

В следующем номере

- Руководство АМЭЕ № 56: Исследовательский компас: введение в исследования в медицинском образовании
- Обучение в медицинских вузах: проблемы и опыт контроля знаний
-

К вопросу о развитии телемедицины в Республике Армения

Купить номер

Оформить подписку

[Скачать квитанцию](#)

Статистический инструментарий для скептического читателя и медицинского исследователя

1 . 2012

- [Святослав Леонидович Плавинский](#)
- [Василий Викторович Власов](#)

Сущности, изучаемые в медицине и биологии, доступны нашему пониманию только как вероятности событий и приближительные оценки. Поэтому получить адекватное представление о физиологических процессах, течении болезней или результативности лечения можно только, правильно применяя статистические методы.

К сожалению, обучение статистике в процессе медицинского образования проводится в недостаточном объеме, поэтому только самые целеустремленные врачи постигают основы статистического анализа самостоятельно или находят возможность пройти обучение на специализированных курсах, подобных летней школе по методологии научных исследований в Первом Московском государственном медицинском университете им. И.М. Сеченова.

Необходимость в использовании статистического инструментария возникает не только у исследователя, планирующего работу и анализирующего полученные результаты. Возникает она и у скептического читателя, каковым по определению является всякий сторонник доказательной медицины. Естественно, что в эпоху широкого распространения компьютеров сложно ожидать, чтобы исследователи складывали и умножали цифры на бумажке. Предполагается, что они будут использовать компьютерные программы, которые не только ускоряют процесс обработки данных, но и предотвращают ряд ошибок связанных с расчетами.

К сожалению, находящиеся под рукой статистические программы не всегда отвечают запросам "доказательного врача". Например, в них нередко отсутствуют расчеты отношения шансов и его доверительного интервала по сгруппированным данным. Подобные упущения постепенно исправляются, но слишком медленно. Поэтому для помощи "доказательным врачам" в Интернете предлагается ряд калькуляторов, привлекательных своей простотой, но не всегда вызывающих особого доверия ввиду их неизвестной реализации.

Проблема с программными комплексами для статистического анализа данных заключается также в том, что их цена обычно оказывается запредельно высокой для индивидуального исследователя. Поэтому в университетах во всем мире все чаще обращают внимание на такую гибкую и мощную систему статистического анализа, как программный комплекс "R" (R - язык программирования для статистической обработки данных и работы с графикой, а также свободная программная среда вычислений с открытым исходным кодом в рамках проекта GNU. Изначально R был разработан сотрудниками статистического факультета Оклендского университета Россом Айхэка (Ross Ihaka) и Робертом Джентльменом (Robert Gentleman) (первая буква их имен - R), на момент 2011 г. язык и среда поддерживаются и развиваются организацией R Foundation. R широко используется как статистическое программное обеспечение для анализа данных и фактически стал стандартом для статистических программ (Fox, Andersen, 2005)).

Строго говоря, "R" является единственной бесплатной статистической программой профессионального уровня. Система свободно доступна для любого заинтересованного исследователя по адресу <http://cran.r-project.org> и работает в настоящее время под основными операционными системами, встречающимися на персональных компьютерах, - под Linux, различными версиями Windows и MacOS X.

Мощь "R" заключается в том, что оно достаточно легко расширяется. Каждый пользователь может написать функции и предоставить их для использования всем желающим. Будучи один раз загруженной на сайт проекта, функция становится доступной точно так же, как и внутренние команды "R". Группы функций часто объединяются в пакеты, которые устанавливаются на компьютер пользователя и полностью становятся частью системы "R". Таким образом, каждый исследователь может подобрать себе такую конфигурацию "R", которая нужна ему для работы. В настоящий момент уже создано более тысячи различных пакетов и их число растет, поскольку в публикациях по статистике становится хорошим тоном после описания алгоритма приводить его реализацию в "R" (т.е. на языке "S"), а также предлагать для загрузки пакет или набор функций, которые реализуют описанные в статье или книге подходы. "R" получил признание и у производителей коммерческих статистических систем. Такие программы, как SPSS, SAS и Statistica1 включают модуль взаимодействия с "R", т.е. дают возможность использовать функции R внутри своего программного комплекса.

В свою очередь, другие разработчики начинают интегрировать "R" с электронными таблицами (Excel) и текстовыми редакторами. В "R" появляются графические интерфейсы пользователя, такие, как "R Commander", причем последний обладает хорошей русификацией.

Однако у специалиста доказательной медицины при использовании "R" могут возникнуть некоторые сложности, связанные, например, с тем, что часто стоящие перед ним задачи (такие, как расчет доверительного интервала долей, отношений шанса и их доверительных интервалов, оценка чувствительности и специфичности диагностического теста с учетом влияния выборочной ошибки и ряд других) не реализованы в базовом пакете "R" или даже в одном дополнительном пакете (хотя ряд инструментов есть в пакете epiTools). По этой причине Общество специалистов доказательной медицины (ОСДМ) заказало С.Л. Плавинскому разработку ряда функций для облегчения выполнения задач, с которыми часто сталкивается специалист доказательной медицины. В результате был создан набор функций (модуль "osdm.R", который может использоваться исследователями, желающими облегчить себе простейшую обработку своих данных. Этот модуль включает следующие функции:

+Доверительные интервалы относительного риска и отношений шансов

- Доверительные интервалы относительного риска: RRn, RRt, RRp, RR.ci

- Доверительные интервалы отношений шансов: ORn, ORt, ORe, ORc, OR.ci

+Доверительные интервалы доли

- Доверительные интервалы долей, рассчитанные разными методами: `clopper.ci`, `agresti.ci`, `wald.ci`, `wilson.ci`, `poisson.ci`
- Суммарная оценка доверительных интервалов разными методами: `one.freq.ci`

+Унивариантный анализ

- Расчет показателей центральной тенденции: `central(x)`
- Стандартная ошибка: `se`
- Одновременный расчет t-теста для равной и неравной дисперсии и тест дисперсий: `sas.t.test`
- Таблица теста Кохрана-Мантеля-

+Ханзеля для таблиц сопряженности: `cmh.SAS`

- Прямая стандартизация: `adjust.direct`
- Непрямая стандартизация: `adjust.indirect`
- Оценка лет потерянной жизни: `YLL`

+Суммарный анализ табличных данных

- Построение таблиц сопряженности: `CrossTable.Rus`
- Анализ таблиц сопряженности: `proc.freq`
- Расчет чувствительности и специфичности: `diagnostic.freq`

+Построение простейших таблиц

- Таблица с процентами в строке: `table.row.pct`
- Таблица с процентами в столбце: `table.col.pct`
- Расчет таблицы средних и ошибок средних: `table.se`
- Расчет таблицы средних, ошибки, минимума и максимума: `se.table.full`
- Таблицы средних и доверительных интервалов для количественной переменной: `quant.ci.table`
- Таблицы доверительных интервалов для качественной переменной: `qual.ci.table`
- Таблицы доверительных интервалов долей целого `table.pct.ci`

+Таблицы с экспортом в MS Word

- Таблицы доверительных интервалов долей целого `table.pct.ci.word`
- Таблицы доверительных интервалов для качественной переменной: `qual.ci.table.word`
- Расчет таблицы средних, ошибки, минимума и максимума: `se.table.full.word`
- Таблицы средних и доверительных интервалов для количественной переменной: `quant.ci.table.word`

+Графики

- Столбиковые диаграммы с доверительными интервалами: `plot.bar` Расчет размера выборки
- Для обеспечения доверительного интервала заданной ширины (нормальная аппроксимация биномиального распределения): `ci.prop.ssize`
- Для обеспечения доверительного интервала заданной ширины (нормальное распределение): `ci.N.ssize`
- Оценка размера эффекта для дисперсионного анализа - `ES.f`
- Оболочка для `rwf.anova.test`, которая берет количество групп из вектора средних `rwf.anova`
- Оболочка для `rwf.chisq.test`, использует только таблицу вероятностей в случае справедливости альтернативной гипотезы `rwf.chisq`

+Утилиты

- Ввод таблиц Excel в R: `fetch.excel`
- Прямой ввод данных в таблицу: `my.matrix`
- Интерпретация отношений шансов: `interpret.or`

Сам модуль не требует установки, однако для его корректной работы требуется наличие в системе "R" ряда пакетов, которые перечислены в сопровождающей модуль документации. Пакеты загружаются с интернет-сайта и устанавливаются один раз. Модуль загружается либо после каждого старта системы вручную, либо может быть прописан в автозагрузчике (в документации написано, как это сделать). После того как модуль установлен, исследователь может легко рассчитать, например, доверительный интервал Клоппера-Пирсона для доли. Для этого в командной строке "R" надо просто написать "`clopper.ci(5, 40)`", если в группе из 40 человек исход наступил у 5. Если надо рассчитать интервал Агрести-Коула, потребуется команда: "`agresti.ci(5,40)`".

Для анализа четырехпольных таблиц надо просто вызвать функцию "`proc.freq()`", и перед исследователем появится четырехпольная интерактивная таблица, куда надо ввести четыре частоты события и нажать кнопку "OK". После этого будет построена таблица сопряжения со всеми процентами, представлены результаты тестирования гипотезы о независимости строк и столбцов, рассчитаны отношения шансов и относительный риск и их доверительные интервалы и даже распечатана табличка, облегчающая интерпретацию этих показателей связи. Аналогичным образом команда "`diagnostic.freq()`" вызовет интерактивную четырехпольную таблицу для заполнения, а затем рассчитает чувствительность и специфичность с доверительными интервалами, положительную и отрицательную предиктивные оценки, а также отношения правдоподобия.

2 утилиты могут оказаться очень полезными для начинающих - "`fetch.excel()`" позволяет легко вводить в систему "R" данные, которые хранятся в формате Excel, а "`my.matrix()`" дает возможность интерактивного ввода в систему R табличных данных. Учитывая особенности "R", например, для того чтобы построить мозаичную диаграмму из пакета "`vcd`" на основании таблицы 3x3, можно воспользоваться следующей командой "`mosaic(my.matrix(3,3))`". Это вызовет интерактивную таблицу для заполнения ячеек, после нажатия на кнопку "OK", в которой будет построена мозаичная диаграмма. Естественно, что эту функцию можно использовать и для других задач, где требуется ввод суммарных данных.

Также этот набор функций обеспечивает создание простых таблиц для количественных и качественных переменных, которые можно экспортировать в текстовый редактор Word (функции создают таблицы в формате HTML, который затем можно открыть в текстовом редакторе). Все функции модуля описаны в документации вместе с примерами их использования.

Конечно, все описанные выше задачи можно решить в "R" и другими способами, но тем не менее модуль предлагает простое решение для автоматизации наиболее частых расчетов и облегчения работы исследователя, причем, что немаловажно, использование как "R", так и модуля "osdm.R" совершенно бесплатно и легально.

Модуль, его описание и инструкции доступны на сайте Общества специалистов доказательной медицины www.osdm.org.

дополнительные материалы

© Издательская группа «ГЭОТАР-Медиа»,
1995-2013 Все права защищены
[Условия пользования сайтом](#)
Издательская группа «ГЭОТАР-Медиа» 115035,
г. Москва, Садовническая ул., д. 9, стр. 4
(метро Новокузнецкая) Телефон: (495) 921-39-07
www.geotar.ru

Журнал основан в 2010 г.
Периодичность: 4 номера в год Свидетельство о регистрации
средства массовой информации от 08 июля 2010 г.:
серия ПИ № ФС 77 - 40698, Федеральная служба по надзору
в сфере связи, информационных технологий и массовых
коммуникаций (Роскомнадзор).

Подписные индексы:
Каталог агентства «Роспечать»:
36185 - для индивидуальных подписчиков,
37141 - для предприятий и организаций.
Каталог «Пресса России»:
89018 - для индивидуальных подписчиков,
89019 - для предприятий и организаций.