

**В.А. Кастрикин, В.П. Леонов**

**О ДОСТОВЕРНОСТИ ПУБЛИКУЕМЫХ РЕЗУЛЬТАТОВ  
ИССЛЕДОВАНИЙ**

V.A. KASTRIKIN, V.P. LEONOV. ON VALIDITY OF PUBLISHING  
RESULTS OF STUDIES

Проведен анализ статей академического журнала “Вопросы ихтиологии” за 1997 год. Во многих работах авторы делают грубые ошибки при статистической обработке данных. Результатом этого является сомнительная научная ценность отдельных декларируемых открытий.

Проведенные нами исследования базируются на материалах статей, опубликованных в журнале “Вопросы ихтиологии” за 1997 год. Цель работы заключалась в оценке качества биометрических выводов анализируемых статей. Приводимые нами результаты реанализа исходных авторских публикаций получены с помощью статистических пакетов SAS 6.04 и Statistica 5.1.

В своей работе мы акцентировали внимание на случаях некачественного применения исследователями методов статистического анализа данных и, как следствие, - на сомнительной надежности получаемых выводов.

По нашему мнению, ситуация с использованием биометрики в настоящее время характеризуется общеизвестным выражением: “Есть 3 вида лжи: ложь, наглая ложь и статистика”, вопреки принятому во всем мире положению, что “статистика есть наука, позволяющая принимать оптимальные решения в условиях недостатка информации”. Нередко исследователь забывает о том, что каждый метод имеет свои ограничения. Так, например, наиболее часто применяемый для проверки гипотез о равенстве генеральных средних t-критерий Стьюдента можно использовать лишь в случае, когда принимаются гипотезы: 1) распределения изучаемых величин подчиняются нормальному закону распределения вероятностей и 2) генеральные дисперсии признака в обеих совокупностях равны. В случае невыполнения второго условия возникает так называемая проблема Беренса-Фишера (Бикел, Доксам, 1983), одно из приближенных решений которой реализовано в пакетах Statgraphics и Statistica. Основываясь на справедливости этих гипотез, производится расчет t-критерия Стьюдента во многих учебниках по биометрии. Кроме того, t-критерий Стьюдента очень чувствителен к наличию в наблюдениях аномальных наблюдений, выбросов. Несмотря на то, что перечисленные условия сильно ограничивают правомочность применения критерия в биологических работах, его используют с завидным постоянством, пренебрегая проверкой условий 1) и 2). К чему это приводит, достаточно подробно проанализировано во многих работах (Бащинский, 1998; Леонов, 1998; Леонов, Ижевский, 1998). И хотя существует множество непараметрических критериев для проверки

аналогичных гипотез (Тюрин, Макаров, 1995), случаи их использования в биологических статьях единичны.

Нередки и ошибки, связанные с анализом связи между переменными. Наиболее популярным среди авторов публикаций является коэффициент корреляции Пирсона, призванный оценить силу линейной связи двух переменных. В этом случае также должно выполняться требование нормальности распределения в обеих исследуемых совокупностях, о проверке которого авторы публикаций также ничего не сообщают. Коэффициент корреляции весьма чувствителен к наличию аномальных наблюдений, а их присутствие может сильно исказить результаты анализа.

В качестве “классических” примеров, содержащих такие нарушения, рассмотрим следующие работы.

В одной из работ, проведенных на биологическом факультете МГУ (Касумян, 1997), автор, вследствие допущенной им ошибки при изучении связи между переменными, пришел к выводу о том, что полученные результаты являются научными открытиями. А между тем, из анализа не были исключены наблюдения (в каждом случае по одному), аномальность которых хорошо видна даже визуально на рисунке, приведенном исследователем (Касумян, 1997, с. 82). “Эксперименты ... обнаружили наличие слабой, но достоверной отрицательной связи между рН растворов (0.1 М) и вкусовой привлекательностью 14 органических (карбоновых) кислот ( $r = -0.69$ ;  $P > 0.99$ ). Положительная связь была обнаружена между величиной молекулярной массы кислот и их вкусовой привлекательностью ( $r = 0.59$ ;  $P > 0.95$ )”.

Мы с сомнением отнеслись к результатам статистического анализа этих исследований и решили перепроверить надежность изложенных выводов. Для этой цели по рисунку, на котором изображены экспериментальные точки, были восстановлены значения переменных и повторно проведен корреляционный и регрессионный анализ.

В публикации автор не указывает, какой именно коэффициент корреляции (Пирсона, Спирмена или др.) он использовал. Однако, основываясь на том, что автор для подтверждения своего предположения приводит модель линейной регрессии, а также на том, что полученные нами при реанализе коэффициенты корреляции Спирмена (-0,48 и 0,48), Кендалла (-0,35 и 0,37) и гамма (-0,36 и 0,37) сильно отличаются от имеющихся в статье цифр, мы с большой уверенностью предположили, что в работе была использована линейная корреляция Пирсона. Проведенная нами критериями Шапиро-Уилки и Лиллифорса проверка на нормальность распределения используемых автором публикации переменных показала следующее. В каждой из анализируемых пар наблюдений, один признак с вероятностью  $P_{\text{дов}} > 0,99$  не отвечает законам нормального распределения. Таким образом, применять корреляцию Пирсона нельзя (Афифи, Эйзен, 1982, с.157; Лакин, 1990, с.215).

Как известно, одной из характеристик качества модели регрессии является устойчивость оценок регрессионных коэффициентов при исключении из анализа отдельных наблюдений. На этом свойстве построены оценки аномальности проверяемых наблюдений с помощью таких статистик, как D-

статистика Кука и расстояние Махаланобиса, реализованные, в частности, в статистических пакетах SAS, SPSS и STATISTICA.

Для проверки массивов на наличие выбросов мы воспользовались D-статистикой Кука. В каждой из тестируемых зависимостей было обнаружено по одному наблюдению, для которого дистанция Кука была на порядок больше, чем значение этого показателя для остальных точек (в одном случае  $D_{\max} = 2,5$  против  $0,0 - 0,1$  для оставшейся совокупности, в другом –  $D_{\max} = 1,5$  против  $0,0 - 0,2$  для других наблюдений). Как видно из приведенных результатов, есть все основания характеризовать эти сильно отличающиеся наблюдения как выбросы и рассматривать их отдельно от изучаемых массивов.

После исключения нами из рассматриваемых совокупностей наблюдений, классифицированных как “аномальные”, была проведена проверка нормальности всех 4-х признаков с помощью критериев Шапиро-Уилки и Лиллифорса, которая показала, что данные цензурированной выборки отвечают законам нормального распределения.

Это позволяет нам использовать в дальнейшей работе корреляцию Пирсона и линейный регрессионный анализ. Корреляционный и регрессионный анализ на цензурированных выборках показали, что коэффициенты корреляции изменились с  $-0,69$  и  $0,59$  до  $-0,35$  и  $0,37$  соответственно (сравните с приведенными выше коэффициентами Кендалла и гамма). Достигнутые уровни значимости регрессионных коэффициентов при этом стали равны  $0,25$  и  $0,21$  соответственно, что означает отсутствие как корреляции, так и регрессии. При этом единственно значимыми параметрами в уравнениях линейной регрессии стали константы. Вывод об отсутствии значимости всего уравнения регрессии в целом подтвердили и результаты дисперсионного анализа этого уравнения.

На рисунке мы проиллюстрировали один из случаев.

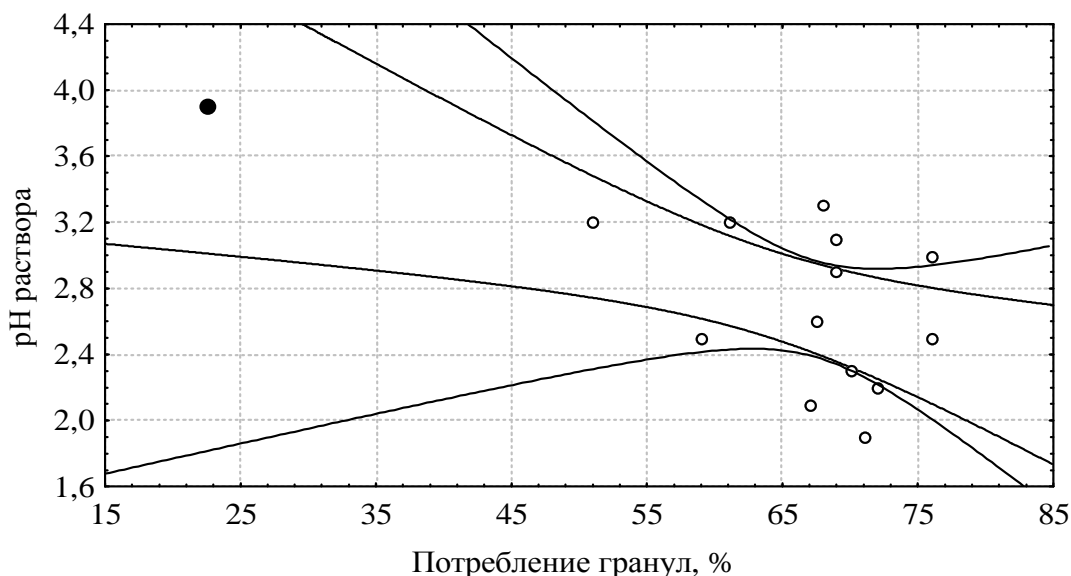


Рис. 95% доверительные интервалы для уравнения линейной регрессии до (пунктиром) и после (сплошная линия) удаления из анализа аномального наблюдения (залитая точка,  $D$  Кука равна  $2,5$ )

Как видно из приведенного рисунка, внутри доверительной зоны, ограниченной пунктирными линиями, можно провести линию, параллельную горизонтальной оси в области значений рН, равных 2,5 - 2,9. Это еще раз подтверждает отсутствие достоверной корреляции и регрессии. Об этом же говорят и полученные по нецензурированным данным непараметрические коэффициенты корреляции.

Таким образом, говорить о связи между изучаемыми явлениями по крайней мере очень и очень сомнительно. Между тем, исследования носят фундаментальный характер и, как утверждает автор, "...представляют важный практический интерес и могут найти применение для решения различных проблем рыболовства и аквакультуры". "Работа выполнена при финансовой поддержке РФФИ (проекты 95-04-11454, 95-04-11754 и 96-04-48218), Международного научного фонда (гранты MV6000 и MV6300) и Европейского научного фонда (грант 1994 г.)". Из рассмотренного примера следует вывод, что грантодатели далеко не всегда контролируют качество научных результатов грантополучателей.

Отсутствие обоснования возможности применения t-критерия Стьюдента можно найти практически в каждом номере рассматриваемого издания. Один из примеров - работа, опубликованная в пятом выпуске (Алексеев, Пичугин, Крысанов. 1997). В таблице на с. 593 авторы приводят результаты проверки гипотез о равенстве генеральных средних 26 признаков в 6 группах, попарно сравнивая наблюдения t-критерием Стьюдента. В публикации допущено множество (более десяти) ошибок первого и второго рода. К примеру, по публикации, гипотеза о равенстве генеральных средних признака  $l_{mх}$  в группах 4 (среднее выборочное равно 36,8; среднее квадратическое - 0,93;  $n=15$ ) и 6 (среднее выборочное равно 36,7; среднее квадратическое - 3,53;  $n=20$ ) отвергается с вероятностью  $P_{дов} > 0,95$ . Проведенные нами (по данным авторов) расчеты показали, что для этого случая гипотеза о равенстве генеральных средних принимается с вероятностью  $P_{дов} > 0,91$ . Так как в описании методик авторы не ссылаются на использование в рассматриваемом нами примере программного обеспечения, можно предположить, что ошибка возникла в результате расчетов, проводимых вручную. Кроме того, вызывает недоумение сам факт использования критерия Стьюдента для задач, на решение которых ориентированы методы дисперсионного анализа.

Нередко в статьях журнала "Вопросы ихтиологии" встречаются работы (Васильева, 1997; Сазонов, 1997), в которых присутствует фраза "выявлены достоверные различия" или близкая к ней по смыслу, но нигде не указывается, каким методом статистического анализа и уровнем значимости пользовался автор. Бывает, что исследователь не делает различий между генеральной средней и выборочной: "...достоверно увеличивает проценты созревания ( $p < 0,01$ ) в 1,34 раза..." (Бурлаков, 1997). Можно продолжить перечисление подобных примеров, но и приведенных, по нашему мнению, вполне достаточно.

Пожалуй, худшим следствием таких работ является возможность дрейфа ошибочно принятой гипотезы из одного исследования в другое и, как результат,

- ситуация, описанная В. Гете в “Фаусте”: “Недостоверна видимость природы в сравнении с данными литературы”.

Из перечисленных выше фактов возникает вопрос: может ли исследователь, неправильно анализирующий результаты эксперимента, грамотно поставить сам эксперимент?

Причины подобных методологических ошибок достаточно подробно описаны (Бащинский, 1998; Леонов, 1998; Леонов, Ижевский, 1998), но можно достаточно уверенно утверждать, что дело не в недостаточности финансирования. Исследования по четырем из пяти приведенных нами примеров финансировались за счет отечественных и иностранных грантов, то есть, авторы работ имели возможность привлечь специалиста по анализу данных, но не сделали этого. Более того, тот уровень анализа, который использовали авторы, отвечает стандартным курсам биометрии, читаемым в отечественных вузах.

### Выводы

Из проведенных нами исследований можно сделать вывод, афористично выраженный в старой русской поговорке: “Доверяй, но проверяй”! Ни статус журнала или учреждения, в котором работает автор статьи, ни ученая степень исследователя, ни звучные имена грантодателей еще не являются гарантией качественно выполненной научной работы. Отмеченные выше недостатки экспериментальной методологии не являются прерогативой только публикаций по ихтиологии, а характерны для статей многих отечественных журналов, диссертаций и монографий биомедицинского профиля. Многочисленные примеры, содержащие подобные ошибки разной степени абсурдности, приведены в “Кунсткамере”. Осознав важность этой проблемы, “...целый ряд научных журналов выступил с инициативой проведения статистического рецензирования статей” (Бащинский, 1998, с.14). Но пока это “в перспективе”, советуем научным сотрудникам критичнее относиться к публикациям в отечественных научных изданиях.

### СПИСОК ЛИТЕРАТУРЫ

*Алексеев С.С., Пичугин М.Ю., Крысанов Ю.Е.* Исследования гольцов *Salvelinus alpinus* (Salmonidae) Забайкалья, внесенных в Красную книгу РСФСР: симпатрические формы из озера Большой Намаракит (морфология, экология, кариология) // *Вопр. ихтиологии.* 1997. Т. 37. Вып. 5. С. 588-602.

*Афифи А., Эйзен С.* Статистический анализ. Подход с использованием ЭВМ. М.: Мир, 1982. 488 с.

*Бащинский С.Е.* Статистика умеет много гитик // *Международный журнал медицинской практики.* 1998. Вып. 4. С.13-15.

*Бикел П., Доксам К.* Математическая статистика. М.: Финансы и статистика, 1983. Вып.1. 254 с.

*Бурлаков А.Б.* Изучение половой специфичности гипофизарных гонадотропинов калкана *Psetta maotica* (Scophthalmidae). V. Особенности действия гонадотропинов in vitro после прохождения гормонозависимого периода // *Вопр. ихтиологии*. 1997. Т. 37. Вып. 6. С. 836-843.

*Васильева Е.Д.* Морфологическая дивергенция двух видов тресковых рыб, *Eleginus navaga* и *E. gracilis* (Gadidae), с дизъюнктивным ареалом // *Вопр. ихтиологии*. 1997. Т. 37. Вып. 6. С. 791-797.

*Касумян А.О.* Вкусовая рецепция и пищевое поведение рыб // *Вопр. ихтиологии*. 1997. Т. 37. Вып. 1. С. 78-93.

"Кунсткамера". <http://www.doktor.ru/doctor/biometr/kk/index.hym>

*Лакин Г.Ф.* Биометрия. М.: Высшая школа, 1990.

*Леонов В.П.* Долгое прощание с лысенковщиной.

<http://www.doktor.ru/doctor/biometr/lib/lis/index.hym>.

*Леонов В.П., Ижевский П.В.* Применение статистики в статьях и диссертациях по медицине и биологии. Часть 1. Описание методов статистического анализа в статьях и диссертациях // *Международный журнал медицинской практики*. 1998. Вып. 4. С. 7-12.

*Сазонов Ю.И.* Новый вид рода *Conocara* (Alerocerphalidae) из Индо-Пацифики // *Вопр. ихтиологии*. 1997. Т. 37. Вып. 6. С. 785-790.

*Тюрин Ю.Н., Макаров А.А.* Анализ данных на компьютере. М.: ИНФРА-М, Финансы и статистика, 1995. 384 с.

Государственный природный  
заповедник «Хинганский»,  
пос. Архара;  
Томский государственный университет

#### SUMMARY

There was conducted the analysis of articles in academic scientific magazines. In many cases the authors of reports made gross errors in data statistical processing. As a result, there was wrong conclusion. There was suggested to make doubt about the articles.