

## ДИСКУССИИ И ОБСУЖДЕНИЯ DISCUSSIONS AND DISPUTATIONS

### Проще простого. Как не надо относиться к статистике

**А.Н. КРИЧЕВЕЦ\***,  
ФГБОУ ВО МГУ имени М.В. Ломоносова, Москва, Россия,  
*ankrich@mail.ru*

*В статье рассматриваются методологические сложности, связанные с распространением статистических способов обработки данных в психологических работах. В частности, происходит упрощение понимания статистики исследователями, что все чаще ведет к принятию гипотезы о наличии эффекта в той ситуации, когда его нет. Приводится краткий обзор публикаций и примеров, где данная проблематика является предметом критического разбора. Упомянуто исследование, которое показало, что при повторении экспериментов, опубликованных в авторитетных журналах, значимость, принятая в научном сообществе, наблюдалась только в 39% работ — такой эффект получил название «публикационный сдвиг». В статье рассматриваются способы решения данной проблемы, в частности метаанализ, благодаря которому возможно подтверждение результатов серии экспериментов, касающихся одной проблематики, при условии доступности информации о дизайне и открытости сырых данных. Понимание вероятностной природы статистических результатов должно стать обязательным для оценки исследований.*

**Ключевые слова:** метаанализ, статистические методы, уровень значимости, публикационный сдвиг.

В середине 90-х гг. прошлого века стал складываться современный способ работы с данными. К этому времени персональные компьютеры были доступны большей части научного сообщества, и

в связи с этим начали распространяться ориентированные на массового пользователя программы, в том числе и программы статистической обработки данных. Доступность статистической обработки

#### Для цитаты:

Кричевец А.Н. Проще простого. Как не надо относиться к статистике // Социальная психология и общество. 2018. Т. 9. № 3. С. 208–212. doi:10.17759/sps.2018090320

\* Кричевец Анатолий Николаевич — доктор философских наук, профессор, ФГБОУ ВО «Московский государственный университет имени М.В. Ломоносова», Москва, Россия, *ankrich@mail.ru*

сопровождалась развитием системы образования в этой сфере, и в настоящий момент для будущих психологов курсы статистики с практикой на компьютерах стали практически повсеместными.

С одной стороны, необходимость понятных формулировок и рецептов обработки данных для тех учащихся в системе массового образования, которым недоступно понимание теоретико-вероятностного и методологического фундамента статистических методов, привела к существенному упрощению «массовой теории».

С другой — процесс превращения статистической обработки данных в массовую деятельность привел к тому, что эффекты, связанные с массовой реализацией случайных испытаний (экспериментов как испытаний случайных искажений закономерных процессов), стали очень заметны и проявились в предсказуемых с точки зрения теории вероятностей последствиях — появлении очень заметного процента результатов, в которых реализовалась ошибка первого рода: принятие гипотезы о наличии эффекта, в то время как реального эффекта нет.

Критика процедур организации массовой работы со статистикой появлялась время от времени в публикациях и прежде, но не приводила к заметным последствиям. Однако в последнее время процесс активизировался. В журналах *Nature* [3] и *Science* [4] вышли работы, в которых статистическое оценивание результатов стало предметом критического разбора. За этим последовало серьезное обсуждение и организационные меры — правда, пока не вполне систематические.

Отмечу здесь посвященный нашему вопросу Бюллетень Американской статистической ассоциации, который начинается с очень точно определяющей

суть вопроса шутки, которую я приведу целиком [5].

Q: Why do so many colleges and grad schools teach  $p = 0,05$ ?

A: Because that's still what the scientific community and journal editors use.

Q: Why do so many people still use  $p = 0,05$ ?

A: Because that's what they were taught in college or grad school.

(Вопрос: Почему так много колледжей и аспирантур учат об уровне значимости  $p = 0,05$ ?

Ответ: Потому что это то, что все еще используется научным сообществом и редакторами журналов.

Вопрос: Почему так много людей использует  $p = 0,05$ ?

Ответ: Потому что их этому научили в колледжах и аспирантурах.)

Здесь отмечен узловой пункт: в упрощенной подаче статистических методов число 0,05 играет существенную роль. Именно благодаря тому, что это число стало *границей принятия решения*, упрощение и стало возможным. В результате мы часто видим в публикациях формулировки: «различие достоверно на уровне значимости 0,05», хотя ни о какой достоверности в данном случае речь идти не может: если искомый эффект отсутствует, то ошибочное принятие решения о его наличии (при традиционной системе принятия решения) будет наблюдаться в среднем один раз из 20 попыток его обнаружения.

Группа, состоящая из двух сотен авторов, деятельность которой была инициирована Американской психологической ассоциацией, отобрала 100 экспериментов, отчеты о которых публиковались в трех весьма авторитетных журналах по психологии, и в контакте с авторами статей повторила эти эксперименты. Ока-

залось, что значимость меньше  $p = 0,05$  наблюдалась при повторениях только в 39% случаев.

Причина такого обескураживающего результата в эффекте, который получил название «публикационный сдвиг». Дело в том, что на поверхности, т. е. опубликованными, оказываются работы, в которых граница  $p = 0,05$  пройдена (в сторону меньшего показателя значимости), а те работы, которые не преодолевают ее, остаются вне видимости, информация о них обычно недоступна. Вследствие этого возможно даже, что некоторые модные направления исследований являются «пузырями», раздутыми благодаря большому количеству экспериментальных «попыток» и публикационному сдвигу, причем чем более модным является направление, тем значительней могут быть искажения.

Естественным способом борьбы с публикационным сдвигом является следующий: должны публиковаться не только результаты удачных экспериментов, которые прошли границу 0,05, но и результаты всех остальных также должны быть доступны в том или ином виде. Только имея полную картину результатов деятельности в области, можно говорить о достоверности.

Процедуры такой обработки результатов давно известны — это так называемый метаанализ. Для того чтобы данное исследование могло быть в него включено, должна быть представлена достаточная информация по дизайну и организации эксперимента. Такие условия предъявляют сейчас в большинстве англоязычных журналов. Замечу, что некоторые журналы сообщают авторам также, что отказ от публикации сырых данных без достаточных оснований, как правило, приведет к отказу в публикации статьи.

Радикальность поворота состоит в изменении статуса отдельного исследования. Мы не занимаемся доказательством гипотез. В нынешнем мире массовых научных процессов мы принимаем участие в массовом движении. Обычно наш вклад ничего не может доказать, а может только несколько усилить метааналитический аргумент в пользу или против гипотезы.

При этом исследователь, адекватный современной ситуации, должен хорошо понимать отличие эксплораторной от конфирматорной деятельности: искать в уже полученных данных какие-то закономерности (эксплораторная стадия) можно, но ценность результата тем ниже, чем больше различных попыток получения статистически значимых результатов по этим данным проделано. Совсем другой статус имеют исследования, авторы которых заранее заявили о плане исследования и способах обработки результатов. Различные типы таких предварительных заявок приняты в некоторых журналах, в том числе в одном отечественном («Теоретическая и экспериментальная психология»). Существует две формы предварительной регистрации, которые практикуются в научных журналах в настоящий момент. Жесткая форма состоит в подаче проекта исследования, который сразу рецензируется. Если он получил высокую оценку от рецензентов, то публикация гарантируется независимо от результата. В более мягкой форме проект фиксируется без рецензирования, далее проводится исследование, и если статья по его результатам признается ценной, то публикация в журнале маркируется специальным уведомлением о том, что проект исследования был представлен заранее. Все *post hoc* интерпретации данных авторы статьи должны в этом случае отмечать как таковые.

Далее, метааналитическое подтверждение серии результатов предпочтительнее подтверждения единственного эксперимента при равенстве характеризующей их значимости, поскольку массовое воспроизведение свидетельствует об отсутствии ошибок дизайна и подбора испытуемых, а также и просто недобросовестности.

Уровень значимости 0,05 — это всего лишь мягкая и достаточно произвольная граница, а не критерий принятия решения. Значимость характеризует *степень* подтверждения гипотезы. Нам нет необходимости принимать окончательное решение относительно наших гипотез, и публикация не должна заканчиваться словами о том, что мы что-то доказали. Мы получили маркированное уровнем значимости подтверждение гипотезы, не более того. Если тема действительно интересна, то надо всячески поощрять повторение экспериментов с публикацией результатов независимо от статистической значимости. Следовало бы,

например, считать воспроизведение в курсовых и дипломных работах уже опубликованных экспериментов полноценной квалификационной работой.

Несмотря на превращение научной деятельности в массовую, упрощение преподаваемых оснований статистических методов не должно переходить определенных границ. Понимание вероятностной природы статистических результатов должно стать обязательным для квалифицированной части сообщества, поскольку, как мы видим, дальнейшее упрощение приводит к тому, что даже эксперты, оценивающие наши работы, оказываются недостаточно образованными.

Замечу в заключение, что вопрос о том, что считать интересным результатом, что публиковать в самых престижных журналах, за что давать Нобелевские премии, нуждается в свете сказанного в новом обсуждении. Я уверен, что сообщество с ним успешно справится.

Изложение деталей проблемы можно найти в публикациях [1; 2].

## ЛИТЕРАТУРА

1. Корнеев А.А., Рассказова Е.И., Кричевец А.Н., Койфман А.Я. Критика методологии проверки нулевой гипотезы: ограничения и возможные пути выхода. Часть I. [Электронный ресурс] // Психологические исследования. 2016. Том 9. № 45, 1. URL: <http://psystudy.ru/index.php/num/2016v9n45/1231-korneev45.html> (дата обращения: 16.06.2018).
2. Корнеев А.А., Рассказова Е.И., Кричевец А.Н., Койфман А.Я. Критика методологии проверки нулевой гипотезы: ограничения и возможные пути выхода. Часть II. Психологические исследования. 2016. Том 9. № 47, 6. URL: <http://psystudy.ru/index.php/num/2016v9n47/1282-korneev47.html> (дата обращения: 16.06.2018).
3. Nuzzo R. Statistical errors // Nature. 2014. Vol. 506. №. 7487. P. 150–152. doi:10.1038/506150a
4. Open Science Collaboration. Estimating the reproducibility of psychological science // Science. 2015. Vol. 349. № 6251. aac4716-1. doi: 10.1126/science.aac4716
5. Wasserstein R.L., Lazar N.A. The ASA's statement on p-values: context, process, and purpose. The American Statistician. doi: 10.1080/00031305.2016.1154108

## Easy as pie. How not to treat statistics

A.N. KRICHEVETS\*,  
Lomonosov Moscow State University, Moscow, Russia,  
ankrich@mail.ru

*The article discusses the methodological difficulties associated with the spread of statistical methods of data processing in psychological works. In particular, there is a simplification of the understanding of statistics by researchers, which increasingly leads to the adoption of the hypothesis of the presence of the effect, in a situation where it is not. There is a brief review of publications and examples where this issue is the subject of critical analysis. Presented a research which showed that in the repetition of experiments published in reputable journals, the significance accepted in the scientific community was observed only in 39% of the works — this effect was called the “publication shift”. The article discusses ways to solve this problem, in particular, meta-analysis, through which it is possible to confirm the results of a series of experiments relating to one problem — subject to the availability of information about the design and openness of raw data. An understanding of the probabilistic nature of statistical results must be mandatory for the evaluation of studies.*

**Keywords:** meta-analysis, statistical methods, significance level, publication shift.

### REFERENCES

1. Korneev A.A., Rasskazova E.I., Krichevec A.N., Kojfman A.YA. Kritika metodologii proverki nulevoj gipotezy: ogranicheniya i vozmozhnye puti vyhoda. CHast' I. [Elektronnyy resurs] [Criticism of the methodology for testing the null hypothesis: limitations and possible ways out. Part I.]. *Psihologicheskie issledovaniya [Psychological research]*, 2016. Vol. 9, no. 45, 1. URL: <http://psystudy.ru/index.php/num/2016v9n45/1231-korneev45.html> (Accessed 16.06.2018).
2. Korneev A.A., Rasskazova E.I., Krichevec A.N., Kojfman A.YA. Kritika metodologii proverki nulevoj gipotezy: ogranicheniya i vozmozhnye puti vyhoda. CHast' II. [Criticism of the methodology for testing the null hypothesis: limitations and possible ways out. Part II.]. *Psihologicheskie issledovaniya [Psychological research]*, 2016. Vol. 9, no. 47, 6. URL: <http://psystudy.ru/index.php/num/2016v9n47/1282-korneev47.html> (Accessed 16.06.2018).
3. Nuzzo R. Statistical errors. *Nature*, 2014. Vol. 506, no. 13, pp. 150–152. doi:10.1038/506150a
4. Open Science Collaboration. Estimating the reproducibility of psychological science. *Science*, 2015. Vol. 349, ISSUE 6251 aac4716-1. doi: 10.1126/science.aac4716
5. Wasserstein R.L. Lazar N.A. The ASA's statement on p-values: context, process, and purpose. *The American Statistician*, 2016. doi: 10.1080/00031305.2016.1154108

#### For citation:

Krichevets A.N. Easy as pie. How not to treat statistics. *Sotsial'naya psikhologiya i obshchestvo [Social Psychology and Society]*, 2018. Vol. 9, no. 3, pp. —212. doi:10.17759/sps.2018090320 (In Russ., abstr. in Engl.).

\* *Krichevets Anatolii N.* — PhD in Philosophy, Professor, Lomonosov Moscow State University, Moscow, Russia, [ankrich@mail.ru](mailto:ankrich@mail.ru)