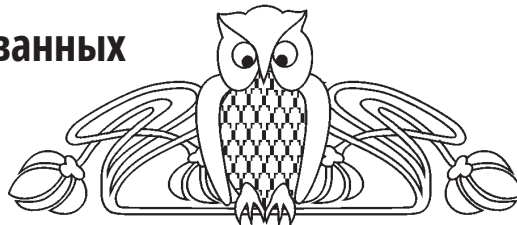




Известия Саратовского университета. Новая серия. Серия: Социология. Политология. 2024. Т. 24, вып. 4. С. 389–398
Izvestiya of Saratov University. Sociology. Politics, 2024, vol. 24, iss. 4, pp. 389–398
<https://soziopolit.sgu.ru> <https://doi.org/10.18500/1818-9601-2024-24-4-389-398>, EDN: GKTLGY

Научная статья
УДК 51-77:303.1

Применение искусственно сгенерированных данных при разработке и внедрении математико-статистических методов в проблемное поле социологии



А. А. Звонок

Луганский государственный педагогический университет, Россия, 291011, г. Луганск, ул. Оборонная, д. 2

Звонок Александр Анатольевич, кандидат философских наук, доцент кафедры социальной педагогики и организации работы с молодежью, al.zvonok@gmail.com, <https://orcid.org/0009-0007-7332-1330>

Аннотация. Статья затрагивает общие вопросы, касающиеся использования искусственно сгенерированных данных в процессе разработки и внедрения математико-статистических методов в социологическую сферу. Кратко охарактеризованы основные направления применения симулированных данных в социологии. Очерчена зона ответственности специалистов-социологов по отношению к математикам при верификации математико-статистических методов и их интеграции в социологические научные отрасли. Дана авторская классификация математико-статистических методов в зависимости от необходимости применять искусственные данные для их верификации. Объяснена классификация симуляций при верификации математико-статистических методов по степени значимости научных проектов, при этом к каждому классу приложен реальный кейс в рамках социологической проблематики. В качестве проекта низкой степени значимости приведен эксперимент с частотными доверительными интервалами А. Крыштановского, использующий искусственные выборки данных социологических опросов. В качестве образца научного кейса средней значимости выступило исследование Дж. К. Ф. де Винтера и Д. Додоу перспектив применения t-критерия Стьюдента для анализа выборок наблюдений, выраженных в порядковой шкале Ликерта. Примером научной симуляции высокого уровня стал собственный опыт автора, связанный с внедрением байесовской методологии в эмпирическую социологию в контексте разработки методов анализа размера эффекта при проведении сравнительных биномиальных социальных экспериментов с бинарными данными. В исследовании поднят вопрос о требованиях к публикации методологических исследований, использующих сгенерированные данные. Приводятся стандарты публикации исследований с симуляциями, начиная от классических соглашений общего профиля и заканчивая жесткими стандартами, принятыми в таких ответственных сферах, как медицинские исследования.

Ключевые слова: сгенерированные данные, симулированные данные, математико-статистические методы, верификация методов, анализ социологических данных, эмпирическая социология

Для цитирования: Звонок А. А. Применение искусственно сгенерированных данных при разработке и внедрении математико-статистических методов в проблемное поле социологии // Известия Саратовского университета. Новая серия. Серия: Социология. Политология. 2024. Т. 24, вып. 4. С. 389–398. <https://doi.org/10.18500/1818-9601-2024-24-4-389-398>, EDN: GKTLGY

Статья опубликована на условиях лицензии Creative Commons Attribution 4.0 International (CC-BY 4.0)

Article

Application of artificially generated data in development and implementation of mathematical-statistical methods in the problem field of sociology

А. А. Zvonok

Lugansk State Pedagogical University, 2 Oboronnaya St., Lugansk 291011, Russia

Aleksandr A. Zvonok, al.zvonok@gmail.com, <https://orcid.org/0009-0007-7332-1330>

Abstract. The article touches on the general issues, related to the use of artificially generated data in the process of development and implementation of mathematical-statistical methods in the sociological sphere. The main directions of application of simulations in sociology are briefly described. The area of responsibility of sociologists in relation to mathematicians in the verification of mathematical-statistical methods and their integration into sociological scientific fields is outlined. The author's classification of mathematical and statistical methods is given depending on the need to use artificial data for their verification. The classification of simulations in the verification of mathematical-statistical methods according to the degree of significance of scientific projects is explained, while each class was accompanied by a real case within the framework of sociological problems. The experiment with frequency confidence intervals by A. Kryshchanovsky, using artificial samples of sociological survey data, was described as a project of low significance. The example of a scientific case of moderate significance was the study by J. C. F. de Winter and D.



Dodou of the prospects for using Student's t-test to analyze samples of observations, expressed in the ordinal Likert scale. The example of high-level scientific simulation was the author's own experience, associated with the introduction of Bayesian methodology into empirical sociology in the context of developing methods for analyzing effect sizes when conducting comparative binomial social experiments with binary data. The study raises questions on the requirements for publishing methodological studies using the generated data. The standards for publishing simulation studies are outlined, ranging from classical general conventions to the important standards, adopted in sensitive areas such as medical research.

Keywords: generated data, simulated data, mathematical-statistical methods, method verification, analysis of sociological data, empirical sociology

For citation: Zvonok A. A. Application of artificially generated data in development and implementation of mathematical-statistical methods in the problem field of sociology. *Izvestiya of Saratov University. Sociology. Politology*, 2024, vol. 24, iss. 4, pp. 389–398 (in Russian). <https://doi.org/10.18500/1818-9601-2024-24-4-389-398>, EDN: GKTLGY

This is an open access distributed under the terms of Creative Commons Attribution 4.0 International License (CC-BY 4.0)

Искусственно сгенерированные (симулированные) данные в рамках социологической проблематики сегодня рассматриваются в двух важных контекстах. Прежде всего, это компьютерное моделирование (симуляция) социальных процессов и ситуаций на основе искусственных данных. Н. Гилберт и К. Г. Троиц датируют возникновение первых подобных симуляций в социальных науках еще 1960-ми гг. и выделяют до семи разновидностей таких моделей: от простейших микросимуляций до сложных самообучающихся систем [1, р. 6–14]. Хотя в отечественной социологии этой области, на наш взгляд, уделяется недостаточное внимание, все же данная тема иногда поднимается в социологических журналах (намного чаще встречаясь в журналах, посвященных математическим и компьютерным наукам). Причины данного противоречия, заключающиеся в субъективных установках отечественных социологов в отношении сложных математических методов, достаточно подробно рассматриваются Ю. Н. Толстой и выходят за рамки настоящего исследования [2]. Другое важное направление применения искусственных данных, которое почти никогда не становилось предметом серьезного обсуждения в социологическом сообществе и которому посвящено наше исследование, – это использование сгенерированных данных при разработке и внедрении математико-статистических методов в социологическую проблематику. Сюда входят как задачи тестирования какого-либо общеизвестного метода (методики, критерия) из математики или другой прикладной научной области на предмет того, насколько он пригоден для конкретной социологической задачи, так и разработка принципиально новых методов анализа, сопровождающаяся их верификацией в контексте социологических кейсов.

На первый взгляд, означенная сфера деятельности лежит за пределами компетенций социологов, относясь к зоне ответственности математиков, искушенных в статистических методах и сведущих в тонкостях математиче-

ского аппарата. Однако математики, как правило, работают с абстрактными числами, не учитывающими реального контекста и особенностей ситуаций, в которых будут применяться разработанные ими методы. И даже специалисты по прикладной математике, призвание которой заключается во внедрении математической методологии в другие области науки и техники, сталкиваются с нехваткой знаний о предметных областях, в которые, по задумке, должны интегрироваться разработанные методы и методики. Таким образом, проблема верификации математико-статистических методов в социологии сегодня находится на стыке социологических и математических наук: для ее решения социологи должны обладать достаточной математической подготовкой, а математики – достаточным объемом знаний о социуме в той степени, которая необходима для эффективного внедрения математико-статистических методов в проблемное поле социологии. Наконец, если подходить к данному вопросу в российской социологии с позиции сухого формализма, текущий паспорт научной специальности ВАК РФ 5.4.1. «Теория, методология и история социологии» прямо содержит в себе такие направления социологических исследований, как «Разработка новых методов математико-статистического анализа эмпирических данных», «Использование приемов и техник анализа, заимствуемых из других дисциплин, в применении к социальным данным» и «Методологические исследования, направленные на совершенствование, развитие, уточнение традиционных, широко применяемых и новых методов, методик, техник сбора и анализа эмпирических данных»¹. Цель текущей работы – раскрыть место и значение искусственно сгенерированных данных в процессе научной работы над вышеозначенными проблемами, продемонстри-

¹ Паспорт научной специальности 5.4.1 «Теория, методология и история социологии» // Высшая аттестационная комиссия при Министерстве науки и высшего образования Российской Федерации. URL: <https://vak.minobrnauki.gov.ru/uploader/loader?type=17&name=92259542002&f=15331> (дата обращения: 18.05.2024).



ровав, что зачастую их решение невозможно без применения симулированных данных.

Контент-анализ М. Харвела, Н. Кохли и Я. Перальта-Торрес шести ведущих международных методологических журналов, включающих такие отрасли, как математическая статистика, психология, биология и технические науки, показал, что за период с 1985 по 2012 гг. в 34,6% научных исследований, опубликованных в данных журналах, использовались искусственно сгенерированные данные [3]. На заре возникновения исследований со сгенерированными данными в связи с развитием компьютерных технологий отношение в научном сообществе к ним было неоднозначным (причиной этому, прежде всего, были слабые возможности первых компьютеров и низкое качество первых генераторов случайных чисел). Однако уже в 1984 г. В. В. Хоук и Ш. Андерсон предложили считать исследования с применением искусственно сгенерированных данных экспериментами и именно так трактовать их со стороны как авторов исследований, так и редакторов научных журналов [4, р. 215]. Современное вычислительное экспериментирование с помощью компьютеров получило признание даже в чистой математике, значительно отличаясь от классической концепции математических доказательств, в которой математический эксперимент рассматривался лишь как предшественник доказательства, и давая эпистемологический статус утверждениям, полученным на основе экспериментальных результатов [5, с. 11]. В прикладных научных областях вопрос о целесообразности использования моделирования с помощью искусственных данных не оспаривался вовсе: сегодня все ответственные научные отрасли, начиная с ядерной физики и заканчивая разработкой перспективных направлений медицинских исследований, применяют верификацию использующихся математико-статистических методов и методик посредством симулированных данных. Данную ситуацию емко описывает широко известное высказывание П. Талла, одного из ведущих методологов прецизионной медицины: «Я всегда считал, что при калибровке расчетных параметров более желательно убивать пациентов, созданных компьютером, нежели реальных» [6, р. 25].

Однако возникает закономерный вопрос на тему того, что считать верификацией математико-статистических методов конкретно в области социологических исследований. Мы считаем, что в социологическом контексте понятие «верификация» никак не противоречит своей традиционной трактовке в естественных

науках, принятой в научно-исследовательских лабораториях. Так, М. Киричевский и С. Комаров определяют верификацию (введение, внедрение) методов и методик измерений (исследований) как процесс подтверждения их соответствия требованиям получения при их использовании достоверной информации при внедрении этих методик или методов в свою практику (под достоверной информацией здесь следует понимать предоставление результатов требуемого уровня качества) [7, с. 263].

Другой важный вопрос заключается в том, действительно ли необходимо применение искусственно сгенерированных данных при указанной верификации конкретно в социологии. В математике, работающей с абстрактными числами, или в медицине, где ошибка в математических расчетах может привести к потере человеческой жизни, ответы на этот вопрос обычно утвердительные. Однако социология, родившись из позитивизма, всегда испытывала тягу к реальным данным и полевой работе, опиралась на практическое изучение общества. Современные технологии обладают огромными возможностями по созданию искусственных данных различного уровня качества и сложности. Рядовые статистические программы способны генерировать простые числа, подчиняющиеся заданным законам распределения вероятностей. Более сложные статистические пакеты из науко-ориентированных языков программирования, таких как R и Python, могут создавать связанные наборы чисел с требуемыми значениями корреляции. Наиболее продвинутые научные библиотеки, такие как scikit-learn, способны создавать многомерные данные с указанными уровнями кластеризации. Хотя современные инструменты и предлагают неисчерпаемое количество возможностей по симулированию данных любой сложности, на наш взгляд, все кейсы верификации математико-статистических методов можно условно разделить на две группы в зависимости от доступности достоверных значений анализируемых параметров для сравнения их с выходными значениями этих методов.

1. Успешная верификация посредством реальных данных возможна в ситуациях, когда исследователь может получить из них достоверную информацию об истинных значениях выходных параметров, продуцируемых тестируемыми математико-статистическими методами. Например, верификация разрабатываемых методов и моделей на основе реальных данных в задачах прогнозирования и классификации, как правило, не представляет особых проблем.



Предположим, нам нужно построить на основе имеющихся реальных наблюдений логистическую регрессионную модель, предсказывающую вероятность того, что студент-социолог сдаст экзамен по математической статистике с первого раза. В процессе верификации построенной модели мы можем разделить изначальную выборку наблюдений на обучающую и проверочную (тестовую) выборку. Первая будет использована для построения модели, факторы (независимые переменные) второй выборки будут загружаться в построенную модель, чтобы получить вероятность бинарной зависимой переменной, которая затем будет сравниваться с известным истинным значением («сдал» или «не сдал»). Таким образом, мы достоверно знаем правильный ответ и можем сверить его с ответом построенной модели.

2. Эффективная верификация с помощью реальных данных невозможна в ситуациях, когда нельзя получить достоверные значения параметров и сравнить их с выходными значениями, производимыми математико-статистическими методами. Предположим, мы методом попарного отбора сформировали две выборки, в одной из которых студенты были записаны на курс развития когнитивных способностей, а в другой – испытуемые продолжали жить своей обычной студенческой жизнью. Через полгода студенты в обеих группах поучаствовали в сдаче экзамена по математической статистике, по итогам которого было подсчитано, сколько студентов в каждой из групп сдали экзамен с первого раза, чтобы проверить, повлиял ли пройденный курс на успеваемость студентов. Это простейшая репрезентация классического эксперимента с контрольной и экспериментальной группами, но нужно понимать, что даже в максимально контролируемом натурном лабораторном эксперименте всегда остается вероятность влияния скрытых неучтенных сторонних факторов, которые могли исказить результаты. Это могут быть, например, неполадки и погрешности в измерительных приборах, смещения в выборках и т.д. И если мы проводим верификацию новой методики сравнения выборок, то у нас попросту отсутствуют достоверные истинные значения, с которыми мы должны сравнивать выходные значения этой методики. В таких случаях только применение искусственных выборок, сгенерированных с известными истинными параметрами генеральной совокупности, обеспечит достоверное информирование о том, насколько выходные значения, получаемые после применения метода или методики, отклоняются от номинальных значений, заложенных в данных.

Кроме того, очевидно, что все методологические проблемы, решаемые с помощью генерации искусственных данных, будут иметь различные уровни сложности, и от этого будут зависеть требования к компетенциям исследователя. М. О'Келли, В. Анисимов, К. Кемпбелл и С. Гамильтон предлагают разделять все ситуации моделирования и симуляции (в том числе в целях верификации математико-статистических моделей и методов) на исследовательские проекты низкой, средней и высокой степени значимости. Научные проекты низкой степени значимости тестируют стандартное поведение уже известной формулы или методики сгенерированными данными (например, проверяя, соответствует ли реальная точность метода запланированной). Проекты средней значимости не только проверяют стандартное поведение метода, но и могут использоваться для обоснования принятия решений по использованию верифицируемого метода в конкретной практической ситуации (не всегда стандартной для этого метода). Проекты высокой значимости используют сгенерированные данные для разработки принципиально новых моделей и методов, ранее не применяемых на практике, включая в себя задачи, решаемые на уровне проектов низкой и средней значимости [8, р. 111–112]. Далее мы проиллюстрируем каждое из выделенных направлений симулированных методологических исследований подробным релевантным примером из социологии.

В качестве социологического кейса низкой степени значимости можно взять недооцененный математический эксперимент А. Крыштановского, проведенный в 2000-х гг. для иллюстрации проблемы доверительных интервалов в частотной статистике и их влияния на точность социологических опросов. Строго говоря, данный эксперимент проводился не с «чистыми» искусственными данными, а с массивом данных, полученным после объединения нескольких реальных опросов ВЦИОМ (но эксперимент очень легко воспроизвести и на полностью симулированных данных). Из искусственно созданной таким образом генеральной совокупности в 40 000 человек были отобраны 100 простых случайных выборок по 500 респондентов, имитируя различные выборочные стратегии. Затем полученные данные в выборках были сопоставлены с генеральной совокупностью. Оценки распределения ответов на пять вопросов анкеты должны были лежать в рамках доверительных интервалов (в эксперименте установлен «золотой стандарт» социологии – 95%). В случае, когда оценивался только один вари-



ант ответа в вопросе – из 100 экспериментальных выборок только не более чем в 5% случаев ответы вышли за границы 95% доверительного интервала, подтверждая эпистемологические основы частотной статистики. Однако когда в вопросе оценивалась не одна градация, а все варианты ответов вместе, в 26 выборках из 100 хотя бы один из вариантов выходил за границы доверительного интервала, означая ошибку вопроса в целом. При рассмотрении всех пяти вопросов анкеты в совокупности выход за пределы доверительных интервалов был зафиксирован уже в 58% выборок [9, с. 678]. Фактически данная проблема воспроизводимости результатов исследований в частотной статистике была известна с ее создания и лишь иллюстрирует тот факт, что данная парадигма изначально создавалась Р. Фишером для проведения экспериментов, а не для комплексных социологических опросов. Для социолога крайне полезно знать возможности и ограничения используемых им подходов и методов, в том числе применяя подопытные эксперименты в своей работе.

Следующий эксперимент с искусственно сгенерированными данными, будучи примером методологической верификации средней степени значимости, относится к одной из самой популярных шкал, используемых в социологии – 5-балльной шкале Ликерта. Исследование Дж. К. Ф. де Винтера и Д. Додоу, проведенное в 2010 г., было посвящено сравнительному анализу эффективности параметрического t-критерия Стьюдента и непараметрического U-критерия Манна-Уитни в контексте сравнения выборок наблюдений, представленных в данной шкале. Будучи фактически порядковой, шкала Ликерта часто трактуется исследователями как интервальная для применения в ее отношении параметрических методов анализа. Однако это не является стандартной ситуацией применения, так как параметрические статистические критерии никогда не предназначались для данных, которые не подчиняются нормальному распределению. В рамках математического эксперимента его авторами было сгенерировано 98 пар искусственных выборок, симулирующих различные ситуации (одинаковый и неравный размер сравниваемых выборок, сравнение выборок малых, средних и больших объемов и т.д.). Полученные результаты доказывают, что оба статистических критерия различий практически эквивалентны в плане применимости в отношении тестируемой шкалы (в наихудших сценариях разница в их эффективности не превышала 10%), даже если у исследователя и нет формальных оснований использовать t-критерий Стьюдента [10]. На

самом деле данное исследование помимо своей основной цели иллюстрирует еще и тот факт, что у непараметрических методов имеется множество проблем и ограничений, вынуждающих аналитиков прибегать к параметрическим методам, даже если это нарушает эпистемологические основания частотной статистики. Непараметрическая статистика часто решает исследовательские задачи нестандартными способами, которые затрудняют интерпретацию получаемых результатов (например, проверяя равенство медиан в нескольких выборках вместо их средних значений). К сожалению, этой проблеме редко уделяется значительное внимание за пределами математического сообщества.

Наконец, в качестве самого сложного кейса – применения сгенерированных данных при верификации новых математико-статистических методов в социологии – мы возьмем наш собственный опыт, полученный в процессе исследования методологических возможностей байесовского подхода для исследований в рамках социологической проблематики. Сегодня одним из наиболее перспективных ответвлений байесовского анализа является синтез байесианства и классической концепции «анализа мощности» Дж. Коэна из частотной статистики, в результате которого исследователь может отказаться от выдвижения и проверки статистических гипотез в пользу прямой апостериорной оценки размера эффекта, оказываемого экспериментальным фактором в исследовании. Ранее нами были продемонстрированы методологические возможности байесовской модели BEST (Bayesian Estimation Supersedes the t Test) Дж. К. Крушке для экспериментальной оценки социальных технологий на основе байесовской реализации размера эффекта d Коэна, значение которого показывает, на сколько стандартных отклонений изменяется значение искомого параметра в одной из выборок в сравнении с другой выборкой под воздействием экспериментального вмешательства [11]. Значение d Коэна на сегодня, возможно, является самым известным и применяемым методом оценки размера эффекта в двух выборках для переменных, выраженных в количественных шкалах, и интерпретируется следующим образом: 0,2–0,5 – «слабый», 0,5–0,8 – «средний», более 0,8 – «сильный» размер эффекта. Однако при этом данный метод – всего лишь один из многих, предложенных Дж. Коэном для разных ситуаций. Например, для вычисления размера эффекта в биномиальных экспериментах с двумя выборками (основанных на бинарных переменных формата «0–1» или «успех – неудача»)



Дж. Коэн предлагает соответствующий параметр h Коэна, интерпретация значений которого идентична значениям d Коэна [12, р. 179–213].

При разработке собственной байесовской модели сравнительных биномиальных экспериментов для применения ее в социологическом контексте мы столкнулись в ряду противоречий, вызванных проблемами поведения математико-статистических методов традиционной статистики при попытке их прямого переноса в область байесовской методологии. С итоговой авторской моделью, получившейся после решения всех возникших методологических проблем, и особенностями ее применения можно ознакомиться в [13].

Одним из наиболее популярных кейсов биномиального социального эксперимента в цифровой социологии и цифровом маркетинге является сравнительное исследование онлайн-контента (так называемое А/В-тестирование), при котором количество отметок «нравится» (или других целевых действий пользователей) представляется в виде биномиальной переменной. Так, количество отметок соответствует числу успехов, а разница между общим количеством просмотров контента и количеством отметок является числом неудач (иначе говоря, пропорция успехов в данном случае является долей пользователей, совершивших целевое действие, из общего количества пользователей, просмотревших контент). Таким образом, различные единицы онлайн-контента предстают в виде «выборок» пользователей с пропорциями успехов и неудач, а суть эксперимента заключается в выяснении того, какой контент обладает большей социальной эффективностью (например, который вызывает больший отклик со стороны аудитории или которым чаще делятся с другими пользователями). В цифровой среде подобные метрики, как правило, имеют достаточно низкие значения в пределах нескольких процентов или даже долей процента: тысяча «лайков» у поста со ста тысячами просмотров в канале мессенджера «Телеграм» или в социальной сети «ВКонтакте» является нормальным явлением, более того, это давно известная особенность экспериментального дизайна А/В-тестирования в цифровой среде [14, р. 15].

Однако при прямой попытке байесовской реализации параметра h Коэна в авторской модели сравнительного биномиального эксперимента по аналогии с подходом, выполненным в модели BEST Дж. К. Крушке с помощью d Коэна, нами было обнаружено странное поведение параметра при очень низких абсолютных

значениях пропорций успехов: апостериорное значение размера эффекта h Коэна оказывалось слишком низким относительно реального эффекта. Для верификации параметра нами с помощью библиотеки РумС языка программирования Python было сгенерировано три набора данных («датасета»), каждый из которых содержал две независимых выборки из 100 000 наблюдений, взятых из распределения Бернулли, причем в каждом из наборов истинное значение пропорции «успехов» в «экспериментальной» выборке представляло собой удвоенное значение «контрольной» выборки ($P_{1-1} = 0,5$, $P_{1-2} = 0,25$; $P_{2-1} = 0,05$, $P_{2-2} = 0,025$; $P_{3-1} = 0,005$, $P_{3-2} = 0,0025$). При генерации применялось начальное значение случайности («random seed») 356757, таким образом, любой исследователь, знакомый с библиотекой РумС и знающий исходные параметры, может в точности воспроизвести наборы данных на своем компьютере. При помощи сгенерированных данных нами тестировалось поведение двух собственных байесовских реализаций методов измерения размера экспериментального эффекта для сравнения выборок биномиальных данных.

Первый вариант – это точная реализация формулы h Коэна из оригинальной работы Дж. Коэна «Статистический анализ мощности для поведенческих наук» (следует заметить, что в байесовском варианте арифметические операции в формуле производятся над апостериорными вероятностными распределениями, что требует использования достаточно высокоуровневых библиотек языков программирования R или Python и значительных вычислительных мощностей компьютера).

Второй вариант – это прямое изменение пропорции доли успехов в «экспериментальной» выборке относительно значения «контрольной». Если ранее упоминавшийся показатель d Коэна для количественных переменных показывал, на сколько стандартных отклонений изменялось значение параметра в одной из выборок, то этот метод показывает, на сколько значений «контрольной» выборки изменяется значение пропорции успехов в «экспериментальной» под воздействием экспериментального фактора. То есть если значение этого показателя 0,25 – это означает, что число успехов в «экспериментальной» выборке на 25% больше, чем в «контрольной», а если значение показателя $-0,15$, то число успехов в «экспериментальной» выборке на 15% меньше, чем в «контрольной».

Результаты симуляции представлены в таблице. При рассмотрении результатов следует обратить внимание на то, что байесовские апо-



Результаты верификации двух методов измерения размера экспериментального эффекта для сравнения выборок биномиальных данных в байесовской реализации с применением искусственно сгенерированных данных

Истинные значения доли успехов в «экспериментальной» / «контрольной» выборках, %	Метод измерения размера экспериментального эффекта			
	h Коэна		Прямое изменение пропорции доли успехов в «экспериментальной» выборке относительно значения «контрольной» выборки	
	95%-ный интервал плотности апостериорного распределения	Среднее значение	95%-ный интервал плотности апостериорного распределения	Среднее значение
50 / 25	(0,512; 0,529)	0,521	(0,967; 1,016)	0,992
5,0 / 2,5	(0,128; 0,146)	0,137	(0,942; 1,134)	1,038
0,5 / 0,25	(0,034; 0,052)	0,043	(0,737; 1,349)	1,038

стериорные оценки значений параметров всегда интервальные, средние значения в таблице приводятся лишь в качестве вспомогательных. За подробным разъяснением особенностей практического применения байесовского подхода читателю предлагается обратиться к ранее упоминаемой статье автора «Байесовское моделирование биномиальных экспериментов в социологии: проблемный анализ», также читатель, владеющий основами научного программирования на языке Python, может самостоятельно проверить поведение тестируемых параметров с помощью кода и данных, доступных в репозитории автора данной работы².

Как мы можем видеть, в процессе уменьшения абсолютных значений доли успехов в «экспериментальной» и «контрольной» выборках сгенерированных данных получаемые значения h Коэна стремительно уменьшаются. При увеличении пропорции с 25 до 50% в первом наборе данных h Коэна ожидаемо показывает «средний» размер экспериментального эффекта, однако при соответствующем увеличении пропорции с 2,5 до 5,0% во втором наборе данных и с 0,25 до 0,5% в третьем метод дает значения, которые следует интерпретировать как «эффект практически отсутствует». В то же время второй метод корректно отображает истинную ситуацию – двойной рост пропорции успехов в наших симулированных биномиальных экспериментах. Таким образом, именно последний метод был включен в окончательное исследование в качестве рабочего (хотя описанные здесь научные изыскания и не вошли в итоговый текст рукописи из-за ограничений на объем публикации). Конечно же, очевидный не-

достаток выбранного метода – необходимость каждый раз интерпретировать значение размера эффекта в зависимости от конкретного исследовательского кейса: для одних ситуаций увеличение исследуемого параметра в экспериментальной группе в сравнении с контрольной на 10% будет являться «слабым» эффектом, для других ситуаций такое же увеличение можно считать «средним» или даже «сильным» эффектом. Однако, как убедительно показывает исследование Ш. Савиловски, даже классические интерпретации размера эффекта Дж. Коэна в настоящее время существенно реинтерпретируются, в том числе в рамках частотной статистики [15].

На самом деле вышеописанный пример использования сгенерированных данных позволяет нам понять следующие важные аспекты верификации математико-статистических методов при их разработке и внедрении в конкретные прикладные научные области, такие как социология.

1. Слепое копирование методов и методик из одной области применения в другую без их предварительного тестирования и дополнительных корректировок, а также понимания получаемых результатов не всегда является приемлемым решением (в данном случае и во все была произведена попытка переноса метода из одной научной парадигмы в другую парадигму). Вышеописанное поведение параметра h Коэна не является проблемой в частотной статистике, для которой он разрабатывался. На практике расчетные значения данного параметра в частотной статистике, в отличие от байесовской статистики, попросту не применяются для прямой оценки экспериментального эффекта в реальных экспериментах, поскольку не являются «статистически значимыми».

² Александр Звонок // GitFlic. URL: https://gitflic.ru/project/aazvonok/verification_case_bayesian_binomial (дата обращения: 18.05.2024).



Они используются для ответа на вопрос: какой планируемый объем выборок (экспериментальной и контрольной) необходим для выявления статистически значимых различий (например, с помощью *t*-критерия Стьюдента) при ожидаемом размере экспериментального эффекта с разумной долей уверенности? Одним из основополагающих эпистемологических оснований частотной статистики является жесткая фиксация размеров выборок до начала эксперимента: нарушение этого положения ломает всю достоверность получаемых результатов, отчего так важно заранее определить оптимальный размер выборок (подробнее с механизмом «поломки» на примере *t*-критерия можно ознакомиться в работе М. Х. Херцога, Г. Фрэнсиса и А. Кларка [16, с. 145–149]). В свою очередь, для байесовской статистики предварительное определение объема наблюдений не имеет такого значения. В рамках байесовского подхода мы можем сосредоточиться на сборе данных, постоянно анализируя их, пока не получим результаты, точность которых достаточна для решения исследовательских задач (из-за этого байесовские методы получают все большую популярность в разного рода мониторинговых исследованиях, требующих быстрые и частые промежуточные «срезы» результатов). На самом деле смешивание и спутывание правил частотной и байесовской статистики, вызванное непониманием их различий, согласно оценке Ф. А. Шродта, является одним из «смертных грехов» современного количественного анализа [17].

2. Если бы в приведенном кейсе верификации методов мы применяли данные реальных экспериментов вместо искусственно сгенерированных, то не знали бы истинные значения параметров в выборках. Это попросту не позволило бы нам обнаружить имеющиеся противоречия (низкие значения показателя *h* Коэна при фактических высоких истинных значениях размера эффекта в выборках). На самом деле этот простой кейс доказывает, что разработка и верификация методов с применением данных реальных исследований возможна только в случаях, когда мы можем получить доступ к достоверным значениям анализируемых параметров, чтобы сравнить их со значениями, которые продуцируют верифицируемые математико-статистические методы. Иногда это можно сделать, обследовав генеральную совокупность и устранив все факторы, искажающие измерения. Иногда мы изначально знаем достоверные выходные значения (например, к какой группе относятся респонденты при разработке метода их классификации). В остальных случаях

требуется применение сгенерированных данных с известными заданными свойствами.

3. Поведение математико-статистических методов при специфическом характере анализируемых данных или даже конкретной исследовательской ситуации способно существенно различаться, что может потребовать дополнительной верификации (как, например, в случае доверительных интервалов опроса в эксперименте А. Крыштановского). Существует возможность возникновения даже технических проблем, которые напрямую не относятся к зоне ответственности социологии, но могут проявиться при попытке проведения реальных прикладных исследований с помощью разработанной методологии. Например, ранее упоминаемая байесовская модель BEST Дж. К. Крушке, предлагаемая как байесовская альтернатива *t*-критерию Стьюдента, разрабатывалась под запросы социально-гуманитарных наук: малые объемы выборок наблюдений и наличие «выбросов» данных. Сложная иерархическая структура модели привела к значительному росту вычислительной сложности: попытка применить модель к большим выборкам в десятки или сотни тысяч наблюдений в эксперименте (что является нормальной исследовательской ситуацией для естественных и технических наук) приведет к скачкообразному увеличению времени вычислений и количеству потенциально возникающих ошибок, сделав невозможным практическое использование модели. На практике ученые-байесианцы, занимающиеся естественнонаучными исследованиями, применяют значительно более простые байесовские модели, недостатки точности которых нивелируются объемами имеющихся наблюдений. Даже не беря во внимание байесовскую статистику, которая относительно слабо известна в социологии, существует достаточно математико-статистических методов в традиционной частотной статистике, которые могут демонстрировать нестандартное поведение в определенных условиях (или применяться в нестандартных ситуациях).

Последний важный вопрос, который необходимо затронуть в контексте темы текущей работы – вопрос публикации методологических исследований, проведенных с использованием симуляций. Первый стандарт публикации исследований, выполненных на основе искусственно сгенерированных данных, был предложен Д. Хоглином и Д. Эндрюсом еще в 1975 г. и уже в 1976 г. принят журналом *Journal of the American Statistical Association* (JASA), что позволило ему со временем полу-



чить признание и в международном научном сообществе. Если не вдаваться в тонкие технические подробности данного соглашения, а проанализировать непосредственно пример, приводимый авторами в оригинальной статье, в научные публикации, содержащие сгенерированные данные, рекомендовалось включать следующую информацию:

1) сведения о дизайне гипотетического исследования, в котором применяются симулированные данные (например, исследователь генерирует две независимых выборки наблюдений для последующего сравнения средних, имитируя контролируемый параллельный эксперимент с экспериментальной и контрольной группой);

2) информация об условиях генерации данных (какому вероятностному распределению подчиняются, его точные заданные характеристики, такие как средние значения или дисперсия, размер ошибки и т.д.);

3) информация о применяемом программном обеспечении и алгоритмах для генерации искусственных данных (наименование и версия программы, использовались ли методы Монте-Карло и т.д.) [18].

Вышеозначенные требования являют собой относительно универсальную концепцию, подходящую для большинства фундаментальных и прикладных научных исследований в различных сферах. Следует отметить, что поскольку разработка современных передовых математико-статистических методов практически невозможна без использования научного программирования, текущие стандарты JASA настоятельно рекомендуют авторам приводить в своих работах программный код или ссылки на репозитории с кодом и данными, позволяющие в точности воспроизвести результаты публикуемого исследования³. Однако большинство научных журналов пока не требуют от своих авторов знаний, умений и навыков научного программирования.

Существуют и более строгие стандарты опубликования исследований со сгенерированными данными, например, наиболее жесткое обнаруженное нами соглашение для медицинских исследований, датированное 2006 г., содержит следующие требования:

«0. Приведение детального протокола всех аспектов симуляции (с обоснованием всех принятых решений).

1. Четкое определение цели и задач симуляции.

³ The JASA Reproducibility Guide // GitHub. URL: <https://jasa-acs.github.io/repro-guide/> (дата обращения: 18.05.2024).

2. Детальное описание процедур симуляции (степень зависимости между сгенерированными выборками данных, допуск ошибок, программное обеспечение, тип генератора случайных чисел, характеристики стартовых значений случайности в случае их наличия (так называемая переменная “random seed”).

3. Указание методов генерации выборок данных.

4. Дизайн гипотетического исследования, моделируемый с помощью сгенерированных данных.

5. Статистические методы, подлежащие верификации посредством сгенерированных данных.

6. Истинные значения симулированных данных.

7. Количество симуляций, которые необходимо выполнить.

8. Критерии оценки эффективности статистических методов для различных сценариев (оценки смещения, точности, охвата).

9. Представление итоговых результатов симуляции» [19, p. 4281].

Однако в социологии, в том числе зарубежной, на текущий момент не существует общепринятых стандартов публикации исследований с симулированными данными, поэтому данный вопрос остается на усмотрение автора исследования.

Итак, нами были продемонстрированы ключевые методологические возможности и ограничения применения искусственно сгенерированных данных при разработке и внедрении математико-статистических методов в социологическое исследовательское поле, однако главной целью данной работы было собственно привлечение внимания к проблеме верификации искомых методов в эмпирической социологии. Современные вызовы в области социологической методологии и методов требуют от социологов овладения надпрофессиональными компетенциями, которые традиционно рассматривались как находящиеся вне социологической проблематики: навыками научного программирования, углубленными математическими знаниями. Однако именно освоение данных областей позволит социологу выйти на уровень действительного аналитика, чьи заключения будут пользоваться авторитетом и доверием со стороны общества и конкретных социальных институтов. Ошибки в разрабатываемых социологических методах, как правило, не связаны с риском потери человеческих жизней, как в случае медицинских или инженерных исследований. Но верификация математико-статистиче-



ских методов, разрабатываемых в социологии или внедряемых в социологическую проблематику из других наук, может значительно повысить доверие к искомым методам и результатам исследования в целом, что особенно важно в прикладной социологии. Например, при проведении коммерческих социологических исследований или принятии управленческих решений с использованием математико-статистических методов анализа социологических данных применение воспроизводимого математического эксперимента с искусственными данными и последующим подтверждением адекватности и эффективности верифицируемых методов позволит добиться большего признания их значимости.

Список литературы

1. Gilbert N., Troitzsch K. G. Simulation for the Social Scientist. New York : Open University Press, 2005. 295 p.
2. Толстова Ю. Н. Математическое моделирование социальных процессов и социология // Социологические исследования. 2018. № 9. С. 104–112. <https://doi.org/10.31857/S013216250001965-4>, EDN: YMRIOD
3. Harwell Dr. M., Kohli Dr. N., Peralta-Torres Ya. A survey of reporting practices of computer simulation studies in statistical research // The American Statistician. 2017. Vol. 72, iss. 4. P. 321–327. <https://doi.org/10.1080/00031305.2017.1342692>
4. Hauck W. W., Anderson S. A survey regarding the reporting of simulation studies // The American Statistician. 1984. Vol. 38, iss. 3. P. 214–216. <https://doi.org/10.1080/00031305.1984.10483206>
5. Зюзьков В. М. Эксперименты в теории чисел. Томск : Изд-во НТЛ, 2019. 348 с.
6. Thall P. F. Bayesian clinical trial design in a cancer center // CHANCE. 2001. Vol. 14, iss. 3. P. 23–28. <https://doi.org/10.1080/09332480.2001.10542279>
7. Киричевский М., Комарова С. Верификация методики измерений в химической лаборатории // Роль технического регулирования и стандартизации в эпоху цифровой экономики : сб. ст. участников III Междунар. науч.-практ. конф. молодых ученых (Екатеринбург, 25 ноября 2021 г.). Екатеринбург : ИД «Ажур», 2021. С. 262–268. EDN: HYPUDV
8. O’Kelly M., Anisimov V., Campbell C., Hamilton S. Proposed best practice for projects that involve modeling and simulation // Pharmaceutical Statistics. 2016. Vol. 16, iss. 2. P. 107–113. <https://doi.org/10.1002/pst.1789>
9. Добренков В. И., Кравченко А. И. Фундаментальная социология : в 15 т. Т. 2. Эмпирическая и прикладная социология. М. : ИНФРА-М, 2004. 986 с.
10. Winter J. de, Dodou D. Five-Point Likert Items: t test versus Mann – Whitney – Wilcoxon // Practical Assessment, Research & Evaluation. 2010. Vol. 15. Art. 11. <https://doi.org/10.7275/bj1p-ts64>
11. Звонок А. А. Байесовская экспериментальная оценка социальных технологий: методологические и методические аспекты // НОМОТНЕТИКА: Философия. Социология. Право. 2024. Т. 49, № 1. С. 26–38. <https://doi.org/10.52575/2712-746X-2024-49-1-26-38>, EDN: DYXJUD
12. Cohen J. Statistical power analysis for the behavioral sciences. New York : Lawrence Erlbaum Associates, 1988. 567 p.
13. Звонок А. А. Байесовское моделирование биномиальных экспериментов в социологии: проблемный анализ // Цифровая социология. 2024. Т. 7, № 1. С. 14–25. <https://doi.org/10.26425/2658-347X-2024-7-1-14-25>, EDN: EUFDXE
14. Sauro J., Lewis J. R. Quantifying the User Experience. Practical Statistics for User Research. Amsterdam, Boston, Heidelberg, London, New York, Oxford, Paris, San Diego, San Francisco, Singapore, Sydney, Tokyo : Elsevier, 2012. 295 p.
15. Sawilowsky S. New effect size rules of thumb // Journal of Modern Applied Statistical Methods. 2009. Vol. 8, iss. 2. P. 467–474. <https://doi.org/10.22237/jmasm/1257035100>
16. Херцог М. Х., Фрэнсис Г., Кларк А. Статистика и планирование эксперимента для непосвященных: как отучить статистику лгать. М. : ДМК Пресс, 2023. 174 с.
17. Schrodt P. A. Seven deadly sins of contemporary quantitative political analysis // Journal of Peace Research. 2014. Vol. 51, iss. 2. P. 287–300. <https://doi.org/10.1177/0022343313499597>
18. Hoaglin D. C., Andrews D. F. The reporting of computation-based results in statistics // The American Statistician. 1975. Vol. 29, iss. 3. P. 122–126. <https://doi.org/10.1080/00031305.1975.10477393>
19. Burton A., Altman D. G., Royston P., Holder R. L. The design of simulation studies in medical statistics // Statistics in Medicine. 2006. Vol. 25, iss. 24. P. 4279–4292. <https://doi.org/10.1002/sim.2673>

Поступила в редакцию 22.05.2024; одобрена после рецензирования 16.06.2024; принята к публикации 09.08.2024; опубликована 29.11.2024

The article was submitted 22.05.2024; approved after reviewing 06.06.2024; accepted for publication 09.08.2024; published 29.11.2024