

DOI: <https://doi.org/10.17816/DD121368>

Проверка гипотез исследования с использованием языка R

И.А. Блохин¹, М.Р. Коденко^{1,2}, Ю.Ф. Шумская^{1,3}, А.П. Гончар¹, Р.В. Решетников¹¹ Научно-практический клинический центр диагностики и телемедицинских технологий, Москва, Российская Федерация² Московский государственный технический университет имени Н.Э. Баумана (национальный исследовательский университет), Москва, Российская Федерация³ Первый Московский государственный медицинский университет имени И.М. Сеченова (Сеченовский Университет), Москва, Российская Федерация

АННОТАЦИЯ

Для современных учёных становятся всё более важными компетенции в области статистической обработки данных. Очевидными преимуществами открытого программного обеспечения (open-source software) для статистического анализа являются доступность и многофункциональность. Наиболее широкими возможностями среди бесплатных решений обладают язык программирования и соответствующее программное обеспечение R, доступное в виде минимального консольного интерфейса или полноценной среды разработки RStudio/Posit.

Предлагаем вашему вниманию практическое руководство по сравнению двух групп с помощью инструментов языка R на примере сопоставления эффективной дозы, полученной при проведении стандартной компьютерной и низкодозной компьютерной томографии при COVID-19, в котором кратко обобщены теоретические подходы к обработке медицинских данных, а также рекомендации по корректной формулировке задач исследования и выбора оптимальных методов статистического анализа.

Основная задача практического руководства — познакомить читателя с интерфейсом Posit и базовым функционалом языка R на практическом примере решения реальной медицинской задачи. Представленный материал может быть полезен на начальном этапе освоения статистического анализа с помощью инструментов языка R.

Ключевые слова: R; RStudio; Posit; статистический анализ; медицинские данные; практическое руководство.

Как цитировать

Блохин И.А., Коденко М.Р., Шумская Ю.Ф., Гончар А.П., Решетников Р.В. Проверка гипотез исследования с использованием языка R // *Digital Diagnostics*. 2023. Т. 4, № 2. С. 238–247. DOI: <https://doi.org/10.17816/DD121368>

DOI: <https://doi.org/10.17816/DD121368>

Hypothesis testing using R

Ivan A. Blokhin¹, Maria R. Kodenko^{1,2}, Yuliya F. Shumskaya^{1,3},
Anna P. Gonchar¹, Roman V. Reshetnikov¹

¹ Moscow Center for Diagnostics and Telemedicine, Moscow, Russian Federation

² Bauman Moscow State Technical University, Moscow, Russian Federation

³ The First Sechenov Moscow State Medical University, Moscow, Russian Federation

ABSTRACT

Competencies in statistical data processing are becoming increasingly important for modern scientists. The apparent advantages of open-source software for statistical analysis are its accessibility and adaptability. The programming language and the corresponding software R, available as a minimalistic console interface or a complete development environment RStudio/Posit, have the widest possibilities among free solutions.

We present a practical guide for comparing two groups using the software R. This study compares the effective doses of standard computed tomography with low-dose computed tomography for COVID-19 patients. The practical guide summarizes theoretical approaches to medical data processing and recommendations for correctly formulating research tasks and selecting optimal statistical analysis methods.

The main goal of the practical guide is to introduce the reader to the Posit interface and the basic functionality of the R language by using a practical example of treating a real medical problem. The presented material can be useful as an introduction to statistical analysis using the programming language R.

Keywords: medical data, Posit, R, RStudio, statistical analysis, tutorial

To cite this article

Blokhin IA, Kodenko MR, Shumskaya YuF, Gonchar AP, Reshetnikov RV. Hypothesis testing using R. *Digital Diagnostics*. 2023;4(2):238–247.

DOI: <https://doi.org/10.17816/DD121368>

Received: 11.01.2023

Accepted: 18.01.2023

Published: 10.04.2023

DOI: <https://doi.org/10.17816/DD121368>

使用R语言进行研究假设检验

Ivan A. Blokhin¹, Maria R. Kodenko^{1,2}, Yuliya F. Shumskaya^{1,3},
Anna P. Gonchar¹, Roman V. Reshetnikov¹

¹ Moscow Center for Diagnostics and Telemedicine, Moscow, Russian Federation

² Bauman Moscow State Technical University, Moscow, Russian Federation

³ The First Sechenov Moscow State Medical University, Moscow, Russian Federation

简评

对于现代科学家来说，统计数据处理的能力越来越重要。用于统计分析的开源软件（open-source software）的明显优势是可用性和多功能性。在免费的解决方案中，R语言和相关软件大有可为，可作为一个最简控制台界面或作为一个完全合格的开发环境RStudio/Posit。

我们提供一份使用R语言工具比较两组数据的实用指南，以COVID-19的标准电子计算机断层扫描和低剂量电脑断层扫描的有效剂量比较为例。本指南简略地总结了医学数据处理的理论方法，以及正确制定研究目标和选择最佳统计分析方法的建议。

本实用指南的主要目的是通过一个解决真实医学问题的实际例子向读者介绍Posit界面和R语言的基本功能。所介绍的材料在借助R语言工具掌握统计分析的初始阶段可以有益处。

关键词：R; RStudio; Posit; 统计分析; 医学数据; 实用手册。

To cite this article

Blokhin IA, Kodenko MR, Shumskaya YuF, Gonchar AP, Reshetnikov RV. 使用R语言进行研究假设检验. *Digital Diagnostics*. 2023;4(2):238–247.

DOI: <https://doi.org/10.17816/DD121368>

收到: 11.01.2023

接受: 18.01.2023

发布日期: 10.04.2023

ВВЕДЕНИЕ

Для современных учёных становятся всё более важными компетенции в области статистической обработки данных. В 1983 году для понимания 2/3 из 760 научных публикаций престижного *New England Journal of Medicine* читателю было достаточно владеть базовыми понятиями описательной статистики (процент, среднее значение и стандартное отклонение) и одним методом статистической проверки гипотез (t -критерий Стьюдента) [1]. Со временем научное сообщество столкнулось с ограниченностью такого подхода к анализу данных. В 2014 году *Journal of American Medical Association* анонсировал специальную серию выпусков, направленную на разбор методов статистического анализа для врачей-клиницистов¹. Сегодня в открытом доступе можно найти обзорные работы не только по базовым аспектам анализа данных [2], но и выбору подходящего метода их обработки [3]. Более того, в 2019 году опубликован подробный чек-лист по использованию статистических методов в биомедицинских исследованиях (*Statistical Analysis Methods in Biomedical Research*) [4]. Однако практическое применение методов анализа данных оставалось бы по-прежнему нетривиальной задачей, если бы не активная разработка программ для его автоматизации.

На фоне большого числа платных программ статистического анализа и наличия коммерческих предложений обработки данных «под ключ» очевидными преимуществами открытого программного обеспечения (*open-source software*) являются доступность и многофункциональность [5]. Наиболее широкими возможностями среди бесплатных решений обладает язык программирования и соответствующее программное обеспечение R [6], доступное в виде минималистичного консольного интерфейса [7] или полноценной среды разработки RStudio/Posit [8].

Цель настоящего материала — познакомить читателя с базовыми операциями в R на примере задачи из реальной практики: сравнения эффективной дозы, полученной пациентами при проведении стандартной компьютерной томографии (КТ) и низкодозной компьютерной томографии (НДКТ) при COVID-19. Актуальность задачи основана на необходимости мониторинга лучевой нагрузки [9], в том числе в связи с нарастанием количества выполняемых лучевых исследований [10] и вследствие этого важности разработки [11] и клинической апробации [12, 13] протоколов НДКТ.

При проведении статистического анализа необходимо придерживаться основополагающих этапов.

Постановка задачи и формулировка нулевой гипотезы анализа

Нулевая гипотеза — отправная точка статистического анализа. Для задач сравнения двух групп нулевая гипотеза (H_0) формулируется следующим образом: «Статистически значимые различия отсутствуют». В настоящем примере проводилось сравнение эффективной дозы облучения пациентов для двух типов КТ-исследования (полнодозной и низкодозной) при COVID-19. Другими словами, формулировка H_0 для данной задачи: «Лучевая нагрузка при КТ и НДКТ сопоставима».

Необходимо также помнить про «альтернативную гипотезу» (H_1) — это гипотеза-антагонист H_0 , они должны быть взаимоисключающими. В данном примере альтернативная гипотеза будет звучать так: «Эффективная доза облучения пациентов с COVID-19 при КТ и НДКТ статистически значимо отличается».

Результатом проверки нулевой гипотезы в статистике является так называемое p -value, т.е. вероятность ошибочно отвергнуть H_0 . Это значение можно интерпретировать следующим образом: «Если мы многократно повторим эксперимент и отвергнем нулевую гипотезу, то ошибемся в p -value из 100% случаев». Например, для p -value 0,03 отвергаем H_0 в пользу H_1 и ошибаемся в 3% случаев. Много это или мало, позволяет решить заранее выбранное граничное значение. Благодаря Роналду Фишеру, в большинстве случаев граничное значение p -value берут равным 0,05 [14]. Возвращаясь к примеру и опираясь на границу 0,05, для p -value 0,03 мы можем уверенно заключить, что сравниваемые выборки разные.

Анализ исходных данных

Выбор метода статистического анализа исходных данных зависит от их типа и характера распределения. Выделяют количественные и качественные данные². Количественные данные характеризуют величину явления или число объектов: например, лучевую нагрузку в милливертах при проведении КТ-исследования органов грудной клетки. Качественные, или категориальные, данные описывают отношение исследуемого явления к определённой группе: например, пол пациента или категория по шкале «КТ0–4».

При анализе данных в дополнение к основной гипотезе исследования всегда необходимо проверять ещё одну H_0 : «Данные распределены нормально». Нормальное распределение — одно из самых важных в области статистики, поскольку часто встречается в возникающих естественным путём явлениях: ему подчиняются рост, вес,

¹ JAMA guide to statistics and methods [Electronic resource]. Режим доступа: <https://jamanetwork.com/collections/44042/jama-guide-to-statistics-and-methods>.

² Медицинская статистика [электронный ресурс]. Общие вопросы статистики. Режим доступа: <https://medstatistic.ru/statistics/statistics3.html>.

размер обуви, а также большое количество других характеристик популяции. Нормальное распределение описывается всего двумя параметрами — средним значением и среднеквадратичным отклонением, и это допущение лежит в основе ряда статистических подходов к проверке гипотез.

Выбор метода проверки распределения на нормальность — задача, не имеющая единственного решения на все случаи жизни. Так, X. Romão и соавт. [15] провели сравнение 33 таких методов и предложили оптимальные решения в зависимости от типа данных. Следует отметить, что выбор метода также зависит от размера исследуемой выборки [16], при этом самыми популярными являются критерий Колмогорова–Смирнова и тест Шапиро–Уилка [17].

Проверка нулевой гипотезы исследования

Помимо типа данных в сравниваемых выборках, корректный метод статистического анализа должен учитывать количество сопоставляемых групп, а также наличие связи между ними. Например, получены ли данные КТ- и НДКТ-исследований для одного и того же пациента или у разных. Различных статистических тестов разработано уже более 50, и для выбора оптимального метода существуют специальные онлайн-ресурсы³.

В нашем примере выборки парные, так как КТ- и НДКТ-данные получены последовательно для одних и тех же пациентов. При нормальном распределении целесообразно использовать парный t-критерий Стьюдента,

а критерий Уилкоксона — при распределении, отличным от нормального.

ПРАКТИЧЕСКИЙ ПРИМЕР

Для проведения статистического анализа были установлены программный пакет R (версия 4.2.2, <https://cloud.r-project.org/>) и интерфейс Posit (версия 353, <https://posit.co/download/rstudio-desktop/>, ранее RStudio).

1. Базовый интерфейс Posit можно разделить на консоль (console), окружение (environment) и файловый менеджер (files) (рис. 1).

Импорт данных осуществляется посредством опции File-Import Dataset. В нашем случае производился импорт таблицы Excel, содержащей данные об эффективных дозах КТ и НДКТ.

После импорта интерфейс программы изменится: в блоке окружения будет отображаться переменная с именем «data» (рис. 2). В левом верхнем квадранте появится также новая область — блок просмотра данных, в котором будет отображаться загруженная таблица. Сравнимые колонки в таблице названы как «Effective Dose (CT)» и «Effective Dose (LDCT)». Использование латиницы в названиях колонок позволяет избежать проблем с кодировкой и некорректным отображением символов в Posit.

Для удобства дальнейшей работы для каждой из этих колонок создадим отдельную переменную, выполнив в консоли команды, приведённые на рис. 3.

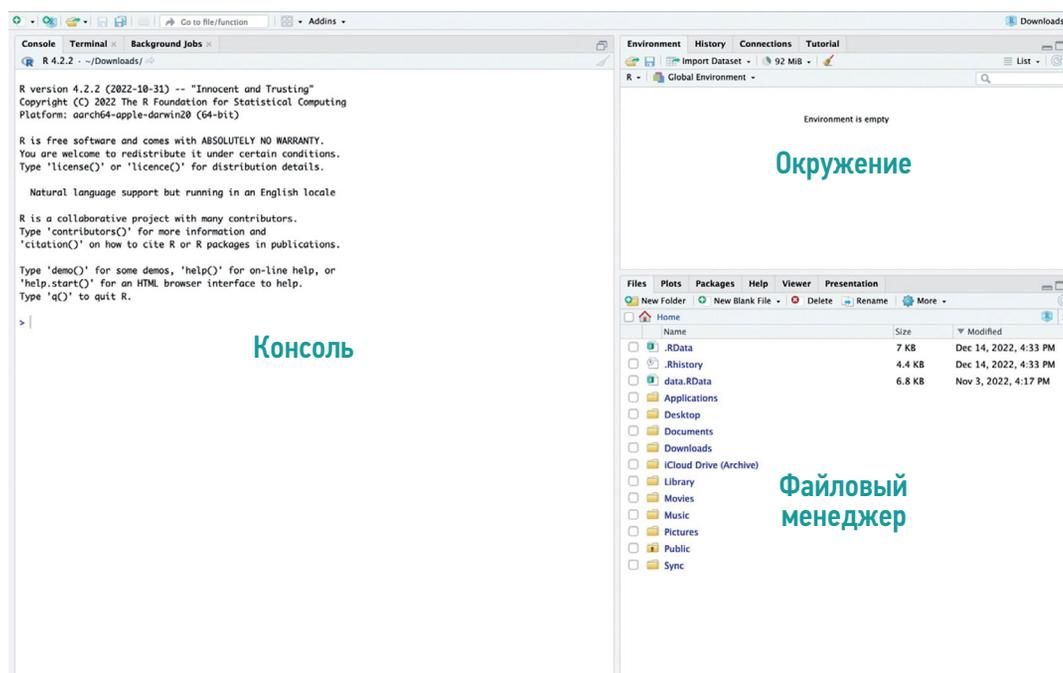


Рис. 1. Интерфейс программы Posit с указанием областей консоли, окружения и файлового менеджера.

³ Statistics online--checks assumptions, interprets results (<https://www.statskingdom.com/>); Медицинская статистика. Выбор метода статистического анализа (<https://medstatistic.ru/calculators/calcchoice.html>).

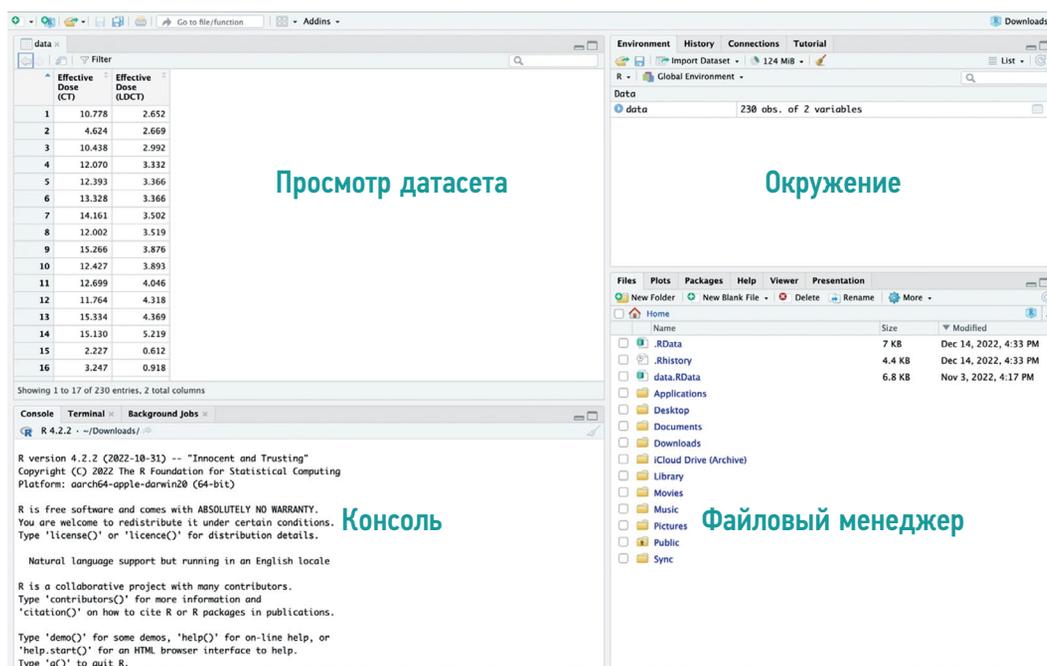


Рис. 2. Интерфейс Posit после импорта файла: в левом верхнем квадранте экрана появилось окно с загруженными столбцами набора данных, а в правом верхнем квадранте указано количество столбцов (variables) и строк (obs., от англ. observations — наблюдения).

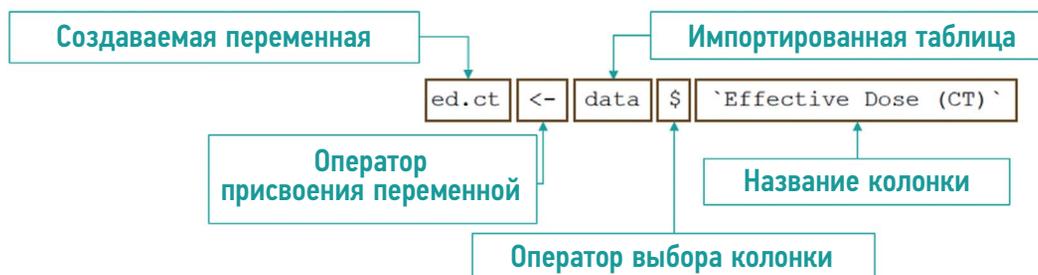


Рис. 3. Создание отдельной переменной для эффективной дозы компьютерной томографии с указанием функции каждого элемента команды.

Использование оператора «\$» (см. рис. 3) отображает выпадающее меню столбцов в импортированной таблице, облегчающее ввод команды. После создания переменных они будут отображаться в правом верхнем квадранте интерфейса как «values» (рис. 4).

2. Проверка нормальности распределения загруженных данных тестом Шапиро–Уилка осуществляется командой `shapiro.test` (рис. 5).

Полученные *p*-value для обеих переменных значительно меньше выбранного граничного значения $p=0,05$, поэтому распределение данных для эффективной дозы как для КТ, так и НДКТ отличается от нормального. Следовательно, для проверки основной гипотезы исследования целесообразно использовать критерий Уилкоксона.

3. Проверка нулевой гипотезы для связанных количественных выборок с распределением данных, отличных от нормального, выполняется командой `wilcox.test` (рис. 6, 7).

Полученное *p*-value значительно меньше выбранного граничного значения $p=0,05$, поэтому H_0 можно отвергнуть в пользу H_1 . Это означает, что лучевая нагрузка при КТ и НДКТ статистически значимо отличается.

ЗАКЛЮЧЕНИЕ

Основная задача изложенного материала — познакомить читателя с интерфейсом Posit и базовым функционалом языка R на практическом примере решения реальной медицинской задачи.

В работе кратко обобщены теоретические подходы к обработке медицинских данных, а также рекомендации по корректной формулировке задач исследования и выбору оптимальных методов статистического анализа.

Представленный материал может быть полезен на начальном этапе освоения статистического анализа с помощью инструментов языка R.

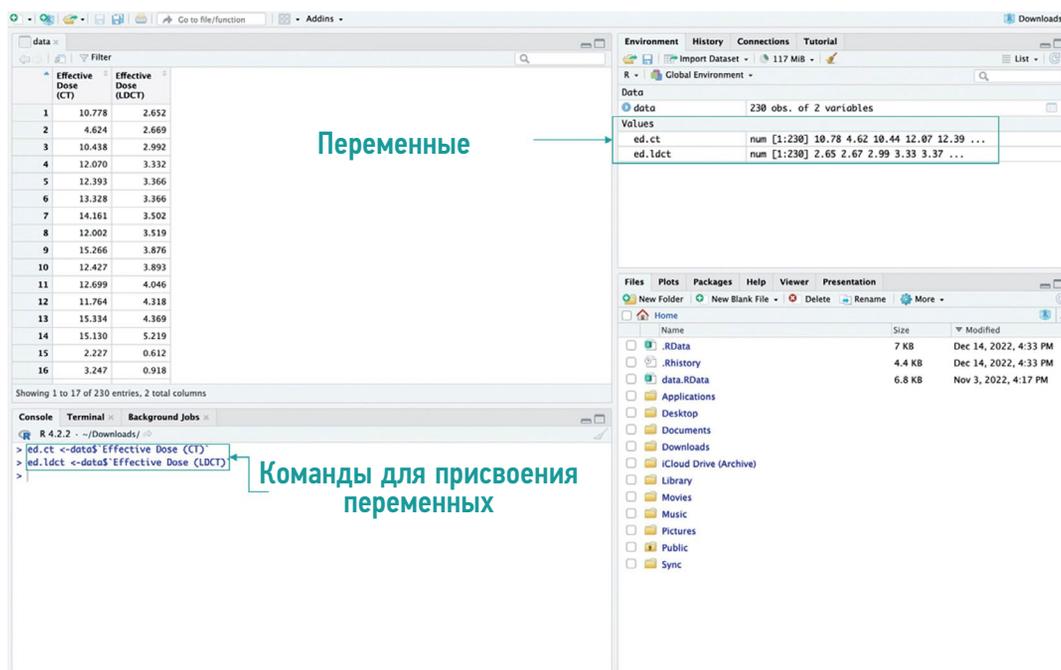


Рис. 4. Окно Posit после импорта файла и присвоения значений переменным: в правом верхнем квадранте появились новые переменные с предпросмотром первых пяти значений в каждой, в левом нижнем квадранте — консольный интерфейс для выполнения команд.

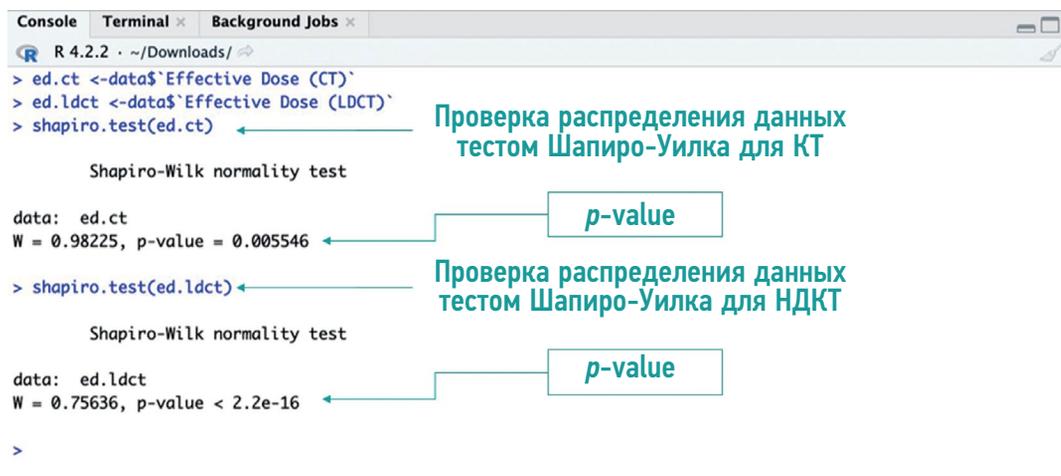


Рис. 5. Область с консольным интерфейсом Posit. Проверка нормальности распределения данных тестом Шапиро–Уилка.

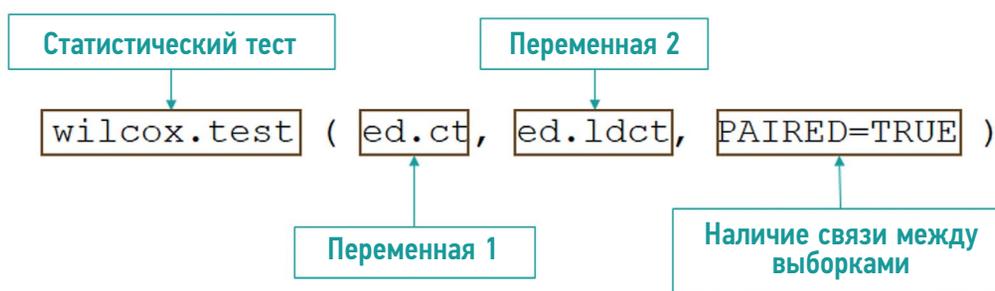


Рис. 6. Проведение теста Уилкоксона с указанием функции каждого элемента команды.

```

R 4.2.2 . ~/Downloads/
> wilcox.test(ed.ct, ed.ldct, PAIRED=TRUE)

Wilcoxon rank sum test with continuity correction

data: ed.ct and ed.ldct
W = 50842, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0
  
```

Проверка нулевой гипотезы исследования с помощью критерия Уилкоксона

Рис. 7. Проверка нулевой гипотезы исследования с помощью критерия Уилкоксона.

ДОПОЛНИТЕЛЬНО

Источник финансирования. Данная статья подготовлена авторским коллективом в рамках научно-практического проекта в сфере медицины (№ ЕГИСУ: 123031500005-2) «Научное обоснование методов лучевой диагностики опухолевых заболеваний с использованием радиомического анализа».

Конфликт интересов. Авторы заявляют об отсутствии явных и потенциальных конфликтов интересов, связанных с публикацией настоящей статьи.

Вклад авторов. Все авторы подтверждают соответствие своего авторства международным критериям ICMJE (все авторы внесли вклад в разработку концепции, проведение поисково-аналитической работы и подготовку статьи, прочли и одобрили финальную версию перед публикацией). Наибольший вклад распределён следующим образом: И.А. Блохин — написание текста статьи, редактирование и утверждение итогового варианта текста рукописи; М.Р. Коденко, Ю.Ф. Шумская, А.П. Гончар — редактирование и утверждение итогового варианта текста рукописи; Р.В. Решетников — редактирование и утверждение итогового варианта текста рукописи, консультативная поддержка.

ADDITIONAL INFORMATION

Funding source. This article was prepared by a group of authors as a part of the medical research project (No. USIS [in the Unified State Information System for Accounting of Research, Development, and Technological Works]: 123031500005-2) «Scientific rationale for diagnostic radiology in oncology using radiomics».

Competing interests. The authors declare that they have no competing interests.

Authors' contribution. All authors made a substantial contribution to the conception of the work, acquisition, analysis, interpretation of data for the work, drafting and revising the work, final approval of the version to be published and agree to be accountable for all aspects of the work. I.A. Blokhin — writing the text of the article, editing and approval of the final version of the text of the manuscript; M.R. Kodenko, Yu.F. Shumskaya, A.P. Gonchar — editing and approval of the final version of the text of the manuscript; R.V. Reshetnikov — editing and approval of the final version of the text of the manuscript, advisory support.

СПИСОК ЛИТЕРАТУРЫ

- Emerson J.D., Colditz G.A. Use of statistical analysis in the New England Journal of Medicine // *New Engl J Med*. 1983. Vol. 309, N 12. P. 709–713. doi: 10.1056/NEJM198309223091206
- Ali Z., Bhaskar S.B. Basic statistical tools in research and data analysis // *Indian J Anaesth*. 2016. Vol. 60, N 9. P. 662–669. doi: 10.4103/0019-5049.190623
- Mishra P., Pandey C.M., Singh U., et al. Selection of appropriate statistical methods for data analysis // *Ann Card Anaesth*. 2019. Vol. 22, N 3. P. 297–301. doi: 10.4103/aca.ACA_248_18
- Dwivedi A.K., Shukla R. Evidence-based statistical analysis and methods in biomedical research (SAMBR) checklists according to design features // *Cancer Rep (Hoboken)*. 2020. Vol. 3, N 4. P. e1211. doi: 10.1002/cnr2.1211
- Rigby P.C., German D.M., Cowen L., et al. Peer review on open-source software projects: Parameters, statistical models, and theory // *ACM Trans Softw Eng Methodol*. 2014. Vol. 23, N 4. P. 35.
- Culpepper S.A., Aguinis H. R is for revolution: A cutting-edge, free, open source statistical package // *Organizational Research Methods*. 2011. Vol. 14, N 4. P. 735–740.
- Ihaka R., Gentleman R. R: A language for data analysis and graphics // *J Computational Graphical Statistics*. 1996. Vol. 5, N 3. P. 299–314. doi:10.1080/10618600.1996.10474713
- Niu G., Segall R.S., Zhao Z., et al. A survey of open source statistical software (OSSS) and their data processing functionalities // *Int J Open Source Software Processes*. 2021. Vol. 12, N 1. P. 1–20. doi: 10.4018/IJOSSP.2021010101
- Shatenok M.P., Ryzhov S.A., Lantukh Z.A., et al. Patient dose monitoring software in radiology // *Digital Diagnostics*. 2022. Vol. 3, N 3. P. 212–230. doi: 10.17816/DD106083
- Druzhinina U.V., Ryzhov S.A., Vodovatov A.V., et al. Coronavirus Disease-2019: Changes in computed tomography radiation burden across Moscow medical facilities // *Digital Diagnostics*. 2022. Vol. 3, N 1. P. 5–15. doi: 10.17816/DD87628

11. Gombolevskiy V., Morozov S., Chernina V., et al. A phantom study to optimise the automatic tube current modulation for chest CT in COVID-19 // *Eur Radiol Exp*. 2021. Vol. 5, N 1. P. 21. doi: 10.1186/s41747-021-00218-0
12. Blokhin I., Gombolevskiy V., Chernina V., et al. Inter-observer agreement between low-dose and standard-dose CT with soft and sharp convolution kernels in COVID-19 pneumonia // *J Clin Med*. 2022. Vol. 11, N 3. P. 669. doi: 10.3390/jcm11030669
13. Blokhin I.A., Gonchar A.P., Kodenko M., et al. Impact of body mass index on the reliability of the CT0–4 grading system: A comparison of computed tomography protocols: 2 // *Digital Diagnostics*. 2022. Vol. 3, N 2. P. 108–118. doi: 10.17816/DD104358
14. Kennedy-Shaffer L. Before $p < 0.05$ to beyond $p < 0.05$: Using history to contextualize p-values and significance

- testing // *Am Stat*. 2019. Vol. 73, Suppl. 1. P. 82–90. doi: 10.1080/00031305.2018.1537891
15. Romão X., Delgado R., Costa A. An empirical power comparison of univariate goodness-of-fit tests for normality // *J Statist Computation Simulat*. 2010. Vol. 80, N 5. P. 545–591. doi: 10.1080/00949650902740824
16. Lumley T., Diehr P., Emerson S., Chen L. The importance of the normality assumption in large public health data sets // *Ann Rev Pub Health*. 2002. Vol. 23, N 1. P. 151–169. doi: 10.1146/annurev.pubhealth.23.100901.140546
17. Mishra P., Pandey C.M., Singh U., et al. Descriptive statistics and normality tests for statistical data // *Ann Card Anaesth*. 2019. Vol. 22, N 1. P. 67–72. doi: 10.4103/aca.ACA_157_18

REFERENCES

1. Emerson J, Colditz GA. Use of statistical analysis in the New England Journal of Medicine. *New Engl J Med*. 1983;309(12):709–713. doi: 10.1056/NEJM198309223091206
2. Ali Z, Bhaskar SB. Basic statistical tools in research and data analysis. *Indian J Anaesth*. 2016;60(9):662–669. doi: 10.4103/0019-5049.190623
3. Mishra P, Pandey CM, Singh U, et al. Selection of appropriate statistical methods for data analysis. *Ann Card Anaesth*. 2019;22(3):297–301. doi: 10.4103/aca.ACA_248_18
4. Dwivedi AK, Shukla R. Evidence-based statistical analysis and methods in biomedical research (SAMBR) checklists according to design features. *Cancer Rep (Hoboken)*. 2020;3(4):e1211. doi: 10.1002/cnr2.1211
5. Rigby PC, German DM, Cowen L, et al. Peer review on open-source software projects: Parameters, statistical models, and theory. *ACM Trans Softw Eng Methodol*. 2014;23(4):35.
6. Culpepper SA, Aguinis H. R is for revolution: A cutting-edge, free, open source statistical package. *Organizational Research Methods*. 2011;14(4):735–740.
7. Ihaka R, Gentleman R. R: A language for data analysis and graphics. *J Computational Graphical Statistics*. 1996;5(3):299–314. doi: 10.1080/10618600.1996.10474713
8. Niu G, Segall RS, Zhao Z, et al. A survey of open source statistical software (OSSS) and their data processing functionalities. *Int J Open Source Software Processes*. 2021;12(1):1–20. doi: 10.4018/IJOSSP.2021010101
9. Shatenok MP, Ryzhov SA, Lantukh ZA, et al. Patient dose monitoring software in radiology. *Digital Diagnostics*. 2022;3(3):212–230. (In Russ). doi: 10.17816/DD106083

10. Druzhinina UV, Ryzhov SA, Vodovатов AV, et al. Coronavirus Disease-2019: Changes in computed tomography radiation burden across Moscow medical facilities. *Digital Diagnostics*. 2022;3(1):5–15. (In Russ). doi: 10.17816/DD87628
11. Gombolevskiy V, Morozov S, Chernina V, et al. A phantom study to optimise the automatic tube current modulation for chest CT in COVID-19. *Eur Radiol Exp*. 2021;5(1):21. doi: 10.1186/s41747-021-00218-0
12. Blokhin I, Gombolevskiy V, Chernina V, et al. Inter-observer agreement between low-dose and standard-dose CT with soft and sharp convolution kernels in COVID-19 pneumonia. *J Clin Med*. 2022;11(3):669. doi: 10.3390/jcm11030669
13. Blokhin IA, Gonchar AP, Kodenko M, et al. Impact of body mass index on the reliability of the CT0–4 grading system: A comparison of computed tomography protocols. *Digital Diagnostics*. 2022;3(2):108–118. (In Russ). doi: 10.17816/DD104358
14. Kennedy-Shaffer L. Before $p < 0.05$ to beyond $p < 0.05$: Using history to contextualize p-values and significance testing. *Am Stat*. 2019;73(Suppl. 1):82–90. doi: 10.1080/00031305.2018.1537891
15. Romão X, Delgado R, Costa A. An empirical power comparison of univariate goodness-of-fit tests for normality. *J Statist Computation Simulat*. 2010;80(5):545–591. doi: 10.1080/00949650902740824
16. Lumley T, Diehr P, Emerson S, Chen L. The Importance of the normality assumption in large public health data sets. *Ann Rev Pub Health*. 2002;23(1):151–169. doi: 10.1146/annurev.pubhealth.23.100901.140546
17. Mishra P, Pandey CM, Singh U, et al. Descriptive statistics and normality tests for statistical data. *Ann Card Anaesth*. 2019;22(1):67–72. doi: 10.4103/aca.ACA_157_18

ОБ АВТОРАХ

* Блохин Иван Андреевич;

адрес: Россия, 127051, Москва, ул. Петровка, д. 24, стр. 1;
ORCID: <https://orcid.org/0000-0002-2681-9378>;
eLibrary SPIN: 3306-1387; e-mail: [BlokhinIA@zdrav.mos.ru](mailto:blokhinIA@zdrav.mos.ru)

Коденко Мария Романовна;

ORCID: <https://orcid.org/0000-0002-0166-3768>;
eLibrary SPIN: 5789-0319; e-mail: KodenkoMR@zdrav.mos.ru

AUTHORS' INFO

* Ivan A. Blokhin;

address: 24/1 Petrovka street, 127051 Moscow, Russia;
ORCID: <https://orcid.org/0000-0002-2681-9378>;
eLibrary SPIN: 3306-1387; e-mail: [BlokhinIA@zdrav.mos.ru](mailto:blokhinIA@zdrav.mos.ru)

Maria R. Kodenko;

ORCID: <https://orcid.org/0000-0002-0166-3768>;
eLibrary SPIN: 5789-0319; e-mail: KodenkoMR@zdrav.mos.ru

* Автор, ответственный за переписку / Corresponding author

Шумская Юлия Федоровна;

ORCID: <https://orcid.org/0000-0002-8521-4045>;
eLibrary SPIN: 3164-5518; e-mail: ShumskayaYF@zdrav.mos.ru

Гончар Анна Павловна;

ORCID: <https://orcid.org/0000-0001-5161-6540>;
eLibrary SPIN: 3513-9531; e-mail: GoncharAP@zdrav.mos.ru

Решетников Роман Владимирович, к.ф.-м.н.;

ORCID: <https://orcid.org/0000-0002-9661-0254>;
eLibrary SPIN: 8592-0558; e-mail: r.reshetnikov@gmail.com

Yuliya F. Shumskaya;

ORCID: <https://orcid.org/0000-0002-8521-4045>;
eLibrary SPIN: 3164-5518; e-mail: ShumskayaYF@zdrav.mos.ru

Anna P. Gonchar;

ORCID: <https://orcid.org/0000-0001-5161-6540>;
eLibrary SPIN: 3513-9531; e-mail: GoncharAP@zdrav.mos.ru

Roman V. Reshetnikov, Cand. Sci. (Phys-Math);

ORCID: <https://orcid.org/0000-0002-9661-0254>;
eLibrary SPIN: 8592-0558; e-mail: r.reshetnikov@gmail.com