

ЭКОЛОГИЯ

(БИОЛОГИЧЕСКИЕ НАУКИ)

Естественные науки. 2025. № 3 (20). С. 31–44.

Yestestvennye nauki = Natural Sciences. 2025; 3 (20): 31–44 (In Russ.)

Научная статья

УДК 615.036.2

doi 10.54398/2500-2805.2025.20.3.004

МЕТОДИЧЕСКИЕ АСПЕКТЫ ВОЗМОЖНОСТЕЙ ПРОГРАММНОЙ СРЕДЫ R В БИОМЕТРИИ ДЛЯ ЭКОЛОГИЧЕСКИХ ИССЛЕДОВАНИЙ

Сафаралиев Ильдар Абсатарович¹✉, Пучков Михаил Юрьевич²

¹ Волжско-Каспийский филиал ВНИРО («КаспНИРХ»), г. Астрахань, Россия

² Астраханский государственный университет им. В. Н. Татищева,
г. Астрахань, Россия

¹ safaraliev.i.a@mail.ru✉

Аннотация. Представлены возможности программной среды R как инструмента для статистического анализа. Приводится краткое сравнение с другими программными статистическими продуктами, показывающее преимущество среды R. В качестве примеров работы в среде R приведены некоторые статистические методы из учебного пособия Г. Ф. Лакина «Биометрия» (1990 г.), которые реализованы в R, свидетельствующие об её актуальности в настоящее время при одновременном использовании с данным языком программирования как в научных исследованиях, так и в образовательном процессе.

Ключевые слова: биометрия, количественная биология, статистические методы, U-критерий Уилкоксона, множественная корреляция, MS Excel, R, RStudio

Для цитирования: Сафаралиев И. А. Пучков М. Ю. Методические аспекты возможностей программной среды R в биометрии для экологических исследований // Естественные науки. 2025. № 3 (20). С. 31–44. <https://doi.org/10.54398/2500-2805.2025.20.3.004>.

IMPLEMENTATION OF THE R SOFTWARE ENVIRONMENT CAPABILITIES USING EXAMPLES OF THE TEXTBOOK BY G. F. LAKIN “BIOMETRICS” FOR ENVIRONMENTAL RESEARCH

Safaraliev Ildar A.¹✉, Puchkov Mikhail Yu.²

¹ Volga-Caspian branch of the VNIRO (“CaspNIRKH”), Astrakhan, Russia

² Astrakhan Tatishchev State University, Astrakhan, Russia

¹ safaraliev.i.a@mail.ru✉

Abstract. The article presents the of the R software environment capabilities as a tool for statistical analysis. A brief comparison with other statistical software products is given, showing

the advantage of the R environment. As examples of work in the R environment, some statistical methods from the textbook by G. F. Lakin “Biometrics” (1990) are given, which are implemented in R and testify to the current relevance of this monograph when used simultaneously with this programming language, both in scientific research and in the educational process.

Keywords: biometrics, quantitative biology, statistical methods, Wilcoxon test (Mann-Whitney), multiple correlation, MS Excel, R, RStudio

For citation: Safaraliev I. A. Puchkov M. Yu. Methodological aspects of the capabilities of the R software environment in biometrics for environmental research. *Yestestvennye nauki = Natural Sciences*. 2025; 3 (20): 31–44. <https://doi.org/10.54398/2500-2805.2025.20.3.004> (In Russ.).

В любых исследованиях по естественным наукам, равно как и в их прикладных областях (медицине, сельском хозяйстве, селекции, охотоведении, лесоводстве, биотехнологии и т. д.), одним из этапов является статистическая обработка данных, результаты которой позволяют математическим языком выразить и оценить надёжность полученных результатов и прийти к верным заключениям. Такое практическое применение математической статистики и теории вероятности реализовалось в биометрии (другое её название — количественная биология) как одной из дисциплин в биологических науках. В основе биометрии лежат адаптированные методы математической статистики и статистические методы, разработанные непосредственно для решения биологических задач.

Биометрия как предмет преподается во всех высших учебных заведениях России, кафедры которых специализируются в биологических дисциплинах. Широкую известность в СССР приобрели работы Н. А. Плохинского «Биометрия» [12] и «Алгоритмы биометрии» [13], В. Ю. Урбах «Статистический анализ в биологических и медицинских исследованиях» [14]. Начиная с 1970-х гг. и по настоящее время популярностью в учебном процессе пользуется пособие Г. Ф. Лакина «Биометрия» [8], которое является наиболее цитируемым изданием по этой тематике¹. В данных пособиях на соответствующем уровне математики для биологов изложены основы статистического анализа и методы, которые использовались в биологических исследованиях, а процедуры расчётов отвечали уровню доступности электронно-вычислительных машин (ЭВМ), т. е. фактически были представлены способы вычислений в ручную, в которых каждое математическое действие необходимо было вычислять отдельно, в лучшем случае при помощи калькулятора или соответствующих таблиц. В совокупности статистический анализ в то время представлял собой достаточно времязатратный этап исследований.

С развитием и широкой доступностью персональных ЭВМ, начиная с 2000-х гг., появилась возможность значительно ускорить и автоматизировать статистический анализ. Стали доступны зарубежные программы: “STATGRAPHICS”, SPSS, “SYSTAT”, “STATISTICA” и др. Но наибольшей популярностью для простого статистического анализа, в частности

¹ URL: <https://elibrary.ru/item.asp?id=32509850> (дата обращения: 09.09.2025).

описательной статистики, стала программа “MS Excel”. Этому способствовала её универсальность в решении множества задач как простейшая база хранения данных, составлении результирующих таблиц, построении графиков, выполнении вычислений. В программу были заложены возможности некоторых видов статистических вычислений и прогнозирования, т. к. функциональность “MS Excel” исходила из её начальных целей в качестве инструмента для решений бухгалтерских и экономических задач, а ряд реализованных методов подходил и для решения задач биометрии. В этой связи появилось достаточно большое количество книг и учебного материала, посвящённого реализации методов биометрии в “MS Excel” [2–4; 7; 9; 15].

Несмотря на широкую популярность “MS Excel”, его возможности как инструмента для полноценного статистического анализа в биологии ограничены. Данный продукт компании “Microsoft” представляет базовый набор статистических функций (СРЗНАЧ, МЕДИАНА, СТАНДОТКЛОН, критерии Стьюдента, дисперсионный анализ и т. д.). Однако для многих сложных биологических задач (многомерная статистика, специфические методы для экологии и генетики и т. д.) этого недостаточно. В естественных науках, где важна воспроизводимость результатов, в данном программном продукте повторение полного процесса расчётов затруднено в отличие от программных языков R или Python, в которых можно записать каждое действие в скрипте, что позволяет воспроизвести в точности и последовательно все этапы вычисления. В “MS Excel” возможно построение только некоторых типов диаграмм (гистограммы, точечные, кривые, круговая). В представлении результатов некоторых современных статистических методов требуется построение таких диаграмм, как PCA-plot, heatmaps, дендрограммы, box-plot, 3-D графики и т. д. Средства “MS Excel” не позволяют построить выше перечисленные диаграммы.

Для серьёзного статистического анализа исследователи часто используют “STATGRAPHICS”, SPSS, “SYSTAT”, “STATISTICA”. К сожалению, данные программы являются платными продуктами, и из-за своей цены в большей степени доступны крупным организациям, научно-исследовательским институтам и университетам. В настоящее время существуют бесплатные решения для статистического анализа, их возможности не уступают платным программам — это языки программирования с открытым исходным кодом, один из них — язык R. R — это язык программирования и интегрированная программная среда, специально разработанная для статистических вычислений и построения графиков. Обладает обширным набором встроенных функций и дополнительных пакетов, расширяющих его возможности. Его можно применять везде, где нужна работа с данными. Основное преимущество R лучше всего проявляется именно при статистическом анализе данных: от вычисления средних величин до серьёзных операций с временными рядами и моделирования.

На сегодняшний момент существует достаточно большое количество литературы по основам и более глубокому использованию среды R. Ниже представлена лишь небольшая часть доступных публикаций по этой теме: «R в действии» Р. И. Кобакова [6]; «Визуализация данных с помощью

«ggplot2» С. Э. Мастицкого [11], «Классификация, регрессия и другие алгоритмы Data Mining с использованием R» В. К. Шитикова, С. Э. Мастицкого [16], «Статистический анализ и визуализация данных с помощью R» С. Э. Мастицкого, В. К. Шитикова [10], «Введение в статистический пакет R: типы переменных, структуры данных, чтение и запись информации, графика» И. С. Зарядова [5]. Но, к сожалению, при достаточно большом выборе книг по обучению работе в R, выбор пособий, посвящённых статистическому анализу в биологических науках, невелик. Мы смогли обнаружить только одну книгу по рассматриваемой теме — «Рандомизация и бутстреп: статистический анализ в биологии и экологии с использованием R» В. К. Шитикова, Г. С. Розенберга [17], хотя в зарубежной литературе их присутствует достаточно большое количество [18; 19; 21; 22].

Несмотря на значительное развитие ИТ-продуктов и наличие современных учебных материалов по статистическому анализу с применением современных программ, до сих пор учебные пособия Н. А. Плохинского [12; 13], В. Ю. Урбах [14] и Г. Ф. Лакина [8] используются в образовательном процессе высшей школы России. Следовательно, становится актуальным вопрос о составлении дополнений к вышеперечисленным учебным пособиям, в которых показана реализация статистических методов средствами современного программирования.

Таким образом, наша цель — на примере отдельных статистических методов из учебного пособия Г. Ф. Лакина «Биометрия» [8] показать их реализацию в программной среде R.

Язык R принадлежит к семейству так называемых высокоуровневых объектно-ориентированных языков программирования. Это означает, что пользователь оперирует с так называемыми объектами, под которыми подразумевается всё, что было создано в сессии работы с R. Выделяют два основных типа объектов:

1. Функции (“function objects”) — это поименованные программы, предназначенные для создания новых объектов или выполнения определённых действий над ними.

2. Объекты данных (“data objects”) предназначены для хранения данных — это отдельные переменные, векторы, матрицы, массивы, списки, факторы и таблицы данных.

Собранные функции, объединённые сходной тематикой, методами обработки данных и предназначенные для коллективного и свободного использования, комплектуются в пакеты (packages). R поставляется со стандартным набором пакетов (base, datasets, utils, grDevices, graphics, methods и т. д.). В них уже содержатся разнообразные функции и наборы данных, доступные по умолчанию, а другие по необходимости пользователю необходимо скачивать с репозитариев и устанавливать самостоятельно (fishmethods, DescTools, ggplot2 и т. д.).

Установочный пакет R можно скачать с официального сайта данного программного продукта² [20], где необходимо выбрать соответствующий пакет в зависимости от операционной среды, которая установлена на компьютере. Помимо установки самого языка R рекомендуется также установить графический интерфейс RStudio, который облегчает работу, организуя аспекты сеанса работы в R на отдельных панелях. Дистрибутив последней версии RStudio доступен на официальном сайте разработчика³. В случае, если у пользователя неподдерживаемая операционная система, например “Windows 7”, то скачивание стабильно работающей архивной версии RStudio возможно по адресу: <https://dailies.rstudio.com/version/2023.03.2+454.pro2/> (дата доступа: 09.09.2025). После успешной установки R и RStudio, во время запуска RStudio, программа открывается в окне, которое разделено на несколько частей, представленных панелью консоли и вкладками, содержащими рабочего пространства, истории выполненных команд, списка доступных пакетов, помощи, построенных графиков и т. д. (рис. 1).

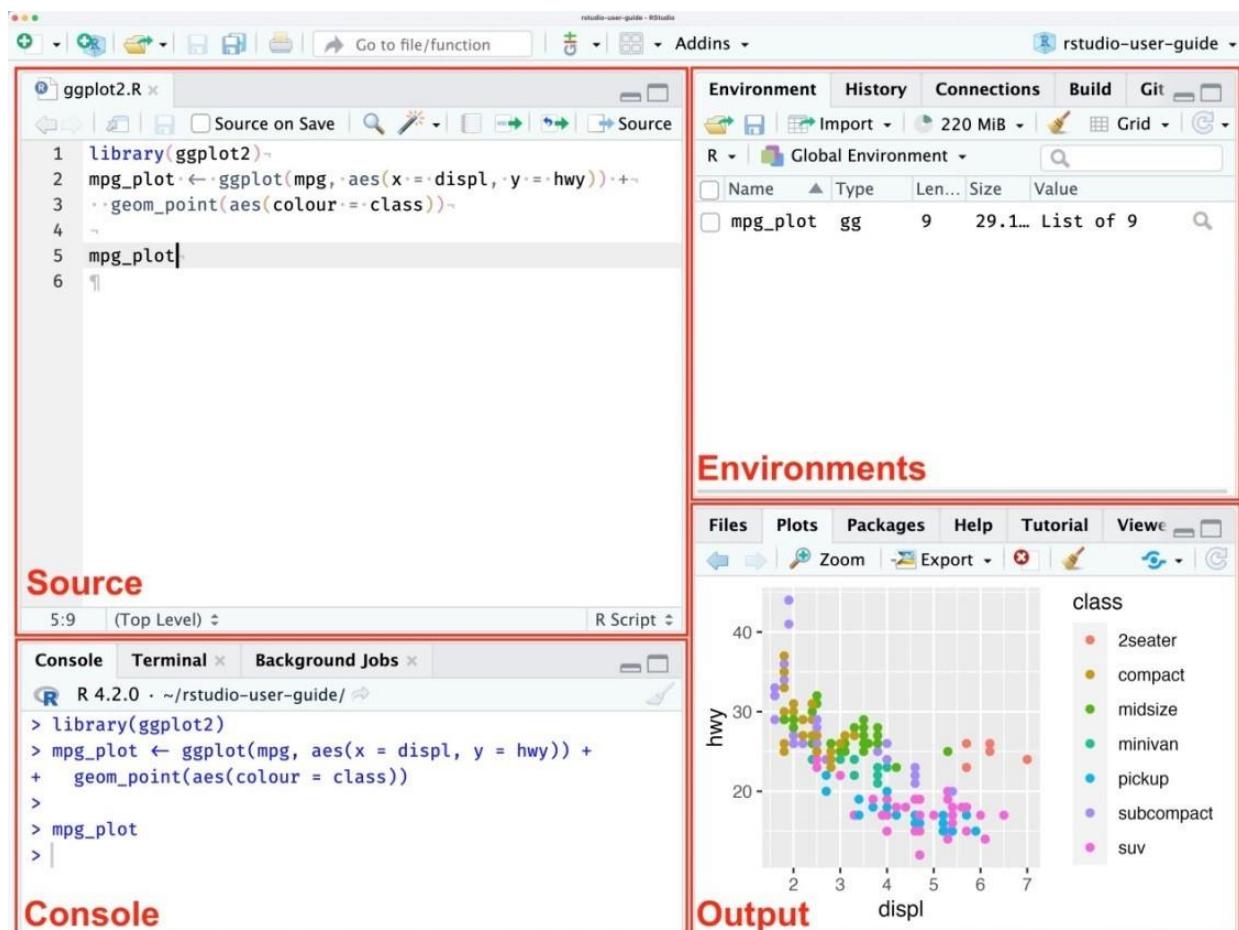


Рисунок 1 — Окно программы RStudio (источник: веб-руководство по работе с RStudio)

² URL: <https://www.r-project.org> (дата доступа: 09.09.2025).

³ URL: <https://posit.co/download/rstudio-desktop/> (дата доступа: 09.09.2025).

Во время сессии работы в RStudio ввод программного кода выполняется в панели “Console” после особого знака/оператора «приглашение» (>) (рис. 1–6). В коде среды R в начале строки часто используется знак «решетка» (#) для создания необходимых примечаний или пояснений к коду. В этом случае содержимое строки не воспринимается программой как команда, т. к. в правиле синтаксиса R содержимое строки после # игнорируется, т. е. не исполняется. Если текст команды более одной строки, то консоль отмечает её продолжение на следующей строке знаком/оператором «плюс» (+), таким образом отмечая, что это одна целая команда.

Рассмотрим пример вычисления в R самого распространённого статистического показателя *средняя*. Данная статистика вычисляется встроенной функцией *mean()*. Для демонстрации работы этой функции воспользуемся данными коров горбеновской породы из колонки «масса тела, кг» таблицы 101 примера 5 на странице 223 пособия [8] и рассчитаем среднее значение выборки. Способ ввода данных, запуск команды вычисления и результаты возвращения функции *mean()* показаны на рисунке 2. Как показывает возвращение функции *mean()* средний вес коров исследуемой выборки составляет 349,05 кг (рис. 2).

The screenshot shows the RStudio interface with the following components:

- Top Bar:** Shows the title "Lakin_in_R - RStudio" and a menu bar with File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help.
- Console:** Displays the R command history:

```
> # ввод данных таблицы 101 примера 5 на странице 223
> x<-c(327, 302, 327, 294, 410, 342, 409, 311, 297,
+      364, 377, 358, 284, 314, 352, 387, 375, 311,
+      332, 262, 333, 381, 320, 295, 345, 440, 405,
+      323, 411, 434, 352, 295, 369, 444, 319, 303,
+      352, 344, 361, 303, 328, 344, 284, 295, 360,
+      244, 279, 303, 329, 323, 287, 337, 295, 339,
+      400, 306, 335, 341, 343, 411, 316, 314, 396,
+      339, 332, 328, 314, 409, 367, 396, 384, 356,
+      446, 338, 300, 360, 324, 245, 345, 368, 397,
+      405, 368, 418, 371, 382, 410, 443, 385, 285,
+      321, 351, 331, 355, 396, 390, 426, 430, 386, 331)
> # расчет среднего значения
> mean(x)
[1] 349.05
>
```
- Environment Panel:** Shows the variable "x" defined as a numeric vector from 1 to 100, containing the values listed in the console.
- Plots Panel:** Shows a small preview of the "About R" page.
- Help Panel:** Shows the "About R" page content, which includes:
 - R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and macOS.
 - See the [R project web page](#) on "What is R?" for more information.
 - Sources and binaries for R and for contributed packages are available from the [Comprehensive R Archive Network \(CRAN\)](#).

Рисунок 2 — Скриншот окна RStudio во время сессии работы в среде R для вычисления средней массы коров горбеновской породы

Любой статистический анализ начинается с вычисления показателей описательной статистики, к которым относятся *средняя, медиана, минимум, максимум, дисперсия, стандартное отклонение* и т. д. Выше перечисленные значения статистик можно вычислить отдельно, применив соответствующую функцию *mean()*, *median()*, *min()*, *max()*, *var()*, *sd()* соответственно. Но в R также существует способ рассчитать эти статистики одномоментно функцией *stat.desc()* из отдельно загружаемого пакета *pastecs*. Для демонстрации работы этой функции используем выборку веса коров горбановской породы, которая использовалась выше для примера работы функции *mean()* (рис. 3).

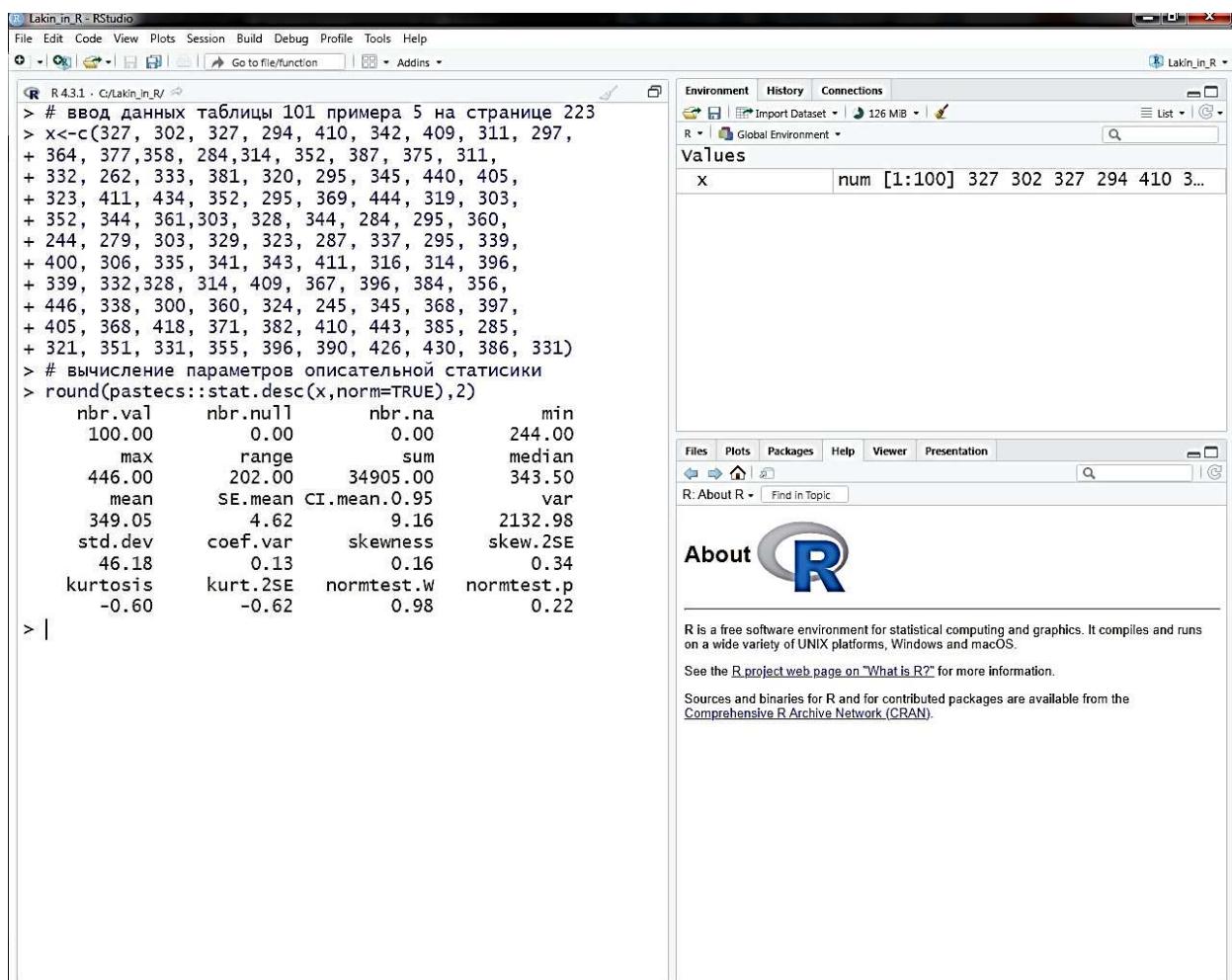


Рисунок 3 — Скриншот окна RStudio во время сессии работы в среде R для вычисления параметров описательной статистики коров горбановской породы

Результатом возвращения функции *stat.desc()* являются: количество вариантов в выборке (*nbr.val*), количество отсутствующих или пропущенных значений (*nbr.null* и *nbr.na*), минимальное и максимальное значения (*min* и *max*), размах выборки (*range*), сумма значений выборки (*sum*), медиана (*median*), средняя арифметическая (*mean*), стандартная ошибка средней (*SE.mean*), 95 % доверительный интервал генеральной средней (*CI.mean.0.95*), дисперсия (*var*), стандартное отклонение (*std.dev*), коэффициент вариации (*coef.var*), асимметрия и её ошибка (*skewness* и *skew.2SE*), эксцесс и его ошибка (*kurtosis*

и $kurt.2SE$), значение статистики теста на нормальность распределение Шапиро-Уилка и значение его доверительной вероятности ($normtest.W$ и $normtest.p$; рис. 3).

В “MS Excel” отсутствуют отдельные функции или надстройки для расчета показателей непараметрических критериев [1]. В среде R представлен широкий выбор методов непараметрической статистики. Рассмотрим способ вычисления значения U-критерия Уилкоксона (Манна – Уитни), который представлен в пособии Г. Ф. Лакина [8], в примере 12 со страницы 131 (данные массы тела лабораторных мышей, разделенные на две группы: опытная и контрольная; рис. 4). Ставится вопрос: есть ли различия в массе тела между двумя группами мышей?

The screenshot shows the RStudio interface with the following components:

- Top Bar:** File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help.
- Left Panel (Code Editor):** Shows R code for loading data and performing a Wilcoxon rank sum test.

```
> # ввод данных примера 12 на странице 131
> x<-c(80,76,76,64,70,68,72,79,83) # опытная группа
> y<-c(70,78,60,80,62,68,73,60,71,66,69) # контрольная группа
> # Расчет U-критерия Уилкоксона (Манна-Уитни)
> wilcox.test(x,y)

Wilcoxon rank sum test with continuity correction

data: x and y
W = 71.5, p-value = 0.1017
alternative hypothesis: true location shift is not equal to 0
> |
```
- Environment Tab:** Displays the variables **x** and **y**.

		values
x	num [1:9]	80 76 76 64 70 68 72...
y	num [1:11]	70 78 60 80 62 68 7...
- Help Tab:** Shows the "About R" page, which includes information about R being a free software environment for statistical computing and graphics, running on various platforms, and links to the R project website and CRAN.

Рисунок 4 — Скриншот окна RStudio во время сессии работы в среде R для вычисления U-критерия Уилкоксона (Манна – Уитни) в оценке различий массы тела у опытной и контрольной групп лабораторных мышей

Возвращение U-критерия Уилкоксона (Манна – Уитни) функцией `wilcox.test()` выводится на консоль: значение критерия — 71,5 (W), величина значимости — 0,1017 (p-value), что указывает на статистическую недостоверность различий в массе тела лабораторных мышей между двумя группами — опытной и контрольной (рис. 4).

Необходимо отметить, что величина U-критерия, рассчитанная функцией *wilcox.test()* отличается от значения в учебном пособии Г. Ф. Лакина [8]. Дело в том, что в коде функции *wilcox.test()* в качестве результирующего значения U-критерия выбирается наибольшая величина из рассчитанных W1 (опыт) и W2 (контроль). Также в процессе вычислений при присвоении индекса ранга используется метод средней (поэтому U-критерий представлен дробным значением); в учебном пособии индекс ранга присваивается порядковым номером варианты двух выборок, числовые значения которых расположены в возрастающем порядке, поэтому U-критерий представлен целым числом. В целом заключения по результатам вычислений критерия Уилкоксона (Манна – Уитни) в пособии Г. Ф. Лакина [8] и среде R совпадают — различия в массе тела лабораторных мышей статистически недостоверны.

Один из трудоёмких и времязатратных для вычисления методов в пособии Г. Ф. Лакина [8] является расчёт значения коэффициента линейной корреляции Пирсона как многоэтапная процедура вычислений, особенно если объёмы выборок значительны. Поэтому в пособии были представлены способы, облегчающие расчёты этого статистического показателя, актуальные для того времени в условиях ограниченной доступности ЭВМ. При использовании языка R эти затруднения, равно как и для других методов, полностью снимаются, наибольшее время затрачивается только для ввода первичных данных.

Коэффициент линейной корреляции Пирсона в среде R можно вычислить двумя функциями *cor()* или *cor.test()*. Для сравнения сопряженности двух рядов признаков *A* и *B* синтаксис этих двух функций одинаковый. В качестве аргументов подаются пути векторов анализируемых выборок, длина которых должна быть равной. Вид команды следующий: *cor(A, B)* или *cor.test(A, B)*. Различие в функциях *cor()* и *cor.test()* состоит в том, что первая функция возвращает только значение коэффициента, а вторая помимо значения коэффициента выполняет тест проверки статистической значимости корреляции.

Рассмотрим работу функции *cor.test()* на данных пособия [8] из примера 5 со страницы 223 (табл. 101; рис. 5).

Результаты возвращения функции *cor.test()* представлены детальным статистическим анализом (рис. 5): наименование проведенного статистического метода (Pearson's product-moment correlation); значение коэффициента корреляции (sample estimates: *cor*) — 0,524; *t* фактическое — 6,0974; число степени свободы (*df*) — 98; 95 % доверительный интервал коэффициента (95 percent confidence interval) — 0,3656854 и 0,6535335. Значение *p-value* = 2,14e-08 указывает, что нулевую гипотезу можно отвергнуть при *p* < 0,001 и принять альтернативную о том, что значение коэффициента для генеральной совокупности не равно нулю.

The screenshot shows the RStudio interface with the following details:

- Top Bar:** File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help.
- Left Panel (Code Editor):**

```
> # ввод данных примера 5 на странице 223 таблицы 101
> mydata<-data.frame (wt.Y=c(327, 302, 327, 294, 410,
+ 342, 409, 311, 297, 364, 377, 358, 284, 314,
+ 352, 387, 375, 311, 332, 262, 333, 381, 320,
+ 295, 345, 440, 405, 323, 411, 434, 352, 295,
+ 369, 444, 319, 303, 352, 344, 361, 303, 328,
+ 344, 284, 295, 360, 244, 279, 303, 329, 323,
+ 287, 337, 295, 339, 400, 306, 335, 341, 343,
+ 411, 316, 314, 396, 339, 332, 328, 314, 409,
+ 367, 396, 384, 356, 446, 338, 300, 360, 324,
+ 245, 345, 368, 397, 405, 368, 418, 371, 382,
+ 410, 443, 385, 285, 321, 351, 331, 355, 396,
+ 390, 426, 430, 386, 331), # масса коров
+ milk_yield.X=c(2325, 1761, 2310,
+ 2035, 2172, 2277, 2784, 1523, 1838, 1984, 1775,
+ 2700, 2241, 1954, 2046, 2323, 1710, 1868, 2166,
+ 1384, 2288, 2249, 1520, 2389, 2012, 3219, 1806,
+ 2803, 2385, 2826, 1832, 2413, 2625, 2614, 2297,
+ 1946, 2278, 2111, 3082, 2478, 2801, 2248, 1085,
+ 2293, 2282, 1736, 1446, 2376, 1937, 1999, 1396,
+ 1819, 2523, 2133, 1918, 1302, 2372, 2688, 2131,
+ 2901, 2151, 1734, 2537, 1979, 2142, 1917, 1873,
+ 2630, 2100, 2493, 2632, 2043, 2358, 2309, 1442,
+ 2696, 2510, 2615, 2715, 2103, 2023, 2162, 2403,
+ 2483, 2016, 2715, 2878, 2431, 3048, 1791, 2554,
+ 2281, 2292, 2340, 2609, 2499, 3013, 2933, 2682,
+ 1689)) # годовой убой
> # вычисление коэффициента корреляции Пирсона:
> cor.test (mydata$wt.Y,mydata$milk_yield.X)

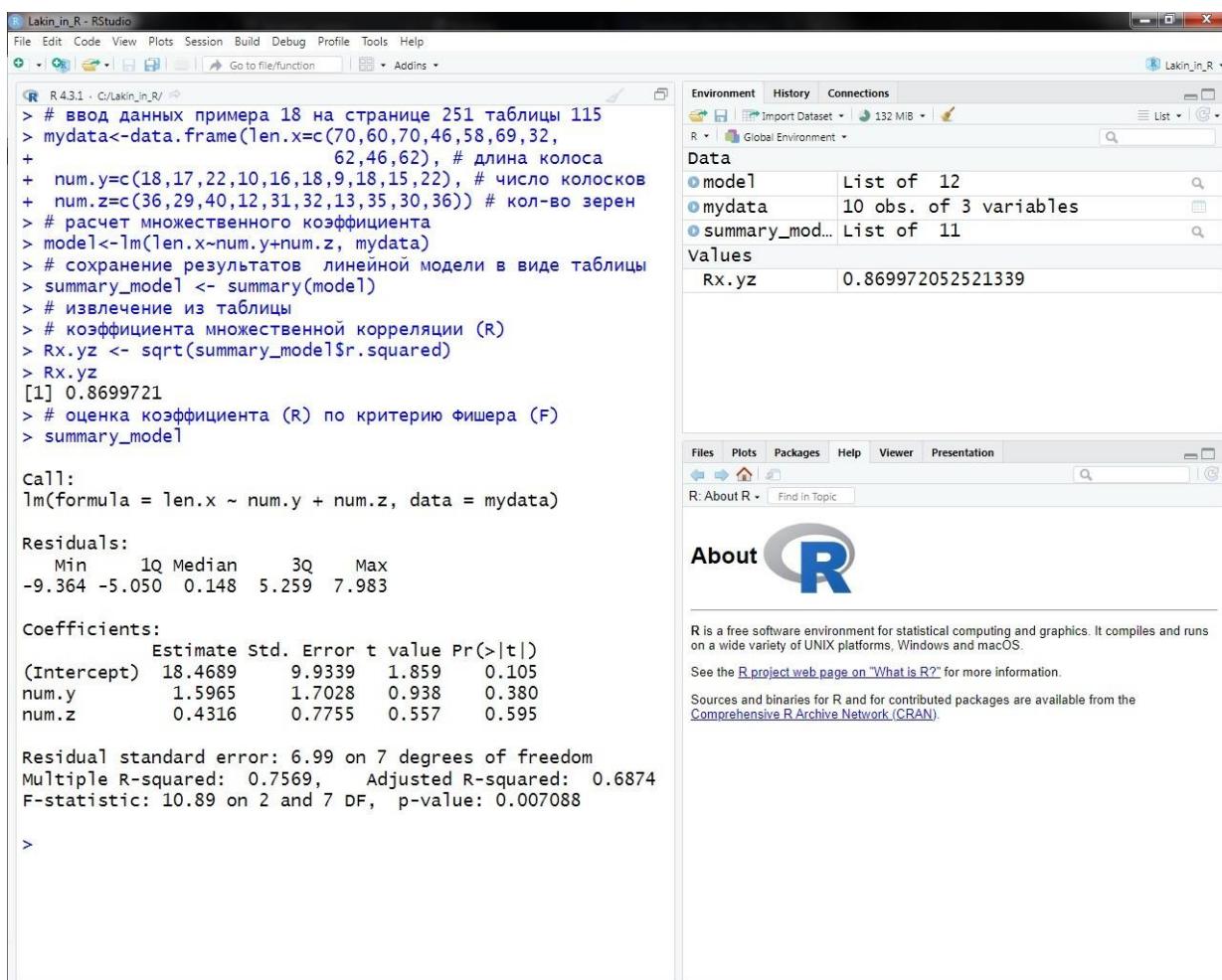
Pearson's product-moment correlation

data: mydata$wt.Y and mydata$milk_yield.X
t = 6.0974, df = 98, p-value = 2.14e-08
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
0.3656854 0.6535335
sample estimates:
cor
0.5244348
```
- Right Panel (Environment View):** Shows the environment with a data frame named "mydata" containing 100 observations of 2 variables.
- Bottom Panel (About R):** Provides information about R, stating it is a free software environment for statistical computing and graphics, compiled and run on various platforms.

Рисунок 5 — Скриншот окна RStudio во время сессии работы в среде R для расчёта коэффициента линейной корреляции Пирсона между массы тела коров горбановской породы и их годовым удоем молока

Рассмотрим способ расчёта множественной корреляции в среде R, который показывает тесноту линейной связи одного признака с совокупностью двух других. Обозначается множественный коэффициент корреляции как $R_{x,yz}$, $R_{y,xz}$ и $R_{z,xy}$, где до точки указывается признак, который изучается, после точки — факторы, влияющие на изучаемый признак. В среде R этот статистический показатель рассчитывается через построение модели линейной регрессии функцией $lm()$ из базового набора функций. Функция $lm()$ имеет следующий синтаксис: $lm(formula, data)$, где $formula$ описывает вид подбираемой модели; $data$ — имя объекта, в котором хранятся анализируемые переменные. Для расчёта коэффициента множественной корреляции $formula$ записывается в следующем виде: $Y \sim X1 + X2$.

В качестве демонстрации расчёта множественной корреляции воспользуемся примером 18 со страницы 251 (табл. 115) [8], в котором представлены данные по озимой ржи: число колосков, количество зёрен и длина колоса. Далее проведём оценку полученной величины корреляции по результатам построения линейной модели с помощью критерия Фишера (F) и его значения p (рис. 6).



The screenshot shows the RStudio interface with the following details:

- Code Editor:** Displays R code for loading data, fitting a linear model, and calculating the coefficient of multiple correlation.
- Environment View:** Shows the global environment with objects like `model`, `mydata`, and `Rx.yz`.
- Plots View:** Shows the "About R" page, which includes the R logo and information about the software.
- Help View:** Shows the "About R" page with links to the R project website and CRAN.

```

R 4.3.1 - C:/Lakin_in_R/
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins
R 4.3.1 - C:/Lakin_in_R/
> # ввод данных примера 18 на странице 251 таблицы 115
> mydata<-data.frame(len.x=c(70,60,70,46,58,69,32,
+                      62,46,62), # длина колоса
+ num.y=c(18,17,22,10,16,18,9,18,15,22), # число колосков
+ num.z=c(36,29,40,12,31,32,13,35,30,36)) # кол-во зерен
> # расчет множественного коэффициента
> model<-lm(len.x~num.y+num.z, mydata)
> # сохранение результатов линейной модели в виде таблицы
> summary_model <- summary(model)
> # извлечение из таблицы
> # коэффициента множественной корреляции (R)
> Rx.yz <- sqrt(summary_model$r.squared)
> Rx.yz
[1] 0.8699721
> # оценка коэффициента (R) по критерию Фишера (F)
> summary_model
Call:
lm(formula = len.x ~ num.y + num.z, data = mydata)

Residuals:
    Min      1Q Median      3Q     Max 
-9.364 -5.050  0.148  5.259  7.983 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 18.4689    9.9339   1.859   0.105    
num.y       1.5965    1.7028   0.938   0.380    
num.z       0.4316    0.7755   0.557   0.595    
Residual standard error: 6.99 on 7 degrees of freedom
Multiple R-squared:  0.7569, Adjusted R-squared:  0.6874 
F-statistic: 10.89 on 2 and 7 DF, p-value: 0.007088

>

```

Рисунок 6 – Скриншот окна RStudio во время сессии работы в среде R для расчёта коэффициента множественной корреляции между признаками озимой ржи (число колосков, количество зёрен и длина колоса) и оценка его значимости

Таким образом, множественный коэффициент корреляции между числом колосков, количеством зёрен и длиной колоса как зависимого признака составляет $R_{x,yz} \approx 0,87$, который значим на высоком уровне ($p\text{-value} = 0,007$).

В качестве примеров представлены лишь несколько статистических методов из учебного пособия Г. Ф. Лакина «Биометрия» [8], реализованных в среде R. Эти примеры показывают, что в совокупности с работой в R данное пособие по-прежнему актуально и в настоящее время. Мы считаем, что необходимо разработать специальное приложение к данному учебному пособию, в котором в виде скриптов на языке R с пояснениями и примечаниями представлены все статистические методы и примеры из этого учебника. Язык R рекомендуем вследствие ряда преимуществ — это бесплатно распространяемый программный продукт, который можно использовать и в частном порядке, и в организациях, обладающий фактически безграничными возможностями в статистике, моделировании и визуализации данных. Использование языка R широко распространено в научной среде по всему миру. Овладение навыками работы в данном программном продукте позволит быть востребованным сотрудником в научных учреждениях. Возможность написания

собственных функций и составления пакетов для проведения математических расчётов с последующим их размещением на репозитариях позволяет исследователям-разработчикам новых методов в статистике или моделировании биологических процессов, свободно распространять и внедрять в научную практику свои разработки по всему миру.

Список литературы

1. Антипина, Н. М. Непараметрический критерий различия для двух независимых выборок в табличном редакторе MS EXCEL / Н. М. Антипина, В. Н. Захаров, Ю. М. Протасов // Информационно-технологический вестник. — 2019. — № 1. — С. 48–58.
2. Баяхов, А. Н. Использование программы Microsoft Excel в биометрии / А. Н. Баяхов, Л. Г. Моисейкина // Устойчивое развитие: наука и практика. — 2016. — № 2 (17). — С. 47–54.
3. Грачев, В. С. Биометрическая обработка данных зоотехнического учета средствами EXCEL с использованием пакета анализа / В. С. Грачев. — Санкт-Петербург : Санкт-Петербургский гос. аграрн. ун-т, 2012. — 48 с.
4. Заплатин, Б. П. Биометрия с Excel / Б. П. Заплатин. — Пенза : Пензенский гос. пед. ун-т им. В. Г. Белинского, 2008. — 21 с.
5. Зарядов, И. С. Введение в статистический пакет R: типы переменных, структуры данных, чтение и запись информации, графика / И. С. Зарядов. — Москва : Российский ун-т дружбы народов, 2010. — 207 с.
6. Кабаков, Р. И. R в действии / Р. И. Кабаков. — 3-е изд. — Москва : ДМК Пресс, 2023. — 768 с.
7. Козак М. Ф. Биометрические методы в научных исследованиях / М. Ф. Козак, М. В. Козак. — Астрахань : Астраханский университет, 2019. — 168 с.
8. Лакин, Г. Ф. Биометрия / Г. Ф. Лакин. — 4-е изд., перераб. и доп. — Москва : Высшая школа, 1990. — 352 с.
9. Лебедько, Е. Я. Биометрия в MS EXCEL / Е. Я. Лебедько, А. М. Хохлов, Д. И. Барановский, О. М. Гетманец. — 3-е изд., стереотип. — Санкт-Петербург : Лань, 2022. — 172 с.
10. Мастицкий С. Э. Статистический анализ и визуализация данных с помощью R / С. Э. Мастицкий, В. К. Шитиков. — 2014. — URL: <https://github.com/ranalytics/r-tutorials> (дата обращения: 09.09.2025).
11. Мастицкий, С. Э. Визуализация данных с помощью ggplot2 / С. Э. Мастицкий. — Москва : ДМК Пресс, 2017. — 222 с.
12. Плохинский, Н. А. Биометрия / Н. А. Плохинский. — Изд. 2-е. — Москва : Московский гос. ун-т, 1970. — 367 с.
13. Плохинский, Н. А. Алгоритмы биометрии / Н. А. Плохинский. — 2-е изд., перераб. и доп. — Москва : Московский гос. ун-т, 1980. — 150 с.
14. Урбах, В. Ю. Статистический анализ в биологических и медицинских исследованиях / В. Ю. Урбах. — Москва : Медицина, 1975. — 295 с.
15. Чаргейшвили, С. В. Использование методов биометрии в MS Excel / С. В. Чаргейшвили, Н. П. Сударев, О. В. Абрампальская. — 2-е изд., доп. — Тверь : Тверская гос. с.-х. акад., 2023. — 98 с.
16. Шитиков, В. К. Классификация, регрессия и другие алгоритмы Data Mining с использованием R / В. К. Шитиков, С. Э. Мастицкий. — 2017. — URL: <https://github.com/ranalytics/data-mining> (дата обращения: 09.09.2025).

17. Шитиков, В. К. Рандомизация и бутстреп: статистический анализ в биологии и экологии с использованием R / В. К. Шитиков, Г. С. Розенберг. — Тольятти : Кассандра, 2013. — 314 с.
18. Logan, M. Biostatistical Design and Analysis Using R: A Practical Guide / M. Logan. — Hoboken : John Wiley and Sons, 2010. — 546 p. — doi: 10.1002/9781444319620.
19. Ogle, D. H. Introductory fisheries analyses with R / D. H. Ogle. — Chapman and Hall/CRC, 2016. — 327 p.
20. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing. — Vienna, Austria, 2023. — URL: <https://www.R-project.org/> (дата обращения: 09.09.2025).
21. Shahbaba, B. Biostatistics with R: An introduction to statistics through biological data / B. Shahbaba. — New York : Springer New York, 2012. — 352 p. — doi: 10.1007/978-1-4614-1302-8.
22. Zuur, A. F. Mixed effects models and extensions in ecology with R / A. F. Zuur, E. N. Ieno, N. J. Walker, A. A. Saveliev, G. M. Smith. — Springer, 2009. — 574 p. — doi: 10.1007/978-0-387-87458-6.

References

1. Antipina, N. M., Zakharov, V. N., Protasov, Yu. M. Nonparametric test of difference for two independent samples in the MS EXCEL spreadsheet editor. *Informatsionno-tehnologicheskiy vestnik = Information Technology Bulletin*. 2019; 1: 48–58.
2. Bayakhov, A. N., Moiseikina, L. G. Using Microsoft Excel in Biometrics. *Ustoychivoye razvitiye: nauka i praktika = Sustainable Development: Science and Practice*. 2016; 2 (17): 47–54.
3. Grachev, V. S. *Biometricheskaya obrabotka dannykh zootehnicheskogo ucheta sredstvami EXCEL s ispolzovaniyem paketa analiza = Biometric processing of zootechnical accounting data by EXCEL using the analysis package*. St. Petersburg: St. Petersburg State Agrarian University; 2012: 48 p.
4. Zaplatin, B. P. *Biometriya s Excel = Biometrics with Excel*. Penza: Penza State Pedagogical University named after V. G. Belinsky; 2008: 21 p.
5. Zaryadov, I. S. *Vvedeniye v statisticheskiy paket R: tipy peremennykh, struk-tury dannykh, chteniye i zapis' informatsii, grafika = Introduction to the R Statistical Package: Variable Types, Data Structures, Reading and Writing Information, Graphics*. Moscow: Peoples' Friendship University of Russia; 2010: 207 p.
6. Kabakov, R. I. *R v deystvii = R in Action*. 3rd ed. Moscow: DMK Press; 2023: 768 p.
7. Kozak, M. F., Kozak, M. V. *Biometricheskiye metody v nauchnykh issledovaniyah = Biometric Methods in Scientific Research*. Astrakhan: Astrakhan University; 2019: 168 p.
8. Lakin, G. F. *Biometriya = Biometrics*. 4th ed., revised and enlarged. Moscow: Vysshaya Shkola; 1990: 352 p.
9. Lebedko, E. Ya., Khokhlov, A. M., Baranovsky, D. I., Getmanets, O. M. *Biometriya v MS EXCEL = Biometrics in MS EXCEL*. 3rd ed., stereotype. St. Petersburg: Lan; 2022: 172 p.
10. Mastitsky, S. E., Shitikov, V. K. *Statisticheskiy analiz i vizualizatsiya dannykh s pomoshchyu R = Statistical Analysis and Visualization of Data with R*. 2014. Available at: <https://github.com/ranalytics/r-tutorials> (accessed: 09.09.2025).
11. Mastitsky, S. E. *Vizualizatsiya dannykh s pomoshchyu ggplot2 = Data Visualization with ggplot2*. Moscow: DMK Press; 2017: 222 p.
12. Plokhinsky, N. A. *Biometriya = Biometrics*. 2nd ed. Moscow: Moscow State University; 1970: 367 p.
13. Plokhinsky, N. A. *Algoritmy biometrii = Biometric Algorithms*. 2nd ed., revised. and add. Moscow: Moscow State University; 1980: 150 p.

14. Urbach, V. Yu. *Statisticheskiy analiz v biologicheskikh i meditsinskikh issledovaniyakh = Statistical Analysis in Biological and Medical Research*. Moscow: Meditsina; 1975: 295 p.
15. Chargeishvili, S. V., Sudarev, N. P., Abrampalskaya, O. V. *Ispolzovaniye metodov biometrii v MS Excel = Using Biometric Methods in MS Excel*. 2nd ed., add. Tver State Agricultural Academy; 2023: 98 p.
16. Shitikov, V. K., Mastitsky, S. E. *Klassifikatsiya, regressiya i drugiye algoritmy Data Mining s ispolzovaniyem R = Classification, Regression, and Other Data Mining Algorithms Using R*. 2017. Available at: <https://github.com/ranalytics/data-mining> (accessed: 09.09.2025).
17. Shitikov, V. K., Rosenberg, G. S. *Randomizatsiya i butstrep: statisticheskiy analiz v biologii i ekologii s ispolzovaniyem R = Randomization and bootstrap: statistical analysis in biology and ecology using R*. Tolyatti: Kassandra; 2013: 314 p.
18. Logan, M. *Biostatistical Design and Analysis Using R: A Practical Guide*. Hoboken: John Wiley and Sons; 2010: 546 p. doi: 10.1002/9781444319620.
19. Ogle, D. H. *Introductory fisheries analyses with R*. Chapman and Hall/CRC; 2016: 327 p.
20. R Core Team. *R: A language and environment for statistical computing. R Foundation for Statistical Computing*. Vienna, Austria; 2023. Available at: <https://www.R-project.org/> (accessed: 09.09.2025).
21. Shahbaba, B. *Biostatistics with R: An introduction to statistics through biological data*. New York: Springer New York; 2012: 352 p. doi: 10.1007/978-1-4614-1302-8.
22. Zuur, A. F., Ieno, E. N., Walker, N. J., Saveliev, A. A., Smith, G. M. *Mixed effects models and extensions in ecology with R*. Springer; 2009: 574 p. doi: 10.1007/978-0-387-87458-6.

Информация об авторах

Сафаралиев И. А. — кандидат биологических наук, заведующий лабораторией;
Пучков М. Ю. — доктор сельскохозяйственных наук, доцент, профессор кафедры.

Information about the authors

Safaraliev I. A. — Candidate of Biological Sciences, Head of The Sturgeon Laboratory;
Puchkov M. Yu. — Doctor of Agricultural Sciences, Professor of the Department.

Вклад авторов

Все авторы сделали эквивалентный вклад в публикации.
Авторы заявляют об отсутствии конфликта интересов.

Contribution of the authors

All authors have made equivalent contributions to publications.
The authors declare that there is no conflict of interest.

Статья поступила в редакцию 05.09.2025; одобрена после рецензирования 12.09.2025; принята к публикации 14.09.2025.

The article was submitted 05.09.2025; approved after reviewing 12.09.2025; accepted for publication 14.09.2025.