

УДК 004.9:61:311

МИНИ-ПРОГРАММА ДЛЯ РАСЧЕТА ДОВЕРИТЕЛЬНОГО ИНТЕРВАЛА ЧАСТОТЫ И ДОЛИ В МЕДИЦИНСКИХ ИССЛЕДОВАНИЯХ С ПОМОЩЬЮ АНДРОИД-ПЛАНШЕТА

¹Перцов С.С., ¹Арутюнов С.Д., ¹Муслов С.А., ²Гржибовский А.М.

¹ГБОУ ВПО «Московский государственный медико-стоматологический Университет (МГМСУ) им. А.И. Евдокимова» Минздрава РФ, Москва, e-mail: muslov@mail.ru;

²Национальный институт общественного здоровья, Осло

Представлена минипрограмма для расчета доверительного интервала частоты и доли в медицинских исследованиях с помощью планшета или смартфона, работающих под управлением операционной системы Андроид. Описаны основные понятия и статистические термины, которые используются для анализа доверительного интервала частоты и доли. Анализируется 95% доверительного интервала. Изложенный материал дает общие сведения о способах расчета доверительных интервалов для частот и долей. Дан краткий обзор наиболее популярных методов расчета доверительного интервала частоты и доли, их основные преимущества и недостатки. Рассмотрен пример применения различных методов оценки доверительного интервала доли для частного случая. Описан алгоритм действия программы, отмечается, что для работы с ней не требуется специальных навыков. Программа предназначена для студентов, аспирантов и ординаторов медицинских вузов, врачей-исследователей, а также слушателей курсов переподготовки медицинских работников.

Ключевые слова: медицинская статистика, доля, доверительный интервал

MINI-PROGRAM FOR CALCULATION OF CONFIDENCE INTERVALS FOR THE PROPORTION IN MEDICAL RESEARCH WITH ANDROID TABLET

¹Pertsov S.S., ¹Arutjunov S.D., ¹Muslov S.A., ¹Grjibovski A.M.

¹A.I. Evdokimov Moscow State Medical Stomatological University (MSMSU), Moscow, e-mail: muslov@mail.ru;

²National Institute of Public Health, Oslo

A mini-program is presented for calculating the confidence interval of frequency and proportion in medical research using a tablet running an Android operating system. The basic concepts and statistical terms that are used to analyze the confidence interval of frequency and proportion are described. A 95% confidence interval is analyzed. A brief overview of the most popular methods for calculating the confidence interval of a share is given, their main advantages and disadvantages. An example of the application of various methods for estimating the confidence interval of a fraction for a particular case is considered. The algorithm of the program's operation is described, it is noted that special skills are not required for working with it. The program is intended for students, post-graduate students and residents of medical schools, doctors-researchers, as well as listeners of retraining courses for medical workers.

Keywords: medical statistics, proportion, confidence interval

В [2] отмечалось, что интервальная выборочная оценка предпочтительнее точечной для описания частоты встречаемости изучаемой характеристики в генеральной совокупности. Но поскольку на практике исследования проводятся с использованием выборочных данных, проекция результатов на генеральную совокупность должна содержать определенный элемент неточности выборочной оценки. Доверительный интервал (ДИ) представляет собой меру точности оцениваемого параметра. Результаты исследований должны указываться вместе с их доверительным интервалом. В биомедицинских исследованиях чаще всего используются 95 доверительные интервалы. Данный доверительный интервал представляет собой область, в которую попадает истинное значение доли в 95 случаях. Другими словами, можно с 95% надежностью сказать, что истинное значение частоты встречаемости

признака в генеральной совокупности будет находиться в пределах 95 доверительного интервала.

В ряде пособий по статистике сообщается [1, 4, 7], что ошибку частоты можно рассчитывать с помощью формулы

$$s = \sqrt{\frac{p(1-p)}{N}} \sim \sqrt{\frac{1}{N}},$$

где p – относительная частота встречаемости признака в выборке, N – количество наблюдений в выборке. Вследствие этого ДИ доли, например, доли умерших при резекции желудка, когда умирает 3 из 4 больных отличается от ДИ для случая, например, когда от данной операции умирает 30 из 40 пациентов. Ряд авторов считают целесообразнее представлять 95 доверительный

интервал для частоты встречаемости признака в виде интервала от

$$p - 1,96\sqrt{\frac{p(1-p)}{N}}$$

до

$$p + 1,96\sqrt{\frac{p(1-p)}{N}}.$$

Также в [4] рекомендуется при малых выборках заменять значение 1,96 на значение t для $N-1$ степеней свободы. Значение t находится по таблицам для t -распределения, имеющимся во многих пособиях по статистике. Именно этот способ подсчета ДИ доли положен в основу работы, представленной в данном сообщении программы.

Кроме этого метода расчета доверительного интервала доли, существует целый ряд других методов, достаточно широко описанных в специальной литературе. Каждый из них имеет свои положительные и отрицательные стороны. Они достаточно подробно рассмотрены в специальном сообщении А.М. Гржибовского [2], а также в [5, 6, 8, 9, 10] и здесь не обсуждаются.

В качестве практического примера приведем результаты подсчета доверительного интервала доли некоторыми из этих методов (табл. 1, рис. 1). Рассмотрим частный случай $N=100$, $M=90$ (то есть $p=M/N=0,9$) и 95 % ДИ. Здесь M – абсолютная частота встречаемости исследуемого признака, N – общее число испытаний.

Из таблицы и графика видно, что наиболее узкий доверительный интервал (11,8 %) дает метод нормальной аппроксимации (Вальда), а наиболее широкий (38,3 %) – метод расчета, использующий распределение хи-квадрат. Последний также дает результат для верхней границы интервала, больший, чем 100 %, что, конечно, абсурдно с точки зрения здравого смысла. В последнем столбце таблицы представлены функции Excel, которые используются при определении границ интервала (все рассмотренные методы могут быть реализованы с помощью стандартных функций в редакторе таблиц Microsoft Windows Excel). Отметим также, что в литературе нет единства в отношении названий приведенных методов.

Расчетная часть программы. Как уже отмечалось, в основу программы Proportion Confidence Interval положена простая формула для расчета границ доверительного интервала

$$\left(p - t\sqrt{\frac{p(1-p)}{N}}; p + t\sqrt{\frac{p(1-p)}{N}}\right),$$

где величина t протабулирована и вшита в память настоящей программы.

Алгоритм работы программы. После запуска программа предлагает к заполнению два поля для параметров N и M . Пользователь заполняет эти поля с помощью всплывающей миниклавиатуры (рис. 2). Далее производится счет программы и открывается окно, в котором выдаются результаты

Доверительный интервал частоты (доли), рассчитанный различными методами

Метод	Нижний предел ДИ	Верхний предел ДИ	Ширина ДИ	Функция Excel
Метод Клоппера-Пирсона	82,38 %	95,10 %	12,7 %	ФРАСПОБР
Нормальная аппроксимация (Вальда)	84,12 %	95,88 %	11,8 %	НОРМСТОБР
Метод Вилсона	82,56 %	94,48 %	11,9 %	НОРМСТОБР
Пуассона хи-квадрат	72,37 %	110,63 %	38,3 %	ХИ2ОБР
Откорректированный метод Вальда	82,39 %	94,65 %	12,3 %	НОРМСТОБР

подсчета: частота (доля), нижняя и верхняя граница доверительного интервала доли и его ширина (рис. 3). Если границы ДИ в результате расчета принимают отрицательные значения или превышают 1, то программа производит корректировку, присваивая им значения 0 или 1, соответственно. Предусмотрено представление результатов расчета в процентах (рис. 4).

Программа выполнена на базе операционной системы Android и может быть реализована с помощью планшетного компьютера или смартфона, работающих под управлением данной системы. Ранее мы уже сообщали об использовании этих популярных вычислительных средств при решении других статистических задач в медицинских исследованиях [3].

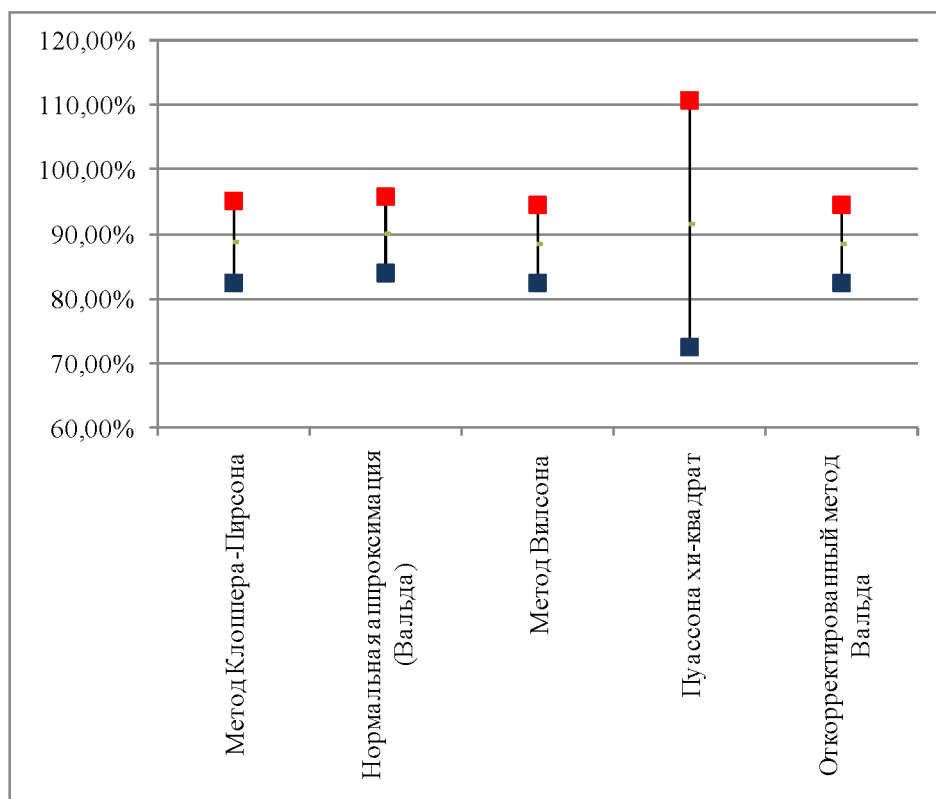


Рис. 1. Доверительный интервал частоты (доли), рассчитанный различными методами

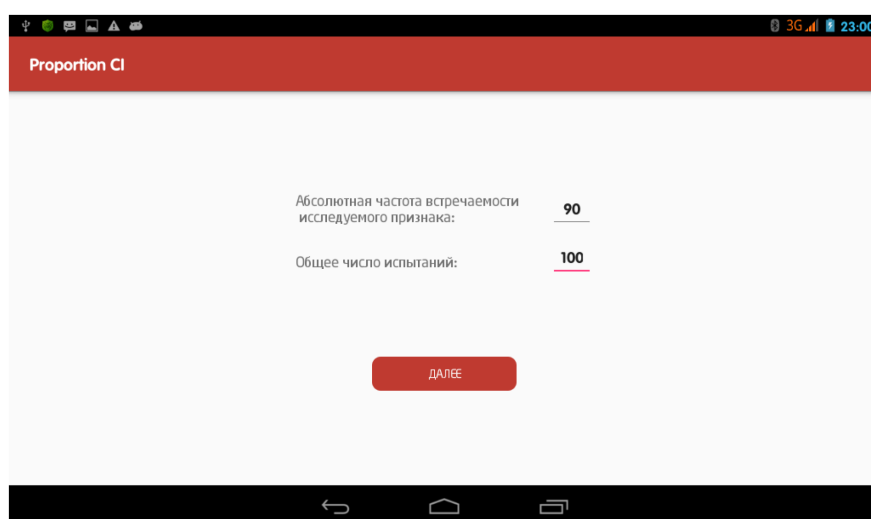


Рис. 2. Окно программы Proportion Confidence Interval для ввода данных.

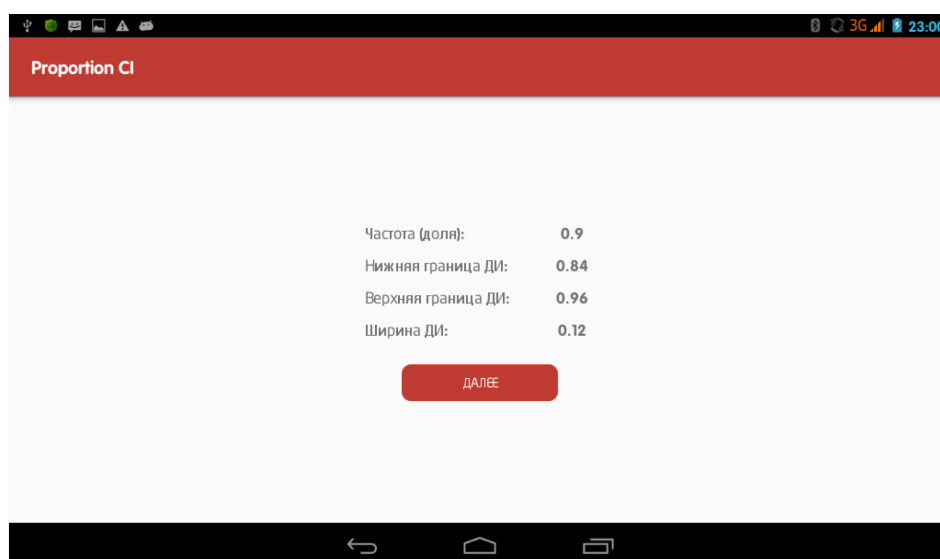


Рис. 3. Окно программы *Proportion Confidence Interval* с результатами расчета

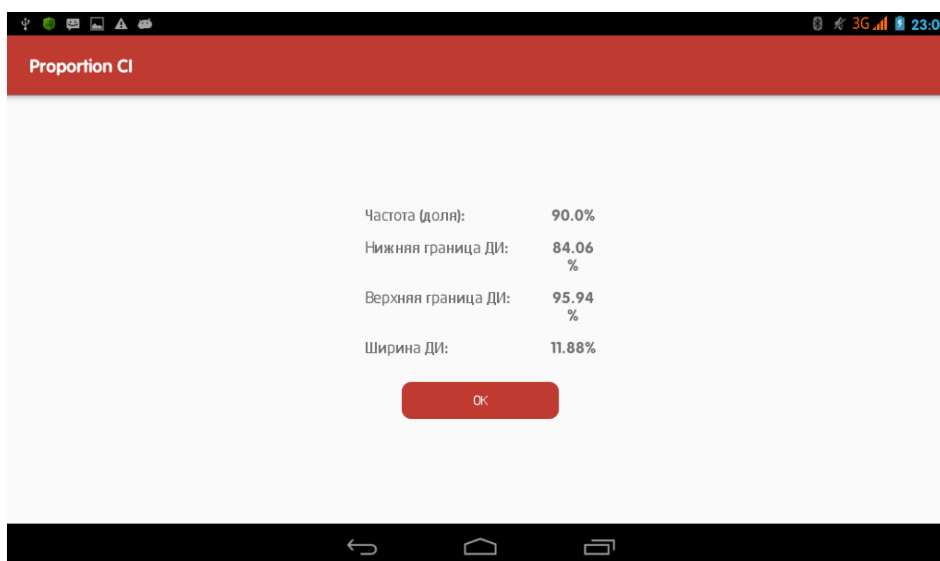


Рис. 4. Окно программы *Proportion Confidence Interval* с результатами расчета в процентах

Как следует из рис. 2 и 3, результаты расчета ДИ представленной программой наиболее близки к полученным в программе Excel методом нормальной аппроксимации (Вальда) результатам: 84,06 и 84,12 – нижняя граница ДИ, 95,94 и 95,88% – верхняя граница ДИ (табл. 1), что позволяет рекомендовать предложенную программу для использования.

Список литературы

1. Гланц С. Медико-биологическая статистика / С. Гланц. – М.: Практика, 1998.
2. Гржибовский А.М. Доверительные интервалы для частот и долей // Экология человека. – 2008. – № 5. – С. 57–60.
3. Перцов С.С., Стюрева Г.М., Муслов С.А., Корнев А.А., Зайцева Н.В., Силицын А.А., Подымов В.Ф., Сидорова Л.Н. Программа анализа медицинских данных (представленных таблицами сопряженности 2x2) для ANDROID планшета // Сборник научных трудов по итогам международной научно-практической конференции (10 марта 2016 г.)

«Актуальные вопросы современных математических и естественных наук». – Екатеринбург, 2016. – С. 12–13.

4. Юнкеров В.И. Медико-статистическая обработка данных медицинских исследований / В.И. Юнкеров, С.Г. Григорьев. – СПб.: ВмедА, 2002. – 266 с.

5. Brown L.D. Interval estimation for a binomial proportion / L.D. Brown, T.T. Cai, A. Dasgupta // *Statistical science*. – 2001. – N 2. – P. 101–133.

6. Garcia-Perez M.A. On the confidence interval for the binomial parameter / M.A. Garcia-Perez // *Quality and quantity*. – 2005. – N 39. – P. 467–481.

7. Motulsky H. *Intuitive biostatistics* // H. Motulsky. – Oxford : Oxford University Press, 1995. – 386 p.

8. Newcombe R.G. Two-Sided Confidence Intervals for the Single Proportion: Comparison of Seven Methods / R.G. Newcombe // *Statistics in Medicine*. – 1998. – N. 17. – P. 857–872.

9. Sauro J. Estimating completion rates from small samples using binomial confidence intervals: comparisons and recommendations / J. Sauro, J.R. Lewis // *Proceedings of the human factors and ergonomics society annual meeting*. – Orlando, FL, 2005.

10. Wald A. Confidence limits for continuous distribution functions // A. Wald, J. Wolfowitz // *Annals of Mathematical Statistics*. – 1939. – № 10. – P. 105–118.