

УДК 598.2; 504.05/.06

РАЗРАБОТКА ПРОГРАММЫ ДЛЯ ПРОВЕДЕНИЯ АНАЛИЗА РАСПРЕДЕЛЕНИЯ ЧИСЛЕННОСТИ ПТИЦ НА ИЗУЧАЕМОЙ ТЕРРИТОРИИ С ИСПОЛЬЗОВАНИЕМ РЕГРЕССИОННОГО АНАЛИЗА

Еремеев В.С., д.т.н., профессор, Федеральное государственное бюджетное образовательное учреждение высшего образования «Мелитопольский государственный университет», г. Мелитополь (272312, Российская Федерация, Запорожская область, г.о. Мелитополь, г. Мелитополь, пр-т Богдана Хмельницкого, д. 18), e-mail: evs1938@yandex.ru

Ступницкий В.С., преподаватель, Федеральное государственное бюджетное образовательное учреждение высшего образования «Мелитопольский государственный университет», г. Мелитополь (272312, Российская Федерация, Запорожская область, г.о. Мелитополь, г. Мелитополь, пр-т Богдана Хмельницкого, д. 18), e-mail: stupnitskiy@gmail.com

Гончаров А.Ю., инженер, начальник отдела технической поддержки пользователей, Федеральное государственное бюджетное образовательное учреждение высшего образования «Мелитопольский государственный университет», г. Мелитополь (272312, Российская Федерация, Запорожская область, г.о. Мелитополь, г. Мелитополь, пр-т Богдана Хмельницкого, д. 18), e-mail: honcharov.alexandr@yandex.ru

Сукманов А.В., к.т.н., доцент, Северо-Кавказский филиал федерального государственного бюджетного образовательного учреждения высшего образования «Белгородский государственный технологический университет им. В.Г. Шухова» (г. Минеральные Воды) (357202, Российская Федерация, Ставропольский край, г. Минеральные Воды, ул. Железнодорожная, д. 24), e-mail: capoaro@yandex.ru

Предложен алгоритм построения уравнения регрессии для обработки результатов мониторинга птиц на исследуемой территории в виде многомерного многочлена методом наименьших квадратов. На алгоритмическом языке Python разработана интерактивная программа `Birds_MLS`, обеспечивающая расчёт коэффициентов уравнения с использованием библиотек NumPy, Pandas, SciPy, Matplotlib, PyQt6. Программа содержит полный цикл вычислений, начиная от введения исходных данных, построения регрессионной модели, и заканчивая проверкой её достоверности и визуализации результатов. Достоверность предлагаемой модели проводится с помощью индекса детерминации R^2 и критерия Фишера F. Величина критериев адекватности определяется путём сравнения известной зависимости функции $u(x, y)$ от двух координат с результатами расчётов, которые получены с помощью уравнения регрессии $v(x, y)$. Тестирующие эксперименты подтвердили адекватность предлагаемого алгоритма при изучении зависимости численности от нескольких переменных в многофакторных задачах.

Ключевые слова: уравнение регрессии, коэффициент корреляции, метод наименьших квадратов, птицы, статистическая обработка

DEVELOPMENT OF A PROGRAM FOR ANALYZING THE DISTRIBUTION OF BIRD NUMBERS IN THE STUDY AREA USING REGRESSION ANALYSIS

Eremeev V.S., D. Sci. (Eng.), Professor, Federal State Budgetary Educational Institution higher education «Melitopol State University», Melitopol (18, Bogdana Khmel'nitskogo Avenue, Melitopol, Zaporozhye Region, 272312, Russian Federation), e-mail: evs1938@yandex.ru

Stupnitsky V.S., Teacher, Federal State Budgetary Educational Institution higher education «Melitopol State University», Melitopol (18, Bogdana Khmel'nitskogo Avenue, Melitopol, Zaporozhye Region, 272312, Russian Federation), e-mail: stupnitskiy@gmail.com

Goncharov A.Yu., Engineer, Federal State Budgetary Educational Institution higher education «Melitopol State University», Melitopol (18, Bogdana Khmel'nitskogo Avenue, Melitopol, Zaporozhye Region, 272312, Russian Federation), e-mail: honcharov.alexandr@yandex.ru

Sukmanov A.V., Cand. Sci. (Eng.), Associate Professor, North Caucasus branch of the Federal State Budgetary Educational Institution of Higher Education «Belgorod State Technological University named after V.G. Shukhov», Mineralnye Vody (24, Zheleznodorodskaya str., Mineralnye Vody, Stavropol Region, 357202, Russian Federation), e-mail: capoaro@yandex.ru

An algorithm for constructing a regression equation for processing the results of bird monitoring in the study area in the form of a multidimensional polynomial by the least squares method is proposed. An interactive program Birds_MLS is developed in the Python algorithmic language, which provides the calculation of the equation coefficients using the NumPy, Pandas, SciPy, Matplotlib, and PyQt6 libraries. The program contains a full cycle of calculations, starting with the introduction of initial data, building a regression model, and ending with checking its validity and visualizing the results. The validity of the proposed model is assessed using the R^2 determination index and the F-Fisher criterion. The adequacy criteria are determined by comparing the known dependence of the $u(x, y)$ function on two coordinates with the results of calculations obtained using the $v(x, y)$ regression equation. The testing experiments confirmed the adequacy of the proposed algorithm when studying the dependence of population size on multiple variables in a multivariate setting.

Keywords: correlation coefficient, least squares method, regression equation, birds, statistical processing

Введение

Азово-Черноморское побережье является одним из важнейшим в Евразии трансконтинентальным миграционным коридором между Африкой, Европой и Азией. Оно обладает уникальным сочетанием прибрежных зон, водно-болотных угодий, степей и модифицированных биотопов, которые обеспечивают птицам идеальные места для отдыха, кормежки и размножения многих видов [1]. На побережье Азовского моря в Херсонской и в Запорожской областях функционируют несколько ветровых электростанций, которые могут оказывать влияние на миграционные процессы птиц в весенний и осенний периоды [2,3]. Вероятность столкновения птиц с ветровыми колёсами турбин определяется плотностью их распределения, высотой полёта, погодными условиями и многими другими факторами. Поэтому разработка и внедрение комплекса информационных технологий по влиянию климатических изменений и антропогенного воздействия на численность птиц и их миграцию в Приазовском регионе с помощью дистанционного зондирования Земли, геоинформационных технологий и математического моделирования имеет большой практический и теоретический интерес.

Математические модели широко применяются в физике, химии, технических науках, экономике, социологии [4,6]. Для проведения обработки результатов мониторинга птиц на исследуемой территории воспользуемся методом регрессионного анализа [7-9]. Предположим, что распределение плотности птиц в некоторой области D определяется неизвестной непрерывной функцией от m переменных $u(x_1, x_2, \dots, x_m)$ с непрерывными частными производными. В качестве переменных могут служить географические координаты, время учёта, высота полёта, полётные характеристики птиц и многие другие факторы [1-3]. Уравнение регрессии, определяющее эту зависимость с некоторой точностью δ , представим в виде многомерного аппроксимирующего многочлена $v(x_1, x_2, \dots, x_m)$. Вид и степень многочлена зависит от решения конкретной задачи. Пусть, например, требуется найти зависимость плотности распределения птиц от географических координат x, y в виде многочлена третьей степени

$$v(x, y) = a_0 + a_x x + a_y y + a_{xy} xy + a_{xx} x^2 + a_{yy} y^2 + a_{xxy} x^2 y + a_{xyy} x y^2 + a_{xxx} x^3 + a_{yyy} y^3 + \delta. \quad (1)$$

Вычисление 10 неизвестных коэффициентов регрессии в формуле (1) потребует информации о численности птиц не менее чем в 11 точках. В более общем случае, когда требуется провести регрессионный анализ зависимости количества птиц от места нахождения, времени их регистрации, высоты полёта и других факторов, математическую модель с m факторами представим в виде

$$v(x_1, x_2, \dots, x_m) = a_0 + a_1 x_1 + a_2 x_2 + \dots + a_m x_m + a_{12} x_1 x_2 + a_{13} x_1 x_3 + a_{11} x_1^2 + a_{22} x_2^2 + \dots + a_{u \dots v} x_1^u \dots x_m^v + \varepsilon \quad (2)$$

где степень многочлена, отвечающая его структуре с выбранными параметрами u и v , определяется исследователем.

Для вычисления k коэффициентов регрессии $a_0, a_1, \dots, a_{12}, a_{13}, \dots, a_{u \dots v}$ потребуются значения плотности распределения птиц не менее чем в $n > k$ точках. Случай $n = k$ исключает возможность проверить адекватность математической модели. Поэтому целесообразно обеспечить выполнение неравенства $n > k$. Известно, что адекватность математической модели возрастает с увеличением числа измерений при выбранном k и с уменьшением парных коэффициентов корреляции между значениями переменных $x_1, x_2 + \dots, x_m, x_1 x_2, x_1 x_3, x_1^2, x_2^2, \dots, x_1^u x_2^{u+1} \dots x_m^v$ в точках измерения [4,6].

Предположим, что в n точках с координатами $x_{1i}, x_{2i}, \dots, x_{mi}, i = 1, 2, \dots, n$, принадлежащих области D , проведены измерения плотности птиц. Результаты измерения определяют величину неизвестной функции $u(x_{1i}, x_{2i}, \dots, x_{mi})$ с некоторой точностью ε . Обозначим полученные значения через $u_i = u(x_{1i}, x_{2i}, \dots, x_{mi}) + \varepsilon$. Расчёт коэффициентов регрессии проведём методом наименьших квадратов (МНК). Потребуем, чтобы их

значения обеспечивали сумме квадратов отклонений $v_i(x_{1i}, x_{2i}, \dots, x_{mi})$ от измеренных величин u_i минимальное значение

$$S = \sum_{i=1}^{i=n} (v_i(x_{1i}, x_{2i}, \dots, x_{mi}) - u_i)^2 = \min. \quad (3)$$

Теоретические аспекты МНК для вычисления коэффициентов регрессии изложены во многих публикациях [4-6]. Большинство из них относятся к уравнениям в виде линейных многочленов. На практике, как и в нашем случае, часто встречаются нелинейные модели, содержащие элементарные функции типа x_1^2 , $x_1 x_2$, $\sin(x_1^3)$... и их комбинации, например $x_2^2 \exp(\sin(x_4))$ и т. д., что исключает возможность использования МНК в линейном варианте математической модели.

Целью настоящей работы является:

- Разработка алгоритма построения нелинейного уравнения регрессии (2) с использованием информации о численности птиц в различных точках исследуемой территории.
- Создание компьютерной программы для проведения регрессионного анализа возможной зависимости численности птиц от географических координат, времени измерения, высоты полёта и других факторов.

1. Алгоритм вычисления коэффициентов регрессии

Введём новые переменные

$$\begin{aligned} z_1 &= 1, z_2 = x_1, z_3 = x_2, \dots, z_i = x_1 x_2, \\ z_{i+1} &= x_1 x_3, \dots, z_j = x_1^2, z_{j+1} = x_2^2, \dots, z_k = x_1^u \dots x_m^v. \end{aligned} \quad (4)$$

Используя формулы (4), преобразуем нелинейное уравнение регрессии (2) с m переменными $x_1, x_2, \dots, x_m$ к линейной форме с k переменными $z_1, z_2, \dots, z_k$

$$v(z_1, z_2, \dots, z_k) = b_1 z_1 + b_2 z_2 + \dots + b_k z_k, \quad (5)$$

коэффициенты которого $b_1, b_2, \dots, b_k$ связаны с коэффициентами уравнения (2) равенствами

$$\begin{aligned} b_1 &= a_0, b_2 = a_1, b_3 = a_2, \dots, b_i = a_{12}, \\ b_{i+1} &= a_{13}, \dots, b_j = a_{11}, b_{i+1} = a_{22}, \dots, b_k = a_{u\dots v} \end{aligned} \quad (6)$$

Каждой узловой точке с координатами $x_{1i}, x_{2i}, \dots, x_{mi}$, где проводятся измерения, сопоставим значения новых переменных $z_{1i}, z_{2i}, \dots, z_{ki}$. В этом случае число птиц, найденное с помощью формулы (5) в узловых точках, будет равно

$$v_i(z_{1i}, z_{2i}, \dots, z_{ki}) = \sum_{j=1}^{j=k} b_j z_{ji}, j = 1, 2, \dots, n. \quad (7)$$

Пусть результаты учёта количества птиц, зафиксированных в i -й узловой точке, известны и равны u_i , $i=1, 2, \dots, n$. Преобразуем формулу (3) с использованием выражений (6) к виду

$$S = \sum_{i=1}^{i=n} \left[\sum_{j=1}^{j=k} (b_j z_{ji}) - u_i \right]^2 = \min \quad (8)$$

Сумма (8) является функцией от k неизвестных коэффициентов $b_1, b_2, \dots, b_k$. Частные производные для неё по коэффициентам $b_1, b_2, \dots, b_k$ равны

$$\begin{aligned} \frac{\partial S}{\partial b_1} &= 2 \sum_{i=1}^{i=n} \left[\sum_{j=1}^{j=k} (b_j z_{ji}) - u_i \right] z_{1i}, \\ \frac{\partial S}{\partial b_2} &= 2 \sum_{i=1}^{i=n} \left[\sum_{j=1}^{j=k} (b_j z_{ji}) - u_i \right] z_{2i}, \\ &\dots \dots \dots \\ \frac{\partial S}{\partial b_k} &= 2 \sum_{i=1}^{i=n} \left[\sum_{j=1}^{j=k} (b_j z_{ji}) - u_i \right] z_{ki} \end{aligned}$$

Минимальное значение S достигается в точках, где $\partial S / \partial b_1, \partial S / \partial b_2, \dots, \partial S / \partial b_k$ равны нулю, откуда получим k уравнений с k неизвестными относительно $b_1, b_2, \dots, b_k$:

[illegible]

Представим уравнения (9) в развёрнутом виде

$$\left\{ \begin{array}{l} b_1 \sum_{i=1}^{i=n} z_{1i}z_{1i} + b_2 \sum_{i=1}^{i=n} z_{2i}z_{1i} + \dots, b_k \sum_{i=1}^{i=n} z_{ki}z_{1i} = \sum_{i=1}^{i=n} u_i z_{1i}, \\ b_1 \sum_{i=1}^{i=n} z_{1i}z_{2i} + b_2 \sum_{i=1}^{i=n} z_{2i}z_{2i} + \dots, b_k \sum_{i=1}^{i=n} z_{ki}z_{2i} = \sum_{i=1}^{i=n} u_i z_{2i}, \\ \vdots \\ b_1 \sum_{i=1}^{i=n} z_{1i}z_{ki} + b_2 \sum_{i=1}^{i=n} z_{2i}z_{ki} + \dots, b_k \sum_{i=1}^{i=n} z_{ki}z_{ki} = \sum_{i=1}^{i=n} u_i z_{ki} \end{array} \right. \quad (10)$$

ИЛИ

[illegible]

где

$$c_{sr} = \sum_{i=1}^{i=n} (z_{ri} z_{si}) , d_s = \sum_{i=1}^{i=n} u_i z_{si},$$

где $s = 1, 2, \dots, k$ – номер строки, $r = 1, 2, \dots, k$ – номер столбца в формуле (11).

Решение системы линейных алгебраических уравнений (11) относительно неизвестных коэффициентов $b_1, b_2, \dots, b_k$ может быть реализовано с использованием стандартных функций в различных алгоритмических языках или с помощью математических пакетов типа Maple, MathCad. Далее с помощью уравнения (5) легко определить количество птиц в зависимости от конкретных значений переменных в области их определения.

Адекватность полученного уравнения регрессии проверяется с помощью индекса детерминации R^2 и критерия Фишера F [5,6].

2. Разработка программы

Для реализации предложенного алгоритма аппроксимации и анализа зависимости плотности действительного распределения птиц $u(x, y)$ от координат территории была разработана интерактивная программа Birds_MLS на языке Python. Программа обеспечивает полный цикл вычислений от введения исходных данных и задания уровня шума до построения регрессионной модели, проверки её достоверности и визуализации результатов с использованием современных научных библиотек, табл. 1.

Таблица 1 – Перечень библиотек, используемых в программе Birds MLS

Библиотека	Назначение
NumPy	Линейная алгебра, реализация МНК по формуле (11)
Pandas	Формирование и отображение таблиц данных
SciPy	Расчёт статистических показателей
Matplotlib	Визуализация графиков и результатов
PyQt6	Разработка графического интерфейса программы

Преобразование нелинейного уравнения регрессии (1) к линейной форме (5) осуществлялось путём введения вектора признаков Z_i с использованием формул (4)-(7).

Архитектура интерфейса реализована в виде многовкладочного приложения, где каждая вкладка соответствует отдельному этапу моделирования и проведения анализа. После запуска программы на дисплее

появляется запрос на выполнение необходимых расчётов с заполнением таблиц и проведение графических построений, рис. 1.

Рисунок 1 – Первая страница программы, обеспечивающей выбор исходных данных и вывод расчётных характеристик

Вкладка «Данные» обеспечивает выполнение следующих операций:

- ввод размера сетки n, m для узловых точек, где задаются координаты,
- устанавливает величину относительной ошибки δ ,
- генерирует синтетические данные.

Вкладка «Модель» обеспечивает нахождение уравнения регрессии. Здесь на основе матрицы признаков Z , построенной по формулам (3)-(6), выполняется вычисление коэффициентов b_i с использованием системы линейных алгебраических уравнений (11) с последующим расчётом коэффициентов регрессии уравнения (1).

Вкладка «Таблицы» содержит информацию о координатах узловых точек x, y , экспериментальных u_i и вычисленных с использованием уравнения регрессии v_i параметрах.

Вкладка «Графики» предназначена для представления результатов вычисления в графическом виде. Все результаты расчётов (таблицы и коэффициенты) экспортируются в файл Excel для последующего анализа. Последняя вкладка «Анализ критического шума» позволяет изучить влияние ошибки измерений на адекватность построенной модели.

Алгоритм функционирования программы можно представить в виде следующих этапов:

- Ввод исходных параметров (размер сетки в случае двух факторов, все коэффициенты, необходимые для вычислений, величину ошибки δ , используемой при анализе устойчивости модели по отношению к шумовым помехам).
- Формирование матрицы признаков Z в соответствии с формулами (6).
- Решение системы линейных алгебраических уравнений (11) методом МНК.
- Получение прогностических значений $v(x, y)$ и генерация данных $v^*(x, y)$ с учётом ошибок на стадии проверки адекватности модели в двухфакторном эксперименте.
- Вычисление индекса детерминации R^2 и критерия Фишера F .
- Формирование таблиц, построение графиков и экспорт данных.
- Проведение расчётов для оценки критического уровня неточности измерения $\delta_{\text{крит}}$.

3. Проверка адекватности нелинейной модели с использованием программы Birds_MLS

Проверка адекватности предлагаемой модели проводилась путём сравнения истинной зависимости некоторого параметра $u(x, y)$ от двух координат, для которой заданы значения $u(x, y)$, равные u_i в нескольких точках с координатами x_i, y_i , с результатами расчётов, которые получены с помощью уравнения регрессии $v(x, y)$ в этих же точках. Пусть истинная зависимость определяется функцией

$$u(x, y) = d_0 + d_x x + d_y y + d_{xy} xy + d_{xx} x^2 + d_{yy} y^2 + d_{xxy} x^2 y + d_{xyy} x y^2 + d_{xxx} x^3 + d_{yyy} y^3 \quad (12)$$

с известными коэффициентами $d_0, d_x, d_y, \dots, d_{yyy}$.

Поскольку измеряемый параметр u_i проводится с некоторой статистической ошибкой, равной δ , то результат измерения u_i^* оказывается равным

$$u_i^* = u_i(1 \pm \delta), \quad (13)$$

где u_i – истинное значение.

Адекватность математической модели зависит от ошибки измерения и выбранной структуры многочлена для уравнения регрессии. Рассмотрим три различных способа построения уравнения регрессии: зависимость исследуемого параметра от входных факторов не содержит максимумов или минимумов, функция имеет не более одной экстремальной точки, функция содержит две и более экстремальных точек.

В первом варианте уравнение регрессии можно представлять в виде многочлена первой или второй степени. В случае однофакторной или двухфакторной зависимости не более чем с одним экстремумом уравнение желательно скать в виде уравнения регрессии с тремя коэффициентами типа

$$v(x) = a_0 + a_x x + a_{xx} x^2, v(x, y) = a_0 + a_x x + a_y y,$$

для которых в литературе существуют относительно простые алгоритмы решения. Использование подобных математических моделей при обработке результатов измерений, когда ожидается зависимость с двумя и более экстремальными точками, приведёт к большим ошибкам.

Здесь необходимо воспользоваться уравнением регрессии с более высокими степенями. Хорошие результаты получаются в тех ситуациях, когда характер зависимости исследуемого параметра $u(x, y)$ известен и уравнение регрессии выбирается с аналогичной структурой.

Проведём анализ достоверности функционирования программы Birds_MLS на двух примерах. Пусть истинная зависимость измеряемого параметра определяется формулой $u(x, y) = d_0 + d_x x + d_y y$ с коэффициентами $d_0 = 1.0$, $d_x = 2.0$, $d_y = 3.0$. Величина этой функции в трёх узловых точках ($i=1, 2, 3$) с координатами $(0,0)$, $(0,1)$, $(1,0)$ равна $u_1 = 1.0$, $u_2 = 4.0$, $u_3 = 3.0$. Измерения, проведённые с абсолютной точностью в этих же точках с координатами, естественно, не отличаются от теоретических значений и дают те же результаты: $u_1^* = 1.0$, $u_2^* = 4.0$, $u_3^* = 3.0$.

Второй тест проводился на примере нелинейной зависимости

$$u(x, y) = d_0 + d_x x + d_y y + d_{xy} xy + d_{xx} x^2 + d_{yy} y^2 + d_{xxy} x^2 y + d_{xyy} x y^2 + d_{xxx} x^3 + d_{yyy} y^3 \quad (14)$$

с коэффициентами $d_0 = 15.0$, $d_x = 1.0$, $d_y = 2.0$, $d_{xy} = -0.3$, $d_{xx} = 0.4$, $d_{yy} = 0.5$, $d_{xxy} = -0.4$, $d_{xyy} = 0.3$, $d_{xxx} = 0.2$, $d_{yyy} = -0.1$.

График зависимости функции (14) от координат x, y приведён на рис. 2.

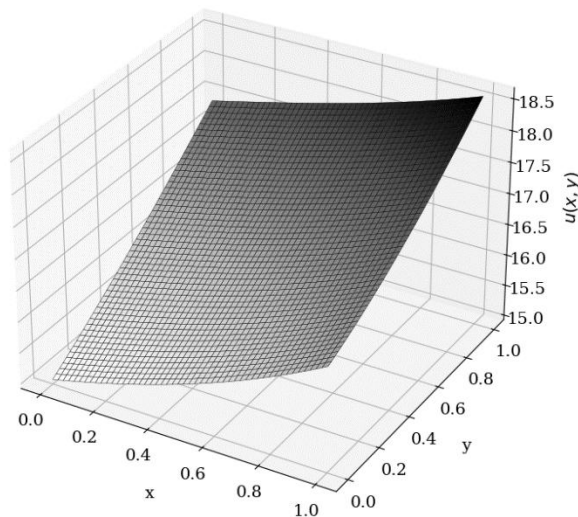


Рисунок 2 – Зависимость функции $u(x, y)$ от координат x, y

Из рис. 2 видно, что функция $u(x, y)$ не содержит экстремальных точек. По мере увеличения независимых переменных она монотонно возрастает в рассматриваемом интервале изменения переменных.

Математическая модель в этом случае относится к первому типу, когда её можно искать в виде многочлена второй или третьей степени. Наилучший результат ожидается при использовании уравнения регрессии с такой же структурой, т. е. в виде

$$v(x, y) = a_0 + a_x x + a_y y + a_{xy} xy + a_{xx} x^2 + a_{yy} y^2 + a_{xxy} x^2 y + a_{xyy} x y^2 + a_{xxx} x^3 + a_{yyy} y^3 + \delta. \quad (15)$$

Любая другая структура математической модели, отличная от многочлена (15), приведёт к появлению дополнительной ошибки. Для вычисления $k = 10$ коэффициентов в формуле (15) необходимы данные о численности птиц в более чем в 10 точках. Пусть в распоряжении исследователя имеется $n = 25$ измерений в точках с координатами, приведёнными в табл. 2.

Таблица 2– Координаты узловых точек x_i, y_i , которые использовались при тестировании функции отклика (16) (первое число в каждой клетке относится к координате x , второе – к координате y).

0;0($i=1$)	0;1($i=2$)	0;2($i=3$)	0;3($i=4$)	0;4($i=5$)
1;0($i=6$)	1;1($i=7$)	1;2($i=8$)	1;3($i=9$)	1;4($i=10$)
2;0($i=11$)	2;1($i=12$)	2;2($i=13$)	2;3($i=14$)	2;4($i=15$)
3;0($i=16$)	3;1($i=17$)	3;2($i=18$)	3;3($i=19$)	3;4($i=20$)
4;0($i=21$)	4;1($i=22$)	4;2($i=23$)	4;3($i=24$)	4;4($i=25$)

Точные значения u_i тестируемой функции $u(x, y)$, вычисленные по формуле (14) для переменных, приведённых в табл. 2 и округлённых до третьего знака, представлены в табл. 3.

Таблица 3 – Точные значения u_i тестируемой функции $u(x, y)$ в узловых точках с координатами x_i, y_i .

15,0($i=1$)	17,4($i=2$)	20,2($i=3$)	22,8($i=4$)	24,6($i=5$)
16,6($i=6$)	18,6($i=7$)	21,6($i=8$)	25,0($i=9$)	28,2($i=10$)
20,2($i=11$)	21,0($i=12$)	23,4($i=13$)	26,8($i=14$)	30,6($i=15$)
27,0($i=16$)	25,8($i=17$)	26,8($i=18$)	29,4($i=19$)	33,0($i=20$)
38,2($i=21$)	34,2($i=22$)	33,0($i=23$)	34,0($i=24$)	36,6($i=25$)

Результаты расчёта значений функции (15) в точках с координатами для данных табл. 2 совпадали с данными табл. 3 с точностью не менее 10^{-15} . Поэтому можно утверждать о корректности предлагаемого алгоритма построения уравнения регрессии в данном случае.

4. Обработка результатов исследования на примере виртуального эксперимента

Проведём обработку виртуального двухфакторного эксперимента при наличии шумовых помех на примере, близком к условиям мониторинга территории ветровой электростанции «Приморск-1» в Запорожской области. Исчерпывающая информация о методе измерений представлена в отчёте [10]. В процессе вечерних наблюдений 24 января 2018 г. на одном из участков с площадью $0,86 \text{ км}^2$ учтено 22 птицы кормового типа, что соответствует средней плотности распределения их на исследуемой площадке около 25,6 птицы на 1 км^2 . Согласно данным из табл. 3 значения функции (14) изменяются в пределах от 15 до 40 со средней величиной, соизмеримой со средней плотностью птиц на участке наблюдения.

Для проведения тестирования предположим, что действительное значение плотности распределения птиц по-прежнему определяются формулой (14) с коэффициентами (15). Результаты измерения, естественно, всегда отличаются от истинных значений. Величина погрешности при оценке численности птиц на исследуемых территориях обычно составляет не менее 30%. Проведём расчёты для случая, когда измерение количества птиц выполняется с некоторой относительной погрешностью δ . Исходные данные u_i^* в узловых точках, отвечающих координатам, представленных в табл. 2 с учётом случайной ошибки измерения, вычислялись по формуле (13), где значение δ для каждого измерения генерировалось случайными числами на отрезке $[-0.3; +0.3]$.

Таблица 4 – Значения тестируемой функции u_i^* в узловых точках, отвечающих координатам, представленных в табл. 2 с учётом случайной ошибки измерения

17,47($i=1$)	16,76($i=2$)	24,55($i=3$)	25,50($i=4$)	18,61($i=5$)
21,34($i=6$)	21,51($i=7$)	25,31($i=8$)	19,42($i=9$)	27,36($i=10$)
18,63($i=11$)	26,38($i=12$)	25,42($i=13$)	31,99($i=14$)	29,56($i=15$)
22,58($i=16$)	26,64($i=17$)	19,79($i=18$)	35,18($i=19$)	35,61($i=20$)
44,12($i=21$)	31,21($i=22$)	42,32($i=23$)	42,02($i=24$)	42,71($i=25$)

Математическую модель процесса построим в виде уравнения регрессии (15), структура которого совпадает с теоретической зависимостью (14). Коэффициенты регрессии $a_0, a_x \dots$ с учётом искажённой информации, представленной в табл. 4, оказались равными:

$$\begin{aligned} a_0 &= 17.7503, a_x = 3.6144, a_y = 0.0316, \\ a_{xy} &= -0.3196, a_{xx} = -2.7892, a_{yy} = 2.2831, \\ a_{xxy} &= -0.4387, a_{xyy} = 0.5651, a_{xxx} = 0.8297, a_{yyy} = -0.5579. \end{aligned} \quad (16)$$

Вычисленные значения функции v_i в узловых точках приведены в табл. 5.

Таблица 5 – Значения уравнения регрессии v_i в узловых точках, отвечающих координатам, представленных в табл. 2 с учётом ошибки измерения

17,75(i=1)	19,51(i=2)	22,48(i=3)	23,33(i=4)	18,70(i=5)	
19,41(i=6)	20,97(i=7)	24,88(i=8)	27,79(i=9)	26,36(i=10)	
20,46(i=11)	20,95(i=12)	24,92(i=13)	29,03(i=14)	29,91(i=15)	
25,89(i=16)	24,44(i=17)	27,59(i=18)	32,01(i=19)	34,34(i=20)	
40,68(i=21)	36,40(i=22)	37,86(i=23)	41,71(i=24)	44,60(i=25)	

Коэффициенты регрессии $a_0, a_x, a_y, \dots, a_{yyy}$ (16), естественно, отличаются от коэффициентов $d_0 = 15,0, d_x = 1,0, d_y = 2,0, \dots, d_{yyy} = -0,1$ многочлена (15), который определяет точное значение исследуемого параметра. Поэтому данные в табл. 3, 4, 5 также различаются.

В приведённом примере различие вызвано только ошибками измерения. Величина индекса детерминации R^2 и критерия Фишера F в данном случае оказались равными 0,72 и 3,5, соответственно. Поскольку $F > F_{кр} = 2,7$ для 5% уровня значимости [5], то математическая модель (15) адекватна. Рассеяние значений v_i в узловых точках относительно величины u_i , характеризующие зависимость (15), представлено на рис. 3.

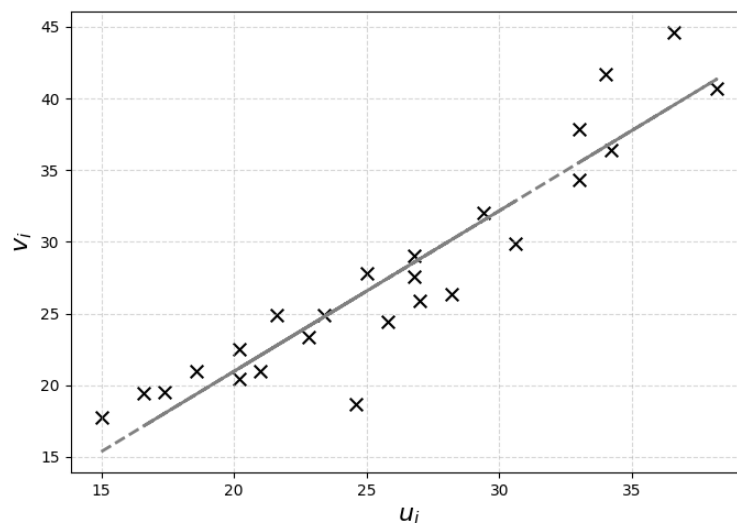


Рисунок 3 – Разброс моделируемых данных v_i , обозначенных точками, относительно точных значений u_i

Зависимость индекса детерминации R^2 и критерия Фишера F от ошибки δ представлена в табл. 6.

Таблица 6 – Влияние ошибки измерения δ на адекватность математической модели.

δ	0	0,1	0,2	0,3	0,4	0,5	0,6
R^2	1	0,96	0,85	0,72	0,70	0,58	0,43
F	∞	33,6	7,9	3,5	3,3	1,9	1,0
Адекватность модели	Да	Да	Да	Да	Да	Нет	Нет

Вычисленное значение критерия Фишера в первых пяти столбцах табл. 6 больше критического значения, равного 2,7 для 5%-го уровня значимости. Зависимость среднеквадратического отклонения параметра $|(u_i^* - v_i)/v_i|$ от погрешности δ представлена на рис. 4.

Как и следовало ожидать, рост ошибки измерения вызывает снижение индекса детерминации R^2 и критерия Фишера F и, следовательно, к уменьшению адекватности математической модели. Из данных табл. 6 и рис. 4 видно, что достоверность уравнения регрессии обеспечивается в области при значениях δ менее 0,45. Поэтому анализируемая модель в этом случае для ошибки $\delta \leq 0,4$ достоверна. При $\delta \geq 0,45$ модель становится неадекватной.

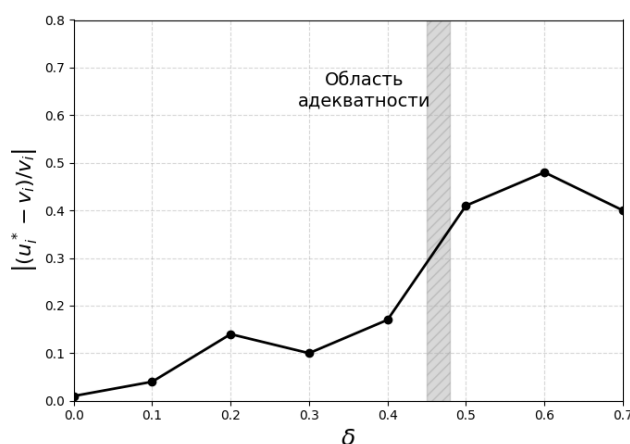


Рисунок 4 – Зависимость среднеквадратического отклонения $|(u_i^* - v_i)/v_i|$ исследуемого параметра от относительной погрешности δ

Заключение

Разработан алгоритм проведения регрессионного анализа результатов измерения численности птиц на исследуемой территории. На алгоритмическом языке Python создана программа Birds_MLS, которая обеспечивает полный цикл вычислений, начиная от введения исходных данных, построения регрессионной модели, и заканчивая проверкой её достоверности и визуализацией результатов с использованием современных научных библиотек NumPy, Pandas, SciPy, Matplotlib, PyQt6. Тестирование предложенного метода построения уравнения регрессии проводилась путём сравнения истинной зависимости, определяемой функцией $u(x, y)$ от двух координат, с результатами расчётов, которые получены с помощью уравнения регрессии $v(x, y)$ в этих же точках. Разработанная программа может быть использована для обработки информации в мониторинговых мероприятиях, полученной при исследовании распределения птиц на исследуемой территории.

Работа выполнена в рамках проекта № FRRS-2023-035 «Оценка состояния сезонных орнитологических комплексов и трансконтинентальных миграций на мониторинговых полигонах Азово-Черноморского региона и Сиваша, разработка программного обеспечения для моделирования и управления природными комплексами» Государственного задания Министерства образования и науки Российской Федерации. Предлагаемый метод нахождения уравнения регрессии рекомендуется для изучения распределения плотности птиц на территории ветровых электростанций и прилегающих к ним буферным зонам.

Авторы выражают большую благодарность кандидату биологических наук доценту В. Д. Сioxину за всестороннюю помощь в проведении настоящей работы и обсуждении полученных результатов.

Список литературы

1. Сioxин В.Д. (2021). Отчет по проведению базового мониторинга сезонных комплексов птиц и летучих мышей с разработкой отдельных позиций менеджмента и минимизации влияния на природные комплексы в границах площадки Ботиевской ВЭС в Приазовском районе Запорожской области / В.Д. Сioxин, П.И. Горлов, И.И. Черничко, А.М. Волох, С.М. Подорожный и др./ – Мелитополь: Научно-производственное предприятие «Ekoresurs i monitorynh», 2021.–280 с.
2. Siokhin V. Identification of “Primorsk-1” wind power plant impact on the ecological situation connected with the behavior of ornithofauna on the Azov Sea coast / V. Siokhin, V. Osadchyi, P. Gorlov, V. Yermieiev / Journal of Physics: Conference Series. – 2021. P. 1–21.
3. Osadchyi V. Development of the information system for forecasting collision between birds and wind arms / V. Osadchyi, V. Siokhin, P. Gorlov, Yermieiev V., Osadcha K./ Eastern-European Journal of Enterprise Technologies. articles/10.15587/1729–4061.2019.174398 Volume 4/2(100), 2019, p.29–40.
4. Орлов А. И. Прикладная статистика / А. И. Орлов. – Изд. «Экзамен». Москва. 2004. – 483 с.

5. Кремер, Н. Ш. Теория вероятностей и математическая статистика. / Н.Ш. Кремер. М.: Изд. «ЮНИТИ», 2004. – 577 с.
6. Шорохова И. С. Статистические методы анализа: учебное пособие / И. С. Шорохова, Н. В. Кисляк, О. С. Мариев; М-во образования и науки Российской Федерации, Уральский федеральный университет. – Екатеринбург: Издательство Уральского университета, 2015. – 300 с. – ISBN 978–5–7996–1633–5.
7. Моделирование информационных систем и технологий: материалы Международной научно-практической конференции, Воронеж, 20 февраля 2025 г. / отв. ред. В. К. Зольников, А. И. Заревич; М-во науки и высшего образования РФ, ФГБОУ ВО «ВГЛУ». – Воронеж, 2025. – 626 с. – Электронный ресурс: [https://vglu.ru/nauka/konferencii/2025/mezhdunarodnaya-nauchno-prakticheskaya konferenciya-modelirovanie-informacionnyh-sistem-i-tehnologij/](https://vglu.ru/nauka/konferencii/2025/mezhdunarodnaya-nauchno-prakticheskaya-konferenciya-modelirovanie-informacionnyh-sistem-i-tehnologij/) электронный. ISBN 978–5–7994–1089–6 –.
8. Яковлев К. Н. Аппроксимация экспериментальных данных методом наименьших квадратов аннотация / К. Н. Яковлев // Перспективы науки и общества в условиях инновационного развития: сборник статей Международной научно-практической конференции, Пермь, 12 марта 2022 года. – Уфа: Общество с ограниченной ответственностью «ОМЕГА САЙНС», 2022. – С. 35–43. – EDN JGOTYP.
9. Антибас, И. Р. Сравнение численного метода и компьютерного моделирования линейной регрессии оптимального профиля среза транспортирующего устройства от величины подачи хлебной массы / И. Р. Антибас, А. Г. Дьяченко, Т. П. Савостина // Научная жизнь. – 2019. – Т. 14, № 9(97). – С. 1432–1440. – DOI 10.35679/1991–9476–2019–14–9–1432–1440. – EDN QUGNQJ.
10. Сіохін, В.Д., (2018). Проведення сучасних моніторингових досліджень та розробка експертного висновку і наукового звіту щодо впливу будівництва ВЕС на природні комплекси довкілля, орнітологічні комплекси та мігруючих птахів, кажанів у межах Приморського району Запорізької області /В.Д. Сіохін, П.І. Горлов, І.К. Поліщук, С.М. Подорожний, О.М. Долинна, І.Б. Сальникова-Буденко та ін.// - Мелітополь: Науково-виробниче підприємство «Ekoresurs i monitorynh», 2018.-149 с.

References

1. Siokhin V. D. (2021). Report on conducting basic monitoring of seasonal bird and bat complexes with the development of individual management positions and minimization of impact on natural complexes within the boundaries of the Botievskaya AES site in the Priazovsky district of the Zaporizhzhia region / V. D. Siokhin, P. I. Gorlov, I. I. Chernichko, A. M. Volokh, S. M. Podorozhny et al. – Melitopol: Ekoresurs i monitorynh Scientific and Production Enterprise, 2021. – 280 p.
2. Siokhin V. Identification of “Primorsk–1” wind power plant impact on the ecological situation connected with the behavior of ornithofauna on the Azov Sea coast / V. Siokhin, V. Osadchyi, P. Gorlov, V. Yermieiev / Journal of Physics: Conference Series. – 2021. P. 1–21.
3. Osadchyi V. Development of the information system for forecasting collision between birds and wind arms / V. Osadchyi, V. Siokhin, P. Gorlov, Yermieiev V., Osadcha K./ Eastern–European Journal of Enterprise Technologies. articles/10.15587/1729–4061.2019.174398 Volume 4/2(100), 2019, p.29–40.
4. Orlov A. I. Applied Statistics / A. I. Orlov. – Ed. «Exam». Moscow. 2004. – 483 p.
5. Kremer N. Sh. Probability Theory and Mathematical Statistics. / N. Sh. Kremer. M.: Publishing House «UNITY», 2004. – 577 p.
6. Shorokhova I. S. Statistical Methods of Analysis: Textbook / I. S. Shorokhova, N. V. Kislyak, and O. S. Mariyev; Ministry of Education and Science of the Russian Federation, Ural Federal University. – Yekaterinburg: Publishing House of the Ural University, 2015. – 300 p. – ISBN 978–5–7996–1633–5.
7. Modeling of Information Systems and Technologies: Proceedings of the International Scientific and Practical Conference, Voronezh, February 20, 2025 / edited by V. K. Zolnikov and A. I. Zarevich; Ministry of Science and Higher Education of the Russian Federation, Voronezh State University of Forestry and Technologies. – Voronezh, 2025. – 626 p. – URL: [https://vglu.ru/nauka/konferencii/2025/mezhdunarodnaya-nauchno-prakticheskaya konferenciya-modelirovanie-informacionnyh-sistem-i-tehnologij/](https://vglu.ru/nauka/konferencii/2025/mezhdunarodnaya-nauchno-prakticheskaya-konferenciya-modelirovanie-informacionnyh-sistem-i-tehnologij/) electronic. ISBN 978–5–7994–1089–6.
8. Yakovlev K. N. Approximation of Experimental Data by the Least Squares Method Abstract / K. N. Yakovlev // Prospects for Science and Society in the Context of Innovative Development: Collection of Articles from the International Scientific and Practical Conference, Perm, March 12, 2022. – Ufa: OMEGA SCIENCE LLC, 2022. – Pp. 35–43. – EDN JGOTYP.
9. Antibas I. R. Comparison of the Numerical Method and Computer Modeling of Linear Regression of the Optimal Profile of the Cutting Device from the Amount of Bread Mass Feed / I. R. Antibas, A. G. Dyachenko, and T. P.

Savostina // Scientific Life. – 2019. – Vol. 14, No. 9(97). – Pp. 1432–1440. – DOI 10.35679/1991–9476–2019–14–9–1432–1440. – EDN QUGNQJ.

10. Siokhin, V. D. (2018). Conducting modern monitoring studies and developing an expert opinion and scientific report on the impact of wind farm construction on natural environmental complexes, Ornithological complexes and migratory birds and bats within the Primorsky District of Zaporizhia region /V. D. Siokhin, P. I. Gorlov, I. K. Polishchuk, S. M. Podorozhny, O. M. Dolinna, I. B. Salnikova-Budenko, etc. // Melitopol: scientific and production enterprise "Ekoresurs i monitorynh", 2018. -149 p. (in Russian)