

Ошибки статистического анализа биомедицинских данных

Леонов Василий Петрович,
редактор сайта
БИОМЕТРИКА

E-MAIL: leo.biostat@gmail.com

(Лекция в Якутске, 13.11.2009)

Ошибки статистического анализа биомедицинских данных

Ошибки не есть лженаука.
Лженаука –
это непризнание ошибок!

П.Л. Капица

Ошибка – это...?

«Словарь русского языка» С.И. Ожегова (М.; Русс. яз., 1986. – 797с.) определяет

это слово так: «ОШИБКА.

Неправильность в действиях, мыслях».

Как тут не вспомнить известное выражение «... разруха не в клозетах, а в головах».

(Михаил Булгаков. Собачье сердце).

Ошибка – это...?

Часто вспоминают слова Дизраели: «Существуют три вида лжи: просто ложь, наглая ложь и статистика». При этом под статистикой обычно понимают ложные наборы статистических данных, взятые чаще всего с потолка и ставящие целью подкрепить чьи-то предвзятые суждения.

Ошибка – это...?

Но бывают и другие случаи, когда используются вполне доброкачественные наборы данных и никаких дурных целей не ставят. Однако неправильное использование статистического аппарата приводит к неверным выводам. Эти ошибки особенно часто происходят в медицине: ведь известно, что врачи не блещут математическими знаниями и поэтому легко впадают в ошибки, когда пользуются статистическими методами. Их жертвами становятся больные, и таких жертв немало, хотя оценить число таких жертв совсем непросто.

Ошибка – это...?

Выделим два основных вида статистических ошибок:

- 1) заведомо ложные данные, результаты анализа которых также заведомо ложны;**
- 2) ложные результаты, как следствие неправильного использования статистического инструментария.**

Ошибка или обман?

**Тьмы низких истин мне дороже
нас возвышающий обман.**

А.С. Пушкин.

Ошибка или обман?

«В науке, как в любой другой профессиональной области, жулики были, есть и, увы, будут. Правде надо смотреть в глаза, какой бы горькой она ни казалась. Раньше у нас главными побудительными мотивами для научного мошенничества были карьеристские устремления и страх. ... взамен появился новый, более коварный стимул для слабых волей научных сотрудников. Это стремление во чтобы то ни стало, получить грант, или финансирование собственных исследований. По мере интеграции отечественной науки в мировую эта опасность будет усугубляться. И чем раньше мы перейдем западный опыт борьбы с тамошними жуликами от науки, тем будет лучше».

Мошенничество в науке

В.С.Логинов, С.И.Дорошенко

<http://www.biometrica.tomsk.ru/obman.htm>

Обзоры ошибок применения статистики в медицине

**Сомнительный результат не может
признаваться в качестве научного.**

Н. Бор.

Обзоры ошибок применения статистики в медицине

В СССР первым обзором таких ошибок, является книга А.Я.Боярского «Статистические методы в экспериментальных медицинских исследованиях». 1955 Медгиз.

Автор на значительном конкретном материале констатировал крайне неблагоприятную ситуацию с применением статистики в экспериментальной медицине и биологии тех лет.

Обзоры ошибок применения статистики в медицине

«Уже беглое ознакомление с состоянием дела показывает, что статистическая обработка экспериментальных данных является наиболее слабым местом во многих исследованиях. ...

Трудно требовать от медика, чтобы он, наряду со знаниями в своей собственной области, был в то же время достаточно вооруженным, скажем, в радиотехнике для конструирования аппаратуры, улавливающей биотоки, или в статистике для нахождения наиболее правильных методов статистической обработки своих экспериментальных данных».

Обзоры ошибок применения статистики в медицине

«И подобно тому, как медику, несомненно, приходится обращаться за содействием к радиотехнику, для правильной статистической обработки экспериментальных данных нередко приходится обращаться к специалисту-статистику».

Обзоры ошибок применения статистики в медицине

«Так или иначе, но бесспорным фактом являются и недостаточная вооруженность медиков статистическими знаниями, и недостаточно высокий научный уровень статистической методики в большинстве их экспериментальных работ».

Эти слова были сказаны еще в 1955г. Полвека назад констатировалось, что медицинские исследования требуют обязательного участия биостатистика.

Обзоры ошибок применения статистики в медицине

Одна из первых зарубежных публикаций, содержащая анализ статистических ошибок в медицинских статьях, относится к 1929 г.

[Dunn, H.L , Application of statistical methods in physiology. *Physiological Reviews*, 1929, v. 9, pp. 275-398.]

В этой статье Dunn сообщил, что примерно половина статей, публикуемых журнале *Physiological Reviews*, содержит примеры ошибочного использования статистики.

Обзоры ошибок применения статистики в медицине

Результаты более общего исследования были опубликованы в 1932 г. Начиная с 60-х годов прошлого века в зарубежной периодике периодически публикуются такие обзоры. Большую работу в этом направлении проводит статистик Douglas G. Altman. В этих обзорах отмечается рост доли публикаций, в которых используется статистика, а также уменьшение доли ошибочного применения этих методов. Такая тенденция характерна не только для европейских журналов, но и для китайских журналов.

Обзоры ошибок применения статистики в медицине

Следует отметить, что в российской медицинской периодике подобные публикации крайне редки. Причем не потому, что такие исследования не ведутся, а по причинам иного порядка. Далеко не каждый журнал отважится опубликовать статью, в которой будет дан критический анализ его публикаций. Однако имеются примеры и иного отношения к такой критике.

Обзоры ошибок применения статистики в медицине

Так в начале 1997 г. автор этих строк направил в журнал «Кардиология» рукопись «Применение методов статистики в кардиологии (по материалам журнала «Кардиология» за 1993-1995 гг.)». Поскольку редакция журнала в течение года никакого ответа автору не дала, то я решил, что рукопись отвергнута и не опубликована. Однако в апреле 2004 г., на семинаре по биометрике в Самаре, одна из слушательниц передала мне копию данной статьи, которая была опубликована. [Леонов В.П. Применение методов статистики в кардиологии (по материалам журнала «Кардиология» за 1993-1995 гг. 1998, Кардиология, с. 55-58.]

Причины возникновения ошибок и меры борьбы с ними

Каждый пишет, как он слышит.
Каждый слышит, как он дышит.
Как он дышит, так и пишет,
не стараясь угодить...

Б.Окуджава.

Причины возникновения ошибок и меры борьбы с ними

То, что такие ошибки проникают на страницы российских журналов и диссертаций, есть результат отсутствия их рецензирования профессиональными биостатистиками.

Гораздо более высокое качество статистических выводов в статьях зарубежных медицинских журналов обеспечивается наличием именно такого статистического рецензирования.

Причины возникновения ошибок и меры борьбы с ними

Причины возникновения таких ошибок присутствуют и на более раннем этапе. Большинство авторов таких публикаций не профессиональные исследователи. Они – прежде всего клиницисты, и поэтому не должны иметь необходимой для исследователя статистической подготовки. Этот вывод можно сделать и на основе содержания квалификационной характеристики биологов и медиков

Причины возникновения ошибок и меры борьбы с ними

В квалификационных характеристиках в разделе «Знание смежных дисциплин» отсутствует упоминание статистики. И это разумно, т.к. **медицинский вуз не может, да и не должен, обеспечивать своим выпускникам профессиональный уровень владения биостатистикой.** Его задача подготовить хороших врачей, врачей, а не исследователей и специалистов по биостатистике.

Причины возникновения ошибок и меры борьбы с ними

Чем можно объяснить многолетнее воспроизводство статистических ошибок в российских публикациях биомедицинской тематики? Основная причина - некритическое отношение специалистов как к собственной исследовательской практике, так и к практике своих коллег.

Результатом этого является неявное следование следующим четырём установкам:

Причины возникновения ошибок и меры борьбы с ними

использовать тот статистический метод, который предложил руководитель (ослушаться нельзя, а если кто-то обвинит в некорректности, то виноват шеф);

применять данный метод потому, что он есть в доступном статистическом пакете (надо же написать в статье, что использовал статистический пакет);

Причины возникновения ошибок и меры борьбы с ними

использовать те методы, которые ранее уже были самым автором использованы (за них никто не поругал, значит можно и впредь их применять);

использовать данный метод потому, что его уже использовал другой автор опубликованной статьи (диссертации) по этой же тематике (раз статью опубликовали (диссертацию защитили), значит можно использовать, ведь в редакциях (диссертационных советах) не дураки же сидят).

Причины возникновения ошибок и меры борьбы с ними

использовать те методы, которые ранее уже были самым автором использованы (за них никто не поругал, значит можно и впредь их применять);

использовать данный метод потому, что его уже использовал другой автор опубликованной статьи (диссертации) по этой же тематике (раз статью опубликовали (диссертацию защитили), значит можно использовать, ведь в редакциях (диссертационных советах) не дураки же сидят).

Причины возникновения ошибок и меры борьбы с ними

Следование этим установкам приводит к генерации цепочек так называемых «мемов», когда несколько авторов последовательно заимствуют друг у друга одну и ту же ошибку.

Меметический анализ описаний методов статистики

Меметика – аналог генетики/

В работе «Эгоистический ген» Ричардом Доукинсом была предложена концепция репликатора в приложении к социокультурным процессам, которая более детально была им описана в его следующей работе «Расширенный фенотип». В данных работах автор впервые ввел понятие **«мема»**, используемого им для описания процессов хранения и распространения отдельных элементов культуры

Меметический анализ описаний методов статистики (см. [ссылку](#))

Часть используемых мемов вполне адекватно описывает несложный статистический инструментарий применяемый авторами. Однако подавляющее большинство меметических описаний несут на себе отпечаток фрагментарности и логической и вербальной незавершенности их конструкции.

В этой ситуации автор уподобляется умельцу, которому в руки попал конструктор типа «Сделай сам» с набором тех или иных элементов, но при том, что в коробке отсутствует программа сборки нужного изделия.

Причины появления и живучести ошибок

1. **Отсутствие знаний по статистике** у авторов публикаций, членов журнальных редколлегий и диссертационных советов, что не позволяет им осознать всю степень ошибочности выводов в таких публикациях.
2. **Деформация цели публикации.** Нередко истинной целью публикации является не стремление ученого обнародовать, легализовать надежные, достоверные результаты исследования, а лишь сам факт быстрой публикации даже ненадежных выводов, как доказательство «научности» результатов.

Причины появления и живучести ошибок

3. **Отдаленность во времени и пространстве возможных негативных последствий** опубликования сомнительных результатов от автора, времени и места публикации. Автор такой публикации, зная о том, что его выводы далеко не бесспорны, понимает, что именно эта неочевидность и не будет способствовать стремительному их внедрению в реальную медицинскую практику. А если такое и случится, то доказать, что основной причиной негативных последствий были именно недоброкачественные выводы автора этой публикации, тоже весьма проблематично.

Причины появления и живучести ошибок

4. Ограниченность доступа широкой читательской аудитории к медицинским публикациям, содержащим ошибочные выводы. Это связано как с ограниченными тиражами медицинской периодики, так и с ограничением доступа к медицинским диссертациям. В ряде медицинских вузов доступ к защищенным диссертациям разрешается ректором вуза.

Причины появления и живучести ошибок

5. Отсутствие доступа для читателей к исходным данным автора публикации. Автор, зная, что любой читатель, сомневающийся в истинности авторских выводов, имеет возможность немедленно (или же спустя некоторое время) получить эти данные, и самостоятельно перепроверить авторские выводы, будет совершенно иначе относиться к процедуре статистического анализа этих данных.

Причины появления и живучести ошибок

6. Незаинтересованность редакций периодических изданий, ректоров медицинских вузов, диссертационных советов, ВАК РФ и дирекций НИИ в повышении качества журнальных публикаций и диссертаций.

Ошибки описания статистических методов

Вырази ложную мысль ясно,
И она сама себя опровергнет.

Л. Вовенарг

Ошибки описания статистических методов

В 25% публикаций вообще нет информации об использовании статистических методов.

Авторы ограничиваются выражениями вида « $p < 0,05$ » или « $p > 0,05$ ».

Ошибки описания статистических методов

Оставшиеся 75% работ содержат в той или иной форме описание использованных методов статистики, обычно в разделе (главе) «Материал и методы». В большинстве работ описание статистических методов и критериев дается в форме шаблонных клише с весьма расплывчатым содержанием. Можно выделить три группы подобных клише.

Ошибки описания статистических методов

Первая группа: просто констатируется сам факт использования статистических методов

Для нее характерны следующие описания:

Результаты **обработаны статистически.**

Обработку данных проводили **статистическими методами.**

Результаты обработаны **методом (методами) вариационной статистики** (варианты: стандартными программами, общепринятыми методами и т.п.).

Результаты **подвергнуты статистической обработке.**

Ошибки описания статистических методов

Пример: В статье «Варьирование активностей цистеиновых катепсинов почек и печени» (БЭБМ, 1995, №12, с.586-589): «**Математическую обработку проводили обычным способом с использованием компьютера**». Далее в статье авторы нигде не упоминают ни одного конкретного метода статистического анализа и статистического критерия, хотя и используют выражения $p < 0,05$ и $p < 0,01$. Попробуйте представить себе аналогичное описание лечения больного: «**Лечение производилось обычным способом с использованием лекарственного препарата**»

Ошибки описания статистических методов

Для второй группы работ (примерно 30%) характерно указание на более конкретные детали: Материал обрабатывали статистически **по методу Лакина** (варианты: по В.Ю. Урбаху; по таблицам Стрелкова).

Достоверность значений определяли по **t-критерию Стьюдента** (варианты: по среднеквадратичному отклонению усредненных значений параметров; методом определения среднего квадратического отклонения от средней величины).

Статистическая обработка материала произведена **с использованием компьютера** IBM PC, (Pentium II) по стандартным программам.

Ошибки описания статистических методов

Очевидно, что для читателя ссылка на определенную фамилию (Лакин, Урбах, Стрелков) ни о чем не говорит, поскольку в книгах данных авторов описано много разнообразных методов. Что сможет узнать врач, например, о методике лечения больных, если автор статьи или диссертации напишет: «**Лечение больным назначалось по Машковскому**»? Делать подобную ссылку все равно, что утверждать наличие методики Машковского, только по той причине, что есть лекарственные справочники с такими именами.

Ошибки описания статистических методов

Третья группа проанализированных работ (примерно 25%) содержит достаточно безграмотные, а подчас и просто абсурдные либо противоречивые сочетания перечисляемых терминов или критериев. В таких описаниях авторы часто вводят свою собственную терминологию, не раскрывая в тексте работы смысла используемых терминов. Вот типичный пример такого описания. [Жерлов Г.К., 1991] «**Достоверность различия между средними арифметическими сравнениями** вариационных рядов устанавливалась по **степени вероятности положительной гипотезы** в соответствии с описанием, приведенным в монографии Г.Ф. Лакина»

Ошибки описания статистических методов

Лечение хронического описторхоза
препаратом из растительного сырья
(попутрилом) и оценка его
эффективности. (Канд. дисс., 14.00.09 -
педиатрия):

Достоверность различия сравниваемых
величин определяли с помощью показателя
точности P по таблице Стьюдента, где он
располагается в зависимости от значений t
и n .

Ошибки описания статистических методов

Влияние некоторых стимуляторов центральной нервной системы на неспецифическую иммунобиологическую реактивность организма.
(Канд. дисс., 14.00.10 - инфекционные болезни):

Достоверность определяли с помощью абсолютного показателя точности « r », который находили по специальной таблице, где он расположен в зависимости от (t) Стьюдента и числа степеней свободы « n ».

Ошибки описания статистических методов

Зырянов Б.Н. и др. Адьювантная химиотерапия рака желудка. Томск, 1996, изд-во ТГУ, 375 с.

Достоверность *различий* средних арифметических определяли по абсолютному показателю точности процентных точек распределения Стьюдента в зависимости от коэффициента достоверности (t) и числа степеней свободы (n).

Ошибки описания статистических методов

Ультразвуковая и радионуклидная диагностика патологии гепатобилиарной системы у детей дошкольного возраста. (Канд. диссер., 14.00.09 - педиатрия, 14.00.19 - лучевая диагностика):

Достоверность **различных** средних арифметических определяли **по абсолютному показателю точности (P) по таблице** процентных точек распределения Стьюдента **в зависимости от коэффициента достоверности (t) и числа степеней свободы (n).**

Ошибки описания статистических методов

Ультразвуковая и радионуклидная диагностика патологии гепатобилиарной системы у детей дошкольного возраста. (Канд. диссер., 14.00.09 - педиатрия, 14.00.19 - лучевая диагностика):

На основании t по таблице Стьюдента определялась вероятность различия (p).
Различие считалось достоверным при $p < 0,05$, т.е. в тех случаях, когда вероятность различия составляла больше 95% . (стр. 40)

Ошибки описания статистических методов

Зырянов Б.Н. и др. Адьювантная химиотерапия рака желудка. Томск, 1996, изд-во ТГУ, 375 с.

На основании критерия t по таблице Стьюдента определялась вероятность различия. Различие считалось достоверным при $p < 0,05$, т.е. в тех случаях, когда вероятность различия составляла больше 95%.(стр. 48)

Ошибки описания статистических методов

С. 30 диссертации «Клиника и дифференциальная диагностика инфекционных экзантем у детей» и сделаем попытку его анализа. «**Для определения значимости выборочных показателей, оценки сущности двух или нескольких показателей,** а также **определения связи между явлениями, полученных** (так в тексте, выделено нами - В.Л.) **в результате выборочных исследований, были использованы следующие формулы: ...**». Каков смысл утверждения – «**определение значимости выборочных показателей**»?

Локализация ошибочных описаний внутри научных школ

Диссертация 7. Диссертация на соискание ученой степени кандидата медицинских наук "Роль поджелудочной железы в регуляции антитрипсиновой и трипсиновой активности крови." Диссертант - Афанасьева А.Н., специальность 14.00.17 - нормальная физиология. Сибирский государственный медицинский университет, Томск - 1990 г.

Диссертация 8. Диссертация на соискание ученой степени доктора биологических наук "Роль трипсина в регуляции моторной функции тонкой кишки." Диссертант - Васильев В.Н., специальность 14.00.17 - нормальная физиология. Сибирский государственный медицинский университет, Томск - 1992 г.

Локализация ошибочных описаний внутри научных школ

Диссертации 9. Диссертация на соискание ученой степени кандидата медицинских наук "Роль поджелудочной железы в регуляции антитрипсиновой и трипсиновой активности крови." Диссертант - Афанасьева А.Н., специальность 14.00.17 - нормальная физиология. Сибирский государственный медицинский университет, Томск - 1990 г.

Диссертация 10. Диссертация на соискание ученой степени доктора медицинских наук "Функционирование калликреин-кининовой системы крови в норме и при патологии." Диссертант - Шумилов С.П., специальность 14.00.16 - пат. Физиология, 14.00.17 - нормальная физиология. Сибирский государственный медицинский университет, Томск - 1995 г.

Локализация ошибочных описаний внутри научных школ

Диссертация 11. Диссертация на соискание ученой степени доктора биологических наук "Механизмы образования и деградации надэпителиального слизистого слоя пищеварительного тракта." Диссертант - Кривова Н.А., специальность 14.00.17 - нормальная физиология. НИИ биологии и биофизики Томского государственного университета, Томск - 1994 г.

Локализация ошибочных описаний внутри научных школ

В этих 5 диссертациях один и тот же человек выступает в роли научного руководителя в кандидатских диссертациях и в роли научного консультанта в докторских диссертациях. Автор же 8-й докторской диссертации выступает в качестве научного руководителя в кандидатской диссертации 7. **Все эти факты позволяют говорить о принадлежности диссертантов к одной научной школе.**

Эти работы достаточно близки по:
срокам выполнения работ: - 1992г., 1993г, 1994г. и 1995г;
содержанию исследования и месту выполнения;
структуре названия работ - **в трех диссертациях одно и то же первое слово, а в двух диссертациях полностью совпадают первые четыре слова.**

Локализация ошибочных описаний внутри научных школ

**Достоверными считались
отличия с уровнем
доверительной
вероятности
 $p < 0,05$**

Локализация ошибочных описаний внутри научных школ

«Вероятности, признанные достаточными для уверенного суждения о генеральных параметрах на основании известных выборочных показателей, называют доверительными вероятностями. Понятие о доверительных вероятностях предложено Р.Фишером. Оно вытекает из принципа, который положен в основу применения теории вероятностей к решению практических задач. Согласно этому принципу, маловероятные события считают практически невозможными, а **события, вероятность которых близка к единице, принимают за почти достоверные. Обычно в качестве доверительных используют вероятности $P = 0,95$; $P = 0,99$; $P = 0,999$** ».

Лакин Г.Ф. Биометрия. М.: Высшая школа, с. 106.

Локализация ошибочных описаний внутри научных школ

Таким образом, 5 авторов, использовавших приведенный выше ошибочный мем, утверждали, что степень их уверенности в своих выводах не более 5%!

Мемы камуфляжных описаний

Внимательное изучение ошибочных описаний в диссертациях выполненного статистического анализа обнаруживает ряд признаков, которые позволяют в ряде случаев говорить о **камуфляжном характере** такого описания. По своему содержанию **такие описания представляются не более чем попытками «онаучить» авторские выводы** перечислением разнообразных статистических методов и терминов.

Локализация ошибочных описаний внутри научных школ

В редакцию «Сибирского медицинского журнала» поступила на рецензию рукопись «Опыт лечения вазилипом больных с острым коронарным синдромом без подъёма сегмента ST» в которой сообщалось, что «Для анализа выживаемости использовали логранговый критерий». Из текста же рукописи следовало, что за период лечения вообще не было смертельных исходов. Возникает вопрос, с какой же целью и как проводился «анализ выживаемости», если все пациенты остались живы? Когда этот вопрос был задан автору рукописи, она не смогла ответить на данный вопрос, сославшись на то, что статистический анализ выполнял другой человек, который не является автором.

Смутно пишут о том, о чём смутно представляют

Смутное представление о том, что означают те или иные величины, используемые в статистическом анализе, приводит к тому, что авторы публикаций путают их между собой. Вот какое, к примеру, описание приведено в статье «Феномен стабильности роста опухоли в организме хозяина». Новый подход к концентрации роста и лечения опухоли.

«**Статистическую обработку полученных данных проводили методом Фишера-Стьюдента. Различия считали достоверными при $t < 0,05$** ». Как видим, авторы этой работы перепутали уровень значимости «p» и критерий Стьюдента «t».

Смутно пишут о том, о чём смутно представляют

Нередко авторы путают между собой и такие величины, как уровень значимости и доверительная вероятность. Вот что, к примеру, можно прочесть в статье (журнал «Генетика») «Анализ взаимосвязи полиморфизма C677T гена метилентетрагидрофолатредуктазы с клиническими проявлениями атеросклероза» (Авторы М.Г. Спиридонова, В.А. Степанов, В.П. Пузырев, Р.С. Карпов.): «Для всех статистических тестов в качестве критерия статистической достоверности рассматривался **уровень значимости более 0,95**».

Смутно пишут о том, о чём смутно представляют

Обратимся к толкованию этого термина в «Статистический словарь. 1989». На с. 542 читаем: «**УРОВЕНЬ ЗНАЧИМОСТИ** - одна из характеристик качества критерия статистической проверки гипотез. На практике часто пользуются след. стандартными значениями альфа: 0,1 , 0,05 , 0,025 , 0,01 , 0,005 , 0,001. Особенно распространенной является величина **УРОВНЯ ЗНАЧИМОСТИ** альфа равная 0,05. Она означает, что в среднем в пяти случаях из ста ошибочно отвергают высказанную гипотезу при пользовании данным *критерием статистическим*».

Смутно пишут о том, о чём смутно представляют

Фактически авторы этой статьи из журнала «Генетика» утверждают, что они верят собственным результатам менее, чем на 5%!

Таким образом, используя уровень значимости более 0,95, авторы ошибочно отвергли высказанные гипотезы более чем в 95 случаях из 100. О какой же «достоверности» можно говорить в этом случае? **Самое печальное в таких ошибках то, что читателю невозможно установить саму природу ошибки.** Или же это ошибка только описания, и тогда мы имеем дело всего лишь с терминологической путаницей, но сами выводы верны, или же данная ошибка присутствовала уже на этапе выполнения анализа, и тогда уже ошибочны сами выводы.

Смутно пишут о том, о чём смутно представляют

Аналогичную путаницу обнаружим и в работах:

{ Кривова Н.А. 1994 } на с. 68. которой читаем «**Достоверными считали различия с уровнем доверительной вероятности меньше 0,05**»;

{ Харьков Е.И. 1987 } на с. 37 которой читаем: «Затем находилась достоверность различий по критерию Стьюдента при **уровне доверительного интервала $P=0,05$** ».

{ Артюхов В.Г., Башарина О.В., Пантак А.А., Свекло Л.С. 2000 } «Достоверность различий оценивали с помощью t критерия Стьюдента при **95% уровне значимости**»;

{ Шварц Я.Ш., Зубахин А.А., М.И. Душкин М.И. 2000 } с. 172-175: «Статистическую обработку проводили с использованием t критерия Стьюдента. при **уровне доверительной вероятности $p < 0,05$** »

Коварный t-критерий Стьюдента

Нужно делать так, как нужно,
а как не нужно - делать не
нужно.

Винни-Пух.

Коварный t-критерий Стьюдента

Наибольшей популярностью при проверке гипотез о равенстве генеральных средних пользуется t-критерий Стьюдента. При чтении статей БЭБМ и «Вестника РАМН» складывается впечатление, что большинство авторов этих журналов знают и используют лишь t-критерий Стьюдента. Например, в выпусках БЭБМ за 1997 г. t-критерий использован в 125 статьях, тогда как корреляционный и дисперсионный анализ применен всего лишь в 15, критерий Колмогорова-Смирнова - в одной, парная линейная регрессия - в трех, точный критерий Фишера - в трех статьях.

Коварный t-критерий Стьюдента

Как известно, использование t-критерия Стьюдента имеет два ограничения, а именно, нормальность распределения в обеих сравниваемых группах, и равенство генеральных дисперсий. Из 1562 проанализированных нами статей, монографий, диссертаций и авторефератов, авторы которых использовали t-критерий Стьюдента, упоминание о проверке нормальности распределения исследуемых признаков было только в 23 работах! О проверке второго ограничения – на равенство генеральных дисперсий упоминалось лишь в одной работе.

Коварный t-критерий Стьюдента

Наши исследования нормальности распределения биомедицинских признаков, проведенные в течение 30 лет более чем на 10 тысячах переменных, показали, что **примерно 75% используемых переменных не подчиняются нормальному распределению.**

Коварный t-критерий Стьюдента

Очевидно, что без наличия исходных данных читателю проверить факт нормальности распределения признаков, анализируемых в таких публикациях, невозможно. Проверка второго условия возможности применения t-критерия, равенства генеральных дисперсий, чаще всего для читателя доступна.

Коварный t-критерий Стьюдента

В этом случае для проверки второго требования – равенства генеральных дисперсий, необходимы лишь по два параметра из каждой из сравниваемых групп. Это **выборочные дисперсии** и **объёмы наблюдений**. Используя их можно вычислить F-критерий Фишера и далее оценить достигнутый уровень значимости «р».

Коварный t-критерий Стьюдента

Чаще всего авторы подобных публикаций приводят в таблицах объем наблюдений «n», и либо стандартное (среднеквадратичное) отклонение «s», либо ошибку среднего «m» для каждой из сравниваемых групп. В первом случае проверку гипотезы о равенстве генеральных дисперсий двух совокупностей можно произвести вычислив F-критерий Фишера по формуле $F = s_1^2/s_2^2$, где s_1^2 максимальная по величине дисперсия, а s_2^2 – минимальная дисперсия.

Во втором же случае используя значение «m» и «n» можно получить значение выборочной дисперсии s^2 по формуле $s^2 = m^2 * n$ для каждой из групп, и далее вычислить F-критерий Фишера.

Коварный t-критерий Стьюдента

В качестве примера такой послепубликационной проверки допустимости применения t-критерия Стьюдента (См. ссылку) рассмотрим табл. 1 из статьи «Поверхностная архитектура эритроцитов периферической крови у психически больных». Новицкий В.В., Рязанцева Н.В., Семин И.Р. Сибирский государственный медицинский университет, Томск. НИИ фармакологии Томского научного центра РАМН. Бюллетень экспериментальной биологии и медицины, вып. 10, 2000, стр. 429-432. В статье сообщается, что «Достоверность различий между сравниваемыми группами оценивали с использованием t критерия Стьюдента. ... Измеряли внешний диаметр клетки и размер центральной впадины у 50 произвольно выбранных дискоцитов ... ».

Коварный t-критерий Стьюдента

Отметим, что в данной работе авторы изучали 4 группы пациентов: здоровые доноры (контрольная группа); больные с непсихотическими расстройствами; больные с умственной отсталостью; больные с параноидной шизофренией. Между тем авторы использовали t-критерий Стьюдента, применяя его поочередно для разных пар групп сравнения, что для данного критерия, без использования специальных поправок, недопустимо. В результате такого приема возникает так называемая **«ошибка множественных сравнений»** приводящая к тому, что **авторы будут чаще обнаруживать желаемое различие, нежели оно существует на самом деле.** Данная ошибка является второй по распространенности после использования критерия Стьюдента без проверки двух имеющихся ограничений. В результате ошибочные выводы вводят в заблуждение не только самих авторов подобных публикаций, но и читателей, на которых к тому же действует магия академических регалий авторов.

**«... не зная законов языка
ирокезского, можешь ли ты делать
такое суждение по сему предмету...»**

**Как блестящие идеи, так и
научные нелепости одинаковым
образом можно облечь во
впечатляющий мундир формул и
теорем.**

В.В.Налимов

**«... не зная законов языка
ирокезского, можешь ли ты делать
такое суждение по сему предмету...»**

В ряде случаев встречаются достаточно одиозные ошибки, которые демонстрируют полнейшее непонимание их авторами сути используемых методов. Чаще всего они связаны с использованием сложных статистических методов. Первый пример заимствован из докторской диссертации В.В. Климова (Клиническая оценка реакции нейтрофилов при острой пневмонии у детей», специальность 14.00.09 - педиатрия, 14.00.36 - аллергология и иммунология. Томск. СГМУ).

«Статистическая обработка данных ... проводилась на основе пакета анализа биомедицинской информации «BMDP». Использованы **вариационный**, корреляционный, регрессионный, факторный, кластерный анализы на ЭВМ ЕС 1033. Проверка достоверности выявленных различий осуществлялась по критерию Стьюдента...».

**«... не зная законов языка
ирокезского, можешь ли ты делать
такое суждение по сему предмету...»**

В диссертации имеется ряд таблиц, содержащих результаты факторного анализа. При этом использованы следующие качественные признаки:

Аномалии иммунитета

Стигмы дисэмбриогенеза

Аллергические диатезы

ЧРЗ (частые респираторные заболевания)

Хронические очаги в носоглотке

Паразитарные инвазии

Преморбидный фон (без иммунопатологии)

Аллергические реакции

Этиология (возбудитель)

Неблагоприятное микросоциальное окружение

«... не зная законов языка ирокезского, можешь ли ты делать такое суждение по сему предмету...»

Факторный же анализ базируется на анализе **непрерывных, количественных признаков, связанных между собой линейными корреляционными соотношениями.**

Следовательно, выводы приведённые в данной докторской диссертации являются некорректными, ложными.

**«... не зная законов языка
ирокезского, можешь ли ты делать
такое суждение по сему предмету...»**

Наша критика в адрес авторов,
безграмотно применяющих критерий
Стьюдента нередко приводит к
курьезным последствиям. В тех
случаях, когда этот критерий
неприменим, авторы используют
другой критерий, однако
незаслуженно приписывая ему
свойства критерия Стьюдента.

«... не зная законов языка ирокезского, можешь ли ты делать такое суждение по сему предмету...»

Проанализируем цитату из доклада «Цитогенетические эффекты ядерно-химического производства». Пленарный доклад на конференции «Медицинские и экологические эффекты ионизирующей радиации». Пузырев В.П., Назаренко С.А., Попова Н.А. (к 15-летию аварии на Чернобыльской АЭС)": Материалы 1 международной научно-практической конференции, 21-22 июня 2001 года. Северск-Томск./ ред. Р.М. Тахауов, Л.В. Капилевич, А.Б. Карпов - Томск, 2001. - 144с.

На 21 странице читаем: «Статистический анализ показал, что по ряду цитогенетических показателей имеется несоответствие распределения эмпирических данных нормальному закону и **поэтому t-критерий Стьюдента не может использоваться для сравнения средних значений выборок. Более адекватным для этой цели является непараметрический критерий однородности Колмогорова-Смирнова. ... приблизительно у 70% обследованных работников СХК и жителей г. Северска отмечается существенное повышение частоты как хромосомных aberrаций нерадиационной природы, так и уровня СХО».**

«... не зная законов языка ирокезского, можешь ли ты делать такое суждение по сему предмету...»

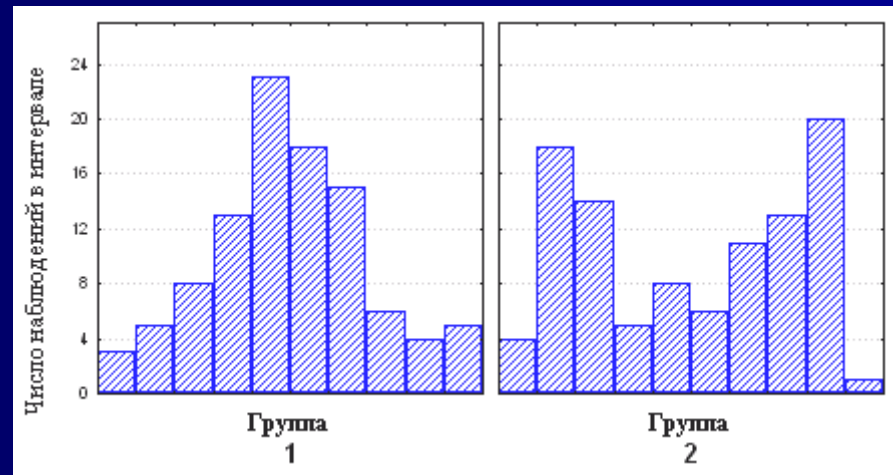
Как видим, используя **критерий Колмогорова-Смирнова**, авторы делают вывод о существенном повышении средней частоты хромосомных аберраций у определенной категории лиц. Однако данный вывод о столь важном показателе весьма сомнителен по следующей причине. **В отличие от критерия Стьюдента, критерий Колмогорова-Смирнова чувствителен к любому различию в параметрах распределений. При этом он не указывает, с каким именно отклонением мы имеем дело**

«... не зная законов языка ирокезского, можешь ли ты делать такое суждение по сему предмету...»

Рассмотренные выше особенности этого критерия проиллюстрируем с помощью массива данных, в котором имеются две группы наблюдений, в каждой из которых по 100 наблюдений. Количественный признак представлен переменной VAR1, группирующий признак - VAR2. Вторая группа была сконструирована из первой группы наблюдений следующим образом. После сортировки признака VAR1 в первой выборке по возрастанию, к первым 50 наблюдениям прибавили одно и то же число, тогда как от следующих 50 наблюдений это же самое число вычли. Полученные при этом новые значения признака VAR1 и составили вторую группу.

«... не зная законов языка ирокезского, можешь ли ты делать такое суждение по сему предмету...»

В результате этой операции обе группы имеют идентичные средние (9,965234) и равные суммы значений (996,5234) переменной VAR1 в каждой из этих групп. Отличие же их заключается в том, что первая группа имеет нормальное распределение с унимодальной гистограммой, тогда как вторая группа имеет ярко выраженное антимодальное распределение (провал в центре и по одной моде слева и справа). Гистограммы распределения этих групп приведены ниже.



**«... не зная законов языка
ирокезского, можешь ли ты делать
такое суждение по сему предмету...»**

Проведем сравнение этих групп, используя непараметрические статистики. Тест Вальда-Вольфовица указывает на статистически значимое различие двух функций распределения

Mean₁=9,965234 Mean₂=9,965234 p=0,000228

О таком же различии говорят и результаты применения теста Колмогорова-Смирнова в пакете S-PLUS 2000. Two-Sample Kolmogorov-Smirnov Test

data: x: V1 in Nazarenko.SAV , and y: V2 in Nazarenko.SAV

D_{ks} = 0.22, p-value = 0.0156

«... не зная законов языка ирокезского, можешь ли ты делать такое суждение по сему предмету...»

Однако все сравнения средних указывают на их идентичность, т.е. **нулевая гипотеза равенства генеральных средних принимается, при отвержении гипотезы идентичности функций распределения!** Таким образом, утверждение авторов этого доклада о том, что *«...приблизительно у 70% обследованных работников СХК и жителей г. Северска отмечается существенное повышение частоты как хромосомных аберраций нерадиационной природы, так и уровня СХО»* вполне может быть ошибочным! Это может быть любым другим различием, например, различием дисперсий, различием мер формы и т.д.

Что же делать, чтобы избежать ошибок?

Свежесть бывает только одна – первая, она же и последняя.
А если осетрина второй свежести, то это означает, что она тухлая!

М.А. Булгаков.
Мастер и Маргарита.

Что же делать, чтобы избежать ошибок?

Уже в начале 20 века стало ясно, что наука становится всё более междисциплинарной. В особенности это проявляется в биологии и медицине, куда все больше проникают методы точных наук.

Поэтому наиболее радикальным средством избежать ошибок в использовании статистических методов является выполнение этого этапа медицинских исследований профессиональным биостатистиком.

Что же делать, чтобы избежать ошибок?

Медик или биолог, занятый экспериментальными исследованиями, конечно должен владеть основными идеями современной прикладной статистики.

Столь же очевидно, что он должен владеть и терминологией данной предметной области. Это облегчит ему общение с профессиональным статистиком. Все этому должны учить исследователей на этапе аспирантуры и докторантуры.

Что же делать, чтобы избежать ошибок?

Но с другой стороны и профессиональный статистик, выполняющий анализ биомедицинских данных, также должен частично проникать в соответствующую предметную область медика или биолога. Такой симбиоз возможен лишь в том случае, когда общение происходит не эпизодически, а постоянно и является рабочей необходимостью.

Достичь же этого возможно только путем создания специализированных лабораторий биостатистики в медицинских НИИ и вузах. Именно по такому пути и идут медицинские исследовательские центры за рубежом.

Что же делать, чтобы избежать ошибок?

Попытки медика самостоятельно овладеть всем тем огромным инструментарием, который доступен сегодня профессиональному статистику, в конечном счете, обречены на неудачу. Статистика столь же узкоспециальная отрасль знания, как и кардиология, гинекология, психиатрия, физиология и т.д.

Нужно стремиться к созданию двухступенчатой системы организации исследований. Экспериментатор, понимающий идейную сторону теории эксперимента, должен получить возможность постоянно взаимодействовать с консультантом. Возникает необходимость в появлении специалистов нового профиля - статистиков-консультантов.

Что же делать, чтобы избежать ошибок?

Но с другой стороны и профессиональный статистик, выполняющий анализ биомедицинских данных, также должен частично проникать в соответствующую предметную область медика или биолога. Такой симбиоз возможен лишь в том случае, когда общение происходит не эпизодически, а постоянно и является рабочей необходимостью.

Достичь же этого возможно только путем создания специализированных лабораторий биостатистики в медицинских НИИ и вузах. Именно по такому пути и идут медицинские исследовательские центры за рубежом.

Умные предпочитают учиться на ошибках других...

С многочисленными примерами
ошибок использования и описания
статистики в медицине и биологии
вы можете ознакомиться в
разделе КУНСКАМЕРА сайта
БИОМЕТРИКА по адресу
<http://www.biometrica.tomsk.ru/kk.htm>

Умные предпочитают учиться на ошибках других...

Желающие пройти дистанционное обучение по статистическому анализу биомедицинских данных могут познакомиться с условиями обучения по адресу http://www.biometrica.tomsk.ru/edu_1.htm