

УДК 31:61

ДОВЕРИТЕЛЬНЫЕ ИНТЕРВАЛЫ ДЛЯ ЧАСТОТ И ДОЛЕЙ

© 2008 г. А. М. Гржибовский

Национальный институт общественного здоровья, г. Осло, Норвегия

В одной из предыдущих публикаций кратко упоминалось описание качественных данных и сообщалось, что их интервальная оценка предпочтительнее точечной для описания частоты встречаемости изучаемой характеристики в генеральной совокупности [4]. Действительно, поскольку исследования проводятся с использованием выборочных данных, проекция результатов на генеральную совокупность должна содержать элемент неточности выборочной оценки. Доверительный интервал представляет собой меру точности оцениваемого параметра. Интересно, что в некоторых книгах по основам статистики для медиков тема доверительных интервалов для частот полностью игнорируется [1, 5]. В данной статье мы рассмотрим несколько способов расчета доверительных интервалов для частот, подразумевая такие характеристики выборки, как бесповторность и репрезентативность, а также независимость наблюдений друг от друга. Под частотой в данной статье понимается не абсолютное число, показывающее, сколько раз встречается в совокупности то или иное значение, а относительная величина, определяющая долю участников исследования, у которых встречается изучаемый признак.

В биомедицинских исследованиях чаще всего используются 95 % доверительные интервалы. Данный интервал представляет собой область, в которую попадает истинное значение доли в 95 % случаев. Другими словами, можно с 95 % надежностью сказать, что истинное значение частоты встречаемости признака в генеральной совокупности будет находиться в пределах 95 % доверительного интервала.

В большинстве пособий по статистике для исследователей от медицины сообщается [3, 6, 7–10, 16], что ошибка частоты рассчитывается с помощью формулы

$$s_p = \sqrt{\frac{p(1-p)}{N}}$$

где p — частота встречаемости признака в выборке (величина от 0 до 1). В большинстве отечественных научных статей указывается значение частоты встречаемости признака в выборке (p), а также ее ошибка (s) в виде $p \pm s$. Целесообразнее, однако, представлять 95 % доверительный интервал для частоты встречаемости признака в генеральной совокупности, который будет включать значения от

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

В некоторых пособиях [10] рекомендуется при малых выборках заменять значение 1,96 на значение t для $N - 1$ степеней свободы, где N — количество наблюдений в выборке. Значение t находится по

В статье описывается и обсуждается расчет доверительных интервалов для частот и долей по методам Вальда, Уилсона, Клоппера – Пирсона, с помощью углового преобразования и по методу Вальда с коррекцией по Агрести – Коуллу. Изложенный материал дает общие сведения о способах расчета доверительных интервалов для частот и долей и призван вызвать интерес читателей журнала не только к использованию доверительных интервалов при представлении результатов собственных исследований, но и к прочтению специализированной литературы перед началом работы над будущими публикациями.

Ключевые слова: доверительный интервал, частота, доля.

таблицам для t -распределения, имеющимся практически во всех пособиях по статистике. Использование распределения t для метода Вальда не дает видимых преимуществ по сравнению с другими методами, рассмотренными ниже [15], и потому некоторыми авторами не приветствуется [16].

Представленный выше метод расчета доверительных интервалов для частот или долей носит имя Вальда в честь Авраама Вальда (Abraham Wald, 1902–1950), поскольку широкое применение его началось после публикации Вальда и Вольфовица в 1939 году [19]. Однако сам метод был предложен Пьером Симоном Лапласом (1749–1827) еще в 1812 году.

Метод Вальда очень популярен, однако его применение связано с существенными проблемами. Метод не рекомендуется при малых объемах выборок, а также в случаях, когда частота встречаемости признака стремится к 0 или 1 (0 % или 100 %), и просто невозможен для частот 0 и 1. Кроме того, аппроксимация нормального распределения, которая используется при расчете ошибки, «не работает» в случаях, когда $n \cdot p < 5$ или $n \cdot (1 - p) < 5$ [16]. Более консервативные статистики считают, что $n \cdot p$ и $n \cdot (1 - p)$ должны быть не менее 10 [8]. Более детальное рассмотрение метода Вальда показало, что полученные с его помощью доверительные интервалы в большинстве случаев слишком узки, то есть их применение ошибочно создает слишком оптимистичную картину, особенно при удалении частоты встречаемости признака от 0,5, или 50 % [13, 15, 17, 18]. К тому же при приближении частоты к 0 или 1 доверительный интервал может принимать отрицательные значения или превышать 1, что выглядит абсурдно для частот. Многие авторы совершенно справедливо не рекомендуют применять данный метод не только в уже упомянутых случаях, но и тогда, когда частота встречаемости признака менее 25 % или более 75 % [10]. Таким образом, несмотря на простоту расчетов, метод Вальда может применяться лишь в очень ограниченном числе случаев. Зарубежные исследователи более категоричны в своих выводах и однозначно рекомендуют не применять этот метод для небольших выборок [11, 13, 15], а ведь именно с такими выборками часто приходится иметь дело исследователям-медикам.

При частотах, не превышающих 25 % или превышающих 75 %, отечественные авторы [7, 9, 10] рекомендуют рассчитывать доверительный интервал с помощью \arcsin -преобразования (оно также часто упоминается как угловое преобразование Фишера), при котором сначала рассчитывается вспомогательная переменная (φ) по формуле:

$$\varphi = 2 \arcsin \sqrt{p},$$

где p — выборочное значение частоты встречаемости признака. Затем рассчитывается стандартная ошибка вспомогательной переменной по формуле:

$$s_{\varphi} = \frac{1}{\sqrt{N}},$$

Поскольку новая переменная имеет нормальное распределение, нижняя и верхняя границы 95 % доверительного интервала для переменной φ будут равны $\varphi - 1,96 S_{\varphi}$ и $\varphi + 1,96 S_{\varphi}$ соответственно, а 95 % доверительный интервал для частоты встречаемости признака в генеральной совокупности будет

$$\text{от } \sin^2 \frac{\varphi - 1,96s_{\varphi}}{2} \quad \text{до } \sin^2 \frac{\varphi + 1,96s_{\varphi}}{2}.$$

Вместо 1,96 для малых выборок рекомендуется подставлять значение t для $N - 1$ степеней свободы [10]. Данный метод не дает отрицательных значений и позволяет более точно оценить доверительные интервалы для частот, чем метод Вальда. Кроме того, он описан во многих отечественных справочниках по медицинской статистике [7, 9, 10], что, правда, не привело к его широкому использованию в медицинских исследованиях. Расчет доверительных интервалов с использованием углового преобразования не рекомендуется при частотах, приближающихся к 0 или 1 [15].

На этом описание способов оценки доверительных интервалов в большинстве книг по основам статистики для исследователей-медиков обычно заканчивается, причем эта проблема характерна не только для отечественной, но и для зарубежной литературы. Оба метода основаны на центральной предельной теореме, которая подразумевает наличие большой выборки.

Принимая во внимание недостатки оценки доверительных интервалов с помощью вышеупомянутых методов, Клоппер (Clopper) и Пирсон (Pearson) предложили в 1934 году способ расчета так называемого точного доверительного интервала с учетом биномиального распределения изучаемого признака [14]. Данный метод доступен во многих онлайн-калькуляторах, однако доверительные интервалы, полученные таким образом, в большинстве случаев слишком широки. В то же время этот метод рекомендуется применять в тех случаях, когда необходима консервативная оценка. Степень консервативности метода увеличивается по мере уменьшения объема выборки, особенно при $N < 15$ [11]. А. Н. Герасимов [2] описывает применение функции биномиального распределения для анализа качественных данных с использованием MS Excel, в том числе и для определения доверительных интервалов, однако расчет последних для частот в электронных таблицах не «затабулирован» в удобном для пользователя виде, а потому, вероятно, и не используется большинством исследователей.

По мнению многих статистиков [11, 13, 17, 18], наиболее оптимальную оценку доверительных интервалов для частот осуществляет метод Уилсона (Wilson), предложенный еще в 1927 году [20], но практически не используемый в отечественных биомедицинских исследованиях. Данный метод не только позволяет оценить доверительные интервалы для очень малых и очень больших частот, но и применим для малого числа наблюдений. В общем виде доверительный интервал по формуле Уилсона имеет вид от

$$\text{до } \frac{p + \frac{z^2_{1-\alpha/2}}{2N} - z_{1-\alpha/2} \sqrt{\frac{p(1-p)}{N} + \frac{z^2_{1-\alpha/2}}{4N^2}}}{1 + \frac{z^2_{1-\alpha/2}}{N}}$$

$$\frac{p + \frac{z^2_{1-\alpha/2}}{2N} + z_{1-\alpha/2} \sqrt{\frac{p(1-p)}{N} + \frac{z^2_{1-\alpha/2}}{4N^2}}}{1 + \frac{z^2_{1-\alpha/2}}{N}}$$

где $Z_{1-\alpha/2}$ принимает значение 1,96 при расчете 95 % доверительного интервала, N – количество наблюдений, а p – частота встречаемости признака в выборке. Данный метод доступен в онлайн-калькуляторах, поэтому его применение не является проблематичным. В. А. Медик и М. С. Токмачев [8] не рекомендуют использовать этот метод при $n \cdot p < 4$ или $n \cdot (1 - p) < 4$ по причине слишком грубого приближения распределения p к нормальному в такой ситуации, однако зарубежные статистики считают метод Уилсона применимым и для малых выборок [13, 15].

Считается, что помимо метода Уилсона метод Вальда с коррекцией по Агрести – Коуллу также дает оптимальную оценку доверительного интервала для частот [11, 13, 15]. Коррекция по Агрести – Коуллу представляет собой замену в формуле Вальда частоты встречаемости признака в выборке (p) на p' , при расчете которой к числителю добавляется 2, а к знаменателю добавляется 4, то есть $p' = (X + 2) / (N + 4)$, где X – количество участников исследования, у которых имеется изучаемый признак, а N – объем выборки [11]. Такая модификация приводит к результатам, очень похожим на результаты применения формулы Уилсона, за исключением случаев, когда частота события приближается к 0 % или 100 %, а выборка мала [13]. Кроме вышеупомянутых способов расчета доверительных интервалов для частот были предложены поправки на непрерывность как для метода Вальда, так и для метода Уилсона для малых выборок, однако исследования показали, что их применение нецелесообразно [15].

Рассмотрим применение вышеописанных способов расчета доверительных интервалов на двух примерах. В первом случае мы изучаем большую выборку, состоящую из 1 000 случайно отобранных участников исследования, из которых 450 имеют изучаемый признак (это может быть фактор риска, исход или любой другой признак), что составляет частоту 0,45, или 45 %. Во втором случае исследование проводится с использованием малой выборки, допустим, всего 20 человек, причем изучаемый признак имеется всего у 1 участника исследования (5 %). Доверительные интервалы по методу Вальда, по методу Вальда с коррекцией по Агрести – Коуллу, по методу Уилсона рассчитывались с помощью онлайн-калькулятора, разработанного Jeff Sauro (<http://www.measuringusability.com/wald.htm>). Доверительные интервалы по методу Уилсона с поправ-

кой на непрерывность рассчитывались с помощью калькулятора, предложенного порталом Wassar Stats: Web Site for Statistical Computation (<http://faculty.vassar.edu/lowry/prop1.html>). Расчеты с помощью углового преобразования Фишера производились «вручную» с использованием критического значения t для 19 и 999 степеней свободы соответственно. Результаты расчетов представлены в таблице для обоих примеров.

Доверительные интервалы, рассчитанные шестью разными способами для двух примеров, описанных в тексте

Способ расчета доверительного интервала	95% ДИ для X=1, N=20, P=0,0500, или 5%	95% ДИ для X=450, N=1000, P=0,4500, или 45%
Вальда	-0,0455–0,2541	0,4192–0,4810
Вальда с коррекцией по Агрести – Коуллу	<,0001–0,2541	0,4194–0,4810
Уилсона	0,0089–0,2361	0,4194–0,4810
Уилсона с коррекцией на непрерывность	0,0026–0,2694	0,4189–0,4815
«Точный метод»		
Клоппера – Пирсона	0,0013–0,2487	0,4189–0,4814
Угловое преобразование	<0,0001–0,1967	0,4193–0,4809

Как видно из таблицы, для первого примера доверительный интервал, рассчитанный по «общепринятому» методу Вальда заходит в отрицательную область, чего для частот быть не может. К сожалению, подобные казусы нередки в отечественной литературе. Традиционный способ представления данных в виде частоты и ее ошибки частично маскирует эту проблему. Например, если частота встречаемости признака (в процентах) представлена как $2,1 \pm 1,4$, то это не настолько «режет глаз», как $2,1 \%$ (95 % ДИ: -0,7; 4,9), хоть и обозначает то же самое. Метод Вальда с коррекцией по Агрести – Коуллу и расчет с помощью углового преобразования дают нижнюю границу, стремящуюся к нулю. Метод Уилсона с поправкой на непрерывность и «точный метод» дают более широкие доверительные интервалы, чем метод Уилсона. Для второго примера все методы дают приблизительно одинаковые доверительные интервалы (различия появляются только в тысячных), что неудивительно, так как частота встречаемости события в этом примере не сильно отличается от 50 %, а объем выборки достаточно велик.

Для читателей, заинтересовавшихся данной проблемой, можно порекомендовать работы R. G. Newcombe [17] и Brown, Cai и Dasgupta [13], в которых приводятся плюсы и минусы применения 7 и 10 различных методов расчета доверительных интервалов соответственно [13]. Из отечественных пособий рекомендуется книга В. А. Медика и М. С. Токмачева [8], в которой помимо подробного описания теории представлены методы Вальда, Уилсона, а также способ расчета доверительных интервалов с

учетом биномиального распределения частот. Кроме бесплатных онлайн-калькуляторов (<http://www.measuringusability.com/wald.htm> и <http://faculty.vassar.edu/lowry/prop1.html>) доверительные интервалы для частот (и не только!) можно рассчитывать с помощью программы CIA (Confidence Intervals Analysis), которую можно загрузить с <http://www.medschool.soton.ac.uk/cia/> [12].

В следующей статье будут рассмотрены одномерные способы сравнения качественных данных.

Список литературы

1. Банержи А. Медицинская статистика понятным языком: вводный курс / А. Банержи. — М.: Практическая медицина, 2007. — 287 с.
2. Герасимов А. Н. Медицинская статистика / А. Н. Герасимов. — М.: Медицинское информационное агентство, 2007. — 475 с.
3. Гланц С. Медико-биологическая статистика / С. Гланц. — М.: Практика, 1998.
4. Гржибовский А. М. Типы данных, проверка распределения и описательная статистика / А. М. Гржибовский // Экология человека — 2008. — № 1. — С. 52–58.
5. Жижин К. С. Медицинская статистика: учебное пособие / К. С. Жижин. — Ростов н/Д: Феникс, 2007. — 160 с.
6. Зайцев В. М. Прикладная медицинская статистика / В. М. Зайцев, В. Г. Лифляндский, В. И. Маринкин. — СПб.: Фолиант, 2003. — 428 с.
7. Лакин Г. Ф. Биометрия / Г. Ф. Лакин. — М.: Высшая школа, 1990. — 350 с.
8. Медик В. А. Математическая статистика в медицине / В. А. Медик, М. С. Токмачев. — М.: Финансы и статистика, 2007. — 798 с.
9. Сергиенко В. И. Математическая статистика в клинических исследованиях / В. И. Сергиенко, И. Б. Бондарева. — М.: ГЭОТАР-МЕД, 2001. — 256 с.
10. Юнкеров В. И. Медико-статистическая обработка данных медицинских исследований / В. И. Юнкеров, С. Г. Григорьев. — СПб.: ВмедА, 2002. — 266 с.
11. Agresti A. Approximate is better than exact for interval estimation of binomial proportions / A. Agresti, B. Coull // American statistician. — 1998. — N 52. — С. 119–126.
12. Altman D. Statistics with confidence // D. Altman, D. Machin, T. Bryant, M. J. Gardner. — London: BMJ Books, 2000. — 240 p.
13. 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.
14. Clopper C. J. The use of confidence or fiducial limits illustrated in the case of the binomial / C. J. Clopper, E. S. Pearson // Biometrika. — 1934. — N 26. — P. 404–413.
15. 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.
16. Motulsky H. Intuitive biostatistics // H. Motulsky. — Oxford: Oxford University Press, 1995. — 386 p.
17. 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.
18. 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.
19. Wald A. Confidence limits for continuous distribution functions // A. Wald, J. Wolfowitz // Annals of Mathematical Statistics. — 1939. — N 10. — P. 105–118.
20. Wilson E. B. Probable inference, the law of succession, and statistical inference / E. B. Wilson // Journal of American Statistical Association. — 1927. — N 22. — P. 209–212.

CONFIDENCE INTERVALS FOR PROPORTIONS

A. M. Grjibovski

National Institute of Public Health, Oslo, Norway

The article presents several methods for calculations confidence intervals for binomial proportions, namely, Wald, Wilson, arcsine, Agresti-Coull and exact Clopper-Pearson methods. The paper gives only general introduction to the problem of confidence interval estimation of a binomial proportion and its aim is not only to stimulate the readers to use confidence intervals when presenting results of own empirical research, but also to encourage them to consult statistics books prior to analysing own data and preparing manuscripts.

Key words: confidence interval, proportion.

Контактная информация:

Гржибовский Андрей Мечиславович — старший советник Национального института общественного здоровья, г. Осло, Норвегия

Адрес: Nasjonalt folkehelseinstitutt, Pb 4404 Nydalen, 0403 Oslo, Norway

Тел.: +47 21076392, +47 45268913; e-mail: angr@fhi.no

Статья поступила 16.04.2008 г.