

МИНИСТЕРСТВО СЕЛЬСКОГО ХОЗЯЙСТВА
РОССИЙСКОЙ ФЕДЕРАЦИИ

ФГБОУ ВПО «КУБАНСКИЙ ГОСУДАРСТВЕННЫЙ
АГРАРНЫЙ УНИВЕРСИТЕТ»

Факультет экологии
Кафедра прикладной экологии

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ В ЭКОЛОГИИ

Курс лекций

по направлению подготовки аспирантов 05.06.01 – Науки о Земле

Краснодар
КубГАУ
2015

Составители: Горковенко Н.Е.

Математическое моделирование в экологии: курс лекций / сост. Н.Е. Горковенко – Краснодар : КубГАУ, 2015. – 45 с.

Курс лекций предназначен для аспирантов по направлению подготовки 05.06.01 – Науки о Земле

Рассмотрено и одобрено методической комиссией факультета экологии 29.06.2015г., протокол № 10.

© Горковенко Н.Е., 2015
© ФГБОУ ВПО «Кубанский
государственный аграрный
университет», 2015

Лекция № 1. История развития математической экологии.

Экология – развивающаяся междисциплинарная область знаний, включающая представления практически всех наук о взаимодействиях живых организмов, включая человека, с окружающей средой. До середины 20 века экология представляла собой одну из биологических дисциплин, а именно, науку о взаимодействии организмов с окружающей средой. Современная экология наряду с этим включает в себя науку и практические методы контроля за состоянием окружающей среды - мониторинг, охрану окружающей среды, учение о биогеоценозах и аторопологических воздействиях на природные экосистемы, эколого-экономические и эколого-социальные аспекты. Все это определяет и предмет математической экологии, объединяющей математически модели и методы, используемые при решении проблем экологии.

Фундаментом математической экологии является математическая теория динамики популяций (См. Статью "Популяций динамика"), в которой фундаментальные биологические представления о динамике численности видов животных, растений, микроорганизмов и их взаимодействии формализованы в виде математических структур, в первую очередь, систем дифференциальных, интегро-дифференциальных и разностных уравнений.

Любая экосистема состоит из нелинейно взаимодействующих подсистем, которые можно упорядочить в некоторую иерархическую структуру. По мере объединения компонентов, или подмножеств, в более крупные функциональные единицы, у этих новых единиц возникают свойства, отсутствующие у составляющих ее компонентов. Такие качественно новые "эмерджентные" свойства экологического уровня или экологической единицы не являются простой суммой свойств компонентов. Следствием является невозможность изучения динамики сложных экосистем путем их иерархического расчленения на подсистемы и последующего изолированного изучения этих подсистем, поскольку при этом неизбежно утрачиваются свойства, определяемые целостностью изучаемой системы.

Воздействие внешних факторов на экологическую систему также нельзя рассматривать независимо друг от друга, так как комбинированное действие нельзя свести к сумме действующих факторов. Тем более сложной задачей является количественное описание реакции сложной системы на комплексное воздействие различных факторов.

Все эти обстоятельства приводят к невозможности описать сложные экосистемы с помощью простых редуцированных "механизменных" моделей. Необходимы либо сложные имитационные модели, объединяющие в одну сложную систему на модельном уровне знания об элементах системы и типах их взаимодействия, либо упрощенные интегрированные модели типа "воздействие - отклик", интегрирующие данные большого числа наблюдений над экосистемой.

Имитационные компьютерные модели включают представления о компонентах систем и их взаимосвязях как в виде собственно математических объектов: формул, уравнений, матриц, логических процедур, так и в виде графиков, таблиц, баз данных, оперативной информации экологического мониторинга. Такие многомерные модели позволяют объединить разнородную информацию об экологической или эколого-экономической системе, "проигрывать" различные сценарии развития и вырабатывать на модели оптимальные стратегии управления, что невозможно делать на реальной системе в силу ее уникальности и ограниченности времени.

Имитационный подход, также как и моделирование экосистем с помощью функций отклик, требуют высоко развитой вычислительной техники, поэтому математическая экология как развитая и практически используемая наука получила распространение только в последние десятилетия 20 века. Широкое применение математического аппарата стимулировало развитие теоретической экологии. Построение математических моделей требует упорядочивания и классификации имеющейся информации об экосистемах, приводит к необходимости планировать систему сбора данных и позволяет объединить на содержательном уровне совокупность физических, химических и биологических сведений и представлений об отдельных происходящих в экосистемах процессах.

Общесистемный подход к моделированию экологических систем

При построении моделей экосистем применяют методы общесистемного анализа. В первую очередь это - выделение из системы отдельных структурных элементов, таких как живые и косные компоненты, среди живых - трофические уровни, виды, возрастные или половые группы, взаимодействие которых и будет определять поведение всей системы. Другой важный элемент - установление характера процессов, в которых участвует каждый элемент (процессы размножения и роста, взаимодействия

типа хищничества, конкуренции и т.д.) Часто в экологическом моделировании используются балансовые компартментальные модели, когда рассматриваются потоки вещества и энергии между составляющими модель компартментами, содержание "вещества" в каждом из которых и представляет собой отдельную переменную системы.

Необходимость описывать экологические взаимодействия послужила толчком для развития системных исследований. По словам одного из инователей общей теории систем Людвиг фон Бергаланфи "работы Вольтерра, Лотки, Гаузе и других по теории популяций принадлежат к классическим трудам общей теории систем. В них впервые была продемонстрирована возможность развития концептуальных моделей для таких явлений как борьба за существование, которые могут быть подвергнуты эмпирической проверке." (Л. Бергаланфи, Общая теория систем. Критический обзор, 1969)

Широко используется принцип изоморфизма, позволяющий сходными математическими уравнениями описывать системы, разные по своей природе, но одинаковые по структуре и типу взаимодействия между элементами, их составляющими.

Работа с имитационной моделью требует знания величин параметров модели, которые могут быть оценены только из наблюдения и эксперимента. Часто приходится разрабатывать новые методики наблюдений и экспериментов с целью установления факторов и взаимосвязей, знание которых позволяет выявить слабые места гипотез и допущений, положенных в основу модели. Весь процесс моделирования – от построения моделей до проверки предсказанных с ее помощью явлений и внедрения полученных результатов в практику – должен быть связан с тщательно отработанной стратегией исследования и строгой проверкой используемых в анализе данных.

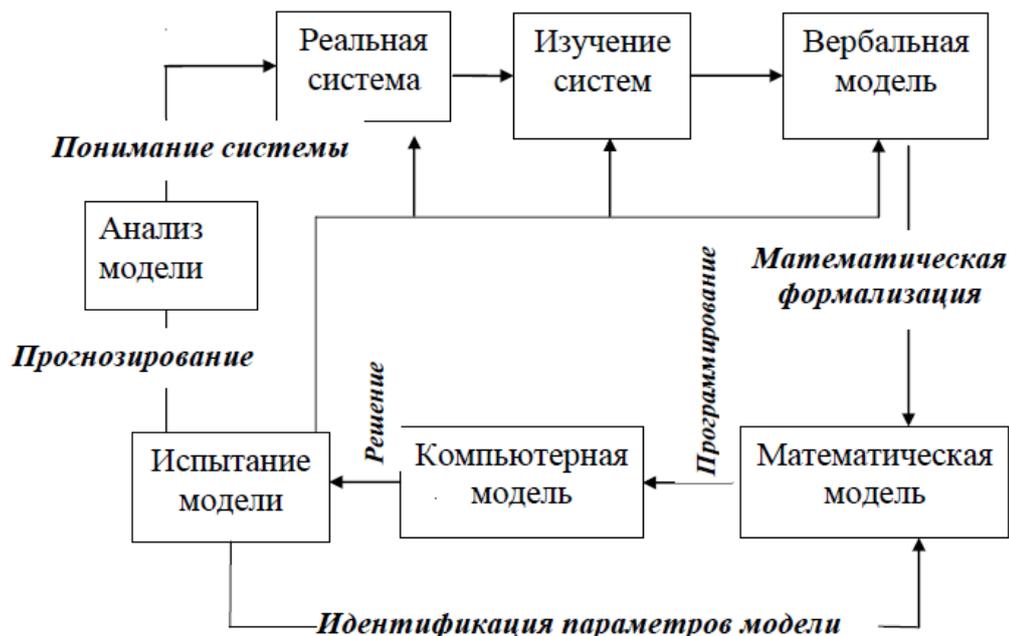
Это положение, справедливое для математического моделирования вообще, особенно важно для такой сложной науки как экология, имеющей дело с разнообразными взаимодействиями между огромным множеством организмов и средой их обитания. Почти все эти взаимодействия динамические в том смысле, что они зависят от времени и постоянно меняются, причем как правило включают в себя положительные и отрицательные обратные связи, то есть являются нелинейными. Сложность экосистем усугубляется с изменчивостью самих живых организмов, которая может проявляться и при взаимодействии организмов друг с другом (например, в процессе конкуренции или хищничества), и в реакции организма на изменения окружающей среды. Эта реакция может выражаться

в изменении скорости роста и воспроизведения и в различной способности к выживанию в сильно различающихся условиях. К этому добавляются происходящие независимо изменения таких факторов среды как климат и характер мест обитания. Поэтому исследование и регулирование экологических процессов представляет собой исключительно сложную задачу.

Экспериментальное и натурное наблюдение экологических процессов осложняется их длительностью. Например, исследования в области земледелия и садоводства связаны главным образом с определением урожайности, а урожай собирают раз в год, так что один цикл эксперимента занимает год и более. Чтобы найти оптимальное количество удобрений и провести другие возможные мероприятия по окультуриванию, может понадобиться несколько лет, особенно когда необходимо рассматривать взаимодействия между экспериментальными результатами и погодой. То же касается процессов, проходящих в аквакультуре, например, при разработке оптимальных режимов содержания рыбоводных прудов. В лесоводстве из-за длительности круговорота урожаев древесины самый непродолжительный эксперимент занимает 25 лет, а долговременные эксперименты могут длиться от 40 до 120 лет. Аналогичные временные масштабы необходимы для проведения исследований с другими природными ресурсами. Поэтому математическое моделирование является необходимым инструментом в экологии, природопользовании и управлении природными ресурсами.

Основные этапы математического моделирования.

Математическое моделирование систем начинается с выбора реальной системы. К реальным системам в экологии относятся – водоем, лесная экосистема, воздушная среда города, экономика города и т.п. Выбор системы для моделирования зависит от множества причин – объективных и субъективных. Решения не всех экологических проблем нуждаются в математическом моделировании. Большое число природоохранных задач может быть решено без привлечения математики, лишь на основе очевидных практических действий в различных экосистемах, промышленности, городском хозяйстве и т.п. В то же время существует большое число важных экологических проблем, которые не могут быть решены без предварительного математического моделирования. В немалой степени постановка задачи математического моделирования зависит от уровня развития экономики страны и уровня отношения общества к экологическим проблемам.



Основные этапы математического моделирования

После того, как поставлена цель – моделирование той или иной реальной системы, первым этапом становится изучение системы. Он включает в себя сбор предварительной информации о моделируемой системе: результаты предыдущих исследований (литературные данные, данные от заказчика и т.п.), постановка собственных экспериментов. Выполнение первого этапа приводит к созданию вербальной модели – словесной модели исследуемой системы (описательный отчет, описательная научная статья). Вербальная модель может давать достаточно полное представление о системе. Многие исследования предметных специалистов – биологов, экологов, химиков ограничиваются созданием вербальной модели системы. Но любая словесная модель даже при очень большом объеме важной и полезной информации имеет существенное ограничение – она не позволяет прогнозировать динамику системы и корректно выявлять управляющие воздействия на систему с целью оптимизации ее функционирования.

Поэтому для современных наук, в том числе и экологии, следующим важным этапом становится создание математической модели системы. Оно начинается с математической формализации. Это – представление в виде математических переменных количественных характеристик элементов системы (численность популяции, концентрация загрязнений, скорость жидкости в водоеме, количество продукции и т.п.). Наряду с переменными определяются параметры, характеризующие интенсивность различных

экологических, биологических, химических и других процессов в экосистеме (коэффициент рождаемости, коэффициент передачи инфекции, константа скорости химической реакции и т.п.). Параметры могут быть константами, а также функциями времени, пространственных переменных и переменных системы. Математическую формализацию и создание математической модели можно определить как два шага – анализ и синтез. Анализ систем: разложение исследуемой системы на подсистемы и элементы, выделение связей между элементами и процессов в системе. Синтез: формулировка математических уравнений на основе выражения связей переменных системы из законов сохранения и гипотез. Результатом синтеза становится математическая модель.

Математическая модель – уравнение или система уравнений на основе выражения связей переменных через законы сохранения (балансовые соотношения) или гипотез (предположения о функционировании элементов системы).

Математическая модель – уравнение или система уравнений на основе выражения связей переменных через законы сохранения (балансовые соотношения) или гипотез (предположения о функционировании элементов системы).

Математические модели в экологии характеризуются комбинацией уравнений, выражающих физические законы о взаимодействии элементов в системе, и математических гипотез о характере зависимости динамики экологических переменных от различных процессов. Так, например, математическая модель процессов в водоеме включает в себя систему уравнений гидродинамики для описания движения жидкой среды, уравнения конвективной диффузии с источниковыми членами, описывающими распространение и физико–химическую трансформацию антропогенных загрязнений, а также уравнения для динамики биотических компонент в водоеме. Уравнения гидродинамики представляют собой систему уравнений в частных производных, выражающих законы сохранения массы, импульса и энергии в единице объема водной среды. Уравнения динамики популяций базируются на гипотезах о взаимодействии различных биотических компонент.

Большинство математических моделей реальных экосистем представляет собой систему из нескольких уравнений (например, дифференциальных), решение которой аналитическими методами невозможно. В этом случае следует применять методы вычислительной математики. Поэтому дальше встает задача реализации математической модели, а именно создание компьютерной программы решения уравнений,

описывающих систему. Программирование – реализация численного метода решения системы уравнений, описывающих систему, с помощью какого-либо языка программирования или стандартного математического пакета. Современный высокий уровень развития вычислительной математики характеризуется наличием большого числа стандартных библиотек программ на различных языках программирования, а также интегрированных математических пакетов (Mathematica, MatLab, Maple, MathCad и т.п.). Кроме этого, в настоящее время развиты специальные программы расчета для различных предметных областей. В частности, таковы пакеты решения задач механики жидкости и газа, так называемые CFD (Computational Fluid Dynamics – Вычислительная гидродинамика) пакеты (FLUENT, StarCD, CFX и др.). Использование указанных современных программных возможностей существенно облегчает решение задач математического моделирования, в том числе и в экологии.

После реализации математической модели – создания собственной компьютерной программы или программы в среде стандартного или специализированного пакета, ключевым становится вопрос о достоверности реализованной модели. Наступает этап, который можно назвать испытанием модели, – этап проверки адекватности созданной математической модели.

Проверка правильности развитой модели начинается с оценки правдоподобия результатов, полученных после расчета по модели. Рассчитанные значения переменных системы должны соответствовать условиям физического и математического правдоподобия: численность популяции должна быть положительной величиной, границы изменения переменных должны соответствовать физическим пределам и т.п. Ошибки в модели, приводящие к неправдоподобным результатам, как правило, легко устраняются. Но правдоподобия расчетных результатов, конечно, недостаточно для того, чтобы говорить о достоверности модели. Основным способом проверки математической модели является сравнение с результатами других расчетных работ и с экспериментальными данными. Реализованная математическая модель может быть также протестирована для некоторых частных случаев, когда возможны аналитические решения задачи. Общей рекомендацией здесь может быть пожелание сравнивать полученное решение со всеми данными, с какими возможно, – любое сравнение с положительным результатом усиливает уверенность в достоверности модели, а также понимание модели и самой моделируемой системы.

Нередко, результаты сравнений с экспериментом и с данными других расчетных работ оказываются первоначально отрицательными, наблюдаются количественные или даже качественные расхождения. Для выявления причин

этого следует критически проанализировать все предыдущие этапы моделирования. Созданная модель может не учитывать каких-то существенных процессов. При первичном формировании модели непросто выделить наиболее важные и несущественные процессы. Такой критический анализ как раз и есть одна из целей математического моделирования.

Следующим ключевым моментом является достоверность параметров модели. Уравнения модели могут быть пригодны для описания ряда однотипных процессов (например, уравнения химической кинетики). Применение различных значений параметров для описания конкретных процессов делает математическую модель вполне определенной моделью конкретной системы. И выбор правильных значений параметров модели играет крайне важную роль в адекватности модели реальной системы. Значения параметров модели обычно известны из результатов научных исследований конкретных процессов – взаимодействия популяций, химических реакций, динамики атмосферы и т.п. И эти значения определяются с различной точностью. Кроме того, параметры могут быть функциями переменных системы (температуры, плотности популяции и т.д.), для выражения которых используются аппроксимации, пригодные в определенных пределах изменения. Таким образом, эти параметры могут содержать в себе ошибки, существенные для результатов моделирования.

К расхождению с экспериментальными данными могут привести ошибочные гипотезы о характере взаимодействия популяций и неправильное написание уравнений, описывающих различные процессы. Источником ошибок модели может быть и выбранный численный алгоритм решения уравнений модели. В зависимости от типа уравнений и характера изменения переменных системы могут выбираться различные численные методы: методы конечных разностей, конечных объемов или конечных элементов. Все численные методы базируются на сеточном разбиении расчетной области. Правильное сеточное разбиение является необходимым условием получения устойчивого численного решения.

Итак, все этапы моделирования могут вносить в модель неточности и быть ответственными за достоверность результатов. Поэтому при наличии 12 расхождений в расчете и эксперименте следует критически отнестись к этим этапам формирования модели и проанализировать их снова. Зачастую время, затраченное на доводку математической модели, может превышать время создания первоначальной модели.

Лекция № 2. Модели популяционной динамики.

Современные математические модели в экологии можно разбить на три класса. Первый – описательные модели: регрессионные и другие эмпирически установленные количественные зависимости, не претендующие на раскрытие механизма описываемого процесса. Примеры таких моделей приведены в (Биология математическая). Они применяются обычно для описания отдельных процессов и зависимостей и включаются как фрагменты в имитационные модели. Второй - модели качественные, которые строятся с целью выяснения динамического механизма изучаемого процесса, способные воспроизвести наблюдаемые динамические эффекты в поведении систем, такие, например, как колебательный характер изменения биомассы или образование неоднородной в пространстве структуры. Обычно эти модели не слишком громоздки, поддающиеся качественному исследованию с применением аналитических и компьютерных методов. Третий класс – имитационные модели конкретных экологических и эколого-экономических систем, учитывающие всю имеющуюся информацию об объекте. Цель построения таких моделей – детальное прогнозирование поведения сложных систем или решение оптимизационной задачи их эксплуатации.

Чем лучше изучен сложная экологическая система, тем более полно может быть обоснована ее математическая модель. При условии тесной связи наблюдений, экспериментального исследования и математического моделирования математическая модель может служить необходимым промежуточным звеном между опытными данными и основанной на них теорией изучаемых процессов. Для решения практических задач можно использовать модели всех трех типов. При этом особенно важны вопросы идентифицируемости (соответствия реальной системе) и управляемости таких моделей.

Обычно при математическом моделировании задача состоит в том, чтобы получить обоснованный прогноз кинетики компонентов экологической системы. При этом делаются различные исходные предположения и преследуются соответствующие цели при изучении моделей, которые один из пионеров математической биологии А.А. Ляпунов сформулировал следующим образом (Ляпунов, 1968, 1972).

А. Биологические характеристики компонентов неизменны, так же как и взаимоотношения между ними. Система считается однородной в пространстве. Изучаются изменения во времени численности (биомассы) компонентов системы.

Б. При сохранении гипотезы однородности вводится предположение о закономерном изменении системы отношений между компонентами. Это может соответствовать либо закономерному изменению внешних условий (например, сезонному), либо заданному характеру эволюций форм, образующих систему. При этом по-прежнему изучается кинетика численности компонентов.

Аппаратом для изучения этих двух классов задач служат системы обыкновенных дифференциальных и дифференциально-разностных уравнений с постоянными (А) и переменными (Б) коэффициентами.

В. Объекты считаются разнородными по своим свойствам и подверженными действию отбора. Предполагается, что эволюция форм определяется условиями существования системы. В этих условиях изучается, с одной стороны, кинетика численности компонентов, с другой - дрейф характеристик популяций. При решении таких задач используют аппарат теории вероятностей. К ним относятся многие задачи популяционной генетики.

Г. Отказ от территориальной однородности и учет зависимости усредненных концентраций от координат. Здесь возникают вопросы, связанные с пространственным перераспределением живых и косных компонентов системы. Например, численность (биомасса) видов - гидробионтов меняется с изменением глубины водоема. Для описания таких систем необходимо привлечение аппарата дифференциальных уравнений в частных производных. В имитационных моделях часто вместо непрерывного пространственного описания применяют разбиение всей системы на несколько пространственных блоков.

Задачи пространственной организации экологических систем представляет особый интерес. До последнего времени предполагали, что пространственная неоднородность распространения видов связана в основном с ландшафтно-климатическими факторами. В последние годы все более глубоко осознается тот факт, что сама пространственная структурированность экологических систем может быть обусловлена не только исходно существующей пространственной неоднородностью, но и спецификой локальных взаимодействий составляющих экосистему популяций между собой и с косными компонентами среды. Возникающие и активно поддерживаемые таким образом пространственные структуры называют экологическими диссипативными структурами.

Биологические популяции и сообщества заведомо являются энергетически "проточными", т.е. далекими от равновесия системами.

колебательные режимы в таких системах давно известны как в лабораторных исследованиях, так и из полевых наблюдений и неплохо исследованы теоретически. Экологические системы подвержены влиянию периодических и нерегулярных геофизических воздействий, их биологические составляющие обладают эндогенными биологическими ритмами (биологические часы). В настоящее время активно решаются проблемы связи между колебательными режимами в локальных (точечных) системах и пространственно-временными структурами в экологических системах. Как и в физических и химических системах, здесь решающую роль играет характер нелинейных взаимодействий, определяющих пути массо- и энергообмена в сложной системе.

Без учета пространственной неоднородности невозможно оценить влияние подвижности особей на регуляцию численности популяций, роль перемещений в синхронизации или затухании колебаний численности, которые имели бы место в отсутствие пространственных перемещений, как направленных, так и случайных – типа диффузии. Современный математический аппарат позволяет выяснить эти вопросы, а также установить связь локальной динамики популяций с крупномасштабными пространственными структурами и долговременной приспособленностью видов и видовых сообществ.

Дискретные модели популяций

Даже в таких популяциях, где особи размножаются несколько лет подряд (млекопитающие и птицы, многолетние растения), наличие сезонов размножения вносит некоторое запаздывание в процессы регуляции численности. Если же взрослые особи, размножающиеся в данном году, редко или никогда не доживают до того, чтобы размножиться в будущем году, как, например, у однолетних растений, мелких грызунов, многих насекомых, это оказывает существенное влияние на динамику их численности.

В этом случае уравнение (7) следует заменить уравнением

$$N_{n+1} = N(x_n), \quad (11)$$

где N_n — численность популяции в году n .

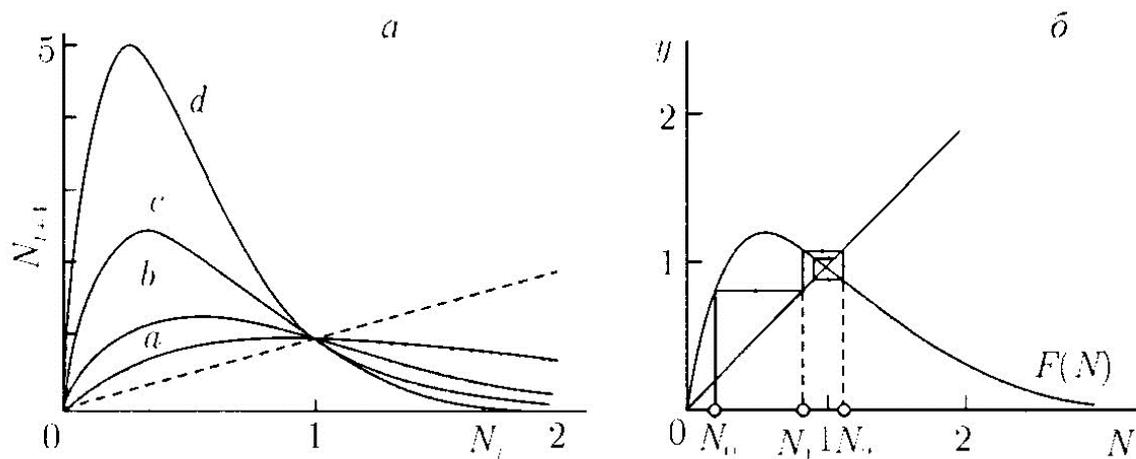


Рис. 5. Модели популяций с неперекрывающимися поколениями, а) Вид одноэкстремальной функции зависимости численности популяции в данный момент времени от численности в предыдущий момент времени. $N_{t+i} = F(N_t)$ б) Определение значений численности популяции в последовательные моменты времени (см. текст) для дискретного аналога логистического уравнения

Наблюдения над динамикой численности показывают, что в таких системах при малых численностях N растет от одной генерации к другой, а при высоких — падает. Это свойство — резко расти при малых N и падать при больших, проявляется в экономике как закон «бумов и спадов». В таких случаях функция F — одноэкстремальная, вид ее изображен на рис. 5а.

Функция такого типа может быть описана с помощью различных формул. Наиболее широко распространена версия дискретного логистического уравнения, предложенная Мораном для численности насекомых (1950) и Рикером для рыбных популяций (1954):

$$N_{t+1} = N_t \exp\{r(1 - N_t/K)\}. \quad (12)$$

Здесь, как и в логистическом уравнении (3), r — константа собственной скорости роста, K — емкость экологической ниши популяции. Ход решения уравнения (12) можно наглядно продемонстрировать графически с помощью диаграммы и лестницы Ламерея. Точка пересечения биссектрисы первого координатного угла $N_{t+1} = N_t$ и функции $F(N_t)$ определяет равновесное состояние системы, аналогичное стационарному состоянию дифференциального уравнения. На рис. 5б показан способ нахождения значений N_t в последовательные моменты времени. Пусть в начальный момент времени $N = N_0$ $F(N_0) = N_1$ задает значение численности в последующий момент времени $t = 1$. Величина N_1 , в свою очередь, определяет значение $F(N_1) = N_2$. И так далее. На рис. 5б изображен случай, когда

траектория сходится к равновесному состоянию, совершая затухающие колебания.

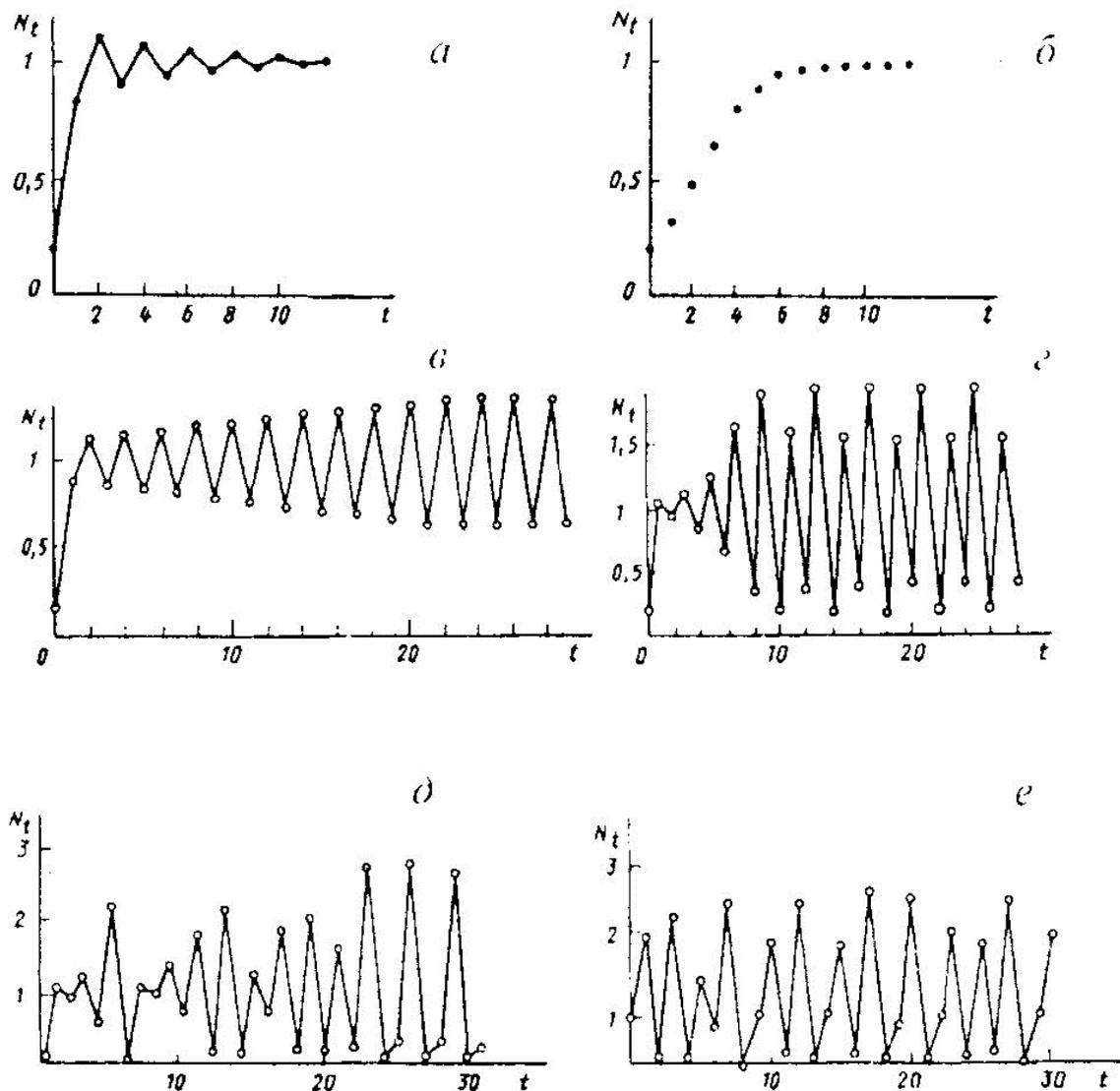


Рис. 6. Типы динамики численности в модели популяции с неперекрывающимися поколениями при разных значениях собственной скорости роста: а — монотонный рост; б — затухающие колебания; в — двухточечный цикл; г — четырехточечный цикл; д, е — квазистохастическое поведение

В зависимости от крутизны графика функции $F(N_i)$ (кривые а, б, в, г на рис. 5) в системе могут возникать самые разнообразные режимы. С ростом γ поведение усложняется. Монотонное стремление к равновесию (рис. 6а) сменяется колебательным (рис. 6б). При дальнейшем увеличении γ

(увеличении крутизны кривой $F(N_i)$) возникают циклы — аналоги предельных циклов для систем дифференциальных уравнений (рис. 6 в, г).

Если g еще больше растет — наблюдается квазистохастическое поведение — хаос (рис. 6 д, е). Модели такого типа являются простейшими детерминированными объектами, демонстрирующими квазистохастическое поведение.

Квазистохастическим поведением могут обладать и переменные в непрерывных нелинейных автономных системах трех и более дифференциальных уравнений. Изображение детерминированного хаоса в популяции из трех видов: хищник-две жертвы — представлено на рис. 12. Таким образом, стохастичность может быть свойством, присущим самим детерминированным природным системам (*Детерминированный хаос*), и не зависит от того, какой математический аппарат, непрерывный или дискретный, используется.

Структурные модели популяций

Основные механизмы, управляющие поведением популяции, могут быть связаны с дифференциацией особей. Таким определяющим фактором является половая структура популяции. Общеизвестна также роль возрастной структуры — так называемые возрастные пирамиды. Например, при моделировании динамики древостоев наибольший интерес представляет не столько численность, сколько запас стволовой древесины, средняя высота и диаметр деревьев (и их дисперсия). Обобщение моделей на случай, когда особь в популяции характеризуется не только возрастом, весом или размером, но одновременно многими факторами, привело к формированию *структурных моделей популяций* (individual based models), которые являются одними из наиболее интенсивно развивающихся за последние 20 лет ветвей математической биологии (Metz J. A. J., Denkmann O. (eds) / *The Dynamics of Physiologically structured populations. Lecture Notes in Biomath.*, 68, Springer, 1986). Теоретическая цель таких моделей — описать поведение популяции в терминах поведения индивидов. Понятно, что расчет структурных моделей требует достаточно продвинутой вычислительной техники.

Структурная модель популяций имеет два уровня описания — индивидуальный и популяционный. На индивидуальном уровне выбирается набор фазовых (структурных) переменных (x_i, X_2, \dots, x_n) . В каждый момент времени индивидум характеризуется возрастом a и состоянием $X = (x_i, x_2, \dots, x_n)$. Затем задается пространство состояний индивидума — область в n -

мерном пространстве, динамика состояний индивидума (обычно, система обыкновенных дифференциальных уравнений), интенсивность гибели индивидуумов, интенсивность рождения от отдельной особи и

распределение новорожденных по пространству состояний, влияние внешней среды и состояния популяции в целом на скорость роста индивидуума и интенсивности рождения и гибели.

Популяционный уровень задается начальным распределением индивидуумов, уравнением на текущее распределение популяции (это уравнение неразрывности или уравнение Колмогорова), общим количеством новорожденных в единицу времени, вычисляемым по интенсивности рождения и текущему распределению популяции.

Таким образом, структурная модель состоит из трех блоков:

- 1) динамическая система, описывающая развитие отдельной особи в популяции;
- 2) уравнение неразрывности на плотность популяции в пространстве состояний;
- 3) интегральное граничное условие на плотность популяции, описывающее процесс возобновления.

Первые структурные модели появились в области микробиологии, где они оказались чрезвычайно плодотворными для использования в биотехнологии с целью оптимизации процесса получения определенных веществ, входящих в состав микроорганизмов, или метаболитов — продуктов жизнедеятельности культивируемых микроорганизмов (Ramkrishna D et.al., 1967; Frederickson A. G., 1976). Наряду с возрастом и размерами микроорганизмов в качестве переменных, характеризующих отдельный микроорганизм, в эти модели входило содержание в нем нуклеиновых кислот и других жизненно важных элементов, определяющих скорость роста биомассы и характерное время деления клеток микроорганизмов, а также содержание вещества, являющегося целевым продуктом биотехнологического процесса. В последние годы структурная теория популяций получила широкое применение при моделировании лесных сообществ.

Лекция № 3. Модели межвидовой конкуренции.

Первые модели динамики популяций – это ряд Фибоначчи (1202), модель экспоненциального роста (1798) Мальтуса, модель ограниченного роста Ферхюльста (1838). К настоящему времени имеется много самых разнообразных дискретных и непрерывных детерминистических и стохастических моделей. В начале 20 века появились первые модели взаимодействия видов. Классической книгой современной математической

экологии является труд В. Вольтерра "Математическая теория борьбы за существование" (Volterra, 1931; Вольтерра, 1976). Развитие теоретической экологии в последующие десятилетия полностью подтвердило глубину и правильность его идей.

Системы, изученные Вольтерра, состоят из нескольких биологических видов и запаса пищи, который используют некоторые из рассматриваемых видов. О компонентах системы формулируются следующие допущения.

1. Пища либо имеется в неограниченном количестве, либо ее поступление с течением времени жестко регламентировано.
2. Особи каждого вида отмирают так, что в единицу времени погибает постоянная доля существующих особей.
3. Хищные виды поедают жертвы, причем в единицу времени количество съеденных жертв всегда пропорционально вероятности встречи особей этих двух видов, т.е. произведению количества хищников на количество жертв.
4. Если имеются пища в неограниченном количестве и несколько видов, которые способны ее потреблять, то доля пищи, потребляемая каждым видом в единицу времени, пропорциональна количеству особей этого вида, взятого с некоторым коэффициентом, зависящим от вида (модели межвидовой конкуренции).
5. Если вид питается пищей, имеющейся в неограниченном количестве, прирост численности вида за единицу времени пропорционален численности вида.
6. Если вид питается пищей, имеющейся в ограниченном количестве, то его размножение регулируется скоростью потребления пищи, т.е. за единицу времени прирост пропорционален количеству съеденной пищи.

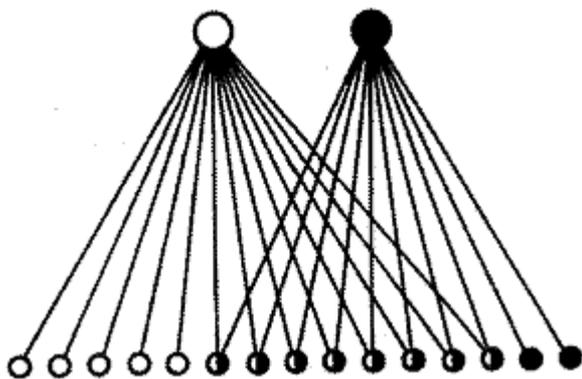
Перечисленные гипотезы позволяют описывать сложные живые системы при помощи систем обыкновенных дифференциальных уравнений, в правых частях которых имеются суммы линейных и билинейных членов. Как известно, такими уравнениями описываются и системы химических реакций. Такое сходство уравнений в химических и экологических моделях позволяет применить для математического моделирования кинетики популяций те же методы исследований, что и для систем химических реакций. Вольтерровские уравнения могут быть получены не только из локального "принципа встреч", ведущего свое происхождение из статистической физики, но и исходя из баланса масс каждого из компонентов биогеоценоза и энергетических потоков между этими компонентами.

Уравнения Вольтерра послужили отправной точкой для создания большинства динамических моделей в экологии вплоть до сегодняшнего дня. Вольтерра изучал сосуществование видов при более широких гипотезах, в частности при изменении внешних условий и с учетом явления

последствия, рассмотрение которого приводит к интегро-дифференциальным уравнениям.

Природные сообщества обладают сложным строением: несколькими уровнями, между которыми существуют разнообразные трофические (пищевые) и топические (не связанные с цепью питания) связи. Структура трофической пирамиды может быть весьма различной в зависимости от климата, почвы, ландшафта, длительности существования биогеоценоза и других факторов.

При анализе биологических сообществ принято строить пищевые или трофические сети, т.е. графы, вершины которых соответствуют видам, входящим в сообщество, а ребра указывают трофические связи между видами. Обычно такие графы - ориентированные: направление дуги между двумя вершинами указывает на тот из видов, который является потребителем другого, т.е. направление дуги совпадает с направлением потока вещества или биомассы в системе.



Пример двухвозрастной трофической пирамиды

Часто трофические графы изображают в виде **трофических пирамид**. В такой пирамиде выделяются **трофические уровни** - группы видов, между которыми невозможны прямые пищевые связи. Уровней может быть несколько: виды, принадлежащие одному уровню, либо находятся в состоянии конкуренции за жизненные ресурсы, либо совместно используют ресурс. На рис. 1 изображен пример двухуровневой трофической пирамиды, взятой из книги Ю. Одум "Основы экологии" (1975). Из 15 видов насекомых (верхний уровень) 5 видов питаются только на одном из двух видов растений, 2 вида - только на втором, в рацион остальных 8 видов насекомых входят оба вида растений. Основные трофические уровни наземных сообществ - это продуценты или автотрофы (растения, аккумулирующие энергию света и вещества субстрата) и гетеротрофы: первичные консументы (травоядные) и вторичные консументы (хищники, питающиеся травоядными). В некоторых случаях трофическая цепь содержит большее

число уровней: например, растения служат пищей насекомым, насекомые поедаются птицами, которые в свою очередь служат пищей более крупным хищным птицам.

Если в структуре сообщества учитывать движение некоторых биогенных элементов и энергии, то в системе обнаруживаются петли обратной связи. Разлагатели (редуценты) - микробы, грибы, бактерии - в процессе своей жизнедеятельности расщепляют сложные органические соединения (экскременты и мертвую органику) на более простые минеральные вещества, необходимые продуцентам. Образование органической биомассы происходит в процессе фотосинтеза с использованием солнечного света из углекислого газа и воды, причем необходимы также элементы, поступающие из почвы: азот, фосфор, калий, магний, железо и многие другие микроэлементы. Общая схема потоков массы и энергии между основными компонентами наземных экосистем изображена на рис.

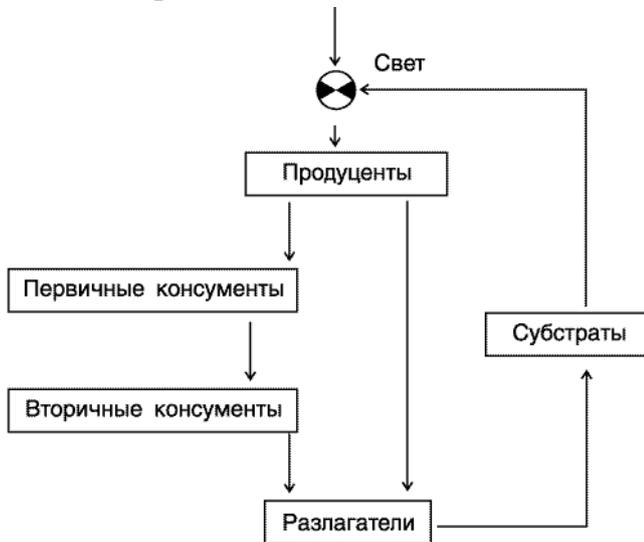


Рис. 2. Общая схема потоков вещества и энергии в экосистеме.

Полную структуру парных взаимодействий между n видами можно изобразить с помощью матрицы S из $n \times n$ элементов. Элемент (i, j) представляет собой знак $+$, $-$ или 0 и показывает влияние i -го вида на j -й. Симметричные пары элементов матрицы S указывают на тип парного взаимодействия между видами:

+ +	симбиоз или мутуализм;
+ -	хищник - жертва (паразит - хозяин);
+ 0	комменсализм;
- -	конкуренция;
- 0	аменсализм;
0 0	нейтрализм.

Взаимодействие между видами можно характеризовать и при помощи знакового ориентированного графа, который строится по следующему правилу. Если вид j влияет каким-либо образом на вид i , то проводится ребро $i \rightarrow j$ и ему приписывается знак этого влияния.

В современной литературе по математической экологии принято считать вольтерровскими моделями сообществ биологических видов системы вида

$$dN_i / dt = N_i \left(\varepsilon_i - \sum_{j=1}^n \gamma_{ij} N_j \right), \quad i = 1, 2$$

где ε_i - скорость естественного прироста или смертности i -го вида в отсутствие всех остальных видов, а знак и абсолютная величина $\gamma_{ij} (i \neq j)$ отражают соответственно характер и интенсивность влияния j -го вида на i -й вид, показатель внутривидового взаимодействия для i -го вида. Матрицу $\Gamma = \|\gamma_{ij}\|$, отражающую структуру связей сообщества, называют **матрицей** сообщества.

С введенной выше знаковой матрицей S она связана соотношением

$$S = - \text{sign } \Gamma.$$

Рассмотрим сообщество, структура которого изображена на рис.2. Компоненты сообщества разобьем на три основные группы. 1. Продуценты с биомассами (или концентрациями) $x_i (i = 1, 2, \dots, m)$ - это в основном зеленые растения. 2. Консументы с концентрациями $y_j (j = 1, 2, \dots, n)$. К этой группе отнесем животных, пожирающих другие организмы и разлагателей, расщепляющих мертвую органику на простые вещества, которые используются продуцентами. 3. Субстраты с концентрациями $C_k (k = 1, 2, \dots, p)$. Это абиотические вещества (в основном продукты жизнедеятельности консументов), используемые продуцентами.

Составим уравнения, отражающие баланс масс каждого из этих компонентов:

$$dx_i / dt = (F_x^i - D_x^i)x_i - \sum_{j=1}^n V_{ij} - y_i + R_x, \quad i = 1, \dots, m,$$

$$dy_j / dt = (F_y^j - D_y^j)y_j - \sum_{r=1}^n v_{jr}y_r + R_y, \quad j = 1, \dots, n,$$

$$dc_k / dt = \sum_{j=1}^n U_{kj}y_j - \sum_{i=1}^m W_{ki}x_i + R_c, \quad k = 1, \dots, p.$$

Здесь F^i , D^i - функции рождаемости и смертности продуцентов и консументов; V_{ij} - функция выедания, описывающая скорость потребления биомассы i -го вида-продуцента единицей биомассы j -го вида-консумента; v_{jr} - функция выедания j -го вида r -м (среди консументов); U_{kj} - интенсивность производства k -го субстрата j -м консументом; W_{ki} - интенсивность потребления k -го субстрата i -м продуцентом; $R_x = R_y = R_c$ - сумма входных и выходных потоков соответствующих компонент. В общем случае все эти функции зависят от параметров внешней среды. Путем преобразования переменных система (2) может быть записана в виде, сходном с уравнениями Вольтерра (1).

Применение этого формализма и его модификаций оказалось особенно успешным при моделировании замкнутых по веществу систем. (Алексеев, 1993). Пример трофической цепи для такой замкнутой системы (например, озерной экосистемы) приведен на рис.:

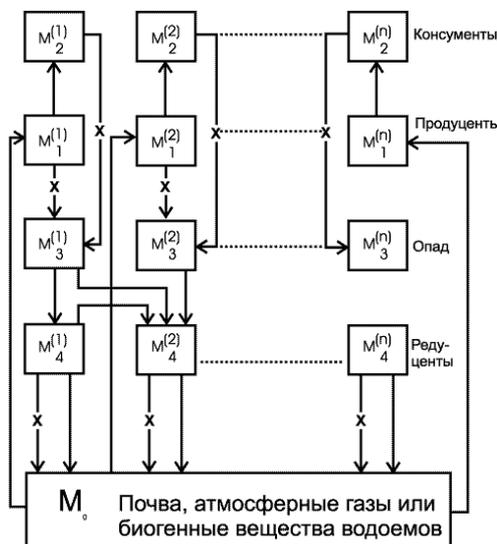


Рис.3. Знакоориентированный граф сообщества из трех видов.

Лекция № 4. Модели популяций и сообществ.

На разных уровнях развития живой материи продукционные процессы проявляют себя по-разному, но их феноменологическое описание всегда включает рождение, рост, взаимодействие с внешней средой, в том числе с другими особями своего вида или других видов, смерть особей. Именно это обстоятельство позволяет применять сходный математический аппарат для описания моделей роста и развития у таких, казалось бы, удаленных друг от друга по лестнице уровней организации живой материи, как клеточная популяция и сообщество видов в экосистеме.

Описание изменения численности популяции во времени составляет предмет *популяционной динамики*. Популяционная динамика является частью *математической биологии*, наиболее продвинутой в смысле формального математического аппарата, своего рода «математическим полигоном» для проверки теоретических идей и представлений о законах роста и эволюции биологических видов, популяций, сообществ. Возможность описания популяций различной биологической природы одинаковыми математическими соотношениями обусловлена тем, что, с динамической точки зрения, рост и отбор организмов в процессе эволюции происходит по принципу «Кинетического совершенства» (Шноль, 1979).

Преимущества математического анализа любых, в том числе популяционных, процессов очевидны. Математическое моделирование не только помогает строго формализовать знания об объекте, но иногда (при хорошей изученности объекта) дать количественное описание процесса, предсказать его ход и эффективность, дать рекомендации по оптимизации управления этим процессом. Это особенно важно для биологических процессов, имеющих прикладное и промышленное значение, — *биотехнологических систем*, *агробиоценозов*, *эксплуатируемых природных экосистем*, продуктивность которых определяется закономерностями роста популяций живых организмов, представляющих собой «продукт» этих биологических систем.

Ряд Фибоначчи

Постановка математических задач в терминах популяционной динамики восходит к глубокой древности. Человеку свойственно рассуждать о предметах, жизненно ему близких, и что может быть ближе, чем законы размножения популяций — людей, животных, растений.

Первая дошедшая до нас математическая модель динамики популяций приводится в книге «Трактат о счете» («*Liber abaci*»), датированной 1202 годом, написанной крупнейшим итальянским ученым Леонардо Фибоначчи

— Леонардо из Пизы (предположительно 1170-1240). В этой книге, представляющей собой собрание арифметических и алгебраических сведений того времени и впоследствии распространившейся в списках по всей Европе, рассматривается следующая задача. «Некто выращивает кроликов в пространстве, со всех сторон обнесенном высокой стеной. Сколько пар кроликов рождается в один год от одной пары, если через месяц пара кроликов производит на свет другую пару, а рожают кролики начиная со второго месяца после своего рождения.» Решением задачи является ряд чисел:

$$1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, 233, 377, \dots (1)$$

(Сам Леонардо опустил первый член ряда.) Два первых числа соответствуют первому и второму месяцу размножения. 12 последующих — месячному приросту поголовья кроликов. Каждый последующий ряд равен сумме двух предыдущих. Ряд (1) вошел в историю как ряд Фибоначчи, а его члены — чисел Фибоначчи.

Это первая известная в Европе рекурсивная последовательность чисел (в которой соотношение между двумя или более членами ряда может быть выражено в виде формулы). Рекуррентная формула для членов ряда Фибоначчи была записана французским математиком Альбертом Гирером в 1634 г.:

$$U_{n+2} = U_{n+1} + U_n.$$

Здесь U представляет собой член последовательности, а нижний индекс — его номер в ряду чисел. В 1753 г. математик из Глазго Роберт Симпсон заметил, что при увеличении порядкового номера членов ряда отношение последующего члена к предыдущему приближается к числу a , называемому «Золотым сечением», равному $1,6180\dots$, или $(1 + \sqrt{5})/2$. В 19 веке о свойствах ряда Фибоначчи и его связи с Золотым сечением много писал французский математик Эдуард Лукас. С тех пор естествоиспытатели наблюдают его закономерности в расположении чешуек на шишках, лепестков в цветке подсолнуха, в спиральных образованиях ракушек моллюсков и других творениях природы. Ряд Фибоначчи и его свойства также используются в вычислительной математике при создании специальных алгоритмов счета.

Уравнение экспоненциального роста

Второй всемирно известной математической моделью, в основу которой положена задача о динамике численности популяции, является классическая модель неограниченного роста — геометрическая прогрессия в дискретном представлении, $An+1 = qAn$, или экспонента, в непрерывном непрерывном Модель предложена Мальтусом в 1798 г. в его классическом

труде «О росте народонаселения». Томас Роберт Мальтус (1766-1834), известный английский демограф и экономист, обратил внимание на тот факт, что численность популяции растет по экспоненте (в геометрической прогрессии), в то время как производство продуктов питания растет со временем линейно (в арифметической прогрессии), из чего сделал справедливый вывод, что рано или поздно экспонента обязательно «обгонит» линейную функцию и наступит голод. На основании этих выводов Мальтус говорит о необходимости ввести ограничения на рождаемость, в особенности для беднейших слоев общества. «Экономический пессимизм», следующий из прогнозов предложенной им модели, в основу которой положен анализ эмпирических данных, Мальтус противопоставлял модным в начале 19 века оптимистическим идеям гуманистов: Жана Жака Руссо, Уильяма Годвина и других, предсказывающих человечеству грядущее счастье и процветание. Можно говорить о том, что Мальтус был первым ученым-«алармистом», который на основании результатов моделирования «бил тревогу» и предупреждал человечество об опасности следования развитию по используемым ранее сценариям прогресса. Во второй половине XX века такую «алармистскую» роль сыграли работы Римского клуба и, в первую очередь, «модель глобального роста» Дж. Форрестера (см. *Математическая экология*).

Обсуждению важности вывода Мальтуса для популяционной динамики Дарвин посвятил несколько страниц своего дневника, указывая, что, поскольку ни одна популяция не размножается до бесконечности, должны существовать факторы, препятствующие такому неограниченному размножению. Среди этих факторов может быть нехватка ресурса (продовольствия), вызывающая конкуренцию внутри популяции за ресурс, хищничество, конкуренция с другими видами. Результатом являются замедление скорости роста популяции и выход ее численности на стационарный уровень.

Ограниченный рост

Впервые системный фактор, ограничивающий рост популяции, описал Ферхюльст в уравнении логистического роста (1848):

$$\frac{dx}{dt} = rx \left(1 - \frac{x}{K}\right) \quad (3)$$

Это уравнение обладает двумя важными свойствами. При малых x численность x возрастает экспоненциально (как в уравнении (2)), при больших — приближается к определенному пределу K . Эта величина, называемая емкостью популяции, определяется ограниченностью пищевых ресурсов, мест для гнездования, многими другими факторами, которые могут

быть различными для разных видов. Таким образом, емкость экологической ниши представляет собой системный фактор, который определяет ограниченность роста популяции в данном ареале обитания.

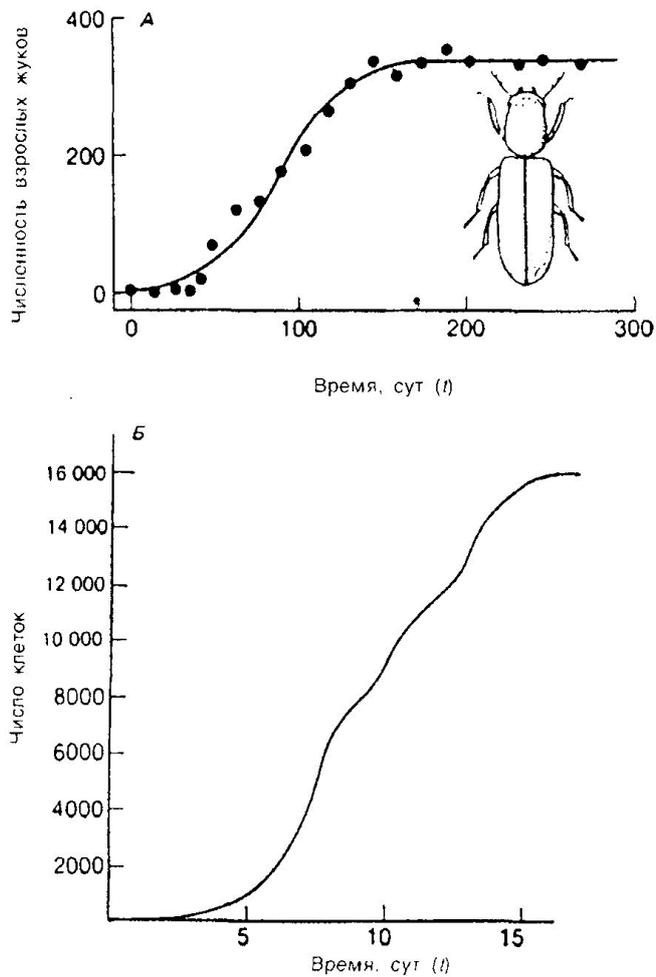


Рис. 1. Ограниченный рост, А) Динамика численности жука *Rhizopertha dominica* в 10-граммовой порции пшеничных зерен, пополняемых каждую неделю. Точки — экспериментальные данные, сплошная линия — логистическая кривая, Б) Динамика численности водоросли *Chlorella* в культуре.

Уравнение (3) можно также переписать в виде

$$\frac{dx}{dt} = rx - \delta x^2, \quad (4)$$

Здесь δ — коэффициент внутривидовой конкуренции (за пищевой ресурс, убежища и т.п.). Уравнение (3) можно решить аналитически. Решение имеет вид:

$$x(t) = \frac{x_0 K e^{rt}}{K - x_0 + x_0 e^{rt}}, \quad (5)$$

Формула (5) описывает кинетическую кривую, то есть зависимость численности популяции от времени. Примеры экспериментально

наблюдаемой динамики популяций, развивающихся по логистическому закону, приведены на рис. 1 а, б. На рис. 1 а сплошной линией представлен график функции (5). Если выражение (5) продифференцировать два раза по t , увидим, что кривая $x(t)$ имеет точку перегиба, с координатами

$$\left(\frac{1}{r} \ln \frac{K - x_0}{x_0}; \frac{K}{2}\right).$$

Ордината представляет собой половину максимальной численности, а абсцисса зависит как от емкости популяции K , так и от константы собственной скорости роста r — чем выше генетические возможности популяции, тем скорее наступает перегиб на кривой численности.

В природе популяции имеют не только максимальную численность, определяемую величиной экологической ниши K , но и минимальную критическую численность L . При падении численности популяции ниже этой критической величины из-за неблагоприятных условий или в результате хищнического промысла восстановление популяции становится невозможным.

Величина нижней критической плотности различна для разных видов. Исследования биологов показали, что она может составлять всего лишь пару особей на тысячу квадратных километров в случае ондатры и сотни тысяч особей для американского странствующего голубя. Заранее трудно было предположить, что столь многочисленный вид уже перешел через критическую границу своей численности и обречен на вымирание. Например, для голубых китов критическая граница численности оказалась равной десяткам-сотням. Хищническое истребление этих гигантских животных привело к тому, что их осталось слишком мало в Мировом океане. И хотя охота на них давно запрещена, надежд на восстановление популяции голубых китов практически нет.

Кривые показателей численности для трех видов китов приведены на рис. 2. Модели, описывающие как внутривидовую конкуренцию, определяющую верхнюю границу численности популяции, так и нижнюю критическую численность популяции, имеют два устойчивых стационарных решения. Одно из них — нулевое для начальных численности, которые ниже наименьшей критической численности популяции. Другое равно K — емкости экологической ниши в случае, когда начальная численность выше наименьшей критической величины. Такими «триггерными» свойствами обладает нелинейное уравнение, предложенное А.Д. Базыкиным:

$$\frac{dx}{dt} = \alpha \frac{\beta x^2}{\beta + \tau x} - dx - \delta x^2, \quad (6)$$

В формуле (6) первый член в правой части описывает размножение двуполой популяции, скорость которого пропорциональна квадрату численности (вероятности встреч особей разного пола) для малых плотностей и пропорциональна числу самок в популяции — для больших плотностей популяции. Второй член описывает смертность, пропорциональную численности, а третий — внутривидовую конкуренцию, подобно тому, как это было в логистическом уравнении (4).

Зависимости численности от времени и скорости прироста от численности представлены на рис. 3 (а,б). Кривые 1-5 соответствуют различным начальным численностям $x=0$ и $x=K$ — устойчивые стационарные состояния, $x=L$ — неустойчивое, разделяющее области влияния устойчивых состояний равновесия.

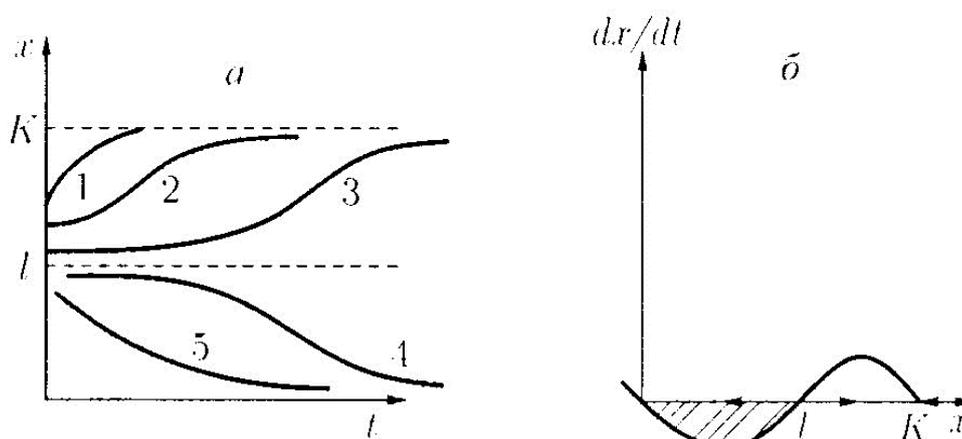


Рис. 3. Модель популяции с нижней критической численностью. Зависимость численности популяции от времени (а) и скорости роста от численности (б) для модели (6). Штриховкой обозначена область вырождения популяции.

Величины L и K различны для разных популяций и могут быть определены из наблюдений и экспериментов.

Из рисунка 3а видно, что скорость восстановления популяции после ее падения, в силу промысла или неблагоприятных условий, зависит от того, насколько близка новая начальная численность к опасной границе L . Если ущерб, нанесенный популяции, невелик (меньше половины емкости экологической ниши), популяция быстро восстанавливается по кривой 1, не имеющей точки перегиба. В случае, когда численность оставшейся популяции близка к критической, восстановление происходит сначала очень медленно, популяция надолго «застревает» вблизи опасной границы, а затем уже, «набрав силы», более быстро выходит на устойчивый стационарный

уровень K (кривая 3). Кривая 2 представляет промежуточный случай. Кривые 4, 5 иллюстрируют вырождение популяции в случае, когда начальная численность опустилась ниже критической границы. Обращает на себя внимание сходство начальных участков кривых 3 и 4. Близость к опасной границе со стороны больших значений (3) и меньших (5) выражается в долгом пребывании системы в неопределенном состоянии, когда малые флуктуации могут легко «перебросить» систему через опасную границу в «благополучную» область возврата к стационарному значению K или, наоборот, — в область вымирания.

В это время сторонний наблюдатель не сможет определить по форме кривой динамики численности, какая судьба ожидает систему. Для самих участников жизненной драмы — нахождения системы вблизи опасной границы — исход не очевиден. Важно понимать, что в этой ситуации чрезвычайно важны любые, даже очень малые усилия, направленные на преодоление критического барьера.

Именно популяции, численность которых близка к нижней критической численности, занесены в Красную книгу. Удастся ли перенести каждый конкретный вид на «Зеленые страницы», куда переносят виды, исчезновение которых удалось предотвратить, — зависит от многих обстоятельств, в частности, как от репродуктивных усилий вида, так и от усилий людей, спасающих эти виды.

Лекция № 5. Статистическая обработка результатов исследований в экологии.

Экологические исследования, как правило, связаны с анализом множества переменных. Необходимо дать оценку этим данным, проанализировать их, визуально представить. Статистическая обработка данных является необходимым этапом при выработке практических рекомендаций для рационального управления экосистемами и ведения экосистемных научных исследований. Определим основные понятия.

Для того чтобы данные можно было подвергнуть статистической обработке их надо правильно получить. Для начала стоит определиться с *целью исследования*. Цели в свою очередь определяют *объект исследования*. Объектом может являться совокупность предметов, явлений и т.п., включённых в эксперимент. В этой совокупности можно выделить первичный элемент, обладающий регистрируемыми признаками. Этим элементом является *единица исследования*.

Формулирование и корректировка признаков производится на основании следующих критериев:

- отбор должен производиться с учётом целей исследования, возможности их обработки и анализа полученных данных;
- отобранных признаков должно быть оптимальное количество (ни много, ни мало);
- необходимо комбинировать признаки для их взаимного дополнения.

Совокупность всех, интересующих нас в данном эксперименте явлений называется *генеральной совокупностью*. Но так как в большинстве случаев мы не имеем возможности проводить сплошное обследование интересующего нас массива, мы пользуемся *выборочным исследованием*. Полученный массив данных при данном способе называется выборкой.

Выборочная совокупность должна иметь несколько свойств:

- Репрезентативность. То есть результаты, полученные в процессе обработки выборочной совокупности, можно было перенести на соответствующую генеральную совокупность.
- Случайность. Для соблюдения репрезентативности выборка должна быть случайной, то есть вероятность выборки каждого единичного объекта генеральной совокупности должна иметь равную вероятность.
- Также она должна быть достаточно полной.

Не менее важен подход к сбору первичных данных для анализа. Как уже было отмечено выше, сбор данных должен быть обусловлен целью исследования и является одним из важнейших этапов анализа системы. Последующее преобразование определяется выбором определённой шкалы.

Под шкалой будем понимать совокупность трёх элементов $\langle E, N, \Psi \rangle$, где E – эмпирическая система; N – числовая система; Ψ – однозначное отображение E на N . Или в словесном выражении:

Шкала – это средство фиксации результатов измерения свойств объектов путём упорядочивания их в определённую числовую систему, в которой отношение между отдельными результатами выражено в соответствующих числах.

Типы шкал:

1. *Абсолютная.* В данной шкале соблюдается соотношение $N1=N2$. Данный тип шкалы является самым «сильным» и в пределах этой шкалы можно производить любые арифметические действия. Примером может являться подсчёт количества особей, а также факт наличия вида на исследуемом участке (эти данные не являются номинальными, как считают некоторые исследователи).

2. *Относительная.* В данной шкале соблюдается соотношение $N1 = a \times N2$, при $a > 0$. Классическим примером является измерение массы в килограммах и в фунтах.

3. *Интервальная.* Соблюдается соотношение $N1 = a \times N2 + b$, при $a > 0$, b – действительное число. В приведённом типе шкал важным является понятие точки отсчёта. Известным примером является соответствие шкалы температуры по Цельсию и шкалы по Фаренгейту.

4. *Порядковая.* Между $N1$ и $N2$ имеется любое монотонное преобразование, то есть можно лишь отметить некоторую упорядоченность объектов относительно друг друга. Примером является шкала Бофорта (штиль, шторм).

5. *Номинальная.* Шкала служит для придания объектам имён. Например, нумерация солдат в строю.

Шкала даёт возможность дать явлениям количественную оценку (квантифицировать). Кроме того, тип шкалы накладывает определённые правила на собираемые данные. Если с помощью некоторого преобразования можно перейти от записи $N1$ к $N2$ то такие шкалы называют шкалами одного типа, а данное преобразование является *допустимым преобразованием*.

Результаты измерений и подсчёта признаков записываются в виде таблицы. Каждое измерение получает свой номер. Признаки по своему характеру могут быть количественными и качественными, причём первые имеют два типа варьирования (изменчивости) – дискретный (разница между соседними значениями не менее единицы) и непрерывный (соседние значения признака могут сколь угодно мало отличаться друг от друга). Для выделения классов в непрерывных данных применяют формулу Стерджеса:

$$m \approx 1 + 1 \log_2 n \approx 1 + 3,322 \cdot \log n \quad ,$$

где m – рекомендуемое число интервалов разбиения; n – общее число наблюдений.

Каждое частное значение, которое принимает признак, называется *вариантой*. Число повторений определённых вариантов называется *частотой*.

Совокупность вариантов называется *рядом распределения*. В зависимости от типа признака выделяют атрибутивный ряд распределения (качественные признаки) и вариационный ряд распределения (количественные признаки).

Под распределением случайной величины понимается совокупность всех возможных значений случайной величины и соответствующих им вероятностей. Всякое соответствие между возможными значениями случайных величин и их вероятностей называется *законом распределения*

случайной величины. В экологии наиболее часто встречаются следующие виды: нормальное (длина тела особей в популяции), экспоненциальное (рост численности населения по Т. Мальтусу), биномиальное (отражает распределение по двум полам в человеческой популяции).

Функция распределения $F(x)$ определяет вероятность того, что случайная величина X примет значение, меньшее фиксированного действительного числа x :

$$F(x) = P(X < x)$$

Для непрерывной случайной величины с дифференцируемой функцией распределения $F(x)$ вводят дифференциальную функцию распределения (или плотность распределения) $f(x) = F'(x)$. Для дискретной случайной величины такой функции (соответствующего аналога) нет.

Описательная статистика

Первичная статистическая обработка проводится с использованием описательной статистики. Некоторые показатели уже знакомы студенту из курса высшей математики. Соотнесём показатели со шкалами. Целесообразно такие показатели как – выборочное среднее (и другие средние), дисперсию, стандартное отклонение, коэффициент вариации, минимум, максимум, размах, мода применять для значений, полученных в абсолютной и относительной шкалах. Для более слабых шкал применяются такие показатели как минимум, максимум, квантили (сюда относятся и квартили, а также медиана, как частный случай), квартильный размах. Для кривой распределения показательными являются показатели асимметрии (отклонение графика от середины) и эксцесса (островершинности графика).

Параметрические и непараметрические критерии

Статистические критерии, с помощью которых можно установить достоверность различия между параметрами (M , σ) вариационных рядов одноимённого признака в двух выборках (или в выборке и генеральной совокупности) называются *параметрическими*. Они используются при предположении, что распределения сравниваемых рядов близки к нормальному.

Критерий Стьюдента (t). Весьма известный критерий, предложенный У. Госсетом. Критерий Стьюдента для сравнения одноимённых параметров (P_1 и P_2) двух вариационных рядов имеет при $n > 20$ в общей форме вид:

$$t = \frac{P_1 - P_2}{m_{P_1 - P_2}},$$

где в знаменателе стоит ошибка разности этих параметров, представляющая собою корень квадратный из суммы квадратов ошибок репрезентативности выборочных параметров:

$$m_{P_1-P_2} = \sqrt{m_{P_1}^2 + m_{P_2}^2},$$

С учётом формул ошибок репрезентативности критерий t приобретает окончательный вид:

- для сравнения средних арифметических:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{m_{M_1}^2 + m_{M_2}^2}},$$

- для сравнения средних квадратических отклонений:

$$t = \frac{\sigma_1 - \sigma_2}{\sqrt{\frac{\sigma_1^2}{2n_1} + \frac{\sigma_2^2}{2n_2}}},$$

Оценка достоверности разницы производится с помощью сравнения полученного значения t со стандартным (t_{st}), взятым из соответствующей таблицы при выбранном уровне достоверности и числе степеней свободы.

Критерий Фишера (F) является более точным критерием сравнения средних квадратических отклонений. Он представляет собой отношение двух дисперсий:

$$F = \frac{\sigma_1^2}{\sigma_2^2},$$

причём в числителе берут большую дисперсию их двух.

Для сравнения двух выборок, распределение которых далеко от нормального или выборки весьма малы, рекомендуется использовать *непараметрические критерии* различия. Рассмотрим некоторые из них, которые активно применяются в биологических науках (в частности в экологических исследованиях).

Критерий χ -квадрат. Критерий был открыт ещё в 1875-1877 гг. Хельмертом, но затем был забыт и открыт уже К. Пирсоном. Рассчитывается по формуле:

$$\chi^2 = \sum \frac{(f - f^*)^2}{f^*},$$

где f – наблюдаемая частота; f^* - ожидаемая частота.

– Если $\chi^2 > \chi_{2_{st.}}$, то нулевая гипотеза об отсутствии различия между теоретическим и эмпирическим распределениями отвергается.

– Если $\chi^2 < \chi^2_{st}$. то нулевая гипотеза об отсутствии различия между теоретическим и эмпирическим распределениями принимается.

Число степеней свободы находится по формуле: $\nu = M - 1$, где M – число классов.

Рассмотрим пример: Проверим гипотезу об отсутствии относительной приуроченности вида к какому-либо местообитанию на примере пчелы *Megachile rotundata* (F.) по 7-летним материалам. $M = 5$; $N = 22905$;

Ожидаемое число особей рассчитываем по формуле:

$$\frac{P_{ij}}{N_j} = \frac{n_j N_j}{N},$$

где N_j – число особей S видов в j -ой выборке. $N_j = \sum n_{ij}$, ($i = 1, 2, \dots, S$),
а n_i – общее число особей одного вида во всех выборках M ; $n_i = \sum n_{ij}$,
($j = 1, 2, \dots, M$), а p_{ij} – доля i -го вида в j -ой выборке.

Оценка различий между ожидаемым и наблюдаемым распределением частот по критерию хи-квадрат

Номер местообитания, j	Объём выборки, N_j	Фактическое число особей i -го вида, $n_{ij}=f$	Ожидаемое число особей i -го вида, $p_{ij}N_j = f^*$	$f-f^*$	$(f-f^*)^2$	$(f-f^*)/f^*$
1	5483	25	24,68	0,32	0,104	0,004
2	1683	7	7,56	-0,56	0,314	0,041
3	2047	14	9,21	4,79	22,944	2,488
4	9578	36	43,09	-7,09	50,298	1,164
5	4114	21	18,52	2,48	6,150	0,332
Сумма	22905	103	103,06			4,029

При $p = 0,05$, $\chi^2_{st} = 9,5$, следовательно гипотеза принимается.

Графическое представление данных

Современную науку невозможно представить без применения графического представления данных. Самые различные формы графического представления уже давно стали средством научного обобщения. Выразительность, доходчивость, лаконичность, универсальность, обозримость графических изображений сделали их незаменимыми в исследовательской работе и при сопоставлении природных явлений, а также в качестве промежуточных ориентиров при вычислениях.

Впервые о технике составления статистических графиков видимо упоминается в работе английского экономиста У. Плейфейра «Коммерческий и политический атлас», опубликованной в 1786 году и положившей начало развитию приёмов и методов графического изображения статистических данных.

При построении графического изображения следует соблюдать ряд требований. Прежде всего, график должен быть достаточно наглядным, так как весь смысл графического изображения как метода анализа в том и состоит, чтобы наглядно изобразить статистические показатели. Кроме того, график должен быть выразительным, доходчивым и понятным.

Для выполнения вышеперечисленных требований каждый график должен включать ряд основных элементов:

1. *Графический образ (основа графика)*. Совокупность точек, фигур, линий и других геометрических знаков, с помощью которых изображаются статистические показатели.

2. *Поле графика*. Часть плоскости, где расположены графические образы.

3. *Пространственные ориентиры графика*. Данный компонент задаётся в виде системы координатных сеток. Наиболее распространена система прямоугольных координат.

4. *Масштабные ориентиры*. Определяются масштабом и системой масштабных шкал. *Масштабом* называют меру перевода числовой величины в графическую. *Масштабной шкалой* называют линию, отдельные точки которой могут быть прочитаны как определённые числа.

5. *Экспликация*. Представляет словесное описание своего содержания. Включает название графика, подписи шкал и пояснения к отдельным частям графика.

Классификация графиков

Существует огромное количество вариантов классификации: по решаемым задачам, по способу построения (диаграммы, статистические карты).

1. *Линейные графики*. Представлены в виде точек, соединённых линиями.

2. *Диаграммы*. Наиболее часто используемые графики после классических линейных. Диаграммы облегчают предварительное изучение первоначальных данных.

– *Диаграммы рассеяния*. Служат для визуальной оценки распределения объектов статистической совокупности. Объекты совокупности представлены в виде точек между осями координат.

- *Линейные диаграммы.* Для характеристики динамики (то есть изменения во времени). По оси *X* – отрезки времени, а по оси *Y* уровни ряда динамики или темпы изменения. Точки соединяются в виде ломаной линии. Являются некоторым развитием идеи линейных графиков.
- *Столбчатые диаграммы.* Столбцы имеют основание и высоту, пропорционально числовым значениям признака.
- *Секторные диаграммы.* Используются для анализа структуры. Обычно используют *круговые диаграммы*, реже в качестве основной фигуры используется прямоугольник.
- *Диаграммы размаха.* Так называемые «ящички-усы». Показывают на одном графике диапазоны значений переменной. Часто используют для визуализации значений медианы и квартилей выборки.

3. *Статистические карты.* Используют для оценки географического размещения явлений. Они включают картограммы и картодиаграммы.

- *Картограммы.* Показывают территориальное распределение признака по отдельным районам. *Фоновые картограммы* разной густотой окраски (или цветом) характеризуют распределение признака на территории. На *точечной картограмме* каждой точке соответствует одно и то же принятое числовое значение. *Фоновые* показывают относительные величины, а *точечные* – абсолютные.
- *Картодиаграмма* – сочетание карты и диаграммы.

4. *Трёхмерные графики.* Дублируют все двухмерные. Применяются или в виде макетов или в виде компьютерного изображения.

- *Графики поверхности.* Применяются для выявления взаимосвязей между большим количеством переменных. Также применяются в картографии.

5. *Пиктографики.* Наблюдения или отдельные испытания представлены в виде картинок со многими элементами (лица Чернова, контуры).

Даже всё перечисленное не охватывает полностью всего разнообразия графического представления. Существует еще немало число специфических графиков: это различные объекты теории графов и топологии, визуализация некоторых многомерных статистических отношений (например, кластерный анализ, факторный анализ и т.п.), методов ординации, структуры некоторых систем искусственного интеллекта и прочее.

Лекция № 6. Общая схема статистического анализа.

Статистические распределения характеризуются наличием более или менее значительной вариации в величине признака у отдельных единиц совокупности. Отсюда возникает вопрос о том, какие же причины

формируют уровень признака в данной совокупности и каков конкретный вклад каждой из них.

Исследования показывают, что вариация каждого изучаемого признака находится в тесной связи с вариациями других признаков, характеризующих исследуемую совокупность единиц. По своему значению для изучения взаимосвязи признаки подразделяются на:

1. *Результативные признаки.* Признаки, которые изменяются под действием других, связанных с ними признаков.

2. *Факторные признаки.* Признаки, обуславливающие изменение результативных признаков.

Под *статистической связью* мы будем понимать зависимость, при которой изменение одной из величин влечёт изменение распределения другой.

По характеру зависимости признаков различают:

1. *Функциональная (полная) связь.* Вид связи, при котором определённому значению факторного признака соответствует одно и только одно значение факторного признака.

2. *Корреляционная (неполная) связь.* Вид связи, при которой статистическая зависимость проявляется в том, что определённому значению факторного признака соответствует лишь среднее значение результативного признака.

Задачей *корреляционного анализа* является количественное определение степени связности между признаками (при парной связи) и между результативными и факторными признаками (при многофакторном анализе).

Корреляционный анализ предваряет различные сложные методы статистического анализа и проявляется в основном в расчёте коэффициентов корреляции.

Термин «корреляция» был введён Ф. Гальтоном в 1886 году. Однако точную формулу для подсчёта коэффициента корреляции предложил его ученик К. Пирсон. Коэффициент характеризует наличие только линейной связи между признаками, обозначаемыми, как правило, символами X и Y .

Формула расчёта коэффициента корреляции построена таким образом, что если связь между признаками имеет линейный характер, то коэффициент Пирсона точно устанавливает тесноту этой связи. Поэтому данный коэффициент ещё называют коэффициентом линейной корреляции Пирсона. Если же связь не линейна, то Пирсоном предлагается использовать, так называемое, корреляционное отношение. Предполагается, что переменные X и Y распределены *нормально*.

В общем виде коэффициент корреляции можно представить следующим образом:

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2}}$$

где $-1 \leq r \leq 1$ (если при расчётах получается величина вне пределов диапазона, то следует искать ошибку в вычислениях); X_i – значения выборки X ; Y_i – значения выборки Y ; \bar{X} – средняя по X ; \bar{Y} – средняя по Y . Знак коэффициента корреляции очень важен для интерпретации полученной связи. Если корреляция положительная, то связь между признаками такова, что увеличению значения первого признака соответствует увеличение значения второго признака. Обратным данному виду связи будет отрицательная корреляция, при которой увеличению значения первого признака соответствует уменьшение значения второго признака. Если взять значения из числителя коэффициента корреляции и разделить его на n (число значений одной из переменных), то мы получим *коэффициент ковариации*. Когда требуется сравнить несколько выборок, то данные собирают в *таблицы корреляции или ковариации*:

	WORK_1	WORK_2	WORK_3	HOBBY_1	HOBBY_2
WORK_1	1,00	0,65	0,65	0,60	0,52
WORK_2	0,65	1,00	0,73	0,69	0,70
WORK_3	0,65	0,73	1,00	0,64	0,63
HOBBY_1	0,60	0,69	0,64	1,00	0,80
HOBBY_2	0,52	0,70	0,63	0,80	1,00

Корреляционная связь между признаками может осуществляться не непосредственно, а косвенно – за счет связи каждого из них в отдельности с каким-либо третьим (четвертым и т.д.) признаком. Например, размеры вегетативных органов обычно сильно коррелируют с высотой растения, и для изучения связи между ними в «чистом» виде необходимо найти способ исключить влияние на эту связь высоты растения.

Если рассчитаны парные коэффициенты корреляции r_{xy} , r_{xz} , r_{yz} между тремя признаками (x, y, z) , то исключить влияние признака z на связь между признаками x и y можно, рассчитав коэффициент частной корреляции:

$$r_{xy} = \frac{r_{xy} - r_{xz} \cdot r_{yz}}{\sqrt{(1 - r_{xz}^2)(1 - r_{yz}^2)}}.$$

В случае когда вы имеем дело с ранжированными данными, то есть по сути со значениями порядковой шкалы, то целесообразно использовать коэффициенты ранговой корреляции. Наиболее часто используются коэффициент Кенделла (τ) и коэффициент Спирмена (ρ):

Коэффициент ранговой корреляции Спирмена используется в случае когда определяется фактическая степень связи между двумя количественными рядами изучаемых признаков и дается оценка близости установленной связи с помощью количественно выраженного коэффициента. Практический расчет коэффициента ранговой корреляции Спирмена включает следующие этапы: 1) сопоставление для каждого из признаков его порядкового номера (ранга) по возрастанию (или убыванию); 2) определение разности рангов каждой пары сопоставляемых значений; 3) возведение в квадрат каждой разности и суммирование полученных результатов.

Вычисляется ρ -Спирмена по формуле:

$$\rho = 1 - \frac{6 \sum_{i=1}^m (r_i - s_i)^2}{n(n^2 - 1)}$$

где r_i – ранг среди ряда чисел (x_i, \dots, x_n); s_i – ранг среди ряда чисел (y_i, \dots, y_n); n – число парных наблюдений.

Тау-Кенделла определяется как:

$$\tau = 1 - \frac{4K}{n(n-1)},$$

где n – общее число рангов; K – число инверсий, т.е. перестановок элементов ряда s_i относительно упорядоченного r_i . Например, $r_i = 1, 2, 3, 4$, а $s_i = 3, 2, 1, 4$ ($m = 4$). Потребуется 3 инверсии: 3-2, 3-1, 2-1, чтобы сопоставить эти два ряда.

При использовании коэффициента ранговой корреляции условно оценивают тесноту связи между признаками, считая значения коэффициента равные 0,3 и менее, показателями слабой тесноты связи; значения более 0,4, но менее 0,7 - показателями умеренной тесноты связи, а значения 0,7 и более – показателями высокой тесноты связи.

Рассмотрим пример. Группу из 10 студентов протестировали двумя разными тестами. Рассчитаем коэффициент Спирмена:

<i>ранг по тесту A</i>	1	2	3	4	5	6	7	8	9	10
<i>ранг по тесту B</i>	2	1	3	4	9	8	10	5	7	6

Для нашего случая $\sum(r - s)^2 = 1 + 1 + 16 + 4 + 9 + 9 + 4 + 16$.
Следовательно:

$$\rho = 1 - \frac{6 \times 60}{10(100 - 1)} = 1 - \frac{360}{990} = 1 - \frac{4}{11} = \frac{7}{11}$$

Дисперсионный анализ

Что делать, когда мы хотим сравнить несколько выборок? Попарно сравнивать параметрическими или непараметрическими критериями? Очень быстро мы утонем в расчётах. Но, разумеется, наука уже знает способ нам помочь. Для сравнения трёх и более выборок используют дисперсионный анализ (ANOVA).

Дисперсионный анализ, основы которого были разработаны Фишером в 1920-1930 гг., позволяет устанавливать не только степень одновременного влияния на признак нескольких факторов и каждого в отдельности, но также их суммарное влияние в любых комбинациях и дополнительный эффект от сочетания разных факторов. Разумеется, и в этом случае остается масса неучтенных факторов, но, во-первых, методика позволяет оценить долю их влияния на общую изменчивость признака, а во-вторых, исследователь обычно имеет возможность выделить несколько ведущих факторов и исследовать именно их воздействие на изменчивость признаков.

Дисперсионный анализ позволяет решить множество задач, когда требуется изучить воздействие природных или искусственно создаваемых факторов на интересующий исследователя признак. Дисперсионный анализ принадлежит к числу довольно трудоемких биометрических методов, однако правильная организация опыта или сбора данных в природных условиях существенно облегчает вычисления.

Идея дисперсионного анализа заключается в разложении общей дисперсии случайной величины на независимые случайные слагаемые, каждое из которых характеризует влияние того или иного фактора или их взаимодействия. Последующее сравнение этих дисперсий позволяет оценить существенность влияния фактора на исследуемую величину. Таким образом, *задача дисперсионного анализа* состоит в том, чтобы выявить ту часть общей изменчивости признака, которая обусловлена воздействием учитываемых факторов, и оценить достоверность делаемого вывода.

Пусть, например, A – исследуемая величина, \bar{A} – среднее значение величины A , учитываемые факторы мы обозначим буквой x , неучитываемые – z , а все факторы вместе – буквой y (или припиской этих букв к соответствующим символам). Неучитываемые факторы составляют «шум» – помехи, мешающие выделить степень влияния учитываемых факторов. Отклонение A от \bar{A} при действии факторов x и z можно представить в виде суммы

$$(A - \bar{A}) = Y = X + Z,$$

где X – отклонение, вызываемое фактором x , Z – отклонение, вызываемое фактором z , Y – отклонение, вызываемое всеми факторами. Кроме того, предположим, что X, Y, Z – являются независимыми случайными величинами, обозначим дисперсии через σ^2_X , σ^2_Y , σ^2_Z , σ^2_A . Тогда имеет место равенство:

$$\sigma^2_A = \sigma^2_X + \sigma^2_Z.$$

Сравнивая дисперсии можно установить степень влияния факторов x и z на величину A , т.е. степень влияния учтенных и неучтенных факторов.

Непрерывным условием дисперсионного анализа является разбивка каждого учитываемого фактора не менее чем на две качественные или количественные градации. Если исследуется влияние одного фактора на исследуемую величину, то речь идет об однофакторном комплексе, если изучается влияние двух факторов – то о двухфакторном комплексе и т.д. Для проведения дисперсионного анализа обязательным условием является нормальное распределение и равные дисперсии совокупности случайных величин.

Для пояснения логической схемы дисперсионного анализа рассмотрим простейший произвольный пример. Предположим, что совокупности возрастающих доз удобрения на разных делянках имеют нормальное распределение и равные дисперсии. Имеется m таких совокупностей (разные делянки), из которых произведены выборки объемом n_1, n_2, \dots, n_m . Обозначим выборку из i -ой совокупности через $(x_{i1}, x_{i2}, \dots, x_{in})$ – урожайность делянок. Тогда все выборки можно записать в виде таблицы, которая называется матрицей наблюдений.

Матрица наблюдений однофакторного дисперсионного комплекса

		Количество элементов совокупности (n)-дозы удобрения					
		1	2	...	J	...	N
Количество совокупностей (m)	1	X_{11}	X_{12}	...	X_{1j}	...	X_{1n}
	2	X_{21}	X_{22}	...	X_{2j}	...	X_{2n}

	I	X_{i1}	X_{i2}	...	x_{ij}	...	x_{in}
	
	m	X_{m1}	X_{m2}	...	x_{mj}	...	x_{mn}

Средние этих выборок обозначим через $\bar{x}_1, \bar{x}_2, K, \bar{x}_i, K, \bar{x}_m$. Для проверки гипотезы о равенстве средних нулевую гипотезу запишем как

$$H_0 = \bar{x}_1 = \bar{x}_2 = K = \bar{x}_i = K = \bar{x}_m, \text{ альтернативную в виде}$$

$$H_0 \neq \bar{x}_1 \neq \bar{x}_2 \neq K \neq \bar{x}_i \neq K \neq \bar{x}_m.$$

Гипотеза H_0 проверяется сравнением внутригрупповых и межгрупповых дисперсий по F -критерию. Если расхождение между ними незначительно, то нулевая гипотеза принимается. В противном случае нулевая гипотеза отвергается и делается заключение о том, что различия в средних обусловлено не только случайностями выборок, но и действием исследуемого фактора.

Для изучаемого признака характерно три типа изменчивости:

1. *Факториальная (или групповая) изменчивость.* Характеризуется тем, что для каждой из совокупностей имеется своя средняя арифметическая (\bar{x}_{i*}). Разница в средних зависит, очевидно, от разного действия факторов;

2. *Остаточная изменчивость.* Характеризуется различными значениями признака внутри каждой градации. Эти различия не зависят от влияния фактора. Видимо, их причина лежит вне опыта, определяется неучитываемыми в данном анализе факторами.

3. *Общая изменчивость.* Заключается в том, что все наблюдения дисперсионного комплекса отличаются друг от друга (или иногда совпадают).

Мерой изменчивости признака в выборке служит сумма квадратов отклонений его значений от средней арифметической $\sum(x - \bar{x})^2$. Эта

величина, отнесенная к числу наблюдений, дает меру рассеяния, именуемую дисперсией, которая и применяется в дисперсионном анализе.

1. Мерой факториальной изменчивости будет сумма квадратов отклонений средних значений групп (\bar{x}_{i*}) от общего среднего

$$\bar{x} = \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n x_{ij}.$$

2. Мера остаточной изменчивости выразится суммой квадратов отклонений всех наблюдений в данной совокупности от среднего значения совокупности:

$$S_z^2 = \sum_{i=1}^m \sum_{j=1}^n (x_{ij} - \bar{x}_{i*})^2.$$

3. Мерой общей изменчивости является сумма квадратов отклонений в дисперсионном комплексе от общего среднего:

$$S_y^2 = \sum_{i=1}^m \sum_{j=1}^n (x_{ij} - \bar{x})^2.$$

Тогда в соответствии с основной идеей дисперсионного анализа можно записать: $S_y^2 = S_x^2 + S_z^2$ или:

$$S_y^2 = n \sum_{i=1}^m (\bar{x}_{i*} - \bar{x})^2 + \sum_{i=1}^m \sum_{j=1}^n (x_{ij} - \bar{x}_{i*})^2.$$

Вычислим факториальную и остаточную дисперсии, как меры соответствующих типов изменчивости признака в дисперсионном комплексе

$$\sigma_x^2 = \frac{S_x^2}{V_x}; \quad \sigma_z^2 = \frac{S_z^2}{V_z} \quad \sigma_y^2 = \frac{S_y^2}{V_y}.$$

В этих формулах фигурируют степени свободы (V_x , V_z , V_y), т.к. дисперсия σ^2 и есть сумма квадратов отклонений в расчете на одну степень свободы.

Число степеней свободы есть количество значений, необходимых для восстановления утраченного.

1. Число степеней свободы для факториальной дисперсии равно числу совокупностей без единицы ($m-1$), т.к. все группы связаны друг с другом лишь одним общим условием – значением средней арифметической всего дисперсионного комплекса (\bar{x}).

2. Число степеней свободы для остаточной дисперсии равно числу наблюдений в комплексе минус число совокупностей ($mn-m$) ибо все наблюдения связаны наличием в каждой группе своей средней арифметической (\bar{x}_{i*}).

3. Число степеней свободы для вычисления общей дисперсии всего комплекса равно числу наблюдений в комплексе без единицы ($mn-1$), ибо все наблюдения связаны только одним общим условием – наличием общей средней (\bar{x}).

Затем необходимо рассчитать доли влияния учтенного и неучтенного факторов как отношения соответствующих сумм квадратов отклонений:

$$\eta_x^2 = \frac{S_x^2}{S_y^2} ; \eta_z^2 = \frac{S_z^2}{S_y^2} .$$

Эти величины представляют собой не что иное, как квадраты корреляционных отношений. В сумме эти показатели должны всегда составлять 1 (100%). Теперь можно ответить на интересующий вопрос: насколько учитываемый фактор ответственен за изменчивость результативного признака и сколько процентов падает на долю неучтенных факторов.

Логическая схема однофакторного дисперсионного комплекса

Компоненты дисперсии	Сумма квадратов	Число степеней свободы	Дисперсии	Степень влияния фактора
Факториальная (межгрупповая)	$n \sum_i (\bar{x}_{i*} - \bar{x})^2$	$m-1$	$\frac{n \sum_i (\bar{x}_{i*} - \bar{x})^2}{m-1}$	$\eta_x^2 = \frac{S_x^2}{S_y^2}$
Остаточная (внутригрупповая)	$\sum_{i=1}^m \sum_{j=1}^n (x_{ij} - \bar{x}_{i*})^2$	$m(n-1)$	$\sum_{i=1}^m \sum_{j=1}^n (x_{ij} - \bar{x}_{i*})^2 \cdot \frac{1}{m(n-1)}$	$\eta_z^2 = \frac{S_z^2}{S_y^2}$
Полная (общая)	$\sum_{i=1}^m \sum_{j=1}^n (x_{ij} - \bar{x})^2$	$mn-1$	$\frac{1}{mn-1} \cdot \sum_{i=1}^m \sum_{j=1}^n (x_{ij} - \bar{x})^2$	

Для проверки достоверности полученного вывода необходимо- Определяют значение критерия Фишера (F), представляющего собой отношение двух д

и с п е р с и й - ф а к т о р и а л ь н о й и о с т а т о ч н о
й $F = \frac{\sigma_x^2}{\sigma_z^2}$ и сравнивают его с табличным в зависимости от числа степеней
свободы $\nu_1 = m - 1$ и $\nu_2 = mn - m$. Для того, чтобы отвергнуть нулевую
гипотезу, необходимо, чтобы полученное значение критерия было больше
табличного. Однофакторный дисперсионный анализ удобно представить в
виде таблицы.

Составитель курса:

Профессор кафедры прикладной

экологии, д. б. н.

_____ Горковенко Н.Е.