

В основе доказательной медицины лежит корректное использование методов биомедицинской статистики. Выбор адекватных статистических критериев и использование репрезентативной выборки обеспечивают правильную интерпретацию результатов исследования. Грамотная статистическая обработка необходима и для публикации в научных изданиях: использование наиболее подходящих для цели исследования статистических методов является обязательным условием публикации в ведущих журналах, индексируемых такими международными базами данных, как Scopus и Web of Science, а также отечественной информационно-аналитической системой Science Index.

Редакция нашего журнала также уделяет этому вопросу самое пристальное внимание: каждая рукопись рецензируется экспертом по биомедицинской статистике. Мы планируем знакомить наших авторов и читателей с наиболее распространенными методами статистической обработки данных, применяемых в биомедицинских исследованиях, в том числе и в травматологии и ортопедии.

Первая публикация посвящена обзору программных продуктов для статистической обработки данных, находящихся в свободном доступе и позволяющих получить необходимый результат, не прибегая к использованию нелегальных версий коммерческих программ, основным недостатком которых является крайне высокая стоимость.

УДК 614

БИОСТАТИСТИЧЕСКИЕ ПРОГРАММЫ СВОБОДНОГО ДОСТУПА

Н.Н. Хромов-Борисов

ФГБУ «Российский научно-исследовательский институт травматологии и ортопедии им. Р.Р. Вредена»
Минздрава России,
ул. Ак. Байкова, д. 8, Санкт-Петербург, Россия, 195427

Реферат

В настоящем сообщении представлены сведения и краткая характеристика современных компьютерных программ для статистического анализа результатов биомедицинских исследований, которые есть в свободном доступе.

Ключевые слова: биостатистика, свободно доступные программы.

Программы общего назначения

Отечественные:

1. **AtteStat** – универсальный пакет программ для профессионального статистического анализа данных и математического моделирования. Программа устанавливается как надстройка к Excel, поэтому перед ее установкой должны быть закрыты все экселевские приложения. На этом же сайте представлено 500-страничное «Руководство пользователя», которое можно использовать как самостоятельное учебное пособие. Поскольку некоторые отечественные научные журналы не принимают ссылки на электронные ресурсы, то кроме ссылки на сайт программы, можно ссылаться

на статьи автора: Гайдышев И.П. Статистика в публикациях // Гений ортопедии. – 2005. – № 4. – С. 155–161. Гайдышев И.П. О перспективах разработки программных продуктов в исследовательском медицинском учреждении // Гений ортопедии. – 2006. – № 2. – С. 95–99. Актуальная версия 13.01.164: http://ilizarov.center/?page_id=82

2. **StatAnt** – Старт новейшего проекта для профессионального статистического анализа данных и математического моделирования. Проект реализует концепцию автономного (stand-alone) приложения, не требующего для своей работы иных компонентов, кроме системной библиотеки Qt: http://ilizarov.center/?page_id=841r

📖 Хромов-Борисов Н.Н. Биостатистические программы свободного доступа. *Травматология и ортопедия России*. 2015; (4):154-159.

✉ Хромов-Борисов Никита Николаевич. Ул. Ак. Байкова, д. 8, Санкт-Петербург, Россия, 195427; e-mail: Nikita.KhromovBorisov@gmail.com

1️⃣ Рукопись поступила: 13.10.2015; принята в печать: 14.11.2015

3. DoctorStat – Статистика для врачей. Программа предназначена для проведения статистических расчетов в медицине. Используются новые технологии для выявления связей между генетикой и болезнью. Актуальная версия 1.9 (19.11.2013): <http://doctorstat.narod.ru/>

4. STADIA – Универсальная статистическая диалоговая система. Доступна бесплатная версия для самообучения и работы. Актуальная версия 8.0: <http://protein.bio.msu.ru/~akula/Podr2-1.htm>

5. StatBase – Программа для статистических расчетов. Реализованы оригинальные методы анализа данных: автоматический подбор наилучшего уравнения нелинейной регрессии, подбор вида распределения, удобное формирование подвыборок. Актуальная версия 2.5.4. (08.10.2012): <http://bioinformatics.ru/Services/statbase.html>

6. ISW – Уникальная система статистического анализа и моделирования одномерных наблюдений. Осуществляет проверку согласия данных с более чем 30 стандартными законами и распределениями, с использованием нескольких критериев и методов оценивания параметров и другие процедуры. Актуальная версия ISW 4.4.1.98 (10.03.2013): <http://ami.nstu.ru/~headrd/ISW.htm>

7. LiTis – Система статистического анализа данных типа времени жизни. Актуальная версия 1.2.0 beta:

<http://amsa.conf.nstu.ru/amsa2015/about-workshop/index.php>

8. SANCT – Структурный анализ таблиц сопряженности. Доступна по запросу от автора: Nikita.KhromovBorisov@gmail.com

9. DiagStat.xls – Статистический контроль качества диагностических тестов с бинарными исходами. Доступна по запросу от автора: Nikita.KhromovBorisov@gmail.com

Зарубежные:

10. R – Программная среда и язык программирования для статистической обработки данных и работы с графиками. С этой программой связаны некоторые указанные ниже программы: **RStudio**, **R2WinBUGS**, **InVivoStat**, **Zelig**. Входит в список лучших открытых программных разработок. Программа интенсивно совершенствуется и обновляется, поэтому стоит регулярно посещать ее сайт. Работе с **R** посвящены уже десятки руководств. Актуальная версия R-3.2.2 (14.08.2015): <http://cran.r-project.org/bin/windows/base/>

11. RStudio – Интерфейс для дружественной работы в **R**. Актуальная версия RStudio-0.98.103 (03.07.2015): <http://www.rstudio.com/>

12. PAST – Универсальный пакет программ для анализа данных. Не требует инсталляции, ею можно пользоваться автономно даже на флэш-накопителях. Там же представлено полезное руководство и некоторые дополнительные обучающие материалы и примеры. При цитировании следует ссылаться на: Hammer, Ø., Harper, D.A.T., Ryan, P.D. PAST: Paleontological statistics software package for education and data analysis // *Palaeontologia Electronica*. – 2001. – Vol. 4. – No. 1. – 9 pp. http://palaeo-electronica.org/2001_1/past/issue1_01.htm

Авторы постоянно совершенствуют программу PAST и расширяют ее возможности, поэтому полезно подписаться на рассылку новостей и обсуждений. Актуальная версия PAST 3.10 (ноябрь 2015): <http://folk.uio.no/ohammer/past/>

13. Instat+ – Статистический пакет общего назначения. Актуальная версия 3.37:

<http://www.reading.ac.uk/ssc/resourcepage/instat.php>

Там же доступно руководство: <http://www.reading.ac.uk/ssc/n/software/instat/tutorial.pdf> Не путать с программой InStat от фирмы GraphPad (<http://www.graphpad.com/scientific-software/instat/>), выпускающей полезную, но коммерческую программу Prism (<http://www.graphpad.com/>).

14. InVivoStat – Для экспериментаторов, работающих с животными. Совмещает сложные и эффективные статистические инструменты из среды **R** с дружественным пользовательским интерфейсом, интуитивно понятным для не статистиков. Особенность: выявляет ошибки в наборах данных до их анализа. Приложение к книге: Bate S.T., Clark R.A. *The Design and Statistical Analysis of Animal Experiments*. – Cambridge: Cambridge University Press, 2014. Актуальная версия 3.3 (14.08.2015): <http://invivostat.co.uk/>

15. MicroSiris – Всесторонний статистический анализ и управление данными с древом решений для пользователей. Актуальная версия 24.8 (20.03.2015): <http://www.microsiris.com/>

16. TANAGRA – Добыча данных, разведочный анализ данных, статистическое обучение, машинное обучение и базы данных. При цитировании следует ссылаться на: Rakotomalala R., TANAGRA: un logiciel gratuit pour l'enseignement et la recherche, in *Actes de EGC'2005, RNTI-E-3, 2005*. – Vol. 2. – P. 697-702, 2005.

<http://eric.univ-lyon2.fr/~ricco/tanagra/en/tanagra.html>

17. Zelig – Статистическое программное обеспечение для всех. Кроме традиционных, включает в себя множество специфических методов: байесовских, основанных на правдоподобии, непараметрических и др. Требуется установка

среды R. Актуальная бета-версия 5.03: <http://zeligproject.org/>

18. Resampling Statistics – рандомизация и бутстреп:

<https://www.uvm.edu/~dhowell/StatPages/Resample3-12-15/ResamplingPackage.zip>

19. LePrep – Вероятности воспроизведения и статистические предсказания.

Актуальная версия LePrep 2.1.0 (01.04.2011): <http://www.univ-rouen.fr/LMRS/Persopage/Lecoutre/PAC.htm>

20. G*Power – Всесторонний анализ мощности статистических критериев. Актуальная версия G*Power 3.1.9.2 (28.03.2014):

<http://www.psych.uni-duesseldorf.de/abteilungen/aap/gpower3>

21. MetaboAnalyst – Набор интерактивных программ для анализа данных метаболомики. Реализованы самые разнообразные методы статистики и машинного обучения. При цитировании следует ссылаться на: Xia J., Sinelnikov I., Han B., Wishart D.S. (2015) MetaboAnalyst 3.0 – making metabolomics more meaningful // Nucl. Acids Res., 2015. – Web Server issue W251–W257. Актуальная версия 3.0 (08.09.2015): <http://www.metaboanalyst.ca/>

22. Mondrian – Система визуализации статистических данных общего назначения. Последняя (бета) версия 1.5b (29.08.2013): <http://www.theusrus.de/Mondrian/>

23. Набор из 29 программ для анализа долей и таблиц сопряженности, сравнения бинарных диагностических тестов, согласия экспертов. Реализованы точные критерии: <http://www.ugr.es/~bioest/software/cmd.php?seccion=props>

Наборы электронных таблиц и надстройки к **Excel**

24. ESCI – Exploratory Software for Confidence Intervals – набор экселевских электронных таблиц для вычисления и моделирования доверительных интервалов. Приложение к книге: Cumming G. Understanding The New Statistics: Effect Sizes, Confidence Intervals, and Meta-Analysis. – New York: Routledge, 2012. – 535 p. <http://www.latrobe.edu.au/psy/research/cognitive-and-developmental-psychology/esci>

25. Confidence intervals for proportions and related quantities – Уникальный набор из 19 экселевских электронных таблиц для построения доверительных интервалов для долей и родственных вычислений. Приложение к книге: Newcomb R.G. Confidence Intervals for Proportions and Related Measures of Effect Size. – CRC Press, 2012. – 468 p. <http://medicine.cf.ac.uk/primary-care-public-health/resources/>

26. Набор из 28 экселевских электронных таблиц. Приложение к онлайн-версии

книги: McDonald J.H. Handbook of Biological Statistics (3rd ed.). 2014; Sparky House Publishing, Baltimore, Maryland.

<http://www.biostathandbook.com/index.html>

На их основе созданы 37 R-скриптов (script): <http://rcompanion.org/rcompanion/> и/или <http://rcompanion.org/documents/RCompanionBioStatistics.pdf>

27. Resampling Stats for Excel – Процедуры бутстреп, перестановок и имитационного моделирования. Надстройка для Excel. Ознакомительная версия предоставляется сроком на один год: <http://www.resample.com/>

28. Reference Value Advisor – Надстройка для Excel для вычисления референсных интервалов. Актуальная версия 2.1 (29.02.2012): <http://www.biostat.envt.fr/spip/spip.php?article63>

29. Real Statistics Using Excel – Реальная статистика, надстройка для Excel. Содержит разнообразные статистические функции и инструменты анализа данных, отсутствующие в Excel: <http://www.real-statistics.com/>

30. Statistician – Дружественный пакет статистических процедур для образования и бизнеса. Надстройка к Excel. Имеется версия Lite (бесплатная) с широким спектром статистических моделей и анализов: <http://www.statisticianaddin.com/>

31. XLStatistics – Набор электронных таблиц для статистического анализа данных в среде Excel. Реализован как пошаговое руководство в виде блок-схем, предназначенных облегчить применение статистики. <http://www.deakin.edu.au/~rodneyc/XLStatistics/>

Интерактивные веб-страницы

32. Веб-страницы для статистических расчетов – одна из наиболее полных и регулярно обновляемых подборок интерактивных и свободно доступных мультиплатформенных статистических программ (15.09.2015): <http://statpages.org/>

33. Statistics 5102 (Geyer, Spring 2012) Examples – 15 страниц интерактивных программ для известных и редких статистических процедур на базе языков R и RWeb: <http://www.stat.umn.edu/geyer/s12/5102/amp/bayes.html>

34. VassarStat – Большой набор программ для разнообразных статистических вычислений: <http://vassarstats.net/>

35. Free Statistics Software (Calculator) – Коллекция калькуляторов, включающая широкий спектр процедур подробного статистического описания и анализа данных с текстами R-скриптов: <http://www.wessa.net/>

36. SISA – Simple Interactive Statistical Analysis – 50 разнообразных интерактивных

статистических процедур, а также экселевские электронные таблицы для 22 распределений: <http://www.quantitativeskills.com/sisa/>

37. QuickCalcs – Набор из 24 наиболее часто применяемых статистических процедур: <http://www.graphpad.com/quickcalcs/>

38. Statistics Calculators – Набор из 104 статистических калькуляторов (29 категорий). 35 миллионов пользователей. Актуальная версия 3.0: <http://www.danielsoper.com/statcalc3/>

39. StatsToDo – Свыше 200 страниц разнообразны статистических программ для клинических исследований и контроля качества с пояснениями:

<http://www.statstodo.com/index.php>

40. A Collection of JavaScript E-labs Learning Objects – Обширная коллекция более 60 статистических процедур от Dr. Hossein Arsham

<http://home.ubalt.edu/ntsbarsh/zero/scientificCal.htm>

41. VassarStat – Большой набор программ для разнообразных статистических вычислений: <http://vassarstats.net/>

42. Statistical Test Calculators – 17 калькуляторов: <http://www.socscistatistics.com/tests/Default.aspx>

43. BoxPlotR – уникальный и полезный веб-инструмент для построения разнообразных и подробных диаграмм типа «короб с усами». <http://boxplot.tyerslab.com/>

Статистико-эпидемиологические программы

44. OpenEpi – Эпидемиологическая статистика. Реализованы разнообразные статистические процедуры общего и специального назначения. Актуальная версия 3.03a (04.05.2015). Работает как online: http://www.openepi.com/Menu/OE_Menu.htm так и автономно: <http://www.openepi.com/Downloads/Downloads.htm>

45. EpiData Software – Осуществляет основные статистические процедуры, графики и разнообразное управление данными. Актуальная версия 2.0 (01.02.2014): <http://www.epidata.dk/>

46. WINPEPI – Набор разнообразных программ общего и специального назначения для работы и обучения эпидемиологов. Актуальная версия 11.50 (08.08.2015): <http://www.brixtonhealth.com/pepi4windows.html>

47. EpiInfo™ 7 – Коллекция инструментов для анализа и визуализации эпидемиологических баз данных. Более миллиона пользователей. Русифицирована версия 7.1.2. Актуальная версия 7.1.5 (19.03.2015): <http://wwwn.cdc.gov/epiinfo/>

Обучающие программы

48. WinStats – Обучающая программа по теории вероятностей и статистике, русифицирована. Актуальная версия от 07.08.2012: <http://math.exeter.edu/rparris/winstats.html>

49. SUStats – *Пакет* программ (апплетов) для понимания статистики:

<http://www.jsc.nildram.co.uk/examples/sustats/SUStats.html>

50. SOCR – Мощный ресурс для обучения теории вероятностей, статистике и статистическим вычислениям, как online: <http://www.socr.ucla.edu/> так и автономно: http://www.socr.ucla.edu/htmls/SOCR_Download.html

51. Keisan Online Calculator/Probability Function – Набор высокоточных калькуляторов для вычисления разнообразных характеристик 21 вероятностного распределения с графикой:

<http://keisan.casio.com/menu/system/000000000540>

52. Statistics101 – Новый подход к обучению студентов социальных наук методам ресемплинга и бутстрепа.

Актуальная версия 3.8 (22.08.2015): <http://www.statistics101.net/>

Бейзовская статистика

53. WinBUGS – Бейзовский анализ сложных статистических моделей с использованием методов Монте-Карло. Последняя версия WinBUGS 1.4.3:

<http://www.mrc-bsu.cam.ac.uk/software/bugs/the-bugs-project-winbugs/>

После установки требуется установить заплатку (patch) и ключ. Некоторые программы работают посредством вызова **WinBUGS** из среды **R**, для чего требуется установить пакет **R2WinBUGS**:

<http://cran.r-project.org/web/packages/R2WinBUGS/index.html> и удобно использовать **RStudio**. При цитировании следует ссылаться на: Lunn D. J., Thomas A., Best N., Spiegelhalter D. WinBUGS – a Bayesian modelling framework: concepts, structure, and extensibility // *Statistics and Computing*. – 2000. – Vol. 10. – N 4. – P. 325–337.

http://www.panic-lab.rutgers.edu/users/kkonst/Papers_03_10_05/content.asp.pdf Lunn D., Spiegelhalter D., Thomas A., Best N. The BUGS project: Evolution, critique and future directions (with discussion), *Statistics in Medicine*. – 2009. – Vol. 28. – No 25. – P. 3049–3082.

54. OpenBUGS – Дальнейшее развитие WinBUGS. Актуальная версия 3.2.3 (20.07.2015): <http://www.openbugs.net/w/FrontPage>

55. LePAC – Уникальная система для статистического анализа экспериментальных данных

бейзовскими методами. Включает в себя несколько программ. Актуальная версия LePAC 2.0.41 (08.11.2012):

<http://www.univ-rouen.fr/LMRS/Persopage/Lecoutre/PAC.htm>

56. Kaplan's statistics pages – 23 обновляемые страницы для интерактивных статистических выводов с использованием языка JavaScript. Каждую страницу можно скопировать на свой компьютер и использовать автономно (11.01.2015): <http://www.printmacroj.com/statistics.htm>

57. Bayes Factor Calculators – 5 интерактивных калькуляторов Бейзова фактора. <http://pcl.missouri.edu/bayesfactor>

Реализован также расширенный пакет для среды **R**:

<https://cran.r-project.org/web/packages/BayesFactor/index.html>

58. JASP – Недавно стартовавший многообещающий проект – альтернатива коммерческому пакету SPSS «без жирка» и «лакомая» альтернатива системе R. Действительно, иначе как изящной, ее не назовешь. Отличается четкой графикой, отчетами, пригодными непосредственно для публикации, и ориентацией на бейзовские методы статистики.

Актуальная версия 0.7.1.12 (2014) <https://jasp-stats.org/>

Следует иметь в виду, что многие зарубежные программы не воспринимают запятые как десятичные разделители и не функционируют с ними. В таком случае надо в Панели управления сменить установку языка с русского на английский.

Конечно, нелегко ориентироваться в таком многообразии статистических программ. Задачу выбора привлекательных для пользователя программ облегчают их описания на соответствующих сайтах. Для многих программ представлены довольно подробные и вполне понятные руководства для пользователей. В Википедии представлены таблицы сравнения многих из перечисленных программ, что, что **поможет потребителю выбрать удовлетворяющую его программу**: https://en.wikipedia.org/wiki/Comparison_of_statistical_packages

Автор этих строк чаще использует следующие программы:

AtteStat – для очень многих задач.

PAST – построение гистограмм и нормальных вероятностных графиков, которые легко редактировать. Проверка согласия данных с нормальным распределением с помощью 4-х надежных критериев. Вычисление основных описательных статистик с доверительными интервалами посредством универсального метода

бутстрепа, проверка гипотез посредством метода Монте-Карло. Причем эти вычисления можно проводить сразу для нескольких выборок или для нескольких критериев.

Instat+ – построение диаграмм «короб с усами», уникальность которых состоит в их большей информативности вследствие того, что на этих диаграммах указываются все экспериментальные точки. Построение графиков одновременно не только для доверительных интервалов, но и для предсказательных интервалов.

ESCI_JSMS.xls – результаты *t*-критерия для двух независимых выборок сопровождаются вычислением для разности средних ДИ и графиком, визуализирующим эти результаты, а также стандартизированного размера эффекта по Коуэну.

LePrep – уникальная возможность вычислять предсказательные вероятности и предсказательные интервалы.

G*Power – анализ мощности

Для анализа счетных (дискретных) данных полезно использовать:

Clopper-Pearson.xls – вычисление точных ДИ для долей по Клопперу-Пирсону.

LePAC – бейзовский анализ счетных данных.

Две программы Ньюкомба: **mover-d.xls** – универсальный метод построения ДИ для разностей параметров и **mover-r.xls** – для отношений параметров.

JASP – бейзовская статистика – вычисление бейзовских факторов для корреляций, бейзовского аналога *t*-критерия, дисперсионного анализа и анализа таблиц сопряженности.

В заключение стоит отметить несколько не слишком дорогих, но хорошо себя зарекомендовавших коммерческих программ, которые отвечают основным запросам биомедицинских исследований:

MedCalc – Включает ROC-анализ, сравнение методов и инструменты статистического контроля качества с хорошей графикой. Доступна русифицированная версия. Актуальная версия 18.5 (14.08.2016): <https://www.medcalc.org/>

GraphPad Prism – удачное сочетание биостатистики, анализа кривых (нелинейная регрессия) и построения научных графиков. Две дополнительные программы: **InStat** – уникальная программа для новичков в статистике. Помогает пользователю выбрать подходящий тест, задавая вопросы о данных, представляет результаты в простых выражениях с минимум статистического жаргона. **StatMate** – анализ статистической мощности и планирования объемов выборок. Актуальная версия 6 (2015): <http://www.graphpad.com/>

StatsDirect – статистическое программное обеспечение исследований в биомедицине, здравоохранении и науках о здоровье. Удобно в использовании с подробной документацией. Представлены редкие (экзотические), но полезные статистические процедуры. Актуальная

версия 3.0.155 (03.09.2015): <http://www.statsdirect.com/>

XLStat – наиболее полная и широко используемая надстройка к Excel. Состоит из разнообразных модулей, которые пользователь может компоновать по своим интересам. <http://www.xlstat.com/en>

СВЕДЕНИЯ ОБ АВТОРЕ

Хромов-Борисов Никита Николаевич – канд. биол. наук, научный редактор журнала «Травматология и ортопедия России», ФГБУ «Российский научно-исследовательский институт травматологии и ортопедии им. Р.Р. Вредена» Минздрава России

FREE BIOSTATISTICAL SOFTWARE

N.N. Khromov-Borisov

*Vreden Russian Research Institute of Traumatology and Orthopedics,
ul. Akademika Baykova, 8, St. Petersburg, Russia, 195427*

Abstract

This report provides information and brief description of current freely available computer programs for statistical analysis of the results in biomedical research.

Key words: biostatistics, free software.

INFORMATION ABOUT AUTHORS:

Khromov-Borisov Nikita N. – scientific editor of journal «Traumatology and Orthopedics of Russia», Vreden Russian Research Institute of Traumatology and Orthopedics.

 **Cite as:** Khromov-Borisov NN. [Free biostatistical software]. *Traumatologiya i ortopediya Rossii*. 2015; (4):154-159. [in Russian]

 *Khromov-Borisov N.N.* Ul. Ak. Baykova, 8, St. Petersburg, Russia, 195427; e-mail: Nikita.KhromovBorisov@gmail.com

 Received: 13.10.2015; Accepted for publication: 14.11.2015