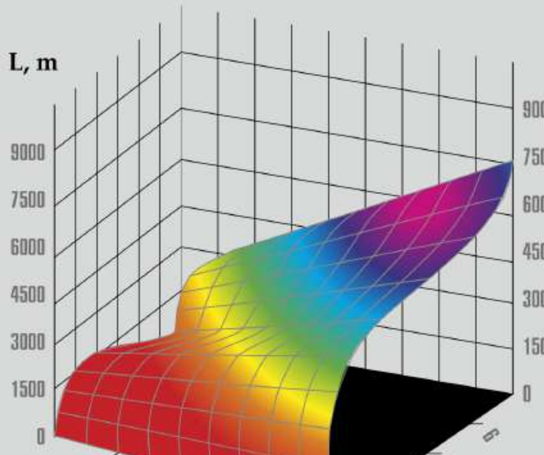
ПЛАНИРОВАНИЕ ЭКСПЕРИМЕНТОВ В ГИДРОБИОЛОГИЧЕСКИХ ИССЛЕДОВАНИЯХ

В. И. ХОЛОДОВ

В. И. ХОЛОДОВ

ПЛАНИРОВАНИЕ ЭКСПЕРИМЕНТОВ В ГИДРОБИОЛОГИЧЕСКИХ ИССЛЕДОВАНИЯХ



Севастополь 2016

РОССИЙСКАЯ АКАДЕМИЯ НАУК
ИНСТИТУТ МОРСКИХ БИОЛОГИЧЕСКИХ ИССЛЕДОВАНИЙ
им. А.О. КОВАЛЕВСКОГО

В.И. ХОЛОДОВ

ПЛАНИРОВАНИЕ ЭКСПЕРИМЕНТОВ В ГИДРОБИОЛОГИЧЕСКИХ ИССЛЕДОВАНИЯХ

Н.Оріанда
издательство
Симферополь
2016

УДК 574.5.001.4
ББК 28.082
Х 73

Рецензенты: д-р биол. наук И.В. Довгаль
д-р биол. наук С.Б. Криворотов
д-р биол. наук А.П. Золотницкий
канд. биол. наук А.Р. Болтачёв

Холодов В.И.
Х 73 **Планирование экспериментов в гидробиологических исследованиях.** – Симферополь: Н.Оріанда, 2016. – 196 с.; 13 ил., 80 табл., библ. 31.
ISBN 978-5-9909071-1-9

Руководство по организации и проведению экспериментальных и полевых гидробиологических и биологических исследований. Подробно на примерах изложена методика планирования активных экспериментов на основе регрессионного и дисперсионного анализов. Описаны методы пассивного эксперимента (полевых наблюдений).
Для научных сотрудников, аспирантов и студентов, специализирующихся в морской биологии, гидробиологии и биологии.

УДК 574.5.001.4
ББК 28.082
Х 73

Kholodov V.I.
Experimental design techniques in hydrobiology. – Simferopol: Н.Оріанда, 2016. – 196 pp.: 13 ill., 80 tabl., bibl. 31.
ISBN 978-5-9909071-1-9

Experimental design techniques and the related organizational principles are given for laboratory and field researches in biology and in hydrobiology. Using examples, the Guide offers a detailed methodology of the active experiment design based on regression analysis and on the analysis of variance. Methods supporting passive experiment, i.e., field observation, are also described.
The Guide is intended for scientists, postgraduates and university students having concern in marine biology, hydrobiology and biology.

Утверждено к печати Учёным советом
ФГБУН «Институт морских биологических исследований
им. А.О. Ковалевского РАН» (протокол от 24 июня 2016 г. № 7)

Ответственный редактор д-р биол. наук В.И. Рябушко

ISBN 978-5-9909071-1-9

СОДЕРЖАНИЕ

CONTENTS

ПРЕДИСЛОВИЕ

Исследователь, приступая к планированию или к выполнению экспериментальных работ, нередко сталкивается с многочисленными трудностями: сложностью изучаемого объекта (процесса), непостоянством его характеристик во времени, зависимостью этих характеристик не от одного, а от многих факторов, среди которых обнаруживаются не только количественные, но и неизмеряемые – качественные факторы. К тому же и природа самого объекта исследования изучена лишь частично. И тем не менее, проводить успешные исследования в такой запутанной ситуации можно, особенно, если исследователь вооружён методологией и методами планирования эксперимента. В число этих методов входят не только методы составления планов экспериментов, но и статистический анализ полученных результатов, их биологическая интерпретация и получение выводов, уточняющих планирование дальнейших исследований.

Дисциплина «Планирование эксперимента» (The Design of Experiment) возникла в 20-ые годы прошлого столетия, благодаря работам Р. Фишера, проводившего исследования по агробиологии. Так, изучая воздействие удобрений на рост растений, он обнаружил, что результат зависит от многих факторов (состава почвы, рельефа, влажности, предшественников, режима внесения удобрения и т.д.), что потребовало от исследователя разработки новых методов экспериментальных исследований. Так было положено начало созданию новой дисциплины, которая в последующие годы вошла в число наиболее быстро развивающихся отраслей науки. В настоящее время методы планирования эксперимента используются практически во всех отраслях промышленности, науки, экономики, а также и в космических исследованиях.

В предлагаемой автором книге рассмотрены методы, наиболее широко применяемые в научных исследованиях. Текст сопровождается конкретными примерами применения методов планирования в исследованиях по морской биологии, что значительно облегчает усвоение материала.

Кроме методов экспериментальных исследований (активных экспериментов), обсуждаются методы полевых исследований (наблюдений или пассивных экспериментов), которые также играют роль важнейших инструментов в гидробиологических исследованиях.

Книга является эффективным методическим руководством для исследователей, специализирующихся как в морской биологии и гидробиологии, так и для широкого круга биологов, а также будет полезной преподавателям биологических факультетов, студентам и аспирантам.

В.И. Рябушко, доктор биол. наук

Глава 1
ОСНОВЫ ПЛАНИРОВАНИЯ ЭКСПЕРИМЕНТА

Действительно ли необходимо планировать эксперименты или целесообразнее их проводить без всякого планирования? Очевидно, что планировать эксперименты (и не только в биологических исследованиях) нужно, прежде всего, потому что любой экспериментатор вольно или невольно в процессе обдумывания эксперимента уже начинает его планировать. Но, не владея методами планирования, он нередко проводит малоэффективные исследования, которые могут закончиться получением ошибочных результатов (артефактов) или формулированием неверных выводов. Типичной является ситуация, когда экспериментатор вначале ставит опыт и, получив результаты, начинает думать о том, как из этих данных получить важную информацию. То есть о статистических методах вспоминают только после получения результатов. После консультаций со специалистами выясняется, что опыт был спланирован неправильно, получены артефакты, а время и средства оказались просто потерянными.

Нетрудно привести множество примеров неудачных экспериментов, но для начала рассмотрим придуманную автором ситуацию, которая иллюстрирует важные положения теории эксперимента. Предположим, что три исследовательские лаборатории Минздрава получили задание: изучить влияние концентрации кислорода вдыхаемого воздуха на трудоспособность человека. Трудоспособность должна измеряться количественно. В данном случае это суммарный вес песка, который испытуемый в течение часа переносит вёдрами с некоторой нижней площадки на верхнюю. Уточним, что лаборатория № 1 расположена на берегу моря; лаборатория № 2 – под водой на глубине 15 м; лаборатория № 3 – в горах на высоте 4500 м. В результате проведенных исследований установлено:

- лаб. № 1 – повышение концентрации кислорода от минимальных значений (10 %) до некоторой величины (назовём её «порог насыщения») ведёт к повышению трудоспособности, после чего увеличение концентрации кислорода не оказывает влияния на трудоспособность;
- лаб. № 2 – при начальной концентрации кислорода (10 %) – трудоспособность близка к максимальной, и она незначительно увеличивается, а затем падает по мере увеличения концентрации кислорода, приводя к летальному исходу при поступлении в лёгкие чистого кислорода;

- лаб. №3 – трудоспособность минимальна при концентрации кислорода, равной 10%; с увеличением концентрации кислорода трудоспособность непрерывно возрастает и достигает максимума при вдыхании чистого кислорода.

Итак, получены разные, более того, противоречащие друг другу результаты. Дальнейшие более глубокие исследования сняли эти противоречия. Выяснилось, что на трудоспособность влияет не концентрация кислорода во вдыхаемом воздухе, а концентрация кислорода, растворённого в артериальной крови, то есть исследовать следовало влияние другого фактора. Концентрация кислорода в крови зависит от атмосферного давления: чем выше давление, тем выше концентрация кислорода в крови. Казалось бы, найдено окончательное решение задачи. Но исследователей смущает неудовлетворительная воспроизводимость результатов, которая может быть вызвана влиянием на трудоспособность других, пока ещё не установленных факторов. В дальнейших исследованиях было обнаружено, что на изучаемую зависимость оказывают влияние такие факторы, как возраст испытуемого, его пол, состояние здоровья, температура окружающей среды. При учёте всех перечисленных факторов результаты воспроизводятся с минимальной ошибкой.

Из данного примера следуют важные выводы:

1. При изучении влияния одного фактора на изучаемый процесс или явление необходимо помнить, что на измеряемую величину могут оказывать влияние и другие факторы. Поэтому, до постановки эксперимента необходимо провести так называемый «анализ априорной информации», одной из задач которого является составление списка факторов, потенциально способных оказывать влияние на изучаемый процесс. Факторы могут быть как количественные (измеряемые количественной мерой), так и качественные (пол животного, стадии развития организма, место взятия пробы, тип экспериментальных установок и т.д.). При изучении однофакторной зависимости значения остальных факторов должны быть стабильными – это условия проведения эксперимента. Полученная зависимость будет частным случаем, соответствующим условиям проведения эксперимента.
2. Изучаемый процесс (явление, характеристика) должен быть также тщательно проанализирован и должна быть выбрана величина, количественно характеризующая изучаемый процесс. Эта величина должна иметь чёткий биологический (или физический) смысл и легко и надёжно измеряться.
3. На процесс, кроме влияний отдельных факторов, могут воздействовать и межфакторные взаимодействия. Например,

характер влияния концентрации кислорода воздуха на трудоспособность человека зависит от величины атмосферного давления. В этом случае говорят о взаимодействии обоих факторов: концентрации кислорода и внешнего давления. Влияние эффекта взаимодействия факторов может быть оценено количественно.

4. Целесообразно сразу поставить многофакторный эксперимент (одновременное варьирование нескольких факторов), который позволит оценить влияние каждого фактора отдельно, а также влияние межфакторных взаимодействий и получить при этом универсальную зависимость, пригодную для разных условий.

Как организовать экспериментальное исследование, какова структура эксперимента и какие бывают эксперименты? Данная книга, прежде всего, предназначена для молодых специалистов, аспирантов и студентов. Поэтому ниже описаны ситуации, возможные для учебного процесса в вузе.

Предположим, что сейчас начало лета и, следовательно, начало практики, во время которой студентам предстоит выполнить в исследовательском институте следующие задания: 1) Исследовать зависимость роста одноклеточных водорослей от интенсивности освещённости; 2) Изучить влияние интенсивности перемешивания воды на выживаемость личинок устриц. В лаборатории завода железобетонных изделий (ЖБИ) требуется исследовать: 3) Зависимость прочности бетона от основных компонентов (цемента, песка); 4) Найти такое сочетание компонентов бетона, при котором его прочность максимальна.

При обдумывании перечисленных задач возникает множество вопросов. Например, в задаче № 1 не ясно, что конкретно необходимо измерять: увеличение численности микроводорослей, увеличение их биомассы, удельную скорость или валовую скорость увеличения численности (биомассы); фотосинтез, первичную продукцию, то есть скорость образования органического вещества? Как конкретно задавать уровни освещённости? Ведь на разном удалении от источника света изменяется и температура. Может, подключить вентилятор? Но есть и другие факторы, влияющие на рост микроводорослей (состав питательной среды, исходная концентрация клеток, объём и форма сосуда, интенсивность перемешивания, возраст культуры и т.д.). Может быть, прежде чем ставить опыт, надо составить список факторов, способных влиять на рост водорослей? Известно, что в процессе опыта культура микроводорослей изменяет свои свойства (развивается, стареет), что в теории эксперимента называют «дрейфом объекта». Как быть в такой сложной ситуации: всё бросить или искать выход?

Задача № 2. На выживаемость личинок устриц оказывают влияние множество факторов: качество воды (источник отбора воды), водоподготовка (отстаивание, фильтрация, стерилизация, продувка воды воздухом), видовой состав и концентрация кормовых микроводорослей; температура, частота смены воды, концентрация личинок в баке, объём бака, форма бака, освещённость и т.д. Кроме этого, сам объект также изменяется во времени, так как личиночный период включает разные стадии. В таком случае, почему ставится задача провести только однофакторный эксперимент? Причины:

1. Ограничены технические возможности.
2. Руководитель и вы не владеете методами проведения многофакторных экспериментов.
3. Нужна именно данная конкретная однофакторная зависимость.

Следует уточнить, что и для получения однофакторной зависимости эксперимент также необходимо планировать, чтобы избежать артефактов или данных, которые невозможно будет использовать как в практической деятельности, так и в теоретических разработках.

Задачи № 3 и № 4. Эти формулировки также требуют обсуждения и уточнения. Действительно, цемент бывает разных марок; он также может быть свежим и старым. Песок бывает морской и речной; крупный и мелкий и т.д. Влияет также и температура воды, и её начальное количество, а также режим последующего увлажнения бетона, перемешивание смеси и её объём и т.д. Так, что для начала нужно изучить специальную литературу (выполнить анализ априорной информации), составить список факторов и возникших вопросов и снова всё обсудить с руководителем. Но здесь мы обращаем внимание читателей на два обстоятельства:

- В данном случае нам предстоит ставить двухфакторные эксперименты, а это шаг вперёд по сравнению с классическим однофакторным подходом, применённым для решения задач № 1 и № 2.
- В задаче № 3 требуется определить только зависимость одной (зависимой) переменной от двух независимых переменных (что нужно для, например, выполнения промежуточных расчётов, то есть интерполяции), а № 4 – это типичная задача оптимизации – для её решения нужно планировать экстремальный эксперимент.

Имеются ли аналогичные элементы во всех 4-х задачах? Можно выделить объект исследования (ОИ), факторы (независимые переменные), влияющие на ОИ; выходную величину или зависимую переменную. Если ОИ изменяет свои свойства с течением времени (во время проведения эксперимента), тогда нужно выделить и дрейф ОИ.

Структуру эксперимента можно изобразить в виде схемы (рис. 1). На схеме ОИ представлен прямоугольником, так называемым «чёрным ящиком», на который воздействуют различные факторы (X, Z, V). Выходной величиной является зависимая переменная Y. Термин «чёрный ящик» взят из кибернетики. Этим термином обозначают ОИ, внутреннее строение которых неизвестно, а свойства ОИ изучают путём воздействия на ОИ управляемыми факторами, уровни (значения) которых можно задавать. Совокупность регулируемых факторов обозначают буквой **X.** Управляемый фактор может быть один **(x)**, либо несколько $(\mathbf{x_1}, \mathbf{x_2}, \mathbf{x_3}, \ldots \mathbf{x_n})$. Воздействуя на объект исследования факторами **X,** экспериментатор следит, точнее, измеряет реакцию ОИ, обозначаемую как **Y.** В теории эксперимента нет однозначного названия этой реакции. Чаще всего её называют зависимая переменная, параметр оптимизации (ПО), отклик, критерий оптимизации, целевая функция, выход чёрного ящика. Получив различные значения **Y** при разных сочетаниях уровней факторов **X,** экспериментатор рассчитывает уравнение зависимости **Y** от **X.** Таблица с набором различных значений (различных уровней) факторов называется планом факторного эксперимента. Руководствуясь этим планом, экспериментатор проводит эксперимент.

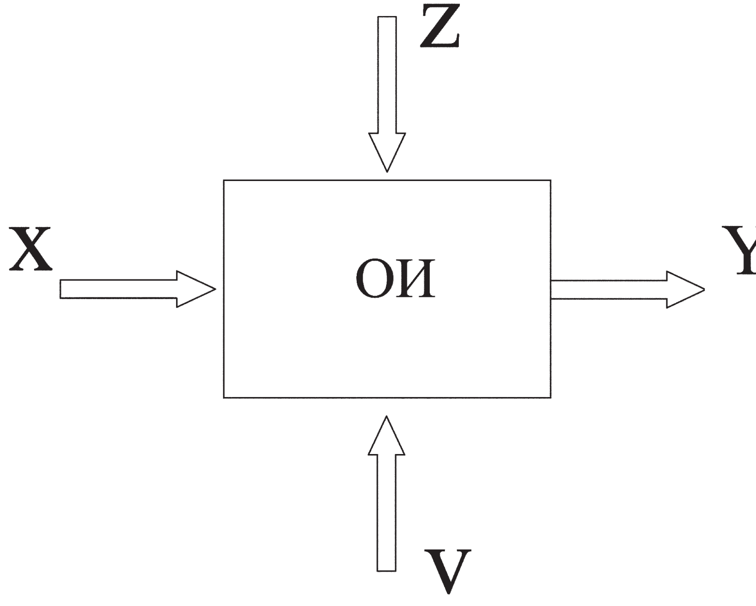


Рис. 1. Схема экспериментального исследования с применением методологии чёрного ящика (объекта исследования ОИ). X – регулируемые (управляемые) факторы; Z – контролируемые, но не регулируемые факторы; V – не контролируемые факторы; Y – выход чёрного ящика (параметр оптимизации).

К сожалению, реальность сложнее описанной упрощённой процедуры. На ОИ действуют ещё и факторы группы **Z** – контролируемые, но не регулируемые факторы, например, температура, дневное освещение, влажность в лаборатории, в которой не установлен кондиционер, атмосферное давление и т.д. Эти факторы изменяются произвольно и могут повлиять на результаты эксперимента. И, тем не менее, существуют методы планирования эксперимента, позволяющие составить план эксперимента, результаты которого будут защищены от искажающих воздействий неуправляемых факторов.

Третья группа факторов – **V,** в которую входят не контролируемые экспериментатором факторы, имеющие случайный и непредсказуемый характер. Об этих факторах мы, как правило, ничего не знаем, например: возможные магнитные бури в атмосфере, возможное воздействие Луны на ОИ, электрическое поле электропроводки в лаборатории, ультра- и инфразвуки и т.д. Эти факторы увеличивают ошибку измерений результатов эксперимента. Но в ряде случаев удаётся защититься и от действия неконтролируемых факторов.

В биологических экспериментах обычно роль объектов исследования играют живые структуры, свойства которых изменяются в течение времени. В непродолжительных экспериментах изменения ОИ (дрейф) можно не учитывать. Однако, если продолжительность эксперимента велика, необходимо предварительно изучить характер дрейфа и, зная функциональную зависимость параметра оптимизации от времени, строить план эксперимента таким образом, чтобы результаты эксперимента были защищены от искажающего действия дрейфа. Если же мы не располагаем математическим описанием дрейфа, в этом случае применяются другие методы защиты результатов эксперимента, о чём будет говориться в последующих разделах.

Итак, основные компоненты эксперимента следующие: объект исследования ОИ, параметр оптимизации ПО, факторы. Рассмотрим подробнее каждый из этих компонентов.

1.1. Объекты исследования, параметры оптимизации, факторы

Живые объекты исследования отличаются высокой сложностью, динамичностью своих структур, нестабильностью характеристик и т.д. В биологических структурах непрерывно протекает процесс деструкции, который компенсируется процессами синтеза. Исследуемые живые организмы – это результаты двух одновременно протекающих процессов – диссимиляции и ассимиляции. Каждый из этих процессов не только подвержен влиянию многих внутренних и внешних факторов, но оба процесса связаны между собой прямыми и обратными связями, как непосредственными, так и опосредованными, реализуемыми через метаболиты, другие организмы и т.д. Механизмы функционирования биохимических, физиологических и экологических процессов нам в лучшем случае известны лишь частично. Однако мы проводим исследования поведения, других характеристик объектов, о строении которых располагаем лишь частичной информацией. В такой ситуации наиболее приемлемой методологией является методология «чёрного ящика». Как было сказано выше, суть её заключается в следующем: воздействуя на неизученный объект исследования, мы измеряем ответные реакции на каждое воздействие. Затем рассчитываем уравнение зависимости реакции от воздействующих факторов. Анализируя полученное уравнение, мы оцениваем роль каждого фактора в формировании величины реакции. В результате у нас складываются представления о закономерностях функционирования ОИ, что даёт нам возможность формулирования гипотез о, например, механизмах функционирования ОИ. Для проверки гипотез мы планируем новые исследования и постановку конкретных экспериментов и т.д., осуществляя таким образом процесс исследования.

1.1.1. Объекты исследования (ОИ) и параметры оптимизации (ПО)

Упрощая ситуацию, можно утверждать, что объект исследования – это то, что мы исследуем, а параметр оптимизации – это то, что мы измеряем. Казалось бы, всё просто и ясно. Однако возникают серьёзные трудности, когда мы задаёмся вопросом: «А что мы конкретно исследуем и что измеряем?» В обычных исследованиях экспериментатор редко задаётся подобными вопросами, так как считается, что это известно и без специального обдумывания. Но при использовании математических методов планирования экспериментов необходимо иметь чёткие ответы на подобные вопросы – иначе с самого начала исследований можно сориентироваться в неверном направлении. Планирование экспериментов требует от экспериментатора более продуманного отношения к исследованию и более чёткого представления и максимально точного формулирования решаемой задачи. Рассмотрим эту проблему на конкретном примере.

Допустим, что мы решили провести исследование роста определённого вида одноклеточных водорослей, выращиваемых в накопительной (непроточной) культуре. Эксперименты будем проводить, допустим, в колбах объёмом 1 л. Что здесь следует считать ОИ: 1) колбу с микроводорослями; 2) культуру микрофитов; 3) процесс роста культуры водорослей; 4) процесс увеличения численности водорослей в накопительной культуре; 5) процесс деления клеток микроводорослей? Если сразу затруднительно ответить на этот вопрос, тогда его надо рассмотреть параллельно с другим вопросом: «Каков должен быть параметр оптимизации?» Нам известно, что процесс увеличения численности клеток в культуре принято характеризовать удельной скоростью роста водорослей, которая в действительности отражает скорость увеличения численности клеток в пересчёте на одну клетку. Численно эта величина равна количеству делений клетки в единицу времени, чаще всего в сутки и рассчитывается по формуле:

$$\mu = \frac{\lg C_t - \lg C_0}{t} \times \lg 2,$$

где: C_0 и C_t – концентрация клеток в начальный момент времени и концентрация клеток через промежуток времени t.

Как следует из формулы, измерять необходимо концентрации клеток в начале и в конце эксперимента, а также продолжительность самого эксперимента. В данном случае параметр оптимизации $\boldsymbol{\mu}$ непосредственно не измеряется, а рассчитывается по другим измеряемым в эксперименте величинам.

Отправной точкой планирования эксперимента является чёткое формулирование целей, причём целей, для которых возможна количественная оценка. В данном случае цель – получить уравнение зависимости величины удельной скорости роста от наиболее существенных факторов среды. Объект исследования – процесс изменения во времени концентрации клеток в накопительной культуре. Точная формулировка ПО и ОИ позволяет сузить объём анализируемой по данному вопросу информации, необходимой для составления списка факторов, способных оказывать влияние на ПО и ОИ.

Итак, ОИ – это кинетика роста численности микроводорослей в накопительной культуре, а ПО – это удельная скорость роста численности культуры в фазе логарифмического роста.

Рассмотрим другой пример: измерение удельных скоростей выделения двустворчатыми моллюсками мидиями азот- и фосфорсодержащих соединений в растворённой форме. Для решения этой задачи предполагается мидий содержать в замкнутых сосудах (без протока) с продувкой воды воздухом. Через определённые интервалы времени будут отбираться пробы воды для выполнения гидрохимических анализов.

С самого начала необходимо чётко определить ОИ, без чего невозможно экспериментальное исследование. Итак, что мы изучаем? Ответ: «мы изучаем мидию» – не корректен ввиду его неконкретности. Другой ответ: «изучаем систему мидия – вода» – тоже неприемлем, так как из него не следует, что конкретно надо измерять и как эти измерения будут выполняться. Гораздо лучше ответ: «ОИ – это **процесс** выделения мидиями растворённых форм NO_3^-, NH_4^+, PO_3^{2-}». В данном случае ясно, что мы изучаем процессы, которые можно охарактеризовать скоростями выделений растворённых веществ, а также составить список факторов, способных влиять на величины этих скоростей. Но, к сожалению, здесь возникают трудности. В действительности процессы выделения растворённых веществ мидиями мы не изучаем, ведь для этого нужно изучать конкретные физиологические механизмы. А в нашу задачу это не входит. Нам важно определить скорости увеличения концентраций (либо уменьшения) перечисленных соединений в воде, в которой находятся мидии. Эти соединения могут выделяться непосредственно мидиями, но могут и вымываться из фекалий мидий; также и микрофлора, которая находится в морской воде, на фекалиях и на створках мидий, может оказывать влияние на концентрации этих соединений и т.д.

Как быть в такой непростой ситуации? Очевидно, что здесь можно рассматривать, по крайней мере, два варианта:

1. Считать, что мы проводим чисто экологическое исследование, поэтому нас интересуют **суммарные скорости поступления** в морскую воду изучаемых соединений в присутствии мидий. Наверное, такая точка зрения справедлива, если мы свои данные будем использовать в расчётах по определению влияний мидий на среду. Но в физиологических исследованиях, при изучении, например, энергобаланса придётся делать оговорки, допущения или вообще не применять для этих случаев наши результаты.
2. Считать, что мы определяем **скорости выделения мидиями** в воду растворённых соединений. Тогда в контрольный сосуд нужно вносить фекалии и створки мидий с микрофлорой, а полученную в контроле поправку вычитать из величин, полученных по измерениям в экспериментальных сосудах.

Возможно, что имеются и другие варианты. Окончательное определение ОИ можно принять после анализа различных вариантов ПО.

1.1.2. Параметр оптимизации

В данной задаче ПО может быть различным и зависеть от выбранных единиц измерения, так как вес мидий можно измерять как общий сырой, общий сухой, сырой вес мягких тканей, их сухой вес, вес органического углерода мягких тканей и т.д. Как выбрать нужный ПО? Ответ зависит от того, где и как мы собираемся применять результаты, и для каких целей мы проводим исследования? Следует пояснить, что мы намерены рассчитывать влияние на экосистему мидийных ферм, на которых выращиваются мидии с определённой суммарной биомассой, например, 100 тонн или 1000 тонн и т.д. В таком случае вес мидий должен быть общим и сырым, так как нам неизвестен вес сухих мягких тканей. Для удобства применения результатов исследований в расчётах, скорости должны быть удельными, то есть скорости должны относиться к единице веса, включая вес раковины. С физиологической точки зрения такое утверждение не является корректным, так как раковина представляет собой метаболически почти инертную структуру и, если мы делим (соотносим) физиологический показатель на активные и неактивные структуры, то мы исходим из неверного предположения, что и те, и другие метаболически одинаково активны. В этом случае можно прийти к неверным выводам. Например, мы определяем удельные скорости потребления кислорода моллюсками, обитающими у открытого берега (шторма – раковины массивные) и в защищённой бухте (вода спокойная – раковины тонкие). При сравнении метаболизма мидий из обоих поселений придём к неверному выводу о замедлении метаболизма у моллюсков открытых берегов. Хотя в действительности метаболизм у них не заторможен, но у них более тяжёлые раковины.

Однако при использовании этих данных только в тех случаях, для которых они предназначены, следует выбрать ПО, характеризующийся удельными скоростями выделения соединений мидиями. При этом вес мидий – общий, сырой (табл. 1, ПО №2). Здесь, как и в первом примере, ПО определяется расчётным путём на основе измерений в течение опыта концентраций исследуемых соединений.

Таблица 1. Характеристика параметров оптимизации.

№ ПО	Название	Размерность	Область определения	Оптимальное значение	Ошибка	Примечание
1	Выживаемость	%	0 – 100	100	1	За 4 сут.
2	Уд. скорость выделения	мкг/ г·час	-0,1 – +5,0	нет	0,03	Сыр. вес мидий

Результаты предварительного эксперимента свидетельствуют о линейности во времени (на протяжении 6 час) процесса увеличения концентрации растворённых форм NO_3^-, NH_4^+, PO_3^{2-} в сосуде с мидиями. Иными словами, эти соединения выделяются в воду с постоянными скоростями, что упрощает расчёт скоростей выделения. Удельные скорости выделения веществ находили по формуле:

$$\textbf{Уд. скор. (мкг/г·час)} = (\mathbf{V}\cdot(\mathbf{C}_{\text{кон}} - \mathbf{C}_0) - \mathbf{V}_{\text{контр}}\cdot(\mathbf{C}_{\text{контроль}} - \mathbf{C}_0))/\mathbf{T}\cdot\mathbf{W},$$

где: $\mathbf{V}$ – объём экспериментальных сосудов, л;
$\mathbf{V}_{\text{контр}}$ – объём контрольного сосуда, л;
$\mathbf{T}$ – продолжительность эксперимента, час;
$\mathbf{W}$ – общий сырой вес мидий, г;
$\mathbf{C}_0$ – концентрация соединений в начале эксперимента, мкг/л;
$\mathbf{C}_{\text{кон}}$ – концентрация соединений в конце эксперимента, мкг/л;
$\mathbf{C}_{\text{контроль}}$ – концентрация соединений в контрольном сосуде в конце эксперимента, мкг/л.

Учитывая, что объёмы каждого сосуда равны 3 л, а продолжительность эксперимента равна 6 час, записываем формулу в окончательном виде (мкг/л·час):

Уд. скор. $= 3(C_{кон} - C_0) - (C_{контроль} - C_0)/6W = ((C_{кон} - C_0) - (C_{контроль} - C_0))/2W = (C_{кон} - C_{контроль})/2W$

Данный параметр оптимизации (Уд. скор $= (C_{кон} - C_{контроль})/2W$) используется далее в разделе о полном факторном эксперименте при обсуждении применения полного факторного эксперимента ПФЭ 2^2 в исследовании процесса выделения растворённых веществ мидиями (стр. 45).

Параметр оптимизации отражает одну из многочисленных характеристик ОИ. При изучении биологических ОИ параметры оптимизации могут характеризовать различные особенности: рост, выживаемость, потребление (выделение) различных соединений, синтез разных соединений, плодовитость, интенсивность нереста, содержание в тканях определённых биохимических компонентов, величину надоя и т.п.

В промышленных экспериментах, например, в исследованиях по биотехнологии, выделяют следующие типы ПО:

1. Экономические: прибыль, рентабельность, затраты на постановку эксперимента.
2. Технико-экономические: производительность, надёжность, КПД, стабильность.
3. Технико-технологические: долговечность, выход продукта, физико-химические характеристики продукта, медико-биологические характеристики.
4. Прочие: психологические, эстетические, статистические.

Иногда достаточно полно описать изучаемый процесс возможно только при использовании нескольких ПО, имеющих различную природу: биологическую, экономическую, технологическую, гигиеническую, эстетическую и т.д. Но эксперимент возможен только при выборе единственного ПО. Остальные параметры оптимизации могут играть роль ограничений, которые вводят при планировании эксперимента.

Существует и другое решение – составление сложного ПО, включающего в качестве компонентов два или более простых ПО с различными единицами измерения. Для объединения таких параметров необходимо принять единую шкалу, независимую от конкретных единиц измерения, иными словами, необходимо стандартизировать исходные ПО. Эту операцию выполняют с помощью метода, применяемого в многомерной статистике, а также в методах математического планирования многофакторных экспериментов. С этой целью для каждого параметра оптимизации $\mathbf{y_u}$ вводят, на основе имеющегося опыта (анализа априорной информации), наилучшее его значение $\mathbf{y_{u0}}$. Разность $\mathbf{y_{ui}}$ – $\mathbf{y_{u0}}$ есть расстояние текущего значения параметра от его оптимального значения. Для освобождения от единиц измерения нужно разделить эту разность на оптимальное значение; $(y_{ui} - y_{u0})/y_{u0}$. Объединение различных ПО выполняют суммированием квадратов их стандартизированных значений $\mathbf{Y=\Sigma((y_{ui}-y_{u0})/y_{u0})^2}$. Чем ближе текущее значение ПО (то есть $\mathbf{y_{ui}}$) к наилучшему значению ($\mathbf{y_{u0}}$), тем ближе значение ПО к 0. Поэтому в оптимальных условиях стандартизированный ПО стремится к 0.

Если исходные параметры оптимизации имеют разную ценность, в этом случае вводят соответствующие весовые коэффициенты $\mathbf{a_u}$, $\mathbf{(0 < a_u \leq 1)}$; $\mathbf{Y=\Sigma(a_u(y_{ui}-y_{u0})/y_{u0})^2}$. Проиллюстрируем составление сложного ПО на примере экспериментального исследования роста и выживаемости личинок мидий на поздних стадиях развития (педивелигеры) в питомнике нашего института. Продолжительность опыта 4 сут, параметры оптимизации:

$ПО_1$ – линейная скорость роста (мкм/сут);

$ПО_2$ – выживаемость за 4 суток (%);

ПО – сложный параметр оптимизации, объединяющий скорость линейного роста и выживаемость.

Таблица 2. Простые ($ПО_1$ и $ПО_2$) и сложный (ПО) параметры оптимизации экспериментального исследования роста и выживаемости личинок мидий *Mytilus galloprovincialis.*

№ опыта	$ПО_1$ n	$ПО_2$ m	$ПО_1$ $(y_{ni} - 17{,}8)/17{,}8$	$ПО_2$ $(y_{mi} - 100)/100$	$ПО=(ПО_1)^2+(ПО_2)^2$ $\Sigma((y_{ui}-y_{u0})/y_{u0})^2$.
1	6,2	100	-0,652	0	0,4251
2	17,8	100	0	0	0
3	5,5	97	-0,69	-0,03	0,4779
4	10,9	93	-0,387	-0,07	0,1549
5	5,9	94	-0,668	-0,06	0,4502
6	12,4	91	-0,303	-0,09	0,1001
7	7,0	91	-0.607	-0,09	0.4494
8	10,2	93	-0.427	-0,07	0,1872

В табл. 2 в первом столбце указаны номера экспериментальных сосудов, в которых находятся личинки мидий. Для каждого сосуда заданы свои условия эксперимента, поэтому каждый сосуд – это отдельный опыт со своим номером, соответствующим номеру сосуда. В столбце $ПО_1$ указаны значения скоростей роста (мкм/сут), полученные в каждом опыте; параметру оптимизации присвоен индекс n, то есть u=n. Столбец для $ПО_2$ (индекс u=m) содержит данные по выживаемости личинок, выраженные в %. В следующих двух столбцах приведены безразмерные стандартизированные значения параметров оптимизации, рассчитанные по формулам, указанным в заголовках столбцов. Цифры 17,8 и 100 соответствуют полученным наилучшим значениям каждого параметра оптимизации. В последней колонке содержатся значения сложного параметра оптимизации.

Наилучшие условия для роста личинок были заданы в опыте № 2, а наихудшие – в опыте № 3. Выживаемость максимальна в опытах № 1 и № 2, а минимальна – в опытах № 6 и № 7. Если принимать в расчёт одновременно оба ПО, тогда для выращивания личинок необходимо обеспечить условия опыта № 2, в котором сложный параметр оптимизации принимает минимальное значение.

При выборе ПО следует проверить соответствие этого ПО следующим требованиям, предъявляемым к параметрам оптимизации (Адлер и др., 1971):

- ПО должен быть количественным (характеризоваться числом, либо числовым рангом).
- Должен быть однозначным (любому набору уровней факторов должен соответствовать только один результат).
- Должен действительно оценивать функционирование ОИ, то есть быть эффективным.
- Быть эффективным и в статистическом смысле (определяться с максимально возможной точностью).
- Отвечать требованию универсальности или полноты (всесторонне характеризовать объект, например, кроме требований технологии и экономики отвечать и требованиям санитарии, эстетики и т.д.).
- Иметь чёткий биологический или физический смысл, что важно при интерпретации результатов

Рассмотрим требования к ПО подробнее. Быть количественным параметром означает возможность количественного измерения ПО в любых экспериментах, то есть всегда значению ПО должно соответствовать определённое число. В реальной ситуации это число может изменяться от некоторого минимального значения до максимального. Диапазон возможных значений ПО называется областью определения. В таблице 2 для выживаемости область определения указана от 0 (погибли все личинки) до 100 % (выжили все). Область определения для скоростей выделения мидиями азот- и фосфорсодержащих растворённых веществ найдена на основе анализа априорной информации: от -0,1 до 5,0 мкг/г·час. Знак минус у нижней границы диапазона говорит о возможном потреблении растворённых веществ (предположительно микрофлорой мидий).

Область определения ПО задаёт множество различных значений, которые может принимать ПО. Но значения могут быть непрерывными (скорости выделения), либо дискретными (количество личинок в 1 мл, количество колоний микроорганизмов в чашке Петри). Обычно область определения ограничена, но может быть и неограниченной, когда значения ПО изменяются от $-\infty$ до $+\infty$, что может наблюдаться для расчётных ПО. Если не удаётся найти требуемый ПО, который можно измерить непосредственно прибором или рассчитать его по формуле, в таком случае можно предложить ранговый ПО, оцениваемый в баллах или в рангах. Например, в баллах можно оценить качество обуви, вина, устриц, блюда, выращенной продукции и т.д. Ранговый ПО также характеризует количественно результаты экспериментальных исследований. Ранговые ПО менее точны обычных ПО но, тем не менее, они широко применяются при планировании многофакторных экспериментов, особенно в опытах с участием экспертов (экспертные оценки).

Требование однозначности: ПО всегда определяется одним числом (величиной прироста, количеством съеденного корма, выходом продукции биотехнологического процесса и т.д.). Если каждому набору условий эксперимента соответствует одно значение ПО – это говорит об однозначности выбранного ПО.

ПО считается эффективным в статистическом смысле, если он определяется с наибольшей возможной точностью. Если ПО рассчитывается по значениям некоторых исходных величин, измеряемых с невысокой точностью, он не будет эффективным в статистическом смысле и от него следует отказаться.

Чёткий биологический или физический смысл выбранного ПО необходим для интерпретации результатов экспериментального исследования. Планируя эксперимент, надо заранее думать о возможных вариантах интерпретации результатов, которые будут получены и о том, где, кем и как эти результаты будут использоваться.

1.1.3. Факторы

Факторы – это всё то, что воздействует на объект исследования и, следовательно, отражается на параметре оптимизации. В контролируемых условиях лаборатории экспериментатор сам, через факторы, воздействует на ОИ, для чего он устанавливает выбранные им факторы на определённых **уровнях.** Математически факторы представляются в виде независимых переменных $\mathbf{x_1, x_2, x_3}$**...** от которых зависят значения ПО, называемые зависимыми переменными $\mathbf{y_1, y_2, y_3}$**...**

Факторы должны быть независимыми друг от друга и управляемыми. Например, если в эксперименте при культивировании микроводорослей увеличивают освещённость культур водорослей, но при этом повышается температура среды – факторы освещенность и температура – зависимы. Аналогично, если при увеличении плотности посадки животных в экспериментальном аквариуме, одновременно уменьшается запас кислорода, приходящийся на единицу живой массы, – в этом случае оба фактора зависимы, а их независимое воздействие на ОИ не может быть изучено на данной экспериментальной установке. Поэтому нужно искать другие технические возможности независимого управления этими факторами. Схожая проблема возникает при попытках получения независимых оценок влияния веса животного и его возраста на метаболизм. Действительно, редко предоставляется возможность независимого варьирования уровней таких факторов как возраст и индивидуальный вес животного. Исключение составляют человек, домашние животные, а также дикие животные, возраст которых можно достаточно точно определить по внешним признакам.

Большое значение имеет и управляемость фактора. Это означает, что для фактора имеется возможность установления его на любом уровне, причём уровень фактора остаётся стабильным на протяжении всего эксперимента независимо от уровней других факторов.

Для каждого фактора должна обеспечиваться высокая точность его измерения. Это необходимо для получения воспроизводимых результатов, ошибки которых должны быть минимальными, а также для получения точных оценок влияний каждого фактора на ОИ.

Фактор должен быть однозначным, иными словами, он не должен быть функцией других факторов. Например, прозрачность воды в культиваторе для микроводорослей не может считаться фактором, так как прозрачность воды зависит от других факторов (концентрации клеток, особенностей перемешивания культуры, возраста культуры, выделения микроводорослями органических веществ и т.д.).

Факторы должны быть совместимыми, что означает, что все комбинации уровней всех факторов возможны и безопасны (не приведут к воспламенению, взрыву, гибели или стрессу подопытных организмов).

По своей природе факторы различаются на две группы: количественные факторы и качественные факторы. Количественные факторы измеряются количественной мерой (температура, объём, концентрация, вес, скорость, давление, освещённость и т.д.). Уровни качественных факторов нельзя задать числом. Например, уровни качественного фактора «тип животного» таковы: окраска раковины; неполовозрелые, половозрелые самцы, половозрелые самки, гермафродиты. Качественный фактор «экстрагент» может задаваться на таких уровнях: ацетон, хлороформ, бензол, этанол, метанол и т.д. Уровни фактора «тип сырья» представляют собой макрофиты, доставленные из различных районов моря. Фактор «способ лечения» представлен уровнями: гомеопатия, физиотерапия, медикаментозное лечение. Фактор «форма экспериментального сосуда» может быть задан уровнями: цилиндр, шар (круглая колба), конус (коническая колба), прямоугольный аквариум и т.д. (табл. 3).

Уровни количественного фактора принято откладывать на числовой горизонтальной оси (X). При проведении однофакторного эксперимента, экспериментатор задаёт единственный фактор $\mathbf{x_1}$ на нескольких уровнях $(\mathbf{x_{1,1}}; \mathbf{x_{1,2}}; \mathbf{x_{1,3}}\ldots)$ и соответственно получает для каждого случая значения ПО $(\mathbf{y_1}, \mathbf{y_2}, \mathbf{y_3}\ldots)$.

Результаты эксперимента позволяют рассчитать уравнение зависимости **y** от **x**, о чём подробно будет говориться ниже. Аналогично рассчитывается уравнение регрессии **y** от многих факторов. Поэтому для экспериментов с количественными факторами применяют экспериментальные планы регрессионного анализа.

Таблица 3. Пример задания факторов (из эксперимента по изучению роста устричной молоди после оседания на субстрат)

№	Наименование фактора	Единица измерения	Ошибка, точность	Область определения	Область интереса	Примечание
1	Объём экспериментального сосуда	л	0,001	0,001-100	1 – 3	–
2	Плотность посадки спата в сосуде	шт/л	–	0 – 1000	15 – 50	–
3	Кол-во корма, выдаваемого за 1 раз	кл/ /экз·сут	1000	0 – ∞	2,8 млн 14 млн.	Сырая масса корма
4	Состав корма	нет	–	**Is:Ph* =1:1 *Is:Ph:Rh* = 2:2:1	*Is:Ph* = 1:1 *Is:Ph:Rh* =2:2:1	Качественный фактор

* **Примечание:** состав корма – *Isochrysis galbana : Phaeodactylum tricornutum* 1:1; *Is. galbana : Ph. tricornutum : Rhodomonas salina* 2:2:1

Уровни качественных факторов невозможно разместить на числовой оси. Действительно, бессмысленно утверждать, что хлороформ больше, чем бензол, или что физиотерапия меньше, чем гомеопатия. Кроме этого, уровни качественного фактора дискретны, поэтому невозможно задавать промежуточные значения качественных уровней, в то время как для количественных факторов это легко выполнимо. Поэтому при работе с качественными факторами, вместо регрессионного анализа применяют дисперсионный анализ и в исследованиях используют планы экспериментов, базирующихся на дисперсионном анализе.

Если в одном эксперименте работают одновременно с количественными и качественными факторами, в этом случае применяют сложные планы. При этом для обработки результатов используют и регрессионный, и дисперсионный анализы.

1.2. Организация экспериментального исследования

Планирование экспериментов (Design of Experiments – англ., Planification d'experimentations – франц.) – это наука, целью которой является получение максимально надёжных выводов минимально дешёвым (во всех смыслах этого слова) способом. При этом эксперимент рассматривается как любой вид деятельности, направленный на получение результата. Получаемые результаты или выводы представляются в стандартной форме, что делает их сопоставимыми с результатами, полученными другими исследователями.

Процесс экспериментального исследования можно разбить на следующие этапы:

1. Формулирование задачи и объекта исследования;
2. Выбор параметра оптимизации (зависимой переменной);
3. Анализ априорной информации;
4. Составление списка факторов, потенциально способных влиять на параметр оптимизации;
5. Составление списка факторов, реально используемых в эксперименте;
6. Задание уровней факторов;
7. Составление плана эксперимента;
8. Проведение эксперимента и получение результатов;
9. Статистическая обработка результатов эксперимента;
10. Интерпретация результатов (апостериорный анализ).

Итак, исследовательская работа начинается с формулирования цели или задачи исследования. Не всегда удаётся сформулировать в окончательном виде цель – в дальнейшем она будет уточняться или даже изменяться. При обдумывании формулировки цели исследователь должен ясно представлять себе, где и как будут применяться результаты его экспериментов. Необходимо также обдумать методы и оборудование, которые необходимо использовать для решения поставленной задачи.

Обычно обдумывание и обсуждение цели исследования, ОИ и ПО выполняются одновременно, так как эти разделы взаимосвязаны. Часто встречающаяся на данном этапе ошибка – поспешность. Но именно на данном этапе не следует экономить время и желательно первый этап (пункты 1 и 2) провести в течение нескольких дней.

В процессе анализа априорной информации (результаты собственных исследований, опубликованные данные и т.д.) основное внимание уделяется выявлению факторов, способных оказывать влияние на ПО. Необходимо также определить диапазоны уровней факторов, представляющих интерес для экспериментального исследования, а также, по возможности, иметь представление о характере зависимости ПО от каждого фактора.

Важно знать, что существует ошибочное мнение в отношении однофакторных экспериментов, утверждающее, что нужно ограничиться учётом только одного фактора, а остальные можно не принимать в рассмотрение. Это – грубая ошибка! Остальные факторы будут воздействовать на ПО и внесут искажения в получаемые результаты. Для защиты результатов от искажений нужно выявить все факторы, потенциально способные оказывать влияние на параметр оптимизации и составить список «подозреваемых факторов». В дальнейшем все факторы, кроме одного, будут удерживаться в течение эксперимента на постоянных уровнях – это обязательное условие проведения однофакторного эксперимента. В список необходимо включать не только «традиционные факторы», такие как температура, освещённость, концентрация веществ, плотность размещения животных в эксперименте и т.д., но и «нетрадиционные»: тип перемешивания воды в экспериментальном сосуде, тип экспериментальной установки, место взятия воды для эксперимента и т.д. Здесь уместно задаться вопросом: «Почему выводы, сделанные на основе результатов данного эксперимента, нельзя будет перенести на природу?» Перечисление причин поможет выявлению новых факторов.

На данном этапе происходит уточнение объекта исследования, находящегося в конкретных условиях и испытуемого на конкретной экспериментальной установке, что может внести изменения в формулировки ОИ и ПО.

Составление списка факторов, реально включаемых в эксперимент, выполняется в процессе анализа списка «подозреваемых факторов». Прежде всего, нужно, хотя бы приблизительно, проранжировать список факторов по силе их возможного влияния на ПО. При этом первое место займёт фактор, предположительно оказывающий наиболее сильное влияние. Возможно, что на данном этапе придётся применить методы отсеивания несущественных факторов, что будет подробнее рассмотрено в последующих разделах.

В эксперимент включаются существенные факторы, хотя не исключается возможность участия и несущественных факторов. Полезно также проанализировать условия и результаты, достигнутые при изучении других, но аналогичных процессов. Возможно, что придётся поставить несложный предварительный эксперимент, который позволит получить представление о величинах ошибок результатов. При составлении плана эксперимента необходимо рассмотреть следующие пункты:

- желаемое число опытов и ограничения на количество опытов;
- желаемые сроки проведения экспериментов;
- примерная продолжительность одного опыта;
- стоимость и затраты труда при проведении одного опыта;

- желательное число уровней для одного фактора;
- возможность выполнения параллельных опытов (повторов) и их желаемое количество;
- желаемая стратегия проведения опытов (например, по одному в день);
- какова ожидаемая зависимость ПО от факторов, заданных в конкретных диапазонах (линейная, нелинейная, точки максимума и минимума, точки перегиба);
- возможное взаимодействие факторов.

На этапе составления плана эксперимента исследователь переходит от языка биологии на язык математики и математической статистики, а возвращение на язык биологии происходит только на этапе интерпретации результатов. Полученные результаты экспериментов обрабатываются статистически – оценивается воспроизводимость результатов, рассчитываются все коэффициенты уравнения регрессии, после чего определяется их статистическая значимость. Незначимые коэффициенты отбрасываются, а уравнение регрессии с оставшимися членами проверяется на адекватность экспериментальным результатам.

Интерпретация результатов зависит от цели исследования, выбранных факторов и, главное, от особенностей плана эксперимента. Если ставилась задача изучения влияния некоторых существенных факторов на выбранный ПО – в этом случае проводится анализ коэффициентов уравнения регрессии. Знак минус, стоящий перед коэффициентом, говорит об отрицательном влиянии данного фактора на ПО и наоборот. Величина коэффициента отражает силу влияния фактора, что даёт возможность количественно оценить величину воздействия каждого фактора, а также всех межфакторных взаимодействий на изучаемый ПО.

Экстремальный эксперимент позволяет быстро найти в сложной ситуации оптимальные условия, в которых ПО принимает наилучшие значения. С помощью отсеивающего эксперимента можно выявить несущественные факторы и освободиться от необходимости их изучения в дальнейших исследованиях. Эксперимент высокой дробности применяют в качестве предварительного эксперимента, позволяющего найти условия проведения более тонкого эксперимента. Последний случай напоминает работу исследователя с микроскопом: поиск при малом увеличении интересного участка на препарате, затем при большем увеличении – обследование этого участка с целью нахождения ещё более интересного участка, который и обследуется детально при максимальном увеличении микроскопа. Аналогично и экспериментальный план высокой дробности позволяет получить общее представление об исследуемой сложной ситуации и выделить в ней перспективную область для дальнейших тщательных исследований, что выполняется с применением планов невысокой дробности, вплоть до применения планов полных факторных экспериментов.

Здесь перечислены только наиболее типичные примеры многофакторных экспериментов. В последующих главах рассматриваются примеры постановки экспериментов, статистической обработки и интерпретации результатов, полученных при экспериментальном решении различных задач.

1.3. Пассивный эксперимент

Для развития науки необходимо постоянное поступление научных фактов, которые добываются либо в процессе экспериментальных исследований, либо в полевых исследованиях (экспедициях). Между этими двумя типами исследований существует принципиальное различие. В первом случае экспериментатор активно управляет факторами, задавая и удерживая их на определённых уровнях в течение всего эксперимента. Во втором случае наблюдатель имеет дело с неуправляемыми факторами в Природе, он может только измерить их значения и записать в полевой журнал. Эта процедура называется «выполнение наблюдений» (в отличие от «проведения экспериментов»). Оба типа исследований в теории эксперимента подразделяют на активные эксперименты и пассивные эксперименты.

При проведении пассивных экспериментов данные измерений заносятся в таблицу, в которой в каждом столбце записывают значения факторов, то есть независимых переменных (например, глубина, освещённость, температура, солёность, концентрации биогенных солей, концентрации фитопланктона по видам и т.д.). В последнем столбце записываются значения зависимой переменной или выходной величины, например, первичной продукции.

Итак, данные наблюдений формируются в виде таблицы (массива), в которой в строках записаны данные по каждой точке отбора проб или по каждому району наблюдений (case), а столбцы – это значения переменных (variable). Данные практически всегда обрабатываются на компьютере с использованием пакетов программ многомерной статистики, например, Statistica 6 (Халафян, 2007). Обработка данных позволяет выявить связи выходной величины с независимыми переменными, исследовать связи между всеми измеряемыми величинами, изучить сходства и различия между разными районами исследований и т.д. Однако существуют ограничения при обработке исходных данных и ограничения на интерпретацию результатов статистической обработки.

Количество и расположение точек отбора проб планируются, исходя из величин ошибок измеряемых величин и желательной точности получаемых результатов. Расположение точек на исследуемом полигоне зависит от расположения «источников неоднородностей», то есть факторов, вносящих искажения, либо отклонения от средних значений измеряемых факторов. Например, при исследовании гидрохимических и гидробиологических особенностей бухты можно столкнуться со следующими источниками неоднородностей: река, впадающая в бухту, труба выпуска загрязнённых вод, место выхода бухты в море, откуда поступает вода с иными значениями измеряемых характеристик, зона пляжа, лодочная станция и т.д. Выявив источники неоднородностей, можно составить план отбора проб воды таким образом, чтобы влияние этих источников на значения измеряемых факторов было не систематическим, а случайным (применение метода рандомизации). Этот метод позволяет избежать ошибочных выводов, получаемых из-за ошибок при взятии повторных проб – «мнимых повторностей» (Козлов, 2003, 2014). В результате мы получим неискажённые оценки средних значений измеряемых параметров по бухте. Но, если нам не нужны обобщённые показатели, характеризующие бухту в целом, в этом случае вместо метода рандомизации следует использовать другие методы, например, планы дисперсионного анализа и т.д.

Обработка данных наблюдений методами многомерной статистики позволяет выявить тонкие особенности, характеризующие сложные связи между факторами. Например, определив коэффициент парной корреляции между двумя параметрами (независимыми или зависимыми переменными), представляющими для нас особый интерес, мы получаем значение коэффициента корреляции близкое к нулю и делаем вывод об отсутствии связи между этими параметрами. Однако, применив факторный анализ, мы можем прийти к противоположному выводу: связи между этими параметрами в действительности существуют, но в одних случаях эти связи положительные, а в других – отрицательные. Но парный коэффициент корреляции даёт только усреднённые характеристики связей, что не всегда достаточно. В данном примере факторный анализ может выявить все случаи, в которых связи между изучаемыми параметрами положительны, отрицательны, либо вовсе отсутствуют, и тем самым, указать нам на закономерности функционирования объекта исследования.

Методы многомерной статистики многообразны. Одних только вариантов факторного анализа существует более 30. Выбор метода зависит от характера решаемых задач и от особенностей результатов наблюдений.

1.4. Некоторые элементы математической статистики

Статистическая обработка результатов экспериментов, рассматриваемых в данной книге, выполняется с применением небольшого числа элементов и приёмов математической статистики.

Среднее арифметическое: $\bar{y} = \frac{\sum y_i}{n}$

Дисперсия: $S^2 = \frac{\sum (y_i - \bar{y})^2}{n-1}$

Стандартное отклонение: $S = \sqrt{\frac{\sum (y_i - \bar{y})^2}{n-1}}$

Ошибка среднего (стандартная ошибка): $S_{\bar{y}} = \frac{S}{\sqrt{n}}$

Доверительный интервал: $y_{\max/\min} = \bar{y} \pm t \frac{S}{\sqrt{n}}$,

где t – критерий Стьюдента

Точность измерения: $p = \frac{S/\sqrt{n}}{\bar{y}} 100\%$

Из приведенных формул следует, что массив первичных данных можно охарактеризовать средним арифметическим, стандартным отклонением и количеством наблюдений. Среднее арифметическое характеризует величину самой измеряемой характеристики, а стандартное отклонение – её качество, точнее её изменяемость или нестабильность. С точки зрения пользователя, среднее арифметическое с минимальным стандартным отклонением лучше, чем с большим, так как это свидетельствует о высокой точности измерения данной величины. Но исследователя, напротив, могут привлекать средние значения с широким диапазоном варьирования, потому что такая нестабильность требует объяснения, а следовательно, поиска причин нестабильности (или неизвестных факторов, воздействующих на изучаемую характеристику). А поиск причин или факторов может привести к открытию... Точно так же при построении графика зависимости двух переменных исследователь, нанося свои данные, получает «облако» точек, что сильно затрудняет выявление связи между переменными. В данном случае целесообразно не пытаться установить связь между переменными, а заняться выявлением причин образования «облака» данных, то есть поиска факторов, приводящих к отклонениям исследуемых переменных. А этот путь может привести к новым интересным и ценным с научной точки зрения результатам.

Не лишним будет напоминание о том, что для статистической обработки используют только первичные данные (непосредственные результаты измерений). Но результаты предварительной статистической обработки первичных данных, а также результаты расчётов, выполненных с использованием первичных данных, не годятся для дальнейшей статистической обработки. Так, составив выборку из средних арифметических, нельзя проводить дальнейшую статобработку этих средних. Точно так же, используя данные первичных измерений, можно составить ряд различных соотношений (отношение размера тела к его весу; отношение содержания белков к содержанию липидов в теле и т.д.), нельзя рассматривать в качестве результатов первичных измерений и обрабатывать их статистически.

Отметим также, что в научных публикациях, в которых содержатся таблицы с результатами, выраженными средними арифметическими, следует средние арифметические приводить с их стандартными ошибками (со знаком ±). Если же средние арифметические представлены на графике – в этом случае на графике откладываются доверительные интервалы. Табличные и графические представления данных преследуют разные цели. Табличные значения представляют читателю значения полученных величин и дают представление о точности, с которой они получены. А график, кроме указания величин, демонстрирует данные в сравнительном аспекте, то есть иллюстрирует изменения зависимой переменной при изменениях независимой переменной. При этом у читателя возникает вопрос: «Действительно ли данные, представленные на графике, различаются, либо они статистически не различимы?» Доверительные интервалы вносят ясность в эту проблему. С некоторой натяжкой можно утверждать, что, если доверительные интервалы сравниваемых величин пересекаются, значит достоверных различий между ними не существует. Верно и обратное. Доверительный интервал и стандартная ошибка различаются принципиально: доверительный интервал указывает поле, в которое попадёт каждое последующее измерение (если мы решим провести дополнительные измерения), с вероятностью 1-α, где α – уровень значимости доверительной вероятности. А стандартная ошибка среднего указывает пределы, в которых изменялась уже измеренная величина. Эти пределы дают представление о точности измерения.

Измерив несколько раз значение переменной у и определив арифметическое среднее и стандартное отклонение, можно задаться требуемой точностью измерений и рассчитать количество измерений, необходимое для обеспечения заданной точности. Допустим, что требуется выполнить измерение с точностью 10%, то есть p=10%, что означает: 10/100=0,1.

Используя формулу для точности измерений, можно записать:

$$\frac{S/\sqrt{n}}{\bar{y}} = 0{,}1 \quad \frac{S}{\sqrt{n}} = 0{,}1\bar{y} \quad \sqrt{n} = \frac{S}{0{,}1\bar{y}} \quad n = \frac{100S^2}{\bar{y}^2}$$

В общем виде формула для расчёта необходимого числа повторов, обеспечивающих приемлемую точность измерений, имеет вид:

$$n = \frac{S^2}{p^2\bar{y}^2}$$

Эта формула весьма полезна при планировании исследовательской работы. Предварительно нужно выполнить несколько определений изучаемой характеристики, найти её среднее арифметическое значение, дисперсию; задаться требуемой точностью измерений и затем рассчитать необходимое количество измерений (достаточное для обеспечения заданной точности). Очевидно, что точность измерений может быть различной и зависит от области исследований. Так, в биохимических исследованиях она может быть довольно высокой – *p=1%*; в физиологических *p=5%*; в экологических *p=20%* может считаться достаточно высокой точностью. Низкая точность измерений в экологии обусловлена рядом причин, в том числе действием экологических механизмов, например, неравномерным созреванием гонад в период размножения, образования различных групп (морф) в популяции, что увеличивает разнообразие и, следовательно, дисперсию и т.д.

Периодически точность измерений следует проверять расчётным путём.

Напомним, что низкая точность измерений характеризует работу как недобросовестную, а завышенная точность говорит о статистической неграмотности исполнителя.

Примеры для усвоения материала.

Вариант 1.

В результате четырёх измерений зависимой переменной у *получены следующие значения: 12; 23; 8; 35. Определить среднее арифметическое, стандартную ошибку среднего арифметического, доверительный интервал и найти необходимое число измерений, обеспечивающее 5% точность измерения (критерий Стьюдента = 3,18).*

Ответ:

среднее арифметическое: 19,5

Стандартное отклонение: 12,12:

Стандартная ошибка среднего: 6,06
Доверительный интервал: 19,5+/-19,28
Число измерений: $n=(100^2 \cdot S^2)/p^2y^2= (10000 \cdot 146,89)/25 \cdot 380,25=155$

Вариант 2.
Дана выборка: 0,56; 0,33; 0,75
Определить: среднее арифметическое, стандартную ошибку среднего арифметического, доверительный интервал.
Найти число измерений, обеспечивающее 10% точность измерений. (Критерий Стьюдента t= 3,18).
Ответ:
Среднее арифметическое = 0,55
Стандартное отклонение = 0,21
Стандартная ошибка среднего = 0,12
Доверительный интервал: 0,55 +/- 0,39
Число измерений = 15

Глава 2.
ПЛАНЫ РЕГРЕССИОННОГО АНАЛИЗА

2.1. Регрессионный анализ

В данной главе рассматриваются планы экспериментов, результаты которых анализируются с применением методов регрессионного анализа. Поэтому, прежде чем применять эти планы, необходимо разобраться в основных положениях регрессионного анализа, его требованиях и ограничениях. Результат эксперимента, проведенного на основе экспериментального плана, базирующегося на регрессионном анализе, представлен в виде уравнения зависимости ПО от интересующих нас факторов (независимых переменных), то есть от переменных, включённых в эксперимент. Это уравнение называется уравнением регрессии, которое имеет вид многочлена, все коэффициенты которого и знаки, стоящие перед ними, несут ценную для исследователя информацию. Кроме этого, уравнение позволяет выполнять интерполяцию, а именно рассчитывать величину ПО при любых промежуточных значениях уровней факторов.

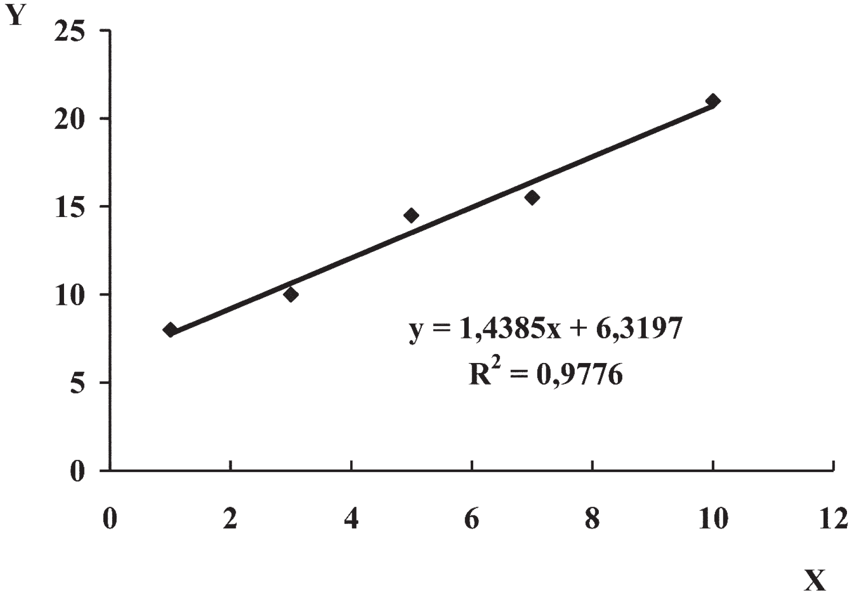


Рис. 2. График и уравнение зависимости **y** от **x**.

Например, при проведении обычного однофакторного эксперимента, в котором единственный фактор **x** варьировался на пяти уровнях, получены, соответственно, пять значений ПО (или **y**). Условия проведения эксперимента и полученные результаты удобно хранить в табличной форме (табл. 4).

Таблица 4. Условия проведения и результаты эксперимента:

№	x	y
1	1	8
2	3	10
3	5	14,5
4	7	15,5
5	10	21

Введя значения **x** и **y** в пакет статистических программ (например, Statistica), можно получить искомое уравнение и график зависимости **y** от **x,** то есть y=f(x), что отражено на рис. 2.

Округлив значения полученного уравнения, перепишем его в каноническом (нормальном) виде: **y =** 6,3 + 1,4 **x.**

Видно, что связь между ПО и изучаемым фактором, при его варьировании в пределах 1 – 10, хорошо описывается линейной зависимостью вида **y =** b_0 + b_1**x.** Свободный член b_0 соответствует значению ПО при условии, что фактор **x** принимает нулевое значение: ПО = b_0= 6,3, если **x =** 0. Коэффициент при втором члене уравнения (b_1 =+1,4) отражает положительное влияние фактора на величину ПО (знак +) и показывает, что значение ПО увеличивается на 1,4 при каждом увеличении значения фактора **x** на единицу.

На данном примере, в упрощённой форме, продемонстрирован регрессионный анализ. Для выполнения полного анализа необходимо провести статистическую проверку воспроизводимости экспериментальных результатов, определить статистическую значимость всех коэффициентов регрессии, после чего все незначимые – отбросить и в заключение выполнить проверку адекватности уравнения (с оставшимися значимыми членами) экспериментальным результатам.

В реальном эксперименте число факторов редко равно 1. Если количество факторов равно двум, тогда в таблице условий и результатов эксперимента добавляется ещё один столбец, в который вносятся значения второго фактора $\mathbf{x_2}$, а уравнение содержит ещё два члена:

$$\mathbf{y = b_0 + b_1x_1 + b_2x_2 + b_{12}x_1x_2}$$

Это уравнение описывает не линию, а поверхность в трехмерном пространстве, координатными осями которого являются вертикальная ось **y** и две горизонтальных оси: $\mathbf{x_1}$ и $\mathbf{x_2}$. Регрессионный анализ результатов многофакторных экспериментов излагается ниже в разделах, описывающих полные и дробные факторные эксперименты. Однако здесь следует отметить, что не любые результаты могут быть обработаны методом регрессионного анализа (РА). Существуют ограничения, которые чаще называют предпосылками.

Предпосылки регрессионного анализа.

1. ПО – случайная величина, подчиняющаяся нормальному закону распределения. Но обычно число повторов каждого опыта не превышает 3 – 5, что не позволяет надёжно определить тип распределения случайной величины, поэтому предполагается, что экспериментальные значения ПО распределены нормально.
2. Дисперсии ПО однородны (значения ПО повторяются, т.е. воспроизводятся с одинаковой дисперсией). Проверка однородности дисперсий ПО приведена в разделе: «Полный факторный эксперимент».
3. Значения (уровни) факторов не случайные величины, т.е. объект управляем. Иначе говоря, все независимые переменные (все факторы) удерживаются на постоянных уровнях в течение всего эксперимента.
4. Факторы не коррелируют между собой – эта предпосылка важна при интерпретации результатов, и она подробно рассматривается в последующих разделах.
5. Заранее известен тип уравнения для описания искомой зависимости. Поэтому, используя РА, рассчитывают значения всех коэффициентов заданного нами типа уравнения.

Следует уточнить, что коэффициенты заданного уравнения определяются методом наименьших квадратов (МНК). С помощью этого метода рассчитываются параметры такой линии (т.е. уравнения), которая наиболее близко проходит к экспериментальным точкам, а именно сумма квадратов отклонений экспериментальных точек от линии будет минимальной. Отклонения могут быть как положительными, так и отрицательными, поэтому при суммировании они взаимно уничтожаются. Избежать этого можно путём возведения отклонений в квадрат. Ниже, для заинтересованных читателей, приведено краткое описание этого метода. Однако это описание можно и опустить без ущерба для усвоения всего последующего материала.

МНК – изобрели Лежандр и Гаусс почти 200 лет тому назад. Это самый популярный среди исследователей метод аппроксимации экспериментальных данных с использованием заранее выбранного типа уравнения.

В самом простом случае – один фактор – линейная модель имеет вид: $y = b_0 + b_1x_1$

Необходимо вычислить: b_0, b_1

Если все точки лежат на прямой, то: $y_i – b_0 – b_1x_1 = 0$.

После проведения эксперимента получили результаты y_i, где i = 1,2,3,...N

$y_i – b_0 – b_1x_1 = \varepsilon_i$ ε_i – некоторый остаток, называемый в математике невязка

Невязка обусловлена ошибкой метода и (или) непригодностью модели, (т.е. уравнения).

Постулируем пригодность модели.

Требуется найти такие коэффициенты регрессии, при которых невязки будут минимальны.

В формате МНК это запишется так: $U = \sum_{i=1}^{n} \varepsilon^2{}_1 \rightarrow \min$

Число опытов должно быть больше числа неизвестных коэффициентов. Система уравнений: $U = \Sigma\varepsilon_i^2 = \Sigma(y_i – b_0 – b_1x_{1i})^2 = \min$ является противоречивой.

Это условие достигается в точке экстремума, где частные производные равны 0:

$đU/đ\ b_0 = 0 \quad đU/đ\ b_1 = 0$ или:

$-2\ \Sigma(y_i – b_0 – b_1x_{1i}) = 0$

$-2\Sigma(y_i – b_0 – b_1x_{1i})\ x_{1i} = 0$

После преобразования:

$Nb_0 + \Sigma\ b_1x_{1i} = \Sigma y_i$

$\Sigma\ b_0x_{1i} + \Sigma\ b_1x_{1i}^2 = \Sigma y_i\ x_{1i}$

Отсюда:

$b_0 = (\Sigma y_i\ \Sigma x_{1i}^2 – \Sigma y_i x_{1i}\ \Sigma\ b_1x_{1i})/(\ N\Sigma x_{1i}^2 – (\Sigma x_{1i})^2$

$b_1 = (N\Sigma y_i x_{1i} – \Sigma y_i\ \Sigma x_{1i})/(\ N\Sigma x_{1i}^2 – (\Sigma x_{1i})^2$

Для проверки правильности расчётов следует проверить равенство

$y_{ср} = b_0 + b_1x_{1ср.}$

Здесь приведена обработка однофакторного эксперимента. Но на практике работают со сложными (многофакторными) объектами, для исследования которых ставят многофакторные эксперименты. В этом случае для расчёта коэффициентов регрессии следует воспользоваться аппаратом матричной алгебры, что описано в конце книги в разделе, излагающем методы пассивного эксперимента. Однако на практике расчёты выполняются на компьютере с использованием статистических пакетов.

При проведении активных экспериментов выполнение расчётов значительно упрощается и поэтому выполняется на калькуляторах, снабжённых несложными статистическими программами, выдающими значения среднего арифметического и стандартного отклонения (либо дисперсии). Эти расчёты подробно объясняются в следующем разделе.

2.2. Полный факторный эксперимент (ПФЭ)

ПФЭ – это эксперимент, в котором реализуются все возможные сочетания уровней факторов.

Выбор плана определяется целью планируемой работы, типом и количеством факторов, числом уровней факторов. На каком количестве уровней следует варьировать фактор? С общих позиций можно утверждать, что различные количественные факторы воздействуют на ПО аналогичным образом (рис. 3), а именно: при низких значениях уровня фактора С (1 – зона лимитирования), увеличение значений фактора приводит к росту ПО (положительное влияние фактора). Затем наступает так называемая «зона насыщения» (2), в пределах которой изменения значений уровней фактора не влекут за собой изменений значений ПО (фактор не оказывает влияния). За верхним пределом «зоны насыщения» наступает «зона угнетения» (3), в пределах которой увеличение уровней фактора приводит к падению значений ПО – отрицательное влияние фактора (рис. 3).

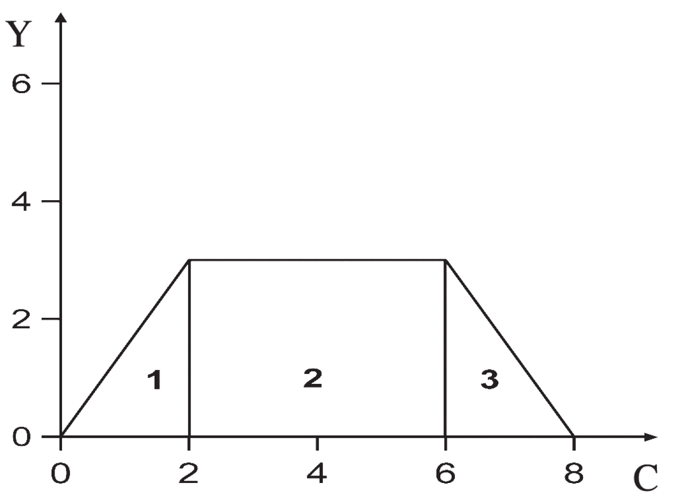


Рис. 3. Влияние фактора С на параметр оптимизации Y.

Таким образом, один и тот же фактор может влиять положительно, отрицательно, либо не влиять вообще – характер влияния зависит от выбора уровней фактора. При планировании эксперимента необходимо выбрать интервал, в рамках которого будет варьироваться фактор. Обычно (по экономическим, техническим и другим причинам), экспериментатора интересует зона лимитирования, ширину интервала которой он планирует установить, а также получить количественную зависимость ПО от данного фактора внутри исследуемой зоны. Если фактор будет варьироваться в пределах одной из перечисленных зон – в этом случае данный фактор достаточно варьировать на двух уровнях, что подтверждается мировым опытом применения методов планирования многофакторных экспериментов. При варьировании фактора в пределах двух или трёх зон, фактор придётся варьировать на трех либо большем количестве уровней.

Формально в качестве экспериментального результата получают зависимость $Y=f(x_1,...,x_n)$. Функциональная зависимость должна быть простой и информативной. Выше говорилось, что лучше всего подходит полиномиальная модель в виде уравнения регрессии:

$$y= b_0 + b_1x_1 + b_2x_2 + b_{12}x_1x_2 + ...+ b_{11}x_1^2 ...$$

Коэффициенты уравнения регрессии количественно характеризуют влияние каждого фактора на объект исследования (ОИ), а также влияния эффектов взаимодействия факторов. В простейшем случае при изучении влияния одного фактора получаем: $y= b_0 + b_1x_1$. Здесь фактор варьируется на двух уровнях. Число разных опытов: $\mathbf{N=2^n=2^1=2}$.

План полного факторного эксперимента ПФЭ 2^2.

Количество факторов: **n=2,** число разных опытов: $\mathbf{N=2^2=4}$. Опыты ставятся в четырех точках (рис. 4). Четырех опытов достаточно для определения четырёх коэффициентов уравнения регрессии: $\mathbf{y= b_0 + b_1x_1 + b_2x_2 + b_{12}x_1x_2}$. Это неполное квадратное уравнение, описывающее поверхность (а не плоскость!).

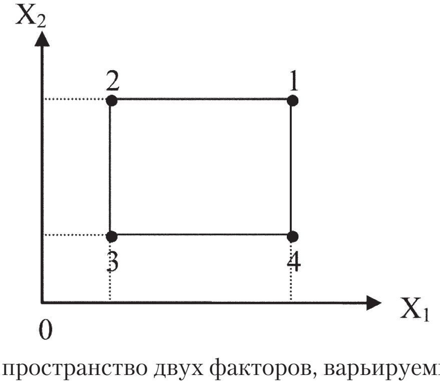


Рис. 4. Факторное пространство двух факторов, варьируемых на двух уровнях.

Величины коэффициентов уравнения зависят от размерности и от силы влияния факторов. Но для выделения чистого влияния фактора необходимо освободиться от размерностей факторов. Эта операция называется «кодированием факторов». Заключается она в следующем: нижний уровень фактора принимается за -1; верхний уровень за +1. Середина между крайними значениями уровней называется базовым (или основным) уровнем и обозначается 0. В результате факторное пространство преобразовывается, становясь симметричным относительно начала координат (рис. 5).

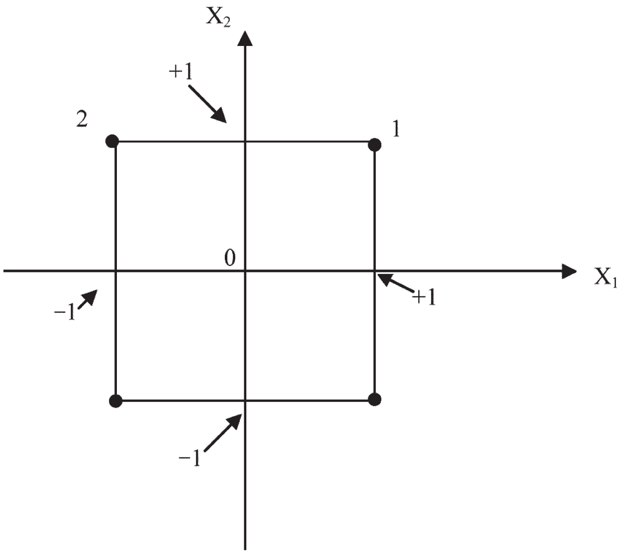


Рис. 5. То же факторное пространство после преобразования путём кодирования переменных.

Перечень координат экспериментальных точек 1, 2, 3, 4 образует матрицу планирования эксперимента (табл. 5).

Итак, весь эксперимент состоит из четырёх разных опытов. В первом столбце перечислены номера четырёх разных опытов (четыре комбинации уровней факторов). Во втором и в третьем столбцах фигурируют значения уровней факторов. Например, в опыте № 2 фактор x_1 задан на нижнем уровне, а фактор x_2 – на верхнем. Каждый опыт повторяется 3 или более раз, что необходимо для выполнения статистической обработки. Каждая строка отражает условия проведения «опытов», сумма которых и составляет «эксперимент».

Представленный план является симметричным относительно начала координат (рис. 5) и обладает следующими важными свойствами:

Таблица 5. План (матрица) полного факторного эксперимента для двух факторов, каждый из которых варьируется на двух уровнях (ПФЭ 2^2)

№	x_1	x_2
1	+1	+1
2	-1	+1
3	-1	-1
4	+1	-1

1. Ортогональность: скалярные произведения столбцов равны 0. Иными словами, сумма построчных парных произведений чисел первого столбца (для x_1) на соответствующие числа второго столбца (x_2) равна 0. Это говорит об отсутствии взаимодействия между столбцами, поэтому коэффициенты уравнения регрессии определяются независимо друг от друга, что очень важно для интерпретации коэффициентов уравнения. Такой план называется **D-оптимальным**.
2. Симметричность: число плюсов равно числу минусов по каждому столбцу. Все коэффициенты уравнения определяются с минимальными и равными ошибками.
3. Свойства нормировки: сумма квадратов элементов столбцов равна числу опытов. Такие планы называются **рототабельными.** Они позволяют рассчитывать ПО с одинаковыми ошибками для всех точек факторного пространства, равноудалённых от центра эксперимента.

2.2.1. Составление плана полного факторного эксперимента

Для составления плана полного факторного эксперимента с числом факторов **k**, варьируемых на **n** уровнях (ПФЭ n^k) необходимо перечислить все возможные сочетания **k** факторов, каждый из которых принимает **n** значений (находится на **n** уровнях). Число таких сочетаний определяется формулой: $g=n^k$. Для двух факторов, варьируемых на двух уровнях $g=2^2=4$; для трёх и четырёх факторов число сочетаний $g=2^3=8$ и $g=2^4=16$. Если факторы варьируются на трёх уровнях, то для двух, трёх и четырёх факторов число сочетаний будет равно соответственно: $g=3^2=9$; $g=3^3=27$; $g=3^4=81$.

В матрице планирования ПФЭ 2^2 (табл. 5) в первом столбце знаки чередуются через одну строку (2^0), а во втором столбце – через две строки (2^1). Если бы факторов было три, тогда в третьем столбце знаки чередовались бы через 4 строки (табл. 6) и т.д.

Если факторы варьируются на трёх уровнях, тогда в первом столбце знаки (-1; 0; +1) должны изменяться через строку, во втором столбце через три строки (3^1), в третьем – через 9 строк (3^2) и т.д.

Для примера построим матрицу планирования ПФЭ2^3 ($g=8$) и ПФЭ3^3 ($g=27$) (табл. 6 и 7).

Таблица 6. Матрица планирования ПФЭ2^3

№	x_1	x_2	x_3
1	+1	+1	+1
2	-1	+1	+1
3	+1	-1	+1
4	-1	-1	+1
5	+1	+1	-1
6	-1	+1	-1
7	+1	-1	-1
8	-1	-1	-1

Матрица планирования для ПФЭ 3^2 g = 9; уравнение содержит 9 членов: $\mathbf{y = b_0 + b_1x_1 + b_2x_2 + b_{12}x_1x_2 + b_{11}x_1^2 + b_{22}x_2^2 + b_{1211}x_1x_2x_1^2 + b_{1222}x_1x_2x_2^2 + b_{121122}x_1x_2x_1^2x_2^2}$**.**

Таблица 7. Матрица планирования ПФЭ3^3.

№	x_1	x_2	x_3
1	+1	+1	+1
2	0	+1	+1
3	-1	+1	+1
4	+1	0	+1
5	0	0	+1
6	-1	0	+1
7	+1	-1	+1
8	0	-1	+1
9	-1	-1	+1
10	+1	+1	0
11	0	+1	0
12	-1	+1	0
13	+1	0	0
14	0	0	0
15	-1	0	0
16	+1	-1	0
17	0	-1	0
18	-1	-1	0
19	+1	+1	-1
20	0	+1	-1
21	-1	+1	-1
22	+1	0	-1
23	0	0	-1
24	-1	0	-1
25	+1	-1	-1
26	0	-1	-1
27	-1	-1	-1

Как было сказано выше, число членов уравнения равно числу разных опытов (число строк в матрице планирования) и эти члены отражают все возможные взаимодействия факторов. ПФЭ 2^3 позволяет получить уравнение, состоящее из 8 членов: $\mathbf{y= b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{23}x_2x_3 + b_{123}x_1x_2x_3}$. По результатам постановки ПФЭ 2^4 получают уравнение регрессии, содержащее 16 членов: $\mathbf{y= b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{14}x_1x_4 + b_{23}x_2x_3 + b_{24}x_2x_4 + b_{34}x_3x_4 + b_{123}x_1x_2x_3 + b_{124}x_1x_2x_4 + b_{134}x_1x_3x_4 + b_{234}x_2x_3x_4 + b_{1234}x_1x_2x_3x_4}$.

Первый член уравнения называется свободным членом. Члены, стоящие перед каждым отдельным фактором, называются главными эффектами. Члены, стоящие перед двойным, тройным и так далее произведениями факторов, называются эффектами двойных, тройных и т.д. взаимодействий факторов. Эффекты взаимодействий отражают степень кривизны поверхности отклика: чем больше поверхность отклоняется от плоскости, тем выше значения эффектов взаимодействий. Часто это происходит в случаях, когда экспериментатор выбирает широкие интервалы варьирования факторов, либо когда факторы тесно связаны. Специалисты по статистике используют сленг, в соответствии с которым свободный член и главные эффекты образуют «голову», а эффекты взаимодействий – «хвост» уравнения регрессии. «Хвост» усиливается либо удлиняется при увеличении кривизны поверхности отклика. Соответственно, если поверхность отклика приближается к плоскости, тогда «хвост» укорачивается, вплоть до полного его «отбрасывания» при отсутствии кривизны.

Для расчёта коэффициентов уравнения регрессии, кроме матрицы планирования, необходимы столбцы для расчёта коэффициентов взаимодействий факторов, а также свободного члена. Последний рассчитывается с помощью столбца, так называемой фиктивной переменной x_0, которая всегда принимает значение +1. Столбцы взаимодействий определяются как произведения соответствующих столбцов. Например, для получения значений для ячеек столбца x_1x_2 (табл. 8) нужно перемножить значения ячейки столбца x_1 на значение соответствующей ячейки столбца x_2. Так для опыта № 2 (вторая строка) перемножаем значение x_1= -1 на x_2=+1 и в результате получаем x_1x_2= -1.

Таблица 8. Полная матрица ПФЭ2^2

№	x_0	x_1	x_2	x_1x_2	$y_1 \dots y_m$	ỳ	S_g	S_g^2	ỷ	(ỳ – ỷ)^2
1	1	+1	+1	+1						
2	1	-1	+1	-1						
3	1	+1	-1	-1						
4	1	-1	-1	+1						
								ΣS_g^2		**Σ (ỳ – ỷ)^2**

В таблице 8 имеются следующие столбцы, предназначенные для проведения статистической обработки результатов: $y_1...y_m$, где **m** – количество повторов или результаты по каждому повтору данного опыта; $\grave{y}$ – средний результат по каждому опыту; $\hat{y}$ – теоретический результат, рассчитанный по уравнению регрессии; S_g и S_g^2 – стандартное отклонение и дисперсия, полученные по результатам каждого опыта.

2.2.2. Статистическая обработка результатов эксперимента

Статистическая обработка результатов эксперимента состоит из трёх разделов:

- проверка воспроизводимости результатов эксперимента;
- расчёт коэффициентов уравнения регрессии и проверка статистической значимости коэффициентов;
- проверка адекватности уравнения регрессии экспериментальным данным.

Итак, прежде всего, выполняется проверка воспроизводимости результатов опытов. Результаты должны повторяться (воспроизводиться) примерно с одинаковой ошибкой – только в этом случае их можно рассматривать в качестве научных фактов и проводить их дальнейший анализ. Проверку воспроизводимости можно выполнять двумя путями: 1) повторить несколько раз каждый опыт (каждую строку) либо повторить одну строку несколько раз. Используя результаты повторов, находят средние арифметические, стандартные отклонения и дисперсии. Все дисперсии суммируют и находят отношение максимальной дисперсии к сумме всех дисперсий. Это отношение называется расчётным критерием Кохрена, который сравнивают с табличным (критическим) значением критерия Кохрена. Если результаты не воспроизводятся (расчётный критерий больше табличного), тогда необходимо выяснить причины: возможно, что действует сильный неучтённый фактор, влияющий на результаты; либо сделано слишком мало повторов.

Коэффициенты уравнения регрессии рассчитываются очень просто: нужно найти алгебраическую сумму средних значений результатов, причём средние значения берутся со знаками, стоящими в столбце x_i. Полученную алгебраическую сумму надо разделить на число разных опытов (число строк в матрице планирования). Например, для вычисления величины $\mathbf{b_1}$ нужно просуммировать арифметические средние результатов опытов с учётом знаков, стоящих в столбце $\mathbf{x_1}$ и, в случае табл. 8, полученную алгебраическую сумму разделить на 4 (число разных опытов).

Коэффициент считается статистически значимым, если его абсолютная величина превышает ошибку этого коэффициента не менее, чем в некоторое число раз, а именно – в величину t_i (критерий Стьюдента). Если фактор не значим, тогда возможно, что уровни взяты в области насыщения; либо интервал варьирования фактора слишком узок. Все статистически незначимые коэффициенты отбрасываются, а уравнение с оставшимися коэффициентами проверяется на адекватность. Если все коэффициенты уравнения статистически значимы – в этом случае проверка на адекватность не выполняется.

Как уже отмечалось, полученное уравнение регрессии (или модель) должно адекватно описывать зависимость ПО от независимых переменных (факторов). Проверка адекватности выполняется путём сравнения теоретических значений результатов опыта (рассчитанных по уравнению) со средними арифметическими результатами, полученными в опытах. Разности между ними не должны превышать среднюю дисперсию результатов опытов, а именно – отношение дисперсии адекватности к средней дисперсии результатов опытов не должно превышать табличное значение критерия Фишера.

Формулы для статистической обработки результатов экспериментов

1. Число повторных опытов: m
2. Число разных опытов: g
3. Всего опытов: $N = m \times g$
4. Оценка воспроизводимости опытов (по критерию Кохрена):
5. $G_{max} = S^2_{gmax} / \Sigma S_g^{\,2}$, где S^2_{gmax} – максимальная дисперсия в столбце; $\Sigma S_g^{\,2}$ – сумма дисперсий столбца.
6. Расчёт коэффициентов уравнения регрессии: $b_i = (\Sigma x_i \grave{y}_g)/g$.
7. Значимость коэффициента уравнения (по Стьюденту): $t_i = b_i / \sqrt{S^2_{(bi)}}$), где: $S^2_{(bi)} = S^2/mg$, где: $S^2 = \Sigma S_g^{\,2}/g$.
8. Проверка адекватности уравнения (модели) по критерию Фишера. $F = S^2_{ад}/S^2$, где: $S^2_{ад} = (\Sigma(\grave{y} - \hat{y})^2/(g-d)$, где: **d** – число оставшихся членов уравнения. Если модель не адекватна, необходимо сузить интервалы варьирования факторов или выбрать более сложный план.

В дальнейшем, при выполнении расчётов с помощью полученных уравнений потребуется перевести натуральные значения факторов в кодированные. В такой операции используются следующие обозначения: x_i – натуральное значение фактора; x'_i – кодированное значение фактора; λ_i – шаг варьирования (величина, прибавление которой к основному уровню фактора даёт значение верхнего уровня, а вычитание от основного – нижний уровень).

Кодированное значение фактора рассчитывается по формуле:
$x'_i = (x_i - x_{i0}) / \lambda_i$,
где: x'_i – кодированное значение;
x_i – натуральное значение;
x_{i0} – среднее значение (середина интервала варьирования фактора).

2.2.3. Пример. Полный факторный эксперимент ПФЭ 2^2.

Выше (стр. 16-17) уже обсуждалась постановка задачи экспериментального определения выделения мидиями растворённых форм азот- и фосфорсодержащих соединений. Параметр оптимизации представлен в виде удельной скорости выделения химических соединений: Уд. скор **=** $(C_{кон} - C_{контроль})/2W$.

Следующий этап – составление списка факторов, способных оказывать влияние на данный ПО, то есть нужно составить «Список подозреваемых факторов» (табл. 9).

«Подозреваемые факторы» – это факторы, способные оказывать влияние на ПО (даже, если мы точно этого не знаем, а только подозреваем, что такая возможность существует).

Таблица 9. Список подозреваемых факторов.

№	Наименование фактора	Единица измерения	Ошибка, точность	Область определения	Область интереса	Примечание
1	Температура морской воды	°C	0,1	0 – 100	5 – 27	В сосуде с животными
2	Вес мидии	г	0,01	0,001 – 50	2,5 – 17	Общий вес
3	Объём экспериментального сосуда	л	0,001	0,1-100	0,5 – 3	
4	Плотность посадки мидий в сосуде	г/л	0,01	0 – 1000	1 – 100	
5	Количество корма, выдаваемого до опыта	мкг/г·сут	0,01	0,01 – 10	0,78–3,87	
6	Продолжительность кормления до эксперимента	сутки	0,04	0 – ∞	0 – 3	
7	Состав корма	нет	-	-	-	Качественный фактор
8	Аэрирование воды	л/мин	-	0 – 100	0 – 10	
9	Перемешивание воды (тип перемешивания)	-	-	-	-	Качественный фактор
10	Место взятия воды для эксперимента	нет	-	-	-	Качественный фактор

11	Обработка воды (размер пор фильтра)	мкм		0-50	0-50	0 - отсутствие фильтрации
12	Состояние гонад (стадия зрелости)	-	-	1 – 6	1-6	Качественный фактор
13	Солёность	‰		0-37	10-18	
14	Освещённость	лк		0-150000	0-50000	
15	Насыщение воды O_2	%		0 -120		
16	Режим адаптации мидий к эксперименту	час	0,08	1 -24	1-12	

Известно, что факторы бывают количественные, то есть их уровни задаются числом (температура, солёность, объём сосуда, вес животного), так и качественные, уровни которых числом не измеряются (фактор: «тип животного»; уровни: неполовозрелый, самец, самка, гермафродит). Список должен быть максимально полным. Фактор считается заданным, если вместе с его названием указана область его определения.

2.2.3.1. Обсуждение списка подозреваемых факторов

Все факторы, перечисленные в таблице 9, должны быть обсуждены и по каждому фактору должно быть принято решение относительно необходимости его включения в эксперимент. Ниже, в целях сокращения объёма раздела, приведено краткое обсуждение.

Первый фактор – *температура*. Её сложно задавать в эксперименте на разных уровнях не только технически, но и не ясно, как адаптировать животных к новым значениям температуры и как потом интерпретировать полученные данные. По-видимому, лучшее решение – это не менять температуру в одном эксперименте, а только указывать её значение в **условиях проведения эксперимента**. Но эксперимент придётся повторять ежемесячно – в этом случае можно охватить весь диапазон температур. Обсуждаемый эксперимент был поставлен 23 июля 2008 года при температуре морской воды 23°С. Следовательно, результат, который будет получен в этом эксперименте, будет пригоден только для температуры 23°С.

Второй фактор – *общий вес мидии*, который возможно варьировать в эксперименте в широких пределах: от самых мелких мидий и до наиболее крупных, которые встречаются редко. Например, мидии в возрасте 3,5 – 4 года достигают 30 г, хотя на мидийной ферме мидия не должна находиться более 2 лет (при достижении веса 13–15 г). Поэтому верхний уровень фактора может быть и 15 г. Важно отметить, что вес мидии **должен учитываться** в эксперименте, так как получаемые результаты будут использоваться для определения влияния мидий по мере их роста на ферме (после метаморфоза осевшей личинки и до достижения мидиями товарного размера). Основную биомассу на ферме образуют мидии размером 30 – 60 мм с общим весом, варьирующим в пределах 2,5 – 17 г. Поэтому нижний уровень фактора равен 2,5 г, а верхний – 17 г.

Следующий фактор – *объём экспериментального сосуда*. Может ли этот фактор сам по себе оказывать влияние на выбранный ПО? Можно предположить, что более значимым является фактор «*плотность посадки*», которая равна отношению веса всех мидий в сосуде к его объёму. Можно было бы проделать специальный опыт по изучению раздельного влияния объёма и плотности посадки, прежде чем включать оба фактора в общий план эксперимента. Либо, если наберётся много факторов для включения в эксперимент, поставить на первом этапе отсеивающий эксперимент, который позволит нам выявить сильные (существенные) и слабые (несущественные) факторы, которые можно исключить из эксперимента, и затем работать только с существенными факторами. В данном случае принято, что объём экспериментального сосуда равен 3 л, что и вошло в **условия постановки эксперимента.**

Плотность посадки мидий в экспериментальном сосуде. Частично этот фактор уже обсуждался выше. Возможно, что он непосредственно слабо влияет на мидий, так как эти животные образуют плотные поселения, поэтому адаптировались к жизни в условиях скученности. Но опосредованно (через недостаток кислорода, избыток аммония в **ограниченном объёме сосуда**) они могут влиять на свой метаболизм. Однако это будут «издержки» экспериментальных условий и к естественным условиям они не имеют отношения. Иными словами, в данном случае экспериментатор рискует получить артефакты. Поэтому плотность посадки в эксперименте нельзя задавать слишком высокой. На первом этапе данный фактор решено не включать в эксперимент, но его значение должно быть одинаковым во всех экспериментальных сосудах, и оно должно быть записано **в условия постановки эксперимента**. Во всех сосудах плотность посадки мидий находилась в пределах 9,7 – 11,3 г/л (общий сырой вес мидий).

Количество выдаваемого корма в сутки. Экспериментально было установлено, что если предварительно (то есть перед экспериментом) мидий не кормить – в этом случае они не будут выделять измеряемые соединения (ПО=0). Этот фактор **необходимо включать** в эксперимент, причём верхний уровень – это кормление почти с избытком, а нижний уровень – «кормить впроголодь». Количество выдаваемого корма должно измеряться, что необходимо для получения количественной оценки влияния данного фактора на ПО. Выдаваться корм должен пропорционально биомассе мидий в аквариуме для кормления. Фактор включается в эксперимент, причём на основе имеющейся информации установлено, что этот фактор должен варьироваться в пределах 0,78 – 3,87 мг/г×сут (количество сырого корма, выдаваемого в сутки на грамм общего сырого веса мидии).

Продолжительность кормления до эксперимента. Для расчёта этой величины необходимо знать время прохождения корма через кишечник. Этот показатель обычно определяют радиоуглеродным методом. Нужно кормить мидий в течение периода примерно равного времени прохождения корма. Кормить мидий значительно дольше – бессмысленно, так как длительное кормление не должно оказывать влияния на выделение мидиями веществ. Однако это предположение нуждается в проверке. Предварительно экспериментально было установлено, что мидии выделяют фекалии менее чем через двое суток после начала поедания корма. Поэтому этот фактор установили равным 2 сут. Данную информацию включили в **условия постановки эксперимента.**

Будет ли зависеть ПО от качественного состава корма? Возможна постановка отдельного эксперимента только с качественными факторами, либо включение этого фактора в отсеивающий эксперимент. На данном этапе решено этот фактор в эксперимент не включать, а вписать его значение **в условия эксперимента**. Качественный состав корма, который принят нами в качестве модели природного корма, состоит из двух видов микроводорослей: изохризиса и дуналиэлы, взятых в соотношении биомассы 1:1. Корм выдавали 1 раз в сутки в течение двух суток (21 и 22 июля) при концентрации 250 тыс. кл./мл и 50 тыс. кл./мл соответственно.

Аэрирование воды. До постановки эксперимента необходимо определить расчётным путём количество O_2 потребляемое мидиями за время эксперимента. Если потребление кислорода мидиями приведёт к понижению концентрации O_2 в воде экспериментального сосуда не более, чем на 50 %, в этом случае этот фактор можно не учитывать, хотя обоснование такого решения необходимо дать в методическом разделе. Следует учитывать, что аэрирование способствует перемешиванию воды в сосуде и вымыванию веществ из фекалий, что может привести к завышению скоростей выделения изучаемых соединений. Необходимо пояснить, что экспериментальные сосуды с мидиями выставлялись в море, что технически затрудняло их аэрирование. Для решения данной задачи был поставлен предварительный опыт по изучению кинетики выделения изучаемых соединений в непроточных сосудах. Опыт показал, что в течение первых шести часов процесс выделения соединений мидиями был линейным, что свидетельствует об отсутствии его угнетения. Таким образом, возможна постановка опыта и без аэрирования воды в сосудах, но при условии, что длительность эксперимента не будет превышать 6 часов.

Перемешивание воды в экспериментальном сосуде. В опытах с водными животными всегда применяют либо перемешивание воды, либо аэрирование. Можно объединить оба фактора, применив эрлифт. Перемешивать воду можно перистальтическими насосами, электромагнитными мешалками, вручную с использованием груши, вклеенной в крышку экспериментального сосуда, встряхивателем, эрлифтом и т.д. Однако в виду отсутствия нужного оборудования пришлось отказаться от изучения влияния этого фактора, что отмечается в методике постановки эксперимента.

Место взятия воды для эксперимента. Известно, что в загрязнённой воде мидии смыкают створки. На этом принципе основан французский прибор «мулиметр» (moule – муль – это мидия). С другой стороны, если в исходной воде концентрации измеряемых соединений высокие, тогда на их фоне труднее определить небольшие изменения концентраций в процессе эксперимента. Можно всегда брать воду из одного наиболее чистого места, что и оговаривается в методической части.

Обработка воды перед экспериментом. Такая обработка ограничивается фильтрацией воды с целью удаления посторонних организмов и возможного неконтролируемого поступления корма. Однако, исследования, обосновывающие необходимость фильтрации воды перед экспериментом, не проводились.

Состояние зрелости гонад. Невозможно варьировать этот фактор независимо от других факторов, так как невозможно узнать до опыта стадию зрелости гонад подопытной мидии. С другой стороны, предварительный эксперимент показал, что наш ПО не зависит от степени зрелости гонады. Поэтому данный фактор не включался ни в эксперимент, ни в условия постановки.

Солёность. Если планируется распространить ожидаемые выводы на Азовское море, на распреснённые лиманы, тогда этот фактор необходимо включать в эксперимент. Но для этого потребуется изучить способность адаптации мидий к понижению солёности (продолжительность периода адаптации и т.д.). К началу эксперимента данный фактор остался неизученным, поэтому результаты эксперимента будут применимы к солёности черноморской воды (17–18‰).

Освещённость. Данный фактор влияет на личинок двустворчатых моллюсков. Влияет ли он на взрослых мидий? Есть сведения о том, что высокая освещённость замедляет рост раковины, но способствует формированию мягких тканей; влияет также продолжительность светового дня (Брегман, Сидоренко, 1979). Этот фактор желательно проверить в отсеивающем эксперименте, но для этого необходимо в течение всего опыта поддерживать освещённость на заданных уровнях. В обсуждаемом эксперименте контроль освещённости не проводился.

Насыщенность воды кислородом. По литературным данным, понижение концентрации кислорода в воде ниже 80 % от насыщения приводит к замедлению роста мидий. Линейность процесса выделения мидиями изучаемых соединений в течение 6 часов позволяет экспериментатору этот фактор не учитывать.

Режим адаптации мидий к экспериментальным условиям. Собранные мидии испытывают стресс. Мидиям нужно дать время для возвращения к нормальной жизнедеятельности. Мидия, находящаяся в нормальном состоянии, всегда прикреплена к субстрату. Но экспериментаторы обычно этот факт игнорируют. Необходимо до эксперимента поместить мидий на индивидуальные подложки (например, предметные стёкла) и после прикрепления мидий к подложкам, мидий вместе с подложками можно переносить в экспериментальные сосуды.

Итак, обсуждение списка подозреваемых факторов позволило уточнить условия проведения эксперимента, методику исследований и выделить факторы, которые требуется включить в эксперимент (табл. 10).

Эксперимент ставится по плану ПФЭ 2^2 в трехкратной повторности, то есть всего 4×3=12 опытов (12 экспериментальных сосудов, объёмом по 3 литра каждый). Для размещения 12 сосудов была выбрана площадка, которая была размечена на 12 мест и места пронумерованы. Для размещения сосудов на площадке был применён метод рандомизации (т.е. случайного размещения сосудов на экспериментальной площадке), а именно: для каждого сосуда выбирали номер места по таблице случайных чисел.

Таблица 10. Список факторов, включаемых в эксперимент

№	Наименование фактора	Обозначение	Размерность	Область определения	Область интереса	Ошибка измерения	Примечание
1	Количество корма, выдаваемого в сут на г мидии	x_1	мг/г×сут	0-1000	0,77-3,87	0,1	Сырой вес
2	Индивидуальный сырой вес мидии	x_2	г	0-40	2,5 – 17	0,1	

Те же факторы в кодированном виде представлены в табл. 11.

Таблица 11. Кодирование факторов, включаемых в эксперимент.

Факторы	x_i	x_1	x_2	x_i
Верхний уровень	$x_{i\,вер}$	3,87	17	+1
Нижний уровень	$x_{i\,ниж}$	0,77	2,5	-1
Базовый уровень	$x_{i\,0}$	2,32	9,75	0
Шаг варьирования	λ_i	1,55	7,25	+1

Эти таблицы можно взять из Интернета. Если в таблице содержатся двузначные, либо трёхзначные числа, тогда нужно ориентироваться на первые цифры каждого числа. Процедура выполняется следующим образом: берут сосуд № 1 и смотрят на первую цифру первого числа таблицы. Если это цифра 5, тогда сосуд № 1 ставят на место № 5 экспериментального полигона. В случае повторения той же цифры в таблице случайных чисел – это число пропускают. Данный метод позволяет преобразовать систематические неоднородности экспериментального полигона в случайные и тем самым избежать искажающих воздействий полигона на результаты эксперимента. План эксперимента, включающий и рандомизированное расположение экспериментальных сосудов, приведен в таблице 12.

Таблица 12. План эксперимента ПФЭ2^2 (первая цифра означает номер сосуда, а вторая – номер места на площадке)

№№ сосудов и мест их размещения			x_1	x_2
1>5	2>7	3>3	+	+
4>10	5>9	6>1	-	+
7>4	8>2	9>6	+	-
10>8	11>12	12>11	-	-

В качестве примера ниже приводится статистическая обработка результатов эксперимента только по определению удельных скоростей выделения мидиями аммонийного азота (табл. 13).

Таблица 13. Матрица планирования эксперимента и расчёта коэффициентов уравнения регрессии

№	x_1	x_2	x_1x_2	y_1	y_2	y_3	ỳ	S_g	S_g^2	ỷ	(ỳ – ỷ)^2
1	+	+	+	-0,0106	0,5234	0,3610	0,2913	0,2737	0,0749	0,2924	0,00000
2	-	+	-	0,6612	0,9635	1,3848	1,0032	0,3634	0,1321	0,9394	0,00410
3	+	-	-	1,3966	1,4784	1,3071	1,3940	0,0857	0,0073	1,3932	0,00000
4	-	-	+	1,8223	2,0047	2,1017	1,9762	0,1419	0,0201	2,0400	0,00407
									ΣS_g^2=0,2344		Σ=0,00817

Примечание: ỳ – среднее значение результатов (ПО); ỷ – теоретическое значение ПО, рассчитанное по уравнению регрессии, составленному из статистически значимых коэффициентов; S_g и S_g^2 – стандартное отклонение и дисперсия

Статистическая обработка результатов эксперимента.

1. Число повторных опытов: m= 3.
2. Число разных опытов: g=4.
3. Всего опытов: N = mg = 3·4 =12.
4. Оценка воспроизводимости опытов (по критерию Кохрена): $G_{max} = S^2_{gmax} / \Sigma S_g^2$, где S^2_{gmax} = 0,1321 (максимальная дисперсия в

столбце); $\Sigma S_g^2 = 0{,}2344$ (сумма дисперсий столбца), $G_{max} = S^2_{gmax} / \Sigma S_g^2 = 0{,}1321/0{,}2344 = 0{,}563$.

Табличное значение критерия Кохрена выбирается для (Приложение IV): $\alpha=0{,}05$; $f_1 = g = 4$; $f_2 = m-1=3-1=2$. В таблице для $\alpha=0{,}05$ выбираем 4-ый столбец и 2-ую строку. Получаем $C_2^4=0{,}7457$. Табличное значение больше расчётного; следовательно, опыты воспроизводимы.

5. Расчёт коэффициентов уравнения регрессии: $b_i = (\Sigma\, x_i \grave{y}_g)/g$

$b_0 = (0{,}2913+1{,}0032+1{,}3940+1{,}9762)/4=1{,}1662$

$b_1 = (0{,}2913-1{,}0032+1{,}3940-1{,}9762)/4=-0{,}3235$

$b_2 = (0{,}2913+1{,}0032-1{,}3940-1{,}9762)/4=-0{,}5503$

$b_{12} = (0{,}2913-1{,}0032-1{,}3940+1{,}9762)/4= -0{,}0324$

Итак, получено уравнение регрессии:

$$\mathbf{y = 1{,}1662 - 0{,}3235x_1 - 0{,}5503x_2 - 0{,}0324x_1x_2}$$

Необходимо выполнить проверку статистической значимости каждого коэффициента этого уравнения. Незначимые коэффициенты должны быть отброшены.

Значимость коэффициента уравнения (по Стьюденту).

$t_i = b_i/\sqrt{S^2(b_i)}$, где: $S^2(b_i) = S^2/mg$; $S^2=\Sigma\, S_g^2/g$. $S^2= 0{,}2344/4 = 0{,}0586$; $S^2(b_i)= 0{,}0586/(3\cdot4)= 0{,}00488$; $\sqrt{S^2(b_i)}=0{,}0699$.

Табличный критерий Стьюдента при $\alpha=0{,}05$ и числе степеней свободы (Приложение I):

$v = g(m-1)= 8$; $\mathbf{t^{\alpha=0,05}_8 = 2{,}31}$ (столбец $\alpha=0{,}05$, строка $v=8$).

$t_0 =1{,}1662/0{,}0699 = 16{,}68$ – больше табличного значения, поэтому $\mathbf{b_0}$ значим;

$t_1 = 0{,}3235/0{,}0699=4{,}628$; $\mathbf{b_1}$ **– значим;** $\mathbf{b_2}$ **– значим**;

$t_{12} =0{,}0324/0{,}0699=0{,}46$; $\mathbf{b_{12}}$ **– не значим.**

Окончательное уравнение регрессии имеет вид:

$$\mathbf{y = 1{,}1662 - 0{,}3235x_1 - 0{,}5503x_2}$$

6. Проверка адекватности уравнения выполняется по критерию Фишера. $F= S^2_{ад}/ S^2$, где $S^2_{ад}= (\Sigma(\grave{y} - \mathring{y})^2/(g-d)$, d – число оставшихся членов уравнения.

Итак, $S^2_{ад}= 0{,}00817/1 = 0{,}00817$; $F=0{,}00817/0{,}0586= 0{,}139$

Табличный критерий Фишера для $\alpha=0{,}05$; $f_1= g - d = 4-1=3$; $f_2 = g(m-1) =4(3-1) =8$ (Приложение II, см. в таблице 3-й столбец, 8-ая строка) $F_{табл}=5{,}3$; $F \le F_{табл}$ $0{,}139<5{,}3$. Следовательно, уравнение адекватно описывает результаты эксперимента.

Апостериорный анализ.

Из уравнения $\mathbf{y = 1{,}1662-0{,}3235x_1 - 0{,}5503x_2}$ следует, что:

1. Межфакторное взаимодействие отсутствует. Иными словами, мелкие и крупные мидии (в пределах интервала варьирования их веса!) одинаково реагируют на изменение интенсивности питания, точнее, сходным образом изменяют скорость выделения аммонийного азота.

2. Оба коэффициента основных эффектов – отрицательны. Это означает, что увеличение количества выдаваемого корма x_1 перед экспериментом (в пределах заданного интервала варьирования) снижает удельную скорость выделения аммонийного азота. Возможно, что выделение аммонийного азота происходит в процессе деструкции (диссимиляции), а не при переваривании пищи. Кроме этого, мелкие мидии (в пересчёте на единицу веса) интенсивнее выделяют азот, чем крупные (x_2).
3. Количественная оценка эффектов. Свободный член b_0 =1,1662 характеризует среднюю скорость выделения аммонийного азота, когда все факторы находятся на средних уровнях (x_1 и x_2 =0) и эта скорость равна 1,1662 мг/г×сут. Определим средний эффект фактора x_1, для чего фактор x_2 оставим на среднем уровне равным 0, а x_1 увеличим до +1 и получим у = 1,1662-0,3235= 0,8427. Изменение составило: (0,3235/1,1662)×100 % = 27,7 %. Полный (или глобальный) эффект фактора x_1 получится, если его варьировать с нижнего уровня (-1) и до верхнего (+1). Очевидно, что полный эффект в 2 раза больше среднего и равен 55,5 %. Аналогично определяется эффекты фактора x_2: средний равен (0,550/1,1662)·100 % =47 %, а полный – 94 %, что говорит о сильной роли фактора «индивидуальный вес мидии», понижающего интенсивность выделения аммонийного азота.
4. Расчёт промежуточных значений (интерполяция). Например, определим удельную скорость выделения аммонийного азота при условии, что x_1= 1 мг/г×сут и x_2 =10 г

Для данных уравнений значения натуральных переменных надо кодировать по формулам: для фактора x_1 формула для кодирования: x_1 = (x'_1 – 2,32)/1,55 = (1 – 2,32)/1,55 = -0,85

Для фактора x_2 формула для кодирования: x_2 = (x'_2 – 9,75)/7,25 = (10 – 9,75)/7,25 = 0,03

Здесь x_1 – кодированное значение переменной

x'_1 – натуральное значение переменной.

Итак, у = 1,1662 – 0,3235·(-0,85) – 0,5503·0,03 = 1,1662+0,2750 – 0,0165 = 1,4247

2.2.4. Пример. Полный факторный эксперимент ПФЭ2^3.

Изучение зависимости выживаемости (в %) личинок устриц от концентрации кормовых водорослей; плотности посадки личинок и температуры морской воды.

Перед постановкой эксперимента был составлен список подозреваемых факторов, и было обсуждено возможное влияние 11 подозреваемых факторов на выживаемость личинок. После обсуждения всех подозреваемых факторов было решено включить в предварительный эксперимент три фактора (табл. 14):

Таблица 14. Список факторов, включаемых в реальный эксперимент

№	Наименование фактора	Обозначение	Размерность	Область определения	Область интереса	Ошибка измерения	Примечание
1	Концентрация кормовых водорослей	x_1	Тысяч кл/мл	0-1000	50-150	5	Камера Горяева, 0,01мл^3
2	Концентрация личинок	x_2	Экз./л	0-50 000	1000-3000	100	Камера Богорова 10 мл
3	Температура воды	x_3	°C	0-100	20-26	0,1	

Кодирование факторов и план эксперимента представлены в табл. 15 и 16.

Таблица 15. Кодирование факторов.

Факторы	x_i	x_1	x_2	x_3	x'_i
Верхний уровень	$x_{i\ вер}$	150	3000	26	+1
Нижний уровень	$x_{i\ ниж}$	50	1000	20	-1
Базовый уровень	$x_{i\ 0}$	100	2000	23	0
Шаг варьирования	λ_i	50	1000	3	+1

Таблица 16. Матрица плана эксперимента ПФЭ 2^3 и расчёта коэффициентов уравнения регрессии.

№	Св. чл.	План эксперим.			Межфакторные взаимодействия				Результаты эксперимента				Статобработка			
	x_0	x_1	x_2	x_3	x_1x_2	x_2x_3	x_1x_3	$x_1x_2x_3$	y_1	y_2	y_3	ỳ	S_g	S_g^2	ỷ	(ỳ – ỷ)^2
1	+1	+1	+1	+1	+1	+1	+1	+1	22,4	10,2	10,9	14,5	6,8	46,4	11,8	7,0
2	+1	-1	+1	+1	-1	+1	-1	-1	53,5	12,4	31,1	32,3	20,6	422,7	32,9	0,33
3	+1	+1	-1	+1	-1	-1	+1	-1	6,9	16,6	4,1	9,2	6,6	43,2	11,8	7,2
4	+1	-1	-1	+1	+1	-1	-1	+1	43,7	19,1	37,6	33,5	12,7	163,6	32,9	0,33
5	+1	+1	+1	-1	+1	-1	-1	-1	23,5	58,9	45,2	42,6	17,9	319,5	40,9	2,72
6	+1	-1	+1	-1	-1	-1	+1	+1	46,9	38,8	45,1	43,6	4,3	18,1	41,2	5,56
7	+1	+1	-1	-1	-1	+1	-1	+1	44,1	47,7	26,0	39,3	11,6	135,3	40,1	2,68
8	+1	-1	-1	-1	+1	+1	+1	-1	51	39,6	26,0	38,9	12,5	156,7	41,2	5,52

$\Sigma S_g^2 = 1305,4$ Σ=28,6

Статистическая обработка.

Число повторных опытов: m = 3
Число разных опытов: g = 8

Всего опытов: $N = mg = 3\cdot 8 = 24$

Оценка воспроизводимости опытов:

$G_{max} = S^2_{gmax} / \Sigma S_g^2 = 422{,}7 / 1305{,}43 = 0{,}22$

Табличный критерий Кохрена выбирается для: $\alpha=0{,}05$; $f_1 = g = 8$; $f_2 = m-1=3-1=2$.

В таблице для $\alpha=0{,}05$ выбираем 8-ой столбец и 2-ую строку. Получаем $C_2^8 = 0{,}815$. Табличное значение больше расчётного; опыты воспроизводимы.

Расчёт коэффициентов уравнения регрессии: $b_i = (\Sigma x_i y_g)/g$

$b_0 = 31{,}71$; $b_1 = -5{,}34$; $b_2 = 1{,}52$; $b_3 = -9{,}35$; $b_{12} = 0{,}64$; $b_{23} = -0{,}48$; $b_{13} = -5{,}18$; $b_{123} = 0{,}99$.

Итак, получено уравнение регрессии:

$y = 31{,}71 - 5{,}34x_1 + 1{,}52\, x_2 - 9{,}35x_3 + 0{,}64\, x_1x_2 - 5{,}18x_1x_3 - 0{,}48\, x_2x_3 + 0{,}99\, x_1x_2x_3$.

Проверка значимости коэффициентов уравнения регрессии (проводится по критерию Стьюдента)

$t_i = b_i / \sqrt{S^2(b_i)}$, где: $S^2(b_i) = S^2/mg$; $S^2 = \Sigma S_g^2/g$

$S^2 = 1305{,}43/8 = 163{,}18$; $S^2(b_i) = 163{,}18/(3\cdot 8) = 6{,}8$; $\sqrt{S^2(b_i)} = 2{,}6$.

Табличный критерий Стьюдента при $\alpha=0{,}05$ и числе степеней свободы: $v = g(m-1) = 16$; $t^{\alpha=0,05}_{16} = 2{,}12$.

$t_0 = 31{,}71/2{,}6 = 12{,}2$ – больше табличного значения, поэтому b_0 – значим;

$t_1 = 5{,}34/2{,}6 = 2{,}05 = 2{,}12$, следовательно b_1 – значим и т.д.

Статистически значимыми являются коэффициенты: b_0, b_1, b_3, b_{13}; окончательное уравнение регрессии имеет вид:

$$\mathbf{y = 31{,}71 - 5{,}34x_1 - 9{,}35x_3 - 5{,}18x_1x_3}$$

2.2.5. Апостериорный анализ

Анализ уравнения: линейные эффекты и эффекты взаимодействий. Интерполяция и экстраполяция: кодирование и переход к натуральным переменным.

Для расчётов кодировать переменные по формуле: $x'_i = (x_i - x_{i0})/\lambda_i$

Итак, уравнение регрессии:

$$\mathbf{y = 31{,}71 - 5{,}34x_1 - 9{,}35x_3 - 5{,}18x_1x_3}$$

Анализ уравнения.

1. В среднем по истечении недели выживает 32 % личинок
2. Изменение концентрации личинок в пределах 1000 – 3000 экз./л не влияет на их выживаемость.
3. Увеличение концентрации корма в пределах 50 – 150 тыс. кл./мл, а также температуры в интервале 21,5 – 26°С понижают выживаемость.
4. Имеется отрицательное межфакторное взаимодействие концентрации корма и температуры воды, которое снижает выжи-

ваемость при низкой температуре и малой концентрации корма, а также при высокой температуре и высокой концентрации корма. Этот эффект положителен при низкой температуре и высокой концентрации корма, а также при высокой температуре и низкой концентрации корма.

5. Наилучшие условия для выживания личинок: низкая концентрация корма (50 тыс. кл/мл) и низкая температура (20°С), когда выживаемость равна: $y = 31{,}71 - 5{,}34(-1) - 9{,}35(-1) - 5{,}18(-1)(-1) = 41{,}22\,\%$. Выживаемость самая низкая при концентрации корма 150 тыс.кл/мл и температуре 26°С: $y = 31{,}71 - 5{,}34(+1) - 9{,}35(+1) - 5{,}18(+1)(+1) = 11{,}84\,\%$.
6. Средние вклады факторов: увеличение температуры снижает выживаемость на 18,7 %, что составляет выше половины средней выживаемости; увеличение концентрации корма снижает выживаемость на 10,6 %, что составляет $(10{,}6/31{,}71)\times100\,\% = 30\,\%$ от средней выживаемости. Межфакторное взаимодействие может понижать, либо повышать среднюю выживаемость на 10 %, в зависимости от сочетания уровней факторов, описанных в п.4.
7. Для увеличения выживаемости личинок нужно понизить, прежде всего, температуру воды, а также концентрацию корма.
8. При использовании данного уравнения необходимо кодировать значения факторов по формуле: $x'_i = (x_i - x_{i0})/\lambda_i$
9. Для опубликования необходимо уравнение в кодированных переменных преобразовать в уравнение с натуральными переменными, подставив в уравнение вместо кодированных переменных их натуральный эквивалент:

$$y = 31{,}71 - 5{,}34x'_1 - 9{,}35x'_3 - 5{,}18x'_1x'_3 = 31{,}71 - 5{,}34(x_1 - x_{10})/\lambda_1 - 9{,}53(x_3 - x_{30})/\lambda_3 - 5{,}18((x_1 - x_{10})/\lambda_1)((x_3 - x_{30})/\lambda_3) = 31{,}71 - 5{,}34(x_1 - 100)/50 - 9{,}53(x_3 - 23{,}75)/2{,}25 - 5{,}18((x_1 - 100)/50)((x_3 - 23{,}75)/2{,}25)$$

10. В уравнении отсутствуют два двойных взаимодействия и одно тройное, поэтому:

- Вместо тройного и двух двойных взаимодействий можно ввести в эксперимент ещё три фактора;
- Либо поставить другой эксперимент, в котором будут расширены интервалы варьирования факторов x_1 и x_2.

Задание для усвоения материала по ПФЭ2^n

Зависимость суточного прироста (мкм/сут) личинок мидий от плотности посадки личинок (x_1), концентрации корма (x_2) и температуры воды (x_3).

Задание: даны ПО – суточный прирост личинок мидий и факторы, возможно, влияющие на ПО. Необходимо установить зависимость ПО от этих факторов.

Список факторов, включаемых в реальный эксперимент.

№	Наименование фактора	Обозначение	Размерность	Область определения	Область интереса	Ошибка измерения	Примечание
1	Плотность посадки личинок	x_1	Экз/л	0-100000	2000-9000	100	Камера Богорова 10 мл
2	Концентрация кормовых водорослей	x_2	Тысяч кл/мл	0-500	30 – 60	100	Камера Горяева 0,01мл^3
3	Температура воды	x_3	°C	0-100	17-19	0,1	

Выполнить самостоятельно кодирование факторов и сверить результаты кодирования с нижеприведенной таблицей.

Кодирование факторов (ответ)

Факторы	x_i	x_1	x_2	x_3	x_i
Верхний уровень	$x_{i\ вер}$	9000	60	19	+1
Нижний уровень	$x_{i\ ниж}$	2000	30	17	-1
Базовый уровень	$x_{i\ 0}$	5500	45	18	0
Шаг варьирования	λ_i	3500	15	1	+1

Проверить правильность цифр, указанных курсивом в Матрице планирования.

Матрица планирования и расчёта коэффициентов уравнения регрессии:

№	x_0	x_1	x_2	x_3	x_1x_2	x_1x_3	x_2x_3	$x_1x_2x_3$	y_1	y_2	y_3	$y_{ср}$	S^2_g	$y_т$	$(y_{ср}- y_т)^2$
1	1	+	+	+	+	+	+	+	6,7	6,3	5,7	*6,2*	*0,25*	*6,2*	*0*
2	1	-	+	+	-	-	+	-	17,4	17,6	18,3	*17,8*	*0,22*	*17,8*	*0*
3	1	+	-	+	-	+	-	-	6,4	5,2	5,0	*5,5*	*0,58*	*5,5*	*0*
4	1	-	-	+	+	-	-	+	11,4	10,9	10,5	*10,9*	*0,20*	*10,9*	*0*
5	1	+	+	-	+	-	-	-	5,6	5,8	6,4	*5,9*	*0,18*	*5,9*	*0*
6	1	-	+	-	-	+	-	+	12,9	11,8	12,4	*12,4*	*0,30*	*12,4*	*0*
7	1	+	-	-	-	-	+	+	7,1	7,5	6,4	*7,0*	*0,31*	*7,0*	*0*
8	1	-	-	-	+	+	+	-	9,4	11,1	10,0	*10,2*	*0,74*	*10,2*	*0*
													Σ2,78		Σ0

Статобработку и апостериорный анализ выполнить самостоятельно

Статистическая обработка (ответ)

Число повторных опытов: m = 3
Число разных опытов: g = 8
Всего опытов: $N = mg = 3 \cdot 8 = 24$
Оценка воспроизводимости опытов:
$G_{max} = S^2_{gmax} / \Sigma S_g^2 = 0,74 / 2,78 = 0,27$
Табличный критерий Кохрена выбирается для:
α=0,05; f_1= g = 8; f_2 = m-1=3-1=2. *В таблице для α=0,05 выбираем 8-ой столбец и 2-ую строку. Получаем* C_2^8= 0,815. *Табличное значение больше расчётного – опыты воспроизводимы.*
Расчёт коэффициентов уравнения регрессии: $b_i = (\Sigma x_i y_g)/g$
$b_0 = 9{,}49$; $b_1 = -3{,}34$; $b_2 = 1{,}088$; $b_3 = 0{,}613$; $b_{12} = -1{,}19$; $b_{23} = 0{,}813$; $b_{13} = -0{,}91$; $b_{123} = -0{,}363$
Итак, получено уравнение регрессии:
$y = 9{,}49 - 3{,}34x_1 + 1{,}088x_2 + 0{,}613x_3 - 1{,}19\, x_1x_2 - 0{,}91x_1x_3 + 0{,}813\, x_2x_3 - 0{,}363\, x_1x_2x_3$
Проверка значимости коэффициентов уравнения регрессии (проводится по критерию Стьюдента)
$t_i = b_i / \sqrt{S^2(b_i)}$, где : $S^2(b_i) = S^2/mg$; $S^2 = \Sigma S_g^2/g$
$S^2 = 2{,}78/8 = 0{,}3475$; $S^2(b_i) = 0{,}3475/(3 \cdot 8) = 0{,}014$; $\sqrt{S^2(b_i)} = 0{,}12$.
Табличный критерий Стьюдента при α=0,05 и числе степеней свободы: v = g(m-1)= 16 *(см. таблицы Приложений).*
$t^{\alpha=0,05}_{16} = 2{,}12$; $t_0 = 9{,}49/0{,}12 = 79$ *– больше табличного значения, поэтому* b_0 *– значим;* $t_1 = 3{,}34/0{,}12 = 27{,}8$; b_1 *(значим) и т.д.* $b_{123} = 0{,}363/0{,}12 = 3{,}025$ *– значим*
Статистически значимыми являются все коэффициенты, поэтому окончательное уравнение регрессии имеет вид:
$y = 9{,}49 - 3{,}34x_1 + 1{,}088x_2 + 0{,}613x_3 - 1{,}19\, x_1x_2 - 0{,}91x_1x_3 + 0{,}813\, x_2x_3 - 0{,}363\, x_1x_2x_3$
Проверка адекватности полученного уравнения опытным данным (по критерию Фишера) в данном случае не выполняется, так как значимыми являются все коэффициенты уравнения регрессии.
Апостериорный анализ.
Анализ уравнения: линейные эффекты и эффекты взаимодействий. Интерполяция и экстраполяция: кодирование и переход к натуральным переменным.
Для расчётов кодировать переменные: $x'_i = (x_i - x_{i0}) / \lambda_i$
<u>Определить прирост личинок</u> при x_1=3000; x_2=45; x_3=17.
x_1 = (3000-5500)/3500 = -0,71; x_2 = (45-45)/15 = 0;
x_3 = (17-18)/1 = – 1; y=10,6 мкм/сут.
Итак, уравнение регрессии:
$y = 9{,}49 - 3{,}34x_1 + 1{,}088x_2 + 0{,}613x_3 - 1{,}19\, x_1x_2 - 0{,}91x_1x_3 + 0{,}813\, x_2x_3 - 0{,}363\, x_1x_2x_3$

Анализ уравнения.

1. *В среднем суточный прирост составляет 9,5 мкм*
2. *Изменение концентрации личинок в пределах 2000 – 9000 экз./л отрицательно влияет на их рост, а именно снижает величину прироста в среднем на 66%.*
3. *Увеличение концентрации корма в пределах 30 – 60 тыс. кл./мл, а также температуры в интервале 17 – 19 ^{0}С положительно влияют на рост личинок*

2.3. Дробный факторный эксперимент ДФЭ $2^{k–q}$

В заголовке раздела указана формула ДФЭ $2^{k–q}$, в которой «k» обозначает общее количество факторов в эксперименте; «q» – число вводимых (то есть дополнительных) факторов.

Почему возникла необходимость разработки дробных факторных экспериментов? Дело в том, что с увеличением количества факторов (k) в эксперименте число разных опытов (g) в ПФЭ2^k и тем более в ПФЭ3^k быстро увеличивается (табл. 17).

Таблица 17. Количество разных опытов в ПФЭ с разным числом факторов

k =	2	3	4	5	6	7
ПФЭ2^k, g =	4	8	16	32	64	128
ПФЭ3^k, g =	9	27	81	243	729	2187

Но при увеличении количества факторов в уравнении регрессии появляются члены взаимодействий высших порядков, которые, как правило, незначительно отличаются от 0, что, например, было показано в последнем примере. Это означает, что столбец для взаимодействий высших порядков можно использовать для введения в эксперимент нового фактора и по данному столбцу рассчитывать эффект воздействия нового фактора на ПО. Конечно, эффект нового фактора окажется смешанным с эффектом выбранного взаимодействия высшего порядка. Но, поскольку последний незначителен, можно считать, что мы определяем почти чистый эффект нового фактора. Таким образом, можно считать, что задача уменьшения числа разных опытов в эксперименте выполнена, причём в новом плане сохранились следующие важные статистические свойства:

1. $\Sigma x_g=0$ свойство симметричного расположения уровней факторов относительно центра эксперимента.
2. $\Sigma x_{ig}x_{jg}=0$ свойство ортогональности: скалярное произведение любых двух вектор-столбцов равно 0 (сумма парных произведений элементов двух столбцов равна нулю).

3. $\Sigma x^2_{ig}=g$ свойство нормировки: сумма квадратов элементов любого столбца равна числу разных опытов.

Дисперсии коэффициентов уравнения регрессии, определяемые с помощью таких планов, минимальны и равны. Свойство ортогональности обеспечивает также независимость определения коэффициентов регрессии. Свойство рототабельности (ошибка предсказанных значений по уравнению регрессии одинакова для точек, равноудалённых от центра эксперимента). Если уравнение регрессии используется для продвижения к оптимуму, в этом случае свойство рототабельности оказывается полезным, так как неизвестно, где в факторном пространстве лежит оптимум, то есть неизвестно, в каком направлении нужно двигаться к оптимуму. Иными словами, неизвестно, как нужно изменять коэффициенты уравнения.

Для уменьшения числа опытов можно использовать части ПФЭ, называемые дробными репликами. При этом, если использовать специальным образом построенные части ПФЭ, кратные 2 (1/2 реплики; 1/4 реплики; 1/8 реплики и т.д.), тогда перечисленные свойства планов сохраняются.

Если в плане для ПФЭ 2^2 использовать столбец двойного взаимодействия для введения дополнительного фактора x_3, ($\mathbf{x_3 = x_1x_2}$) – получим план, называемой полурепликой от ПФЭ 2^3, который обозначается как **ДФЭ$\mathbf{2^{3-1}}$.** Почему полуреплика? Потому что в качестве экспериментального плана взята половина плана ПФЭ2^3; другую половину можно получить, приравняв $\mathbf{x_3 = -x_1x_2}$.

В таком укороченном плане разные эффекты имеют одинаковые столбцы, поэтому эффекты смешаны. Например, столбец $\mathbf{x_1}$ повторяет столбец $\mathbf{x_2x_3}$ и т.д. Следовательно, свойства плана сохраняются, но при этом теряется его информативность. Эта потеря компенсируется большим выигрышем в объёме исследований (табл. 18).

Таблица 18. Сокращение числа опытов при выборе реплик высокой дробности.

Число факторов	Дробная реплика	Условные обозначения	Число опытов ДФЭ	Число опытов ПФЭ
3	1/2 реплики от ПФЭ2^3	ДФЭ 2^{3-1}	4	8
4	1/2 реплики от ПФЭ2^4	ДФЭ 2^{4-1}	8	16
5	1/4 реплики от ПФЭ2^5	ДФЭ 2^{5-2}	8	32
6	1/4 реплики от ПФЭ2^6	ДФЭ 2^{6-2}	16	64
7	1/8 реплики от ПФЭ2^7	ДФЭ 2^{7-3}	16	128
14	1/1024 реплики от ПФЭ2^{14}	ДФЭ 2^{14-10}	16	16384
15	1/2048 реплики от ПФЭ2^{15}	ДФЭ 2^{15-11}	16	32768

Определение условий смешивания эффектов в матрице планирования ДФЭ.

Условия смешивания эффектов задаются **определяющим контрастом (ОК),** который получается из **генерирующего соотношения (ГС).**

ГС – это соотношение, показывающее какое факторное взаимодействие заменено вводимым фактором.

ОК – это соотношение, получаемое от умножения ГС на выражение, стоящее в левой части ГС.

Итак, ГС**:** $\mathbf{x_4=x_1x_2x_3}$ Для определения ОК умножим обе части уравнения на $\mathbf{x_4}$:

ОК: $\mathbf{x_4x_4= x_1x_2x_3x_4\,;\,1= x_1x_2x_3x_4}$

С помощью определяющего контраста находим, с каким эффектом смешан интересующий нас эффект. Для этого умножаем интересующий эффект на обе части ОК.

Например, для $\mathbf{x_1}$: $\mathbf{x_1 = x^2{}_1x_2x_3x_4}$; $\mathbf{x_1= x_2x_3x_4}$, так как $\mathbf{x^2{}_i}$ всегда равен 1.

Для $\mathbf{x_1x_2}$: $\mathbf{x_1x_2 = x_3x_4}$

Если в план эксперимента включить ещё один фактор, то получим четверть-реплику от **ПФЭ2**^{5}. Такой план называется **ДФЭ2**$^{5\text{-}2}$**.** Вводим дополнительный фактор $\mathbf{x_5}$; ГС: $\mathbf{x_5= x_1x_2}$, тогда ОК: $\mathbf{1= x_1x_2x_5}$ Перемножим оба ОК: $\mathbf{1 = x^2{}_1x^2{}_2x_3x_4\ x_5 = x_3x_4\ x_5}$.

Общий ОК получается в результате сложения правых частей всех трёх ОК:

$\mathbf{1= x_1x_2x_3x_4= x_1x_2x_5=x_3x_4\ x_5}$

В этом случае оценка коэффициента b_1 оказывается смешанной с оценками ещё трёх коэффициентов**:** $\mathbf{b_1\rightarrow\beta_1+ \beta_{234}+ \beta_{25}+ \beta_{1345}}$

2.3.1. Реплики большой дробности

При выборе 1/8 реплики $2^{6\text{-}3}$ можно воспользоваться вектор-столбцами трёх взаимодействий:

$x_4= x_1x_2\ x_5= x_1x_3\ x_6= x_2x_3$
$x_4= x_1x_3\ x_5= x_2x_3\ x_6= x_1x_2x_3$
$x_4= x_1x_2\ x_5= x_2x_3\ x_6= x_1x_2x_3$
$x_4= x_1x_2\ x_5= x_1x_3\ x_6= x_1x_2x_3$

Для каждого из этих решений можно сделать 6 перестановок. Всего получается 24 возможности выбора 1/8 реплики (если всюду выбираем положительные знаки).

Наименее удачен первый вариант, так как здесь все линейные эффекты смешиваются с парными взаимодействиями.

Если для нас важно сохранить взаимодействие x_1x_2, в этом случае нужно выбрать второй вариант и т.д. Допустим, мы выбрали **4-й вариант**, причём известно, что наиболее важный фактор – это x_4, тогда записываем:

$\mathbf{x_4 = x_1x_2x_3\,;\, x_5 = x_1x_2\,;\, x_6 = x_1x_3}$. Это записаны ГС для 1/8 реплики.

ОК: $\mathbf{1 = x_1x_2x_3x_4 \quad 1 = x_1x_2x_5 \quad 1 = x_1x_3x_6}$

Если попарно перемножить эти определяющие контрасты, получим:

$\mathbf{1 = x_3x_4x_5 \quad 1 = x_2x_4x_6 \quad 1 = x_2x_3x_5x_6}$. Произведение трёх (первых) ОК равно: $\mathbf{1 = x_1x_4x_5x_6}$.

Обобщающий определяющий контраст:

$\mathbf{1 = x_1x_2x_3x_4 = x_1x_2x_5 = x_1x_3x_6 = x_3x_4x_5 = x_2x_4x_6 = x_2x_3x_5x_6 = x_1x_4x_5x_6}$

Получается следующая система смешивания <u>(если мы пренебрежём коэффициентами, начиная с четверных взаимодействий):</u>

$\mathbf{b_1 \rightarrow \beta_1 + \beta_{25} + \beta_{36} + \beta_{234} + \beta_{456}}$
$\mathbf{b_2 \rightarrow \beta_2 + \beta_{15} + \beta_{46} + \beta_{134} + \beta_{356}}$
$\mathbf{b_3 \rightarrow \beta_3 + \beta_{16} + \beta_{45} + \beta_{124} + \beta_{256}}$
$\mathbf{b_4 \rightarrow \beta_4 + \beta_{35} + \beta_{26} + \beta_{123} + \beta_{156}}$
$\mathbf{b_5 \rightarrow \beta_5 + \beta_{12} + \beta_{34} + \beta_{236} + \beta_{146}}$
$\mathbf{b_6 \rightarrow \beta_6 + \beta_{13} + \beta_{24} + \beta_{235} + \beta_{145}}$

Дробные реплики могут быть насыщенными, либо ненасыщенными. Если число опытов равно числу определяемых линейных эффектов, включая свободный член, то такой план называется **насыщенным** (вместо всех взаимодействий ввели новые факторы); если число опытов больше – то план **ненасыщенный** (например, ПФЭ); если число опытов меньше числа определяемых линейных эффектов, то план называется **сверхнасыщенным.**

Реплика, полученная при введении в план нового фактора за счёт высшего взаимодействия, **называется главной полурепликой**.

С увеличением дробности реплики (1/4; 1/8; 1/16 и т.д.) разрешающая способность плана уменьшается, так как получаемые оценки коэффициентов уравнения регрессии смешиваются с эффектами других факторов и их взаимодействий. **Разрешающая способность реплики** – это число несмешанных линейных эффектов. Разрешающая способность реплики соответствует минимальному количеству факторов, записанных в ОК. Если линейные эффекты смешаны с эффектами взаимодействий максимально высокого порядка, то такая реплика обладает максимальной разрешающей способностью и соответствует она **главной реплике**. Реплики высокой дробности применяют на первых этапах экспериментальных исследований. Такие планы служат для отбора наиболее существенных факторов, либо для предварительной оптимизации условий путём отбора наиболее интересных вариантов, которые будут исследоваться далее более подробно (аналогия с микроскопированием).

Схема расчётов обобщённого определяющего контраста **ООК**:

1) ООК для 1/2 реплики (вводим один фактор):

1. Записываем ГС;

2. Умножив левую и правую часть ГС на выражение, записанное в левой части, получаем ОК.

2) ООК для 1/4 реплики (вводим два фактора):
1. Записываем два ГС;
2. Из двух ГС соответственно получаем два ОК;
3. Перемножаем оба ОК и получаем третий ОК;
4. Взяв все три ОК, получаем ООК.

3) ООК для 1/8 реплики (вводим три фактора):
1. Записываем три ГС;
2. Из трёх ГС получаем соответственно три ОК;
3. Перемножаем попарно все ОК и получаем ещё три ОК;
4. Перемножаем три начальных ОК (из п.2) и получаем ещё один ОК;
5. Записываем ООК, у которого в правой части записаны все ОК (семь штук).

Разбиение плана эксперимента на блоки.

При наличии источников неоднородности (различий в сырье, дрейфа ОИ и т.д.) и невозможности постановки всех опытов одновременно, матрицу планирования можно разбить на блоки и ставить опыты по каждому блоку отдельно.

Допустим, что мы проводим эксперимент по плану ПФЭ2^3. Разбиваем этот план на два блока в соответствии со знаками для столбца тройного взаимодействия. При этом в первый блок войдут все опыты, которым в столбце тройного взаимодействия соответствует знак плюс, а во второй блок – знак минус. В результате опыты первого блока выполняются в первый день (или с одним сырьём), а опыты, вошедшие во второй блок, – реализуем в последующий период (или с другим сырьём). При этом эффект дрейфа ОИ, либо эффект от неоднородности сырья будет смешан с эффектом тройного взаимодействия, а линейные эффекты будут защищены.

2.3.2. Пример ДФЭ 2^{k-q}

Во втором примере было установлено, что в результате исследований зависимости выживаемости личинок устриц (y) от трёх факторов (x_1 – концентрация корма; x_2 – плотность посадки личинок; x_3 – температура морской воды) получено уравнение регрессии:

$$\mathbf{y = 31{,}71 - 5{,}34x_1 - 9{,}35x_3 - 5{,}18x_1x_3}$$

Статистически незначимыми оказались тройное взаимодействие $x_1x_2x_3$ и два двойных взаимодействия, включающих фактор x_2 – x_1x_2 и x_2x_3. Вместо столбцов для этих взаимодействий будем вводить новые факторы:

1. $x_4 = x_1x_2x_3$. Получаем полуреплику от ПФЭ 2^4, которая обозначается ДФЭ2^{4-1}.

Итак, ГС: $x_4=x_1x_2x_3$. Для определения ОК умножим обе части уравнения на x_4:

ОК: $\mathbf{x_4x_4= x_1x_2x_3x_4}$ $\mathbf{1= x_1x_2x_3x_4}$

С помощью определяющего контраста находим, с каким эффектом смешан интересующий нас эффект. Для этого умножаем интересующий эффект на обе части ОК.

Например, для x_1: $x_1= x^2{}_1x_2x_3x_4$; $x_1= x_2x_3x_4$; для $x_2= x_1x^2{}_2x_3x_4 = x_1x_3x_4$; $x_3 = x_1x_2x_4$

$b_1 \rightarrow \beta_1 + \beta_{234}$

$b_2 \rightarrow \beta_2 + \beta_{134}$

$b_3 \rightarrow \beta_3 + \beta_{124}$

Для x_1x_2: $x_1x_2= x^2{}_1x^2{}_2x_3x_4 = x_3x_4$

$x_1x_3= x^2{}_1x_2x^2{}_3x_4 = x_2x_4$

$x_1x_4= x^2{}_1x_2x_3x^2{}_4 = x_2x_3$

$x_2x_4= x_1x^2{}_2x_3x^2{}_4 = x_1x_3$ (смотри вторую строку)

$b_{12} \rightarrow \beta_{12} + \beta_{34}$

$b_{13} \rightarrow \beta_{13} + \beta_{24}$

$b_{14} \rightarrow \beta_{14} + \beta_{23}$

$b_{24} \rightarrow \beta_{24} + \beta_{13}$, что совпадает с $b_{13} \rightarrow \beta_{13} + \beta_{24}$

Для $x_1x_2x_3 = x_4$ (как записано в генерирующем соотношении). Оценки всех тройных взаимодействий смешаны с оценками линейных эффектов.

2. Вводим ещё один фактор: $x_5 = x_1x_2$. Получаем четверть-реплику от ПФЭ2^5. Такой план называется ДФЭ2^{5-2} .

Вводим дополнительный ГС: $x_5= x_1x_2$, (всего два ГС: $x_4= x_1x_2x_3$ и $x_5 = x_1x_2$), тогда два ОК:

$1= x_1x_2x_3x_4$ и $1= x_1x_2x_5$ Перемножим оба ОК: $1 = x^2{}_1x^2{}_2x_3x_4x_5 = x_3x_4x_5$.

Обобщённый **ООК** получается в результате сложения правых частей всех трёх ОК:

$\mathbf{1 = x_1x_2x_3x_4= x_1x_2x_5 = x_3x_4x_5}$. По этому суммарному ОК будем определять условия смешиваемости оценок линейных эффектов факторов:

$x_1 = x^2{}_1x_2x_3x_4 + x^2{}_1x_2x_5 + x_1x_3x_4x_5 = x_2x_3x_4 + x_2x_5 + x_1x_3x_4x_5$.

$b_1 \rightarrow \beta_1 + \beta_{25} + \beta_{234} + \beta_{1345}$

$x_2= x_1x_3x_4 + x_1x_5 + x_2x_3x_4x_5$

$b_2 \rightarrow \beta_1 + \beta_{15} + \beta_{134} + \beta_{2345}$

$x_3= x_1x_2x_4 + x_1x_2x_3x_5 + x_4x_5$

$b_3 \rightarrow \beta_3 + \beta_{45} + \beta_{124} + \beta_{1235}$

$x_4 = x_1x_2x_3 + x_1x_2x_4x_5 + x_3x_5$

$b_4 \rightarrow \beta_4 + \beta_{35} + \beta_{123} + \beta_{1245}$

$x_5= x_1x_2+ x_1x_2x_3x_4x_5+ x_3x_4$

$b_5 \rightarrow \beta_5+ \beta_{34}+ \beta_{12}+ \beta_{1234}$

Статистическая обработка результатов, полученных при постановке ДФЭ, проводится точно так же, как и результатов, полученных при проведении ПФЭ.

2.3.3. Контрольная работа по ДФЭ

Как называется реплика ДФЭ2^{7-3}?
Ответ: 1/8 реплики от ПФЭ2^{7}.

Написать уравнение регрессии, которое рассчитывается по результатам этого эксперимента.

Ответ: уравнение для ПФЭ2^4: $Y= b_0 +b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4+b_{12}x_1x_2+b_{13}x_1x_3+b_{14}x_1x_4+b_{23}x_2x_3+b_{24}x_2x_4+b_{34}x_3x_4+b_{123}x_1x_2x_3+b_{124}x_1x_2x_4+b_{234}x_2x_3x_4+b_{134}x_1x_3x_4 +b_{1234}x_1x_2x_3x_4$

Определить условия смешиваемости оценок для двух линейных коэффициентов, двух парных взаимодействий и двух тройных взаимодействий (любых).

Ответ: Вводим три фактора: $x_5= x_1x_2x_3x_4\ x_6= x_1x_3x_4\ x_7= x_2x_3x_4.$ *Получим уравнение регрессии:* $Y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 + b_5x_5 + b_6x_6 b_7x_7 + b_{12}x_1x_2 + b_{13}x_1x_3+b_{14}x_1x_4+b_{23}x_2x_3+b_{24}x_2x_4+b_{34}x_3x_4+b_{123}x_1x_2x_3+b_{124}x_1x_2x_4+$

ОК: $1= x_1x_2x_3x_4x_5\ 1= x_1x_3x_4x_6\ 1= x_2x_3x_4x_7$

Перемножаем попарно эти три ОК: $1= x_2x_5x_6; 1= x_1x_5x_7; 1= x_1x_2x_6x_7;$

Перемножаем первые три ОК: $1 = x_3x_4x_5x_6x_7$

ООК: $1= x_1x_2x_3x_4x_5 = x_1x_3x_4x_6 = x_1x_2x_6x_7 = x_2x_5x_6 = x_1x_5x_7 = x_1x_2x_6x_7= x_3x_4x_5x_6x_7$

Условия смешиваемости: $b_1 \rightarrow \beta_1 + \beta_{2345} + \beta_{346} + \beta_{267} + \beta_{1256} + \beta_{57} + \beta_{267} + \beta_{134567}$

ДФЭ2^{8-4}

Написать условное обозначение 1/4 реплики от ПФЭ2^5 и 1/8 реплики от ПФЭ2^6.

Ответ: ДФЭ2^{5-2}; ДФЭ2^{6-3};

Написать уравнение регрессии, получаемое в результате реализации 1/8 реплики от ПФЭ2^6.

Ответ: $Y= b_0 +b_1x_1+ b_2x_2+ b_3x_3+ b_4x_4+ b_5x_5 + b_6x_6 + b_{12}x_1x_2$

Определить условия смешиваемости оценок коэффициентов в ДФЭ2^{6-3}

Условия смешиваемости:

Генерирующие соотношения: $x_4= x_1x_2x_3; x_5= x_2x_3; x_6= x_1x_3$

ОК: $1= x_1x_2x_3x_4\ 1= x_2x_3x_5\ 1= x_1x_3x_6$

Перемножаем попарно: $1= x_1x_4x_5\ 1= x_2x_4x_6\ 1= x_1x_2x_5x_6$

Перемножаем три первых ОК: $1= x_3x_4x_5x_6$

ООК: $1 = x_1x_2x_3x_4 = x_2x_3x_5 = x_1x_3x_6 = x_1x_4x_5 = x_2x_4x_6 = x_1x_2x_5x_6 = x_3x_4x_5x_6$

Условия смешиваемости: $b_1 \to \beta_1 + \beta_{1235} + \beta_{36} + \beta_{45} + \beta_{1246} + \beta_{256} + \beta_{267} + \beta_{13456}$

Вопросы для закрепления материала) по ДФЭ 2^{k-q} (усложнённые варианты)

Вариант 1.

Как называется реплика ДФЭ2^{7-3}? Написать уравнение регрессии, которое рассчитывается по результатам этого эксперимента. Определить условия смешиваемости оценок для двух линейных коэффициентов, двух парных взаимодействий и двух тройных взаимодействий (любых).

Написать условное обозначение 1/16 реплики от ПФЭ2^8.

Ответ:

1/8 реплики от ПФЭ2^7.

Уравнение ПФЭ2^4имеет вид:

$Y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{14}x_1x_4 + b_{23}x_2x_3 + b_{24}x_2x_4 + b_{34}x_3x_4 + b_{123}x_1x_2x_3 + b_{124}x_1x_2x_4 + b_{234}x_2x_3x_4 + b_{134}x_1x_3x_4 + b_{1234}x_1x_2x_3x_4$

Вводим три фактора: $x_5 = x_1x_2x_3x_4\, x_6 = x_1x_3x_4\, x_7 = x_2x_3x_4$

Получим уравнение регрессии: $Y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 + b_5x_5 + b_6x_6 + b_7x_7 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{14}x_1x_4 + b_{23}x_2x_3 + b_{24}x_2x_4 + b_{34}x_3x_4 + b_{123}x_1x_2x_3 + b_{124}x_1x_2x_4 +$

ОК: $1 = x_1x_2x_3x_4x_5\; 1 = x_1x_3x_4x_6\; 1 = x_2x_3x_4x_7$

Перемножаем попарно эти три ОК: $1 = x_2x_5x_6\; 1 = x_1x_5x_7\; 1 = x_1x_2x_6x_7$

Перемножаем первые три ОК: $1 = x_3x_4x_5x_6x_7$

ООК: $1 = x_1x_2x_3x_4x_5 = x_1x_3x_4x_6 = x_1x_2x_6x_7 = x_2x_5x_6 = x_1x_5x_7 = x_1x_2x_6x_7 = x_3x_4x_5x_6x_7$

Условия смешиваемости: $b_1 \to \beta_1 + \beta_{2345} + \beta_{346} + \beta_{267} + \beta_{1256} + \beta_{57} + \beta_{267} + \beta_{134567}$

ДФЭ2^{8-4}

Вариант 2.

Написать условное обозначение 1/4 реплики от ПФЭ2^5 и 1/8 реплики от ПФЭ2^6

Написать уравнение регрессии, получаемое в результате реализации 1/8 реплики от ПФЭ2^6.

Определить условия смешиваемости оценок коэффициентов в ДФЭ2^{6-3}

Ответ:

ДФЭ2^{5-2}; ДФЭ2^{6-3};

$Y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 + b_5x_5 + b_6x_6 + b_{12}x_1x_2$

Условия смешиваемости:
Генерирующие соотношения: $x_4 = x_1x_2x_3$ $x_5 = x_2x_3$ $x_6 = x_1x_3$
ОК: $1 = x_1x_2x_3x_4$ $1 = x_2x_3x_5$ $1 = x_1x_3x_6$
Перемножаем попарно: $1 = x_1x_4x_5$ $1 = x_2x_4x_6$ $1 = x_1x_2x_5x_6$
Перемножаем три первых ОК: $1 = x_3x_4x_5x_6$
ООК: $1 = x_1x_2x_3x_4 = x_2x_3x_5 = x_1x_3x_6 = x_1x_4x_5 = x_2x_4x_6 = x_1x_2x_5x_6 = x_3x_4x_5x_6$
Условия смешиваемости: $b_1 \rightarrow \beta_1 + \beta_{1235} + \beta_{36} + \beta_{45} + \beta_{1246} + \beta_{256} + \beta_{267} + \beta_{13456}$

2.3.4. Пример применения методов планирования экспериментов в гидробиологических исследованиях

Постановка задачи

При изучении водных сообществ исследователи обычно определяют такие интегральные показатели экосистемы, как численность и биомасса организмов, концентрации биогенных элементов, гидрохимические характеристики и т.д. Многие из этих показателей представляют собой результаты нескольких одновременно протекающих процессов, которые обычно не учитываются. Но для выяснения причин и факторов, определяющих биомассы, продуктивность конкретных видов, устойчивость экосистем необходимо проводить исследования этих процессов. Сложность изучения обменных процессов в экосистемах значительно облегчается путём исследования простых экспериментальных сообществ в проточных установках. При этом скорости выделения и потребления веществ различными компонентами сообщества, а также эффекты взаимодействий между видами можно определить, используя методы математического планирования экспериментов при варьировании биомасс взаимодействующих организмов. Для определения нетрофического влияния иглокожих на продукционные процессы в прибрежных экосистемах изучено взаимодействие автотрофного (микрофитобентос твёрдого грунта) и гетеротрофного (морские звёзды патирии) компонентов простого экспериментального сообщества посредством внешних метаболитов (Рябушко, Холодов, 1985).

Итак, цель исследований: изучение нетрофических взаимодействий между продуцентами и консументами.

Задачи:

1. Изучение влияния одного компонента экспериментального сообщества на другой и обратно:

- Определение скоростей выделения (потребления) биогенных элементов каждым компонентом без воздействий на него другого компонента.
- То же самое, но под воздействием другого компонента.

2. Одновременное влияние компонентов друг на друга:
- Зависимость скоростей выделения (потребления) биогенов при одновременном взаимном влиянии обоих компонентов друг на друга.
3. Определение условий составления экспериментальных сообществ, сбалансированных по выделяемым и потребляемым соединениям.

Методика исследований

Опыты проводили в августе 1980 г. на биостанции «Восток» Института биологии моря АН СССР (Владивосток) при температуре морской воды (17,6-19,8°С), соответствующей природным условиям.

Морских звёзд патирий и камни, обросшие микрофлорой размером 3-5 см, собирали с глубины 1 метр за 2 часа до начала опытов. Микрофитобентос был представлен, главным образом, пенатными диатомовыми водорослями.

Экспериментальные сосуды объёмом по 0,8 л, общее количество 27 шт. Распределение экспериментальных сосудов: 12 – со звёздами; 12 – с продуцентами; 3 – контроль, то есть без живых компонентов (рис. 6).

Прокачивание воды через сосуды и её перемешивание в сосудах осуществляли многоканальными перистальтическими насосами. Скорость перемешивания на порядок превосходила скорость прокачивания.

Продолжительность эксперимента: 18 – 24 час. Отбор проб воды из экспериментальных сосудов производили через 2 – 5 час.

Опыты ставили по схеме ПФЭ2^2

Варьируемые факторы:

X_1 – масса сырая водорослей (мг), с уровнями: 14,3 (-1) и 19,3 (+1);

X_2 – масса сухая звёзд (г): 2,82 (-1) и 7,77 (+1).

Сырую массу микрофитобентоса вычисляли, используя номограммы (Численко, 1968), а также по содержанию хлорофилла «а» в пробах.

В проточных сосудах концентрации биогенных соединений зависят от нескольких взаимосвязанных процессов: поступления соединений с притоком воды, выделения или потребления их гидробионтами, а также выноса этих соединений с вытекающей водой. Следовательно, изменение во времени (t) концентрации соединений в первом (C_1) и во втором (C_2) респирометрах могут быть описаны системой уравнений:

$$dC_1(t)/dt = V_1(C_0 - C_1(t))/W_1 - R_1/W_1$$
$$dC_2(t)/dt = V_2(C_1(t) - C_2(t))/W_2 - R_2/W_2 ,$$

где: C_0- начальная концентрация во втекающей воде; V_1, V_2 – скорости протока через первый и второй респирометры; W_1 и W_2 – объёмы респирометров; R_1 R_2 – скорости потребления (выделения) соединений организмами.

Эти уравнения проинтегрированы и рассчитаны скорости R_1 и R_2 для кислорода, аммиака и фосфатов.

Результаты

Выделение и потребление кислорода (продукционные и деструкционные процессы).

Соотношение выделения и потребления кислорода (выделение O_2/потребление O_2) отражает соотношение продукции (фотосинтеза) и деструкции (дыхания) в системе.

Почти везде фотосинтез преобладает над деструкцией, что свидетельствует о высокой активности микроводорослей (рис. 7). Отметим, что биомасса микроводорослей была мала по сравнению с биомассой морских звёзд.

В опытах, в которых вода протекала в направлении от продуцента к консументу отмечено значительное повышение концентрации кислорода на выходе из системы из-за интенсивного фотосинтеза микроводорослей (табл. 19). При этом потребление звёздами O_2 не зависит от концентрации O_2. Это можно объяснить тем, что интенсивность дыхания морских звёзд слабо зависит от концентрации O_2. Однако общая скорость выделения O_2 системой сильно зависит от биомассы звёзд.

При обратном направлении тока воды (от морской звезды к микрофитобентосу) в сосуды с микроводорослями поступала обеднённая кислородом вода. Интенсивность выделения микроводорослями O_2 слабо зависит от биомассы звёзд – источников метаболитов. Общая скорость выделения кислорода системой значительно выше, чем в предыдущем опыте.

В серии опытов с взаимным обменом воды между обоими компонентами сообщества невозможно определить скорость выделения или потребления кислорода каждым компонентом в отдельности, поэтому измерялись интегральные характеристики системы. В системе продуцировалось в 4 раза больше кислорода, чем его потреблялось. Получено следующее уравнение регрессии в системе с взаимным обменом компонентов метаболитами (**y**, мл/час):

$$\mathbf{y = 0{,}0379 + 0{,}0063x_1 - 0{,}0621x_2 - 0{,}0128x_1x_2}$$

Положительный свободный член уравнения показывает, что в системе преобладают продукционные процессы. В целом скорость выделения O_2 в системе возрастает при увеличении биомассы микроводорослей и резко снижается при увеличении биомассы морских звёзд. Максимальное продуцирование O_2 наблюдается в присутствие мелких звёзд. Крупные животные интенсивно потребляют кислород, поэтому в системе деструкционные процессы преобладают над продукционными.

Аммоний (процессы деструкции)

В поступающей воде содержится мало NH_4^+, поэтому микрофлора потребляет NH_4^+ с небольшой скоростью. Морские звёзды экскретируют значительные количества аммония, и на выходе из системы отмечаются высокие концентрации аммиака. Скорость выделения NH_4^+ из системы сильно зависит от биомассы компонентов (мкг-ат./час):

$\mathbf{y = 0,8102 + 0,023x_1 + 0,1377x_2 – 0,335x_1x_2}$ (движение от водорослей к звезде).

При обратном направлении потока воды микрофлора интенсивно потребляла аммиак:

$\mathbf{y = 0,74 + 0,1915x_1 - 0,303x_2 – 0,025x_1x_2}$

В замкнутой системе в целом выделялся аммиак:

$\mathbf{y = 0,0243 + 0,07x_1 - 0,155x_2 – 0,077x_1x_2}$

Из результатов опытов с односторонним действием одного компонента на другой следует, что выделения морских звёзд оказывают сильное влияние на скорость потребления NH_4^+ микроводорослями, в то время как обратное влияние значительно слабее. Увеличение массы микрофлоры не вызывает снижения скоростей выделения аммиака из системы. При любых сочетаниях уровней факторов микроводоросли потребляли аммоний, причём скорость потребления в большей мере зависит от массы животных, чем от массы микрофитобентоса. В замкнутой системе всегда идёт накопление аммония, что свидетельствует о несбалансированности системы по аммонию.

Ортофосфаты

Микрофлора во всех опытах потребляет фосфаты. Известно, что звёзды вначале выделяют фосфат (первые 2 – 4 часа), а потом (через 5 – 7 час) потребляют его. Причём это не связано с развитием микроорганизмов в экспериментальных сосудах и трубопроводах. Во всех экспериментах процессы потребления PO_4 в системе значительно преобладают над выделением, причём микрофитобентос потребляет фосфаты интенсивнее, чем звёзды.

При движении воды от микрофитобентоса к звёздам (**У**, мкг-ат/час):

$\mathbf{y = 0,014 – 0,005x_1 – 0,007x_2 + 0,0002x_1x_2}$

При обратном направлении:

$$y = 0{,}0193 + 0{,}0077x_1 - 0{,}0067x_2 - 0{,}0042x_1x_2$$

В замкнутой системе:

$$y = 0{,}009 + 0{,}0235x_1 - 0{,}012x_2 + 0{,}014x_1x_2$$

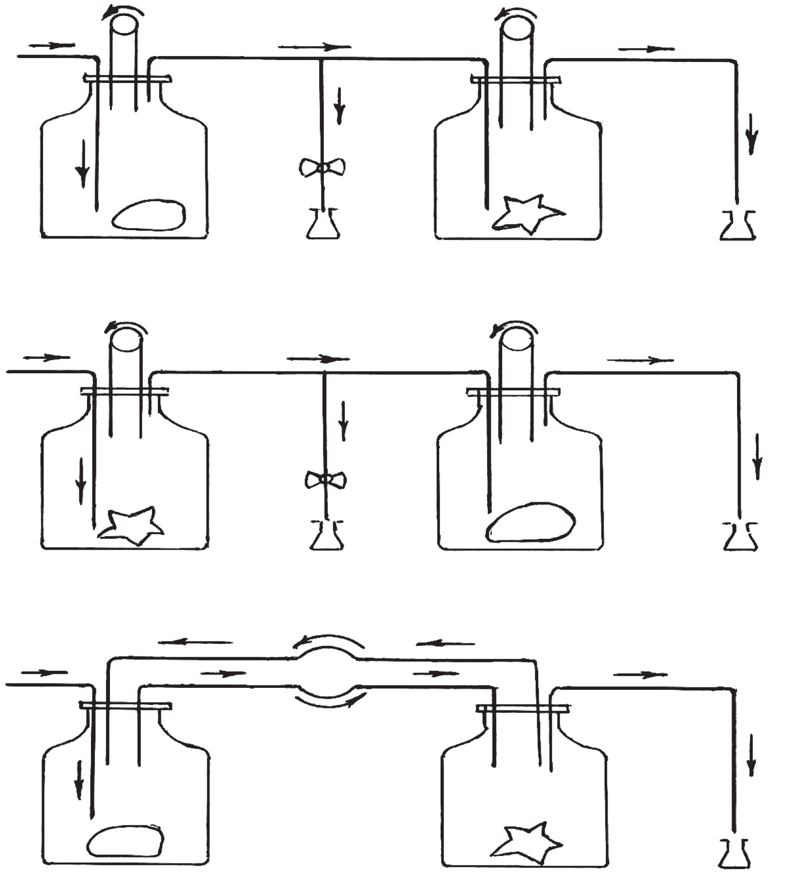


Рис. 6. Схемы проточных экспериментальных установок для изучения взаимного влияния морской звезды и микрофитобентоса посредством внешних метаболитов: вверху – поступление воды от микрофитобентоса к морской звезде; в середине – противоположное направление потока воды; внизу – взаимный обмен воды между компонентами

В последнем случае эффект взаимодействия биомасс обоих компонентов очень велик (больше свободного члена). Установлено, что процессы потребления ортофосфатов в системе ослабляются как фитобентосом, так и морскими звёздами. Поэтому в эксперименте с взаимным обменом воды потребление PO_4 идёт значительно слабее, что обусловлено отрицательными связями между компонентами экспериментального сообщества. Из последнего уравнения следует, что скорость потребления ортофосфатов в «разбалансированной системе», то есть при больших различиях в метаболической активности обоих компонентов, значительно снижается.

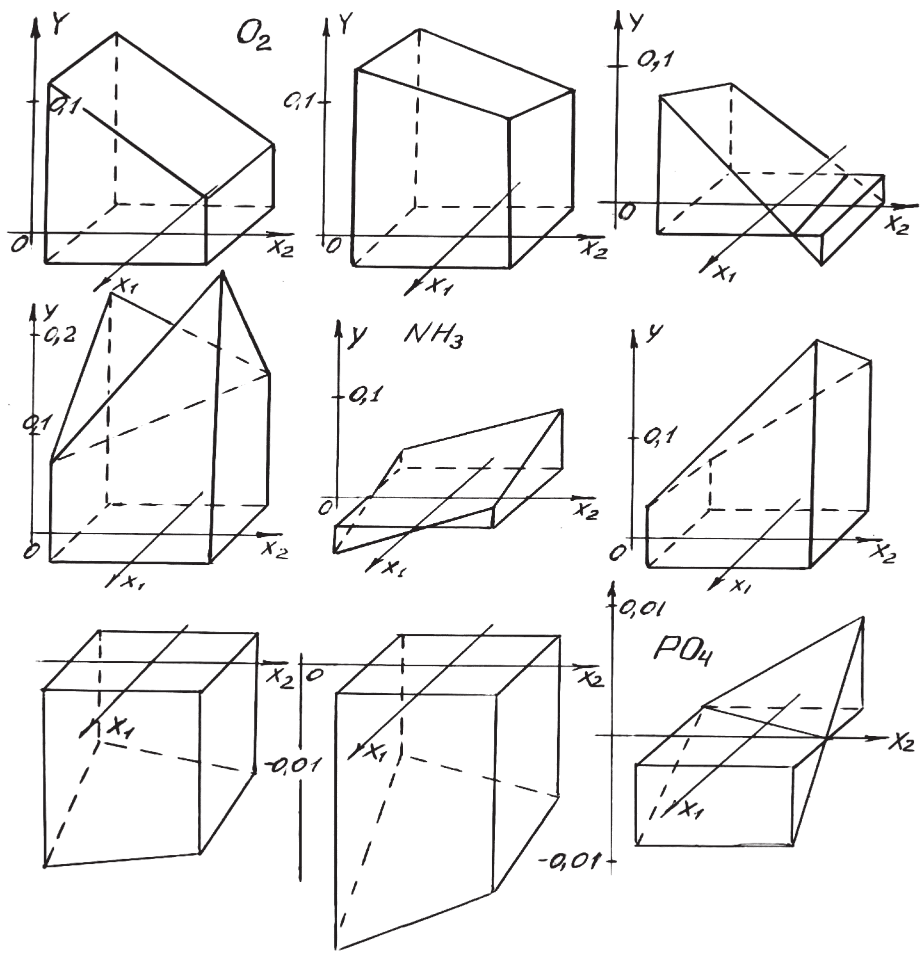


Рис. 7. Скорости выделения кислорода, аммонийного азота и ортофосфатов в системе «продуценты – консументы» (микрофитобентос – морские звёзды). Порядок поступления воды (по столбцам) соответствует схемам, указанным по строкам рисунка 6.

Таблица 19. Скорость потребления (-) или выделения (+) кислорода, аммиака, ортофосфата в системе взаимодействия морских звёзд (З) и микрофитобентоса камней (В) через метаболиты.

Уровень варьирования биомасс		O_2, мл/час			NH_4^+, мкг-ат/час			PO_4^{2-} мкг-ат/час		
В	З	В>З	З>В	В↔З	В>З	З>В	В↔З	В>З	З>В	В↔З
+1	+1	0,048	0,133	-0,031	0,261	0,040	0,545	-0,013	-0,017	0,007
+1	-1	0,014	0,175	0,119	0,072	-0,048	0,080	-0,014	-0,020	-0,006
-1	+1	0,055	0,108	-0,018	0,118	0,094	0,251	-0,012	-0,013	+0,008
-1	-1	0,127	0,146	0,081	0,197	0,003	0,095	-0,009	-0,009	-0,002

Сравнительный анализ энергетического, азотистого и фосфорного метаболизма в экспериментальном сообществе

Потребление O_2 и выделение NH_4^+ отражает интенсивность протекания деструктивных процессов, а потребление PO_4^{2-} – продукционных. Оба компонента экспериментального сообщества связаны между собой положительными связями через выделение и потребление O_2 и CO_2, а также NH_4^+. Отношение обоих компонентов к PO_4^{2-} имеет более сложный характер, а именно: при односторонних воздействиях одного компонента на другой наблюдается конкуренция за PO_4^{2-}, а при взаимном воздействии – симбиоз.

В экспериментах выявлено сильное взаимодействие между компонентами системы, в результате чего происходит либо усиление, либо ослабление их метаболической активности. Взаимодействие приводит к тому, что скорость продуцирования биогенов или их выделения не определяется как результирующая или алгебраическая сумма всех скоростей, полученных в опытах с односторонним воздействием метаболитов.

Конструирование искусственных сообществ, сбалансированных по экзометаболитам

В сбалансированном экспериментальном сообществе метаболиты, продуцируемые одним компонентом, потребляются другим компонентом. Поэтому скорость продуцирования изучаемого метаболита замкнутой системой, содержащей сбалансированное сообщество, оказывается равной нулю. Следовательно, необходимо в уравнении регрессии для замкнутой системы, полученном по одному соединению, приравнять к нулю левую часть этого уравнения. Зафиксировав уровень одного фактора, определяют значение другого фактора. Выполнив эту процедуру по каждому из трёх соединений, можно проследить за изменениями соотношений биомасс компонентов сообщества.

$\mathbf{x_1}$ – сырая масса водорослей: 14,3 (-1) и 19,3 (+1) мг; $\mathbf{x_0}$ = 16,8 мг $\Delta\mathbf{x}$=2,5

$\mathbf{x_2}$- сухая масса звёзд: 2,82 (-1) и 7,77 (+1)г $\mathbf{x_0}$= 5,295 г $\Delta\mathbf{x}$=2,475

По кислороду: $0 = 0{,}0379 + 0{,}0063\mathbf{x_1} - 0{,}0621\mathbf{x_2} - 0{,}0128\mathbf{x_1 x_2}$. Примем, что $\mathbf{x_1} = 0$ (т.е. 16,8), тогда уравнение будет иметь вид: $0 = 0{,}0379 - 0{,}0621\mathbf{x_2}$

$\mathbf{x_2}$ = 0,0379: 0,621= 0,061 (это кодированное значение). Переводим в натуральную форму: $x^\wedge = x_0 + x\cdot\Delta\mathbf{x_0}$= 5,295 + 0,061·2,475= 5,45г.

Таким образом, система будет полностью сбалансированной по кислороду, если сырая биомасса микроводорослей равна **16,8 мг**, а сухая биомасса звёзд: **5,45 г**.

По аммонию: $0 = 0{,}0243 + 0{,}07\mathbf{x_1} - 0{,}155\mathbf{x_2} - 0{,}077\mathbf{x_1 x_2}$; $\mathbf{x}_1$= 0 (т.е. 16,8 мг). Тогда:

$0 = 0{,}0243 - 0{,}155\mathbf{x_2}$**;** $\mathbf{x_2}$=0,0243:0,155=0,157. В натуральных значениях: $x^\wedge = x_0 + x\cdot\Delta\mathbf{x}$ = 5,295 + 0,157·2,475 = 5,684г.

Система сбалансирована по аммонию при биомассе микрофлоры равной **16,8 мг** и биомассе звёзд равной **5,684 г.**

По фосфатам: $0 = 0{,}009 + 0{,}0235\mathbf{x_1} - 0{,}012\mathbf{x_2} + 0{,}014\mathbf{x_1 x_2}$ $\mathbf{x_1}$= 0 (т.е. 16,8 мг). Тогда:

$0=0{,}009-0{,}012\mathbf{x_2}$**;** $\mathbf{x_2}$ = 0,009 : 0,012= 0,75 (это кодированное значение). Натуральное:

$x^\wedge = x_0 + x\cdot\Delta\mathbf{x}$= 5,295 + 0,75·2,475= 7,15 г.

Система сбалансирована по фосфатам при биомассе микроводорослей равной **16,8 мг** и биомассе звёзд равной **7,15 г**.

Вывод. Система микрофлора-звезда становится сбалансированной (производимый метаболит будет полностью потребляться), если, при содержании микроводорослей в количестве 16,8 мг сырого веса, биомасса звезды (сух. вес) соответственно будет равна:

По кислороду 5,45 г

По аммиаку: 5,684 г

По фосфатам: 7,15 г

Возможные приложения к поликультуре:

1. Изучение влияния одного компонента сообщества на другой и взаимодействия обоих компонентов.
2. Экспериментальный поиск наиболее сочетаемых компонентов сообщества, то есть перспективных для выращивания в поликультуре. Иными словами, компонентов, которые стимулируют или, как минимум, не угнетают рост друг друга.

3. Разработка методик конструирования сбалансированных сообществ.

В заключение следует отметить, что данное исследование, основанное на применении методов планирования экспериментов, представляет, главным образом, методический интерес, так как не учитывает трофические связи и воспроизводство самих компонентов. Методический интерес представляет также определение условий, при которых устанавливаются концентрации элементов на постоянных уровнях. Эти концентрации зависят с одной стороны от поступления их в систему и продуцирования в самой системе, а с другой – от деструкции и выноса их из системы. Концентрацию этих элементов можно принять в качестве модели биомасс и тем самым определить (используя арсенал планирования экспериментов) биомассы компонентов сообщества, условия его стационарного функционирования и трофическую ёмкость акватории.

2.4. Планирование отсеивающих экспериментов

Успех исследования зависит от полноты списка факторов, влияющих на изучаемый объект. Выше говорилось о том, что всякое исследование либо эксперимент начинается с формулирования задачи и цели исследования. Второй обязательный этап – анализ априорной информации с составлением полного списка факторов, которые могут повлиять на ОИ. Полный список совершенно необходим для экстремальных экспериментов, которые применяют при поиске оптимальных условий. Однако технически очень трудно проводить громоздкий эксперимент, в котором задействованы многие факторы. Возникает задача выделения наиболее значимых факторов и отсева несущественных факторов. Эту процедуру можно выполнять разными способами, например, взять дробную реплику высокой дробности, позволяющую определять линейные (главные) эффекты факторов, что позволит затем отсеять факторы с несущественными главными эффектами и тем самым значительно сократить количество факторов. С оставшимися факторами можно будет поставить эксперимент по схеме ПФЭ, либо взять полуреплику от ПФЭ.

Например, для проведения ПФЭ 2^3 необходимо поставить восемь разных опытов, что даёт возможность рассчитать значения восьми коэффициентов, из которых – один свободный член, а семь остальных – характеризуют влияние исследуемых факторов (табл. 20).

Таблица 20. Матрица планирования отсеивающего эксперимента для семи факторов.

№	x_0	x_1	x_2	x_3	$x_4 = x_1x_2$	$x_5 = x_2x_3$	$x_6 = x_1x_3$	$x_7 = x_1x_2x_3$
1	+1	+1	+1	+1	+1	+1	+1	+1
2	+1	-1	+1	+1	-1	+1	-1	-1
3	+1	+1	-1	+1	-1	-1	+1	-1
4	+1	-1	-1	+1	+1	-1	-1	+1
5	+1	+1	+1	-1	+1	-1	-1	-1
6	+1	-1	+1	-1	-1	-1	+1	+1
7	+1	+1	-1	-1	-1	+1	-1	+1
8	+1	-1	-1	-1	+1	+1	+1	-1

Применяя этот план, предполагают, что роль межфакторных взаимодействий мала по сравнению с эффектами влияния отдельных факторов. Это предположение часто оказывается справедливым для тройных и, тем более, четверных и взаимодействий более высоких порядков.

При постановке ПФЭ 2^4 появляется возможность проверить влияние 15 факторов. Но особенно надёжной будет проверка факторов, вводимых вместо одного четверного и четырёх тройных взаимодействий. Итак, можно утверждать, что ПФЭ 2^4 можно надёжно использовать для проверки статистической значимости 9 факторов: 4 линейных коэффициентов + 4 тройных взаимодействий и одного четверного взаимодействия. Кроме этого, данный эксперимент позволит оценить влияние ещё 6 парных взаимодействий, что немаловажно, так как нередко статистически незначимый фактор оказывается значимым при взаимодействии с другим фактором.

Более эффективным, хотя и сравнительно сложным, является метод случайного баланса. Матрицы планирования по методу случайного баланса составляются компьютерной программой. Компьютер синтезирует планы следующим образом: он выбирает случайно строки из планов ДФЭ и ПФЭ, рассчитанных на работу с большим числом факторов, и из этих строк строит новый план, при этом программа выполняет следующие условия:

1. Максимизация количества ортогональных вектор-столбцов.
2. Минимизация коэффициента корреляции между столбцами.
3. Сохранение симметричности столбца: $\Sigma x_i=0$.

В результате получается сверхнасыщенный, иногда насыщенный план и на основе этого плана ставится эксперимент без повторных опытов (см. Приложение VII). Если в эксперименте число факторов меньше, чем в опубликованной матрице планирования, тогда лишние столбцы из плана вычёркивают.

2.4.1. Обработка результатов эксперимента, поставленного по методу случайного баланса

Первый этап. После реализации эксперимента строят гистограммы рассеяния, для чего проводится горизонтальная ось, на которой отмечают точками положения для факторов. Из каждой точки восстанавливают перпендикуляры. Слева от перпендикуляра наносят результаты, соответствующие значению нижнего уровня фактора, а справа – верхнего (рис. 8).

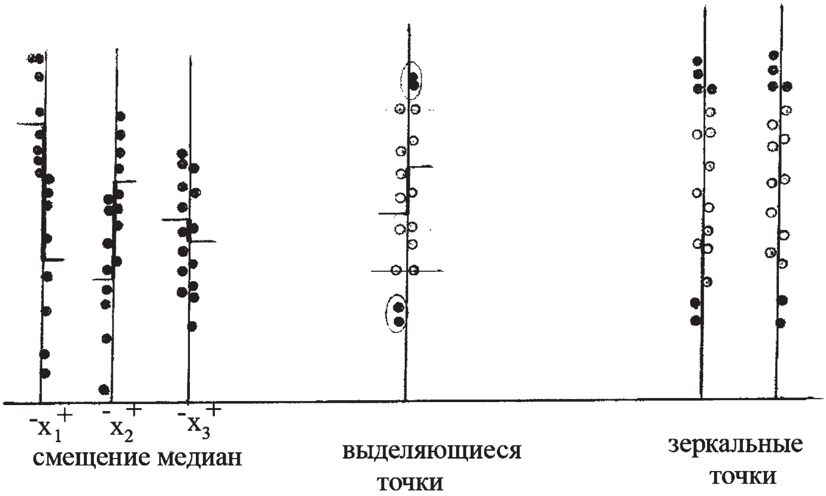


Рис. 8. Гистограммы рассеяния результатов эксперимента, поставленного по методу случайного баланса (положение медиан отмечено горизонтальными отрезками).

Вначале факторы ранжируются по величине смещения медиан (расстояние между двумя медианами). Напомним, что медиана делит выборку на две равные части. Например, если количество результатов равно 8, то медиану проводят между 4 и 5 точками (Википедия). Для этого находят среднее арифметическое 4-го и 5-го результатов и откладывает его на вертикали данного фактора (рис. 8). Существенными факторами на первом этапе выделения факторов следует считать такие факторы, для которых смещение медиан превосходит таковое для остальных факторов (см. рис. 8). Затем выбирают факторы по числу выделяющихся точек. Потом по полуразностям средних арифметических результатов опытов, в которых фактор находится на верхнем и на нижнем уровнях (см. рис. 8).

Продолжение первого этапа. Цель этого этапа – определить эффекты выделенных факторов и вычесть их из результатов эксперимента. После выполнения данной процедуры разброс результатов сглаживается, так как оставшиеся факторы слабее влияют на результаты.

Выполняется данная процедура следующим образом: рассчитывают уравнение регрессии **выделенных факторов** (без свободного члена) и производят статистическую проверку их значимости. При этом число факторов должно быть не более трёх. Для нахождения коэффициентов регрессии составляют новую матрицу, содержащую только выделенные факторы. Для этого выписывают столбцы выделенных факторов, в которых строки повторяются. Затем матрицу переписывают таким образом, чтобы каждая строка встречалась только один раз, а против каждой строки выписывают соответствующие значения ПО. Получаем матрицу планирования ПФЭ, с помощью которой рассчитываем уравнения регрессии без свободного члена (таблица 22).

Таблица 21. Матрица выделенных факторов.

x_i	x_j	Y
-	+	Y_1
+	-	Y_2
-	-	Y_3
+	-	Y_4
+	+	Y_5
-	+	Y_6
-	-	Y_7
+	-	Y_8
+	+	Y_9
+	+	Y_{10}
-	+	Y_{11}
-	-	Y_{12}
-	-	Y_{13}
-	+	Y_{14}
+	+	Y_{15}
+	-	Y_{16}

Таблица 22. Матрица ПФЭ2^2

x_1	x_2	x_1x_2	Результаты (матрица Y)				Среднее
-	–	+	Y_3	Y_7	Y_{12}	Y_{13}	$\hat{Y}_1$
+	-	-	Y_2	Y_4	Y_8	Y_{16}	$\hat{Y}_2$
-	+	-	Y_1	Y_6	Y_{11}	Y_{14}	$\hat{Y}_3$
+	+	+	Y_5	Y_9	Y_{10}	Y_{15}	$\hat{Y}_4$

Расчёт коэффициентов регрессии.

Коэффициенты рассчитываются по следующим формулам:

$b_1 = (- \hat{Y}_1 + \hat{Y}_2 - \hat{Y}_3 + \hat{Y}_4)/4$; $b_2 = (- \hat{Y}_1 - \hat{Y}_2 + \hat{Y}_3 + \hat{Y}_4)/4$; $b_{12} = (+ \hat{Y}_1 - \hat{Y}_2 - \hat{Y}_3 + \hat{Y}_4)/4$.

Но, если в исходной матрице случайного баланса отсутствуют некоторые комбинации знаков, встречающиеся в ПФЭ, тогда коэффициент находят по формуле:

$b_i = (\acute{Y}_+ - \acute{Y}_-)/2$, где: $\acute{Y}_+$ – это среднее арифметическое результатов при x_i=+1

Коэффициенты взаимодействия рассчитываются, как обычно, с помощью вектор-столбца взаимодействия.

По критерию Стьюдента проверяется статистическая значимость коэффициентов.

В итоге получается уравнение регрессии без свободного члена. Это уравнение позволяет рассчитать величину эффекта выделенных факторов: $\Delta Y_u = b_1(x_i+1) + b_2(x_j+1) + ...$

Здесь, если x = -1, то при прибавлении 1 получится 0; если x = +1, то будет 2. Далее выполняется операция «снятия эффекта»:

ΔY_u вычитаем из каждого значения Y_u и получаем новое значение Y'_u: $Y'_u = Y_u - \Delta Y_u$

Так мы исключаем влияние выделенных факторов на величину Y_u. Далее столбец Y_u заменяется столбцом Y'_u и все операции повторяются (см. Второй этап).

Второй этап

Построение гистограммы рассеяния, но уже без выделенных на 1-ом этапе факторов.

Выделение наиболее сильных факторов.

Построение матрицы ПФЭ.

Расчёт коэффициентов уравнения регрессии.

Проверка статистической значимости коэффициентов и т.д.

Выделение существенных факторов продолжают до тех пор, пока рассеяние результатов не станет малым. Точнее говоря, пока дисперсия рассеяния результатов будет лишь незначительно превышать дисперсию воспроизводимости результатов:

$F_p = (S^2_{расс}\{Y_u\})/(S^2_{воспр}\{Y\}) \le F$ (N-1; f), где:

$S^2_{расс}\{Y_u\} = [\Sigma Y^2_u - (\Sigma Y_u)^2/N]/(N-1)$

$S^2_{воспр}\{Y\} = (\Sigma(Y_{ui} - \acute{Y}_g)^2)/(m-1)$, где m – число повторов одного варианта (одной строки).

Однако на некотором этапе может получиться так: значимые линейные эффекты не выделяются, а рассеяние всё ещё велико. Это происходит от наличия парных взаимодействий невыделенных факторов (линейные эффекты слабые, но парные взаимодействия – сильные). Наличие парных взаимодействий обнаруживается при сравнении гистограмм рассеяния. Если точки двух сравниваемых гистограмм образуют у одного из концов перпендикуляра одинаковый, а у другого – зеркально отражённый рисунок, следует предположить взаимодействие между этими факторами и выделить эти факторы.

2.4.2. Априорное ранжирование факторов (Стандартная анкета)

Если для отсеивания несущественных факторов постановка многофакторного эксперимента технически невозможна, либо очень затруднительна, применяют метод априорного ранжирования факторов (табл. 23). Этот метод позволяет учитывать коллективное мнение специалистов. Число специалистов должно быть больше 10. Но при работе с высококвалифицированными специалистами их число может быть порядка 5 – 10.

Метод включает следующие этапы:

1. Составление списка факторов.
2. Составление стандартной анкеты.
3. Составление матрицы рангов (обработка стандартной анкеты).
4. Расчёт коэффициента конкордации и проверка его значимости.
5. Построение средней априорной диаграммы рангов.
6. Проверка с помощью χ^2-критерия гипотезы о неравномерном распределении рангов.
7. Отсеивание несущественных факторов.

Ниже приведено общее описание каждого из перечисленных этапов, после чего эти этапы рассматриваются на конкретном примере.

Таблица 23. Стандартная анкета

Фактор	Операционное определение	Размерность	Область определения	Ранг*
X_1	Концентрация металла в растворе. Определяется весовым анализом с точностью 0,5 %	г/л	20 – 40	3
………	……………………………...	………………	……………	…………
$X_к$	……………………………..	………………	……………	…………

* **Примечание:** ранги проставляют исследователи. Если некоторые факторы влияют одинаково, таким факторам исследователи приписывают одинаковый ранг.

Таблица 24. Матрица рангов

№ исследователя	ФАКТОРЫ						
	X_1	X_2	X_3	X_4	X_5	X_6	X_7
1	1	2	6	4	7	3	5
2	1	2	7	6	3	5	4
3	7	1	6	4	2	5	3
4	3	1	5	6	4	7	2
5	1	2	6	4	5	7	3
Сумма рангов	13	8	30	24	21	27	17
Отклонение суммы рангов от среднего, (Δ_i)	-7	-12	10	4	1	7	-3
Квадраты отклонений, $(\Delta_i)^2$	49	144	100	16	1	49	9

Расчёт коэффициента конкордации (согласованности мнений экспертов).

1. Сумма квадратов отклонений: $S = \Sigma\,(\Delta_i)^2$
2. Коэффициент конкордации: $W = (12S)/(m^2(k^3\text{-}k))$, где: k – число факторов; m – число исследователей (экспертов). (Для нашей таблицы рангов W = 0,526).

Если некоторые факторы имеют одинаковые ранги, то W находят по формуле:

$W = S/[1/12(m^2(k^3\text{-}k) - m\,\Sigma T_i)]$, где: $T_i = 1/12\,\Sigma(t_j^3\text{-}t_j)$; t_j – j-е число одинаковых рангов в i-том ранжировании. W сравнивается с табличным значением коэффициента конкордации. W > Wтабл. Если таблица распределения коэффициента конкордации отсутствует, а число факторов к>7, то можно пользоваться критерием «хи квадрат».

$\chi^2_R = S/(mk(k+1)/12)$. Если есть одинаковые ранги, то:

$\chi^2_R = S/[(mk(k+1)/12) - \Sigma T_i/(m\text{-}1)]$.

Число степеней свободы: k-1. Должно быть $\chi^2_R > \chi^2$.

Итак, если коэффициент конкордации значим, тогда следует перейти к построению средней априорной диаграммы рангов.

Средняя априорная диаграмма рангов

По вертикали откладываются суммы рангов, а на горизонтали – факторы. Проверка с помощью χ^2-критерия гипотезы о неравномерном распределении рангов. Число степеней свободы k-1 – это строка в таблице для χ^2-критерия. Столбец выбирают с числом степеней свободы 1-α, где α – уровень значимости.

Отсеивание несущественных факторов

В случае равномерного распределения факторов все факторы включают в эксперимент. Если распределение неравномерное и убывание величин влияний факторов быстрое, тогда часть факторов с большими суммами рангов отсеивают. Если же убывание очень медленное – в этом случае все факторы включают в эксперимент.

Пример применения метода стандартной анкеты для априорного ранжирования факторов: «Выбор существенных факторов, влияющих на рост мидий».

Экспертам предложена анкета (табл. 25) со списком факторов, влияющих на рост мидий. Факторы необходимо проранжировать по силе их влияния на рост мидий.

Статистическая обработка результатов анкетирования.

1. Обозначения:

1.1. Количество исследователей, m = 10, ($1 \le i \le 10$);

1.2. Количество факторов, k = 20, ($1 \le j \le 20$);

1.3. a_{ij} – ранг i-го фактора у j-го исследователя;

2. Средняя сумма рангов:

$$T = \frac{1}{k}\sum_{j=1}^{k}\sum_{i=1}^{m} a_{ij} = \frac{2085{,}5}{20} = 104{,}275$$

3. Отклонения суммы рангов каждого фактора от средней суммы рангов:

$$\Delta_i = \sum_{i=1}^{10} a_{ij} - T$$

Δ_1 = -65,295; Δ_2 = -56,295; Δ_3 = – 24,595; Δ_4 = -41,795; Δ_5 = -34,295; Δ_6 = -48,795; Δ_7 = -12,595; Δ_8 = -27,795; Δ_9 = 15,705; Δ_{10} = -6,795; Δ_{11} = 43,705; Δ_{12} = -11,795;

Δ_{13} = 5,205; Δ_{14} = 70,70; Δ_{15} = 27,705; Δ_{16} = 27,205; Δ_{17} = -5,295; Δ_{18} = 49,205

Δ_{19} = 25,205; Δ_{20} = 70,205

4. Сумма квадратов отклонений:

S=Σ $(\Delta_i)^2$ = 31161,23; $S = \sum_{i=1}^{20} (\Delta_i)^2 = 31161{,}23$

χ^2_R для «связанных» рангов:

$$\chi^2{}_R = \frac{S}{\frac{mk(k+1)}{12} - \frac{\sum_{i=1}^{m} T_i}{m-1}},$$

где: $T_i = \frac{\sum_{i=1}^{m}(t_j{}^3 - t_j)}{12}$,

где: t_j – j- ое число одинаковых рангов в i-ом ранжировании.

Для 1-го исследователя имеем:

Связанные ранги	5,5	3	10,5	13	15,5	19,5
t_j	2	3	2	3	2	2

Для t_1: $2^3 – 2 = 6$; $3^3 -3 = 24$; $2^3 – 2 = 6$; $3^3 -3 = 24$; $2^3 – 2 = 6$; $2^3 – 2 = 6$;

$T_1 = (6+24+6+24+6+6)/12=72/12 = 6$.

Аналогично:

$T_2 =0$; $T_3 = 8$; $T_4 = 0$; $T_5 = 5,5$; $T_6 = 8$; $T_7 = 8,5$; $T_8 = 0,5$; $T_9 = 0,5$; $T_{10} =13$; $\Sigma T_i =50$.

$\chi^2_R= 31161,8/[10*20(20+1)/12 – 50/(10-1)]=31161,8/(350 – 5,56)= 90,6$

Число степеней свободы для χ^2 при p=0,95; f = k-1=19; $\chi^2 = 10,1$

$\chi^2< \chi^2_R$ (10,1<90,6).

Распределение факторов неравномерное, поэтому возможно отсеивание слабых факторов.

5. Коэффициент конкордации для случая «связанных рангов»:

$W= S/[m^2(k^3 – k)/12 -m\Sigma T_i]= 31161,2/[10^2 (20^3 – 20)/12 – 10\cdot 50] =31161,2/66000= 0,472$

Определяем статистическую значимость коэффициента конкордации по критерию χ^2:

χ^2расч = $m(k-1)W = 10(20-1)\cdot 0,472= 89,68$

Табличное значение равно 10,1; $\chi^2< \chi^2_R$. Коэффициент конкордации статистически значим; предположение о согласованности мнений специалистов подтверждается.

6. Строим гистограмму ранжирования факторов.
7. Отбираем наиболее существенные факторы:

- Температура морской воды;
- Возраст мидий;
- Скорость потребления пищи;
- Общий вес тела;
- Состав пищи;
- Продуктивность фитопланктона на акватории фермы;
- Состояние гонад.

8. Факторы, слабо влияющие на рост мидий:

- Субстрат, к которому прикрепились мидии;
- Концентрация аммиака в морской воде;
- Ориентация субстрата в пространстве;
- Освещённость;
- Прозрачность воды.

Таблица 25. Стандартная анкета

Фактор	Наименование фактора	Размерность	Область определения	Область интереса	Ранг*
x_1	Температура морской воды	°C	0 – 30	6 – 27	
x_2	Возраст мидии	год	0,1 – 25	0,5 – 4	
x_3	Состояние гонад	Стадия зрелости	1 – 6	1 – 6	
x_4	Масса тела (сырая)	г	0,1 – 75	15 – 25	
x_5	Состав пищи	–	Качественный фактор	–	
x_6	Скорость потребления пищи	мг/экз.·сут.	0 – 120	40 – 70	
x_7	Прибойность	–	Качественный фактор	–	
x_8	Продуктивность фитопланктона	гС/м²·сут.	0 – 15	0,5 – 3	
x_9	Плотность популяции (посадки)	экз./дм²	0 – 100	2 – 10	
x_{10}	Концентрация O_2	Насыщение, %	0 – 120	50 – 100	
x_{11}	Освещённость	лк	0 – 100 000	5000-50000	
x_{12}	Местоположение в друзе	–	Качественный фактор	Край, центр	
x_{13}	Солёность	‰	0 – 40	12 – 18	
x_{14}	Способ выращивания	Субстрат жёсткий, гибкий	Качественный фактор	–	
x_{15}	Прозрачность воды, диск Секки	м	0.5 – 20	2 – 10	
x_{16}	Концентрация H_2S	мг/л	1 – 5	0,5 – 1	
x_{17}	Время оседания молоди	месяц	I – XII	I – XII	
x_{18}	Положение субстрата в пространстве	–	Вертикальное горизонтальное наклонное	Вертикальное горизонтальное наклонное	
x_{19}	Глубина	м	0 – 80	1 – 20	
x_{20}	Концентрация NH_4	мкг/л	0 – 500	50 – 200	

*__Примечание:__ ранг каждого фактора проставляют эксперты

Таблица 26. Матрица рангов

Исследователь	Факторы											
	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}
1	5,5	3	4	3	5,5	1	3	7	12	9	10,5	13
2	3	2	11	1	9	6	10	12	13	8	18	14
3	1	4,5	4,5	4,5	10	10	14	10	12,5	7,5	17,0	12,5
4	1	7	12	13	4	2	17	8	5	9	11	3

5	1	9	8	13	5	6	13	7	18	3,5	18	10
6	5	2,5	6,5	8	4,5	2,5	6,5	1	10,5	4,5	12,5	15,5
7	3,5	1,5	3,5	1,5	9	9	9	12	6	6	17	6
8	10	8,5	4	8,5	2	3	5	12	14	17	11	6
9	3	1,5	13	1,5	11	5,5	12	5,5	10	19	16,5	7
10	2	8,5	13	8,5	10,5	10,5	2	2	10	14	16,5	5,8
$\sum$	39	48	79,5	62,5	70,5	55,5	91,5	76,5	120	97,5	148	92,5

Таблица 26. (продолжение)

Исследователь	Факторы							
	x_{13}	x_{14}	x_{15}	x_{16}	x_{17}	x_{18}	x_{19}	x_{20}
1	13	15,5	8	19,5	10,5	18,8	13	19,5
2	4	15	19	16	5	17	4	20
3	2	18	15	7,5	4,5	10	20	16
4	10	20	14	6	16	10	18	15
5	2	18	13	3,5	13	18	18	18
6	18,5	18,5	14	18,5	10,5	18,5	12,5	18,5
7	17	19,5	19,5	12	14,5	17	14,5	12
8	7	19	15	13	1	18	16	20
9	16,5	15	9	19	4	14	8	19
10	16,5	5,5	12	16,5	20	5,5	5,5	16,5
$\sum$	109,5	175	132	131,5	99	159,8	120,5	174,5

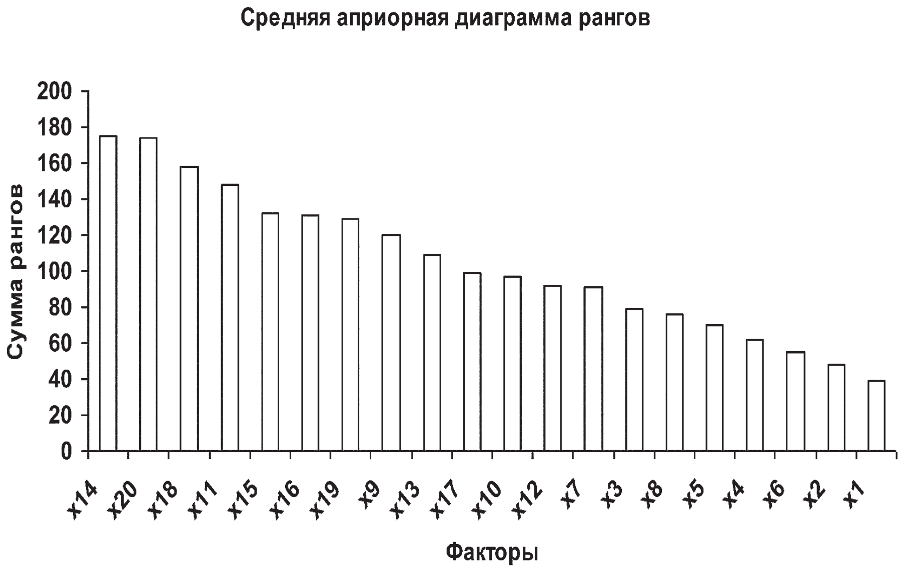


Рис. 9. Средняя априорная диаграмма рангов.

2.5. Защита экспериментальных результатов от воздействия дрейфа объекта исследования

В биологических исследованиях объекты исследования, как правило, представлены живыми организмами, которые в течение эксперимента могут изменять свои характеристики и иначе реагировать на задаваемые факторы, что, в свою очередь, влияет на результаты эксперимента. На стр. 63 обсуждается отсеивание незначимых факторов с использованием матрицы планирования для ПФЭ 2^4. Для реализации данного эксперимента необходимо поставить 16 разных опытов (табл. 27), а если эксперимент выполняется в трехкратной повторности, тогда число экспериментальных сосудов будет равным 48. Может оказаться, что такое количество невозможно обработать в течение одного дня. Но в последующие дни объект может изменить свои свойства. Поэтому экспериментальный план приходится разбивать на блоки и в один день обрабатывать по одному блоку. Допустим, что имеется возможность в один день обработать половину результатов, в этом случае достаточно план разбить всего на два блока. Но разбиение на блоки необходимо произвести таким образом, чтобы эффект дрейфа объекта не исказил получаемые результаты.

Таблица 27. Матрица планирования и расчёта основных эффектов и эффектов взаимодействий ПФЭ 2^4

№	x_0	x_1	x_2	x_3	x_4	x_1x_2	x_2x_3	x_1x_3	x_1x_4	x_2x_4	x_3x_4	$x_1x_2x_3$	$x_1x_2x_4$	$x_2x_3x_4$	$x_1x_3x_4$	$x_1x_2x_3x_4$
1	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1	+1
2	+1	-1	+1	+1	+1	-1	+1	-1	-1	+1	+1	-1	-1	+1	-1	-1
3	+1	+1	-1	+1	+1	-1	-1	+1	+1	-1	+1	-1	-1	-1	+1	-1
4	+1	-1	-1	+1	+1	+1	-1	-1	-1	-1	+1	+1	+1	-1	-1	+1
5	+1	+1	+1	-1	+1	+1	-1	-1	+1	+1	-1	-1	+1	-1	-1	-1
6	+1	-1	+1	-1	+1	-1	-1	+1	-1	-1	-1	+1	-1	-1	+1	+1
7	+1	+1	-1	-1	+1	-1	+1	-1	+1	+1	-1	+1	-1	+1	-1	+1
8	+1	-1	-1	-1	+1	+1	+1	+1	-1	-1	-1	-1	+1	+1	-1	-1
9	+1	+1	+1	+1	-1	+1	+1	+1	-1	-1	-1	+1	-1	-1	+1	-1
10	+1	-1	+1	+1	-1	-1	+1	-1	+1	-1	-1	-1	+1	-1	+1	+1
11	+1	+1	-1	+1	-1	-1	-1	+1	-1	+1	-1	-1	+1	+1	-1	+1
12	+1	-1	-1	+1	-1	+1	-1	-1	+1	+1	-1	+1	-1	+1	+1	-1
13	+1	+1	+1	-1	-1	+1	-1	-1	-1	-1	+1	-1	-1	+1	+1	+1
14	+1	-1	+1	-1	-1	-1	-1	+1	+1	-1	+1	+1	+1	+1	-1	-1
15	+1	+1	-1	-1	-1	-1	+1	-1	-1	+1	+1	+1	+1	-1	+1	-1
16	+1	-1	-1	-1	-1	+1	+1	+1	+1	+1	+1	-1	-1	-1	-1	+1

Данное условие легко выполняется, если эффект дрейфа приравнять к эффекту, определяемому с помощью какого-либо столбца матрицы планирования (см. табл. 27). Можно ожидать, что наименее значимым является эффект четверного взаимодействия, к которому приравнивается дрейф ОИ:

Дрейф = $x_1x_2x_3x_4$

Разбиение на два блока осуществляется путём выбора для блока № 1 всех строк данной матрицы, которым в столбце $x_1x_2x_3x_4$ соответствует +1, а для блока № 2 –1 (или наоборот). Поэтому в блок № 1 войдут опыты: 1, 4, 6, 7, 10, 11, 13, 16. Второй блок образуют строки: 2, 3, 5, 8, 9, 12, 14, 15.

Если разбиения на два блока не достаточно, тогда каждый из блоков можно разбить на подблоки, выбирая из блока строки, содержащие в столбце, допустим, $x_1x_3x_4$ единицу со знаком плюс или со знаком минус. Блок № 1 включает две строки, соответствующие $x_1x_2x_3x_4 = +1$ и $x_1x_3x_4 = +1$. Блок № 2: $x_1x_2x_3x_4 = +1$ и $x_1x_3x_4 = -1$. Блок № 3: $x_1x_2x_3x_4 = -1$ и $x_1x_3x_4 = +1$. Блок № 4: $x_1x_2x_3x_4 = -1$ и $x_1x_3x_4 = -1$. Поэтому в блок № 1 войдут опыты со следующими номерами: 1, 6, 10, 13; в блок № 2: 4, 7, 11, 16. Блок № 3: 3, 9, 12, 15. Блок № 4: 2, 5, 8, 14.

Итак, результаты опытов можно обрабатывать за четыре этапа, при этом без риска искажения конечных результатов за счёт изменения во времени свойств объекта исследования. Но за этот выигрыш пришлось заплатить цену в виде двух столбцов, затраченных на борьбу с искажающим воздействием дрейфа ОИ. Вместо проверки значимости 15 эффектов, в данном случае имеется возможность проверки 13 эффектов. Или же вместо проверки значимости девяти факторов, выполняется проверка семи факторов.

Дрейф объекта исследования происходит от того, что неуправляемые внутренние либо внешние факторы, воздействуя на ОИ, изменяют его свойства во времени, причём изменения свойств ОИ происходят некоторым упорядоченным образом, а это значит, что изменения ОИ во времени можно описать математически.

Важно отметить, что изменение характеристик ОИ может происходить не только во времени, но и по какой-либо другой оси координат, например, по глубине. Так, макрофиты, взятые для эксперимента на разных глубинах, могут иметь различные биохимические либо физиологические характеристики. В этом случае можно воспользоваться латинскими планами, составив из разных проб макрофитов элиминирующую группировку и защитив таким образом результаты эксперимента от искажающего воздействия дрейфа.

Характеристики дрейфа не изменяются при варьировании управляемых факторов (дрейф не зависит от управляемых факторов). Если известно, что ОИ не стабилен и подвержен дрейфу, тогда значение Y изменяется при постоянных уровнях управляемых факторов. Для измерения характеристик дрейфа необходимо получить кривую зависимости параметра оптимизации ПО (или Y) от времени: Y=f(T) при постоянных управляемых факторах. Затем эту кривую описывают полиномиальным уравнением. Обычно дрейф имеет вид плавной кривой, поэтому его можно представить полиномом невысокой степени, или другой гладкой функцией, например, экспонентой или квадратичной функцией.

Экспериментальный план строится ортогонально к уравнению, описывающему дрейф. Методы проведения эксперимента в условиях полиномиального дрейфа (построение плана и реализация эксперимента) изложены в источниках, размещённых в Интернете, где рассматриваются случаи линейного, экспоненциального, дискретного, нелинейного дрейфа.

2.6. Планирование экстремальных экспериментов (Метод крутого восхождения Бокса-Уилсона)

Эксперименты, проводимые с целью нахождения оптимальных условий протекания изучаемого процесса либо получения оптимальных результатов, называют экстремальным экспериментом.

В таких экспериментах ставится задача достижения максимума (либо минимума) ПО. Иными словами, самого лучшего результата или экстремума функции: <u>максимум</u> выживаемости организмов, темпов роста, содержания пигментов, продукции, надёжности функционирования, или <u>минимум</u> себестоимости, смертности организмов, уродств, дорогих реактивов, энергоёмкости технологического процесса.

Существуют различные экспериментальные методы оптимизации, но наиболее распространённым является «Метод крутого восхождения Бокса – Уилсона» (Адлер, Маркова, Грановский, 1971).

Следует отметить, что задачи оптимизации очень широко распространены в народном хозяйстве, в науке, в быту. Всегда, при поиске рационального, экономного решения либо решения, базирующегося на здравом смысле, человек вольно или невольно занимается оптимизацией. Например, как рационально потратить зарплату, распределить своё свободное время; составить своё меню и т.д. Для того, чтобы решить эти задачи, нужно принять во внимание целый ряд факторов, которые наиболее чётко влияют на стоящую задачу оптимизации.

На производстве, в научных исследованиях экспериментальная оптимизация сводится к поиску таких сочетаний уровней факторов, при которых ПО достигает максимальных или минимальных значений. Но для того, чтобы достичь желаемых результатов, нужно:

- Выявить существенные факторы, влияющие на ПО;
- Найти оптимальные значения уровней этих факторов.

Оптимизировать какой-либо процесс можно двумя путями:

1. На основе имеющейся теории исследуемого явления, процесса. Например, через математическое описание, построение математической модели и последующую оптимизацию модели численными методами.
2. Второй подход – эмпирический; он не требует разработанной теории, поэтому он значительно дешевле. Однако поиск лучших вариантов методом проб и ошибок требует больших временных и финансовых затрат. Метод Бокса-Уилсона (градиентный метод, или метод крутого восхождения) позволяет значительно сократить время поиска оптимума.

Эмпирически двигаться к оптимуму можно разными путями:

- последовательным перебором факторов, то есть варьировать уровни одного фактора при фиксированных уровнях остальных факторов; затем изменять уровни какого-либо другого фактора до тех пор, пока ПО улучшается и т.д.
- сразу варьировать уровни всех факторов в нужном направлении. Но для этого нужно знать это направление.

Основное в эмпирическом методе – это последовательное улучшение отклика (результата). Оказывается, что можно, даже не зная механизма процесса (не владея теорией), двигаться кратчайшим путём к оптимальной области, то есть к таким значениям факторов, при которых величина отклика принимает наилучшие значения (максимальное либо минимальное). Движение к оптимуму осуществляется при одновременном изменении уровней всех факторов таким образом, чтобы при новых значениях факторов отклик (измеряемая величина) стал лучше. Иными словами, осуществляется переход в факторном пространстве в новую область, расположенную ближе к области оптимума (рис. 10). Как определяется направление движения? Дело в том, что экспериментальные точки выбираются в некотором объёме (области) факторного пространства, поэтому появляется возможность исследовать это пространство сразу в нескольких точках и по результатам составить представление о распределении «хороших» и «плохих» мест в данном объёме (области). А на основе такого представления не трудно прогнозировать направление дальнейшего перемещения, которое приведёт к улучшению отклика.

Линейные члены уравнения регрессии описывает градиент, то есть направление, в котором нужно двигаться, изменяя уровни факторов пропорционально величинам коэффициентов уравнения регрессии и в соответствии со знаками коэффициентов. Поэтому данный метод называется градиентным. Выражение для градиента имеет вид:

$$grady(x_i) = \nu y(x_i) = \frac{dy(x_i)}{dx_1} i + \frac{dy(x_i)}{dx_2} j + ... + \frac{dy(x_i)}{dx_k} k$$

νy – обозначение градиента;

dy/dx_k – частная производная функции по k-тому фактору.

i, j, k – единичные векторы в направлении координатных осей.

Следовательно, необходимо получить такое уравнение регрессии, которое не будет отличаться от записи градиента. Иными словами, опыт нужно провести так, чтобы все взаимодействия стали незначимыми. Для этого нужно сузить интервалы варьирования факторов, но не настолько, чтобы линейные члены уравнения стали незначимыми.

Метод крутого восхождения был изобретён в 1951 г. Боксом и Уилсоном и сразу получил широкую популярность, так как позволял добиваться экономического эффекта без серьёзных затрат финансов и времени. С тех пор было опубликовано много работ по оптимизации разнообразных задач, и даже у многих сложилось представление о планировании эксперимента как о предмете, целью которого является оптимизация различных процессов и других объектов исследования.

Метод состоит из следующих этапов:

1. Выбор параметра оптимизации (ПО).
2. Выбор основного уровня факторов и интервалов варьирования.
3. Проведение ПФЭ или ДФЭ.
4. Принятие решений после первой серии опытов.
5. Крутое восхождение.
6. Принятие решений после крутого восхождения.

Реализация перечисленных выше шести этапов:

1. Методика выбора параметра оптимизации описана в разделе о ПФЭ.
2. Основной уровень выбирается либо по результатам анализа априорных сведений, либо из интуитивных соображений, либо по результатам предварительных опытов. При выборе интервалов варьирования факторов необходимо учитывать, что *узкие* интервалы варьирования приводят к неоправданным затратам; *широкие* интервалы могут привести к «перепрыгиванию» через экстремум и его потере, а также может обнаружиться, что слишком обширный участок поверхности отклика нельзя аппроксимировать плоскостью.
3. Если число факторов невелико (2-4), тогда проводят ПФЭ; при большем числе факторов ставят ДФЭ.
4. В первой серии опытов интервалы варьирования задают широкими. После постановки эксперимента и получения уравнения регрессии возможны следующие варианты:

- Все линейные коэффициенты значимы. Модель адекватна, поэтому возможно движение по градиенту.
- Не все линейные коэффициенты значимы. Нужно расширить интервалы по незначимым коэффициентам, либо увеличить число повторных опытов для того, чтобы уменьшить дисперсию воспроизводимости.
- Линейная модель – неадекватна. Адекватность не обязательно рассчитывать по критерию Фишера: если значим хоть один эффект взаимодействия, то линейная модель – не адекватна. Поэтому нужно перенести центр эксперимента в условия наилучшего опыта и поставить новый эксперимент, уменьшив при этом интервалы варьирования.

5. Схема крутого восхождения представлена на рисунке 11. $AB/\Delta x = tg\alpha$. Отклик: $AB = tg\alpha\Delta x = b_1\Delta x_1$. Вводим переменный шаг варьирования k, (k изменяется от 0 до ∞). Переменный шаг: $k \cdot b_1 \cdot \Delta x_1$ должен быть пропорциональным как к величинам коэффициентов регрессии, так и к размерам интервалов варьирования натуральных переменных. Например, в эксперименте выяснилось, что существенными являются четыре фактора.

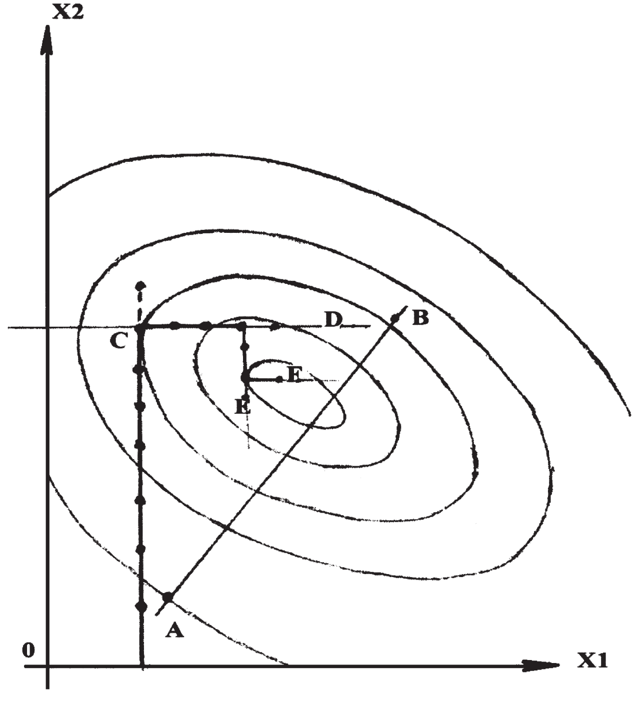


Рис. 10. Поиск оптимальной области методом крутого восхождения в двухфакторном пространстве (подробности в тексте).

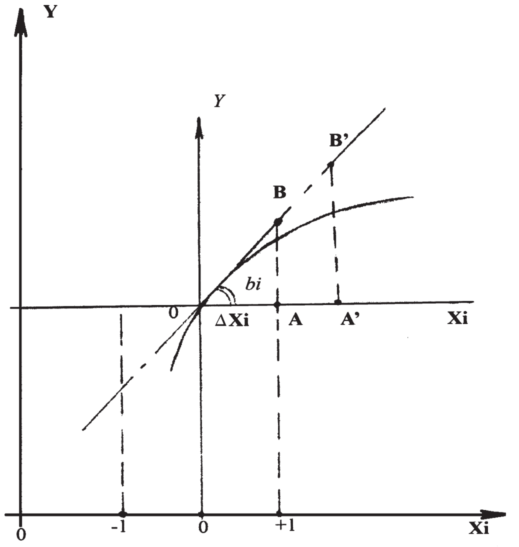


Рис. 11. Определение шага крутого восхождения.

Таблица 28. Условия планирования

Факторы	x_1	x_2	x_3	x_4
Базовый уровень, x_0	6,5	85	350	7,75
Интервал варьирования, Δx_i	5,5	15	150	1,25
Линейные коэффициенты регрессии, b_i	-8,75	+3,1	-21,8	-10
Произведение $b_i \cdot \Delta x_i$	-48	45	-3260	-12,5
Шаг крутого восхождения, $\lambda_{i\ крут\ восх}$	-1,47	13,9	-100	-0,38
Округлённый $\lambda_{i\ крут\ восх}$	-1,5	15	-100	-0,5

Расчёт шагов крутого восхождения

В таблице 28 произведение $b_3\Delta x_3$ является максимальным по абсолютной величине. Принимаем его в качестве базового и выбираем для него шаг крутого восхождения, который назовём базовым. Шаги крутого восхождения по остальным факторам рассчитываются из соображения пропорциональности по формуле:

$$\lambda_{i\ крут\ восх} = [(b_i \cdot \Delta x_i)/(b_{баз}\Delta x_{баз})]\ \lambda_{крут\ восх\ баз}$$

6. Мысленные опыты. Мысленно прибавляем к базовому уровню каждого фактора $\lambda_{i\ крут\ восх}$ и подсчитываем результат (Y). Но шаги у нас рассчитаны в натуральных значениях, поэтому их надо перевести в кодированные значения по формуле:

$\lambda_{i\ \text{крут восх кодир}} = (\lambda_{i\ \text{крут восх}} - x_{0i})/\Delta x_i$

Можно и шаг крутого восхождения в натуральных значениях прибавлять к базовому значению фактора, выраженному в натуральных значениях. А затем уже полученные новые уровни факторов перевести в кодированные значения и подставить их в уравнение регрессии. Важно подчеркнуть, что шаг всегда прибавляется с учётом знака. Когда Y_n = (1,5 – 2) от $Y_{\text{начальное}}$, тогда в точке x_n ставим новый эксперимент и определяем новый градиент, причём интервалы варьирования и шаги крутого восхождения уменьшаем. Затем совершаем второй этап крутого восхождения и т.д. до тех пор, пока линейные эффекты становятся незначимыми, а эффекты взаимодействий возрастают.

Примечание. На каждом этапе выполняется проверка значимости линейных коэффициентов, воспроизводимости опытов и адекватности модели. Матрица планирования строится как для ПФЭ, но без взаимодействий факторов.

Алгоритм поиска оптимума методом «Крутого восхождения»:

1. Постановка ПФЭ или ДФЭ, получение адекватного уравнения регрессии со значимыми линейными коэффициентами.
2. Расчёт шагов крутого восхождения. Для этого составляется список факторов (натуральные значения), включаемых в поиск оптимума с указанием базового, верхнего и нижнего уровней, а также интервала варьирования.
3. Расчёт произведения интервалов варьирования факторов (натуральные значения) на соответствующие коэффициенты регрессии: $b_i \cdot \Delta x_i$. За основу принимается максимальное произведение $b_i \cdot \Delta x_i$. Определение шага крутого восхождения по формуле:
4. $\lambda_{i\ \text{крут восх}} = [(b_i \cdot \Delta x_i)/(b\text{баз} \cdot \Delta x\text{баз})]\ \lambda_{\text{крут восх баз}}$. Округление шагов.
5. Переход к «мысленным опытам». Шаги крутого восхождения (натуральные) прибавить к основным уровням (натуральным); либо перевод в кодированную форму: $\lambda_{i\ \text{крут восх кодир}} = (\lambda_{i\ \text{крут восх}} - x_{0i})/\ \Delta x_i$
6. Кодированные шаги (с учётом их знака) прибавить к значениям факторов (к основным уровням=0) и подсчёт значения отклика. Если новые значения уровней факторов представлены в натуральных значениях, тогда их необходимо кодировать, подставить в уравнение регрессии и подсчитать значение отклика.

7. Когда отклик изменится в два раза, выполняется вновь постановка ПФЭ или ДФЭ, после чего вся процедура повторяется. Шаги крутого восхождения при этом необходимо уменьшить, чтобы избежать «перепрыгивания» через экстремум.

2.6.1. Пример применения метода «крутого восхождения»

Данный пример носит формальный, чисто учебный характер и предназначен для освоения последовательности вычислений. Единицы измерений ПО и факторов опущены.

Краткое описание предварительных исследований. В лабораторных условиях изучали влияние ряда факторов на синтез некоторых пигментов культурами одноклеточных водорослей.

ПО – скорость синтеза пигментов.

После проведения цикла исследований, в которые входил и метод стандартной анкеты, постановка ДФЭ, было установлено, что накопление пигментов определяется, главным образом, тремя факторами: x_1, x_2, x_3.

С этими факторами был поставлен ПФЭ2^3, причём с таким расчётом, чтобы линейное уравнение было адекватным, то есть интервалы варьирования факторов были достаточно узкими (табл. 29).

Таблица 29. Уровни факторов (натуральные значения)

Наименование	x_1	x_2	x_3
Базовый уровень, x_0	3	0,9	40
Интервал варьирования, Δx_i	1	0,5	20
Верхний уровень	4	1,4	60
Нижний уровень	2	0,4	20

Получено уравнение регрессии:
$y = 2042 - 482\,x_1 - 242x_2 + 122x_3$

В таблице 30 наибольшее значение имеет произведение $b_3\Delta x_3$, поэтому шаг крутого восхождения, равный 2,5 принят за основу ($\lambda_{\text{крут восх баз}}$); шаги по другим факторам рассчитаны по формуле: $\lambda_{\text{i крут восх}} = [(b_i \cdot \Delta x_i)/(b_{\text{баз}} \cdot \Delta x_{\text{баз}})]\,\lambda_{\text{крут восх баз}}$

$\lambda_{\text{1 крут восх}} = 2{,}5(-482/2440) = -0{,}5$

$\lambda_{\text{2 крут восх}} = 2{,}5(-121/2440) = -0{,}125$

В 5-ом опыте факторы принимают предельные значения, поэтому решено реализовать опыты № 4 и № 5. Получены значения ПО: 3280 и 3490.

Это наилучшие результаты, полученные в процессе всего исследования. Поэтому крутое восхождение считается удачным.

Таблица 30. Расчёт крутого восхождения (исходным является базовый уровень)

Условия планирования	x_1	x_2	x_3
Коэффициенты уравнения регрессии, b_i	-482	-242	122
Произведение $b_i \cdot \Delta x_i$	-482	-121	2440
Шаг крутого восхождения, $\lambda_{i\ \text{крут восх}}$	-0,5	-0,125	2,5
Округлённый $\lambda_{i\ \text{крут восх}}$	-0,5	-0,1	2,5
Мысленные опыты:			
1	2,5	0,8	42,5
2	2,0	0,7	45,0
3	1,5	0,6	47,5
4	1,0	0,5	50,0
5	0,5	0,4	52,5

Проверка мысленных опытов 4 и 5; кодируем переменные:

Опыт № 4. $x'_i = (x_i - x_{i0})/\Delta x_i$
$x_1 = (1\text{-}3)/1 = -2$ $x_2 = (0{,}5\text{-}0{,}9)/0{,}5 = -0{,}8$ $x_3 = (50\text{-}40)/20 = 0{,}5$
$y = 2042 - 482(-2) - 242\,(-0{,}8) + 122 \cdot 0{,}5 = 3260$

Опыт № 5
$x_1 = (0{,}5\text{-}3)/1 = -2{,}5$ $x_2 = (0{,}4\text{-}0{,}9)/0{,}5 = -1$ $x_3 = (52{,}5\text{-}40)/20 = 0{,}625$
$y = 2042 - 482(-2{,}5) - 242\,(-1) + 122 \cdot 0{,}625 = 3565{,}25$

2.7. Центральное композиционное планирование (ЦКП)

Данный тип экспериментальных планов завершает всю последовательность планов экспериментов, основанных на регрессионном анализе. ЦКП эффективно использовать после нахождения области факторного пространства, в котором ПО принимает оптимальные значения. Уравнение регрессии, рассчитываемое по результатам ЦКП, содержит квадратичные члены и, поэтому описывает существенно нелинейные поверхности отклика. Исследование такой поверхности позволяет определить область экстремума, то есть, наиболее оптимальные значения факторов. Очевидно, что и в области экстремума кривизна поверхности отклика возрастает; её невозможно описать поверхностью, получаемой с помощью планов 1-го порядка. Расчёт адекватности модели даёт отрицательный результат. Для тщательного исследования области экстремума приходится применять планы 2-го порядка, в которых факторы варьируются на трех уровнях, например, ПФЭ3^2, либо с ещё большим числом уровней. В данном случае эффективным также является применение метода ЦКП. Выигрыш в количестве опытов виден из таблицы 31.

Таблица 31. Сравнение требуемого количества опытов при реализации ПФЭ3^n и ЦКП при работе с числом факторов равным n

Число факторов, n	2	3	4	5	6
ПФЭ3^n	9	27	81	243	729
ЦКП	9	15	25	43	72

Для сокращения количества опытов Боксом и Уилсоном предложена идея достройки плана вокруг основного уровня серии опытов (ПФЭ) – это и есть ЦКП (рис. 12).

Точки ПФЭ называются главной композицией.

Точки на осях факторного пространства называются звёздными точками α.

В центре факторного пространства – центральная точка.

Расстояние между центральной и звёздной точками называется звёздным плечом (таб. 32).

План ЦКП состоит из: 2^n (ПФЭ)+ 2n (звёздные точки) + центральная точка.

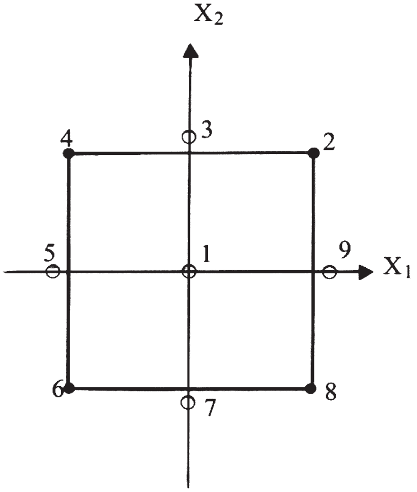


Рис 12. Положение экспериментальных точек в факторном пространстве при центральном композиционном планировании (ЦКП). Главная композиция N=2^n; точки: 2, 4, 6, 8. Центральня точка N=1; точка 1. Звёздные точки N=2n; точки: 3, 5, 7, 9.

Таблица 32. Звёздные плечи для числа факторов от 2 до 5

n	2	3	4	5
α.	1,000	1,215	1,414	1,547

Построение матрицы планирования эксперимента по методу ЦКП:

1. Построить главную композицию, т.е. ПФЭ 2^n, состоящую из 2^n строк.

2. Добавить 2n строк со звёздными точками.
3. Добавить 1 строку с центральной точкой (все координаты равны 0).

В таблице 33 представлен план эксперимента ЦКП для трёх факторов.

Таблица 33. Центральное композиционное ортогональное планирование для трёх факторов (план эксперимента)

№ опыта	x_1	x_2	x_3
1	-1	-1	+1
2	-1	+1	-1
3	+1	-1	-1
4	+	+1	+1
5	-1	-1	-1
6	-1	+1	+1
7	+1	-1	+1
8	+1	+1	-1
9	- α.	0	0
10	+ α.	0	0
11	0	- α.	0
12	0	+ α.	0
13	0	0	- α.
14	0	0	+ α.
15	0	0	0

Если число факторов более трёх, в этом случае план эксперимента следует строить с помощью таблицы 34.

Таблица 34. Матрица планирования и расчёта коэффициентов регрессии для ЦКП в общем виде (n – число факторов)

№	X_0	X_1	X_2	…	X_n	X_1X_2	…	$X_{n-1}X_n$	X_1^2	X_2^2	…	X_n^2
1	1	-1	-1	…	-1	+1		+1	+1	+1	…	+1
2	1	+1	-1	…	-1	-1		+1	+1	+1	…	+1
3	1	-1	+1	…	-1	-1		+1	+1	+1	…	+1
4	1	+1	+1	…	-1	+1		+1	+1	+1	…	+1
…	…	…	…	…	…	…	…	…	…	…	…	…
2^n	1	+1	+1	…	+1	+1	+1	+1	+1	+1	…	+1
2^n+1	1	- α	0	…	0	0	0	0	α^2	0	…	0
2^n+2	1	+α	0	…	0	0	0	0	α^2	0	…	0
2^n+3	1	0	-α	…	0	0	0	0	0	α^2	…	0
2^n+4	1	0	+α	…	0	0	0	0	0	α^2	…	0
…	…	…	…	…	…	…	…	…	…	…	…	…
2^n+2n-1	1	0	0	…	-α	0	0	0	0	0	…	α^2
2^n+2n	1	0	0	…	+α	0	0	0	0	0	…	α^2
2^n+2n+1	1	0	0	…	0	0	0	0	0	0	…	0

В плане имеется нарушение ортогональности двух типов:

1. $\Sigma x_0 x_i^2 \neq 0$
2. $\Sigma xi^2 xj^2 \neq 0$

Нарушение ортогональности устраняется, если ввести $'X_i^2 = x_i^2 - \Theta$, где: $\Theta = (2^n + 2\alpha^2)/(2^n + 2n + 1)$

Для n=2 имеем: $\Theta = (2^2+2\times1^2)/(2^2+2\times2+1)=6/9=2/3$.

Подставляем в план вместо $x_i^2 = 1 - \Theta$, (где $x_i^2=1$).

Для n=2 имеем: 1 – 2/3=1/3.

В тех местах, где $x_i^2 = \alpha^2$ нужно сделать: $\alpha^2 - \Theta = 1 - 2/3 = 1/3$.

В местах, где $x_i^2 = 0$ нужно сделать: $0 - \Theta = 0-2/3= -2/3$

Для 2-х факторного плана получим такую матрицу (табл. 35):

Таблица 35. Матрица ЦКП для двух факторов

№	x_0	x_1	x_2	x_1x_2	$'x_1^2$	$'x_2^2$
1	1	-1	-1	+1	1|3	1|3
2	1	+1	-1	-1	1|3	1|3
3	1	-1	+1	-1	1|3	1|3
4	1	+1	+1	+1	1|3	1|3
5	1	+α (+1)	0	0	1|3	-2|3
6	1	-α (-1)	0	0	1|3	-2|3
7	1	0	+α (+1)	0	-2|3	1|3
8	1	0	-α (-1)	0	-2|3	1|3
9	1	0	0	0	-2|3	-2|3

Следует обратить внимание, что в вектор-столбцах для квадратичных членов стоят преобразованные значения: $'x_1^2 = x_1^2 - \Theta$ и $'x_2^2 = x_2^2 - \Theta$.

Расчёт коэффициентов уравнения регрессии выполняется по формулам:

$b_0 = \Sigma y_u / g$,

где g – количество разных опытов (число строк)

$$b_i = \frac{\sum x_{iu} y_u}{\sum x^2_{iu}} \quad b_{ij} = \frac{\sum x_{iu} x_{ju} y_u}{\sum (x_{iu} x_{ju})^2} \quad b_{ii} = \frac{\sum x^2_{iu} y_u}{\sum (x^2_{iu})^2}$$

Итак, в данной матрице вместо x_i^2 подставлены $x_i^2 - \Theta$, поэтому, если вернуться к обычным кодированным факторам, свободный член изменится: он будет равен: $b'_0 = b_0 - \Sigma b_{ii} x'^2_i$. Для двух факторов $b'_0 = b_0 - (b_{11}\times2/3 + b_{22}\times2/3) = b_0 - 2/3(b_{11} + b_{22})$.

Уравнение для любого числа факторов имеет вид:

$Y = (b_0 - \Sigma b_{ii} x'^2_i) + \Sigma b_i x_i + \Sigma b_{ij} x_i x_j + \Sigma b_{ii} x_i^2$

Дисперсии:

$S^2\{b_0\} = S^2/N$ $S^2 = (\Sigma S^2_g)/g$ (см. ПФЭ)

$S^2\{b_i\} = S^2/((2^n + 2\alpha^2)m)$
$S^2\{b_{ij}\} = S^2/2^n m$
$S^2\{b_{ii}\} = S^2/2\alpha^4 m$

Дисперсии коэффициентов не равны. Дисперсии предсказанных значений отличаются для разных точек факторного пространства.

Расчёт воспроизводимости, значимости и адекватности выполняется также как и в ПФЭ.

В центральном, композиционном, ротатабельном униформ планировании выбирают число нулевых точек таким образом, чтобы сохранилось постоянное значение дисперсии предсказания в окрестностях центра эксперимента (табл. 36). При этом $\alpha = 2^{n/4}$.

Таблица 36. Матрица эксперимента ЦКП для трёх факторов

№	x_0	x_1	x_2	x_3	x_1x_2	x_1x_3	x_2x_3	x'^2_1	x'^2_2	x'^2_3
1	1	+1	+1	+1	+1	+1	+1	0,2698	0,2698	0,2698
2	1	-1	+1	+1	-1	-1	+1	0,2698	0,2698	0,2698
3	1	+1	-1	+1	-1	+1	-1	0,2698	0,2698	0,2698
4	1	-1	-1	+1	+1	-1	-1	0,2698	0,2698	0,2698
5	1	+1	+1	-1	+1	-1	-1	0,2698	0,2698	0,2698
6	1	-1	+1	-1	-1	+1	-1	0,2698	0,2698	0,2698
7	1	+1	-1	-1	-1	-1	+1	0,2698	0,2698	0,2698
8	1	-1	-1	-1	+1	+1	+1	0,2698	0,2698	0,2698
9	1	+1,215	0	0	0	0	0	0,746	-0,7302	-0,7302
10	1	-1,215	0	0	0	0	0	0,746	-0,7302	-0,7302
11	1	0	+1,215	0	0	0	0	-0,7302	0,746	-0,7302
12	1	0	-1,215	0	0	0	0	-0,7302	0,746	-0,7302
13	1	0	0	+1,215	0	0	0	-0,7302	-0,7302	0,746
14	1	0	0	-1,215	0	0	0	-0,7302	-0,7302	0,746
15	1	0	0	0	0	0	0	-0,7302	-0,7302	-0,7302

Здесь: $\alpha = 1{,}215$; $\alpha^2 = 1{,}4762$; $x'^2_i = (2^3 + 2\times1{,}215^2)/(2^3 + 2\times3 + 1) = 0{,}7302 = \Theta$.

$1 - \Theta = 1 - 0{,}7302 = 0{,}2698$.

$\alpha^2 - \Theta = 1{,}4762 - 0{,}7302 = 0{,}7460$.

2.7.1. Пример применения ЦКП

Постановка задачи.

Известно, что морские организмы оказывают влияние на другие гидробионты метаболитами, выделяемыми в среду. К настоящему времени этот вопрос продолжает оставаться малоизученным. В данной работе сделана попытка изучения внешнеметаболического влияния растительноядных животных на некоторые физиологические функции растений. В качестве подопытных организмов выбраны представители массовых видов, обитающих в Чёрном море: двустворчатые моллюски – мидии и зелёная водоросль – ульва.

Известно, что освещённость – важнейший фактор, определяющий физиологические процессы растений. Изучаемое влияние должно быть представлено как функция освещённости и количества животных (или их массы) – доноров метаболитов.

Формулировка задачи: изучить влияние выделений мидий на рост водоросли ульвы при различной интенсивности освещённости.

Параметр оптимизации – величина прироста стандартного фрагмента таллома ульвы (в % за 4 суток).

Пояснение: стандартный фрагмент таллома Ø20 мм вырезался пробойником из средней части таллома, лишённой спороносителей.

Условия проведения опыта:

1. Общая продолжительность опыта – 4 суток;
2. Продолжительность светового периода в течение суток: 16 час;
3. Температура воды: 15 °С;
4. Объём экспериментальных сосудов: 3 л;
5. Перемешивание воды в сосудах осуществлялось микрокомпрессорами, т.е. непрерывной продувкой воздухом.
6. Вода для экспериментов бралась с поверхности моря на расстоянии 2 мили от берега;
7. Мидий ежедневно отсаживали на 2 часа в общую ёмкость для кормления одноклеточными водорослями, после чего мидий брали произвольно и переносили в экспериментальные сосуды в количестве, соответствующем плану эксперимента.

Выбор плана эксперимента. Предварительный опыт, поставленный по плану ПФЭ2^2, оказался недостаточным для получения адекватного уравнения регрессии. Поэтому было решено для проведения эксперимента выбрать план ЦКП для двух факторов (табл. 37, 38, 39).

Таблица 37. Список факторов, включаемых в эксперимент

№	Наименование фактора	Обозначение	Размерность	Область интереса	Ошибка измерения	Примечание
1	Освещённость	x_1	люкс	1000-6000	100	
2	Количество мидий доноров	x_2	штук	0 – 8	0	

Таблица 38. Кодирование факторов

Факторы	$\tilde{x}_i$	$\tilde{x}_1$	$\tilde{x}_2$	x_i
Верхний уровень	$\tilde{x}_{i верх}$	6000	8	+1
Нижний уровень	$\tilde{x}_{i нижн}$	1000	0	-1
Средний уровень	$\tilde{x}_{io}$	3500	4	0
Шаг варьирования	λ_i	2500	4	+1

Таблица 39. Матрица планирования, расчёта коэффициентов регрессии и результаты опытов

№	x_0	x_1	x_2	x_1x_2	x_1^2	x_2^2	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y	S_g^2	$\hat{y}$	$(\hat{y}-y)^2$
1	1	-1	-1	+1	1/3	1/3	38,65	8,42	16,51	5,74	33,03	31,45	11,84	20,81	176,73	16,79	16,14
2	1	+1	-1	-1	1/3	1/3	45,14	0	49,82	41,28	85,15	58,67	36,84	45,27	654,49	43,21	4,24
3	1	-1	+1	-1	1/3	1/3	24,44	44,51	12,55	24,88	41,59	31,72	14,27	27,71	153,36	14,43	176,24
4	1	+1	+1	+1	1/3	1/3	92,56	45,98	90,59	87,13	145,20	68,36	92,16	88,85	910,59	77,54	128,07
5	1	-1	0	0	1/3	-2/3	21,39	10,30	18,76	15,44	2,36	0	37,45	15,10	160,91	15,61	5,83
6	1	+1	0	0	1/3	-2/3	25,37	88,80	75,52	72,08	70,73	87,40	26,57	63,78	716,47	60,38	11,60
7	1	0	-1	0	-2/3	1/3	15,60	13,52	14,32	41,49	21,85	55,72	16,40	25,56	271,92	15,32	104,78
8	1	0	+1	0	-2/3	1/3	40,20	36,19	9,77	40,03	33,22	1,74	0	23,02	336,72	31,30	68,61
9	1	0	0	0	-2/3	-2/3	0	28,0	21,28	0	18,46	0	23,02	12,97	155,13	23,31	107,04
															Σ 3536,32		Σ 622,55

Статистическая обработка результатов эксперимента ЦКП.

Число повторных опытов: m **= 7.**

Число разных опытов: g **= 9.**

Всего опытов: N **= mg = 7·9 = 63.**

Звёздное плечо **α =1.**

Оценка воспроизводимости опытов:

$\mathbf{G_{max} = S^2_{gmax} / \Sigma S_g^{\,2} = 910{,}59 / 3536{,}32 = 0{,}257.}$

Табличный критерий Кохрена выбирается для:

α=0,05; f_1= g = 9; f_2 = m-1=7-1=6. В таблице для α=0,05 выбираем 9-ой столбец и 6-ую строку. Получаем $\mathbf{C_6^{\,9} = 0{,}368}$**.** Табличное значение больше расчётного; опыты воспроизводимы.

Расчёт коэффициентов уравнения регрессии выполняется по формулам:

$'b_0 = (\Sigma y_u)/g$, где g – количество разных опытов (число строк)

$b_i = (\Sigma x_{iu} y_u)/\Sigma x^2_{iu}$ $b_{ij} = (\Sigma x_{iu} x_{ju} y_u)/\Sigma\,(x_{iu} x_{ju})^2$

$b_{ii} = (\Sigma\, 'x^2_{iu} y_u)/\Sigma\,('x^2_{iu})^2$

$'b_0 = 323{,}07/9 =$ **35,9** $b_1 = 134{,}29/6 =$ **22,38** $b_2 = 47{,}95/6 =$ **7,99**

$b_{12} = (20{,}8-45{,}2-27{,}7+88{,}8)/4 =$ **9,18**

b_{11} = (6.93+15.09+9.24+29.62+5+21.26-17.04-15.41-8.71)/(6·1/9 + 3*4/9) = 45,98/2=22,99 = **23**

b_{22} = (6.93+15.09+9.24+29.62-10,12-42,73+8,52+7,67-8,69)/2 = 7,92 = **8**

Итак, получено уравнение регрессии:

$$\mathbf{y = 35{,}9 + 22{,}38x_1 + 8x_2 + 9{,}18x_1x_2 + 23x_1^{\,2} + 8x_2^{\,2}}$$

Проверка значимости коэффициентов уравнения регрессии (проводится по критерию Стьюдента).

Дисперсии коэффициентов для b_0:

$S^2\{b_0\} = S^2/N$; $S^2 = (\Sigma\, S^2_g)/g$; $S^2 = 3536{,}32/9 = 392{,}92$ $S^2\{b_0\} = 6{,}23$.

$t_0 = b_0 / \sqrt{S^2(b_0)} = 35{,}9/2{,}5 = 14{,}36$.

Табличный критерий Стьюдента при α=0,05 и числе степеней свободы: **v =** g(m-1)= 54

$t_i = 2{,}0 < 14{,}6$ Следовательно, b_0 – значим.

$S^2\{b_i\} = S^2/((2^n + 2\alpha^2)m) = 392{,}9/(2^2+2\cdot 1^2)7 = 392{,}9/42 = 9{,}33$ $S\{b_i\} = \sqrt{S^2\{b_i\}} = \sqrt{9{,}33} = 3{,}05$

Проверяем значимость b_2: $t_2 = 8/3{,}05 = 2{,}63$ $t_{табл} = 2{,}0 < 2{,}63$ b_2 –значим

b_1 – тем более значим.

Проверяем значимость коэффициента при парном взаимодействии:

$S^2\{b_{ij}\} = S^2/2^n m = 392{,}9/(2^2 \cdot 7) = 392{,}9/28 = 14$ $S\{b_{ij}\} = \sqrt{S^2\{b_{ij}\}} = \sqrt{14} = 3{,}74$

t_{12} =9,18/3,74=2,45 что больше табличного равного 2,0. Следовательно, b_{12} – значим.

Проверяем значимость коэффициентов квадратичных членов:

$S^2\{b_{ii}\} = S^2/2\alpha^4 m = 392{,}9/2\cdot 7 = 392{,}9/14 = 28$; $S\{b_{ii}\} = \sqrt{S^2\{b_{ii}\}} = \sqrt{28} = 5{,}3$

Проверяем b_{22}, как самый малый коэффициент:

t_{22} =8/5,3=1,51<2,0. Следовательно, b_{22} статистически не значим.

Уравнение в окончательном виде имеет вид:

$$\mathbf{y = 35{,}9 + 22{,}38x_1 + 8x_2 + 9{,}18x_1x_2 + 23x_1^2}$$

Проверка адекватности данного уравнения (выполняется так же, как и ПФЭ):

$F_{расч} = S^2_{ад} / S^2$, где $S^2_{ад} = (\Sigma(y_g - y'_g)^2/(g-d)$, где: d – число оставшихся членов в уравнении регрессии, включая свободный член.

$S^2_{ад}$=622,55/(9-5)= 155,6; $F_{расч}$=155,6/392,9 = 0,392

$F_{табл}$ для $\alpha = 0{,}05$ и f_1= g-d = 9-5=4 (столбец); f_2=g(m-1)=54 (строка)

$F_{табл}$ =2,5, что больше расчётного значения. Следовательно, уравнение адекватно описывает экспериментальные результаты.

Интерпретация результатов.

В уравнении в квадратичном члене фигурирует преобразованный фактор x_1. Для того, чтобы вернуться к обычным кодированным переменным, нужно внести изменения в соответствии с формулой: $b'_0 = b_0 - \Sigma\, b_{ii}\, x'^2_i$.

Для двух факторов $b'_0 = b_0 - (b_{11}\cdot 2/3 + b_{22}\cdot 2/3) = b_0 - 2/3(b_{11} + b_{22})$ = 35,9 – 2/3 (23+8)= 35,9 – 20,5= 15,4. Поэтому уравнение, предназначенное для использования обычных кодированных переменных, имеет вид:

$$\mathbf{y = 15{,}4 + 22{,}38x_1 + 8x_2 + 9{,}18x_1x_2 + 23x_1^2}$$

Все факторы положительные, что говорит о том, что эксперимент поставлен в зоне лимитирования роста водоросли светом и метаболитами мидий. При увеличении уровней обоих факторов возможно ускорение роста ульвы.

Положительное влияние метаболитов мидий заметно на фоне такого сильного фактора, как освещённость. Для получения более чёткого представления о роли обоих факторов построим кривые зависимости величины прироста от освещённости в отсутствие мидий и при максимальном их количестве равном 8 шт.

Для построения графика (рис. 13) освещённость задаём на уровнях:

1000; 2250; 3500; 4750; 6000 люкс

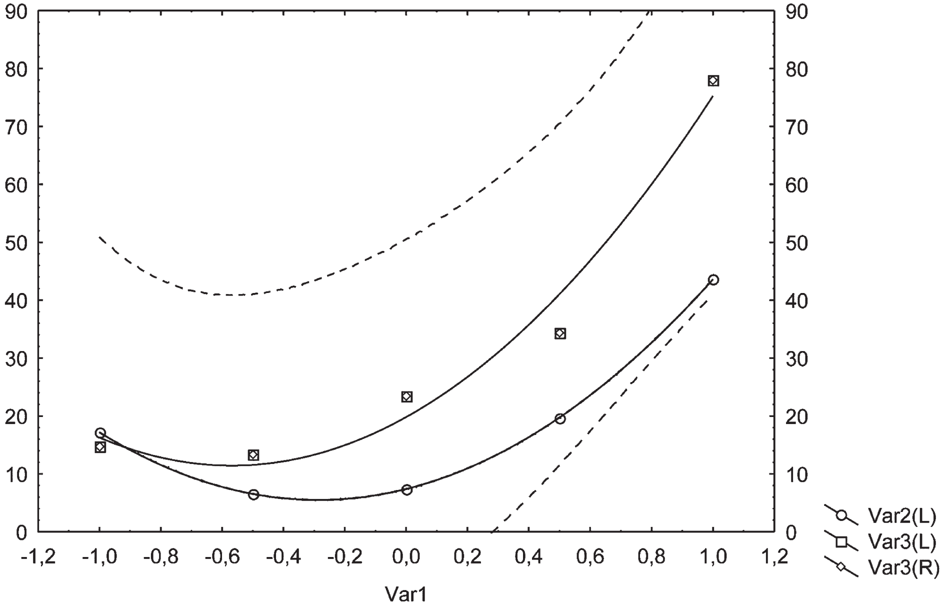


Рис. 13. Кривые зависимости прироста ульвы (% за 4 сут.) от освещённости в отсутствие мидий – доноров метаболитов (нижняя сплошная) и в присутствии мидий (верхняя сплошная)

Для подстановки значений уровней фактора натуральные значения переводим в кодированные по формуле:

$\mathbf{x_i = (x'_i - x'_{i0}) / \lambda'_i}$

При $\mathbf{x'_1 =}$ 1000 лк имеем: $\mathbf{x_1 =}$ (1000 – 3500)/2500 =-1;
2250 $\mathbf{x_1 =}$ (2250 – 3500)/2500 =-0,5;
3500 $\mathbf{x_1 =}$ (3500 – 3500)/2500 =0;
4750 $\mathbf{x_1 =}$ (4750 – 3500)/2500 =0,5;
6000 $\mathbf{x_1 =}$ (6000 – 3500)/2500 =1.

1) В отсутствие мидий ($\mathbf{x_2}$ = -1) при подстановке вместо фактора $\mathbf{x_1}$ его кодированных значений получим:

$x_1 = -1$; $x_2 = -1$; y = 15,4-22,38-8+9,18+23=17,2;
$x_1 = -0,5$; $x_2 = -1$; y = 15,4-11,2-8+4,59+5,75=6,55;
$x_1 = 0$; $x_2 = -1$; y = 15,4-8 = 7,4;
x_1 =0,5; $x_2 = -1$; y = 15,4 + 11,2-8 – 4,59+5,75= 19,76;
$x_1 = +1$; $x_2 = -1$; y = 15,4+22,38-8-9,18+23=43,6.

2) При максимальном количестве мидий ($\mathbf{x_2}$ = +1):

$x_1 = -1$; $x_2 = +1$; y = 15,4-22,38+8-9,18+23=14,8;

x_1 = -0,5; x_2 = +1; y = 15,4-11,2+8-4,59+5,75=13,36;
x_1 = 0; x_2 = +1; y = 15,4+8=23,4;
x_1 = +0,5; x_2 = +1; y = 15,4+11,2+8+4,59+5,75=34,9;
x_1 = +1; x_2 = +1; y = 15,4+22,38+8+9,18+23=77,96.

При низких значениях освещённости метаболиты мидий не оказывают влияния на рост ульвы. Это влияние проявляется в максимальной мере при высокой освещённости, благоприятной для фотосинтеза.

Глава 3

ПЛАНЫ ДИСПЕРСИОННОГО АНАЛИЗА

3.1. Дисперсионный анализ (ДА)

Если бы все факторы, действующие в окружающем нас мире, были только количественными (без качественных), тогда не нужно было бы изобретать ДА, потому что, при планировании экспериментов, можно было бы ограничиться только регрессионным анализом.

В первой главе отмечалось, что цель регрессионного анализа – это количественное изучение зависимости величины Y от величины X. Именно от **величины** X, что означает, что значения X можно увеличивать или уменьшать. В соответствии с этими изменениями и изменяется Y. На основании этой связи и определяют расчётным путём зависимость Y от X, то есть функцию Y=f(X).

Однако качественный фактор невозможно ни уменьшить, ни увеличить. Уровни качественного фактора произвольно откладывают на горизонтальной оси (оси фактора). Поэтому при работе с качественным фактором определение функции Y=f(X) становится невозможным.

Примеры качественных факторов:

1) Судейская бригада на международных соревнованиях по фигурному катанию состоит из судей; здесь фактор: «судья». Этот фактор принимает уровни: американец, японец, русский, француз, канадец и т.д. Здесь перечислены качественные уровни.

2) В химических реакторах можно использовать разные катализаторы. Фактор «тип катализатора» принимает качественные уровни (платина, окись титана и т.д.).

3) В производстве изделий из пластмасс используют разные типы наполнителей и красителей. Здесь «наполнитель» и «краситель» – качественные факторы.

4) При экстрагировании веществ из сырья используют различные растворители: спирт, бензол, ацетон, хлороформ, эфир. Здесь качественный фактор: «тип растворителя».

5) Лекарственные препараты испытывают на разных типах больных (качественный фактор). Типы больных выделяют в зависимости от вида болезней, которыми переболели больные; состояния их сердечно-сосудистой системы; возраста; пола; упитанности и т.д.

6) Тип подопытного животного, уровни: самцы или самки; до линьки или после линьки; половозрелые или не половозрелые и т.д.

ДА – статистический метод, с помощью которого производится разложение суммарной дисперсии на составляющие. Факторы могут быть как количественные, так и качественные, но они должны быть обязательно независимы. Важно, чтобы результаты были нормально распределены. Уточним, что речь идёт о дисперсии результатов, полученных в опытах. Ясно, что при варьировании факторов изменяются и значения результатов, на которые влияют и неучтённые факторы.

В многофакторных экспериментах с помощью ДА определяется величина дисперсии, обусловленная действием каждого фактора в отдельности, а также действием их взаимодействий и оценивается статистическая значимость этих эффектов на фоне ошибки эксперимента.

Допустим, что в одинаковых условиях выполнили n – наблюдений, что позволило получить n – результатов: y_1; y_2; y_3 ... y_n – рассеяние которых можно оценить дисперсией S^2.

$S^2 = (\Sigma(y_i - \tilde{y})^2/(n-1)$ – выборочная дисперсия, или оценка дисперсии генеральной совокупности: $\tilde{y}$ – среднее арифметическое. Дисперсия генеральной совокупности обозначена как Σ^2.

Если некоторый фактор «А» оказывает влияние на результат, то есть y = f(A), тогда общая дисперсия результатов будет состоять из дисперсии, обусловленной фактором А и ошибки измерения: $S^2_{общее} = S^2_{а} + S^2_{ошибки}$. Здесь $S^2_{а}$ – часть общей дисперсии, которая обусловлена воздействием фактора А.

В многофакторном случае (если модель линейна):

y = f(A, B, C), $S^2_{общее} = S^2_{а} + S^2_{в} + S^2_{с} + S^2_{ошибки}$

Если модель нелинейная, тогда:

$S^2_{общее} = S^2_{а} + S^2_{в} + S^2_{с} + S^2_{ав} + S^2_{ас} + S^2_{вс} + S^2_{авс} + S^2_{ошибки}$

Общую дисперсию можно разложить на составляющие. Разложение дисперсии основано на её свойстве: **для вероятностно независимых случайных величин дисперсии обладают свойством аддитивности,** иными словами, дисперсию можно разложить на слагаемые (Шефе, 1963). Аддитивность дисперсий – не просто результат независимости случайных величин, но и, в связи с эти свойством, при многофакторном эксперименте, независимость факторов является строго обязательным условием применения дисперсионного анализа.

В дисперсионном анализе, в отличие от регрессионного анализа, нет требования непрерывности поверхности отклика, что позволяет работать с качественными факторами. Поверхность отклика представляет собой многомерную решётку, а значение «y» существует только в узлах решётки.

Качественные факторы, сами по себе, могут иметь сложную структуру. Например, фактор эксперт состоит из качеств: национальность, образование, религия, научная школа и т.д.

Основные типы задач, решаемых с помощью ДА (Маркова, Лисенков, 1973):

1) Элиминирующий эксперимент – это задача исключения влияния источника неоднородностей (качественных факторов), например, различных партий сырья или реактивов, аппаратов, методик, исполнителей, индивидуальных свойств организмов, неоднородностей среды, в которой находятся подопытные животные и т.д. Получаемый в эксперименте результат будет защищён от влияния выбранного качественного фактора.

Модель такова: $y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \varepsilon_{ijk}$

Данное уравнение означает, что рассматриваемый конкретный экспериментальный результат y_{ijk} складывается (алгебраическая сумма) из общего среднего результата μ, эффектов экспериментальных факторов α, β и γ, находящихся на указанных уровнях, и ещё от остатка ε, в который входят влияния всех неучтённых в эксперименте факторов.

Допустим, что γ_k – это индивидуальность больного. Нужно так спланировать эксперимент, чтобы исключить (элиминировать) влияние индивидуальности больного.

2) Сравнительный или сопоставительный эксперимент. Допустим, что фактор А – это аминокислоты: a_1, a_2, a_3, ... Нужно расположить эти дискретные элементы в ряд по возрастанию некоторого признака в них. Сравнительный эксперимент позволит корректно провести исследование для достижения этой цели.

3) Неполноблочный эксперимент.

Если, например, в биохимических исследованиях источником неоднородности является различие в партии, допустим, водорослей (сырья), то считается, что опыты, поставленные на одной и той же партии, проводятся в однородных условиях. Партии водорослей формируют понятие блока. Если партии водорослей не настолько велики, чтобы на них провести все варианты экспериментов, тогда возникает задача преодоления трудностей, связанных с ограниченным размером блока. Эта задача решается путем постановки неполноблочного эксперимента.

4) Отсеивающий эксперимент с качественными факторами.

При разработке, допустим, новых методик, подбора какой-то рецептуры и т.д., сталкиваются с большим числом качественных и количественных факторов, заданных на многих уровнях. Например, растворитель – качественный фактор. Беря различные типы растворителей, тем самым задают различные уровни этого фактора.

Допустим, создаётся новый материал. Для этого в качестве факторов выбираются:

полимер (A) с уровнями a_1 a_2 a_3 ...
наполнитель (B) с уровнями $в_1$ $в_2$ $в_3$ ...
краситель (C) с уровнями c_1 c_2 c_3 ...
антиоксидант (D) с уровнями d_1 d_2 d_3 ...

Нужно отсеять неинтересные варианты, то есть сочетания уровней различных факторов. Таких сочетаний может быть много, поэтому невозможно проверить их все, вследствие чего возникает необходимость сведения объёма работы к минимуму, и в то же время свести к минимуму вероятность пропуска интересных вариантов и таким образом получить надёжные выводы.

5) Экстремальный эксперимент. Оптимизация.

Такой эксперимент возможен в ДА, если все факторы качественные; либо часть факторов качественная, а часть – количественная:

- Качественные факторы. При планировании экспериментов с дискретными уровнями качественных факторов нужно иметь в виду, что поверхность отклика представляет собой многомерную решётку, поэтому крутое восхождение – невозможно. В таком случае используются сложные комбинаторные планы для дискретных уровней. Основное преимущество – сокращение объёма перебора вариантов.
- Качественные и количественные факторы. Качественная составляющая отделяется от эффекта количественных факторов. Это даёт возможность крутого восхождения по количественным факторам.

Допустим, что возникла необходимость в постановке эксперимента с 6 факторами, каждый из которых варьируется на пяти уровнях. В этом случае потребуется испытать $5^6 = 15625$ комбинаций, а для этого надо поставить такое же количество опытов, что практически невозможно. Экспериментальный план нужно составить так, чтобы каждый уровень фактора встречался с другим уровнем других факторов хотя бы только один раз, что существенно сократит количество разных опытов.

3.1.1. Однофакторный дисперсионный анализ (однофакторный эксперимент)

Составление плана.

План эксперимента составляется на основе однофакторного дисперсионного анализа.

Искомая модель: $y_{im} = \mu + \alpha_i + \varepsilon_{im}$

y_{im} – зависимая переменная (исследуемая характеристика, параметр оптимизации);

μ – генеральное среднее эксперимента;

α_{i} – эффект i-того уровня фактора А;

ε_{im} – остаточный член, с помощью которого оценивается ошибка эксперимента, которая формируется за счёт воздействия всех неучтённых факторов. Если ошибка эксперимента окажется слишком большой – необходимо выявить причины этого и дополнительно включить ещё фактор, который может существенно влиять на результат.

Общая (суммарная) дисперсия складывается из дисперсии, обусловленной уровнем фактора А и ошибкой эксперимента:

$S^2_{общее} = S^2_{a} + S^2_{ошибки}$

В табличной форме составляется план для однофакторного эксперимента. В таблице 40 столбцы – это варианты опытов, а строки – это повторности. Фактор варьируется на k уровнях (число вариантов опыта). Каждый вариант повторяется n раз. Повторы должны обеспечить получение репрезентативных результатов, например, их можно будет экстраполировать на всё озеро, а не только на отдельный его участок. Если же пробы взяты на одном участке, то они не могут дать информацию о всём озере, поэтому их называют мнимыми повторностями (Козлов, 2003, 2014). Сбор проб для повторностей должен планироваться, например, с применением метода рандомизации, либо латинских планов.

Таблица 40. Экспериментальный план на основе однофакторного дисперсионного анализа

Уровни фактора А

Повторы	1	2	i	…………..	k
1	y_{11}	y_{21}	y_{i1}	…………..	yk_{1}
2	y_{12}	y_{22}	y_{i2}	…………..	y_{k2}
. . . m . . . n	y_{1m} y_{1n}	y_{2m} y_{2n}	y_{im} y_{in}	…………..	y_{km} y_{kn}
ΣA_{i}	A_{1}	A_{2}	A_{i}	…………..	A_{k}
$A_{i\ среднее}$	$\tilde{y}_{1}$	$\tilde{y}_{2}$	$\tilde{y}_{i}$	…………..	$\tilde{y}_{k}$

Предварительная обработка данных.

По каждому столбцу вычисляем средние: $\tilde{y}_1$, $\tilde{y}_2$, $\tilde{y}_3$, ... $\tilde{y}_k$.

Формулируется гипотеза H_0: *Между столбцами нет различия*, то есть математические ожидания средних значений всех результатов равны между собой:

$my_1 = my_2 = my_3 = ... = my_k$

1. Итоги: $A_{i\ сред} = (\Sigma\ y_{im})/n$; ($A_{i\ сред}$ это $\tilde{y}_i$).
2. Общий итог: $G = \Sigma\ \Sigma\ y_{im}$ (сумма результатов всех наблюдений) $G = \Sigma A_i$.
3. Средний итог $G_{средн} = (\Sigma\ \Sigma\ y_{im})/kn$.

Вычислительная процедура для дисперсионного анализа.

1. Подсчёт общей суммы квадратов:

$SS_{общ} = \Sigma\Sigma y^2_{im} - G^2/N$

G^2/N – корректирующий или поправочный член. $N=k \cdot n$.

2. Сумма квадратов итогов по столбцам:

$SS_a = \Sigma((A^2_i)/n) - G^2/N$.

3. Ошибка опыта:

$SS_{ошиб} = SS_{общ} - SS_a$.

Таблица 41. Сводная таблица дисперсионного анализа

Источник дисперсии	Число степеней свободы	Сумма квадратов	Средний квадрат	Критерий Фишера
Фактор А α_i	k -1	SS_a	$MS_a = SS_a/(k-1)$	$F_{эксп} = MS_a / MS_{ошиб}$
Ошибка ε_{im}	N-k=k(n-1)	$SS_{ошиб}$	$MS_{ошиб} = SS_{ошиб} / k(n-1)$	$F_{эксп} > F_{табл}$
Итог	kn – 1	$SS_{общ}$		

Если $F_{эксп} > F_{табл}$, тогда гипотеза H_0 не проходит (средний квадрат фактора А значимо отличается от среднего квадрата ошибки). В таком случае определяем, какие средние отличаются друг от друга. Это делается с помощью рангового критерия Дункана (Приложение III).

1. Определение доверительных интервалов для средних:

$A'_i +/- t_{1-\alpha/2}\sqrt{(MS_{ошиб}/n)}$.

2. Нахождение средних, значимо отличающихся друг от друга. Для этого нужно упорядочить средние, т.е.: A'_i.
3. Определение нормированных ошибок:

$S_н = \sqrt{(MS_{ошиб}/n)}$.

Далее следует обратиться к табличным значениям значимых рангов (Хикс, 1967, Приложение III) Из таблицы Дункана значимых рангов с выбранным уровнем значимости α, числом v, равным числу степеней свободы среднего квадрата ошибки k(n-1) и p=2, 3, . . . k следует выписать k -1 значений рангов.

4. Эти значения рангов нужно умножить на $S_н$, что необходимо для получения группы из k -1 наименьших значимых рангов.
5. Проверка получаемых разностей между средними, начиная с крайних; разность максимального и минимального значений среднего сравнивается с наименьшим значимым рангом при p= k и т.д., пока не будут исследованы все k(n-1) /2 возможные пары.

Пример проведения однофакторного дисперсионного анализа

Выполнили измерения концентрации растворённого органического вещества морской воды в трёх разных акваториях (табл. 42).

Таблица 42. Средние значения концентраций растворённого органического вещества (РОВ) из разных акваторий

Испытание \ акватории	1	2	3
$A'_{i.}$	16	2	10

Наблюдения проводились в трёх разных акваториях (k = 3); каждое испытание повторялось 6 раз (данные по каждому повтору не приведены).

1. Упорядочить средние по возрастанию и присвоить им номера:

2	10	16
2	3	1

2. Данные из таблицы дисперсионного анализа (здесь таблица не приведена)

Число степеней свободы: k(n-1) = 3·(6-1) =15.

Средний квадрат ошибки: $MS_{ошиб}$= 14,8 (из таблицы ДА).

3. Нормированная ошибка: $S_н = \sqrt{(14{,}8/6)} = 1{,}57$.
4. Из таблицы Дункана v= 15 и 5%-го уровня значимости значимые ранги p:

p	2	3
	3,01	3,16

5. Наименьшие значимые ранги (НЗР), (умноженные на нормированную ошибку 1,57) равны:

p	2	3
НЗР	4,73	4,96

Наибольшее против наименьшего (1-е против 2-го, т.е.:16–2=14): 14>4,96 (различаются).

Наибольшее против третьего, (1-е против 3-го): 16 – 10 =6>4,73 (различаются).

Обнаруживается значимая разность между всеми испытаниями.

Пример практического применения ДА.

Исследуется влияние типа аквариума на выживаемость личинок устриц (%) в течение 21 дня.

Тип аквариума – это фактор А, заданный на четырёх уровнях: a_1, a_2, a_3, a_4. То есть:

$i = 1, ...4; k=4$.

В данном случае выполняется сравнительный эксперимент (сравнивается выживаемость личинок по пяти повторным наблюдениям, n=5). Общее число наблюдений: $N=k \cdot n=4 \cdot 5=20$.

Модель эксперимента: $y_{im} = \mu + \alpha_i + \varepsilon_{im}$; $S^2_{общее} = S^2_a + S^2_{ошибки}$.

Результаты наблюдений представлены в таблице 43.

Формулируем гипотезу: «Между разными значениями выживаемости нет значимого различия».

Если средние отличаются друг от друга не более чем дисперсия внутри каждой выборки, тогда статистически значимого различия между средними нет.

Таблица 43. Выживаемость личинок устриц (%) в аквариумах четырех типов

Повтор	a_1	a_2	a_3	a_4
1	56	64	45	42
2	55	61	46	39
3	62	50	45	45
4	59	55	39	43
5	60	56	43	41

Для упрощения дальнейших расчётов, вычтем из каждого значения число 50, что не окажет влияния на выводы:

Таблица 44. Данные таблицы 43 после преобразования

Повтор	a_1	a_2	a_3	a_4
1	6	14	-5	-8
2	5	11	-4	-11
3	12	0	-5	-5
4	9	5	-11	-7
5	10	6	-7	-9
ΣA_i	42	36	-32	-40
‘A_i	8,4	7,2	-6,4	-8

Общий итог: $G=\sum_{i=1}^{k} A_i = \sum_{i=1}^{k} \sum_{m=1}^{n} y_{im} = 6$

Сумма квадратов наблюдений:

$$\sum_{i=1}^{k} \sum_{m=1}^{n} y^2_{im} = 1340$$

1. Общая сумма квадратов $SS_{общ} = \Sigma \Sigma y^2_{im} - G^2/kn = 1340 - 6^2/20 =$ 1338,2.

2. Сумма квадратов итогов по столбцам: $SS_a = \Sigma((A^2_i)/n) - G^2/N =$ $42^2/5 + +36^2/5 + (-32)^2/5 + (-40)^2/5 - 6^2/20 = 1135,0$.

3. Ошибка опыта: $SS_{ошиб} = SS_{общ} - SS_a = 1338,2 - 1135,0 = 203,2$.

Таблица 45. Сводная таблица результатов дисперсионного анализа

Источник дисперсии	Число степ свободы	Сумма квадратов SS	Средний квадрат, MS	F-критерий	Вывод
A	k-1=4-1=3	1135,0	378,3	F=378,3/12,7= =29,8	$F_{табл} < F$ Гипотеза отверга-ется
Ошибка	к(n-1) = 4(5-1)=16	203,2	12,7	$F_{табл}$=3,24 ст своб 3, 16	
Сумма	кn-1= 20-1=19	1138,2			

С помощью рангов из таблицы Дункана определяем, какие именно значения выживаемости отличаются друг от друга.

Установление доверительных пределов для средних:

A'_i +/- $t_{1-\alpha/2}\sqrt{(MS_{ошиб}/n)}$, где $t_{1-\alpha/2} = t_{0,975;\,16} = 2,12$.

A'_i+/-2,12$\sqrt{12,7/5}$ = A'_i+/-3,37.

A'_1=8,4 +/-3,37 = 5,03 – 11,77 (в преобразованном виде); A'_2; A'_3; A'_4.

В натуральных значениях: 55,03 – 61,77. Необходимо упорядочить средние:

A'_i: A'_4=-8; A'_3=- 6,4; A'_2= 7,2; A'_1=8,4.

Подсчитываем нормированные ошибки:

Нормированная ошибка: $S_н = \sqrt{12,7/5} = 1,59$.

Обращаемся к табличным значениям. У нас четыре типа аквариумов. Сколько будет сравнений? Каждое значение сравнивается с тремя остальными. Значит ранг = 3; v= k(n-1) = 4(5-1) =16.

2	3	4
3,00	3,15	3,23

Наименьшие значимые ранги (ранг, умноженный на нормированную ошибку):

4,77; 5,01; 5,13.

Сравниваем наибольшее с наименьшим:

8,4 – (-8) = 16,4.

Наибольший ранг равен 5,13 16≥5,13

$A'_1 \geq A'_4$ ($\alpha_1 \geq \alpha_4$) то есть $\alpha_1 \neq \alpha_4$.

Аналогично сравниваем A_1 и A_3:

8,4 – (-6,4) = 14,8 14,8≥5,01 $\alpha_1 \neq \alpha_3$

A_1 и A_3: 8,4 – 7,2 = 1,2 1,2<4,77 $\alpha_1 = \alpha_3$.
Сравниваем A_2 с A_4 и с A_3:
A_2 с A_4: 7,2 – (8,0)= 15,2 15,2>5,01 $\alpha_2 \neq \alpha_4$.
A_2 с A_3: 7,2 – (-6,4) =13,6 13,6>4,77 $\alpha_2 \neq \alpha_3$.
Сравниваем A_3 с A_4:
-6,4 – (-8,0) =1,6 1,6<4,77 $\alpha_2 = \alpha_4$.
Выводы по результатам дисперсионного анализа данных:

1. Статистически значимые различия не обнаружены между выживаемостью в аквариумах следующих типов: 1 и 2; 3 и 4;
2. Лучшие значения выживаемости получены в аквариумах 3 и 4 типов. Рекомендации по выбору самого лучшего следует сделать на основании экономического и технологического анализов.

Пример применения однофакторного дисперсионного анализа при изучении влияния типа корма на линейный рост молоди мидий.

Эксперимент проводили в лаборатории ИМБИ (в период с 9.01. 2014 г. – 10.02.2014 г.). В качестве корма использовали 2 вида диатомовых водорослей *Cylindroteca closterium* и *Phaeodactylum tricornutum*. Суммарная биомасса корма всегда составляла 56 г/л. Качественный состав корма (уровни качественного фактора) задавался по следующей схеме:

1-ый уровень – 100 % *C closterium* + 0 % *Ph tricornutum*
2-ой уровень – 75 % *C closterium* + 25 % *Ph tricornutum*
3-ий уровень – 50 % *C closterium* + 50 % *Ph tricornutum*
4-ый уровень – 25 % *C closterium* + 75 % *Ph tricornutum*
5-ый уровень – 0 % *C closterium* + 100 % *Ph tricornutum*

Смена воды и подача корма осуществлялась ежедневно. Эксперимент проводили в трех повторах. Спат мидий по 6 экз. поместили в чашки Петри и установили в ванне с морской водой, объемом 2 л. при постоянной аэрации. Первые сутки – адаптация и прикрепление спата. Температура воды в течение эксперимента изменялась от 15,1°С до 17,2 °С. Среднее значение температуры 16,4 °С.

Статистическая обработка результатов эксперимента

Итак, тип корма – качественный фактор А, заданный на пяти уровнях: a_1, a_2, a_3, a_4, $a_{5,}$ то есть: i= 1, ...5; k=5. Это – сравнительный эксперимент (сравниваются темпы роста по трём повторным наблюдениям, n=3). Общее число наблюдений: N=k×n=5×3=15.

Модель эксперимента: $y_{im} = \mu + \alpha_i + \varepsilon_{im}$; $S^2_{общее} = S^2_a + S^2_{ошибки}$

Результаты эксперимента (таб. 46):

Таблица 46. Линейный прирост мидий за месяц, мм

Повторы	Уровни				
	a_1	a_2	a_3	a_4	a_5
1	1,792	1,991	2,341	2,4	1,925
2	1,35	1,808	2,325	2,259	1,859
3	1,783	2,083	2,383	2,167	1,9
ΣA_i	4,925	5,882	7,049	6,826	5,684
$‘A_i$	1,642	1,961	2,350	2,275	1,895

Формулируем гипотезу H_0: «Между различными значениями линейного прироста нет значимого различия».

Если средние отличаются друг от друга не более, чем дисперсия внутри каждой выборки, тогда значимого различия между средними нет.

Общий итог: $G=\sum_{i=1}^{k} A_i=\sum_{i=1}^{k}\sum_{m=1}^{n} y_{im} = 30{,}366$

Сумма квадратов наблюдений:

$$\sum_{i=1}^{k}\sum_{m=1}^{n} y^2_{im} = 62{,}680$$

1. Общая сумма квадратов $SS_{общ}= \Sigma\Sigma\ y^2_{im} - G^2/kn = 62{,}680 - 30{,}366^2/15= 62{,}680 - 61{,}473 = 1{,}207$
2. Сумма квадратов итогов по столбцам: $SS_a= \Sigma((A^2_i)/n) - G^2/N = 4{,}925^2/3 + 5{,}882^2/3 + (7{,}049)^2/3 + (6{,}826)^2/3 +5{,}684^2/3 - 30{,}366^2/15=8{,}085 + 11{,}533 + 16{,}563 + 15{,}531 + 10{,}769 - 61{,}473 = 1{,}008$
3. Ошибка опыта: $SS_{ошиб} = SS_{общ} - SS_a= 1{,}207 - 1{,}008 = 0{,}199$

Таблица 47. Сводная таблица дисперсионного анализа:

Источник дисперсии	Число степеней свободы	Сумма квадратов	Средний квадрат	Критерий Фишера
Фактор А α_i	k -1=5-1=4	SS_a=1,008	$MS_a= SS_a/(k-1)= =0{,}252$	$F_{эксп}= MS_a/ MS_{ошиб}= =0{,}252/0{,}02=12{,}6$
Ошибка ε_{im}	N-k=k(n-1)= 5(3-1)=10	$SS_{ошиб}= =0{,}199$	$MS_{ошиб}= SS_{ошиб}/ /k(n-1)=0{,}02$	$F_{эксп}>F_{табл}$ 12,6>5,2
Итог	kn – 1=14	$SS_{общ}$=1,207		Гипотеза отвергается

4. Если $F_{эксп}>F_{табл}$, то гипотеза H_0 не проходит (средний квадрат фактора А значимо отличается от среднего квадрата ошибки). В таком случае определяем, какие средние отличаются друг от друга. Это делается с помощью рангового критерия Дункана.

5. Какие средние значимо отличаются друг от друга? Для этого нужно упорядочить средние, т.е.: $'A_3$; $'A_4$; $'A_2$; $'A_5$; $'A_1$.
   2,350 2,275 1,961 1,895 1,642
6. Подсчитываем нормированные ошибки:

$S_н = \sqrt{(MS_{ошиб}/n)} = \sqrt{(0,02/3)}=0,082$ и устанавливаем доверительные интервалы для средних: A'_i +/- $t_{1-\alpha/2}\sqrt{(MS_{ошиб}/n)}$, где $t_{1-\alpha/2}= t_{0,975;\,5}$ $=2,57$ $'A_i$ +/- $2,57\cdot 0,08 = A'_i$ +/- 0,206:

$'A_3$= 2,350+/- 0,206=2,556 – 2,144

$'A_4$=2,275+/- 0,206=2,481 – 2,069

Обращаемся к табличным значениям. Из таблицы Дункана значимых рангов с выбранным уровнем значимости $\alpha=0,05$, числом v, равным числу степеней свободы среднего квадрата ошибки k(n-1) =10 и p=2, 3, . . . k выписываем k -1 значений рангов (то есть четыре значения).

2	3	4	5
3,15	3,30	3,37	3,43

7. Умножаем эти значения рангов на $S_н$=0,082 и получаем группу из k-1 наименьших значимых рангов:
   0,258; 0,271; 0,276; 0,281
8. Проверяем наблюдаемые нами разности между средними, начиная с крайних; разность максимального и минимального значений среднего сравниваем с наименьшим значимым рангом при p= k и т.д., пока не будут исследованы все k(n-1) /2 возможные пары:

- сравнение максимального и минимального значений средних: 2,350–1,642=0,708 0,708>0,281, следовательно, $'A_3 > 'A_1$
- сравнение $'A_3$ и $'A_5$: 2,350-1,895=0,455 0,455>0,276; $'A_3 > 'A_5$
- сравнение $'A_3$ и $'A_2$: 2,350-1,961=0,389 0,389>0,271; $'A_3 > 'A_2$
- сравнение $'A_3$ и $'A_4$: 2,350-2,275=0,075 0,075<0,258; $'A_3$ и $'A_4$ статистически не различаются.
- сравнение $'A_2$ и $'A_1$: 1,961-1,642=0,319 0,319>0,276; $'A_2 > 'A_1$
- сравнение $'A_2$ и $'A_5$: 1,961-1,895=0,066 0,066<0,271; $'A_2$ и $'A_5$ статистически не различимы.
- сравнение $'A_1$ и $'A_5$: 1,895-1,642=0,253 0,253<0,258; $'A_1$ и $'A_5$ статистически не различимы.

Выводы дисперсионного анализа результатов эксперимента по изучению влияния состава корма на линейный рост мидий.

1. Темпы роста минимальны при использовании монокультур и не зависят от вида испытанных кормовых водорослей.

2. Максимальные темпы линейного роста получены в 3-ем и 4-ом вариантах, причём результаты, полученные в обоих вариантах, статистически не различаются.
3. Темпы линейного роста мидий при использовании кормовой смеси во втором варианте превосходят темпы роста при мидий, питавшихся кормом 1-го варианта, но статистически не отличаются от темпов роста при кормлении кормом 5-го варианта.
4. Темпы линейного роста мидий при использовании кормовой смеси во втором варианте ниже темпов роста, полученных в 3-ем и 4-ом вариантах.

3.1.2. Двухфакторный эксперимент на конкретном примере

Таблица 48. Исследование влияния типа аквариума и режима продувки воды на выживаемость личинок устриц.

Тип аквариума	A	a_1, a_2, a_3
Продувка (есть, нет)	B	b_1, b_2
Выживаемость, %	Y	y_1, y_2, y_3 (три наблюдения)

Таблица 49. Таблица условий и результатов наблюдений:

B\A	a_1	a_2	a_3	B_j	B'_j
b_1	4 6 Σ 15 9	8 10 Σ 25 7	2 5 Σ 13 6	53	5,9
b_2	-6 -5 Σ-15 -4	0 -4 Σ-9 -5	-8 -7 Σ-21 -6	-45	-5
A_i	0	16	-8	G=8	
A'_i	0	2,67	-1,33		

Вычислительная процедура для подсчёта сумм квадратов отклонений:

Количество вариантов испытаний: $n=a \cdot b=3 \cdot 2=6$

Количество повторов: $m=3$.

Общее число наблюдений: $N=18$.

$G= \Sigma A_i= \Sigma B_j=8$.

Поправочный член: $G^2/N = 8^2/18 = 3,56$.

Сумма квадратов итогов по столбцам: $SS_a = (0^2+ 16^2+(-8)^2)/6 - 3,56 = 49,7$, где: 6 – число элементов в столбце.

Сумма квадратов итогов по строкам: $SS_b = (53^2 + (-45)^2)/9 - 3,56 = 535,55$.

Сумма итогов: $SS_a + SS_b =49,77 + 533,55 = 583,32$.

Сумма квадратов наблюдений: $\Sigma\Sigma y_{ij}^2=622$.

Общий итог: $SS_{общ} = \Sigma\Sigma y_{ij}^{2} - G^2/N = 622 - 3{,}56 = 618{,}44$.

Сумма квадратов итогов по испытаниям: $SS_{испытан} = 15^2/3 + 25^2/3 + \ldots + (-21)^2/3 - G^2/N = 585{,}11$.

Остаток: $SS_{ост} = SS_{общ} - SS_{испытан} = 618{,}44 - 585{,}11 = 33{,}33$.

Сумма квадратов взаимодействий: $SS_{ab} = SS_{испытан} - (SS_a + SS_b)$, так как: $SS_{испытан} = SS_a + SS_b + SS_{ab}$.

$SS_{ab} = 585{,}11 - 583{,}32 = 1{,}79$.

Таблица 50. Результаты дисперсионного анализа двухфакторного эксперимента с тремя повторами и с учётом межфакторного взаимодействия.

Источники дисперсии	Число степеней свободы	Сумма квадратов SS	Средний квадрат MS	$F_{эксп}$	$F_{табл}$	H_0: $\alpha_i = 0$ $\beta i = 0$ $\alpha_i\,\beta_i = 0$
A	a – 1 3 – 1 = 2	49,77	49,77/2= =24,88	24,88/2,78= =8,95	6,93 f_1=2 f_2=12	$F_{эксп} > F_{табл}$ $\alpha_i \neq 0$
B	b – 1 2 – 1 = 1	535,55	535,55/1= 535,55	535,55/2,78= =191,92	9,33 f_1=1 f_2=12	$F_{эксп} > F_{табл}$ $\beta_i \neq 0$
AB	(a-1)(b-1) 2*1=2	1,79	1,79/2=0,89	0,89/2,78= =0,32	F < 1	$\alpha_i\,\beta_i = 0$
Ошибка	N – n 18 – 6 = 12	33,33	33,33/12 =2,78			
Сумма	N-1 18 – 1 = 17	618,44				

Так как $F_{эксп} > F_{табл}$ в обоих случаях, то гипотеза $\alpha_i = 0$ и $\beta_i = 0$ отвергается, т.е. средние в обоих случаях статистически различаются, точнее говоря, средние квадраты обоих факторов значимо отличаются от среднего квадрата ошибки.

Для определения лучшего типа аквариума и лучшего режима барботирования нужно, как в однофакторном эксперименте, выполнить сравнение средних по критерию Дункана.

Тип аквариума:

1	2	3
0	2,67	-1,33
3	1	2
-1,33	0	2,67

Нормированная ошибка: $S_н = \sqrt{(MS_{ошиб}/n)} = \sqrt{(2{,}78/6)} = 0{,}68$

Всего типов аквариумов три, поэтому рангов 2.

2 3

3,08 3,23 – это ранги из таблицы Дункана.

Умножим ранги на нормированную ошибку и получим НЗР:

2,09 2,20 НЗР.

Сравниваем тип аквариумов 2 и 3:

2,67 – (-1,33) = 4,00>2,20 (при α = 0,05), т.е. 2 лучше, чем 3.

Сравниваем 2 и 1: 2,67 – 0 = 2,67>2,09, т.е. 2 лучше, чем 1.

Сравниваем 1 и 3:

0 – (-1,33) = 1,33<2,09, т.е. 1 и 3 не различаются.

Режимы барботирования не сравниваются, так как уже известно, что режим 1 (наличие продувки) лучше режима 2 (отсутствие продувки).

Таким образом, лучший тип аквариума – 2; лучший режим барботирования – 1.

Важное дополнение к обработке результатов данного эксперимента.

Сотрудницей ИМБИ Д.С. Балычевой была обнаружена ошибка в подсчёте суммы квадратов наблюдений ($\Sigma\Sigma y_{ij}^2$=622, что неверно).

В действительности $\Sigma\Sigma y_{ij}^2$=678 и результаты обработки, а также выводы существенно меняются. Однако в данной книге наряду с правильными расчётами (см. ниже) приведена статистическая обработка, использующая ошибочный подсчёт. Благодаря этому читатель сможет проследить всю процедуру обработки данных по схеме двухфакторного дисперсионного анализа, а также формулирование выводов.

Вычислительная процедура для подсчёта сумм квадратов отклонений:

Количество вариантов испытаний: n=a·b=3·2=6.

Количество повторов: m=3.

Общее число наблюдений: N=18.

$G = \Sigma A_i = \Sigma B_j = 8$.

Поправочный член: $G^2/N = 8^2/18 = 3{,}56$.

Сумма квадратов итогов по столбцам: $SS_a = (0^2 + 16^2 + (-8)^2)/6 - 3{,}56 = 49{,}7$, где: 6 – число элементов в столбце.

Сумма квадратов итогов по строкам: $SS_b = (53^2 + (-45)^2 /9 - 3{,}56 = 535{,}55$.

Сумма итогов: $SS_a + SS_b = 49{,}77 + 533{,}55 = 583{,}32$.

Сумма квадратов наблюдений: $\Sigma\Sigma y_{ij}^2 = 678$.

Общий итог: $SS_{общ} = \Sigma\Sigma y_{ij}^2 - G^2/N = 678 - 3{,}56 = 674{,}44$.

Сумма квадратов итогов по испытаниям: $SS_{испытан} = 15^2/3 + 25^2/3 + ...+ (-21)^2/3 - G^2/N = 585{,}11$.

Остаток: $SS_{ост} = SS_{общ} - SS_{испытан} = 674{,}44 - 585{,}11 = 89{,}33$.

Сумма квадратов взаимодействий: $SS_{ab} = SS_{испытан} - (SS_a + SS_b)$, так как: $SS_{испытан} = SS_a + SS_b + SS_{ab}$.

$SS_{ab} = 585{,}11 - 583{,}32 = 1{,}79$.

Таблица 51. Результаты дисперсионного анализа 2-х факторного эксперимента с тремя наблюдениями и с учётом взаимодействия.

Источники дисперсии	Число степеней свободы	Сумма квадратов SS	Средний квадрат MS	$F_{эксп}$	$F_{табл}$	H_0: $\alpha_i = 0$ $\beta i = 0$ $\alpha_i\ \beta_i = 0$
A	a – 1 3 – 1 = 2	49,77	49,77/2= =24,88	24,88/7,44= =3,34	6,93 f_1=2 f_2=12	$F_{эксп}<F_{табл}$ $\alpha_i = 0$
B	b – 1 2 – 1 = 1	535,55	535,55/1= 535,55	535,55/7,44= =71,98	9,33 f_1=1 f_2=12	$F_{эксп}>F_{табл}$ $\beta_i \neq 0$
AB	(a-1)(b-1) 2·1=2	1,79	1,79/2=0,89	0,89/7,44= =0,12	F < 1	$\alpha_i\ \beta_i = 0$
Ошибка	N – n 18 – 6 = 12	89,33	89,33/12 =7,44			
Сумма	N-1 18 – 1 = 17	618,44				

Поскольку для фактора «тип аквариума» $F_{эксп}<F_{табл,}$ то в данном случае гипотеза $\alpha_i = 0$ подтверждается. Но во втором случае $F_{эксп}>F_{табл}$, поэтому гипотеза и $\beta_i = 0$ отвергается, т.е. средние в первом случае статистически не различаются, а во втором – различаются, точнее говоря, средние квадраты фактора B значимо отличаются от среднего квадрата ошибки.

Для того, чтобы определить лучший тип барботирования нужно, как в однофакторном эксперименте, выполнить сравнение средних по критерию Дункана.

Однако, в данном случае режимы барботирования не сравниваются, так как уже известно, что режим 1 (наличие продувки) лучше 2-го (отсутствие продувки).

Таким образом, все типы аквариумов одинаковы, а лучший режим барботирования – первый.

3.2. Неполные классификации дисперсионного анализа

Если в опытах реализуются все возможные совокупности условий, задаваемых выбранной схемой эксперимента, тогда такой экспериментальный план называют полной классификацией ДА. В противном случае экспериментальный план является неполной классификацией ДА. Например, латинский квадрат – это 1/n от ПФЭ n^3. 1. Полные классификации ДА (небольшое число факторов).

- Однофакторный ДА
- Двухфакторный ДА
- Трехфакторный ДА

1. Неполные классификации ДА (сокращение перебора вариантов)
2.1. Без ограничения на рандомизацию.
2.2. С ограничением на рандомизацию.
2.2.1. Ортогональные планы:
- Латинские планы.
- Латинские и гипер – греко-латинские планы.

2.2.2. Неортогональные планы.
- Неполноблочные планы (блок-схемы).
- Сбалансированные и частично сбалансированные блок-схемы).

3.2.1. Латинские квадраты

Латинским квадратом n×n называют квадратную таблицу, состоящую из n элементов, чисел или букв, такую, что каждый из элементов встречается в точности 1 раз в каждой строке и в каждом столбце. Напомним, что разные латинские буквы соответствуют разным уровням качественного фактора. Например, n=3:

A B C
B C A
C A B

Это **диагональный квадрат**, в котором каждая буква соответствует уровню качественного фактора, допустим, «вид корма», где А – цистозира; В – ульва; С – зостера. Для практических целей квадрат рандомизируют, т.е. случайным образом переносят столбцы и строки, не нарушая их.

Квадрат может быть записан **в канонической форме** (J_n), где буквы в первой строке и в первом столбце следуют в алфавитном порядке. Такие квадраты также называют **стандартными**:

A B C D
B D A C
C A D B
D C B A

Сопряжёнными стандартными квадратами называют такую пару стандартных квадратов, для которой строки одного являются столбцами другого.

Возможное число квадратов (табл. 52) зависит от размера квадрата (n – размер, Jn – число стандартных (канонических); N – общее число квадратов).

Таблица 52. Количество канонических латинских квадратов (Jn) и общее число латинских квадратов (N) размером n.

n	2	3	4	5	6	7
J_n	1	1	4	56	9408	16 942 080
N	2	12	576	161 280	812 851 200	…………

В каждой канонической форме имеется: Jn = n!(n-1)! квадратов.

Общее число квадратов: $N = J_n n!(n\text{-}1)!$

Латинский квадрат – это 1/n от ПФЭ n^3. Если имеется 3 фактора на n уровнях, то число опытов ПФЭ равно n^3, а в латинском квадрате n^2 опытов.

Достоинства латинских квадратов:

- Возможность постановки элиминирующего эксперимента;
- Сокращение перебора вариантов;
- Линейные эффекты оцениваются независимо.

Модель, исследуемая латинскими квадратами, является линейной:
$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \varepsilon_{ijk}$

Межфакторные взаимодействия не оцениваются, поэтому латинские квадраты применяют для исследования качественных факторов, которые не взаимодействуют.

Однако, если модель не линейна, то латинский квадрат достраивается до ПФЭ с помощью сбалансированного ряда латинских квадратов.

Сбалансированный ряд латинских квадратов – это набор квадратов, в котором каждая буква появляется в каждой возможной позиции один и только один раз, например:

$$\begin{matrix} C_1 & C_2 & C_3 \\ C_3 & C_1 & C_2 \\ C_2 & C_3 & C_1 \end{matrix} \qquad \begin{matrix} C_2 & C_3 & C_1 \\ C_1 & C_2 & C_3 \\ C_3 & C_1 & C_2 \end{matrix} \qquad \begin{matrix} C_3 & C_1 & C_2 \\ C_2 & C_3 & C_1 \\ C_1 & C_2 & C_3 \end{matrix}$$

Каждый из приведённых квадратов является 1/3 от ПФЭ 3^3; в сумме они дают ПФЭ.

3.2.2. Статистический анализ латинских квадратов без повторных опытов

Таблица 53. Латинский квадрат 3х3 (три фактора на трех уровнях).

A/B	b_1	b_2	b_3	Итоги по строкам
a_1	C_1 Y_1	C_2 Y_2	C_3 Y_3	A_1
a_2	C_3 Y_4	C_1 Y_5	C_2 Y_6	A_2
a_3	C_2 Y_7	C_3 Y_8	C_1 Y_9	A_3
Итоги по столбцам	B_1	B_2	B_3	G

Это план трёхфакторного эксперимента. Факторы А и В могут быть количественными, либо качественными и каждый из них варьируется на трёх уровнях, например: $a_1\, a_2\, a_3$. Фактор С – качественный на трёх уровнях: C_1, C_2, C_3.

Предполагаем, что линейная модель адекватна. Эффекты взаимодействий не определяются; они входят в остаточный член.

1. Подсчитываем итоги (суммы) по строкам (A_1, A_2, A_3); по столбцам (B_1, B_2, B_3); по элементам квадрата (C_1, C_2, C_3).
2. Общий итог: $G = A_1 + A_2 + A_3 = B_1 + B_2 + B_3 = C_1 + C_2 + C_3 = \Sigma\, y_{ijk}$.
3. Подсчитываем вспомогательную сумму квадратов отклонений:
4. $SS_{общ} = \Sigma\Sigma\, y^2_{ijk} - G^2/n^2$, n=3 (число наблюдений).
5. $SS_a = \Sigma(A^2_i/n) - G^2/n^2$ (Сумма квадратов итогов по строкам).
6. $SS_b = \Sigma(B^2_j/n) - G^2/n^2$ (Сумма квадратов итогов по столбцам).
7. $SS_c = \Sigma(C^2_j/n) - G^2/n^2$ (Сумма квадратов итогов по латинским буквам).
8. Остаточный член: $SS_{ост} = SS_{общ} - (SS_a + SS_b + SS_c)$.

Таблица 54. Сводная таблица дисперсионного анализа латинского квадрата 3 × 3

Источник дисперсии	Число степеней свободы	MS	$F_{эксп}$	$F_{табл}$	H_0: $\alpha_i = 0$ $\beta j = 0$ $\gamma_k = 0$
A	n – 1	$SS_a/(n-1)$	$MS_a/ MS_{ост}$	Один	$F^a_{эксп} > F_{табл}$ $\alpha_i \neq 0$
B	n – 1	$SS_b/(n-1)$	$MS_b/ MS_{ост}$	и тот	$F^b_{эксп} > F_{табл}$ $\beta_j \neq 0$
C	n – 1	$SS_c/(n-1)$	$MS_c/ MS_{ост}$	же	$F^c_{эксп} < F_{табл}$ $\gamma_k = 0$
Остаток	(n – 1)(n – 2)	$SS_{ост}/(n-1)(n-2)$			
Сумма	$n^2 - 1$				

Для значимых источников дисперсии наступает второй этап статистического анализа по ранговому критерию Дункана (так же, как и для одно- и двухфакторных экспериментов ДА).

3.3. Применение латинского квадрата при поиске оптимальных условий для роста личинок гигантской устрицы (*Пиркова А.В., Ладыгина Л.В.*)

Технология выращивания устриц в полноцикличных морских хозяйствах базируется на использовании посадочного материала, то есть устричной молоди (спата), который получают в специальных питомниках (Пиркова, Холодов, Ладыгина, 1998; Ладыгина, Пиркова, 2002; Пиркова, Ладыгина, 2004). Определение оптимальных условий для выращивания личинок устрицы *Crassostrea gigas* в питомнике и подращивание спата для его последующего размещения на ферме является важным этапом в биотехнике культивирования этого моллюска. В предварительных исследованиях было установлено, что для оптимизации условий выращивания личинок, прежде всего, требуется определить экспериментальным путём оптимальные значения концентрации корма, его качественный состав и плотность посадки личинок в выростных баках.

В качестве первого этапа поиска оптимальных условий принят отбор наиболее перспективных сочетаний трёх важных факторов. С этой целью был поставлен эксперимент по схеме латинского квадрата 4×4, который проводился в два этапа: 1- личинки на стадии велигера; 2 – личинки на стадии великонхи. В таблицах 55 и 56 приведен план эксперимента по изучению зависимости суточного прироста (мкм/сут) личинок устрицы на стадии велигера и великонхи при варьировании двух количественных факторов (концентрации корма и плотности посадки личинок) и одного качественного (состава корма).

Исследовали влияние трех факторов:

x_1 – плотность посадки личинок (уровни 5, 10, 15 и 20);

x_2 – концентрация корма (уровни 50, 10, 150 и 200);

x_3 – качественный фактор (уровни A, B, C и D). A – *Isochrysis galbana;* B – *Isochrysis galbana* + *Phaeodactilum tricornutum* + *Dunaliella viridis;* C – *Isochrysis galbana* + *Dunaliella viridis;* D – *Isochrysis galbana* + *Phaeodactilum tricornutum*

Обработка результатов выполнена в два этапа по расчётным схемам, предложенным Ч. Р. Хиксом (Хикс, 1967):

1) дисперсионный анализ (см. Приложение A, B. Приложения приведены в конце данного раздела);

2) регрессионный анализ (см. Приложение C).

Данный подход позволяет на первом этапе ориентировочно выявить оптимальную область, а на втором – выполнить более подробное исследование оптимальной области, что необходимо для определения оптимальных значений исследуемых факторов.

В результате математической обработки данных было показано, что среднесуточный прирост гигантской устрицы на стадии велигера не зависит от заданной в опыте концентрации корма ($F_{эксп.} = 2{,}92 < F_{табл.} = 4{,}8$) и плотности посадки личинок ($F_{эксп.} = 1{,}62 < F_{табл.} = 4{,}8$).

Таблица 55. План эксперимента по исследованию зависимости суточного прироста (мкм/сут) личинок *Crassostrea gigas* на стадии велигера от концентрации корма, его состава и плотности посадки личинок

Концентрация корма, тыс.кл./мл	Плотность личинок, тыс. лич./л			
	5	10	15	20
50	11,1 A 1 1,96	5,7 B 5 12,86	10,0 C 9 8,82	7,4 D 13 5,58
100	7,5 B 2 25,72	9,7 C 6 17,64	8,9 D 10 11,16	10,5 A 14 3,92
150	5,0 C 3 26,46	11,1 D 7 16,74	8,4 A 11 5,88	4,5 B 15 38,58
200	6,3 D 4 22,32	10,1 A 8 7,84	3,3 B 12 51,44	3.4 C 16 35,28

Примечание: в верхнем левом углу каждой ячейки: среднесуточный прирост личинок, мкм/сут; в правом нижнем – сырая биомасса водорослей, мг/л

Таблица 56. План эксперимента по исследованию зависимости суточного прироста (мкм/сут) личинок *Crassostrea gigas* на стадии великонхи от концентрации корма, его состава и плотности посадки личинок

Концентрация корма, тыс.кл/мл	Плотность личинок, тыс. лич./л			
	3	5	7	9
50	19,6 A 1 13,61	26,1 B 5 2,29	18,3 C 9 13,94	22,4 D 13 10,15
100	21,5 B 2 4,58	14,9 C 6 6,97	18,7 D 10 20,30	15,5 A 14 27,22
150	10,4 C 3 41,82	14,2 D 7 30,45	14,9 A 11 40,83	21,1 B 15 6,87
200	10,7 D 4 40,60	12,1 A 8 54,44	21,5 B 12 9,16	11,4 C 16 55,76

A – *Isochrysis galbana + Tetraselmis suecica;* B – *Isochrysis galbana + Chaetoceros calcitrans;* C – *Chaetoceros calcitrans + Tetraselmis suecica;* D – *Isochrysis galbana + Chaetoceros calcitrans + Tetraselmis suecica*;

Фактор, значимо влияющий на среднесуточный прирост личинок на стадии велигера, – состав корма (Fэксп. = 5,34 > Fтабл. = 4,8) (табл.3 и Приложение А).

Для проверки различия средних значений для значимых факторов (состав корма) применяли множественный ранговый критерий Дункана.

Таблица 57. Результаты дисперсионного анализа латинского квадрата 4×4

Источник дисперсии	Число степеней свободы	Средний квадрат, Ms	Критерий Фишера Fэксп.	Критерий Фишера Fтабл.	Соответствие
A	3 (n-1)	8,98	2,92	4,8	Fэкс. < Fтабл.
B	3 (n-1)	4,96	1,62	4,8	Fэкс. >Fтабл.
C	3 (n-1)	16,39	5,34	4,8	Fэкс. > Fтабл.
Ошибка	6 (n-1) (n-2)	3,07		$f_1 = 3$ $f_2 = 6$ $\alpha = 0,05$	
Итого	15 (n^2-1)				

Проверка по критерию Дункана выполнялась следующим образом:

1) подсчитывалась нормированная ошибка –

$S_H = S_A\sqrt{\frac{Msocm}{4}} = \sqrt{\frac{3,07}{4}} = 0,77$

2) доверительные интервалы средних, по формуле:

$\pm t_{\alpha}^{n-1} \cdot S_n \cdot [t_{0,05}^{3} = 3,18]$; $\pm t = 3,18 \cdot 0,77 = 2,44$

по латинским буквам:

$C_1 = 10,02 \pm 0,77$ $C_3 = 7,02 \pm 0,77$

$C_2 = 5,25 \pm 0,77$ $C_4 = 8,42 \pm 0,77$

3) из таблицы Дункана значимых рангов для уровня значимости 0,05 и числа степеней свободы, равного числу степеней свободы среднего квадрата ошибки и р =2, 3, 4 выписывали n – 1 значимых рангов

р	2	3	4
ранги	3,46	3,58	3,64

4) находили группу наименьших значений рангов (НЗР)

р	2	3	4
НЗР	2,66	2,75	2,80

5) упорядочили средние значения по возрастанию по латинским буквам

C_1	C_4	C_3	C_2
10,02	8,42	7,02	5,25

6) Подсчитывалась разность между средним, начиная с крайних. Разность максимального и минимального значений сравнивалась и наибольшим значимым рангом при р = n, т.е.

а) $\overline{C}_1 - \overline{C}_2 = 4,77 > 2,8$
в) $\overline{C}_4 - \overline{C}_2 = 3,17 > 2,75$
с) $\overline{C}_1 - \overline{C}_4 = 1,6 < 2,75$
d) $\overline{C}_4 - \overline{C}_3 = 1,4 < 2,66$
e) $\overline{C}_1 - \overline{C}_3 = 3,0 > 2,66$
f) $\overline{C}_3 - \overline{C}_2 = 1,77 < 2,66$

Таким образом, между всеми значениями средних различия статистически значимы, кроме C_2 и C_3 (смеси водорослей В и С).

Путем сравнения максимального и минимального значений с наименьшим значимым рангом показано, что величина среднесуточного прироста личинок достоверно различается при варьировании состава корма.

Максимальный прирост был отмечен при использовании корма, состоящего из *I. galbana,* и смешанного корма из *I. galbana+Ph. tricornutum*. Отмеченный прирост достоверно выше, чем при корме, состоящем из *I. galbana+ P. tricornutum + D. viridis* и *I. galbana +D. viridis*. Следовательно, *D. viridis* не является подходящим кормом для личинок устрицы на стадии велигера. Наблюдалась обратная зависимость прироста личинок от биомассы корма (r = – 0,81, р = 0,05). Максимальный среднесуточный прирост был отмечен при концентрации корма 50 тыс. кл./мл, состоящем из *I. galbana* (биомасса 1,96 х 10^6 мг/л). Примерно такой же прирост личинок –10,5 мкм/сут. при плотности посадки 20 тыс. лич./л и концентрации корма *I. galbana* 100 тыс. кл./мл. Среднее значение биомассы водорослей, рассчитанное на 1 тыс. личинок, соответствует 0, $392 \cdot 10^6$ мг/л, т.е. расход биомассы водорослей в другом опыте в два раза меньше. Следовательно, оптимальные условия для роста велигеров устрицы следующие: состав корма – *I. galbana* в пределах 50-100 тыс. кл. /мл и плотность посадки от 5 до 20 тыс. лич./л.

В эксперименте с личинками устриц на стадии великонхи суммарная концентрация корма была задана в диапазоне 50-200 тыс. кл./мл, плотность посадки личинок 3, 5, 7 и 9 тыс. лич./л (табл. 54).

Концентрация микроводорослей и состав корма оказались значимыми факторами для роста личинок на стадии великонхи: Fэксп. = 34,01 > Fтабл. = 4,8 и F эксп. = 43,22 > Fтабл. = 4,8 (для числа степеней свободы 3 и 6 и 5% уровней значимости) (табл. 58 и Приложение В).

Проведена проверка средних значений по факторам – концентрация корма и состав корма. Для проверки различия средних значений для значимых факторов применяли множественный ранговый критерий Дункана.

Таблица 58. Результаты дисперсионного анализа латинского квадрата 4×4

Источник дисперсии	Число степеней свободы	Средний квадрат, Ms	Критерий Фишера Fэксп.	Критерий Фишера Fтабл.	Соответствие
A	3 (n-1)	45,91	34,01	4,8	Fэксп. >Fтабл.
B	3 (n-1)	5,82	4,31	4,8	Fэксп. < Fтабл.
C	3 (n-1)	58,35	43,2	4,8	Fэксп. >Fтабл.
Ошибка	6 (n-1) (n-2)	1,35		$f_1 = 3$	
Итого	15 (n^2-1)			$f_2 = 6$ $\alpha = 0.05$	

Значения по строкам и по латинским буквам были расположены в порядке возрастания их значений:

A_1	A_2	A_3	A_4
21,6	17,7	15,5	13,9
C_2	C_4	C_1	C_3
22,6	16,5	15,5	13,8

Подсчитанная нормированная ошибка – S_n равнялась 0,29. Доверительные интервалы средних:

$\pm t_{\alpha}^{n-1} \cdot S_n \cdot [t_{0,05}^{3} = 3,18]; \pm t = 3,18 \cdot 0,29 = 0,92$

а) доверительный интервал по строкам:

$A_1 = 21,6 \pm 0,92$ $A_3 = 15,5 \pm 0,92$

$A_2 = 17,7 \pm 0,92$ $A_4 = 13,9 \pm 0,92$

б) доверительный интервал по латинским буквам:

$C_1 = 15,5 \pm 0,92$ $C_3 = 13,8 \pm 0,92$

$C_2 = 22,6 \pm 0,92$ $C_4 = 16,5 \pm 0,92$

Из таблицы Дункана значимых рангов для уровня значимости 0,05 и числа степеней свободы, равного числу степеней свободы среднего квадрата ошибки и p =2, 3, 4, выписывали n – 1 значимых рангов:

p	2	3	4
ранги	3,46	3,58	3,64

находили группу наименьших значений рангов (НЗР)

p	2	3	4
НЗР	1,00	1,04	1,06

Подсчитывалась разность между средними наибольшими и наименьшими значениями средних и сравнивалась с наименьшим значимым рангом

а) по строкам:

$\overline{A}_1 - \overline{A}_4 = 7{,}7 > 1{,}06$ $\quad$ $\overline{A}_2 - \overline{A}_4 = 3{,}8 > 1{,}04$

$\overline{A}_1 - \overline{A}_3 = 6{,}4 > 1{,}04$ $\quad$ $\overline{A}_2 - \overline{A}_3 = 2{,}5 > 1{,}00$

$\overline{A}_1 - \overline{A}_2 = 3{,}9 > 1{,}00$ $\quad$ $\overline{A}_3 - \overline{A}_4 = 1{,}3 > 1{,}00$

б) по латинским буквам

$\overline{C}_2 - \overline{C}_3 = 8{,}8 > 1{,}06$ $\quad$ $\overline{C}_4 - \overline{C}_3 = 2{,}7 > 1{,}04$

$\overline{C}_2 - \overline{C}_1 = 7{,}1 > 1{,}04$ $\quad$ $\overline{C}_4 - \overline{C}_1 = 1{,}0 > 1{,}00$

$\overline{C}_2 - \overline{C}_4 = 6{,}1 > 1{,}00$ $\quad$ $\overline{C}_1 - \overline{C}_3 = 1{,}7 > 1{,}00$

Определив разность между средними значениями и величинами наименьших значимых рангов, выявили достоверные различия темпа роста личинок при разных уровнях концентрации и состава корма.

Максимальный среднесуточный прирост личинок отмечен при концентрации корма 50 тыс. кл/мл и составлял 21,6 мкм/сут. при 100 тыс. кл./мл – 17,7 мкм/сут.; при 150 тыс. кл./мл – 15,2 мкм/сут. и при 200 тыс. кл./мл – 13,9 мкм/сут. Различия всех значений средних статистически достоверны. Коэффициент корреляции прироста и концентрации корма составил r = – 0,86 (P = 0,05), то есть наблюдается обратная связь темпов роста и концентрации (биомассы) корма.

Состав корма – значимый фактор для поддержания оптимального темпа роста личинок. Наибольший среднесуточный прирост (22,6 мкм/сут.) отмечен при составе корма *I. galbana* + *C. calcitrans*; затем при составе корма из трех видов микроводорослей: *I. galbana*+ *C. calcitrans* + *T. suecica* (16 мкм/сут.); далее при *I. galbana* + *T.suecica* (15,5 мкм/сут.) и *C. calcitrans* + *T. suecica* (13,8 мкм/сут.). Между всеми значениями средних величин прироста различия статистически достоверны.

Фактор «плотность посадки» оказался незначимым для среднесуточного прироста личинок, так как не выявлено достоверных различий среднесуточного прироста личинок при плотности посадки от 3 до 9 тыс. лич./л. При сравнении темпа роста личинок в разных опытах (среднесуточный прирост 21,5 и 21,1 мкм/сут. соответственно), в условиях плотности посадки 3 и 9 тыс. лич./л и одинаковом составе корма (*I. galbana*+ *C. calcitrans)* среднее значение биомассы, рассчитанное на 1 тыс. личинок, в опыте оказалось в два раза выше.

Следовательно, оптимальными условиями для роста личинок устрицы *C. gigas* на стадии великонхи являются: концентрация корма в пределах 50–150 тыс.кл./мл; корм, состоящий из *I. galbana* + *C. calcitrans* в соотношении клеток 1:1; плотность посадки личинок в пределах 3–9 тыс. лич./л.

В зависимости от способа культивирования плотность посадки личинок гигантской устрицы может задаваться от 10 тыс. лич./л (на стадии велигера) до 1-2 тыс. лич./л (на стадии великонхи) или 800 и 1000 тыс. лич./л соответственно в закрытой или проточной системе при достаточном количестве корма.

По результатам ПФЭ 2^2 получено уравнение зависимости среднесуточного прироста личинок на стадии велигера от плотности посадки личинок и концентрации корма. Были исследованы факторы:

1) плотность посадки личинок, – X_1
нижний уровень (–1) – 3 тыс. лич./ л
верхний уровень (+ 1) – 7 тыс. лич./ л;
2) концентрация корма – X_2
верхний уровень (+ 1) – 70 тыс.кл./мл
нижний уровень (– 1) – 30 тыс.кл./ мл;
Состав корма: *I. galbana*+ *C. calcitrans* (2: 1).

Результаты исследования влияния плотности посадки (X_1) и концентрации корма (X_2) на прирост личинок *C. gigas* приведены в табл.59.

Таблица 59. План эксперимента ПФЭ 2^2 и результаты исследования влияния плотности посадки (X_1) и концентрации корма (X_2) на прирост личинок *Crassostrea gigas*

№ опыта	X_1	X_2	$X_1 \cdot X_2$	y			ŷ	Sg	Sg^2
1	+	+	+	3,59	3,69	4,11	3,80	0,28	0,0784
2	-	+	-	3,27	3,22	3,64	3,38	0,23	0,0529
3	+	-	-	1,96	2,19	1,87	2,01	0,17	0,0289
4	-	-	+	2,67	2,89	2,57	2,71	0,16	0,0256

$\Sigma Sg^2 = 0,1858$

Дальнейшая статистическая обработка результатов сводилась к проверке воспроизводимости результатов опыта по критерию Фишера:

$G_{max} = 0,4220 < G_{табл.} = 0,74$ ($\alpha = 0,05$)

Результаты воспроизводимы.

Затем были произведены расчеты коэффициентов уравнения регрессии:

$\Delta y_{1,2} = b_0 + b_1x_1 + b_2x_2 + b_{1,2} \cdot x_1x_2$
$b_0 = 2,975$; $b_1 = -0,07$;
$b_2 = 0,62$; $b_{1,2} = 0,265$

Проверка значимости коэффициентов уравнения по критерию Стьюдента

$$t_1 = \frac{b_1}{\sqrt{S^2\langle bi\rangle}}\text{; где: } S^2\{b_i\} = \frac{S^2}{mg}\text{; } S^2 = \frac{\Sigma S^2 g}{g}$$

Критерий Стьюдента при x=0,05 и числе степеней свободы 12 равен 2,18

$S^2 = 0,04645$; $S^2\{b_i\} = 0,0029$; $\sqrt{0,0029}=0,0539$

$t_0 = 55,19 > 2.18$ – значим;

$t_1 = -1,3 < 2,18$ – не значим;

$t_2 = 11,5 > 2,18$ – значим;

$t_{1,2} = 4,92 > 2.18$ – значим.

Окончательно уравнение регрессии зависимости среднесуточного прироста личинок на стадии велигера от плотности посадки личинок и концентрации корма имеет вид:

$\Delta y_{1,2} = 2,98 + 0,62 \cdot X_2 + 0,27 \cdot X_1 \cdot X_2$

Проведена проверка адекватности данного уравнения экспериментальным данным:

	разность	квадрат
$\hat{y}_{g1} = 2,98 + 0,62(+1) + 0,27(+1) = 3,87$	0,07	0,0049
$\hat{y}_{g2} = 2,98 + 0,62(+1) + 0,27(-1) = 3,33$	0,05	0,0025
$\hat{y}_{g3} = 2,98 + 0,62(-1) + 0,27(-1) = 2,09$	0,08	0,0064
$\hat{y}_{g4} = 2,98 + 0,62(-1) + 0,27(-1) = 2,09$	0,08	0,0064
		$\Sigma = 0,0202$

$$F = \frac{S^2 ag}{S^2};\ S_{ag}^2 = \frac{m}{g-d} \cdot \Sigma(\hat{y}_{g1} - \hat{y}_{g2})^2 = 2 \cdot 0,0202 = 0,0404$$

$$S^2 = \frac{\Sigma S^2 g}{g} = 0,4645;\ F = 0,87$$

Следовательно, уравнение адекватно описывает экспериментальные данные.

Разность между фактическими данными и рассчитанными по уравнению лежит в пределах от 0,05 до 0,08 мкм/сут. Концентрация корма в данном случае – наиболее важный фактор для роста личинок, превышающий в 2,3 раза величину влияния взаимодействия изучаемых факторов. Среднесуточный прирост личинок возрастает с увеличением концентрации корма. Заданная в опыте плотность посадки личинок оказалась несущественным фактором для их роста. Увеличение плотности посадки оказывает относительно небольшое положительное влияние только при одновременном повышении концентрации корма (положительное взаимодействие факторов X_1 и X_2). Максимальный среднесуточный прирост отмечен при плотности посадки 3 тыс. лич./л и концентрации корма 70 тыс. кл./мл.

На выживаемость личинок на стадии велигер изученные факторы не оказывают существенного влияния (табл. 60).

Таблица 60. Зависимость выживаемости личинок *C. gigas* на стадии велигера от плотности их посадки (X_1) и концентрации корма (X_2)

№ опыта	X_1	X_2	$X_1 \cdot X_2$	y	ŷ	Sg	S_g^2
1	+	+	+	74,6; 87,7; 89,3	83,87	8,06	64,96
2	-	+	-	80,6; 91,7; 92,6	88,3	6,68	44,62
3	+	-	-	80,6; 80,6; 4,6	78,6	346	11,97
4	-	-	+	87,0; 78,1; 75,8	80,3	5,92	35,05

$\Sigma S_g^2 = 156{,}6$

Дальнейшая статистическая обработка результатов проводилась в следующей последовательности:

1) оценка воспроизводимости результатов опыта по критерию Фишера

$G_{max} = 0,4148 < G_{табл.} = 0,74$ ($\alpha = 0,05$)

Опыт воспроизводим.

2) расчет коэффициентов уравнения регрессии

$b_0 = 82,77$ $b_2 = 3,32$

$b_{1} = -1,53$ $b_{1,2} = -0,68$

3) проверка значимости коэффициентов

$S^2 = 39{,}15$; $S^2\{b_i\} = 244685$; $\sqrt{2{,}44685} = 1{,}5642$

$t_0 = 53{,}06 > 2,18$ – значим;

$t_1 = 0{,}98 < 2,18$ – не значим;

$t_2 = 2{,}13 < 2,18$ – не значим;

$t_{1,2} = 0{,}44 < 2,18$ – не значим

Уравнение зависимости выживаемости личинок на стадии велигера от изучаемых факторов имеет вид:

$\Delta Y_{1,2} = 82{,}77$.

Следовательно, выживаемость личинок не зависит от концентрации корма и плотности посадки.

Однако наблюдается тенденция зависимости выживаемости личинок на стадии велигера от концентрации водорослей. Проверка значимости коэффициента по критерию Стьюдента показала, что: $t_2 = 2{,}13 < t_{st} = 2{,}18$.

Темп роста личинок на поздних стадиях развития описывается уравнением регрессии (Приложение С):

$\Delta Y_{1,2} = 18{,}86 + 4{,}37X_1$.

Таким образом, среднесуточный прирост личинок гигантской устрицы зависит в основном от заданной в опыте плотности их посадки. Максимальный прирост отмечен при плотности 3 тыс. лич./л, а минимальный – при 7 тыс. лич./л.

Уравнение регрессии выживаемости личинок, проходящих поздние стадии развития, имеет следующий вид:

$\Delta Y_{1,2} = 80{,}72 - 2{,}70 X_1 + 7{,}82X_2$

Концентрация водорослей является определяющим фактором, влияющим на выживаемость личинок на стадиях великонхи и педивелигера. Его влияние в 2,8 раза превышает влияние фактора плотности посадки личинок. Из уравнения следует, что выживаемость личинок максимальная при низкой плотности посадки и высоком уровне концентрации корма. Выживаемость максимальна при концентрации корма 70 тыс. кл./мл, и при плотности посадки личинок – 3 тыс. лич./ л.

ПРИЛОЖЕНИЕ А

Дисперсионный анализ результатов эксперимента по исследованию зависимости суточного прироста (мкм/сут) личинок *C. gigas* на стадии велигер от концентрации корма, его качественного состава и плотности посадки личинок.

1. Итоги и среднее значение по строкам, столбцам и латинским буквам латинского квадрата 4×4.

1.1. Итоги по строкам

$A_1 + A_2 + A_3 + A_4 = 34{,}2+36{,}6+29{,}0+23{,}1 = 122{,}9$

1.2. Итоги по столбцам

$B_1 + B_2 + B_3 + B_4 = 29{,}9+36{,}6 + 30{,}6 + 25{,}8 = 122{,}9$

1.3. Итоги по латинским буквам

$C_1 + C_2 + C_3 + C_4 = 40{,}1 + 21{,}0 + 28{,}1 + 33{,}7 = 122{,}9$

1.4. Общий итог

$\hat{G} = A_1 + A_2 + A_3 + A_4 = B_1 + B_2 + B_3 + B_4 = C_1 + C_2 + C_3 + C_4 = \sum^{16} y_{ijk}^2$

$\hat{G}^2 = (122{,}9)^2 = 15104{,}41$

2. Расчет вспомогательной суммы квадратических отклонений:

$A_1 = 123{,}21+32{,}49+100+54{,}76 = 310{,}46$

$A_2 = 56{,}25+94{,}09+ 79{,}21+110{,}25 = 339{,}8$

$A_3 = 25+123{,}21+ 70{,}56+ 20{,}25 = 239{,}02$

$A_4 = 39{,}69 +102{,}01 +10{,}89 +11{,}56 = 164{,}15$

$\boldsymbol{\Sigma} = 1053{,}43$

$B_1 = 123{,}21+56{,}25 +25 +39{,}69 = 244{,}15$

$B_2 = 32{,}49 +94{,}09 +123{,}21 + 102{,}01 = 351{,}8$

$B_3 = 100 +79{,}21 +70{,}56 +10{,}89 = 260{,}66$

$B_4 = 54{,}76 +110{,}25 +20{,}25 +11{,}56 = 196{,}82$

$\boldsymbol{\Sigma} = 1053{,}43$

3. Определение общей суммы квадратов:

$$SS_{общ.} = 1053,43 - \frac{15104,41}{16} = 109,41$$

4. Сумма квадратов отклонений по исследуемым факторам:

4.1 $SS_a = \overset{4}{\boldsymbol{\Sigma}}$; $SS_a = \frac{A^2}{n} - \frac{G^2}{n^2}$

$\frac{A_1^2}{n} = \frac{(34{,}2)^2}{4} = 292{,}41$ $\frac{A_2^2}{n} = \frac{(36{,}6)^2}{4} = 334{,}89$

$\frac{A_3^2}{n} = \frac{(29{,}0)^2}{4} = 210{,}25$ $\frac{A_4^2}{n} = \frac{(23{,}1)^2}{4} = 133{,}40$

$\overset{4}{\boldsymbol{\Sigma}} = \frac{A^2}{4} = 970{,}95$

$SS_a = 970{,}95 - 944{,}02 = 26{,}93$

4.2 $SS_b = \overset{4}{\boldsymbol{\Sigma}}$; $SS_b = \frac{B^2}{n} - \frac{G^2}{n^2}$

$\frac{B_1^2}{n} = \frac{(29{,}9)^2}{4} = 223{,}50$ $\frac{B_2^2}{n} = \frac{(36{,}6)^2}{4} = 334{,}89$

$\frac{B_3^2}{n} = \frac{(30{,}6)^2}{4} = 234{,}09$ $\frac{B_4^2}{n} = \frac{(25{,}8)^2}{4} = 166{,}41$

$\overset{4}{\boldsymbol{\Sigma}} = \frac{B^2}{4} = 958{,}89$

$SS_b = 958{,}89 - 944{,}02 = 14{,}87$

4.3 $SS_c = \overset{4}{\boldsymbol{\Sigma}}$; $SS_c = \frac{C^2}{n} - \frac{G^2}{n^2}$

$\frac{C_1^2}{n} = \frac{(40{,}1)^2}{4} = 402{,}0$ $\frac{C_2^2}{n} = \frac{(21{,}0)^2}{4} = 110{,}25$

$\frac{C_3^2}{n} = \frac{(28{,}1)^2}{4} = 197{,}04$ $\frac{C_4^2}{n} = \frac{(33{,}7)^2}{4} = 283{,}92$

$\overset{4}{\boldsymbol{\Sigma}} = \frac{C^2}{4} = 993{,}21$

$SS_c = 993{,}21 - 944{,}02 = 49{,}19$

5. Определение остаточного члена:

$SS_{ост}$ – остаточная сумма квадратов служит для оценки ошибки эксперимента

$SS_{ост.} = SS_{общ} - (SS_а + SS_в + SS_с)$
$SS_{ост.} = 109{,}41 - (26{,}93 + 14{,}87 + 49{,}19) = 18{,}42$

Таблица 61. Результаты дисперсионного анализа латинского квадрата 4×4

Источник дисперсии	Число степеней свободы	Средний квадрат, Ms	Критерий Фишера Fэксп.	Критерий Фишера Fтабл.	Соответствие
A	3 (n-1)	8,98	2,92	4,8	Fэкс. < Fтабл.
B	3 (n-1)	4,96	1,62	4,8	Fэкс. >Fтабл.
C	3 (n-1)	16,39	5,34	4,8	Fэкс. > Fтабл.
Ошибка	6 (n-1) (n-2)	3,07		$f_1 = 3$ $f_2 = 6$ $\alpha = 0{,}05$	
Итого	15 (n^2-1)				

Примечание: $Ms = \frac{SS}{n-1}$; $Fэксп. = \frac{Ms}{Msocm}$

Таким образом, среднесуточный прирост личинок *Crassostrea gigas* на стадии велигер не зависит от заданной в опыте концентрации корма и плотности посадки личинок, но достоверно зависит от состава корма.

Для проверки различия средних значений для значимых факторов (плотность посадки и состав корма) применяли множественный ранговый критерий Дункана. Проверка по критерию Дункана выполнялась следующим образом:

1) подсчитывалась нормированная ошибка – $S_Н = SA\sqrt{Msocm/4} = \sqrt{3{,}07/4} = 0{,}77$.

2) доверительные интервалы средних, по формуле:

$$\pm t_{\alpha}^{n-1} \cdot S_n[t^{3}_{0,05} = 3.18]; \pm t = 3{,}18 \cdot 0{,}77 = 2{,}44$$

по латинским буквам:

$C_1 = 10{,}02 \pm 0{,}77$ $C_3 = 28{,}1 \pm 0{,}77$
$C_2 = 5{,}25 \pm 0{,}77$ $C_4 = 8{,}42 \pm 0{,}77$

3) из таблицы Дункана значимых рангов для уровня значимости 0,05 и числа степеней свободы, равного числу степеней свободы среднего квадрата ошибки и р =2, 3, 4 выписывали n – 1 значимых рангов:

р	2	3	4
ранги	3,46	3,58	3,64

4) находили группу наименьших значений рангов (НЗР)

р	2	3	4
НЗР	2,66	2,75	2,80

5) упорядочили средние значения по возрастанию по латинским буквам

C_1	C_4	C_3	C_2
10,02	8,42	7,02	5,25

$\overline{C}_1 - \overline{C}_2 = 4{,}77 > 2{,}8$ $\quad$ $\overline{C}_4 - \overline{C}_3 = 1{,}4 < 2{,}66$
$\overline{C}_4 - \overline{C}_2 = 3{,}17 > 2{,}75$ $\quad$ $\overline{C}_2 - \overline{C}_3 = 3{,}0 > 2{,}66$
$\overline{C}_1 - \overline{C}_4 = 1{,}6 < 2{,}75$ $\quad$ $\overline{C}_3 - \overline{C}_2 = 1{,}77 < 2{,}66$

Таким образом, между всеми значениями средних различия статистически значимы, кроме C_2 и C_3 (смеси водорослей B и C).

ПРИЛОЖЕНИЕ В

Дисперсионный анализ результатов эксперимента по исследованию зависимости суточного прироста (мкм/сут.) личинок *C. gigas* на стадии великонхи от плотности их посадки, концентрации корма и его состава.

1. Итоги и среднее значение по строкам, столбцам и латинским буквам латинского квадрата 4×4.

1.1. Итоги по строкам

$A_1 + A_2 + A_3 + A_4 = 86{,}4 + 70{,}6 + 60{,}6 + 55{,}7 = 273{,}3$

1.2. Итоги по столбцам

$B_1 + B_2 + B_3 + B_4 = 62{,}2 + 67{,}3 + 73{,}4 + 70{,}4 = 273{,}3$

1.3. Итоги по латинским буквам

$C_1 + C_2 + C_3 + C_4 = 62{,}1 + 90{,}2 + 55{,}0 + 66{,}0 = 273{,}3$

1.4. Общий итог:

$\hat{G} = A_1 + A_2 + A_3 + A_4 = B_1 + B_2 + B_3 + B_4 = C_1 + C_2 + C_3 + C_4 = \sum^{16} y^2_{ijk}$

$\hat{G}^2 = (273{,}3)^2 = 74692{,}89$

2. Расчет вспомогательной суммы квадратических отклонений:

$A_1 = 384{,}16 + 681{,}21 + 334{,}89 + 501{,}76 = 1902{,}02$

$A_2 = 462{,}25 + 222{,}01 + 349{,}69 + 240{,}25 = 1274{,}2$

$A_3 = 108{,}17 + 201{,}64 + 222{,}01 + 445{,}21 = 977{,}03$

$A_4 = 114{,}19 + 146{,}41 + 462{,}25 + 129{,}96 = 853{,}11$

$\Sigma = 5006{,}36$

$B_1 = 384{,}16 + 462{,}25 + 108{,}1 + 114{,}89 = 1069{,}07$

$B_2 = 681{,}21 + 222{,}01 + 201{,}64 + 146{,}41 = 1251{,}2$

$B_3 = 334{,}89 + 349{,}69 + 222{,}01 + 462{,}25 = 1368{,}84$

$B_4 = 501{,}6 + 240{,}25 + 445{,}21 + 129{,}96 = 1317{,}18$

$\Sigma = 5006{,}36$

3. Определение общей суммы квадратов:

$$SS_{общ.} = 5006,36 - \frac{74692,89}{16} = 338,036$$

4. Сумма квадратов отклонений по исследуемым факторам:

4.1 $SS_a = \overset{4}{\Sigma}$; $SS_a = \frac{A^2}{n} - \frac{G^2}{n^2}$

$$\frac{A_1^2}{n} = \frac{(86,4)^2}{4} = 1866,22 \qquad \frac{A_2^2}{n} = \frac{(70,6)^2}{4} = 1246,09$$

$$\frac{A_3^2}{n} = \frac{(60,6)^2}{4} = 918,09 \qquad \frac{A_4^2}{n} = \frac{(55,77)^2}{4} = 775,62$$

$$\overset{4}{\Sigma} = \frac{A^2}{4} = 4806,04$$

$SS_a = 4806,04 - 4668,3 = 137,74$

4.2 $SS_b = \overset{4}{\Sigma}$; $SS_b = \frac{B^2}{n} - \frac{G^2}{n^2}$

$$\frac{B_1^2}{n} = \frac{(62,2)^2}{4} = 967,21 \qquad \frac{B_2^2}{n} = \frac{(67,3)^2}{4} = 1132,32$$

$$\frac{B_3^2}{n} = \frac{(73,4)^2}{4} = 1346,89 \qquad \frac{B_4^2}{n} = \frac{(70,4)^2}{4} = 1239,04$$

$$\overset{4}{\Sigma} = \frac{B^2}{4} = 4685,46$$

$SS_b = 4685,46 - 4668,3 = 17,16$

4.3 $SS_c = \overset{4}{\Sigma}$; $SS_c = \frac{C^2}{n} - \frac{G^2}{n^2}$

$$\frac{C_1^2}{n} = \frac{(62,1)^2}{4} = 964,1 \qquad \frac{C_2^2}{n} = \frac{(90,2)^2}{4} = 2034,0$$

$$\frac{C_3^2}{n} = \frac{(55,0)^2}{4} = 56,25 \qquad \frac{C_4^2}{n} = \frac{(66,0)^2}{4} = 1089,9$$

$$\overset{4}{\Sigma} = \frac{C^2}{4} = 4843,35$$

$S_c = 4843,35 - 4668,3 = 15,05$

5. Определение остаточного члена:

$SS_{ост}$ – остаточная сумма квадратов служит для оценки ошибки эксперимента

$SS_{ост.} = SS_{общ} – (SS_а + SS_в + SS_с)$

$SS_{ост.} = 338,06 – (137,74 + 17,16 + 175,05) = 8,11$

Таблица 62. Результаты дисперсионного анализа латинского квадрата 4×4

Источник дисперсии	Число степеней свободы	Средний квадрат, Ms	Критерий Фишера Fэксп.	Критерий Фишера Fтабл.	Соответствие
А	3 (n-1)	45,91	34,01	4,8	Fэксп. >Fтабл.
В	3 (n-1)	5,82	4,31	4,8	Fэксп. < Fтабл.
С	3 (n-1)	58,35	43,2	4,8 $f_1 = 3$ $f_2 = 6$ $\alpha = 0.05$	Fэксп. >Fтабл.
Ошибка	6 (n-1) (n-2)	1,35			
Итого	15 (n^2-1)				

Примечание: А – плотность личинок; В – концентрация корма; С – состав корма; $Ms = \frac{SS}{n-1}$; $Fэксп. = \frac{M_s}{M_{s.ост}}$

Следовательно, среднесуточный прирост личинок *Crassostrea gigas* на стадии великонхи не зависит от плотности посадки личинок, но достоверно зависит заданной концентрации корма и его состава.

Для проверки различия средних значений для значимых факторов (плотность посадки и состав корма) применяли множественный ранговый критерий Дункана. Проверка по критерию Дункана выполнялась следующим образом:

1) подсчитывалась нормированная ошибка – $S_n = S_A \sqrt{M_{s ост}/4} = \sqrt{1,35/4} = 0,29$

2) доверительные интервалы средних, по формуле:

$\pm t_{\alpha}^{n-1} \cdot S_n [t^{0,05}_{3} = 3,18]$; $\pm t = 3,18 \cdot 0,29 = 0,92$

а) доверительный интервал по строкам:

$A_1 = 21,6 \pm 0,92$ $A_3 = 15,5 \pm 0,92$

$A_2 = 17,7 \pm 0,92$ $A_4 = 13,9 \pm 0,92$

б) по латинским буквам:

$C_1 = 15,5 \pm 0,92$ $C_3 = 13,8 \pm 0,92$

$C_2 = 22,6 \pm 0,92$ $C_4 = 16,5 \pm 0,92$

3) из таблицы Дункана значимых рангов для уровня значимости 0,05 и числа степеней свободы, равного числу степеней свободы среднего квадрата ошибки и р =2, 3, 4, выписывали n – 1 значимых рангов

р	2	3	4
ранги	3,46	3,58	3,64

находили группу наименьших значений рангов (НЗР)

р	2	3	4
НЗР	1,00	1,04	1,06

упорядочили средние значения по возрастанию

A_1	A_2	A_3	A_4
21,6	17,7	15,5	13,9

4) сравнивали наибольшее и наименьшее значение по строкам

$\overline{A}_1 - \overline{A}_4 = 7{,}7 > 1{,}06$ $\overline{A}_2 - \overline{A}_4 = 3{,}8 > 1{,}04$
$\overline{A}_1 - \overline{A}_3 = 6{,}4 > 1{,}04$ $\overline{A}_2 - \overline{A}_3 = 2{,}5 > 1{,}00$
$\overline{A}_1 - \overline{A}_2 = 3{,}9 > 1{,}00$ $\overline{A}_3 - \overline{A}_4 = 1{,}3 > 1{,}00$

Таким образом, между всеми значениями средних различия статистически достоверны. Среднесуточный прирост личинок *C. gigas* достоверно различается при всех заданных в опыте четырех уровнях плотности посадки.

б) по латинским буквам

C_2	C_4	C_1	C_3
22,6	16,5	15,5	13,8

$\overline{C}_2 - \overline{C}_3 = 8{,}8 > 1{,}06$ $\overline{C}_4 - \overline{C}_3 = 2{,}7 > 1{,}04$
$\overline{C}_2 - \overline{C}_1 = 7{,}1 > 1{,}04$ $\overline{C}_4 - \overline{C}_1 = 1{,}0 > 1{,}00$
$\overline{C}_2 - \overline{C}_4 = 6{,}1 > 1{,}00$ $\overline{C}_1 - \overline{C}_3 = 1{,}7 > 1{,}00$

Между всеми значениями средних различия достоверны, кроме C_1 и C_4 (смеси водорослей A и D).

Предварительный анализ показал, что значения прироста максимальны при минимальной плотности посадки и при корме, состоящем из смеси A (*Isochrysis galbana*+ *Chaetoceros calcitrans*).

Полученные результаты применяли в качестве базовых для проведения регрессионного анализа.

ПРИЛОЖЕНИЕ С

Регрессионный анализ результатов эксперимента по исследованию зависимости суточного прироста (мкм/сут.) личинок *C. gigas* на стадии великонхи от плотности их посадки, концентрации корма и его состава

Для исследования взят экспериментальный план полного факторного эксперимента 2^2 (ПФЭ 2^2).

Факторы: 1) плотность посадки личинок, – X_1

нижний уровень (+ 1) – 3 тыс. лич./ л

верхний уровень (- 1) – 7 тыс. лич./ л;

2) концентрация корма – X_2
нижний уровень (+ 1) – 100 тыс.кл./ мл.
верхний уровень (- 1) – 70 тыс.кл./ мл.
Состав корма: *Isochrysis galbana*+ *Chaetoceros calcitrans* (2: 1)

Таблица 63. План эксперимента ПФЭ 2^2 и результаты исследования влияния плотности посадки (X_1) и концентрации корма (X_2) на прирост личинок *C. Gigas* на стадии велигер

№ опыта	X_1	X_2	$X_1 \cdot X_2$	y	ŷ	S_g	S_g^2
1	+	+	+	26,81; 22,82; 23,03	24,22	2,25	5,0625
2	-	+	-	11,2; 14,14; 15,65	13,66	2,26	5,1076
3	+	-	-	21,07; 21,46; 24,22	22,25	1,72	2,9584
4	-	-	+	16,31; 14,91; 14,74	15,32	0,86	0,7396

$\Sigma S_g^2 = 13{,}8681$

1. Оценка воспроизводимости опыта по критерию Фишера:

$$G_{max} = \frac{S^2_{gmax}}{\Sigma S^2 g} = 0,37 < G_{табл.} = 0,74 \ (\alpha = 0,05)$$

2. Зависимость среднесуточного прироста от плотности посадки личинок и концентрации корма на стадии великонхи описывается уравнением регрессии:

$$\Delta y_{1,2} = b_0 + b_1 x_1 + b_2 x_2 + b_{1,2} \cdot x_1 x_2$$

Расчет коэффициентов уравнения регрессии:

$$b_0 = \frac{24{,}22+13{,}66+22{,}25+15{,}32}{4} = 18{,}8625;$$

$$b_1 = \frac{24{,}22-13{,}66+22{,}25-15{,}32}{4} = 4{,}37;$$

$$b_2 = \frac{24{,}22+13{,}66-22{,}25-15{,}32}{4} = 0{,}0775;$$

$$b_{1,2} = \frac{24{,}22-13{,}66-22{,}25+15{,}32}{4} = 0{,}9075;$$

3. Проверка значимости коэффициентов:

$$S^2 = \frac{13{,}8825}{4} = 3{,}467025; \ S^2\{b_i\} = \frac{3{,}467025}{4} = 0{,}216689;$$

$$S = \sqrt{0{,}216689} = 0{,}4655$$

$$t_0 = \frac{18,8625}{0,4655} = 40,52 > 2,18 \text{ – значим;}$$

$$t_1 = \frac{4,36}{0,4655} = 9,39 > 2,18 \text{ – значим;}$$

$$t_2 = \frac{0,0775}{0,4655} = 0,17 < 2,18 \text{ – не значим;}$$

$$t_{1,2} = \frac{0,9075}{0,4655} = 1,95 < 2,18 \text{ – не значим}$$

4. На основании произведенных расчетов получено уравнение регрессии:

$\Delta y_{1,2} = 18,86 + 4,37\ X_1$

5. Проверка адекватности данного уравнения экспериментальным данным

	разность	квадрат
$\hat{y}_{g1} = 18,86+4,37(+1) = 23,23$	0,99	0,98
$\hat{y}_{g2} = 18,86 + 4,37(-1) = 14,49$	0,83	0,69
$\hat{y}_{g3} = 18,86 + 4,37(+1) = 23,23$	0,98	0,96
$\hat{y}_{g4} = 18,86 + 4,37(-1) = 14,49$	0,83	0,69
		$\Sigma = 3,32$

$$F = \frac{S^2_{ag}}{S^2};\ S^2_{ag} = 2\cdot 3,32 = 6,64;\ S^2 = \frac{13,8681}{4} = 3,47;$$

$$F = \frac{6,64}{3,47} = 1,92;$$

Данное уравнение адекватно описывает экспериментальные данные.

Следовательно, среднесуточный прирост личинок *C. gigas* на стадии великонхи зависит от заданной в опыте плотности их посадки.

Таблица 64. Зависимость выживаемости личинок C. gigas на стадии великонхи от плотности их посадки (X_1) и от концентрации корма (X_2)

№ опыта	X_1	X_2	$X_1 \cdot X_2$	y			$\hat{y}$	S_g	S_g^2
1	+	+	+	88,54	84,31	86,26	86,37	2,12	4,4944
2	-	+	-	84,96	94,30	92,82	90,67	5,02	25,2004
3	+	-	-	68,15	70,0	70,86	69,67	1,38	1,9004
4	-	-	+	70,8	83,16	4,43	76,13	6,35	40,3225

$\Sigma S_g^2 = 71,9217$

1. Оценка воспроизводимости опыта по критерию Фишера:

$$G_{max} = \frac{S^2_{gmax}}{\Sigma S^2 g} = 0,56 < G_{табл.} = 0,74\ (\alpha = 0,05)$$

Опыт воспроизводим.

2. Зависимость выживаемости личинок от плотности их посадки и концентрации корма на стадии велигера описывается уравнением регрессии:

$\Delta y_{1,2} = b_0 + b_1 \cdot X_1 + b_2 \cdot X_2 + b_{1,2} \cdot X_1 \cdot X_2$

Расчет коэффициентов уравнения регрессии:

$b_0 = 80,72$ $\qquad b_2 = 7,815$

$b_1 = -2,695$ $\qquad b_{1,2} = 0,535$

3. Проверка значимости коэффициентов:

$$S^2 = \frac{71,9217}{4} = 17,9804;\ S^2\{b_i\} = \frac{17,9804}{16} = 1,1238;\ \sqrt{1,1238} = 1,06$$

$$t_0 = \frac{80,72}{1,06} = 76,15 > 2,18 \text{ – значим;}$$

$$t_1 = \frac{-2,695}{1,06} = -2,54 < 2,18 \text{ – не значим;}$$

$$t_2 = \frac{7,815}{1,06} = 7,37 > 2,18 \text{ – значим;}$$

$$t_{1,2} = \frac{0,535}{1,06} = 0,505 < 2,18 \text{ – не значим}$$

4. Получено следующее уравнение регрессии:

$\Delta y_{1,2} = 80,72 - 2,70 \cdot X_1 + 7,82 \cdot X_2$

Проверка адекватности данного уравнения экспериментальным данным

	разность	квадрат
$\hat{y}_{g1} = 80,72 - 2,70(+1) + 7,82(+1) = 85,84$	0,53	0,2809
$\hat{y}_{g2} = 80,72 - 2,70(-1) + 7,82(+1) = 92,24$	0,55	0,3025
$\hat{y}_{g3} = 80,72 - 2,70(+1) + 7,82(-1) = 70,2$	0,53	0,2809
$\hat{y}_{g4} = 80,72 - 2,70(-1) + 7,82(-1) = 75,6$	0,53	0,2809
		$\Sigma = 1,1452$

$$F = \frac{S^2_{ag}}{S^2};\ S^2_{ag} = 2 \cdot 1,1452 = 2,2904;\ S^2 = \frac{71,9217}{4} = 17,98;$$

$$F = \frac{2,2904}{17,98} = 0,13;$$

Данное уравнение адекватно описывает экспериментальные данные.

Таким образом, выживаемость личинок на стадии великонхи зависит как от плотности их посадки, так и от концентрации корма.

3.4. Греко-латинские квадраты

Два латинских квадрата называются ортогональными, если при наложении одного квадрата на другой <u>каждая пара</u> букв встречается один и только один раз.

Комбинация двух ортогональных квадратов называется латинским квадратом 2-го порядка.

Если элементы первого квадрата обозначить латинскими, а другого – греческими буквами и совместить их – в этом случае новый квадрат будет называться греко-латинским квадратом. Здесь каждая латинская буква встречается только один раз с каждой греческой буквой:

A	B	C	α	β	γ	Aα	Bβ	Cγ
B	C	A	γ	α	β	Bγ	Cα	Aβ
C	A	B	β	γ	α	Cβ	Aγ	Bα

Элементы греко-латинского квадрата образуют двойную элиминирующую группировку. Число ортогональных квадратов равно n – 1. Например, для n = 4 имеем n – 1 = 3:

1	2	3	4	1	2	3	4	1	2	3	4
2	1	4	3	3	4	1	2	4	3	2	1
3	4	1	2	4	3	2	1	2	1	4	3
4	3	2	1	2	1	4	3	3	4	1	2

Однако для **n = 6** ортогональных квадратов не существует.

Пример греко-латинского квадрата 3×3 в качестве плана эксперимента:

A/B	b_1	b_2	b_3
a_1	Aα	Bβ	Cγ
a_2	Bγ	Cα	Aβ
a_3	Cβ	Aγ	Bα

В этом эксперименте 4 фактора заданы на трех уровнях, это – ПФЭ 3^4, число опытов = 81.

Но в данном случае ставятся только 9 опытов, то есть греко-латинский квадрат – это $1/ n^2$ от ПФЭ.

Применение: 1) Сокращение числа перебора вариантов; 2) Оптимизация условий (при работе с качественными факторами на решётке); 3) Отсев по уровням качественных факторов.

Гипер-греко-латинские квадраты (ГГЛК)

Совмещение трех ортогональных квадратов образует ГГЛК.

A/B	b_1	b_2	b_3	b_4
a_1	Aα1	Bβ2	Cγ3	Dδ4
a_2	Bγ4	Aδ3	Dα2	Cβ1
a_3	Cδ2	Dγ1	Aβ4	Bα3
a_4	Dβ3	Cα4	Bδ1	Aγ2

Всего 16 опытов для пяти факторов вместо $4^5 = 1024$, т.е. $1/n^3$ от ПФЭ 4^5.

Планы столь большой дробности применяются на первых этапах исследования. При этом выделяются те комбинации, которые затем будут детально исследованы.

Таблица 65. Сводная таблица дисперсионного анализа ГГЛК 4×4

Источник дисперсии	Число степ свободы	SS	MS	$F_{эксп}$	$F_{табл}$	$\alpha_i = 0$ $\beta j = 0$ $\gamma_k = 0$ $\delta_l = 0$
A	n – 1	SS_a	MS_a	$MS_a / MS_{ост}$	Один и тот же для всех строк	$\alpha_i = 0?$
B	n – 1	SS_b	MS_b	$MS_b / MS_{ост}$		$\beta j = 0?$
C	n – 1	SS_c	MS_c	$MS_c / MS_{ост}$		$\gamma_k = 0?$
D	n – 1	SS_d	MS_d	$MS_d / MS_{ост}$		$\delta_l = 0?$
E	n – 1	SS_e	MS_e	$MS_e / MS_{ост}$		$\xi_e = 0?_{ee}$
Остаток	(n – 1)(n – 4)	$SS_{ост}$	$f_1=n-1$ $f_2=(n-1)(n-4)$			
Сумма	$n^2 – 1$					

Преимущества и недостатки латинских и греко-латинских квадратов

Преимущества:

Исследования качественных факторов на многих уровнях.

Эффективный способ планирования для исключения влияния источников неоднородностей, а также для сокращения числа перебора вариантов.

Статистический анализ проводится просто. Данные хорошо интерпретируются благодаря ортогональности планов и независимости линейных эффектов друг от друга.

Недостатки:

Возможно исследование только линейных моделей.

Все факторы должны варьироваться на одинаковом количестве уровней.

3.5. Латинские кубы

Латинским кубом 1-го порядка размера n называют кубическую таблицу, составленную из n элементов, расположенных в n^3 позициях, такую, что каждый элемент входит в таблицу n^2 раз и встречается в каждой из трёх плоскостей параллельных координатным плоскостям одинаковое для всех элементов число раз равное n (табл. 66).

Таблица 66. Латинский куб 3×3×3:

d_1				d_2				d_3			
	b_1	b_2	b_3		b_1	b_2	b_3		b_1	b_2	b_3
a_1	A	B	C	a_1	B	C	A	a_1	C	A	B
a_2	B	C	A	a_2	C	A	B	a_2	A	B	C
a_3	C	A	B	a_3	A	B	C	a_3	B	C	A

Видно, что каждый элемент встречается в трёх горизонтальных плоскостях, в трёх профильных, в трёх фронтальных. В каждой плоскости встречается по 3 раза.

Латинский куб n×n×n состоит из сбалансированного ряда латинских квадратов, число которых равно n. Изучается влияние четырех факторов на n уровнях. В отличие от греко-латинского квадрата, когда также изучается влияние четырех факторов, в данном случае три фактора (по осям) считаются главными (A, B, C), а один фактор (D) составляет элиминирующую группировку. В греко-латинском квадрате главные факторы A и B, а элементы квадрата (латинская и греческая буквы) образуют двойную элиминирующую группировку.

С помощью латинского куба исследуется нелинейная модель (если есть повторные опыты) $N = (ПФЭ\ n^4) / n$.

Количество опытов в греко-латинских квадратах: $N = (ПФЭn^4)/n^2$ (линейная модель). Возникает вопрос: что же выбрать – латинский куб или греко-латинский квадрат? Ответ зависит от возможностей экспериментатора, а именно: возможно ли ограничиться линейной моделью (или нужно оценить и взаимодействия), а также от количества опытов, для которых имеется возможность постановки.

Если совместить два ортогональных латинских куба, тогда получится латинский куб 2-го порядка или греко-латинский куб 1-го порядка. Если совместить три и более куба, в этом случае получится гипер-греко-латинский куб. В кубах 1-го порядка все факторы устанавливаются на n уровнях. В кубах 2-го порядка один фактор на n^2 уровнях, а остальные – на n уровнях. Эффект фактора на n^2 уровнях определяется с точностью в n раз меньшей, чем остальные факторы.

Статистический анализ латинского куба 1-го порядка 4×4×4 (размер n=4, r=1).

Таблица 67. Латинский куб 1-го порядка 4×4×4

		b_1	b_2	b_3	b_4			b_1	b_2	b_3	b_4
d_1	a_1	C_1	C_2	C_3	C_4	d_2	a_1	C_2	C_3	C_4	C_1
	a_2	C_3	C_3	C_4	C_1		a_2	C_3	C_4	C_1	C_2
	a_3	C_3	C_4	C_1	C_2		a_3	C_4	C_1	C_2	C_3
	a_4	C_4	C_1	C_2	C_3		a_4	C_1	C_2	C_3	C_4

		b_1	b_2	b_3	b_4			b_1	b_2	b_3	b_4
d_3	a_1	C_3	C_4	C_1	C_2	d_4	a_1	C_4	C_1	C_2	C_3
	a_2	C_4	C_1	C_2	C_3		a_2	C_1	C_2	C_3	C_4
	a_3	C_1	C_2	C_3	C_4		a_3	C_2	C_3	C_4	C_1
	a_4	C_2	C_3	C_4	C_1		a_4	C_3	C_4	C_1	C_2

	Уровни				
Факторы	A	a_1	a_2	a_3	a_4
	B	b_1	b_2	b_3	b_4
	C	c_1	c_2	c_3	c_4
	D	d_1	d_2	d_3	d_4

Эксперимент без повторных опытов

Линейная модель: $y_{ijko} = \mu + \alpha_i + \beta_j + \gamma_k + \delta_o + \varepsilon_{ijko}$.

1. Подсчитываем итоги и средние значения. Далее рассчитываются вспомогательные суммы квадратов.
2. SS_1 сумма квадратов результатов всех наблюдений: $SS_1 = \Sigma\Sigma\Sigma\,(y_{ijko})^2$.
3. SS_2 – сумма квадратов итогов по абсциссе, деленная на n^2: $SS_2 = (\Sigma A^2_i)/\ n^2$.
4. SS_3 – сумма квадратов итогов по ординате, деленная на n^2: $SS_3 = (\Sigma B^2_j)/\ n^2$.
5. SS_4 - сумма квадратов итогов по аппликате, деленная на n^2: $SS_4 = (\Sigma C^2_k)/\ n^2$.
6. SS_5 – сумма квадратов итогов по латинской букве, деленная на n^2: $SS_5 = (\Sigma D^2_o)/\ n^2$.

7. SS_6 – корректирующий член, равный квадрату общего итога, деленному на n^3, то есть на число опытов $N= n^3$:
   $SS_6 = G^2/ n^3$.
8. $SS_{общ}$ – общая сумма квадратов, которая равна разности между суммой квадратов всех наблюдений и корректирующим членом:
   $SS_{общ} = SS_1 – SS_6$.
9. $SS_{ост}$ – остаточная сумма квадратов служит для оценки ошибки эксперимента:
   $SS_{ост} = SS_{общ} – (SS_A + SS_B + SS_C + SS_D)$.

Таблица 68. Сводная таблица дисперсионного анализа латинского куба 1-го порядка (без повторных опытов).

Источник дисперсии	Число степ. свободы	Сумма квадратов SS	Средний кв. MS	$F_{эксп}$	Гипотезы
A (абсцисса)	$n – 1$	$SS_A=SS_2 – SS_6$	$SS_A/(n – 1)$	$MS_A/ MS_{ост}$	$\alpha_i = 0$?
B (ордината)	$n – 1$	$SS_B=SS_3 – SS_6$	$SS_B/(n – 1)$	$MS_B/ MS_{ост}$	$\beta j = 0$?
C (аппликата)	$n – 1$	$SS_C=SS_4 – SS_6$	$SS_C/(n – 1)$	$MS_C/ MS_{ост}$	$\gamma_k = 0$?
D (латинская буква)	$n – 1$	$SS_D=SS_5 – SS_6$	$SS_D/(n – 1)$	$MS_D/ MS_{ост}$	$\delta_l=0$?
Остаток (ошибка)	$n^3 – 4n + 3$	$SS_{ост} = SS_{общ} – (SS_a + SS_B + SS_C + SS_d)$.	$SS_{ост}/(n^3 – 4n + 3)$		
Сумма	$n^3 – 1$	$SS_{общ} = SS_1 – SS_6$			

Если гипотеза H_0 отвергается (то есть между средними имеется различие), то на втором этапе статистического исследования средние сравниваются по ранговому критерию Дункана.

Эксперимент с повторными опытами позволяет проверить гипотезу об отсутствии эффектов взаимодействия.

3.6. Сложные (комбинированные) планы

Это планы, построенные на базе латинских квадратов и прямоугольников.

Рассмотрим следующую задачу:

$Y = F(x_1, x_2, ..., x_n; z_1, z_2, ... z_m; \varepsilon_{l...m})$, где:

x_i – количественные факторы $i=1,2,...n$;

z_j – качественные факторы $j= 1,2,...m$.

Нужно построить планы, которые позволили бы:

1. увеличить число количественных факторов $n>2$;
2. варьировать количественные факторы только на двух уровнях, (что достаточно, если справедлива гипотеза об отсутствии взаимодействий);
3. исключить нарушающее влияние качественных факторов при определении линейных эффектов количественных факторов;
4. совершить, в случае необходимости, движение по градиенту для количественных факторов;
5. построить оптимальный перебор комбинаций уровней качественных факторов, если $m>1$;
6. не превысить число опытов по сравнению с ПФЭ.

Для того, чтобы выполнить все перечисленные условия, нужно синтезировать сложные планы из планов 2^n и латинских квадратов n x n.

Латинские квадраты, совмещённые с планами типа 2^n, где $n = 2k$ ($k=2, 3, ...$).

Пусть имеется четыре количественных фактора и один качественный. Совместим ПФЭ2^4 и латинский квадрат 4×4.

Количественные факторы: x_1, x_2, x_3, x_4 с уровнями -1 и +1.

Качественный фактор: x_5 с уровнями A, B, C, D.

Тогда план будет иметь вид:

		$x_1(+1)$		$x_1(-1)$	
		$x_2(+1)$	$x_2(-1)$	$x_2(+1)$	$x_2(-1)$
$x_4(+1)$	$x_3(+1)$	A	B	C	D
	$x_3(-1)$	C	D	A	B
$x_3(-1)$	$x_3(+1)$	B	A	D	C
	$x_3(-1)$	D	C	B	A

Перепишем этот план в другой форме (табл.69), для того, чтобы определить, какие взаимодействия количественных факторов оказываются смешанными с качественными факторами (соответствующие столбцы не ортогональны, т.е. произведения столбцов не равны 0).

Из таблицы 69 следует, что с фактором x_5 смешаны следующие взаимодействия:

x_1x_3; x_2x_4; $x_1x_2x_3x_4$. Эти взаимодействия невозможно оценить, поэтому следует считать их потерянными.

Если на ПФЭ наложен греко-латинский план, то мы можем ввести ещё один качественный фактор. При этом число потерянных взаимодействий возрастает, так как другой качественный фактор смешивается с другими взаимодействиями.

Таблица 69. Проверка смешиваемости качественного фактора с эффектами количественных факторов, и с эффектами взаимодействий количественных факторов

№	x_1	x_2	x_3	x_4	x_1x_2	x_1x_3	x_1x_4	x_2x_4	x_2x_3	x_3x_4	x_{123}	x_{124}	x_{134}	x_{234}	x_{1234}	x_5
1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	A
2	+	-	+	+	-	+	+	-	-	+	-	-	+	-	-	B
3	-	+	+	+	-	-	-	+	+	+	-	-	-	+	-	C
4	-	-	+	+	+	-	-	-	-	+	+	+	-	-	+	D
5	+	+	-	+	+	-	+	+	-	-	-	+	-	-	-	C
6	+	-	-	+	-	-	+	-	+	-	+	-	-	+	+	D
7	-	+	-	+	-	+	-	+	-	-	+	-	-	-	+	A
8	-	-	-	+	+	+	-	-	+	-	+	+	+	+	-	B
9	+	+	+	-	+	+	-	-	+	-	+	-	-	-	-	B
10	+	-	+	-	-	+	-	+	-	-	-	+	-	+	+	A
11	-	+	+	-	-	-	+	-	+	-	-	+	+	-	+	D
12	-	-	+	-	+	-	+	+	-	-	+	-	+	+	-	C
13	+	+	-	-	+	-	-	-	-	+	-	-	+	+	+	D
14	+	-	-	-	-	-	-	+	+	+	+	+	+	-	-	C
15	-	+	-	-	-	+	+	-	+	+	+	+	-	+	-	B
16	-	-	-	-	+	+	+	+	+	+	-	-	-	-	+	A

В общем случае в ПФЭ 2^n имеется ($2^n - n - 1$) взаимодействий. При совмещении с латинским квадратом смешивается ($n - 1$) взаимодействие, а с m квадратами: m ($n - 1$) взаимодействие. В данном случае при наложении гипер-греко-латинского квадрата теряется $3\ (4 - 1) = 9$ взаимодействий. Всего же имеется $2^4 - 4 - 1 = 11$ взаимодействий. Поэтому независимо оценить мы сможем только два взаимодействия и линейные эффекты.

Нужно помнить, что вместо несмешанных взаимодействий можно вводить дополнительные факторы, т.е. переходить к ДФЭ, оценивать линейные эффекты и совершать **крутое восхождение** по количественным факторам.

Так, с латинским квадратом 4×4 можно совместить ДФЭ$^{12-8}$. А ДФЭ$^{56-50}$ можно совместить с латинским квадратом 8×8. Статистическая обработка обычная: регрессионный анализ и дисперсионный анализ.

3.6.1. Пример: зависимость среднесуточного прироста спата мидий (мкм/сут.) от концентрации корма, плотности посадки особей и состава корма

Количественные факторы:

А – концентрация корма (50, 10, 150, 200 тыс. кл./мл);

В – плотность посадки спата (0,5; 1; 2; 3; тыс. экз./л).

Качественный фактор: латинские буквы – состав корма (A; B; C; D).

Таблица 70. План и результаты эксперимента (Пиркова, Ладыгина, 2004).

a\b	0,5	1	2	3	Итог
50	A 21,0	B 36,1	C 47,7	D 32,9	a_1
100	B 38,3	C 43,7	D 31,5	A 30,1	a_2
150	C 58,9	D 38,5	A 61,6	B 50,6	a_3
200	D 73,9	A 65,4	B 63,6	C 64,1	a_4
Итог	b_1	b_2	b_3	b_4	

Статистическая обработка результатов эксперимента

1) Подсчитываем итоги и средние значения по строкам, столбцам и латинским буквам.

a_1= 21,0+36,1+47,7+32,9=137,7 среднее: 34,4
a_2= 38,3+43,7+31,5+30,1=143,6 35,9
a_3= 58,9+38,5+61,6+50,6=209,6 52,4
a_4= 73,9+65,4+63,6+64,1=267,0 66,8
Σ=757,9

b_1=21,0+38,3+58,9+73,9=192,1 среднее: 48,0
b_2=36,1+43,7+38,5+65,4=183,7 45,9
b_3=47,7+31,5+61,6+63,6=204,4 51,1
b_4=32,9+30,1+50,6+64,1=177,7 44,4
Σ=757,9

c_1= 21,0+30,1+61,6+65,4=178,1 среднее: 44,5
c_2= 36,1+38,3+50,6+63,6=188,6 47,2
c_3= 47,7+43,7+58,9+64,1=214,4 53,6
c_4=32,9+31,5+38,5+73,9=176,8 44,2
Σ=757,9

Общий итог G=757,9; G^2 =757,9^2=574412,41.

2) Вспомогательная сумма квадратов отклонений:

$SS_{общ}=\Sigma\Sigma\, y^2_{ijk} - G^2/n^2$;

$SS_{общ}$=39575,27 – 574412,41/16 = 39575,27 –3590,76 = 3674,49.

Сумма квадратов итогов по строкам $SS_a = \Sigma(a^2_i/n) - G^2/n^2$.

SS_a = (137,7^2)/4 + (143,6^2)/4 + (209,6^2)/4 + (267^2)/4 – 3590,76 = 4740,32 + 5155,24 + 10983,04 + 17822,25 – 3590,76 = 3870,83 – 3590,78 = 280,05.

Сумма квадратов итогов по столбцам $SS_b = \Sigma(b^2_j/n) - G^2/n^2$;

SS_b =36001,16 – 3590,76= 10,38.

Сумма квадратов итогов по латинским буквам $SS_c = \Sigma(c^2_j/n) - G^2/n^2$;
SS_c =36128,79-3590,76 = 228,01.

Остаточный член: $SS_{ост} = SS_{общ} - (SS_a + SS_b + SS_c)$;
$SS_{ост}$ = 3674,49 – (280,05+10,38+228,01)=546,05.

Таблица 71. Сводная таблица дисперсионного анализа латинского квадрата 4×4

Источник дисперсии	Число Степеней свободы	Сумма квадратов	Средний квадрат MS	Критерий Фишера (эксперим)	Критерий Фишера (табличный)	$\alpha_i = 0$ $\beta j = 0$ $\gamma_k = 0$
a	n-1=3	2800	$SS_a/(n-1)$ 2800/3= 933	$MS_a/MS_{ост}$ 933/91=10,3	F=4,8 $f_1=3$; $f_2=6$	10,3>4,8 $\alpha_i \neq 0$
b	n-1=3	100	$SS_b/(n-1)$ 100/3=33,5	$MS_b/MS_{ост}$ 33,5/91=0,36	F=4,8	0,36<4,8 $\beta j = 0$
c	n-1=3	288	$SS_c/(n-1)$ 228/3=76	$MS_c/MS_{ост}$ 76/91=0,84	F=4,8	0,84<4,8 $\gamma_k = 0$
Остаток (ошибка)	(n – 1)(n – 2) = 6	546	$SS_{ост}/(n-1)(n-2)$ 546/6=91			
Сумма	n^2- 1=15		$SS_{общ}$=3674			

Таким образом, рост спата мидий зависит от концентрации кормовых водорослей и не зависит от качественного состава исследованного корма и плотности посадки спата в эксперименте.

Сравнение средних значений, полученных при разных концентрациях корма, с помощью рангового критерия Дункана

1) Нормированная ошибка:
$S_н = \sqrt{(MS_{ошиб}/N)} = \sqrt{(91/16)} = 2,38$.
2) Из таблицы для рангового критерия Дункана выписываем n – 1 = 3 рангов (число степеней свободы равно числу среднего квадрата ошибки =6):

3,46 3,58 3,64

3) Наименьшие значимые ранги (ранги умножить на $S_н$ =2,38):
НЗР 8,23 8,52 8,66
4) Упорядочивание средних значений:

a_1	a_2	a_3	a_4
34,4	35,9	52,4	66,8

5) Сравнение средних значений
Наибольшее против наименьшего ($a_4 - a_1$) 66,8-34,4=32,4>8,66 – различие значимо
$a_4 - a_2$= 66,8 – 35,9=30,9>8,52 – различие значимо;
$a_4 - a_3$= 66,8 – 52,4= 14,4>8,23 – различие значимо;

$a_3 - a_1 = 52{,}4 - 34{,}4 = 18{,}0 > 8{,}66$ – различие значимо;
$a_3 - a_2 = 52{,}4 - 35{,}9 = 16{,}5 > 8{,}52$ – различие значимо;
$a_2 - a_1 = 35{,}9 - 34{,}4 = 1{,}5 < 8{,}66$ – различие не значимо.

Выводы:
1. Скорость роста спата мидий не зависит от качественного состава использованного в эксперименте корма и не зависит от изменения плотности посадки спата в пределах от 0,5 до 3 тысяч экз. на литр.
2. Скорость роста зависит от концентрации корма при её варьировании от 50 до 200 тыс. клеток на мл. Максимальные значения скорости роста получены при 200 тыс. кл. на мл.
3. Рекомендуется выращивать спат при плотности посадки 3000 экз./л; при использовании наиболее дешёвого корма в концентрации 200 тыс. кл/мл.

3.6.2. Пример применения сложного плана

Исследование потока углерода через организмы гарпактицид.

Радиоуглеродным методом измерены параметры потока углерода растворённого и твёрдого органического вещества (скорость переноса углерода, размеры обменных фондов углерода в теле, время оборота углерода в обменных фондах, интенсивность потребления углерода) через тело тигриопусов, обитающих в супралиторальных ваннах Кольского побережья Баренцева моря (Холодов, 1981). С использованием одного из методов математического планирования экспериментов – сложного плана, полученного совмещением латинского прямоугольника 2×4 и плана ПФЭ2^3, установлено, что интенсивность потребления углерода зависит от природы источника углерода и в значительно меньшей мере от количества доступного немеченого корма (многоклеточной водоросли энтероморфы). Изменение температуры в пределах 7 – 10°С и солёности от 25 до 34‰ не оказывают влияния на интенсивность потребления углерода.

Было установлено, что интенсивность выведения углерода из тела тигриопусов зависит от типа субстрата и от количества немеченой (нерадиоактивной) энтероморфы, помещённой в экспериментальный сосуд с животными, выделяющими ^{14}C. Интересно оценить эффекты влияния этих факторов (температуры, солёности, типа источника углерода, количества немеченой энтероморфы в сосуде) на интенсивность выведения углерода мечеными тигриопусами. Так как в число перечисленных факторов входит качественный – тип меченого субстрата, то для решения данной задачи применили сложный план, представляющий собой комбинацию латинского прямоугольника 2×4 и ПФЭ2^3 (таблица 72). Такой план позволяет исследовать влияние одного качественного фактора на четырёх уровнях и трёх количественных факторов на двух уровнях.

Таблица 72. Сложный план для одного качественного и трёх количественных факторов, варьируемых на 2 уровнях*

Фактор	x_1(+1)		x_1(-1)	
	x_2(+1)	x_2(-1)	x_2(+1)	x_2(-1)
x_3(-1)	1 A	2 B	3 C	4 D
x_3(+1)	5 D	6 C	7 B	8 A

* Здесь и в таблицах 73 и 74: x_1 – количество энтероморфы в опытном сосуде; x_2 – солёность; x_3 – температура воды; латинские буквы – тип кормового субстрата, которым метились животные. Цифры указывают номер экспериментального сосуда.

Латинский прямоугольник 2×4 представляет собой половину латинского квадрата 4×4. Расположение элементов (букв) в данном прямоугольнике таково, что с каждым уровнем количественного фактора сочетаются все уровни качественного фактора. Равным образом с каждым уровнем качественного фактора сочетаются все уровни всех количественных факторов. Поэтому, каким бы ни было влияние источников неоднородностей, оно в равной мере скажется при подсчёте итогов по любому фактору. Опыт выполнен в трёх повторностях. Факторы задавались на следующих уровнях: x_1- количество (масса) немеченой энтероморфы в опытном сосуде 10–100 мг сухой массы; x_2 – солёность воды 25–34‰; x_3 – температура воды: 7–10°С; x_4 – тип субстрата, которым метились животные: А – гидролизат водорослей; В – мочевина; С – энтероморфа; D – глицин.

Таблица 73. Расчётная матрица и результаты эксперимента, поставленного по сложному плану, приведенному в таблице 72.

№	x_0	x_1	x_2	x_3	x_1x_2	x_1x_3	x_2x_3	$x_1x_2x_3$	x_4	y
1	+1	+1	+1	-1	+1	-1	-1	-1	A	0,363
2	+1	+1	-1	-1	-1	-1	+1	+1	B	0,573
3	+1	-1	+1	-1	-1	+1	-1	+1	C	0,531
4	+1	-1	-1	-1	+1	+1	+1	-1	D	0,072
5	+1	+1	+1	+1	+1	+1	+1	+1	D	0,079
6	+1	+1	-1	+1	-1	+1	-1	-1	C	0,573
7	+1	-1	+1	+1	-1	-1	+1	-1	B	0,590
8	+1	-1	-1	+1	+1	-1	-1	+1	A	0,259

Каждая из восьми ячеек прямоугольника (табл. 72) представляет собой экспериментальный сосуд объёмом 1 л. Таким образом, например, вторая ячейка нижней строки прямоугольника соответствует опытному сосуду №6, в который помещены тигриопусы, предварительно помеченные энтероморфой. В этом сосуде находится 100 мг сухой энтероморфы; солёность воды 25‰, температура 10°С. Экспериментальные результаты (средние арифметические) вместе с планом эксперимента ПФЭ2^3 и с вектор-столбцами для расчёта уравнения регрессии приведены в таблице 73.

При проверке условий смешиваемости эффектов количественных факторов с качественным фактором, выяснилось, что с качественным фактором смешаны все парные взаимодействия количественных факторов. Свободными являются линейные эффекты и эффект тройного взаимодействия.

Уравнение регрессии, рассчитанное по данным таблицы 73:

$$У=0{,}380+0{,}017x_1+0{,}011x_2-0{,}005x_3-0{,}187x_1x_2-0{,}062\,x_1x_3-0{,}052x_2x_3-0{,}019\,x_1x_2x_3$$

Проверка значимости коэффициентов регрессии выявила два статистически незначимых коэффициента: $0{,}011x_2$ и $0{,}005x_3$, поэтому окончательно уравнение имеет вид:

$$У=0{,}380+0{,}017x_1-0{,}187x_1x_2-0{,}062\,x_1x_3-0{,}052x_2x_3-0{,}019\,x_1x_2x_3$$

Из данного уравнения следует, что количество вносимой в сосуд энтероморфы (x_1) ведёт к увеличению интенсивности выведения радиоуглерода из тканей животных, причём линейный эффект составляет 9 % среднего значения (свободного члена). Изменения температуры и солёности воды в выбранных интервалах не оказывают влияния на интенсивность выведения. Однако парные взаимодействия всех количественных факторов довольно значительны, что объясняется влиянием качественного фактора, эффект которого смешан с эффектами всех парных взаимодействий. Отрицательное тройное взаимодействие свидетельствует о том, что все три количественных фактора, взятые на верхних уровнях, приводят к снижению интенсивности выведения метки (радиоуглерода), а на нижних – напротив, увеличивают интенсивность выведения метки из тела рачков.

Все эффекты количественных факторов слишком малы в сравнении с эффектом качественного фактора, роль которого выявляется дисперсионным анализом (таблицы 74 и 75).

Таблица 74. Результаты исследований, выполненных по сложному плану, с целью изучения влияния четырёх факторов на интенсивность выведения тигриопусами радиоуглерода.

Фактор	x_1 (+1)		x_1 (-1)	
	x_2 (+1)	x_2 (-1)	x_2 (+1)	x_2 (-1)
x_3 (-1)	0,346; 0,379; 0,363	0,525; 0,590; 0,604	0,517; 0,545; 0,531	0,084; 0,064; 0,068
x_3 (+1)	0,082; 0,082; 0,073	0,516; 0,592; 0,611	0,597; 0,579; 0,594	0,248; 0,269; 0,261

Таблица 75. Сводные данные результатов дисперсионного анализа сложного плана, полученного совмещением плана ПФЭ2^3 с 2×4 латинским прямоугольником

Источник дисперсии	Число степеней свободы	Средний квадрат	Критерий Фишера
Количество энтероморфы в сосуде, x_1	2-1	0,00691	0,239
Солёность, x_2	2-1	0,00276	0,095
Температура воды, x_3	2-1	0,00053	0,018
Взаимодействие x_1x_2	(2-1)·(2-1)	0,83750	28,940
Взаимодействие x_1x_3	(2-1)·(2-1)	0,10524	3,636
Тип кормового субстрата	2-1	1,00632	34,770
Ошибка внутри ячейки	8(k*-1)	0,02800	-

*k – количество повторов

В последнем столбце таблицы 75 величина критерия Фишера отражает степень влияния каждого фактора, а также двух взаимодействий факторов. В данном случае значения критериев Фишера, рассчитанные как отношения средних квадратов к среднему квадрату внутри ячейки (ошибке воспроизводимости), использованы для проверки гипотезы о значимости всех линейных эффектов и двух парных взаимодействий. Критическое значение критерия Фишера при 95%-ном уровне значимости доверительной вероятности и при степенях свободы f_1=1 и f_2 =16 равно 4,5. Таким образом, значимыми оказываются эффекты типа кормового субстрата и парного взаимодействия первых двух факторов.

Данное исследование показало, что регрессионный анализ – более чувствительный метод, позволяющий оценивать сравнительно слабые эффекты. Вместе с тем отчётливо проявляется влияние качественного фактора, или источника радиоуглерода, на интенсивность выведения углерода из тканей тела. Эффект этого фактора превосходит эффект фактора x_1, то есть наиболее мощного количественного фактора, в 145 раз.

Глава 4
ПАССИВНЫЙ ЭКСПЕРИМЕНТ

Выполнение полевых исследований во время экспедиций – важнейший способ получения ценных научных фактов, дополняющий экспериментальные исследования. Оба направления не входят во взаимное противоречие, напротив – способствуют развитию друг друга, выдвигая новые гипотезы и стимулируя постановку новых задач и проведение очередных, более глубоких исследований.

Структурную схему пассивного эксперимента легко можно получить, преобразуя схему активного эксперимента (рис. 1), для чего нужно из схемы убрать **X** – регулируемые (управляемые) факторы. Однако частично некоторыми факторами можно управлять и в пассивном эксперименте. Например, можно задавать глубину размещения подопытных гидробионтов, глубину и ориентацию экспериментальных субстратов, плотность размещения животных в экспериментальном садке, а также их возраст, индивидуальный вес и т.д. Но, в основном, наблюдатель имеет дело с неуправляемыми факторами: он только измеряет их значения и записывает в журнал. В тот же массив данных заносятся и выходные данные объекта исследования (зависимая переменная). Такими характеристиками могут быть: первичная продукция, биомасса, численность, потребление кислорода, количество осетров, попавших в сеть в качестве прилова, и т.д. В качестве примеров из других отраслей можно привести: спрос на определённый товар, объём реализованной продукции, процент безработных в районе, процент углерода в выплавляемом чугуне в доменной печи и т.д. При планировании сбора материала для повторов необходимо избегать получения «мнимых повторностей» (Козлов, 2003, 2014).

Несмотря на то, что в полевых исследованиях управлять факторами затруднительно, а чаще – невозможно, тем не менее, сбор материала должен выполняться по определённому плану. Например, при изучении суточной динамики первичной продукции в бухтах на разных горизонтах (глубинах). В таком эксперименте можно задавать глубину размещения экспериментальных сосудов и продолжительность экспозиции этих сосудов в воде. Для планирования данного пассивного эксперимента целесообразно использовать латинский квадрат, как, например, в таблице 55. В таком плане латинские буквы будут обозначать различные бухты (элиминирующая группировка); по горизонтали задаются различные уровни фактора «глубина», а по вертикали – продолжительность экспозиции экспериментального сосуда.

В результате эксперимента будут получены: 1) суточная динамика первичной продукции на разных горизонтах независимо от места проведения эксперимента; 2) зависимость первичной продукции от глубины, независимо от места проведения эксперимента. То есть, будут получены универсальные зависимости, которые можно использовать для общих подходов при изучении первичной продукции в прибрежной зоне моря. Однако при более детальном исследовании первичной продукции в конкретной бухте придётся использовать другие планы.

В настоящее время данные практически всегда обрабатываются на компьютере. Однако, ввиду широкого применения регрессионного анализа в научных исследованиях, в проведении активных и пассивных экспериментов, в данном разделе рассмотрена ручная обработка данных наблюдений, что позволит читателю детально ознакомиться с вычислительной процедурой регрессионного анализа.

На стр. 33-36 изложен регрессионный анализ с использованием метода наименьших квадратов (МНК) для одного фактора. Но на практике приходится работать с многофакторными объектами. В этом случае для расчёта коэффициентов регрессии потребуется воспользоваться аппаратом матричной алгебры.

Регрессионный анализ с применением аппарата матричной алгебры

Уравнение $y_i = b_0 + b_1 x_{1i}$ в матричной форме имеет вид:

$$Y = XB \tag{1}$$

Система уравнений для МНК запишется в виде:

$$X^T XB = X^T Y \tag{2}$$

Эта система получилась после умножения слева обеих частей уравнения (1) на X^T.

Вектор-столбец коэффициентов уравнения рассчитываем по формуле:

$$B = (X^T X)^{-1} X^T Y \tag{3}$$

Ниже решение уравнения (3) подробно рассматривается на конкретном примере.

4.1. Постановка задачи применения регрессионного анализа

Определение методом пассивного эксперимента зависимости концентрации клеток фитопланктона от суммарной концентрации биогенных солей в морской воде (x_1) и освещённости (x_2) в слое 0 – 30 метров.

Рассматриваемый в данном разделе пример взят не из Природы, а придуман автором с целью демонстрации вычислительной процедуры регрессионного анализа. (Этот пример не обязателен для изучения, особенно для читателей, не владеющих аппаратом матричной алгебры.)

Известно, что x_1 и x_2 – основные факторы, определяющие распределение фитопланктона в фотической зоне моря. Причём, свет проникает сверху, а биогенные соединения поступают в фотический слой из нижних слоёв. Оба фактора образуют по вертикали противоположно направленные градиенты. В нижней части фотического слоя развитие фитопланктона ограничивает недостаток освещённости, а в верхней – биогенов. Однако в данном слое численность фитопланктона почти линейно увеличивается с глубиной, что позволяет выбрать для аппроксимации данных линейную модель.

Следует отметить, что данный пример упрощает реальную ситуацию и составлен специально для учебных целей.

С борта научно-исследовательского судна батометром берутся пробы воды с разных глубин, количество клеток подсчитывается под микроскопом с помощью камеры Горяева (объём 0,01 мл) с точностью 20 %, освещённость измеряется подводным пиранометром (в 100 лк). Количество клеток, указанное в таблице 76, необходимо умножить на 10000 кл/литр.

Таблица 76. Условия проведения наблюдений и полученные результаты (подробности в тексте)

№	x_1	x_2	y_1	y_2	y_3	$y_{ср}$	$y_{теор}$
1	20	50	16	15	17	15	15,6
2	20	44	14	19	15	16	17,7
3	24	36	18	22	23	21	20,34
4	28	26	20	23	26	23	23,46
5	36	24	22	26	27	25	27,32
6	48	20	28	25	31	28	27,96
7	70	14	28	30	35	31	32,4
8	110	10	36	42	39	39	38,4

Проверка гипотезы о нормальности распределения зависимой переменной.

4.1.2. Вычислительная процедура регрессионного анализа

Как уже упоминалось, регрессионный анализ проводится в предположении о том, что результаты наблюдений являются случайными нормально распределёнными величинами, а дисперсии при повторах различных опытов – однородны. Следовательно, необходимо проверить гипотезу о нормальности распределения зависимой переменной и гипотезу об однородности дисперсий в повторностях всех наблюдений.

В доступной литературе трудно найти подробное изложение несложного метода проверки нормальности. В данном методе определение нормальности распределения результатов наблюдения определяется путём сравнения расчётного критерия нормальности ($W_{расч}$) с табличным критерием (W). Расчётный критерий нормальности вычисляют по формуле: $W_{расч} = b^2/S^2$

Определение величин b и S^2.

Вспомогательная величина b вычисляется по формуле:

$$b = a_n(y_n - y_1) + (y_{n-1} - y_2) + ... + a_{n-k+1}(y_{n-k+1} - y_k) = \sum_{i=1}^{k} a_{n-i+1}(y_{n-i+1} - y_i)$$

здесь: n=24; k=n/2 = 12 (так как n – чётное число; если n – не чётное, то k=(n-1)/2).

Коэффициенты a_{n-i+1}, используемые при проверке на нормальность с помощью критерия W, берутся из таблицы Приложения VI.

Для n=24, k=12 ai равны:

a_{24}= 0,4493
a_{23}= 0,3098
a_{22}= 0,2554
a_{21}= 0,2145
a_{20}= 0,1807
a_{19}= 0,1512
a_{18}= 0,1245
a_{17}= 0,0997
a_{16}= 0,0764
a_{15}= 0,0539
a_{14}= 0,0321
a_{13}= 0,0107

b = 04493·28 + 0,3098·24 + 0,2554·21 + 0,2145·19 + 0,1807·14 + 0,1512·12 + 0,1245·9 + +0,0997·8 + 0,0764·5 + 0,0539·4 + 0,0321·3 + 0,0107·2 = 35,914; b^2 = 128,500.

Расчёт S^2.

Составление упорядоченной выборки значений зависимой переменной:

$14 < 15 \leq 15 < 16 < 17 < 18 < 19 < 20 < 22 \leq 22 < 23 \leq 23 < 25 < 26 \leq 26 < 27 < 28 \leq 28 < 30 < 31 < 35 < 36 < 39 < 42$

Расчёт величины S^2:

$$S^2 = \sum_{i=1}^{n} y_i^2 - \frac{(\sum_{i=1}^{n} y_i)^2}{n}$$

$\sum_{i=1}^{n} y_i^2$ = 196 + 225 + 225 + 256 + 289 + 324 + 362 + 400 + 484 + 484 + + 529 + 625 + 72 + 672 + 729 + 784 + 784 + 900 + 961 + 1250 + 1296 + + 1521 + 1764 = 16369;

$(\sum_{i=1}^{n} y_i)^2/n = 212^2/24 = 44944/24 = 1873;$

S^2= 16369 – 1873 = 14496.

Расчётный критерий нормальности: $W_{расч}=b^2/S^2$=128500/14496 = 8,8.

Табличный критерий (для α=0,05 и n=24, Приложение VII) W=0,916; Wрасч≥W. Следовательно, гипотеза о нормальном распределении зависимой переменной принимается.

Оценка воспроизводимости результатов наблюдений

В данном пассивном эксперименте число параллельных наблюдений одинаково во всех точках проведения замеров и равно 3. Поэтому гипотеза об однородности дисперсий во все выборках проверяется с помощью критерия Кохрена G:

$$G = \frac{S^2{}_{g\max}}{\sum_{m=1}^{m} S^2{}_g}$$

$$S^2{}_{gm} = \frac{\sum_{i=1}^{m} (y_{gi} - y_{gcp})^2}{m-1}$$

(y_{gcp} – приведены в таблице 76)

Расчёт дисперсий по каждой точке взятия проб воды:

$S^2{}_{g1}$= 1/2 (1+0+4) = 2,5
$S^2{}_{g2}$= 1/2 (4+9+1) = 7
$S^2{}_{g3}$= 1/2 (9+1+4) = 7
$S^2{}_{g4}$= 1/2 (9+0+9) = 9
$S^2{}_{g5}$= 1/2 (9+1+4) = 7
$S^2{}_{g6}$= 1/2 (0+9+9) = 9
$S^2{}_{g7}$= 1/2 (9+1+16) = 13
$S^2{}_{g8}$= 1/2 (9+0+9) = 9

$\sum S^2_g = 63{,}5$; $G_{max} = 13 / 63{,}5 = 0{,}205$.

Табличное значение С при: $\alpha=0{,}05$; N=8; m-1=2 $C^8_{(3-1)}= 0{,}515$ $0{,}205<0{,}515$

Следовательно, дисперсии однородны (наблюдения воспроизводимы, иными словами, наблюдения повторяются с одинаковой ошибкой).

Расчёт коэффициентов уравнения регрессии.

Из таблицы 76 следует, что $y_{ср}$ изменяется примерно линейно (хотя факторы изменяются не линейно). Поэтому в качестве искомой модели выбирается линейная регрессия:

$y = b_0 + b_1x_1 + b_2x_2$

Построение матрицы **X**

Эта матрица должна содержать три столбца (x_0, x_1, x_2).

Таблица 77. Матрица **X** (натуральные значения переменных)

x_0	x_1	x_2
1	20	50
1	20	44
1	24	36
1	28	26
1	36	24
1	48	20
1	70	14
1	110	10

Для упрощения дальнейших расчётов кодируем переменные по формулам:

$X_1= (x_1 – 44)/4$ и $X_2= (x_2 – 28)/2$.

Получаем матрицу **X** с кодированными значениями переменных (табл. 78).

Таблица 78. Матрица **X** (кодированные значения переменных)

X_0	X_1	X_2
1	-6	11
1	-6	8
1	-5	4
1	-4	0
1	-2	-2
1	1	-4
1	9	-7
1	11	-9

Определяем элементы уравнения (3):

$$X^T \times X = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -6 & -6 & -5 & -4 & -2 & 1 & 9 & 11 \\ 11 & 8 & 4 & 0 & -2 & -4 & -7 & -9 \end{bmatrix} \times \begin{bmatrix} 1 & -6 & 11 \\ 1 & -6 & 8 \\ 1 & -5 & 4 \\ 1 & -4 & 0 \\ 1 & -2 & -2 \\ 1 & 1 & -4 \\ 1 & 9 & -7 \\ 1 & 11 & -9 \end{bmatrix} = \begin{bmatrix} 8 & -2 & 1 \\ -6 & 320 & -296 \\ 1 & -296 & 351 \end{bmatrix}$$

Проверка вырожденности информационной матрицы (X^TX)

Известно, что матрица называется вырожденной, если её определитель равен нулю.

$det(X^TX)$ **=** 8·320·351+2·296·1+2·296·1-320-4·351-296²·8 = 896000 + 592 + 592-320-1204-700000 = 195660

Информационная матрица (X^TX) не вырождена, следовательно, регрессионный анализ в данном случае применим.

Определение обратной матрицы $(X^TX)^{-1}$ по формуле:

$$(X^TX)^{-1}=C_{(X^TX)}/\,det(X^TX),$$

где: $C_{(X^TX)}$ – присоединённая матрица

$$C_{(X^TX)} = \begin{bmatrix} A^1{}_1 & A^2{}_1 & A^3{}_1 \\ A^1{}_2 & A^2{}_2 & A^3{}_2 \\ A^1{}_3 & A^2{}_3 & A^3{}_3 \end{bmatrix}$$

Здесь A^i_j – алгебраические дополнения (миноры со знаками);

i – номер строки

j – номер столбца

Если i+j – чётное число, берётся знак +. В противном случае – знак минус.

Расчёт алгебраических дополнений:

$$A^1{}_1 = \begin{bmatrix} 320 & -296 \\ -296 & 351 \end{bmatrix} =11200+8770=19970$$

$$A^1{}_2 = -\begin{bmatrix} -2 & -296 \\ 1 & 351 \end{bmatrix} = 702-296 = 406$$

$$A^1{}_3 = \begin{bmatrix} -2 & 320 \\ 1 & -296 \end{bmatrix} = 592-320 = 272$$

$$A^2{}_1 = -\begin{bmatrix} -2 & 1 \\ -296 & 351 \end{bmatrix} = 702 - 296 = 406$$

$$A^2{}_2 = \begin{bmatrix} 8 & 1 \\ 1 & 351 \end{bmatrix} = 3608 - 1 = 3607$$

$$A^2{}_3 = -\begin{bmatrix} 8 & -2 \\ 1 & -296 \end{bmatrix} = 2378 - 2 = 2376$$

$$A^3{}_1 = \begin{bmatrix} -2 & 1 \\ 320 & -296 \end{bmatrix} = 272$$

$$A^3{}_2 = -\begin{bmatrix} 8 & 1 \\ -2 & -296 \end{bmatrix} = 2376$$

$$A^3{}_3 = \begin{bmatrix} 8 & -2 \\ -2 & 320 \end{bmatrix} = 2560 - 4 = 2556$$

Обратная матрица:

$$(X^TX)^{-1} = C_{(X^TX)}^{T} / \det(X^TX)$$

$$(X^T X)^{-1} = \begin{bmatrix} 0{,}102 & 0{,}00204 & 0{,}00140 \\ 0{,}00204 & 0{,}0185 & 0{,}0122 \\ 0{,}00140 & 0{,}0122 & 0{,}0131 \end{bmatrix}$$

$$X^T Y = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -6 & -6 & -5 & -4 & -2 & 1 & 9 & 11 \\ 11 & 8 & 4 & 0 & -2 & -4 & -7 & -9 \end{bmatrix} \times \begin{bmatrix} 15 \\ 16 \\ 21 \\ 23 \\ 25 \\ 28 \\ 31 \\ 39 \end{bmatrix} = \begin{bmatrix} 198 \\ 303 \\ -453 \end{bmatrix}$$

Коэффициенты уравнения регрессии рассчитываются по формуле:
$B = (X^TX)^{-1} X^TY$

$$B = \begin{bmatrix} 0{,}102 & 0{,}00204 & 0{,}00140 \\ 0{,}00204 & 0{,}0185 & 0{,}0122 \\ 0{,}00140 & 0{,}0122 & 0{,}0131 \end{bmatrix} \times \begin{bmatrix} 198 \\ 303 \\ -453 \end{bmatrix} = \begin{bmatrix} 20{,}18 \\ 0{,}48 \\ -1{,}96 \end{bmatrix}$$

Итак, уравнение регрессии: $y=20,18+0,48x_1-1,96x_2$.
Проверка значимости коэффициентов уравнения регрессии.
Данная проверка выполняется по критерию Стьюдента:

$t_i = [b_i] / S\{b_i\}$, $S\{b_i\}= (X^TX)^{-1} \cdot S^2\{y\}$

$$S^2\{y\} = \frac{1}{n}\sum_{g=1}^{n} S^2{}_g = \frac{1\times 63{,}5}{8} = 7{,}9$$

$S\{b_0\} = 7{,}9{\cdot}0{,}102 + 7{,}9{\cdot}0{,}00204 + 7{,}9{\cdot}0{,}00140 = 0{,}877$
$S\{b_1\} = 7{,}9{\cdot}(0{,}00204 + 0{,}0185 + 0{,}0122) = 0{,}0409$
$S\{b_2\} = 7{,}9{\cdot}(0{,}00140 + 0{,}0122 + 0{,}0131) = 0{,}0323$

$t_0 = 20{,}4 / 0{,}877 = 25{,}3$; $t_1 = 0{,}12 / 0{,}0409 = 3$; $t_0 = 0{,}30 / 0{,}032 = 9{,}4$.

Табличный критерий Стьюдента при $\alpha=0{,}05$ и числе степеней свободы $n(m-1) = 8{\cdot}2 = 16$ равен: $t_{табл} = 2{,}12$, что ниже расчётных критериев Стьюдента для коэффициентов уравнения. Таким образом, все коэффициенты уравнения регрессии статистически значимы.

В этом случае проверка адекватности полученного уравнения экспериментальным данным не выполняется. Тем не менее, ниже производится данная проверка, так как цель данного раздела – полное изложение вычислительной процедуры регрессионного анализа.

Проверка адекватности уравнения регрессии результатам наблюдений

Проверка адекватности производится по критерию Фишера.

$F= S^2_{ad} / S^2$; $$S^2 = \frac{\sum_{1}^{8} S^2{}_g}{8} = \frac{63{,}5}{8} = 7{,}95$$

$$S^2{}_{ad} = \frac{m}{n-d}\sum_{g=1}^{n}(y_{cpg} - y_{теорg})^2$$

$m=3$; $n=8$; $d=3$; $[y_{ср.g} - y_{теор.g}] = \delta y_g$ $g=1; 2...8$

$\delta y_1 = 0{,}6$ $(\delta y_1)^2 = 0{,}36$
$\delta y_2 = 1{,}7$ $(\delta y_2)^2 = 2{,}89$

$\delta y_3 = 0{,}66$ $(\delta y_3)^2 = 0{,}435$
$\delta y_4 = 0{,}46$ $(\delta y_4)^2 = 0{,}21$
$\delta y_5 = 2{,}3$ $(\delta y_5)^2 = 5{,}30$
$\delta y_6 = 0{,}1$ $(\delta y_6)^2 = 0{,}01$
$\delta y_7 = 1{,}4$ $(\delta y_7)^2 = 1{,}96$
$\delta y_8 = 0{,}6$ $(\delta y_4)^2 = 0{,}36$

$$\sum_{g=1}^{8} (\delta y_g)^2 = 11{,}52$$

$S^2_{ad} = 3/(8\text{-}3)\cdot 11{,}52 = 6{,}9$

$F_{расч} = 6{,}9/7{,}95 = 0{,}867$

Табличное значение критерия Фишера для α=0,05 и чисел степеней свободы n-d = 8 – 3 = 5; n(m-1)=16 равен: $F_{табл}$= 2,9 $F_{расч} < F_{табл}$ Следовательно, уравнение адекватно.

Это уравнение пригодно для задач интерполяции, то есть вычисления численности клеток фитопланктона в случаях, когда фактор $\mathbf{x_1}$ изменяется в пределах 20 – 110, а фактор $\mathbf{x_2}$ – в пределах 10 – 50 условных единиц. За пределами этих интервалов (задачи экстраполяции) точность определения **y** быстро снижается. Второе замечание: план сбора наблюдений (таблица 76) – не ортогонален, то есть коэффициент корреляции между столбцами для факторов $\mathbf{x_1}$ и $\mathbf{x_2}$ не равен нулю. Поэтому коэффициенты уравнения регрессии получены не независимо друг от друга (оценки их величины смешаны). Следовательно, величины коэффициентов нельзя интерпретировать в качестве показателей силы влияний соответствующих факторов на величину численности клеток фитопланктона.

4.2. Обработка на персональном компьютере (ПК) данных, полученных при проведении пассивного эксперимента

Данные наблюдений формируются в виде таблицы, в которой в строках указаны наблюдения (или станции, или точки отбора проб в море (case)), часто указан порядковый № наблюдения. В столбцах записаны значения измеряемых факторов или переменных (variable).

Существуют различные пакеты программ, но самым полным является пакет Statistica. Возьмём этот пакет за основу (Халафян, 2007).

Данные в Statistica организованы в виде электронной таблицы (ЭТ). Существуют различные операции для работы с ЭТ:

- Операции изменения структуры ЭТ (добавления, удаления, копирования, перемещения переменных и наблюдений из электронной таблицы);
- Операции по заданию спецификаций (т.е. имён, форматов);
- Операции с выделенным блоком значений (вырезать, копировать, вставить, очистить) и т.д.

После введения данных начинается их обработка и анализ. Предварительный анализ – это визуализация исходных данных, то есть их просмотр в виде графиков и различных диаграмм (Graphs, Histogrammes), так как ряды цифр из таблицы наше сознание не способно анализировать. При этом можно обнаружить интересные закономерности, которые затем можно будет исследовать более тщательно. Можно вывести на экран матрицу графиков, то есть много графиков, отражающих парные связи различных переменных.

Затем получают основные статистические параметры или «Основные статистики» (Basic Statistic Tables).

Mean – среднее арифметическое;

Median – медиана;

Mode – мода;

Std.dev. – стандартное отклонение;

Std.err.of mean – стандартная ошибка среднего;

Limits of Confidence – доверительные интервалы;

Varience – коэффициент вариации;

Skewness – ассиметрия, то есть мера симметричности распределения. Если распределение симметрично, то ассиметрия = 0. Нормальное распределение симметрично.

Kurtosis – эксцесс – мера остроты пика. При нормальном распределении эксцесс=0

Связи, существующие между переменными, отражают коэффициенты парной корреляции (Correlation Matrix). Надо отметить, что коэффициент корреляции отражает упрощённую линейную связь, поэтому корреляционный анализ можно рассматривать в качестве предварительного анализа. Здесь необходимо дать некоторые разъяснения. Допустим, что измеряемые нами переменные (здесь они названы факторами) зависят от некоторых общих факторов, влияющих на измеряемые переменные, поэтому, как следствие зависимости нескольких факторов от одного общего фактора, проявляется кажущаяся зависимость этих нескольких факторов друг от друга. Например, с изменением глубины изменяются: освещённость, фотосинтез, температура, численность клеток, первичная продукция, концентрация хлорофилла в воде. Поэтому, при рассмотрении корреляционной матрицы (КМ) окажется, что все перечисленные факторы связаны друг с другом. КМ представляет собой квадратную таблицу размером к·к, где к – число переменных (факторов). Элементами диагонали являются 1, а остальные элементы – это коэффициенты парной корреляции, рассчитываемые по формуле:

$$r = \sum (x_i - x')(y_i - \tilde{y}) \Big/ \sqrt{\sum (x_i - x')^2 \sum (y_i - \tilde{y})^2}$$

Если зависимость между переменными линейна, то r может достигнуть +1 или -1.

Статистически значимые коэффициенты корреляции выделяются цветом. Большие значения r по абсолютной величине говорят о тесной связи между переменными, и можно предположить, что эти переменные входят в единую систему, которую нужно исследовать.

Далее, в зависимости от задачи, применяют тот или иной тип многомерного статистического анализа.

Дисперсионный анализ (Analysis of Variance, One-way ANOVA)

При необходимости сравнения средних применяют ДА. Например, при изучении многолетних изменений загрязнения моря требуется установить, изменяется (либо нет) загрязнение моря по годам.

Регрессионный анализ (Regression Analysis)

Линейная регрессионная модель (Linear Regression). Изучается зависимость одного параметра от одной или нескольких переменных:

$Y = bx$; $Y = b_0 + b_1x_1 + b_2x_2 + \ldots$

Программа рассчитывает методом наименьших квадратов коэффициенты и их доверительные интервалы; определяется значимость коэффициентов; оценивается адекватность полученного уравнения.

Это позволяет свернуть большой объём информации к одному уравнению. Но нужно помнить, что, если исходные переменные коррелированы, тогда коэффициенты уравнения нельзя интерпретировать.

Нелинейная регрессионная модель (Nonlinear Regression).

Если зависимость не линейная, то последовательно добавляются в уравнение нелинейные члены, например, квадраты переменных $Y = b_0 + b_1x_1 + b_2x_2 + b_{11}x^2_1 + b_{22}x^2_2 + \ldots$

Или можно задать более сложную зависимость, например, экспоненциальную регрессию:

$Y = c + \exp(b_0 + b_1x_1 + b_{11}x^2_1 + \ldots)$.

Канонический анализ (Canonical Analysis)

Канонический анализ исследует зависимости между группами переменных. Например, мы решили изучить связи между структурой и функцией фитопланктонного сообщества. Для этого мы выбираем переменные, характеризующие функцию (синтез органического вещества): концентрации пигментов (белки фотосинтетического аппарата), первичную продукцию, интенсивность фотосинтеза, а с другой стороны выбираем структурные показатели: численность и биомасса различных видов фитопланктона. Канонический анализ даст несколько зависимостей и коэффициенты множественной корреляции, отражающие тесноту связей между группами переменных.

Аналогично, можно проанализировать связь между симптомами разных заболеваний и уровнями неблагоприятных факторов или же между уловами различных видов рыб и наборами факторов, характеризующих процесс лова (тип орудия лова, судна, размер ячеи ловчей сети, глубина лова, температура морской воды и т.д.). В школе можно проанализировать связь между навыками по трём дисциплинам и отметками по пяти предметам:

$$a_1y_1 + a_2y_2 + a_3y_3 = b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 + b_5x_5$$

Параметры a и b подбираются таким образом, чтобы R стремился к 1. Вначале a_1 задаётся максимальным, потом a_2 и т.д.

Выводы бывают иногда неожиданными. Например, социологи пришли к выводу о том, что для большего получения удовлетворения от работы нужно в свободное время чаще ходить в театр и меньше по вечерам сидеть на диване.

Дискриминантный анализ (Discriminant Analysis)

Дискриминантный и кластерный анализы применяются для задач классификации многомерных наблюдений, в то время как факторный анализ направлен в основном на исследование связей между переменными, хотя последний также может быть применён для задач классификации.

Цель дискриминантного анализа: на основе измерения различных характеристик объекта исследования (ОИ) отнести его к той или иной группе, т.е. классифицировать. Задача дискриминации – это задача различения.

Каждое наблюдение над ОИ записывается в виде к-мерного вектора $X(x_1, x_2,...x_к)$. По значениям координат вектора объект X относят к той или иной совокупности объектов. Например, яблоки можно рассортировать по величине, цвету, вкусу (кислый – сладкий), твёрдости, лёжкости и т.д. Когда признаков много и много объектов, то их рассортировать вручную – сложно.

Кластерный анализ (Cluster Analysis)

Cluster – гроздь, скопление. Метод выполняет разбиение множества объектов на группы разной степени сходства, то есть создание кластера объектов.

Цель кластерного анализа – большое число исходных переменных представить гораздо меньшим числом объектов (кластеров), что выполняется путём группировки переменных. Кластерный анализ применяется обычно для выделения групп объектов, на основе их сходства по измеренным признакам.

Часто применяют кластерный анализ на первых этапах с целью разбиения огромных массивов данных на родственные группы. Это могут быть как объекты, так и факторы, которые ведут себя сходным образом. Затем можно проводить исследования отдельных групп, либо наоборот: взять по одному представителю из каждой группы, который и будет отражать поведение всех членов группы.

Программа работает так: вначале разбивает все объекты на группы, внутри которых объекты отличаются минимально. Затем некоторые группы объединяются в надгруппы, в которых все объекты обладают меньшей степенью сходства. Затем происходит объединение по ещё более различимым признакам и т.д. В конце концов, все группы объединяются в один большой кластер. Аналогия: виды> роды> семейства> отряды> классы> типы> царство.

На каждом уровне объединения указана степень сходства (родства).

Сделав такое разбиение, можно задаться степенью сходства и выбрать соответствующие группы.

Пример практического применения кластерного анализа взят из:

http://www.statsoft.ru/solutions/ExamplesBase/branches/detail.php?ELEMENT_ID=1573

В данном примере исследуются автомобили (22 марки) и водители с целью разбиения их на классы, каждый из которых соответствует определенной рисковой группе. Каждый автомобиль с водителем характеризуется набором переменных (цена автомобиля, его возраст, возраст водителя, его стаж и т.д.). Наблюдения, попавшие в одну группу, характеризуются одинаковой вероятностью наступления страхового случая, которая впоследствии оценивается страховщиком. Мера близости, определяемая евклидовым расстоянием, является геометрическим расстоянием в n-мерном пространстве и вычисляется следующим образом:

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

Чем меньше расстояние между каждым конкретным случаем, тем теснее связь внутри кластера (рис. 14).

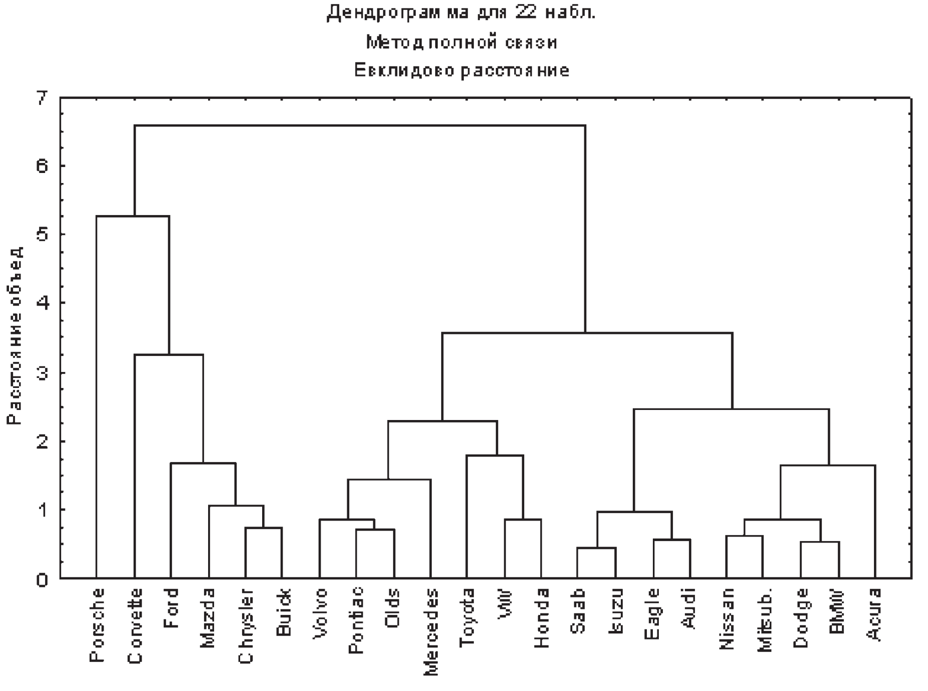


Рис. 14. Кластеры, сформированные из 22 марок автомобилей

Рассматривать диаграмму следует начинать сверху для каждого автомобиля. При перемещении сверху вниз автомобили, которые наиболее близко связаны, образуют кластеры. Каждый узел диаграммы, приведенной выше, представляет объединение двух или более кластеров, положение узлов на вертикальной оси определяет расстояние, на котором были объединены соответствующие кластеры. Чем меньше расстояние, тем теснее связаны автомобили.

4.3. Факторный анализ

Существует несколько десятков вариантов факторного анализа. Рассмотрим один из вариантов, который называется «Метод главных компонент» (Principal Components). Иногда этот метод вычленяют в самостоятельный метод. МГК позволяет решить целый ряд задач. Ниже приведено обсуждение наиболее часто встречающихся задач.

Допустим, что при выполнении регрессионного анализа данных с целью нахождения переменных (факторов), сильно и слабо влияющих на изучаемый процесс, обнаружились слабые факторы, которые можно было бы отбросить. Но корреляционная матрица показала, что переменные сильно коррелируют, поэтому интерпретация коэффициентов уравнения регрессии – невозможна. Но в этом случае можно предположить, что переменные зависят от небольшого числа общих причин (факторов), которые невозможно измерить непосредственно. Например, наступление зимы измерить непосредственно нельзя. Но можно зафиксировать последовательное уменьшение продолжительности светового дня, уменьшение освещённости, допустим, в полдень; понижение температуры воздуха, увеличение его влажности, усиление осадков; направление и средняя скорость ветра также меняется и т.д. Все эти факторы последовательно изменяются во времени, отражая определённую тенденцию, в данном случае, наступление зимы. Такие тенденции и называются главными компонентами, либо главными факторами. Другой пример тенденции – влияние глубины, которое проявляется в целом ряде коррелированных изменений большого числа переменных. Ещё пример тенденции – онтогенез, или индивидуальное развитие, которое непосредственно измерить невозможно. Но можно определить увеличение размеров тела и веса с возрастом, а также замедление метаболизма, понижение соотношения пластического и энергетического обмена и т.д.

Основная цель МГК – выявление наиболее общих свойств изучаемого процесса, вызывающих коррелированное изменение измеряемых признаков. Кроме этой задачи, можно выделить ещё, по крайней мере, четыре:

- Сжатие информации, или описание изучаемого процесса числом ГК, значительно меньшим количества измеряемых признаков;
- Исследование связей между признаками и главными компонентами, что позволяет проанализировать природу признаков;
- Классификация наблюдений, либо исследуемых объектов;
- Прогнозирование поведения (изменений) ОИ, что выполняется на основе уравнения регрессии, построенного по полученным главным компонентам.

Одна из задач, как сказано выше, – выявление ГК. Именно они и объяснят истинные причины изменений измеряемых переменных, то есть причины разброса данных и вклад каждой из этих причин в общую дисперсию.

Эти причины (то есть главные компоненты) представляют в виде линейных комбинаций, составленных из анализируемых переменных. В терминах матричной алгебры они являются собственными векторами корреляционной матрицы. Из собственных векторов строится ортогональная система координат, в которой первая ось идёт в направлении максимального рассеяния исходных данных. Таким образом, первая ось отражает наиболее сильную тенденцию, и о её вкладе в объяснение суммарной дисперсии говорит собственное число, которое и показывает, какую часть общей дисперсии объясняет главная компонента.

Вторая ось идёт в направлении ортогональном к первой ГК, причём в направлении наибольшего изменения (за исключением первой оси) переменных и т.д.

Каждая из исходных переменных вносит разный вклад в формирование ГК. Этот вклад отражён соответствующим коэффициентом и называется нагрузкой. Примеры ГК будут приведены при рассмотрении конкретных случаев применения этого метода.

Так как ГК ортогональны, поэтому представляется важным нахождение уравнения регрессии исследуемых переменных от полученных главных компонент. Это поможет уточнить природу изучаемых переменных и выяснить, что и как влияет на них. В пространстве ГК можно разместить исходные переменные произвести, таким образом, их разбивку на сходные группы, иными словами, выполнить классификацию исходных переменных, что позволит в дальнейшей работе использовать только отдельных представителей этих групп и тем самым сократить объём работы.

4.3.1. Пример применения факторного анализа

Цель работы: анализ влияния состава грунта на численность разных групп зообентоса и исследование связей между организмами макробентоса, эумейобентоса и псевдомейобентоса (Холодов, Киселёва, 1985). Иными словами, изучение связей между организмами разных размерных групп.

Справка:
Микробентос, размеры менее 100 мк.
Эумейобентос, размеры 100 мк – 1 мм.
Псевдомейобентос, молодь макробентоса, размеры 100 мк – 1 мм.
Макробентос, размеры более 1 мм. Сюда входят все крупные животные.

Пробы отбирали дночерпателем в районе Ялты на глубинах от 20 до 125 м. Измеряли 12 показателей (переменных): глубина, компоненты гранулометрического состава грунта (содержание гранул разного размера в %), а также количество видов по разным систематическим группам. Работали на 38 станциях. Было собрано 150 проб.

Измеряемые переменные:
x_1 – глубина;
x_2 – крупная фракция грунта;
x_3 – средняя фракция;
x_4 – мелкая фракция грунта;
x_5 – самая мелкая фракция грунта;
x_6 – макробентос (количество видов);
x_7 – нематоды (количество видов);
x_8 – общая численность макробентоса;
x_9 – массовые виды макробентоса (численность);
x_{10} – моллюски (численность);
x_{11} – общая численность эумейобентоса;
x_{12} – численность мелких нематод.

После обработки проб данные были оформлены в виде таблицы, состоящей из 150 строк и 12 столбцов. Таблица данных была обработана методом главных компонент.

Однако предварительно, в качестве визуализации основных показателей, было решено проанализировать характер зависимости видового разнообразия от гранулометрического состава грунта. С этой целью построили трехмерные графики распределения числа видов на грунтах, состоящих из разных фракций. В основании графиков лежат треугольные диаграммы гранулометрического состава грунта. В вершинах диаграммы: крупные (более 0,05 мм), средние (0,005–0,05 мм) и мелкие (менее 0,005 мм) гранулы. В результате стало ясно, что основная масса видов тяготеет к грунту, содержащему в основном мелкие и средние фракции, особенно чётко эта тенденция проявляется при анализе распределения полихет: чем реже встречается крупная фракция в грунте, тем больше отмечается видов полихет.

Можно сказать, что предварительный анализ позволил получить картину приуроченности разных групп организмов к различным типам грунтов.

Корреляционную матрицу проанализировали методом корреляционных плеяд (КП). КП – это группа изучаемых переменных, связанных между собой определённым уровнем коэффициента корреляции (Терентьев, 1959). Вначале выбирают переменные, связанные между собой наиболее тесно, то есть с большим (по абсолютному значению) коэффициентом корреляции. В данном случае выявились виды, входящие в один общий биотоп, а также виды, зависимые друг от друга. Установлены виды среди моллюсков, червей и ракообразных, связанных многочисленными связями с другими видами. Выявились также и виды, в том числе массовые (нематоды), численность которых почти не связана с численностями других организмов.

Для более детального исследования были вычислены ГК, причём 5 из них объяснили 88% суммарной дисперсии (всей изменчивости данных).

Наиболее типичную тенденцию в распределении организмов характеризует первая ГК (объясняет 25% общей дисперсии), которая выделяет условия обитания макробентоса и связанных с ним нематод, а также обилие моллюсков. В соответствии с первой строкой таблицы 79, первая ГК имеет вид:

$$1\text{-ая ГК} = 0{,}102{\cdot}x_1 - 0{,}211{\cdot}x_2 + 0{,}075{\cdot}x_3 + 0{,}248{\cdot}x_4 + 0{,}055{\cdot}x_5 + 0{,}105{\cdot}x_6 + 0{,}329{\cdot}x_7 + 0{,}509{\cdot}x_8 + 0{,}507{\cdot}x_9 + 0{,}441{\cdot}x_{10} - 0{,}007{\cdot}x_{11} + 0{,}0027{\cdot}x_{12}$$

Величина коэффициента и его знак указывают на роль соответствующего фактора в формировании тенденции. Эта тенденция показывает, что в грунте этого биотопа преобладают средние гранулы, а крупные отсутствуют. Благодаря обилию фазеолины интенсивно продуцируется фазеолиновый ил, который и формирует грунт, а следовательно, способствует расселению живущих в нём животных. Данную тенденцию можно назвать: «Влияние фазеолины на распределение бентосных организмов».

Таблица 79. Коэффициенты (нагрузки) первых пяти главных компонент, рассчитанные по видовому разнообразию и численности разных групп зообентоса, обитающих на грунтах различного гранулометрического состава

Главная компонента	Глубина, м	Размер гранул, мм				Количество видов		Численность				
								макробентос			эумейобентос	
		>0,1	0,1-0,01	0,01 0,001	< 0,001	макбентос	нематод	общая	массов.	моллюск	общая	нематоды
	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}
1	0,102	-0,2	0,07	0,25	0,05	0,10	0,33	0,51	0,51	0,44	-0,01	0,003
2	-0,14	0,47	-0,3	-0,32	-0,05	-0,07	-0,2	0,20	0,20	0,18	-0,44	-0,44
3	-0,59	0,12	0,26	0,36	-0,39	0,40	0,06	0,08	0,06	0,08	0,19	0,24
4	0,097	0,34	-0,5	0,17	-0,33	-0,30	-0,2	0,11	0,12	0,07	0,45	0,40
5	-0,23	0,13	0,01	-0,39	0,61	-0,43	0,33	-0,0	-0,0	0,15	0,15	0,26

Вторая ГК выделяет тенденцию, объясняющую 20% всей дисперсии. Она отражает условия, неблагоприятные для эумейобентоса, а именно: пониженное содержание мелких фракций и повышенное крупных ведёт к понижению общей численности нематод, хотя их видовое разнообразие почти не снижается.

Третья ГК (15% всей дисперсии) выделяет условия максимального видового разнообразия макробентоса, что никак не связано с численностью этих видов. Условия таковы: малые глубины, низкое содержание в грунте мелких фракций. Иными словами, на мелководье грунты без мелких фракций заселены многими видами макробентоса. На больших глубинах – тенденция обратная. Здесь проявляется межфакторное взаимодействие, а именно – характер влияния глубины на видовое разнообразие макробентоса зависит от гранулометрического состава грунта.

Итак, анализ структуры донных сообществ показал, что изменение численности исследованных видов объясняется связями, существующими между группами бентоса. Наиболее сильные связи – это связи внутриценотические и, прежде всего, связи, существующие между группами мейобентоса. Затем выделяются связи между различными группами макробентоса. Более слабыми, но всё же чётко определяемыми являются межценотические связи: трофические и связи, обусловленные общим биотопом молодых и взрослых животных, относящихся к малоподвижным формам.

4.3.2. Пример применения метода главных компонент

Выполнено исследование распределения основных групп фитопланктона: диатомовых, перидиниевых, кокколитофорид и мелких жгутиковых в фотическом слое вихревого образования Саргассова моря. Работы выполнялись с борта научно-исследовательского судна «Михаил Ломоносов» (Бурлакова, Холодов, Кузьменко, 1983).

Исследования проводили на четырёх станциях меридионального и шести станциях широтного направлений, на которых измеряли 20 гидрологических, гидрохимических и гидробиологических факторов: H – глубина, м; t – температура воды, °С; S – солёность, ‰; T – показатель ослабления излучения света, м$^{-1}$; N – концентрация нитратов, мкг-ат/л; P – концентрация фосфатов, мкг-ат/л; Si – концентрация силикатов, мкг-ат/л; Cl – содержание хлорофилла «*a*», мг/м^3; Pr – первичная продукция, мгC/м^3·день; $C_{орг}$ – концентрация взвешенного органического углерода, мг/м^3; $N_{пер}$ – численность перидиниевых, млн. кл/м^3; $B_{пер}$ – биомасса перидиниевых, мгC/м^3; $N_{диат}$ – численность диатомовых, млн. кл/м^3; $B_{диат}$ – биомасса диатомовых, мгC/м^3; $N_{кокк}$ – численность кокколитофорид, млн. кл/м^3; $B_{кокк}$ – биомасса кокколитофорид мгC/м^3; $N_{жгут}$ – численность мелких жгутиковых, кл/м^3; $B_{жгут}$ – биомасса мелких жгутиковых, мгC/м^3; $N_{общ}$ – общая численность клеток фитопланктона; $B_{общ}$ – общая биомасса фитопланктона. МГК был применён для определения основных тенденций в распределении биотических и абиотических характеристик экосистемы.

Таблица 80. Коэффициенты (нагрузки) первых четырёх главных компонент, рассчитанные по 20 гидрологическим, гидрохимическим и гидробиологическим факторам на полигоне в Саргассовом море

Признак	Главные компоненты			
	Первая	Вторая	Третья	Четвёртая
H	0,537	0,759	-0,035	-0,042
$N_{пер}$	-0,780	-0,000	-0,011	0,022
$B_{пер}$	-0,738	-0,201	-0,279	0,191
$N_{диат}$	-0,664	0,064	-0,104	-0,261
$B_{диат}$	-0,412	-0,484	-0,031	0,215
$N_{кокк}$	-0,709	0,540	-0,092	-0,235
$C_{орг}$	-0,023	0,451	-0,078	-0,080
$B_{кокк}$	-0,763	0,298	-0,247	-0,255
$N_{жгут}$	-0,546	0,472	0,459	0,382
$B_{жгут}$	-0,305	0,609	0,205	0,389
$N_{общ}$	-0,873	0,420	0,147	0,048
$B_{общ}$	-0,885	-0,118	-0,261	0,079
t	-0,347	-0,796	-0,043	0,198
S	-0,441	-0,519	0,234	0,072
Pr	-0,113	-0,327	0,461	-0,557
N	0,333	0,414	-0,299	0,232
P	-0,036	-0,001	0,680	-0,301
Cl	0,295	0,632	-0,240	-0,028
Si	0,038	0,340	0,537	0,415
T	-0,389	0,704	0,027	-0,292

Главные компоненты (табл. 80) вносят разный вклад в объяснение всей изменчивости исследуемой системы признаков: первая извлекает 28,6% суммарной дисперсии; вторая – 22%. Вклады каждой из последующих малы и различаются незначительно. Следовательно, в системе довольно чётко проявляются две тенденции, которые и подлежат тщательному анализу.

Кроме этих наиболее важных двух тенденций, действуют многие другие «нетипичные тенденции», имеющие частное значение, что и обусловливает низкие абсолютные величины большинства коэффициентов парной корреляции между изучаемыми факторами. Интерпретация «нетипичных тенденций» весьма затруднительна и требует участия квалифицированных специалистов в исследуемой области.

Итак, знак перед коэффициентом указывает на возрастание (+) либо убывание (-) фактора в рамках данной тенденции (компоненты). Абсолютная величина фактора отражает величину его изменения. Первая главная компонента иллюстрирует основную закономерность распределения факторов в системе: с увеличением глубины

(H=0,537) несколько возрастает концентрация нитратов (N=0,333) и сильно снижается общие численность ($N_{общ}$= – 0,873) и биомасса ($B_{общ}$ – 0,885) фитопланктона, что происходит, главным образом, за счёт снижения численности и биомассы перидиниевых и кокколитофорид. Однако, содержание органического углерода, кремния и фосфора практически не зависит ни от глубины, ни от распределения фитопланктона. Тенденцию, вскрываемую 1-ой ГК, можно охарактеризовать как «закономерность распределения факторов по глубине фотического слоя».

Вторая главная компонента противопоставляет факторы, связанные с глубиной, факторам, коррелирующим с температурой воды. Очевидно, что в этой ГК отражено распределение факторов в пробах воды, взятых с больших глубин, где отмечается высокая концентрация мелких жгутиковых и кокколитофорид и низкая – диатомовых. Температура и солёность в этих пробах – понижены. Речь идёт о слое минимальной прозрачности, проходящем на скачке плотности вод. Таким образом, 22 % наблюдаемой изменчивости факторов обусловлено наличием слоя сезонного скачка плотности. Следует отметить, что в этом слое, кроме мелких жгутиковых и кокколитофорид, обнаружено повышенное содержание углерода взвешенного органического вещества, нитратов и кремния. Перидиниевые и диатомовые существенно не влияют на показатели оптической плотности на данном участке экосистемы.

На исследованном полигоне содержание фосфатов, как правило, было низким. Однако в некоторых пробах отмечены высокие концентрации фосфатов и силикатов, но содержание нитратов в них было пониженным. Из фитопланктона в таких пробах обнаруживаются только мелкие жгутиковые, что отражено в третьей главной компоненте. Отметим, что таких проб воды было сравнительно мало, и взяты они были на средних глубинах. Мы затрудняемся объяснить механизм формирования таких вод, тем не менее, МГК выявил наличие вод данного типа, что даёт повод для формулирования новых гипотез и планирования более глубоких исследований.

ЗАКЛЮЧЕНИЕ

В данной небольшой монографии изложены принципы и структура экспериментального исследования, а также обсуждаются различные подходы к организации и к выполнению полевых исследований. Достаточно подробно на примерах изложены схемы постановки сравнительно несложных, но эффективных экспериментов и даётся описание статистической обработки результатов экспериментов с последующей биологической интерпретацией полученных результатов.

Организация полевых исследований (расположение станций отбора проб и их общее количество и т.д.) зависит от поставленных целей и задач. Данные полевых наблюдений должны быть сгруппированы в виде таблиц, в которых отмечаются значения уровней всех факторов, а пропуски в таблицах не допускаются. Последующая статистическая обработка данных позволяет выявить связи между отдельными факторами, группами факторов, выполнить их классификацию, а также обнаружить тенденции в совместных изменениях многих факторов, что отражает динамику структуры изучаемого объекта (например, экосистемы) в зависимости от внешних или внутренних воздействий (сезонности, глубины акватории, стадии сукцессии, онтогенеза и т.д.).

О планировании эксперимента говорят как о науке, целью которой является получение наиболее экономичным способом наиболее надёжных выводов. При этом выводы (или результаты) представляются в стандартной форме, что делает их сопоставимыми с результатами, полученными другими исследователями. Иными словами, планировать эксперимент означает выбирать оптимальную схему эксперимента, дающую возможность получить надёжную информацию об объекте минимальным числом опытов.

Планирование эксперимента позволяет:

1. Получать результаты в виде универсального математического описания изучаемого объекта с учётом действия многих факторов, что даёт возможность применять результаты для различных ситуаций;
2. Уменьшить ошибку эксперимента и исключить влияние факторов, вносящих искажения в результат;

3. Получать для исследуемых объектов математическое описание, обладающее рядом оптимальных свойств;
4. На основе чётких формализованных правил принимать на разных этапах исследований решения о дальнейших действиях;
5. Быстро осуществлять оптимизацию процессов, не прибегая к дорогостоящим исследованиям механизмов этих процессов;
6. Получать количественные оценки влияний каждого фактора, а также их взаимодействий на изучаемый процесс;
7. Составить поэтапную программу исследований, в которой на первых этапах выполняются ориентировочные исследования с привлечением большого числа факторов, зафиксированных на многих уровнях. На заключительных этапах проводятся подробные эксперименты с участием небольшого числа наиболее существенных факторов.

К настоящему времени накоплено огромное количество работ, рассматривающих различные аспекты применения методов планирования экспериментов в различных областях человеческой деятельности, включая космические исследования. Обширная литература содержится и в Интернете (достаточно в поисковик ввести два слова: «планирование эксперимента»). Часто эта литература недоступна для пользователей с недостаточной математической и статистической подготовкой. В данной краткой монографии приведены основные, ставшие классическими, учебники по планированию, а также рассмотрены только наиболее широко применяемые планы экспериментов. Однако следует отметить, что в России и за рубежом разработано много различных планов, рассчитанных на решение задач разных типов, различной сложности и учитывающих специфические требования исследователей. Каждый план обладает своими отличительными свойствами и, прежде всего, соответствует определённому критерию оптимальности.

Всего применяют свыше 20 различных критериев оптимальности планов, которые подразделяются на две основные группы. К первой группе относят критерии, связанные с ошибками оценок коэффициентов уравнения регрессии, а ко второй – с ошибками в определении поверхности отклика.

Критерии первой группы важны при решении задач оптимизации и выделения существенных факторов. Например:

D-оптимальные планы – позволяют получить наиболее точные значения коэффициентов регрессии, причём их дисперсии (следовательно, и ошибки) равны для всех коэффициентов. Планы этого типа являются ортогональными. Поэтому коэффициенты в уравнении регрессии определяются независимо от влияний других факторов, что важно при проведении научных исследований.

А-оптимальные планы – обеспечивают требование минимальности средней (либо суммарной) дисперсии всех коэффициентов регрессии.

Е-оптимальные планы – дают возможность получать коэффициенты регрессии с заданной точностью. В этом случае ошибки коэффициентов не превосходят заданные пределы.

Критерии второй группы применяются либо в задачах интерполяции, либо в задачах экстраполяции.

Для задач интерполяции:

G-оптимальные планы – гарантируют, что в интерполяционных расчётах не окажется точек на поверхности отклика, для которых ошибка предсказанного значения будет слишком большой.

Q-оптимальные планы – обеспечивают минимальность средней дисперсии расчётных значений независимой переменной (минимальность средней дисперсии поверхности отклика).

Планы, отвечающие требованиям униформности, обеспечивают постоянное и минимальное значение ошибки расчётного значения зависимой переменной.

Для задач экстраполяции используются рототабельные планы, обеспечивающие получение предсказанных (экстраполированных) значений отклика (зависимой переменной), которые не зависят от направления, задаваемого из центральной точки экспериментального плана. Поэтому ошибка предсказанного значения одинакова для всех точек, равноудалённых от центра эксперимента. Эта особенность важна при осуществлении движения к оптимуму методом крутого восхождения. Экспериментальные точки для рототабельных планов выбираются на поверхности шара. В частности, планы ПФЭ являются рототабельными, так как экспериментальные точки ПФЭ лежат в вершинах куба, а куб, как известно, вписывается в шар.

Планирование пассивных экспериментов позволяет повысить эффективность полевых исследований и выполнить глубокий анализ собранных данных, изучить связи между измеренными показателями экосистемы, выявить основные тенденции в их изменениях. Нередко полевые исследования приводят к выдвижению гипотез, например, о механизмах функционирования экосистем, либо о механизмах взаимодействий различных компонентов экосистем. С использованием методов планирования многофакторных экспериментов оказывается возможной проверка гипотез. И наоборот: выводы, полученные в экспериментальных исследованиях (активные эксперименты), стимулируют и уточняют планирование полевых исследований.

Наиболее эффективная стратегия гидробиологических исследований – разумное сочетание активных и пассивных экспериментов.

СПИСОК ЛИТЕРАТУРЫ

1. *Адлер Ю.П., Маркова Е.В., Грановский Ю.В.* Планирование эксперимента при поиске оптимальных условий. М. Наука, 1971. – 283 с.

2. *Адлер Ю.П., Горский В.Г.* Планирование промышленных экспериментов. М., Металлургия, 1974 – 264 с.

3. *Асатурян В.И.* Теория планирования эксперимента: для втузов. М. Радио и связь, 1983. – 248 с.

4. *Бурлакова З. П., Холодов В. И., Кузьменко Л. В.* Структура сообщества фитопланктона в районе вихревого образования в Саргассовом море // Экология моря, 1983, вып. 13. с. 3–13.

5. *Гинберг А.М., Грановский Ю.В., Федотова Р.Я., Калмуцкий В.С.* Оптимизация технологических процессов в гальванотехнике. М., «Машиностроение», 1972 – 128 с.

6. *Джонсон Н.* и др. Статистика и планирование эксперимента в технике и науке: Методы планирования эксперимента. М. Мир, 1981. – 520 с.

7. *Ладыгина Л.В., Пиркова А.В.* Оптимизация биотехники культивирования личинок гигантской устрицы *Crassostrea gigas* Th. в питомнике //Экология моря. – 2002. – Вып. 60. – С. 60–64.

8. *Маркова Е.В.* (разработчик). Руководство по применению латинских планов при планировании эксперимента с качественными факторами. Челябинск, 1971. (материал изложен в рецептурном плане).

9. *Маркова Е.В., Лисенков А.Н.* Планирование эксперимента в условиях неоднородностей. М. Наука, 1973.

10. *Мельников С.В.* и др. Планирование эксперимента в исследованиях сельскохозяйственных процессов. М. Колос, 1980. – 168 с.

11. *Монтгомери Д.К.* Планирование эксперимента и анализ данных. Л. Судостроение, 1980 – 384 с.

12. *Налимов В.В.* Теория эксперимента, М. Наука, 1971.

13. *Налимов В.В., Голикова Т.И.* Логические основания планирования эксперимента. М. Металлургия, 1976 – 128 с.

14. *Пиркова А.В., Холодов В.И., Ладыгина Л.В.* Оптимизация некоторых элементов культивирования личинок мидии *Mytilus galloprovincialis* Lam. // Гидробиол. журн. – 1998. – Т. 34, № 1. – С. 57–61.

15. *Пиркова А.В., Ладыгина Л.В.* Определение оптимальных условий роста и выживаемости личинок устрицы *Crassostrea gigas* (Th.) на

разных стадиях развития //Рыбное хоз-во Украины. – 2004. – № 6. – С. 174–177.

16. *Рябушко В.И., Холодов В.И.* Нетрофическое взаимодействие морских звёзд и микрофлоры каменистого грунта // Биология моря. – 1985. – № 5. –С. 18–26.

17. *Спиридонов А.А.* Планирование эксперимента при исследовании технологических процессов. М. Машиностроение. 1981 – 184 с.

18. *Терентьев П.В.* Метод корреляционных плеяд // Вести ЛГУ 1959 – 1959. – № 9. Сер. Биол., вып. 2, С 137–141.

19. *Халафян А.А.* STATISTICA 6 // Статистический анализ данных. М. Изд-во БИНОМ, 2007. – 512 с.

20. *Холодов В.И.* Исследование потоков углерода через организмы гарпактицид *Tigriopus brevicornis* методами математического планирования экспериментов // Экология моря, 1981, вып. 6, с. 29–37.

21. *Холодов В.И., Киселёва М.И.* Статистический анализ влияния различных грунтов на плотность поселения организмов макро- и мейобентоса. Биология моря, 1985, № 2, с. 17–25.

22. *Холодов В.И., Пиркова А.В., Ладыгина Л.В.* Выращивание мидий и устриц в Чёрном море. Севастополь, 2010 – 421 с.

23. *Хикс Ч.Р.* Основные принципы планирования эксперимента. М. Мир, 1967. – 203 с.

24. *Численко Л.Л.* Номограммы для определения веса водных организмов по размерам и форме тела. – Л.: Наука, 1968. – 106 с.

25. *Шефе Г.* Дисперсионный анализ. Перевод с английского Б.А. Севастьянова, В.П. Чистякова. М. Государственное издательство физико-математической литературы, 1963. – 625 с.

26. *Robert K., Gerard A.* Bivalve hatchery technology: The current situation for the Pacific oyster *Crassostrea gigas* and the scallop *Pecten maximus* in France // Aquatic Living Resources. – 1999. – Vol. 12, № 2. – P. 121–130.

СПИСОК РЕКОМЕНДОВАННОЙ ЛИТЕРАТУРЫ

1. *Козлов М.В.* Мнимые повторности (pseudoreplication) в экологических исследованиях: проблема, не замеченная российскими учеными. Журнал общей биологии. 2003 – Т. 64, № 4. – С. 292–307.

2. *Козлов М.В.* Планирование экологических исследований: теория и практические рекомендации. КМК, 2014. – 171 с.

3. *Любищев А.А.* Дисперсионный анализ в биологии. М.: Изд-во Моск. ун-та, 1986. – 200 с.

4. *Розенберг Г.С., Гелашвили Д.Б.* (ред.) Проблемы экологического эксперимента (Планирование и анализ наблюдений). Тольятти: СамНЦ РАН; Кассандра, 2008. – 274 с.

5. *Sheiner S.M., Gurevich J.* (eds). Design and analysis of ecological experiments. NY: Oxford Univ. Press. – 2001. – 413 p.

БЛАГОДАРНОСТИ

Приношу искреннюю благодарность коллегам к.б.н. А.В. Пирковой и к.б.н. Л.В. Ладыгиной за многолетнее научное сотрудничество, совместные экспериментальные исследования, а также за полезные советы и ценные замечания, сделанные при прочтении рукописи.

Благодарю своих многочисленных коллег и друзей, принимавших активное участие в постановке совместных экспериментов и в выполнении полевых исследований с применением методов планирования экспериментов:

д.б.н. М.И. Киселёву, д.б.н. В.В. Мурину, д.б.н. А.А. Калугину-Гутник, д.б.н. В.И. Рябушко, к.б.н. З.П. Бурлакову, д.б.н. Л.И. Рябушко, к.х.н. Л.Н. Пропп, к.б.н. В.П. Парчевского, д.б.н. Нгуен Так Ан, А.Г. Короткова.

Благодарю научных сотрудников Отдела аквакультуры и морской фармакологии ИМБИ РАН за дружескую поддержку, оптимизм, увлечённость исследованиями, что и явилось стимулом для написания данного пособия.

Выражаю благодарность д.б.н. И.В. Довгалю за ценные замечания и советы, которые способствовали улучшению содержания книги.

ПРИЛОЖЕНИЕ I

Критические значения t-критерия Стьюдента
при уровне значимости 0,10; 0,05; 0,01

Число степен свобод. f.	Уровень знач. α			Число степен свобод f.	Уровень знач. α		
	0,10	0,05	0,01		0,10	0,05	0,01
1	6,3138	12,706	63,657	18	1,7341	2,1009	2,8784
2	2.9200	4,3027	9,9248	19	1,7291	2,0930	2,8609
3	2,3534	3,1825	5,8409	20	1,7247	2,0860	2,8453
4	2,1318	2,7764	4,6041	21	1,7207	2,0796	2,8314
5	2,0150	2,5706	4,0321	22	1,7171	2,0739	2,8188
6	1,9432	2,4469	3,7074	23	1,7139	2,0687	2,8073
7	1,8946	2,3646	3,4995	24	1,7109	2,0639	2,7969
8	1,8595	2,3060	3,3554	25	1,7081	2,0595	2,7874
9	1,8331	2,2622	3,2498	26	1,7056	2,0555	2,7787
10	1,8125	2,2281	3,1693	27	1,7033	2,0518	2,7707
11	1,7959	2,2010	3,1058	28	1,7011	2,0484	2,7633
12	1,7823	2,1788	3,0545	29	1,6991	2,0452	2,7564
13	1,7709	2,1604	3,0123	30	1,6973	2,0423	2,7500
14	1,7613	2,1448	2,9768	40	1,6839	2,0211	2,7045
15	1,7530	2,1315	2,9467	60	1,6707	2,0003	2,6603
16	1,7459	2,1199	2,9208	120	1,6577	1,9799	2,6174
17	1,7396	2,1098	2,8982	∞	1,6449	1,9600	2,5758

ПРИЛОЖЕНИЕ II

Таблица значений F-критерия Фишера при уровне значимости α =0,05

k1 / k2	1	2	3	4	5	6	8	12	24	∞
1	161,45	199,50	215,72	224,57	230,17	233,97	238,89	243,91	249,04	254,32
2	18,51	19,00	19,16	19,25	19,30	19,33	19,37	19,41	19,45	19,50
3	10,13	9,55	9,28	9,12	9,01	8,94	8,84	8,74	8,64	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,04	5,91	5,77	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,82	4,68	4,53	4,36
6	5,99	5,14	4,76	4,53	4,39	4,28	4,15	4,00	3,84	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,73	3,57	3,41	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,44	3,28	3,12	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,23	3,07	2,90	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,07	2,91	2,74	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	2,95	2,79	2,61	2,40
12	4,75	3,88	3,49	3,26	3,11	3,00	2,85	2,69	2,50	2,30
13	4,67	3,80	3,41	3,18	3,02	2,92	2,77	2,60	2,42	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,70	2,53	2,35	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,64	2,48	2,29	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,59	2,42	2,24	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,55	2,38	2,19	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,51	2,34	2,15	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,48	2,31	2,11	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,45	2,28	2,08	1,84
21	4,32	3,47	3,07	2,84	2,68	2,57	2,42	2,25	2,05	1,81
22	4,30	3,44	3,05	2,82	2,66	2,55	2,40	2,23	2,03	1,78
23	4,28	3,42	3,03	2,80	2,64	2,53	2,38	2,20	2,00	1,76
24	4,26	3,40	3,01	2,78	2,62	2,51	2,36	2,18	1,98	1,73
25	4,24	3,38	2,99	2,76	2,60	2,49	2,34	2,16	1,96	1,71
26	4,22	3,37	2,98	2,74	2,59	2,47	2,32	2,15	1,95	1,69
27	4,21	3,35	2,96	2,73	2,57	2,46	2,30	2,13	1,93	1,67
28	4,20	3,34	2,95	2,71	2,56	2,44	2,29	2,12	1,91	1,65
29	4,18	3,33	2,93	2,70	2,54	2,43	2,28	2,10	1,90	1,64
30	4,17	3,32	2,92	2,69	2,53	2,42	2,27	2,09	1,89	1,62
35	4,12	3,26	2,87	2,64	2,48	2,37	2,22	2,04	1,83	1,57
40	4,08	3,23	2,84	2,61	2,45	2,34	2,18	2,00	1,79	1,51
45	4,06	3,21	2,81	2,58	2,42	2,31	2,15	1,97	1,76	1,48
50	4,03	3,18	2,79	2,56	2,40	2,29	2,13	1,95	1,74	1,44
60	4,00	3,15	2,76	2,52	2,37	2,25	2,10	1,92	1,70	1,39
70	3,98	3,13	2,74	2,50	2,35	2,23	2,07	1,89	1,67	1,35
80	3,96	3,11	2,72	2,49	2,33	2,21	2,06	1,88	1,65	1,31
90	3,95	3,10	2,71	2,47	2,32	2,20	2,04	1,86	1,64	1,28
100	3,94	3,09	2,70	2,46	2,30	2,19	2,03	1,85	1,63	1,26
125	3,92	3,07	2,68	2,44	2,29	2,17	2,01	1,83	1,60	1,21
150	3,90	3,06	2,66	2,43	2,27	2,16	2,00	1,82	1,59	1,18
200	3,89	3,04	2,65	2,42	2,26	2,14	1,98	1,80	1,57	1,14
300	3,87	3,03	2,64	2,41	2,25	2,13	1,97	1,79	1,55	1,10
400	3,86	3,02	2,63	2,40	2,24	2,12	1,96	1,78	1,54	1,07
500	3,86	3,01	2,62	2,39	2,23	2,11	1,96	1,77	1,54	1,06
1000	3,85	3,00	2,61	2,38	2,22	2,10	1,95	1,76	1,53	1,03
∞	3,84	2,99	2,60	2,37	2,21	2,09	1,94	1,75	1,52	1,00

ПРИЛОЖЕНИЕ III

Ранговый критерий Дункана
Значимые стьюдентизированные ранги для нового множественного рангового критерия при 5% уровне значимости (Хикс, 1967)

№\p	2	3	4	5	6	7	8	9	10	12	14	16	18	20	50	100
1	18,0	18,0	18,0	18,0	18,0	18,0	18,0	18	18	18	18	18	18	18	18	18,0
2	6,09	6,09	6,09	6,09	6,09	6,09	6,09	-	-	-	-	-	-	-	-	6,09
3	4,50	4,50	4,50	4,50	4,50	4,50	4,50	-	-	-	-	-	-	-	-	4,50
4	3,93	4,01	4,02	4,02	4,02	4,02	4,02	-	-	-	-	-	-	-	-	4,02
5	3,64	3,74	3,79	3,83	3,83	3,83	3,83	-	-	-	-	-	-	-	-	3,83
6	3,46	3,58	3,64	3,68	3,68	3,68	3,68	-	-	-	-	-	-	-	-	3,68
7	3,35	3,47	3,54	3,58	3,60	3,61	3,61	-	-	-	-	-	-	-	-	3,61
8	3,26	3,39	3,47	3,52	3,55	3,56	3,56	-	-	-	-	-	-	-	-	3,56
9	3,20	3,34	3,41	3,47	3,50	3,52	3,52	-	-	-	-	-	-	-	-	3,52
10	3,15	3,30	3,37	3,43	3,46	3,47	3,47	-	-	-	-	-	-	-	-	3,48
11	3,11	3,27	3,35	3,39	3,43	3,44	3,45	-	-	-	-	-	-	-	-	3,48
12	3,08	3,23	3,33	3,36	3,40	3,42	3,44	3,46	-	-	-	-	-	-	-	3,48
13	3,06	3,21	3,30	3,35	3,38	3,41	3,42	3,44	-	-	-	-	-	-	-	3,47
14	3,03	3,18	3,27	3,33	3,37	3,39	3,41	3,42	-	-	-	-	-	-	-	3,47
15	3,01	3,16	3,25	3,31	3,36	3,38	3,40	3,42	-	-	-	-	-	-	-	3,47
16	3,00	3,15	3,23	3,30	3,34	3,37	3,59	3,41	-	-	-	-	-	-	-	3,47
17	2,98	3,13	3,22	3,28	3,33	3,36	3,38	3,40	-	-	-	-	-	-	-	3,47
18	2,97	3,12	3,21	3,27	3,32	3,35	3,37	3,39	-	-	-	-	-	-	-	3,47
19	2.96	3,11	3,19	3,26	3,31	3,35	3,37	3,39	-	-	-	-	-	-	-	3,47
20	2,95	3,10	3,18	3,25	3,30	3,34	3,36	3,38	3,4	-	-	-	-	-	-	3,47
22	2,93	3,08	3,17	3,24	3,29	3,32	3,35	3,37	-	-	-	-	-	-	-	3,47
24	2,92	3,07	3,15	3,22	3,28	3,31	3,34	3,37	-	-	-	-	-	-	-	3,47
26	2,91	3,06	3,14	3,21	3,27	3,30	3,34	3,36	-	-	-	-	-	-	-	3,47
28	2,90	3,04	3,13	3,20	3,26	3,30	3,33	3,35	-	-	-	-	-	-	-	3,47
30	2,89	3,04	3,12	3,20	3,25	3,29	3,32	3,35	-	-	-	-	-	-	-	3,47
40	2,86	3,01	3,10	3,17	3,22	3,27	3,30	3,33	-	-	-	-	-	-	-	3,47
50	2,83	2,98	3,08	3,14	3,20	3,24	3,28	3,31	-	-	-	-	-	-	-	3,48
100	2,80	2,95	3,05	3,12	3,18	3,22	3,26	3,29	-	-	-	-	-	-	-	3,53

ПРИЛОЖЕНИЕ IV

Критические значения коэффициента Кохрена (G-критерия) для доверительной вероятности р = 95 % и числе степеней свободы

Число измерений k	Число степеней свободы, ν										
	1	2	3	4	5	6	8	10	16	36	∞
2	9985	9750	9392	9057	8772	8534	8159	7880	7541	6602	5000
3	9669	8709	0797	7454	7071	6771	6333	6025	5466	4748	3333
4	9065	7679	6841	6287	5895	5598	5175	4884	4366	3720	2500
5	8412	6838	5981	5441	5065	4783	4387	4118	3645	3066	2000
6	7808	6161	5321	4803	4447	4184	3817	3568	3135	2612	1667
7	7271	5612	4800	4307	3974	3726	3384	3154	2756	2278	1429
8	6798	5157	4377	3910	3595	3362	3043	2829	2462	2022	1250
9	6385	4775	4027	3584	7276	3067	2768	2568	2226	1820	1111
10	6020	4450	3733	3311	3029	2823	2541	2353	2032	1655	1000
12	5410	3924	3264	2880	2624	2439	2187	2020	1737	1403	0833
15	4709	3346	2758	2419	2195	2034	1815	1671	1429	1144	0667
20	3894	2705	2205	1921	1735	1602	1422	1303	1108	0879	0500
24	3434	2354	1907	1656	1493	1374	1216	1113	0942	0743	0417
30	2929	1980	1593	1377	1237	1137	1001	0921	0771	0604	0333
40	2370	1576	1259	1082	0968	0887	5950	0713	0595	0462	0250
60	1737	1131	0895	0765	0682	0623	0552	0497	0411	0316	0167
120	0998	0632	0495	0419	0371	0337	0292	0266	0218	0165	0083
∞	0000	0000	0000	0000	0000	0000	0000	0000	0000	0000	0000

Все значения G-критерия меньше единицы, поэтому в таблице приведены лишь десятичные знаки, следующие после запятой, перед которой при пользовании таблицей нужно ставить ноль целых.

Например, при k = 6, = 3 имеем $G_{0,95} = 0,5321$

ПРИЛОЖЕНИЕ V

W-критерий Шапиро-Уилкса,
используемого для проверки на нормальность

n\α	1	2	3	10	50
3	0,753	0,756	0,767	0,789	0,959
4	0,687	0,707	0,748	0,792	0,935
5	0,686	0,715	0,762	0,806	0,927
6	0,713	0,743	0,788	0,826	0,927
7	0,730	0,760	0,803	0,838	0,928
8	0,749	0,778	0,816	0,851	0,932
9	0,764	0,791	0,829	0,859	0,935
10	0,781	0,806	0,842	0,869	0,938
11	0,792	0,817	0,850	0,878	0,940
12	0,805	0,828	0,859	0,883	0,943
13	0,814	0,837	0,866	0,889	0,945
14	0,825	0,846	0,874	0,895	0,947
15	0,835	0,855	0,881	0,901	0,950
16	0,844	0,863	0,887	0,905	0,952
17	0,851	0,869	0,892	0,910	0,954
18	0,858	0,874	0,896	0,914	0,956
19	0,863	0,879	0,904	0,917	0,957
20	0,868	0,884	0,905	0,920	0,959
21	0,873	0,888	0,908	0,923	0,960
22	0,878	0,892	0,911	0,926	0,961
23	0,881	0,895	0,914	0,928	0,962
24	0,884	0,898	0,916	0,930	0,963
25	0,888	0,901	0,918	0,931	0,964
26	0,891	0,904	0,920	0,933	0,965
27	0,894	0,906	0,923	0,935	0,965
28	0,896	0,908	0,924	0,936	0,966
29	0,898	0,910	0,926	0,937	0,966
30	0,900	0,912	0,927	0,938	0,967
31	0,902	0,914	0,929	0,940	0,967
32	0,904	0,915	0,930	0,941	0,968
33	0,906	0,917	0,931	0,942	0,968
34	0,908	0,919	0,933	0,943	0,969
35	0,910	0,920	0,934	0,944	0,969
36	0,912	0,922	0,935	0,945	0,970
37	0,914	0,924	0,936	0,946	0,970
38	0,916	0,925	0,938	0,947	0,971
39	0,917	0,927	0,939	0,948	0,971
40	0,919	0,928	0,940	0,949	0,972
41	0,920	0,929	0,941	0,950	0,972
42	0,922	0,930	0,942	0,951	0,972
43	0,923	0,932	0,943	0,951	0,973
44	0,924	0,933	0,944	0,952	0,973
45	0,926	0,934	0,945	0,953	0,973
46	0,927	0,935	0,945	0,953	0,974
47	0,928	0,936	0,946	0,954	0,974
48	0,929	0,937	0,947	0,954	0,974
49	0,929	0,937	0,947	0,955	0,974
50	0,930	0,938	0,947	0,955	0,974

ПРИЛОЖЕНИЕ VI

Коэффициенты a_{n-i+1}, используемые при проверке на нормальность с помощью W-критерия Шапиро-Уилкса

i\n	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	0,707	0,687	0,664	0,643	0,623	0,605	0,589	0,574	0,560	0,548	0,536	0,525	0,515	0,506	0,497	0,489
2	0	0,168	0,241	0,281	0,303	0,316	0,324	0,329	0,331	0,333	0,333	0,332	0,331	0,329	0,327	0,325
3	0	0	0	0,088	0,140	0,174	0,198	0,214	0,226	0,235	0,241	0,246	0,249	0,252	0,254	0,256
4						0,056	0,095	0,122	0,143	0,159	0,171	0,180	0,188	0,194	0,199	0,203
5								0,040	0,069	0,092	0,110	0,124	0,135	0,145	0,152	0,159
6										0,030	0,054	0,073	0,089	0,101	0,111	0,120
7												0,024	0,043	0,059	0,073	0,084
8														0,020	0,036	0,050
9																0,016
i/n	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34
1	0,481	0,473	0,464	0,459	0,454	0,449	0,445	0,441	0,437	0,433	0,429	0,425	0,422	0,419	0,416	0,413
2	0,323	0,321	0,319	0,316	0,313	0,310	0,307	0,304	0,302	0,299	0,297	0,294	0,292	0,290	0,288	0,285
3	0,256	0,257	0,258	0,257	0,256	0,255	0,254	0,253	0,252	0,251	0,250	0,249	0,248	0,246	0,245	0,244
4	0,206	0,209	0,212	0,213	0,214	0,215	0,215	0,215	0,215	0,215	0,215	0,215	0,215	0,214	0,214	0,213
5	0,164	0,169	0,174	0,176	0,179	0,181	0,182	0,184	0,185	0,186	0,186	0,187	0,187	0,188	0,188	0,188
6	0,127	0,133	0,140	0,144	0,148	0,151	0,154	0,156	0,158	0,160	0,162	0,163	0,164	0,165	0,166	0,167
7	0,093	0,101	0,109	0,115	0,120	0,125	0,128	0,132	0,135	0,137	0,140	0,142	0,143	0,145	0,146	0,148
8	0,061	0,071	0,080	0,088	0,094	0,100	0,105	0,109	0,113	0,116	0,119	0,122	0,124	0,127	0,128	0,130
9	0,030	0,042	0,053	0,062	0,070	0,076	0,082	0,088	0,092	0,097	0,100	0,104	0,107	0,109	0,112	0,114
10	0	0,015	0,026	0,037	0,046	0,054	0,061	0,067	0,073	0,078	0,082	0,086	0,090	0,093	0,096	0,099
11	0	0	0	0,012	0,023	0,032	0,040	0,048	0,054	0,060	0,065	0,070	0,074	0,078	0,081	0,084
12						0,011	0,020	0,028	0,036	0,042	0,048	0,054	0,059	0,063	0,067	0,071
13								0,009	0,018	0,025	0,032	0,038	0,044	0,049	0,053	0,057
14										0,008	0,016	0,023	0,029	0,034	0,040	0,044
15												0,008	0,014	0,021	0,026	0,031
16														0,007	0,013	0,019
17																0,006
i/n	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50
1	0,410	0,407	0,404	0,402	0,399	0,396	0,394	0,392	0,389	0,387	0,385	0,383	0,381	0,379	0,377	0,375
2	0,283	0,281	0,279	0,277	0,275	0,274	0,272	0,270	0,268	0,267	0,265	0,264	0,262	0,260	0,259	0,257
3	0,243	0,242	0,240	0,239	0,238	0,237	0,236	0,235	0,233	0,232	0,231	0,230	0,229	0,228	0,227	0,226
4	0,213	0,212	0,212	0,211	0,210	0,210	0,209	0,209	0,208	0,207	0,207	0,206	0,205	0,205	0,204	0,203
5	0,188	0,188	0,188	0,188	0,188	0,188	0,188	0,187	0,187	0,187	0,187	0,186	0,186	0,186	0,185	0,185
6	0,167	0,168	0,168	0,169	0,169	0,169	0,169	0,169	0,170	0,170	0,170	0,170	0,170	0,169	0,169	0,169
7	0,149	0,150	0,151	0,151	0,152	0,153	0,153	0,154	0,154	0,154	0,155	0,155	0,155	0,155	0,155	0,155
8	0,132	0,133	0,134	0,136	0,137	0,138	0,138	0,139	0,140	0,141	0,141	0,142	0,142	0,142	0,143	0,143
9	0,117	0,118	0,120	0,121	0,123	0,125	0,126	0,126	0,127	0,128	0,129	0,129	0,130	0,131	0,131	0,132
10	0,101	0,104	0,106	0,108	0,109	0,111	0,112	0,114	0,115	0,116	0,117	0,118	0,119	0,120	0,121	0,122
11	0,087	0,090	0,092	0,095	0,097	0,099	0,100	0,102	0,104	0,105	0,106	0,107	0,109	0,110	0,111	0,111
12	0,074	0,077	0,080	0,082	0,085	0,087	0,089	0,091	0,093	0,094	0,096	0,097	0,099	0,100	0,101	0,102
13	0,061	0,065	0,068	0,071	0,073	0,076	0,078	0,080	0,082	0,084	0,086	0,088	0,089	0,091	0,092	0,093
14	0,048	0,052	0,056	0,059	0,062	0,065	0,068	0,070	0,072	0,075	0,077	0,078	0,080	0,082	0,083	0,085
15	0,036	0,040	0,044	0,048	0,052	0,055	0,058	0,060	0,063	0,065	0,067	0,069	0,071	0,072	0,075	0,076
16	0,024	0,029	0,033	0,037	0,041	0,044	0,048	0,051	0,053	0,056	0,058	0,061	0,063	0,065	0,067	0,069
17	0,012	0,017	0,022	0,026	0,031	0,034	0,038	0,041	0,044	0,047	0,050	0,052	0,055	0,057	0,059	0,061
18	0	0,006	0,011	0,016	0,020	0,024	0,028	0,032	0,035	0,038	0,041	0,044	0,047	0,049	0,051	0,053
19	0	0	0	0,005	0,010	0,015	0,019	0,023	0,026	0,030	0,033	0,036	0,039	0,041	0,044	0,046
20	0					0,005	0,009	0,014	0,018	0,021	0,025	0,028	0,031	0,034	0,036	0,039
21	0							0,005	0,009	0,013	0,016	0,020	0,023	0,026	0,029	0,031
22	0									0,004	0,008	0,012	0,015	0,019	0,022	0,024

ПРИЛОЖЕНИЕ VII

Матрицы планирования экспериментов по методу
случайного баланса (Гинберг и др., 1972)
а) число факторов k=19; число строк g=32

№	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}	x_{17}	x_{18}	x_{19}
1	-	-	+	+	+	-	+	+	+	-	+	-	-	-	-	-	-	-	-
2	+	-	-	+	+	-	+	+	-	-	+	+	-	+	+	-	+	+	-
3	-	+	-	+	+	+	-	+	+	+	+	-	+	-	-	-	+	+	+
4	+	+	+	+	+	-	+	-	+	-	+	+	-	+	+	+	-	+	-
5	-	-	-	-	+	+	-	+	+	+	-	-	-	+	+	+	+	-	+
6	+	-	+	-	+	-	+	-	-	-	+	-	-	+	+	-	-	+	+
7	-	+	+	-	+	-	+	+	-	-	-	+	+	-	-	-	+	+	-
8	+	+	-	-	+	+	-	-	-	-	+	-	+	-	+	+	+	-	-
9	-	-	+	-	-	-	-	+	+	+	-	-	+	-	+	+	+	-	+
10	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+
11	-	+	-	-	-	+	+	+	+	+	+	+	-	+	+	-	-	+	-
12	+	+	+	-	-	-	-	+	-	+	+	-	+	+	-	+	-	-	-
13	-	-	-	+	-	+	+	-	+	-	-	+	+	+	-	+	-	+	++
14	+	-	+	+	-	-	-	-	+	+	+	-	-	-	-	+	+	-	-
15	-	+	+	+	-	+	+	+	-	+	+	+	-	-	+	+	-	+	+
16	+	+	-	+	-	+	+	+	+	+	-	+	-	-	+	-	+	-	+
17	-	-	+	+	-	+	-	+	-	-	+	-	+	+	-	-	+	-	-
18	+	-	-	+	-	+	-	-	-	+	+	+	+	+	+	-	-	-	-
19	-	+	-	+	-	+	-	-	-	-	+	+	-	-	+	-	+	+	+
20	+	+	+	+	-	-	+	-	+	-	-	+	-	-	+	+	-	-	-
21	-	-	-	-	-	-	+	-	-	-	-	-	+	-	-	+	+	+	-
22	+	-	+	-	-	+	-	+	-	-	-	-	-	+	-	-	-	+	-
23	-	+	+	-	-	+	-	-	-	+	-	-	-	-	+	-	+	+	+
24	+	+	-	-	-	+	+	+	-	+	-	+	+	+	+	-	-	+	-
25	-	-	+	-	+	-	+	+	-	+	+	-	-	+	-	+	+	-	-
26	+	-	-	-	+	-	+	+	+	-	+	+	+	+	-	-	+	-	-
27	-	+	-	-	+	+	-	-	+	+	-	+	-	+	-	+	+	+	+
28	+	+	+	-	+	-	+	-	-	-	-	-	+	+	-	-	-	+	+
29	-	-	-	+	+	+	-	-	+	+	+	+	+	-	-	-	-	-	+
30	+	-	+	+	+	-	-	+	+	+	-	-	+	+	-	+	-	-	+
31	-	+	+	+	+	-	-	-	+	-	-	+	-	+	+	+	-	+	-
32	+	+	-	+	+	+	+	-	+	+	-	+	+	-	+	+	+	-	+

Примечание: для столбцов 1-5 $r = 0$;
для столбцов 1-19 $r \leq 0{,}51$

б) число факторов k=12; число строк g=24

№	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}
1	-	+	-	+	+	+	+	-	+	+	+	+
2	-	-	+	-	-	-	-	-	+	+	+	-
3	-	+	+	-	+	-	-	-	-	-	+	+
4	+	-	-	-	+	+	-	-	-	+	-	-
5	+	-	-	+	-	-	-	+	-	+	+	+
6	-	+	-	-	+	-	-	-	+	-	-	-
7	+	+	-	-	-	+	+	-	+	-	-	+
8	-	-	+	+	+	-	-	+	+	+	-	-
9	-	-	+	+	-	-	+	+	+	-	+	+
10	-	+	+	-	-	-	+	+	+	-	+	+
11	+	+	-	+	+	+	-	-	-	+	-	+
12	-	-	-	-	-	+	+	+	-	+	+	-
13	+	+	+	+	+	+	-	-	-	+	-	+
14	-	-	-	+	-	+	+	-	-	-	-	-
15	-	-	-	+	-	+	-	+	+	-	-	+
16	+	-	-	-	+	-	-	-	+	+	+	-
17	+	+	+	+	-	+	-	+	+	+	-	+
18	+	-	+	-	+	-	+	+	-	-	-	-
19	+	-	+	-	-	-	+	+	-	-	+	+
20	-	+	-	+	-	+	+	+	+	+	-	-
21	-	+	+	-	+	+	+	-	-	-	+	+
22	+	-	+	+	+	-	-	-	+	-	+	-
23	+	+	+	+	-	+	+	+	-	-	-	+
24	+	+	-	-	+	-	+	+	-	-	+	-

Примечание: для столбцов 1-4 $r = 0$;
для столбцов 1-7 $r \leq 0{,}50$;
для столбцов 1-2 $r \leq 0{,}542$

в) число факторов k=14; число строк g=16

№	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}
1	-	-	+	+	+	-	+	-	+	-	+	+	-	+
2	+	-	-	-	-	-	-	-	-	+	-	-	-	+
3	-	+	-	+	+	-	-	+	+	+	+	-	+	-
4	+	+	+	+	+	+	-	-	+	+	-	+	-	+
5	-	-	-	-	-	+	+	+	-	-	+	-	-	-
6	+	-	+	-	+	+	+	-	+	+	+	-	+	+
7	-	+	+	-	-	+	+	-	+	+	-	+	+	-
8	+	+	-	-	-	-	-	+	-	-	+	-	-	-
9	-	-	+	-	-	+	-	+	-	+	+	+	+	+
10	+	-	-	-	-	+	-	-	-	-	-	-	+	+
11	-	+	-	-	+	-	+	-	+	-	-	-	-	-
12	+	+	+	-	+	-	-	-	-	-	-	-	+	-
13	-	-	-	+	+	+	-	+	-	+	+	+	+	-
14	+	-	+	+	-	-	+	+	+	-	+	+	-	-
15	-	+	+	+	-	-	+	+	+	+	-	+	+	+
16	+	+	-	+	+	+	+	+	-	-	-	+	-	+

Примечание: для столбцов 1-4 $r = 0$;
для столбцов 1-10 $r \leq 0,517$;
для столбцов 1-14 $r \leq 0,700$

Знак информационной продукции **16+**

Научное издание

Валентин Иванович Холодов

**ПЛАНИРОВАНИЕ ЭКСПЕРИМЕНТОВ
В ГИДРОБИОЛОГИЧЕСКИХ ИССЛЕДОВАНИЯХ**

Макет и компьютерная вёрстка *А.В.Пинчук*
Художественный редактор *Н.В.Дымникова*

«Н.Оріанда»™
ИП Пинчук А.В. 23 №008849391 от 15 ноября 2014 г.

Подписано к печати с оригинал-макета 2.12.2016.
Формат 70x100 $^{1}/_{16}$. Гарнитура «Петербург». Печать офсетная.
Усл. печ. л. 16,3. Тираж 300 экз. Заказ № .

«Н.Оріанда»™ 295051, Россия, Республика Крым,
г. Симферополь, ул. Калинина, 25
Тел.: 8 (3652) 60-49-19, 8 (978) 726-76-68
E-mail: n-orianda@mail.ru, http://n-orianda.ru

Полиграфическое исполнение:
ООО «Фирма «Салта» ЛТД»
Россия, Республика Крым, г. Симферополь, ул. Коммунальная, 24/3
Тел.: 8 (3652) 24-84-72. www.saltaprint.com