

Министерство сельского хозяйства Российской Федерации  
ФГБОУ ВО «Красноярский государственный аграрный университет»

*И.А. Шадрин, Н.В. Фомина*

## **МЕТОДЫ ЭКОЛОГИЧЕСКИХ ИССЛЕДОВАНИЙ**

*Методические указания по выполнению самостоятельных работ*

*Электронное издание*

Красноярск 2016

*Рецензент*

*И.А. Чаплыгина, канд. биол. наук, доцент*

**Шадрин, И.А.**

**Методы экологических исследований:** методические указания по выполнению самостоятельных работ / И.А. Шадрин, Н.В. Фомина; Краснояр. гос. аграр. ун-т. – Красноярск, 2016. – 84 с.

Представлены темы лабораторных и самостоятельных работ по курсу «Методы экологических исследований» и рекомендации по их выполнению, библиографический список.

Предназначено для студентов всех форм обучения по направлению подготовки 35.04.03 (110100.62) «Агрохимия и агропочвоведение».

Печатается по решению редакционно-издательского совета  
Красноярского государственного аграрного университета

© Шадрин И.А., Фомина Н.В., 2016  
© ФГБОУ ВО «Красноярский государственный  
аграрный университет», 2016

## ОГЛАВЛЕНИЕ

|   |    |
|---|----|
| Введение.....   | 4  |
| Работа № 1. Методика планирования экспериментов. Методики первичной обработки информации.....   | 5  |
| Работа № 2. Оценка степени адекватности модели, абсолютной и относительной погрешности.....   | 10 |
| Работа № 3. Корреляционно-регрессивный анализ. Анализ данных, измерения, расчёты.....   | 18 |
| Работа № 4. Многофакторный дисперсионный анализ.....  | 25 |
| Работа № 5. Построение модели видовой структуры биоценоза и её исследование (базисная модель двувидового леса с возобновляемым ресурсом)..... | 30 |
| Работа № 6. Исследование процессов динамики экологических систем.....   | 32 |
| Работа № 7. Многомерные модели на примере кластерного анализа   | 35 |
| Работа № 8. Динамические статистические модели.....   | 39 |
| Работа № 9. Матричные модели.....   | 48 |
| Работа № 10. Пространственные распределения организмов....  | 53 |
| Работа № 11. Регрессионный анализ.....  | 61 |
| Работа № 12. Многофакторные модели.....   | 66 |
| Работа № 13. Оптимизационные модели.....  | 74 |
| Работа № 14. Теоретико-игровые модели.....  | 78 |
| Литература.....   | 83 |

## ВВЕДЕНИЕ

Математическая статистика в экологии – интенсивно развивающаяся область научной деятельности, которая является одним из наиболее результативных проявлений интегративных тенденций в экологической науке.

Цель преподавания данного раздела дисциплины заключается в обучении студентов основам знаний и методов математической статистики для использования возможностей моделирования при решении экологических задач.

**Студент по окончании обучения должен:**

***знать***

- основы теории систем и терминологии;
- основные методы математической статистики и системного анализа экосистем;

***уметь***

- грамотно формулировать задачи и вести совместную работу с математиками и другими специалистами при решении сложных задач;
- применять навыки математического моделирования экосистем.

## РАБОТА № 1. МЕТОДИКА ПЛАНИРОВАНИЯ ЭКСПЕРИМЕНТОВ. МЕТОДИКИ ПЕРВИЧНОЙ ОБРАБОТКИ ИНФОРМАЦИИ

Научные исследования и разработки, осуществляемые методом полевого эксперимента, включают три основных этапа: планирование; проведение полевых опытов, наблюдений и учетов; обработку и обобщение полученных данных.

*Планирование – это определение задачи и объектов исследования, разработка схемы эксперимента, выбор земельного участка и оптимальной структуры полевого опыта.*

Период, предшествующий исследованию, включает: выбор темы, определение задачи и объекта исследования; изучение современного состояния вопроса; выдвижение рабочей гипотезы или ряда конкурирующих гипотез; разработку схемы и методики эксперимента.

Необходимо четко сформулировать цель исследования, построить логическую модель изучаемого явления и правильно выбрать стратегию, которая определяет методы и приемы исследования.

Следующий этап планирования – изучение литературы по избранной проблеме и выдвижение рабочей гипотезы или ряда конкурирующих гипотез. Рабочая гипотеза служит отправным пунктом для составления схемы или ряда схем будущих опытов и разработки программы исследования. В программе указывают схемы опытов, основные элементы методики и техники эксперимента, наблюдения и учеты.

Сложным и ответственным этапом планирования является разработка схемы и методики опыта, выбор полевых и лабораторных наблюдений (анализов) и учетов для оценки и объяснения действия изучаемых факторов. Надежность результатов эксперимента и соответствие их поставленной задаче зависят от правильного решения основного вопроса планирования-разработки рациональной схемы полевого опыта.

Особое внимание при планировании следует обратить на правильное сочетание основных элементов методики и в зависимости от целей исследования, схемы опыта, земельного участка и технических возможностей установить наиболее рациональное направле-

ние, форму и площадь делянки, повторность, систему расположения повторений, делянок и вариантов.

**Задание 1.** Приближенные оценки основных статистических показателей.

**Материалы и оборудование.** Расчетные таблицы, рабочий материал, калькуляторы.

**Пояснение.** Классические работы Р. Фишера открыли новую страницу в истории биометрии: они показали, что планирование экспериментов и обработка их результатов – это две тесно связанные между собой задачи статистического анализа. Это открытие легло в основу разработки теории планирования экспериментов, которая в настоящее время находит применение не только при проведении сельскохозяйственных опытов, на базе которых она возникла, но и в различных областях биологии, медицины, антропологии, в сфере других научно-практических дисциплин, включая и социально-экономические исследования.

Планирование экспериментов стало самостоятельным разделом биометрии, которому посвящена огромная литература. В начальном курсе биометрии невозможно осветить все аспекты теории экспериментов. Здесь будут рассмотрены лишь некоторые общие положения, относящиеся к этой сложной и многогранной проблеме.

Термин «эксперимент» (от лат. *experimentum* – опыт) означает искусственно организуемый комплекс условий, в которых испытывают воздействие того или иного фактора или одновременно нескольких факторов на результативный признак. В земледелии это полевые опыты; в животноводстве – опыты по кормлению животных, по уходу за ними; в педагогике – опыты по проверке новых методов обучения и воспитания учащихся; в фармакологии – испытание эффективности новых лечебных препаратов; в медицине – проверка разных способов лечения больных и т. д.

Исследовательская работа сводится не только к экспериментам, ее проводят и вне их на основе непосредственных наблюдений. Так что выражение «планирование исследований» оказывается более емким, а следовательно, и более подходящим, чем введенный Р. Фишером (1930) термин «планирование экспериментов». Конечно, и термин «эксперимент» можно применять в более широком смысле, понимая под ним любые испытания, проводимые исследователем в отношении изучаемого объекта. При всем разнообразии методов исследовательской работы задача планирования сводится к тому, что-

бы при возможно минимальных объемах наблюдений получать достаточно полную информацию об изучаемых объектах.

С варьированием результатов наблюдений связана *повторность вариантов опыта*, позволяющая повысить точность оценок генеральных параметров, надежность выводов, которые делает исследователь на основании выборочных показателей. Под повторностью в полевом опыте понимают число одноименных делянок для каждого варианта опыта. В лабораторных условиях повторность может выражаться числом одинаковых проб серий одновременных испытаний, измерений и т. п. повторений одного и того же варианта опыта. Очевидно, чем шире диапазон варьирования признака, тем больше должна быть и повторность опыта, и, наоборот, при слабом варьировании учитываемого признака число вариантов опыта, т. е. их повторность, уменьшается. В такой же зависимости от размаха варьирования признаков находится и организация планирования минимально допустимого числа испытаний.

Прежде чем наметить необходимый объем выборки, надо определить *среднюю величину и ее ошибку* для варьирующего признака – характеристики, которые позволяют использовать показатель точности выборочной средней при решении этой задачи. Приближенное значение средней арифметической  $x$  можно определить по полусумме лимитов:

$$X_{cp} = (x_{min} + x_{max})/K,$$

а среднее квадратическое отклонение  $s_x$  – по разности лимитов, отнесенной к коэффициенту  $K$ , который устанавливают в зависимости от объема выборки ( $n$ ) с помощью формулы

$$S_x = (x_{min} - x_{max})/K.$$

**Задача 1.** Зная лимиты  $x_{min} = 9,0$  мг% и  $x_{max} = 14,7$  мг% кальция в сыворотке крови обследованной группы обезьян ( $n = 100$ ), определите основные характеристики для этой выборки.

Величину ошибки средней  $Sx'$  можно определить по следующей приближенной формуле:

$$Sx' = (x_{min} - x_{max})/K\sqrt{N}.$$

Отсюда возможно рассчитать показатель точности  $C_s$  выборочной средней  $X_{ср}$ :

$$C_s = 100 \cdot S_{x'} / X_{ср}.$$

Намечаемый таким образом объем выборки можно считать вполне достаточным для получения надежных оценок генеральных параметров (при условии, что совокупность, из которой взята выборка, распределяется по нормальному закону).

**Выводы.** Провести приближенную оценку основных статистических показателей. Записать результаты и объяснить их.

**Задание 2.** Определение необходимого объема выборки.

**Материалы и оборудование.** Расчетные таблицы, рабочий материал, калькуляторы.

**Пояснение.** Элементарная логика и практический опыт подсказывают, что неразумно стремиться к неоправданно большому числу испытаний, если убедительный результат можно получить при минимально допустимом объеме выборки. Необходимая численность выборки  $n$ , отвечающая точности, с которой намечено получить средний результат, зависит от величины ошибки выборочной средней и определяется по формуле

$$n = t^2 S_x^2 / \Delta^2,$$

где  $t$  – нормированное отклонение, с которым связан тот или иной уровень значимости ( $\alpha$ );  $S_x^2$  – выборочная дисперсия;  $\Delta = t s_x'$  – величина, определяющая границы доверительного интервала (здесь  $s_x' = \sqrt{S_x^2/n}$  – ошибка выборочной средней);  $K = \Delta/S_x$ .

**Задача 1.** Случайная выборка девяти вариантов характеризуется средней  $\bar{x} = 12,1 \pm 0,68$ . Точность выборочной средней оказалась недостаточно высокой:  $C_s = 5,62$ .

Какое число испытаний  $n$  нужно провести, чтобы ошибку средней уменьшить вдвое?

**Задача 2.** При определении необходимого объема выборки для получения статистически достоверной разности между средними  $(x_1 - x_2) = d$  применяют формулу

$$n = (t/\Delta)^2 \cdot (S_1^2/\alpha + S_2^2).$$



Здесь  $\Delta = t \cdot sd$ , где  $sd$  – заданная величина ошибки для разности сравниваемых средних;  $S_1^2$  и  $S_2^2$  – дисперсии для сравниваемых выборок, причем  $S_1^2$  – дисперсия для большей выборки;  $a = n_1/n_2$  – отношение объема большей выборки к объему меньшей выборки. При  $n_1 = n_2$  формула принимает следующий вид:

$$n = (t/\Delta)^2 \cdot (S_1^2 + S_2^2).$$

Изучали влияние лечебного препарата на массу тела лабораторных мышей. Были получены следующие результаты. Характеристики опытной группы ( $n = 9$ ):  $x_1 = 74,1$  г;  $S_1^2 = 37,86$ ; контрольной группы ( $n_2 = 11$ ):  $x_2 = 68,8$  г.;  $S_2^2 = 44,36$ .

Разность между  $x_1$  и  $x_2$ , равная  $5,3 \pm 2,89$ , оказалась статистически недостоверной. Определите число наблюдений  $n$ , которое необходимо провести при уменьшении ошибки разности вдвое, т. е.  $sd = 2,87/2 = 1,445$ . Примем  $t = 2$ .

**Задача 3.** При альтернативной группировке данных, когда численность выборочных групп выражают в долях единицы, планируемый объем наблюдений определяют по формуле

$$n = t^2 p(1 - p) / \Delta^2,$$

где  $p$  – доля вариантов, обладающих данным признаком;  $\Delta = t \cdot s_p$ .

Если доли выражают в процентах от общего числа наблюдений, формула принимает следующий вид:

$$n = t^2 p(100 - p) / \Delta^2.$$

По предварительным данным, число гельминтоносителей среди лиц, проживающих в  $N$ -м населенном пункте, равно 8%. Определить необходимое число наблюдений, при котором величина максимальной ошибки  $\Delta$  не превысит 4% для уровня значимости, равного 0,05, и соответственно  $t = 2$ .

**Выводы.** Определить необходимый объем выборки. Записать результаты и объяснить их.

**Задача 4.** Занесите в словарь новые понятия и обоснуйте обязательность определения необходимого объема выборки.

## РАБОТА № 2. ОЦЕНКА СТЕПЕНИ АДЕКВАТНОСТИ МОДЕЛИ, АБСОЛЮТНОЙ И ОТНОСИТЕЛЬНОЙ ПОГРЕШНОСТИ

**Материалы и оборудование.** Расчетные таблицы, рабочий материал, калькуляторы.

**Пояснение.** К количественным относят признаки, которые могут быть охарактеризованы количественно, – урожай с делянки, число, высота и масса растений, содержание белка и клейковины в зерне и т. д. Различают два вида количественной изменчивости: непрерывную и прерывистую или дискретную. В первом случае значения признака выражены мерами объема, длины, массы и т. д., во втором различия между единицами наблюдения выражаются целыми числами, между которыми нет и не может быть переходов, например число зерен в колосе и т. д.

Выборки, состоящие из 20–30 единиц наблюдения, называют малыми, а выборки большего объема – большими. После изучения выборочная совокупность представляет собой ряд варьирующих значений признака, записанных в той последовательности, в какой они были получены. Статистические характеристики вычисляются по формулам следующей таблицы.

Таблица 1

| Показатель                                   | Малые выборки,<br>несгруппированные<br>данные | Большие выборки,<br>сгруппированные данные |
|--|---|--|
| Средняя арифметическая                       | $X_{cp} = \sum x/n$                           | $X_{cp} = \sum fx/n$                       |
| Дисперсия                                    | $S^2 = \sum (x - X_{cp})^2 / (n - 1)$         | $S^2 = \sum f(x - X_{cp})^2 / (n - 1)$     |
| Стандартное отклонение                       | $s = \sqrt{S^2}$                              |  |
| Коэффициент вариации                         | $V = (s / X_{cp}) \cdot 100$                  |  |
| Ошибка средней                               | $S_x = s / \sqrt{n}$                          |  |
| Относительная ошибка средней                 | $S_{x\%} = (S_x / X_{cp}) \cdot 100$          |  |
| Доверительный интервал для среднего значения | $X_{cp} \pm tS_x$                             |  |
| Степень свободы                              | $n - 1$                                       |  |

В таблице 1 через  $X$  обозначены отдельные значения признака в малых выборках и групповые средние в больших выборках;  $f$  – частота, численность группы;  $n$  – объем выборки;  $t$  – теоретическое значение критерия Стьюдента.

Для вычисления средней арифметической и суммы квадратов (числитель дисперсии) в таблице дано несколько формул. Все они дают практически одинаковые результаты.

При вычислениях исходные даты целесообразно преобразовать так, чтобы отбросить лишние цифры и опустить запятые. Последние потом вновь восстанавливают. Преобразование (кодирование) может осуществляться вычитанием от результатов измерений одного и того же числа  $A$ , умножением или делением исходных дат на одно и то же число  $K$ , а также одновременным проведением двух действий.

При работе с преобразованными (закодированными) датами необходимо иметь в виду, что вычитание или прибавление условной средней, т. е. изменение начала отсчета, не оказывает влияния на сумму квадратов и поправка необходима лишь при определении средней арифметической.

**Задание 1.** Расчет основных характеристик варьирующих объектов (малые выборки, несгруппированные данные).

**Задача 1.** При определении содержания фосфора в растительном материале получены следующие результаты (в г  $P_2O_5$  на 100 г сухого вещества): 0,55; 0,63; 0,49; 0,57; 0,50; 0,53; 0,57. Необходимо вычислить  $X$ ,  $S^2$ ,  $S$ ,  $V$ ,  $S_x$ ,  $S_x \%$ .

**Выводы.** Занесите в словарь новые понятия и правила анализа малых выборок. Занесите результаты в таблицу 2.

Таблица 2

| $X$ | $X_{cp}$ | $S^2$ | $S$ | $V$ | $S_x$ | $S_x \%$ |
|-----|----------|-------|-----|-----|-------|----------|
|     |          |       |     |     |       |          |

**Задача 2.** В вегетационном опыте получены урожаи томатов по параллельным сосудам (г/сосуд): 578; 564; 539; 604; 551; 468. Необходимо вычислить  $X$ ,  $S^2$ ,  $S$ ,  $V$ ,  $S_x$ ,  $S_x \%$  и 95% доверительный интервал для среднего значения совокупности ( $X \pm t S_x$ ).

Занесите в словарь новые понятия и правила анализа малых выборок.

**Выводы.** Занесите результаты в таблицу 3.

Таблица 3

| X | $X_{cp}$ | $S^2$ | S | V | $S_x$ | $S_x \%$ | $X \pm t S_x$ |
|---|----------|-------|---|---|-------|----------|---------------|
|   |          |       |   |   |       |          |               |

**Задание 2.** Вычисление статистических характеристик выборки при количественной изменчивости признака (большие выборки, сгруппированные данные).

**Материалы и оборудование.** Расчетные таблицы, рабочий материал, калькуляторы.

**Пояснение.** При большом числе исходных наблюдений результаты необходимо представить в виде систематизированного вариационного ряда. Систематизация сводится к распределению отдельных значений по группам или классам. Число групп зависит от объема выборки: при 30–60 наблюдениях рекомендуется выделить 6–7 групп, при 60–100 наблюдениях – 7–8, а если число наблюдений более 100, то выделяют 8–15 групп. Ориентировочно число групп равно корню квадратному из общего числа наблюдений.

Необходимо иметь в виду, что выделением неоправданно большого числа групп можно затушевать общую картину распределения случайными отклонениями, а если взять слишком мало групп (меньше 5–6), то нельзя выяснить характерную особенность распределения изучаемого признака в совокупности.

После установления числа групп необходимо определить величину интервала, верхнюю и нижнюю границу каждой группы, групповые или средние значения вариантов и частоты.

Величину интервала, т. е. промежутки, на которые разбивается ряд варьирующих признаков, определяют по формуле

$$i = (X_{max} - X_{min})/k.$$

При интервальной группировке допускают, что в каждом интервале включены варианты, имеющие одинаковое значение варьирующего признака, равное центральному значению каждой группы.

На правильное определение интервала для групп (классов) следует обратить серьезное внимание. Величина промежутка между границами соседних групп должна быть всегда одной и той же, границы групп необходимо наметить так, чтобы одно и то же значение не по-

вторялось в двух классах. Конец каждой группы должен быть меньше начала следующей на величину, равную принятой точности измерения. Если, например, первая группа заканчивается величиной 60, то следующая должна начинаться цифрой 61, а если первая группа заканчивается цифрой 60,5, то следующая должна начинаться цифрой 60,6 и т. д. Не обязательно за начало первой группы брать минимальное значение признака. Лучше за начало принять целое число с таким расчетом, чтобы минимальная варианта попала примерно в середину первого класса.

Например, если установлено значение интервала  $i = 10$ , а признак варьирует от 45,4 до 115,2, то начала групп можно установить следующие: 40, 50, 60, ..., 100.

При непрерывной изменчивости срединные или групповые значения вариант устанавливаются прибавлением к началу каждой группы половины интервала. В нашем примере для первой группы срединное значение равно  $40 + y = 45$ , для второй  $50 + y = 55$  и т. д.

Иногда удобнее установить сначала групповые варианты, а затем определить границы классов. Начало группы находят вычитанием от групповой варианты половины значения интервала, а конец – прибавлением половины интервала, уменьшенного на величину, равную точности измерения. Например, если установлены групповые варианты, равные 45, 55, 65 и т. д., то при  $i = 10$ , то началами групп будут соответственно 40; 50 и т. д., а концами групп 49,9; 59,9 и т. д.

Частоту встречаемости признака в каждой группе устанавливают путем разности исходных дат по классам. Чтобы избежать ошибок и сэкономить время при распределении вариант по группам, рекомендуется не искать одинаковые варианты совокупности, а разносить их подряд по группам, что не одно и то же. Для разности целесообразно пользоваться одним из следующих способов.

*Способ «штрихов».* В исходных данных зачеркивают первую дату и заносят ее в соответствующую строчку (группу) рабочей таблицы, отмечая вертикальной чертой. Затем зачеркивают вторую дату и также переносят ее в таблицу.

В таблице первые четыре даты в каждом классе отмечают вертикальными черточками, а пятую – в виде диагонали.

*Способ «конвертиков».* Первые четыре даты в каждой группе изображают точками по углам квадрата; следующие четыре даты (с 5-й по 8-ю) отмечают в виде сторон квадрата, соединяющих ранее нанесенные точки, 9-ю и 10-ю даты – в виде диагоналей.

Таким образом, каждый десяток отмечают в виде «конвертика».

Сумма частот всех групп должна быть равна объему выборки  $n$ . Правильность разности проверяют повторным составлением рабочей таблицы. После определения групповых вариантов и разности дат по группам непрерывный вариационный ряд будет трансформирован в прерывистый или дискретный. При этом исходные даты, попав в соответствующие группы, приравниваются по величине к групповым средним значениям, которые и используются для расчетов средней арифметической и других показателей. Такая трансформация непрерывного ряда в прерывистый связана с потерей части информации, и поэтому метод расчета статистических характеристик по сгруппированным данным не является абсолютно точным. Однако для больших выборок погрешности метода незначительны и ими можно пренебречь.

Чтобы наглядно представить закономерность распределения изучаемого признака в совокупности, вариационные ряды принято изображать графически в виде ступенчатого графика-гистограммы или полигона – ломаной линией, соединяющей средние значения групп. Графическое изображение вариационного ряда называется кривой распределения.

**Задача 1.** При измерении высоты стебля льна (см) получены следующие результаты:

|      |       |       |       |       |      |      |      |       |       |
|------|-------|-------|-------|-------|------|------|------|-------|-------|
| 90,1 | 109,9 | 99,1  | 100,1 | 115,3 | 68,0 | 70,4 | 72,3 | 73,0  | 70,1  |
| 76,2 | 82,2  | 80,0  | 68,4  | 69,4  | 74,0 | 72,2 | 69,4 | 80,0  | 59,2  |
| 79,9 | 81,4  | 84,0  | 108,2 | 83,3  | 81,7 | 99,4 | 98,0 | 102,4 | 101,7 |
| 45,4 | 59,1  | 60,1  | 63,3  | 78,2  | 87,0 | 94,7 | 91,5 | 88,2  | 90,1  |
| 72,4 | 68,5  | 80,7  | 81,2  | 84,4  | 77,0 | 79,8 | 81,6 | 84,3  | 50,2  |
| 70,7 | 67,0  | 100,4 | 103,4 | 69    | 72,4 | 74,4 | 66,1 | 67,3  | 52,0  |
| 79,1 | 78,0  | 83,9  | 92,2  | 93,2  | 81,3 | 82,0 | 86,4 | 89,1  | 93,5  |
| 77,0 | 76,1  | 88,1  | 89,7  | 94,1  | 82,0 | 80,1 | 81,0 | 77,0  | 80,0  |
| 92,1 | 91,5  | 93,1  | 79,0  | 73,5  | 84,4 | 79,7 | 84,0 | 79,6  | 84,1  |
| 89,4 | 85,4  | 76,7  | 90,0  | 79,0  | 83,0 | 91,0 | 87,2 | 80,3  | 54,7  |

Необходимо вычислить  $\bar{X}$ ,  $S^2$ ,  $S$ ,  $V$ ,  $S_x$ ,  $S_x$  %.

**Выводы.** Занесите результаты в таблицы 4, 5.

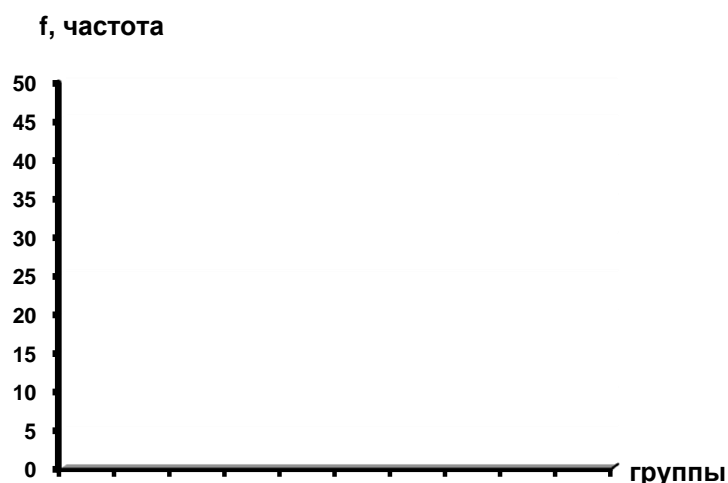
Таблица 4

| Группа | Разноска дат | Частота f | Групповые варианты X | По исходным данным |                |                 |
|--------|--------------|-----------|----------------------|--------------------|----------------|-----------------|
|        |              |           |                      | fX                 | X <sup>2</sup> | fX <sup>2</sup> |
|        |              |           |                      |                    |                |                 |

Таблица 5

| X | X <sub>ср</sub> | S <sup>2</sup> | S | V | S <sub>x</sub> | S <sub>x</sub> % |
|---|-----------------|----------------|---|---|----------------|------------------|
|   |                 |                |   |   |                |                  |

Постройте гистограмму и полигон распределения 50 растений льна по технической длине стебля:



**Задача 2.** Занесите в словарь новые понятия и правила анализа больших выборок.

**Задание 3.** Вычисление статистических характеристик выборки при качественной изменчивости признака.

**Материалы и оборудование.** Расчетные таблицы, рабочий материал, калькуляторы.

**Пояснение.** К качественным относят такие признаки, которые выражаются в каких-то качествах, не поддающихся количественному измерению: разные сельскохозяйственные культуры, разные виды болезней, окраска зерна или цветков, форма плода, наличие или отсутствие признаков или реакции на воздействие и т. д. Наиболее час-

то при изучении качественных признаков встречается случай, когда изучаемая совокупность представлена объектами только с двумя градациями – признак есть и признака нет, т. е. имеется две возможности, две альтернативы. Такое распределение называется альтернативным (двоичным).

Сводные статистические характеристики вычисляются по формулам следующей таблицы.

Таблица 6

| Показатель                                  | Формула                           |
|---|-----------------------------------|
| Доля признака:<br>при $k = 2$               | $P = n_1/N, q = 1-p$              |
| при $k > 2$                                 | $P_k = n_k/N$                     |
| Стандартное отклонение:<br>при $k = 2$      | $S = \sqrt{pq}$                   |
| при $k > 2$                                 | $LgS = (lgp_1 + \dots + lgp_k)/k$ |
| Коэффициент вариации                        | $V_p = (s/ s_{max}) \cdot 100$    |
| Ошибка доли                                 | $S_x = s/\sqrt{N}$                |
| Доверительный интервал<br>для доли признака | $p \pm t_{sp}$                    |
| Степень свободы                             | $N - 1$                           |

В таблице 6  $p_1, p_2 \dots p_n$  и  $q$  обозначают доли признака в совокупности;  $n_1, n_2 \dots n_k$  – численность группы;  $N$  – объем выборки;  $k$  – число градаций признака;  $t$  – теоретическое значение критерия Стьюдента.

Вычисления сводных характеристик выборки при качественной изменчивости складываются из распределения исходных наблюдений по группам (классам), определения среднего значения доли, изменчивости признака и доверительного интервала, в пределах которого находится значение доли в генеральной совокупности. При вычислении коэффициента вариации следует иметь в виду, что максимально воз-



можная изменчивость «макс» при двух градациях признаков равна 0,500 (50,0%), трех – 0,333 (33,3%), четырех – 0,250 (25,0%), пяти – 0,200 (20,0%) и шести – 0,167 (16,7%).

**Задача 1.** При просмотре 500 растений льна было обнаружено 50 растений, пораженных фузариозом. Определить 95% и 99% доверительные интервалы для генеральной доли пораженных растений в совокупности.

**Выводы.** Занесите результаты в таблицы; определите 95% и 99% доверительные интервалы для генеральной доли пораженных растений.

**Задача 2.** После распределения зерен озимой пшеницы по стекловидности получены данные (штук, зерен): полностью стекловидные  $n_1 = 658$ ; частично стекловидные  $n_2 = 102$ ; мучнистые  $n_3 = 60$ . Определить процентное содержание каждой группы зерен в выборке и их доверительные интервалы в генеральной совокупности с 1%-м уровнем значимости.

**Выводы.** Занесите результаты в таблицы; определите процентное содержание каждой группы зерен в выборке и их доверительные интервалы в генеральной совокупности с 1%-м уровнем значимости.

**Задача 3.** Занесите в словарь новые понятия и правила анализа выборок (качественные признаки).

## РАБОТА № 3. КОРРЕЛЯЦИОННО-РЕГРЕССИВНЫЙ АНАЛИЗ. АНАЛИЗ ДАННЫХ, ИЗМЕРЕНИЯ, РАСЧЁТЫ

**Материалы и оборудование.** Расчетные таблицы, рабочий материал, калькуляторы.

**Пояснение.** В научных исследованиях редко приходится иметь дело с точными и определенными функциональными связями, когда каждому значению одной величины соответствует строго определенное значение другой величины. Здесь чаще встречаются такие соотношения между переменными, когда каждому значению признака  $X$  соответствует не одно, а множество возможных значений признака  $Y$ , т. е. их распределение. Такие связи, обнаруживаемые лишь при массовом изучении признаков, в отличие от функциональных называются стохастическими (вероятностными) или корреляционными.

Регрессионный и ковариационный анализы приобретают все большее значение в современных исследованиях по биологии и агрономии. Под регрессией понимается изменение результативного признака  $Y$  (функции) при определенном изменении одного или нескольких факториальных (аргументов).

Связь между функцией и аргументом выражается уравнением регрессии или корреляционным уравнением. При простой регрессии уравнение кратко обозначается  $Y = f(x)$  и при множественной  $Y = f(X, Z, V \dots)$ . Если степень связи между признаками велика, то по уравнению регрессии можно предсказать значение результативного признака для определенных значений факториальных признаков. Для оценки тесноты (силы) связи используют коэффициенты корреляции и корреляционное отношение.

Совместное применение методов корреляции, регрессии и дисперсионного анализа для уточнения эксперимента получило название ковариационного анализа. Слово ковариация составлено из начальных букв слова «корреляция» и из слова «вариация».

Суть ковариационного анализа сводится к следующему. Если между результативным признаком  $Y$  и сопутствующим эксперименту неизучаемым признаком  $X$  имеет место значимая линейная связь, то методом ковариации можно статистически выровнять условия проведения опыта в отношении признака  $X$  и тем самым заметно снизить ошибку эксперимента и получить больше информации об изучаемом явлении.

Под линейной (прямолинейной) корреляционной зависимостью между двумя признаками  $X$  и  $Y$  понимают такую зависимость, которая носит линейный характер и выражается уравнением прямой линии  $Y = a + bX$ . Это уравнение называется уравнением регрессии  $Y$  на  $X$ , а соответствующая ему прямая линия – выборочной линией регрессии  $Y$  на  $X$ . Когда при одинаковых приращениях аргумента функция имеет неодинаковые изменения, регрессия называется криволинейной.

В качестве числового показателя простой линейной корреляции, указывающего на тесноту (силу) и направление связи  $X$  с  $Y$ , используют коэффициент корреляции, обозначаемый буквой  $r$ . Он является безразмерной величиной, изменяющейся в области  $-1 < r < +1$ . Коэффициент корреляции рассчитывают по формуле

$$r = \frac{\sum (X - X_{cp}) \cdot (Y - Y_{cp})}{\sqrt{\sum (X - X_{cp})^2 \cdot \sum (Y - Y_{cp})^2}}.$$

Если каждой величине  $X$  соответствует только определенная величина  $Y$ , то корреляционная связь переходит в функциональную, которую можно считать частным случаем корреляционной. При полных связях, когда корреляционная связь превращается в функциональную, значение коэффициента корреляции равно для положительных, или прямых, связей  $+1,0$ , для отрицательных, или обратных, связей  $-1,0$ . Чем ближе  $r$  к  $+1$  или  $-1$ , тем теснее прямолинейная корреляционная связь; она ослабевает с приближением  $r$  к  $0$ . Когда  $r = 0$ , между  $X$  и  $Y$  нет линейной связи, но криволинейная зависимость может существовать.

Квадрат коэффициента корреляции ( $r^2$ ) называется коэффициентом детерминации и обозначается  $d_{yx}$ . Он показывает долю (%) тех изменений, которые в данном явлении зависят от изучаемого фактора. Коэффициент детерминации является более непосредственным и прямым способом выражения зависимости одной величины от другой, и в этом отношении он предпочтительнее коэффициента корреляции. В случаях, где известно, что зависимая переменная  $Y$  находится в причинной связи с независимой переменной  $X$ , значение  $r^2$  показывает ту долю элементов в вариации  $Y$ , которая определена влиянием  $X$ . Поэтому когда употребляют, например, выражение «50%

колебаний в урожае вызывается колебаниями в выпадении осадков», то здесь 50% – коэффициент детерминации.

Считается, что при  $r < 0,3$  корреляционная зависимость между признаками слабая,  $r = 0,3-0,7$  – средняя, а при  $r > 0,7$  – сильная.

Для оценки надежности выборочного коэффициента корреляции вычисляют его ошибку и критерий существенности.

Стандартную ошибку коэффициента корреляции определяют по формуле

$$s_r = \sqrt{\frac{1 - r^2}{n - 2}},$$

где  $s_r$  – ошибка коэффициента корреляции;  $r$  – коэффициент корреляции;  $n$  – численность выборки, т. е. число пар значений, по которым вычислен выборочный коэффициент корреляции.

Из формулы следует, что коэффициенты корреляции, близкие к единице, оказываются всегда точнее коэффициентов корреляции, близких к нулю. С увеличением числа объектов исследования  $s_r$  также будет всегда уменьшаться, а точность в определении  $r$  – возрастать.

Критерий существенности коэффициента корреляции рассчитывают по формуле

$$t_r = \frac{r}{s_r},$$

если  $t_{\text{факт}} > t_{\text{теор}}$ , то корреляционная связь существенна, а когда  $t_{\text{факт}} < t_{\text{теор}}$  – несущественна. Теоретическое значение критерия  $t$  находят по таблице Стьюдента, принимая 5%-й уровень значимости.

Коэффициент корреляции указывает на направление и степень сопряженности в изменчивости признаков, но не позволяет судить о том, как количественно меняется результативный признак при изменении факториального на единицу измерения, что важно в познавательных и практических целях. В подобных случаях на помощь приходит регрессионный анализ. Его основная задача – определить формулу корреляционной зависимости, т. е. уравнение прямой линии.

Уравнение линейной регрессии  $Y$  по  $X$  имеет вид

$$Y = Y_{cp} - b_{yx} (X - X_{cp}).$$

Коэффициент регрессии  $b_{yx}$  показывает, как изменяется  $Y$  при изменении  $X$  на единицу измерения, и выражается в единицах  $Y$ , а  $b_{xy}$  указывает регрессию  $X$  на  $Y$  и выражается в единицах  $X$ .

Таким образом, *коэффициентом линейной регрессии называется число, показывающее, в каком направлении и на какую величину изменяется в среднем признак  $Y$  (функция) при изменении признака  $X$  (аргумента) на единицу измерения.* Коэффициенты регрессии имеют знак коэффициента корреляции.

Произведение коэффициентов регрессии равно квадрату коэффициента корреляции:

$$b_{yx} b_{xy} = r^2.$$

Этой формулой можно пользоваться как проверочной при вычислении коэффициентов регрессии.

Ошибку коэффициента регрессии вычисляют по формуле

$$s_{byx} = s_r \sqrt{\frac{\sum (Y - Y_{cp})^2}{\sum (X - X_{cp})^2}};$$
$$s_{bxy} = s_r \sqrt{\frac{\sum (X - X_{cp})^2}{\sum (Y - Y_{cp})^2}}.$$

Критерий существенности коэффициента регрессии определяют по формуле

$$t_b = \frac{b}{s_b}.$$

Если определен критерий существенности для коэффициента корреляции, он может быть использован и для оценки значимости коэффициента регрессии, так как  $t_b = t_r$ .

Корреляция может быть изображена графически в виде линии регрессии. Для построения графика по оси абсцисс откладывают значения признака  $X$ , по оси ординат – значения признака  $Y$  и каждое наблюдение над двумя переменными с точкой с координатами  $(X, Y)$ . Такой график называется «точечной диаграммой» или «корреляционным полем».

**Задача 1.** Провести корреляционный и регрессионный анализ данных таблицы 1, в которой представлены данные по определению относительно» влажности ( $X$ ) и липкости ( $Y$ ) чернозема.

Таблица 1

| Номер пары | Значение признаков |                    | $X^2$ | $Y^2$ | $XY$ |
|------------|--------------------|--------------------|-------|-------|------|
|            | $X, \%$            | $Y, \text{г/см}^2$ |       |       |      |
| 1          | 19,9               | 0,0                |       |       |      |
| 2          | 20,9               | 0,6                |       |       |      |
| 3          | 26,1               | 1,1                |       |       |      |
| 4          | 29,4               | 1,2                |       |       |      |
| 5          | 30,5               | 1,7                |       |       |      |
| 6          | 40,3               | 1,7                |       |       |      |
| 7          | 44,8               | 2,6                |       |       |      |
| 8          | 47,8               | 3,4                |       |       |      |
| 9          | 55,6               | 4,2                |       |       |      |
| 10         | 58,3               | 5,8                |       |       |      |
| 11         | 64,5               | 6,3                |       |       |      |
| 12         | 76,6               | 7,3                |       |       |      |
| Сумма      |                    |                    |       |       |      |

**Решение:**

1. Вычисляют шесть вспомогательных величин, записывая цифры под расчетной таблицей 1.
2. Определяют коэффициент корреляции, регрессии и уравнение регрессии.
3. Вычисляют ошибки, критерий значимости и доверительные интервалы.
4. По уравнению регрессии рассчитывают усредненные теоретические значения  $Y$  для экстремальных величин  $X$  и строят теоретическую линию регрессии  $Y$  по  $X$ .

5. Найденные точки наносят на график и, соединяя их прямой, получают теоретическую линию регрессии  $Y$  по  $X$ . На графике целесообразно указать уравнение регрессии, коэффициент регрессии и корреляции, доверительную зону для истинной линии регрессии в совокупности. Чтобы отграничить доверительную зону, необходимо вверх и вниз от теоретической линии регрессии отложить величину одной (68%-я зона) или двух (95%-я зона) ошибок отклонения от регрессии и соединить найденные точки пунктирными линиями. Область, заключенная между этими линиями, и называется доверительной зоной регрессии.

**Выводы.** Записать результат и объяснить его.

**Задача 2.** Определена пораженность льна фузариозом (ряд  $Y$ ) в зависимости от интервала между посевом на одном и том же поле восприимчивых к грибным патогенам (фузариозу) сортов льна (ряд  $X$  в таблице 2). Провести корреляционный и регрессионный анализ данных.

**Решение:**

1. Составляем расчетную таблицу и вычисляем вспомогательные величины, записывая их под таблицей 2.

2. Определяют коэффициент корреляции, регрессии и уравнение регрессии  $Y$  по  $X$ .

3. Вычисляют ошибки, критерий значимости и доверительные интервалы для  $r$ ,  $b_{yX}$  и проверяют  $H_0$ .

4. По найденному уравнению регрессии рассчитывают теоретические усредненные значения  $Y$  или двух крайних величин  $X$  и строят линию регрессии  $Y$  по  $X$ .

5. Найденные точки наносят на график и соединяют прямой линией. Из таблицы 2 на график последовательно переносят исходные даты и указывают основные статистические показатели. Экспериментальные точки, которые отмечены кружками, достаточно хорошо ложатся на линию прямолинейной регрессии.

**Выводы.** Записать результат и объяснить его.

Таблица 2 – Расчет вспомогательных величин для вычисления корреляции и регрессии  $Y$  по  $X$

| Номер пары | Значение признаков |         | $X^2$ | $Y^2$ | $XY$ |
|------------|--------------------|---------|-------|-------|------|
|            | $X$ ,годы          | $Y$ , % |       |       |      |
| 1          | 1                  | 88      |       |       |      |
| 2          | 2                  | 76      |       |       |      |
| 3          | 2                  | 70      |       |       |      |
| 4          | 7                  | 5       |       |       |      |
| 5          | 6                  | 12      |       |       |      |
| 6          | 5                  | 28      |       |       |      |
| 7          | 3                  | 45      |       |       |      |
| 8          | 4                  | 45      |       |       |      |
| 9          | 6                  | 9       |       |       |      |
| 10         | 3                  | 62      |       |       |      |
| Сумма      |                    |         |       |       |      |

**Задача 3.** Провести корреляционный и регрессионный анализ для выборочной совокупности, в которой представлены результаты определения содержания гумуса и подвижных форм фосфатов в пахотном слое легкосуглинистой дерново-подзолистой почвы.

**Решение:**

1. Группируют данные в корреляционную таблицу (решетку), состоящую из столбцов и строк, количество которых соответствует числу групп для ряда  $X$  (столбцы) и ряда  $Y$  (строки). При  $n = 64$  целесообразно выделить 6-8 групп. Определяют для ряда  $X$  и  $Y$  величину интервала группировки и число групп.

2. Составляют расчетную таблицу и проводят вспомогательные вычисления. В таблице вместо границ групп проставляют их середины и преобразуют  $X$  и  $Y$  по соотношениям.

3. Вычисляют выборочный коэффициент корреляции, регрессии и уравнение регрессии  $Y$  по  $X$ .

4. Определяют ошибки, критерий значимости, доверительные интервалы для  $r$ ,  $b_{yx}$  и проверяют  $H_0$ .

5. По найденному уравнению регрессии рассчитывают средние теоретические значения  $yx$  для экстремальных групповых значений  $X$  и строят теоретическую линию регрессии  $Y$  по  $X$ . Построив на графике точки, проводят через них теоретическую линию регрессии  $Y$  по  $X$ ; пунктирными линиями указывают доверительную зону регрессии для 68%-го уровня.

**Выводы.** Записать результат и объяснить его.



## РАБОТА № 4. МНОГОФАКТОРНЫЙ ДИСПЕРСИОННЫЙ АНАЛИЗ

**Пояснение.** Дисперсионный анализ широко используется для планирования эксперимента и статистической обработки его данных. Если в недалеком прошлом считали, что роль математика состоит лишь в анализе экспериментальных данных, то работы Р.А. Фишера коренным образом изменили эту точку зрения, и в настоящее время статистическое планирование опыта в соответствии с требованиями дисперсионного анализа и математическая интерпретация результатов – неперенные условия успешного получения ответов на вопросы, интересующие экспериментатора. Статистически обоснованный план эксперимента определяет и метод математического анализа результатов. Поэтому современный эксперимент нельзя правильно спланировать, не зная основ дисперсионного анализа.

При дисперсионном анализе одновременно обрабатывают данные нескольких выборок (вариантов), составляющих единый статистический комплекс, оформленный в виде специальной рабочей таблицы. Структура статистического комплекса и его последующий анализ определяются схемой и методикой эксперимента.

*Сущностью дисперсионного анализа является расчленение общей суммы квадратов отклонений и общего числа степеней свободы на части – компоненты, соответствующие структуре эксперимента, и оценка значимости действия и взаимодействия изучаемых факторов по F-критерию.*

Если обрабатывают однофакторные статистические комплексы, состоящие из нескольких независимых выборок, то общая изменчивость результативного признака, измеряемая общей суммой квадратов  $S_v$ , расчленяется на два компонента: варьирование между выборками (вариантами)  $S_v$  и внутри выборок  $S_z$ . Следовательно, в общей форме изменчивость признака может быть представлена выражением:

$$S_v = S_p + S_v + S_z.$$

Здесь вариация между выборками (вариантами) представляет ту часть общей дисперсии, которая обусловлена действием изучаемых факторов, а дисперсия внутри выборок характеризует случайное варьирование изучаемого признака, т. е. ошибку эксперимента.

Общее число степеней свободы ( $N-1$ ) также расчленяется на две части – степени свободы для вариантов ( $l-1$ ) и для случайного варьирования ( $N-l$ ):

$$N - 1 = (l - 1) + (N - l).$$

Суммы квадратов отклонений по данным полевого опыта – статистического комплекса с  $l$ -вариантами и  $n$ -повторениями – находят обычно в такой последовательности. В исходной таблице определяют суммы по повторениям  $P$ , вариантам  $V$  и общую сумму всех наблюдений  $UX$ .

Затем вычисляют:

- 1) общее число наблюдений  $N = ln$ ;
- 2) корректирующий фактор (поправку)  $C = (UX)^2 : N$ ;
- 3) общую сумму квадратов  $C_V = UX^2 - C$ ;
- 4) сумму квадратов для повторений  $C_P = UP^2 : l - C$ ;
- 5) сумму квадратов для вариантов  $C_V = UV^2 : n - C$ ;
- 6) сумму квадратов для ошибки (остаток)  $C_Z = C_Y - C_P - C_V$ .

Две последние суммы квадратов  $C_V$  и  $C_Z$  делят на соответствующие им степени свободы, т. е. приводят к сравниваемому виду – одной степени свободы вариации. В результате получают два средних квадрата (дисперсии)

$$\begin{aligned} \text{вариантов } s_v^2 &= C_V / (l - 1); \\ \text{ошибки } s^2 &= C_Z / (n-1)(l - 1). \end{aligned}$$

Эти средние квадраты и используют в дисперсионном анализе для оценки значимости действия изучаемых факторов. Оценка проводится путем сравнения дисперсии вариантов  $s_v^2$  с дисперсией ошибки  $s^2$  по критерию  $F - s_v^2 / s^2$ . Таким образом, за базу – единицу сравнения – принимают средний квадрат случайной дисперсии, которая определяет случайную ошибку эксперимента. При этом проверяемой нулевой гипотезой служит предположение: все выборочные средние являются оценками одной генеральной средней и, следовательно, различия между ними несущественны. Если  $F_\phi = s_v^2 / s^2 < F_m$ , то нулевая гипотеза не отвергается; между всеми выборочными средними нет существенных различий, и на этом проверка заканчивается. Нулевая гипотеза отвергается, когда  $F_\phi = s_v^2 / s^2 > F_m$ . В этом случае дополни-

тельно оценивают существенность частных различий по НСР и определяют, между какими средними имеются значимые различия.

Теоретическое значение критерия  $F$  для принятого в исследовании уровня значимости находят по таблицам с учетом числа степеней свободы для дисперсии вариантов и случайной дисперсии.

**Задача 1.** Обработка опытов, проведенных методом рендомизированных повторений.

В двухфакторном опыте  $3 \times 4$ , поставленном в четырех рендомизированных повторениях, изучено действие трех градаций орошения (0 – без орошения, 1 – умеренное и 2 – обильное орошение) и четырех доз азота (0 – без азота, 1 – 60, 2–120, 3 – 240 фунтов на акр) на урожай семян хлопчатника (табл. 1). Провести дисперсионный анализ данных.

Таблица 1 – Влияние орошения и доз азота на урожай семян хлопчатника (центнера с 1 акра)

| Орошение $A$ | Дозы азота $B$ | Повторения, $X$ |    |     |    | Суммы $V$ | Средние |
|--------------|----------------|-----------------|----|-----|----|-----------|---------|
|              |                | I               | II | III | IV |           |         |
| 0            | 0              | 19              | 20 | 15  | 15 |           |         |
|              | 1              | 20              | 20 | 20  | 18 |           |         |
|              | 2              | 18              | 20 | 18  | 18 |           |         |
|              | 3              | 20              | 19 | 18  | 19 |           |         |
| 1            | 0              | 32              | 29 | 18  | 21 |           |         |
|              | 1              | 40              | 39 | 33  | 34 |           |         |
|              | 2              | 39              | 38 | 40  | 37 |           |         |
|              | 3              | 44              | 42 | 40  | 39 |           |         |
| 2            | 0              | 30              | 31 | 21  | 17 |           |         |
|              | 1              | 42              | 35 | 28  | 33 |           |         |
|              | 2              | 38              | 38 | 36  | 35 |           |         |
|              | 3              | 48              | 51 | 50  | 48 |           |         |
| Суммы        |                |                 |    |     |    |           |         |

**Решение.** Дисперсионный анализ двухфакторного опыта с тремя градациями фактора  $A$  – орошения ( $l_A = 3$ ) и четырьмя градациями фактора  $B$  – доз азота ( $l_B = 4$ ), поставленного в четырех повторениях ( $n = 4$ ), складывается из следующих этапов.

1. В таблице 1 определяют суммы и средние. Правильность вычислений проверяют по соотношению  $UP = UV = UX$ .

2. Определяют суммы квадратов отклонений:

- 1) общее число наблюдений  $N = l_A l_B n$ ;
- 2) корректирующий фактор (поправку)  $C = (YX)^2 : N$ ;
- 3) общую сумму квадратов  $C_V = YX^2 - C$ ;
- 4) сумму квадратов для повторений  $C_P = YP^2 : l - C$ ;
- 5) сумму квадратов для вариантов  $C_V = YV^2 : n - C$ ;
- 6) сумму квадратов для ошибки (остаток)  $C_Z = C_Y - C_P - C_V$ .

3. Составляют таблицу 3x4, в которую вписывают суммы урожаев по вариантам (из табл. 1), и находят необходимые для расчета главных эффектов суммы  $A$  и  $B$  (табл. 2).

Таблица 2 – Определение главных эффектов и взаимодействий

| Орошение $A$ | Доза азота, $B$ |   |   |   | Суммы $A$ |
|--------------|-----------------|---|---|---|-----------|
|              | 0               | 1 | 2 | 3 |           |
| 0            |                 |   |   |   |           |
| 1            |                 |   |   |   |           |
| 2            |                 |   |   |   |           |
| Суммы $B$    |                 |   |   |   |           |

Составляют таблицу дисперсионного анализа и определяют значимость действия и взаимодействия изучаемых факторов по F-критерию (табл. 3).

Таблица 3 – Результаты дисперсионного анализа двухфакторного опыта 3x4, проведенного методом рандомизированных блоков

| Дисперсия           | Сумма квадратов | Степени свободы | Средний квадрат | $F_\phi$ | $F_m$ |
|---------------------|-----------------|-----------------|-----------------|----------|-------|
| Общая               |                 |                 |                 |          |       |
| Повторений          |                 |                 |                 |          |       |
| Орошения $A$        |                 |                 |                 |          |       |
| Азота $B$           |                 |                 |                 |          |       |
| Взаимодействия $AB$ |                 |                 |                 |          |       |
| Остаток (ошибки)    |                 |                 |                 |          |       |

4. Для оценки существенности частных различий определяют

$$S_x = \sqrt{s^2/n},$$

$$S_d = \sqrt{2s^2/n}.$$

5. Оценка существенности главных эффектов и взаимодействия по НСР. В этом примере частные средние опираются на  $n = 4$ , а средние для главного эффекта  $A$  – на  $nl_b = 4 \times 4 = 16$  и средние для главного эффекта  $B$  – на  $nl_A = 4 \times 3 = 12$  наблюдений. Вычисляют  $Sd$  и  $НСР_{05}$  для главных эффектов: для фактора  $A$ ,  $B$  и  $AB$ :

$$S_d = \sqrt{2s^2/nl_b},$$

$$НСР_{05} = t_{05}S_d.$$

В заключение составляют итоговую таблицу (табл. 4). В таблице 4 показаны три значения  $НСР_{05}$ : одно для оценки существенности частных различий между средними, а два других для оценки существенности разности средних по фактору  $A$  и по фактору  $B$ , т. е. оценки главных эффектов орошения и азота.

Таблица 4 – Действие орошения и доз азота на урожай семян хлопчатника (центнера с 1 акра)

| Орошение $A$           | Доза азота, $B$ |    |     |     | Средние по фактору $A$ |
|------------------------|-----------------|----|-----|-----|------------------------|
|                        | 0               | 60 | 120 | 240 |                        |
| Без орошения           |                 |    |     |     |                        |
| Умеренное              |                 |    |     |     |                        |
| Обильное               |                 |    |     |     |                        |
| Средние по фактору $B$ |                 |    |     |     |                        |

## РАБОТА № 5. ПОСТРОЕНИЕ МОДЕЛИ ВИДОВОЙ СТРУКТУРЫ БИОЦЕНОЗА И ЕЁ ИССЛЕДОВАНИЕ (базисная модель двувидового леса с возобновляемым ресурсом)

**Материалы и оборудование.** Расчетные таблицы, рабочий материал, калькуляторы.

**Пояснение.** Главной задачей является построение сосредоточенной модели развития леса, которая учитывала бы все основные взаимодействия и приводила к верным результатам. Модель должна удовлетворять двум требованиям:

- возможность ее использования для анализа и прогноза;
- применение модели в качестве основы для более сложной пространственно-распределенной модели, причем с учетом локально-экологической направленности.

Начнем с базисной модели без учета возраста, которая отражает основные взаимодействия и может быть исследована методами теории динамических систем.

Будем характеризовать популяцию вектором плотности биомассы  $u = \{u_i(t)\}$ ,  $i = л$  (лиственные породы),  $i = х$  (хвойные). Мы используем двухуровневую систему трофических взаимодействий типа «ресурс – потребитель»: почва – конкурирующий между собой двувидовой лес. Состояние почвы будем описывать третьей переменной – обобщенным показателем плодородия  $P(t)$ . Он характеризует количество перегноя, структуру микрофлоры, дисперсность почвы и другие факторы. Система в таком случае имеет вид

$$\begin{aligned} P &= \sum (A_i - V_i)u_i + B(P^0 - P), \\ i &= (x, л), \\ u &= (k_i C_i V_i - D_i)u_i \end{aligned}$$

Помимо этой системы нужно указать начальные данные для трех искомым функций:  $u_x$ ,  $u_l$ ,  $P$ .

Использованы следующие обозначения:

$P$  – обобщенный показатель плодородия – плотность ресурса ( $кг/м^2$ );

$u_l$  – плотность биомассы лиственных пород ( $кг/м^2$ );

$u_x$  – плотность биомассы хвойных пород ( $кг/м^2$ );

$A_i$  – коэффициент восстановления почвы за счет опада  $i$ -й породы;

$B$  – коэффициент самовосстановления почвы (1/год);  
 $P^\circ$  – асимптотическое значение плодородия в отсутствие леса (кг/м<sup>2</sup>);  
 $V_i$  – скорость потребления ресурса (трофическая функция) (1/год);  
 $C_i$  – поправочный множитель, описывающий конкуренцию;  
 $\kappa_i$  – коэффициент роста  $i$ -й породы;  
 $D_i$  – коэффициент естественной смертности деревьев (1/год);  
 $W$  – влияние внешних факторов, чаще пагубное, поэтому с отрицательным знаком, (кг/(год-м<sup>2</sup>));  
 $t^\circ$  – среднее время взросления молодого леса (год).

**Задание.** Построить, по приведенным данным, сосредоточенную модель развития леса. В процессе исследования получены следующие результаты:  $u_n = 150$  кг/м<sup>2</sup>;  $u_x = 78$  кг/м<sup>2</sup>;  $A_i = 0,5$ ;  $B = 2,0$  (1/год);  $P^\circ = 10$  кг/м<sup>2</sup>;  $V_i = 3,2$  (1/год);  $C_i = 0,2$ ;  $\kappa_i = 3,0$ ;  $D_i = 5,0$  (1/год);  $W = -4,0$ ; (кг/(год-м<sup>2</sup>));  $t^\circ = 80$  лет.

**Решение.** Система написана в формализме трофических взаимодействий. Если поправочный коэффициент, описывающий конкуренцию  $C$  подставить в уравнения, то часть системы, описывающая взаимодействия между компонентами биоценоза, окажется записанной в вольтеровском формализме «встреч и эквивалентов». В таком случае коэффициент естественной смертности окажется включенным в линейный член естественного прироста. Аналогично можно представить и уравнение для плодородия в том же формализме за исключением члена самовосстановления почвы, которое происходит достаточно медленно, а также внешнего воздействия (чаще всего отрицательного).

Восстановление почвы в лесных массивах происходит эффективней, чем на открытых безлесных пространствах, что учитывается выполняемым в массивах условием  $A_i - V_i \geq 0$ .

Система описывает симбиотические отношения: чем плодороднее почва, тем лучше растут деревья, и чем больше деревьев, тем быстрее идет процесс восстановления почвы. Подобные системы типа «ресурс потребитель» хорошо изучены в математической биологии.

**Выводы.** Записать результат и объяснить его.

## РАБОТА № 6. ИССЛЕДОВАНИЕ ПРОЦЕССОВ ДИНАМИКИ ЭКОЛОГИЧЕСКИХ СИСТЕМ

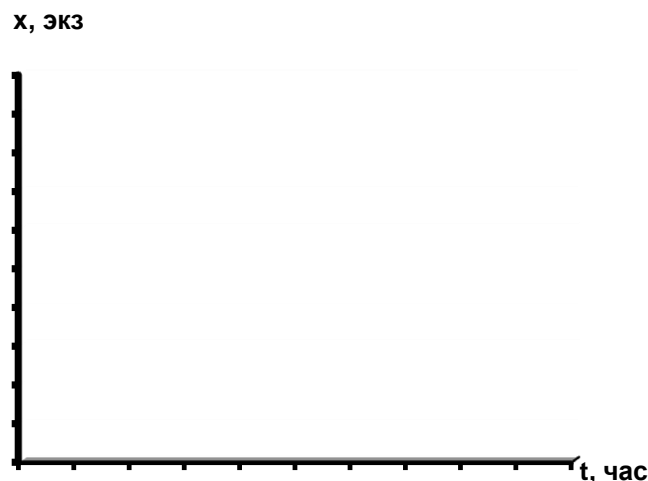
**Материалы и оборудование.** Расчетные таблицы, рабочий материал, калькуляторы.

**Пояснение.** *Уравнение Ферхюльста.* Базовой моделью, описывающей ограниченный рост, является модель Ферхюльста (1848):

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

Здесь параметр  $K$  носит название «емкости популяции» и выражается в единицах численности (или концентрации). Он не имеет какого-либо простого физического или биологического смысла и носит системный характер, то есть определяется целым рядом различных обстоятельств, среди них ограничения на количество субстрата для микроорганизмов, доступного объема для популяции клеток ткани, пищевой базы или убежищ для высших животных.

**Задача 1. Расчет ограниченного роста популяции.** В эксперименте с культурой инфузорий *Paramecium caudatum* получены следующие результаты:  $r = 0,007$ ,  $x = 45$  экз.,  $t = 1, 10, 20, 40, 60, 80, 100, 120, 140, 160, 180, 200$  часов,  $K = 1000$  экз. Необходимо вычислить численность популяции за каждый момент времени ( $x$ ) и нанести данные на график:



**Выводы.** Занесите в словарь новые понятия и правила расчета ограниченного роста популяции.



**Пояснение.** *Неограниченный рост. Экспоненциальный рост.* Одно из фундаментальных предположений, лежащих в основе всех моделей роста, – пропорциональность скорости роста численности популяции, будь то популяция зайцев или популяция клеток. Для многих одноклеточных организмов или клеток, входящих в состав клеточных тканей, размножение – это просто деление, то есть удвоение числа клеток через определенный интервал времени, называемый характерным временем деления. Для сложно организованных растений и животных размножение происходит по более сложному закону, но в простейшей модели можно предположить, что скорость размножения вида пропорциональна численности этого вида.

Математически это записывается с помощью дифференциального уравнения, линейного относительно переменной  $x$ , характеризующей численность (концентрацию) особей в популяции:

$$\frac{dx}{dt} = R \cdot x.$$

Здесь  $R$  в общем случае может быть функцией как самой численности, так и времени, или зависеть от других внешних и внутренних факторов.

Следующий закон был сформулирован Томасом Робертом Мальтусом (1766–1834) в книге «О росте народонаселения» (1798). Согласно формуле, если коэффициент пропорциональности  $R = r = \text{const}$  (как это предполагал Мальтус), численность будет расти неограниченно по экспоненте:

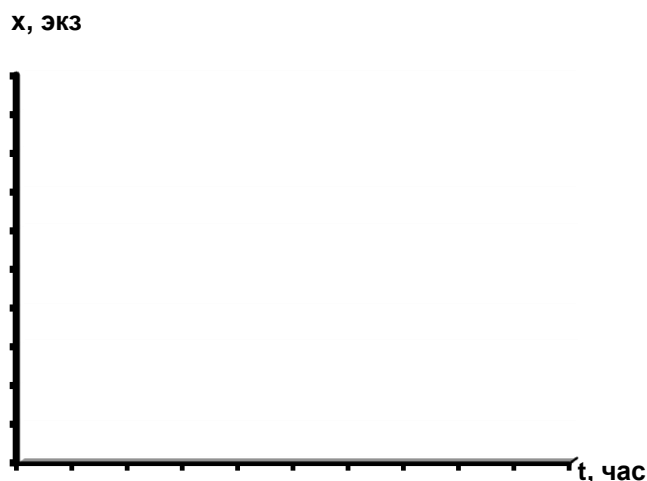
$$x = x_0 e^{rt}.$$

Для большинства популяций существуют ограничивающие факторы, и по тем или иным причинам рост популяции прекращается. Единственное исключение представляет человеческая популяция, которая на протяжении всего исторического времени растет даже быстрее, чем по экспоненте.

Закон экспоненциального роста справедлив на определенной стадии роста для популяций клеток в ткани, водорослей или бактерий в культуре.

## Задача 2. Расчет неограниченного роста популяции.

В эксперименте с культурой водорослей *Anabaena sp.* получены следующие результаты:  $r = 0,003$ ;  $x_0 = 2000$  экземпляров,  $t = 2, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150$  часов. Необходимо вычислить численность популяции за каждый момент времени ( $x$ ) и нанести данные на график:



**Выводы.** Занесите в словарь новые понятия и правила расчета неограниченного роста популяции.

## РАБОТА № 7. МНОГОМЕРНЫЕ МОДЕЛИ НА ПРИМЕРЕ КЛАСТЕРНОГО АНАЛИЗА

**Материалы и оборудование.** Расчетные таблицы, рабочий материал, калькуляторы.

**Пояснение.** Когда все входы количественные, альтернативной многомерной описательной моделью является модель, основанная на кластерном анализе. Кластерный анализ сопутствует самым разнообразным методам обнаружения структур, присущих сложным совокупностям данных. Основу данных чаще всего составляет выборка объектов, каждый из которых описывается набором отдельных переменных. Задача заключается в объединении переменных или элементов данной группы в такие кластеры, чтобы элементы внутри одного кластера обладали высокой степенью «естественной близости между собой, а сами кластеры были «достаточно отличны» один от другого. И подход к проблеме, и результаты принципиально зависят от того, какой смысл вкладывает исследователь в выражения «естественная близость» и «достаточно отличны».

В общем случае кластерный анализ предполагает, что о структуре ничего неизвестно или известно лишь немного. Все, что имеется в нашем распоряжении, это совокупность данных, в этом случае является обнаружение некоей категорной структуры, которая соответствовала бы наблюдениям, и проблема часто формулируется как задача отыскания «естественных групп». Сущностью кластерного анализа можно было бы с равным успехом считать и отыскание подходящего смысла для терминов «естественные группы» и «естественные ассоциации».

Кластерный анализ представляет собой попытку сгруппировать выборочные точки многомерного пространства в отдельные множества, которые, как мы надеемся, будут соответствовать наблюдаемым свойствам выборки. Группы точек могут быть в свою очередь сгруппированы в крупные множества, так что в конечном счете все точки оказываются иерархически классифицированными. Одним из простейших типов кластерного анализа является анализ по методу «одного звена», который был предложен Снитом как удобный способ представления таксономических связей в форме дендрограмм. Связи между  $n$  выборками выражаются через таксономические расстояния между каждой парой выборок, измеренные в некотором разумном масштабе. Метод заключается в такой сортировке выборок, которая

определяет кластеры по возрастающему набору пороговых расстояний ( $d_1, d_2, \dots, d_n$ ) (рис. 1).

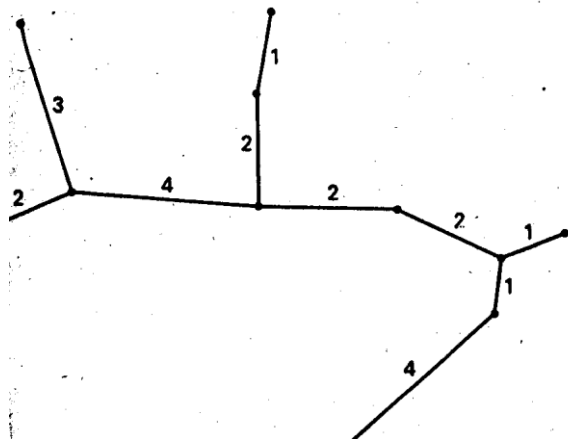


Рис. 1. Дерево минимальной протяженности с целочисленными длинами отрезков.

### Задача 1. Почвы Лейк-Дистрикта.

В качестве примера применения кластерного анализа, основанного на методах одного звена и дерева минимальной протяженности, мы рассмотрим анализ свойств 25 почв из национального парка Лейк-Дистрикт, применявшихся при исследовании реакции платана и березы на содержание питательных веществ в почве. Почвы были выбраны так, чтобы перекрыть как можно более широкий спектр химических свойств и, в частности, свойств, связанных с содержанием фосфатов. Прежде чем использовать почвы в экспериментах по выяснению реакции платана и березы различного происхождения, необходимо было установить диапазон изменчивости свойств почв и возможность их объединения в кластеры.

В табл. 1 приведены определенные для каждой из 25 почв значения семи переменных, а именно: потерь при прокаливании; количества фосфора, участвующего в изотопном обмене; фосфатазной активности; количество экстрагируемого железа; общего содержания фосфора; содержания азота и pH.

Таблица 1 – Значение семи переменных для 25 почв Лейк-Дистрикта.

| № п/п | Потери при прокаливании, % сухого веса | Количество фосфора, принявшего участие в изотопном обмене, мкг на 1г сухого веса | Фосфатная активность, мкг фенола на 1 г сухого веса почвы, 13 °С, 3 ч | Количество экстрагируемого железа, мг на 100г сухого веса | Общее содержание фосфора, % сухого веса | Общее содержание азота, % сухого веса | pH   |
|-------|--|--|---|---|---|---------------------------------------|------|
| 1     | 15,21                                  | 70,6   | 467,1   | 1400  | 0,12                                    | 0,63                                  | 4,53 |
| 2     | 33,27                                  | 67,5   | 1059,8  | 460   | 0,15                                    | 1,19                                  | 4,90 |
| 3     | 68,09                                  | 1700,3   | 3309,7  | 1200  | 0,36                                    | 2,30                                  | 4,82 |
| 4     | 32,89                                  | 168,1  | 1392,9  | 2100  | 0,17                                    | 1,29                                  | 4,84 |
| 5     | 19,87                                  | 102,7  | 71,3  | 920   | 0,14                                    | 0,73                                  | 7,93 |
| 6     | 16,46                                  | 32,5   | 367,0   | 1100  | 0,06                                    | 0,52                                  | 3,78 |
| 7     | 10,56                                  | 192,9  | 352,4   | 1000  | 0,10                                    | 0,33                                  | 4,59 |
| 8     | 15,63                                  | 118,4  | 300,2   | 1900  | 0,11                                    | 0,61                                  | 4,16 |
| 9     | 11,15                                  | 101,4  | 308,4   | 1300  | 0,11                                    | 0,47                                  | 5,13 |
| 10    | 16,25                                  | 232,5  | 306,2   | 1600  | 0,12                                    | 0,66                                  | 4,43 |
| 11    | 9,94                                   | 51,4   | 212,3   | 1800  | 0,10                                    | 0,37                                  | 4,70 |
| 12    | 70,63                                  | 150,3  | 627,7   | 590   | 0,15                                    | 1,81                                  | 3,65 |
| 13    | 9,0                                    | 9,8  | 129,7   | 95  | 0,01                                    | 0,21                                  | 3,63 |
| 14    | 19,71                                  | 297,7  | 467,9   | 2200  | 0,08                                    | 0,63                                  | 4,04 |
| 15    | 26,02                                  | 83,9   | 618,3   | 2800  | 0,08                                    | 0,88                                  | 3,93 |
| 16    | 11,84                                  | 168,9  | 375,8   | 750   | 0,07                                    | 0,45                                  | 5,89 |
| 17    | 10,71                                  | 127,3  | 330,3   | 910   | 0,13                                    | 0,43                                  | 4,56 |
| 18    | 8,3                                    | 107,4  | 241,4   | 880   | 0,08                                    | 0,31                                  | 4,74 |
| 19    | 12,67                                  | 188,7  | 516,4   | 1300  | 0,05                                    | 0,33                                  | 4,40 |
| 20    | 15,92                                  | 203,6  | 336,9   | 1500  | 0,08                                    | 0,52                                  | 4,13 |
| 21    | 12,92                                  | 170,6  | 319,6   | 1600  | 0,06                                    | 0,44                                  | 4,05 |
| 22    | 7,54                                   | 53,8   | 315,7   | 890   | 0,05                                    | 0,28                                  | 4,70 |
| 23    | 21,96                                  | 104,3  | 578,8   | 1900  | 0,12                                    | 0,81                                  | 4,11 |
| 24    | 88,78                                  | 107,6  | 1156,8  | 290   | 0,06                                    | 0,99                                  | 3,19 |
| 25    | 72,19                                  | 174,7  | 1061,3  | 690   | 0,14                                    | 2,32                                  | 3,93 |

Между каждыми двумя почвами можно вычислить расстояние в евклидовом пространстве по формуле

$$d_{ij} = [\sum(x_{ki} - x_{kj})^2]^{1/2},$$

где  $d_{ij}$  – евклидово расстояние между  $i$ -й и  $j$ -й почвами,  $k_i$  – значение  $k$ -й случайной переменной для  $i$ -й почвы, нормализованное путем

вычитания среднего по 25 почвам и деления на величину стандартного отклонения по 25 почвам.

Поскольку вычисления необходимо проделать для всех возможных пар почв, т.е.  $n(n-1)/2$  (в данном случае для 300 пар), ясно, что без ЭВМ здесь не обойтись. Кластерный анализ удобно проводить с использованием статистических программ, например, STATISTIKA или STATGRAF.

**Выводы.** Построить дерево минимальной протяженности по 25 почвам. Записать результаты и объяснить их.

## РАБОТА № 8. ДИНАМИЧЕСКИЕ СТАТИСТИЧЕСКИЕ МОДЕЛИ

**Материалы и оборудование.** Расчетные таблицы, рабочий материал, калькуляторы.

**Пояснение.** Изучение динамики систем базируется на теории сервомеханизмов, разработанной тоже сравнительно недавно, а всякое практическое использование динамических моделей зависит от способности современных быстродействующих вычислительных машин решать большое число (сотни) уравнений за короткие промежутки времени. Эти уравнения являются более или менее сложными математическими описаниями того, как функционирует имитируемая система, и даются они в форме выражений для «уровней» различных типов, «темпа» изменения которых регулируется управляющими функциями. Уравнения для уровней описывают накопления в системе таких величин, как вес, численность организмов или количество энергии, а уравнения для темпов управляют изменением этих уровней во времени. Управляющие функции отражают правила, явные или неявные, которые, как предполагается, регулируют функционирование системы.

Как известно, математические модели данной системы могут отображать эту систему лишь с той степенью точности, с какой уравнения, описывающие свойства компонентов модели, отображают свойства компонентов реальной системы. Популярность динамических моделей обязана большой гибкости методов, применяемых для описания динамики систем, которая включает нелинейные реакции компонентов на регулирующие переменные, а также положительные и отрицательные обратные связи. Однако такая гибкость имеет и свои недостатки. Например, обычно невозможно учесть уравнения для всех компонентов системы, так как даже при наличии современных ЭВМ имитация быстро становится слишком сложной. Поэтому необходимо иметь некоторую абстракцию, основанную на здравом смысле и на допущениях относительно того, какие из многих **компонент** системы в действительности управляют ее функционированием.

При использовании системной динамики в моделировании мы выделяем три главных этапа. Во-первых, нужно установить, какое именно динамическое свойство системы представляет интерес, и сформулировать гипотезы о взаимодействиях, порождающих данное свойство; этот шаг можно отождествить с этапом постановки задачи;

он, очевидно, тесно связан с теми стадиями системного анализа, которые предшествуют самому этапу моделирования, хотя мы и имеем дело уже с детализированным поведением и взаимодействиями переменных системы. Во-вторых, машинная имитационная модель должна быть построена таким образом, чтобы она дублировала элементы поведения и взаимодействий, определенные как существенные для системы. В-третьих, когда мы убедимся в том, что поведение модели достаточно близко к поведению реальной системы, мы используем модель, чтобы понять последовательность изменений, наблюдаемых в реальной системе, и предложить эксперименты, которые можно поставить на стадии оценки потенциальных стратегий, т. е. на следующем этапе системного анализа.

Одна из привлекательных особенностей динамических моделей заключается в возможности использовать диаграммы связей для представления основных взаимодействий в сложной системе.

При изменении параметров экологического объекта с течением времени широко используются модели, называемые *рядами динамики*. При изменении целевой функции  $Y$  от фактора  $X$ , в качестве которого рассматривается время или другой фактор, не зависящий от  $Y$ , ряды динамики позволяют наглядно представить процесс в виде графиков или таблиц.

По времени, отражаемому в динамических рядах, они разделяются на *моментные* и *интервальные*. В моментных рядах уровни ряда выражают величину  $Y$  на определенную дату, например динамика числа заповедников (табл. 1) на конец года.

Таблица 1 – Число заповедников на конец года

| Год  | Число заповедников |
|------|--------------------|
| 1990 | 72                 |
| 1991 | 76                 |
| 1992 | 79                 |
| 1993 | 85                 |
| 1994 | 90                 |

Уровни моментных рядов динамики суммировать не имеет смысла, поскольку суммирование будет включать одну и ту же величину несколько раз, но разность уровней имеет определенный смысл.



В интервальных рядах уровни ряда выражают размеры явления за определенный промежуток времени. Отличительной особенностью интервальных рядов абсолютных величин является возможностью суммировать уровни следующих друг за другом по периодам, поскольку их можно рассматривать как итог за длительный период времени, например лесовосстановление (табл. 2).

Таблица 2 – Посадка леса

| Год  | Посадка леса, га |
|------|------------------|
| 2010 | 566              |
| 2011 | 521              |
| 2012 | 447              |
| 2013 | 428              |
| 2014 | 391              |

По полноте времени, отражаемого в рядах динамики, их можно разделить на ряды *полные и неполные*. В полных рядах даты или периоды следуют друг за другом с равным интервалом, в неполных – в последовательности времени равный интервал не соблюдается.

Большинство статистических характеристик основано на абсолютном или относительном сравнении уровней динамических рядов показателей динамики: *абсолютный прирост показателя, темпы роста и прироста*. Сравнимый уровень называют *текущим*, а уровень, с которым производится сравнение – *базисным*. За базисный уровень часто принимается либо предыдущий уровень, либо начальный в данном динамическом ряду.

Если производится сравнение каждого уровня с предыдущим, то получают цепные показатели динамики. Если каждый уровень сравнивается с начальным или каким-либо одним, принятым за базу сравнения, то получают базисные показатели.

*Абсолютным приростом* называется разность последующего и предыдущего уровней ряда динамики:

$$\Delta y = y_i - y_{i-1},$$

где  $y_i$  – текущий уровень ряда динамики;  $y_{i-1}$  – предыдущий уровень.

За весь период, описываемый рядом, абсолютный прирост будет равен

$$\Delta y = Y_n - Y_i = \Sigma \Delta = \Sigma (y_i - y_{i-1}),$$

где  $y_n$  – последний уровень ряда;  $y_i$  – первый уровень.

*Темпом роста* называется отношение последующего уровня к предыдущему или к базовому. Темп роста в виде коэффициентов вычисляется по следующим формулам:

$$\begin{aligned} T_p &= y_i / y_{i-1} - \text{цепные темпы роста,} \\ T_{m.p.} &= y_i / y_1 - \text{базисные темпы роста,} \\ T_{m.p.} &= Y_n / Y_1 - \text{темпы роста за весь период.} \end{aligned}$$

*Темпом прироста* называется отношение абсолютного прироста к базисному уровню

$$T_{p.n.} = \Delta y / y_{i-1}.$$

Для выявления основных тенденций развития экологических процессов используются методы выравнивания (сглаживания) рядов. Рассмотрим их.

1. *Метод укрупнения интервалов* применяется тогда, когда экологические показатели за короткие промежутки времени в силу влияния различных факторов, действующих на них, то повышаются, то понижаются. Из-за этого не видна основная тенденция развития изучаемого явления. Рассматривая приведенные в таблице 3 данные по объему сброса сточных вод, мы видим значительные колебания этого показателя, обусловленную главным образом изменением объема производства на данном предприятии.

Таблица 3 – Данные по объему сброса качественных сточных вод

| Год  | Объем сброса сточных вод объектом, тыс. м <sup>3</sup> | Год  | Объем сброса сточных вод объектом, тыс. м <sup>3</sup> |
|------|--|------|--|
| 1980 | 125,5  | 1988 | 169,5  |
| 1981 | 130,8  | 1989 | 162,4  |
| 1982 | 140,2  | 1990 | 186,8  |
| 1983 | 107,5  | 1991 | 181,2  |
| 1984 | 152,1  | 1992 | 168,2  |
| 1985 | 121,1  | 1993 | 222,5  |
| 1986 | 171,2  | 1994 | 195,7  |
| 1987 | 147,9  | 1995 | 140,1  |

Если укрупнить интервалы времени до пятилетнего, то получим новый ряд динамики (табл. 4), показывающий последовательное увеличение качественного сброса сточных вод. Здесь же определяется и среднегодовой сброс завода за пятилетие.

Таблица 4 – Данные по объему сброса сточных вод за пятилетие

| Пятилетие, гг. | Качественный сброс сточных вод предприятием, тыс. м |       |
|----------------|---|-------|
|                | 1981–1985   | 651,7 |
| 1986–1990      | 837,8   | 167,6 |
| 1991–1995      | 907,7   | 181,5 |

2. *Метод скользящей средней* используется при выявлении основной тенденции развития при укрупненных интервалах времени вместо каждого уровня данного ряда берутся средние из уровней рядом стоящих лет. Полученная средняя охватывает группу из некоторого числа уровней: трех, пяти, семи и т. д., в середине которой находится взятый.

Вместо каждого такого уровня берется средняя, в которой сглаживаются случайные отклонения. Эта средняя будет скользящей, поскольку период осреднения все время меняется: из него вычитается один член и прибавляется следующий.

**Задача 1.** Произвести расчет скользящей средней для статистики по качественному сбросу сточных вод предприятием.

Таблица 5 – Данные по объему сброса качественных сточных вод

| Год  | Сброс сточных вод объектом, тыс. м <sup>3</sup> | 5-летняя скользящая |         | Год  | Сброс сточных вод объектом, тыс. м <sup>3</sup> | 5-летняя скользящая |         |
|------|---|---------------------|---------|------|---|---------------------|---------|
|      |   | Сумма               | Средняя |      |   | Сумма               | Средняя |
| 1970 | 81,2  |                     |         | 1983 | 107,5   |                     |         |
| 1971 | 78,7  |                     |         | 1984 | 152,1   |                     |         |
| 1972 | 92,2  |                     |         | 1985 | 121,1   |                     |         |
| 1973 | 82,5  |                     |         | 1986 | 171,2   |                     |         |
| 1974 | 85,6  |                     |         | 1987 | 147,9   |                     |         |
| 1975 | 103,7   |                     |         | 1988 | 169,5   |                     |         |
| 1976 | 125,0   |                     |         | 1989 | 162,4   |                     |         |
| 1977 | 102,6   |                     |         | 1990 | 186,8   |                     |         |
| 1978 | 134,7   |                     |         | 1991 | 181,2   |                     |         |
| 1979 | 119,5   |                     |         | 1992 | 168,2   |                     |         |
| 1980 | 125,5   |                     |         | 1993 | 222,5   |                     |         |
| 1981 | 130,8   |                     |         | 1994 | 195,7   |                     |         |
| 1982 | 140,2   |                     |         | 1995 | 140,1   |                     |         |

**Решение.** Производим расчет 5-летних средних и заполняем таблицу 5. Скользящая средняя дает более или менее плавное изменение уровней. Проводим центрирование, заключающееся в нахождении средней из средних для отнесения полученного уровня к определенной дате.

Наиболее эффективным способом выявления основной тенденции развития является *аналитическое выравнивание* (определение тренда). При этом уровни ряда динамики выражаются в виде функции времени.

Аналитическое выравнивание является предпосылкой для применения других приемов углубленного изучения развития экологических процессов во времени, изучения варьирования данных в динамике, их связи с другими явлениями.

Аналитическое выравнивание состоит в подборе для данного ряда динамики теоретической кривой, выражающей основные черты фактической динамики. Здесь часто применяют МНК.

Проводим выравнивание ряда динамики по прямой:

$$y_i = a_0 + a_1 t.$$

По МНК имеем систему нормальных уравнений

$$\begin{aligned} n \cdot a_0 + a_1 \Sigma t &= \Sigma y, \\ a_0 \Sigma t + a_1 \Sigma t^2 &= \Sigma ty, \end{aligned}$$

где  $n$  – число членов ряда динамики.

Система уравнений упрощается, если  $t$  подобрать так, чтобы их сумма равнялась нулю, т. е. начало отсчета времени перенести в середину рассматриваемого периода. Тогда

$$\begin{aligned} a_0 &= \Sigma y/n, \\ a_1 &= \Sigma ty/ \Sigma t^2. \end{aligned}$$

Если число уровней четное, то условное обозначение принимает вид, как показано в таблице 6.

Таблица 6 – Условное обозначение времени

| Год  | Уровень |
|------|---------|
| 1989 | -7      |
| 1990 | -5      |
| 1991 | -3      |
| 1992 | -1      |
| 1993 | +1      |
| 1994 | +3      |
| 1995 | +5      |
| 1996 | +7      |

При нечетном числе членов ряда отсчет ведется от середины, взятой за ноль. Значение  $\Sigma t^2$  при четном числе уровней:

$$\Sigma t^2 = n(n^2 - 1)/3.$$

При нечетном:

$$\Sigma t^2 = n(n^2 - 1)/12.$$

**Задача 2.** По данным таблицы 7 найти уравнение динамики

$$y = a_0 + a_1 t.$$

Таблица 7 – Расчетные значения для определения уравнения динамики

| Год   | Процент загрязнения воздуха от уровня ПДК ( $y_i$ ) | $t_i$ | $t_i^2$ | $y_i t_i$ | Теоретические значения $y_i$ |
|-------|---|-------|---------|-----------|------------------------------|
| 1987  | 39,4  | -9    |         |           |                              |
| 1988  | 39,8  | -7    |         |           |                              |
| 1989  | 40,0  | -5    |         |           |                              |
| 1990  | 40,6  | -3    |         |           |                              |
| 1991  | 41,4  | -1    |         |           |                              |
| 1992  | 41,9  | +1    |         |           |                              |
| 1993  | 41,9  | +3    |         |           |                              |
| 1994  | 42,0  | +5    |         |           |                              |
| 1995  | 42,6  | +7    |         |           |                              |
| 1996  | 43,1  | +9    |         |           |                              |
| Сумма |   |       |         |           |                              |

**Решение.** Вычисляем параметры  $t_i^2$ ,  $y_i t_i$  и заносим их в таблицу 7.

Из таблицы находим: при  $n = 10$

$$a_0 = \Sigma y/n,$$

$$a_1 = \Sigma ty / \Sigma t^2.$$

Записать уравнение прямой при данных условиях.

По полученному уравнению находим теоретические значения процента загрязнения воздуха от уровня ПДК для каждого периода времени.

Мерой колебания уровней динамического ряда выступает средний квадрат отклонений фактических уровней ряда от пере-

менных уровней, исчисляемых по тренду. Эта величина подобна дисперсии, исчисляемой в рядах распределения с той разницей, что отсчет отклонений ведется не от средней (постоянной для данного ряда), а от переменной средней – выровненных уровней. Мера колеблемости определяется по формуле

$$\sigma_t^2 = 1/n \cdot (\Sigma(y_i - y'_{it})^2).$$

Относительная мера колеблемости (своеобразный коэффициент вариации) определяется по формулам

$$V_t = \sigma_t / y_{cp.t}$$
$$y_{cp.t} = 1/n \Sigma y_i$$

а в процентах

$$V_{t\%} = V_t \cdot 100.$$

Эта величина служит критерием правильности выбора уравнения тренда.

**Выводы.** По приведенным данным составить уравнение динамики, найти теоретические значения процента загрязнения воздуха от уровня ПДК для каждого периода времени и определить меры колеблемости.

## РАБОТА № 9. МАТРИЧНЫЕ МОДЕЛИ

**Материалы и оборудование.** Расчетные таблицы, рабочий материал, калькуляторы.

**Пояснение.** Рассмотренное в предыдущей лабораторной работе семейство динамических моделей – это пример лишь одного из возможных подходов к моделированию экологических систем. Такие модели предоставляют исследователю почти полную свободу при описании тех элементов системы, которые он считает существенными для понимания основных взаимосвязей между переменными и блоками, выделенными при описании системы. Эти модели стремятся обычно к максимальной «реалистичности» – к тому, чтобы как можно точнее описать языком математики те или иные физические, химические или биологические процессы, иногда даже в ущерб математической простоте и элегантности. Платой за это часто оказывается необходимость увеличивать число блоков модели (с тем, чтобы учесть относительно малые изменения в поведении системы) и те трудности, с которыми приходится сталкиваться при получении несмещенных и достоверных оценок модельных параметров. Матричные модели представляют собой семейство таких моделей, реалистичность которых в известной мере принесена в жертву тем преимуществам, которые дает специфика математического описания модели. При этом дедуктивная логика чистой математики позволяет «модельеру» проверить следствия принятых им допущений без всякого «экспериментирования» с моделью, требующего затрат машинного времени, хотя ЭВМ и могут использоваться для некоторых вычислений.

Термин «матрица» применяется математиками для обозначения прямоугольной таблицы чисел; так, матрица

$$A = \begin{vmatrix} 1 & 0 & 3 \\ 2 & 4 & 8 \\ 1 & 3 & 2 \end{vmatrix}$$

– это матрица из трех строк и трех столбцов или матрица  $3 \times 3$ . Каждое из 9 записанных значений называется **элементом**, и если матрица как целое есть  $A$ , то  $a_{ij}$  – это элемент  $i$ - строки и  $j$ -го столбца  $A$ . По-



этому символическая запись матрицы  $A$  выглядит следующим образом:

$$A = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}.$$

Условимся использовать для обозначения матрицы заглавные латинские буквы, выделенные жирным шрифтом, а для обозначения отдельных элементов этой матрицы – соответствующие строчные буквы с надлежащими индексами. В матричной алгебре, оперирующей этими таблицами чисел, часто встречаются матрицы специального вида. Например, матрица с одинаковым числом строк и столбцов называется **квадратной**, а следующим трем видам квадратных матриц

$$I = \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix}, \quad O = \begin{vmatrix} 0 & 0 \\ 0 & 0 \end{vmatrix}, \quad A = \begin{vmatrix} 3 & -2 \\ -2 & 5 \end{vmatrix}$$

присвоены специальные названия:  $I$  называется **единичной матрицей**,  $O$  – **нулевой**,  $A$  – **симметричной матрицей**. Особую роль играют элементы матрицы, стоящие на пересечении первой строки и первого столбца, второй строки и второго столбца и т. д.; они образуют так называемую **главную диагональ** матрицы. В **единичной**, или **тождественной**, матрице главная диагональ состоит из единиц, а все остальные элементы – нули. В симметричной матрице элементы главной диагонали могут принимать любые значения, а недиагональные элементы связаны соотношением  $a_{ij} = a_{ji}$ . Наконец, матрицы, состоящие только из одной строки или одного столбца

$$a = [1 \quad 3 \quad 2] \text{ или } b = \begin{vmatrix} 2 \\ 4 \\ 6 \end{vmatrix}$$

называют соответственно **вектор-строкой** и **вектор-столбцом**. Условимся обозначать такие матрицы строчными латинскими буквами, выделенными жирным шрифтом. Матрица, состоящая лишь из одного элемента, например,  $D = 3$  называется скаляром.

Записывать массивы чисел в виде матриц удобно потому, что это позволяет оперировать с ними так же, как и с обычными числами, т. е. скалярами. Например, сложение или вычитание двух матриц состо-

ит в сложении или вычитании всех соответствующих элементов этих матриц. Умножение и деление матриц не столь просты, но тоже представляют собой вполне определенные математические операции. Матричная алгебра является одной из важнейших областей современной математики, и экологам, которые намереваются применять в своих исследованиях системный анализ, стоит ознакомиться с данным разделом математики.

Квадратные матрицы обладают одним важным свойством, а именно: для любой такой матрицы существуют собственные числа и собственные векторы, которые удовлетворяют уравнению

$$Av = \lambda v,$$

где  $A$  – квадратная матрица,  $v$  – вектор-столбец, а  $\lambda$  – скаляр. В общем случае, если  $A$  есть матрица  $n \times n$ , существует  $n$  собственных чисел  $\lambda$ , причем среди них могут быть повторяющиеся, отрицательные и мнимые.

Один из ранних вариантов матричной модели был разработан Льюисом и Лесли как детерминистская модель, предсказывающая будущую возрастную структуру популяции самок по известной структуре в настоящий момент времени и гипотетическим коэффициентам выживания и плодовитости. Популяцию разбивают на  $n + 1$  возрастных групп (т. е. 0, 1, 2, 3, ...,  $n$ , причем каждая группа состоит из особей одного возраста), так что самая старшая группа, или группа, в которой все доживающие до данного возраста животные умирают, имеет номер  $n$ .

Пожалуй, наиболее известные модификации матричных моделей относятся к рассмотрению размерных классов или дискретных стадий внутри популяции. Ашер, например, применял такие модели для анализа управления и эксплуатации лесов, где деревья классифицировались не только по возрасту, но и по размеру. В отличие от него Лефкович использовал матричные модели для анализа популяций насекомых-вредителей; при этом структура популяций насекомых определялась различными стадиями жизненного цикла насекомого.

Матричные модели могут применяться и для анализа таких динамических процессов, как круговорот питательных веществ и поток энергии в экосистемах; предпосылкой для такого моделирования служит естественная подразделенность экосистемы на компартменты

– входящие в ее состав виды или трофические уровни. Потери экосистем трактуются при этом как разность между суммой входов и выходов и запасом Марковские модели отличаются тем, что сумма элементов в каждой строке матрицы равна единице.

**Задача 1. Выживание и плодовитость синего кита.** Наша задача по практическому применению матричных моделей аналогична основной модели Лесли. Он был построен Ашером на основе данных по синему киту (*Balaenoptera musculus*), собранных Лоузом и Эренфельдом в 30-е годы еще до того, как произошли резкие изменения коэффициентов выживания и вид фактически вымер.

Самки синего кита достигают половой зрелости в возрасте между четырьмя и семью годами, а продолжительность беременности составляет примерно один год. Рождается единственный детеныш, которого самка выкармливает в течение семи месяцев и за это время не может забеременеть вновь. Учитывая это обстоятельство, а также то, что киты совершают миграции, можно сделать вывод, что самка рождает не более одного детеныша за два года. Численности самцов и самок примерно одинаковы, и это отношение почти не меняется с возрастом. Предполагается, что синие киты живут не более 40 лет. Для первых десяти лет жизни коэффициенты выживания по оценкам равны 87% популяции за каждый год (приведенная ниже матрица соответствует оценке 87% за каждые два года).

Если шаг по времени положить равным двум годам, то матрица Лесли для популяции синего кита будет иметь следующий вид:

|      |      |      |      |      |      |      |  |
|------|------|------|------|------|------|------|--|
| 0    | 0    | 0,19 | 0,44 | 0,50 | 0,50 | 0,45 |  |
| 0,87 | 0    | 0    | 0    | 0    | 0    | 0    |  |
| 0    | 0,87 | 0    | 0    | 0    | 0    | 0    |  |
| 0    | 0    | 0,87 | 0    | 0    | 0    | 0    |  |
| 0    | 0    | 0    | 0,87 | 0    | 0    | 0    |  |
| 0    | 0    | 0    | 0    | 0,87 | 0    | 0    |  |
| 0    | 0    | 0    | 0    | 0    | 0,87 | 0,80 |  |

Элементы первой строки равны среднему числу потомков женского пола, производимых одной самкой за два года, причем максимальная плодовитость характерна для возрастных классов 8–9 и 10–11 лет. Для возрастных классов 12 лет и старше плодовитость несколько ниже – она уменьшается в среднем до 0,45 на одну самку. Естественная смертность принимается равной 13% популяции за каждый двухлетний период в течение первых десяти лет жизни. Ко-

ээффициент выживания китов возрастной группы 12 лет и старше полагается равным 80%.

**Решение.** Вычислите главное собственное число  $\lambda$  и соответствующий ему собственный вектор данной матрицы **a**.

Собственное число указывает на то, что численность популяции способна возрастать, а натуральный логарифм его дает оценку внутренней скорости прироста:

$$r = \ln\lambda.$$

Поскольку собственное число близко к 1, внутренняя скорость прироста, а следовательно, и доля животных, которых можно отлавливать каждый год, могут быть довольно малы. Если при эксплуатации  $N$  превысит расчетное значение, то численность вида неизбежно будет убывать, за исключением случаев, когда гомеостатические механизмы изменяют соответствующим образом параметры выживания и плодовитости.

## РАБОТА № 10. ПРОСТРАНСТВЕННЫЕ РАСПРЕДЕЛЕНИЯ ОРГАНИЗМОВ

**Материалы и оборудование.** Расчетные таблицы, рабочий материал, калькуляторы.

**Пояснение.** Все семейства моделей, которые мы рассматривали в предыдущих лабораторных работах, были детерминистскими. Это значит, что при одних и тех же начальных условиях результат моделирования одинаков и предсказывается математическими соотношениями, задающими модель. Детерминистские модели – это закономерное развитие той ветви математики, которую мы изучали еще на первых курсах институтов, в частности, так называемой «прикладной математики», т. е. математики, применяемой в физике. Такие модели являются необходимыми аналогами тех физических процессов, где имеет место взаимно однозначное соответствие между причиной и следствием.

Существует, однако, и более позднее направление математики, которое позволяет описывать связи с помощью вероятностей, в результате чего исход моделируемой реакции определен неоднозначно. Модели, которые оперируют вероятностями, называют стохастическими; такие модели особенно важны при учете изменчивости и сложности экологических систем. Разумеется, вероятностные элементы можно ввести почти в любой тип моделей, например, в динамические модели и, в частности, в исследование устойчивости этих моделей к вариациям их основных параметров. Но в данный момент мы остановились на моделях, которые стохастичны изначально. Поскольку развитие статистической теории прошло уже долгий исторический путь, нам придется остановиться лишь на некоторых типах моделей.

**Задача 1. Пространственные распределения организмов.** Одним из простейших применений стохастических моделей к решению экологических проблем является описание пространственных распределений живых организмов.

В экологии часто возникает необходимость предсказать число организмов, обитающих на какой-то определенной площади или на данном объекте (например, на дереве, листе или семени). При этом средняя плотность организмов, вообще говоря, неизвестна, и потому нам нужна математическая модель, которая давала бы эффективную меру среднего числа встречаемых особей, а также меру вариабельности встречаемости и структуру этой вариабельности. Если мы за-

будем на время о существовании полностью равномерного распределения, т. е. распределения, при котором каждый выборочный объект содержит одно и то же, без вариаций, число особей, простейшая гипотеза, которую мы сможем сформулировать, будет состоять в том, что каждый организм имеет постоянную, хотя и неизвестную, вероятность встретиться в выборке и что особи никак не влияют друг на друга или, иными словами, что вероятности встречаемости организмов независимы. Не вдаваясь в детали вывода (который на самом деле можно найти в большинстве учебников по элементарной теории вероятностей), заметим, что при сделанных допущениях вероятности обнаружения в выборке 0, 1, 2, 3 ...,  $x$  особей задаются следующей последовательностью:

$$e^{-m}, e^{-m} m/1, e^{-m} m^2/2, e^{-m} m^3/3, \dots, e^{-m} m^x/x,$$

где  $e$  – основание натурального (неперова) логарифма.

Эта последовательность широко известна в статистической литературе как ряд (или распределение) Пуассона. Согласно этому распределению, среднее число особей в выборке и их дисперсия равны  $m$ . Проверить пригодность распределения Пуассона для описания пространственного распределения организмов (а следовательно, и справедливость гипотезы, что вероятность встречаемости отдельного организма постоянна и не зависит от присутствия остальных особей) можно либо путем сравнения наблюдаемой частоты встречаемости с частотой, которая задается теоретическим распределением, либо сравнением среднего и дисперсии наблюдаемых частот с соответствующими теоретическими значениями.

Если пуассоновское распределение отвергается, это дает основание формулировать какие-то альтернативные гипотезы относительно распределения. Хотя при этом мы и руководствуемся тем, что уже известно о данной системе, всегда существует бесконечное множество вероятных гипотез, и потому наш поиск адекватной математической модели должен опираться прежде всего на экологию проблемы, а не на удобство математического описания. Кроме того, модель не должна быть переопределенной и содержать больше параметров, чем можно оценить исходя из разумной совокупности данных. Многие распределения, отличные от пуассоновского, порождают модели, предполагающие весьма специфические отклонения от «чистой»

случайности, и их применимость зависит, таким образом, от экологической сути задачи.

Если особи стремятся избегать тесных контактов, особенно когда число их возрастает, характерной особенностью распределения будет его регулярность и равномерная распространенность особей, причем дисперсия числа особей оказывается меньше его среднего значения. Относительно равномерное распространение часто бывает обусловлено, например, территориальным поведением животных, а расселение прикрепленных беспозвоночных может быть регулярным лишь на небольших участках дна реки или озера. Такого рода регулярные распределения можно аппроксимировать (положительным) биномиальным распределением, для которого ожидаемое распределение частот задается разложением бинома

$$n (q + p)^k,$$

где  $n$  – число выборочных объектов;  $p$  – вероятность того, что данная точка выборочного объекта занята какой-то особью;  $k$  – максимально возможное число особей, которое способен содержать выборочный объект. На практике параметры  $k$ ,  $p$  и  $q$  оценивают по выборкам из популяций. Чтобы проверить соответствие между фактическими результатами и положительным биномиальным распределением, сравнивают наблюдаемые и ожидаемые частоты, используя критерий  $\chi^2$ .

Когда пространственное распределение особей не является ни чисто случайным, ни регулярным, а дисперсия числа особей в выборке больше среднего их числа в выборке, распределение обычно называют «контагиозным», подразумевая, что в пространственной структуре имеются скопления, или «пятна» организмов, и нерегулярные «пустоты». Разумеется, неравномерности распределения особей может способствовать множество факторов внешней среды, а некоторые виды стремятся к агрегации, образуя скопления даже независимо от влияния этих факторов. Результирующая структура распределения зависит от размера групп и расстояния между ними, от пространственного распределения групп и распределения особей внутри группы.

Для неравномерных распределений предлагалось несколько математических моделей, и, пожалуй, наиболее известной из них является отрицательное биномиальное распределение. Оно задается разложением бинома

$$(q - p)^{-k},$$

где  $p = \mu/k$ ,  $q = 1 + p$ .

Параметрами этого распределения служат арифметическое среднее  $\mu$  и показатель степени  $k$ . Этот показатель связан с характером пространственного распределения особей, а его обратное значение  $1/k$  служит мерой избыточной дисперсии, или сгруппированности особей в популяции. Когда  $k$  стремится к бесконечности,  $1/k$  приближается к нулю, а распределение сходится к ряду Пуассона, т. е. к чисто случайной встречаемости особей. Если же  $k$  приближается к нулю, а  $1/k$  стремится к бесконечности, то распределение сходится к логарифмическому ряду.

Имеются различные критерии для подбора отрицательного биномиального распределения, причем эффективность этих критериев зависит от размера выборки, среднего числа особей на одну выборку и от того, насколько трудно оценить параметр  $k$ .

Недостатком отрицательного биномиального распределения является то, что оно может быть получено в самых разных биологических моделях. Это распределение может иметь место, например, в модели, где есть истинная контагиозность в том смысле, что присутствие в выборке одной особи повышает вероятность обнаружить в ней и другую особь. Можно показать, что к отрицательному биномиальному распределению размера популяции приводит также рост популяции с постоянными коэффициентам рождаемости и смертности особей и постоянной скоростью иммиграции. Отрицательному биномиальному распределению подчиняются и случайно распределенные группы особей, с числом индивидов в группах, распределенным по логарифмическому закону, а также популяции, состоящие из нескольких подпопуляций, внутри которых распределение случайно, но с различной вероятностью встречаемости.

По мере роста числа организмов на выборку или, наоборот, по мере увеличения числа выборок все эти распределения приближаются к нормальному, которое лежит в основе многих основных идей и критериев математической статистики. Даже при довольно малом числе выборок приближение к нормальному распределению можно получить сравнительно простым преобразованием данных наблюдений, а отыскание нужного преобразования само по себе может дать представление о природе изучаемого распределения организмов. Следует упомянуть, что существуют и другие подходы к исследова-



нию пространственных распределений, основанные на измерении расстояний между соседними особями.

**Задача 1. Численность *Helobdella* в выборках из озера.**

В таблице 1 приведено число пиявок *Helobdella* в 103 выборках, взятых из пресноводного озера, а в таблице 2 – соответствующие этим выборкам частоты.

**Решение.** Общее число обнаруженных особей равно  $\Sigma x$  (см. табл. 1, 2), так что среднее их число на выборку рассчитывается как

$$X_{cp} = \Sigma x_i / n,$$

где  $\Sigma x_i$  – сумма особей во всех выборках, n – число выборок.

Таблица 1 – Численность *Helobdella* в 103 выборках из пресноводного озера

|   |   |   |   |   |   |   |   |   |   |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 1 | 0 | 0 | 1 | 3 | 0 | 0 | 1 |
| 3 | 2 | 0 | 2 | 2 | 2 | 0 | 0 | 0 | 2 |
| 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| 1 | 1 | 0 | 0 | 0 | 0 | 6 | 2 | 2 | 0 |
| 1 | 0 | 2 | 0 | 1 | 0 | 1 | 2 | 1 | 0 |
| 0 | 1 | 4 | 0 | 0 | 0 | 1 | 0 | 4 | 0 |
| 2 | 1 | 1 | 1 | 5 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 2 | 1 | 1 | 0 | 0 | 2 |
| 0 | 0 | 0 | 1 | 0 | 2 | 1 | 0 | 0 | 1 |
| 0 | 1 | 8 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 0 | 1 | 0 | - | - | - | - | - | - | - |

Предполагая, что вероятность встретить любую особь в выборке остается постоянной и независимой, частота встречаемости 0, 1, 2, 3, ..., x организмов задается соответствующим членом ряда Пуассона:

$$P_{(x)} = e^{-m} m^x / x!$$

где  $m = \bar{x}$ .

Отсюда ожидаемая частота отсутствия индивидов в выборке есть  $ne^{-m} = 103(0,442418) = 45,57$ , и т. д. для остальных членов ряда.

Вычислите наблюдаемые и ожидаемые частоты и занесите их в таблицу 2.

Таблица 2 – Сравнение наблюдаемых частот *Helobdella* с частотами, ожидаемыми по распределению Пуассона

| Число особей | Наблюдаемая частота | Ожидаемая частота | Разность частот |
|--------------|---------------------|-------------------|-----------------|
| 0            |                     |                   |                 |
| 1            |                     |                   |                 |
| 2            |                     |                   |                 |
| 3            |                     |                   |                 |
| 4            |                     |                   |                 |
| 5            |                     |                   |                 |
| 6            |                     |                   |                 |
| 7            |                     |                   |                 |
| 8            |                     |                   |                 |

В случае если исходы выборок ожидаемых частот будут иметь малые значения, то их допустимо объединить, чтобы минимальное значение ожидаемой частоты было равно 1. Адекватность пуассоновского распределения как модели распределения *Helobdella* в озерных выборках проверяется затем с помощью критерия  $\chi^2$ , описанного в любом учебнике по элементарной статистике. Этот критерий вычисляется как

$$\chi^2 = \sum (Q_i - E_i)^2 / E_i$$

где  $Q_i$  – наблюдаемая частота, характерная для класса  $i$ ,  $E_i$  – ожидаемая частота, характерная для класса  $i$ .

Сравните полученную величину с табличными значениями для  $(k-1) = 4$ .

Другая проверка адекватности распределения состоит в сравнении дисперсии наблюдаемой численности организмов со средним числом особей на выборку. Для рассматриваемого нами примера расчет вести со 102 степенями свободы:

$$\chi^2 = \frac{s^2(n-1)}{m},$$

где  $s^2$  – дисперсия,  $n$  – число выборок,  $m$  – среднее число организмов на выборку.

Можно заметить, что здесь дается два разных толкования вероятностей – обстоятельство, которое может привести к некоторой путанице. Прежде всего мы определили стохастическую модель, в которой вероятность встречаемости особи постоянна и которая давала в среднем значение особи на выборку. Чтобы проверить справедливость этой модели, мы применили статистический критерий согласия наблюдаемых частот с частотами, ожидаемыми исходя из модели. Результаты этой проверки выражаются как вероятность (не путать с вероятностью встречаемости отдельного организма!) того, что подобные выборочные данные могли бы получиться для популяции, если бы та обладала свойствами, присущими модели. По соглашению, любая гипотетическая модель, для которой эта найденная по данным наблюдений вероятность оказывается меньше 0,05, отвергается.

Для того чтобы к данным нашего примера подобрать отрицательное биномиальное распределение, нам нужно прежде всего оценить параметр  $k$  в выражении

$$P_{(x)} = (1 + \mu/k)^{-k} (k + x - 1)! / x! (k - 1)! (\mu / \mu + k)^x,$$

где  $P_{(x)}$  есть вероятность обнаружения  $x$  особей в выборке, а параметры  $\mu$  и  $k$  оцениваются по распределению выборочных частот и с помощью статистик  $x'$  и  $s^2$ . Предварительную оценку  $k$  можно получить из уравнения

$$K = x'^2 / (s^2 - x'),$$

но затем эта оценка должна использоваться как начальное значение для подстановки в уравнение максимального правдоподобия:

$$n \cdot \ln(1 + x'/k) = \sum(A_{(x)}/k + x),$$

где  $n$  – общее число выборок,  $\ln$  – натуральный логарифм,  $A_{(x)}$  – общее число выборок с исходом, превосходящим  $x$ . Опробуются различные значения  $k$  до тех пор, пока равенство не будет приближенно выполняться.

Отдельные члены частотного распределения  $P_{(x=1, 2, 3, \dots, 8)}$  последовательно вычисляются по этапам, представленным ниже:

$$P_{(x=0)} = (1 - x'/k)^{-k},$$

$$P_{(x=1)} = (k/1) \cdot (x'/x + k),$$

$$P_{(x=2)} = (k + 1/2) \cdot (x'/x + k),$$

$$P_{(x=3)} = (k + 2/3) \cdot (x'/x + k) \text{ и т. д.}$$

Полученные значения ожидаемых частот занесите в таблицу 3, здесь же (для удобства) укажите наблюдаемые частоты и разность между ними и ожидаемыми значениями. Вычислите критерий  $\chi^2$  с 4 степенями свободы (т. е.  $n - 2$ , т. к. модель содержит теперь два параметра,  $\mu$  и  $k$ ).

Таблица 3 – Сравнение наблюдаемых частот с частотами, ожидаемыми исходя из отрицательного биномиального распределения

| Число особей | Наблюдаемая частота | Ожидаемая частота | Разность |
|--------------|---------------------|-------------------|----------|
| 0            |                     |                   |          |
| 1            |                     |                   |          |
| 2            |                     |                   |          |
| 3            |                     |                   |          |
| 4            |                     |                   |          |
| 5            |                     |                   |          |
| 6            |                     |                   |          |
| 7            |                     |                   |          |
| 8            |                     |                   |          |
| Сумма        |                     |                   |          |

**Выводы.** Сделайте выводы по проделанной работе и обоснуйте их.

## РАБОТА № 11. РЕГРЕССИОННЫЙ АНАЛИЗ

**Материалы и оборудование.** Расчетные таблицы, рабочий материал, калькуляторы.

**Пояснение.** При построении математических зависимостей могут быть две формы связей между функцией и переменными: *функциональная* и *регрессионная*. Если функциональные связи точно выражаются аналитическими уравнениями, то регрессионные связи выражаются уравнениями лишь приближенно. В общем случае можно сказать, что связь между функцией и аргументами будет тогда **функциональной**, когда будут учтены все аргументы, определяющие значения функции.

Уравнение регрессии составляется исследователем на основе характера связи между функцией и аргументами. Вопрос о форме связи решается, как правило, поэтапно.

Вначале рассматривается линейная форма связи вида

$$Y = b_0 + b_1 X_1 + b_2 X_2 + \dots + b_n X_n,$$

где  $X_i$  – факторы ( $i = 1, 2, \dots, n$ ), так как такая форма связи часто встречается на практике и для нее разработан хороший математический аппарат. При этом могут решаться следующие задачи:

- установление точности определения коэффициентов уравнения регрессии  $b_i$  в виде значений дисперсий  $S^2(b_i)$  или величины доверительных интервалов;
- установление значимости коэффициента  $b_i$ ;
- проверка адекватности установленной формы связи и экспериментальных данных.

При установлении тесноты связи между  $Y$  и  $X$  решается задача установления строгости соблюдения функциональной зависимости между изменениями  $Y$  и  $X$ . Для оценки тесноты связи между случайными переменными величинами используются следующие показатели:

а) в случае линейной формы связи

- коэффициент парной корреляции  $r_{yx}$  или  $r_{xy}$ , характеризующий строгость соблюдения пропорциональности, т. е. близость исследуемой формы связи с линейной;

– коэффициент частной корреляции  $r_{yx_1x_2\dots}$ , характеризующий тесноту связи между изучаемыми переменными при условии, что влияние остальных факторов исключается;

– коэффициент множественной корреляции  $R_{yx_1x_2\dots x_n}$ , характеризующий суммарное влияние всех факторов на величину  $Y$ ;

б) *в случае нелинейной формы связи*

– корреляционное отношение  $r$ , которое является характеристикой, насколько строго соблюдается функциональная связь между исследуемыми переменными. Этот показатель применим и для оценки тесноты связи в случае линейной формы связи. В этом случае он равен абсолютному значению коэффициента парной корреляции;

– множественное корреляционное отношение  $R_{yx_1x_2\dots x_n}$ , которое является характеристикой тесноты связи между  $Y$  и  $X$ . Аппарат корреляционно-регрессионного анализа используется в двух направлениях:

1) для проведения статистического анализа результатов наблюдений пассивных экспериментов, в которых независимые переменные  $X_i$  не могут изменяться экспериментатором, т.е. не регулируются. В результате такого анализа решение вопроса о виде формы связи не является окончательным, т.е. можно принять в качестве математической модели процесса большое число уравнений регрессии, которые могут удовлетворять полученным экспериментальным данным;

2) совместно с методом наименьших квадратов для планирования статистических экспериментов и анализа их результатов. В этом случае планирование экспериментов осуществляется в соответствии с принятым видом уравнения связи  $Y$  и  $X$ .

В соответствии с числом учитываемых независимых переменных  $X_i$  и характером связи между  $Y$  и  $X$  различают:

а) *по количеству исследуемых переменных*

- парный корреляционно-регрессионный анализ;
- множественный корреляционно-регрессионный анализ;

б) *в зависимости от формы связи*

- линейный корреляционно-регрессионный анализ;
- нелинейный корреляционно-регрессионный анализ.

**Метод наименьших квадратов.** Широкое распространение в практике математического моделирования получили уравнения регрессии вида

$$y = f(x),$$

где  $x$  – величина, которая рассматривается как случайная независимая переменная;  $y$  – случайная зависимая величина. При линейной форме связи эту зависимость можно выразить уравнением прямой

$$Y = b_0 + b_1X,$$

для построения которого требуется проведение экспериментов в объеме  $n$ , в каждом из которых должна фиксироваться пара значений  $(x; y)$ . Результаты эксперимента представляются либо в виде таблицы, либо в виде графиков.

**Задача 1.** В результате эксперимента зафиксированы пары значений  $(x_i, y_i)$ , приведенных в таблице 1.

Таблица 1 – Результаты эксперимента

|       |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
|-------|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $x_i$ | 1 | 3 | 2 | 5 | 2 | 5 | 6 | 2 | 3 | 6 | 4 | 1 | 3 | 4 | 6 | 5 | 1 | 4 |
| $y_i$ | 3 | 5 | 3 | 4 | 3 | 6 | 7 | 2 | 3 | 8 | 6 | 2 | 4 | 4 | 8 | 6 | 1 | 5 |

Вычислите:

1. Коэффициенты уравнения регрессии.
2. Постройте уравнение регрессии вида  $y = b_0 + b_1x$ .
3. Заполните статистическую таблицу эксперимента.

**Решение.** Для вычисления коэффициентов уравнения регрессии составьте статистическую таблицу 2.

Таблица 2 – Статистическая таблица эксперимента

| Номер опыта | Значения $X_i$ | Значения $Y_i$ | $X_iY_i$ | $X_i^2$ | $Y_i^2$ |
|-------------|----------------|----------------|----------|---------|---------|
| 1           |                |                |          |         |         |
| 2           |                |                |          |         |         |
| 3           |                |                |          |         |         |
| 4           |                |                |          |         |         |
| 5           |                |                |          |         |         |
| 6           |                |                |          |         |         |
| 7           |                |                |          |         |         |
| 8           |                |                |          |         |         |
| 9           |                |                |          |         |         |
| 10          |                |                |          |         |         |
| 11          |                |                |          |         |         |

| 1     | 2 | 3 | 4 | 5 | 6 |
|-------|---|---|---|---|---|
| 12    |   |   |   |   |   |
| 13    |   |   |   |   |   |
| 14    |   |   |   |   |   |
| 15    |   |   |   |   |   |
| 16    |   |   |   |   |   |
| 17    |   |   |   |   |   |
| 18    |   |   |   |   |   |
| Сумма |   |   |   |   |   |

По вычисленным суммам определяем

$$b_0 = (\sum y_i \cdot \sum x_i^2 - \sum x_i y_i \cdot \sum x_i) / (n \cdot \sum x_i^2 - (\sum x_i)^2);$$

$$b_1 = (\sum x_i y_i - \sum x_i \cdot \sum y_i) / (n \cdot \sum x_i^2 - (\sum x_i)^2).$$

Определяем коэффициент корреляции:

$$r = (\sum x_i y_i - \sum x_i \cdot \sum y_i) / \sqrt{[n \cdot \sum x_i^2 - (\sum x_i)^2] \cdot [n \cdot \sum y_i^2 - (\sum y_i)^2]}.$$

**Выводы.** Вычислить, насколько тесно значения X, Y связаны друг с другом.

При нелинейной форме связи могут быть использованы два подхода:

- первый – когда нелинейная форма связи представляется в виде линеаризованной функции;
- второй – когда используется итерационный нелинейный метод наименьших квадратов.

В первом случае исследователь сначала выбирает форму нелинейной связи, затем её линеаризует, преобразуя члены уравнения регрессии. Затем используется метод наименьших квадратов (МНК) для линеаризованного уравнения, откуда определяются коэффициенты уравнения регрессии. Полученное уравнение регрессии затем вновь преобразуется в нелинейную форму.

**Задача 2.** В результате многолетних исследований зависимости толщины слоя ила после разлива на пойменных лугах от толщи снежного покрова получены данные, показанные в таблице 3.



Таблица 3 – Изменение толщины слоя ила от величины снежного покрова

| Толщина снежного покрова $X$ , см | Толщина слоя ила $Y$ , см |
|-----------------------------------|---------------------------|
| 1                                 | 0,5                       |
| 2                                 | 1,0                       |
| 3                                 | 1,4                       |
| 4                                 | 1,7                       |
| 5                                 | 1,8                       |
| 6                                 | 1,9                       |
| 7                                 | 2,0                       |

Требуется найти зависимость между толщиной снежного покрова и толщиной слоя ила.

**Решение.** Предполагаем зависимость между  $X$  и  $Y$  вида  $y = ab^x$ .

Линеаризуем уравнение при  $y' = \ln y$ ;  $x' = x$ ;  $a' = \ln a$  и  $b' = \ln b$ , тогда  $y' = a' + xb'$ .

Составляем статистическую таблицу.

Таблица 4 – Вычисление данных для линеаризации уравнения регрессии

|          | Показатель |     |         |                 |       |
|----------|------------|-----|---------|-----------------|-------|
|          | $X$        | $Y$ | $\ln y$ | $x \cdot \ln y$ | $x^2$ |
| Значения | 1          | 0,5 |         |                 |       |
|          | 2          | 1,0 |         |                 |       |
|          | 3          | 1,4 |         |                 |       |
|          | 4          | 1,7 |         |                 |       |
|          | 5          | 1,8 |         |                 |       |
|          | 6          | 1,9 |         |                 |       |
|          | 7          | 2,0 |         |                 |       |
| $\Sigma$ |            |     |         |                 |       |

Составляем систему нормальных уравнений МНК:

$$na' + b' \Sigma x = y',$$

$$a'x + \Sigma x^2 = xy'.$$

**Выводы.** После преобразований получите значения  $a$ ,  $b$  и составьте уравнение регрессии для зависимости толщины слоя ила от толщины снежного покрова.

## РАБОТА № 12. МНОГОФАКТОРНЫЕ МОДЕЛИ

**Материалы и оборудование.** Расчетные таблицы, рабочий материал, калькуляторы.

**Пояснение.** В практике исследователя нередки случаи, когда функция отклика (цели)  $Y$  зависит не от одного, а многих факторов. Установление формы связи в таких случаях начинают, как правило, с рассмотрения линейной регрессии вида:

$$y = b_0 + b_1x_1 + b_2x_2 + \dots + b_kx_k.$$

В этом случае результаты наблюдений представляют в виде уравнений или матрицы результатов наблюдений. Для решения системы необходимо, чтобы количество опытов было не менее  $(k + 1)$ , т. е.  $n \geq k + 1$ .

*Задачей множественного регрессионного анализа* является построение такого уравнения прямой в  $k$ -мерном пространстве, отклонения результатов наблюдений  $x_{ij}$  от которой были бы минимальными. Используя для этого метод наименьших квадратов, получаем систему нормальных уравнений или в виде матрицы

$$(X^T X)B + X^T Y.$$

где  $B$  – вектор-столбец коэффициентов уравнения регрессии;  $X$  – матрица значений факторов;  $Y$  – вектор-столбец функций отклика;  $X^T$  – транспонированная матрица  $X$ .

**Задача 1.** В результате проведенных исследований влияния мощности гумусового слоя почвы ( $X_1$ ) и количества внесенного сложного состава минерального удобрения ( $X_2$ ) на урожайность зерновой культуры ( $Y$ ) (табл. 1) получить уравнения:

$$b_0 + x_{11}b_1 + x_{21}b_2 = y_1,$$

$$b_0 + x_{12}b_1 + x_{22}b_2 = y_2,$$

$$b_0 + x_{13}b_1 + x_{23}b_2 = y_3.$$

Установить форму связи урожайности с факторами  $x_1$  и  $x_2$  в виде линейного уравнения регрессии.

**Решение.** Представляем результаты опытов в виде матриц  $X$ ,  $X^T$ ,  $Y$ . Вычисляем  $X^T X$ ,  $X^T Y$ ,  $(X^T X)^{-1}$ .

Определяем коэффициенты уравнения регрессии  $B = (X^T X)^{-1} \cdot X^T Y$ .  
Отсюда находим  $b_0, b_1, b_2$ .

Таблица 1 – Экспериментальный материал исследования

| № | Уровни факторов |       | Значение функции $y$ при параллельных опытах |       |       | Опытное среднее значение $y_i$ |
|---|-----------------|-------|--|-------|-------|--------------------------------|
|   | $x_1$           | $x_2$ | $y_1$  | $y_2$ | $y_3$ |                                |
| 1 | 1,0             | 0,2   | 18,2   | 18,6  | 18,7  |                                |
| 2 | 2,0             | 0,4   | 21,6   | 23,4  | 23,7  |                                |
| 3 | 2,5             | 0,3   | 22,0   | 23,0  | 22,5  |                                |

Произведите расчет среднего значения  $y_i$ .

**Выводы.** Запишите полученное уравнение регрессии, какое оно имеет вид.

**Пояснение.** Для проверки значимости уравнения регрессии необходимо при заданных значениях  $(x_1, x_2)$  провести несколько экспериментов, чтобы для данного значения  $(x_1, x_2)$  получить некоторое среднее значение функции  $y$ . В этом случае экспериментальный материал представляется, например, в виде таблицы.

Число параллельных опытов, как правило, должно быть  $k \geq 3$ . Проверка значимости уравнения регрессии проводится по F-критерию. Для этого вычисляем остаточную дисперсию:

$$S_{ост}^2(y) = \Sigma(y_i - y'_i)^2 / (n - k - 1),$$

а затем вычисляем  $F_B$ -статистику:

$$F_B = S'^2 / S_{ост}^2.$$

которую сравниваем с табличным значением  $F_{\alpha \cdot k_1 \cdot k_2}$  при уровне значимости  $\alpha$  и числе степеней свободы  $k_1 = n - 1$ ;  $k_2 = n - k - 1$ . Гипотеза о значимости уравнения регрессии принимается при условии

$$F_B \geq F_{\alpha \cdot k_1 \cdot k_2}.$$

Значимость коэффициентов регрессии проверяют по t-критерию. Статистику

$$T = b_i / S'(b_i)$$

сравнивают с  $t_{\alpha, k1}$  при уровне значимости  $\alpha$  и степени свободы  $k = n - k - 1$ . Погрешность коэффициента регрессии определяется по формуле

$$S(B) = \sqrt{S^2_{ост}(y) \cdot c_{ij}}$$

где  $c_{ij}$  – диагональный элемент матрицы  $(X^T X)^{-1}$ . Доверительный интервал для коэффициентов регрессии определяем по формуле

$$b_i - t_{\alpha; k1} S'(b_i) \leq \beta_i \leq b_i + t_{\alpha; k1} S'(b_i),$$

где  $\beta$  – значение коэффициента регрессии в генеральной совокупности.

**Задача 2.** По результатам опытов, приведённым в таблице 1, получено уравнение регрессии. Занесите данные в таблицу. Проверьте значимость уравнения регрессии.

**Решение.** Данные представим в виде, удобном для вычислений (табл. 2).

Таблица 2 – Результаты проведенных опытов

| №     | Уровни факторов |       | Значение |         | Опытное среднее $y_i$ | Значение $y'_i$ из уравнения регрессии | $y_i - y'^2_i$ | $(y_i - y'_i)^2$ |
|-------|-----------------|-------|----------|---------|-----------------------|--|----------------|------------------|
|       | $x_1$           | $x_2$ | $y_i$    | $y_i^2$ |                       |  |                |                  |
| 1     |                 |       |          |         |                       |  |                |                  |
| 2     |                 |       |          |         |                       |  |                |                  |
| 3     |                 |       |          |         |                       |  |                |                  |
| 4     |                 |       |          |         |                       |  |                |                  |
| 5     |                 |       |          |         |                       |  |                |                  |
| 6     |                 |       |          |         |                       |  |                |                  |
| 7     |                 |       |          |         |                       |  |                |                  |
| 8     |                 |       |          |         |                       |  |                |                  |
| 9     |                 |       |          |         |                       |  |                |                  |
| Сумма |                 |       | -        | -       | -                     | -                                      | -              |                  |

Определяем остаточную дисперсию

$$S_{ост}^2(y) = \Sigma(y_i - y'_i)^2 / (n - k - 1)$$

и дисперсию для  $y$ :

$$S^2(y) = 1/(n - 1)[\Sigma y_i^2 - 1/n(\Sigma y_i)^2],$$

Вычисляем  $F_B$ -статистику:

$$F_B = S^2 / S_{ост}^2$$

**Выводы.** Исходя из полученных данных сделайте вывод о значимости уравнения регрессии.

*Множественный корреляционный анализ.* При множественном корреляционном анализе можно вычислить два типа парных коэффициентов регрессии:

1)  $r_{yxi}$  – коэффициент, определяющий тесноту связи между функцией отклика  $y$  и одним из факторов  $x$ ;

2)  $r_{xixm}$  – коэффициент, показывающий на связь между двумя факторами  $x_i$  и  $x_m$ . Их величины вычисляются по формулам

$$\begin{aligned} r_{yxi} &= Q_{xiy} / \sqrt{Q_{xi} Q_y}, \\ Q_{xiy} &= \Sigma x_i y - 1/n \cdot (\Sigma x_i) \cdot (\Sigma y), \\ Q_{xi} &= \Sigma x_i^2 - 1/n \cdot (\Sigma x_i)^2, \\ Q_y &= \Sigma y^2 - 1/n \cdot (\Sigma y)^2, \end{aligned}$$

где

$$r_{xixm} = (\Sigma x_i x_m - 1/n \cdot (\Sigma x_i)(\Sigma x_m)) / \sqrt{[\Sigma x_i^2 - 1/n \cdot (\Sigma x_i)^2] \cdot [\Sigma x_m^2 - 1/n \cdot (\Sigma x_m)^2]}.$$

Если ввести обозначения

$$\begin{aligned} Q_{xixm} &= \Sigma x_i x_m - 1/n \cdot (\Sigma x_i)(\Sigma x_m), \\ Q_{xm} &= \Sigma x_m^2 - 1/n \cdot (\Sigma x_m)^2, \end{aligned}$$

то

$$r_{xixm} = Q_{xixm} / \sqrt{Q_{xi} \cdot Q_{xm}}.$$

Проверка значимости коэффициентов корреляции может производиться:

а) сравнением статистического значения  $r_e$  с табличным  $r_{\alpha; k}$ , значимость коэффициента корреляции устанавливается исходя из условия

$$r_e \geq r_{\alpha; k},$$

где  $r_{\alpha; k}$  выбирается по специальной таблице при уровне значимости  $\alpha$  и числе степеней свободы  $k = n - 2$ ;

б) сравнением статистики  $t_e$  с табличным значением t-критерия при уровне значимости  $\alpha$  и числе степеней свободы  $k = n - 2$ , значимость коэффициента корреляции устанавливается исходя из условия

$$t_B = r_B / S'(r) \geq t_{\alpha; k}$$

где величина  $t_{\alpha; k}$  выбирается по таблице, а среднеквадратичное отклонение коэффициента корреляции  $S'(r)$  определяется по формуле

$$S'(r) = (1/\sqrt{n-1}) \cdot (1 - r_B^2).$$

Последовательно вычисляя значения всех парных коэффициентов, строим матрицу коэффициентов корреляции:

$$R_k = \begin{vmatrix} 1 & r_{yx1} & r_{yx2} & r_{yxk} \\ r_{x1y} & 1 & r_{x1x2} & r_{x1xk} \\ \dots & \dots & \dots & \dots \\ r_{xky} & r_{xkx1} & \dots & 1 \end{vmatrix}.$$

С помощью матрицы  $R_k$  вычисляют частные коэффициент корреляции, показывающие степень влияния одного из фактора  $x_i$  на функцию отклика  $Y$  при условии, что остальные факторы имеют постоянные значения.

Частные коэффициенты определяются по формуле

$$r_{y,x_1,x_2,\dots,x_k} = D_{1i} / \sqrt{D_{11} \cdot D_{ii}}$$

где  $D_{1j}$  – определитель матрицы  $R_k$ , образованный вычеркиванием первой строки и  $j$ -го столбца для каждого определителя соответственно. Аналогично вычисляются и определители  $D_{11}$  и  $D_{jj}$ . Значимость коэффициентов частной корреляции и доверительный интервал вычисляются так же, как и для коэффициентов парной корреляции, но число степеней свободы для критерия  $t_{a,k}$  принимается равным  $k = (n - 2) - p - 1$ , где  $(p - 1)$  – порядок частного коэффициента парной корреляции.

Если необходимо изучить степень тесноты связи между функцией отклика  $Y$  и несколькими факторами  $x_1, x_2, \dots, x_p$ . ( $p < k$ ) используют коэффициент множественной корреляции  $R_M$ , который всегда положителен и изменяется в пределах  $0 < R_M < 1$ . Чем ближе значение  $R_M$  к единице, тем лучше качество предсказания полученной моделью процесса, по наблюдениям за которым получены статистические данные. Коэффициент множественной корреляции исчисляется по формуле

$$R_M = \sqrt{(1 - (n - 1)/(n - 2) - p - 1) \cdot [1 - (\sum(y_i - y_i')^2 / \sum(y_i - y_i')^2)]}$$

или

$$R_M = \sqrt{(1 - D/D_{11})},$$

где  $D$  - определитель корреляционной матрицы.

Если  $R_M$  возвести в квадрат, то величина  $R_M^2$  называется множественным коэффициентом детерминации и показывает, какая часть дисперсии функции отклика объясняется вариацией линейной комбинации выбранных факторов.

При ( $p \leq k$ ),  $x_1, x_2, \dots, x_p$  значимость  $R_M$  можно проверить:

а) по  $t$ -критерию

$$t_{RM} = R_M / S_{RM};$$

б) по  $F$ -критерию

$$F_{RM} = \frac{R_M^2 \{n - p - 1\}}{(1 - R_M^2) p},$$

где

$$k_i = n - p - 1$$

и

$$k_2 = p.$$

**Задача 3.** Для предыдущего примера, в котором используются данные таблиц 1, 2, вычислить коэффициенты корреляции.

**Решение.** С целью облегчения вычислений результаты работы сведите в таблицу 3.

Таблица 3 – Результаты вычислений исследования

| №                  | Уровни факторов |       | $y_i$ | $y_i^2$ | $x_1 y_i$ | $x_2 y_i$ | $x_1^2$ | $x_2^2$ | $x_1 x_2$ |
|--------------------|-----------------|-------|-------|---------|-----------|-----------|---------|---------|-----------|
|                    | $x_1$           | $x_2$ |       |         |           |           |         |         |           |
| 1                  |                 |       |       |         |           |           |         |         |           |
| 2                  |                 |       |       |         |           |           |         |         |           |
| 3                  |                 |       |       |         |           |           |         |         |           |
| 4                  |                 |       |       |         |           |           |         |         |           |
| 5                  |                 |       |       |         |           |           |         |         |           |
| 6                  |                 |       |       |         |           |           |         |         |           |
| 7                  |                 |       |       |         |           |           |         |         |           |
| 8                  |                 |       |       |         |           |           |         |         |           |
| 9                  |                 |       |       |         |           |           |         |         |           |
| Сумма,<br>$\Sigma$ |                 |       |       |         |           |           |         |         |           |

Вычислим вспомогательные числа:

$$\begin{aligned} Q_{x_1 y} &= \Sigma x_1 y - 1/n \cdot (\Sigma x_1) \cdot (\Sigma y), \\ Q_{x_2 y} &= \Sigma x_2 y - 1/n \cdot (\Sigma x_2) \cdot (\Sigma y), \\ Q_{x_1} &= \Sigma x_1^2 - 1/n \cdot (\Sigma x_1)^2, \\ Q_{x_2} &= \Sigma x_2^2 - 1/n \cdot (\Sigma x_2)^2, \\ Q_y &= \Sigma y^2 - 1/n \cdot (\Sigma y)^2. \end{aligned}$$

Вычисляем коэффициент, определяющий степень связи между функцией  $Y$  и фактором  $x_1$  и  $x_2$ :



$$r_{yx1} = Q_{x1y} / \sqrt{Q_{x1} Q_y},$$

$$r_{yx2} = Q_{x2y} / \sqrt{Q_{x2} Q_y}.$$

Определяем тесноту связи между факторами  $x_1$  и  $x_2$ :

$$r_{x1x2} = Q_{x1x2} / \sqrt{Q_{x1} \cdot Q_{x2}}.$$

Составляем корреляционную матрицу  $R$ , определяем частные коэффициенты корреляции:

$$r_{yx1,x2} = D_{12} / \sqrt{D_{12} \cdot D_{22}},$$

$$r_{yx2,x1} = D_{21} / \sqrt{D_{22} \cdot D_{11}}.$$

Вычисляем коэффициент множественной корреляции  $R_M$ . Величина множественного коэффициента детерминации равна  $R_M^2$ .

**Выводы.** Сделайте заключение, какова доля дисперсии функции отклика объясняется вариацией линейной комбинации факторов  $x_1$  и  $x_2$ .

## РАБОТА № 13. ОПТИМИЗАЦИОННЫЕ МОДЕЛИ

**Материалы и оборудование.** Расчетные таблицы, рабочий материал, калькуляторы.

**Пояснение.** Столь странное слово «оптимизация» придумано для того, чтобы обозначить отыскание максимума либо минимума какого-то математического выражения или функции, когда некоторые их переменные мы можем изменять в определенных пределах. Если бы мы хотели найти только максимум, мы могли бы назвать этот процесс максимизацией – словом, которое в конце концов более приемлемо. И наоборот, отыскивая только минимум, мы могли бы использовать слово минимизация. Математически одну из этих операций всегда можно превратить в другую, так что в том, что оба процесса рассматриваются как один, есть определенная логика.

Совершенно естественным, однако, является желание сформулировать модель так, чтобы облегчить отыскание оптимальной комбинации ключевых переменных, и основные математические формулировки такого рода были разработаны независимо в тех ранних приложениях математических методов к практическим задачам, которые известны сейчас под названием «исследование операций». Здесь возникают терминологические трудности из-за того, что нахождение общего класса решений таких задач называлось «математическим программированием» еще до того, как слово «программирование» вошло в обиход в смысле написания инструкций для ЭВМ. Тот факт, что большая часть математического программирования сейчас тесно связана с ЭВМ, вносит лишь дополнительную путаницу!

В наиболее простой для описания форме математическое программирование известно под названием «линейное программирование». В этой модели мы начинаем с линейной целевой функции

$$Y = a_1x_1 + a_2x_2 + \dots + a_nx_n = \sum a_i x_i$$

и хотим найти максимум или минимум этой функции при одном или более ограничениях, которые также выражены в виде линейных функций, хотя исходно это могут быть просто неравенства, например,

$$b_1x_1 + b_3x_3 \geq Z.$$

Часто имеются неявные ограничения, состоящие в том, что  $x_i$  не могут быть отрицательными.

Когда переменных только две, задачи оптимизации такого рода довольно легко решаются графическими методами. Для более чем двух переменных задача сильно усложняется, и обычный подход к ее решению предполагает использование так называемого «симплекс-метода».

Одно из преимуществ оптимизационных моделей состоит в том, что они всегда освещают два немаловажных аспекта проблемы. Полученное решение дает значения переменных целевой функции, при которых эта функция достигает максимума или минимума, в зависимости от того, как поставлена задача. Однако, помимо этого, метод указывает и то ограничение, которое нужно ослабить, чтобы улучшить оптимальное значение целевой функции. В результате этого экспериментатор может тщательнее проверить постановку задачи и, в частности, свои оценки коэффициентов при переменных в целевой функции и природу указанного ограничения. Если окажется, что можно улучшить оценки или ослабить ограничение, то он сможет найти еще лучшее решение.

Как и прежде, проиллюстрировать структуру модели и ее применение, по-видимому, легче всего на решении простой задачи.

**Задача 1. Оптимальные стратегии хищника.** Предлагается простая экологическая задача, которую можно сформулировать в виде модели линейного программирования. Постулируется существование хищника, гнездо которого находится в точке А, и двух потенциальных источников пищи, расположенных на участках В и С. Время, необходимое хищнику для того, чтобы добраться до участков В и С и возвратиться с единицей добычи, полагается равным двум и трем минутам соответственно. С другой стороны, на участке В хищник затрачивает на поимку единицы добычи  $x_1$  две минуты, тогда как на участке С ему требуется лишь одна минута, чтобы поймать единицу добычи  $x_2$ . Энергетическая ценность одной единицы  $x_1$  оценивается в 25 Дж, а единицы  $x_2$  – в 30 Дж. Если мы введем теперь ограничение, состоящее в том, что на путь из гнезда в любой из участков и обратно хищник может затрачивать не более 120 мин в сутки и что на поиск жертв он может тратить не более 80 мин в сутки, то мы приходим к классической задаче линейного программирования. Упомянутые ограничения записываются в виде неравенств:

$$n_1x_1 + n_2x_2 \leq 120 \text{ (для времени в пути)}$$

и

$$m_1x_1 + m_2x_2 \leq 80 \text{ (для времени поиска пищи).}$$

Нужно записать также неявные ограничения  $x_1, x_2 \geq 0$ , так как хищник не может поймать отрицательное число жертв. При этих ограничениях мы хотим максимизировать целевую функцию

$$Z = (25x_1 + 30x_2) \text{ Дж.}$$

Эту частную задачу довольно легко решить графически, воспользовавшись записанными в виде неравенств ограничениями.

Постройте график по полученным неравенствам, отметьте, что таким образом, все имеющие смысл решения будут лежать в четырехугольнике  $OPQR$  на рис. 1, а максимум целевой функции достигается в точке, которая наиболее удалена от начала координат.

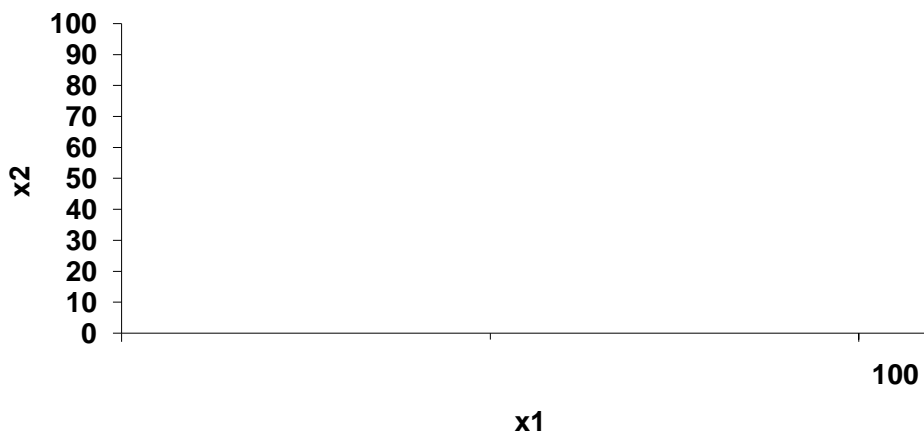


Рис 1. Графическое решение задачи линейного программирования

**Выводы.** Вычислите координаты точки с помощью графика, рассчитайте максимальное значение целевой функции  $Z$ . Обоснуйте полученный результат.

**P.S.** Студент может самостоятельно проверить, как влияет ослабление одного или обоих ограничений на целевую функцию, имея

в виду, что самым важным при оптимизации часто является нахождение того ограничения, ослабление которого позволяет найти еще лучшее решение.

Данный пример, разумеется, очень прост, и графический метод решения обычно можно применять лишь в том случае, когда целевая функция и ограничения зависят от двух переменных; правда, такие решения нередко можно найти и для случая нескольких ограничений.

## РАБОТА № 14. ТЕОРЕТИКО-ИГРОВЫЕ МОДЕЛИ

**Материалы и оборудование.** Расчетные таблицы, рабочий материал, калькуляторы.

**Пояснение.** Широкое применение в исследовании операций может найти и так называемая *теория игр*. Существует важное различие между собственно теорией игр и операционными играми, которые также стали использоваться для решения экологических проблем в последнее время. Термин «операционная игра» по существу обозначает моделирование некоторого очень сложного и изменчивого процесса принятия решений с помощью игры, проводимой по определенным образом закодированным правилам; в ней могут принимать участие и люди, но лучше проводить ее с воображаемыми участниками, решения которых определяются путем соответствующих вычислений. Теория же игр связана главным образом с методами и принципами выбора наилучшей стратегии в определенной ситуации конкурентной борьбы. Практических примеров конкурентного поведения в экологии, торговле и промышленности можно найти сколько угодно. Последние примеры большей частью оказываются слишком сложными для непосредственного анализа, хотя уже достигнуты кое-какие успехи в анализе ситуаций, в которых конкурирующие компании добиваются заключения контрактов, предлагая различные условия поставки. Теоретические исследования обычно посвящаются исключительно простым и искусственно подобранным ситуациям, однако именно по этим причинам они позволяют выявить некоторые наиболее важные аспекты реальных проблем. Чтобы получить некоторое представление о теории игр, рассмотрим следующий простой пример.

Допустим, что два лица А и В участвуют в определенной игре, конкретные детали которой нас не интересуют. Игроки стремятся к тому, чтобы выиграть каждую партию при ограничениях, налагаемых правилами игры. Целью, или конечным состоянием, каждой игры является выигрыш для А, выигрыш для В или ничья. В заключение каждой игры один из игроков платит другому определенную сумму денег, размер которой зависит от того, как закончилась игра. Ничья означает нулевые платежи, но сумма, которую получает игрок А в случае выигрыша, может зависеть от характера его победы. Как уже говорилось ранее, в теории игр нас интересуют не детали отдельных шагов или ходов какой-либо конкретной игры, а общий план

действий, или *стратегия*, которую нужно выбрать. Допустим, что игроку А известны три различные стратегии:  $A_1$ ,  $A_2$  и  $A_3$ , одну из которых он может выбрать, в то время как у игрока В имеются два возможных плана действий:  $B_1$  и  $B_2$ . Допустим теперь, что нам известно, какую сумму в конечном счете один игрок выплачивает другому, если выбрана какая-либо данная пара стратегий: например, игрок В выигрывает 20 очков у игрока А, если последний выбирает стратегию  $A_1$ , а игрок В выбирает стратегию  $B_1$ . Результаты для шести комбинаций стратегий можно представить в виде матрицы, в которой показаны суммы, выплачиваемые игроком В игроку А при выборе каждой пары стратегий.

*Платежная матрица* для данного примера имеет следующий вид:

|         |       | Игрок А |       |       |
|---------|-------|---------|-------|-------|
|         |       | $A_1$   | $A_2$ | $A_3$ |
| Игрок В | $B_1$ | -20     | -10   | 10    |
|         | $B_2$ | 20      | 30    | 20    |

Если нас интересует игрок В, то сразу видно, что стратегия  $B_2$  для него плоха, поскольку он всегда проигрывает и должен платить игроку А. Если же игрок В выбирает стратегию  $B_1$  то он выигрывает, когда игрок А выбирает стратегию  $A_1$  или стратегию  $A_2$ , но проигрывает 10 очков, когда игрок А поступает достаточно разумно и выбирает стратегию  $A_3$ . Наилучшей стратегией для игрока В всегда является  $B_1$  В то же время игроку А лучше всего придерживаться стратегии  $A_2$ , если игрок В выбирает стратегию  $B_2$ . Однако по изложенным выше причинам маловероятно, чтобы игрок В поступил таким образом. Если он выбирает стратегию  $B_1$  то игрок А получает максимальный выигрыш, действуя по плану  $A_3$ . Таким образом, наилучшими стратегиями для А и В оказываются такие, в результате выбора которых игрок В проигрывает игроку А 10 очков. Эта величина называется *ценой игры*. Если, как в данном случае, игроку лучше придерживаться одного и того же плана в каждой партии, то говорят, что он использует *чистую стратегию*.

Оптимальные стратегии можно найти с помощью полезного правила – *принципа минимакса*, согласно которому целесообразно выбирать стратегию, минимизирующую максимально возможный

проигрыш. Так, для игрока А наибольшими проигрышами являются +20, +10 и –10 при стратегиях  $A_1$ ,  $A_2$  и  $A_3$  соответственно. Наименьшим проигрышем здесь является –10. Следовательно, минимаксной стратегией является  $A_3$ . Применение аналогичных рассуждений к игроку В показывает, что для него максимальными проигрышами являются +10 и +30 при стратегиях  $B_1$  и  $B_2$  соответственно. Таким образом, минимаксной стратегией для игрока В является  $B_1$ .

Итак, получен тот же результат, что и ранее. Если стратегии игрока А соответствуют столбцам платежной матрицы, то для выбора оптимальной стратегии нужно найти элемент, являющийся наибольшим в своем ряду и наименьшим в своем столбце. Если такой элемент существует, то говорят, что это *седловая точка* и соответствующие стратегии являются наилучшими для двух рассматриваемых игроков.

Однако часто бывает так, что седловая точка отсутствует, т. е. не существует какой-либо пары стратегий, каждая из которых является оптимальной для данного игрока при всех обстоятельствах. В этом случае наилучший метод состоит в принятии игроками *смешанной стратегии*, которая означает, что каждый игрок случайным образом производит выбор определенного плана из числа всех возможных, причем вероятность выбора зависит от конкретного набора чисел, входящих в платежную матрицу. Полное решение таких задач может быть довольно сложным и осуществлено несколькими способами. Один из них состоит в превращении всей проблемы в задачу линейного программирования и использовании соответствующего метода решения этой преобразованной задачи.

**Задача 1.** Вычислите наиболее оптимальную стратегию поведения для задачи «Не сегодня ли день рождения жены?» (см. табл. 1).

Таблица 1 – Матрица игры, указывающая стратегии и исходы для задачи «Не сегодня ли день рождения жены?»

|     | Стратегия  | Природа                  |                       |
|-----|------------|--------------------------|-----------------------|
|     |            | День рождения не сегодня | День рождения сегодня |
| Муж | Без цветов | 0                        | -10                   |
|     | С цветами  | 1                        | 1,5                   |



Числа в таблице показывают исход игры с точки зрения мужа для каждой комбинации стратегий двух игроков.

Данная игра обладает той характерной особенностью, которая в теории игр называется «седловой точкой». Попросту говоря, когда наибольший из минимумов по строке совпадает с наименьшим из максимумов по столбцу, игра имеет седловую точку и игрокам всегда следует придерживаться той чистой стратегии, которая стоит на пересечении соответствующих строки и столбца.

**Выводы.** Выбрать: а) наилучшую стратегию игрока; б) наилучшую стратегию с учетом «седловой точки».

**Пояснение.** Поиск седловых точек является весьма важным моментом теории игр; вероятность того, что эти седловые точки будут существовать у случайно выбранной матрицы, велика для матриц невысоких порядков. В тех же случаях, когда никакой седловой точки не существует, можно показать, что решение следует искать в смешанных стратегиях. Это означает, что должны применяться две или более стратегии и что вероятности, с которыми данные стратегии применяются, могут быть вычислены по матрице игры. Каждый раз во время игры выбор стратегии должен осуществляться случайно, но с фиксированными вероятностями для соответствующих стратегий.

Уместно, конечно, спросить, почему мы рассматриваем природу как злого противника, стремящегося минимизировать выигрыш партнера, будь то человек, животное или растение. И все же в ситуациях, когда у нас не хватает знаний о реакции живых организмов или внешней среды на выбор стратегий, которые будут давать наилучший результат хотя бы в среднем, стоит применять некую комбинацию стратегий, консервативную в том смысле, что она минимизирует ущерб, причиняемый при наихудших стратегиях, которые может применить природа.

**Задача 2.** Стратегии ловли на удочку и питания.

Представим себе, что существование некоего вида рыб, питающихся у поверхности воды, зависит главным образом от наличия трех видов летающих насекомых. Эти виды – обозначим их через  $x_1$ ,  $x_2$  и  $x_3$  – у поверхности воды представлены не одинаково, а с частотами соответственно  $15n$ ,  $5n$  и  $n$ . Иными словами, насекомых  $x_2$  в пять раз больше чем  $x_3$ , а  $x_1$  – в три раза больше, чем  $x_2$ .

Допустим, что некто ловит рыбу на один из этих трех видов насекомых, насаживая их на крючок. Тогда исход игры с точки зрения рыбы для каждого из возможных сочетаний стратегий питания и

применяемой наживки могут быть такими, как приведенные в таблицу 2.

Таблица 2 – Матрица игры для стратегий ловли на удочку и питания

| Стратегия     |       | Рыболов использует в качестве наживки |       |       |
|---------------|-------|---------------------------------------|-------|-------|
|               |       | $x_1$                                 | $x_2$ | $x_3$ |
| Рыба питается | $x_1$ | -2                                    | 0     | 0     |
|               | $x_2$ | 0                                     | -6    | 0     |
|               | $x_3$ | 0                                     | 0     | -30   |

**Задача 3.** Измененная стратегия ловли на удочку и питания.

Если рыболов иногда начинает использовать приманку, которая может быть принята рыбой за любое из трех насекомых, но которая вдвое чаще вызывает подозрение у рыбы, то матрица игры изменится (см. табл. 3).

Таблица 3 – Матрица игры для измененных стратегий ловли на удочку и питания

| Стратегия     |       | Рыболов использует в качестве наживки |       |       |          |
|---------------|-------|---------------------------------------|-------|-------|----------|
|               |       | $x_1$                                 | $x_2$ | $x_3$ | приманка |
| Рыба питается | $x_1$ | -2                                    | 0     | 0     | -1       |
|               | $x_2$ | 0                                     | -6    | 0     | -3       |
|               | $x_3$ | 0                                     | 0     | -30   | -15      |

**Выводы.** Выбрать наилучшую стратегию игрока для обоих случаев из задач 2 и 3 и обосновать решение.

## ЛИТЕРАТУРА

1. Антонов, А.В. Системный анализ / А.В. Антонов. – М.: Высш. шк., 2008. – 454 с.
2. Гринин, А.С. Математическое моделирование в экологии / А.С. Гринин, Н.А. Орехов, В.Н. Новиков. – М.: ЮНИТИ-ДАНА, 2003. – 269 с.
3. Джефферс, Дж. Введение в системный анализ: применение в экологии / Дж. Джефферс. – М., 1981. – 256 с.
4. Образцов, А.С. Системный метод: применение в земледелии / А.С. Образцов. – М., 1991. – 289 с.
5. Ризниченко, Г.Ю. Математические модели в биофизике и экологии / Г.Ю. Ризниченко. – М., 2003. – 184 с.
6. Ризниченко, Г.Ю. Лекции по математическим моделям в биологии. – ч. 1, 2. / Г.Ю. Ризниченко. – М., НИЦ, 2002. – 300 с.
7. Спицнадель, В.Н. Основы системного анализа / В.Н. Спицнадель. – СПб., 2000. – 326 с.
8. Хомяков, П.М. Системный анализ / П.М. Хомяков. – М.: КомКнига, 2007. – 216 с.

# МЕТОДЫ ЭКОЛОГИЧЕСКИХ ИССЛЕДОВАНИЙ

*Методические указания по выполнению самостоятельных работ*

*Игорь Александрович Шадрин  
Наталья Валентиновна Фомина*

*Электронное издание*

Редактор Е.А. Андреева

Подписано в свет 20.06.2016. Регистрационный номер 403  
Редакционно-издательский центр Красноярского государственного аграрного университета  
660017, Красноярск, ул. Ленина, 117  
e-mail: rio@kgau.ru