

ПРИМЕНЕНИЕ МАТЕМАТИЧЕСКИХ МЕТОДОВ В ЭКОЛОГИЧЕСКИХ ИССЛЕДОВАНИЯХ ОБЪЕКТОВ РАЗЛИЧНОГО РАНГА

Черных К.Д.

с. Кинель-Черкассы Самарской области, ГБОУ СОШ №1 «ОЦ», 9 класс

Научный руководитель: Зимовец Т.И., ГБОУ СОШ №1 «ОЦ»

На современном этапе развития научных исследований становится явно недостаточно провести простое описание изучаемого явления. Обязательным требованием является такая форма результата, которая может быть применена на практике в виде математической модели, программного продукта и т.п. Учитывая это, в качестве основного метода исследования были выбраны математические методы.

В настоящей работе приводятся методы математической статистики и корреляционно-регрессионный анализ, примененные для анализа экологических опытов.

Прежде чем выполнять математические расчёты в любой системе, необходимо отметить, что математический анализ отличается объективностью и динамичностью.

В настоящее время становится всё более закономерным увеличение доли математических методов в современных исследованиях, именно поэтому тема исследования актуальна.

Целью работы является обоснование применения основных математических методов в экологических исследованиях.

Реализация цели осуществляется через ряд задач:

- изучить литературные источники по теме исследований;
- собрать статистический материал, провести опыты и наблюдения;
- обработать данные и написать научную работу.

В работе обобщены и систематизированы данные, полученные из фондовых источников Кинель-Черкасского района, НИИ «ВолгоГипрозема», и собственных наблюдений, выполненных на базе школьного научного общества.

Статистическая обработка данных экологического опыта

В экологических исследованиях отдельных растительных объектов, планктона, растительных сообществ, для выяснения правильности измерений и их объективности применяют различные математические методы. Более подробно остановимся на статистических методах [7, 10].

По завершении опыта и после получения всех данных встаёт вопрос о достаточ-

ности измеренных значений для того, чтобы делать выводы и выдвигать гипотезы. Хотя и говорят, что существуют в порядке возрастания ненадежности: вруны, заклётые лжецы и статистики, но рано или поздно приходится обращаться к статистике для истолкования полученных данных. Статистический анализ не только подтверждает полученные результаты, но, что более важно, он может реально помочь выполнению программы исследований. При умелом использовании статистики статистика поможет определить, действительно ли данные означают то, что хотелось бы получить, помогает полнее осмыслить исследование, чего иногда нельзя добиться иным путем, и часто может сэкономить время и труд, намечая последующие опыты.

Статистически достоверные числа

Для опыта требуется проведение серии измерений конечной длины стеблей растения, например, молодых растений бальзамина, произрастающих в различных экологических условиях. Мы тщательно измерили длину 11 стеблей растений линейкой с миллиметровой шкалой и каждое измерение указали в миллиметрах.

Результаты этого опыта приведены в табл. 1.

Таблица 1

Результаты опытов по измерению длины стеблей растений

Окончательная длина стебля, мм	Окончательная длина стебля, мм
14	16
15	19
13	20
11	19
10	15
18	

Независимо от вопроса о статистической достоверности этих данных, следует рассмотреть вопрос о точности полученных числовых данных.

При измерении длины стебля очевидно наличие какой-то неустранимой погрешности, которую обычно считают равной примерно $\pm 0,5$ мм. Среднюю длину стебля в ис-

следуемых образцах получают в каждом случае делением суммы всех полученных значений на 11. Точное деление показывает, что ни в одном случае среднее не является целым числом. Средняя длина стебля равна 15,636... Еще один момент, который следует учесть, – это то, что измерения длины стебля, как бы их ни выражали, используют как таковые.

Если бы потребовалось довести измерения длины до крайнего значения, ее можно было бы выразить с точностью до третьего или даже до шестого знака после запятой. Это было бы бессмысленно, особенно потому, что мы знаем о наличии неустранимой погрешности измерения. Поскольку эта ошибка может составлять $\pm 0,5$ мм, вполне оправдано указание среднего значения только до первого знака после запятой или в пределе до ошибки измерения, т. е. указание, что средняя длина стебля составляла 15,6 мм.

Вопрос о том, до какого знака следует рассчитывать среднюю величину – один из наиболее часто возникающих перед учеными; стандартной процедурой стало вести расчет до минимально допустимого числа знаков. За исключением особых случаев, все измерения даются не более чем до второго знака после запятой, но даже это может быть рискованным. Правильным является использование наименьшего возможного числа, причем надо быть готовым дать обоснование более чем двум цифрам после запятой [4, 7].

Рендомизация

Помимо трудностей, связанных с получением однородных экспериментальных систем, должен быть решен вопрос о размещении повторностей. Даже при наилучших условиях вполне возможны небольшие колебания микроклимата (температура, свет, влажность, ветер и т. д.) на различных участках площади. Например, в полевых опытах на рост растений могут сильно влиять разница в плодородии почвы, рельеф местности, степень загрязнения и т. д. В лаборатории у многих термостатов или камер для выращивания также имеются различия в микроклимате от точки к точке. Поскольку необходимо иметь достаточное число повторностей для каждой переменной, чтобы численность популяции была статистически достоверной, крайне важно, чтобы все неконтролируемые переменные были устранимы. Это и называется рендомизацией.

Предположим, что опыт включает 4 переменных (т.е. 4 разных концентрации ростового вещества, 4 разных сочетания азотистых субстратов, 4 разные длины волн излучения и т.д.). Предположим далее,

что все они размещаются на одной и той же площади поля, термостата или камеры для выращивания. Допустим, наконец, что для каждой переменной принято по 4 повторности (или серии повторностей).

Площадь, где ведется опыт, сначала делят на 16 равных частей. Затем нумеруют все повторности (табл. 2).

Таблица 2

Переменная №	Повторность №
1	а,б,в,г
2	а,б,в,г
3	а,б,в,г
4	а,б,в,г

Один из способов избежать пристрастности в размещении повторностей заключается в том, что в шапку кладут 16 пронумерованных кусочков бумаги и размещают каждую из 16 повторностей согласно вытянутому номеру.

Второй способ называется методом латинского квадрата. Этот квадрат представляет собой распределение из n переменных X n повторностей таким образом, что каждая из букв встречается один раз в каждом ряду и один раз в каждом столбце. В рассмотренном выше примере $n \times n$ размещение будет 4×4 (табл. 3).

Таблица 3

1	2	3	4
а б в г	а б в г	а б в г	а б в г
б а г в.	б в г а	б г а в	б а г в
в г б а	в г а б	в а г б	в г а б
г в а б	г а б в	г в б а	г в б а

Могут быть латинские квадраты 5×5 и т. д., построенные таким же образом.

Разновидностью латинского квадрата является так называемый греко-латинский квадрат, где каждая из переменных имеет свой индекс. Снова взяв тот же пример из четырех переменных, можно использовать греко-латинский квадрат, для рендомизации групп повторностей в рамках переменных (табл. 4) [4, 10].

Таблица 4

а1	б2	в3	г4
б4	а3	г2	в1
в2	г1	а4	б3
г3	в4	б1	а2

Квадрат такого типа можно также увеличить до 5×5 , 7×7 , 10×10 и т.д., но только не 6×6 .

Имеются еще варианты, которыми можно пользоваться. Некоторые ученые пользуются латинским квадратом, в котором

размещение повторностей производится по ходу какой-либо шахматной фигуры, например коня или слона. Они удобны в тех случаях, когда возможно постоянное отклонение в определенной позиции. Примером может служить квадрат 5X5, полученный ходом коня (табл.5).

Таблица 5

Латинский квадрат	Греко-латинский квадрат	Квадрат ходом коня
а б в г д	а1 б2 в3 г4 д5	а б в г д
б а д в	б4 в5 г1 д2 а3	г д а б в
г в г а д	в2 г3 д4 а5 б1	б в г д а
б г д б	г5 д1 а2 б3 в4	д а б в г
а в д в г	д3 а4 б5 в1 г2	в г д а б

Третий способ рендомизации – использование случайных чисел. Таблицы случайных чисел состоят из ряда однозначных чисел, в которых каждое число встречается примерно с одной и той же частотой и распределение которого не подчиняется никаким закономерностям. Таблицы случайных чисел в перестановках, начиная с 6 и выше, можно найти в большинстве книг по математике. Пример можно видеть в табл. 6.

Таблица 6

5	5	6	7	1
4	1	2	8	2
9	3	3	2	9
7	9	7	4	3
1	6	9	6	5
6	4	4	3	6
8	7	8	1	7
3	2	1	9	4
2	8	5	5	8

Стандартные отклонения

В любой серии измерений имеется высокая степень вероятности, что не все из полученных значений будут одинаковыми в любом данном варианте. Частично это обусловлено, как мы уже видели ранее, изменчивостью, свойственной самому биологическому объекту вследствие его генетической структуры, небольшими колебаниями в окружающих условиях и в самих вариантах. К этим источникам изменчивости или неопределенности можно еще добавить ошибки измерения [5].

Обычно довольно важно уметь выразить эту изменчивость математическими или статистическими терминами так, чтобы иметь некоторое представление о рассеивании значений в ту или иную сторону от средней величины. Наиболее частым критерием рассеивания изменчивости служит так на-

зываемое стандартное (или средне-квадратичное) отклонение. Пример вычисления стандартного отклонения приведен ниже. Предположим, что подсчитывается число растений на площади 1 м, чтобы получить представление о растительном покрове большой территории. Допуская, что климат и т. д. совершенно одинаковы на всей территории, исследователь выделит 10 участков по 1 м и подсчитывает число растений на каждом из них. Полезно иметь представление о том, каковы могут быть отклонения в численности (табл. 7).

Таблица 7

Число растений на 1 м²

130	60
120	170
160	130
140	140
120	90

Среднее число растений получают, сложив все цифры в колонке и разделив их на *n*, т. е. число измерений: 1260: 10=126.

Хотя среднее число равно 126 растений на единицу площади, ясно, что имеются колебания от 170 до 60 растений/м². Вычислив стандартное отклонение, можно выразить эту изменчивость.

Стандартное отклонение (часто обозначаемое сокращенно *S.0.*) получают, разделив сумму всех отклонений от средней величины, возведенных в квадрат, на число наблюдений минус 1, и извлекая затем квадратный корень из этого числа. Математически это будет выглядеть так:

$$D.S.=\sum \Delta.$$

Возвращаясь к рассмотренному примеру можно составить приводимую ниже табл. 8. Следует помнить, что эта операция должна быть проделана со всеми использованными числами или значениями вне зависимости от того, будут ли отклонения отрицательными или положительными, поскольку при возведении в квадрат знаки устраняются. Сумма третьей колонки составляет 8740. Это число делят затем на число измерений, уменьшенное на единицу, т. е. на *n-1 = 9*; отсюда 8740:9=971. Теперь для того, чтобы получить стандартное отклонение, нужно извлечь квадратный корень из 971. Для наших целей можно округлить квадратный корень из 971 до 31. Это число 31 и будет являться стандартным отклонением для десяти определений числа растений на 1 м. В таблицах его обычно приводят как среднее число плюс или минус стандартное отклонение, т. е. 126+31 растений/м².

Таблица 8

Число растений на 1 м ²	Отклонения от среднего (Δ)	Квадрат отклонения от среднего (Δ^2)	Число растений на 1 м ²	Отклонения от среднего (Δ)	Квадрат отклонения от среднего (Δ^2)
130	4	16	60	66	4356
120	6	36	170	44	1396
160	34	1196	130	4	16
140	14	196	140	14	196
120	6	36	90	36	1296

Статистическая достоверность

В большинстве экологических опытов необходимо знать не только масштабы изменчивости в серии повторностей, но и уметь судить о том, насколько «реальны» различия, выявленные между переменными или среди них. В некоторых случаях это видно на глаз. Например, если вариант а способствует росту, а вариант б нет, то можно полагать, что разница между а и б реальна (если это действительно единственные переменные величины). Ответы типа «да» или «нет», хотя желательны, но их редко получают и, как правило, получают ответ типа «может быть». Оценку этого «может быть» называют определением статистической достоверности. В основном, статистическая достоверность свидетельствует о том, что имеется вероятность реальности наблюдаемых различий. Экологи, осознающие неизбежную изменчивость биологического материала и прекрасно зная, что незначительная разница в схеме опыта и условиях его проведения может сильно повлиять на результаты, решили принять лишь определенные уровни вероятности в качестве указания на реальность различий. Например, если в серии повторных опытов установлено, что вариант а отличается от варианта б только один раз из десяти, было бы глупо настаивать, что вариант а действительно отличается от варианта б. Если же, с другой стороны, вариант а дал результаты, последовательно и значительно отличающиеся от результатов в варианте б, то вероятно, что они действительно различны. Вопрос заключается лишь в том, где подвести черту [4,5,7,10].

Имеется ряд статистических методов, которыми можно пользоваться для оценки полученных данных, причем большинство их достаточно просто, чтобы оправдать их применение в большинстве экологических опытов. Математическое обоснование разных методов не входит в задачи данного исследования.

Здесь можно проиллюстрировать лишь два важных примера. Предположим, что проводится сравнение влияния интенсивности накопления свинца на удлинение стебля фасоли. Вариант а состоял из 25 растений, выращиваемых при содержании ниже кларка, а в варианте б растения выращивали при содержании выше кларка. Были приняты достаточные меры для контролирования факторов, как генетических, так и внешней среды, и рендомизации повторностей. После измерения длины стеблей были вычислены средняя длина стебля и стандартное отклонение от среднего:

Варианта а $15,0 \pm 1,4$ см

вариант б $8,0 \pm 0,8$ см

Разница между этими средними ($15,0 - 8,0$) равна $7,0$, а среднее из двух стандартных отклонений равно $1,1$. Разделив разность между средними на среднее стандартное отклонение, мы получим меру отношения оцениваемой разницы, которая в данном опыте составляет $7,0:1,1 = 6,36$. Это очень большая разница, показывающая, что различия между повторностями малы по сравнению с разницей между вариантами. Чтобы определить, насколько они малы, следует обратиться к табл. 9, где указаны пределы статистической достоверности для отношений различий.

Таблица 9

Отношение оцениваемых различий	Уровень статистической достоверности, %	Отношение оцениваемых различий	Уровень статистической достоверности, %
1,28	20	2,58	1
1,64	10	2,81	0,5
1,96	5	3,09	0,2
2,33	2	3,29	0,1

Можно видеть, что разница между вариантами а и б статистически достоверна при уровне еще меньшем, чем 0,1%.

Во втором примере ставится опыт по сравнению влияния двух источников питательных веществ на рост гриба. Снова, при достаточном числе повторностей и внимательном подходе к рендомизации и другим факторам, опыт проводят и определяют вес сухого вещества, рассчитывают, как обычно, стандартное отклонение и получают следующие результаты:

вариант а ... 18,1 ± 2,1 мг сухого вещества
вариант б... 14,7 ± 1,9 мг сухого вещества

Здесь разница между средними равна 3,4 мг, а среднее стандартное отклонение равно 2,0. Отношение оцениваемых различий равно 3,4:2,0=1,70. Справка по таблице статистической достоверности покажет, что это отношение значимо достоверно где-то между 10%-м и 20%-м уровнями.

Итак, в первом опыте разница была достоверной при уровне, лучшем 0,1%-го, в то время как во втором опыте результаты были достоверны примерно при 15%-м уровне. Это значит, что если бы первый опыт повторили 1000 раз, то, вероятно, в 999 случаях установили бы разницу между вариантами. Короче говоря, можно быть уверенным, что наблюдаемая разница, видимо, реальна. Во втором примере повторение опыта, скорее всего, покажет, что в одном из семи опытов не будет установлено никакой разницы. Короче, нельзя быть уверенным, что разница реальна.

В экологии принято, что достоверными результатами можно считать только те, для которых уровень статистической достоверности (которая фактически является мерой вероятности) равен по меньшей мере 5%, т. е., что только в одном из каждых 20 опытов не будет установлено разницы. Вероятность, что разница реальна, будет возрастать при уменьшении уровня до 1%, 0,1% и т. д.

Корреляционно-регрессионный анализ в экологических системах

В экологических исследованиях более высокого ранга для выявления закономерных взаимных влияний используется корреляционно-регрессионный анализ [8] и метод построения классификационных образов [9].

Построение уравнения регрессии сводится к оценке ее параметров. Для оценки

параметров регрессий, линейных по параметрам, применяется метод наименьших квадратов. Он позволяет получить такие оценки параметров, при которых сумма квадратов отклонений фактических значений результативного признака $y(x_i) = a + bx$ от теоретических значений минимальна, т. е.

$$\sum (y - \hat{y}(x))^2 \rightarrow \min. \quad (1)$$

Тесноту связи изучаемых явлений оценивает линейный коэффициент парной корреляции, принимающий значение от [-1;1]:

$$r_{xy} = b \frac{\sigma_x}{\sigma_y} = \frac{\text{cov}(x, y)}{\sigma_x \sigma_y} \quad (2)$$

Для оценки качества полученной модели используется коэффициент детерминации, который показывает долю дисперсии, объясняемую регрессией, в общей дисперсии результативного признака.

Оценка статистической значимости полученного уравнения регрессии проводится путем прохождения F-теста, который состоит в проверке гипотезы Но об отсутствии статистической значимости уравнения регрессии и показателя тесноты связи. Для этого выполняется сравнение фактического $F_{\text{факт}}$ и критического (табличного) $F_{\text{табл}}$ значения F-критерия Фишера.

Если $F_{\text{табл}} < F_{\text{факт}}$, то гипотеза Но о случайной природе оцениваемых характеристик отклоняется и признается их статистическая значимость и надежность.

Для оценки статистической значимости коэффициентов регрессии и корреляции рассчитывается t-критерий Стьюдента и доверительные интервалы каждого из показателей.

Сравнение табличного и фактического значения критерия Стьюдента осуществляется аналогично тому, как это делается с критерием Фишера.

Для расчета доверительного интервала определяется предельная ошибка для каждого показателя. Тогда доверительные границы можно установить в виде:

$$\gamma = a \pm \Delta_a; \gamma_b = b \pm \Delta_b \quad (3)$$

В случае, если строится множественная регрессия заболеваемости по факторам $y = f(x_1, x_2, \dots, X_p)$, система нормальных уравнений МНК имеет вид:

$$\begin{cases} \sum y = na + b_1 \sum x_1 + b_2 \sum x_2 + \dots + b_p \sum x_p, \\ \sum yx_1 = a \sum x_1 + b_1 \sum x_1^2 + b_2 \sum x_2 x_1 + \dots + b_p \sum x_p x_1, \\ \dots \\ \sum yx_p = a \sum x_p + b_1 \sum x_1 x_p + b_2 \sum x_2 x_p + \dots + b_p \sum x_p^2. \end{cases} \quad (4)$$

Однако по коэффициентам чистой регрессии нельзя сказать, какой из факторов сильнее влияет на результат. Для этого используется уравнение регрессии в стандартизованном масштабе

$$t_y = \beta_1 t_{x_1} + \beta_2 t_{x_2} + \dots + \beta_p t_{x_p},$$

где $t_y = \frac{y - \hat{y}}{\sigma}$ – стандартизованные вели-

чины, β_i – стандартизованные коэффициенты регрессии.

Связь между коэффициентами чистой регрессии и стандартизованными коэффициентами описывается соотношением

$$b_i = \beta_i \frac{\sigma_y}{\sigma_{x_i}}.$$

Индекс множественной корреляции R [0; 1] определяется как корень квадратный из доли объясненной дисперсии, а для стандартизованных коэффициентов может быть найден по формуле

$$R_{yx_1x_2\dots x_p} = \sqrt{\sum \beta_i r_{yx_i} \geq \max r_{yx_i}} \quad (6)$$

При линейной зависимости коэффициент множественной корреляции можно определить через матрицу парных коэффициентов корреляции:

$$R_{yx_1x_2\dots x_p} = \sqrt{1 - \frac{\Delta r}{\Delta r_{11}}}, \quad (7)$$

где

$$\Delta r = \begin{vmatrix} 1 & r_{yx_1} & r_{yx_2} & \dots & r_{yx_p} \\ r_{yx_1} & 1 & r_{x_1x_2} & \dots & r_{x_1x_p} \\ r_{yx_2} & r_{x_2x_1} & 1 & \dots & r_{x_2x_p} \\ \dots & \dots & \dots & \dots & \dots \\ r_{yx_p} & r_{x_px_1} & r_{x_px_2} & \dots & 1 \end{vmatrix}$$

$$\Delta r_{11} = \begin{vmatrix} 1 & r_{x_1x_2} & \Delta & r_{x_1x_p} \\ r_{x_2x_1} & 1 & \Delta & r_{x_2x_p} \\ \Delta & \Delta & \Delta & \Delta \\ r_{x_px_1} & r_{x_px_2} & \Delta & 1 \end{vmatrix}$$

Для определения влияния на заболеваемость конкретного фактора x , при неизменном уровне других факторов применяются частные коэффициенты корреляции.

Качество построенной модели в целом определяется коэффициентом множественной детерминации, определяемым как квадрат индекса множественной корреляции $R^2_{yx_1x_2\dots x_p}$.

Статистическую значимость уравнения множественной регрессии в целом оценивается с помощью F-критерия Фишера.

Оценка значимости коэффициентов чистой регрессии с помощью t-критерия Стьюдента сводится к вычислению значения данного критерия.

При построении уравнения множественной регрессии может возникнуть проблема мультиколлинеарности факторов, их тесной линейной связанности. Примем, что две переменные явно коллинеарны, т.е. находятся между собой в линейной зависимости, если $r_{x_1x_2} > 0,7$. Для оценки мультиколлинеарности факторов строится матрица межфакторной корреляции.

Основным методом, позволяющим установить взаимосвязь между уровнем заболеваемости и внешними факторами, является корреляционно-регрессионный анализ. Опыт многочисленных исследований показывает, что наиболее удобными являются парные регрессионные модели, а из множественных – те, в которых участвует не более двух или трех переменных. В качестве уравнений связи уровня заболеваемости (далее обозначается $y(x_i)$) и внешних факторов (далее x_i) могут применяться как линейные, так и нелинейные формы зависимости.

Однако многие нелинейные модели сводятся к линейным, поэтому общая теория будет рассматриваться для линейных систем.

Важным фактором, влияющим на уровень заболеваемости, является наличие в почве района различных микроэлементов в разнообразных количествах. Однако построение многофакторной модели в данном случае нерационально в силу высокой трудоемкости модели и ее низкой точности.

По данным проведенных исследований сельскохозяйственных угодий района, для Кинель-Черкасского района имеются непрерывные данные по содержанию в почве микро- и макроэлементов за период с 1997 по 2011 гг. [1,3,6].

Исходные данные для построения регрессионной модели приведены в приложении 1.

Для уточнения картины заболеваемости была проведена работа по установлению связи между основными геохимическими параметрами территории и уровнем заболеваемости за длительный период (1997–2011 гг.) с помощью множественной линейной регрессии [9]. В качестве действующих

на заболеваемость факторов $\{X_i\}$, были выбраны 8 рядов макро- и микроэлементов, концентрация которых в почвах района измерялась в мг/кг, а заболеваемость – Y – случаев в год. Анализируя заболеваемость **группы кардиологические заболевания**, отмечается наиболее высокий уровень корреляции между заболеваемостью и концентрацией азота в почвах (0,830; прямая), а также с концентрацией серы (-0,841; обратная). Значительная корреляция (-0,746; обратная) между кардиологическими заболеваниями и кобальтом, однако ее значение ниже. Связь оценивается как существенная.

Анализируя заболеваемость **группы новообразования**, отмечается наиболее высокий уровень корреляции между заболеваемостью и концентрацией азота (0,915; прямая), а также с концентрацией серы (-0,883; обратная). Корреляция между новообразованиями и кобальтом (-0,578; обратная) низка, что особенно интересно, так как большинство исследователей на других территориях получают другие результаты. Помимо регрессии по калию и сере, объясняющей 89% вариации заболеваемости, автором была проведена попытка построения четырехмерной регрессии по калию (X_2), сере (X_4) марганцу (X_5), кобальту (X_8). Объясняющие факторы покрывают 91,1% дисперсии заболеваемости. Стандартизированные коэффициенты регрессии говорят о том, что все перечисленные факторы влияют на снижение заболеваемости Новообразованиями, т.к. все коэффициенты имеют отрицательный знак. Наиболее эффективным регулятором является сера (-1,013), на втором месте по значимости – калий (-0,322), далее марганец (-0,129). Крайне незначительное влияние имеет кобальт.

Анализируя заболеваемость **группы болезни органов дыхания**, не выявляется яркой корреляции между проявлениями заболеваний этой группы и концентрациями микроэлементов. Это свидетельствует в пользу того, что причины заболеваний этой группы нужно искать в климатической составляющей. Тот факт, что некоторая связь все же имеется, объясняется тем, что иммунная реакция организма зависит от наличия в почве и в питании человека микро- и макроэлементов.

Анализируя заболеваемость **группы болезни органов пищеварения**, отмечается наиболее высокий уровень корреляции между заболеваемостью и концентрацией азота в почвах (0,821; прямая), а также с концентрацией серы (-0,881; обратная). Корреляция между болезнями органов пищеварения и кобальтом (-0,515; обратная), а также между болезнями органов пищеварения и марганцем в почвах (0,333; прямая).

Анализируя заболеваемость **группы инфекционные и паразитарные**, отмечается наиболее высокий уровень корреляции между заболеваемостью и концентрацией калия в почвах (0,494; прямая), однако этот уровень недостаточен для объяснения. Отсюда следует, что объяснение группе инфекционные и паразитарные связано миграциями паразитических элементов из Похвистневского природно-эпидемического очага с перемещающимися животными, прежде всего мышевидными грызунами и лисами.

Анализируя группы диагнозов **кардиологические заболевания, новообразования и болезни органов пищеварения**, прослеживается следующая закономерность: для всех этих групп диагнозов большое значение имеет концентрация калия и серы.

Построение классификационных образов

Для более полного и комплексного анализа связи заболеваемости населения с ландшафтами можно провести построение классификационных образов на основе алгоритма кратчайшего расстояния [9].

В качестве объектов здесь выступают типы местности Кинель-Черкасского района: А – приречный, В – плакорный, С – пойменный, D – надпойменно-террасовый [2]. Для каждого типа местности вычислено среднее по выборке значение вероятности нахождения особи в различных состояниях – «здоров», «умер», «острый», «хроник» – рассчитанных на основе марковских процессов. Именно вследствие осреднения на средних величинах не наблюдается столь же значительных различий между типами местности, как в случае анализа территориального размещения вероятностей.

В качестве пространства состояний выбрана плоскость (x, y) , где x – вероятность нахождения организма в определенном состоянии через 1 год, y – вероятность нахождения организма в том же состоянии через 2 года после начала прогнозирования.

В качестве меры различия выбрано классическое евклидово расстояние между двумя объектами вида

$$R(k, l) = \sqrt{\sum_{i=1}^n (x_i^{(k)} - x_i^{(l)})^2}$$

где k, l – сравниваемые объекты или классы, x -признак объекта, n – количество признаков сравнения (приложение 2).

Выводы

Был выполнен обзор существующих литературных источников по проблеме применения математических методов в экологии и выбраны для примера некоторые из них.

Экологические факторы влияют на объёмы разного ранга и их объективная оценка возможна только с помощью математических методов.

Выявленные, математически достоверные связи, между отдельными заболеваниями и геохимическим фоном в почвах Кинель-Черкасского района, можно использовать в практических целях. Например, для внесения соответствующих удобрений. Это повлияет на уровень заболеваемости и затрагивает сразу два национальных проекта «Развитие АПК» и «Здоровье».

Подобного рода работы очень актуальны: системны, разноплановы, обоснованы. Другое дело, что реализация научных разработок – дело дорогостоящее и требует взвешенных, коллективных решений на уровне региональных властей. В любом случае, дорога пройдёт идущий.

Список литературы

1. Агрохимия / Под ред. Г.Б. Ягодина. 2-е изд., перераб. и доп. – М.: Агропромиздат, 1989. – 654 с.

2. Зейдис И.М., Кружалин В.И., Симонов Ю.Г. и др. Общие свойства динамики геосистем // Вестник МГУ. – 2001. – Сер. 5 – №4. – С. 3 – 9.

3. Клиническая оценка лабораторных тестов: Пер. с англ. / Под ред. Н.Т. Тица. – М.: Медицина, 1986. – 480 с.

4. Лапко А.В. Статистические методы моделирования и принятия решений в развивающихся медико-биологических системах. /А.В. Лапко, О.М. Новиков, Л.С. Поликарпов; Отв. ред. К.Р. Седов, А.М. Федотов; АН СССР, Сиб. отделение, Краснояр. науч. центр, ВЦ. – Новосибирск: Наука, Сибирское отд-ние, 1991. – 216 с.

5. Лядов В.Р. Основы теории вероятностей и математической статистики: для студентов мед. ВУЗов. – СПб.: Фонд «Инициатива», 1998. – 107 с.

6. Матвеев Н.М., Павловский В.А., Прохорова Н.В. Экологические основы аккумуляции, тяжёлых металлов сельскохозяйственными растениями в лесостепном и степном Поволжье. – Самара: Изд-во Сам. ун-т, 1997. – 215 с.

7. Урбах В.Ю. Статистический анализ в биологических и медицинских исследованиях. – М.: Медицина, 1975. – 295 с.

8. Ферстер Э., Рёниц Б. Методы корреляционного и регрессионного анализа: Руководство для экономистов / Пер. с нем. – М.: Финансы и статистика, 1983. – 302 с.

9. Шиманчик И.П. Некоторые аспекты моделирования медико-географической ситуации как случайного процесса // Шестое Сибирское совещание по климатологическому мониторингу. Материалы совещания. – Томск, 2005. – С. 447–450.

10. Юнкеров В.И., Григорьев С.Г. Математико-статистическая обработка данных медицинских исследований. – СПб.: ВМедА, 2002. – 266 с.

Приложение 1

Влияние микроэлементов на заболеваемость в Кинель-Черкасском районе

№	Показатель	Ед. изм	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	Содержание азота в почвах района (X4)	мг/кг	350,0	315,0	250,0	200,0	226,0	360,0	279,0	418,0	370,0	628,0	898,0	640,0	620,0	592,0	513,0
2	Содержание калия в почвах района (X2)	мг/кг	10,921	11,123	11,358	10,526	9,985	8,522	8,510	10,286	9,654	9,000	8,215	8,000	13,383	13,571	13,312
3	Содержание фосфора в почвах района (X3)	мг/кг	6,512	6,076	6,369	6,210	5,980	5,810	5,830	6,100	5,987	5,874	5,526	5,421	9,591	9,884	8,846
4	Уд. содержание серы в почвах района (X4)	мг/кг	3,156	3,269	3,382	3,383	3,256	3,356	3,351	3,022	3,122	2,889	2,665	2,512	2,331	2,411	2,651
5	Уд. содержание марганца в почвах района (X5)	мг/кг	11,540	8,360	14,320	16,190	9,830	19,700	11,959	17,094	13,665	14,975	10,088	14,675	15,791	19,182	17,689
6	Уд. содержание меди в почвах района (X6)	мг/кг	0,196	0,202	0,198	0,169	0,174	0,119	0,111	0,156	0,198	0,154	0,211	0,178	0,178	0,184	0,179
7	Уд. содержание цинка в почвах района (X7)	мг/кг	1,235	1,198	1,211	1,633	1,663	1,451	1,722	1,954	1,813	1,497	1,265	1,069	1,320	1,341	1,673

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
8	Уд. содержание кобальта в почвах района (X84)	мг/кг	0,158	0,152	0,164	0,169	0,172	0,154	0,147	0,145	0,139	0,152	0,142	0,137	0,148	0,152	0,145
9	Заболеваемость КАРДИО (YD)	сл/год	4541	4720	3559	3950	4205	4752	5178	6267	7401	7656	7739	7906	7388	7185	7422
10	Заболеваемость НОВЫЕ (Y2)	сл/год	1401	1345	1203	1227	1283	1231	1200	1586	1405	1610	1895	1818	1712	1645	1498
11	Заболеваемость БОД (Y3)	сл/год	14478	17202	15745	13580	12256	13751	14299	13081	16374	15094	17556	17816	17488	18132	18251
12	Заболеваемость БОП (Y4)	сл/год	3925	4072	4705	4836	5063	5238	5477	5866	5960	7667	8468	8207	7665	8541	8614
13	Заболеваемость ПАРАЗ (Y5)	сл/год	2696	2459	2727	2237	2569	2431	2179	2368	2393	2508	2292	2372	2427	2601	2697
14	Общая заболеваемость (Y6)	сл/год	41325	40795	44114	41126	44562	46897	49871	45710	59764	59884	70216	71247	68736	65530	69225

Приложение 2

Классификационные образы по типам местности

