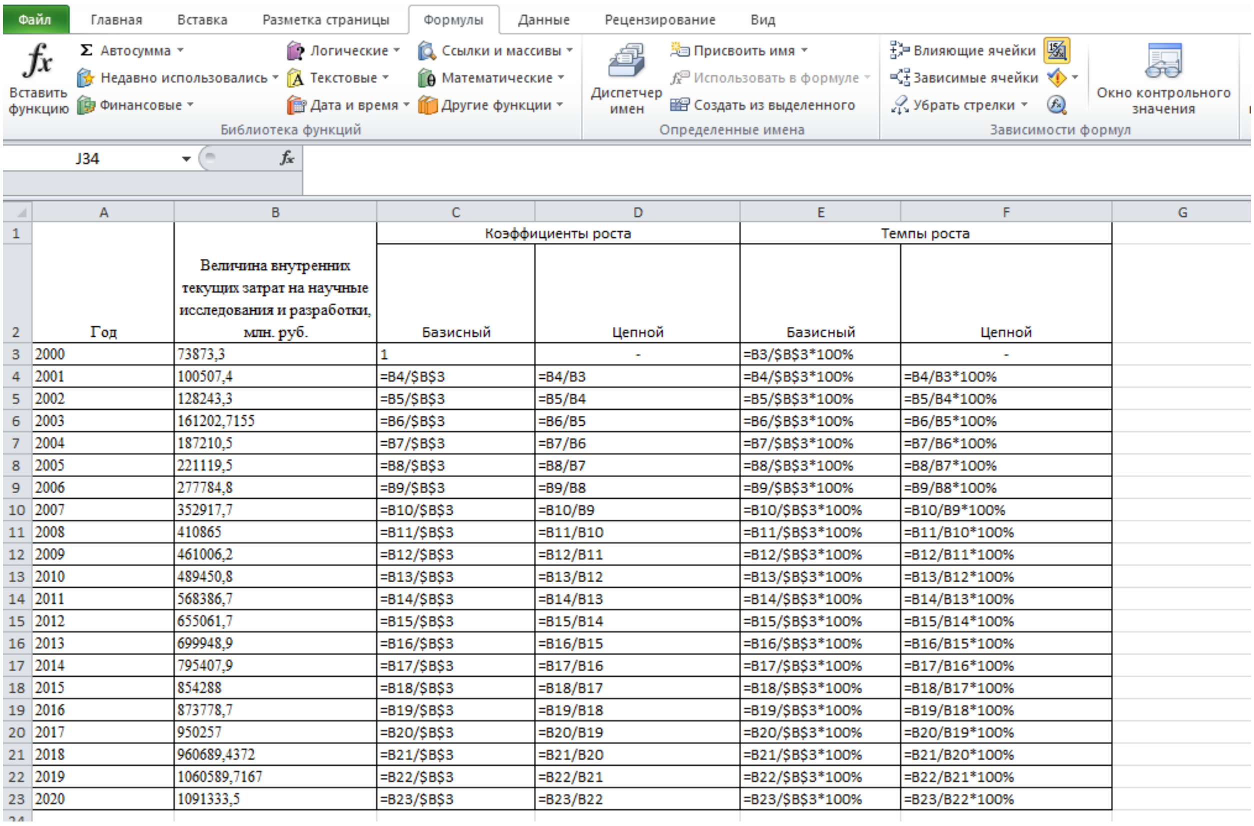
$$T_{рб} = \frac{Y_n}{Y_0} * 100\%$$ - базисный темп роста

$$T_{рц} = \frac{Y_n}{Y_{n-1}} * 100\%$$ - цепной темп роста

$$K_{р_б} = \frac{Y_n}{Y_0}$$, - базисный коэффициент роста

$$K_{р_ц} = \frac{Y_n}{Y_{n-1}}$$ - цепной коэффициент роста

Представленное в формульном виде решение выглядит следующим образом:




| | A | B | C | D | E | F |
|---|---|---|---|---|---|---|
| 1 | | | Коэффициенты роста | | Темпы роста | |
| 2 | Год | Величина внутренних текущих затрат на научные исследования и разработки, млн. руб. | Базисный | Цепной | Базисный | Цепной |
| 3 | 2000 | 73873,3 | 1 | - | =B3/$B$3*100% | - |
| 4 | 2001 | 100507,4 | =B4/$B$3 | =B4/B3 | =B4/$B$3*100% | =B4/B3*100% |
| 5 | 2002 | 128243,3 | =B5/$B$3 | =B5/B4 | =B5/$B$3*100% | =B5/B4*100% |
| 6 | 2003 | 161202,7155 | =B6/$B$3 | =B6/B5 | =B6/$B$3*100% | =B6/B5*100% |
| 7 | 2004 | 187210,5 | =B7/$B$3 | =B7/B6 | =B7/$B$3*100% | =B7/B6*100% |
| 8 | 2005 | 221119,5 | =B8/$B$3 | =B8/B7 | =B8/$B$3*100% | =B8/B7*100% |
| 9 | 2006 | 277784,8 | =B9/$B$3 | =B9/B8 | =B9/$B$3*100% | =B9/B8*100% |
| 10 | 2007 | 352917,7 | =B10/$B$3 | =B10/B9 | =B10/$B$3*100% | =B10/B9*100% |
| 11 | 2008 | 410865 | =B11/$B$3 | =B11/B10 | =B11/$B$3*100% | =B11/B10*100% |
| 12 | 2009 | 461006,2 | =B12/$B$3 | =B12/B11 | =B12/$B$3*100% | =B12/B11*100% |
| 13 | 2010 | 489450,8 | =B13/$B$3 | =B13/B12 | =B13/$B$3*100% | =B13/B12*100% |
| 14 | 2011 | 568386,7 | =B14/$B$3 | =B14/B13 | =B14/$B$3*100% | =B14/B13*100% |
| 15 | 2012 | 655061,7 | =B15/$B$3 | =B15/B14 | =B15/$B$3*100% | =B15/B14*100% |
| 16 | 2013 | 699948,9 | =B16/$B$3 | =B16/B15 | =B16/$B$3*100% | =B16/B15*100% |
| 17 | 2014 | 795407,9 | =B17/$B$3 | =B17/B16 | =B17/$B$3*100% | =B17/B16*100% |
| 18 | 2015 | 854288 | =B18/$B$3 | =B18/B17 | =B18/$B$3*100% | =B18/B17*100% |
| 19 | 2016 | 873778,7 | =B19/$B$3 | =B19/B18 | =B19/$B$3*100% | =B19/B18*100% |
| 20 | 2017 | 950257 | =B20/$B$3 | =B20/B19 | =B20/$B$3*100% | =B20/B19*100% |
| 21 | 2018 | 960689,4372 | =B21/$B$3 | =B21/B20 | =B21/$B$3*100% | =B21/B20*100% |
| 22 | 2019 | 1060589,7167 | =B22/$B$3 | =B22/B21 | =B22/$B$3*100% | =B22/B21*100% |
| 23 | 2020 | 1091333,5 | =B23/$B$3 | =B23/B22 | =B23/$B$3*100% | =B23/B22*100% |

Средний темп роста определяется по формуле средней геометрической:

$$\overline{T_p} = \sqrt[n-1]{T_{рц_1} * T_{рц_2} * ... * T_{рц_{n-1}}}$$

Таким образом, для исходных данных вычисление будет следующим:

$$\overline{T_p} = \sqrt[21-1]{136{,}05\% * 127{,}6\% * 125{,}7\% * ... * 110{,}4\% * 102{,}9\%} = 114{,}41\%$$

Коэффициенты и темпы прироста

$T_{пр_б} = T_{р_б} - 100\%$ - базисный темп прироста

$T_{пр_ц} = T_{р_ц} - 100\%$ - цепной темп прироста

$K_{пр_б} = K_{р_б} - 1$ - базисный коэффициент прироста

$K_{пр_ц} = K_{р_ц} - 1$ - цепной коэффициент прироста

Результаты вычислений для заданных исходных данных приведены в таблице:

| Год | Величина внутренних текущих затрат на научные исследования и разработки, млн. руб. | Коэффициенты роста | | Темпы роста | |
|---|---|---|---|---|---|
| | | Базисный | Цепной | Базисный | Цепной |
| 2000 | 73873,3 | 1-1=0 | - | 100,00% - 100,00% = 0 | - |
| 2001 | 100507,4 | 1,36-1=0,36 | =1,36-1=0,36 | 136,05%-100% = 36% | 136,05%-100% = 36% |
| 2002 | 128243,3 | 1,74-1=0,74 | 1,28-1=0,28 | 173,60%-100% = 73,6% | 127,60%-100% = 27,6% |
| 2003 | 161202,7 | 2,18-1=1,18 | 1,26-1=0,26 | 218,22%-100% = 118,22% | 125,70%-100% = 25,7% |
| 2004 | 187210,5 | 1,53 | 0,16 | 153,42% | 16,13% |
| 2005 | 221119,5 | 1,99 | 0,18 | 199,32% | 18,11% |
| 2006 | 277784,8 | 2,76 | 0,26 | 276,03% | 25,63% |
| 2007 | 352917,7 | 3,78 | 0,27 | 377,73% | 27,05% |
| 2008 | 410865,0 | 4,56 | 0,16 | 456,18% | 16,42% |

| | | | | | |
|---|---|---|---|---|---|
| 2009 | 461006,2 | 5,24 | 0,12 | 524,05% | 12,20% |
| 2010 | 489450,8 | 5,63 | 0,06 | 562,55% | 06,17% |
| 2011 | 568386,7 | 6,69 | 0,16 | 669,41% | 16,13% |
| 2012 | 655061,7 | 7,87 | 0,15 | 786,74% | 15,25% |
| 2013 | 699948,9 | 8,47 | 0,07 | 847,50% | 06,85% |
| 2014 | 795407,9 | 9,77 | 0,14 | 976,72% | 13,64% |
| 2015 | 854288,0 | 10,56 | 0,07 | 1056,42% | 7,40% |
| 2016 | 873778,7 | 10,83 | 0,02 | 1082,81% | 2,28% |
| 2017 | 950257,0 | 11,86 | 0,09 | 1086,33% | 8,75% |
| 2018 | 960689,4 | 12,00 | 0,01 | 1200,46% | 1,10% |
| 2019 | 1060589,7 | 13,36 | 0,10 | 1335,69% | 10,40% |
| 2020 | 1091333,5 | 13,77 | 0,03 | 1377,30% | 2,90% |

**Средний темп прироста** определяется по формуле:

$$\overline{\mathrm{T}_{\text{пр}}} = \mathrm{T}_{\text{р}} - 100\% = 114{,}41 - 100\% = 14{,}41\%$$

Таким образом, в течение рассматриваемого временного интервала темп роста составляет свыше 100%, что свидетельствует о постоянном увеличении значения показателя как по сравнению с базисным периодом, так и об увеличении каждого последующего значения временного ряда по сравнению с предыдущим.

Максимальный темп прироста по сравнению с начальным значением ряда наблюдается в 2020 году. Анализ цепных приростов позволяет сделать вывод о том, что наибольшее изменение величины внутренних текущих затрат на научные исследования и разработки (на 10,4%) наблюдается в 2019 году по сравнению с 2018 годом.

В среднем темп роста внутренних текущих затрат на научные исследования и разработки составляет 114,41%, а темп прироста 14,41%.

**Абсолютное значение одного процента** прироста определим по формуле:

$$|\mathrm{A}| = \frac{\Delta_{\text{цепн}}}{\mathrm{T}_{\text{пр}_{\text{ц}}}}$$

Для вычислений воспользуемся таблицей, содержащей необходимые данные для расчета:

| Год | Величина внутренних текущих затрат на научные исследования и разработки, млн. руб. | Абсолютный цепной прирост, млн. руб. | Цепной тест прироста | Абсолютное значение одного процента прироста, млн. руб. |
|---|---|---|---|---|
| 2000 | 73873,3 | - | - | - |
| 2001 | 100507,4 | 26634,1 | 36,05% | 26634,1/36,05 =738,81 |
| 2002 | 128243,3 | 27735,9 | 27,60% | 27735,9/27,60 =769,37 |
| 2003 | 161202,7 | 32959,4 | 25,70% | 32959,4/25,70 = 914,27 |
| 2004 | 187210,5 | 26007,8 | 16,13% | 721,44 |
| 2005 | 221119,5 | 33909,0 | 18,11% | 940,61 |
| 2006 | 277784,8 | 56665,3 | 25,63% | 1571,85 |
| 2007 | 352917,7 | 75132,9 | 27,05% | 2084,13 |
| 2008 | 410865,0 | 57947,3 | 16,42% | 1607,41 |
| 2009 | 461006,2 | 50141,2 | 12,20% | 1390,88 |
| 2010 | 489450,8 | 28444,6 | 6,17% | 789,03 |
| 2011 | 568386,7 | 78935,9 | 16,13% | 2189,62 |
| 2012 | 655061,7 | 86675,0 | 15,25% | 2404,30 |
| 2013 | 699948,9 | 44887,2 | 6,85% | 1245,14 |
| 2014 | 795407,9 | 95459,0 | 13,64% | 2647,96 |
| 2015 | 854288,0 | 58880,1 | 7,40% | 1633,29 |
| 2016 | 873778,7 | 19490,7 | 2,28% | 540,66 |
| 2017 | 950257,0 | 76478,3 | 8,75% | 2121,45 |
| 2018 | 960689,4 | 10432,4 | 1,10% | 289,39 |
| 2019 | 1060589,7 | 99900,3 | 10,40% | 2771,16 |
| 2020 | 1091333,5 | 30743,8 | 2,90% | 852,81 |

Наибольшее абсолютное значение, соответствующее процентному приросту, наблюдается в 2019 году и составляет 2771,16 млн. руб. Наименьшее значение показателя имеет место 2018 году: абсолютное значение, которое соответствует процентному приросту величины внутренних текущих затрат на научные исследования и разработки по сравнению с 2017 годом, в данном периоде составляет 289.39 млн. руб.

**Пример 2.2.**

Имеются данные о численности населения Белгородской, Брянской и Владимирской области на конец года:

| | Численность населения, тыс. чел. | | |
|---|---|---|---|
| | Белгородская область | Брянская область | Владимирская область |
| 2000 | 1507,0 | 1407,9 | 1558,0 |
| 2001 | 1508,1 | 1391,4 | 1539,2 |
| 2002 | 1511,9 | 1375,0 | 1520,1 |
| 2003 | 1513,9 | 1360,2 | 1509,6 |
| 2004 | 1511,7 | 1344,1 | 1497,6 |
| 2005 | 1511,7 | 1327,7 | 1486,5 |
| 2006 | 1514,2 | 1312,7 | 1475,9 |
| 2007 | 1520,1 | 1303,3 | 1466,8 |
| 2008 | 1526,3 | 1294,3 | 1457,9 |
| 2009 | 1531,8 | 1286,5 | 1449,8 |
| 2010 | 1532,4 | 1275,3 | 1441,1 |
| 2011 | 1536,1 | 1264,4 | 1431,9 |
| 2012 | 1541,0 | 1253,6 | 1421,7 |
| 2013 | 1544,1 | 1242,6 | 1413,3 |
| 2014 | 1547,9 | 1233,0 | 1405,6 |
| 2015 | 1550,1 | 1225,8 | 1397,2 |
| 2016 | 1552,9 | 1220,5 | 1389,6 |
| 2017 | 1549,9 | 1211,0 | 1378,3 |
| 2018 | 1547,4 | 1200,2 | 1365,8 |
| 2019 | 1549,2 | 1192,5 | 1358,4 |
| 2020 | 1541,3 | 1182,7 | 1342,1 |

На основании данных таблицы рассчитайте средний уровень временного ряда для каждого из представленных субъектов Российской Федерации, сформулируйте выводы.

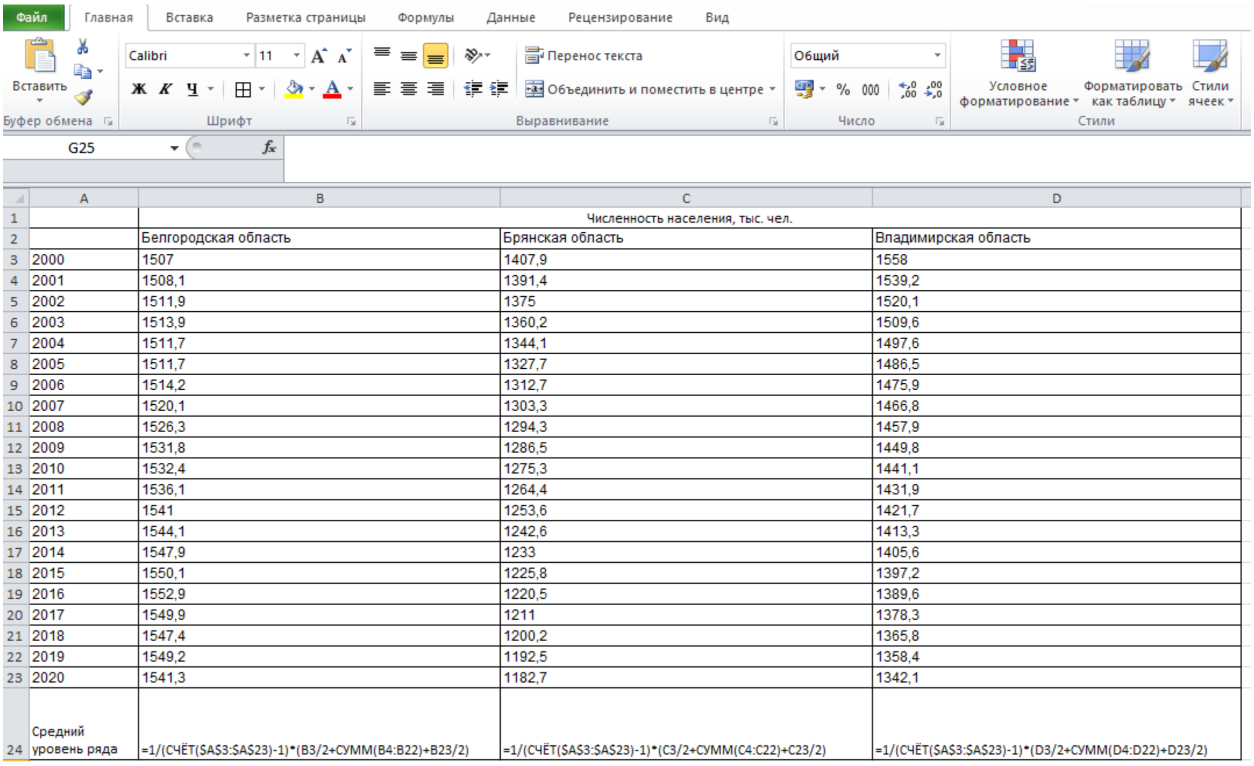
**Решение.**

Поскольку временной ряд, характеризующий изменение численности населения, рассчитывается на конец года, то есть в конкретный момент времени, можно сделать вывод о том, что совокупность значений ряда образует моментный временной ряд. Значения приведены ежегодные, следовательно, представленные ряды являются равномерными

Таким образом, для расчета среднего уровня ряда динамики необходимо применить формулу:

$$\bar{Y} = \frac{1}{n-1}(\frac{Y_1}{2} + Y_2 + Y_3 + ...Y_{n-1} + \frac{Y_n}{2})$$

Представленное в формульном виде решение выглядит следующим образом:




| | A | B | C | D |
|---|---|---|---|---|
| 1 | | | Численность населения, тыс. чел. | |
| 2 | | Белгородская область | Брянская область | Владимирская область |
| 3 | 2000 | 1507 | 1407,9 | 1558 |
| 4 | 2001 | 1508,1 | 1391,4 | 1539,2 |
| 5 | 2002 | 1511,9 | 1375 | 1520,1 |
| 6 | 2003 | 1513,9 | 1360,2 | 1509,6 |
| 7 | 2004 | 1511,7 | 1344,1 | 1497,6 |
| 8 | 2005 | 1511,7 | 1327,7 | 1486,5 |
| 9 | 2006 | 1514,2 | 1312,7 | 1475,9 |
| 10 | 2007 | 1520,1 | 1303,3 | 1466,8 |
| 11 | 2008 | 1526,3 | 1294,3 | 1457,9 |
| 12 | 2009 | 1531,8 | 1286,5 | 1449,8 |
| 13 | 2010 | 1532,4 | 1275,3 | 1441,1 |
| 14 | 2011 | 1536,1 | 1264,4 | 1431,9 |
| 15 | 2012 | 1541 | 1253,6 | 1421,7 |
| 16 | 2013 | 1544,1 | 1242,6 | 1413,3 |
| 17 | 2014 | 1547,9 | 1233 | 1405,6 |
| 18 | 2015 | 1550,1 | 1225,8 | 1397,2 |
| 19 | 2016 | 1552,9 | 1220,5 | 1389,6 |
| 20 | 2017 | 1549,9 | 1211 | 1378,3 |
| 21 | 2018 | 1547,4 | 1200,2 | 1365,8 |
| 22 | 2019 | 1549,2 | 1192,5 | 1358,4 |
| 23 | 2020 | 1541,3 | 1182,7 | 1342,1 |
| 24 | Средний уровень ряда | =1/(СЧЁТ($A$3:$A$23)-1)*(B3/2+СУММ(B4:B22)+B23/2) | =1/(СЧЁТ($A$3:$A$23)-1)*(C3/2+СУММ(C4:C22)+C23/2) | =1/(СЧЁТ($A$3:$A$23)-1)*(D3/2+СУММ(D4:D22)+D23/2) |

В результаты вычислений было выявлено, что среднее значение численности населения за период с 2000 по 2020 год в Белгородской области составило 1531,24 тыс.чел., в Брянской области 1280, 47 тыс. чел, во Владимирской области 1442,82 тыс. чел.

Таким образом, становится очевидным, что наибольшее среднее значение уровня временного ряда среди данных субъектов наблюдается в Белгородской области, а наименьшее – в Брянской области.

**Пример 2.3**

В таблице приведены данные о количестве малых предприятий в Российской Федерации:

| | 1999 | 2000 | 2001 | 2002 | 2003 |
|---|---|---|---|---|---|
| Число малых предприятий (по старой методологии) | 54 | 89 | 125 | - | - |
| Число малых предприятий (по новой методологии) | - | - | 90 | 138 | 169 |

На основании представленных данных провести процедуру смыкания абсолютным и относительным способом.

**Решение.**

Произведем смыкание двух динамических рядов в один. Для реализации абсолютного способа выполним следующие действия:

1. По данным за 2001 г. рассчитаем коэффициент пересчета - соотношение величин, вычисленных по новой и старой методологии:

$$k = 90/125 = 0,72$$

2. Данные за 1999 г. и 2000 г. переведем в вид, соответствующий новой методологии расчета, для чего умножим показатели этих лет на полученный коэффициент пересчета. Таким образом, временной ряд после процедуры смыкания примет следующий вид:

| | 1999 | 2000 | 2001 | 2002 | 2003 |
|---|---|---|---|---|---|
| Сомкнутый динамический ряд (с использованием абсолютного способа смыкания) | = 0,72*54 = 39 | =0,72*89 = 64 | 90 | 138 | 169 |

Для реализации относительного способа выполним следующие действия:

1. Год, в котором произошли методологические изменения, примем за 100 % (в данном примере 2001)
2. Остальные уровни пересчитаем в процентах к нему: для уровней, исчисленных по старой методологии, в качестве базы сравнения выступают значение показателя в 2001 г., определенное по старой методологии (это - 125), для уровней, рассчитанных по новой методологии, - значение показателя в 2001 г., исчисленное по новой методологии (это - 90).

Расчеты и результаты вычислений с использованием относительного способа смыкания приведены в таблице:

| | 1999 | 2000 | 2001 | 2002 | 2003 |
|---|---|---|---|---|---|
| Сомкнутый динамический ряд (с относительного способа смыкания) | 54/125 * 100% ≈ 43% | 89/125 * 100% ≈ 71% | 100% | 138/90 * 100% ≈ 153% | 169/90 * 100 % ≈ 188% |

Таким образом, для смыкания временных рядов можно использовать абсолютный и относительный способ. Применение конкретного метода должно осуществляться с учетом конкретных исследовательских задач.

**Пример 2.4.**

В таблице приведены данные ряда динамики по некоторому наблюдению:

| Номер наблюдения | Значение уровня временного ряда |
|---|---|
| 1 | 4,88 |
| 2 | 4,69 |
| 3 | 4,79 |
| 4 | 4,84 |
| 5 | 4,69 |
| 6 | 4,88 |
| 7 | 4,91 |
| 8 | 4,65 |
| 9 | 4,89 |
| 10 | 5,75 |

Последнее наблюдение ставится под сомнение. Произвести обработку результатов измерений по обнаружению грубых погрешностей, используя статистический критерий Романовского. Уровень значимости принять равным 0,99.

**Решение.**

Число наблюдений составляет менее 20, таким образом, применение критерия Романовского для данного временного ряда представляется возможным.

Проверка по критерию Романовского предполагает определение всех характеристик без учета подозрительного результата.

Таким образом, вычислим значение параметра $\overline{y}$:

$$\overline{y} = \frac{4{,}88 + 4{,}69 + 4{,}79 + 4{,}84 + 4{,}69 + 4{,}88 + 4{,}91 + 4{,}65 + 4{,}89}{9} = 4{,}8$$

Для нахождения $\delta_y$ воспользуемся вспомогательной таблицей:

| Исходное значение | Отклонение от среднего | Квадраты отклонений |
|---|---|---|
| 4,88 | 0,08 | 0,01 |
| 4,69 | -0,11 | 0,01 |
| 4,79 | -0,01 | 0,00 |
| 4,84 | 0,04 | 0,00 |
| 4,69 | -0,11 | 0,01 |
| 4,88 | 0,08 | 0,01 |
| 4,91 | 0,11 | 0,01 |
| 4,65 | -0,15 | 0,02 |
| 4,89 | 0,09 | 0,01 |

Тогда $\delta_y$ составит:

$$\delta_y = \sqrt{\frac{1{,}02}{9-1}} = 0{,}36$$

Найдем значение параметра $\beta_t$:

$$\beta_t = \frac{|5{,}75 - 4{,}8|}{0{,}36} = 2{,}66$$

При заданных условиях $\beta_q = 2{,}62$.

Так как $\beta_t \geq \beta_q$ (2,66>2,62), то значение уровня ряда считается промахом и отбрасывается. Таким образом, последний результат измерения необходимо не учитывать.

**Пример 2.5.**

Временной ряд задан в табличной форме:

| Номер наблюдения | Значение уровня временного ряда |
|---|---|
| 1 | 1,6 |
| 2 | 1,9 |
| 3 | 2,1 |
| 4 | 2,4 |
| 5 | 4,5 |
| 6 | 2,8 |
| 7 | 3,1 |
| 8 | 3,3 |
| 9 | 3,6 |
| 10 | 3,8 |

С помощью метода Ирвина проверить наличие аномальных значений во втором и пятом наблюдении.

**Решение.**

Вычислим среднее значение временного ряда:

$$\overline{y} = \frac{1{,}6 + 1{,}9 + 2{,}1 + 2{,}4 + 4{,}5 + 2{,}8 + 3{,}1 + 3{,}3 + 3{,}6 + 3{,}8}{10} = 2{,}91$$

Для нахождения $\delta_y$ воспользуемся вспомогательной таблицей:

| Исходное значение | Отклонение от среднего | Квадраты отклонений |
|---|---|---|
| 1,6 | -1,31 | 1,7161 |
| 1,9 | -1,01 | 1,0201 |
| 2,1 | -0,81 | 0,6561 |
| 2,4 | -0,51 | 0,2601 |
| 4,5 | 1,59 | 2,5281 |
| 2,8 | -0,11 | 0,0121 |
| 3,1 | 0,19 | 0,0361 |
| 3,3 | 0,39 | 0,1521 |
| 3,6 | 0,69 | 0,4761 |

Тогда $\delta_y$ составит:

$$\delta_y = \sqrt{\frac{7{,}649}{10 - 1}} = 0{,}92$$

Исследуем на аномальные значения точки t = 2 и t = 5:

Найдем значение параметра $\lambda_2$:

$$\lambda_2 = \frac{|1{,}9 - 1{,}6|}{0{,}92} = 0{,}33$$

Найдем значение параметра $\lambda_5$:

$$\lambda_5 = \frac{|4{,}5 - 2{,}4|}{0{,}92} = 2{,}28$$

Так как $\lambda_2 = 0{,}33$, а значение $\lambda_\alpha = 1{,}5$ (при уровне значимости 95% и t=10) , можно сделать вывод о том, что для t=2 значение уровня 1,9 является нормальным (0,33<1,5).

Так как $\lambda_5 = 2{,}28$, а значение $\lambda_\alpha = 1{,}5$ (при уровне значимости 95% и t=10) , можно сделать вывод о том, что для t=5 значение уровня 1,9 является аномальным (2,28>1,5).

**Пример 2.6.**

Временной ряд задан в табличной форме:

| Номер наблюдения | Значение уровня временного ряда |
|---|---|
| 1 | 36 |
| 2 | 65 |
| 3 | 40 |
| 4 | 41,5 |
| 5 | 42,5 |
| 6 | 51 |
| 7 | 44 |
| 8 | 46,5 |
| 9 | 38 |
| 10 | 33 |

Провести проверку на наличие грубых ошибок по критерию Диксона при доверительной вероятности 0,95, если известно, что распределение показателя соответствует нормальному.

**Решение.**

Предварительно упорядочим временной ряд по возрастанию. Тогда отсортированный ряд примет следующий вид:

| Уровень ряда | Значение |
|---|---|
| 1 | 33 |
| 2 | 36 |
| 3 | 38 |
| 4 | 40 |
| 5 | 41,5 |
| 6 | 42,5 |
| 7 | 44 |
| 8 | 46,5 |
| 9 | 51 |
| 10 | 65 |

В вариационном ряду выглядит сомнительно наибольшее значение ряда 65. Тогда для одного одностороннего выброса найдем коэффициент Диксона:

$$r = \frac{65 - 51}{65 - 36} = 0,38$$

Вычисленное значение коэффициента Диксона сравним с табличным значением. Для заданного объема выборки (n=10) $r_q$ составляет 0,477

Так как $r < r_q$, нулевая гипотеза об отсутствии грубой погрешности выполняется, следовательно, результат наблюдения 65 является нормальным.

**Пример 2.7.**

Временной ряд задан в табличной форме:

| Номер наблюдения | Значение уровня временного ряда |
|---|---|
| 1 | 3,42 |
| 2 | 3,43 |
| 3 | 3,44 |
| 4 | 3,45 |
| 5 | 3,46 |
| 6 | 3,47 |
| 7 | 4,21 |

Провести проверку седьмого наблюдения по критерию Райта и правилу «трех сигм».

**Решение.**

Вычислим среднее значение временного ряда без учета подозрительного результата:

$$\overline{y} = \frac{3{,}42 + 3{,}43 + 3{,}44 + 3{,}45 + 3{,}46 + 3{,}47 + 4{,}21}{7} = 3{,}55$$

Для нахождения $\delta_y$ воспользуемся вспомогательной таблицей:

| Исходное значение | Отклонение от среднего | Квадраты отклонений |
|---|---|---|
| 3,42 | -0,02 | 0,0006 |
| 3,43 | -0,01 | 0,0002 |
| 3,44 | 0,00 | 0,0000 |
| 3,45 | 0,01 | 0,0000 |
| 3,46 | 0,02 | 0,0002 |
| 3,47 | 0,03 | 0,0006 |
| 4,21 | 0,66 | 0,4300 |

Тогда $\delta_y$ составит:

$$\delta_y = \sqrt{\frac{0{,}5034}{7-1}} = 0{,}29$$

Согласно правилу «трех сигм», сравним подозрительное значение с утроенным значением $\delta_y$:

$$|4{,}21 - 3{,}55| < 3 * 0{,}29$$

Таким образом, можно сделать вывод о том, что седьмое наблюдение является нормальным.

Согласно критерию Райта, подозрительное значение для семи наблюдений сравнивается со значением $4 * \delta_y$:

$$|4{,}21 - 3{,}45| < 4 * 0{,}29$$

Таким образом, можно сделать вывод о том, что по критерию Райта седьмое наблюдение также признается нормальным.

**Пример 2.8**

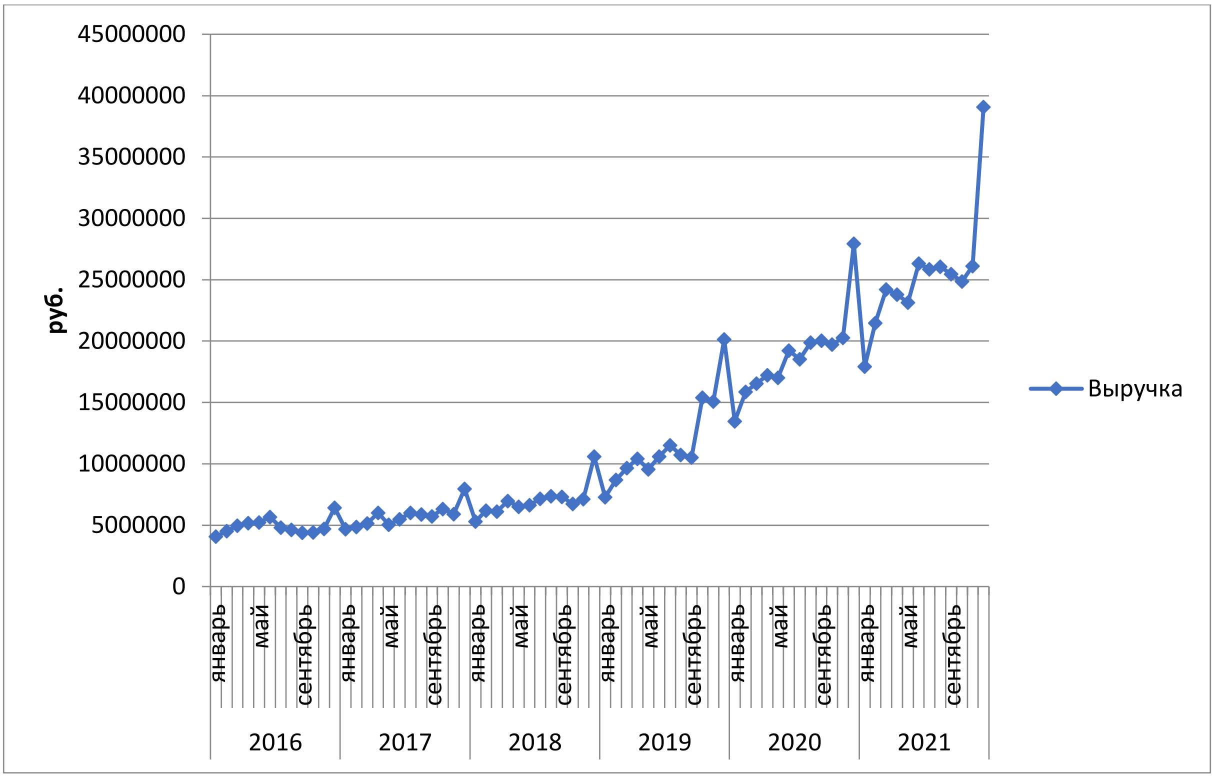
Имеются данные о выручке некоторого предприятия за 2016-2021 гг.:

| Год | Месяц | Выручка, руб. | Год | Месяц | Выручка, руб. |
|---|---|---|---|---|---|
| 2016 | январь | 4071000 | 2019 | январь | 7273000 |
| | февраль | 4519000 | | февраль | 8687000 |
| | март | 4953000 | | март | 9652000 |
| | апрель | 5169000 | | апрель | 10412000 |
| | май | 5238000 | | май | 9543000 |
| | июнь | 5677000 | | июнь | 10585000 |
| | июль | 4797000 | | июль | 11500000 |
| | август | 4637000 | | август | 10720000 |
| | сентябрь | 4397000 | | сентябрь | 10530000 |
| | октябрь | 4412000 | | октябрь | 15399000 |
| | ноябрь | 4703000 | | ноябрь | 15073000 |
| | декабрь | 5437000 | | декабрь | 20130000 |
| 2017 | январь | 4681000 | 2020 | январь | 13453000 |
| | февраль | 4866000 | | февраль | 15870000 |
| | март | 5144000 | | март | 16552000 |
| | апрель | 6010000 | | апрель | 17230000 |
| | май | 5034000 | | май | 17015000 |
| | июнь | 5489000 | | июнь | 19239000 |
| | июль | 6015000 | | июль | 18526000 |
| | август | 5866000 | | август | 19876000 |
| | сентябрь | 5713000 | | сентябрь | 20033000 |
| | октябрь | 6323000 | | октябрь | 19716000 |
| | ноябрь | 5897000 | | ноябрь | 20270000 |
| | декабрь | 7963000 | | декабрь | 27927000 |
| 2018 | январь | 5302000 | 2021 | январь | 17932000 |
| | февраль | 6200000 | | февраль | 21469000 |
| | март | 6123000 | | март | 24223000 |
| | апрель | 6978000 | | апрель | 23803000 |
| | май | 6499000 | | май | 23154000 |
| | июнь | 6632000 | | июнь | 26328000 |
| | июль | 7160000 | | июль | 25841000 |
| | август | 7360000 | | август | 26067000 |
| | сентябрь | 7317000 | | сентябрь | 25475000 |
| | октябрь | 6741000 | | октябрь | 24856000 |
| | ноябрь | 7129000 | | ноябрь | 26118000 |
| | декабрь | 10599000 | | декабрь | 39065000 |

На основании данных таблицы, проведите графический анализ временного ряда, сформулируйте выводы.

**Решение**

С помощью Ms Excel построим график зависимости выручки от времени (Вставка – Диаграммы-График):




По графику видно, что динамический ряд имеет периодические колебания, интервал между которыми составляет 12 месяцев, то есть один год. В каждом году также имеют место менее заметные колебания, повторяющиеся с определенной периодичностью. Таким образом, можно сделать вывод о том, что подобные колебания связаны с сезонностью реализации определенных товаров предприятия. Пик продаж ежегодно наблюдается в декабре.

Также по графику видно, что для представленного временного ряда амплитуда колебаний постоянно увеличивается. Таким образом, для описания данного ряда динамики, характеризующего изменение выручки предприятия, целесообразно использовать мультипликативную модель. Применение данного типа модели позволит максимально точно описать динамические изменения ряда и спрогнозировать выручку предприятия в краткосрочной перспективе.

# Глава 3. ОСНОВНЫЕ ТЕНДЕНЦИИ В УПРАВЛЕНИИ ФИРМОЙ

## 3.1. Основные критерии проверки гипотезы о существовании тренда

Начальным этапом решения задачи анализа и прогнозирования временных рядов является исследование графика изменения исследуемого показателя. Современные программные средства позволяют решить данную задачу быстро и максимально точно.

В результате графического отображения временного ряда присутствие тренда не всегда является очевидным. В таком случае для выявления тенденции и определения тренда проводят дополнительный анализ, направленный на подтверждение или опровержение утверждения о существовании тренда.

Сущность основных подходов к решению данной задачи заключается в статистической проверке гипотез [77]. Критерии выявления элементов ряда основываются на проверке гипотезы о случайности ряда, то есть (3.1):

$$H_0 : My(t) = a = \mathrm{const} \tag{3.1}$$

Наиболее часто используемые методы проверки гипотезы о наличии тренда приставлены на рис. 3.1.

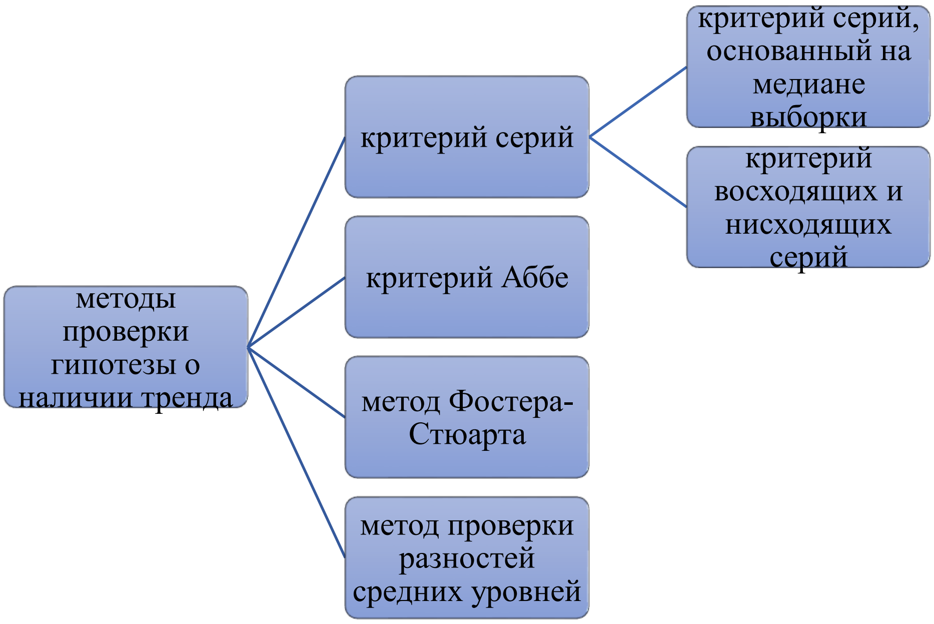



*Рис.3.1. Основные методы проверки гипотезы о наличии тренда*

**Расчет критерия серий, основанного на медиане выборки**, предполагает выполнение следующих шагов.

Предварительным этапом расчета является преобразование исходного ряда $y_1, y_2, ..., y_n$ в ранжированный по возрастанию вариационный ряд $y_1, y_2, ..., y_n$, где $y_1$ - наименьшее значение среди исходного ряда.

На первом шаге определяется медиана данного ранжированного ряда. В случае нечетного значения ряда n $(n = 2m + 1)$ медиана определяется как соответствующее серединное значение уровня , в противном случае (т.е при n=2m) медиана рассчитывается как среднее арифметическое серединных значений временного ряда [78].

На втором шаге образуется последовательность $\delta_i$ из плюсов и минусов по правилу, выраженному следующей системой (3.2):

$$\delta_i = \begin{cases} +, \text{ если } y_t > Me, \; t = 1,2,...n \\ -, \text{ если } y_t < Me, \; t = 1,2,...n \end{cases} \qquad (3.2)$$

Если же значение уровня $y_t$ равно медиане, то данное значение пропускается. Очевидно, что общее число знаков плюс и минус заранее не известно, поэтому индекс i может принимать значения от 1 до k, причем k < n.

Далее определяется число серий в совокупности $\nu(n)$. Серия представляет собой последовательность подряд идущих плюсов или минусов, даже в том случае если число членов последовательности единично, то есть имеет место только один плюс или минус.

Затем определяется последовательность самой длинной серии $\tau_{max}(n)$

Сущность проверка гипотезы с использованием данного критерия состоит в следующем: если у временного ряда отсутствует систематическая составляющая, то есть динамический ряд является случайным, протяженность самой длинной серии $\tau_{max}(n)$ не должна быть слишком большой, а общее число серий $\nu(n)$ не должно быть слишком маленьким.

Для того, чтобы гипотеза о случайности ряда не была отвергнута, необходимо выполнение следующей системы неравенств (3.3):

$$\begin{cases} \nu(n) > \left[\frac{1}{2}(n+1-1{,}96\sqrt{n-1}\right] \\ \tau_{max}(n) < [3{,}3(\lg n+1)] \end{cases} \qquad (3.3)$$

Таким образом, если хотя бы одно из неравенств нарушается, гипотеза о наличии тренда принимается с вероятностью 95%.

Квадратные скобки в системе неравенств означают целую часть числа. При этом целой частью числа А называется целое число, ближайшее к А, но не превосходящее его.

Как и метод проверки гипотезы, основанный на медиане выборки, **критерий восходящих и нисходящих серий** предполагает образование последовательности из плюсов и минусов, однако подход к образованию серий при использовании данного метода иной [79].

Вспомогательная последовательность для динамического ряда $y_1, y_2, ..., y_n$ определяется, исходя из следующих условий (3.4):

$$\delta_i = \begin{cases} +, \text{ если } y_{t+1} - y_t > 0, \ t = 1,2,...n \\ -, \text{ если } y_{t+1} - y_t < 0, \ t = 1,2,...n \end{cases} \qquad (3.4)$$

Если два соседних уровня ряда одинаковые, учитывается только одно значение временного ряда. Таким образом, элементы последовательности принимают знак «+» в том случае, если последующее значение временного ряда больше предыдущего, и «–», если предыдущее значение временного ряда больше последующего.

Как и при использовании критерия, основанного на медиане выборки, общее число знаков «+» и «–» также заранее неизвестно. Индекс i может принимать значения от 1 до k, причем $k<n-1$

Аналогично рассмотренному выше варианту расчета критерия, осуществляется подсчет общего числа серий $\nu(n)$ и протяженность самой длинной серии $\tau_{max}(n)$. Серия, состоящая из плюсов, является возрастающей (восходящей), а их минусов – убывающей (нисходящей).

При уровне значимости $a \in [0{,}05; 0{,}0975]$гипотеза об отсутствии тренда принимается в случае, если выполняется следующая система неравенств (3.5):

$$\begin{cases} \nu(n) > \left[\frac{1}{3}(2n-1) - 1{,}96\sqrt{\frac{16n-29}{90}}\right] \\ \tau_{max}(n) < \tau_0(n) \end{cases} \tag{3.5}$$

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

*Табл. 3.1. Значение $\tau_0(n)$, применяемое для расчета критерия восходящих и нисходящих серий*

| Длина ряда динамики (n) | Значение $\tau_0(n)$ |
|---|---|
| $n \le 26$ | 5 |
| $26 < n \le 153$ | 6 |
| $153 < n \le 170$ | 7 |

Если нарушается хотя бы одно из представленных условий, гипотеза о наличии тренда принимается с вероятностью 95%.

**Критерий Аббе** [80] (критерий квадратов последовательных разностей) для проверки стохастической независимости предполагает расчет величины (3.6):

$$\gamma(n) = \frac{q^2(n)}{s'^2(n)} \tag{3.6}$$

где

$$q^2(n) = \frac{1}{2(n-1)}\sum_{i=1}^{n-1}(y_{i+1} - y_i)^2,$$

$$s'^2(n) = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2,$$

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i.$$

Для того, чтобы гипотеза о стохастической независимости уровней временного ряда отвергалась, необходимо выполнение следующего условия (3.7)

$$\gamma(n) \le \gamma_\alpha^{min}(n) \tag{3.7}$$

Если число уровней ряда составляет более 60, величина $\gamma_\alpha^{min}(n)$ находится по формуле (3.8):

$$\gamma_{\alpha}^{\min}(n) = 1 + \frac{u_{\alpha}}{\sqrt{n + 0{,}5(1 + u_{\alpha}^{2})}}, \qquad (3.8)$$

где $u_{\alpha}$ - α-квантиль нормированного нормального распределения.

Величина $\gamma_{\alpha}^{\min}(n)$ для n≤60 может быть определена на основе уже существующих статистических расчетов. В табл. 3.2. приведены значения $\gamma_{\alpha}^{\min}(n)$ для различных уровней значимости при $n \in [4;20]$

*Табл. 3.2. Значение $\gamma_{\alpha}^{\min}(n)$, применяемое для расчета критерия Аббе*

| Длина ряда динамики (n) | $\gamma_{\alpha}^{\min}(n)$ | | |
|---|---|---|---|
| | α = 0,001 | α = 0,01 | α = 0,05 |
| 4 | 0,295 | 0,313 | 0,390 |
| 5 | 0,208 | 0,269 | 0,410 |
| 6 | 0,182 | 0,281 | 0,445 |
| 7 | 0,185 | 0,307 | 0,468 |
| 8 | 0,202 | 0,331 | 0,491 |
| 9 | 0,221 | 0,354 | 0,512 |
| 10 | 0,241 | 0,376 | 0,531 |
| 11 | 0,260 | 0,396 | 0,548 |
| 12 | 0,278 | 0,414 | 0,564 |
| 13 | 0,295 | 0,431 | 0,578 |
| 14 | 0,311 | 0,447 | 0,591 |
| 15 | 0,327 | 0,461 | 0,603 |
| 16 | 0,341 | 0,474 | 0,614 |
| 17 | 0,355 | 0,487 | 0,624 |
| 18 | 0,368 | 0,499 | 0,633 |
| 19 | 0,381 | 0,510 | 0,642 |
| 20 | 0,393 | 0,520 | 0,650 |

Определение наличия тренда в исходном временном ряду **методом проверки разностей средних уровней** предполагает выполнение четырёх основных этапов [81].

На первом этапе исходный временной ряд $y_1, y_2, y_3, ..., y_n$ разбивается на две примерно равные по числу уровней части: в первой части – $n_1$ первых уровней исходного ряда, во второй – $n_2$ остальных уровней. Общее число уровней n равно сумме уровней $n_1$ и $n_2$.

На втором этапе для каждой из этих частей вычисляются средние значения и дисперсии (3.9):

$$\bar{y}_1 = \frac{\sum_{t=1}^{n_1} y_t}{n_1},;$$

$$S_1^2 = \frac{\sum_{t=1}^{n_1} (y_t - \bar{y}_1)^2}{n_1 - 1}$$

$$\bar{y}_2 = \frac{\sum_{t=n_1+1}^{n} y_t}{n_2};$$

$$S_2^2 = \frac{\sum_{t=1}^{n_1} (y_t - \bar{y}_2)^2}{n_2 - 1} \qquad (3.9)$$

На третьем этапе осуществляется проверка однородности дисперсий обеих частей ряда с помощью F-критерия Фишера (3.10):

$$F = \begin{cases} S_1^2 / S_2^2, & \text{если } S_1^2 > S_2^2 \\ S_2^2 / S_1^2, & \text{если } S_1^2 < S_2^2 \end{cases} \qquad (3.10)$$

Сущность данной проверки заключается в сравнении расчетного значения этого критерия с табличным (критическим) значением критерия Фишера $F_q$ с заданным уровнем значимости q (приложение 2) . В качестве q чаще всего берут значения 0,1 (десятипроцентная ошибка), 0,05 (пятипроцентная ошибка), 0,01 (однопроцентная ошибка). Величина P = 1 – q называется доверительной вероятностью

Если расчетное значение F меньше табличного Fq, то гипотеза о равенстве дисперсий принимается и переходят к четвертому этапу. Если F больше или равно Fq, гипотеза о равенстве дисперсий отклоняется и делается вывод, что данный метод для определения наличия тренда ответа не дает.

На четвертом этапе осуществляется проверка гипотезы об отсутствии тренда с использованием t-критерия Стьюдента.

Для этого определяется расчетное значение критерия Стьюдента по формуле (3.11):

$$t = \frac{|\bar{y}_1 - \bar{y}_2|}{\bar{S} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \qquad (3.11)$$

где $\overline{S}$·– среднеквадратическое отклонение разности средних (3.12):

$$\overline{S} = \sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}} \qquad (3.12)$$

Если расчетное значение t меньше табличного значения статистики Стьюдента $t_q$ с заданным уровнем значимости q, гипотеза принимается, т.е. тренда нет, в противном случае тренд есть (приложение 3).

Число степеней свободы i для определения табличного значения $t_q$ находится по формуле (3.13):

$$i = n_1+n_2-2 \qquad (3.13)$$

Применение метода проверки разностей средних уровней возможно только для рядов динамики с монотонной тенденцией.

**Метод Фостера-Стюарта** обладает относительно широкими возможностями практического применения. Помимо тренда самого временного ряда, с помощью данного метода можно установить наличие тренда дисперсии временного ряда: если тренда дисперсии нет, то разброс уровней ряда является постоянным; если дисперсия увеличивается, то ряд "раскачивается" и т.д.

Реализация данного метода предполагает выполнение четырех основных этапов [82].

На первом этапе осуществляется сравнение каждого уровня исходного временного ряда, начиная со второго уровня, со всеми предыдущими, при этом определяются две числовые последовательности (3.14):

$$k_t = \begin{cases} 1, & \text{если } y_t \text{ больше всех предыдущих уровней} \\ 0, & \text{в противном случае} \end{cases}$$

$$l_t = \begin{cases} 1, & \text{если } y_t \text{ меньше всех предыдущих уровней} \\ 0, & \text{в противном случае} \end{cases} \qquad (3.14)$$

На втором этапе вычисляются величины S и D, характеризующие изменения среднего значения и дисперсии временного ряда (3.15 и 3.16):

$$S = \sum_{t=2}^{n} k_t + l_t \tag{3.15}$$

$$D = \sum_{t=2}^{n} k_t - l_t . \tag{3.16}$$

Величина S характеризует изменение дисперсии временного ряда и может принимать значения от 0 до n-1 (когда ряд монотонно изменяется).

Величина D характеризует изменение среднего значения временного ряда и изменяется от 1-n (когда ряд монотонно убывает) до n-1 (когда ряд монотонно возрастает). Величины S и D являются случайными с математическим ожиданием равным $\mu$ для значения S и равным 0 для значения D.

На третьем этапе осуществляется проверка гипотезы о случайности отклонения величин s и d от их математических ожиданий с помощью t-критерия Стьюдента (3.17):

$$t_S = \frac{|S - \mu|}{\sigma_1},$$

$$t_D = \frac{|D - 0|}{\sigma_2},$$

$$\sigma_1 = \sqrt{2 * \ln(n) - 3{,}4253},$$

$$\sigma_2 = \sqrt{2 * \ln(n) - 0{,}8456},$$

$$\mu = \sigma_2^2 \tag{3.17}$$

где $\mu$ - оценка математического ожидания величины S для случайного временного ряда;

$\sigma_1$ – оценка среднеквадратического отклонения S для случайного временного ряда;

$\sigma_2$ – оценка среднеквадратического отклонения D для случайного временного ряда

Формулы для нахождения $\sigma_1$ и $\sigma_2$ следует применять в случае, если число уровней временного ряда составляет более 50 элементов, в противном случае целесообразно применять готовые статистические таблицы (табл. 3.3.)

*Табл. 3.3. Значения $\sigma_1$ и $\sigma_2$ для проверки гипотезы о существовании тренда методом Фостера-Стюарта*

| n | $\sigma_1$ | $\sigma_2$ |
|---|---|---|
| 10 | 1,288 | 1,964 |
| 15 | 1,512 | 2,152 |
| 20 | 1,677 | 2,279 |
| 25 | 1,791 | 2,373 |
| 30 | 1,882 | 2,447 |
| 35 | 1,956 | 2,509 |
| 40 | 2,019 | 2,561 |
| 45 | 2,072 | 2,606 |
| 50 | 2,121 | 2,645 |

На четвертом этапе происходит сравнение найденных параметров $t_S$ и $t_D$ с табличным значением $t_q$. Если табличное значение больше расчетного, можно сделать вывод о том, что тренд отсутствует: например, если $t_S > t_q$, а $t_D < t_q$, то тренд дисперсии есть, а тренда временного ряда нет [83].

## 3.2. Основные типы тенденций и уравнений тренда

В данном параграфе рассматриваются далеко не все известные в математике линии и их уравнения, а лишь набор их сравнительно простых форм, которых достаточно для отображения и анализа большинства встречающихся на практике тенденций временных рядов. При этом желательно всегда выбирать из нескольких типов линий, достаточно близко выражающих тенденцию, более простую линию. Данный принцип обоснован тем, что чем сложнее уравнение линии тренда, чем большее число параметров оно содержит, тем при равной степени приближения труднее дать надежную оценку этих параметров по ограниченному числу уровней ряда и тем больше ошибка оценки этих параметров, ошибки прогнозируемых уровней [84].

**Прямолинейный тренд и его свойства.** Самым простым типом линии тренда является прямая линия, описываемая линейным (т.е. первой степени) уравнением тренда (3.18):

$$\hat{y}_i = a + b * t_i, \qquad (3.18)$$

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

a - свободный член уравнения, численно равный среднему выравненному уровню для момента или периода времени, принятого за начало отсчета, т.е. для $t_i = 0$;

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

$t_i$ - номера моментов или периодов времени, к которым относятся уровни временного ряда (год, квартал, месяц, дата).

Среднее изменение уровней ряда за единицу времени - главный параметр и константа прямолинейного тренда. Следовательно, этот тип тренда подходит для отображения тенденции примерно равномерных изменений уровней: равных в среднем абсолютных приростов или абсолютных сокращений уровней за равные промежутки времени. Практика показывает, что такой характер динамики встречается достаточно часто. Причина близких к равномерному абсолютных изменений уровней ряда состоит в следующем: многие явления, как, например, урожайность сельскохозяйственных культур, численность населения региона, города, сумма дохода населения, среднее потребление какого-либо продовольственного товара и др., зависят от большого числа различных факторов. Одни из них влияют в сторону ускоренного роста изучаемого явления, другие - в сторону замедленного роста, третьи - в направлении сокращения уровней и т.д. Влияние разнонаправленных и разноускоренных (замедленных) сил факторов взаимно усредняется, частично взаимно погашается, а равнодействующая их влияний приобретает характер, близкий к равномерной тенденции. Итак, равномерная тенденция динамики (или застоя) - это результат сложения влияния большого количества факторов на изменение изучаемого показателя [84].

Основные свойства тренда в форме прямой линии представлены на рис. 3.2.

- равные изменения за равные промежутки времени;
- если средний абсолютный прирост - положительная величина, то относительные приросты или темпы прироста постепенно уменьшаются;
- если среднее абсолютное изменение - отрицательная величина, то относительные изменения или темпы сокращения постепенно увеличиваются по абсолютной величине снижения к предыдущему уровню;
- если тенденция к сокращению уровней, а изучаемая величина является по определению положительной, то среднее изменение *b* не может быть больше среднего уровня *а;*
- при линейном тренде ускорение, т.е. разность абсолютных изменений за последовательные периоды, равно нулю.

*Рис. 3.2. Основные свойства прямолинейного тренда*

**Параболический тренд и его свойства**

Под названием параболического будем иметь в виду тренд, выраженный параболой II порядка с уравнением (3.19):

$$\hat{y}_i = a + b*t + c*t^2 \qquad (3.19)$$

Параболы III порядка и более высоких порядков редко применимы для выражения тенденции динамики и слишком сложны для получения надежных оценок параметров при ограниченной длине временного ряда. Прямую линию, с точки зрения математики, можно также считать одним из видов парабол - параболой I порядка, которая уже рассмотрена выше.

Значения параметров параболы II порядка таковы: свободный член а - это средний (выравненный) уровень тренда на момент или период, принятый за начало отсчета времени, т.е. t = 0; b - это средний за весь период среднегодовой прирост, который уже не является константой, а изменяется равномерно со средним ускорением, равным 2с, которое и служит константой, главным параметром параболы II порядка.

Следовательно, тренд в форме параболы II порядка применяется для отображения таких тенденций динамики, которым свойственно примерно постоянное ускорение абсолютных изменений уровней. Процессы такого рода встречаются на практике гораздо реже, чем процессы с равномерным изменением, но, с другой стороны, любое отклонение процесса от строго равномерного прироста (или сокращения) уровней можно интерпретировать как наличие ускорения. Более того, существует строгое математическое правило: чем выше порядок параболы, тем ближе линия тренда к уровням исходного временного ряда. Если это правило довести до крайнего предела, то любой ряд из n уровней может быть точно отображен параболой (n-1)-го порядка, так как через любые две точки проходит одна прямая, через три точки - одна парабола II порядка и т.д. Такое «приближение» линии тренда к эмпирическому ряду, содержащему как тенденцию, так и колебания, нельзя считать достижением научного анализа. Напротив, применяя параболу более высокого порядка там, где сущность процесса этого не требует, а только ради уменьшения остаточной суммы отклонений (или их квадратов) отдельных уровней от тренда, исследователь уходит от цели, смешивая тренд с колебаниями [85]. Парабола II порядка, как уравнение тренда, применяется к различным процессам, которые на некотором, как правило, непродолжительном, этапе развития имеют примерно постоянное ускорение абсолютного прироста уровней. Такими бывают рост населения отдельных городов или регионов, ускоренное увеличение объема продукции в фазе циклического подъема и т.д.

Основные свойства тренда в форме параболы II порядка представлены на рисунке 3.3.

неравные, но равномерно возрастающие или равномерно убывающие абсолютные изменения за равные промежутки времени;

парабола, рассматриваемая относительно ее математической формы, имеет две ветви: восходящую с увеличением уровней признака и нисходящую с их уменьшением. Но относительно статистики по содержанию изучаемого процесса изменений трендом, выражающим определенную тенденцию развития, чаще всего можно считать только одну из ветвей: либо восходящую, либо нисходящую. В особых, более конкретных, ситуациях мы не отрицаем возможности объединения обеих ветвей в единый тренд;

так как свободный член уравнения *a* как значение показателя в начальный момент (период) отсчета времени, как правило, величина положительная, то характер тренда определяется знаками параметров *b* и *c:*

при *b >0* и с > 0 имеем восходящую ветвь, т.е. тенденцию к ускоренному росту уровней;

при *b <0* и с<0 имеем нисходящую ветвь - тенденцию к ускоренному сокращению уровней;

при *b* > 0 и с<0 имеем либо восходящую ветвь с замедляющимся ростом уровней, либо обе ветви параболы, восходящую и нисходящую, если их по существу можно считать единым процессом;

при *b <0* и с>0 имеем либо нисходящую ветвь с замедляющимся сокращением уровней, либо обе ветви - нисходящую и восходящую, если их можно считать единой тенденцией;

при параболической форме тренда, в зависимости от соотношений между его параметрами, цепные темпы изменений могут либо уменьшаться, либо некоторое время возрастать, но при достаточно длительном периоде рано или поздно темпы роста обязательно начинают уменьшаться, а темпы сокращения уровней при *b <0* и с<0 обязательно начинают возрастать (по абсолютной величине относительного изменения).

*Рис. 3.3. Основные свойства тренда в форме параболы II порядка*

В тех случаях, когда по существу изучаемого процесса допустимо считать единым трендом обе ветви параболы, представляет большой интерес решение задачи о нахождении того периода или момента времени, когда уровень тренда достигает максимума (когда b>0, c<0) или минимума (если b<0, c>0). Экстремальная точка параболы $\hat{y}_i = a + b*t + c*t^2$ достигается при нулевом значении первой производной (3.20):

$$\frac{df}{dt} = (a + bt + ct^2)' = b + 2ct \tag{3.20}$$

Из b+2ct=0 равенства имеем: t=-b/2c. Например, если $\hat{y}_i = 100 + 20t - 2t^2$, то максимум парабола имеет при t=-5.

Максимальное значение уровня тренда при t=5 составит 150.

Если имеем параболу при b<0, а c>0, например, $\hat{y}_i = 100 - 20t + 2t^2$, то минимальное значение тренда достигается при t=5, и это минимальное значение составит 50.

**Экспоненциальный тренд и его свойства**

Экспоненциальным трендом называют тренд, выраженный уравнением (3.21):

$$\hat{y}_i = ak^{t_i} \tag{3.21}$$

или формулой (3.22):

$$\hat{y}_i = \exp[\ln a + \ln k t_i] \tag{3.22}$$

Свободный член экспоненты а равен выравненному уровню, т.е. уровню тренда в момент или период, принятый за начало отсчета времени, т.е. при t=0. Основной параметр экспоненциального тренда k является постоянным темпом изменения уровней (ценным). Если k>1, имеем тренд с возрастающими уровнями, причем это возрастание не просто ускоренное, а с возрастающим ускорением и возрастающими производными всех более высоких порядков. Если k<1, то имеем тренд, выражающий тенденцию постоянного, но замедляющегося сокращения уровней, причем замедление непрерывно усиливается. Экстремума экспонента не имеет и при $t \to \infty$ стремится либо к $\infty$ при k > 1, либо к 0 при k<1.

Экспоненциальный тренд характерен для процессов, развивающихся в среде, не создающей никаких ограничений для роста уровня. Из этого следует, что на практике он может развиваться только на ограниченном промежутке времени, так как любая среда рано или поздно создает ограничения, любые ресурсы со временем исчерпаемы. Однако практика показала что, например, численность населения Земли на протяжении 1950-1985 гг. возрастала примерно по экспоненте со среднегодовым темпом роста $k \approx 1{,}018$ и за это время возросла вдвое - с 2,5 до 5 млрд. чел. В настоящее время темп роста населения постепенно уменьшается.

Экспоненциальный рост объема реализации и производства происходит при возникновении новых видов продукции и их освоении промышленностью: при появлении цветных телевизоров, видеомагнитофонов, пейджеров и т.п., но когда производство начинает наполнять рынок, приближаться к спросу, экспоненциальный рост прекращается.

Основные свойства экспоненциального тренда представлены на рис. 3.4.

Абсолютные изменения уровней тренда пропорциональны самим уровням.

Экспонента экстремумов не имеет: при $k > 1$ тренд стремится к $+\infty$, при $k<1$ тренд стремится к нулю.

Уровни тренда представляют собой геометрическую прогрессию: уровень периода с номером $t = m$ есть $ak^m$

При $k > 1$ тренд отражает ускоряющийся неравномерно рост уровней, при $k < 1$ тренд отражает замедляющееся неравномерно уменьшение уровней.

*Рис. 3.4. Основные свойства экспоненциального тренда*

**Гиперболический тренд и его свойства**

Из различных форм гипербол рассмотрим только наиболее простую (3.23):

$$\hat{y}_i = a + \frac{b}{t} \qquad (3.23)$$

Если основной параметр гиперболы b>0, то этот тренд выражает тенденцию замедляющегося снижения уровней и при t→∞, ŷ→a. Таким образом, свободный член гиперболы - это предел, к которому стремится уровень тренда.

Такая тенденция наблюдается, например, при изучении процесса снижения затрат любого ресурса (труда, материалов, энергии) на единицу данного вида продукции или ее себестоимости в целом. Затраты ресурса не могут стремиться к нулю, значит, экспонента не соответствует сущности процесса; нужно применить гиперболическую формулу тренда.

Если параметр b<0, то с возрастанием t, т.е. с течением времени, уровни тренда возрастают и стремятся к величине a при t→∞ [86].

Такой характер динамики присущ, например, показателям КПД двигателей или иных преобразователей энергии (трансформатор тока, фотоэлемент и т.п.). По мере развития научно-технического прогресса эти КПД постепенно повышаются, но никогда не могут превысить определенного предела для каждого типа двигателя и не могут превысить 100% в принципе для любого преобразователя энергии. При расчете гиперболического тренда нельзя нумеровать года от середины ряда, так как значения 1/ti должны быть всегда положительными. Основные свойства гиперболического тренда представлены на рис. 3.5.

Абсолютный прирост или сокращение уровней, ускорение абсолютных изменений, темп изменения - все эти показатели не являются постоянными. При *b>0* уровни замедленно уменьшаются, отрицательные абсолютные изменения, а также положительные ускорения тоже уменьшаются, цепные темпы изменения растут и стремятся к 100%.

При *b<0* уровни замедленно возрастают, положительные абсолютные изменения, а также отрицательные ускорения и цепные темпы роста замедленно уменьшаются, стремясь к 100%.

*Рис. 3.5. Основные свойства гиперболического тренда*

Таким образом, гиперболический тренд описывает тенденцию такого процесса, показатели которого со временем затухают, т.е. происходит переход от движения к застою.

**Логарифмический тренд и его свойства**

Если изучаемый процесс приводит к замедлению роста какого-то показателя, но при этом рост не прекращается, не стремится к какому-либо ограниченному пределу, то гиперболическая форма тренда уже не подходит. Тем более не подходит парабола с отрицательным ускорением, по которой замедляющийся рост перейдет со временем в снижение уровней. В указанном случае тенденция изменения лучше всего отображается логарифмической формой тренда (3.24):

$$\hat{y}_i = a + b \ln t_i \tag{3.24}$$

Логарифмы возрастают значительно медленнее, чем сами числа (номера периодов ti), но рост логарифмов неограничен. Подбирая начало отсчета периодов (моментов) времени, можно найти такую скорость снижения абсолютных изменений, которая наилучшим образом отвечает фактическому временному ряду.

Примером тенденций, соответствующих логарифмическому тренду, может служить динамика на отдельных этапах анализа урожайности или валового сбора какой-то культуры в данном регионе, пока новое агротехническое достижение не придаст снова тенденции ускорения. Основные свойства логарифмического тренда представлены на рис. 3.6.

Если b>0, то уровни возрастают, но с замедлением, а если b<0, то уровни тренда уменьшаются, тоже с замедлением.

Абсолютные изменения уровней по модулю всегда уменьшаются со временем.

Ускорения абсолютных изменений имеют знак, противоположный самим абсолютным изменениям, а по модулю постепенно уменьшаются.

Темпы изменения (цепные) постепенно приближаются к 100% при $t \to \infty$

*Рис. 3.6. Основные свойства логарифмического тренда*

Логарифмический тренд, как и гиперболический, отражает постепенно затухающий процесс изменений. Различие состоит в том, что затухание по гиперболе происходит быстро при приближении к конечному пределу, а при логарифмическом тренде затухающий процесс продолжается без ограничения гораздо медленнее.

**Логистический тренд и его свойства**

Логистическая форма тренда подходит для описания такого процесса, при котором изучаемый показатель проходит полный цикл развития. Начиная, как правило, от нулевого уровня, значения уровней растут сначала медленно, затем темп роста ускоряется, в середине цикла ускорение становится нулевым, т.е. рост происходит по линейному тренду, затем, в завершающей части цикла, рост замедляется по гиперболе по мере приближения к предельному значению показателя.

Примером такого цикла динамики может служить изменение доли грамотного населения в стране, например в России, с 1800 г. до наших дней. В некоторых зарубежных программах для компьютеров логистическая кривая называется S-образной кривой.

С одной стороны, логистическую тенденцию можно считать объединением трех разных по типу тенденций: параболической с ускоряющимся ростом на первом этапе, линейной - на втором и гиперболической с замедляющимся ростом - на третьем этапе. Но есть доводы и в пользу рассмотрения всего цикла развития как особого единого типа тенденции со сложными, переменными свойствами, но постоянным направлением изменений в сторону увеличения или уменьшения уровней.

Рассмотрение таких временных рядов, как проявление единой логистической тенденции, позволяет уже на первом этапе рассчитать всю траекторию развития, определить сроки перехода от ускоренного роста к замедленному, что чрезвычайно важно при планировании производства или реализации нового вида товара, спрос на который будет проходить все этапы логистической тенденции вплоть до насыщения рынка.

Так, например, обеспеченность населения в России автомобилями в конце 1980-х годов находилась на начальном этапе логистической кривой, и это означало, что предстоит еще ряд лет или даже десятилетий ускоренного роста спроса. В то же время обеспеченность фотоаппаратами уже достигла этапа замедления роста, и это означало, что расширять производство или импорт прежних типов фотоаппаратов не следует. Расширение их рынка возможно было только для принципиально новых типов фотоаппаратов, насыщенность которыми еще находится в самом начале первого этапа [87].

В вышеописанном диапазоне изменения уровней, т.е. от нуля до единицы, уравнение логистического тренда имеет вид (3.25):

$$\hat{y}_i = \frac{1}{e^{a_0 + a_1 t_i} + 1} \tag{3.25}$$

При $a_0>0$, $a_1<0$, с ростом номеров периодов времени $t_i$ получаем логистическую тенденцию роста уровней, причем если нужно начать рост почти от нулевой величины, то $a_0$ должно быть примерно равно 10. Чем больше модуль $a_1$, тем быстрее возрастание уровней. При $a_0<0$ и $a_1>0$ имеем логистический тренд со снижением уровней, причем если снижение должно начаться почти от единицы, то $a_0$ должно быть примерно равно -10. Чем больше $a_1$, тем быстрее будут снижаться уровни.

Если же диапазон изменения уровней ограничен не нулем и единицей, а любыми значениями, определяемыми исходя из существа задачи, обозначаемыми $y_{max}$ и $y_{min}$, то формула логистического тренда принимает вид (3.26):

$$\hat{y}_i = \frac{y_{max} - y_{min}}{e^{a_0 + a_1 t_i} + 1} + y_{min} \tag{3.26}$$

## 3.3. Определение типа тренда в развитии фирмы и оценка его параметров

При изучении методов распознавания типа тренда не следует забывать о существе изучаемого процесса, который отображается временным рядом. Как правило, тип тренда должен соответствовать характерным особенностям процесса. На практике зачастую используется графический анализ для распознавания типа тенденции.

Графическое изображение во многих случаях позволяет приближенно выявить тип тенденции временного ряда. Но для этого следует соблюдать правила построения графика, сущность которых представлена на рис. 3.7.

Точное соблюдение масштаба как по величине уровней ряда, так и по времени

Временные интервалы откладывают по оси абсцисс, величины уровней - по оси ординат.

По каждой оси следует установить такой масштаб, чтобы ширина графика была примерно в 1,5 раза больше его высоты.

Если уровни ряда на всем протяжении периода много больше нуля и между собой различаются не более чем на 20-30%, то следует обозначить перерыв на оси ординат, увеличить масштаб так, чтобы меньший из уровней ненамного превышал разрыв оси.

Если уровни ряда различаются в десятки, сотни и тысячи раз, ось ординат следует разметить в логарифмическом масштабе, чтобы равные отрезки означали различие уровней в одинаковое число раз. Интерпретация вида графика будет другой: при линейном масштабе график, близкий к прямой линии, означает линейную тенденцию, а при логарифмическом масштабе оси ординат прямая линия показывает экспоненциальную тенденцию.

Необходимо строго соблюдать равенство промежутков времени на равных отрезках оси абсцисс. Логарифмический масштаб по времени не рекомендуется, так как он крайне затруднит интерпретацию графика.

*Рис. 3.7. Основные правила построения графика для корректного распознавания типа тенденции*

Следует отметить, что не всегда график позволяет выбрать тип линии тренда. Трудно графически отличить параболу от экспоненты, логарифмическую кривую от гиперболы и т.д. Оценка типа тренда по типу графика включает субъективные моменты, что может привести к ошибке. Есть много способов объективной, статистико-математической оценки пригодности того или иного типа линии. Весьма популярен его выбор с помощью перебора на персональном компьютере всех имеющихся в пакете программ статистического анализа типов линий либо по наименьшему среднему квадратическому отклонению, либо по наименьшему модулю отклонений фактических уровней от расчетных по проверяемой линии. Недостатки данной методики заключаются в том, что, во-первых, не все пакеты программ статистического анализа содержат достаточный выбор линий тренда, но главное состоит в том, что, чем больше параметров содержит уравнение тренда, тем меньше и отклонений отдельных уровней от тренда. Парабола II порядка, а тем более III и более высоких порядков всегда при таком подходе «лучше», чем прямая или экспонента [88].

Но «преимущество» параболы над прямой может быть невелико. Следовательно, нужно применить опять же статистико-математические критерии существенности уменьшения среднего отклонения при переходе от прямой к параболе.

Предположим, что предварительная гипотеза о типе тренда выбрана на основе теоретических соображений об изучаемом процессе и на основе графического изображения. Для того чтобы проверить данную гипотезу, необходимо сформулировать ее математически. Так, гипотеза о том, что тренд является прямой линией, означает, что на всем периоде временной ряд в среднем сохраняет постоянную величину абсолютного изменения уровней. Гипотеза о параболе II порядка означает, что на всем периоде (в среднем) имеется постоянная величина ускорения абсолютных изменений. Гипотеза об экспоненциальном тренде подтвердится, если можно будет доказать, что на периоде сохраняется постоянная величина (в среднем) цепного темпа изменений.

Для указанных трех типов линий предлагается методика статистической проверки гипотез, разработанная М.С. Каяйкиной и А.И. Манеллей [89, 90] (рис. 3.8).


*Рис. 3.8. Методика проверки статистических гипотез о типе тренда*

Таким образом, последовательная реализация представленных на рис. 3.8 этапов позволяет сформулировать вывод о типе линии тренда [91]. Однако, это не является конечным результатом работы с обнаружением типа тенденции.

## 3.4. Метод наименьших квадратов как инструмент оценки параметров тренда

Основой методики проверки статистических гипотез о типе тренда является метод наименьших квадратов, который дает оценки параметров, отвечающие принципу максимального правдоподобия. Суть данного принципа состоит в том, что сумма квадратов отклонений фактических уровней от тренда (от выравненных по уравнению тренда уровней) должна быть минимальной для данного типа уравнения [92].

Данная методика близка к методике корреляционно-регрессионного анализа связей - парной регрессии. Однако между ними есть и принципиальные различия: выступающий при расчете уравнения тренда в качестве независимой переменной ряд номеров периодов или моментов времени не является случайной варьирующей переменной регрессионного анализа. Ряд значений времени – это жестко упорядоченный ряд величин, и, следовательно, не может быть речи о корреляции между ним и значениями зависимой переменной - варьирующих уровней показателя, изменяющегося во времени. Нередко применяемые в литературе и в программах для компьютеров коэффициенты корреляции со временем или фактических уровней с выравненными (т.е. тоже упорядоченными) уровнями тренда таковыми на самом деле не являются и не могут измерять какой-либо «тесноты связи». Чем длиннее период, охватываемый рядом, тем автоматически становятся больше так называемые коэффициенты корреляции при той же самой скорости роста уровней и той же самой силе колебаний. Таким образом, эти лжекоэффициенты не могут характеризовать соотношение между ролью факторов тенденции и ролью факторов колеблемости.

Для уравнения **прямой линии тренда** величина параметров a и b определяется по методу наименьших квадратов путем приравнивания частных первых производных функции $f(a,b) = \sum_{i=1}^{n} y_i - \hat{y}_i = \sum_{i=1}^{n} (y_i - a - bt_i)^2$ к нулю. Тогда система нормальный уравнений имеет вид (3.26):

$$\begin{cases} na + b\sum_{i=1}^{n} t_i = \sum_{i=1}^{n} y_i \\ a\sum_{i=1}^{n} t_i + b\sum_{i=1}^{n} t_i^2 = \sum_{i=1}^{n} y_i t_i \end{cases} \tag{3.26}$$

Решая эти уравнения с двумя неизвестными по данным фактического временного ряда $y_i$, получаем значения a и b. Если номера периодов (моментов) времени отсчитываются от начала ряда так, что первый период (момент) обозначен номером t=1, то свободный член a есть уровень тренда для предыдущего периода (момента), а не первого в ряду, как часто ошибочно полагают. Для первого периода уровень тренда $\hat{y}$ равен a+b, для второго $\hat{y}$=a+2b и т.д.

Однако рациональнее начало отсчета времени перенести в середину ряда, т.е. при нечетном n - на период (момент) с номером (n+1)/2, а при четном числе уровней ряда - на середину между периодом с номером n/2 и (n/2)+1. В последнем случае все номера периодов ti будут дробными. При нумерации периодов времени точно от середины ряда половина номеров ti будет отрицательными числами (аналогично годам до нашей эры), а половина - положительными, их сумма равна нулю. В таком случае система нормальных уравнений МНК распадается на два уравнения с одним неизвестным в каждом (3.27 и 3.28):

$$na = \sum_{i=1}^{n} y_i \tag{3.27}$$

$$b\sum_{i=1}^{n} t_i^{\,2} = \sum_{i=1}^{n} y_i t_i \tag{3.28}$$

Откуда имеем (3.29 и 3.30):

$$a = \frac{\sum_{i=1}^{n} y_i}{n} = \bar{y} \tag{3.29}$$

$$b = \frac{\sum_{i=1}^{n} y_i t_i}{\sum_{i=1}^{n} t_i^{\,2}} \tag{3.30}$$

Зачастую многие компьютерные программы не предусматривают такого упрощения, и нумерация периодов (моментов) в них производится с начала ряда, с номера t=1, причем пользователь об этом не предупреждается [93]. При расчетах без компьютера следует применить упрощенный прием. Знаменатель в формуле (3.30) при нумерации периодов от середины ряда вычисляется устно при n≤10 или по формуле (3.31):

$$\sum_{i=-\frac{n+1}{2}}^{+\frac{n+1}{2}} t_i^{\,2} = \frac{n^3-3}{12} \tag{3.31}$$

**Уравнение параболического (II порядка) тренда**

Для вычисления параметров a, b, c по методу наименьших квадратов три частные производные функции $f(a,b,c)=\sum_{i=1}^{n}(y_i-\hat{y})^2$ приравниваются к нулю, и после преобразований получаем систему трех уравнений с тремя неизвестными (3.32):

$$\begin{cases} na + b\sum_{i=1}^{n} t_i + c\sum_{i=1}^{n} t_i^2 = \sum_{i=1}^{n} y_i \\ a\sum_{i=1}^{n} t_i + b\sum_{i=1}^{n} t_i^{\,2} + c\sum_{i=1}^{n} t_i^{\,3} = \sum_{i=1}^{n} y_i t_i \\ a\sum_{i=1}^{n} t_i^{\,2} + b\sum_{i=1}^{n} t_i^{\,3} + c\sum_{i=1}^{n} t_i^{\,4} = \sum_{i=1}^{n} y_i t_i^{\,2} \end{cases} \tag{3.32}$$

При переносе начала отсчета периодов (моментов) времени в середину ряда суммы нечетных степеней номеров этих периодов $\sum t_i$ и $\sum t_i^{\,3}$ обращаются в нуль. При этом второе уравнение обращается в уравнение с одним неизвестным, откуда (3.33):

$$b = \frac{\sum_{i=1}^{n} y_i t_i}{\sum_{i=1}^{n} t_i^2} \tag{3.33}$$

Образуется система из двух уравнений с двумя неизвестными (3.34):

$$\begin{cases} na + c\sum_{i=1}^{n} t_i^2 = \sum_{i=1}^{n} y_i \\ a\sum_{i=1}^{n} t_i^{\,2} + c\sum_{i=1}^{n} t_i^{\,4} = \sum_{i=1}^{n} y_i t_i^{\,2} \end{cases} \tag{3.34}$$

В данной системе уравнений (3.35 и 3.36):

$$\sum_{i=1}^{n} t_i^2 = \frac{n^3 - n}{12} \quad (3.35)$$

$$\sum_{i=1}^{n} t_i^4 = \frac{3n^5 - 10n^3 + 7n}{240} \quad (3.36)$$

**Гиперболическое уравнение тренда** отличается от линейного уравнения тем, что вместо первой степени включает номера периодов времени (моментов) в минус первой степени: $1/t_i$. Соответственно нормальные уравнения метода наименьших квадратов получат вид (3.37):

$$\begin{cases} na + b\sum_{i=1}^{n}\frac{1}{t_i} = \sum_{i=1}^{n} y_i \\ a\sum_{i=1}^{n}\frac{1}{t_i} + b\sum_{i=1}^{n}\frac{1}{t_i^2} = \sum_{i=1}^{n}\frac{y_i}{t_i} \end{cases} \quad (3.27)$$

Однако при этом нельзя, в отличие от линейного тренда, переносить начало отсчета периодов времени в середину, так как гипербола не имеет постоянного параметра изменения уровней на протяжении всего периода, и все величины $1/t_i$ должны быть положительными.

## 3.5. Логарифмирование как инструмент оценки параметров экспоненциального, логарифмического и логистического уравнений тренда

Данные типы трендов объединены в одну группу в связи с необходимостью при оценке их параметров прибегать к логарифмированию. При расчете логарифмического уравнения тренда логарифмируют номера периодов (моментов) времени, а при расчете параметров экспоненциального и логистического трендов - сами уровни. Поскольку отрицательные числа не имеют действительных логарифмов, если нужно логарифмировать номера периодов времени, то нельзя переносить начало их отсчета в середину ряда [94]. Если же сами уровни могут принимать отрицательные значения, например, уровни финансового результата от реализации, уровни температуры воздуха или почвы, то необходимо перенести начало отсчета уровней на величину, алгебраически меньшую реального наименьшего уровня. По окончании расчета тренда нетрудно восстановить обычные единицы измерения.

Для экспоненциального уравнения тренда формула уравнения имеет вид (3.28):

$$\hat{y}_i = ak^{t_i} \tag{3.28}$$

Для нахождения параметров *a* и *k* уравнение логарифмируем (3.29):

$$\ln \hat{y}_i = \ln a + t_i \ln k \tag{3.29}$$

В такой форме, т.е. для логарифмов, уравнение соответствует линейному, следовательно, метод наименьших квадратов дает для логарифмов *a* и *k* нормальные уравнения, аналогичные таковым для параметров *a* и *b* линейного тренда (3.30):

$$\begin{cases} n \ln a + \ln k \sum_{i=1}^{n} t_i = \sum_{i=1}^{n} \ln y_i \\ \ln a \sum_{i=1}^{n} t_i + \ln k \sum_{i=1}^{n} t_i^2 = \sum_{i=1}^{n} t_i \ln y_i \end{cases} \tag{3.30}$$

Так как номера периодов времени не логарифмируются, можно перенести начало отсчета в середину ряда и упростить систему (3.31):

$$\begin{cases} n \ln a = \sum_{i=1}^{n} \ln y_i \\ \ln k \sum_{i=1}^{n} t_i^2 = \sum_{i=1}^{n} t_i \ln y_i \end{cases} \tag{3.31}$$

Отсюда получаем систему (3.32):

$$\begin{cases} \ln a = \dfrac{\sum_{i=1}^{n} \ln y_i}{n} \\ \ln k = \dfrac{\sum_{i=1}^{n} t_i \ln y_i}{\sum_{i=1}^{n} t_i^2} \end{cases} \tag{3.32}$$

Особенность **логарифмического уравнения тренда** заключается в том, что логарифмировать необходимо номера периодов (моментов) времени. Следовательно, все номера должны быть положительными числами. Однако это вовсе не означает, что нумерацию следует начинать с числа 1. Дело в том, что величина логарифма быстро возрастает при переходе от единицы к двум: натуральный логарифм единицы равен нулю, а логарифм двух равен 0,693, имеем рост на 0,693; в то же время логарифм четырех равен 1,386, а логарифм пяти равен 1,609, имеем прирост лишь на 0,223 и т.д. Если и уровень изучаемого ряда вначале возрастает втрое быстрее, чем между четвертым и пятым периодом, тогда нумерация от единицы допустима. Если же уменьшение прироста уровней происходит значительно медленнее, нумерацию периодов (моментов) следует начинать не с единицы, а с большего числа.

**Логистическое уравнение тренда** имеет вид (3.33):

$$\hat{y}_i = \frac{y_{max} - y_{min}}{e^{a_0 + a_1 t_i} + 1} + y_{min} \qquad (3.33)$$

При расчете этого уравнения логарифмируют величину, производную от уровней ряда, но не номера периодов (моментов) времени, эту нумерацию поэтому рациональнее проводить от середины ряда.

Особенностью логистического тренда является этап обоснования значений максимального и минимального уровней временного ряда.

Это обоснование осуществляется на основе, во-первых, уровней фактического ряда, во-вторых, теоретических, т.е. внешних по отношению к статистике, соображений. относящихся к содержанию изучаемого процесса.

Уравнение логистического тренда в общем виде непосредственно логарифмировать невозможно [95]. Для этого необходимо преобразовать его в форму (3.34):

$$\frac{\hat{y}_{max} - \hat{y}_{min}}{\hat{y}_i - y_{min}} - 1 = e^{a_0 + a_1 t} \qquad (3.34)$$

Произведем замену переменной (3.35):

$$\frac{\hat{y}_{max} - \hat{y}_{min}}{\hat{y}_i - y_{min}} - 1 = \hat{\xi}, \hat{\xi} = e^{a_0 + a_1 t} \qquad (3.35)$$

Тогда получим (3.36):

$$\ln \hat{\xi} = a_0 + a_1 t_i \qquad (3.36)$$

Условие метода наименьших квадратов (3.37):

$$\sum_{i=1}^{n}(\ln\xi_i - \ln\hat{\xi}_i)^2 \to \min \quad (3.37)$$

Тогда получим (3.38):

$$\sum_{i=1}^{n}(\ln\xi_i - a_0 - a_1 t_i)^2 \to \min \quad (3.38)$$

После вычисления частных производных по $a_0$ и по $a_1$, получаем нормальные уравнения МНК для логистической кривой, аналогичные таковым для прямой линии, так как заменой на ζ фактически проведена линеаризация функции логистической кривой (3.39):

$$\begin{cases} na_0 + a_1\sum_{i=1}^{n} t_i = \sum_{i=1}^{n}\ln\xi_i \\ a_0\sum_{i=1}^{n} t_i + a_1\sum_{i=1}^{n} t_i^2 = \sum_{i=1}^{n} t_i \ln\xi_i \end{cases} \quad (3.39)$$

При переносе начала отсчета периодов (моментов) времени в середину ряда система упрощается до двух уравнений с одним неизвестным в каждом из них (3.40):

$$\begin{cases} na_0 = \sum_{i=1}^{n}\ln\xi_i \\ a_1\sum_{i=1}^{n} t_i^2 = \sum_{i=1}^{n} t_i \ln\xi_i \end{cases} \quad (3.40)$$

Таким образом, коэффициенты могут быть найдены (3.41):

$$\begin{cases} a_0 = \dfrac{\sum_{i=1}^{n}\ln\xi_i}{n} \\ a_1 = \dfrac{\sum_{i=1}^{n} t_i \ln\xi_i}{\sum_{i=1}^{n} t_i^2} \end{cases} \quad (3.41)$$

## 3.6. Сглаживание рядов, характеризующих динамику изменения показателей деятельности фирмы

Одной из важнейших задач исследования динамических рядов является определение общей тенденции развития.

Тенденция представляет собой направление развития явления или процесса во времени под влиянием определённых факторов.

Для того, чтобы выявить общую тенденцию изменения процесса или явления во времени, сократив при этом колебания уровней ряда, применяются различные методы сглаживания динамических рядов. К числу простейших можно отнести **метод сглаживания с использованием скользящей средней** и **метод укрупнения интервалов**.

Сущность всех методов сглаживания состоит в замене исходных уровней ряда расчетными параметрами. По сравнению с фактическими значениями ряда они обладают значительно меньшей колеблемостью, что позволяет четче проследить общую тенденцию и сформулировать общие выводы.

**Метод укрупнения интервалов**

Сущность данного метода сосотоит в переходе от менее крупных интервалов к более крупным: от месячных - к квартальным, от квартальных - к полугодовым, от полугодовых – к годовым и т.д.

Нахождение уровней укрупненных рядов осуществляют путем поиска среднего значения уровня или суммированием значений за определенный промежуток времени

Стоит отметить, что метод укрупнения интервалов, является самым простым, но вместе с тем и наименее точным.

Применение данного метода возможно только для интервальных рядов, для моментных временных рядов используется метод скользящей средней [96].

При **применении метода простой скользящей средней** новый динамический ряд строится из простых средних арифметических. При этом сначала для временного ряда определяется интервал сглаживания m.

Для вычисления сглаженных уровней временного ряда $\overline{y_i}$ применяется следующая формула (3.42):

$$\overline{y_i} = \frac{\sum_{t=1}^{m} y_t}{m} \qquad (3.42)$$

После вычисления значений средней для m первых уровней $y_1, y_2 ... y_m$ переходят к вычислению последующих $y_2, y_3 ... y_{m+1}$ и т.д. Таким образом, интервал как бы «скользит» по исходному динамическому ряду с шагом, равным единице.

Если число уровней, взятых для расчета средней, нечетное, средняя записывается в уровень находящийся по середине. Если число уровней четное, то средняя будет относиться к промежутку между серединными значениями. Для ликвидации этого сдвига применяют способ центрирования.

Центрирование представляет собой процесс определения средней из средних для отнесения полученного уровня к конкретному уровню.

Существенным недостатком метода скользящей средней является «укорачивание» ряда полученных значений с каждого конца по сравнению с исходным на m/2 при четном интервале сглаживания и на (m-1)/2 для нечетного интервала.

Для восстановления потерянных значений временного ряда нужно вычислить прирост на последнем участке скольжения, а затем восстановить значения ряда динамики путем последовательного прибавления среднего абсолютного прироста к последнему сглаженному значению. Аналогичным образом восстанавливаются первые члены ряда.

Если для процесса характерно нелинейное развитие, то более надежным способом является использование взвешенной скользящей средней, которое осуществляется по полиномам 2-го и 3-го порядков. В некоторых случаях использование простой скользящей средней является неинформативным, поскольку сглаживание данным способом приводит к проявлению тенденции только в самом общем виде, исключая при этом важные для экономического анализа относительно мелкие изгибы линий [97].

При применении взвешенной скользящей средней каждому уровню в пределах интервала сглаживания приписывается определенный вес. Его величина зависит от расстояния, рассчитанного от данного уровня до середины интервала сглаживания

Для определения значений взвешенных скользящих средних внутри каждого периода уровни описываются полиномом в степени p (3.43):

$$\overline{y_t} = \sum_{j=0}^{p} a_j t^j \qquad (3.43)$$

Каждому значению уровня ряда соответствует свой весовой коэффициент. Все весовые коэффициенты характеризуются свойствами:

1. Они симметричны относительно центрального уровня.
2. Сумма всех весов с учетом общего множителя равна единице.
3. Наличие положительных и отрицательных весов позволяет сглаженной кривой сохранять различные изгибы кривой тренда [98].

Весовые коэффициенты сглаживания ряда динамики по полиномам 2-го и 3-го порядка для различных активных участков сглаживания приведены в табл. 3.4.

*Табл. 3.4. Весовые коэффициенты, применяемые в методе взвешенной скользящей средней*

| Длина активного участка | Весовые коэффициенты |
|---|---|
| 5 | $\frac{1}{35}\left[-3;+12;+17;+12;-3\right]$ |
| 7 | $\frac{1}{21}\left[-2;+3;+6+7;+6;+3;-2\right]$ |
| 9 | $\frac{1}{231}\left[-21;+14;+39;+54;+59;+54;+39;+14;-21\right]$ |

Таким образом, в рамках данного параграфа были рассмотрены простейшие методы сглаживания временных рядов. Выбор способа сглаживания зависит от специфики рассматриваемого ряда, его вида, а также от конкретных исследовательских задач.

## Контрольные вопросы

1. Что такое тенденция?
2. Каковы основные критерии проверки гипотезы о существовании тренда
3. В чем состоит сущность критерия Аббе?
4. Реализацию каких этапов предполагает т расчет критерия серий, основанного на медиане выборки?
5. Каковы возможности применения Фостера – Стюарта?
6. Что является самым простым типом линии тренда?
7. Каковы свойства прямолинейного тренда?

8. Как можно охарактеризовать параболу II порядка при $b<0$ и $c>0$?

9. Для каких экономических процессов характерен экспоненциальный рост?

10. Как с помощью графического анализа можно распознать тип тренда?

11. В чем состоит методика статистической проверки гипотез, разработанная М.С. Каяйкиной и А.И. Манеллей?

12. Что является основой методики проверки статистических гипотез о типе тренда?

13. Что объединяет оценку параметров экспоненциального, логарифмического и логистического уравнений тренда?

14. В чем состоит суть метода укрупнения интервалов?

15. В каких случаях применяется процедура центрирования? В чем состоит ее суть?

16. В каких случаях целесообразно использование взвешенной скользящей средней?

## ПРИМЕРЫ РЕШЕНИЯ ЗАДАЧ

**Пример 3.1.**

В таблице приведен временной ряд:

| t | Значение временного ряда $y_i$ |
|---|---|
| 1 | 174 |
| 2 | 168 |
| 3 | 164 |
| 4 | 167 |
| 5 | 171 |
| 6 | 178 |
| 7 | 184 |
| 8 | 186 |
| 9 | 194 |
| 10 | 201 |

С помощью критерия серия, основанного на медиане выборки, проверить гипотезу о наличии тренда

**Решение.**

На первом этапе решения исходный ряд данных был проранжирован по возрастанию:

| t* | Значение временного ряда $y_i*$ |
|---|---|
| 1 | 164 |
| 2 | 167 |
| 3 | 168 |
| 4 | 171 |
| 5 | 174 |
| 6 | 178 |
| 7 | 184 |
| 8 | 186 |
| 9 | 194 |
| 10 | 201 |

Затем была определена медиана данного вариационного ряда Ме. Поскольку число звеньев четное, медиана определяется как средняя из двух центральных значений:

$$\text{Me} = \frac{174+178}{2} = 176$$

В исходной выборке вместо каждого значения уровня ряда будем ставить «+», если значение больше медианы, и «-», если значение ряда меньше медианы. В случае, если значение уровня ряда равно медиане, знак проставляться не будет:

| Значение временного ряда $y_i$ | Серии |
|---|---|
| 174 | - |
| 168 | - |
| 164 | - |
| 167 | - |
| 171 | - |
| 178 | + |
| 184 | + |
| 186 | + |
| 194 | + |
| 201 | + |

Таким образом, число серий равно 2, самая длинная последовательность серий равна 5.

Проанализируем, с учетом полученных данных, систему неравенств

$$\begin{cases} 2 > \left[\dfrac{1}{2}(10+1-1{,}96\sqrt{10-1}\right] \\ 5 < \left[3{,}3(\lg n+1)\right] \end{cases}$$

Поскольку первое неравенство неверно, гипотеза об отсутствии тренда отвергается. Стоит отметить, что критерий серий, основанный на медиане, улавливает только монотонное изменение среднего (оценки математического ожидания).

**Пример 3.2.**

В таблице приведен временной ряд:

| t | Значение временного ряда $y_i$ |
|---|---|
| 1 | 200 |
| 2 | 194 |
| 3 | 190 |
| 4 | 193 |
| 5 | 197 |
| 6 | 204 |
| 7 | 210 |
| 8 | 193 |

С помощью критерия восходящих и нисходящих серий проверить гипотезу о наличии тренда.

**Решение.**

На основании исходных данных составим последовательность из плюсов и минусов, исходя из условий системы:

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

Таким образом, получим:

| Значение временного ряда $y_i$ | Серии |
|---|---|
| 200 | |
| 194 | - |
| 190 | - |
| 189 | - |
| 197 | + |
| 204 | + |
| 210 | + |
| 193 | - |

Исходный временной ряд состоит из 3 серий, максимальная длина серии составляет 3

Проанализирует полученную систему:

$$\begin{cases} 3 > \left[\frac{1}{3}(2*8-1)-1{,}96\sqrt{\frac{16*8-29}{90}}\right] \\ 3 < 5 \end{cases}$$

Таким образом, гипотеза о наличии тренда принимается. Стоит отметить, что критерий "восходящих" и "нисходящих" серий улавливает смещение оценки математического ожидания монотонного и периодического характера. Это более мощный критерий по сравнению с критерием серий, основанном на медиане выборки.

**Пример 3.3**

Ниже представлен временной ряд, отражающий изменений признака во времени:

| t | Значение временного ряда $y_i$ |
|---|---|
| 1 | 313 |
| 2 | 307 |
| 3 | 303 |
| 4 | 306 |
| 5 | 310 |
| 6 | 317 |
| 7 | 323 |

Проверить стохастическую независимость с помощью критерия квадратов последовательных разностей (критерия Аббе).

**Решение.**

Для вычислений построим вспомогательную таблицу:

| t | Значение временного ряда $y_i$ | $y_{i+1}-y_i$ | $(y_{i+1}-y_i)^2$ | $y_i-\bar{y}$ | $(y_i-\bar{y})^2$ |
|---|---|---|---|---|---|
| 1 | 313 | | | 1,714 | 2,939 |
| 2 | 307 | -6 | 36 | -4,286 | 18,367 |
| 3 | 303 | -4 | 16 | -8,286 | 68,653 |
| 4 | 306 | 3 | 9 | -5,286 | 27,939 |
| 5 | 310 | 4 | 16 | -1,286 | 1,653 |
| 6 | 317 | 7 | 49 | 5,714 | 32,653 |
| 7 | 323 | 6 | 36 | 11,714 | 137,224 |
| | | $\sum$ | 162 | $\sum$ | 289,429 |

Найдем необходимые параметры для расчета:

$$q^2(n) = \frac{1}{2(7-1)}162 = 13{,}5$$

$$s'^2(n) = \frac{1}{7-1}289{,}429 = 48{,}238,$$

При заданных исходных данных $\gamma(n)$ составит:

$$\gamma(n) = \frac{13{,}5}{48{,}238} = 0{,}28$$

Сравним найденное значение с критическим при уровне значимости 95%:

$$0{,}28 < 0{,}468$$

Следовательно, не подтверждается гипотеза о постоянстве центра группирования. То есть имеют место систематические расхождения между результатами наблюдений.

**Пример 3.4**

В таблице приведены данные об изменении значений уровней временного ряда

| Год | Значение временного ряда $y_i$ |
|---|---|
| 2016 | 75,3 |
| 2017 | 73,8 |
| 2018 | 72,1 |
| 2019 | 70,9 |
| 2020 | 68,5 |
| 2021 | 67,1 |
| 2022 | 65,9 |

На основании представленных данных, постройте линейную модель тренда, сформулируйте вывод

**Решение.**

Построим вспомогательную таблицу для расчетов:

| Год | Значение временного ряда $y_i$ | $t_i$ | $y*t_i$ | $t_i^2$ | $\hat{y}$ | $y_i - \hat{y}$ |
|---|---|---|---|---|---|---|
| 2016 | 85,3 | -3 | -255,9 | 9 | 85,36 | -0,06 |
| 2017 | 83,8 | -2 | -167,6 | 4 | 83,74 | 0,06 |
| 2018 | 82,1 | -1 | -82,1 | 1 | 82,13 | -0,03 |
| 2019 | 80,9 | 0 | 0 | 0 | 80,51 | 0,39 |
| 2020 | 78,5 | 1 | 78,5 | 1 | 78,90 | -0,40 |
| 2021 | 77,1 | 2 | 154,2 | 4 | 77,29 | -0,19 |
| 2022 | 75,9 | 3 | 227,7 | 9 | 75,67 | 0,23 |
| $\sum$ | 563,6 | 0 | -45,2 | 28 | 563,6 | 0 |

Для данного тренда параметры a и b:

$$a = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{563{,}6}{7} = 80{,}51$$

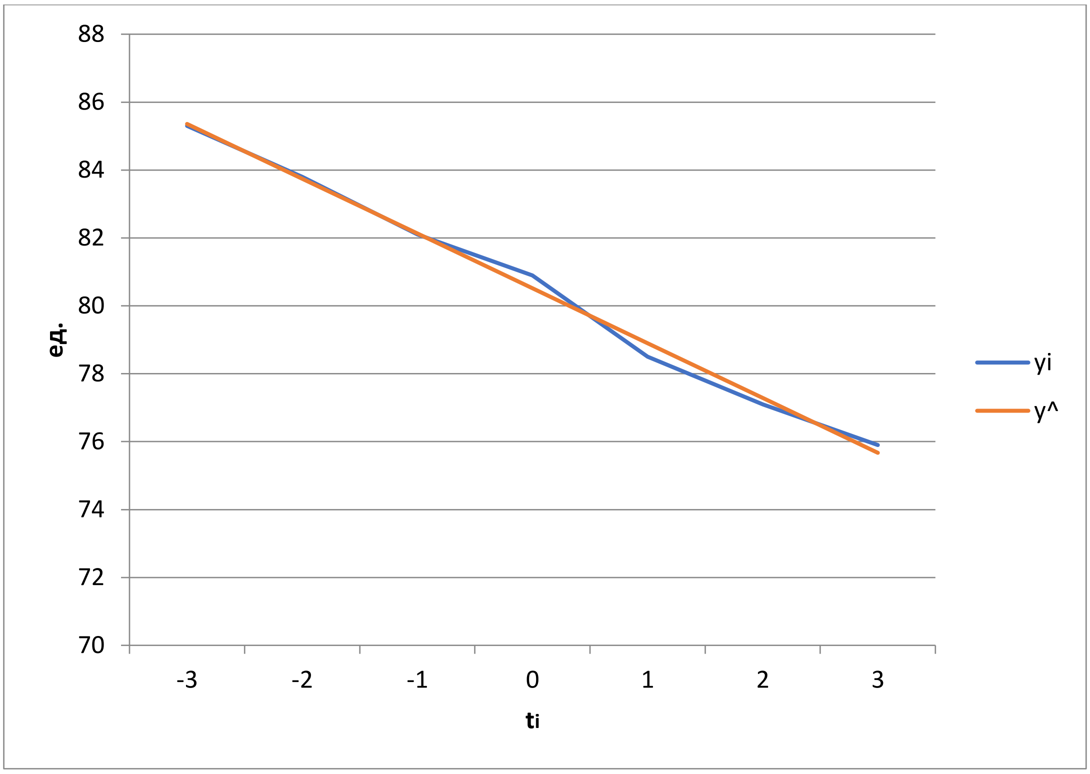
$$b = \frac{\sum_{i=1}^{n} y_i t_i}{\sum_{i=1}^{n} t_i^2} = \frac{-45{,}2}{28} = -1{,}61$$

Таким образом, уравнение тренда имеет вид:

$$\hat{y} = 80{,}51 - 1{,}61t$$

Таким образом, в среднем значение уровня ряда сокращалось на 1,61 единиц в год.

Сумма уровней тренда равна сумме фактических уровней, что свидетельствует о корректности данной модели. Также это может быть подтверждено методом графического анализа:




Близость тренда к заданным уровням временного ряда свидетельствует о том, что линейная модель максимально полно описывает исходный ряд динамики.

**Пример 3.5**

В таблице приведены данные об изменении значений уровней временного ряда

| Год | Уровень временного ряда $y_i$ |
|---|---|
| 2015 | 317 |
| 2016 | 326 |
| 2017 | 340 |
| 2018 | 367 |
| 2019 | 392 |
| 2020 | 414 |
| 2021 | 449 |
| 2022 | 495 |

На основании представленных данных, постройте параболическую модель тренда, представьте прогноз на 3 периода, сформулируйте вывод

**Решение.**

Для вычислений построим вспомогательную таблицу:

| Год | Значение временного ряда $y_i$ | $t_i$ | $y*t_i$ | $y*t_i^2$ | $\hat{y}$ | $y_i-\hat{y}$ |
|---|---|---|---|---|---|---|
| 2015 | 317 | -3,5 | -1109,5 | 3883,25 | 316,42 | 0,58 |
| 2016 | 326 | -2,5 | -815 | 2037,50 | 327,16 | -1,16 |
| 2017 | 340 | -1,5 | -510 | 765,00 | 342,68 | -2,68 |
| 2018 | 367 | -0,5 | -183,5 | 91,75 | 362,99 | 4,01 |
| 2019 | 392 | 0,5 | 196 | 98,00 | 388,08 | 3,92 |
| 2020 | 414 | 1,5 | 621 | 931,50 | 417,97 | -3,97 |
| 2021 | 449 | 2,5 | 1122,5 | 2806,25 | 452,63 | -3,63 |
| 2022 | 495 | 3,5 | 1732,5 | 6063,75 | 492,09 | 2,91 |
| $\sum$ | 3100 | 0 | 1054 | 3883,25 | ≈3100 | ≈0 |

Вычислим параметр b:

$$b=\frac{\sum_{i=1}^{n} y_i t_i}{\sum_{i=1}^{n} t_i^2}=\frac{1054}{\frac{8^2-8}{12}}\approx 25{,}095$$

Для нахождения параметров a и c воспользуемся системой уравнений:

$$\begin{cases} 8a+42c=3100 \\ 42a+c\dfrac{3*8^5-10*8^3+7*8}{240}=16677 \end{cases}$$

Таким образом, получим:

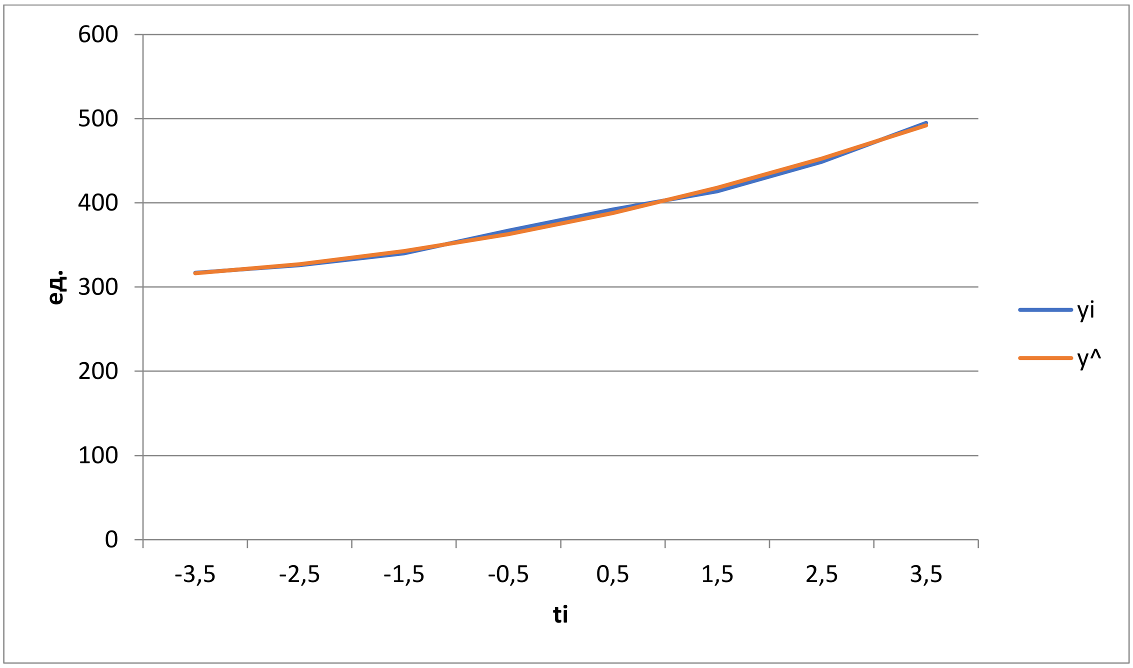
$$\begin{cases} a+5{,}25c=387{,}5 \\ a+9{,}25c=397{,}07 \end{cases}$$

Решая данную систему, получим с = 2,393; а= 374,939

Таким образом, уравнение тренда имеет вид:

$$\hat{y}_i = 374{,}939 + 25{,}095 * t + 2{,}393 * t^2$$

Сравним графически исходный временной ряд и параболический тренд:

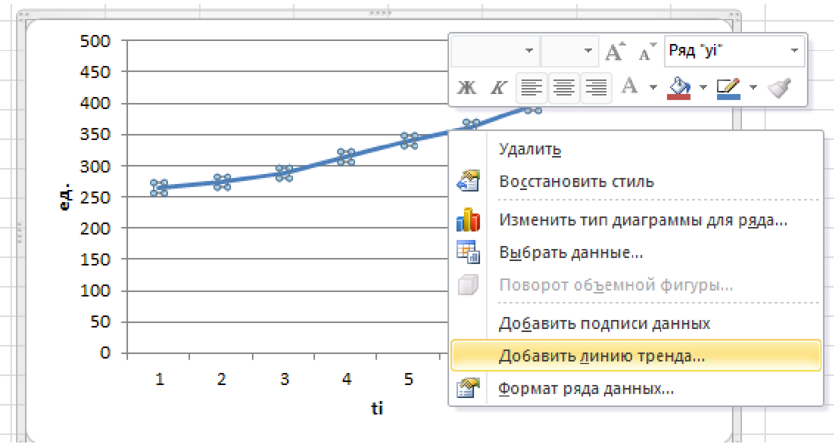



Близость тренда к заданным уровням временного ряда свидетельствует о том, что параболическая модель максимально полно описывает исходный ряд динамики.

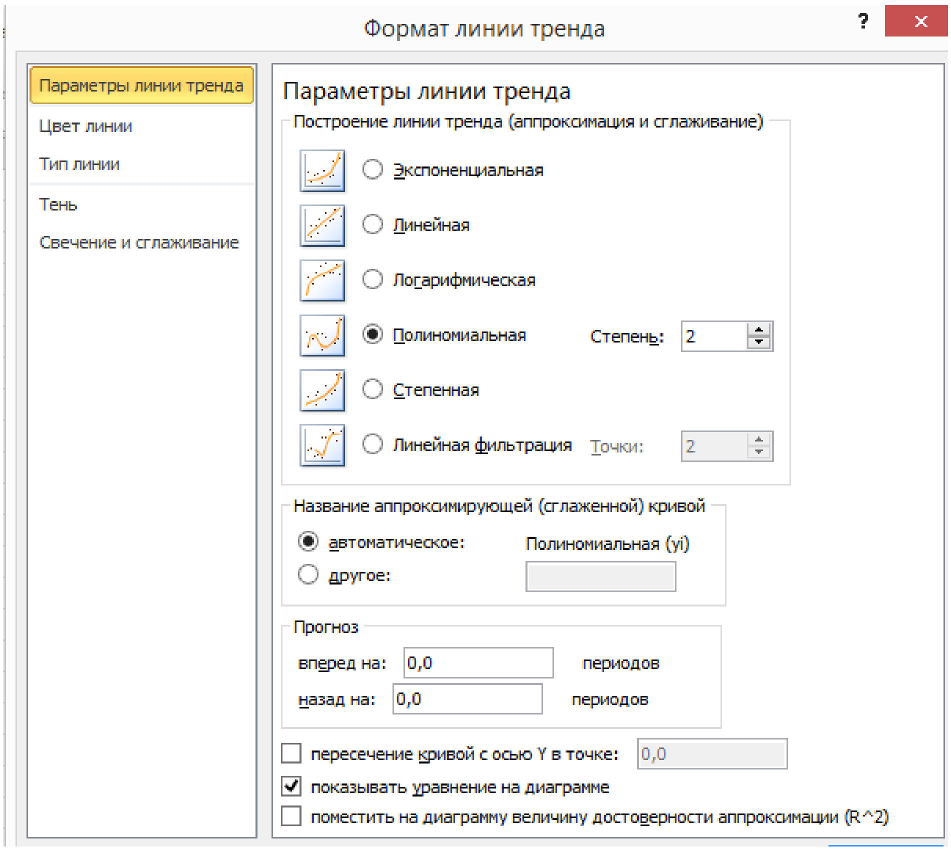
Результаты моделирования можно интерпретировать следующим образом: значение уровней ряда возрастало в номинальной оценке ускоренно, со средним ускорением: 2*2,393 = 4,786 единиц в год, средний за весь период прирост показателя составил 25,095 единиц в год, среднее значение уровня в середине анализируемого периода составляет 374,939 единиц.

Рассмотрим результаты моделирования в Ms Excel. Как уже было отмечено, автоматическое моделирование предусматривает нумерацию лет от начала с номера t = 1, поэтому уравнение имеет иной вид.

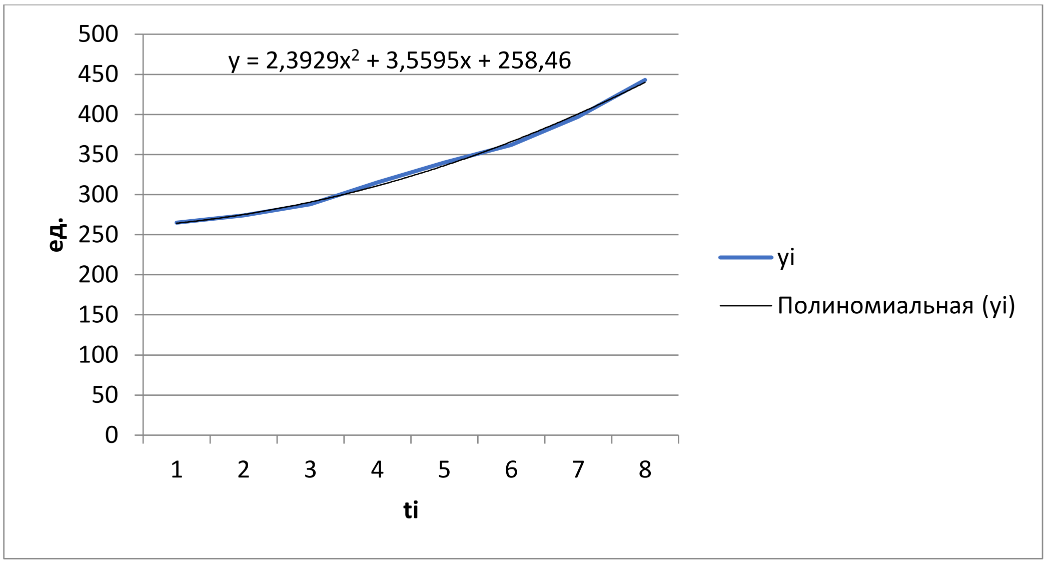
Для построения модели тренда с помощью Ms Excel необходимо построить исходный ряд, а затем, нажав на построенный график, добавить тренд:




Затем необходимо выбрать нужную линию тренда и указать желаемые результаты отображения на графике:



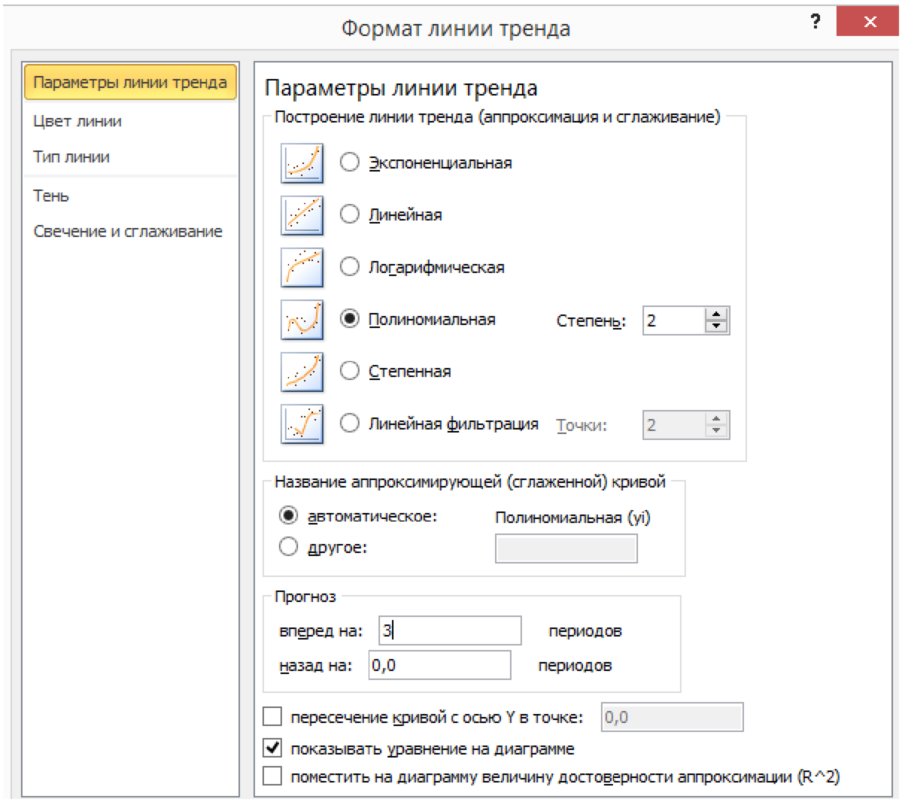

В результате была построена линия тренда следующего вида:



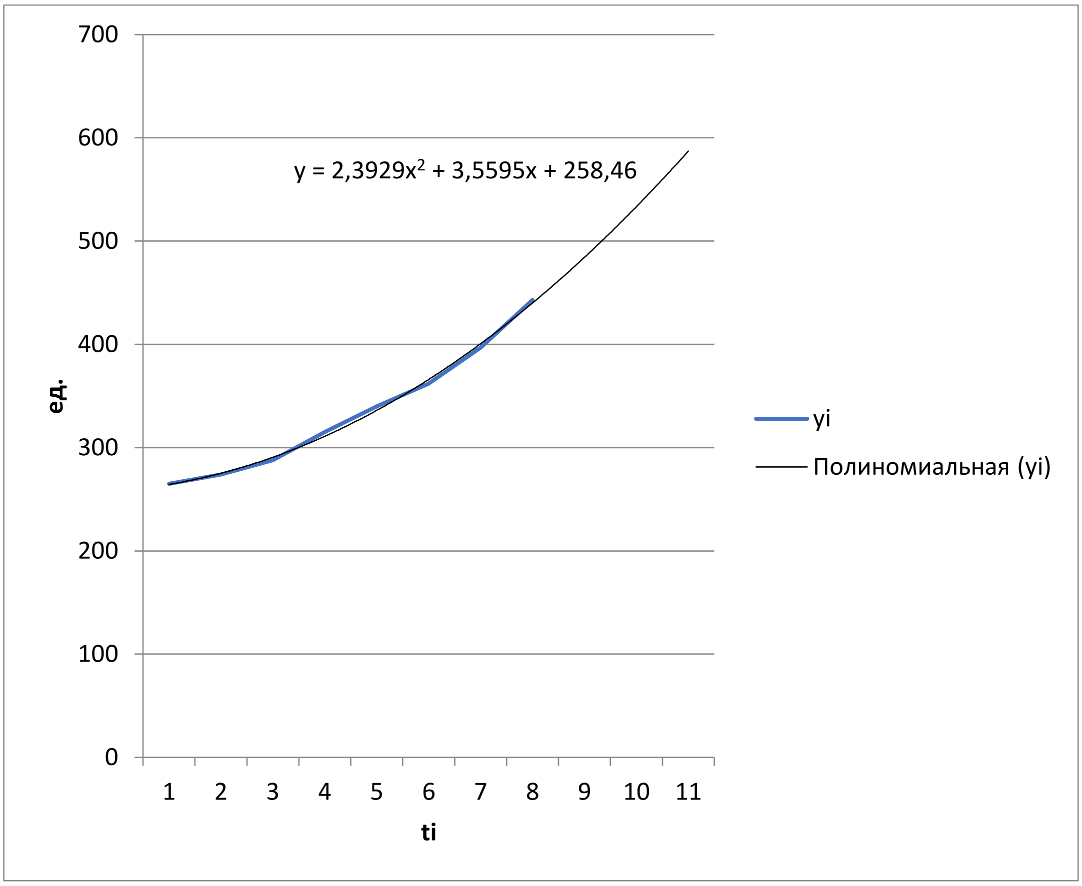

Разница в результатах вычисления обусловлена тем, что за нулевой уровень при построении линии тренда выбраны различные значения параметра t.

Осуществить прогнозирование можно путем подстановки в найденное уравнение тренда последующих значений ti.

Кроме того, можно воспользоваться встроенными средствами MS Excel. Для этого в окне ввода параметров линии тренда необходимо задать желаемый интервал прогнозирования:



В результате получим графическое отображение прогнозных значений на 3 заданных периода:




Стоит отметить, что с увеличением горизонта прогнозирования, точность прогноза снижается.

**Пример 3.6**

В таблице приведены данные об изменении значений уровней временного ряда:

| Год | Уровень временного ряда $y_i$ |
|---|---|
| 2016 | 468 |
| 2017 | 401 |
| 2018 | 367 |
| 2019 | 340 |
| 2020 | 328 |
| 2021 | 326 |
| 2022 | 325 |

На основании представленных данных, постройте гиперболическую модель тренда, сформулируйте вывод

**Решение.**

Для вычислений построим вспомогательную таблицу:

| Год | Значение временного ряда $y_i$ | $t_i$ | $\frac{1}{t_i}$ | $\frac{1}{t_i^2}$ | $\frac{y_i}{t_i}$ | $\hat{y}$ | $y_i - \hat{y}$ |
|---|---|---|---|---|---|---|---|
| 2016 | 491 | 1 | 1,000 | 1,000 | 491,000 | 497,300 | -6,300 |
| 2017 | 424 | 2 | 0,500 | 0,250 | 212,000 | 410,382 | 13,619 |
| 2018 | 390 | 3 | 0,333 | 0,111 | 130,000 | 381,409 | 8,591 |
| 2019 | 362 | 4 | 0,250 | 0,063 | 90,500 | 366,922 | -4,922 |
| 2020 | 351 | 5 | 0,200 | 0,040 | 70,200 | 358,230 | -7,230 |
| 2021 | 349 | 6 | 0,167 | 0,028 | 58,167 | 352,436 | -3,436 |
| 2022 | 348 | 7 | 0,143 | 0,020 | 49,714 | 348,297 | -0,297 |
| $\sum$ | 2715 | 28 | 2,593 | 1,512 | 1101,581 | ≈2715 | ≈0 |

Для данных исходных данный система нормальных уравнений имеет вид:

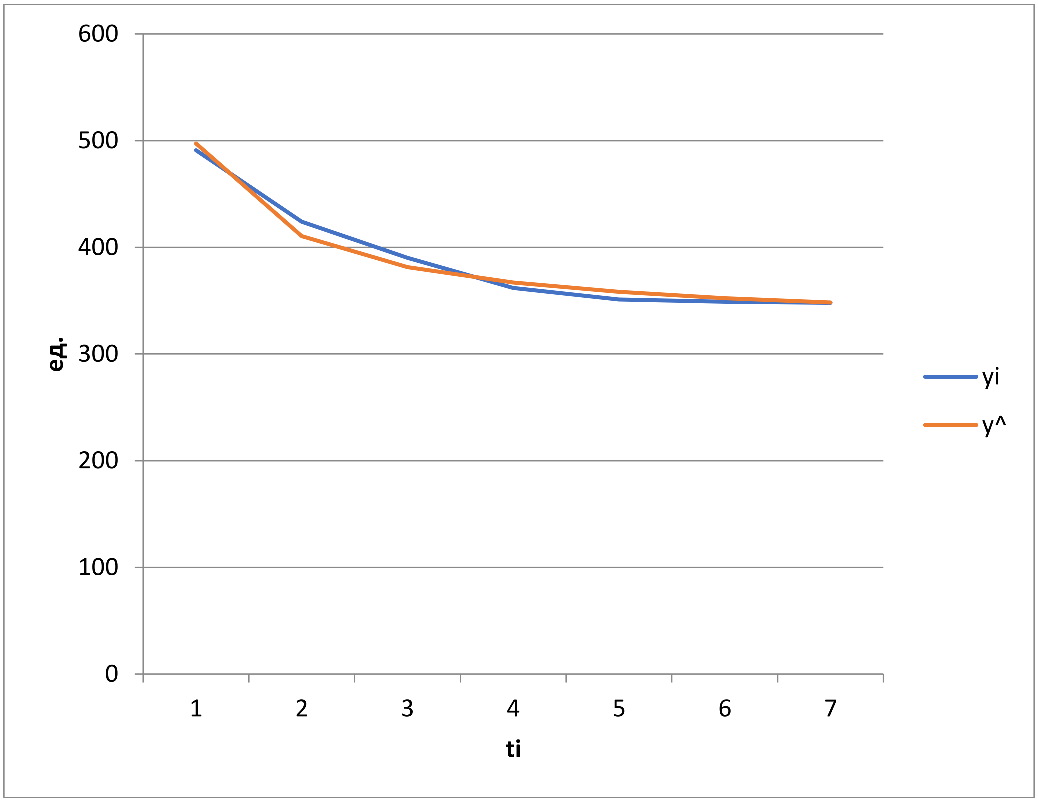
$$\begin{cases} 7a + 2{,}593b = 2715 \\ 2{,}593a + 1{,}512b = 1101{,}581 \end{cases}$$

Решая данную систему, получим a = 323,463; b= 173,837

Таким образом, уравнение тренда имеет вид:

$$\hat{y} = 323{,}463 + \frac{173{,}837}{t}$$

График представлен ниже:




Таким образом, величина 323,463 – предельное значение уровня ряда, к которому стремится величина показателя $y_i$.

**Пример 3.7**

С помощью Ms Excel рассчитать параметры экспоненциального тренда по данным временного ряда, приведенным в таблице

| Год | Значение временного ряда $y_i$ |
|---|---|
| 2017 | 2527 |
| 2018 | 3060 |
| 2019 | 3727 |
| 2020 | 4430 |
| 2021 | 5241 |
| 2022 | 6160 |

**Решение**

Для расчёта параметров экспоненциального тренда построим вспомогательную таблицу

| Год | Значение временного ряда $y_i$ | $t_i$ | $y_i t_i$ | $\ln y_i$ | $t_i \ln y_i$ | $\hat{y}$ | $y_i - \hat{y}$ |
|---|---|---|---|---|---|---|---|
| 2017 | 2540 | -2,5 | -6350 | 7,840 | -19,600 | 2567,319 | -27,319 |
| 2018 | 3073 | -1,5 | -4609,5 | 8,030 | -12,046 | 3066,740 | 6,260 |
| 2019 | 3740 | -0,5 | -1870 | 8,227 | -4,113 | 3663,314 | 76,686 |
| 2020 | 4443 | 0,5 | 2221,5 | 8,399 | 4,200 | 4375,939 | 67,061 |
| 2021 | 5254 | 1,5 | 7881 | 8,567 | 12,850 | 5227,192 | 26,808 |
| 2022 | 6173 | 2,5 | 15432,5 | 8,728 | 21,820 | 6244,039 | -71,039 |
| $\sum$ | 25223 | 0 | 12705,5 | 49,791 | 3,111 | 25144,544 | 78,456 |

Согласно данным таблицы:

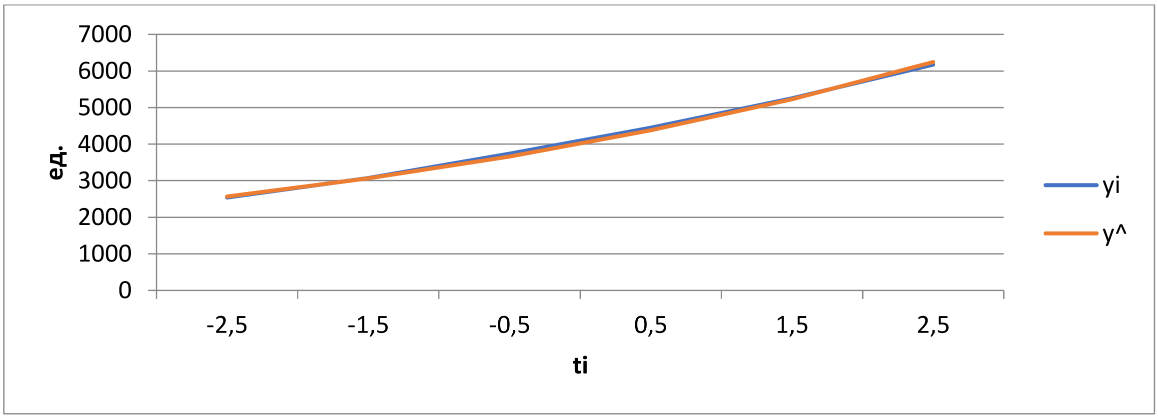
$$\ln a = \frac{49{,}791}{6}$$, следовательно, a = 4003,803

$$\ln b = \frac{3{,}111}{17{,}5}$$, следовательно, b = 1,195

Таким образом, уравнение тренда имеет вид:

$$\hat{y} = 4003{,}803 * 1{,}195^t$$

Сравним тренд с исходным уровнем ряда:




Результаты вычислений показали, что исходное и расчетное значение уровня ряда динамики отличаются друг от друга. Также разница между уровнями видна на графике. Такая неточность может быть объяснена, в частности, малым числом наблюдений для построения уравнения экспоненциального тренда

**Пример 3.8.**

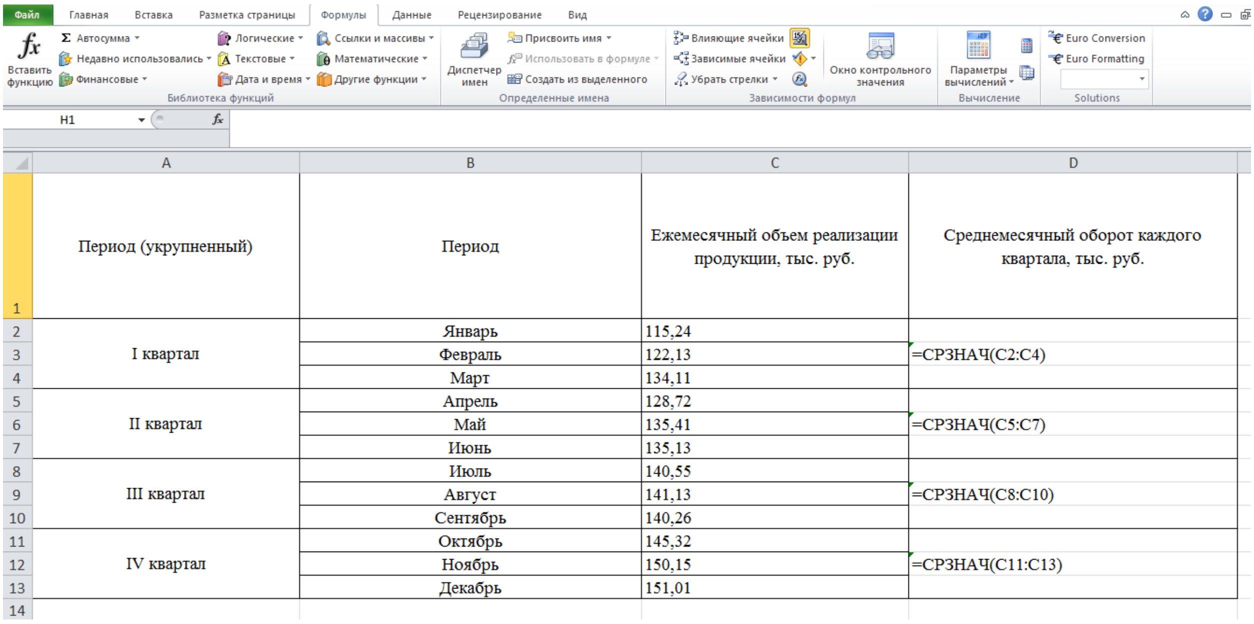
В таблице приведены данные о ежемесячном объеме реализации продукции. На основании исходным данных, проведите сглаживание ряда динамики методом укрупнения интервалов, используя при этом трехзвенные интервалы

| Период | Объем реализации продукции, тыс. руб. |
|---|---|
| Январь | 115,24 |
| Февраль | 122,13 |
| Март | 134,11 |
| Апрель | 128,72 |
| Май | 135,41 |
| Июнь | 135,13 |
| Июль | 140,55 |
| Август | 141,13 |
| Сентябрь | 140,26 |
| Октябрь | 145,32 |
| Ноябрь | 150,15 |
| Декабрь | 151,01 |

**Решение**

Для проведения сглаживания временных рядов, вычислим среднемесячное значение для каждого квартала (поскольку применяются трехзвенные интервалы, а исходя из логики исходных данных три последовательных месяца представляют собой квартал)

Представленное в формульном виде решение выглядит следующим образом:

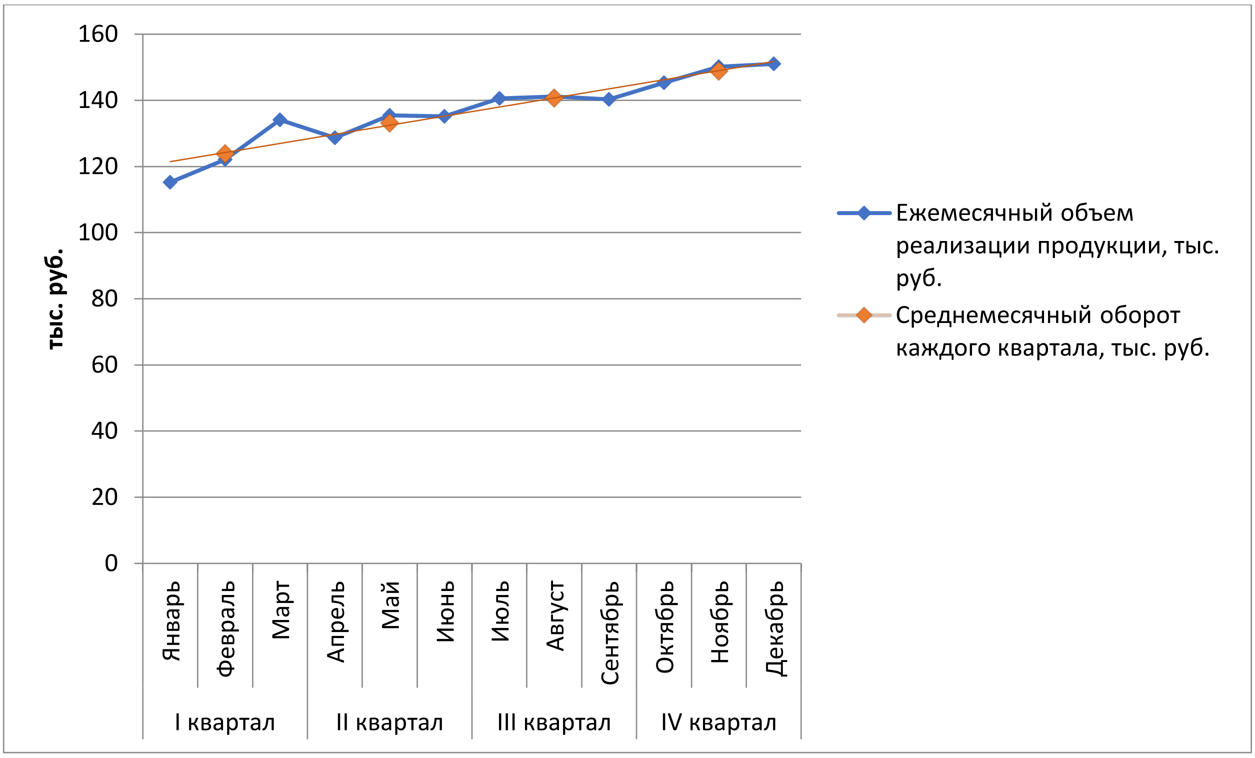


| Период (укрупненный) | Период | Ежемесячный объем реализации продукции, тыс. руб. | Среднемесячный оборот каждого квартала, тыс. руб. |
|---|---|---|---|
| I квартал | Январь | 115,24 | |
| | Февраль | 122,13 | =СРЗНАЧ(C2:C4) |
| | Март | 134,11 | |
| II квартал | Апрель | 128,72 | |
| | Май | 135,41 | =СРЗНАЧ(C5:C7) |
| | Июнь | 135,13 | |
| III квартал | Июль | 140,55 | |
| | Август | 141,13 | =СРЗНАЧ(C8:C10) |
| | Сентябрь | 140,26 | |
| IV квартал | Октябрь | 145,32 | |
| | Ноябрь | 150,15 | =СРЗНАЧ(C11:C13) |
| | Декабрь | 151,01 | |

Результаты вычислений для заданных исходных данных приведены в таблице:

| Период (укрупненный) | Период | Ежемесячный объем реализации продукции, тыс. руб. | Среднемесячный оборот каждого квартала, тыс. руб. |
|---|---|---|---|
| I квартал | Январь | 115,24 | = (115,24 + 122,12+134,11)/3 = 123,83 |
| | Февраль | 122,13 | |
| | Март | 134,11 | |
| II квартал | Апрель | 128,72 | 133,09 |
| | Май | 135,41 | |
| | Июнь | 135,13 | |
| III квартал | Июль | 140,55 | 140,65 |
| | Август | 141,13 | |
| | Сентябрь | 140,26 | |
| IV квартал | Октябрь | 145,32 | 148,83 |
| | Ноябрь | 150,15 | |
| | Декабрь | 151,01 | |

Для более наглядного представления результатов сглаживания методов укрупнения интервалов, постоим график, на котором отображен исходный и сглаженный с использованием трехзвенных интервалов временной ряд




В результате процедуры сглаживания был получен новый ряд, колеблемость которого стала значительно меньше по сравнению с исходным рядом.

**Пример 3.9**

По исходным данным проведите сглаживание ряда динамики методом простой скользящей средней с пятизвенным интервалом

| | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| | ,3 | ,6 | ,9 | ,4 | ,6 | ,8 | ,6 | ,9 | ,2 | ,4 | ,1 | ,3 | ,7 | ,9 | ,2 | ,5 |

**Решение**

Для построения сглаженного ряда воспользуемся формулой:

$$\overline{y}_i = \frac{\sum\limits_{t=1}^{m} y_t}{m}$$

По данной формуле будут получены сглаженные методом простой скользящей средней 13 уровней нового ряда.

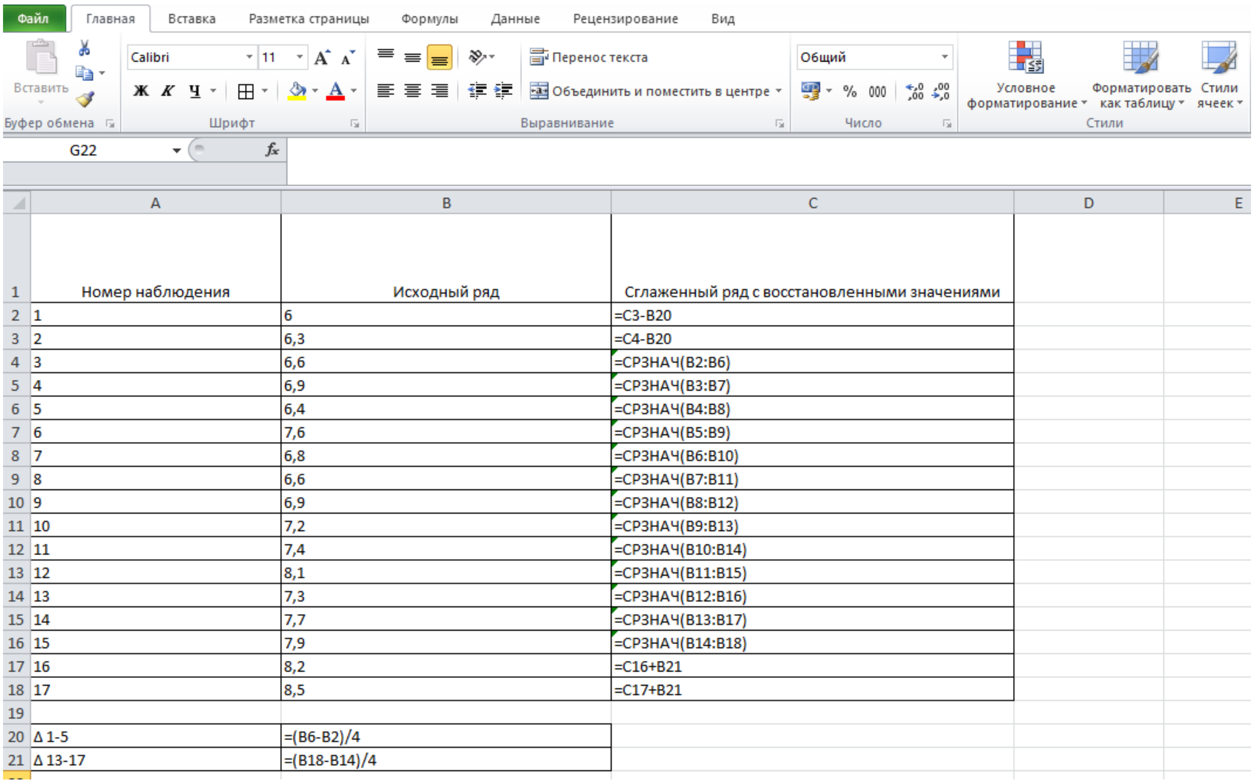
Вычислим значение среднего абсолютного прироста для первого и последнего участка скольжения:

$$\Delta_{1-5} = \frac{y_5 - y_1}{m-1}$$

$$\Delta_{13-17} = \frac{y_{17} - y_{13}}{m-1}$$

Для восстановления потерянных значений на последнем участке временного ряда последовательно прибавим величину среднего абсолютного прироста к последнему сглаженному значению. Аналогичным образом восстановим первые члены ряда.

Представленное в формульном виде решение выглядит следующим образом:



| | A | B | C |
|---|---|---|---|
| 1 | Номер наблюдения | Исходный ряд | Сглаженный ряд с восстановленными значениями |
| 2 | 1 | 6 | =C3-B20 |
| 3 | 2 | 6,3 | =C4-B20 |
| 4 | 3 | 6,6 | =СРЗНАЧ(B2:B6) |
| 5 | 4 | 6,9 | =СРЗНАЧ(B3:B7) |
| 6 | 5 | 6,4 | =СРЗНАЧ(B4:B8) |
| 7 | 6 | 7,6 | =СРЗНАЧ(B5:B9) |
| 8 | 7 | 6,8 | =СРЗНАЧ(B6:B10) |
| 9 | 8 | 6,6 | =СРЗНАЧ(B7:B11) |
| 10 | 9 | 6,9 | =СРЗНАЧ(B8:B12) |
| 11 | 10 | 7,2 | =СРЗНАЧ(B9:B13) |
| 12 | 11 | 7,4 | =СРЗНАЧ(B10:B14) |
| 13 | 12 | 8,1 | =СРЗНАЧ(B11:B15) |
| 14 | 13 | 7,3 | =СРЗНАЧ(B12:B16) |
| 15 | 14 | 7,7 | =СРЗНАЧ(B13:B17) |
| 16 | 15 | 7,9 | =СРЗНАЧ(B14:B18) |
| 17 | 16 | 8,2 | =C16+B21 |
| 18 | 17 | 8,5 | =C17+B21 |
| 19 | | | |
| 20 | Δ 1-5 | =(B6-B2)/4 | |
| 21 | Δ 13-17 | =(B18-B14)/4 | |

Таким образом, величина средних абсолютных приростов для исходных данных составит:

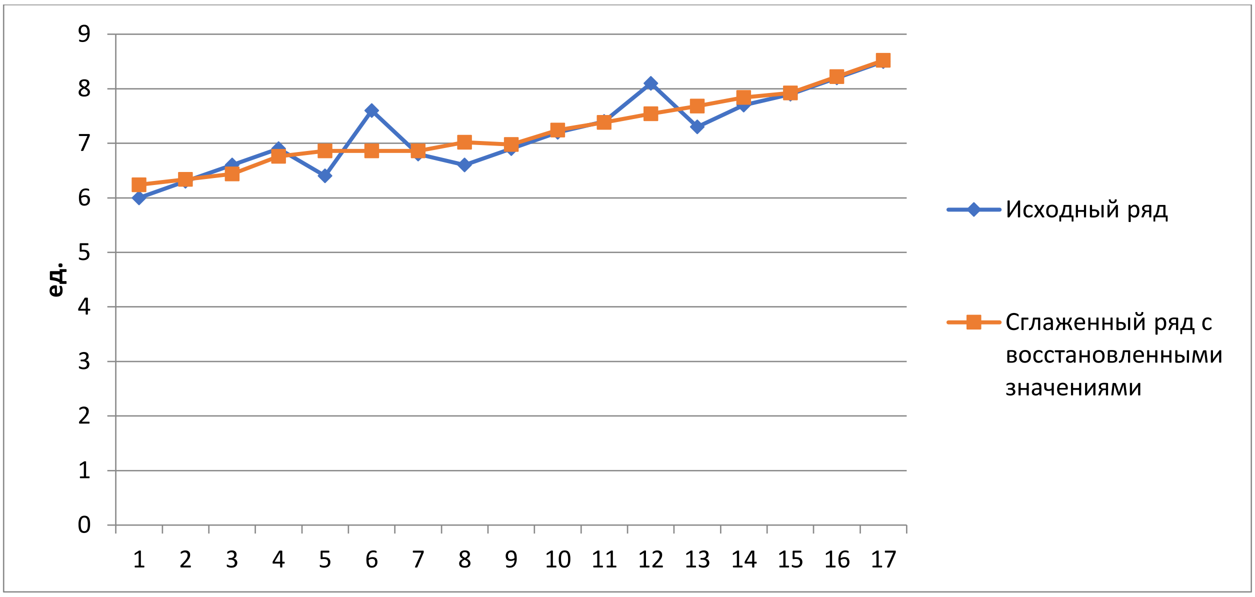
$$\Delta_{1-5} = \frac{6{,}4-6}{4} = 0{,}1$$

$$\Delta_{13-17} = \frac{8{,}5-7{,}3}{4} = 0{,}3$$

Результаты вычислений для заданных исходных данных приведены в таблице:

| Номер наблюдения | Исходный ряд | Сглаженный ряд с восстановленными значениями |
|---|---|---|
| 1 | 6 | 6,34-0,1=6,24 |
| 2 | 6,3 | 6,44-0,1=6,34 |
| 3 | 6,6 | =(6+6,3,+6,6+6,9+6,4)/5=6,44 |
| 4 | 6,9 | =(6,3,+6,6+6,9+6,4+7,6)/5 =6,76 |
| 5 | 6,4 | =(6,6+6,9+6,4+7,6+6,8)/5 =6,86 |
| 6 | 7,6 | 6,86 |
| 7 | 6,8 | 6,86 |
| 8 | 6,6 | 7,02 |
| 9 | 6,9 | 6,98 |
| 10 | 7,2 | 7,24 |
| 11 | 7,4 | 7,38 |
| 12 | 8,1 | 7,54 |
| 13 | 7,3 | 7,68 |
| 14 | 7,7 | 7,84 |
| 15 | 7,9 | =(7,3+7,7+7,9+8,2+8,5)/5 =7,92 |
| 16 | 8,2 | 7,92 + 0,3 = 8,22 |
| 17 | 8,5 | 8,22+0,3 = 8,52 |

Для более наглядного представления результатов, постоим график, на котором отображен исходный и сглаженный методом простой скользящей средней с использованием пятизвенных интервалов временной ряд:




По графику видно, что получившийся ряд является более сглаженным по сравнению с исходными динамическим рядом.

**Пример 3.10**

По исходным данным проведите сглаживание ряда динамики методом простой скользящей средней с четырехзвенным интервалом

| | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
| 6 | 6,3 | 6,6 | 6,9 | 6,4 | 7,6 | 6,8 | 6,6 | 6,9 | 7,2 | 7,4 | 8,1 | 7,3 | 7,7 | 7,9 | 8,2 | 8,5 |

**Решение**

Нахождение сглаженных значений также осуществляется по формуле:

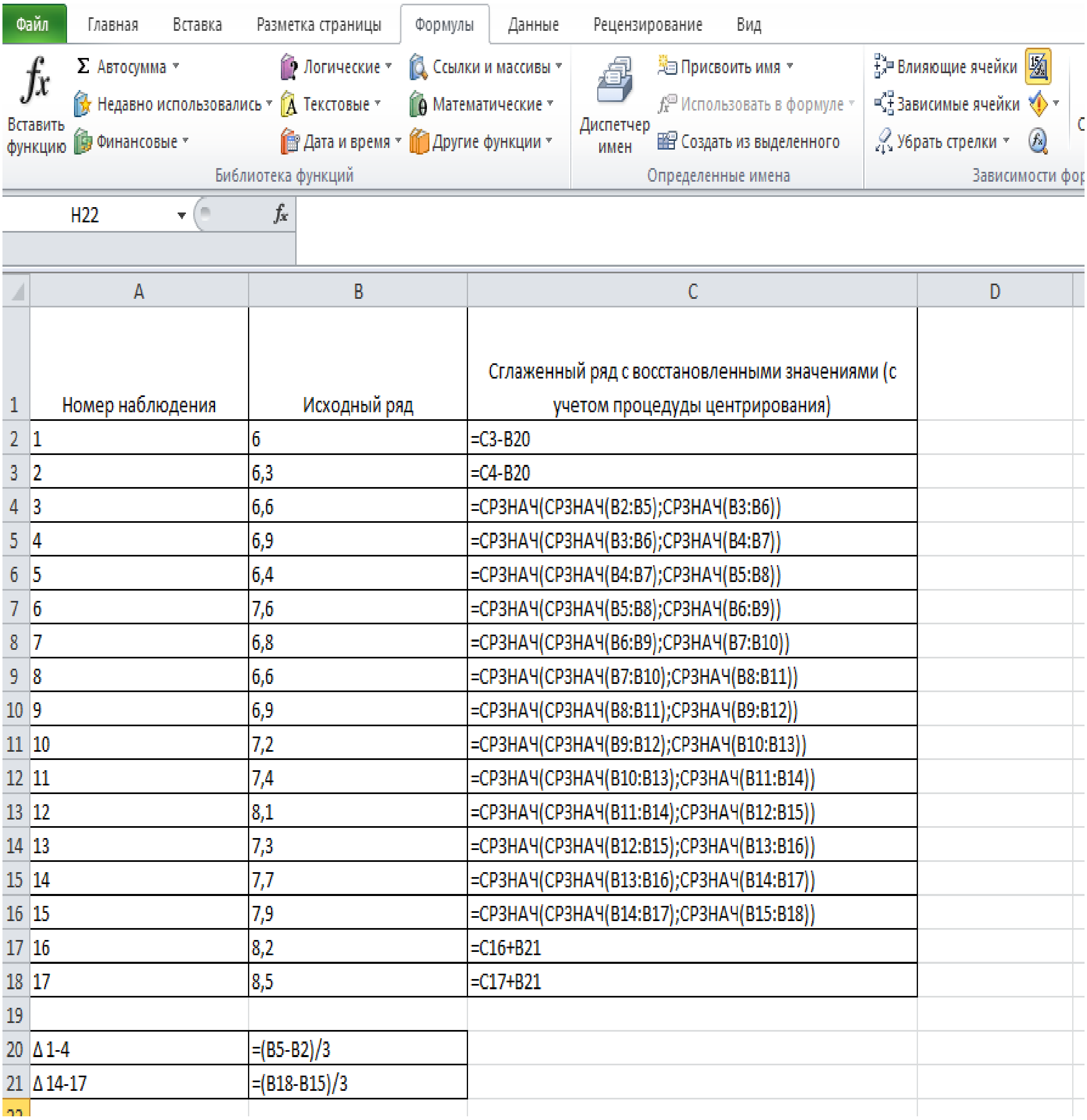
$$\overline{y_i} = \frac{\sum_{t=1}^{m} y_t}{m}$$

Полученные значения нового уровня ряда находятся между исходными уровнями. Для решения данной проблемы проводится процедура центрирования, когда для каждых соседних сглаженных значений находится среднее арифметическое.

Поскольку в данном примере используется четырехзвенное сглаживание, можно сделать вывод о том в результате процедуры сглаживания произойдет «укорачивание» ряда по 2 уровня с каждого конца.

Восполнение недостающих значений производится также с использованием средних абсолютных приростов

Представленное в формульном виде решение выглядит следующим образом:




| | A | B | C | D |
|---|---|---|---|---|
| 1 | Номер наблюдения | Исходный ряд | Сглаженный ряд с восстановленными значениями (с учетом процедуды центрирования) | |
| 2 | 1 | 6 | =C3-B20 | |
| 3 | 2 | 6,3 | =C4-B20 | |
| 4 | 3 | 6,6 | =СРЗНАЧ(СРЗНАЧ(B2:B5);СРЗНАЧ(B3:B6)) | |
| 5 | 4 | 6,9 | =СРЗНАЧ(СРЗНАЧ(B3:B6);СРЗНАЧ(B4:B7)) | |
| 6 | 5 | 6,4 | =СРЗНАЧ(СРЗНАЧ(B4:B7);СРЗНАЧ(B5:B8)) | |
| 7 | 6 | 7,6 | =СРЗНАЧ(СРЗНАЧ(B5:B8);СРЗНАЧ(B6:B9)) | |
| 8 | 7 | 6,8 | =СРЗНАЧ(СРЗНАЧ(B6:B9);СРЗНАЧ(B7:B10)) | |
| 9 | 8 | 6,6 | =СРЗНАЧ(СРЗНАЧ(B7:B10);СРЗНАЧ(B8:B11)) | |
| 10 | 9 | 6,9 | =СРЗНАЧ(СРЗНАЧ(B8:B11);СРЗНАЧ(B9:B12)) | |
| 11 | 10 | 7,2 | =СРЗНАЧ(СРЗНАЧ(B9:B12);СРЗНАЧ(B10:B13)) | |
| 12 | 11 | 7,4 | =СРЗНАЧ(СРЗНАЧ(B10:B13);СРЗНАЧ(B11:B14)) | |
| 13 | 12 | 8,1 | =СРЗНАЧ(СРЗНАЧ(B11:B14);СРЗНАЧ(B12:B15)) | |
| 14 | 13 | 7,3 | =СРЗНАЧ(СРЗНАЧ(B12:B15);СРЗНАЧ(B13:B16)) | |
| 15 | 14 | 7,7 | =СРЗНАЧ(СРЗНАЧ(B13:B16);СРЗНАЧ(B14:B17)) | |
| 16 | 15 | 7,9 | =СРЗНАЧ(СРЗНАЧ(B14:B17);СРЗНАЧ(B15:B18)) | |
| 17 | 16 | 8,2 | =C16+B21 | |
| 18 | 17 | 8,5 | =C17+B21 | |
| 19 | | | | |
| 20 | Δ 1-4 | =(B5-B2)/3 | | |
| 21 | Δ 14-17 | =(B18-B15)/3 | | |

Величина средних абсолютных приростов для исходных данных составит:

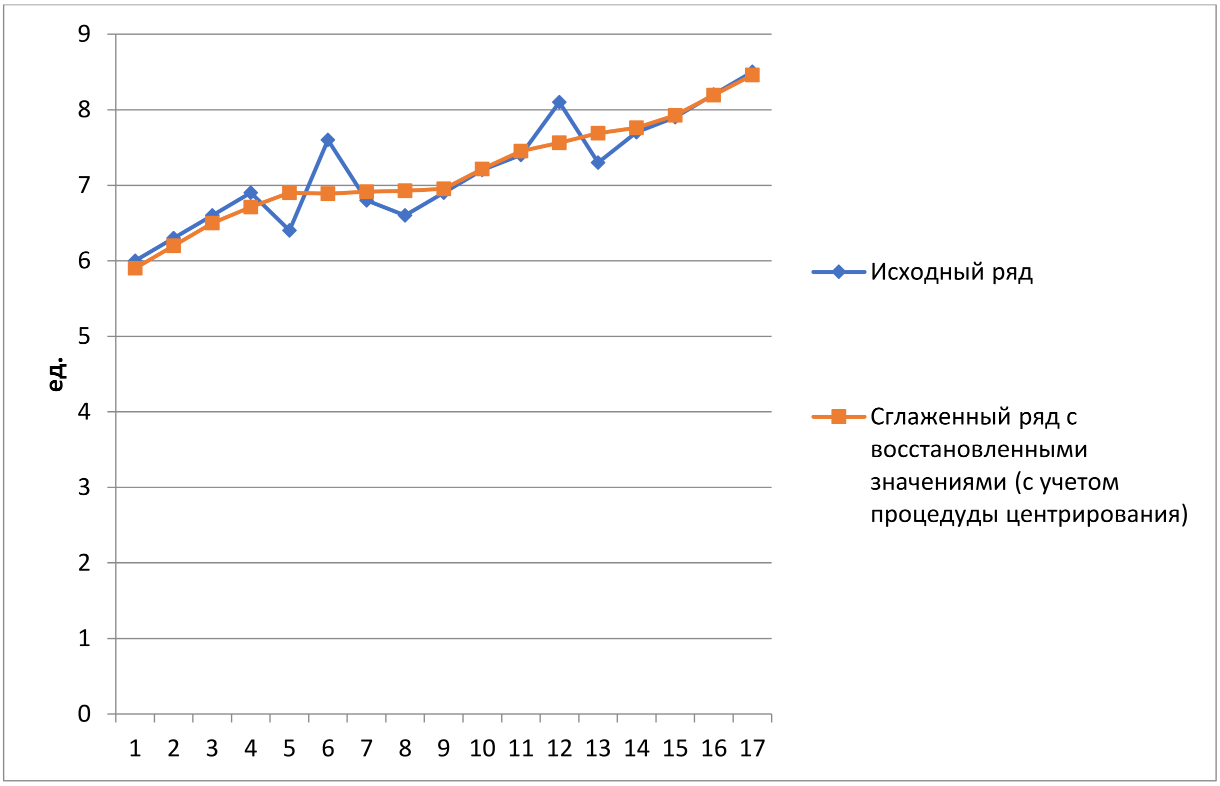
$$\Delta_{1-4} = \frac{6{,}9 - 6}{3} = 0{,}3$$

$$\Delta_{14-17} = \frac{8{,}5 - 7{,}7}{3} = 0{,}27$$

Результаты вычислений для заданных исходных данных приведены в таблице:

| Номер наблюдения | Исходный ряд | Сглаженный ряд с восстановленными значениями (с учетом процедуры центрирования) |
|---|---|---|
| 1 | 6 | 6,2-0,3=5,9 |
| 2 | 6,3 | 6,5-0,3=6,2 |
| 3 | 6,6 | = ((6+6,3+6,6+6,9)/4+ (6,3+6,6+6,9+6,4)/4)/2 = 6,5 |
| 4 | 6,9 | =((6,3+6,6+6,9+6,4)/4 + (6,6+6,9+6,4+7,6)/4)/2 = 6,7 |
| 5 | 6,4 | 6,90 |
| 6 | 7,6 | 6,89 |
| 7 | 6,8 | 6,91 |
| 8 | 6,6 | 6,93 |
| 9 | 6,9 | 6,95 |
| 10 | 7,2 | 7,21 |
| 11 | 7,4 | 7,45 |
| 12 | 8,1 | 7,56 |
| 13 | 7,3 | 7,69 |
| 14 | 7,7 | 7,76 |
| 15 | 7,9 | =(7,3+7,7+7,9+8,2)/4+(7,7+7,9+8,2+8,5)/4)/2 = 7,93 |
| 16 | 8,2 | =7,93+0,27 = 8,19 |
| 17 | 8,5 | =8,19+0,27=8,46 |

Для более наглядного представления результатов, постоим график, на котором отображен исходный и сглаженный методом простой скользящей средней с использованием четырехзвенных интервалов временной ряд:




Сравнивая полученные результаты данной задачи и задачи 6, можно сформулировать следующий вывод: чем меньше интервал сглаживания, тем ближе рассчитанный ряд к исходному временному ряду.

**Пример 3.11**

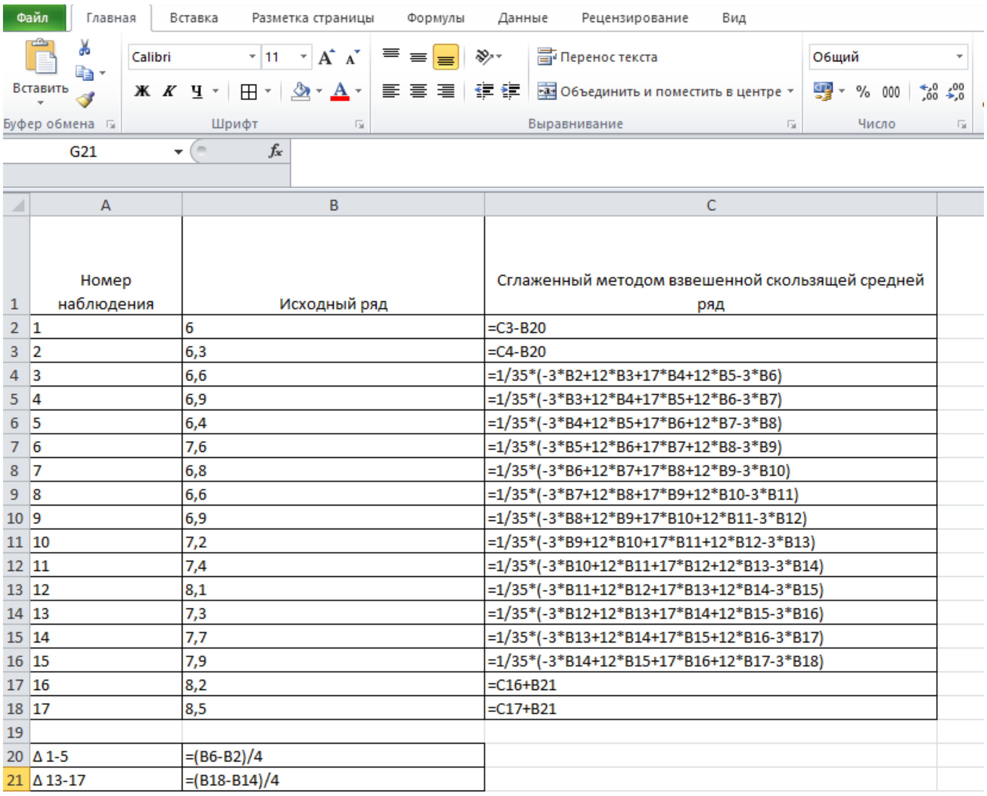
По исходным данным проведите пятизвенное сглаживание ряда динамики методом взвешенной скользящей средней

| | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
| 6 | 6,3 | 6,6 | 6,9 | 6,4 | 7,6 | 6,8 | 6,6 | 6,9 | 7,2 | 7,4 | 8,1 | 7,3 | 7,7 | 7,9 | 8,2 | 8,5 |

**Решение**

Сглаживание методом взвешенной скользящей средней предполагает применение весовых коэффициентов сглаживания.

Представленное в формульном виде решение выглядит следующим образом:

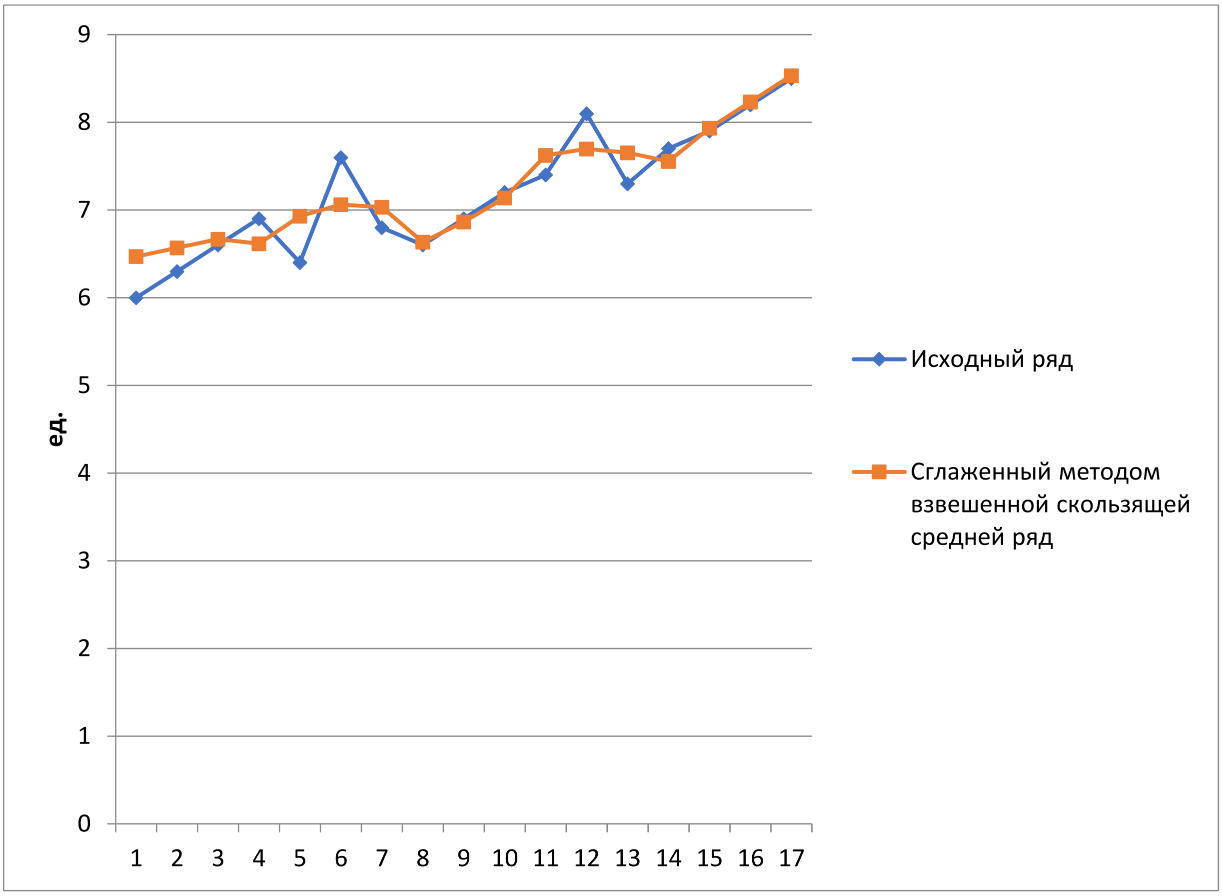



| | A | B | C |
|---|---|---|---|
| 1 | Номер наблюдения | Исходный ряд | Сглаженный методом взвешенной скользящей средней ряд |
| 2 | 1 | 6 | =C3-B20 |
| 3 | 2 | 6,3 | =C4-B20 |
| 4 | 3 | 6,6 | =1/35*(-3*B2+12*B3+17*B4+12*B5-3*B6) |
| 5 | 4 | 6,9 | =1/35*(-3*B3+12*B4+17*B5+12*B6-3*B7) |
| 6 | 5 | 6,4 | =1/35*(-3*B4+12*B5+17*B6+12*B7-3*B8) |
| 7 | 6 | 7,6 | =1/35*(-3*B5+12*B6+17*B7+12*B8-3*B9) |
| 8 | 7 | 6,8 | =1/35*(-3*B6+12*B7+17*B8+12*B9-3*B10) |
| 9 | 8 | 6,6 | =1/35*(-3*B7+12*B8+17*B9+12*B10-3*B11) |
| 10 | 9 | 6,9 | =1/35*(-3*B8+12*B9+17*B10+12*B11-3*B12) |
| 11 | 10 | 7,2 | =1/35*(-3*B9+12*B10+17*B11+12*B12-3*B13) |
| 12 | 11 | 7,4 | =1/35*(-3*B10+12*B11+17*B12+12*B13-3*B14) |
| 13 | 12 | 8,1 | =1/35*(-3*B11+12*B12+17*B13+12*B14-3*B15) |
| 14 | 13 | 7,3 | =1/35*(-3*B12+12*B13+17*B14+12*B15-3*B16) |
| 15 | 14 | 7,7 | =1/35*(-3*B13+12*B14+17*B15+12*B16-3*B17) |
| 16 | 15 | 7,9 | =1/35*(-3*B14+12*B15+17*B16+12*B17-3*B18) |
| 17 | 16 | 8,2 | =C16+B21 |
| 18 | 17 | 8,5 | =C17+B21 |
| 19 | | | |
| 20 | Δ 1-5 | =(B6-B2)/4 | |
| 21 | Δ 13-17 | =(B18-B14)/4 | |

Результаты вычислений для заданных исходных данных приведены в таблице:

| Номер наблюдения | Исходный ряд | Сглаженный методом взвешенной скользящей средней ряд |
|---|---|---|
| 1 | 6 | =6,57-0,1=6,47 |
| 2 | 6,3 | =6,67-0,1=6,57 |
| 3 | 6,6 | =1/35*(-3*6+12*6,3+17*6,6+12*6,9-3*6,4)= 6,67 |
| 4 | 6,9 | =1/35*(-3*6,3+12*6,6+17*6,9+12*6,4-3*7,6)= 6,62 |
| 5 | 6,4 | 6,93 |
| 6 | 7,6 | 7,06 |
| 7 | 6,8 | 7,03 |
| 8 | 6,6 | 6,63 |
| 9 | 6,9 | 6,87 |
| 10 | 7,2 | 7,14 |
| 11 | 7,4 | 7,62 |
| 12 | 8,1 | 7,70 |
| 13 | 7,3 | 7,65 |
| 14 | 7,7 | 7,55 |
| 15 | 7,9 | =1/35*(-3*7,3+12*7,7+17*7,9+12*8,2-3*8,5)= 7,93 |
| 16 | 8,2 | =7,93+0,3=8,23 |
| 17 | 8,5 | =8,23+0,3=8,53 |

На графике представлен исходный и сглаженный методом взвешенной скользящей средней временной ряд:




Применение метода сглаженной взвешенной средней целесообразно в тех случаях, когда изменения ряда описываются нелинейной функцией. По графику видно, что использование данного способа позволило построить сглаженный ряд, достаточно точно описывающий интервалы роста и спада.

# Глава 4. ОЦЕНКА ПЕРИОДИЧЕСКОЙ КОМПОНЕНТЫ В ДИНАМИКЕ ИЗМЕНЕНИЯ ПОКАЗАТЕЛЕЙ ДЕЯТЕЛЬНОСТИ ФИРМЫ

## 4.1. Методы выявления периодической компоненты

Для проверки предположения о существенности периодической компоненты ряда динамики целесообразно использовать такие критерии случайности, которые имеют наибольшую мощность относительно альтернативной гипотезы о цикличности ряда. Наиболее простым для применения и зрительно понятным является критерий «пиков» и «ям». В основе этого критерия лежит подсчет числа экстремальных точек ряда p, который осуществляется следующим образом(4.1):

$$p_t = \begin{cases} 1, \quad \text{если} \quad y_{t-1} < y_t < y_{t+1} \\ \text{или} \quad y_{t-1} > y_t > y_{t+1} \\ 0 \quad \text{в} \quad \text{противном} \quad \text{случае} \end{cases} \tag{4.1}$$

где t=n+1,

n – число наблюдений в ряду динамики.

Для случайного ряда математическое ожидание числа экстремальных точек (4.2):

$$\bar{p} = \frac{2 \cdot (n-2)}{3} \tag{4.2}$$

Проверка гипотезы сводится к сравнению $\bar{p}$ с расчетным значением $\hat{p}$. Если эти значения близки, то можно отказаться от дальнейшей проверки и признать ряд случайным. Если же $\hat{p}$ и $\bar{p}$ значительно отличаются друг от друга, то проводится дальнейшая проверка гипотезы, основанная на подсчете фаз различной длины [99].

Фазой называется интервал между двумя соседними уровнями, для которых $p_t = 1$. Для определения длины фазы l достаточно просто найти разности индексов двух соседних экстремальных точек, затем подсчитать число фаз $N_1$, $N_2$, $N_3$ длин $l_1 = 1$, $l_2 = 2, l_3 = 3$. Теоретическое значение числа фаз длины l для случайного ряда следующее (4.3):

$$\hat{N}_l = \frac{2 \cdot (n - l - 2) \cdot (l^2 - 3 \cdot l + 1)}{(l + 3)!} \qquad (4.3)$$

Естественная процедура проверки случайности сводится к сравнению наблюдаемых значений $N_1$, $N_2$, $N_3$ с теоретическим значением $N_l$. Однако при небольшом числе наблюдений n критерий $\chi^2$ здесь непосредственно использовать нельзя, так как в этом случае длины фаз $l_i$ не являются независимыми. Доказано, что при разбиении длины фазы на три группы: $l_1 = 1$, $l_2 = 2$, $l_3 = 3$ (две степени свободы) - статистика $\chi^2$ может быть использована в обычной форме ( $v = 2{,}5$) при $\chi^2 = 6{,}3$. Расчетные значения $\chi^2$ в случае трех групп длин фазы определяются по формуле (4.4):

$$\chi^2 = \sum_{l=1}^{3} \frac{(N_l - \hat{N}_l)}{N_l} \qquad (4.4)$$

Если $\chi^2 \geq 6{,}3$, то колебания исходного ряда нельзя считать чисто случайными и ряд содержит периодическую составляющую. Этот критерий весьма чувствителен к периодическим колебаниям и имеет практически нулевую эффективность относительно альтернативы наличия тренда, поэтому он может применяться непосредственно к исходному ряду динамики в отличие от других критериев, которые требуют, чтобы из ряда динамики предварительно была выделена систематическая составляющая [100]. После того как установлена периодическая составляющая, проводится ее анализ.

## 4.2. Сезонная составляющая в деятельности фирмы. Индексы сезонности

При рассмотрении квартальных или месячных данных многих социально-экономических явлений часто обнаруживаются определенные, постоянно повторяющиеся колебания, которые существенно не изменяются за длительный период времени. Они являются результатом влияния природно-климатических условий, общих экономических факторов, а также ряда многочисленных разнообразных факторов, которые частично являются регулируемыми. В статистике периодические колебания, которые имеют определенный и постоянный период, равный годовому промежутку, носят название сезонных колебаний, или сезонных волн, а динамический ряд в этом случае называют тренд-сезонным, или просто сезонным рядом динамики.

Классификация наиболее распространенных методов измерения сезонной волны представлена на рис. 4.1.

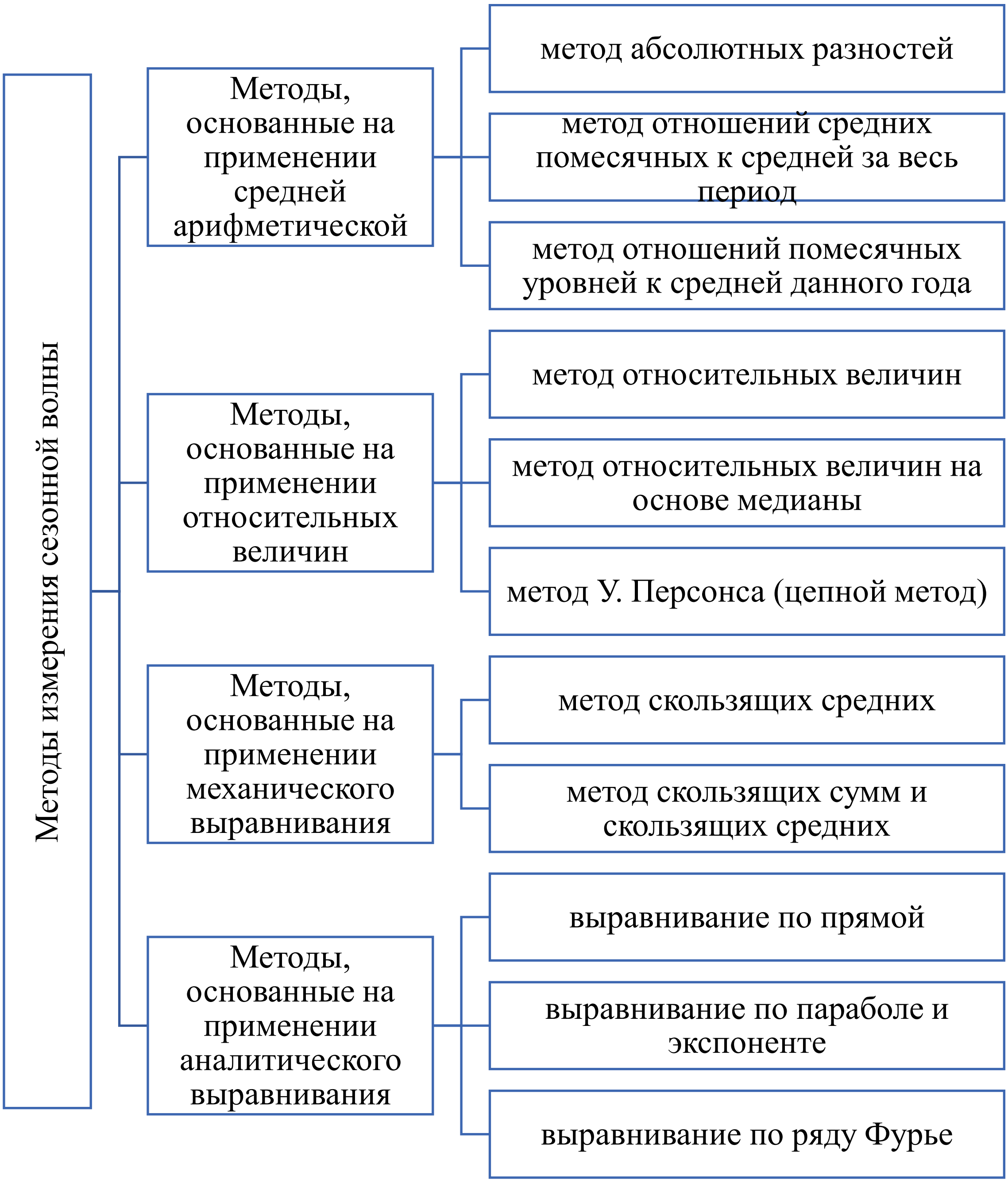



*Рис. 4.1 Методы измерения сезонной волны*

Сезонные колебания характеризуются специальными показателями, которые называются индексами сезонности ($I_S$). Совокупность этих показателей отражает сезонную волну. Индексами сезонности являются процентные отношения фактических внутригодовых уровней к постоянной или переменной средней.

Для выявления сезонных колебаний обычно берут данные за несколько лет, распределенные по месяцам. Данные за несколько лет (не менее трех) используют для того, чтобы выявить устойчивую сезонную волну, которая бы не отражала случайные условия одного года.

Для вычисления индексов сезонности применяют различные методы. Если ряд динамики не содержит ярко выраженной тенденции в развитии, то индексы сезонности вычисляют непосредственно по эмпирическим данным без их предварительного выравнивания (рис. 4.2).


*Рис. 4.2. Схема расчета индексов сезонности в случае, если ряд динамики не содержит ярко выраженной тенденции в развитии*

Таким образом, расчет будет производиться по формуле (4.5):

$$I_S = \frac{\bar{y}_i}{\bar{y}} \cdot 100\% , \qquad (4.5)$$

где $\bar{y}_i$ - средняя величина уровня для каждого одноименного периода

$\bar{y}$ - общее среднее значение за весь период наблюдения.

Если ряд динамики содержит определенную тенденцию в развитии, то прежде чем вычислить сезонную волну фактические данные нужно обработать так, чтобы выявить общую тенденцию. Обычно для этого прибегают к аналитическому выравниванию ряда динамики. Схема, используемая в данном случае, отражена на рис. 4.3.


*Рис. 4.3. Схема расчета индексов сезонности в случае, если ряд динамики содержит тенденцию в развитии*

В общем виде формулу расчета индекса сезонности данным способом можно записать так (4.6):

$$I_S = \left[\sum \frac{y_i}{\bar{y}_t}\right] : n \qquad (4.6)$$

Для сопоставления величины сезонных колебаний по различным территориям или периодам может быть рассчитан коэффициент сезонности, представляющий собой среднеквадратическое отклонение (4.7):

$$k_s = \sigma = \sqrt{\frac{(i_c - 1)^2}{n}}, \qquad (4.6)$$

Где $i_c$ - индекс сезонности для каждого периода (квартала, месяца, дня и т.д.)

n – число таких периодов (кварталов, месяцев, дней и т.д.)

Стоит отметить, что величина $k_s$ прямо пропорциональна величине сезонных колебаний [101]. Соответственно, чем больше данное значение, тем выше величина сезонных колебаний.

## 4.3. Оценка параметров сезонности на основе абсолютных приростов

В п. 2.5. были рассмотрены основные структурные элементы временного ряда. На основе анализа имеющихся данных был сформулирован вывод: в случае, если амплитуда сезонных колебаний с течением времени не меняется, предполагается использование аддитивной модели.

Стоит отметить, что расчет параметров такой модели осуществляется на основе абсолютных отклонений.

Способ построения аддитивной модели опирается на данные о наличии либо отсутствии.

Схема корректировки модели, предполагающей отсутствие тренда, с учетом сезонной составляющей, представлена на рис.4.4.

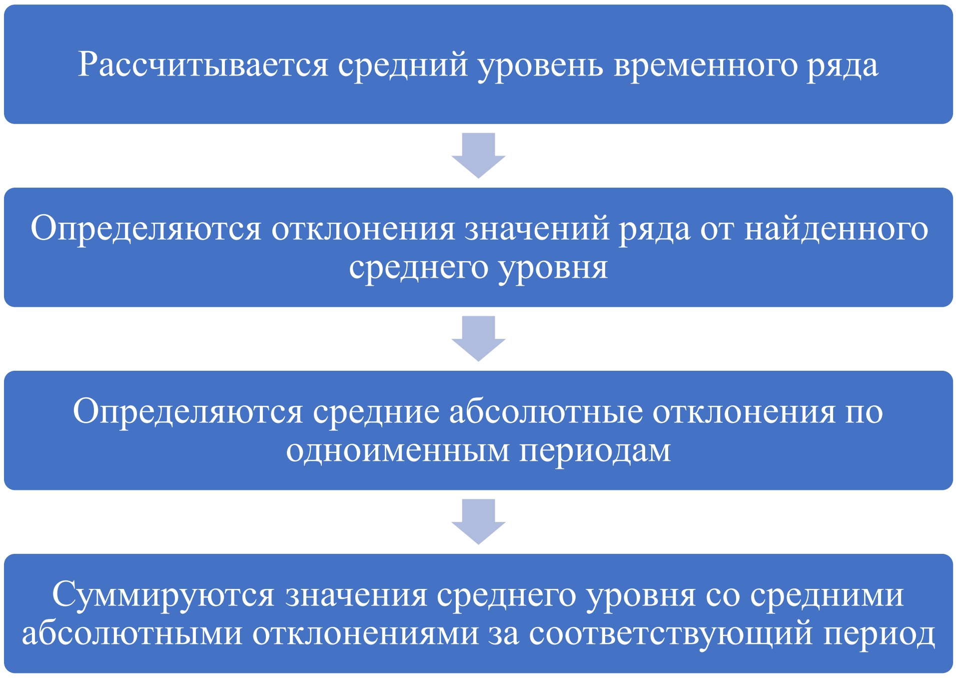



*Рис. 4.4 - Схема корректировки модели, предполагающей отсутствие тренда, с учетом сезонной составляющей*

В случае, если временной ряд содержит тенденцию, построение аддитивной модели с учетом сезонной составляющей осуществляется с использованием теоретических кривых [102] (рис. 4.5).


*Рис. 4.5. Схема корректировки модели, предполагающей отсутствие тренда, с учетом сезонной составляющей*

## Контрольные вопросы

1. В чем состоит суть основных методов выявления периодической компоненты?
2. Назовите методы, основанные на применении средней арифметической.
3. Какие методы измерения сезонной составляющей основаны на применении механического выравнивания?
4. Опишите схему корректировки модели, предполагающей отсутствие тренда, с учетом сезонной составляющей.
5. В чем состоит особенность построения модели с учетом сезонной составляющей, если имеет место наличие тренда?
6. Что такое индекс сезонности?
7. Различается ли расчет коэффициента сезонности для моделей с наличием тренда и без него?
8. С какой целью рассчитываются коэффициенты сезонности? Каков характер связи между данным параметром и величиной сезонных колебаний?
9. Как осуществляется расчет параметров аддитивной модели на основе абсолютных отклонений?

## ПРИМЕРЫ РЕШЕНИЯ ЗАДАЧ

**Пример 4.1.**

В таблице представлены данные о поквартальной выручке некоторого предприятия за трехлетний период:

| Год | Квартал | Объем продаж, млн. руб. |
|---|---|---|
| 2020 | I | 7 |
| | II | 8,7 |
| | III | 11,9 |
| | IV | 28,6 |
| 2021 | I | 11,6 |
| | II | 9,1 |
| | III | 10,7 |
| | IV | 25,9 |
| 2022 | I | 8,6 |
| | II | 8,9 |
| | III | 11 |
| | IV | 23,3 |

Предполагая, что представленный ряд не имеет общей тенденции развития, рассчитайте индексы сезонности, постройте график.

**Решение**

Рассчитаем средний уровень ввода в действие жилых домов по одноименным кварталам:

$$\bar{y}_{I} = \frac{7 + 11{,}6 + 8{,}6}{3} = 9{,}07 \text{ млн. руб.}$$

$$\bar{y}_{II} = \frac{8{,}7 + 9{,}1 + 8{,}9}{3} = 8{,}9 \text{ млн. руб.}$$

$$\bar{y}_{III} = \frac{11{,}9 + 10{,}7 + 11}{3} = 11{,}2 \text{ млн. руб.}$$

$$\bar{y}_{IV} = \frac{28{,}6 + 25{,}9 + 23{,}3}{3} = 25{,}93 \text{ млн. руб.}$$

Среднее значение уровня для всех наблюдений оставит:

$$\bar{y}_{I}=\frac{165{,}3}{12}=13{,}78 \text{ млн. руб.}$$

Для нахождения индексов сезонности построим вспомогательную таблицу:

| Год | Квартал | Объем продаж, млн. руб. | Среднее поквартальное значение, млн. руб. | Индекс сезонности | $i_c$ -1 | $(i_c -1)^2$ |
|---|---|---|---|---|---|---|
| 2020 | I | 7 | 9,07 | 0,66 | -0,34 | 0,12 |
| | II | 8,7 | 8,90 | 0,65 | -0,35 | 0,13 |
| | III | 11,9 | 11,20 | 0,81 | -0,19 | 0,03 |
| | IV | 28,6 | 25,93 | 1,88 | 0,88 | 0,78 |
| 2021 | I | 11,6 | 9,07 | 0,66 | -0,34 | 0,12 |
| | II | 9,1 | 8,90 | 0,65 | -0,35 | 0,13 |
| | III | 10,7 | 11,20 | 0,81 | -0,19 | 0,03 |
| | IV | 25,9 | 25,93 | 1,88 | 0,88 | 0,78 |
| 2022 | I | 8,6 | 9,07 | 0,66 | -0,34 | 0,12 |
| | II | 8,9 | 8,90 | 0,65 | -0,35 | 0,13 |
| | III | 11 | 11,20 | 0,81 | -0,19 | 0,03 |
| | IV | 23,3 | 25,93 | 1,88 | 0,88 | 0,78 |

В формульном виде решение выглядит следующим образом:

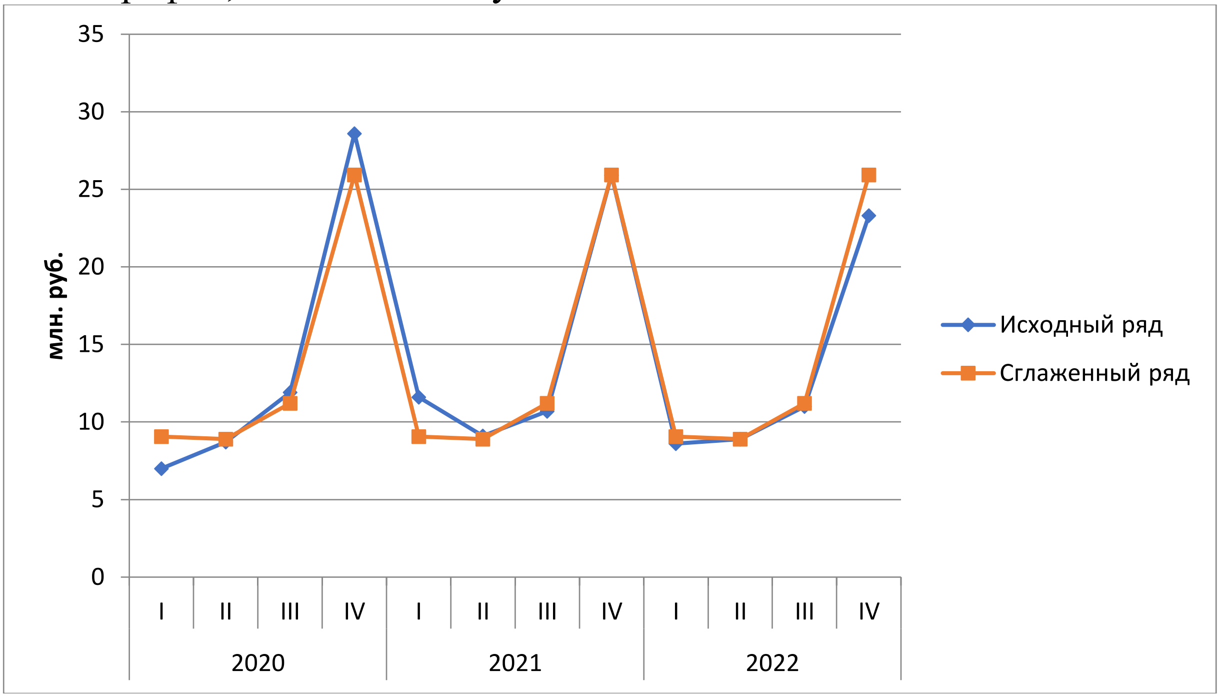
| | A | B | C | D | E | F | G |
|---|---|---|---|---|---|---|---|
| 1 | Год | Квартал | Объем продаж, млн. руб. | Среднее поквартальное значение, млн. руб. | Индекс сезонности | $i_c$ -1 | $(i_c$ -1)^2 |
| 2 | | I | 7 | =(C2+C6+C10)/3 | =D2/СРЗНАЧ($C$2:$C$13) | =E2-1 | =F2^2 |
| 3 | | II | 8,7 | =(C3+C7+C11)/3 | =D3/СРЗНАЧ($C$2:$C$13) | =E3-1 | =F3^2 |
| 4 | | III | 11,9 | =(C4+C8+C12)/3 | =D4/СРЗНАЧ($C$2:$C$13) | =E4-1 | =F4^2 |
| 5 | 2020 | IV | 28,6 | =(C5+C9+C13)/3 | =D5/СРЗНАЧ($C$2:$C$13) | =E5-1 | =F5^2 |
| 6 | | I | 11,6 | =(C2+C6+C10)/3 | =D6/СРЗНАЧ($C$2:$C$13) | =E6-1 | =F6^2 |
| 7 | | II | 9,1 | =(C3+C7+C11)/3 | =D7/СРЗНАЧ($C$2:$C$13) | =E7-1 | =F7^2 |
| 8 | | III | 10,7 | =(C4+C8+C12)/3 | =D8/СРЗНАЧ($C$2:$C$13) | =E8-1 | =F8^2 |
| 9 | 2021 | IV | 25,9 | =(C5+C9+C13)/3 | =D9/СРЗНАЧ($C$2:$C$13) | =E9-1 | =F9^2 |
| 10 | | I | 8,6 | =(C2+C6+C10)/3 | =D10/СРЗНАЧ($C$2:$C$13) | =E10-1 | =F10^2 |
| 11 | | II | 8,9 | =(C3+C7+C11)/3 | =D11/СРЗНАЧ($C$2:$C$13) | =E11-1 | =F11^2 |
| 12 | | III | 11 | =(C4+C8+C12)/3 | =D12/СРЗНАЧ($C$2:$C$13) | =E12-1 | =F12^2 |
| 13 | 2022 | IV | 23,3 | =(C5+C9+C13)/3 | =D13/СРЗНАЧ($C$2:$C$13) | =E13-1 | =F13^2 |

Найденные индексы сезонности свидетельствуют о том, что I, II, III квартале средние значения соответствующих уровней меньше среднегодового значения на 34%, 35% и 19%. Средние значения за IV квартал превышают среднегодовые параметры на 88%.

Для данного временного интервального ряда коэффициент сезонности составит:

$$k_s = \sigma = \sqrt{\frac{1{,}06}{12}} = 0{,}297$$

График, сглаженный с учетом индексов сезонности:




**Пример 4.2.**

В таблице представлены данные о поквартальном изменении цены на некоторый товар:

| Год | Квартал | Цена, руб. |
|---|---|---|
| 2019 | I | 3482 |
| | II | 3253,3 |
| | III | 3120,3 |
| | IV | 4541 |
| 2020 | I | 3721 |
| | II | 3397,7 |
| | III | 3787,3 |
| | IV | 4386,3 |
| 2021 | I | 5176,7 |
| | II | 4554 |
| | III | 3870 |
| | IV | 5053 |
| 2022 | I | 5301,3 |
| | II | 4422 |
| | III | 4134 |
| | IV | 5146,7 |

Предполагая наличие общей тенденции в динамике, рассчитайте индексы сезонности, постройте график.

**Решение**

Для выравнивания примем линейную функцию и перенесём начало координат в середину ряда динамики. Параметры регрессии определим из системы нормальных уравнений.

Для расчета построим вспомогательную таблицу:

| Год | Квартал | Исходные данные | t | $t^2$ | y*t | y^ | $I_{сез}$ | $I_{сез\ ср.}$ | Выравненный уровень с учетом сезонности |
|---|---|---|---|---|---|---|---|---|---|
| | I | 3482 | -15 | 225 | -52230 | 3411,107 | 1,021 | 1,084 | 3697,602 |
| | II | 3253,3 | -13 | 169 | -42292,9 | 3517,515 | 0,925 | 0,938 | 3298,797 |
| | III | 3120,3 | -11 | 121 | -34323,3 | 3623,922 | 0,861 | 0,876 | 3175,073 |
| 2019 | IV | 4541 | -9 | 81 | -40869 | 3730,329 | 1,217 | 1,101 | 4106,619 |
| | I | 3721 | -7 | 49 | -26047 | 3836,737 | 0,970 | 1,084 | 4158,980 |
| | II | 3397,7 | -5 | 25 | -16988,5 | 3943,144 | 0,862 | 0,938 | 3697,961 |
| | III | 3787,3 | -3 | 9 | -11361,9 | 4049,551 | 0,935 | 0,876 | 3547,985 |
| 2020 | IV | 4386,3 | -1 | 1 | -4386,3 | 4155,959 | 1,055 | 1,101 | 4575,183 |
| | I | 5176,7 | 1 | 1 | 5176,7 | 4262,366 | 1,215 | 1,084 | 4620,357 |
| | II | 4554 | 3 | 9 | 13662 | 4368,774 | 1,042 | 0,938 | 4097,125 |
| | III | 3870 | 5 | 25 | 19350 | 4475,181 | 0,865 | 0,876 | 3920,898 |
| 2021 | IV | 5053 | 7 | 49 | 35371 | 4581,588 | 1,103 | 1,101 | 5043,747 |
| | I | 5301,3 | 9 | 81 | 47711,7 | 4687,996 | 1,131 | 1,084 | 5081,735 |
| | II | 4422 | 11 | 121 | 48642 | 4794,403 | 0,922 | 0,938 | 4496,289 |
| | III | 4134 | 13 | 169 | 53742 | 4900,810 | 0,844 | 0,876 | 4293,810 |
| 2022 | IV | 5146,7 | 15 | 225 | 77200,5 | 5007,218 | 1,028 | 1,101 | 5512,311 |

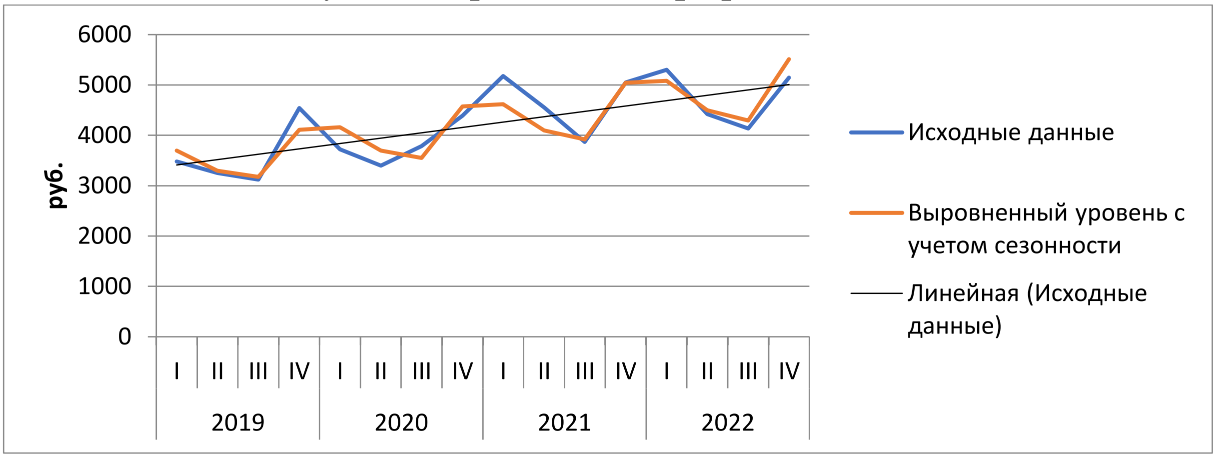
Тогда для заданного ряда параметры составят:

$$\begin{cases} a = \dfrac{\sum\limits_{i=1}^{n} y_i}{n} = 4209{,}16 \\ b = \dfrac{\sum\limits_{i=1}^{n} y_i t_i}{\sum\limits_{i=1}^{n} {t_i}^2} = 53{,}2 \end{cases}$$

Расчеты в формульном виде приведены ниже:

| | A | B | C | D | E | F | G | H | I | J |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Год | Квартал | Исходные данные | t | t^2 | yt | y^ | Iсез | Iсез. Ср. | Выровненный уровень с учетом сезонности |
| 2 | | I | 3482 | -15 | =D2^2 | =C2*D2 | =СРЗНАЧ($C$2:$C$17)+СУММ($F$2:$F$17)/СУММ($E$2:$E$17)*D2 | =C2/G2 | =СРЗНАЧ(H2;H6;H10;H14) | =G2*I2 |
| 3 | | II | 3253,3 | -13 | =D3^2 | =C3*D3 | =СРЗНАЧ($C$2:$C$17)+СУММ($F$2:$F$17)/СУММ($E$2:$E$17)*D3 | =C3/G3 | =СРЗНАЧ(H3;H7;H11;H15) | =G3*I3 |
| 4 | | III | 3120,3 | -11 | =D4^2 | =C4*D4 | =СРЗНАЧ($C$2:$C$17)+СУММ($F$2:$F$17)/СУММ($E$2:$E$17)*D4 | =C4/G4 | =СРЗНАЧ(H4;H8;H12;H16) | =G4*I4 |
| 5 | 2019 | IV | 4541 | -9 | =D5^2 | =C5*D5 | =СРЗНАЧ($C$2:$C$17)+СУММ($F$2:$F$17)/СУММ($E$2:$E$17)*D5 | =C5/G5 | =СРЗНАЧ(H5;H9;H13;H17) | =G5*I5 |
| 6 | | I | 3721 | -7 | =D6^2 | =C6*D6 | =СРЗНАЧ($C$2:$C$17)+СУММ($F$2:$F$17)/СУММ($E$2:$E$17)*D6 | =C6/G6 | =СРЗНАЧ(H2;H6;H10;H14) | =G6*I6 |
| 7 | | II | 3397,7 | -5 | =D7^2 | =C7*D7 | =СРЗНАЧ($C$2:$C$17)+СУММ($F$2:$F$17)/СУММ($E$2:$E$17)*D7 | =C7/G7 | =СРЗНАЧ(H3;H7;H11;H15) | =G7*I7 |
| 8 | | III | 3787,3 | -3 | =D8^2 | =C8*D8 | =СРЗНАЧ($C$2:$C$17)+СУММ($F$2:$F$17)/СУММ($E$2:$E$17)*D8 | =C8/G8 | =СРЗНАЧ(H4;H8;H12;H16) | =G8*I8 |
| 9 | 2020 | IV | 4386,3 | -1 | =D9^2 | =C9*D9 | =СРЗНАЧ($C$2:$C$17)+СУММ($F$2:$F$17)/СУММ($E$2:$E$17)*D9 | =C9/G9 | =СРЗНАЧ(H5;H9;H13;H17) | =G9*I9 |
| 10 | | I | 5176,7 | 1 | =D10^2 | =C10*D10 | =СРЗНАЧ($C$2:$C$17)+СУММ($F$2:$F$17)/СУММ($E$2:$E$17)*D10 | =C10/G10 | =СРЗНАЧ(H2;H6;H10;H14) | =G10*I10 |
| 11 | | II | 4554 | 3 | =D11^2 | =C11*D11 | =СРЗНАЧ($C$2:$C$17)+СУММ($F$2:$F$17)/СУММ($E$2:$E$17)*D11 | =C11/G11 | =СРЗНАЧ(H3;H7;H11;H15) | =G11*I11 |
| 12 | | III | 3870 | 5 | =D12^2 | =C12*D12 | =СРЗНАЧ($C$2:$C$17)+СУММ($F$2:$F$17)/СУММ($E$2:$E$17)*D12 | =C12/G12 | =СРЗНАЧ(H4;H8;H12;H16) | =G12*I12 |
| 13 | 2021 | IV | 5053 | 7 | =D13^2 | =C13*D13 | =СРЗНАЧ($C$2:$C$17)+СУММ($F$2:$F$17)/СУММ($E$2:$E$17)*D13 | =C13/G13 | =СРЗНАЧ(H5;H9;H13;H17) | =G13*I13 |
| 14 | | I | 5301,3 | 9 | =D14^2 | =C14*D14 | =СРЗНАЧ($C$2:$C$17)+СУММ($F$2:$F$17)/СУММ($E$2:$E$17)*D14 | =C14/G14 | =СРЗНАЧ(H2;H6;H10;H14) | =G14*I14 |
| 15 | | II | 4422 | 11 | =D15^2 | =C15*D15 | =СРЗНАЧ($C$2:$C$17)+СУММ($F$2:$F$17)/СУММ($E$2:$E$17)*D15 | =C15/G15 | =СРЗНАЧ(H3;H7;H11;H15) | =G15*I15 |
| 16 | | III | 4134 | 13 | =D16^2 | =C16*D16 | =СРЗНАЧ($C$2:$C$17)+СУММ($F$2:$F$17)/СУММ($E$2:$E$17)*D16 | =C16/G16 | =СРЗНАЧ(H4;H8;H12;H16) | =G16*I16 |
| 17 | 2022 | IV | 5146,7 | 15 | =D17^2 | =C17*D17 | =СРЗНАЧ($C$2:$C$17)+СУММ($F$2:$F$17)/СУММ($E$2:$E$17)*D17 | =C17/G17 | =СРЗНАЧ(H5;H9;H13;H17) | =G17*I17 |

Результаты расчетов в графическом виде:




По результатам проведенных расчетов можно сделать вывод о том, что относительно линии тренда разброс индексов сезонности меньше, чем по отношению к среднегодовому изменению цен на некотором предприятии за четыре года.

**Пример 4.3.**

Имеются данные об изменении значений временного ряда:

| Год | Квартал | y |
|---|---|---|
| 2020 | I | 24 |
| | II | 25,7 |
| | III | 28,9 |
| | IV | 45,6 |
| 2021 | I | 28,6 |
| | II | 26,1 |
| | III | 27,7 |
| | IV | 42,9 |
| 2022 | I | 25,6 |
| | II | 25,9 |
| | III | 28 |
| | IV | 40,3 |

Предполагая отсутствие тренда, скорректировать ряд динамики на сезонную составляющую.

**Решение.**

Для выполнения поставленной задачи составим вспомогательную таблицу:

| Год | Квартал | y | Абсолютное отклонение от среднего уровня, Δ | Среднее абсолютное отклонение от среднегодового уровня, Δ ср. | Выравненный уровень с учетом сезонности |
|---|---|---|---|---|---|
| 2020 | I | 24 | -6,775 | -4,71 | 26,07 |
| | II | 25,7 | -5,075 | -4,88 | 25,90 |
| | III | 28,9 | -1,875 | -2,58 | 28,20 |
| | IV | 45,6 | 14,825 | 12,16 | 42,93 |
| 2021 | I | 28,6 | -2,175 | -4,71 | 26,07 |
| | II | 26,1 | -4,675 | -4,88 | 25,90 |
| | III | 27,7 | -3,075 | -2,58 | 28,20 |
| | IV | 42,9 | 12,125 | 12,16 | 42,93 |
| 2022 | I | 25,6 | -5,175 | -4,71 | 26,07 |
| | II | 25,9 | -4,875 | -4,88 | 25,90 |
| | III | 28 | -2,775 | -2,58 | 28,20 |
| | IV | 40,3 | 9,525 | 12,16 | 42,93 |

Для вычислений найдем средний уровень ряда:

$$\bar{y} = \frac{\sum_{t=1}^{n} y_t}{n} = 30{,}78$$

Средние поквартальные отклонения от среднегодового уровня составят:

$$\overline{\Delta}_{I} = \frac{-6{,}78 + (-2{,}175) + (-5{,}178)}{3} = -4{,}71$$

$$\overline{\Delta}_{II} = \frac{-5{,}075 + (-4{,}675) + (-4{,}878)}{3} = -4{,}88$$

$$\overline{\Delta}_{III} = \frac{-1{,}875 + (-3{,}075) + (-2{,}775)}{3} = -2{,}58$$

$$\overline{\Delta}_{IV} = \frac{14{,}825 + 12{,}125 + 9{,}525}{3} = 12{,}16$$

Таким образом, поквартальные данные выравненные с учетом сезонности:

$$\overline{y_{\mathrm{I}}} = 30{,}78 + (-4{,}71)$$

$$\overline{y_{\mathrm{II}}} = 30{,}78 + (-4{,}88) = 25{,}9$$

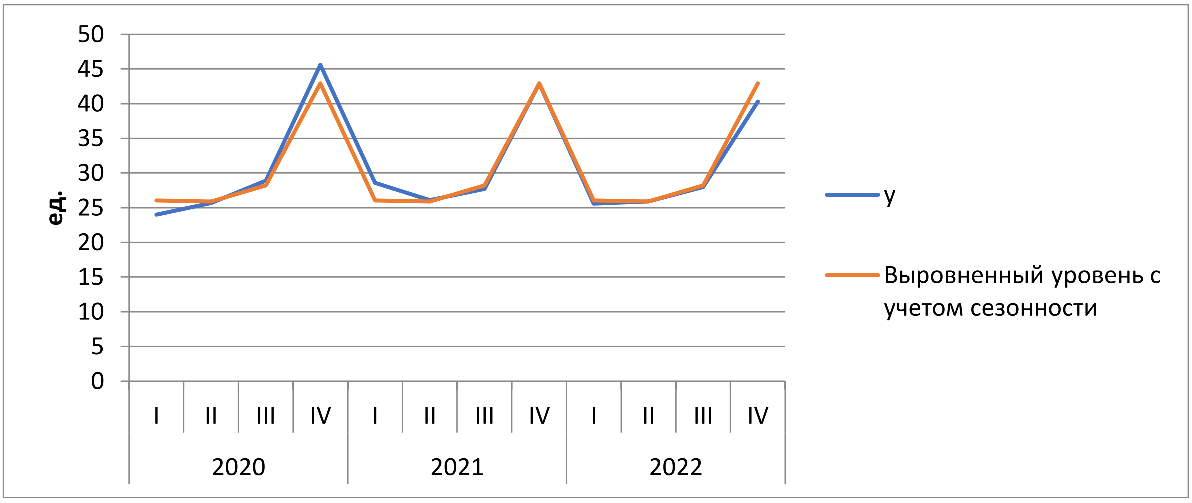
$$\overline{y_{\mathrm{III}}} = 30{,}78 + (-2{,}58) = 28{,}2$$

$$\overline{y_{\mathrm{IV}}} = 30{,}78 + 12{,}16 = 42{,}93$$

Вычисления в формульном виде:

| | A | B | C | D | E | F |
|---|---|---|---|---|---|---|
| 1 | Год | Квартал | y | Абсолютное отклонение от среднего уровня, Δ | Среднее абсолютное отклонение от среднегдового уровня | Выровненный уровень с учетом сезонности |
| 2 | | I | 24 | =C2-СРЗНАЧ($C$2:$C$13) | =СРЗНАЧ(D2;D6;D10) | =СРЗНАЧ($C$2:$C$13)+E2 |
| 3 | | II | 25,7 | =C3-СРЗНАЧ($C$2:$C$13) | =СРЗНАЧ(D3;D7;D11) | =СРЗНАЧ($C$2:$C$13)+E3 |
| 4 | | III | 28,9 | =C4-СРЗНАЧ($C$2:$C$13) | =СРЗНАЧ(D4;D8;D12) | =СРЗНАЧ($C$2:$C$13)+E4 |
| 5 | 2020 | IV | 45,6 | =C5-СРЗНАЧ($C$2:$C$13) | =СРЗНАЧ(D5;D9;D13) | =СРЗНАЧ($C$2:$C$13)+E5 |
| 6 | | I | 28,6 | =C6-СРЗНАЧ($C$2:$C$13) | =СРЗНАЧ(D2;D6;D10) | =СРЗНАЧ($C$2:$C$13)+E6 |
| 7 | | II | 26,1 | =C7-СРЗНАЧ($C$2:$C$13) | =СРЗНАЧ(D3;D7;D11) | =СРЗНАЧ($C$2:$C$13)+E7 |
| 8 | | III | 27,7 | =C8-СРЗНАЧ($C$2:$C$13) | =СРЗНАЧ(D4;D8;D12) | =СРЗНАЧ($C$2:$C$13)+E8 |
| 9 | 2021 | IV | 42,9 | =C9-СРЗНАЧ($C$2:$C$13) | =СРЗНАЧ(D5;D9;D13) | =СРЗНАЧ($C$2:$C$13)+E9 |
| 10 | | I | 25,6 | =C10-СРЗНАЧ($C$2:$C$13) | =СРЗНАЧ(D2;D6;D10) | =СРЗНАЧ($C$2:$C$13)+E10 |
| 11 | | II | 25,9 | =C11-СРЗНАЧ($C$2:$C$13) | =СРЗНАЧ(D3;D7;D11) | =СРЗНАЧ($C$2:$C$13)+E11 |
| 12 | | III | 28 | =C12-СРЗНАЧ($C$2:$C$13) | =СРЗНАЧ(D4;D8;D12) | =СРЗНАЧ($C$2:$C$13)+E12 |
| 13 | 2022 | IV | 40,3 | =C13-СРЗНАЧ($C$2:$C$13) | =СРЗНАЧ(D5;D9;D13) | =СРЗНАЧ($C$2:$C$13)+E13 |

Динамика сезонных колебаний отражена на графике:



# Глава 5. КОРРЕЛЯЦИЯ И РЕГРЕССИЯ

## 5.1. Регрессионный анализ связанных временных рядов

Многомерные временные ряды, показывающие зависимость результативного признака от одного или нескольких факторных, называют связными рядами динамики. Применение методов наименьших квадратов для обработки рядов динамики не требует выдвижения никаких предположений о законах распределения исходных данных.

Однако при использовании метода наименьших квадратов для обработки связных рядов следует учитывать наличие автокорреляции (авторегрессии), которая не учитывалась при обработке одномерных рядов динамики, поскольку ее наличие способствовало более плотному и четкому выявлению тенденции развития рассматриваемого социально-экономического явления во времени.

Выявление автокорреляции в уровнях ряда динамики. В рядах динамики экономических процессов между уровнями, особенно близко расположенными, существует взаимосвязь.

Ее удобно представить в виде корреляционной зависимости между рядами $y_1$, $y_2$, $y_3$, …, $y_n$ и этим же рядом, сдвинутым относительно первоначального положения на k моментов времени $y_{1+h}$, $y_{2+h}$, $y_{3+h}$, …, $y_{n+h}$. Временное смещение $L$ называется сдвигом, а само явление взаимосвязи - автокорреляцией.

Автокорреляционная зависимость особенно существенна между последующими и предшествующими уровнями ряда динамики. Поскольку классические методы математической статистики применимы лишь в случае независимости отдельных членов ряда между собой, то при анализе нескольких взаимосвязанных рядов динамики важно установить наличие и степень их автокорреляции (рис.5.1).

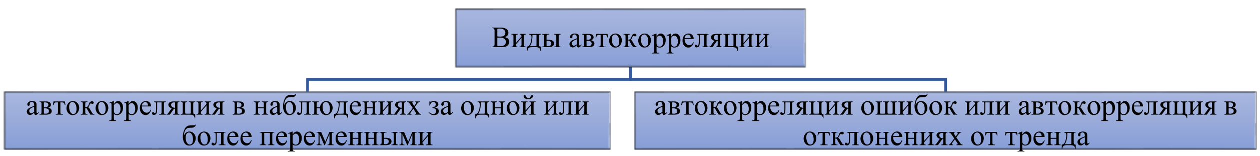



*Рис. 5.1. Виды автокорреляции*

Наличие последней приводит к искажению величин средних квадратических ошибок коэффициентов регрессии, что затрудняет построение доверительных интервалов для коэффициентов регрессии, а также проверку их значимости.

Автокорреляцию измеряют при помощи нециклического коэффициента автокорреляции, который может рассчитываться не только между соседними уровнями, т.е. сдвинутыми на один период, но и между сдвинутыми на любое число единиц времени (L). Этот сдвиг, именуемый временным лагом, определяет и порядок коэффициентов автокорреляции: первого порядка (при $L = 1$), второго порядка (при $L = 2$) и т.д. Однако наибольший интерес для исследования представляет вычисление нециклического коэффициента (первого порядка), так как наиболее сильные искажения результатов анализа возникают при корреляции между исходными уровнями ряда ($y_{t1}$) и теми же уровнями, сдвинутыми на одну единицу времени, т.е. $y_{t-1}$ (или $y_{t+1}$).

Тогда формулу коэффициента автокорреляции можно записать следующим образом (5.1):

$$r_a = \frac{\overline{y_t \cdot y_{t+1}} - \bar{y}_t \cdot \bar{y}_{t+1}}{\sigma_{y_t} \cdot \sigma_{y_{t+1}}} \tag{5.1}$$

Если значение последнего уровня ($y_n$) ряда мало отличается от первого ($y_1$), то сдвинутый ряд не укорачивается, его можно условно дополнить, приняв $y_n = y_1$. Тогда $y_t = y_{t+1}$ и $\sigma_{y_t} = \sigma_{y_{t+1}}$, поскольку рассчитываются они для одного и того же ряда. При такой замене, т.е. если $\bar{y}_t = \bar{y}_{t+1}$ и $\sigma_{y_t} = \sigma_{y_{t+1}}$ формула коэффициента автокорреляции примет вид (5.2 или 5.3):

$$r_a = \frac{\overline{y_t \cdot y_{t+1}} - (\bar{y}_t)^2}{\sigma^2_{y_t}} \tag{5.2}$$

$$r_a = \frac{\sum y_t \cdot y_{t+1} - n \cdot (\bar{y}_t)^2}{\sum y_t^2 - n \cdot (\bar{y}_t)^2} \tag{5.3}$$

Если ряд динамики состоит из уровней, среднее значение которых равно нулю ($\bar{y} = 0$), то выражение (5.3) значительно упрощается (5.4):

$$r_a = \frac{\sum_{t=1}^{n-1} y_t \cdot y_{t+1}}{\sum_{t=1}^{n} y_t^2} \tag{5.4}$$

Для суждения о наличии или отсутствии автокорреляции в исследуемом ряду фактическое значение коэффициентов автокорреляции сопоставляется с табличным (критическим) для 5%-го или 1%-го уровня значимости (вероятность допустить ошибку при принятии нулевой гипотезы о независимости уровней ряда).

Одна из специальных таблиц, в которой определена критическая область проверяемой гипотезы (об отсутствии автокорреляции), составленная Р. Андерсеном в 1942 г., приведена в приложении 4.

Если фактическое значение коэффициента автокорреляции меньше табличного, то гипотеза об отсутствии автокорреляции в ряду может быть принята. Когда же фактическое значение больше табличного, можно сделать вывод о наличии автокорреляции в ряду динамики.

Одним из способов выявления автокорреляции в отклонениях от тренда или от регрессионной модели является использование критерия Дарбина-Уотсона, который рассчитывается по формуле (5.5):

$$d = \frac{\sum_{t=1}^{n} (l_{t+1} - l_t)^2}{\sum_{t=1}^{n} l_t^2} \tag{5.5}$$

Теоретическое основание применения этого критерия обусловлено тем, что в динамических рядах как сами наблюдения, так и отклонения от них распределяются в хронологическом порядке.

При условии, что отклонения уровней от тенденции (так называемые остатки) случайны, значения $D$, лежащие в интервале 0 - 4, всегда будут находиться ближе к 2. Если автокорреляция положительная, то $D < 2$; отрицательная $-2 \le D \le 4$. Следовательно, оценки, получаемые по критерию, являются не точечными, а интервальными. Их значения для трех уровней значимости ($\alpha = 0{,}01, \alpha = 0{,}025$ и $\alpha = 0{,}05$) с учетом числа наблюдений даны в специальных таблицах (приложение 5)

Способы исключения или уменьшения автокорреляции (авторегрессии) в рядах динамики приведены на рис. 5.2.

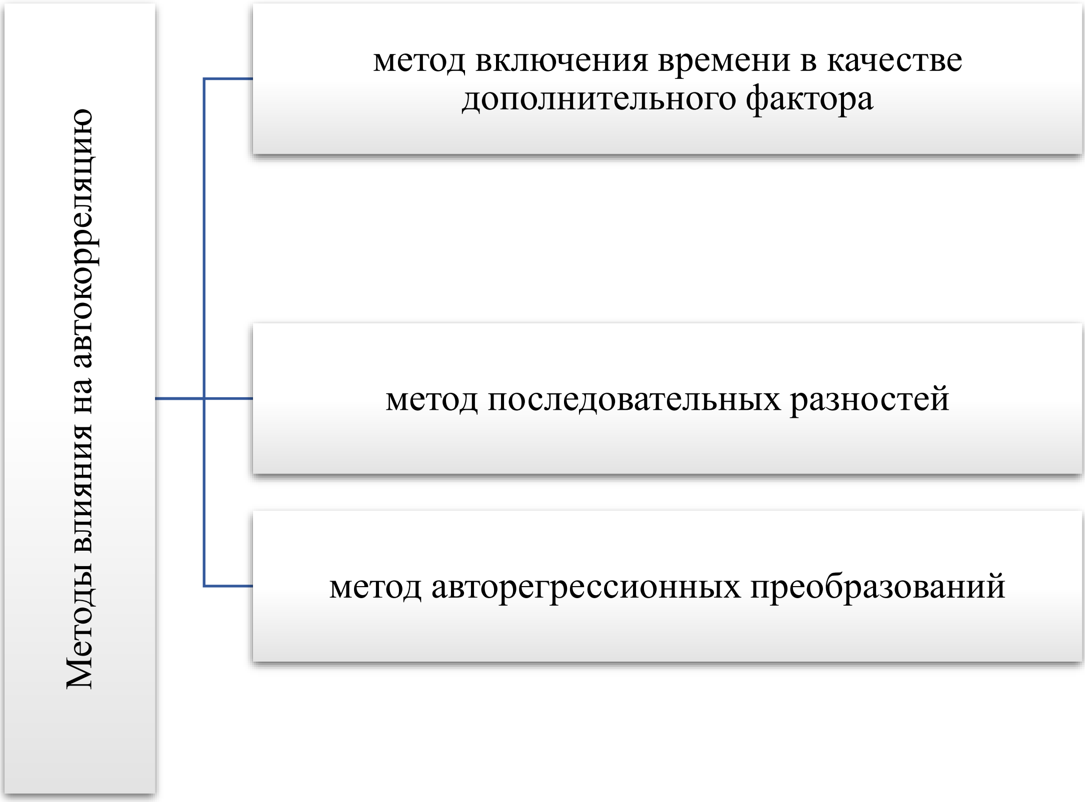



*Рис. 5.2. Способы исключения или уменьшения авторегрессии*

Рассмотрим суть **метода включения времени в качестве дополнительного фактора**. В соответствии с теоремой, доказанной Фришем и Воу, время вводится в систему связных динамических рядов в явной форме в качестве дополнительного фактора, и эта процедура называется введением фактора времени в уравнение регрессии. Уровни исходных динамических рядов могут быть представлены показателями в любой форме, в том числе логарифмической, а время всегда вводится в линейной форме. Считается, что введение фактора времени исключает основную тенденцию развития всех явлений, представленных исследуемыми рядами динамики. Доказано, что введение времени аналогично использованию отклонения фактических данных от трендов.

Применение метода наименьших квадратов к обработке многомерных временных рядов не отличается от методологии применения его к обычным статистическим рядам. В рассматриваемом случае минимизируется следующее выражение (5.6):

$$S = \sum\left[y_i - f\left(x_1, x_2,..., x_n, t\right)\right]^2 \rightarrow \min \qquad (5.6)$$

При исключении автокорреляции **методом последовательных разностей** подвергаются обработке методом наименьших квадратов не сами уровни исходных рядов $y_t, y_{t+1},..., y_{t+n}$ и $x_t, x_{t+1},..., x_{t+n}$ а последовательные разности между ними:

| | |
|---|---|
| $\Delta y_1 = y_t - y_{t-1};$ | $\Delta x_1 = x_t - x_{t-1};$ |
| $\Delta y_2 = y_{t-1} - y_{t-2};$ | $\Delta x_2 = x_{t-1} - x_{t-2};$ |
| ………………. | ………………. |
| ………………. | ………………. |
| $\Delta y_k = y_{t-k} - y_{t-k-1};$ | $\Delta x_k = x_{t-k} - x_{t-k-1}.$ |

При использовании этого метода исходят из предположения, что все разности между уровнями динамических рядов, начиная с первой, будут содержать только случайную компоненту. Причем первые разности содержат случайную компоненту в линейной форме, вторые - описываемую параболой 2-го порядка, третьи – показательной функцией.

**Метод авторегрессионных преобразований** заключается в том, что определяют уравнение связи между отклонениями от тенденций двух связных рядов динамики:

| | |
|---|---|
| $y_1 - \bar{y}_{t1};$ | $x_1 - \bar{x}_{t1};$ |
| $y_2 - \bar{y}_{t2};$ | $x_2 - \bar{x}_{t2};$ |
| ………………. | ………………. |
| ………………. | ………………. |
| $y_n - \bar{y}_{tn};$ | $x_n - \bar{x}_{tn}.$ |

В этом случае также получают уравнения регрессии, не искаженные влиянием автокорреляции.

Введение времени в качестве дополнительной переменной является наиболее действенным способом обработки связных рядов динамики. Во всяком случае, при линейной связи между исследуемыми рядами этот способ более точен, чем использование последовательных разностей или отклонений от трендов.

При обработке методом наименьших квадратов последовательных разностей или отклонений от трендов исследователь имеет дело с чисто случайными величинами, взаимосвязь между которыми является часто весьма сомнительной, так как исключение в обоих случаях тенденций нарушает существование причинно-следственной связи между явлениями.

## 5.2. Корреляция временных рядов

При изучении развития явления во времени часто возникает необходимость оценить степень взаимосвязи в изменениях уровней двух или более рядов динамики различного содержания, но связанных между собой.

Эта задача решается методами коррелирования (рис. 5.3).


*Рис. 5.3. Методы коррелирования*

**Расчет парного коэффициента корреляции по уровням ряда динамики**. Этот расчет правильно показывает тесноту связи между рядами динамики лишь в том случае, если в каждом из них отсутствует автокорреляция.

В этом случае величину коэффициента корреляции находят по формуле (5.6):

$$r = \frac{\overline{x \cdot y} - \bar{x} \cdot \bar{y}}{\sigma_x \cdot \sigma_y} \qquad (5.6)$$

где $x_i$ - уровни факторного ряда динамики

$y_i$ - уровни результативного ряда динамики.

Следовательно, прежде чем коррелировать ряды динамики (по уровням), необходимо проверить каждый из рядов на наличие или отсутствие в них автокорреляции (при помощи коэффициента автокорреляции). В случае наличия автокорреляции между уровнями ряда последняя должна быть устранена.

**Расчет парного коэффициента корреляции по отклонениям фактических уровней от выравненных по уравнению (тренду)**. Этот способ состоит в том, что коррелируют не сами уровни, а отклонения фактических уровней от выравненных, отражающих тренд, т.е. коррелируют остаточные величины. Для этого каждый ряд динамики выравнивают по определенной, характерной для него аналитической формуле, затем из эмпирических уровней вычитают выравненные (т.е. находят $d_x = x_t - \bar{x}_t$; $d_y = y_t - \bar{y}_t$) и определяют тесноту связи между рассчитанными отклонениями ($d_x$ и $d_y$) по формуле (5.7):

$$r_{d_x d_y} = \frac{\sum d_x \cdot d_y}{\sqrt{\sum d_x^2 \cdot \sum d_y^2}} \qquad (5.7)$$

Расчет парного коэффициента корреляции по абсолютным отклонениям уровней ряда динамики. Исключить влияние автокорреляции можно путем вычитания из каждого уровня предшествующего ему, т.е. находя разности уровней ($y_i - y_{i-1}$). Алгебраически легко показать, что при переходе от уровней к их разностям исключается влияние общей тенденции на колеблемость. При этом при изменении уровней по прямой можно коррелировать первые разности, при изменении по параболе n-го порядка - n-е разности. Формула коэффициента разностей, используемая для измерения тесноты связи между исследуемыми рядами, имеет вид (5.8):

$$r_{\Delta_x \Delta_y} = \frac{\sum \Delta_x \cdot \Delta_y}{\sqrt{\sum \Delta_x^2 \cdot \sum \Delta_y^2}} \quad (5.8)$$

Коэффициент корреляции, рассчитанный для измерения тесноты зависимости изменения уровней двух рядов, является своего рода средним, обобщающим показателем. Однако для длительного периода эта зависимость не является постоянной, она может меняться во времени. Поэтому чтобы судить о том, в какие периоды зависимость между изменениями уровней двух рядов слабая или сильная, рекомендуется рассчитывать серию скользящих коэффициентов корреляции для определенного интервала времени.

## Контрольные вопросы

1. Какие ряды динамики называются связанными?
2. Требует ли применение методов наименьших квадратов для обработки рядов динамики выдвижение предположений о законах распределения исходных данных?
3. В чем состоит суть регрессии?
4. Что означает автокорреляционная зависимость?
5. В каком случае автокорреляционная зависимость особенно существенна?
6. Какие существуют виды автокорреляции?
7. Как автокорреляция оказывает влияние на результаты регреммионного моделирования?
8. С помощью каких инстурментов может быть рассчитана автокорреляция?
9. Для чего используется таблица Р. Андерсена?
10. В чем состоит сущность статистики Дарбина-Уотсона?

11. В чем состоит теоретическое основание применения статистики Дарбина-Уотсона?

12. Каким образом можно оказывать влияние на автокорреляцию?

13. В чем состоит суть метода включения времени в качестве дополнительного фактора?

14. Что представляет собой метод авторегрессионных преобразований?

15. Отличается ли процедура применения метода наименьших квадратов к обработке многомерных временных рядов по сравнению с обычными статистическими рядами?

16. Какие существуют методы коррелирования?

17. Каким образом производится расчет парного коэффициента корреляции по отклонениям фактических уровней от выравненных по уравнению?

## ПРИМЕРЫ РЕШЕНИЯ ЗАДАЧ

Пример 5.1

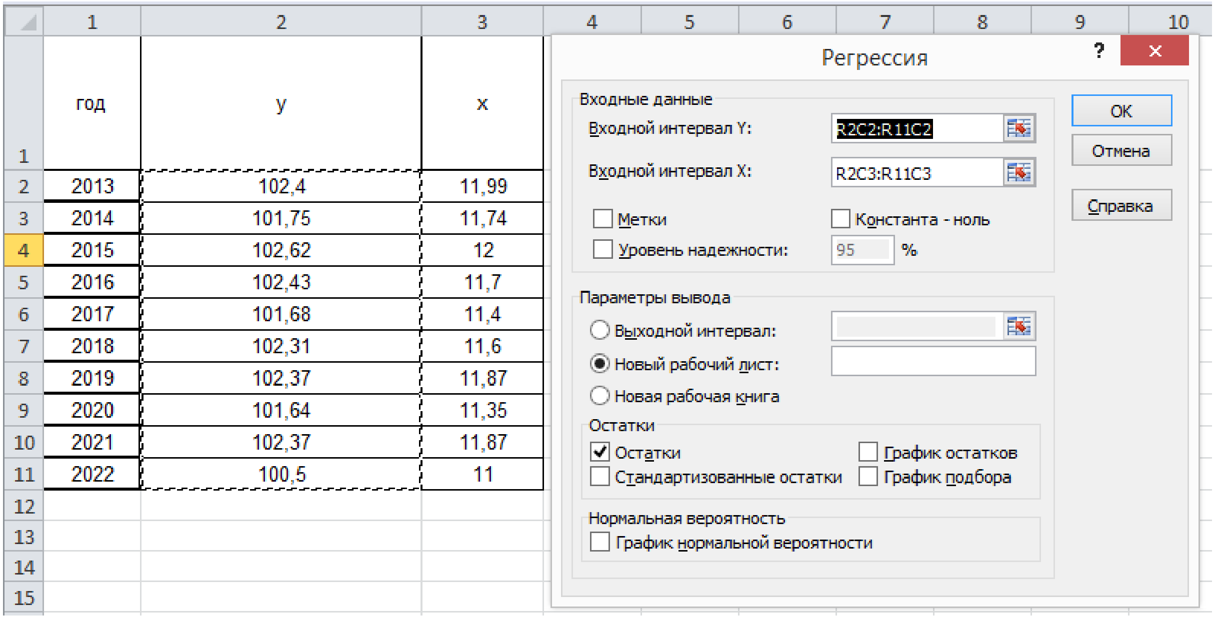
Имеются данные о некоторых независимых переменных и результирующем факторе.

| год | y | x |
|---|---|---|
| 2013 | 102,4 | 11,99 |
| 2014 | 101,75 | 11,74 |
| 2015 | 102,62 | 12 |
| 2016 | 102,43 | 11,7 |
| 2017 | 101,68 | 11,4 |
| 2018 | 102,31 | 11,6 |
| 2019 | 102,37 | 11,87 |
| 2020 | 101,64 | 11,35 |
| 2021 | 102,37 | 11,87 |
| 2022 | 100,5 | 11 |

На основании представленных данных с помощью Ms Excel проверить наличие автокорреляции в линейной модели, описывающей изменение результирующего признака под влиянием независимой переменной.

**Решение.**

Для построения модели воспользуемся инструментом «Регрессия» в пакете «Анализ данных» (вкладка «Данные»). Также отметим поле «остатки» для их автоматизированного расчета:




В результате вывода остатков получим:

| | | | |
|---|---|---|---|
| 22 | ВЫВОД ОСТАТКА | | |
| 23 | | | |
| 24 | *Наблюдение* | *Предсказанное Y* | *Остатки* |
| 25 | 1 | 102,6157047 | -0,215704689 |
| 26 | 2 | 102,1654793 | -0,415479327 |
| 27 | 3 | 102,6337137 | -0,013713704 |
| 28 | 4 | 102,0934433 | 0,336556731 |
| 29 | 5 | 101,5531728 | 0,126827165 |
| 30 | 6 | 101,9133531 | 0,396646875 |
| 31 | 7 | 102,3995965 | -0,029596516 |
| 32 | 8 | 101,4631278 | 0,176872237 |
| 33 | 9 | 102,3995965 | -0,029596516 |
| 34 | 10 | 100,8328123 | -0,332812256 |

Для расчета статистики Дарбина-Уотсона воспользуемся вспомогательной таблицей:

| $t$ | $l_{t-1}$ | $l_t$ | $l_t^2$ | $(l_t-l_{t-1})^2$ |
|---|---|---|---|---|
| 1 | | -0,2157 | 0,0465 | 0,0465 |
| 2 | -0,2157 | -0,4155 | 0,1726 | 0,0399 |
| 3 | -0,4155 | -0,0137 | 0,0002 | 0,1614 |
| 4 | -0,0137 | 0,3366 | 0,1133 | 0,1227 |
| 5 | 0,3366 | 0,1268 | 0,0161 | 0,0440 |
| 6 | 0,1268 | 0,3966 | 0,1573 | 0,0728 |
| 7 | 0,3966 | -0,0296 | 0,0009 | 0,1817 |
| 8 | -0,0296 | 0,1769 | 0,0313 | 0,0426 |
| 9 | 0,1769 | -0,0296 | 0,0009 | 0,0426 |
| 10 | -0,0296 | -0,3328 | 0,1108 | 0,0919 |
| $\sum$ | | | 0,6498 | 0,8462 |

Таким образом,

$$d = \frac{0,8462}{0,6498} = 1,302$$

Не обращаясь к таблице критических точек Дарбина-Уотсона, можно пользоваться «грубым» правилом и считать, что автокорреляция остатков отсутствует, если $1,5 < d < 2,5$. Таким образом, данное условие соблюдается, автокорреляция остатков отсутствует.

# Глава 6. ПРОГНОЗИРОВАНИЕ ДЕЯТЕЛЬНОСТИ ФИРМЫ С ИСПОЛЬЗОВАНИЕМ СТАТИСТИЧЕСКИХ МЕТОДОВ

## 6.1. Сущность экстраполяции как метода прогнозирования

Исследование динамики социально-экономических явлений, выявление и характеристика основной тенденции развития и моделей взаимосвязи дают основание для прогнозирования - определения будущих размеров уровня экономического явления.

Важное место в системе методов прогнозирования занимают статистические методы. Применение прогнозирования предполагает, что закономерность развития, действующая в прошлом (внутри ряда динамики), сохранится и в прогнозируемом будущем, т.е. прогноз основан на экстраполяции. Экстраполяция, проводимая в будущее, называется перспективной и в прошлое - ретроспективной. Обычно, говоря об экстраполяции рядов динамики, подразумевают чаще всего перспективную экстраполяцию.

Теоретической основой распространения тенденции на будущее является известное свойство социально-экономических явлений, называемое инерционностью. Именно инерционность позволяет выявить сложившиеся взаимосвязи как между уровнями динамического ряда, так и между группой связных рядов динамики. На основе рядов динамики получаются весьма надежные прогнозы, если уровни ряда динамики сопоставимы и получены на основе единой методологии.

Применение экстраполяции в прогнозировании базируется на следующих предпосылках:

- развитие исследуемого явления в целом описывается плавной кривой;
- общая тенденция развития явления в прошлом и настоящем не претерпит серьезных изменений в будущем.

Поэтому надежность и точность прогноза зависят от того, насколько близкими к действительности окажутся эти предположения, а также как точно удастся охарактеризовать выявленную в прошлом закономерность. Экстраполяцию следует рассматривать как начальную стадию построения окончательных прогнозов. Механическое использование экстраполяции может стать причиной погрешности и неправильных выводов. Всегда следует учитывать все необходимые условия, предпосылки и гипотезы, связывая их с содержательным экономико-теоретическим анализом.

Разумеется, чем шире раздвигаются временные рамки прогнозирования, тем очевиднее становится недостаточность простого экстраполяционного метода (изменение тенденций, неизвестны точки поворота кривых, влияние новых факторов и т.д.). В этом случае динамичность экономических явлений и процессов вступает в противоречие с инерционностью их развития. Так как анализируемые экономические ряды динамики нередко относительно короткие, то и временной горизонт экстраполяции не может быть бесконечным. Поэтому чем короче срок экстраполяции (период упреждения), тем более надежные и точные результаты (при прочих равных условиях) дает прогноз. За короткий период не успевают сильно измениться условия развития явления и характер его динамики.

Экстраполяцию в общем виде можно представить формулой (6.1):

$$\hat{y}_{i+T} = f\left(y_i, T, a_j\right), \tag{6.1}$$

где $\hat{y}_{i+T}$ – прогнозируемый уровень;

$y_i$ – текущий уровень прогнозируемого ряда;

$T$ – период упреждения;

$a_j$ – параметр уравнения тренда.

В зависимости от того, какие принципы и исходные данные положены в основу прогноза, выделяют следующие элементарные методы экстраполяции:

- средний абсолютный прирост;
- средний темп роста;
- экстраполяцию на основе выравнивания рядов по какой-либо аналитической формуле.

Прогнозирование по среднему абсолютному приросту может быть выполнено в том случае, если есть уверенность считать общую тенденцию линейной, т.е. метод основан на предположении о равномерном изменении уровня (под равномерностью понимается стабильность абсолютных приростов).

Для нахождения интересующего нас аналитического выражения тенденции на любую дату t необходимо определить средний абсолютный прирост и последовательно прибавлять его к последнему уровню ряда столько раз, на сколько периодов экстраполируется ряд, т.е. экстраполяцию можно сделать по следующей формуле (6.2):

$$\hat{y}_{i+t} = y_i + \overline{\Delta} \cdot t, \tag{6.2}$$

где $\hat{y}_{i+t}$ – экстраполируемый уровень, $(i+1)$ - номер этого уровня (года);

$i$ – номер последнего уровня (года) исследуемого периода, за который рассчитан $\overline{\Delta}$;

$t$ – срок прогноза (период упреждения);

$\overline{\Delta}$ – средний абсолютный прирост.

Однако следует иметь в виду, что использование среднего абсолютного прироста для прогноза возможно только при следующем условии (6.3):

$$\sigma^2_{ост.} \le \rho^2, \tag{6.3}$$

$$\rho^2 = \frac{1}{2} \cdot \frac{\sum \Delta_i^2}{n}$$

$$\sigma^2_{ост.} = \frac{\sum \left(y_i - \overline{y}_{\overline{\Delta}}\right)^2}{n}$$

Прогнозирование по среднему темпу роста осуществляется в случае, когда есть основание считать, что общая тенденция ряда характеризуется показательной (экспоненциальной) кривой. Для нахождения тенденции необходимо определить средний коэффициент роста, возведенный в степень, соответствующую периоду экстраполяции, т.е. по формуле (6.4):

$$\hat{y}_{i+t} = y_i \cdot \overline{k}_p^{\,t}, \tag{6.4}$$

где $y_i$ – последний уровень ряда динамики;

$t$ – срок прогноза;

$\overline{k}_p$ – средний коэффициент роста.

Если же ряду динамики свойственна иная закономерность, то данные, полученные при экстраполяции на основе среднего темпа роста, будут отличаться от данных, рассчитанных другими способами экстраполяции.

Рассмотренные способы экстраполяции тренда, будучи простейшими, в то же время являются и самыми приближенными.

Экстраполяция дает возможность получить точечное значение прогноза. Точное совпадение фактических данных и прогностических точечных оценок, полученных путем экстраполяции кривых, характеризующих тенденцию, имеет малую вероятность. Возникновение отклонений фактических уровней ряда динамики от выравненных по уравнению тренда объясняется следующими причинами:

- выбранная для прогнозирования кривая не является единственно возможной для описания тенденции. Можно подобрать такую кривую, которая дает более точные результаты;
- построение прогноза осуществляется на основе ограниченного числа исходных данных. Кроме того, каждый исходный уровень обладает еще случайной компонентой. Поэтому и кривая, по которой осуществляется экстраполяция, будет содержать случайную компоненту;
- тенденция характеризует лишь движение среднего уровня ряда динамики, поэтому отдельные наблюдения от него отклоняются. Если такие отклонения наблюдались в прошлом, то они будут наблюдаться и в будущем. Любой статистический прогноз носит приближенный характер. Поэтому целесообразно определение доверительных интервалов прогноза.

Величина доверительного интервала определяется следующим образом (6.5):

$$\hat{y}_t \pm t_\alpha \cdot \sigma_{\bar{y}_t} \qquad (6.5)$$

где $\sigma_{\bar{y}_t}$ – средняя квадратическая ошибка тренда;

$\hat{y}_t$ – расчетное значение уровня;

$t_\alpha$ – доверительная величина.

При анализе рядов динамики иногда приходится прибегать к определению некоторых неизвестных уровней внутри данного ряда динамики, т.е. к интерполяции.

Интерполяция может производиться на основе среднего абсолютного прироста, среднего темпа роста и с помощью аналитического выравнивания. Она также основана на том или ином предположении о тенденции изменения уровней, но характер этого прогноза несколько иной: здесь уже не приходится предполагать, что тенденция, характерная для прошлого, сохранится и в будущем [103 – 106].

При интерполяции считается, что ни выявленная тенденция, ни ее характер не претерпели существенных изменений в том промежутке времени, уровень (уровни) которого нам не известен. Такое предположение обычно является более обоснованным, чем предположение о будущей тенденции.

## 6.2. Прогнозирование на основе аналитического выравнивания тренда

Наиболее распространенным методом прогнозирования считают аналитическое выражение тренда. При этом для выхода за границы исследуемого периода достаточно продолжить значения независимой переменной времени (t).

При таком подходе к прогнозированию предполагается, что размер уровня, характеризующего явление, формируется род воздействием множества факторов, причем не представляется возможным выделить отдельно их влияние. В связи с этим ход развития связывается не с какими-либо конкретными факторами, а с течением времени, т.е. $y = f(t)$.

Процедура разработки прогноза с использованием кривых роста включает в себя следующие этапы, отраженные на рис. 6.1.

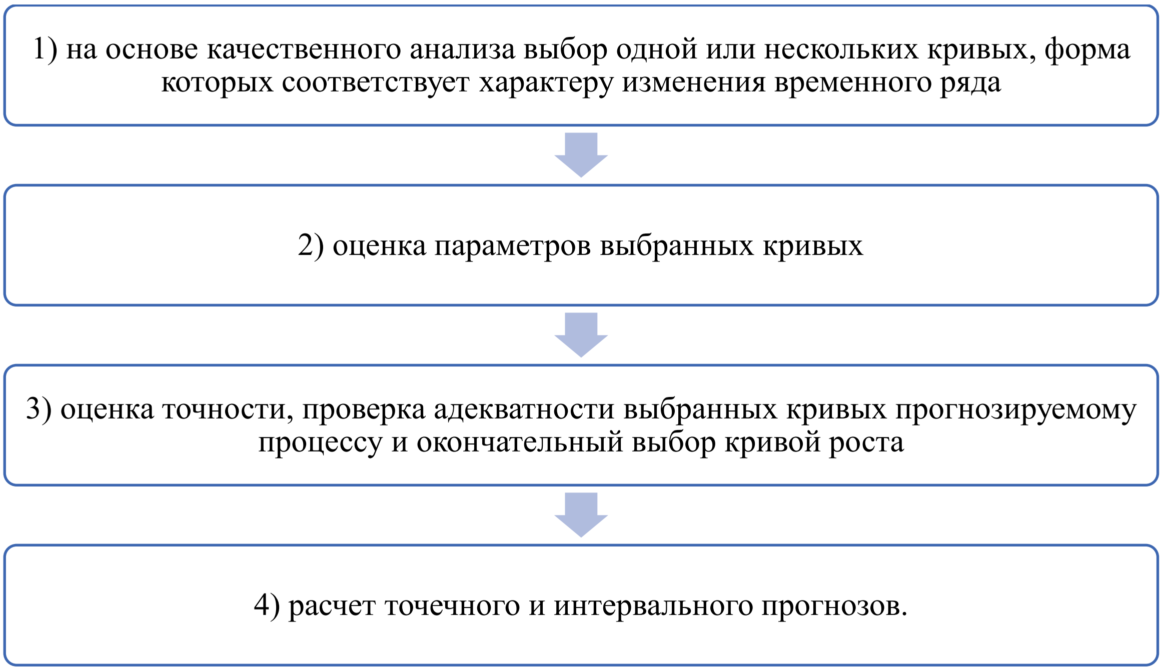



*Рис. 6.1. Схема прогнозирования с использованием кривых роста*

Независимо от метода построения моделей их качество оценивается на основе исследования свойств остаточной компоненты E(t), t=1,2,...,N, т.е. величины расхождений на участке аппроксимации (построения модели) между фактическими уровнями и их расчетными значениями.

Качество модели определяется ее адекватностью исследуемому процессу и точностью. Адекватность характеризуется наличием определенных статистических свойств, а точность - степенью близости к фактическим данным.

Модель считается адекватной, если ряд остатков обладает свойствами: независимости уровней, их случайности, соответствия нормальному закону распределения и равенства нулю средней ошибки.

При проверке независимости (отсутствии автокорреляции) определяется отсутствие в ряду остатков систематической составляющей.

Точечный прогноз заключается в получении прогнозного значения результирующего признака, которое определяется путем подстановки в уравнение регрессии соответствующего значения независимой переменной.

Интервальный прогноз заключается в построении доверительного интервала прогноза, т. е. нижней и верхней границ интервала, содержащего точную величину для прогнозного значения с заданной вероятностью

При построении доверительного интервала прогноза используется стандартная ошибка прогноза [104] (6.6):

$$M_{\hat{y}_p} = \sqrt{\frac{\Sigma(y-\hat{y})^2}{n-m-1}} * \sqrt{1+\frac{1}{n}+\frac{(x_p-\bar{x})^2}{\Sigma(x-\bar{x})^2}} \qquad (6.6)$$

Строится доверительный интервал прогноза (6.7):

$$\gamma_{\hat{y}_p} = \hat{y}_p \pm t_{табл} . M_{\hat{y}_p} \qquad (6.7)$$

## 6.3. Количественная оценка достоверности прогноза развития фирмы

Количественно степень достоверности прогноза характеризуется показателями оправдываемости (степень соответствия прогнозов фактическим условиям) и предсказуемости (устойчивость прогноза с определенной вероятностью по отношению к ошибкам измерения), а также ошибками 1-го и 2-го рода.

Оправдываемость прогноза определяется по формуле (6.8):

$$O = \frac{N_{наст.пр.}}{N_{пр}} \qquad (6.8)$$

где O – показатель опрадываемости прогноза;

$N_{наст.пр.}$- число наступивших событий, по которым дан прогноз;

$N_{пр}$ – число событий, по которым дан прогноз.

Предсказуемость прогноза определяется по формуле:

$$П = \frac{N_{наст.пр.}}{N_{наст.}} 100\% \qquad (6.9)$$

где П – показатель предсказуемости прогноза;

$N_{наст.}$ – число наступивших событий

Если событие было предсказано, но не наступило, то имеет место ошибка 1-го рода α, которая определяется по формуле (6.10):

$$\alpha = \frac{N_{\text{пр.}} - N_{\text{наст.пр.}}}{N_{\text{пр.}}} 100\% \qquad (6.10)$$

Если событие не было предсказано, но наступило, то имеет место ошибка 2-го рода β, которая определяется по формуле (6.11):

$$\beta = \frac{N_{\text{наст.}} - N_{\text{наст.пр.}}}{N_{\text{наст.}}} 100\% \qquad (6.11)$$

Таким образом, для оценки степени достоверности прогноза развития фирмы используются различные количественные параметры. Данный этап оценки результатов прогнозирования необходим для того, чтобы скорректировать планово-прогнозные мероприятия в организации для достижения более достоверных результатов, учитывающих совокупность внешних факторов в управлении фирмой.

## Контрольные вопросы

1. В каких случаях возможно прогнозирование на основе экстраполяции?
2. Что такое интерполяция?
3. Какие методы экстраполяции Вам известны?
4. Каким образом осуществляется прогнозирование на основе аналитического выравнивания тренда?
5. Опишите схему прогнозирования с использованием кривых роста
6. Что такое адекватность как один из признаков качественной модели?
7. Как осуществляется точечное прогнозирование?
8. Что составляет основу интервального прогнозирования?
9. Как вычисляется стандартная ошибка прогноза?
10. Что представляет собой доверительный интервал прогнозирования?
11. Какие параметры используются для количественной оценки достоверности прогноза?

# ПРИМЕРЫ РЕШЕНИЯ ЗАДАЧ

## Пример 6.1

В таблице представлены данные об изменении выручки фирмы за восьмилетний период:

| Год | 2016 | 2017 | 2018 | 2019 | 2020 | 2021 | 2022 | 2023 |
|---|---|---|---|---|---|---|---|---|
| Выручка, млн. руб. | 5,2 | 6,1 | 6,4 | 6,5 | 7,1 | 7,3 | 7,4 | 7,8 |

На основании представленных данных, рассчитайте прогнозное значение выручки организации на основе экстраполяции по сложившемуся среднегодовому темпу роста в 2024 году.

**Решение.**

Найдем цепные коэффициента роста для заданного временного ряда:

| y | $k_ц$ |
|---|---|
| 5,2 | 1 |
| 6,1 | 1,173 |
| 6,4 | 1,049 |
| 6,5 | 1,016 |
| 7,1 | 1,092 |
| 7,3 | 1,028 |
| 7,4 | 1,014 |
| 7,8 | 1,054 |

Для нахождения среднегодового темпа роста воспользуемся формулой:

$$T_{ср.г.} = \sqrt[n-1]{k_{ц1} * k_{ц2} * k_{ц3} * .... * k_{цn}} * 100\% = \sqrt[8-1]{1{,}5} * 100\% = 105{,}96\%$$

Аналогичные результаты вычисления будут получены при использовании в расчетах базисного темпа роста:

$$T_{ср.г.} = \sqrt[n-1]{\frac{y_n}{y_1}} * 100\% = \sqrt[8-1]{\frac{7,8}{5,2}} * 100\% = 105,96\%$$

Таким образом, можно сделать вывод о том, что за исследуемый промежуток времени наблюдается рост выручки фирмы в среднем на 5,96%. Положительное значение темпа прироста свидетельствует о том, что при сохранении существующей тенденции в будущем ожидается рост анализируемого параметра.

Тогда ожидаемое значение выручки в 2024 году составит:

$$y_{n+1} = y_n * k_{ср.г.} = 7,8 * 1,0596 = 8,27 \text{ (млн. руб.)}$$

**Пример 6.2**

В таблице представлены данные о численности рабочих некоторого предприятия и его выручке

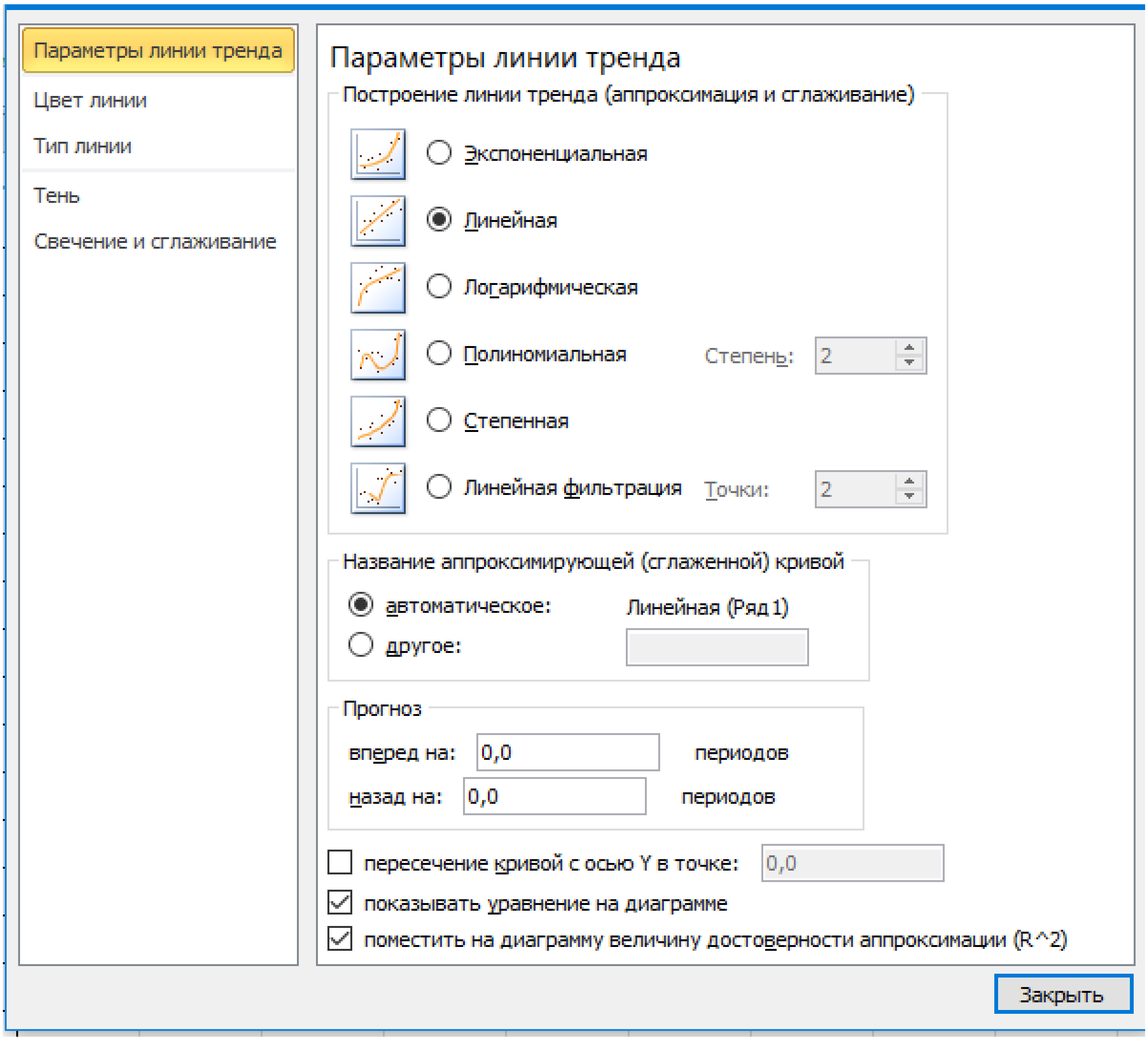
| Год | Численность рабочих, чел. | Выручка, тыс. руб. |
|---|---|---|
| 2005 | 137 151 | 76 933 |
| 2006 | 145 777 | 87 306 |
| 2007 | 153 765 | 105 369 |
| 2008 | 160 722 | 111 705 |
| 2009 | 164 065 | 126 816 |
| 2010 | 176 773 | 159 299 |
| 2011 | 160 101 | 162 175 |
| 2012 | 180 071 | 169 152 |
| 2013 | 188 362 | 197 073 |
| 2014 | 189 236 | 186 445 |
| 2015 | 193 251 | 204 354 |
| 2016 | 212 412 | 223 439 |
| 2017 | 202 765 | 234 112 |
| 2018 | 202 307 | 235 553 |
| 2019 | 222 830 | 239 092 |
| 2020 | 230 918 | 246 917 |
| 2021 | 239 191 | 257 214 |
| 2022 | 238 699 | 285 091 |

На основании приведенных данных постройте модель влияния численности рабочих на величину выручки предприятия, оцените значимость модели и коэффициентов полученного уравнения, описывающего влияние независимой переменной на результирующий признак. Рассчитайте ожидаемое значение выручки при расширении масштабов производства с использованием труда 245 000 рабочих.

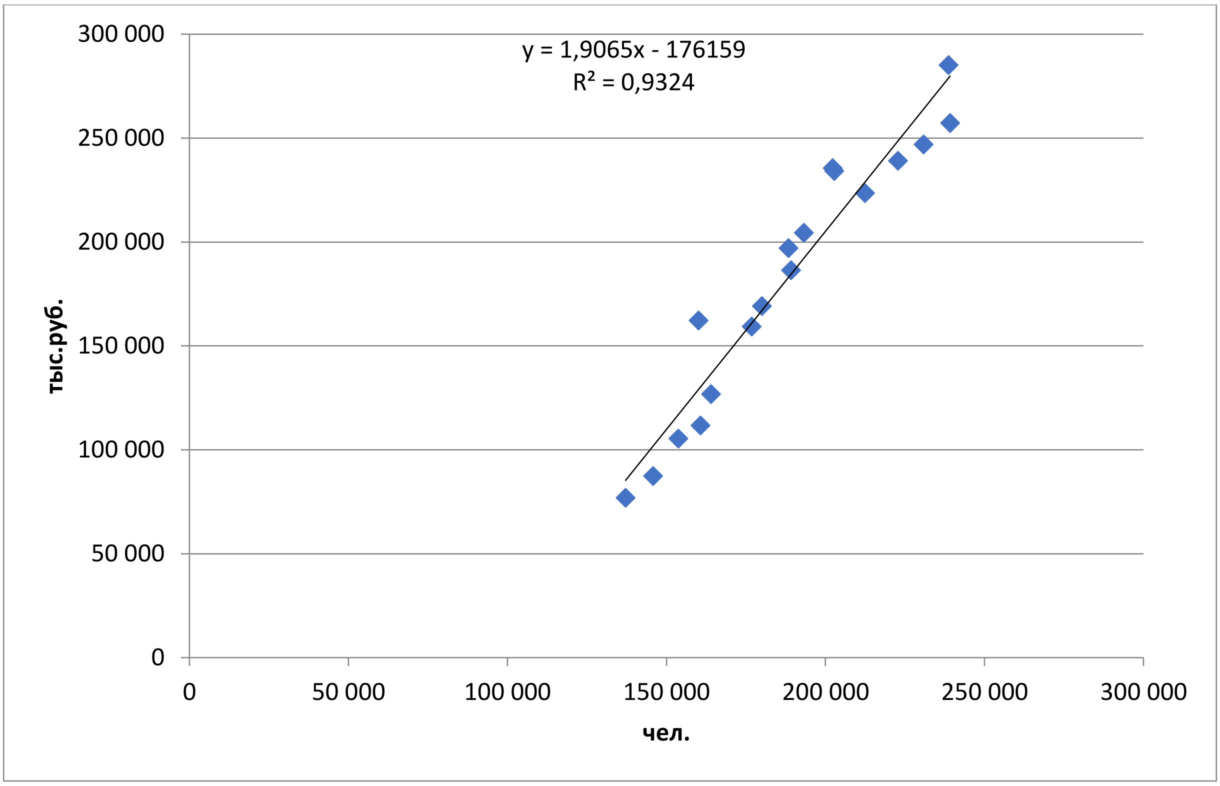
**Решение**

Для решения задачи предварительно определим характер влияния независимой переменной на результирующий признак.

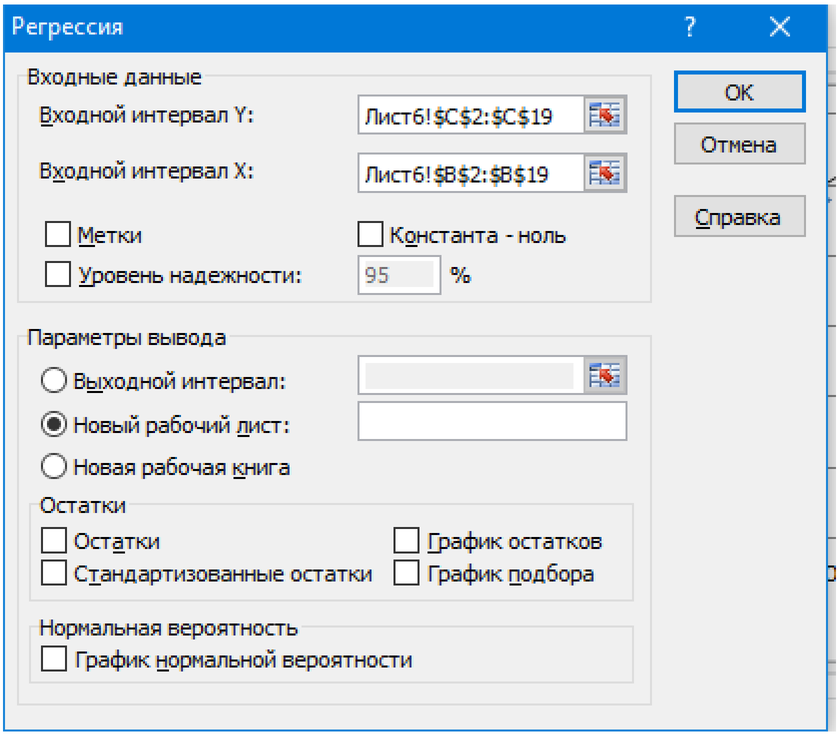
Для этого построим точечную диаграмму. По графику можно предположить, что имеет место линейная зависимость, поэтому добавим линейный тренд:




В результате получим график вида:




Для оценки характеристик линейного уравнения воспользуемся пакетом «Анализ данных» (пункт «Регрессия»):



В результате была получены следующие результаты регрессионного моделирования:

| | A | B | C | D | E | F | G | H | I |
|---|---|---|---|---|---|---|---|---|---|
| 1 | ВЫВОД ИТОГОВ | | | | | | | | |
| 2 | | | | | | | | | |
| 3 | *Регрессионная статистика* | | | | | | | | |
| 4 | Множеств | 0,965591263 | | | | | | | |
| 5 | R-квадрат | 0,932366486 | | | | | | | |
| 6 | Нормиро | 0,928139392 | | | | | | | |
| 7 | Стандарт | 16789,70528 | | | | | | | |
| 8 | Наблюде | 18 | | | | | | | |
| 9 | | | | | | | | | |
| 10 | Дисперсионный анализ | | | | | | | | |
| 11 | | *df* | *SS* | *MS* | *F* | *Значимость F* | | | |
| 12 | Регрессия | 1 | 62177153037 | 62177153037 | 220,5691081 | 8,86882E-11 | | | |
| 13 | Остаток | 16 | 4510307256 | 281894203,5 | | | | | |
| 14 | Итого | 17 | 66687460293 | | | | | | |
| 15 | | | | | | | | | |
| 16 | | *Коэффициенты* | *Стандартная ошибка* | *t-статистика* | *P-Значение* | *Нижние 95%* | *Верхние 95%* | *Нижние 95,0%* | *Верхние 95,0%* |
| 17 | Y-пересеч | -176159,3721 | 24556,76734 | -7,173557076 | 2,21435E-06 | -228217,3933 | -124101,3509 | -228217,3933 | -124101,3509 |
| 18 | Перемен | 1,906462519 | 0,12836775 | 14,85156921 | 8,86882E-11 | 1,634335046 | 2,178589992 | 1,634335046 | 2,178589992 |

Коэффициент детерминации составил 93,24 %, что свидетельствует о том, что изменение результирующей переменной на 93,24 % объясняется изменением независимой переменной.

Модель в целом адекватна, так как значимость F-критерия Фишера меньше пятипроцентной величины. Коэффициенты линейного уравнения также значимы, так как р-значение меньше 5%.

Таким образом, уравнение, описывающее изменение выручки под влиянием изменения численности рабочих:

$$y = 1{,}91x - 176159$$

Для нахождения прогнозного значения выручки подставим в уравнение ожидаемое значение независимой переменной:

$$y = 1{,}91 * 245000 - 176159 = 291791 \text{ (тыс. руб.)}$$

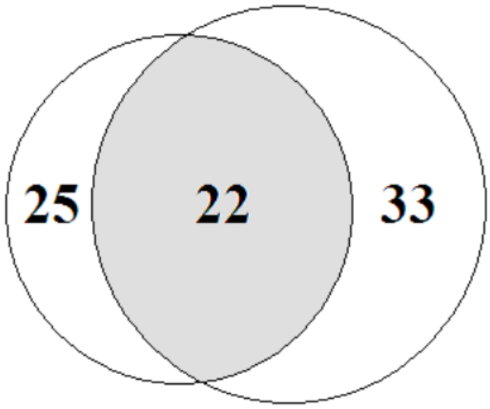
Таким образом, при расширении масштабов производства с увеличением численности рабочих до 245 000 человек можно ожидать величину выручки 291791 тыс. руб.

**Пример 6.3**

Необходимо предоставить руководителю фирмы отчет о достоверности прогнозов во втором полугодии 2023 года, если за анализируемый период прогнозировали появление 25 новых угроз, а в итоге системой мониторинга был о обнаружено 33 новые угрозы, причем 22 из них совпали с прогнозами специалистов.

**Решение**

Для упрощения расчетов и наглядного представления исходных данных, изобразим исходные условия для решения задачи с помощью кругов Эйлера.




Таким образом, $N_{\text{пр.}} = 25, N_{\text{наст.}} = 33, N_{\text{наст.пр.}} = 22$.

Для заданных исходных данных оправдываемость прогноза может быть определена следующим образом:

$$\text{О} = \frac{22}{25} 100\% = 88\%$$

Полученный результат свидетельствует о том, что с точки зрения выявленных угроз среди прогнозируемых ранее, прогноз оправдан на 88%.

Определим предсказуемость прогноза:

$$\Pi = \frac{22}{33} 100\% = 67\%$$

Поскольку из 33 выявленных угроз только 22 были указаны в прогнозе, можно сделать вывод о том, что предсказуемость составляет только 67%.

Определим ошибку 1-го рода:

$$\alpha = \frac{25 - 22}{25} 100\% = 12\%$$

Полученные результат свидетельствует о том, что 12% угроз в плане были указаны ошибочно и компания понесла необоснованные затраты на их выявление, прогнозирование, оценку и защиту от них.

Определим, какой процент новых угроз не был отражен в прогнозных показателях фирмы:

$$\beta = \frac{33 - 22}{33} 100\% = 33\%$$

Ошибка второго рода свидетельствует о том, что около трети угроз фирмы не были спрогнозированы ранее. Таким образом, организацией не были предприняты меры для их минимизации, что могло стать причиной причинения определенного ущерба компании.

# Глава 7. ОСНОВЫ КЛАСТЕРНОГО АНАЛИЗА

## 7.1. Введение в кластерный анализ

Деятельность фирмы в современных условиях сопряжена с огромным количеством внутренних и внешних факторов, оказывающих влияние на вектор ее развития. Принятие управленческих решений в условиях неполной и стохастической информации обуславливает необходимость применения инструментов многомерного анализа, в частности, кластеризации.

Сущность многомерных статистических методов в деятельности фирмы состоит в том, чтобы выбрать такую вероятностную модель, которая бы наилучшим образом описывала реальное изменение показателей деятельности организации.

По своей сути кластерный анализ представляет собой процесс исследования, с помощью которого набор каких-либо объектов (исследуемых параметров) объединяется или группируется в относительно небольшие группы, именуемые кластерами, которые имеют черты внутригруппового сходства и набор четко выраженных отличий с другими кластерами.

Применение для решения задач кластерного анализа пакетов прикладных статистических программ значительно упрощает исследование, поскольку перегруппировка кластеров при изменении, добавлении или исключении какого-либо признака осуществляется без необходимости проводить дополнительные массивные вычисления вручную.

Наиболее распространенными узкоспециализированными пакетами для проведения различных видов статистического анализа являются STADIA, STATGRAPHICS, SPSS, STATISTICA. Все перечисленные программные продукты обладают широким набором статистических функций, которые применяются для решения самого широкого спектра исследовательских задач. Более того, применение вышеназванных пакетов возможно не только в экономических, но и в различных социологических, географических, биологических и других исследованиях. Все это еще раз доказывает высокую значимость статистических методов в различных областях профессиональной деятельности.

При выборе статистического пакета для решения конкретной прикладной задачи следует руководствоваться следующими положениями.

Во-первых, программный продукт должен обладать статистическим разнообразием для эффективной работы с нужным массивом исходных данных. Для работы с информацией в зависимости от целей исследования используются различные статистические инструменты, грамотный выбор которых является необходимым условием решения конкретной задачи пользователя.

Во-вторых, статистический пакет должен обладать широкими графическими возможностями. В частности, в ряде программ имеют место встроенные графические редакторы, существует опция демонстрации отдельных элементов графика. Также имеется возможность экспорта полученных графических материалов в текстовые документы, презентации и т.д., что создает дополнительные условия представления полученных результатов исследования не только пользователям, осуществляющим работу в узкоспециализированных программах, но и широкому кругу лиц, заинтересованных в доступном и понятном представлении результатов статистических исследований.

В–третьих, важно, чтобы каждому конкретному пользователю было удобно работать в выбранном статистическом пакете, осуществлять экспорт/импорт данных, а также их реструктуризацию. Вопрос эргономичности интерфейса является достаточно субъективным, поэтому человеку, осуществляющему обработку статистической информации, необходимо руководствоваться собственными предпочтениями и взглядами, а не популярностью или распространенностью продукта в широком доступе.

Впервые кластерный анализ как метод исследования нашел свое применение в социологии. Само слово «cluster» происходит от английского определения понятий «скопление» и «гроздь». Родоначальником кластерного исследования можно назвать Роберта Триона, который в 1939 году выпустил труд, систематизирующий представления о предмете, сущности и базовых методах кластеризации. Сущность кластерного анализа, вне зависимости от области его применения, можно определить следующим образом: имеет место некоторое число объектов, которое требуется разбить на подмножества, отличающиеся максимальным сходством внутри исследуемых групп (кластеров) и существенными различиями между самими кластерами. По сути кластерный анализ представляет собой метод классификации, предполагающий исследование структуры образованных групп.

Основное преимущество кластеризации как метода состоит в том, что выявление групп происходит не по какому-либо одному признаку, а по их совокупности, что позволят проводить всестороннее изучение анализируемых объектов.

Также стоит отметить, что данный математико-статистический метод никак не ограничивает область рассматриваемых объектов, поскольку применим для данных практически любой природы. Данное свойство является наиболее ценным, когда анализируются и исследуются разнообразные признаки. Применение иных эконометрических методов и приемов в таком случае может быть крайне затруднительным.

Применение кластерного анализа позволяет существенно сократить объем данных, сохраняя при этом существенные свойства и признаки выделенных массивов данных. Кроме того, использование данного метода возможно в сочетании с другими качественными методами эконометрики.

Также использование кластеризации активно используется при анализе рядов динамики, позволяя выделять временные интервалы с общими характеристиками.

При проведении кластерного анализа есть возможность применять данный метод циклически, дополняя каждый цикл анализа уточняющей информацией.

Следует отметить, что кластерный анализ имеет также ряд недостатков и ограничений. В частности, качество кластерного анализа во многом обусловлено тщательностью отбора критериев разбиения. Также сведение массива исходных данных к объединениям внутри кластерных групп приводит к некоторым искажениям и отсутствию учета индивидуальных признаков. При кластеризации важно, чтобы соблюдался целый ряд условий, основные их которых представлены на рис. 7.1.

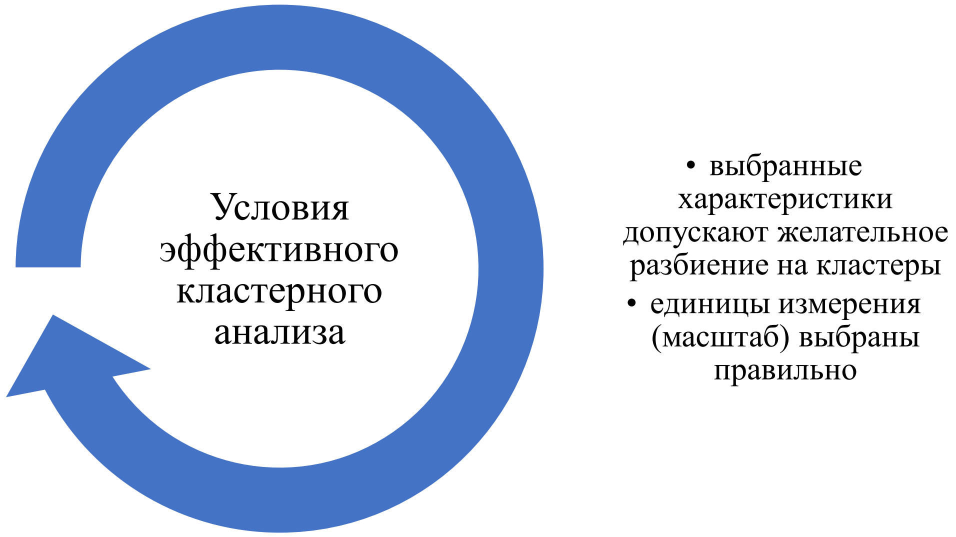



*Рис. 7.1. Условия эффективного кластерного анализа*

Выбор масштаба является крайне значимым этапом проведения кластеризации. Как правило, на предварительном этапе исходные данные, характеризующие признак, нормализуются путем вычитания среднего и делением на стандартное отклонение. При такой операции дисперсия равна 1.

Основная задача проведения кластерного анализа состоит в том, чтобы разбить на основании данных, содержащихся в множестве X, множество объектов G на m кластеров таким образом, чтобы каждый из анализируемых объектов $G_i$ принадлежал только одному из определенных подмножеств разбиения. Важно обеспечение сходства между объектами, принадлежащими одному кластеру, и наличие существенных межкластерных различий.

По сути решение задачи кластерного анализа сводится к разбиению совокупности, которая удовлетворяет условию оптимальности в рамках конкретной исследовательской задачи. По сути критерий оптимальности может представлять собой некоторый функционал, выражающий уровни желательности различных разбиений и группировок, который называют целевой функцией. В роли целевой функции может быть использована внутригрупповая сумма квадратов отклонений.

Важнейшими характеристиками кластера являются его размер, центр, радиус, среднеквадратическое отклонение (рис 7.2)

**Центр кластера**

- среднее геометрическое место точек в пространстве переменных

**Радиус кластера**

- максимальное расстояние точек от центра кластера. Кластеры могут быть перекрывающимися. Такая ситуация возникает, когда обнаруживается перекрытие кластеров. В этом случае невозможно при помощи математических процедур однозначно отнести объект к одному из двух кластеров. Такие объекты называют спорными ( по мере сходства моут быть отнесены к нескольким кластерам)

**Размер кластера**

- может быть определен либо по радиусу кластера, либо по среднеквадратичному отклонению объектов для этого кластера. Объект относится к кластеру, если расстояние от объекта до центра кластера меньше радиуса кластера. Если это условие выполняется для двух и более кластеров, объект является спорным.

*Рис. 7.2. Основные характеристики кластера*

При возникновении проблемы неопределенности, она может быть решена аналитиком или экспертом.

Как уже было отмечено выше, проблема масштаба является значимой при проведении кластерного исследования. Предположим, что набор данных содержит сведения о двух признаках x и y. При этом x принадлежит диапазону от 100 до 700, а y - от 0 до 1. В таком случае корректный расчет расстояний между точками, характеризующими положение объектов, становится невозможным, поскольку переменная, имеющая большие значения, т.е. переменная x, будет практически полностью доминировать над переменной с малыми значениями, т.е. переменной y.

Для устранения данной проблемы используется процедура предварительной стандартизации данных.

Стандартизация (standardization) или нормирование (normalization) приводит значения всех преобразованных переменных к единому диапазону значений путем выражения через отношение этих значений к некой величине, отражающей определенные свойства конкретного признака. В статистике фирмы применяются различные способы проведения данной процедуры, например, по формулам (7.1-7.4)

$$z = \frac{x - \overline{x}}{\sigma} \quad (7.1)$$

$$z = \frac{x}{\overline{x}} \quad (7.2)$$

$$z = \frac{x}{x_{max}} \quad (7.3)$$

$$z = \frac{x - \overline{x}}{x_{max} - x_{min}} \quad (7.4)$$

где $\overline{x}$ - среднее значение признака;

$\sigma$ - среднеквадратическое отклонение x;

$x_{max}$ –наибольшее значение признака;

$x_{min}$ –наименьшее значение признака.

Среднеквадратическое отклонение σ определяется, исходя из числа наблюдений, и может быть определено по формуле 7.5:

$$\sigma = \sqrt{\frac{1}{n-1}\sum^{n}(x_i - \overline{x})^2} \qquad (7.5)$$

Основных видов стандартизации в кластерном анализе представлены в табл. 7.1.

*Табл. 7.1. Основные виды стандартизации в кластерном анализе*

| Стандартизация | Расчет |
|---|---|
| Z-шкалы (Z-Scores) | Из значений переменных вычитается их среднее, и эти значения делятся на стандартное отклонение. |
| Разброс от -1 до 1 | Линейным преобразованием переменных добиваются разброса значений от -1 до 1 |
| Разброс от 0 до 1 | Линейным преобразованием переменных добиваются разброса значений от 0 до 1. |
| Максимум 1 | Значения переменных делятся на их максимум. |
| Среднее 1 | Значения переменных делятся на их среднее |
| Стандартное отклонение 1 | Значения переменных делятся на стандартное отклонение. |

Помимо процедуры стандартизации, на практике также применяется вариант решения существующей проблемы путем корректировки параметров с учетом коэффициента важности, который представляет собой весовую характеристику, отражающую значимость конкретной переменной. При определении данного коэффициента может использоваться метод экспертных оценок, предусматривающий проведение экспертного опроса специалистом конкретной предметной области.

Полученные произведения нормированных переменных на соответствующие веса дают возможность адекватной оценки расстояний между точками в многомерном пространстве с учетом неодинакового веса переменных.

## 7.2. Понятие близости объектов и их характеристика

Вне зависимости от используемого метода классификации и подходов к определению кластеров, проблема измерения близости объектов является неизбежной. При этом основу данного вопроса составляют два основных положения: неоднозначность выбора способа нормировки и определение расстояния между объектами.

Допустим, имеются данные о размере выручки предприятий и численности персонала, по ним оформляется корреляционное поле. Масштабы по осям выбираются произвольно (рис.7.3).

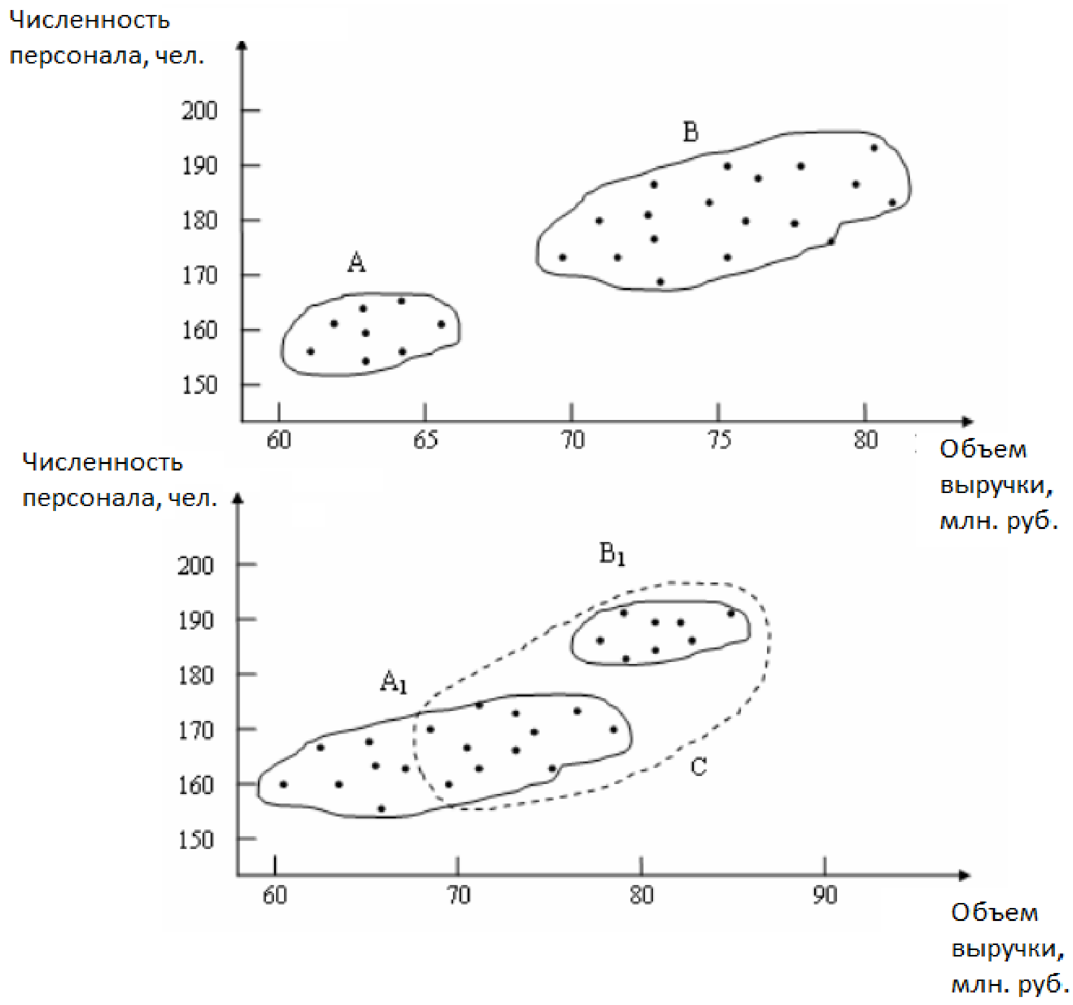



*Рис. 7.3. Корреляционное поле*

На рис. 7.3а выделяются классы А – более мелкие предприятия, В – более крупные предприятия. На рис. 7.3b выделяются классы А1 (включает А и В) и В1(часть В). Класс предприятий С (пунктирная линия) на рис. 7.3б не выделен, так как расстояния между ближайшими объектами классов A1 и B1 существенно больше, чем внутренние расстояния в A1, предприятия A почти никакими алгоритмами к группе B1 не присоединяются.

В данном примере определение расстояний между объектами невозможно, поскольку признаки (численность персонала и объем выручки) представлены в различных единицах измерения. Требуется нормировка показателей, переводящая их в безразмерные величины. Только после этого измерение близости объектов становится возможным.

В кластерном анализе для количественной оценки сходства вводится понятие метрики. Сходство или различие между классифицируемыми объектами устанавливается в зависимости от метрического расстояния между ними. Если каждый объект описывается k признаками, то он может быть представлен как точка в k-мерном пространстве, и сходство с другими объектами будет определяться как соответствующее расстояние.

Метрикой между анализируемыми объектами в пространстве принято называть такую величину $d_{ab}$, которая бы удовлетворяла аксиомам, приведенным на рисунке 7.4.

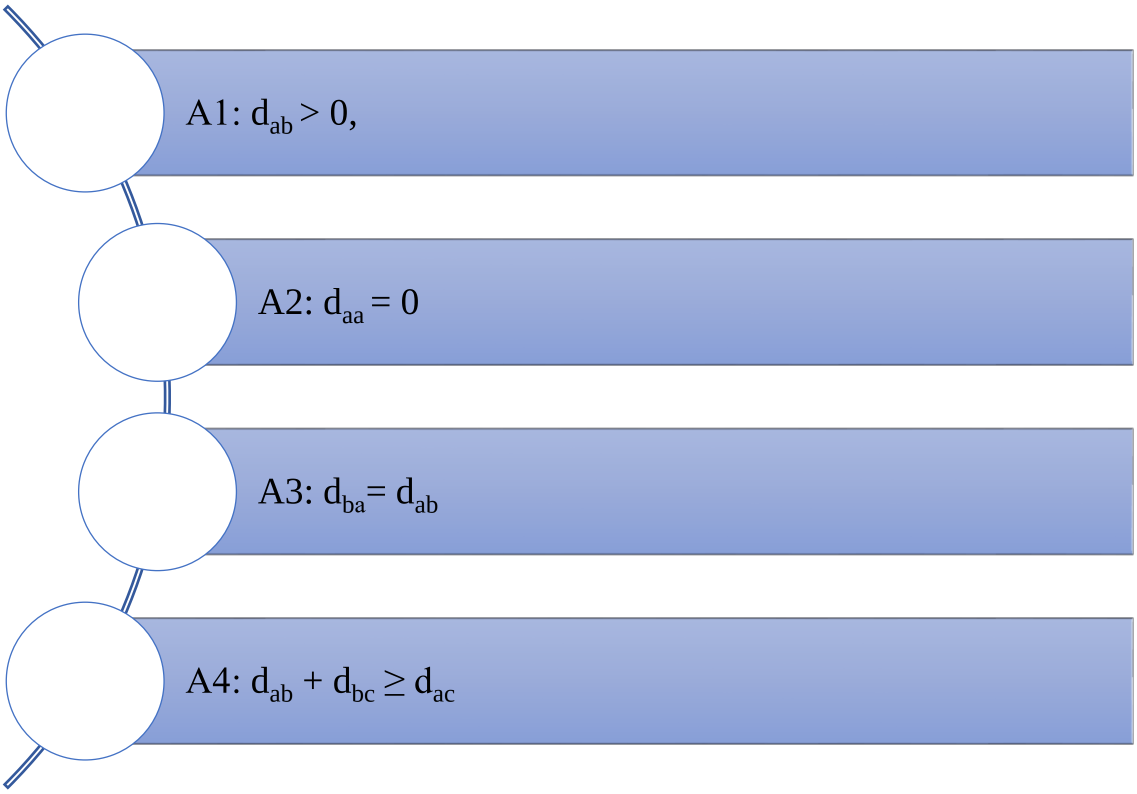



*Рис. 7.4 – Аксиомы метрики*

В качестве меры близости, характеризующей степень сходства, применяется величина $\mu_{ab}$, для которой существует конкретный предел, и которая возрастает с усилением степени близости объектов (рис. 7.5).

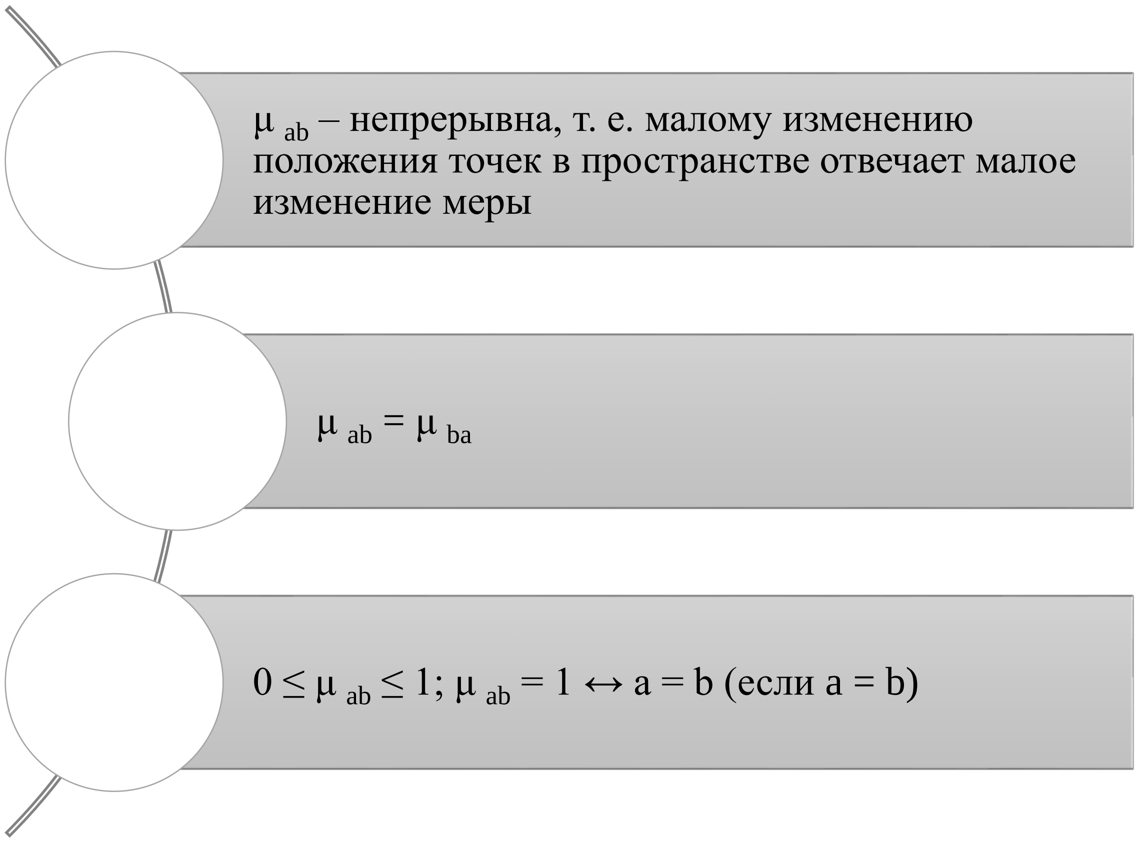



*Рис. 7.5 – Условия меры близости (сходства) объектов*

Для перехода от метрики к расстоянию близости объектов применяют формулу 7.6:

$$\mu = \frac{1}{1+d} \tag{7.6}$$

Объединение или метод древовидной кластеризации используется при формировании кластеров несходства или расстояния между объектами. Эти расстояния могут определяться в одномерном или многомерном пространстве. Например, если требуется кластеризовать организации или фирмы, то можно принять во внимание количество работников предприятия, годовой объем выручки, их рентабельность и т.д. Наиболее прямой путь вычисления расстояний между объектами в многомерном пространстве состоит в вычислении евклидовых расстояний. Если имеется двух- или трёхмерное пространство, то данная мера представляет собой реальное геометрическое расстояние между объектами в пространстве.

Данный метод анализа является самым простым с точки зрения вычислений, однако, применение такого алгоритма не дает возможность оценки «реальности» такого расстояния или его производного характера. В зависимости от задач исследования, могут применяться различные характеристики объектов, что требует подбора адекватных методов оценки.

**Евклидово расстояние** наиболее часто используется в качестве метрики кластерного анализа и представляет собой простое геометрическое расстояние, определяемое в многомерном пространстве. С точки зрения геометрии применение данного метода целесообразно в том случае, если для объектов характерно шарообразное скопление.

**Квадрат евклидова расстояния.** Возведение в квадрат стандартного евклидова расстояния позволяет придать больший вес более отдаленным друг от друга объектам.

**Обобщенное степенное расстояние** является универсальной метрикой и является значимой характеристикой только с математической точки зрения.

**Расстояние Чебышева** следует применять только в том случае, что необходимо идентифицировать два объекта как различные в том случае, если они отличаются только по определенному критерию.

**Манхэттенское расстояние** («расстояние городских кварталов», «хэмминговое» расстояние, «сити-блок» расстояние) рассчитывается как среднее разностей по координатам. В большинстве случаев данная мера расстояния приводит к результатам, аналогичным расчетам евклидова расстояния. Однако, для этой меры влияние отдельных выбросов меньше, чем при использовании евклидова расстояния, поскольку здесь координаты не возводятся в квадрат.

**Процент несогласия** рассчитывается в том случае, если анализируются категориальные данные.

Основные способы определения близости между объектами представлены в таблице 7.2.

*Табл. 7.2. Основные способы определения меры близости при проведении кластерного анализа*

| Метрика кластерного анализа | Формула для расчета метрики |
|---|---|
| Линейное расстояние | $d_{Lij} = \sum_{l=1}^{m} \left\lvert x_i^l - x_j^l \right\rvert$ |
| Евклидово расстояние | $d_{Eij} = (\sum_{l=1}^{m} \left(x_i^l - x_j^l\right)^2)^{\frac{1}{2}}$ |
| Квадрат евклидова расстояния | $d^2{}_{Eij} = \sum_{l=1}^{m} \left(x_i^l - x_j^l\right)^2$ |
| Обобщенное степенное расстояние | $d_{Pij} = (\sum_{l=1}^{m} \left(x_i^l - x_j^l\right)^p)^{\frac{1}{p}}$ |
| Расстояние Чебышева | $d_{ij} = \max_{1 \le i, j \le l} \left\lvert x_i - x_j \right\rvert$ |
| Манхэттенское расстояние | $d_H(x_i, x_j) = \sum_{l=1}^{k} \left\lvert x_i^l - x_j^l \right\rvert$ |

При осуществлении кластерного анализа важно понимать, что современные методы кластеризации могут основываться на работе как с количественными, так и с неколичественными данными. На формальном уровне единицей анализа является поименованная сущность (объект данных), описываемая произвольным набором элементарных свойств (качеств). Другими словами, сущность определяется как подмножество во множестве свойств / качеств. Свойство, в свою очередь, определяет, посредством своей встречаемости, группу сущностей, и, следовательно, может рассматриваться как подмножество во множестве сущностей. На практике набор данных существует как последовательность записей, каждая из которых описывает один объект. Качества могут принадлежать к различным группам. Эти группы могут служить аналогами переменных («полей» - в терминах баз данных), а качества, им принадлежащие, - значениям переменных. Но группы, с одной стороны, могут иметь более одного значения для каждой записи, а с другой стороны, их существование в общем случае необязательно. Более того, группы качеств могут существовать динамически и приобретать различный смысл в процессе анализа.

### 7.3. Методы, используемые в кластерном анализе

Существуют различные методы кластерного анализа, применяемые на практике (рис. 7.6).


*Рис. 7.6. Методы кластерного анализа*

Иерархические и неиерархические методы кластеризации отличаются применяемыми алгоритмами и подходами. Используя различные методы кластерного анализа, у исследователя появляется возможность получения различных результатов при одинаковом наборе статистических данных.

Сущность иерархической кластеризации состоит в том, что меньшие кластеры объединяются в группы больших размеров последовательно, либо происходит обратный процесс разбиения больших кластеров на меньшие группы.

**Для иерархических агломеративных методов** (Agglomerative Nesting, AGNES) характерно поэтапное объединение исходных элементов в кластеры. При таком подходе происходит поэтапное сокращение общего числа групп. Алгоритм объединения при использовании данного метода предполагает соединение объектов в кластеры до тех пор, пока все они не будут составлять одну группу (рис. 7.7).

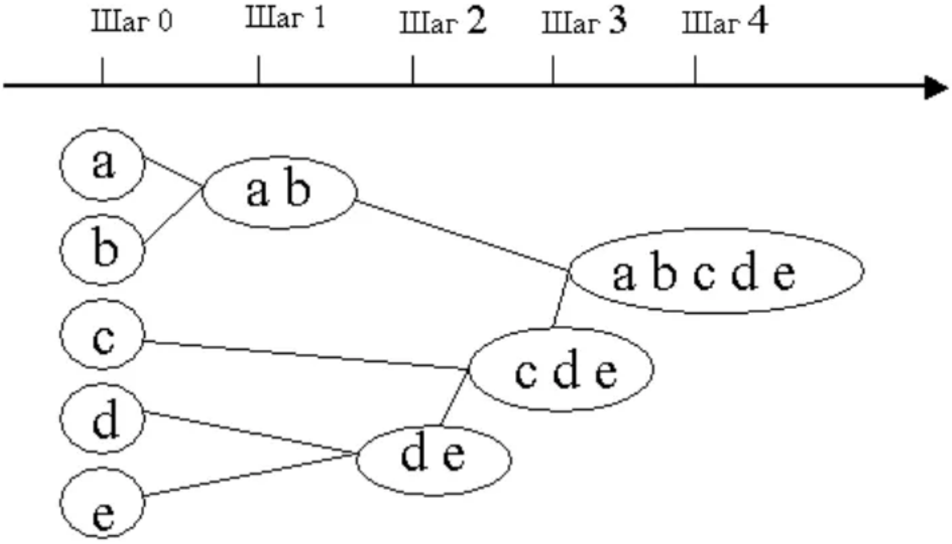



*Рис. 7.7. Алгоритм реализации агломеративных методов*

**Принцип работы иерархических дивизимных методов** (DIvisive ANAlysis, DIANA) логически противоположен принципам кластеризации с опорой на агломеративные методы. При использовании дивизимных методов процесс формируется от обратного: предполагается, что изначально объекты принадлежат одному кластеру, затем с каждым последующим шагом производит разбиение на меньшие группы (рис. 7.8).

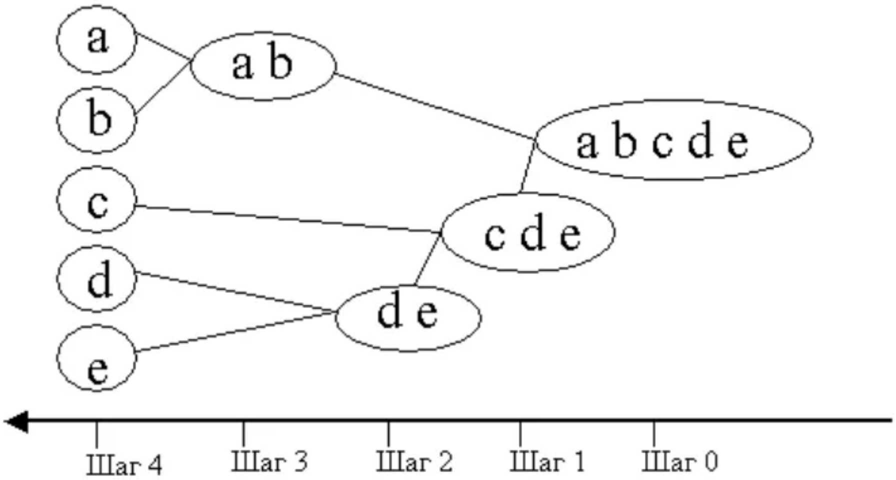



*Рис. 7.8. Алгоритм реализации дивизимных методов*

Применение иерархических методов кластеризации целесообразно в том случае, если исходный объем характеристик для описания кластеров является относительно небольшим. Очевидным преимуществом данных методов является их наглядность, что особенно удобно при необходимости представления результатов исследования в графическом виде.

В результате реализации иерархических алгоритмов становится возможным построение дендрограмм. Само слово в переводе с греческого означает «дерево». С помощью данного инструмента в древовидной форме отражаются результаты кластеризации.

Дендрограммы (древовидные схемы, деревья объединения, деревья иерархической структуры групп) применяются для характеристики отдельных точек и кластеров по отношению друг к другу и демонстрируют в виде особого графика последовательность осуществления объединений в результате кластеризации. Каждый уровень дендрограммы соответствует конкретному шагу поэтопного укрупнения числа кластерных групп.

Дендрограммы могут быть представлены в вертикальном (рис. 7.9) и горизонтальном виде (рис. 7.10).

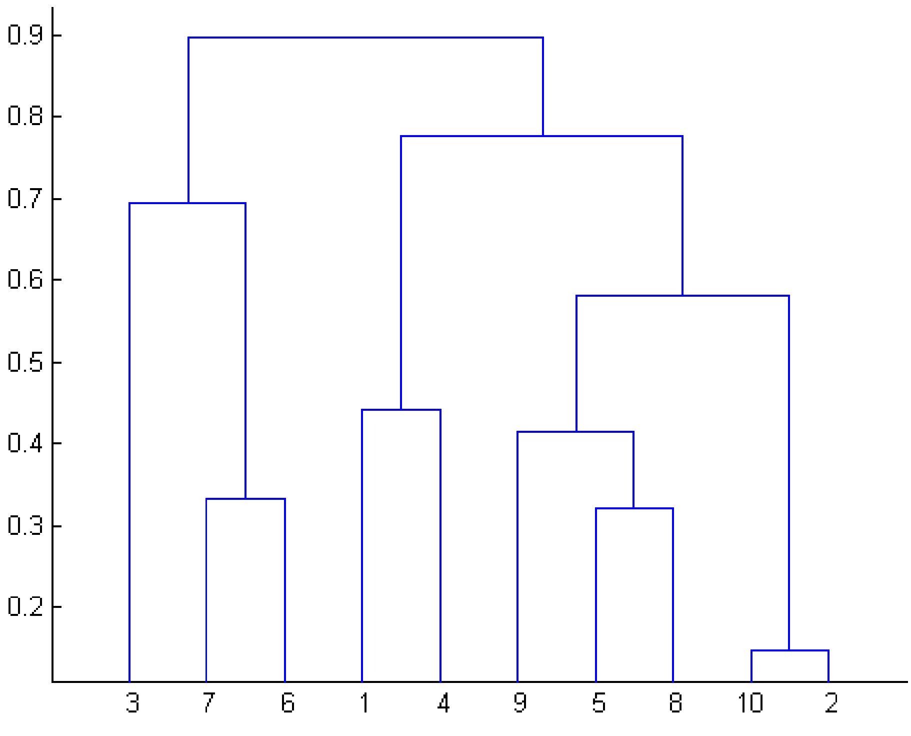



*Рис. 7.9. Пример вертикальной дендрограммы*

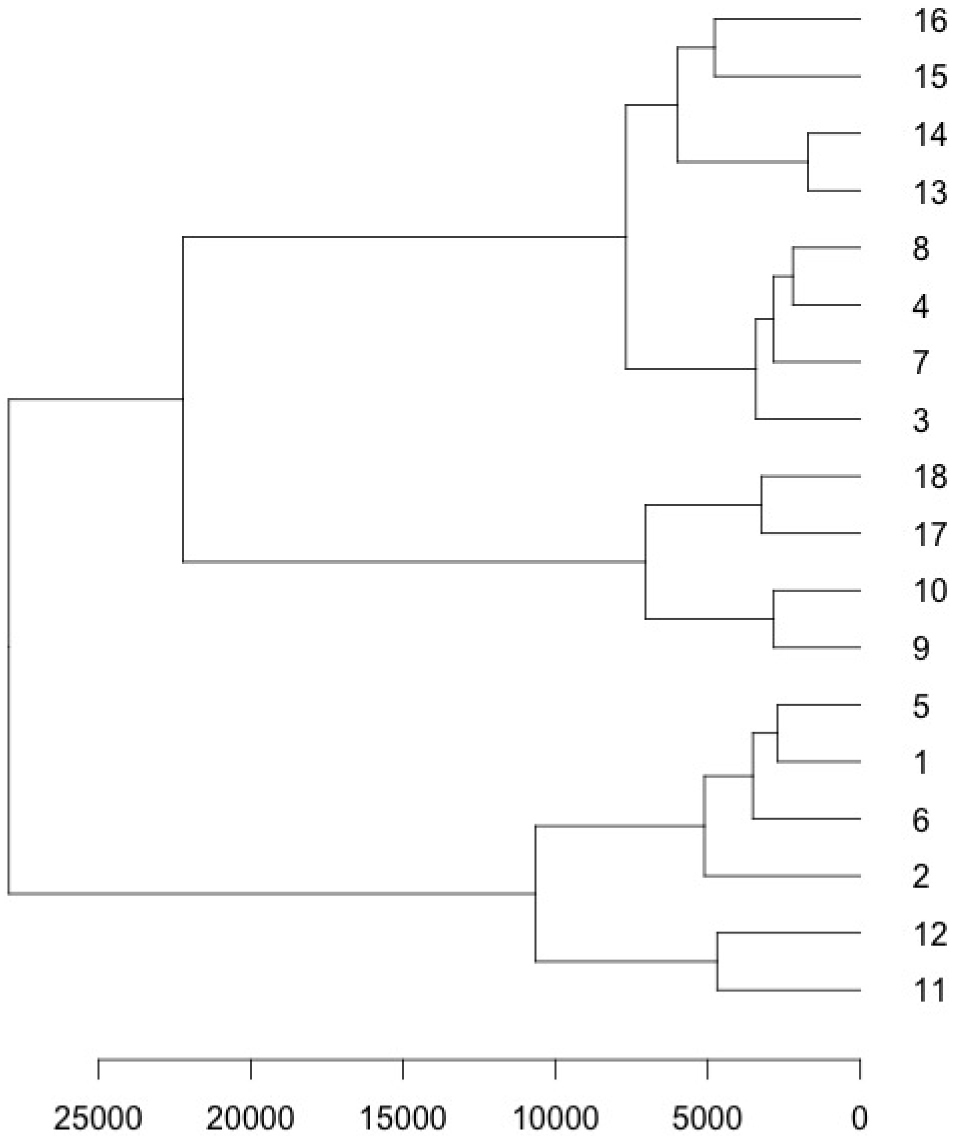



*Рис. 7.10. Пример горизонтальной дендрограммы*

При осуществлении кластеризации важно понимать, каким способом определяется расстояние между объектами и каким образом происходит объединение элементов в группу.

**Метод одиночной связи («ближайшего» соседа)** предполагает наиболее близкое расположение друг по отношению к другу объектов в кластерах по сравнению с соответствующим расстоянием связи (7.7):

$$\rho_{min}(K_i, K_j) = \min_{x_i \in K_i, x_{ij} \in K_j} \rho(x_i, x_{j)} \qquad (7.7)$$

Метод ближайшего соседа применяется при построении кластеров, которые, как правило, связаны между собой не системными связями, а отдельными элементами, оказавшимися на минимальном расстоянии друг от друга.

Методом, противоположным данному, **является метод полной связи («дальнего» соседа)**. Его реализация предполагает оценку межкластерных расстояний по величине, характеризующей наиболее отдаленное положение всех остальных пар объектов друг от друга (7.8):

$$\rho_{maa}(K_i, K_j) = \max_{x_i \in K_i, x_{ij} \in K_j} \rho(x_i, x_{j)} \qquad (7.8)$$

**Метод Варда** был открыт в 1963 году. В качестве расстояния между кластерами берется прирост суммы квадратов расстояний объектов до центров кластеров, получаемый в результате их объединения. В отличие от других методов кластерного анализа для оценки расстояний между кластерами, здесь используются методы дисперсионного анализа.

Каждый шаг реализации кластерного алгоритма предполагает объединять такие два кластера, которые приводят к минимальному увеличению целевой функции, т.е. внутригрупповой суммы квадратов. В результате реализации данного метода происходит создание малых групп, а основой реализации способа является объединение близко расположенных кластеров

**Метод невзвешенного попарного среднего (метод невзвешенного попарного арифметического среднего)** предполагает, что в качестве расстояния между двумя кластерами берется среднее расстояние между всеми парами объектов в них. Этот метод следует использовать, если объекты существенно отличаются друг от друга, в случаях присутствия кластеров «цепочного» типа, а также при предположении неравных размеров кластеров.

**Метод взвешенного попарного среднего (метод взвешенного попарного арифметического среднего)** отличается от предыдущего рассмотренного метода тем, что численность объектов кластера в нем является весовой характеристикой. В остальном алгоритм реализации метода аналогичен. Способ взвешенного попарного арифметического среднего следует использовать в том случае, если выдвинута гипотеза о различном размере предполагаемых кластерных групп.

**Невзвешенный центроидный метод (метод невзвешенного попарного центроидного усреднения**) предполагает использование в качестве расстояния между двумя кластерами расстояние между центрами тяжести соответствующих групп.

**Взвешенный центроидный метод (метод взвешенного попарного центроидного усреднения)**. Этот метод похож на предыдущий, разница состоит лишь в том, что для учета разницы между размерами кластеров (число объектов в них) используются веса. Следует использовать в том случае, если выдвинута гипотеза о существенных различиях в размере предполагаемых кластерных групп.

Помимо иерархических методов, на практике часто используются **итерационные приемы кластеризации**.

Наиболее популярным из них является **метод k-средних**. Впервые полный алгоритм быстрой кластеризации был рассмотрен в работе 1978 года Хартигана и Вонга (Hartigan and Wong). Отличие метода k-средних от иных иерархических методов состоит в необходимости выдвижения гипотезы о числе кластеров до начала проведения анализа.

Метод k-средних имеет свои преимущества и недостатки (рис.7.11).

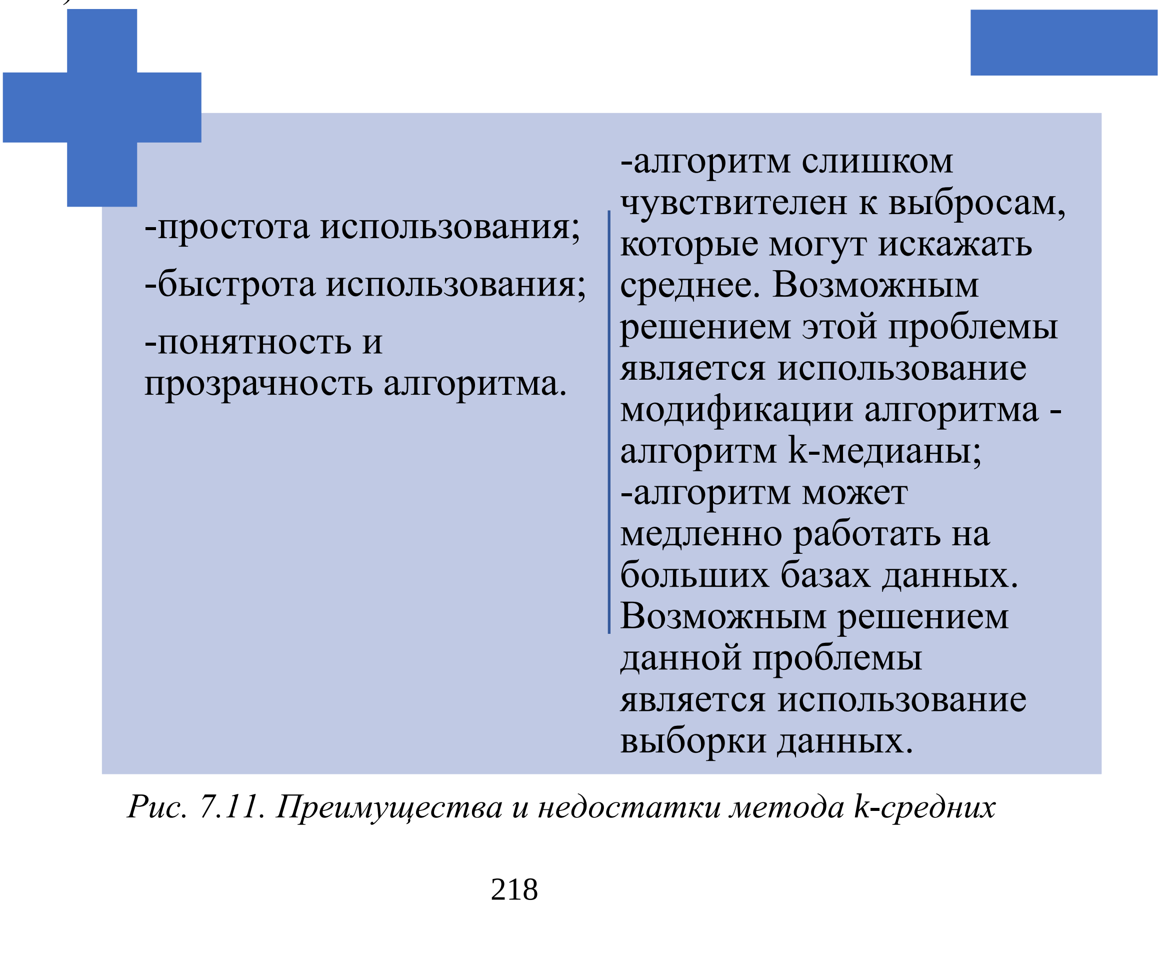



*Рис. 7.11. Преимущества и недостатки метода k-средних*

Алгоритм реализации метода представлен на рис. 7.12.


*Рис. 7.12. Алгоритм реализации метода k-средних*

Суть принципа реализации алгоритма k-средних состоит в следующем: происходит построение кластеров, расстояние между которыми является наибольшим. Выбор числа кластерных групп k может опираться на предыдущие результаты анализа, теоретические положения в конкретной предметной области и т.д.

Процесс итерации прекращается, когда границы кластеров не перестанут изменяться от итерации к итерации, т.е. на каждой итерации в каждом кластере будет оставаться один и тот же набор записей.

После получений результатов кластерного анализа методом k-средних необходимо проверять адекватность проведенной процедуры. Суть такой проверки состоит в оценке значимости различий вычисленных кластерных групп. Анализ различий основывается на расчете средних значений групп. Если процедура дала качественные результаты, средние характеристики каждого из выявленных кластеров должны существенно отличаться друг от друга [108].

Пример реализации алгоритма для двух кластерных групп представлен на рис. 7.13 [109].

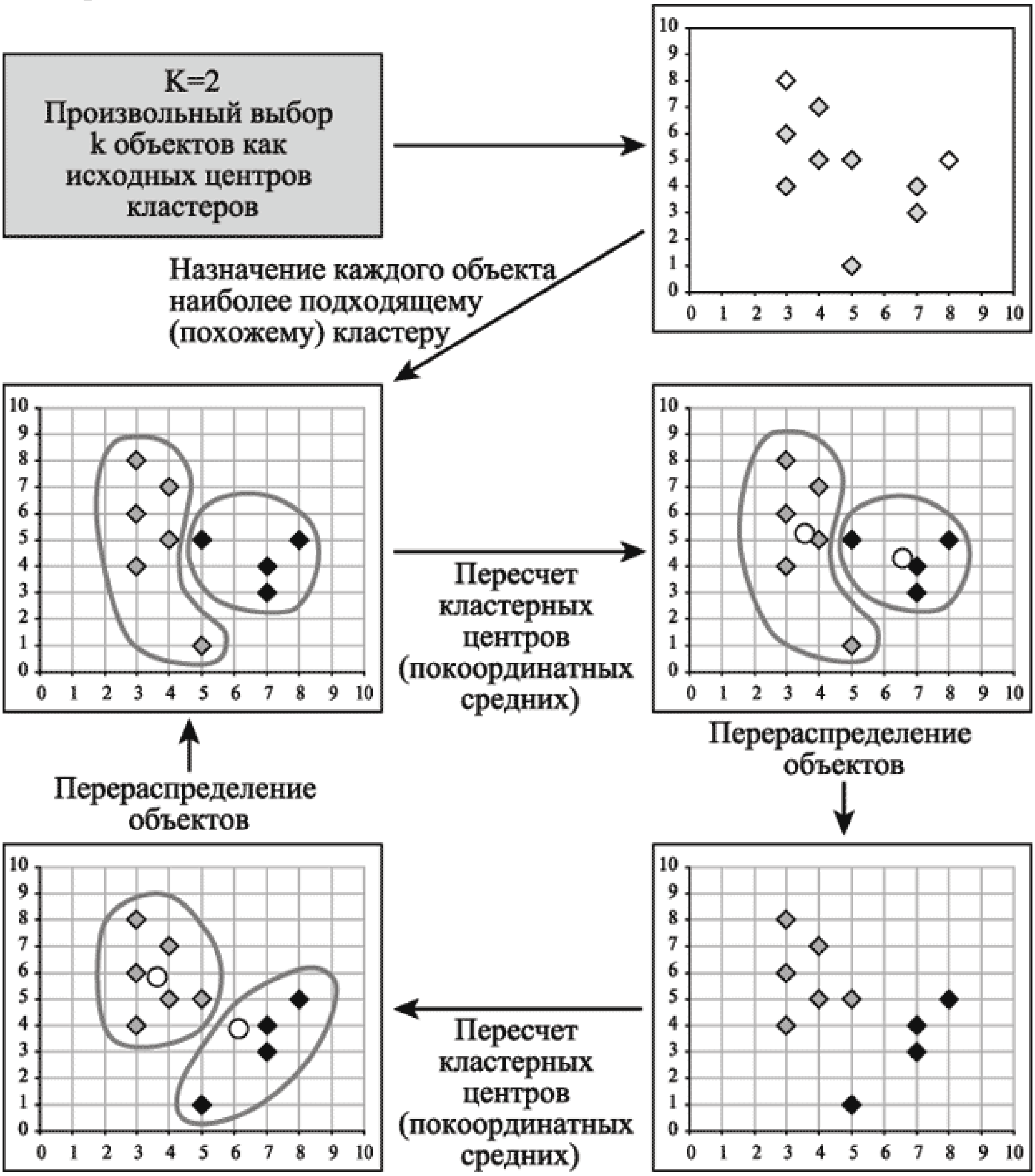



*Рис. 7.13. Алгоритм реализации метода k-средних для k=2*

## 7.4. Особенности решения задачи кластерного анализа в программе Statistica

Рассмотрим основные шаги реализации кластерного анализа с использованием программы в программе Statistica.

Начальным этапом кластеризации является ввод необходимых переменных в окне при открытии программы (рис. 7.14)

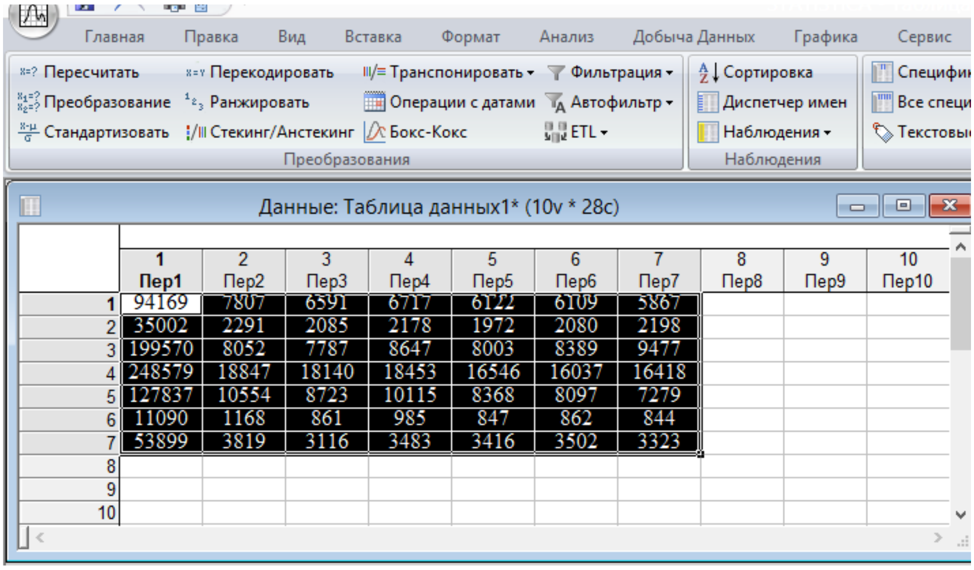



*Рис. 7.14. Ввод исходных данных*

Если применяются разнородные данные, их необходимо стандартизировать для приведения измерений к общей безразмерной величине

Для этого во вкладке «Данные» нужно выбрать пункт «Стандартизировать» (рис. 7.15)

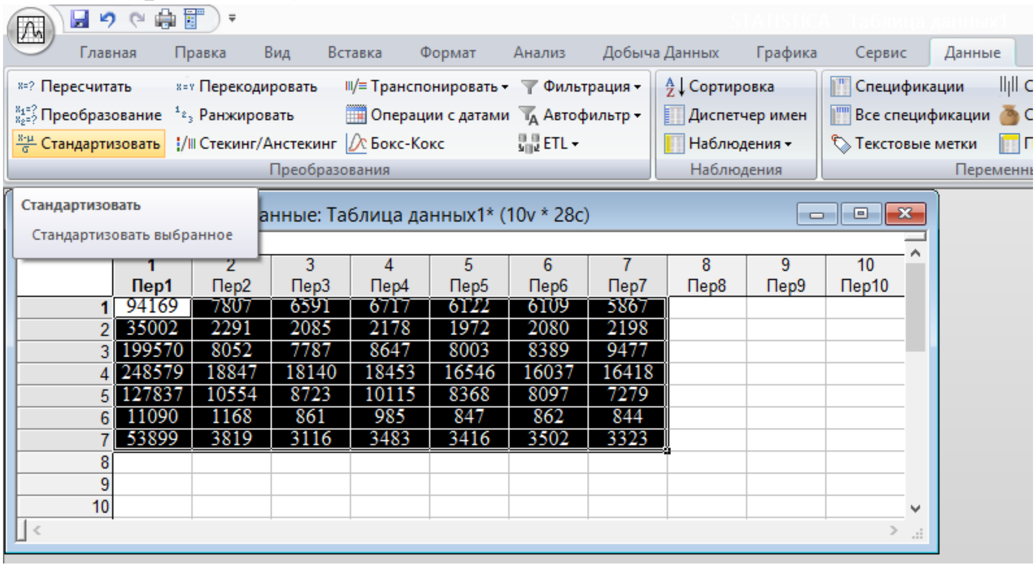



*Рис. 7.15. Ввод исходных данных*

При определении объекта стандартизации можно выбрать «переменные» или «наблюдения» (рис. 7.16).

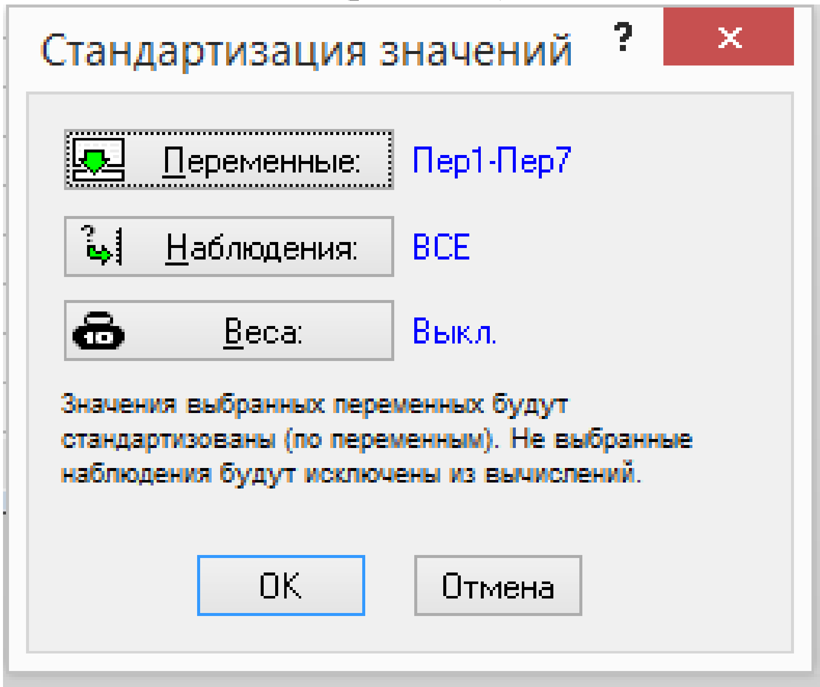



*Рис. 7.16. Выбор объекта стандартизации*

Также существует возможность задать вес для каждого из наблюдений во вкладке «Веса» (рис. 7.17).

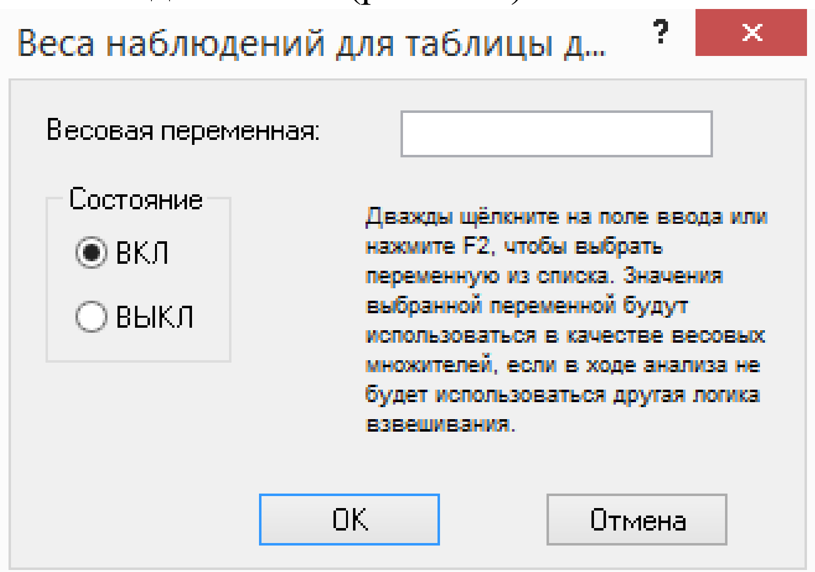



*Рис. 7.17. Ввод веса наблюдений*

Если выбрать стандартизацию по переменным, общий вид данных после стандартизации примет следующий вид (рис.7.18):

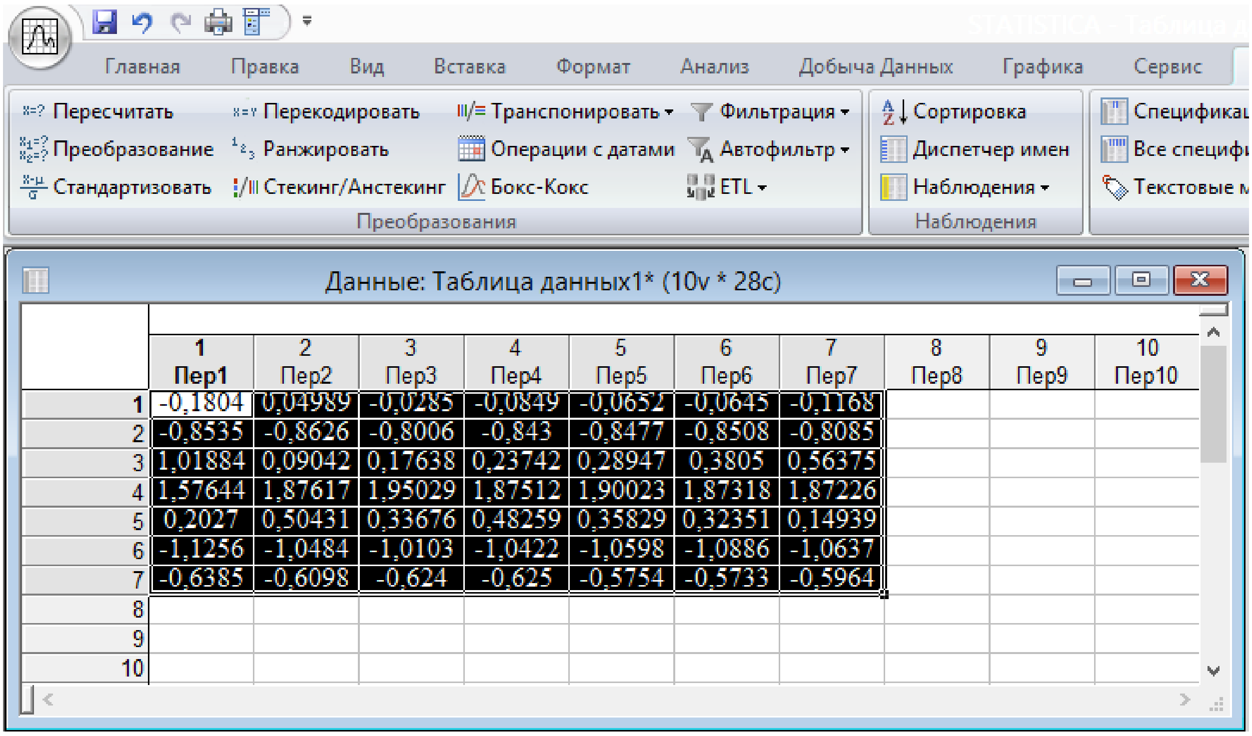



*Рис. 7.18. Массив исходных данных после стандартизации*

Для начала осуществления непосредственно кластерного анализа во вкладке «Анализ» главной панели был выбран «Многомерный анализ», а затем «Кластерный анализ» (рис. 7.19).

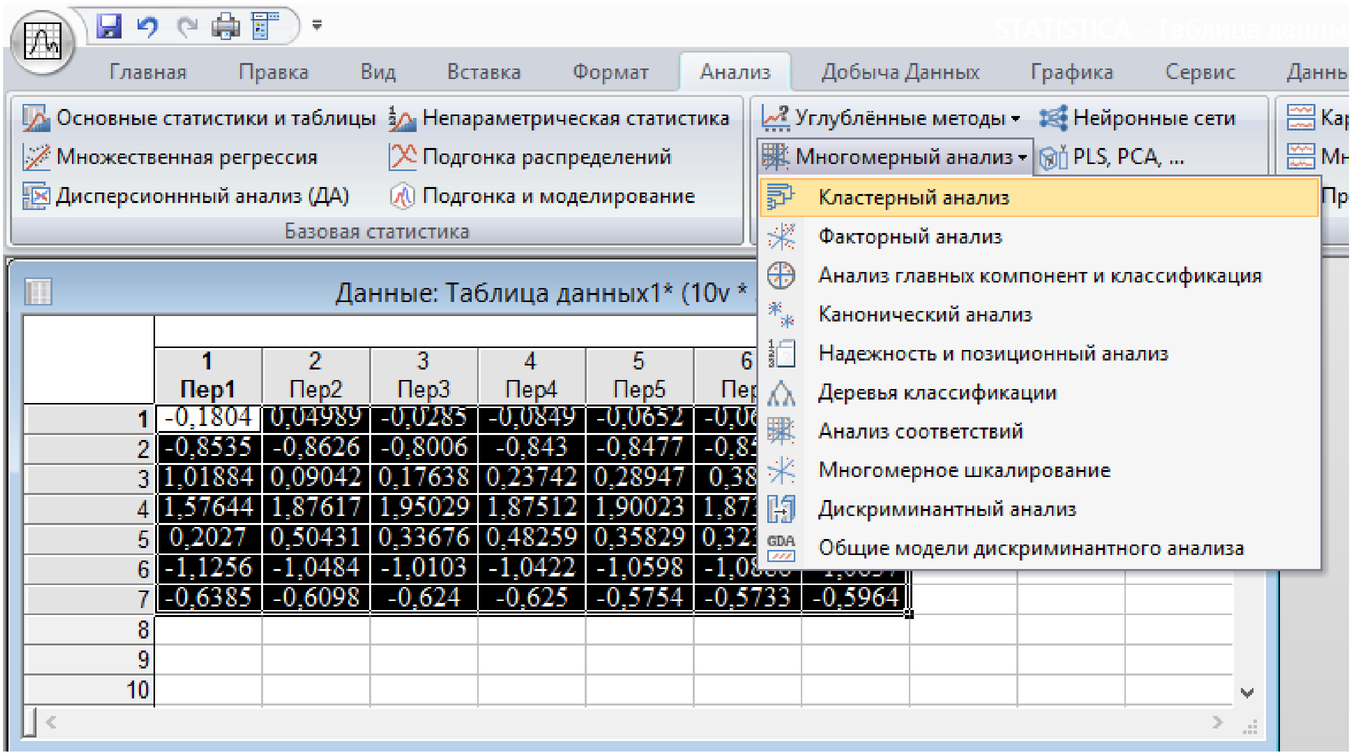



*Рис. 7.19. Выбор кластерного анализа как метода исследования*

На следующем шаге осуществляется выбор методов кластеризации из предложенных программой вариантов (рис.7.20):

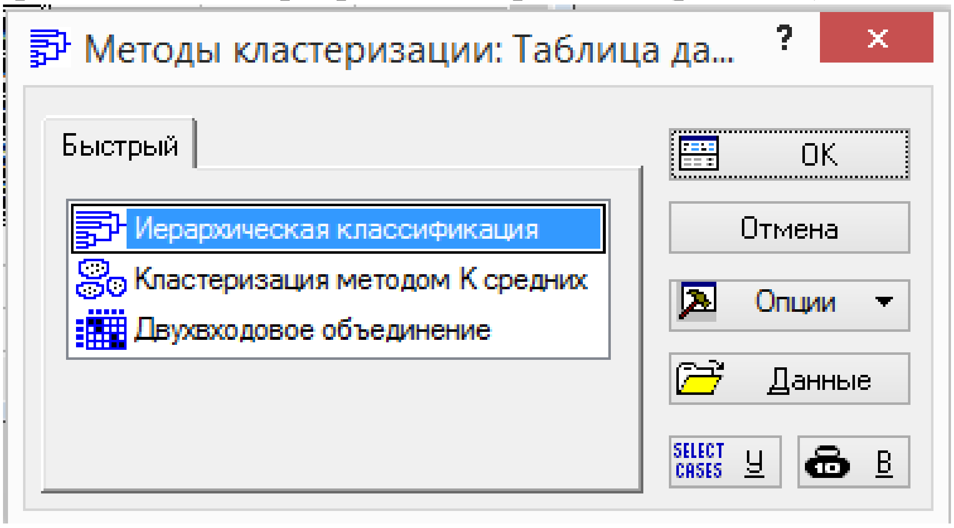



*Рис. 7.20. Выбор метода кластеризации*

При выборе иерархической классификации будет отображено следующее окно, в котором появится возможность выбора переменных для анализа (рис.7.21):

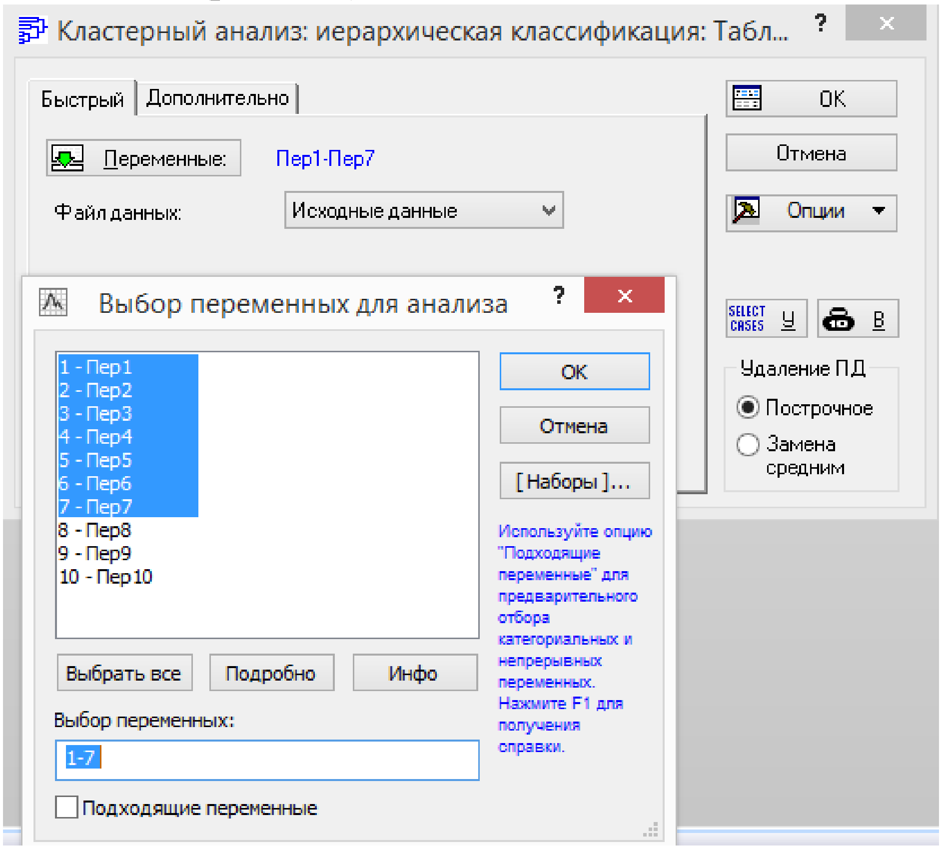



*Рис. 7.21. Выбор переменных для осуществления иерархической кластеризации*

В случае, если объектом кластеризации будут выступать не столбцы, а строки, во вкладке «Дополнительно» есть возможность изменения данного параметра (рис.7.22):

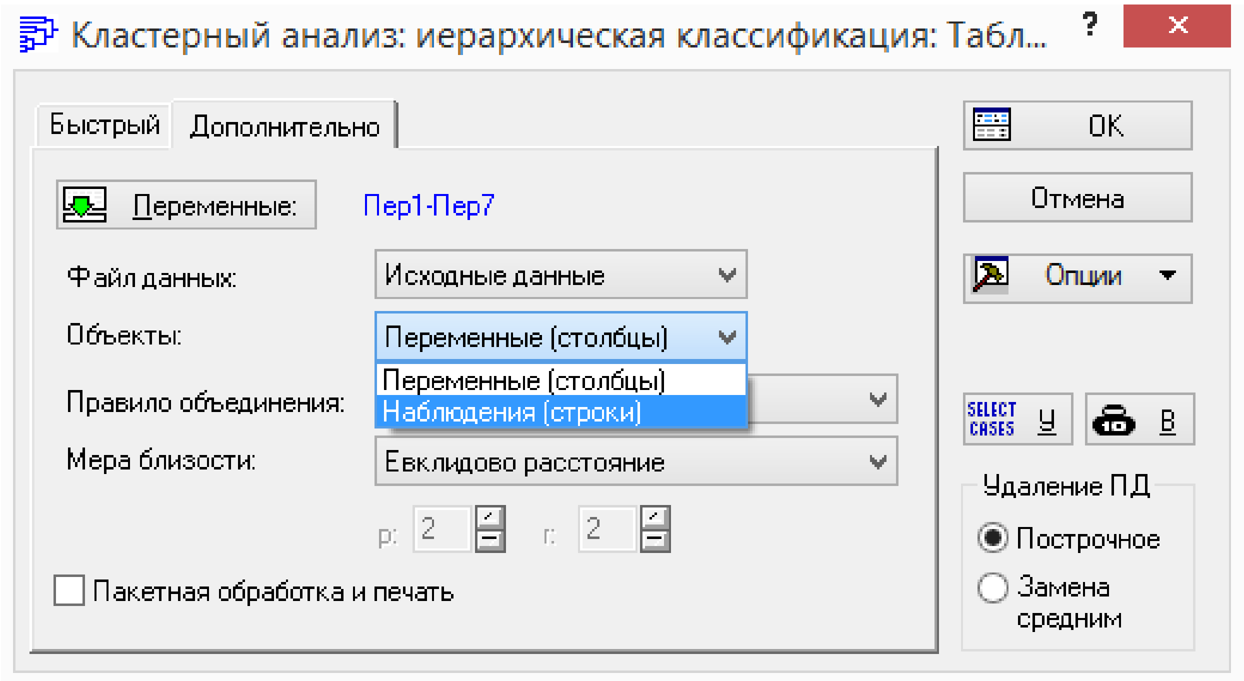



*Рис. 7.22. Изменение объекта иерархической кластеризации*

Кроме того, на данной вкладке есть возможность выбора правила объединения (рис.7.23):

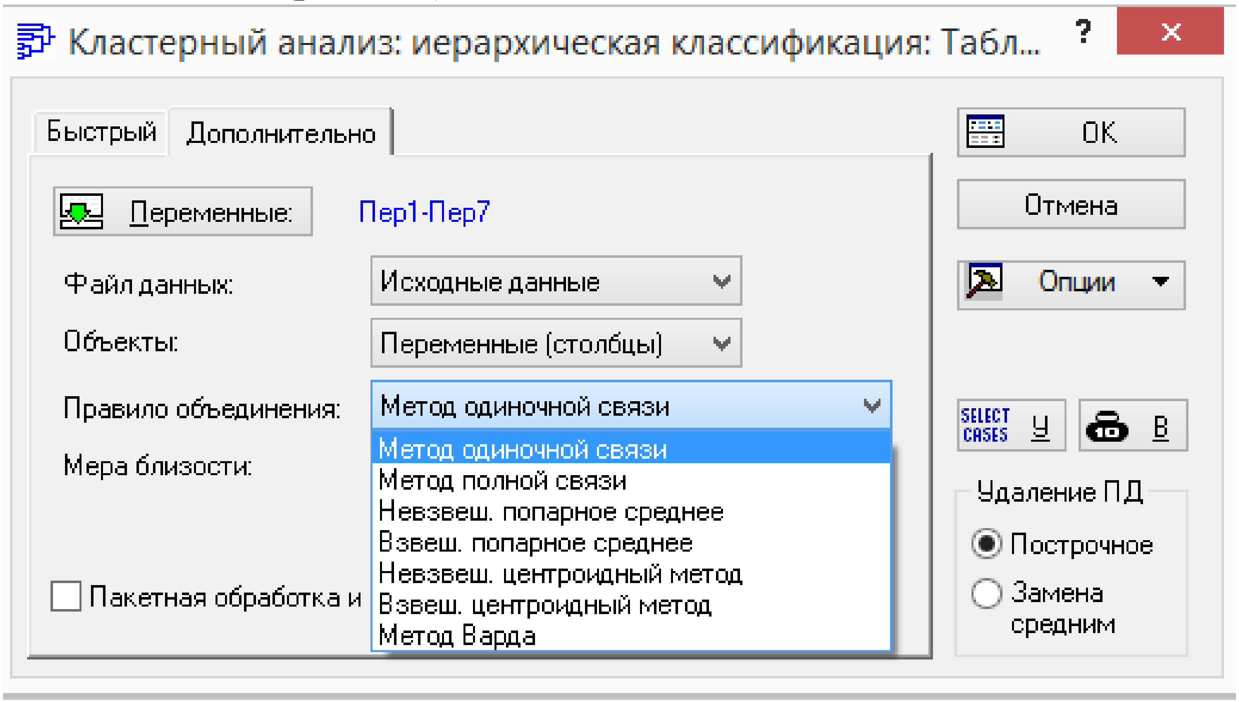



*Рис. 7.23. Выбора правила объединения*

В качестве вариантов мер близости программа представляет возможность выбора одного варианта из 7 альтернатив (рис.7.24):

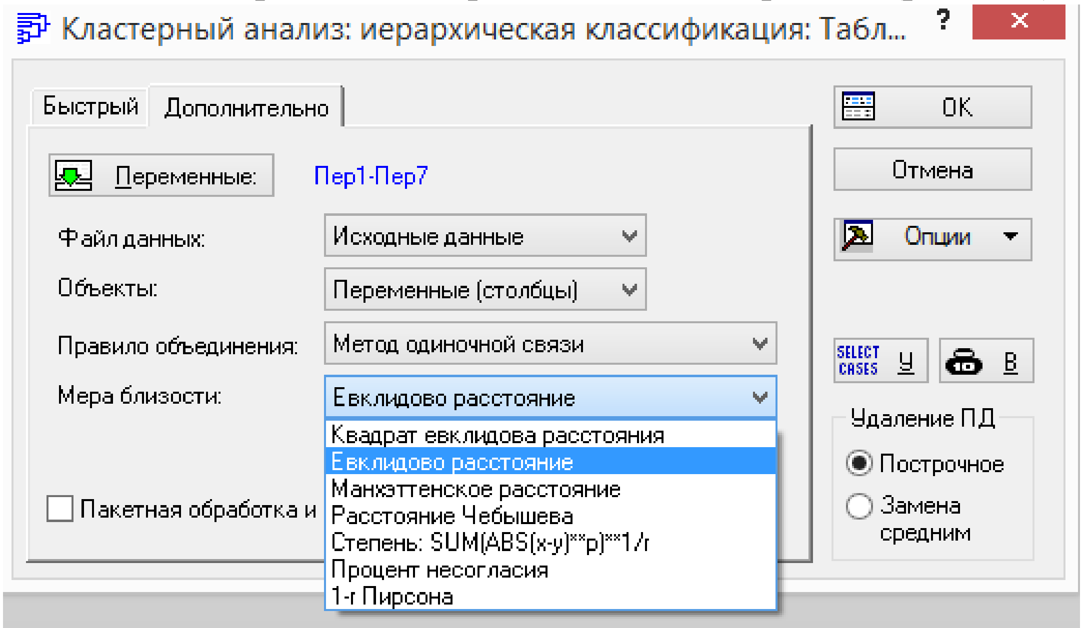



*Рис. 7.24. Выбор меры близости*

Результаты кластеризации по переменным-столбцам по методу одиночной связи с использованием Евклидова расстояния в качестве меры близости (рис.7.25):

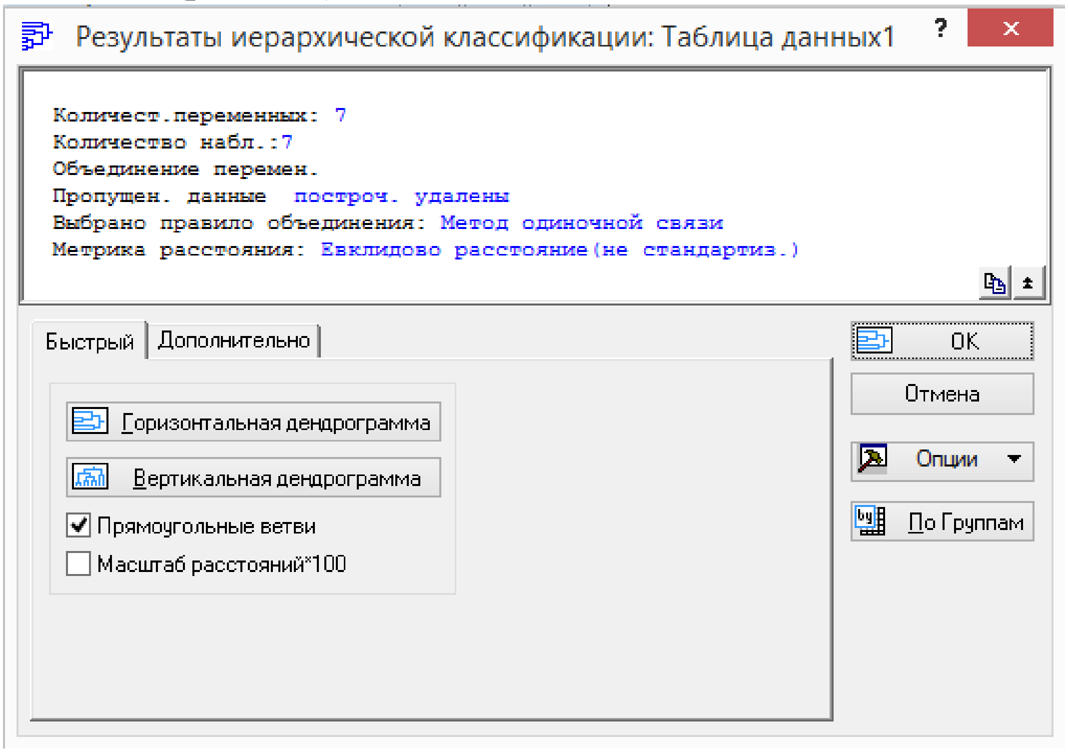



*Рис. 7.25. Окно вывода быстрых результатов иерархической кластеризации*

Горизонтальная дендрограмма примет вид (7.26):

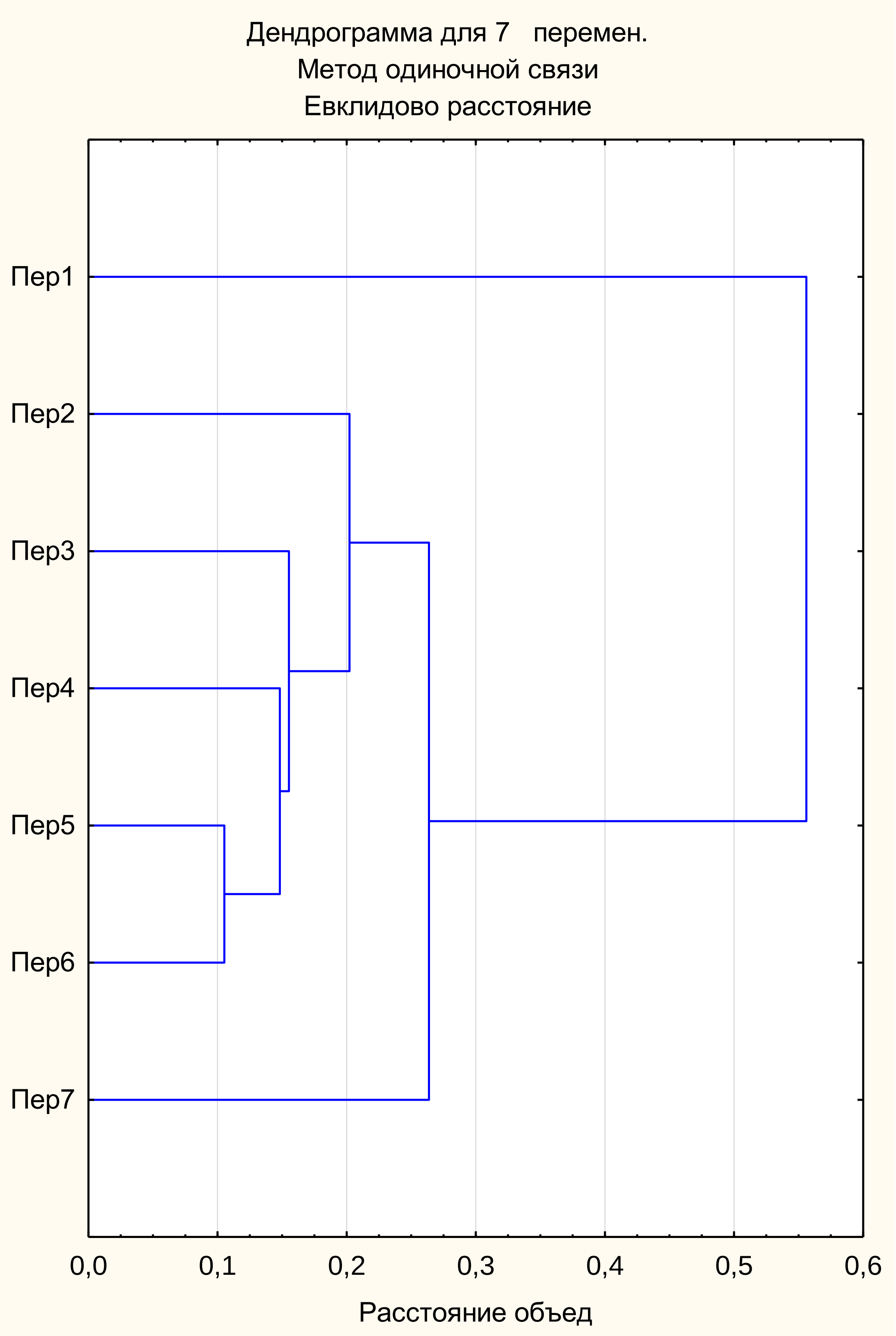



*Рис. 7.26. Вывод результатов иерархической кластеризации в виде горизонтальной дендрограммы*

При использовании вертикальной дендрограммы (7.27):

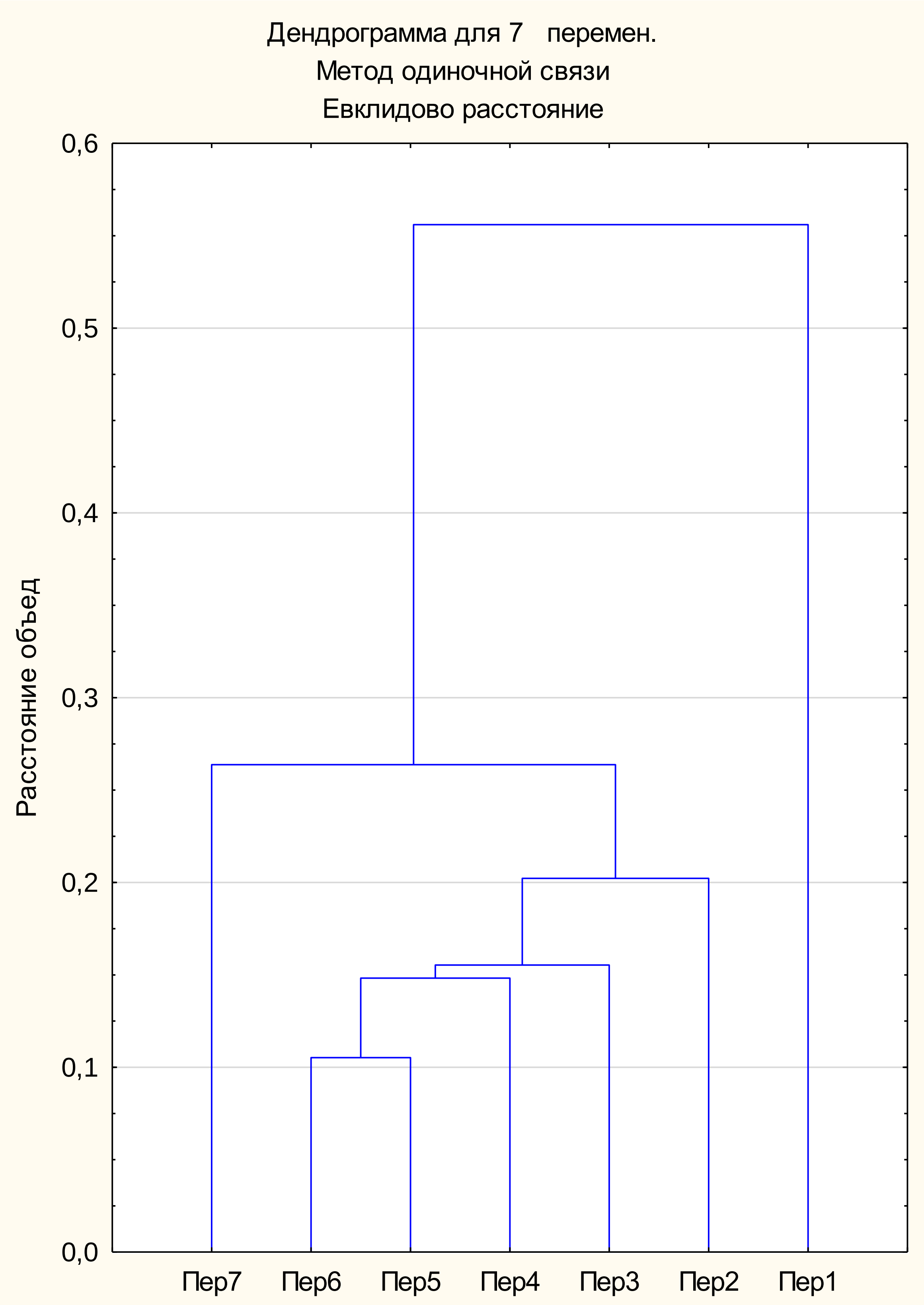



*Рис. 7.27. Вывод результатов иерархической кластеризации в виде вертикальной дендрограммы*

Во вкладке «Дополнительно» по результатам иерархической классификации можно рассмотреть иные результаты анализа (рис.7.28):

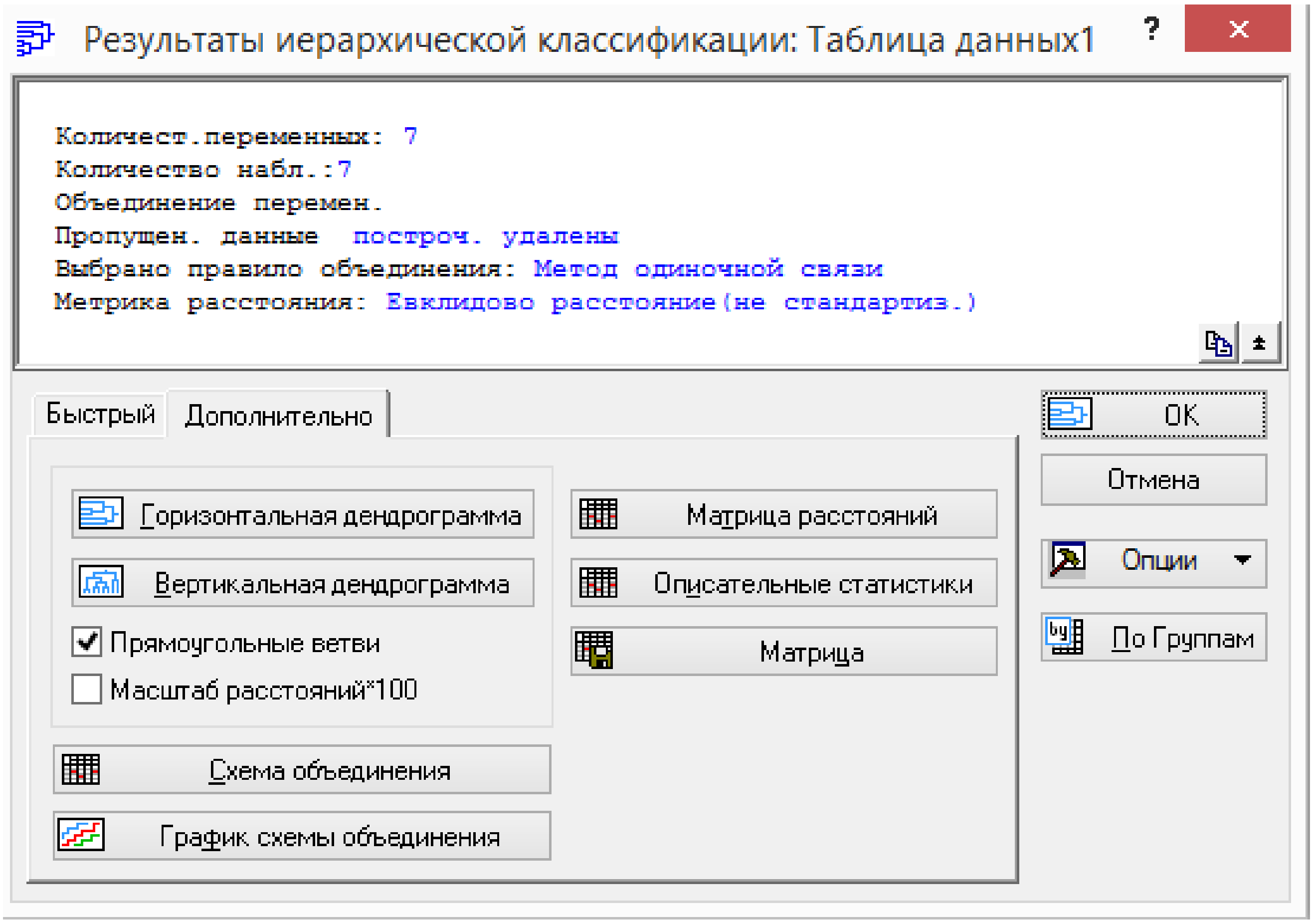



*Рис. 7.28. Окно вывода дополнительных результатов иерархической кластеризации*

Матрица расстояний для представленного массива исходных данных (рис. 7.29):


| | Евклидово расстояние (Таблица данных1) | | | | | | |
|---|---|---|---|---|---|---|---|
| перемен. | Пер1 | Пер2 | Пер3 | Пер4 | Пер5 | Пер6 | Пер7 |
| Пер1 | 0,00 | 1,05 | 0,95 | 0,89 | 0,83 | 0,73 | 0,56 |
| Пер2 | 1,05 | 0,00 | 0,23 | 0,20 | 0,28 | 0,36 | 0,62 |
| Пер3 | 0,95 | 0,23 | 0,00 | 0,19 | 0,16 | 0,25 | 0,45 |
| Пер4 | 0,89 | 0,20 | 0,19 | 0,00 | 0,15 | 0,23 | 0,47 |
| Пер5 | 0,83 | 0,28 | 0,16 | 0,15 | 0,00 | 0,11 | 0,35 |
| Пер6 | 0,73 | 0,36 | 0,25 | 0,23 | 0,11 | 0,00 | 0,26 |
| Пер7 | 0,56 | 0,62 | 0,45 | 0,47 | 0,35 | 0,26 | 0,00 |

*Рис. 7.29. Матрица расстояний*

Описательные статистики (рис. 7.30):

| перемен. | Средние и станд. отклонения (Таблица данных1) Среднее | Стд.откл |
|---|---|---|
| **Пер1** | -0,000000 | 1,000000 |
| Пер2 | -0,000000 | 1,000000 |
| Пер3 | -0,000000 | 1,000000 |
| Пер4 | -0,000000 | 1,000000 |
| Пер5 | -0,000000 | 1,000000 |
| Пер6 | -0,000000 | 1,000000 |
| Пер7 | 0,000000 | 1,000000 |

*Рис. 7.30. Описательные статистики*

Матрица данных имеет вид (рис. 7.31):

| | Таблица данных1 1 Пер1 | 2 Пер2 | 3 Пер3 | 4 Пер4 | 5 Пер5 | 6 Пер6 | 7 Пер7 |
|---|---|---|---|---|---|---|---|
| **Пер1** | 0,00000 | 1,05008 | 0,95227 | 0,89136 | 0,82622 | 0,72745 | 0,55599 |
| Пер2 | 1,05008 | 0,00000 | 0,22934 | 0,20225 | 0,27623 | 0,36471 | 0,61738 |
| Пер3 | 0,95227 | 0,22934 | 0,00000 | 0,19139 | 0,15530 | 0,24553 | 0,45024 |
| Пер4 | 0,89136 | 0,20225 | 0,19139 | 0,00000 | 0,14823 | 0,22602 | 0,47012 |
| Пер5 | 0,82622 | 0,27623 | 0,15530 | 0,14823 | 0,00000 | 0,10522 | 0,35257 |
| Пер6 | 0,72745 | 0,36471 | 0,24553 | 0,22602 | 0,10522 | 0,00000 | 0,26378 |
| Пер7 | 0,55599 | 0,61738 | 0,45024 | 0,47012 | 0,35257 | 0,26378 | 0,00000 |
| Средние | -0,00000 | -0,00000 | -0,00000 | -0,00000 | -0,00000 | -0,00000 | 0,00000 |
| Стд.откл | 1,00000 | 1,00000 | 1,00000 | 1,00000 | 1,00000 | 1,00000 | 1,00000 |
| N набл. | 7,00000 | | | | | | |
| Матрица | 3,00000 | | | | | | |

*Рис. 7.31. Матрица данных*

Схема объединения по методу одиночной связи (рис. 7.32):

Схема объединения (Таблица данных1)
Метод одиночной связи
Евклидово расстояние

| расст. объедин. | Объект 1 | Объект 2 | Объект 3 | Объект 4 | Объект 5 | Объект 6 | Объект 7 |
|---|---|---|---|---|---|---|---|
| **,1052218** | Пер5 | Пер6 | | | | | |
| ,1482306 | Пер4 | Пер5 | Пер6 | | | | |
| ,1552966 | Пер3 | Пер4 | Пер5 | Пер6 | | | |
| ,2022543 | Пер2 | Пер3 | Пер4 | Пер5 | Пер6 | | |
| ,2637829 | Пер2 | Пер3 | Пер4 | Пер5 | Пер6 | Пер7 | |
| ,5559943 | Пер1 | Пер2 | Пер3 | Пер4 | Пер5 | Пер6 | Пер7 |

*Рис. 7.32. Схема объединения по заданным критериям кластеризации (правила объединения и меры близости)*

Пошаговый график объединения (рис. 7.33):

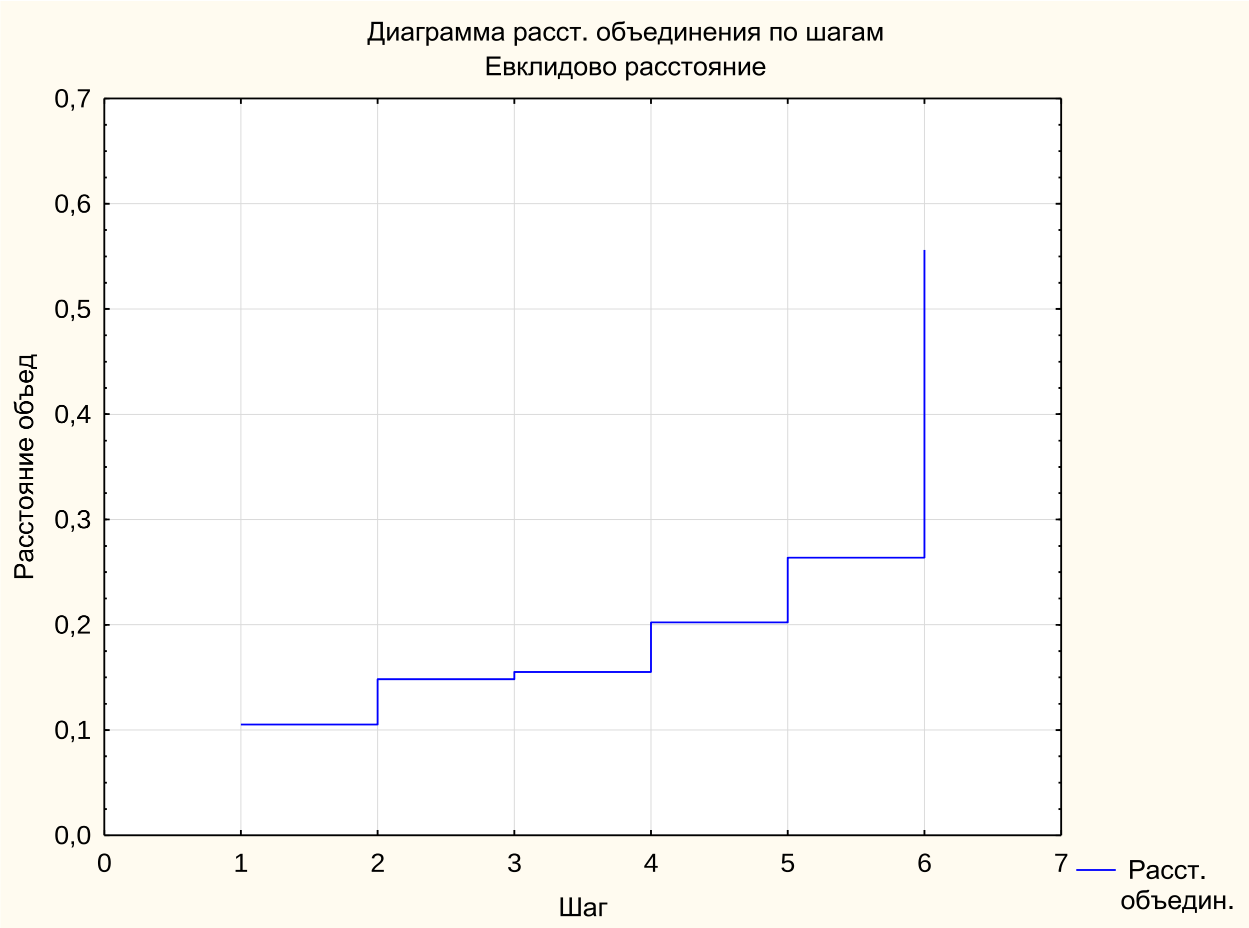



*Рис. 7.33. График пошагового формирования диаграммы расстояний объединения*

Также в практической работе часто используется метод к-средних (7.34)

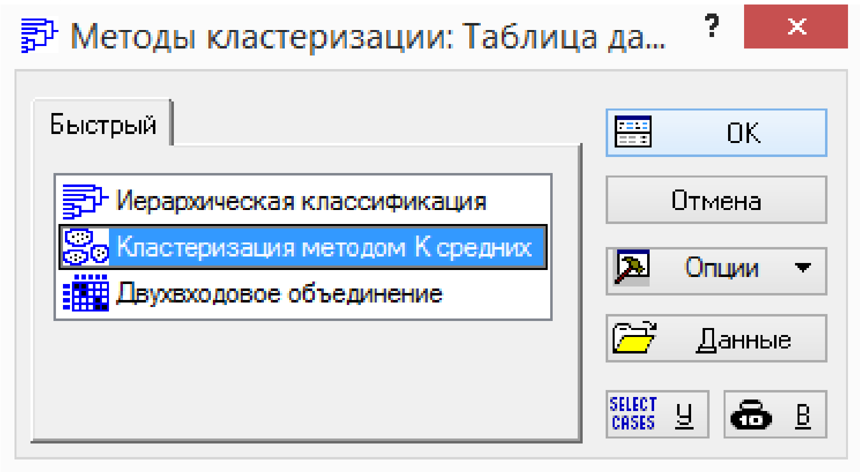



*Рис. 7.34. Окно выбора варианта кластеризации методом k-средних*

При выборе данного метода необходимо определить объект кластеризации (рис. 7.35):

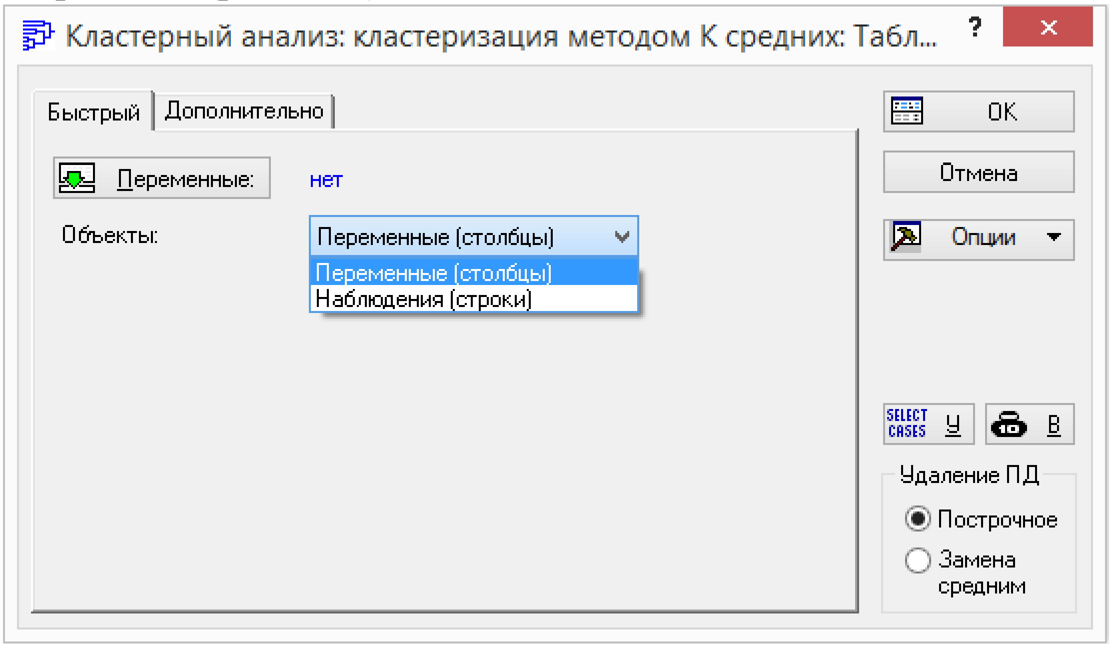



*Рис. 7.35. Определение объекта кластеризации*

А также во вкладке «Дополнительно» задать желаемое (рассчитанное заранее или определенное по результатам предварительного графического анализа) число кластерных групп (рис. 7.36):

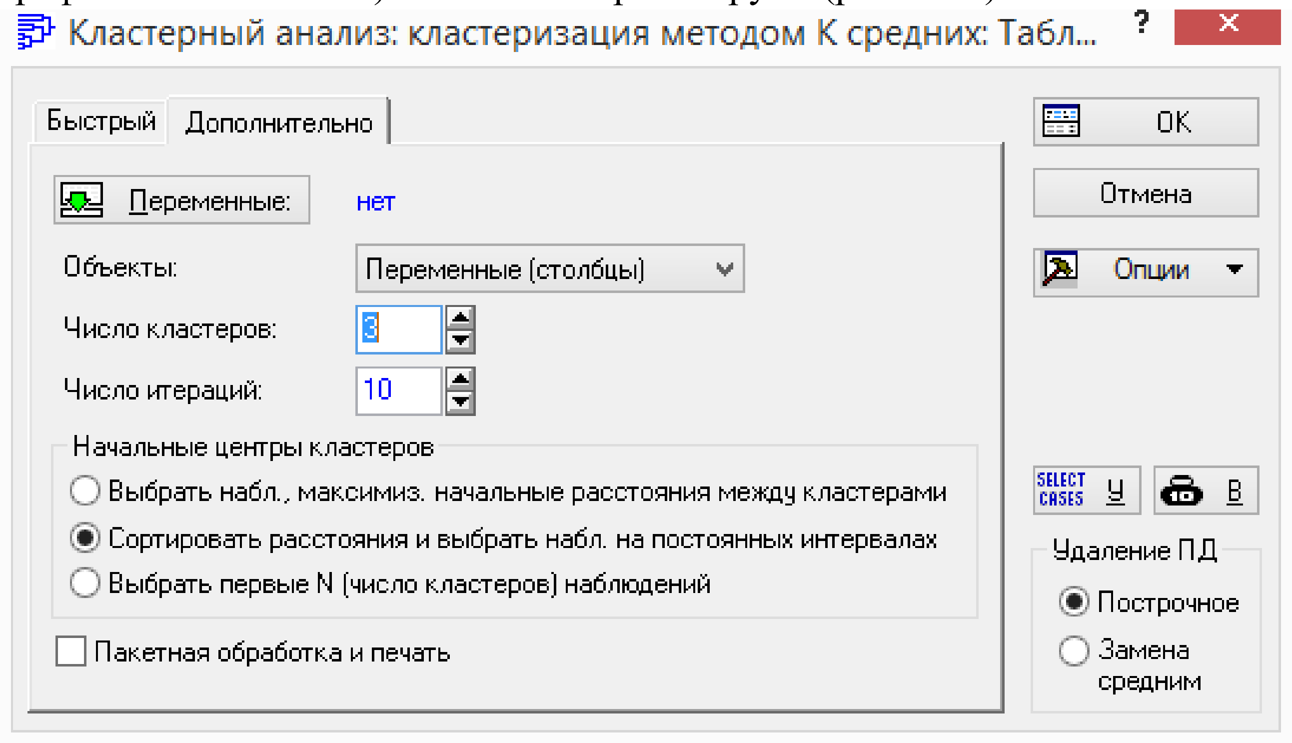



*Рис. 7.36. Определение числа кластеров и итераций*

В результате открывается окно выбора переменных для анализа (рис. 7.37):

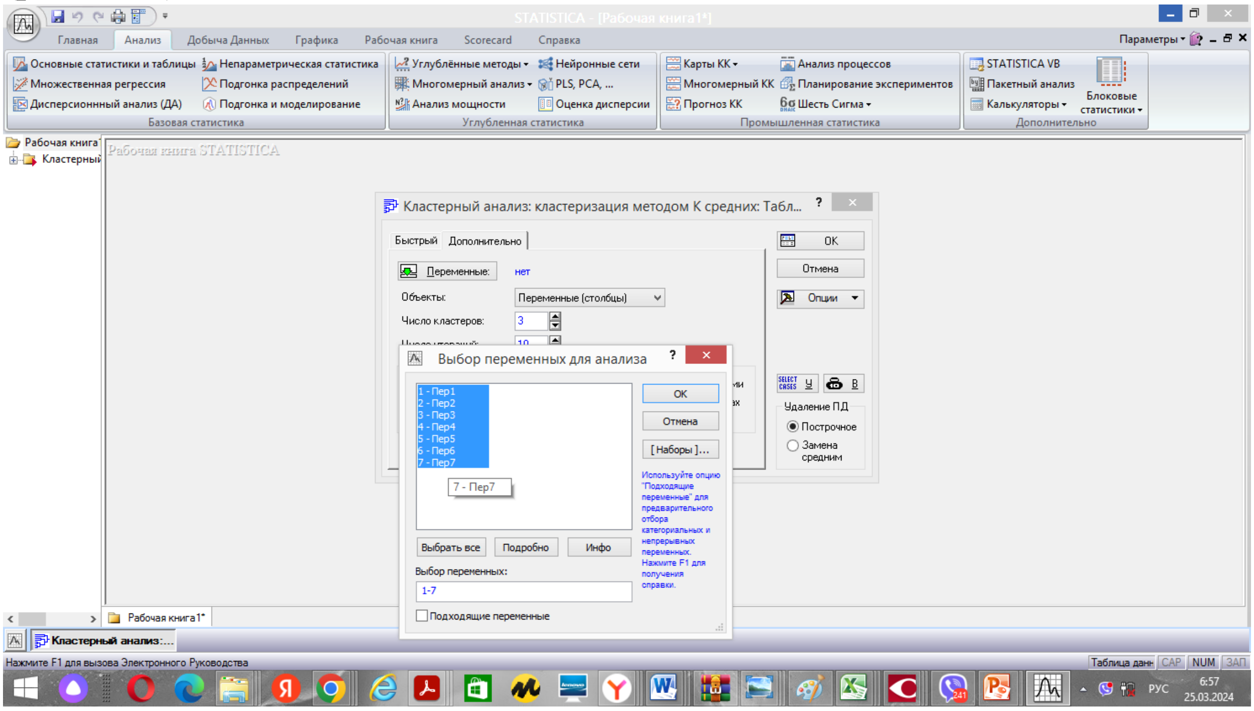



*Рис. 7.37. Выбор переменных для анализа*

Результаты выполнения заданных расчетов (рис. 7.38):

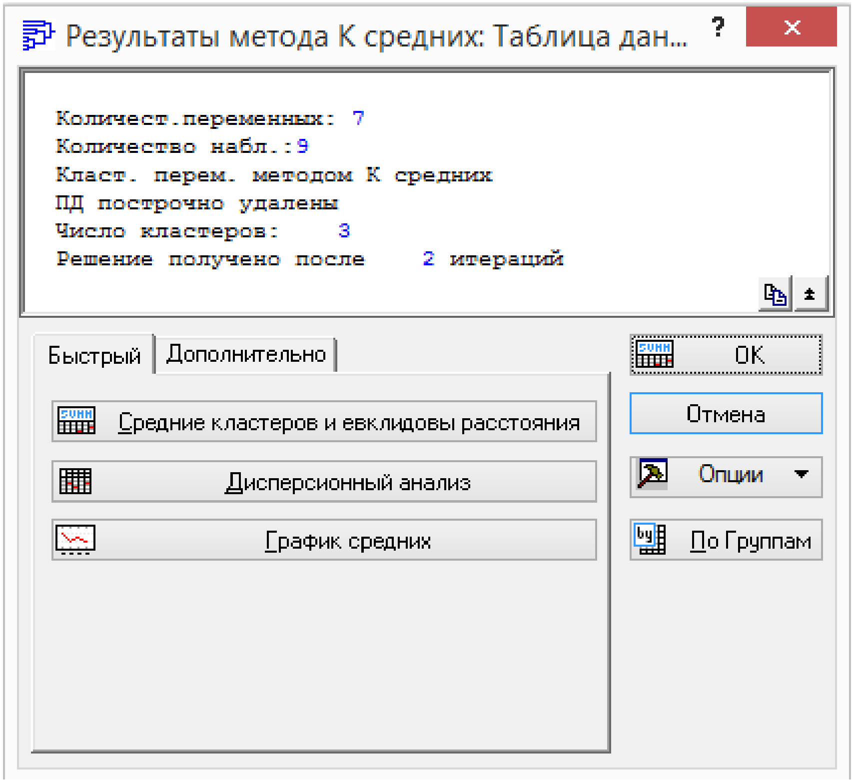



*Рис. 7.38. Результаты кластеризации методом k-средних*

График средних при трехкластерном разбиении примет вид (рис. 7.39):

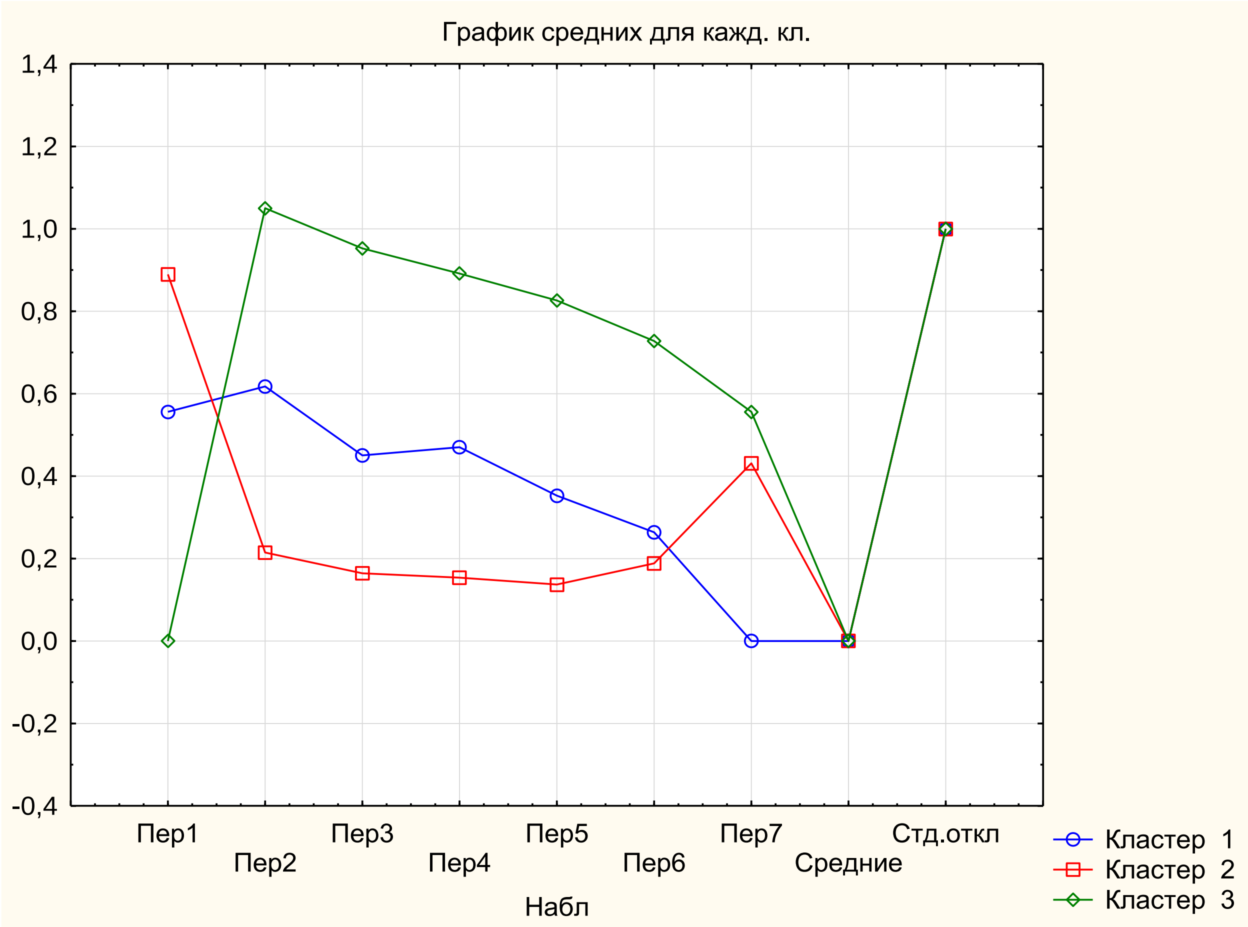



*Рис. 7.39. График средних кластеров*

Результаты дисперсионного анализа (рис. 7.40):

| Наблюд. | Между SS | сс | Внутри SS | сс | F | значим. p |
|---|---|---|---|---|---|---|
| | Дисперсионный анализ (Таблица данных4) | | | | | |
| **Пер1** | 0,688716 | 2 | 0,059991 | 4 | 22,96050 | 0,006420 |
| Пер2 | 0,641377 | 2 | 0,072755 | 4 | 17,63117 | 0,010379 |
| Пер3 | 0,537881 | 2 | 0,038638 | 4 | 27,84208 | 0,004492 |
| Пер4 | 0,485721 | 2 | 0,032662 | 4 | 29,74260 | 0,003970 |
| Пер5 | 0,404555 | 2 | 0,039625 | 4 | 20,41935 | 0,007958 |
| Пер6 | 0,242417 | 2 | 0,078179 | 4 | 6,20160 | 0,059465 |
| Пер7 | 0,187929 | 2 | 0,070751 | 4 | 5,31242 | 0,074806 |
| Средние | 0,000000 | 2 | 0,000000 | 4 | 2,11960 | 0,235695 |
| Стд.откл | 0,000000 | 2 | 0,000000 | 4 | 0,00000 | 1,000000 |

*Рис. 7.40. Результаты дисперсионного анализа*

Во вкладке «Дополнительно» можно рассмотреть другие характеристики результатов кластеризации методом k-средних (рис. 7.41):

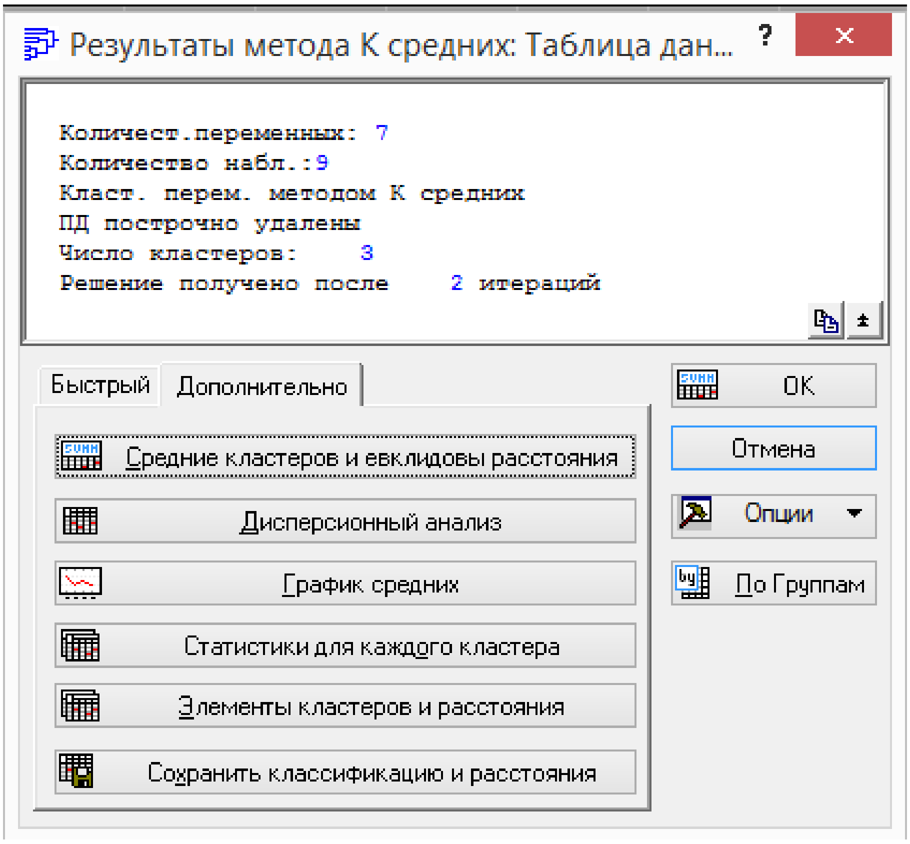



*Рис. 7.41. Дополнительные характеирстики анализа*

Рассмотренный в рамках данного параграфа массив исходных данных не может быть использован для кластеризации (ввиду небольшого объема исходных данных), а приведенные положения используются для демонстрации существующих возможностей программного продукта Statistica в процессе осуществления кластерного анализа. Более подробно пример кластеризации с последующей интерпретацией полученных значений приведен в примерах к главе 7.

## Контрольные вопросы

1. Чем обусловлена необходимость осуществления кластерного анализа в рамках анализа деятельности субъектов на мезоуровне?
2. В чем состоит сущность применения многомерных статистических метода при анализе деятельности фирмы?
3. В чем состоит сущность метода кластеризациию?

4. Как применение пакетов прикладных статистических программ оказывает влияние на процесс проведения многомерного анализа?
5. Какие программные продукты используются при проведении кластерного анализа?
6. В рамках каких исследовний применяются программные продукты STADIA, STATGRAPHICS, SPSS, STATISTICA?
7. Зачем программным продуктам необходимо обладать статистическим разнообразием для работы с исходными данными?
8. Какие исследовательские задачи могут быть решены благодаря широким графическим возможностям прикладных программных продуктов STADIA, STATGRAPHICS, SPSS, STATISTICA?
9. Объясните роль эргономичности интерфейса при выборе программного продукта для осуществления кластерного анализа.
10. В какой области был впервые применен метод кластерного анализа?
11. Что означает слово «cluster» в переводу с английского языка?
12. Кого можно назвать родоначальником кластерного исследования?
13. В каком году был выпущен первый труд, систематизирующий представления о предмете, сущности и базовых методах кластеризации?
14. Как можно охарактеризовать сущность кластерного анализа?
15. В чем состоит основное преимущество кластеризации перед другими методами исследования?
16. Для каких данных применим метод кластерного анализа?
17. Сохраняются ли существенные свойства и признаки массивов данных при осуществлении процедуры кластеризации?
18. Возможно ли применение кластерного анализа в сочетании с другими методами эконометрики?
19. Каковы возможности кластеризации при анализе рядов динамики?
20. Возможно ли циклическое проведение процедуры кластеризации? Для сего применяется данный прием?
21. Перечислите основные недостатки метода кластерного анализа
22. Соблюдение каких условий необходимо для успешного проведения процедуры кластеризации?
23. Сформулируйте основную задачу кластерного анализа

24. Что характеризует центр кластера?
25. Что представляет собой радиус кластера?
26. Какие методы определения размеров кластера Вам известны?
27. Как решается проблема неопределенности при осуществлении кластерного анализа?
28. В чем состоит сущность проблемы масштаба?
29. Охарактеризуйте стандартизацию как прием проведения исходных данных к сопоставимому виду
30. Как осуществляется стандартизация с использованием Z-шкалы?
31. Какие другие методы стандартизации Вам известны? Охарактеризуйте их
32. Каких результатов можно добиться за счет корректировки параметров с учетом коэффициента важности?
33. Какие методы могут применяться исследователем при корректировке параметров с учетом коэффициента важности?
34. Какие два основные положения должны учитываться при измерении близости объектов?
35. Что понимается под термином «метрика» в кластерном анализе?
36. Перечислите основные аксиомы метрики
37. Назовите условия меры близости (сходства) объектов.
38. В каких случаях целесообразно использование Евклидова расстояния в качестве метрики кластерного анализа?
39. Чего позволяет добиться использование в кластерном анализе в качестве меры близости квадрата евклидова расстояния?
40. С какой точки зрения значимо применение в расчётах обобщенного степенного расстояния?
41. Какой метод следует применять, если необходимо идентифицировать два объекта как различные в том случае, если они отличаются только по определенному критерию?
42. Возможно ли применение процедуры кластерного анализа по отношению к качественным данным?
43. По каким критериям могут быть классифицированы методы кластерного анализа?
44. Чем отличаются иерархические и неиерархические методы кластеризации?

45. Каким образом осуществляется кластеризация с использованием иерархических агломеративных методов?
46. В чем состоит принцип работы иерархических дивизимных методов?
47. В виде какого графического объекта отражаются результаты кластеризации?
48. В каком виде могут быть представлены дердрограммы?
49. В чем состоит сущность метода «ближайшего» соседа?
50. Как реализуется метод полной связи («дальнего» соседа)?
51. Какая мера используется в качестве кластерного расстояния при применении метода Варда?
52. В каком случае следует применять метод невзвешенного попарного среднего (метод невзвешенного попарного арифметического среднего)?
53. Чем отличается метод взвешенного попарного арифметического среднего от метода невзвешенного попарного арифметического среднего?
54. Объясните черты сходства и различия невзвешенного и взвешенного центроидного метода
55. Какие итерационные приемы кластеризации Вам известны?
56. Назовите преимущества и недостатки метода k-средних.
57. Каков алгоритм реализации метода k-средних?
58. В чем состоит суть принципа реализации алгоритма k-средних?
59. В каком случае процесс итерации при реализации метода k-средних прекращается?

## ПРИМЕРЫ РЕШЕНИЯ ЗАДАЧ

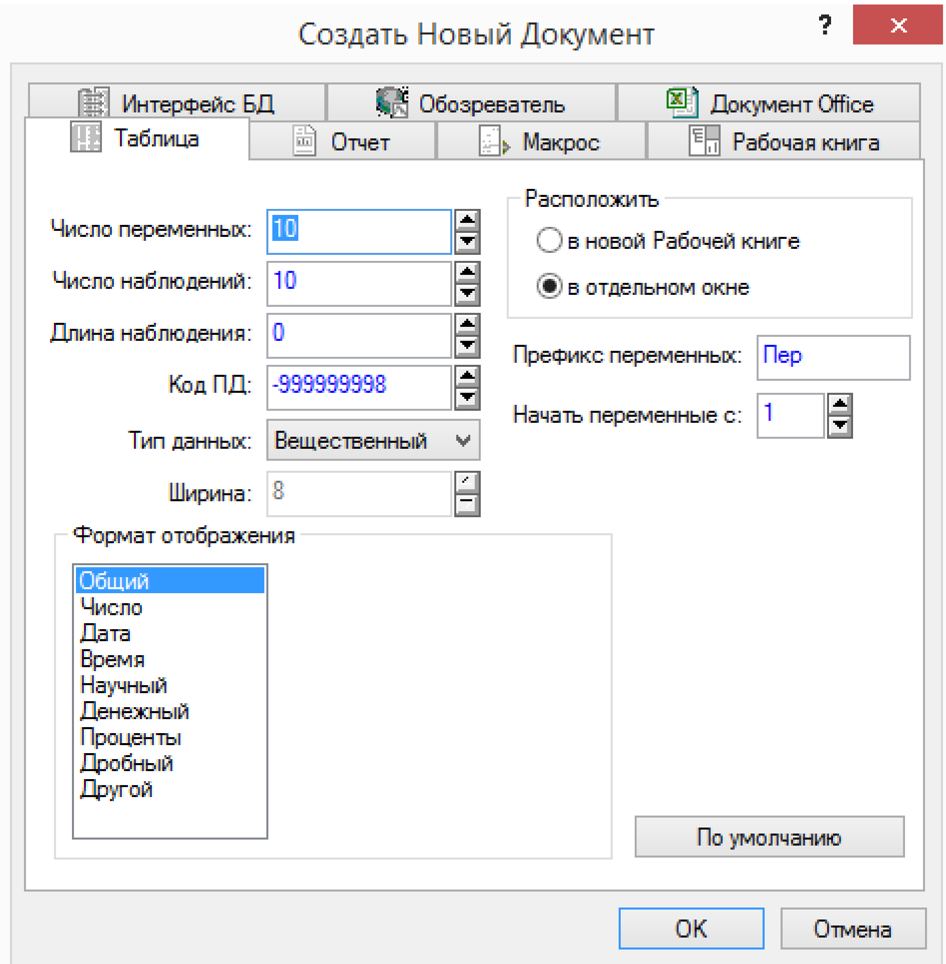
### Пример 7.1

Имеются данные о показателях, характеризующих применение 82 фирмами различных цифровых технологий (приложение 6 и 7).

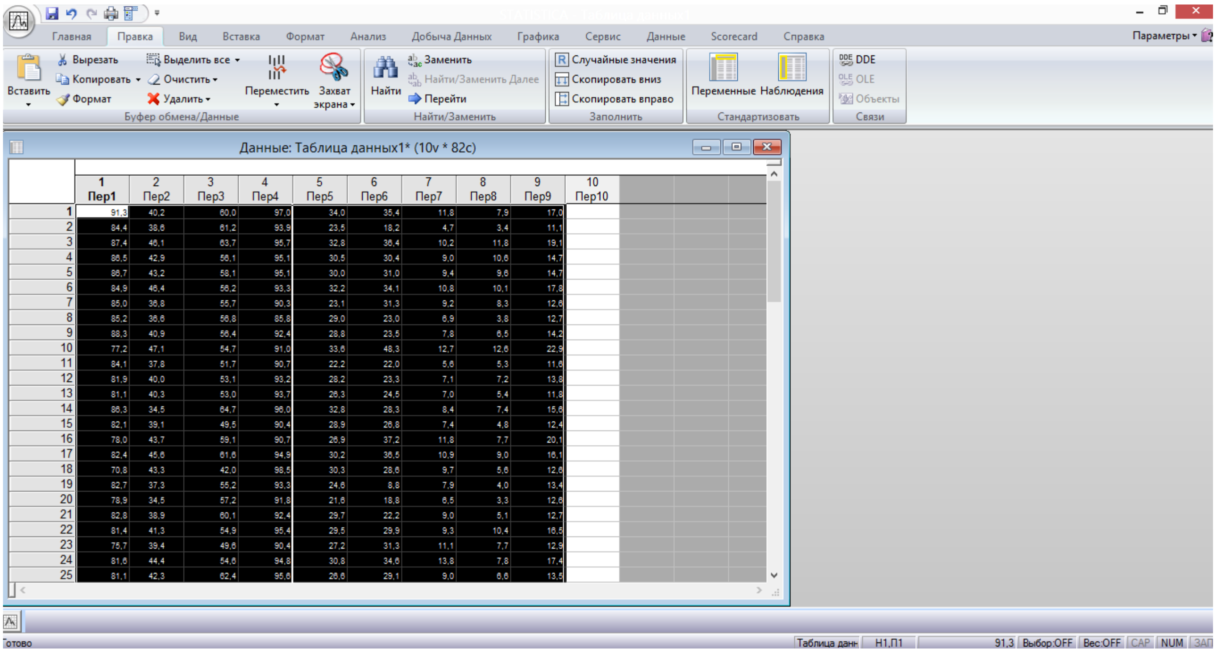
На основании представленных данных определите кластеры, используя Евклидово расстояние в качестве меры близости объектов. Сравните результаты иерархической кластеризации и группировки методом k-средних

**Решение.**

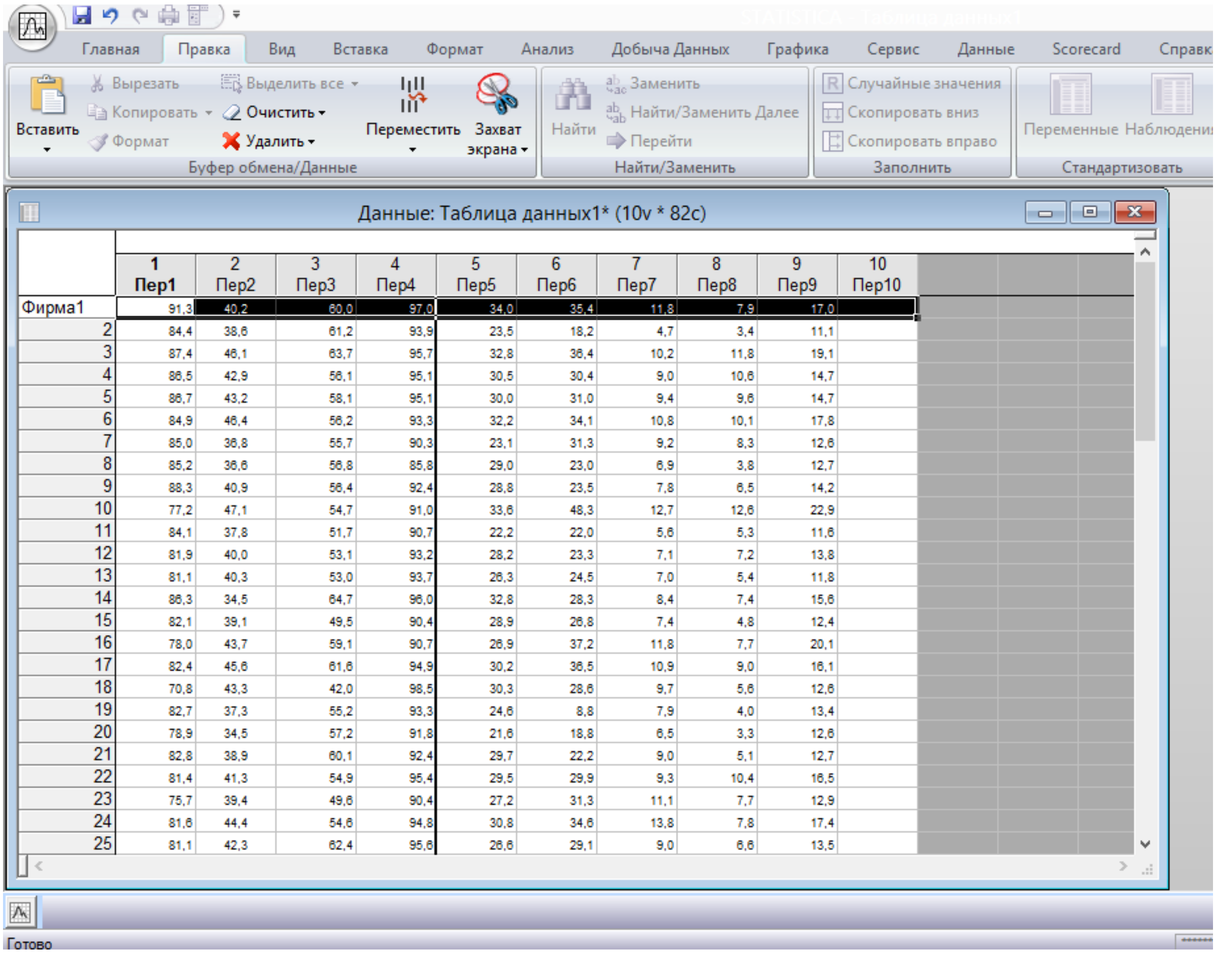
На первом этапе введем исходный массив данных в программу Statistica через команду «Создать». Для заданных данных заполнение диалогового окна создания нового документа будет выглядеть следующим образом:



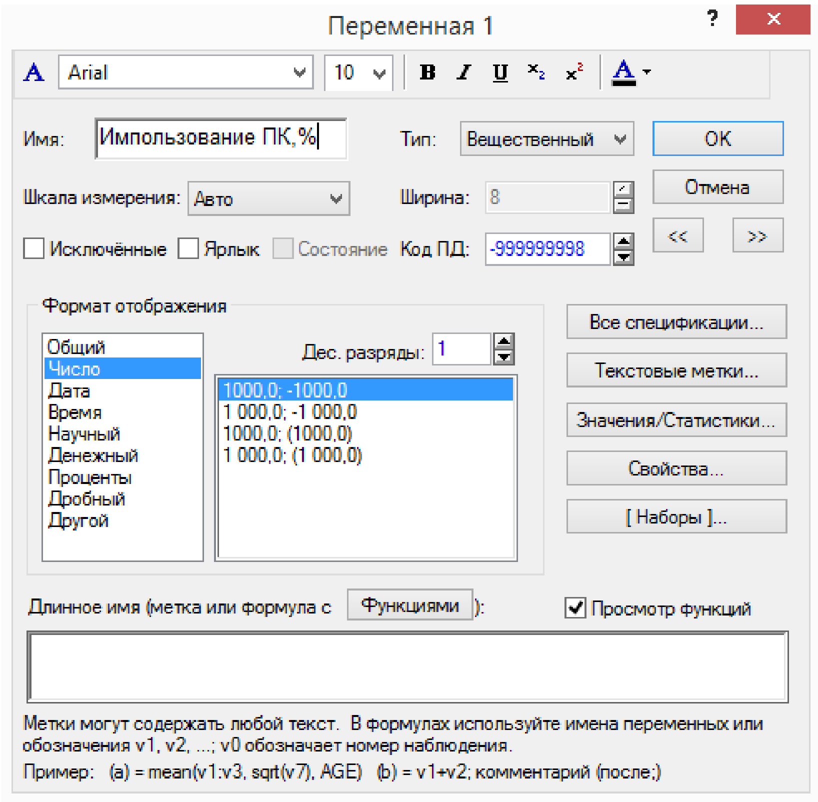
Также возможно копирование исходного массива из Excel:




Переименование наблюдений возможно после двойного щелчка левой кнопки мыши в конкретной ячейке:



Для переименования переменных необходимо после двойного щелчка левой кнопки задать желаемое название:



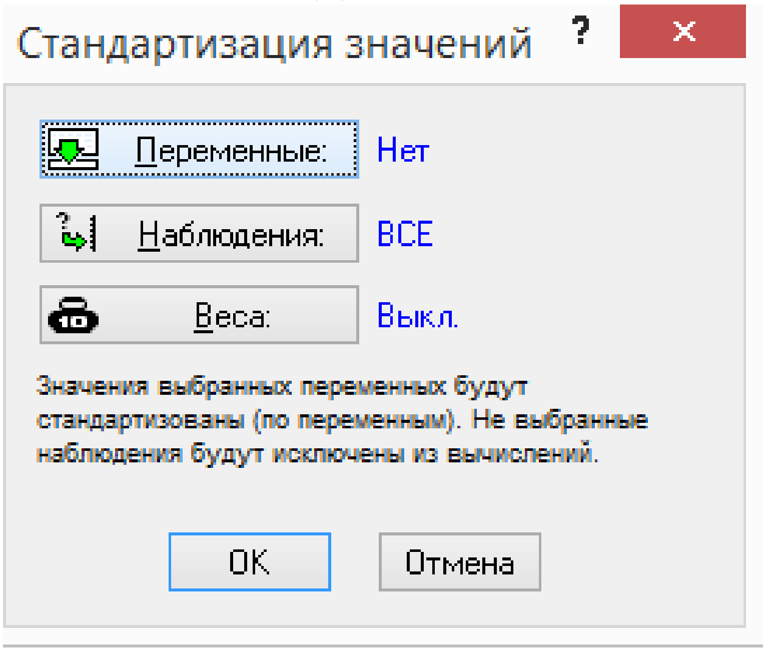

После переименования переменной 1 массив примет следующий вид:

Данные: Таблица данных1* (10v * 82c)

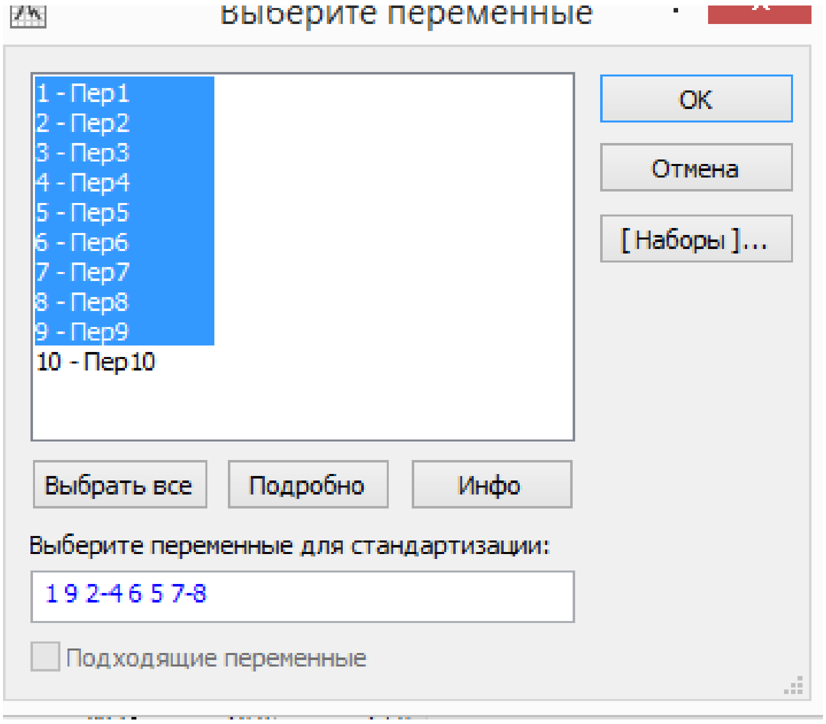
| | 1 Использование ПК,% | 2 Пер2 | 3 Пер3 | 4 Пер4 | 5 Пер5 | 6 Пер6 | 7 Пер7 | 8 Пер8 | 9 Пер9 | 10 Пер10 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 91,3 | 40,2 | 60,0 | 97,0 | 34,0 | 35,4 | 11,8 | 7,9 | 17,0 | |
| 2 | 84,4 | 38,6 | 61,2 | 93,9 | 23,5 | 18,2 | 4,7 | 3,4 | 11,1 | |
| 3 | 87,4 | 46,1 | 63,7 | 95,7 | 32,8 | 36,4 | 10,2 | 11,8 | 19,1 | |
| 4 | 86,5 | 42,9 | 56,1 | 95,1 | 30,5 | 30,4 | 9,0 | 10,6 | 14,7 | |
| 5 | 86,7 | 43,2 | 58,1 | 95,1 | 30,0 | 31,0 | 9,4 | 9,6 | 14,7 | |
| 6 | 84,9 | 46,4 | 56,2 | 93,3 | 32,2 | 34,1 | 10,8 | 10,1 | 17,8 | |
| 7 | 85,0 | 36,8 | 55,7 | 90,3 | 23,1 | 31,3 | 9,2 | 8,3 | 12,6 | |
| 8 | 85,2 | 36,6 | 56,8 | 85,8 | 29,0 | 23,0 | 6,9 | 3,8 | 12,7 | |
| 9 | 88,3 | 40,9 | 56,4 | 92,4 | 28,8 | 23,5 | 7,8 | 6,5 | 14,2 | |
| 10 | 77,2 | 47,1 | 54,7 | 91,0 | 33,6 | 48,3 | 12,7 | 12,6 | 22,9 | |
| 11 | 84,1 | 37,8 | 51,7 | 90,7 | 22,2 | 22,0 | 5,6 | 5,3 | 11,6 | |
| 12 | 81,9 | 40,0 | 53,1 | 93,2 | 28,2 | 23,3 | 7,1 | 7,2 | 13,8 | |
| 13 | 81,1 | 40,3 | 53,0 | 93,7 | 26,3 | 24,5 | 7,0 | 5,4 | 11,8 | |
| 14 | 86,3 | 34,5 | 64,7 | 96,0 | 32,8 | 28,3 | 8,4 | 7,4 | 15,6 | |
| 15 | 82,1 | 39,1 | 49,5 | 90,4 | 28,9 | 26,8 | 7,4 | 4,8 | 12,4 | |
| 16 | 78,0 | 43,7 | 59,1 | 90,7 | 26,9 | 37,2 | 11,8 | 7,7 | 20,1 | |
| 17 | 82,4 | 45,6 | 61,6 | 94,9 | 30,2 | 36,5 | 10,9 | 9,0 | 16,1 | |
| 18 | 70,8 | 43,3 | 42,0 | 98,5 | 30,3 | 28,6 | 9,7 | 5,6 | 12,6 | |
| 19 | 82,7 | 37,3 | 55,2 | 93,3 | 24,6 | 8,8 | 7,9 | 4,0 | 13,4 | |
| 20 | 78,9 | 34,5 | 57,2 | 91,8 | 21,6 | 18,8 | 6,5 | 3,3 | 12,6 | |
| 21 | 82,8 | 38,9 | 60,1 | 92,4 | 29,7 | 22,2 | 9,0 | 5,1 | 12,7 | |
| 22 | 81,4 | 41,3 | 54,9 | 95,4 | 29,5 | 29,9 | 9,3 | 10,4 | 16,5 | |
| 23 | 75,7 | 39,4 | 49,6 | 90,4 | 27,2 | 31,3 | 11,1 | 7,7 | 12,9 | |
| 24 | 81,6 | 44,4 | 54,6 | 94,8 | 30,8 | 34,6 | 13,8 | 7,8 | 17,4 | |

Аналогичным образом могут переименованы и другие параметры массива исходных данных.

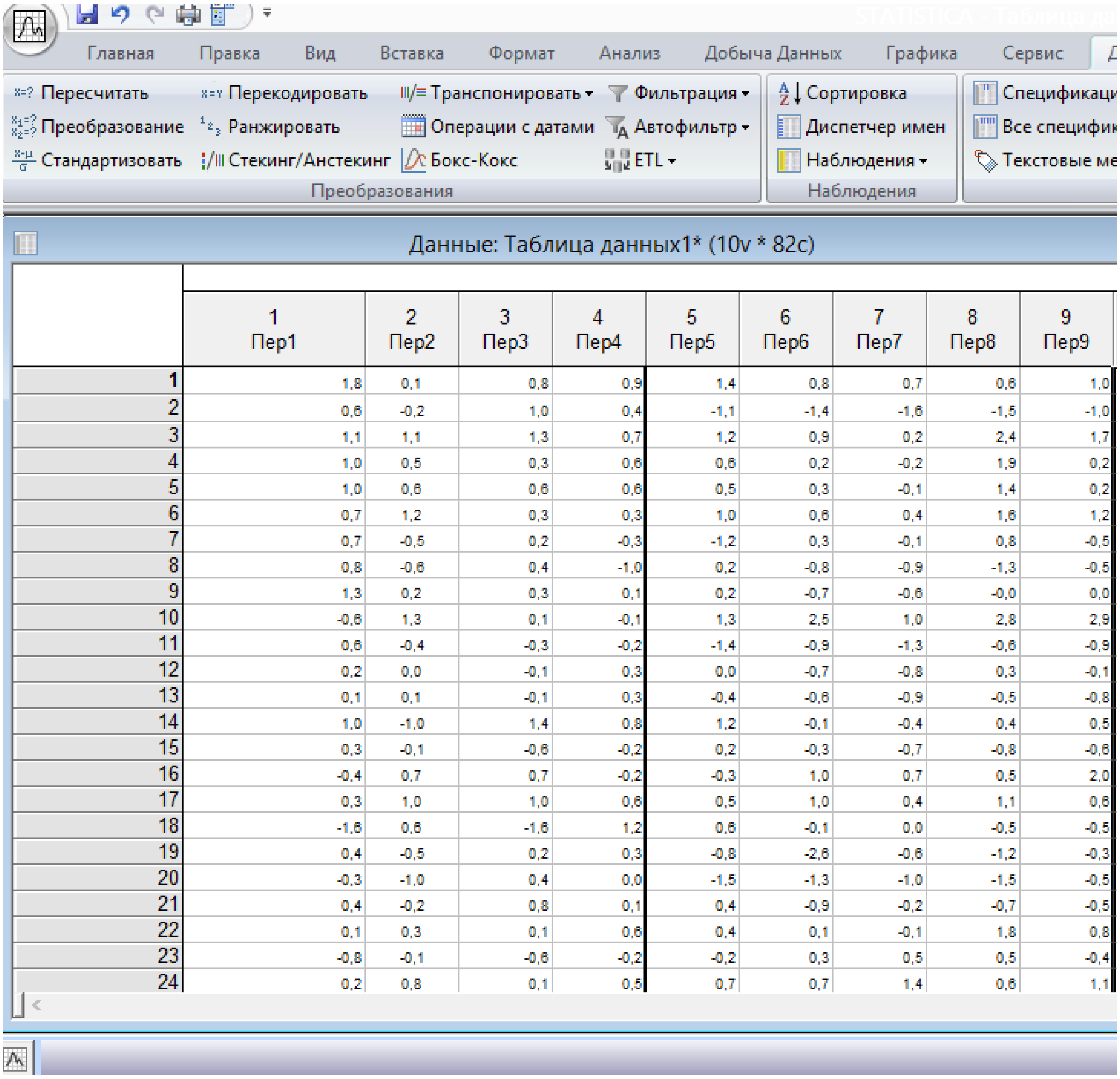
Затем данные необходимо стандартизировать, нажав кнопку «Стандартизация» во вкладке «Данные»:




Поскольку в данном примере 9 переменных, выделяем нужный массив:

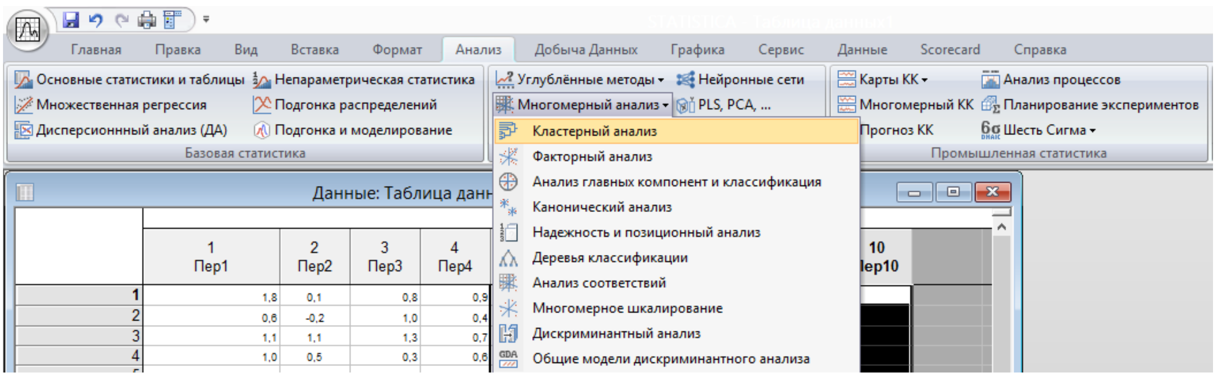



После проведения процедуры стандартизации массив будет выглядеть следующим образом:

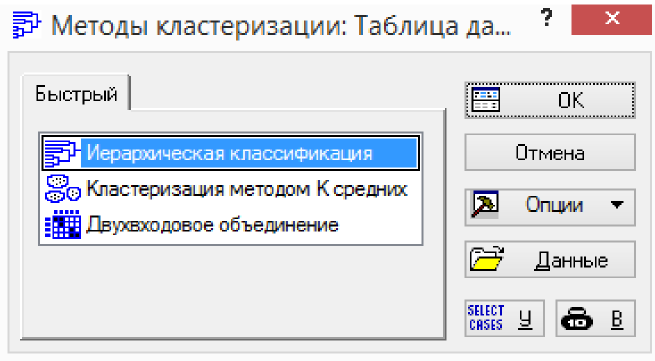



| | 1 Пер1 | 2 Пер2 | 3 Пер3 | 4 Пер4 | 5 Пер5 | 6 Пер6 | 7 Пер7 | 8 Пер8 | 9 Пер9 |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 1,8 | 0,1 | 0,8 | 0,9 | 1,4 | 0,8 | 0,7 | 0,6 | 1,0 |
| 2 | 0,6 | -0,2 | 1,0 | 0,4 | -1,1 | -1,4 | -1,6 | -1,5 | -1,0 |
| 3 | 1,1 | 1,1 | 1,3 | 0,7 | 1,2 | 0,9 | 0,2 | 2,4 | 1,7 |
| 4 | 1,0 | 0,5 | 0,3 | 0,6 | 0,6 | 0,2 | -0,2 | 1,9 | 0,2 |
| 5 | 1,0 | 0,6 | 0,6 | 0,6 | 0,5 | 0,3 | -0,1 | 1,4 | 0,2 |
| 6 | 0,7 | 1,2 | 0,3 | 0,3 | 1,0 | 0,6 | 0,4 | 1,6 | 1,2 |
| 7 | 0,7 | -0,5 | 0,2 | -0,3 | -1,2 | 0,3 | -0,1 | 0,8 | -0,5 |
| 8 | 0,8 | -0,6 | 0,4 | -1,0 | 0,2 | -0,8 | -0,9 | -1,3 | -0,5 |
| 9 | 1,3 | 0,2 | 0,3 | 0,1 | 0,2 | -0,7 | -0,6 | -0,0 | 0,0 |
| 10 | -0,6 | 1,3 | 0,1 | -0,1 | 1,3 | 2,5 | 1,0 | 2,8 | 2,9 |
| 11 | 0,6 | -0,4 | -0,3 | -0,2 | -1,4 | -0,9 | -1,3 | -0,6 | -0,9 |
| 12 | 0,2 | 0,0 | -0,1 | 0,3 | 0,0 | -0,7 | -0,8 | 0,3 | -0,1 |
| 13 | 0,1 | 0,1 | -0,1 | 0,3 | -0,4 | -0,6 | -0,9 | -0,5 | -0,8 |
| 14 | 1,0 | -1,0 | 1,4 | 0,8 | 1,2 | -0,1 | -0,4 | 0,4 | 0,5 |
| 15 | 0,3 | -0,1 | -0,6 | -0,2 | 0,2 | -0,3 | -0,7 | -0,8 | -0,6 |
| 16 | -0,4 | 0,7 | 0,7 | -0,2 | -0,3 | 1,0 | 0,7 | 0,5 | 2,0 |
| 17 | 0,3 | 1,0 | 1,0 | 0,6 | 0,5 | 1,0 | 0,4 | 1,1 | 0,6 |
| 18 | -1,6 | 0,6 | -1,6 | 1,2 | 0,6 | -0,1 | 0,0 | -0,5 | -0,5 |
| 19 | 0,4 | -0,5 | 0,2 | 0,3 | -0,8 | -2,6 | -0,6 | -1,2 | -0,3 |
| 20 | -0,3 | -1,0 | 0,4 | 0,0 | -1,5 | -1,3 | -1,0 | -1,5 | -0,5 |
| 21 | 0,4 | -0,2 | 0,8 | 0,1 | 0,4 | -0,9 | -0,2 | -0,7 | -0,5 |
| 22 | 0,1 | 0,3 | 0,1 | 0,6 | 0,4 | 0,1 | -0,1 | 1,8 | 0,8 |
| 23 | -0,8 | -0,1 | -0,6 | -0,2 | -0,2 | 0,3 | 0,5 | 0,5 | -0,4 |
| 24 | 0,2 | 0,8 | 0,1 | 0,5 | 0,7 | 0,7 | 1,4 | 0,6 | 1,1 |

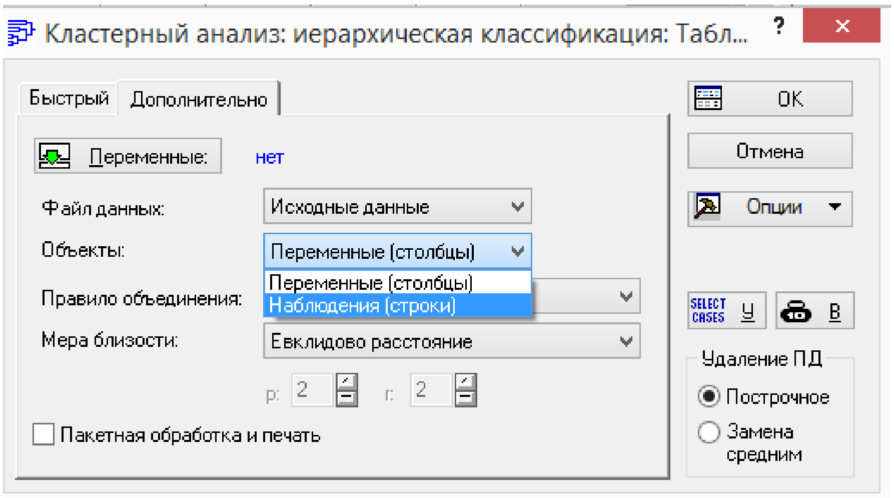
Непосредственно для начала осуществления кластерного анализа необходимо выбрать команду Анализ/Многомерный анализ/Кластерный анализ:



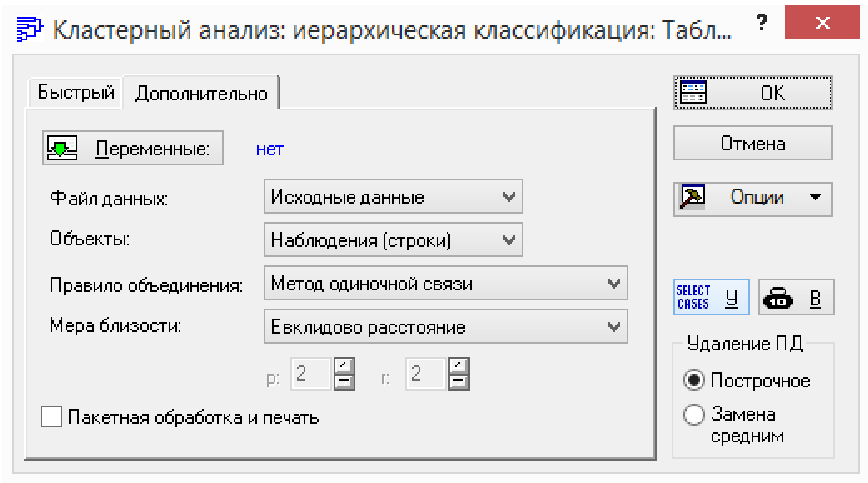
Для выбора иерархической классификации выбирается соответственный вариант в диалоговом окне:



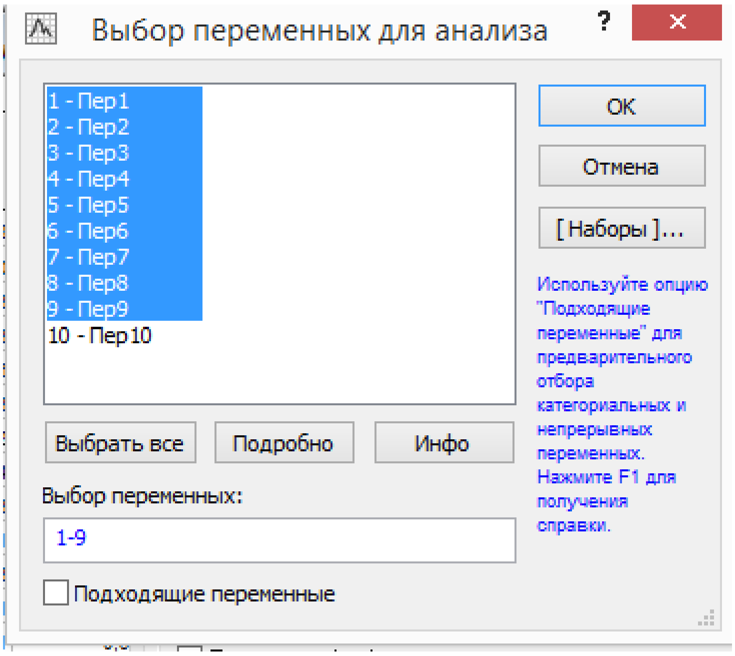

Поскольку в данном примере объектом классификации являются фирмы, а не признаки цифрового развития, во вкладке «Дополнительно» был изменен объект анализа:



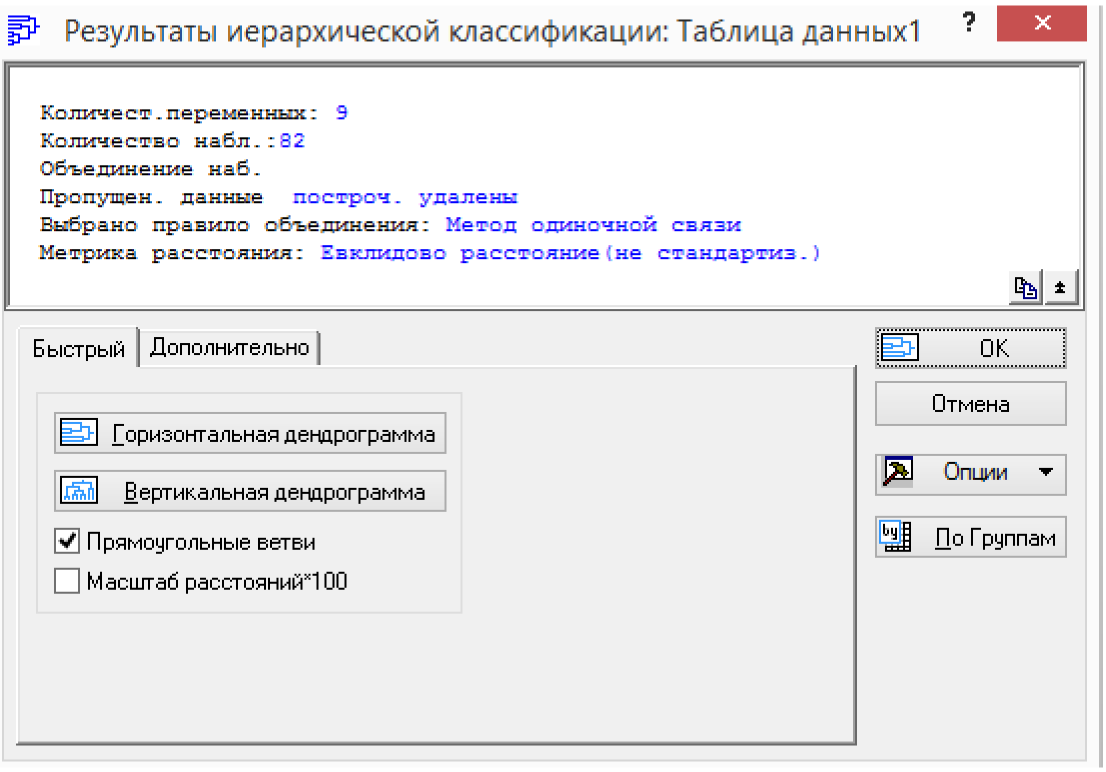

В соответствии с заданными условиями кластеризации правило объединения и мера близости остается без изменения:



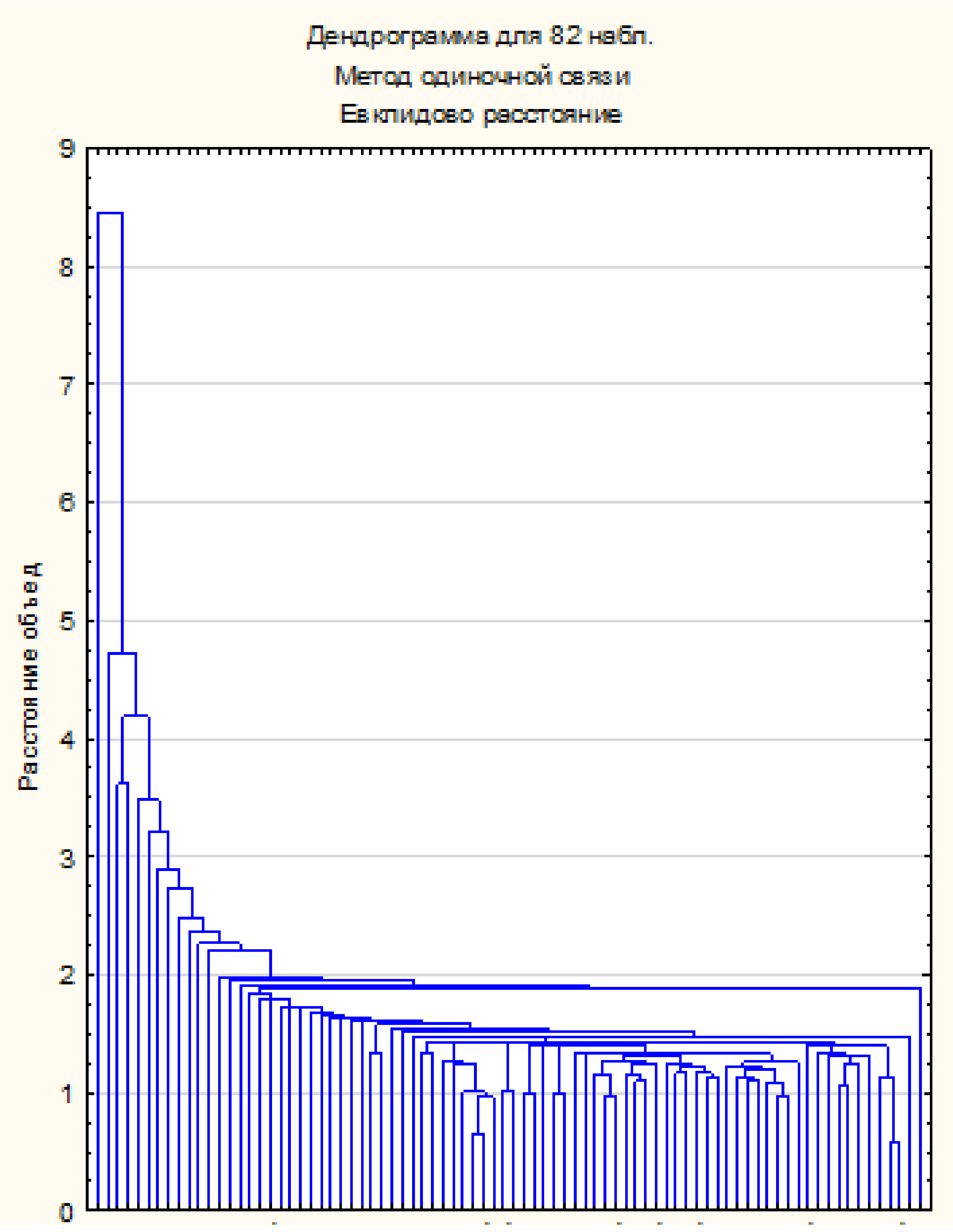
После нажатия кнопки «Ок» открывается поле выбора переменных (в исходном примере 1-9):



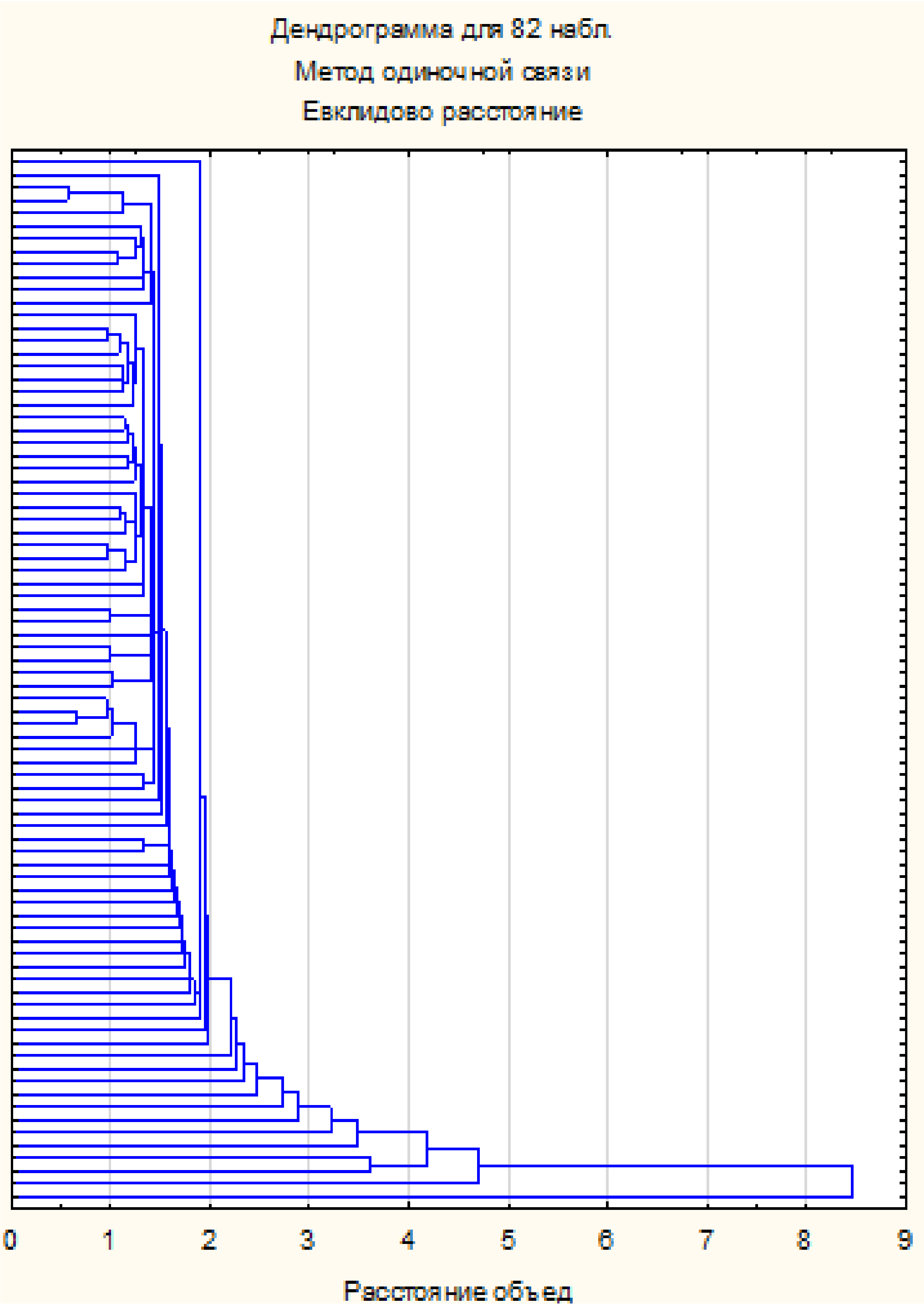

Результаты иерархической классификации отображаются в следующем окне:



Вертикальная дендрограмма, построенная по результатам кластерного анализа, будет иметь вид:



Горизонтальная дендрограмма отличается от вертикальной только расположением дерева кластера относительно осей:



Оптимальное число групп (интервалов) может быть определено по формуле Стерджесса:

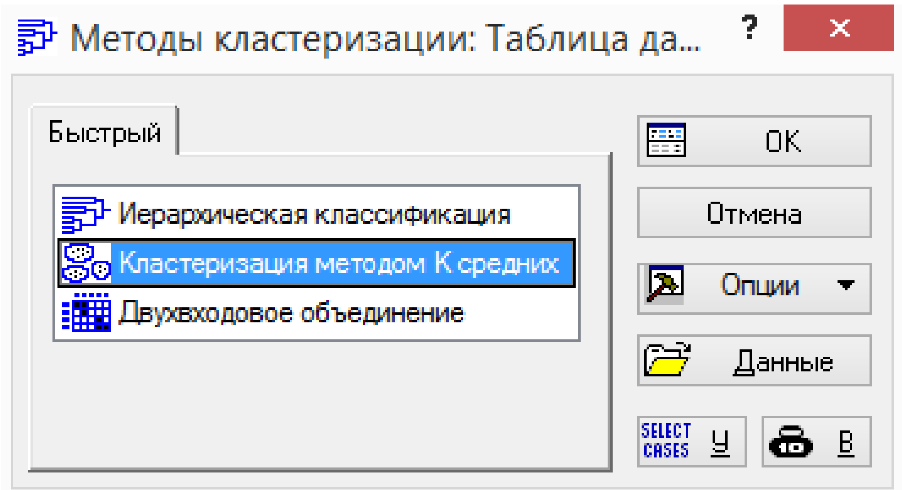
$$K = 1+3.322 \lg n,$$

где n - число единиц совокупности

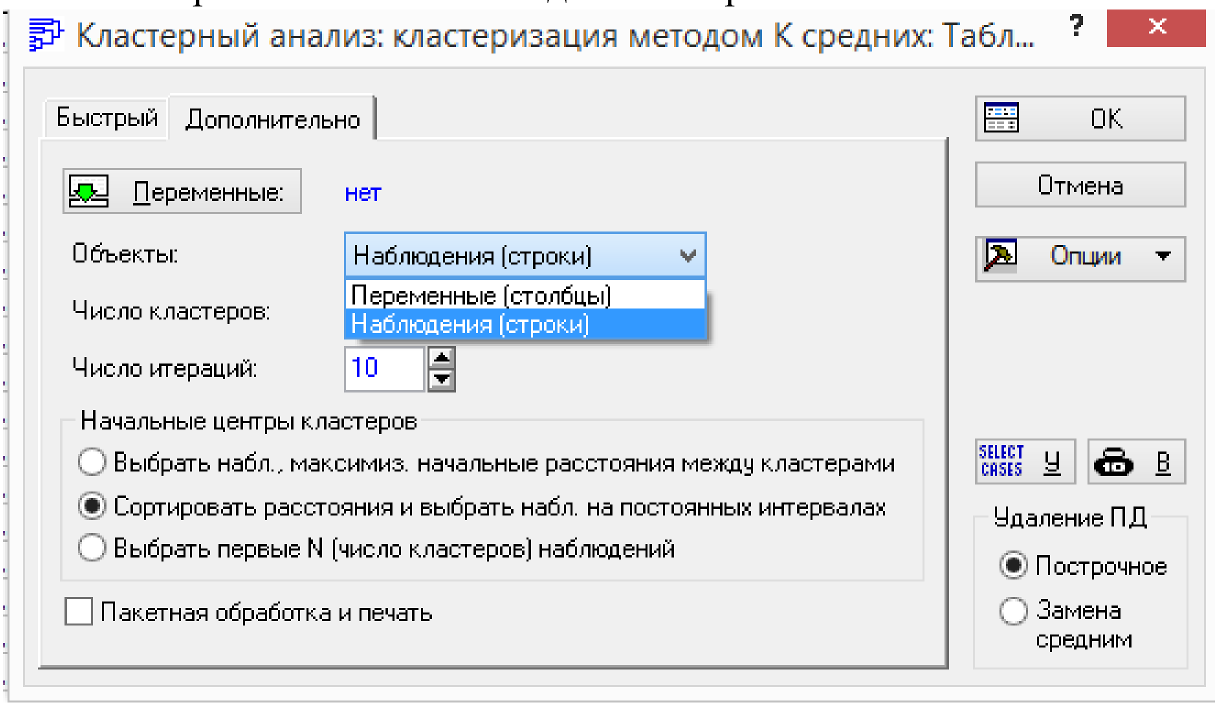
Для 82 наблюдений:

$$K = 1+3.322 \lg 82 \approx 7$$

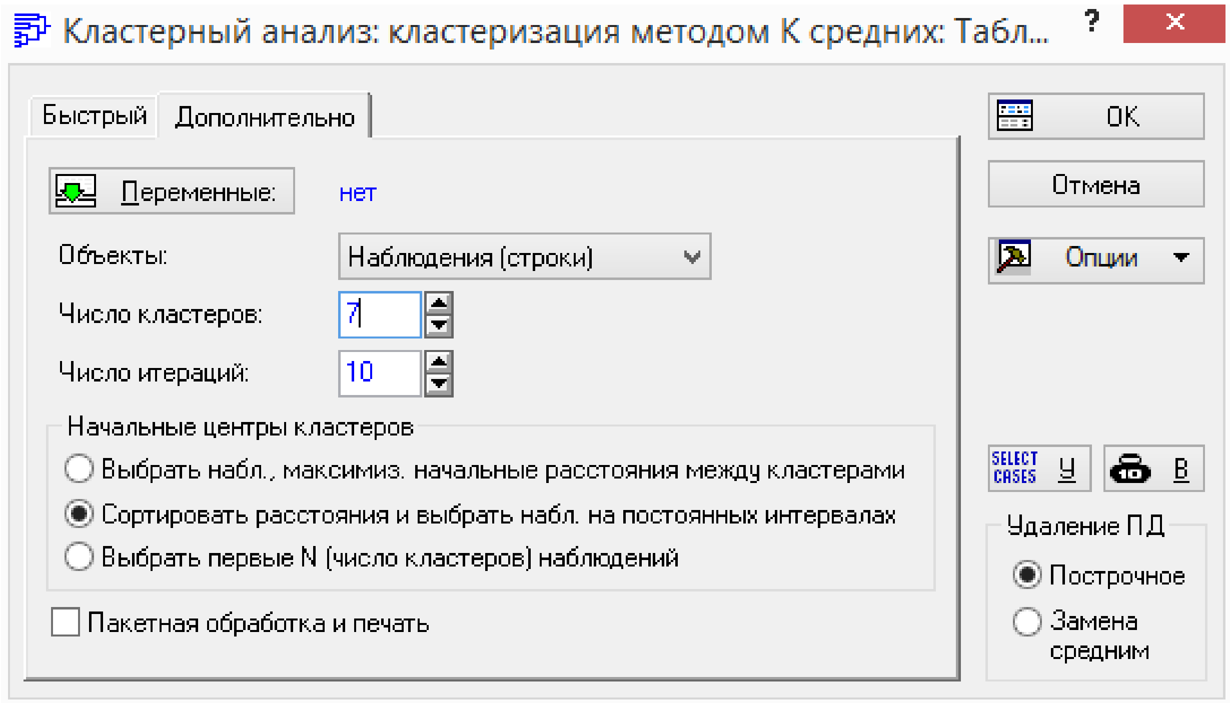
Проверим гипотезу о том, что 82 фирмы целесообразно разбить на 7 кластеров методом k-средних.



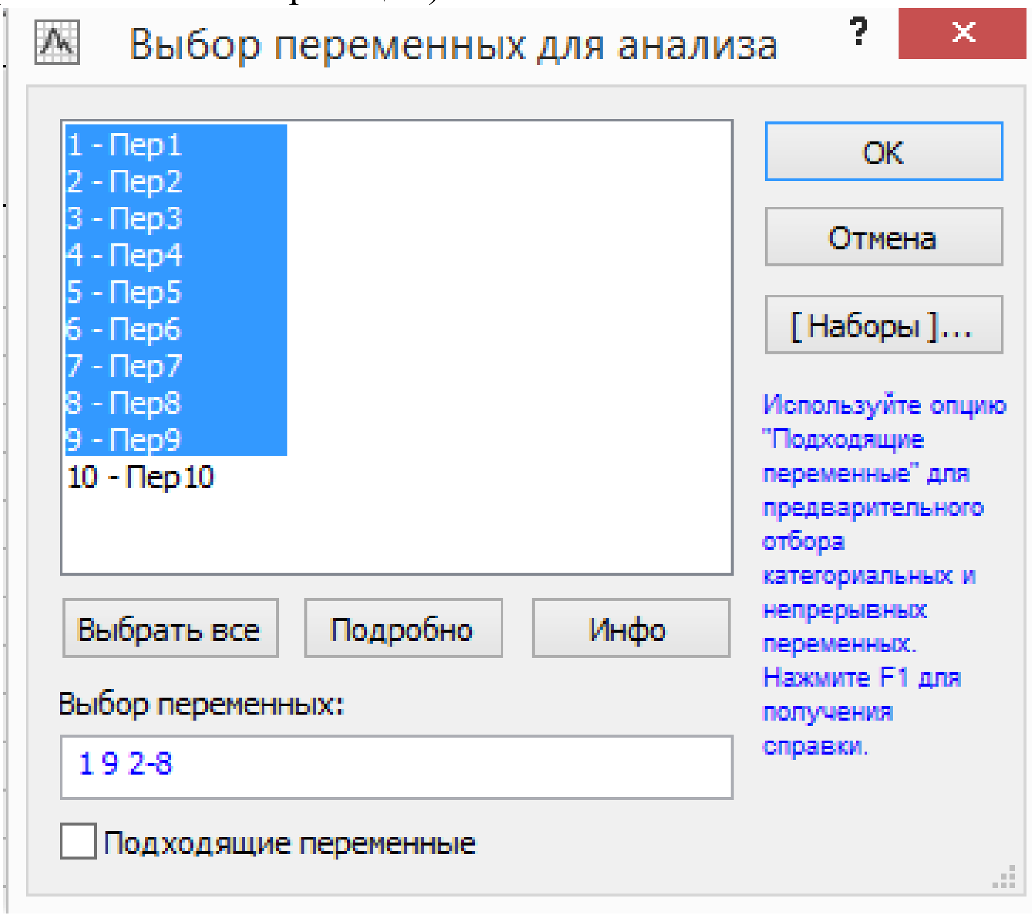

Выбирается объект наблюдений – строки:



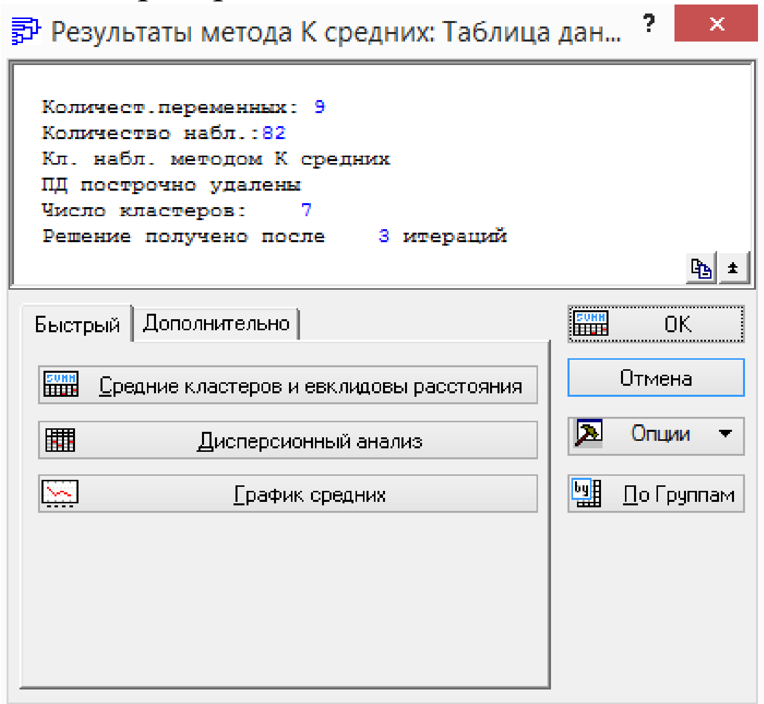

Число кластеров определяется равным 7, итерации остаются без изменений:




Затем также выбираются переменные для анализа (как и при иерархической кластеризации):



Результаты кластеризации будут отображены на появившемся диалоговом пространстве:




Средние кластеров по итогам разбиения фирм на 7 кластеров:

| перемен. | Средн.класт. (Таблица данных1) Кластер Но. 1 | Кластер Но. 2 | Кластер Но. 3 | Кластер Но. 4 | Кластер Но. 5 | Кластер Но. 6 | Кластер Но. 7 |
|---|---|---|---|---|---|---|---|
| **Пер1** | 0,320580 | -0,330278 | 0,497450 | 0,751342 | -0,10730 | -1,34963 | -3,42826 |
| Пер9 | 0,873779 | -0,130030 | 0,041997 | 1,251814 | -0,91273 | -1,47731 | -1,55609 |
| Пер2 | 1,009038 | -0,352792 | 0,275533 | 0,852756 | -0,58451 | 1,10257 | -2,97125 |
| Пер3 | 0,572725 | -0,406028 | 0,570291 | 0,650641 | -0,22131 | 0,12836 | -3,61685 |
| Пер4 | 0,187642 | -0,074128 | 0,345567 | 0,330155 | -0,09845 | -3,59776 | -1,54280 |
| Пер5 | 0,431786 | -0,295878 | 0,276627 | 0,970073 | -0,84832 | 3,97484 | -1,10274 |
| Пер6 | 1,003116 | 0,213688 | -0,424273 | 0,816765 | -1,11656 | 3,69928 | -0,71189 |
| Пер7 | 0,798880 | 0,026246 | -0,107668 | 0,351295 | -0,93380 | 6,57651 | -0,53325 |
| Пер8 | -0,325109 | 0,172068 | -0,257568 | 1,541505 | -0,78017 | 2,46272 | -0,94463 |

Полученные Евклидовы расстояния для межкластерного деления:

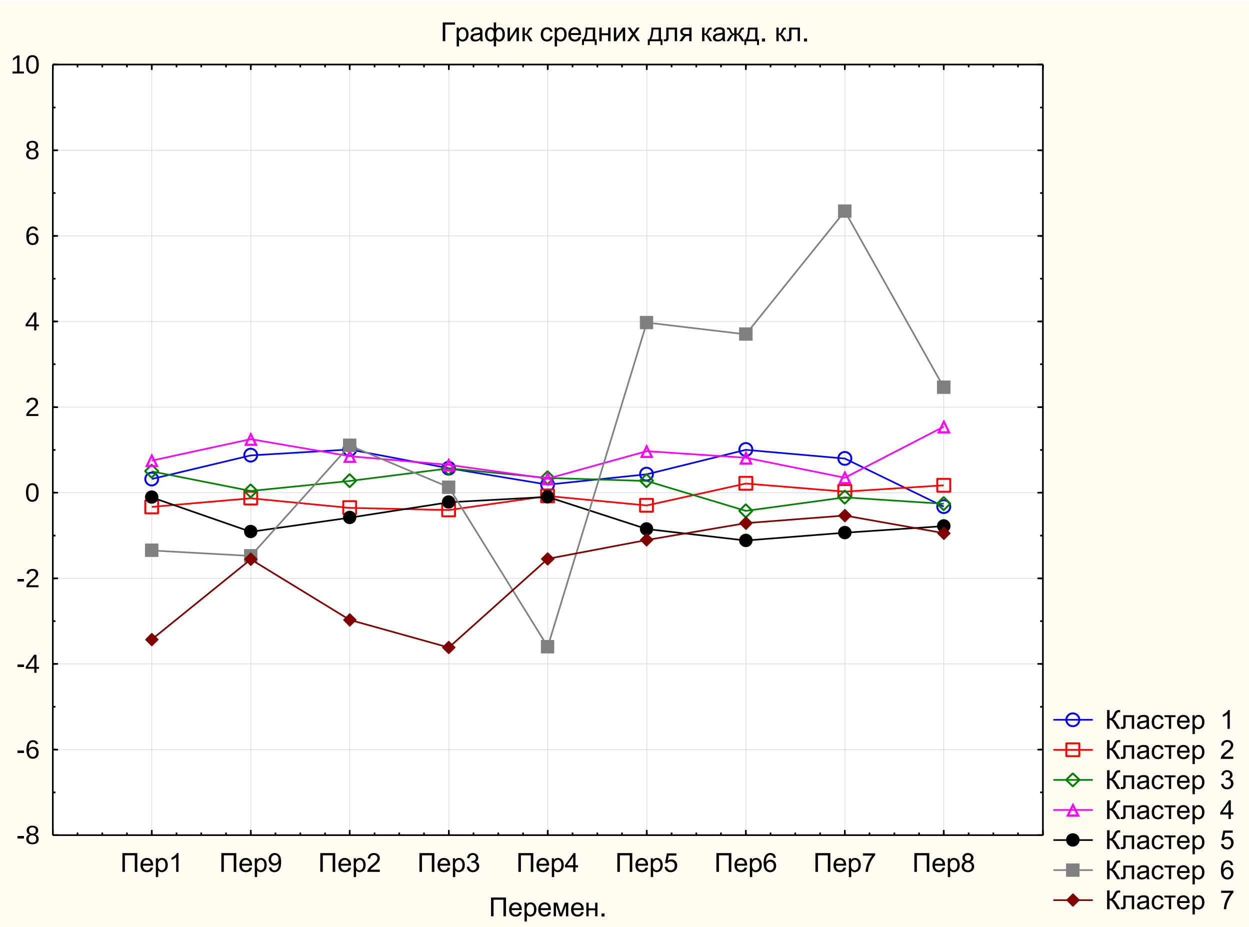
| Кластер Номер | Евклидовы расст. между кластерами (Таблица данных1) Расстояния под диагональю Квадраты расстояний над диагональю Но. 1 | Но. 2 | Но. 3 | Но. 4 | Но. 5 | Но. 6 | Но. 7 |
|---|---|---|---|---|---|---|---|
| **Но. 1** | 0,000000 | 0,701017 | 0,463780 | 0,487592 | 1,774156 | 9,31425 | 7,08912 |
| Но. 2 | 0,837267 | 0,000000 | 0,352904 | 1,084454 | 0,517106 | 10,69059 | 3,78017 |
| Но. 3 | 0,681014 | 0,594057 | 0,000000 | 0,815151 | 0,615674 | 11,65643 | 5,80481 |
| Но. 4 | 0,698278 | 1,041371 | 0,902857 | 0,000000 | 2,502023 | 9,39615 | 8,36583 |
| Но. 5 | 1,331975 | 0,719101 | 0,784649 | 1,581779 | 0,000000 | 14,49469 | 3,46345 |
| Но. 6 | 3,051926 | 3,269647 | 3,414151 | 3,065315 | 3,807189 | 0,00000 | 16,28570 |
| Но. 7 | 2,662540 | 1,944267 | 2,409317 | 2,892374 | 1,861033 | 4,03555 | 0,00000 |

Результаты дисперсионного анализа:

| | Дисперсионный анализ (Таблица данных1) | | | | | |
|---|---|---|---|---|---|---|
| перемен. | Между SS | сс | Внутри SS | сс | F | значим. p |
| **Пер1** | 51,44237 | 6 | 29,55763 | 75 | 21,75511 | 0,000000 |
| Пер9 | 52,02806 | 6 | 28,97194 | 75 | 22,44761 | 0,000000 |
| Пер2 | 57,60490 | 6 | 23,39510 | 75 | 30,77830 | 0,000000 |
| Пер3 | 57,49674 | 6 | 23,50326 | 75 | 30,57913 | 0,000000 |
| Пер4 | 23,98046 | 6 | 57,01954 | 75 | 5,25707 | 0,000145 |
| Пер5 | 48,80720 | 6 | 32,19280 | 75 | 18,95113 | 0,000000 |
| Пер6 | 60,55880 | 6 | 20,44120 | 75 | 37,03231 | 0,000000 |
| Пер7 | 68,50046 | 6 | 12,49954 | 75 | 68,50297 | 0,000000 |
| Пер8 | 51,05866 | 6 | 29,94134 | 75 | 21,31612 | 0,000000 |

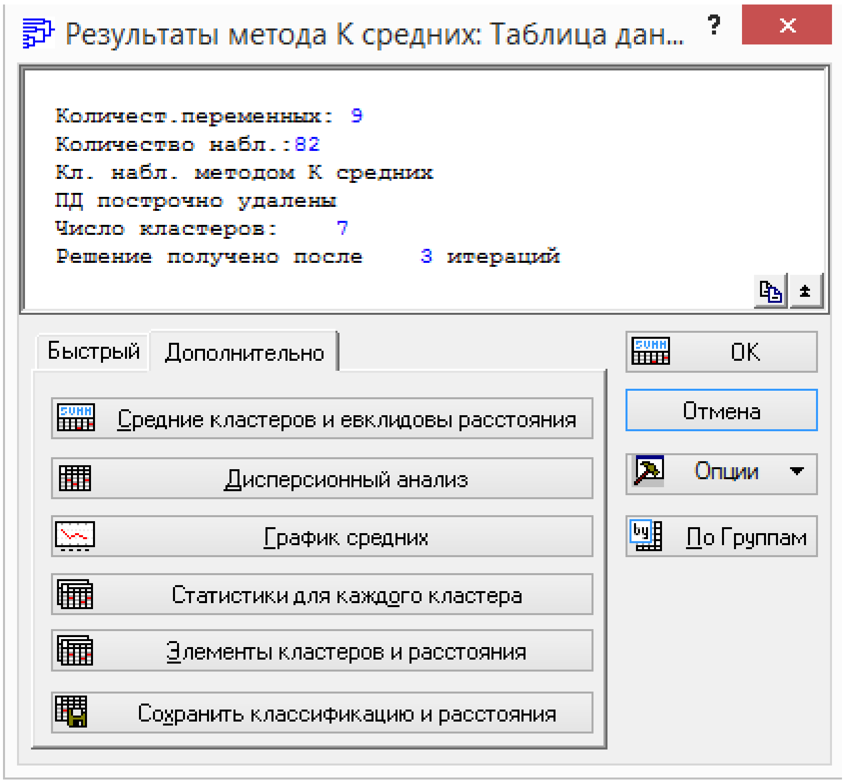
Поскольку p-значение меньше 5%, деление совокупности фирм на 7 кластеров можно считать значимым.

Подробнее график средних характеристик каждого их кластеров:




По результатам анализа данного графика можно, например, сделать вывод о том, что использование Интернета в фирмах 6 кластера является наименьшим по сравнению с другими организациями, однако в фирмах данной группы наибольшая доля используемых передовых производственных технологий, корпоративных сетей, электоронного документооборота и максимальный объем выручки от использования передовых производственных технологий.

Дополнительные характеристики кластеризации методом k-средних включают статистики для каждого кластера, элементы кластеров и расстояния, возможность сохранения классификации и расстояния:




Статистики для каждого кластера содержат информацию о средних по каждой из переменных, входящих в состав группы, величину дисперсии и стандартного отклонения:

| | Описат.статистики для кластера 1 (Таблица данных1) Кластер содержит 11 набл. | | |
|---|---|---|---|
| перемен. | **Среднее** | Стандарт отклон. | Дисперс. |
| **Пер1** | 0,320580 | 0,522744 | 0,273261 |
| Пер9 | 0,873779 | 0,793706 | 0,629969 |
| Пер2 | 1,009038 | 0,754524 | 0,569307 |
| Пер3 | 0,572725 | 0,466862 | 0,217961 |
| Пер4 | 0,187642 | 0,512410 | 0,262564 |
| Пер5 | 0,431786 | 0,817779 | 0,668763 |
| Пер6 | 1,003116 | 0,368238 | 0,135599 |
| Пер7 | 0,798880 | 0,444195 | 0,197309 |
| Пер8 | -0,325109 | 0,833872 | 0,695343 |

Кластер 2:

| перемен. | Описат.статистики для кластера 2 (Таблица данных1) Кластер содержит 21 набл. | | |
|---|---|---|---|
| | Среднее | Стандарт отклон. | Дисперс. |
| Пер1 | -0,330278 | 0,558456 | 0,311873 |
| Пер9 | -0,130030 | 0,559875 | 0,313460 |
| Пер2 | -0,352792 | 0,461953 | 0,213401 |
| Пер3 | -0,406028 | 0,514961 | 0,265185 |
| Пер4 | -0,074128 | 0,915885 | 0,838845 |
| Пер5 | -0,295878 | 0,478942 | 0,229386 |
| Пер6 | 0,213688 | 0,401990 | 0,161596 |
| Пер7 | 0,026246 | 0,413269 | 0,170791 |
| Пер8 | 0,172068 | 0,521994 | 0,272478 |

Кластер 3:

| перемен. | Описат.статистики для кластера 3 (Таблица данных1) Кластер содержит 16 набл. | | |
|---|---|---|---|
| | Среднее | Стандарт отклон. | Дисперс. |
| Пер1 | 0,497450 | 0,409204 | 0,167448 |
| Пер9 | 0,041997 | 0,472673 | 0,223420 |
| Пер2 | 0,275533 | 0,552858 | 0,305652 |
| Пер3 | 0,570291 | 0,518405 | 0,268744 |
| Пер4 | 0,345567 | 0,441212 | 0,194668 |
| Пер5 | 0,276627 | 0,792769 | 0,628483 |
| Пер6 | -0,424273 | 0,481165 | 0,231520 |
| Пер7 | -0,107668 | 0,403414 | 0,162743 |
| Пер8 | -0,257568 | 0,719581 | 0,517797 |

Кластер 4:

| перемен. | Описат.статистики для кластера 4 (Таблица данных1) Кластер содержит 12 набл. | | |
|---|---|---|---|
| | Среднее | Стандарт отклон. | Дисперс. |
| Пер1 | 0,751342 | 0,639493 | 0,408951 |
| Пер9 | 1,251814 | 0,803210 | 0,645146 |
| Пер2 | 0,852756 | 0,569600 | 0,324444 |
| Пер3 | 0,650641 | 0,446240 | 0,199130 |
| Пер4 | 0,330155 | 0,558116 | 0,311494 |
| Пер5 | 0,970073 | 0,706622 | 0,499315 |
| Пер6 | 0,816765 | 0,701874 | 0,492627 |
| Пер7 | 0,351295 | 0,394306 | 0,155477 |
| Пер8 | 1,541505 | 0,629413 | 0,396161 |

Кластер 5:

| перемен. | Описат.статистики для кластера 5 (Таблица данных1) Кластер содержит 18 набл. Среднее | Стандарт отклон. | Дисперс. |
|---|---|---|---|
| **Пер1** | -0,10730 | 0,749005 | 0,561008 |
| Пер9 | -0,91273 | 0,583841 | 0,340870 |
| Пер2 | -0,58451 | 0,544342 | 0,296308 |
| Пер3 | -0,22131 | 0,756447 | 0,572211 |
| Пер4 | -0,09845 | 0,643446 | 0,414023 |
| Пер5 | -0,84832 | 0,566875 | 0,321347 |
| Пер6 | -1,11656 | 0,546607 | 0,298779 |
| Пер7 | -0,93380 | 0,319547 | 0,102110 |
| Пер8 | -0,78017 | 0,540545 | 0,292189 |

Кластер 6:

| перемен. | Описат.статистики для кластера 6 (Таблица данных1) Кластер содержит 1 набл. Среднее | Стандарт отклон. | Дисперс. |
|---|---|---|---|
| **Пер1** | -1,34963 | 0,00 | 0,00 |
| Пер9 | -1,47731 | 0,00 | 0,00 |
| Пер2 | 1,10257 | 0,00 | 0,00 |
| Пер3 | 0,12836 | 0,00 | 0,00 |
| Пер4 | -3,59776 | 0,00 | 0,00 |
| Пер5 | 3,97484 | 0,00 | 0,00 |
| Пер6 | 3,69928 | 0,00 | 0,00 |
| Пер7 | 6,57651 | 0,00 | 0,00 |
| Пер8 | 2,46272 | 0,00 | 0,00 |

Кластер 7:

| перемен. | Описат.статистики для кластера 7 (Таблица данных1) Кластер содержит 3 набл. Среднее | Стандарт отклон. | Дисперс. |
|---|---|---|---|
| **Пер1** | -3,42826 | 1,421310 | 2,02012 |
| Пер9 | -1,55609 | 0,283148 | 0,08017 |
| Пер2 | -2,97125 | 0,348646 | 0,12155 |
| Пер3 | -3,61685 | 0,188121 | 0,03539 |
| Пер4 | -1,54280 | 3,480817 | 12,11609 |
| Пер5 | -1,10274 | 0,517127 | 0,26742 |
| Пер6 | -0,71189 | 0,970141 | 0,94117 |
| Пер7 | -0,53325 | 0,782099 | 0,61168 |
| Пер8 | -0,94463 | 0,472452 | 0,22321 |

Логично, что при формулировке выводов анализа исследователю важно понимать, какие именно фирмы относятся к той или иной кластерной группе. Для этого нужно рассмотреть информацию об элементах кластера и расстояния. В левом столбце таблицы указаны те фирмы, которые входят в конкретную группу.

Состав 1 кластера:

| | Элементы кластера номер 1 (Таблица данных1) и расстояния до центра кластера. Кластер содержит 11 набл. |
|---|---|
| | **объедин.** |
| **16** | 0,605268 |
| 24 | 0,443797 |
| 28 | 0,607433 |
| 50 | 0,639634 |
| 59 | 0,274056 |
| 60 | 0,392050 |
| 67 | 1,080086 |
| 68 | 0,356016 |
| 79 | 0,578811 |
| 80 | 0,453941 |
| 82 | 0,806846 |

Состав 2 кластера:

| | Элементы кластера номер 2 (Таблица данных1) и расстояния до центра кластера. Кластер содержит 21 набл. |
|---|---|
| | **объедин.** |
| **7** | 0,578300 |
| 18 | 0,881268 |
| 23 | 0,300810 |
| 27 | 0,440301 |
| 32 | 0,449717 |
| 33 | 0,421870 |
| 34 | 0,417834 |
| 35 | 0,445946 |
| 40 | 0,339058 |
| 41 | 1,186571 |
| 44 | 0,346150 |
| 45 | 0,524509 |
| 47 | 0,591295 |
| 55 | 0,404806 |
| 56 | 0,259853 |
| 57 | 0,439751 |
| 72 | 0,719171 |
| 74 | 0,460457 |
| 76 | 0,363827 |
| 78 | 0,325055 |
| 81 | 0,602127 |

Состав 3 кластера:

| | Элементы кластера номер 3 (Таблица данных1)<br>и расстояния до центра кластера.<br>Кластер содержит 16 набл. |
|---|---|
| | объедин. |
| 9 | 0,355440 |
| 12 | 0,422938 |
| 14 | 0,685169 |
| 21 | 0,337478 |
| 25 | 0,386128 |
| 48 | 0,411649 |
| 49 | 0,480433 |
| 52 | 0,604985 |
| 53 | 0,408404 |
| 58 | 0,347207 |
| 61 | 0,495333 |
| 65 | 0,434268 |
| 69 | 0,370896 |
| 73 | 0,831408 |
| 75 | 0,791058 |
| 77 | 0,721797 |

Состав 4 кластера:

| | Элементы кластера номер 4 (Таблица данных1)<br>и расстояния до центра кластера.<br>Кластер содержит 12 набл. |
|---|---|
| | объедин. |
| 1 | 0,618841 |
| 3 | 0,446415 |
| 4 | 0,523603 |
| 5 | 0,486806 |
| 6 | 0,169145 |
| 10 | 1,060731 |
| 17 | 0,361137 |
| 22 | 0,521240 |
| 26 | 0,786596 |
| 29 | 0,379885 |
| 43 | 0,596948 |
| 71 | 0,653395 |

Состав 6 кластера:

| | Элементы кластера номер 6 (Таблица данных1)<br>и расстояния до центра кластера.<br>Кластер содержит 1 набл. |
|---|---|
| | объедин. |
| 42 | 0,00 |

Состав 5 кластера:

| | Элементы кластера номер 5 (Таблица данных1) и расстояния до центра кластера. Кластер содержит 18 набл. |
|---|---|
| | объедин. |
| 2 | 0,619434 |
| 8 | 0,647949 |
| 11 | 0,350618 |
| 13 | 0,371313 |
| 15 | 0,522965 |
| 19 | 0,611208 |
| 20 | 0,450108 |
| 30 | 0,599378 |
| 31 | 0,665829 |
| 36 | 0,692314 |
| 46 | 0,651262 |
| 51 | 0,376626 |
| 54 | 0,458228 |
| 62 | 0,488208 |
| 63 | 1,174966 |
| 64 | 0,344646 |
| 66 | 0,377545 |
| 70 | 0,436121 |

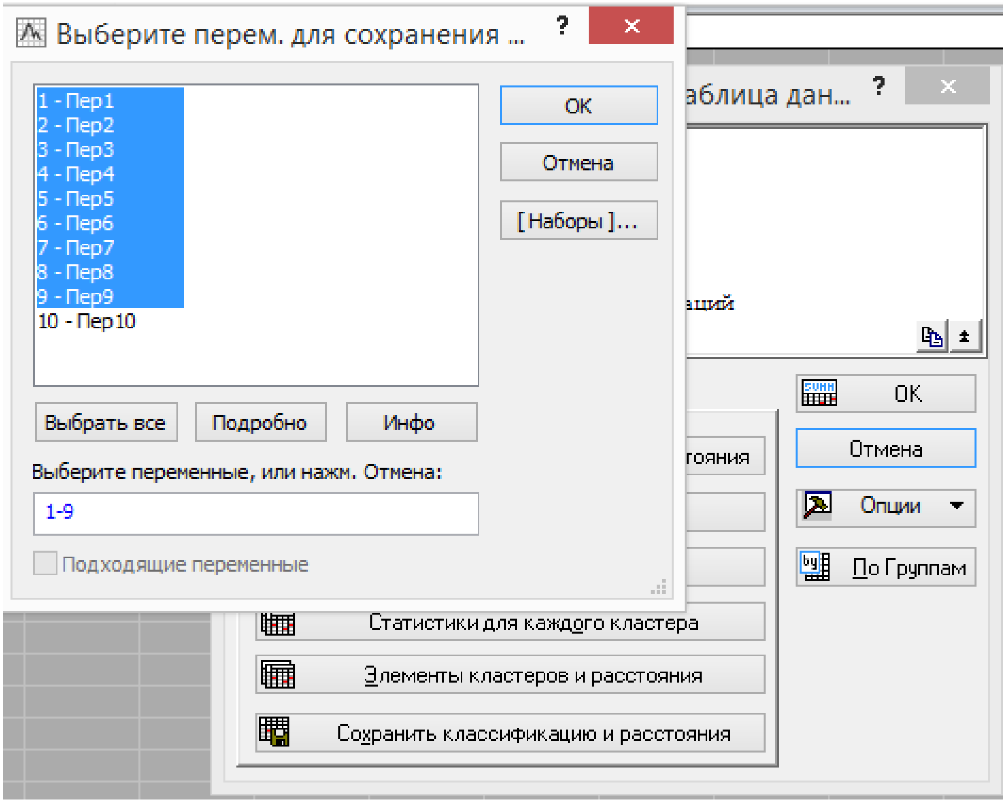
Состав 7 кластера:

| | Элементы кластера номер 7 (Таблица данных1) и расстояния до центра кластера. Кластер содержит 3 набл. |
|---|---|
| | объедин. |
| 37 | 1,395691 |
| 38 | 1,087465 |
| 39 | 0,719475 |

Представленные результаты кластеризации методом k-средних свидетельствуют о том, что разбивка фирм на кластерные группы неравномерна. Подобные результаты подтверждаются и результатами иерархической кластеризации. На дендрограмме видно, что фирма 42 находится обособленно от других элементов кластера. При расстоянии объединения 3 получается разбивка на кластеры, аналогичные группировке методом k-средних.

Наиболее многочисленной группой является кластер №2, который объединяет в себе элементы наблюдений по 21 фирме. При управлении развитием большого числа фирм важно понимать черты сходства и различия каждой из анализируемых групп.

Если необходимо рассмотреть общие результаты классификации и полученные расстояния, нужно нажать на соответствующую кнопку и выбрать переменные:




Тогда отобразится окно:

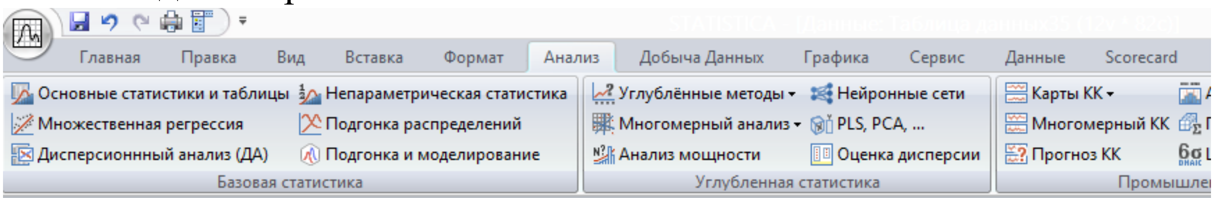
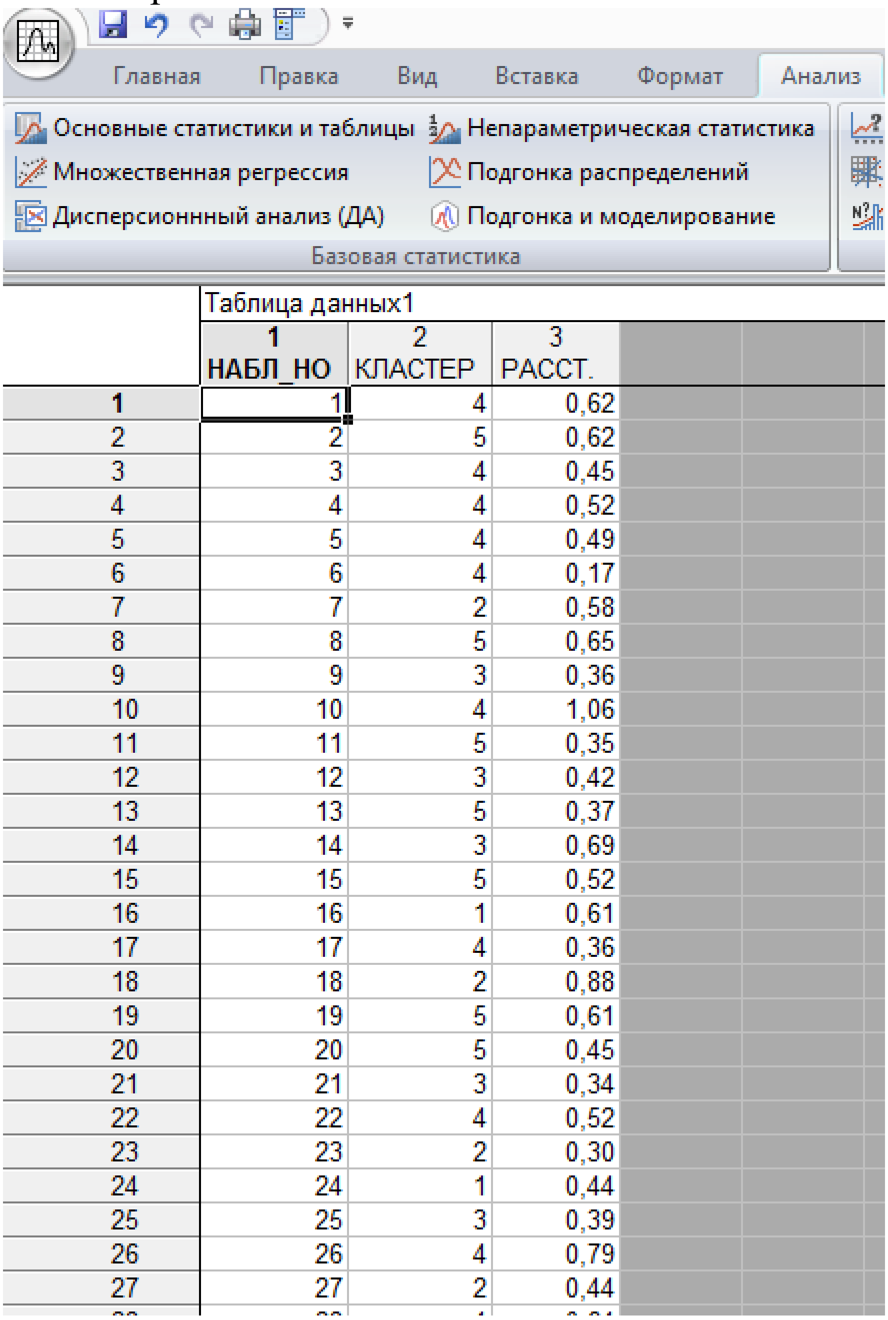



Таблица данных1

| | 1 Пер1 | 2 Пер2 | 3 Пер3 | 4 Пер4 | 5 Пер5 | 6 Пер6 | 7 Пер7 | 8 Пер8 | 9 Пер9 | 10 НАБЛ_НО | 11 КЛАСТЕР | 12 РАССТ. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1,8 | 0,1 | 0,8 | 0,92903 | 1,4 | 0,8 | 0,7 | 0,6 | 1,0 | 1 | 4 | 0,62 |
| 2 | 0,6 | -0,2 | 1,0 | 0,382997 | -1,1 | -1,4 | -1,6 | -1,5 | -1,0 | 2 | 5 | 0,62 |
| 3 | 1,1 | 1,1 | 1,3 | 0,700049 | 1,2 | 0,9 | 0,2 | 2,4 | 1,7 | 3 | 4 | 0,45 |
| 4 | 1,0 | 0,5 | 0,3 | 0,594365 | 0,6 | 0,2 | -0,2 | 1,9 | 0,2 | 4 | 4 | 0,52 |
| 5 | 1,0 | 0,6 | 0,6 | 0,594365 | 0,5 | 0,3 | -0,1 | 1,4 | 0,2 | 5 | 4 | 0,49 |
| 6 | 0,7 | 1,2 | 0,3 | 0,277313 | 1,0 | 0,6 | 0,4 | 1,6 | 1,2 | 6 | 4 | 0,17 |
| 7 | 0,7 | -0,5 | 0,2 | -0,25111 | -1,2 | 0,3 | -0,1 | 0,8 | -0,5 | 7 | 2 | 0,58 |
| 8 | 0,8 | -0,6 | 0,4 | -1,04374 | 0,2 | -0,8 | -0,9 | -1,3 | -0,5 | 8 | 5 | 0,65 |
| 9 | 1,3 | 0,2 | 0,3 | 0,118787 | 0,2 | -0,7 | -0,6 | -0,0 | 0,0 | 9 | 3 | 0,36 |
| 10 | -0,6 | 1,3 | 0,1 | -0,12781 | 1,3 | 2,5 | 1,0 | 2,8 | 2,9 | 10 | 4 | 1,06 |
| 11 | 0,6 | -0,4 | -0,3 | -0,18065 | -1,4 | -0,9 | -1,3 | -0,6 | -0,9 | 11 | 5 | 0,35 |
| 12 | 0,2 | 0,0 | -0,1 | 0,259699 | 0,0 | -0,7 | -0,8 | 0,3 | -0,1 | 12 | 3 | 0,42 |
| 13 | 0,1 | 0,1 | -0,1 | 0,347769 | -0,4 | -0,6 | -0,9 | -0,5 | -0,8 | 13 | 5 | 0,37 |
| 14 | 1,0 | -1,0 | 1,4 | 0,752891 | 1,2 | -0,1 | -0,4 | 0,4 | 0,5 | 14 | 3 | 0,69 |
| 15 | 0,3 | -0,1 | -0,6 | -0,23349 | 0,2 | -0,3 | -0,7 | -0,8 | -0,6 | 15 | 5 | 0,52 |
| 16 | -0,4 | 0,7 | 0,7 | -0,18065 | -0,3 | 1,0 | 0,7 | 0,5 | 2,0 | 16 | 1 | 0,61 |
| 17 | 0,3 | 1,0 | 1,0 | 0,559137 | 0,5 | 1,0 | 0,4 | 1,1 | 0,6 | 17 | 4 | 0,36 |
| 18 | -1,6 | 0,6 | -1,6 | 1,19324 | 0,6 | -0,1 | 0,0 | -0,5 | -0,5 | 18 | 2 | 0,88 |
| 19 | 0,4 | -0,5 | 0,2 | 0,277313 | -0,8 | -2,6 | -0,6 | -1,2 | -0,3 | 19 | 5 | 0,61 |
| 20 | -0,3 | -1,0 | 0,4 | 0,013103 | -1,5 | -1,3 | -1,0 | -1,5 | -0,5 | 20 | 5 | 0,45 |
| 21 | 0,4 | -0,2 | 0,8 | 0,118787 | 0,4 | -0,9 | -0,2 | -0,7 | -0,5 | 21 | 3 | 0,34 |

Если рассмотрение переменных исследователя не интересует, нажав кнопку «Отмена» появляются результаты, не отражающие данные сведения, а только состав групп и принадлежность отдельных фирм к кластерам:




| | 1 НАБЛ_НО | 2 КЛАСТЕР | 3 РАССТ. |
|---|---|---|---|
| 1 | 1 | 4 | 0,62 |
| 2 | 2 | 5 | 0,62 |
| 3 | 3 | 4 | 0,45 |
| 4 | 4 | 4 | 0,52 |
| 5 | 5 | 4 | 0,49 |
| 6 | 6 | 4 | 0,17 |
| 7 | 7 | 2 | 0,58 |
| 8 | 8 | 5 | 0,65 |
| 9 | 9 | 3 | 0,36 |
| 10 | 10 | 4 | 1,06 |
| 11 | 11 | 5 | 0,35 |
| 12 | 12 | 3 | 0,42 |
| 13 | 13 | 5 | 0,37 |
| 14 | 14 | 3 | 0,69 |
| 15 | 15 | 5 | 0,52 |
| 16 | 16 | 1 | 0,61 |
| 17 | 17 | 4 | 0,36 |
| 18 | 18 | 2 | 0,88 |
| 19 | 19 | 5 | 0,61 |
| 20 | 20 | 5 | 0,45 |
| 21 | 21 | 3 | 0,34 |
| 22 | 22 | 4 | 0,52 |
| 23 | 23 | 2 | 0,30 |
| 24 | 24 | 1 | 0,44 |
| 25 | 25 | 3 | 0,39 |
| 26 | 26 | 4 | 0,79 |
| 27 | 27 | 2 | 0,44 |

Использование возможностей программного продукта Statistica позволяет быстро и наглядно рассмотреть систему взаимосвязей элементов, выявить кластерные группы по результатам расчетов либо графического анализа, сформулировать вывод о значимости результатов исследования.

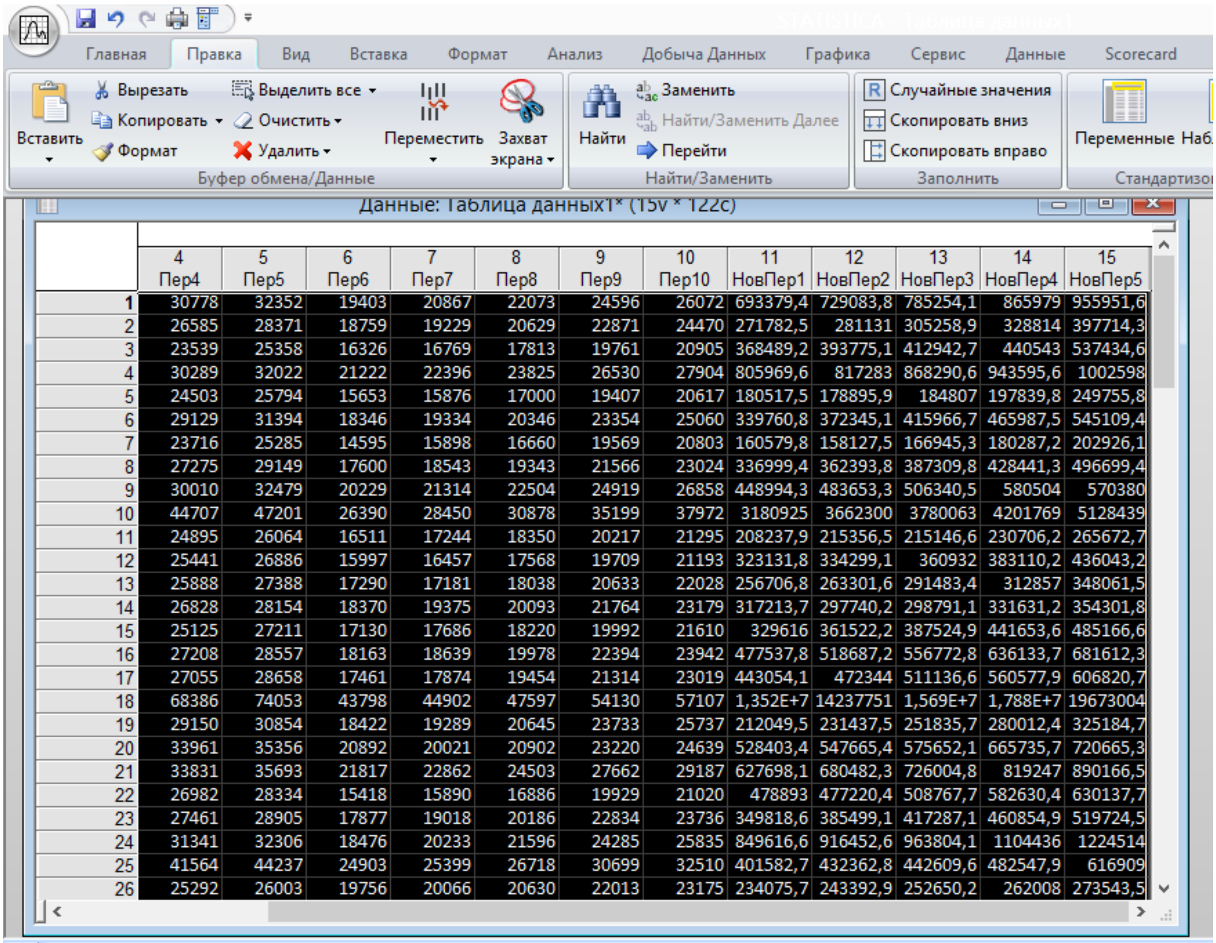
**Пример 7.2**

Имеются данные о показателях, характеризующих некоторые показатели производственной деятельности, 27 фирмами (приложение 8-10).

На основании представленных данных определите кластеры, используя квадрат Евклидова расстояния в качестве меры близости объектов. Сравните результаты иерархической кластеризации и группировки методом k-средних

**Решение.**

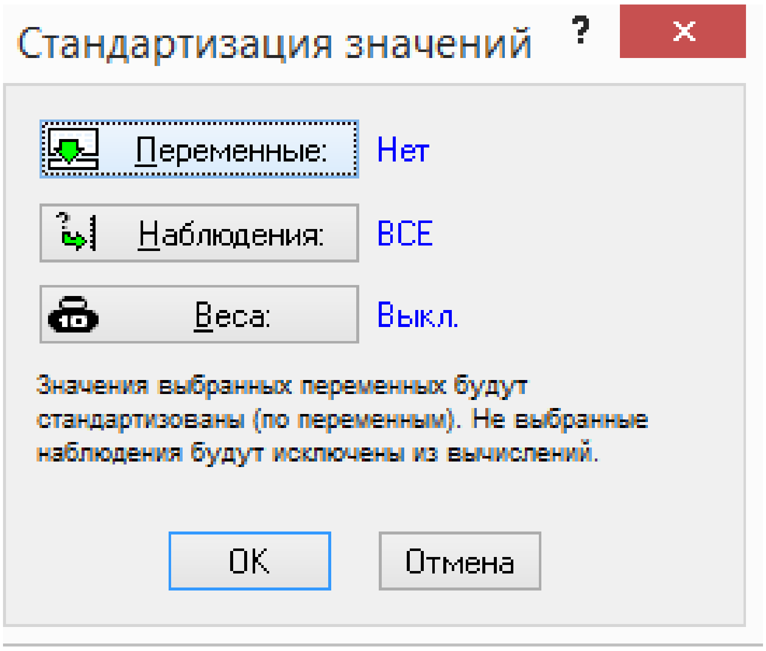
На первом этапе введем исходный массив данных в программу Statistica :



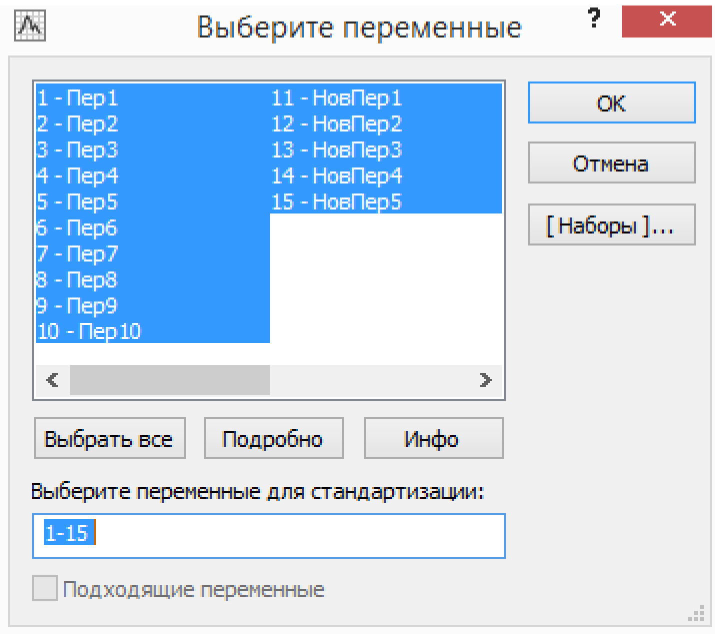
Данные: Таблица данных1* (15v * 122c)

| | 4 Пер4 | 5 Пер5 | 6 Пер6 | 7 Пер7 | 8 Пер8 | 9 Пер9 | 10 Пер10 | 11 НовПер1 | 12 НовПер2 | 13 НовПер3 | 14 НовПер4 | 15 НовПер5 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 30778 | 32352 | 19403 | 20867 | 22073 | 24596 | 26072 | 693379,4 | 729083,8 | 785254,1 | 865979 | 955951,6 |
| 2 | 26585 | 28371 | 18759 | 19229 | 20629 | 22871 | 24470 | 271782,5 | 281131 | 305258,9 | 328814 | 397714,3 |
| 3 | 23539 | 25358 | 16326 | 16769 | 17813 | 19761 | 20905 | 368489,2 | 393775,1 | 412942,7 | 440543 | 537434,6 |
| 4 | 30289 | 32022 | 21222 | 22396 | 23825 | 26530 | 27904 | 805969,6 | 817283 | 868290,6 | 943595,6 | 1002598 |
| 5 | 24503 | 25794 | 15653 | 15876 | 17000 | 19407 | 20617 | 180517,5 | 178895,9 | 184807 | 197839,8 | 249755,8 |
| 6 | 29129 | 31394 | 18346 | 19334 | 20346 | 23354 | 25060 | 339760,8 | 372345,1 | 415966,7 | 465987,5 | 545109,4 |
| 7 | 23716 | 25285 | 14595 | 15898 | 16660 | 19569 | 20803 | 160579,8 | 158127,5 | 166945,3 | 180287,2 | 202926,1 |
| 8 | 27275 | 29149 | 17600 | 18543 | 19343 | 21566 | 23024 | 336999,4 | 362393,8 | 387309,8 | 428441,3 | 496699,4 |
| 9 | 30010 | 32479 | 20229 | 21314 | 22504 | 24919 | 26858 | 448994,3 | 483653,3 | 506340,5 | 580504 | 570380 |
| 10 | 44707 | 47201 | 26390 | 28450 | 30878 | 35199 | 37972 | 3180925 | 3662300 | 3780063 | 4201769 | 5128439 |
| 11 | 24895 | 26064 | 16511 | 17244 | 18350 | 20217 | 21295 | 208237,9 | 215356,5 | 215146,6 | 230706,2 | 265672,7 |
| 12 | 25441 | 26886 | 15997 | 16457 | 17568 | 19709 | 21193 | 323131,8 | 334299,1 | 360932 | 383110,2 | 436043,2 |
| 13 | 25888 | 27388 | 17290 | 17181 | 18038 | 20633 | 22028 | 256706,8 | 263301,6 | 291483,4 | 312857 | 348061,5 |
| 14 | 26828 | 28154 | 18370 | 19375 | 20093 | 21764 | 23179 | 317213,7 | 297740,2 | 298791,1 | 331631,2 | 354301,8 |
| 15 | 25125 | 27211 | 17130 | 17686 | 18220 | 19992 | 21610 | 329616 | 361522,2 | 387524,9 | 441653,6 | 485166,6 |
| 16 | 27208 | 28557 | 18163 | 18639 | 19978 | 22394 | 23942 | 477537,8 | 518687,2 | 556772,8 | 636133,7 | 681612,3 |
| 17 | 27055 | 28658 | 17461 | 17874 | 19454 | 21314 | 23019 | 443054,1 | 472344 | 511136,6 | 560577,9 | 606820,7 |
| 18 | 68386 | 74053 | 43798 | 44902 | 47597 | 54130 | 57107 | 1,352E+7 | 14237751 | 1,569E+7 | 1,788E+7 | 19673004 |
| 19 | 29150 | 30854 | 18422 | 19289 | 20645 | 23733 | 25737 | 212049,5 | 231437,5 | 251835,7 | 280012,4 | 325184,7 |
| 20 | 33961 | 35356 | 20892 | 20021 | 20902 | 23220 | 24639 | 528403,4 | 547665,4 | 575652,1 | 665735,7 | 720665,3 |
| 21 | 33831 | 35693 | 21817 | 22862 | 24503 | 27662 | 29187 | 627698,1 | 680482,3 | 726004,8 | 819247 | 890166,5 |
| 22 | 26982 | 28334 | 15418 | 15890 | 16886 | 19929 | 21020 | 478893 | 477220,4 | 508767,7 | 582630,4 | 630137,7 |
| 23 | 27461 | 28905 | 17877 | 19018 | 20186 | 22834 | 23736 | 349818,6 | 385499,1 | 417287,1 | 460854,9 | 519724,5 |
| 24 | 31341 | 32306 | 18476 | 20233 | 21596 | 24285 | 25835 | 849616,6 | 916452,6 | 963804,1 | 1104436 | 1224514 |
| 25 | 41564 | 44237 | 24903 | 25399 | 26718 | 30699 | 32510 | 401582,7 | 432362,8 | 442609,6 | 482547,9 | 616909 |
| 26 | 25292 | 26003 | 19756 | 20066 | 20630 | 22013 | 23175 | 234075,7 | 243392,9 | 252650,2 | 262008 | 273543,5 |

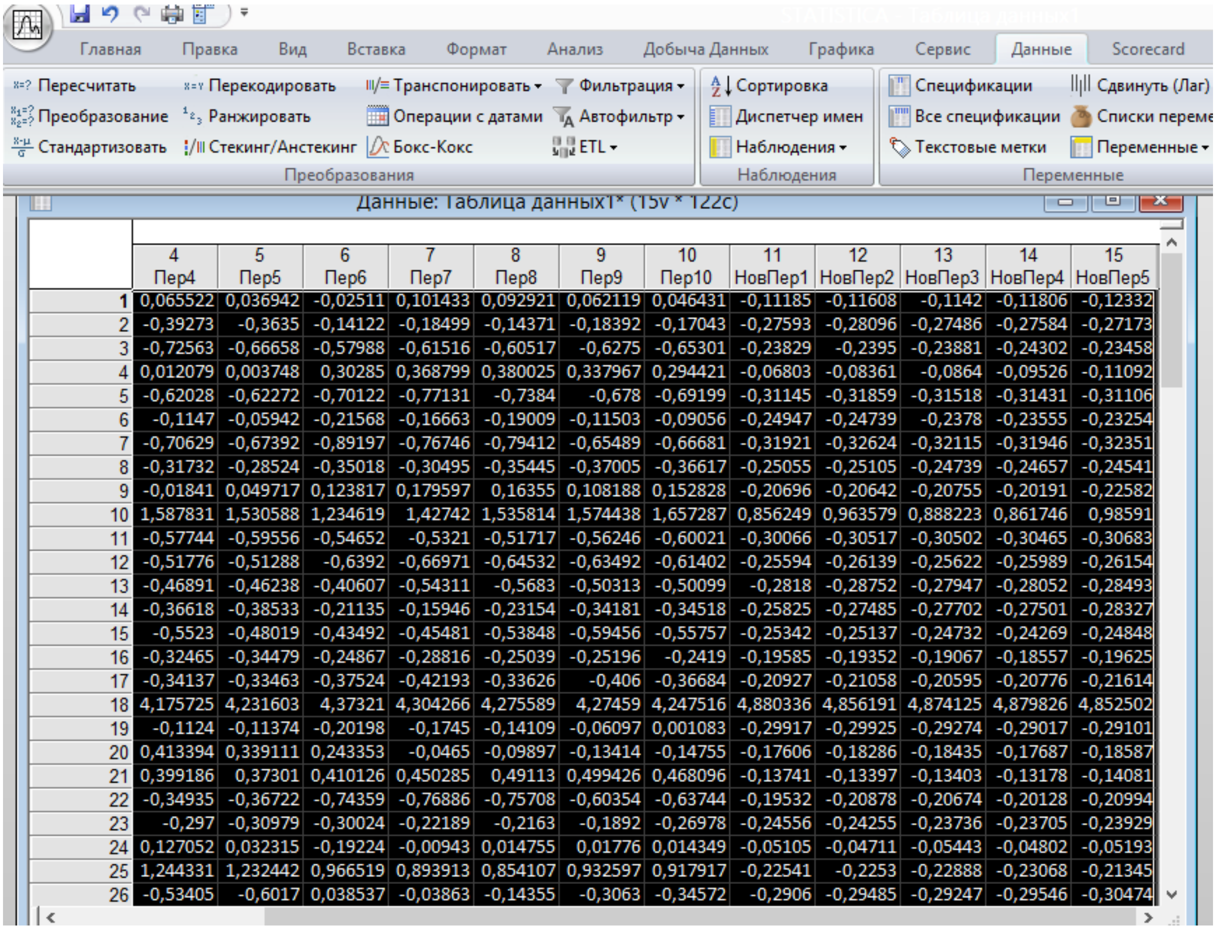
Поскольку предполагается, что используемые характеристики признака носят разнородный характер, их необходимо стандартизировать, нажав кнопку «Стандартизация» во вкладке «Данные»:




Поскольку в данном примере 15 переменных, выделяем нужный массив:

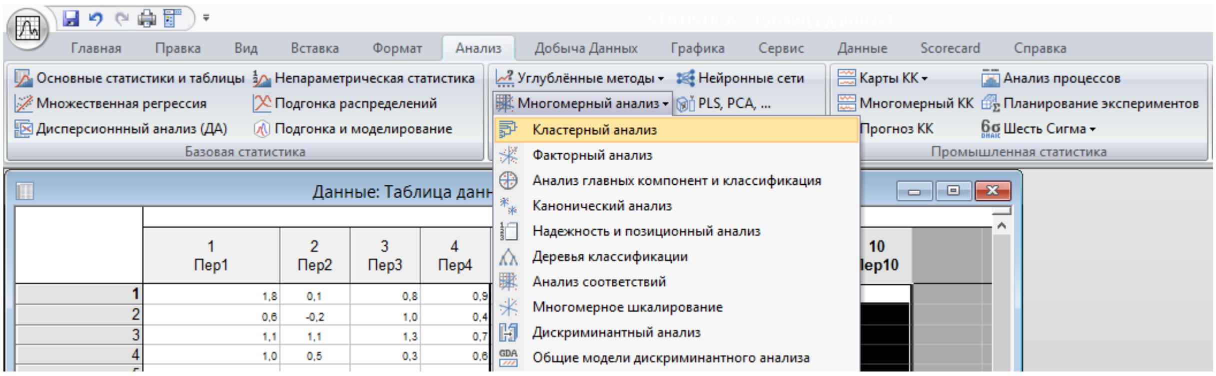



После проведения процедуры стандартизации массив будет выглядеть следующим образом:

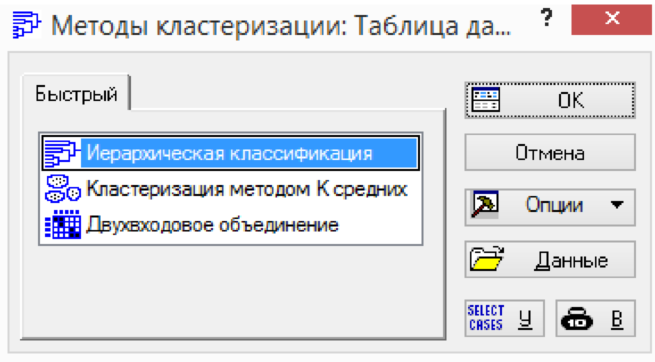



| | 4 Пер4 | 5 Пер5 | 6 Пер6 | 7 Пер7 | 8 Пер8 | 9 Пер9 | 10 Пер10 | 11 НовПер1 | 12 НовПер2 | 13 НовПер3 | 14 НовПер4 | 15 НовПер5 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0,065522 | 0,036942 | -0,02511 | 0,101433 | 0,092921 | 0,062119 | 0,046431 | -0,11185 | -0,11608 | -0,1142 | -0,11806 | -0,12332 |
| 2 | -0,39273 | -0,3635 | -0,14122 | -0,18499 | -0,14371 | -0,18392 | -0,17043 | -0,27593 | -0,28096 | -0,27486 | -0,27584 | -0,27173 |
| 3 | -0,72563 | -0,66658 | -0,57988 | -0,61516 | -0,60517 | -0,6275 | -0,65301 | -0,23829 | -0,2395 | -0,23881 | -0,24302 | -0,23458 |
| 4 | 0,012079 | 0,003748 | 0,30285 | 0,368799 | 0,380025 | 0,337967 | 0,294421 | -0,06803 | -0,08361 | -0,0864 | -0,09526 | -0,11092 |
| 5 | -0,62028 | -0,62272 | -0,70122 | -0,77131 | -0,7384 | -0,678 | -0,69199 | -0,31145 | -0,31859 | -0,31518 | -0,31431 | -0,31106 |
| 6 | -0,1147 | -0,05942 | -0,21568 | -0,16663 | -0,19009 | -0,11503 | -0,09056 | -0,24947 | -0,24739 | -0,2378 | -0,23555 | -0,23254 |
| 7 | -0,70629 | -0,67392 | -0,89197 | -0,76746 | -0,79412 | -0,65489 | -0,66681 | -0,31921 | -0,32624 | -0,32115 | -0,31946 | -0,32351 |
| 8 | -0,31732 | -0,28524 | -0,35018 | -0,30495 | -0,35445 | -0,37005 | -0,36617 | -0,25055 | -0,25105 | -0,24739 | -0,24657 | -0,24541 |
| 9 | -0,01841 | 0,049717 | 0,123817 | 0,179597 | 0,16355 | 0,108188 | 0,152828 | -0,20696 | -0,20642 | -0,20755 | -0,20191 | -0,22582 |
| 10 | 1,587831 | 1,530588 | 1,234619 | 1,42742 | 1,535814 | 1,574438 | 1,657287 | 0,856249 | 0,963579 | 0,888223 | 0,861746 | 0,98591 |
| 11 | -0,57744 | -0,59556 | -0,54652 | -0,5321 | -0,51717 | -0,56246 | -0,60021 | -0,30066 | -0,30517 | -0,30502 | -0,30465 | -0,30683 |
| 12 | -0,51776 | -0,51288 | -0,6392 | -0,66971 | -0,64532 | -0,63492 | -0,61402 | -0,25594 | -0,26139 | -0,25622 | -0,25989 | -0,26154 |
| 13 | -0,46891 | -0,46238 | -0,40607 | -0,54311 | -0,5683 | -0,50313 | -0,50099 | -0,2818 | -0,28752 | -0,27947 | -0,28052 | -0,28493 |
| 14 | -0,36618 | -0,38533 | -0,21135 | -0,15946 | -0,23154 | -0,34181 | -0,34518 | -0,25825 | -0,27485 | -0,27702 | -0,27501 | -0,28327 |
| 15 | -0,5523 | -0,48019 | -0,43492 | -0,45481 | -0,53848 | -0,59456 | -0,55757 | -0,25342 | -0,25137 | -0,24732 | -0,24269 | -0,24848 |
| 16 | -0,32465 | -0,34479 | -0,24867 | -0,28816 | -0,25039 | -0,25196 | -0,2419 | -0,19585 | -0,19352 | -0,19067 | -0,18557 | -0,19625 |
| 17 | -0,34137 | -0,33463 | -0,37524 | -0,42193 | -0,33626 | -0,406 | -0,36684 | -0,20927 | -0,21058 | -0,20595 | -0,20776 | -0,21614 |
| 18 | 4,175725 | 4,231603 | 4,37321 | 4,304266 | 4,275589 | 4,27459 | 4,247516 | 4,880336 | 4,856191 | 4,874125 | 4,879826 | 4,852502 |
| 19 | -0,1124 | -0,11374 | -0,20198 | -0,1745 | -0,14109 | -0,06097 | 0,001083 | -0,29917 | -0,29925 | -0,29274 | -0,29017 | -0,29101 |
| 20 | 0,413394 | 0,339111 | 0,243353 | -0,0465 | -0,09897 | -0,13414 | -0,14755 | -0,17606 | -0,18286 | -0,18435 | -0,17687 | -0,18587 |
| 21 | 0,399186 | 0,37301 | 0,410126 | 0,450285 | 0,49113 | 0,499426 | 0,468096 | -0,13741 | -0,13397 | -0,13403 | -0,13178 | -0,14081 |
| 22 | -0,34935 | -0,36722 | -0,74359 | -0,76886 | -0,75708 | -0,60354 | -0,63744 | -0,19532 | -0,20878 | -0,20674 | -0,20128 | -0,20994 |
| 23 | -0,297 | -0,30979 | -0,30024 | -0,22189 | -0,2163 | -0,1892 | -0,26978 | -0,24556 | -0,24255 | -0,23736 | -0,23705 | -0,23929 |
| 24 | 0,127052 | 0,032315 | -0,19224 | -0,00943 | 0,014755 | 0,01776 | 0,014349 | -0,05105 | -0,04711 | -0,05443 | -0,04802 | -0,05193 |
| 25 | 1,244331 | 1,232442 | 0,966519 | 0,893913 | 0,854107 | 0,932597 | 0,917917 | -0,22541 | -0,2253 | -0,22888 | -0,23068 | -0,21345 |
| 26 | -0,53405 | -0,6017 | 0,038537 | -0,03863 | -0,14355 | -0,3063 | -0,34572 | -0,2906 | -0,29485 | -0,29247 | -0,29546 | -0,30474 |

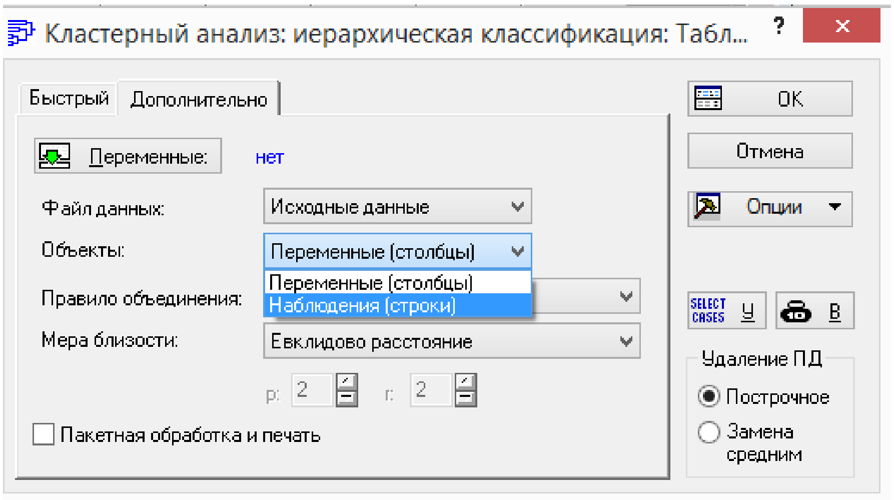
Непосредственно для начала осуществления кластерного анализа необходимо выбрать команду Анализ/Многомерный анализ/Кластерный анализ:



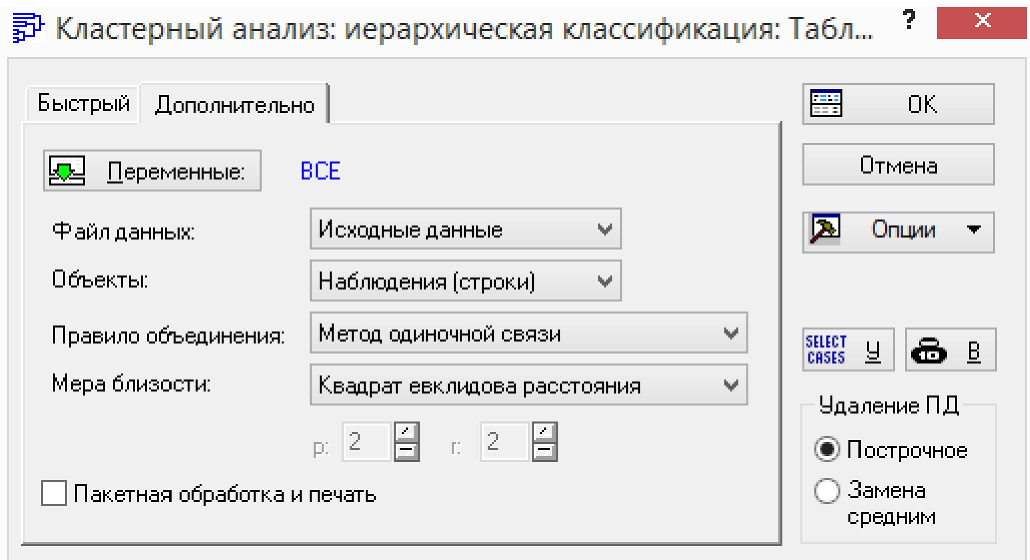

Для выбора иерархической классификации выбирается соответственный вариант в диалоговом окне:



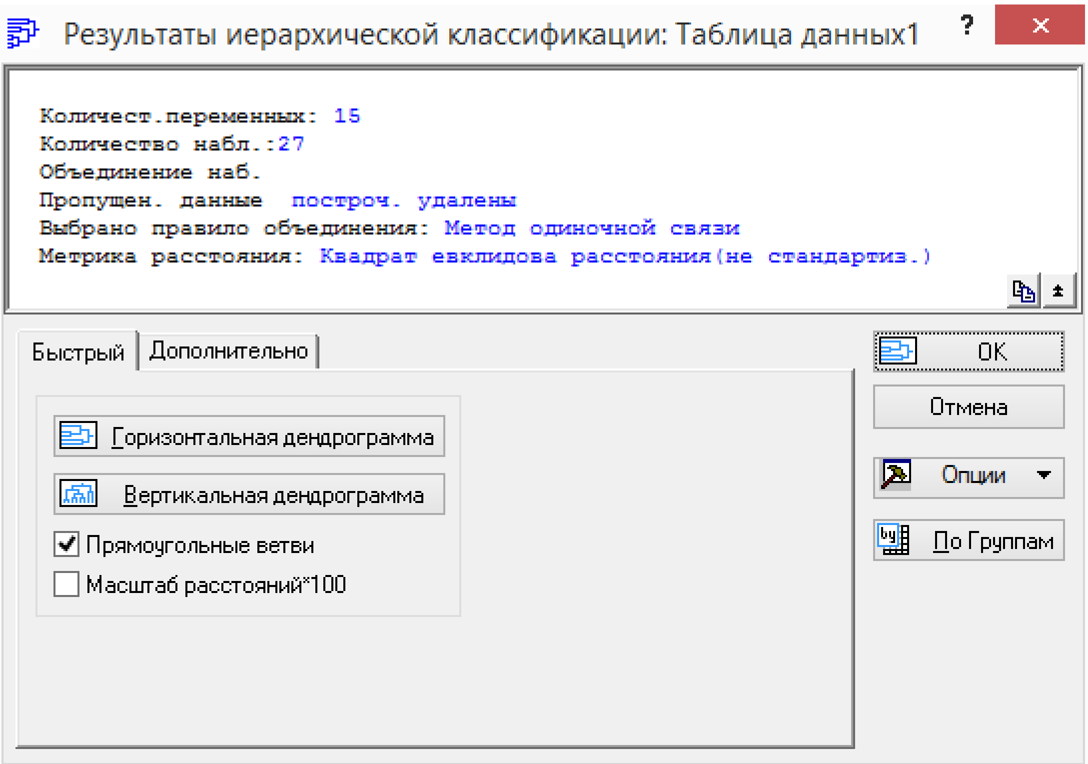

Поскольку в данном примере объектом классификации являются фирмы, а не значения независимых переменных, во вкладке «Дополнительно» был изменен объект анализа:



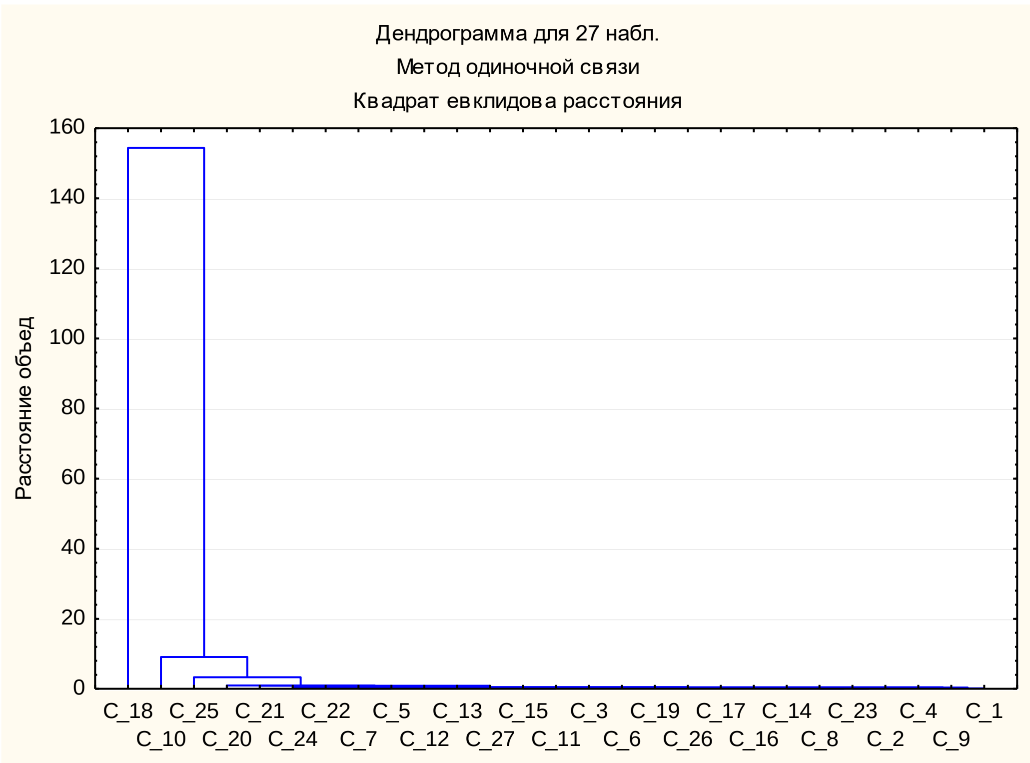

В соответствии с заданными условиями кластеризации, была изменена применяемая мера близости:



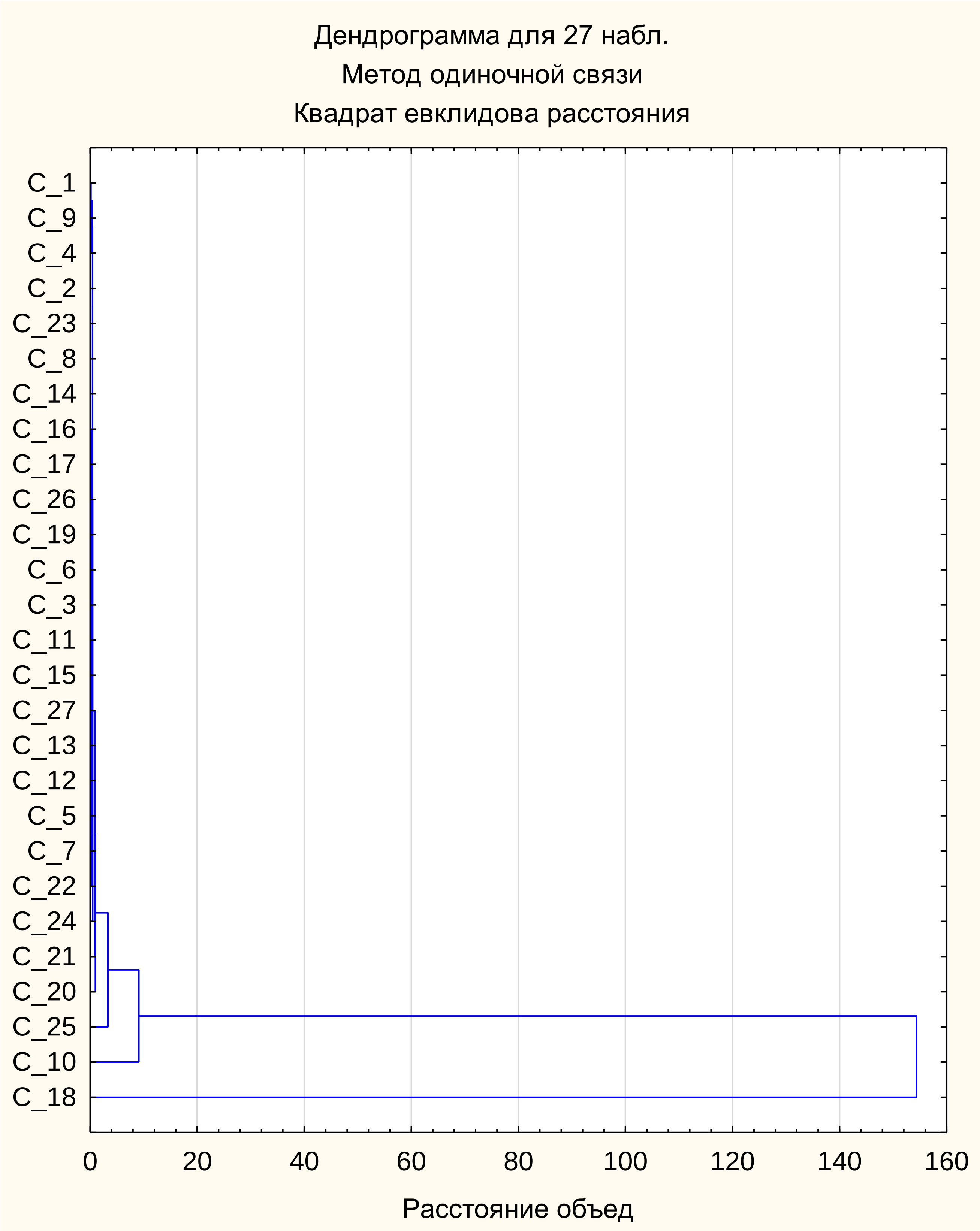
Результаты иерархической классификации отображаются в следующем окне:




Вертикальная дендрограмма, построенная по результатам кластерного анализа, будет иметь вид:



Горизонтальная дендрограмма:



Оптимальное число групп (интервалов) может быть определено по формуле Стерджесса:

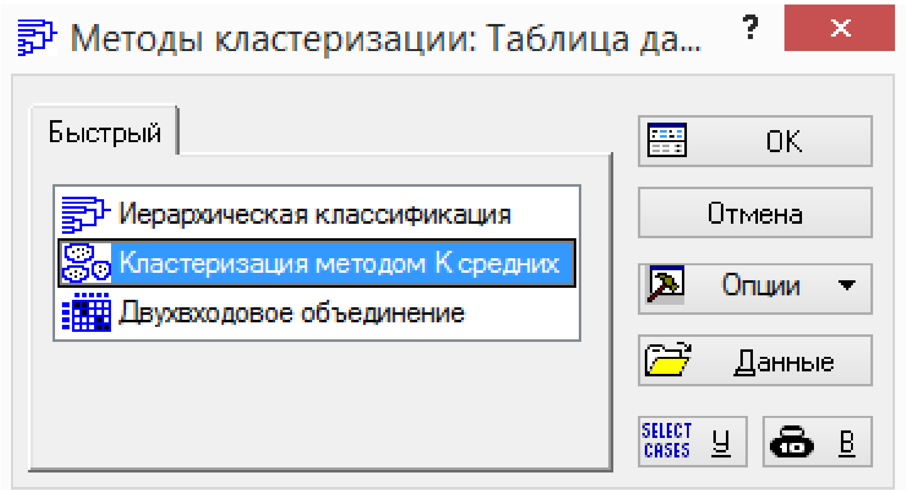
$$K = 1+3.322 \lg n,$$

где n - число единиц совокупности

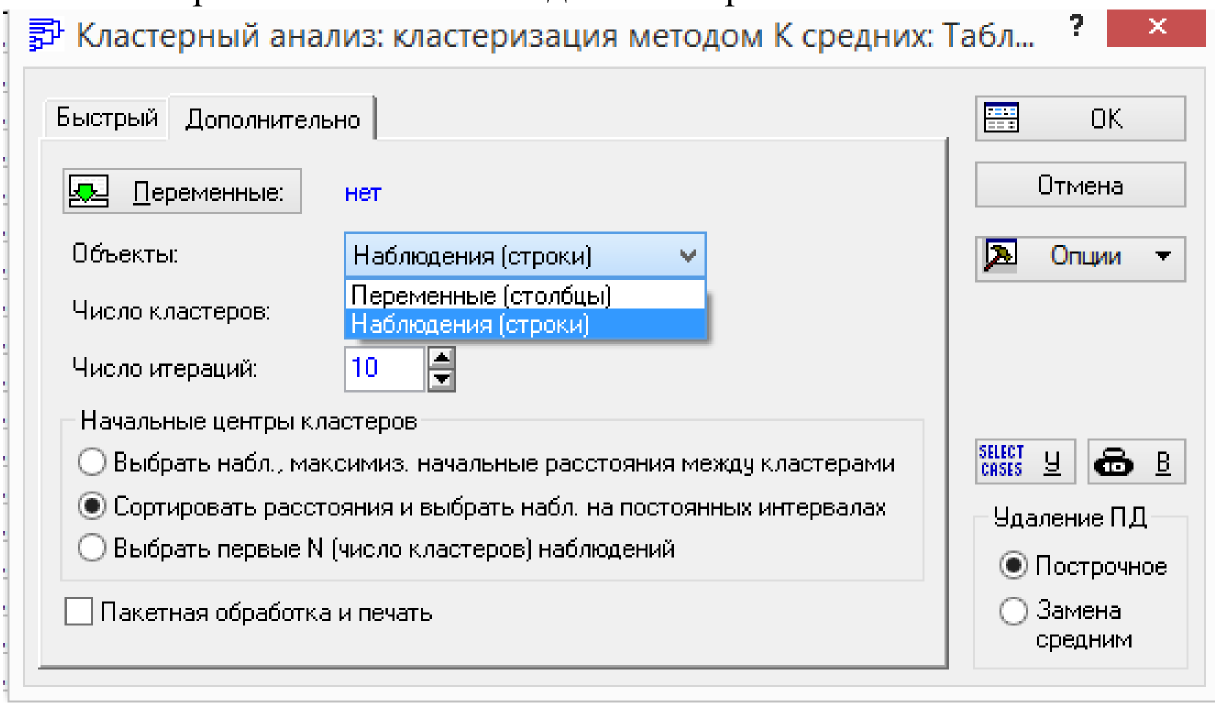
Для 27 фирм:

$$K = 1+3.322 \lg 27 \approx 6$$

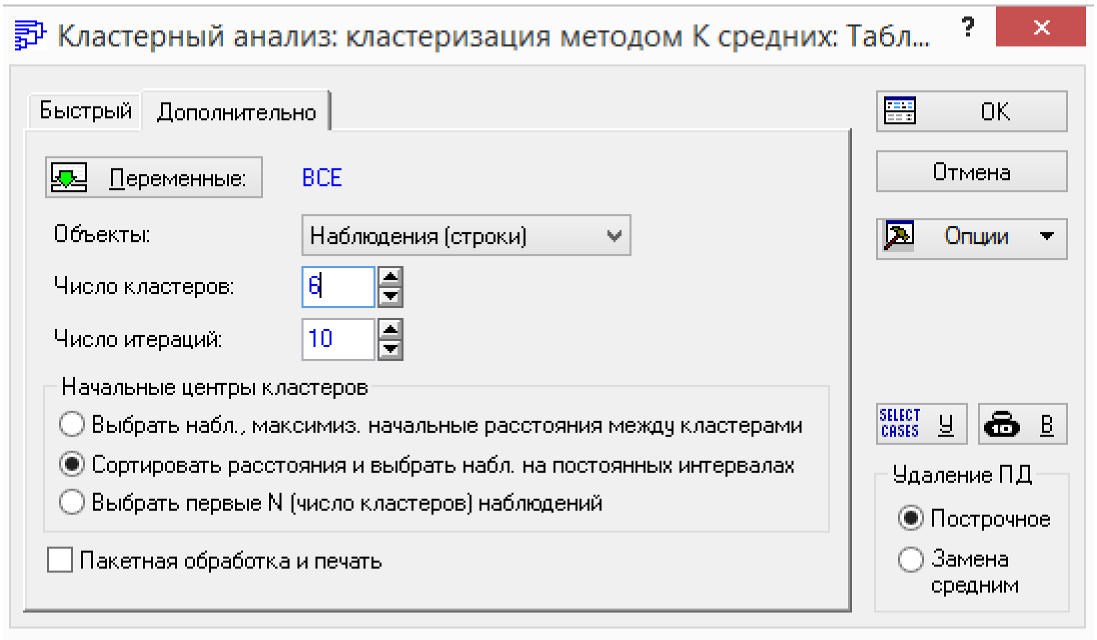
Проверим гипотезу о том, что 27 фирм целесообразно разбить на 6 кластеров методом k-средних.



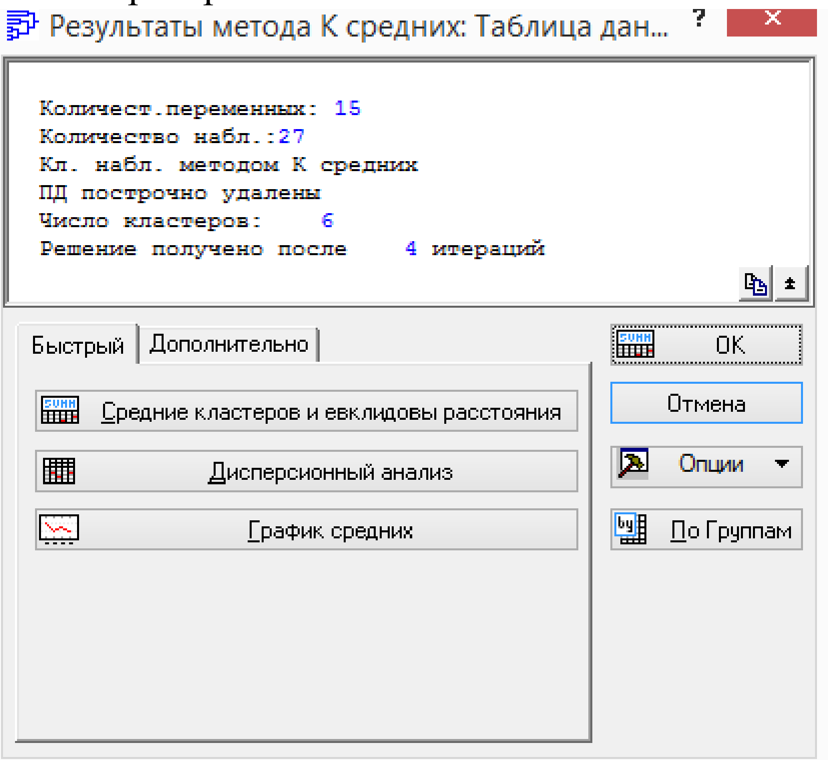

Выбирается объект наблюдений – строки:




Число кластеров определяется равным 7, итерации остаются без изменений:




Результаты кластеризации будут отображены на появившемся диалоговом пространстве:




Средние кластеров по итогам разбиения фирм на 6 кластеров:

| перемен. | Средн.класт. (Таблица данных1) Кластер Но. 1 | Кластер Но. 2 | Кластер Но. 3 | Кластер Но. 4 | Кластер Но. 5 | Кластер Но. 6 |
|---|---|---|---|---|---|---|
| **Пер1** | 1,265170 | 4,209947 | 0,311462 | -0,276488 | 1,166421 | -0,603774 |
| Пер2 | 1,627487 | 4,155233 | 0,250653 | -0,264010 | 1,033343 | -0,603248 |
| Пер3 | 1,560804 | 4,272174 | 0,198024 | -0,274736 | 1,027617 | -0,567039 |
| Пер4 | 1,587831 | 4,175725 | 0,174353 | -0,267334 | 1,244331 | -0,578480 |
| Пер5 | 1,530588 | 4,231603 | 0,160506 | -0,276583 | 1,232442 | -0,559036 |
| Пер6 | 1,234619 | 4,373209 | 0,211008 | -0,219827 | 0,966519 | -0,603457 |
| Пер7 | 1,427420 | 4,304265 | 0,210723 | -0,197058 | 0,893913 | -0,634293 |
| Пер8 | 1,535814 | 4,275589 | 0,205731 | -0,199262 | 0,854107 | -0,633505 |
| Пер9 | 1,574438 | 4,274590 | 0,174712 | -0,220748 | 0,932597 | -0,605301 |
| Пер10 | 1,657287 | 4,247516 | 0,162845 | -0,218115 | 0,917917 | -0,606199 |
| НовПер1 | 0,856249 | 4,880336 | -0,140063 | -0,232570 | -0,225412 | -0,276128 |
| НовПер2 | 0,963579 | 4,856192 | -0,144587 | -0,234211 | -0,225296 | -0,281048 |
| НовПер3 | 0,888223 | 4,874125 | -0,145306 | -0,231070 | -0,228884 | -0,277360 |
| НовПер4 | 0,861746 | 4,879826 | -0,144777 | -0,229700 | -0,230681 | -0,276668 |
| НовПер5 | 0,985911 | 4,852501 | -0,157350 | -0,233230 | -0,213455 | -0,278434 |

Полученные Евклидовы расстояния для межкластерного деления:

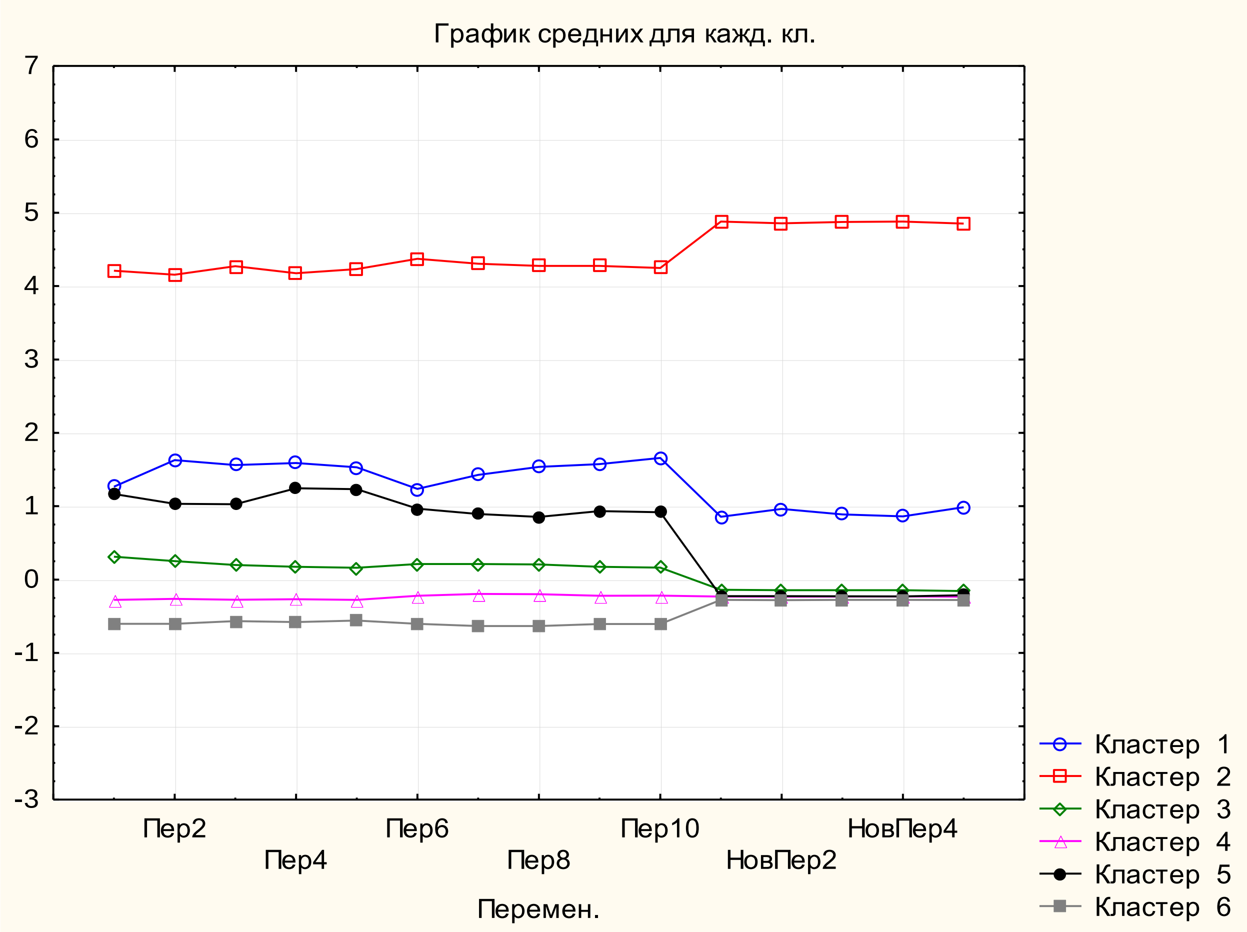
| Кластер Номер | Евклидовы расст. между кластерами (Таблица данных1) Расстояния под диагональю Квадраты расстояний над диагональю Но. 1 | Но. 2 | Но. 3 | Но. 4 | Но. 5 | Но. 6 |
|---|---|---|---|---|---|---|
| **Но. 1** | 0,000000 | 10,29117 | 1,50909 | 2,47243 | 0,60647 | 3,42337 |
| Но. 2 | 3,207985 | 0,00000 | 19,30119 | 22,13452 | 15,60228 | 24,52316 |
| Но. 3 | 1,228448 | 4,39331 | 0,00000 | 0,13830 | 0,46396 | 0,44032 |
| Но. 4 | 1,572396 | 4,70473 | 0,37189 | 0,00000 | 1,09014 | 0,08801 |
| Но. 5 | 0,778760 | 3,94997 | 0,68115 | 1,04410 | 0,00000 | 1,77351 |
| Но. 6 | 1,850235 | 4,95209 | 0,66357 | 0,29666 | 1,33173 | 0,00000 |

Результаты дисперсионного анализа:

| перемен. | Дисперсионный анализ (Таблица данных1) Между SS | сс | Внутри SS | сс | F | значим. p |
|---|---|---|---|---|---|---|
| **Пер1** | 25,21524 | 5 | 0,784761 | 21 | 134,951 | 0,000000 |
| Пер2 | 25,26880 | 5 | 0,731201 | 21 | 145,143 | 0,000000 |
| Пер3 | 25,58825 | 5 | 0,411751 | 21 | 261,009 | 0,000000 |
| Пер4 | 25,38466 | 5 | 0,615337 | 21 | 173,264 | 0,000000 |
| Пер5 | 25,47457 | 5 | 0,525435 | 21 | 203,628 | 0,000000 |
| Пер6 | 25,56671 | 5 | 0,433287 | 21 | 247,827 | 0,000000 |
| Пер7 | 25,59459 | 5 | 0,405405 | 21 | 265,160 | 0,000000 |
| Пер8 | 25,58952 | 5 | 0,410484 | 21 | 261,827 | 0,000000 |
| Пер9 | 25,55813 | 5 | 0,441870 | 21 | 242,932 | 0,000000 |
| Пер10 | 25,54620 | 5 | 0,453804 | 21 | 236,433 | 0,000000 |
| НовПер1 | 25,92685 | 5 | 0,073147 | 21 | 1488,688 | 0,000000 |
| НовПер2 | 25,92581 | 5 | 0,074194 | 21 | 1467,618 | 0,000000 |
| НовПер3 | 25,93028 | 5 | 0,069721 | 21 | 1562,038 | 0,000000 |
| НовПер4 | 25,92984 | 5 | 0,070158 | 21 | 1552,287 | 0,000000 |
| НовПер5 | 25,92984 | 5 | 0,070160 | 21 | 1552,244 | 0,000000 |

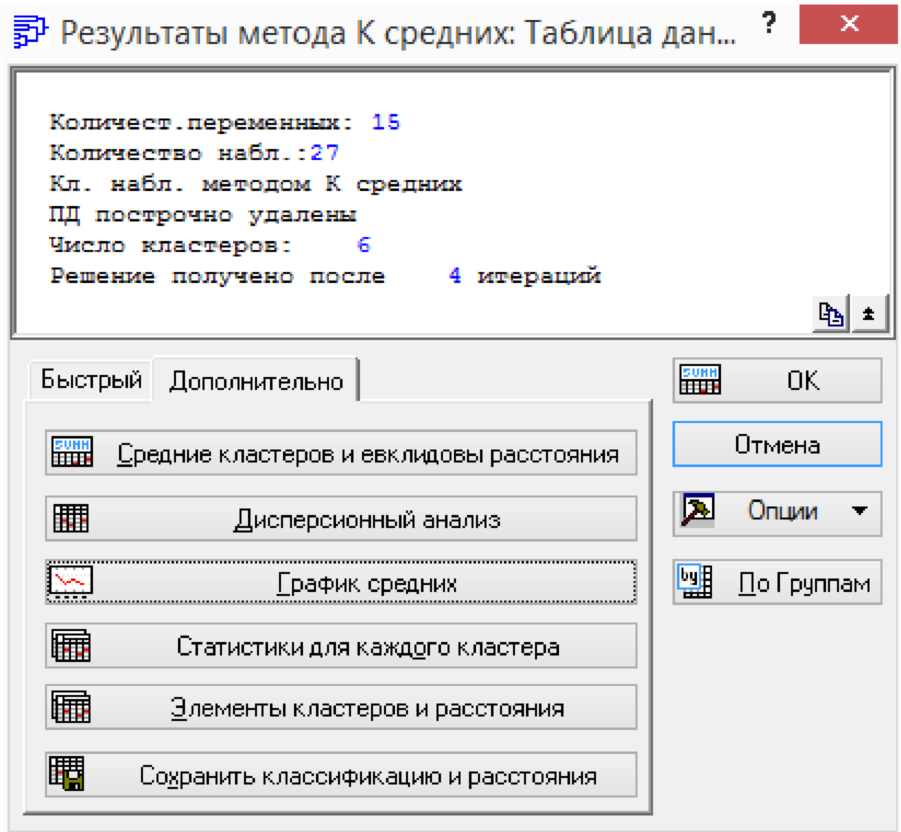
Поскольку р-значение меньше 5%, деление совокупности фирм на 6 кластеров можно считать значимым.

Подробнее график средних характеристик каждого их кластеров:




По графику видно, что средние значения для кластера 2 являются максимальными по всем рассматриваемым переменным. Наименьшие средние характерны для кластера 6. Разница средних по переменным 10-15 в кластерах 3, 4, 5 и 6 является наименьшей.

Дополнительные характеристики кластеризации методом k-средних включают статистики для каждого кластера, элементы кластеров и расстояния, возможность сохранения классификации и расстояния:




Статистики для каждого кластера содержат информацию о средних по каждой из переменных, входящих в состав группы, величину дисперсии и стандартного отклонения:

| | Описат.статистики для кластера 1 (Таблица данных1) Кластер содержит 1 набл. | | |
|---|---|---|---|
| перемен. | **Среднее** | Стандарт отклон. | Дисперс. |
| **Пер1** | 1,265170 | 0,00 | 0,00 |
| Пер2 | 1,627487 | 0,00 | 0,00 |
| Пер3 | 1,560804 | 0,00 | 0,00 |
| Пер4 | 1,587831 | 0,00 | 0,00 |
| Пер5 | 1,530588 | 0,00 | 0,00 |
| Пер6 | 1,234619 | 0,00 | 0,00 |
| Пер7 | 1,427420 | 0,00 | 0,00 |
| Пер8 | 1,535814 | 0,00 | 0,00 |
| Пер9 | 1,574438 | 0,00 | 0,00 |
| Пер10 | 1,657287 | 0,00 | 0,00 |
| НовПер1 | 0,856249 | 0,00 | 0,00 |
| НовПер2 | 0,963579 | 0,00 | 0,00 |
| НовПер3 | 0,888223 | 0,00 | 0,00 |
| НовПер4 | 0,861746 | 0,00 | 0,00 |
| НовПер5 | 0,985911 | 0,00 | 0,00 |

Кластер 2:

| перемен. | Описат.статистики для кластера 2 (Таблица данных1) Кластер содержит 1 набл. Среднее | Стандарт отклон. | Дисперс. |
|---|---|---|---|
| **Пер1** | 4,209947 | 0,00 | 0,00 |
| Пер2 | 4,155233 | 0,00 | 0,00 |
| Пер3 | 4,272174 | 0,00 | 0,00 |
| Пер4 | 4,175725 | 0,00 | 0,00 |
| Пер5 | 4,231603 | 0,00 | 0,00 |
| Пер6 | 4,373209 | 0,00 | 0,00 |
| Пер7 | 4,304265 | 0,00 | 0,00 |
| Пер8 | 4,275589 | 0,00 | 0,00 |
| Пер9 | 4,274590 | 0,00 | 0,00 |
| Пер10 | 4,247516 | 0,00 | 0,00 |
| НовПер1 | 4,880336 | 0,00 | 0,00 |
| НовПер2 | 4,856192 | 0,00 | 0,00 |
| НовПер3 | 4,874125 | 0,00 | 0,00 |
| НовПер4 | 4,879826 | 0,00 | 0,00 |
| НовПер5 | 4,852501 | 0,00 | 0,00 |

Кластер 3:

| перемен. | Описат.статистики для кластера 3 (Таблица данных1) Кластер содержит 5 набл. Среднее | Стандарт отклон. | Дисперс. |
|---|---|---|---|
| **Пер1** | 0,311462 | 0,333082 | 0,110944 |
| Пер2 | 0,250653 | 0,222322 | 0,049427 |
| Пер3 | 0,198024 | 0,206753 | 0,042747 |
| Пер4 | 0,174353 | 0,213908 | 0,045757 |
| Пер5 | 0,160506 | 0,179703 | 0,032293 |
| Пер6 | 0,211008 | 0,167692 | 0,028121 |
| Пер7 | 0,210723 | 0,200907 | 0,040364 |
| Пер8 | 0,205731 | 0,234084 | 0,054795 |
| Пер9 | 0,174712 | 0,247278 | 0,061146 |
| Пер10 | 0,162845 | 0,234780 | 0,055122 |
| НовПер1 | -0,140063 | 0,054226 | 0,002941 |
| НовПер2 | -0,144587 | 0,049826 | 0,002483 |
| НовПер3 | -0,145306 | 0,049910 | 0,002491 |
| НовПер4 | -0,144777 | 0,043672 | 0,001907 |
| НовПер5 | -0,157350 | 0,047669 | 0,002272 |

Кластер 4:

| перемен. | Описат.статистики для кластера 4 (Таблица данных1) Кластер содержит 10 набл. Среднее | Стандарт отклон. | Дисперс. |
|---|---|---|---|
| **Пер1** | -0,276488 | 0,119807 | 0,014354 |
| Пер2 | -0,264010 | 0,154853 | 0,023979 |
| Пер3 | -0,274736 | 0,116503 | 0,013573 |
| Пер4 | -0,267334 | 0,186076 | 0,034624 |
| Пер5 | -0,276583 | 0,183598 | 0,033708 |
| Пер6 | -0,219827 | 0,116688 | 0,013616 |
| Пер7 | -0,197058 | 0,122212 | 0,014936 |
| Пер8 | -0,199262 | 0,106594 | 0,011362 |
| Пер9 | -0,220748 | 0,139643 | 0,019500 |
| Пер10 | -0,218115 | 0,149197 | 0,022260 |
| НовПер1 | -0,232570 | 0,071463 | 0,005107 |
| НовПер2 | -0,234211 | 0,074155 | 0,005499 |
| НовПер3 | -0,231070 | 0,071038 | 0,005046 |
| НовПер4 | -0,229700 | 0,073009 | 0,005330 |
| НовПер5 | -0,233230 | 0,072383 | 0,005239 |

Кластер 5:

| перемен. | Описат.статистики для кластера 5 (Таблица данных1) Кластер содержит 1 набл. Среднее | Стандарт отклон. | Дисперс. |
|---|---|---|---|
| **Пер1** | 1,166421 | 0,00 | 0,00 |
| Пер2 | 1,033343 | 0,00 | 0,00 |
| Пер3 | 1,027617 | 0,00 | 0,00 |
| Пер4 | 1,244331 | 0,00 | 0,00 |
| Пер5 | 1,232442 | 0,00 | 0,00 |
| Пер6 | 0,966519 | 0,00 | 0,00 |
| Пер7 | 0,893913 | 0,00 | 0,00 |
| Пер8 | 0,854107 | 0,00 | 0,00 |
| Пер9 | 0,932597 | 0,00 | 0,00 |
| Пер10 | 0,917917 | 0,00 | 0,00 |
| НовПер1 | -0,225412 | 0,00 | 0,00 |
| НовПер2 | -0,225296 | 0,00 | 0,00 |
| НовПер3 | -0,228884 | 0,00 | 0,00 |
| НовПер4 | -0,230681 | 0,00 | 0,00 |
| НовПер5 | -0,213455 | 0,00 | 0,00 |

Кластер 6:

| перемен. | Описат.статистики для кластера 6 (Таблица данных1) Кластер содержит 9 набл. Среднее | Стандарт отклон. | Дисперс. |
|---|---|---|---|
| **Пер1** | -0,603774 | 0,162713 | 0,026476 |
| Пер2 | -0,603248 | 0,199273 | 0,039710 |
| Пер3 | -0,567039 | 0,121761 | 0,014826 |
| Пер4 | -0,578480 | 0,122827 | 0,015087 |
| Пер5 | -0,559036 | 0,107754 | 0,011611 |
| Пер6 | -0,603457 | 0,157425 | 0,024783 |
| Пер7 | -0,634293 | 0,117010 | 0,013691 |
| Пер8 | -0,633505 | 0,105502 | 0,011131 |
| Пер9 | -0,605301 | 0,052180 | 0,002723 |
| Пер10 | -0,606199 | 0,064207 | 0,004123 |
| НовПер1 | -0,276128 | 0,043908 | 0,001928 |
| НовПер2 | -0,281048 | 0,042972 | 0,001847 |
| НовПер3 | -0,277360 | 0,042333 | 0,001792 |
| НовПер4 | -0,276668 | 0,042653 | 0,001819 |
| НовПер5 | -0,278434 | 0,041706 | 0,001739 |

Логично, что при формулировке выводов анализа исследователю важно понимать, какие именно фирмы относятся к той или иной кластерной группе. Для этого нужно рассмотреть информацию об элементах кластера и расстояния. В левом столбце таблицы указаны те фирмы, которые входят в конкретную группу.

Кластеры 1 и 2 представлены одной фирмой:

| Наблюд. | Элементы кластера номер 1 (Таблица данных1) и расстояния до центра кластера. Кластер содержит 1 набл. **объедин.** |
|---|---|
| **C_10** | 0,00 |

| Наблюд. | Элементы кластера номер 2 (Таблица данных1) и расстояния до центра кластера. Кластер содержит 1 набл. **объедин.** |
|---|---|
| **C_18** | 0,00 |

Состав 3 кластера:

| Наблюд. | Элементы кластера номер 3 (Таблица данных1) и расстояния до центра кластера. Кластер содержит 5 набл. объедин. |
|---|---|
| C_1 | 0,123481 |
| C_4 | 0,119557 |
| C_9 | 0,143479 |
| C_20 | 0,204566 |
| C_21 | 0,223848 |

Состав 4 кластера:

| Наблюд. | Элементы кластера номер 4 (Таблица данных1) и расстояния до центра кластера. Кластер содержит 10 набл. объедин. |
|---|---|
| C_2 | 0,073770 |
| C_6 | 0,128002 |
| C_8 | 0,087177 |
| C_14 | 0,079575 |
| C_16 | 0,062229 |
| C_17 | 0,125247 |
| C_19 | 0,103318 |
| C_23 | 0,038927 |
| C_24 | 0,210166 |
| C_26 | 0,158755 |

Состав 5 кластера:

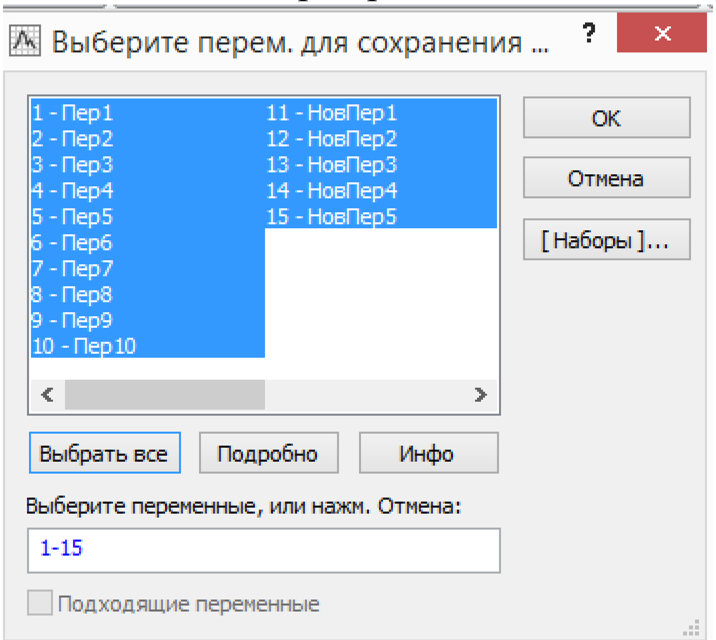
| Наблюд. | Элементы кластера номер 5 (Таблица данных1) и расстояния до центра кластера. Кластер содержит 1 набл. объедин. |
|---|---|
| C_25 | 0,00 |

Состав 6 кластера:

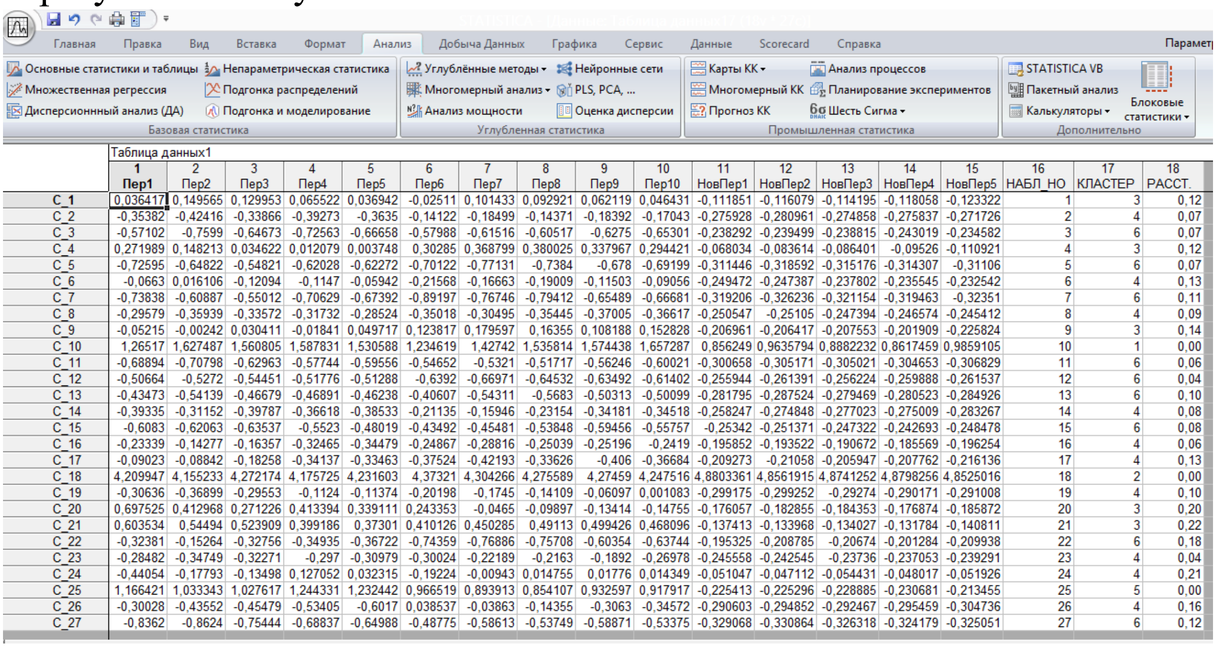
| Наблюд. | Элементы кластера номер 6 (Таблица данных1) и расстояния до центра кластера. Кластер содержит 9 набл. объедин. |
|---|---|
| C_3 | 0,071783 |
| C_5 | 0,073780 |
| C_7 | 0,112710 |
| C_11 | 0,061157 |
| C_12 | 0,042396 |
| C_13 | 0,095558 |
| C_15 | 0,076757 |
| C_22 | 0,184242 |
| C_27 | 0,120883 |

По результатам расчетом можно сделать вывод, что 4 кластер является самым многочисленным и включает в себя 10 фирм.

Общие результаты классификации и полученные расстояния при сохранении предполагают выбор переменных:




В результате получим:




| | 1 Пер1 | 2 Пер2 | 3 Пер3 | 4 Пер4 | 5 Пер5 | 6 Пер6 | 7 Пер7 | 8 Пер8 | 9 Пер9 | 10 Пер10 | 11 НовПер1 | 12 НовПер2 | 13 НовПер3 | 14 НовПер4 | 15 НовПер5 | 16 НАБЛ_НО | 17 КЛАСТЕР | 18 РАССТ. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| С_1 | 0,036417 | 0,149565 | 0,129953 | 0,065522 | 0,036942 | -0,02511 | 0,101433 | 0,092921 | 0,062119 | 0,046431 | -0,111851 | -0,116079 | -0,114195 | -0,118058 | -0,123322 | 1 | 3 | 0,12 |
| С_2 | -0,35382 | -0,42416 | -0,33866 | -0,39273 | -0,3635 | -0,14122 | -0,18499 | -0,14371 | -0,18392 | -0,17043 | -0,275928 | -0,280961 | -0,274858 | -0,275837 | -0,271726 | 2 | 4 | 0,07 |
| С_3 | -0,57102 | -0,7599 | -0,64673 | -0,72563 | -0,66658 | -0,57988 | -0,61516 | -0,60517 | -0,6275 | -0,65301 | -0,238292 | -0,239499 | -0,238815 | -0,243019 | -0,234582 | 3 | 6 | 0,07 |
| С_4 | 0,271989 | 0,148213 | 0,034622 | 0,012079 | 0,003748 | 0,30285 | 0,368799 | 0,380025 | 0,337967 | 0,294421 | -0,068034 | -0,083614 | -0,086401 | -0,09526 | -0,110921 | 4 | 3 | 0,12 |
| С_5 | -0,72595 | -0,64822 | -0,54821 | -0,62028 | -0,62272 | -0,70122 | -0,77131 | -0,7384 | -0,678 | -0,69199 | -0,311446 | -0,318592 | -0,315176 | -0,314307 | -0,31106 | 5 | 6 | 0,07 |
| С_6 | -0,0663 | 0,016106 | -0,12094 | -0,1147 | -0,05942 | -0,21568 | -0,16663 | -0,19009 | -0,11503 | -0,09056 | -0,249472 | -0,247387 | -0,237802 | -0,235545 | -0,232542 | 6 | 4 | 0,13 |
| С_7 | -0,73838 | -0,60887 | -0,55012 | -0,70629 | -0,67392 | -0,89197 | -0,76746 | -0,79412 | -0,65489 | -0,66681 | -0,319206 | -0,326236 | -0,321154 | -0,319463 | -0,32351 | 7 | 6 | 0,11 |
| С_8 | -0,29579 | -0,35939 | -0,33572 | -0,31732 | -0,28524 | -0,35018 | -0,30495 | -0,35445 | -0,37005 | -0,36617 | -0,250547 | -0,25105 | -0,247394 | -0,246574 | -0,245412 | 8 | 4 | 0,09 |
| С_9 | -0,05215 | -0,00242 | 0,030411 | -0,01841 | 0,049717 | 0,123817 | 0,179597 | 0,16355 | 0,108188 | 0,152828 | -0,206961 | -0,206417 | -0,207553 | -0,201909 | -0,225824 | 9 | 3 | 0,14 |
| С_10 | 1,26517 | 1,627487 | 1,560805 | 1,587831 | 1,530588 | 1,234619 | 1,42742 | 1,535814 | 1,574438 | 1,657287 | 0,856249 | 0,9635794 | 0,8882232 | 0,8617459 | 0,9859105 | 10 | 1 | 0,00 |
| С_11 | -0,68894 | -0,70798 | -0,62963 | -0,57744 | -0,59556 | -0,54652 | -0,5321 | -0,51717 | -0,56246 | -0,60021 | -0,300658 | -0,305171 | -0,305021 | -0,304653 | -0,306829 | 11 | 6 | 0,06 |
| С_12 | -0,50664 | -0,5272 | -0,54451 | -0,51776 | -0,51288 | -0,6392 | -0,66971 | -0,64532 | -0,63492 | -0,61402 | -0,255944 | -0,261391 | -0,256224 | -0,259888 | -0,261537 | 12 | 6 | 0,04 |
| С_13 | -0,43473 | -0,54139 | -0,46679 | -0,46891 | -0,46238 | -0,40607 | -0,54311 | -0,5683 | -0,50313 | -0,50099 | -0,281795 | -0,287524 | -0,279469 | -0,280523 | -0,284926 | 13 | 6 | 0,10 |
| С_14 | -0,39335 | -0,31152 | -0,39787 | -0,36618 | -0,38533 | -0,21135 | -0,15946 | -0,23154 | -0,34181 | -0,34518 | -0,258247 | -0,274848 | -0,277023 | -0,275009 | -0,283267 | 14 | 4 | 0,08 |
| С_15 | -0,6083 | -0,62063 | -0,63537 | -0,5523 | -0,48019 | -0,43492 | -0,45481 | -0,53848 | -0,59456 | -0,55757 | -0,25342 | -0,251371 | -0,247322 | -0,242693 | -0,248478 | 15 | 6 | 0,08 |
| С_16 | -0,23339 | -0,14277 | -0,16357 | -0,32465 | -0,34479 | -0,24867 | -0,28816 | -0,25039 | -0,25196 | -0,2419 | -0,195852 | -0,193522 | -0,190672 | -0,185569 | -0,196254 | 16 | 4 | 0,06 |
| С_17 | -0,09023 | -0,08842 | -0,18258 | -0,34137 | -0,33463 | -0,37524 | -0,42193 | -0,33626 | -0,406 | -0,36684 | -0,209273 | -0,21058 | -0,205947 | -0,207762 | -0,216136 | 17 | 4 | 0,13 |
| С_18 | 4,209947 | 4,155233 | 4,272174 | 4,175725 | 4,231603 | 4,37321 | 4,304266 | 4,275589 | 4,27459 | 4,247516 | 4,8803361 | 4,8561915 | 4,8741252 | 4,8798256 | 4,8525016 | 18 | 2 | 0,00 |
| С_19 | -0,30636 | -0,36899 | -0,29553 | -0,1124 | -0,11374 | -0,20198 | -0,1745 | -0,14109 | -0,06097 | 0,001083 | -0,299175 | -0,299252 | -0,29274 | -0,290171 | -0,291008 | 19 | 4 | 0,10 |
| С_20 | 0,697525 | 0,412968 | 0,271226 | 0,413394 | 0,339111 | 0,243353 | -0,0465 | -0,09897 | -0,13414 | -0,14755 | -0,176057 | -0,182855 | -0,184353 | -0,176874 | -0,185872 | 20 | 3 | 0,20 |
| С_21 | 0,603534 | 0,54494 | 0,523909 | 0,399186 | 0,37301 | 0,410126 | 0,450285 | 0,49113 | 0,499426 | 0,468096 | -0,137413 | -0,133968 | -0,134027 | -0,131784 | -0,140811 | 21 | 3 | 0,22 |
| С_22 | -0,32381 | -0,15264 | -0,32756 | -0,34935 | -0,36722 | -0,74359 | -0,76886 | -0,75708 | -0,60354 | -0,63744 | -0,195325 | -0,208785 | -0,20674 | -0,201284 | -0,209938 | 22 | 6 | 0,18 |
| С_23 | -0,28482 | -0,34749 | -0,32271 | -0,297 | -0,30979 | -0,30024 | -0,22189 | -0,2163 | -0,1892 | -0,26978 | -0,245558 | -0,242545 | -0,23736 | -0,237053 | -0,239291 | 23 | 4 | 0,04 |
| С_24 | -0,44054 | -0,17793 | -0,13498 | 0,127052 | 0,032315 | -0,19224 | -0,00943 | 0,014755 | 0,01776 | 0,014349 | -0,051047 | -0,047112 | -0,054431 | -0,048017 | -0,051926 | 24 | 4 | 0,21 |
| С_25 | 1,166421 | 1,033343 | 1,027617 | 1,244331 | 1,232442 | 0,966519 | 0,893913 | 0,854107 | 0,932597 | 0,917917 | -0,225413 | -0,225296 | -0,228885 | -0,230681 | -0,213455 | 25 | 5 | 0,00 |
| С_26 | -0,30028 | -0,43552 | -0,45479 | -0,53405 | -0,6017 | 0,038537 | -0,03863 | -0,14355 | -0,3063 | -0,34572 | -0,290603 | -0,294852 | -0,292467 | -0,295459 | -0,304736 | 26 | 4 | 0,16 |
| С_27 | -0,8362 | -0,8624 | -0,75444 | -0,68837 | -0,64988 | -0,48775 | -0,58613 | -0,53749 | -0,58871 | -0,53375 | -0,329068 | -0,330864 | -0,326318 | -0,324179 | -0,325051 | 27 | 6 | 0,12 |

У кластерного анализа в статистике фирмы много направлений применения: возможно исследование отдельных направлений хозяйственной деятельности, разнородных характеристик цифрового развития, трудовых параметров и т.д.

# Глава 8. ДИСКРИМИНАНТНЫЙ АНАЛИЗ В СТАТИСТИКЕ ФИРМЫ

## 8.1. Основы дискриминантного анализа

Задача дискриминантного анализа опирается на следующий набор данных. Допустим, что имеются сведения о n наблюдениях, каждое из которых можно охарактеризоваться по k признакам. В таком случае каждое конкретное наблюдение может быть описано с помощью вектора x, характеристики которого являются набором случайных величин (8.1):

$$X = (X_1, X_2, \ldots X_k)^T \tag{8.1}$$

Суть задачи дискриминации состоит в том, чтобы разбить все множество реализации анализируемой величины на определенное количество областей $R_i$, затем каждое из новых анализируемых наблюдений относится к какой-либо области опираясь на правило, которое признано в рамках решения конкретной исследовательской задачи решающим. Предполагается, что заранее информация о принадлежности объекта к области недоступна, либо требует значительных затрат ресурсов на ее получение.

Выбор правила осуществления дискриминантного анализа должен базироваться на принципе оптимальности, который представляет собой минимизацию средних потерь от неправильной классификации, исходя из априорных вероятностей pi извлечения объекта из группы Ri. Решающее правило считается наилучшим в определенном смысле слова, если никакое другое правило не может дать меньшей величины функции потерь.

Значения априорных вероятностей могут быть известны заранее, и определены пользователями заблаговременно (по результатам предварительного анализа), либо заданы в процессе ввода данных модуль. В качестве средних потерь чаще всего принимают вероятность ложной классификации наблюдения.

Построение решающего правила также можно рассматривать как задачу поиска областей R, которые не пересекаются между собой. Дискриминантные функции в этом случае дают определение этих областей путем задания их границ в многомерном пространстве.

В процессе дискриминантного анализа автоматически вычисляются функции классификации, предназначенные для определения той группы, к которой наиболее вероятно принадлежит новый объект. При этом важно равенство количества функций классификации заданной величине кластерных групп.

Считается, что принадлежность наблюдения в определенной группе является объясненной в том случае, если функция классификация максимальна, либо значение апостериорной функции является наибольшим.

Дискриминантный анализ используется для исследования различий заранее заданных групп объектов исследования (фирм, категорий товаров и т.д.). Переменная является группирующей в том случае, если она делит совокупность объектов исследования на конкретные классы или категории. Дискриминантный анализ используется для того, чтобы оценить межгрупповые развития по определенным признакам. Если признак применяется для выявления различий между группами, они именуются дискриминационными переменными.

Важно грамотное использование шкалирования при анализе: Группирующая переменная должна принадлежать номинальной шкале, а зависимые характеристики быть метрическими. Соблюдение данного условия необходимо для обеспечения высокой точности производимых расчетов. В практической деятельности возможно, чтобы группирующая переменная принадлежала также к порядковой школе, а дискриминационные - к шкале любого типа, однако, надо четко понимать конкретные исследовательские задачи и объекты аналитической практики.

В результате проведения дискриминантного анализа осуществляется построение модели (дискриминантной функции), общий вид которой (8.2):

$$D = b_0 + b_i X_1 + b_2 X_2 + \ldots + b_k X_k \tag{8.2}$$

где D – группирующая (зависимая) переменная;

$b_k$ – коэффициенты дискриминантной функции;

$b_o$ – свободный член (константа);

$x_n$ – дискриминационные (независимые) переменные.

Данная модель позволяет определять принадлежность каждого конкретного объекта к группе, опираясь на базовые характеристики исследования, признанные исследователем как значимые.

Основные цели дискриминантного анализа отражены на рисунке 8.1.


*Рис. 8.1. Основные виды дискриминантного анализа*

Целесообразность дискриминантного анализа определяется на основе исследовательских задач.

К статистикам, используемым в дискриминантном анализе, относятся следующие коэффициенты и показатели.

**Каноническая корреляция** используется для измерения степени связи между дискриминантными показателями и группами. Это мера связи между единственной дискриминирующей функцией и набором фиктивных переменных, которые определяют принадлежность к данной группе.

**Центроид (средняя точка).** Центроид – это средние значения для дискриминантных показателей конкретной группы. Центроидов столько, сколько групп, т.е. один центроид для каждой группы. Средние группы для всех функций – это групповые центроиды.

**Классификационная матрица.** Иногда ее называют смешанной матрицей, или матрицей предсказания. Классификационная матрица содержит ряд правильно классифицированных и ошибочно классифицированных случаев. Верно классифицированные случаи лежат на диагонали матрицы, поскольку предсказанные и фактические группы одни и те же. Элементы, не лежащие по диагзнали матрицы, представляют случаи, классифицированные ошибочно. Сумма элементов, лежащих на диагонали, деленная на общее количество случаев, дает коэффициент результативности.

**Коэффициенты дискриминантной функции**. Коэффициенты дискриминантной функции (ненормированные) – это коэффициенты переменных, когда они измерены в первоначальных единицах.

**Дискриминантные показатели**. Сумма произведений ненормированных коэффициентов дискриминантной функции на значения переменных, добавленная к постоянному члену.

**Собственное (характеристическое) значение.** Для каждой дискриминантной функции собственное значение – это отношение межгрупповой суммы квадратов к внутри-групповой сумме квадратов. Большие собственные значения указывают на функции более высокого порядка.

**F – статистика и ее значимость**. Значения F -статистики вычисляют однофакторный дисперсионный анализ, разбивая на группы независимую переменную. Каждый предиктор, в свою очередь, служит в ANOVA метрической зависимой переменной.

**Средние группы и групповые стандартные отклонения.** Эти показатели вычисляют для каждого предиктора каждой группы.

**Объединенная межгрупповая корреляционная матрица.** Объединенную межгрупповую корреляционную матрицу вычисляют усреднением отдельных ковариационных матриц для всех групп.

**Нормированные коэффициенты дискриминантных функций.** Коэффициенты дискриминантных функций используют как множители для нормированных переменных, т.е. переменных с нулевым средним и дисперсией, равной 1.

**Структурные коэффициенты корреляции**. Также известны как дискри-минантные нагрузки, представляют собой линейные коэффициенты корреляции между предикторами и дискриминантной функцией.

**Общая корреляционная матрица.** Если при вычислении корреляций наблюдения обрабатывают так, как будто они взяты из одной выборки, то в результате получают общую корреляционную матрицу.

**Коэффициент λ Уилкса.** Иногда называемый U-статистикой, коэффициент λ Уилкса для каждого предиктора – это отношение внутри групповой суммы квадратов к общей сумме квадратов. Его значение варьирует от 0 до 1. Большое значение λ (около 1) указывает на то, что средние групп не должны различаться. Малые значения λ (около 0) указывают на то, что средние групп различаются.

## 8.2. Реализация процедуры дискриминантного анализа

Процедура выполнения дискриминантного анализа состоит из шести основных шагов (рис. 8.2).

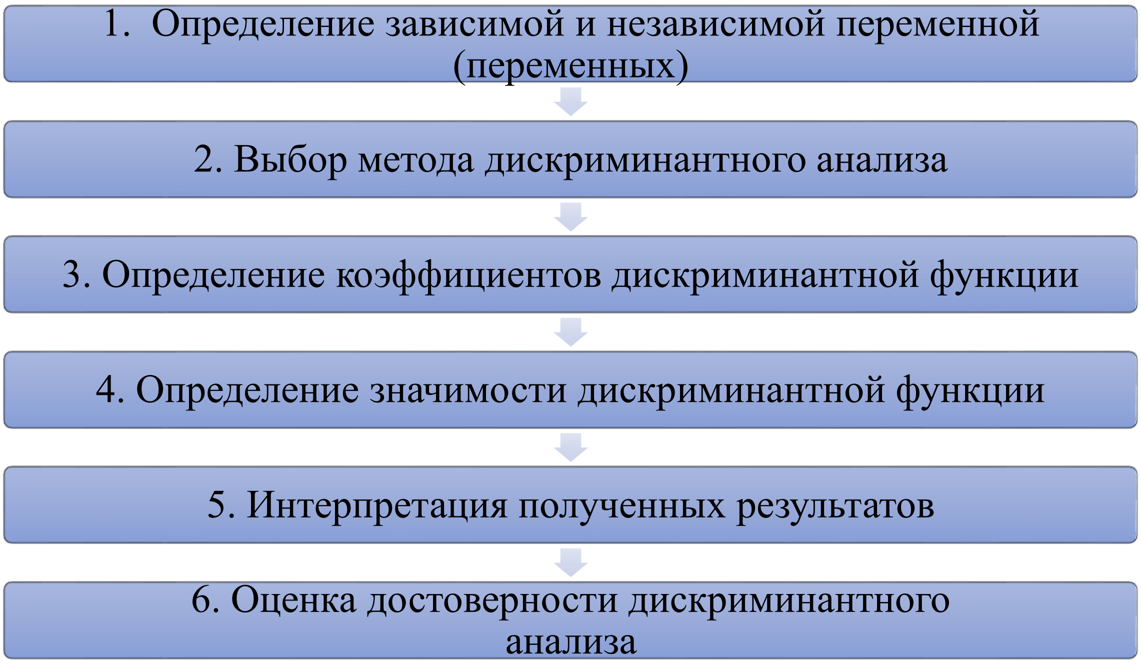



*Рис. 8.2. Этапы процедуры дискриминантного анализа*

На первом шаге анализа  определяется, что является независимой переменной, а что результирующим фактором. Зависимая переменная должна состоять из двух или больше взаимоисключающих и взаимно исчерпывающих категорий. Если зависимая переменная измерена с помощью интервальной или относительной шкалы, то ее следует, в первую очередь, перевести к статусу категориальной. Например, отношение к фирме, измеренное по девятибалльной шкале, можно категоризировать как неблагоприятное (1, 2, 3, 4), нейтральное (5) и благоприятное (6, 7, 8, 9). Также возможно формирование одинаковых групп точками отсечек по результатам графического анализа графика распределения значений зависимой переменной. Предикторы следует выбирать, исходя из теоретической модели или ранее проверенного исследования, или, в случае поискового исследования, из интуиции и опыта исследователя.

Далее выборку делят на две части. Одна из них – анализируемая выборка – используется для вычисления дискриминантной функции. Другая часть – проверочная выборка используется для проверки результатов дискриминантного моделирования.

Если исследователь сталкивается с необходимостью работы с большой по объему выборкой, ее можно разбить на две меньшие равные совокупности, первая из которых будет анализируемой, а вторая проверочной.  После анализ повторяется, однако выполняемые выборками функции меняются: анализируемая становится проверочной, а проверочная анализируемой. Таким образом, осуществляется двойная перекрестная проверка результатов моделирования.

Часто распределение количества случаев в анализируемой и проверочной выборки явствует из распределения в общей выборке. Рассмотрим конкретный пример. Предположим, что 50% фирм выборки развиваются стабильно, а в деятельности другой половины организаций часто возникают проблемы под влиянием внешних факторов. Таким образом, ½ анализируемых фирм более подвержены риску, а другая ½ - менее.  Или предположим, что четверть организации менее подвержены внешним риска и угрозам. В таком случае выбор проверочной и анализируемой выборки должен руководствоваться теми же соотношениями (25% : 75%).

Для выбора предикторов в дискриминантной функции можно использовать два метода (рис. 8.3.)


*Рис. 8.3 Методы выбора предикторов в дискриминантно функции*

На втором этапе осуществляется выбор метода дискриминантного анализа, который описывается числом категорий, имеющихся у зависимой переменной. Выделяют дискриминантный анализ для двух групп и множественный анализ: при осуществлении первого анализируется 2 категории, второй базируется на анализе трех и более категорий.

Главное отличие между ними заключается в том, что при наличии двух групп возможно вывести только одну дискриминантную функцию. Используя множественный дискриминантный анализ, можно вычислить несколько функций.

Рассмотрим случай для двух дискриминантных переменных. Тогда определение коэффициентов дискриминантной функции будет сведено к следующему.

Функция f(X) называется канонической дискриминантной функцией, а величины $x_1$ и $x_2$ – дискриминантными переменными (8.3):

$$f(x) = a_1X_1 + a_2X_2 \tag{8.3}$$

Дискриминантная функция может быть как линейной, так и нелинейной. Выбор вида этой функции зависит от геометрического расположения разделяемых классов в пространстве дискриминантных переменных.

Коэффициенты дискриминантной функции ($a_i$) определяются таким образом, чтобы $\bar{f}_1(X)$ и $\bar{f}_2(X)$ как можно больше отличались между собой.

Вектор коэффициентов дискриминантной функции *(*A*)* определяется по формуле (8.4):

$$A = S_*^{-1}(\overline{X}_1 - \overline{X}_2) \tag{8.4}$$

Полученные значения коэффициентов подставляют в формулу и для каждого объекта в обоих множествах вычисляют дискриминантные функции f(X*)*, затем находят среднее значение для каждой группы *(*$\bar{f}_k$).

Таким образом, каждому *i*-му наблюдению, которое первоначально описывалось *m*-переменными, будет соответствовать одно значение дискриминантной функции, и размерность признакового пространства снижается.

Перед тем как приступить непосредственно к процедуре классификации, нужно определить границу, разделяющую два множества. Такой величиной может быть значение функции, равноудаленное от $\bar{f}_1$ и $\bar{f}_2$ (8.5):

$$c = \frac{1}{2}(\bar{f}_1 + \bar{f}_2) \qquad (8.5)$$

Величина с называется константой дискриминации.

Объекты, расположенные над разделяющей поверхностью $f(x) = a_1x_1 + a_2x_2+\ldots+ a_px_p=c$ находятся ближе к центру множества $M_1$, следовательно, могут быть отнесены к первой группе, а объекты, расположенные ниже этой поверхности, ближе к центру второго множества, т.е. относятся ко второй группе.

Если граница между группами будет выбрана как сказано выше, то в этом случае суммарная вероятность ошибочной классификации будет минимальной.

Важнейшим этапом дискриминантного исследования является анализ значимости результатов моделирования.

Бессмысленно интерпретировать результаты анализа, если определенные дискриминантные функции не являются статистически значимыми. Поэтому следует выполнить статистическую проверку нулевой гипотезы о равенстве средних всех дискриминантных функций во всех группах генеральной совокупности. В программе SPSS эта проверка базируется на коэффициенте лямбда ($\lambda$) Уилкса. Если одновременно проверяют несколько функций, как в случае множественного дискриминантного анализа, то коэффициент $\lambda$ является суммой одномерных $\lambda$ для каждой функции. Уровень значимости оценивают, исходя из преобразования $\lambda$ -статистики в статистику хи-квадрат (исходя из распределения хи-квадрат, которому подчиняется $\lambda$ -статистика). Если нулевую гипотезу отклоняют, что указывает на значимую дискриминацию, то можно продолжать интерпретировать результаты.

Для интерпретации дискриминантных весов используется процедура, аналогичная множественному регрессионному анализу.

Значение коэффициента для конкретного предиктора зависит от других предикторов, включенных в дискриминантную функцию. Знаки коэффициентов условны, но они указывают, какие значения переменной приводят к большим и маленьким значениям функции и связывают их с конкретными группами.

При наличии мультиколлинеарности между независимыми переменными не существует однозначной меры относительной важности предикторов для дискриминации между группами. Помня об этом предостережении, можно получить некоторое представление об относительной важности переменных, изучив абсолютные значения нормированных коэффициентов дискриминантной функции. Как правило, предикторы с относительно большими нормированными коэффициентами вносят больший вклад в дискриминирующую мощность функции по сравнению с предикторами, имеющими меньшие коэффициенты.

Некоторое представление об относительной важности предикторов можно также получить, изучив структурные коэффициенты корреляции, которые также называют каноническими или дискриминантными нагрузками. Эти линейные коэффициенты корреляции между каждым из предикторов и дискриминантной функцией представляют дисперсию, которую предиктор делит вместе с функцией. Как и нормированные коэффициенты, эти коэффициенты корреляции следует использовать осторожно.

При интерпретации результатов дискриминантного анализа также может помочь разработка характеристической структуры для каждой группы посредством описания каждой группы через групповые средние для предикторов.

Оценка достоверности дискриминантного анализа является заключительным этапом исследования.

Как уже говорилось, данные разбивают случайным образом на две подвыборки. Анализируемую часть выборки используют для вычисления дискриминантной функции, а проверочную – для построения классификационной матрицы.

Дискриминантные веса, определенные анализируемой выборкой, умножают на значения независимых переменных в проверочной выборке, чтобы получить дискриминантные показатели для случаев в этой выборке. Затем случаи распределяют по группам, исходя из дискриминантных показателей и соответствующего правила принятия решения. Например, при дискриминантном анализе двух групп случай может быть отнесен к группе с самым близким по значению центроидом. Затем, сложив элементы, лежащие на диагонали матрицы, и разделив полученную сумму на общее количество случаев, можно определить коэффициент результативности или процент верно классифицированных случаев. Полезно сравнить процент случаев, верно классифицированных с помощью дискриминантного анализа, с процентом случаев, который можно получить случайным образом. Для равных по размеру групп процент случайной классификации равен частному от деления единицы на количество групп. Превысит ли и насколько количество верно классифицированных случаев их случайное количество? Здесь нет общепринятого подхода, хотя некоторые авторы считают, что точность классификации, достигнутая с помощью дискриминантного анализа, должна быть, по крайней мере, на 25% выше, чем точность, которую можно достичь случайным образом.

Большинство программ для выполнения дискриминантного анализа также определяют классификационную матрицу, исходя из анализируемой выборки. Поскольку программы учитывают даже случайные вариации в данных, то полученные результаты всегда точнее, чем классификация данных на основе проверочной выборки.

**Контрольные вопросы**

1. В чем состоит задача дискриминантного анализа?
2. Какие данные необходимы для проведения дискриминантного анализа?
3. В чем состоит суть принципа оптимальности?
4. Что такое априорная вероятность и как она определяется?
5. В каком случае принадлежность наблюдения в определенной группе является объясненной?
6. В каком случае переменная является группирующей?
7. Какой шкале должна принадлежать группирующая пременная?
8. Какой шкале должны зависимые пременные?
9. Опишите общий вид модели дискриминантной функции.
10. Каковы основные цели осуществления дискриминантного анализа.
11. Для чегоо используется каноническая корреляции при осуществлении процедуры дискриминантного анализа?

12. Что представляет собой центроид?
13. Что представляет собой классификационная матрица и каковы цели ее построения?
14. В каких единицах измерения представлены ненормированные коэффициенты дискриминантной функции?
15. Какие этапы процедуры дискрисинантного анализа Вам известны?
16. Что представляет собой собственное (характеристическое значение) и каким способон оно рассчитывается?
17. Что представляет собой F-статистика?
18. На какие группы принято разбивать совокупность при проверке результатов дисперсонного анализа с использованием F-статистики?
19. Какой вид корреляционной матрицы вычисляют усреднением отдельных ковариационных матриц для всех групп?э
20. Что представляют собой нормированные коэффициенты дискриминантных функций?
21. Объясните значение структурных коэффициентов корреляции при осуществлении процедуры дискриминантного анализа
22. Какая матрица получается в том случае, если при вычислении корреляций наблюдения обрабатывают так, как будто они взяты из одной выборки?
23. В каких пределах может меняться коэффициент Уилкса?
24. Каковы варианты интерпретации коэффициента Уилкса?
25. С помощью каких шагов реализуется процедура дискриминантного анализа?
26. Для каких целей нужна анализируемая выборка?
27. Что можно оценить с помощью проверочной выборки?
28. Какова процедура реализации прямого метода выбора предикторов?
29. Как вводятся предикторы при пошаговом дискриминантном анализе?
30. В каких случаях целесообразно применять пошаговый метод дискрминантного анализа?
31. Когда следует применять прямой метод при осуществении процедуры дискриминантного нанлиза?
32. Какая процедура используется для интерпретации дискриминантных весов?

# Глава 9. ФАКТОРНЫЙ АНАЛИЗ

## 9.1. Основные элементы факторного анализа

Факторный анализ является процедурой, которая используется для того, чтобы с помощью меньшего количества переменных, именуемых факторами, описать процесс или явление, характеристики которого ранее были представлены большим количеством величин.

Объединение переменных в факторы должно осуществляться на основе подходов, представленных на рисунке 9.1.

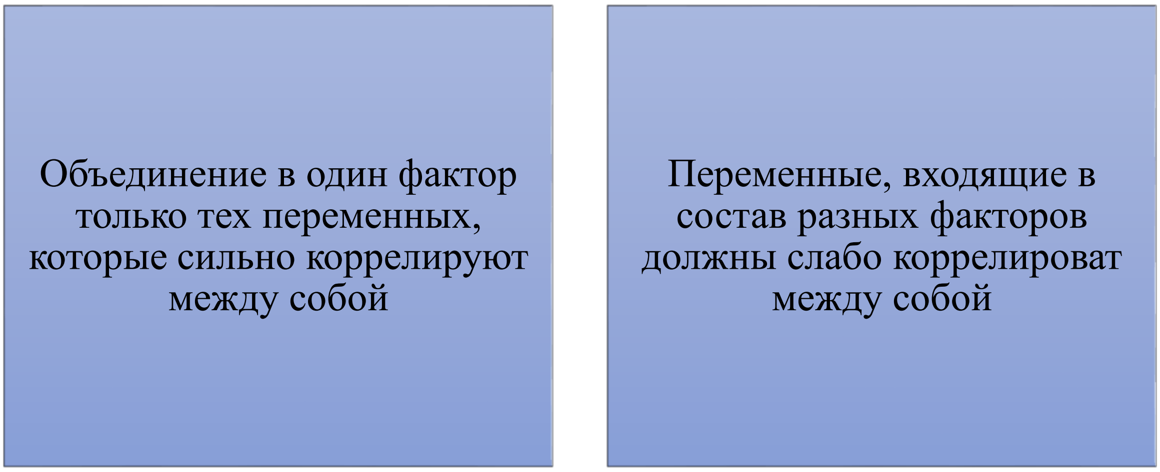



*Рис. 9.1. Принципы объединения переменных в факторы*

При осуществлении факторного анализа важно понимать, что означает термин «фактор». Под определением данного понятия принято понимать скрытые переменные, именуемые в некоторых источниках «латентные» Факторы не подлежат непосредственному изменению, однако, тесно связаны с измеряемыми параметрами, которые являются проявлением данных факторов.

Сущность факторного анализа состоит в предположении, что имеет место ряд величин, благодаря которым проявляются отношения различного характера между переменными. Также предполагается, что данные величины не известны исследователю.

Для объяснения структуры связей между признаками, подлежащими анализу $x^{(1)}$, $x^{(2)}$, $x^{(3)}$….$x^{(p)}$используется их линейная или иная зависимость от меньшего числа иных факторов, которые непосредственно не измерялись ($f^{(1)},\ldots, f^{(m)}$ ($m < p$)). Данные факторы принято называть общими.

На основании вышеизложенного можно сделать вывод о том, что в широком смысле факторный анализ представляет собой совокупность методов и моделей, основное назначение которых состоит в выявлении, построении и исследовании факторов, определенных на основе их внешних проявлений.

В узком смысле данный термин используется для обозначения методов, которые используются для выявления скрытых, ненаблюдаемых факторов. За счет них объясняются имеющиеся корреляционные матрицы, построенные на основе известных результатов количественных наблюдаемых переменных.

Абсолютное большинство моделей предполагает такое их построение, при котором факторы, определенные по результатам анализа, не коррелируют друг с другом. Также предполагается, что отсутствует возможность однозначного восстановления каждого из значений признака на основе определенных по результатам исследования общих факторов $f(m)$. На каждую переменную также оказывает влияние случайная составляющая e(j). За счет нее и возникает статистическая связь между x(j) и f (m).

Любой факторный анализ по сути сводится к выявлению общих факторов и минимизации степени зависимости признака x(j) от собственных случайных компонент e(j). В последующем происходит интерпретация результатов вычисления $f(m)$.

Поскольку факторный анализ является модельной схемой, абсолютное достижение постеленной цели невозможно. Таким образом, происходит приближенное вычисление. Результаты факторного анализа можно считать успешными в том случае, если удалось объяснить как можно большее число исходных переменных минимальным числом определенных факторов, которые, в зависимости от условий моделирования, могут выполнять роль исходных причин, либо агрегированных теоретических конструкций.

Особенности осуществления факторного анализа должны базироваться на понимании конкретных условий исследования. На практике принято выделять 2 основных уровня (рис.9.2).

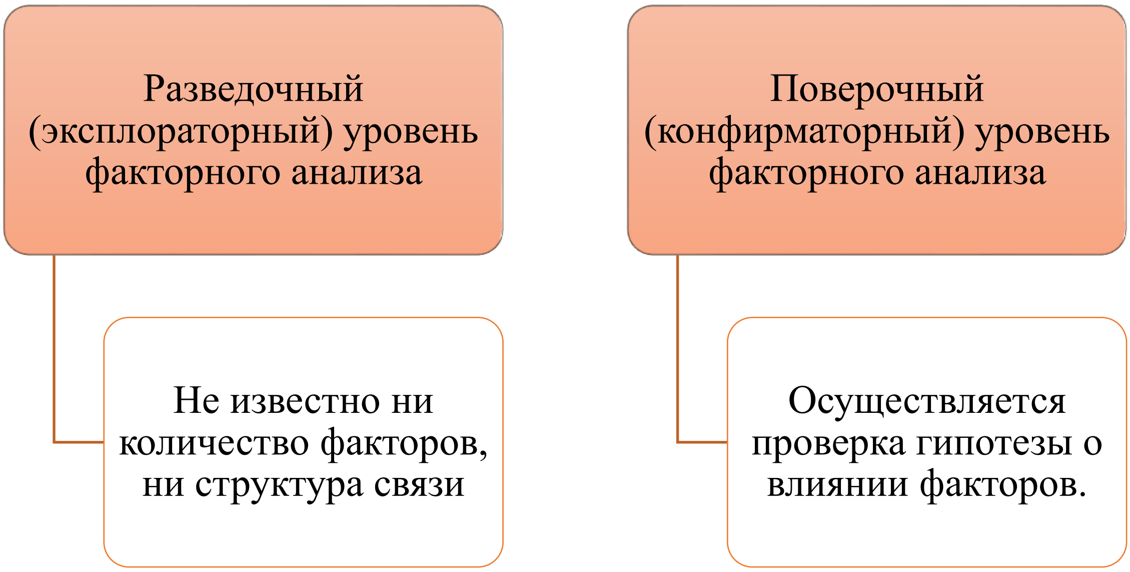



*Рис. 9.2. Основные уровни факторного анализа*

Методы факторного анализа направлены на то, чтобы снизить размерность массива исходных данных, который описывает процесс или явление, на основе сжатия. Достижение данного результата осуществляется, как правило, за счет наличия корреляционных связей в исходном массиве признаков.

Основные этапы осуществления метода представлены на рис. 9.3.


*Рис. 9.3. Основные этапы факторного анализа*

Выделяют 2 основных типа целей в рамках осуществления факторного анализа (рис. 9.4).

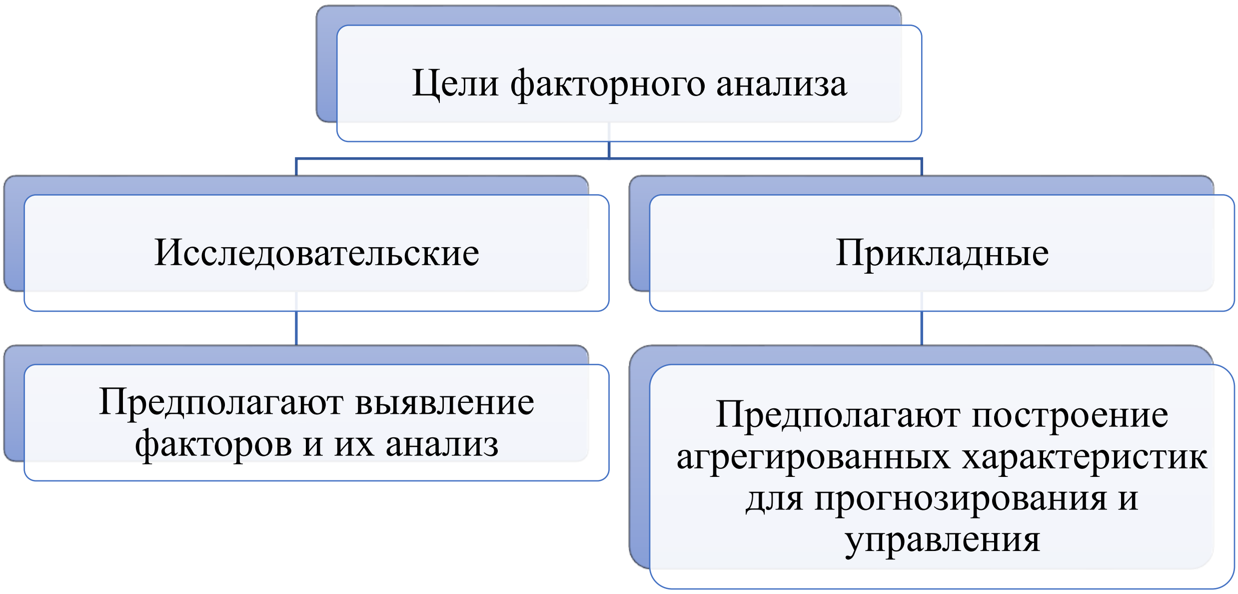



*Рис. 9.4. Основные этапы факторного анализа*

Традиционные представления о факторном анализе при всем многообразии решаемых с его помощью задач сводятся к двум основным положениям:

- ✓ Сущность основных процессов заключается в их многообразных, но одновременно с этим простых проявлениях. Они могут быть объяснены при помощи нескольких обобщающих факторов
- ✓ Идея постижения сущности наблюдаемых процессов и явлений реализуется с помощью бесконечного приближения.

Основываясь на данных началах, с помощью факторного анализа появляется возможность получения относительно простых моделей, объясняющих взаимосвязи на причинном уровне. Метод факторного анализа активно используется как в теоретических, так и практических исследованиях, в том числе в области экономики.

Для того, чтобы решение задачи факторного анализа стало возможным, необходимо, чтобы соблюдались основные условия применения метода (рис. 9.5).

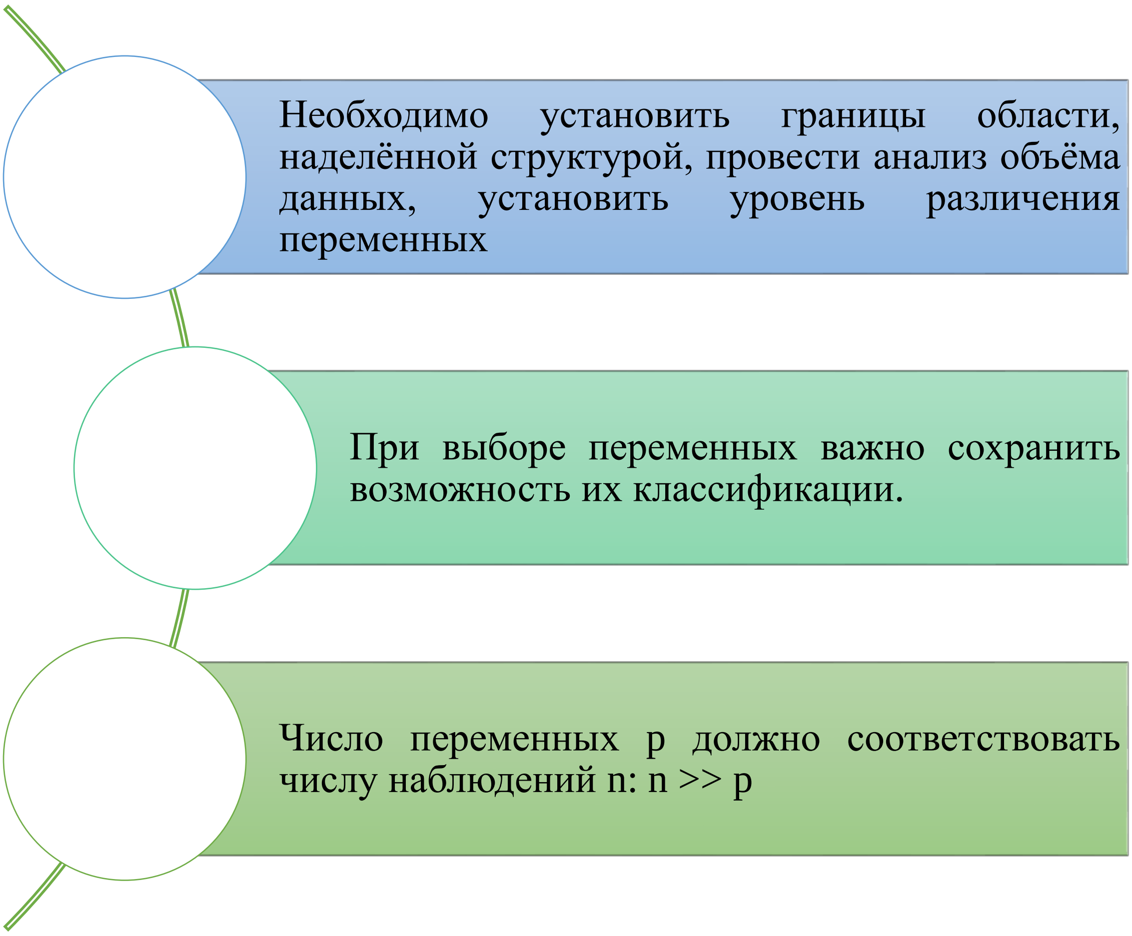



*Рис. 9.5. Основные условия осуществления факторного анализа*

С помощью факторного анализа становится возможным определение количественных отношений между переменными. Модель можно выразить как при помощи коэффициентов, так и в процентном отношении.

Существуют определенные объекты, по отношению которым возможно проведение процедуры факторного анализа.

К ним можно отнести:

✓ Вектор отклонений (9.1):

$$y = x - \overline{x}, \tag{9.1}$$

причем $E(x - \overline{x}) = E(y = 0)$

В таком случае исходный материал для анализа представлен матрицей ковариаций (8.2):

$$К = E(yy') \tag{9.2}$$

✓ Нормированный вектор (8.3):

$$z = S^{-1}(x - \overline{x}) = S^{-1}y \qquad (9.3)$$

где S-диагональная матрица стандартных отклонений, E(z = 0).

В таком случае исходный материал представлен ковариационной матрицей R, которая в данном случае равна корреляционной.

✓ Наблюдаемый вектор X без поправки на среднее.

Матрица наблюдений при традиционном подходе представляется в виде (8.4):

$$X = (x_{ik}), \qquad (9.4)$$

Где $x_{ik}$ значение k -го признака для i -го объекта, в виде линейных комбинаций значений $f_{it}$ факторов $f_t$ на объектах с невязками $e_{ik}$ (8.5):

$$x_{ik} = a_{1k}f_{i1} + ... + a_{tk}f_{it} + e_{ik}, \qquad (9.5)$$

причем i = 1….n,

k = 1….p,

t = 1 ….m.

$a_{tk}$ - нагрузки факторов на признак k.

Соотношения факторного анализа формально воспроизводят запись модели множественных регрессий, в которой под $f^{(i)}$ (i = 1,2,…,m) понимаются так называемые объясняющие переменные (факторы-аргументы).

Однако принципиальное отличие модели факторного анализа от регрессионных схем состоит в том, что переменные $f^{(j)}$ , выступающие в роли аргументов в моделях регрессии, не являются непосредственно наблюдаемыми в моделях факторного анализа, в то время как в регрессионном анализе значения $f^{(j)}$ измеряются на статистически обследованных объектах.

В матричном виде модель выглядит следующим образом (9.6):

$$X=AF+X \qquad (9.6)$$

В ходе факторного анализа необходимо оценить минимальное число факторов, определить векторы факторных нагрузок, и вычислить значения факторов для каждого наблюдаемого объекта. Поскольку число обобщенных факторов предполагается существенно меньше числа исходных признаков, данная задача не имеет однозначного решения. В зависимости от того, какие условия накладываются на обобщенные факторы, существуют различные модели и методы факторного анализа. Можно выделить два основных подхода к построению факторных моделей. В первом случае обобщенные факторы должны выделять большую часть суммарной дисперсии исходных факторов, во втором - обобщенные факторы должны наилучшим образом описывать ковариацию между исходными факторами. Первый метод, по сути, есть рассмотренный ранее метод главных компонент (при условии, что мы строим обобщенные факторы, как линейные комбинации исходных), отличный лишь условиями нормировки, накладываемыми на коэффициенты преобразования. Методы второй группы принято называть каноническими методами факторного анализа.

### 9.2. Метод главных компонент в факторном анализе

Суть метода главных компонент состоит в поиске обобщенных факторов $f^{(1)}, f^{(2)}, \ldots, f^{(m)}$ как линейных комбинаций исходных признаков $\xi_1, \xi_2, \ldots, \xi_k$ (9.7):

$$f^{(i)} = \beta^{(i)T}\xi = \beta_1^{(i)}\xi_1 + \beta_2^{(i)}\xi_2 + \ldots + \beta_k^{(i)}\xi_k \tag{9.7}$$

где $\beta^{(i)}$ – векторы неизвестных коэффициентов преобразования, $i = \overline{1,m}$, $(m \le k)$,

$\xi = (\xi_1, \xi_2, \ldots, \xi_k)$ – центрированная многомерная случайная величина с матрицей ковариаций A.

Обобщенные факторы $f^{(1)}, f^{(2)}, \ldots, f^{(m)}$ в рамках реализации метода предполагают некоррелированными случайные величины с нулевым математическим ожиданием и единичной дисперсией. Важно, что выбор линейных комбинаций является процедурой, которая зависит от ряда условий, основное из которых состоит в том, первый из обобщённых факторов объяснял бы наиболее существенную часть дисперсии исходных признаков.

Выбор каждого последующего обобщающего признака базируется на тех же принципах: он должен объяснять максимальную долю оставшейся дисперсии. Кроме того, необходимо, чтобы не наблюдалась межфакторная корреляция.

Результат построения по данной схеме приводит к тому, что в качестве векторов $\beta^{(i)}$, $i=\overline{1,m}$ принимаются собственные векторы матрицы ковариаций А, систематизированные в порядке убывания.

Абсолютно линейное преобразование величин $\xi_j$, $j=\overline{1,k}$ в факторы $f^{(i)}$, $i=\overline{1,k}$ наблюдается, если m=k. В таком случае векторы факторных нагрузок $\alpha^{(i)}$ также должны являться собственными векторами матрицы ковариаций $A$.

Условный вид нормировок для векторов факторных нагрузок обусловлен тем, что дисперсия, которая объясняется фактором, равна соответствующему собственному значению (9.8):

$$D\left(\sum_{j=1}^{k}\alpha_j^{(i)}f^{(i)}\right)=\sum_{j=1}^{k}(\alpha_j^{(i)})^2 D(f^{(i)})\equiv\lambda_i \quad\Rightarrow\quad |\alpha^{(i)}|^2=\lambda_i \qquad (9.8)$$

Векторы $\beta^{(i)}$ должны удовлетворять условию (9.9):

$$D(f^{(i)})=D\left(\sum_{j=1}^{k}\beta_j^{(i)}\xi_j\right)=\beta^{(i)^{\mathrm{T}}}A\beta^{(i)}=1,\ i=\overline{1,m} \qquad (9.9)$$

Поскольку данные векторы также являются собственными векторами матрицы ковариаций (9.10):

$$|\beta^{(i)}|^2=1/\lambda_i,\quad \alpha^{(i)}=\lambda_i\,\beta^{(i)} \text{ или } \alpha^{(i)}=A\beta^{(i)},\ i=\overline{1,m} \qquad (9.10)$$

Кроме того, координаты вектора факторных нагрузок являются ковариациями между фактором и соответствующим исходным параметром (9.11):

$$\mathrm{cov}(\xi_j,f^{(s)})=\mathrm{cov}(\xi_j,\sum_{i=1}^{k}\beta_i^{(s)}\xi_i)=\sum_{i=1}^{k}\mathrm{cov}(\xi_j,\xi_i)\beta_i^{(s)}=\alpha_j^{(s)} \qquad (9.11)$$

Критерии отбора главных факторов в этом случае аналогичны рассмотренным ранее в методе главных компонент.

## 9.3. Каноническая модель факторного анализа

Канонической называют такую модель, для котороый выполняется ряд условий.

Пусть $A=M(\xi\xi^{\mathrm{T}})$ - ковариационная матрица исходных признаков, а $\Sigma=M(\varepsilon\varepsilon^T)$ - диагональная матрица ковариаций характерных факторов, тогда справедливо следующее соотношение (9.12):

$$A=M((\alpha f+\varepsilon)(\alpha f+\varepsilon)^{\mathrm{T}})=\alpha M(ff^{\mathrm{T}})\alpha^{\mathrm{T}}+\Sigma=\alpha\alpha^{\mathrm{T}}+\Sigma \qquad (9.12)$$

или (9.13)

$$\begin{cases} \operatorname{cov}(\xi_i, \xi_j) = \sum\limits_{s=1}^{m} \alpha_i^{(s)} \alpha_j^{(s)}, \quad i = \overline{1,k}, \quad j = \overline{1,k}, \quad i \neq j \\ D(\xi_i) = \sum\limits_{s=1}^{m} \left(\alpha_i^{(s)}\right)^2 + D(\varepsilon_i), \quad i = \overline{1,k} \end{cases} \tag{9.13}$$

Таким образом, ковариации исходных признаков полностью воспроизводятся матрицей нагрузок, а для воспроизведения их дисперсий нужны также дисперсии характерных факторов. Так как (9.14):

$$M(\xi f^T) = M((\alpha f + \varepsilon) f^T) = \alpha M(f f^T) = \alpha, \tag{9.14}$$

то также, как и в методе главных компонент (9.15)

$$\operatorname{cov}(\xi_j, f^{(s)}) = \alpha_j^{(s)}, \quad j = \overline{1,k}, \quad s = \overline{1,m} \tag{9.15}$$

Если соблюдаются условия системы уравнений (9.13), то модель факторного анализа является канонической. Число уравнений в системе может быть определено по формуле (9.16):

$$k(k+1)/2 \tag{9.16}$$

Число неизвестных параметров равно (9.17):

$$km + k = k(m+1). \tag{9.17}$$

В случае, если решение системы существует, оно может рыть рассчитано с точностью до ортогонального преобразования матрицы факторных нагрузок.

Задача факторного анализа обладает единственным решением, если для матрицы факторных нагрузок существуют дополнительные ограничения. Поскольку ортогональное преобразование порядка $m$ однозначно определяется заданием $m(m-1)/2$ элементов, то необходимо, соответственно, обозначить дополнительные $m(m-1)/2$ условий на параметры модели. Обычно полагают, что для элементов матрицы $\alpha = (\alpha^{(1)}, \alpha^{(2)}, \ldots, \alpha^{(m)})$ выполнено одно из следующих условий

- матрица $\alpha^T \Sigma^{-1} \alpha$ должна быть диагональной;
- матрица $B^T \alpha$, где $B$ заданная матрица порядков $k \times m$, должна быть нижней треугольной матрицей.

Задача канонического факторного анализа в общем случае не имеет аналитического решения и может быть решена только численно, используя ту или иную итерационную процедуру.

## 9.4. Параметры факторного моделирования при применении метода максимального правдоподобия

Решение задачи методом максимального правдоподобия предполагает допущение, что обобщенные и характерные факторы имеют нормальное распределение (то есть для исходных признаков характерно многомерное нормальное распределение), а матрица факторных нагрузок удовлетворяет условию (9.18):

$$\alpha^{T}\Sigma^{-1}\alpha = J \tag{9.18}$$

где J представляет собой диагональную матрицу, элементы которой расположены в порядке убывания.

Число дополнительных условий моделирования составляет $m(m-1)/2$. Таким образом, количество условий на параметры теперь равно $k(k+1)/2 + m(m-1)/2$, а число неизвестных параметров по прежнему равно $k(m+1)$.

Чтобы задача не была неопределенной, требуется, чтобы $k(k+1)/2 + m(m-1)/2 \geq k(m+1)$ или $(k-m)^2 \geq (k+m)$.

Так как при $(k-m)^2 = (k+m)$ решение теряет статистическую значимость, то максимальное число параметров для данной модели должно удовлетворять условию (9.18):

$$(k-m)^2 > (k+m). \tag{9.18}$$

Метод максимального правдоподобия предполагает, что оценки матрицы факторных нагрузок и дисперсионной матрицы характерных факторов должны удовлетворять системе уравнений (9.19):

$$\begin{cases} J\alpha = \Sigma^{-1}(\overline{A} - \Sigma)\alpha \\ \mathrm{diag}\overline{A} = \mathrm{diag}(\alpha\alpha^{T} + \Sigma) \end{cases} \tag{9.19}$$

Из (9.19) следует вывод, что оценки векторов факторных нагрузок $\alpha^{(j)}$ должны быть собственными векторами матрицы $(\overline{A} - \Sigma)\Sigma^{-1}$, а оценки диагональных элементов $\lambda_j$ матрицы J должны представлять собой соответствующие собственные значения данной матрицы. В качестве собственных векторов берутся m векторов, соответствующих наибольшим собственным значениям $\lambda_j$.

Условия нормировки для собственных векторов (9.20):

$$\alpha^{(j)^{T}}\Sigma^{-1}\alpha^{(j)} = \lambda_j. \tag{9.20}$$

При заданном числе факторов m решение системы можно осуществить с применением численной итерационной процедуры.

Так как данный метод предполагает оценки правдоподобия, то можно построить критерий отношения правдоподобия для проверки значимости полученной модели (9.21)

$$\eta=-\left(n-1-\frac{1}{6}(2k+5)-\frac{2}{3}m\right)\left(\ln\frac{|\overline{A}|}{|\hat{\alpha}\hat{\alpha}^{T}+\hat{\Sigma}|}-Sp(\overline{A}(\hat{\alpha}\hat{\alpha}^{T}+\hat{\Sigma})^{-1})+k\right) \quad (9.21)$$

При истинности $H_0$: допустимо представление исходных признаков в виде m-факторной модели, статистика $\eta$ асимптотически имеет распределение $\chi^2$ с числом степеней свободы и может быть определена по формуле 9.22:

$$\nu=\frac{1}{2}((k-m)^2-(k+m)) \quad (9.22)$$

Если нулевая гипотеза (допустимо представление исходных признаков в виде m- факторной модели) отвергается критерием, то следует перейти к рассмотрению модели с числом обобщенных факторов равным m+1 в том случае, если такое число параметров модели является допустимым.

## 9.5. Значимость факторных признаков

Зачастую определение факторов анализа не является заключительной целью исследования. Также требуется оценить вес фактора в формировании дисперсии каждого признака, а также опредеелить возможность интерпретации полученных результатов с точки зрения здравого смысла

Пусть имеется m- факторная модель. Пусть $f^{(p)}$ – некоторый фактор, $p\le m$. Общностью фактора $f^{(p)}$ (или вкладом фактора в суммарную дисперсию) признаков называется число $V^{(p)}=\sum_{j=1}^{k}\left[\alpha_j^{(p)}\right]^2$.

Величина $V_0=\sum_{p=1}^{m}V^{(p)}$ является суммарной общностью факторов $f^{(1)},...,f^{(m)}$. Отношение $V^{(p)}/V_0$ называется долей фактора $f^{(p)}$ в суммарной общности.

Пусть $\alpha_{s_1}^{(p)}, \alpha_{s_2}^{(p)}, \ldots, \alpha_{s_l}^{(p)}$ факторные нагрузки фактора $f^{(p)}$ соответствующие признакам $\xi_{(s_1)}, \ldots, \xi_{(s_l)}$. Тогда коэффициент информативности признаков $\xi_{(s_1)}, \ldots, \xi_{(s_l)}$ K может быть определен по формуле 9.23:

$$K_u = \frac{\sum_{j=1}^{l} \left(\alpha_{s_j}^{(p)}\right)^2}{\left\|\alpha^{(p)}\right\|^2} \tag{9.23}$$

Данное число позволяет определить, какой вклад в признаки $\xi_{(s_1)}, \ldots, \xi_{(s_l)}$ вносит в фактор $f^{(p)}$. Если можно выделить какую либо группу признаков, коэффициент информативности которой для фактора $f^{(p)}$ существенно выше коэффициента информативности оставшихся признаков, то первая группа признаков и будет определять содержание фактора.

Предположим, что рассматриваются следующие статистические характеристики деятельности фирмы.

$X_1$ – уровень выработки в год на одного работника;

$X_2$ – уровень фондоотдачи

$X_3$ – размер оборотных производственных средств,

$X_4$ – размер затрат на выпуск единицы продукции,

$X_5$ – численность персонала,

$X_6$ – рентабельность продукции,

$X_7$ – уровень энерговооруженности труда.

Предполагается, что на предыдущем этапе анализа были определены факторы $f^{(1)}, f^{(2)}$, а также найдены факторные нагрузки $\alpha^{(1)} = (0{,}9; 0{,}8; 0{,}1; 0{,}8; 0{,}3; 0{,}7; 0{,}2)$, $\alpha^{(2)} = (0{,}1; 0{,}4; 0{,}8; 0{,}3; 0{,}7; 0{,}2; 0{,}6)$.

В таком случае коэффициенты нормативности признаков составят:

$$K_{f^{(1)}}(\xi_1, \xi_2, \xi_4, \xi_6) = \frac{0{,}9^2 + 0{,}8^2 + 0{,}8^2 + 0{,}7^2}{0{,}9^2 + 0{,}8^2 + 0{,}1^2 + 0{,}8^2 + 0{,}3^2 + 0{,}7^2 + 0{,}2^2} \approx 0{,}948,$$

$$K_{f^{(2)}}(\xi_3, \xi_5, \xi_7) = \frac{0{,}8^2 + 0{,}7^2 + 0{,}6^2}{0{,}1^2 + 0{,}4^2 + 0{,}8^2 + 0{,}3^2 + 0{,}7^2 + 0{,}2^2 + 0{,}6^2} \approx 0{,}832.$$

Название факторов $f^{(1)}, f^{(2)}$ следует выбирать, опираясь на формирующие их признаки. Фактор $f^{(1)}$ условно можно назвать характеризующим эффективность производства, а фактор $f^{(2)}$ определяющим размеры производства.

Если факторные нагрузки обладают более или менее равномерным распределением, то задача интерпретации фактора усложняется. В этом случае целесообразно прибегать к вращению факторов.

Обобщенные факторы определяются с точностью до ортогонального преобразования. Следовательно, можно реализовывать ортогональное вращение факторного пространства, добиваясь такого расположения осей, при котором факторы допускают наиболее содержательную интерпретацию.

Как правило, используют две основные стратегии вращения. Формулы для их вычисления приведены в табл. 9.1.

*Табл. 9.1. Методы вращения в факторном анализе*

| Применяемых метод | Сущность подхода в рамках реализации метода | Формула |
|---|---|---|
| Метод варимакс | При вращении максимизируют величину, которая характеризует различие столбцов матрицы факторных нагрузок | $V=\sum_{j=1}^{m}\frac{1}{k}\sum_{i=1}^{k}\left(\left(\alpha_i^{(j)}\right)^2-\frac{1}{k}\sum_{s=1}^{k}\left(\alpha_s^{(j)}\right)^2\right)^2$ |
| Метод квартимакс | При вращении максимизируют величину, которая характеризует различие строк матрицы факторных нагрузок | $Q=\sum_{i=1}^{k}\frac{1}{m}\sum_{j=1}^{m}\left(\left(\alpha_i^{(j)}\right)^2-\frac{1}{m}\sum_{s=1}^{m}\left(\alpha_i^{(s)}\right)^2\right)^2$ |

Учитывая, что метод варимакс основан на упрощении столбцов, а квартимакс на упрощении строк матрицы факторных нагрузок, можно построить обобщенный метод, основанный на максимизации величины $\alpha Q+\beta V$. В частности при $\alpha=\beta$ получим критерий, используемый в методе вращения биквартимакс.

Если C представляет собой матрицу поворота, то в новой системе координат матрица факторных нагрузок определится как: $\alpha'=\alpha C$. Обычно в процедуре вращения осуществляют последовательное попарное ортогональное вращение осей. В этом случае полная матрица преобразования будет равна произведению матриц попарных поворотов. Например, для трехфакторной модели при вращении против часовой стрелки будут справедливы следующие формулы (9.24 и 9.25)

$$C=C_{12}*C_{23}*C_{31} \tag{9.24}$$

$$C=\begin{pmatrix}\cos\varphi_1 & -\sin\varphi_1 & 0\\ \sin\varphi_1 & \cos\varphi_1 & 0\\ 0 & 0 & 1\end{pmatrix}\cdot\begin{pmatrix}1 & 0 & 0\\ 0 & \cos\varphi_2 & -\sin\varphi_2\\ 0 & \sin\varphi_2 & \cos\varphi_2\end{pmatrix}\cdot\begin{pmatrix}\sin\varphi_3 & 0 & \cos\varphi_3\\ 0 & 1 & 0\\ \cos\varphi_3 & 0 & -\sin\varphi_3\end{pmatrix} \tag{9.25}$$

Для того, чтобы избежать влияния на результаты вращения переменных с большой общностью, осуществляют нормировку факторных нагрузок. Производят данную процедуру путем деления координат векторов факторных нагрузок на соответствующие корни из общностей [107].

После того как найдены оценки параметров факторной модели, как правило, необходимо вычислить оценки значений обобщенных факторов для каждого наблюдения. Эти значения могут быть использованы, например, в задачах классификации или в задачах регрессионного анализа.

Если для факторного анализа использовался метод главных компонент, то задача оценивания значений факторов решается путем выражения главных факторов через линейную комбинацию исходных признаков (9.26):

$$f^{(j)}=\beta^{(j)T}\xi=\beta_1^{(j)}\xi_1+\beta_2^{(j)}\xi_2+\ldots+\beta_k^{(j)}\xi_k,\quad j=\overline{1,m}, \tag{9.26}$$

где векторы $\beta^{(j)}$, $j=\overline{1,m}$ - собственные векторы матрицы ковариаций A, упорядоченные по убыванию собственных значений $\lambda_j$ и удовлетворяющие условию $|\beta^{(i)}|^2=1/\lambda_i$.

Пусть X матрица значений исходных признаков размеров $n \times k$, B - матрица, составленная из оценок векторов $\beta^{(j)}$, $j = \overline{1,m}$ размером $k \times m$, F- матрица оценок значений обобщенных факторов размеров $n \times m$, тогда (9.27):

$$F = XB \tag{9.27}$$

Если, кроме того, использовалась процедура вращения факторов с матрицей преобразования C, то справедлива формула 9.28:

$$F = XBC \tag{9.28}$$

Для оценивания значений факторов канонической модели факторного анализа используют либо метод Бартлетта, либо метод Томпсона.

Согласно методу Бартлетта модель для каждого наблюдения $X^{(i)}, i = \overline{1,n}$ рассматривается как регрессия величины $X^{(i)}$ на факторы $\alpha^{(j)}$, $j = \overline{1,m}$ с неизвестными коэффициентами $f_i^{(j)}$ (9.29):

$$X^{(i)} = f_i^{(1)}\alpha^{(1)} + f_i^{(2)}\alpha^{(2)} + \ldots + f_i^{(m)}\alpha^{(m)} + \varepsilon \tag{9.29}$$

Соответственно для значения $X_j^{(i)}$, которое, по сути, есть значение признака $\xi_j$ в i-ом наблюдении справедливо (9.30):

$$X_j^{(i)} = f_i^{(1)}\alpha_j^{(1)} + f_i^{(2)}\alpha_j^{(2)} + \ldots + f_i^{(m)}\alpha_j^{(m)} + \varepsilon_j, \quad j = \overline{1,k}, \tag{9.30}$$

Для оценки факторов следует применять обобщенный метод наименьших квадратов, согласно которому:

$$\hat{f}_i = (\hat{f}_i^{(1)}, \hat{f}_i^{(2)}, \ldots, \hat{f}_i^{(m)})^T = (\alpha^T \Sigma^{-1} \alpha)^{-1} \alpha^T \Sigma^{-1} X^{(i)} \tag{9.31}$$

В методе Томпсона предполагается, что оценки обобщенных факторов есть линейная комбинация исходных признаков, коэффициенты которой определяются из условия минимума среднеквадратичного отклонения оценки от фактора, то есть (9.32):

$$\hat{f}^{(j)} = \beta^{(j)T}\xi = \xi^T\beta^{(j)} = \beta_1^{(j)}\xi_1 + \beta_2^{(j)}\xi_2 + \ldots + \beta_k^{(j)}\xi_k, \quad j = \overline{1,m}, \tag{9.32}$$

где вектор коэффициентов удовлетворяет условию (9.33):

$$M(f^{(j)} - \hat{f}^{(j)})^2 \to \min \tag{9.33}$$

Таким образом, коэффициенты являются коэффициентами линейной среднеквадратичной регрессии и могут быть выражены через ковариации величин $\xi_1, \xi_2, \ldots, \xi_k$ и $f^{(j)}$ (9.34)

$$\beta^{(j)} = A^{-1}(\xi) M(f^{(j)} \cdot \xi) \tag{9.34}$$

здесь: $A$ - матрица ковариаций величин $\xi_1, \xi_2, \ldots, \xi_k$, $M(f^{(j)} \cdot \xi)$ - вектор ковариаций величины $f^{(j)}$ и величин $\xi_1, \xi_2, \ldots, \xi_k$.
Соответственно (9.35)

$$\hat{f}^{(j)} = \xi^T A^{-1}(\xi) M(f^{(j)} \cdot \xi) \tag{9.35}$$

Учитывая, что для канонической модели факторного анализа $A(\xi) = \alpha\alpha^T + \Sigma$ и $M(f^{(j)} \cdot \xi) = \alpha^{(j)}$ получим следующую оценку для $f^{(j)}$ (9.36):

$$\hat{f}^{(j)} = \xi^T (\alpha\alpha^T + \Sigma)^{-1} \alpha^{(j)}, \quad j = \overline{1, m} \tag{9.36}$$

Соответственно оценка для вектора обобщенных факторов:

$$\hat{f}^T = \xi^T (\alpha\alpha^T + \Sigma)^{-1} \alpha \tag{9.37}$$

Рассмотренные в рамках данного параграфа положения отражают основные положения факторных оценок и могут использоваться для оценки значимости результатов факторного анализа

## 9.6. Особенности решения задачи факторного анализа в программе Statitica

Рассмотрим основные шаги реализации факторного анализа с использованием программы в программе Statistica.

Начальным этапом кластеризации является ввод необходимых переменных в окне при открытии программы (рис. 9.6)

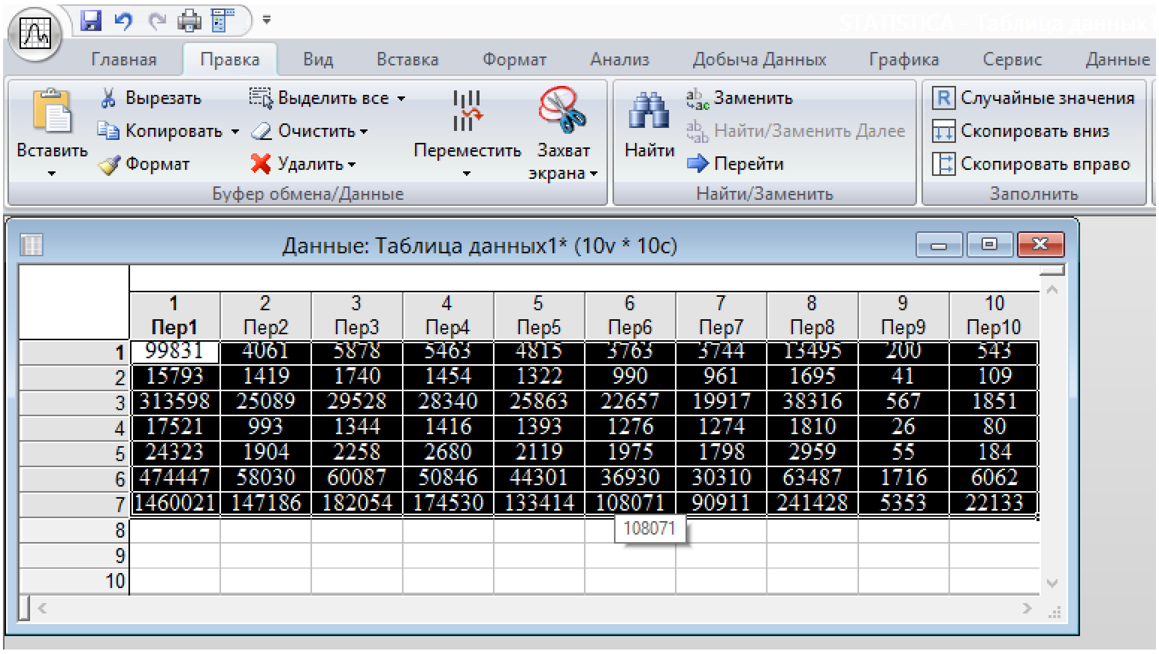


*Рис. 9.6. Ввод исходных данных*

Для начала осуществления процедуры факторного анализа методом главных компонент во вкладке «Анализ» главной панели был выбран «Многомерный анализ», а затем «Факторный анализ» (рис. 9.7).

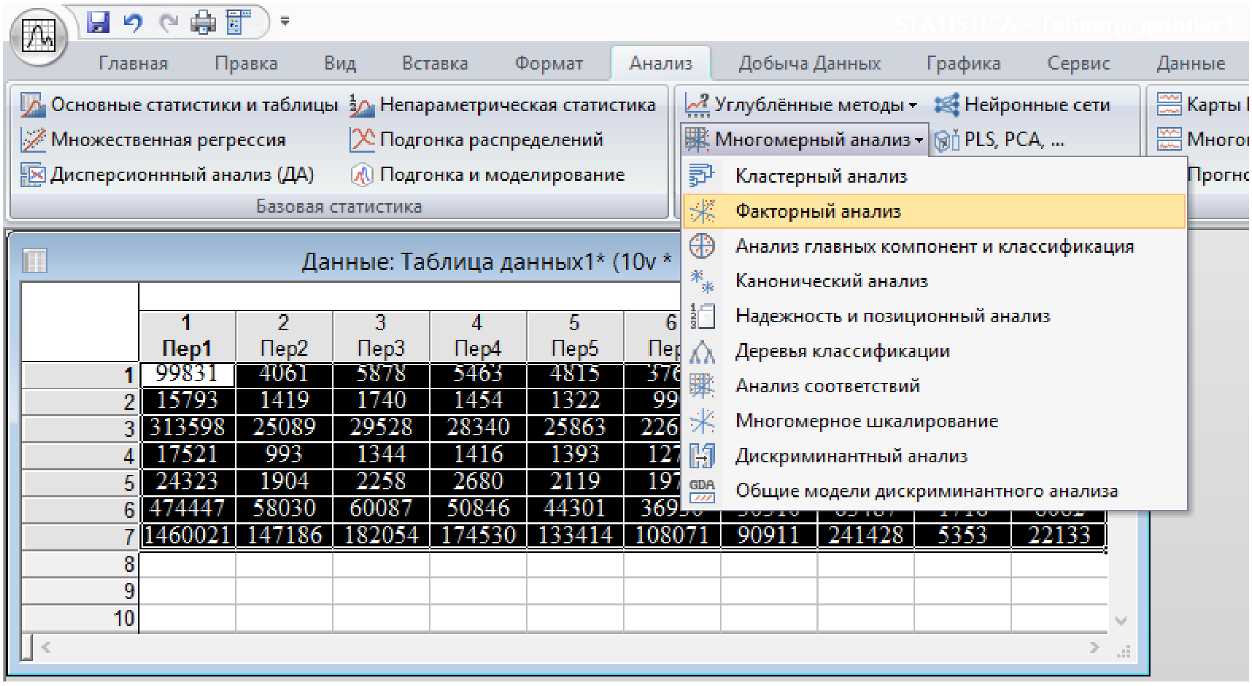



*Рис. 9.7. Выбор метода факторного анализа*

На следующем шаге осуществляется выбор переменных для анализа (рис.9.8):

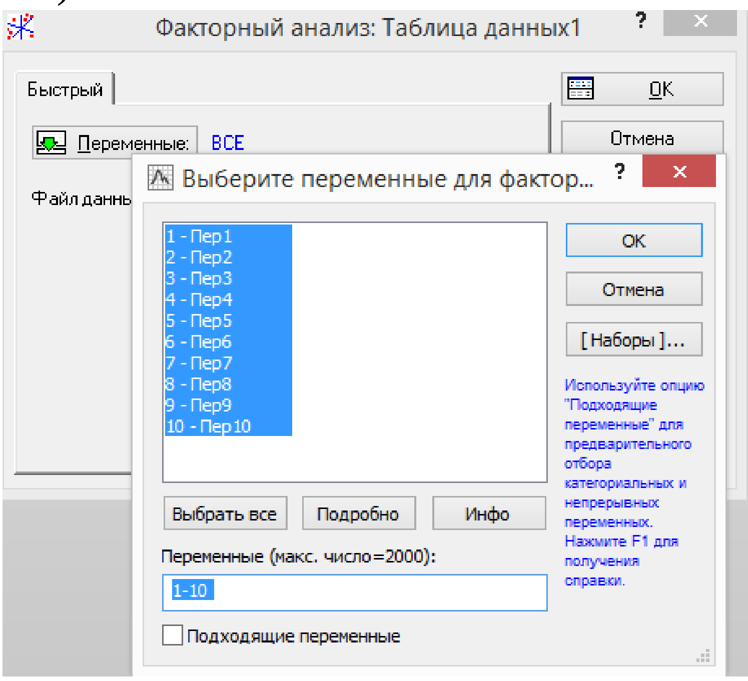



*Рис. 9.8. Окно выбора переменных факторного анализа*

Также существует возможность выбора информации о том, какими данными является массив: исходными характеирстиками или корреляционной матрицей (рис. 9.9):

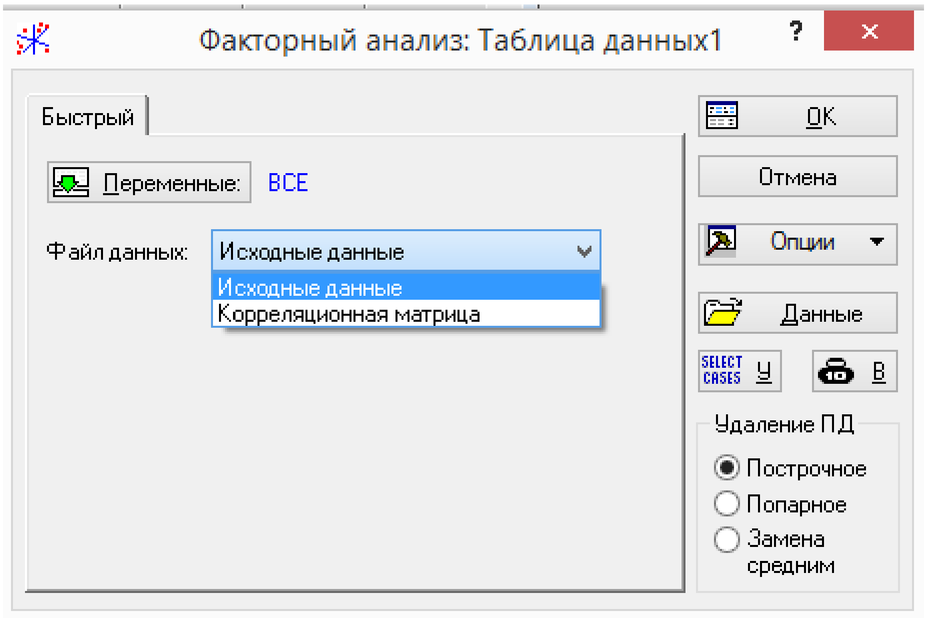



*Рис. 9.9. Выбор информации об исходных данных*

На следующем этапе факторного моделирования открывается окно, в котором необходимо задать параметры осуществления программой расчетов (рис. 9.10):

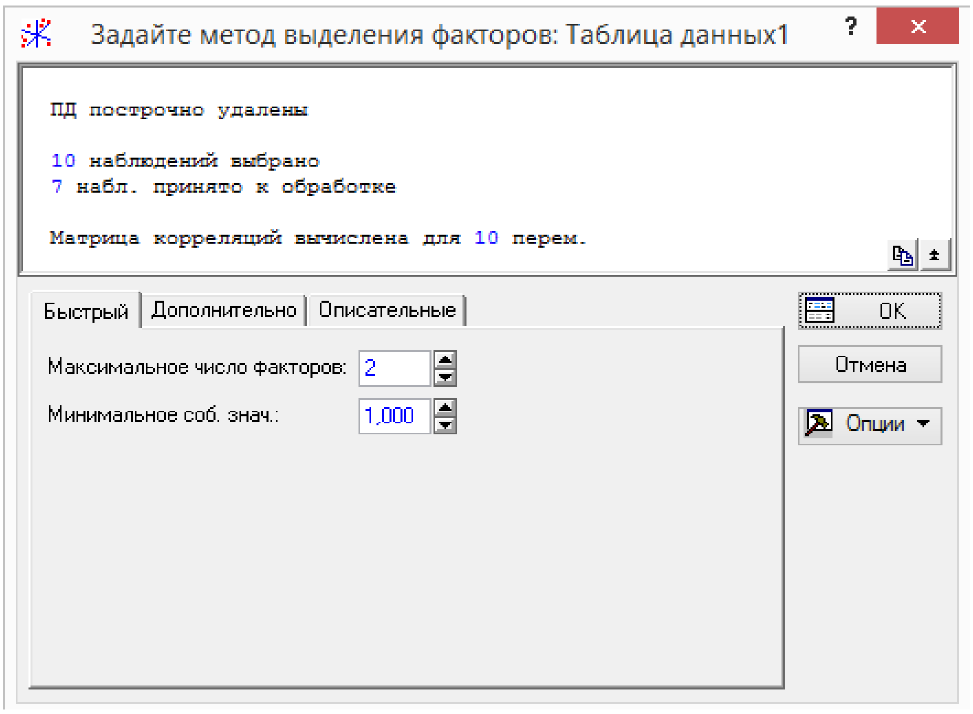



*Рис. 9.10. Окно определения параметров факторного моделирования*

На вкладке «Дополнительно» указывается метод выделения факторов. По умолчанию применяетмя метод главных компонент (рис. 9.11):

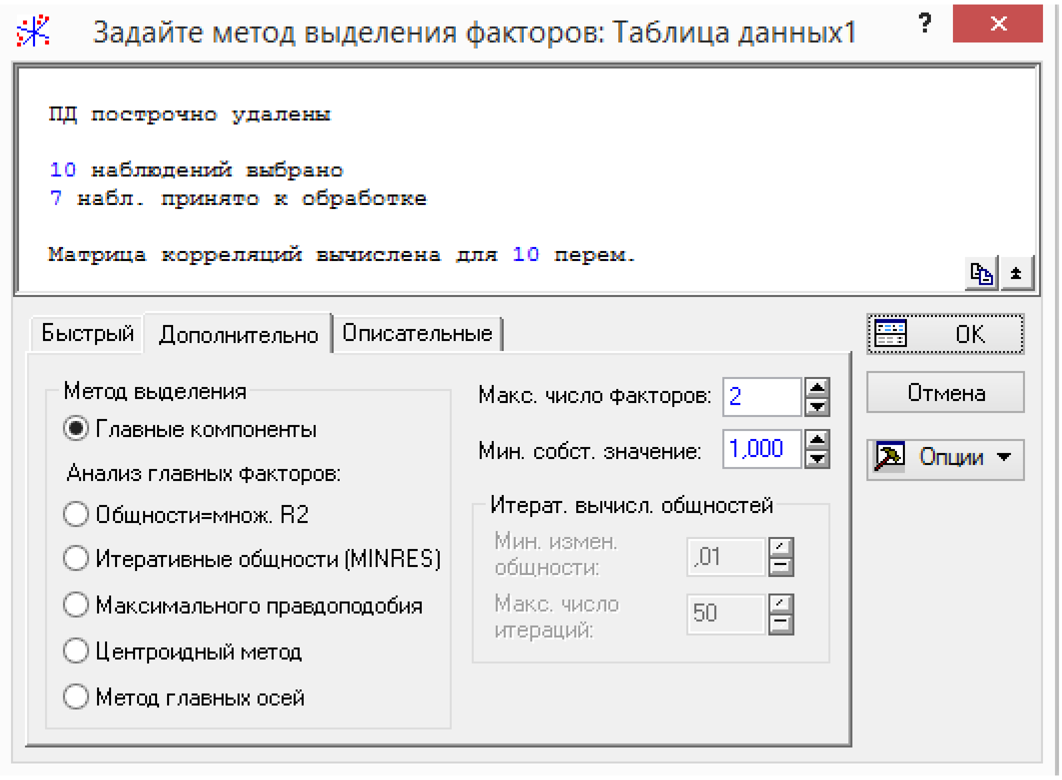



*Рис. 9.11. Выбор метода выделения факторов анализа*

Вкладка «Описательные» используется для просмотра характеристик корреляции, средних и стандартных отклонений, а также для построения множественной регрессии (рис. 9.12)

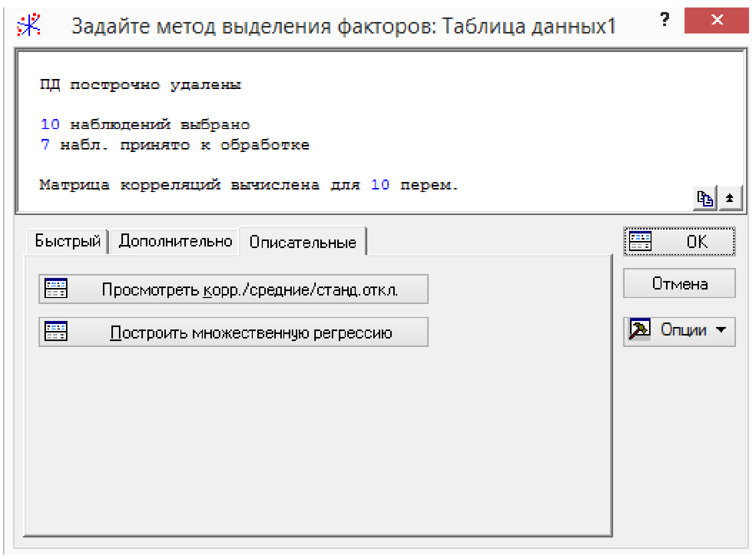



*Рис. 9.12. Выбор метода выделения факторов анализа*

Окно просмотра описательных статистик представлено на рисунке 9.13:

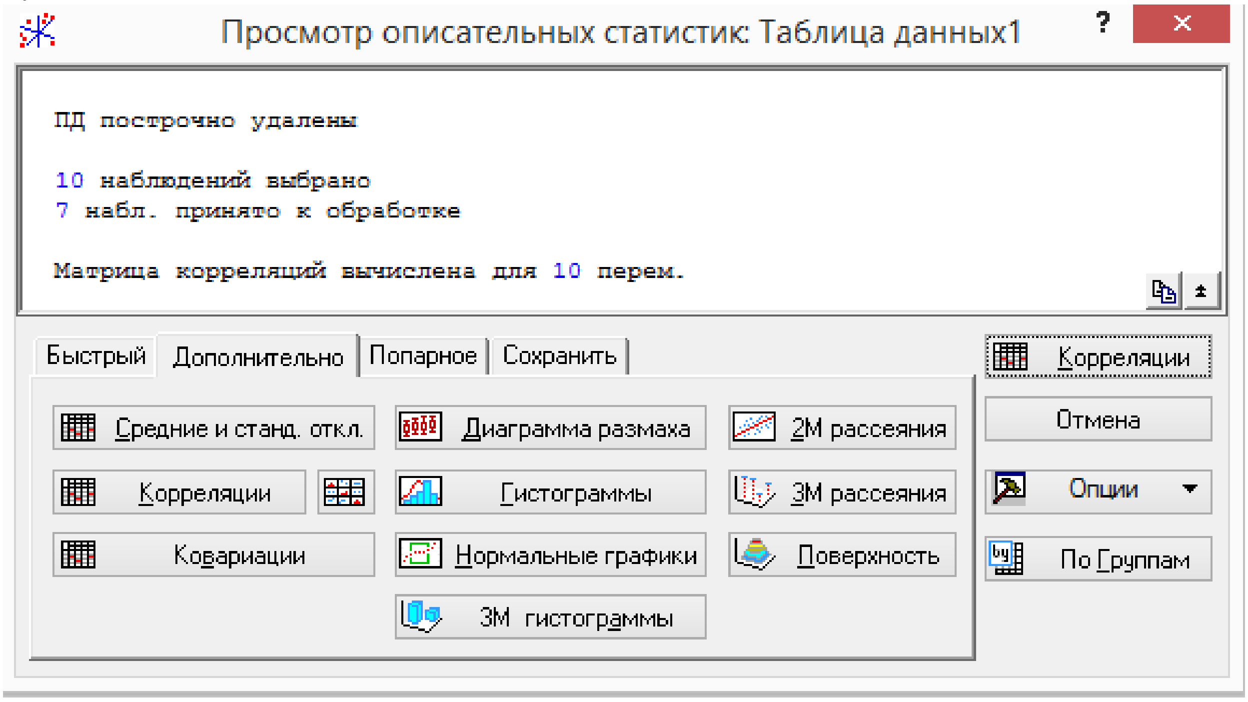



*Рис. 9.13. Описательные статистики*

Результаты вычислений средний и стандартных отклонений (рис. 9.14):

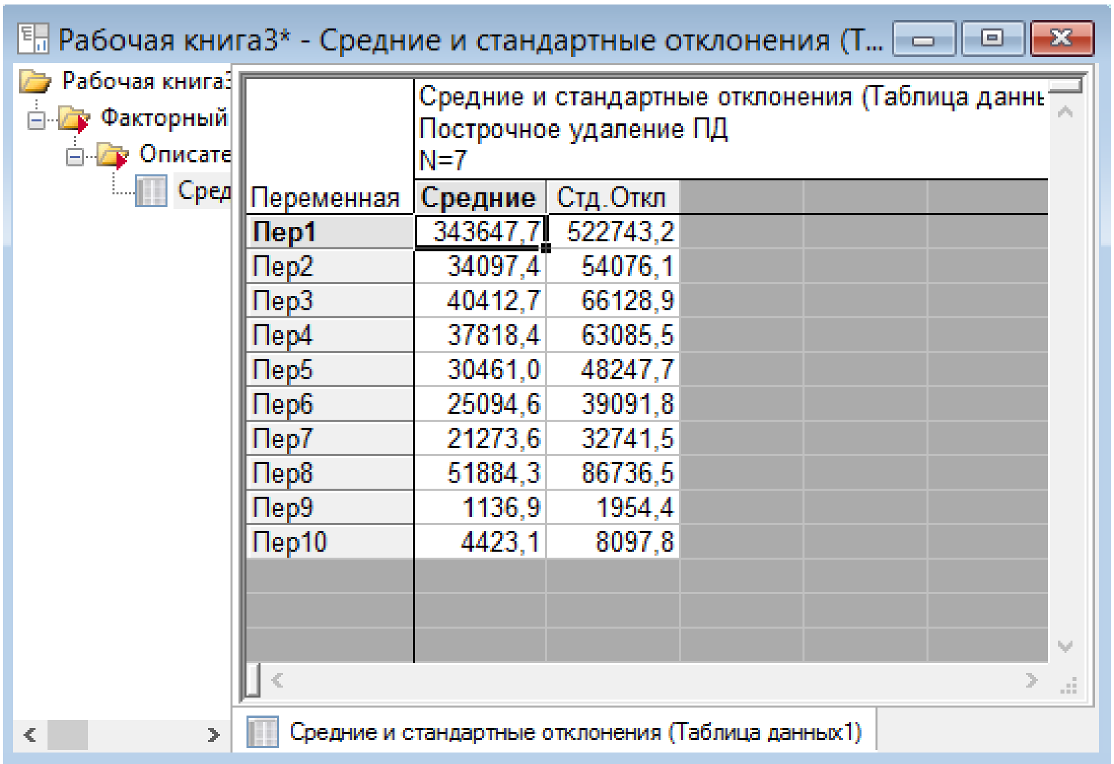



*Рис. 9.14. Окно вывода результатов средних и стандартных отклонений*

Данные по корреляции (рис. 9.15):

| Переменная | Корреляции (Таблица данных1)<br>Построчное удаление ПД<br>N=7<br>Пер1 | Пер2 | Пер3 | Пер4 | Пер5 | Пер6 | Пер7 | Пер8 | Пер9 | Пер10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Пер1 | 1,00 | 1,00 | 1,00 | 1,00 | 1,00 | 1,00 | 1,00 | 1,00 | 0,99 | 0,99 |
| Пер2 | 1,00 | 1,00 | 1,00 | 0,99 | 1,00 | 1,00 | 1,00 | 0,99 | 1,00 | 0,99 |
| Пер3 | 1,00 | 1,00 | 1,00 | 1,00 | 1,00 | 1,00 | 1,00 | 1,00 | 1,00 | 1,00 |
| Пер4 | 1,00 | 0,99 | 1,00 | 1,00 | 1,00 | 1,00 | 1,00 | 1,00 | 1,00 | 1,00 |
| Пер5 | 1,00 | 1,00 | 1,00 | 1,00 | 1,00 | 1,00 | 1,00 | 1,00 | 1,00 | 0,99 |
| Пер6 | 1,00 | 1,00 | 1,00 | 1,00 | 1,00 | 1,00 | 1,00 | 1,00 | 0,99 | 0,99 |
| Пер7 | 1,00 | 1,00 | 1,00 | 1,00 | 1,00 | 1,00 | 1,00 | 1,00 | 0,99 | 0,99 |
| Пер8 | 1,00 | 0,99 | 1,00 | 1,00 | 1,00 | 1,00 | 1,00 | 1,00 | 1,00 | 1,00 |
| Пер9 | 0,99 | 1,00 | 1,00 | 1,00 | 1,00 | 0,99 | 0,99 | 1,00 | 1,00 | 1,00 |
| Пер10 | 0,99 | 0,99 | 1,00 | 1,00 | 0,99 | 0,99 | 0,99 | 1,00 | 1,00 | 1,00 |

*Рис. 9.15. Окно вывода результатов корреляции*

Результаты вычисления ковариации для n = 7 (рис. 9.16):

| Переменная | Ковариации (Таблица данных1)<br>Построчное удаление ПД<br>N=7<br>Пер1 | Пер2 | Пер3 | Пер4 | Пер5 | Пер6 | Пер7 | Пер8 | Пер9 | Пер10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Пер1 | 2,732605E+11 | 2,812728E+10 | 3,450821E+10 | 3,291890E+10 | 2,520314E+10 | 2,041805E+10 | 1,710651E+10 | 4,521587E+10 | 1,015852E+09 | 4,198833E+09 |
| Пер2 | 2,812728E+10 | 2,924222E+09 | 3,568277E+09 | 3,392361E+09 | 2,602115E+09 | 2,108353E+09 | 1,763842E+09 | 4,643513E+09 | 1,052242E+08 | 4,338251E+08 |
| Пер3 | 3,450821E+10 | 3,568277E+09 | 4,373030E+09 | 4,168345E+09 | 3,189055E+09 | 2,582304E+09 | 2,161908E+09 | 5,718827E+09 | 1,290314E+08 | 5,334918E+08 |
| Пер4 | 3,291890E+10 | 3,392361E+09 | 4,168345E+09 | 3,979784E+09 | 3,040367E+09 | 2,461218E+09 | 2,061505E+09 | 5,466962E+09 | 1,229928E+08 | 5,093926E+08 |
| Пер5 | 2,520314E+10 | 2,602115E+09 | 3,189055E+09 | 3,040367E+09 | 2,327837E+09 | 1,885791E+09 | 1,579243E+09 | 4,171043E+09 | 9,393522E+07 | 3,882339E+08 |
| Пер6 | 2,041805E+10 | 2,108353E+09 | 2,582304E+09 | 2,461218E+09 | 1,885791E+09 | 1,528169E+09 | 1,279785E+09 | 3,375097E+09 | 7,599248E+07 | 3,138970E+08 |
| Пер7 | 1,710651E+10 | 1,763842E+09 | 2,161908E+09 | 2,061505E+09 | 1,579243E+09 | 1,279785E+09 | 1,072006E+09 | 2,828221E+09 | 6,359610E+07 | 2,627728E+08 |
| Пер8 | 4,521587E+10 | 4,643513E+09 | 5,718827E+09 | 5,466962E+09 | 4,171043E+09 | 3,375097E+09 | 2,828221E+09 | 7,523227E+09 | 1,688527E+08 | 7,001779E+08 |
| Пер9 | 1,015852E+09 | 1,052242E+08 | 1,290314E+08 | 1,229928E+08 | 9,393522E+07 | 7,599248E+07 | 6,359610E+07 | 1,688527E+08 | 3,819838E+06 | 1,580929E+07 |
| Пер10 | 4,198833E+09 | 4,338251E+08 | 5,334918E+08 | 5,093926E+08 | 3,882339E+08 | 3,138970E+08 | 2,627728E+08 | 7,001779E+08 | 1,580929E+07 | 6,557356E+07 |

*Рис. 9.16. Окно вывода результатов ковариации*

Построение диаграммы размаха предполагает при выборе данной операции выбор переменных. Окно выбора представлено на рис. 9.17:

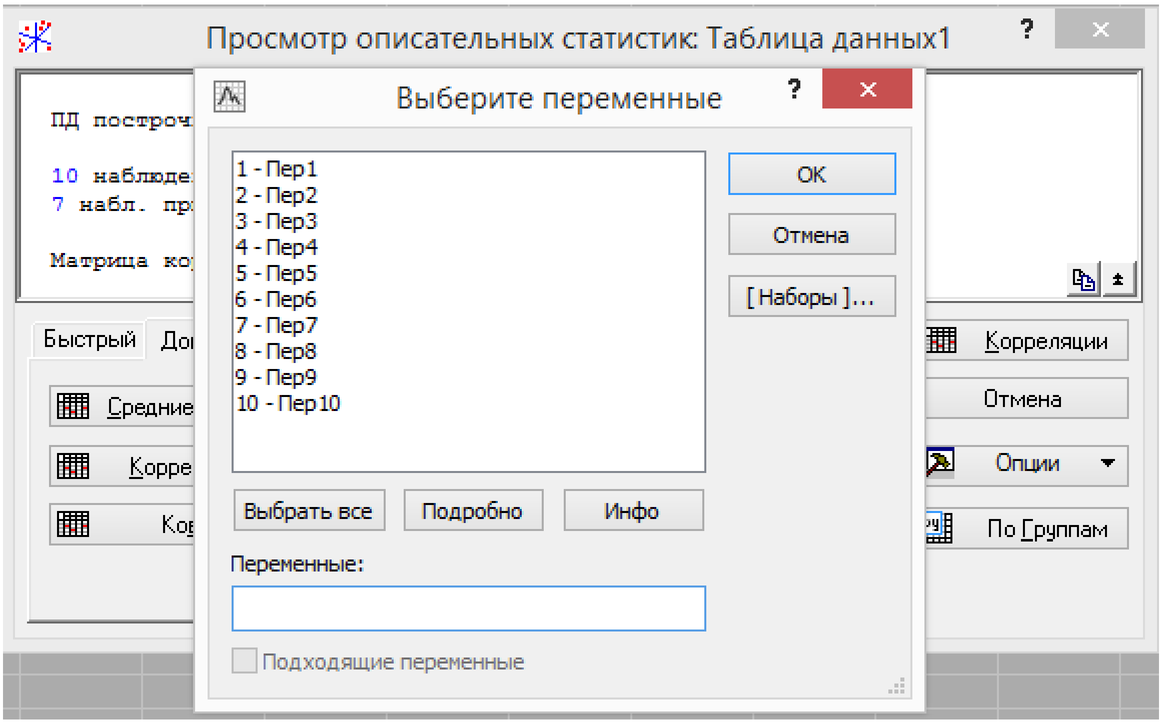



*Рис. 9.17. Выбор переменных для построения диаграммы размаха*

Для построении диаграммы размаха могут быть выбраны следующие варианты (рис. 9.18):

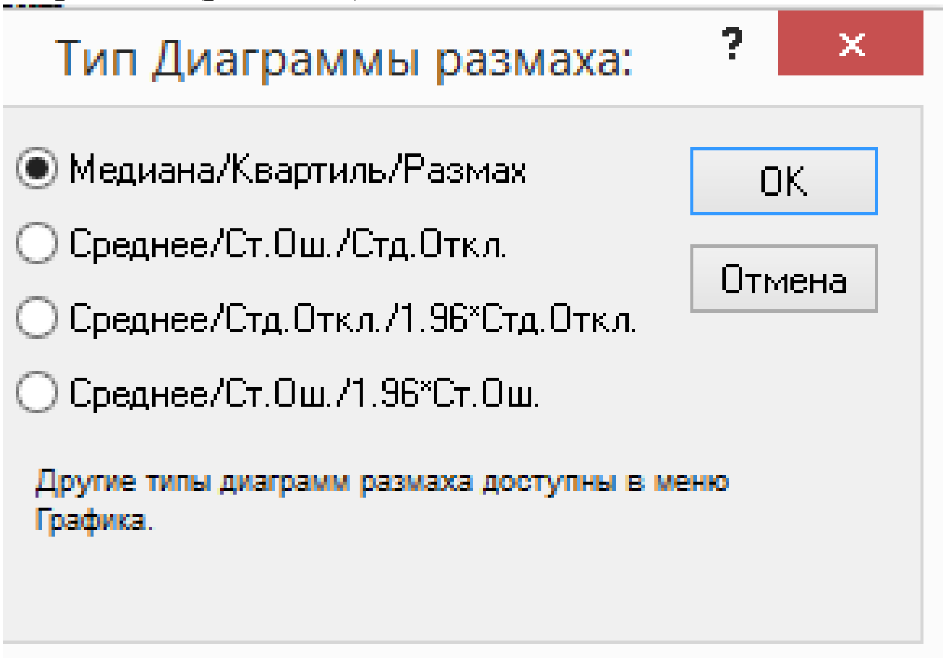



*Рис. 9.18. Выбор типа диаграммы размаха*

При выборе первого варианта (Медиана/Квартиль/Размах) диаграмма примет следующий вид (рис. 9.19):

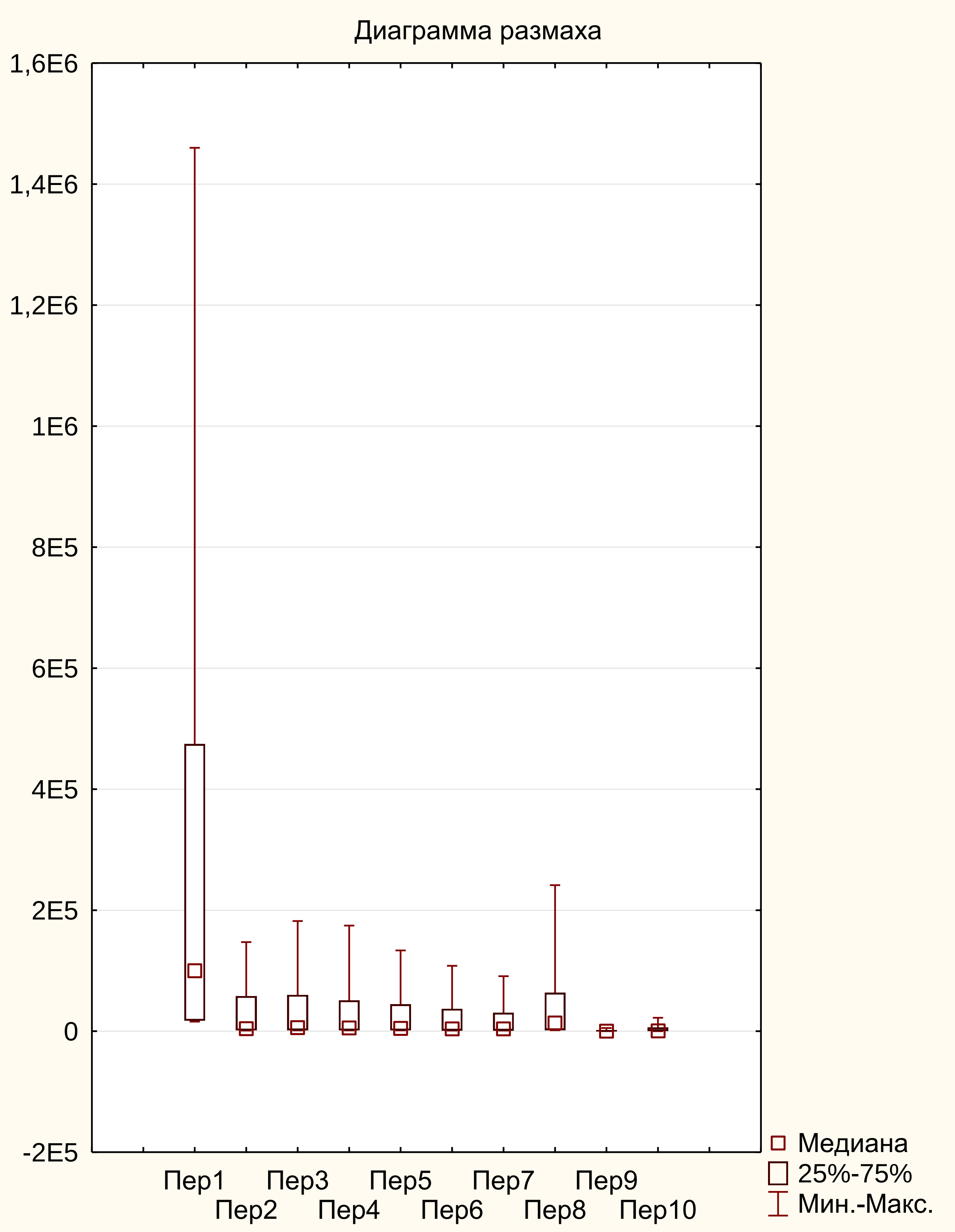



*Рис. 9.19. Диаграмма размаха при выборе первого варианта*

При выборе второго варианта (рис. 9.20):

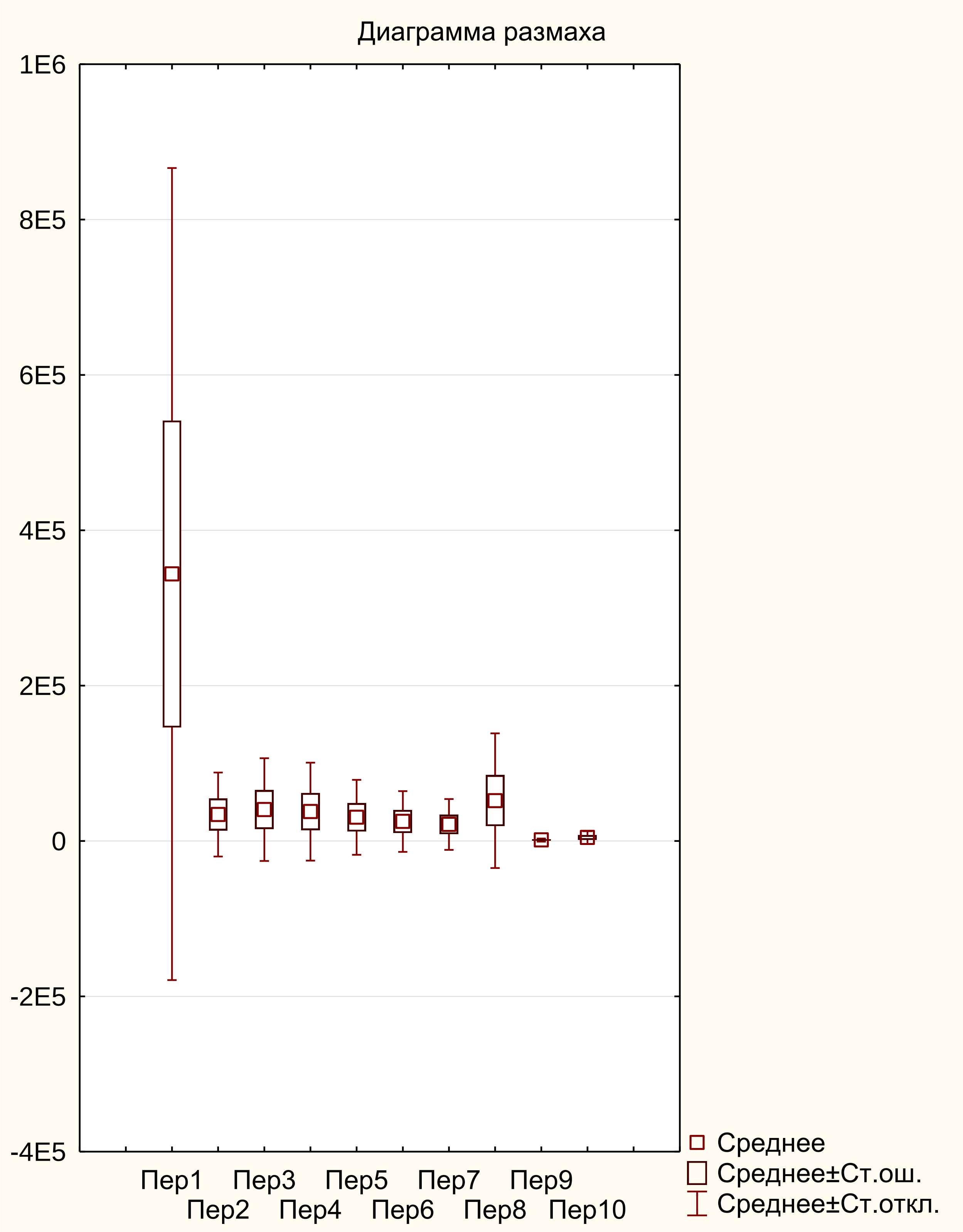



*Рис. 9.20. Диаграмма размаха при выборе второго варианта*

При выборе третьего варианта (рис. 9.21):

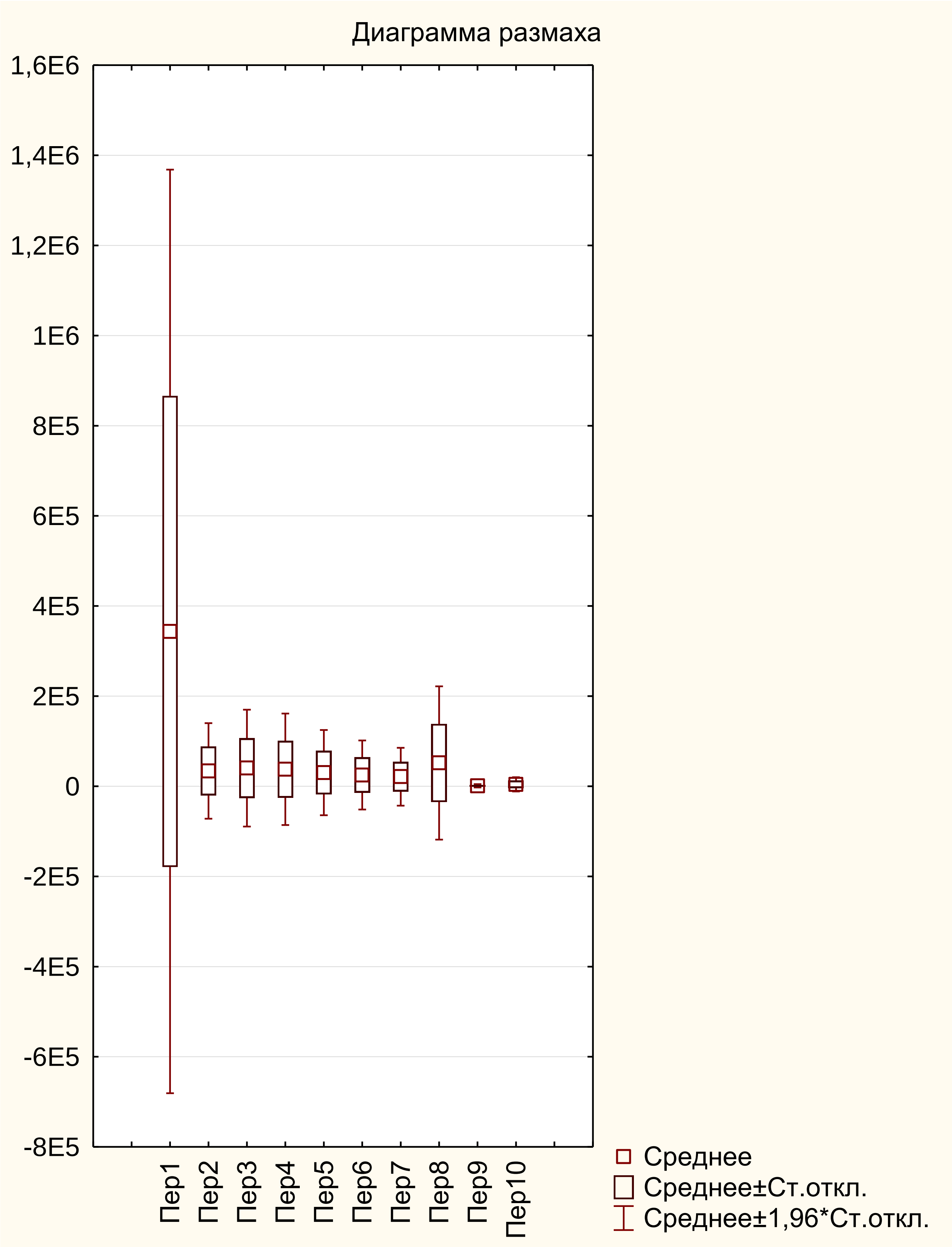



*Рис. 9.21. Диаграмма размаха при выборе третьего варианта*

При выборе четвертого варианта (рис. 9.22):

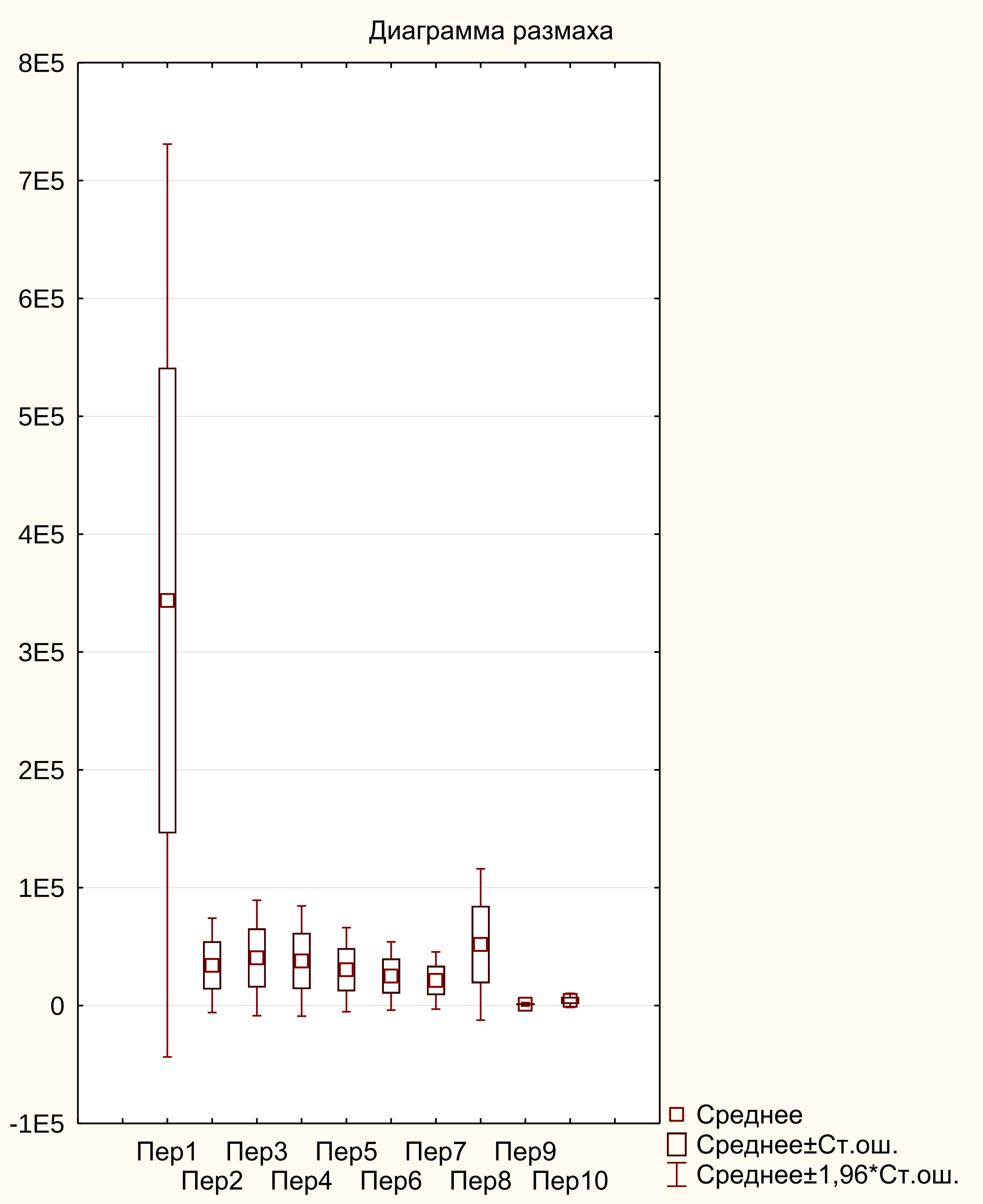



*Рис. 9.22. Диаграмма размаха при выборе четвертого варианта*

Гистограммы по каждой из десяти рассматриваемых переменных примут вид (рис. 9.23-9.33)

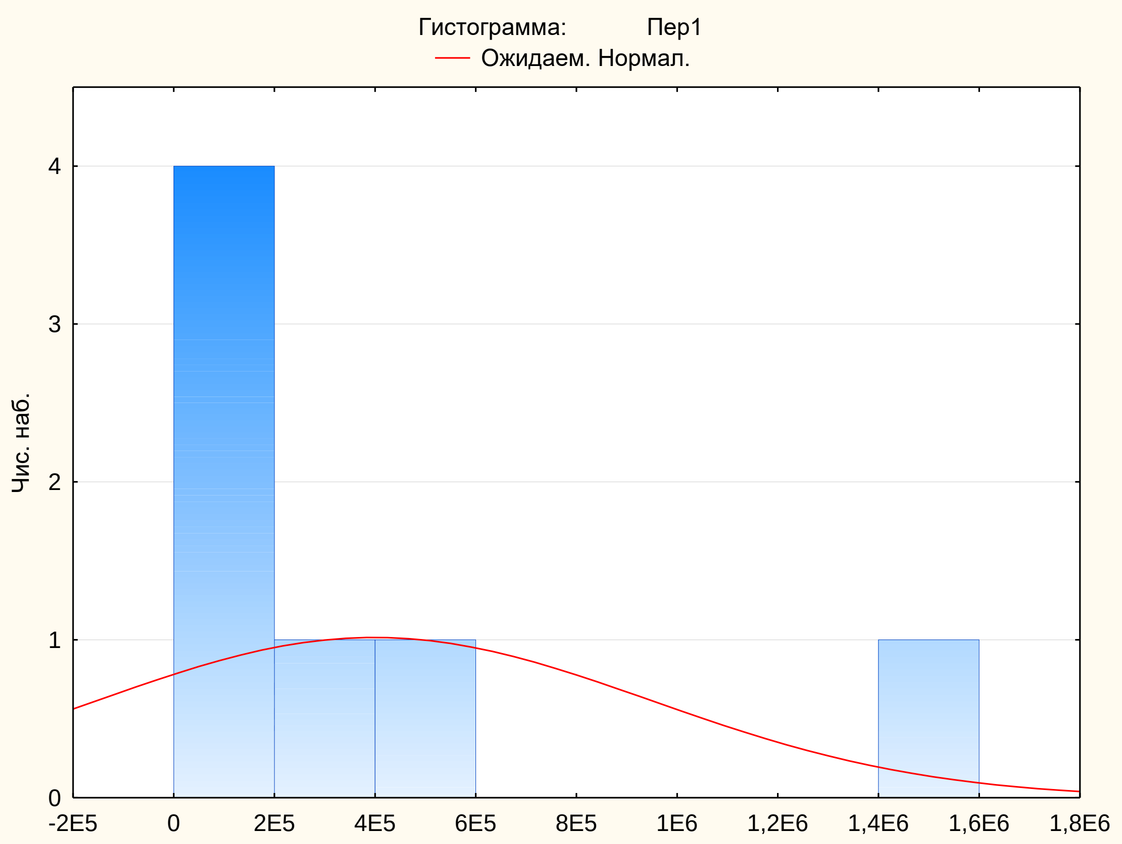



*Рис. 9.23. Результаты построения гистограммы по переменной 1*

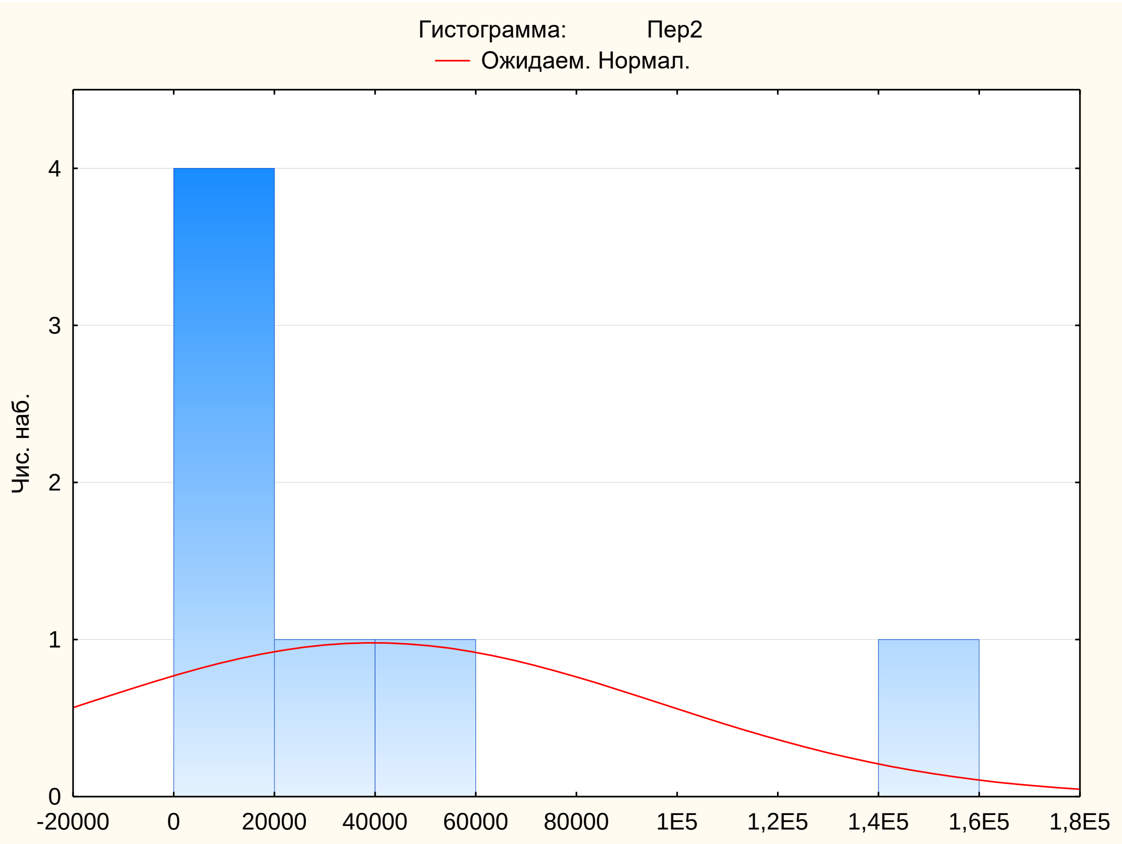



*Рис. 9.24. Результаты построения гистограммы по переменной 2*

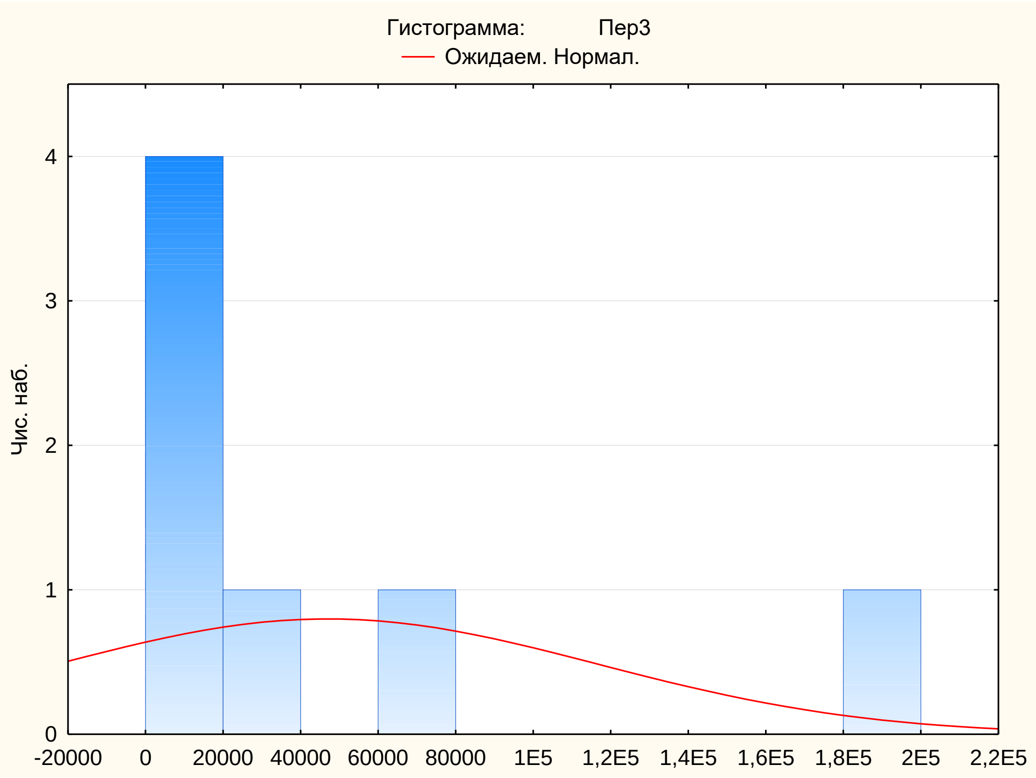



*Рис. 9.25. Результаты построения гистограммы по переменной 3*

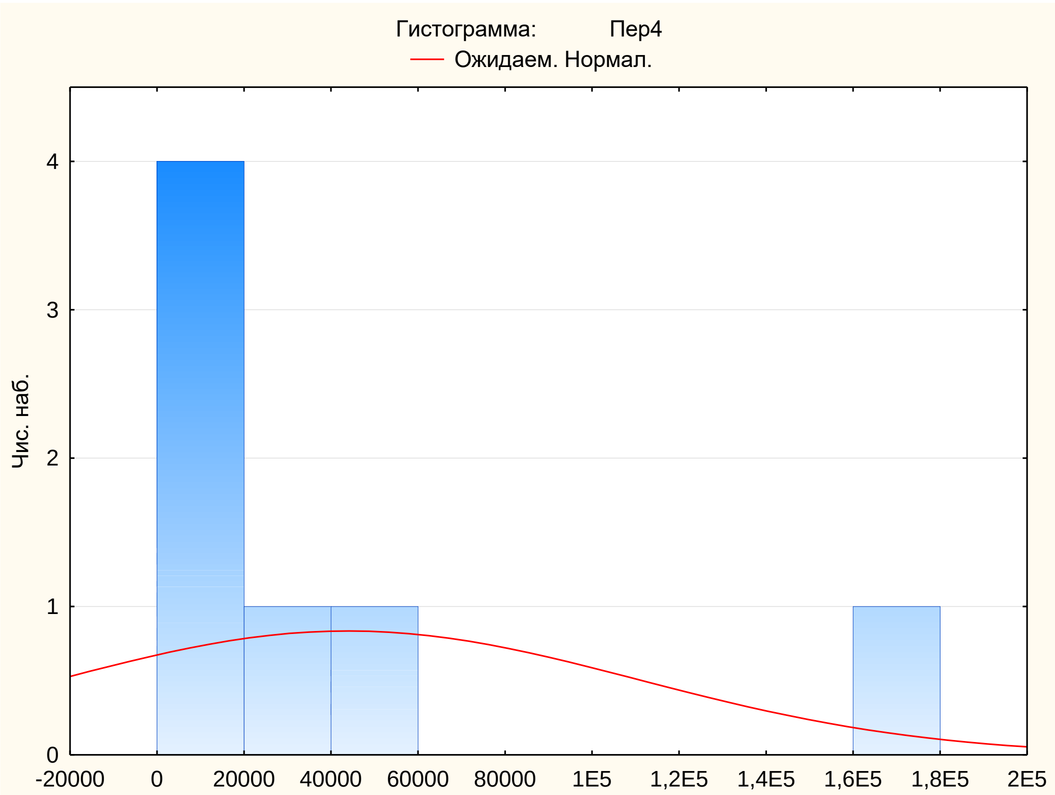



*Рис. 9.26. Результаты построения гистограммы по переменной 4*

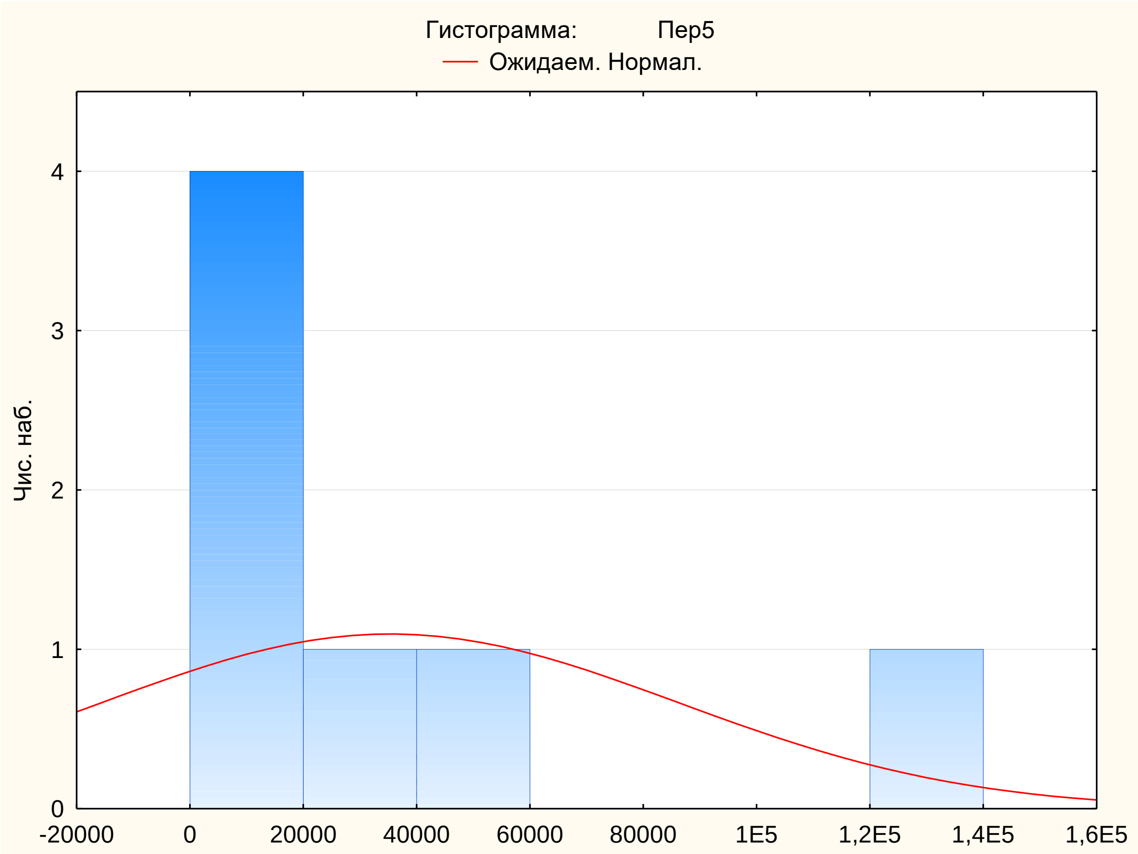



*Рис. 9.27. Результаты построения гистограммы по переменной 5*

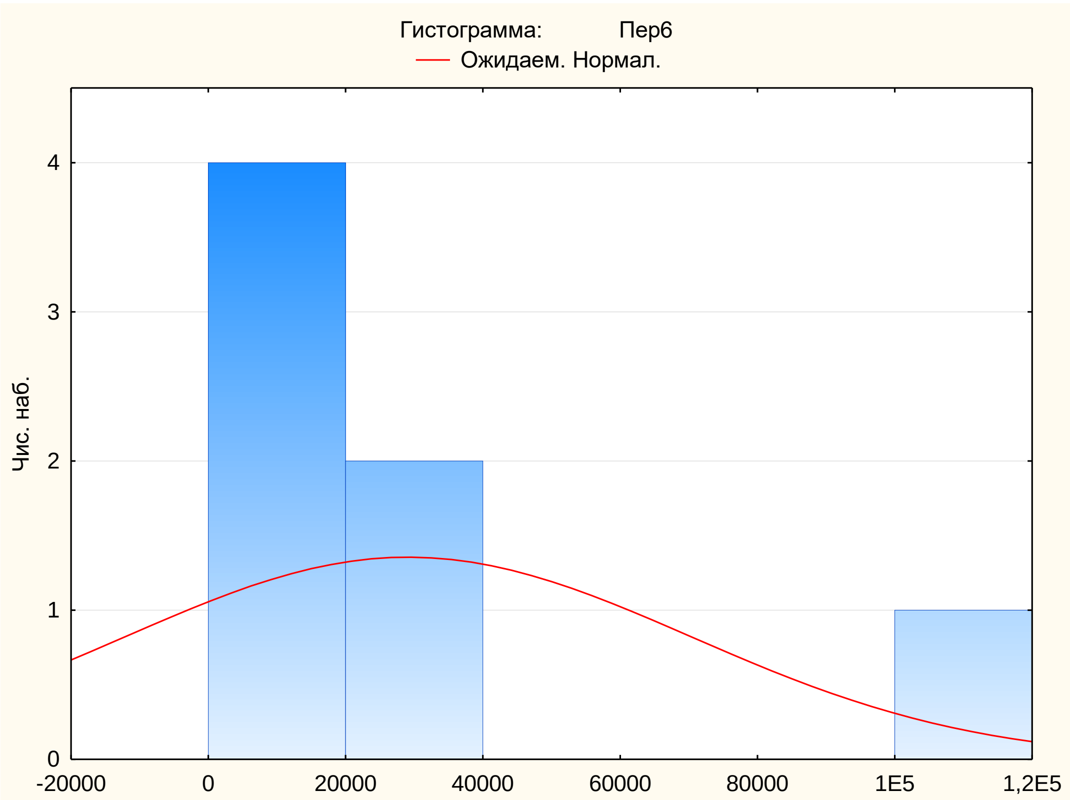



*Рис. 9.28. Результаты построения гистограммы по переменной 6*

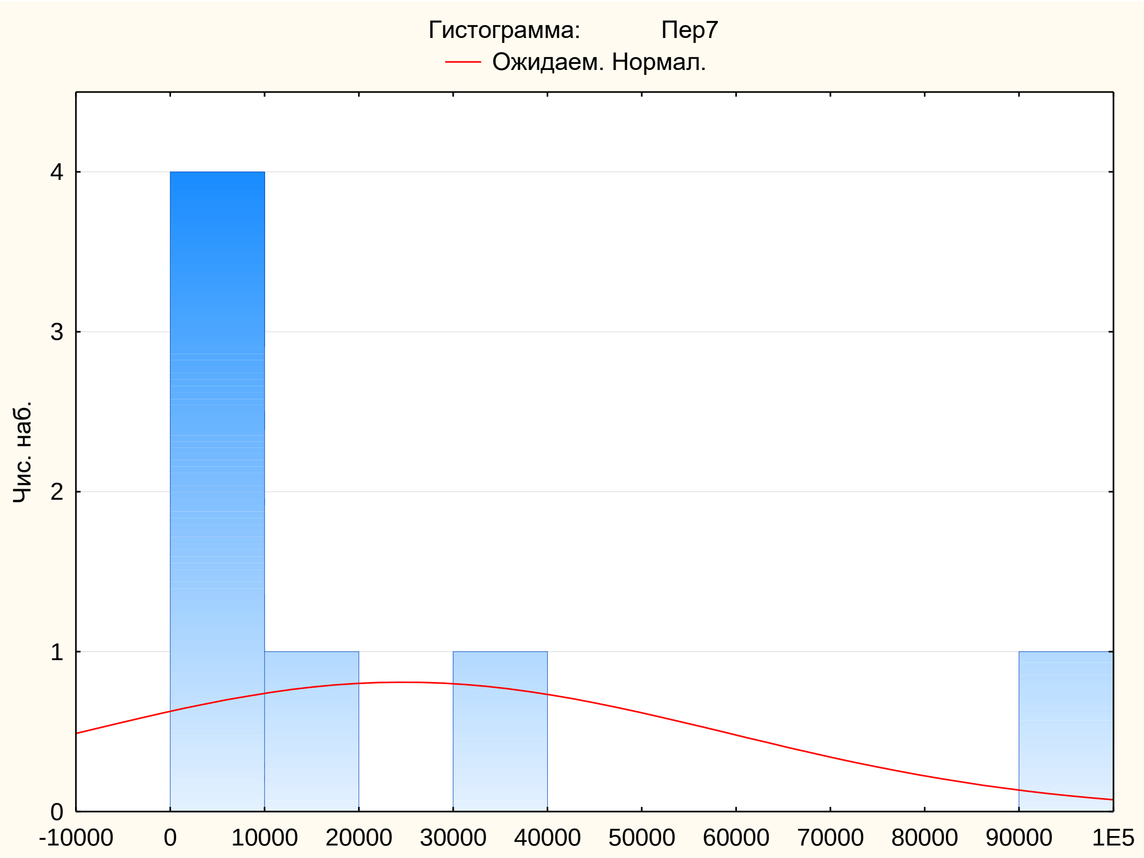



*Рис. 9.29. Результаты построения гистограммы по переменной 7*

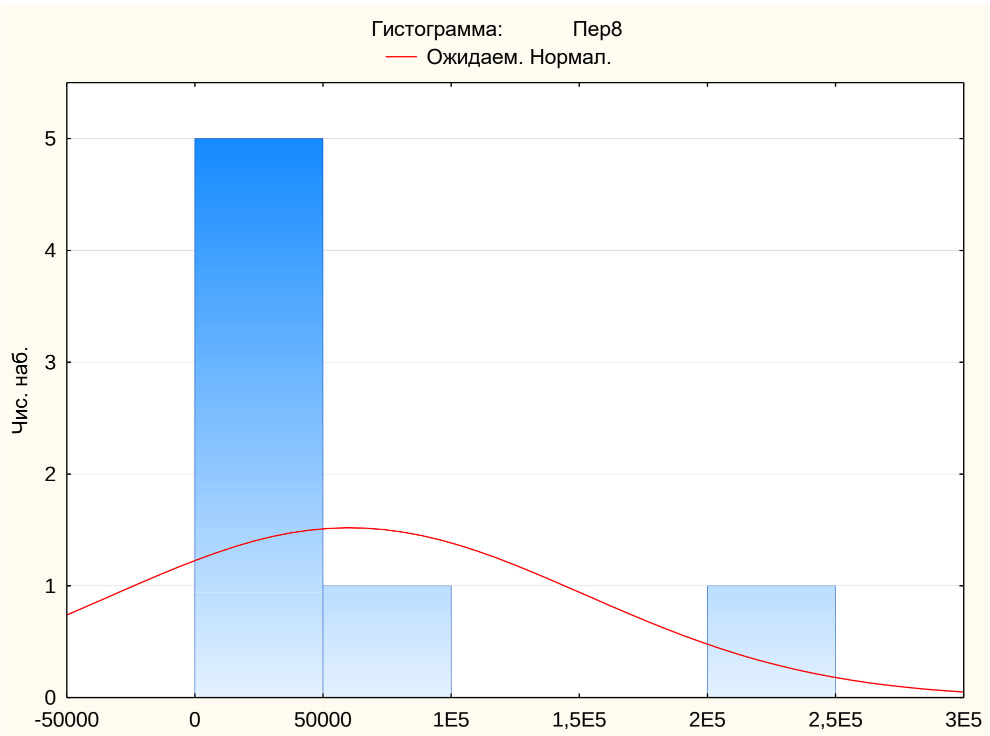



*Рис. 9.30. Результаты построения гистограммы по переменной 8*

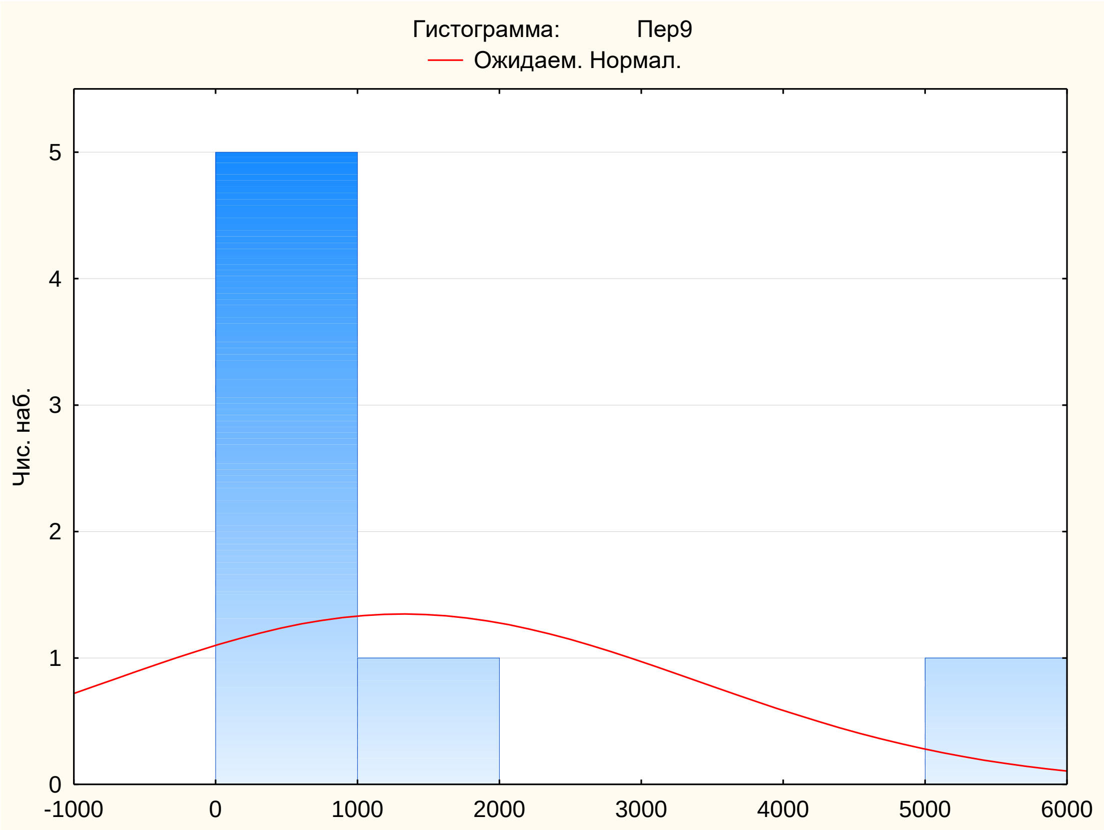



*Рис. 9.31. Результаты построения гистограммы по переменной 9*

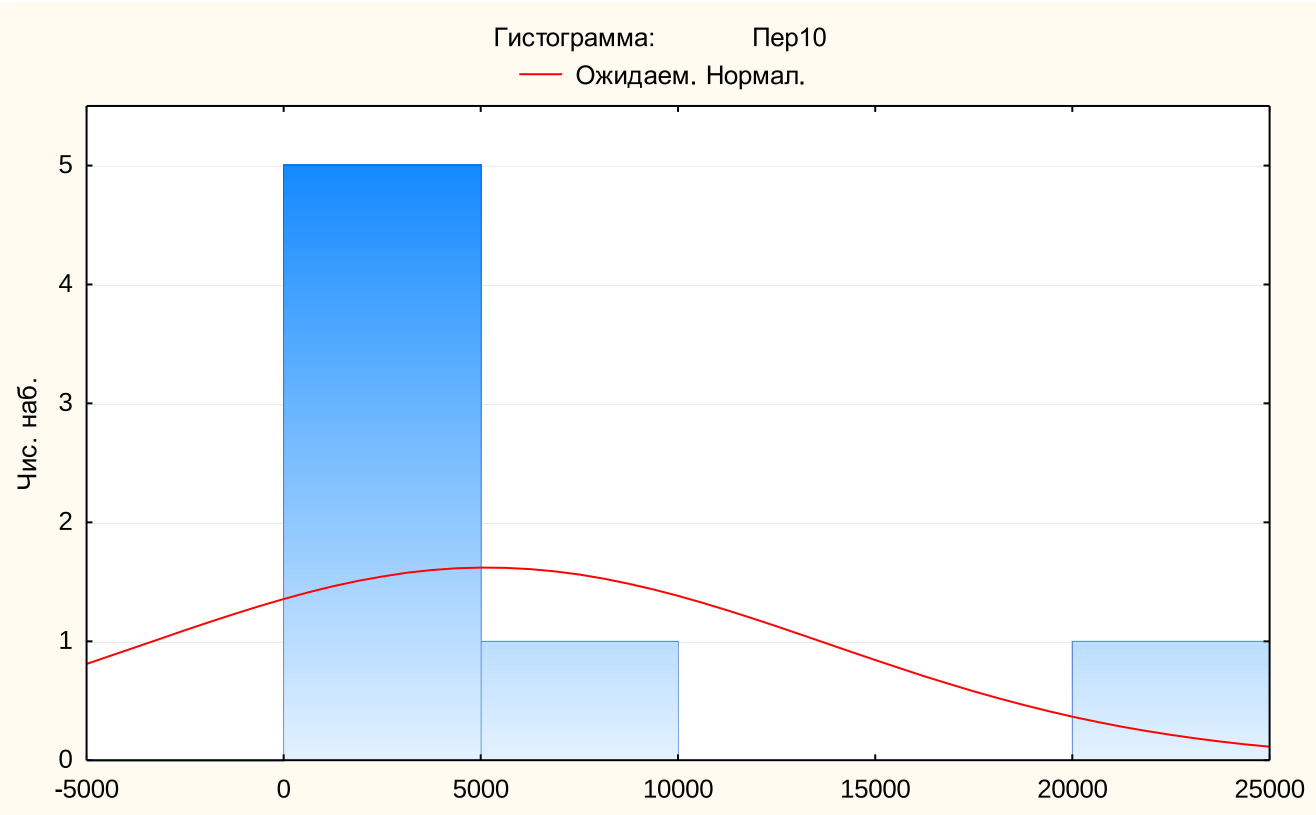



*Рис. 9.32. Результаты построения гистограммы по переменной 10*

Нормальные графики также строятся по каждой из выбранных переменных. Пример результата построения для переменной 1 представлен на рис.9.33

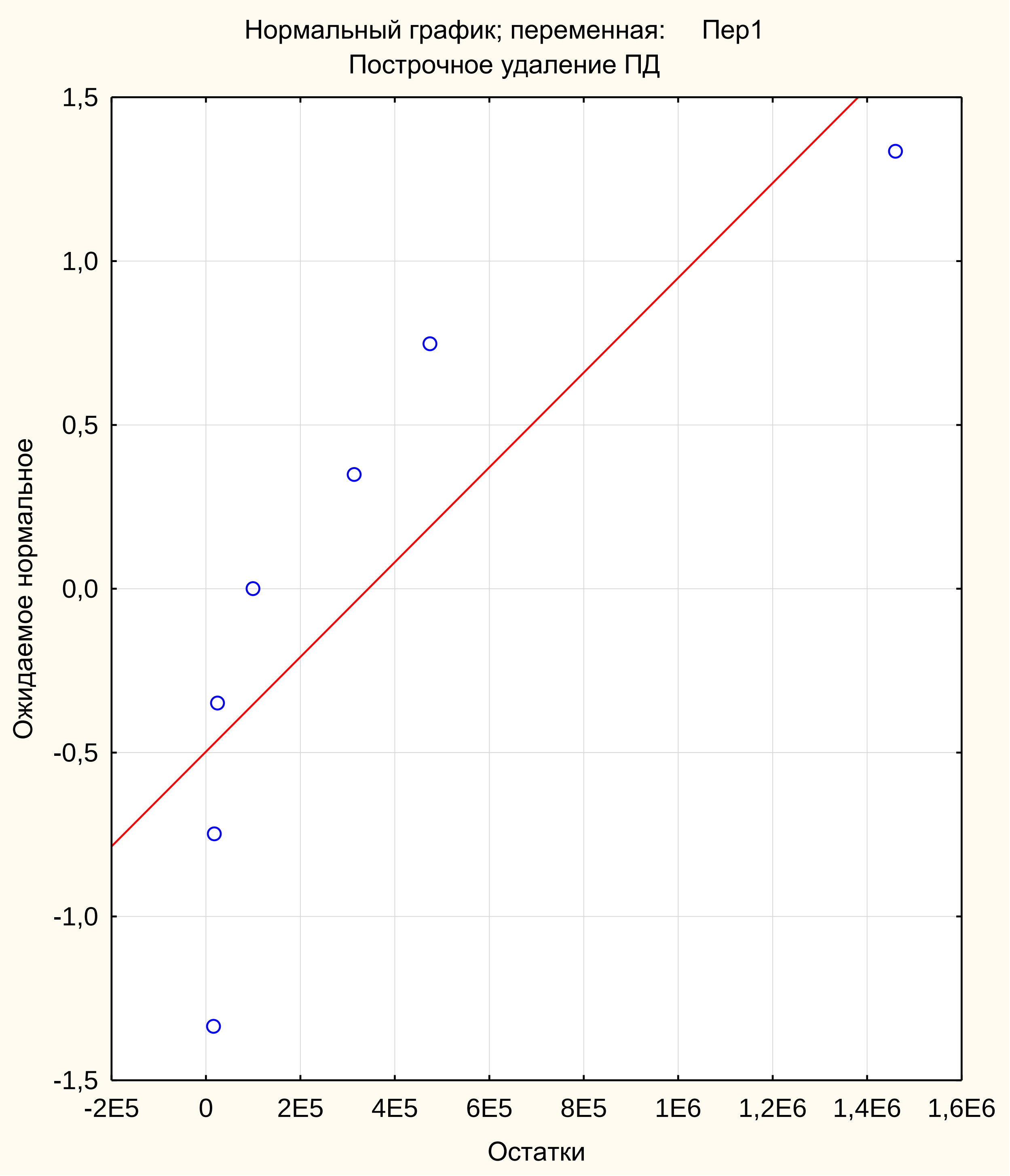



*Рис. 9.33. Результаты построения нормального графика по переменной 1*

Построение трехмерной гистограммы предполагает выбор переменных для расчетов (рис. 9.34).

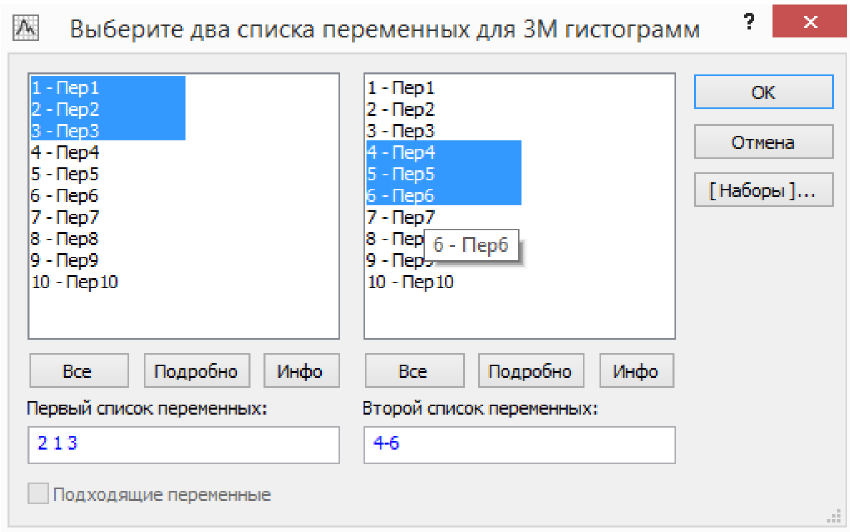



*Рис. 9.34. Окно выбора переменных для построения 3М гистограмм*

При выборе данных переменных результаты построения представлены на рис. 9.35:

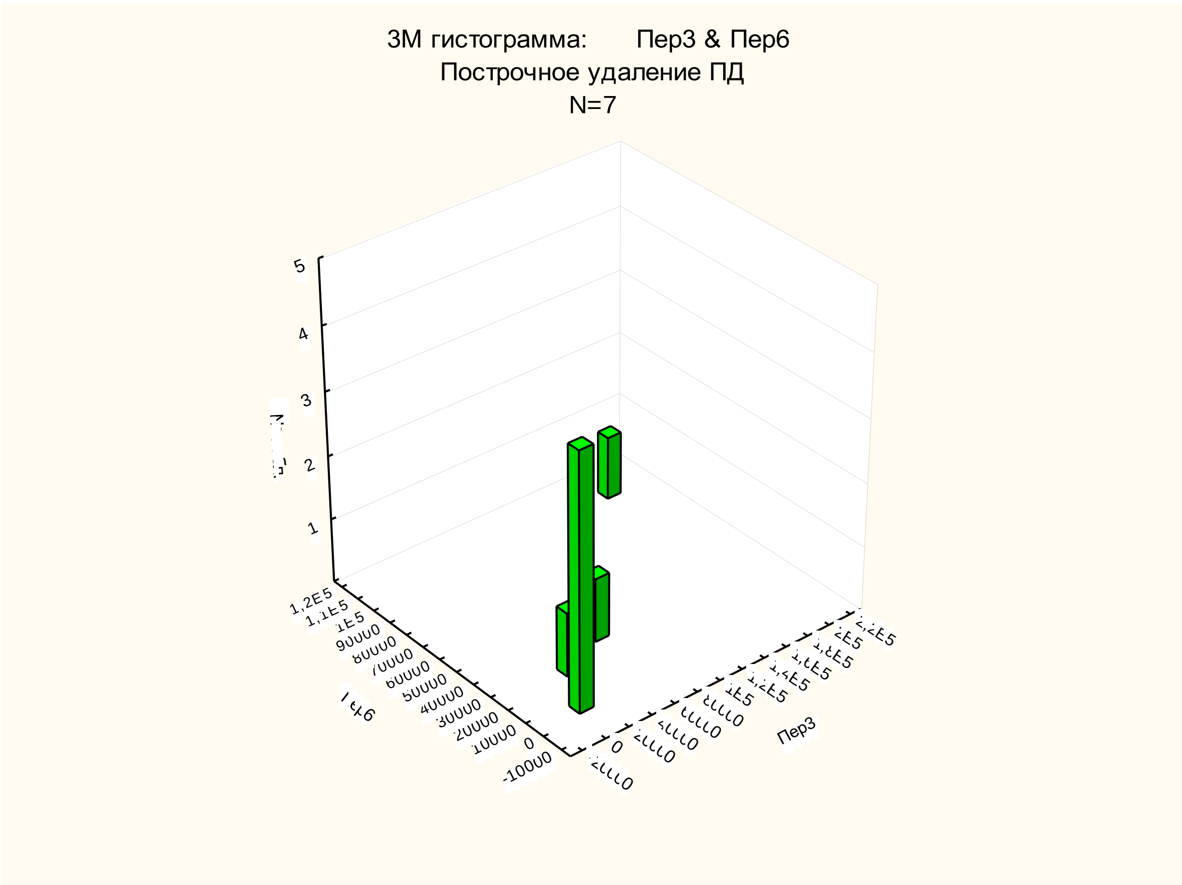



*Рис. 9.35. Результаты построения 3М гистограммы*

Построение двухмерной диаграммы рассеяния также предполагает предварительных выбор переменных для анализа (рис. 9.36):

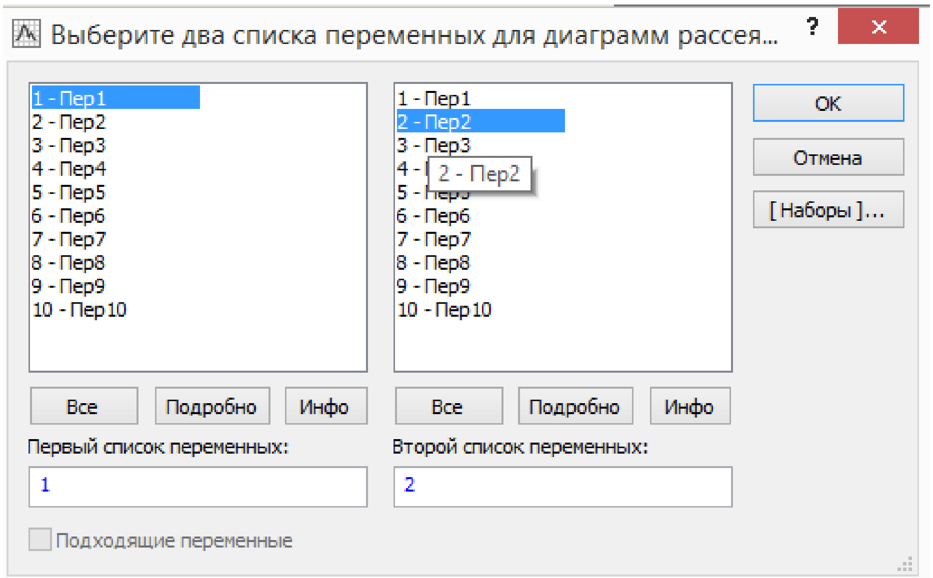



*Рис. 9.36 – Окно выбора переменных для построения 2М диаграммы рассеяния*

Результаты построения 2М пространстве представлены на рис. 9.37.:

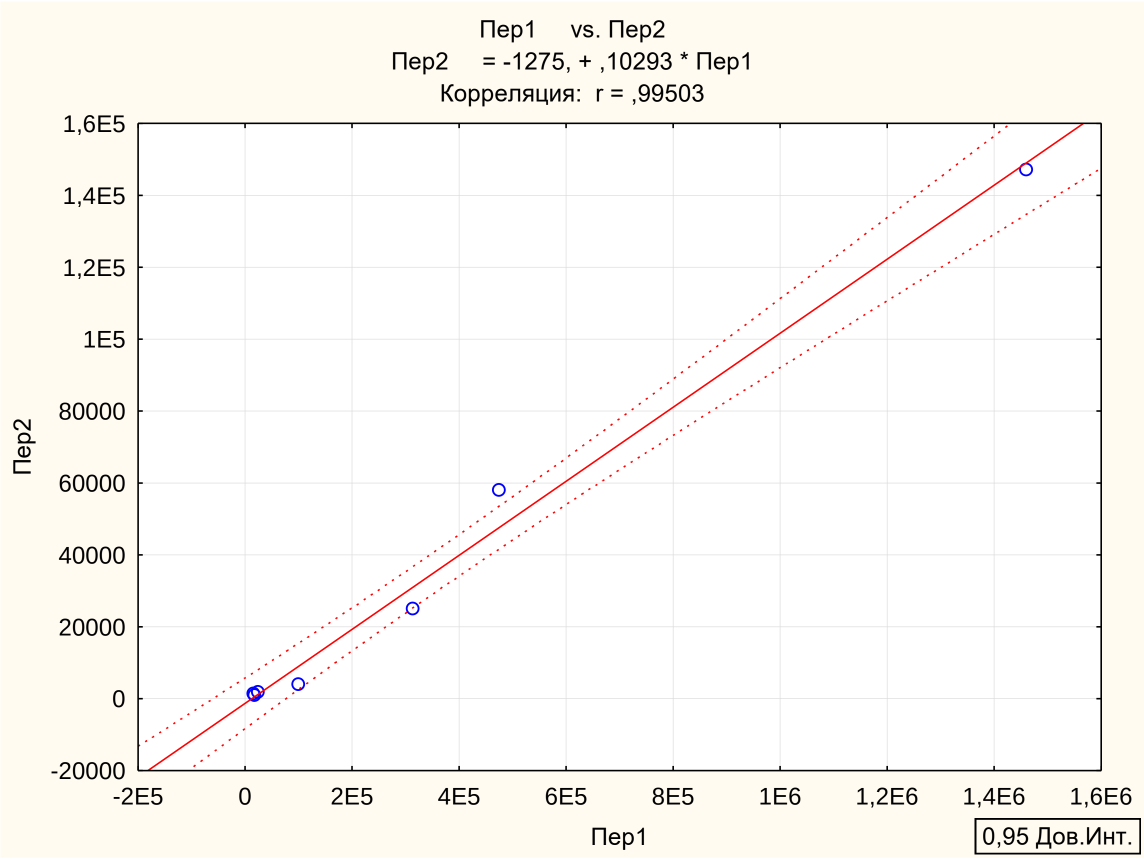



*Рис. 9.37. Результаты построения 2М диаграммы рассеяния*

Выбор переменных для построения 3М диаграммы рассеяния (рис. 9.38):

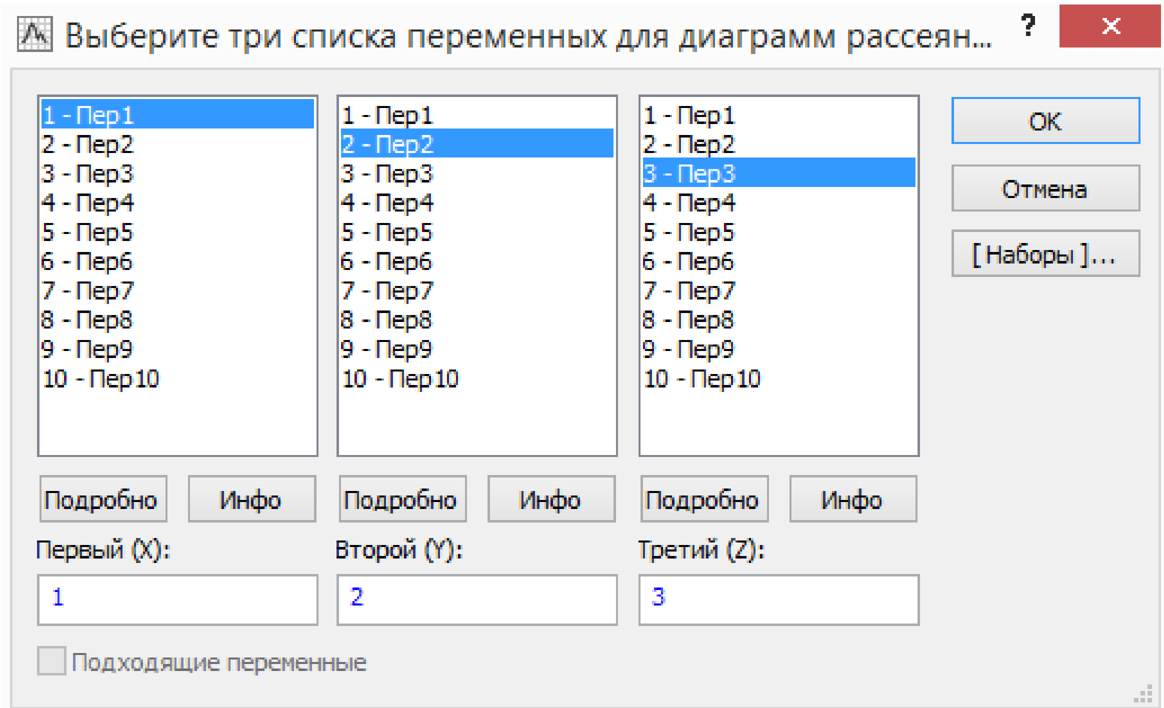



*Рис. 9.38. Окно выбора переменных для построения 3М диаграммы рассеяния*

Результаты построения при заданных параметров анализа (рис. 9.39):

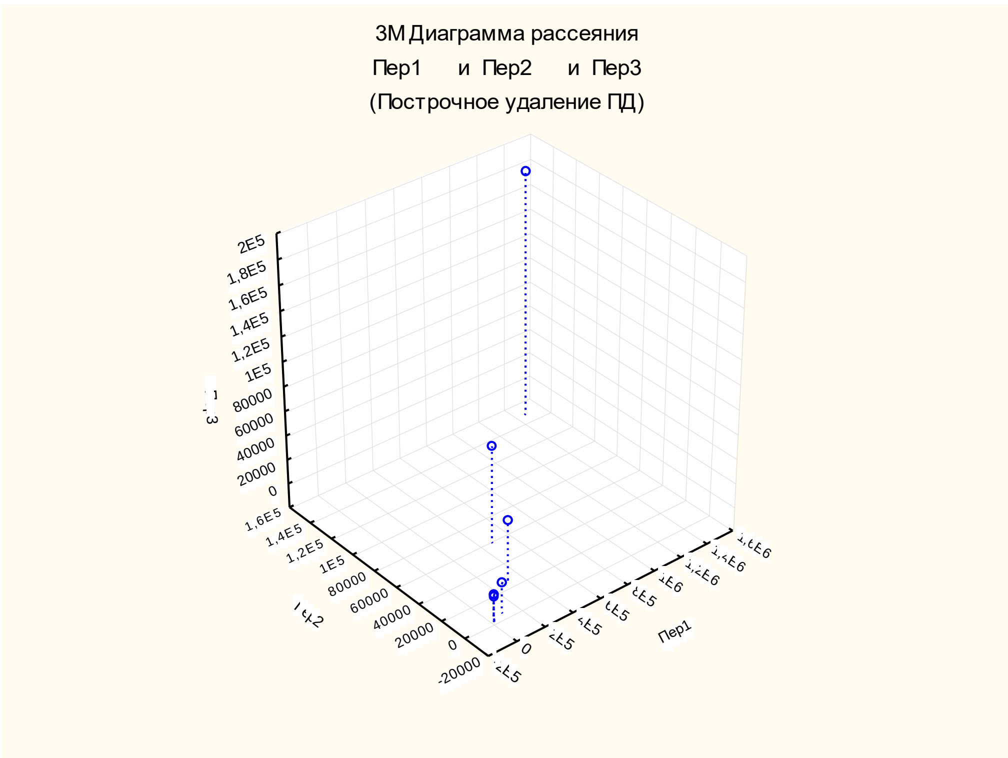



*Рис. 9.39. Результаты построения 3М диаграммы рассеяния*

В программе «Statistica» также есть возможность сохранения построенной матрицы (рис. 9.40):

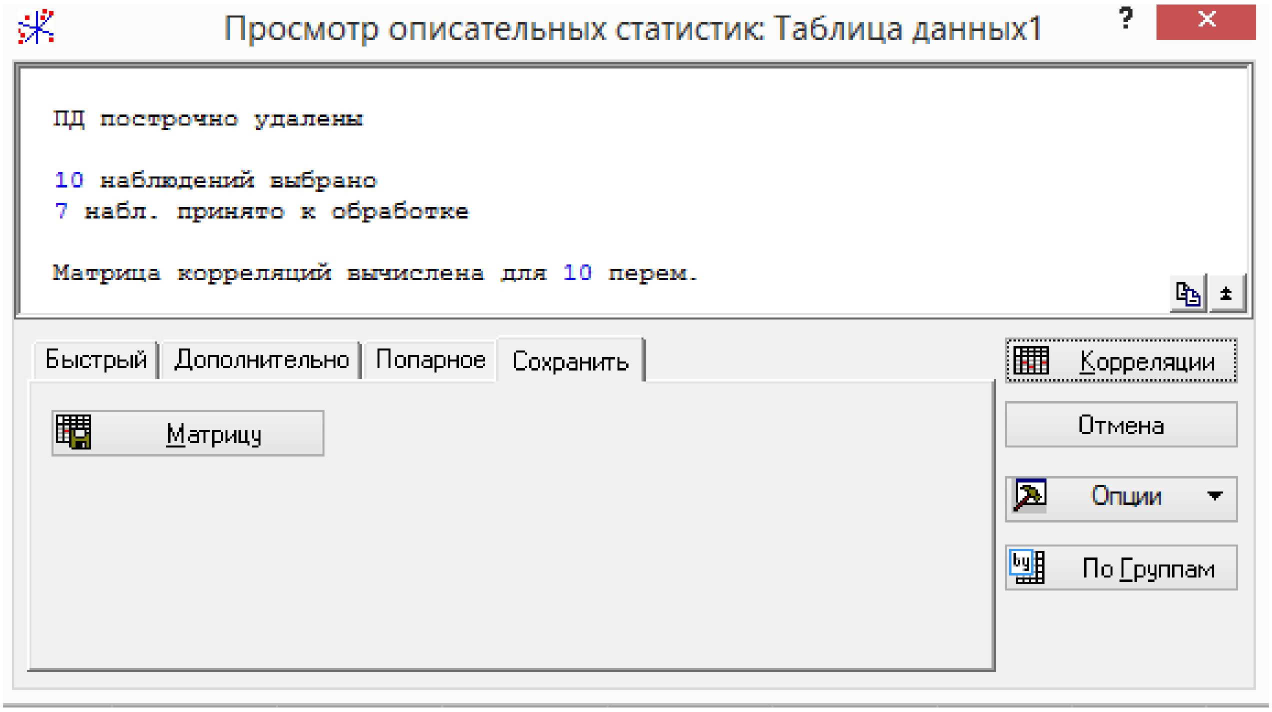



*Рис. 9.40. Окно сохранения матрицы*

Результаты построения при заданных данных представлены на рис. 9.41:

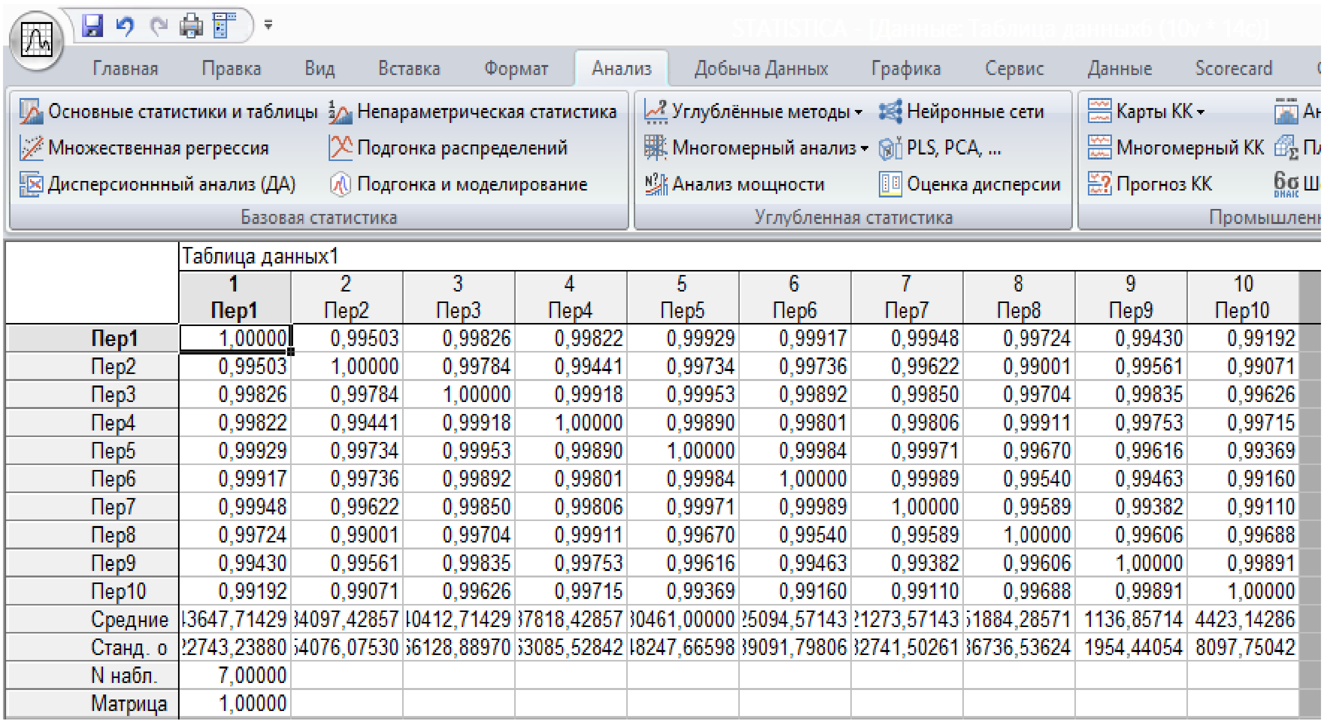



Таблица данных1

| | 1 Пер1 | 2 Пер2 | 3 Пер3 | 4 Пер4 | 5 Пер5 | 6 Пер6 | 7 Пер7 | 8 Пер8 | 9 Пер9 | 10 Пер10 |
|---|---|---|---|---|---|---|---|---|---|---|
| **Пер1** | 1,00000 | 0,99503 | 0,99826 | 0,99822 | 0,99929 | 0,99917 | 0,99948 | 0,99724 | 0,99430 | 0,99192 |
| Пер2 | 0,99503 | 1,00000 | 0,99784 | 0,99441 | 0,99734 | 0,99736 | 0,99622 | 0,99001 | 0,99561 | 0,99071 |
| Пер3 | 0,99826 | 0,99784 | 1,00000 | 0,99918 | 0,99953 | 0,99892 | 0,99850 | 0,99704 | 0,99835 | 0,99626 |
| Пер4 | 0,99822 | 0,99441 | 0,99918 | 1,00000 | 0,99890 | 0,99801 | 0,99806 | 0,99911 | 0,99753 | 0,99715 |
| Пер5 | 0,99929 | 0,99734 | 0,99953 | 0,99890 | 1,00000 | 0,99984 | 0,99971 | 0,99670 | 0,99616 | 0,99369 |
| Пер6 | 0,99917 | 0,99736 | 0,99892 | 0,99801 | 0,99984 | 1,00000 | 0,99989 | 0,99540 | 0,99463 | 0,99160 |
| Пер7 | 0,99948 | 0,99622 | 0,99850 | 0,99806 | 0,99971 | 0,99989 | 1,00000 | 0,99589 | 0,99382 | 0,99110 |
| Пер8 | 0,99724 | 0,99001 | 0,99704 | 0,99911 | 0,99670 | 0,99540 | 0,99589 | 1,00000 | 0,99606 | 0,99688 |
| Пер9 | 0,99430 | 0,99561 | 0,99835 | 0,99753 | 0,99616 | 0,99463 | 0,99382 | 0,99606 | 1,00000 | 0,99891 |
| Пер10 | 0,99192 | 0,99071 | 0,99626 | 0,99715 | 0,99369 | 0,99160 | 0,99110 | 0,99688 | 0,99891 | 1,00000 |
| Средние | 3647,71429 | 4097,42857 | 0412,71429 | 7818,42857 | 0461,00000 | 5094,57143 | 1273,57143 | 1884,28571 | 1136,85714 | 4423,14286 |
| Станд. о | 2743,23880 | 4076,07530 | 6128,88970 | 3085,52842 | 8247,66598 | 9091,79806 | 2741,50261 | 6736,53624 | 1954,44054 | 8097,75042 |
| N набл. | 7,00000 | | | | | | | | | |
| Матрица | 1,00000 | | | | | | | | | |

*Рис. 9.41. Результаты построения диаграммы*

Результаты факторного анализа представлены на рис. 9.42:

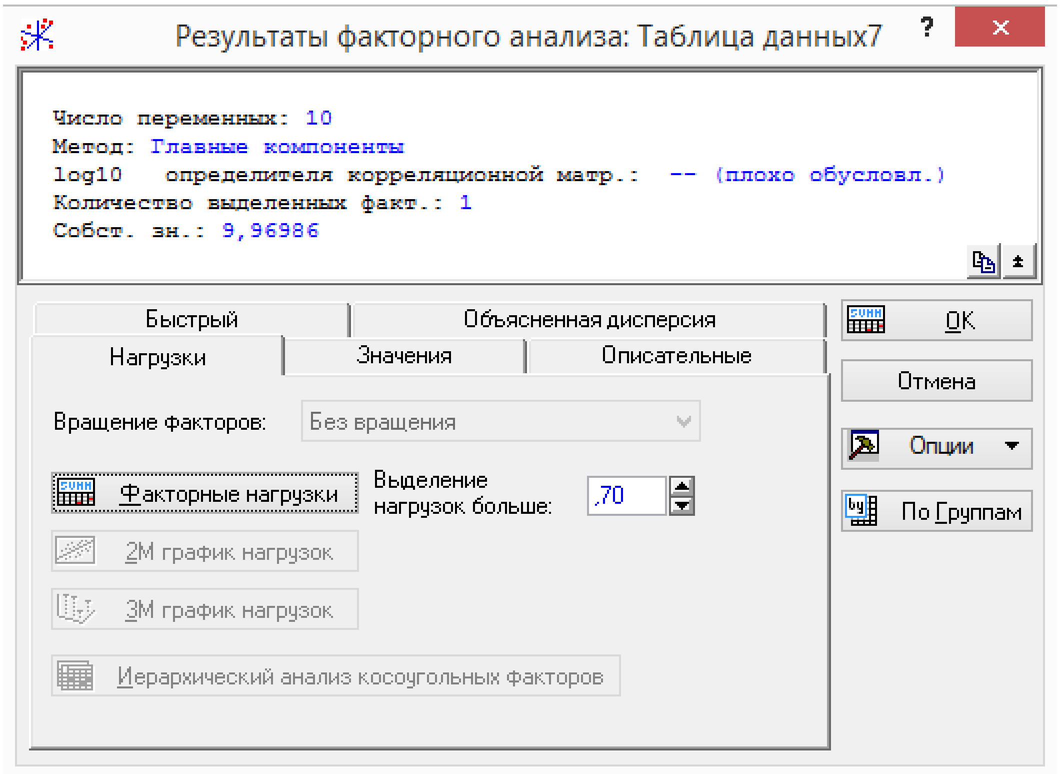
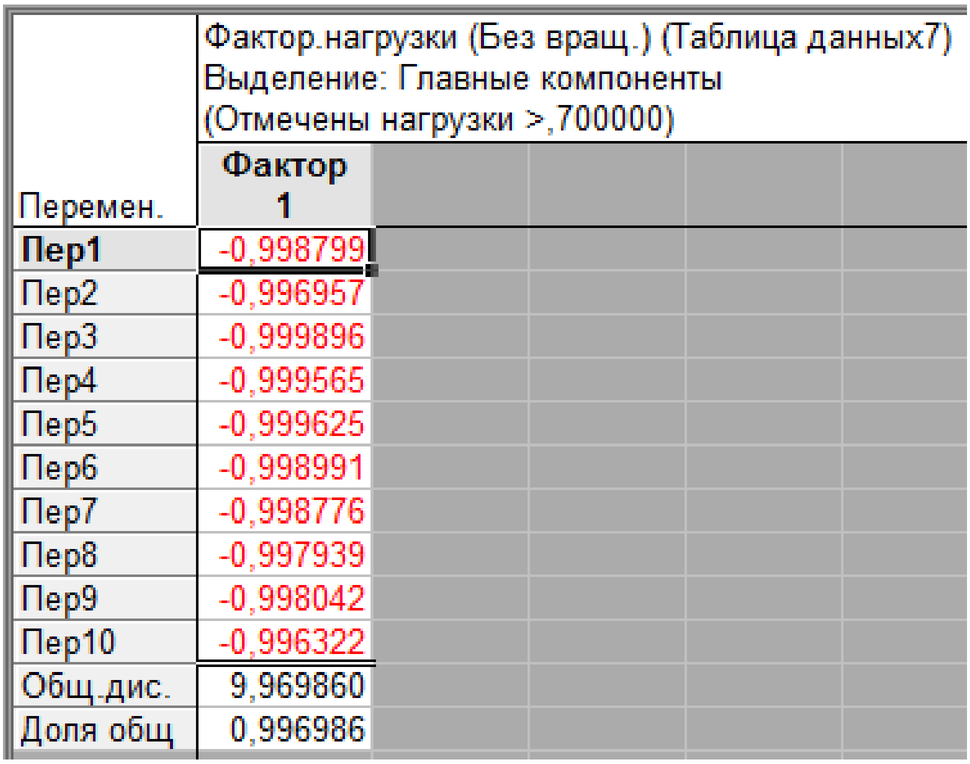



*Рис. 9.42. Результаты факторного анализа*

Факторные нагрузки будут выглядеть следующим образом (рис. 9.43):



Фактор.нагрузки (Без вращ.) (Таблица данных7)
Выделение: Главные компоненты
(Отмечены нагрузки >,700000)

| Перемен. | Фактор 1 |
|---|---|
| **Пер1** | -0,998799 |
| Пер2 | -0,996957 |
| Пер3 | -0,999896 |
| Пер4 | -0,999565 |
| Пер5 | -0,999625 |
| Пер6 | -0,998991 |
| Пер7 | -0,998776 |
| Пер8 | -0,997939 |
| Пер9 | -0,998042 |
| Пер10 | -0,996322 |
| Общ.дис. | 9,969860 |
| Доля общ | 0,996986 |

*Рис. 9.43. Результаты вычисления факторных нагрузок*

Результаты вычисления коэффициентов факторов представлены на рис. 9.44:

| Перемен. | Коэффициенты факторов (Таблица данных7)<br>Вращение: Без вращ.<br>Выделение: Главные компоненты<br>Фактор 1 |
|---|---|
| Пер1 | -0,100182 |
| Пер2 | -0,099997 |
| Пер3 | -0,100292 |
| Пер4 | -0,100259 |
| Пер5 | -0,100265 |
| Пер6 | -0,100201 |
| Пер7 | -0,100180 |
| Пер8 | -0,100096 |
| Пер9 | -0,100106 |
| Пер10 | -0,099933 |

*Рис. 9.44. Коэффициенты факторов*

Значения факторов при выделении по методу главных компонент (рис. 9.45):

| Набл. | Значения факторов (Таблица данных7)<br>Вращение: Без вращ.<br>Выделение: Главные компоненты<br>Фактор 1 |
|---|---|
| 1 | 0,50785 |
| 2 | 0,59148 |
| 3 | 0,15053 |
| 4 | 0,59176 |
| 5 | 0,57638 |
| 6 | -0,26989 |
| 7 | -2,14810 |

*Рис. 9.45. Значения факторов*

Основная задача данного параграфа состоит в возможности демонстрации возможностей программного продукта для решения задачи факторного анализа. Исходные данные, используемые в примере, являются произвольными величинами и не могут быть использованы для выполнения реальных статистически значимых вычислений.

## Контрольные вопросы

1. Для чего используется процедура факторного анализа?
2. Какие подходы при объединении переменных в факторы Вам известны?
3. Какие переменные называются латентными?
4. Какое предположение определяет сущность факторного анализа?
5. Какие факторы принято называть общими?
6. Что понимают в широком смысле слова при использовании термина «факторный анализ»?
7. Что понимают в узком смысле слова при использовании термина «факторный анализ»?
8. Существует ли возможность однозначного восстановления каждого из значений признака на основе определенных по результатам исследования общих факторов? Объясните, почему.
9. За счет чего возникает статистическая связь между x(j) и f (m)?
10. Возможно ли абсолютное достижение постеленной цели при проведении процедуры факторного анализа? Почему?
11. В каком случае результаты факторного анализа можно считать успешными?
12. Назовите основные уровни проведения процедуры факторного анализа.
13. За счет чего становится возможным снижение размерности массива исходных данных, который описывает процесс или явление?
14. Перечислите основные этапы осуществления процедуры факторного анализа
15. Что включают в себя исследовательские цели факторного анализа?
16. Что предполагает реализация прикладных целей факторного анализа?
17. Сформулируйте основные положения факторного анализа.
18. Каковы основные условия осуществления факторного анализа?
19. По отношению к каким объектам возможно проведение процедуры факторного анализа?
20. В каком виде представляется матрица наблюдений при традиционном подходе?

21. В чем состоит принципиальное отличие модели факторного анализа от регрессионных схем?
22. Какие параметры необходимо определить в ходе реализации факторного анализа?
23. Какие основные подходы к построению факторных моделей Вам известны?
24. В чем состоит сущность метода главных компонент?
25. В чем состоит основное условие выбора линейных комбинаций?
26. Каков условный вид нормировок для векторов факторных нагрузок?
27. Какую модель факторного анализа называют канонической?
28. Назовите условия существования канонической модели факторного анализа?
29. В каком случае задача факторного анализа предполагает единственное решение?
30. Имеет ли задача канонического факторного анализа аналитическое решение?
31. Какое допущение предполагает решение задачи методом максимального правдоподобия?
32. С помощью какой системы уравнений описываются оценки матрицы факторных нагрузок и дисперсионной матрицы?
33. Каковы условия нормировки для собственных векторов?
34. Что предполагает применение численной итерационной процедуры?
35. Что делать, если нулевая гипотеза о допустимости представления исходных признаков в виде m- факторной модели отвергается?
36. Является ли определение факторов анализа заключительной целью исследования?
37. Как рассчитывается суммарная общность факторов?
38. Что представляет собой коэффициент информативности признаков?
39. Что можно оценить с помощью коэффициента информативности признака?
40. Как осуществляется интерпретация факторов, если факторные нагрузки обладают равномерным распределением?
41. В каком случае необходимо прибегать к процессу вращения факторов?

42. С какой точностью определяются обобщенные факторы?
43. Как можно реализовывать ортогональное вращение факторного пространства?
44. В чем состоит сузность метода варимакс?
45. Назовите черты сходства и различия методов варимакс и квартимакс
46. Какую процедуру осуществляют для того, чтобы избежать влияния на результаты вращения переменных с большой общностью?
47. Как осуществляется нормировка факторных нагрузок?
48. Как реализуется задача оценивания значений факторов, если в рамках анализа использовался метод главных компонент?
49. Какие методы оценивания значений факторов канонической модели факторного анализа Вам известны?
50. В чем состоит сущность метода Бартлетта?
51. Каким образом реализуется метод Томпсона?
52. Приведите пример факторного анализа в рамках исследования показателей деятельности организации

# Глава 10. ДИСПЕРСИОННЫЙ АНАЛИЗ

## 10.1. Сущность дисперсионного анализа в статистике фирмы

**Дисперсионный анализ** представляет собой статистический метод, с помощью которого осуществляется анализ степени влияния факторов на результаты эксперимента, путем исследования значимости различий в средних значениях.

Сущность дисперсионного анализа (analysis of variance, ANOVA) заключается в том, чтобы разбить дисперсию измеряемого признака на отдельные элементы, описывающие влияние каждого отдельного фактора, а также их взаимодействия.

Последующее сравнение данных элементов позволяет выявить долю вариации исходных данных, обусловленную факторным влиянием и влиянием случайных отклонений.

Соответственно, становится возможной оценка значимости каждого из рассматриваемых факторов, а также их вариантов комбинаций.

Фактором при осуществлении дисперсионного анализа называется переменная, которая, предположительно, может оказывать значимый вклад на формирование конечного результата.

Дисперсионный анализ называется однофакторным, если рассматривается зависимость только от одного фактора, и многофакторным, если анализируется влияние двух или более признаков на результат.

Применительно к статистике фирмы могут быть рассмотрены следующие ситуации: предположим, что требуется построение модели объяснения выручки фирм тем, что они расположены в различных городах страны.

В данном случае переменная «месторасположение фирмы» будет выполнять роль анализируемого фактора. Уровнем фактора является конкретное его значение (например, наименование населенного пункта расположения фирмы).

Откликом в дисперсионном анализе принято называть значение измеряемого признака (в рамках рассматриваемого примера откликов выступают конкретные значения выручки фирм).

## 10.2. Однофакторный дисперсионный анализ

Для однофакторного дисперсионного анализа модель выглядит следующим образом (10.1)::

$$y_{ij} = \mu_j + \varepsilon_{ij} = \mu + \alpha_j + \varepsilon_{ij}, \quad i = \overline{1, n_j}, \quad j = \overline{1, k} \qquad (10.1)$$

где: $y_{ij}$ - i-ое наблюдаемое значение отклика в *j*-ой группе (для *j*-го уровня фактора);

$\mu$ - среднее значение отклика по всем уровням фактора (среднее по всей совокупности);

$\mu_j$ - среднее значение отклика для *j*-го уровня фактора;

$\alpha_j = \mu_j - \mu$ - дифференциальный эффект среднего, который соответствует *j*-му уровню фактора;

$\varepsilon_{ij}$ - независимые случайные величины с математическим ожиданием равным нулю и одинаковой дисперсией $\sigma^2$.

Очевидно, что при заданных величинах $\mu_j$, величины $\alpha_j$ и $\mu$ определяются неоднозначно, поэтому необходимо наложить дополнительное условие, устанавливающие связь между этими величинами. Обычно используют одно из следующих условий (10.2 или 10.3)

$$: \sum_{i=1}^{k} \alpha_i = 0, \qquad (10.2)$$

$$\sum_{i=1}^{k} n_i \alpha_i = 0, \ \alpha_k = 0. \qquad (10.3)$$

Выражение $y_{ij} = \mu + \alpha_j + \varepsilon_{ij}$ можно представить в виде (10.4)

$$y_{ij} = \mu + (\mu_j - \mu) + (y_{ij} - \mu_j) \qquad (10.4)$$

или:

$$y_{ij} - \mu = (\mu_j - \mu) + (y_{ij} - \mu_j). \qquad (10.5)$$

Математический смысл данного выражения заключается в следующем: отклонение наблюдаемого значения отклика для j-ой группы складывается из суммы двух слагаемых. Одним из них выступает отклонение отклика от среднего значения *j*-ой группы: $(y_{ij} - \mu_j)$, а вторым отклонение среднего значения *j*-ой группы от среднего значения всей совокупности: $(\mu_j - \mu)$.

Как правило, разложение общей дисперсии для выборочных данных, представляется в виде равенства сумм квадратов соответствующих отклонений:

$$SS_T = SS_B + SS_R \tag{10.6}$$

где:

$k$ – число уровней фактора,

$n_j$ – число наблюдений для $j$-го уровня фактора,

$n = \sum_{j=1}^{k} n_j$ - общее число наблюдений.

Общая сумма квадратов отклонений (10.7):

$$SS_T = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij} - \bar{y})^2 \tag{10.7}$$

Сумма квадратов отклонений групповых средних от общего среднего (10.8):

$$SS_B = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(\bar{y}_i - \bar{y})^2 = \sum_{i=1}^{k} n_i(\bar{y}_i - \bar{y})^2 \tag{10.8}$$

Выражение 10.8 также называется эффектом фактора (суммой квадратов эффекта )

Внутригрупповая (остаточная) сумма квадратов отклонений определяется по формуле 10.9:

$$SS_R = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_i)^2 \tag{10.9}$$

Выражение 10.9 также называется остаточным эффектом (эффектом ошибок )

Таким образом, в разложении дисперсии на составляющие заключена основная идея дисперсионного анализа: общая вариация переменной, обусловленная влиянием фактора и измеренная суммой $SS_T$, складывается из двух компонент: $SS_B$ и $SS_R$, которые используются для описания изменчивости переменной между уровнями фактора ($SS_B$) и внутри уровней фактора ($SS_R$).

При осуществлении дисперсионного анализа анализируются не сами значения суммы квадратов отклонений, а средние квадраты. Они получаются путем деления сумм квадратов отклонений на соответствующее число степеней свободы.

Число степеней свободы для суммы квадратов случайных величин определяется как общее число линейно независимых слагаемых .

Для полной суммы квадратов $SS_T = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij} - \bar{y})^2$ число степеней свободы $\nu_T = n - 1$, так как при ее расчете используются n наблюдений, которые связаны между собой единственным уравнением для общего выборочного среднего всей совокупности.

Для суммы квадратов эффекта фактора $SS_B = \sum_{i=1}^{k} n_i(\bar{y}_i - \bar{y})^2$ число степеней свободы $\nu_B = k - 1$, так как при ее расчете используются $k$ групповых средних, связанных между собой также одним уравнением для общего выборочного среднего всей совокупности.

Для суммы квадратов ошибок $SS_R = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_i)^2$ число степеней свободы $\nu_R = n - k$, ибо при его расчете используются $n$ наблюдений, связанных между собой $k$ уравнениями для выборочных средних $k$ групп.

Соответственно выражения для средних квадратов отклонений, которые являются оценками соответствующих дисперсий, имеют вид (10.10-10.12).

Оценка общей дисперсии?

$$MS_T = \frac{1}{n-1}\sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij} - \bar{y})^2 \tag{10.10}$$

Оценка межгрупповой дисперсии (10.11):

$$MS_B = \frac{1}{k-1}\sum_{i=1}^{k} n_i(\bar{y}_i - \bar{y})^2 \tag{10.11}$$

Оценка остаточной дисперсии (10.12)

$$MS_R = \frac{1}{n-k}\sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_i)^2 \tag{10.12}$$

При условии истинности нулевой гипотезы $H_0 : \mu_1 = \mu_2 = \ldots = \mu_k$, статистики $MS_B$ и $MS_R$, являются несмещенными оценками одной и той же дисперсии $\sigma^2$.

В таких условиях сущность проверки нулевой гипотезы состоит в анализе равенства дисперсий на основе F-отношения (10.13):

$$F = \frac{MS_B}{MS_R} = \frac{n-k}{k-1}\frac{SS_B}{SS_R} \tag{10.13}$$

Если нулевая гипотеза верна, статистика $F$ в случае нормального распределения величин $\varepsilon_{ij}$ обладает распределением Фишера с $\nu_1 = k-1$ и $\nu_2 = n-k$ числом степеней свободы.

В случае, если наблюдаемое значение больше или равно критического значения распределения Фишера с уровнем $\alpha$ и с числом степеней свободы $\nu_1 = k-1$ и $\nu_2 = n-k$ (то есть $F_{набл} \geq F_{кр}$), нулевая гипотеза отклоняется и считается, что средние для различных уровней фактора значимо различаются.

Если применяются порядковые данные, непараметрической альтернативой однофакторного дисперсионного анализа являются ранговый дисперсионный анализ Краскела–Уоллиса.

Его основой является однофакторный дисперсионный анализ, однако вместо исходных значений переменных анализируются их ранговые характеристики.

Если обозначить через $R_{ij}$ ранг элемента $x_{ij}$, в общем вариационном ряду значений отклика, то величины $\overline{R}_j = \frac{1}{n_j}\sum_{i=1}^{n_j} R_{ij}$ будут определять средние ранги для элементов j-ой группы, а величина $\overline{R} = \frac{1}{n}\sum_{i=1}^{n_j}\sum_{j=1}^{k} R_{ij} = \frac{n+1}{2}$ средний ранг всей совокупности. Таким бразом, величина $\sum_{j=1}^{k} n_j(\overline{R}_j - \overline{R})^2$ будет использоваться для характеристики межгруппового разброса рангов.

Если нулевая гипотеза о равенстве средних рангов верна, статистика примет вид (10.14):

$$H = \frac{12}{n(n+1)}\sum_{j=1}^{k} n_j(\overline{R}_j - \overline{R})^2 \qquad (10.14)$$

Приближенно она будет равна распределению Хи-квадрат с $k-1$ степенью свободы.

В случае, если соблюдается неравенство $\mathrm{H}_{набл} \geq \mathrm{H}_{кр}$, нулевая гипотеза отклоняется и считается, что средние ранги для различных уровней фактора значимо различаются.

## 10.3. Многофакторный дисперсионный анализ

Если производится анализ двух и более различных факторов на результаты наблюдений, имеет место ***многофакторный дисперсионный анализ***.

Примером двухфакторной модели в статистике фирмы может служить следующая модель: Например, двухфакторная модель будет применяться при построении объяснения различий в доходах фирм, обусловленных как месторасположением организаций, так и спецификой их вида деятельности.

Рассмотрим подробнее влияние на величину X фактора A, имеющего kуровней, и фактора B,имеющего m уровней.

Построение двухфакторной модели опирается на выражение (10.15):

$$y_{ijl} = \mu_{ij} + \varepsilon_{ijl} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijl}, \quad l = \overline{1, n_{ij}}, \quad i = \overline{1, k}, \quad j = \overline{1, m}, \qquad (10.15)$$

где: $y_{ijl}$ - *l*-ое наблюдаемое значение отклика для *i*-го уровня фактора A и *j*-го уровня фактора B;

$\mu$ - среднее значение отклика по всей совокупности (генеральное среднее);

$\mu_{ij}$ - среднее значение отклика для *i*-го уровня фактора A и *j*-го уровня фактора B;

$\alpha_i = \mu_{i*} - \mu$ - главный эффект *i*-го уровня фактора A ($\mu_{i*}$ - среднее значение отклика для *i*-го уровня фактора A);

$\beta_j = \mu_{*j} - \mu$ - главный эффект *j*-го уровня фактора B ($\mu_{*j}$ - среднее значение отклика для *j*-го уровня фактора B);

$\gamma_{ij} = \mu_{ij} - \mu_{i*} - \mu_{*j} + \mu$ - эффект взаимодействия *i*-го уровня фактора A и *j*-го уровня фактора B;

$\varepsilon_{ijl}$ - независимые случайные величины с математическим ожиданием равным нулю и одинаковой дисперсией $\sigma^2$.

Средние могут определяться через величины $\mu_{ij}$, например, как взвешенные средние:

$$\mu = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{m}\sum_{l=1}^{n_{ij}} M(y_{ijl}) = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{m} n_{ij}\mu_{ij};$$

$$\mu_{i*} = \frac{1}{n_{i*}}\sum_{j=1}^{m}\sum_{l=1}^{n_{ij}} M(y_{ijl}) = \frac{1}{n_{i*}}\sum_{j=1}^{m} n_{ij}\mu_{ij};$$

$$\mu_{*j} = \frac{1}{n_{*j}}\sum_{i=1}^{k}\sum_{l=1}^{n_{ij}} M(y_{ijl}) = \frac{1}{n_{*j}}\sum_{i=1}^{k} n_{ij}\mu_{ij};$$

$$n = \sum_{i=1}^{k}\sum_{j=1}^{m} n_{ij},$$

$$n_{i*} = \sum_{j=1}^{m} n_{ij},$$

$$n_{*j} = \sum_{i=1}^{k} n_{ij}.$$

Возможны и другие варианты определения средних величин.

Выражение $y_{ijl} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijl}$ можно представить в виде (10.16):

$$y_{ijl} - \mu = (\mu_{i*} - \mu) + (\mu_{*j} - \mu) + (\mu_{ij} - \mu_{i*} - \mu_{*j} + \mu) + (x_{ijl} - \mu_{ij}) \qquad (10.16)$$

Данное соотношение говорит о том, что отклонение наблюдаемого значения отклика складывается из суммы четырех слагаемых: отклонения отклика от среднего значения для *i, j*-го набора уровней факторов А и В $(y_{ijl} - \mu_{ij})$, главных эффектов *i*-го уровня фактора А и *j*-го уровня фактора В и эффекта взаимодействия.

Таким образом, дисперсия отклика может быть представлена в виде суммы четырех дисперсий, одна из которых характеризует внутригрупповую изменчивость для *i, j*-го набора уровней факторов А и В, а остальные соответствующие эффекты.

НМК оценки параметров модели двухфакторного дисперсионного анализа достаточно просто получить только в случае пропорциональных частот, то есть при условии (10.17):

$$\frac{n_{ij}}{n} = \frac{n_{i*}}{n}\cdot\frac{n_{*j}}{n} \quad \forall\, i = \overline{1,k},\; j = \overline{1,m}. \qquad (10.17)$$

Данные условия выполняются, например, если количества наблюдений $n_{ij}$ для каждого сочетания уровней факторов совпадают.

Если соблюдается данное условие, эксперимент принято считать сбалансированным.

В случае пропорциональных частот исследователь получает понятные МНК-оценки для параметров модели вне зависимости от условий, накладываемых на параметры модели:

$$\hat{\mu} = \bar{y},$$

$$\hat{\alpha}_i = \bar{y}_{i*} - \bar{y},$$

$$\hat{\beta}_j = \bar{y}_{*j} - \bar{y},$$

$$\hat{\gamma}_{ij} = \bar{y}_{ij} - (\bar{y}_{i*} + \bar{y}_{*j}) + \bar{y},$$

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{m}\sum_{l=1}^{n_{ij}} y_{ijl},$$

$$\bar{y}_{i*} = \frac{1}{n_{i*}}\sum_{j=1}^{m}\sum_{l=1}^{n_{ij}} y_{ijl},$$

$$\bar{y}_{*j} = \frac{1}{n_{*j}}\sum_{i=1}^{k}\sum_{l=1}^{n_{ij}} y_{ijl},$$

$$\bar{y}_{ij} = \frac{1}{n_{ij}}\sum_{l=1}^{n_{ij}} y_{ijl}$$

В случае пропорциональных также справедливо следующее разложение общей суммы квадратов отклонений на составляющие (10.18):

$$SS_T = SS_A + SS_B + SS_{AB} + SS_R, \qquad (10.18)$$

где:

$SS_T = \sum_{i=1}^{k}\sum_{j=1}^{m}\sum_{l=1}^{n_{ij}} (y_{ijl} - \bar{y})^2$ – общая, или полная, сумма квадратов отклонений;

$SS_A = \sum_{i=1}^{k}\sum_{j=1}^{m}\sum_{l=1}^{n_{ij}} (\bar{y}_{i*} - \bar{y})^2 = \sum_{i=1}^{k} n_{i*}(\bar{y}_{i*} - \bar{y})^2$ – сумма квадратов отклонений средних по уровням фактора A от общего среднего, или сумма квадратов главных эффектов A (можно и так: сумма квадратов, соответствующих эффекту фактора A);

$SS_B = \sum_{i=1}^{k}\sum_{j=1}^{m}\sum_{l=1}^{n_{ij}} (\bar{y}_{*j} - \bar{y})^2 = \sum_{j=1}^{m} n_{*j}(\bar{y}_{*j} - \bar{y})^2$ – сумма квадратов отклонений средних по уровням фактора B от общего среднего, или сумма квадратов главных эффектов B;

$SS_{AB} = \sum_{i=1}^{k}\sum_{j=1}^{m}\sum_{l=1}^{n_{ij}} (\bar{y}_{ij} - \bar{y}_{*j} - \bar{y}_{i*} + \bar{y})^2 = \sum_{i=1}^{k}\sum_{j=1}^{m} n_{ij}(\bar{y}_{ij} - \bar{y}_{*j} - \bar{y}_{i*} + \bar{y})^2$ – сумма квадратов взаимодействия эффектов A и B;

$SS_R = \sum_{i=1}^{k}\sum_{j=1}^{m}\sum_{l=1}^{n_{ij}}(y_{ijl} - \bar{y}_{ij})^2$ – остаточная сумма квадратов отклонений.

Число степеней свободы для сумм квадратов $SS_A$ и $SS_B$ равно соответственно $\nu_A = k-1$ и $\nu_B = m-1$. Число степеней свободы для суммы квадратов взаимодействия эффектов $SS_{AB}$ равно $\nu_{AB} = (\kappa-1)(m-1)$. Число степеней свободы суммы квадратов остатков $SS_R$ равно $\nu_R = n - km$. Соответственно средние суммы квадратов будут равны:

$$MS_A = \frac{SS_A}{k-1},$$

$$MS_B = \frac{SS_B}{m-1},$$

$$MS_{AB} = \frac{SS_{AB}}{(k-1)(m-1)},$$

$$MS_R = \frac{SS_R}{n-km}.$$

Так как в рамках построения двухфакторной модели рассматриваются различные эффекты от влияния факторов, необходимо осуществлять проверку гипотез значимости различных выявленных эффектов с помощью инструментов статистического анализа.

При условии истинности $H_0$:«эффект незначим» средний квадрат эффекта является несмещенной оценкой дисперсии $\sigma^2$, так же как и величина $MS_R$.

Поэтому в качестве статистик критериев проверки гипотез о значимости соответствующих эффектов можно использовать отношения средней суммы квадратов эффектов к средней сумме квадратов остатков.

При нормальном распределении остатков данные статистики имеют распределение Фишера с параметрами, определяемыми числами степеней свободы соответствующих сумм, участвующих в отношении.

В том случает, если наблюдаемое значение статистики $F_{набл} \geq F_{кр}$, где $F_{кр}$ - критическая точка распределения Фишера уровня $\alpha$ (или квантиль уровня $1-\alpha$) с числом степеней свободы $\nu_1$ и $\nu_2$, то нулевая гипотеза отклоняется и считается, что средние для различных уровней фактора значимо различаются.

В случае непропорциональных частот оценки МНК параметров модели, при заданных $k+m-1$ условиях на параметры модели, могут быть получены численно, однако в этом случае нарушается условие ортогональности сумм квадратов эффектов.

Соответственно, независимая одновременная проверка всех гипотез двухфакторного анализа не представляется возможной.

Однако всегда существует возможность проверки значимости любого из выявленных эффектов (либо линейной комбинации эффектов), рассматривая соответствующую однофакторную модель и выделив соответствующую сумму квадратов. Затем можно исключить данную сумму квадратов из общей суммы квадратов и проверить значимость другого эффекта и т.д.

Таким образом, происходит последовательное разложение суммы квадратов на составляющие.

Можно использовать и другой подход, выделив на основе однофакторной модели сумму квадратов всех эффектов, за исключением одного, а затем уже исследовать значимость этого эффекта на основе оставшейся суммы квадратов.

В зависимости от того, как формируется разложение суммы квадратов на составляющие, различаются и статистики критериев для проверки гипотез о значимости эффектов в несбалансированной многофакторной модели дисперсионного анализа, при этом общий принцип формирования статистики критерия на основе F-отношения сумм квадратов эффекта и остатков остается прежним.

В многофакторной несбалансированной модели принято рассматривать 3 основных типа разложения суммы квадратов.

При этом в анализе будут рассматриваться следующие обозначения.

$R(\mu)$ – остаточная сумма квадратов для модели содержащий только параметр $\mu$ - общее среднее;

$R(\mu, A)$ – остаточная сумма квадратов для модели учитывающей эффект фактора А, т.е. для модели, содержащей параметр $\mu$ и параметры $\alpha_i, i=\overline{1,k}$;

$R(\mu, B)$ – остаточная сумма квадратов для модели учитывающей эффект фактора В, т.е. для модели, содержащей параметр $\mu$ и параметры $\beta_j$, $j = \overline{1, m}$;

$R(\mu, A, B)$ – остаточная сумма квадратов для модели учитывающей эффекты факторов А и В, без учета взаимодействия факторов;

$R(\mu, A, B, AB)$ – остаточная сумма квадратов для полной модели, учитывающей как эффекты факторов А и В, так и эффект взаимодействия факторов АВ.

Следует отметить, что записи вида: $R(\mu, A)$, $R(\mu, A, B)$, $R(\mu, A, B, AB)$ однозначно характеризуют не только остаточную сумму квадратов, но и саму модель, поэтому в дальнейшем под этой записью будем понимать, как остаточную сумму квадратов, так и модель, в зависимости от контекста.

Под записью $SS(A \mid B)$, будем понимать сумму квадратов, соответствующих эффекту фактору А, после того как из разложения была удалена сумма квадратов, соответствующая эффекту фактора В.

**Тип I** называют также последовательной суммой квадратов. Разложение зависит от порядка эффектов в модели. Каждый последующий эффект скорректирован на предыдущие эффекты, эффекты взаимодействия оцениваются после эффектов факторов. Разложение является аддитивным по отношению к общей сумме квадратов.

Пример формирования сумм квадратов на примере двухфакторной модели (в качестве первого фактора выбирается фактор А) приведен ниже.

Суммы квадратов I типа, для эффектов модели, где в качестве первого фактора выбран А будут иметь вид:

эффект А: $SS(A) = R(\mu) - R(\mu, A)$;

эффект В: $SS(B \mid A) = R(\mu, A) - R(\mu, A, B)$;

эффект АВ: $SS(AB \mid A, B) = R(\mu, A, B) - R(\mu, A, B, AB)$.

Все суммы определяются однозначно, вне зависимости от условий, накладываемых на параметры модели. Таким образом, чтобы получить данные суммы потребуется построить 4 различные модели (хотя можно сократить число моделей, используя для построения сумм соответствующие оценочные функции).

В **типе II** суммы квадратов каждого эффекта в модели, корректируются по всем остальным "подходящим" эффектам, то есть, вычисляются после удаления из общей суммы квадратов сумм квадратов "подходящих" эффектов. Под "подходящим" понимается любой эффект, который не содержит исследуемый эффект. Разложение не зависит от порядка эффектов в модели и не является аддитивным по отношению к общей сумме квадратов.

Суммы квадратов II типа, для эффектов двухфакторной модели будут иметь вид:

эффект A: $SS(A|B) = R(\mu, B) - R(\mu, A, B)$;

эффект B: $SS(B|A) = R(\mu, A) - R(\mu, A, B)$;

эффект AB: $SS(AB|A, B) = R(\mu, A, B) - R(\mu, A, B, AB)$.

Все суммы определяются однозначно, вне зависимости от условий, накладываемых на параметры модели. Также, как и в предыдущем случае, чтобы получить данные суммы потребуется построить 4 различные модели.

При использовании типа III суммы квадратов каждого эффекта в модели, корректируются по всем остальным эффектам и являются ортогональными к суммам квадратов, содержащим исследуемый эффект. Разложение не зависит от порядка эффектов в модели и не является аддитивным по отношению к общей сумме квадратов.

Суммы квадратов **типа III**, для эффектов двухфакторной модели будут иметь вид:

эффект A: $SS(A|B, AB) = R(\mu, B, AB) - R(\mu, A, B, AB)$;

эффект B: $SS(B|A, AB) = R(\mu, A, AB) - R(\mu, A, B, AB)$;

эффект AB: $SS(AB|A, B) = R(\mu, A, B) - R(\mu, A, B, AB)$.

Суммы квадратов $R(\mu, A, AB)$ и $R(\mu, B, AB)$ зависят от условий, накладываемых на параметры модели.

Для выполнения условий ортогональности, параметры модели должны удовлетворять условиям:

$$\sum_{i=1}^{k} \alpha_i = 0,$$

$$\sum_{j=1}^{m} \beta_j = 0,$$

$$\sum_{i=1}^{k} \gamma_{ij} = 0, \ j = \overline{1, m-1}$$

$$\sum_{j=1}^{m}\gamma_{ij} = 0, i = \overline{1, k-1}$$

$$\sum_{i=1}^{k}\sum_{j=1}^{m}\gamma_{ij} = 0, i = \overline{1, k},\ j = \overline{1, m}.$$

Как и при использовании двух других типов, получение данных суммы должно опираться на построение 4х различных моделей. Также, как и в предыдущих случаях, чтобы получить данные суммы, потребуется построить 4 различные модели.

## 10.4. Дисперсионный анализ для повторных наблюдений

На практике часто исследователи сталкиваются с проблемой соответствия различным уровням фактора одних и тех же объектов.

Метод дисперсионного анализа может быть применен для анализа чистой прибыли фирм за различные периоды времени. Однако, сделать вывод о том, что исходные данные не являются взаимокоррелируемыми величинами однозначно нельзя.

В такой ситуации анализ эффектов фактора опирается на дисперсионный анализ, предполагающий исключение влияния зависимостей выборок, которые связаны с повторным рассмотрением одних и тех же объектов.

Данный метод - дисперсионный анализ зависимых выборок (повторных наблюдений). С его помощью возможно уменьшение общей дисперсии данных за счет исключения составляющей индивидуальных различий из остаточной дисперсии. Таким образом, повышается мощность критерия, что оказывает положительное влияние на модель.

Если имеется k-выборок, объем каждой из которых равен n, соответствующих k различным уровням фактора *W*, влияние которого на значения наблюдаемой переменной рассматривается.

Предполагается, что *j*-ое наблюдаемое значение отклика в каждой группе соответствует одному и тому же объекту наблюдения. В простейшей однофакторной модели дисперсионного анализа повторных измерений исходят из следующей модели порождения данных (10.19):

$$y_{ij} = \mu_{ij} + \varepsilon_{ij} = \mu + w_i + \tau_j + \varepsilon_{ij} \quad i = \overline{1,k}, \quad j = \overline{1,n}, \qquad (10.19)$$

где: $y_{ij}$ - *j*-ое наблюдаемое значение отклика в *i*-ой группе;

$\mu_{ij}$ - среднее значение отклика для *i*-го наблюдения в *j*-ой группе;

$\varepsilon_{ij}$ - независимые случайные величины с математическим ожиданием равным нулю и одинаковой дисперсией $\sigma^2$;

$\mu$ - среднее значение отклика;

$w_i = \mu_{i*} - \mu$ - эффект *i*-го уровня фактора *W*, $\mu_{i*} = \frac{1}{n}\sum_{j=1}^{n}\mu_{ij}$ - среднее для *i*-го уровня фактора;

$\tau_j = \mu_{*j} - \mu$ - эффект индивидуальности *j*-го наблюдения, $\mu_{*j} = \frac{1}{k}\sum_{i=1}^{k}\mu_{ij}$ - среднее для *j*-го объекта наблюдения;

$\varepsilon_{ij}$ - независимые случайные величины с математическим ожиданием равным нулю и одинаковой дисперсией $\sigma^2$.

Уравнение (10.19) можно рассматривать как соответствующее классической модели двухфакторного дисперсионного анализа, без компоненты, учитывающей взаимодействие факторов и числом наблюдений для каждой ячейки $n_{ij} = 1$ (что соответствует сбалансированному плану).

Таким образом, его можно записать в виде 10.20

$$y_{ij} - \mu_{*j} = (\mu_{i*} - \mu) + \varepsilon_{ij} = w_i + \varepsilon_{ij}, \quad i = \overline{1,k}, \quad j = \overline{1,n}. \tag{10.20}$$

Для того, чтобы исключить эффект индивидуальности наблюдений и анализировать только влияние фактора *W* на результаты наблюдений, мы должны перейти к рассмотрению величин $y_{ij} - \mu_{*j}$. МНК оценкой параметра $\mu_{*j}$ является выборочное среднее $\bar{y}_{*j} = \frac{1}{k}\sum_{i=1}^{k} y_{ij}$. Соответственно, сумма квадратов наблюдений за вычетом суммы квадратов, соответствующих индивидуальным различиям, будет равна $\sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{*j})^2$.

Если обозначить $z_{ij} = y_{ij} - \bar{y}_{*j}$, то разложение для суммы квадратов $\sum_{i=1}^{k}\sum_{j=1}^{n} z_{ij}^2$ будет иметь вид (10.21):

$$\sum_{i=1}^{k}\sum_{j=1}^{n} z_{ij}^2 = \sum_{i=1}^{k} n\bar{z}_i^2 + \sum_{i=1}^{k}\sum_{j=1}^{n}(z_{ij} - \bar{z}_i)^2 . \qquad (10.21)$$

Число степеней свободы для суммы $SS_T = \sum_{i=1}^{k}\sum_{j=1}^{n} z_{ij}^2$ равно $nk - n$, для суммы $SS_W = \sum_{i=1}^{k} n(\bar{z}_i - \bar{z})^2 = \sum_{i=1}^{k} n\bar{z}_i^2$ равно $k-1$, для суммы $SS_R = \sum_{i=1}^{k}\sum_{j=1}^{n}(z_{ij} - \bar{z}_i)^2$ равно $nk - n - k + 1 = (n-1)(k-1)$. Соответственно, средние суммы квадратов отклонений будут равны (10.22 и 10.23):

$$MS_W = \frac{SS_W}{k-1} = \frac{1}{k-1}\sum_{i=1}^{k} n\bar{z}_i^2 \qquad (10.22)$$

$$MS_R = \frac{SS_R}{(n-1)(k-1)} = \frac{1}{(n-1)(k-1)}\sum_{i=1}^{k}\sum_{j=1}^{n}(z_{ij} - \bar{z}_i)^2 \qquad (10.23)$$

F-статистика для проверки гипотезы $H_0 : w_1 = w_2 = \ldots = w_k = 0$ будет иметь вид: $F = \dfrac{MS_W}{MS_R}$. Если наблюдаемое значение статистики $F_{набл} \geq F_{кр}$, где $F_{кр}$ - критическая точка распределения Фишера уровня $\alpha$ (или квантиль уровня $1-\alpha$) с числом степеней свободы $\nu_1 = k-1$ и $\nu_2 = (n-1)(k-1)$, то нулевая гипотеза отклоняется и считается, что средние для различных уровней фактора $W$ значимо различаются.

Данный критерий можно получить и как критерий для проверки соответствующей гипотезы в модели двухфакторного дисперсионного анализа.

Для модели (10.20) справедливо следующей разложение для суммы квадратов отклонений :

$$\sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij} - \bar{y})^2 = \sum_{i=1}^{k} n(\bar{y}_{i*} - \bar{y})^2 + \sum_{j=1}^{n} k(\bar{y}_{*j} - \bar{y})^2 + \sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{*j} - \bar{y}_{i*} + \bar{y})^2$$

При данном разложении сумма квадратов, которая раньше в двухфакторной модели являлась суммой квадратов взаимодействий, стала остаточной суммой квадратов отклонений, а сумма квадратов, которая была остаточной суммой в двухфакторной модели, обратилась в ноль.

Сумма квадратов $\sum_{i=1}^{k} n(\bar{y}_{i*} - \bar{y})^2$ соответствует эффекту фактора $W$, а сумма квадратов $\sum_{j=1}^{n} k(\bar{y}_{*j} - \bar{y})^2$ соответствует эффекту индивидуальных различий. Таким образом, имеем (10.24 и 10.25):

$$SS_W = \sum_{i=1}^{k} n(\bar{y}_{i*} - \bar{y})^2, \ \nu_W = k-1 \qquad (10.24)$$

$$SS_R = \sum_{i=1}^{k}\sum_{j=1}^{n} (y_{ij} - \bar{y}_{*j} - \bar{y}_{i*} + \bar{y})^2, \ \nu_W = (n-1)(k-1) \qquad (10.25)$$

Таким образом, для расчета статистики можно не осуществлять переход к центрированным по наблюдениям данным, однако, необходимо грамотно определить вид рассчитываемой остаточной суммы квадратов.

Если предположить, что дополнительно все $n$ объектов разбиты на $m$ групп по $p_j$ объектов в каждой, $j = \overline{1,m}$ ($\sum_{j=1}^{m} p_j = n$), соответствующих различным уровням фактора $A$, то на результаты наблюдений могут оказывать влияние фактор $W$ повторных наблюдений, фактор $A$, воздействующий на слои объектов, и фактор индивидуальных различий каждого объекта. Соответственно, модель порождения данных принимает вид (10.26):

$$y_{ijl} = \mu_{ijl} + \varepsilon_{ijl} = \mu + w_i + \alpha_j + \tau_{jl} + \beta_{ij} + \varepsilon_{ijl} \quad i = \overline{1,k}, \quad j = \overline{1,m}, \quad l = \overline{1,p_j}$$
,(10.26)

где: $y_{ijl}$ - $l$-ое наблюдаемое значение отклика в $i$-ой группе в $j$-ом слое;

$\mu_{ijl} = \mu + w_i + \alpha_j + \tau_{jl} + \beta_{ij}$ - среднее для $l$-го наблюдаемого значения отклика в $i$-ой группе в $j$-ом слое;

$\mu$ - среднее значение отклика;

$w_i$ - эффект $i$-го уровня фактора $W$;

$\alpha_j$ - эффект $j$-го уровня фактора $A$;

$\tau_{jl}$ - эффект индивидуальности $l$-го наблюдения в $j$-ом слое;

$\beta_{ij}$ - эффект взаимодействия $i$-го уровня фактора $W$ и $j$-го уровня фактора $A$;

$\varepsilon_{ijl}$ - независимые случайные величины с математическим ожиданием равным нулю и одинаковой дисперсией $\sigma^2$.

Для того, чтобы существовала возможность однозначно определять параметры модели, необходимо на параметры модели наложить дополнительные условия. В качестве этих условий примем:

$$\sum_{i=1}^{k} w_i = 0,$$

$$\sum_{j=1}^{m} \alpha_j = 0,$$

$$\sum_{i=1}^{k} \beta_{ij} = 0,\ j = \overline{1, m-1},$$

$$\sum_{j=1}^{m} \beta_{ij} = 0, i = \overline{1, k-1},\ \sum_{i=1}^{k}\sum_{j=1}^{m} \beta_{ij} = 0.$$

Обозначим: $\mu_{*jl} = \frac{1}{k}\sum_{i=1}^{k} \mu_{ijl} = \mu + \alpha_j + \tau_{jl}$ - среднее для $l$-го объекта наблюдения в $j$-ом слое. Таким образом, уравнение примет вид (10.27):

$$y_{ijl} - \mu_{*jl} = w_i + \beta_{ij} + \varepsilon_{ijl} \quad i = \overline{1,k}, \quad j = \overline{1,n}, \quad l = \overline{1, p}_j. \tag{10.27}$$

Таким образом, на величину $y_{ijl} - \mu_{*jl}$ могут оказывать влияние только главный эффект фактора $W$ и эффект взаимодействия факторов $W$ и $A$. Соответствующее (6) разложение суммы квадратов в случае равного числа наблюдений $p_j$ для каждого слоя имеет вид (10.28):

$$\sum_{i=1}^{k}\sum_{j=1}^{m}\sum_{l=1}^{p_j}(y_{ijl} - \bar{y}_{*jl})^2 = \sum_{i=1}^{k} n(\bar{y}_{i**} - \bar{y})^2 + \sum_{i=1}^{k}\sum_{j=1}^{m} p_j(\bar{y}_{ij*} - \bar{y}_{*j*} - \bar{y}_{i**} + \bar{y})^2 +$$
$$+ \sum_{i=1}^{k}\sum_{j=1}^{m}\sum_{l=1}^{p_j}(y_{ijl} - \bar{y}_{*jl} - (\bar{y}_{ij*} - \bar{y}_{*j*}))^2 \tag{10.28}$$

Сумма квадратов $SS_W = \sum_{i=1}^{k} n(\bar{y}_{i**} - \bar{y})^2$ соответствует эффекту фактора $W$, сумма квадратов $SS_{WA} = \sum_{i=1}^{k}\sum_{j=1}^{m} p_j(\bar{y}_{ij*} - \bar{y}_{*j*} - \bar{y}_{i**} + \bar{y})^2$ соответствует эффекту взаимодействия факторов $W$ и $A$, а сумма квадратов $SS_R=\sum_{i=1}^{k}\sum_{j=1}^{m}\sum_{l=1}^{p_j}(y_{ijl}-\bar{y}_{*jl}-(\bar{y}_{ij*}-\bar{y}_{*j*}))^2$ является остаточной суммой квадратов. Числа степеней свободы соответствующих сумм квадратов равны:

$\nu_W=k-1;$

$\nu_{WA}=(k-1)(m-1);$

$\nu_R=kn-(n-m)-(km-m)-m=(k-1)(n-m)$ (если заданы средние $\bar{y}_{*j*}$, $j=\overline{1,m}$, то независимо можно определить только $km-m$ величин $\bar{y}_{ij*}$ и $n-m$ величин $\bar{y}_{*jl}$), $\nu_W=k-1$, $\nu_{WA}=(k-1)(m-1)$ .

Соответственно, средние суммы квадратов отклонений будут равны (10.29):

$$MS_W=\frac{SS_W}{k-1},\ MS_{WA}=\frac{SS_{WA}}{(k-1)(m-1)},\ MS_R=\frac{SS_R}{(k-1)(n-m)} \qquad (10.29)$$

F-статистика для проверки гипотезы $H_0: w_1=w_2=\ldots=w_k=0$ будет иметь вид: $F=\frac{MS_W}{MS_R}$. Если наблюдаемое значение статистики $F_{набл}\geq F_{кр}$, где $F_{кр}$ - критическая точка распределения Фишера уровня $\alpha$ (или квантиль уровня $1-\alpha$) с числом степеней свободы $\nu_1=k-1$ и $\nu_2=m(k-1)(p-1)$, то нулевая гипотеза отклоняется и считается, что эффект фактора *W* значим.

F-статистика для проверки гипотезы $H_0:\beta_{11}=\beta_{12}=\ldots=\beta_{km}=0$ соответственно будет иметь вид: $F=\frac{MS_{WA}}{MS_R}$. Если наблюдаемое значение статистики $F_{набл}\geq F_{кр}$, где $F_{кр}$ - критическая точка распределения Фишера уровня $\alpha$ (или квантиль уровня $1-\alpha$) с числом степеней свободы $\nu_1=(k-1)(m-1)$ и $\nu_2=(k-1)(n-m)$, то нулевая гипотеза отклоняется и считается, что эффект взаимодействия факторов *W* и *A* значим.

На основе разложения невозможно построить критерий для проверки гипотезы $H_0:\alpha_1=\alpha_2=\ldots=\alpha_k=0$. Для этого необходимо использовать разложение суммы квадратов отклонений, которая может быть получена из исходной (соответствующей выражению (5)) за вычетом суммы квадратов отклонений, соответствующей выражению

(6). То есть, для проверки гипотезы $H_0:\alpha_1=\alpha_2=\ldots=\alpha_k=0$ используем разложение, которое соответствует уравнениям (10.30):

$$\mu_{*jl}-\mu=\alpha_j+\tau_{jl} \quad i=\overline{1,k}, \quad j=\overline{1,n}, \quad l=\overline{1,p_j}. \tag{10.30}$$

Для модели разложение суммы квадратов имеет вид (10.31):

$$k\sum_{j=1}^{m}\sum_{l=1}^{p_j}(\bar{y}_{*jl}-\bar{y})^2=k\sum_{j=1}^{m}p_j(\bar{y}_{*j*}-\bar{y})^2+k\sum_{j=1}^{m}\sum_{l=1}^{p_j}(\bar{y}_{*jl}-\bar{y}_{*j*})^2 \tag{10.31}$$

где:

$SS_A=k\sum_{j=1}^{m}p_j(\bar{y}_{*j*}-\bar{y})^2$ - сумма квадратов, соответствующая эффекту фактора *A;*

$SS_{R_1}=k\sum_{i=1}^{k}\sum_{j=1}^{m}(\bar{y}_{*jl}-\bar{y}_{*j*})^2$ - остаточная сумма квадратов отклонений, которая есть не что иное, как сумма квадратов эффекта индивидуальных различий. Число степеней свободы для суммы $SS_A$: $\nu_A=m-1$, для остаточной суммы: $\nu_{R_1}=n-m$. Таким образом, F-отношение для проверки гипотезы $H_0:\alpha_1=\alpha_2=\ldots=\alpha_k=0$ будет иметь вид: $F=\frac{MS_A}{MS_{R_1}}=\frac{n-m}{m-1}\frac{SS_A}{SS_{R_1}}$. Если наблюдаемое значение статистики $F_{набл}\geq F_{кр}$, где $F_{кр}$ - критическая точка распределения Фишера уровня $\alpha$ (или квантиль уровня $1-\alpha$) с числом степеней свободы $\nu_1=m-1$ и $\nu_2=n-m$, то нулевая гипотеза отклоняется и считается, что эффект фактора *A* значим.

Таким образом, в дисперсионном анализе повторных измерений, по сути, используются две модели.

Первая модель в качестве источника данных использует центрированные по объектам наблюдений данные и исследует различия в повторных наблюдениях и соответствующие эффекты взаимодействия.

Вторая модель в качестве источника данных использует средние по объектам наблюдения данные и исследует прочие эффекты. Если число наблюдений в слоях различное, то также как и в многофакторной модели дисперсионного анализа, следует использовать разложения сумм квадратов различных типов. Используемый тип разложения определяется исследователем, исходя из конкретной ситуации. Для большинства приложений дисперсионного анализа рекомендуется использовать III тип разложения.

Непараметрическим аналогом дисперсионного анализа повторных измерений является тест Фридмана. В тесте Фридмана предварительно ранжируются наблюдения для каждого из объектов наблюдения. Пусть $r_{ij}$, $i=\overline{1,k}$, $j=\overline{1,n}$ - ранг $j$-го наблюдения в $i$-ой группе. Обозначим: $\overline{R}_i=\sum_{j=1}^{n} r_{ij}$ – средний ранг для i-ой группы, $\overline{R}=\dfrac{k+1}{2}$ - средний ранг всех наблюдений. Тогда выборочная дисперсия для средних рангов групп будет равна (10.32):

$$D=\frac{1}{(k-1)}\sum_{i=1}^{k}(\overline{R}_i-\overline{R})^2 \qquad (10.32)$$

Максимальное значение дисперсии достигается, когда все наблюдения проранжированы согласованно, в этом случае средние ранги принимают целые значения от 1 до $k$. Соответствующее наибольшее значение дисперсии при этом равно (10.33):

$$D_{\max}=\frac{k(k+1)}{12} \qquad (10.33)$$

Отношение $W=\dfrac{D}{D_{\max}}=\dfrac{12}{k^3-k}\sum_{i=1}^{k}(\overline{R}_i-\overline{R})^2$ называется коэффициентом конкордации Кендалла, оно является характеристикой согласованности рангов наблюдений. Значения коэффициента конкордации меняются от 0 до 1, значение 1 соответствует, что все наблюдения проранжированы одинаково, то есть ранги для всех групп максимально различаются. Значение коэффициента конкордации равное нулю соответствует ситуации, когда средние ранги всех групп совпадают. С коэффициентом конкордации связана статистика Фридмана, определяемая по формуле 10.34:

$$S=n(k-1)W=\frac{12n}{k(k+1)}\sum_{i=1}^{k}(\overline{R}_i-\overline{R})^2 \qquad (10.34)$$

При условии истинности нулевой гипотезы (средние ранги по группам не различаются) статистика $S$ имеет распределение Фридмана. При больших $k,n$ статистика Фридмана приближенно имеет распределение Хи-квадрат с $k-1$ степенью свободы. Если наблюдаемое значение статистики $S_{набл}\geq S_{кр}$, где $S_{кр}$ - критическая точка распределения Хи-квадрат с числом степеней свободы $k-1$ уровня $\alpha$(или квантиль уровня $1-\alpha$), то нулевая гипотеза отклоняется и считается, что средние ранги для различных уровней повторных измерений значимо различаются.

Для двух зависимых групп вместо теста Фридмана следует использовать тест Уилкоксона для зависимых выборок (аналог теста Манна-Уитни для независимых выборок).

### 10.5. Апостериорные множественные сравнения средних

Если в результате дисперсионного анализа получается, что средние значения отклика для разных уровней фактора, различаются, данный результат не является окончательным.

В дальнейшем подразумевается анализ того, для каких уровней фактора средние больше, для каких меньше, а для каких одинаковы. Основная процедура дисперсионного анализа не дает возможности ответить на эти вопросы.

Наиболее простым способом решения данной проблемы является проведение серии попарных сравнений с использованием t-критерия, используя в качестве оценки дисперсии величину $MS_R$ - оценку внутригрупповой дисперсии, полученную в ходе дисперсионного анализа.

Такой подход реализуется в так называемом ***методе наименьшей*** значимой разности (LSD). Статистика критерия LSD для проверки гипотезы равенства средних $\mu_i$ и $\mu_j$ имеет вид (10.35):

$$t=\frac{\overline{\mu}_i-\overline{\mu}_j}{\sqrt{MS_R(1/n_i+1/n_j)}} \tag{10.35}$$

Если наблюдаемое значение статистики $|t_{набл}|\geq t_{кр}$, где $t_{кр}$ - критическая точка распределения Стьюдента уровня $\alpha/2$ (или квантиль уровня $1-\alpha/2$) с числом степеней свободы $\nu=n-k$, то нулевая гипотеза отклоняется и принимается гипотеза $H_1:\mu_1\neq\mu_2$.

Существуют разные подходы к решению данной проблемы. Один из них – уменьшить уровень значимости при попарном сравнении так, чтобы вероятность хотя бы одного отклонения нулевой гипотезы равнялось заданному уровню значимости.

Такой подход реализуется в принципе Бонферрони множественных сравнений, в котором при каждом попарном сравнении задается уровень значимости $\alpha / C_k^2$, где $C_k^2$ - число сравнений.

Данная величина гарантирует, что вероятность отклонение нулевой гипотезы (при ее истинности) хотя бы в одном из $C_k^2$ сравнений не превзойдет $\alpha$. Однако, принцип Бонферрони является чересчур консервативным, он приводит к существенному снижению мощности критерия.

LSD – критерий и критерий Бонферрони занимают полярные позиции в ряду критериев множественных сравнений. Среди остальных критериев множественного сравнения средних можно выделить критерии множественных сравнений Шеффе, Ньюмена-Келса, Тьюки и другие.

В методе множественных сравнений Шеффе для проверки гипотезы равенства средних $\mu_i$ и $\mu_j$ используется статистика (10.36):

$$F = \frac{(\bar{\mu}_i - \bar{\mu}_j)^2}{(k-1) MS_R (1/n_i + 1/n_j)} \tag{10.36}$$

где $MS_R$ – оценка внутригрупповой (остаточной) дисперсии, полученная в ходе дисперсионного анализа. Если наблюдаемое значение статистики $F_{набл} \geq F_{кр}$, где $F_{кр}$ - критическая точка распределения Фишера уровня $\alpha$ (или квантиль уровня $1-\alpha$) с числом степеней свободы $\nu_1 = k-1$ и $\nu_2 = n-k$, то нулевая гипотеза отклоняется и принимается гипотеза $H_1 : \mu_i \neq \mu_j$.

Критерий Шеффе также относится к достаточно консервативным критериям, то есть обладает малой мощностью. Более мощными, соответственно, более чувствительными являются критерии Тьюки, Ньюмена-Келса, Дункана.

В методе множественных сравнений Тьюки (или достоверно значимой разности – HSD) для проверки гипотезы $H_0 : \mu_i = \mu_j$ против альтернативы $H_1 : \mu_i \neq \mu_j$ используется статистика (10.37):

$$t_R = \frac{|\bar{\mu}_i - \bar{\mu}_j|}{\sqrt{MS_R (1/n_i + 1/n_j)/2}} \tag{10.37}$$

Ее значения сравниваются с критическими точками уровня $\alpha$ распределения стьюдентизированного размаха с $\nu_1 = k$ и $\nu_2 = n - k$ степенями свободы.

Если наблюдаемое значение статистики $t_{R\,набл} \geq t_{R\,кр}$, где $t_{R\,кр}$ - критическая точка распределения стьюдентизированного размаха уровня $\alpha$ (или квантиль уровня $1-\alpha$) с числом степеней свободы $\nu_1 = k$ и $\nu_2 = n - k$, то нулевая гипотеза отклоняется и принимается гипотеза $H_1 : \mu_i \neq \mu_j$.

Если объемы выборок различаются сильно, то рекомендуется использовать HSD критерий Тьюки для неравных выборок (критерий Spjovoll-Stoline). Статистика критерия в этом случае имеет вид (10.38)

$$t_R = \frac{\left|\overline{\mu}_i - \overline{\mu}_j\right|}{\sqrt{MS_R / \min(n_i, n_j)}} \tag{10.38}$$

Критические точки определяются также, как и для критерия HSD Тьюки.

В ***критерии Ньюмана-Келса*** используется та же статистика, что и в критерии Тьюки, однако по другому определяются критические точки. В качестве критических точек критерия Ньюмана-Келса используются критические точки распределения стьюдентизированного размаха с $\nu_1 = r$ и $\nu_2 = n - k$ степенями свободы, где $r$ - число средних расположенных между $\overline{\mu}_i$ и $\overline{\mu}_j$ в вариационном ряду выборочных средних, включая $\overline{\mu}_i$ и $\overline{\mu}_j$. Например, если сравниваются значения $\overline{\mu}_{(i)}$ и $\overline{\mu}_{(i+1)}$ вариационного (упорядоченного) ряда средних, то $r = 2$, если сравниваются значения $\overline{\mu}_{(i)}$ и $\overline{\mu}_{(i+2)}$, то $r = 3$ и так далее.

В пакете STATISTICA используется модифицированный вариант критерия Ньюмана-Келса, в котором в качестве статистики критерия используется величина (10.39):

$$t_R = \frac{\left|\overline{\mu}_i - \overline{\mu}_j\right|}{\sqrt{MS_R \frac{1}{k} \sum_{l=1}^{k} \frac{1}{n_l}}} \tag{10.39}$$

Аналогичная статистика используется и в *критерии Дункана*, но в качестве критических точек берутся точки D-распределения Дункана с $\nu_1 = r$ и $\nu_2 = n - k$ степенями свободы, где $r$ - число средних расположенных между $\overline{\mu}_i$ и $\overline{\mu}_j$ в вариационном ряду выборочных средних, включая $\overline{\mu}_i$ и $\overline{\mu}_j$.

Методы множественного сравнения средних можно использовать не только для проверки гипотез о попарном различии средних, а также для проверки гипотез о различии средних для любых выбранных наборов групп. В силу этого, основная гипотеза в данных методах в общем случае имеет вид: $H_0 : \sum_{i=1}^{k} c_i \mu_i = 0$, где $c_i, i = \overline{1,k}$ некоторые заданные константы, удовлетворяющие условию $\sum_{i=1}^{k} c_i = 0$. Например, при $c_3 = c_4 = \ldots = c_k = 0, c_1 = 1, c_2 = -1$, мы будем проверять гипотезу $H_0 : \mu_1 - \mu_2 = 0$ или $\mu_1 = \mu_2$. При $c_1 = 1, c_2 = -1/2, c_3 = -1/2$, $c_4 = c_5 = \ldots = c_k = 0$, будем проверять гипотезу $H_0 : \mu_1 = \frac{1}{2}(\mu_2 + \mu_3)$, то есть, гипотезу однородности первой и совокупности второй и третьей групп и т.д. Линейные комбинации вида: $\alpha(\mu_1 - \mu_2)$, $\alpha(\mu_1 - \frac{1}{2}(\mu_2 + \mu_3))$, то есть величины, пропорциональные разности между средними от средних, называются контрастами.

Критерии LSD, Шеффе, HSD Тьюки легко модифицировать под проверку гипотезы $H_0 : \sum_{i=1}^{k} c_i \mu_i = 0$. Например, статистика LSD критерия для проверки гипотезы $H_0 : \sum_{i=1}^{k} c_i \mu_i = 0$ будет иметь вид (10.40):

$$t = \frac{\sum_{i=1}^{k} c_i \overline{\mu}_i}{\sqrt{MS_R \sum_{i=1}^{k} (c_i^2 / n_i)}} \tag{10.40}$$

Критическими точками статистики, по прежнему, будут являться квантили распределения Стьюдента уровня $1 - \alpha/2$ с числом степеней свободы $\nu = n - k$.

## Контрольные вопросы

1. В чем состоит сущность дисперсионного анализа?
2. Что называется фактором в рамках осуществления дисперсионного анализа
3. В каком случае дисперсионный анализ является однофакторным?
4. Приведите пример дисперсионного анализа в статистике фирмы
5. Опишите модель однофакторного дисперсионного анализа
6. По какой формуле вычисляется сумма квадратов отклонений групповых средних от общего среднего?
7. В чем состоит общая идея дисперсионного анализа?
8. Что называется остаточным эффектом (эффектом ошибок) дисперсионного анализа?
9. Врно ли утверждение о том, что при осуществлении дисперсионного анализа анализируются не сами значения суммы квадратов отклонений, а средние квадраты? Объясните, почему.
10. Каким образом рассчитывается число степеней свободы для суммы квадратов случайных величин?
11. В чем заключается сущность проверки нулевой гипотезы?
12. С помощью какого сравнения нулевая гипотеза отклоняется и считается, что средние для различных уровней фактора значимо различаются?
13. Что является непараметрической альтернативой однофакторного дисперсионного анализа?
14. Что является основой дисперсионного анализа Краскела–Уоллиса?
15. Для каких данных применяется анализа Краскела–Уоллиса?
16. Чему приближенно равна статистика при принятии нулевой гипотезы о равенстве средних рангов?
17. При соблюдении какого неравенства считается, что средние ранги для различных уровней фактора значимо различаются?
18. В каком случае имеет место многофакторный дисперионный анализ данных?
19. Приведите пример двухфакторной дисперсионной модели в статистике фирмы?
20. В каком виде может быть представлена дисперсия отклика?
21. Когда эксперимент принято считать сбалансированным?

22. Какие МНКоценки получает исследователь в случае пропорциональных частот?
23. Зачем необходимо осуществлять проверку гипотез значимости различных выявленных эффектов с помощью инструментов статистического анализа?
24. Что можно использовать в качестве статистик критериев проверки гипотез о значимости соответствующих эффектов?
25. Какое распределение имеют данные статистики при нормальном распределении остатков?
26. Что означает нарушение условия ортогональности сумм квадратов эффектов?
27. Возможна ли независимая одновременная проверка всех гипотез двухфакторного анализа? Почему?
28. Объясните процедуру последовательного разложения суммы квадратов на составляющие.
29. В зависимости от какого условия различаются статистики критериев для проверки гипотез о значимости эффектов в несбалансированной многофакторной модели дисперсионного анализа?
30. Какие основные типы разложения суммы квадратов в рамках осуществления дисперсионного анализа Вам известны?
31. В чем состоит проблема соответствия различным уровням фактора одних и тех же объектов? Приведите примеры
32. С какой целью используется дисперсионный анализ зависимых выборок (повторных наблюдений)?
33. К какой процедуре прибегает исследователь длч того, чтобы исключить эффект индивидуальности наблюдений и анализировать только влияние фактора на результаты наблюдений?
34. Обязательно ли в рамках дисперсионного анализа осуществлять переход к центрированным по наблюдениям данным?
35. Какие две основные модели используются в дисперсионном анализе повторных измерений?
36. Что является непараметрическим аналогом дисперсионного анализа повторных измерений?
37. Перечислите основные этапы проведения теста Фридмана.
38. В каком случае достигается максимальное значение дисперсии?
39. Для чего в рамках дисперионного анализа используются апостериорные множественные сравнения средних?

## МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ ПО ВЫПОЛНЕНИЮ КУРСОВЫХ РАБОТ (ПРОЕКТОВ)

### 1. ОБЩИЕ ПОЛОЖЕНИЯ

Курсовая работа (проект) представляет собой вид учебной и научно- исследовательской работы студента, является индивидуальным, завершенным трудом, отражающим знания, навыки и умения студента, полученные в ходе освоения дисциплины.

Тема курсовой работы (проекта) не может носить описательного характера, в ее формулировку должна быть заложена исследовательская проблема. Курсовая работа (проект) подготавливает студента к выполнению более сложной задачи – написанию выпускной квалификационной или дипломной работ.

Рациональнее темы курсовых работ (проектов), выполняемых студентами за весь период обучения, подбирать таким образом, чтобы они вместе с выпускной работой составляли единую систему последовательно усложняемых и взаимосвязанных работ. При защите работы студент учится не только правильно излагать свои мысли, но и аргументированно отстаивать, защищать выдвигаемые выводы и решения. Формулировка темы должна быть по возможности краткой и соответствовать содержанию работы.

### 2. ЦЕЛИ И ЗАДАЧИ ВЫПОЛНЕНИЯ КУРСОВЫХ РАБОТ (ПРОЕКТОВ)

Основной целью выполнения курсовой работы (проекта) является развитие мышления, творческих способностей студента, привитие навыков самостоятельной работы, связанной с поиском, систематизацией и обобщением научной и учебной литературы, углублённым изучением определенного вопроса, темы, раздела учебной дисциплины, формирование умений анализировать и критически оценивать исследуемый научный и практический материал, овладение методами современных научных исследований.

Курсовая работа (проект) представляет собой:

- изложение результатов исследования с учетом вопросов теории и практики в проделах выбранной темы;

- авторский труд, самостоятельное творчество студента, формирование его личной позиции и практического подхода к выбранной теме;

- отражение умения студентом логично, аргументировано, ясно, последовательно и кратко излагать свои мысли.

Основные отличия курсовой работы (проекта) от контрольной работы:

- курсовая работа требует более глубокого анализа проблемы, поэтому её минимальный требуемый объем значительно больше,

- обязательно включает практический раздел, направленный на отработку факто- логического материала, в курсовой работе должно найти отражение взаимосвязи теоретических положений с практикой;

- контроль за ходом написания курсовой работы осуществляется кафедрой.

Научно-консультационную и методическую помощь студенту оказывает научный руководитель. Работа над избранной темой требует от студента знаний основ методологии исследования, творческого мышления, прилежания и профессионализма.

Задачами выполнения курсовых работ (проектов) являются:

- систематизация, закрепление, углубление и расширение приобретенных студентом знаний, умений, навыков по учебным дисциплинам профессиональной подготовки;
- овладение методами научных исследований;
- формирование навыков решения творческих задач в ходе научного исследования, художественного творчества или проектирования по определенной теме;
- овладение современными методами поиска, обработки и использования информации.
- подготовка к написанию дипломной работы (материалы курсовых работ могут входить в дипломную работу).

При выполнении курсовых работ (проектов) студент должен продемонстрировать способности:

– выдвинуть научную (рабочую) гипотезу;
– собрать и обработать информацию по теме;
– изучить и критически проанализировать полученные материалы;
– систематизировать и обобщить имеющуюся информацию;

– самостоятельно решить поставленные творческие задачи;
– логически обосновать и сформулировать выводы, предложения и рекомендации.

Особенности курсовых работ (проектов) в зависимости от курса обучения проявляются в постепенном усложнении объектов и методов исследования (проектирования).

Количество курсовых работ (проектов), наименование дисциплин, по которым они предусматриваются, определяется учебным планом. Общее число курсовых работ (проектов) по дисциплинам учебного плана не может превышать 5-6 на весь период обучения, если иное не предусматривается государственным образовательным стандартом и примерным учебным планом по соответствующей специальности (направлению). Курсовая работа (проект) рассматривается как вид учебной работы по дисциплине и выполняется в пределах часов, отводимых на ее изучение. Курсовые работы (проекты) рассматриваются как форма отчетности.

Полные названия курсовых работ (проектов) вносятся в экзаменационные ведомости, зачетные книжки студентов и в приложения к дипломам.

Согласно номенклатуре дел курсовые работы (проекты) учитываются и хранятся на кафедре в течение пяти лет. По истечении указанного срока все курсовые работы (проекты), не представляющие учебно-методической ценности, списываются по акту и уничтожаются.

## 3. СТРУКТУРА И ЭТАПЫ ВЫПОЛНЕНИЯ КР (КП)

Тематику курсовых работ (проектов) разрабатывает кафедра в учебном году, предшествующем выполнению курсовой работы (проекта).

Выбор и утверждение темы курсовой работы (проекта) происходит в следующем порядке:

- тематика курсовых работ (проектов) сообщается студентам;
- студент может выбрать тему курсовой работы (проекта) из числа тем, предложенных кафедрой;
- студент может также самостоятельно предложить тему курсовой работы (проекта) с обоснованием ее целесообразности;

- тематика курсовых работ (проектов) на предстоящий учебный год утверждается на заседании кафедры, о чем в протоколе заседания делается соответствующая запись.

Студент выполняет курсовую работу (проект) по утвержденной теме под руководством преподавателя, являющегося его научным руководителем.

Темы курсовых работ (проектов) утверждаются на заседании кафедры и подтверждаются соответствующими заявлениями студентов о выборе темы.

Руководителем курсовой работы (проекта) по дисциплине учебного плана является, как правило, лектор, ведущий данную дисциплину, преподаватель, ведущий практические занятия. Руководителем курсовой работы (проекта) по специальным дисциплинам, дисциплинам специализации может быть назначен приглашенный специалист, выполняющий учебную нагрузку на условиях почасовой оплаты на условиях почасовой оплаты.

Научный руководитель составляет задание на курсовую работу (проект), осуществляет ее текущее руководство. Текущее руководство курсовой работой (проектом) включает систематические консультации с целью оказания организационной и научно-методической помощи студенту, контроль за осуществлением выполнения работы в соответствии с планом – графиком, проверку содержания и оформления завершенной работы.

Завершенная курсовая работа (проект) передается студентом на кафедру за неделю до защиты для ее анализа.

Написание работы - процесс, включающий в себя ряд взаимосвязанных этапов:

1. Выбор темы. Рекомендованная тематика курсовых работ содержится в рабочих программах дисциплин, по которым формой промежуточной аттестации является курсовая работа (проект). При выборе темы курсовой работы (проекта) можно рекомендовать студенту четко определить круг своих интересов и выполнять весь комплекс курсовых работ (в рамках соответствующих учебных дисциплин) по одной проблематике. Это позволит существенно повысить качество выполняемых курсовых работ (проектов) и даст возможность студенту лучше подготовиться к выполнению выпускной квалификационной работы.

2. Разработка структуры и оформление содержания. Структура работы должна быть согласована с научным руководителем.

3. Сбор, анализ и обобщение материалов исследования, написание текста работы:

- сбор материалов, необходимых для выполнения курсовой работы (проекта), посредством использования литературных источников, нормативных актов, директивных документов и документации предприятия (организации) по рассматриваемой в работе проблематике;
- систематизация и обработка собранного материала по каждому из разрабатываемых в курсовой работе (проекту) вопросу или проблеме. На базе систематизированного материала формируются основные направления анализа. Одновременно выясняется необходимость сбора дополнительной информации по отдельному вопросу или вопросам;
- сбор дополнительной информации и разработка аналитической части курсовой работы (проекта). На этом этапе выявляются негативные моменты и недостатки функционирования объекта исследования;
- разработка и обоснование предложений по основным направлениям деятельности объекта исследования. На основе разработанных предложений и рекомендаций формулируются соответствующие выводы

4. Оформление работы и её представление для проверки.

5. Защита курсовой работы. Работа предоставляется на кафедру (руководителю) заранее, не позднее, чем за 10 дней до защиты.

Методологической основой курсовой работы (проекта) являются законодательные акты Российской Федерации по экономике, в целом, и по изучаемой дисциплине, в частности, программные документы и решения правительства РФ по хозяйственным вопросам.

По выбранной теме курсовой работы (проекта) рекомендуется использовать данные Росстата, материалы Института исследования товародвижения и конъюнктуры оптового рынка (ИТКОР), учебную специальную литературу, монографии, брошюры, статьи. Целесообразно изучить зарубежный опыт применительно к рассматриваемой теме. Важным условием успешного раскрытия избранной темы является ознакомление с материалами, опубликованными в периодических изданиях и др.

Желательно, чтобы курсовой проект выполнялся на материалах предприятия или организации по месту работы студентов заочной формы обучения или по месту прохождения производственной практики студентов очной формы обучения. В качестве основы написания курсовой работы (проекта) могут быть использованы материалы, собранные для курсовых работ по смежным дисциплинам, изученным ранее, а также материалы, собранные в ходе учебной и производственной практик

## 4. ФОРМЫ И ПОРЯДОК АТТЕСТАЦИИ КУРСОВЫХ РАБОТ (ПРОЕКТОВ)

Формами аттестации студента по результатам выполнения курсовой работы являются зачёт (зачтено/не зачтено), а по результатам курсового проекта дифференцированный зачёт ("отлично" - "хорошо" - "удовлетворительно" - "неудовлетворительно"). Форма аттестации по курсовым работам (проектам) по дисциплинам учебного плана вносится в рабочий учебный план специальности (направления) и утверждается Ученым советом института.

Аттестация всех курсовых работ (проектов) должна быть проведена до начала экзаменационной сессии, в сроки, указанные рабочим учебным планом специальности (направления).

Аттестация по курсовым работам (проектам) производится в виде ее защиты перед группой и научным руководителем работы (проекта).

Решение об оценке курсовой работы (проекта) принимается преподавателем по результатам трех рейтингов, проводимых в течение семестра, для которых деканатом выдается отдельная ведомость, аналогичная ведомости текущего рейтинг-контроля, а также по итогам анализа предъявленной курсовой работы (проекта), доклада студента и его ответов на вопросы. Оценка по курсовой работе (проекту) вносится в экзаменационную ведомость, зачетную книжку студента научным руководителем.

Студент, по неуважительной причине не предоставивший в установленный срок или не защитивший курсовую работу (проект), считается имеющим академическую задолженность. Научный руководитель курсовой работы (проекта) проставляет в экзаменационную ведомость неудовлетворительную оценку. В случае наличия уважительных причин, подтвержденных документально, распоряжением по институту (факультету) студенту устанавливаются индивидуальный порядок и сроки выполнения и защиты курсовой работы (проекта). Курсовая работа, оцененная неудовлетворительно перерабатывается студентом и возвращается на проверку тому же преподавателю.

Критериями оценки курсовой работы являются*:*

– актуальность и степень разработанности темы;
– творческий подход и самостоятельность в анализе, обобщениях и выводах;
– полнота охвата первоисточников и исследовательской литературы;
– уровень овладения методикой исследования;
– научная обоснованность и аргументированность обобщений, выводов и рекомендаций;
– научный стиль изложения;
– соблюдение всех требований к оформлению курсовой работы и сроков ее исполнения.

## 5. ТРЕБОВАНИЯ К СОДЕРЖАНИЮ КР (КП)

Курсовая работа (проект) имеет ряд структурных элементов: введение, теоретическая часть, практическая часть, заключение.

Разработка введения. Во-первых, во введении следует обосновать актуальность избранной темы курсовой работы (проекта), раскрыть ее теоретическую и практическую значимость, сформулировать цели и задачи работы.

Во-вторых, во введении, а также в той части работы, где рассматривается теоретический аспект данной проблемы, автор должен дать, хотя бы кратко, обзор литературы, изданной по этой теме.

Введение должно подготовить читателя к восприятию основного текста работы. Оно состоит из обязательных элементов, которые необходимо правильно сформулировать. В первом предложении называется тема курсовой работы.

Актуальность исследования (почему это следует изучать?). Актуальность исследования рассматривается с позиций социальной и практической значимости. В данном пункте необходимо раскрыть суть исследуемой проблемы и показать степень ее проработанности в различных трудах. Здесь же можно перечислить источники информации, используемые для исследования (Информационная база исследования может быть вынесена в первую главу).

Цель исследования (какой результат будет получен?). Цель должна заключаться в решении исследуемой проблемы путем ее анализа и практической реализации. Цель всегда направлена на объект.

Объект исследования (что будет исследоваться?). Объект предполагает работу с понятиями. В данном пункте дается определение экономическому явлению, на которое направлена исследовательская деятельность. Объектом может быть личность, среда, процесс, структура, хозяйственная деятельность предприятия (организации).

Предмет исследования (как, через что будет идти поиск?). Здесь необходимо дать определение планируемым к исследованию конкретным свойствам объекта или способам изучения экономического явления. Предмет исследования направлен на практическую деятельность и отражается через результаты этих действий.

Задачи исследования (как идти к результату?), пути достижения цели. Задачи соотносятся с гипотезой. Определяются они исходя из целей работы. Формулировки задач необходимо делать как можно более тщательно, поскольку описание их решения должно составить содержание глав и параграфов работы. Как правило, формулируются 3-4 задачи.

Примерный перечень рекомендуемых задач:

1. «На основе теоретического анализа литературы разработать...» (ключевые понятия, основные концепции).
2. «Определить... » (выделить основные условия, факторы, причины, влияющие на объект исследования).
3. «Раскрыть... » (выделить основные условия, факторы, причины, влияющие на предмет исследования).
4. «Разработать... » (средства, условия, формы, программы).
5. «Апробировать…» (что разработали) и дать рекомендации...

Методы исследования (как исследовали?): дается краткое перечисление методов исследования через запятую без обоснования.

Структура работы – это завершающая часть введения (что в итоге в работе/проекте представлено).

В завершающей части в назывном порядке перечисляются структурные части работы (проекта), например: «Структура работы соответствует логике исследования и включает в себя введение, теоретическую часть, практическую часть, заключение, список литературы, 5 приложений».

Здесь допустимо дать развернутую структуру курсовой работы (проекта) и кратко изложить содержание глав. (Чаще содержание глав курсовой работы излагается в заключении).

Таким образом, введение должно подготовить к восприятию основного текста работы.

Краткие комментарии по формулированию элементов введения представлены в таблице 1.

Таблица 1 – Комментарии по формулированию элементов введения

| Элемент введения | Комментарий к формулировке |
|---|---|
| Актуальность темы | Почему это следует изучать?<br>Раскрыть суть исследуемой проблемы и показать степень ее проработанности. |
| Цель исследования | Какой результат будет получен?<br>Должна заключаться в решении исследуемой проблемы путем ее анализа и практической реализации. |
| Объект исследования | Что будет исследоваться?<br>Дать определение явлению или проблеме, на которое направлена исследовательская деятельность. |
| Предмет исследования | Как и через что будет идти поиск?<br>Дать определение планируемым к исследованию конкретным свойствам объекта или способам изучения явления или проблемы. |
| Задачи работы | Как идти к результату?<br>Определяются исходя из целей работы и в развитие поставленных целей. Формулировки задач необходимо делать как можно более тщательно, поскольку описание их решения должно составить содержание глав и параграфов работы. Рекомендуется сформулировать 3 – 4 задачи. |
| Методы исследования | Как изучали?<br>Краткое перечисление методов через запятую без обоснования. |
| Элемент введения | Комментарий к формулировке |

| | |
|---|---|
| Структура работы (завершающая часть введения) | Что в итоге в работе/проекте представлено.<br>Краткое изложение перечня и/или содержания глав работы/проекта. |

Разработка основной части курсовой работы/проекта. Основная часть обычно состоит из двух-трех разделов: в первом содержатся теоретические основы темы; дается история вопроса, уровень разработанности вопроса темы в теории и практике посредством сравнительного анализа литературы.

В теоретической части рекомендуется излагать наиболее общие положения, касающиеся данной темы, а не вторгаться во все проблемы в глобальном масштабе. Теоретическая часть предполагает анализ объекта исследования и должна содержать ключевые понятия, историю вопроса, уровень разработанности проблемы в теории и практике. Излагая содержание публикаций других авторов, необходимо обязательно давать ссылки на них с указанием номеров страниц этих информационных источников.

Вторым разделом является практическая часть, которая должна носить сугубо прикладной характер. В ней необходимо описать конкретный объект исследования, привести результаты практических расчетов и направления их использования, а также сформулировать направления совершенствования, либо вынести их в отдельных – третий раздел курсовой работы (проекта).

Важно глубоко изучить наиболее существенные с точки зрения задач курсовой работы (проекта) стороны и особенности.

Разработка заключения. По окончании исследования подводятся итоги по теме. Заключение носит форму синтеза полученных в работе результатов. Его основное назначение - резюмировать содержание работы, подвести итоги проведенного исследования. В заключении излагаются полученные выводы и их соотношение с целью исследования, конкретными задачами, гипотезой, сформулированными во введении.

Проведенное исследование должно подтвердить или опровергнуть гипотезу исследования. В случае опровержения гипотезы даются рекомендации по возможному совершенствованию деятельности в свете исследуемой проблемы.

Составление списка литературы. В список источников и литературы включаются источники, изученные Вами в процессе подготовки работы, в т.ч. те, на которые Вы ссылаетесь в тексте курсовой работы/проекта.

Список используемой литературы должен содержать не менее 20 источников (не менее 10 книг и 10-15 материалов периодической печати), с которыми работал автор курсовой работы (проекта).

Список используемой литературы включает в себя:

- нормативные правовые акты;
- научную литературу и материалы периодической печати;
- практические материалы.

## 6. ТРЕБОВАНИЯ К ОФОРМЛЕНИЮ КР (КП)

Курсовые работы (проекты) следует оформлять в печатном виде с использованием компьютера и принтера распечатывать на одной стороне листа белой бумаги формата А4. Рукописное оформление работы не допускается (разрешается вписывать черными чернилами отдельные слова, формулы, условные знаки, а также выполнять отдельные иллюстрации).

Вне зависимости от способа выполнения работы качество напечатанного текста и оформления иллюстраций, таблиц, распечаток с ЭВМ должно удовлетворять требованию их четкого воспроизведения. При выполнении отчета необходимо соблюдать равномерную плотность, контрастность и четкость изображения по всему отчету. В отчете должны быть четкие, не расплывшиеся линии, буквы, цифры и знаки.

Расположение текста должно обеспечивать соблюдение следующих полей:

- левое поле - не менее 30 мм;
- правое поле - не менее 10 мм;
- верхнее поле - не менее 20 мм;
- нижнее поле - не менее 20 мм.

**Все страницы курсовой работы (проекта), включая приложения, должны быть пронумерованы** арабскими цифрами сквозной нумерацией по всему тексту. Первой страницей является титульный лист, на котором номер страницы не проставляется. Порядковый номер страницы помещается в нижнем правом углу колонтитула.

Структура выпускной квалификационной работы состоит из следующих элементов:

1. *Титульный лист,* образец которого представлен в приложении А
2. *Пояснительная записка*:
   - Содержание (см. Приложение Б) - включает в себя перечень частей ВКР с указанием страниц, соответствующих началу каждой части работы;
   - Введение - раскрывает актуальность выбранной темы исследования, степень разработанности темы, цели, задачи, объект, предмет, гипотезу и методы исследования, структуру работы;
   - Основная часть - состоит из нескольких глав, содержащих параграфы;
   - Заключение - подводятся основные итоги работы, обобщаются полученные результаты, освещаются рекомендации по конкретному использованию результатов выпускной квалификационной работы и направления дальнейших исследований;
   - Список использованных источников - он включает литературу, используемую при подготовке текста: цитируемую, упоминаемую, а также имеющую непосредственное отношение к исследуемой теме. Полнота списка зависит от тщательности сбора публикаций. Правильно составленный и грамотно оформленный список свидетельствует о том, насколько автор знаком с литературой по теме исследования. Важным компонентом является работа автора с литературой последних трех-пяти лет, как показатель ориентированности автора в современном состоянии научной изученности темы исследования. Библиографический список должен включать не менее 20 источников.
   - Приложения (если таковые имеются).

Оформление заголовков и основного текста.

Текст работы следует разделять на разделы, подразделы и пункты. Разделы и подразделы должны иметь заголовки. Наименования структурных элементов отчета "СОДЕРЖАНИЕ", "ВВЕДЕНИЕ», "ЗАКЛЮЧЕНИЕ", "СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ", "ПРИЛОЖЕНИЕ" служат заголовками структурных элементов работы (проекта). Заголовки структурных элементов (введение, заключение, главы и т.п.) следует располагать в середине строки без точки в конце и печатать прописными (заглавными) буквами, не подчеркивая, полужирный шрифт не применяется.

Разделы основной части пояснительной записки работы (проекта) должны иметь порядковые номера в пределах всего документа, обозначенные арабскими цифрами без точки и записанные с абзацного отступа. Подразделы должны иметь нумерацию в пределах каждого раздела. В конце номера подраздела точка не ставится.

Каждый раздел следует начинать с нового листа (страницы). Расстояние между заголовками раздела или подраздела приблизительно 1.5-2 см. Расстояние между заголовками раздела и текстом должно быть равно 1,5-2 см.

Расположение текста должно обеспечивать следующих полей:

- левое поле – не менее 30 мм;
- правое поле – не менее 10 мм;
- верхнее поле – не менее 20 мм;
- нижнее поле – не менее 20 мм.

Все страницы курсовой работы (проекта), включая приложения, должны быть пронумерованы арабскими цифрами, шрифт Times New Roman, 12 пт. Порядковый номер страницы помещается в нижнем правом углу колонтитула.

Оформление заголовков раздела (ВВЕДЕНИЕ, ГЛАВА и т.д.):

- междустрочный интервал - 1,5;
- шрифт Times New Roman;
- **написание - прописные (заглавные) буквы;**
- **полужирный шрифт не применяется**;
- размер шрифта 14 пт;
- **режим выравнивания - по центру**;
- отступ в начале абзаца - 15 мм.

Оформление заголовков подраздела и подпункта (1.1, 1.2 и т.д.):

- междустрочный интервал - 1,5;
- шрифт Times New Roman;
- **написание - первая заглавная, остальные строчные буквы;**
- **полужирный шрифт не применяется;**
- размер шрифта 14 пт;
- **режим выравнивания - слева;**

– отступ в начале абзаца - 15 мм.

Оформление основного текста работы (проекта):

– междустрочный интервал - 1,5;
– шрифт Times New Roman;
– **полужирный шрифт не применяется**;
– размер шрифта 14 пт (для таблиц допускается 12 пт);
– **режим выравнивания - по ширине**;
– отступ в начале абзаца - 15 мм.

Разрешается использовать компьютерные возможности акцентирования внимания на определенных терминах, формулах, теоремах, применяя шрифты разной гарнитуры.

Числовые значения величин в тексте следует указывать с необходимой степенью точности, при этом в ряду величин осуществляется выравнивание числа знаков после запятой. Округление числовых значений величин до первого, второго, третьего и т.д. десятичного знака для величин одного наименования должны быть одинаковыми. Например: 1,50; 1,75; 2,00.

Оформление списков.

Внутри пунктов или подпунктов раздела могут быть приведены перечисления, которые записываются с абзацного отступа. **Перед каждой позицией перечисления следует ставить дефис**, а при необходимости ссылки в тексте ВКР на один из элементов перечисления вместо дефиса ставятся строчные буквы в порядке русского алфавита, начиная с буквы а (за исключением букв ё, з, й, о, ч, ъ, ы, ь). Для дальнейшей детализации перечислений необходимо использовать арабские цифры, после которых ставится скобка, а запись производится с абзацного отступа.

Примеры приведены на рисунках 1 и 2.

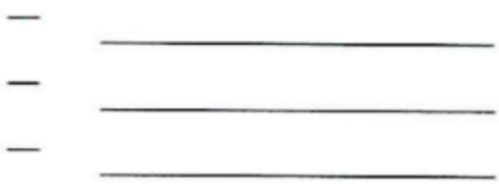


Рисунок 1 – Пример оформления списка

а) ____________

    1) ____________

    2) ____________

б) ____________

в) ____________

Рисунок 2 – Пример оформления списка при необходимости дальнейшей ссылки на один из его элементов

Оформление формул.

Уравнения и формулы следует выделять из текста в отдельную строку. Выше и ниже каждой формулы или уравнения должно быть оставлено не менее одной свободной строки. Пояснения символов и числовых коэффициентов, входящих в формулу, если они не пояснены ранее в тексте, должны быть приведены непосредственно под формулой. Пояснения каждого символа следует давать с новой строки в той последовательности, в которой символы приведены в формуле. Первая строка пояснения должна начинаться со слова "где" без двоеточия после него.

Формулы должны нумероваться сквозной нумерацией арабскими цифрами, которые записывают на уровне формулы в крайнем положении справа в круглых скобках. Одну формулу обозначают - (1).

Ссылки в тексте на порядковые номера формул дают в скобках, например "... в формуле (1)".

Пример:

Плотность каждого образца ρ, кг/м$^3$, вычисляют по формуле

$$\rho = \frac{m}{V}, \tag{1}$$

где $m$ - масса образца, кг;

$V$ - объем образца, м$^3$.

Оформление таблиц.

Таблицу следует располагать непосредственно после текста, в котором она упоминается впервые. При ссылке следует писать слово "таблица" с указанием ее номера.

Все таблицы должны иметь название и порядковую нумерацию. Таблицы нумеруются арабскими цифрами сквозной нумерацией в пределах всей работы (за исключением таблиц приложений). Номер таблицы следует проставлять в левом верхнем углу над таблицей после слова Таблица, без знака №, например, "Таблица 1". В приложениях таблицы обозначают отдельной нумерацией арабскими цифрами с добавлением перед цифрой обозначения приложения, например "Таблица В.1", если она приведена в приложении В.

Название таблицы должно отражать ее содержание, быть точным кратким. **Наименование таблицы следует помещать над таблицей слева, без абзацного отступа в одну строку с ее номером через тире**.

Таблицы выравниваются по центру страницы и оформляются в соответствии с рисунком 3. Выше и ниже каждой таблицы должно быть оставлено не менее одной свободной строки.


Рисунок 3 – Оформление таблиц

В каждой таблице следует указывать единицы измерения показателей и период времени к которому относятся данные. если единица измерения в таблице является общей для всех числовых данных то ее приводят в заголовке таблицы после ее названия.

Таблицу с большим числом строк допускается переносить на другой лист (страницу). При переносе части таблицы на другой лист (страницу) слово "Таблица", ее номер и наименование указывают один раз слева над первой частью таблицы и указывают номер таблицы (рисунок 4).

Расчет влияния каждого фактора представлен в таблице 1.

Таблица 1 - Расчет влияния факторов на деятельность производственного предприятия

| Фактор | Расчет |
|---|---|
| 1. Выручка от реализации | $\Delta Пп_в = (В_1 – В_0) \times R1_0 / 100$ (6)<br>где $\Delta Пп_в$ - изменение суммы прибыли от продаж за счет изменения |

21

Продолжение таблицы 1

| | |
|---|---|
| | объемов выручки; $В_1$ и $В_0$ - соответственно выручка от продажи в отчетном (1) и базисном (0) периодах; $R1_0$ - рентабельность продаж в базисном периоде |
| 2. Себестоимость реализованной продукции | $\Delta Пп_с = В_1 \times (УС_1 - УС_0) / 100$ (7)<br>где $УС_1$ и $УС_0$ – соответственно уровни себестоимости в отчетном и базисном периодах. |
| 3. Коммерческие расходы | $\Delta Пп_{кр} = В_1 \times (УКР_1 - УКР_0)/100$ (8)<br>где $УКР_1$ и $УКР_0$ – соответственно уровни коммерческих расходов в отчетном и базисном периодах. |
| 4. Управленческие расходы | $\Delta Пп_{ур} = В_1 \times (УУР_1 - УУР_0)/100$ (9)<br>где $УУР_1$ и $УУР_0$ – соответственно уровни управленческих расходов в отчетном и базисном периодах. |

Рисунок 4 – Оформление при делении таблиц

Оформление иллюстраций и графической части.

Весь графический материал (схемы, диаграммы, фотографии, чертежи и т.п.), расположенный по тексту работы (не включая приложения), следует нумеровать арабскими цифрами сквозной нумерацией. Если рисунок один, то он обозначается "Рисунок 1". Графики, схемы, диаграммы, располагаются в работе непосредственно после текста, имеющего на них ссылку, или на следующей странице. Поясняющие данные помещают под иллюстрацией, а **ниже по центру печатают слово "Рисунок", его номер, а через знак "–" и его наименование**. Иллюстрации каждого приложения обозначают отдельной нумерацией арабскими цифрами с добавлением перед цифрой обозначения приложения. Например, "Рисунок А.3 – Детали прибора".

Пример оформления иллюстраций представлен на рисунке 5.

Структура продаж товаров основных товарных групп магазина представлена на рисунке 1.

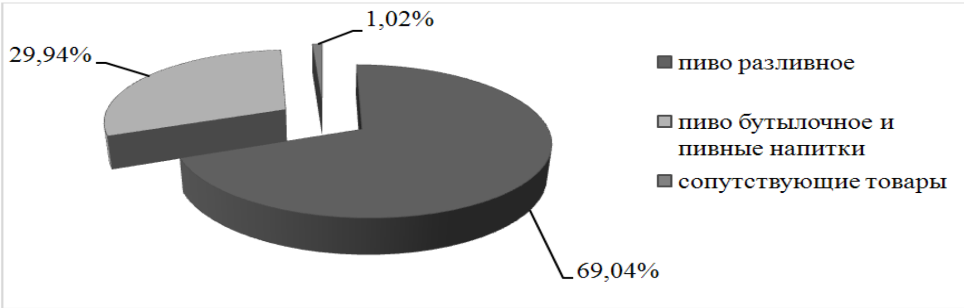



Рисунок 1 – Структура продаж магазина «Золотая кружка» (ИП Медведев К.Д.) по данным на 2015г., %

Рисунок 5 – Пример оформления иллюстраций и графической части

При ссылках на иллюстрации следует писать "... в соответствии с рисунком 2".

Выше и ниже каждого рисунка должно быть оставлено не менее одной свободной строки.

Оформление приложений.

Материал, дополняющий текст документа, допускается помещать в приложениях. Приложения могут быть, например, графический материал, таблицы большого формата, расчеты, описания аппаратуры и приборов, описания алгоритмов и программ задач, решаемых на ЭВМ и т.д. Приложения располагают в порядке появления ссылок на них в тексте документа. В тексте документа на все приложения должны быть даны ссылки.

**Каждое приложение следует начинать с новой страницы с указанием наверху посередине страницы слова "ПРИЛОЖЕНИЕ" и его обозначения**.

Приложения обозначают заглавными буквами русского алфавита, начиная с А. Буквы Ё, З, Й, О, Ч, Ь, Ы, Ъ для обозначения приложений НЕ используются.

**Приложение должно иметь заголовок**, который записывают симметрично относительно текста (выравнивание по тексту) с прописной (заглавной) буквы с новой строки.

Пример оформления приложения представлен на рисунке 6.

ПРИЛОЖЕНИЕ А

Основные группы показателей экономической эффективности хозяйственной деятельности предприятия

Таблица А.1 - Основные группы показателей экономической эффективности хозяйственной деятельности предприятия

| Показатели | Характеристика | Способ расчета |
|---|---|---|
| | I. Производительность труда | |
| 1. Выработка | Отражает количество продукции, произведенной в единицу рабочего времени или приходящееся на одного среднесписочного работника в месяц, квартал, год | Отношение количества произведенной продукции к затратам рабочего времени на производство этой продукции |
| 2. Трудоемкость | Величина, обратная выработке, характеризует затраты труда на производство единицы продукции | Отношение затрат труда к объему продукции |

Рисунок 6 – Пример оформления приложения

Оформление библиографического списка используемой литературы.

Список используемой литературы содержит перечень источников, используемых обучающимся при работе над темой работы (проекта). Список используемой литературы нумеруется арабскими цифрами, после которых ставится скобка, запись производится с абзацного отступа. Сведения об источниках следует располагать в порядке появления ссылок на источники в тексте работы (проекта).

При написании работы обучающийся обязан давать ссылку на источник, библиографическое описание которого должно приводиться в списке используемых источников. Порядковый номер ссылки в тексте работы заключают в квадратные скобки.

Для каждого учебника, книги обязательно должен быть указан уникальный номер книжного издания ISBN, для периодических изданий – ISSN, для электронных ресурсов – ссылка (URL) и дата обращения.

## 7. ПРОЦЕДУРА ЗАЩИТЫ КР (КП)

Курсовая работа (проект), выполненная с соблюдением рекомендуемых требований, оценивается и допускается к защите. Защита должна производиться до начала экзамена или зачета по дисциплине.

Процедура защиты курсовой работы/проекта включает в себя:

- выступление студента по теме и результатам работы (5-8 мин),
- ответы на вопросы аудитории и научного руководителя работы.

Окончательная оценка за курсовую работу (проект) выставляется преподавателем после защиты.

Результаты защиты курсового проекта оцениваются по четырехбалльной системе: «отлично», «хорошо», «удовлетворительно», «неудовлетворительно», а при выполнении курсовой работы по двухбалльной системе: «зачтено» или «незачтено».

Положительная оценка по той дисциплине, по которой предусматривается курсовая работа (проект), выставляется только при условии успешной сдачи курсовой работы (проекта) на оценку не ниже «удовлетворительно».

К защите курсовой работы (проекта) предъявляются следующие требования:

1. Глубокая теоретическая проработка исследуемых проблем на основе анализа экономической литературы.
2. Умелая систематизация цифровых данных в виде таблиц и графиков с необходимым анализом, обобщением и выявлением тенденций развития исследуемых явлений и процессов.
3. Критический подход к изучаемым фактическим материалам с целью поиска направлений совершенствования деятельности.
4. Аргументированность выводов, обоснованность предложений и рекомендаций.
5. Логически последовательное и самостоятельное изложение материала.
6. Оформление материала в соответствии с установленными требованиями.

Для выступления на защите необходимо заранее подготовить и согласовать с руководителем тезисы доклада и иллюстрационный материал в виде презентации.

Рекомендуемые структура, объем и время доклада приведены в таблице 2.

Таблица 2 – Структура, объем и время доклада

<table>
<tr><th>№</th><th>Структура доклада</th><th>Объем</th><th>Время</th></tr>
<tr><td>1.</td><td>Представление темы работы.</td><td rowspan="3">До 1,5 страниц</td><td rowspan="3">До 2 минут</td></tr>
<tr><td>2.</td><td>Актуальность темы.</td></tr>
<tr><td>3.</td><td>Цель работы.</td></tr>
<tr><td>4.</td><td>Постановка задачи, результаты ее решения и сделанные выводы (по каждой из задач, которые были поставлены для достижения цели курсовой работы/ проекта).</td><td>До 6 страниц</td><td>До 7 минут</td></tr>
<tr><td>5.</td><td>Перспективы и направления дальнейшего исследования данной темы.</td><td>До 0,5 страницы</td><td>До 1 минуты</td></tr>
</table>

При составлении тезисов необходимо учитывать ориентировочное время доклада на защите, которое составляет 8-10 минут. Доклад целесообразно строить не путем изложения содержания работы по главам, а по задачам, то есть, раскрывая логику получения значимых результатов. В докладе обязательно должно присутствовать обращение к иллюстративному материалу, который будет использоваться в ходе защиты работы. Объем доклада должен составлять 7-8 страниц текста в формате Word, размер шрифта 14, полуторный интервал [110].

## ПРИМЕРНАЯ ТЕМАТИКА КУРСОВЫХ РАБОТ ПО ДИСЦИПЛИНЕ

1. Роль статистических методов в управлении фирмой
2. Моделирование вектора развития фирмы с использованием статистических методов
3. Прогнозирование деятельности фирмы
4. Управление промышленным производством на основе статистических методов
5. Управление сельским хозяйством на основе статистических методов
6. Корреляционно-регрессионный анализ как инструмент проектирования вектора развития фирмы
7. Точечное прогнозирование на микроуровне
8. Интервальное прогнозирование в управлении фирмой
9. Анализ тренда развития отдельных направлений деятельности фирмы
10. Прогнозирование научно-инновационного развития организаций
11. Сезонная составляющая в деятельности организаций
12. Анализ устойчивости развития фирмы: экономический аспект
13. Анализ устойчивости развития фирмы: социальный аспект
14. Анализ устойчивости развития фирмы: экологический аспект
15. Оценка инновационной составляющей деятельности фирмы с использованием статистических методов
16. Статистический анализ и прогнозирование доходов предприятия
17. Исследование роли трудовых ресурсов фирмы с использованием статистических методов

18. Кластерный анализ в управлении фирмой
19. Факторный анализ в управлении фирмой
20. Дисперсионный анализ в управлении фирмой
21. Анализ влияния заработной платы сотрудников фирмы на эффективность их работы с использованием статистических методов
22. Прогнозирование процессов на микроуровне с использованием производственных моделей (на примере функции Кобба–Дугласа)
23. Прогнозирование деятельности фирмы на основе метода экстраполяции
24. Исследование особенностей управления торговыми фирмами с использованием статистических методов
25. Исследование особенностей управления туристическими фирмами с использованием статистических методов
26. Управление малыми предприятиями на основе статистических методов
27. Использование статистических методов в управлении производственным процессом
28. Использование статистических методов в управлении качеством

# ЗАКЛЮЧЕНИЕ

В настоящее время в управлении фирмой активно применяются различные статистические методы. Возрастает актуальность повышения качества прогнозных исследований на микроуровне. Все это обусловливает необходимость углубленного изучения и разработки основных проблем применения статистических методов, с которыми можно столкнуться в процессе управления организацией.

В процессе систематизированного научно обоснованного прогнозирования развития социально-экономических процессов происходило развитие методологии прогнозирования как совокупности методов, приемов и способов мышления, позволяющих на основе анализа ретроспективных данных, экзогенных и эндогенных связей объекта прогнозирования, а также их измерений в рамках рассматриваемого явления или процесса вывести суждения определенной достоверности относительно его будущего развития.

Исследование различных классификационных схем методов прогнозирования позволяет выделить в качестве основных классов фактографические, экспертные и комбинированные методы, специализация которых обусловлена спецификой целей и задач, количеством и качеством исходной информации, периодом упреждения прогноза.

Прогнозы должны предшествовать планам, содержать оценку хода, последствий выполнения (или невыполнения) планов, охватывать все, что не поддается планированию, решению. Прогноз и план различаются способами оперирования информацией о будущем.

В первой главе пособия были рассмотрены теоретические аспекты применения статистических методов в управлении фирмой, в последующих девяти главах – даны практические аспекты моделирования и прогнозирования на микроуровне.

Приведенные в пособии положения могут быть использованы на лекционных и практических занятиях по дисциплине «Статистические методы в управлении фирмой».

# БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Шорохова, И.С. Статистические методы анализа : [учеб. пособие] / И. С. Шорохова, Н. В. Кисляк, О. С. Мариев; М-во образования и науки Рос. Федерации, Урал. федер. ун-т. – Екатеринбург : Изд-во Урал. ун-та, 2015. – 300 с. – ISBN 978-5-7996-1633-5

2. Курмангалиева А.К. Статистические методы анализа : [учеб. пособие]. – Костанай, КГУ им. А. Байтурсынова, 2018 – 112 с. – ISBN 978-601-7955-08-3

3. Басовский, Л. Е. Прогнозирование и планирование в условиях рынка : учеб. пособие / Л. Е. Басовский. – М. : ИНФРА-М, 2023. – 260 с. – (Высшее образование: Бакалавриат). – ISBN 978-5-16-004198-8.

4. Герасимов, А. Н. Социально-экономическое прогнозирование : учеб. пособие / А. Н. Герасимов, Е. И. Громов, Ю. С. Скрипниченко. – М. : СтГАУ : Агрус, 2017. – 144 с. – ISBN 978-5-9596-1294-8.

5. Почекутова, Е. Н. Прогнозирование и планирование : учеб.-метод. пособие / Е. Н. Почекутова, А. П. Феденко. – Красноярск : СФУ, 2016. – 126 с. – ISBN 978-5-7638-3439-0.

6. Кулешова, Е. С. Макроэкономическое планирование и прогнозирование : учеб. пособие / Е. С. Кулешова. – 2-е изд., доп. – Томск : Эль-Контент, 2015. – 178 с. – ISBN 978-5-4332-0252-8.

7. Петросов, А. А. Стратегическое планирование и прогнозирование / Петросов А.А. – М. :МГГУ, 2001. – 464 с.: ISBN 5-7418-0145-5.

8. Прогнозирование и планирование в условиях рынка [Электронный ресурс] : учеб. пособие / Т. Н. Бабич [и др.]. – М. : ИНФРА-М, 2020. – 336 с. – (Высшее образование: Бакалавриат). – URL: www.dx.doi.org/10.12737/2517. – ISBN 978-5-16-004577-1 (дата обращения: 15.03.2023).

9. Судакова, А. Е. Бюджетное планирование и прогнозирование : учеб. пособие / А. Е. Судакова, Г. А. Агарков, А. А. Тарасьев. - 2-е изд., стер. – М. : ФЛИНТА : Изд-во Урал. ун-та, 2022.– 308 с. - ISBN 978-5-9765-5024-7 (ФЛИНТА), ISBN 978-5-7996-2922-9 (Изд-во Урал. ун-та).

10. Основы экономического прогнозирования [Электронный ресурс] : учебное пособие / Е.В. Смирнова, Е.В. Чмышенко, И.Ю. Цыганова; Оренбургский гос. ун-т. – Оренбург: ОГУ, 2019. – 145 с. ISBN 978-5-7410-2425-6

11. Социально-экономическое прогнозирование: учебное пособие / Ю. А. Антохина, А. М. Колесникова, С. Н. Медведева ; М-во образования и науки Российской Федерации, Федеральное гос. бюджетное образовательное учреждение высш. образования Санкт-Петербургский гос. ун-т аэрокосмического приборостроения. - Санкт-Петербург : ГУАП, 2016. – 177 с. : табл.; 21 см.; ISBN 978-5-8088-1103-4.

12. Виноградская, Н. А. Управление производством : методы экономического прогнозирования и планирования : практикум / Н. А. Виноградская, Е. Н. Елисеева, О. О. Скрябин. – М. : Изд. Дом МИСиС, 2013. - 96 с. - ISBN 978-5-87623-687-6

13. Орехов, А. М. Методы экономических исследований : учебное пособие / А.М. Орехов. — 2-е изд. — Москва : ИНФРА-М, 2023. — 344 с. — (Высшее образование: Бакалавриат). - ISBN 978-5-16-005748-4

14. Бабешко, Л. О. Эконометрика и эконометрическое моделирование : учебник / Л.О. Бабешко, М.Г. Бич, И.В. Орлова. — 2-е изд., испр. и доп. – Москва : ИНФРА-М, 2023. – 387 с. : ил. – (Высшее образование: Бакалавриат). – DOI 10.12737/1141216. – ISBN 978-5-16-016417-5

15. Лычкина, Н. Н. Имитационное моделирование экономических процессов : учебное пособие / Н.Н. Лычкина. – Москва : ИНФРА-М, 2022. – 254 с. – (Высшее образование: Бакалавриат). – DOI 10.12737/724. – ISBN 978-5-16-017094-7

16. Бакланова, И. И. Теория вероятности : учебно-методическое пособие / И. И. Бакланова, Е. В. Матвеева, Л. А. Медведков. - Йошкар-Ола : Поволжский государственный технологический университет, 2017. - 64 с. - ISBN 978-5-8158-1801-9

17. Бочаров, П. П. Теория вероятностей. Математическая статистика [Электронный ресурс] / П. П. Бочаров, А. В. Печинкин. – 2-е изд. - Москва : ФИЗМАТЛИТ, 2005. - 296 с. - ISBN 5-9221-0633-3

18. Коган, Е. А. Теория вероятностей и математическая статистика : учебник / Е.А. Коган, А.А. Юрченко. – М. : ИНФРА-М, 2020. – 250 с. – (Среднее профессиональное образование). – ISBN 978-5-16-015649-1

19. Бакланова, И. И. Теория вероятности : учебно-методическое пособие / И. И. Бакланова, Е. В. Матвеева, Л. А. Медведков. – Йош-

кар-Ола: Поволжский государственный технологический университет, 2017. – 64 с. – ISBN 978-5-8158-1801-9

20. Теория вероятностей и математическая статистика : учебное пособие / Л.Г. Бирюкова, Г.И. Бобрик, Р.В. Сагитов [и др.] ; под ред. В.И. Матвеева. – 2-е изд., испр. и доп. – М.: ИНФРА-М, 2020. – 289 с. – (Среднее профессиональное образование). – ISBN 978-5-16-015712-2

21. Мхитарян, В. С. Теория вероятностей и математическая статистика [Электронный ресурс] : учеб. пособие / В. С. Мхитарян, Е. В. Астафьева, Ю. Н. Миронкина, Л. И. Трошин; под ред. В. С. Мхитаряна. – 2-е изд., перераб. и доп. – М. : Московский финансово-промышленный университет «Синергия», 2013. – (Университетская серия). – ISBN 978-5-4257-0106-0

22. Лычкина, Н. Н. Имитационное моделирование экономических процессов : учебное пособие / Н.Н. Лычкина. – Москва : ИНФРА-М, 2022. – 254 с. – (Высшее образование: Бакалавриат). – DOI 10.12737/724. – ISBN 978-5-16-017094-7

23. Булыгина, О. В. Имитационное моделирование в экономике и управлении : учебник / О.В. Булыгина, А.А. Емельянов, Н.З. Емельянова ; под ред. д-ра экон. наук, проф. А.А. Емельянова. – Москва : ИНФРА-М, 2021. – 592 с. – (Высшее образование: Бакалавриат). – DOI 10.12737/textbook_5b5ab5571bd995.05564317. – ISBN 978-5-16-014523-5

24. Гулай, Т.А. Теория вероятностей и математическая статистика [Электронный ресурс] : учебное пособие / Т.А. Гулай, А.Ф. Долгополова, Д.Б. Литвин, С.В. Мелешко. – 2-е изд., доп. – Ставрополь: АГРУС, 2013. – 260 с

25. Осипов, Г. В. Моделирование социальных явлений и процессов с применением математических методов : учебное пособие / Г. В. Осипов, В. А. Лисичкин ; под общ. ред. В. А. Садовничего. – Москва : Норма : ИНФРА-М, 2022. – 192 с. : ил. – (Социальные науки и математика). – ISBN 978-5-91768-533-5

26. Криволапов, С. Я. Теория вероятностей в примерах и задачах на языке R : учебник / С.Я. Криволапов. – М.: ИНФРА-М, 2023. – 412 с. + Доп. материалы [Электронный ресурс]. – (Высшее образование). – DOI 10.12737/1898404. – ISBN 978-5-16-017941-4

27. Ларионова, И. А. Статистика. Анализ временных рядов: учебное пособие / И. А. Ларионова. – М.: ИД МИСиС, 2001. – 73 с.

28. Воейко, О. А. Анализ временных рядов и прогнозирование : практикум / О. А. Воейко. – М.; Берлин : Директ-Медиа, 2019. – 175 с. – ISBN 978-5-4499-0178-1

29. Ярушкина, Н. Г. Интеллектуальный анализ временных рядов : учебное пособие / Н.Г. Ярушкина, Т.В. Афанасьева, И.Г. Перфильева. – М. : ИД «ФОРУМ»: ИНФРА-М, 2018. – 160 с. – (Высшее образование). – ISBN 978-5-8199-0496-1

30. Карпенко, Н. В. Эконометрика. Анализ и прогнозирование временного ряда : учебное пособие / Н. В. Карпенко. – М.: РУТ (МИИТ), 2018. – 132 с.

31. Замедлина, Е. А. Статистика: учебное пособие / Е.А. Замедлина – М.: РИОР : ИНФРА-М, 2019. – 160 с. – (Среднее профессиональное образование). – ISBN 978-5-369-01303-8

32. Шумак, О. А. Статистика: Учебное пособие / О.А. Шумак, А.В. Гераськин. – М.: ИЦ РИОР: НИЦ Инфра-М, 2019. – 311 с.: ил.; – (Высшее образование: Бакалавриат). – ISBN 978-5-369-01048-8

33. Статистика : учебник / В.В. Глинский, В.Г. Ионин, Л.К. Серга [и др.] ; под ред. В.Г. Ионина. – 4-е изд., перераб. и доп. – М.: ИНФРА-М, 2023. – 355 с. – (Высшее образование: Бакалавриат). – DOI 10.12737/25127. – ISBN 978-5-16-012070-6

34. Сергеева, И. И. Статистика: учебник / И.И. Сергеева, Т.А. Чекулина, С.А. Тимофеева. – 2-е изд., испр. и доп. — Москва : ФОРУМ : ИНФРА-М, 2021. – 304 с. – (Среднее профессиональное образование). – ISBN 978-5-8199-0888-4

35. Годин, А. М. Статистика : учебник для бакалавров / А. М. Годин. - 12-е изд., стер. – Москва : Издательско-торговая корпорация «Дашков и К°», 2020. – 410 с. – ISBN 978-5-394-03485-5

36. Тимофеева, И. Ю. Статистика. Часть 1. Общая теория статистики: Учебное пособие / Тимофеева И.Ю., Лаврова Е.В., Полякова О.Е. – М.:НИЦ ИНФРА-М, 2018. – 104 с. (Высшее образование)ISBN 978-5-16-107041-3

37. Сидоренко, М. Г. Статистика : учебное пособие / М.Г. Сидоренко. – М. : ФОРУМ : ИНФРА-М, 2022. – 160 с. – (Высшее образование). – ISBN 978-5-91134-160-2

38. Иода, Е. В. Статистика: Учебное пособие / Иода Е.В. - М.:Вузовский учебник, НИЦ ИНФРА-М, 2018. – 303 с. ISBN 978-5-9558-0144-5

39. Ивченко, Ю. С. Статистика: Учебное пособие / Ю.С. Ивченко. - Москва : ИЦ РИОР: ИНФРА-М, 2018. – 375 с.: – (Высшее образование). – ISBN 978-5-369-00636-8

40. Непомнящая, Н. В. Статистика: общая теория статистики, экономическая статистика. Практикум/НепомнящаяН.В., ГригорьеваЕ.Г. – Краснояр.: СФУ, 2015. – 376 с.: ISBN 978-5-7638-3185-6

41. Гужова, О. А. Статистика в управлении социально-экономическими процессами : учебное пособие / О.А. Гужова, Ю.А. Токарев. – М. : ИНФРА-М, 2020. – 172 с. – (Высшее образование: Бакалавриат). – ISBN 978-5-16-012151-2

42. Мелкумов, Я. С. Социально-экономическая статистика : учебное пособие / Я.С. Мелкумов. – 2-е изд., перераб. и доп. – М. : ИНФРА-М, 2023. – 186 с. – (Высшее образование: Бакалавриат). - ISBN 978-5-16-005424-7

43. Экономическая статистика. Практикум: учебное пособие / Ю.Н. Иванов, Г.Л. Громыко, А.Н. Воробьев [и др.] ; под ред. д-ра экон. наук, проф. Ю.Н. Иванова. – М. : ИНФРА-М, 2022. – 176 с. – (Высшее образование: Бакалавриат). – DOI 10.12737/23950. - ISBN 978-5-16-012772-9

44. Батракова, Л. Г. Социально-экономическая статистика : учебник / Л. Г. Батракова. – М.: Логос, 2020. – 480 с. – ISBN 978-5-98704-657-9

45. Гусева, Е. Н. Теория вероятностей и математическая статистика [Электронный ресурс] : Уч. пособ. / Е. Н. Гусева. - 5-е изд., стереотип. – М.: Флинта, 2011. – 220 с. – ISBN 978-5-9765-1192-7.

46. Непомнящая, Н. В. Статистика: общая теория статистики, экономическая статистика. Практикум/НепомнящаяН.В., ГригорьеваЕ.Г. – Краснояр.: СФУ, 2015. – 376 с.: ISBN 978-5-7638-3185-6

47. Теория вероятностей и математическая статистика : учебное пособие / Л.Г. Бирюкова, Г.И. Бобрик, Р.В. Сагитов [и др.] ; под ред. В.И. Матвеева. – 2-е изд., испр. и доп. – М. : ИНФРА-М, 2021. – 289 с. – (Высшее образование: Бакалавриат). – DOI 10.12737/18865. - ISBN 978-5-16-011793-5

48. Теоретическая метрология : методические указания к практическим занятиям / сост. И. А. Петюль, В. Д. Борозна. – Витебск : УО «ВГТУ», 2017. – 66 с.

49. Белько, И. В. Теория вероятностей, математическая статистика, математическое программирование : учебное пособие / И.В. Белько, И.М. Морозова, Е.А. Криштапович. – М.: ИНФРА-М, 2022. – 299 с. : ил. – (Высшее образование: Бакалавриат). – ISBN 978-5-16-011748-5

50. Эрастов, В. Е. Метрология, стандартизация и сертификация : учебное пособие / В.Е. Эрастов. – 2-е изд., перераб. и доп. — Москва : ИНФРА-М, 2022. – 196 с. – (Высшее образование: Бакалавриат). – DOI 10.12737/23696. – ISBN 978-5-16-012324-0

51. Белокопытов, В. И. Организация, планирование и обработка результатов эксперимента : учебное пособие / В. И. Белокопытов. – Красноярск : Сиб. федер. ун-т, 2020. – 132 с. – ISBN 978-5-7638-4297-5

52. Шпаков, П. С. Математическая обработка результатов измерений / П. С. Шпаков, Ю. Л. Юнаков. – Красноярск : СФУ, 2014. - 410 с. – ISBN 978-5-7638-3077-4.

53. Гальянов, А. В. Математическая обработка результатов измерений в горном деле : учебное пособие / А. В. Гальянов. – 2-е изд., испр. и доп. – М.; Вологда : Инфра-Инженерия, 2022. – 292 с. - ISBN 978-5-9729-0815-8

54. Организация производства и управление предприятием : учебник / под ред. О.Г. Туровца. – 3-е изд. – М.: ИНФРА-М, 2022. – 506 с. – (Среднее профессиональное образование). – ISBN 978-5-16-015612-5

55. Пискажова, Т. В. Математическое моделирование объектов и систем управления : учебное пособие / Т. В. Пискажова, Т. В. Донцова, Г. Б. Даныкина. – Красноярск : Сиб. федер. ун-т, 2020. – 230 с. – ISBN 978-5-7638-4184-8

56. С.С. Бондарчук, И.С. Бондарчук. Б81 Статобработка экспериментальных данных в MS Excel: учебное пособие. – Томск: Издательство Томского государственного педагогического университета, 2018. – 433

57. Жуков, В. И. Методология математического моделирования управления социальными процессами : монография / В.И. Жуков, Г.С. Жукова. – М.: ИНФРА-М, 2019. – 207 с. – (Научная мысль). – ISBN 978-5-16-108296-6

58. Двойцова, И. Н. Основы математического моделирования социально-экономических процессов : учебное пособие / И. Н. Двойцова. – Железногорск : ФГБОУ ВО Сибирская пожарно-спасательная академия ГПС МЧС России, 2022. – 112 с.

59. Инструментальные средства математического моделирования: учебное пособие / Золотарев А.А., Бычков А.А., Золотарева Л.И. – Ростов н/Д: Издательство ЮФУ, 2011. – 90 с. ISBN 978-5-9275-0887-7

60. Колпаков, В. Ф. Экономико-математическое и эконометрическое моделирование: компьютерный практикум : учеб. пособие / В.Ф. Колпаков. – М. : ИНФРА-М, 2018. — 396 с. – (Высшее образование: Бакалавриат). – www.dx.doi.org/10.12737/24417. - ISBN 978-5-16-010967-1

61. Гусева, Е. Н. Экономическо-математическое моделирование [Электронный ресурс] : Уч. пособ. / Е. Н. Гусева. – 2-е изд., стереотип. - Москва : Флинта : МПСИ, 2011. – 216 с. – ISBN 978-5-89349-976-6 (Флинта), ISBN 978-5-9770-0256-1 (МПСИ)

62. Кундышева, Е. С. Экономико-математическое моделирование : учебник / Е. С. Кундышева ; под ред. Б. А. Суслакова. – 4-е изд. - Москва : Дашков и К°, 2012. – 424 с. - ISBN 978-5-394-01716-2

63. Введение в математическое моделирование : учебное пособие / В. Н. Ашихмин, М. Б. Гитман, И. Э. Келлер [и др.] ; под. ред. П. В. Трусова. – Москва : Логос, 2020. – 440 с. – ISBN 978-5-98704-637-1

64. Орлова, И. В. Экономико-математическое моделирование: практическое пособие по решению задач / И.В. Орлова, М.Г. Бич. – 3-е изд., испр. и доп. – Москва : Вузовский учебник : ИНФРА-М, 2023. – 190 с. – ISBN 978-5-9558-0527-6

65. Федосеев, В.В. Математическое моделирование в экономике и социологии труда. Методы, модели, задачи: учеб. пособие для студентов вузов, обучающихся по специальностям 080104 «Экономика труда», 080116 «Математические методы в экономике» / В.В. Федосеев. – М.: ЮНИТИ-ДАНА, 2017. – 167 с. – ISBN 978-5-238-01114-8

66. Математическое моделирование и проектирование : учеб. пособие / А.С. Коломейченко, И.Н. Кравченко, А.Н. Ставцев, А.А. Полухин ; под ред. А.С. Коломейченко. – М.: ИНФРА-М, 2018. – 181 с. – (Высшее образование: Магистратура). – www.dx.doi.org/10.12737/textbook_59688803c3cb35.15568286. – ISBN 978-5-16-012890-0

67. Гаврилов, Л. П. Информационные технологии в коммерции : учебное пособие / Л.П. Гаврилов. – 2-е изд., перераб. и доп. – Москва : ИНФРА-М, 2022. – 369 с. + Доп. материалы [Электронный ресурс]. – (Высшее образование: Бакалавриат). – DOI 10.12737/1085795. – ISBN 978-5-16-016187-7

68. Математическое моделирование и проектирование : учебное пособие / А.С. Коломейченко, И.Н. Кравченко, А.Н. Ставцев, А.А. Полухин ; под ред. А.С. Коломейченко. – М.: ИНФРА-М, 2021. – 181 с. – (Среднее профессиональное образование). – ISBN 978-5-16-015651-4

69. Тарасик, В. П. Математическое моделирование технических систем : учебник / В.П. Тарасик. – Минск : Новое знание ; Москва : ИНФРА-М, 2022. – 592 с. – (Высшее образование: Бакалавриат). – ISBN 978-5-16-011996-0

70. Назарова, Ю. Н. Математическое моделирование в экономике : практикум : специальность : 38.05.01 «Экономическая безопасность». Специализация : «Судебная экономическая экспертиза» / Ю. Н. Назарова. – Волгоград : ФГБОУ ВО Волгоградский ГАУ, 2019. – 68 с.

71. Введение в математическое моделирование : учебное пособие / В. Н. Ашихмин, М. Б. Гитман, И. Э. Келлер [и др.] ; под. ред. П. В. Трусова. – Москва : Логос, 2020. – 440 с. – ISBN 978-5-98704-637-1

72. Математическое моделирование и проектирование : учебное пособие / А.С. Коломейченко, И.Н. Кравченко, А.Н. Ставцев, А.А. Полухин ; под ред. А.С. Коломейченко. – Москва : ИНФРА-М, 2021. – 181 с. – (Среднее профессиональное образование). – ISBN 978-5-16-015651-4

73. Колпаков, В. Ф. Экономико-математическое и эконометрическое моделирование: компьютерный практикум : учеб. пособие / В.Ф. Колпаков. – М.: ИНФРА-М, 2018. – 396 с. – (Высшее образование: Бакалавриат). – www.dx.doi.org/10.12737/24417. – ISBN 978-5-16-0109671

74. Белько, И. В. Теория вероятностей, математическая статистика, математическое программирование : учебное пособие / И.В. Белько, И.М. Морозова, Е.А. Криштапович. – М.: ИНФРА-М, 2022. – 299 с. : ил. – (Высшее образование: Бакалавриат). – ISBN 978-5-16-011748-5

75. Михалева, М. Ю. Математическое моделирование и количественные методы исследований в менеджменте : учеб. пособие / М.Ю. Михалева, И.В. Орлова. – М.: Вузовский учебник : ИНФРА-М, 2018. – 296 с. – (Высшее образование: Магистратура). — www.dx.doi.org/10.12737/textbook_5b03f73021f562.03199866. - ISBN 978-5-9558-0607-5

76. Орлова, И. В. Экономико-математическое моделирование: практическое пособие по решению задач / И.В. Орлова, М.Г. Бич. – 3-е изд., испр. и доп. – М.: Вузовский учебник : ИНФРА-М, 2023. – 190 с. - ISBN 978-5-9558-0527-6

77. Анализ и диагностика финансово-хозяйственной деятельности предприятия : учебник / под ред. А.П. Гарнова. — Москва : ИНФРА-М, 2023. – 366 с. + Доп. материалы [Электронный ресурс]. — (Высшее образование: Бакалавриат). – DOI 10.12737/8240. - ISBN 978-5-16-009995-8

78. Погорелова, М. Я. Экономический анализ: теория и практика: Учебное пособие / Погорелова М.Я. – М.:ИЦ РИОР, НИЦ ИНФРА-М, 2018. – 290 с.(Высшее образование: Бакалавриат). - ISBN 978-5-369-01295-6

79. Гулай, Т.А. Теория вероятностей и математическая статистика [Электронный ресурс]: учебное пособие / Т.А. Гулай, А.Ф. Долгополова, Д.Б. Литвин, С.В. Мелешко. – 2-е изд., доп. – Ставрополь: АГРУС, 2013. – 260 с.

80. Пахунова, Р. Н. Общая и прикладная статистика : учебник для студентов высшего профессионального образования / П.Ф. Аскеров, Р.Н. Пахунова, А.В. Пахунов; под общ. ред. Р.Н. Пахуновой. – Москва : ИНФРА-М, 2022. — 272 с. + Доп. материалы [Электронный ресурс]. – (Высшее образование: Бакалавриат). – DOI 10.12737/748. – ISBN 978-5-16-006669-1

81. Прикладная математическая статистика : учебное пособие / сост. А. А. Мицель. - Томск : Томский государственный университет систем управления и радиоэлектроники, 2016. - 113 с. - Текст : электронный. – URL: https://znanium.com/catalog/product/1845838 (дата обращения: 27.03.2023)

82. Полякова, В. В. Прикладная статистика: методы анализа эмпирической информации : учебно-методическое пособие / В. В. Полякова, Н. В. Шаброва ; Министерство науки и высшего образования

Российской Федерации, Уральский федеральный университет им. первого Президента России Б. Н. Ельцина. – Екатеринбург : Изд-во Уральского ун-та, 2020. – 188 с. – ISBN 978-5-7996-3021-8

83. Зайцев, В. М. Прикладная медицинская статистика: учебное пособие / В. М. Зайцев, В. Г. Лифляндский, В. И. Маринкин. – 2-е изд. – Санкт-Петербург : ООО «Издательство ФОЛИАНТ», 2006. – 432 с. - ISBN 5-93929-135-X

84. Григорьев, А. А. Методы и алгоритмы обработки данных : учебное пособие / А.А. Григорьев, Е.А. Исаев. –2-е изд., перераб. и доп. — Москва : ИНФРА-М, 2022. – 383 с. + Доп. материалы [Электронный ресурс]. – (Высшее образование: Бакалавриат). – DOI 10.12737/1032305. – ISBN 978-5-16-015581-4

85. Статистический анализ данных, моделирование и исследование вероятностных закономерностей. Компьютерный подход/ЛемешкоБ.Ю., ЛемешкоС.Б., ПостоваловС.Н. и др. – Новосибирск : НГТУ, 2011. – 888 с.: ISBN 978-5-7782-1590-0.

86. Эконометрика: учебник / В.Н. Афанасьев, Т.В. Леушина, Т.В. Лебедева, А.П. Цыпин; под ред. проф. В.Н. Афанасьева; Оренбургский гос. ун-т. – Оренбург: ОГУ, 2012. – 402 с.

87. Лемешко, Б. Ю. Критерии проверки гипотез о случайности и отсутствии тренда. Руководство по применению : монография / Б.Ю. Лемешко, И.В. Веретельникова. – Москва : ИНФРА-М, 2021. – 221 с. – (Научная мысль). – DOI 10.12737/1587437. – ISBN 978-5-16-017054-1

88. Теория статистики : учебник / под ред. проф. Г.Л. Громыко. – 4-е изд., перераб. и доп. — Москва : ИНФРА-М, 2021. – 465 с. – (Высшее образование: Бакалавриат). – DOI 10.12737/textbook_5d0734d6e23853.79720708. – ISBN 978-5-16-014914-1

89. Альсова, О. К. Исследование временных рядов в среде R : учебное пособие / О. К. Альсова. – Новосибирск : Изд-во НГТУ, 2021. – 88 с. – ISBN 978-5-7782-4337-8

90. Афанасьев В.Н., Юзбашев М.М. Анализ временных рядов и прогнозирование: Учебник. – М.: Финансы и статистика, 2001 – 228 с. – ISBN 5-279-02419-8

91. Каяйкина М. С. Статистические методы изучения урожайности (на примере совхоза Ленинградской области). –Л.,1969. 106 с.

92. Манелля, А. И. Измерение устойчивости производства продукции земледелия // Статистический анализ развития АПК. – М.: Наука, 1992. С. 60-73.

93. Двойцова, И. Н. Основы математического моделирования социально-экономических процессов : учебное пособие / И. Н. Двойцова. – Железногорск : ФГБОУ ВО Сибирская пожарно-спасательная академия ГПС МЧС России, 2022. – 112 с.

94. Валентинов, В. А. Эконометрика / Валентинов В.А., – 3-е изд. – Москва :Дашков и К, 2016. – 436 с.: ISBN 978-5-394-02111-4

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

96. Балдин К.В., Быстров О.Ф., Соколов М.М. Эконометрика: Учеб. пособие для вузов. – 2-е изд., перераб. и доп. – М. : ЮНИТИ-ДАНА, 2017. – 254 с. – ISBN 978-5-238-00702-7

97. Середа, В. А. Эконометрика : учебное пособие / В. А. Середа, А. В. Литаврин, Н. Л. Собачкина. – Красноярск : Сиб. федер. ун-т, 2018. – 148 с. – ISBN 978-5-7638-3996-8

98. Едронова, В. Н. Экономический анализ развития территорий : учебник / В.Н. Едронова. – Москва : Магистр : ИНФРА-М, 2023. – 328 с. - ISBN 978-5-9776-0547-2

99. Пахунова, Р. Н. Общая и прикладная статистика : учебник для студентов высшего профессионального образования / П.Ф. Аскеров, Р.Н. Пахунова, А.В. Пахунов ; под общ. ред. Р.Н. Пахуновой. – Москва : ИНФРА-М, 2022. – 272 с. + Доп. материалы [Электронный ресурс]. – (Высшее образование: Бакалавриат). – DOI 10.12737/748. - ISBN 978-5-16-006669-1

100. Сергеева, И. И. Статистика : учебник / И.И. Сергеева, Т.А. Чекулина, С.А. Тимофеева. – 2-е изд., испр. и доп. – М.: ФОРУМ : ИНФРА-М, 2021. – 304 с. – (Среднее профессиональное образование). – ISBN 978-5-8199-0888-4

101. Лысенко, С. Н. Общая теория статистики : учебное пособие / С. Н. Лысенко, И. А. Дмитриева. – изд. испр. и доп. – М.: Вузовский учебник : ИНФРА-М, 2022. – 219 с. - ISBN 978-5-9558-0115-5

102. Ефимова, М. Р. Общая теория статистики: Учебник / М.Р. Ефимова, Е.В. Петрова, В.Н. Румянцев. – 2-е изд., испр. и доп. – М.:

ИНФРА-М, 2011. – 416 с. (Высшее образование). ISBN 978-5-16-004265-7.

103. Рыжикова, Т. Н. Маркетинг: экономика, финансы, контроллинг : учебное пособие / Т.Н. Рыжикова. – Москва : ИНФРА-М, 2023. – 225 с. – (Высшее образование: Бакалавриат). — DOI 10.12737/24399. - ISBN 978-5-16-012515-2

104. Статистическое моделирование и прогнозирование : учеб. пособие / Е. М. Марченко [и др.] ; Владим. гос. ун-т им А. Г. и Н. Г. Столетовых. – Владимир : Изд-во ВлГУ, 2018. – 100 с. ISBN 978-5-9984-0861-8

105. Дайитбегов, Д. М. Компьютерные технологии анализа данных в эконометрике: Монография / Д.М. Дайитбегов. – 3-е изд., испр. и доп. – М.: Вузовский учебник: НИЦ Инфра-М, 2018. – XIV, 587 с.: – (Научная книга). – ISBN 978-5-9558-0275-6

106. Айвазян, С. А. Эконометрика – 2: продвинутый курс с приложениями в финансах: Учебник / Айвазян С.А., Фантаццини Д. – М.: Магистр, НИЦ ИНФРА-М, 2018. – 944 с. – ISBN 978-5-9776-0333-1

107. Буреева Н.Н. Мноомерный статистический анализ с использованием ППП «Statistica». Учебно-методический матриал по программе повышения квалификации «Применение программных средств в научных исследованиях и преподавании математики и механики. – Нижний Новгород, 2007 – 112 с.

108. А.В. Кугаевских, Д.И. Муромцев, О.В. Кирсанова. Классические методы машинного обучения. – СПб: Университет ИТМО, 2022. – 53 с.

109. Чубукова, И. А. Data Mining : курс лекций / И. А. Чубукова. - Москва : ИНТУИТ, 2016. - 337 с. - ISBN 978-5-94774-819-2. - Текст : электронный. - URL: https://znanium.ru/catalog/product/2136992 (дата обращения: 26.03.2024).

110. Обеспечение экономической безопасности предприятия [Электронный ресурс] : учеб. пособие / авт.-сост.: С. А. Грачев, М. А. Гундорова ; Владим. гос. ун-т им. А. Г. и Н. Г. Столетовых. – Владимир : Изд-во ВлГУ, 2022. – 420 с. – ISBN 978-5-9984-1666-8

# ПРИЛОЖЕНИЯ

***Приложение 1***

Интегральная функция нормированного нормального распределения

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-t^2/2} dt$$

| Z | 0,08 | 0,06 | 0,04 | 0,02 | 0 |
|---|---|---|---|---|---|
| -3,5 | 0,00017 | 0,00019 | 0,0002 | 0,00022 | 0,00023 |
| -3,4 | 0,00025 | 0,00027 | 0,00029 | 0,00031 | 0,00034 |
| -3,3 | 0,00036 | 0,00039 | 0,00042 | 0,00045 | 0,00048 |
| -3,2 | 0,00052 | 0,00056 | 0,0006 | 0,00064 | 0,00069 |
| -3,1 | 0,00074 | 0,00079 | 0,00085 | 0,0009 | 0,00097 |
| -3 | 0,00104 | 0,00111 | 0,00118 | 0,00126 | 0,00135 |
| -2,9 | 0,0014 | 0,0015 | 0,0016 | 0,0017 | 0,0019 |
| -2,8 | 0,002 | 0,0021 | 0,0023 | 0,0024 | 0,0026 |
| -2,7 | 0,0027 | 0,0029 | 0,0031 | 0,0033 | 0,0035 |
| -2,6 | 0,0037 | 0,0039 | 0,0041 | 0,0044 | 0,0047 |
| -2,5 | 0,0049 | 0,0052 | 0,0055 | 0,0059 | 0,0062 |
| -2,4 | 0,0066 | 0,0069 | 0,0073 | 0,0078 | 0,0082 |
| -2,3 | 0,0087 | 0,0091 | 0,0096 | 0,0102 | 0,0107 |
| -2,2 | 0,0113 | 0,0119 | 0,0125 | 0,0132 | 0,0139 |
| -2,1 | 0,0146 | 0,0154 | 0,0162 | 0,017 | 0,0179 |
| -2 | 0,0188 | 0,0197 | 0,0207 | 0,0217 | 0,0228 |
| -1,9 | 0,0239 | 0,025 | 0,0262 | 0,0274 | 0,0287 |
| -1,8 | 0,0301 | 0,0314 | 0,0329 | 0,0344 | 0,0359 |
| -1,7 | 0,0375 | 0,0392 | 0,0409 | 0,0427 | 0,0446 |
| -1,6 | 0,0465 | 0,0485 | 0,0505 | 0,0526 | 0,0548 |
| -1,5 | 0,0571 | 0,0594 | 0,0618 | 0,0643 | 0,0668 |
| -1,4 | 0,0694 | 0,0721 | 0,0749 | 0,0778 | 0,0808 |
| -1,3 | 0,0838 | 0,0869 | 0,0901 | 0,0934 | 0,0968 |
| -1,2 | 0,1003 | 0,1038 | 0,1075 | 0,1112 | 0,1151 |
| -1,1 | 0,119 | 0,123 | 0,1271 | 0,1314 | 0,1357 |
| -1 | 0,1401 | 0,1446 | 0,1492 | 0,1539 | 0,1587 |
| -0,9 | 0,1635 | 0,1685 | 0,1736 | 0,1788 | 0,1841 |
| -0,8 | 0,1894 | 0,1949 | 0,2005 | 0,2061 | 0,2119 |
| -0,7 | 0,2177 | 0,2236 | 0,2297 | 0,2358 | 0,242 |
| -0,6 | 0,2483 | 0,2546 | 0,2611 | 0,2676 | 0,2743 |
| -0,5 | 0,281 | 0,2877 | 0,2946 | 0,3015 | 0,3085 |
| -0,4 | 0,3156 | 0,3228 | 0,33 | 0,3372 | 0,3446 |
| -0,3 | 0,352 | 0,3594 | 0,3669 | 0,3745 | 0,3821 |
| -0,2 | 0,3897 | 0,3974 | 0,4052 | 0,4129 | 0,4207 |
| -0,1 | 0,4286 | 0,4364 | 0,4443 | 0,4522 | 0,4602 |
| 0 | 0,4681 | 0,4761 | 0,484 | 0,492 | 0,5 |
| z | 0 | 0,02 | 0,04 | 0,06 | 0,08 |

| | | | | | |
|---|---|---|---|---|---|
| 0 | 0,5 | 0,508 | 0,516 | 0,5239 | 0,5319 |
| 0,1 | 0,5398 | 0,5478 | 0,5557 | 0,5636 | 0,5714 |
| 0,2 | 0,5793 | 0,5871 | 0,5948 | 0,6026 | 0,6103 |
| 0,3 | 0,6179 | 0,6225 | 0,6331 | 0,6406 | 0,648 |
| 0,4 | 0,6554 | 0,6628 | 0,67 | 0,6772 | 0,6844 |
| 0,5 | 0,6915 | 0,6985 | 0,7054 | 0,7123 | 0,719 |
| 0,6 | 0,7257 | 0,7324 | 0,7389 | 0,7454 | 0,7517 |
| 0,7 | 0,758 | 0,7642 | 0,7704 | 0,7764 | 0,7823 |
| 0,8 | 0,7881 | 0,7939 | 0,7995 | 0,8051 | 0,8106 |
| 0,9 | 0,8159 | 0,8212 | 0,8264 | 0,8315 | 0,8365 |
| 1 | 0,8413 | 0,8461 | 0,8505 | 0,8554 | 0,8599 |
| 1,1 | 0,8643 | 0,8686 | 0,8729 | 0,877 | 0,881 |
| 1,2 | 0,8849 | 0,8888 | 0,8925 | 0,8962 | 0,8997 |
| 1,3 | 0,9032 | 0,9066 | 0,9099 | 0,9131 | 0,9162 |
| 1,4 | 0,9192 | 0,9222 | 0,9251 | 0,9279 | 0,9306 |
| 1,5 | 0,9332 | 0,9357 | 0,9382 | 0,9406 | 0,9429 |
| 1,6 | 0,9452 | 0,9474 | 0,9495 | 0,9515 | 0,9535 |
| 1,7 | 0,9554 | 0,9573 | 0,9591 | 0,9608 | 0,9625 |
| 1,8 | 0,9641 | 0,9656 | 0,9671 | 0,9686 | 0,9699 |
| 1,9 | 0,9713 | 0,9726 | 0,9738 | 0,975 | 0,9761 |
| 2 | 0,9773 | 0,9783 | 0,9793 | 0,9803 | 0,9812 |
| 2,1 | 0,9821 | 0,983 | 0,9838 | 0,9846 | 0,9854 |
| 2,2 | 0,9861 | 0,9868 | 0,9875 | 0,9881 | 0,9887 |
| 2,3 | 0,9893 | 0,9898 | 0,9904 | 0,9909 | 0,9913 |
| 2,4 | 0,9918 | 0,9922 | 0,9927 | 0,9931 | 0,9934 |
| 2,5 | 0,9938 | 0,9941 | 0,9945 | 0,9943 | 0,9951 |
| 2,6 | 0,9953 | 0,9956 | 0,9959 | 0,9961 | 0,9963 |
| 2,7 | 0,9965 | 0,9967 | 0,9969 | 0,9971 | 0,9973 |
| 2,8 | 0,9974 | 0,9976 | 0,9977 | 0,9979 | 0,998 |
| 2,9 | 0,9981 | 0,9983 | 0,9984 | 0,9985 | 0,9986 |
| 3 | 0,99865 | 0,99874 | 0,99882 | 0,99889 | 0,99896 |
| 3,1 | 0,99903 | 0,9991 | 0,99915 | 0,99921 | 0,99926 |
| 3,2 | 0,99931 | 0,99936 | 0,9994 | 0,99954 | 0,99948 |
| 3,3 | 0,99952 | 0,99955 | 0,99958 | 0,99961 | 0,99964 |
| 3,4 | 0,99966 | 0,99969 | 0,99971 | 0,99973 | 0,99975 |
| 3,5 | 0,99977 | 0,99978 | 0,9998 | 0,99981 | 0,99983 |

## *Приложение 2*

# Критические точки распределения Стьюдента

| Число степеней свободы $k$ | Уровень значимости $\alpha$ (двусторонняя критическая область) | | | | | |
|---|---|---|---|---|---|---|
| | 0.10 | 0.05 | 0.02 | 0.01 | 0.002 | 0.001 |
| 1 | 6.31 | 12.7 | 31.82 | 63.7 | 318.3 | 637.0 |
| 2 | 2.92 | 4.30 | 6.97 | 9.92 | 22.33 | 31.6 |
| 3 | 2.35 | 3.18 | 4.54 | 5.84 | 10.22 | 12.9 |
| 4 | 2.13 | 2.78 | 3.75 | 4.60 | 7.17 | 8.61 |
| 5 | 2.01 | 2.57 | 3.37 | 4.03 | 5.89 | 6.86 |
| 6 | 1.94 | 2.45 | 3.14 | 3.71 | 5.21 | 5.96 |
| 7 | 1.89 | 2.36 | 3.00 | 3.50 | 4.79 | 5.40 |
| 8 | 1.86 | 2.31 | 2.90 | 3.36 | 4.50 | 5.04 |
| 9 | 1.83 | 2.26 | 2.82 | 3.25 | 4.30 | 4.78 |
| 10 | 1.81 | 2.23 | 2.76 | 3.17 | 4.14 | 4.59 |
| 11 | 1.80 | 2.20 | 2.72 | 3.11 | 4.03 | 4.44 |
| 12 | 1.78 | 2.18 | 2.68 | 3.05 | 3.93 | 4.32 |
| 13 | 1.77 | 2.16 | 2.65 | 3.01 | 3.85 | 4.22 |
| 14 | 1.76 | 2.14 | 2.62 | 2.98 | 3.79 | 4.14 |
| 15 | 1.75 | 2.13 | 2.60 | 2.95 | 3.73 | 4.07 |
| 16 | 1.75 | 2.12 | 2.58 | 2.92 | 3.69 | 4.01 |
| 17 | 1.74 | 2.11 | 2.57 | 2.90 | 3.65 | 3.95 |
| 18 | 1.73 | 2.10 | 2.55 | 2.88 | 3.61 | 3.92 |
| 19 | 1.73 | 2.09 | 2.54 | 2.86 | 3.58 | 3.88 |
| 20 | 1.73 | 2.09 | 2.53 | 2.85 | 3.55 | 3.85 |
| 21 | 1.72 | 2.08 | 2.52 | 2.83 | 3.53 | 3.82 |
| 22 | 1.72 | 2.07 | 2.51 | 2.82 | 3.51 | 3.79 |
| 23 | 1.71 | 2.07 | 2.50 | 2.81 | 3.59 | 3.77 |
| 24 | 1.71 | 2.06 | 2.49 | 2.80 | 3.47 | 3.74 |
| 25 | 1.71 | 2.06 | 2.49 | 2.79 | 3.45 | 3.72 |
| 26 | 1.71 | 2.06 | 2.48 | 2.78 | 3.44 | 3.71 |
| 27 | 1.71 | 2.05 | 2.47 | 2.77 | 3.42 | 3.69 |
| 28 | 1.70 | 2.05 | 2.46 | 2.76 | 3.40 | 3.66 |
| 29 | 1.70 | 2.05 | 2.46 | 2.76 | 3.40 | 3.66 |
| 30 | 1.70 | 2.04 | 2.46 | 2.75 | 3.39 | 3.65 |
| 40 | 1.68 | 2.02 | 2.42 | 2.70 | 3.31 | 3.55 |
| 60 | 1.67 | 2.00 | 2.39 | 2.66 | 3.23 | 3.46 |
| 120 | 1.66 | 1.98 | 2.36 | 2.62 | 3.17 | 3.37 |
| $\infty$ | 1.64 | 1.96 | 2.33 | 2.58 | 3.09 | 3.29 |
| | 0.05 | 0.025 | 0.01 | 0.005 | 0.001 | 0.0005 |
| | Уровень значимости $\alpha$ (односторонняя критическая область) | | | | | |

*Приложение 3*

# Критические точки распределения Фишера

($k_1$ и $k_2$ — число степеней свободы большей и меньшей дисперсии соответственно)

Уровень значимости $\alpha$ =0.01

| $k_2$ \ $k_1$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 4052 | 4999 | 5403 | 5625 | 5764 | 5889 | 5928 | 5981 | 6022 | 6056 | 6082 | 6106 |
| 2 | 98.49 | 99.01 | 90.17 | 99.25 | 99.33 | 99.30 | 99.34 | 99.36 | 99.36 | 99.40 | 99.41 | 99.42 |
| 3 | 34.12 | 30.81 | 29.46 | 28.71 | 28.24 | 27.91 | 27.67 | 27.49 | 27.34 | 27.23 | 27.13 | 27.05 |
| 4 | 21.20 | 18.00 | 16.69 | 15.98 | 15.52 | 15.21 | 14.98 | 14.80 | 14.66 | 14.54 | 14.45 | 14.37 |
| 5 | 16.26 | 13.27 | 12.06 | 11.39 | 10.97 | 10.67 | 10.45 | 10.27 | 10.15 | 10.05 | 9.96 | 9.89 |
| 6 | 13.74 | 10.92 | 9.78 | 9.15 | 8.75 | 8.47 | 8.26 | 8.10 | 7.98 | 7.87 | 7.79 | 7.72 |
| 7 | 12.25 | 9.55 | 8.45 | 7.85 | 7.46 | 7.19 | 7.00 | 6.84 | 6.71 | 6.62 | 6.54 | 6.47 |
| 8 | 11.26 | 8.65 | 7.59 | 7.01 | 6.63 | 6.37 | 6.19 | 6.03 | 5.91 | 5.82 | 5.74 | 5.67 |
| 9 | 10.56 | 8.02 | 6.99 | 6.42 | 6.06 | 5.80 | 5.62 | 5.47 | 5.35 | 5.26 | 5.18 | 5.11 |
| 10 | 10.04 | 7.56 | 6.55 | 5.99 | 5.64 | 5.39 | 5.21 | 5.06 | 4.95 | 4.85 | 4.78 | 4.71 |
| 11 | 9.86 | 7.20 | 6.22 | 5.67 | 5.32 | 5.07 | 4.88 | 4.74 | 4.63 | 4.54 | 4.46 | 4.40 |
| 12 | 9.33 | 6.93 | 5.95 | 5.41 | 5.06 | 4.82 | 4.65 | 4.50 | 4.39 | 4.30 | 4.22 | 4.16 |
| 13 | 9.07 | 6.70 | 5.74 | 5.20 | 4.86 | 4.62 | 4.44 | 4.30 | 4.19 | 4.10 | 4.02 | 3.96 |
| 14 | 8.86 | 6.51 | 5.56 | 5.03 | 4.69 | 4.46 | 4.28 | 4.14 | 4.03 | 3.94 | 3.86 | 3.80 |
| 15 | 8.68 | 6.36 | 5.42 | 4.89 | 4.56 | 4.32 | 4.14 | 4.00 | 3.89 | 3.80 | 3.73 | 3.67 |
| 16 | 8.53 | 6.23 | 5.29 | 4.77 | 4.44 | 4.20 | 4.03 | 3.89 | 3.78 | 3.69 | 3.61 | 3.55 |
| 17 | 8.40 | 6.11 | 5.18 | 4.67 | 4.34 | 4.10 | 3.93 | 3.79 | 3.68 | 3.59 | 3.52 | 3.45 |

Уровень значимости $\alpha$ =0.05

| $k_2$ \ $k_1$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 161 | 200 | 216 | 225 | 230 | 234 | 237 | 239 | 241 | 242 | 243 | 244 |
| 2 | 18.5 | 19.00 | 19.16 | 19.25 | 19:30 | 19.33 | 19.36 | 19.37 | 19.38 | 19.39 | 19.40 | 19.41 |
| 3 | 10.13 | 9.55 | 9.28 | 9.12 | 9.01 | 8.94 | 8.88 | 8.84 | 8.81 | 8.78 | 8.76 | 8.74 |
| 4 | 7.71 | 6.94 | 6.59 | 6.39 | 6.26 | 6.16 | 6.09 | 6.04 | 6.00 | 5.96 | 5.93 | 5.91 |
| 5 | 6.61 | 5.79 | 5.41 | 5.19 | 5.05 | 4.95 | 4.88 | 4.82 | 4.78 | 4.74 | 4.70 | 4.68 |
| 6 | 5.99 | 5.14 | 4.76 | 4.53 | 4.39 | 4.28 | 4.21 | 4.15 | 4.10 | 4.06 | 4.03 | 4.00 |
| 7 | 5.59 | 4.74 | 4.35 | 4.12 | 3.97 | 3.87 | 3.79 | 3.73 | 3.68 | 3.63 | 3.60 | 3.57 |
| 8 | 5.32 | 4.46 | 4.07 | 3.84 | 3.69 | 3.58 | 3.50 | 3.44 | 3.39 | 3.34 | 3.31 | 3.28 |
| 9 | 5.12 | 4.26 | 3.86 | 3.63 | 3.48 | 3.37 | 3.29 | 3.23 | 3.18 | 3.13 | 3.10 | 3.07 |
| 10 | 4.96 | 4.10 | 3.71 | 3.48 | 3.33 | 3.22 | 3.14 | 3.07 | 3.02 | 2.97 | 2.94 | 2.91 |
| 11 | 4.84 | 3.98 | 3.59 | 3.36 | 3.20 | 3.09 | 3.01 | 2.95 | 2.90 | 2.86 | 2.82 | 2.79 |
| 12 | 4.75 | 3.88 | 3.49 | 3.26 | 3.11 | 3.00 | 2.92 | 2.85 | 2.80 | 2.76 | 2.72 | 2.69 |
| 13 | 4.67 | 3.80 | 3.41 | 3.18 | 3.02 | 2.92 | 2.84 | 2.77 | 2.72 | 2.67 | 2.63 | 2.60 |
| 14 | 4.60 | 3.74 | 3.34 | 3.11 | 2.96 | 2.85 | 2.77 | 2.70 | 2.65 | 2.60 | 2.56 | 2.53 |
| 15 | 4.54 | 3.68 | 3.29 | 3.06 | 2.90 | 2.79 | 2.70 | 2.64 | 2.59 | 2.55 | 2.51 | 2.48 |
| 16 | 4.49 | 3.63 | 3.24 | 3.01 | 2.85 | 2.74 | 2.66 | 2.59 | 2.54 | 2.49 | 2.45 | 2.42 |
| 17 | 4.45 | 3.59 | 3.20 | 2.96 | 2.81 | 2.70 | 2.62 | 2.55 | 2.50 | 2.45 | 2.41 | 2.38 |

***Приложение 4***

## Таблица 5%-го и 1%-го уровней вероятности коэффициентов корреляции ( $r_a$ )

| Размер выборки | Положительные значения | | Отрицательные значения | |
|---|---|---|---|---|
| | 5%-й уровень | 1 %-й уровень | 5%-й уровень | 1 %-й уровень |
| 5 | 0,253 | 0,297 | -0,753 | -0,798 |
| 6 | 0,354 | 0,447 | -0,708 | -0,863 |
| 7 | 0,370 | 0,510 | -0,674 | -0,799 |
| 8 | 0,371 | 0,531 | -0,625 | -0,764 |
| 9 | 0,366 | 0,533 | -0,593 | -0,737 |
| 10 | 0,360 | 0,525 | -0,564 | -0,705 |
| 11 | 0,353 | 0,515 | -0,539 | -0,679 |
| 12 | 0,348 | 0,505 | -0,516 | -0,655 |
| 13 | 0,341 | 0,495 | -0,497 | -0,634 |
| 14 | 0,335 | 0,485 | -0,479 | -0,615 |
| 15 | 0,328 | 0,475 | -0,462 | -0,597 |
| 20 | 0,299 | 0,432 | -0,399 | -0,524 |
| 25 | 0,276 | 0,398 | -0,356 | -0,473 |
| 30 | 0,257 | 0,370 | -0,324 | -0,433 |
| 35 | 0,242 | 0,347 | -0,300 | -0,401 |
| 40 | 0,229 | 0,329 | -0,279 | -0,376 |
| 45 | 0,218 | 0,313 | -0,262 | -0,256 |
| 50 | 0,208 | 0,301 | -0,248 | -0,339 |

***Приложение 5***

## Распределение критерия Дарбина-Уотсона для положительной автокорреляции (для 5%-го уровня значимости)

| $n$ | $V=1$ | | $V=2$ | | $V=3$ | | $V=4$ | | $V=5$ | |
|---|---|---|---|---|---|---|---|---|---|---|
| | $d_1$ | $d_2$ | $d_1$ | $d_2$ | $d_1$ | $d_2$ | $d_1$ | $d_2$ | $d_1$ | $d_2$ |
| 15 | 1,08 | 1,36 | 0,95 | ,54 | 0,82 | 1,75 | 0,69 | 1,97 | 0,56 | 2,21 |
| 16 | 1,10 | 1,37 | 0,98 | 1,54 | 0,86 | 1,73 | 0,74 | 1,93 | 0,62 | 2,15 |
| 17 | 1,13 | 1,38 | 1,02 | 1,54 | 0,90 | 1,71 | 0,78 | 1,90 | 0,67 | 2,10 |
| 18 | 1,16 | 1,39 | 1,05 | 1,53 | 0,93 | 1,69 | 0,82 | 1,87 | 0,71 | 2,06 |
| 19 | 1,18 | 1,40 | 1,08 | 1,53 | 0,97 | 1,68 | 0,86 | 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 | 1,89 |
| 27 | 1,32 | 1,47 | 1,24 | 1,56 | 1,16 | 1,65 | 1,08 | 1,76 | 1,01 | 1,86 |
| 28 | 1,33 | 1,48 | 1,26 | 1,56 | 1,18 | 1,65 | 1,10 | 1,75 | 1,03 | 1,85 |
| 29 | 1,34 | 1,48 | 1,27 | 1,56 | 1,20 | 1,65 | 1,12 | 1,74 | 1,05 | 1,84 |
| 30 | 1,35 | 1,49 | 1,28 | 1,57 | 1,21 | 1,65 | 1,14 | 1,74 | 1,07 | 1,83 |
| 31 | 1,36 | 1,50 | 1,30 | 1,57 | 1,23 | 1,65 | 1,16 | 1,74 | 1,09 | 1,83 |
| 32 | 1,37 | 1,50 | 1,31 | 1,57 | 1,24 | 1,65 | 1,18 | 1,73 | 1,11 | 1,82 |
| 33 | 1,38 | 1,51 | 1,32 | 1,58 | 1,26 | 1,63 | 1,19 | 1,73 | 1,13 | 1,81 |
| 34 | 1,39 | 1,51 | 1,33 | 1,58 | 1,27 | 1,65 | 1,21 | 1,73 | 1,15 | 1,81 |
| 35 | 1,40 | 1,52 | 1,34 | 1,58 | 1,28 | 1,65 | 1,22 | 1,73 | 1,16 | 1,80 |
| 36 | 1,41 | 1,52 | 1,35 | 1,55 | 1,29 | 1,65 | 1,24 | 1,73 | 1,18 | 1,80 |
| 37 | 1,42 | 1,53 | 1,36 | 1,59 | 1,31 | 1,66 | 1,25 | 1,72 | 1,19 | 1,80 |
| 38 | 1,43 | 1,54 | 1,37 | 1,59 | 1,32 | 1,66 | 1,26 | 1,72 | 1,21 | 1,79 |
| 39 | 1,43 | 1,54 | 1,38 | 1,60 | 1,33 | 1,66 | 1,27 | 1,72 | 1,22 | 1,79 |
| 40 | 1,44 | 1,54 | 1,39 | 1,60 | 1,34 | 1,66 | 1,29 | 1,72 | 1,23 | 1,79 |
| 45 | 1,48 | 1,57 | 1,43 | 1,62 | 1,38 | 1,67 | 1,34 | 1,72 | 1,29 | 1,78 |
| 50 | 1,50 | 1,59 | 1,46 | 1,63 | 1,42 | 1,67 | 1,38 | 1,72 | 1,34 | 1,77 |
| 55 | 1,53 | 1,60 | 1,49 | 1,64 | 1,45 | 1,68 | 1,41 | 1,72 | 1,38 | 1,77 |
| 60 | 1,55 | 1,62 | 1,51 | 1,65 | 1,48 | 1,69 | 1,44 | 1,73 | 1,41 | 1,77 |
| 65 | 1,57 | 1,63 | 1,54 | 1,66 | 1,50 | 1,70 | 1,47 | 1,73 | 1,44 | 1,77 |
| 70 | 1,58 | 1,64 | 1,55 | 1,67 | 1,52 | 1,70 | 1,49 | 1,74 | 1,46 | 1,77 |
| 75 | 1,60 | 1,65 | 1,57 | 1,68 | 1,54 | 1,71 | 1,51 | 1,74 | 1,49 | 1,77 |
| 80 | 1,61 | 1,66 | 1,59 | 1,69 | 1,56 | 1,72 | 1,53 | 1,74 | 1,51 | 1,77 |
| 85 | 1,62 | 1,67 | 1,60 | 1,70 | 1,57 | 1,72 | 1,55 | 1,75 | 1,52 | 1,77 |
| 90 | 1,63 | 1,68 | 1,61 | 1,70 | 1,59 | 1,73 | 1,57 | 1,75 | 1,54 | 1,78 |
| 95 | 1,64 | 1,69 | 1,62 | 1,71 | 1,60 | 1,73 | 1,58 | 1,75 | 1,56 | 1,78 |
| 100 | 1,65 | 1,69 | 1,63 | 1,72 | 1,61 | 1,74 | 1,59 | 1,76 | 1,57 | 1,78 |

***Приложение 6***

Исходные данные для решения задачи кластерного анализа по показателям $x_1$-$x_5$

| | Использование ПК, % | Использование серверов, % | Использование облачных технологий, % | Использование сети Internet,% | Использование корпоративных сетей, % |
|---|---|---|---|---|---|
| Фирма1 | 88,561 | 38,994 | 58,2 | 94,09 | 32,98 |
| Фирма2 | 81,868 | 37,442 | 59,364 | 91,083 | 22,795 |
| Фирма3 | 84,778 | 44,717 | 61,789 | 92,829 | 31,816 |
| Фирма4 | 83,905 | 41,613 | 54,417 | 92,247 | 29,585 |
| Фирма5 | 84,099 | 41,904 | 56,357 | 92,247 | 29,1 |
| Фирма6 | 82,353 | 45,008 | 54,514 | 90,501 | 31,234 |
| Фирма7 | 82,45 | 35,696 | 54,029 | 87,591 | 22,407 |
| Фирма8 | 82,644 | 35,502 | 55,096 | 83,226 | 28,13 |
| Фирма9 | 85,651 | 39,673 | 54,708 | 89,628 | 27,936 |
| Фирма10 | 74,884 | 45,687 | 53,059 | 88,27 | 32,592 |
| Фирма11 | 81,577 | 36,666 | 50,149 | 87,979 | 21,534 |
| Фирма12 | 79,443 | 38,8 | 51,507 | 90,404 | 27,354 |
| Фирма13 | 78,667 | 39,091 | 51,41 | 90,889 | 25,511 |
| Фирма14 | 83,711 | 33,465 | 62,759 | 93,12 | 31,816 |
| Фирма15 | 79,637 | 37,927 | 48,015 | 87,688 | 28,033 |
| Фирма16 | 75,66 | 42,389 | 57,327 | 87,979 | 26,093 |
| Фирма17 | 79,928 | 44,232 | 59,752 | 92,053 | 29,294 |
| Фирма18 | 68,676 | 42,001 | 40,74 | 95,545 | 29,391 |
| Фирма19 | 80,219 | 36,181 | 53,544 | 90,501 | 23,862 |
| Фирма20 | 76,533 | 33,465 | 55,484 | 89,046 | 20,952 |
| Фирма21 | 80,316 | 37,733 | 58,297 | 89,628 | 28,809 |
| Фирма22 | 78,958 | 40,061 | 53,253 | 92,538 | 28,615 |
| Фирма23 | 73,429 | 38,218 | 48,112 | 87,688 | 26,384 |
| Фирма24 | 79,152 | 43,068 | 52,962 | 91,956 | 29,876 |
| Фирма25 | 78,667 | 41,031 | 60,528 | 92,732 | 25,802 |
| Фирма26 | 82,838 | 42,486 | 56,066 | 91,18 | 38,606 |
| Фирма27 | 78,667 | 35,987 | 50,925 | 89,046 | 27,548 |
| Фирма28 | 77,988 | 50,052 | 52,574 | 90,501 | 33,271 |
| Фирма29 | 80,413 | 39,382 | 54,805 | 88,367 | 30,555 |
| Фирма30 | 78,376 | 32,107 | 43,359 | 87,591 | 22,504 |
| Фирма31 | 72,071 | 37,442 | 43,553 | 88,076 | 27,742 |
| Фирма32 | 75,175 | 36,569 | 49,082 | 90,501 | 24,832 |
| Фирма33 | 77,794 | 39,091 | 53,156 | 91,374 | 25,026 |
| Фирма34 | 76,824 | 35,696 | 50,828 | 85,748 | 28,906 |
| Фирма35 | 75,272 | 38,315 | 51,41 | 91,18 | 27,742 |
| Фирма36 | 67,124 | 37,054 | 47,239 | 88,561 | 24,056 |
| Фирма37 | 67,221 | 24,638 | 25,996 | 59,461 | 21,34 |
| Фирма38 | 56,648 | 21,243 | 27,548 | 97 | 21,728 |

| | | | | | |
|---|---|---|---|---|---|
| Фирма39 | 51,022 | 21,437 | 24,832 | 84,972 | 25,123 |
| Фирма40 | 77,309 | 37,345 | 45,202 | 87,979 | 26,481 |
| Фирма41 | 77,503 | 35,696 | 44,62 | 70,228 | 27,742 |
| Фирма42 | 70,325 | 44,717 | 53,253 | 69,161 | 43,165 |
| Фирма43 | 87,203 | 48,209 | 60,528 | 92,441 | 27,936 |
| Фирма44 | 75,175 | 34,144 | 46,463 | 91,083 | 24,347 |
| Фирма45 | 76,048 | 36,569 | 53,641 | 91,956 | 27,548 |
| Фирма46 | 71,004 | 29,973 | 45,59 | 91,956 | 21,049 |
| Фирма47 | 74,884 | 38,509 | 47,918 | 95,739 | 26,966 |
| Фирма48 | 80,122 | 42,874 | 57,909 | 92,829 | 28,421 |
| Фирма49 | 78,861 | 40,352 | 62,177 | 93,12 | 29,779 |
| Фирма50 | 79,54 | 42,195 | 57,618 | 89,725 | 35,987 |
| Фирма51 | 76,339 | 35,114 | 52,962 | 90,986 | 23,571 |
| Фирма52 | 82,159 | 42,777 | 57,23 | 91,956 | 34,338 |
| Фирма53 | 84,778 | 38,606 | 58,103 | 93,799 | 27,257 |
| Фирма54 | 72,556 | 33,368 | 46,269 | 91,568 | 23,086 |
| Фирма55 | 73,526 | 39,188 | 51,992 | 86,524 | 26,287 |
| Фирма56 | 76,145 | 34,823 | 49,664 | 86,233 | 25,608 |
| Фирма57 | 75,369 | 33,756 | 50,343 | 87,591 | 25,802 |
| Фирма58 | 78,861 | 40,546 | 56,842 | 86,136 | 29,197 |
| Фирма59 | 81,965 | 45,008 | 57,424 | 93,023 | 29,488 |
| Фирма60 | 76,145 | 43,456 | 57,23 | 88,561 | 26,966 |
| Фирма61 | 80,704 | 42,583 | 55,969 | 90,598 | 27,548 |
| Фирма62 | 79,831 | 36,084 | 57,618 | 91,374 | 24,347 |
| Фирма63 | 76,145 | 28,227 | 39,867 | 77,309 | 23,668 |
| Фирма64 | 81,092 | 37,927 | 53,738 | 87,688 | 21,922 |
| Фирма65 | 80,51 | 37,151 | 52,477 | 92,247 | 28,809 |
| Фирма66 | 77,988 | 38,315 | 54,514 | 89,725 | 23,28 |
| Фирма67 | 77,406 | 34,726 | 50,246 | 83,323 | 28,518 |
| Фирма68 | 82,353 | 42,68 | 60,14 | 90,016 | 27,354 |
| Фирма69 | 81,286 | 39,964 | 51,895 | 87,785 | 29,682 |
| Фирма70 | 81,577 | 36,957 | 55,193 | 86,524 | 21,534 |
| Фирма71 | 81,868 | 47,918 | 61,692 | 82,741 | 30,555 |
| Фирма72 | 69,84 | 31,525 | 42,292 | 87,3 | 26,093 |
| Фирма73 | 83,905 | 38,121 | 49,761 | 86,524 | 31,622 |
| Фирма74 | 80,801 | 37,151 | 49,664 | 90,307 | 25,123 |
| Фирма75 | 77,891 | 46,463 | 58,006 | 93,12 | 21,437 |
| Фирма76 | 77,6 | 40,74 | 51,216 | 86,136 | 24,735 |
| Фирма77 | 79,346 | 42,971 | 55,096 | 90,404 | 22,698 |
| Фирма78 | 78,57 | 37,733 | 51,701 | 90,404 | 25,123 |
| Фирма79 | 81,383 | 47,918 | 60,043 | 93,411 | 26,287 |
| Фирма80 | 83,711 | 47,239 | 60,528 | 89,628 | 29,294 |
| Фирма81 | 79,637 | 33,174 | 54,029 | 89,725 | 21,437 |
| Фирма82 | 84,584 | 47,53 | 55,096 | 91,956 | 24,832 |

## *Приложение 7*

# Исходные данные для решения задачи кластерного анализа по показателям $x_6$-$x_9$

| | Доля используемых передовых производственных технологий, % | Объем выручки от использования ППТ, млн. руб. | Доля организация, внедривших в свою деятельность элементы ЭД, % | Использование фирмами элементов систем искусственного интеллекта, % |
|---|---|---|---|---|
| Фирма1 | 34,338 | 11,446 | 7,663 | 16,49 |
| Фирма2 | 17,654 | 4,559 | 3,298 | 10,767 |
| Фирма3 | 35,308 | 9,894 | 11,446 | 18,527 |
| Фирма4 | 29,488 | 8,73 | 10,282 | 14,259 |
| Фирма5 | 30,07 | 9,118 | 9,312 | 14,259 |
| Фирма6 | 33,077 | 10,476 | 9,797 | 17,266 |
| Фирма7 | 30,361 | 8,924 | 8,051 | 12,222 |
| Фирма8 | 22,31 | 6,693 | 3,686 | 12,319 |
| Фирма9 | 22,795 | 7,566 | 6,305 | 13,774 |
| Фирма10 | 46,851 | 12,319 | 12,222 | 22,213 |
| Фирма11 | 21,34 | 5,432 | 5,141 | 11,252 |
| Фирма12 | 22,601 | 6,887 | 6,984 | 13,386 |
| Фирма13 | 23,765 | 6,79 | 5,238 | 11,446 |
| Фирма14 | 27,451 | 8,148 | 7,178 | 15,132 |
| Фирма15 | 25,996 | 7,178 | 4,656 | 12,028 |
| Фирма16 | 36,084 | 11,446 | 7,469 | 19,497 |
| Фирма17 | 35,405 | 10,573 | 8,73 | 15,617 |
| Фирма18 | 27,742 | 9,409 | 5,432 | 12,222 |
| Фирма19 | 8,536 | 7,663 | 3,88 | 12,998 |
| Фирма20 | 18,236 | 6,305 | 3,201 | 12,222 |
| Фирма21 | 21,534 | 8,73 | 4,947 | 12,319 |
| Фирма22 | 29,003 | 9,021 | 10,088 | 16,005 |
| Фирма23 | 30,361 | 10,767 | 7,469 | 12,513 |
| Фирма24 | 33,562 | 13,386 | 7,566 | 16,878 |
| Фирма25 | 28,227 | 8,73 | 6,402 | 13,095 |
| Фирма26 | 41,516 | 9,118 | 9,118 | 20,079 |
| Фирма27 | 33,756 | 10,185 | 7,081 | 16,199 |
| Фирма28 | 32,592 | 11,737 | 6,693 | 15,035 |
| Фирма29 | 36,181 | 11,155 | 9,991 | 18,527 |
| Фирма30 | 24,444 | 5,917 | 6,79 | 12,61 |
| Фирма31 | 22,213 | 7,857 | 5,723 | 9,118 |
| Фирма32 | 36,375 | 11,252 | 6,693 | 15,229 |
| Фирма33 | 26,384 | 7,663 | 7,663 | 12,998 |
| Фирма34 | 30,264 | 10,185 | 8,051 | 12,028 |
| Фирма35 | 33,271 | 10,573 | 5,141 | 14,744 |
| Фирма36 | 22,407 | 8,245 | 3,589 | 11,446 |
| Фирма37 | 23,474 | 8,924 | 3,686 | 9,7 |
| Фирма38 | 29,682 | 9,215 | 5,529 | 9,797 |
| Фирма39 | 14,938 | 5,141 | 3,977 | 8,342 |
| Фирма40 | 32,98 | 9,797 | 7,954 | 14,259 |

| | | | | |
|---|---|---|---|---|
| Фирма41 | 32,786 | 10,379 | 7,372 | 15,423 |
| Фирма42 | 56,357 | 28,421 | 11,543 | 9,506 |
| Фирма43 | 29,682 | 11,155 | 8,342 | 17,363 |
| Фирма44 | 28,227 | 8,633 | 6,014 | 12,707 |
| Фирма45 | 26,384 | 8,245 | 8,924 | 12,222 |
| Фирма46 | 20,37 | 6,208 | 5,529 | 11,252 |
| Фирма47 | 25,026 | 9,603 | 6,402 | 15,908 |
| Фирма48 | 16,975 | 9,021 | 6,014 | 13,58 |
| Фирма49 | 30,361 | 10,185 | 6,402 | 15,326 |
| Фирма50 | 37,248 | 11,446 | 4,753 | 17,46 |
| Фирма51 | 19,4 | 6,402 | 6,596 | 10,088 |
| Фирма52 | 28,13 | 9,118 | 6,693 | 15,617 |
| Фирма53 | 27,451 | 9,118 | 7,178 | 13,095 |
| Фирма54 | 21,049 | 6,402 | 5,044 | 9,797 |
| Фирма55 | 29,197 | 8,342 | 5,82 | 15,326 |
| Фирма56 | 28,712 | 8,342 | 6,596 | 14,453 |
| Фирма57 | 24,735 | 7,178 | 7,857 | 12,61 |
| Фирма58 | 27,063 | 9,506 | 6,111 | 14,259 |
| Фирма59 | 38,412 | 11,64 | 6,208 | 16,684 |
| Фирма60 | 34,241 | 11,64 | 4,268 | 15,326 |
| Фирма61 | 18,527 | 9,312 | 6,499 | 16,975 |
| Фирма62 | 19,885 | 7,372 | 6,305 | 10,088 |
| Фирма63 | 13,289 | 5,335 | 3,492 | 6,111 |
| Фирма64 | 18,818 | 7,178 | 4,365 | 11,349 |
| Фирма65 | 25,026 | 7,178 | 4,365 | 12,901 |
| Фирма66 | 16,49 | 6,693 | 4,171 | 12,998 |
| Фирма67 | 36,278 | 10,379 | 1,843 | 19,788 |
| Фирма68 | 37,539 | 10,67 | 6,887 | 16,49 |
| Фирма69 | 27,257 | 9,7 | 6,79 | 13,289 |
| Фирма70 | 16,781 | 6,499 | 4,656 | 12,416 |
| Фирма71 | 31,428 | 10,961 | 8,342 | 17,557 |
| Фирма72 | 28,033 | 8,245 | 5,044 | 11,058 |
| Фирма73 | 23,571 | 11,64 | 2,328 | 13,968 |
| Фирма74 | 30,458 | 10,185 | 5,626 | 10,864 |
| Фирма75 | 25,996 | 9,7 | 6,596 | 11,446 |
| Фирма76 | 27,354 | 9,894 | 6,014 | 13,289 |
| Фирма77 | 25,317 | 9,409 | 2,619 | 13,774 |
| Фирма78 | 29,973 | 11,252 | 6,305 | 12,804 |
| Фирма79 | 30,943 | 11,931 | 6,693 | 13,677 |
| Фирма80 | 35,89 | 14,162 | 6,111 | 16,005 |
| Фирма81 | 32,592 | 8,051 | 6,014 | 11,834 |
| Фирма82 | 40,837 | 9,506 | 4,171 | 12,028 |

***Приложение 8***

Исходные данные для решения задачи кластерного анализа по показателям $x_1$-$x_5$

| | x1 | x2 | x3 | x4 | x5 |
|---|---|---|---|---|---|
| Фирма 1 | 28355,33 | 29608,58 | 30104,07 | 30808,78 | 32384,35 |
| Фирма 2 | 25400,38 | 25361,34 | 26428,4 | 26611,59 | 28399,37 |
| Фирма 3 | 23755,73 | 22875,85 | 24011,99 | 23562,54 | 25383,36 |
| Фирма 4 | 30139,11 | 29598,57 | 29356,33 | 30319,29 | 32054,02 |
| Фирма 5 | 22582,56 | 23702,68 | 24784,76 | 24527,5 | 25819,79 |
| Фирма 6 | 27577,55 | 28620,59 | 28136,11 | 29158,13 | 31425,39 |
| Фирма 7 | 22488,47 | 23993,97 | 24769,75 | 23739,72 | 25310,29 |
| Фирма 8 | 25839,81 | 25840,82 | 26451,43 | 27302,28 | 29178,15 |
| Фирма 9 | 27684,66 | 28483,46 | 29323,29 | 30040,01 | 32511,48 |
| Фирма 10 | 37659,62 | 40549,51 | 41327,29 | 44751,71 | 47248,2 |
| Фирма 11 | 22862,84 | 23260,24 | 24146,12 | 24919,9 | 26090,06 |
| Фирма 12 | 24243,22 | 24598,57 | 24813,79 | 25466,44 | 26912,89 |
| Фирма 13 | 24787,76 | 24493,47 | 25423,4 | 25913,89 | 27415,39 |
| Фирма 14 | 25101,08 | 26195,17 | 25963,94 | 26854,83 | 28182,15 |
| Фирма 15 | 23473,45 | 23906,88 | 24101,08 | 25150,13 | 27238,21 |
| Фирма 16 | 26312,29 | 27444,42 | 27801,77 | 27235,21 | 28585,56 |
| Фирма 17 | 27396,37 | 27846,82 | 27652,63 | 27082,06 | 28686,66 |
| Фирма 18 | 59957,9 | 59262,2 | 62594,53 | 68454,39 | 74127,05 |
| Фирма 19 | 25759,73 | 25769,74 | 26766,74 | 29179,15 | 30884,85 |
| Фирма 20 | 33361,33 | 31558,53 | 31212,18 | 33994,96 | 35391,36 |
| Фирма 21 | 32649,62 | 32535,5 | 33194,16 | 33864,83 | 35728,69 |
| Фирма 22 | 25627,6 | 27371,34 | 26515,49 | 27008,98 | 28362,33 |
| Фирма 23 | 25922,9 | 25928,9 | 26553,53 | 27488,46 | 28933,91 |
| Фирма 24 | 24743,72 | 27184,16 | 28026 | 31372,34 | 32338,31 |
| Фирма 25 | 36911,88 | 36151,12 | 37145,11 | 41605,56 | 44281,24 |
| Фирма 26 | 25805,78 | 25277,25 | 25517,49 | 25317,29 | 26029 |
| Фирма 27 | 21747,73 | 22117,1 | 23167,14 | 23903,88 | 25549,52 |

***Приложение 9***

Исходные данные для решения задачи кластерного анализа по показателям $x_6$-$x_{10}$

| | x6 | x7 | x8 | x9 | x10 |
|---|---|---|---|---|---|
| Фирма 1 | 19422,4 | 20887,87 | 22095,07 | 24620,6 | 26098,07 |
| Фирма 2 | 18777,76 | 19248,23 | 20649,63 | 22893,87 | 24494,47 |
| Фирма 3 | 16342,33 | 16785,77 | 17830,81 | 19780,76 | 20925,91 |
| Фирма 4 | 21243,22 | 22418,4 | 23848,83 | 26556,53 | 27931,9 |
| Фирма 5 | 15668,65 | 15891,88 | 17017 | 19426,41 | 20637,62 |
| Фирма 6 | 18364,35 | 19353,33 | 20366,35 | 23377,35 | 25085,06 |
| Фирма 7 | 14609,6 | 15913,9 | 16676,66 | 19588,57 | 20823,8 |
| Фирма 8 | 17617,6 | 18561,54 | 19362,34 | 21587,57 | 23047,02 |
| Фирма 9 | 20249,23 | 21335,31 | 22526,5 | 24943,92 | 26884,86 |
| Фирма 10 | 26416,39 | 28478,45 | 30908,88 | 35234,2 | 38009,97 |
| Фирма 11 | 16527,51 | 17261,24 | 18368,35 | 20237,22 | 21316,3 |
| Фирма 12 | 16013 | 16473,46 | 17585,57 | 19728,71 | 21214,19 |
| Фирма 13 | 17307,29 | 17198,18 | 18056,04 | 20653,63 | 22050,03 |
| Фирма 14 | 18388,37 | 19394,38 | 20113,09 | 21785,76 | 23202,18 |
| Фирма 15 | 17147,13 | 17703,69 | 18238,22 | 20011,99 | 21631,61 |
| Фирма 16 | 18181,16 | 18657,64 | 19997,98 | 22416,39 | 23965,94 |
| Фирма 17 | 17478,46 | 17891,87 | 19473,45 | 21335,31 | 23042,02 |
| Фирма 18 | 43841,8 | 44946,9 | 47644,6 | 54184,13 | 57164,11 |
| Фирма 19 | 18440,42 | 19308,29 | 20665,65 | 23756,73 | 25762,74 |
| Фирма 20 | 20912,89 | 20041,02 | 20922,9 | 23243,22 | 24663,64 |
| Фирма 21 | 21838,82 | 22884,86 | 24527,5 | 27689,66 | 29216,19 |
| Фирма 22 | 15433,42 | 15905,89 | 16902,89 | 19948,93 | 21041,02 |
| Фирма 23 | 17894,88 | 19037,02 | 20206,19 | 22856,83 | 23759,74 |
| Фирма 24 | 18494,48 | 20253,23 | 21617,6 | 24309,29 | 25860,84 |
| Фирма 25 | 24927,9 | 25424,4 | 26744,72 | 30729,7 | 32542,51 |
| Фирма 26 | 19775,76 | 20086,07 | 20650,63 | 22035,01 | 23198,18 |
| Фирма 27 | 16853,84 | 16951,94 | 18244,23 | 20053,03 | 21807,79 |

***Приложение 10***

Исходные данные для решения задачи кластерного анализа по показателям $x_{11}$-$x_{15}$

| | x11 | x12 | x13 | x14 | x15 |
|---|---|---|---|---|---|
| Фирма 1 | 694072,8 | 729812,9 | 786039,4 | 866845 | 956907,6 |
| Фирма 2 | 272054,3 | 281412,1 | 305564,2 | 329142,8 | 398112 |
| Фирма 3 | 368857,7 | 394168,9 | 413355,6 | 440983,5 | 537972 |
| Фирма 4 | 806775,6 | 818100,3 | 869158,9 | 944539,2 | 1003600 |
| Фирма 5 | 180698 | 179074,8 | 184991,8 | 198037,6 | 250005,6 |
| Фирма 6 | 340100,6 | 372717,4 | 416382,7 | 466453,5 | 545654,5 |
| Фирма 7 | 160740,4 | 158285,6 | 167112,2 | 180467,5 | 203129 |
| Фирма 8 | 337336,4 | 362756,2 | 387697,1 | 428869,7 | 497196,1 |
| Фирма 9 | 449443,3 | 484137 | 506846,8 | 581084,5 | 570950,4 |
| Фирма 10 | 3184106 | 3665962 | 3783843 | 4205971 | 5133568 |
| Фирма 11 | 208446,1 | 215571,9 | 215361,7 | 230936,9 | 265938,4 |
| Фирма 12 | 323454,9 | 334633,4 | 361292,9 | 383493,3 | 436479,2 |
| Фирма 13 | 256963,5 | 263564,9 | 291774,9 | 313169,9 | 348409,6 |
| Фирма 14 | 317530,9 | 298037,9 | 299089,9 | 331962,8 | 354656,1 |
| Фирма 15 | 329945,6 | 361883,7 | 387912,4 | 442095,3 | 485651,8 |
| Фирма 16 | 478015,3 | 519205,9 | 557329,6 | 636769,8 | 682293,9 |
| Фирма 17 | 443497,2 | 472816,3 | 511647,7 | 561138,5 | 607427,5 |
| Фирма 18 | 13534384 | 14251989 | 15703970 | 17899398 | 19692677 |
| Фирма 19 | 212261,5 | 231668,9 | 252087,5 | 280292,4 | 325509,9 |
| Фирма 20 | 528931,8 | 548213,1 | 576227,8 | 666401,4 | 721386 |
| Фирма 21 | 628325,8 | 681162,8 | 726730,8 | 820066,2 | 891056,7 |
| Фирма 22 | 479371,9 | 477697,6 | 509276,5 | 583213 | 630767,8 |
| Фирма 23 | 350168,4 | 385884,6 | 417704,4 | 461315,8 | 520244,2 |
| Фирма 24 | 850466,2 | 917369,1 | 964767,9 | 1105540 | 1225739 |
| Фирма 25 | 401984,3 | 432795,2 | 443052,2 | 483030,4 | 617525,9 |
| Фирма 26 | 234309,8 | 243636,3 | 252902,9 | 262270 | 273817 |
| Фирма 27 | 135374,7 | 145699,7 | 151670 | 164392,7 | 197326,7 |

*Учебное электронное издание*

СТАТИСТИЧЕСКИЕ МЕТОДЫ В УПРАВЛЕНИИ ФИРМОЙ

Учебное пособие

Авторы-составители:
ФРАЙМОВИЧ Денис Юрьевич
БЫКОВА Маргарита Леонидовна

*Издается в авторской редакции*

***Системные требования:*** Intel от 1,3 ГГц; Windows XP/7/8/10;
Adobe Reader; дисковод CD-ROM.

**Тираж 25 экз.**

Владимирский государственный университет
имени Александра Григорьевича и Николая Григорьевича Столетовых
Изд-во ВлГУ
rio.vlgu@yandex.ru

Институт экономики и туризма
кафедра экономики инноваций и финансов
margarita93@bk.ru