Л.Н. Мамадалиева, И.М. Хаконова

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ БИОЛОГИЧЕСКИХ ПРОЦЕССОВ

Учебное пособие

Майкоп - 2019

УДК [330.44:57] (07)
ББК 22.18+28.0
М 22

Печатается по решению научно-технического совета
федерального государственного бюджетного образовательного
учреждения высшего образования
«Майкопский государственный технологический университет»

Рецензенты:
кандидат физико-математических наук, доцент Паланджянц Л.Ж.
кандидат физико-математических наук, доцент Шевякова О.П.

М-22 Мамадалиева Л.Н., Хаконова И.М. Математическое моделирование биологических процессов: учебное пособие. – Майкоп: ИП Кучеренко В. О., 2019 – 148 с.
ISBN 978-5-907004-42-9

Учебное пособие составлено в соответствии с ФГОС ВО и рабочей программой направления подготовки бакалавров «Ландшафтная архитектура». Содержит краткий обзор основных статистических понятий, подробные методические указания к решению задач прогнозирования, примеры применения соответствующего математического аппарата к решению конкретных биологических задач, вопросы для самоконтроля, задачи для самостоятельного решения, задания к расчетно-графической работе, тест, темы рефератов.

Пособие предназначено для преподавателей и студентов очной и заочной формы обучения.

ISBN 978-5-907004-42-9

УДК [330.44:57] (07)
ББК 22.18+28.0

ОГЛАВЛЕНИЕ

ПРЕДИСЛОВИЕ

Развитие современной науки и производства вызывает потребность в поиске новых современных методов, средств и форм обучения, которые выражают интересы личности, адекватные тенденциям интенсивного развития общества.

Квалификация бакалавра технологического профиля зависит от того, как он умеет ориентироваться в реальных условиях производства и рынка. Эти условия часто бывают подвержены случайным воздействиям. На конечный результат деятельности влияет большое число случайных и неконтролируемых факторов. Для того чтобы рационально выбрать конструктивные параметры системы управления, необходимо изучить ее реакцию на непрерывно воздействующие случайные возмущения, а единственным аппаратом, пригодным для такого исследования, является аппарат математики.

Поэтому естественно, что в учебные планы технологических направлений вузов вошло изучение дисциплины «Математическое моделирование биологических процессов», который логически дополняет и продолжает раздел «Теория вероятностей и математическая статистика».

Необходимо, чтобы содержание этого раздела:

а) отражало современные достижения науки и техники;

б) соответствовало квалификации выпускников – бакалавров-технологов;

в) учитывало бы массовый характер высшего образования;

г) помогало обучающимся рационально и творчески думать;

д) определяло возможность применять знания и умения, полученные при обучении моделированию биологических процессов, в работе над дипломными проектами и в дальнейшей практической деятельности;

е) учитывало возможность внедрения в учебный процесс более совершенных способов и средств обучения, повышающих качество и ускоряющих процесс обучения.

В связи с этим, в первой главе пособия рассматриваются вопросы сбора и первичной обработки экспериментальных данных. Моделирование динамического ряда бакалавры начинают с наблюдения или эксперимента. В результате этого появляются некоторые опытные данные. При обсуждении организации эксперимента должен быть раскрыт объект статистического изучения, следовательно:

- рассмотрены основные вопросы, касающиеся сбора первичных данных, которые будут подвергнуты систематизации и обобщению;

- выявлены основные принципы организации и проведения наблюдения;

- определена цель сбора статистической информации, которая предусматривает дальнейшее решение вопроса о выявлении наличия функциональной зависимости в исследуемом динамическом ряду от времени.

Без систематизации и обобщения разрозненных сведений в них бывает трудно разобраться. Поэтому предлагается изучение методики построения вариационных рядов, их графическое представление, расчет выборочных характеристик данных.

Вторая глава содержит теоретические сведения и практические задачи, связанные с рядами динамики – рядами показателей, изменяющихся во времени. Это эмпирические прообразы случайных процессов. Изучение основных характеристик динамических рядов дает возможность определить закономерности их развития.

Третья глава посвящена решению задач выявления основных тенденций развития динамического ряда, которые достигаются путем их обработки математическими методами, анализа изменения его уровней, выявления наличия трендовой зависимости, вычисления параметров трендов. После получения нескольких функциональных зависимостей возможно проверить точность моделей, их адекватность исследуемому процессу и расчет прогноза поведения процесса в будущем.

Четвертая глава охватывает вопросы применения математического аппарата систем массового обслуживания к построению вероятностных моделей биологических процессов.

В каждой главе содержатся подробные методические указания к решению задач прогнозирования, примеры применения соответствующего математического аппарата к решению конкретных биологических задач. А также вопросы для самоконтроля, задачи для самостоятельного решения, задания к расчетно-графической работе, тест, темы рефератов.

Пособие может использоваться преподавателями в качестве дидактических материалов для обучения бакалавров очной и заочной форм обучения одноименной дисциплине.

ВВЕДЕНИЕ

Целью освоения дисциплины Моделирование биологических процессов является формирование у бакалавров представлений об использовании идей и методов математики в современных технологиях, а также формирование у обучающихся математического мышления при работе с данными экологических исследований и экспериментов, знакомство с основными методами математической обработки биологических и экологических данных, приемами анализа, хранения и интерпретации биологической экологической информации, предсказание поведения процессов в будущем.

Математическое моделирование подразумевает создание описания процесса с помощью математических объектов (графиков, стохастических матриц, функций, уравнений и т.д.) для упрощения его исследования, получения о нем новых знаний, анализа и оценки возможностей принятия производственных решений. Построение математической модели необходимо для исследования характеристик системы математическими методами.

Задачами изучения дисциплины являются освоение методов количественной оценки параметров исследуемых процессов, формирование умений содержательно интерпретировать и анализировать полученные результаты, развитие навыков математического мышления, подготовка к применению математических методов для решения практических задач общего и профессионального характера.

В результате освоения данной дисциплины студент должен

- *знать:* цель, основные задачи и области применения методов математического моделирования в сфере биотехнических систем и технологий; особенности биологических объектов моделирования и методики экспериментальной оценки их свойств; классификацию моделей по свойствам;

- *уметь:* адекватно ставить задачи исследования и оптимизации сложных объектов на основе методов математического моделирования; осуществлять формализацию и алгоритмизацию функционирования исследуемой системы; выбирать класс модели и оптимизировать ее структуру в зависимости от поставленной задачи, свойств моделируемого объекта и условий проведения эксперимента; рассчитывать параметры и основные характеристики моделей любого из рассмотренных классов;

- *владеть:* навыками выбора адекватных методов исследования моделей; навыками принятия адекватных решений по результатам исследования моделей.

Дисциплина входит в перечень дисциплин по выбору, устанавливаемых вузом вариативной части математического и естественно-научного цикла. Ее изучение способствует формированию фундаментальных и прикладных знаний.

Для освоения дисциплины необходимы знания элементарной математики и информатики, изучаемые в курсе общеобразовательной школы, а также знания, полученные в курсе дисциплины «Математика». Знания, полученные при изучении данной дисциплины, требуются в дальнейшем для успешного овладения таких дисциплин как урбоэкология и мониторинг, ботаника, почвоведение, основы лесопаркового хозяйства, декоративная дендрология, декоративное растениеводство, дендрометрия и др.

В результате освоения дисциплины бакалавр должен обладать следующими компетенциями: способностью использовать основные законы естественнонаучных дисциплин в профессиональной деятельности; способностью применять современные методы исследования в области биологии; готовностью выполнить расчеты и проектирование деталей и узлов в соответствии с техническим заданием; владение современными методами обработки, анализа и интерпретации экологической информации при проведении исследований биологический процессов; способность обосновать выбор методов экспериментальной работы, интерпретировать и представить результаты экспериментов по биологическим ресурсам.

МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ ДЛЯ ПРЕПОДАВАТЕЛЕЙ

Изучение методов математического моделирования биологических процессов имеет большое значение не только для будущей профессиональной деятельности бакалавров, но и при выполнении курсовых, выпускных квалификационных работ, а также изучении дисциплин, использующих методы математики.

В результате изучения дисциплины студент должен получить ясное представление о том, какое место занимает математическое моделирование в изучении биологических процессов. Изучение дисциплины вооружит обучающихся пониманием основ математического моделирования биологических систем.

При рассмотрении вопросов, касающихся сбора первичных данных, обучающимся необходимо разъяснить основные принципы организации и проведения наблюдения, а также научить решать практические задачи, встающие перед наблюдателем. Первичные данные в дальнейшем систематизируются и обобщаются. Материалы, полученные в процессе наблюдения, могут быть положены в основу практических занятий по последующим темам курса (сводка и группировка, абсолютные и относительные величины, средние величины, графическое представление данных, выявление наличия трендовой зависимости, аналитическое нахождение параметров тренда и т.д.).

Первые аудиторные практические занятия целесообразно провести по докладной системе. Практика показывает, что бакалавры обычно испытывают затруднения в выступлениях на первых занятиях без специальной подготовки. Поэтому можно предложить темы рефератов с краткими докладами. Выступление с докладом мобилизует внимание всей аудитории, вызывает желание высказываться по обсуждаемым вопросам.

Практические занятия могут быть посвящены рассмотрению различных видов таблиц из периодической печати, Интернет-ресурсов или данных, составленных преподавателем и студентами. Следует рассмотреть методику компоновки статистических данных, представленных в текстовой форме, в табличную. Составленные вариационные ряды подвергаются дальнейшему статистическому изучению – вычислению показателей изменения уровней ряда, нахождению средних показателей. Полезно в процессе занятий не ограничиваться техникой построения таблиц, а предложить студентам проанализировать каждую из них и сформулировать практические выводы по объектам и процессам, рассмотренным в них.

После изучения каждой главы целесообразно провести контрольную работу в виде кратких ответов обучающихся на вопросы, помещенные в конце главы. По окончании изучения всего материала дисциплины преподаватель может предложить контрольные вопросы в виде тестов.

Задания для самостоятельной внеаудиторной работы могут состоять из проектирования какого-либо наблюдения, проведения наблюдения (например, замер длины листа растения); реферата о своей практической работе по статистике, рецензии на какую-либо статью по вопросам статистического наблюдения.

МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ ДЛЯ СТУДЕНТОВ

Советы по планированию и организации времени, необходимого на изучение раздела

В процессе изучения математического моделирования биологических процессов студенту предлагается прослушать курс лекций и посетить практические занятия. Лекционные занятия имеют целью изложить изучаемый материал, сообщить студентам систему определенных знаний, ориентированных на специфику профессиональной подготовки бакалавров. На практических занятиях предполагается решение практических задач, формирование навыков действий по заданным алгоритмам построения математических моделей исследуемых процессов и умений конструировать самостоятельно модели, отражающие суть явлений и процессов реальной действительности, обучить умениям создавать самостоятельно алгоритмы решения практических задач. Самостоятельная работа студента включает в себя самоконтроль знаний, полученных на лекционных и практических занятиях, домашняя подготовка к предстоящим практическим занятиям, коллоквиумам, воспроизведение по памяти определений, формулировок теорем, выводов, самостоятельный поиск в дополнительной литературе информации, необходимой для успешного освоения раздела.

Описание последовательности действий студента

После изучения определенной темы на лекционном занятии и решения достаточного количества практических задач студенту рекомендуется воспроизвести по памяти определения терминов, формулы, выводы. В случае необходимости надо еще раз разобраться в материале лекции, разыскать и усвоить дополнительные сведения из других источников, рекомендованных преподавателем, решить ряд задач.

Иногда недостаточность усвоения того или иного вопроса выясняется только при изучении дальнейшего материала. В этом случае рекомендуется вернуться назад и повторить плохо изученный раздел.

Важным критерием усвоения теории является умение решать задачи по изучаемым темам. Однако распространенной ошибкой является то, что благополучное решение задач воспринимается студентами как признак усвоения теории. Часто правильное решение задачи получается в результате применения механически заученных формул, без понимания сути процесса. Можно сказать, что умение решать задачи является необходимым, но недостаточным условием хорошего знания теории. В конечном итоге целью изучения дисциплины является усвоение системы определенных знаний, позволяющих:

- формировать научно обоснованные взгляды и убеждения;
- развивать логическое и вариативное мышление;
- приобрести умение принять решение в различных жизненных ситуациях, используя опыт, накопленный при решении математических задач;
- развить навыки анализа полученных результатов по обработке исследуемых процессов.

Рекомендации по использованию материалов учебного пособия

Содержание пособия позволяет эффективно организовать и поддерживать аудиторную и самостоятельную работу студента и сохранить преемственность в преподавании дисциплины.

Учебно-методические и учебные материалы, включаемые в пособие, отражают современный уровень развития науки, предусматривают логически последовательное изложение учебного материала, использование современных методов и технических средств интенсификации учебного процесса, позволяющих студентам глубоко осваивать учебный материал и получать навыки по его использованию на практике.

Рекомендации по работе с литературой

Изучая материал по учебному пособию, следует переходить к следующему вопросу только после правильного понимания предыдущего, производя на бумаге все вычисления (в том числе и те, которые ради краткости опущены).

Особое внимание следует обращать на определение основных понятий. Студент должен подробно разбирать примеры, которые поясняют определения, и уметь строить аналогичные примеры самостоятельно.

При изучении материала полезно вести конспект, в который рекомендуется вписывать определения, формулы, уравнения и т.д. на полях конспекта следует отмечать вопросы, выделенные для получения письменной или устной консультации преподавателя.

Письменное оформление работы имеет исключительно важное значение. Записи должны быть сделаны аккуратно и расположены в определенном порядке. Хорошее внешнее оформление конспекта по изученному материалу приучит к необходимому в работе порядку и позволит избежать многочисленных ошибок, которые происходят из-за небрежных беспорядочных записей.

Выводы, полученные в виде формул, рекомендуется подчеркнуть в конспекте или обвести в рамку, чтобы при перечитывании конспекта они выделялись и лучше запомнились. Многим студентам помогает в работе составление листа, содержащего важнейшие и наиболее часто употребляемые формулы. Такой лист не только помогает запомнить формулы, но и может служить постоянным справочником для студента.

Разъяснения по работе с тестовой системой

Тестовая система дает возможность определить соответствие уровня подготовки студентов требованиям ФГОС ВО по изучаемой математической дисциплине.

Тесты сгруппированы по темам, соответствующим ФГОС ВО по направлениям подготовки бакалавров; каждое задание имеет целью отработать вполне определенные навыки и понятия. Результаты тестов позволяют судить о степени усвоения студентами соответствующих разделов, наглядно показывают пробелы в знаниях студентов не только по всей теме в целом, но и по отдельным ее фрагментам, правилам, понятиям, на желательное направление дальнейшей работы; помогают сформулировать вопросы для постановки их перед преподавателем.

Тесты позволяют активизировать образовательную деятельность студентов на занятиях, разнообразить процесс образования, получать наглядную картину успеваемости, экономить время, отводимое на другие виды учебной деятельности. Задача тестирования – помочь студенту успешно усвоить программу раздела, а также научить четко и логически мыслить, владеть математическим языком.

Тестирование может предлагаться перед традиционными письменными контрольными работами для обобщения изученного материала.

Перед ответом на каждый вопрос студент должен определить и вспомнить соответствующий теоретический материал (формулы, определения). При необходимости выполнить письменные решения.

Не следует приступать к выполнению тестовых заданий, не решив достаточного количества задач по материалу, соответствующих этому заданию. Неумение правильно ответить на задание теста чаще всего вызывается именно невыполнением именно этого требования.

Тесты должны выполняться самостоятельно. Несамостоятельно выполненная работа не дает возможности преподавателю указать студенту на недостатки в его работе, в усвоении им учебного материала, в результате чего студент не приобретает необходимые знания и умения и может оказаться неподготовленным к устному зачету или экзамену.

Разъяснения по выполнению домашних заданий

Выполнение домашнего задания – одна из форм образовательной деятельности студента, которая способствует успешному усвоению изучаемого материала и в конечном итоге, помогает достижению целей и задач изучения раздела.

Домашнее задание следует начинать выполнять, изучив соответствующий раздел или тему лекции. При необходимости следует обратиться к учебнику, рекомендованному преподавателем. Рекомендуется воспроизвести по памяти определения, теоремы, выводы. В случае необходимости, нужно еще раз разобраться в изучаемом материале.

После этого можно приступать к выполнению домашнего задания. При решении задач нужно обосновывать каждый этап решения, исходя из теоретических положений курса. Если студент видит несколько путей решения, то он должен сравнить их и выбрать самый лучший. Полезно до начала вычислений составить краткий план решения.

Решение задач и примеров следует излагать подробно, вычисления располагать в стогом порядке, отделяя вспомогательные вычисления от основных.

Решение каждой задачи должно быть сначала в общем виде и сопровождаться выводом формулы. Затем в полученную формулу подставляют числовые значения (если они даны). В промежуточных вычислениях не следует вводить приближенные вычисления корней. Решение должно доводиться до ответа, требуемого условием.

ГЛАВА 1. ПРЕДВАРИТЕЛЬНАЯ ОБРАБОТКА ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ

1.1. Основные понятия

В статистике различают генеральную и выборочную совокупности. *Генеральная совокупность* - это множество однородных, но индивидуально различимых объектов. Например, при изучении числа ветвей у дуба, генеральной совокупностью будут данные о числе ветвей всех без исключения деревьев, растущих на планете. При статистическом изучении биологических явлений мы не сможем иметь дело с генеральной совокупностью в её полном объёме, а будем располагать лишь какой-то, как правило, очень небольшой, её частью. Такая отобранная тем или иным способом часть генеральной совокупности получила название *выборки*. На основании изучения свойств выборки исследователь формирует представление о свойствах генеральной совокупности. Основное требование, предъявляемое к любой выборке, это её *репрезентативность* (от лат. represento - представляю). Выборка считается репрезентативной, если она получена путём случайного отбора и правильно представляет пропорции генеральной совокупности.

Объем статистической совокупности - численность элементов совокупности, взятых для исследования.

Традиционно в статистике количество элементов генеральной совокупности обозначают буквой N, а число наблюдений, образующих выборку - буквой n.

Каждый отдельно взятый элемент данного множества называется *единицей статистической совокупности*. Единицы статистической совокупности характеризуются общими свойствами, именуемыми в статистике признаками. Признаки могут быть количественными (в совокупности деревьев в парке – количество лип, кленов и т.д.) и качественными (например, в совокупности парковых насаждений – деревья, кустарники, цветы). Качественная определенность совокупности устанавливается в каждом конкретном статистическом исследовании в соответствии с его целями и познавательными задачами. Отличие количественных признаков от качественных состоит в том, что первые можно выразить итоговыми значениями, вторые — только числом единиц совокупности. Количественные признаки можно разделить на дискретные и непрерывные.

Непрерывные признаки теоретически могут принимать любые возможные значения в пределах между минимальным и максимальным показателем признака. К ним относятся масса, линейные размеры и температура организма, содержание биохимических и неорганических веществ в его тканях и т.п. В качестве примера дискретных признаков можно привести количество лепестков растения. Отсюда следует, что непосредственно наблюдаемые значения дискретного признака могут характеризоваться лишь целыми числами, тогда как при непрерывном варьировании значения признака могут быть как целыми, так и дробными.

Исследование математическими методами различных процессов начинают с наблюдения или эксперимента. В результате этого появляются некоторые опытные данные. При организации эксперимента должен быть раскрыт объект статистического изучения. Следовательно

- рассмотрены основные вопросы, касающиеся сбора первичных данных, которые будут подвергнуты систематизации и обобщению;
- выявлены основные принципы организации и проведения наблюдения;
- определена цель сбора статистической информации.

Без систематизации и обобщения разрозненных сведений в них бывает трудно разобраться. Поэтому появляется необходимость представления результатов наблюдений в виде таблиц. Сведенные в таблицу данные позволяют подмечать некоторые характерные черты изучаемых процессов.

Грамотно составленные таблицы являются важным средством изложения и анализа обработанного статистического материала. Таблица представляет собой расположенные по определенной системе ряды чисел. Статистическая таблица дает возможность систематизировать изложение статистических данных. Сведенные в таблицу, они приобретают обозримый вид, появляется возможность на их основании делать те или иные выводы. Собранный в таблицу материал выглядит компактно. Основное преимущество табличной формы изложения статистических данных в том, что именно при помощи таблиц легче всего осуществить сравнение, сопоставление и анализ чисел.

Табличная запись данных наблюдений компактна, наглядна, удобна для дальнейшего изучения. Сведенные в таблицу данные (таблица 1.1) позволяют провести предварительный сравнительный анализ (можно изучить, имеется ли рост показателя в зависимости от времени, сделать предварительную оценку скорости роста показателя и т.д.).

Пример. Наблюдения за весенним прилетом уток в течение 2006-2010 годов.

Таблица 1.1 - Данные наблюдений

Год	Дата
2006	12.04
2007	18.04
2008	2.05

2009	16.04
2010	28.04

Если в статистической таблице расположить наблюдаемые данные в хронологической последовательности, можно получить временной динамический ряд, в котором каждому моменту или периоду времени соответствует статистический показатель, характеризующий изучаемый объект.

1.2. Вариационные ряды

Подвергая варьирующий признак статистическому анализу, мы отвлекаемся от конкретного содержания и оперируем лишь абстрактными значениями. При этом смысл анализа сводится к выявлению общих свойств генеральной совокупности, отраженных в конкретной выборке. В связи с этим может оказаться полезной группировка первичных данных. Группировка считается успешной, если удовлетворяет следующим требованиям:

1) в ходе группировки не теряется важная информация;

2) в результате группировки первичные данные принимают более компактную форму, что облегчает восприятие и последующие вычисления.

Вариационный ряд – ряд, представленный в порядке возрастания значений изучаемого признака. Различные значения признака (случайной величины X) называются вариантами (обозначаем их через x_i).

Вариационные ряды подразделяются на интервальные и безынтервальные. Безынтервальный вариационный ряд целесообразно строить в тех случаях, когда разница между минимальным и максимальным значением вариантов невелика, особенно, если признак варьирует дискретно. Последовательность его построения рассмотрим на примере данных о числе глазков на 40 клубнях картофеля. Результаты подсчёта:

7, 9, 5, 11, 9, 7, 8, 5, 8, 7, 7, 10, 6, 7, 5, 7, 8, 6, 5, 7, 7, 6, 5, 6, 4, 5, 5, 10, 5, 8, 7, 6, 7, 5, 7, 7, 6, 5,6, 6.

Является ли данная выборка репрезентативной, если из всего количества исследуемых людей выбирались самые большие результаты?, самые маленькие?

- Исследуемый показатель является количественным или качественным показателем?
- Исследуемый показатель является дискретным или непрерывным?

Находим минимальное и максимальное значение признака:

$$x_{min} = 4,\ x_{max} = 11.$$

Записываем все возможные, в пределах минимального и максимального, значения, и указываем, как часто принимает ту или иную величину анализируемый признак x. Абсолютные частоты признака обозначаются n_i.

Таблица 1.2.1 - Вариационный ряд числа глазков в клубнях картофеля (n = 40)

Значение признака, x_i	4	5	6	7	8	9	10	11
Абсолютная частота, n_i	1	10	8	12	4	2	2	1
Накопленная частота s_i (по абсолютным частотам)	1	11	19	31	35	37	39	40
Относительная частота, w_i	$\frac{1}{40} = 0.025$	$\frac{10}{40} = 0.25$	0,2	0,3	0,1	0,05	0,05	0,025
Накопленная частота (по относительным частотам)	W_1=0,025	W_1+W_2=0,275	W_1+W_2+ W_3=0,475	0,775	0,875	0,925	0,975	1

Полученный вариационный ряд полностью соответствует требованиям, предъявляемым к группировке данных, поскольку первоначальный материал стал более «обозримым», и в результате простейших преобразований мы сделали явной ранее завуалированную информацию. В частности, хорошо видно, что большая часть значений признака сгруппирована в пределах 5-7 глазков, тогда как крайние значения вариант встречаются относительно редко.

Процедуру построения интервального вариационного ряда разберём на следующем примере.

Пусть даны сведения о количестве гусениц на одном дереве.

Таблица 1.2.2 - Данные наблюдений

Число гусениц	3	4	8	11	12	13	14	16	17	18	19	21	22	24	27	28
Число деревьев	1	1	1	1	2	1	1	2	3	1	1	1	1	1	1	1

Разобьем варианты на отдельные интервалы, т.е. проведем их группировку.

Число интервалов m следует брать не очень большим, чтобы после группировки ряд не был громоздким, и не очень малым, чтобы не потерять особенности распределения признака.

Согласно формуле Стерджеса, рекомендуемое число интервалов m=1+3,322·1gn, а величина интервала (интервальная разность, ширина интервала)

$$k = \frac{x_{max} - x_{min}}{m},$$

$$n=20,$$

$$lg20 = \frac{\ln 20}{ln10} = \frac{2{,}996}{2{,}303} = 1{,}301,$$

$$m=1+3{,}322 \cdot 1{,}301=5,$$

$$k = \frac{28-3}{5} = 5.$$

За начало первого интервала рекомендуется брать величину

$$x_{нач} = x_{min} - \frac{k}{2}.$$

$$x_{нач} = 3 - \frac{k}{2} = 3 - 2{,}5 = 0{,}5.$$

Таблица 1.2.3 – Расчет относительной и накопленной частот

Число гусениц, x_i	0-5	6-10	11-15	16-20	21-28
Число деревьев (абсолютная частота), n_i	2	1	5	7	5
Накопленная частота (по абсолютным частотам), s_i	2	3	8	15	20
Относительная частота w_i	0,1	0,05	0,25	0,35	0,25
Накопленная частота (по относительным частотам)	0,1	0,15	0,40	0,75	1,00

Примечание: если имеются граничные значения (например, 10), то их можно относить к одному из соседних интервалов, но можно договориться всегда считать их правыми границами соответствующих интервалов.

Относительная частота вычисляется по формуле

$$w_i = \frac{n_i}{n}, \sum w_i = 1.$$

Вопросы: на каком количестве деревьев насчитывается

1) менее 5 гусениц? (2);

2) менее 10 гусениц? (2+1=3);

3) менее 15 гусениц? (2+1+5=8).

Соответствующие дроби 0,1; 0,15; 0,40… называются накопленными частотами. Общий вид накопленных частот показан в таблице 1.2.4.

Таблица 1.2.4 – Накопленные частоты

Значение признака	x_1	x_2	x_3	…	x_k
Накопленная частота	w_1	$w_1 + w_2$	$w_1 + w_2 + w_3$	…	$w_1 + w_2 + w_3 + \ldots + w_k = 1$

1.3. Графическое представление статистических данных

Для выработки умений конструировать статистические модели, точно отражающие поведение исследуемой системы, необходимо научиться обрабатывать исходные статистические данные, выявлять наиболее существенные связи и отношения между элементами системы и экспериментально проверять полученную модель. В решении этой задачи помогут элементы наглядной статистики. Понятие «наглядная статистика» В.Д. Селютин трактует как совокупность приемов наблюдений и анализа структуры, взаимосвязей и тенденций развития статистических совокупностей и случайных явлений с помощью графических изображений статистических сведений [18]. К элементам наглядной статистики отнесем следующие диаграммы: координатную (полигон) и столбчатую (гистограмму), а также кумулятивную кривую.

Полигон служит для изображения дискретного вариационного ряда и представляет собой ломаную, в которой концы отрезков прямой имеют координаты (x_i, y_i), где x_i значение признака, y_i - абсолютная, относительная или накопленная частота. Если исследуется интервальный вариационный ряд, то x_i – середина i-го интервала.

Пример. По данным таблицы 1.2.2 построить полигон абсолютных частот.

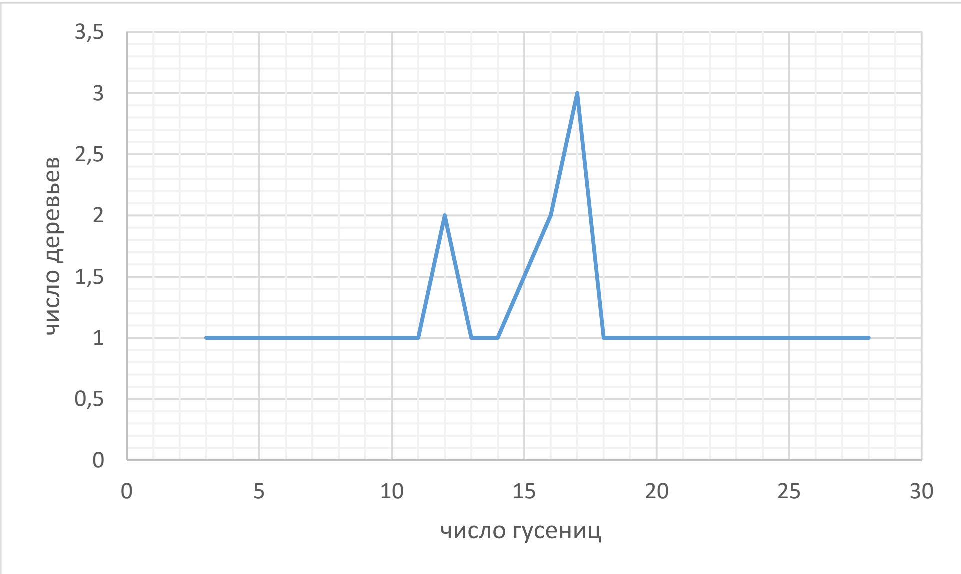


Рисунок 1.3.1 – Полигон абсолютных частот

Гистограмма служит только для изображения интервальных вариационных рядов и представляет собой ступенчатую фигуру из прямоугольников с основаниями, равными интервалам значений и высотами, равными частотам интервалов (абсолютной, относительной или накопленной).

Пример. По данным таблицы 1.2.3 построить гистограмму.

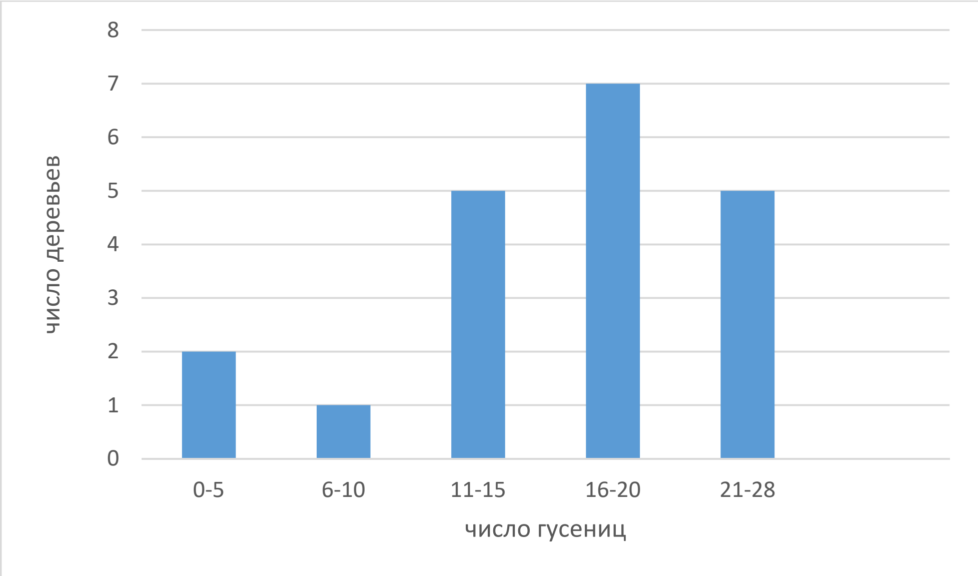


Рисунок 1.3.2. – Число гусениц на деревьях

По данному графику легко увидеть, что меньшее число гусениц - до 5 штук было только на одном дереве, больше всего – от 25 до 30 – на двух. Самое большое число гусениц -15-20 штук было на семи деревьях.

Кумулятивная кривая (кумулята) — линия накопленных частот, которая есть эмпирический прообраз графика функции распределения $y = F(x)$.

Эмпирической функцией распределения

$$y = F(x)$$

называется относительная частота того, что признак (случайная величина X) примет значение, меньшее заданного x.

Используя расчет накопленных частот из таблицы 1.2.3, построим кумулятивную кривую.

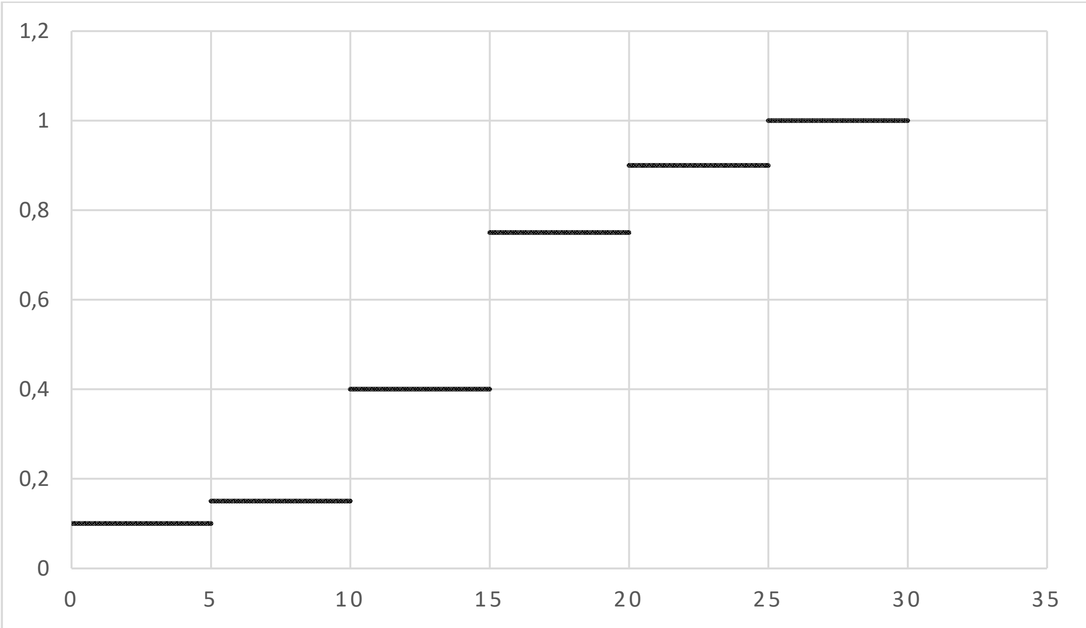


Рисунок 1.3.3 – Кумулятивная кривая

1.4. Выборочные характеристики данных наблюдений

Рассмотрим некоторые статистические характеристики исследуемого показателя.

Мода (M_o) — значение во множестве наблюдений, которое встречается наиболее часто.

Пример. По данным наблюдений найти моду

x_i	2	3	5	6	8	9	11	12	13
n_i	1	3	2	4	6	7	2	2	1

$M_o = 9$, так как это значение признака встречается чаще остальных значений – 7 раз.

Для нахождения моды интервального ряда сначала находим интервал, в котором абсолютная частота наибольшая.

Формула для вычисления моды интервального ряда

$$M_o = x_{\mathrm{Mo}} + \Delta x \frac{n_{\mathrm{Mo}} - n_{\mathrm{Mo}-1}}{(n_{\mathrm{Mo}} - n_{\mathrm{Mo}-1}) + (n_{\mathrm{Mo}} - n_{\mathrm{Mo}+1})}. \qquad \textbf{(1.4.1)}$$

где x_{M0} – начало модального интервала;

Δx – ширина интервала;

n_{Mo} – абсолютная частота в модальном интервале;

$n_{\mathrm{Mo}-1}$ - абсолютная частота в интервале, предшествующем модальному;

$n_{\mathrm{Mo}+1}$ - абсолютная частота в интервале, следующем за модальным.

Медиана – это варианта, которая находится в середине упорядоченного вариационного ряда. Медиана показывает количественную границу значения варьирующего признака, которой достигла половина единиц совокупности. Медиана делит ряд на 2 части, одна из которых имеет значения признака, меньшее, чем медиана, а другая – большее.

Для нахождения медианы данные наблюдений нужно упорядочить по возрастанию.

Для четного числа наблюдений в середине ряда находятся два значения. Медианой будет их среднее арифметическое.

Пример. По данным наблюдений найти медиану.

30, 28, 12, 32, 32, 34, 35, 37, 38.

Упорядочим ряд по возрастанию:

12, 28, 30, 32, 34, 35, 37, 38.

Число вариант четное – 8. В середине ряда находятся варианты 32 и 34. Находим их среднее арифметическое:

$$M_e = \frac{32+34}{2} = 33.$$

Для нечетного числа наблюдений находим значение признака, слева и справа от которого находится одинаковое число вариант. Номер наблюдения, соответствующий модальному находим по формуле

$$\frac{n+1}{2}. \tag{1.4.2}$$

Пример. Пусть дан ряд:

12, 14, 14, 15, 17, 17, 17, 18, 19, 20.

Составим вариационный ряд:

x_i	12	14	15	17	18	19	20
n_i	1	2	1	3	1	1	1

Здесь количество вариант нечетное –7. Согласно формуле (1.4.2), при n=7 $\frac{n+1}{2} = 4$. На четвертом месте находится варианта, равная модальному значению. *Mo*=17.

Если данные наблюдений сгруппированы, то нужно найти интервал, в котором находится половина наблюдений (ориентируемся по накопленным частотам). Медиану вычисляем по формуле

$$Me = x_{Me} + \Delta x \frac{\frac{N}{2} - S_{Me}}{n_{Me}}, \qquad \textbf{(1.4.3)}$$

где x_{Me} – начало интервала,

Δx – ширина интервала,

N – количество наблюдений,

S_{Me} – накопленная частота на начало интервального ряда,

n_{Me} - частота в медианном интервале.

Среднее значение ряда (выборочное среднее) можно вычислить по формуле

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i n_i, \;\; i = \overline{1, n}. \qquad \textbf{(1.4.4)}$$

где x_i – значение признака, n_i – частота повторения значения x_i, n – общее число наблюдений.

Сравним гистограммы двух случайных величин (рисунок 1.4): высота декоративного растения Хоста, посаженного по 12 штук на 1 м^2 (ряд 1) и посаженного по 9 штук на 1 м^2 (ряд 2). Горизонтальная ось – высота растения, вертикальная – количество растений.

Чем плотнее посажено растение, тем его высота больше. По графику (ряд 1) видно, что большинство частот группируется ближе к среднему значению. Здесь «рассеяние» вариант невелико.

В связи с этим рассмотрим характеристику – выборочную дисперсию.

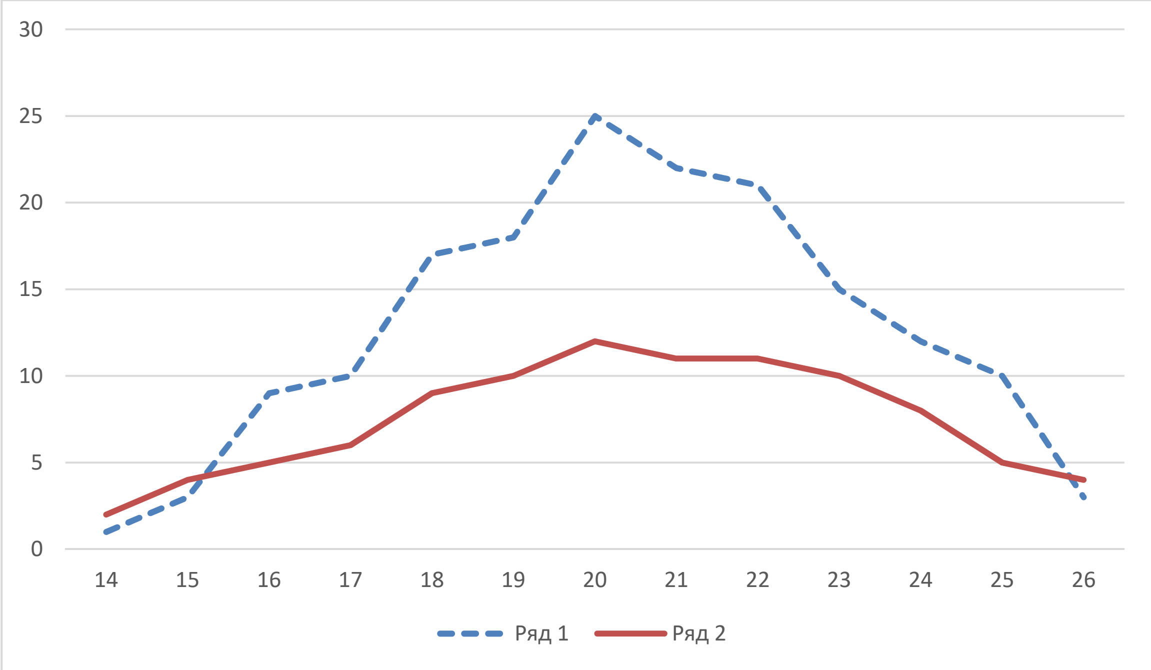


Рисунок 1.4 – Распределение высоты растения

Выборочная дисперсия

$$D = \frac{1}{n}\sum_{i=1}^{n} {x_i}^2 n_i - (\bar{x})^2. \qquad \textbf{(1.4.5)}$$

Несмещенная оценка для дисперсии

$$\overline{D} = \frac{n}{n-1} D. \qquad \textbf{(1.4.6)}$$

Среднеквадратическое отклонение

$$\sigma = \sqrt{\overline{D}}. \qquad \textbf{(1.4.7)}$$

Дисперсия и среднеквадратическое отклонение характеризуют меру разброса данных наблюдений от среднего значения. Среднеквадратическое отклонение измеряется в тех же единицах, что и сама случайная величина, а дисперсия - в квадратах этой единицы измерения.

Свойства среднего значения

1. Среднее значение постоянной равна самой постоянной.

2. Если все элементы выборки умножить на одно и то же число, то и среднее значение умножается на это число

$$\overline{cx} = c\bar{x},$$

т.к. $\frac{\sum_{i=1}^{n}(cx_i)n_i}{n} = c\frac{\sum_{i=1}^{n}x_i}{n}$.

3. Если ко всем элементам выборки прибавить одно и то же число, то к среднему значению прибавится это же число

$$\overline{x+c} = \bar{x} + c.$$

4. Среднее значение отклонения выборочных вариант от средней равно нулю

$$\overline{x-\bar{x}} = 0.$$

5. Среднее значение алгебраической суммы выборок равно алгебраической сумме средних

$$\overline{x \pm y} = \bar{x} \pm \bar{y}.$$

Свойства дисперсии

1. Дисперсия постоянной равна нулю

$$D(C) = 0.$$

2. Если все элементы выборки умножить на одно и то же число, то дисперсия будет умножена на квадрат этого числа

$$D(cx) = c^2 D(x).$$

3. Дисперсия равна разности между средним значением от квадрата значений выборки и квадратом среднего

$$D(x) = \overline{x^2} - \bar{x}^2.$$

Пример. Вычислить моду, медиану, среднее значение, дисперсию и среднеквадратическое отклонение по данным из таблицы 1.2.3. Укажем середины интервалов.

Таблица 1.4 – Интервальный ряд

Число гусениц	0-5	6-10	11-15	16-20	21-28	$\sum$
Середина интервала, x_i	2,5	7,5	12,5	17,5	24,5	
Число деревьев (абсолютная частота), n_i	2	1	5	7	5	
Накопленные частоты	2	3	8	15	20	
$x_i n_i$	5	7,5	62,5	122,5	122,5	**320**
x_i^2	6,25	56,25	156,25	306,25	600,25	
$x_i^2 n_i$	12,5	56,25	781,25	2143,75	3001,25	**5995**

По формуле (1.4.1) находим моду, предварительно определив, что интервал, в котором находится самая большая частота – 7, это интервал (16-20).

x_{M0} =16; Δx =5; n_{Mo} =7; $n_{\text{Mo}-1}$ =5; $n_{\text{Mo}+1}$ =5.

$$M_o = 16 + 5\frac{7-5}{(7-5)+(7-5)} \approx 19.$$

Для нахождения медианы по формуле (1.4.3) при количестве наблюдений n=20, $\frac{n}{2} = 10$ – эта частота по накопленным частотам находится в интервале (16-20). x_{Me} =16; Δx =5; S_{Me} =8; n_{Me+1} =7.

$$Me = 16 + 5\frac{10-8}{7} = 17.$$

Среднее значение ряда вычислим по формуле (1.4.4)

$$\bar{x} = \frac{320}{20} = 16.$$

Дисперсия (формула (1.4.5)):

$$D = \frac{5995}{20} - 16^2 = 43{,}75.$$

Несмещенная оценка для дисперсии (формула (1.4.6)):

$$\overline{D} = 20 \cdot \frac{43{,}75}{19} = 46{,}05.$$

Среднеквадратическое отклонение (формула (1.4.7)):

$$\sigma = \sqrt{46{,}05} = 6{,}8.$$

Задачи к главе 1

1.1. Измерение длины листа пшеницы в мм представлено в таблице

21	23	18	26	14	26	23	16	19	29	18	36	31	26	17	18
30	27	31	31	32	23	21	22	29	15	22	17	22	25	20	36
34	13	22	18	14	30	15	26	30	21	34	28	29	22	12	15
22	27	36	25	28	30	30	22	24	18	28	21	27	25	27	32
36	24	30	14	19	26	19	22	28	23	33	27	25	11	26	24
22	29	32	24	23	25	19	19	23	34	32	18	25	24	20	21

Построить интервальный вариационный ряд, рассчитать абсолютные, относительные, накопленные частоты, начертить гистограмму, кумуляту. Вычислить модальное значение, медиану, среднее значение, дисперсию и среднеквадратическое отклонение.

1.2. На основе наблюдений 100 растений толстянки запутанной получены данные о числе чашелистиков:

16,14,15,15,10,12,11,7,6,6,7,6,5,6,7,8,6,6,7,10,8,7,8,7,9,8,10,10,10
11,10,11,10,10,12,12,10,11,13,12,12,13,12,13,15,14,16,15,10,11,

11,10,9,7,10,10,9,10,11,11,10,9,8,8,7,6,8,9,7,7,8,8,8,7,9,9,11,10,9,
8,9,9,8,8,10,11,12,12,13,12,12,13,13,12,20,18,19,20,20,20.

Построить интервальный вариационный ряд, рассчитать абсолютные, относительные, накопленные частоты, начертить гистограмму, кумуляту. Вычислить модальное значение, медиану, среднее значение, дисперсию и среднеквадратическое отклонение.

Вопросы к главе 1

Вариант 1

1. Генеральная совокупность – это …………………………...
2. Объем совокупности – это……………………………………..
3. Полигон частот – это ……………………………………….
4. Выборочная дисперсия – это………………………………..
5. Выборочная совокупность – это……………………………
6. Репрезентативная выборка – это……………………………
7. Гистограмма – это……………………………………………...
8. Свойства выборочной средней……………………………

Вариант 2

1. Вариационный ряд – это……………………………………….
2. Эмпирическая функция – это…………………………………
3. Свойства выборочной дисперсии…………………………..
4. Относительные частоты – это………………………………
5. Варианты - это………………………………………………
6. Выборочная средняя – это…………………………………..
7. Исправленная дисперсия – это ……………………………..
8. Медиана – это ……………………………………………….

ГЛАВА 2. ПОНЯТИЕ И КЛАССИФИКАЦИЯ РЯДОВ ДИНАМИКИ

2.1. Виды временных рядов

Математические модели, при построении которых используются данные, характеризующие один объект за ряд последовательных моментов времени, называются *моделями временных рядов*.

Изучением изменения во времени числовых показателей различных явлений занимается раздел статистики, называемый Ряды динамики. Если расположить наблюдаемые данные в хронологической последовательности, можно получить временной динамический ряд, в котором каждому моменту или периоду времени соответствует статистический показатель, характеризующий изучаемый объект.

Динамический ряд называется *интервальным*, если он характеризует величину наблюдаемого явления за какой-то промежуток (период, интервал) времени.

Примером интервального ряда служат данные о количестве диатомовых водорослей с 1968 по 2009 годы в озере Нарочь в составе фитопланктонного сообщества[1]:

---

[1]По материалам статьи Михеева Т.М., Лукьянова Е.В. Многолетняя динамика показателей структурной организации фитопланктонных сообществ Нарочанских озер (Беларусь). - / Вестник БГУ. Сер. 2. 2011. № 1 – С.49.

Таблица 2.1.1 – Количество водорослей в озере

1968-1980	1981-1991	1992-2009
120	80	75

Динамический ряд называется *моментным,* если он характеризует величину наблюдаемого явления на определенный момент времени.

Например, данные о количестве инфузорий в сенном настое с 20 по 70 сутки[2]:

Таблица 2.1.2 – Количество инфузорий в сенном настое

Сутки	20	25	30	35	40	45	50	55	60	65	70
Количество, экз. в 1 см^3	80	100	150	160	200	390	900	2100	1700	1300	250

Моменты или периоды времени обозначаются через t, а числовые показатели, характеризующие изучаемое явление – называют уровнем ряда (обозначают y).

Рассмотрим показатели, характеризующие ряды динамики.

Цепные показатели характеризуют скорость роста уровня от периода к периоду. Базисные показатели характеризуют окончательный результат всех изменений в уровнях ряда от

[2] По материалам статьи Корсак М.Н. Особенности сезонной динамики, структуры и продуктивности фитопланктона Учинского водохранилища в 1998—2001 гг.

периода, к которому относится базисный уровень, до данного периода.

2.2. Показатели изменения уровней ряда динамики

Абсолютный прирост Δ_i показывает, на сколько данный уровень ряда изменился по сравнению с предыдущим или базовым.

Ускорение – разность между абсолютным изменением за данный период и абсолютным изменением за предыдущий период.

Рассчитывается только для цепных показателей. Отрицательная величина ускорения показывает на замедление роста.

Темпы роста характеризуют скорость изменения показателя в единицу времени, выраженную в процентах.

Темпы прироста показывают, на сколько процентов один уровень больше или меньше другого уровня

$$T_n^ц = \frac{y_i - y_{i-1}}{y_{i-1}} \cdot 100\%, \quad T_n^б = \frac{y_i - y_0}{y_0} \cdot 100\%$$

или $$T_n = T_p - 100\%.$$

Таблица 2.2.1 – Цепные и базисные показатели

Показатели	Абсолютный прирост	Ускорение	Темп роста	Темп прироста
Цепные	$\Delta_i^{ц} = y_i - y_{i-1}$	$\Delta_i^{/} = \Delta_i - \Delta_{i-1}$	$T_p^{ц} = \frac{y_i}{y_{i-1}} \cdot 100\%$	$T_n^{ц} = T_p^{ц} - 100\ \%$
базисные	$\Delta_i^{б} = y_i - y_0$	-	$T_p^{б} = \frac{y_i}{y_0} \cdot 100\%$	$T_n^{б} = T_p^{б} - 100\ \%$

где y_i - уровень сравниваемого периода, y_{i-1} - уровень непосредственно предшествующего периода, y_0 - уровень базисного периода.

Пример. По данным таблицы 2.1.2 вычислить основные показатели ряда динамики.

Покажем, как получены числа в таблице 2.2.2 на примере 30 суток.

Абсолютный цепной прирост $\Delta_i^{ц}$: 150-100=50 (млн.);

абсолютный базисный прирост $\Delta_i^{б}$: 150-80=70 (млн.);

ускорение $\Delta_i^{/}$: 50-20=30 (сравниваются цепные показатели за 30 и 25 сутки);

темп роста цепной $T_p^{ц}$: (150:100)·100%=150%;

темп роста базисный $T_p^{б}$: (150:80) · 100%=188%;

темп прироста цепной $T_n^{ц}$: 150-100=50(%);

темп прироста базисный $T_n^{б}$: 188-100=88(%).

Из таблицы видим: по цепным абсолютным приростам до 55 суток количество инфузорий в настое увеличивалось по сравнению с предыдущими сутками, а с 60 по 70 - уменьшалось. Базисные абсолютные приросты показывают, что с начала наблюдений количество инфузорий в настое увеличивалось по сравнению с 20 сутками.

Отрицательное ускорение на 35, 60 и 70 сутки показывает, что этот рост замедлялся, положительное – что в основном рост происходил.

Самый большой темп прироста наблюдался с 45 на 50 сутки: 131%, самое большое замедление прироста - с 65 по 70 сутки (-81%).

По сравнению с началом наблюдений самый маленький базисный темп прироста наблюдался на 25 сутки: 25%, самый большой (2525%) на 55 сутки.

При изучении двух взаимосвязанных явлений используют показатели, называемые *коэффициентами опережения*, которые вычисляются как отношения темпов роста или темпов прироста

$$k_{on} = \frac{T_p^{/}}{T_p^{//}} = \frac{T_n^{/}}{T_n^{//}}.$$

Пример. При переходе от менее развитых травянистых сообществ к более развитым наблюдается прямая зависимость количества биомассы (таблица 2.2.3)[3].

[3] А. Д. Покаржевский Геохимическая экология наземных животных. - М.: «Наука»,1983.

Таблица 2.2.2 - Расчет показателей динамического ряда

Сутки	Количество экз. в 1 см3	Абсолютный прирост		Ускорение $\Delta_i^{/}$	Темп роста, %		Темп прироста, %	
		По сравнению с предыдущими сутками (цепной) $\Delta_i^{ц}$	По сравнению с 20 сутками (базисный) $\Delta_i^{б}$		По сравнению с предыдущими сутками (цепной) $T_p^{ц}$	По сравнению с 20 сутками (базисный) $T_p^{б}$	По сравнению с предыдущими сутками (цепной) $T_n^{ц}$	По сравнению с 20 сутками (базисный)
20	80	-	-	-	-	-	-	-
25	100	20	20	-	125	125	25	25
30	150	50	70	30	150	188	50	88
35	160	10	80	-40	107	200	7	100
40	200	40	120	30	125	250	25	150
45	390	190	310	150	195	488	95	388
50	900	510	820	320	231	1125	131	1025
55	2100	1200	2020	690	133	2625	33	2525
60	1700	-400	1620	-1600	81	2125	-19	2025
65	1300	-400	1220	0	76	1625	-24	1525
70	250	-1050	170	-650	19	313	-81	213

Таблица 2.2.3 – Количество биомассы, г/м^2

Компонент	Корни	Двукрылые	Темп роста биомассы корней, % (базисный) $T_p^б$	Темп роста биомассы животных, % (базисный) $T_p^б$	Коэффициент опережения $k_{оп}$
Полярная пустыня	29	0,01	-	-	-
Тундра арктическая	350	0,1	1207	1000	1,21
Тундра типичная	2300	0,4	7931	4000	1,98

Для тундры типичной (при условии принятия за базис количества биомассы в полярной пустыне) базисный темп роста биомассы корней

$$T_p' = \frac{2300}{29} \cdot 100\% = 7931\%,$$

базисный темп роста биомассы двукрылых

$$T_p'' = \frac{0,4}{0,01} \cdot 100\% = 4000\%,$$

коэффициент опережения

$$k_{оп} = \frac{7931}{4000} = 1,98.$$

Оба рассматриваемых явления получили значительное развитие, но так как коэффициент опережения больше единицы, то базисные темпы роста биомассы корней больше базисных темпов роста биомассы двукрылых.

2.3. Средние показатели динамических рядов

Для *интервального ряда средний уровень* рассчитывается как среднее арифметическое

$$\overline{y} = \frac{\sum_{i=1}^{n} y_i}{n},$$

где n – число уровней ряда.

Пример.

Таблица 2.3.1 – Число потомков в популяции долгоносика, на 1 самку

Возраст, недели	4-5	5-6	6-7	7-8	8-9	9-10
Количество	17	19	12	10	10	9

$$\overline{y} = \frac{17+19+12+10+10+9}{6} = 13.$$

Средний уровень моментного ряда характеризует его среднее между начальным и конечным моментами. Начальный и конечный уровни находятся на границе изучаемого интервала, они наполовину относятся к предыдущему и последующему интервалам; и наполовину к изучаемому. Уровни, относящиеся к моментам внутри рассматриваемого интервала, относятся только к нему. Отсюда *формула хронологической средней для моментного ряда с равномерными интервалами*

$$\overline{y}_{xp} = \frac{\frac{y_1}{2} + y_2 + y_3 + ... + y_{n-1} + \frac{y_n}{2}}{n-1}.$$

Пример.

Таблица 2.3.2 – Число занятых гнезд в некоторой местности

Год	2000	2001	2002	2003	2004	2005
Количество	270	300	290	320	325	380

Среднее количество занятых гнезд за период с 2000 по 2005 годы:

$$\overline{y} = \frac{\frac{270}{2} + 300 + 290 + 320 + 325 + \frac{380}{2}}{6-1} = 312.$$

Средний абсолютный прирост показывает, на сколько в среднем за единицу времени должен измениться уровень ряда, чтобы от начального уровня достигнуть конечный уровень

$$\overline{\Delta} = \frac{\sum_{i=1}^{n} \Delta_i^{ц}}{n-1}, \qquad \textbf{(2.3)}$$

где n – число уровней ряда.

Покажем, что $\overline{\Delta} = \frac{y_n - y_0}{n-1}$.

Действительно, так как $\overline{\Delta_i^{ц}} = y_i - y_{i-1}$, то

$$\overline{\Delta} = \frac{(y_1 - y_0) + (y_2 - y_1) + ... + (y_{n-1} - y_{n-2}) + (y_n - y_{n-1})}{n-1} = \frac{y_n - y_0}{n-1}.$$

Если цепные абсолютные приросты примерно одинаковы, то это говорит о линейном характере развития ряда. При этом средний абсолютный прирост можно использовать для прогноза значения ряда в момент времени *n+1*

$$y_{n+1} = y_n + \overline{\Delta}.$$

Пример. При наблюдении за ростовой активностью в процессе вегетации растения производилось внесение удобрения, содержащего медь.

Таблица 2.3.3 – Высота доминирующего побега

День	Высота, см	Цепные абсолютные приросты
5.06.	26,3	-
8.06.	32,4	6,1
11.06.	38,8	6,4
14.06.	45,0	6,2
17.06.	51,1	6,1
20.06.	57,7	6,6
23.06.	63,6	5,9

Цепные абсолютные приросты приблизительно равны между собой. Тогда, согласно формуле (2.3)

$$\overline{\Delta} = \frac{63{,}6 - 26{,}3}{7 - 1} = 6{,}2.$$

Получено, что за 3 дня в среднем прирост составляет 6,3 см.

Задачи к главе 2

2.1. По данным таблицы вычислить цепные и базисные абсолютные приросты, темпы роста, темпы прироста, ускорение. Изобразить ряд динамики графически.

Время, год	Плотность популяции
5	150
10	350
15	280
20	370
25	410
30	360
35	350
40	320
45	310
50	340
55	360

2.2. По данным таблицы определить средний абсолютный прирост, средний темп роста и прироста численности одноклеточного.

Часы	4	8	12	16	20	24
Рост численности популяции одноклеточного	3	5	11	15	30	67

2.3. Имеются данные наблюдений об изменении численности двух видов простейших при совместном культивировании. Найти коэффициенты опережения. Сделать вывод.

Время, суток	Paramecium aurelia	Paramecium caudatum
2	16	12
4	38	17
6	59	14
8	74	13
10	77	11
12	80	8
14	81	5
16	85	3
18	89	2

2.4. Найти средний уровень интервального ряда:

Недели	4-5	5-6	6-7	7-8	8-9	9-10
Выживание жука, на 10 особей	88	84	80	78	75	73

2.5. Определить вид ряда динамики и его средний показатель:

Год	2007	2008	2009	2010	2011	2012	2013
Число бездомных собак	102	93	104	113	128	116	110

2.6. Получены данные годичного прироста осевого побега у сеянцев клена

Возраст дерева	1 год	2 год	3 года	4 года	5 лет
Прирост, см	6,5	1,3	1,6	1,5	1,7

В какой период наблюдался равномерный прирост? Рассчитать средний абсолютный прирост за год.

Вопросы к главе 2

1. Какой ряд называется интервальным?
2. Какой ряд называется моментным?
3. Что выражает показатель ряда – абсолютный прирост?
4. Как вычислить ускорение?
5. Что характеризует темп роста?
6. Что показывает темп прироста?
7. Как вычислить средний уровень для интервального ряда?
8. Запишите формулу для вычисления средний уровень для моментного ряда с равномерными интервалами.

ГЛАВА 3. СТАТИСТИЧЕСКИЕ МЕТОДЫ НАХОЖДЕНИЯ ТЕНДЕНЦИИ РАЗВИТИЯ ПОКАЗАТЕЛЯ ВО ВРЕМЕНИ

3.1. Типы трендовых моделей

В рядах динамики при их изучении может быть выявлена некоторая закономерность. На процессы, изменяемые во времени, влияют несколько факторов. Это могут быть конъюнктурные и сезонные колебания; изменения, вызванные природными катаклизмами и различные незначительные факторы. Однако можно выделить основную тенденцию развития, которую называют трендом. Тренд – это функция, зависящая от времени. Прогнозирование с помощью трендовых моделей включает в себя следующие этапы:

1) выявление наличия трендовой зависимости;
2) выбор одной или нескольких кривых, форма которых соответствует динамике временного ряда;
3) нахождение параметров выбранных кривых;
4) проверка адекватности и точности выбранных моделей и окончательный выбор кривой;
5) расчет прогнозируемого значения временного ряда;
6) расчет доверительного интервала в точке прогноза.

Предположить, к какому типу трендовых моделей относится исследуемый ряд динамики, можно, проанализировав его графическое изображение. При построении нужно строго соблюдать масштаб, чтобы не исказить наглядности. Далее можно применить метод сглаживания или укрупнения динамических интервалов, который заключается в механическом выравнивании уровней ряда с использованием соседних уровней. Расчет трехдневных скользящих сумм: 1-я скользящая сумма равна объему продаж за 1,2 и 3 дни недели; 2-я – за 2,3 и 4 дни и т.д. Аналогично выполняется расчет пятидневных скользящих сумм. Скользящая средняя по трехдневным скользящим суммам будет относиться ко 2-му дню каждой трехдневки, по пятидневным- к 3-му дню каждой пятидневки.

Пример. Имеются данные о численности фитопланктона в озере, полученные через равные промежутки времени в период с 26 марта по 22 октября (таблица 3.1.1).

По виду эмпирического графика (рисунок 3.1) можно предположить, что здесь имеется квадратичная зависимость.

Кривые роста могут подразделяться на 3 группы.

К *первой группе* отнесем кривые, описывающие процессы с монотонным развитием без пределов роста. Такой процесс может быть описан одной из функций

- $\breve{y}_t = a_0 + a_1 t$,
- $\breve{y}_t = a_0 + a_1 t + a_2 t^2$,

- $\hat{y}_t = a_0 + a_1 t + a_2 t^2 + a_3 t^3$,
- $\hat{y}_t = ab^t$.

Таблица 3.1.1 – Расчет скользящих средних

Время, *t*	**Численность фитопланктона, *y***	**Трехдневные скользящие суммы**	**Трехдневные скользящие средние**
1	90	-	-
2	100	270	90
3	80	275	92
4	95	265	88
5	90	1685	562
6	1500	11590	3863
7	10000	20500	6833
8	9000	28500	9500
9	9500	24500	8167
10	6000	25000	8333
11	9500	24500	8167
12	9000	26500	8833
13	8000	25500	8500
14	8500	23500	7833
15	7000	17500	5833
16	2000	9900	3300
17	900	3700	1233
18	800	1900	633
19	200	1400	467
20	400	-	-

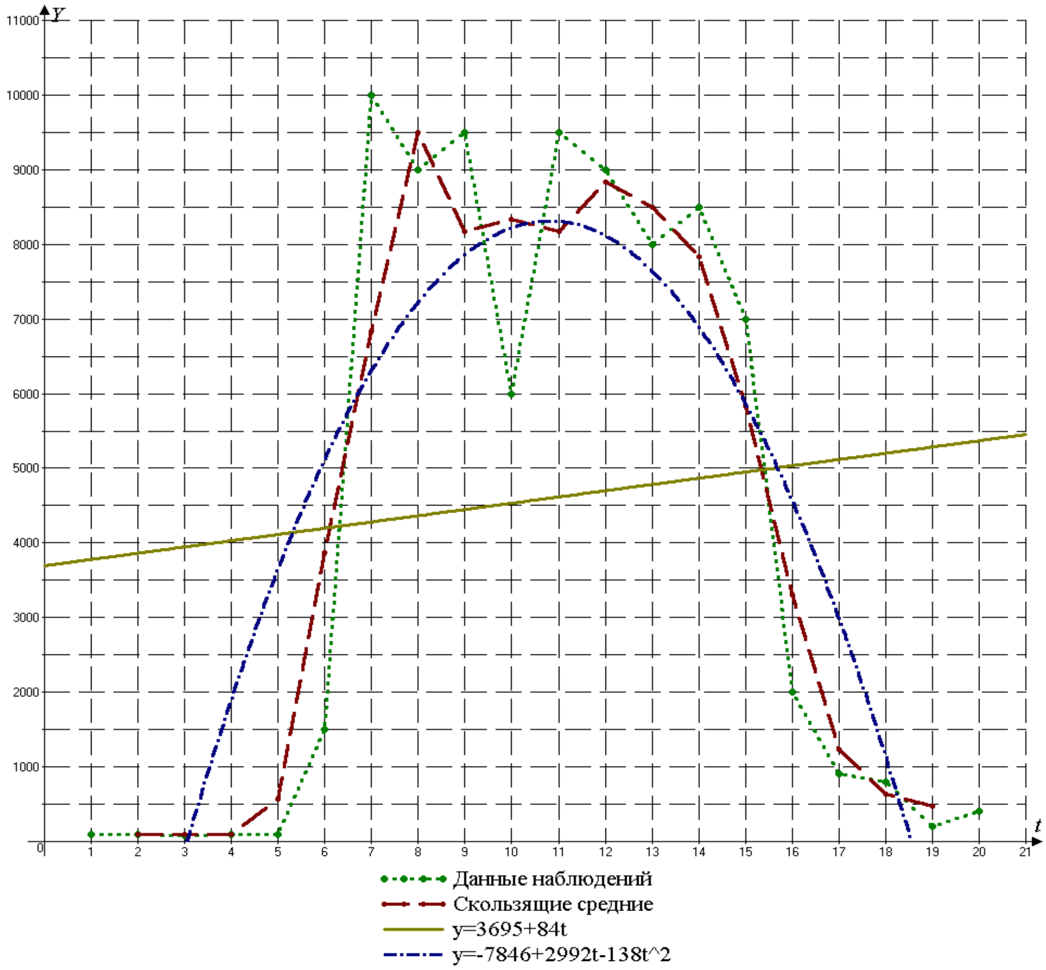


Рисунок 3.1 – Графики протекания процесса

Ко *второй* – кривые, имеющие предел роста в исследуемом периоде. Такие кривые называют кривыми насыщения

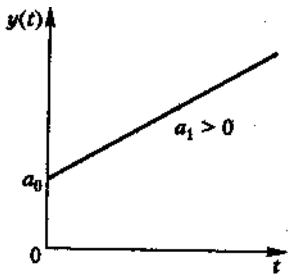
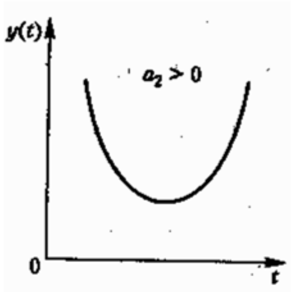
- $\hat{y}_t = k + ab^t$.

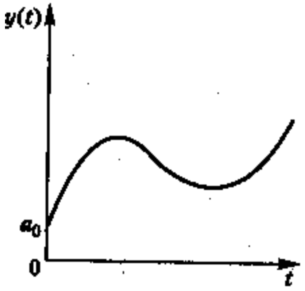
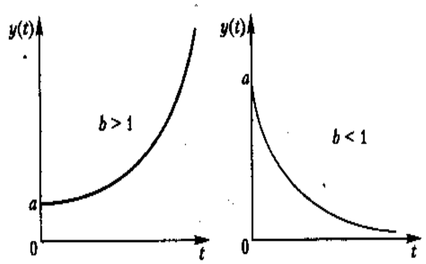
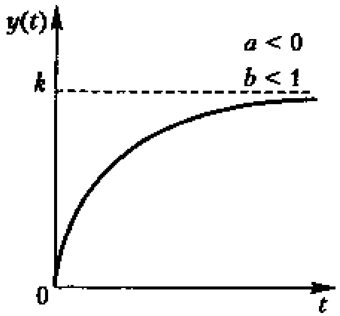
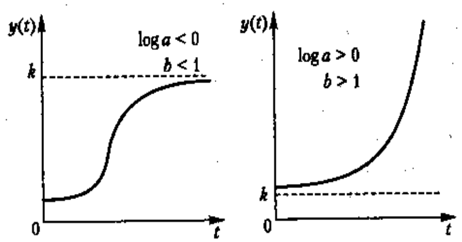
Здесь k– горизонтальная асимптота (см. таблицу 3.1.2).

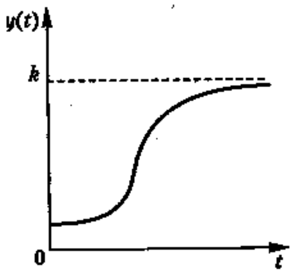
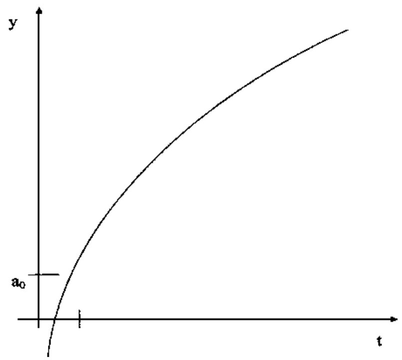
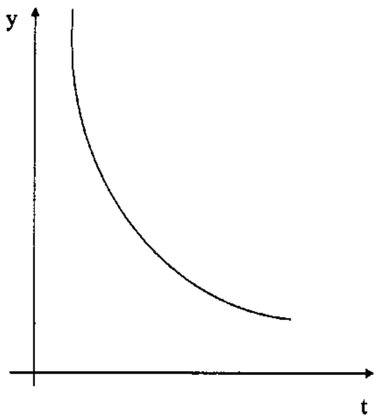
К *третьей* – кривые насыщения, имеющие точки перегиба. Это S-образные кривые

- $\hat{y}_t = ka^{b^t}$,
- $\dfrac{1}{y_t} = ka^{b^t}$.

Таблица 3.1.2 - Виды трендовых моделей.

Функция	Аналитическое выражение	График	Выбор функции	Значение коэффициентов	Применение (примеры)
Линейная	$\hat{y}_t = a_0 + a_1 t$		Примерно постоянное абсолютное ускорение	a_0– начальный уровень тренда в момент или период, принятый за начало отсчета времени *t;* a_1 –среднегодовой абсолютный прирост (среднее изменение за единицу времени); константа тренда	Урожайность для области, республики, региона
Парабола второго порядка	$\hat{y}_t = a_0 + a_1 t + a_2 t^2$		Ускоренное или замедленное изменение уровней при постоянном ускорении	a_0, a_1 – те же, a_2– квадратический параметр, равный половине ускорения; константа тренда	Для равноускоренных процессов: прогрессирующее поступление нового высокопроизводительного оборудования, увеличение среднесуточного прироста живого веса с возрастом; при снятии ограничений (в распределении доходов, в уровне оплаты труда, при повышении цены на продукцию); а так же производство устаревшей продукции, ликвидируемой отрасли с/х

Кубическая парабола	$\hat{y}_t = a_0 + a_1 t + a_2 t^2 + a_3 t^3$		Постоянные третьи разности	a_0, a_1,a_2– те же, a_3 – изменение ускорения роста	
Экспонента	$\hat{y}_t = ab^t$		Абсолютный прирост пропорционален достигнутому уровню или постоянные темпы роста и прироста	*a*-характеризует начальные условия,*b*–постоянный темп изменения в разах, константа тренда. Если *b*>1- тенденция ускоряющегося возрастания уровней; *b*<1 – замедляющееся снижение уровней	Размножение организмов при отсутствии ограничений со стороны среды: кормов, пространства, хищников, болезней
Модифицированная экспонента	$\hat{y}_t = k + ab^t$		Рост уровней ряда замедляется и стремится к некоторому пределу	*k* определяется свойствами прогнозируемого процесса или задается экспертным путем	
Кривая Гомперца	$\hat{y}_t = ka^{b^t}$		Наличие экстремальных значений	*y=k*– асимптота функции, *a>0, 0<b<1*	

Логистическая кривая	$\widehat{y_t} = \dfrac{y_{max} - y_{min}}{e^{a_0 + a_1 t} + 1} + y_{min}$		Имеет два перегиба – от ускоряющегося роста к равномерному и от равномерного роста к замедлению. Точка симметрии совпадает с точкой перегиба	*a*и *b* параметры тренда	Развитие нового производства, процесс насыщения рынка новыми товарами, рост численности особей в популяции
Логарифмическая	$\hat{y}_t = a_0 + a_1 \log t$		Замедляющийся рост уровней при достаточно большом *t*. Кривая не имеет экстремума		Рост спортивных достижений, повышение продуктивности животных
гиперболическая	$\hat{y}_t = a_0 + \dfrac{a_1}{t}$			$a_1>0$ – замедление снижения уровней, $a_1<0$ – медленно возрастающая функция	Для характеристики связи удельных расходов сырья, материалов, топлива с объемом выпускаемой продукции, зависимости времени обращения товаров от величины товарооборота

Рассмотрим методы, позволяющие выявить наличие тренда, определить вид функциональной зависимости, проверить его на адекватность.

3.2. Выявление наличия тренда

3.2.1. Метод разности средних

1. Количество уровней ряда делится на две примерно равные части

$$n_1 \approx n_2, \quad n_1 + n_2 = n.$$

Находим средние уровни каждой части

$$\bar{y}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1} y_i, \quad \bar{y}_2 = \frac{1}{n_2}\sum_{i=n_1+1}^{n_2} y_i. \qquad \textbf{(3.2.1)}$$

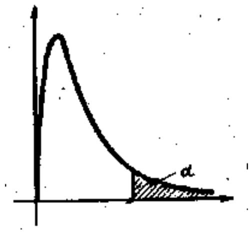
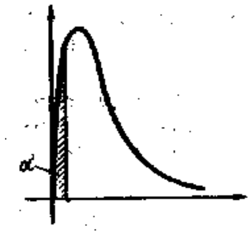
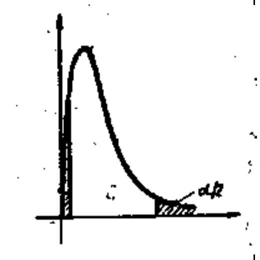
2. Находим дисперсии σ_1^2 и σ_2^2

$$\sigma_1^2 = \frac{1}{n_1 - 1}\sum_{i=1}^{n_1}\left(y_i - \bar{y}_1\right)^2; \quad \sigma_2^2 = \frac{1}{n_2 - 1}\sum_{i=n_1+1}^{n}\left(y_i - \bar{y}_2\right)^2. \qquad \textbf{(3.2.2)}$$

3. Проверяем гипотезу о равенстве дисперсий σ_1^2 и σ_2^2 с помощью F-критерия Фишера. В качестве нулевой гипотезы берем H_0: $\sigma_1^2 = \sigma_2^2$. В качестве альтернативных гипотез одну из H_1. Для уровня значимости α=0,05 из таблицы Приложения 1 находим $F_{кр}$.

$$F_{\text{расч}} = \sigma_1^2 / \sigma_2^2 \qquad \textbf{(3.2.3)}$$

Таблица 3.2.1 – Выбор гипотезы

H₀	***H₁***	***Fкр***	**график**	**Выбор гипотезы**
$\sigma_1^2 = \sigma_2^2$	$\sigma_1^2 > \sigma_2^2$	Для $F_{кр}^{пр}$: $\nu_1 = n_1\text{-}1;$ $\nu_2 = n_2\text{-}1$		H_0: $F_{расч} < F_{кр}^{пр}$ H_1: $F_{расч} > F_{кр}^{пр}$
H_0	H_1	$F_{кр}$	график	Выбор гипотезы
$\sigma_1^2 = \sigma_2^2$	$\sigma_1^2 < \sigma_2^2$	Для $F_{кр}^{лев}$: $\nu_1 = n_2\text{-}1;$ $\nu_2 = n_1\text{-}1$		H_0: $F_{расч} > F_{кр}^{лев}$ H_1: $0 < F_{расч} < F_{кр}^{лев}$
$\sigma_1^2 = \sigma_2^2$	$\sigma_1^2 \neq \sigma_2^2$	Для $F_{кр}^{пр}$: $\nu_1 = n_1\text{-}1;\ \nu_2 = n_2\text{-}1$, для $F_{кр}^{лев}$: $\nu_1 = n_2\text{-}1;\ \nu_2 = n_1\text{-}1$ при уровне значимости α/2.		H_0: $F_{кр}^{лев} < F_{расч} < F_{кр}^{пр}$ H_1: $0 < F_{расч} < F_{кр}^{лев}$ или $F_{расч} > F_{кр}^{пр}$

4. Если с помощью *F*-критерия Фишера установлено равенство дисперсий σ_1^2 и σ_2^2, вычисляем *S*- средневзвешенное значение среднеквадратичных отклонений σ_1 и σ_2 относительно $\bar{y}_1$ и $\bar{y}_2$

$$S = \sqrt{\frac{(n_1 - 1)^2 \sigma_1^2 + (n_2 - 1)^2 \sigma_2^2}{n - 2}}. \qquad \textbf{(3.2.4)}$$

5. Воспользуемся критерием Стьюдента для проверки гипотезы о равенстве средних $\bar{y}_1$ и $\bar{y}_2$

$$t_{расч} = \frac{\overline{y}_1 - \overline{y}_2}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}. \qquad \textbf{(3.2.5)}$$

$t_{теор}$ находим из таблицы Приложения 2 при $f=n-1$.

Если $t_{расч} \geq t_{теор}$, то считается, что между первой и второй частями динамического ряда существует зависимость, а значит, существует тенденция развития исследуемой системы. При $t_{расч} < t_{теор}$ тенденция развития отсутствует.

Пример.

Таблица 3.2.2 - Пятидневные скользящие средние

№	Данные наблюдений	Пятидневные суммы	Пятидневные скользящие средние
1.	249	-	-
2.	230	-	-
3.	320	1559	311,8
4.	460	1549	309,8
5.	300	1734	346,8
6.	239	1930	386
7.	415	1999	399,8
8.	516	2242	448,4
9.	529	2573	514,6
10.	543	2675	535
11.	570	2671	-
12.	517	2639	-

Проверка методом разности средних исходного ряда не показывает наличие связи, поэтому, сначала выровняем ряд по пятидневной скользящей средней и по выровненному ряду проверим наличие тренда.

1. $n=8,\ n_1=n_2=4$.

Средние уровни каждой части согласно формуле (3.2.1)

$$\bar{y}_1 = \frac{311{,}8 + 309{,}8 + 346{,}8 + 386}{4} = 339,$$

$$\bar{y}_2 = \frac{399{,}8 + 448{,}4 + 514{,}6 + 535}{4} = 477{,}5.$$

2. По формуле (3.2.2) находим дисперсии

$$\sigma_1^2 = \frac{(311{,}8-339)^2 + (309{,}8-339)^2 + (346{,}8-339)^2 + (386-339)^{2^2}}{4-1} = 1287$$

Аналогично находим σ_2^2

$$\sigma_2^2 = 1220.$$

3. Проверим гипотезу H_0: можно ли при уровне значимости α=0,05 (или с вероятностью 95%) считать статистически незначимым различие между оценками σ_1^2=1287 и σ_2^2=1220. В качестве альтернативной гипотезы H_1 берем гипотезу $\sigma_1^2 > \sigma_2^2$.

$v_1 = v_2 = n_1 - 1 = n_2 - 1 = 4 - 1 = 3$.

Из таблицы значений F-критерия Фишера (Приложение 1) при уровне значимости α=0,05 $F_{кр}^{пр}$ =9,28.

По формуле (3.2.3)

$$F_{расч}=\frac{1287}{1220}=1{,}06.$$

Так как 1,06<9,28, то $F_{расч}< F_{кр}^{пр}$. Значит, принимается гипотеза H_0 о равенстве дисперсий, то есть с вероятностью 95% можно считать, что различия между дисперсиями σ_1^2 и σ_2^2 незначимые.

4. Найдем S по формуле (3.2.4)

$$S=\sqrt{\frac{(4-1)^2\cdot 1287+(4-1)^2\cdot 1220}{8-2}}=61{,}3.$$

5. Согласно формуле (3.2.5)

$$t_{расч}=\frac{339-474{,}5}{61{,}3\sqrt{\frac{1}{4}+\frac{1}{4}}}=3{,}125.$$

Из таблицы значений t-критерия Стьюдента (Приложение 2) при f=8-1=7 при уровне значимости α=0,05 $t_{теор}$=2,36. Итак, $t_{расч}>t_{теор}$, поэтому существует тенденция развития динамического ряда.

3.2.2. Метод Фостера-Стьюарта

Для проверки существования тренда к слабо проявляющимся тенденциям можно применить более чувствительный метод – метод Фостера-Стьюарта, который считается более надежным и не требует дополнительных проверок.

Основу метода составляют характеристики

$$W=\sum_{t=1}^{n} w_t \text{ и } D=\sum_{t=1}^{n} d_t ,$$

где $w_t = u_t + v_t$, $d_t = u_t - v_t$, а параметры u_t и v_t находятся по формулам

$$u_t = \begin{cases} 1, & \text{если} \quad y_t > y_{t-1},...,y_1; \\ 0 & \text{в} \quad \text{остальных} \quad \text{случаях}; \end{cases} \qquad \textbf{(3.2.6)}$$

$$v_t = \begin{cases} 1, & \text{если} \quad y_t < y_{t-1},...,y_1; \\ 0 & \text{в} \quad \text{остальных} \quad \text{случаях} \end{cases}$$

Из приведенных формул следует, что $0 \le W \le n-1$. Если $y_1 = y_2 = ... = y_n$, то $W=0$. Если $y_1 < y_2 < ... < y_n$, то $W=n-1$. Аналогично $-(n-1) \le D \le n-1$.

Показатель *D* применяется для определения тенденции изменения во времени y_t. По *t*-критерию Стьюдента проверяется гипотеза об отсутствии тенденции в дисперсии (*W*). Находим

$$t_{расч_1} = \frac{D}{\sigma_2} \qquad \textbf{(3.2.7)}$$

(значение σ_2 берем из таблицы Приложения 6). Теоретическое значение *t*-критерия Стьюдента определяют по таблице Приложения 2. Если $\left| t_{расч_1} \right| \ge t_{теор}$, то гипотеза об отсутствии тенденции развития в средней (об отсутствии тренда) отклоняется. Другими словами, *тренд существует*.

Показатель W применяется для определения тенденции изменения во времени дисперсии S_t. Находим

$$t_{расч_2} = \frac{W - \overline{W}}{\sigma_1} \qquad \textbf{(3.2.8)},$$

$\overline{W}$, σ_1 и σ_2 находятся из таблицы Приложения 6.

Если $\left|t_{расч\ 2}\right| \geq t_{теор}$, то гипотеза об отсутствии тенденции в дисперсии отклоняется.

Пример. Проверим динамический ряд: 2,4,6,8,10,12,14,16,18,20 на наличие трендовой зависимости методом Фостера-Стьюарта. Составим таблицу, используя формулы (3.2.6):

Таблица 3.2.1 – Расчет показателей u_t и w_t

t	y_t	u_t	v_t	w_t	d_t
1	2	0	0	0	0
2	4	1	0	1	1
3	6	1	0	1	1
4	8	1	0	1	1
5	10	1	0	1	1
6	12	1	0	1	1
7	14	1	0	1	1
8	16	1	0	1	1
9	18	1	0	1	1
10	20	1	0	1	1
$\sum$	**-**	**-**	**-**	**9**	**9**

$W=\sum w_t=9$; $D=\sum d_t=9$,

Из Приложения 6 находим

$$\overline{W}=3{,}86;\ \sigma_1=1{,}29,\ \sigma_2=1{,}96.$$

По формуле (3.2.7)

$$t_{расч_1}=\frac{9}{1{,}96}=4{,}59.$$

По формуле (3.2.8)

$$t_{расч_2}=\frac{9-3{,}86}{1{,}29}=3{,}98;$$

$t_{теор}=2{,}31$ (см. Приложение 5 при $n=10,\ v=n-2$).

$t_{расч_1}>t_{теор},\ t_{расч_2}>t_{теор}$. Значит, гипотеза о наличии тенденции развития в средних и в дисперсии принимается.

3.3. Аналитическое нахождение параметров тренда

После того, как установлено, что существует тенденция развития динамического ряда и выявлен характер кривой, вычислим параметры кривой.

Рассмотрим один из методов, применяемых для расчета параметров кривой – метод наименьших квадратов, суть которого состоит в определении зависимости *y* от *t* путем минимизации суммы квадратов отклонений фактических (эмпирических) значений *y* от теоретических значений *y*, определяемых предполагаемым уравнением

$$s = \sum_{i=1}^{n} \left(y_i - \hat{y}_i \right)^2 \rightarrow \min .$$

Пусть искомая функция имеет вид

$$\hat{y} = a_0 + \sum_{p=1}^{k} a_p t^p .$$

Известно, что необходимым условием существования минимума функции является равенство нулю ее частных производных по неизвестным параметрам $a_0, a_1, a_2, ..., a_k$. Составляя разности $y_i - \hat{y}_i = y_i - a_0 - \sum_{p=1}^{k} a_p t^p$ и приравнивая к нулю частные производные $\frac{\partial S}{\partial a_p}$, получим в общем виде систему нормальных уравнений относительно $a_0, a_1, a_2, ..., a_k$

$$\begin{cases} na_0 + a_1 \sum t + a_2 \sum t^2 + ... + a_k \sum t^k = \sum y; \\ a_0 \sum t + a_1 \sum t^2 + a_2 \sum t^3 ... + a_k \sum t^{k+1} = \sum yt; \\ .... \\ a_0 \sum t^k + a_1 \sum t^{k+1} + a_2 \sum t^{k+2} ... + a_k \sum t^{2k} = \sum yt^k , \end{cases}$$

где n – число членов ряда динамики, k- степень, с которой переменная t входит в уравнение $y = a_0 + \sum_{p=1}^{k} a_p t^p$.

Ниже приведем системы уравнений, с помощью которых вычисляются параметры моделей по методу наименьших квадратов.

Для линейного уравнения вида $\hat{y} = a_0 + a_1 t$ система примет вид

$$\begin{cases} na_0 + a_1 \sum t = \sum y; \\ a_0 \sum t + a_1 \sum t^2 = \sum yt. \end{cases} \quad \textbf{(3.3.1)}$$

Для параболы $\hat{y} = a_0 + a_1 t + a_2 t^2$

$$\begin{cases} na_0 + a_1 \sum t + a_2 \sum t^2 = \sum y; \\ a_0 \sum t + a_1 \sum t^2 + a_2 \sum t^3 = \sum yt; \\ a_0 \sum t^2 + a_1 \sum t^3 + a_2 \sum t^4 = \sum yt^2. \end{cases} \quad \textbf{(3.3.2)}$$

Для гиперболы $\hat{y} = a_0 + \frac{a_1}{t}$

$$\begin{cases} na_0 + a_1 \sum \frac{1}{t} = \sum y; \\ a_0 \sum \frac{1}{t} + a_1 \sum \frac{1}{t^2} = \sum \frac{y}{t}. \end{cases} \quad \textbf{(3.3.3)}$$

Для логарифмической зависимости $\hat{y} = a_0 + a_1 \ln t$

$$\begin{cases} na_0 + a_1 \sum \ln t = \sum y; \\ a_0 \sum \ln t + a_1 \sum (\ln t)^2 = \sum y \ln t. \end{cases} \quad \textbf{(3.3.4)}$$

Для экспоненты $\hat{y} = ab^t$ после ее приведения к линейному виду

$$\begin{cases} n \ln a + \ln b \sum t = \sum \ln y; \\ \ln a \sum t + \ln b \sum t^2 = \sum t \ln y. \end{cases} \quad \textbf{(3.3.5)}$$

Для модифицированной экспоненты $\hat{y} = k + ab^t$ метод наименьших квадратов непосредственно не используется. Можно применить метод трех сумм. Ряд y_t, $t = \overline{1,n}$, разбивается на 3 равные части: $\sum_1 y_t$, $\sum_2 y_t$, $\sum_3 y_t$. Пусть m – число уровней в каждом отрезке. Тогда имеем 3 уравнения относительно 3-х неизвестных *a, b,* и *k*

$$\sum_1 y_t = \sum_{t=0}^{m-1}\left(k + ab^t\right) = mk + a\frac{b^m - 1}{b-1};$$

$$\sum_2 y_t = \sum_{t=m}^{2m-1}\left(k + ab^t\right) = mk + ab^m\frac{b^m - 1}{b-1};$$

$$\sum_3 y_t = \sum_{t=2m}^{3m-1}\left(k + ab^t\right) = mk + ab^{2m}\frac{b^m - 1}{b-1}.$$

Вычитая из второго уравнения первое и из третьего – второе, получим

$$\sum_2 y_t - \sum_1 y_t = a\frac{\left(b^m - 1\right)^2}{b-1};$$

$$\sum_3 y_t - \sum_2 y_t = ab^m\frac{\left(b^m - 1\right)^2}{b-1}.$$

Откуда

$$a=\left(\sum{}_2 y_t-\sum{}_1 y_t\right)\frac{b-1}{\left(b^m-1\right)^2};$$

$$b=\sqrt[m]{\frac{\sum{}_3 y_t-\sum{}_2 y_t}{\sum{}_2 y_t-\sum{}_1 y_t}};$$

$$k=\frac{1}{n}\left(\sum{}_1 y_t-\frac{b^m-1}{b-1}a\right).$$

Замечание. Приведенные системы для линейной, параболической и гиперболической зависимость можно упростить, если перенести начало координат в середину ряда динамики. Если до переноса *t=1,2,3,…*, то после переноса:

- для четного числа наблюдений *t=…,-5, -3, -1, 1, 3, 5,…;*

- для нечетного числа наблюдений *t=…,-3,-2, -1, 0, 1 2, 3,…*

При этом $\sum t=\sum t^3=\sum \frac{1}{t}=0$ и количество слагаемых в уравнениях уменьшается.

Пример. По трехдневным скользящим средним из примера пункта 3.1 (таблица 3.1.1) рассчитать коэффициенты a_0 и a_1 линейного тренда $y=a_0+a_1t$ и b_0, b_1, b_2 параболического тренда $y=b_0+b_1t+b_2t^2$.

Решение. Число наблюдений *n=18*.

Расчетные данные запишем в таблицу 3.1.2.

Подставим полученные значения в систему (3.3.1.)

$$\begin{cases} 18a_0+189a_1=82327; \\ 189a_0+2469a_1=904978; \end{cases}$$

Таблица 3.3.1 – Данные наблюдений

№	Время, *t*	Трехдневные скользящие средние
1	2	90
2	3	92
3	4	88
4	5	562
5	6	3863
6	7	6833
7	8	9500
8	9	8167
9	10	8333
10	11	8167
11	12	8833
12	13	8500
13	14	7833
14	15	5833
15	16	3300
16	17	1233
17	18	633
18	19	467

Используя метод Крамера для решения системы линейных уравнений, получим

$$\Delta = \begin{vmatrix} 18 & 189 \\ 189 & 246 \end{vmatrix} = 8721,$$

Таблица 3.3.2 – Расчетные данные для линейного и параболического трендов

№	*y*	*t*	t^2	t^3	t^4	*yt*	yt^2
1	90	2	4	8	16	180	360
2	92	3	9	27	81	276	828
3	88	4	16	64	256	352	1408
4	562	5	25	125	625	2810	14050
5	3863	6	36	216	1296	23178	139068
6	6833	7	49	343	2401	47831	334817
7	9500	8	64	512	4096	76000	608000
8	8167	9	81	729	6561	73503	661527
9	8333	10	100	1000	10000	83330	833300
10	8167	11	121	1331	14641	89837	988207
11	8833	12	144	1728	20736	105996	1271952
12	8500	13	169	2197	28561	110500	1436500
13	7833	14	196	2744	38416	109662	1535268
14	5833	15	225	3375	50625	87495	1312425
15	3300	16	256	4096	65536	52800	844800
16	1233	17	289	4913	83521	20961	356337
17	633	18	324	5832	104976	11394	205092
18	467	19	361	6859	130321	8873	168587
$\sum$	**82327**	**189**	**2469**	**36099**	**562665**	**904978**	**10712526**

$$\Delta_0 = \begin{vmatrix} 82327 & 189 \\ 904978 & 246 \end{vmatrix} = 32224521,$$

$$\Delta_1 = \begin{vmatrix} 18 & 189 \\ 189 & 246 \end{vmatrix} = 8721,$$

$a_0 = 3695, a_1 = 84.$

Итак, линейный тренд

$$\hat{y}_1 = 3695 + 84t\,. \qquad \textbf{(3.3.6)}$$

Составим систему уравнений для нахождения параметров параболического тренда, используя формулы (3.3.2).

Подставим полученные значения из расчетной таблицы 3.3.2 в систему

$$\begin{cases} 18b_0 + 189b_1 + 2469b_2 = 82327, \\ 189b_0 + 2469b_1 + 36099b_2 = 904978, \\ 2469b_0 + 36099b_1 + 562665b_2 = 10712526. \end{cases}$$

Δ=90140256, Δ_0=-707218134444, Δ_1=269696500578, Δ_2=-12483489402.

b_0=-7846, b_1=2992, b_2=-138.

Параболический тренд

$$\hat{y}_2 = -7846 + 2992t - 138t^2 . \qquad \textbf{(3.3.7)}$$

Графики полученных функций – на рисунке 3.3.

Какой же из прогнозов более точный: вычисленный по линейной модели тренда или по параболической? Для ответа на этот вопрос нужно проверить модели на адекватность и точность.

Рассмотрим еще один пример – рост численности особей в популяции. Этот процесс часто развивается по закону от ускоренного роста к равномерному и от равномерного – замедленному. Графически это выглядит как S-образная кривая. Ее уравнение

$$\hat{y}_i = \frac{y_{max} - y_{min}}{e^{a_0 + a_1 t} + 1} + y_{min}. \qquad \textbf{(3.3.8)}$$

Метод наименьших квадратов к отысканию параметров тренда непосредственно не применяется. Его уравнение преобразуется к виду

$$e^{a_0+a_1t} = \frac{y_{max}-y_{min}}{y_i-y_{min}} - 1. \qquad \textbf{(3.3.9)}$$

Введем обозначение $g_i = e^{a_0+a_1t}$. Тогда $lng_i = a_0 + a_1t.$

По условию метода наименьших квадратов минимизируется сумма квадратов разностей логарифмов g_i и $\widehat{g_i}$

$$\sum_{i=1}^{n}\left(lng_i - ln\widehat{g_i}\right)^2 \to min,$$

где g_i – вычисленные по расчетным данным значения уровней ряда, $\widehat{g_i}$ – теоретические значения. Или

$$\sum_{i=1}^{n}\left(lng_i - (a_0 + a_1t)\right)^2 \to min.$$

После нахождения частных производных по a_0 и a_1, получаем систему уравнений как для линейной функции

$$\begin{cases} na_0 + a_1\sum t_i = \sum g_i\,; \\ a_0\sum t_i + a_1\sum t_i^2 = \sum t_i g_{i.} \end{cases}$$

Приведенные системы для линейной зависимость можно упростить, если перенести начало координат в середину ряда динамики (правило переноса см. в п. 3.3.). При этом и количество слагаемых в уравнениях уменьшается.

Тогда неизвестные a_0 и a_1 можно вычислить по формулам

$$a_0 = \frac{\sum lng_i}{n},$$

(3.3.10)

$$a_1 = \frac{\sum t_i^{/} lng_i}{\sum \left(t_i^{/}\right)^2},$$

где n – число наблюдений.

Пример расчета численности дрожжей в зависимости от времени. Данные наблюдений приведены во втором и третьем столбцах таблицы 3.3.3.

Алгоритм

1) Выберем в качестве минимального значения $y_{min} = 20$, максимального - $y_{max} = 600$.

2) Рассчитаем по формуле (3.3.9) g_i: $g_1 = \frac{600-20}{25-20} = 115$ и т.д.

3) Вычислим lng_i. $lng_1 = ln115 = 4{,}74$ и т.д.

4) Перенесем значение $t = 0$ в середину ряда.

5) Вычислим $t^{/} lng_i$.

6) Вычислим $(t^{/})^2$.

7) Рассчитаем по формулам (3.3.10) параметры a_0 и a_1 при n=16-1=15, т.к. число расчетных значений уменьшилось на единицу.

$$a_0 = \frac{1{,}81}{15} = 0{,}12,$$

$$a_1 = \frac{-150{,}84}{280} = -0{,}54.$$

8) Вычислим $a_0 + a_1 t_i^/$

для $t_1^/ = 0{,}12 - 0{,}54 \cdot (-7) = 3{,}9.$

9) Вычислим $\widehat{g_i} = e^{a_0 + a_1 t_i^/}$.

10) Вычислим $\widehat{y_i} = \frac{y_{max} - y_{min}}{\widehat{g_i} + 1} + y_{min}$

$$\widehat{y_1} = \frac{600 - 20}{49{,}40 + 1} + 20 = 31{,}51 \text{ и т.д.}$$

11) В последнем столбце таблицы находится разница между наблюдаемыми и теоретическими значениями

$$e_i = y_i - \widehat{y_i}.$$

Все расчеты представим в таблице.

Таблица 3.3.3 - Расчет параметров логистической кривой

n	Время, ч	y_i	g_i	lng_i	$t^/$	$t^/ lng_i$	$(t^/)^2$	$a_0 + a_1 t_i^/$	$\widehat{g_i}$	$\widehat{y_i}$	e_i
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
1	1	25	115	4,74	-7	-33,18	49	3,90	49,40	31,51	-6,51
2	2	50	18,33	2,91	-6	-17,46	36	3,36	28,79	39,47	10,53
3	3	60	13,50	2,60	-5	-13	25	2,82	16,78	52	8
4	4	75	9,55	2,26	-4	-9,04	16	2,28	9,78	73,81	1,19
5	5	10	6,25	1,83	-3	-5,49	9	1,74	5,70	106,5	6,57
6	6	15	3,46	1,24	-2	-2,48	4	1,20	3,32	154,2	4,26
7	7	26	1,42	0,35	-1	-0,35	1	0,66	1,93	217,9	42,05
8	8	30	1,07	0,07	0	0	0	0,12	1,13	292,3	-7,7
9	9	41	0,49	-0,71	1	-0,71	1	-0,42	0,66	362,4	47,60
10	10	46	0,32	-1,14	2	-2,28	4	-0,96	0,38	440,2	20,29
11	11	49	0,23	-1,47	3	-4,41	9	-1,50	0,22	495,4	-5,41
12	12	53	0,14	-1,97	4	-7,88	16	-2,04	0,13	533,2	-3,27
13	13	55	0,09	-2,41	5	-12,05	25	-2,58	0,08	557,0	-7,01
14	14	57	0,05	-2,99	6	-17,94	36	-3,12	0,04	577,6	-7,69
15	15	58	0,03	-3,51	7	-24,57	49	-3,66	0,03	583,1	1,83
16	16	60	0	-	-	-	-	-	-	-	-
	$\sum$	**-**	**-**	**1,81**	**0**	**-150,84**	**280**	**-**	**-**	**-**	**-**

Ниже приведены графики эмпирической и теоретической кривой. Видим, что наблюдаемые и расчетные значения близки друг к другу, и точки эмпирического графика располагаются достаточно близко к теоретической кривой.

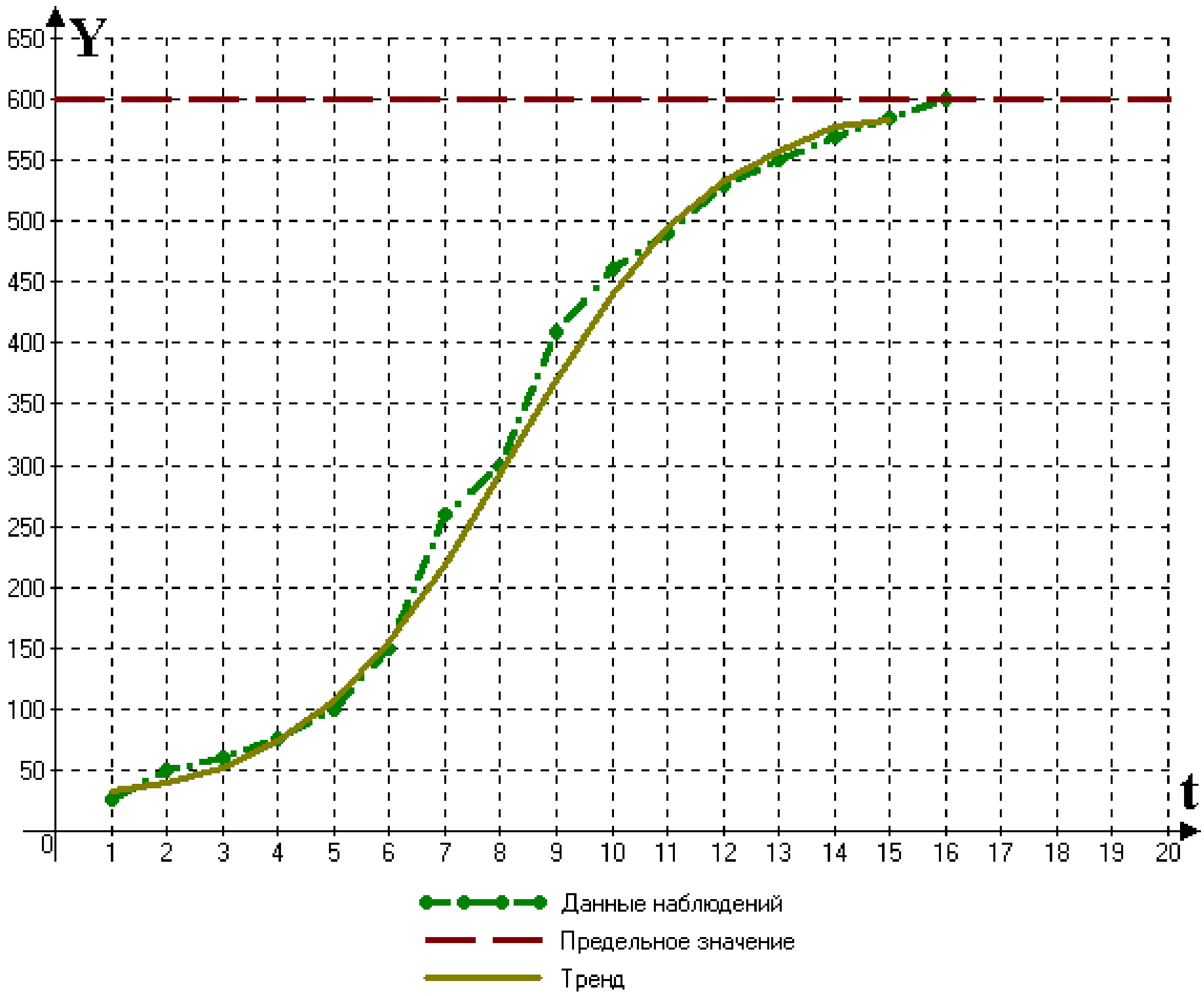


Рисунок 3.3 – Эмпирический и теоретический графики

3.4. Проверка точности модели

Проводится с целью оценки ошибки в подборе аппроксимирующей функции.

1) *Оценка стандартной ошибки* (среднее квадратическое отклонение)

$$S_{\hat{y}_i} = \sqrt{\frac{1}{n-(m+1)}\sum e_i^2}, \qquad \textbf{(3.4.1)}$$

где n – число наблюдений, m – число параметров модели.

2) *Средняя относительная ошибка аппроксимации*

$$|\bar{e}| = \frac{1}{n}\sum\left|\frac{e_i}{y_i}\right| \cdot 100\%. \qquad \textbf{(3.4.2)}$$

3) *Коэффициент сходимости*

$$\varphi^2 = \frac{\sum(e_i)^2}{\sum(y_i - \bar{y})^2}, \qquad \textbf{(3.4.3)}$$

где y_i – наблюдаемые значения, $\overline{y}$ – среднее значение из наблюдаемых. Эта величина показывает долю изменения у, объясняемую изменением включенных в модель факторов. Составляются отношения суммы квадратов разностей эмпирических и теоретических значений функций к сумме квадратов разностей между эмпирическими значениями и средним между эмпирическими.

4) *Коэффициент детерминации*

$$D = 1 - \varphi^2. \qquad \textbf{(3.4.4)}$$

5) *Коэффициент корреляции*

$$R = \sqrt{D}\ . \qquad \textbf{(3.4.5)}$$

Коэффициент корреляции показывает тесноту связи наблюдаемых значений и значений функции, вычисленной по выведенной формуле. Если выбор стоит между несколькими теоретическими моделями, то та модель точнее описывает исследуемый процесс, у которой коэффициент корреляции по модулю ближе к единице.

<u>Вывод</u>: модели, для которых $S_{\hat{y}}$, $\overline{e}$, φ^2минимальны, аD и R - максимальны, точнее отображают исследуемый процесс.

Пример. Найдем значения $S_{\hat{y}}$, $\overline{e}$, φ^2, D и R для полученных ранее линейной зависимости (3.3.6) и параболы (3.3.7).

1) Рассчитаем теоретические значения $\hat{y}_1$ и $\hat{y}_2$, подставляя t_i в уравнения $\hat{y}_1 = 3695 + 84t$ и
$\hat{y}_2 = -7846 + 2992t - 138t^2$:

Таблица 3.4.1 – Проверка точности модели

№	t	y	$\hat{y}_1$	$\hat{y}_2$	e_1	e_2	e_1^2	e_2^2	$\lvert e_{i1} / y_i \rvert$	$\lvert e_{i2}/y_i \rvert$
1	2	90	3781	-2414	90-3781=-	90-(-	3691^2=13623481	2504^2=	|−3691/	|2504/
2	3	92	3863	-112	-3771	204	14220441	41616	41	2,22
3	4	88	4031	1914	-3943	-1826	15547249	3334276	44,81	20,75
4	5	562	4115	3664	-3553	-3102	12623809	9622404	6,32	5,52
5	6	3863	4199	5138	-336	-1275	112896	1625625	0,09	0,33
6	7	6833	4283	6336	2550	497	6502500	247009	0,37	0,07
7	8	9500	4367	7258	5133	2242	26347689	5026564	0,54	0,24
8	9	8167	4451	7904	3716	263	13808656	69169	0,46	0,03
9	10	8333	4535	8274	3798	59	14424804	3481	0,46	0,01
10	11	8167	4619	8368	3548	-201	12588304	40401	0,43	0,02
11	12	8833	4703	8186	4130	647	17056900	418609	0,47	0,07
12	13	8500	4787	7728	3713	772	13786369	595984	0,44	0,09
13	14	7833	4871	6994	2962	839	8773444	703921	0,38	0,12
14	15	5833	4955	5984	878	-151	770884	22801	0,15	0,03
15	16	3300	5039	4698	-1739	-1398	3024121	1954404	0,53	0,42
16	17	1233	5123	3136	-3890	-1903	15132100	3621409	3,15	1,54
17	18	633	5207	1298	-4574	-665	20921476	442225	7,23	1,05
18	19	467	5291	-816	-4824	1283	23270976	1646089	10,33	2,75
∑	**189**	**82327**	**82220**	**83538**	**107**	**-1211**	**232536099**	**35686003**	**158,16**	**63,07**

Таблица 3.4.2 – Расчет показателей моделей

№	t	y	$\hat{y}_1$	$\hat{y}_2$	$y-\bar{y}$	$(y-\bar{y})^2$
1	2	90	3781	-2414	90-4574=-4484	(-4484)2=20106256
2	3	92	3863	-112	92-4574=-4482	4482^2=20088324
3	4	88	4031	1914	-4486	20124196
4	5	562	4115	3664	-4012	16096144
5	6	3863	4199	5138	-711	505521
6	7	6833	4283	6336	2259	5103081
7	8	9500	4367	7258	4926	24265476
8	9	8167	4451	7904	3593	12909649
9	10	8333	4535	8274	3759	14130081
10	11	8167	4619	8368	3593	12909649
11	12	8833	4703	8186	4259	18139081
12	13	8500	4787	7728	3926	15413476
13	14	7833	4871	6994	3259	10621081
14	15	5833	4955	5984	1259	1585081
15	16	3300	5039	4698	-1274	1623076
16	17	1233	5123	3136	-3341	11162281
17	18	633	5207	1298	-3941	15531481
18	19	467	5291	-816	-4107	16867449
$\sum$	**189**	**82327**	**82220**	**83538**		**237181383**

Например, при t=2 $\hat{y}_1(2) = 3695 + 84 \cdot 2 = 3863$ и $\hat{y}_2(2) = -7846 + 2992 \cdot 2 - 138 \cdot 2^2 = -112$ и т.д.

$e_1 = y - \hat{y}_1$; $e_2 = y - \hat{y}_2$.

1) *Оценку стандартной ошибки* вычислим по формуле (3.4.1)

$$S_1 = \sqrt{\frac{1}{18-(2+1)} \cdot 232536099} = 3937;$$

$$S_2 = \sqrt{\frac{1}{18-(3+1)} \cdot 35686003} = 1597.$$

2) *Средняя относительная ошибка* (формула (3.4.2)):

$$|\overline{e_1}| = \frac{1}{18} \cdot 158{,}16 \cdot 100\% = 879\%,$$

$$|\overline{e_2}| = \frac{1}{18} \cdot 63{,}07 \cdot 100\% = 350\%.$$

3) *Коэффициент сходимости*

$\overline{y} = \frac{\sum y_t}{n} = \frac{82327}{18} = 4574 -$ среднее среди наблюдаемых значений.

Согласно формуле (3.4.3) и расчетным данным из таблиц 3.4.1 и 3.4.2,

$$\varphi_1{}^2 = \frac{232536099}{237181383} = 0{,}98;$$

$$\varphi_2{}^2 = \frac{35686003}{237181383} = 0{,}15.$$

4) Коэффициент детерминации

$$D = 1 - \phi^2;$$

$$D_1 = 1 - 0{,}98 = 0{,}02;$$
$$D_2 = 1 - 0{,}15 = 0{,}85.$$

5) Коэффициент корреляции

$$R = \sqrt{D};$$

$$R_1 = \sqrt{0{,}02} = 0{,}14;$$
$$R_2 = \sqrt{0{,}85} = 0{,}92.$$

Вывод: так как $S_2 < S_1$, $\hat{e}_2 < \hat{e}_1$, $\varphi_2{}^2 < \varphi_1{}^2$, а $D_2 > D_1$, $R_2 > R_1$, то параболическая функция более точно описывает исследуемый процесс.

3.5. Проверка адекватности модели

Качество модели определяется ее адекватностью исследуемому процессу. Адекватность характеризуется наличием определенных статистических свойств. Модель прогнозирования считается адекватной, если она учитывает существенную закономерность исследуемого процесса. Закономерность исследуемого процесса находит отражение в наличии определенных статистических свойств остаточной компоненты. Остаточная компонента получается после выделения из исследуемого ряда тренда $\hat{y}_t$ и периодической составляющей e_t

$$y_t = \hat{y}_t + e_t .$$

Составляющая e_t находится по формуле

$$e_t = y_t - \acute{y}_t ,$$

где y_t - фактические уровни, $\acute{y}_t$ - расчетные.

Компонента $\hat{y}_t$ найдена правильно тогда и только тогда, когда случайная компонента e_t удовлетворяет следующим условиям, доказывающим, что она носит случайный характер:

1) случайность,

2) соответствие распределения остаточной компоненты нормальному закону распределения,

3) равенство нулю математического ожидания,

4) независимость значений ряда остатков.

1) *Проверка на случайность* предполагает проверку независимости величины e_t от времени. Производится с помощью критерия «пиков» и «впадин»: значения e_t сравнивается с двумя соседними с ним. Если оно больше или меньше их ($e_{t-1} > e_t < e_{t+1}$ или $e_{t-1} < e_t > e_{t+1}$), то точка считается поворотной. Далее подсчитывается сумма поворотных точек P. Если $P > \left(\bar{p} - 1{,}96\sqrt{s^2}\right)$, то ряд остатков является случайным.

$\bar{p} = \frac{2n-1}{3}$- математическое ожидание числа пиков,

$$s^2 = \frac{16n-29}{90} \text{ - дисперсия.}$$

Таким образом,

$$P > \left(\frac{2n-4}{3} - 1{,}96\sqrt{\frac{16n-29}{90}}\right). \qquad \textbf{(3.5.1)}$$

Учитывается целая часть числа в скобках.

2) *Проверка соответствия распределения остаточной компоненты нормальному закону распределения.*

Нормальное распределение (называемое также распределением Гаусса) - одно из важнейших понятий в математической статистике. Оно характеризуется тем, что крайние значения признака в нем встречаются достаточно редко, а значения, близкие к средней величине – часто. Нормальное распределение возникает, когда исследуемая случайная величина представляет собой сумму большого числа независимых случайных величин, каждая из которых играет в общем эксперименте незначительную роль. Графически нормальное распределение имеет колоколообразную форму, значения моды, медианы и среднего арифметического равны между собой. Было установлено, что многие биологические параметры подчиняются этому закону. При этом, чем больше объем выборки, тем более полученное эмпирическое распределение приближается к нормальному.

Основными свойствами ряда остатков являются их симметричность относительно тренда и преобладание малых по абсолютной величине ошибок над большими. В этой связи определяются коэффициенты асимметрии A_s (мера «скошенности» от математического ожидания) и эксцесса - $Э_s$ (мера «скученности») наблюдений около моды

$$A_s = \frac{\left[\frac{1}{n}\sum_{i=1}^{n} e_i^{\,3}\right]}{\left(\sqrt{\frac{1}{n}\sum_{i=1}^{n} e_i^{\,2}}\right)^3}; \qquad \textbf{(3.5.2)}$$

$$Э_s = \frac{\frac{1}{n}\sum_{i=1}^{n} e_i^{\,4}}{\left(\frac{1}{n}\sum_{i=1}^{n} e_i^{\,2}\right)^2} - 3. \qquad \textbf{(3.5.3)}$$

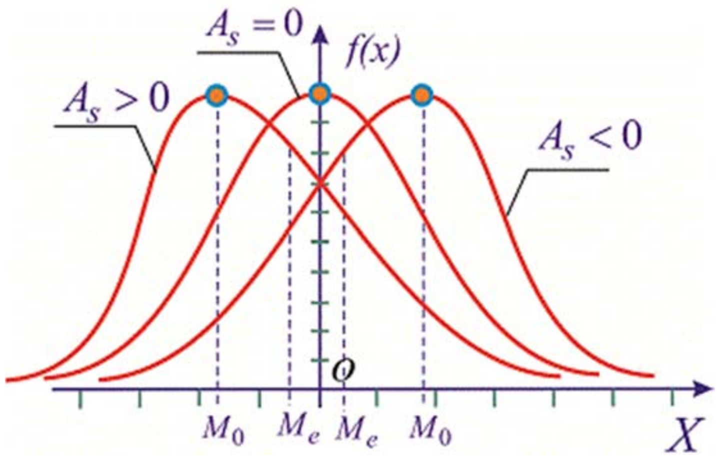


Рисунок 3.5.1 - Асимметрия

Если коэффициент асимметрии отрицателен, то либо большая часть значений случайной величины, либо мода находятся левее математического ожидания, и наоборот, если асимметрия положительна, то правее.

Эксцесс случайной величины - величина, характеризующая степень островершинности или плосковершинности распределения, т.е. степень «выпада».

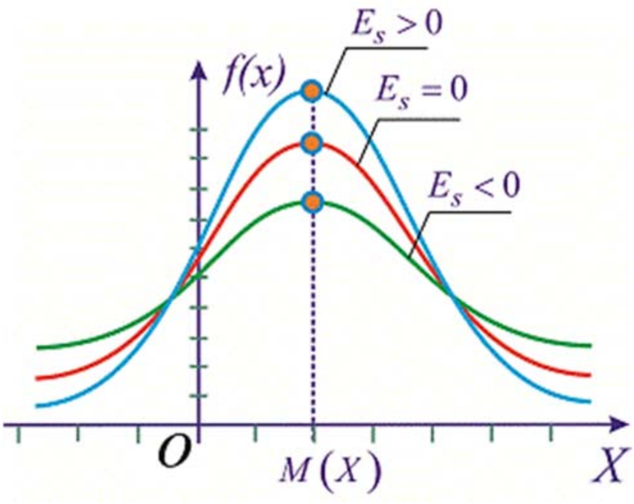


Рисунок 3.5.2 - Эксцесс

Для нормального закона распределения $E_s = 0$. Если эксцесс положительный $(E_s > 0)$, то на графике функция распределения имеет острую вершину и для отрицательных значений $(E_s < 0)$ более пологую. Таким образом, можно наглядно установить отклонения заданного закона от нормального.

Далее вычисляется среднеквадратическая ошибка выборочной характеристики асимметрии

$$S_A = \sqrt{\frac{6(n-2)}{(n+1)(n+3)}} \qquad \textbf{(3.5.4)}$$

и среднеквадратическая ошибка выборочной характеристики эксцесса

$$S_Э = \sqrt{\frac{24n(n-2)(n-3)}{(n+1)^2(n+3)(n+5)}}. \qquad \textbf{(3.5.5)}$$

Если

$$|A_s| \leq 1{,}5S_A \text{ и } |Э_s| \leq 1{,}5S_Э, \qquad \textbf{(3.5.6)}$$

то распределение ряда остатков не противоречит нормальному закону.

Если

$$|A_s| \geq 2S_A \text{ и } |Э_s| \geq 2S_Э, \qquad \textbf{(3.5.7)}$$

то данные не являются нормальными. Их применение в дальнейшем не рекомендуется.

Если $1{,}5\ S_A < |A_s| < 2\ S_A$ и $1{,}5\ S_Э < |Э_s| < 2\ S_Э$, то можно воспользоваться *RS*-критерием:

$$RS = \frac{e_{max} - e_{min}}{S},$$

где e_{max} и e_{min} - максимальный и минимальный уровни ряда остатков, S - среднеквадратическое отклонение остатков.

Если значение критерия попадает между табулированными границами с заданным уровнем значимости (Приложение 3), то гипотеза о нормальном распределении ряда остатков принимается.

3) *Проверка равенства нулю математического ожидания* производится на основе *t*-критерия Стьюдента

$$t_{расч} = \left| \frac{1}{n} \sum_{i=1}^{n} e_i \right| \frac{\sqrt{n}}{S}, \qquad \textbf{(3.5.8)}$$

где S - среднеквадратическое отклонение остатков

$$S = \sqrt{\frac{\sum e_i{}^2}{n} - (\overline{e})^2}. \qquad \textbf{(3.5.9)}$$

Гипотеза о равенстве нулю принимается, если $t_{расч} < t_{таб}$ при $k = n - 1$ степенях свободы.

4) *Проверка независимости значений ряда остатков* подразумевает выявление отсутствия автокорреляции.

Присутствие автокорреляции ряда остатков может свидетельствовать о наличии ошибок измерений или о неправильном выборе функции, которая не отражает воздействие важного фактора.

Под автокорреляцией последовательности значений понимается корреляция между членами этой последовательности, передвинутой, в данном случае, на одну единицу. Последствия, вызываемые автокорреляцией остатков:

- недооценка дисперсии остатков функции регрессии;
- наличие ошибки при оценке выборочной дисперсии параметров регрессии, которая препятствует применению метода наименьших квадратов для нахождения трендовой модели динамического ряда.

Критерием, позволяющим установить наличие автокорреляции, является *критерий Дарбина-Уотсона*

$$d = \frac{\sum (e_t - e_{t-1})^2}{\sum e_t^{\ 2}}. \qquad \textbf{(3.5.10)}$$

Значения критерия d находятся в интервале [0,4] и сравниваются со значениями $d_н$ и $d_в$, взятыми из таблицы Приложения 4. Здесь α=5% - уровень значимости, n – объем выборки.

Вывод:

1) $d<d_н$- автокорреляция есть;
2) $d>d_в$- автокорреляции нет;
3) $d_н<d<d_в$ - при выбранном уровне значимости нельзя прийти к определенному выводу.

Если $2<d<4$, то для проверки нужно найти $d'=4-d$.

Если автокорреляция ряда остатков присутствует, то это значит, что исследуемый процесс подвержен цикличности.

Заключение. По результатам проверки всех свойств в целом можно сделать вывод о соответствии (адекватности) модели исследуемому процессу и возможности ее использования для прогнозирования.

Пример. В предыдущем исследовании теоретических моделей на точность получено, что более точно исследуемый процесс описывает параболический тренд. Поэтому проверим эту модель на адекватность.

1) *Проверим на случайность остаточную компоненту* полученного ранее параболического тренда

$$\hat{y}_2 = -7846 + 2992t - 138t^2.$$

При n=18 согласно формуле (3.5.1)

$$\frac{2n-4}{3} - 1{,}96\sqrt{\frac{16n-29}{90}} = \frac{2\cdot 18-4}{3} - 1{,}96\sqrt{\frac{16\cdot 18-29}{90}} = 7{,}35.$$

Расчеты в таблице 3.5.1 показали, что P=5.

Так как 5<7,35, значит, ряд остатков не является случайным.

2) *Проверим соответствие распределения остаточной компоненты нормальному закону распределения.* Найдем значения коэффициентов асимметрии и эксцесса, руководствуясь формулами (3.5.2) и (3.5.3) и их оценками, вычисленными по формулам (3.5.4) и (3.5.5):

Таблица 3.5.1 – Расчеты для проверки модели на адекватность

№	t	y_t	$\hat{y}_t$	$y_i - \bar{y}_i$	P	e_t^2	e_t^3	e_t^4	$e_t - e_{t-1}$	$(e_t - e_{t-1})^2$
1	2	90	-2414	2504	-	6270016	15700120064	3,93131E+13	-	-
2	3	92	-112	204	0	41616	8489664	1731891456	-500	250000
3	4	88	1914	-1826	0	3334276	-6088387976	1,11174E+13	-2030	4120900
4	5	562	3664	-3102	1	9622404	-29848697208	9,25907E+13	-1276	1628176
5	6	3863	5138	-1275	0	1625625	-2072671875	2,64266E+12	1827	3337929
6	7	6833	6336	497	0	247009	122763473	61013446081	1772	3139984
7	8	9500	7258	2242	1	5026564	11269556488	2,52663E+13	1745	3045025
8	9	8167	7904	263	0	69169	18191447	4784350561	-1979	3916441
9	10	8333	8274	59	0	3481	205379	12117361	-204	41616
10	11	8167	8368	-201	1	40401	-8120601	1632240801	-260	67600
11	12	8833	8186	647	0	418609	270840023	1,75233E+11	848	719104
12	13	8500	7728	772	0	595984	460099648	3,55197E+11	125	15625
13	14	7833	6994	839	1	703921	590589719	4,95505E+11	67	4489
14	15	5833	5984	-151	0	22801	-3442951	519885601	-990	980100
15	16	3300	4698	-1398	0	1954404	-2732256792	3,81969E+12	-1247	1555009
16	17	1233	3136	-1903	1	3621409	-6891541327	1,31146E+13	-505	255025
17	18	633	1298	-665	0	442225	-294079625	1,95563E+11	1238	1532644
18	19	467	-816	1283	-	1646089	2111932187	2,70961E+12	1948	3794704
$\sum$			**83538**	**-1211**	**5**	**35686003**	**-17386410263**	**1,91865E+14**	**579**	**28404371**

$$A_s = \frac{\frac{1}{18}\cdot(-17386410263)}{\left(\sqrt{\frac{35686003}{18}}\right)^3} = -0{,}035;$$

$$Э_s = \frac{\frac{1}{18}\cdot 1{,}91865\cdot 10^{14}}{\left(\frac{1}{18}\cdot 35686003\right)^2} - 3 = -0{,}28;$$

$$S_A = \sqrt{\frac{6(18-2)}{(18+1)(18+3)}} = 0{,}491;$$

$$S_Э = \sqrt{\frac{24\cdot 18(18-2)(18-3)}{(18+1)^2(18+3)(18+5)}} = 0{,}77\ .$$

$$1{,}5S_A = 1{,}5\cdot 0{,}491 = 0{,}7365;$$

$$1{,}5S_Э = 1{,}5\cdot 0{,}77 = 1{,}155.$$

$$|A_s| = 0{,}035 < 0{,}7365;$$

$$|Э_s| = 0{,}28 < 1{,}155.$$

В нашем исследовании выполнены условия (3.5.6). Значит, распределение остаточной компоненты соответствует нормальному закону.

3) *Проверка гипотезы о равенстве нулю математического ожидания:*

Для расчета среднеквадратического отклонения остатков S по формуле (3.5.9) предварительно вычислим математическое ожидание остаточной компоненты:

$$\overline{e} = \frac{\sum e_i}{n};$$

$$\overline{e} = \frac{-1211}{18} = -67{,}28;$$

$$\overline{e}^2 = (-67{,}28)^2 = 4526{,}3;$$

$$S = \sqrt{\frac{35686003}{18} - 4526{,}3} = 1406.$$

Согласно формуле (3.5.8)

$$t_{расч} = \left| \frac{1}{18} \cdot (-1211) \right| \cdot \frac{\sqrt{18}}{1406} = 0{,}203;$$

$t_{таб} = 2{,}11$ при k=18-1=17, α=0,95 (Приложение 2).

$t_{расч} < t_{таб}$, поэтому гипотеза о равенстве нулю математического ожидания принимается.

4) Проверка независимости значений ряда остатков производится по формуле (3.5.10):

$$d = \frac{28404371}{35686003} = 0{,}796 \;;$$

при k=2, n=18 $d_н$=1,05; $d_в$=1,54 (Приложение 4);

$$d < d_н.$$

Автокорреляция есть.

Вывод: по результатам проверки заключаем, что параболическая модель исследуемого тренда по двум из четырех пунктов адекватна исследуемому процессу. Так как из двух исследуемых моделей более точно данные наблюдений описывает парабола, то решение об использовании этой модели для прогнозирования поведения биологического процесса принимает исследователь. Возможен вариант отыскания параметров другой функции и сравнение ее показателей с параболой.

3.6. Расчет доверительного интервала в точке прогноза

После того, как получено уравнение кривой, по которой развивается исследуемый процесс, можно спрогнозировать значение исследуемого показателя в момент времени *t*, подставляя соответствующее значение *t* в уравнение. Такой прогноз называется *точечным*.

Далее можно рассчитать, в каком интервале будет находиться полученное прогнозное значение, то есть вычислить *интервальную оценку*.

Доверительный интервал вычисляется по формуле

$$\hat{y}_t - t_\alpha \frac{S_{\hat{y}}}{\sqrt{n}} < \hat{y}_t < \hat{y}_t + t_\alpha \frac{S_{\hat{y}}}{\sqrt{n}}, \qquad \mathbf{(3.6.1)}$$

где

$$S_{\hat{y}} = \sqrt{\frac{1}{n-(m+1)} \sum e_i^2}. \qquad \mathbf{(3.6.2)}$$

$S_{\hat{y}}$ - среднеквадратическое отклонение фактических наблюдений от расчетных (оценка стандартной ошибки),

n – число наблюдений,

m – число коэффициентов модели,

t_α - табличное значение *t*-критерия Стьюдента при уровне значимости α и числе степеней свободы $f=n-1$.

Итогом работы по прогнозированию экономических процессов будет служить его обобщенная характеристика, которая включает:

1) вид уравнения тренда;
2) оценки адекватности и точности модели;
3) точечные прогнозные оценки;
4) интервальные прогнозные оценки.

Задачи к главе 3

3.1. По данным наблюдений за изменением популяции простейших в изолированной культуре изобразить графически динамический ряд. По виду кривой сделать предположение, какая функция описывает данную зависимость.

Время, сутки	2	4	6	8	10	12	14	16
Paramecium aurelia	18	75	100	95	87	84	82	81

3.2. По данным таблицы показать общую тенденцию в динамике междугородного телефонного обмена путем исчисления скользящей средней. Изобразить графически исходные и выровненные ряды.

Месяцы	июль	август	сентябрь	октябрь	ноябрь	декабрь	январь	февраль	март	апрель	май	июнь
Сухая биомасса, мг	1,2	0,9	1,7	1,4	0,6	1,1	0,2	0,4	0,5	1,3	2,8	0,8

3.3. Выровнять тренд по трехдневной скользящей средней

Время, *ч*	0	5	10	15	20	25	30	35	40	45
Количество живых клеток	9	18	400	550	550	530	340	225	175	30

По графику сделать прогноз на 50 сутки.

3.4. По данным таблицы 3.5.2 произвести выравнивание по трехдневной и пятидневной скользящей средней. Построить на одном рисунке графики эмпирической кривой и сглаженные кривые.

3.5. Проверить наличие трендовой зависимости:

Годы	2002	2003	2004	2005	2006	2007
Данные наблюдений	18	21	26	22	25	28

а) методом разности средних;

б) методом Фостера-Стьюарта.

С какой вероятностью при проверке методом Фостера-Стьюарта можно утверждать, что тренд существует?

Таблица 3.5.2 – Данные для задачи 3.4

№ варианта	***t*, время**	**Эмпирические данные, y_i**															
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**
1	1	37	42	33	45	58	55	56	70	69	74	71	86	70	92	68	93
2	2	261	242	332	472	312	251	427	528	541	465	582	529	524	509	563	522
3	3	38	44	50	57	66	76	84	94	102	124	134	148	-	-	-	-
4	4	4,3	3,2	5,7	7,0	9,2	6,7	7,5	8,9	10,5	12,6	15,0	12,5	14,6	-	-	-
5	5	249	230	320	460	300	239	415	516	529	453	570	517	512	497	551	510
6	6	48	54	60	67	76	86	94	104	112	134	144	158	170	181	192	204
7	7	21	23	27	32	30	30	35	39	30	34	36	39	38	40	41	42
8	8	16	18	20	19	22	21	23	22	25	24	26	27	27	28	30	32
9	9	34	39	30	33	53	50	52	66	65	67	70	82	66	80	64	83
10	10	43	48	38	41	62	59	61	75	74	76	79	91	75	89	73	92
11	11	6	8	11	14	17	20	23	26	29	32	35	38	-	-	-	-
12	12	7	8	12	10	14	13	15	16	19	19	20	23	23	22	25	27
13	13	2	3	7	5	9	8	10	11	14	14	15	17	18	18	20	22
14	14	6	8	10	9	12	11	13	12	15	14	16	17	17	18	20	22
15	15	54	57	62	65	67	69	70	74	78	80	83	84	89	92	-	-
16	16	34	36	39	44	52	55	59	65	69	72	76	83	88	90	-	-
17	17	20	18	15	19	26	24	30	28	33	37	36	38	42	42	45	48
18	18	19	25	22	21	23	23	28	26	32	34	33	33	36	37	39	41
19	19	19	22	27	26	30	32	36	37	39	42	45	47	50	52	54	57
20	20	25	28	33	32	36	38	42	43	45	48	51	53	56	58	60	63

3.6. В результате исследования зависимости роста биомассы дрожжей от времени получили следующие данные:

t, часов	1	2	3	4	5	6	7	8	9
Биомасса дрожжей, г	2	3	4,7	7	11	16	25	35	50

Построить график. Определить характер зависимости. Определить коэффициенты тренда.

3.7. Определить коэффициенты тренда $y = a_0 + a_1 t + a_2 t^2$

t	1	2	3	4	5	6	7	8	9	10
y	14	15	17	20	24	30	48	49	59	67

Спрогнозировать значение *y* при *t*=11.

3.8. Определить коэффициенты тренда $y = a_0 + a_1 t + a_2 t^2$

t,годы	2002	2003	2004	2005	2006	2007
y	18	21	26	22	25	28

Спрогнозировать значение *y* в 2009 году.

3.9. Имеются данные о росте числа особей в популяции за некоторый период времени:

t, сутки	1	2	3	4	5	6
y,число особей	12	30	80	100	200	1000

По этим данным получена зависимость

$$\hat{y} = 73t^2 - 357t + 375.$$

Проверить:

1) на случайность остаточную компоненту данного тренда;

2) соответствие распределения остаточной компоненты нормальному закону распределения;

3) равенство нулю математического ожидания;

4) независимость значений ряда остатков.

3.10. Динамика роста однолетнего растения отражена в таблице:

***t*,недели**	2	4	6	8	10	12
***y*,сухая масса (г)**	5	11	19	35	44	70

По этим данным получено линейное уравнение регрессии

$$\hat{y}_t = 6{,}3t - 13{,}3\ .$$

Проверить данные модели на точность. Найти:

1) оценку стандартной ошибки (среднее квадратическое отклонение);
2) среднюю относительную ошибку аппроксимации;
3) коэффициент сходимости;
4) коэффициент детерминации;
5) коэффициент корреляции.

3.11. По данным задачи **2.9.** **с**делать точечный прогноз сухой массы при *t=14* и рассчитать доверительный интервал в точке прогноза с вероятностью 95%.

Задание к расчетно-графической работе

По данным таблицы (3.5.3)

1) найти коэффициенты параболического тренда;
2) рассчитать параметры логистической кривой;
3) проверить модели на адекватность, выбрать лучшую;
4) проверить точность выбранной модели;
5) спрогнозировать значение y при t=21.
6) рассчитать доверительный интервал в точке прогноза.

Таблица 3.5.3 - Данные для расчетно-графической работы

***t*, время**	**вариант**																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	5	90	40	40	80	30	20	70	180	25	30	76	38	12	80	60	90	12	23	
2	10	110	60	50	85	50	30	80	200	30	35	94	42	30	85	80	96	14	30	
3	60	170	100	55	90	60	35	90	210	40	50	98	50	40	90	120	100	15	36	
4	200	250	180	60	100	70	50	95	300	60	65	100	65	50	104	170	103	18	38	
5	430	360	270	70	110	75	60	97	380	110	100	136	100	80	146	225	120	19	40	
6	450	500	380	95	120	85	90	105	475	160	150	148	150	110	180	290	130	20	46	
7	690	620	490	115	130	110	130	130	630	210	200	200	200	150	230	350	150	22	57	
8	700	780	600	165	140	130	180	170	650	280	250	259	258	200	280	405	180	30	66	
9	760	890	680	250	185	160	230	200	750	340	365	330	340	240	330	490	216	43	84	
10	790	920	740	345	250	215	270	230	820	385	450	400	400	280	386	540	270	58	116	
11	820	930	800	405	300	285	310	270	880	440	540	460	464	336	420	570	310	70	140	
12	850	960	830	460	360	330	330	300	910	480	600	512	502	350	450	610	360	81	160	
13	860	975	880	510	420	380	350	330	950	510	648	550	550	380	490	640	430	88	176	
14	870	980	895	530	460	410	365	360	970	550	700	580	567	400	527	650	500	90	180	
15	885	985	910	550	480	430	390	400	975	570	730	650	600	418	548	665	560	91	182	
16	890	1000	915	565	520	460	400	410	985	580	750	687	610	440	560	678	630	93	188	
17	900			570	540	480	405	425	990	585	758	720	616	452	568	700	670	94	194	
18					570	500	410	430	1000	590	774	735	640	468	580		700	96	200	
19					575			440		600	780	750	650	474	588		740	98		
20					600			450			790	770	660	480	600		750	100		

Вопросы к главе 3

1. Что такое тренд? Для чего он применяется?
2. Что нужно для нахождения тренда?
3. Как произвести сглаживание графика эмпирической кривой?
4. Как рассчитать пятидневные скользящие средние?
5. Охарактеризуйте три группы кривых роста.
6. Для чего применяют метод разности средних?
7. Для чего применяют метод наименьших квадратов?
8. Запишите систему уравнений для нахождения параметров линейной зависимости.
9. Сколько параметров у параболического тренда? Сколько уравнений нужно составить для их отыскания?
10. Какая кривая может охарактеризовать рост популяции?
11. Какие характеристики вычисляются для проверки точности модели?
12. Какие статистические свойства исследуются для проверки модели на адекватность?

ГЛАВА 4. СИСТЕМЫ МАССОВОГО ОБСЛУЖИВАНИЯ

4.1. Построение вероятностных моделей с помощью теории систем массового обслуживания

Системы массового обслуживания (СМО) – это системы, предназначенные для обслуживания какого-либо потока заявок. Методы изучения этих систем применяются в биологии, медицине, физике, астрономии, технике.

Теория марковских и пуассоновских процессов – основа для описания случайных процессов, протекающих в СМО.

Элементы теории случайных процессов в биологии позволяют моделировать экологические системы с учетом влияния самых разнообразных факторов: размножения, гибели, миграций, мутаций, конкуренции и т.д. В теории массового обслуживания широко распространен специальный класс случайных процессов так называемые процессы гибели и размножения. Название это связано с рядом биологических задач, в которых этот процесс служит математической моделью изменения численности биологических популяций. Система массового обслуживания представляет систему дискретного типа с конечным множеством состояний. Переход системы из одного состояния в другое происходит скачками, в момент, когда осуществляется какое-то событие. В системе массового обслуживания предпочтение имеет марковский процесс.

Объект, занимающийся обслуживанием заявок, называется *каналом обслуживания*.

4.2. Построение вероятностных моделей при изучении марковских случайных процессов

Примером марковского случайного процессом может служить детская настольная игра с такими правилами: фишка игрока должна пройти некоторое число пунктов от 1 до n. Переход из одного пункта в другой каждый раз определяется исходом бросания игральной кости. Фишка проходит столько пунктов, сколько очков выпало на игральной кости. Из любого пункта i фишка с некоторой вероятностью p_{ij} переходит в один из пунктов j независимо от попадания в пункт i. Победителем становится игрок, первым достигший пункта n.

Марковские случайные процессы описываются с помощью матриц вероятностей перехода. С матрицами мы знакомимся в курсе алгебры, а о вероятности состояний – в курсе теории вероятностей. Здесь понятие «матрица» рассматривается с вероятностной точки зрения.

Определение. Пусть в момент времени n система находится в состоянии S_k. Условную вероятность p_{jk} будем называть вероятностью перехода из состояния S_j в состояние S_k. Вероятности перехода p_{jk} могут быть записаны в виде *матрицы вероятностей перехода*

$$P = \begin{pmatrix} p_{11} & p_{12} & p_{13} & \cdots \\ p_{21} & p_{22} & p_{23} & \cdots \\ p_{31} & p_{32} & p_{33} & \cdots \\ \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots \end{pmatrix},$$

где первый индекс означает номер строки, второй – номер столбца. P – квадратная матрица с неотрицательными элементами, причем сумма элементов каждой строки равна единице. Такие матрицы называют стохастическими. Они вполне описывают марковский случайный процесс.

Возникает вопрос о способе связи элементов матрицы с вероятностями состояний системы. Рассмотрим на конкретном примере последовательность построения матрицы вероятностей перехода. Опишем модель внутривидовой борьбы с незатухающими колебаниями: размер популяции имеет состояния: наибольшее количество особей, обитающих в данной местности S_1 и наименьшее S_0. Через равные промежутки времени подсчитывается количество особей в популяции. Известны вероятности перехода системы из состояния в состояние. Опишем вероятности состояния S_0: если при проверке оказалось, что в популяции насчитывается минимальное количество особей, то вероятность того, что система перейдет в состояние S_1, равна 0,6, а вероятность, что останется в состоянии S_0 – 0,4. Если система находится в состоянии S_1 – в популяции максимальное число особей, то вероятность, что это состояние перейдет в S_0 равна 0,3, а останется в этом состоянии - 0,7.

Рассмотрим схему работы этой динамической системы, подвергающейся воздействию факторов случайного характера

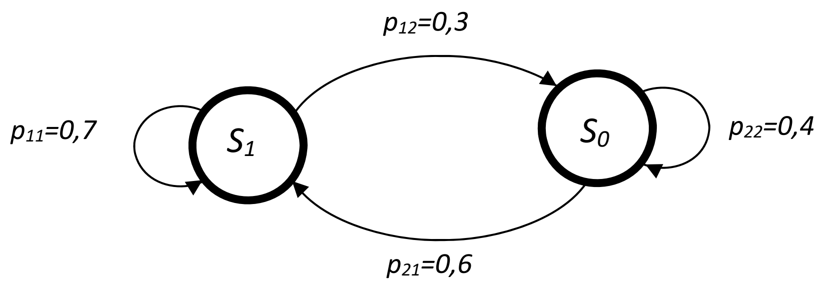


Рисунок 4.2.1 - Граф состояний динамической системы

Такую схему, в которой изображены состояния системы, переходы из состояния в состояние и указаны вероятности переходов будем называть *графом состояний*.

Обратим внимание на тот факт, что каждая вероятность состояния имеет два нижних индекса. В алгебре при записи элемента матрицы принято первым индексом обозначать номер строки, в которой находится этот элемент, а вторым – номер столбца. При записи матрицы вероятностей перехода руководствуемся тем же правилом. Только здесь первый индекс означает, в каком состоянии система находится в момент проверки, а второй индекс – в какое состояние она переходит.

Таким образом, в нашем примере вероятность p_{10} находится над стрелкой, соединяющей состояния S_1 и S_0, и количественно описывает возможность перехода из S_1 в S_0. Матрица вероятностей перехода в общем виде будет выглядеть так

$$\begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix},$$

а конкретно для нашего примера

$$\begin{pmatrix} 0{,}7 & 0{,}3 \\ 0{,}6 & 0{,}4 \end{pmatrix}.$$

В первой строке находятся вероятности перехода из первого состояния, а во второй – из второго. Сумма вероятностей по строкам должна равняться единице.

Приведем классификацию случайных процессов по времени и состоянию.

Таблица 4.2.1 - Классификация случайных процессов по времени и состоянию

по времени	
дискретные	Система, в которой протекает процесс, меняет свои состояния только в моменты $t_1, t_2, \ldots, t_n$, число которых конечно
непрерывные	Переходы системы из состояния в состояние происходят в любой момент времени t в течение наблюдаемого периода
по состоянию	
дискретные	Множество значений случайной величины в сечении конечно (счетное)
непрерывные	Множество значений случайных величин в сечении непрерывно (несчетное)

Пример марковского случайного процесса с дискретным состоянием и дискретным временем.

Известно, что сток рек имеет 4 состояния: первое – самый низкий уровень воды, четвертое – самый высокий, второе и третье – средние между ними. А также известно, что первое и четвертое состояния никогда не следуют по годам друг за другом, а остальные переходы возможны. Переходы из состояния в состояние имеют вероятности:

- из первого состояния снова в первое (за засушливым годом снова следует засушливый) p_{11}=0,2; из первого во второе - p_{12}=0,4; аналогично p_{13}=0,4 и p_{14}=0;

- из второго состояния: p_{21}=0,2; p_{22}=0,4; p_{23}=0,3; p_{24}=0,1;

- из третьего состояния: p_{31}=0,1; p_{32}=0,4; p_{33}=0,4; p_{34}=0,1;

- из четвертого состояния: p_{41}=0; p_{42}=0,4; p_{43}=0,5; p_{44}=0,1.

Если вода в реке достигает самого высокого уровня, то объявляются чрезвычайные меры. Пусть в первый год наблюдался самый низкий уровень воды. Следует ли готовиться к объявлению чрезвычайных мер через два года?

<u>Этап формализации</u>: создадим математическую модель ситуации, описанной в задаче. Определим, что необходимо исследовать марковский случайный процесс с дискретным состоянием и дискретным временем.

<u>Этап внутримодельного решения</u>: выразим взаимосвязи и закономерности наблюдаемого случайного процесса математическим языком: вычерчивая граф состояний системы,

дерево решений, составляя стохастическую матрицу вероятностей перехода из состояния в состояние. Абстрагируясь от природного явления, исследуем свойства и характеристики случайного процесса с помощью математических методов.

Изобразим условие задачи в виде графа (рисунок 4.2.2), в котором обозначим через S_1 – первое состояние, через S_2, S_3 и S_4 – второе, третье и четвертое состояния соответственно. Над каждой стрелкой запишем вероятности перехода из состояния в состояние, осуществляемого по стрелке.

Составим матрицу вероятностей перехода:

$$P = \begin{pmatrix} 0,2 & 0,4 & 0,4 & 0 \\ 0,2 & 0,4 & 0,3 & 0,1 \\ 0,1 & 0,4 & 0,4 & 0,1 \\ 0 & 0,4 & 0,5 & 0,1 \end{pmatrix}.$$

В первой строке запишем вероятности перехода из первой градации в первую (0,2), во вторую (0,4), в третью (0,4) и в четвертую (0). Аналогично записываются вероятности перехода из второго состояния – вторая строка, из третьего – третья строка и из четвертого – четвертая строка.

Для облегчения поиска искомой вероятности построим дерево решений. Опишем принцип его построения. В первой строке записывается исходное состояние, с которого начинается наблюдение – состояние S_1 (в первый наблюдаемый год зафиксирован самый низкий уровень воды); во второй

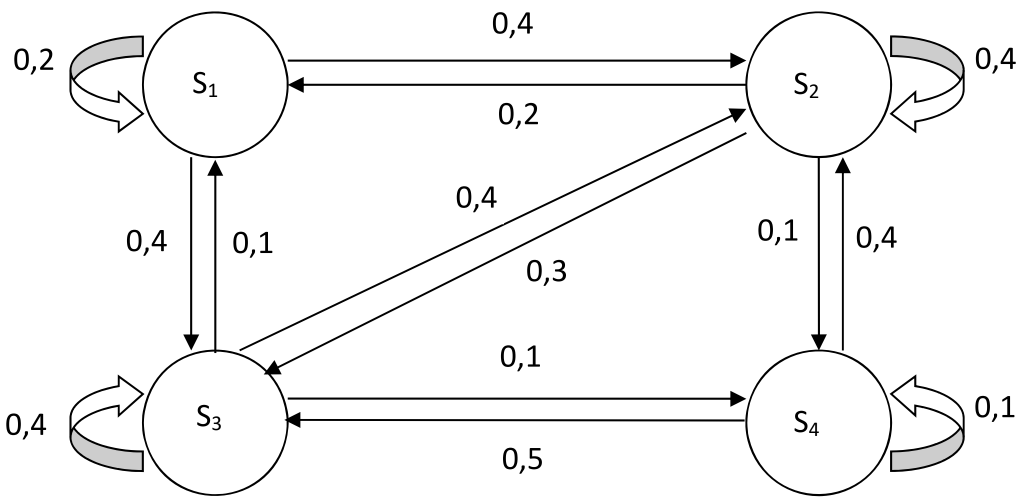


Рисунок 4.2.2 - Граф состояний системы

строке – все возможные состояния во втором году, которые могут наблюдаться после состояния S_1; в третьей строке – все возможные состояния в третьем году.

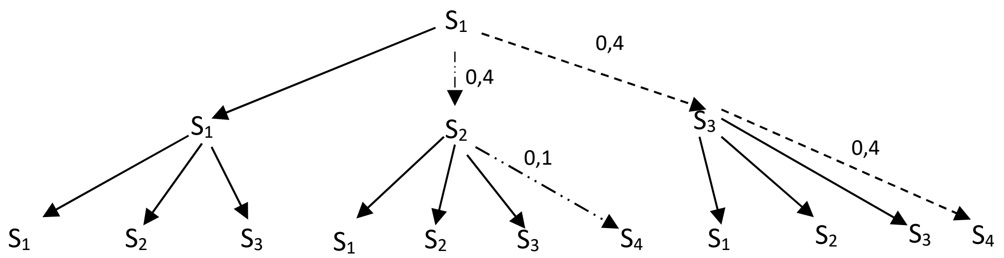


Рисунок 4.2.3 - Дерево решений

Из схемы на рисунке 4.2.3 видно, что из состояния S_1 в состояние S_4 за два года от начала наблюдений можно прийти двумя путями:

1) $S_1 \rightarrow S_2 \rightarrow S_4$;

2) $S_1 \rightarrow S_3 \rightarrow S_4$.

Выделим эти пути разными стрелками и над каждой из них подпишем, чему равна вероятность соответствующего перехода. Теперь по этим стрелкам вычислим условные вероятности состояний.

Условная вероятность того, что в первый год наблюдался самый низкий уровень воды и через два года после этого можно наблюдать уровень воды, равный S_4,

- по первому пути $p(S_2 / S_1)p(S_4 / S_2) = 0{,}4 \cdot 0{,}1 = 0{,}04$;
- по второму пути $p(S_3 / S_1)p(S_4 / S_3) = 0{,}4 \cdot 0{,}1 = 0{,}04$.

Полная вероятность того, что если в начале наблюдения уровень воды находился в состоянии S_1 и через два года после этого он стал равен S_4, равна сумме двух вычисленных условных вероятностей, то есть

$$0{,}04 + 0{,}04 = 0{,}08 .$$

Итак, мы определили, что вероятность наблюдать самый высокий уровень воды через два года после наблюдения самого низкого уровня равна 0,08.

Этап интерпретации: переведем полученные в решении задачи числовые характеристики на язык, применяемый в производственной деятельности. Вывод: так как мала вероятность того, что через два года сток воды в реке будет самым высоким, не стоит готовиться к объявлению чрезвычайных мер.

Переход из S_j в S_k ровно за n шагов может быть осуществлен различными путями

$$S_j \to S_{j_1} \to S_{j_2} \to \ldots \to S_{j_{k-1}} \to S_{j_k} .$$

Условная вероятность того, что переход совершился по некоторому фиксированному пути при условии, что в какой-то момент система находилась в состоянии S_j, равна

$$p\left(S_{j_1} / S_j\right) p\left(S_{j1} / S_{j2}\right) .. \; p\left(S_{j_k} / S_{j\,k-1}\right).$$

Сумма условных вероятностей для всех возможных путей равна вероятности того, что система в момент $r+n$ находится в состоянии S_k при условии, что в момент r она находилась в состоянии S_j.

Будем обозначать эту вероятность $p_{jk}^{(n)}$.

В частности,

$$p_{jk}^{(1)} = p\left(S_k / S_j\right);$$

$$p_{jk}^{(2)} = \sum_{v} p\left(S_v / S_j\right) p\left(S_v / S_k\right).$$

По индукции найдем рекуррентную формулу

$$p_{jk}^{(n+1)} = \sum_{v} p\left(S_v / S_j\right) p^{n}\left(S_v / S_k\right);$$

индукция по m показывает, что

$$p_{jk}^{(m+n)} = \sum_{v} p^{m}\left(S_v / S_j\right) p^{n}\left(S_v / S_k\right). \qquad \textbf{(4.2)}$$

Равенство (4.2) показывает, что за первые m шагов система переходит из S_j в некоторое промежуточное состояние, а за последующие n шагов – из S_v в S_k. Выполнение этого равенства есть характерное свойство цепи Маркова.

4.3. Функционирование систем, приводящее к уравнениям Колмогорова

Пример. Система имеет два канала обслуживания. Пусть состояния этой системы обслуживания такие:

- оба канала свободны (обозначим это состояние S_0),
- первый занят, второй свободен (S_1),
- второй занят, первый свободен (S_2),
- оба заняты (S_3).

Нарисуем граф состояний этой системы:

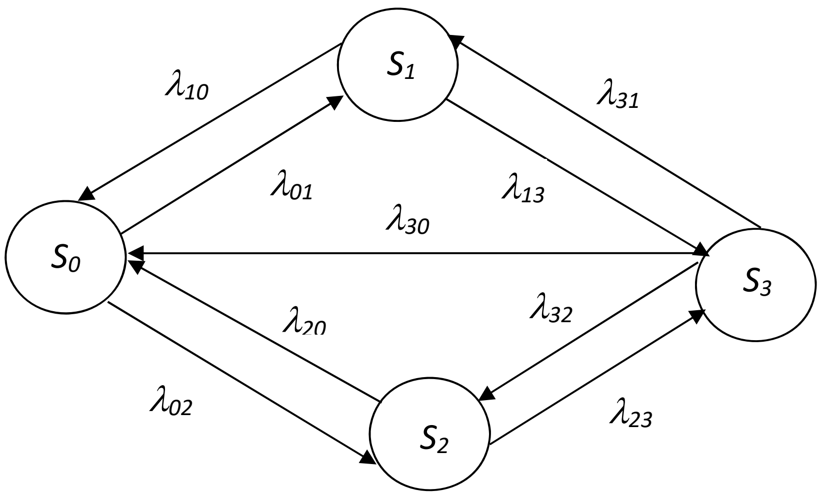


Рисунок 4.3 - Граф состояний СМО

Здесь стрелками обозначены переходы из состояния в состояние. λ_{ij} - интенсивность, с которой система переходит из состояния S_i в S_j. Например, под действием λ_{13} система переходит из состояния S_1 - первый канал занят в состояние S_2.

p_i – вероятность того, что СМО в момент времени t находится в состоянии S_i.

$$\sum_{i=0}^{3} p_i = 1.$$

Пусть в момент времени t СМО находится в состоянии S_0, то есть оба канала свободны. Пусть с момента времени t прошел малый промежуток времени Δt. Найдем вероятность того, что через время Δt СМО снова находится в состоянии S_0. Это произойдет, если:

а) система не вышла из состояния S_0;

б) она находилась в одном из состояний S_1, S_2 или S_3 и вернулась в состояние S_0.

Рассмотрим случай а). Система выйдет из состояния S_0 под действием λ_{01} в состояние S_1 с вероятностью $\lambda_{01}\Delta t$, или под действием λ_{02} в состояние S_2 с вероятностью $\lambda_{02}\Delta t$. Вероятность противоположного события – система не выйдет из состояния S_0 равна

$$1-(\lambda_{01}\Delta t+\lambda_{02}\Delta t)=1-(\lambda_{01}+\lambda_{02})\Delta t.$$

Итак, вероятность того, что система находилась в состоянии S_0 и не вышла из него, по формуле произведения вероятностей равна

$$p_0(t)[1-(\lambda_{01}+\lambda_{02})\Delta t].$$

Рассмотрим случай б). Вероятность того, что система находилась в состоянии S_1, равна $p_1(t)$ и вероятность того, что она из S_1 перейдет в S_0, равна $\lambda_{10}\Delta t$. Т.о., вероятность того, что через промежуток времени Δt система окажется в состоянии S_0, равна

$$p_1(t)\lambda_{10}\Delta t.$$

Аналогично, вероятность того, что система окажется в состоянии S_0 после того, как она была в состоянии S_2, равна

$$p_2(t)\lambda_{20}\Delta t.$$

И, наконец, вероятность того, что система окажется в состоянии S_0 после того, как она была в состоянии S_3, равна

$$p_3(t)\lambda_{30}\Delta t.$$

В совокупности случай а) и случай б) дадут нам искомую вероятность того, что через время Δt СМО снова находится в состоянии S_0. Согласно формуле сложения вероятностей

$$p_0(t+\Delta t)=p_1(t)\lambda_{10}\Delta t+p_2(t)\lambda_{20}\Delta t+p_3(t)\lambda_{30}\Delta t+p_0(t)[1-(\lambda_{01}+\lambda_{02})\Delta t];$$
$$p_0(t+\Delta t)=p_1(t)\lambda_{10}\Delta t+p_2(t)\lambda_{20}\Delta t+p_3(t)\lambda_{30}\Delta t+p_0(t)-p_0(t)(\lambda_{01}+\lambda_{02})\Delta t];$$
$$p_0(t+\Delta t)-p_0(t)=p_1(t)\lambda_{10}\Delta t+p_2(t)\lambda_{20}\Delta t+p_3(t)\lambda_{30}\Delta t-p_0(t)(\lambda_{01}+\lambda_{02})\Delta t].$$

Разделим обе части равенства на Δt

$$\frac{p_0(t+\Delta t)-p_0(t)}{\Delta t}=p_1(t)\lambda_{10}+p_2(t)\lambda_{20}+p_3(t)\lambda_{30}-p_0(t)(\lambda_{01}+\lambda_{02}).$$

Пусть $\Delta t\to 0$, тогда

$$\lim_{\Delta t\to 0}\frac{p_0(t+\Delta t)-p_0(t)}{\Delta t}=p_1(t)\lambda_{10}+p_2(t)\lambda_{20}+p_3(t)\lambda_{30}-p_0(t)(\lambda_{01}+\lambda_{02}).$$

Так как $\lim\limits_{\Delta t \to 0} \dfrac{p_0(t+\Delta t)-p_0(t)}{\Delta t} = p_0'(t)$, то

$$p_0' = p_1\lambda_{10} + p_2\lambda_{20} + p_3\lambda_{30} - p_0(\lambda_{01} + \lambda_{02}).$$

Получено *дифференциальное уравнение Колмогорова* для нахождения вероятности p_0 состояния СМО.

Аналогично получим уравнения для вероятностей того, что СМО находится в состоянии S_1, S_2, S_3.

Получим систему

$$\begin{cases} p_0' = p_1\lambda_{10} + p_2\lambda_{20} + p_3\lambda_{30} - p_0\lambda_{01} - p_0\lambda_{02}; \\ p_1' = p_0\lambda_{01} + p_3\lambda_{31} - p_1\lambda_{13} - p_1\lambda_{10}; \\ p_2' = p_0\lambda_{02} + p_3\lambda_{32} - p_2\lambda_{20} - p_2\lambda_{23}; \\ p_3' = p_1\lambda_{13} + p_2\lambda_{23} - p_3\lambda_{30} - p_3\lambda_{31} - p_3\lambda_{32}. \end{cases} \qquad \mathbf{(4.3.1)}$$

Для нахождения неизвестных вероятностей к системе необходимо добавить нормировочное условие

$$\sum_{i=1}^{3} p_i = 1.$$

Правило составления уравнений Колмогорова: в левой части каждого уравнения стоит производная p_i' вероятности *i*-го состояния, а в правой – столько слагаемых, сколько стрелок связано с этим состоянием. Если стрелка *входит* в состояние, то слагаемое со знаком «+», если стрелка *выходит* из состояния, то знак «–». Каждое слагаемое равно интенсивности потока, переводящей систему по данной стрелке, умноженной на вероятность того состояния, из которого выходит эта стрелка.

Чтобы решить систему (4.3.1), необходимо задать начальные условия. В нашем примере – это вероятности состояния системы в начальный момент времени *t=0*. Пусть при *t=0* оба канала свободны, тогда

$$p_0(0)=1,\ p_1(0)=p_2(0)=p_3(0)=0\,.$$

Когда случайный процесс длится достаточно долгое время, говорят о предельном поведении вероятностей $p_i(t)$ при $t\to\infty$. Если поток событий является пуассоновским с интенсивностями $\lambda_{ij}=const$, то существуют *предельные вероятности состояний*

$$p_i=\lim_{t\to\infty}p_i(t);\ (i=1,\dots,n),$$

независящие от того, в каком состоянии система находилась в начальный момент времени. В этом случае в системе устанавливается *предельный стационарный режим*, при котором система переходит из состояния в состояние с вероятностями, которые не меняются со временем. Физический смысл этих предельных вероятностей таков: это *среднее относительное время* пребывания системы в данном состоянии.

Если предельные вероятности принимают конкретные значения, то есть $p_i = const$, то в системе (4.3.1) $p_i' = 0$. Тогда система (4.3.1) примет вид:

$$\begin{cases} p_0(\lambda_{01} + \lambda_{02}) = p_1\lambda_{10} + p_2\lambda_{20} + p_3\lambda_{30}; \\ p_1(\lambda_{13} + \lambda_{10}) = p_0\lambda_{01} + p_3\lambda_{31}; \\ p_2(\lambda_{20} + \lambda_{23}) = p_0\lambda_{02} + p_3\lambda_{32}; \\ p_3(\lambda_{31} + \lambda_{32}) = p_1\lambda_{13} + p_2\lambda_{23} - p_3\lambda_{30}; \\ \sum_{i=0}^{3} p_i = 1. \end{cases} \qquad \textbf{(4.3.2)}$$

Нетрудно заметить, что в полученной системе *в левой части уравнений находится произведение предельной вероятности состояния на сумму интенсивностей потоков, выходящих из этого состояния, а в правой части – сумма произведений интенсивностей всех потоков, входящих в это состояние на предельные вероятности состояний, из которых идут эти потоки.* Систему уравнений (4.3.2) легко составить непосредственно по размеченному графу состояний.

4.4. Процессы «гибели и размножения»

В биологии с помощью теории процессов гибели и размножения описывается изменение численности популяции. Процесс гибели и размножения – это процесс с непрерывным временем и дискретным состоянием.

Процессы «гибели и размножения» - это марковские процессы с дискретными состояниями и непрерывным временем. Термин происходит от биологических задач, связанных с изучением численности популяций, и изображается размеченным графом состояний:

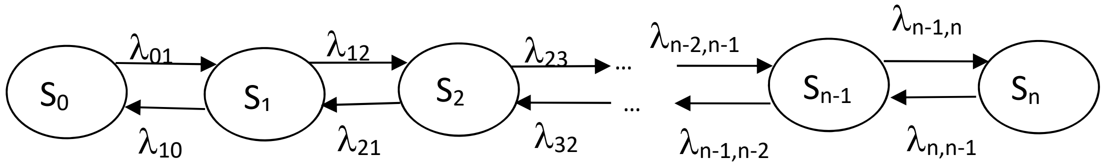


Рисунок 4.4.1 – Граф состояний процесса гибели и размножения

Здесь каждое из состояний, кроме первого и последнего, связано с соседними состояниями прямой и обратной связью.

В соответствии с системой уравнений **(4.3.2)**, имеем систему для предельных вероятностей состояний:

$$\begin{cases} p_0\lambda_{01} = p_1\lambda_{10}; \\ p_1(\lambda_{12} + \lambda_{10}) = p_0\lambda_{01} + p_2\lambda_{21}; \\ p_2(\lambda_{21} + \lambda_{23}) = p_1\lambda_{12} + p_3\lambda_{32}; \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots; \\ p_{n-1}\lambda_{n-1,n} = p_n\lambda_{n,n-1}; \\ \sum_{i=0}^{n} p_i = 1. \end{cases} \quad \textbf{(4.4.1)}$$

Раскроем скобки во втором уравнении полученной системы:

$$p_1\lambda_{12} + p_1\lambda_{10} = p_0\lambda_{01} + p_2\lambda_{21}.$$

Заменим в соответствии с первым уравнением в правой части второго уравнения $p_0\lambda_{01}$ на $p_1\lambda_{10}$, получим:

$$p_1\lambda_{12} + p_1\lambda_{10} = p_1\lambda_{10} + p_2\lambda_{21}.$$

Сократим в обеих частях уравнения слагаемое $p_1\lambda_{10}$:

$$p_1\lambda_{12} = p_2\lambda_{21}.$$

Проведя аналогичные преобразования в остальных уравнениях системы (4.4.1), получим систему

$$\begin{cases} p_0\lambda_{01} = p_1\lambda_{10}; \\ p_1\lambda_{12} = p_2\lambda_{21}; \\ \dots\dots\dots\dots\dots\dots; \\ p_{n-1}\lambda_{n-1,n} = p_n; \\ \sum_{i=0}^{n} p_i = 1. \end{cases} \qquad \mathbf{(4.4.2)}$$

Выразим в системе (4.4.2) все вероятности через p_0:

$$p_1 = \frac{\lambda_{01}}{\lambda_{10}} p_0,\ p_2 = \frac{\lambda_{12}\lambda_{01}}{\lambda_{21}\lambda_{10}} p_0,\ \dots,\ p_n = \frac{\lambda_{n-1,n}\dots\lambda_{12}\lambda_{01}}{\lambda_{n,n-1}\dots\lambda_{21}\lambda_{10}} p_0$$

Легко заметить, что числители коэффициентов при p_0 содержат произведения всех интенсивностей, переводящих из состояния p_0 к данному состоянию слева направо, а знаменатели - произведения всех интенсивностей, переводящих от данного состояния к состоянию p_0 справа налево.

Подставим полученные значения в выражение $\sum_{i=0}^{n} p_i = 1$:

$$p_0 + \frac{\lambda_{01}}{\lambda_{10}} p_0 + \frac{\lambda_{12}\lambda_{01}}{\lambda_{21}\lambda_{10}} p_0 + \dots + \frac{\lambda_{n-1,n}\dots\lambda_{12}\lambda_{01}}{\lambda_{n,n-1}\dots\lambda_{21}\lambda_{10}} p_0 = 1$$

Выразим p_0:

$$p_0 = \frac{1}{1 + \frac{\lambda_{01}}{\lambda_{10}} + \frac{\lambda_{12}\lambda_{01}}{\lambda_{21}\lambda_{10}} + \ldots + \frac{\lambda_{n-1,n}\ldots\lambda_{12}\lambda_{01}}{\lambda_{n,n-1}\ldots\lambda_{21}\lambda_{10}}} \qquad \textbf{(4.4.3)}$$

Т.о., получена система уравнений:

$$\begin{cases} p_0 = \dfrac{1}{1 + \frac{\lambda_{01}}{\lambda_{10}} + \frac{\lambda_{12}\lambda_{01}}{\lambda_{21}\lambda_{10}} + \ldots + \frac{\lambda_{n-1,n}\ldots\lambda_{12}\lambda_{01}}{\lambda_{n,n-1}\ldots\lambda_{21}\lambda_{10}}}; \\ p_1 = \dfrac{\lambda_{01}}{\lambda_{10}} p_0; \\ p_2 = \dfrac{\lambda_{12}\lambda_{01}}{\lambda_{21}\lambda_{10}} p_0; \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots; \\ p_n = \dfrac{\lambda_{n-1,n}\ldots\lambda_{12}\lambda_{01}}{\lambda_{n,n-1}\ldots\lambda_{21}\lambda_{10}} p_0. \end{cases} \qquad \textbf{(4.4.4)}$$

Пример. Процесс гибели и размножения представлен графом (рисунок 4.4.2). Найти предельные вероятности состояний системы.

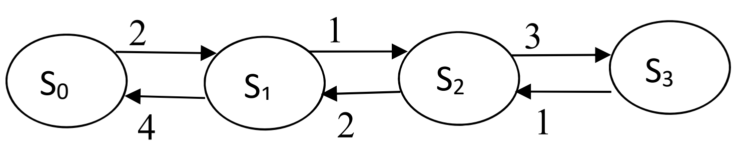


Рисунок 4.4.2 - Граф состояний к примеру

Решение. Согласно формулам (4.4.4),

$p_0 = \left(1 + \frac{2}{4} + \frac{1\cdot 2}{2\cdot 4} + \frac{3\cdot 1\cdot 2}{1\cdot 2\cdot 4}\right)^{-1} = \left(\frac{10}{4}\right)^{-1} = 0{,}4;$

$$p_1 = \frac{2}{4} = 0{,}2;$$

$$p_2 = \frac{1 \cdot 2}{2 \cdot 4} = 0{,}1;$$

$$p_3 = \frac{3 \cdot 1 \cdot 2}{1 \cdot 2 \cdot 4} = 0{,}3.$$

Т. о., система находится в состоянии S_0 40% времени, в S_1 - 20%, в S_2 – 10%, в S_3 – 30%.

Задачи к главе 4

4.1. Через фиксированные промежутки времени проводится фиксация состояния биологической системы, которая может находиться в одном из трех состояний: S_1, S_2, S_3. Задана матрица вероятностей перехода:

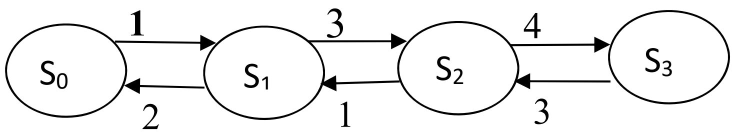
$$P = \begin{pmatrix} 0{,}6 & 0{,}1 & p_{13} \\ 0{,}3 & p_{22} & 0{,}5 \\ 0{,}41 & 0{,}19 & p_{33} \end{pmatrix}.$$

Найти неизвестные элементы матрицы P и составить граф состояний.

4.2. В некоторой местности весной сухая погода сменяется дождливой. Если идет дождь, то с вероятностью 0,6 он будет идти на следующий день и с вероятностью 0,4 будет сухая погода. Если погода сухая, то с вероятностью 0,7 она останется сухой на следующий день и с вероятностью 0,3 пойдет дождь. Начертить граф состояний, составить матрицу перехода.

Вычислить вероятность того, что погода будет дождливая в ближайшую пятницу, если известно, что в среду она была дождливая.

4.3. Граф состояний системы, описываемой процессами гибели и размножения, имеет вид



Вычислить предельные вероятности состояния системы.

Вопросы к главе 4

1. Что такое марковский случайный процесс?
2. Что описывают вероятности перехода из состояния в состояние?
3. Как составляется матрица вероятностей перехода?
4. Как строится граф состояний динамической системы?
5. Как построить дерево решений?
6. Сформулируйте правило составления уравнений Колмогорова.
7. Охарактеризуйте три этапа решения практических задач с помощью математического моделирования.
8. Что такое системы массового обслуживания? Как они применяются для моделирования биологических процессов?
9. Начертите граф состояний процесса гибели и размножения.

ТЕСТ

Вариант 1

1) *Статистической совокупностью называют*

a) множество единиц, обладающих массовостью, однородностью, взаимозависимостью состояния отдельных единиц и наличием вариации;
b) множество изучаемых разнородных объектов;
c) группа зафиксированных случайных событий.

2) *Назовите основные элементы статистических рядов распределения*

a) варианта и частота;
b) генеральная совокупность;
c) кумулята.

3) *В результате исследования биологического процесса получены данные на конец каждого месяца:*

янв	февр	март	апр	май	июнь
105	123	96	115	160	142

Определите среднее значение исследуемого показателя за 1 полугодие.

a) 123,5;
b) 116;
c) 132.

4) *Для графического изображения интервальных вариационных рядов используют*

a) гистограмму;
b) полигон;
c) круговую диаграмму.

5) *Генеральная совокупность – это*

a) совокупность всех объектов, относительно которых исследователь намерен делать выводы при изучении конкретной проблемы;
b) совокупность объектов в двух и более выборках;
c) совокупность объектов, полученная путем выбора по несколько представителей из каждой группы.

6) *Вариационный ряд – это*

a) представленный в порядке возрастания ряд значений изучаемого признака;
b) ряд, построенный по качественному признаку;
c) ряд, полученный путем удаления из данных наблюдений, значений, ниже модального.

7) Объем статистической совокупности – это

a) количество групп, на которые разбита совокупность;
b) численность элементов совокупности, взятых для исследования;
c) объем биологического объекта, занимаемый им в пространстве.

8) Выборочная дисперсия – это

a) мера разброса данных наблюдений от среднего значения*;*
b) разложение света в спектр;
c) оценка среднеквадратического отклонения.

9) Репрезентативная выборка – это

a) выборка, построенная на основе разной вероятности отобранных объектов;

b) выборка, полученная путём случайного отбора и правильно представляющая пропорции генеральной совокупности;

c) выборка, среднеквадратическое отклонение которой равна нулю.

10) Гистограмма – это

a) круг, разделенный на секторы, относительный размер которых пропорционален численным значениям;

b) графическое представление данных наблюдений в виде ломаной линии;

c) ступенчатая фигура из прямоугольников с основаниями, равными интервалам значений и высотами, равными частотам интервалов.

11) Эмпирическая функция распределения выражает

a) относительную частоту того, что признак (случайная величина X) примет значение, меньшее заданного x;

b) функцию, которая определяет для каждого значения выборки ее вероятность;

c) долю каждого наблюдаемого значения в выборке.

12. Эксцесс – это

a) крайние значения в выборке;

b) величина, характеризующая степень островершинности или плосковершинности распределения;

c) величина, характеризующая степень отдаленности значений выборки от среднего.

13) Задано распределение частот выборки объема $n = 20$:

x_i	2	6	12
n_i	3	10	7

Укажите правильное распределение относительных частот.

a)

x_i	*2*	*6*	*12*
n_i	*0,1*	*0,5*	*0,3*

b)

x_i	*2*	*6*	*12*
n_i	*0,10*	*0,50*	*0,30*

c)

x_i	*2*	*6*	*12*
n_i	*0,15*	*0,50*	*0,35*

d)

x_i	*2*	*6*	*12*
n_i	*0,50*	*0,15*	*0,30*

14) Выборка задана таблицей распределения

x_i	*1*	*3*	*4*	*5*	*7*
n_i	*2*	*1*	*3*	*2*	*2*

Найдите выборочную дисперсию.

a) $D = 4{,}25$;
b) $D = 4{,}10$;
c) $D = 3{,}89$;
d) $D = 1{,}08$.

15) Что такое математическая модель временного ряда

a) математическая модель, при построении которой используются данные, характеризующие один объект за ряд последовательных моментов времени;
b) математическая модель, при построении которой используются данные, изучающие несколько объектов за определенный момент времени;
c) математическая модель, описывающая повторяемость данных через небольшой промежуток времени.

Вариант 2

1) Варианты – это

a) различные значения признака (случайной величины *X*);
b) изменяющееся значение наблюдаемого признака;
c) числа, показывающие, сколько раз в наблюдении встречается данная величина.

2) Полигоном частот называется
a) прямая линия на координатной плоскости;
b) ломаная линия, отрезки которой соединяют точки ($x_i;n_i$);
c) ломаная линия, отрезки которой соединяют точки ($x_i;y_i$);
d) прямая линия, отрезки которой соединяют точки ($x_i;n_i$).

3) Наблюдаемые значения рассматриваемого признака называются

a) отношениями;
b) группировкой;
c) частотами;
d) вариантами.

4) Совокупность задана таблицей распределения

x_i	2	4	5	6
n_i	8	9	10	3

Найти дисперсию.
a) 4,8;
b) 2,8;
c) 3,8;
d) 1,8.

5) Статистическим распределением выборки называют

a) перечень частот или относительных частот;

b) варианты и частоты или относительные частоты;
c) перечень вариант и соответствующих им частот или относительных частот;
d) перечень вариант относительных частот.

6) Ступенчатую фигуру, состоящая из прямоугольников, основаниями которых служат частичные интервалы длиною h, а высоты равны отношению n_i / h (плотность частоты), называют

a) графиком функции;
b) полигоном частот;
c) галлаграммой частот;
d) гистограммой частот.

7) Оценку, которая определяется одним числом, называют

a) координатной;
b) точечной;
c) интервальной;
d) надежной.

8) Дана выборка объема N. Если каждый элемент выборки увеличить в 5 раз, то выборочное среднее $\bar{x}$ ...

a) увеличится в 5 раз;
b) увеличится в 25 раз;
c) не изменится;
d) уменьшится в 5 раз.

9) Среднее арифметическое значений признака генеральной совокупности называют

a) выборочной средней;
b) генеральной оценкой;
c) выборочной оценкой;

d) генеральной средней.

10) Асимметрия – это

a) мера «скошенности» наблюдений около моды;
b) показатель, характеризующий скученность данных наблюдений около медианы;
c) величина, характеризующая степень отдаленности значений выборки от значений теоретической функции распределения.

11) Относительная частота события А определяется формулой W(A) = m/n, где

a) *m* – число элементарных исходов, благоприятствующих А, *n* – число всех возможных элементарных исходов испытания;
b) m – число элементарных исходов, благоприятствующих А, *n* – число появлений события;
c) *m* – общее число испытаний, *n* – число всех возможных элементарных исходов испытания;
d) *m* – число появлений события, *n* – общее число испытаний.

12) Выборка задана таблицей распределения

x_i	2	4	5	6
n_i	8	9	10	3

Найдите выборочную дисперсию.
a) $D = 0{,}8$;
b) $D = 1{,}8$;
c) $D = 0{,}08$;
d) $D = 1{,}08$.

13) Дана выборка объема N. Если каждый элемент выборки уменьшить в 5 раз, то выборочное среднее $\bar{x}$

a) увеличится в 5 раз;

b) увеличится в 25 раз;
c) не изменится;
d) уменьшится в 5 раз.

14) Модой называется
a) среднее значение в выборке;
b) значение показателя, который имеет наибольшую частоту;
c) значение показателя, который находится в середине вариационного ряда.

15) Медиана – это
a) варианта, которая находится в середине упорядоченного вариационного ряда;
b) величина, характеризующая общее количество наблюдений в вариационном ряде;
c) значение признака, который чаще других встречается в наблюдении.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Гильдерман, Ю.И.* Математизация биологии / Ю.И. Гильдерман. – М.: Знание, 1969. – 50 с.

2. *Грейг-Смит, П.* Количественная экология растений / П. Грейг-Смит . – М.: Мир, 1967. - 359 с.

3. *Гроссман, С., Тернер, Дж.* Математика для биологов / С. Гроссман, Дж. Тернер. – М.: Высшая школа, 1983. – 383 с.

4. *Елисеева, И.И.; Юзбашев, М.М.* Общая теория статистики: учебник / Под ред. чл. корр. РАН И.И. Елисеевой. – М.: Финансы и статистика, 2002. – 480 с: ил.

5. *Зайцев, Г.Н.* Математика в экспериментальной ботанике / Г.Н. Зайцев. – М.: Наука, 1990. – 296 с.

6. *Ивантер, Э. В., Коросов, А. В.* Элементарная биометрия: учеб. пособие / Э. В. Ивантер, А. В. Коросов. — Петрозаводск: Изд-во ПетрГУ, 2010. — 104 с.

7. Имитационное моделирование URL: http://www.center-yf.ru (дата обращения: 2.05.2014 г.).

8. *Калинина, В.Н.; Панкин, В.Ф.* Математическая статистика / В.Н. Калинина, В.Ф. Панкин. - М.: Дрофа, 2002. - 246 с.

9. *Красс, М.С.; Чупрынов, Б.П.* Математика для экономистов / М.С. Красс, Б.П. Чупрынов. - СПб.: Питер, 2004. – 464 с.: ил.

10. *Малков, П. Ю.* Количественный анализ биологических данных: Учебное пособие / П.Ю. Малков. - Горно-Алтайск: РИО ГАГУ, 2009. 71 с.

11. *Мамадалиева, Л.Н.* Ряды динамики: учеб. - методич. пособие / Л.Н. Мамадалиева. – Майкоп: Магарин О.Г., 2011. – 41 с.

12. *Плохинский, Н.А.* Алгоритмы биологии / Н.А. Плохинский. – М.: Изд-во Московского университета, 1980. – 150 с.

13. *Просветов, Г.И.* Математические методы и модели в экономике: задачи и решения: учеб.- практич. пособие. / Г.И. Просветов. – М.: Альфа-Пресс, 2008. – 344 с.

14. *Разговоров, А.В.* Сборник задач и упражнений по статистике связи / А.В. Разговоров. - М.: Государственное издательство литературы по вопросам связи и радио, 1963. – 160 с.

15. Регрессионные модели в биофизических исследованиях: учеб. пособие для вузов / Е.А. Калаева, В.Г. Артюхов. – Воронеж: Издательско-полиграфический центр Воронежского государственного университета, 2007. – 36 с.

16. *Урбах, В.Ю*. Биометрические методы / В.Ю. Урбах. – М.: Наука, 1964. – 415 с.

17. *Шмидт, В.М.* Математические методы в ботанике: учеб. Пособие / В.М. Шмидт. – Л.: Изд. Ленинградского университета, 1984. – 288 с.

Приложение 1

Распределение Фишера (F-критерий)

α=0,05		**число степеней свободы v_1**																		
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**15**	**20**	**24**	**30**	**40**	**60**	**120**
число степеней свободы v_2	**1**	161	200	216	225	230	234	237	239	271	242	243	244	246	248	249	250	251	252	253
	2	18,5	19,0	19,2	19,2	19,3	19,3	19,4	19,4	19,4	19,4	19,4	19,4	19,4	19,4	19,5	19,5	19,5	19,5	19,5
	3	10,1	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,76	8,74	8,70	8,66	8,64	8,62	8,59	8,57	8,55
	4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,94	5,91	5,86	5,80	5,77	5,75	5,72	5,69	5,66
	5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,71	4,68	4,62	4,56	4,53	4,50	4,46	4,43	4,40
	6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,03	4,00	3,94	3,87	3,84	3,81	3,77	3,74	3,70
	7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,60	3,57	3,51	3,44	3,41	3,38	3,34	3,30	3,27
	8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,31	3,28	3,22	3,15	3,12	3,08	3,04	3,01	2,97
	9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,10	3,07	3,01	2,94	2,90	2,86	2,83	2,79	2,75
	10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,94	2,91	2,85	2,77	2,74	2,70	2,66	2,62	2,58
	11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,82	2,79	2,72	2,65	2,61	2,57	2,53	2,49	2,45
	12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75	2,72	2,69	2,62	2,54	2,51	2,47	2,43	2,38	2,34
	13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,63	2,60	2,53	2,46	2,42	2,38	2,34	2,30	2,25
	14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,57	2,53	2,46	2,39	2,35	2,31	2,27	2,22	2,18
	15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,51	2,48	3,52	3,37	3,29	3,21	3,13	3,05	2,96
	16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,46	2,42	2,35	2,28	2,24	2,19	2,15	2,11	2,06
	17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45	2,41	2,38	2,31	2,23	2,19	2,15	2,10	2,06	2,01
	18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,37	2,34	2,27	2,19	2,15	2,11	2,06	2,02	1,97
	19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38	2,34	2,31	2,23	2,16	2,11	2,07	2,03	1,98	1,93
	20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,31	2,28	2,20	2,12	2,08	2,04	1,99	1,95	1,90
	22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30	2,26	2,23	2,15	2,07	2,03	1,98	1,94	1,89	1,84
	24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25	2,21	2,18	2,11	2,03	1,98	1,94	1,89	1,84	1,79
	26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22	2,18	2,15	2,07	1,99	1,95	1,90	1,85	1,80	1,75
	28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24	2,19	2,15	2,12	2,04	1,96	1,91	1,87	1,82	1,77	1,71
	30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	2,13	2,09	2,01	1,93	1,89	1,84	1,79	1,74	1,68
	40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08	2,04	2,00	1,92	1,84	1,79	1,74	1,69	1,64	1,58
	60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99	1,95	1,92	1,84	1,75	1,70	1,65	1,59	1,53	1,47
	120	3,92	3,07	2,68	2,45	2,29	2,17	2,09	2,02	1,96	1,91	1,87	1,83	1,75	1,66	1,61	1,55	1,50	1,43	1,35

Приложение 2

Распределение Стьюдента (*t*-распределение)

Число степеней свободы *f* = *n* - *1*	**n**	**Доверительная вероятность**			
		0,90	**0,95**	**0,99**	**0,999**
1	2	6,3137515148	12,7062047364	63,6567411629	636,619249432
2	3	2,91998558036	4,30265272991	9,92484320092	31,599054577
3	4	2,3533634348	3,18244630528	5,84090929976	12,9239786366
4	5	2,13184678134	2,7764451052	4,60409487142	8,61030158138
5	6	2,01504837267	2,57058183661	4,03214298356	6,86882663987
6	7	1,94318028039	2,44691184879	3,70742802132	5,95881617993
7	8	1,89457860506	2,36462425101	3,49948329735	5,40788252098
8	9	1,85954803752	2,30600413503	3,35538733133	5,04130543339
9	10	1,83311293265	2,26215716274	3,24983554402	4,78091258593
10	11	1,81246112281	2,22813885196	3,16927266718	4,5868938587
11	12	1,7958848187	2,20098516008	3,10580651322	4,43697933823
12	13	1,78228755565	2,17881282966	3,05453958834	4,31779128361
13	14	1,77093339599	2,16036865646	3,01227583821	4,22083172771
14	15	1,76131013577	2,14478668792	2,97684273411	4,14045411274
15	16	1,75305035569	2,13144954556	2,94671288334	4,0727651959
16	17	1,74588367628	2,11990529922	2,92078162235	4,0149963326
17	18	1,73960672608	2,10981557783	2,89823051963	3,96512626361
18	19	1,73406360662	2,10092204024	2,87844047271	3,92164582001
19	20	1,72913281152	2,09302405441	2,86093460645	3,88340584948
20	21	1,72471824292	2,08596344727	2,84533970978	3,84951627298
21	22	1,72074290281	2,07961384473	2,83135955802	3,81927716303
22	23	1,71714437438	2,0738730679	2,8187560606	3,79213067089
23	24	1,71387152775	2,06865761042	2,80733568377	3,76762680377
24	25	1,71088207991	2,06389856163	2,79693950477	3,74539861893
25	26	1,70814076125	2,05953855275	2,78743581368	3,72514394948
26	27	1,70561791976	2,05552943864	2,77871453333	3,70661174331
27	28	1,70328844572	2,05183051648	2,77068295712	3,68959171334
28	29	1,70113093427	2,0484071418	2,76326245546	3,67390640062
29	30	1,69912702653	2,04522964213	2,75638590367	3,6594050194
30	31	1,69726089436	2,0422724563	2,74999565357	3,645958635
40	41	1,68385101139	2,021075383	2,70445926743	3,55096576086
60	61	1,67064886465	2,00029782106	2,66028303115	3,4602004692
120	121	1,65765089935	1,97993040505	2,61742114477	3,37345376507

Приложение 3

Таблица критических уровней RS-критерия

Количество наблюдений	Граница RS-критерия	
	нижняя	верхняя
10	2,67	3,69
15	2,96	4,14
20	3,18	4,49
25	3,34	4,71
30	3,47	4,89
…	…	…

Приложение 4

Критические значения статистики Дарбина-Уотсона при 5%-ном уровне значимости, *k* – число параметров модели

n		**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
***k*=1**	$d_н$	0,61	0,70	0,76	0,82	0,88	0,93	0,97	1,01	1,05	1,08	1,11	1,13	1,16	1,18	1,20
	$d_в$	1,40	1,36	1,33	1,32	1,32	1,32	1,33	1,34	1,33	1,36	1,37	1,38	1,39	1,40	1,41
***k*=2**	$d_н$		0,47	0,36	0,63	0,70	0,66	0,81	0,86	0,91	0,95	0,98	1,01	1,05	1,07	1,10
	$d_в$		1,90	1,78	1,70	1,64	1,60	1,58	1,56	1,55	1,54	1,53	1,54	1,54	1,54	1,54

Приложение 5

Таблица значений $\overline{W}$, σ_1 и σ_2
для метода Фостера-Стьюарта

n	$\overline{W}$	σ_1	σ_2
10	3,86	1,29	1,96
15	3,64	1,52	2,15
20	5,19	1,68	2,28
25	5,63	1,79	2,37
30	5,99	1,88	2,45
35	6,29	1,96	2,51
40	6,56	2,02	2,56
45	6,79	2,07	2,61
50	7,00	2,12	2,64
55	7,19	2,16	2,68

ГЛОССАРИЙ

Абсолютный прирост Δ_i показывает, на сколько данный уровень ряда изменился по сравнению с предыдущим или базовым.

Базисные показатели ряда динамики характеризуют окончательный результат всех изменений в уровнях ряда от периода, к которому относится базисный уровень, до данного периода.

Вариационный ряд – ряд, представленный в порядке возрастания значений изучаемого признака.

Генеральная совокупность - это множество однородных, но индивидуально различимых объектов.

Гистограмма - ступенчатая фигура из прямоугольников с основаниями, равными интервалам значений и высотами, равными частотам интервалов (абсолютной, относительной или накопленной).

Дискретные случайные процессы, классифицируемые по времени - процессы, для которых в система меняет свои состояния только в моменты $t_1, t_2, \ldots, t_n$, число которых конечно.

Дискретные случайные процессы, классифицируемые по состоянию – процессы, для которых множество значений случайной величины в сечении конечно (счетное).

Интервальный динамический ряд характеризует величину наблюдаемого явления за какой-то промежуток (период, интервал) времени.

Канал обслуживания - *о*бъект, занимающийся обслуживанием - заявок.

Коэффициент корреляции показывает тесноту связи наблюдаемых значений и значений функции, вычисленной по выведенной формуле.

Коэффициент опережения вычисляется как отношения темпов роста или темпов прироста. Используют при изучении двух взаимосвязанных явлений.

Коэффициент сходимости показывает долю изменения y, объясняемую изменением включенных в модель факторов.

Кумулята — линия накопленных частот, которая есть эмпирический прообраз графика функции распределения y=F(x).

Моментный динамический ряд характеризует величину наблюдаемого явления на определенный момент времени.

Непрерывные случайные процессы, классифицируемые по времени - процессы, для которых переходы системы из состояния в состояние происходят в любой момент времени tв течение наблюдаемого периода

Непрерывные случайные процессы, классифицируемые по состоянию - процессы, для которых множество значений случайных величин в сечении непрерывно (несчетное).

Полигон служит для изображения дискретного вариационного ряда и представляет собой ломаную, в которой концы отрезков прямой имеют координаты (x_i, y_i), где– значение признака, y_i - абсолютная, относительная или накопленная частота. Если исследуется интервальный вариационный ряд, то x_i – середина i-го интервала.

Предельные вероятности - это среднее относительное время пребывания системы в данном состоянии.

Процессы «гибели и размножения» - это марковские процессы с дискретными состояниями и непрерывным временем.

Репрезентативная выборка – выборка, полученная путём случайного отбора и правильно представляет пропорции генеральной совокупности.

Системы массового обслуживания – это системы, предназначенные для обслуживания какого-либо потока заявок.

Темпы прироста показывают, на сколько процентов один уровень больше или меньше другого уровня.

Темпы роста характеризуют скорость изменения показателя в единицу времени, выраженную в процентах.

Ускорение – разность между абсолютным изменением за данный период и абсолютным изменением за предыдущий период. Рассчитывается только для цепных показателей.

Цепные показатели ряда динамики характеризуют скорость роста уровня от периода к периоду.

Эмпирическая функция распределения - относительная частота того, что признак (случайная величина X) примет значение, меньшее заданного x.

ТЕМЫ РЕФЕРАТОВ

1. Математические модели биологических систем.

2. Математик А.Н. Колмогоров. Его вклад в теорию массового обслуживания.

3. Прогнозирование биологических процессов с помощью математических методов.

4. Марковские случайные процессы. Их применение в биологии.

5. Процессы «гибели и размножения».

6. Инструментальные средства автоматизации моделирования.

7. Ошибки в экологических исследованиях.

8. Определение шансов. Использование вероятностей в биологии.

9. Разбираемся в результатах исследования.

10. Доверительные интервалы: их функции, формулы, факторы влияния, толкование.

11. Статистические исследования: взгляд изнутри.

12. Десять критериев для проведения хорошего эксперимента и десять распространенных статистических ошибок.

Мамадалиева Людмила Николаевна
Хаконова Ирина Магомедовна

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ БИОЛОГИЧЕСКИХ ПРОЦЕССОВ

Учебное пособие

Подписано в печать 15.03.19. Формат бумаги 60х84/16. Бумага офсетная.
Печать цифровая. Гарнитура Таймс. Усл. п.л. 9,2. Тираж 300. Заказ 004.

Отпечатано с готового оригинал-макета
на участке оперативной полиграфии.
ИП Кучеренко В.О. 385008, г. Майкоп, ул. Пионерская, 403/33.
Тел. для справок 8-928-470-36-87. E-mail: slv01.maykop.ru@gmail.com