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

В. Е. КРЫЛОВ Н. В. МУРАВЬЕВА

ОБЩАЯ ТЕОРИЯ СТАТИСТИКИ

Учебное пособие

ВлГУ
ИЗДАТЕЛЬСТВО

Владимир 2020

УДК 311
ББК 60.6
К85

Рецензенты:
Кандидат экономических наук, доцент
зав. кафедрой экономики и финансов Финансового университета
при Правительстве Российской Федерации (Владимирский филиал)
Д. В. Кузнецов

Генеральный директор ООО «Хрустальное небо»
В. Н. Козырев

К85 **Крылов, В. Е.** Общая теория статистики : учеб. пособие / В. Е. Крылов, Н. В. Муравьева ; Владим. гос. ун-т им. А. Г. и Н. Г. Столетовых. – Владимир : Изд-во ВлГУ, 2020. – 243 с.
ISBN 978-5-9984-1113-7

Содержит все необходимые дидактические требования, предусмотренные Примерной программой дисциплины «Статистика», разработанной в соответствии с требованиями Государственного образовательного стандарта третьего поколения («3+» и «3++»). Пособие может быть использовано при изучении ряда смежных дисциплин, а также не только в учебных целях, но и в практической работе организаций, осуществляющих экономическую деятельность.

Предназначено для студентов всех форм обучения и всех направлений подготовки укрупненной группы специальностей 38.00.00 «Экономика и управление» (уровень бакалавриата). Содержание пособия полностью соответствует требованиям Государственного образовательного стандарта, Государственным требованиям к минимуму содержания и уровню подготовки выпускников по укрупненной группе специальностей 38.00.00 «Экономика и управление».

Ил. 51. Табл. 68. Библиогр.: 20 назв.

УДК 311
ББК 60.6

ISBN 978-5-9984-1113-7

ВВЕДЕНИЕ

Современные реалии диктуют нам свои условия. Огромное информационное пространство и взаимодействие общества в различных сферах деятельности усилило интерес к статистике и как к науке, и как к сфере деятельности в целом.

Статистика владеет огромным количеством методов, позволяющих проанализировать эмпирические данные, на основании этих данных построить модель развития экономического показателя и на основании модели дать прогноз его развития. Изучению этих методов посвящено учебное пособие.

Текст пособия разбит на параграфы, каждый из которых, в свою очередь, разделен на пункты. Для удобства нумерация пунктов в тексте пособия сквозная. После теоретического материала приведены практические задачи, позволяющие обучающемуся на практике отработать знания, умения и навыки. Необходимые для решения задач справочные таблицы приведены в приложении.

Определения, формулировки теорем и свойств выделены в тексте курсивом.

Понятия, которые встречаются в тексте первый раз или положения курса, на которые следует обратить внимание, выделены жирным шрифтом. Начало доказательства в большинстве случаев обозначается символом $\triangleright$, а конец – символом $\square$.

Текст пособия снабжен достаточным количеством подробно решенных примеров и задач, позволяющих самостоятельно разобраться в материале. Для его усвоения необходимы стандартные знания линейной алгебры, аналитической геометрии, математического анализа, а также теории вероятностей и математической статистики.

Параграф 1 посвящен рассмотрению общих вопросов. В нем рассматриваются определения понятия «статистика», исследуются цели и основные задачи статистики. Рассматриваются основные термины статистки. Далее вводятся основные принципы и методы статистики. Кратко излагается история возникновения и развития статистики в мире и в России. Отдельное внимание уделено организации статистики в Российской Федерации на современном этапе.

В параграфе 2 рассматривается основной метод организации статистического наблюдения – выборочный метод. Определяются основные способы сводки и группировки эмпирических данных, а также их графическое представление. Далее вводятся основные числовые характеристики, позволяющие анализировать выборочную совокупность. Завершается параграф изложением методики определения интервальных оценок основных характеристик генеральной совокупности, полученных по соответствующим точечным характеристикам выборочной совокупности.

В параграфе 3 вводится алгоритм проверки статистических гипотез. Он приведен на примерах проверок гипотез о распределении генеральной совокупности по нормальному закону и закону Пуассона.

Очень важным в курсе представляется параграф 4. В нем изучаются вопросы построения моделей зависимости между двумя показателями. Описываются способы оценки формы, тесноты связи между признаками. Вводятся показатели точности моделей, а также методы отбора наиболее точной модели. Параграф 4 завершается подробным алгоритмом построения прогноза, начиная с обработки эмпирических данных и заканчивая прогнозом результативного показателя по данному факторному признаку.

Обобщением параграфа 4 является параграф 5. В нем излагается методика построения прогноза не по двухфакторной, а по многофакторной регрессионной модели. Этот параграф, как и предыдущий, завершается алгоритмом построения прогноза.

Ряды динамики, являющиеся частным случаем парной регрессии, рассматриваются в параграфе 6. Этот параграф носит, прежде всего, прикладной характер. Его изучение позволит решать множество практических задач, связанных не только с экономикой. Основная задача – получить математический аппарат, позволяющий осуществить прогноз (форсайт) показателя в определенный момент времени. Для этого определяются факторы (случайные и неслучайные), оказывающие влияние на формирование значений уровней ряда динамики. На основании этих факторов строятся аналитические модели ряда динамики, из них выбирается наиболее точная модель и по ней осуществляется прогноз.

Заключительный параграф 7. Он посвящен индексному методу, позволяющему получить относительные характеристики сравниваемых явлений. Индексный метод является важным статистическим инструментом, которым должен владеть экономист.

§ 1. СТАТИСТИКА: ОСНОВНЫЕ ПОНЯТИЯ

1.1. Термин «статистика»

В современном обществе важную роль в механизме управления экономикой выполняет статистика. Независимо от уровня и стадии экономического развития, характера политической системы, статистика на протяжении сотен лет своего существования всегда выступала как необходимый и эффективный инструмент государственного управления и одновременно как наука, исследующая количественную сторону массовых явлений.

Особенность статистики заключается в том, что статистические данные сообщаются в количественной форме, т. е. статистика говорит языком цифр, отображающих общественную жизнь во всем многообразии ее проявлений. При этом статистику прежде всего интересуют те выводы, которые можно сделать на основе анализа надлежащим образом собранных и обработанных цифровых данных.

Выполняя самые разнообразные функции сбора, систематизации и анализа сведений, характеризующих экономическое и социальное развитие общества, она всегда играла роль главного поставщика факторов для управленческих, научно-исследовательских и прикладных практических нужд различного рода структур, организаций и населения.

Понятие «статистика» очень многогранно, нельзя остановиться на какой-то одной трактовке. Сейчас существует более тысячи определений этого слова в зависимости от того, в широком или узком смысле требуется объяснение. Статистика как термин корни свои берет от латинского слова «статус» – определенное положение вещей. Приведем несколько примеров известных определений статистики: Статистика – обширная область знаний, включающая в себя знания из математики, физики, экономики, излагающая вопросы сбора, сдвиги количественных изменений в процессах. В некоторых случаях статистику определяют, как совокупность цифровых сведений, характеризующих определенные явления или процессы в жизни. Статистика выступает и как отрасль деятельности по сбору, обработке, анализу и публикации определенных сведений. Есть определение статистики как параметра ряда случайных величин. Еще раз отметим, что определение зависит от того, в каком контексте вы хотите использовать данный термин.

Мы придерживаемся следующего определения. Статистикой называется отрасль знаний, объединяющая принципы и методы работы с числовыми данными, или отрасль практической деятельности, направленной на сбор, обработку, анализ и интерпретацию числовых данных, характеризующих массовые явления.

Итак, статистика имеет дело, прежде всего с количественной стороной явлений и процессов общественной жизни.

Данные статистических справочников языком цифр характеризуют размеры и количественные соотношения (объемы, структуру, темпы развития и т.п.) явлений общественной жизни и проявляющиеся в них закономерности. Общей чертой сведений, составляющих статистику, служит то, что они всегда относятся не к одному единичному (индивидуальному) явлению, а охватывают сводными характеристиками целый ряд таких явлений или, как говорят, их совокупность.

Индивидуальное явление отличается от совокупности своей неразложимостью на самостоятельно существующие и аналогичные друг другу составные элементы. Совокупность же состоит именно из таких элементов. Исчезновение одного из элементов совокупности не уничтожает ее как таковую. Так, население страны остается ее населением, даже если одно из входящих в его состав лицо покинуло страну и переселилось за ее пределы.

Разные совокупности и их единицы в реальности сочетаются и переплетаются друг с другом подчас в весьма сложных комплексах. Так, говоря о промышленности, статистика рассматривает ее как совокупность предприятий, каждое из которых образует одну из входящих в нее единиц. Обратившись далее к исследованию предприятия, мы находим на нем совокупность рабочих, станков и т.п. В совокупности станков отдельный станок образует одну из единиц, но производимую на нем продукцию можно представить, как совокупность изделий и т. д.

Специфическая черта статистики состоит в том, что во всех случаях ее данные относятся к совокупности. Характеристики отдельных индивидуальных явлений попадают в поле ее зрения лишь в качестве основания для получения сводных характеристик совокупности. В этом состоит связь учета (бухгалтерского, первичного, хозяйственного и т.п.). Например, регистрация брака имеет определенное значение для данной конкретной пары, вступающей в него. К статистике же относятся лишь сводные данные о числе заключенных браков. Величина прибыли от конкретной сделки интересует конкретного предпринимателя как составляющая его дохода. Сумма прибыли от всех сделок является статистической характеристикой финансовых результатов деятельности предприятий.

Таким образом, статистику образуют сводные характеристики совокупностей объектов и явлений, относящихся к жизни общества, или, шире, тех или иных совокупностей вообще.

В ряде случаев термин «статистика» употребляется в несколько более узком смысле, связанном с обработкой результатов серии индивидуальных наблюдений. «Статистикой» называют некоторый параметр у, зависящий от значений x_1, x_2, …, x_n случайной величины X. В этом смысле термин «статистика» применяется главным образом в математической статистике.

Под статистикой понимают также процесс собирания и обработки данных, необходимых для получения статистики в обоих рассмотренных смыслах.

1.2. Предмет статистики, ее задачи

Предметом статистики является исследование количественной стороны массовых общественных явлений в неразрывной связи с их качественной стороной, или их содержанием, в конкретных условиях места и времени.

Главной задачей статистики на современном этапе является совершенствование экономического анализа статистической информации.

Перечислим другие задачи статистики.

1. Обеспечение органов управления государством, регионами, отраслями и отдельными предприятиями своевременной полной и достоверной информацией, необходимой для принятия решения.

2. Информирование общественности о явлениях и процессах, происходящих в обществе.

3. Получение объективной информации о деятельности хозяйственных структур с учетом теневого сектора.

4. Создание автоматизированных баз данных о деятельности текущих хозяйственных структур с возможностью санкционированного доступа к ним для получения информации, необходимой для решения текущих хозяйственных задач.

5. Прогнозирование развития важных социально-экономических процессов и явлений.

6. Распространение выборочных обследований во всех секторах общественной и экономической жизни.

7. Проведение организационно-методологической работы по постепенному переходу на систему национальных счетов.

8. Приведение системы статистических показателей к сопоставимому виду при международных сравнениях.

9. Совершенствование статистической отчетности.

10. Повышение достоверности статистических данных.

11. Изучение уровня и структуры массовых социально-экономических явлений и процессов.

12. Изучение динамики массовых социально-экономических явлений и процессов.

13. Выявление взаимосвязей между социально-экономическими явлениями и процессами.

14. Развитие информационной системы государственной статистики, взаимодействующей с другими информационными системами страны.

15. Обеспечение сбора, хранения, обработки и защиты официальной статистической информации.

16. Проведение работ по обеспечению конфиденциальности статистической информации, предоставляемой предприятиями и организациями в органы государственной и ведомственной статистики.

Отдельно рассмотрим задачи статистического исследования. Статистическое исследование включает в себя: разработку программы статистического наблюдения (определение объекта, единицы и формы наблюдения, разработку методик расчета запрашиваемых показателей и предполагаемые результаты обработки полученных данных); сбор массовых данных о статистической совокупности (непосредственно статистическое наблюдение); обработку данных (сводку, группировку); анализ полученной информации.

Таким образом, в задачи статистического исследования входят: разработка методологии статистического изучения того или иного процесса или явления; проведение статистического наблюдения; осуществление статистического анализа полученных результатов наблюдения.

Последняя и важнейшая ступень статистического исследования, является статистический анализ. Задачи статистического анализа: определение уровня или масштабы исследуемого явления или процесса; характеристика структуры наблюдаемого объекта; исследование динамики явления для выявления закономерности происходящих процессов; сравнительный анализ исследуемого объекта наблюдения с аналогичными (с зарубежными, с нормативом и т.п.); выявление взаимосвязи основного объекта исследования с другими объектами.

Различают экономическую и социальную статистику.

Экономическая статистика занимается изучением экономических процессов с точки зрения анализа развития экономики в обществе. Здесь собирают, обрабатывают и анализируют информацию о данной сфере общества. С помощью экономической статистики можно оценить положение экономики в целом по стране.

Социальная статистика исследует изменения социального характера, количественные и качественные. Общество, а также совокупность процессов и явлений, с ним связанных, являются предметом для изучения. Социальная статистика позволяет выявить развитие социальных условий населения. Структура населения, его состав, уровень доходов населения, бытовые условия – лишь небольшой перечень показателей для социальной статистики.

1.3. Основные понятия статистики

Основными понятиями статистики являются: статистическая совокупность, единица совокупности, признак, статистический показатель, система статистических показателей.

Статистическая совокупность – *это совокупность социально-экономических объектов или явлений общественной жизни, объединенных некоторой качественной основой, общей связью, но отличающимися друг от друга отдельными признаками.*

Примерами статистической совокупности могут служить: совокупность семей, совокупность предприятий, организаций, фирм и т.п.

Различают однородные и разнородные статистические совокупности. *Совокупность называется* ***однородной****, если один или несколько изучаемых признаков ее объектов являются общими для всех единиц. Если в совокупность входят явления разного типа, считается* ***разнородной***.

Единицей совокупности *называется ее первичный элемент, являющийся носителем признаков, подлежащих регистрации, и основой ведущегося при обследовании счета.*

Признак *статистической совокупности - качественная особенность ее единицы*.

По характеру отображения свойств единиц изучаемой совокупности, признаки делятся на две основные группы.

Первая группа признаков: признаки, имеющие непосредственное количественное выражение (например, возраст, стаж работы, средний заработок и т. д.). Они могут быть ***дискретными*** *и* ***непрерывными***.

Вторая группа признаков: признаки, не имеющие непосредственного количественного выражения. В этом случае отдельные единицы совокупности различаются своим содержанием. Например, профессии различаются характером труда: учитель, столяр, швея и т. д. Такие признаки обычно называют ***атрибутивными***.

Особенностью статистических исследований является то, что в нем изучаются только ***варьирующие признаки***, т. е. признаки, принимающие различные значения или имеющие различные количественные уровни у отдельных единиц совокупности. ***Вариация*** – это изменение («колеблемость») величины либо значение признака при переходе от одной единицы совокупности к другой.

Статистический показатель *– это понятие, отображающее количественные характеристики (размеры) соотношения признаков общественных явлений*. Статистические показатели могут быть ***объемными*** (численность населения) и ***расчетными*** (средний возраст).

Система статистических показателей *– это совокупность статистических показателей, отражающая взаимосвязи, которые объективно существуют между явлениями*. Для каждой общественно-экономической формации характерна определенная система взаимосвязи общественных явлений.

Система статистических показателей охватывает все стороны жизни общества на различных уровнях: макроуровень (страна, регион) и микроуровень (предприятие, семья).

Системы статистических показателей имеют следующие особенности: они носят исторический характер, статистические показатели меняются при изменении условий жизни населения; методология статистических показателей непрерывно совершенствуется.

1.4. Принципы организации статистической деятельности

Долгое время в период перехода России на рыночные отношения статистическое ведомство в России руководствовалось в своей деятельности основополагающими принципами официальной статистики в странах с экономикой переходного периода, одобренными Статистической комиссией ООН в 1994 году. Этот документ аккумулировал все главные принципы, которыми руководствовались страны с рыночной экономикой и которые отражались в их законах о статистике. Эти принципы легли в основу Федерального закона «Об официальном статистическом учете и системе государственной статистики в Российской Федерации» №282-ФЗ.

Принцип 1. Актуальность, объективность и доступность. Официальные статистические данные, отвечающие требованиям практической значимости, должны составляться и предоставляться официальными органами статистики на основе беспристрастности для обеспечения права граждан на открытое получение информации.

Принцип 2. Профессионализм. Решение о методах и порядке сбора, обработки, хранения и предоставления статистических данных принимается, строго руководствуясь принципами профессионализма, включая научные принципы и профессиональную этику.

Принцип 3. Использование статистических стандартов. Для обеспечения правильной трактовки данных статистические ведомства должны предоставлять информацию в соответствии с научными стандартами, установленными для источников, методов и порядка составления статистики.

Принцип 4. Правильное использование и интерпретация статистических данных. Статистические ведомства имеют право комментировать искажения в трактовке и использовании статистических данных.

Принцип 5. Эффективность статистических наблюдений. Статистические данные могут собираться из любых источников информации, будь то статистические обзоры или данные административного характера. При этом источники данных должны выбираться с учетом качества, своевременности, эффективности затрат, а также нагрузки на респондентов.

Принцип 6. Конфиденциальность. Индивидуальные данные, собираемые статистическими ведомствами, должны быть строго конфиденциальными и использоваться исключительно в целях статистики, независимо от того, касаются ли они физических или юридических лиц.

Принцип 7. Законодательство и гласность. Законы, нормативно-правовые документы, определяющие работу статистической системы, подлежат обнародованию.

Принцип 8. Координация на национальном уровне. Для достижения последовательности и эффективности функционирования статистической системы в рамках государств должна проводится координация работы статистических ведомств.

Принцип 9. Координация на международном уровне. Использование статистическими ведомствами каждой отдельной страны международных концепций, классификаций и методов способствует достижению последовательности и эффективности работы статистических систем на всех официальных уровнях.

Принцип 10. Международное сотрудничество в области статистики. Двустороннее и многостороннее сотрудничество в области статистики способствует усовершенствованию официальных статистических систем во всех странах.

Основные принципы официальной статистики ориентируют национальные статистические ведомства на обеспечение высокого качества статистической информации, объективность и доверие со стороны общества.

1.5. Основные методы статистики

Несмотря на разнообразие сфер применения статистики, имеются общие методы статистической работы.

Метод статистики – это целая совокупность приемов, пользуясь которыми статистика исследует свой предмет. Она включает в себя три основных метода.

Первым методом считается ***статистическое наблюдение***, которое заключается в сборе первичного статистического материала, в научно организованной регистрации всех существующих фактов, относящихся к рассматриваемому объекту.

Второй метод называется методом ***группировок***. Он дает возможность все собранные в результате массового статистического наблюдения все факты подвергать систематизации и классификации.

Третий метод – метод ***обобщающих показателей***. Он позволяет характеризовать изучаемые явления и процессы при помощи статистических величин (абсолютных, относительных, средних).

Поскольку статистика имеет дело с количественными характеристиками, она широко применяет в своих исследованиях положения и методы математики. Особенно широкое применение находят в статистике теория вероятностей и математическая статистика, которые занимаются изучением абстрактных множеств единиц и действующих в них общих количественных закономерностей. Установленные этими отраслями математики законы, правила и методы статистика использует при решении своих специфических задач. В частности, важную роль играет в статистике закон больших чисел. Закон больших чисел выражает общий принцип, в силу которого в большом числе явлений при некоторых общих условиях почти устраняется влияние случайного фактора. Достигается это в результате того, что в большом числе случаев происходит взаимопогашение индивидуальных отклонений величин одного и того же вида от общей их меры.

Опираясь на закон больших чисел, статистика выявляет характерные для определенных условий закономерности, типичные количественные соотношения и уровни явлений.

1.6. Основные этапы развития статистики как науки в мире

Слово «статистика» происходит от латинского слова status – состояние, положение вещей. Первоначально оно употреблялось в значении «политическое состояние». Отсюда произошло итальянское слово stato – государство и statista – знаток государства. В научный обиход слово «статистика» вошло в XVIII веке и первоначально употреблялось в значении «государствоведение».

Исторически развитие статистики было связано с развитием государств, с потребностями государственного управления. Хозяйственные и военные нужды уже в древний период развития человечества требовали наличия данных о населении, его составе, имущественном положении. С целью налогообложения организовывались переписи населения, проводился учет земель и т. д.

Первые статистические сведения содержатся в китайских документальных источниках еще во II веке до нашей эры.

В Древнем Риме проводились учеты свободных граждан и их имущества. Учёт осуществлялся по полу и возрасту, собирались сведения о состоянии промышленности и сельского хозяйства. В античном мире был организован учет родившихся: молодые люди, достигшие 18 лет, вносились в списки военнообязанных, а по достижении 20 лет – в списки полноправных граждан. Составлялись земельные кадастры, в которые вносились сведения о строениях, рабах, скоте, инвентаре, получаемых доходах.

Появились описания государств. За триста лет до нашей эры Аристотель (384 – 322 гг. до н.э.) составил описание 157 городов и государств своего времени.

«Книга страшного суда» (1086 г.) – свод материалов всеобщей переписи населения Англии и его имущества (включает данные о 240000 дворов). Перепись была завершена к концу 1086 г. и её результаты в виде огромного массива списков и отчётов были предоставлены королю. В дальнейшем они хранились в казначействе Английского королевства в Винчестере. Помимо этого уже к 1088 г. на основании этой документации были составлены два тома «Книги страшного суда», в которые в компактной форме вошли важнейшие сведения, полученные в результате переписи, отсортированные по графствам: имена владельцев поместья на дату проведения переписи и на 1066 г.; имена иных держателей поместья, если владелец передавал его в условное держание; площадь пахотной земли; количество пахотных бригад (измеряемых упряжками из восьми быков) на домениальных землях владельца и на землях крестьян; количество крестьян различных категорий (вилланов, коттариев, сервов, свободных и сокменов), проживающих на территории поместья; размеры пастбищ, лугов и лесов, относящихся к поместью; количество мельниц и мест для рыболовства; денежная оценка хозяйства поместья на дату проведения переписи и на 1066 г.; размеры наделов свободных крестьян и сокменов в границах поместья по состоянию на дату проведения переписи и на 1066 г.; потенциальная возможность повышения продуктивности поместья.

Возникновение и развитие международной торговли вызвало потребность в информации об иностранных государствах, их населении, городах, торговле и т. д. Такого рода сведения стало собирать с XIII в. правительство Венецианской республики.

В XVI в. в Венеции, Голландии появляются сборники, характеризующие политическое устройство, население, основные занятия, производимую продукцию в странах, с которыми устанавливалась торговля. В дальнейшем и другие страны стали издавать такие же справочники.

Со временем собирание данных о массовых общественных явлениях приобрело регулярный характер. С середины XIX в. были выработаны первые правила переписей населения и начата регулярность их проведения в развитых странах.

В процессе практических статистических работ начали складываться определённые правила сбора и обработки данных, приёмы анализа информации. Появляется необходимость теоретического научного осмысления накопленной практики. Начали складываться и исторические черты познания массовых явлений и формы их количественного измерения.

У истоков статистики как науки, а не только практической деятельности стояли две школы: английская научная школа политических арифметиков и немецкая описательная школа.

Английская научная школа политических арифметиков возникла в середине XVIIв. и ставила целью изучать общественные явления с помощью числовых характеристик. В центре исследования были статистические методы, теория статистики. Явления изучались не в статике, а в динамике. Предметом статистического изучения являлись не отдельные, а массовые общественные явления, поскольку закономерность может проявиться лишь при достаточно большом объёме анализируемой совокупности.

Школа английских арифметиков имела два направления: демографическое, представленное Д. Граунтом (1620 – 1674) и Э. Галлеем (1656 – 1742), и статистико-экономическое, разработанное В. Петти (1623 – 1687).

Английские учёные впервые не описывали социально-экономические явления, а давали им числовую оценку. Конкретными цифрами они стремились охарактеризовать состояние и развитие общества, показать закономерности развития общественных явлений на основе изучения массовых данных. Наибольшее развитие школа политических арифметиков получила в XVII и XVIIIвв. в Англии, Голландии, Франции. История показала, что именно эта научная школа явилась истоком современной теории статистики.

Представители описательной статистики стремились систематизировать существующие способы описания государств, создать теорию плохого описания, разработать её детальную схему. Однако они вели описание только в словесной форме, без цифр, вне динамики и связи явлений, т. е. без отражения особенностей развития государства. Собирался информационный материал, который впоследствии не анализировался. Описывался последний период, предмет и методы науки не были чётко определены. В трудах немецких учёных описывались государства, их устройство, быт и нравы населения, климат, финансы, армия, религия.

Основанием описательной школы был немец Г. Конрринг (1606 – 1681), который разработал систему описания государственного устройства. Дальнейшее развитие направление получило в работах Г. Ахенвалля (1719 – 1772). Он привел описание политического состояния и достопримечательностей государств. А. Шлицер (1735 – 1809) опроверг представление Ахенвалля и считал, что предметом статистики является все общество. Школа просуществовала более 150 лет, не меняя своих теоретических основ. Содержание, задачи, предмет изучения статистики в понимании представителей этого направления были далеки от современного взгляда на статистику как на науку.

В первой половине XIXв. возникло третье направление статистической науки – статистико-математическое. Особый вклад в это направление внёс бельгийский статистик А. Кетле (1796 – 1874). По правилам, разработанным А. Кетле, с середины XIX в. в развитых странах проводятся регулярные переписи населения. Он стал основоположником учения о средних величинах. По инициативе учёного для координации развития статистики проводились международные статистические конгрессы, в 1885 г. основан международный статистический институт, существующий до настоящего времени.

В XIX в. развитию статистической методологии способствовали также труды английских учёных Ф. Гальтона (1822 - 1911), К. Пирсона (1857 - 1936), У. Госсета (1876 – 1937), Р. Фишера (1890 – 1962), внёсших значительный вклад в разработку теории корреляции, изучения взаимосвязей явлений. Ф. Гальтон применил статистические методы к проблеме наследственности. К. Пирсон разрабатывал вопросы количественной оценки связи между явлениями. У. Госсет, писавший под псевдонимом Стьюдент, разработал тео рию малой выборки.

С начала XX в. при социально-экономических исследованиях уровня жизни населения, покупательского спроса, качества продукции начали применяться методы теории вероятностей, составляющей одну из отраслей прикладной математики. Наиболее известным учёным в этой области является Р.Фишер.

1.7. Основные этапы развития статистики как науки в России

На Руси первыми статистическими источниками были летописи, в которых упоминается о сборе различной информации в IX-XI вв.: возникновении и развитии городских поселений, расположенных на водных путях, о наличии в них храмов, церквей, монастырей, жилых строений. Статистические сведения приводятся в летописи по Ипатьевскому списку и Лаврентьевской летописи 963 - 981 гг.

Начало государственной статистики в России можно отнести к концу XII – началу XIII в., хотя первые переписи земель и населения с постоянно усложнявшейся программой проводились еще в Киевской Руси (IX – XII вв.). С XIII века к переписям прибегали часто. «Книги сошного письма» и «сказки» - основа для писцовых книг конца XV и начала XVI вв.

Особое внимание в истории развития статистики в России стоит уделить периоду реформ Петра I. Многочисленные изменения во всех сферах государственного устройства определили потребность в учете. Сенат был центром проведения работ по статистике. Сюда собирали отчеты со всех ведомств государства Российского. Уже в восемнадцатом веке подавались сведения о тех, кто родился и умер. И тогда же начали переписывать рабочих различных фабрик.

Реформы Петра I (1672-1725), которыми были охвачены все основные направления общественной жизни: экономика страны, административное управление, армия, культура и быт населения, а также войны вызывали потребность в полном и точном учете материальных ресурсов и населения. В этот период высший правительственный орган - Сенат - через систему коллегий не только руководил экономикой страны, но и являлся центром по проведению важнейших статистических работ, там собирались полученные материалы обследований, отчеты подведомственных коллегиям производств и заведений, а также местной администрации.

Петровская реформа налоговой системы связана с появлением новой единицы, ею стала «душа» мужского пола, что потребовало подушной переписи населения – ревизии. Первая ревизия была объявлена 26 ноября 1718 г., ревизию проводила армия.

В начале XVIII в. в России зарождался и текущий учет населения. Так, в 1702 г. был издан указ о подаче в Патриарший Духовный приказ приходскими священниками недельных ведомостей о родившихся и умерших. В первой половине XVIII в. проводились уже переписи рабочих фабрик и мануфактур.

Новый этап в истории развития статистики в России пришелся на девятнадцатый век. Официальным годом рождения статистики принято считать 1802 год. Именно тогда министерства, по Манифесту Александра I, стали сдавать письменные отчеты. В 1811 году при Министерстве внутренних дел создали Статистическое отделение. Руководителем назначили К.Ф. Германа (1767 – 1838).

В России, как и в зарубежных странах, существовали определенные направления (школы) в изучении статистики. Описательную школу представлял В.Н. Татищев (1686 – 1750). Он предложил определенные правила проведения ревизий, единый учет населения. М.В. Ломоносов (1711 – 1765) усовершенствовал систему, предложенную В.Н. Татищевым. В работах М.В. Ломоносова были уже и аналитические данные. Также яркими представителями русской описательной школы являются И.К. Кириллов (1689-1737), И.И. Голиков (1735-1801), С.Н. Плещев (1752-1802), М.И. Чулков (1740-1793) Собранные ими материалы стали источником сведений по экономической теории России с древних времён до XVIII в.

К.Ф. Герман написал книгу «Всеобщая история статистики. Для обучающихся сей науке». Здесь статистика рассматривалась именно как наука. В истории развития статистики большое значение имеют работы К.И. Арсеньева (1789-1856), в которых он утверждал, что статистика в состоянии дать адекватную характеристику жизни государства. Основы статистики в России создал Д.П. Журавский (1810 - 1856). Он дал системное изложение основ теоретической базы статистики как науки, определение статистической науки, уделил большое внимание вопросу достоверности данных, методу группировок, раскрыл принцип единства количественного и качественного анализа. Экономическую и судебную статистику дополнили своими трудами А.Н. Радищев (1749 – 1802), Н.П. Огарев (1813 – 1877), А.И. Герцен (1812 – 1870).

Особую роль в формировании статистики как науки сыграло земство. Именно во времена земства были созданы статистические бюро. Наиболее известными представителями земской статистки являются В.И. Орлов (1848 -1885) и А.П. Шликевич (род. 1849).

В России развитие математической статистики интенсивно проходило с начала XX в. Появились исследования А.В. Монтовича о кривых распределения; Е.Е. Слуцкого, А.А. Чупрова (1874 - 1926) о корреляционном анализе. Продолжателем А.А. Чупрова стал Н.К. Дружинин - один из ведущих специалистов по математической статистике двадцатого столетия и истории статистической науки. В 1949 г. был издан учебник Н.К. Дружинина, в котором статистика определена как наука о количественных закономерностях массовых явлений, как учение о тех принципах, на которых основывается сбор обработки этих сведений. Н.К. Дружинин последовательно отстаивал в своих трудах мысль, что статистические методы применимы не только в общественных науках, но и науках о природе.

Свою роль в истории статистики сыграли представители академической школы статистики, характерной особенностью которой было стремление заменить изучение государства изучением общества. Основоположниками этой школы явились Э.Ю. Янсон (1835-1893), А.И. Чупров (1842-1908), А.А. Чупров (1874-1926), Н.А. Каблуков (1849-1919) и А.А. Кауфман (1864-1919).

Особенно интенсивно развивалась советская статистика. Анализу подвергались промышленность, народное хозяйство, население, строительство, сельское хозяйство, государственный бюджет и остальные сферы.

Исторический опыт советской статистики как науки был обобщён в трудах В.И. Хотимского (1892-1937), В.С. Немчинова (1894-1964), В.Н. Старовского (1905-1975), А.Я. Боярского (1906-1985), Б.С. Ястремского (1877-1962), Л.В. Некрама (1886-1949) и других учёных.

Начальный этап советской статистики (1917-1930 гг.) отличается исключительной интенсивностью: проводится большое число специально организованных, статистических переписей и обследований, плодотворно работают различные научные коллективы, строится первый баланс народного хозяйства.

В годы Великой Отечественной войны перед советской статистикой стояли задачи по оперативному учету трудовых, материальных ресурсов, перемещение производственных сил страны в восточные районы.

В послевоенный период внимание статистической науки было приковано к вопросу о предмете статистики, её соотношении с математической статистикой. В 1954 г. этот вопрос обсуждался на научном совещании, которое ещё раз подтвердило значение статистики как самостоятельной общественной науки. После совещания вышли в свет новые монографии, учебники по общей теории статистики. В это время значительный вклад в теорию индексного метода был внесён учёными С.М. Югенбергом, Г.И. Баклановым, Л.С. Казинцовым, В.Е. Адамовым и др. Большим шагом вперёд к развитию статистической науки послужило комплексное применение, наряду со статистическими, экономико-математических методов и широкое использование компьютерной техники в анализе социально-экономических явлений.

1.8. Организация статики в Российской Федерации на современном этапе

В соответствии со ст. 71 Конституции РФ руководство статистикой в стране осуществляет Госкомстат как федеральный орган исполнительной власти.

Госкомстат 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.

Федеральная служба государственной статистики – федеральный орган исполнительной власти РФ, который выполняет действия по сбору и формированию официальной статистической информации об экономическом, социальном, экологическом и демографическом положении страны, а также выполняет функции по контролю и надзору в области государственной статистической деятельности на всей территории Российской Федерации.

Служба статистики предназначена для сбора и анализа информации. Она формирует информационную базу, на основе которой принимаются обоснованные управленческие решения. Служба эта сравнима с информационной базой предприятия, действующей в масштабах страны. Информация необходима органам власти для совершенствования налоговой, таможенной, инвестиционной политики.

По данным статистики можно определить, насколько высока налоговая нагрузка на предприятия, в каком состоянии находится та или иная отрасль, как развивается государство, улучшается или ухудшается положение в отдельных сферах жизни.

Деятельность Госкомстата РФ регулирует федеральный закон «Об официальном статистическом учете и системе государственной статистики в Российской Федерации» от 29.11.2007 № 282-ФЗ.

Основные задачи и функции Федеральной службы государственной статистики. Главной задачей можно назвать удовлетворение потребности в информации органов власти и различных организаций – СМИ, коммерческих компаний, научных сообществ.

Основными функциями Федеральной службы государственной статистики являются:

1) обеспечение хранения государственных информационных ресурсов и защиты конфиденциальной и отнесенной к государственной тайне статистической информации;

2) разработка и совершенствование научно обоснованной официальной статистической методологии для проведения статистических наблюдений и формирования статистических показателей, обеспечение соответствия указанной методологии международным стандартам;

3) представление в установленном порядке статистической информации гражданам, Президенту Российской Федерации, Правительству Российской Федерации, Федеральному Собранию Российской Федерации, органам государственной власти, средствам массовой информации, другим организациям, в том числе международным;

4) сбор статистической информации и формирование на её основе официальной статистической отчетности;

5) разработка и совершенствование системы статистических показателей, характеризующих состояние экономики и социальной сферы;

6) развитие информационной системы государственной статистики, обеспечение её совместимости и взаимодействия с другими государственными информационными системами;

7) контроль за выполнением организациями и гражданами, осуществляющими предпринимательскую деятельность без образования

юридического лица, законодательства Российской Федерации в области государственной статистики;

8) реализация обязательств Российской Федерации, вытекающих из членства в международных организациях и участия в международных договорах, осуществление международного сотрудничества в области статистики.

Госкомстат РФ осуществляет два вида наблюдения за деятельностью организаций:

1) сплошное наблюдение для малого и среднего бизнеса проводится раз в пять лет;

2) выборочное наблюдение проводится постоянно, и состав отчетности может меняться из года в год.

Помимо периодических мероприятий, Госкомстат собирает постоянную отчетность. Для этого установлены определенные отчеты. Базой для их подтверждения является бухгалтерская отчетность. Поэтому, вне зависимости от масштабов и вида деятельности, все компании, которые обязаны составлять бухгалтерскую (финансовую) отчетность, должны сдать ее экземпляр в территориальный орган статистики до 31 марта года, следующего за отчетным периодом.

Существующая в России система государственной статистики имеет 3 уровня.

1. Федеральный: центральный аппарат и его подведомственные учреждения.

2. Территориальный: 82 территориальных органов государственной статистики, по числу субъектов РФ. Санкт-Петербургский комитет объединяет два субъекта Федерации: г. Санкт-Петербург и Ленинградскую область.

3. Районный: порядка 2000 районных (городских) отделов статистики.

Основные печатные издания Госкомстата России являются: ежегодники, «Российская Федерация», «Регионы России»; журнал «Вопросы статистики» и другое, а также ознакомиться со статистическими данными можно через сеть intenet.

Наряду с государственными статистическими службами существует ведомственная статистика, которая ведется в министерствах, ведомствах, на предприятиях, в объединениях и фирмах всевозможных отраслей экономики. Ведомственная статистика занимается сбором, обработкой и анализом статистической информации. Эта информация необходима для принятия руководством управленческих решений, для планирования деятельности организации или органа власти. На малых предприятиях такой работой обычно занимается либо главный бухгалтер, либо непосредственно сам руководитель. На крупных предприятиях, в которых разветвлена собственная региональная структура или имеется большая численность работающих, обработкой и анализом статистической информации занимаются целые отделы или управления. К такой работе привлекаются специалисты в сфере статистики, математики, бухгалтерского учета и экономического анализа, менеджеры и технологи. Подобная команда, вооруженная современными средствами вычислительной техники, опираясь на методологию, предлагаемую теорией статистики, и применяя современные методики анализа, помогает строить эффективные стратегии развития бизнеса, а также эффективно формировать деятельность органов государственной власти. Управлять сложными социальными и экономическими системами, не обладая полной, достоверной и оперативной статистической информацией, невозможно.

Схема органов Государственной статистики

§ 2. ВЫБОРОЧНЫЙ МЕТОД

2.1. Выборка. Классификация выборок

В основе практически всех исследований социально – экономических явлений лежит выборочный метод. Он состоит в определении сводных характеристик не всех, а лишь части членов генеральной совокупности, взятых на выборку. К нему прибегают в тех случаях, когда сплошное обследование нельзя осуществить, например, из-за того, что генеральная совокупность имеет бесконечное число элементов (членов).

Например, для определения качества продукции, изготавливаемой предприятием, отбирается сравнительно небольшая ее часть и испытывается. Тогда, показатели качества испытанной партии принимаются за приближенный уровень качества всей производимой предприятием продукцией.

Теоретической основой выборочного метода является Закон больших чисел.

Дадим математическое определение выборочной совокупности. ***Выборкой объема*** *n из генеральной совокупности объема N с функцией плотности распределения вероятностей f(x) называется последовательность*

$$x_1,\ x_2, \ldots,\ x_n$$

значений случайной величины (признака, фактора) X, соответствующих n независимым повторениям эксперимента.

В статистических исследованиях применяются следующие виды выборок.

Собственно-случайной *называется выборка, элементы которой отобраны без какой-либо систематизации, наудачу. Если генеральная совокупность некоторым образом упорядочена* (например, она может быть упорядочена по абонентам телефонной сети), *то выборка, образованная из ее элементов есть* ***механическая***. Здесь отбор осуществляется в соответствие с выбранной пропорцией через равные интервалы. *В случае объединения генеральной совокупности в несколько типических групп* (примерами таких групп может быть деление населения города на мужчин и женщин, или в качестве типических групп можно рассматривать граждан не достигших трудоспособного возраста, граждан трудоспособного возраста и пенсионеров), *применяют* ***типическую выборку***. *Если отбор осуществляется сериями* (они и являются единицами совокупности), *то выборка называ-*

ется ***серийной***. Наконец, *если при отборе единиц генеральной совокупности на выборку применяется несколько вышеуказанных способов* (скажем, серийным и механическим способом), *то выборка носит название* ***комбинаторной***. Каждый из способов отбора характеризуется своими методами и особенностями расчета числовых характеристик. В учебном пособии речь идет в основном о случайных и механических выборках.

Отметим также, что отбор элементов на выборку осуществляется ***повторным и бесповторным способами***. В первом случае, после изучения интересующих нас характеристик, элемент возвращается в генеральную совокупность и может быть вновь взят на выборку. Если отбор бесповторный, то отобранная единица генеральной совокупности в нее не возвращается и не участвует в дальнейшем отборе.

2.2. Сводка и группировка выборочных данных

После того как исследование произведено, показания сняты, информация получена, возникает задача систематизации эмпирических данных

Самым простым способом систематизации данных является их расположение в неубывающем порядке,

$$x_{\min} = x^{(1)} \leq x^{(2)} \leq \ldots \leq x^{(n)} = x_{\max} .$$

Такая форма записи выборочных данных называется ***вариационным рядом*** или ***вариационным распределением***.

Определим ***размах выборки*** как разность между ее максимальным и минимальным элементами

$$R = x_{\max} - x_{\min} . \tag{2.1}$$

Пример 2.1. Имеются сведения о решении контрольной задачи ста студентами (данные приведены в секундах):

99	93	104	100	104	100	108	112	89	97
112	102	104	108	105	104	98	116	120	100
112	102	116	108	96	102	100	91	96	92
96	102	100	99	107	97	96	108	107	101
101	116	99	90	104	94	100	107	96	103
92	104	97	98	110	103	110	105	104	113
108	97	104	98	102	106	107	110	101	110
94	105	88	96	97	94	120	119	104	103
108	96	91	103	102	100	106	90	91	95
106	113	95	105	102	102	104	102	89	103

Необходимо: а) определить объем выборки; б) построить вариационный ряд; в) определить размах выборочной совокупности.

а) Очевидно, что объем выборки равен числу ее элементов. Значит, в рассматриваемом случае

$$n = 100.$$

б) следуя определению, строим вариационный ряд:

88	89	89	90	90	91	91	91	92	92
93	94	94	94	95	95	96	96	96	96
96	96	96	97	97	97	97	97	98	98
98	99	99	99	100	100	100	100	100	100
100	101	101	101	102	102	102	102	102	102
102	102	102	103	103	103	103	103	104	104
104	104	104	104	104	104	104	104	105	105
105	105	106	106	106	107	107	107	107	108
108	108	108	108	108	110	110	110	110	112
112	112	113	113	116	116	116	119	120	120

в) Из (2.1) следует, что размах

$$R = 120 - 88 = 32.$$

Как видим, вариационное распределение позволяет систематизировать выборочные данные. Однако, при большом объеме выборке информация, представленная в виде вариационного распределения, трудно воспринимается, ее практически невозможно анализировать. Кроме того, существенным недостатком вариационного ряда является «повторяемость» элементов.

Проблему повторяющихся элементов можно преодолеть, представляя результаты наблюдений в виде статистического ряда или статистического распределения. Назовем частотой

$$n_x = n_{x_i} = n_i$$

элемента

$$x = x_i$$

называется число его повторений. Очевидно, что объем выборки равен сумме частот элементов,

$$\sum_x n_x = \sum_{i=1}^{n} n_{x_i} = \sum_{i=1}^{n} n_i \ .$$

Статистическим рядом или ***статистическим распределением*** называется такой способ записи выборки, при котором ее элементы представляются в виде пар чисел

$$(x;n_x).$$

Также пару «случайная величина и ее частота» будем обозначать следующим образом: $(x_i;n_{x_i})$ или $(x_i;n_i)$. Обычно, вариационный ряд представляется в виде ***статистической таблицы***:

x	x_1	x_2	x_3	…
n_x	n_1	n_2	n_3	…

Статистическая таблица также может иметь и другую, вертикальную ориентацию. *Строка (столбец) статистической таблицы, в которой записаны значения признака называется ее* ***подлежащим****, а строка (столбец), в которой представлены частоты –* ***сказуемым****.*

Пример 2.2. Представить выборку из примера 2.1 в виде статистического ряда.

Используем построенный в примере 2.1 вариационный ряд, подсчитываем частоту каждого из элементов выборки, получаем:

x	88	89	90	91	92	93	94	95	96	97	98	99	100	101
n_x	1	2	2	3	2	1	3	2	7	5	3	3	7	3
x	102	103	104	105	106	107	108	110	112	113	116	119	120	
n_x	9	5	10	4	3	4	6	4	3	2	3	1	2	

По сравнению с вариационным, статистическое распределение является более компактным способ записи эмпирических данных, при этом без потерь информации. Но проблема «длины» при большом объеме выборке остается непреодолимой.

При большом объеме выборки ее элементы объединяются в группы (интервалы, разряды), представляя результаты наблюдения в виде ***группированного статистического ряда***. Для этого интервал, содержащий все элементы выборочной совокупности разбивается на k частичных непересекающихся интервала как правило равной длины b вида $(\alpha;\beta)$. На практике, число интервалов берется от 6 до 20,

$$6 \le k \le 20.$$

Число интервалов группировки и их длина связаны соотношением:

$$b \approx \frac{R}{k}. \quad (2.2)$$

При этом необходимо учесть, что должно выполняться неравенство:

$$x_{\min} + k \cdot b \geq x_{\max}.$$

Число интервалов также можно подсчитать и по ***формуле Стерджесса***[7]:

$$k \approx 1 + [3{,}222 \lg n]^{8}.$$

Для каждого интервала группировки $(\alpha; \beta)$ определяем:

1. ***Середину интервала***

$$x = \frac{\alpha + \beta}{2};$$

2. ***Частоту интервала*** n_x, равную количеству элементов, попавших в данный интервал; при этом, частота первого интервала равна числу элементов из отрезка $[\alpha; \beta]$ (учитывается и левая и правая граница); для остальных интервалов частота численно равна числу элементов из полуинтервала $(\alpha; \beta]$ (левая граница не учитывается);

3. ***Накопленную частоту интервала*** $s_x = s_i$, равную сумме частот интервалов с первого по данный включительно,

$$s_x = s_i = \sum_{j=1}^{i} n_j;$$

очевидно, что накопленная частота последнего интервала равна объему выборки;

4. ***Относительную частоту***, равную отношению частоты интервала к объему выборки,

$$n'_x = \frac{n_x}{n};$$

относительная частота показывает долю (процентную концентрацию) элементов выборочной совокупности, находящихся в данном интервале;

5. ***Относительную накопленную частоту***

[7] Стерджесс Герберт (1882-1958) – американский статистик.
[8] [*] – целая часть числа *

$$s'_x = s'_i = \sum_{j=1}^{i} n'_j ,$$

равную сумме относительных частот интервалов с первого по данный включительно; из определения следует, что относительная накопленная частота определяет долю (процентную концентрацию) элементов выборочной совокупности, находящуюся в интервалах с первого по данный включительно; при расчете относительной накопленной частоты удобно пользоваться формулой:

$$s'_x = s'_i = \sum_{j=1}^{i} n'_j = \sum_{j=1}^{i} \frac{n_j}{n} = \frac{\sum_{j=1}^{i} n_j}{n} = \frac{s_x}{n};$$

относительная накопленная частота последнего интервала группировки всегда равна 1;

6. Наконец, для каждого интервала находим величину

$$h_x = \frac{n_x}{b}. \qquad (2.3)$$

В некоторых случаях берутся интервалы группировки, в которых левая и правая граница не ограничена числом. Такой интервал называется ***открытым***. Его виртуальную середину следует искать так: если открыта правая граница, то к левой необходимо прибавить половину постоянной длины имеющихся закрытых интервалов, если же открыта левая граница, то из правой необходимо эту половину длины вычесть.

Пример 2.3. По данным примера 2.1 построить группированный статистический ряд.

Вначале определим число интервалов группировки и их длину. Размах выборки равен 32. Конечно, количество интервалов желательно подобрать таким, чтобы длина выражалась целым числом. Число 32 имеет два делителя $6 \le k \le 20$: это 8 и 16. Выбираем $k = 8$. Тогда по формуле (2.2) $b = \frac{32}{8} = 4$. Проверим: $x_{\min} + k \cdot b = 88 + 8 \cdot 4 = 120 = x_{\max}$. Итак, данная выборка разбивается на 8 интервалов длины 4,

$$k = 8,\ b = 4:$$

$$(88;92),\ (92;96),\ (96;100),\ (100;104),\ (104;108),\ (108;112),\ (112;116),\ (116;120).$$

С помощью формулы Стерджесса получаем:

$$k \approx 1 + [3{,}222 \lg 100] = 1 + [6{,}444] = 1 + 6 = 7.$$

Находим середины каждого из интервалов группировок:

$$90, 94, 98, 102, 106, 110, 114, 118.$$

Переходим к расчету частоты. Для этого все элементы выборочной совокупности распределяем по интервалам группировки. Удобно рассуждения вести в таблице (табл. 2.1). Значения частот получены в последнем столбце таблицы.

Остальные расчеты производим согласно определению. Поскольку длина интервалов равна 4, то

$$h_x = \frac{n_x}{4}.$$

Таблица 2.1

Интервал	**Элементы выборочной совокупности**	**Итого**
$(88;92)$	92, 88, 91, 90, 91,90, 89, 91, 89, 92	10
$(92;96)$	96, 94, 93, 96, 95, 96, 96, 94, 94, 96, 96, 96, 95	13
$(96;100)$	99, 97, 100, 99, 97, 100, 99, 98, 98, 97, 100, 97, 100, 98, 100, 100, 97, 100	18
$(100;104)$	101, 102, 102, 102, 104, 104, 104, 104, 103, 104, 104, 102, 102, 102, 104, 102, 103, 102, 104, 102, 104, 101, 104, 101, 103, 103, 103	27
$(104;108)$	108, 108, 106, 105, 108, 108, 105, 105, 107, 106, 108, 107, 106, 108, 107, 105, 107	17
$(108;112)$	112, 112, 110, 110, 112, 110, 110	7
$(112;116)$	116, 113, 116, 113	5
$(116;120)$	120, 119, 120	3

Итак, группированный статистический ряд для данной совокупности имеет вид (табл. 2.2).

Для примера проанализируем четвертую строку группированного ряда. За время от 100до 104 секунд контрольную задачу решило 27 студентов, что составляет 27% от общего объема выборки $\left(n'_x = \frac{27}{100}\right)$. Число студентов, решивших задачу за время от 88 до 104 секунд, составляет 68 человек $(s_x = 68)$ или 68% от общего числа испытуемых $\left(s'_x = \frac{68}{100}\right)$.

С помощью понятия накопленной частоты и относительной накопленной частоты можно определить число элементов и относительные характеристики интервалов, являющихся объединением имеющихся интервалов группировки.

Таблица 2.2

k	$(\alpha;\beta)$	X	n_x	s_x	n'_x	s'_x	h_x
1	$(88;92)$	90	10	10	$\frac{10}{100}$	$\frac{10}{100}$	$\frac{10}{4}$
2	$(92;96)$	94	13	23	$\frac{13}{100}$	$\frac{23}{100}$	$\frac{13}{4}$
3	$(96;100)$	98	18	41	$\frac{18}{100}$	$\frac{41}{100}$	$\frac{18}{4}$
4	$(100;104)$	102	27	68	$\frac{27}{100}$	$\frac{68}{100}$	$\frac{27}{4}$
5	$(104;108)$	106	17	85	$\frac{17}{100}$	$\frac{85}{100}$	$\frac{17}{4}$
6	$(108;112)$	110	7	92	$\frac{7}{100}$	$\frac{92}{100}$	$\frac{7}{4}$
7	$(112;116)$	114	5	97	$\frac{5}{100}$	$\frac{97}{100}$	$\frac{5}{4}$
8	$(116;120)$	118	3	100	$\frac{3}{100}$	1	$\frac{3}{4}$

Пример 2.4. Предположим, что если студент решит задачу за время от 88 до 96 секунд, то он получит оценку «отлично»; если время решения задачи составит от 96 до 108 секунд – «хорошо»; оценку «удовлетворительно» студент получит, если задача будет решена за время от 108 до 116 секунд; в остальных случаях студент получает оценку «неудовлетворительно». Определить распределение студентов по выставленной оценке. Для каждой группы найти число студентов и процентную долю получивших данную оценку от общего числа студентов.

Время от 88 до 96 секунд включает в себя интервалы 1 и 2 (см. табл. 2.2). Значит, общее число элементов, принадлежащих этим интервалам равно накопленной частоте второго интервала,

$$n_{\text{"}отлично\text{"}} = s_{94} = 23.$$

Процент учащихся, получивших эту оценку, определяется по относительной накопленной частоте:

$$n'_{"отлично"} = n'_{94} = \frac{23}{100} \text{ или } 23\%.$$

Чтобы определить числу студентов, решивших задачу на оценку «хорошо», необходимо из накопленной частоты пятого интервала вычесть «отличников»:

$$n_{"хорошо"} = s_{106} - n_{94} = 85 - 23 = 62.$$

Аналогично находим процент учащихся, получивших оценку «хорошо»:

$$n'_{"хорошо"} = n'_{106} - n'_{94} = \frac{85}{100} - \frac{23}{100} = \frac{62}{100} \text{ или } 62\%.$$

Расчет студентов, получивших удовлетворительную оценку, а также их доли, производится аналогично:

$$n_{"удовлетвор\,ительно"} = s_{114} - s_{106} = 97 - 85 = 12,$$

$$n'_{"удовлетвор\,ительно"} = n'_{114} - n'_{106} = \frac{97}{100} - \frac{85}{100} = \frac{12}{100} \text{ или } 12\%.$$

Наконец, осуществляем подсчет студентов в абсолютном и относительном выражении, получивших оценку «неудовлетворительно»:

$$n_{"неудовлетв\,орительно"} = s_{118} - s_{114} = 100 - 97 = 3,$$

$$n'_{"неудовлетв\,орительно"} = n'_{118} - n'_{114} = 1 - \frac{97}{100} = \frac{3}{100} \text{ или } 3\%.$$

Результаты вычислений собираем в таблицу.

Оценка	Время решения задачи, с.	Количество студентов, решивших задачу, чел.	Процент студентов, решивших задачу, %
«отлично»	88 – 96	23	23
«хорошо»	96 – 108	62	62
«удовлетворительно»	108 – 116	12	12
«неудовлетворительно»	118 – 120	3	3

Итак, мы можем найти число элементов выборочной совокупности и их долю, попавших в интервал группировки, состоящих из исходных разбиений. Решить эту задачу для произвольного интервала имеющимися методами не представляется возможным. Мы ее решим в следующем пункте.

Разбиение выборочной совокупности на интервалы равной длины применяется в тех случаях, когда значения признака распределены равномерно. Можно считать, что этому требованию удовлетворяет выборка из примера 2.1. Действительно, согласно статистическому ряду, построенному в примере 2.2, разброс частот значения выборки крайне незначителен. Если же этот принцип нарушается, то имеет смысл производить группировку в интервалы неравной длины. Например, производится группировка по признаку

$$X=\{\text{число книг в семье}\}.$$

Опираясь на данные предыдущих исследований можно предположить, что наибольшее число домашних библиотек не превышает 500 томов. Очень редко встречаются семьи, в домашних библиотеках которых от 500 до 5000 книг. Домашние библиотеки с числом томов свыше 5000 практически не встречаются. Поэтому, целесообразно установить интервалы:

$$(0;50),\ (50;100),\ (100;200),\ (200;300),\ (300;500),\ (500;700),$$

$$(700;1000),\ (1000;2000),\ (2000;5000),\ (5000;10000),\ (10000;+\infty).$$

2.3. Графическое представление выборки

Графическое представление выборки позволяет наглядно представить результаты наблюдений. Кроме того, с помощью графиков можно получить ценную информацию о статистической совокупности. Существует огромное множество всевозможных диаграмм, графиков, распределений, являющихся графическим представлением выборочной совокупности. Мы рассмотрим самые основные.

Полигоном частот *статистического распределения называется ломаная с вершинами в точках с координатами* $(x;n_x)$.

Пример 14.5. Построить полигон частот для статистического ряда из примера 2.2.

Следуя определению и используя статистическую таблицу примера 2.2, строим полигон частот (рис. 2.1).

Аналогично стоится ***полигон частот*** *для случая* ***группированного статистического ряда*** *- ломаная с вершинами в точках с координатами* $(x;n_x)$*, где x – середина, а* n_x *- частота каждого из интервалов группировки.*

Пример 2.6. Построить полигон частот для группированного статистического ряда из примера 2.3.

Искомый полигон частот (рис. 2.2) строим, исходя из определения и табл. 2.2.

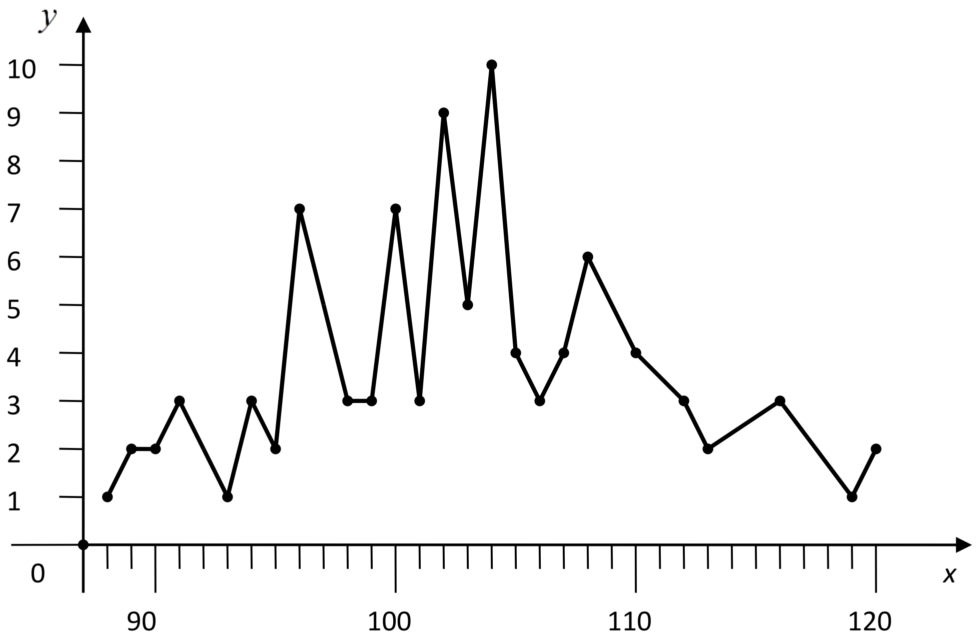

Рис. 2.1

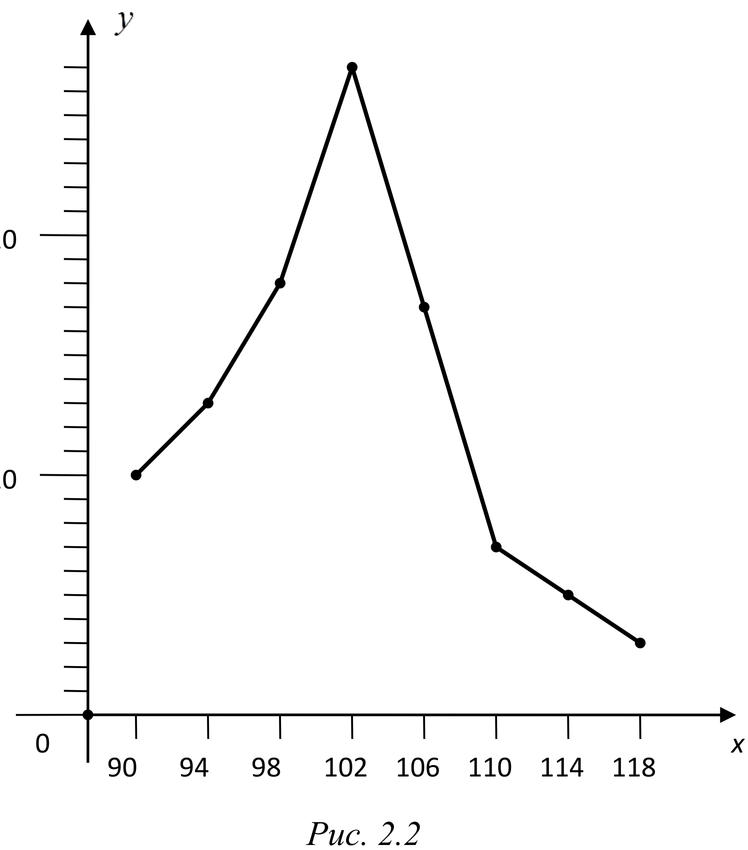

Рис. 2.2

Аналогично строится ***полигон относительных частот*** – *ломаная с вершинами в точках с координатами* $(x; n'_x)$.

Гистограммой частот *группированного распределения называется ступенчатая фигура, состоящая из прямоугольников, построенных на интервалах группировки, площадь каждого из которых равна частоте интервала,*

$$S_x = n_x .$$

Для того, чтобы построить гистограмму, необходимо знать высоту h_x каждого из ее прямоугольников. По определению

$$S_x = b \cdot h_x = n_x ,$$

откуда

$$h_x = \frac{n_x}{b} .$$

Пример 2.7. Поострить гистограмму частот для группированного статистического ряда из примера 2.3.

Высоты каждого из прямоугольников гистограммы найдены в табл. 2.1. Строим ее, исходя из определения (рис. 2.3).

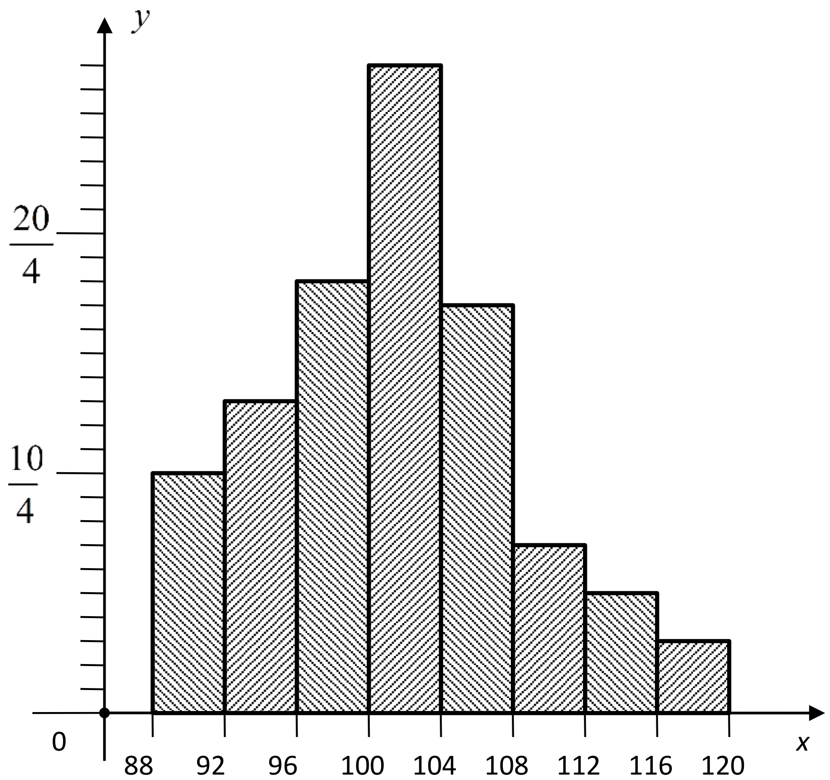

Рис. 2.3

Аналогично строится ***гистограмма относительных частот***. В этом случае площадь каждого ее прямоугольника равна относительной частоте,

$$S_x = n'_x.$$

Полигоном относительных накопленных частот *или* ***кумулятивной кривой,*** *или* ***кумулятой*** *называется ломаная с вершинами в точках с координатами*

$$\left(x + \frac{b}{2}; s'_x\right).$$

Из определения следует, что абсцисса каждой точки кумулятивной кривой – правая граница интервалов группировки.

Пример 2.8. Поострить кумулятивную кривую для группированного статистического ряда из примера 2.3.

При построении используем вторую и седьмую колонки группированного статистического ряда (табл. 2.2). Кумулятивная кривая изображена на рис. 2.4.

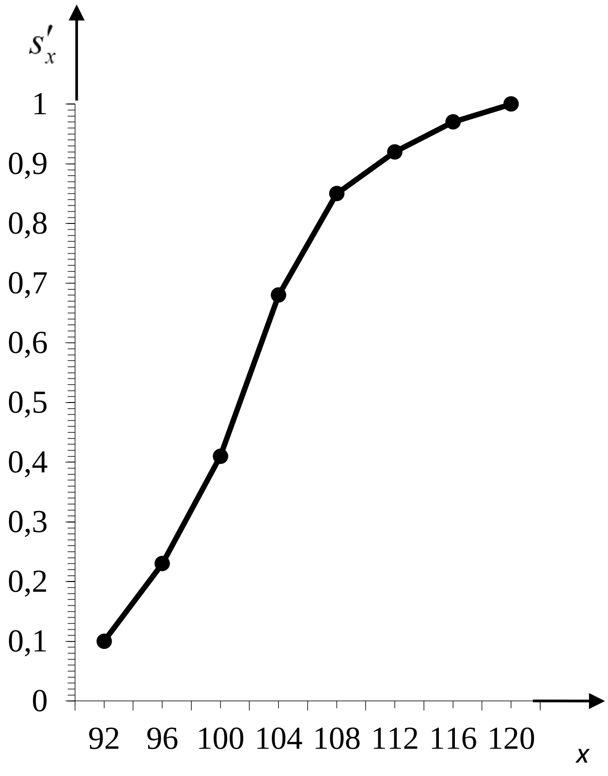

Рис. 2.4

Построенная в примере 2.8 кривая представляет собой лишь часть кумуляты. Полностью построенная кумулятивная кривая должна соединяться с началом координат, причем значения случайной величины должны быть в одном масштабе (рис. 2.5).

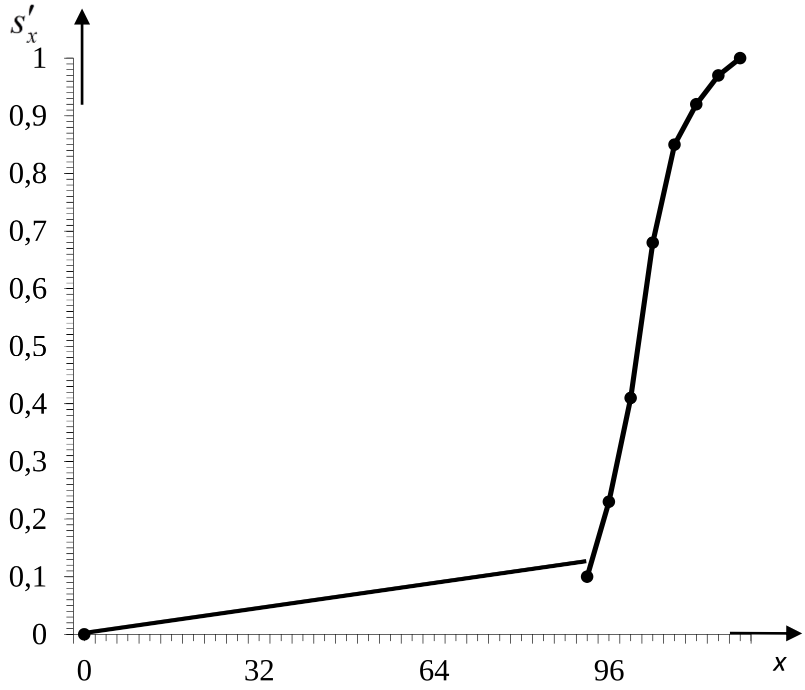

Рис. 2.5

Пусть число

$$p \in [0;1].$$

Выборочной квантилью *порядка* $p = p_x$ *называется абсцисса* x_p *точки, лежащей на кумулятивной кривой, ордината которой равна* p. Исходя из определения следует, что выборочная квантиль определяет интервал

$$(x_{\min}; x_p),$$

в котором содержится доля элементов выборочной совокупности, равная p.

Пример 2.9. Для выборки из примера 2.1 определить квантиль порядка p = 0,54.

Действуем согласно определению. Для этого на оси ординат значение 0,54. Из этой точки восстанавливаем перпендикуляр до пересечения с кумулятивной кривой, а затем из полученной точки опускаем перпендикуляр до пересечения с осью абсцисс (рис. 2.6). Из рисунка видно, что искомая квантиль

$$x_{0,54} \approx 103.$$

Сформулируем ответ в терминах примера 2.1. Имеем: 54% учащихся с минимальным временем написания задачи справились с ней за время от 88 до 103 секунд.

Очевидно, что процесс можно произвести и в обратную сторону, то есть по заданной квантили x_p найти ее порядок *p*. В рассматриваемом примере для $x_p = 103$ порядок

$$p = p_{102} \approx 0{,}54.$$

Если рассуждать в условиях примера 13.1, то тем самым мы получили, что задачу в течение 88 – 103 секунд решило 54% студентов.

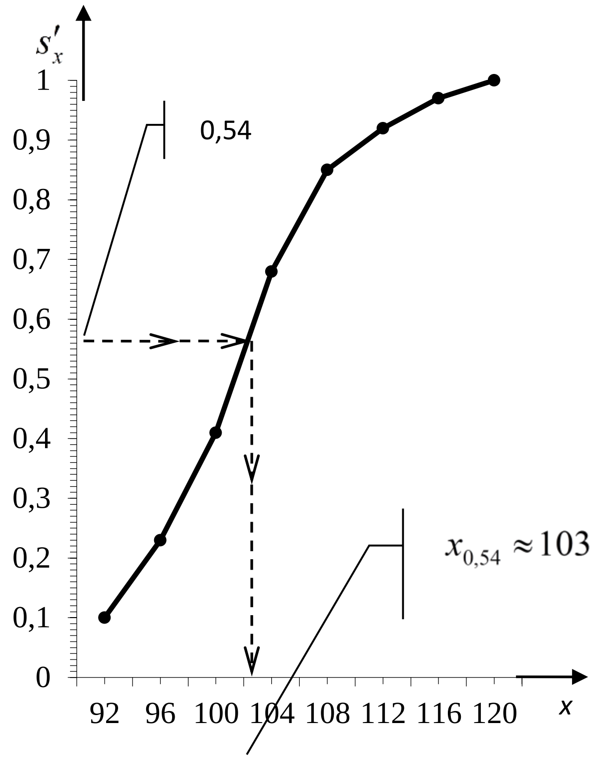

Рис. 2.6

Можно проценты перевести и в абсолютные единицы и рассуждать в них. Поскольку в примере 2.1 объем выборки равен 100, то 54 процентам соответствует 54 студента.

При нахождении порядка квантили необходимо быть аккуратным и внимательным. Например, если объем выборки равен 70, то цена деления оси ординат - $\frac{1}{70}$. Тогда, 27% соответствует не 27 делений (как в случае с объемом *n*=100), а $\frac{70 \cdot 27}{100} = 18{,}9 \approx 19$ делений.

В пункте 2.2 мы определяли долю элементов выборочной совокупности, попавших в интервал, состоящий из нескольких интервалов группировки. С помощью понятия квантиль можно определить эту долю для произвольного интервала $(\alpha;\beta)$. Другими словами, можно найти вероятность $p(\alpha \le X \le \beta)$ попадания значения случайной величины в интервал $(\alpha;\beta)$. Она равна

$$p(\alpha \le X \le \beta) = p_\beta - p_\alpha. \qquad (2.4)$$

Пример 2.10. Пользуясь определением квантили, найти долю элементов выборки из примера 2.1, принадлежащих интервалу $(103;115)$.

Делаем соответствующие построения на кумулятивной кривой (рис. 2.7), затем применяем формулу (2.4). Получаем:

$$p(103 \le X \le 115) = p_{114} - p_{103} \approx 0{,}98 - 0{,}54 = 0{,}44 \text{ или } 44\%.$$

Получаем: 44% студентов справились с задачей за время от 103 до 115 секунд. Проценты можно перевезти и в абсолютные единицы: за указанное время задачу решило приблизительно 44 студента.

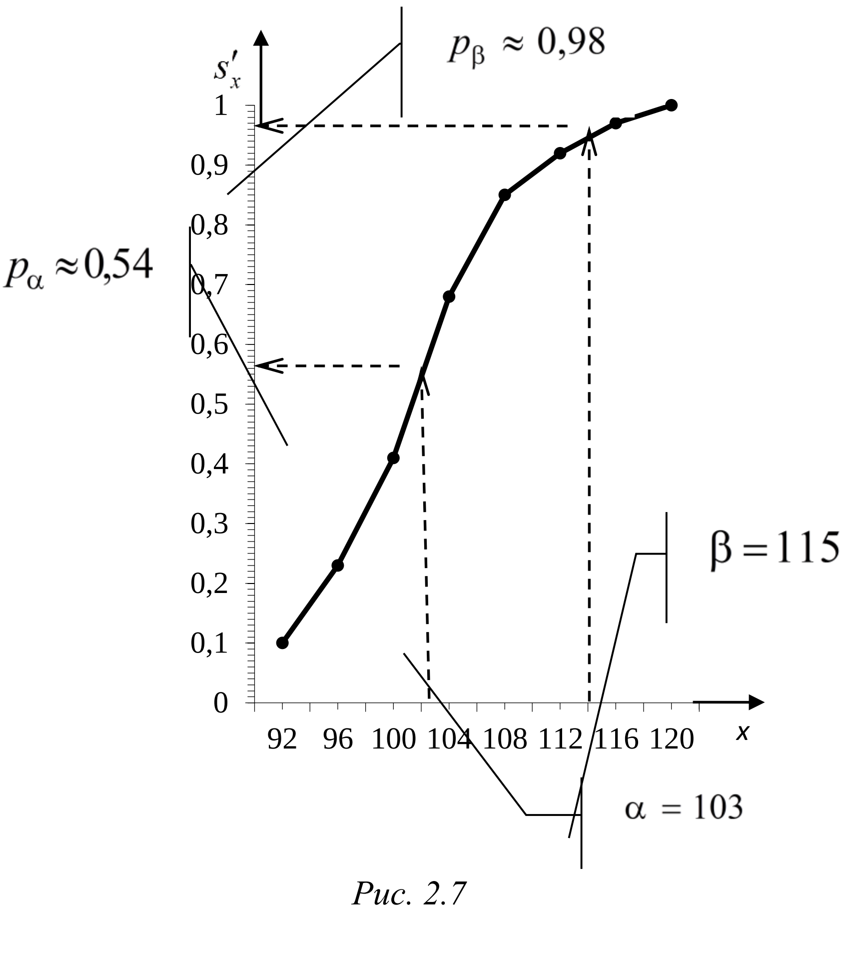

Рис. 2.7

2.4. Квартили, децили перцентили, мода, медиана

С помощью понятия квантиль в вариационных рядах можно отыскать значение признака у любой по порядку следования единицы ранжирования ряда. Существует ряд фиксированных таких единиц. Они делят вариационный ряд на четыре, десять и сто равных частей и носят название ***квартилей***, ***децилей*** и ***перцентилей***.

Различают нижние $Q_p^н$ и верхние $Q_p^в$ квартили, децили и перцентили. Нижние отделяют 25% ($p = \frac{1}{4}$), 10% ($p = \frac{1}{10}$) и 1% ($p = \frac{1}{100}$) единиц выборочной совокупности с минимальным значением признака, а верхние – такое же количество процентов, но с максимальным значением признака. По сути, верхние и нижние квартили, децили и перцентили являются квантилями порядка $p = \frac{1}{4}, \frac{1}{10}, \frac{1}{100}, \frac{3}{4}, \frac{9}{10}, \frac{99}{100}$ соответственно. Их можно найти и с помощью кумулятивной кривой. Но существуют и расчетные формулы для нахождения этих величин.

Расчет $Q_p^н$ начинается с определения интервала, содержащего нижние квартиль, дециль и перцентиль. Считается, что интервал $\left(x_p^н; x_p^н + b_p^н\right)$ с нижней границей $x_p^н$, частотой $n_p^н$ и длиной $b_p^н$ содержит параметр p,если для него первый раз, начиная с интервала номер один, разность между произведением объема выборочной совокупности на p и накопленной частотой s_x, становится отрицательным числом,

$$n \cdot p - s_x < 0.$$

Если s_{x-1} - накопленная частота интервала, предшествующего интервалу, в котором содержится данный параметр, то

$$Q_p^н = x_p^н + b_p^н \cdot \frac{n \cdot p - s_{x-1}}{n_p^н}. \tag{2.5}$$

Аналогично рассчитываются верхние квартили, децили и перцентили. Интервал $\left(x_p^в; x_p^в + b_p^в\right)$ с нижней границей $x_p^в$, частотой $n_p^в$ и длиной $b_p^в$ содержит верхнюю квартиль, дециль или перцентиль, если для него первый раз

$$n \cdot (1 - p) - s_x < 0;$$

$$Q_p^в = x_p^в + b_p^в \cdot \frac{n \cdot (1 - p) - s_{x-1}}{n_p^в}. \tag{2.6}$$

Медианой *me называется значение случайной величины, делящее выборочную совокупность на две равные по количеству элементов части*. Расчет медианы вариационного распределения производится по формуле:

$$me = \begin{cases} x^{\left(\frac{n-1}{2}\right)}, \textit{если} & n \; - \; \textit{нечетное число} \\ \frac{1}{2} \cdot \left(x^{\left(\frac{n}{2}\right)} + x^{\left(\frac{n}{2}+1\right)} \right), \textit{если} & n \; - \; \textit{четное число} \end{cases}.$$

Вычисление медианы для случая группированного статистического ряда производится аналогично (2.5) и (2.6). Только здесь $p = \frac{1}{2}$. Также как и в предыдущих случаях, необходимо искать разность между полусуммой частот $\frac{n}{2}$ и накопленной частотой. И тот интервал с нижней границей x_{me}, длиной b_{me} и частотой n_{me}, для которого первый раз

$$\frac{n}{2} - s_{me} < 0,$$

будет медианным, то есть содержать медиану. Тогда, медиана

$$me = x_{me} + b_{me} \cdot \frac{\frac{n}{2} - s_{me-1}}{n_{me}},$$

где s_{me-1} - накопленная частота интервала, предшествующего медианному.

Мода *определяется как значение случайной величины mo, имеющее максимальную частоту*. Расчет моды группированного распределения начинается с определения модального интервала. Модальным считается интервал, частота которого – максимальна. Величина моды подсчитывается по формуле:

$$mo = x_{mo} + b_{mo} \cdot \frac{n_{mo} - n_{mo-1}}{\left(n_{mo} - n_{mo-1}\right) + \left(n_{mo} - n_{mo+1}\right)},$$

где x_{mo}, b_{mo} и n_{mo} - соответственно нижняя граница, длина и частота модального интервала, n_{mo-1} и n_{mo+1} - частоты интервалов, предшествующего и следующего за модальным.

Пример 2.11. Найти величину моды выборочной совокупности из примера 2.1.

Согласно статистическому распределению, полученному для данной выборки (пример 2.2), максимальная частота $n_{\max} = n_{104} = 10$ соответствует $x = 104$. Следовательно,

$$mo = 104.$$

Пример 2.12. Найти значение медианы для вариационного распределения выборки из примера 2.1.

Так как объем выборки n=100 – четное число, то медиана

$$me = \frac{1}{2}\cdot\left(x^{\left(\frac{100}{2}\right)} + x^{\left(\frac{100}{2}+1\right)}\right) = \frac{1}{2}\cdot\left(x^{(50)} + x^{(51)}\right) = \frac{1}{2}\cdot(102+102) = 102.$$

Пример 2.13. Определить нижние и верхние квартили, децили, перцентили, моду и медиану для группированного ряда из примера 2.3 (табл. 2.2).

Составляем расчетную таблицу (табл. 2.3). В ней выделяем первые отрицательные разности и по ним смотрим интервалы, содержащие искомые параметры. Также выделяем максимальную частоту, которой соответствует модальный интервал. Напомним, что в рассматриваемой выборке объем равен 100, а длина каждого интервала группировки равна 4.

Таблица 2.3

$(\alpha;\beta)$	X	n_x	s_x	$50-s_x$	$p=\frac{1}{4}$		$p=\frac{1}{10}$		$p=\frac{1}{100}$	
					$25-s_x$	$75-s_x$	$10-s_x$	$90-s_x$	$1-s_x$	$99-s_x$
(88;92)	90	10	10	40	15	65	0	80	-9	89
(92;96)	94	13	23	27	2	52	-13	67	-22	74
(96;100)	98	18	41	9	-16	34	-31	49	-31	55
(100;104)	102	27	68	-18	-43	7	-58	22	-67	31
(104;108)	106	17	85	-35	-60	-10	-75	5	-84	14
(108;112)	110	7	92	-42	-67	-17	-82	-2	-91	7
(112;116)	114	5	97	-47	-72	-22	-87	-7	-96	2
(116;120)	118	3	100	-50	-75	-25	-90	-10	-99	-1

Получаем (в терминах задачи):

$$mo = 100 + 4 \cdot \frac{27-18}{(27-18)+(27-17)} \approx 101{,}79 \text{ (с.)};$$

$$me = 100 + 4 \cdot \frac{9}{27} \approx 101{,}33 \text{ (с.)};$$

$$Q^{н}_{\frac{1}{4}} = 96 + 4 \cdot \frac{2}{18} \approx 96{,}44 \text{ (с.)},\ Q^{в}_{\frac{1}{4}} = 104 + 4 \cdot \frac{7}{17} \approx 105{,}65 \text{ (с.)};$$

$$Q^{н}_{\frac{1}{10}} = 92 + 4 \cdot \frac{0}{13} = 92 \text{ (с.)},\ Q^{в}_{\frac{1}{10}} = 108 + 4 \cdot \frac{5}{7} \approx 110{,}86 \text{ (с.)};$$

$$Q^{н}_{\frac{1}{100}} = 88 + 4 \cdot \frac{1-0}{10} = 88{,}40 \text{ (с.)},\ Q^{в}_{\frac{1}{100}} = 116 + 4 \cdot \frac{2}{3} \approx 118{,}67 \text{ (с.)}.$$

Делаем выводы.

1. Максимальное число студентов решило задачу за время 101,79 с.

2. Половина студентов решила задачу за время от 88 до 101, 33 с., а другая половина – за время от 101,33 до 120 с.

3. За время от 88 до 96,44, от 88 до 92, от 88 до 88, 40 с. задачу решило соответственно 25%, 10% и 1% соответственно, а такая же доля студентов с максимальным временем решения задачи принадлежит интервалам: от 105,65 до 120 с., от 110,86 до 120 с., от 118, 67 до 120 с.

Заметим, что рассчитанные в примерах 2.11 и 2.13 значения моды и медианы хоть и ненамного, но все же отличаются от аналогичных характеристик, полученных в примере 2.13.

2.5. Средние величины

Наиболее распространенной формой представления статистических показателей, используемых в социально – экономических исследованиях, является средняя величина. Она дает обобщающую характеристику признака в статистической совокупности в конкретных условиях места и времени. Показатель в форме средней величины выражает типичные черты признака и дает обобщающую характеристику однотипных явлений по одному из варьирующих признаков. Средняя отражает уровень этого признака, отнесенный к единице совокупности. Важнейшее свойство средней заключается в том, что она отражает то общее, что присуще всем единицам исследуемой совокупности.

Отдельные значения признака могут изменяться под влиянием ряда факторов, среди которых могут быть как основные, так и случайные. Сущность средней заключается в том, что в ней взаимопогашаются изменения, вызванные действием случайных факторов и учитываются изменения, вызванные действием факторов основных.

В зависимости от вида распределения расчетные формулы средней бывают простыми и взвешенными. Простая формула применяется для вариационного ряда, а взвешенная – для статистического.

В табл. 2.4 приведены формулы основных видов средней.

Таблица 2.4

	Наименование	Формула	
		Простая	Взвешенная
1	Средняя арифметическая	$\bar{x} = \dfrac{\sum_x x}{n}$	$\bar{x} = \dfrac{\sum_x x \cdot n_x}{\sum_x n_x}$
2	Средняя квадратическая (средний квадрат)	$\overline{x^2} = \sqrt{\dfrac{\sum_x x^2}{n}}$	$\overline{x^2} = \sqrt{\dfrac{\sum_x x^2 \cdot n_x}{\sum_x n_x}}$
3	Средняя геометрическая	$\bar{x} = \sqrt[n]{\prod_x x}$	$\bar{x} = \sqrt[\sum_x n_x]{\prod_x x^{n_x}}$
4	Средняя гармоническая	$\bar{x} = \dfrac{n}{\sum_x \frac{1}{x}}$	$\bar{x} = \dfrac{\sum_x n_x}{\sum_x \frac{1}{x} \cdot n_x}$

Обратим внимание на то, что формула средней арифметической очень напоминает определение математического ожидания дискретной случайной величины. Перечислим свойства средней арифметической.

Свойство 2.1. *Средняя арифметическая постоянной величины равна этой постоянной.*

Свойство 2.2. *Алгебраическая сумма линейных отклонений (разностей) индивидуальных значений признака от средней арифметической, равна нулю.*

Свойство означает, что все отклонения от средней в ту или иную сторону, обусловленные случайными причинами, взаимно погашаются.

Свойство 2.3. *Сумма квадратов отклонений индивидуальных значений признака от средней арифметической, есть число минимальное.*

Свойство 2.4. *Если значения признака каждой единицы совокупности (все усредняемые варианты) увеличить или уменьшить на одно и то же постоянное число с, то и со средней арифметической произойдут те же изменения.*

Свойство 2.5. *Если значения признака единицы совокупности умножить или разделить на постоянное число с, то средняя арифметическая увеличится или уменьшится в с раз.*

Свойство 2.6. *Если частоту каждого значения признака разделить на постоянное число с, то средняя арифметическая не изменится.*

Целесообразность выбора того или иного вида средней зависит от следующих условий: цели усреднении, вида распределения, уровня измерения признака, вычислительных соображений.

2.6. Показатели вариации значений признака

Для характеристики выборочной совокупности оказывается недостаточным знание только средней величины признака. Например, рассмотрим два предприятия «А» и «Б» с одинаковым числом работающих (по 100 человек). Заработная плата на них характеризуется следующими данными (табл. 2.5):

Таблица 2.5

	Категория работающих	Предприятие «А»		Предприятие «Б»	
		Заработная плата, руб.	Количество работающих, чел.	Заработная плата, руб.	Количество работающих, чел.
1	Дирекция	10000	10	8000	5
2	Обслуживающий персонал	5000	20	5000	10
3	Рабочие и служащие	6000	70	2000	85

Среднюю заработную плату можно рассчитать по формуле средней арифметической взвешенной. Имеем по каждому предприятию:

$$\bar{x}_А = \frac{10000 \cdot 10 + 5000 \cdot 20 + 6000 \cdot 70}{100} = 6200 \text{ (руб.)};$$

$$\bar{x}_Б = \frac{8000 \cdot 5 + 5000 \cdot 100 + 2000 \cdot 85}{100} = 6200 \text{ (руб.)}.$$

Из расчета средней заработной платы можно сделать вывод об одинаковых условиях туда на предприятиях. Однако, это не так.

Анализируя табл. 2.5 можно заметить, что разброс значений заработной платы от средней на предприятии «А» гораздо меньше, чем разброс значений от средней на предприятии «Б».

Итак, возникает необходимость ввода числовых характеристик, определяющих меру разброса (вариации, концентрации, колеблемости, рассеяния, изменяемости) значений признака от средней.

Одним из показателей вариации является коэффициент вариации R.

Расчетные формулы характеристик вариации значения признака будем записывать только для взвешенного случая (он более распространен; кроме того, простые расчетные формулы легко получить из взвешенных).

Средним линейным отклонением *называется средний модуль отклонения значения признак от средней,*

$$d = \frac{\sum_x \left| x - \bar{x} \right| \cdot n_x}{\sum_x n_x}.$$

Дисперсия *представляет собой средний квадрат отклонения значения признака от средней,*

$$\sigma_x^2 = \frac{\sum_x \left(x - \bar{x} \right)^2 \cdot n_x}{\sum_x n_x}. \qquad (2.7)$$

Дисперсия может быть рассчитана по упрощенной формуле расчета дисперсии или формуле разностей:

$$\sigma_x^2 = \overline{x^2} - \left(\bar{x} \right)^2, \qquad (2.8)$$

Формула разностей читается так: дисперсия равна разности среднего квадрата и квадрата среднего значения случайной величины X. Докажем (2.8). Используя введенные обозначения и формулу средней арифметической, получаем: $\sigma_x^2 = \frac{\sum_x \left(x - \bar{x} \right)^2 \cdot n_x}{\sum_x n_x} = \frac{\sum_x \left(x^2 - 2 \cdot x \cdot \bar{x} + \left(\bar{x} \right)^2 \right) \cdot n_x}{\sum_x n_x} =$

$$=\frac{\sum_x x^2\cdot n_x+\sum_x\left(-2\cdot x\cdot\overline{x}\right)\cdot n_x+\sum_x\left(\overline{x}\right)^2\cdot n_x}{\sum_x n_x}=\frac{\sum_x x^2\cdot n_x-2\cdot\overline{x}\cdot\sum_x x\cdot n_x+\left(\overline{x}\right)^2\cdot\sum_x n_x}{\sum_x n_x}=$$

$$=\frac{\sum_x x^2\cdot n_x}{\sum_x n_x}-2\cdot\overline{x}\cdot\frac{\sum_x x\cdot n_x}{\sum_x n_x}+\left(\overline{x}\right)^2\cdot\frac{\sum_x n_x}{\sum_x n_x}=\overline{x^2}-2\cdot\overline{x}\cdot\overline{x}+\left(\overline{x}\right)^2=\overline{x^2}-2\cdot\left(\overline{x}\right)^2+\left(\overline{x}\right)^2=$$

$$=\overline{x^2}-\left(\overline{x}\right)^2.$$

Перечислим некоторые свойства дисперсии.

Свойство 2.7. *Дисперсия от постоянной величины равна нулю.*

Свойство 2.8. *Если значения признака уменьшить или увеличить на постоянное число с, то она от этого не изменится.*

Свойство 2.9. *Если все значения признака увеличить или уменьшить в с раз, то дисперсия возрастет или уменьшится в* c^2 *раз.*

Свойство 2.10. *Сумма квадратов отклонений индивидуальных значений признака от их средней* $\overline{x}$ *меньше суммы квадратов отклонений от любого данного числа* $c\neq\overline{x}$.

В ряде случаев возникает необходимость измерить вариацию альтернативного признака. Обозначим w – доля единиц, обладающая альтернативным признаком. Тогда, дисперсия альтернативного показателя равна

$$\sigma_w^2=w\cdot\left(1-w\right). \qquad (2.9)$$

Наряду с формулой (2.7) в статистике рассчитывают несмещенную или исправленную оценку дисперсии:

$$\delta_x^2=\frac{\sum_x\left(x-\overline{x}\right)^2\cdot n_x}{n-1}.$$

Очевидна связь между σ_x^2 и δ_x^2:

$$\sigma_x^2=\delta_x^2\cdot\frac{n-1}{n}\ \text{или}\ \delta_x^2=\sigma_x^2\cdot\frac{n}{n-1}.$$

Среднеквадратическое отклонение *есть арифметический квадратный корень из дисперсии,*

$$\sigma_x=\sqrt{\sigma_x^2}.$$

Отметим, что среднее линейное отклонение и среднеквадратическое отклонение являются именованными величинами, а дисперсия – неименованная величина.

Мерой разброса является также ***коэффициент вариации***, *равный отношению среднеквадратического отклонения к средней, выраженному в процентах*,

$$K_{\sigma_x} = \frac{\sigma_x}{\overline{x}} \cdot 100\% \, .$$

Считается, что если коэффициент вариации меньше 33%, то выборочная совокупность является однородной, корректно составленной, и по ней можно производить анализ, делать выводы и прогнозы, выдвигать гипотезы. В противном случае возникает необходимость нового выборочного обследования, так как имеющаяся выборочная совокупность не является однородной. Коэффициент вариации позволяет сопоставить между собой средние отклонения значений признака в одной и той же совокупности с различными средними (чем больше отклонение, тем большее значение будет иметь коэффициент вариации).

Относительное линейное отклонение *определяется как отношение среднего линейного отклонения к средней, выраженное в процентах*,

$$K_d = \frac{d}{\overline{x}} \cdot 100\% \, .$$

Отношение размаха вариации к средней, выраженное в процентах, есть ***коэффициент осцилляции***,

$$K_R = \frac{R}{\overline{x}} \cdot 100\% \, .$$

К показателям вариации можно отнести асимметрию и эксцесс. ***Асимметрия*** *равна среднему кубу отклонения значений признака от средней, деленному на третью степень среднеквадратического отклонения*,

$$as = \frac{1}{\sigma_x^3} \cdot \frac{\sum\limits_x \left(x - \overline{x}\right)^3 \cdot n_x}{\sum\limits_x n_x} \, . \tag{2.10}$$

Эксцесс определяется как *средняя четвертая степень отклонения значений признака от средней, деленная на четвертую степень среднеквадратического отклонения, минус число 3*:

$$es = \frac{1}{\sigma_x^4} \cdot \frac{\sum_x \left(x - \bar{x}\right)^4 \cdot n_x}{\sum_x n_x} - 3. \tag{2.11}$$

Пример 2.14. Для выборки из примера 2.1 найти среднюю арифметическую и показатели вариации. Расчет дисперсии произвести двумя способами: по определению и пользуясь формулой разностей.

Все расчеты удобно производить в таблице:

x	n_x	$x \cdot n_x$	$x^2 \cdot n_x$	$x - \bar{x}$	$\lvert x - \bar{x}\rvert \cdot n_x$	$\left(x - \bar{x}\right)^2 \cdot n_x$	$\left(x - \bar{x}\right)^3 \cdot n_x$	$\left(x - \bar{x}\right)^4 \cdot n_x$
90	10	900	81000	-11,36	113,6	1290,496	-14660,03456	166537,9926
94	13	1222	114868	-7,36	95,68	704,2048	-5182,947328	38146,49233
98	18	1764	172872	-3,36	60,48	203,2128	-682,795008	2294,191227
102	27	2754	280908	0,64	17,28	11,0592	7,077888	4,52984832
106	17	1802	191012	4,64	78,88	366,0032	1698,254848	7879,902495
110	7	770	84700	8,64	60,48	522,5472	4514,807808	39007,93946
114	5	570	64980	12,64	63,2	798,848	10097,43872	127631,6254
118	3	354	41772	16,64	49,92	830,6688	13822,32883	230003,5518
Σ	100	10136	1032112	-	539,52	4727,04	9614,1312	611506,2252

Теперь находим интересующие нас характеристики.

1. Средняя арифметическая (среднее время решения задачи студентами из имеющейся выборки, выборочное среднее)

$$\bar{x} = \frac{\sum_x x \cdot n_x}{\sum_x n_x} = \frac{10136}{100} = 101{,}36 \text{ (с.)}. \tag{2.12}$$

2. Среднее линейное отклонение:

$$d = \frac{\sum_x \lvert x - \bar{x}\rvert \cdot n_x}{\sum_x n_x} = \frac{539{,}52}{100} = 5{,}3952 \approx 5{,}40 \text{ (с.)}.$$

3. Дисперсия, рассчитанная по определению:

$$\sigma_x^2 = \frac{\sum_x \left(x - \bar{x}\right)^2 \cdot n_x}{\sum_x n_x} = \frac{4727{,}04}{100} = 47{,}2704. \tag{2.13}$$

4. Дисперсия, рассчитанная по формуле разностей:

$$\sigma_x^2 = \overline{x^2} - \left(\overline{x}\right)^2 = \frac{1032112}{100} - (101{,}36)^2 = 47{,}2704.$$

Результат полностью совпадает с (2.13).

5. Несмещенная оценка дисперсии:

$$\delta_x^2 = \frac{\sum_x \left(x - \overline{x}\right)^2 \cdot n_x}{n-1} = \frac{4727{,}04}{100-1} \approx 47{,}7479.$$

6. Среднеквадратическое отклонение:

$$\sigma_x = \sqrt{\sigma_x^2} = \sqrt{47{,}2704} \approx 6{,}8753 \approx 6{,}88 \text{ (с.)}. \qquad (2.14)$$

7. Коэффициент вариации:

$$K_{\sigma_x} = \frac{\sigma_x}{\overline{x}} \cdot 100\% = \frac{6{,}8753}{101{,}36} \cdot 100\% \approx 6{,}7831\%.$$

8. Относительное линейное отклонение:

$$K_d = \frac{d}{\overline{x}} \cdot 100\% = \frac{5{,}3952}{101{,}36} \cdot 100\% \approx 5{,}3228\%.$$

9. Коэффициент осцилляции:

$$K_R = \frac{R}{\overline{x}} \cdot 100\% = \frac{32}{101{,}36} \cdot 100\% \approx 31{,}5706\%.$$

10. Асимметрия:

$$as = \frac{1}{\sigma_x^3} \cdot \frac{\sum_x \left(x - \overline{x}\right)^3 \cdot n_x}{\sum_x n_x} = \frac{1}{(6{,}8753)^3} \cdot \frac{9614{,}1312}{100} \approx 0{,}2958. \qquad (2.15)$$

11. Эксцесс:

$$es = \frac{1}{\sigma_x^4} \cdot \frac{\sum_x \left(x - \overline{x}\right)^4 \cdot n_x}{\sum_x n_x} - 3 = \frac{1}{(6{,}8753)^4} \cdot \frac{611506{,}2252}{100} - 3 \approx -0{,}2633. \qquad (2.16)$$

2.7. Метод моментов

Вычисление числовых характеристик выборки показателей ее разброса при больших значениях признака x и их частот n_x может быть крайне затруднительным. Метод моментов упрощает и облегчает подсчет выборочных характеристик в указанных случаях.

В начале находим ***шаг варьирования*** h_x, *то есть разность между любыми двумя соседними значениями признака*. Будем считать, что для выборки h_x - постоянное число.

Далее, определяем *величину* ***ложного нуля*** *– значение случайной величины* C_x*, имеющее максимальную частоту*.

Каждому значению признака x ставим в соответствие ***условный вариант*** α_x, равный

$$\alpha_x = \frac{x - C_x}{h_x}.$$

Условные моменты m_1, m_2, m_3 и m_3 определяются по формулам:

$$m_1 = \frac{\sum_x \alpha_x \cdot n_x}{\sum_x n_x},\; m_2 = \frac{\sum_x \alpha_x^2 \cdot n_x}{\sum_x n_x},\; m_3 = \frac{\sum_x \alpha_x^3 \cdot n_x}{\sum_x n_x},\; m_4 = \frac{\sum_x \alpha_x^4 \cdot n_x}{\sum_x n_x}. \qquad (2.17)$$

Выразим x через условный вариант, шаг варьирования и ложный нуль:

$$x = \alpha_x \cdot h_x + C_x. \qquad (2.18)$$

Подставим выражение (2.18) в формулу средней арифметической:

$$\bar{x} = \frac{\sum_x x \cdot n_x}{\sum_x n_x} = \frac{\sum_x (\alpha_x \cdot h_x + C_x) \cdot n_x}{\sum_x n_x} = \frac{\sum_x (\alpha_x \cdot h_x \cdot n_x + C_x \cdot n_x)}{\sum_x n_x} = \frac{\sum_x \alpha_x \cdot h_x \cdot n_x + \sum_x C_x \cdot n_x}{\sum_x n_x} =$$

$$= \frac{h_x \cdot \sum_x \alpha_x \cdot n_x + C_x \cdot \sum_x n_x}{\sum_x n_x} = h_x \cdot \frac{\sum_x \alpha_x \cdot n_x}{\sum_x n_x} + C_x \cdot \frac{\sum_x n_x}{\sum_x n_x} = h_x \cdot m_1 + C_x \cdot 1 = m_1 \cdot h_x + C_x.$$ Таким образом, метод моментов дает следующую формулу расчета средней арифметической:

$$\bar{x} = m_1 \cdot h_x + C_x.$$

Применим метод моментов к расчету дисперсии. Для этого выражение (2.18) подставим в (2.7) и используем (2.17):

$$\sigma_x^2 = \frac{\sum_x (x - \bar{x})^2 \cdot n_x}{\sum_x n_x} =$$

$$= \frac{\sum_x ((\alpha_x \cdot h_x + C_x) - (m_1 \cdot h_x + C_x))^2 \cdot n_x}{\sum_x n_x} = \frac{\sum_x (\alpha_x \cdot h_x - m_1 \cdot h_x)^2 \cdot n_x}{\sum_x n_x} = \frac{\sum_x (\alpha_x - m_1)^2 \cdot h_x^2 \cdot n_x}{\sum_x n_x} =$$

$$= h_x^2 \cdot \frac{\sum_x (\alpha_x - m_1)^2 \cdot n_x}{\sum_x n_x} = h_x^2 \cdot \frac{\sum_x (\alpha_x^2 - 2 \cdot \alpha_x \cdot m_1 + m_1^2) \cdot n_x}{\sum_x n_x} =$$

$$= h_x^2 \cdot \frac{\sum\limits_x \left(\alpha_x^2 \cdot n_x - 2 \cdot \alpha_x \cdot m_1 \cdot n_x + m_1^2 \cdot n_x\right)}{\sum\limits_x n_x} = h_x^2 \cdot \frac{\sum\limits_x \alpha_x^2 \cdot n_x - 2 \cdot m_1 \cdot \sum\limits_x \alpha_x \cdot n_x + m_1^2 \cdot \sum\limits_x n_x}{\sum\limits_x n_x} =$$

$$= h_x^2 \cdot \left(\frac{\sum\limits_x \alpha_x^2 \cdot n_x}{\sum\limits_x n_x} - 2 \cdot m_1 \cdot \frac{\sum\limits_x \alpha_x \cdot n_x}{\sum\limits_x n_x} + m_1^2 \cdot \frac{\sum\limits_x n_x}{\sum\limits_x n_x} \right) = h_x^2 \cdot \left(m_2 - 2 \cdot m_1 \cdot m_1 + m_1^2 \cdot 1\right) =$$

$= h_x^2 \cdot \left(m_2 - 2 \cdot m_1^2 + m_1^2\right) = h_x^2 \cdot \left(m_2 - m_1^2\right) = \left(m_2 - m_1^2\right) \cdot h_x^2$. Итак, дисперсия

$$\sigma_x^2 = \left(m_2 - m_1^2\right) \cdot h_x^2 .$$

В формуле асимметрии (2.10) выразим средний куб отклонения значений признака от средней методом моментов. Также как и в случае с дисперсией $\frac{\sum\limits_x \left(x - \bar{x}\right)^3 \cdot n_x}{\sum\limits_x n_x} = \frac{\sum\limits_x \left(\left(\alpha_x \cdot h_x + C_x\right) - \left(m_1 \cdot h_x + C_x\right)\right)^3 \cdot n_x}{\sum\limits_x n_x} =$

$$= \frac{\sum\limits_x \left(\alpha_x \cdot h_x - m_1 \cdot h_x\right)^3 \cdot n_x}{\sum\limits_x n_x} = \frac{\sum\limits_x \left(\alpha_x - m_1\right)^3 \cdot h_x^3 \cdot n_x}{\sum\limits_x n_x} = h_x^3 \cdot \frac{\sum\limits_x \left(\alpha_x - m_1\right)^3 \cdot n_x}{\sum\limits_x n_x} =$$

$$= h_x^3 \cdot \frac{\sum\limits_x \left(\alpha_x^3 - 3 \cdot \alpha_x^2 \cdot m_1 + 3 \cdot \alpha_x \cdot m_1^2 - m_1^3\right) \cdot n_x}{\sum\limits_x n_x} =$$

$$= h_x^3 \cdot \frac{\sum\limits_x \left(\alpha_x^3 \cdot n_x - 3 \cdot \alpha_x^2 \cdot m_1 \cdot n_x + 3 \cdot \alpha_x \cdot m_1^2 \cdot n_x - m_1^3 \cdot n_x\right)}{\sum\limits_x n_x} =$$

$$= h_x^3 \cdot \frac{\sum\limits_x \alpha_x^3 \cdot n_x - 3 \cdot m_1 \sum\limits_x \alpha_x^2 \cdot n_x + 3 \cdot m_1^2 \cdot \sum\limits_x \alpha_x \cdot n_x - m_1^3 \cdot \sum\limits_x n_x}{\sum\limits_x n_x} =$$

$$= h_x^3 \cdot \left(\frac{\sum\limits_x \alpha_x^3 \cdot n_x}{\sum\limits_x n_x} - 3 \cdot m_1 \cdot \frac{\sum\limits_x \alpha_x^2 \cdot n_x}{\sum\limits_x n_x} + 3 \cdot m_1^2 \cdot \frac{\sum\limits_x \alpha_x \cdot n_x}{\sum\limits_x n_x} - m_1^3 \cdot \frac{\sum\limits_x n_x}{\sum\limits_x n_x} \right) =$$

$$= h_x^3 \cdot \left(m_3 - 3 \cdot m_1 \cdot m_2 + 3 \cdot m_1^2 \cdot m_1 - m_1^3 \cdot 1\right) = h_x^3 \cdot \left(m_3 - 3 \cdot m_1 \cdot m_2 + 3 \cdot m_1^3 - m_1^3\right) =$$

$= \left(m_3 - 3 \cdot m_2 \cdot m_1 + 2 \cdot m_1^3\right) \cdot h_x^3$. Итак, асимметрию с помощью метода моментов подсчитывают так:

$$as = \frac{1}{\sigma_x^3} \cdot \left(m_3 - 3 \cdot m_2 \cdot m_1 + 2 \cdot m_1^3\right) \cdot h_x^3 .$$

Аналогично получаем формулу эксцесса. Средняя четвертая степень отклонения значения признака от средней равна $\frac{\sum_x \left(x-\bar{x}\right)^4 \cdot n_x}{\sum_x n_x} =$

$$= \frac{\sum_x \left(\left(\alpha_x \cdot h_x + C_x\right) - \left(m_1 \cdot h_x + C_x\right)\right)^4 \cdot n_x}{\sum_x n_x} = \frac{\sum_x \left(\alpha_x \cdot h_x - m_1 \cdot h_x\right)^4 \cdot n_x}{\sum_x n_x} =$$

$$= \frac{\sum_x \left(\alpha_x - m_1\right)^4 \cdot h_x^4 \cdot n_x}{\sum_x n_x} = h_x^4 \cdot \frac{\sum_x \left(\alpha_x - m_1\right)^4 \cdot n_x}{\sum_x n_x} =$$

$$= h_x^4 \cdot \frac{\sum_x \left(\alpha_x^4 - 4 \cdot \alpha_x^3 \cdot m_1 + 6 \cdot \alpha_x^2 \cdot m_1^2 - 4 \cdot \alpha_x \cdot m_1^3 + m_1^4\right) \cdot n_x}{\sum_x n_x} =$$

$$= h_x^4 \cdot \frac{\sum_x \left(\alpha_x^4 \cdot n_x - 4 \cdot \alpha_x^3 \cdot m_1 \cdot n_x + 6 \cdot \alpha_x^2 \cdot m_1^2 \cdot n_x - 4 \cdot \alpha_x \cdot m_1^3 \cdot n_x + m_1^4 \cdot n_x\right)}{\sum_x n_x} =$$

$$= h_x^4 \cdot \frac{\sum_x \alpha_x^4 \cdot n_x - 4 \cdot m_1 \cdot \sum_x \alpha_x^3 \cdot n_x + 6 \cdot m_1^2 \cdot \sum_x \alpha_x^2 \cdot n_x - 4 \cdot m_1^3 \cdot \sum_x \alpha_x \cdot n_x + m_1^4 \cdot \sum_x n_x}{\sum_x n_x} =$$

$$= h_x^4 \cdot \left(\frac{\sum_x \alpha_x^4 \cdot n_x}{\sum_x n_x} - 4 \cdot m_1 \cdot \frac{\sum_x \alpha_x^3 \cdot n_x}{\sum_x n_x} + 6 \cdot m_1^2 \cdot \frac{\sum_x \alpha_x^2 \cdot n_x}{\sum_x n_x} - 4 \cdot m_1^3 \cdot \frac{\sum_x \alpha_x \cdot n_x}{\sum_x n_x} + m_1^4 \cdot \frac{\sum_x n_x}{\sum_x n_x} \right) =$$

$$= h_x^4 \cdot \left(m_4 - 4 \cdot m_1 \cdot m_3 + 6 \cdot m_1^2 \cdot m_2 - 4 \cdot m_1^3 \cdot m_1 + m_1^4 \cdot 1\right) =$$

$$= h_x^4 \cdot \left(m_4 - 4 \cdot m_1 \cdot m_3 + 6 \cdot m_1^2 \cdot m_2 - 4 \cdot m_1^4 + m_1^4\right) =$$

$$= \left(m_4 - 4 \cdot m_3 \cdot m_1 + 6 \cdot m_2 \cdot m_1^2 - 3 \cdot m_1^4\right) \cdot h_x^4.$$

Итак, эксцесс рассчитывается по формуле:

$$es = \frac{1}{\sigma_x^4} \cdot \left(m_4 - 4 \cdot m_3 \cdot m_1 + 6 \cdot m_2 \cdot m_1^2 - 3 \cdot m_1^4\right) \cdot h_x^4 - 3.$$

Пример 2.15. Для выборки из примера 2.1 найти среднюю арифметическую, дисперсию, асимметрию и эксцесс методом моментов.

Анализируем таблицу 2.2 (пример 2.3). Видим, что все соседние значения случайной величины отличаются друг от друга на 4. Следовательно, шаг варьирования

$$h_x = 4.$$

Далее, максимальной частоте $n_x = 27$ соответствует значение $x = 102$, следовательно

$$C_x = 102.$$

Значит, условный вариант

$$\alpha_x = \frac{x-102}{4}.$$

Находим условные варианты и производим необходимые вычисления для расчета условных моментов

x	n_x	α_x	$\alpha_x \cdot n_x$	$\alpha_x^2 \cdot n_x$	$\alpha_x^3 \cdot n_x$	$\alpha_x^4 \cdot n_x$
90	10	-3	-30	90	-270	810
94	13	-2	-26	52	-104	208
98	18	-1	-18	18	-18	18
102	27	0	0	0	0	0
106	17	1	17	17	17	17
110	7	2	14	28	56	112
114	5	3	15	45	135	405
118	3	4	12	48	192	768
Σ	100	-	-16	298	8	2338

Находим условные моменты по формулам (2.17):

$$m_1 = \frac{-16}{100} = -0{,}16,\ m_2 = \frac{298}{100} = 2{,}98,\ m_3 = \frac{8}{100} = 0{,}08,\ m_4 = \frac{2338}{100} = 23{,}38.$$

Получаем:

$$\bar{x} = -0{,}16 \cdot 4 + 102 = 101{,}36;$$

$$\sigma_x^2 = \left(2{,}98 - (-0{,}16)^2\right) \cdot 4^2 = 47{,}2704;$$

$$as = \frac{1}{(6{,}8753)^3} \cdot \left(0{,}08 - 3 \cdot 2{,}98 \cdot (-0{,}16) + 2 \cdot (-0{,}16)^3\right) \cdot 4^3 \approx 0{,}2958;$$

$$es = \frac{1}{(6{,}8753)^4} \cdot \left(23{,}38 - 4 \cdot 0{,}08 \cdot (-0{,}16) + 6 \cdot 2{,}98 \cdot (-0{,}16)^2 - 3 \cdot (-0{,}16)^4\right) \cdot 4^4 - 3 \approx -0{,}2633.$$

Заметим, что полученные результаты полностью совпадают с результатами примера 2.14.

2.8. Доверительная вероятность. Доверительный интервал

В предыдущих пунктах мы рассмотрели характеристики выборочной совокупности. Задачей любого статистического исследования является прежде всего получение и изучение сводных характеристик генеральной совокупности. В данном пункте будет показано: как с помощью выборочного метода можно по оценкам выборочной совокупности делать выводы параметрах генеральной совокупности.

Все рассмотренные выборочные характеристики выражаются одним числом. Они называются ***точечными оценками***. Их величина зависит от множества факторов (например, от объема выборки) и могут сильно отличаться от истинной величины оцениваемого параметра, что может привести к грубым ошибкам. Это вызывает необходимость оценивать точность и надежность полученных по выборке результатов, что производится при помощи ***интервальных оценок***.

Рассмотрим вывод интервальной оценки для математического ожидания (или генеральной средней или среднего значения генеральной совокупности) a.

Очевидно, что выборочная средняя $\bar{x}$ тем точнее описывает параметр a, чем меньше величина отклонения $\left|\bar{x}-a\right|$. Другими словами, если выбрать положительное число $\delta > 0$ и записать неравенство

$$\left|\bar{x}-a\right| < \delta,\ \delta > 0,$$

то оценка будет тем точнее, чем меньше величина δ. Раскроем последнее неравенство: $-\delta < \bar{x}-a < \delta$. Теперь от каждой части неравенства отнимем выборочную среднюю: $-\delta-\bar{x} < \bar{x}-a-\bar{x} < \delta-\bar{x}$, $-\delta-\bar{x} < -a < \delta-\bar{x}$. Умножим все три части неравенства на -1 (знак неравенства сменится на противоположный), получаем:

$$\bar{x}-\delta < a < \bar{x}+\delta.$$

Поучаем, что неизвестная величина математического ожидания принадлежит интервалу:

$$a \in \left(\bar{x}-\delta; \bar{x}+\delta\right). \tag{2.19}$$

Интервал (2.19) называется ***доверительным интервалом***, а величина δ - ***предельная ошибка выборки***.

Говорить о принадлежности математического ожидания доверительному интервалу можно лишь с определенной вероятностью γ или надежностью. Она называется ***доверительной вероятностью***. Иногда вместо γ используют ***уровень значимости*** $\alpha = 1-\gamma$.

Предельная ошибка равна

$$\delta = t \cdot \mu.$$

Постоянная величина t зависит от доверительной вероятности γ и определяется как аргумент функции Лапласа, если ее значение равно γ,

$$\Phi(t) = \gamma .$$

Значения функции Лапласа приведены в табл. 1 прил. В частности

t=3, если γ=0,9973,

t=2, если γ=0,9545,

t=1, если γ=0,6827.

Средняя ошибка выборки μ в зависимости от типа отбора подсчитывается следующим образом:

$\mu = \sqrt{\dfrac{\sigma_x^2}{n}}$ - если отбор случайный повторный;

$\mu = \sqrt{\dfrac{\sigma_x^2}{n} \cdot \left(1 - \dfrac{n}{N}\right)}$ - если отбор случайный бесповторный.

Введенные формулы средней ошибки наглядно показывают, что при бесповторном отборе длина доверительного интервала уменьшается. Это происходит потому, что в этом случае величина средней ошибки выборки умножается на $\sqrt{\cdot\left(1 - \dfrac{n}{N}\right)} < 1$. Следовательно, применение бесповторного отбора сужает границы доверительного интервала, оценка математического ожидания становится более точной. Однако, при бесповторном отборе требуется знание объема генеральной совокупности, что порой установить невозможно. При применении повторного отбора значения N не требуется.

Часто возникает необходимость определить доверительный интервал, в котором находится ***генеральная доля признака***. Имеем, генеральная доля признака принадлежит интервалу, $p \in (w - \delta_w; w + \delta_w)$.

При этом

$$w = \frac{m}{n}$$

– выборочная доля признака, m – количество элементов выборочной совокупности, обладающих интересующим нас признаком. Предельная ошибка доли равна

$$\delta_w = t \cdot \mu_w .$$

Особенности расчета средней ошибки обусловлены видом дисперсии. в данном случае она равна (2.9). Имеем:

$$\mu_w = \sqrt{\frac{w \cdot (1-w)}{n}}$$ - если отбор случайный повторный;

$$\mu_w = \sqrt{\frac{w \cdot (1-w)}{n} \cdot \left(1 - \frac{n}{N}\right)}$$ - если отбор случайный бесповторный.

Прежде чем приступить к выборочному обследованию, необходимо определить объем выборочной совокупности n. Во – первых он должен быть достаточным, чтобы по найденным выборочным характеристикам можно было бы получить достоверные (несмещенные) оценки генеральной совокупности. Во-вторых, необходимо, чтобы этот объем был бы минимальным; это необходимо для упрощения анализа выборочной информации и расчетов. Другими словами, объем должен быть минимальным, но достаточным для получения несмещенных оценок генеральной совокупности. Можно сказать, и так: выборочная совокупность должна быть самой минимальной моделью, с помощью которой можно было бы иметь адекватное представление о генеральной совокупности. Эта задача называется проблемой ***репрезентативности выборочных данных***. Обычно заранее задаются: 1) значением дисперсии или среднеквадратического отклонения (например, значение дисперсии можно взять из предыдущих исследований); 2) величиной предельной ошибкой выборки; 3) доверительной вероятностью, с которой необходимо гарантировать полученные результаты. Задача формулируется так: объем выборки должен быть таким, чтобы

$$t \cdot \mu \le \delta .$$

Для повторного отбора имеем: $t \cdot \sqrt{\frac{\sigma_x^2}{n}} \le \delta$. Разделим обе части неравенства на t (поскольку t – положительное число, то знак неравенства при этом не изменится): $\sqrt{\frac{\sigma_x^2}{n}} \le \frac{\delta}{t}$. Возводим обе части полученного неравенства в квадрат: $\frac{\sigma_x^2}{n} \le \frac{\delta^2}{t^2}$. Выражаем объем выборки:

$$n \ge \frac{t^2 \cdot \sigma_x^2}{\delta^2}. \tag{2.20}$$

Итак, если отбор случайный повторный, то неравенство (2.20) определяет минимальный объем выборочной совокупности.

Решение проблемы репрезентативности выборочных данных для случая бесповторного отбора аналогично. Получаем: $t \cdot \sqrt{\frac{\sigma_x^2}{n} \cdot \left(1 - \frac{n}{N}\right)} \le \delta$;

$\sqrt{\frac{\sigma_x^2}{n} \cdot \left(1 - \frac{n}{N}\right)} \le \frac{\delta}{t}$; $\frac{\sigma_x^2}{n} \cdot \left(1 - \frac{n}{N}\right) \le \frac{\delta^2}{t^2}$; $\sigma_x^2 \cdot \left(1 - \frac{n}{N}\right) \le n \cdot \frac{\delta^2}{t^2}$; $\sigma_x^2 - \sigma_x^2 \cdot \frac{n}{N} \le n \cdot \frac{\delta^2}{t^2}$;

$\sigma_x^2 \le \sigma_x^2 \cdot \frac{n}{N} + n \cdot \frac{\delta^2}{t^2}$; $\sigma_x^2 \le n \cdot \left(\frac{\sigma_x^2}{N} + \frac{\delta^2}{t^2}\right)$;

$$n \ge \frac{\sigma_x^2}{\frac{\sigma_x^2}{N} + \frac{\delta^2}{t^2}}. \qquad (2.21)$$

Значит, минимальный объем выборочной совокупности определяется с помощью неравенства (2.20).

Пример 2.16. Для определения средней массы изделия в партии готовой продукции было отобрано 100 изделий методом случайного повторного отбора. Оказалось, что средняя масса изделия равна 15 г. при среднеквадратическом отклонении 2 г. С вероятностью 0,9973 найти границы интервала, в котором будет находиться средняя масса изделий всей партии.

Имеем: $n = 100$, $\bar{x} = 15$, $\sigma = 2$; для вероятности γ=0,9973 постоянная величина t=3.

Для повторного отбора средняя ошибка выборки равна

$$\mu = \sqrt{\frac{2^2}{100}} = 0{,}2.$$

Значит, предельная ошибка

$$\delta = 3 \cdot 0{,}2 = 0{,}6.$$

Определяем нижнюю и верхнюю границы доверительного интервала:

$$\bar{x} - \delta = 15 - 0{,}6 = 14{,}4 \text{ (г.)};$$

$$\bar{x} + \delta = 15 + 0{,}6 = 15{,}6 \text{ (г.)}.$$

Итак, с вероятностью 0,9973 можно утверждать, средний вес изделия все партии находится в пределах от 14,4 до 15,6 г. ($15 \pm 0{,}6$ г.).

Пример 2.17. Для определения среднего возраста жителей города с населением 250 тысяч человек было осуществлено бесповторное обследование 1000 человек. Его результаты представлены в таблице:

Возраст, лет.	0 - 10	10 - 20	20 - 30	30 - 40	40 - 50	50 - 60	60 - 70	70 - 80	80 и выше
Количество жителей, человек.	20	80	100	200	250	120	100	80	50

Необходимо с вероятностью 0,9545 найти границы интервала, в котором будет находиться средний возраст жителей города.

Имеем: $N = 250000$, $n = 1000$, γ=0,9545, t=2. Для нахождения границ доверительного интервала необходимо определить выборочную среднюю и дисперсию. Дисперсию найдем по формуле разностей (14.8). Середину последнего интервала определяем в соответствии с рассуждением пункта 13.2. Расчеты удобно производить в таблице:

x	n_x	$x \cdot n_x$	$x^2 \cdot n_x$
5	20	100	500
15	80	1200	18000
25	100	2500	62500
35	200	7000	245000
45	250	11250	506250
55	120	6600	363000
65	100	6500	338000
75	80	6000	450000
85	50	4250	361250
Σ	1000	45400	2344500

Получаем: выборочная средняя (средний возраст жителей города, рассчитанный по выборке):

$$\bar{x} = \frac{\sum_x x \cdot n_x}{\sum_x n_x} = \frac{45400}{1000} = 45{,}4 \text{ (года)};$$

дисперсия

$$\sigma_x^2 = \overline{x^2} - \left(\bar{x}\right)^2 = \frac{2344500}{1000} - \left(45{,}4\right)^2 = 283{,}34.$$

Определяем среднюю ошибку выборки:

$$\mu = \sqrt{\frac{\sigma_x^2}{n} \cdot \left(1 - \frac{n}{N}\right)} = \sqrt{\frac{283{,}34}{1000} \cdot \left(1 - \frac{1000}{250000}\right)} \approx 0{,}5312.$$

Отсюда, предельная ошибка

$$\delta = t \cdot \mu = 2 \cdot 0{,}5312 = 1{,}0624 \approx 1{,}06.$$

Подсчитываем нижнюю и верхнюю границы доверительного интервала:

$$\bar{x} - \delta = 45{,}4 - 1{,}06 = 44{,}34 \text{ (года)};$$

$$\bar{x} + \delta = 45{,}4 + 1{,}06 = 46{,}46 \text{ (года)}.$$

С вероятностью 0,9545 заключаем, что средний возраст жителей города находится в пределах от 44,34 до 46, 46 года.

Пример 2.18. По данным примера 2.17 определить вероятностью 0,9545 границы интервала, в котором будет находиться доля жителей города возраста от двадцати до шестидесяти лет.

Согласно данным (пример 2.17), интересующему нас возрасту отвечают интервалы номер 1, 2, 3 и 4. Значит, число элементов выборочной совокупности, обладающих интересующим нас признаком, равно

$$m = 100 + 200 + 250 + 120 = 670.$$

Тогда, выборочная доля

$$w = \frac{m}{n} = \frac{670}{1000} = 0{,}67.$$

Для доверительной вероятности γ=0,9545 постоянная величина t=2. Тогда, предельная ошибка доли равна

$$\delta_w = t \cdot \mu_w = t \cdot \sqrt{\frac{w \cdot (1 - w)}{n} \cdot \left(1 - \frac{n}{N}\right)} = 2 \cdot \sqrt{\frac{0{,}67 \cdot (1 - 0{,}67)}{1000} \cdot \left(1 - \frac{1000}{250000}\right)} \approx 0{,}0297.$$

Отсюда, границы доверительного интервала следующие:

$$w - \delta_w = 0{,}67 - 0{,}0297 = 0{,}6403 \text{ или } 64{,}03\%;$$

$$w + \delta_w = 0{,}67 + 0{,}0297 = 0{,}6997 \text{ или } 69{,}97\%.$$

Делаем вывод с вероятностью 0,9545: в городе проживает от 64,03% до 69,97% жителей города возраста 20 – 60 лет.

В примере 2.18 выборочная доля нашлась легко, так как возраст жителей города укладывался в интервалы разбиения выборочной совокупности. На практике порой требуется определить выборочную долю элементов, попавших в произвольный интервал. В этом случае, нахождение w производится с помощью построения кумулятивной кривой и формулы (2.4).

Пример 2.19. С целью определения среднего размера вкладов отделением Сбербанка планируется организовать бесповторную выборку из общего числа вкладчиков 1000 человек. По данных предыдущих исследований была установлена дисперсия, равная 15. Каков должен быть объем выборки, если результаты необходимо гарантировать с вероятностью 0,6827, а предельная ошибка не должна превышать 0,5?

Для вероятности γ=0,6827 постоянная величина t=1. Применяя формулу (2.21), определяем необходимый объем выборочной совокупности:

$$n \geq \frac{\sigma_x^2}{\frac{\sigma_x^2}{N} + \frac{\delta^2}{t^2}} = \frac{15}{\frac{15}{1000} + \frac{(0{,}5)^2}{1^2}} \approx 56{,}60.$$

Берем минимальное целое число, удовлетворяющее неравенству. Получаем: чтобы результаты гарантировать с вероятностью 0,6827 и предельной ошибкой, не превышающей 0,5, необходимо обследовать по крайней мере 57 вкладчиков.

§ 3. ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ

3.1. Статистическая гипотеза

В параграфе 2 были рассмотрены статистические процедуры, связанные с нахождением выборочных характеристик. Здесь вводятся понятия, с помощью которых можно принимать определенные решения, делать прогнозы и выводы.

Кроме того, полученные в параграфе знания вместе с определениями пункта 2.8 позволят полностью сформировать представление о генеральной совокупности.

Статистической гипотезой *H называется гипотеза о виде неизвестного распределения или о параметрах известного распределения.*

Например, статистической можно считать гипотезу о том, что распределенная по нормальному закону генеральная совокупность имеет определенное значение математического ожидания a или среднеквадратического отклонения σ (при этом, в первом случае об σ, а во втором – об a ничего неизвестно). Такие ***гипотезы*** называются ***параметрическими***. Примером ***непараметрической гипотезы*** может служить гипотеза о том, что генеральная совокупность, которой принадлежит выборка, распределена по закону Пуассона.

Проверяемая гипотеза носит название ***нулевой*** и обозначается H_0. Вместе с ней выдвигается ***альтернативная*** или ***конкурирующая гипотеза*** H_1. Она противоположна нулевой, $H_1 = \overline{H_0}$.

Принятие или отвержение нулевой гипотезы производится с ***доверительной вероятностью*** γ. Иногда вместо γ используют ***уровень значимости*** $\alpha = 1-\gamma$. Обычно, для предварительного анализа вероятность γ принимают равной 0,95. Окончательные выводы осуществляются с вероятностью 0,99.

3.2. Статистический критерий

Для проверки нулевой гипотезы используется специально подобранная случайная величина K, точное либо приближенное значение которой приводится в таблице. Она называется ***статистическим критерием***. Для K фиксируется ***критическая область***, то есть совокупность значения критерия, при котором нулевую гипотезу отвергают. Точка $K_{крит}$ называется ***критической точкой*** (***критическим значением критерия***), если она отделяет ***критическую область*** от ***области принятия гипотезы*** (рис. 3.1). При использовании некоторых статистических критериев возникает также ***зона неопределенности***. Зона неопределенности представляет собой совокупность значений критерия, при которых нулевую гипотезу невозможно ни принять, ни отвергнуть.

Рис. 3.1

Принятие или отвержение нулевой гипотезы на практике происходит следующим образом.

1. Выбирается статистический критерий *K*.
2. Подсчитывается наблюдаемое значение критерия $K_{набл}$.
3. Выбирается уровень значимости α.
4. По таблице распределения критерия *K* для данного уровня значимости находится критическое значение критерия $K_{крит}$.
5. Вывод о принятии или отвержении гипотезы H_0 производится на основе сравнения:

если $K_{набл} < K_{крит}$, то гипотезу H_0 принимают,

если $K_{набл} > K_{крит}$, то гипотезу H_0 отвергают.

При проверке статистических гипотез возможны следующие ошибки:

1) принять гипотезу, когда она неверна;
2) отвергнуть гипотезу, когда она верна.

Чтобы избежать подобных ошибок, необходимо гипотезу проверить другим критерием. Можно нулевую гипотезу проверить и тем же критерием, но изменив выборку, скажем, увеличив объем исходной выборочной совокупности.

3.3. Критерий согласия Пирсона[9]

Рассмотрим критерий, который наиболее часто используется при проверке следующей непараметрической гипотезы: генеральная совокупность, которой принадлежит данная выборка, распределена по закону A. Он называется ***критерием согласия Пирсона*** или ***критерием хи - квадрат***[10] (обозначается χ^2).

В основе критерия Пирсона лежит сравнение эмпирических частот n_x значения x признака X (интервала группировки с серединой x) и теоретических частот n_x^*, рассчитанных в предположении, что выборочная совокупность подчиняется закону распределения A.

Теоретическая частота n_x^* интервала группировки $(\alpha;\beta)$ с серединой x может быть рассчитана следующим образом:

$$n_x^* = n \cdot p_x, \tag{3.1}$$

при этом

$$p_x = p(\alpha < X < \beta) \tag{3.2}$$

- вероятность того, что значение случайной величины, распределенной по закону A, попадет в указанный интервал. Вероятность p_x можно интерпретировать и как вероятность того, что значение случайной величины будет равно x – данной выборочной характеристике признака.

Наблюдаемое значение $\chi^2_{набл}$ критерия Пирсона подсчитывается по формуле:

$$\chi^2_{набл} = \sum_x \frac{\left(n_x - n_x^*\right)^2}{n_x^*}. \tag{3.3}$$

Критическое значение критерия находится по таблице в зависимости от уровня значимости

$$\alpha = 1 - \gamma$$

и числа степеней свободы

$$\nu = k - s - 1, \tag{3.4}$$

где k – число интервалов группировки (в случае статистического ряда k – количество значений рассматриваемого признака), s – число параметров распределения A,

$$\chi^2_{крит} = \chi(\alpha;\nu),\ \alpha = 1 - \gamma,\ \nu = k - s - 1^{11}. \tag{3.5}$$

---

[9] Пирсон Чарлз (1857 - 1936) – выдающийся английский математик, статистик, биолог и философ, основатель математической статистики.

[10] Назван по букве греческого алфавита χ – читается «хи».

[11] Некоторые значения $\chi^2_{крит}$ приведены в табл. 2 прил.

Теорема 3.1. *Если наблюдаемое значение критерия согласия Пирсона меньше критического,*

$$\chi^2_{набл} < \chi^2_{крит},$$

то это с доверительной вероятностью γ (уровнем значимости $\alpha = 1 - \gamma$*) говорит о том, что генеральная совокупность, которой принадлежит выборка, подчиняется закону распределения А. Если же*

$$\chi^2_{набл} > \chi^2_{крит},$$

то гипотеза о том, что генеральная совокупность распределена по закону А, отвергается с вероятностью γ.

3.4. Проверка статистической гипотезы о нормальном распределении генеральной совокупности

Применим критерий согласия Пирсона к проверке следующей гипотезы: генеральная совокупность, которая принадлежит выборка, распределена по нормальному (гауссову) закону распределения.

Нормальное или гауссово[12] распределение непрерывной случайной величины является пожалуй, самым широко используемым в социально – экономических исследованиях видом распределения. Там где присутствуют понятие «шанс», «риск», «случай» - непременно появляется этот вид распределения. Приведем несколько примеров.

Рассмотрим физическое понятие – броуновское движение, то есть хаотическое движение молекул жидкости или газа. Распространенной моделью такого движения является следующий опыт. На наклонную поверхность в шахматном порядке набивают гвозди. Затем по этой поверхности сверху вниз начинают сыпать дробь. Дробинки, соударяясь с гвоздями и друг с другом, начинают двигаться хаотично. Нас интересует не сам эксперимент, а его итог. На нижней части поверхности образуется «куча» из дробинок. Мы можем повторять опыт, изменять размер дробинок, скорость их передвижения по поверхности. Но «куча» будет иметь один и тот же вид (если конечно сознательно не искажать опыт, например, сыпля дробинки с одного края поверхности), такой, какой на рис. 3.2.

[12] Гаусс Карл Фридрих (1777 – 1855) – великий немецкий математик, создатель современной математической науки, внесший также фундаментальный вклад в астрономию и геодезию.

Итак, во многих опытах случайная величина X описывается с помощью колоколообразной кривой. Такое распределение носит название нормальное. Карл Гаусс впервые получил для него функцию плотности. Поэтому распределение называется также гауссовым.

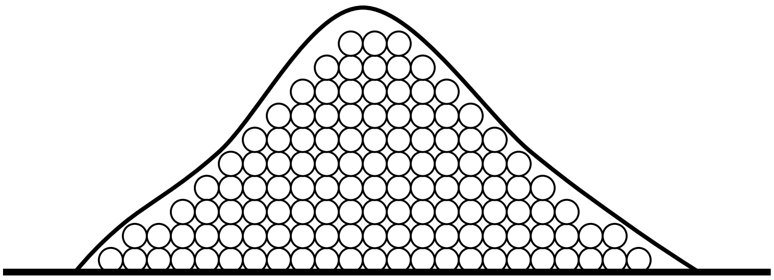

Рис. 3.2

Непрерывная случайная величина X распределена по ***нормальному (гауссову) закону распределения****, если ее функция плотности вероятности имеет вид*:

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-a)^2}{2\sigma^2}},\ -\infty<x<+\infty.$$

График функции $f(x)$ изображен на рис. 3.3.

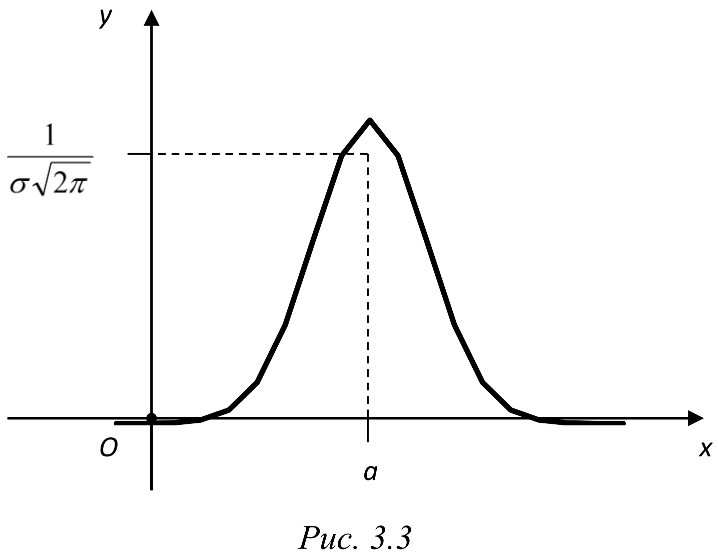

Рис. 3.3

Интегральная функция нормального распределения имеет вид:

$$F(x)=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{(x-a)^2}{2\sigma^2}}dx.$$

График функции $F(x)$ построен на рис. 3.4.

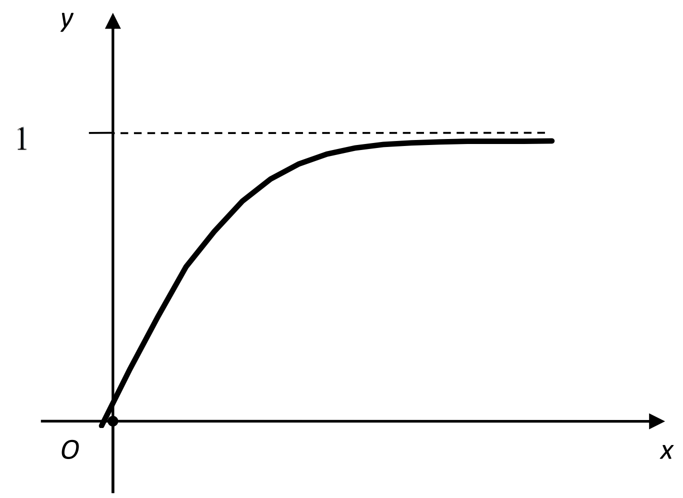

Рис. 3.4

Смысл параметров a и σ в уравнениях функции плотности и интегральной функции раскрывается в теореме:

Теорема 3.2. *Непрерывная случайная величина X, имеющая нормальное (гауссово) распределение, характеризуется следующими числовыми характеристиками:*

$$M(X)=a\,,\ D(X)=\sigma^2\ \sigma(X)=\sigma\,.$$

Таким образом, в формуле (3.4.)

$$s=2,$$

откуда

$$\nu=k-3\,. \tag{3.6}$$

Теорема 3.3. *Вероятность попадания значения нормально распределенной непрерывной случайной величины в интервал* $(x_1;x_2)$ *равна*

$$p(x_1<X<x_2)=\frac{1}{2}\cdot\left(\Phi\left(\frac{x_2-a}{\sigma}\right)-\Phi\left(\frac{x_1-a}{\sigma}\right)\right).$$

Итак, если предположить, что выборочная совокупность распределена нормально, то теоретическая частота каждого интервала группировки $(\alpha;\beta)$ равна

$$n_x^*=\frac{n}{2}\cdot(\Phi(y_2)-\Phi(y_1)),\ y_2=\frac{\beta-\bar{x}}{\sigma},\ y_1=\frac{\alpha-\bar{x}}{\sigma}\,. \tag{3.7}$$

Приведем алгоритм проверки нулевой гипотезы.

1. Сравнение полигона частот с графиком функции плотности идеального нормального распределения. Сравниваем полигон частот выборочного распределения с графиком функции плотности нормального распределения (см. рис. 3.3). Если они похожи (речь идет не о совпадении, а лишь о внешнем сходстве), то переходим к следующему пункту алгоритма. В противном случае, например, если полигоны частот выглядят также как на рис. 3.5 сразу утверждается, что с помощью данной выборки проверить нулевую гипотезу невозможно.

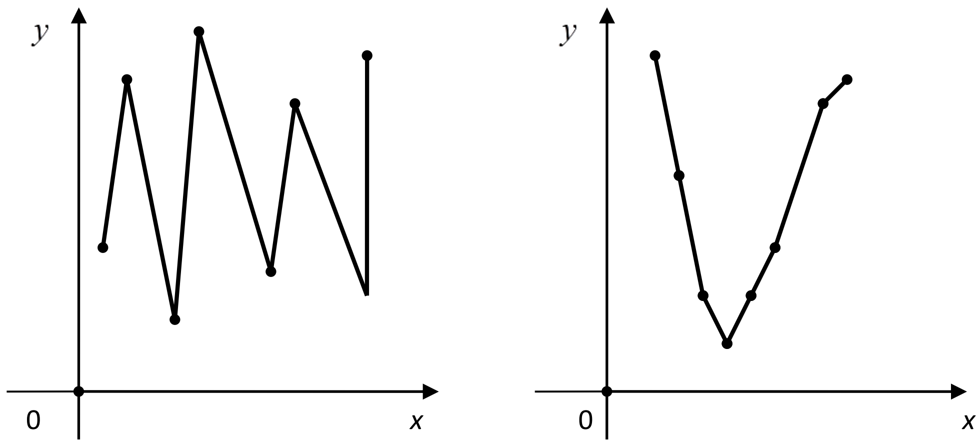

Рис. 3.5

2. Сравнение кумулятивной кривой с графиком интегральной функции нормального распределения. Следующим этапом проверки гипотезы является сравнение полигона относительных частот с графиком интегральной функции нормального распределения (см. рис. 3.4). Выводы аналогичны выводам п. 1 алгоритма.

3. Определение вида распределения. На исследуемое распределение оказывает влияние множество случайных факторов. Поэтому полигон часто выборки может отличаться от графика функции плотности идеального нормального распределения (см. рис. 3.3).

Во – первых, возможно смещение графика эмпирического распределения относительно прямой $x = a$, распределение становится несимметричным (рис. 3.6).

В зависимости от смещения вершины эмпирического распределения влево или вправо от точки $\left(a; \frac{1}{\sigma\sqrt{2\pi}}\right)$ различают ***левостороннее*** и ***правостороннее*** распределения. На рис. 3.6 цифрой I обозначено левостороннее распределение, а цифрой II – правостороннее.

Аналитическим способом определения степени несимметричности эмпирического распределения является расчет асимметрии: чем ближе значение асимметрии к нулю, тем ближе имеющееся распределение к идеальному нормальному. Кроме того, если значение асимметрии отрицательно, то распределение правостороннее, если же величина асимметрии – положительное число, то имеет место левостороннее распределение.

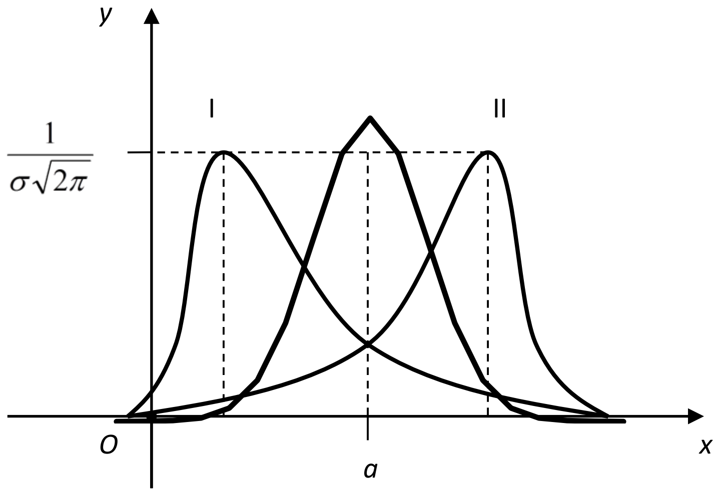

Рис. 3.6

Далее, график эмпирического распределения может быть смещен вдоль прямой $x = a$ вверх или вниз относительно точки $\left(a; \frac{1}{\sigma\sqrt{2\pi}}\right)$ (рис. 3.7).

На рис. 3.7 цифрой I обозначено ***островершинное распределение***, а цифрой II - ***туповершинное распределение***.

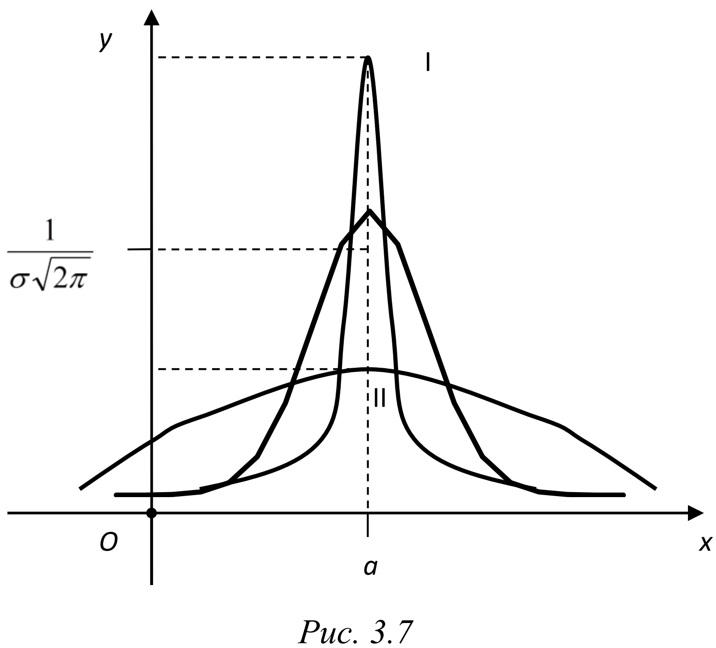

Рис. 3.7

Этот тип искажения можно определить с помощью эксцесса: чем ближе величина эксцесса к нулю, тем ближе распределение к идеальному нормальному. Если значение эксцесса – отрицательное число, то распределение туповершинное, а если положительное – островершинное.

Итак, в зависимости от знака асимметрии и эксцесса мы имеем четыре вида распределения, представленные в табл. 3.1.

Таблица 3.1

as / *es*		<0	>0
		правостороннее	**левостороннее**
<0	**Туповершинное**	правостороннее туповершинное	левостороннее туповершинное
>0	**Островершинное**	правостороннее островершинное	левостороннее островершинное

На рис. 3.8 показано правостороннее островершинное распределение. Значит, у него

$$as < 0, \ es > 0.$$

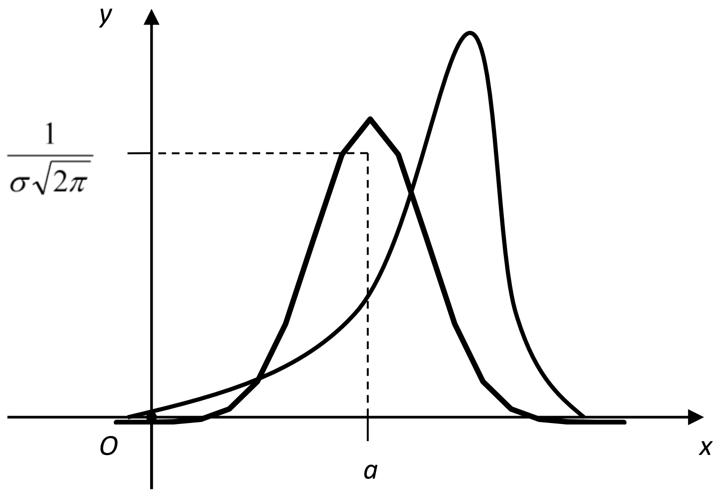

Рис. 3.8

Заметим, что каков бы не был вид распределения, площадь криволинейной трапеции, образованной графиком эмпирического распределения и осью абсцисс, равна 1.

4. Определение типа распределения. В зависимости от значения асимметрии и эксцесса определяем один из четырех типов распределения:

1. если

$$as = es = 0$$

- ***идеальное нормальное распределение***;

2. если

$$|\,as\,| < 0{,}1\,,\ \ |\,es\,| < 1$$

- ***нормальное распределение***;

3. если

$$|\,as\,| < 0{,}5\,,\ \ |\,es\,| < 0{,}5$$

- ***распределение, близкое к нормальному***;

4. если

$$|\,as\,| < 1\,,\ \ |\,es\,| < 1$$

- ***распределение нормального типа***.

На практике необходимо определять тип распределения, подставляя последовательно значения асимметрии и эксцесса в условия 1 – 4, начиная с условия 1.

Возникает вопрос: может ли быть значение асимметрии и эксцесса по модулю больше 1? Да, может. Но в этом случае график эмпирического распределения не имеет ничего общего с графиком функции плотности нормального распределения (рис. 3.3). Другими словами, если первый и второй пункты алгоритма пройдены, то, скорее всего, значения $|\,as\,| > 1$ и $|\,es\,| > 1$ говорят о допущенной в расчетах ошибке.

5. Нахождение теоретических частот. Для каждого интервала группировки по формуле (3.7) находим теоретическую частоту.

Проверка правильности произведенных расчетов осуществляется следующим образом. Во – первых, значение теоретической частоты не может выражаться отрицательным числом. Далее, значение теоретической частоты не может сильно отличаться от соответствующей эмпирической частоты. Наконец, несложно доказать, что сумма теоретических частот близка к сумме эмпирических частот, то есть к объему выборки,

$$\sum_x n_x^* \approx \sum_x n_x = n\,.$$

Если эти условия нарушается, то в расчете теоретической частоты допущена ошибка.

В основном ошибки на этом этапе происходят из-за неправильного пользования табл. 1 прил. Часто не применяется свойство нечетности функции Лапласа $\Phi(x)$ и минус не учитывается в расчетах.

6. Построение кривой теоретических частот. Для того, чтобы еще раз убедиться в правильности произведенных расчетов, необходимо на одном чертеже построить полигон частот выборки и ***кривую теоретических частот***, то есть кривую, определенную точками $\left(x; n_x^*\right)$. Будем считать, что расчеты верны, если два графика не сильно отличаются друг от друга как формой, так и расстоянием между точками $\left(x; n_x^*\right)$ и $\left(x; n_x\right)$. Кроме того, кривая теоретических частот не должна противоречить выводам пункта 3 алгоритма.

7. Проверка нулевой гипотезы. С доверительной вероятностью γ (уровнем значимости α=1-γ) выдвигаем гипотезу о том, что генеральная совокупность, которой принадлежит выборка распределена по нормальному (гауссову) закону распределения.

Для этого, пользуясь табл. 2 прил., находим критическое значение (3.5) критерия Пирсона в соответствие с уровнем значимости и числом степеней свободы (3.6).

Затем, находим наблюдаемое значение критерия хи – квадрат по формуле (3.3).

Сравнение критического и наблюдаемого значений критерия согласия, а также выводы о принятии или отвержении нулевой гипотезы, производятся с помощью теоремы 3.1.

Пример 3.1. С вероятностью 0,95 проверить гипотезу о том, что генеральная совокупность, которой принадлежит выборка из примера 2.1, подчиняется нормальному закону распределения.

Группированный статистический ряд данной выборки построен в примере 2.3.

Применяем вышеизложенный алгоритм к проверке данной статистической гипотезы.

1. Сравнение рис. 2.2 и 3.3 говорит о том, что построенные на них графики схожи друг с другом.

2. Сравниваем кумулятивную кривую выборки (рис. 2.4 или 2.5) с графиком интегральной функции нормального распределения (рис. 3.4). Видим, что они похожи друг на друга.

3. Значения асимметрии (2.15) и эксцесса (2.16) выборки были определены в примере 2.14 согласно определению и получены методом моментов в примере 2.15:

$$as = 0{,}2958,\ es = -0{,}2633. \tag{3.8}$$

Так как

$$as = 0{,}2958 > 0\,,\ es = -0{,}2633 < 0\,,$$

то определенное выборкой распределение является (см. табл. 3.1) левосторонним туповершинным.

4. Значения (3.8) не отвечают ни условию 1, ни условию 2 пункта 3 алгоритма. Значит, эмпирическое распределение не является ни идеальным нормальным, ни нормальным. Но, поскольку одновременно

$$|as| = |0{,}2958| = 0{,}2958 < 0{,}5, \quad |es| = |-0{,}2633| = 0{,}2633 < 0{,}5,$$

то условие 3 выполнено. Поэтому, рассматриваемое распределение есть близкое к нормальному.

5. Переходим к расчету теоретических частот. Имеем:

$$n = 100,$$

значит согласно (3.7) теоретические частоты рассчитываются по формуле:

$$n_x^* = 50 \cdot (\Phi(y_2) - \Phi(y_1)). \tag{3.9}$$

Средняя выборочная и среднеквадратическое отклонение найдены в примерах 2.14 и 2.15. Они соответственно равны (2.12) и (2.14):

$$\bar{x} = 101{,}36, \quad \sigma_x = 6{,}88.$$

Следовательно, в (3.7)

$$y_2 = \frac{\beta - 101{,}36}{6{,}88}, \quad y_1 = \frac{\alpha - 101{,}36}{6{,}88}.$$

Находим значения аргумента функции Лапласа y_2 и y_1, округляем их до сотых. Затем, пользуясь табл. 1 прил., определяем значения функции Лапласа $\Phi(y_2)$ и $\Phi(y_1)$. Наконец, подставляем их в (3.9), получаем значения теоретических частот. Вычисления удобно производить в таблице:

k	$(\alpha;\beta)$	x	n_x	y_2	$\Phi(y_2)$	y_1	$\Phi(y_1)$	n_x^*
1	$(88;92)$	90	10	-1,36	-0,8262	-1,94	-0,9476	6,070
2	$(92;96)$	94	13	-0,78	-0,5646	-1,36	-0,8262	13,080
3	$(96;100)$	98	18	-0,20	-0,1585	-0,78	-0,5646	20,305
4	$(100;104)$	102	27	0,38	0,2960	-0,20	-0,1585	22,725
5	$(104;108)$	106	17	0,97	0,6679	0,38	0,2960	18,595
6	$(108;112)$	110	7	1,55	0,8789	0,97	0,6679	10,550
7	$(112;116)$	114	5	2,13	0,9668	1,55	0,8789	4,395
8	$(116;120)$	118	3	2,71	0,9933	2,13	0,9668	1,325
Σ	-	-	100	-	-	-	-	97,045

Сравнение столбцов 4 и 9 таблицы убеждает нас в правильности произведенных вычислений.

6. Строим полигон частот и кривую теоретических частот (рис. 3.9). На рис. 3.9 полигон частот изображен пунктиром, а кривая теоретических частот – сплошной линией.

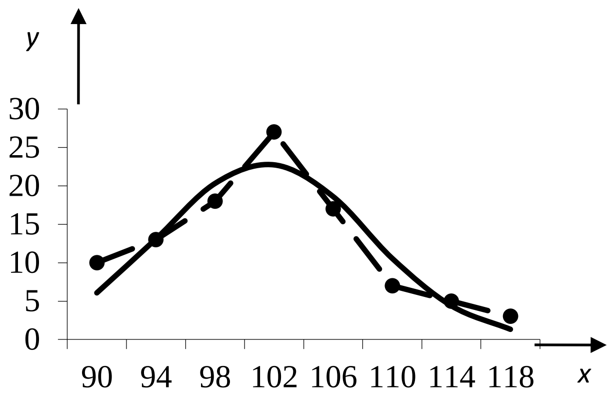

Рис. 3.9

Сравнение графиков наглядно подтверждает правильность произведенных вычислений.

7. С вероятностью 0,95 выдвигаем гипотезу о нормальном распределении генеральной совокупности, которой принадлежит исследуемая выборка.

Уровень значимости

$$\alpha = 1 - \gamma = 1 - 0{,}95 = 0{,}05.$$

Так как количество интервалов группировки

$$k = 8,$$

то число степеней свободы согласно (15.6) равно

$$\nu = k - 3 = 8 - 3 = 5.$$

С помощью табл. 2 прил. находим критическое значение критерия Пирсона:

$$\chi^2_{крит} = \chi(\alpha;\nu) = \chi(0{,}05;5) = 11{,}1.$$

Расчет наблюдаемого значения критерия Пирсона удобно производить в таблице:

k	x	n_x	n_x^*	$\frac{\left(n_x - n_x^*\right)^2}{n_x^*}$
1	90	10	6,070	2,5445
2	94	13	13,080	0,0005
3	98	18	20,305	0,2617
4	102	27	22,725	0,8042
5	106	17	18,595	0,1368
6	110	7	10,550	1,1945
7	114	5	4,395	0,0833
8	118	3	1,325	2,1175
Σ	-	100	97,045	7,1430

Имеем

$$\chi^2_{набл} = 7{,}143.$$

Сравниваем критическое и наблюдаемое значения, получаем: поскольку

$$7{,}143 < 11{,}1, \ \chi^2_{набл} < \chi^2_{крит},$$

то гипотеза о том, что генеральная совокупность, которой принадлежит выборка из примера 2.1, распределена по нормальному закону, принимается с вероятностью 0,95.

3.5. Проверка статистической гипотезы о распределении Пуассона генеральной совокупности

Применим критерий согласия Пирсона для проверки гипотезы о распределении генеральной совокупности, которой принадлежит выборка, по закону Пуассона или по закону редких чисел.

Дискретная случайная величина X, принимающая только целые, неотрицательные значения (причем последовательность этих значений теоретически неограниченна), имеет ***распределение Пуассона*** *(подчиняется* ***закону редких чисел****), если ее закон распределения имеет вид*:

X	0	1	…	M	…
p_X	$p(0)$	$p(1)$	…	$p(m)$	…

$$p(m) = \frac{\lambda^m}{m!} \cdot e^{-\lambda}. \qquad (3.10)$$

Теорема 3.4. *Дискретная случайная величина X, распределенная по закону Пуассона, характеризуется следующими числовыми характеристиками:*

$$M(X) = D(X) = \lambda\,,\ \sigma(X) = \sqrt{\lambda}\,.$$

Итак, параметр λ является одновременно математическим ожиданием и дисперсией. Поэтому, число степеней свободы (3.4) равно

$$\nu = k - s - 1 = k - 1 - 1 = k - 2\,.$$

В качестве λ берем выборочную среднюю, рассчитанную по формуле средней арифметической,

$$\lambda = \bar{x}\,.$$

Теорема 3.5 может служить одним из способов проверки гипотезы о распределении случайной величины по закону Пуассона. Если математическое ожидание приближенно равно дисперсии,

$$M(X) \approx D(X),$$

то есть основания утверждать, что случайная величина распределена по закону Пуассона; резкое различие этих характеристик свидетельствует против этой гипотезы.

Вероятность того, что дискретная случайная величина X, распределенная по закону Пуассона примет значение, не меньшее m, равна

$$p(X \geq m) = 1 - p(X < m) = 1 - (p(0) + p(1) + \ldots + p(m-1)).$$

Следуя полученному результату определим вероятность $p(X > 0)$того, что распределенная по закону редких чисел случайная величина примет положительное значение:

$$p(X > 0) = 1 - e^{-\lambda}\,.$$

Итак, алгоритм проверки нулевой гипотезы выглядит следующим образом.

1. Сравнение выборочной средней и дисперсии. Сравниваем выборочную среднюю и дисперсию. Если их значения приближенно равны, то переходим к следующему пункту алгоритма. В противном случае констатируем, что с помощью данной выборки проверить гипотезу о распределении генеральной совокупности по закону Пуассона невозможно.

2. Нахождение теоретических частот. По формуле (3.10) находим вероятность p_x того, что случайная величина примет значение, равное x. Эти вычисления можно производить как с помощью калькулятора, так и с помощью табл. 3 прил. Затем, умножаем p_x на объем выборки согласно (3.1). Получаем значения теоретических частот. Также, как и при проверке предыдущей гипотезы

$$\sum_x n_x^* \approx \sum_x n_x = n.$$

3. Построение кривой теоретических частот. Строим на одном чертеже полигон частот выборки и кривую теоретических частот, то есть кривую, определенную точками $(x; n_x^*)$. Будем считать, что расчеты верны, если два графика не сильно отличаются друг от друга формой, так расстоянием между точками $(x; n_x^*)$ и $(x; n_x)$.

4. Проверка нулевой гипотезы. С доверительной вероятностью γ (уровнем значимости α=1-γ) выдвигаем гипотезу о том, что генеральная совокупность, которой принадлежит выборка подчиняется распределению Пуассона.

Для этого, пользуясь табл. 2 Приложения, находим критическое значение (3.5) критерия Пирсона в соответствие с уровнем значимости и числом степеней свободы (3.6).

Затем, находим наблюдаемое значение критерия хи – квадрат по формуле (3.3).

Принятие или отвержение нулевой гипотезы производится с помощью теоремы 15.1.

Пример 3.2. Выборочная совокупность представлена статистическим рядом:

x	0	1	2	3	4	5	6	7
n_x	4	10	15	21	19	14	10	7

С вероятностью 0,99 проверить гипотезу о том, что генеральная совокупность, которой принадлежит выборка, подчиняется закону распределения Пуассона.

1. Рассчитаем выборочную среднюю и дисперсию. Дисперсию найдем по формуле разностей.

k	x	n_x	$x \cdot n_x$	$x^2 \cdot n_x$
1	0	4	0	0
2	1	10	10	10
3	2	15	30	60
4	3	21	63	189
5	4	19	76	304
6	5	14	70	350
7	6	10	60	360
8	7	7	49	343
Σ	-	100	358	1616

Получаем:

$$\bar{x} = \frac{\sum_x x \cdot n_x}{\sum_x n_x} = \frac{358}{100} = 3{,}58;$$

$$\overline{x^2} = \frac{\sum_x x^2 \cdot n_x}{\sum_x n_x} = \frac{1616}{100} = 16{,}16;$$

$$\sigma_x^2 = \overline{x^2} - \left(\bar{x}\right)^2 = 16{,}16 - (3{,}58)^2 = 3{,}3436.$$

Сравнение выборочной средней и дисперсии показывает, что

$$\bar{x} \approx \sigma_x^2.$$

2. Используем формулу (3.5):

$$p_x = \frac{\lambda^x}{x!} \cdot e^{-\lambda} = \frac{(3{,}58)^x}{x!} \cdot e^{-3{,}58}.$$

Далее, определяем значения теоретических частот (табл. 3.2):

$$n_x^* = n \cdot p_x = 100 \cdot p_x.$$

Таблица 3.2

k	x	n_x	p_x	n_x^*
1	0	4	0,0279	2,79
2	1	10	0,0998	9,98
3	2	15	0,1786	17,86
4	3	21	0,2132	21,32
5	4	19	0,1908	19,08
6	5	14	0,1366	13,66
7	6	10	0,0815	8,15
8	7	7	0,0417	4,17
Σ	-	100	0,9701	97,01

3. На одном чертеже (рис. 3.10) строим полигон частот (пунктирная ломаная) и кривую теоретических частот (сплошная линия). Видим, что графики наглядно подтверждают правильность произведенных вычислений.

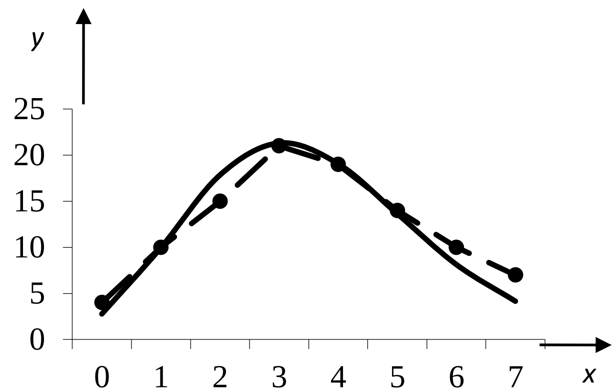

Рис. 3.10

4. С вероятностью 0,99 выдвигаем гипотезу о распределении генеральной совокупности, которой принадлежит исследуемая выборка, по закону Пуассона.

Уровень значимости

$$\alpha = 1 - \gamma = 1 - 0{,}99 = 0{,}01.$$

Так как число значений случайной величины

$$k = 8,$$

то число степеней свободы согласно (15.6) равно

$$\nu = k - 2 = 8 - 2 = 5.$$

С помощью табл. 2 прил. находим критическое значение критерия Пирсона:

$$\chi^2_{крит} = \chi(\alpha; \nu) = \chi(0{,}01; 6) = 16{,}8.$$

Расчет наблюдаемого значения критерия Пирсона удобно производить в табл. 3.3.

Таблица 3.3

k	x	n_x	n_x^*	$\frac{\left(n_x - n_x^*\right)^2}{n_x^*}$
1	0	4	2,79	0,527336
2	1	10	9,98	0,000042
3	2	15	17,86	0,458958
4	3	21	21,32	0,004710
5	4	19	19,08	0,000324
6	5	14	13,66	0,008448
7	6	10	8,15	0,419618
8	7	7	4,17	1,923387
Σ	-	100	97,01	3,342823

Имеем

$$\chi^2_{набл} = 3{,}342823.$$

Сравниваем критическое и наблюдаемое значения, получаем: поскольку

$$3{,}342823 < 16{,}8, \ \chi^2_{набл} < \chi^2_{крит},$$

то гипотеза о том, что генеральная совокупность, которой принадлежит выборка, подчиняется закону распределения Пуассона, принимается с вероятностью 0,99.

§ 4. ПАРНАЯ РЕГРЕССИЯ

4.1. Введение. Функциональная и корреляционная зависимость

Важнейшей задачей, возникающей в практике экономических исследований, является изучение связей между явлениями, моделирование этих связей, прогнозирование значений одного признака от другого или других экономических показателей.

С одним из способов моделирования зависимости одного признака от другого мы познакомились в школе. Это функциональная зависимость. Напомним еще раз ее определение.

Пусть $x = \{X\}$ и $x = \{X\}$ - два множества элементов любой природы. Правило f, согласно которому каждому элементу $x \subset X$ ставится в соответствие не более одного элемента $y \subset Y$, называется ***функцией,***

$$y = f(x). \tag{4.1}$$

Зависимость между множествами, определенная этим правилом, есть ***функциональная зависимость***. На рис. 4.1 представлен пример функции как отображения

$$X \xrightarrow{f} Y.$$

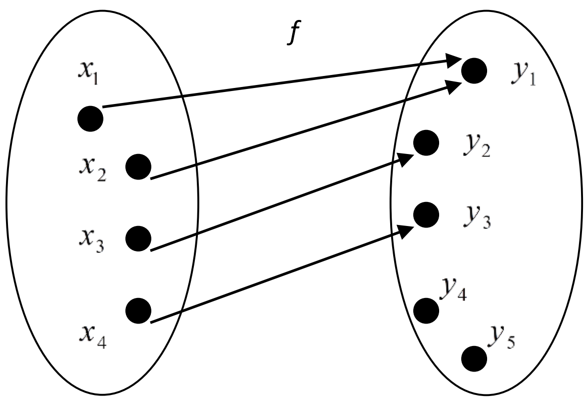

Рис. 4.1

Величина x называется переменной или аргументом, *а* y – значением функции.

Самыми известными способами задания функции являются графический и аналитический, при котором функция представляется в виде уравнения (4.1).

Наиболее известными в экономике являются следующие функции:

1. Линейная функция

$$y = a \cdot x + b; \qquad (4.2)$$

2. Степенная функция

$$y = b \cdot x^a;$$

3. Параболическая функция

$$y = a \cdot x^2 + b \cdot x + c; \qquad (4.3)$$

4. Гиперболическая функция

$$y = a \cdot \frac{1}{x} + b;$$

5. Показательная функция

$$y = b \cdot a^x; \qquad (4.4)$$

6. Дробно-линейная функция

$$y = \frac{a \cdot x + b}{c \cdot x + d}.$$

Однако не все виды зависимости могут быть идентифицированы как функциональные. Например, пусть x – товарооборот торгующей организации, y – сумма ее издержек обращения. Если эмпирические данные о величинах товарооборота и суммах издержек обращения для нескольких торгующих организаций изобразить на координатной плоскости в виде точек с координатами $(x; y)$, то мы получим прямолинейно вытянутое облако (рис. 4.2).

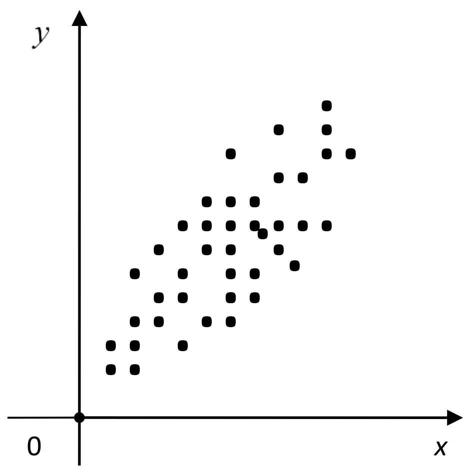

Рис. 4.2

Итак, ***корреляционной*** *называется такой вид зависимости между множествами X и Y, согласно которому каждому элементу* $x \subset X$ *ставится в соответствие множество элементов* $y \subset Y$*, распределенный по известному закону распределения.*

Если для каждого значения *x* найти ***групповую среднюю*** $\overline{y}_x$ [13] (на рис. 4.3 групповая средняя выделена крупной точкой в «срезе» корреляционного облака, соответствующего *x*), то полученная зависимость групповой средней от *x* будет уже функциональной.

Определенная таким образом функциональная зависимость будет приближенно, в среднем, описывать корреляционную. Она называется ***регрессионной зависимостью*** или ***регрессией*** *y* ***на*** *x*. Ее уравнение – ***уравнение регрессии*** *y* ***на*** *x*,

$$\overline{y}_x = f(x), \tag{4.5}$$

а график – ***линия регрессии*** *y* ***на*** *x*.

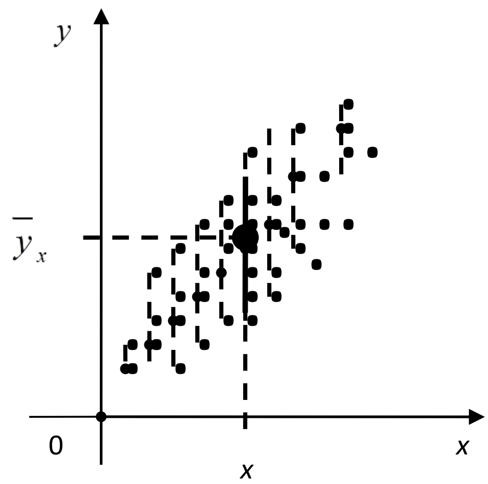

Рис. 4.3

В примере с торгующими организациями видно, что регрессионная зависимость является линейной прямой (возрастающей) с уравнением регрессии *y* на *x* вида (4.2), причем в нем коэффициент $a > 0$ (рис. 4.4).

[13] Ее определение будет дано в п. 4.3.

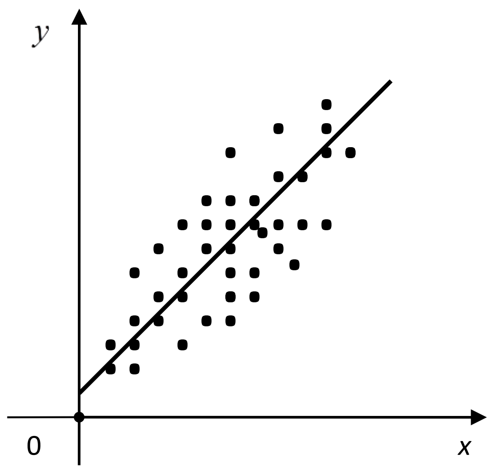

Рис. 4.4

Вместе с регрессией y на x может быть построена ***регрессия** x **на** y*. Она определяется групповыми средними $\bar{x}_y$, рассчитанными для каждого значения y (рис. 4.5). Уравнение регрессии x на y имеет общий вид:

$$\bar{x}_y = \varphi(y).$$

Ее график может отличаться от линии регрессии y на x.

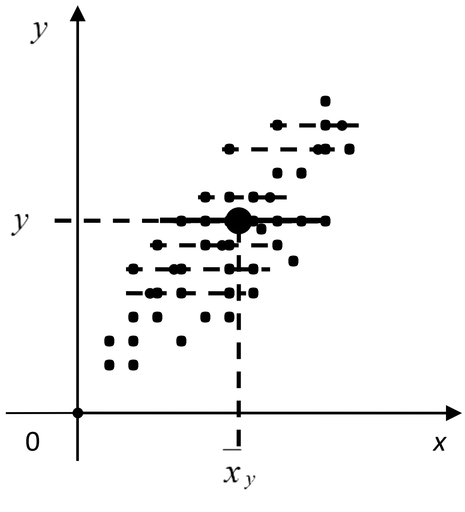

Рис. 4.5

Будем называть *переменную, входящую в правую часть уравнения регрессии **объясняющей**. Переменная, выражающаяся через объясняющую переменную, то есть находящаяся в левой части уравне-*

ния регрессии – ***объясняемая*** *или* ***зависимая***. Соответственно, *признак, состоящий из объясняющих переменных называется* ***факторным****, а признак, состоящий из объясняемых переменных –* ***результативным***.

Перечислим некоторые уравнения регрессии x на y:

1. линейная

$$x = a' \cdot y + b'; \tag{4.6}$$

2. параболическая

$$x = a' \cdot y^2 + b' \cdot y + c'; \tag{4.7}$$

3. показательная

$$x = b' \cdot (a')^y. \tag{4.8}$$

4.2. Способы задания корреляционной зависимости

Рассмотрим основные способы задания корреляционной зависимости.

Обозначим: x и y – соответственно значения факторов X и Y (если группировочные признаки представлены в виде интервалов, то в качестве x и y рассматриваем их середины); n_x и n_y - частоты значений признаков x и y; n_{xy} - частота пары $(x; y)$.

Эмпирические данные принято представлять в виде ***корреляционной таблицы*** (табл. 4.1):

Таблица 4.1

Y \ X	x_1	x_2	…	x_k	n_y
y_1	$n_{x_1y_1}$	$n_{x_2y_1}$	…	$n_{x_ky_1}$	$n_{y_1} = \sum_{j=1}^{k} n_{x_jy_1}$
y_2	$n_{x_1y_2}$	$n_{x_2y_2}$	…	$n_{x_ky_2}$	$n_{y_2} = \sum_{j=1}^{k} n_{x_jy_2}$
⋮	⋮	⋮	⋮	⋮	⋮
y_m	$n_{x_1y_m}$	$n_{x_2y_m}$	…	$n_{x_ky_m}$	$n_{y_m} = \sum_{j=1}^{k} n_{x_jy_m}$
n_x	$n_{x_1} = \sum_{i=1}^{m} n_{x_1y_i}$	$n_{x_1} = \sum_{i=1}^{m} n_{x_2y_i}$	…	$n_{x_1} = \sum_{i=1}^{m} n_{x_ky_i}$	$n = \sum_{j=1}^{k} n_{x_j} = \sum_{i=1}^{m} n_{y_i}$

В первой строке и первом столбце корреляционной табл. 4.1 указаны значения признака, в последней строке и последнем столбце – соответственно их частоты, а на пересечении строк и столбцов – частоты пар $(x;y)$. Условимся, что если в корреляционной таблице не стоит числа, то это значит, что пары $(x;y)$ в данном случае не существует. Обратим внимание на то, что сумма частот пар в фиксированной строке дает величину частоты данного значения x,

$$n_x = \sum_y n_{xy} \, , \tag{4.9}$$

а для фиксированного столбца

$$n_y = \sum_x n_{xy} \, . \tag{4.10}$$

Объем выборки можно рассчитать и как сумму частот по каждому из группировочных признаков, и как сумму пар,

$$n = \sum_x n_x = \sum_y n_y = \sum_x \sum_y n_{xy} \, .$$

Пример 4.1. Дана зависимость между признаками X и Y:

y \ x	(0,5;1,5)	(1,5;2,5)	(2,5;3,5)	(3,5;4,5)	(4,5;5,5)
(23;33)				37	3
(33;43)			13	6	
(43;53)		13	10		
(53;63)	17	1			

Построить корреляционную таблицу.

Ищем середины каждого из интервалов группировки, по формулам (4.9) и (4.10) находим частоты группировочных признаков и объем выборочной совокупности. Получаем корреляционную таблицу (табл. 4.2):

Таблица 4.2

y \ x	1	2	3	4	5	n_y
28				37	3	40
38			13	6		19
48		13	10			23
58	17	1				18
n_x	17	14	23	43	3	100

Множество точек на координатной плоскости, соответствующих значениям признаков X и Y, называется ***корреляционным полем данных*** *или* ***коррелограммой***. По сути, коррелограмма нами была изображена на рис. 4.6.

Пример 4.2. По данным примера 4.1 построить корреляционное поле данных.

На координатной плоскости (рис. 4.6) отмечаем точки, координаты которых $(x;y)$ соответствуют парам значений признаков из табл. 4.2. Тем самым, получаем коррелограмму.

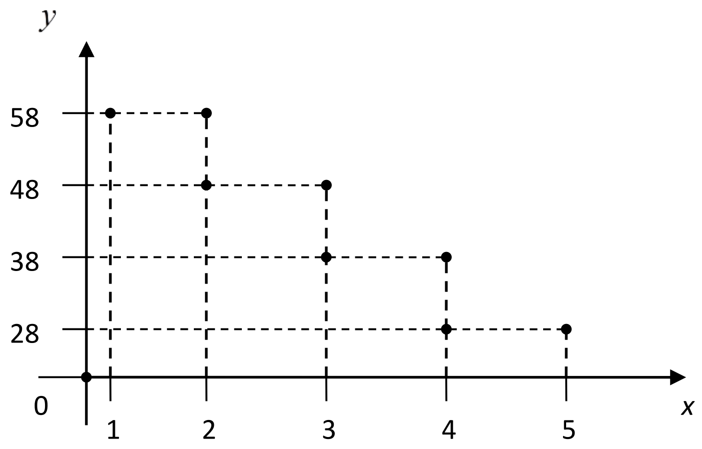

Рис. 4.6

Корреляционное поле данных позволяет наглядно увидеть взаимоположение пар $(x;y)$. Однако, коррелограмма не несет информации о частоте пар. Указанный недостаток устранен в следующем графическом представлении эмпирических данных.

Корреляционной поверхностью *называется совокупность точек с координатами* $(x;y;n_{xy})$.Корреляционную поверхность удобно строить так. Каждой паре $(x;y)$ на плоскости xOy ставим в соответствие прямоугольник с вершинами в точках с координатами

$$(x;y;0),\ (x-h_x;y;0),\ (x;y-h_y;0),\ (x-h_x;y-h_y;0),$$

где h_x и h_y - шаги варьирования рассматриваемых признаков[14]. Эти прямоугольники затем достраиваются до параллелепипеда высотой n_{xy}.

[14] См. пункт 2.7.

Пример 4.3. Построить корреляционную поверхность по эмпирическим данным примера 4.1.

Из табл. 4.2 заключаем, что

$$h_x = 1, \; h_y = 10. \tag{4.11}$$

Значит, каждый прямоугольник будет определятся вершинами:

$$(x; y; 0), \; (x-1; y; 0), \; (x; y-10; 0), \; (x-1; y-10; 0).$$

Следуя определению, строим корреляционную поверхность (рис. 4.7).

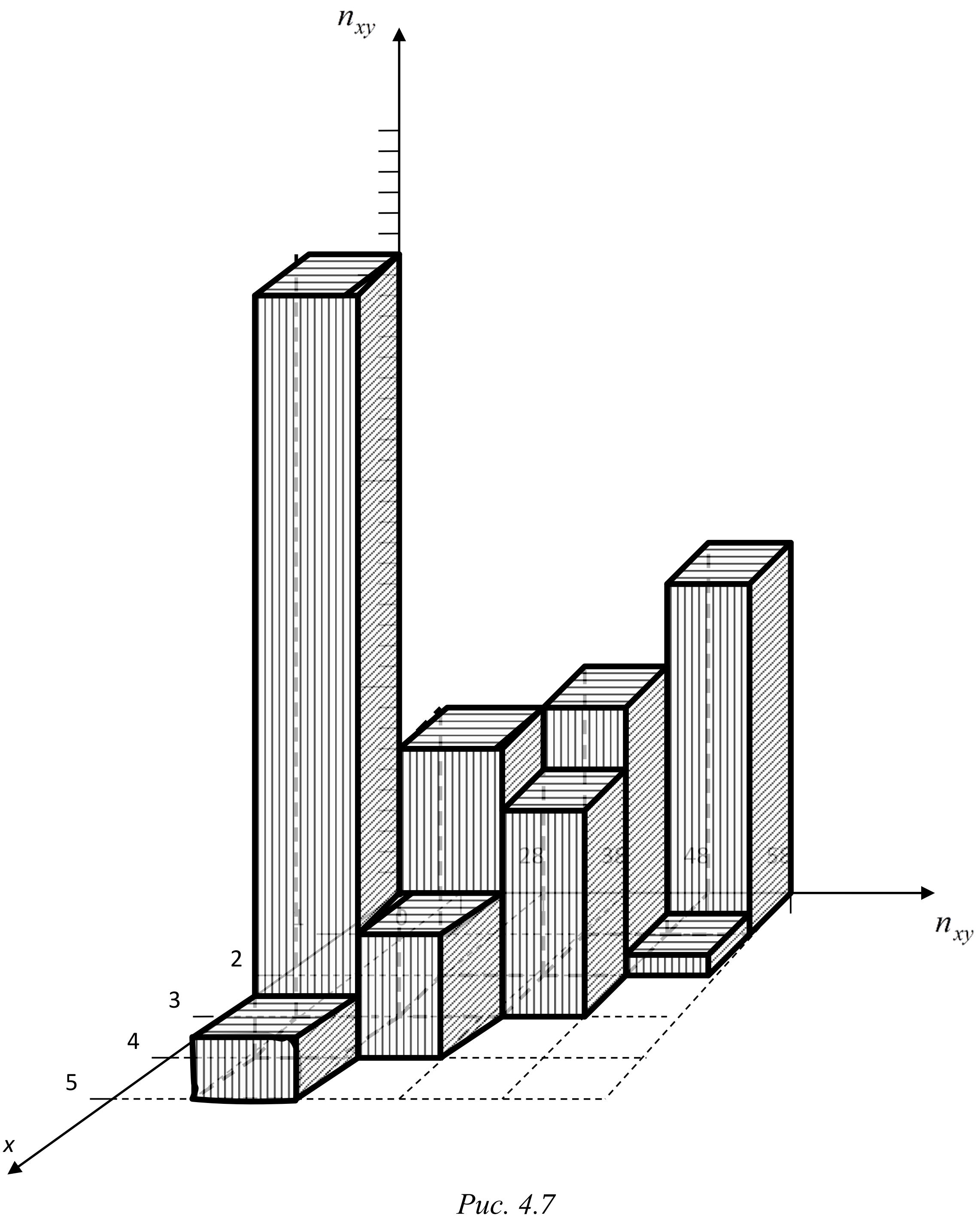

Рис. 4.7

4.3. Эмпирические линии регрессии

Определим групповую среднюю. Для фиксированного значения *x* ***групповая средняя*** $\overline{y}_x$ подсчитывается следующим образом:

$$\overline{y}_x = \frac{\sum\limits_y y \cdot n_{xy}}{n_x}. \qquad (4.12)$$

Аналогично рассчитывается групповая средняя

$$\overline{x}_y = \frac{\sum\limits_x x \cdot n_{xy}}{n_y}. \qquad (4.13)$$

Пример 4.4. По данным примера 4.1 найти групповые средние $\overline{y}_x$ и $\overline{x}_y$.

Фиксируем значения *x* признака *X* (столбцы табл. 4.2) и, следуя (4.12), находим групповые средние:

$$x = 1,\ \overline{y}_x = \frac{58 \cdot 17}{17} = 58;$$

$$x = 2,\ \overline{y}_x = \frac{48 \cdot 13 + 58 \cdot 1}{14} \approx 48{,}71;$$

$$x = 3,\ \overline{y}_x = \frac{38 \cdot 13 + 48 \cdot 10}{23} \approx 42{,}35;$$

$$x = 4,\ \overline{y}_x = \frac{28 \cdot 37 + 38 \cdot 6}{43} \approx 29{,}40;$$

$$x = 5,\ \overline{y}_x = \frac{28 \cdot 3}{3} = 28.$$

Полученные результаты запишем в табл. 4.3.

Таблица 4.3

x	1	2	3	4	5
$\overline{y}_x$	58	48,71	42,35	29,40	28

Действуем аналогично для поиска $\overline{x}_y$. Здесь фиксируются значения *y* признака *Y* (строки). Имеем согласно (4.13):

$$y = 28,\ \overline{x}_y = \frac{4 \cdot 37 + 5 \cdot 3}{40} \approx 4{,}08;$$

$$y = 38,\ \overline{x}_y = \frac{3 \cdot 13 + 4 \cdot 6}{19} \approx 3{,}32;$$

$$y = 48,\ \overline{x}_y = \frac{2 \cdot 13 + 3 \cdot 10}{23} \approx 2{,}43;$$

$$y = 58,\ \overline{x}_y = \frac{1 \cdot 17 + 2 \cdot 1}{18} \approx 1{,}06.$$

Полученные групповые средние записываем в табл. 4.4.

Таблица 4.4

y	28	38	48	58
$\bar{x}_y$	4,08	3,32	2,43	1,06

Первоначальные выводы о форме корреляционной зависимости можно сделать, построив эмпирические линии регрессии. ***Эмпирической линией регрессии** y **на** x называется ломаная с вершинами в точках с координатами* $(x; \bar{y}_x)$. Аналогично строится ***эмпирическая линия регрессии** x **на** y – ломаная с вершинами в точках с координатами* $(\bar{x}_y; y)$.

Пример 4.5. По данным примера 4.1 построить эмпирические линии регрессии и сделать первоначальные выводы о форме корреляционной зависимости.

Групповые средние нами были получены в примере 4.4. Используя табл. 4.3 и 4.4 на рис. 4.8 строим эмпирические линии регрессии y на x (сплошная линия) и x на y (пунктирная линия).

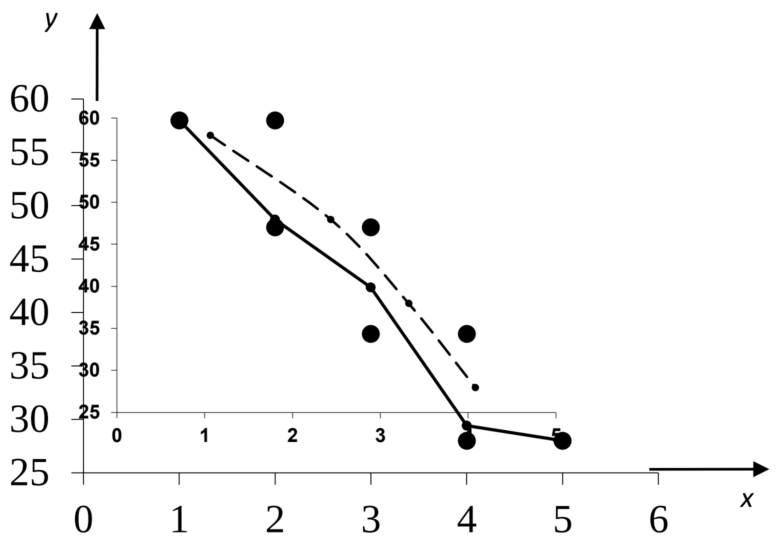

Рис. 4.8

Анализируем рис. 4.8. Видим, что с ростом значений x, значения y практически монотонно убывают. Следовательно, скорее всего, имеет место линейная обратная корреляционная зависимость.

4.4. Коэффициент линейной корреляции и его свойства

Окончательные выводы о форме корреляционной зависимости осуществляются с помощью коэффициента линейной корреляции.

Коэффициент регрессии *y* ***на*** *x* определяется как отношение ковариации к дисперсии признака *X*:

$$\rho_{y/x} = \frac{\mu}{\sigma_x^2}\,[15], \qquad (4.14)$$

при этом ***ковариация***

$$\mu = \overline{x \cdot y} - \bar{x} \cdot \bar{y}, \qquad (4.15)$$

где

$$\overline{x \cdot y} = \frac{\sum\limits_x \sum\limits_y x \cdot y \cdot n_{xy}}{n},$$

а $\bar{x}$ и $\bar{y}$ - выборочные средние признаков *X* и *Y* соответственно, а σ_x^2 - дисперсия признака *X*.

Аналогично определяется коэффициент регрессии *x* на *y*:

$$\rho_{x/y} = \frac{\mu}{\sigma_y^2}, \qquad (4.16)$$

причем σ_y^2 - дисперсия признака *Y*.

Ковариация в (4.14) и (4.16) может быть рассчитана и методом моментов[16]. Действительно, если C_x и h_x, C_y и h_y – соответственно ложный нуль и шаг варьирования признаков *X* и *Y*, то применяя соответствующие формулы п. 3.14.7, получаем:

$$\overline{x \cdot y} = \frac{\sum\limits_x \sum\limits_y x \cdot y \cdot n_{xy}}{n} = \frac{\sum\limits_x \sum\limits_y (\alpha_x \cdot h_x + C_x) \cdot (\alpha_y \cdot h_y + C_y) \cdot n_{xy}}{n} =$$

$$= \frac{\sum\limits_x \sum\limits_y (\alpha_x \cdot \alpha_y \cdot h_x \cdot h_y + \alpha_x \cdot h_x \cdot C_y + \alpha_y \cdot h_y \cdot C_x + C_x \cdot C_y) \cdot n_{xy}}{n} =$$

$$= \frac{\sum\limits_x \sum\limits_y (\alpha_x \cdot \alpha_y \cdot h_x \cdot h_y \cdot n_{xy} + \alpha_x \cdot h_x \cdot C_y \cdot n_{xy} + \alpha_y \cdot h_y \cdot C_x \cdot n_{xy} + C_x \cdot C_y \cdot n_{xy})}{n} =$$

$$= \frac{\sum\limits_x \sum\limits_y \alpha_x \cdot \alpha_y \cdot h_x \cdot h_y \cdot n_{xy}}{n} + \frac{\sum\limits_x \sum\limits_y \alpha_x \cdot h_x \cdot C_y \cdot n_{xy}}{n} + \frac{\sum\limits_x \sum\limits_y \alpha_y \cdot h_y \cdot C_x \cdot n_{xy}}{n} + \frac{\sum\limits_x \sum\limits_y C_x \cdot C_y \cdot n_{xy}}{n} =$$

$$= \frac{h_x \cdot h_y \cdot \sum\limits_x \sum\limits_y \alpha_x \cdot \alpha_y \cdot n_{xy}}{n} + \frac{h_x \cdot C_y \cdot \sum\limits_x \sum\limits_y \alpha_x \cdot n_{xy}}{n} + \frac{h_y \cdot C_x \cdot \sum\limits_x \sum\limits_y \alpha_y \cdot n_{xy}}{n} + \frac{C_x \cdot C_y \cdot \sum\limits_x \sum\limits_y n_{xy}}{n}.$$

---

[15] ρ – буква греческого алфавита, читается – «ро».

[16] См. пункт 2.7.

При этом: $\dfrac{h_x \cdot C_y \cdot \sum\limits_x \sum\limits_y \alpha_x \cdot n_{xy}}{n} = h_x \cdot C_y \cdot \dfrac{\sum\limits_x \sum\limits_y \alpha_x \cdot n_{xy}}{n} = h_x \cdot C_y \cdot \dfrac{\sum\limits_x \alpha_x \cdot \sum\limits_y n_{xy}}{n} =$

$= h_x \cdot C_y \cdot \dfrac{\sum\limits_x \alpha_x \cdot n_x}{n} = h_x \cdot C_y \cdot m_1^x$; $\dfrac{h_y \cdot C_x \cdot \sum\limits_x \sum\limits_y \alpha_y \cdot n_{xy}}{n} = \dfrac{h_y \cdot C_x \cdot \sum\limits_y \sum\limits_x \alpha_y \cdot n_{xy}}{n} =$

$= h_y \cdot C_x \cdot \dfrac{\sum\limits_y \sum\limits_x \alpha_y \cdot n_{xy}}{n} = h_y \cdot C_x \cdot \dfrac{\sum\limits_y \alpha_y \cdot \sum\limits_x n_{xy}}{n} = h_y \cdot C_x \cdot \dfrac{\sum\limits_y \alpha_y \cdot n_y}{n} = h_y \cdot C_x \cdot m_1^y$;

$\dfrac{C_x \cdot C_y \cdot \sum\limits_x \sum\limits_y n_{xy}}{n} = \dfrac{C_x \cdot C_y \cdot n}{n} = C_x \cdot C_y$.

Значит,

$$\overline{x \cdot y} = \frac{h_x \cdot h_y}{n} \cdot \sum_x \sum_y \alpha_x \cdot \alpha_y \cdot n_{xy} + h_x \cdot C_y \cdot m_1^x + h_y \cdot C_x \cdot m_1^y + C_x \cdot C_y .$$

Принимая во внимание, что

$$\bar{x} = m_1^x \cdot h_x + C_x,\ \bar{y} = m_1^y \cdot h_y + C_y,$$

$\bar{x} \cdot \bar{y} = \left(m_1^x \cdot h_x + C_x\right) \cdot \left(m_1^y \cdot h_y + C_y\right) = m_1^x \cdot h_x \cdot m_1^y \cdot h_y + h_x \cdot C_y \cdot m_1^x + h_y \cdot C_x \cdot m_1^y + C_x \cdot C_y$,

получаем:

$\mu = \overline{x \cdot y} - \bar{x} \cdot \bar{y} = \dfrac{h_x \cdot h_y}{n} \cdot \sum\limits_x \sum\limits_y \alpha_x \cdot \alpha_y \cdot n_{xy} + h_x \cdot C_y \cdot m_1^x + h_y \cdot C_x \cdot m_1^y + C_x \cdot C_y -$

$- m_1^x \cdot h_x \cdot m_1^y \cdot h_y - h_x \cdot C_y \cdot m_1^x - h_y \cdot C_x \cdot m_1^y - C_x \cdot C_y = \dfrac{h_x \cdot h_y}{n} \cdot \sum\limits_x \sum\limits_y \alpha_x \cdot \alpha_y \cdot n_{xy} -$

$- m_1^x \cdot h_x \cdot m_1^y \cdot h_y = \dfrac{h_x \cdot h_y}{n} \cdot \sum\limits_x \sum\limits_y \alpha_x \cdot \alpha_y \cdot n_{xy} - \left(m_1^x \cdot h_x\right) \cdot \left(m_1^y \cdot h_y\right) = \dfrac{h_x \cdot h_y}{n} \cdot \sum\limits_x \sum\limits_y \alpha_x \cdot \alpha_y \cdot n_{xy} -$

$- \left(\bar{x} - C_x\right) \cdot \left(\bar{y} - C_y\right)$,

$$\mu = \frac{h_x \cdot h_y}{n} \cdot \sum_x \sum_y \alpha_x \cdot \alpha_y \cdot n_{xy} - \left(\bar{x} - C_x\right) \cdot \left(\bar{y} - C_y\right). \qquad (4.17)$$

Коэффициентом линейной корреляции *r переменных x и y называется среднее геометрическое коэффициентов регрессии, имеющее их знак,*

$$r = \pm\sqrt{\rho_{y/x} \cdot \rho_{x/y}} .$$

Исходя из определений (4.14) и (4.16) коэффициентов регрессии, коэффициент линейной корреляции может быть записан в виде:

$$r = \frac{\mu}{\sigma_x \cdot \sigma_y} . \qquad (4.18)$$

С помощью (4.18) можно выразить коэффициенты регрессии через коэффициент линейной корреляции. Действительно, согласно (4.14) и (4.18) $\rho_{y/x} = \frac{\mu}{\sigma_x^2} = \frac{\mu}{\sigma_x \cdot \sigma_y} \cdot \frac{\sigma_y}{\sigma_x} = r \cdot \frac{\sigma_y}{\sigma_x}$,

$$\rho_{y/x} = r \cdot \frac{\sigma_y}{\sigma_x}. \qquad (4.19)$$

Аналогично доказывается, что

$$\rho_{x/y} = r \cdot \frac{\sigma_x}{\sigma_y}. \qquad (4.20)$$

Без доказательства перечислим основные свойства коэффициента линейной корреляции.

Свойство 4.1. *Коэффициент линейной корреляции по абсолютной величине не превосходит единицы,*

$$|r| \le 1.$$

Свойство 4.2. *Значение коэффициента линейной корреляции равно*

$$|r| = \pm 1$$

тогда и только тогда, когда имеет место линейная функциональная зависимость.

Свойство 4.3. *Если между признаками X и Y отсутствует корреляционная зависимость, то*

$$r = 0.$$

В зависимости от r имеем следующую интерпретацию связи (табл. 4.5).

Таблица 4.5

Значение r	Интерпретация связи
$r = -1$	Линейная функциональная обратная
$-1 < r < -0{,}5$	Линейная обратная
$-0{,}5 \le r < 0$	Нелинейная
$r = 0$	Отсутствует
$0 < r \le 0{,}5$	Нелинейная
$0{,}5 < r < 1$	Линейная прямая
$r = 1$	Линейная функциональная прямая

Пример 4.6. Рассчитать коэффициент линейной корреляции и сделать выводы о форме корреляционной зависимости между признаками X и Y из примера 4.1. Расчет ковариации произвести двумя способами: по определению (4.15) и пользуясь методом моментов (4.17).

В примере 4.5 мы предположили с высокой степенью вероятности наличие между признаками X и Y из примера 4.1 линейной обратной корреляционной зависимости. Это значит, (табл. 4.5), что при правильном расчете

$$-1 < r < -0{,}5.$$

Причем

$$r \neq -1,$$

так как имеющаяся зависимость очевидно не функциональная. Эмпирические линии регрессии близки к прямой (см. рис. 4.8). Окончательно получаем: при правильном расчете, значение коэффициента линейно корреляции должно быть близким к -1.

Находим средние значения и дисперсии (по формуле разностей) каждого из признаков.

X	n_x	$x \cdot n_x$	$x^2 \cdot n_x$
1	17	17	17
2	14	28	56
3	23	69	207
4	43	172	688
5	3	15	75
Σ	100	301	1043

Получаем:

$$\bar{x} = \frac{\sum_x x \cdot n_x}{\sum_x n_x} = \frac{301}{100} = 3{,}01;$$

$$\overline{x^2} = \frac{\sum_x x^2 \cdot n_x}{\sum_x n_x} = \frac{1043}{100} = 10{,}43;$$

$$\sigma_x^2 = \overline{x^2} - \left(\bar{x}\right)^2 = 10{,}43 - (3{,}01)^2 = 1{,}3699;$$

$$\sigma_x = \sqrt{\sigma_x^2} = \sqrt{1{,}3699} \approx 1{,}1704.$$

Теперь рассчитаем аналогичные числовые характеристики признака Y:

y	n_y	$y \cdot n_y$	$y^2 \cdot n_y$
28	40	1120	31360
38	19	722	27436
48	23	1104	52992
58	18	1044	60552
Σ	100	3990	172340

Отсюда:

$$\bar{y} = \frac{\sum_y y \cdot n_y}{\sum_y n_y} = \frac{3990}{100} = 39{,}9;$$

$$\overline{y^2} = \frac{\sum_y y^2 \cdot n_y}{\sum_y n_y} = \frac{172340}{100} = 1723{,}4;$$

$$\sigma_y^2 = \overline{y^2} - (\bar{y})^2 = 1723{,}4 - (39{,}9)^2 = 131{,}39;$$

$$\sigma_y = \sqrt{\sigma_y^2} = \sqrt{131{,}39} \approx 11{,}4625.$$

Найдем среднее значение произведения $\overline{x \cdot y}$. Согласно табл. 4.2

$$\overline{x \cdot y} = \frac{\sum_x \sum_y x \cdot y \cdot n_{xy}}{n} =$$

$$= \frac{4 \cdot 28 \cdot 37 + 5 \cdot 28 \cdot 3 + 3 \cdot 38 \cdot 13 + 4 \cdot 38 \cdot 6 + 2 \cdot 48 \cdot 13 + 3 \cdot 48 \cdot 10 + 1 \cdot 58 \cdot 17 + 2 \cdot 58 \cdot 1}{100} = \frac{10748}{100} =$$

$$= 107{,}48.$$

Отсюда, согласно (4.15)

$$\mu = \overline{x \cdot y} - \bar{x} \cdot \bar{y} = 107{,}48 - 3{,}01 \cdot 39{,}9 = -12{,}619.$$

Переходим к методу моментов. Анализ табл. 4.2 говорит, что

$$h_x = 1, C_x = 4, \; h_y = 10, \; C_y = 28.$$

Следуя соответствующим определениям, находим условные варианты α_x и α_y. Удобнее условные варианты поместить в таблицу, аналогичную 4.2

α_y \ α_x	-3	-2	-1	0	1	n_y
0				37	3	40
1			13	6		19
2		13	10			23
3	17	1				18
n_x	17	14	23	43	3	100

Теперь находим

$$\sum_x \sum_y \alpha_x \cdot \alpha_y \cdot n_{xy} =$$

$$= 0 \cdot 0 \cdot 37 + 1 \cdot 0 \cdot 3 + (-1) \cdot 1 \cdot 13 + 0 \cdot 1 \cdot 6 + (-2) \cdot 2 \cdot 13 + (-1) \cdot 2 \cdot 10 + (-3) \cdot 3 \cdot 17 + (-2) \cdot 3 \cdot 1 = -244$$

.

Значит, по формуле (4.17)

$$\mu = \frac{h_x \cdot h_y}{n} \cdot \sum_x \sum_y \alpha_x \cdot \alpha_y \cdot n_{xy} - \left(\bar{x} - C_x\right) \cdot \left(\bar{y} - C_y\right) = \frac{1 \cdot 10}{100} \cdot (-244) - (3{,}01 - 4) \cdot (39{,}9 - 28) = -12{,}619$$

Используя (4.18), получаем:

$$r = \frac{-12{,}619}{1{,}1704 \cdot 11{,}4625} \approx -0{,}9406. \qquad (4.21)$$

Сравнивая найденную величину коэффициента линейной корреляции со значениями, помещенными в табл. 4.5, окончательно убеждаемся, что имеет место линейная обратная корреляционная зависимость.

4.5. Корреляционное отношение и его свойства

Следующим важным показателем корреляционной зависимости является корреляционное отношение η[17]. Оно характеризует тесноту связи между признаками *X* и *Y*.

Корреляционным отношением *y и x называется отношение межгруппового среднеквадратического отклонения признака Y к его общему среднеквадратическому отклонению,*

$$\eta_{y/x} = \frac{\delta_{y/x}}{\sigma_y}.$$

При этом ***межгрупповое среднеквадратическое отклонение*** *равно корню из межгрупповой дисперсии признака Y,*

$$\delta_{y/x} = \sqrt{\delta^2_{y/x}}\ .$$

Межгрупповая дисперсия *признака Y равна среднему квадрату отклонения групповой средней* $\bar{y}_x$ *от выборочной средней* $\bar{y}$,

$$\delta^2_{y/x} = \frac{\sum_x \left(\bar{y}_x - \bar{y}\right)^2 \cdot n_x}{n}.$$

Общее среднеквадратическое отклонение считается также, как и в пункте 2.6.

[17] Буква греческого алфавита, читается – «эта».

Аналогично определяется ***корреляционное отношение*** *x* и *y*:

$$\eta_{x/y} = \frac{\delta_{x/y}}{\sigma_x},$$

$$\delta_{x/y} = \sqrt{\delta^2_{x/y}}\ ,\ \delta^2_{x/y} = \frac{\sum\limits_{y}\left(\bar{x}_y - \bar{x}\right)^2 \cdot n_y}{n}.$$

Укажем свойства корреляционного отношения.

Свойство .4.4. *Корреляционное отношение не меньше нуля, но не больше единицы,*

$$0 \le \eta_{y/x} \le 1.$$

Свойство 4.5. *Корреляционное отношение*

$$\eta_{y/x} = 0,$$

тогда и только тогда, когда отсутствует связь между признаками X и Y.

Свойство 4.6. *Корреляционное отношение*

$$\eta_{y/x} = 1$$

в том и только в том случае, когда признаки X и Y связаны функциональной зависимостью.

Свойство 4.7. *Корреляционное отношение не меньше модуля коэффициента линейной корреляции,*

$$\eta_{y/x} \ge |\, r\,|.$$

Свойство 4.8. *Выполнение равенства*

$$\eta_{y/x} = |\, r\,|$$

является необходимым и достаточным условием того, чтобы корреляционная зависимость была точно линейной.

Заметим, что сформулированные условия 16.4 – 16.8 справедливы и в отношении корреляционного отношения $\eta_{x/y}$.

Характер связи определяется так (табл. 4.6):

Таблица 4.6

Значение $\eta_{y/x}$	Характер связи
$\eta_{y/x} = 0$	Отсутствует
$0 < \eta_{y/x} \le 0{,}3$	Практически отсутствует
$0{,}3 < \eta_{y/x} \le 0{,}5$	Слабая
$0{,}5 < \eta_{y/x} \le 0{,}7$	Умеренная
$0{,}7 < \eta_{y/x} < 1$	Сильная
$\eta_{y/x} = 1$	Функциональная

Пример 4.7. Рассчитав корреляционные отношения $\eta_{y/x}$ и $\eta_{x/y}$, сделать вывод о тесноте корреляционной связи между признаками *X* и *Y* из примера 4.1.

Исходя из соответствующих свойств корреляционного отношения можно предположить, что правильно рассчитанные корреляционные отношения отвечают неравенствам:

$$0{,}9406 < \eta_{y/x} < 1, \; 0{,}9406 < \eta_{x/y} < 1.$$

Средние значения и среднеквадратические отклонения каждого из признаков были получены в примере 4.6, а величины групповых средних – в примере 4.4 (табл. 4.3 и 4.4)

Расчеты осуществим в таблице:

x	n_x	$\overline{y}_x$	$\overline{y}_x - \overline{y}$	$\left(\overline{y}_x - \overline{y}\right)^2 \cdot n_x$
1	17	58	18,1	5569,37
2	14	48,71	8,81	1086,6254
3	23	42,35	2,45	138,0575
4	43	29,4	-10,5	4740,75
5	3	28	-11,9	424,83
Σ	100	-	-	11959,6329

Имеем:

$$\delta^2_{y/x} = \frac{11959{,}6329}{100} = 119{,}596329, \; \delta_{y/x} = \sqrt{119{,}596329} \approx 10{,}9360,$$

$$\eta_{y/x} = \frac{10{,}9360}{11{,}4625} \approx 0{,}9541.$$

Аналогичные расчеты делаем для второго признака:

y	n_y	$\overline{x}_y$	$\overline{x}_y - \overline{x}$	$\left(\overline{x}_y - \overline{x}\right)^2 \cdot n_y$
28	40	4,08	1,07	45,796
38	19	3,32	0,31	1,8259
48	23	2,43	-0,58	7,7372
58	18	1,06	-1,95	68,445
Σ	100	-	-	123,8041

Следовательно:

$$\delta^2_{x/y} = \frac{123{,}8041}{100} = 1{,}238041, \; \delta_{x/y} = \sqrt{1{,}238041} \approx 1{,}1127,$$

$$\eta_{x/y} = \frac{1{,}1127}{1{,}1704} \approx 0{,}9507.$$

С помощью табл. 4.6 заключаем, что связь является сильной.

4.6. Оценка качества эмпирических данных

Прежде чем приступить к построению уравнения регрессии, необходимо быть уверенным в том, что имеющиеся эмпирические данные статистически значимы. Другими словами, нужно убедиться, что между значениями признаков имеет место статистически существенная зависимость. В противном случае, если гипотеза не подтверждается, приходится признать, что качество эмпирических данных недостаточно для построения уравнения регрессии, позволяющего осуществить точный прогноз результативного показателя. В этом случае необходимо рассмотреть другую выборочную совокупность.

Проверка гипотезы о наличии между признаками X и Y статистически существенной зависимости производится с помощью t – критерия Стьюдента[18].

Правило проверки гипотезы. *Если наблюдаемое значение критерия больше критического,*

$$t_{набл} > t_{крит},$$

то это с доверительной вероятностью γ (*уровнем значимости* $\alpha=1-\gamma$) *говорит о наличии между признаками* X *и* Y *статистически существенной зависимости. Наблюдаемое значение критерия Стьюдента равно*

$$t_{набл} = \sqrt{\frac{r^2}{1-r^2}\cdot(n-2)}. \tag{4.22}$$

Критическое значение рассчитывается по таблице[19] *в зависимости от уровня значимости* α *и числа степеней свободы* ν,

$$t_{крит} = t(\alpha;\nu),\ \ \alpha = 1-\gamma,\ \ \nu = n-2. \tag{4.23}$$

Пример 4.8. С вероятностью 0,95 проверить гипотезу о наличии статистически существенной зависимости между признаками X и Y из примера 4.1.

Объем выборочной совокупности $n = 100$. Коэффициент линейной корреляции был рассчитан в примере 4.6: он равен (4.21). Согласно формуле (4.22) наблюдаемое значение критерия Стьюдента равно

$$t_{набл} = \sqrt{\frac{(-0{,}9406)^2}{1-(-0{,}9406)^2}\cdot(100-2)} \approx 27{,}4256.$$

---

[18] Госсет, Уильям Сили (1867 - 1937) – известный английский ученый – статистик, работавший под псевдонимом Стьюдент.

[19] Некоторые критические значения критерия Стьюдента приведены в табл. 4 прил.

Для вероятности $\gamma = 0{,}95$ уровень значимости $\alpha = 1 - 0{,}95 = 0{,}05$. Число степеней свободы для $n = 100$ равно $\nu = 100 - 2 = 98$. Согласно табл. 4 прил. критическое значение (4.23) критерия Стьюдента равно

$$t_{крит} = t(0{,}05;98) = 1{,}99.$$

Поскольку

$$27{,}4256 > 1{,}99,\ t_{набл} > t_{крит},$$

правило проверки гипотезы выполняется.

Это значит, что с вероятностью 0,95 можно говорить о наличии между признаками статистически существенной зависимости. Другими словами, эмпирические данные примера 16.1 являются статистически значимыми, их качество удовлетворительно, регрессионная модель может быть построена.

4.7. Метод наименьших квадратов

Среди большого числа методов, позволяющих найти параметры уравнения регрессии, самым простым является метод наименьших квадратов (МНК), Сформулированный и впервые примененный Лежандром[20] и Гауссом[21]. Объясним суть МНК.

Пусть $y = y(x)$ - уравнение регрессии y на x. Изобразим ее график на рис. 4.9.

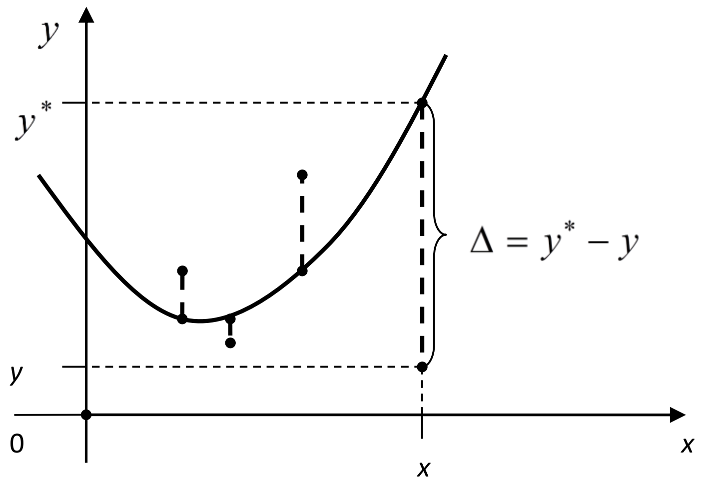

Рис. 4.9

[20] Лежандр Адриен Мари (1752 - 1833) – французский математик.

[21] Гаусс Карл Фридрих (1777 - 1855) – великий немецкий математик, создатель современной математической науки, внесший также фундаментальный вклад в астрономию и геодезию.

Поскольку связь корреляционная, то эмпирическое значение признака y, соответствующее данному значению x, может отличаться от теоретического y_x^*, рассчитанного по уравнению регрессии,

$$y_x^* = y(x). \tag{4.24}$$

Разность между теоретическим и эмпирическим значениями признака Y, соответствующих данному значению x признака X, называется ***отклонением*** *или* ***невязкой****,*

$$\Delta_x = y_x^* - y. \tag{4.25}$$

В дальнейшем будем рассматривать квадрат невязок,

$$\Delta_x^2 = \left(y_x^* - y\right)^2.$$

Ясно, что линия регрессии тем точнее будет описывать корреляционную зависимость, чем меньше будет сумма квадратов невязок, найденная по каждому значению x,

$$\sum_x \Delta_x^2 \to \min.$$

Уравнение регрессии (4.24) однозначно определяется своими параметрами a, b, c,…Значит, сумму $\sum_x \Delta_x^2$ можно воспринимать как функцию от переменных a, b, c,…,

$$f = f(a,b,c,...) = \sum_x \Delta_x^2. \tag{4.26}$$

Итак, поиск параметров уравнения регрессии свелся к известной задаче, решаемой методами математического анализа: найти минимальное значение функции (4.26). Экстремум функции (4.26) определяется путем нахождения частных производных функции по каждой из переменной и приравнивания их к нулю. Тем самым, поиск параметров уравнения регрессии свелся к решению системы:

$$\begin{cases} f_a' = 0 \\ f_b' = 0 \\ f_c' = 0 \\ \vdots \end{cases}. \tag{4.27}$$

Система (4.27) носит название ***системы нормальных уравнений***.

Решив систему нормальных уравнений, мы получим параметры уравнения регрессии y на x. Аналогичную систему можно получить и для поиска параметров уравнения регрессии x на y.

4.8. Уравнение регрессии

Пользуясь методом наименьших квадратов, найдем параметры a и b уравнения (4.2) прямой линии регрессии y на x.

Теоретическое значение признака Y в этом случае рассчитывается по формуле:

$$y_x^* = a \cdot x + b.$$

Отсюда, квадрат невязки

$$\Delta_x^2 = \left(y_x^* - y\right)^2 = (a \cdot x + b - y)^2.$$

Значит, функция (4.26) для данного случая имеет вид:

$$f = f(a,b) = \sum (a \cdot x + b - y)^2. \qquad (4.28)$$

В общем виде, система нормальных уравнений следующая:

$$\begin{cases} f_a' = 0 \\ f_b' = 0 \end{cases}. \qquad (4.29)$$

Находим частные производные функции (4.28) по каждой из переменных, делаем соответствующие преобразования, используем (4.12), а затем приравниваем их к 0.

Первое уравнение системы (4.29): $f_a' = \left(\sum (a \cdot x + b - y)^2\right)_a' =$

$$= \sum \left((a \cdot x + b - y)^2\right)_a' = \sum 2 \cdot (a \cdot x + b - y) \cdot (a \cdot x + b - y)_a' = \sum 2 \cdot (a \cdot x + b - y) \cdot x =$$

$$= 2 \cdot \sum x \cdot (a \cdot x + b - y) = 2 \cdot \sum \left(a \cdot x^2 + b \cdot x - x \cdot y\right) = 2 \cdot \left(\sum a \cdot x^2 + \sum b \cdot x - \sum x \cdot y\right) =$$

$$= 2 \cdot \left(\sum_x a \cdot x^2 \cdot n_x + \sum_x b \cdot x \cdot n_x - \sum_x \sum_y x \cdot y \cdot n_{xy}\right) =$$

$$= 2 \cdot \left(a \cdot \sum_x x^2 \cdot n_x + b \cdot \sum_x x \cdot n_x - \sum_x \sum_y x \cdot y \cdot n_{xy}\right) =$$

$$= 2 \cdot \left(\left(\sum_x x^2 \cdot n_x\right) \cdot a + \left(\sum_x x \cdot n_x\right) \cdot b - \sum_x \sum_y x \cdot y \cdot n_{xy}\right) =$$

$$= 2 \cdot \left(\left(\sum_x x^2 \cdot n_x\right) \cdot a + \left(\sum_x x \cdot n_x\right) \cdot b - \sum_x \left(x \cdot \sum_y y \cdot n_{xy}\right)\right) =$$

$$= 2 \cdot \left(\left(\sum_x x^2 \cdot n_x\right) \cdot a + \left(\sum_x x \cdot n_x\right) \cdot b - \sum_x x \cdot \frac{\sum_y y \cdot n_{xy}}{n_x} \cdot n_x\right) =$$

$$= 2 \cdot \left(\left(\sum_x x^2 \cdot n_x\right) \cdot a + \left(\sum_x x \cdot n_x\right) \cdot b - \sum_x x \cdot \overline{y}_x \cdot n_x\right) = 0.$$

Отсюда

$$\left(\sum_x x^2 \cdot n_x\right)\cdot a + \left(\sum_x x \cdot n_x\right)\cdot b - \sum_x x \cdot \overline{y}_x \cdot n_x = 0$$

или

$$\left(\sum_x x^2 \cdot n_x\right)\cdot a + \left(\sum_x x \cdot n_x\right)\cdot b = \sum_x x \cdot \overline{y}_x \cdot n_x\,. \qquad (4.30)$$

Аналогично определяем второе уравнение системы (4.29): $f_b' =$

$$= \left(\sum (a\cdot x + b - y)^2\right)_b' = \sum\left((a\cdot x + b - y)^2\right)_b' = \sum 2\cdot(a\cdot x + b - y)\cdot(a\cdot x + b - y)_b' =$$

$$= \sum 2\cdot(a\cdot x + b - y)\cdot 1 = 2\cdot\sum(a\cdot x + b - y) = 2\cdot\left(\sum a\cdot x + \sum b - \sum y\right) =$$

$$= 2\cdot\left(\sum_x a\cdot x\cdot n_x + \sum_x b\cdot n_x - \sum_x\sum_y y\cdot n_{xy}\right) = 2\cdot\left(a\cdot\sum_x x\cdot n_x + b\cdot\sum_x n_x - \sum_x\sum_y y\cdot n_{xy}\right) =$$

$$= 2\cdot\left(\left(\sum_x x\cdot n_x\right)\cdot a + \left(\sum_x n_x\right)\cdot b - \sum_x\sum_y y\cdot n_{xy}\right) =$$

$$= 2\cdot\left(\left(\sum_x x\cdot n_x\right)\cdot a + \left(\sum_x n_x\right)\cdot b - \sum_x \frac{\sum_y y\cdot n_{xy}}{n_x}\cdot n_x\right) = 2\cdot\left(\left(\sum_x x\cdot n_x\right)\cdot a + n\cdot b - \sum_x \overline{y}_x\cdot n_x\right) = 0\,,$$

$$\left(\sum_x x\cdot n_x\right)\cdot a + n\cdot b - \sum_x \overline{y}_x\cdot n_x = 0\,,$$

$$\left(\sum_x x\cdot n_x\right)\cdot a + n\cdot b = \sum_x \overline{y}_x\cdot n_x\,. \qquad (4.31)$$

Собирая уравнения (4.30) и (4.31) в (4.29), получаем систему:

$$\begin{cases} \left(\sum_x x^2\cdot n_x\right)\cdot a & + & \left(\sum_x x\cdot n_x\right)\cdot b & = & \sum_x x\cdot\overline{y}_x\cdot n_x \\ \left(\sum_x x\cdot n_x\right)\cdot a & + & n\cdot b & = & \sum_x \overline{y}_x\cdot n_x \end{cases}. \qquad (4.32)$$

Действуя аналогично, получаем систему нормальных уравнений для нахождения параметров a' и b' уравнения (4.6) прямой линии регрессии x на y[22]:

$$\begin{cases} \left(\sum_y y^2\cdot n_y\right)\cdot a' & + & \left(\sum_y y\cdot n_y\right)\cdot b' & = & \sum_y y\cdot\overline{x}_y\cdot n_y \\ \left(\sum_y y\cdot n_y\right)\cdot a' & + & n\cdot b' & = & \sum_y \overline{x}_y\cdot n_y \end{cases}. \qquad (4.33)$$

Стоит заметить, что в уравнениях (4.32) и (4.33)

$$\sum_x x\cdot n_x = \sum_y \overline{x}_y\cdot n_y\,. \qquad (4.34)$$

---

[22] Обязательно проделайте эти вычисления самостоятельно.

Действительно

$$\sum_y \overline{x}_y \cdot n_y = \sum_y \frac{\sum_x x \cdot n_{xy}}{n_y} \cdot n_y = \sum_y \sum_x x \cdot n_{xy} = \sum_x \sum_y x \cdot n_{xy} = \sum_x x \cdot \sum_y n_{xy} =$$

$$= \sum_x x \cdot n_x .$$

Аналогично, доказывается, что

$$\sum_y y \cdot n_y = \sum_x \overline{y}_x \cdot n_x . \qquad (4.35)$$

Наконец, имеет место равенство:

$$\sum_x x \cdot \overline{y}_x \cdot n_x = \sum_y y \cdot \overline{x}_y \cdot n_y = \sum_x \sum_y x \cdot y \cdot n_{xy} . \qquad (4.36)$$

Делая соответствующие преобразования, получаем: $\sum_x x \cdot \overline{y}_x \cdot n_x =$

$$\sum_x x \cdot \frac{\sum_y y \cdot n_{xy}}{n_x} \cdot n_x = \sum_x x \cdot \sum_y y \cdot n_{xy} = \sum_y \sum_x x \cdot y \cdot n_{xy} = \sum_y y \cdot \sum_x x \cdot n_{xy} =$$

$$= \sum_y y \cdot \frac{\sum_x x \cdot n_{xy}}{n_y} \cdot n_y = \sum_y y \cdot \overline{x}_y \cdot n_y .$$

Пример 4.8. Найти параметры уравнения прямой линии регрессии y на x и x на y для корреляционной зависимости из примера 4.1. Результаты вычислений представить графически.

В примерах 4.6 и 4.7 мы установили, что связь между признаками X и Y является тесной линейной. Значит, уравнения регрессии имеют общий вид (4.2) и (4.6). При этом, поскольку связь обратная, то угловые коэффициенты должны выражаться отрицательными числами,

$$a < 0,\ a' < 0 .$$

Частично, расчеты были осуществлены в примере 16.6, групповые средние были найдены в примере 4.4 (табл. 4.3 и 4.4).

Расчеты будем производить в таблице:

x	n_x	$x \cdot n_x$	$x^2 \cdot n_x$	$\overline{y}_x$	$\overline{y}_x \cdot n_x$	$x \cdot \overline{y}_x \cdot n_x$
1	17	17	17	58	986	986
2	14	28	56	48,71	681,94	1363,88
3	23	69	207	42,35	974,05	2922,15
4	43	172	688	29,4	1264,2	5056,8
5	3	15	75	28	84	420
Σ	100	301	1043	-	3990,19	10748,83

Устанавливаем вид системы (16.32) нормальных уравнений:

$$\begin{cases} 1043\cdot a + 301\cdot b = 10748{,}83 \\ 301\cdot a + 100\cdot b = 3990{,}19 \end{cases}.$$

Решаем систему, скажем, методом Крамера. Получаем:

$$a \approx -9{,}2097,\ b \approx 67{,}6232.$$

Итак, искомое уравнение прямой линии регрессии y на x имеет вид:

$$y = -9{,}2097\cdot x + 67{,}6232. \qquad (4.37)$$

Для того, чтобы найти теоретические значения y_x^*, необходимо подставить имеющиеся значения x в (4.37). Результаты вычислений представим в таблице (табл. 4.7):

Таблица 4.7

x	1	2	3	4	5
y_x^*	58,4135	49,2038	39,9941	30,7844	21,5747

Теперь найдем параметры уравнения (4.6).

Определяем коэффициенты системы нормальных уравнений:

y	n_y	$y\cdot n_y$	$y^2\cdot n_y$	$\bar{x}_y$	$\bar{x}_y\cdot n_y$	$y\cdot\bar{x}_y\cdot n_y$
28	40	1120	31360	4,08	163,2	4569,6
38	19	722	27436	3,32	63,08	2397,04
48	23	1104	52992	2,43	55,89	2682,72
58	18	1044	60552	1,06	19,08	1106,64
Σ	100	3990	172340	-	301,25	10756

Составляем систему нормальных уравнений вида (16.33):

$$\begin{cases} 172340\cdot a' + 3990\cdot b' = 10756 \\ 3990\cdot a' + 100\cdot b' = 301{,}25 \end{cases}.$$

Решаем ее:

$$a' \approx -0{,}0962,\ b' \approx 6{,}8506.$$

Значит, уравнение регрессии x на y имеет вид:

$$x = -0{,}0962\cdot y + 6{,}8506. \qquad (4.38)$$

Подставляем значения y в (4.38), получаем теоретические значения x_y^* (табл. 4.8).

Таблица 4.8

y	28	38	48	58
x_y^*	4,157	3,195	2,233	1,271

На одном чертеже (рис. 4.10) строим: эмпирические данные; по точкам из табл. 4.7 строим прямую регрессии y на x (сплошная линия), а по точкам из табл. 4.8 – прямую регрессии x на y (пунктирная линия).

Параметры уравнения прямой линии регрессии можно найти и иначе, не только решая системы (4.32) и (4.33). Разделим левую и правую части каждого из уравнения системы (4.32) на объем выборки n:

$$\begin{cases} \dfrac{\sum_x x^2 \cdot n_x}{n} \cdot a + \dfrac{\sum_x x \cdot n_x}{n} \cdot b = \dfrac{\sum_x x \cdot \overline{y}_x \cdot n_x}{n} \\ \dfrac{\sum_x x \cdot n_x}{n} \cdot a + \dfrac{n}{n} \cdot b = \dfrac{\sum_x \overline{y}_x \cdot n_x}{n} \end{cases}. \quad (4.39)$$

С учетом определений арифметической средней, среднего квадрата, а также формул (4.35) и (4.36):

$$\frac{\sum_x x^2 \cdot n_x}{n} = \frac{\sum_x x^2 \cdot n_x}{\sum_x n_x} = \overline{x^2}; \quad \frac{\sum_x x \cdot n_x}{n} = \frac{\sum_x x \cdot n_x}{\sum_x n_x} = \overline{x}; \quad \frac{n}{n} = 1;$$

$$\frac{\sum_x x \cdot \overline{y}_x \cdot n_x}{n} = \frac{\sum_x \sum_y x \cdot y \cdot n_{xy}}{n} = \overline{x \cdot y}; \quad \frac{\sum_x \overline{y}_x \cdot n_x}{n} = \frac{\sum_y y \cdot n_y}{\sum_y n_y} = \overline{y}.$$

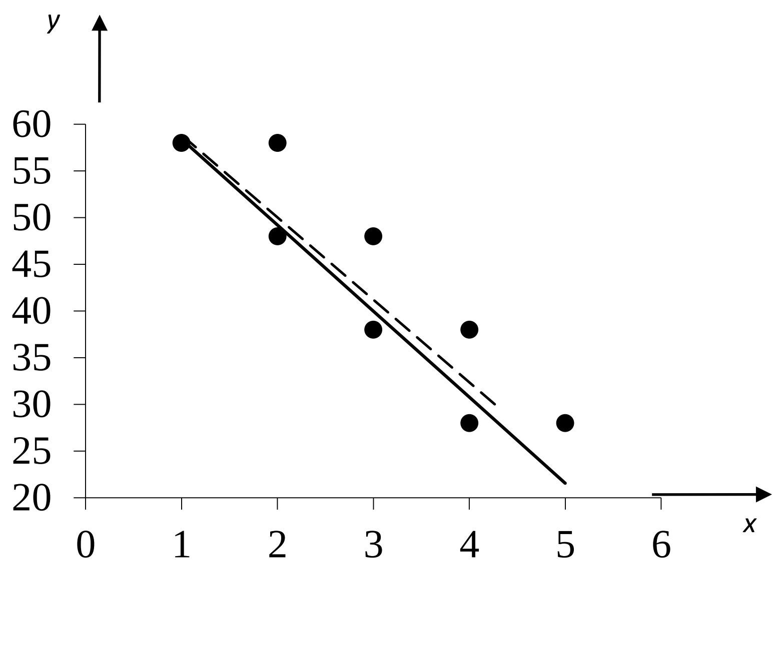

Рис. 4.10

С учетом этих преобразований, система (4.39) перепишется в виде:

$$\begin{cases} \overline{x^2} \cdot a & + & \overline{x} \cdot b & = & \overline{x \cdot y} \\ \overline{x} \cdot a & + & 1 \cdot b & = & \overline{y} \end{cases}.$$

Решим ее методом Крамера. Определитель матрицы системы согласно формуле (2.8) равен:

$$\Delta = \begin{vmatrix} \overline{x^2} & \overline{x} \\ \overline{x} & 1 \end{vmatrix} = \overline{x^2} \cdot 1 - \overline{x} \cdot \overline{x} = \overline{x^2} - \left(\overline{x}\right)^2 = \sigma_x^2 .$$

В силу (4.15)

$$\Delta_1 = \begin{vmatrix} \overline{x \cdot y} & \overline{x} \\ \overline{y} & 1 \end{vmatrix} = \overline{x \cdot y} \cdot 1 - \overline{x} \cdot \overline{y} = \mu .$$

Делая несложные преобразования, находим

$$\Delta_2 = \begin{vmatrix} \overline{x^2} & \overline{x \cdot y} \\ \overline{x} & \overline{y} \end{vmatrix} = \overline{x^2} \cdot \overline{y} - \overline{x} \cdot \overline{x \cdot y} = \overline{x^2} \cdot \overline{y} - \overline{x} \cdot \overline{x \cdot y} \pm \left(\overline{x}\right)^2 \cdot \overline{y} =$$

$$= \left(\overline{x^2} \cdot \overline{y} - \left(\overline{x}\right)^2 \cdot \overline{y}\right) - \left(\overline{x} \cdot \overline{x \cdot y} - \left(\overline{x}\right)^2 \cdot \overline{y}\right) = \left(\overline{x^2} - \left(\overline{x}\right)^2\right) \cdot \overline{y} - \left(\overline{x \cdot y} - \overline{x} \cdot \overline{y}\right) \cdot \overline{x} =$$

$$= \sigma_x^2 \cdot \overline{y} - \mu \cdot \overline{x} .$$

Используя (4.14), получаем:

$$a = \frac{\Delta_1}{\Delta} = \frac{\mu}{\sigma_x^2} = \rho_{y/x} ;$$

$$b = \frac{\Delta_2}{\Delta} = \frac{\sigma_x^2 \cdot \overline{y} - \mu \cdot \overline{x}}{\sigma_x^2} = \frac{\sigma_x^2 \cdot \overline{y}}{\sigma_x^2} - \frac{\mu \cdot \overline{x}}{\sigma_x^2} = \overline{y} - \frac{\mu}{\sigma_x^2} \cdot \overline{x} = \overline{y} - \rho_{y/x} \cdot \overline{x} .$$

Теперь подставим найденные значения параметров в уравнение (4.2):

$$y = \rho_{y/x} \cdot x + \left(\overline{y} - \rho_{y/x} \cdot \overline{x}\right),$$

$$y - \overline{y} = \rho_{y/x} \cdot \left(x - \overline{x}\right). \tag{4.40}$$

Найденное уравнение (4.40) является альтернативным способом определения уравнения прямой линии регрессии *y* на *x*. Аналогично можно получить вид уравнения регрессии *x* на *y*[23]:

$$x - \overline{x} = \rho_{x/y} \cdot \left(y - \overline{y}\right). \tag{4.41}$$

[23] Обязательно проделайте вывод уравнения самостоятельно.

Пример 4.9. По данным примера 4.1 получить вид уравнения регрессии в виде (4.40) и (4.41). Сравнить полученные результаты с выводами примера 4.8.

Числовые характеристики признаков X и Y были определены в примере 4.6.

Пользуясь определением (4.19) и (4.20) определим коэффициенты регрессии:

$$\rho_{y/x} = r \cdot \frac{\sigma_y}{\sigma_x} = -0{,}9406 \cdot \frac{11{,}4625}{1{,}1704} \approx -9{,}2119.$$

$$\rho_{x/y} = r \cdot \frac{\sigma_x}{\sigma_y} = -0{,}9406 \cdot \frac{1{,}1704}{11{,}4625} \approx -0{,}0960.$$

Теперь устанавливаем вид уравнения регрессии:

$$y - 39{,}9 = -9{,}2119 \cdot (x - 3{,}01) \text{ или } y = -9{,}2119 \cdot x + 67{,}6279;$$

$$x - 3{,}01 = -0{,}0960 \cdot (x - 39{,}9) \text{ или } y = -0{,}0960 \cdot x + 6{,}8421.$$

Заметим, что уравнения регрессии почти не отличаются от (4.37) и (4.38).

Мы осуществили вывод уравнения линейной регрессии. Уравнение для любого другого типа регрессии получаются точно также, путем решения системы нормальных уравнений. В табл. 4.9 представлены системы нормальных уравнений для нахождения параметров уравнений регрессии для линейного, параболического и показательного случаев. Попробуйте с помощью метода наименьших квадратов вывести эти системы самостоятельно.

Таблица 4.9

y на x	x на y	
$y = a \cdot x + b$	$x = a' \cdot y + b'$	**Линейная**
$\begin{cases} \left(\sum_x x^2 \cdot n_x\right) \cdot a + \left(\sum_x x \cdot n_x\right) \cdot b = \sum_x x \cdot \bar{y}_x \cdot n_x \\ \left(\sum_x x \cdot n_x\right) \cdot a + n \cdot b = \sum_x \bar{y}_x \cdot n_x \end{cases}$	$\begin{cases} \left(\sum_y y^2 \cdot n_y\right) \cdot a' + \left(\sum_y y \cdot n_y\right) \cdot b' = \sum_y y \cdot \bar{x}_y \cdot n_y \\ \left(\sum_y y \cdot n_y\right) \cdot a' + n \cdot b' = \sum_y \bar{x}_y \cdot n_y \end{cases}$	
$y - \bar{y} = \rho_{y/x} \cdot (x - \bar{x})$, $\rho_{y/x} = r \cdot \dfrac{\sigma_y}{\sigma_x}$	$x - \bar{x} = \rho_{x/y} \cdot (y - \bar{y})$, $\rho_{x/y} = r \cdot \dfrac{\sigma_x}{\sigma_y}$	
$y = a \cdot x^2 + b \cdot x + c$	$x = a' \cdot y^2 + b' \cdot y + c'$	**Параболическая**
$\begin{cases} \left(\sum_x x^4 \cdot n_x\right) \cdot a + \left(\sum_x x^3 \cdot n_x\right) \cdot b + \left(\sum_x x^2 \cdot n_x\right) \cdot c = \sum_x x^2 \cdot \bar{y}_x \cdot n_x \\ \left(\sum_x x^3 \cdot n_x\right) \cdot a + \left(\sum_x x^2 \cdot n_x\right) \cdot b + \left(\sum_x x \cdot n_x\right) \cdot c = \sum_x x \cdot \bar{y}_x \cdot n_x \\ \left(\sum_x x^2 \cdot n_x\right) \cdot a + \left(\sum_x x \cdot n_x\right) \cdot b + n \cdot c = \sum_x \bar{y}_x \cdot n_x \end{cases}$	$\begin{cases} \left(\sum_y y^4 \cdot n_y\right) \cdot a' + \left(\sum_y y^3 \cdot n_y\right) \cdot b' + \left(\sum_y y^2 \cdot n_y\right) \cdot c' = \sum_y y^2 \cdot \bar{x}_y \cdot n_y \\ \left(\sum_y y^3 \cdot n_y\right) \cdot a' + \left(\sum_y y^2 \cdot n_y\right) \cdot b' + \left(\sum_y y \cdot n_y\right) \cdot c' = \sum_y y \cdot \bar{x}_y \cdot n_y \\ \left(\sum_y y^2 \cdot n_y\right) \cdot a' + \left(\sum_y y \cdot n_y\right) \cdot b' + n \cdot c' = \sum_y \bar{x}_y \cdot n_y \end{cases}$	
$y = b \cdot (a)^x$	$x = b' \cdot (a')^y$	**Показательная**
$\begin{cases} \left(\sum_x x^2 \cdot n_x\right) \cdot \ln a + \left(\sum_x x \cdot n_x\right) \cdot \ln b = \sum_x x \cdot (\ln \bar{y}_x) \cdot n_x \\ \left(\sum_x x \cdot n_x\right) \cdot \ln a + n \cdot \ln b = \sum_x \ln \bar{y}_x \cdot n_x \end{cases}$	$\begin{cases} \left(\sum_y y^2 \cdot n_y\right) \cdot \ln a' + \left(\sum_y y \cdot n_y\right) \cdot \ln b' = \sum_y y \cdot (\ln \bar{x}_y) \cdot n_y \\ \left(\sum_y y \cdot n_y\right) \cdot \ln a' + n \cdot \ln b' = \sum_y \ln \bar{x}_y \cdot n_y \end{cases}$	

4.9. Проверка точности регрессионной модели

Рассмотрим статистические показатели точности и качества регрессионной модели. С их помощью можно также отобрать и наиболее точную модель, если корреляционная зависимость не является линейной и стоит проблема выбора из нескольких вариантов.

Первым показателем точности уравнения регрессии y на x является ***средняя ошибка аппроксимации***, равная

$$\bar{\varepsilon}_{y/x} = \frac{1}{n} \cdot \sum_{x} \left| \frac{y - y_x^*}{y} \right| \cdot 100\% \text{[24]}.$$

Напомним, что y и y_x^* - соответственно эмпирическое и теоретическое (рассчитанное по уравнению регрессии) значения результативного фактора, соответствующие данному значению факторного признака. Средняя ошибка аппроксимации определяет степень удаленности эмпирических значений от теоретических. На практике, если $\bar{\varepsilon}_{y/x}$ для уравнения регрессии превышает 15%, то такая модель отвергается.

Пример 4.10. Оценить точность линейной модели (4.37).

Все теоретические значения y_x^* были найдены в примере 4.8 и помещены в табл. 4.7. Необходимо учесть, что каждому значению y может соответствовать несколько значений x. Игнорирование этого факта приводит к ошибкам. Все необходимые вычисления производим в таблице.

y	x	y_x^*	$y - y_x^*$	$\left\lvert \frac{y - y_x^*}{y} \right\rvert$
28	4	30,7844	-2,7844	0,0994
	5	21,5743	6,4257	0,2295
38	3	39,9941	-1,9941	0,0525
	4	30,7844	7,2156	0,1899
48	2	49,2038	-1,2038	0,0251
	3	39,9941	8,0059	0,1668
58	1	58,4135	-0,4135	0,0071
	2	49,2038	8,7962	0,1517
Σ	-	-	-	0,9220

[24] ε – буква греческого алфавита, читается – «эпсилон».

Получаем:

$$\bar{\varepsilon}_{y/x} = \frac{1}{100} \cdot 0{,}9220 \cdot 100\% = 0{,}9220\%,$$

что говорит об очень высокой точности линейной модели (4.37).

Аналогично вычисляется средняя ошибка аппроксимации для уравнения регрессии x на y:

$$\bar{\varepsilon}_{x/y} = \frac{1}{n} \cdot \sum_{y} \left| \frac{x - x_y^*}{x} \right| \cdot 100\%$$ [25].

Важным показателем качества регрессионной модели y на x является ***индекс детерминации*** или ***индекс причинности*** R^2. Корень из индекса детерминации равен:

$$R = \sqrt{1 - \frac{\sigma_\varepsilon^2}{\sigma_y^2}}. \tag{4.42}$$

При этом σ_ε^2 – ***остаточная дисперсия***, отображающая вариацию признака Y от всех остальных, кроме X, факторов. Остаточная дисперсия равна:

$$\sigma_\varepsilon^2 = \frac{\sum_y \left(y - y^*\right)^2 \cdot n_y}{\sum_y n_y},$$

выровненные значения y^* подсчитываются для каждого y по формуле:

$$y^* = \frac{\sum_x y_x^* \cdot n_x}{\sum_x n_x}.$$

Общая дисперсия σ_y^2 подсчитывается также как и дисперсия признака Y – по формуле (2.7):

$$\sigma_y^2 = \frac{\sum_y \left(y - \bar{y}\right)^2 \cdot n_y}{\sum_y n_y}.$$

Значит,

$$R^2 = 1 - \frac{\sum_y \left(y - y^*\right)^2 \cdot n_y}{\sum_y \left(y - \bar{y}\right)^2 \cdot n_y}. \tag{4.43}$$

Индекс детерминации определяет часть вариации результативного признака, обусловленного влиянием признака X. Можно сказать

---

[25] Докажите, что для уравнения (4.38) $\bar{\varepsilon}_{x/y} \approx 1{,}4834\%$.

и так: с помощью индекса детерминации можно найти степень влияния признака X на Y.

При значениях тесноты связи свыше 0,7 зависимость y от x признается высокой, а при величинах больше 0,9 – очень высокой (см. табл. 4.6). Это в соответствии с показаниями R^2 означает, что более половины общей вариации результативного фактора обусловлено влиянием факторного признака, включенного в модель. Последнее позволяет считать уравнение регрессии y на x точно описывающим корреляционную зависимость, регрессионная модель признается пригодной для практического применения. При показателях $\eta_{y/x} < 0{,}7$ величина R^2 будет всегда меньше 0,5. Это значит, что на долю признака, включенного в модель, приходится меньше 50% влияния на результативный показатель. Такие регрессионные модели практического значения не имеют.

Пример 4.11. Для модели (4.37) найти индекс детерминации.

Строим расчетную таблицу ($\overline{y} = 39{,}9$).

y	n_y	x	n_x	$\sum_x n_x$	y_x^*	$y_x^* \cdot n_x$	y^*	$\left(y - y^*\right)^2 \cdot n_y$	$\left(y - \overline{y}\right)^2 \cdot n_y$
28	40	4	43	46	30,7844	1323,729	30,1837	190,7490	5664,4
		5	3		21,5743	64,7229			
38	19	3	23	66	39,9941	919,8643	33,9938	304,9369	68,59
		4	43		30,7844	1323,729			
48	23	2	14	37	49,2038	688,8532	43,4789	470,1381	1509,03
		3	23		39,9941	919,8643			
58	18	1	17	31	58,4135	993,0295	54,2543	252,5474	5896,98
		2	14		49,2038	688,8532			
Σ	100	-	-	-	-	-	-	1218,3714	13139

Получаем

$$R^2 = 1 - \frac{1218{,}3714}{13139} \approx 0{,}9073.$$

Найденное значение индекса детерминации говорит о сильной степени влияния факторного признака на результативный.

Очевидно, что чем меньше величина средней ошибки аппроксимации, тем точнее уравнение регрессии будет описывать эмпирические данные. С другой стороны, наибольший интерес представляет модель, в которой влияние факторного признака на результативный максимально. Следовательно, наиболее точной признается та регрессионная модель, у которой

$$\bar{\varepsilon}_{y/x} \to \min, \; R^2 \to \max. \qquad (4.44)$$

4.10. Прогноз результативного показателя

После того, как выбрана наиболее точная регрессионная модель, по ней можно осуществить прогноз результативного показателя по данному факторному признаку. При этом, если требуется спрогнозировать значение y по заданному значению x, то необходимо воспользоваться уравнением регрессии y на x. В случае необходимости прогноза x по имеющемуся значению y, применяем уравнение регрессии x на y.

Пример 4.12. По данным примера 46.1 осуществить прогноз значения y для $x = 5{,}75$ и x, если $y = 64$.

В первом случае подставляем значение x в уравнение (4.37):

$$y = -9{,}2097 \cdot 5{,}75 + 67{,}6232 \approx 14{,}67.$$

Для получения второго прогноза, воспользуемся уравнением (4.38):

$$x = -0{,}0962 \cdot 64 + 6{,}8506 \approx 0{,}69.$$

4.11. Алгоритм построения прогноза

Перечислим основные этапы построения регрессионной модели и осуществления по ней прогноза.

1. Производим все требуемые расчеты. Для этого составляем одну или несколько расчетных таблиц.

2. Для каждого из признаков находим среднюю арифметическую, дисперсию и среднеквадратическое отклонение. Также определяем среднее значение произведения факторного на результативный признак. Здесь же получаем значение ковариации (для контроля лучше ковариацию подсчитать двумя способами – по формулам (4.15) и (4.17)). Наконец, вычисляем коэффициенты регрессии, коэффициент линейной корреляции, корреляционные отношения.

3. Проверяем гипотезу о статистической значимости эмпирических данных (оцениваем качество эмпирических данных). Если она отвергается, то необходимо перейти к другой выборке.

4. С помощью построения эмпирических линий регрессии делаем первоначальные выводы о форме корреляционной зависимости.

5. Используя коэффициент линейной корреляции делаем окончательный вывод о форме корреляционной зависимости.

6. Применяя корреляционное отношение, определяем степень тесноты корреляционной связи.

7. Определяем вид уравнений регрессии y на x и x на у. Желательно, особенно в нелинейном случае, установить несколько уравнений. Параметры уравнений линейной регрессии лучше найти двумя способами: решая системы (4.32) и (4.33), а также с помощью (4.40) и (4.41).

8. Рассчитав для каждой модели среднюю ошибку аппроксимации и индекс детерминации, отбираем наиболее точную регрессионную модель.

9. Осуществляем прогноз результативного показателя по заданному значению факторного признака.

Пример 4.13. В табл. 4.10 приведены результаты изучения зависимости между признаками X и Y.

Таблица 4.10

y \ x	(5;15)	(15;25)	(25;35)	(35;45)	(45;55)	(55;65)	(65;75)	(75;85)
(−1,5;3,5)			9	8	7	5	2	1
(3,5;8,5)		17	10	4	3	3	2	
(8,5;13,5)		14	7					
(13,5;18,5)	10	11	7					
(18,5;23,5)	10	9	6					
(23,5;28,5)	5							

Определить значение y для $x = 2$ и величину x для $y = 20{,}4$.

Производим предварительный этап работы. На нем табл. 4.10 преобразовываем в корреляционную табл. 4.11, строим корреляционное поле данных (рис. 4.11) и корреляционную поверхность (рис. 4.12).

Таблица 4.11

y \ x	10	20	30	40	50	60	70	80	n_y
1			9	8	7	5	2	1	32
6		17	10	4	3	3	2		39
11		14	7						21
16	10	11	7						28
21	10	9	6						25
26	5								5
n_x	25	51	39	12	10	8	4	1	150

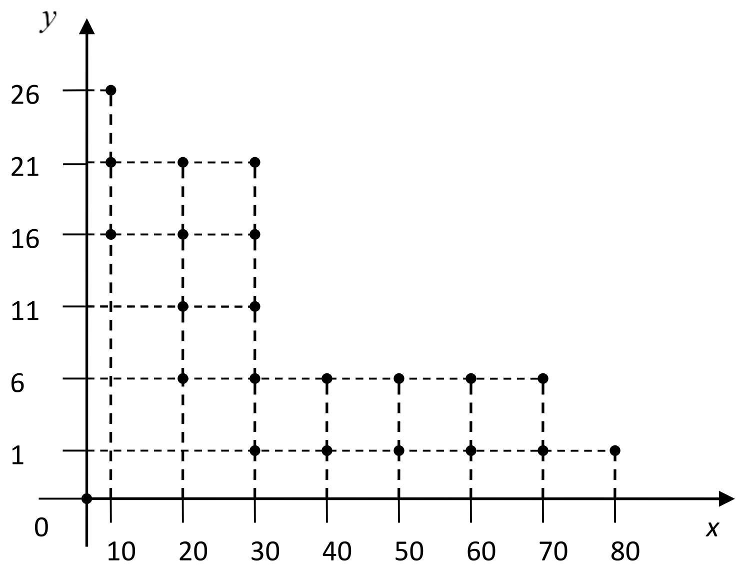

Рис. 4.11

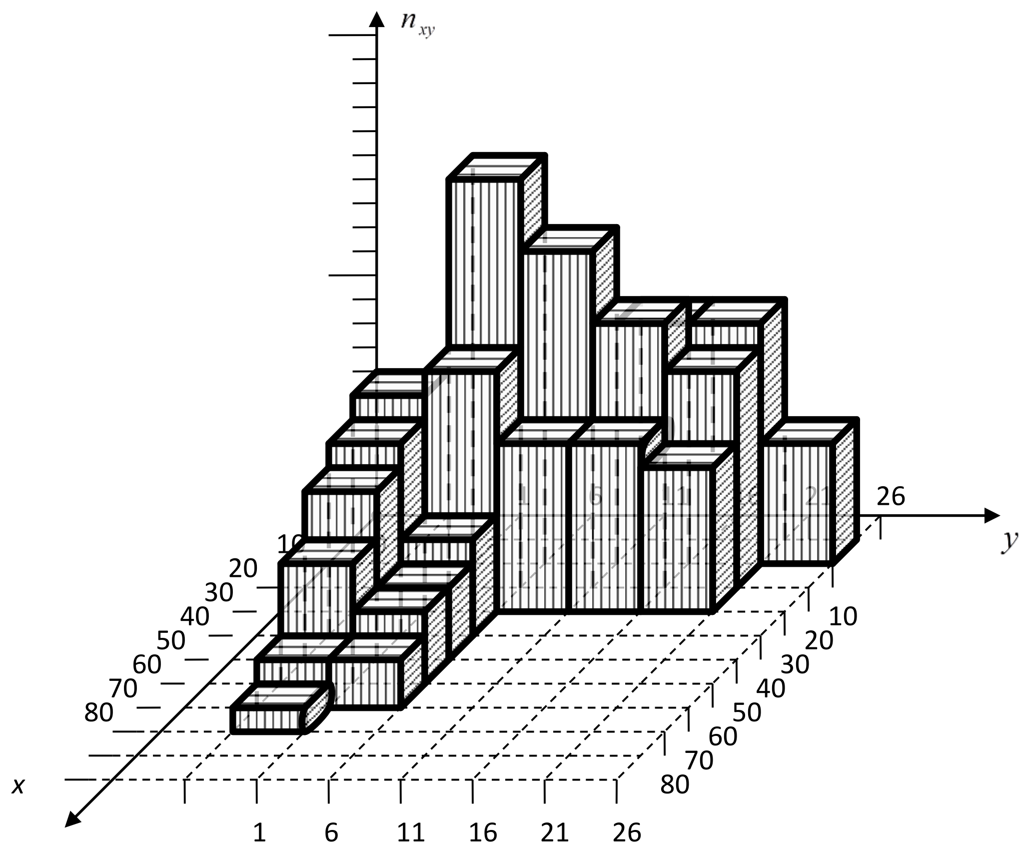

Рис. 4.12

1. Производим все необходимые вычисления в табл. 4.12, 4.13 и 4.14. В клетке табл. 3.20, стоящей на пересечении строки y и столбца x указаны следующие данные:

$y \cdot n_{xy}$		$x \cdot n_{xy}$
	n_{xy}	
$x \cdot y \cdot n_{xy}$		$\alpha_x \cdot \alpha_y \cdot n_{xy}$

Последний столбец табл. 4.13 и 4.14 заполняется только тогда, когда становятся известными средние значения $\bar{y}$ и $\bar{x}$. Обратим внимание на практическое совпадение сумм чисел, стоящих в столбце 3 табл. 4.13 и 8 табл. 4.14, 8 табл. 3.13 и 3 табл. 4.14, а также столбцов 9 в обеих таблицах. Это говорит о правильности произведенных вычислений, что согласуется с формулами (4.34) – (4.36).

2. Используя табл. 4.12 – 4.14, подсчитываем требуемые числовые характеристики[26].

1) Среднее значение, дисперсия и среднеквадратическое отклонение признака *X*:

$$\bar{x} = \frac{4260}{150} = 28{,}4; \ \overline{x^2} = \frac{157000}{150} \approx 1046{,}6667; \ \sigma_x^2 = 240{,}1067; \ \sigma_x \approx 15{,}4954.$$

2) Среднее значение, дисперсия и среднеквадратическое отклонение признака *Y*:

$$\bar{y} = \frac{1600}{150} \approx 10{,}6667; \ \overline{y^2} = \frac{25550}{150} \approx 170{,}3333; \ \sigma_y^2 = 56{,}5548; \ \sigma_y \approx 7{,}5203.$$

3) Среднее значение произведения признаков:

$$\overline{x \cdot y} = \frac{33910}{150} \approx 226{,}0667\,^{27}.$$

[26] Проверьте правильность вычислений самостоятельно.

[27] Обратите внимание, что число, стоящее в числителе дроби, практически равно сумме элементов, стоящих в столбце 9 в табл. 3.21 и 3.22, равенство (16.36) выполняется.

Таблица 4.12

Y \ X		10	20	30	40	50	60	70	80	n_y	$\sum_x x\cdot n_{xy}$	$\bar{x}_y$
		-1	0	1	2	3	4	5	6			
1	-1			9 270 9 270 -9	8 320 8 320 -16	7 350 7 350 -21	5 300 5 300 -20	2 140 2 140 -10	1 80 1 80 -6	32	1460	45,625
6	0		102 340 17 2040 0	60 300 10 1800 0	24 160 4 960 0	18 150 3 900 0	18 180 3 1080 0	12 140 2 840 0		39	1270	32,5641
11	1		154 280 14 3080 0	77 210 7 2310 7						21	490	23,3333
16	2	160 100 10 1600 -20	176 220 11 3520 0	112 210 7 3360 14						28	530	18,9286
21	3	210 100 10 2100 -30	189 180 9 3780 0	126 180 6 3780 18						25	460	18,4
26	4	130 50 5 1300 -20								5	50	10
n_x		25	51	39	12	10	8	4	1	150		
$\sum_y y\cdot n_{xy}$		500	621	384	32	25	23	14	1	Σ		
$\bar{y}_x$		20	12,1765	9,8462	2,6667	2,5	2,875	3,5	1			
$x\cdot y\cdot n_{xy}$		5000	12420	11520	1280	1250	1380	980	80	33910		
$\alpha_x\cdot\alpha_y\cdot n_{xy}$		-70	0	30	-16	-21	-20	-10	-6	-113		

Таблица 4.13

x	n_x	$x \cdot n_x$	$x^2 \cdot n_x$	$x^3 \cdot n_x$	$x^4 \cdot n_x$	$\bar{y}_x$	$\bar{y}_x \cdot n_x$	$x \cdot \bar{y}_x \cdot n_x$	$x^2 \cdot \bar{y}_x \cdot n_x$	$\ln \bar{y}_x$	$\ln \bar{y}_x \cdot n_x$	$x \cdot \ln \bar{y}_x \cdot n_x$	$(\bar{y}_x - \bar{y})^2 \cdot n_x$
10	25	250	2500	25000	250000	20	500	5000	50000	2,9957	74,8925	748,925	2177,7622
20	51	1020	20400	408000	8160000	12,1765	621,0015	12420,03	248400,6	2,4995	127,4725	2549,49	116,2543
30	39	1170	35100	1053000	31590000	9,8462	384,0018	11520,05	345601,6	2,2871	89,1969	2675,907	26,2556
40	12	480	19200	768000	30720000	2,6667	32,0004	1280,016	51200,64	0,9808	11,7696	470,784	768
50	10	500	25000	1250000	62500000	2,5	25	1250	62500	0,9163	9,1629	458,1454	666,9499
60	8	480	28800	17280000	103680000	2,875	23	1380	82800	1,0561	8,4488	506,928	485,6847
70	4	280	19600	1372000	96040000	3,5	14	980	68600	1,2528	5,0112	350,784	205,4464
80	1	80	6400	512000	40960000	1	1	80	6400	0	0	0	93,4451
Σ	150	4260	157000	7116000	373900000	-	1600,0037	33910,1	915502,86	-	325,9565	7760,968	4539,7982

Таблица 4.14

y	n_y	$y \cdot n_y$	$y^2 \cdot n_y$	$y^3 \cdot n_y$	$y^4 \cdot n_y$	$\bar{x}_y$	$\bar{x}_y \cdot n_y$	$y \cdot \bar{x}_y \cdot n_y$	$y^2 \cdot \bar{x}_y \cdot n_y$	$\ln \bar{x}_y$	$\ln \bar{x}_y \cdot n_y$	$y \cdot \ln \bar{x}_y \cdot n_y$	$(\bar{x}_y - \bar{x})^2 \cdot n_y$
1	32	32	32	32	32	45,6250	1460	1460	1460	3,8205	122,256	122,256	9494,42
6	39	234	1404	8424	50544	32,5641	1269,9999	7619,9994	45719,9964	3,4832	135,8448	815,0688	676,2494
11	21	231	2541	27951	307461	23,3333	489,9993	5389,9923	59289,9153	3,1499	66,1479	727,6269	539,1004
16	28	448	7168	114688	1835008	18,9286	530,0008	8480,0128	135680,2048	2,9407	82,3396	1317,4336	2511,8077
21	25	525	11025	231525	4862025	18,4	460	9660	202860	2,9124	72,81	1529,01	2500
26	5	130	3380	87880	2284880	10	50	1300	33800	2,3026	11,513	299,338	1692,8
Σ	150	1600	25550	470500	9339950	-	4260	33910,0045	478810,1165	-	490,9113	4810,7333	17414,3775

4) Ковариация:

а) Определение

$$\mu = 226{,}0667 - 28{,}4 \cdot 10{,}6667 \approx -76{,}8676.$$

б) Метод моментов

$$\mu = \frac{10 \cdot 5}{150} \cdot (-113) - (28{,}4 - 20) \cdot (10{,}6667 - 6) \approx -76{,}8676.$$

5) Коэффициенты регрессии:

$$\rho_{y/x} = \frac{-76{,}8676}{240{,}1067} \approx -0{,}3201, \quad \rho_{x/y} = \frac{-76{,}8676}{56{,}5548} \approx -1{,}3592.$$

6) Коэффициент линейной корреляции

$$r = \frac{-76{,}8676}{15{,}4954 \cdot 7{,}5203} \approx -0{,}6596.$$

7) Корреляционные отношения:

$$\delta_{y/x} = \sqrt{\frac{4539{,}7982}{150}} \approx 5{,}5014, \quad \sigma_y = 7{,}5203,$$

$$\eta_{y/x} = \frac{5{,}5014}{7{,}5203} \approx 0{,}7315;$$

$$\delta_{x/y} = \sqrt{\frac{17414{,}3775}{150}} \approx 10{,}7748, \quad \sigma_x \approx 15{,}4954,$$

$$\eta_{x/y} = \frac{10{,}7748}{15{,}4954} \approx 0{,}6954.$$

3. Гипотезу о статистической значимости эмпирических данных проверим с доверительной вероятностью 0,99. Наблюдаемое значение t-критерия Стьюдента равно

$$t_{набл} = \sqrt{\frac{(-0{,}6954)^2}{1-(-0{,}6954)^2} \cdot (150-2)} \approx 10{,}6772.$$

Переходим к расчету критического значения. Имеем:

$$\alpha = 1 - 0{,}99 = 0{,}01, \quad \nu = 150 - 2 = 148.$$

Критическое значение

$$t_{крит} = t(0{,}01; 148) \approx 2{,}58$$

согласно табл. 4 прил.

Поскольку

$$10{,}6772 > 2{,}58, \quad t_{набл} > t_{крит},$$

правило проверки гипотезы выполняется. Это значит, что с вероятностью 0,99 можно говорить о наличии между признаками статистически существенной зависимости.

4. Строим эмпирические линии регрессии (рис. 4.13). Анализ эмпирических линий регрессии показывает, что корреляционная зави-

симость между признаками X и Y не обладает высокой степенью линейности.

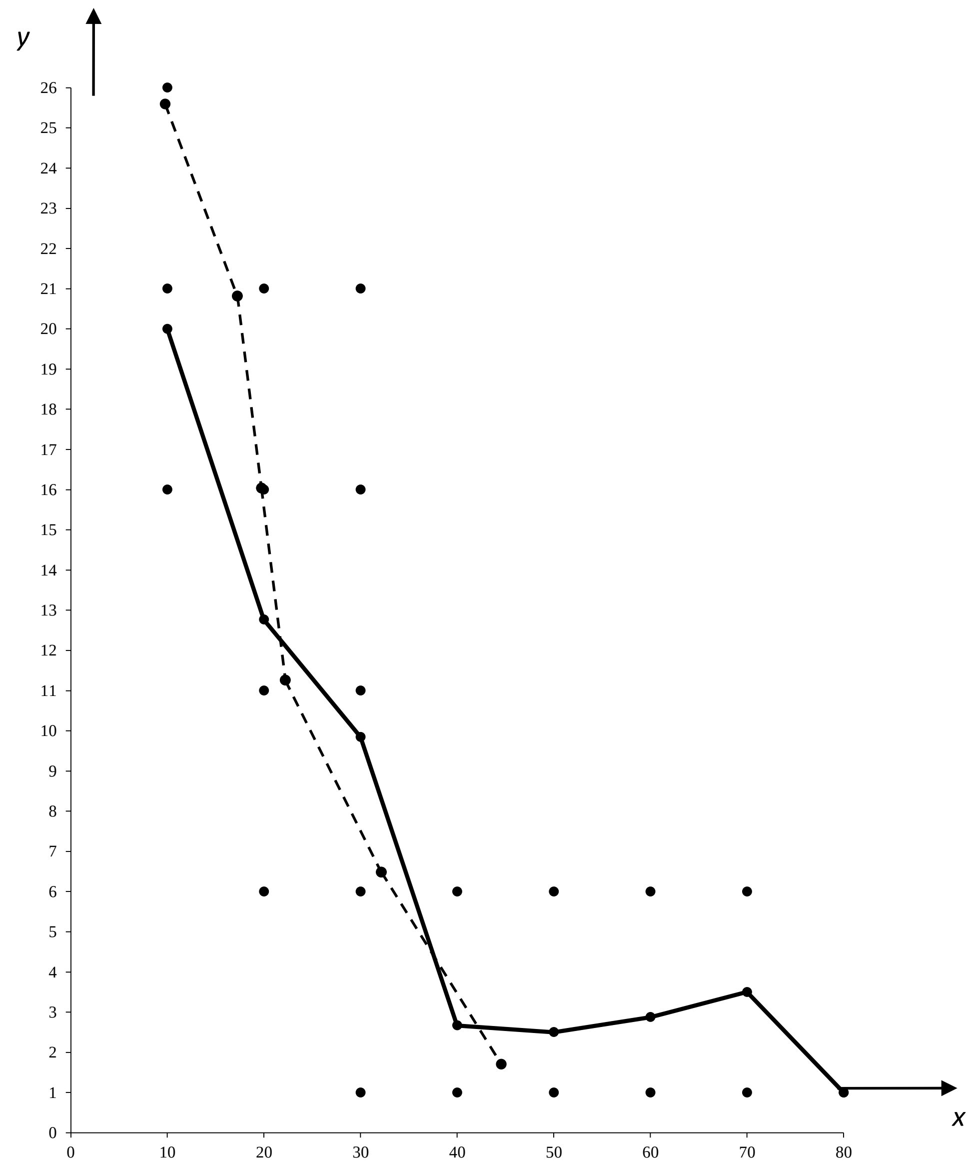

Рис. 4.13

5. Анализируем табл. 4.5. Получаем: корреляционная зависимость между признаками X и Y является линейной, обратной, поскольку

$$-1 < -0{,}6596 < -0{,}5 .$$

6. Так как значение корреляционного отношения

$$0{,}7 < 0{,}7315 < 1,$$

то связь признает сильной согласно табл. 4.6.

7. Пользуясь табл. 4.13 и 4.14, составляем системы нормальных уравнений и, решая их, получаем параметры уравнений регрессии *y* на *x* и *x* на *y* в линейном, параболическом и показательном случаях. В линейном случае применим еще один, упрощенный способ (4.40) и (4.41). Для каждой из регрессионных моделей по расчетным точкам строим графики.

Линейная корреляционная зависимость

Регрессия *y* на *x*. Система нормальных уравнений:

$$\begin{cases} 157000\cdot a + 4260\cdot b = 33910{,}1 \\ 4260\cdot a + 150\cdot b = 1600{,}0037 \end{cases}, \quad \begin{cases} a \approx -0{,}3201 \\ b \approx 19{,}7586 \end{cases},$$

$$y = -0{,}3201\cdot x + 19{,}7586.$$

Нахождение параметров уравнения прямой линии регрессии по формуле (4.40):

$$\rho_{y/x} = -0{,}6596\cdot\frac{7{,}5203}{15{,}4954} \approx -0{,}3201,$$

$$y - 10{,}6667 = -0{,}3201\cdot(x - 28{,}4),$$

$$y = -0{,}3201\cdot x + 19{,}7587.$$

Теоретические значения признака *Y*:

x	10	20	30	40	50	60	70	80
y_x^*	16,5568	13,3558	10,1548	6,9538	3,7528	0,5518	-2,6492	-5,8502

Регрессия x на y. Система нормальных уравнений:

$$\begin{cases} 25550\cdot a' + 1600\cdot b' = 33910{,}0045 \\ 1600\cdot a' + 150\cdot b' = 4260 \end{cases}, \quad \begin{cases} a' \approx -1{,}3591 \\ b' \approx 42{,}8974 \end{cases},$$

$$x = -1{,}3591\cdot y + 42{,}8974.$$

Нахождение параметров уравнения прямой линии регрессии по формуле (4.41):

$$\rho_{x/y} = -0{,}6596\cdot\frac{15{,}4954}{7{,}5203} \approx -1{,}3592,$$

$$x - 28{,}4 = -1{,}3592\cdot(y - 10{,}6667),$$

$$x = -1{,}3592\cdot y + 42{,}8979.$$

Теоретические значения признака *X*:

y	1	6	11	16	21	26
x_y^*	41,5383	34,7428	27,9473	21,1518	14,3563	7,5608

Строим чертеж (рис. 4.14). На рис. 4.14 точками показаны эмпирические данные, сплошной линией – прямая регрессии y на x, а пунктирной – прямая регрессии x на y.

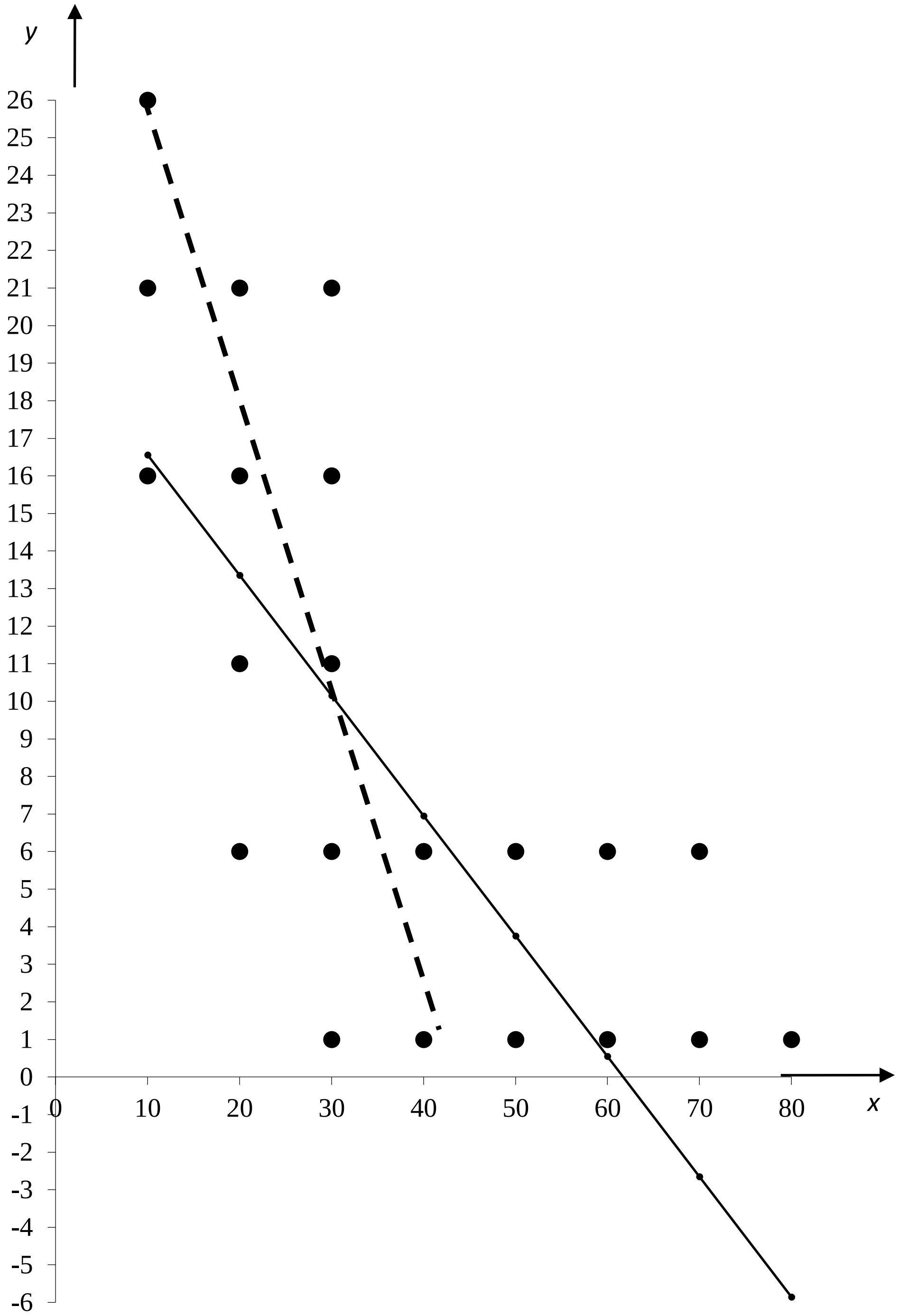

Рис. 4.14

Параболическая корреляционная зависимость

Регрессия y на x. Система нормальных уравнений:

$$\begin{cases} 373900000\cdot a + 7116000\cdot b + 157000\cdot c = 91550{,}86 \\ 7116000\cdot a + 157000\cdot b + 4260\cdot c = 33910{,}1 \\ 157000\cdot a + 4260\cdot b + 150\cdot c = 1600{,}0037 \end{cases}, \quad \begin{cases} a \approx 0{,}0068 \\ b \approx -0{,}8191, \\ c \approx 26{,}8502 \end{cases}$$

$$y = 0{,}0068\cdot x^2 - 0{,}8192\cdot x + 26{,}8502.$$

Теоретические значения признака Y:

x	10	20	30	40	50	60	70	80
y_x^*	19,3392	13,1882	8,3972	4,9662	2,8952	2,1842	2,8332	4,8422

Регрессия x на y. Система нормальных уравнений:

$$\begin{cases} 9339950\cdot a' + 470500\cdot b' + 25550\cdot c' = 478810{,}1165 \\ 470500\cdot a' + 25550\cdot b' + 1600\cdot c' = 33910{,}0045 \\ 25550\cdot a' + 1600\cdot b' + 150\cdot c' = 4260 \end{cases}, \quad \begin{cases} a' \approx 0{,}0604 \\ b' \approx -2{,}7696, \\ c' \approx 47{,}6470 \end{cases}$$

$$x = 0{,}0604\cdot y^2 - 2{,}7696\cdot y + 47{,}6470.$$

Теоретические значения признака X:

y	1	6	11	16	21	26
x_y^*	44,9379	33,2074	24,5019	18,8214	16,1659	16,5354

Строим чертеж (рис. 4.15). На рис. 4.15 точками показаны эмпирические данные, сплошной линией – параболическая линия регрессии y на x, а пунктирной – параболическая линия регрессии x на y.

Показательная корреляционная зависимость.

Регрессия y на x. Система нормальных уравнений:

$$\begin{cases} 157000\cdot \ln a + 4260\cdot \ln b = 7760{,}9356 \\ 4260\cdot \ln a + 150\cdot \ln b = 325{,}9566 \end{cases}, \quad \begin{cases} \ln a \approx -0{,}0415 \\ \ln b \approx 3{,}3529 \end{cases}, \quad \begin{cases} a \approx 0{,}9593 \\ b \approx 28{,}5855 \end{cases}[28],$$

$$y = 28{,}5855\cdot (0{,}9593)^x.$$

Теоретические значения признака Y:

x	10	20	30	40	50	60	70	80
y_x^*	18,8665	12,4519	8,2182	5,4240	3,5799	2,3627	1,5594	1,0292

[28] Для того, чтобы решить уравнение вида $\ln x = k$, необходимо осуществить операцию потенцирование: $e^{\ln x} = e^k$; откуда $x = e^k$.

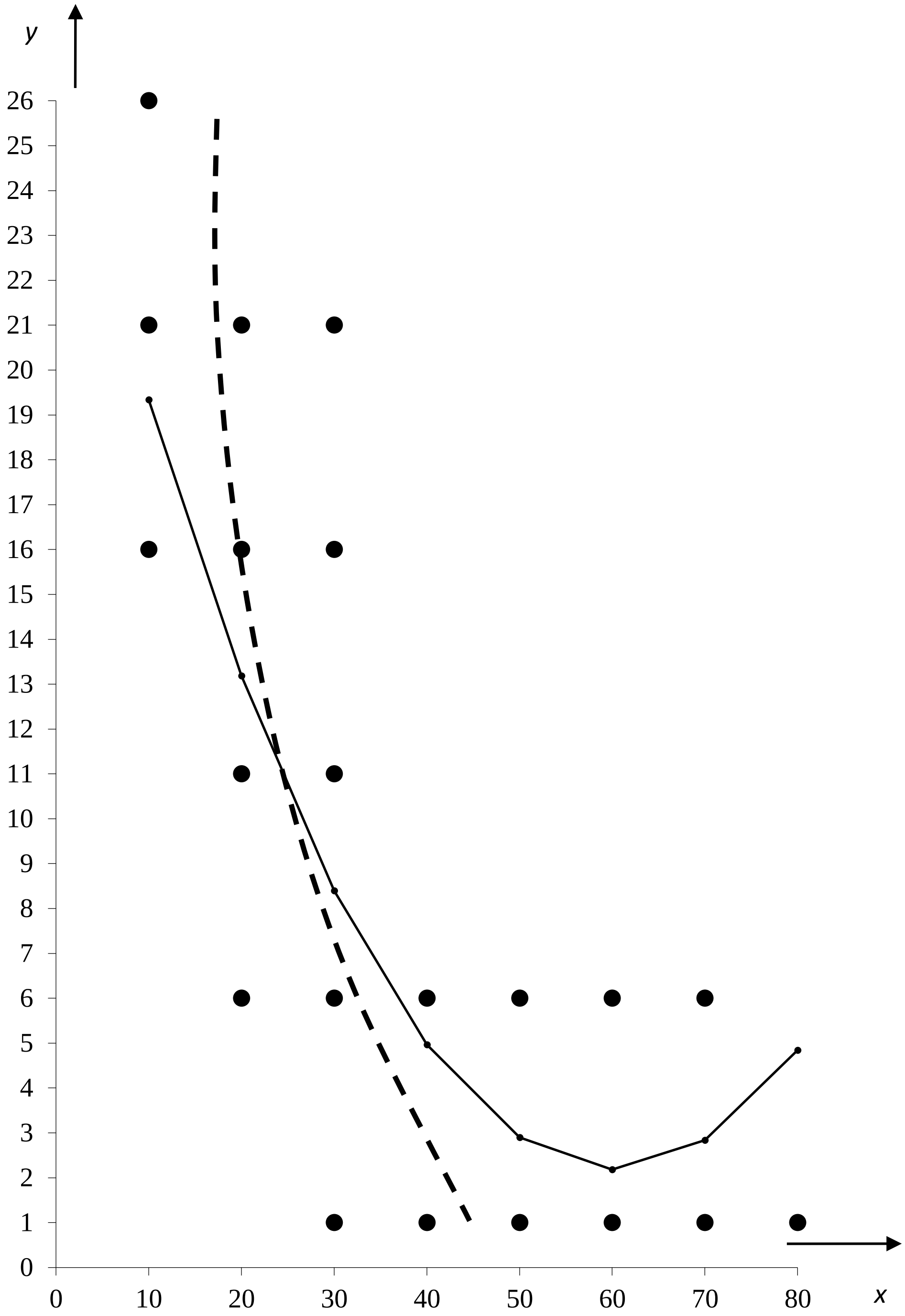

Рис. 4.15

Регрессия x на y. Система нормальных уравнений:

$$\begin{cases} 25550 \cdot \ln a' + 1600 \cdot \ln b' = 4810{,}7333 \\ 1600 \cdot \ln a' + 150 \cdot \ln b' = 490{,}9113 \end{cases}, \quad \begin{cases} \ln a' \approx -0{,}0502 \\ \ln b' \approx 3{,}8079 \end{cases}, \quad \begin{cases} a' \approx 0{,}9510 \\ b' \approx 45{,}0557 \end{cases},$$

$$x = 45{,}0557 \cdot (0{,}9510)^y.$$

Строим чертеж (рис. 4.16). На рис. 4.16 точками показаны эмпирические данные, сплошной линией – показательная линия регрессии y на x, а пунктирной – показательная линия регрессии x на y.

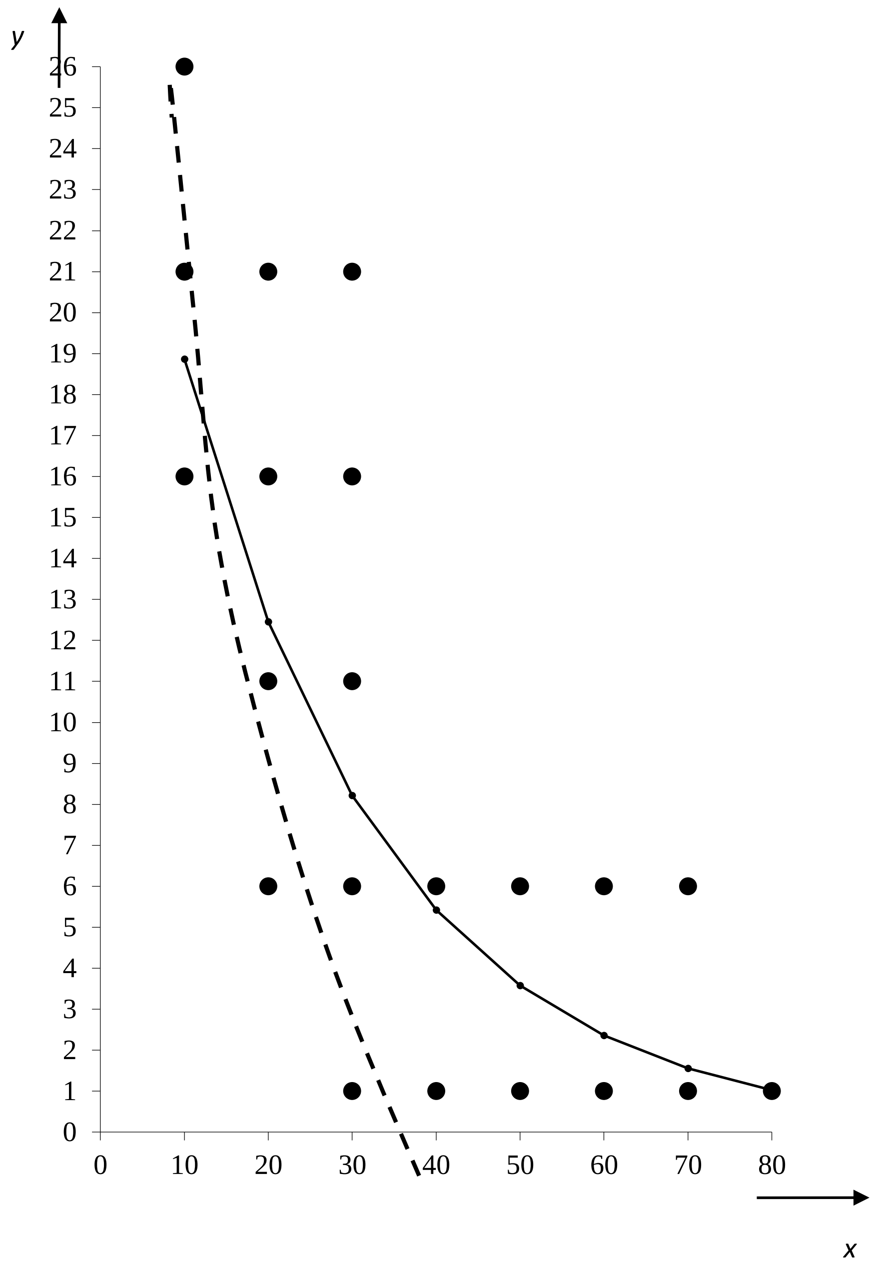

Рис. 4.16

Теоретические значения признака X:

y	1	6	11	16	21	26
x_y^*	42,8480	33,3298	25,9260	20,1668	15,6870	12,2023

8. Рассчитаем для каждой модели среднюю ошибку аппроксимации и индекс детерминации. Для этого составляем расчетные таблицы:

						Линейная модель				
y	n_y	$(y-\overline{y})^2 \cdot n_y$	x	n_x	$\sum_x n_x$	y_x^*	$\left\lvert \frac{y-y_x^*}{y} \right\rvert$	$y_x^* \cdot n_x$	y^*	$(y-y^*)^2 \cdot n_y$
1	32	2990,2428	30	39	74	10,1548	9,1548	396,0372	6,8240	1085,4183
			40	12		6,9538	5,9538	83,4456		
			50	10		3,7528	2,7528	37,528		
			60	8		0,5518	0,4482	4,4144		
			70	4		-2,6492	3,6492	-10,5968		
			80	1		-5,8502	6,8502	-5,8502		
6	39	849,3455	20	51	124	13,3558	1,2260	681,1458	9,6127	509,0111
			30	39		10,1548	0,6925	396,0372		
			40	12		6,9538	0,1590	83,4456		
			50	10		3,7528	0,3745	37,528		
			60	8		0,5518	0,9080	4,4144		
			70	4		-2,6492	1,4415	-10,5968		
11	21	2,3329	20	51	90	13,3558	0,2142	681,1458	11,9687	19,7060
			30	39		10,1548	0,0768	396,0372		
16	28	796,4345	10	25	115	16,5568	0,0348	413,92	12,9661	257,7252
			20	51		13,3558	0,1653	681,1458		
			30	39		10,1548	0,3653	396,0372		
21	25	2669,4272	10	25	115	16,5568	0,2116	413,92	12,9661	1613,5835
			20	51		13,3558	0,3640	681,1458		
			30	39		10,1548	0,5164	396,0372		
26	5	1175,5504	10	25	25	16,5568	0,3632	413,92	16,5568	445,8701
Σ	150	8483,3333	-	-	-	-	35,9221	-	-	3931,3142

y	n_y	$(y-\bar{y})^2 \cdot n_y$	x	n_x	$\sum\limits_x n_x$	**Параболическая модель**				
						y_x^*	$\left\lvert \dfrac{y-y_x^*}{y} \right\rvert$	$y_x^* \cdot n_x$	y^*	$(y-y^*)^2 \cdot n_y$
1	32	2990,2428	30	39	74	8,3972	7,3972	327,4908	6,0768	824,7762
			40	12		4,9662	3,9662	59,5944		
			50	10		2,8952	1,8952	28,952		
			60	8		2,1842	1,1842	17,4736		
			70	4		2,8332	1,8332	11,3328		
			80	1		4,8422	3,8422	4,8422		
6	39	849,3455	20	51	124	13,1882	1,1980	672,5982	9,0116	353,7261
			30	39		8,3972	0,3995	327,4908		
			40	12		4,9662	0,1723	59,5944		
			50	10		2,8952	0,5175	28,952		
			60	8		2,1842	0,6360	17,4736		
			70	4		2,8332	0,5278	11,3328		
11	21	2,3329	20	51	90	13,1882	0,1989	672,5982	11,1121	0,2639
			30	39		8,3972	0,2366	327,4908		
16	28	796,4345	10	25	115	19,3392	0,2087	483,48	12,9006	268,9759
			20	51		13,1882	0,1757	672,5982		
			30	39		8,3972	0,4752	327,4908		
21	25	2669,4272	10	25	115	19,3392	0,0791	483,48	12,9006	1640,007
			20	51		13,1882	0,3720	672,5982		
			30	39		8,3972	0,6001	327,4908		
26	5	1175,5504	10	25	25	19,3392	0,2562	483,48	19,3392	221,8313
Σ	150	8483,3333	-	-	-	-	26,17185	-	-	3309,58

y	n_y	$(y-\bar{y})^2 \cdot n_y$	x	n_x	$\sum_x n_x$	**Показательная модель**				
						y_x^*	$\left\lvert \frac{y-y_x^*}{y} \right\rvert$	$y_x^* \cdot n_x$	y^*	$(y-y^*)^2 \cdot n_y$
1	32	2990,2428	30	39	74	8,2182	7,218246	320,5116	6,0482	815,5014
			40	12		5,4240	4,424049	65,08858		
			50	10		3,5799	2,579876	35,79876		
			60	8		2,3627	1,362721	18,90177		
			70	4		1,5594	0,559398	6,237591		
			80	1		1,0292	0,029204	1,029204		
6	39	849,3455	20	51	124	12,4519	1,0753	635,0455	8,7225	289,0577
			30	39		8,2182	0,3697	320,5116		
			40	12		5,4240	0,0960	65,08858		
			50	10		3,5799	0,4034	35,79876		
			60	8		2,3627	0,6062	18,90177		
			70	4		1,5594	0,7401	6,237591		
11	21	2,3329	20	51	90	12,4519	0,1320	635,0455	10,6173	3,0756
			30	39		8,2182	0,2529	320,5116		
16	28	796,4345	10	25	115	18,8665	0,1792	471,6613	12,4106	360,7472
			20	51		12,4519	0,2218	635,0455		
			30	39		8,2182	0,4864	320,5116		
21	25	2669,4272	10	25	115	18,8665	0,1016	471,6613	12,4106	1844,447
			20	51		12,4519	0,4071	635,0455		
			30	39		8,2182	0,6087	320,5116		
26	5	1175,5504	10	25	25	18,8665	0,2744	471,6613	18,8665	254,4375
Σ	150	8483,3333	-	-	-	-	22,12799	-		3567,266

Теперь, пользуясь соответствующими формулами находим среднюю ошибку аппроксимации $\bar{\varepsilon}_{y/x}$ и индекс детерминации R^2[29]:

	Линейная модель	**Параболическая модель**	**Показательная модель**
$\bar{\varepsilon}_{y/x}$	23,9481	17,4479	14,7520
R^2	0,5366	0,6099	0,5795

Сравнивая соответствующие строки последней таблицы, получаем: наиболее точной является показательная модель. По ней возможен наиболее точный прогноз.

9. Осуществляем прогноз:

$$x = 2,\ y = 28,5855 \cdot (0,9593)^2 \approx 26,31;$$

$$y = 20,4,\ x = 45,0557 \cdot (0,9510)^{20,4} \approx 16,18.$$

[29] Обязательно проделайте соответствующие вычисления самостоятельно.

§ 5. МНОЖЕСТВЕННАЯ РЕГРЕССИЯ

5.1. Введение

В параграфе 4 были рассмотрены случаи между одним факторным и результативным показателями. Такая регрессия называется ***парной***.

Однако, в жизни мы сталкиваемся с тем, что число факторных признаков может быть больше 1. Например, стоимость квартиры не зависит не от одного показателя (скажем – жилой площади). На стоимость квартиры влияют также и год постройки дома, материал, из которого изготовлен дом, район, в котором находится дом и многое другое. Более того, есть более значимые признаки, оказывающие большое влияние на результативный фактор, а есть и незначимые. Скажем известно, что наличие ремонта в квартире мало влияет на ее стоимость.

Итак, *если число факторных признаков* $k \geq 2$, *то регрессия называется* ***многофакторной*** *или* ***множественной***. В этом случае уравнение регрессии между факторными признаками x_1, x_2,..., X_k и результативным Y имеет следующий общий вид:

$$y = f(x_1; x_2; ...; x_k).$$

Перечислим некоторые многофакторные регрессионные модели, широко применяемые в экономике:

1. линейная функция:

$$y = a_1 \cdot x_1 + a_2 \cdot x_2 + ... + a_k \cdot x_k + b = \sum_{i=1}^{k} a_i \cdot x_i + b, \tag{5.1}$$

в частности, если $k = 2$, то

$$y = a_1 \cdot x_1 + a_2 \cdot x_2 + b; \tag{5.2}$$

2. степенная функция

$$y = b \cdot x_1^{\alpha_1} \cdot x_2^{\alpha_2} \cdot ... \cdot x_k^{\alpha_k} = b \prod_{i=1}^{k} x_i^{\alpha_i}, \tag{5.3}$$

частным случаем которой является производственная функция Кобба[30] – Дугласа[31]

$$y = b \cdot x_1^{\alpha} \cdot x_2^{\beta}, \tag{5.4}$$

---

[30] Кобб Чарльз (1875 - 1949) – американский математик и экономист.

[31] Дуглас Пол Говард (1892 - 1976) – американский математик и статистик.

примененная при анализе развития американской экономики в 20 – 30 гг. XX века и выражающая национальный доход y через коэффициент размерности b, объемы приложенного труда x_1 и капитала x_2 и $0 \le \alpha \le 1$ и $0 \le \beta \le 1$ – коэффициенты эластичности по труду и капиталу;

3. параболическая функция

$$y = a_1 \cdot x_1^2 + a_2 \cdot x_2^2 + .. + a_k \cdot x_k^2 + b = \sum_{i=1}^{k} a_i \cdot x_i^2 + b;$$

4. гиперболическая функция

$$y = a_1 \cdot \frac{1}{x_1} + a_2 \cdot \frac{1}{x_2} + ... a_k \cdot \frac{1}{x_k} + b = \sum_{i=1}^{k} a_i \cdot \frac{1}{x_i} + b;$$

5. показательная функция

$$y = a^{\sum_{i=1}^{k} a_i \cdot x_i + b}$$ [32].

К специальным экономическим функциям можно отнести функцию полезности, которая чаще всего встречается в двух видах:

1. логарифмический вид

$$y = \sum_{i=1}^{k} a_i \cdot \ln(x_i - c_i),\ a_i > 0,\ x_i > 0,\ c_i \ge 0;$$

2. функция постоянной эластичности

$$y = \sum_{i=1}^{k} \frac{a_i}{1 - b_i} \cdot (x_i - c_i)^{1-b_i},\ a_i > 0,\ 0 < b_i < 1,\ x_i > 0,\ c_i \ge 0.$$

Множественная регрессия решает те же задачи и отвечает на те вопросы, что и парная.

Эмпирические данные выборки объема n принято записывать в виде таблицы, в которой Y – результативный признак со значениями y_i, а X_1, X_2,…, X_k - факторные признаки со значениями x_{ij}, i=1,2,…, n, j=1,2,…k (табл. 5.1):

Таблица 5.1

	Y	X_1	X_2	…	X_k
1	y_1	x_{11}	x_{12}	…	x_{1k}
2	y_2	x_{21}	x_{22}	…	x_{2k}
⋮	⋮	⋮	⋮	⋮	⋮
N	y_n	x_{n1}	x_{n2}	…	x_{nk}

[32] В качестве основания степени часто ставится число е или 10.

5.2. Оценка силы и тесноты корреляционной связи

Оценка тесноты связи производится с помощью ***коэффициента множественной корреляции*** $R_{y/x_1,x_2,\dots,x_k}$. Он обладает рядом свойств, аналогичных свойствам коэффициента линейной корреляции[33] и корреляционного отношения[34]. В частности, если

$$R_{y/x_1,x_2,\dots,x_k} \geq 0{,}5,$$

то связь – линейная. Если же

$$R_{y/x_1,x_2,\dots,x_k} \geq 0{,}7,$$

то связь – линейная и сильная.

Коэффициент множественной корреляции рассчитывается также как и (16.42):

$$R_{y/x_1,x_2,\dots,x_k} = \sqrt{1-\frac{\sigma_\varepsilon^2}{\sigma_y^2}}, \tag{5.5}$$

при этом σ_ε^2 - ***остаточная дисперсия***, σ_y^2 - ***общая дисперсия***. Остаточную дисперсию можно найти, подсчитав определитель симметричной матрицы, составленной из ***парных коэффициентов линейной корреляции*** r_{yx_i} и ***межфакторных коэффициентов линейной корреляции*** $r_{x_ix_j}$:

$$\sigma_\varepsilon^2 = \begin{vmatrix}
1 & r_{yx_1} & r_{yx_2} & r_{yx_3} & r_{yx_4} & r_{yx_5} & \cdots & r_{yx_{k-1}} & r_{yx_k} \\
r_{yx_1} & 1 & r_{x_1x_2} & r_{x_1x_3} & r_{x_1x_4} & r_{x_1x_5} & \cdots & r_{x_1x_{k-1}} & r_{x_1x_k} \\
r_{yx_2} & r_{x_1x_2} & 1 & r_{x_2x_3} & r_{x_2x_4} & r_{x_2x_5} & \cdots & r_{x_2x_{k-1}} & r_{x_2x_k} \\
r_{yx_3} & r_{x_1x_3} & r_{x_2x_3} & 1 & r_{x_3x_4} & r_{x_3x_5} & \cdots & r_{x_3x_{k-1}} & r_{x_3x_1} \\
r_{yx_4} & r_{x_1x_4} & r_{x_2x_4} & r_{x_3x_4} & 1 & r_{x_4x_5} & \cdots & r_{x_4x_{k-1}} & r_{x_4x_k} \\
r_{yx_5} & r_{x_1x_5} & r_{x_2x_5} & r_{x_3x_5} & r_{x_4x_5} & 1 & \cdots & r_{x_5x_{k-1}} & r_{x_5x_k} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
r_{yx_{k-1}} & r_{x_1x_{k-1}} & r_{x_2x_{k-1}} & r_{x_3x_{k-1}} & r_{x_4x_{k-1}} & r_{x_5x_{k-1}} & \cdots & 1 & r_{x_{k-1}x_k} \\
r_{yx_k} & r_{x_1x_k} & r_{x_2x_k} & r_{x_3x_1} & r_{x_4x_k} & r_{x_5x_k} & \cdots & r_{x_{k-1}x_k} & 1
\end{vmatrix}.$$

Остаточная дисперсия численно равна определителю матрицы, составленной из межфакторных коэффициентов линейной корреляции $r_{x_ix_j}$:

---

[33] См. пункт 4.4.

[34] См. пункт 4.5.

$$\sigma_y^2 = \begin{vmatrix} 1 & r_{x_1x_2} & r_{x_1x_3} & r_{x_1x_4} & r_{x_1x_5} & \dots & r_{x_1x_{k-1}} & r_{x_1x_k} \\ r_{x_1x_2} & 1 & r_{x_2x_3} & r_{x_2x_4} & r_{x_2x_5} & \dots & r_{x_2x_{k-1}} & r_{x_2x_k} \\ r_{x_1x_3} & r_{x_2x_3} & 1 & r_{x_3x_4} & r_{x_3x_5} & \dots & r_{x_3x_{k-1}} & r_{x_3x_k} \\ r_{x_1x_4} & r_{x_2x_4} & r_{x_3x_4} & 1 & r_{x_4x_5} & \dots & r_{x_4x_{k-1}} & r_{x_4x_k} \\ r_{x_1x_5} & r_{x_2x_5} & r_{x_3x_5} & r_{x_4x_5} & 1 & \dots & r_{x_5x_{k-1}} & r_{x_5x_k} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ r_{x_1x_{k-1}} & r_{x_2x_{k-1}} & r_{x_3x_{k-1}} & r_{x_4x_{k-1}} & r_{x_5x_{k-1}} & \dots & 1 & r_{x_{k-1}x_k} \\ r_{x_1x_k} & r_{x_2x_k} & r_{x_3x_k} & r_{x_4x_k} & r_{x_5x_k} & \dots & r_{x_{k-1}x_k} & 1 \end{vmatrix}.$$

Парные и межфакторные коэффициенты линейной корреляции рассчитываются также как и коэффициент линейной корреляции (4.18):

$$r_{yx_i} = \frac{\overline{y \cdot x_i} - \overline{y} \cdot \overline{x_i}}{\sigma_y \cdot \sigma_{x_i}} = \frac{\mu_{yx_i}}{\sigma_y \cdot \sigma_{x_i}}; \; r_{x_ix_j} = \frac{\overline{x_i \cdot x_j} - \overline{x_i} \cdot \overline{x_j}}{\sigma_{x_i} \cdot \sigma_{x_j}} = \frac{\mu_{x_ix_j}}{\sigma_{x_i} \cdot \sigma_{x_j}}. \qquad (5.6)$$

Они обладают свойствам, аналогичными свойствам r.

Для проверки правильности произведенных вычислений пользуемся неравенством:

$$R_{y/x_1,x_2,\dots,x_k} \ge r_{yx_i} \text{ для всех } i = 1,2,\dots,k. \qquad (5.7)$$

Получим расчетную формулу нахождения коэффициента множественной регрессии для случая $k = 2$. Вначале, пользуясь правилом нахождения определителя матрицы второго и третьего порядка, вычисляем остаточную и общую дисперсию:

$$\sigma_\varepsilon^2 = \begin{vmatrix} 1 & r_{yx_1} & r_{yx_2} \\ r_{yx_1} & 1 & r_{x_1x_2} \\ r_{yx_2} & r_{x_1x_2} & 1 \end{vmatrix} = 1 + 2 \cdot r_{yx_1} \cdot r_{yx_2} \cdot r_{x_1x_2} - r_{yx_1}^2 - r_{yx_2}^2 - r_{x_1x_2}^2;$$

$$\sigma_y^2 = \begin{vmatrix} 1 & r_{x_1x_2} \\ r_{x_1x_2} & 1 \end{vmatrix} = 1 - r_{x_1x_2}^2.$$

Найденные значения подставляем в (5.5): $R_{y/x_1,x_2} = \sqrt{1 - \frac{\sigma_\varepsilon^2}{\sigma_y^2}} =$

$$= \sqrt{1 - \frac{1 + 2 \cdot r_{yx_1} \cdot r_{yx_2} \cdot r_{x_1x_2} - r_{yx_1}^2 - r_{yx_2}^2 - r_{x_1x_2}^2}{1 - r_{x_1x_2}^2}} = \sqrt{\frac{r_{yx_1}^2 + r_{yx_2}^2 - 2 \cdot r_{yx_1} \cdot r_{yx_2} \cdot r_{x_1x_2}}{1 - r_{x_1x_2}^2}},$$

$$R_{y/x_1,x_2} = \sqrt{\frac{r_{yx_1}^2 + r_{yx_2}^2 - 2 \cdot r_{yx_1} \cdot r_{yx_2} \cdot r_{x_1x_2}}{1 - r_{x_1x_2}^2}}. \qquad (5.8)$$

Пример 5.1. Дана зависимость между факторами x_1, x_2 и Y (табл. 5.2):

Таблица 5.2

Y	x_1	x_2
3176	2496	209
3066	1962	201
2941	783	177
1997	1319	136
1865	1142	175

Найдя коэффициент множественной корреляции, сделать выводы о форме и тесноте корреляционной связи.

В табл. 5.3 делаем все требуемые расчеты:

Таблица 5.3

i	y	y^2	x_1	x_1^2	$y\cdot x_1$	x_2	x_2^2	$y\cdot x_2$	$x_1\cdot x_2$
1	3176	10086976	2496	6230016	7927296	209	43681	663784	521664
2	3066	9400356	1962	3849444	6015492	201	40401	616266	394362
3	2941	8649481	783	613089	2302803	177	31329	520557	138591
4	1997	3988009	1319	1739761	2634043	136	18496	271592	179384
5	1865	3478225	1142	1304164	2129830	175	30625	326375	199850
Σ	13045	35603047	7702	13736474	21009464	898	164532	2398574	1433851

Рассчитываем среднее значение, дисперсию и среднеквадратическое отклонение признака *Y*:

$$\bar{y} = \frac{\sum_y y}{n} = \frac{13045}{5} = 2609;$$

$$\overline{y^2} = \frac{\sum_y y^2}{n} = \frac{35603047}{5} = 7120609{,}4;$$

$$\sigma_y^2 = \overline{y^2} - \left(\bar{y}\right)^2 = 7120609{,}4 - (2609)^2 = 313728{,}4;$$

$$\sigma_y = \sqrt{\sigma_y^2} = \sqrt{313728{,}4} \approx 560{,}1146.$$

Аналогичные характеристики рассчитаем для признака x_1:

$$\overline{x_1} = \frac{\sum_{x_1} x_1}{n} = \frac{7702}{5} = 1540{,}4;$$

$$\overline{x_1^2} = \frac{\sum_{x_1} x_1^2}{n} = \frac{13736474}{5} = 2747294{,}8$$

$$\sigma_{x_1}^2 = \overline{x_1^2} - \left(\overline{x_1}\right)^2 = 2747294{,}8 - (1540{,}4)^2 = 374462{,}64;$$

$$\sigma_{x_1} = \sqrt{\sigma_{x_1}^2} = \sqrt{374462{,}64} \approx 611{,}9335.$$

Теперь переходим к расчету среднего значения, дисперсии и среднеквадратического отклонения признака x_2:

$$\overline{x_2}=\frac{\sum\limits_{x_2}x_2}{n}=\frac{898}{5}=179{,}6;$$

$$\overline{x_2^2}=\frac{\sum\limits_{x_2}x_2^2}{n}=\frac{164532}{5}=32906{,}4$$

$$\sigma_{x_2}^2=\overline{x_2^2}-\left(\overline{x_2}\right)^2=32906{,}4-(179{,}6)^2=650{,}24;$$

$$\sigma_{x_2}=\sqrt{\sigma_{x_2}^2}=\sqrt{650{,}24}\approx 25{,}4998.$$

Определяем средние значения произведений:

$$\overline{y\cdot x_1}=\frac{\sum\limits_{y}\sum\limits_{x_1}y\cdot x_1}{n}=\frac{21009464}{5}=4201892{,}8;$$

$$\overline{y\cdot x_2}=\frac{\sum\limits_{y}\sum\limits_{x_2}y\cdot x_2}{n}=\frac{2398574}{5}=479714{,}8;$$

$$\overline{x_1\cdot x_2}=\frac{\sum\limits_{x_1}\sum\limits_{x_2}x_1\cdot x_2}{n}=\frac{1433851}{5}=286770{,}2.$$

Теперь, пользуясь (5.6) находим величины парных коэффициентов линейной корреляции r_{yx_1} и r_{yx_2}, а также межфакторный коэффициент линейной корреляции $r_{x_1x_2}$; полученные значения затем возводим в квадрат:

$$r_{yx_1}=\frac{\overline{y\cdot x_1}-\overline{y}\cdot\overline{x_1}}{\sigma_y\cdot\sigma_{x_1}}=\frac{4201892{,}8-2609\cdot 1540{,}4}{560{,}1146\cdot 611{,}9335}\approx 0{,}5339,\ r_{yx_1}^2\approx 0{,}2850;$$

$$r_{yx_2}=\frac{\overline{y\cdot x_2}-\overline{y}\cdot\overline{x_2}}{\sigma_y\cdot\sigma_{x_2}}=\frac{479714{,}8-2609\cdot 179{,}6}{560{,}1146\cdot 25{,}4998}\approx 0{,}7798,\ r_{yx_2}^2\approx 0{,}6081;$$

$$r_{x_1x_2}=\frac{\overline{x_1\cdot x_2}-\overline{x_1}\cdot\overline{x_2}}{\sigma_{x_1}\cdot\sigma_{x_2}}=\frac{286770{,}2-1540{,}4\cdot 179{,}6}{611{,}9335\cdot 25{,}4998}\approx 0{,}6482,\ r_{x_1x_2}^2\approx 0{,}4202.$$

Теперь, подставив в (5.8), находим величину коэффициента множественной корреляции:

$$R_{y/x_1,x_2}=\sqrt{\frac{0{,}2850+0{,}6081-2\cdot 0{,}5339\cdot 0{,}7798\cdot 0{,}6482}{1-0{,}4202}}\approx 0{,}7807.$$

Поскольку

$$R_{y/x_1,x_2}=0{,}7807>0{,}5339=r_{yx_1},\ R_{y/x_1,x_2}=0{,}7807>0{,}7798=r_{yx_2},$$

условия (5.7) выполнены, то есть все основания полагать, что вычисления осуществлены верно.

Делаем выводы о форме и тесноте корреляционной связи: поскольку

$$R_{y/x_1,x_2} = 0{,}7807 > 0{,}7,$$

то связь между рассматриваемыми факторными признаками x_1, x_2 и результативным Y является линейной и тесной.

5.3. Статистическая значимость эмпирических данных

Также как и в случае с парной регрессией, нахождение параметров уравнения многофакторной регрессии предваряет проверка гипотезы о статистической значимости эмпирических данных (оценка качества эмпирических данных). Проверка гипотезы о наличии между факторными признаками x_1, $x_2,\ldots$, X_k и результативным Y статистически существенной зависимости производится с помощью F – критерия Фишера[35].

Правило проверки гипотезы. *Если наблюдаемое значение критерия больше критического,*

$$F_{набл} > F_{крит},$$

то это с доверительной вероятностью γ (уровнем значимости α=1-γ) говорит о наличии между факторными признаками x_1, $x_2,\ldots$, X_k и результативным Y статистически существенной зависимости. Наблюдаемое значение критерия Фишера равно

$$F_{набл} = \frac{\frac{1}{k}\cdot R^2_{y/x_1,x_2,\ldots,x_k}}{\frac{1}{n-k-1}\cdot\left(1-R^2_{y/x_1,x_2,\ldots,x_k}\right)}.$$

Критическое значение рассчитывается по таблице[36] *в зависимости от уровня значимости α и числа степеней свободы ν_1 и ν_2,*

$$F_{крит} = F(\alpha;\nu_1;\nu_2),\ \alpha = 1-\gamma,\ \nu_1 = k,\ \nu_2 = n-k-1.$$

Пример 5.2. С вероятностью 0,95 проверить гипотезу о статистический значимости эмпирических данных из примера 5.1.

Имеем:

$$\gamma = 0{,}95,\ \alpha = 1-0{,}95 = 0{,}05;$$

$$k = 2,\ n = 5,\ \nu_1 = 2,\ \nu_2 = 5-2-1 = 2.$$

---

[35] Фишер Рональд Эймлер (1890 - 1962) – английский статистик, биолог – эволюционист и генетик, один из основателей современной статистики.

[36] Некоторые критические значения критерия Фишера приведены в табл. 5 прил.

$$F_{крит} = F(0{,}05;2;2) = 19{,}00.$$

Наблюдаемое значение критерия Фишера находим, используя результаты примера 5.1:

$$F_{набл} = \frac{\frac{1}{2}\cdot(0{,}7807)^2}{\frac{1}{5-2-1}\cdot\left(1-(0{,}7807)^2\right)} \approx 1{,}5536.$$

Так как

$$1{,}5536 < 19{,}00,$$

то с вероятностью 0,95 утверждаем, что гипотеза о статистической значимости эмпирических данных из примера 5.1 не подтверждается.

Важным показателем качества эмпирических данных является ***множественный коэффициент детерминации****, равный квадрату коэффициента множественной корреляции*,

$$R = R^2_{y/x_1,x_2,\dots,x_k}.$$

Множественный коэффициент детерминации показывает, какая доля вариации результативного признака обусловлена влиянием факторных признаков.

К R можно применить принцип Парето[37]. Согласно ему, значимыми можно считать те факторные признаки, суммарное влияние которых на результативный равно 80%. Признаки, влияющие в сумме на Y в пределах 20% можно считать незначимыми. Это означает, что для наиболее значимых признаков

$$R \geq 0{,}8.$$

Пример 5.3. Найти множественный коэффициент детерминации и сделать соответствующие выводы о степени влияния факторных признаков из примера 5.1 на результативный.

Согласно примеру 5.1 и определению

$$R = R^2_{y/x_1,x_2} = (0{,}7807)^2 \approx 0{,}6095 \text{ или } 60{,}95\%.$$

Значит, имеющиеся факторные признаки оказывают на Y суммарное влияние в пределах 60,95%, что является недостаточным. Имеются и другие факторные признаки, степень влияния которых на результативный равна приблизительно 39,05%.

5.4. Уравнение регрессии

[37] Парето Вильфредо (1848 - 1923) – итальянский инженер, экономист и социолог. Один из основоположников терии элит.

Уравнение многофакторной модели регрессии может быть получено методом наименьших квадратов[38] путем решения системы нормальных уравнений.

Установим вид системы нормальных уравнений для линейного случая (5.1). Функция (4.26) имеет вид:

$$f = f(a_1, a_2, ..., a_k, b) = \sum \Delta^2 = \sum (y^* - y)^2 = \sum \left(\sum_{i=1}^{k} a_i \cdot x_i + b - y \right)^2 .$$

Частная производная по переменной a_j равна: $f'_{a_j} =$

$$= \left(\sum \left(\sum_{i=1}^{k} a_i \cdot x_i + b - y \right)^2 \right)'_{a_j} = \sum \left(\left(\sum_{i=1}^{k} a_i \cdot x_i + b - y \right)^2 \right)'_{a_j} =$$

$$= \sum \left((a_1 \cdot x_1 + a_2 \cdot x_2 + ... + a_{j-1} \cdot x_{j-1} + a_j \cdot x_j + a_{j+1} \cdot x_{j+1} + ... + a_k \cdot x_k + b - y)^2 \right)'_{a_j} =$$
$$= \sum 2x_j \cdot \left(a_1 \cdot x_1 + a_2 \cdot x_2 + ... + a_{j-1} \cdot x_{j-1} + a_j \cdot x_j + a_{j+1} \cdot x_{j+1} + ... + a_k \cdot x_k + b - y \right) =$$
$$= 2\sum x_j \cdot \left(a_1 \cdot x_1 + a_2 \cdot x_2 + ... + a_{j-1} \cdot x_{j-1} + a_j \cdot x_j + a_{j+1} \cdot x_{j+1} + ... + a_k \cdot x_k + b - y \right) =$$
$$= 2\sum \left(a_1 \cdot x_1 \cdot x_j + ... + a_{j-1} \cdot x_{j-1} \cdot x_j + a_j \cdot x_j^2 + a_{j+1} \cdot x_j \cdot x_{j+1} + ... + a_k \cdot x_j \cdot x_k + b \cdot x_j - y \cdot x_j \right) =$$
$$= 2(\sum a_1 \cdot x_1 \cdot x_j + \sum a_2 \cdot x_2 \cdot x_j + ... + \sum a_{j-1} \cdot x_{j-1} \cdot x_j + \sum a_j \cdot x_j^2 + \sum a_{j+1} \cdot x_j \cdot x_{j+1} + ... +$$
$$+ \sum a_k \cdot x_j \cdot x_k + \sum b \cdot x_j - \sum y \cdot x_j) = 2((\sum x_1 \cdot x_j) \cdot a_1 + (\sum x_2 \cdot x_j) \cdot a_2 + ... +$$
$$+ (\sum x_{j-1} \cdot x_j) \cdot a_{j-1} + (\sum x_j^2) \cdot a_j + (\sum x_j \cdot x_{j+1}) \cdot a_{j+1} + ... + (\sum x_j \cdot x_k) \cdot a_k + (\sum x_j) \cdot b -$$
$$- \sum y \cdot x_j) = 0,$$

$$(\sum x_1 \cdot x_j) \cdot a_1 + (\sum x_2 \cdot x_j) \cdot a_2 + ... + (\sum x_{j-1} \cdot x_j) \cdot a_{j-1} + (\sum x_j^2) \cdot a_j + (\sum x_j \cdot x_{j+1}) \cdot a_{j+1} + ... +$$
$$+ (\sum x_j \cdot x_k) \cdot a_k + (\sum x_j) \cdot b = \sum y \cdot x_j, \quad j = 1, 2, ..., k. \tag{5.9}$$

Теперь находим частную производную функции по переменной *b*:

$$f'_b = \left(\sum \left(\sum_{i=1}^{k} a_i \cdot x_i + b - y \right)^2 \right)'_b = \sum \left(\left(\sum_{i=1}^{k} a_i \cdot x_i + b - y \right)^2 \right)'_b =$$

$$= \sum \left((a_1 \cdot x_1 + a_2 \cdot x_2 + ... + a_{j-1} \cdot x_{j-1} + a_j \cdot x_j + a_{j+1} \cdot x_{j+1} + ... + a_k \cdot x_k + b - y)^2 \right)'_b =$$
$$= \sum 2 \cdot \left(a_1 \cdot x_1 + a_2 \cdot x_2 + ... + a_{j-1} \cdot x_{j-1} + a_j \cdot x_j + a_{j+1} \cdot x_{j+1} + ... + a_k \cdot x_k + b - y \right) =$$
$$= 2 \cdot \sum \left(a_1 \cdot x_1 + a_2 \cdot x_2 + ... + a_{j-1} \cdot x_{j-1} + a_j \cdot x_j + a_{j+1} \cdot x_{j+1} + ... + a_k \cdot x_k + b - y \right) =$$
$$= 2 \cdot \left(\sum a_1 \cdot x_1 + \sum a_2 \cdot x_2 + ... + \sum a_j \cdot x_j + ... + \sum a_k \cdot x_k + \sum b - \sum y \right) =,$$
$$= 2 \cdot \left((\sum x_1) \cdot a_1 + (\sum x_2) \cdot a_2 + ... + (\sum x_j) \cdot a_j + ... + (\sum x_k) \cdot a_k + (\sum 1) \cdot b - \sum y \right) = 0,$$

$$(\sum x_1) \cdot a_1 + (\sum x_2) \cdot a_2 + ... + (\sum x_j) \cdot a_j + ... + (\sum x_k) \cdot a_k + n \cdot b = \sum y. \tag{5.10}$$

[38] См. пункт 4.7.

Объединение условий (5.9) и (5.10) определяет систему нормальных уравнений. Для компактности запишем ее в матричном виде:

$$A \times X = B, \tag{5.11}$$

где

$$X = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ \vdots \\ a_k \\ b \end{pmatrix},$$

$$A = \begin{pmatrix} \sum x_1^2 & \sum x_1 \cdot x_2 & \sum x_1 \cdot x_3 & \sum x_1 \cdot x_4 & \dots & \sum x_1 \cdot x_k & \sum x_1 \\ \sum x_1 \cdot x_2 & \sum x_2^2 & \sum x_2 \cdot x_3 & \sum x_2 \cdot x_4 & \dots & \sum x_2 \cdot x_k & \sum x_2 \\ \sum x_1 \cdot x_3 & \sum x_2 \cdot x_3 & \sum x_3^2 & \sum x_3 \cdot x_4 & \dots & \sum x_3 \cdot x_k & \sum x_3 \\ \sum x_1 \cdot x_4 & \sum x_2 \cdot x_4 & \sum x_3 \cdot x_4 & \sum x_4^2 & \dots & \sum x_4 \cdot x_k & \sum x_4 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \sum x_1 \cdot x_k & \sum x_2 \cdot x_k & \sum x_3 \cdot x_k & \sum x_4 \cdot x_k & \dots & \sum x_k^2 & \sum x_k \\ \sum x_1 & \sum x_2 & \sum x_3 & \sum x_4 & \dots & \sum x_k & n \end{pmatrix}, \quad B = \begin{pmatrix} \sum y \cdot x_1 \\ \sum y \cdot x_2 \\ \sum y \cdot x_3 \\ \sum y \cdot x_4 \\ \vdots \\ \sum y \cdot x_k \\ \sum y \end{pmatrix}.$$

В частности, если число факторных признаков $k = 2$, то система нормальных уравнений для нахождения параметров уравнения (5.2) регрессии имеет вид:

$$\begin{cases} \left(\sum x_1^2\right) \cdot a_1 + \left(\sum x_1 \cdot x_2\right) \cdot a_2 + \left(\sum x_1\right) \cdot b = \sum y \cdot x_1 \\ \left(\sum x_1 \cdot x_2\right) \cdot a_1 + \left(\sum x_2^2\right) \cdot a_2 + \left(\sum x_2\right) \cdot b = \sum y \cdot x_2 \\ \left(\sum x_1\right) \cdot a_1 + \left(\sum x_2\right) \cdot a_2 + n \cdot b = \sum y \end{cases}. \tag{5.12}$$

Заметим, что (5.11) можно решать не только путем решения соответствующей системы, но при большом числе факторных признаков уравнение (5.11) даже рациональнее решить методом обратной матрицы.

Пример 5.4. Найти параметры уравнения регрессии по данным примера 5.1.

Составляем систему (5.12) (все соответствующие вычисления были осуществлены в табл. 5.3):

$$\begin{cases} 13736474 \cdot a_1 + 1433851 \cdot a_2 + 7702 \cdot b = 21009464 \\ 1433851 \cdot a_1 + 164532 \cdot a_2 + 898 \cdot b = 2398574 \\ 7702 \cdot a_1 + 898 \cdot a_2 + 5 \cdot b = 13045 \end{cases}.$$

Решая систему методом Крамера, получаем:

$$a_1 \approx 0{,}0448, \; a_2 \approx 16{,}4324, \; b \approx -411{,}3107.$$

Итак, искомое уравнение имеет вид:

$$y = 0{,}0448 \cdot x_1 + 16{,}4324 \cdot x_2 - 411{,}3107. \qquad (5.13)$$

Параметры уравнения (5.2) можно также найти и другим способом, не требующим решения системы (5.12). Он во многом аналогичен выводу уравнения (4.40).

Вначале, разделим правую и левую части каждого из уравнений системы (5.12) на объем выборки *n*. Применим знакомые обозначения:

$$\frac{\sum x_1^2}{n} = \overline{x_1^2}, \; \frac{\sum x_1 \cdot x_2}{n} = \overline{x_1 \cdot x_2}, \; \frac{\sum x_1}{n} = \overline{x_1}, \; \frac{\sum x_2}{n} = \overline{x_2} \; \frac{\sum x_2^2}{n} = \overline{x_2^2}, \; \frac{n}{n} = 1,$$

$$\frac{\sum y \cdot x_1}{n} = \overline{y \cdot x_1}, \; \frac{\sum y \cdot x_2}{n} = \overline{y \cdot x_2}, \; \frac{\sum y}{n} = \overline{y}.$$

Получаем систему:

$$\begin{cases} \overline{x_1^2} \cdot a_1 + \overline{x_1 \cdot x_2} \cdot a_2 + \overline{x_1} \cdot b = \overline{y \cdot x_1} \\ \overline{x_1 \cdot x_2} \cdot a_1 + \overline{x_1^2} \cdot a_2 + \overline{x_2} \cdot b = \overline{y \cdot x_2} \\ \overline{x_1} \cdot a_1 + \overline{x_2} \cdot a_2 + 1 \cdot b = \overline{y} \end{cases}$$

Из третьего уравнения выразим *b*, подставим полученное выражение в первое и второе уравнения:

$$\begin{cases} \overline{x_1^2} \cdot a_1 + \overline{x_1 \cdot x_2} \cdot a_2 + \overline{x_1} \cdot \left(\overline{y} - \overline{x_1} \cdot a_1 - \overline{x_2} \cdot a_2\right) = \overline{y \cdot x_1} \\ \overline{x_1 \cdot x_2} \cdot a_1 + \overline{x_1^2} \cdot a_2 + \overline{x_2} \cdot \left(\overline{y} - \overline{x_1} \cdot a_1 - \overline{x_2} \cdot a_2\right) = \overline{y \cdot x_2}; \\ b = \overline{y} - \overline{x_1} \cdot a_1 - \overline{x_2} \cdot a_2 \end{cases}$$

$$\begin{cases} \left(\overline{x_1^2} - \left(\overline{x_1}\right)^2\right) \cdot a_1 + \left(\overline{x_1 \cdot x_2} - \overline{x_1} \cdot \overline{x_2}\right) \cdot a_2 = \overline{y \cdot x_1} - \overline{y} \cdot \overline{x_1} \\ \left(\overline{x_1 \cdot x_2} - \overline{x_1} \cdot \overline{x_2}\right) \cdot a_1 + \left(\overline{x_2^2} - \left(\overline{x_2}\right)^2\right) \cdot a_2 = \overline{y \cdot x_2} - \overline{y} \cdot \overline{x_2}. \\ b = \overline{y} - \overline{x_1} \cdot a_1 - \overline{x_2} \cdot a_2 \end{cases}$$

Очевидно, что

$$\left(\overline{x_1^2} - \left(\overline{x_1}\right)^2\right) = \sigma_{x_1}^2, \; \left(\overline{x_2^2} - \left(\overline{x_2}\right)^2\right) = \sigma_{x_2}^2;$$

$$\overline{x_1 \cdot x_2} - \overline{x_1} \cdot \overline{x_2} = \mu_{x_1x_2}, \; \overline{y \cdot x_1} - \overline{y} \cdot \overline{x_1} = \mu_{yx_1}, \; \overline{y \cdot x_2} - \overline{y} \cdot \overline{x_2} = \mu_{yx_2}.$$

Тем самым, нами установлен вид системы нормальных уравнений:

$$\begin{cases} \sigma_{x_1}^2 \cdot a_1 + \mu_{x_1x_2} \cdot a_2 = \mu_{yx_1} \\ \mu_{x_1x_2} \cdot a_1 + \sigma_{x_2}^2 \cdot a_2 = \mu_{yx_2}. \\ b = \overline{y} - \overline{x_1} \cdot a_1 - \overline{x_2} \cdot a_2 \end{cases}$$

Решаем ее методом Крамера. Определитель матрицы системы

$$\Delta=\begin{vmatrix}\sigma_{x_1}^2 & \mu_{x_1x_2}\\ \mu_{x_1x_2} & \sigma_{x_2}^2\end{vmatrix}=\sigma_{x_1}^2\cdot\sigma_{x_2}^2-\left(\mu_{x_1x_2}\right)^2=\sigma_{x_1}^2\cdot\sigma_{x_2}^2\cdot\left(1-\frac{\left(\mu_{x_1x_2}\right)^2}{\sigma_{x_1}^2\cdot\sigma_{x_2}^2}\right)=\sigma_{x_1}^2\cdot\sigma_{x_2}^2\cdot\left(1-r_{x_1x_2}^2\right).$$

Определитель

$$\Delta_1=\begin{vmatrix}\mu_{yx_1} & \mu_{x_1x_2}\\ \mu_{yx_2} & \sigma_{x_2}^2\end{vmatrix}=\mu_{yx_1}\cdot\sigma_{x_2}^2-\mu_{x_1x_2}\cdot\mu_{yx_2},$$

а

$$\Delta_2=\begin{vmatrix}\sigma_{x_1}^2 & \mu_{yx_1}\\ \mu_{x_1x_2} & \mu_{yx_2}\end{vmatrix}=\mu_{yx_2}\cdot\sigma_{x_1}^2-\mu_{x_1x_2}\cdot\mu_{yx_1}.$$

Получаем: $a_1=\dfrac{\Delta_1}{\Delta}=\dfrac{\mu_{yx_1}\cdot\sigma_{x_2}^2-\mu_{x_1x_2}\cdot\mu_{yx_2}}{\sigma_{x_1}^2\cdot\sigma_{x_2}^2\cdot\left(1-r_{x_1x_2}^2\right)}=\dfrac{\mu_{yx_1}\cdot\sigma_{x_2}^2-\mu_{x_1x_2}\cdot\mu_{yx_2}}{\sigma_{x_1}^2\cdot\sigma_{x_2}^2\cdot\sigma_y\cdot\left(1-r_{x_1x_2}^2\right)}\cdot\sigma_y=$

$$=\frac{\dfrac{\mu_{yx_1}\cdot\sigma_{x_2}^2}{\sigma_{x_1}^2\cdot\sigma_{x_2}^2\cdot\sigma_y}-\dfrac{\mu_{x_1x_2}\cdot\mu_{yx_2}}{\sigma_{x_1}^2\cdot\sigma_{x_2}^2\cdot\sigma_y}}{\left(1-r_{x_1x_2}^2\right)}\cdot\sigma_y=\frac{\dfrac{\mu_{yx_1}}{\sigma_y\cdot\sigma_{x_1}}\cdot\dfrac{\sigma_{x_2}^2}{\sigma_{x_1}\cdot\sigma_{x_2}^2}-\dfrac{\mu_{x_1x_2}}{\sigma_{x_1}\cdot\sigma_{x_2}}\cdot\dfrac{\mu_{yx_2}}{\sigma_y\cdot\sigma_{x_2}}\cdot\dfrac{1}{\sigma_{x_1}}}{\left(1-r_{x_1x_2}^2\right)}\cdot\sigma_y=$$

$$=\frac{r_{yx_1}\cdot\dfrac{1}{\sigma_{x_1}}-r_{x_1x_2}\cdot r_{yx_2}\cdot\dfrac{1}{\sigma_{x_1}}}{\left(1-r_{x_1x_2}^2\right)}\cdot\sigma_y=\frac{r_{yx_1}-r_{x_1x_2}\cdot r_{yx_2}}{\left(1-r_{x_1x_2}^2\right)}\cdot\frac{\sigma_y}{\sigma_{x_1}};\quad a_2=\frac{\Delta_2}{\Delta}=\frac{\mu_{yx_2}\cdot\sigma_{x_1}^2-\mu_{x_1x_2}\cdot\mu_{yx_1}}{\sigma_{x_1}^2\cdot\sigma_{x_2}^2\cdot\left(1-r_{x_1x_2}^2\right)}=$$

$$=\frac{\mu_{yx_2}\cdot\sigma_{x_1}^2-\mu_{x_1x_2}\cdot\mu_{yx_1}}{\sigma_{x_1}^2\cdot\sigma_{x_2}^2\cdot\sigma_x\cdot\left(1-r_{x_1x_2}^2\right)}\cdot\sigma_x=\frac{\dfrac{\mu_{yx_2}\cdot\sigma_{x_1}^2}{\sigma_{x_1}^2\cdot\sigma_{x_2}^2\cdot\sigma_x}-\dfrac{\mu_{x_1x_2}\cdot\mu_{yx_1}}{\sigma_{x_1}^2\cdot\sigma_{x_2}^2\cdot\sigma_x}}{\left(1-r_{x_1x_2}^2\right)}\cdot\sigma_x=$$

$$=\frac{\dfrac{\mu_{yx_2}}{\sigma_y\cdot\sigma_{x_2}}\cdot\dfrac{\sigma_{x_1}^2}{\sigma_{x_2}\cdot\sigma_{x_1}^2}-\dfrac{\mu_{x_1x_2}}{\sigma_{x_1}\cdot\sigma_{x_2}}\cdot\dfrac{\mu_{yx_1}}{\sigma_y\cdot\sigma_{x_1}}\cdot\dfrac{1}{\sigma_{x_2}}}{\left(1-r_{x_1x_2}^2\right)}\cdot\sigma_x=\frac{r_{yx_2}\cdot\dfrac{1}{\sigma_{x_2}}-r_{x_1x_2}\cdot r_{yx_1}\cdot\dfrac{1}{\sigma_{x_2}}}{\left(1-r_{x_1x_2}^2\right)}\cdot\sigma_x=$$

$$=\frac{r_{yx_2}-r_{x_1x_2}\cdot r_{yx_1}}{\left(1-r_{x_1x_2}^2\right)}\cdot\frac{\sigma_x}{\sigma_{x_2}}.$$

Итак, параметры a_1, a_2 и b уравнения (5.2) находятся по формулам:

$$a_1=\frac{r_{yx_1}-r_{x_1x_2}\cdot r_{yx_2}}{\left(1-r_{x_1x_2}^2\right)}\cdot\frac{\sigma_y}{\sigma_{x_1}},\quad a_2=\frac{r_{yx_2}-r_{x_1x_2}\cdot r_{yx_1}}{\left(1-r_{x_1x_2}^2\right)}\cdot\frac{\sigma_y}{\sigma_{x_2}},\quad b=\overline{y}-\overline{x_1}\cdot a_1-\overline{x_2}\cdot a_2,\qquad(5.14)$$

Пример 5.5. Найти параметры уравнения регрессии для эмпирических данных из примера 5.1 по формулам (5.14).

Все требуемые в (5.14) значения были подсчитаны в примере 5.1. Имеем:

$$a_1=\frac{0{,}5339-0{,}6482\cdot0{,}7798}{(1-0{,}4202)}\cdot\frac{560{,}1146}{611{,}9335}\approx0{,}0449;$$

$$a_2=\frac{0{,}7798-0{,}6482\cdot0{,}5339}{(1-0{,}4202)}\cdot\frac{560{,}1146}{25{,}4998}\approx16{,}4315;$$

$$b = 2609 - 1540{,}4 \cdot 0{,}0449 - 179{,}6 \cdot 16{,}4315 \approx -411{,}2435.$$

Уравнение регрессии имеет вид:

$$y = 0{,}0449 \cdot x_1 + 16{,}4315 \cdot x_2 - 411{,}2435.$$

Заметим, что оно отличается от (5.13). Это произошло из – за ошибок округления. Стоит отметить, что округлений при решении системы нормальных уравнений гораздо меньше, чем в случае (5.13). Поэтому, параметры, найденные с помощью (5.12) более точные. Уравнение (5.13) будем использовать для дальнейших рассуждений.

Есть еще один способ определения параметров уравнения линейной модели. Предположим, что мы располагаем эмпирическими данными о факторных признаках x_1, $x_2, \ldots, X_k$ и результативном Y (табл. 5.1). Подставим значения факторных признаков в (5.1), получим набор теоретических значений y_i^*, $i = 1,2,\ldots,k$, результативного показателя:

$$\begin{cases} y_1^* = a_1 \cdot x_{11} + a_2 \cdot x_{12} + \ldots + a_k \cdot x_{1k} + b \\ y_2^* = a_1 \cdot x_{21} + a_2 \cdot x_{22} + \ldots + a_k \cdot x_{2k} + b \\ \vdots \\ y_n^* = a_1 \cdot x_{n1} + a_2 \cdot x_{n2} + \ldots + a_k \cdot x_{nk} + b \end{cases}. \qquad (5.15)$$

Последнюю систему можно переписать в матричном виде. А именно, если

$$Y^* = \begin{pmatrix} y_1^* \\ y_2^* \\ \vdots \\ y_n^* \end{pmatrix}$$

- матрица – столбец теоретических значений результативного показателя,

$$X = \begin{pmatrix} x_{11} & x_{12} & \ldots & x_{1k} & 1 \\ x_{21} & x_{22} & \ldots & x_{2k} & 1 \\ \vdots & \vdots & \ldots & \vdots & \vdots \\ x_{n1} & x_{n2} & \ldots & x_{nk} & 1 \end{pmatrix}$$

- матрица, составленная из эмпирических значений факторных признаков,

$$A = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_k \\ b \end{pmatrix}$$

- матрица – столбец неизвестных параметров уравнения (5.1), то (5.15) перепишется в матричном виде:

$$Y^* = X \times A. \tag{5.16}$$

Тогда, если

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

- матрица – столбец эмпирических значений результативного фактора, то величина невязки согласно (5.16) равна

$$\Delta = Y - Y^* = Y - X \times A.$$

Применяя соответствующие матричные преобразования, устанавливаем вид квадрата невязки: $\Delta^2 = \Delta^t \times \Delta = (Y - X \times A)^t \times (Y - X \times A) =$
$= \left(Y^t - (X \times A)^t\right) \times (Y - X \times A) = \left(Y^t - A^t \times X^t\right) \times (Y - X \times A) = Y^t \times Y - Y^t \times X \times A - A^t \times X^t \times Y +$
$+ A^t \times X^t \times X \times A = Y^t \times Y - 2 \cdot A^t \times X^t \times Y + A^t \times X^t \times X \times A.$

Также как и в пункте 4.7, нахождение параметров уравнения регрессии сводится к определению минимума функции

$$f(A) = Y^t \times Y - 2 \cdot A^t \times X^t \times Y + A^t \times X^t \times X \times A.$$

Производная

$$f'_A = -2 \cdot X^t \times Y + 2 \cdot X^t \times X \times A.$$

Приравниваем ее к нулю:

$$-2 \cdot X^t \times Y + 2 \cdot X^t \times X \times A = 0,$$
$$X^t \times X \times A = X^t \times Y.$$

Отсюда выражаем A:

$$A = \left(X^t \times X\right)^{-1} \times \left(X^t \times Y\right). \tag{5.17}$$

Итак, с помощью (5.17) можно получить параметры линейной модели (5.1) многофакторной регрессии. Заметим, что этот способ гораздо эффективнее решения системы (5.11) при больших количествах факторных признаков, $k \geq 3$.

Пример 5.6. Найти параметры уравнения регрессии для эмпирических данных из примера 5.1 по формуле (5.17).

Имеем (см. таблицу 5.2):

$$Y = \begin{pmatrix} 3176 \\ 3066 \\ 2941 \\ 1997 \\ 1865 \end{pmatrix}, \; X = \begin{pmatrix} 2496 & 209 & 1 \\ 1962 & 201 & 1 \\ 783 & 177 & 1 \\ 1319 & 136 & 1 \\ 1142 & 175 & 1 \end{pmatrix}, \; X^t = \begin{pmatrix} 2496 & 1962 & 783 & 1319 & 1142 \\ 209 & 201 & 177 & 136 & 175 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix}.$$

Теперь находим произведение матриц[39]

$$X^t \times X = \begin{pmatrix} 2496 & 1962 & 783 & 1319 & 1142 \\ 209 & 201 & 177 & 136 & 175 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix} \times \begin{pmatrix} 2496 & 209 & 1 \\ 1962 & 201 & 1 \\ 783 & 177 & 1 \\ 1319 & 136 & 1 \\ 1142 & 175 & 1 \end{pmatrix} = \begin{pmatrix} 13736474 & 1433851 & 7702 \\ 1433851 & 164532 & 898 \\ 7702 & 898 & 5 \end{pmatrix}.$$

Находим матрицу, обратную к найденной[40]. Определитель

$$\Delta = \det\left(X^t \times X\right) = 17648788603.$$

Ввиду симметричности матрицы $X^t \times X$, алгебраические дополнения равны:

$$A_{11} = \begin{vmatrix} 164532 & 898 \\ 898 & 5 \end{vmatrix} = 16256;$$

$$A_{12} = -\begin{vmatrix} 1433851 & 898 \\ 7702 & 5 \end{vmatrix} = -252859 = A_{21};$$

$$A_{13} = \begin{vmatrix} 1433851 & 164532 \\ 7702 & 898 \end{vmatrix} = 20372734 = A_{31};$$

$$A_{22} = \begin{vmatrix} 13736474 & 7702 \\ 7702 & 5 \end{vmatrix} = 9361566;$$

$$A_{23} = \begin{vmatrix} 13736474 & 1433851 \\ 7702 & 898 \end{vmatrix} = -1291833250 = A_{32};$$

$$A_{33} = \begin{vmatrix} 13736474 & 1433851 \\ 1433851 & 164532 \end{vmatrix} = 204160849967.$$

Отсюда,

$$\left(X^t \times X\right)^{-1} = \frac{1}{\Delta} \cdot \begin{pmatrix} A_{11} & A_{21} & A_{31} \\ A_{12} & A_{22} & A_{32} \\ A_{13} & A_{23} & A_{33} \end{pmatrix} =$$

---

[39] Обязательно проделайте вычисления самостоятельно.

[40] Обязательно проделайте вычисления самостоятельно.

$$
= \frac{1}{17648788603} \cdot \begin{pmatrix} 16256 & -252859 & 20372734 \\ -252859 & 9361566 & -1291833250 \\ 20372734 & -1291833250 & 204160849967 \end{pmatrix} \approx
$$

$$
\approx \begin{pmatrix} 0{,}00000092 & -0{,}00001433 & 0{,}00115434 \\ -0{,}00001433 & 0{,}00053044 & -0{,}07319671 \\ 0{,}00115434 & -0{,}07319671 & 11{,}56798093 \end{pmatrix}.
$$

Матрица

$$
X^t \times Y = \begin{pmatrix} 2496 & 1962 & 783 & 1319 & 1142 \\ 209 & 201 & 177 & 136 & 175 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix} \times \begin{pmatrix} 3176 \\ 3066 \\ 2941 \\ 1997 \\ 1865 \end{pmatrix} = \begin{pmatrix} 21009464 \\ 2398574 \\ 13045 \end{pmatrix}.
$$

Получаем согласно (17.17):

$$
A = \begin{pmatrix} a_1 \\ a_2 \\ b \end{pmatrix} = \begin{pmatrix} 0{,}00000092 & -0{,}00001433 & 0{,}00115434 \\ -0{,}00001433 & 0{,}00053044 & -0{,}07319671 \\ 0{,}00115434 & -0{,}07319671 & 11{,}56798093 \end{pmatrix} \times \begin{pmatrix} 21009464 \\ 2398574 \\ 13045 \end{pmatrix} = \begin{pmatrix} 0{,}04482680 \\ 16{,}43240279 \\ -411{,}31074629 \end{pmatrix}
$$

,

$$
a_1 \approx 0{,}0448,\ a_2 \approx 16{,}4324,\ b \approx -411{,}3107.
$$

$$
y = 0{,}0448 \cdot x_1 + 16{,}4324 \cdot x_2 - 411{,}3107,
$$

что полностью совпадает с (5.13).

Уравнения некоторых нелинейных многофакторных моделей могут быть получены путем приведения к линейному виду. Это процесс называется ***линеаризацией***.

Например, рассмотрим степенную модель (5.3). Прологарифмируем ее:

$$
\ln y = \ln\left(b \cdot x_1^{\alpha_1} \cdot x_2^{\alpha_2} \cdot \ldots \cdot x_k^{\alpha_k}\right) = \ln b + \alpha_1 \cdot \ln x_1 + \alpha_2 \cdot \ln x_2 + \ldots + \alpha_k \cdot \ln x_k .
$$

Получаем вспомогательное уравнение:

$$
y' = a'_1 \cdot x'_1 + a'_2 \cdot x'_2 + \ldots + a'_1 \cdot x'_1 + b',
$$

где

$$
y' = \ln y,\ x'_i = \ln x_i \quad i = 1,2,\ldots,k, \tag{5.18}
$$

$$
a'_i = \alpha_i, \tag{5.19}
$$

$$
b' = \ln b. \tag{5.20}
$$

Найдя параметры вспомогательного уравнения, получим соответствующие параметры и исходного.

Пример 5.7. Найти параметры уравнения регрессии (5.4) для эмпирических данных из примера 5.1.

Вначале, найдем параметры вспомогательного уравнения

$$y' = a_1' \cdot x_1' + a_2' \cdot x_2' + b'.$$

Для этого произведем преобразования эмпирических данных согласно (5.18):

y	$y' = \ln y$	x_1	$x_1' = \ln x_1$	x_2	$x_2' = \ln x_2$
3176	8,0634	2496	7,8224	209	5,3423
3066	8,0281	1962	7,5817	201	5,3033
2941	7,9865	783	6,6631	177	5,1761
1997	7,5994	1319	7,1846	136	4,9127
1865	7,5310	1142	7,0405	175	5,1648

Остальные вычисления аналогичны произведенным в примере 5.6[41]. Имеем:

$$Y = \begin{pmatrix} 8{,}0634 \\ 8{,}0281 \\ 7{,}9865 \\ 7{,}5994 \\ 7{,}5310 \end{pmatrix}, \quad X = \begin{pmatrix} 7{,}8824 & 5{,}3423 & 1 \\ 7{,}5817 & 5{,}3033 & 1 \\ 6{,}6631 & 5{,}1761 & 1 \\ 7{,}1846 & 4{,}9127 & 1 \\ 7{,}0405 & 5{,}1648 & 1 \end{pmatrix},$$

$$X^t = \begin{pmatrix} 7{,}8824 & 7{,}5817 & 6{,}6631 & 7{,}1846 & 7{,}0405 \\ 5{,}3423 & 5{,}3033 & 5{,}1761 & 4{,}9127 & 5{,}1648 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix};$$

$$X^t \times X = \begin{pmatrix} 264{,}2585 & 188{,}1461 & 36{,}2925 \\ 188{,}1461 & 134{,}2673 & 25{,}8992 \\ 36{,}2925 & 25{,}8992 & 5 \end{pmatrix};$$

$$\left(X^t \times X\right)^{-1} = \begin{pmatrix} 1{,}6313 & -2{,}2576 & -0{,}1472 \\ -2{,}2576 & 11{,}9521 & -45{,}5239 \\ -0{,}1472 & -45{,}5239 & 237{,}0753 \end{pmatrix};$$

$$X^t \times Y = \begin{pmatrix} 284{,}7788 \\ 203{,}2215 \\ 39{,}2084 \end{pmatrix};$$

---

[41] Обязательно проделайте их самостоятельно

$$A=\begin{pmatrix}a_1'\\a_2'\\b'\end{pmatrix}=\begin{pmatrix}0{,}0125\\1{,}1120\\1{,}9911\end{pmatrix}.$$

Теперь применим формулы (5.19) и (5.20):

$$\alpha=a_1'=0{,}0125,\ \beta=a_2'=1{,}1120,\ b=e^{b''}=e^{1{,}9911}\approx 7{,}3234.$$

Итак, нами получена двухфакторная показательная модель вида (5.4):

$$y=7{,}3234\cdot x_1^{0{,}0125}\cdot x_2^{1{,}1120}. \qquad (5.21)$$

5.5. Показатели качества регрессионной модели

В данном пункте будут рассмотрены критерии, позволяющие оценить качество регрессионной модели.

Точность модели, также как и в пункте 4.9, определяется с помощью величины средней ошибки аппроксимации

$$\bar{\varepsilon}=\frac{1}{n}\cdot\sum_{x_1,x_2,\ldots,x_k}\left|\frac{y-y^*}{y}\right|\cdot 100\%,$$

где y и y^* - соответственно эмпирическое и теоретическое (рассчитанное по уравнению регрессии) значения результативного фактора, соответствующие данным значениям факторных признаков. Величина $\bar{\varepsilon}$ не должна превышать 15%.

Пример 5.8. Найти среднюю ошибку аппроксимации линейной модели (5.13).

Теоретические значения y^* получаем, подставляя эмпирические данные (табл. 5.2) в уравнение (5.13). Вычисления удобно производить в таблице:

y	x_1	x_2	y^*	$\left\|\frac{y-y^*}{y}\right\|$
3176	2496	209	3134,8817	0,0129
3066	1962	201	2979,5013	0,0282
2941	783	177	2532,3025	0,1390
1997	1319	136	1882,5869	0,0573
1865	1142	175	2515,5209	0,3488
Σ	-	-	-	0,5862

Получаем:

$$\bar{\varepsilon}=\frac{1}{5}\cdot 0{,}5862\cdot 100\%=11{,}724\%, \qquad (5.22)$$

что говорит об удовлетворительной точности построенной модели.

С помощью средней ошибки аппроксимации можно отобрать наиболее точную регрессионную модель: ей соответствует минимальное значение $\bar{\varepsilon}$.

Пример 5.9. Какая из регрессионных моделей является наиболее точной: (5.13) или (5.21)?

Средняя ошибка аппроксимации для линейной модели (5.13) была найдена в примере 5.8. Найдем этот показатель для (5.21). Составляем расчетную таблицу.

y	x_1	x_2	y^*	$\left\lvert\frac{y-y^*}{y}\right\rvert$
3176	2496	209	3070,3012	0,0333
3066	1962	201	2931,0654	0,0440
2941	783	177	2515,5390	0,1447
1997	1319	136	1888,9092	0,0541
1865	1142	175	2495,6972	0,3382
Σ	-	-	-	0,6143

Отсюда, величина средней ошибки аппроксимации равна:

$$\bar{\varepsilon} = \frac{1}{5} \cdot 0{,}6143 \cdot 100\% = 12{,}286\% .$$

Сравнивая полученный результат с (5.22), заключаем: поскольку минимальное значение средней ошибки аппроксимации соответствует линейной модели, то она признается наиболее точной регрессионной моделью для данных примера 5.1.

С помощью расчета ***дельта-коэффициентов*** можно ответить на вопрос: в какой мере каждый из факторных признаков влияет на результативный? Если зависимость является линейной с уравнением (5.1), то для данного факторного признака X_i значение дельта – коэффициента подсчитывается по формуле Δ_i:

$$\Delta_i = a_i \cdot \frac{r_{yx_i}}{R} \cdot \frac{\sigma_{x_i}}{\sigma_y}, \ i = 1,2,...,k .$$

Очевидно, что суммарное влияние факторных признаков, включенных в модель, на результативный, равно 1,

$$\sum_{i=1}^{k} \Delta_i = 1 .$$

Это равенство полезно при проверке правильности произведенных вычислений. Докажем его для случая $k=2$. Имеем согласно определению, (5.14) и (5.8): $\Delta_1+\Delta_1=a_1\cdot\frac{r_{yx_1}}{R}\cdot\frac{\sigma_{x_1}}{\sigma_y}+a_2\cdot\frac{r_{yx_2}}{R}\cdot\frac{\sigma_{x_2}}{\sigma_y}=$

$$=\frac{r_{yx_1}-r_{x_1x_2}\cdot r_{yx_2}}{\left(1-r_{x_1x_2}^2\right)}\cdot\frac{\sigma_y}{\sigma_{x_1}}\cdot\frac{r_{yx_1}}{R}\cdot\frac{\sigma_{x_1}}{\sigma_y}+\frac{r_{yx_2}-r_{x_1x_2}\cdot r_{yx_1}}{\left(1-r_{x_1x_2}^2\right)}\cdot\frac{\sigma_y}{\sigma_{x_2}}\cdot\frac{r_{yx_2}}{R}\cdot\frac{\sigma_{x_2}}{\sigma_y}=\frac{r_{yx_1}\cdot\left(r_{yx_1}-r_{x_1x_2}\cdot r_{yx_2}\right)}{\left(1-r_{x_1x_2}^2\right)}\cdot\frac{1}{R}+$$

$$+\frac{r_{yx_2}\cdot\left(r_{yx_2}-r_{x_1x_2}\cdot r_{yx_1}\right)}{\left(1-r_{x_1x_2}^2\right)}\cdot\frac{1}{R}=\frac{1}{R}\cdot\left(\frac{r_{yx_1}^2-r_{x_1x_2}\cdot r_{yx_1}\cdot r_{yx_2}}{1-r_{x_1x_2}^2}+\frac{r_{yx_2}^2-r_{x_1x_2}\cdot r_{yx_1}\cdot r_{yx_2}}{1-r_{x_1x_2}^2}\right)=$$

$$=\frac{1}{R}\cdot\frac{r_{yx_1}^2+r_{yx_2}^2-2\cdot r_{x_1x_2}\cdot r_{yx_1}\cdot r_{yx_2}}{1-r_{x_1x_2}^2}=\frac{1}{R}\cdot\left(\sqrt{\frac{r_{yx_1}^2+r_{yx_2}^2-2\cdot r_{x_1x_2}\cdot r_{yx_1}\cdot r_{yx_2}}{1-r_{x_1x_2}^2}}\right)^2=\frac{1}{R}\cdot R_{y/x_1,x_2}^2=$$

$$=\frac{1}{R}\cdot R=1.$$

Пример 5.10. Рассчитать дельта – коэффициенты для линейной модели (5.13).

Множественный индекс детерминации найден в примере 5.3,

$$R=0{,}6095,$$

а необходимые вычисления осуществлены в примере 5.1.

Имеем:

$$\Delta_1=a_1\cdot\frac{r_{yx_1}}{R}\cdot\frac{\sigma_{x_1}}{\sigma_y}=0{,}0448\cdot\frac{0{,}5339}{0{,}6095}\cdot\frac{611{,}9335}{560{,}1146}\approx0{,}0429 \text{ или } 4{,}29\%,$$

$$\Delta_2=a_2\cdot\frac{r_{yx_2}}{R}\cdot\frac{\sigma_{x_2}}{\sigma_y}=16{,}4324\cdot\frac{0{,}7798}{0{,}6095}\cdot\frac{25{,}4998}{560{,}1146}\approx0{,}9571 \text{ или } 95{,}71\%.$$

Итак, фактор x_1 влияет на результативный фактор Y в пределах 4,29%, а фактор x_2 - в пределах 95,71%.

Частный коэффициент детерминации показывает, насколько процентов вариация результативного показателя объясняется вариацией факторного признака X_i, входящего в множественное уравнение регрессии. Он рассчитывается по формуле:

$$d_i=a_i\cdot\frac{\sigma_{x_i}}{\sigma_y}\cdot r_{yx_i},\ i=1,2,...,k.$$

Сравнивая формулы дельта – коэффициентов и частных коэффициентов детерминации видим, что

$$d_i=R\cdot\Delta_i, \tag{5.23}$$

а для проверки правильности произведенных вычислений легко получаем, что

$$\sum_{i=1}^{k}d_i=R.$$

Пример 5.11. Рассчитать частные коэффициенты детерминации для линейной модели (5.13).

Применяя (5.23) и результаты примера 5.10, получаем:

$$d_1 = R \cdot \Delta_1 = 0{,}6095 \cdot 0{,}0429 \approx 0{,}0261 \text{ или } 2{,}61\%;$$

$$d_2 = R \cdot \Delta_2 = 0{,}6095 \cdot 0{,}9571 \approx 0{,}5834 \text{ или } 58{,}34\%.$$

Значение d_1 свидетельствует о том, что 2,61% изменения результативного показателя объясняется вариацией признака x_1. На 58,34% изменение признака *Y* объясняется изменением x_2.

Следующий показатель, называемый ***средним коэффициентом эластичности***, позволяет ответить на вопрос: на сколько процентов изменится результат *Y* при изменении фактора X_i на один процент? Для линейной модели он рассчитывается следующим образом:

$$e_i = a_i \cdot \frac{\overline{x_i}}{\overline{y}}, \; i = 1,2,\ldots,k .$$

Пример 5.12. Рассчитать средние коэффициенты эластичности для линейной модели (5.13).

Имеем:

$$e_1 = a_1 \cdot \frac{\overline{x_1}}{\overline{y}} = 0{,}0448 \cdot \frac{1540{,}4}{2609} \approx 0{,}0265 \text{ или } 2{,}65\%;$$

$$e_2 = a_2 \cdot \frac{\overline{x_2}}{\overline{y}} = 16{,}4324 \cdot \frac{179{,}6}{2609} \approx 1{,}1312 \text{ или } 113{,}12\%.$$

Итак, при изменении фактора x_1 на 1% результат изменится на 2,65%, а изменение фактора x_2 на 1% влечет за собой изменение результата на 113,12%.

5.6. Прогноз результативного показателя

После того, как определена многофакторная регрессионная модель, по ней можно осуществить прогноз результативного показателя по данным значениям факторных признаков.

Пример 5.13. По модели (5.13) произвести прогноз результативного показателя, если $x_1 = 3500$, $x_2 = 250$.

Подставляем имеющиеся значения факторных признаков в (5.13), получаем

$$y = 0{,}0448 \cdot 3500 + 16{,}4324 \cdot 250 - 411{,}3107 = 3853{,}589.$$

5.7. Проблема формирования признакового пространства

Прежде чем начать обрабатывать выборочные данные, анализировать их качество, определять тип регрессионной модели и находить ее параметры, необходимо сформировать ***признаковое пространство***. Речь идет о множестве факторных признаков

$$\{X_1; X_2; ...; X_k\}.$$

С одной стороны, их количество не должно быть большим, необходимо исключить из рассмотрения незначимые факторные признаки, оказывающие на результативный минимальное влияние. В противном случае вычисления могут быть чрезвычайно громоздкими. Кроме того, при большом числе факторных признаков очень трудно пользоваться многофакторной моделью.

С другой стороны, сформированная совокупность факторных признаков должна максимально полно описывать результативный фактор, оказывать на него максимальное влияние.

Рассмотрим способы формирования признакового пространства. Они разделены на две группы. Первая группа способов применяется в тех случаях, когда регрессионная модель еще не построена и идет речь о первичном формировании пространства факторных показателей. Вторая группа способов применяется в тех случаях, когда модель уже построена и возникает необходимость понижения размерности, уменьшения числа факторных признаков.

Способ 1 (Модель не построена). Между факторными признаками не должно наблюдать ни корреляционной, ни тем более функциональной зависимости (***принцип отсутствия автокорреляции***). Если принцип отсутствия автокорреляции нарушается для пары факторных признаков X_i и X_j, то один из них следует исключить из рассмотрения. Считается, что между факторными признаками X_i и X_j имеет место автокорреляция, если межфакторный коэффициент линейной корреляции

$$r_{x_i x_j} \geq 0{,}8 .$$

Пример 5.14. Выяснит факт наличия или отсутствия автокорреляции между факторными признаками из примера 5.1.

Поскольку межфакторный коэффициент линейной корреляции

$$r_{x_1 x_2} = 0{,}6482 < 0{,}8 ,$$

значит, между факторными признаками автокорреляция отсутствует.

Способ 2 (Модель не построена). Этот способ основан на анализе парного коэффициента линейной корреляции. Из свойства 4.7 следует, что чем больше по абсолютной величине будет значение парного коэффициента линейной корреляции r_{yx_i}, тем теснее будет связь между факторным признаком X_i и результативным Y. Итак, ***значимость факторного признака*** X_i (степень его влияния на результативный показатель) может быть определена путем проверки гипотезы о статистической значимости соответствующего коэффициента парной корреляции, что производится при помощи t – критерия Стьюдента.

Правило проверки гипотезы. *Если наблюдаемое значение критерия больше критического,*

$$t_{набл} > t_{крит},$$

то это с доверительной вероятностью γ (*уровнем значимости* α=1-γ) *говорит о значимости коэффициента парной корреляции, а следовательно - факторного признака* X_i. *Наблюдаемое значение критерия Стьюдента равно*

$$t_{набл} = \sqrt{\frac{r_{yx_i}^2}{1-r_{yx_i}^2}\cdot(n-2)}.$$

Критическое значение рассчитывается по таблице[42] *в зависимости от уровня значимости* α *и числа степеней свободы* v,

$$t_{крит} = t(\alpha;v),\ \alpha = 1-\gamma\ ,\ v = n-2.$$

Пример 5.15. С вероятностью 0,95 проверить гипотезу о статистической значимости факторных признаков из примера 5.1.

Вначале определим критическое значение критерия Стьюдента. Поскольку $\gamma = 0{,}95$, то $\alpha = 1-0{,}95 = 0{,}05$. Объем выборочной совокупности $n = 5$, следовательно, число степеней свободы $v = 5-2 = 3$. Используя таблицу 5 Приложения, находим критическое значения критерия:

$$t_{крит} = t(0{,}05;3) = 3{,}18.$$

Парные коэффициенты линейной корреляции были найдены в примере 5.1. Наблюдаемые значения соответственно равны:

$$X_1,\ t_{набл} = \sqrt{\frac{r_{yx_1}^2}{1-r_{yx_1}^2}\cdot(n-2)} = \sqrt{\frac{(0{,}5339)^2}{1-(0{,}5339)^2}\cdot(5-2)} \approx 1{,}09;$$

$$X_2,\ t_{набл} = \sqrt{\frac{r_{yx_2}^2}{1-r_{yx_2}^2}\cdot(n-2)} = \sqrt{\frac{(0{,}7798)^2}{1-(0{,}7798)^2}\cdot(5-2)} \approx 2{,}16.$$

---

[42] Некоторые критические значения критерия Стьюдента приведены в табл. 4 прил.

Видим, что в обоих случаях неравенство из правила проверки гипотезы оказалось нарушенным. Влияние факторных признаков на результативный с вероятностью 0,95 признается незначительным.

Способ 3 (Модель не построена). Способ 3 образования признакового пространства применяется в тех случаях, когда изначально число факторных признаков невелико или с помощью критерия Стьюдента невозможно отобрать факторные признаки (как это произошло в примере 5.15). Этот способ целиком основывается на свойстве 4.7 корреляционного отношения. Согласно ему, чем больше абсолютная величина коэффициента парной корреляции r_{yx_i}, тем теснее связь между факторным признаком X_i и результативным Y, а следовательно, тем значимее данный факторный признак X_i. Итак, признаковое пространство формируется из тез факторных признаков, которым соответствуют наибольшие значения (по абсолютной величине) парных коэффициентов линейной корреляции. Соответственно, факторные признаки с минимальными величинами (по абсолютной величине) r_{yx_i} исключаются из рассмотрения.

Пример 5.16. Исключить из рассмотрения один из факторных признаков x_1 или x_2 примера 5.1.

Минимальное значение

$$\min_i r_{yx_i} = \min\left(r_{yx_1}; r_{yx_2}\right) = \min(0{,}5339;07798) = 0{,}5339 = r_{yx_1}.$$

Соответствует признаку x_1. Значит, он исключается из рассмотрения.

Способ 4 (Модель построена). Этот способ основан на применении показателей качества регрессионной модели[43]. Они позволяют определить факторные признаки, включенные в модель, оказывающие наибольшее влияние на результативный показатель. Таковыми признаются показатели, которым соответствуют наибольшие значения дельта – коэффициентов, честных коэффициентов детерминации и средних коэффициентов эластичность. Согласно принципу Парето суммарная степень влияния факторных признаков на результативный должна быть в пределах 80%. Следовательно, для отобранных признаков

$$\sum_i \Delta_i \geq 0{,}8.$$

[43] См. пункт 5.5.

Пример 5.17. Исключить из рассмотрения один из факторных признаков x_1 или x_2 примера 5.1.

Определим минимальные значения дельта – коэффициентов, частных коэффициентов детерминации и средних коэффициентов эластичности:

$$\min_i \Delta_i = \min(\Delta_1; \Delta_2) = \min(0{,}0429; 0{,}9571) = 0{,}0429 = \Delta_1;$$

$$\min_i d_i = \min(d_1; d_2) = \min(0{,}0261; 0{,}5834) = 0{,}0261 = d_1;$$

$$\min_i e_i = \min(e_1; e_2) = \min(0{,}0265; 1{,}1312) = 0{,}0265 = e_1.$$

Видим, что минимальные значения соответствуют признаку x_1. Он подлежит исключению.

После того, как с помощью способа 4 исключен один или несколько факторных признаков, можно перейти к регрессионной модели меньшей размерности.

Пример 5.18. Исключив один из факторных признаков в примере 5.1, перейти к модели с парной регрессией.

Из примеров 5.16 и 5.17 следует, что из рассмотрения исключается признак x_1. Тем самым от многофакторной модели можно перейти к случаю парной корреляции, общий вид уравнения регрессии которой следующий:

$$y = f(x_2).$$

Пользуясь табл. 4.5 и 4.6, а также свойством 4.7 корреляционного отношения, получаем: поскольку парный коэффициент линейной корреляции $r_{yx_2} = 0{,}7798 > 0{,}7$, то связь между признаками x_2 и Y является линейно прямой, тесной.

Уравнение регрессии будем искать в виде (4.40),

$$y - \overline{y} = \rho_{y/x_2} \cdot \left(x - \overline{x_2}\right), \text{ где } \rho_{y/x_2} = r_{yx_2} \cdot \frac{\sigma_y}{\sigma_{x_2}}$$

согласно (4.20). Все необходимые расчеты были произведены в примере 5.1. Имеем:

$$\rho_{y/x_2} = 0{,}7798 \cdot \frac{560{,}1146}{25{,}4998} \approx 17{,}1297;$$

$$y - 2609 = 17{,}1297 \cdot (x - 179{,}6),$$

$$y = 17{,}1297 \cdot x - 467{,}4941.$$

5.8. Алгоритм построения прогноза

Перечислим основные этапы построения многофакторной регрессионной модели и осуществления по ней прогноза.

1. Для всех имеющихся в распоряжении факторных признаков и результативного показателя подсчитываем: среднее значение; дисперсию; среднеквадратическое отклонение. Здесь же находим средние значения произведений признаков.

2. Определяем значения парных и межфакторных коэффициентов линейной корреляции.

3. С помощью способов 1 – 3 пункта 5.7 формируем признаковое пространство.

4. Находим для признакового пространства коэффициент множественной корреляции и множественный коэффициент детерминации.

5. Проверяем полноту признакового пространства. В идеале множественный коэффициент детерминации $R \geq 0{,}8$.

6. С доверительной вероятностью γ проверяем гипотезу о статистической значимости эмпирических данных.

7. По величине коэффициента множественной корреляции делаем вывод о степени линейности и тесноты корреляционной зависимости.

8. Находим параметры уравнения множественной регрессии. Если связь нелинейная, то рекомендуется построить несколько нелинейных многофакторных моделей.

9. Определяем показатели качества многофакторной регрессионной модели, делаем соответствующие выводы. Если построено несколько моделей, то с помощью величины средней ошибки определяем наиболее точную из них.

10. В случае необходимости, понижаем размерность уже имеющегося признакового пространства. Для этого необходимо воспользоваться способом 4 пункта 5.7. Стоит иметь в виду, что для оставшихся факторных признаков должно выполняться неравенство:

$$\sum_i \Delta_i \geq 0{,}8 .$$

11. Производим прогноз факторного признака по включенным в модель факторным.

Пример 5.19. Имеются следующие показатели по десяти предприятиям некоторой отрасли (на 31.12.2019):

Номер предприятия	Стоимость основных промышленно-производственных фондов, тыс. руб.	Валовая продукция предприятия в оптовых ценах, тыс. руб.	Среднесписочная численность промышленно-производственного персонала, чел.	Среднесписочная численность рабочих, чел.
1	4999	5349	420	331
2	6929	6882	553	486
3	6902	7046	570	498
4	10097	7248	883	789
5	8097	5256	433	359
6	11116	14090	839	724
7	4880	3525	933	821
8	7355	5431	526	428
9	10066	7680	676	607
10	7884	8226	684	619

Приняв стоимость основных промышленно-производственных основных фондов за результативный признак, а остальные показатели – за факторные признаки, необходимо получить формулу зависимости результативного показателя от отобранных в модель факторных признаков.

Запишем эмпирические данные (объем выборки n = 10) в виде таблицы:

	Y	X_1	X_2	X_3
1	4999	5349	420	331
2	6929	6882	553	486
3	6902	7046	570	498
4	10097	7248	883	789
5	8097	5256	433	359
6	11116	14090	839	724
7	4880	3525	933	821
8	7355	5431	526	428
9	10066	7680	676	607
10	7884	8226	684	619

1. Все необходимые расчеты осуществлены в табл. 5.4. Под таблицей рассчитаем средние значения, дисперсии (по формуле разностей) и средеквадратические отклонения каждого из признаков.

$$Y\colon\ \overline{y} = \frac{\sum_y y}{n} = \frac{78325}{10} = 7832{,}5,$$

$$\overline{y^2} = \frac{\sum_y y^2}{n} = \frac{653107157}{10} = 65310715{,}3,$$

$$\sigma_y^2 = \overline{y^2} - \left(\overline{y}\right)^2 = 65310715{,}3 - (7832{,}5)^2 = 3962659{,}45,$$

$$\sigma_y = \sqrt{\sigma_y^2} = \sqrt{3962659{,}45} \approx 1990{,}6430.$$

$$X_1\colon\ \overline{x_1} = \frac{\sum_{x_1} x_1}{n} = \frac{70733}{10} = 7073{,}3,$$

$$\overline{x_1^2} = \frac{\sum_{x_1} x_1^2}{n} = \frac{572877843}{10} = 57287784{,}3,$$

$$\sigma_{x_1}^2 = \overline{x_1^2} - \left(\overline{x_1}\right)^2 = 57287784{,}3 - (7073{,}3)^2 = 7256211{,}41,$$

$$\sigma_{x_1} = \sqrt{\sigma_{x_1}^2} = \sqrt{7256211{,}41} \approx 2693{,}7356.$$

$$X_2\colon\ \overline{x_2} = \frac{\sum_{x_2} x_2}{n} = \frac{6517}{10} = 651{,}7,$$

$$\overline{x_2^2} = \frac{\sum_{x_2} x_2^2}{n} = \frac{4550205}{10} = 455020{,}5,$$

$$\sigma_{x_2}^2 = \overline{x_2^2} - \left(\overline{x_2}\right)^2 = 455020{,}5 - (651{,}7)^2 = 30307{,}61,$$

$$\sigma_{x_2} = \sqrt{\sigma_{x_2}^2} = \sqrt{30307{,}61} \approx 174{,}0908.$$

$$X_3\colon\ \overline{x_3} = \frac{\sum_{x_3} x_3}{n} = \frac{5662}{10} = 566{,}2,$$

$$\overline{x_3^2} = \frac{\sum_{x_3} x_3^2}{n} = \frac{3478174}{10} = 347817{,}4,$$

$$\sigma_{x_2}^2 = \overline{x_3^2} - \left(\overline{x_3}\right)^2 = 347817{,}4 - (566{,}2)^2 = 27234{,}96,$$

$$\sigma_{x_3} = \sqrt{\sigma_{x_3}^2} = \sqrt{27264{,}96} \approx 165{,}0302.$$

Таблица 5.4

	y	y^2	x_1	x_1^2	$y \cdot x_1$	x_2	x_2^2	$y \cdot x_2$	x_3	x_3^2	$y \cdot x_3$	$x_1 \cdot x_2$	$x_1 \cdot x_3$	$x_2 \cdot x_3$
1	4999	24990001	5349	28611801	26739651	420	176400	2099580	331	109561	1654669	2246580	1770519	139020
2	6929	48011041	6882	47361924	47685378	553	305809	3831737	486	236196	3367494	3805746	3344652	268758
3	6902	47637604	7046	49646116	48631492	570	324900	3934140	498	248004	3437196	4016220	3508908	283860
4	10097	101949409	7248	52533504	73183056	883	779689	8915651	789	622521	7966533	6399984	5718672	696687
5	8097	65561409	5256	27625536	42557832	433	187489	3506001	359	128881	2906823	2275848	1886904	155447
6	11116	123565456	14090	198528100	156624440	839	703921	9326324	724	524176	8047984	11821510	10201160	607436
7	4880	23814400	3525	12425625	17202000	933	870489	4553040	821	674041	4006480	3288825	2894025	765993
8	7355	54096025	5431	29495761	39945005	526	276676	3868730	428	183184	3147940	2856706	2324468	225128
9	10066	101324356	7680	58982400	77306880	676	456976	6804616	607	368449	6110062	5191680	4661760	410332
10	7884	62157456	8226	67667076	64853784	684	467856	5392656	619	383161	4880196	5626584	5091894	423396
Σ	78325	653107157	70733	572877843	594729518	6517	4550205	52232475	5662	3478174	45525377	47529683	41402962	3976057

2. Вычисляем межфакторные и парные коэффициенты линейной корреляции:

$$r_{yx_1} = \frac{\overline{y \cdot x_1} - \overline{y} \cdot \overline{x_1}}{\sigma_y \cdot \sigma_{x_1}} = \frac{5947295{,}8 - 7832{,}5 \cdot 7073{,}3}{1990{,}6430 \cdot 2693{,}7356} \approx 0{,}7593,$$

$$r^2_{yx_1} = (0{,}7593)^2 \approx 0{,}5665;$$

$$r_{yx_2} = \frac{\overline{y \cdot x_2} - \overline{y} \cdot \overline{x_2}}{\sigma_y \cdot \sigma_{x_2}} = \frac{5223247{,}5 - 7832{,}5 \cdot 651{,}7}{1990{,}6430 \cdot 174{,}0908} \approx 0{,}3428,$$

$$r^2_{yx_2} = (0{,}3428)^2 \approx 0{,}1175;$$

$$r_{yx_3} = \frac{\overline{y \cdot x_3} - \overline{y} \cdot \overline{x_3}}{\sigma_y \cdot \sigma_{x_3}} = \frac{4552537{,}7 - 7832{,}5 \cdot 566{,}2}{1990{,}6430 \cdot 165{,}0302} \approx 0{,}3585,$$

$$r^2_{yx_3} = (0{,}3585)^2 \approx 0{,}1285;$$

$$r_{x_1x_2} = \frac{\overline{x_1 \cdot x_2} - \overline{x_1} \cdot \overline{x_2}}{\sigma_{x_1} \cdot \sigma_{x_2}} = \frac{4752968{,}3 - 7073{,}3 \cdot 651{,}7}{2693{,}7356 \cdot 174{,}0908} \approx 0{,}3058,$$

$$r^2_{x_1x_2} = (0{,}3058)^2 \approx 0{,}0935;$$

$$r_{x_1x_3} = \frac{\overline{x_1 \cdot x_3} - \overline{x_1} \cdot \overline{x_3}}{\sigma_{x_1} \cdot \sigma_{x_3}} = \frac{4140296{,}2 - 7073{,}3 \cdot 566{,}2}{2693{,}7356 \cdot 165{,}0302} \approx 0{,}3046,$$

$$r^2_{X_1X_3} = (0{,}3046)^2 \approx 0{,}0928;$$

$$r_{x_2x_3} = \frac{\overline{x_2 \cdot x_3} - \overline{x_2} \cdot \overline{x_3}}{\sigma_{x_2} \cdot \sigma_{x_3}} = \frac{397605{,}7 - 651{,}7 \cdot 566{,}2}{174{,}0908 \cdot 165{,}0302} \approx 0{,}9959,$$

$$r^2_{x_2x_3} = (0{,}9959)^2 \approx 0{,}9918.$$

3. Займемся формированием признакового пространства.

Принцип отсутствия автокорреляции нарушается для признаков x_2 и X_3, поскольку

$$r_{x_2x_3} = 0{,}9959 > 0{,}8.$$

Один из факторных признаков следует удалить.

С вероятностью 0,95 оценим статистическую значимость каждого из имеющихся факторных признаков. Согласно табл. 4 прил. критическое значение критерия Стьюдента для уровня значимости α = 1 - 0,95 = 0,05 и числа степеней свободы ν =10 – 2 = 8 равно

$$t_{крит} = t(0{,}05;8) = 2{,}31.$$

Вычислим наблюдаемые значения:

$$X_1: \; t_{набл} = \sqrt{\frac{r_{yx_1}^2}{1-r_{yx_1}^2}\cdot(n-2)} = \sqrt{\frac{0,5665}{1-0,5665}\cdot(10-2)} \approx 3,2998;$$

$$X_2: \; t_{набл} = \sqrt{\frac{r_{yx_2}^2}{1-r_{yx_2}^2}\cdot(n-2)} = \sqrt{\frac{0,1175}{1-0,1175}\cdot(10-2)} \approx 1,0322;$$

$$X_3: \; t_{набл} = \sqrt{\frac{r_{yx_3}^2}{1-r_{yx_3}^2}\cdot(n-2)} = \sqrt{\frac{0,1285}{1-0,1285}\cdot(10-2)} \approx 1,0862.$$

Видим, что только для признака X_1 выполняется правило проверки гипотезы. Следовательно, он однозначно включается в модель.

Наибольшие значения парных коэффициентов линейной корреляции соответствуют признакам X_1 и X_3 ($r_{yx_1} = 0,7593, r_{yx_3} = 0,3585$), а минимальное – признаку X_2 ($r_{yx_2} = 0,3428$).

С учетом вышесказанного окончательно получаем: признаковое пространство для рассматриваемого случая имеет вид -

$$\{X_1; X_3\}. \qquad (5.24)$$

4. Коэффициент множественной корреляции равен согласно (5.8):

$$R_{y/x_1,x_3} = \sqrt{\frac{r_{yx_1}^2 + r_{yx_3}^2 - 2\cdot r_{yx_1}\cdot r_{yx_3}\cdot r_{x_1x_3}}{1-r_{x_1x_3}^2}} =$$

$$= \sqrt{\frac{0,5665 + 0,1285 - 2\cdot 0,7593\cdot 0,3585\cdot 0,3046}{1-0,0928}} \approx 0,7709,$$

неравенство (5.7) выполнено.

Множественный коэффициент детерминации

$$R = R_{Y/X_1,X_3}^2 = (0,7709)^2 \approx 0,5943.$$

5. Значение множественного коэффициента детерминации показывает, что факторные признаки из сформированного признакового пространства, влияют на результативный в пределах 59,43%. Это не очень сильное влияние. Величина R в нашем случае меньше требуемых 80%. Признаковое пространство требует включения дополнительных факторных признаков.

6. С вероятностью 0,95 выдвинем гипотезу о статистической значимости эмпирических данных. Поскольку n = 10, k =2, то α=1-0,95 = 0,05 $\nu_1 = 2$, $\nu_2 = 10-2-1 = 7$. Согласно табл. 5 прил.

$$F_{крит} = F(0,05;2;7) = 4,46.$$

Наблюдаемое значение равно

$$F_{набл} = \frac{\frac{1}{2}\cdot(0{,}7709)^2}{\frac{1}{10-2-1}\cdot\left(1-(0{,}7709)^2\right)} \approx 5{,}1275.$$

Правило проверки гипотезы выполнено. Поэтому с вероятностью 0,95 признается наличие статистически существенной зависимости между факторными признаками x_1 и X_3 и результативным Y. Многофакторная регрессионная модель может быть построена.

7. Так как

$$R_{y/x_1,x_3} = 0{,}7709 > 0{,}7,$$

то найденное значение указывает на высокую степень тесноты и линейности корреляционной зависимости.

8. Линейная модель, описывающая корреляционную зависимость, имеет следующий общий вид:

$$y = a_1\cdot x_1 + a_2\cdot x_3 + b.$$

Система (5.11) для нахождения ее параметров имеет вид:

$$\begin{cases} \left(\sum\limits_{x_1} x_1^2\right)\cdot a_1 + \left(\sum\limits_{x_1}\sum\limits_{x_3} x_1\cdot x_3\right)\cdot a_2 + \left(\sum\limits_{x_1} x_1\right)\cdot b = \sum\limits_{y}\sum\limits_{x_1} y\cdot x_1 \\ \left(\sum\limits_{x_1}\sum\limits_{x_3} x_1\cdot x_3\right)\cdot a_1 + \left(\sum\limits_{x_3} x_3^2\right)\cdot a_2 + \left(\sum\limits_{x_3} x_3\right)\cdot b = \sum\limits_{y}\sum\limits_{x_3} y\cdot x_3 \\ \left(\sum\limits_{x_1} x_1\right)\cdot a_1 + \left(\sum\limits_{x_3} x_3\right)\cdot a_2 + n\cdot b = \sum\limits_{y} y \end{cases}.$$

Пользуясь табл. 5.4, получаем систему нормальных уравнений:

$$\begin{cases} 572877843\cdot a_1 + 41402962\cdot a_2 + 70733\cdot b = 594729518 \\ 41402962\cdot a_1 + 3478174\cdot a_2 + 5662\cdot b = 45525377 \\ 70733\cdot a_1 + 5662\cdot a_2 + 10\cdot b = 78325 \end{cases}.$$

Решаем ее:

$$a_1 = 0{,}5295,\ a_2 = 1{,}6921,\ b = 3129{,}0615.$$

Итак, искомое уравнение регрессии имеет вид:

$$y = 0{,}5295\cdot x_1 + 1{,}6921\cdot x_3 + 3129{,}0615. \qquad (5.25)$$

Теперь найдем параметры линейной многофакторной модели по формулам (5.14):

$$a_1 = \frac{r_{yx_1} - r_{x_1x_3}\cdot r_{yx_3}}{1 - r_{x_1x_3}^2}\cdot\frac{\sigma_y}{\sigma_{x_1}} = \frac{0{,}7593 - 0{,}3046\cdot 0{,}3585}{1 - 0{,}0928}\cdot\frac{1990{,}6430}{2693{,}7356} \approx 0{,}5295$$

$$a_2 = \frac{r_{yx_3} - r_{x_1x_3} \cdot r_{yx_1}}{1 - r_{x_1x_3}^2} \cdot \frac{\sigma_y}{\sigma_{x_3}} = \frac{0{,}3585 - 0{,}3046 \cdot 0{,}7593}{1 - 0{,}0928} \cdot \frac{1990{,}6430}{165{,}0302} \approx 1{,}6921,$$

$$b = \overline{y} - a_1 \cdot \overline{x_1} - a_2 \cdot \overline{x_3} = 7835{,}5 - 0{,}5295 \cdot 7073{,}3 - 1{,}6921 \cdot 566{,}2 \approx 3129{,}0615.$$

Получаем тот же результат.

9. Произведем оценку качества регрессионной модели.

Найдем среднюю ошибку аппроксимации. Для этого, подставив значения факторных признаков, соответствующих данному значению y в модель, получаем теоретические значения y^*. Вычисления производим в табл. 5.5.

Итак, значение средней ошибки аппроксимации равно

$$\bar{\varepsilon} = \frac{1}{10} \cdot 1{,}6801 \cdot 100\% = 16{,}801\%,$$

что говорит о низкой точности модели.

Таблица 5.5

y	x_1	x_3	$y - y^*$	$\left\lvert \frac{y - y^*}{y} \right\rvert$
4999	5349	331	6672,0838	0,3347
6929	6882	486	7708,8693	0,1126
6902	7046	498	7824,4743	0,1337
10097	7248	789	8461,0588	0,1620
8097	5256	359	6644,8366	0,1793
11116	14090	724	12009,5096	0,0804
4880	3525	821	6574,3001	0,3472
7355	5431	428	6894,8649	0,0626
10066	7680	607	8339,5446	0,1715
7884	8226	619	8642,1934	0,0962
Σ	-	-	-	1,6801

Определим значения дельта – коэффициентов. Имеем:

$$\Delta_1 = a_1 \cdot \frac{r_{yx_1}}{R} \cdot \frac{\sigma_{x_1}}{\sigma_y} = 0{,}5295 \cdot \frac{0{,}7593}{0{,}5943} \cdot \frac{2693{,}7356}{1990{,}6430} \approx 0{,}9154 \text{ или } 91{,}54\%,$$

$$\Delta_3 = a_2 \cdot \frac{r_{yx_3}}{R} \cdot \frac{\sigma_{x_3}}{\sigma_y} = 1{,}6921 \cdot \frac{0{,}3585}{0{,}5943} \cdot \frac{165{,}0302}{1990{,}6430} \approx 0{,}0846 \text{ или } 8{,}46\%.$$

Сумма дельта – коэффициентов равна 1, следовательно, есть все основания полагать, что вычисления произведены верно. Итак, признак x_1 (валовая продукция предприятия) влияет на признак Y (стоимость промышленно - производственных основных фондов) в пределах 91,54%, а степень влияния признака X_3 (среднесписочная численность работающих) равна 8,46%.

Частные коэффициенты детерминации рассчитаем по формулам (5.23). Имеем:

$$d_1 = R \cdot \Delta_1 = 0{,}5943 \cdot 0{,}9154 \approx 0{,}5440 \text{ или } 54{,}4\%,$$

$$d_3 = R \cdot \Delta_3 = 0{,}5943 \cdot 0{,}0846 \approx 0{,}0503 \text{ или } 5{,}03\%.$$

Значение d_1 свидетельствует о том, что 54,4% изменения результативного показателя объясняется вариацией признака x_1. На 5,03% изменение признака *Y* объясняется изменением X_3.

Найдем величины средних коэффициентов эластичности:

$$e_1 = a_1 \cdot \frac{\overline{x_1}}{\overline{y}} = 0{,}5295 \cdot \frac{7073{,}3}{7835{,}5} \approx 0{,}4782 \text{ или } 47{,}82\%,$$

$$e_3 = a_2 \cdot \frac{\overline{x_3}}{\overline{y}} = 1{,}6921 \cdot \frac{566{,}2}{7835{,}5} \approx 0{,}1223 \text{ или } 12{,}23\%.$$

Таким образом, изменение валовой продукции предприятия на 1% влечет за собой изменение стоимости промышленно – производственных основных фондов на 47,82%, а вследствие изменения численности рабочих на 1%, изменение стоимости промышленно – производственных основных фондов 12,23%.

10. Понизим размерность признакового пространства (5.24). Для этого, исключим из (5.24) один из факторных признаков. Поскольку одновременно минимум дельта – коэффициента и среднего коэффициента эластичности соответствует признаку X_3,

$$\min_i \Delta_i = \min(\Delta_1; \Delta_3) = \min(0{,}9154; 0{,}0846) = 0{,}0846 = \Delta_3,$$

$$\min_i d_i = \min(d_1; d_3) = \min(0{,}5440; 0{,}0503) = 0{,}0503 = d_3,$$

$$\min_i e_i = \min(e_1; e_3) = \min(0{,}4782; 0{,}1223) = 0{,}1223 = e_3,$$

то он исключается из (5.24). Заметим, что степень влиянии оставшегося признака отвечает условию $\sum_i \Delta_i \geq 0{,}8$, так как $\Delta_1 = 0{,}9154$.

Итак, от множественной регрессии мы перешли к парной, общий вид уравнения которой есть

$$y = f(x_1).$$

Так как $r_{yx_1} = 0{,}7593 > 0{,}7$, то связь признается линейной, тесной, прямой. Уравнение прямой линии регрессии найдем по формулам (16.40):

$$\rho_{y/x_1} = r_{yx_1} \cdot \frac{\sigma_{x_1}}{\sigma_y} = 0{,}7593 \cdot \frac{2693{,}7356}{1990{,}6430} \approx 1{,}0275;$$

$$y - \bar{y} = \rho_{y/x_1} \cdot (x_1 - \overline{x_1});$$

$$y - 7835{,}5 = 1{,}0275 \cdot (x_1 - 7073{,}3);$$

$$y = 1{,}0275 \cdot x_1 + 567{,}6842. \qquad (5.26)$$

11. Произведем прогноз стоимости промышленно производственных основных фондов при валовом выпуске продукции в 8 млн. руб. и при численности рабочих 600 человек.

Если прогноз производить по модели (5.25), то

$$y = 0{,}5295 \cdot 8000 + 1{,}6921 \cdot 600 + 3129{,}0615 = 8380{,}322 \text{ (тыс.руб.)}.$$

Если же прогнозировать значение результативного показателя по модели (5.26), то

$$y = 1{,}0275 \cdot 8000 + 567{,}6842 = 8787{,}684 \text{ (тыс. руб.)}.$$

§ 6. РЯДЫ ДИНАМИКИ

6.1. Основные определения и классификация рядов динамики

Исследование динамики социально – экономических процессов производится с помощью рядов динамики. Рядом динамики называется такой способ записи случайной величины (признака, фактора) Y, при котором ее значения (уровни) $y = y_i = y(t) = y(t_i)$ приводятся в зависимости от времени $t = t_i$. Обычно ряды динамики записываются в виде таблицы (табл. 6.1)

Таблица 6.1

t	t_1	t_2	$\cdots$	t_n
Y	y_1	y_2	$\cdots$	y_n

Уровни ряда динамики могут быть приедены на определенный момент времени (например, на начало года, на конец месяца и т. д.). Также их значения приводятся на интервал времени (декада, квартал, год и т. д.). В первом случае ряд динамики называется моментным, а во втором – интервальным. Различают также ряды динамики с равноотстоящими и неравноотстоящими по времени уровнями.

Пример 6.1. Имеются данные (табл. 6.2) о количестве выпускаемой продукции предприятием в 2010 – 2018 годах (тыс. ед.). Определить тир ряда динамики.

Таблица 6.2

Годы	2010	2011	2012	2013	2014	2015	2016	2017	2018
Количество выпускаемой продукции, тыс. ед.	23,4	25,0	27,4	28,1	28,5	29,3	30,0	29,6	29,4

Так как значения уровней ряда динамики (объем выпускаемой продукции) приводятся не на конкретную дату, а на интервал времени (год), а временная разница между уровнями постоянна, то его можно считать интервальным с равноотстоящими по времени уровнями.

Пример 6.2. Дан объем продаж (тыс. руб.) торгующей организации в 2014 – 2019 годах (табл. 6.3); данные приведены на конец года. Определить тип ряда динамики.

Таблица 6.3

Годы	2014	2015	2016	2017	2018	2019
Объем продаж, тыс. руб.	80900	75600	82500	83800	85000	86700

Ряд динамики является интервальным с равноотстоящими по времени уровнями, так как есть указание, что данные приведены на конец года (то есть на 31 декабря – конкретную дату).

Пример 6.3. В табл. 6.4. приведены сведения о площади складских помещений (тыс кв. м.) производственного предприятия в 2013 – 2019 годах. Определить тип ряда динамики. Определить тип ряда динамики.

Таблица 6.4

Годы	2013	2015	2016	2019
Площадь складских помещений, тыс. кв. м.	47,4	50,5	54,8	66,0

Ряд динамики – интервальный. Действительно, его уровни соответствуют интервалам времени (год). Временная разность между уровнями не является постоянным числом. Между первым и вторым уровнем она составляет два года, между вторым и третьим – один год, между третьим и четвертым – три года. Окончательно, мы рассматриваемый интервальный ряд динамики с не равноотстоящими по времени уровнями.

Пример 6.4. Пусть мы располагаем данными о числе работающих на предприятии (чел.) в 2019 году, приведенными на начало месяца (табл. 6.5). Определить тип ряда динамики.

Таблица 6.5

Месяцы	Январь 2018 г.	Март	Апрель	Июнь	Июль	Сентябрь	Декабрь	Январь 2019 г.
Число работающих, чел.	784	761	766	1014	1106	1262	1591	1600

Этот ряд динамики является моментным (значения уровней приведены на определенную дату – начало месяца) с неравноотстоящими уровнями (показатели не следуют друг за другом через равные промежутки времени).

Помимо табл. 6.1, эмпирические данные могут быть представлены графически – в виде точек с координатами $(t_i; y_i)$ на координатной плоскости *tOy*.

Пример 6.5. В табл. 6.6. представлены данные о прибылях от продаж торгующей организации в 2019 году (тыс. руб.).

Таблица 6.6

Месяцы	Прибыль, тыс. руб.
Январь	374,6
Февраль	245,5
Март	304,6
Апрель	171,1
Май	210,8
Июнь	321,3
Июль	244,7
Август	345,6
Сентябрь	495,4
Октябрь	523,2
Ноябрь	385,3
Декабрь	274,2

Представить эмпирические данные в виде табл. 6.1 и графически.

Рассматриваемый ряд динамики является интервальным с равноотстоящими уровнями[75]. Присваивая каждому месяцу свой порядковый номер, получаем представление данных в виде табл. 6.1 (табл. 6.7).

Таблица 6.7

t_i	1	2	3	4	5	6	7	8	9	10	11	12
y_i	374,6	245,5	304,6	171,1	210,8	321,3	244,7	345,6	495,4	523,2	385,3	274,2

[75] Подумайте, почему.

Используя данные табл. 6.7, представляем эмпирические данные в виде точек (рис. 6.1).

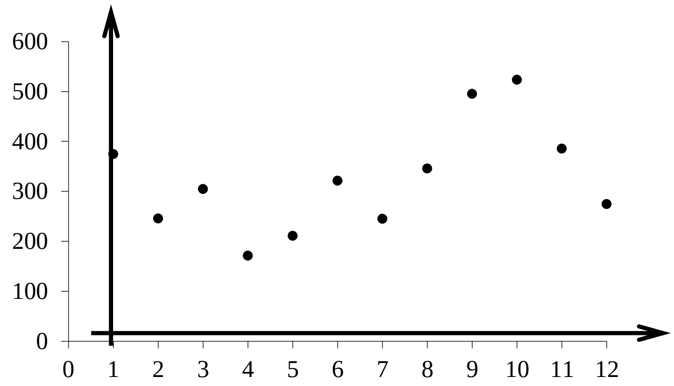

Рис. 6.1

6.2. Сравнение уровней ряда динамики

Сравнение уровней ряда динамики производится двумя способами: цепным и базисным. При первом способе сравнения каждый уровень y_i сравнивается с предыдущим ему уровнем y_{i-1}. При втором способе выбирается базисный уровень y_0 (он не обязательно является первым уровнем, более того, базисный уровень может не принадлежать ряду динамики) и все уровни сравниваются с ним.

Рассмотрим основные числовые характеристики, позволяющие сравнивать уровни ряда динамики.

Для определения абсолютной величины скорости изменения значений уровней ряда динамики используется абсолютный прирост Δ_i, который в зависимости от способа сравнения (цепной или базисный) равен разности между значением данного уровня и предыдущим ему или базисным:

$$\Delta_i^{Ц} = y_i - y_{i-1},\ \Delta_i^{Б} = y_i - y_0 .$$

Абсолютный прирост отвечает на вопрос: на сколько единиц изменился данный уровень ряда динамики по сравнению с предыдущим или базисным? Например, количество выпускаемой продукции в примере 6.1 в 2013 году увеличилось по сравнению с 2012 годом на

$$\Delta_4^{Ц} = y_4 - y_3 = 28{,}1 - 27{,}4 = 0{,}7 \text{ (тыс. ед.).}$$

В примере 6.1. возьмем в качестве базисного показатели 2010 года. Получаем: по сравнению с 2010 годом объем выпускаемой продукции в 2018 году увеличился на

$$\Delta_9^Б = y_9 - y_0 = 29{,}4 - 23{,}4 = 6{,}0 \text{ (тыс. ед.)}$$

Относительная скорость роста (снижения) уровней определяется с помощью коэффициента роста K_i, который определяется как отношение значения данного уровня к предыдущему ему (цепной способ сравнения) или базисному (базисный способ сравнения):

$$K_i^Ц = \frac{y_i}{y_{i-1}}, \; K_i^Б = \frac{y_i}{y_0}.$$

Так, коэффициент роста объемов продаж в примере 6.2. составил

$$K_3^Ц = \frac{y_3}{y_2} = \frac{82500}{75600} \approx 1{,}0913.$$

Это значит, что в 2016 году по сравнению с 2015 годом объем продаж торгующей организации в 1,0913. В примере 6.2. за базисный уровень примем объем продаж 2014 года. Коэффициент роста

$$K_1^Б = \frac{y_2}{y_0} = \frac{75600}{80900} \approx 0{,}9345.$$

Это значит, что объем продаж в 2015 снизился и составил 0,9345 от показателей 2014 года.

Коэффициент роста, выраженный в процентах, называется темпом роста,

$$T = K \cdot 100\%.$$

Этот показатель может быть рассчитан как цепным, так и базисным способом.

Используя предыдущие рассуждения получаем:

$$T_3^Ц = K_3^Ц \cdot 100\% = 1{,}0913 \cdot 100\% = 109{,}13\%, \; T_1^Б = K_1^Б \cdot 100\% = 0{,}9345 \cdot 100\% = 93{,}45\%.$$

Получаем: объем продаж торгующей организации в 2016 году по сравнению с 2015 годом составил 109,13%, а в 2015 году – лишь 93,45% от уровня 2014 года.

С помощью определения темпов прироста ($T_{пр,i}$) можно узнать: на сколько процентов изменится данный уровень ряда динамики по сравнению с предыдущим или базисным. В зависимости от способа сравнения темп прироста определяется как отношение абсолютного прироста данного уровня к предыдущему или базисному уровням, выраженное в процентах:

$$T_{пр,i}^Ц = \frac{\Delta_i^Ц}{y_{i-1}} \cdot 100\%, \; T_{пр,i}^Б = \frac{\Delta_i^Б}{y_0} \cdot 100\%.$$

Нетрудно доказать, что темп прироста равен темпу роста, уменьшенному на 100%,

$$T_{пр} = T - 100\%.$$

Рассуждения проведем, используя формулу расчетов цепных темпов прироста. Имеем:

$$T^{Ц}_{пр,i} = \frac{\Delta^{Ц}_{i}}{y_{i-1}} \cdot 100\% = \frac{y_i - y_{i-1}}{y_{i-1}} \cdot 100\% = \left(\frac{y_i}{y_{i-1}} - \frac{y_{i-1}}{y_{i-1}}\right) \cdot 100\% = \left(K^{Ц}_{i} - 1\right) \cdot 100\% = K^{Ц}_{i} \cdot 100\% - 100\% =$$

$= T^{Ц}_{i} - 100\%$. Формула доказана. Эту упрощенную формулу расчета темпов прироста будем использовать в дальнейших вычислениях. Так, следуя примеру 6.2 и предыдущим вычислениям, получаем:

$$T^{Ц}_{пр,3} = T^{Ц}_{3} - 100\% = 109{,}13\% - 100\% = 9{,}13\%,$$

$$T^{Б}_{пр,1} = T^{Б}_{1} - 100\% = 93{,}45\% - 100\% = -6{,}55\%.$$

Таким образом, объем продаж торгующей организации в 2016 году по сравнению с 2015 годом увеличился на 9,13%, а в 2015 году по сравнению с 2014 годом – сократился на 6,55%.

Показатель одного процента прироста $|\%|$ определяется как результат деления абсолютного прироста к темпам прироста. Расчет этого показателя имеет экономический смысл только на цепной основе. Следовательно,

$$|\%|_{i} = \frac{\Delta^{Ц}_{i}}{T^{Ц}_{пр,i}}.$$

Данный показатель применяется в тех случаях, когда абсолютную скорость изменения уровней ряда динамики надо перевести в относительную и наоборот. Действительно,

$$T^{Ц}_{пр,i} = \frac{\Delta^{Ц}_{i}}{|\%|_{i}}, \quad \Delta^{Ц}_{i} = |\%|_{i} \cdot T^{Ц}_{пр,i}.$$

Абсолютное значение одного процента прироста можно рассчитать по - другому. Для этого преобразуем исходную формулу: $|\%|_{i} = \frac{\Delta^{Ц}_{i}}{T^{Ц}_{пр,i}} = |\%|_{i} = \frac{\Delta^{Ц}_{i}}{\frac{\Delta^{Ц}_{i}}{y_{i-1}} \cdot 100} = \frac{y_{i-1}}{100}$. Итак, абсолютное значение одного процента прироста может быть найдено как результат деления предыдущего данному уровня на 100,

$$|\%|_{i} = 0{,}01 \cdot y_{i-1}.$$

Пример 6.6. Повести анализ уровней ряда динамики из примера 6.5 цепным и базисным способами. За базисный уровень принять прибыль предприятия, полученную им в январе 2019 года.

Все расчеты производим в табл. 6.8. Обязательно проделайте их самостоятельно. Заметим, что при цепном способе сравнения у первого уровня нет предыдущего, расчет невозможен. Поэтому, в соответствующих клетках таблицы мы ставим прочерк.

Таблица 6.8

Месяцы	Прибыль, тыс. руб.	Абсолютный прирост, тыс. руб.		Коэффициент роста		Темп роста, %		Темп прироста, %		Абсолютное значение одного процента прироста, тыс. руб.
		С пред. месяцем	С январем 2019 г.	С пред. месяцем	С январем 2019 г.	С пред. месяцем	С январем 2019 г.	С пред. Месяцем	С январем 2019 г.	
Январь	374,6	-	0	-	1	-	100	-	0	-
Февраль	245,5	-129,1	-129,1	0,6554	0,6554	65,54	65,54	-34,46	-34,46	3,746
Март	304,6	59,1	-70,0	1,2407	0,8131	124,07	81,31	24,07	-18,69	2,455
Апрель	171,1	-133,5	-203,5	0,5617	0,4568	56,17	45,68	-43,83	-54,32	3,046
Май	210,8	39,7	-163,80	1,2320	0,5627	123,20	56,27	23,20	-43,73	1,711
Июнь	321,3	110,5	-53,3	1,5242	0,8577	152,42	85,77	52,42	-14,23	2,108
Июль	244,7	-76,6	-129,9	0,7616	0,6532	76,16	65,32	-23,84	-34,68	3,213
Август	345,6	100,9	-29,0	1,4123	0,9226	141,23	92,26	41,23	-7,74	2,447
Сентябрь	495,4	149,8	120,8	1,4334	1,3225	143,34	132,25	43,34	32,25	3,456
Октябрь	523,2	27,8	148,6	1,0561	1,3967	105,61	139,67	5,61	39,67	4,954
Ноябрь	385,3	-137,9	10,7	0,7364	1,0286	73,64	102,86	-26,36	2,86	5,232
Декабрь	274,2	-111,1	-100,4	0,7117	0,7320	71,17	73,20	-28,83	-26,80	3,853

Заметим, что при анализе уровней ряда динамики сначала необходимо выбрать способ сравнения. Сравнение и цепным и базисным способом не производится. Выбор способа сравнения зависит от множества параметров: цели анализа, характера показателей, желания показать успехи или скрыть недостатки в работе организации и т. д. Также может быть исчислен только один показатель (например, абсолютный прирост). Также отметим, что все характеристики можно изображать графически в виде ломаных, абсциссы точек которых – время, а ординаты – соответствующие числовые характеристики уровней ряда динамики.

Пример 6.7. Построить ломаные, показывающие темпы прироста объемов продаж торгующей организации (данные см. пример 6.5).

Строим по соответствующим точкам ломаные (рис. 6.2). Точки, соответствующие цепным темпам прироста, соединены сплошной линией, пунктирной линией соединены точки, соответствующие базисным темпам прироста. Обратим внимание на различный характер этих ломаных.

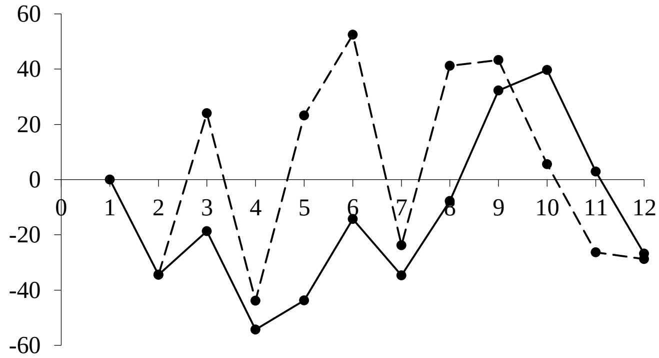

Рис. 6.2

6.3. Средние значения уровней ряда динамики и его числовых характеристик

Формула расчета среднего значения уровней ряда динамики $\overline{y}$ зависит от его вида. Среднее значение уровней моментного ряда динамики называется средней хронологической.

Среднее значение интервального ряда динамики с равноотстоящими уровнями подсчитывается по формуле средней арифметической простой:

$$\overline{y} = \frac{\sum_i y_i}{n}.$$

Так, в примере 2.1 среднегодовой выпуск продукции в 2010 – 2018 годах составил

$$\overline{y} = \frac{23{,}4 + 25{,}0 + 27{,}4 + 28{,}1 + 28{,}5 + 29{,}3 + 30{,}0 + 29{,}6 + 29{,}4}{9} \approx 27{,}856 \text{ (тыс. ед.).}$$

Среднее значение интервального ряда динамики с не равноотстоящими по времени уровнями подсчитывается по формуле средней арифметической взвешенной:

$$\overline{y} = \frac{\sum_i y_i \cdot t_i}{\sum_i t_i},$$

при этом t_i - временная разность между данным уровнем y_i и следующим за ним y_{i+1}. Условимся, что для последнего уровня значение временной разности будет всегда равно 1,

$$t_n = 1.$$

Например, среднегодовая площадь складских помещений в 2013 – 2019 годах (пример 6.3) равна

$$\bar{y} = \frac{23{,}4 \cdot 2 + 50{,}5 \cdot 1 + 54{,}8 \cdot 3 + 66{,}0 \cdot 1}{2 + 1 + 3 + 1} = \frac{375{,}7}{7} \approx 53{,}671 \text{ (тыс. кв. м.).}$$

Для ряда с равностоящими уровнями средняя хронологическая равна:

$$\bar{y} = \frac{\frac{1}{2} \cdot y_1 + y_2 + \ldots + y_{n-1} + \frac{1}{2} \cdot y_n}{n-1}.$$

В примере 6.2 среднегодовой объем продаж в 2014 – 2019 составил

$$\bar{y} = \frac{\frac{1}{2} \cdot 80900 + 75600 + 82500 + 83800 + 85000 + \frac{1}{2} \cdot 86700}{6-1} = \frac{410700}{5} = 82140 \text{ (тыс. руб.).}$$

Средняя хронологическая для ряда динамики с не равноотстоящими уровнями подсчитывается следующим образом:

$$\bar{y} = \frac{\frac{y_1 + y_2}{2} \cdot t_1 + \frac{y_2 + y_3}{2} \cdot t_2 + \ldots + \frac{y_{n-1} + y_n}{2} \cdot t_{n-1}}{t_1 + t_2 + \ldots + t_{n-1}},$$

t_i - временная разность между уровнями y_i и y_{i+1}. Используя эту формулу, получаем величину среднемесячной численности работающих на предприятии в 2019 году:

$$\bar{y} = \left(\frac{784 + 761}{2} \cdot 2 + \frac{761 + 766}{2} \cdot 1 + \frac{766 + 1014}{2} \cdot 2 + \frac{1014 + 1106}{2} \cdot 1 + \frac{1106 + 1262}{2} \cdot 2 + \right.$$
$$\left. + \frac{1262 + 1591}{2} \cdot 3 + \frac{1591 + 1600}{2} \cdot 1\right) : (2 + 1 + 2 + 1 + 2 + 3 + 1) = \frac{26783}{12} \approx 1116 \text{ (чел.).}$$

Средний абсолютный прирост рассчитывается как средняя арифметическая цепных абсолютных приростов:

$$\bar{\Delta} = \frac{\sum_i \Delta_i^{Ц}}{n-1}.$$

Его можно рассчитать по другому: $\bar{\Delta} = \frac{\sum_i \Delta_i^{Ц}}{n-1} = \frac{\Delta_2^{Ц} + \Delta_3^{Ц} + \ldots + \Delta_n^{Ц}}{n-1} =$

$$= \frac{(y_2 - y_1) + (y_3 - y_2) + \ldots + (y_n - y_{n-1})}{n-1} = \frac{y_n - y_1}{n-1},$$

$$\bar{\Delta} = \frac{y_n - y_1}{n-1}.$$

Средний коэффициент роста ряда динамики подсчитывается как средняя геометрическая его цепных коэффициентов роста,

$$\bar{K} = \sqrt[n-1]{\prod_i K_i^{Ц}}.$$

Получим упрощенную формулу расчета среднего коэффициента роста: $\overline{K} = \sqrt[n-1]{\prod_i K_i^Ц} =$

$$= \sqrt[n-1]{K_2^Ц \cdot K_3^Ц \cdot \ldots \cdot K_n^Ц} = \sqrt[n-1]{\frac{y_2}{y_1} \cdot \frac{y_3}{y_2} \cdot \ldots \cdot \frac{y_n}{y_{n-1}}} = \sqrt[n-1]{\frac{y_n}{y_1}},$$

$$\overline{K} = \sqrt[n-1]{\frac{y_n}{y_1}}.$$

Средний темп роста равен среднему коэффициенту роста, выраженному в процентах,

$$\overline{T} = \overline{K} \cdot 100\%.$$

Соответственно, средний темп прироста равен:

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

Пример 6.8. Определить средние значения и показатели вариации ряда динамики из примера 6.5.

Так как ряд динамики – интервальный с равноотстоящими уровнями, то его средний уровень может быть подсчитан по формуле:

$$\overline{y} = \frac{374{,}6 + 245{,}5 + 304{,}6 + 171{,}1 + 210{,}8 + 321{,}3 + 244{,}7 + 345{,}6 + 495{,}4 + 523{,}2 + 385{,}3 + 274{,}2}{12} =$$

$$= \frac{3896{,}3}{12} \approx 324{,}692 \text{ (тыс. руб.).}$$

Итак, среднемесячная прибыль торгующей организации в 2019 году составила 324,692 тысяч рублей.

Средний абсолютный прирост

$$\overline{\Delta} = \frac{274{,}2 - 374{,}6}{12 - 1} \approx -9{,}127 \text{ (тыс. руб.).}$$

Итак, в 2019 году ежемесячное сокращение прибыли составило в среднем 9,127 тыс. руб.

Средний коэффициент роста:

$$\overline{K} = \sqrt[12-1]{\frac{274{,}2}{374{,}6}} \approx 0{,}9720.$$

Средний темп роста:

$$\overline{T} = 0{,}9720 \cdot 100\% = 97{,}2\%.$$

Средний темп прироста:

$$\overline{T}_{пр} = 97{,}2\% - 100\% = -2{,}8\%.$$

Получаем: месячная прибыль торгующей организации в 2019 году составляла 0,9720 от показателей предыдущего месяца или 97,2%. Следовательно прибыли ежемесячно сокращалась в среднем на 2,8%.

6.4. Аналитическая модель ряда динамики

Основополагающей задачей исследования рядов динамики является построение прогноза их уровней в интересующий нас момент (интервал) времени. Прогноз, произведенный в будущее, называется перспективным. Иногда возникает необходимость теоретической оценки эмпирических уровней ряда динамики. В этом случае мы имеем дело с ретроспективным прогнозом.

Мощным методом прогнозирования является построение моделирование уровней ряда динамики в виде уравнения

$$y = y(t).$$

Чтобы установить его общий вид, определим факторы, оказывающие влияние на формирование значений уровней ряда динамики.

1. Долговременные факторы, формирующие общую, в длительной перспективе тенденцию развития признака. Результат действия этих факторов моделируется в виде неслучайной монотонной функции тренда $f_1(t)$. На рис. 6.3 представлен линейный возрастающий тренд.

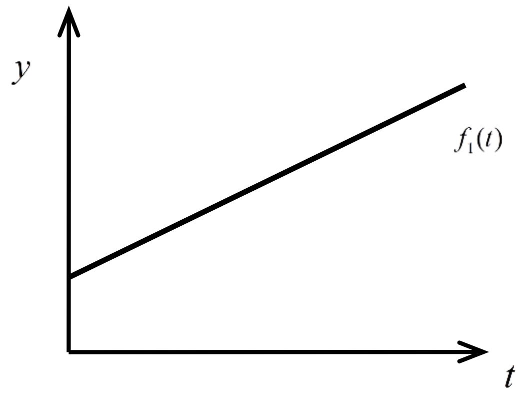

Рис. 6.3

2. Сезонные факторы, формирующие периодические колебания исследуемого признака в определенные моменты / интервалы времени (сезоны), причем колебание признака в данный сезон можно считать постоянной величиной. Результат действия этих факторов моделируется с помощью неслучайной периодической, как правило тригонометрической, функции $f_2(t)$. На рис. 6.4. пунктирной линией изображена линия тренда, а сплошной – бегущая по ней сезонная волна.

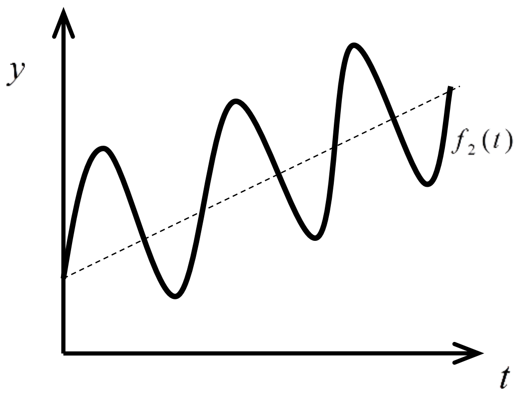

Рис. 6.4

3. Циклические факторы, формирующие периодические колебания за достаточно большой период времени (свыше пяти лет) и имеющие экономическую, социальную или астрофизическую природу. Результат действия этих факторов моделируется с помощью неслучайной периодической функции $f_3(t)$. На рис. 6.5. показана корректировка модели ряда динамики с учетом влияния циклических факторов. На нем штрих – пунктирной линией показана линия тренда, пунктирном – циклическая волна, а сплошной – результирующее действие и долговременных и циклических и сезонных факторов.

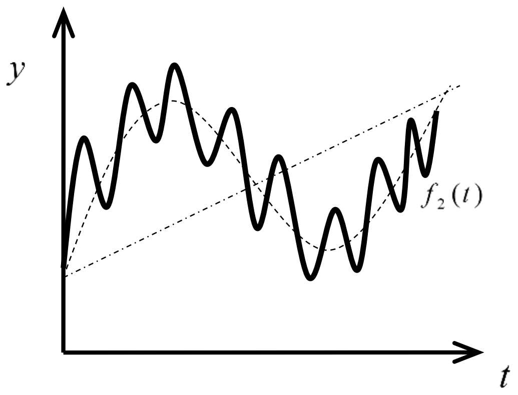

Рис. 6.5

4. Кроме неслучайных, на формирование значений уровней ряда динамики оказывают влияние случайные факторы, не поддающиеся учету и регистрации. Случайные факторы моделируются с помощью случайной величины $\varepsilon(t)$. На рис. 6.6 точечным пунктиром показана линия тренда, штрих – пунктирной линией – влияние циклической составляющей, пунктирной – сезонная волна. Сплошной линией пока-

зано влияние трех вышеперечисленных неслучайных факторов, скорректированное с учетом действия на них случайной составляющей.

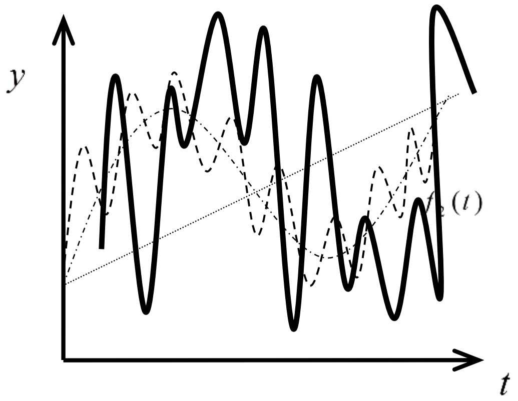

Рис. 6.6

Исходя из вышесказанного можно сделать вывод, что модель ряда динамики является аддитивной и представляет собой композицию неслучайной и случайной составляющих:

$$y = f(t) + \varepsilon(t)\,. \tag{6.1}$$

В свою очередь, неслучайную составляющую можно представить в виде мультипликативной модели:

$$f(t) = f_1(t)\cdot f_2(t)\cdot f_3(t)\,. \tag{6.2}$$

Итак, прежде чем произвести прогноз по модели (6.1), необходимо

1. выяснить, есть ли исследуемом ряде динамики неслучайная составляющая (если таковой нет, то ряд динамики представляет собой случайный набор чисел, моделирование и прогнозирование на их основе невозможно);
2. определить, какая из неслучайных составляющих (долговременная, циклическая, сезонная) входит в (6.2);
3. для каждой из неслучайных составляющих, входящих в (6.2), подобрать функцию, наиболее точно и адекватно описывающую эмпирические данные;
4. среди нескольких моделей неслучайной составляющей отобрать наиболее точную;
5. по модели, найденной в п. 4 произвести точечный прогноз уровня ряда динамики;
6. проверить гипотезу о наличии в ряде динамики случайной составляющей;

7. определить качество случайной составляющей;

8. если показатели качества неслучайной составляющей (п.7) признаются удовлетворительными, то точечный прогноз ряда динамики (п.5) уточняется с помощью интервального прогноза.

Объем учебника не позволяет рассмотреть все пункты алгоритма построения прогноза. Мы остановимся на следующих вопросах:

1. рассмотрим подходы к построению функции тренда;

2. определим простейшие способы моделирования влияния сезонных факторов;

3. получим модели неслучайной составляющей (без учета влияния циклических факторов);

4. введем показатели точности модели неслучайной составляющей;

5. определим формулы получения точечного и интервального прогнозов уровней ряда динамики.

Полученный нами прогноз может быть использован для оперативных прогнозов (длина прогнозного периода не превышает трех лет, что вполне достаточно для оперативного анализа и принятия управленческих решений).

6.5. Функция тренда

Перечисли модели тренда, часто встречающиеся в экономической практике:

1. линейная модель:

$$f_1(t) = a \cdot t + b;$$

2. параболическая модель:

$$f_1(t) = a \cdot t^2 + b \cdot t + c;$$

3. показательная модель:

$$f_1(t) = b \cdot a^t.$$

Наиболее эффективным инструментом моделирования долговременных факторов является корреляционно – регрессионный анализ (параграф 4). Параметры уравнения функции тренда могут быть найдены методом наименьших квадратов путем решения соответствующей системы нормальных уравнений. Приведем вид системы нормальных уравнений для линейного, параболического и показательного случаев.

1. Линейная модель:

$$\begin{cases} \left(\sum_t t^2\right)\cdot a + \left(\sum_t t\right)\cdot b = \sum_t\sum_y t\cdot y \\ \left(\sum_t t\right)\cdot a + n\cdot b = \sum_y y \end{cases}.$$

2. Параболическая модель:

$$\begin{cases} \left(\sum_t t^4\right)\cdot a + \left(\sum_t t^3\right)\cdot b + \left(\sum_t t^2\right)\cdot c = \sum_t\sum_y t^2\cdot y \\ \left(\sum_t t^3\right)\cdot a + \left(\sum_t t^2\right)\cdot b + \left(\sum_t t\right)\cdot c = \sum_t\sum_y t\cdot y \\ \left(\sum_t t^2\right)\cdot a + \left(\sum_t t\right)\cdot b + n\cdot c = \sum_y y \end{cases}.$$

3. Показательная модель:

$$\begin{cases} \left(\sum_t t^2\right)\cdot \ln a + \left(\sum_t t\right)\cdot \ln b = \sum_t\sum_y t\cdot \ln y \\ \left(\sum_t t\right)\cdot \ln a + n\cdot \ln b = \sum_y \ln y \end{cases}.$$

Пример 6.9. По данным примера 6.5 построить линейную, параболическую и показательную модель функции тренда. Результаты представить графически.

Все необходимые расчеты производим в табл. 6.9.
Для каждой функции тренда составляем систему нормальных уравнений, решаем ее, получаем вид каждой из моделей.

1. Линейная модель. Система нормальных уравнений:

$$\begin{cases} 650\cdot a + 78\cdot b = 27142{,}6 \\ 78\cdot a + 12\cdot b = 3896{,}3 \end{cases}, \begin{cases} a \approx 12{,}7038 \\ b \approx 242{,}1167 \end{cases}.$$

Модель функции тренда:

$$f_1(t) = 12{,}7038\cdot t + 242{,}167. \tag{6.3}$$

2. Параболическая модель. Система нормальных уравнений:

$$\begin{cases} 60710\cdot a + 6084\cdot b + 650\cdot c = 236334{,}6 \\ 6084\cdot a + 650\cdot b + 78\cdot c = 27142{,}6 \\ 650\cdot a + 78\cdot b + 12\cdot c = 3896{,}3 \end{cases}, \begin{cases} a \approx 1{,}2502 \\ b \approx -3{,}5484 \\ c \approx 280{,}0386 \end{cases}.$$

Модель функции тренда:

$$f_1(t) = 1{,}2502\cdot t^2 - 3{,}5484\cdot t + 280{,}0386. \tag{6.4}$$

3. Показательная модель. Система нормальных уравнений:

$$\begin{cases} 650\cdot \ln a + 78\cdot \ln b = 452{,}6190 \\ 78\cdot \ln a + 12\cdot \ln b = 68{,}7923 \end{cases}, \begin{cases} a \approx 1{,}0390 \\ b \approx 240{,}8326 \end{cases}.$$

Модель функции тренда:

$$f_1(t) = 240{,}8326 \cdot (1{,}0390)^t \, . \tag{6.5}$$

Таблица 6.9

	t	t^2	$t^3{}_{t^2}$	t^4	y	$\ln y$	$t \cdot y$	$t^2 \cdot y$	$t \cdot \ln y$
1	1	1	1	1	374,6	5,9259	374,6	374,6	5,9259
2	2	4	8	16	245,5	5,5033	491,0	982,0	11,0066
3	3	9	27	81	304,6	5,7190	913,8	2741,4	17,1570
4	4	16	64	256	171,1	5,1422	684,4	2737,6	20,5688
5	5	25	125	625	210,8	5,3509	1054,0	5270,0	26,7545
6	6	36	216	1296	321,3	5,7724	1927,8	11566,8	34,6344
7	7	49	343	2401	244,7	5,5000	1712,9	11990,3	38,5000
8	8	64	512	4096	345,6	5,8453	2764,8	22118,4	46,7624
9	9	81	729	6561	495,4	6,2054	4458,6	40127,4	55,8486
10	10	100	1000	10000	523,2	6,2600	5232,0	52320,0	62,6000
11	11	121	1331	14641	385,3	5,9540	4238,3	46621,3	65,4940
12	12	144	1728	201736	274,2	5,6139	3290,4	39484,8	67,3668
$\sum$	78	650	6084	60710	3896,3	68,7923	27142,6	236334,6	452,6190

Для построения графика найдем теоретические значения $f_1(t)$ уровней ряда, соответствующие эмпирическим уровням (ретроспективный прогноз). Для этого значение t подставляем в правую часть каждого из уравнений. Результаты расчетов помещаем в табл. 6.10.

Таблица 6.10

t	y	$f_1(t)$		
		Линейная модель (6.3)	Параболическая модель (6.4)	Показательная модель (6.5)
1	374,6	254,8205	277,7404	250,2251
2	245,5	267,5243	277,9425	259,9838
3	304,6	280,2281	280,6449	270,1232
4	171,1	292,9319	285,8477	280,6580
5	210,8	305,6357	293,5509	291,6037
6	321,3	318,3395	303,7544	302,9762
7	244,7	331,0433	316,4582	314,7923
8	345,6	343,7471	331,6624	327,0692
9	495,4	356,4509	349,3670	339,8249
10	523,2	369,1547	369,5719	353,0781
11	385,3	381,8585	392,2771	366,8481
12	274,2	394,5623	417,4827	381,1552

Строим чертеж. Линейный тренд построен на рис. 6.7, параболический – на рис. 6.8, а показательный – на рис. 6.9.

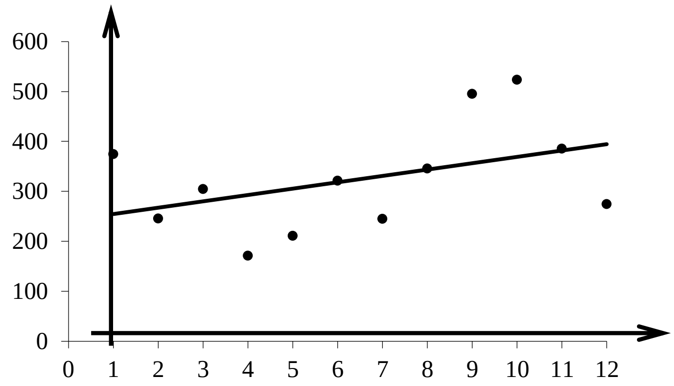

Рис. 6.7

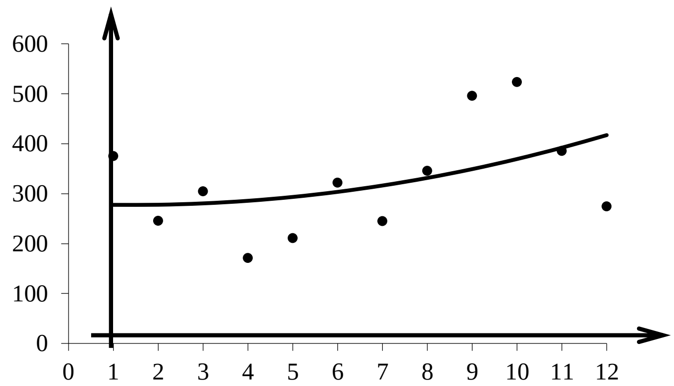

Рис. 6.8

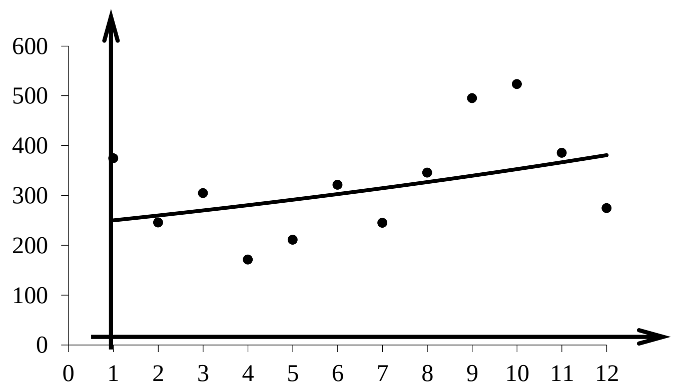

Рис. 6.9

Кроме решения систем нормальных уравнений, вид функции тренда устанавливается с помощью средних показателей числовых характеристик ряда динамики.

А именно, тренд может быть представлен в виде линейной модели, в которой использован средний абсолютный прирост,

$$f_1(t) = f_1(n+\tau) = \overline{\Delta} \cdot \tau + y_n. \tag{6.6}$$

Также тренд может быть смоделирован в показательной модели, в которой применен средний коэффициент роста,

$$f_1(t) = f_1(n+\tau) = y_n \cdot \left(\overline{K}\right)^{\tau}. \tag{6.7}$$

В моделях (6.6) и (6.7) величина называется периодом упреждения.

Пример 6.10. Построить модель функции тренда для ряда динамики из примера 6.5 в виде (6.6) и (6.7). Результаты представить графически.

Используя расчеты из примера 6.8, получаем:

$$f_1(t) = f_1(12+\tau) = -9{,}127 \cdot \tau + 274{,}2; \tag{6.8}$$

$$f_1(t) = f_1(12+\tau) = 274{,}2 \cdot \left(0{,}9720\right)^{\tau}. \tag{6.9}$$

Следуя формулам (6.8) и (6.9) находим теоретические значения уровней ряда динамики (табл. 6.11).

Таблица 6.11

t	τ	y	$f_1(t)$	
			Модель (6.8)	Модель (6.9)
1	-11	374,6	374,597	374,7473
2	-10	245,5	365,470	364,2544
3	-9	304,6	356,343	354,0553
4	-8	171,1	347,216	344,1417
5	-7	210,8	338,089	334,5058
6	-6	321,3	328,962	325,1396
7	-5	244,7	319,835	316,0357
8	-4	345,6	310,708	307,1867
9	-3	495,4	301,581	298,5855
10	-2	523,2	292,454	290,2251
11	-1	385,3	283,327	282,0988
12	0	274,2	274,200	274,2000

Модели (6.8) и (6.9) графически представлены на рис. 6.10 и 6.11 соответственно.

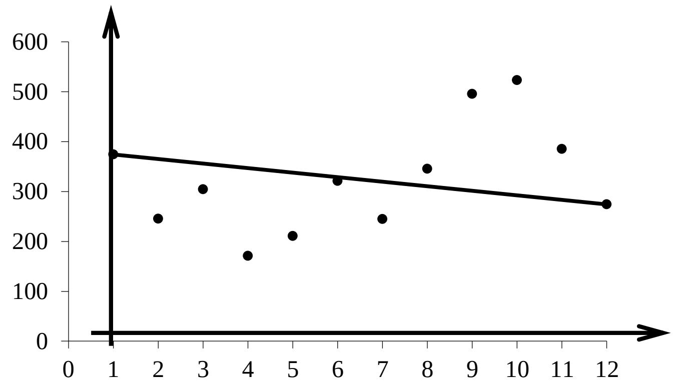

Рис. 6.10

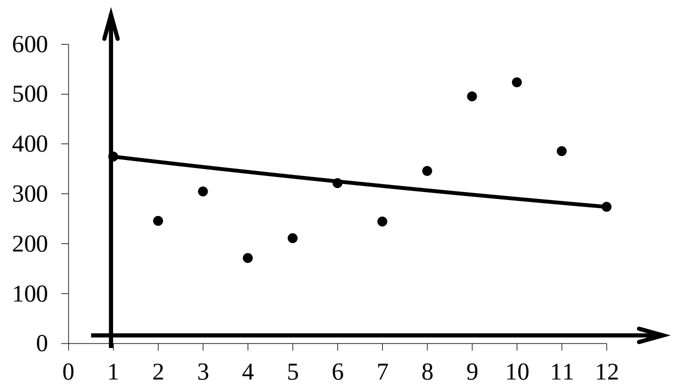

Рис. 6.11

6.6. Индексы сезонности

Одним из простейших способов учета сезонных факторов в модели ряда динамики является расчет индексов сезонности:

$$I_s = \frac{y_s}{\overline{y}},$$

где y_s – значение уровня ряда динамики в данный момент/интервал времени (сезон), $\overline{y}$ - среднее значение уровней ряда динамики.

Пример 6.11. Построить индексы сезонности для уровней ряда динамики из примера 6.5.

Среднее значение уровней было найдено в примере 6.8:

$$\overline{y} = 324{,}692.$$

Разделив значение уровня ряда динамики $y = y_s$ на $\overline{y}$, получаем индексы сезонности (табл. 6.12).

Таблица 6.12

t	$y = y_s$	I_s
1	374,6	1,1537
2	245,5	0,7561
3	304,6	0,9381
4	171,1	0,5270
5	210,8	0,6492
6	321,3	0,9896
7	244,7	0,7536
8	345,6	1,0644
9	495,4	1,5258
10	523,2	1,6114
11	385,3	1,1867
12	274,2	0,8445

Заметим, что расчет индексов сезонности по одному периоду времени (одному кварталу, одному году и т. д.) на практике по возможности не производится. Индексы сезонности, рассчитываются за $k \geq 3$ периодов времени:

$$I_s = I_{s,k} = \frac{y_{s,k}}{\overline{y}_k},$$

при этом

$$y_{s,k} = \frac{\sum_k y_s}{k}$$

- среднее значение уровней ряда динамики по каждому сезону,

$$\overline{y}_k = \frac{\sum_k \overline{y}}{k}$$

- среднее значение уровней ряда динамики за k периодов. Это необходимо для сглаживания влияния случайных факторов на значения уровней ряда динамики.

Пример 6.12. Предположим, что мы располагаем значениями прибылей торгующей организации за 2017 – 2019 г.г. (тыс. руб.). Данные приведены в табл. 6.13.

Таблица 6.13

Месяцы	Прибыль, тыс. руб.		
	2017 г.	2018 г.	2019 г.
Январь	102,6	262,3	374,6
Февраль	105,8	301,3	245,5
Март	150,3	315,7	304,6
Апрель	215,7	200,8	171,1
Май	222,6	195,6	210,8
Июнь	228,4	220,7	321,3
Июль	199,6	233,4	244,7
Август	172,8	300,2	345,6
Сентябрь	170,4	350,5	495,4
Октябрь	230,4	406,8	523,2
Ноябрь	244,6	350,9	385,3
Декабрь	250,8	360,1	274,2

Найти индексы сезонности за указанные три года.

Все необходимые вычисления производим в табл. 6.14. Обязательно проделайте их самостоятельно. Для сравнения, в последней колонке табл. 6.14 найдены значения индексов сезонности, рассчитанные по одному 2019 году (см. пример 6.11).

Таблица 6.14

t	$y = y_s$			$y_{s,k}$	$I_{s,k}$	I_s
	2017	2018	2019			
1	102,6	262,3	374,6	246,5000	0,9159	1,1537
2	105,8	301,3	245,5	217,5333	0,8083	0,7561
3	150,3	315,7	304,6	256,8667	0,9544	0,9381
4	215,7	200,8	171,1	195,8667	0,7278	0,5270
5	222,6	195,6	210,8	209,6667	0,7791	0,6492
6	228,4	220,7	321,3	256,8000	0,9542	0,9896
7	199,6	233,4	244,7	225,9000	0,8394	0,7536
8	172,8	300,2	345,6	272,8667	1,0139	1,0644
9	170,4	350,5	495,4	338,7667	1,2588	1,5258
10	230,4	406,8	523,2	386,8000	1,4372	1,6114
11	244,6	350,9	385,3	326,9333	1,2148	1,1867
12	250,8	360,1	274,2	295,0333	1,0963	0,8445
$\sum$	2294,0	3498,3	3896,3			
$\bar{y}$	191,1667	291,5250	324,6917	$\bar{y}_k =$ 269,1278		

6.7. Модель неслучайной составляющей

Рассмотрим способы моделирования неслучайной составляющей. Напомним, что мы рассматриваем только долговременные и сезонные факторы.

При первом способе неслучайная составляющая представляется в виде произведения индекса сезонности на функцию тренда:

$$f(t) = I_s \cdot f_1(t). \quad (6.10)$$

Пример 6.13. Построить модель неслучайной составляющей ряда динамики из примера 6.5. в виде (6.10). Результаты представим графически.

Индексы сезонности были найдены в примере 6.11, а вид функции тренда установлен в примерах 6.9 и 6.10. Имеем:

1. функция тренда – (6.3):

$$f(t) = I_s \cdot (12{,}7038 \cdot t + 242{,}167); \quad (6.11)$$

2. функция тренда – (6.4):

$$f(t) = I_s \cdot (1{,}2502 \cdot t^2 - 3{,}5484 \cdot t + 280{,}0386); \quad (6.12)$$

3. функция тренда – (6.5):

$$f(t) = I_s \cdot \left(240{,}8326 \cdot (1{,}0390)^t\right); \quad (6.13)$$

4. функция тренда – (6.8):

$$f(t) = I_s \cdot f_1(12+\tau) = I_s \cdot (-9{,}127 \cdot \tau + 274{,}2); \quad (6.14)$$

5. функция тренда – (6.9):

$$f(t) = I_s \cdot f_1(12+\tau) = I_s \cdot \left(274{,}2 \cdot (0{,}9720)^\tau\right). \quad (6.15)$$

Для построения графиков по формулам (6.11) – (6.15) находим теоретические значения уровней ряда динамики – ретроспективные прогнозы. Индексы сезонности представлены в табл. 6.12, а значения функции тренда – в табл. 6.10 и 6.11. Все расчеты производим в табл. 6.15. Обязательно проделайте их самостоятельно.

По найденным теоретическим значениям неслучайной составляющей строим графики. Они представлены на рис. 6.12 – 6.16. На одном чертеже показаны эмпирические данные, линии тренда и ломаные $f(t)$. Заметим, насколько ближе значения $f(t)$ к теоретическим показателям. Это говорит о том, что точность $f(t)$ значительно точнее модели функции тренда $f_1(t)$. Это значит, что прогноз, произведенный по модели неслучайной составляющей будет значительно выше, чем прогноз, произведенный по модели функции тренда. Аналитические вопросы оценки точности модели будут обсуждаться в следующем пункте.

Таблица 6.15

t	y	I_s	Модель (6.11)		Модель (6.12)		Модель (6.13)		Модель (6.14)		Модель (6.15)	
			$f_1(t)$	$f(t)$	$f_1(t)$	$f(t)$	$f_1(t)$	$f(t)$	$f_1(t)$	$f(t)$	$f_1(t)$	$f(t)$
1	374,6	1,1537	254,8205	293,9886	277,7404	320,4315	250,2251	288,6869	374,597	432,1758	374,7473	432,3492
2	245,5	0,7561	267,5243	202,2754	277,9425	210,1526	259,9838	196,5741	365,470	276,3323	364,2544	275,4132
3	304,6	0,9381	280,2281	262,8875	280,6449	263,2785	270,1232	253,4079	356,343	334,2924	354,0553	332,1463
4	171,1	0,5270	292,9319	154,3637	285,8477	150,6306	280,6580	147,8958	347,216	182,9693	344,1417	181,3492
5	210,8	0,6492	305,6357	198,4281	293,5509	190,5822	291,6037	189,3181	338,089	219,4977	334,5058	217,1714
6	321,3	0,9896	318,3395	315,0139	303,7544	300,5811	302,9762	299,8111	328,962	325,5254	325,1396	321,7429
7	244,7	0,7536	331,0433	249,4866	316,4582	238,4947	314,7923	237,2392	319,835	241,0396	316,0357	238,1763
8	345,6	1,0644	343,7471	365,8821	331,6624	353,0192	327,0692	348,1303	310,708	330,7155	307,1867	326,9675
9	495,4	1,5258	356,4509	543,8563	349,3670	533,0480	339,8249	518,4891	301,581	460,1383	298,5855	455,5679
10	523,2	1,6114	369,1547	594,8460	369,5719	595,5183	353,0781	568,9406	292,454	471,2525	290,2251	467,6610
11	385,3	1,1867	381,8585	453,1374	392,2771	465,5007	366,8481	435,3251	283,327	336,2137	282,0988	334,7562
12	274,2	0,8445	394,5623	333,2050	417,4827	352,5611	381,1552	321,8828	274,200	231,5599	274,2000	231,5599

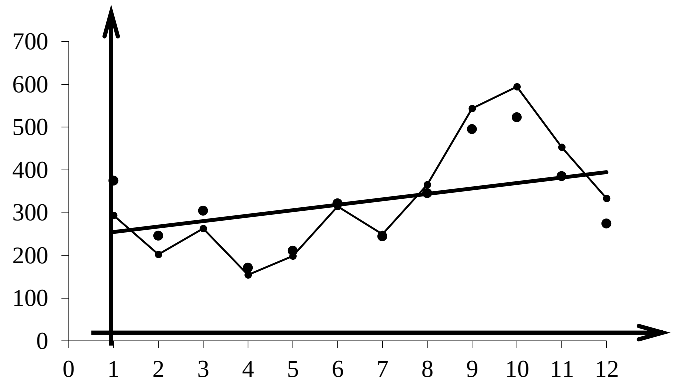

Рис. 6.12

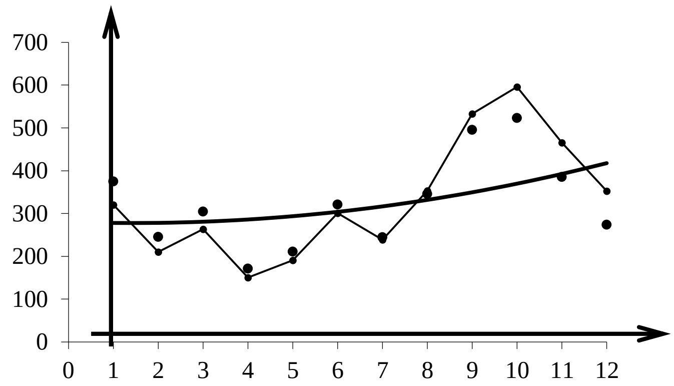

Рис. 6.13

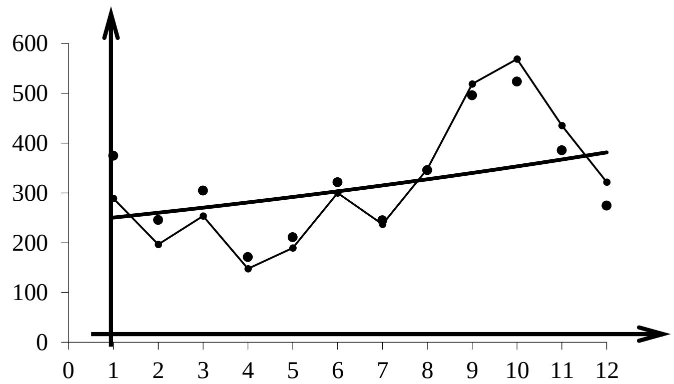

Рис. 6.14

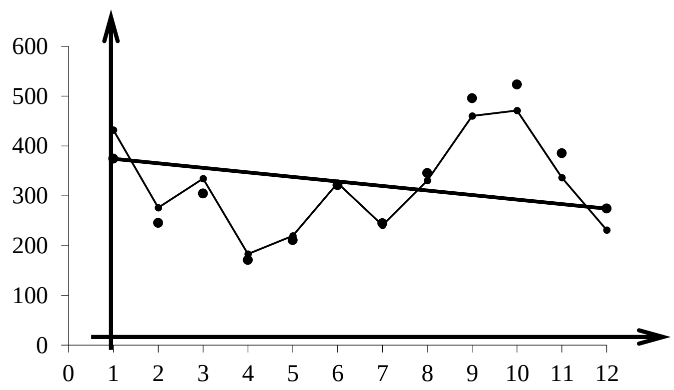

Рис. 6.15

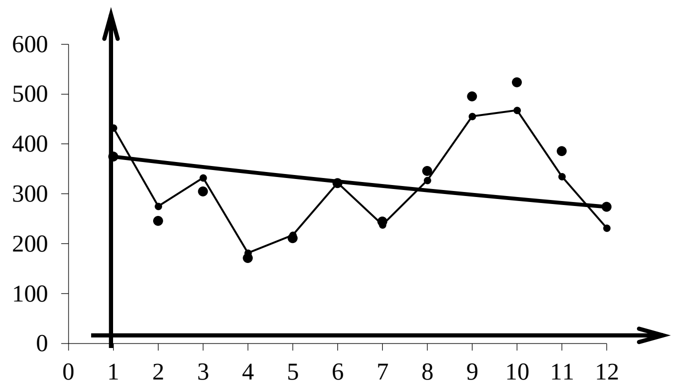

Рис. 6.16

Аналитическим способом учета сезонных факторов является представление неслучайной составляющей в виде тригонометрического ряда или уравнения Фурье[76]:

$$f(t) = a + \sum_{k=1}^{m} \left(b_k \cdot \cos(k \cdot \alpha_t) + c_k \cdot \sin(k \cdot \alpha_t)\right). \qquad (6.16)$$

В формуле (6.16) m – число слагаемых (гармоник) в тригонометрическом ряде, на практике берется не более четырех гармоник,

$$m \le 4.$$

Приведем вид уравнения Фурье для одной, двух и трех гармоник:

$$m=1:\ f(t) = a + b_1 \cdot \cos\alpha_t + c_1 \cdot \sin\alpha_t;$$

$$m=2:\ f(t) = a + b_1 \cdot \cos\alpha_t + c_1 \cdot \sin\alpha_t + $$
$$+ b_2 \cdot \cos 2\alpha_t + c_1 \cdot \sin 2\alpha_t;$$

$$m=3:\ f(t) = a + b_1 \cdot \cos\alpha_t + c_1 \cdot \sin\alpha_t +$$

[76] Фурье, Жан Батист Жозеф (1768 - 1830) – французский математик и физик.

$$+b_2 \cdot \cos 2\alpha_t + c_1 \cdot \sin 2\alpha_t +$$
$$+b_3 \cdot \cos 3\alpha_t + c_3 \cdot \sin 3\alpha_t.$$

Угол α_t связан с t соотношением:

$$\alpha_t = \frac{2\pi}{n} \cdot (t-1).$$

Например, при

$$n = 12$$

угол

$$\alpha_t = \frac{\pi}{6} \cdot (t-1). \qquad (6.17)$$

Параметры уравнения (6.16) могут быть найдены с помощью метода наименьших квадратов по формулам:

$$a = \frac{\sum y}{n}, \; b_k = \frac{2\sum y \cdot \cos(k \cdot \alpha_t)}{n}, \; c_k = \frac{2\sum y \cdot \sin(k \cdot \alpha_t)}{n}.$$

Пример 6.14. Построить модель неслучайной составляющей ряда динамики из примера 6.5 в виде уравнения Фурье. Число гармоник принять равным 1, 2 и 3. Результаты вычислений представить графически.

Углы определяем по формуле (6.17). Все необходимые вычисления производим в табл. 6.16. Обязательно проделайте их самостоятельно. Получаем:

$$a = \frac{3896{,}3}{12} \approx 324{,}6917;$$
$$b_1 = \frac{2 \cdot (-7{,}7290)}{12} \approx -0{,}9548, \; c_1 = \frac{2 \cdot (-694{,}9591)}{12} \approx -115{,}8265;$$
$$b_2 = \frac{2 \cdot (-179{,}75)}{12} \approx -29{,}9583, \; c_2 = \frac{2 \cdot 172\text{б}7719}{12} \approx 23{,}7953;$$
$$b_3 = \frac{2 \cdot 146{,}2}{12} \approx 24{,}3667, \; c_3 = \frac{2 \cdot 299{,}1}{12} = 49{,}8500.$$

Тем самым мы установили уравнения Фурье для рассматриваемого ряда динамики:

Таблица 6.16

t	y	$m=1$					$m=2$					$m=3$				
		α_t	$\cos\alpha_t$	$\sin\alpha_t$	$y\cdot\cos\alpha_t$	$y\cdot\sin\alpha_t$	$2\alpha_t$	$\cos2\alpha_t$	$\sin2\alpha_t$	$y\cdot\cos2\alpha_t$	$y\cdot\sin2\alpha_t$	$3\alpha_t$	$\cos3\alpha_t$	$\sin3\alpha_t$	$y\cdot\cos3\alpha_t$	$y\cdot\sin3\alpha_t$
1	374,6	0	1	0	374,6	0	0	1	0	374,6	0	0	1	0	374,6	0
2	245,5	$\pi/6$	0,866	0,5	212,6092	122,75	$\pi/3$	0,5	0,866	122,75	212,6092	$\pi/2$	0	1	0	245,5
3	304,6	$\pi/3$	0,5	0,866	152,3	263,7013	$2\pi/3$	-0,5	0,866	-152,3	263,7913	π	-1	0	-304,6	0
4	171,1	$\pi/2$	0	1	0	171,1	π	-1	0	-171,1	0	$3\pi/2$	0	-1	0	-171,1
5	210,8	$2\pi/3$	-0,5	0,866	-105,4	182,5582	$4\pi/3$	-0,5	-0,866	-105,4	-182,5582	0	1	0	210,8	0
6	321,3	$5\pi/6$	-0,866	0,5	-278,2540	160,65	$5\pi/3$	0,5	-0,866	160,65	-278,2540	$\pi/2$	0	1	0	321,3
7	244,7	π	-1	0	-244,7	0	0	1	0	244,7	0	π	-1	0	-244,7	0
8	345,6	$7\pi/6$	-0,866	-0,5	-299,2984	-172,8	$\pi/3$	0,5	0,866	172,8	299,2984	$3\pi/2$	0	-1	0	-345,6
9	495,4	$4\pi/3$	-0,5	-0,866	-247,7	-429,0290	$2\pi/3$	-0,5	0,866	-247,7	429,0290	0	1	0	495,4	0
10	523,2	$3\pi/2$	0	-1	0	-523,2	π	-1	0	-523,2	0	$\pi/2$	0	1	0	523,2
11	385,3	$5\pi/3$	0,5	-0,866	192,65	-333,6796	$4\pi/3$	-0,5	-0,866	-192,65	-333,6796	π	-1	0	-385,3	0
12	274,2	$11\pi/6$	0,866	-0,5	237,4642	-137,1	$5\pi/3$	0,5	-0,866	137,1	-237,4642	$3\pi/2$	0	-1	0	-274,2
$\sum$	3896,3	-	-	-	-5,7290	-694,9591	-	-	-	-179,75	172,7719	-	-	-	146,2	299,1

$$m=1:\ f(t)=324{,}6917-0{,}9548\cdot\cos\alpha_t-115{,}8265\cdot\sin\alpha_t; \qquad (6.18)$$

$$m=2:\ f(t)=324{,}6917-0{,}9548\cdot\cos\alpha_t-115{,}8265\cdot\sin\alpha_t-$$
$$-29{,}9583\cdot\cos 2\alpha_t+23{,}7953\cdot\sin 2\alpha_t; \qquad (6.19)$$

$$m=3:\ f(t)=324{,}6917-0{,}9548\cdot\cos\alpha_t-115{,}8265\cdot\sin\alpha_t-$$
$$-29{,}9583\cdot\cos 2\alpha_t+23{,}7953\cdot\sin 2\alpha_t+; \qquad (6.20)$$
$$+24{,}3667\cdot\cos 3\alpha_t+49{,}8500\cdot\sin 3\alpha_t.$$

Строим чертеж. Для этого в модели (6.18) – (6.20) подставляем значения синусов и косинусов углов, соответствующих каждому периоду ряда динамики, $t=1,2,\ldots,12$. Все они приведены в табл. 6.16. Получаем теоретические значения ряда динамики (ретроспективные прогнозы). Расчеты произведены в табл. 6.17. Обязательно проделайте их самостоятельно.

Таблица 6.17

t	y	$f(t)$		
		Модель (6.18)	Модель (6.19)	Модель (6.20)
1	374,6	323,7368	293,7785	318,1452
2	245,5	265,9515	275,9098	325,7598
3	304,6	223,9056	263,8222	239,4556
4	171,1	208,8652	238,8235	188,9735
5	210,8	224,8604	214,9020	239,2687
6	321,3	267,6053	227,6886	277,5386
7	244,7	325,6465	295,6882	271,3215
8	345,6	383,4318	393,3902	343,5402
9	495,4	425,4778	465,3944	489,7611
10	523,2	440,5182	470,4765	520,3265
11	385,3	424,5230	414,5646	390,1980
12	274,2	381,7780	341,8614	292,0114

На рис. 6.17 модель (6.18) показана точечной линией, модель (6.19) – пунктирной, а модель (6.20) – сплошной. Из рисунка следует, что наиболее близка к эмпирическим данным модель Фурье с числом гармоник m=3. Видимо, она будет среди этих трех моделей наиболее точной моделью неслучайной составляющей.

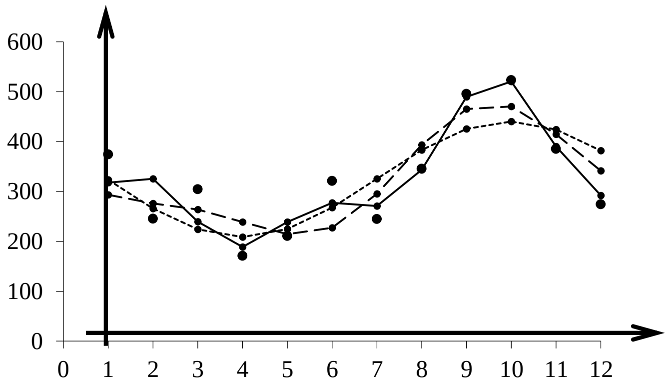

Рис. 6.17

6.8. Точность модели неслучайной составляющей

Прежде чем производить прогноз уровней ряда динамики, необходимо оценить точность модели неслучайной составляющей. Очевидно, что чем точнее модель, тоем точнее будет прогноз.

Точность модели будем определять также как и в параграфе 4 – с помощью величины средней ошибки аппроксимации:

$$\bar{\varepsilon} = \frac{1}{n}\sum_t \left|\frac{y - f(t)}{y}\right| \cdot 100\% .$$

Напомним, что если

$$\bar{\varepsilon} > 15\% ,$$

то точность прогноза будет неудовлетворительной, модель следует исключить из рассмотрения. Наиболее точной моделью неслучайной составляющей (и, соответственно, по которой возможен наиболее точный прогноз) признается та, которой соответствует минимальное значение средней ошибки аппроксимации.

Пример 6.15. Среди моделей неслучайной составляющей, найденных в примерах 6.13 и 6.14 найти наиболее точные. Сравнение произвести среди моделей (6.11) – (6.13), (6.14) – (6.15) и (6.18) – (6.20).

Вычисления производим в табл. 6.18. Теоретические значения уровней рядов динамики для каждой из моделей неслучайной составляющей получены в табл. 6.15 и 6.17.

Таблица 6.18

t	y	Модель (6.11)		Модель (6.12)		Модель (6.13)	
		$f(t)$	$\left\|\frac{y-f(t)}{y}\right\|$	$f(t)$	$\left\|\frac{y-f(t)}{y}\right\|$	$f(t)$	$\left\|\frac{y-f(t)}{y}\right\|$
1	374,6	293,9886	0,2152	320,4315	0,1446	288,6869	0,2293
2	245,5	202,2754	0,1761	210,1526	0,1440	196,5741	0,1993
3	304,6	262,8875	0,1369	263,2785	0,1357	253,4079	0,1681
4	171,1	154,3637	0,0978	150,6306	0,1196	147,8958	0,1356
5	210,8	198,4281	0,0587	190,5822	0,0959	189,3181	0,1019
6	321,3	315,0139	0,0196	300,5811	0,0645	299,8111	0,0669
7	244,7	249,4866	0,0196	238,4947	0,0254	237,2392	0,0305
8	345,6	365,8821	0,0587	353,0192	0,0215	348,1303	0,0073
9	495,4	543,8563	0,0978	533,048	0,0760	518,4891	0,0466
10	523,2	594,846	0,1369	595,5183	0,1382	568,9406	0,0874
11	385,3	453,1374	0,1761	465,5007	0,2082	435,3251	0,1298
12	274,2	333,2050	0,2152	352,5611	0,2858	321,8828	0,1739
$\sum$	-	-	1,4086	-	1,4594	-	1,3766
$\bar{\varepsilon}$	-	-	11,74	-	12,16	-	11,47

Таблица 6.18 (продолжение)

t	y	Модель (6.14)		Модель (6.15)	
		$f(t)$	$\left\|\frac{y-f(t)}{y}\right\|$	$f(t)$	$\left\|\frac{y-f(t)}{y}\right\|$
1	374,6	432,1758	0,1537	432,3492	0,1542
2	245,5	276,3323	0,1256	275,4132	0,1218
3	304,6	334,2924	0,0975	332,1463	0,0904
4	171,1	182,9693	0,0694	181,3492	0,0599
5	210,8	219,4977	0,0413	217,1714	0,0302
6	321,3	325,5254	0,0132	321,7429	0,0014
7	244,7	241,0396	0,0150	238,1763	0,0267
8	345,6	330,7155	0,0431	326,9675	0,0539
9	495,4	460,1383	0,0712	455,5679	0,0804
10	523,2	471,2525	0,0993	467,661	0,1062
11	385,3	336,2137	0,1274	334,7562	0,1312
12	274,2	231,5599	0,1555	231,5599	0,1555
$\sum$	-	-	1,0122	-	1,0118
$\bar{\varepsilon}$	-	-	8,44	-	8,43

Таблица 6.18 (продолжение)

t	y	Модель (6.18)		Модель (6.19)		Модель (6.20)	
		$f(t)$	$\left\lvert\frac{y-f(t)}{y}\right\rvert$	$f(t)$	$\left\lvert\frac{y-f(t)}{y}\right\rvert$	$f(t)$	$\left\lvert\frac{y-f(t)}{y}\right\rvert$
1	374,6	323,7368	0,1358	293,7785	0,2158	318,1452	0,1507
2	245,5	265,9515	0,0833	275,9098	0,1239	325,7598	0,3269
3	304,6	223,9056	0,2649	263,8222	0,1339	239,4556	0,2139
4	171,1	208,8652	0,2207	238,8235	0,3958	188,9735	0,1045
5	210,8	224,8604	0,0667	214,902	0,0195	239,2687	0,1351
6	321,3	267,6053	0,1671	227,6886	0,2914	277,5386	0,1362
7	244,7	325,6465	0,3308	295,6882	0,2084	271,3215	0,1088
8	345,6	383,4318	0,1095	393,3902	0,1383	343,5402	0,0060
9	495,4	425,4778	0,1411	465,3944	0,0606	489,7611	0,0114
10	523,2	440,5182	0,1580	470,4765	0,1008	520,3265	0,0055
11	385,3	424,523	0,1018	414,5646	0,0760	390,198	0,0127
12	274,2	381,778	0,3923	341,8614	0,2468	292,0114	0,0650
$\sum$	-	-	2,1720	-	2,0112	-	1,2767
$\bar{\varepsilon}$	-	-	18,10	-	16,76	-	10,64

На основании произведенных расчетов делаем выводы.

1. Среди моделей (6.11) – (6.13) наиболее точной является модель (6.13), так как

$$\min \bar{\varepsilon} = \min(11{,}74; 12{,}16; 11{,}47) = 11{,}47.$$

2. Среди моделей (6.14) и (6.15) наиболее точной является модель (6.15), так как

$$\min \bar{\varepsilon} = \min(8{,}44; 8{,}43) = 8{,}43.$$

3. Среди моделей (6.18) – (6.20) наиболее точной является модель (6.20), так как

$$\min \bar{\varepsilon} = \min(18{,}10; 16{,}76; 10{,}64) = 10{,}64.$$

4. Среди всех рассмотренных моделей неслучайной составляющей наиболее точной признается модель (6.15), так как

$$\min \bar{\varepsilon} = \min(11{,}47; 8{,}43; 10{,}64) = 8{,}43.$$

6.9. Точечный прогноз. Интервальный прогноз

После того, как отобрана наиболее точная модель неслучайной составляющей, по ней может быть осуществлен прогноз модели неслучайной составляющей. Для этого в правую часть уравнения подставляем прогнозируемое значение t.

Пример 6.16. Среди наиболее точных моделей неслучайной составляющей (пример 6.15) ряда динамики из примера 6.5, спрогнозировать прибыль от продаж торгующей организации на апрель 2020 года.

Прогнозируемый период является шестнадцатым от начала отсчета (январь 2019 г.),

$$t = 16.$$

Получаем:

(6.13): $y_{04.20} = f(16) = I_4 \cdot f_1(16) = 0{,}5270 \cdot \left(240{,}8326 \cdot (1{,}0390)^{16}\right) \approx 234{,}086$ (тыс. руб.);

(6.15): $y_{04.20} = f(16) = I_4 \cdot f_1(16) = I_4 \cdot f_1(12+4) = 0{,}5270 \cdot \left(274{,}2 \cdot (0{,}9720)^{4}\right) \approx 128{,}986$ (тыс. руб.);

(6.20): $y_{04.20} = f(16) = f(4) \approx 188{,}974$ (тыс. руб.).

Последний результат берем из табл. 6.17, так как уравнение Фурье – циклическая функция.

Итак, нами получен прогноз уровней ряда динамики, произведенный по модели неслучайной составляющей. Он выражается одним числом, поэтому называется точечным. Однако, при этом не было учтено влияние случайных факторов (см. п. 6.4). Этот недостаток устранен в интервальном прогнозе. Утверждается, что в прогнозный период времени t, которому соответствует период упреждения

$$\tau = t - n,$$

с доверительной вероятностью γ (уровнем значимости $\alpha = 1 - \gamma$) значение уровня ряда динамики y принадлежит интервалу:

$$y \in \left(f(t) - t_{крит} \cdot \Delta_\tau;\ f(t) + t_{крит} \cdot \Delta_\tau\right),$$

при этом $f(t)$ - точечный прогноз ряда динамики, произведенный по модели неслучайной составляющей; $t_{крит}$ - критическое значение критерия Стьюдента, определяемое по таблице в зависимости от уровня значимости и числа степеней свободы ν,

$$t_{крит} = t(\alpha; \nu),\ \alpha = 1 - \gamma,\ \nu = n - 2;$$

Δ_τ - ошибка прогноза,

$$\Delta_\tau = \delta \cdot \sqrt{1 + \frac{1}{n} + \frac{\left(\tau - \bar{t}\right)^2}{\sum_{i=1}^{n} t_i^2 - \frac{\left(\sum_{i=1}^{n} t_i\right)^2}{n}}},$$

среднее значение

$$\bar{t} = \frac{\sum_{i=1}^{n} t_i}{n},$$

а среднеквадратическая ошибка

$$\delta = \sqrt{\frac{\sum_{i=1}^{n} \left(y_i - f(t_i)\right)^2}{n-2}}.$$

Пример 6.17. С вероятностью 0,95 уточнить точечный прогноз прибыли торгующей организации в апреле 2020 г. (пример 6.16).

По табл. 4 прил. находим критическое значение критерия Стьюдента для доверительной вероятности $\gamma = 0{,}95$ (уровня значимости $\alpha = 1 - 0{,}95 = 0{,}05$) и числа степеней свободы $\nu = 12 - 2 = 10$:

$$t_{крит} = t(0{,}05;10) = 2{,}23.$$

Для апреля 2020 г.

$$t = 16.$$

Ему соответствует период упреждения

$$\tau = t - n = 16 - 12 = 4.$$

В качестве модели неслучайной составляющей мы рассматриваем наиболее точную модель – (6.15). Теоретические значения неслучайной составляющей приведены в табл. 6.15. Все необходимые расчеты производим в табл. 6.19.

Таблица 6.19

i	t_i	t_i^2	y_i	$f(t_i)$	$y - f(t_i)$	$(y - f(t_i))^2$
1	1	1	374,6	432,3492	-57,7492	3334,9701
2	2	4	245,5	275,4132	-29,9132	894,7995
3	3	9	304,6	332,1463	-27,5463	758,7986
4	4	16	171,1	181,3492	-10,2492	105,0461
5	5	25	210,8	217,1714	-6,3714	40,5947
6	6	36	321,3	321,7429	-0,4429	0,1962
7	7	49	244,7	238,1763	6,5237	42,5587
8	8	64	345,6	326,9675	18,6325	347,1701
9	9	81	495,4	455,5679	39,8321	1586,5962
10	10	100	523,2	467,661	55,539	3084,5805
11	11	121	385,3	334,7562	50,5438	2554,6757
12	12	144	274,2	231,5599	42,6401	1818,1781
$\sum$	78	650				14568,1645

Получаем:

$$\bar{t} = \frac{78}{12} = 6{,}5;$$

среднеквадратическая ошибка

$$\delta = \sqrt{\frac{14568{,}1645}{12-2}} \approx 38{,}1683;$$

$$\sqrt{1 + \frac{1}{12} + \frac{(4-6{,}5)^2}{650 - \frac{(78)^2}{12}}} \approx 1{,}0396;$$

ошибка прогноза

$$\Delta_4 = 38{,}1683 \cdot 1{,}0396 \approx 39{,}6798;$$

$$t_{крит} \cdot \Delta_4 = 2{,}23 \cdot 39{,}6798 \approx 88{,}486 \text{ (тыс. руб.)}.$$

Используя точечный прогноз из примера 6.17, получаем: прибыль от продаж торгующей организации в апреле 2020 года составляет

$$128{,}986 \pm 88{,}486 \text{ (тыс. руб.)}$$

Обратим внимание на достаточно тривиальный характер произведенного прогноза. Такой результат связан прежде всего с недостаточной точностью модели неслучайной составляющей ($8{,}43\%$).

§ 7. ИНДЕКСНЫЙ МЕТОД

7.1. Основные определения

Экономический индекс – сложный относительный показатель, который характеризует изменение исследуемого экономического признака по сравнению с некоторым эталонным уровнем. Если в качестве него выступает уровень некоторого предыдущего периода, то индекс – динамический. Когда под эталонным понимается уровень того же периода, но на другой территории, то индекс – территориальный. Сравниваемый и эталонный уровни называются соответственно отчетным и базисным.

Обозначим p – цена (цена реализации, цена продаж и т. д.). Пусть также z – себестоимость продукции. Показатель q, характеризующий объем выпуска, продаж и т. д., назовем объемом. Тогда произведение цены на объем $p \cdot q$ очевидно назвать товарооборотом, а произведение себестоимости на объем $z \cdot q$ - себестоимость.

Пусть также T – время, необходимое на производство продукции в объеме q. Тогда отношение

$$w = \frac{q}{T}$$

есть производительность труда. Производительность труда определяет количество продукции, изготавливаемой за единицу времени. Величина, обратная к w- трудоемкость

$$t = \frac{1}{w} = \frac{T}{q} \cdot$$

Трудоемкость показывает время, необходимое на производство единицы продукции.

Все показатели, относящиеся к отчетному уровню будем обозначать индексом 1, а показатели базисного уровня снабжены индексом 0.

7.2. Индивидуальные индексы

Для рассмотрения отдельных элементов совокупности используются индивидуальные индексы. Они рассчитываются как отношение показателей отчетного уровня к соответствующим показателям базисного.

Отсюда, индивидуальные индексы цены, себестоимости и объема рассчитываются по соответствующим формулам:

$$i_p = \frac{p_1}{p_0},$$

$$i_z = \frac{z_1}{z_0},$$

$$i_q = \frac{q_1}{q_0}.$$

Исходя из введенных определений, получаем индивидуальные индексы товарооборота и себестоимости, а также их связь с индивидуальными индексами цены, себестоимости и объема:

$$i_{p \cdot q} = \frac{p_1 \cdot q_1}{p_0 \cdot q_0} = \frac{p_1}{p_0} \cdot \frac{q_1}{q_0} = i_{p \cdot} \cdot i_q,$$

$$i_{zq} = \frac{z_1 \cdot q_1}{z_0 \cdot q_0} = \frac{z_1}{z_0} \cdot \frac{q_1}{q_0} = i_z \cdot i_q.$$

Если

$$i_T = \frac{T_1}{T_0}$$

- индивидуальный индекс суммарного времени, то индивидуальный индекс производительности труда подсчитывается по формуле и имеет следующую связь с индивидуальными индексами суммарного времени и объема:

$$i_w = \frac{w_1}{w_0} = \frac{q_1}{T_1} : \frac{q_0}{T_0} = \frac{q_1}{T_1} \cdot \frac{T_0}{q_0} = \frac{q_1}{q_0} \cdot \frac{T_0}{T_1} = \frac{q_1}{q_0} : \frac{T_1}{T_0} = i_q : i_T$$

$$i_t = \frac{t_1}{t_0} = \frac{T_1}{q_1} : \frac{T_0}{q_0} = \frac{T_1}{q_1} \cdot \frac{q_0}{T_0} = \frac{T_1}{T_0} \cdot \frac{q_0}{q_1} = \frac{T_1}{T_0} : \frac{q_1}{q_0} = i_T : i_q = \frac{1}{i_w}.$$

Пример 7.1. Торгово-промышленное предприятие производит и реализует три вида товаров А, В и С. Его деятельность в 2018 и 2019 годах характеризуется следующими данными (табл. 7.1):

Таблица 7.1

Товар	Себестоимость, руб./ед.		Цена изделия, тыс. руб./ед.		Объем выпуска (продаж), тыс. ед.		Трудоемкость, час./ед.	
	2018	2019	2018	2019	2018	2019	2018	2019
А	515	624	0,85	0,87	5,8	6,4	0,54	0,57
В	243	243	0,44	0,45	7,0	7,5	0,38	0,39
С	178	169	0,38	0,35	10,0	14,0	0,46	0,41

Рассчитать индивидуальные индексы себестоимости, цены, объема, затрат, товарооборота, трудоемкости и производительности труда.

В качестве базисного периода берем показатели 2018 года, а в качестве отчетного – показатели 209 года.

Товар А:

$$i_z = \frac{624}{515} \approx 1{,}2117 \text{ или } 121{,}17\%;$$

$$i_p = \frac{0{,}87}{0{,}85} \approx 1{,}0235 \text{ или } 102{,}35\%;$$

$$i_q = \frac{6{,}4}{5{,}8} \approx 1{,}1034 \text{ или } 110{,}34\%;$$

$$i_{zq} = 1{,}2117 \cdot 1{,}1034 \approx 1{,}3370 \text{ или } 133{,}70\%;$$

$$i_{p \cdot q} = 1{,}0235 \cdot 1{,}1034 \approx 1{,}1293 \text{ или } 112{,}93\%;$$

$$i_t = \frac{0{,}57}{0{,}54} \approx 1{,}0556 \text{ или } 105{,}56\%;$$

$$i_w = \frac{1}{1{,}0556} \approx 0{,}9473 \text{ или } 94{,}73\%.$$

Товар B:

$$i_z = \frac{243}{243} \approx 1 \text{ или } 100\%;$$

$$i_p = \frac{0{,}45}{0{,}44} \approx 1{,}0227 \text{ или } 102{,}27\%;$$

$$i_q = \frac{7{,}5}{7{,}0} \approx 1{,}0714 \text{ или } 107{,}14\%;$$

$$i_{zq} = 1 \cdot 1{,}0714 \approx 1{,}0714 \text{ или } 107{,}14\%;$$

$$i_{p \cdot q} = 1{,}0227 \cdot 1{,}0714 \approx 1{,}0957 \text{ или } 109{,}57\%;$$

$$i_t = \frac{0{,}39}{0{,}38} \approx 1{,}0263 \text{ или } 102{,}63\%;$$

$$i_w = \frac{1}{1{,}0263} \approx 0{,}9744 \text{ или } 97{,}44\%.$$

Товар C:

$$i_z = \frac{169}{178} \approx 0{,}9494 \text{ или } 94{,}94\%;$$

$$i_p = \frac{0{,}35}{0{,}38} \approx 0{,}9211 \text{ или } 92{,}11\%;$$

$$i_q = \frac{14{,}0}{10{,}0} \approx 1{,}4 \text{ или } 140\%;$$

$$i_{zq} = 0{,}9494 \cdot 1{,}4 \approx 1{,}3292 \text{ или } 132{,}92\%;$$

$$i_{p \cdot q} = 0{,}9211 \cdot 1{,}4 \approx 1{,}2895 \text{ или } 128{,}95\%;$$

$$i_t = \frac{0{,}41}{0{,}46} \approx 0{,}8913 \text{ или } 89{,}3\%;$$

$$i_w = \frac{1}{0{,}8913} \approx 1{,}1220 \text{ или } 112{,}20\%.$$

Итак, себестоимость в 2019 году по сравнению с 2018 годом на товар A возросла на 21,17%, на товар B себестоимость не изменилась, а на товар C снизилась на 5,06%. В свою очередь цена на товар A увеличилась в отчетном периоде по сравнению с базисным на 2,35%, для товара В рост цены составил 2,27%, а на товар С произошло снижение цены на 7,89%. Объем выпуска (продаж) товара А в 2019 году по сравнению с 2018 годом вырос 10,34%, производство товара В также увеличилось (на 7,14%), а производство товара С увеличилось в 1,4 раза – на 40%. Затраты на производство товара А возросли на 33,7%, также выросли затраты на производство товара В – на 7,14%, рост затрат на товар С составил 32,92%. Увеличение товарооборота по каждому товару в 2019 году по сравнению с 2018 годом выглядит следующим образом: для товара А – на 12,93%, для товара В – на 9,57%, а для товара С – на 28,95%. Изменение трудоемкости по каждому товару в 2019 году по сравнению с 2018 годом составило: по товару А трудоемкость увеличилась на 5,56%, по товару В – увеличилась на 2,63%, а по товару С трудоемкость сократилась на 10,7%. Соответственно, производительность труда по товарам А и В уменьшилась на 5,27% и 2,56% соответственно, а по товару С производительность труда увеличилась на 12,2%.

7.3. Сводные индексы

Для измерения суммарного изменения экономических показателей, характеризующих работу одного предприятия, рынка, территории и т. д. используют сводные индексы.

Сводный индекс цены I_p подсчитывается по формуле

$$I_p = \frac{\sum p_1 \cdot q_1}{\sum p_0 \cdot q_1}.$$

Индекс цен, записанный в таком виде называется индексом цены Пааше[77]. В нем в качестве веса выступает объем отчетного периода. Также в качестве веса может использоваться объем базисного периода:

$$I_p = \frac{\sum p_1 \cdot q_0}{\sum p_0 \cdot q_0}.$$

В этом случае индекс цены называется индексом цены Ласпейерса[78].

[77] Пааше, Герман (1851 - 1925) – немецкий экономист, статистик и политик.

[78] Ласпейерс, Эрнст Луи Этьен (1834 - 1913) – немецкий экономист и статистик, юрист, философ.

Сводный индекс физического объема равен

$$I_q = \frac{\sum p_0 \cdot q_1}{\sum p_0 \cdot q_0}.$$

Наконец, сводный индекс товарооборота равен отношению суммарного товарооборота отчетного периода к суммарному товарообороту базисного периода:

$$I_{pq} = \frac{\sum p_1 \cdot q_1}{\sum p_0 \cdot q_0}.$$

Укажем связь между веденными индексами:

$$I_{pq} = \frac{\sum p_1 \cdot q_1}{\sum p_0 \cdot q_0} = \frac{\dfrac{\sum p_1 \cdot q_1}{\sum p_0 \cdot q_1}}{\dfrac{\sum p_0 \cdot q_0}{\sum p_0 \cdot q_1}} =$$

$$= \frac{\sum p_1 \cdot q_1}{\sum p_0 \cdot q_1} \cdot \frac{\sum p_0 \cdot q_1}{\sum p_0 \cdot q_0} = I_p \cdot I_q,$$

$$I_{pq} = I_p \cdot I_q.$$

При расчете сводных индексов можно пользоваться формулами средних. Рассмотрим индекс цен Пааше. Делая очевидные преобразования и используя определение индивидуального индекса цены, получаем: $I_p = \frac{\sum p_1 \cdot q_1}{\sum p_0 \cdot q_1} = \frac{\sum p_1 \cdot q_1}{\sum \frac{p_0}{p_1} \cdot p_1 \cdot q_1} =$

$$= \frac{\sum p_1 \cdot q_1}{\sum \frac{1}{\frac{p_1}{p_0}} \cdot p_1 \cdot q_1} = \frac{\sum p_1 \cdot q_1}{\sum \frac{1}{i_p} \cdot p_1 \cdot q_1}.$$

Итак, сводный индекс цены (в интерпретации Пааше) может быть подсчитан по формуле средней гармонической взвешенной:

$$I_p = \frac{\sum p_1 \cdot q_1}{\sum \frac{1}{i_p} \cdot p_1 \cdot q_1}.$$

В ней в качестве веса выступают величины товарооборотов отчетного периода[79]. Ее целесообразно использовать тогда, когда известны изменения цен по каждому товару и значения товарооборотов отчетного периода.

[79] Получите формулу сводного индекса цены в формулировке Ласпейерса.

Покажем, что сводный индекс цен может быть найден по формуле средней арифметической взвешенной:

$$I_q = \frac{\sum p_0 \cdot q_1}{\sum p_0 \cdot q_0} = \frac{\sum p_0 \cdot q_1 \cdot \frac{q_0}{q_0}}{\sum p_0 \cdot q_0} = \frac{\sum p_0 \cdot q_0 \cdot \frac{q_1}{q_0}}{\sum p_0 \cdot q_0} =$$

$$= \frac{\sum p_0 \cdot q_0 \cdot \frac{q_1}{q_0}}{\sum p_0 \cdot q_0} = \frac{\sum p_0 \cdot q_0 \cdot i_q}{\sum p_0 \cdot q_0} = \frac{\sum i_q \cdot p_0 \cdot q_0}{\sum p_0 \cdot q_0},$$

$$I_q = \frac{\sum i_q \cdot p_0 \cdot q_0}{\sum p_0 \cdot q_0}.$$

Он применяется тогда, когда есть информация об изменения объемов по каждому товару и значения товарооборотов базисного периода.

Кроме определенных выше, рассчитываются также: сводный индекс себестоимости

$$I_z = \frac{\sum z_1 \cdot q_1}{\sum z_0 \cdot q_1};$$

сводный индекс объема, взвешенный по себестоимости

$$I_q^z = \frac{\sum z_0 \cdot q_1}{\sum z_0 \cdot q_0};$$

сводный индекс затрат на производство

$$I_{zq} = \frac{\sum z_1 \cdot q_1}{\sum z_0 \cdot q_0} = I_z \cdot I_q^z.$$

При расчете индексов себестоимости и объема можно также воспользоваться формулами средних. Мы предлагаем читателю получить их самостоятельно.

С помощью индексов цены и себестоимости можно вычислить величину экономии (в этом случае значение показателя будет отрицательным) или перерасхода (значение положительное) E_p покупателей от изменения цены, а также величину экономии (перерасход) E_z производителей от изменения себестоимости:

$$E_p = \sum p_1 \cdot q_1 - \sum p_0 \cdot q_1;$$

$$E_z = \sum z_1 \cdot q_1 - \sum z_0 \cdot q_1.$$

Пример 7.2. По данным примера 7.1 найти сводные индексы и величины экономии покупателей от изменения цен и потребителей от изменения себестоимости.

Все необходимые вычисления производим в таблице (табл. 7.2).

Таблица 7.2

Товар	Z		p		q		Расчетные графы					
	z_0	z_1	p_0	p_1	q_0	q_1	$p_0 \cdot q_0$	$p_0 \cdot q_1$	$p_1 \cdot q_1$	$z_0 \cdot q_0$	$z_0 \cdot q_1$	$z_1 \cdot q_1$
A	515	624	0,85	0,87	5,8	6,4	4,93	5,44	5,568	2987	3296	3993,6
B	243	243	0,44	0,45	7,0	7,5	3,08	3,30	3,375	1701	1822,5	1822,5
C	178	169	0,38	0,35	10,0	14,0	3,80	5,32	4,900	1780	2492	2366
Σ	-	-	-	-	-	-	11,81	14,06	13,843	6468	7610,5	8182,1

Получаем: сводный индекс цен

$$I_p = \frac{13{,}843}{14{,}06} \approx 0{,}9846 \text{ или } 98{,}46\%.$$

Это значит, что цена на рассматриваемые товары в 2019 году снизилась на 1,54% по сравнению с 2018 годом. Сводный индекс объема равен

$$I_q = \frac{14{,}06}{11{,}81} \approx 1{,}1905 \text{ или } 119{,}85\%.$$

Следовательно, рост объемов продаж в 2019 году по сравнению с 2018 годом составил 19,85%. Сводный индекс товарооборота

$$I_{pq} = 0{,}9846 \cdot 1{,}1905 \approx 1{,}1722 \text{ или } 117{,}22\%,$$

что говорит о росте товарооборота в 2019 году по сравнению с 2018 г.

Вычислим теперь сводные индексы себестоимости, объема, взвешенного по себестоимости и затрат:

$$I_z = \frac{8182{,}1}{7610{,}5} \approx 1{,}0751 \text{ или } 107{,}51\%;$$

$$I_q^z = \frac{7610{,}5}{6468} \approx 1{,}1766 \text{ или } 117{,}66\%;$$

$$I_{zq} = 1{,}0751 \cdot 1{,}1766 \approx 1{,}2650 \text{ или } 126{,}50\%.$$

Итак, в 2019 году по сравнению с 2018 годом: себестоимость выросла на 7,51%, объем производства также вырос – на 17,66%; увеличение себестоимости и производства сказалось на росте затрат – они увеличились на 26,50%.

Определим величину экономии от изменения цены:

$$E_p = 13{,}843 - 14{,}06 = -0{,}217 \text{ (млн. руб.).}$$

Получаем: экономия покупателей от изменения цены составила 217 тысяч рублей. Экономия от изменения себестоимости

$$E_z = 8182{,}1 - 7610{,}5 = 571{,}6 \text{ (тыс. руб.)}.$$

Заключаем: наблюдается перерасход от изменения себестоимости в размере 571,6 тысяч рублей.

Пример 7.3. Пусть мы располагаем значениями товарооборотов торгующей организации за 2019 и 2019 годы (млн. руб.), а также информацией об изменении цены и объема (%) в 2019 году по сравнению с 2018 годом (табл. 7.3).

Таблица 7.3

Товар	Товарооборот, млн. руб.		Изменение цены, %	Изменение объемов продаж, %
	2018	2019		
A	4,93	5,568	+2,35	+10,34
B	3,08	3,375	+2,27	+7,14
C	3,80	4,900	-7,89	+40,0

Найти сводные индексы цены и объема.

Для расчета сводных индексов здесь целесообразно воспользоваться средними формулами. Получаем:

$$I_p = \frac{5{,}568 + 3{,}375 + 4{,}900}{\frac{5{,}568}{1{,}0235} + \frac{3{,}375}{1{,}0227} + \frac{4{,}900}{0{,}9211}} \approx 0{,}9846 \text{ или } 98{,}46\%;$$

$$I_q = \frac{1{,}1034 \cdot 4{,}93 + 1{,}0714 \cdot 3{,}08 + 1{,}4 \cdot 3{,}80}{4{,}93 + 3{,}08 + 3{,}80} \approx 1{,}1905 \text{ или } 119{,}05\%.$$

Сводный индекс производительности труда может быть подсчитан двумя способами. Сводный индекс производительности труда, взвешенный по трудоемкости, подсчитывается по формуле:

$$I_w^t = \frac{\sum t_0 \cdot q_1}{\sum t_1 \cdot q_1}.$$

Также рассчитывается сводный индекс производительности труда, взвешенный по выработке:

$$I_w^q = \frac{\sum q_1 \cdot p}{\sum T_1} : \frac{\sum q_0 \cdot p}{\sum T_0},$$

при этом $T_1 = t_1 \cdot q_1$ и $T_0 = t_0 \cdot q_0$ – суммарное время затраченное на выпуск изделий каждого вида в отчетном и базисном периодах соответственно, p - сопоставимые цены, то есть цены либо отчетного, либо базисного периодов, либо средние цены.

Используя взаимосвязь производительности труда и трудоемкости, определяем сводный индекс трудоемкости:

$$I_t = \frac{1}{I_w^z} = \frac{\sum t_1 \cdot q_1}{\sum t_0 \cdot q_1}$$

Легко показать, что сводный индекс производительности труда, взвешенный по трудоемкости, может быть подсчитан по формуле средней арифметической, в которой в качестве веса выступает суммарное время изготовления каждого вида продукции в отчетном периоде. Действительно $I_w^z = \frac{\sum t_0 \cdot q_1}{\sum t_1 \cdot q_1} = \frac{\sum t_0 \cdot q_1 \cdot \frac{t_1}{t_1}}{\sum t_1 \cdot q_1} = \frac{\sum \frac{t_0}{t_1} \cdot t_1 \cdot q_1}{\sum t_1 \cdot q_1} = \frac{\sum \frac{1}{i_t} \cdot t_1 \cdot q_1}{\sum t_1 \cdot q_1} =$

$= \frac{\sum i_w \cdot t_1 \cdot q_1}{\sum t_1 \cdot q_1} = \frac{\sum i_w \cdot T_1}{\sum T_1}$,

$$I_w^z = \frac{\sum i_w \cdot T_1}{\sum T_1}.$$

Сводный индекс суммарного времени, необходимого на выпуск всей продукции, равен:

$$I_T = \frac{\sum t_1 \cdot q_1}{\sum t_0 \cdot q_0} = \frac{\sum T_1}{\sum T_0}.$$

Наконец, сводный индекс объема, взвешенный по трудоемкости

$$I_q^t = \frac{\sum t_0 \cdot q_1}{\sum t_0 \cdot q_0}.$$

Причем

$$I_q^t = I_w^t \cdot I_T,$$

так как $I_q^t = \frac{\sum t_0 \cdot q_1}{\sum t_0 \cdot q_0} = \frac{\sum t_0 \cdot q_1}{\sum t_0 \cdot q_0} \cdot \frac{\sum t_1 \cdot q_1}{\sum t_1 \cdot q_1} = \frac{\sum t_0 \cdot q_1}{\sum t_1 \cdot q_1} \cdot \frac{\sum t_1 \cdot q_1}{\sum t_0 \cdot q_0} = I_w^t \cdot I_T$.

Пример 7.4. Предприятие производит товары четырех типов: A, B, C и D. В табл. 1.4 представлены данные об объеме выпуска (тыс. шт.), времени производства единицы товара (мин.) и цены реализации (у.е.) в 2018 – 2019 гг.

Таблица 7.4

Товар	2018 г.			2019 г.		
	Объем выпуска, тыс. шт.	Время производства одной единицы товара, мин.	Цена реализации единицы товара, у.е.	Объем выпуска, тыс. шт.	Время производства одной единицы товара, мин.	Цена реализации единицы товара, у.е.
A	245	12	10	324	10	11
B	147	10	15	178	8	16
C	200	7	20	215	5	18
D	315	6	23	300	8	25

Рассчитать: сводные индексы производительности труда, взвешенные по трудоемкости и выработке, сводный индекс трудоемкости; сводный индекс суммарного времени; сводный индекс физического объема, взвешенный по трудоемкости.

Все необходимые расчеты производим в табл. 7.5. В качестве сопоставимых цен берем средние цены по каждому товару:

$$p = \frac{p_0 \cdot q_0 + p_1 \cdot q_1}{q_0 + \cdot q_1}.$$

Получаем:

$$I_T = \frac{8139}{7700} \approx 1{,}0570 \text{ или } 105{,}70\%;$$

$$I_w^t = \frac{8973}{8139} \approx 1{,}1025 \text{ или } 10{,}25\%;$$

$$I_w^q = \frac{17566{,}4379}{8139} : \frac{16291{,}5999}{7700} \approx 1{,}0195 \text{ или } 110{,}95\%;$$

$$I_t = \frac{1}{1{,}1025} \approx 0{,}9070 \text{ или } 90{,}70\%;$$

$$I_q^t = 1{,}1025 \cdot 1{,}0570 \approx 1{,}1653 \text{ или } 116{,}53\%.$$

Таблица 7.5

Товар	q		t		p		$T_0 = t_0 \cdot q_0$	$t_0 \cdot q_1$	$T_1 = t_1 \cdot q_1$	p	$q_0 \cdot p$	$q_1 \cdot p$
	q_0	q_1	t_0	t_1	p_0	p_1						
A	245	324	12	10	10	11	2940	3888	3240	10,8612	2660,9940	3519,0288
B	147	178	10	8	15	16	1470	1780	1424	15,5477	2285,5119	2767,4906
C	200	215	7	5	20	18	1400	1505	1075	18,9639	3722,7800	4077,2385
D	315	300	6	8	23	25	1890	1800	2400	23,9756	7552,3140	7192,6800
Σ	-	-	-	-	-	-	7700	8973	8139	-	16291,5999	17566,4379

Делаем выводы. Суммарное время на выпуск продукции в 2019 году по сравнению с 2018 годом увеличилось на 5,70%. За счет изменения трудоемкости производительность труда выросла на 10,25%, а за счет изменения выработки – на 1,95%. В 2019 году произошло сокращение трудоемкости производства на 9,30%. Выработка вросла на 16,53%.

7.4. Индексы переменного, постоянного состава и структурных сдвигов

Вышеперечисленные индексы позволяли оценить показатели производства разных товаров и товарных групп, выпускаемых одним предприятием, продажи продукции на одном рынке и т. д. С помощью следующих групп индексов можно определить эффективность работы нескольких предприятий, производящих однотипную продукцию, продаж одного и того же товара на нескольких рынках.

Первым в данной системе индексов является индекс цен переменного состава $I_p^{ПС}$. Он рассчитывается как отношение средних цен отчетного $\overline{p}_1$ и базисного $\overline{p}_0$ периодов:

$$I_p^{ПС} = \overline{p}_1 : \overline{p}_1 = \frac{\sum p_1 \cdot q_1}{\sum q_1} : \frac{\sum p_0 \cdot q_0}{\sum q_0} .$$

Изменение индивидуальных цен, а также изменение и специфика реализации (производства) в различных местах продажи (производства) учитывается индексом структурных сдвигов $I^{СТР}$:

$$I^{СТР} = \frac{\sum p_0 \cdot q_1}{\sum q_1} : \frac{\sum p_0 \cdot q_0}{\sum q_0} .$$

Наконец, изменение цен без учета структуры производится с помощью индекса цен фиксированного состава $I_p^{ФС}$, который рассчитывается также, как индекс цен Пааше:

$$I_p^{ФС} = \frac{\sum p_1 \cdot q_1}{\sum p_0 \cdot q_1} .$$

Между введенными индексами установим связь: $I_p^{ФС} \cdot I^{СТР} =$

$$= \frac{\sum p_1 \cdot q_1}{\sum p_0 \cdot q_1} \cdot \frac{\sum p_0 \cdot q_1}{\sum q_1} : \frac{\sum p_0 \cdot q_0}{\sum q_0} = \frac{\sum p_1 \cdot q_1}{\sum q_1} : \frac{\sum p_0 \cdot q_0}{\sum q_0} = I_p^{ПС} ,$$

$$I_p^{ПС} = I_p^{ФС} \cdot I^{СТР} .$$

Пример 7.5. Производится анализ реализации товара А в трех регионах. Цены на товар и объем реализации в 2018 – 2019 годах приведены в табл. 7.6.

Таблица 7.6

Регион	2018		2019	
	Цена, д.е./ед. прод.	Продано, тыс. ед.	Цена, д.е./ед. прод.	Продано, тыс. ед.
1	7,80	120	7,90	150
2	8,00	103	8,15	95
3	7,65	116	7,70	118

Производим вычисления в табл. 7.7.

Таблица 7.7

Регион	2018		2019		$p_0 \cdot q_0$	$p_0 \cdot q_1$	$p_1 \cdot q_1$
	p_0	q_0	p_1	q_1			
1	7,80	120	7,90	150	936	1170	1185
2	8,00	103	8,15	95	824	760	774,25
3	7,65	116	7,70	118	887,4	902,7	908,6
Σ	-	339	-	363	2647,4	2832,7	2867,85

Получаем: индекс цен переменного состава

$$I_p^{ПС} = \frac{2867{,}85}{363} : \frac{2647{,}4}{339} \approx 1{,}0116 \text{ или } 101{,}16\%;$$

индекс цен фиксированного состава

$$I_p^{ФС} = \frac{2867{,}85}{2832{,}7} \approx 1{,}0124 \text{ или } 101{,}24\%;$$

индекс структурных сдвигов

$$I^{СТР} = \frac{2832{,}7}{363} : \frac{2647{,}4}{339} \approx 0{,}9992 \text{ или } 99{,}92\%.$$

Видим, что средняя цена выросла на 1,16% в 2019 году по сравнению с 2018 годом. Однако, рассчитанный индекс цен фиксированного состава указывает на рост цен на 1,24%. Такое несоответствие объясняется тем, что при расчете индекса цен фиксированного состава не учтено изменение структуры реализации товара: в 2019 году по более дорогой цене было продано товаров меньше, чем в 2018 году. В целом, за счет изменения структуры цены снизились на 0,08%, что и привело к росту цен в среднем лишь на 1,16%.

7.5. Территориальные индексы

Для сравнения социально – экономических явлений в пространстве служат территориальные индексы. Их особенностью является то, что в качестве сравниваемой (аналога отчетного периода) может выступать как та, так и другая территория (регион).

Один из способов сравнения заключается в том, что в качестве веса выступают объемы Q обеих территорий. Именно, если А и В – рассматриваемые регионы, то

$$Q = q_A + q_B.$$

Тогда, территориальный индекс цен $I_p^{B/A}$ равен

$$I_p^{B/A} = \frac{\sum p_B \cdot Q}{\sum p_A \cdot Q}.$$

Пример 7.6. По ценам и объемам реализации товаров трех видов в двух регионах А и В рассчитать территориальный индекс цен. Исходные данные приведены в табл. 7.8.

Таблица 7.8

Товар	Регион А		Регион В	
	Цена, д.е./ед. прод.	Продано, тыс. ед.	Цена, д.е./ед. прод.	Продано, тыс. ед.
1	10,0	175,50	10,7	200,00
2	14,0	38,45	14,2	40,04
3	8,4	118,22	8,3	150,25

Все необходимые расчеты производим в табл. 7.9.

Таблица 7.9

Товар	Регион А		Регион В		Q	$p_A \cdot Q$	$p_B \cdot Q$
	p_A	q_A	p_B	q_B			
1	10,0	175,50	10,7	200,00	375,50	3755,000	4017,850
2	14,0	38,45	14,2	40,04	78,49	1098,860	1114,558
3	8,4	118,22	8,3	150,25	268,47	2255,148	2228,301
Σ	-	-	-	-	-	7109,008	7360,709

Тогда, территориальный индекс цен

$$I_p^{B/A} = \frac{7360{,}709}{7109{,}008} \approx 1{,}0354$$ или 103,54%,

что говорит о том, что цены на товары в регионе В на 3,54% выше, чем в регионе А.

В рассмотренном территориальном индексе не учитывается соотношение весов сравниваемых территорий. Этот недостаток устранен в формуле:

$$I_p^{B/A} = \frac{\sum p_B \cdot q_B}{\sum \overline{p} \cdot q_B} : \frac{\sum p_A \cdot q_A}{\sum \overline{p} \cdot q_A},$$

где

$$\overline{p} = \frac{\sum p \cdot q}{Q} = \frac{p_A \cdot q_A + p_B \cdot q_B}{q_A + q_B}.$$

Используя данный метод, определяется индекс физического объема реализации:

$$I_q^{B/A} = \frac{\sum \overline{p} \cdot q_B}{\sum \overline{p} \cdot q_A}.$$

Пример 7.7. По данным примера 7.6 найти территориальный индекс цен и индекс физического объема реализации.

Составим расчетную табл. 7.10.

Таблица 7.10

Товар	Регион А		Регион В		Q	$p_A \cdot q_A$	$p_B \cdot q_B$	$\overline{p}$	$\overline{p} \cdot q_A$	$\overline{p} \cdot q_B$
	p_A	q_A	p_B	q_B						
1	10,0	175,50	10,7	200,00	375,50	1755,000	2140,000	10,3728	1820,4264	2074,5600
2	14,0	38,45	14,2	40,04	78,49	538,300	568,568	14,1020	542,2219	564,6441
3	8,4	118,22	8,3	150,25	268,47	993,048	1247,075	8,3440	986,4277	1253,6860
Σ	-	-	-	-	-	3286,348	3955,643		3349,0760	3892,8901

Получаем:

$$I_p^{B/A} = \frac{3955{,}643}{3892{,}8901} : \frac{3286{,}348}{3349{,}0760} \approx 1{,}0355 \text{ или } 103{,}55\%.$$

Итак, средняя цена на товары в регионе В выше на 3,55%, чем в регионе А.

Индекс физического объема реализации равен:

$$I_q^{B/A} = \frac{3892{,}8901}{3349{,}0760} \approx 1{,}1624 \text{ или } 116{,}24\%.$$

Следовательно, реализовано товара в регионе В в среднем на 16,24% больше, чем в регионе А.

7.6. Цепные индексы

При исследовании динамики социально – экономических явлений применяются цепные индексы. Приведем систему индексов на примере сводных индексов цен за n периодов.

1. Цепные индексы цен с переменными весами:

$$I_p^{1/0}=\frac{\sum p_1\cdot q_1}{\sum p_0\cdot q_1},\ I_p^{2/1}=\frac{\sum p_2\cdot q_2}{\sum p_1\cdot q_2},\ldots,\ I_p^{n/n-1}=\frac{\sum p_n\cdot q_n}{\sum p_{n-1}\cdot q_n}.$$

2. Цепные индексы цен с постоянными весами.

$$I_p^{1/0}=\frac{\sum p_1\cdot q_0}{\sum p_0\cdot q_0},\ I_p^{2/1}=\frac{\sum p_2\cdot q_0}{\sum p_1\cdot q_0},\ldots,\ I_p^{n/n-1}=\frac{\sum p_n\cdot q_0}{\sum p_{n-1}\cdot q_0}.$$

3. Базисные индексы цен с переменными весами.

$$I_p^{1/0}=\frac{\sum p_1\cdot q_1}{\sum p_0\cdot q_1},\ I_p^{2/0}=\frac{\sum p_2\cdot q_2}{\sum p_0\cdot q_2},\ldots,\ I_p^{n/0}=\frac{\sum p_n\cdot q_n}{\sum p_0\cdot q_n}.$$

4. Базисные индексы цен с постоянными весами.

$$I_p^{1/0}=\frac{\sum p_1\cdot q_0}{\sum p_0\cdot q_0},\ I_p^{2/0}=\frac{\sum p_2\cdot q_0}{\sum p_0\cdot q_0},\ldots,\ I_p^{n/0}=\frac{\sum p_n\cdot q_0}{\sum p_0\cdot q_0}.$$

Пример 7.8. По данным о ценах и объемах продаж товаров A, B и C в 2015 – 2019 годах (табл. 7.11) рассчитать цепные индексы цен.

Таблица 7.11

Товар	2015		2016		2017		2018		2019	
	Цена, руб./ед. прод.	Продано, тыс. ед.	Цена, руб./ед. прод.	Продано, тыс. ед.	Цена, руб./ед. прод.	Продано, тыс. ед.	Цена, руб./ед. прод.	Продано, тыс. ед.	Цена, руб./ед. прод.	Продано, тыс. ед.
A	120	187	123	230	127	300	126	170	125	150
B	89	128	90	127	96	128	90	130	94	142
C	20	250	19	300	17	287	17	320	14	350

Эмпирические данные представим с учетом обозначений (табл. 7.12), в качестве базисного периода принимаем 2015 год.

Таблица 7.12

Товар	**2015** *n=0*		**2016** *n=1*		**2017** *n=2*		**2018** *n=3*		**2019** *n=4*	
	p_0	q_0	p_1	q_1	p_2	q_2	p_3	q_3	p_4	q_4
A	120	187	123	230	127	300	126	170	125	150
B	89	128	90	127	96	128	90	130	94	142
C	20	250	19	300	17	287	17	320	14	350

Все необходимые расчеты производим в табл. 7.13.

Таблица 7.13

Товар	$p_0 \cdot q_0$	$p_0 \cdot q_1$	$p_0 \cdot q_2$	$p_0 \cdot q_3$	$p_0 \cdot q_4$	$p_1 \cdot q_0$	$p_1 \cdot q_1$	$p_1 \cdot q_2$
A	22440	27600	24000	20400	18000	23001	28290	24600
B	11392	11303	11392	11570	12638	11520	11430	11520
C	5000	6000	5740	6400	7000	4750	5700	5453
Σ	38832	44903	41132	38370	37638	39271	45420	41573

Таблица 7.13 (продолжение)

Товар	$p_2 \cdot q_0$	$p_2 \cdot q_2$	$p_2 \cdot q_3$	$p_3 \cdot q_0$	$p_3 \cdot q_3$	$p_3 \cdot q_4$	$p_4 \cdot q_0$	$p_4 \cdot q_4$
A	23749	25400	21590	23562	21420	18900	23375	18750
B	12288	12288	12480	11520	11700	12780	12032	13348
C	4250	4879	5440	4250	5440	5950	3500	4900
Σ	40287	42567	39510	39332	38560	37630	38907	36998

Получаем:

1. Цепные индексы цен с переменными весами:

$$I_p^{1/0} = \frac{45420}{44903} \approx 1{,}0115,\ I_p^{2/1} = \frac{42567}{41573} \approx 1{,}0239,\ I_p^{3/2} = \frac{38560}{39510} \approx 0{,}9760,$$

$$I_p^{4/3} = \frac{36998}{37630} \approx 0{,}9832.$$

2. Цепные индексы цен с постоянными весами.

$$I_p^{1/0} = \frac{39271}{38832} \approx 1{,}0113,\ I_p^{2/1} = \frac{40287}{39271} \approx 1{,}0259,\ I_p^{3/2} = \frac{39332}{40287} \approx 0{,}9763,$$

$$I_p^{4/3} = \frac{38907}{39332} \approx 0{,}9892.$$

3. Базисные индексы цен с переменными весами.

$$I_p^{1/0} = \frac{45420}{44903} \approx 1{,}0115,\ I_p^{2/0} = \frac{42567}{41132} \approx 1{,}0349,\ I_p^{3/0} = \frac{38560}{38370} \approx 1{,}0050,$$

$$I_p^{4/0} = \frac{36998}{37638} \approx 0{,}9830.$$

4. Базисные индексы цен с постоянными весами.

$$I_p^{1/0} = \frac{39271}{38832} \approx 1{,}0113,\ I_p^{2/0} = \frac{40287}{38833} \approx 1{,}0374,\ I_p^{3/0} = \frac{39332}{38833} \approx 1{,}0128,$$

$$I_p^{4/0} = \frac{38907}{38833} \approx 1{,}0019.$$

ЗАДАЧИ ДЛЯ САМОСТОЯТЕЛЬНОЙ РАБОТЫ

Задача 1. По данным табл. 1 необходимо:

1) построить группированный статистический ряд;
2) начертить полигон частот, гистограмму и кумулятивную кривую;
3) найти квантиль порядка p = 0,3+0,01 k, где k – номер своего варианта;
4) по кумулятивной кривой найти вероятность $p(\alpha \le X \le \beta)$ нахождения значения случайной величины в интервале $(\alpha;\beta)$(данные см. в табл. 2);
5) по величине максимального и минимального элементов, а также объема выборки (табл. 3), найти число интервалов группировки и длину интервала.

Задача 2. По данным табл. 1 необходимо найти нижние и верхние квартили, децили, перцентили, моду и медиану. Дать анализ полученных результатов.

Задача 3. По данным табл. 1 следуя определению необходимо вычислить:

1) среднее значение;
2) среднее линейное отклонение;
3) дисперсию (произвести расчет дисперсии также и по формуле разностей);
4) получить несмещенную оценку дисперсии;
5) среднее квадратическое отклонение;
6) коэффициент вариации;
7) среднее относительное отклонение;
8) коэффициент осцилляции;
9) асимметрию;
10) эксцесс.

Задача 4. По данным табл. 1 методом моментов вычислить:

1) среднее значение;
2) дисперсию;
3) асимметрию;
4) эксцесс.

Сравнить результаты вычислений с результатами задачи 3.

Таблица 1

Вариант 1	**Интервал**	(4;5)	(5;6)	(6;7)	(7;8)	(8;9)
	Частота	1	3	10	5	1
Вариант 2	**Интервал**	(10,5;13,5)	(13,5;16,5)	(16,5;19,5)	(19,5;22,5)	(22,5;25,5)
	Частота	10	30	52	5	3
Вариант 3	**Интервал**	(10;12)	(12;14)	(14;16)	(16;18)	(18;20)
	Частота	5	10	15	13	7
Вариант 4	**Интервал**	(5,5;6,5)	(6,5;7,5)	(7,5;8,5)	(8,5;9,5)	(9,5;10,5)
	Частота	11	25	34	9	1
Вариант 5	**Интервал**	(0;2)	(2;4)	(4;6)	(6;8)	(8;10)
	Частота	5	10	15	7	3
Вариант 6	**Интервал**	(10;17)	(17;24)	(24;31)	(31;38)	(38;45)
	Частота	7	10	18	4	1
Вариант 7	**Интервал**	(5;10)	(10;15)	(15;20)	(20;25)	(25;30)
	Частота	2	6	7	3	2
Вариант 8	**Интервал**	(4,4;8,4)	(8,4;12,4)	(12,4;16,4)	(16,4;20,4)	(20,4;24,4)
	Частота	16	19	27	17	11
Вариант 9	**Интервал**	(5;25)	(25;45)	(45;65)	(65;85)	(85;105)
	Частота	3	6	8	5	3
Вариант 10	**Интервал**	(6;8)	(8;10)	(10;12)	(12;14)	(14;16)
	Частота	17	20	29	18	16
Вариант 11	**Интервал**	(1,5;2,5)	(2,5;3,5)	(3,5;4,5)	(4,5;5,5)	(5,5;6,5)
	Частота	14	27	39	18	2
Вариант 12	**Интервал**	(10;15)	(15;20)	(20;25)	(25;30)	(30;35)
	Частота	1	4	10	3	2
Вариант 13	**Интервал**	(6;16)	(16;26)	(26;36)	(36;46)	(46;56)
	Частота	2	3	20	15	10
Вариант 14	**Интервал**	(1;5)	(5;9)	(9;13)	(13;17)	(17;21)
	Частота	5	8	10	9	8
Вариант 15	**Интервал**	(4,5;6,5)	(6,5;8,5)	(8,5;10,5)	(10,5;12,5)	(12,5;14,5)
	Частота	1	2	12	4	1
Вариант 16	**Интервал**	(1;6)	(6;11)	(11;16)	(16;21)	(21;26)
	Частота	5	15	20	8	2
Вариант 17	**Интервал**	(6;7)	(7;8)	(8;9)	(9;10)	(10;11)
	Частота	10	15	20	4	1
Вариант 18	**Интервал**	(3;11)	(11;19)	(19;27)	(27;35)	(35;43)
	Частота	1	10	15	13	11
Вариант 19	**Интервал**	(1;7)	(7;13)	(13;19)	(19;25)	(25;31)
	Частота	1	6	14	3	1
Вариант 20	**Интервал**	(10;13)	(13;16)	(16;19)	(19;22)	(22;25)
	Частота	2	10	7	5	1

Таблица 2

Варианты										
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
α	5,4	17	12,4	7,0	2,3	21	10,5	9	32	9,7
β	8,7	21	16,5	10,2	9,8	44	21,5	23	71	15,1
Варианты										
	11	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
α	3,15	16,5	20	10,5	6,75	9,3	7,4	18	9	15,5
β	5,75	31,5	37,5	20	11,3	22	10,8	40	30	20,5

Таблица 3

Варианты										
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
$x_{\min}$	12	10,5	1,0	4	6	1	6,0	7	4,75	100
$x_{\max}$	17	38,5	5,6	24	51	11	8,5	22	9,75	115
n	50	80	200	80	100	30	40	100	100	40
Варианты										
	11	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
$x_{\min}$	1	4,5	4	6,5	1,5	3,9	25	14,0	3,25	1,3
$x_{\max}$	601	11,5	44	8,4	11,3	25,6	140	72,5	10	5,4
n	20	25	115	25	40	80	200	100	25	40

Задача 5. Для выборок и табл. 1 необходимо:

1.с вероятностью $\gamma = 0{,}85 + 0{,}0015\ k$ (k – номер своего варианта) найти интервал, в котором заключено математическое ожидание a (расчет произвести для случая а) повторного, б) бесповторного отбора из генеральной совокупности объема $N = 250 + 10\ k$, k – номер своего варианта);

2. с вероятностью $\gamma = 0{,}85 + 0{,}0025\ k$ (k – номер своего варианта) определить границы интервала, в котором заключена генеральная доля признака p, m – число элементов выборочной совокупности, принадлежащих интервалам со второго по четвертый включительно (расчет произвести для случая а) повторного, б) бесповторного отбора из генеральной совокупности объема $N = 250 + 15\ k$, k – номер своего варианта);

3.каков должен быть объем выборочной совокупности, если найденное значение дисперсии уменьшить на 10%, вероятность γ –

увеличить на 20%, а предельную ошибку – уменьшить в 1,35 раза (расчет произвести для случая а) повторного, б) бесповторного отбора из генеральной совокупности объема $N = 250 + 20\ k$, k – номер своего варианта).

Задача 6. С уровнем значимости $\alpha = 0{,}05$ проверить гипотезу о нормальном распределении генеральной совокупности, которой принадлежит выборка из табл. 1.

Задача 7. Дана зависимость между признаками X и Y. Необходимо:

1) произвести все необходимые вычисления (рассчитать среднее значение и показатели вариации по определению и методом моментов);
2) построить эмпирические линии регрессии и сделать первоначальные выводы о форме корреляционной связи;
3) определить величину коэффициента линейной корреляции (по определению и методом моментов) и сделать выводы о форме корреляционной зависимости;
4) найти значение корреляционного отношения и сделать выводы о тесноте корреляционной связи;
5) с вероятностью 0,95 проверить гипотезу о статистической значимости эмпирических данных;
6) установить вид уравнения регрессии y на x и x на y в предположении прямой (расчет коэффициентов произвести двумя способами), параболической и показательной регрессионной моделей;
7) с помощью величины средней ошибки аппроксимации и индекса детерминации отобрать наиболее точную модель;
8) построить на одном чертеже эмпирические данные и линии регрессии;
9) произвести прогноз значения y по заданному значению x и спрогнозировать величину x по y.

Вариант 1

$y \backslash x$	10	15	20	25	30	n_y
1	2					2
2	5	3				8
3	4	1	10			15
4		10	5	4		19
5				5	1	6
n_x	11	14	15	9	1	50

x=36,7; y=5,3

Вариант 2

$y \backslash x$	0	1	2	3	4	n_y
11					1	1
12		1		10		11
13	5	6	7	2		20
14	2	5	8			15
15	3					3
n_x	10	12	15	12	1	50

x=6,2; y=15,3

Вариант 3

$y \backslash x$	1,3	2,3	3,3	4,3	5,3	n_y
2,5				1	5	6
12,5			7	3		10
22,5		2	4			6
32,5		6				6
42,5	10	2				12
n_x	10	10	11	4	5	40

x=1,1; y=50,4

Вариант 4

$y \backslash x$	11	13	15	17	19	n_y
90	3					3
91	5	7				12
92	2	1	5			8
93	5	2	4	5		16
94			3		8	11
n_x	15	10	12	5	8	50

x=20,6; y=100,3

Вариант 5

$y \backslash x$	6	16	26	36	46	n_y
10				5	3	8
20			6	3		9
30		2	2	4		8
40	9	7				16
50	1	8				9
n_x	10	17	8	12	3	50

x=50,5; y=51,2

Вариант 6

$y \backslash x$	10	13	16	19	22	n_y
0,01	10	11				21
0,02	15	13				28
0,03		14	6			20
0,04		2	4	20		26
0,05					5	5
n_x	25	40	10	20	5	100

x=7,5; y=0,36

Вариант 7

$y \backslash x$	0,1	0,2	0,3	0,4	0,5	n_y
10					5	5
20			3	5		8
30		7	10			17
40	3	11	2			16
50	4					4
n_x	7	18	15	5	5	50

x=0,58; y=52,3

Вариант 8

$y \backslash x$	14	20	26	32	38	n_y
0,08	3					3
0,09	7	5				12
0,10		5	4			9
0,11		2	6	2		10
0,12					6	6
n_x	10	12	10	2	6	40

x=39,1; y=0,15

Вариант 9

$y \backslash x$	0,01	0,04	0,07	0,10	0,13	n_y
3	7					7
4	7					7
5	6	3	1			10
6	5	2	3	5		15
7				5	6	11
n_x	25	5	4	10	6	50

x=0,21; y=7,7

Вариант 10

$y \backslash x$	0,2	0,3	0,4	0,5	0,6	n_y
0				3	3	6
10			7	7		14
20		3	8	3		14
30	5	5	2			12
40	2	2				4
n_x	7	10	17	13	3	50

x=0,8; y=42,3

Вариант 11

$y \backslash x$	0	0,15	0,30	0,45	0,60	n_y
15					3	3
16					7	7
17			5	4		9
18		4	4	6		14
19	2	4	11			17
n_x	2	8	20	10	10	50

x=0,72; y=20,1

Вариант 12

$y \backslash x$	3,0	6,5	10,0	13,5	16,5	n_y
4,5	4					4
4,7	6	5				11
4,9	5	5	4			14
5,1			1	7		8
5,3					3	3
n_x	15	10	5	7	3	40

x=18,2; y=5,5

Вариант 13

$y \backslash x$	0,11	0,12	0,13	0,14	0,15	n_y
11			7	3	2	12
13			5	5		10
15		2	8			10
17	2	3				5
19	3					3
n_x	5	5	20	8	2	40

x=0,18; y=19,7

Вариант 14

$y \backslash x$	0,01	0,14	0,27	0,40	0,53	n_y
12,0	3	5				8
13,5		7	10			17
15,0		3	3			6
16,5			2	3		5
18,0					4	4
n_x	3	15	15	3	4	40

x=0,06; y=18,4

Вариант 15

$y \backslash x$	2	7	12	17	22	n_y
0,25				3	4	7
0,50			7	2	1	10
0,75		3	8	3		14
1,00		5				5
1,25	2	2				4
n_x	2	10	15	8	5	40

x=23,7; y=15

Вариант 16

$y \backslash x$	4,2	4,3	4,4	4,5	4,6	n_y
11	10	7	5			22
21		4	6			10
31		1	5	2		8
41			4	2	1	7
51				2	1	3
n_x	10	12	20	6	2	50

x=5,4; y=54,5

Вариант 17

$y \backslash x$	1	6	11	16	21	n_y
2				2	1	3
10			13	2		15
18			6	5		11
26		7	3			10
34	7	3	1			11
n_x	7	10	23	9	1	50

x=22; y=44,6

Вариант 18

$y \backslash x$	6,4	6,5	6,6	6,7	6,8	n_y
21,5	4	3				7
22,0	8	6	11			25
22,5		4	5	10		19
23,0		5	14	5	7	31
23,5			10	4	4	18
n_x	12	18	40	19	11	100

x=7,5; y=25

Вариант 19

$y \backslash x$	0,4	0,6	0,8	1,0	1,2	n_y
11,1			5	1	2	8
11,2		4	4	7	4	19
11,3		2	1	6		9
11,4	2	1	7			10
11,5	3	1				4
n_x	5	8	17	14	6	50

x=2; y=12,3

Вариант 20

$y \backslash x$	20	30	40	50	60	n_y
1,14	9					9
1,19		5				5
1,24		3	5			8
1,29		2	5	2		9
1,34			1	5	3	9
n_x	9	10	11	7	3	40

x=61,5; y=1,5

Задача 8. Дана зависимость между признаками X_1, X_2 и Y (табл. 4). Необходимо:

1) вычислить множественный коэффициент корреляции и сделать выводы о форме и силе корреляционной зависимости;
2) с помощью F – критерия Фишера с вероятностью 0,95 оценить статистическую значимость эмпирических данных;
3) вычислить значение общего индекса детерминации;
4) двумя способами получить уравнение линейной модели множественной регрессии;
5) по величине средней ошибки аппроксимации оценить точность линейной модели;
6) подсчитать дельта – коэффициенты;
7) найти значения коэффициентов эластичности;
8) исключить из модели один из факторных признаков и перейти к модели с парной регрессией.

Таблица 4

Варианты										
Y	1		2		3		4		5	
	X_1	X_2	X_1	X_2	X_1	X_2	X_1	X_2	X_1	X_2
11,3	0,21	1	13	7,5	0,07	1,6	0,9	23	10	4,8
14,2	0,35	6	14	8,2	0,06	1,7	0,7	15	9	3,5
13,6	0,33	8	16	8,6	0,08	1,9	0,4	17	6	2,1
14,3	0,35	5	17	8,7	0,09	2,3	0,5	16	3	2,7
15,1	0,40	7	19	8,8	0,12	2,1	0,3	14	1	1,8
Варианты										
Y	6		7		8		9		10	
	X_1	X_2	X_1	X_2	X_1	X_2	X_1	X_2	X_1	X_2
17,5	21	1,7	2,01	4,15	1,01	11,5	1,04	32	4,40	3,15
12,3	30	1,6	1,82	4,20	1,15	11,8	2,00	33	4,00	3,16
6,3	76	1,8	1,70	4,35	1,14	11,9	4,15	30	2,15	3,19
4,0	84	1,9	1,45	4,62	1,21	15,2	6,01	39	1,75	3,25
0,6	120	2,0	1,33	4,99	1,20	17,3	6,24	45	0,80	3,71
Варианты										
Y	11		12		13		14		15	
	X_1	X_2	X_1	X_2	X_1	X_2	X_1	X_2	X_1	X_2
1,11	1,60	0,75	1,32	75	14,4	5,18	7,17	13,13	71,5	3,15
1,35	1,90	0,76	1,44	84	14,6	5,01	7,95	13,15	31,4	3,14
1,70	3,00	0,77	1,71	35	12,2	4,32	8,54	13,20	20,8	3,00
1,91	3,17	0,93	1,99	48	10,4	4,00	10,01	13,19	10,3	2,50
2,23	3,19	1,01	2,01	33	8,8	3,97	14,30	13,18	0,8	2,49
Варианты										
Y	16		17		18		19		20	
	X_1	X_2	X_1	X_2	X_1	X_2	X_1	X_2	X_1	X_2
41,5	1,13	10,5	13	0,11	31,5	8,81	35	5,7	0,15	3,4
35,7	0,99	8,4	12	0,10	32,3	8,95	33	5,3	0,83	3,9
28,8	0,85	7,5	10	0,09	33,0	9,23	29	5,0	1,51	4,3
24,9	0,84	5,8	7	0,06	35,4	9,24	28	4,5	2,15	4,9
19,7	0,63	4,2	3	0,05	37,0	9,25	24	4,4	2,25	5,3

Задача 9. Деятельность предприятия в 2019 году характеризовалась данными, помещенными в таблицу 5.

Каждому варианту соответствует свой показатель.

Вариант 1. Уровень у - среднесписочная численность работающих, тыс. чел.

Вариант 2. Уровень y – количество занятых на производстве (на конец месяца), чел.

Вариант 3. Уровень y – число принятых на работу (на конец месяца), чел.

Вариант 4. Уровень y –число уволенных (на конец месяца), чел.

Вариант 5. Уровень y – количество прогулов без уважительных причин, чел. час.

Вариант 6. Уровень y – среднемесячная заработная плата, тыс. руб.

Вариант 7. Уровень y – среднемесячный размер премии, руб.

Вариант 8. Уровень y – среднемесячные отчисления на социальные нужды, тыс. руб.

Вариант 9. Уровень y – уровень компьютеризации (на конец месяца), комп./тыс. чел.

Вариант 10. Уровень y – площадь производственных помещений (на начало месяца), тыс. кв. м.

Вариант 11. Уровень y – площадь административных и хозяйственных помещений (на начало месяца), тыс. кв. м.

Вариант 12. Уровень y – площадь складских помещений (на начало месяца), тыс. кв. м.

Вариант 13. Уровень y – основные производственные фонды (на конец месяца), млн. руб.

Вариант 14. Уровень y – объем выпускаемой продукции, млн. руб.

Вариант 15. Уровень y – прибыль, млн. руб.

Вариант 16. Уровень y – число бракованных изделий, ед.

Вариант 17. Уровень y – издержки производства (на конец месяца), тыс. руб.

Вариант 18. Уровень y – общепроизводственные расходы (на конец месяца), тыс. руб.

Вариант 19. Уровень y – управленческие расходы (на конец месяца), тыс. руб.

Вариант 20. Уровень y – амортизация основных фондов (на конец месяца), тыс. руб.

Для каждого ряда динамики (табл. 5) необходимо:

1.определить тип ряда динамики;

2.произвести анализ уровней ряда динамики цепным и базисным способами (за базисный принять уровень января 2019г.);

3.рассчитать средние характеристики уровней ряда динамики.

Задача 10. Для своего ряда динамики (табл. 5) необходимо:

1) установить вид линейного, параболического и показательного трендов; результаты представить графически;

2) найти индексы сезонности;

3) построить модель неслучайной составляющей ряда динамики тремя способами (с помощью функции тренда и индексов сезонности, в виде линейной и показательной моделей в которых используется средний абсолютный прирост и средний коэффициент роста, а также в виде уравнения Фурье, число гармоник принять равными 1, 2 и 3), результаты представить графически;

4) определить наиболее точную модель неслучайной составляющей, построенной в п. 3;

5) по найденной в п. 4 модели произвести точечный прогноз уровней ряда динамики на январь, февраль и март 2020 г.

6) с помощью интервального прогноза уточнить точечный прогноз из п. 5 (доверительную вероятность принять равной 0,95).

Таблица 5

Месяц	**Вариант (показатели деятельности предприятия)**									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Январь	1,200	1000	25	10	0,50	7,300	500	79	15	5,172
Февраль	1,250	850	20	7	0,70	7,381	480	81	23	5,178
Март	1,300	930	35	6	1,40	7,560	600	83	25	5,200
Апрель	1,380	980	40	15	0,50	8,320	650	78	26	5,215
Май	1,400	970	42	20	0,40	8,403	540	77	27	5,322
Июнь	1,390	953	35	4	0,90	7,700	800	95	32	5,417
Июль	1,370	940	30	10	0,30	7,050	560	115	44	5,417
Август	1,350	948	25	15	0,20	7,400	565	123	50	5,500
Сентябрь	1,500	997	150	95	0,85	7,703	810	149	76	5,800
Октябрь	1,680	1000	200	73	0,71	7,800	815	94	85	6,300
Ноябрь	1,900	1320	210	80	0,62	8,500	821	96	87	6,512
Декабрь	2,000	1450	215	97	0,95	9,230	930	100	92	7,020

Окончание табл. 5

Месяц	Вариант (показатели деятельности предприятия)									
	11	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
Январь	0,51	6,00	23,8	41,3	1,171	0,015	71,2	37,7	10,32	29,90
Февраль	0,52	6,20	23,9	35,2	1,182	0,019	72,4	39,5	11,37	29,95
Март	0,53	6,15	24,0	34,3	1,199	0,023	73,8	42,7	11,42	29,96
Апрель	0,45	6,14	24,2	33,7	1,200	0,025	72,4	45,8	10,95	31,40
Май	0,48	6,15	25,0	36,2	1,205	0,031	72,3	47,9	15,60	31,41
Июнь	0,47	6,10	24,8	37,0	1,207	0,042	71,7	43,5	10,51	31,73
Июль	0,45	6,05	24,7	36,8	1,178	0,035	70,5	42,4	10,48	31,84
Август	0,45	6,03	24,4	37,1	1,164	0,031	69,4	41,9	10,74	31,97
Сентябрь	0,46	6,50	24,9	40,5	1,512	0,054	72,3	49,5	10,52	32,05
Октябрь	0,56	6,72	25,2	47,5	1,542	0,042	73,7	51,3	11,95	36,06
Ноябрь	0,71	8,84	25,9	49,9	1,637	0,059	75,8	53,2	12,40	38,74
Декабрь	0,92	9,10	26,3	59,8	1,724	0,071	81,4	55,1	19,98	39,15

Задача 11. По данным табл. 7 (варианты см. в табл. 6) необходимо:

1) вычислить индивидуальные и сводные индексы себестоимости, цен, объема, затрат и товарооборота;

2) определить сводные индексы цен и объема, используя средние взвешенные формулы, результаты сравнить с вычислениями п. 1;

3) определить величину экономии предприятия от изменения;

4) себестоимости и величину экономии покупателя от изменения цен.

Задача 12. По данным табл. 7 (варианты см. в табл. 6) необходимо:

1) вычислить индивидуальный индекс производительности труда;

2) определить сводный индекс производительности труда, взвешенный по трудоемкости;

3) найти сводный индекс производительности труда, взвешенный по выработке;

4) подсчитать средний индекс производительности труда, взвешенный по трудоемкости.

Задача 13. Имеются данные о продаже товара в трех регионах (варианты см. в табл. 8). Необходимо определить индексы цен переменного и фиксированного состава, а также индекс структурных сдвигов. Дать анализ полученных результатов.

Задача 14. По ценам и объемам реализации трех групп товаров (варианты см. в табл. 9) в двух регионах А и Б найти территориальный индекс цен двумя способами и индекс физического объема реализации.

Таблица 6

Распределение вариантов в задачах 11 – 14

Вариант	**Задача 11**	**Задача 12**	**Задача 13**	**Задача 14**
	Товары	**Товары**	**Регионы**	**Товары**
1	1, 2, 3	4, 5, 6	A, B, C	IV, V, VI
2	1, 2, 4	3, 5, 6	A, B, D	III, V, VI
3	1, 2, 5	3, 4, 6	A, B, E	III, IV, VI
4	1, 2, 6	3, 4, 5	A, B, F	III, IV, V
5	1, 3, 4	2, 5, 6	A, C, D	II, V, VI
6	1, 3, 5	2, 4, 6	A, C, E	II, IV, VI
7	1, 3, 6	2, 4, 5	A, C, F	II, IV, V
8	1, 4, 5	2, 3, 6	A, D, E	II, III, VI
9	1, 4, 6	2, 3, 5	A, D, F	II, III, V
10	1, 5, 6	2, 3, 4	A, E, F	II, III, IV
11	2, 3, 4	1, 5, 6	B, C, D	I, V, VI
12	2, 3, 5	1, 5, 6	B, C, E	I, IV, VI
13	2, 3, 6	1, 4, 6	B, C, F	I, IV, V
14	2, 4, 5	1, 3, 6	B, D, E	I, III, VI
15	2, 4, 6	1, 3, 5	B, D, F	I, III, V
16	2, 5, 6	1, 3, 4	B, E, F	I, III, IV
17	3, 4, 5	1, 2, 6	C, D, E	I, II, VI
18	3, 4, 6	1, 2, 5	C, D, F	I, II, V
19	3, 5, 6	1, 2, 4	C, E, F	I, II, IV
20	4, 5, 6	1, 2, 3	D, E, F	I, II, III

Таблица 7

Показатели работы предприятия в 2018 – 2019 годах

Товар	**Себестоимость, руб./ шт.**		**Цена изделия, руб./шт.**		**Объем выпуска, тыс. шт.**		**Время на производство, ед. прод./ч.**	
	2018	**2019**	**2018**	**2019**	**2018**	**2019**	**2018**	**2019**
1	0,78	0,54	1,20	1,15	23	24	0,20	0,15
2	0,95	1,01	1,75	2,02	95	88	0,05	0,07
3	1,20	1,25	2,05	2,15	78	53	0,15	0,21
4	1,75	1,50	5,48	5,30	110	200	0,10	0,09
5	0,23	0,19	2,50	2,01	244	253	0,08	0,10
6	0,18	0,19	1,02	1,95	51	85	0,79	0,75

Таблица 8

Регион	**2018 г.**		**2019 г.**	
	Цена, руб./шт.	**Продано, тыс. шт.**	**Цена, руб./шт.**	**Продано, тыс. шт.**
A	3,40	120	3,52	145
B	5,80	71	4,95	68
C	0,85	234	1,15	250
D	2,54	95	2,52	87
E	1,95	50	2,15	64
F	10,25	15	9,99	21

Таблица 9

Товар	**Регион А**		**Регион Б**	
	Цена, у. е./шт.	**Продано, тыс. шт.**	**Цена, у. е./шт.**	**Продано, тыс. шт.**
I	0,11	115	0,08	141
II	0,45	71	0,97	97
III	1,25	14	1,14	13
IV	2,75	77	3,01	90
V	0,07	131	0,05	157
VI	43,40	21	43,50	29

ЗАКЛЮЧЕНИЕ

Пособие было составлено в соответствии с учебной программой дисциплины «Статистика». Авторы рассматривали основные понятия общей теории статистики как отрасли статистической науки, группировку статистических данных и ее роль в анализе информации, абсолютные, относительные и средние величины, статистические распределения, статистические гипотезы, выборочное наблюдение, ряды динамики и их использование в экономико-статистическом исследовании.

В качестве дидактического материала в пособии даны задачи для самостоятельного решения.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Аскеров, П. Ф. Общая и прикладная статистика : учеб. для студентов высшего профессионального образования / под общ. ред. Р.Н. Пахуновой. – М. : ИНФРА-М, 2017. – 272 с.

2. Бурова, О. А. Статистика : Сборник задач / О. А. Бурова. – 2-е изд. – М. : МИСИ-МГСУ, 2017. – 127 с.

3. Венецкий И. Г., Кильдишев Г. С. Основы теории вероятностей и математической статистики. – М. : Статистика, 1968. – 360 с.

4. Громыко, Г. Л. Теория статистики : практикум / Г.Л. Громыко. – 5-е изд., испр. и доп. – М. : ИНФРА-М, 2018. – 238 с. – (Высшее образование: Бакалавриат).

5. Даймейнд С. Мир вероятностей. – М. : Статистика, 1970. – 154 с.

6. Дубров А. М. Последовательный анализ в статистической обработке информации. – М. : Статистика, 1976. – 160 с.

7. Карасев А. И. Теория вероятностей и математическая статистика. – М. : Статистика, 1979. – 279 с.

8. Кельберт М. Я., Сухов Ю. М. Вероятность и статистика в примерах и задачах. Т. I. Основные понятия теории вероятностей и математической статистики : учеб. пособие. – 2-е изд. – М. : МЦНМО, 2017. – 486 с.

9. Кельберт М. Я., Сухов Ю. М. Вероятность и статистика в примерах и задачах. Т. 2: Марковские цепи как отправная точка теории случайных процессов. – М. : МЦНМО, 2017. – 560 с.

10. Кочетков Е. С., Смерчинская С. О., Соколов В. В. Теория вероятностей и математическая статистика : учебник. – 2-е изд., испр. и перераб. – М. : ФОРУМ : ИНФРА-М, 2018. – 240 с.

11. Крыньский Х. Э. Математика для экономистов. – М. : Статистика, 1970. – 580 с.

12. Математический энциклопедический словарь / гл. ред. Ю. В. Прохоров. – М. : Сов. энциклопедия, 1988. – 847 с.

13. Панова, А. В. Статистика туризма : учеб. пособие / А. В. Панова. – М. : НИЦ ИНФРА-М, 2016. - 248 с.

14. Радионова М. В., Сапожников П. Н., Макаров А. А. Теория вероятностей, математическая статистика в примерах, задачах и тестах : учеб. пособие. – М. : КУРС, НИЦ ИНФРА-М, 2016. – 496 с.

15. Теория статистики : учебник / под ред. проф. Г. Л. Громыко. – 3-е изд., перераб. и доп. – М. : ИНФРА-М, 2017. – 476 с. – (Высшее образование: Бакалавриат).

16. Тимофеева, И. Ю. Статистика. Ч. 1. Общая теория статистики : учеб. пособие / И. Ю. Тимофеева, Е. В. Лаврова, О. Е. Полякова. – М. : НИЦ ИНФРА-М, 2018. – 104 с.

17. Хастингс Н., Пикок Дж. Справочник по статистическим распределениям. – М. : Статистика, 1980. – 96 с.

18. Шапкин А. С., Шапкин В. А. Задачи с решениями по высшей математике, теории вероятностей, математической статистике, математическому программированию. – 8-е изд. – М. : Дашков и К, 2017. – 432 с.

19. Шумак, О. А. Статистика : учеб. пособие / О. А. Шумак, А. В. Гераськин. – М. : ИЦ РИОР : НИЦ Инфра-М, 2019. – 311 с. – (Высшее образование: Бакалавриат).

20. Яковлев, В. Б. Практикум по общей теории статистики : учеб. пособие / В. Б. Яковлев, О. А Яковлева. – М. : НИЦ ИНФРА-М, 2016. – 382 с.

ПРИЛОЖЕНИЯ

Таблица П1

Значения функции Лапласа $\Phi(x) = \dfrac{2}{\sqrt{2\pi}} \int\limits_0^x e^{-\frac{t^2}{2}} dt$

Целые и деся-тые доли X	Сотые доли x									
	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	0,0080	0,0160	0,0239	0,0319	0,0399	0,0478	0,0558	0,0638	0,0717
0,1	0,0797	0,0876	0,0955	0,1034	0,1113	0,1192	0,1271	0,1350	0,1428	0,1507
0,2	0,1585	0,1683	0,1174	0,1819	0,1897	0,1974	0,2025	0,2128	0,2205	0,2282
0,3	0,2358	0,2434	0,2510	0,2586	0,2661	0,2737	0,2812	0,2886	0,2960	0,3035
0.4	0,3108	0,3182	0,3255	0,3328	0,3401	0,3473	0,3545	0,3616	0,3668	0,3579
0,5	0,3829	0,3889	0,3969	0,4039	0,4108	0,4177	0,4245	0,4313	0,4338	0,4448
0,6	0,4515	0,4581	0,4647	0,4713	0,4778	0,4843	0,4907	0,4971	0,5035	0,5098
0,7	0,5161	0,5223	0,5285	0,5346	0,5407	0,5467	0,5527	0,5587	0,5646	0,5705
0,8	0,5763	0,5821	0,5878	0,5935	0,5991	0,6047	0,6102	0,6157	0,6211	0,6265
0,9	0,6319	0,6372	06424	0,6476	0,6528	0,6579	0,6629	0,6679	0,6729	0,6778
1,0	0,6827	0,6875	0,6923	0,6970	0,7017	0,7063	0,7109	0,7154	0,7199	0,7243
1,1	0,7287	0,7330	0,7373	0,7415	0,7457	0,7499	0,7540	0,7580	0,7620	0,7660
1,2	0,7699	0,7737	0,7775	0,7813	0,7850	0,7887	0,7923	0,7959	0,7994	0,8029
1,3	0,8064	0,8098	0,8132	0,8165	0,8198	0,8230	0,8262	0,8293	0,8324	0,8355
1,4	0,8385	0,8415	0,8444	0,8473	0,8501	0,8529	0,8557	0,8584	0,8611	0,8638
1,5	0,8664	0,8690	0,8715	0,8740	0,8764	0,8789	0,8812	0,8836	0,8859	0,8882
1,6	0,8904	0,8926	0,8948	0,8969	0,8990	0,9011	0,9031	0,9051	0,9070	0,9090
1,7	0,9109	0,9127	0,9146	0,9164	0,9181	0,9199	0,9216	0,9233	0,9249	0,9265
1,8	0,9281	0,9297	0,9312	0,9327	0,9342	0,9357	0,9371	0,9385	0,9399	0,9412
1,9	0,9426	0,9439	0,9451	0,9464	0,9476	0,9488	0,9500	0,9512	0,9523	0,9534
2,0	0,9545	0,9556	0,9566	0,9576	0,9586	0,9596	0,9606	0,9616	0,9625	0,9634
2,1	0,9643	0,9651	0,9660	0,9668	0,9676	0,9684	0,9692	0,9700	0,9707	0,9715
2,2	0,9722	0,9729	0,9736	0,9743	0,9749	0,9756	0,9762	0,9768	0,9774	0,9780
2,3	0,9786	0,9791	0,9797	0,9802	09807	0,9812	0,9817	0,9822	0,9827	0,9832
2,4	0,9836	0,9841	0,9845	0,9849	0,9853	0,9857	0,9861	0,9865	0,9869	0,9872
2,5	0,9876	0,9879	0,9883	0,9886	0,9889	0,9892	0,9895	0,9898	0,9901	0,9904
2,6	0,9907	0,9910	0,9912	0,9915	0,9917	0,9920	0,9922	0,9924	0,9926	0,9928
2,7	0,9931	0,9933	0,9935	0,9937	0,9939	0,9940	0,9942	0,9944	0,9946	0,9947
2,8	0,9949	0,9951	0,9952	0,9953	0,9955	0,9956	0,9958	0,9959	0,9960	0,9961
2,9	0,9963	0,9964	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972
3,0	0,9973	0,9974	0,9975	0,9976	0,9976	0,9977	0,9978	0,9979	0,9979	0,9980

Таблица П1. Окончание

Целые и десятые доли X	сотые доли x									
	0	1	2	3	4	5	6	7	8	9
3,1	0,9981	0,9981	0,9982	0,9983	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986
3,2	0,9986	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990
3,3	0,9990	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993
3,4	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995	0,9995
3,5	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9997	0,9997
3,6	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9998	0,9998	0,9998
3,7	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998
3,8	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,9	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
4,0	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999

Таблица П2

Некоторые значения $\chi^2_{крит} = \chi(\alpha;\nu)$ критерия Пирсона

ν / α	1	2	3	4	5	6	7	8	9	10
0,05	3,84	5,99	7,82	9,49	11,1	12,6	14,1	15,5	16,9	18,3
0,01	6,64	9,21	11,3	13,3	15,1	16,8	18,5	20,1	21,7	23,2

Таблица П3

Таблица значений функции Пуассона $p(x) = \frac{\lambda^m}{m!} \cdot e^{-\lambda}$

λ / m	0,1	0,2	0,3	0,4	0,5	0,6
0	0,904837	0,818731	0,740818	0,670818	0,606531	0,548812
1	0,090484	0,163746	0,222245	0,268128	0,203165	0,329287
2	0,004524	0,16375	0,033337	0,053626	0,075816	0,098786
3	0,000151	0,001091	0,003334	0,007150	0,012636	0,019757
4	0,000004	0,000055	0,000250	0,000715	0,001580	0,002964
5	0,000000	0,000002	0,000015	0,000057	0,000158	0,000356
6	0,000000	0,000000	0,000001	0,000004	0,000013	0,000035
7	0,000000	0,000000	0,000000	0,000000	0,000001	0,000003

Таблица П3. Продолжение

m \ λ	0,7	0,8	0,9	1,0	2,0	3,0
0	0,496586	0,449329	0,406570	0,367879	0,135335	0,049787
1	0,347610	0,359463	0,365913	0,367879	0,2701671	0,149361
2	0,121663	0,143785	0,164661	0,183940	0,270671	0,224042
3	0,028388	0,038343	0,049398	0,061313	0,180447	0,224042
4	0,004968	0,007669	0,011115	0,015328	0,090324	0,168031
5	0,000695	0,001227	0,002001	0,003066	0,036089	0,100819
6	0,000081	0,000164	0,000300	0,000511	0,012030	0,050409
7	0,000008	0,000019	0,000039	0,000073	0,003437	0,021604
8	0,000001	0,000002	0,000004	0,000009	0,000859	0,008102
9	0,000000	0,000000	0,000000	0,000001	0,000191	0,002701
10	0,000000	0,000000	0,000000	0,000000	0,000038	0,000810
11	0,000000	0,000000	0,000000	0,000000	0,000007	0,000221
12	0,000000	0,000000	0,000000	0,000000	0,000001	0,000055
13	0,000000	0,000000	0,000000	0,000000	0,000000	0,000013
14	0,000000	0,000000	0,000000	0,000000	0,000000	0,000003
15	0,000000	0,000000	0,000000	0,000000	0,000000	0,000001

Таблица П3. Окончание

m \ λ	4,0	5,0	6,0	7,0	8,0	9,0	10,0
0	0,018316	0,006738	0,002479	0,000192	0,000335	0,000123	0,000045
1	0,073263	0,033690	0,014873	0,006383	0,002684	0,001111	0,000454
2	0,146525	0,084224	0,044618	0,022341	0,010735	0,004998	0,002270
3	0,195367	0,140374	0,089235	0,052129	0,028626	0,014994	0,007567
4	0,195367	0,175467	0,133853	0,091226	0,057252	0,033737	0,018917
5	0,156293	0,175467	0,160623	0,127717	0,091604	0,060727	0,037833
6	0,104194	0,146223	0,160623	0,149003	0,122138	0,091090	0,063055
7	0,059540	0,104445	0,137677	0,149003	0,139587	0,117116	0,090079
8	0,029770	0,065278	0,103278	0,103377	0,139587	0,131756	0,112600
9	0,013231	0,036266	0,068838	0,101405	0,124077	0,131756	0,125110
10	0,005292	0,018133	0,041303	0,070983	0,099662	0,118580	0,125110
11	0,001925	0,008242	0,022529	0,045171	0,072190	0,097020	0,113740
12	0,000642	0,003434	0,011262	0,026350	0,048127	0,072765	0,094780
13	0,000197	0,001321	0,005199	0,014181	0,029616	0,050376	0,072908
14	0,000056	0,000472	0,002228	0,007094	0,016924	0,032384	0,052077
15	0,000015	0,000157	0,000891	0,003311	0,009026	0,019431	0,034718
16	0,000004	0,000049	0,000334	0,001448	0,004513	0,010930	0,021699
17	0,000001	0,000014	0,000118	0,000596	0,002124	0,005786	0,012764
18	0,000000	0,000004	0,000039	0,000232	0,000944	0,002893	0,007091
19	0,000000	0,000001	0,000012	0,000085	0,000397	0,001370	0,003732
20	0,000000	0,000000	0,000004	0,000030	0,000159	0,000617	0,001866
21	0,000000	0,000000	0,000001	0,000010	0,000061	0,000264	0,000889
22	0,000000	0,000000	0,000000	0,000003	0,000022	0,000108	0,000404
23	0,000000	0,000000	0,000000	0,000001	0,000008	0,000042	0,000176
24	0,000000	0,000000	0,000000	0,000000	0,000003	0,000016	0,000073
25	0,000000	0,000000	0,000000	0,000000	0,000001	0,000006	0,000029
26	0,000000	0,000000	0,000000	0,000000	0,000000	0,000002	0,000011
27	0,000000	0,000000	0,000000	0,000000	0,000000	0,000001	0,000004
28	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	0,000001

Таблица П4

Некоторые значения $t_{крит} = t(\alpha;\nu)$ критерия Стьюдента

α \ ν	1	2	3	4	5	6	7	8	9	10
0,05	12,70	4,30	3,18	2,78	2,57	2,45	2,36	2,31	2,26	2,23
0,01	63,70	9,92	5,84	4,60	4,03	3,71	3,50	3,36	3,25	3,17
α \ ν	11	12	13	14	15	16	17	18	19	20
0,05	2,20	2,18	2,16	2,14	2,13	2,12	2,11	2,10	2,09	2,09
0,01	3,11	3,05	3,01	2,98	2,95	2,92	2,90	2,88	2,86	2,85
α \ ν	21	22	23	24	25	26	27	28	29	30
0,05	2,08	2,07	2,07	2,06	2,06	2,06	2,05	2,05	2,05	2,04
0,01	2,83	2,82	2,81	2,80	2,79	2,78	2.77	2,76	2,76	2,75
α \ ν	40	60	80	120	∞					
0,05	2,02	2,00	1,99	1,98	1,96					
0,01	2,70	2,66	2,64	2,62	2,58					

Таблица П5

Некоторые значения $F_{крит} = F(0{,}05; \nu_1; \nu_2)$ критерия Фишера

ν_2 \ ν_1	1	2	3	4	5	6	8	12	24	∞
1	161,40	199,50	215,70	224,60	230,20	234,00	238,90	243,90	249,00	25.'UO
2	18,51	19,00	19,16	19,25	19,30	19,33	19,37	19,41	19,45	19,50
3	10,13	9,55	9,28	9,12	9,01	8,94	8,84	8,74	8,64	8,53
4	7,71	6.9!	6,59	6,39	6,26	6,16	6,04	5,91	5,77	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,82	4,68	4,53	4,30
6	5,99	5,14	4,76	4,53	4,39	4,28	4,15	4,00	3,84	3,67
7	5,59	4.74	4,35	4,12	3,97	3,87	3,73	3,57	3,41	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,44	3,28	3,12	2,99
9	5,12	4,20	3,86	3,63	3,48	3,37	3,23	3,07	2,90	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,07	2,91	2,74	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	2,95	2,79	2,61	2,40
12	4,75	3,88	3,49	3,26	3,11	3,00	2,85	2,69	2,50	2,30
13	4.67	3,80	3,41	3,18	3,02	2,92	2,77	2,60	2,42	2,21
14	4,60	3,7 I	3,34	3,11	2,96	2,85	2,70	2,53	2,35	2,13
15	4,45	3,68	3,29	3,06	2,90	2,79	2,64	2,48	2,29	2,07
16	4,41	3,63	3,24	3,01	2,83	2,74	2,59	2,42	2,24	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,55	2,38	2,19	1,96
18	4,41	3,5й	3,16	2,93	2,77	2,66	2,51	2,34	2,15	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,48	2,31	2,11	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,45	2,28	2,08	1,84
21	4,32	3,47	3,07	2,84	2,68	2,57	2,42	2,25	2,05	1,82
22	4,30	3,44	3,05	2,82	2,66	2,55	2,40	2,23	2,03	1,78
23	4,28	3,42	3,03	2,80	2,64	2,53	2,38	2,20	2,00	1,76
24	4,26	3,40	3,01	2,78	2,62	2,51	2,36	2,18	1,98	1,73
25	4,24	3,38	2,99	2,76	2,00	2,49	2,34	2,16	1,96	1,71
26	4,22	3,37	2,98	2,74	2,59	2,47	2,32	2,15	1,95	1,69
27	4,21	3,35	2,96	2,73	2,57	2,46	2,30	2,13	1,93	1,67
28	4,19	3,34	2,95	2,71	2,56	2,44	2,29	2,12	1,91	1,65
29	4,18	3,33	2,93	2,70	2,54	2,43	2,28	2,10	1,90	1,64
30	4,17	3,32	2,92	2,69	2,53	2,42	2,27	2,09	1,89	1,62
60	4,00	3,15	2,70	2,52	2,37	2,25	2,10	1,92	1,70	1,39
∞	3,84	2,99	2,00	2,37	2,21	2,09	1,94	1,75	1,52	1,03

ОГЛАВЛЕНИЕ

Учебное издание

КРЫЛОВ Василий Евгеньевич
МУРАВЬЕВА Надежда Викторовна

ОБЩАЯ ТЕОРИЯ СТАТИСТИКИ

Учебное пособие

Издается в авторской редакции

Подписано в печать 22.05.20.
Формат 60×84/16. Усл. печ. л. 14,18. Тираж 50 экз.
Заказ
Издательство
Владимирского государственного университета
имени Александра Григорьевича и Николая Григорьевича Столетовых.
600000, Владимир, ул. Горького, 87.