

ТЕОРИЯ И МЕТОДОЛОГИЯ

АЛЕКСАНДР АНАТОЛЬЕВИЧ ЗВОНОК

кандидат философских наук,
доцент Луганского государственного
педагогического университета,
Луганск, Россия;
e-mail: al.zvonok@gmail.com
ORCID: 0009-0007-7332-1330



Проблема сравнения социологических данных, выраженных в порядковых шкалах: решение методом байесовской порядковой пробит-регрессии

УДК: 51-77:303.092.55

DOI: 10.24412/2079-0910-2026-3-202-218

Статья раскрывает сущность и особенности сравнения порядковых данных в эмпирической социологии посредством метода байесовской порядковой пробит-регрессии. Охарактеризованы основные проблемы статистического анализа и теоретической интерпретации его результатов, возникающие при работе с данными, выраженными в порядковых шкалах: ограниченность спектра проводимых математических операций с порядковыми данными, контринтуитивность выводов, получаемых с помощью распространенных методов непараметрической статистики. Затронуты ограничения классических решений указанных проблем, включая применение параметрических методов на порядковых данных и сознательную трактовку порядковых шкал как интервальных. В рамках исследования с применением современного высокоуровневого инструмента байесовского анализа РуМС построена базовая байесовская пробит-модель для сравнения двух выборок, выраженных в порядковых шкалах. Рассмотрен ряд методологических преимуществ искомой байесовской модели относительно традиционно применяемых методов частотной статистики: отсутствие жестких требований к форме распределения данных, интуитивно понятная теоретическая интерпретация результатов сравнения и возможность применения байесовской версии анализа мощности для измерения значимого размера эффекта различий между анализируемыми выборками наблюдений. На примере искусственно сгенерированных выборок данных, представленных в виде 5-балльной шкалы Ликерта, продемонстрировано эмпирическое превосходство байесовской порядковой пробит-регрессии над ее частотной версией, а также такими популярными непараметрическими критериями различий, как U-критерий Манна — Уитни и критерий меди-

ан Муда. Полученные результаты свидетельствуют о значительном потенциале применения байесовской порядковой пробит-модели в сравнительных социологических исследованиях в условиях, когда размеры выборочных совокупностей слишком малы или ожидаемый размер эффекта слишком низок для его выявления традиционными методами.

Ключевые слова: социологические данные, порядковая шкала, шкала Ликерта, порядковая пробит-регрессия, анализ мощности, размер эффекта, байесовский анализ, байесовская модель, РуМС.

Введение

Порядковые шкалы, такие как шкала Ликерта, играют ключевую роль в социологических исследованиях, поскольку позволяют измерять субъективные установки, оценки и мнения респондентов. Свою популярность в социологии порядковые шкалы снискали благодаря таким особенностям, как удобство применения, интуитивная понятность для опрашиваемых, гибкость и универсальность, а также возможность кодирования результатов подобных опросов для последующей статистической обработки. Все это обусловило огромную популярность порядковых шкал в дескриптивных социологических исследованиях, где главная цель — зафиксировать и визуализировать распределение мнений. Однако сравнение данных, выраженных в порядковых шкалах, сопряжено с рядом методологических проблем. Основные сложности в этом контексте связаны с природой порядковых данных и ограничениями их математической обработки. Как правило, параметрические статистические методы не подходят для порядковых данных, а непараметрические методы имеют свои ограничения и особенности интерпретации, что приводит к их некорректному применению. Распространенность некорректных практик применения непараметрических методов для сравнения данных, выраженных в порядковых шкалах, и неправильная интерпретация их результатов столь обширны, что подобные случаи сегодня происходят даже в исследованиях, которые опубликованы в ведущих журналах, входящих в международные наукометрические базы [Zhang, Zhang, 2009].

Указанная проблема особенно актуальна для социологии, в рамках которой подготовка специалистов обычно не предусматривает получения углубленных математических знаний в области теории вероятностей и математической статистики в той же мере, в какой производится подготовка на технических и естественнонаучных специальностях. Такая ситуация связана с наличием в социологической среде устойчивого стереотипа о социологии как о «гуманитарной» науке, в которой математические методы имеют вспомогательный характер [Шуметов, Лясковская, 2014, с. 80–81]. К сожалению, анализ 93 образовательных программ на социологических факультетах 86 российских вузов, проведенный в 2022 г. В.В. Зыряновым, подтверждает актуальность означенной проблемы: в 58% вузов статистическая подготовка будущих социологов производилась в недостаточном объеме [Зырянов, 2022]. Подобные тенденции негативно влияют на развитие социологической науки, ведь верификация, апробация и внедрение новых математико-статистических методов в конкретные отрасли социологии, а также модификация существующих методик анализа под конкретные социологические кейсы и выработка корректных практик по их применению — все вышеозначенные задачи входят в зону ответственности методологии социологических исследований. Фундаментальная математика сосре-

доточена на работе с абстрактными числами, без учета особенностей контекста и реальных ситуаций, а прикладным математикам недостает знаний о предметных областях, в обиход которых входят разработанные математико-статистические методы. Поэтому этот род научной деятельности все еще остается прерогативой социологов-исследователей [Звонок, 2024, с. 390].

Целью текущего исследования является подробное раскрытие сущности и особенностей применения в социологии метода сравнения порядковых данных, основанного на синтезе двух актуальных направлений “data science”: порядковой пробит-регрессии и современного байесовского анализа. Вобрав в себя преимущества обоих подходов, указанный метод представляет собой эффективный инструмент сравнения выборок социологических данных, выраженных в порядковых шкалах, количественно отражающий меру этих изменений с учетом их статистической значимости и применимый в социологических исследованиях повышенной важности.

Проблема порядковых шкал и ограничения классических решений в сравнительных социологических исследованиях

Порядковые шкалы традиционно являются чрезвычайно удобными для применения в рамках опросных полевых методов сбора социологической информации и, как правило, на интуитивном уровне хорошо воспринимаются и понимаются респондентами. Это обусловило широкое применение таких опросных конструкций, как шкала Ликерта или семантический дифференциал, практически во всех социальных науках, от социологии до эконометрии. Однако в математическом контексте данные, выраженные в порядковых шкалах, являются особенно сложными для статистической обработки и последующей теоретической интерпретации.

Как известно, порядковая шкала является более продвинутой версией номинальной шкалы и позволяет не только классифицировать объекты, но и упорядочивать их по принципу «больше-меньше», однако, в отличие от интервальных шкал, большинство арифметических операций в рамках порядковой шкалы неосуществимо с точки зрения здравого смысла. Интервалы, разделяющие категории, не равны по величине, поэтому нахождение таких характеристик, как среднее арифметическое, для порядковых шкал не имеет адекватной интерпретации. Категории «никогда», «очень редко» и «редко» могут стоять одна за другой, но расстояние между первой и второй категорией несоизмеримо с расстоянием между второй и третьей категорией. По сути, мы можем только подсчитать процентное отношение объектов в порядковой выборке и определить медиану [Лукашин, Рахлина, 2012, с. 17–18].

Самым простым и очевидным исторически сложившимся решением для сравнения порядковых данных в частотной статистике является применение непараметрических методов, основанных на рангах, медианах и других нестандартных значениях. Однако, во-первых, как правило, указанные методы обладают значительно меньшей мощностью в сравнении с их параметрическими аналогами. Во-вторых, несмотря на отсутствие условия «нормальности» данных, непараметрическая статистика имеет более высокий показатель специальных условий, что повышает требования к компетенциям применяющего их исследователя. Например, даже хорошо апробированный U-критерий Манна — Уитни в частных случаях может быть эквивалентен тесту равенства медиан, однако в целом проверяет гипотезу о стоха-

стическом доминировании [Hart, 2001]. Наконец, теоретическая интерпретация результатов, получаемых при помощи непараметрических методов анализа, является существенно более сложной и менее интуитивной в конкретных прикладных исследованиях. Интерпретируя различия медиан так же как средних или неправильно понимая концепцию стохастического доминирования, исследователь может прийти к совершенно неверным выводам в исследовании, в то время как математические расчеты выдают статистически корректные значимые результаты.

Другая важная проблема порядковых шкал состоит в том, что данные, выраженные в них, во многих случаях не получается соотнести с каким-либо известным законом распределения вероятностей, и потому с точки зрения строгой теории частотной статистики они не могут подвергаться анализу с помощью более совершенных и точных параметрических математико-статистических методов. Это существенно ограничивает разработку методов для анализа порядковых данных, зачастую или порождая одиночные решения для частных случаев, или вынуждая исследователей прибегать к различным теоретическим допущениям о природе этих данных. Так, подход «рассмотрение порядковых данных как интервальных» применяется с середины XX в. в таких прикладных областях, как психометрия, где считается практически классическим решением, хотя и жестко осуждается консервативно настроенными сторонниками статистического анализа [Knapp, 1990]. Эмпирическая применимость такого «волевого» решения подтверждается многочисленными математико-статистическими экспериментами с искусственно сгенерированными данными как в простом измерении мощности параметрических методов [DeWees et al., 2020], так и при сравнении параметрических и непараметрических критериев [Winter, Dodou, 2010]. При этом эффективность параметрических методов возрастает с увеличением количества категорий на порядковой шкале, что было продемонстрировано на примере факторного анализа [Зангиева, Ротмистров, 2018, с. 44–45].

Однако подобная философия инструментализма в области статистического анализа является нарушением эпистемологических основ частотной статистики и существенно снижает доверие к результатам соответствующих научных исследований, даже если не брать во внимание потенциальную возможность некорректных теоретических интерпретаций их результатов из-за игнорирования онтологии данных. Дело в том, что за каждой существующей парадигмой статистического анализа стоит определенный пласт философских и эпистемологических построений: различных интерпретаций вероятности, доверительных интервалов и т. д. Если исследователь применяет определенный подход к статистическому анализу данных (частотный, байесовский или фидуциальный), то он должен и принимать соответствующую эпистемологию.

Метод порядковой пробит-регрессии для сравнения порядковых данных

Концепция применения статистических методов регрессии не для задач прогнозирования или классификации, но в рамках сравнительных исследований получила распространение, прежде всего, в научных областях, где неверная интерпретация результатов статистического анализа связана с явными и прямыми негативными последствиями, — в экспериментальной и социальной медицине. Результаты

исследований в этой сфере свидетельствуют, что применение к дискретным данным (к которым можно отнести и данные, выраженные в порядковых шкалах) *t*-критерия Стьюдента и непараметрических тестов, таких как *U*-критерий Манна — Уитни или его расширений, в задаче сравнения двух групп часто приводит к некорректным выводам. В качестве замены искомым методов было предложено использовать различные регрессионные модели [McElduff *et al.*, 2010].

Метод порядковой регрессии применяется для анализа взаимосвязи между зависимой переменной, измеренной в порядковой шкале, и одной или несколькими независимыми переменными (предикторами). В порядковой пробит-модели предполагается, что в основе наблюдаемой порядковой переменной y (например, шкалы Ликерта, закодированной как 0, 1, 2, 3, 4) лежит непрерывная латентная (ненаблюдаемая) переменная y^* . По сути, можно сказать, что y^* — это истинная, но ненаблюдаемая величина, а y — ее зашумленное и дискретизированное измерение. Латентная переменная y^* имеет стандартное нормальное распределение и определяется следующим образом:

$$y^* = \beta X + \epsilon, \epsilon \sim \text{Normal}(0, 1),$$

где β — коэффициент, измеряющий сдвиг латентной переменной между группами, X — предиктор, ϵ — случайная ошибка.

Пороговые значения θ , разбивающие латентную переменную y^* на интервалы, соответствующие наблюдаемым k -категориям, всегда имеют количество, равное $k - 1$, и занимают определенные положения на непрерывной шкале [Greene, 2018, p. 868–875]. С наглядной визуализацией подобной порядковой пробит-модели можно ознакомиться ниже (см. рис. 1).

Заметим, что традиционно порядковая пробит-модель применяется для предсказания вероятности принадлежности зависимой переменной к одной из порядковых категорий. Однако в контексте предлагаемого исследования уместно сделать акцент именно на сдвиге латентной переменной между выборками, то есть в какой выборке наблюдений ответы в целом имеют более высокие значения. Такой подход существенно упрощает модель, так как позволяет использовать один бинарный предиктор принадлежности к группе, например, 0 и 1. Положительный коэффициент β будет свидетельствовать о более высоких значениях во второй выборке порядковых наблюдений, отрицательный коэффициент — о более низких значениях. Таким образом, нам следует концептуально переосмыслить порядковую пробит-регрессию, рассматривая ее как статистический критерий различий между выборками порядковых данных.

Здесь необходимо сделать очень важное замечание о ключевом преимуществе порядковой пробит-регрессии в сравнении с прочими видами порядковой регрессии именно в задаче сравнения двух выборок порядковых данных. В пробит-модели ненаблюдаемая переменная имеет стандартное нормальное распределение. Соответственно, значения сдвига латентной переменной порядковой пробит-модели уже являются стандартизированными [Agresti, 2010, p. 122–123].

Как нами уже было замечено, одной из главных проблем сравнения порядковых данных с помощью классических методов, таких как непараметрические тесты, является сложность и неоднозначность интерпретации результатов подобного анализа, а также проблемы с количественным измерением разницы, так называемого

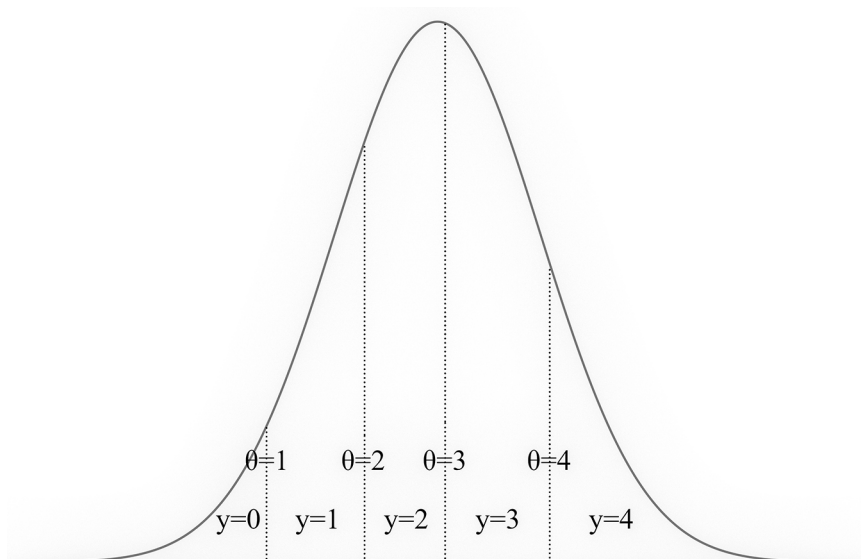


Рис. 1. Пороговые значения латентной переменной y^* для пяти порядковых категорий наблюдаемой переменной y

Fig. 1. Threshold values of latent variable y^* for five ordinal categories of observed variable y

«размера эффекта», если изъясняться в терминологии концепции анализа мощности Дж. Коэна. Высокую популярность в сравнительных исследованиях получил разработанный им параметр d Коэна: разница средних значений двух выборок интервальных нормально распределенных данных, разделенная на их объединенное стандартное отклонение (то есть оценки d Коэна стандартизованы). Иначе говоря, d Коэна обозначает то, на сколько объединенных стандартных отклонений различается показатель в сравниваемых выборках, причем его интерпретация имеет универсальный характер: 0,2 — «слабый эффект», 0,5 — «средний эффект», 0,8 — «сильный эффект» [Cohen, 1988, р. 20–27]. Эта универсальность достигается благодаря характеристикам нормального распределения, в том числе широко известному правилу «трех сигм». Позднее Ш.С. Савиловски расширил эти интерпретации размера эффекта для методов, использующих вероятностные алгоритмы Монте-Карло: 0,01 — «очень слабый эффект», 0,2 — «слабый эффект», 0,5 — «средний эффект», 0,8 — «сильный эффект», 1,2 — «очень сильный эффект» и 2,0 — «огромный эффект» [Sawilowsky, 2009].

К сожалению, для выборок данных, выраженных в порядковых шкалах, не существует настолько же универсального и интерпретируемого показателя размера эффекта, позволяющего проводить их сравнение в чистом виде. Однако сдвиг латентной переменной между сравниваемыми выборками концептуально можно рассматривать как d Коэна (разница, выраженная в объединенных стандартных отклонениях, так как сдвиг происходит на шкале, общей для сравниваемых выборок) и, так как латентная переменная имеет стандартное нормальное распределение, «размер эффекта» при сравнении этих выборок также можно интерпретировать, используя существующие соглашения, указанные выше. Иначе говоря, вычисляя значение коэффициента β как «размера эффекта», мы получаем достаточно полную и точную,

стандартизированную и легко интерпретируемую информацию о силе и направленности изменений в сравниваемых наборах данных, выраженных в общей для них порядковой шкале, вне зависимости от фактического количества категорий на ней.

Идея анализа мощности Дж. Коэна, несмотря на свой явный потенциал, содержит один существенный недостаток, вызванный особенностями частотной статистики. Так, значение d Коэна 0,5 и пройденная проверка на статистическую значимость с помощью t -критерия Стьюдента вовсе не доказывают, что истинный размер эффекта в сравниваемых выборках можно интерпретировать как «средний». Далее мы продемонстрируем, как переход к байесовской парадигме статистического анализа устраняет обнаруженный недочет, позволяя получать статистически значимые оценки размера эффекта в сравнительных байесовских моделях.

Построение байесовской порядковой пробит-модели

В рамках решения задач текущего исследования нами было принято решение о применении научной библиотеки байесовского анализа PyMC языка программирования *Python* для построения байесовской порядковой пробит-модели. PyMC использует для порядковых моделей современный алгоритм марковских цепей Монте-Карло (МЦМК) *No-U-Turn Sampling*, который автоматически адаптируется к сложной геометрии, в отличие от традиционных МЦМК-алгоритмов программ байесовского анализа JAGS и BUGS [Hoffman, Gelman, 2014]. С полным подробно прокомментированным исходным кодом модели на языке программирования *Python*, а также наборами искусственных данных для ее верификации можно ознакомиться в репозитории автора предлагаемой работы¹. Размещение исходного кода полностью соответствует принципам воспроизводимости результатов исследований, принятым в ведущих международных методологических журналах.

Прежде всего, следует кратко рассмотреть основные особенности байесовского статистического вывода, которые применялись при создании модели в искомом исследовании. Фактически байесовский подход придерживается иного понимания вероятности и основан на отличающихся от традиционной статистики эпистемологических принципах, являясь отдельным направлением философии науки (так как байесовская модель рационального познания может применяться и вне сферы математических вычислений). Некоторые философы даже рассматривают байесовскую парадигму как отдельную форму научного вывода, представляющую собой синтез традиционных методов дедукции, индукции и абдукции [Светлов, 2021, с. 19–21]. Важной идеей современного байесовского статистического вывода, которую следует обозначить, является оперирование вероятностными распределениями параметров вместо точечных оценок. Результатом байесовского анализа являются апостериорные распределения — это полные вероятностные описания параметров после наблюдения данных. Они не нуждаются в проверке статистической значимости в том классическом смысле, какой придается подобным проверкам в частотной статистике, так как уже содержат в себе всю информацию о неопределенности и не опираются на гипотетические повторные выборки, как концепция p -значений. Другая особенность байесовского подхода — это процесс уменьшения незнания об

¹ Режим доступа: https://gitflic.ru/project/aazvonok/bayesian_ordered_probit_model (дата обращения: 14.08.2025).

исследуемом объекте путем перехода от предшествующих опыту априорных знаний и предположений к апостериорным знаниям, представляющим собой совокупность априорных установок и данных, полученных в результате наблюдений. Из нашего теоретического рассмотрения концепции порядковой пробит-регрессии мы видим, что наиболее важные априорные гиперпараметры будущей байесовской модели уже известны и могут быть без изменений имплементированы в модель:

1. Латентная переменная имеет стандартное нормальное распределение. Это давно устоявшийся стандарт для реализаций порядковой пробит-регрессии.
2. Наблюдаемые значения обеих выборок представляют собой упорядоченные категориальные данные, имеющие количество порогов разделения, равное количеству категорий, уменьшенному на 1.
3. Предиктор первой выборки устанавливает значение β в качестве отправной точки сравнения (нулевое значение), предиктор второй выборки будет отражать систематический сдвиг β в общем латентном пространстве относительно первой группы.

Единственным существенным априорным гиперпараметром порядковой пробит-модели, в отношении которого не существует явных соглашений, является расположение пороговых значений на шкале латентной переменной. В представленной модели априорные установки для пороговых параметров были равномерно рассредоточены нами по шкале латентной переменной в интервале $[-3, 3]$. Такая спецификация отражает предположение о том, что пороги упорядочены и расположены симметрично относительно нуля, что соответствует стандартизованному латентному пространству. Таким образом, для пяти категорий в качестве априорных на шкале латентной переменной будет автоматически установлено четыре пороговых значения $[-3, -1, 1, 3]$, для шести категорий пять пороговых значений $[-3, -1,5, 0, 1,5, 3]$ и т. д. В строгих научных исследованиях, особенно имеющих характер фундаментальных и требующих высокого уровня обоснованности результатов, исследователи-байесианцы часто устанавливают неинформативные априорные распределения вида $\text{Normal}(0, 10000)$, позволяя данным говорить самим за себя. Однако Э. Гельман, Д. Симпсон и М. Бетанкурт аргументированно обосновывают разумный компромисс между генеративностью априоров (способностью генерировать только данные, соответствующие нашему пониманию задачи) и их прогностической эффективностью (способностью корректно предсказывать новые данные) в прикладных научных областях [Gelman et al., 2017]. Следовательно, наши априорные установки балансируют на грани подобного компромисса, обеспечивая при этом устойчивость модели к выбросам данных.

Интерпретация результатов и проверка гипотез в байесовском анализе существенно отличаются от принятых в частотной статистике. Тем не менее зачастую принципы и соглашения, принятые в частотной статистике, могут заимствоваться байесианцами попросту по той причине, что являются общепринятыми. Например, в рамках байесовской парадигмы концептуально приемлемо применять доверительные интервалы значительно меньше 95%. В то же время на практике большинство исследователей все равно придерживаются значения 95%, потому что оно принимается научным сообществом, особенно в социальных науках. В нашем исследовании для подтверждения статистической значимости сдвига латентной переменной в байесовской порядковой пробит-регрессии применяется три способа проверки значимости различий между порядковыми выборками:

1. Простейшим байесовским тестом на статистическую значимость является проверка того, что 95-процентный интервал плотности апостериорного распределения, или ИПАР (“highest density interval”, HDI), коэффициента β между 2,5- и 97,5-перцентилем не затрагивает нулевое значение.
2. Более современной байесовской альтернативой является установление границ зоны практической эквивалентности, или ЗПЭ (“region of practical equivalence”, ROPE), то есть определенной области вокруг нулевого значения, апостериорная вероятность попадания в которую будет вероятностью результата «не выявлено значимых статистических различий». Заметим, что в отличие от частотной статистики, где нулевую гипотезу можно только опровергнуть, байесовская эпистемология позволяет ее принять, например, если апостериорная вероятность ЗПЭ составляет более 95%. Если же ЗПЭ составляет менее 5%, можно говорить о наличии значимых различий между выборками с высокой степенью уверенности. Дж.К. Крушке в рамках развития теоретической концепции зоны практической эквивалентности предлагает соглашение об установлении пределов ЗПЭ, представляющих половину от «слабого» размера эффекта тестируемого параметра. Таким образом, для классической интерпретации d Коэна оптимальное значение ЗПЭ будет $[-0,1, 0,1]$ [Kruschke, 2018, p. 6]. Как нами было показано ранее, величина сдвига латентной переменной в порядковой пробит-регрессии концептуально эквивалентна d Коэна, поэтому разумным будет придерживаться аналогичных параметров ЗПЭ и для тестирования коэффициента β .
3. В качестве более мощного одностороннего теста в байесовском анализе может быть использована проверка предположения о том, что площадь апостериорного распределения коэффициента β на отрезке $[-\infty, 0]$ или $[0, \infty]$ составляет более 95%. При такой проверке статистической гипотезы, как правило, устанавливается 0 в качестве «референсного» значения и вычисляются площади апостериорного распределения по обе стороны от него. Однако технически многие инструменты байесовского анализа позволяют устанавливать и ненулевые значения, например, для проверки гипотез о предполагаемом «среднем» или «сильном» размере эффекта. Таким образом, становится возможным применять байесовский анализ мощности для получения статистически значимых количественных выводов, например, «с апостериорной вероятностью 95% латентный сдвиг β между выборками превышает 0,2, что с высокой степенью уверенности свидетельствует о наличии статистически значимых различий как минимум слабой величины».

Верификация байесовской модели на сгенерированных данных

Одним из главных препятствий в разработке статистических методов сравнения выборок порядковых данных является их непараметрический характер. Проблема в том, что для верификации любых математико-статистических методов нужно знать истинные значения выходных параметров для их сравнения с теми, которые порождаются в результате работы соответствующего метода. Выборки реальных данных часто неприменимы в этом случае, так как в настоящих исследованиях неизбежны различные скрытые отклонения и смещения, которые невозможно предсказать и

устранить в полной мере без знания полной информации о генеральных совокупностях. Для верификации и оценки большинства статистических методов, основанных на анализе выборок наблюдений, можно сформировать произвольное количество данных, взятых из определенного закона вероятностного распределения с требуемыми параметрами, что мы можем видеть на соответствующих исследованиях [Горбунова и др., 2010]. Но для верификации байесовской порядковой пробит-модели мы не можем использовать такие параметры выборок, как их среднее, медиана или стандартное отклонение, так как метод измеряет величину сдвига латентной переменной, которая не наблюдается напрямую и не устанавливается каким-либо определенным законом распределения вероятностей.

Исходя из вышесказанного, в рамках текущего исследования нами было принято решение смоделировать специальный набор искусственных порядковых данных, состоящий из двух выборок с заданным сдвигом латентной переменной. Сначала нами, посредством библиотеки научных вычислений *NumPy* языка программирования *Python* и с использованием стартового значения случайности 62650, была сгенерирована случайная выборка из 10 000 наблюдений — чисел 0, 1, 2, 3, 4 (представляющих собой ответы, выраженные в классической пятибалльной шкале Ликерта гипотетического исследования) со следующим распределением вероятностей: $P_0 = 0,1$, $P_1 = 0,25$, $P_2 = 0,3$, $P_3 = 0,25$, $P_4 = 0,1$. Затем была сгенерирована независимая выборка такого же объема с распределением вероятностей $P_0 = 0,05$, $P_1 = 0,14$, $P_2 = 0,23$, $P_3 = 0,35$, $P_4 = 0,23$. Согласно расчетам, такие вводные параметры должны были имитировать сдвиг латентной переменной β между сгенерированными выборками, приближенный к 0,5 — «среднему» размеру эффекта в рамках концепции анализа мощности.

Далее с помощью научной библиотеки *statsmodels* языка программирования *Python* методом максимального правдоподобия нами была построена частотная порядковая пробит-модель и проведена ее оптимизация алгоритмом Бройдена — Флетчера — Гольдфарба — Шанно. Характерная особенность библиотеки *statsmodels* состоит в том, что она производит построение и подгонку регрессионных моделей в полуавтоматическом режиме: это решение исключило возможность ошибок, допущенных автором исследования на этапе вычислений, позволяя использовать получаемые результаты в качестве «эталонных». Результаты сравнения сгенерированных выборок с помощью полученной модели показали статистически значимый сдвиг латентной переменной $\beta = 0,504$ при 95-процентном частотном доверительном интервале $[0,474, 0,533]$, что полностью соответствует заложенным в тестовые данные ожиданиям. Примечательно, что наша байесовская порядковая пробит-модель после применения ее на сгенерированном наборе данных выдала идентичный результат: среднее значение апостериорного распределения $\beta = 0,504$ при 95% интервале плотности апостериорного распределения $[0,474, 0,533]$. Прежде всего, такие значения являются следствием того, что наша байесовская модель использует слабоинформативные регуляризующие априорные гиперпараметры, оказывающие минимальное субъективное влияние на результаты анализа, обеспечивая научную объективность модели. Однако теоретическая интерпретация этих результатов в частотном и байесовском анализе будет различаться. В частотной интерпретации $\beta = 0,504$ является оценкой фиксированного, но неизвестного параметра. Доверительный интервал $[0,474, 0,533]$ означает, что при гипотетическом многократном повторении эксперимента 95% таких интервалов накрыли бы истинное значение

параметра. В байесовской интерпретации β — это случайная величина, а интервал $[0,474, 0,533]$ — это область, содержащая 95% вероятностной массы апостериорного распределения. При этом среднее значение апостериорного распределения $\beta = 0,504$ используется только как вспомогательный параметр для упрощения восприятия результатов анализа.

Для дальнейшей оценки пределов чувствительности байесовской модели нами из подготовленного двухвыборочного набора данных методом простой случайной выборки было отобрано девять соответствующих пар независимых выборок меньших размеров: $n_1 = 15$, $n_2 = 20$, $n_3 = 25$, $n_4 = 30$, $n_5 = 35$, $n_6 = 40$, $n_7 = 45$, $n_8 = 50$ и $n_9 = 100$. Учитывая, что при столь малых размерах выборок и «генеральной совокупности» $n = 10\,000$ даже теоретически рассчитанная погрешность с 95-процентной доверительной вероятностью может достигать 25%, отбирались только те пары выборок, вычисленный результат анализа сдвига латентной переменной в которых (с помощью ранее построенной частотной модели *statsmodels*) находился в рамках доверительного интервала $[0,474, 0,533]$. Так как мы проверяем именно мощность методов сравнения порядковых данных, это решение нивелировало очень высокий риск попадания в тестовые выборки «зашумленных» данных, что могло бы исказить результаты анализа.

Затем искомая байесовская порядковая пробит-модель была применена к каждому из полученных наборов данных (входные параметры: алгоритм *No-U-Turn Sampling*, размер семпла 10 000) с последовательным проведением трех видов байесовской проверки статистических гипотез (см. рис. 2).

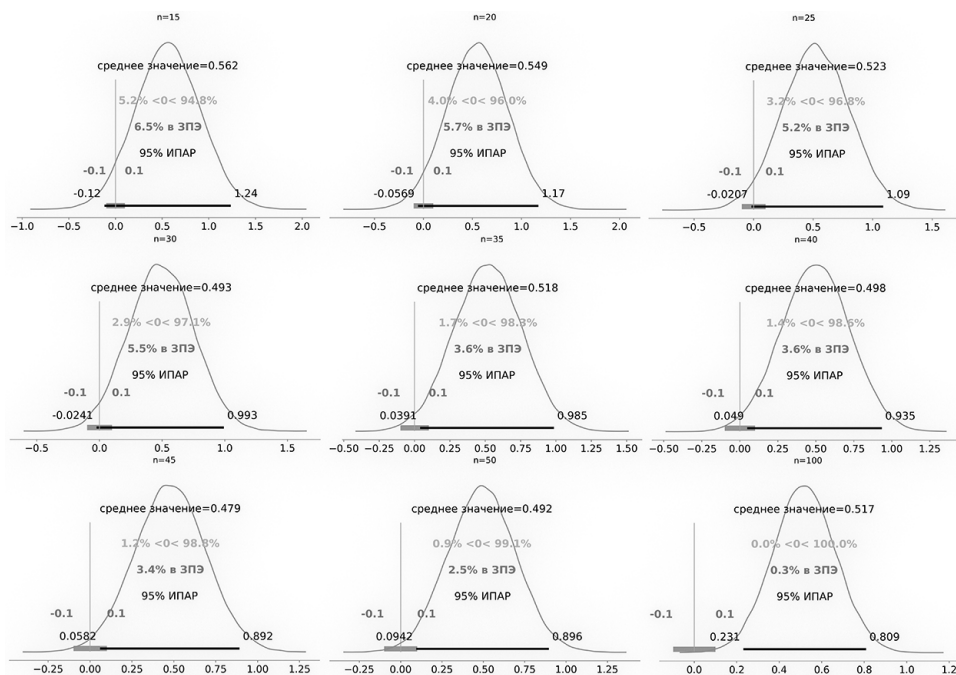


Рис. 2. Апостериорное распределение коэффициента β при различных размерах сгенерированных выборок

Fig. 2. Posterior distribution of the coefficient β for different sizes of generated samples

Далее нами было принято решение применить ко всем подготовленным наборам данных три метода анализа из частотной статистики: ранее реализованную модель частотной порядковой пробит-регрессии, односторонний вариант U-критерия Манна — Уитни и критерий медиан Муда. Для всех частотных методов был установлен пороговый уровень значимости $\alpha = 0,05$. Если не считать порядковую пробит-модель, два вышеуказанных непараметрических критерия хорошо зарекомендовали себя с середины XX в. в задаче выявления статистически значимых различий порядковых выборок и хотя бы один из них присутствует в большинстве существующих на сегодня программ и пакетов статистического анализа. С результатами применения всех указанных методов на тестовых парах независимых выборок от 15 до 100 наблюдений для каждой выборки в соответствующей паре можно ознакомиться ниже (см. табл. 1).

Как мы можем видеть из полученных значений, эффективность метода частотной пробит-регрессии оказалась приблизительно на одном уровне с двусторонними проверками статистических различий в байесовской реализации пробит-модели через 95% ИПАР и через апостериорную вероятность ЗПЭ $[-0,1, 0,1]$. Хотя частотная версия метода рассчитывает значение β при любых размерах выборок, полученные значения попросту не проходили проверку на статистическую значимость до выборок объема $n = 35$, что формально не дает права признавать различия между этими выборками статистически значимыми. Односторонний U-критерий Манна — Уитни стабильно демонстрировал статистически значимое стохастическое доминирование в одной из сравниваемых выборок уже начиная с $n = 30$, что обуславливает его высокую популярность среди прочих непараметрических методов. К сожалению, критерий медиан Муда показал в тестировании неустойчивые результаты, определив статистически значимые различия медиан в выборке $n = 35$, но затем снова не выявляя различий в медианах выборок до $n = 50$. Однако указанный факт все же демонстрирует нам важный методический аспект: равенство медиан сравниваемых выборок вовсе не означает равенство самих выборок. Медиана отражает только центральное положение данных, но не учитывает их форму распределения, дисперсию, моду и другие параметры. Наконец, наилучшую чувствительность на тестовых данных проявил метод односторонней байесовской проверки через вычисление апостериорной вероятности β на отрезке $[0, \infty]$: вероятность 95,99% была вычислена уже при размерах сравниваемых выборок $n = 20$ и стабильно увеличивалась с увеличением объема данных наблюдений, таким образом, показав более высокую точность, чем даже односторонний U-критерий Манна — Уитни. При этом, в отличие от последнего, байесовское одностороннее тестирование через вычисление вероятности интервалов апостериорного распределения обладает и методологическими преимуществами, позволяя применять байесовский анализ мощности. Так, рассчитывая вероятности на произвольных отрезках, таких как $[0,2, \infty]$, $[0,5, \infty]$ или $[0,8, \infty]$, можно было бы сделать выводы о статистически значимом размере эффекта.

Таким образом, верификация байесовской порядковой пробит-модели на наборах данных, выраженных в традиционной для социологии пятибалльной шкале Ликерта, демонстрирует высокий потенциал ее использования в сравнительных социологических исследованиях повышенной важности. При этом в исследованиях, связанных с принятием важных управленческих решений или выделением финансирования на значимые социальные проекты, байесовский подход предоставляет

Табл. 1. Результаты сравнения выборок порядковых данных различными методами
 Table 1. Results of comparison of ordinal data samples by different methods

Метод сравнения		Байесовская порядковая пробит-модель				Частотная порядковая пробит-модель	U-критерий Манна — Уитни	Критерий медиан Муда
Библиотека Python		PyMC				statsmodels	SciPy	SciPy
Критерии оценки		Среднее значение коэффициента β	95% ИПАР коэффициента β не включает 0	Вероятность ЗПЭ $[-0,1, 0,1]$, %	Вероятность $\beta > 0$, %	Коэффициент β , р-значение	р-значение	р-значение
Размер выборки	n = 15	0,562	$[-0,120, 1,235]$ (незн.)	6,45 (незн.)	94,82 (незн.)	$\beta = 0,5219$, р = 0,185 (незн.)	0,0997 (незн.)	0,4256 (незн.)
	n = 20	0,549	$[-0,057, 1,171]$ (незн.)	5,68 (незн.)	95,99	$\beta = 0,4947$, р = 0,143 (незн.)	0,0854 (незн.)	0,3428 (незн.)
	n = 25	0,523	$[-0,021, 1,088]$ (незн.)	5,25 (незн.)	96,80	$\beta = 0,4991$, р = 0,098 (незн.)	0,0507 (незн.)	0,2579 (незн.)
	n = 30	0,493	$[-0,024, 0,993]$ (незн.)	5,47 (незн.)	97,15	$\beta = 0,5089$, р = 0,065 (незн.)	0,0424	0,1957 (незн.)
	n = 35	0,518	$[0,039, 0,985]$	3,59	98,30	$\beta = 0,5108$, р = 0,045	0,0272	0,0463
	n = 40	0,498	$[0,049, 0,935]$	3,60	98,58	$\beta = 0,5007$, р = 0,035	0,0290	0,2636 (незн.)
	n = 45	0,479	$[0,058, 0,892]$	3,43	98,81	$\beta = 0,4879$, р = 0,029	0,0154	0,1987 (незн.)
	n = 50	0,492	$[0,094, 0,896]$	2,47	99,15	$\beta = 0,4895$, р = 0,022	0,0058	0,0049
	n = 100	0,517	$[0,231, 0,809]$	0,28	99,97	$\beta = 0,5193$, р = 0,001	0,0001	0,0011

возможность применения концепции байесовского анализа мощности, которой лишены традиционные методы частотной статистики. Однако применение байесовской парадигмы анализа данных требует от исследователя проводить его в границах соответствующей эпистемологии и с учетом особенностей теоретической интерпретации результатов, получаемых с использованием байесовских методов.

Выводы

В рамках раскрытия сущности и особенностей применения в социологии метода байесовской порядковой пробит-регрессии для задачи сравнения данных, выраженных в порядковых шкалах, нами были получены следующие значимые научные результаты.

Байесовская порядковая пробит-модель предоставляет ряд методологических преимуществ в задаче сравнения выборок порядковых данных относительно традиционных методов частотной статистики. Во-первых, этот метод не имеет жестких требований к форме распределения данных и других специфических условий, характерных для параметрических и некоторых непараметрических частотных критериев различий. Во-вторых, теоретическая интерпретация результатов сравнения выборок в контексте порядковой пробит-модели основана на интуитивно понятном линейном сдвиге латентной переменной вместо неявно интерпретируемых различий медиан или стохастического доминирования. Благодаря стандартному нормальному распределению величина сдвига латентной переменной является стандартизированной и ее рассчитанные значения концептуально эквивалентны параметру d Коэна для интервальных данных. В-третьих, применение байесовского подхода позволяет получать статистически значимые заключения о размере эффекта различий между порядковыми выборками, количественная интерпретация которых также может проводиться с использованием соглашений анализа мощности Дж. Коэна без дополнительных поправок и модификаций.

Верификация байесовской порядковой-пробит модели со слабоинформативными априорными пороговыми распределениями на искусственно сгенерированных парах выборок данных, представленных в виде пятибалльной шкалы Ликерта, продемонстрировала ее эмпирическое превосходство над методами частотной статистики, традиционно применяемыми в задаче сравнения порядковых данных. Хотя двусторонние байесовские проверки гипотез о различиях выборок проявили чувствительность, эквивалентную возможностям частотной порядковой пробит-модели, одностороннее байесовское тестирование показало наилучшие результаты, превзойдя метод частотной порядковой пробит-регрессии, односторонний U -критерий Манна — Уитни и критерий медиан Муда. Установленный факт говорит о высоком потенциале применения байесовской порядковой пробит-модели в сравнительных социологических исследованиях, основанных на малых выборках, или исследованиях, сосредоточенных на выявлении слабых различий между группами.

Таким образом, сравнение порядковых социологических данных с использованием метода байесовской порядковой пробит-регрессии видится нами как перспективное направление анализа, рекомендуемое для применения в эмпирической социологии в ситуациях, связанных с принятием значимых управленческих решений, или когда получаемые научные результаты могут оказывать серьезные социальные последствия. Однако практическое внедрение указанного метода в социологическую практику связано с повышенной технологической сложностью его реализации и требует от исследователей дополнительных компетенций в области научного программирования, а также теоретических знаний основ байесовской эпистемологии.

Литература

- Горбунова А.А., Лемешко Б.Ю., Лемешко С.Б. Критерии проверки гипотез об однородности дисперсий при наблюдаемых законах, отличных от нормального // Материалы X международной конференции «Актуальные проблемы электронного приборостроения» (АПЭП-2010) (г. Новосибирск, 22–24 сентября 2010 г.). Т. 6. Новосибирск: [б. и.], 2010. С. 36–41.
- Зангиева И.К., Ротмистров А.Н. Сравнительный анализ способов проведения факторного анализа на порядковых переменных // Мониторинг общественного мнения: экономические и социальные перемены. 2018. № 3. С. 29–46. DOI: 10.14515/monitoring.2018.3.02.
- Звонок А.А. Применение искусственно сгенерированных данных при разработке и внедрении математико-статистических методов в проблемное поле социологии // Известия Саратовского ун-та. Новая серия. Сер.: Социология. Политология. 2024. Т. 24. Вып. 4. С. 389–398. DOI: 10.18500/1818-9601-2024-24-4-389-398.
- Зырянов В.В. Социальная статистика в социологическом образовании // Социологические исследования. 2022. № 2. С. 129–141. DOI: 10.31857/S013216250017138-4.
- Лукашин Ю.П., Рахлина Л.И. Современные направления статистического анализа взаимосвязей и зависимостей. М.: ИМЭМО РАН, 2012. 54 с.
- Светлов В.А. Научный вывод. Байесовская парадигма. М.: Ленанд, 2021. 200 с.
- Шуметов В.Г., Ляскова О.В. О применении количественных методов анализа в прикладных социологических исследованиях: проблемы и пути их решения // Среднерусский вестник общественных наук. 2014. Т. 32. № 2. С. 74–84.
- Agresti A. Analysis of Ordinal Categorical Data. Hoboken: Wiley, 2010. 405 p.
- Cohen J. Statistical Power Analysis for the Behavioral Sciences. New York: Lawrence Erlbaum Associates, 1988. 567 p.
- DeWees T.A., Mazza G.L., Golafshar M.A., Dueck A.C. Investigation Into the Effects of Using Normal Distribution Theory Methodology for Likert Scale Patient-Reported Outcome Data From Varying Underlying Distributions Including Floor / Ceiling Effects // Value Health. 2020. Vol. 23. No. 5. P. 625–631. DOI: 10.1016/j.jval.2020.01.007.
- Gelman A., Simpson D., Betancourt M. The Prior Can Often Only Be Understood in the Context of the Likelihood // Entropy. 2017. Vol. 19. No. 10. P. 555. DOI: 10.3390/e19100555.
- Greene W.H. Econometric Analysis. New York: Pearson, 2018. 1126 p.
- Hart A. Mann-Whitney Test Is Not Just a Test of Medians: Differences in Spread Can Be Important // British Medical Journal. 2001. Vol. 323. P. 391–393. DOI: 10.1136/bmj.323.7309.391.
- Hoffman M.D., Gelman A. The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo // Journal of Machine Learning Research. 2014. Vol. 15. P. 1593–1623.
- Knapp T.R. Treating Ordinal Scales as Interval Scales: An Attempt To Resolve the Controversy // Nursing Research. 1990. Vol. 39. No. 2. P. 121–123. DOI: 10.1097/00006199-199003000-00019.
- Kruschke J.K. Rejecting or Accepting Parameter Values in Bayesian Estimation // Advances in Methods and Practices in Psychological Science. 2018. Vol. 1. P. 270–280. DOI: 10.1177/2515245918771304.
- McElduff F., Cortina-Borja M., Chan S.-K., Wade A. When T-Tests or Wilcoxon-Mann-Whitney Tests Won't Do // Advances in Physiology Education. 2010. Vol. 34. No. 3. P. 128–133. DOI: 10.1152/advan.00017.2010.
- Sawilowsky S.S. New Effect Size Rules of Thumb // Journal of Modern Applied Statistical Methods. 2009. Vol. 8. No. 2. P. 597–599. DOI: 10.22237/jmasm/1257035100.
- Winter J.C.F., Dodou D. Five-Point Likert Items: T Test Versus Mann-Whitney-Wilcoxon // Practical Assessment, Research & Evaluation. 2010. Vol. 15. No. 11. Available at: <https://openpublishing.library.umass.edu/pare/article/1558/galley/1509/view/> (date accessed: 14.08.2025). DOI: 10.7275/bj1p-ts64.
- Zhang B., Zhang Y. Mann-Whitney U Test and Kruskal-Wallis Test Should Be Used for Comparisons of Differences in Medians, Not Means: Comment on the Article by Van der Helm-van Mil et al. // Arthritis & Rheumatology. 2009. Vol. 60. No. 5. P. 1565. DOI: 10.1002/art.24497.

The Problem of Comparing Sociological Ordinal Data: a Bayesian Ordinal Probit Model Solution

ALEKSANDR A. ZVONOK

Lugansk State Pedagogical University,
Lugansk, Russia;
e-mail: al.zvonok@gmail.com
ORCID: 0009-0007-7332-1330

The article explores the essence and specifics of comparing ordinal data in empirical sociology using the Bayesian ordinal probit regression method. It characterizes the main challenges of statistical analysis and theoretical interpretation of its results when working with data measured on ordinal scales: the limited range of mathematical operations applicable to ordinal data and the counterintuitive conclusions derived from commonly used nonparametric statistical methods. The limitations of classical solutions to these problems are outlined, including the application of parametric methods to ordinal data and the deliberate treatment of ordinal scales as interval ones. Within the framework of the study, a simple Bayesian probit model for comparing two samples measured on ordinal scales was constructed using the modern high-level Bayesian analysis tool PyMC. Several methodological advantages of the proposed Bayesian model over traditionally used frequentist methods are considered: the absence of strict requirements regarding data distribution, an intuitively clear theoretical interpretation of comparison results, and the applicability of Bayesian power analysis to measure the statistically significant effect size of differences between the analyzed samples. Using artificially generated datasets represented as 5-point Likert scales, the empirical superiority of the Bayesian ordinal probit model over its frequentist counterpart, as well as over popular nonparametric difference tests, such as the Mann-Whitney U test and Mood's median test, is demonstrated. The findings indicate significant potential for the application of the Bayesian ordinal probit model in comparative sociological research, particularly when sample sizes are too small or the expected effect size is too subtle to be detected by traditional methods.

Keywords: sociological data, ordinal scale, Likert scale, ordered probit model, power analysis, effect size, Bayesian analysis, Bayesian model, PyMC.

References

- Agresti, A. (2010). *Analysis of Ordinal Categorical Data*, Hoboken: Wiley.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*, New York: Lawrence Erlbaum Associates.
- DeWees, T.A., Mazza, G.L., Golafshar, M.A., Dueck, A.C. (2020). Investigation into the Effects of Using Normal Distribution Theory Methodology for Likert Scale Patient-Reported Outcome Data From Varying Underlying Distributions Including Floor / Ceiling Effects, *Value in Health*, 23 (5), 625–631. DOI: 10.1016/j.jval.2020.01.007.
- Gelman, A., Simpson, D., Betancourt, M. (2017). The Prior Can Often Only Be Understood in the Context of the Likelihood, *Entropy*, 19 (10), 555. DOI: 10.3390/e19100555.
- Gorbunova, A.A., Lemesenko, B.Yu., Lemesenko, S.B. (2010). Kriterii proverki gipotez ob odnorodnosti dispersii pri nablyudayemykh zakonakh, otlichnykh ot normal'nogo [Criteria for testing homogeneity of variance hypotheses under observed non-normal distributions], in *Materialy X Mezhdunarodnoy konferentsii "Aktual'nyye problemy elektronnoy priborostroeniya" (APEP-2010)* [Pro-

ceedings of the X International conference “Current problems of electronic instrument engineering” (APEP-2010)], vol. 6 (pp. 36–41), Novosibirsk (in Russian).

Greene, W.H. (2018). *Econometric Analysis*, New York: Pearson.

Hart, A. (2001). Mann-Whitney Test Is Not Just a Test of Medians: Differences in Spread Can Be Important, *British Medical Journal*, no. 323, 391–393. DOI: 10.1136/bmj.323.7309.391.

Hoffman, M.D., Gelman, A. (2014). The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo, *Journal of Machine Learning Research*, no. 15, 1593–1623.

Knapp, T.R. (1990). Treating Ordinal Scales as Interval Scales: An Attempt to Resolve the Controversy, *Nursing Research*, 39 (2), 121–123. DOI: 10.1097/00006199-199003000-00019.

Kruschke, J.K. (2018). Rejecting or Accepting Parameter Values in Bayesian Estimation, *Advances in Methods and Practices in Psychological Science*, no. 1, 270–280. DOI: 10.1177/2515245918771304.

Lukashin, Yu.P., Rakhlina, L.I. (2012). *Sovremennyye napravleniya statisticheskogo analiza vzaimosvyazey i zavisimostey* [Modern trends in statistical analysis of relationships and dependencies], Moskva: IMEMO RAN (in Russian).

McElduff, F., Cortina-Borja, M., Chan, S.-K., Wade, A. (2010). When T-Tests or Wilcoxon-Mann-Whitney Tests Won't Do, *Advances in Physiology Education*, 34 (3), 128–133. DOI: 10.1152/advan.00017.2010.

Sawilowsky, S.S. (2009). New Effect Size Rules of Thumb, *Journal of Modern Applied Statistical Methods*, 8 (2), 597–599. DOI: 10.22237/jmasm/1257035100.

Shumetov, V.G., Lyaskovskaya, O.V. (2014). O primeneniі kolichestvennykh metodov analiza v prikladnykh sotsiologicheskikh issledovaniyakh: problemy i puti ikh resheniya [On the use of quantitative analysis methods in applied sociological research: Problems and solutions], *Srednerusskiy vestnik obshchestvennykh nauk*, 32 (2), 74–84 (in Russian).

Svetlov, V.A. (2021). *Nauchnyy vyvod. Bayesovskaya paradigma* [Scientific inference: Bayesian paradigm], Moskva: Lenand (in Russian).

Winter, J.C.F., Dodou, D. (2010). Five-Point Likert Items: T Test Versus Mann-Whitney-Wilcoxon, *Practical Assessment, Research & Evaluation*, 15 (11). DOI: 10.7275/bj1p-ts64.

Zangieva, I.K., Rotmistrov, A.N. (2018). Sravnitel'nyy analiz sposobov provedeniya faktornogo analiza na poryadkovykh peremennykh [Comparative analysis of factor analysis methods for ordinal variables], *Monitoring obshchestvennogo mneniya: Ekonomicheskiye i sotsial'nyye peremeny*, no. 3, 29–46 (in Russian). DOI: 10.14515/monitoring.2018.3.02.

Zhang, B., Zhang, Y. (2009). Mann-Whitney U Test and Kruskal-Wallis Test Should Be Used for Comparisons of Differences in Medians, Not Means: Comment on the Article by Van der Helm-van Mil et al., *Arthritis & Rheumatology*, 60 (5), 1565. DOI: 10.1002/art.24497.

Zvonok, A.A. (2024). Primeneniye iskusstvenno sgenerirovannykh dannykh pri razrabotke i vnedrenii matematiko-statisticheskikh metodov v problemnoye pole sotsiologii [Application of artificially generated data in development and implementation of mathematical-statistical methods in the problem field of sociology], *Izvestiya Saratovskogo universiteta. Novaya seriya. Ser.: Sotsiologiya. Politologiya*, 24 (4), 389–398 (in Russian). DOI: 10.18500/1818-9601-2024-24-4-389-398.

Zyryanov, V.V. (2022). Sotsial'naya statistika v sotsiologicheskoy obrazovaniy [Social statistics in sociological education], *Sotsiologicheskiye issledovaniya*, no. 2, 129–141 (in Russian). DOI: 10.31857/S013216250017138-4.